Available online at www.sciencedirect.com
ScienceDirect
Statistical Methodology 4 (2007) 341–353
Statistical Methodology
Official Journal of the International Indian Statistical Association
www.elsevier.com/locate/stamet

The generalized von Mises distribution

Riccardo Gatto[a,*], Sreenivasa Rao Jammalamadaka[b]

[a] *Institute of Mathematical Statistics and Actuarial Science, University of Berne, Sidlerstrasse 5, 3012 Berne, Switzerland*
[b] *Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106, USA*

Received 4 January 2006; received in revised form 27 June 2006; accepted 11 November 2006

Abstract

A generalization of the von Mises distribution, which is broad enough to cover unimodality as well as multimodality, symmetry as well as asymmetry of circular data, is discussed here. We study this distribution in some detail and discuss its many features, some inferential and computational aspects, and provide some important results including characterization properties for this distribution.

Keywords: Circular distribution; Entropy; Exponential family; Fourier expansion; Maximum likelihood; Normal distribution; Offset distribution; Symmetry and asymmetry; Trigonometric moment; Unimodality and multimodality

1. Introduction and significance of the model

Maksimov [11] discusses a wide class of absolutely continuous circular distributions with continuous densities of the form

$$g(\theta) \propto \exp\left\{\sum_{j=1}^{k} a_j \cos j\theta + b_j \sin j\theta\right\}, \tag{1}$$

for $\theta \in [0, 2\pi)$ and for some constants $a_1, a_2, \ldots, b_1, b_2, \ldots \in \mathrm{R}$. In this paper we will analyze this class paying particular attention to the case where $k = 2$, which leads to an important

* Corresponding author. Tel.: +41 31 631 8807; fax: +41 31 631 3870.
E-mail address: gatto@stat.unibe.ch (R. Gatto).
URL: http://www.stat.unibe.ch/~gatto (R. Gatto).

1572-3127/$ - see front matter
doi:10.1016/j.stamet.2006.11.003

extension of the Circular Normal or von Mises (vM) density, and which we re-express as

$$f(\theta \mid \mu_1, \mu_2, \kappa_1, \kappa_2) = \frac{1}{2\pi G_0(\delta, \kappa_1, \kappa_2)} \exp\{\kappa_1 \cos(\theta - \mu_1) + \kappa_2 \cos 2(\theta - \mu_2)\}, \quad (2)$$

for $\theta \in [0, 2\pi)$, $\mu_1 \in [0, 2\pi)$, $\mu_2 \in [0, \pi)$, $\delta = (\mu_1 - \mu_2) \bmod \pi$, $\kappa_1, \kappa_2 > 0$ and where the normalizing constant is given by

$$G_0(\delta, \kappa_1, \kappa_2) = \frac{1}{2\pi} \int_0^{2\pi} \exp\{\kappa_1 \cos\theta + \kappa_2 \cos 2(\theta + \delta)\} \mathrm{d}\theta. \quad (3)$$

We will call (2) the "Generalized von Mises" (GvM) density and will denote any circular random variable θ with this density by $\theta \sim \mathrm{GvM}(\mu_1, \mu_2, \kappa_1, \kappa_2)$. Besides Maksimov [11], there have been brief mentions of this distribution in Yfantis and Borgman [16], who focus on numerical aspects, and in Kato and Shimizu [9], who consider a particular distribution on the cylinder for which the conditional density of the angular component given the height in the cylinder turns out to have the GvM form (2).

The well-known von Mises density is obtained by taking the sum in the exponent of (1) just for $k = 1$, which yields

$$f(\theta \mid \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\theta - \mu)\}, \quad (4)$$

for $\theta \in [0, 2\pi)$, $\mu \in [0, 2\pi)$, $\kappa > 0$ and where $I_r(z) = (2\pi)^{-1} \int_0^{2\pi} \cos r\theta \exp\{z \cos\theta\} \mathrm{d}\theta$, $z \in \mathrm{C}$, is the modified Bessel function I of integer order r (see e.g. [1]). The GvM allows for greater flexibility in terms of asymmetry and bimodality, compared to the vM distribution, this latter being always circularly symmetric and unimodal with density dropping exponentially on either side from the center.

