Maximum likelihood characterization of rotationally symmetric distributions on the sphere

Mitia Duerinckx[a], Christophe Ley[*,b]

[a] *Université Libre de Bruxelles, Département de Mathématique*
[b] *Université Libre de Bruxelles, E.C.A.R.E.S. and Département de Mathématique*

Abstract

A classical characterization result, which can be traced back to Gauss, states that the maximum likelihood estimator (MLE) of the location parameter equals the sample mean for any possible univariate samples of any possible sizes n if and only if the samples are drawn from a Gaussian population. A similar result, in the two-dimensional case, is given in von Mises (1918) for the Fisher-von Mises-Langevin (FVML) distribution, the equivalent of the Gaussian law on the unit circle. Half a century later, Bingham and Mardia (1975) extend the result to FVML distributions on the unit sphere $\mathcal{S}^{k-1} := \{\mathbf{v} \in \mathbb{R}^k : \mathbf{v}'\mathbf{v} = 1\}$, $k \geq 2$. In this paper, we present a general MLE characterization theorem for a large subclass of rotationally symmetric distributions on $\mathcal{S}^{k-1}$, $k \geq 2$, including the FVML distribution.

AMS (2000) subject classification. Primary 62H05 62E10; secondary 60E05.
Keywords and phrases: Cauchy's functional equation, Characterization theorem, Fisher-von Mises-Langevin distribution, Maximum likelihood estimator, Rotationally symmetric distributions on the sphere.

1. Introduction.

In probability and statistics, a characterization theorem occurs whenever a given law is the only one which satisfies a certain property. Besides their evident mathematical interest *per se*, such characterization results also pave the way for generalizations and extensions which deepen our understanding of the laws under scrutiny (see, e.g., Kagan, Linnik and Rao 1973). A famous example of such a result is due to Gauss (see Gauss 1809), who proved that the normal is the only continuous distribution for which the sample mean is always (that is, for all samples) the maximum likelihood estimator (MLE) of the location parameter. A modern version of this result and an extension to the general k-dimensional setup are given in Azzalini and Genton (2007). This MLE characterization of the Gaussian distribution, which is particularly useful as it means that the intuitively reasonable estimator of the location parameter always coincides with the most efficient one, has motivated researchers to determine in non-linear contexts which distributions enjoy this remarkable property.

An important such non-linear setup are the so-called *spherical distributions*, that is, distributions of random k-vectors taking values on the surface of the unit sphere $\mathcal{S}^{k-1} := \{\mathbf{v} \in \mathbb{R}^k : \mathbf{v}'\mathbf{v} = 1\}$, $k \geq 2$. In general, the spherical distribution of a random unit vector $\mathbf{X}$ depends only on its distance–in a sense to be made precise–from a fixed point $\boldsymbol{\mu} \in \mathcal{S}^{k-1}$. This parameter $\boldsymbol{\mu}$ can be viewed as a "north pole" or "mean direction" for the problem under study, and hence corresponds to the location parameter for spherical distributions. Although this field of research is as old as mathematical

[*] Corresponding author

statistics themselves and involves several problems of different natures (astrophysics, biology, geology, medicine, or meteorology, to cite but a few), its permanent study has only started in the 1950s, due to the pioneering paper Fisher (1953). We refer the reader to Mardia (1972), Mardia (1975a), Watson (1982 and 1983) or Mardia and Jupp (2000) for an overview on the literature of spherical distributions.

One particular spherical distribution has received a lot of attention in the literature: the Fisher-von Mises-Langevin, hereafter abbreviated as FVML, distribution. According to Watson (1983), it is referred to as von Mises (1918) for the two-dimensional, as Fisher (1953) for the three-dimensional, and as Langevin (1905) for the general k-dimensional setup, hence the terminology. Its probability density function (pdf) is given by

$$f_k(\mathbf{x};\boldsymbol{\mu},\kappa) = C_k(\kappa)\exp(\kappa\,\boldsymbol{\mu}'\mathbf{x}), \quad \mathbf{x}\in\mathcal{S}^{k-1},$$

where $\boldsymbol{\mu}\in\mathcal{S}^{k-1}$ is the location parameter and $\kappa>0$ some concentration parameter, and the normalization constant $C_k(\kappa)$ is equal to

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

with $I_{k/2-1}$ the modified Bessel function of the first kind and of order $k/2-1$. The FVML distribution is considered as the spherical analogue of the Gaussian distribution, which explains its central role in the literature.

A reason for this analogy is simple: the FVML is the only spherical distribution for which the (spherical) sample mean is always (that is, for all samples) the maximum likelihood estimator of the location parameter. This property has been established in the two-dimensional case by von Mises (1918), in dimension three by Arnold (1941) and, using a simpler method, by Breitenberger (1963), and finally the result has been proved for any dimension in Bingham and Mardia (1975). In fact, the method developed in Breitenberger (1963) allows the production of MLE characterizations for various three-dimensional spherical distributions, as explained, e.g., in Mardia (1975b). In this paper, our goal is similar: we aim to propose, for *any* dimension, a general MLE characterization theorem, valid for several spherical distributions, including the FVML, within the family of *rotationally symmetric distributions* introduced by Saw (1978) (see Section 2 for a concrete definition). As we shall see, one of the most interesting features of our method is that for the characterization to hold it suffices to have samples of a fixed size n instead of samples of all sizes, as usually required in the literature. In particular, our method thus allows to weaken some conditions of Bingham and Mardia (1975) in case of the FVML distribution, as they need all sample sizes for their main theorem to hold and later on only briefly remark, without a formal proof, that this requirement could be weakened by restricting to sample sizes $n=3k$ and $n=4k$, $k\in\mathbb{N}_0$.

