Hacettepe Journal of Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 50 (6) (2021), 1838 – 1854
DOI : 10.15672/hujms.788296

Research Article

A new extension of von Mises-Fisher distribution

Meisam Moghimbeygi[1], Mousa Golalizadeh*[2]

[1] *Department of Mathematics, Faculty of Mathematics and Computer Science, Kharazmi University, Tehran, Iran*

[2] *Department of Statistics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran*

Abstract

Spherical distributions, including the von Mises-Fisher density, have received a great attention in the literature because of their usefulness to model circular data lying on the unit sphere. However, there is a paucity of research on proposing spherical densities possing multimode tuned with a single parameter. To fill in this gap, we extend von Mises-Fisher distribution to construct a new density. Moreover, some of the important statistical properties of the proposed distribution including the estimation of parameters are highlighted. To evaluate the performance of the proposed distribution, some simulation studies and analyzing three real-life examples are presented.

Mathematics Subject Classification (2020). 62Exx, 62H11, 62E15

Keywords. Directional statistics, von Mises-Fisher distribution, spherical data, multimodality, Bessel function

1. Introduction

There are many unimodal probability densities that describe the random behavior of some phenomena. However, fitting unimodal densities might not produce satisfactory models when there is an indication that the histogram of the directional data looks multimodal. The bimodal and multimodal distributions might, instead, show up a better fitting in these cases. So far there have been many new developments in the literature regarding multimodal data fitting in the Euclidean space but the same is not true for the random variables whose domain is in the non-Euclidean space. The statistics dealing with this topic are called non-linear statistics, and the corresponding data are referred to as the manifold-valued data. A well-known field of non-linear statistics is called directional statistics which consists of circular and spherical statistics as described by [15]

The von Mises-Fisher (vMF) distribution is a well-known density for modeling data lying on the surface of the unit sphere. Following [15], the probability density function

*Corresponding Author.
Email addresses: m.moghimbeygi@khu.ac.ir (M. Moghimbeygi), golalizadeh@modares.ac.ir (M. Golalizadeh)
Received: 31.08.2020; Accepted: 29.07.2021

($p.d.f.$) of the vMF distribution for a random variable X, taking its value on the surface of a general unit sphere in $\mathbb{R}^d$, i.e. $\mathbb{S}^{d-1}$, is given by

$$f(x;\mu,\kappa)=\frac{\kappa^{d/2-1}}{(2\pi)^{d/2}I_{d/2-1}(\kappa)}e^{\kappa \mathrm{x}^T\mu}, \tag{1.1}$$

where μ is the mean direction with $||\mu||=1$, $\kappa \geq 0$ is a concentration parameter and $I_\nu(.)$ denotes the modified Bessel function of the first kind and order ν [3]. This distribution is known as von Mises and Fisher for the case in which the points lying on the unit circle ($\mathbb{S}^1$) and the unit sphere ($\mathbb{S}^2$), respectively. It was of interest to extend von Mises distributions over ten years or so. For example, Gatto and Jammalamadaka [9] studied unimodal and multimodal von Mises distribution and provided some important properties of this density. Umbach and Jammalamadaka [19] used weight function to introduce asymmetry in circular distribution. Abe and Pewsey [1] presented a class of symmetric circular distribution using duplication and cosine perturbation. They also studied the skew circular model using sine function, reported in [2]. A set of unimodal and multimodal circular distributions along with model fitting programs can be found in [16]. By generalizing the Fisher-Bingham and vMF distributions, Kim et al. [14] introduced two kinds of distributions on the sphere. Their distributions are multi-unimodal distributions that are located on one or several small-subsphere with mode(s).

Another version of circular density with a multimode feature is formed by changing the period of trigonometric functions. For instance, Yfantis and Borgman [23] presented four-parameters bimodal distribution for circular data. Also, Kato and Jones [11] introduced a four-parameter extended family of densities related to the wrapped Cauchy distribution. It is worth mentioning that most of the multimodal densities were obtained by extending von Mises distribution. Watson [20] surveyed the relationships between the different distributions on the circle and the sphere. He also compared vMF, Brownian, and angular Gaussian distributions in the arithmetical and graphical methods.

Research on modeling directional data was not just restricted to modeling data lying on the circumference of the unit circle. Many research has been done on developing the models on the surface of the unit sphere. One can consult, for example, Fisher et al. [8] for a comprehensive treatment of spherical data and related topics. Unlike multimodal distributions, there are many literature dealing with various unimodal densities on the unit sphere. An exceptional example is the Kent [13] distribution where the multimodality occurs by setting the parameters of the density to some specific values. As expected, the research on proposing multimodal distributions on a high dimensional sphere (S^{d-1}) is few. Wood [21] modified the Fisher distribution by doubling the longitude to model bimodal data. Also, particular value of a parameter in the Bingham distributions, useful to model the axial data, leads to bimodal distributions (see, e.g. [15], p. 181).

We are interested in constructing bimodal density based on the vMF distribution. In particular, we aim to introduce a bimodal density to model spherical data having two modes. To this end, we use Azzalini [5]'s method to propose a non-symmetric distribution by a tractable symmetric distribution. Mathematically, he suggested to construct a non-symmetric distribution f using a symmetric distribution g and an absolutely continuous distribution function H, such that H' is symmetric about 0, through invoking the expression

$$f(\mathrm{x})=2H(\lambda\mathrm{x})g(\mathrm{x}), \tag{1.2}$$

where λ is any real-valued parameter.

The remainder of this paper is organized as follows. In Section 2, we introduce extended von Mises-Fisher (EvMF) distribution and highlight some of its important properties. It also includes a procedure to generate data from this density. Section 3 gives statistical inference on the parameters of the EvMF. Simulation studies along with analyzing real-life data are provided in Section 4. The paper ends with a general conclusion and more technical proof of some mathematical results.

2. Extended von Mises-Fisher distribution

In this section, we are going to extend the vMF distribution on $\mathbb{S}^{d-1}$ using the expression in Equation (1.2). Let us assume that X is a random vector on $\mathbb{S}^{d-1}$ with the mean direction μ and the concentration parameter κ. We can construct a new density function via considering f as

$$f(\mathrm{x}) \propto 2(1-\lambda \mathrm{x}^T\mu)g(\mathrm{x}),$$

where $g(\mathrm{x})$ is the density function of vMF distribution with mean direction μ and concentration parameter κ given by Equation (1.1). We call this new distribution as EvMF in the text and mathematically as EvMF (μ, κ). It has the following *p.d.f*

$$f(x;\mu,\kappa,\lambda) = C_d(\kappa,\lambda)(1-\lambda \mathrm{x}^T\mu)e^{\kappa \mathrm{x}^T\mu}, \tag{2.1}$$

where $|\lambda| \leq 1$ is a skew control parameter and $C_d(\kappa,\lambda)$ is a normalized constant. It is straightforward to show that the normalizing constant in Equation (2.1) is given by

$$C_d(\kappa,\lambda) = \frac{\kappa^{d/2-1}}{(2\pi)^{d/2}(I_{\frac{d}{2}-1}(\kappa)-\lambda I_{\frac{d}{2}}(\kappa))}.$$

Following [15, p.179], it is seen that our proposed distribution in Equation (2.1) is rotationally symmetric.

It is of interest to derive the density given by Equation (2.1) in terms of the spherical polar coordinates [15]. To achieve this, let us write X and μ, as

$$X = (\cos\theta_1, \sin\theta_1\cos\theta_2, ..., \sin\theta_1\cdots\sin\theta_{d-2}\cos\theta_{d-1}, \sin\theta_1\cdots\sin\theta_{d-2}\sin\theta_{d-1})^T$$

and

$$\mu = (\cos\alpha_1, \sin\alpha_1\cos\alpha_2, ..., \sin\alpha_1\cdots\sin\alpha_{d-2}\cos\alpha_{d-1}, \sin\alpha_1\cdots\sin\alpha_{d-2}\sin\alpha_{d-1})^T,$$

where

$$\sin\theta_0 = \sin\alpha_0 = \cos\theta_d = \cos\alpha_d = 1.$$

Then, the probability element, say V, of $(\theta_1, ..., \theta_{d-1})$ from X is given by

$$d_{\mathbb{S}^{d-1}}V = \sin^{d-2}\theta_1\sin^{d-3}\theta_2\cdots\sin\theta_{d-2}d\theta_1\cdots d\theta_{d-1},$$

where $d_{\mathbb{S}^{d-1}}V$ is the volume element of $\mathbb{S}^{d-1}$ which is, in fact, the Jacobian of the transformation from the Cartesian coordinates to the spherical one where $0 \leq \theta_j \leq \pi$ for $j = 1, .., d-2$ and $0 \leq \theta_{d-1} < 2\pi$.