Another wide and important class of absolutely continuous distributions for circular data is the wrapped α-stable (WαS) class, which derives from the characteristic function of α-stable distributions in the real line. The WαSdensity admits the Fourier series

$$g(\theta) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{j=1}^{\infty} \exp\{-\tau^\alpha j^\alpha\} \cos\left\{ j(\theta - \mu) - \tau^\alpha j^\alpha \beta \tan \frac{\alpha\pi}{2} \right\},$$

for $\theta \in [0, 2\pi)$, $\alpha \in (0, 1) \cup (1, 2]$, $\mu \in [0, 2\pi)$, $\beta \in [-1, 1]$ and $\tau > 0$. These densities are unimodal and have different tail behaviors, according to the value of α, and can be circularly symmetric, left- or right-skewed, according to the value of β. For more details about WαSdistributions and inference, refer to [5,6]. However, in comparison with GvM densities, WαSdensities cannot be bimodal and do not share the theoretical properties given in Section 2.

As previously stated, the GvM densities can be symmetric, asymmetric, unimodal or bimodal. We now give some basic results in relation to the possible shapes of the densities together with some graphical illustrations. Let us first consider the hypothesis $\mathrm{H}_0 : \mu_2 = \mu_1 \bmod \pi$, i.e. $\mathrm{H}_0 : \delta = 0$, with $\kappa_1, \kappa_2 > 0$ implicitly assumed. The density is now circularly symmetric around μ_1, which can be assumed equal to 0 without loss of generality. By differentiation, we obtain that the critical points of the density satisfy

$$\frac{\kappa_1}{4\kappa_2} \sin\theta + \sin\theta \cos\theta = 0.$$

Table 1
Critical points of the GvM density under $H_0 : \mu_2 = \mu_1 \bmod \pi$ and for $\kappa_1 < 4\kappa_2$

Argument values	Type	Density values
$\mu_1 - \pi$	Maximum	$\{2\pi G_0(0, \kappa_1, \kappa_2)\}^{-1} \exp\{-\kappa_1 + \kappa_2\}$
$\mu_1 - \arccos(-\frac{\kappa_1}{4\kappa_2})$	Minimum	$\{2\pi G_0(0, \kappa_1, \kappa_2)\}^{-1} \exp\{-\kappa_2 - \frac{\kappa_1^2}{8\kappa_2}\}$
μ_1	Maximum	$\{2\pi G_0(0, \kappa_1, \kappa_2)\}^{-1} \exp\{\kappa_1 + \kappa_2\}$
$\mu_1 + \arccos(-\frac{\kappa_1}{4\kappa_2})$	Minimum	$\{2\pi G_0(0, \kappa_1, \kappa_2)\}^{-1} \exp\{-\kappa_2 - \frac{\kappa_1^2}{8\kappa_2}\}$

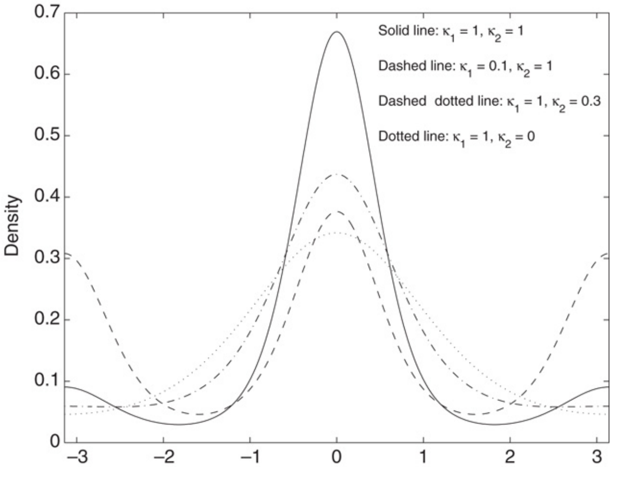

Fig. 1. Some symmetric GvM densities with $\mu_1 = 0$.

