Electronic Journal of Statistics
Vol. 15 (2021) 6344–6381
ISSN: 1935-7524
https://doi.org/10.1214/21-EJS1946

# The entropy based goodness of fit tests for generalized von Mises-Fisher distributions and beyond

**Nikolai Leonenko**

*School of Mathematics, Cardiff University, Senghennydd Road, Cardiff, Wales, UK, CF24 4AG.*
*e-mail:* leonenkon@cardiff.ac.uk

**Vitalii Makogin***

*Institute of Stochastics, Ulm University, Ulm, 08069 Germany*
*e-mail:* vitalii.makogin@uni-ulm.de

**and**

**Mehmet Siddik Cadirci**

*School of Mathematics, Cardiff University, Senghennydd Road, Cardiff, Wales, UK, CF24 4AG.*
*e-mail:* cadircims@cardiff.ac.uk

**Abstract:** We introduce some new classes of unimodal rotational invariant directional distributions, which generalize von Mises–Fisher distribution. We propose three types of distributions, one of which represents axial data. For each new type we provide formulae and short computational study of parameter estimators by the method of moments and the method of maximum likelihood. The main goal of the paper is to develop the goodness of fit test to detect that sample entries follow one of the introduced generalized von Mises–Fisher distribution based on the maximum entropy principle. We use $k$th nearest neighbour distances estimator of Shannon entropy and prove its $L^2$-consistency. We examine the behaviour of the test statistics, find critical values and compute power of the test on simulated samples. We apply the goodness of fit test to local fiber directions in a glass fibre reinforced composite material and detect the samples which follow axial generalized von Mises–Fisher distribution.

**MSC2020 subject classifications:** Primary 62H11; secondary 28D20.
**Keywords and phrases:** Directional distribution, generalized von Mises–Fisher distribution, goodness of fit test, entropy estimation, maximum entropy principle, nearest neighbour estimator.

Received April 2021.

## 1. Introduction

Directional distributions characterize randomness in unit vectors (directions). Spherical data sets appear in a wide range of problems arising from Earth sci-

*The research is supported by DFG Grant 390879134.

ences [44], biology [39], and material science [16]. Directional data are important in cosmology and astrophysics, for instance, in results of the Laser Interferometer Gravitational-Wave Observatory [1] and the Alpha Magnetic Spectrometer on the International Space Station [2]. Further applications and modern state of the art on statistical theory on directional data can be found in Ley and Verdebout [30], Pewsey and García-Portugués [43] and references therein.

In this paper, we consider some classes of random unit vectors with values on sphere $\mathbb{S}^{d-1} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\}$, which have the absolutely continuous directional distributions with respect to the uniform distribution on $\mathbb{S}^{d-1}$. We denote by $\mathbf{a}^{\mathrm{T}}\mathbf{b}$ the scalar product of vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ and by $\|\mathbf{a}\|$ the Euclidean norm of $\mathbf{a} \in \mathbb{R}^d$.

The von Mises–Fisher distribution is a fundamental isotropic distribution which is widely used in directional statistics [e.g. 36, p. 168]. It belongs to the exponential family of distributions, is rotational invariant and has a density proportional to $\exp(\kappa\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}), \mathbf{x} \in \mathbb{S}^{d-1}$, such that random vectors are concentrated with rate $\kappa \in \mathbb{R}$ along direction $\boldsymbol{\mu} \in \mathbb{S}^{d-1}$. The rotational invariant family of distributions is actively studied nowadays by Cutting, Paindaveine and Verdebout [12], Duerinckx and Ley [17], García-Portugués, Paindaveine and Verdebout [23], Paindaveine and Verdebout [40].

Among several important properties of the von Mises–Fisher distribution we focus on maximum entropy characterization, that is, the von Mises–Fisher distribution has maximum entropy in the class of continuous distributions on $\mathbb{S}^{d-1}$ with a given value of $\mathbb{E}(X)$ [see 35]. The von Mises-Fisher distribution is widely used for analysis of neutrino arrival directions recorded by the IceCube Neutrino Observatory, see e.g. [11, 14, 31] and arrival directions of ultrahigh energy cosmic rays recorded by the Pierre Auger Observatory, see e.g. [3, 9, 27].

There are several generalizations, including the Fisher–Bingham distribution with a density proportional to $\exp(\kappa\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x} + \mathbf{x}^{\mathrm{T}}A\mathbf{x}), \mathbf{x} \in \mathbb{S}^{d-1}$ [see 35], and the generalized von Mises–Fisher distribution of order $k$ (GvMFk) introduced in Gatto and Jammalamadaka [24], having the density proportional to $\exp\left(\sum_{j=1}^{k} \kappa_j (\boldsymbol{\mu}_j^{\mathrm{T}}\mathbf{x})^{r_j}\right)$, where $\boldsymbol{\mu}_j \in \mathbb{S}^{d-1}, \kappa_j \in \mathbb{R}, r_j \in \mathbb{N}$ $(j = 1, \ldots, k)$, and $r_1 \leq \ldots \leq r_k$.

In this paper, we introduce a new generalization of the von Mises–Fisher distribution, which stays in the exponential family and is rotational invariant with one mode. In contrast to the generalized von Mises–Fisher distribution of order $k$ with integer power $r \in \mathbb{N}$, we consider densities with arbitrary positive power $r \in \mathbb{R}_+$. The motivation of such choice is to provide the analogue of a generalized Gaussian distribution for random vectors on the unit sphere. To do so we introduce the following three types of distributions of order $\alpha \in \mathbb{R}_+$, whose densities $f$ are proportional to

$$\begin{aligned}
\text{Type I, } \mathrm{GvMF}_{1,d}(\alpha, \kappa, \boldsymbol{\mu}): &\quad f(\mathbf{x}) \propto \exp\left(\frac{\kappa}{\alpha}(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x})^{<\alpha>}\right), \mathbf{x} \in \mathbb{S}^{d-1}, \\
\text{Type II, } \mathrm{GvMF}_{2,d}(\alpha, \kappa, \boldsymbol{\mu}): &\quad f(\mathbf{x}) \propto \exp\left(\frac{\kappa}{2^{\alpha}\alpha}\|\mathbf{x} - \boldsymbol{\mu}\|^{2\alpha}\right), \mathbf{x} \in \mathbb{S}^{d-1}, \\
\text{Axial Type, } \mathrm{GvMF}_{3,d}(\alpha, \kappa, \boldsymbol{\mu}): &\quad f(\mathbf{x}) \propto \exp\left(\frac{\kappa}{\alpha}|\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}|^{\alpha}\right), \mathbf{x} \in \mathbb{S}^{d-1},
\end{aligned}$$

where $\kappa > 0$ is a concentration parameter, and $\boldsymbol{\mu} \in \mathbb{S}^{d-1}$ is a mean direction parameter. In the paper we denote by $x^{<\alpha>} = |x|^{\alpha}\mathrm{sgn}(x), x \in \mathbb{R}$.

Apart from studying the properties, simulations and parameter estimation for distributions $\mathrm{GvMF}_{j,d}$ $(j = 1, 2, 3)$, we develop the goodness of fit test based on the estimation of the Shannon entropy and independent identically distributed (i.i.d.) sample. These tests exploit the maximum entropy principle, which is also proved in the paper as the spherical analogue of the results by Lutwak, Yang and Zhang [34].

To do so, we employ the entropy estimators $\hat{H}_{N,k}$ derived from $k$th nearest neighbour distances. Starting from the pioneering paper [28], which proves by direct probability methods the consistency of $\hat{H}_{N,1}$ for random vectors with values in Euclidean space, many authors considered extending the class of admissible distributions and improved the convergence of $\hat{H}_{N,k}$, see Berrett, Samworth and Yuan [6], Bulinski and Dimitrov [8], Delattre and Fournier [13], Evans [18], Evans, Jones and Schmidt [19], Gao, Oh and Viswanath [22], Goria et al. [25], Leonenko, Pronzato and Savani [29], and the references therein. In [32, 38], the $k$th nearest neighbour entropy estimation is generalized for hyperspherical distributions.

Unlike the above mentioned works, the limit theory for point processes with a fixed $k$ allows to prove the $L^p$-consistency of functionals of $k$th nearest neighbour distances for a wider class of distributions. The nearest neighbours method of estimation of the Shannon entropy for manifolds, including spheres, was developed by Penrose and Yukich [42]. In the present paper, we continue their work and prove the $L^2$-consistency of $\hat{H}_{N,k}$, as $N \to \infty$ with arbitrary $k \geq 1$ and for a random vector on a Riemannian manifold if its density is bounded and has compact support, see Theorem 4.6. Therefore, we show that $\hat{H}_{N,k}$ is a consistent estimator for the samples from the introduced generalized von Mises–Fisher distributions.

From the recent papers, we mention [6], where the efficient entropy estimation is provided via the weighted $k$th nearest neighbour distances with $k = k_N$ depending on sample size $N$. Moreover, Berrett and Samworth [5] introduced a non-parametric entropy based test of independence for multidimensional data. Lund and Jammalamadaka [33] considered the entropy based test of goodness of fit for the von Mises distribution on the circle and use a different entropy estimate. Our study is motivated, particularly, by the work of Cadirci et al. [10], where the entropy based goodness of fit test for generalized Gaussian distribution is given.

We verify our theoretical results by computational study on simulated samples and show the inflation of variances of $\hat{H}_{N,k}$ as $k$ grows, which confirms the conclusion of [6]. Moreover, we detect the evidence of generalized von Mises–Fisher distributions in real world data by the presented entropy based goodness of fit test. Particularly, we find the evidence in 3D images of a glass fibre reinforced composite material, where fiber directions follow a generelized von Mises–Fisher distribution of axial type.

The manuscript is organized as follows. In Section 2, we revise the basic facts for von Mises-Fisher distribution. In Section 3, we introduce our three

types of generalized von Mises-Fisher distribution and compute their moments. Section 4 is devoted to the Shannon entropy of generalized von Mises-Fisher distributions and we show the maximum entropy principle for them in Section 4.1. Then we discuss the statistical estimation of an entropy and prove the $L^2$ convergence of $k$th nearest neighbour estimator for random variables on compact manifolds (Section 4.2). In Section 5, we formulate the maximum likelihood estimators (Section 5.1) and estimators by the method of moments for distributions $\mathrm{GvMF}_{1,d}, \mathrm{GvMF}_{2,d}$, and $\mathrm{GvMF}_{3,d}$ (Section 5.2). In Section 6 we develop goodness of fit tests based on the maximum entropy principle for the introduced distributions. Results of numerical experiments on simulated samples are given in Section 7. We present the method of simulations in Section 7.1, entropy estimation in Section 7.2, and study of the test statistics in Section 7.3. In Section 8, we detect the generalized von Mises-Fisher distributions in a real data set. The study of the tests' asymptotic distribution are given in Section 9. The numerical results of parameter estimation and some auxiliary material are given in the Appendix.

## 2. Preliminaries

In this section we provide some known facts needed for the sequel. Let $\sigma(d\mathbf{x})$ be spherical measure on the sphere $\mathbb{S}^{d-1}$. It can be written in polar coordinates $\mathbf{x} = (1, \mathbf{u}), \mathbf{u} \in \mathbb{S}^{d-1}$ as $\sigma(d\mathbf{x}) = 2^{-1}\pi^{-d/2}\Gamma(d/2)\, d\mathbf{u}$. Further, we use Lemma 2.5.1 from [20] for computation of integrals with respect to $\sigma$. Namely, let $g : \mathbb{R} \to \mathbb{R}_+$ be a non-negative Borel function and $\mathbf{a} \in \mathbb{S}^{d-1}$, then

$$\int_{\mathbf{x}^{\mathrm{T}}\mathbf{x}=1} g(\mathbf{a}^{\mathrm{T}}\mathbf{x})\sigma(d\mathbf{x}) = \frac{2\pi^{(d-1)/2}}{\Gamma\left((d-1)/2\right)} \int_{-1}^{1} g(y)(1-y^2)^{\frac{d-3}{2}} dy. \tag{1}$$

We take $\kappa \in \mathbb{R}$, $\boldsymbol{\mu} \in \mathbb{S}^{d-1}, d \geq 2$ and consider further the probability densities with respect to the measure $\sigma$.

**Definition 2.1.** *A unit random vector* $\mathbf{X}$ *has the* $(d-1)$*-dimensional von Mises-Fisher distribution* $\mathrm{vMF}_{1,d}(\boldsymbol{\mu}, \kappa)$ *if its probability density function is* $f_{\mathbf{X}}(\mathbf{x}) = (\kappa/2)^{d/2-1}\left(2\pi^{d/2} I_{d/2-1}(\kappa)\right)^{-1} \exp(\kappa \boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}), \mathbf{x} \in \mathbb{S}^{d-1}$ *where* $I_\nu$ *is the modified Bessel function of order* $\nu \geq 0$*, see [e.g. 26, (A5)]*

In the case $d = 3$ von Mises-Fisher distribution $M_{1,3}(\boldsymbol{\mu}, \kappa)$ is called *Fisher distribution* and its density simplifies to $\kappa/(4\pi \sinh \kappa) \exp(\kappa \boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}), \mathbf{x} \in \mathbb{S}^2$.

The density of $\mathrm{vMF}_{1,d}$ can be written in the alternative form. We say that a random vector $\mathbf{X}$ has the *von Mises-Fisher distribution* $\mathrm{vMF}_{2,d}(\boldsymbol{\mu}, \kappa)$ if its density function is $f_{\mathbf{X}}(\mathbf{x}) = e^{\kappa}\left(\kappa/2\right)^{d/2-1}\left(2\pi^{d/2} I_{d/2-1}(\kappa)\right)^{-1} \exp\left(-\kappa\|\mathbf{x} - \boldsymbol{\mu}\|^2/2\right)$, $\mathbf{x} \in \mathbb{S}^{d-1}$. Indeed, $\frac{\kappa}{2}\|\mathbf{x} - \boldsymbol{\mu}\|^2 = \frac{\kappa}{2}\|\mathbf{x}\|^2 + \frac{\kappa}{2}\|\boldsymbol{\mu}\|^2 - \kappa\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x} = \kappa - \kappa\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}$ for $\mathbf{x}, \boldsymbol{\mu} \in \mathbb{S}^{d-1}$.

Let us recall the standard directional statistics.

**Definition 2.2.** *Let* $\mathbf{X}$ *be random vector with values in* $\mathbb{S}^{d-1}$ *and* $\mathbb{E}\mathbf{X} \neq \mathbf{0}$*. The mean direction of* $\mathbf{X}$ *is the vector* $\mathbb{E}\mathbf{X}/\|\mathbb{E}\mathbf{X}\|$*, while the mean resultant length is* $\|\mathbb{E}\mathbf{X}\|$.

The mean resultant length is invariant and the mean direction is equivariant under rotation. Formally, let $\mathbf{U} \in SO(d)$ be a rotation matrix, then $\|\mathbb{E}\mathbf{U}\mathbf{X}\| = \|\mathbb{E}\mathbf{X}\|$ and $\mathbb{E}\mathbf{U}\mathbf{X}/\|\mathbb{E}\mathbf{U}\mathbf{X}\| = \mathbf{U}\mathbb{E}\mathbf{X}/\|\mathbb{E}\mathbf{X}\|$.

Consider the class of distributions on $\mathbb{S}^{d-1}$ with rotational symmetry, that is their distribution functions have a form $f(\mathbf{x}) = g(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x})$, $\mathbf{x}, \boldsymbol{\mu} \in \mathbb{S}^{d-1}$, e.g. [7]. Such random vectors $\mathbf{X}$ posses a *tangent-normal decomposition*

$$\mathbf{X} = (\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X})\boldsymbol{\mu} + \sqrt{1-(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X})^2}\mathbf{Y}, \tag{2}$$

where $\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X}$ and $\mathbf{Y}$ are independent, $\boldsymbol{\mu} \perp \mathbf{Y}$, and $\mathbf{Y}$ is uniformly distributed on the tangent space $\mathbb{S}^{d-1}_{\boldsymbol{\mu}} := \{\mathbf{y} \in \mathbb{S}^{d-1} | \boldsymbol{\mu}^{\mathrm{T}}\mathbf{y} = 0\}$. It follows from (2), that the mean resultant length is $\|\mathbb{E}\mathbf{X}\| = \mathbb{E}[\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X}]$ and the mean direction equals $\boldsymbol{\mu}$.

## 3. Generalized von Mises-Fisher distributions

In this section we introduce our generalizations of the von Mises-Ficher distribution. We call $\kappa \in \mathbb{R}$ a concentration parameter and $\boldsymbol{\mu} \in \mathbb{S}^{d-1}$ a mean direction parameter.

**Definition 3.1.** *A unit random vector* $\mathbf{X}$ *has the* $(d-1)$*-dimensional I-type generalized von Mises-Fisher distribution* $\mathrm{GvMF}_{1,d}(\alpha, \kappa, \boldsymbol{\mu})$ *of order* $\alpha > 0$ *if its probability density function is*

$$f_{\mathbf{X}}(\mathbf{x}) = c_{1,d}(\kappa, \alpha) \exp\left(\frac{\kappa}{\alpha}(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x})^{<\alpha>}\right), \mathbf{x} \in \mathbb{S}^{d-1}, \tag{3}$$

*where*

$$c^{-1}_{1,d}(\kappa, \alpha) = \frac{2\pi^{(d-1)/2}}{\Gamma\left((d-1)/2\right)} \int_0^1 \left(e^{\frac{\kappa}{\alpha}y^{\alpha}} + e^{-\frac{\kappa}{\alpha}y^{\alpha}}\right)(1-y^2)^{\frac{d-3}{2}} dy. \tag{4}$$

As an analogue of von Mises-Fisher distribution in the form $\mathrm{vMF}_{2,d}$, we introduce the following class.

**Definition 3.2.** *A unit random vector* $\mathbf{X}$ *has the* $(d-1)$*-dimensional II-type generalized von Mises-Fisher distribution* $\mathrm{GvMF}_{2,d}(\alpha, \kappa, \boldsymbol{\mu})$ *of order* $\alpha > 0$ *if its probability density function is*

$$f_{\mathbf{X}}(\mathbf{x}) = c_{2,d}(\kappa, \alpha) \exp\left(-\frac{\kappa}{2^{\alpha}\alpha}\|\mathbf{x} - \boldsymbol{\mu}\|^{2\alpha}\right), \mathbf{x} \in \mathbb{S}^{d-1}, \tag{5}$$

*where*

$$c^{-1}_{2,d}(\kappa, \alpha) = \frac{2\pi^{(d-1)/2}}{\Gamma\left((d-1)/2\right)} \int_0^1 \left(e^{-\frac{\kappa}{\alpha}(1-y)^{\alpha}} + e^{-\frac{\kappa}{\alpha}(1+y)^{\alpha}}\right)(1-y^2)^{\frac{d-3}{2}} dy. \tag{6}$$

In the case of $\alpha = 1$, the introduced distributions $\mathrm{GvMF}_{1,d}$ and $\mathrm{GvMF}_{2,d}$ become the von Mises-Fisher distributions $\mathrm{vMF}_{1,d}$ and $\mathrm{vMF}_{2,d}$ respectively.

If we do not distinguish opposite directions we deal with axes. Commonly used technique in this case is to consider symmetric density functions $f$ such that $f(\mathbf{x}) = f(-\mathbf{x}), \mathbf{x} \in \mathbb{S}^{d-1}$. Since our motivation is to stay in the class of rotational invariant densities and to generalize the von Mises-Fisher distribution, we propose the following model for an axial data.

**Definition 3.3.** *A unit random vector* $\mathbf{X}$ *has the* $(d-1)$*-dimensional axial generalized von Mises-Fisher distribution* $\mathrm{GvMF}_{3,d}(\alpha,\kappa,\boldsymbol{\mu})$ *(or distribution of axial type) of order* $\alpha>0$ *if its probability density function is*

$$f_{\mathbf{X}}(\mathbf{x})=c_{3,d}(\kappa,\alpha)\exp\left(\frac{\kappa}{\alpha}|\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}|^{\alpha}\right),\mathbf{x}\in\mathbb{S}^{d-1}, \tag{7}$$

*where*

$$c_{3,d}^{-1}(\kappa,\alpha)=\frac{4\pi^{(d-1)/2}}{\Gamma\left((d-1)/2\right)}\int_0^1 e^{\frac{\kappa}{\alpha}y^{\alpha}}(1-y^2)^{\frac{d-3}{2}}dy. \tag{8}$$

*Remark* 3.1. We find $c_{j,d},j=1,2,3$ by checking $\int_{\mathbb{S}^{d-1}}f_{\mathbf{X}}(\mathbf{x})\sigma(d\mathbf{x})=1$. for example, the constant $c_{2,d}^{-1}(\kappa,\alpha)$ equals

$$\begin{aligned}&\int_{\mathbb{S}^{d-1}}\exp\left(-\frac{\kappa}{2^{\alpha}\alpha}\|\mathbf{x}-\boldsymbol{\mu}\|^{2\alpha}\right)\sigma(d\mathbf{x})=\int_{\mathbb{S}^{d-1}}\exp\left(-\frac{\kappa}{\alpha}\left(1-\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}\right)^{\alpha}\right)\sigma(d\mathbf{x})\\&\overset{(1)}{=}\frac{2\pi^{(d-1)/2}}{\Gamma\left((d-1)/2\right)}\int_{-1}^1\exp\left(-\frac{\kappa}{\alpha}\left(1-y\right)^{\alpha}\right)(1-y^2)^{\frac{d-3}{2}}dy.\end{aligned}$$

*Remark* 3.2. As usual for axial distributions, parameter $\boldsymbol{\mu}$ is defined up to a sign, in a sense that $\mathrm{GvMF}_{3,d}(\alpha,\kappa,\boldsymbol{\mu})$ and $\mathrm{GvMF}_{3,d}(\alpha,\kappa,-\boldsymbol{\mu})$ are equal. In the case $\alpha=2$, the generalized von Mises-Fisher distribution of axial type reduces to the Watson distribution.