The outline of the paper is as follows. In Section 2, we establish and prove our general characterization result, and in Section 3 we discuss some examples of distributions on the sphere, including of course the FVML, in the light of our main theorem.

2. The general characterization theorem

Let $\mathbf{X}_1,\ldots,\mathbf{X}_n$ be i.i.d. random k-vectors on the unit sphere $\mathcal{S}^{k-1}$. As announced in the Introduction, we suppose that the common distribution of the $\mathbf{X}_i$'s is rotationally symmetric, which means that their common pdf (with respect to the usual surface measure on spheres) is of the form

$$\mathbf{x}\mapsto f_{\boldsymbol{\mu}}(\mathbf{x}) = c_{k,f_1}f_1(\mathbf{x}'\boldsymbol{\mu}), \quad \mathbf{x}\in\mathcal{S}^{k-1},$$

where $\boldsymbol{\mu}\in\mathcal{S}^{k-1}$ is a location parameter and $f_1:[-1,1]\to\mathbb{R}_0^+$ is a continuous and (strictly) monotone increasing function, called *angular function*. This definition reflects the fact that the distribution of each $\mathbf{X}_i$ depends only on the angle between it and the location parameter $\boldsymbol{\mu}$. Aside from the fact that it encompasses many well-known spherical distributions, including the FVML obtained for

$f_1(t) = \exp(\kappa t)$ with $\kappa > 0$, the family of rotationally symmetric distributions satisfies a natural requirement: it is invariant for the actual choice of "north pole". This implies that the family falls within the much more general class of *statistical group models* (see for instance Chang 2004) and thus enjoys all the advantages of this class. This explains why it is interesting to produce a characterization theorem valid for a large subclass (to be defined below) of rotationally symmetric distributions.

The log-likelihood function for $\mathbf{X}_1, \ldots, \mathbf{X}_n$ reads

$$L_{\boldsymbol{\mu}}(\mathbf{X}_1, \ldots, \mathbf{X}_n) = n \log(c_{k,f_1}) + \sum_{i=1}^{n} \log(f_1(\mathbf{X}_i'\boldsymbol{\mu})).$$

In order to maximize this function under the constraint $||\boldsymbol{\mu}|| = 1$ with $||\boldsymbol{\mu}|| = (\boldsymbol{\mu}'\boldsymbol{\mu})^{1/2}$ (recall that we are working on the unit sphere $\mathcal{S}^{k-1}$), we need to introduce a Lagrangian multiplier $\lambda \in \mathbb{R}$, yielding the function $L_{\boldsymbol{\mu}}(\mathbf{X}_1, \ldots, \mathbf{X}_n) + \lambda(1 - ||\boldsymbol{\mu}||)$. Assuming that f_1 is differentiable, the related likelihood equations for $\boldsymbol{\mu}$ and λ correspond to

$$\begin{cases} \sum_{i=1}^{n} \mathbf{X}_i \varphi_{f_1}(\mathbf{X}_i'\boldsymbol{\mu}) = 2\lambda\boldsymbol{\mu} \\ \boldsymbol{\mu}'\boldsymbol{\mu} = 1, \end{cases} \tag{2.1}$$

where $\varphi_{f_1} = f_1'/f_1$ with f_1' the derivative of f_1. Note that, unlike Breitenberger (1963), we directly differentiate w.r.t. the location parameter $\boldsymbol{\mu}$ and not w.r.t. its spherical angles. From (2.1), we deduce the following equation which defines the MLE $\hat{\boldsymbol{\mu}}_{f_1}$ of $\boldsymbol{\mu}$ under the angular density f_1:

$$\hat{\boldsymbol{\mu}}_{f_1} = \frac{\sum_{i=1}^{n} \mathbf{X}_i \varphi_{f_1}(\mathbf{X}_i'\hat{\boldsymbol{\mu}}_{f_1})}{||\sum_{i=1}^{n} \mathbf{X}_i \varphi_{f_1}(\mathbf{X}_i'\hat{\boldsymbol{\mu}}_{f_1})||}.$$

Now, let g_1 be another differentiable angular density, and suppose that $\hat{\boldsymbol{\mu}}_{g_1}$, the MLE of $\boldsymbol{\mu}$ under g_1, coincides with $\hat{\boldsymbol{\mu}}_{f_1}$, yielding the equality

$$\frac{\sum_{i=1}^{n} \mathbf{X}_i \varphi_{f_1}(\mathbf{X}_i'\hat{\boldsymbol{\mu}}_{f_1})}{||\sum_{i=1}^{n} \mathbf{X}_i \varphi_{f_1}(\mathbf{X}_i'\hat{\boldsymbol{\mu}}_{f_1})||} = \hat{\boldsymbol{\mu}}_{f_1} = \hat{\boldsymbol{\mu}}_{g_1} = \frac{\sum_{i=1}^{n} \mathbf{X}_i \varphi_{g_1}(\mathbf{X}_i'\hat{\boldsymbol{\mu}}_{g_1})}{||\sum_{i=1}^{n} \mathbf{X}_i \varphi_{g_1}(\mathbf{X}_i'\hat{\boldsymbol{\mu}}_{g_1})||}, \tag{2.2}$$

where, as above, $\varphi_{g_1} = g_1'/g_1$. In what follows, we aim to prove, under some further mild conditions on f_1 (and g_1), that equation (2.2) holds for any sample $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathcal{S}^{k-1}$ (the $\mathbf{x}_i$'s are realizations of the random unit vectors $\mathbf{X}_i$, $i = 1, \ldots, n$) of fixed size $n \geq 3$ if and only if $\varphi_{g_1} = d\varphi_{f_1}$ for some positive real constant d, as this yields the desired general MLE characterization result (see the paragraph right before Theorem 2.1).