If $\kappa_1 < 4\kappa_2$, then there are two trivial and two non-trivial critical points in $[-\pi, \pi)$. All these critical points are displayed in Table 1 for a general μ_1, together with their natures and the corresponding values of the GvM density. (As usually, $\arccos : [-1, 1] \to [0, \pi]$.) If the above inequality is not satisfied, then there remain only the two trivial critical points. In Fig. 1 we can see various symmetric GvM densities for $\mu_1 = 0$. The dotted density with $\kappa_1 = 1$ and $\kappa_2 = 0$ is a vM one and has one maximum at 0 and one minimum at π, and the three other densities are bimodal. The dashed–dotted density with $\kappa_1 = 1$ and $\kappa_2 = 0.3$ resembles a unimodal heavy-tailed one, although there are two minima in the tails (as $1 < 4 \cdot 0.3$).

Consider now a general δ and, again without loss of generality, take $\mu_1 = 0$. It is straightforward to see that the extrema are the solutions in $\theta \in [-\pi, \pi)$ of the equation

$$(1 - 2\sin^2\delta)\sin\theta\cos\theta - 2\sin\delta\cos\delta\sin^2\theta + \rho\sin\theta + \sin\delta\cos\delta = 0, \tag{5}$$

where $\rho = \kappa_1/(4\kappa_2)$. In terms of $x = \sin\theta$, these extrema can be obtained by the solutions in $x \in [-1, 1]$ of the equations

$$\pm\left(1 - 2\sin^2\delta\right)x\sqrt{1 - x^2} - 2\sin\delta\cos\delta x^2 + \rho x + \sin\delta\cos\delta = 0.$$

Alternatively, these extrema can be found by computing the roots in $x \in [-1, 1]$ of the fourth-degree polynomial

$$x^4 - 4\rho\sin\delta\cos\delta x^3 + (\rho^2 - 1)x^2 + 2\rho\sin\delta\cos\delta x + \sin^2\delta\cos^2\delta.$$

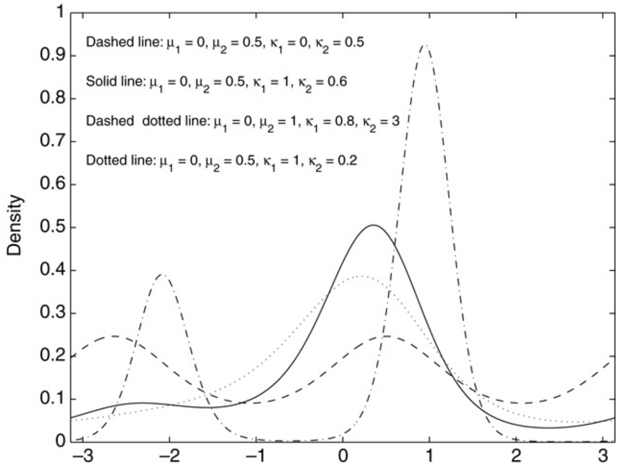

Fig. 2. Some GvM densities.