*Remark* 3.3. For $d=3$ and $\alpha=1$, one can represent an expansion of the von Mises-Fisher $\mathrm{vMF}_{1,3}(\boldsymbol{\mu},\kappa)$ density into the series of orthogonal functions $Y_l^m,-l\le m\le l,l=0,1,2,\ldots$, on the sphere (real spherical harmonics, see [e.g. 26, p. 437], [e.g. 37, S. 13.2]). For example, [26, (5)] gives

$$f(\mathbf{x})=\sum_{l=0}^{\infty}\sqrt{\frac{2l+1}{4\pi}}\frac{I_{l+1/2}(\kappa)}{I_{1/2}(\kappa)}Y_l^0(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}),\mathbf{x}\in\mathbb{S}^2,$$

which can be used potentially for computational purposes. However, it is difficult for general $\alpha$ to express coefficients of expansions in terms of some known special functions (this is true even for $\alpha=2$, see formulae (7) and (8) in [26]).

Let us consider the moments characteristics of $\mathrm{GvMF}_{j,d}(\alpha,\kappa,\boldsymbol{\mu}),j=1,2,3$ distributions. Denote by

$$A_1(\kappa,\alpha,\beta)=\int_0^1 e^{\frac{\kappa}{\alpha}y^{\alpha}}y^{\beta}(1-y^2)^{\frac{d-3}{2}}dy, \tag{9}$$

$$A_2(\kappa,\alpha,\beta)=\int_0^2 e^{-\frac{\kappa}{\alpha}y^{\alpha}}(2-y)^{\frac{d-3}{2}}y^{\frac{d-3}{2}+\beta}dy. \tag{10}$$

**Proposition 3.1.** *Let* $\beta\ge0$ *and* $\mathbf{X}\sim\mathrm{GvMF}_{1,d}(\alpha,\kappa,\boldsymbol{\mu})$*, then*

$$\mathbb{E}\left((\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X})^{<\beta>}\right)=\frac{A_1(\kappa,\alpha,\beta)-A_1(-\kappa,\alpha,\beta)}{A_1(\kappa,\alpha,0)+A_1(-\kappa,\alpha,0)}. \tag{11}$$

*Proof.* Let $f$ be the density of the form (3), then

$$\begin{aligned}
&\mathbb{E}\left((\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X})^{<\beta>}\right) = \int_{\mathbb{S}^{d-1}} (\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x})^{<\beta>} f(\mathbf{x})\sigma(d\mathbf{x}) \\
&\overset{(1)}{=} c_{1,d}(\kappa,\alpha)\frac{2\pi^{(d-1)/2}}{\Gamma\left((d-1)/2\right)}\int_{-1}^{1} y^{<\beta>}\exp\left(\frac{\kappa}{\alpha}y^{<\alpha>}\right)(1-y^2)^{\frac{d-3}{2}}dy \\
&= c_{1,d}(\kappa,\alpha)\frac{2\pi^{(d-1)/2}}{\Gamma\left((d-1)/2\right)}\int_{0}^{1}\left(e^{\frac{\kappa}{\alpha}y^{\alpha}} - e^{-\frac{\kappa}{\alpha}y^{\alpha}}\right)y^{\beta}(1-y^2)^{\frac{d-3}{2}}dy \\
&= \frac{\int_0^1\left(e^{\frac{\kappa}{\alpha}y^{\alpha}} - e^{-\frac{\kappa}{\alpha}y^{\alpha}}\right)y^{\beta}(1-y^2)^{\frac{d-3}{2}}dy}{\int_0^1\left(e^{\frac{\kappa}{\alpha}y^{\alpha}} + e^{-\frac{\kappa}{\alpha}y^{\alpha}}\right)(1-y^2)^{\frac{d-3}{2}}dy}.
\end{aligned}$$

□

**Proposition 3.2.** *Let $\beta \geq 0$ and $\mathbf{X} \sim \mathrm{GvMF}_{2,d}(\alpha,\kappa,\boldsymbol{\mu})$, then*

$$\mathbb{E}\|\mathbf{X}-\boldsymbol{\mu}\|^{2\beta} = 2^{\beta}A_2(\kappa,\alpha,\beta)/A_2(\kappa,\alpha,0). \tag{12}$$

*Proof.* Let $f$ be the density of the form (5), then

$$\begin{aligned}
&\mathbb{E}\|\mathbf{X}-\boldsymbol{\mu}\|^{2\beta} = \int_{\mathbb{S}^{d-1}}\|\mathbf{x}-\boldsymbol{\mu}\|^{2\beta}f(\mathbf{x})\sigma(d\mathbf{x}) = \int_{\mathbb{S}^{d-1}}(2-2\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x})^{\beta}f(\mathbf{x})\sigma(d\mathbf{x}) \\
&\overset{(1)}{=} c_{2,d}(\kappa,\alpha)\frac{2\pi^{(d-1)/2}}{\Gamma\left((d-1)/2\right)}\int_{-1}^{1}(2-2y)^{\beta}\exp\left(-\frac{\kappa}{\alpha}(1-y)^{\alpha}\right)(1-y^2)^{\frac{d-3}{2}}dy \\
&= 2^{\beta}\frac{\int_{-1}^{1}(1-y)^{\beta}e^{-\frac{\kappa}{\alpha}(1-y)^{\alpha}}(1-y^2)^{\frac{d-3}{2}}dy}{\int_{-1}^{1}e^{-\frac{\kappa}{\alpha}(1-y)^{\alpha}}(1-y^2)^{\frac{d-3}{2}}dy} = 2^{\beta}\frac{\int_0^2 e^{-\frac{\kappa}{\alpha}z^{\alpha}}(2-z)^{\frac{d-3}{2}}z^{\frac{d-3}{2}+\beta}dz}{\int_0^2 e^{-\frac{\kappa}{\alpha}z^{\alpha}}(2-z)^{\frac{d-3}{2}}z^{\frac{d-3}{2}}dz}.
\end{aligned}$$

□

Similarly, we get the moments for the distribution $\mathrm{GvMF}_{3,d}(\alpha,\kappa,\boldsymbol{\mu})$.

**Proposition 3.3.** *Let $\beta \geq 0$ and $\mathbf{X} \sim \mathrm{GvMF}_{3,d}(\alpha,\kappa,\boldsymbol{\mu})$, then*

$$\mathbb{E}\left(|\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X}|^{\beta}\right) = A_1(\kappa,\alpha,\beta)/A_1(\kappa,\alpha,0). \tag{13}$$

*Remark* 3.4. Particularly, the mean direction of $\mathbf{X}_j \sim \mathrm{GvMF}_{j,d}(\alpha,\kappa,\boldsymbol{\mu})$ $(j = 1,2)$ is $\boldsymbol{\mu}$. Note that, the mean direction of $\mathbf{X}_3 \sim \mathrm{GvMF}_{3,d}(\alpha,\kappa,\boldsymbol{\mu})$ is not defined and its mean resultant length $\|\mathbb{E}(\mathbf{X}_3)\| = \mathbb{E}(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X}_3)$ equals 0. The mean resultant length of $X_1$ equals $\|\mathbb{E}(\mathbf{X}_1)\| = \mathbb{E}(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X}_1)$ and computed by (11) with $\beta = 1$. The mean resultant length of $\mathbf{X}_2$ equals $\|\mathbb{E}(\mathbf{X}_2)\| = \mathbb{E}(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X}_2) = 1 - A_2(\kappa,\alpha,1)/A_2(\kappa,\alpha,0)$.

## 4. Entropy

In this section we find the entropy of generalized von Mises–Fisher distributions, and show the maximum entropy principle for them. Then we discuss the statistical estimation of an entropy and prove the $L^2$-convergence of the $k$th nearest neigbour estimator for random variables on compact manifolds.

### *4.1. Maximum entropy principle for generalized von-Mises Fisher distributions*

Recall that the entropy of a continuous random vector $\mathbf{X} \in \mathbb{S}^{d-1}$ with a density $f$ is $H(\mathbf{X}) = -\int_{\mathbb{S}^{d-1}} (\log f(\mathbf{x}))f(\mathbf{x})\sigma(d\mathbf{x})$. For a density version $f$ we denote its support by $\text{supp}f = \{\mathbf{x} \in \mathbb{S}^{d-1} : f(\mathbf{x}) > 0\}$. Clearly, the integral in $H(\mathbf{X})$ is taken over $\text{supp}f$.

**Theorem 4.1.** *Let* $\mathbf{X}_j \sim \text{GvMF}_{j,d}(\alpha, \kappa, \boldsymbol{\mu})$, $j = 1, 2, 3$, *then*

$$H(\mathbf{X}_1) = -\log c_{1,d}(\kappa, \alpha) - \frac{\kappa}{\alpha}\mathbb{E}\left((\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X}_1)^{<\alpha>}\right), \tag{14}$$

$$H(\mathbf{X}_2) = -\log c_{2,d}(\kappa, \alpha) + \frac{\kappa}{2^{\alpha}\alpha}\mathbb{E}\|\mathbf{X}_2 - \boldsymbol{\mu}\|^{2\alpha} \tag{15}$$

$$H(\mathbf{X}_3) = -\log c_{3,d}(\kappa, \alpha) - \frac{\kappa}{\alpha}\mathbb{E}|\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X}_3|^{\alpha}. \tag{16}$$

*Proof.* Let $\mathbf{X}_1$ have density $f_1$, then the entropy of $\mathbf{X}_1$ equals

$$-\int_{\mathbb{S}^{d-1}} (\log f_1(\mathbf{x}))f_1(\mathbf{x})\sigma(d\mathbf{x}) = -\log c_{1,d}(\kappa, \alpha)\int_{\mathbb{S}^{d-1}} f_1(\mathbf{x})\sigma(d\mathbf{x})$$
$$-\frac{\kappa}{\alpha}\int_{\mathbb{S}^{d-1}} (\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x})^{<\alpha>}f_1(\mathbf{x})\sigma(d\mathbf{x}) = -\log c_{1,d}(\kappa, \alpha) - \frac{\kappa}{\alpha}\mathbb{E}\left((\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X}_1)^{<\alpha>}\right).$$

The cases of $\mathbf{X}_2$ and $\mathbf{X}_3$ are similar. □

**Theorem 4.2.** *Let a unit random vector* $\mathbf{Z} \in \mathbb{S}^{d-1}$ *have a generalized von Mises-Fisher distribution* $\text{GvMF}_{1,d}(\alpha, \kappa, \boldsymbol{\mu})$. *Then* $\mathbf{Z}$ *has the maximum entropy value over all continuous random variables* $\mathbf{X}$ *on* $\mathbb{S}^{d-1}$ *with*

$$\mathbb{E}\left((\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X})^{<\alpha>}\right) = \mathbb{E}\left((\boldsymbol{\mu}^{\mathrm{T}}\mathbf{Z})^{<\alpha>}\right). \tag{17}$$

*Proof.* Let $\mathbf{X}$ be a random unit vector on $\mathbb{S}^{d-1}, d \geq 2$ such that (17) holds true. Let $f$ and $f^*$ be the densities of $\mathbf{X}$ and $\mathbf{Z}$ respectively. By Jensen's inequality,

$$\int_{\mathbb{S}^{d-1}} f(\mathbf{x})\log f^*(\mathbf{x})\sigma(d\mathbf{x}) - \int_{\mathbb{S}^{d-1}} f(\mathbf{x})\log f(\mathbf{x})\sigma(d\mathbf{x})$$
$$= \int_{\mathbb{S}^{d-1}} f(\mathbf{x})\log\frac{f^*(\mathbf{x})}{f(\mathbf{x})}\sigma(d\mathbf{x}) \leq \log\left(\int_{\mathbb{S}^{d-1}} f(\mathbf{x})\frac{f^*(\mathbf{x})}{f(\mathbf{x})}\sigma(d\mathbf{x})\right) = 0$$

with equality if and only if $f = f^*$ almost everywhere with respect to $\sigma$. So, $H(\mathbf{X}) = -\int_{\mathbb{S}^{d-1}} f(\mathbf{x})\log f(\mathbf{x})\sigma(d\mathbf{x}) \leq -\int_{\mathbb{S}^{d-1}} f(\mathbf{x})\log f^*(\mathbf{x})\sigma(d\mathbf{x})$. In this case $f^*(\mathbf{x}) = \log c_{1,d}(\kappa, \alpha) + \frac{\kappa}{\alpha}(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x})^{<\alpha>}, \mathbf{x} \in \mathbb{S}^{d-1}$ and hence

$$H(\mathbf{X}) \leq -\int_{\mathbb{S}^{d-1}} f(\mathbf{x})\log f^*(\mathbf{x})\sigma(d\mathbf{x}) = -\log c_{1,d}(\kappa, \alpha)$$
$$-\frac{\kappa}{\alpha}\int_{\mathbb{S}^{d-1}} (\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x})^{<\alpha>}f(\mathbf{x})\sigma(d\mathbf{x}) = -\log c_{1,d}(\kappa, \alpha) - \frac{\kappa}{\alpha}\mathbb{E}\left[(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X})^{<\alpha>}\right]$$
$$= -\log c_{1,d}(\kappa, \alpha) - \frac{\kappa}{\alpha}\mathbb{E}\left[(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{Z})^{<\alpha>}\right] = -\log c_{1,d}(\kappa, \alpha)$$

$$-\frac{\kappa}{\alpha}\int_{\mathbb{S}^{d-1}}(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x})^{<\alpha>}f^*(\mathbf{x})\sigma(d\mathbf{x}) = -\int_{\mathbb{S}^{d-1}} f^*(\mathbf{x})\log f^*(\mathbf{x})\sigma(d\mathbf{x}) = H(\mathbf{Z}).$$

□

The maximum entropy principle for generalized von Mises-Fisher distribution of type II has the following form.

**Theorem 4.3.** *Let a unit random vector* $\mathbf{Z} \in \mathbb{S}^{d-1}$ *have a generalized von Mises-Fisher distribution* $\mathrm{GvMF}_{2,d}(\alpha,\kappa,\boldsymbol{\mu})$. *Then* $\mathbf{Z}$ *has the maximum entropy value over all continuous random variables* $\mathbf{X}$ *on* $\mathbb{S}^{d-1}$ *with* $\mathbb{E}\|\mathbf{X}-\boldsymbol{\mu}\|^{2\alpha} = \mathbb{E}\|\mathbf{Z}-\boldsymbol{\mu}\|^{2\alpha}$.

*Proof.* Let $f$ and $f^*$ be the densities of $\mathbf{X}$ and $\mathbf{Z}$ respectively. The proof is similar to Theorem 4.2. Indeed, $f^*(\mathbf{x}) = \log c_{2,d}(\kappa,\alpha) - \frac{\kappa}{2^\alpha\alpha}\|\mathbf{x}-\boldsymbol{\mu}\|^{2\alpha}, \mathbf{x}\in\mathbb{S}^{d-1}$ and hence

$$H(\mathbf{X}) \le -\int_{\mathbb{S}^{d-1}} f(\mathbf{x})\log f^*(\mathbf{x})\sigma(d\mathbf{x}) = -\log c_{2,d}(\kappa,\alpha) + \frac{\kappa}{2^\alpha\alpha}\mathbb{E}\|\mathbf{X}-\boldsymbol{\mu}\|^{2\alpha}$$
$$= -\log c_{2,d}(\kappa,\alpha) + \frac{\kappa}{2^\alpha\alpha}\mathbb{E}\|\mathbf{Z}-\boldsymbol{\mu}\|^{2\alpha} = -\int_{\mathbb{S}^{d-1}} f^*(\mathbf{x})\log f^*(\mathbf{x})\sigma(d\mathbf{x}) = H(\mathbf{Z}).$$

□

**Theorem 4.4.** *Let a unit random vector* $\mathbf{Z} \in \mathbb{S}^{d-1}$ *have an axial generalized von Mises-Fisher distribution* $\mathrm{GvMF}_{3,d}(\alpha,\kappa,\boldsymbol{\mu})$. *Then* $\mathbf{Z}$ *has the maximum entropy value over all continuous random variables* $\mathbf{X}$ *on* $\mathbb{S}^{d-1}$ *with* $\mathbb{E}|\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X}|^\alpha = \mathbb{E}|\boldsymbol{\mu}^{\mathrm{T}}\mathbf{Z}|^\alpha$.

*Proof.* The proof is similar to Theorems 4.2 and 4.3. □

### *4.2. Entropy estimation*

In this section, we describe the method of entropy estimation for random unit vectors. Actually, we extend the phase-space $\mathbb{S}^{d-1}$ to the arbitrary compact Riemannian manifold. Let $m,d\in\mathbb{N}$, $m\le d$, and $\mathcal{M}$ be a $m$-dimensional $C^1$ manifold embedded in $\mathbb{R}^d$ with the atlas $((U_i,g_i), i\in I_0)$, i.e., for each $y\in\mathcal{M}$ there exists an open subset $U_i$ of $\mathbb{R}^m$ and a continuously differentiable injection $g_i: U_i\to\mathbb{R}^d$, such that $y\in g_i(U)\subset\mathcal{M}$, and $g_i$ is an open map from $U_i$ to $\mathcal{M}$, and the linear map $g_i'(u)$ has full rank for all $u\in U_i$.

For bounded measurable $h:\mathcal{M}\to\mathbb{R}$, the integral $\int_{\mathcal{M}} h(y)\nu(dy)$ is defined by

$$\int_{\mathcal{M}} h(y)\nu(dy) = \sum_{i\in I_0}\int_{U_i} h(g_i(x))\psi_i(g_i(x))\sqrt{\det(J_{g_i}(x))'(J_{g_i}(x))}dx,$$

where $\nu$ is a $\sigma$-finite measure on $\mathcal{M}$, $J_{g_i}$ is the Jacobian of $g_i$ and $\{\psi_i, i\in I_0\}$ is the partition of unity, see [42, pp. 3–4] and [4, Chapter 2] for more detailed setting.

Let $f : \mathcal{M} \to \mathbb{R}_+$ be a probability density of independent random elements $X, X_i, i \in \mathbb{N}$ with values in $\mathcal{M}$, i.e., $\int_{\mathcal{M}} f(x)\nu(dx) = 1$. Let a natural number $k$ be fixed. Denote by $\mathcal{X}_N = \{X_1, \dots, X_N\}$, $N \geq k$ the samples of the first $N$ elements. The entropy of $X$ equals $H(X) = -\int_{\mathcal{M}} \log(f(x))f(x)\nu(dx)$. Let $F$ be a finite subset of $\{X_i, i \geq k\}$ and $\rho_k(x, F)$ be the Euclidean distance between $x$ and its $k$th nearest neighbour in $F \setminus \{x\}$. Let $\gamma \approx 0.5772$ be the Euler–Mascheroni constant.

**Definition 4.1.** *The kth nearest neighbour estimator of the entropy $H(X)$ is defined by*

$$\widehat{H}_{N,k}(\mathcal{X}_N) = \frac{1}{N}\sum_{i=1}^{N} \log\left(\rho_k^m(X_i, \mathcal{X}_N)V_m(N-1)e^{-\psi(k)}\right), \tag{18}$$

*where $\psi(k) = \sum_{j=1}^{k-1} j^{-j} - \gamma$, $V_m = \pi^{m/2}\Gamma^{-1}(1 + m/2)$.*

For example, if $k = 1$, then the nearest neighbour distances estimator is $\widehat{H}_{N,1}(\mathcal{X}_N) = \frac{m}{N}\sum_{i=1}^{N} \log \rho_1(X_i, \mathcal{X}_N) + \log V_m + \gamma + \log(N-1)$. We start the proof of $L^2$-consistency of $\widehat{H}_{N,k}$ by writing down the particular case of Theorem 3.1 from [42] for the functional $\xi(x, \mathcal{X}) := \log\left(e^{-\psi(k)}V_m\rho_k^m(x, \mathcal{X})\right)$.

**Theorem 4.5.** *Let $k \geq 1$, put $q = 1$ or $q = 2$. Suppose there exists $p \geq q$ such that*

$$\sup_{N \geq k} \mathbb{E}\left|\xi\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathcal{X}_N\right)\right|^p < \infty. \tag{19}$$

*Then as $N \to \infty$ we have $L^q$ convergence*

$$\frac{1}{N}\sum_{x \in \mathcal{X}_N} \xi\left(N^{\frac{1}{m}}x, N^{\frac{1}{m}}\mathcal{X}_N\right) \to \int_{\mathcal{M}} \mathbb{E}[\xi\left(\mathbf{0}, \mathcal{P}_{f(x)}\right)]f(x)\nu(dx), \tag{20}$$

*where $\mathcal{P}_\lambda$ denotes a homogeneous Poisson point process of intensity $\lambda > 0$ in $\mathbb{R}^m$ (embedded in $\mathbb{R}^d$).*

For the bounded random variables $X_i, i \geq 1$ and $\rho_k(x, \mathcal{X}_N)$, we generalize Lemma 7.8 from [42], which was proved for the case $k = 1$.

**Lemma 4.1.** *Let $f$ be bounded and have a compact support on $\mathcal{M}$, then for any $\delta \in (0, m)$ it holds $\sup_{N \geq k} \mathbb{E}\left[\rho_k^\delta\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathcal{X}_N\right)\right] < \infty$.*

*Proof.* The proof is very similar to [42, Lemma 7.8]. Recall that $\mathcal{M}$ has the atlas $((U_i, g_i), i \in I_0)$, where $I_0 = \{1, \dots, i_0\}$, and there exist $\delta_i, x_i, i \in I_0$ such that $\mathcal{M} \in \cup_{i \in I_0} B_{\delta_i}(y_i)$.