In order to establish this general MLE characterization theorem, we proceed in several steps. We start by fixing the sample size to $n = N$ with $N \geq 3$. Since we are working in a $k(\geq 2)$-dimensional setup and have the full freedom of choice among all samples $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathcal{S}^{k-1}$, we can restrict our attention to samples taking their values on some two-dimensional subspace of $\mathcal{S}^{k-1}$, and hence study equation (2.2) on a circle $\mathcal{C}$. Let $\mathbf{u}_1$ and $\mathbf{u}_2$ denote the orthogonal unit vectors that span $\mathcal{C}$, with

$$\mathbf{u}_1 = \frac{\sum_{i=1}^{N} \mathbf{x}_i \varphi_{f_1}(\mathbf{x}_i'\hat{\boldsymbol{\mu}}_{f_1})}{||\sum_{i=1}^{N} \mathbf{x}_i \varphi_{f_1}(\mathbf{x}_i'\hat{\boldsymbol{\mu}}_{f_1})||}. \tag{2.3}$$

The vectors $\mathbf{x}_1, \ldots, \mathbf{x}_N$ thence can be expressed in terms of N angles $\alpha_1, \ldots, \alpha_N$:

$$\begin{cases} \mathbf{x}_1 = \cos(\alpha_1)\mathbf{u}_1 + \sin(\alpha_1)\mathbf{u}_2 \\ \vdots \\ \mathbf{x}_N = \cos(\alpha_N)\mathbf{u}_1 + \sin(\alpha_N)\mathbf{u}_2. \end{cases}$$

Of course, all N-tuples $(\alpha_1, \ldots, \alpha_N)$ cannot arise in this way: by projecting equation (2.3) onto both $\mathbf{u}_1$ and $\mathbf{u}_2$, these angles satisfy the two conditions

$$\begin{cases} \cos(\alpha_1)\varphi_{f_1}(\cos(\alpha_1)) + \ldots + \cos(\alpha_N)\varphi_{f_1}(\cos(\alpha_N)) > 0 & (2.4) \\ \sin(\alpha_1)\varphi_{f_1}(\cos(\alpha_1)) + \ldots + \sin(\alpha_N)\varphi_{f_1}(\cos(\alpha_N)) = 0, & (2.5) \end{cases}$$

and these conditions are clearly necessary and sufficient for the N-tuple $(\alpha_1, \ldots, \alpha_N)$ to be admissible. With this "angular notation", the starting point of our investigation, namely the equality $\hat{\boldsymbol{\mu}}_{f_1} = \hat{\boldsymbol{\mu}}_{g_1}$, takes on the guise

$$\sin(\alpha_1)\varphi_{g_1}(\cos(\alpha_1)) + \ldots + \sin(\alpha_N)\varphi_{g_1}(\cos(\alpha_N)) = 0. \tag{2.6}$$

So the initial equality (2.2), which needs to hold for all samples $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathcal{S}^{k-1}$, has been turned into equality (2.6) which needs to hold for all samples of angles $\alpha_1, \ldots, \alpha_N$ satisfying conditions (2.4) and (2.5).

Remark 2.1 *Rewriting the functional equation in (2.2) in terms of angles shows that if a solution can be found for $n = N$, then the solution holds for $n = N+1$ by choosing $\mathbf{x}_{N+1}$ such that $\alpha_{N+1} = 0$. In other words, once the minimal value of N for which the problem can be solved is determined, the problem is solvable for all fixed sample sizes $n \geq N$.*

Definition 2.1 *We call* minimal necessary sample size (MNSS) for MLE characterization *the minimal value of N for which the MLE characterization can be proved under a given set of conditions on the angular function f_1.*

Now, define the odd functions $H_{f_1}(t) = t\varphi_{f_1}(\sqrt{1-t^2})$ and $H_{g_1}(t) = t\varphi_{g_1}(\sqrt{1-t^2})$ for $t \in [-1, 1]$. With these definitions in hand, and in view of the conditions (2.4) and (2.5) as well as equation (2.6), we are able to prove an intermediate result, namely that for $N = 3$ there exists a constant $d \in \mathbb{R}_0^+$ such that $\varphi_{g_1}(t) = d\varphi_{f_1}(t)$ for all $t \in [0, 1]$.

Lemma 2.1 *Let f_1 and g_1 be two continuously differentiable angular functions over $[-1, 1]$, and suppose that H_{f_1} is invertible over $[-1, 1]$. Then equation (2.6) holds for all triples of angles $\alpha_1, \alpha_2, \alpha_3$ satisfying (2.4) and (2.5) if and only if there exists a constant $d \in \mathbb{R}_0^+$ such that $\varphi_{g_1}(t) = d\varphi_{f_1}(t)\, \forall t \in [0, 1]$.*