Let $Y = (y_1, \ldots, y_d)^T$, taking its value on $\mathbb{S}^{d-1}$, is distributed as EvMF(μ_0, λ, κ), where the d-dimensional vector $\mu_0 = (1, 0, \ldots, 0)^T$, is known as the North pole in directional statistics terminology. Furthermore, assume M is a rotation matrix featuring a particular property in which pre-multiplying it on μ_0, leads to rotating the North pole to μ, i.e., $M\mu_0 = \mu$. Then, the variable $X = MY$ is distributed as EvMF(μ, λ, κ). In particular, if we use the spherical polar coordinate of X, and consider μ as the North pole, then the components (angles) of the coordinates are independent in distribution and the density for the first angle is given by

$$g(\theta_1) = C_d^{y_1}(\kappa,\lambda)\{1-\lambda\cos\theta_1\}e^{\kappa\cos\theta_1}\sin^{d-2}\theta_1, \tag{2.2}$$

where

$$C_d^{y_1}(\kappa,\lambda) = \left[\int_{-1}^{1}(1-\lambda t)(1-t^2)^{(d-3)/2}e^{\kappa t}dt\right]^{-1} = \frac{(\frac{\kappa}{2})^{d/2-1}}{\Gamma(\frac{d-1}{2})\Gamma(\frac{1}{2})\{I_{d/2-1}(\kappa)-\lambda I_{d/2}(\kappa)\}}. \tag{2.3}$$

Now, it is worths briefly point out the procedure on simulating from our proposed distribution. Assume one is going to generate a sample from EvMF(μ, λ, κ). Following our discussion presented at the beginning of Section 2, we recommend to simulate, first, from EvMF(μ_0, λ, κ) and then employ the transformation $X = MY$, to get the random samples from EvMF(μ, λ, κ). Note that the density of Y is defined on the North pole, and so straightforward to simulate from it using the spherical coordinates. We have seen that the distribution of Y, in terms of the spherical coordinates, is the product of the density of θ_1, given by Equation (2.2), and the distribution of $(\theta_2, \dots, \theta_{d-1})^T$, which is the uniform distribution on $(d-1)$-dimensional unit sphere. Furthermore, θ_1, and $(\theta_2, \dots, \theta_{d-1})^T$, are independent. Finally, to generate θ_1, one could simulate from Equation (2.2) using many popular sampling procedures such as the acceptance-rejection method.

Note that there is another technique to simulate from EvMF(μ_0, λ, κ). Similar to [18] and [22] in generating sample from vMF distribution, we can use $Y^T = (t, (1-t^2)^{1/2} v)$ where t is a random variable with the *p.d.f* given by

$$C_d(\kappa, \lambda)(1-\lambda t)(1-t^2)^{(d-3)/2} e^{\kappa t}, \qquad -1 \le t \le 1.$$

Here, v is uniformly distributed over the space of the $(d-1)$-dimensional unit sphere. See, for example, [8], for more details on this topic.

Finally, it worths to point out, briefly, to choosing a suitable transformation matrix. Following [15], a convenient choice for rotation matrix (M) to rotate unit vector μ_0 to unit vector μ is

$$\frac{(\mu_0 + \mu)(\mu_0 + \mu)^T}{1 + \mu_0^T \mu} - I_d,$$

where $\mu_0 \neq \mu$. Now, we describe a procedure to compute the mean and variance of the EvMF distribution. Using the transformation $X = MY$, the mean of the random variable X is derived as

$$E(X) = M \times E(Y) = \rho_1 M \mu_0 = \rho_1 \mu, \tag{2.4}$$

where ρ_1 is the mean of the first coordinate, say Y_1, for the random variable Y, given by

$$\rho_1 = E_{Y_1}(y_1) = \frac{\frac{I_{d/2}(\kappa)}{I_{d/2-1}(\kappa)} - \lambda\left\{1 - \frac{d-1}{\kappa}\frac{I_{d/2}(\kappa)}{I_{d/2-1}(\kappa)}\right\}}{1 - \lambda\frac{I_{d/2}(\kappa)}{I_{d/2-1}(\kappa)}}. \tag{2.5}$$

Also, the second moment of the random variable Y_1 can be obtained through the following equalities:

$$\rho_2 = E_{Y_1}(y_1^2) = \{1 + \frac{d-1}{\kappa}\frac{\lambda - (1 + \frac{d\lambda}{\kappa})\frac{I_{d/2}(\kappa)}{I_{d/2-1}(\kappa)}}{1 - \lambda\frac{I_{d/2}(\kappa)}{I_{d/2-1}(\kappa)}}\}. \tag{2.6}$$

To derive the variance of the random variable X, we first find $E_{Y_i}(y_i^2)$ for $i = 2, \dots, d$. Since the random variable Y_1 is independent of Y_i, $i = 2, \dots, d$, we have

$$E_{Y_i}(y_i^2) = \frac{1-\rho_2}{d-1}.$$

The following standard integral equality is helpful to obtain $E_{Y_i}(y_i^2)$ and other expressions in the subsequent sections:

$$\int_0^\pi \sin^n x dx = \frac{n-1}{n}\int_0^\pi \sin^{n-2} x dx.$$

Now, for any $Y_m = (\sin\theta_1 \sin\theta_2 \cdots \sin\theta_{m-1}\cos\theta_m)$, $m = 2, \ldots, d$, we have

$$\begin{aligned}
E_{Y_m}(y_m^2) &= \frac{C_d(\kappa,\lambda)}{C_d^{y_1}(\kappa,\lambda)} \int_0^{2\pi}\int_0^{\pi}\cdots\int_0^{\pi} y_m^2 g(\theta_1) d_{\mathbb{S}^{d-1}} V \\
&= \frac{C_d(\kappa,\lambda)}{C_d^{y_1}(\kappa,\lambda)}(1-\rho_2)\int_0^{2\pi}\int_0^{\pi}\cdots\int_0^{\pi} \sin^{d-1}\theta_2 \cdots \sin^{d+3-m}\theta_{m-2}\cos^2\theta_{m-1} \\
&\quad \sin^{d-m}\theta_{m-1}\cdots\sin\theta_{d-2} d\theta_2\cdots d\theta_{d-1} \\
&= \frac{C_d(\kappa,\lambda)}{C_d^{y_1}(\kappa,\lambda)}(1-\rho_2)\frac{d-2}{d-1}\frac{d-3}{d-2}\cdots\frac{d+2-m}{d+3-m}(1-\frac{d+1-m}{d+2-m}) \\
&\quad \int_0^{2\pi}\int_0^{\pi}\cdots\int_0^{\pi} \sin^{d-3}\theta_2\cdots\sin\theta_{d-2}d\theta_2\cdots d\theta_{d-1} \\
&= (1-\rho_2)\frac{d-2}{d-1}\frac{d-3}{d-2}\cdots\frac{d+2-m}{d+3-m}(1-\frac{d+1-m}{d+2-m}) \\
&= \frac{1-\rho_2}{d-1}.
\end{aligned}$$

Since $E_{(Y_2,\ldots,Y_d)}(\cos\theta_i) = 0$, then, $E_{(Y_i,Y_j)}(y_i y_j) = 0$ for $i \neq j$. Hereafter, if necessary we omit the subscripts, indicating corresponding variables for $E(.)$, to ease the complexity of notations. To calculate the variance matrix of the random variable X, we can use the matrix algebra as

$$\begin{aligned}
\mathrm{Var}(X) &= M \times \mathrm{Var}(Y) \times M^T \\
&= M \times \{E(YY^T) - E(Y)E(Y^T)\} \times M^T \\
&= M \begin{bmatrix} \rho_2 & 0 & \cdots & 0 \\ 0 & \frac{1-\rho_2}{d-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1-\rho_2}{d-1} \end{bmatrix} M^T - \rho_1^2 M \times \mu_0\mu_0^T M^T \\
&= \rho_2 M \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} M^T + \frac{1-\rho_2}{d-1} M\{I_d - \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}\}M^T - \rho_1^2\mu\mu^T \\
&= (\rho_2 - \rho_1^2)\mu\mu^T + \frac{1-\rho_2}{d-1}(I_d - \mu\mu^T), \qquad (2.7)
\end{aligned}$$

where I_d is the identity matrix of order d.