Denote $A_i = B_{\delta_i} \setminus \cup_{j<i} B_{\delta_j}(y_j)$. Since $\mathrm{supp}(f)$ is bounded then there exist $i_0 \in \mathbb{N}$ and constant $C > 0$ such that

$$\mathbb{E}\left[N^{\frac{\delta}{m}}\rho_k^\delta(X_1, \mathcal{X}_N)\right] = N^{\frac{\delta}{m}-1}\mathbb{E}\left(\sum_{x \in \mathcal{X}_N} \rho_k^\delta(x, \mathcal{X}_N)\right)$$

$$\leq N^{\frac{\delta}{m}-1}\left[\sum_{i=1}^{i_0}\sum_{x\in A_i\cap\mathcal{X}_N}\rho_k^{\delta}(x,A_i\cap\mathcal{X}_N)+C\right]. \tag{21}$$

Now we prove that $\sum_{x\in\mathcal{Y}}\rho_k^{\delta}(x,\mathcal{Y})\leq C_i\left[\mathrm{card}(\mathcal{Y})\right]^{1-\delta/m}$ for all finite $\mathcal{Y}\subset A_i$, where $C_i>0$. Let $\mathcal{Y}\subset A_i$ and $y_j\in\mathcal{Y}$ be a the $j$th nearest neighbour of $x\in\mathcal{Y}$. Taking $z_j\in\mathcal{Y}$ such that $g_i^{-1}(z_j)$ to be $j$-th nearest neighbour of $g_i^{-1}(x)$ in $g_i^{-1}(\mathcal{Y})$, we have from [42, Lemma 4.1] that

$$\begin{aligned}\rho_k(x,\mathcal{Y})&=\max\{\|y_1-x\|,\dots,\|y_k-x\|\}\leq\max\{\|z_1-x\|,\dots,\|z_k-x\|\}\\&\leq C_i\max\left\{\left\|g_i^{-1}(z_1)-g_i^{-1}(x)\right\|,\dots,\left\|g_i^{-1}(z_k)-g_i^{-1}(x)\right\|\right\}\\&=C_i\rho_k(g_i^{-1}(x),g_i^{-1}(\mathcal{Y})).\end{aligned}$$

Thus, from [46, Lemma 3.3] we have for any $\delta\in(0,m)$

$$\begin{aligned}&\sum_{x\in A_i\cap\mathcal{X}_N}\rho_k^{\delta}(x,A_i\cap\mathcal{X}_N)\\&\leq C_i[\mathrm{diam}(g_i^{-1}(A_i\cap\mathcal{X}_N))]^{\delta}[\mathrm{card}(g_i^{-1}(A_i\cap\mathcal{X}_N))]^{1-\frac{\delta}{m}}\leq\tilde{C}_iN^{1-\frac{\delta}{m}}.\end{aligned}$$

Hence, the right-hand side of (21) is bounded above uniformly. □

Now we prove the $L^2$-convergence of the $k$th nearest neighbour estimator $\widehat{H}_{N,k}(\mathcal{X}_N)=N^{-1}\sum_{x\in\mathcal{X}_N}\xi\left(N^{1/m}x,N^{1/m}\mathcal{X}_N\right)$, which is an extension from the case $k=1$ to $k\geq1$ of Theorem 2.4. from [42].

**Theorem 4.6.** *Suppose $f$ is bounded and has compact support. Then for every fixed $k\geq1$*

$$\mathbb{E}\left[\widehat{H}_{N,k}(\mathcal{X}_N)-H(X)\right]^2\to0\quad\text{as } N\to\infty. \tag{22}$$

*Proof.* We apply Theorem 4.5. Repeating the lines of the proof of [10, Theorem 3], we have that $\int_{\mathcal{M}}\mathbb{E}[\xi\left(\mathbf{0},\mathcal{P}_{f(x)}\right)]f(x)\nu(dx)=-\int_{\mathcal{M}}(\log f(x))f(x)\nu(dx)=H(X)$. Second, we check condition (19). Note that for every $\delta\in(0,1)$ and $p>1$ there exists $C>0$ such that $|\log t|^p\leq Ct^{-\delta}\mathbb{1}_{[0,1]}(t)+Ct^{\delta}\mathbb{1}_{[1,\infty)}(t),t>0$. Then

$$\begin{aligned}&\mathbb{E}\left|\xi\left(N^{\frac{1}{m}}X_1,N^{\frac{1}{m}}\mathcal{X}_N\right)\right|^p\leq2^{p-1}\left|\log V_m-\psi(k)\right|^p\\&+2^{p-1}\mathbb{E}\left|\log\rho_k^m\left(N^{\frac{1}{m}}X_1,N^{\frac{1}{m}}\mathcal{X}_N\right)\right|^p\leq2^{p-1}\left|\log V_m-\psi(k)\right|^p\\&+2^{p-1}C\mathbb{E}\rho_k^{-\delta}\left(N^{\frac{1}{m}}X_1,N^{\frac{1}{m}}\mathcal{X}_N\right)\mathbb{1}_{[0,1]}\left(\rho_k^{\delta}\left(N^{\frac{1}{m}}X_1,N^{\frac{1}{m}}\mathcal{X}_N\right)\right) &(23)\\&+2^{p-1}C\mathbb{E}\rho_k^{\delta}\left(N^{\frac{1}{m}}X_1,N^{\frac{1}{m}}\mathcal{X}_N\right)\mathbb{1}_{[1,\infty)}\left(\rho_k^{\delta}\left(N^{\frac{1}{m}}X_1,N^{\frac{1}{m}}\mathcal{X}_N\right)\right). &(24)\end{aligned}$$

Term (23) is finite because

$$\begin{aligned}&\sup_{N\geq k}\mathbb{E}\rho_k^{-\delta}\left(N^{\frac{1}{m}}X_1,N^{\frac{1}{m}}\mathcal{X}_N\right)\mathbb{1}_{[0,1]}\left(\rho_k^{\delta}\left(N^{\frac{1}{m}}X_1,N^{\frac{1}{m}}\mathcal{X}_N\right)\right)\\&\leq\sup_{N\geq k}\mathbb{E}\rho_1^{-\delta}\left(N^{\frac{1}{m}}X_1,N^{\frac{1}{m}}\mathcal{X}_N\right)<\infty,\end{aligned} \tag{25}$$

where (25) is ensured by [42, Lemma 7.5] if $f$ is bounded and $\delta \in (0, m)$. Thus, applying Lemma 4.1 we get that $\sup_{N\geq k} \mathbb{E}\rho_k^{\delta}\left(N^{\frac{1}{m}}X_1, N^{\frac{1}{m}}\mathcal{X}_N\right) < \infty$ if $0 < \delta < m$. Hence, (19) is satisfied. □

The 2-dimensional sphere $\mathbb{S}^2$ is a compact manifold with $d = 3$, $m = 2$ and $\nu = \sigma$. Thus, Theorem 4.3 is valid for all bounded densities on $\mathbb{S}^2$, $k$th nearest neighbour estimator has the form $\widehat{H}_{N,k}(\mathcal{X}_N) = 2N^{-1}\sum_{i=1}^{N}\log\rho_k(X_i, \mathcal{X}_N) - \psi(k) + \log(N-1) + \log\pi$, and $\widehat{H}_{N,k}(\mathcal{X}_N) \to H(X)$ in $L^2(\Omega)$. This yields, that $\widehat{H}_{N,k}(\mathcal{X}_N)$ is a consistent estimator of the Shannon entropy.

## 5. Estimation of parameters

### 5.1. Fisher's maximum likelihood estimation

Let $\mathcal{X}_N = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ be a random sample. From direct calculations we write down the log-likelihood $l(\mathcal{X}_N)$ for random samples from the introduced generalized von Mises-Fisher distributions. Let $\mathcal{X}_{j,N}$ follow $\mathrm{GvMF}_{j,d}(\alpha, \kappa, \boldsymbol{\mu})$ distribution, then $l(\mathcal{X}_{1,N}) = N\log c_{1,d}(\kappa,\alpha) + (\kappa/\alpha)\sum_{i=1}^{N}(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}_i)^{<\alpha>}$, $l(\mathcal{X}_{2,N}) = N\log c_{2,d}(\kappa,\alpha) - 2^{-\alpha}(\kappa/\alpha)\sum_{i=1}^{N}\|\mathbf{x}_i - \boldsymbol{\mu}\|^{2\alpha}$, and $l(\mathcal{X}_{3,N}) = N\log c_{3,d}(\kappa,\alpha) + (\kappa/\alpha)\sum_{i=1}^{N}|\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}_i|^{\alpha}$. In each case, we use numerical methods to find the maximum likelihood estimates $(\hat{\alpha}_L, \hat{\kappa}_L, \hat{\boldsymbol{\mu}}_L)$ of $(\alpha, \kappa, \boldsymbol{\mu})$ which maximize the log-likelihood $l(\mathcal{X}_{j,N})$.

The problem becomes easier when parameter $\alpha$ is known. In such a case, we can derive maximum likelihood estimates taking derivatives of $l(\mathcal{X}_N)$. We have the estimates of $\boldsymbol{\mu}$ as $\kappa$ as

- Let $\mathcal{X}_N \sim \mathrm{GvMF}_{1,d}(\alpha, \kappa, \boldsymbol{\mu})$, then $\hat{\boldsymbol{\mu}}_L = \arg\max_{\boldsymbol{\mu}\in\mathbb{S}^{d-1}}\sum_{i=1}^{N}(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}_i)^{<\alpha>}$,

$$\frac{A_1(\hat{\kappa}_L,\alpha,\alpha) - A_1(-\hat{\kappa}_L,\alpha,\alpha)}{A_1(\hat{\kappa}_L,\alpha,0) + A_1(-\hat{\kappa}_L,\alpha,0))} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\boldsymbol{\mu}}_L^{\mathrm{T}}\mathbf{x}_i)^{<\alpha>}.$$

- Let $\mathcal{X}_N \sim \mathrm{GvMF}_{2,d}(\alpha, \kappa, \boldsymbol{\mu})$, then $\hat{\boldsymbol{\mu}}_L = \arg\min_{\boldsymbol{\mu}\in\mathbb{S}^{d-1}}\sum_{i=1}^{N}\|\mathbf{x}_i - \boldsymbol{\mu}\|^{2\alpha}$, and $2^{\alpha}A_2(\hat{\kappa}_L,\alpha,\alpha)/A_2(\hat{\kappa}_L,\alpha,0) = N^{-1}\sum_{i=1}^{N}\|\mathbf{x}_i - \hat{\boldsymbol{\mu}}_L\|^{2\alpha}$.
- Let $\mathcal{X}_N \sim \mathrm{GvMF}_{3,d}(\alpha, \kappa, \boldsymbol{\mu})$, then $\hat{\boldsymbol{\mu}}_L = \arg\max_{\boldsymbol{\mu}\in\mathbb{S}^{d-1}}\sum_{i=1}^{N}|\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}_i|^{\alpha}$, and $A_1(\hat{\kappa}_L,\alpha,\alpha)/A_1(\hat{\kappa}_L,\alpha,0) = N^{-1}\sum_{i=1}^{N}|\hat{\boldsymbol{\mu}}_L^{\mathrm{T}}\mathbf{x}_i|^{\alpha}$.

### 5.2. Method of moments

In this section we consider parameter estimation of generalized von Mises-Fisher distributions based on the moments estimation.

In the case of non-axial random vector $\mathbf{X} \in \mathbb{S}^{d-1}$ we assume that $\|\mathbb{E}\mathbf{X}\| \neq 0$. We know from Definition 2.2 that $\hat{\boldsymbol{\mu}} := \bar{\mathcal{X}}_N/\|\bar{\mathcal{X}}_N\|$, where $\bar{\mathcal{X}}_N = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i$, is the natural estimator for mean direction parameter $\boldsymbol{\mu}$. In order to find estimates for parameters $\alpha$ and $\kappa$ we need at least two more moment statistics. A standard

approach involves the use of the resultant length. For the second relations we choose $\mathbb{E}(\operatorname{sgn}\{\mathbf{X}_1{}^{\mathrm{T}}\mathbb{E}\mathbf{X}_1\})$ and $\mathbb{E}(\|\mathbf{X}_2-\boldsymbol{\mu}\|^4)$ for vectors $\mathbf{X}_j\sim \mathrm{GvMF}_{j,d}(\alpha,\kappa,\boldsymbol{\mu})$ $(j=1,2)$. By Remark 3.4 one can get the estimators $\hat\kappa,\hat\alpha$ as a solution of the following equations

$$\frac{A_1(\hat\kappa,\hat\alpha,1)-A_1(-\hat\kappa,\hat\alpha,1)}{A_1(\hat\kappa,\hat\alpha,0)+A_1(-\hat\kappa,\hat\alpha,0)}=\|\bar{\mathcal{X}}_{1,N}\|,$$
$$\frac{A_1(\hat\kappa,\hat\alpha,0)-A_1(-\hat\kappa,\hat\alpha,0)}{A_1(\hat\kappa,\hat\alpha,0)+A_1(-\hat\kappa,\hat\alpha,0)}=\frac{\sum_{i=1}^N\operatorname{sgn}(\mathbf{x}_i^{\mathrm{T}}\bar{\mathcal{X}}_{1,N})}{N\|\bar{\mathcal{X}}_{1,N}\|};$$
$$\frac{A_2(\hat\kappa,\hat\alpha,1)}{A_2(\hat\kappa,\hat\alpha,0)}=1-\|\bar{\mathcal{X}}_{2,N}\|,\quad \frac{A_2(\hat\kappa,\hat\alpha,2)}{A_2(\hat\kappa,\hat\alpha,0)}=\frac{1}{4N}\sum_{i=1}^N\left\|x_i-\frac{\bar{\mathcal{X}}_{2,N}}{\|\bar{\mathcal{X}}_{2,N}\|}\right\|^4$$

for the samples $\mathcal{X}_{1,N}\sim \mathrm{GvMF}_{1,d}(\alpha,\kappa,\boldsymbol{\mu})$ and $\mathcal{X}_{2,N}\sim \mathrm{GvMF}_{2,d}(\alpha,\kappa,\boldsymbol{\mu})$, respectively.

*Remark* 5.1. If the parameter $\alpha$ is known, we can reduce the problem of the moment estimation to the solution of one equation. Namely,

$$\frac{A_1(\hat\kappa,\alpha,1)-A_1(-\hat\kappa,\alpha,1)}{A_1(\hat\kappa,\alpha,0)+A_1(-\hat\kappa,\alpha,0)}=\|\bar{\mathcal{X}}_{1,N}\|,\ \text{and}\ \frac{A_2(\hat\kappa,\alpha,1)}{A_2(\hat\kappa,\alpha,0)}=1-\|\bar{\mathcal{X}}_{2,N}\|.$$

In the case of a symmetrically distributed random vector $\mathbf{X}\in\mathbb{S}^{d-1}$, $\mathbb{E}\mathbf{X}=0$ and the value $\mathbb{E}\mathbf{X}/\|\mathbb{E}\mathbf{X}\|$ is not defined. Recall the tangent-normal decomposition (2) of a random vector $\mathbf{X}\in\mathbb{S}^{d-1}$, that is $\mathbf{X}=\boldsymbol{\mu}\xi+\sqrt{1-\xi^2}\mathbf{Y}$, where $\boldsymbol{\mu}\in\mathbb{S}^{d-1}$ is a mean direction parameter, $\xi$ is a random variable on $[-1,1]$ independent of a uniformly distributed random vector $\mathbf{Y}\in\mathbb{S}^{d-1}_{\boldsymbol{\mu}}$ such that $\boldsymbol{\mu}\perp\mathbf{Y}$.

To find relations which determine the parameter $\boldsymbol{\mu}$ and distribution $\xi$ we consider an orientation tensor given by

$$T(\mathbf{X})=\mathbf{X}\mathbf{X}^{\mathrm{T}}=\xi^2\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}+\xi\sqrt{1-\xi^2}\,(\boldsymbol{\mu}\mathbf{Y}^{\mathrm{T}}+\mathbf{Y}\boldsymbol{\mu}^{\mathrm{T}})+(1-\xi^2)\mathbf{Y}\mathbf{Y}^{\mathrm{T}}. \tag{26}$$

Therefore, the mean orientation tensor is $\mathbb{E}T(\mathbf{X})=\mathbb{E}[\mathbf{X}\mathbf{X}^{\mathrm{T}}]=\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}\mathbb{E}\xi^2+(1-\mathbb{E}\xi^2)\mathbb{E}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}$.

**Theorem 5.1.** *Let a random vector* $\mathbf{X}$ *has a representation as above, i.e.,* $\mathbf{X}=\boldsymbol{\mu}\xi+\sqrt{1-\xi^2}\mathbf{Y}$. *Then*

$$\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}=\sqrt{\frac{d-1}{d\mathbb{E}[\mathbf{X}^{\mathrm{T}}(\mathbb{E}T(\mathbf{X}))\mathbf{X}]-1}}\left(\mathbb{E}T(\mathbf{X})-\frac{1}{d}I_d\right)+\frac{1}{d}I_d, \tag{27}$$

$$\mathbb{E}\xi^2=\frac{1}{d}+\sqrt{\frac{d-1}{d}}\sqrt{\mathbb{E}[\mathbf{X}^{\mathrm{T}}(\mathbb{E}T(\mathbf{X}))\mathbf{X}]-\frac{1}{d}}, \tag{28}$$

$$\mathbb{E}\xi^4=\mathbb{E}\left[\mathbf{X}^{\mathrm{T}}\mathbb{E}T(\mathbf{X})\mathbf{X}-\frac{1-\mathbb{E}\xi^2}{d-1}\right]^2\frac{d-1}{d\mathbb{E}[\mathbf{X}^{\mathrm{T}}(\mathbb{E}T(\mathbf{X}))\mathbf{X}]-1}. \tag{29}$$

*where* $I_d$ *is* $d\times d$ *identity matrix.*

*Proof.* Let $U_{\boldsymbol{\mu}} \in SO(d)$, such that $\boldsymbol{\mu} = U_{\boldsymbol{\mu}}\mathbf{e}_x$, where $\mathbf{e}_x = (1, 0, \dots, 0)^{\mathrm{T}}$. Denote by $\tilde{\mathbf{Y}} = U_{\boldsymbol{\mu}}^{-1}\mathbf{Y}$. The vector $\tilde{\mathbf{Y}}$ is uniformly distributed on $\mathbb{S}^{d-1}_{\mathbf{e}_x}$ with the first coordinate equal 0. Then $\mathbf{X} = U_{\boldsymbol{\mu}}\left(\xi\mathbf{e}_x + \sqrt{1-\xi^2}\tilde{\mathbf{Y}}\right)$ and $\mathbb{E}\mathbf{X}\mathbf{X}^{\mathrm{T}} = U_{\boldsymbol{\mu}}\left(\mathbf{e}_x\mathbf{e}_x^{\mathrm{T}}\mathbb{E}\xi^2 + (1-\mathbb{E}\xi^2)\mathbb{E}\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^{\mathrm{T}}\right)U_{\boldsymbol{\mu}}^{-1}$. It follows from the symmetry of $\tilde{\mathbf{Y}}$ that $\mathbb{E}\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^{\mathrm{T}} = \frac{1}{d-1}\begin{pmatrix} 0 & 0 \\ 0 & I_{d-1}\end{pmatrix}$. Therefore, $\mathbb{E}\mathbf{X}\mathbf{X}^{\mathrm{T}}$ equals

$$U_{\boldsymbol{\mu}}\left(\mathbf{e}_x\mathbf{e}_x^{\mathrm{T}}\mathbb{E}\xi^2 + \frac{1-\mathbb{E}\xi^2}{d-1}(I_d - \mathbf{e}_x\mathbf{e}_x^{\mathrm{T}})\right)U_{\boldsymbol{\mu}}^{\mathrm{T}} = \mathbb{E}\xi^2\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}} + \frac{1-\mathbb{E}\xi^2}{d-1}(I_d - \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}). \tag{30}$$

Thus, $\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}} = (d-1)(d\mathbb{E}\xi^2 - 1)^{-1}\left(\mathbb{E}\mathbf{X}\mathbf{X}^{\mathrm{T}} - (1-\mathbb{E}\xi^2)/d - 1\right)$. From (30), we have that $\mathbf{X}^{\mathrm{T}}\mathbb{E}T(\mathbf{X})\mathbf{X}$ equals

$$\mathbf{X}^{\mathrm{T}}\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X}\mathbb{E}\xi^2 + \frac{1-\mathbb{E}\xi^2}{d-1}(1 - \mathbf{X}^{\mathrm{T}}\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X}) = \xi^2\mathbb{E}\xi^2 + \frac{1-\mathbb{E}\xi^2}{d-1}(1-\xi^2),$$

and $\mathbb{E}[\mathbf{X}^{\mathrm{T}}\mathbb{E}T(\mathbf{X})\mathbf{X}] = (\mathbb{E}\xi^2)^2 + (1-\mathbb{E}\xi^2)^2/(d-1)$. This yields

$$\mathbb{E}\xi^2 = \frac{1}{d} + \sqrt{\frac{d-1}{d}}\sqrt{\mathbb{E}[\mathbf{X}^{\mathrm{T}}(\mathbb{E}\mathbf{X}\mathbf{X}^{\mathrm{T}})\mathbf{X}] - \frac{1}{d}}$$

and

$$\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}} = \sqrt{\frac{d-1}{d\mathbb{E}[\mathbf{X}^{\mathrm{T}}(\mathbb{E}\mathbf{X}\mathbf{X}^{\mathrm{T}})\mathbf{X}] - 1}}\left(\mathbb{E}\mathbf{X}\mathbf{X}^{\mathrm{T}} - \frac{1}{d}I_d\right) + \frac{1}{d}I_d.$$

Finally, $\mathbb{E}\left[\mathbf{X}^{\mathrm{T}}\mathbb{E}T(\mathbf{X})\mathbf{X} - (1-\mathbb{E}\xi^2)/(d-1)\right]^2 = \left(d\mathbb{E}\xi^2 - 1\right)^2(d-1)^{-2}\mathbb{E}\xi^4$. □

Note, that for an axial vector $\mathbf{X}$, the random variable $\xi$ has a symmetric distribution on $[-1, 1]$, therefore $\mathbb{E}\xi = 0$. So, if $\xi$ has two-dimensional parametric distribution, one get from Theorem 5.1 the parameter estimates for $\xi$.