PROOF. The sufficiency part being trivial, we only prove the necessity part of the equivalence. Remember that we can freely choose the observations $\mathbf{x}_1, \mathbf{x}_2$ and $\mathbf{x}_3$, and hence also the associated angles $\alpha_1, \alpha_2, \alpha_3$ provided that conditions (2.4) and (2.5) are satisfied. We consider $\alpha_1, \alpha_2 \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ such that $\sin(\alpha_1)$ and $\sin(\alpha_2)$ are of opposite signs. Therefore, $H_{f_1}(\sin\alpha_1)$ and $H_{f_1}(\sin\alpha_2)$ are of opposite signs too, and their sum is in $H_{f_1}([-1, 1]) = [-H_{f_1}(1), H_{f_1}(1)]$. Hence, the intermediate value theorem ensures that there exists $\alpha_3 \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ such that (2.5) holds. All three angles lying thus in the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$ (but not all three simultaneously at the boundary of this interval–otherwise condition (2.4) could not be satisfied), the positiveness of φ_{f_1} (inherited from the fact that f_1 is monotone increasing) readily allows to see that condition (2.4) is satisfied, too, hence this choice of angles can be used in equation (2.6). Now note that, with our notations and choice of angles, condition (2.5) can be expressed as

$$H_{f_1}(\sin(\alpha_1)) + H_{f_1}(\sin(\alpha_2)) + H_{f_1}(\sin(\alpha_3)) = 0. \tag{2.7}$$

In the same fashion, rewriting equation (2.6) in terms of H_{g_1} yields

$$H_{g_1}(\sin(\alpha_1)) + H_{g_1}(\sin(\alpha_2)) + H_{g_1}(\sin(\alpha_3)) = 0. \tag{2.8}$$

Now, the invertibility of H_{f_1} allows us to deduce from (2.7) that

$$\sin(\alpha_3) = H_{f_1}^{-1}(-H_{f_1}(\sin(\alpha_1)) - H_{f_1}(\sin(\alpha_2))),$$

which, injected into equation (2.8) and bearing in mind that H_{f_1} and H_{g_1} are odd, leads to

$$\underbrace{H_{g_1} \circ H_{f_1}^{-1}}_{=:H}(H_{f_1}(\sin(\alpha_1)) + H_{f_1}(\sin(\alpha_2))) = \underbrace{H_{g_1} \circ H_{f_1}^{-1}}_{=:H}(H_{f_1}(\sin(\alpha_1))) + \underbrace{H_{g_1} \circ H_{f_1}^{-1}}_{=:H}(H_{f_1}(\sin(\alpha_2))).$$

Let us suppose, w.l.o.g., that $\alpha_1 \in [0, \frac{\pi}{2}]$ and $\alpha_2 \in [-\frac{\pi}{2}, 0]$ (the inverse situation would conduct to exactly the same result). Setting $a = H_{f_1}(\sin(\alpha_1))$ and $b = -H_{f_1}(\sin(\alpha_2))$, the latter equation finally can be written as

$$H(a-b) = H(a) - H(b) \quad \forall a, b \in [0, H_{f_1}(1)].$$

One easily recognizes that this equation is a particular form of the celebrated Cauchy functional equation which has been extensively discussed in the mathematical literature; see, for instance, Aczél and Dhombres (1989) under the assumption of continuity at a point for H. The latter function being continuous by hypothesis (continuous differentiability of f_1 and g_1), a similar proof allows to conclude here that $H(s) = ds \, \forall s \in [0, H_{f_1}(1)]$ for some real constant d, hence that $H(s) = ds \, \forall s \in [-H_{f_1}(1), H_{f_1}(1)]$ since H is an odd function.

We thus have, for $s = H_{f_1}(t)$, that $H_{g_1}(t) = dH_{f_1}(t) \, \forall t \in [-1,1]$. Finally, by definition of H_{f_1} and H_{g_1}, this yields $\varphi_{g_1}(t) = d\varphi_{f_1}(t) \, \forall t \in [0,1]$ with $d > 0$. This concludes the necessity part, and consequently the claim holds. □

In view of Remark 2.1, the result of Lemma 2.1 is valid for any sample size $N \geq 3$. Note that, in case φ_{f_1} is even (implying that φ_{g_1} is even via (2.5) and (2.6)), this lemma clearly entails that $\varphi_{g_1}(t) = d\varphi_{f_1}(t)$ over the complete interval $[-1,1]$, hence yields the desired result. For example, for $f_1(t) = \exp(\kappa t)$ with $\kappa > 0$, the angular function of the FVML distribution, $\varphi_{f_1}(t) = \kappa$, hence Lemma 2.1 is sufficient for the MLE characterization of the FVML in any dimension (see Theorem 2.1 below).

If no parity assumptions are made on φ_{f_1}, Lemma 2.1 is evidently not sufficient to obtain the desired general characterization theorem. Actually, we have to strengthen it into Lemma 2.2 below, for which we need to define the following two odd functions: $\tilde{H}_{f_1}(t) = t\varphi_{f_1}(-\sqrt{1-t^2})$ and $\tilde{H}_{g_1}(t) = t\varphi_{g_1}(-\sqrt{1-t^2})$ for $t \in [-1,1]$. These functions will allow us to obtain constraints on φ_{f_1} and φ_{g_1} over $[-1,0]$, which constitutes the missing piece in our characterization puzzle. At first sight, this seems quite easy: choosing $\alpha_1 \in [\frac{\pi}{2}, \pi]$ and $\alpha_2, \alpha_3 \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ turns condition (2.5) into

$$\tilde{H}_{f_1}(\sin(\alpha_1)) + H_{f_1}(\sin(\alpha_2)) + H_{f_1}(\sin(\alpha_3)) = 0 \tag{2.9}$$

instead of (2.7) and equation (2.6) into

$$\tilde{H}_{g_1}(\sin(\alpha_1)) + H_{g_1}(\sin(\alpha_2)) + H_{g_1}(\sin(\alpha_3)) = 0 \tag{2.10}$$