To calculate the mean vector and the variance matrix of the random variable X, we can also use the tangent-normal decomposition. In particular, since the EvMF distribution is rotationally symmetric, then its mean and variance can be directly derived from Equations (9.3.33) and (9.3.34) in [15]. The equations for any random variable on $\mathbb{S}^{d-1}$ with the rotational symmetry property and the mean direction μ are given as

$$E[X] = E[t]\mu$$

and

$$\mathrm{Var}(X) = \mathrm{Var}(t)\mu\mu^T + \frac{1-E[t^2]}{d-1}(I_d - \mu\mu^T),$$

where $t = X^T\mu$.

Now, we provide some limiting properties of the quantities given so far that are useful for subsequent results. Let us first recall the important inequality (11) from [4] given by

$$\frac{\kappa}{\nu+1+(\kappa^2+(\nu+1)^2)^{1/2}} \leq \frac{I_{\nu+1}(\kappa)}{I_\nu(\kappa)} \leq \frac{\kappa}{\nu+(\kappa^2+(\nu+2)^2)^{1/2}},$$

for any $\kappa, \nu \geq 0$. Then, we have the following results for the high concentration case $(\kappa \to +\infty)$:

$$\lim_{\kappa \to +\infty} \rho_1 = 1, \qquad \lim_{\kappa \to +\infty} \rho_2 = 1.$$

We can also look at the low concentration case, i.e. $\kappa \to 0$. The results are

$$\lim_{\kappa \to 0} \rho_1 = \frac{-\lambda}{d}, \qquad \lim_{\kappa \to 0} \rho_2 = \frac{d+1-2\lambda}{d}.$$

Technical proof of these limiting expressions are given in the Appendix.

As it is common in the directional distribution, we are also interested in the properties of the EvMF for particular cases of d. Let $d = 2$, and considering the random variable $\theta = \arccos(X)$, the *p.d.f* given by Equation (2.1) turns to

$$\frac{\{1 - \lambda\cos(\theta-\alpha)\}e^{\kappa\cos(\theta-\alpha)}}{2\pi(I_0(\kappa) - \lambda I_1(\kappa))} = C_2(\kappa,\lambda)\{1-\lambda\cos(\theta-\alpha)\}e^{\kappa\cos(\theta-\alpha)},$$

where $C_2(\kappa, \lambda) = \{2\pi(I_0(\kappa) - \lambda I_1(\kappa))\}^{-1}$. Interestingly, this density is quite similar to the cosine perturbed von Mises distribution introduced by [1] in the symmetric case.

Let define $\psi = \kappa^{1/2}(\theta - \alpha)$. Considering the trivial equality $1 - \cos(\kappa^{-1/2}\psi) = \frac{1}{2}\kappa^{-1}\psi^2 + O(\kappa^{-2})$, for the low concentrate case $(\kappa \to 0)$, we have

$$\begin{aligned}
\{1-\lambda\cos(\theta-\alpha)\}e^{\kappa\cos(\theta-\alpha)} &\propto \{1+\lambda(\frac{\psi^2}{2\kappa}-1)\}e^{-\kappa(\frac{\psi^2}{2\kappa}-1)} \\
&= e^{\kappa}\{1-\lambda+\frac{\lambda}{2\kappa}\psi^2\}e^{-\frac{\psi^2}{2}} \\
&\propto (1-\lambda)f_1(\psi) + \lambda f_2(\psi),
\end{aligned}$$

where $f_1(\psi)$ is the *p.d.f* of $\psi \sim N(0,1)$, in fact the standard normal distribution, and $f_2(\psi)$ is the Bimodal Extended Generalized Gamma (BEGG) probability function with parameters $\alpha = 1$, $\beta = 2$, $\delta_{0,1} = 2$, $\eta = 2\sqrt{2}\kappa$ and $\varepsilon = \mp(1 - 1/2\kappa)$. For more information on the BEGG distribution, see for example [6].

To recall Equation (2.1) for the particular case $d = 3$, one can easily show that *p.d.f* of the random variable X is given as

$$\frac{\kappa\{1-\lambda \mathrm{x}^T\mu\}e^{\kappa \mathrm{x}^T\mu}}{4\pi\sinh\kappa\{1+\lambda(\frac{1}{\kappa}-\coth\kappa)\}}. \tag{2.8}$$

However, we prefer to use the spherical coordinate (θ, ϕ) as the interested random variable rather than the unit length vector X. Then, the density in Equation (2.8) turns to a distribution with the mean direction (α, β) and the concentration parameter κ. We call it the extended Fisher distribution with the *p.d.f* given by

$$\frac{\kappa\{1-\lambda(\cos\theta\cos\alpha+\sin\theta\sin\alpha\cos(\phi-\beta))\}e^{\kappa\{\cos\theta\cos\alpha+\sin\theta\sin\alpha\cos(\phi-\beta)\}}\sin\theta}{4\pi\sinh\kappa\{1+\lambda(\frac{1}{\kappa}-\coth\kappa)\}}. \tag{2.9}$$

In this case, if $\alpha = \pi/2$, the distribution is symmetric about (α, β). Moreover, there are two local modes, two saddle points and a local minimum in this case. To clarify this last remark, let us take the derivatives of density given by Equation (2.9) with respect to θ and ϕ, separately, and set the results to zero. Then, solving the ultimate expressions leads to the following solutions:

$$\begin{aligned}\theta_1 &= \pi/2,\\ \theta_2 &= \arcsin\left(\frac{\sqrt{4\lambda^2+\kappa^2}}{2\kappa\lambda}+\frac{1}{2\lambda}-\frac{1}{\kappa}\right),\\ \phi_1 &= \beta,\\ \phi_2 &= \arccos\left(\frac{1}{\lambda}-\frac{1}{\kappa}\right)+\beta,\end{aligned}$$

where (θ_1, ϕ_1) is the local minimum, (θ_1, ϕ_2) are the local maximums and (θ_2, ϕ_1) are the saddle points provided the inequality $|1/\lambda - 1/\kappa| < 1$ holds. Note that one gets two answers in using the functions arcsin and arccos, and so, she finally has five pairs of critical points. Moreover, the point (θ_1, ϕ_1) will be a local maximum for the density if $|1/\lambda - 1/\kappa| \geq 1$.

A schematic representation of the EvMF density for some particular values of its parameters is plotted in Figure 1. Some remarks highlighted above can be seen in this plot. Note that the *p.d.f* in Equation (2.9) has a more sophisticated form if $\alpha \neq \pi/2$. However, it can be shown that the density still has only one maximum point, which cannot be written in an explicit form.

Many other properties of the density given in Equation (2.1) require a deep mathematical understanding which is behind the scope of this paper. We leave them for future research and now concentrate on statistical inference on the parameters of the EvMF distribution.

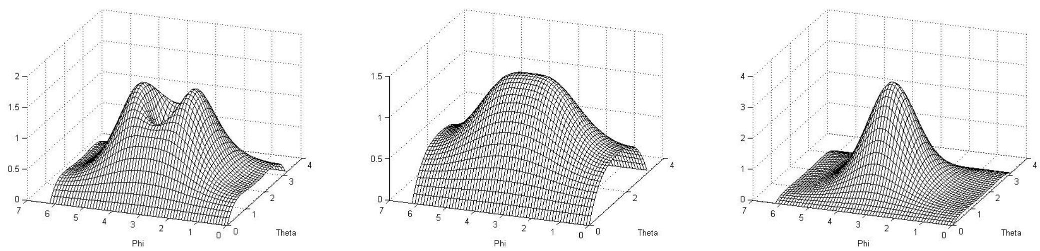

Figure 1. The plot of *p.d.f.* for the EvMF distribution while its parameters are set on some specific values. In particular, α, β are fixed at $\pi/2$ and π, respectively. Also, from left to right, we set (λ, κ) to $(0.9, 2), (0.5, 1), (0.3, 2)$.