We apply Theorem 5.1 for the random sample $\mathcal{X}_N = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ from $\mathrm{GvMF}_{3,d}(\alpha, \kappa, \boldsymbol{\mu})$. Then $\mathbb{E}\xi^2 = \mathbb{E}(\boldsymbol{\mu}'\mathbf{x}_1)^2$ and $\mathbb{E}\xi^4 = \mathbb{E}(\boldsymbol{\mu}'\mathbf{x}_1)^4$ are given in Proposition 3.3.

**Corollary 5.1.1.** *Let $\mathcal{X}_N \sim \mathrm{GvMF}_{3,d}(\alpha, \kappa, \boldsymbol{\mu})$. Denote by $\bar{T} = \frac{1}{N}\sum_{i=1}^N \mathbf{x}_i\mathbf{x}_i{}^{\mathrm{T}}$ and $\bar{V} = \frac{1}{N}\sum_{i=1}^N \mathbf{x}_i^{\mathrm{T}}\bar{T}\mathbf{x}_i$. Then the solutions $\hat{\boldsymbol{\mu}}, \hat{\kappa}, \hat{\alpha}$ of the following equations are the estimators of the parameters $\boldsymbol{\mu}, \kappa, \alpha$ by the method of moments.*

$$\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^{\mathrm{T}} = \sqrt{\frac{d-1}{d\bar{V}-1}}\left(\bar{T} - \frac{1}{d}I_d\right) + \frac{1}{d}I_d, \quad \frac{A_1(\hat{\kappa}, \hat{\alpha}, 2)}{A_1(\hat{\kappa}, \hat{\alpha}, 0)} = \frac{1}{d} + \sqrt{\frac{d-1}{d}}\sqrt{\bar{V} - \frac{1}{d}},$$
$$\frac{A_1(\hat{\kappa}, \hat{\alpha}, 4)}{A_1(\hat{\kappa}, \hat{\alpha}, 0)} = \frac{d-1}{(d\bar{V}-1)N}\sum_{i=1}^N\left(\mathbf{x}_i{}^{\mathrm{T}}\bar{T}\mathbf{x}_i - \frac{1}{d} - \frac{\sqrt{d\bar{V}-1}}{d\sqrt{d-1}}\right)^2.$$

## 6. Goodness of fit test based on the maximum entropy principle

In this section we provide the statistical test for verification that a random sample follows a generalized von Mises-Fisher distribution. The methodology for all three introduced distributions is very similar. For simplicity, we provide a detailed explanation for the Type-II distribution.

Denote by $\mathrm{GvMF}_{2,d}$ the class of generalized von Mises-Fisher distributions $\mathrm{GvMF}_{2,d}(\alpha,\kappa,\boldsymbol{\mu})$, $\alpha>0$, $\kappa>0$ and $\boldsymbol{\mu}\in\mathbb{S}^{d-1}$. Let $\mathcal{X}_N=\{\mathbf{x}_1,\dots,\mathbf{x}_N\}$ be a random sample of vectors on a sphere $\mathbb{S}^{d-1}$ and $\mathbf{x}_j\overset{d}{=}\mathbf{X}, j=1,\dots,N$ with unknown distribution.

Let $\mathbf{Z}\sim\mathrm{GvMF}_{2,d}(\alpha,\kappa,\boldsymbol{\mu})$. From (10), (12), and Theorem 4.3 we know that $H(\mathbf{Z})\geq H(\mathbf{X})$ for all continuous random vectors $\mathbf{X}\in\mathbb{S}^{d-1}$ with $\mathbb{E}\|\mathbf{X}-\boldsymbol{\mu}\|^{2\alpha}=\mathbb{E}\|\mathbf{Z}-\boldsymbol{\mu}\|^{2\alpha}=2^{\alpha}\frac{A_2(\kappa,\alpha,\alpha)}{A_2(\kappa,\alpha,0)}$. Using Theorem 4.1, we get that

$$\inf_{\substack{\alpha,\kappa>0,\\ \boldsymbol{\mu}\in\mathbb{S}^{d-1}}}\left\{-\log c_{2,d}(\kappa,\alpha)+\frac{\kappa}{\alpha}\frac{A_2(\kappa,\alpha,\alpha)}{A_2(\kappa,\alpha,0)}\,\middle|\,\mathbb{E}\|\mathbf{X}-\boldsymbol{\mu}\|^{2\alpha}=\frac{2^{\alpha}A_2(\kappa,\alpha,\alpha)}{A_2(\kappa,\alpha,0)}\right\} \tag{31}$$

does not exceed $H(\mathbf{X})$. Moreover, equality in (31) appears if and only if $\mathbf{X}$ belongs to some distribution from the family $\mathrm{GvMF}_{2,d}$. We substitute now the unobservable value $\mathbb{E}\|\mathbf{X}-\boldsymbol{\mu}\|^{2\alpha}$ by its statistical counterpart $\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_i-\boldsymbol{\mu}\|^{2\alpha}$ and define the statistic $S_2(\mathcal{X}_N)$ by

$$\inf_{\substack{\alpha,\kappa>0,\\ \boldsymbol{\mu}\in\mathbb{S}^{d-1}}}\left\{-\log c_{2,d}(\kappa,\alpha)+\frac{\kappa}{\alpha}\frac{A_2(\kappa,\alpha,\alpha)}{A_2(\kappa,\alpha,0)}\,\middle|\,\frac{\sum_{i=1}^{N}\|\mathbf{x}_i-\boldsymbol{\mu}\|^{2\alpha}}{N2^{\alpha}}=\frac{A_2(\kappa,\alpha,\alpha)}{A_2(\kappa,\alpha,0)}\right\}.$$

Under the condition $N^{-1}2^{-\alpha}\sum_{i=1}^{N}\|\mathbf{x}_i-\boldsymbol{\mu}\|^{2\alpha}=A_2(\kappa,\alpha,\alpha)/A_2(\kappa,\alpha,0)$, we have

$$\log c_{2,d}(\kappa,\alpha)-\frac{\kappa}{\alpha}\frac{A_2(\kappa,\alpha,\alpha)}{A_2(\kappa,\alpha,0)}=\log c_{2,d}(\kappa,\alpha)-\frac{\kappa}{\alpha}\frac{\sum_{i=1}^{N}\|\mathbf{x}_i-\boldsymbol{\mu}\|^{2\alpha}}{N2^{\alpha}}=\frac{l_2(\mathcal{X}_N)}{N}.$$

Thus,

$$S_2(\mathcal{X}_N)=-\frac{1}{N}\sup_{\alpha,\kappa>0,\boldsymbol{\mu}\in\mathbb{S}^{d-1}}\left\{l_2(\mathcal{X}_N)\,\middle|\,\frac{\sum_{i=1}^{N}\|\mathbf{x}_i-\boldsymbol{\mu}\|^{2\alpha}}{N2^{\alpha}}=\frac{A_2(\kappa,\alpha,\alpha)}{A_2(\kappa,\alpha,0)}\right\}.$$

Let us consider unconditional maximization of log-likelihood $l_2(\mathcal{X}_N)$. Partial derivative with respect to $\kappa$ equals

$$\begin{aligned}\frac{\partial l_2(\mathcal{X}_N)}{\partial\kappa}&=\frac{\partial}{\partial\kappa}\left(\log\frac{\Gamma\left(\frac{d-1}{2}\right)}{2\pi^{\frac{d-1}{2}}}-\log A_2(\kappa,\alpha,0)-\frac{\kappa}{\alpha 2^{\alpha}}\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_i-\boldsymbol{\mu}\|^{2\alpha}\right)\\&=\frac{1}{\alpha}\frac{A_2(\kappa,\alpha,\alpha)}{A_2(\kappa,\alpha,0)}-\frac{1}{\alpha 2^{\alpha}N}\sum_{i=1}^{N}\|\mathbf{x}_i-\boldsymbol{\mu}\|^{2\alpha},\end{aligned}$$

where we used $\frac{\partial}{\partial\kappa}A_2(\kappa,\alpha,0) = -\frac{1}{\alpha}A_2(\kappa,\alpha,\alpha)$. Thus, the supremum in $S_2(\mathcal{X}_N)$ with respect to $\kappa$ coincides with the unconditional supremum of $l_2(\mathcal{X}_N)$ and $S_2(\mathcal{X}_N) = -N^{-1}\sup\{l_2(\mathcal{X}_N)|\alpha,\kappa>0,\boldsymbol{\mu}\in\mathbb{S}^{d-1}\}$.

Let $\Theta_0$ be a compact subset of $\mathbb{R}^2_+$ large enough to contain all values of parameters $(\alpha,\kappa)$ appearing in practice. Consider the following hypotheses

- $H_{2,0}: \mathcal{X}_N \sim \mathrm{GvMF}_{2,d}$, for some $(\alpha,\kappa)\in\Theta_0$, and $\boldsymbol{\mu}\in\mathbb{S}^{d-1}$,
- $H_{2,1}: \mathcal{X}_N \not\sim \mathrm{GvMF}_{2,d}$ for all $(\alpha,\kappa)\in\Theta_0$, and $\boldsymbol{\mu}\in\mathbb{S}^{d-1}$.

Since $\Theta_0$ is compact, maximum likelihood estimators $\hat{\alpha}_L$, $\hat{\kappa}_L$ are consistent. We proved in Theorem (4.6) that the $k$th nearest neighbour estimator $\widehat{H}_{N,k}$ of $H(\mathbf{X})$ is $L^2$-consistent for any $k\in\mathbb{N}$. Thus, we test $H_{2,0}$ vs. $H_{2,1}$ with the statistic

$$\hat{T}^L_{2,k}(\mathcal{X}_N) := -\log c_{2,d}(\hat{\kappa}_L,\hat{\alpha}_L) + \frac{\hat{\kappa}_L}{\hat{\alpha}_L}\frac{A_2(\hat{\kappa}_L,\hat{\alpha}_L,\hat{\alpha}_L)}{A_2(\hat{\kappa}_L,\hat{\alpha}_L,0)} - \widehat{H}_{N,k} \tag{32}$$

which tends in probability to 0, as $N\to\infty$. We reject $H_{2,0}$ with level of significance $\beta$ if $|\hat{T}^L_{2,k}(\mathcal{X}_N)|\geq x_\beta$, where $x_\beta$ is a critical value determined by $\mathbb{P}_{H_0}(|\hat{T}^L_{2,k}(\mathcal{X}_N)|\geq x_\beta)\leq\beta$.

*Remark* 6.1. It is easy to see that maximum likelihood estimates of $\alpha,\kappa$, estimator $\hat{H}_{N,k}$ and the statistics $\hat{T}^L_{2,k}$ are rotational invariant. Furthermore, by Slutsky's theorem, we can replace the maximum likelihood estimates of $\alpha,\kappa$ in (32) by any consistent estimates $\hat{\alpha},\hat{\kappa}$. Indeed, if $\mathbf{x}_1\sim\mathrm{GvMF}_{2,d}(\alpha,\kappa,\boldsymbol{\mu})$ under hypothesis $H_{2,}$, then $2^{\hat{\alpha}}A_2(\hat{\kappa},\hat{\alpha},\hat{\alpha})/A_2(\hat{\kappa},\hat{\alpha},0)\to 2^{\alpha}A_2(\kappa,\alpha,\alpha)/A_2(\kappa,\alpha,0) = \mathbb{E}\|\mathbf{x}_1-\boldsymbol{\mu}\|^{\alpha}$ and $\hat{T}_{2,k}(\mathcal{X}_N) := -\log c_{2,d}(\hat{\kappa},\hat{\alpha}) + (\hat{\kappa}/\hat{\alpha})A_2(\hat{\kappa},\hat{\alpha},\hat{\alpha})/A_2(\hat{\kappa},\hat{\alpha},0) - \widehat{H}_{N,k}\to H(\mathbf{x}_1) - H(\mathbf{x}_1) = 0$ in probability as $N\to\infty$.

The critical values $x_\beta$ can be found by Monte Carlo simulations of test statistics $\hat{T}^L_{2,N}$ or $\hat{T}_{2,N}$.

The goodness of fit test for the axial generalized von-Mises distribution and the distribution of the I-type are constructed similarly to II-type distributions. Let $\Theta_0$ be a compact subset of $\mathbb{R}^2_+$ large enough to contain all values of parameters $(\alpha,\kappa)$ appearing in practice. Let $j=1,3$ and consider the following hypotheses

- $H_{j,0}: \mathcal{X}_N \sim \mathrm{GvMF}_{j,d}$, for some $(\alpha,\kappa)\in\Theta_0$ and $\boldsymbol{\mu}\in\mathbb{S}^{d-1}$,
- $H_{j,1}: \mathcal{X}_N \not\sim \mathrm{GvMF}_{j,d}$ for all $(\alpha,\kappa)\in\Theta_0$ and $\boldsymbol{\mu}\in\mathbb{S}^{d-1}$.

For testing $H_{1,0}$ vs. $H_{1,1}$ and $H_{3,0}$ vs. $H_{3,1}$ we use the statistics $\hat{T}_{1,N}$ and $\hat{T}_{3,N}$, respectively, which are given by

$$\hat{T}_{1,k}(\mathcal{X}_N) := -\log c_{1,d}(\hat{\kappa},\hat{\alpha}) - \frac{\hat{\kappa}}{\hat{\alpha}}\frac{A_1(\hat{\kappa},\hat{\alpha},\hat{\alpha}) - A_1(-\hat{\kappa},\hat{\alpha},\hat{\alpha})}{A_1(\hat{\kappa},\hat{\alpha},0) + A_1(-\hat{\kappa},\hat{\alpha},0)} - \widehat{H}_{N,k}, \tag{33}$$

$$\hat{T}_{3,k}(\mathcal{X}_N) := -\log c_{3,d}(\hat{\kappa},\hat{\alpha}) - \frac{\hat{\kappa}}{\hat{\alpha}}\frac{A_1(\hat{\kappa},\hat{\alpha},\hat{\alpha})}{A_1(\hat{\kappa},\hat{\alpha},0)} - \widehat{H}_{N,k}, \tag{34}$$

where $\hat{\alpha},\hat{\kappa}$ are some consistent estimates of $\alpha,\kappa$.

*Remark* 6.2. Note, that our goodness of fit tests do not detect some particular generalized von Mises-Fisher distribution but tell whether a sample belongs to the parametric family $\mathrm{GvMF}_{j,d}, j = 1, 2, 3$.

## 7. Numerical experiments

In this section, we provide the simulation method for $\mathrm{GvMF}_{j,d}$ distributions and study the behaviour of the test statistic $\hat{T}_{j,k}$ on simulated samples, $j = 1, 2, 3$.

### *7.1. Simulation*

Let $\mathbf{X}_j \sim \mathrm{GvMF}_{j,d}(\alpha, \kappa, \boldsymbol{\mu}), j = 1, 2, 3$. Due to the tangent-normal decomposition $\mathbf{X}_j = (\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X}_j)\boldsymbol{\mu} + \sqrt{1-(\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X}_j)^2}\mathbf{Y_j}$, where $\mathbf{Y_j}$ is uniformly distributed on $\mathbb{S}^{d-2}_{\boldsymbol{\mu}}$, $j = 1, 2, 3$. So, in order to simulate $\mathbf{X}_j$ we can easily simulate random vectors $\mathbf{Y_j}$ and independent random variables $\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X}_j$. The probability densities $f_j$ of $\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X}_j$ are given in [30, (2.22)] or can be found by applying (1):

$$f_1(y) = \frac{2\pi^{(d-1)/2}c_{1,d}(\kappa,\alpha)}{\Gamma\left((d-1)/2\right)} \exp\left(\frac{\kappa}{\alpha}y^{<\alpha>}\right)(1-y^2)^{\frac{d-3}{2}}, y \in [-1,1], \tag{35}$$

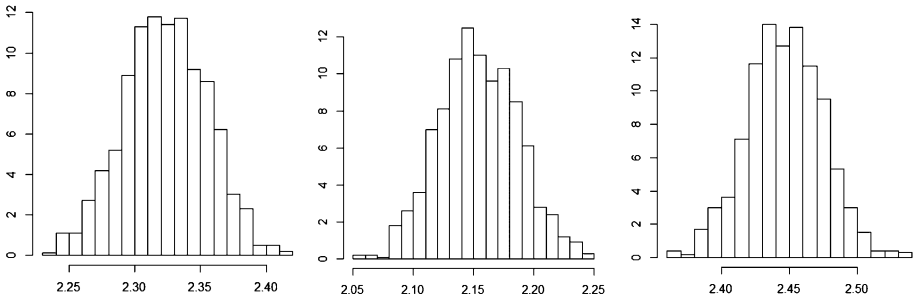
$$f_2(y) = \frac{2\pi^{(d-1)/2}c_{2,d}(\kappa,\alpha)}{\Gamma\left((d-1)/2\right)} \exp\left(-\frac{\kappa}{\alpha}(1-y)^{\alpha}\right)(1-y^2)^{\frac{d-3}{2}}, y \in [-1,1], \tag{36}$$

$$f_3(y) = \frac{2\pi^{(d-1)/2}c_{3,d}(\kappa,\alpha)}{\Gamma\left((d-1)/2\right)} \exp\left(\frac{\kappa}{\alpha}|y|^{\alpha}\right)(1-y^2)^{\frac{d-3}{2}}, y \in [-1,1]. \tag{37}$$

Applying the described simulation procedure, we generate several samples of generalized von Mises–Fisher distributions with 1000 sample points on a 2-dimensional sphere. In the Appendix, we present the locations of the sample points on $\mathbb{S}^2$ in Figures 16, 18, 20, 22, 24, 26. We want to emphasize, that larger values of the parameter $\kappa$ correspond to more concentrated samples along the direction $\boldsymbol{\mu}$. The histograms for samples of random variables $\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X}_j$ $(j = 1, 2, 3)$ and the plots of the probability densities $f_1, f_2, f_3$ are in Figures 15, 17, 19, 21, 23, 25.

We provide a short computational study of the parameter estimation methods from sections 5.1 and 5.2 in the Appendix.

The method of moments is preferable for Types I and II, if the computing power plays a decisive role. For axial data this method has no such advantages because we need to operate with an orientation tensor. The experiments show that the speed of convergence $\hat{\alpha} \to \alpha$ and $\hat{\kappa} \to \kappa$ depend on the values of $\alpha$, and $\kappa$ and $\hat{\kappa} \to \kappa$ faster than $\hat{\alpha} \to \alpha$. We conclude also that the errors of maximum likelihood estimators are generally less than moment estimators. Comparing the errors by the distribution type, we observe that samples of Type II very often carry the smallest error.



(a) Type I (b) Type II (c) Axial type

FIG 1. *Histograms of* $\hat{H}_{N,3}(\mathcal{X}_N)$, $\mathcal{X}_N \sim \mathrm{GvMF}_{j,3}(\alpha,\kappa)$ *with* $\alpha = 1.5$ *and* $\kappa = 2$.

### *7.2. Entropy estimation*

In this section, we apply the $k$th nearest neighbour estimator (18) to the simulated set of samples. We compute estimates $\widehat{H}_{N,k}(\mathcal{X}_N)$ and their sample variances $\mathrm{sVar}(\widehat{H}_{N,k}(\alpha,\kappa))$ for $k = 1,2,3,4,5$, $\kappa \in \{0.1, 0.5, 1, 1.5, 2, 2.5, 3, 4, 5, 6, 7\}$, and $\alpha \in \{0.5, 1, 1.5, 2, 2.5, 3\}$. In Figure 1, we illustrate the distribution of $\widehat{H}_{N,k}(\mathcal{X}_N)$ with $k = 3$ by histograms for samples simulated from distributions $\mathrm{GvMF}_{j,3}(1.5, 2), j = 1, 2, 3$.

In order to choose the right value of $k$ we compare the sample variances $\mathrm{sVar}(\widehat{H}_{N,k}(\alpha,\kappa))$ for $k = 1,2,3,4,5$. The minimum and maximum values of $\mathrm{sVar}(\widehat{H}_{N,k}(\alpha,\kappa))$ with respect to $\alpha$ and $\kappa$ are presented in Table 1. Our results confirm the conclusion in [6], that is, the asymptotic variance of $\widehat{H}_{N,k}$ decreases rapidly up to $k = 3$. One can observe that the $k$th nearest neighbour estimates depend on values $\alpha$ and $\kappa$. Although, the sample variances are quite small for all examined values of $\alpha$ and $\kappa$ and sample size $N = 1000$.