rather than into (2.8). By Lemma 2.1, we know that $H_{g_1}(t) = dH_{f_1}(t) \, \forall t \in [-1,1]$; (2.9) and (2.10) then readily show that $\tilde{H}_{g_1}(\sin(\alpha_1)) = d\tilde{H}_{f_1}(\sin(\alpha_1))$, which will allow to achieve our goal. However, an important factor is missing here: actually, we do not have this equality for all values of $\alpha_1 \in [\frac{\pi}{2}, \pi]$, and consequently our constraints on φ_{f_1} and φ_{g_1} will not cover the entire interval $[-1,0]$. One can easily imagine a situation in which $\alpha_1 \in [\frac{\pi}{2}, \pi]$ and $\alpha_2 \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ but where α_3 rather needs to lie inside $[\frac{\pi}{2}, \frac{3\pi}{2}]$ in order to fulfill both conditions (2.4) and (2.5) (think of an angular function f_1 for which φ_{f_1} takes high values on $[-1, -1+\epsilon]$ and low values on $[1-\epsilon, 1]$ for some small $\epsilon > 0$; values of α_1 near $\frac{\pi}{2}$ could no more satisfy condition (2.5)). This choice of α_3 requires using $\tilde{H}_{f_1}$ and $\tilde{H}_{g_1}$ and hence prevents from writing (2.9) and (2.10). Thus we need to tackle the problem from a slightly different angle. One possibility would consist in adding conditions on f_1 via $\tilde{H}_{f_1}$, but these might be difficult to check in practice and could rule out certain rotationally symmetric distributions. An alternative solution, which we shall use here, rests in simply considering a sample size N greater than 3.

Lemma 2.2 *Let f_1 and g_1 be two continuously differentiable angular functions over $[-1,1]$, and suppose that H_{f_1} is invertible over $[-1,1]$. Fix $N = 3 + \lceil (H_{f_1}(1))^{-1} \max_{t\in[0,1]} \tilde{H}_{f_1}(t) - 1 \rceil$. Then equation (2.6) holds for all N-tuples of angles $\alpha_1, \ldots, \alpha_N$ satisfying (2.4) and (2.5) if and only if there exists a constant $d \in \mathbb{R}_0^+$ such that $\varphi_{g_1}(t) = d\varphi_{f_1}(t) \, \forall t \in [-1,1]$.*

Proof. Again the sufficiency part of the equivalence is trivial and we concentrate our attention on the necessity part. As stated in Remark 2.1, we can choose all $\alpha_i = 0$ for $i > 3$ in order to be left with only the three angles α_1, α_2 and α_3. Lemma 2.1 then tells us that there exists a positive constant d such that $\varphi_{g_1}(t) = d\varphi_{f_1}(t)\,\forall t \in [0,1]$. It remains to obtain the same result over $[-1,0]$. To do so, fix $\alpha_1 \in [\frac{\pi}{2}, \pi]$. For $k = \lceil (H_{f_1}(1))^{-1}\tilde{H}_{f_1}(\sin(\alpha_1)) - 1\rceil \leq N-3$, choose $\alpha_2 = \cdots = \alpha_{k+1} = -\frac{\pi}{2}$. We then have that

$$\sin(\alpha_1)\varphi_{f_1}(\cos(\alpha_1)) + \ldots + \sin(\alpha_{k+1})\varphi_{f_1}(\cos(\alpha_{k+1})) = \tilde{H}_{f_1}(\sin(\alpha_1)) - kH_{f_1}(1) \in [0, H_{f_1}(1)).$$

The intermediate value theorem ensures that we can find $\alpha_{k+2} \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ such that $\sin(\alpha_1)\varphi_{f_1}(\cos(\alpha_1)) + \ldots + \sin(\alpha_{k+2})\varphi_{f_1}(\cos(\alpha_{k+2})) = 0$. Finally, take $\alpha_{k+3} = \cdots = \alpha_N = 0$: condition (2.5) then holds trivially. Moreover, since $k+3 \leq N$, it is clear that

$$\cos(\alpha_1) + \ldots + \cos(\alpha_N) \geq \cos(\alpha_1) + \cos(\alpha_N) = \cos(\alpha_1) + 1 > 0,$$

and hence, in view of $\varphi_{f_1} > 0$, condition (2.4) is satisfied too by this choice of angles. Now, rewriting condition (2.5) and equality (2.6) in terms of $H_{f_1}, H_{g_1}, \tilde{H}_{f_1}$ and $\tilde{H}_{g_1}$ respectively leads to

$$\tilde{H}_{f_1}(\sin(\alpha_1)) - kH_{f_1}(1) + H_{f_1}(\sin(\alpha_{k+2})) = 0$$

and

$$\tilde{H}_{g_1}(\sin(\alpha_1)) - kH_{g_1}(1) + H_{g_1}(\sin(\alpha_{k+2})) = 0.$$

Since $\varphi_{g_1} = d\varphi_{f_1}$ over $[0,1]$, and hence $H_{g_1} = dH_{f_1}$ over $[-1,1]$, it follows that $\tilde{H}_{g_1}(\sin(\alpha_1)) = d\tilde{H}_{f_1}(\sin(\alpha_1))\,\forall \alpha_1 \in [\frac{\pi}{2}, \pi]$, and similar manipulations as in the proof of Lemma 2.1 reveal that $\varphi_{g_1}(t) = d\varphi_{f_1}(t)\,\forall t \in [-1,0]$, which concludes the proof. □

The statement of this lemma begs for a comment regarding the sample size N. Actually, whenever $\max_{t\in[0,1]} \tilde{H}_{f_1}(t) \leq H_{f_1}(1)$, N equals 3 and the requirement of Lemma 2.2 is not stronger than the one of Lemma 2.1. One is tempted to say that then we have a problem, as discussed just before Lemma 2.2. However, as stated there, the fact that $N=3$ is in general not sufficient to obtain a result over the whole interval $[-1,1]$ is due to a lack of knowledge about φ_{f_1}, which might imply that $\tilde{H}_{f_1}$ takes much higher values than H_{f_1}. This problematic constellation is precisely ruled out here, as we know that $\tilde{H}_{f_1}(t) \leq H_{f_1}(1)$ over $[0,1]$. Therefore, Lemma 2.2 covers all possible cases. One might then wonder why we have introduced Lemma 2.1, as the more general Lemma 2.2 supersedes it. The reasons are mainly twofold: on the one hand, our present approach is more constructive and hence easier to follow, and, on the other hand, Lemma 2.1 with its weaker conditions contains several interesting rotationally symmetric distributions (such as, e.g., the FVML) and therefore is interesting *per se*.