3. Statistical inference on EvMF distribution

In this section, we consider the statistical inference on the distribution given by Equation (2.1) while all parameters are assumed to be unknown. We confine ourselves to the point estimation in this paper. The interval estimation and testing hypothesis are left for future research. We first concentrate on the maximum likelihood method and then follow the method of moment.

3.1. Maximum likelihood method

Let $\mathrm{x}_1, \ldots, \mathrm{x}_n$ be a random sample on $\mathbb{S}^{d-1}$, from EvMF(μ, λ, κ). Then, the logarithm of the likelihood function for the parameters, say $\ell(\mu, \kappa, \lambda)$, can be expressed as

$$\ell(\mu,\kappa,\lambda) \propto n(d/2-1)\log\kappa - n\log(I_{\frac{d}{2}-1}(\kappa) - \lambda I_{\frac{d}{2}}(\kappa)) + \textstyle\sum_{i=1}^n \log\{1-\lambda \mathrm{x}_i^T\mu\} + \kappa\sum_{i=1}^n \mathrm{x}_i^T\mu.$$

Differentiating $\ell(\mu, \kappa, \lambda)$ with respect to μ, κ and λ, noting that $\mu^T\mu = 1$ and $\kappa > 0$, the maximum likelihood estimators are derived through solving the following equalities:

$$\hat{\mu} = \frac{\hat{\kappa}\sum_{i=1}^n \mathrm{x}_i - \sum_{i=1}^n \frac{\hat{\lambda}\mathrm{x}_i}{1-\hat{\lambda}\mathrm{x}_i^T\hat{\mu}}}{||\hat{\kappa}\sum_{i=1}^n \mathrm{x}_i - \sum_{i=1}^n \frac{\hat{\lambda}\mathrm{x}_i}{1-\hat{\lambda}\mathrm{x}_i^T\hat{\mu}}||}, \tag{3.1}$$

$$(\frac{d}{2}-1)\frac{1}{\hat{\kappa}} - \frac{I'_{\frac{d}{2}-1}(\hat{\kappa}) - \hat{\lambda}I'_{\frac{d}{2}}(\hat{\kappa})}{I_{\frac{d}{2}-1}(\hat{\kappa}) - \hat{\lambda}I_{\frac{d}{2}}(\hat{\kappa})} = -\frac{1}{n}\sum_{i=1}^n \mathrm{x}_i^T\hat{\mu}, \tag{3.2}$$

$$\frac{I_{d/2}(\hat{\kappa})}{I_{\frac{d}{2}-1}(\hat{\kappa}) - \hat{\lambda}I_{\frac{d}{2}}(\hat{\kappa})} = \frac{1}{n}\sum_{i=1}^n \frac{\mathrm{x}_i^T\hat{\mu}}{1-\hat{\lambda}\mathrm{x}_i^T\hat{\mu}}. \tag{3.3}$$

As seen, there is not any explicit expression for either of these equations to derive the estimators. Hence, one has to follow some iterative algorithms to derive the estimators numerically. From computational viewpoint, it is important to implement a feasible procedure. We suggest the following steps in turn:

Step 1. Given an initial value for λ_0 and taking $\mu_0 = \sum_{i=1}^n \mathrm{x}_i/||\sum_{i=1}^n \mathrm{x}_i||$, estimate κ from Equation (3.2).
Step 2. Using λ_0 and the estimated κ gained from **Step 1**, estimate μ using Equation (3.1).
Step 3. Using the updated μ and κ, estimate λ via invoking Equation (3.3).
Step 4. Iterate Equations (3.1-3.3) with the updated parameters until a convergence criterion is satisfied.

We discuss a procedure to choose some specific initial values to guarantee the convergence closed to the end of the next subsection. It can also be used here to derive the maximum likelihood estimators for the parameters of the EvMF distribution.

3.2. Method of moment

Now, we concentrate on estimating the parameters, i.e. μ, λ and κ, of the EvMF distribution using the method of moments. To recall Equations (2.4) to (2.7), one only needs to employ the first two moments to derive the estimates of the parameters. Note that to distinct our estimators in this section with the corresponding estimators through the maximum likelihood method, we insert a tilde over each estimator in the following expressions.

Based on a random sample $(\mathrm{x}_1, \ldots, \mathrm{x}_n)$, from the EvMF($\mu$, λ, κ) distribution and using the Equations (2.4) and (2.7), the moment estimators are derived by solving the following equalities:

$$\tilde{\mu} = \bar{\mathrm{x}}/\tilde{\rho}_1, \tag{3.4}$$

$$\frac{1}{n}\sum_{i=1}^n \mathrm{x}_i\mathrm{x}_i^T = \tilde{\rho}_2\tilde{\mu}\tilde{\mu}^T + \frac{1-\tilde{\rho}_2}{2}(I_3 - \tilde{\mu}\tilde{\mu}^T). \tag{3.5}$$

Considering the constraint $\tilde{\mu}^T\tilde{\mu} = 1$, it is seen that

$$\tilde{\rho}_1 = ||\bar{\mathrm{x}}||, \tag{3.6}$$

$$\tilde{\rho}_2 = \frac{1}{n}\sum_{i=1}^n (\mathrm{x}_i^T\tilde{\mu})^2. \tag{3.7}$$

Note that Equation (3.7) can also be derived via pre-multiplication and post-multiplication of the Equation (3.5) by $\tilde{\mu}^T$ and $\tilde{\mu}$, respectively. Now, plugging Equation (3.6) into Equation (3.4), the moment estimator of μ, i.e. $\tilde{\mu}$ is given. Then, inserting $\tilde{\mu}$ into Equation

(3.7) and recalling Equation (2.6), one only needs to solve the resulted equalities to derive the moment estimators of λ and κ.

In summary, to obtain the moment estimators of the EvMF, one can first derive $\tilde{\mu} = \bar{\mathrm{x}}/||\bar{\mathrm{x}}||$, based on the aforementioned discussion. Then, we have intuitively a value for $\tilde{\rho}_1$. To solve Equation (2.5) in term of λ, and assuming an estimate for $\tilde{\kappa}$, we have the following the moment estimator for λ:

$$\tilde{\lambda} = \frac{\frac{I_{d/2}(\tilde{\kappa})}{I_{d/2-1}(\tilde{\kappa})} - \tilde{\rho}_1}{1 - \frac{I_{d/2}(\tilde{\kappa})}{I_{d/2-1}(\tilde{\kappa})}(\frac{d-1}{\tilde{\kappa}} + \tilde{\rho}_1)}. \tag{3.8}$$

Clearly, this last expression along with Equation (3.7), while inserting $\tilde{\mu}$, should be iterated to derive the moment estimators for both $\tilde{\lambda}$ and $\tilde{\kappa}$.

From a computational viewpoint, one requires an initial value for either $\tilde{\lambda}$ or $\tilde{\kappa}$, to obtain the corresponding moment estimators. We have proposed a simple trick to choose the reasonable values which guarantee the iterative algorithm related to either the maximum likelihood or method of moment estimators. First, following the equality (9.6.26) given by [3], we have

$$I_{d/2+1}(\tilde{\kappa}) = I_{d/2-1}(\tilde{\kappa}) - \frac{d}{\tilde{\kappa}} I_{d/2}(\tilde{\kappa}). \tag{3.9}$$

Then, using Equations (2.5) and (2.6), we can write

$$\frac{\tilde{\rho}_1}{\frac{1-\tilde{\rho}_2}{d-1}} = \tilde{\kappa}\frac{I_{d/2}(\tilde{\kappa}) - \lambda(I_{d/2-1}(\tilde{\kappa}) - \frac{d-1}{\tilde{\kappa}} I_{d/2}(\tilde{\kappa}))}{I_{d/2}(\tilde{\kappa}) - \lambda(I_{d/2-1}(\tilde{\kappa}) - \frac{d}{\tilde{\kappa}} I_{d/2}(\tilde{\kappa}))}.$$

Now, utilizing Equation (3.9), we will have

$$\begin{aligned} \frac{(d-1)\tilde{\rho}_1}{1-\tilde{\rho}_2} &= \tilde{\kappa}\frac{I_{d/2}(\tilde{\kappa}) - \lambda(I_{d/2+1}(\tilde{\kappa}) + \frac{1}{\tilde{\kappa}} I_{d/2}(\tilde{\kappa}))}{I_{d/2}(\tilde{\kappa}) - \lambda I_{d/2+1}(\tilde{\kappa})} \\ &= \tilde{\kappa} - \lambda\frac{I_{d/2}(\tilde{\kappa}))}{I_{d/2}(\tilde{\kappa}) - \lambda I_{d/2+1}(\tilde{\kappa})}. \end{aligned}$$

An initial value for $\tilde{\kappa}$, can be set up based on this last expression. In particular, to set $\lambda = 0$, we will get

$$\tilde{\kappa} = \frac{(d-1)\tilde{\rho}_1}{1-\tilde{\rho}_2},$$

which can be considered as the initial value for $\tilde{\kappa}$, to start any iterative algorithm to derive the moment estimators for the parameters described above. Note that this value is, in fact, the middle point of the confidence band achieved from the fixed point iteration method proposed by [17].