TABLE 1
*Sample variance* $sVar(\widehat{H}_{N,k}(\alpha,\kappa))$ *for distribution of type I*

| Distribution | | $k=1$ | $k=2$ | $k=3$ | $k=4$ | $k=5$ |
|---|---|---|---|---|---|---|
| $\mathrm{GvMF}_{1,3}(\alpha,\kappa)$ | $\min_{\alpha,\kappa}(\mathrm{sVar})$ | 0.00214 | 0.00092 | 0.00058 | 0.00042 | 0.00034 |
| | $\max_{\alpha,\kappa}(\mathrm{sVar})$ | 0.00388 | 0.00239 | 0.00208 | 0.00185 | 0.00152 |
| $\mathrm{GvMF}_{2,3}(\alpha,\kappa)$ | $\min_{\alpha,\kappa}(\mathrm{sVar})$ | 0.00212 | 0.00093 | 0.00059 | 0.00043 | 0.00034 |
| | $\max_{\alpha,\kappa}(\mathrm{sVar})$ | 0.00404 | 0.00304 | 0.00275 | 0.00152 | 0.00257 |
| $\mathrm{GvMF}_{3,3}(\alpha,\kappa)$ | $\min_{\alpha,\kappa}(\mathrm{sVar})$ | 0.00212 | 0.00093 | 0.00059 | 0.00043 | 0.00034 |
| | $\max_{\alpha,\kappa}(\mathrm{sVar})$ | 0.00334 | 0.00207 | 0.00170 | 0.00152 | 0.00144 |

Thus, we choose $k = 3$ for computations in the next sections.

### *7.3. Test statistic*

In this section, we present our study of the goodness of fit tests from Section 6 and their test statistics $\hat{T}_{j,k}(\mathcal{X}_N), j = 1, 2, 3$ from (32), (33) and (34) with $k = 3$

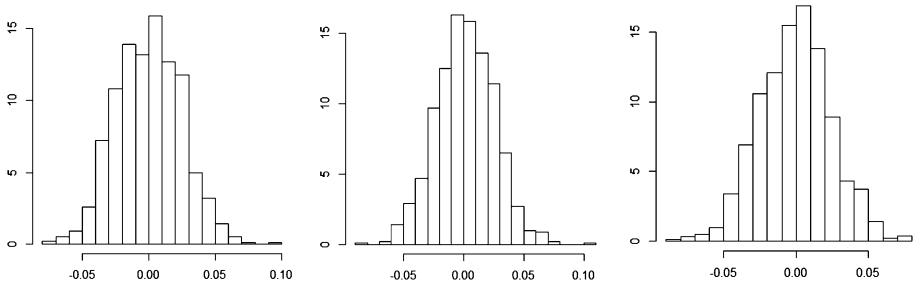
TABLE 2
*Sample variances* $sVar(\hat{T}_{j,3}(\alpha,\kappa))$

| | The method of moments | | | Maximum likelihood method | | |
|---|---|---|---|---|---|---|
| Distribution | $\mathrm{GvMF}_{1,3}$ | $\mathrm{GvMF}_{2,3}$ | $\mathrm{GvMF}_{3,3}$ | $\mathrm{GvMF}_{1,3}$ | $\mathrm{GvMF}_{2,3}$ | $\mathrm{GvMF}_{3,3}$ |
| $\min_{\alpha,\kappa}(\mathrm{sVar})$ | 0.000563 | 0.000538 | 0.000573 | 0.000558 | 0.000535 | 0.000544 |
| $\max_{\alpha,\kappa}(\mathrm{sVar})$ | 0.001014 | 0.001558 | 0.000734 | 0.000665 | 0.000688 | 0.000667 |

and $N = 1000$. We compute $\hat{T}^L_{j,k}(\mathcal{X}_N)$ and $\hat{T}^M_{j,k}(\mathcal{X}_N)$ separately with maximum likelihood estimates and estimates be the method of moments of parameters $\alpha, \kappa$, respectively.

For comparison of different types of estimates, we look on the sample variances $\mathrm{sVar}(\hat{T}^M_{j,3}(\alpha,\kappa))$ and $\mathrm{sVar}(\hat{T}^L_{j,3}(\alpha,\kappa))$ of $\hat{T}^M_{j,3}(\mathcal{X}_N)$ and $\hat{T}^L_{j,3}(\mathcal{X}_N)$, respectively, for all combinations of parameters $\kappa \in \{0.1, 0.5, 1, 1.5, 2, 2.5, 3, 4, 5, 6, 7\}$ and $\alpha \in \{0.5, 1, 1.5, 2, 2.5, 3\}$. Numbers in the Table 2, where the minimum and maximum values of $\mathrm{sVar}(\hat{T}^M_{j,3}(\alpha,\kappa))$ and $\mathrm{sVar}(\hat{T}^L_{j,3}(\alpha,\kappa))$ are presented, demonstrate significant benefits of the maximum likelihood method over the method of moments for distributions of I and II types. For axial data, one can also prefer $\hat{T}^L_{j,k}$. Additionally, one can observe from Table 2 and tables with errors of estimates $\hat{\kappa}^L, \hat{\kappa}^M, \hat{\alpha}^L, \hat{\alpha}^M$ that the statistics $\hat{T}^M_{j,k}$ and $\hat{T}^L_{j,k}$ are much more accurate than estimators of parameters, and they have small variances even for small $\alpha, \kappa$ in contrast to $\hat{\alpha}, \hat{\kappa}$, whose deviations are large.

In Section 6, we choose the two-sided test with rejection criteria $|\hat{T}^L_{j,k}| > x_{\beta,j}$. We confirm this choice by the histograms of $\hat{T}^L_{j,3}$ with $\alpha = 1.5$ and $\kappa = 2$, see Figure 2.




(a) $\mathrm{GvMF}_{1,3}$ (b) $\mathrm{GvMF}_{2,3}$ (c) $\mathrm{GvMF}_{3,3}$

FIG 2. *Histograms of* $\hat{T}^L_{j,3}(\mathcal{X}_N), \mathcal{X}_N \sim \mathrm{GvMF}_{j,3}(\alpha,\kappa)$ *with* $\alpha = 1.5$ *and* $\kappa = 2$. *The maximal sample variance is 0.0006118*

We see that the statistics $\hat{T}^L_{j,3}$ have approximately symmetric distribution with mode at 0. Therefore, we put rejection region as $(-\infty, -x_\beta) \cup [x_\beta, +\infty)$, where critical values $x_\beta$ are obtained as a samples quantiles $\mathbb{P}_{H_0}(|\hat{T}_{i,3}| > x_{\beta,j}) \leq \beta, j = 1, 2, 3$ The corresponding values of $x_{\beta,j}$ with significance level $\beta = 0.05$ are presented in Table 3 for $\hat{T}^L_{1,3}$, in Table 4 for $\hat{T}^L_{2,3}$, and in Table 5 for $\hat{T}^L_{3,3}$.

We provide also the study of the goodness of fit test's power for the samples

TABLE 3
*Critical values $x_{\beta,1}$ for test statistic $\hat{T}^L_{1,3}$ and $\beta = 0.05$, with respect to $\alpha$ (rows) and $\kappa$ (columns), multiplied by $10^2$.*

| | 0.1 | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.5 | 4.884 | 4.626 | 5.128 | 5.281 | 5.069 | 4.960 | 4.726 | 4.999 | 5.132 | 5.360 | 5.301 |
| 1 | 4.727 | 4.792 | 4.879 | 4.736 | 4.757 | 4.843 | 4.920 | 5.217 | 4.951 | 5.278 | 5.091 |
| 1.5 | 4.935 | 4.914 | 4.657 | 4.731 | 4.745 | 4.879 | 4.849 | 4.995 | 5.152 | 4.990 | 5.009 |
| 2 | 4.916 | 4.873 | 4.839 | 4.982 | 4.801 | 4.987 | 4.752 | 4.934 | 4.829 | 5.076 | 5.010 |
| 2.5 | 4.916 | 4.921 | 5.276 | 4.932 | 4.854 | 4.852 | 4.711 | 4.700 | 4.860 | 4.983 | 5.373 |
| 3 | 4.687 | 4.821 | 4.783 | 4.626 | 4.926 | 4.704 | 4.656 | 4.690 | 4.631 | 4.706 | 4.844 |

TABLE 4
*Critical values $x_{\beta,2}$ for test statistic $\hat{T}^L_{2,3}$ and $\beta = 0.05$, with respect to $\alpha$ (rows) and $\kappa$ (columns), multiplied by $10^2$.*

| | 0.1 | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.5 | 4.921 | 4.795 | 4.750 | 4.962 | 4.711 | 5.015 | 4.742 | 5.182 | 5.430 | 5.374 | 5.388 |
| 1 | 4.718 | 4.688 | 4.858 | 4.937 | 4.996 | 4.829 | 4.870 | 5.065 | 5.234 | 5.287 | 5.449 |
| 1.5 | 4.698 | 4.916 | 4.943 | 4.714 | 4.824 | 4.988 | 4.964 | 4.613 | 4.975 | 5.267 | 5.263 |
| 2 | 4.920 | 4.989 | 4.852 | 4.611 | 4.890 | 4.869 | 5.147 | 5.037 | 4.941 | 5.423 | 5.217 |
| 2.5 | 5.066 | 4.753 | 5.034 | 4.805 | 4.768 | 4.930 | 4.973 | 5.346 | 5.283 | 5.138 | 5.074 |
| 3 | 4.713 | 4.417 | 4.731 | 4.907 | 5.001 | 5.007 | 4.870 | 5.023 | 5.015 | 5.083 | 5.116 |

TABLE 5
*Critical values $x_{\beta,3}$ for test statistic $\hat{T}^L_{3,3}$ and $\beta = 0.05$, with respect to $\alpha$ (rows) and $\kappa$ (columns), multiplied by $10^2$.*

| | 0.1 | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.5 | 4.932 | 5.095 | 4.843 | 4.863 | 5.040 | 4.995 | 5.000 | 5.477 | 5.810 | 5.219 | 5.527 |
| 1 | 4.956 | 4.793 | 4.636 | 4.912 | 4.896 | 4.873 | 4.959 | 5.251 | 5.196 | 5.565 | 5.917 |
| 1.5 | 4.620 | 4.873 | 4.781 | 4.892 | 4.839 | 5.006 | 4.806 | 4.824 | 5.016 | 5.239 | 5.370 |
| 2 | 4.727 | 5.049 | 4.804 | 4.567 | 4.811 | 4.651 | 4.842 | 4.999 | 4.836 | 5.162 | 5.191 |
| 2.5 | 4.925 | 4.799 | 4.938 | 4.831 | 4.735 | 4.834 | 5.024 | 4.739 | 4.917 | 4.899 | 4.703 |
| 3 | 4.960 | 4.901 | 4.904 | 4.649 | 5.037 | 4.843 | 4.898 | 4.852 | 4.828 | 4.895 | 4.974 |

from Fisher-Bingham distribution. In these simulations we put $\boldsymbol{\mu}_1 = (1,0,0)^{\mathrm{T}}$, $\boldsymbol{\mu}_2 = (0, \sqrt{2}/2, \sqrt{2}/2)^{\mathrm{T}}$, and consider the following series of hypotheses.

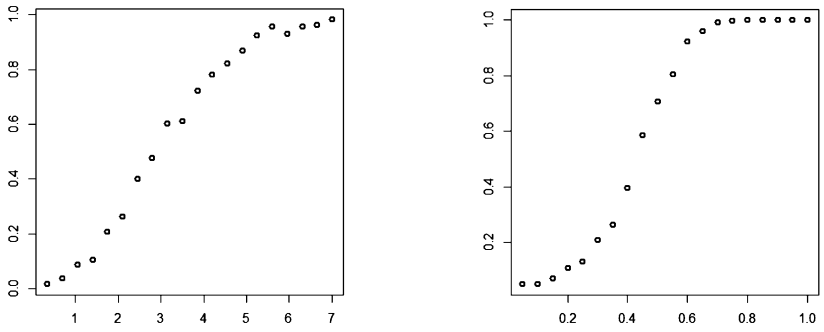
For type I:

- $H^1_0 : \mathcal{X}_N \sim \mathrm{GvMF}_{1,3}$,
- $H^1_{1,j} : \mathcal{X}_N$ has the Fisher-Bingham distribution with density $\propto \exp(3\boldsymbol{\mu}_1^{\mathrm{T}}\mathbf{x} + 0.35j(\boldsymbol{\mu}_2^{\mathrm{T}}\mathbf{x})^2)$, $\mathbf{x} \in \mathbb{S}^2$, $j = 1, \ldots, 20$.

For axial type:

- $H^2_0 : \mathcal{X}_N \sim \mathrm{GvMF}_{3,3}$,
- $H^2_{1,j} : \mathcal{X}_N$ has the Fisher-Bingham distribution with density proportional to $\exp(0.05j(\boldsymbol{\mu}_1^{\mathrm{T}}\mathbf{x}) + 6(\boldsymbol{\mu}_2^{\mathrm{T}}\mathbf{x})^2)$, $\mathbf{x} \in \mathbb{S}^2$, $j = 1, \ldots, 20$.

For each $j = 1, \ldots, 20$, we simulate 500 samples $\mathcal{X}^1_{j,N}$ under $H^1_{1,j}$ and $\mathcal{X}^2_{j,N}$ under $H^2_{1,j}$ with sample size $N = 1000$. We use the simulation procedure from the R package 'Directional'. In order to simplify commutations, we reject $H^1_0$ and $H^2_0$ if $|\hat{T}^L_{1,3}(\mathcal{X}^1_{j,N})| > x^{(1)}_\beta$ and $|\hat{T}^L_{2,3}(\mathcal{X}^2_{j,N})| > x^{(2)}_\beta$ respectively, where critical values




(a) Type I, $0.35j$ on the $x-$axis. (b) Axial type, $0.05j$ on the $x-$axis.

FIG 3. *Powers of goodness of fit tests* $H_0^1$ *vs* $H_{1,j}^1$ *(left) and* $H_0^2$ *vs* $H_{1,j}^2$ *(right),* $j = 1, \dots, 20$.

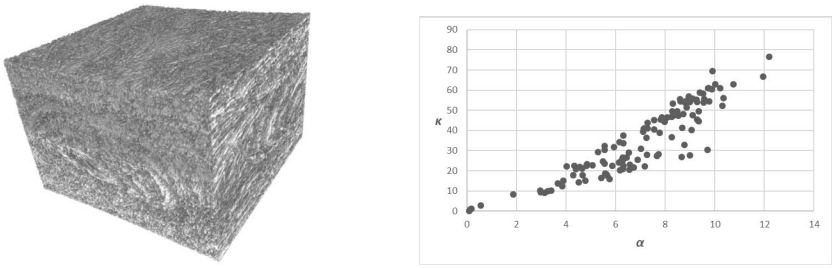
$x_\beta^{(1)} = 0.05373$ and $x_\beta^{(2)} = 0.05917$ are taken as maximum of $x_\beta$ from Tables 3 and 5 for significance level $\beta = 0.05$.

The ratios of rejections $H_0^1$ and $H_0^2$ are presented in Figure 3.

## 8. Application to a real data set

In this section, we apply the introduced goodness of fit test to the data set consists of fiber directions in a glass fibre reinforced composite material. The 3D-images of a fibre composite obtained by micro computed tomography and are provided by the Institute for Composite Materials (IVW) in Kaiserslautern, Germany, see Fig. 4b (left). The detailed description of the material can be found in [45] and it was the object of studies in [15] and [16], where the regions of anomaly behaviour of the fibres were found. The data set is provided by Prof. Claudia Redenbach (TU Kaiserslautern) and consists of local direction of fibres estimated by the tools of MAVI software [21]. Each data set entry $Y_k, k = ([1, 97] \times [1, 80] \times [1, 64]) \cap \mathbb{N}^3$ is the average of fibre local directions in small observation windows $\tilde{W}$ with $75 \times 75 \times 75$ voxels each. Note that some of such windows can be empty, or they might contain not enough material for direction computation. We denote by $J_W$ the collection of indexes $k$ such that $Y_k$ is non-empty. In the considered data set $|J_W| = 430741$ and its precise construction is in [15].

The estimating procedure of directions in MAVI software produces vectors on a unit sphere which are not necessarily symmetrically distributed. However, the fibres are not oriented, therefore we expect an axial distribution of their directions. We propose the symmetrization of the original sample by $\mathcal{X}_N = \{\mathbf{X_k} = \mathbf{Y_k}\xi_\mathbf{k}, \mathbf{k} \in J_W\}$, where $\xi_\mathbf{k}, \mathbf{k} \in J_W$ are i.i.d random variables with $\mathbb{P}(\xi_\mathbf{k} = +1) = \mathbb{P}(\xi_\mathbf{k} = -1) = \frac{1}{2}$. We separate the whole material into blocks $W_\mathbf{l}$, each of size $16 \times 15 \times 16$, such that $J_\mathbf{l} = J_W \cap [l_1, l_1 + 16) \times [l_2, l_2 + 15) \times [l_3, l_2 + 16)$ and consider subsamples $\mathcal{X}_\mathbf{l} = \{\mathbf{X_k}, \mathbf{k} \in J_\mathbf{l}\}$ with simple sizes $2736 \leq |\mathcal{X}_\mathbf{l}| \leq 3745$. For each subsample $\mathcal{X}_\mathbf{l}$ we provide the introduced goodness




(a) 3D image of a glass fibre reinforced composite material. 970×1469×1217 voxels, spacing: 4$\mu$m.

(b) Maximum likelihood estimates of $\alpha$ and $\kappa$ for each subsample $\mathcal{X}_{\mathbf{l}}$

FIG 4. *Testing on a glass fibre reinforced composite material.*

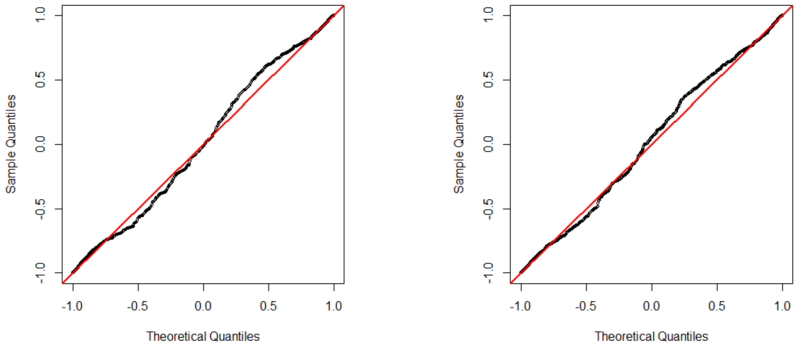
of fit test for distributions $\mathrm{GvMF}_{3,3}$, i.e., we test $H_{0,\mathbf{l}} : \mathcal{X}_{\mathbf{l}} \sim \mathrm{GvMF}_{3,3}$ vs. $H_{1,\mathbf{l}} : \mathcal{X}_{\mathbf{l}} \not\sim \mathrm{GvMF}_{3,3}$.

At first, we provide maximum likelihood estimation of parameters $\alpha$ and $\kappa$ (for the variety of their values $\hat{\alpha}_{\mathbf{l}}$ and $\hat{\kappa}_{\mathbf{l}}$ see Figure 4b). Second, we need to simulate the samples of statistics $\hat{T}_{3,k}(\mathcal{X}_{\mathbf{l}})$ under hypotheses $\mathcal{X}_{\mathbf{l}} \sim \mathrm{GvMF}_{3,3}(\hat{\alpha}_{\mathbf{l}}, \hat{\kappa}_{\mathbf{l}}, \cdot)$ based on samples sizes $|\mathcal{X}_{\mathbf{l}}|$. Unfortunately, our computational resources were limited and we have to group simulations with close values of $\hat{\alpha}_{\mathbf{l}}$ and $\hat{\kappa}_{\mathbf{l}}$. One can observe that the majority of $\hat{\alpha}_{\mathbf{l}}$ belongs to the interval $[3, 11]$ and the ratios $\frac{\hat{\kappa}_{\mathbf{l}}}{\hat{\alpha}_{\mathbf{l}}}$ are mostly in $[3, 7]$. Therefore, we simulate 800 samples $\mathcal{Y}_N \sim \mathrm{GvMF}_{3,3}(\alpha, \kappa, \cdot)$ each of size $N = 3500$ for all combinations of $\alpha \in \{4, 6, 8, 10\}$ and $\frac{\kappa}{\alpha} \in \{4, 6\}$ to obtain the corresponding empirical distributions of $\hat{T}_{3,3}(\mathcal{Y}_N)$.

Then we compute statistics $\hat{T}_{3,3}(\mathcal{X}_{\mathbf{l}})$ for each $\mathbf{l}$ and their $p$-values. We obtain that our goodness of fit test rejects almost all hypotheses $H_{0,\mathbf{l}}$ with significance level 0.05, and detects 3 regions with directional distributions $\mathrm{GvMF}_{3,3}$ (see Table 6 for the samples $\mathcal{X}_{\mathbf{l}}$ with $p$-values greater than 0.01). In order to illustrate how tight the fitted distributions are, we present for two blocks $W_{\mathbf{l}}$ with $\mathbf{l} = (49, 61, 1)$ and $\mathbf{l} = (49, 61, 1)$, the QQ-plots for samples $\{\hat{\boldsymbol{\mu}}_{\mathbf{l}}^{\mathrm{T}} \mathcal{X}_{\mathbf{l}}\}$ and distribution $f_3$ defined in (37) with parameters $\hat{\alpha}_{\mathbf{l}}, \hat{\kappa}_{\mathbf{l}}$, see Figure 5.