Now, Lemma 2.1 and Lemma 2.2, under the same conditions on f_1 but for different minimal sample sizes N, yield the desired result, namely that (2.2) entails that there exists a positive constant d such that $\varphi_{g_1}(t) = d\varphi_{f_1}(t)\,\forall t \in [-1,1]$. Solving this first-order differential equation is easy and leads to $g_1(t) = c(f_1(t))^d$, where $c > 0$ is a normalizing constant. Thus, up to the exponent d and the corresponding normalizing constant c we retrieve the original angular function f_1, which allows us to state our general MLE characterization theorem.

Theorem 2.1 (MLE characterization theorem of rotationally symmetric distributions) *Let f_1 and g_1 be two continuously differentiable angular functions over $[-1,1]$ associated with rotationally symmetric distributions over $\mathcal{S}^{k-1}$, $k \geq 2$. Suppose that H_{f_1} is invertible over $[-1,1]$.*

- *(i) Fix $N = 3 + \lceil (H_{f_1}(1))^{-1}\max_{t\in[0,1]} \tilde{H}_{f_1}(t) - 1\rceil$. Then $\hat{\boldsymbol{\mu}}_{f_1} = \hat{\boldsymbol{\mu}}_{g_1}$ for all samples of fixed sample size $n \geq N$ if and only if there exist constants $c, d \in \mathbb{R}_0^+$ such that $g_1(t) = c(f_1(t))^d\,\forall t \in [-1,1]$.*
- *(ii) Fix $N = 3$ and suppose that f_1 is such that φ_{f_1} is even. Then $\hat{\boldsymbol{\mu}}_{f_1} = \hat{\boldsymbol{\mu}}_{g_1}$ for all samples of fixed sample size $n \geq N$ if and only if there exist constants $c, d \in \mathbb{R}_0^+$ such that $g_1(t) = c(f_1(t))^d\,\forall t \in [-1,1]$.*

Note that, when $N = 3$, we have reached the MNSS for MLE characterization. In all other cases, the value of N constitutes a good upper bound on the MNSS, which seems to be difficult to obtain in general. It is however interesting to remark that the MNSS, in view of our upper bounds, will always be finite, contrarily to the Euclidean case (see Duerinckx *et al.* 2012).

Finally note that, for (strictly) monotone decreasing angular functions f_1 and g_1, all the above results can be retrieved up to some obvious modifications.

3. Examples

In this final section, we shall discuss some examples of rotationally symmetric distributions in the light of our findings and show whether they are MLE-characterizable or not. These examples include the *FVML distribution*, the *median distribution*, the *linear, logarithmic and logistic spherical distributions* (used in Ley et al. 2013), the *wrapped normal distribution* (extensively described in Mardia 1972) and a family of symmetric distributions on the unit circle (introduced in Jones and Pewsey 2005). All these distributions possess continuously differentiable angular functions.

Let us start with the FVML distribution, whose angular function is of the form $f_1(t) = e^{\kappa t}$ with $\kappa > 0$. One immediately sees that $\varphi_{f_1}(t) = \kappa$ and $H_{f_1}(t) = \kappa t$, and hence all conditions for Theorem 2.1(ii) are fulfilled, showing that the FVML can be characterized by its MLE, which here coincides with the standardized sample mean $\sum_{i=1}^n \mathbf{x}_i / ||\sum_{i=1}^n \mathbf{x}_i||$. Since our result holds true for any fixed sample size $n \geq 3$ (and does not necessitate any other sample size), we generalize the result of Bingham and Mardia (1975), as already announced in the Introduction.

Next, we consider what we call the "median distribution", whose angular function reads $f_1(t) = ce^{-a \arccos t}$ with $a > 0$ and c a normalizing constant. Evident computations show that $\varphi_{f_1}(t) = \frac{a}{\sqrt{1-t^2}}$ and $H_{f_1}(t) = a\,\mathrm{sign}(t)$. The latter function is not invertible over $[-1, 1]$, consequently our MLE characterization theorem does not apply in this particular case. However, and this explains our terminology, Purkayastha (1991) has established in dimensions 2 and 3 that this rotationally symmetric distribution can be characterized by its MLE, which coincides with the spherical median defined in Fisher (1985). For the method and the related assumptions and conditions, see Purkayastha (1991).

Thirdly, we investigate the case of the linear spherical distribution, with angular function $f_1(t) = t + a$, $a > 1$. It follows that $\varphi_{f_1}(t) = \frac{1}{t+a}$ and $H_{f_1}(t) = \frac{t}{a+\sqrt{1-t^2}}$. One notices that φ_{f_1} is not even, and since the derivative

$$\frac{\mathrm{d}}{\mathrm{d}t} H_{f_1}(t) = \frac{a + \frac{1}{\sqrt{1-t^2}}}{(a + \sqrt{1-t^2})^2} > 0 \quad \forall t \in [-1, 1],$$

this example falls into the category of Theorem 2.1(i), and hence is MLE-characterizable.