Now, one can iteratively repeat the following steps, along with some reasonable convergence criteria, to derive the moment estimators of λ and κ:

Step 1. Obtain the value for $\tilde{\lambda}$ using Equation (3.8).
Step 2. Solve the equality

$$\frac{1}{n}\sum_{i=1}^{n}(\mathrm{x}_i^T\bar{\mathrm{x}}/||\bar{\mathrm{x}}||)^2 = \{1 + \frac{d-1}{\tilde{\kappa}}\frac{\tilde{\lambda} - (1+\tilde{\lambda}\frac{d}{\tilde{\kappa}})\frac{I_{d/2}(\tilde{\kappa})}{I_{d/2-1}(\tilde{\kappa})}}{1 - \tilde{\lambda}\frac{I_{d/2}(\tilde{\kappa})}{I_{d/2-1}(\tilde{\kappa})}}\},$$

to compute the value for $\tilde{\kappa}$.
Step 3. Iterate **Step 1** and **Step 2** with updated parameters to reach pre-specified convergence criteria.

4. Simulation studies and real data application

This section concerns conducting a simulation study on evaluating the statistical inference on the parameters of the EvMF distribution. We also consider analyzing the real-life data using the EvMF and some other directional densities in this section.

We first generate data from the EvMF distribution and then estimate the parameters using both point estimation procedures discussed in this paper. The simulation studies are done via altering the location (α, β), the concentration (κ) and skewness (λ) parameters for the reasonable ranges of their values to showcase their performance on our statistical inference on the EvMF distribution. Also, to show the effect of changing the sample size on the ultimate results, we consider $n = 20, 50, 100, 200$. The procedure is also done for the small, middle and large values of κ and λ via changing the location parameters α and β, on different values. Furthermore, we set the simulation run on 1000. Finally, to evaluate the accuracy of the estimators, we calculate the absolute bias (ABias) and mean square errors (MSE) after estimating the parameters in each set of the simulation run and scenario. For brevity of notation, we write MLE and MM to stand for the maximum likelihood and method of moment estimators, respectively, throughout this section.

We have reported the results gained from our simulation studies in three tables. In particular, the results are shown in Tables 1 for $\alpha = \beta = \pi/3$, $\kappa = 2$ and $\lambda = 0.1$, in Tables 2 for $\alpha = \pi/4$, $\beta = \pi/8$, $\kappa = 5$ and $\lambda = 0.5$ and in Tables 3 for $\alpha = \pi/10$, $\beta = \pi/6$, $\kappa = 8$ and $\lambda = 0.9$. As seen, there is no sign of a monotone change in the values of the ABias and MSE throughout these results. Hence, stating a general remark is not straightforward, although we need to provide a clear comparison. It sounds that the ABias as well as MSE values of the MM estimators for the location parameters, i.e. (α, β), are higher than those for the MLE estimators. However, this is not the case when λ increases. As expected, as the sample size increases the MLE has less ABias and MSE for all scenarios.

Table 1. The absolute bias and MSE of the estimators in estimating the parameters of the EvMF distribution for different sample sizes, using two methods of estimations and considering pre-defined quantities set as follows: $n = 20, 50, 100, 200$, $\alpha = \pi/3$, $\beta = \pi/3$, $\kappa = 2$ and $\lambda = 0.1$.

n	20		50		100		200	
Method	MLE	MM	MLE	MM	MLE	MM	MLE	MM
ABias(α)	0.023	0.023	0.003	0.003	0.005	0.005	0.004	0.004
MSE(α)	0.047	0.041	0.020	0.020	0.009	0.009	0.005	0.006
ABias(β)	0.097	0.127	0.009	0.012	0.008	0.011	0.001	0.003
MSE(β)	0.281	0.194	0.023	0.026	0.013	0.015	0.007	0.007
ABias(κ)	0.162	0.896	0.028	0.889	0.009	0.805	0.001	0.657
MSE(κ)	0.310	1.313	0.134	1.252	0.059	0.944	0.030	0.607
ABias(λ)	0.058	0.247	0.048	0.214	0.043	0.190	0.035	0.145
MSE(λ)	0.004	0.274	0.003	0.225	0.002	0.182	0.002	0.122

The story for estimating the concentration (κ) and skewness (λ) parameters are relatively the same as the location parameters. However, unlike the MLE, the ABias and MSE criteria for the MM estimator very remarkably. Generally, the MLE has less values for the ABias and MSE for all scenarios. The exception is when λ increases in which the MM has the same or better performance in comparison with the MLE.

As mentioned before, there is not a clear sign of a particular pattern in estimating the parameters and the method of estimation. Nonetheless, we can claim that there is less bias to estimate the location parameters regardless of the method of estimation. But, the bias is relatively high for both the concentration and skewness parameters. This case is significant for the large λ. Moreover, there is a considerable change in the values of the MSE in these situations. Recalling Equation (3.3), one sees that the estimation of λ is highly dependent on its corresponding estimation for the κ and vice versa. Hence, extra care is required in estimating the concentration and skewness parameters of the EvMF distribution.

Table 2. The absolute bias and MSE of the estimators in estimating the parameters of the EvMF distribution for different sample sizes, using two methods of estimations and considering pre-defined quantities set as follows: $n = 20, 50, 100, 200$, $\alpha = \pi/4$, $\beta = \pi/8$, $\kappa = 5$ and $\lambda = 0.5$.

n	20		50		100		200	
Method	MLE	MM	MLE	MM	MLE	MM	MLE	MM
ABias(α)	0.022	0.010	0.003	0.003	0.002	0.002	0.002	0.003
MSE(α)	0.015	0.014	0.006	0.006	0.003	0.003	0.002	0.002
ABias(β)	0.003	0.003	0.003	0.003	0.002	0.002	0.001	0.002
MSE(β)	0.025	0.030	0.011	0.012	0.007	0.008	0.003	0.003
ABias(κ)	0.179	0.685	0.057	0.643	0.042	0.518	0.030	0.447
MSE(κ)	1.673	1.169	0.439	1.113	0.199	0.797	0.106	0.608
ABias(λ)	0.087	0.208	0.80	0.181	0.074	0.124	0.053	0.094
MSE(λ)	0.009	0.118	0.008	0.100	0.006	0.076	0.004	0.061

Table 3. The absolute bias and MSE of the estimators in estimating the parameters of the EvMF distribution for different sample sizes, using two methods of estimations and considering pre-defined quantities set as follows: $n = 20, 50, 100, 200$ $\alpha = \pi/10$, $\beta = \pi/6$, $\kappa = 8$ and $\lambda = 0.9$.

n	20		50		100		200	
Method	MLE	MM	MLE	MM	MLE	MM	MLE	MM
ABias(α)	0.020	0.015	0.019	0.008	0.010	0.008	0.002	0.002
MSE(α)	0.014	0.011	0.007	0.005	0.003	0.003	0.001	0.001
ABias(β)	0.034	0.024	0.020	0.012	0.005	0.008	0.002	0.005
MSE(β)	0.199	0.167	0.087	0.050	0.034	0.026	0.021	0.012
ABias(κ)	0.291	0.486	0.272	0.478	0.223	0.382	0.145	0.331
MSE(κ)	1.592	0.941	0.715	0.929	0.402	0.763	0.220	0.626
ABias(λ)	0.027	0.007	0.018	0.009	0.018	0.012	0.006	0.006
MSE(λ)	0.001	0.009	0.001	0.005	0.001	0.003	0.001	0.003