TABLE 6
*Results of goodness of fit tests $H_{0,\mathbf{l}}$ vs. $H_{1,\mathbf{l}}$ for fiber directions in glass fibre reinforced composite material.*

| $l_1$ | $l_2$ | $l_3$ | $\lvert\mathcal{X}_{\mathbf{l}}\rvert$ | $\hat{\alpha}$ | $\hat{\kappa}$ | $\hat{H}_{N,3}$ | $\hat{T}_{3,3}(\mathcal{X}_{\mathbf{l}})$ | $p$-value |
|---|---|---|---|---|---|---|---|---|
| 49 | 46 | 1 | 3434 | 8.80 | 53.90 | 0.4369 | 0.02344 | 0.0775 |
| 49 | 61 | 1 | 3222 | 8.53 | 47.62 | 0.7132 | 0.01976 | 0.1234 |
| 49 | 16 | 17 | 3474 | 8.84 | 53.63 | 0.4690 | 0.02987 | 0.0263 |
| 49 | 61 | 17 | 3364 | 10.22 | 60.91 | 0.4628 | 0.02946 | 0.0263 |
| 1 | 46 | 65 | 3319 | 7.25 | 36.45 | 1.0455 | 0.02057 | 0.1275 |




(a) $\mathbf{l} = (49, 61, 1), \hat{\alpha}_{\mathbf{l}} = 8.53, \hat{\kappa}_{\mathbf{l}} = 47.62$ (b) $\mathbf{l} = (1, 46, 65), \hat{\alpha}_{\mathbf{l}} = 7.25, \hat{\kappa}_{\mathbf{l}} = 36.45$

FIG 5. *QQ plots for samples* $\hat{\boldsymbol{\mu}}_{\mathbf{l}}^{\mathrm{T}} \mathcal{X}_{\mathbf{l}}$ *and distributions with density* $f_3$ (37).

## 9. Asymptotic behaviour of the test statistics

In this section we analyse the convergence in distribution of $d_N(\widehat{T}_{j,k}(\mathcal{X}_N) - a_N)$ as $N \to \infty$ with suitable sequences $\{d_n\}_{n\geq 1}$ and $\{a_n\}_{n\geq 1}$.

Denote by $E_{j,d}(\alpha, \kappa)$ the entropy values for the $\mathrm{GvMF}_{j,d}(\alpha, \kappa, \cdot)$ distribution. Theorem 4.1 and Propositions 3.1, 3.2, 3.3 give

$$E_{1,d}(\alpha, \kappa) = -\log c_{1,d}(\kappa, \alpha) - \frac{\kappa}{\alpha} \frac{A_1(\kappa, \alpha, \alpha) - A_1(-\kappa, \alpha, \alpha)}{A_1(\kappa, \alpha, 0) + A_1(-\kappa, \alpha, 0)}, \tag{38}$$

$$E_{2,d}(\alpha, \kappa) = -\log c_{2,d}(\kappa, \alpha) + \frac{\kappa}{\alpha} \frac{A_2(\kappa, \alpha, \alpha)}{A_2(\kappa, \alpha, 0)}, \tag{39}$$

$$E_{3,d}(\alpha, \kappa) = -\log c_{3,d}(\kappa, \alpha) - \frac{\kappa}{\alpha} \frac{A_1(\kappa, \alpha, \alpha)}{A_1(\kappa, \alpha, 0)}. \tag{40}$$

Therefore, statistics $\widehat{T}_{j,k}(\mathcal{X}_N)$ have the form $T_{j,k}(\mathcal{X}_N) = E_{j,d}(\hat{\alpha}, \hat{\kappa}) - \widehat{H}_{N,k}$. Considering distributional convergence, we could apply the delta method for the first component $E_{j,d}(\hat{\alpha}, \hat{\kappa})$ in order to show that $\sqrt{N}(E_{j,d}(\hat{\alpha}, \hat{\kappa}) - E_{j,d}(\alpha, \kappa)) \xrightarrow{d} N(0, \sigma_{1,j}^2)$, $N \to \infty$ for some $\sigma_{1,j} > 0$ and asymptotically normal estimators $\hat{\kappa}$ and $\hat{\alpha}$. Meanwhile for the entropy component $\widehat{H}_{N,k}$, we know that $\sqrt{N}(\widehat{H}_{N,k} - \mathbb{E}\widehat{H}_{N,k}) \xrightarrow{d} N(0, \sigma_{2,j}^2)$, $N \to \infty$ under some additional conditions, c.f. [6, 13, 41]. However, the covariance between two components $E_{j,d}(\hat{\alpha}, \hat{\kappa}))$ and $\widehat{H}_{N,k}$ seems to be analytically intractable and the decay of $\mathbb{E}\widehat{H}_{N,k}$ remains unknown. That is why, we can not provide analytical derivation of the asymptotic distribution of $\widehat{T}_{j,k}(\mathcal{X}_N)$ and use Monte-Carlo simulations.

We consider the case of known $\alpha = 1.5$, unknown $\kappa$ and $\boldsymbol{\mu}$. We simulate 1000 samples $\mathcal{X}_{j,N}$ from $\mathrm{GvMF}_{j,3}(\alpha, \kappa, \boldsymbol{\mu})$ for each combination of $N \in 50\{2, \ldots, 20\}$, $\kappa \in \{0.5, 2, 6\}$ and $\boldsymbol{\mu} = (0, 0, 0)^{\mathrm{T}}$. For each sample we estimate the parameter $\kappa$ by the method of moments (Remark 5.1) and by the maximum likelihood

estimator (Section 5.1). The entropy is estimated via $\widehat{H}_{N,3}$. Therefore, we obtain 1000 values of $\widehat{T}^M_{j,3}$ and $\widehat{T}^L_{j,3}$ for each combination of parameters $N, \kappa, \boldsymbol{\mu}$.

At first, we investigate the rate of convergence of $\mathbb{E}\widehat{T}^M_{j,3}(\mathcal{X}_N)$ and $\mathbb{E}\widehat{T}^L_{j,3}(\mathcal{X}_N)$ as $N \to \infty$. We examine the model

$$\log|\mathbb{E}\widehat{T}_{j,3}(\mathcal{X}_N)| = a_{j,\kappa} + b_{j,\kappa}\log N - \frac{1}{2}\log N, \quad \mathbf{Var}(\widehat{T}_{j,3}(\mathcal{X}_N)) = \frac{\sigma_{j,\kappa}}{N}$$

via the standard linear regression based on described Monte-Carlo simulations (see Figure 6). Table 7 shows that the slope values $b_{j,\kappa}$ are negative or close to zero for all three types of distributions, which means that the decay of $\mathbb{E}\widehat{T}_{j,3}(\mathcal{X}_N)$ is faster or equal than $N^{-1/2}$.

TABLE 7
*Slope values b in Log-Log regression* $\log|\mathbb{E}\widehat{T}_{j,3}(\mathcal{X}_N)| = a_{j,\kappa} + b_{j,\kappa}\log N - \frac{1}{2}\log N$.

| Statistic | $\widehat{T}^M_{1,3}$ | $\widehat{T}^L_{1,3}$ | $\widehat{T}^M_{2,3}$ | $\widehat{T}^L_{2,3}$ | $\widehat{T}^M_{3,3}$ | $\widehat{T}^L_{3,3}$ |
|---|---|---|---|---|---|---|
| $\kappa = 0.5$ | -0.7177 | -0.7313 | -0.3846 | -0.3864 | -0.4576 | -0.4698 |
| $\kappa = 2$ | -0.7611 | -0.9725 | -0.2940 | -0.3040 | -0.2368 | -0.2003 |
| $\kappa = 6$ | -0.2522 | -0.2492 | 0.0120 | 0.0264 | -0.2088 | -0.2201 |

It is interesting to investigate the impact of each component $E_{j,d}(\alpha,\kappa))$ and $\widehat{H}_{N,3}$ in the statistic $\widehat{T}_{j,3}$. Figures 7 and 8 show the similar Log-Log regression for biases $\mathbb{E}E_{j,d}(\alpha,\widehat{\kappa}) - E_{j,d}(\alpha,\kappa)$ and $\mathbb{E}\widehat{H}_{N,3} - E_{j,d}(\alpha,\kappa)$, respectively. On can observe that the decay for the entropy component is definitely slower than $N^{-1/2}$ for $\kappa = 6$, which means that $\sqrt{N}(\widehat{H}_{N,k} - E_{j,d}(\alpha,\kappa))$ does not converge to $N(0,\sigma^2_{2,j}), N \to \infty$. Meanwhile, the decay of $\mathbb{E}E_{j,d}(\widehat{\alpha},\widehat{\kappa}))$ is close to $N^{-1/2}$. Although the results for each component separately, the decay of the mean $\mathbb{E}\widehat{T}_{j,3}(\mathcal{X}_N)$ is still faster that $N^{-1/2}$, which can be explained by strong correlation between two components (see Figure 9).

We present values of $\sqrt{N}\sqrt{\mathbf{Var}(\widehat{T}_{j,3}(\mathcal{X}_N))}$ as box-plots in Figure 10 and deduce that they are approximately a constant with the mean value $\sigma^2 = 0.7842$.

Figure 2 suggests that the distribution of $\widehat{T}_{j,3}(\mathcal{X}_N)$ is asymptotically normal for large $N$. The corresponding Q-Q plots for $\widehat{T}^L_{j,3}(\mathcal{X}_N)$ with $N = 1000$ and $\kappa = 2$ in Figure 11 confirm that the distributions belong to a Gaussian family. We verify the assumption about an asymptotic Gaussian distribution of test statistics $\widehat{T}^L_{j,3}$ via Shapiro-Wilk test for normality and record the $p$-value returned by the test. Figure 12 shows how these $p$-values behave as $N$ increases. One can observe that the normal hypothesis cannot be rejected for samples of size $N = 500$ or more.

Thus, we confirm numerically that the limiting distribution of $\widehat{T}^L_{j,3}$ is Gaussian and its variance decays with order $N^{-1}$. We use this fact to construct the approximation of the empirical critical values $x_\beta$ for tests from Section 7.3 by

$$x_\beta \approx x^a_\beta := \frac{z_\beta\sigma + \mu_{j,\kappa}}{\sqrt{N}}, \tag{41}$$

where $z_\beta$ is a quantile of a standard normal law. Then, $\mu_{j,\kappa}$ must be approximately equal to $\sqrt{N}(x_\beta - z_\beta\sigma)$. We put these values with $\beta = 0.05$ in Figure

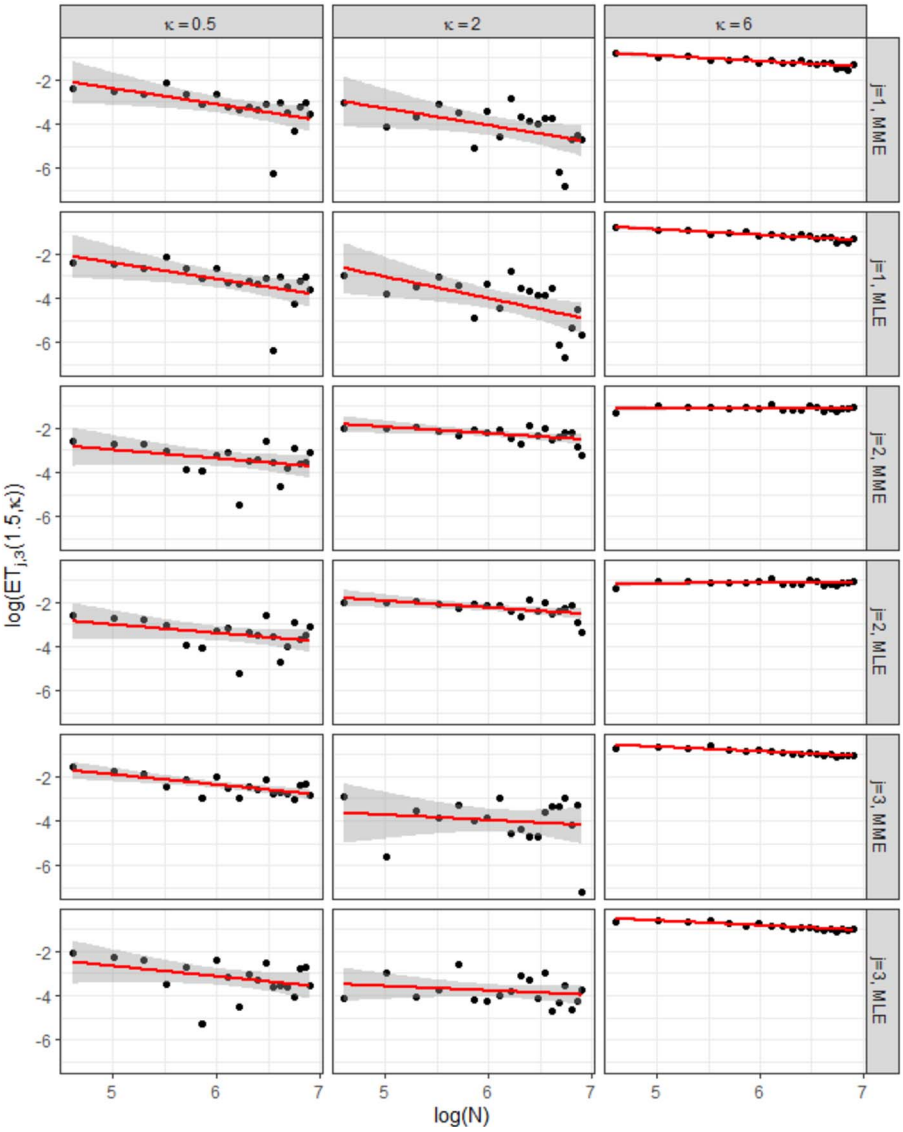



FIG 6. *Log-Log regression* $\log|\mathbb{E}\widehat{T}_{j,3}(\mathcal{X}_N)| = a_{j,\kappa} + b_{j,\kappa}\log N - \frac{1}{2}\log N$.

13 and compute $\mu_{j,\kappa}$ as $\max_{N\in[500,1000]}\sqrt{N}(x_{0.05} - z_{0.05}\sigma)$ (see Table 8). In

TABLE 8
*Values of* $\mu_{j,\kappa}$ *for* $\beta = 0.05$.

| Statistic | $\widehat{T}^L_{1,3}$ | $\widehat{T}^L_{2,3}$ | $\widehat{T}^L_{3,3}$ |
|---|---|---|---|
| $\kappa = 0.5$ | 0.0555 | 0.1077 | 0.0394 |
| $\kappa = 2$ | 0.1499 | 0.2557 | 0.1487 |
| $\kappa = 6$ | 0.3813 | 0.5392 | 0.4984 |

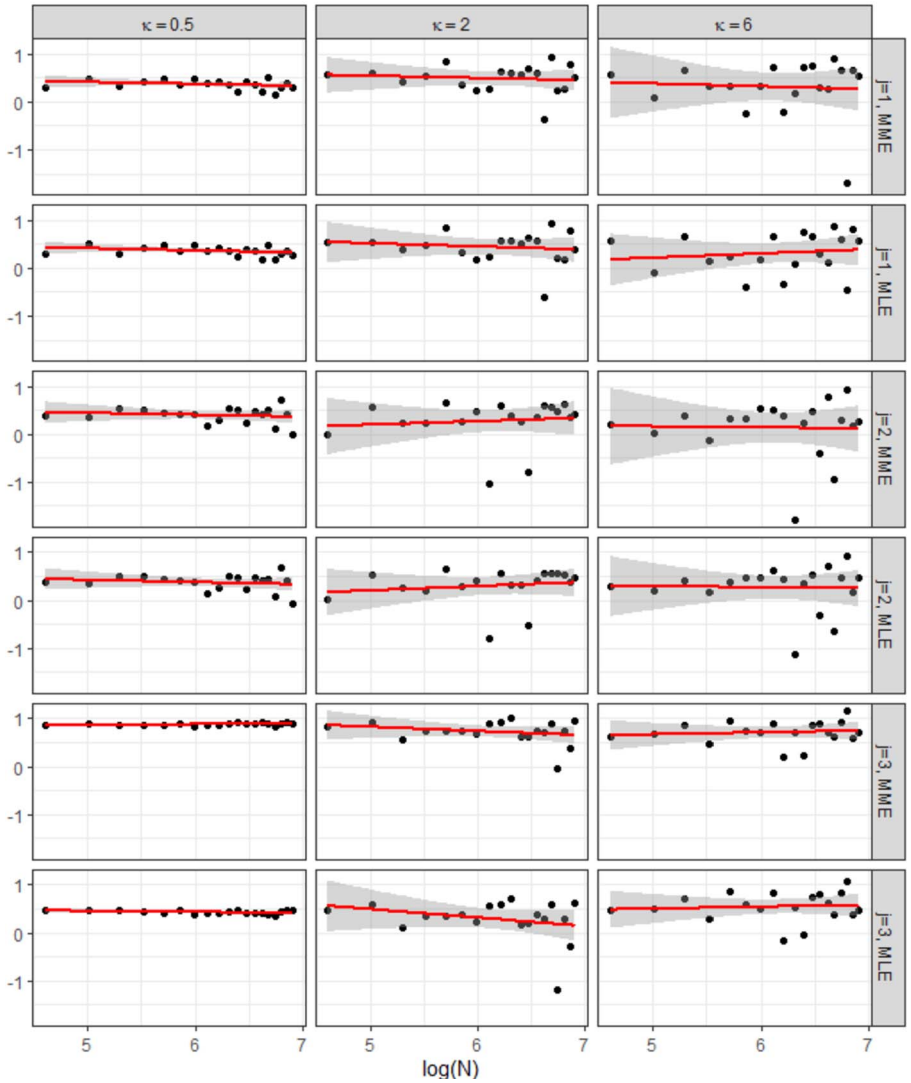



FIG 7. *Log-Log regression* $\log|\mathbb{E}E_{j,d}(\alpha,\widehat{\kappa}) - E_{j,d}(\alpha,\kappa)| = a_{j,\kappa} + b_{j,\kappa}\log N - \frac{1}{2}\log N$.

such case, $x^a_{0.05} \geq x_{0.05}$ and the I-type errors of the tests based on approximated quantiles $x^a_{0.05}$ are less or equal 0.05 but remain separated from 0, see Figure 14.

## Appendix

Here we present tables of mean square errors of estimates of $\kappa$ and $\alpha$ and plots with realizations of the generalized von Mises-Fisher distributions.

For each type of distributions $\mathrm{GvMF}_{1,3}$, $\mathrm{GvMF}_{2,3}$, $\mathrm{GvMF}_{3,3}$ we simulate 1000

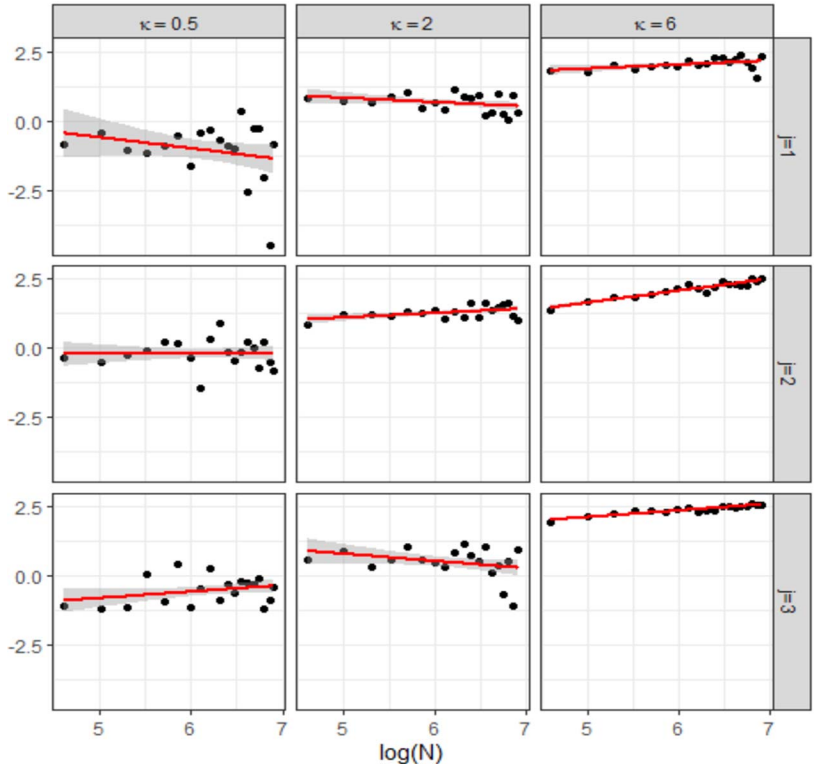



FIG 8. *Log-Log regression* $\log|\mathbb{E}\widehat{H}_{N,3} - E_{j,d}(\alpha,\kappa)| = a_{j,\kappa} + b_{j,\kappa}\log N - \frac{1}{2}\log N$.