Fourthly, we analyze the logarithmic spherical distribution with associated angular function $f_1(t) = \log(t+a)$, $a > 2$. In this case, we have $\varphi_{f_1}(t) = \frac{1}{(t+a)\log(t+a)}$ and $H_{f_1}(t) = \frac{t}{(a+\sqrt{1-t^2})\log(a+\sqrt{1-t^2})}$. Again, φ_{f_1} is not even, and straightforward calculations reveal that

$$\frac{\mathrm{d}}{\mathrm{d}t} H_{f_1}(t) = \frac{a \log(a + \sqrt{1-t^2}) + \frac{1}{\sqrt{1-t^2}} \log(a + \sqrt{1-t^2}) + \frac{t^2}{\sqrt{1-t^2}}}{(a + \sqrt{1-t^2})^2 (\log(a + \sqrt{1-t^2}))^2} > 0 \quad \forall t \in [-1, 1].$$

Theorem 2.1(i) thus applies to this example as well.

Fifthly and lastly for the set of spherical distributions described in Ley et al. (2013), we consider the logistic spherical distribution with angular function $f_1(t) = \frac{a \exp(-b \arccos(t))}{(1+a \exp(-b \arccos(t)))^2}$, where a and b are chosen in such a way that f_1 satisfies the conditions of an angular function. Calculations show that $\varphi_{f_1}(t) = \frac{b(1-a \exp(-b \arccos t))}{(1+a \exp(-b \arccos t))\sqrt{1-t^2}}$ and $H_{f_1}(t) = \frac{b(1-a \exp(b \arccos(\sqrt{1-t^2})))}{(1+a \exp(b \arccos(\sqrt{1-t^2})))}\,\mathrm{sign}(t)$. Once more, φ_{f_1} does not happen to be even, and the derivative

$$\frac{\mathrm{d}}{\mathrm{d}t} H_{f_1}(t) = \frac{2a^2 b^2}{\sqrt{1-t^2}} \frac{\exp(-b \arccos \sqrt{1-t^2})}{(1 + a \exp(-b \arccos \sqrt{1-t^2}))^2}$$

is strictly positive over $[-1,1]$, hence the logistic spherical distributions satisfy the conditions of Theorem 2.1(i).

Sixthly, we analyze the wrapped normal distribution, whose angular function reads $f_1(t) = \frac{1}{2\pi}\vartheta_3(\arccos(t)/2, e^{-\sigma^2/2})$ in terms of the Jacobi theta function $\vartheta_3(z,q) := \sum_{n=-\infty}^{\infty} q^{n^2} e^{2niz}$. Straightforward calculations yield $\varphi_{f_1}(t) = -\frac{1}{2}\frac{1}{\sqrt{1-t^2}}\frac{\vartheta_3'(\frac{1}{2}\arccos(t), e^{-\sigma^2/2})}{\vartheta_3(\frac{1}{2}\arccos(t), e^{-\sigma^2/2})}$, where ϑ_3' denotes the derivative of ϑ_3 w.r.t. its first argument. Using the identity $\frac{\vartheta_3'(z,q)}{\vartheta_3(z,q)} = -4\sum_{n=1}^{\infty}\frac{q^{2n-1}}{2q^{2n-1}}\frac{\sin(2z)}{\cos(2z)+q^{4n-2}+1}$ (see Whittaker and Watson 1990, p. 489), our expression simplifies to (the non-even function)

$$\varphi_{f_1}(t) = 2\sum_{n=1}^{\infty}\frac{\left(e^{-\sigma^2/2}\right)^{2n-1}}{2t\left(e^{-\sigma^2/2}\right)^{2n-1} + \left(e^{-\sigma^2}\right)^{2n-1} + 1} = 2\sum_{n=1}^{\infty}\frac{1}{2t + \left(e^{-\sigma^2/2}\right)^{2n-1} + \left(e^{\sigma^2/2}\right)^{2n-1}}.$$

This series is uniformly convergent and defines a positive function over $[-1,1]$. We then have

$$H_{f_1}(t) = 2\sum_{n=1}^{\infty}\frac{t}{2\sqrt{1-t^2} + \left(e^{-\sigma^2/2}\right)^{2n-1} + \left(e^{\sigma^2/2}\right)^{2n-1}}$$

which uniformly converges over $[-1,1]$. The series of the derivatives w.r.t. t writes

$$2\sum_{n=1}^{\infty}\frac{\frac{2}{\sqrt{1-t^2}} + \left(e^{-\sigma^2/2}\right)^{2n-1} + \left(e^{\sigma^2/2}\right)^{2n-1}}{\left(2\sqrt{1-t^2} + \left(e^{-\sigma^2/2}\right)^{2n-1} + \left(e^{\sigma^2/2}\right)^{2n-1}\right)^2}$$

which uniformly converges over every compact of $(-1,1)$; hence the differentiation under the summation sign is allowed. Since this derivative is strictly positive over $(-1,1)$, H_{f_1} is invertible, and Theorem 2.1(i) thus applies to this example too.