Here, it worths mentioning some remarks in conjunction with the criticism on estimating the concentration and skewness parameters of the EvMF distribution. We emphasized on the constraint $|1/\lambda - 1/\kappa|$ leading to either the unimodal or bimodal distributions, while discussing the properties of the EvMF density. As one can see, this constraint will affect the performance of the estimators for both κ and λ. In particular, if $|1/\lambda - 1/\kappa|$ is relatively big, which is the case in Table 3, there is no guarantee on deriving the good estimator for the concentration parameter (κ). The EvMF density has two modes in this case and so one, in fact, encounters with a mixture of directional distributions with two different concentration parameters. Hence, take them the same and then estimate it from only one density, might not be a reasonable strategy. Albeit, there is no excuse for our unrealistic procedure in this case and we should seek a proper inferential method to more accurately estimate κ. This is a topic which we left for our future research. In contrast, if the value of $|1/\lambda - 1/\kappa|$, is relatively small, which is the case in Table 1, one expects more accurate estimates for κ and λ if the MLE method is employed to estimate the concentration and skewness parameters of the EvMF distribution. Note that for a very small value of κ the density is unimodal. So, one expects not to see the skewness parameter appearing in the distribution. Hence, an attempt to estimate λ might result in reasonable estimation. Generally, extra care is required in estimating the concentration and skewness parameters of the EvMF distribution when treating real-life examples.

Now, we consider analyzing real-life data using the EvMF distributions. To have a basis for comparison, we also employ some other directional densities in fitting these data. They are the vMF, Kent, Wood, angular central Gaussian (ACG) and the mixture of vMF (movMF) with two clusters. The interested readers can consult [15] for more details on these distributions. There are three sets of data, briefly described below:

Data set 1: Measurements of the orientation of the dendritic field at various sites in the retinas of 6 cats, in response to different visual stimuli. The effect of the visual stimulus was recorded as two angles and these effects were applied in three directions. The sample size of data, which was initially analyzed by [12], is 94. The data worth to study because it is a mixture of three groups of responses to different visual stimuli.

Data set 2 and 3: The data set was collected from a survey on household expenditures in four commodity groups and give the expenses of 20 single men and 20 single women on four commodity groups. The data were analyzed by [10] in which they normalized three commodity groups (housing, food and service) while fitting mixtures of vMF on $\mathbb{S}^2$. We divide the entire data into two new sets: housing, goods and food as data set 2 and housing, food and service as data set 3. In reality, the data were collected from two different groups males and females, or in another word, they come from two clusters. So, the distribution of the entire data might be properly fitted by a bimodal density.

Data set 1 can be found in the Appendix B15 of [8] and data sets 2 and 3 are available in package **HSAUR2** [7]. A schematic representation of data set 1, 2 and 3 are plotted in Figure 2. As seen, groups are separated into two distinct clusters for data set 2 and 3 and so a bimodal spherical density might be an appropriate distribution to model the data here. Due to the dispersion of dataset 1, we may have the symmetric unimodal model. An example of this kind of model can be seen in the middle of Figure 1.

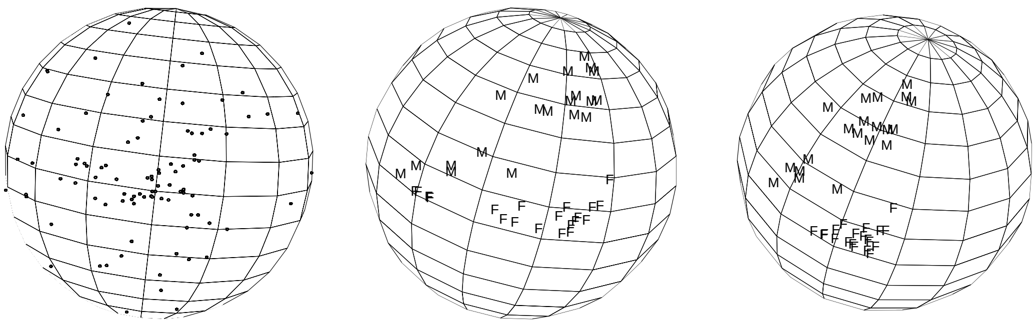

Figure 2. Position of unit sized vectors for data set 1, 2 and 3 (left to right) over the unit sphere.

After fitting these groups of data by implementing different spherical densities, the Akaike information criterion (AIC) and Bayesian information criterion (BIC) were calculated. Four spherical distributions considered here are the vMF, Kent, Wood, angular central Gaussian (ACG) and mixture of vMF (movMF) with two components. We, also, fit the data sets using our proposed density, i.e., EvMF. The results are shown in Table 4 with the AIC appeared first and the BIC inside the bracket.

As seen in Table 4, there is a better fitting of the EvMF density versus other distributions while using data set 2. For data set 3, the considered criteria is the smallest if one uses movMF. Leaving this density aside, the EvMF distribution fits data reasonably better than the other densities. Although fitting data using the movMF is better in some cases, the general theory of the mixture distributions for directional statistics makes it intractable here. Apart from this, choosing the numbers of mixing components and parameters for each mixing densities are two critical issues that one needs to deal with while using the mixture distributions. Moreover, statistical inferences on the parameters can be misleading on using the mixture distributions when the data involves highly asymmetric observations. Data sets 1 is a special case in which the movMF and Wood are doing better than the others. In fact, treating the former density needs extra investigation. The latter distribution is only suitable for the axial data, although we were not interested in dealing with the axial distribution in this paper.

Table 4. The values of AIC (BIC) criteria after fitting the vMF and EvMF, Kent, Wood, ACG and movMF distributions when different data sets are used. See the text for more details.

Data set	vMF	EvMF	Kent	Wood	ACG	movMF
1	421(429)	405(415)	416(429)	186(199)	464(480)	176(194)
2	63.8(68.9)	51.9(58.7)	64.5(72.9)	62.6(71.1)	146(157)	62.0(73.8)
3	28.0(33.0)	14.4(21.2)	17.3(25.7)	35.0(43.5)	91.3(101)	4.92(16.7)

Based on the results given in Table 4, one can see that the EvMF distribution is fitting all three sets of data better than the vMF density. Hence, it is of interest to see the extent by which the skewness control parameter affect the estimate of the other parameters. We reported the estimates of the parameters after fitting these distributions when the values of three real data sets are used as the spherical observations in Table 5. As seen, the estimation of the mean parameter is relatively the same for both densities. However, ignoring the skewness leads to an underestimate of the concentration parameter for all three data sets. In other words, using the EvMF distribution to fit the data turns to estimate κ as twice as bigger than the case in which we ignore λ.

Table 5. The estimations of the parameters after fitting the vMF and EvMF distributions when three real data set are used as the spherical observations.

	Data set 1				Data set 2				Data set 3			
Parameters	α	β	κ	λ	α	β	κ	λ	α	β	κ	λ
vMF	2.81	-1.00	1.55	-	0.67	0.73	8.28	-	0.57	0.71	12.97	-
EvMF	2.81	-1.01	3.31	0.96	0.66	0.73	16.6	≈ 1	0.56	0.72	25.94	≈ 1

5. Conclusion

Invoking some unimodal distributions might not be suitable for describing the random behavior of some spherical data. Particularly, if more than one mode is appearing in the histogram of data under study, taking into account a multimodal spherical density is recommended. To overcome this obstacle, we proposed an extended version of the von Misses-Fisher distribution to tackle multimodality in this paper. Some properties of this density have been investigated and the connection with previously proposed distributions on the sphere was highlighted. We also detailed two procedures to estimate the parameters of the distribution. To evaluate the performance of our estimators, we conducted simulation studies. The results of simulations showed that the estimation through the maximum likelihood method outperforms that given by the method of moment. In analyzing the real-life data set, the flexibility of EvMF was shown by using the AIC and BIC criteria.