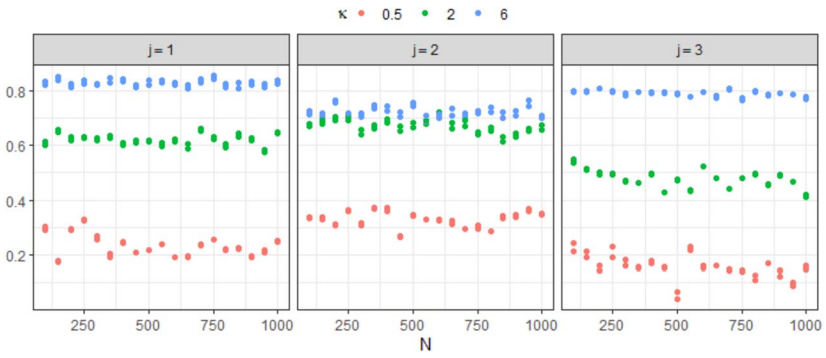



FIG 9. *Correlations between* $\widehat{H}_{N,3}$ *and* $E_{j,d}(1.5,\widehat{\kappa})$.

samples with $N = 1000$ entries each for several values of $\alpha \in \{0.5, 1, 1.5, 2, 2.5, 3\}$ and $\kappa \in \{0.5, 1, 1.5, 2, 2.5, 3, 4, 5, 6\}$. For each sample we compute maximum likelihood estimates $\hat{\boldsymbol{\mu}}_L, \hat{\alpha}_L, \hat{\kappa}_L$ and moment estimates $\hat{\boldsymbol{\mu}}_M, \hat{\alpha}_M, \hat{\kappa}_M$. We present the sample mean square errors of $\hat{\alpha}_L$ and $\hat{\alpha}_M$ in Tables 9 (type I), 11 (type II),

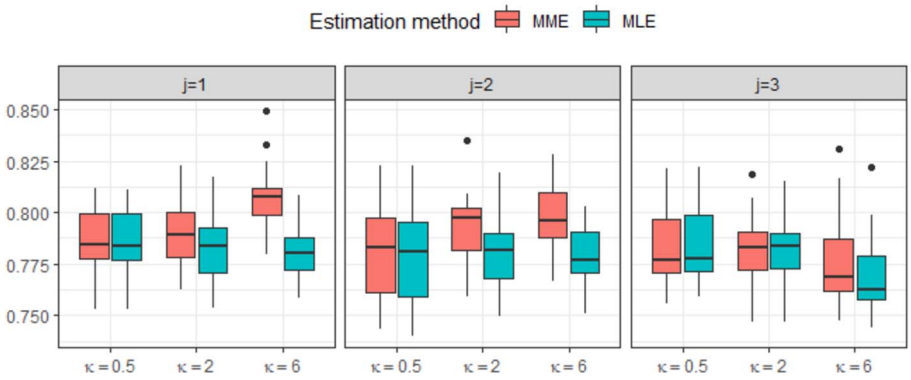



FIG 10. *Values of* $\sqrt{N\mathbf{Var}(\widehat{T}_{j,3}(\mathcal{X}_N))}$

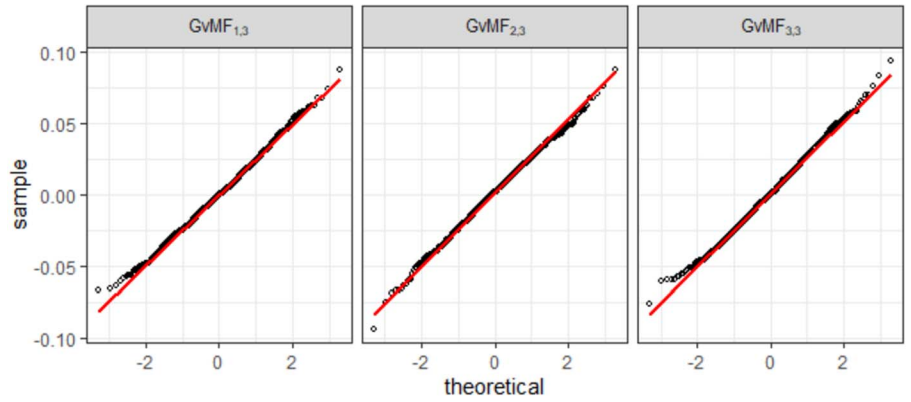



FIG 11. *Q-Q plots of empirical distribution of* $\widehat{T}^L_{j,3}(\mathcal{X}_N)$ *with* $\kappa = 2$ *and a Gaussian distribution.*

and 13 (axial type). The mean square errors of $\hat{\kappa}_L$ and $\hat{\kappa}_M$ can be found in Tables 10 (Type I), 12 (Type II), and 14 (axial data). We group error values of $\hat{\kappa}_L$, $\hat{\kappa}_M$ and $\hat{\alpha}_L$, $\hat{\alpha}_M$ in order to decide which method is more appropriate for parameter estimation.

Let us illustrate the generalized von Mises-Fisher distributions on 2-dimensional sphere by several samples with 1000 entries. For all samples we fix mean direction $\boldsymbol{\mu} = (0, \sqrt{2}/2, \sqrt{2}/2)$. For different values of $\alpha$ and $\kappa$ we present locations of samples entries on a unit sphere: for Type I, see Figure 16 (with $\alpha = 0.5$) and Figure 18 (with $\alpha = 1.5$); for Type 2, Figures 20 and 22 with $\alpha = 0.5$ and $\alpha = 1.5$, respectively, and the samples of axial data are presented in Figures 24 (with $\alpha = 0.5$) and 26 (with $\alpha = 1.5$). The corresponding histograms and probability densities of random variables $\boldsymbol{\mu}^{\mathrm{T}}\mathbf{X}_i, i = 1, 2, 3$ can be found in Figures 15 ($\alpha = 0.5$) and 17 ($\alpha = 1.5$) for Type I, in Figures 19 ($\alpha = 0.5$) and 21 ($\alpha = 1.5$) for Type II, and in Figures 23 ($\alpha = 0.5$) and 25 ($\alpha = 1.5$) for axial data.

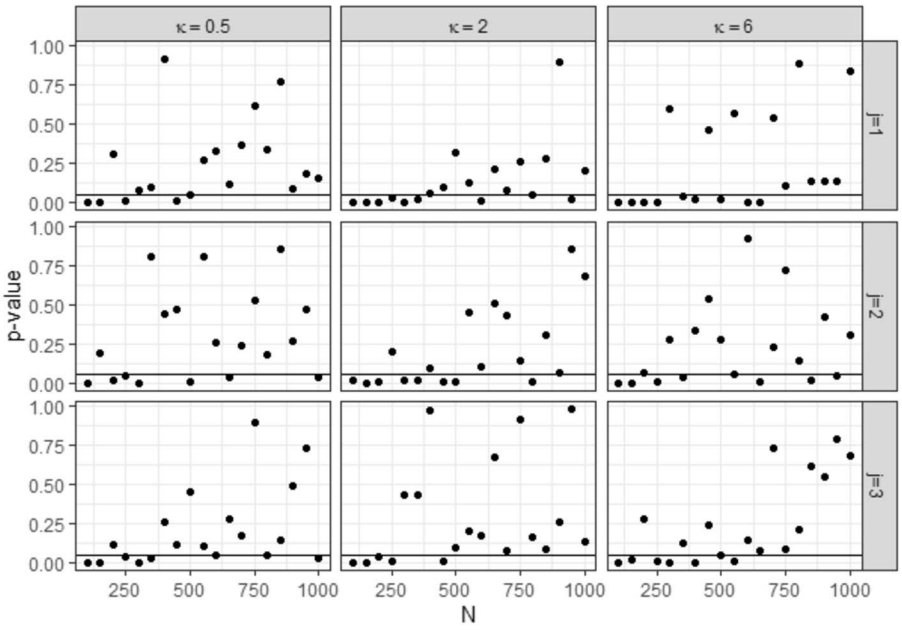



FIG 12. *Shapiro-Wilk p-values as N increases for different values of κ (1000 repetitions).*

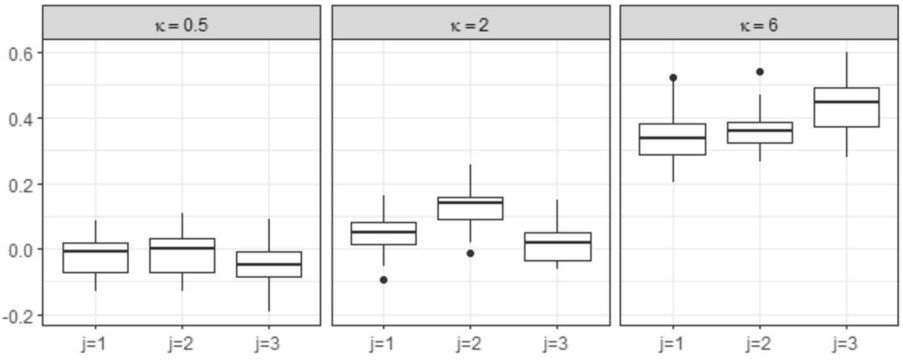



FIG 13. *Values of* $\sqrt{N}(x_\beta - z_\beta\sigma)$ *for* $\beta = 0.05$.

## Acknowledgment

The authors are grateful to Prof. Katja Schladitz for the help with the real data, Martin Gurka and Sebastian Nissle (Institut für Verbundwerkstoffe, Kaiserslautern) for permission to reuse the tomographic images, Prof. Claudia Redenbach for providing the data set of fiber directions. We would like to thank the Editor, Domenico Marinucci, and the anonymous referee for their insightful comments and suggestions, that led to an improvement of a previous version of

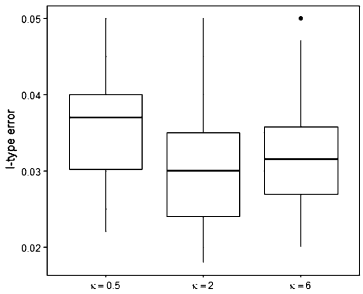



FIG 14. *I-type error of the test based on approximated critical values.*

this work.

## References

[1] ABADIE, J., ABBOTT, B., ABBOTT, R., ABERNATHY, M. et al. (2011). Directional limits on persistent gravitational waves using LIGO S5 science data. *Physical Review Letters* **107** 271102.

[2] AGUILAR, M., CAVASONZA, L. A., AMBROSI, G., ARRUDA, L., ATTIG, N. et al. (2020). The Alpha Magnetic Spectrometer (AMS) on the international space station: Part II–Results from the first seven years. *Physics reports*.

[3] ANCHORDOQUI, L. A., BERGMAN, D. R., BERTAINA, M. E. et al. (2020). Performance and science reach of the Probe of Extreme Multimessenger Astrophysics for ultrahigh-energy particles. *Physical Review D* **101** 023012. MR4232708

[4] BERGER, M. and GOSTIAUX, B. (1988). *Differential Geometry: Manifolds, Curves, and Surfaces. Graduate Texts in Mathematics* **115**. Springer-Verlag, New York. MR0917479

[5] BERRETT, T. B. and SAMWORTH, R. J. (2019). Nonparametric independence testing via mutual information. *Biometrika* **106** 547–566. MR3992389

[6] BERRETT, T. B., SAMWORTH, R. J. and YUAN, M. (2019). Efficient multivariate entropy estimation via $k$-nearest neighbour distances. *Ann. Statist.* **47** 288–318. MR3909934

[7] BINGHAM, M. and MARDIA, K. (1975). Maximum likelihood characterization of the von Mises distribution. In *A Modern Course on Statistical Distributions in Scientific Work* 387–398. Springer.

[8] BULINSKI, A. and DIMITROV, D. (2019). Statistical estimation of the Shannon entropy. *Acta Math. Sin. (Engl. Ser.)* **35** 17–46. MR3917990 MR3917990

[9] CACCIANIGA, L. (2019). Anisotropies of the highest energy cosmic-ray events recorded by the Pierre Auger Observatory in 15 years of operation. In *36th International Cosmic Ray Conference* **358** 206. SISSA Medialab.

[10] Cadirci, M. S., Evans, D., Leonenko, N. and Makogin, V. (2020). Entropy-based test for generalized Gaussian distributions. arXiv preprint arXiv:2010.06284.
[11] Carpio, J. and Gago, A. (2017). Roadmap for searching cosmic rays correlated with the extraterrestrial neutrinos seen at IceCube. *Physical Review D* **95** 123009.
[12] Cutting, C., Paindaveine, D. and Verdebout, T. (2020). On the power of axial tests of uniformity on spheres. *Electron. J. Stat.* **14** 2123–2154. MR4097051 MR4097051
[13] Delattre, S. and Fournier, N. (2017). On the Kozachenko-Leonenko entropy estimator. *J. Statist. Plann. Inference* **185** 69–93. MR3612672 MR3612672
[14] Denton, P. B. and Tamborra, I. (2018). Exploring the properties of choked gamma-ray bursts with IceCube's high-energy neutrinos. *The Astrophysical Journal* **855** 37.
[15] Dresvyanskiy, D., Karaseva, T., Mitrofanov, S., Redenbach, C., Schwaar, S., Makogin, V. and Spodarev, E. (2019). Application of clustering methods to anomaly detection in fibrous media. *IOP Conference Series: Materials Science and Engineering* **537** 022001.
[16] Dresvyanskiy, D., Karaseva, T., Makogin, V., Mitrofanov, S., Redenbach, C. and Spodarev, E. (2020). Detecting anomalies in fibre systems using 3-dimensional image data. *Stat. Comput.* **30** 817–837. MR4108679
[17] Duerinckx, M. and Ley, C. (2012). Maximum likelihood characterization of rotationally symmetric distributions on the sphere. *Sankhya A* **74** 249–262. MR3021559
[18] Evans, D. (2008). A law of large numbers for nearest neighbour statistics. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **464** 3175–3192. MR2448521
[19] Evans, D., Jones, A. J. and Schmidt, W. M. (2002). Asymptotic moments of near-neighbour distance distributions. *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.* **458** 2839–2849. MR1987515 MR1987515
[20] Fang, K. T. and Zhang, Y. T. (1990). *Generalized Multivariate Analysis*. Springer-Verlag, Berlin; Science Press Beijing, Beijing. MR1079542
[21] Fraunhofer ITWM, Department of Image Processing (2005). MAVI – Modular Algorithms for Volume Images. http://www.mavi-3d.de.
[22] Gao, W., Oh, S. and Viswanath, P. (2018). Demystifying fixed $k$-nearest neighbor information estimators. *IEEE Trans. Inform. Theory* **64** 5629–5661. MR3832327 MR3832327
[23] García-Portugués, E., Paindaveine, D. and Verdebout, T. (2020). On optimal tests for rotational symmetry against new classes of hyperspherical distributions. *J. Amer. Statist. Assoc.* **115** 1873–1887. MR4189764 MR4189764
[24] Gatto, R. and Jammalamadaka, S. R. (2007). The generalized von Mises distribution. *Stat. Methodol.* **4** 341–353. MR2380560 MR2380560
[25] Goria, M. N., Leonenko, N. N., Mergel, V. V. and Novi Inver-

ARDI, P. L. (2005). A new class of random vector entropy estimators and its applications in testing statistical hypotheses. *J. Nonparametr. Stat.* **17** 277–297. MR2129834 MR2129834

[26] JAMMALAMADAKA, S. R. and TERDIK, G. H. (2019). Harmonic analysis and distribution-free inference for spherical distributions. *Journal of Multivariate Analysis* **171** 436–451. MR3910506

[27] KHANIN, A. and MORTLOCK, D. J. (2016). A Bayesian analysis of the 69 highest energy cosmic rays detected by the Pierre Auger Observatory. *Monthly Notices of the Royal Astronomical Society* **460** 2765–2778.

[28] KOZACHENKO, L. F. and LEONENKO, N. N. (1987). Sample estimate of the entropy of a random vector. *Problems of Information Transmission* **23** 95–101. MR0908626

[29] LEONENKO, N., PRONZATO, L. and SAVANI, V. (2008). A class of Rényi information estimators for multidimensional densities. *Ann. Statist.* **36** 2153–2182. Corrected by Leonenko, N. and Pronzato, L. Correction: A class of Rényi information estimators for multidimensional densities. Ann. Statist. 38, 3837–3838. MR2458183 MR2766870

[30] LEY, C. and VERDEBOUT, T. (2017). *Modern Directional Statistics. Chapman & Hall/CRC Interdisciplinary Statistics Series*. CRC Press, Boca Raton, FL. MR3752655 MR3752655

[31] LEYTON, M., DYE, S. and MONROE, J. (2017). Exploring the hidden interior of the Earth with directional neutrino measurements. *Nature communications* **8** 1–11.

[32] LI, S., MNATSAKANOV, R. M. and ANDREW, M. E. (2011). $k$-nearest neighbor based consistent entropy estimation for hyperspherical distributions. *Entropy* **13** 650–667. MR2784148 MR2784148

[33] LUND, U. and JAMMALAMADAKA, S. R. (2000). An entropy-based test for goodness of fit on the von Mises distribution. *J. Statist. Comput. Simulation* **67** 319–332. MR1806900 MR1806900

[34] LUTWAK, E., YANG, D. and ZHANG, G. (2004). Moment-entropy inequalities. *Ann. Probab.* **32** 757–774. MR2039942 MR2039942

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

[36] MARDIA, K. V. and JUPP, P. E. (2000). *Directional Statistics. Wiley Series in Probability and Statistics*. John Wiley & Sons, Ltd., Chichester. MR1828667

[37] MARINUCCI, D. and PECCATI, G. (2011). *Random Fields on the Sphere: Representation, Limit Theorems and Cosmological Applications. London Mathematical Society Lecture Note Series* **389**. Cambridge University Press, Cambridge Representation, limit theorems and cosmological applications. MR2840154

[38] MISRA, N., SINGH, H. and HNIZDO, V. (2010). Nearest neighbor estimates of entropy for multivariate circular distributions. *Entropy* **12** 1125–1144. MR2653296

[39] MOURITSEN, H. and MOURITSEN, O. (2000). A Mathematical Expectation Model for Bird Navigation based on the Clock-and-Compass Strategy.

*Journal of Theoretical Biology* **207** 283 - 291.

[40] Paindaveine, D. and Verdebout, T. (2020). Inference for spherical location under high concentration. *Ann. Statist.* **48** 2982–2998. MR4152631 MR4152631

[41] Penrose, M. D. and Yukich, J. E. (2011). Laws of large numbers and nearest neighbor distances. In *Advances in directional and linear statistics* 189–199. Physica-Verlag/Springer, Heidelberg. MR2767541 MR2767541

[42] Penrose, M. D. and Yukich, J. E. (2013). Limit theory for point processes in manifolds. *Ann. Appl. Probab.* **23** 2161–2211. MR3127932 MR3127932

[43] Pewsey, A. and García-Portugués, E. (2021). Recent advances in directional statistics. *TEST*. MR4242171

[44] Pischiutta, M., Rovelli, A., Salvini, F., Di Giulio, G. and Ben-Zion, Y. (2013). Directional resonance variations across the Pernicana Fault, Mt Etna, in relation to brittle deformation fields. *Geophysical Journal International* **193** 986-996.

[45] Wirjadi, O., Godehardt, M., Schladitz, K., Wagner, B., Rack, A., Gurka, M., Nissle, S. and Noll, A. (2014). Characterization of multilayer structures of fiber reinforced polymer employing synchrotron and laboratory X-ray CT. *International Journal of Materials Research* **105** 645–654.