Finally, we consider the family of symmetric distributions on the unit circle introduced in Jones and Pewsey (2005), with angular functions of the form $f_1(t) = \gamma(\kappa,\psi)(1+\tanh(\kappa\psi)t)^{1/\psi}$, where $\kappa \geq 0$, $\psi \in \mathbb{R}$ and where γ is a normalization constant. Fixing $\psi = 1$, -1 or 0 (by continuity) yields respectively the cardioid, the wrapped Cauchy or the von Mises distribution. Straightforward calculations give $\varphi_{f_1}(t) = \frac{\tanh(\kappa\psi)}{\psi(1+\tanh(\kappa\psi)t)}$ and $H_{f_1}(t) = \frac{\tanh(\kappa\psi)t}{\psi(1+\tanh(\kappa\psi)\sqrt{1-t^2})}$. Similar manipulations as before yield

$$\frac{\mathrm{d}}{\mathrm{d}t}H_{f_1}(t) = \frac{\tanh(\kappa\psi)}{\psi}\frac{(1+\frac{\tanh(\kappa\psi)}{\sqrt{1-t^2}})}{(1+\tanh(\kappa\psi)\sqrt{1-t^2})^2},$$

which is strictly positive over $[-1,1]$ provided that $\psi \geq 0$. Our characterization theorem thus applies to this example whenever $\psi \geq 0$. Note that, for $\psi > 0$ (e.g., for the cardioid distribution), φ_{f_1} is not even, hence Theorem 2.1(i) has to be used, whereas, for $\psi = 0$ corresponding to the FVML case, we retrieve exactly the results of our first example in this section. Finally, for $\psi < 0$ (e.g., for the wrapped Cauchy distribution), our result does not apply.

Acknowledgements

Christophe Ley thanks the Fonds National de la Recherche Scientifique, Communauté française de Belgique, for support via a Mandat de Chargé de Recherche.

References

[1] Aczél, J. and Dhombres, J. (1989) *Functional equations in several variables with applications to mathematics, information theory and to the natural and social sciences.* Encyclopedia Math. Appl. **31**. Cambridge: Cambridge University Press.

[2] Arnold, K.J. (1941) *On spherical probability distributions.* Ph.D. thesis Massachusetts Institute of Technology.

[3] Azzalini, A. and Genton, M.G. (2007) On Gauss's characterization of the normal distribution. *Bernoulli* **13**, 169–174.

[4] Bingham, M.S. and Mardia, K.V. (1975) Characterizations and applications. In *Statistical Distributions for Scientific Work, Vol. 3, ed. G.P. Patil, S. Kotz and J.K. Ord, Reidel, Dordrecht and Boston, 387–398.*

[5] Breitenberger, E. (1963) Analogues of the normal distribution on the circle and the sphere. *Biometrika* **50**, 81–88.

[6] Chang, T. (2004) Spatial statistics. *Stat. Sci.* **19**, 624–635.

[7] Duerinckx, M., Ley, C. and Swan, Y. (2012). Maximum likelihood characterization of distributions. arXiv:1207.2805.

[8] Fisher, R.A. (1953) Dispersion on a sphere. *Proc. Roy. Soc. London ser. A* **217**, 295–305.

[9] Fisher, N.I. (1985) Spherical medians. *J. Roy. Stat. Soc. B* **47**, 342–348.

[10] Gauss, C.F. (1809) *Theoria motus corporum coelestium in sectionibus conicis solem ambientium.* Hamburg: Perthes et Besser. English translation by C.H. Davis, reprinted by Dover, New York (1963).

[11] Jones, M.C. and Pewsey, A. (2005) A family of symmetric distributions on the circle. *J. Am. Stat. Assoc.* **100**, 1422–1428.

[12] Kagan, A.M., Linnik, Y.V. and Rao, C.R. (1973) *Characterization problems in mathematical statistics.* Wiley, New York.

[13] Langevin, P. (1905) Sur la théorie du magnétisme. *J. Phys.* **4**, 678–693; Magnétisme et théorie des électrons. *Ann. Chim. Phys.* **5**, 70–127.

[14] Ley, C., Swan, Y., Thiam, B. and Verdebout, T. (2013) Optimal R-estimation of a spherical location. *Stat. Sinica*, **23**, to appear..

[15] Mardia, K.V. (1972) *Statistics of directional data.* Academic Press, London.

[16] Mardia, K.V. (1975a) Statistics of directional data. *J. Roy. Stat. Soc. B* **37**, 349–393.

[17] Mardia, K.V. (1975b) Characterization of directional distributions. In *Statistical Distributions for Scientific Work, Vol. 3, ed. G.P. Patil, S. Kotz and J.K. Ord, Reidel, Dordrecht and Boston, 365–385.*

[18] Mardia, K.V. and Jupp, P.E. (2000) *Directional statistics.* Wiley, Chichester.

[19] Purkayastha, S. (1991) A rotationally symmetric directional distribution: obtained through maximum likelihood characterization. *Sankhyā Ser. A* **53**, 70–83.

[20] Saw, J.G. (1978) A family of distributions on the m-sphere and some hypothesis tests. *Biometrika* **65**, 69–73.

[21] Von Mises, R. (1918) Uber die Ganzzahligkeit der Atomgewichte und verwandte Fragen. *Physikalische Zeitschrift* **19**, 490–500.

[22] Watson, G.S. (1982) Distributions on the circle and sphere. *J. Appl. Probab.* **19**, 265–280.

[23] Watson, G.S. (1983) *Statistics on spheres.* Wiley, New York.

[24] Whittaker, E.T. and Watson, G.N. (1990) *A Course in Modern Analysis* 4th ed., Cambridge University Press.

Mitia Duerinckx
Département de Mathématique,
Université Libre de Bruxelles
Campus de la Plaine, boulevard du triomphe, CP 210
B-1050 Bruxelles Belgium
mduerinc@ulb.ac.be

Christophe Ley
E.C.A.R.E.S. and
Département de Mathématique,
Université Libre de Bruxelles
Campus de la Plaine, boulevard du triomphe, CP 210
B-1050 Bruxelles Belgium
chrisley@ulb.ac.be