However, the mixture of vMF distribution has a better performance in some particular cases. This needs to be investigated in more details.

The research reported in this paper can be extended further in some other directions. The methodology employed in this paper was initially constructed based on a density existing in the Euclidean space. However, extending other spherical distributions can be a good proposal. Moreover, seeking some proper inferential procedures for estimating the concentration and skewness parameters of the EvMF distribution in order to be applicable smoothly for both unimodal and multimodal cases, is a possible option for future research.

References

[1] T. Abe and A. Pewsey, *Sine-skewed circular distributions*, Statist. Papers **52** (3), 683-707, 2011.

[2] T. Abe and A. Pewsey, *Symmetric circular models through duplication and cosine perturbation*, Comput. Statist. Data Anal. **55** (12), 3271-3282, 2011.

[3] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*, Dover Publications, 1965.

[4] D.E. Amos, *Computation of modified Bessel functions and their ratios*, Math. Comp. **28** (125), 239-251, 1974.

[5] A. Azzalini, *A class of distributions which includes the normal ones*, Scand. Actuar. J. **12** (2), 171-178, 1985.

[6] M.N. Çankaya, Y.M. Bulut, F.Z. Dogru and O. Arslan, *A bimodal extension of the generalized gamma distribution*, Rev. Colomb. Estad. **38** (2), 371-384, 2015.

[7] B.S. Everitt and T. Hothorn, *HSAUR2: A Handbook of Statistical Analyses Using R*, 2nd ed., R Package Version 1.1-6, 2013.

[8] N.I. Fisher, T. Lewis and B.J. Embleton, *Statistical Analysis of Spherical Data*, Cambridge University Press, 1987.

[9] R. Gatto and S.R. Jammalamadaka, *The generalized von Mises distribution*, Stat. Methodol. **4** (3), 341-353, 2007.

[10] K. Hornik and B. Grün, *movmf: An R package for fitting mixtures of von Mises-Fisher distributions*, J. Stat. Softw. **58** (10), 1-31, 2014.

[11] S. Kato and M. Jones, *An extended family of circular distributions related to wrapped Cauchy distributions via brownian motion*, Bernoulli **19** (1), 154-171, 2013.

[12] J. Keilson, D. Petrondas, U. Sumita and J. Wellner, *Significance points for some tests of uniformity on the sphere*, J. Stat. Comput. Simul. **17** (3), 195-218, 1983.

[13] J.T. Kent, *The Fisher-Bingham distribution on the sphere*, J. R. Stat. Soc. Ser. B. Stat. Methodol. **44** (1), 71-80, 1982.

[14] B. Kim, S. Huckemann, J. Schulz and S. Jung, *Small-sphere distributions for directional data with application to medical imaging*, Scand. J. Stat. **46** (4), 1047-1071, 2019.

[15] K.V. Mardia, and P.E. Jupp, *Directional Statistics*, John Wiley & Sons, 2000.

[16] A. Pewsey, M. Neuhäuser and G.D. Ruxton, *Circular Statistics in R*, Oxford University Press, 2013.

[17] A. Tanabe, K. Fukumizu, S. Oba, T. Takenouchi and S. Ishii, *Parameter estimation for von MisesFisher distributions*, Comput. Statist. **22** (1), 145-157, 2007.

[18] G. Ulrich, *Computer generation of distributions on the m-sphere*, J. R. Stat. Soc. Ser. C. Appl. Stat. **33**, 158-163, 1984.

[19] D. Umbach and S.R. Jammalamadaka, *On introducing asymmetry into circular distributions*, Pak. J. Stat. Oper. Res. **8** (3), 531-535, 2012.

[20] G.S. Watson, *Distributions on the circle and sphere*, J. Appl. Probab. **19**, 265-280, 1982.

[21] A.T. Wood, *A bimodal distribution for the sphere*, J. R. Stat. Soc. Ser. C. Appl. Stat. **31** (1), 52-58, 1982.

[22] A.T. Wood, *Simulation of the von Mises-Fisher distribution*, Comm. Statist. Simulation Comput. **23** (1), 157-164, 1994.

[23] E. Yfantis and L. Borgman, *An extension of the von Mises distribution*, Comm. Statist. Theory Methods **11** (15), 1695-1706, 1982.

Appendix

To calculate the normalized constant, ρ_1 and ρ_2, we first calculate some important integrals. Given modified Bessel function integral representation as, see e.g. [3]

$$I_\nu(\kappa) = \frac{(\frac{\kappa}{2})^\nu}{\Gamma(\frac{1}{2})\Gamma(\nu+\frac{1}{2})}\int_{-1}^{1}(1-t^2)^{\nu-\frac{1}{2}}e^{\kappa t}dt, \tag{5.1}$$

and the derivative properties of Bessel function, we have

$$\begin{aligned}\frac{\partial I_\nu(\kappa)}{\partial\kappa} &= \frac{\partial}{\partial\kappa}\frac{(\frac{\kappa}{2})^\nu}{\Gamma(\frac{1}{2})\Gamma(\nu+\frac{1}{2})}\int_{-1}^{1}(1-t^2)^{\nu-\frac{1}{2}}e^{\kappa t}dt\\ &= \frac{\frac{\nu}{2}(\frac{\kappa}{2})^{\nu-1}}{\Gamma(\frac{1}{2})\Gamma(\nu+\frac{1}{2})}\int_{-1}^{1}(1-t^2)^{\nu-\frac{1}{2}}e^{\kappa t}dt + \frac{(\frac{\kappa}{2})^\nu}{\Gamma(\frac{1}{2})\Gamma(\nu+\frac{1}{2})}\int_{-1}^{1}t(1-t^2)^{\nu-\frac{1}{2}}e^{\kappa t}dt\\ &= \frac{\nu}{\kappa}I_\nu(\kappa) + \frac{(\frac{\kappa}{2})^\nu}{\Gamma(\frac{1}{2})\Gamma(\nu+\frac{1}{2})}\int_{-1}^{1}t(1-t^2)^{\nu-\frac{1}{2}}e^{\kappa t}dt.\end{aligned}$$

From Equation (9.6.26) of [3], we can write

$$\int_{-1}^{1}t(1-t^2)^{\nu-\frac{1}{2}}e^{\kappa t}dt = \frac{\Gamma(\frac{1}{2})\Gamma(\nu+\frac{1}{2})}{(\frac{\kappa}{2})^\nu}I_{\nu+1}(\kappa). \tag{5.2}$$

Also, we have

$$\begin{aligned}\int_{-1}^{1}t^2(1-t^2)^{\nu-\frac{1}{2}}e^{\kappa t}dt &= \int_{-1}^{1}(1-t^2)^{\nu-\frac{1}{2}}e^{\kappa t}dt - \int_{-1}^{1}(1-t^2)^{\nu+1-\frac{1}{2}}e^{\kappa t}dt\\ &= \frac{\Gamma(\frac{1}{2})\Gamma(\nu+\frac{1}{2})}{(\frac{\kappa}{2})^\nu}\{I_\nu(\kappa) - \frac{2\nu+1}{\kappa}I_{\nu+1}(\kappa)\} \qquad (5.3)\end{aligned}$$

and

$$\begin{aligned}\int_{-1}^{1}t^3(1-t^2)^{\nu-\frac{1}{2}}e^{\kappa t}dt &= \int_{-1}^{1}t(1-t^2)^{\nu-\frac{1}{2}}e^{\kappa t}dt - \int_{-1}^{1}t(1-t^2)^{\nu+1-\frac{1}{2}}e^{\kappa t}dt\\ &= \frac{\Gamma(\frac{1}{2})\Gamma(\nu+\frac{1}{2})}{(\frac{\kappa}{2})^\nu}\{I_{\nu+1}(\kappa) - \frac{2\nu+1}{\kappa}I_{\nu+2}(\kappa)\}. \qquad (5.4)\end{aligned}$$

To calculate the normalized constant, we use the density given in Equation (2.1), in which its integration should be one. To compute the integral, we prefer to recall the spherical coordinates. Without loss of generality, let us take $\alpha_1 = 0$. Then, we have

$\mu = [1, 0, \ldots, 0]^T$. Now, as discussed at the beginning of Section 2, we can write

$$\begin{aligned}
1/C_d(\kappa,\lambda) &= \int_{\mathbb{S}^{d-1}} (1-\lambda x^T\mu) e^{\kappa x^T \mu} d_{\mathbb{S}^{d-1}} V \\
&= \int_{\mathbb{S}^{d-2}} \int_0^{\pi} \sin^{d-2}\theta_1 (1-\lambda\cos\theta) e^{\kappa\cos\theta_1} d\theta_1 d_{\mathbb{S}^{d-2}} V \\
&= \int_{\mathbb{S}^{d-2}} \int_{-1}^{1} (1-\lambda t)(1-t^2)^{(d-3)/2} e^{\kappa t} dt d_{\mathbb{S}^{d-2}} V \\
&= \frac{2\pi^{(d-1)/2}}{\Gamma(\frac{d-1}{2})} \Big\{ \int_{-1}^{1} (1-t^2)^{(d-3)/2} e^{\kappa t} dt - \lambda \int_{-1}^{1} t(1-t^2)^{(d-3)/2} e^{\kappa t} dt \Big\}.
\end{aligned}$$

As seen, taking $\nu = d/2 - 1$ in Equations (5.1) and (5.2), the normalized constant is obtained via calculating the corresponding Bessel functions.