[46] Yukich, J. E. (2006). *Probability Theory of Classical Euclidean Optimization Problems*. Springer. MR1632875

TABLE 9
*Mean square errors of $\hat{\alpha}_M$ (top raw) and $\hat{\alpha}_L$ (bottom raw) for* $\mathrm{GvMF}_{1,3}$ *distribution, for different values of $\alpha$ (rows) and $\kappa$ (columns)*

| | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|---|
| 0.5 | 0.01131 | 0.00309 | 0.00229 | 0.00334 | 0.00720 | 0.00234 | 0.00067 | 0.00044 | 0.00046 |
| | 0.01103 | 0.00294 | 0.00209 | 0.00290 | 0.00600 | 0.01028 | 0.01568 | 0.02000 | 0.02222 |
| 1 | 0.27260 | 0.03760 | 0.01778 | 0.01126 | 0.00766 | 0.00633 | 0.00686 | 0.00950 | 0.02210 |
| | 0.21929 | 0.03207 | 0.01518 | 0.00838 | 0.00569 | 0.00480 | 0.00508 | 0.00644 | 0.00850 |
| 1.5 | 2.89363 | 0.44954 | 0.12728 | 0.07166 | 0.04616 | 0.03330 | 0.02049 | 0.01535 | 0.01465 |
| | 2.75421 | 0.20300 | 0.07216 | 0.03613 | 0.02355 | 0.01725 | 0.01081 | 0.00904 | 0.00793 |
| 2 | 8.34069 | 2.90354 | 0.89834 | 0.44383 | 0.27728 | 0.16970 | 0.08593 | 0.05445 | 0.04011 |
| | 10.1685 | 1.30990 | 0.30051 | 0.13914 | 0.08601 | 0.05726 | 0.03554 | 0.02187 | 0.01650 |
| 2.5 | 8.68665 | 6.79656 | 3.96727 | 2.18844 | 1.23084 | 0.62982 | 0.35996 | 0.18528 | 0.13112 |
| | 15.4429 | 4.46642 | 1.21787 | 0.42840 | 0.23206 | 0.16395 | 0.08478 | 0.05083 | 0.03633 |
| 3 | 10.1863 | 8.74483 | 6.31636 | 6.33024 | 3.48461 | 2.84763 | 1.35829 | 0.62283 | 0.50274 |
| | 18.7129 | 8.33385 | 3.35711 | 1.24416 | 0.60680 | 0.36062 | 0.20265 | 0.11569 | 0.08699 |

TABLE 10
*Mean square error of $\hat{\kappa}_M$ (top raw) and $\hat{\kappa}_L$ (bottom raw) for* $\mathrm{GvMF}_{1,3}$ *distribution, for different values of $\alpha$ (rows) and $\kappa$ (columns)*

| | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|---|
| 0.5 | 0.01899 | 0.01699 | 0.01898 | 0.02602 | 0.03563 | 0.03419 | 0.03749 | 0.04394 | 0.06683 |
| | 0.01717 | 0.01598 | 0.01727 | 0.02396 | 0.03306 | 0.04987 | 0.07889 | 0.10651 | 0.13889 |
| 1 | 0.22319 | 0.07076 | 0.06534 | 0.06965 | 0.06449 | 0.06737 | 0.08770 | 0.11201 | 0.19431 |
| | 0.11194 | 0.05895 | 0.05552 | 0.05215 | 0.05029 | 0.05277 | 0.06947 | 0.08617 | 0.10606 |
| 1.5 | 2.11049 | 0.64790 | 0.26637 | 0.25165 | 0.23415 | 0.22145 | 0.20812 | 0.20797 | 0.23012 |
| | 1.19008 | 0.19097 | 0.14496 | 0.12205 | 0.11508 | 0.11355 | 0.10874 | 0.12384 | 0.12786 |
| 2 | 3.68444 | 4.15468 | 1.72159 | 1.18887 | 1.05376 | 0.82951 | 0.63399 | 0.54231 | 0.49912 |
| | 3.19240 | 0.92650 | 0.39620 | 0.29374 | 0.27322 | 0.24016 | 0.24388 | 0.20779 | 0.19860 |
| 2.5 | 2.17772 | 6.54718 | 7.27056 | 5.81158 | 4.41608 | 2.47149 | 2.26686 | 1.43704 | 1.34411 |
| | 4.02490 | 2.52365 | 1.19600 | 0.61967 | 0.50278 | 0.50154 | 0.41390 | 0.35365 | 0.35024 |
| 3 | 1.87215 | 5.16488 | 7.11001 | 13.3072 | 10.1227 | 11.0565 | 7.43739 | 4.25858 | 4.56458 |
| | 4.29139 | 3.63187 | 2.47833 | 1.40486 | 0.99006 | 0.80562 | 0.74920 | 0.62553 | 0.61910 |

TABLE 11
*Mean square error of $\hat{\alpha}_M$ (top raw) and $\hat{\alpha}_L$ (bottom raw) for* $\mathrm{GvMF}_{2,3}$ *distribution, for different values of $\alpha$ (rows) and $\kappa$ (columns)*

| | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|---|
| 0.5 | 0.04439 | 0.00892 | 0.00384 | 0.00240 | 0.00200 | 0.00176 | 0.00174 | 0.00208 | 0.00112 |
| | 0.03719 | 0.00571 | 0.00244 | 0.00131 | 0.00111 | 0.00082 | 0.00069 | 0.00073 | 0.00072 |
| 1 | 0.11476 | 0.02662 | 0.01238 | 0.00779 | 0.00620 | 0.00508 | 0.00415 | 0.00424 | 0.00384 |
| | 0.15498 | 0.02641 | 0.01200 | 0.00721 | 0.00555 | 0.00469 | 0.00369 | 0.00357 | 0.00341 |
| 1.5 | 0.20015 | 0.05429 | 0.02785 | 0.01742 | 0.01432 | 0.01291 | 0.00981 | 0.00956 | 0.00925 |
| | 0.30670 | 0.05952 | 0.02937 | 0.01765 | 0.01473 | 0.01303 | 0.01006 | 0.01004 | 0.00962 |
| 2 | 0.30415 | 0.09795 | 0.05169 | 0.03786 | 0.02962 | 0.02579 | 0.02137 | 0.02306 | 0.02314 |
| | 0.40585 | 0.09406 | 0.05020 | 0.03618 | 0.02640 | 0.02466 | 0.01993 | 0.02191 | 0.02198 |
| 2.5 | 0.56165 | 0.12324 | 0.08668 | 0.06643 | 0.05643 | 0.05414 | 0.05298 | 0.04757 | 0.05055 |
| | 0.51124 | 0.12344 | 0.06832 | 0.05349 | 0.04539 | 0.04304 | 0.04235 | 0.03748 | 0.04087 |
| 3 | 0.71927 | 0.18736 | 0.14068 | 0.11660 | 0.11100 | 0.11534 | 0.09450 | 0.10280 | 0.10102 |
| | 0.52430 | 0.16199 | 0.09598 | 0.08051 | 0.07280 | 0.07200 | 0.06493 | 0.07062 | 0.06307 |

TABLE 12
*Mean square error of $\hat{\kappa}_M$ (top raw) and $\hat{\kappa}_L$ (bottom raw) for* $\mathrm{GvMF}_{2,3}$ *distribution, for different values of $\alpha$ (rows) and $\kappa$ (columns)*

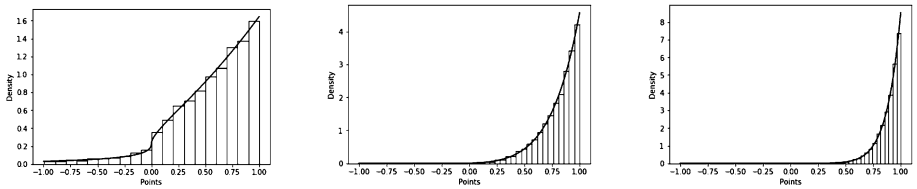
| | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|---|
| 0.5 | 0.00593 | 0.00788 | 0.01080 | 0.01751 | 0.03210 | 0.05088 | 0.14597 | 0.39689 | 0.26720 |
| | 0.00585 | 0.00619 | 0.00815 | 0.01132 | 0.02080 | 0.02932 | 0.06803 | 0.16595 | 0.29599 |
| 1 | 0.00460 | 0.00611 | 0.00912 | 0.01508 | 0.02487 | 0.03871 | 0.08568 | 0.18878 | 0.31815 |
| | 0.00504 | 0.00602 | 0.00885 | 0.01432 | 0.02260 | 0.03695 | 0.07654 | 0.16276 | 0.28533 |
| 1.5 | 0.00382 | 0.00496 | 0.00712 | 0.01159 | 0.02027 | 0.03192 | 0.07540 | 0.16707 | 0.28675 |
| | 0.00433 | 0.00497 | 0.00715 | 0.01170 | 0.02050 | 0.03221 | 0.07713 | 0.17453 | 0.29832 |
| 2 | 0.00507 | 0.00512 | 0.00658 | 0.01078 | 0.01733 | 0.02968 | 0.07776 | 0.17039 | 0.33426 |
| | 0.00594 | 0.00515 | 0.00663 | 0.01090 | 0.01729 | 0.02992 | 0.07590 | 0.16294 | 0.32125 |
| 2.5 | 0.00725 | 0.00705 | 0.00885 | 0.01145 | 0.01878 | 0.03171 | 0.08975 | 0.19040 | 0.36456 |
| | 0.00797 | 0.00761 | 0.00847 | 0.01155 | 0.01850 | 0.03021 | 0.07865 | 0.16314 | 0.29699 |
| 3 | 0.01072 | 0.01052 | 0.01256 | 0.01447 | 0.02044 | 0.03274 | 0.09333 | 0.24011 | 0.45512 |
| | 0.01004 | 0.01033 | 0.01179 | 0.01420 | 0.02058 | 0.03070 | 0.07711 | 0.17748 | 0.30782 |

TABLE 13
*Mean square error of $\hat{\alpha}_M$ (top raw) and $\hat{\alpha}_L$ (bottom raw) for* $\mathrm{GvMF}_{3,3}$ *distribution, for different values of $\alpha$ (rows) and $\kappa$ (columns)*

| | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|---|
| 0.5 | 0.18987 | 0.05705 | 0.02550 | 0.01188 | 0.00814 | 0.00512 | 0.00236 | 0.00128 | 0.00045 |
| | 0.16463 | 0.04162 | 0.02490 | 0.01907 | 0.01631 | 0.01637 | 0.02073 | 0.02427 | 0.02521 |
| 1 | 0.72857 | 0.23713 | 0.11809 | 0.07490 | 0.05300 | 0.03753 | 0.02852 | 0.01925 | 0.01203 |
| | 1.99071 | 0.21501 | 0.08130 | 0.04956 | 0.03309 | 0.02432 | 0.02189 | 0.01923 | 0.02272 |
| 1.5 | 1.41746 | 0.69305 | 0.28007 | 0.16580 | 0.09577 | 0.07204 | 0.04872 | 0.03733 | 0.03224 |
| | 7.55889 | 1.11221 | 0.28031 | 0.14598 | 0.08200 | 0.06144 | 0.04110 | 0.03296 | 0.02763 |
| 2 | 2.16193 | 1.34087 | 0.66370 | 0.37039 | 0.20955 | 0.15980 | 0.09065 | 0.05605 | 0.04741 |
| | 11.8352 | 4.29336 | 1.22665 | 0.40504 | 0.20974 | 0.16143 | 0.08807 | 0.05484 | 0.04622 |
| 2.5 | 3.14653 | 2.42570 | 1.51950 | 0.79157 | 0.47397 | 0.33253 | 0.17312 | 0.11837 | 0.08002 |
| | 17.0535 | 7.75701 | 3.11424 | 1.16675 | 0.52106 | 0.34421 | 0.18537 | 0.12071 | 0.08197 |
| 3 | 4.87133 | 4.04307 | 2.64681 | 1.72811 | 1.07167 | 0.70276 | 0.33226 | 0.20734 | 0.15167 |
| | 17.9717 | 10.7576 | 5.75811 | 2.78415 | 1.28941 | 0.74024 | 0.35797 | 0.22610 | 0.15511 |

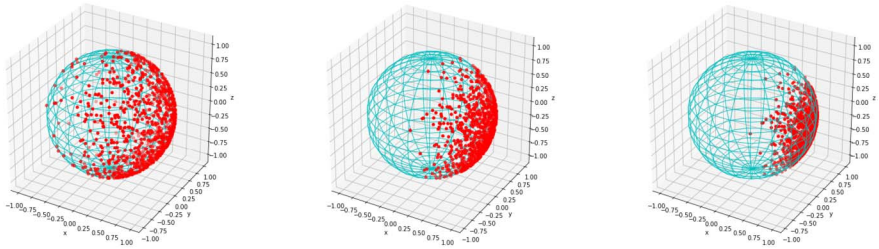
TABLE 14
*Mean square error of $\hat{\kappa}_M$ (top raw) and $\hat{\kappa}_L$ (bottom raw) for* $\mathrm{GvMF}_{3,3}$ *distribution, for different values of $\alpha$ (rows) and $\kappa$ (columns)*

| | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|---|
| 0.5 | 0.06602 | 0.05435 | 0.05186 | 0.04292 | 0.04582 | 0.04249 | 0.05097 | 0.05318 | 0.06447 |
| | 0.06010 | 0.04156 | 0.04885 | 0.05101 | 0.05939 | 0.06571 | 0.10154 | 0.12560 | 0.15621 |
| 1 | 0.14994 | 0.12997 | 0.13578 | 0.14216 | 0.14670 | 0.12929 | 0.14336 | 0.14140 | 0.14276 |
| | 0.64104 | 0.14258 | 0.10734 | 0.10834 | 0.10368 | 0.09266 | 0.11498 | 0.12950 | 0.17055 |
| 1.5 | 0.22813 | 0.22610 | 0.21028 | 0.22102 | 0.19741 | 0.19829 | 0.21212 | 0.23031 | 0.24706 |
| | 1.96361 | 0.54431 | 0.25810 | 0.22148 | 0.18410 | 0.17962 | 0.18814 | 0.21156 | 0.21623 |
| 2 | 0.15945 | 0.29881 | 0.35199 | 0.36035 | 0.31712 | 0.33244 | 0.31035 | 0.30316 | 0.32061 |
| | 2.32459 | 1.70401 | 0.86157 | 0.45843 | 0.35382 | 0.35891 | 0.30941 | 0.30274 | 0.31870 |
| 2.5 | 0.15577 | 0.37208 | 0.55335 | 0.55512 | 0.51438 | 0.51368 | 0.48383 | 0.49249 | 0.42680 |
| | 3.15115 | 2.58508 | 1.76622 | 0.99471 | 0.65380 | 0.58423 | 0.53832 | 0.50869 | 0.43904 |
| 3 | 0.14980 | 0.44026 | 0.67499 | 0.86090 | 0.84880 | 0.85450 | 0.69410 | 0.67099 | 0.66948 |
| | 3.34039 | 3.09945 | 2.79798 | 1.99287 | 1.32447 | 1.07203 | 0.79730 | 0.75204 | 0.69418 |



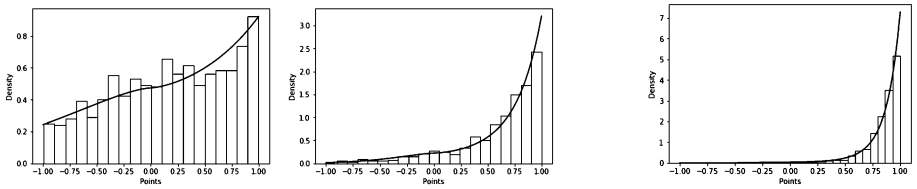

(a) $\kappa = 1$ (b) $\kappa = 4$ (c) $\kappa = 8$

FIG 15. *Density $f_1$ from (35) for $\alpha = 0.5$.*



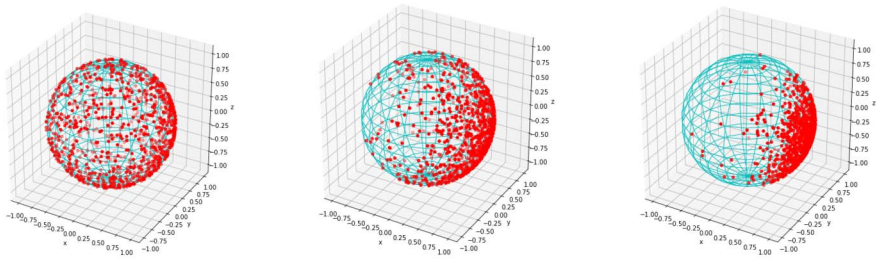
**(a) $\kappa = 1$** **(b) $\kappa = 4$** **(c) $\kappa = 8$**

FIG 16. *Realisations of $X \sim \mathrm{GvMF}_{1,3}(\kappa, \alpha, \boldsymbol{\mu})$ with $\alpha = 0.5$.*



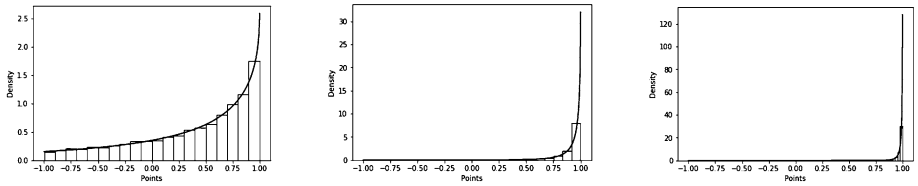

(a) $\kappa = 1$ (b) $\kappa = 4$ (c) $\kappa = 8$

FIG 17. *Density $f_1$ from (35) for $\alpha = 1.5$.*



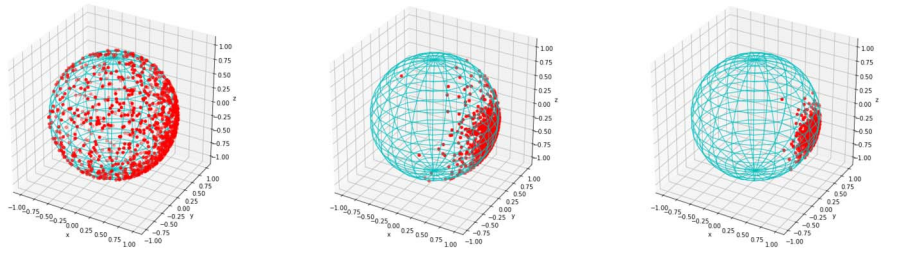
**(a) $\kappa = 1$** **(b) $\kappa = 4$** **(c) $\kappa = 8$**

FIG 18. *Realisations of $X \sim \mathrm{GvMF}_{1,3}(\kappa, \alpha, \boldsymbol{\mu})$ with $\alpha = 1.5$.*



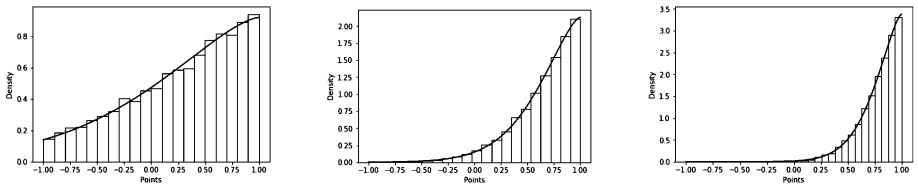

(a) $\kappa = 1$ (b) $\kappa = 4$ (c) $\kappa = 8$

FIG 19. *Density* $f_2$ *from* (36) *for* $\alpha = 0.5$.



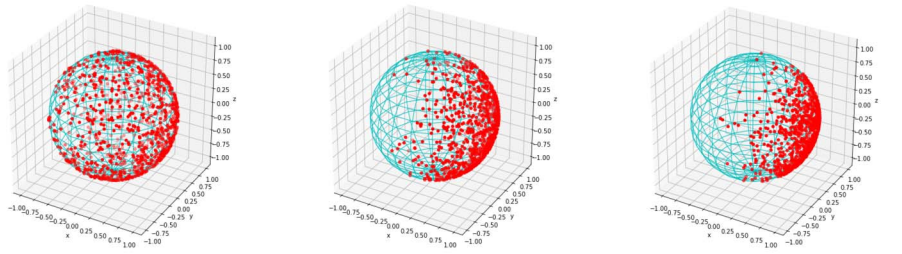
**(a)** $\boldsymbol{\kappa = 1}$ **(b)** $\boldsymbol{\kappa = 4}$ **(c)** $\boldsymbol{\kappa = 8}$

FIG 20. *Realisations of* $X \sim \mathrm{GvMF}_{2,3}(\kappa, \alpha, \boldsymbol{\mu})$ *with* $\alpha = 0.5$.



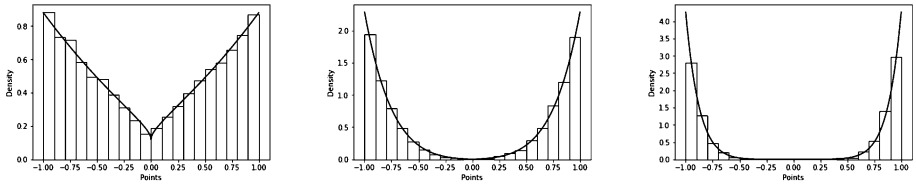

(a) $\kappa = 1$ (b) $\kappa = 4$ (c) $\kappa = 8$

FIG 21. *Density* $f_2$ *from* (36) *for* $\alpha = 1.5$.



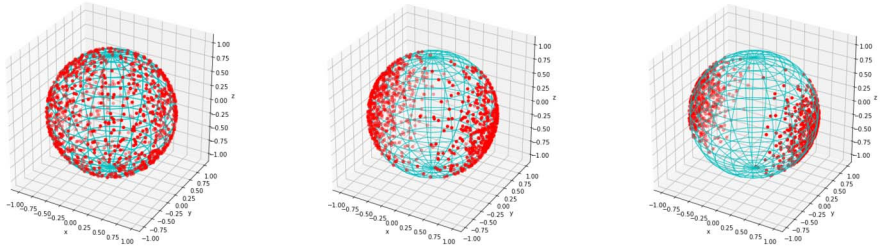
**(a)** $\boldsymbol{\kappa = 1}$ **(b)** $\boldsymbol{\kappa = 4}$ **(c)** $\boldsymbol{\kappa = 8}$

FIG 22. *Realisations of* $X \sim \mathrm{GvMF}_{2,3}(\kappa, \alpha, \boldsymbol{\mu})$ *with* $\alpha = 1.5$.



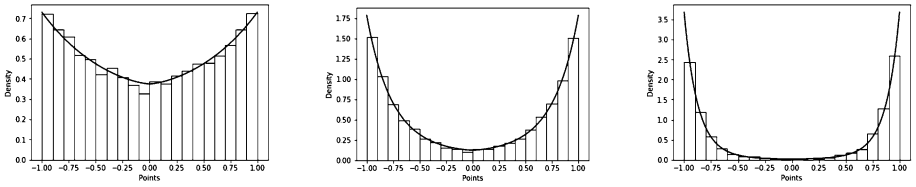

(a) $\kappa = 1$ (b) $\kappa = 4$ (c) $\kappa = 8$

FIG 23. *Density $f_3$ from (37) for $\alpha = 0.5$.*



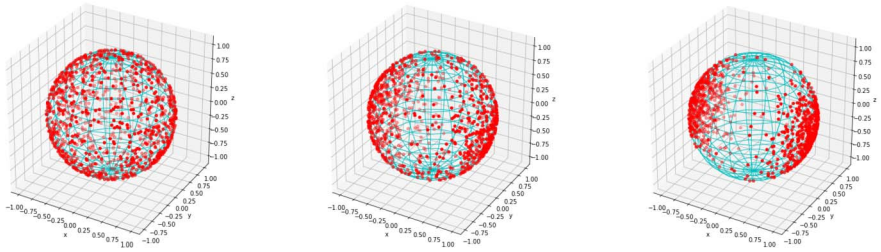
**(a) $\kappa = 1$** **(b) $\kappa = 4$** **(c) $\kappa = 8$**

FIG 24. *Realisations of $X \sim \mathrm{GvMF}_{3,3}(\kappa, \alpha, \boldsymbol{\mu})$ with $\alpha = 0.5$.*




(a) $\kappa = 1$ (b) $\kappa = 4$ (c) $\kappa = 8$

FIG 25. *Density $f_3$ from (37) for $\alpha = 1.5$.*



**(a) $\kappa = 1$** **(b) $\kappa = 4$** **(c) $\kappa = 8$**

FIG 26. *Realisations of $X \sim \mathrm{GvMF}_{3,3}(\kappa, \alpha, \boldsymbol{\mu})$ with $\alpha = 1.5$.*