Now, we describe procedure to compute ρ_1 and ρ_2. Similar to the above discussion, taking $\nu = d/2 - 1$ and using Equations (5.2)-(5.4) and Equations (9.6.26) of [3], we have

$$\begin{aligned}
\rho_1 = E_{y_1}(t) &= C_d^{y_1}(\kappa,\lambda) \int_{-1}^{1} t(1-\lambda t)(1-t^2)^{(d-3)/2} e^{\kappa t} dt \qquad (5.5) \\
&= C_d^{y_1}(\kappa,\lambda) \{ \int_{-1}^{1} t(1-t^2)^{(d-3)/2} e^{\kappa t} dt - \lambda \int_{-1}^{1} t^2(1-t^2)^{(d-3)/2} e^{\kappa t} dt \} \\
&= C_d^{y_1}(\kappa,\lambda) \frac{\Gamma(\frac{d-1}{2})\Gamma(\frac{1}{2})}{(\frac{\kappa}{2})^{d/2-1}} \Big\{ I_{d/2}(\kappa) - \lambda \Big\{ I_{d/2-1}(\kappa) - \frac{d-1}{\kappa} I_{d/2}(\kappa) \Big\} \Big\} \\
&= \frac{I_{d/2}(\kappa) - \lambda \Big\{ I_{d/2-1}(\kappa) - \frac{d-1}{\kappa} I_{d/2}(\kappa) \Big\}}{I_{d/2-1}(\kappa) - \lambda I_{d/2}(\kappa)} \\
&= \frac{\frac{I_{d/2}(\kappa)}{I_{d/2-1}(\kappa)} - \lambda \Big\{ 1 - \frac{d-1}{\kappa} \frac{I_{d/2}(\kappa)}{I_{d/2-1}(\kappa)} \Big\}}{1 - \lambda \frac{I_{d/2}(\kappa)}{I_{d/2-1}(\kappa)}}.
\end{aligned}$$

Moreover, it is seen that

$$\begin{aligned}
\rho_2 = E_{y_1}(t^2) &= C_d^{y_1}(\kappa,\lambda) \int_{-1}^{1} t^2(1-\lambda t)(1-t^2)^{(d-3)/2} e^{\kappa t} dt \\
&= C_d^{y_1}(\kappa,\lambda) \{ \int_{-1}^{1} t^2(1-t^2)^{(d-3)/2} e^{\kappa t} - \lambda \int_{-1}^{1} t^3(1-t^2)^{(d-3)/2} e^{\kappa t} dt \} \\
&= C_d^{y_1}(\kappa,\lambda) \frac{\Gamma(\frac{d-1}{2})\Gamma(\frac{1}{2})}{(\frac{\kappa}{2})^{d/2-1}} \Big\{ I_{d/2-1}(\kappa) - \frac{d-1}{\kappa} I_{d/2}(\kappa) \\
&\quad -\lambda \Big\{ I_{d/2}(\kappa) - \frac{d-1}{\kappa} I_{d/2+1}(\kappa) \Big\} \Big\} \\
&= \frac{I_{d/2-1}(\kappa) - \frac{d-1}{\kappa} I_{d/2}(\kappa) - \lambda \Big\{ I_{d/2}(\kappa) - \frac{d-1}{\kappa} I_{d/2+1}(\kappa) \Big\}}{I_{d/2-1}(\kappa) - \lambda I_{d/2}(\kappa)} \\
&= 1 - \frac{d-1}{\kappa} \frac{I_{d/2}(\kappa) - \lambda I_{d/2+1}(\kappa)}{I_{d/2-1}(\kappa) - \lambda I_{d/2}(\kappa)} \\
&= 1 - \frac{d-1}{\kappa} \frac{I_{d/2}(\kappa) - \lambda \{ I_{d/2-1}(\kappa) - \frac{d}{\kappa} I_{d/2}(\kappa) \}}{I_{d/2-1}(\kappa) - \lambda I_{d/2}(\kappa)} \\
&= 1 + \frac{d-1}{\kappa} \frac{\lambda - (1 + \frac{d\lambda}{\kappa}) \frac{I_{d/2}(\kappa)}{I_{d/2-1}(\kappa)}}{1 - \lambda \frac{I_{d/2}(\kappa)}{I_{d/2-1}(\kappa)}}.
\end{aligned}$$

To prove the convergency of ρ_1 and ρ_2 for small values of κ, we first show an important equality. Using the transformation $u = (t+1)/2$, and invoking the integral by part, it can be shown that, for any $m > 0$,

$$\begin{aligned}\int_{-1}^{1} t^m (1-\lambda t)(1-t^2)^{\nu-1} dt &= 2^{2\nu-1} \int_0^1 (2u-1)^m (1+\lambda-2\lambda u)(1-u)^{\nu-1} u^{\nu-1} du \\ &= \sum_{i=0}^{m} 2^{2\nu+i-1} (-1)^{m-i} \Big\{ (1+\lambda) \int_0^1 (1-u)^{\nu-1} u^{\nu+i-1} du \\ &- 2\lambda \int_0^1 (1-u)^{\nu-1} u^{\nu+i} du \Big\} \\ &= \sum_{i=0}^{m} 2^{2\nu+i-1} (-1)^{m-i} \frac{\Gamma(\nu)\Gamma(\nu+i)}{\Gamma(2\nu+i)} (1+\lambda-2\lambda \frac{\nu+i}{2\nu+i}).\end{aligned}$$

It is seen that the functions ρ_1 and ρ_2 are equivalent to the particular version of the above integral for $m = 0, 1, 2$ while κ is getting close to zero. In particular, to recall Equation (5.6) and the integral representation of the normalizing constant given by Equation (2.3), we can write

$$\begin{aligned}\lim_{\kappa\to 0} \rho_1 &= \lim_{\kappa\to 0} \frac{\int_{-1}^{1} t(1-\lambda t)(1-t^2)^{(d-3)/2} e^{\kappa t} dt}{\int_{-1}^{1} (1-\lambda t)(1-t^2)^{(d-3)/2} e^{\kappa t} dt} \\ &= \frac{\int_{-1}^{1} t(1-\lambda t)(1-t^2)^{(d-3)/2} dt}{\int_{-1}^{1} (1-\lambda t)(1-t^2)^{(d-3)/2} dt} = \frac{-\lambda}{d}.\end{aligned}$$

Similarly, we have

$$\begin{aligned}\lim_{\kappa\to 0} \rho_2 &= \lim_{\kappa\to 0} \frac{\int_{-1}^{1} t^2(1-\lambda t)(1-t^2)^{(d-3)/2} e^{\kappa t} dt}{\int_{-1}^{1} (1-\lambda t)(1-t^2)^{(d-3)/2} e^{\kappa t} dt} \\ &= \frac{\int_{-1}^{1} t^2(1-\lambda t)(1-t^2)^{(d-3)/2} dt}{\int_{-1}^{1} (1-\lambda t)(1-t^2)^{(d-3)/2} dt} = \frac{d+1-2\lambda}{d}.\end{aligned}$$