Consequently the density cannot have more than two modes. This last computation is easy to do with e.g. *Matlab*'s routine `roots`, which re-expresses the problem in the search of the eigenvalues of the companion matrix. Then we transform these roots back to $\theta = \arcsin x, \pi - \arcsin x$ and retain only the values θ which satisfy (5). (As usual, $\arcsin : [-1, 1] \to [-\pi/2, \pi/2]$.) In the general case, we add μ_1 to the results and we obtain extrema in $[\mu_1 - \pi, \mu_1 + \pi)$. If for example $\mu_1 = 0, \delta = 0$, and $\rho < 1$, we find the solutions $\theta = 0, \pi, \pi - \arcsin \pm\sqrt{1-\rho^2}$, which (by noting that $\pm \arccos -\rho = \pi - \arcsin \pm\sqrt{1-\rho^2}$, when $0 < \rho < 1$) satisfy (5), and the solutions $\theta = \arcsin \pm\sqrt{1-\rho^2}$, which do not satisfy (5). The solutions which satisfy (5) are indeed the ones of Table 1. In Fig. 2 we can see some typical asymmetric GvM densities. The dashed density with $\kappa_1 = 0$ is a kind of vM density with double frequency, the solid density has a minor bump in the left tail, the dashed–dotted density has two modes with light tails, and the dotted density is left-skewed and heavy-tailed.

In order to compute a GvM density we must evaluate the constant G_0 $(= G_0^{(2)})$ in (3). This can be done with the expansion

$$G_0(\delta, \kappa_1, \kappa_2) = I_0(\kappa_1) I_0(\kappa_2) + 2 \sum_{j=1}^{\infty} I_{2j}(\kappa_1) I_j(\kappa_2) \cos 2j\delta, \tag{6}$$

where $\delta \in [0, \pi)$ and $\kappa_1, \kappa_2 > 0$. It can be justified by applying twice the Fourier expansion

$$\mathrm{e}^{\kappa \cos\theta} = I_0(\kappa) + 2 \sum_{j=1}^{\infty} I_j(\kappa) \cos j\theta$$

to the integrand of G_0 in (3), thus giving

$$G_0(\delta, \kappa_1, \kappa_2) = I_0(\kappa_1) I_0(\kappa_2) + \frac{2}{\pi} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} I_j(\kappa_1) I_k(\kappa_2) \int_0^{2\pi} \cos j\theta \, \cos 2k(\theta + \delta) \, \mathrm{d}\theta.$$

Then (6) follows directly by noting that

$$\int_0^{2\pi} \cos j\theta \, \cos 2k(\theta + \delta) \, d\theta = \begin{cases} 0, & \text{if } j \neq 2k, \\ \pi \cos 2k\delta, & \text{if } j = 2k. \end{cases}$$

One may check that $G_0(\delta, \kappa_1, 0) = I_0(\kappa_1), \forall \delta \in [0, \pi)$.

In fact, the Maksimov distribution (1) can also be re-expressed as

$$f(\theta \mid \mu_1, \ldots, \mu_k, \kappa_1, \ldots, \kappa_k) = \frac{1}{2\pi G_0^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k)} \exp\left\{\sum_{j=1}^{k} \kappa_j \cos j(\theta - \mu_j)\right\}, \quad (7)$$

where $\kappa_1, \ldots, \kappa_k > 0$, $\mu_1 \in [0, 2\pi)$, $\mu_2 \in [0, \pi), \ldots, \mu_k \in [0, 2\pi/k)$,

$$G_0^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k) = \frac{1}{2\pi} \int_0^{2\pi} \exp\{\kappa_1 \cos\theta + \kappa_2 \cos 2(\theta + \delta_1) + \cdots + \kappa_k \cos k(\theta + \delta_{k-1})\} \mathrm{d}\theta,$$

and where $\delta_1 = (\mu_1 - \mu_2) \bmod \pi, \delta_2 = (\mu_1 - \mu_3) \bmod (2\pi/3), \ldots, \delta_{k-1} = (\mu_1 - \mu_k) \bmod (2\pi/k)$. The density (7) could be called a "Generalized von Mises density of order k" (GvM_k) and a circular random variable θ with this density could be denoted as $\theta \sim \mathrm{GvM}_k(\mu_1, \ldots, \mu_k, \kappa_1, \ldots, \kappa_k)$. However cases of $k > 2$ lead to some practical difficulties, e.g. in the computation of the normalizing constants $G_0^{(k)}$ and in the computation of the estimators. Moreover, the GvM_2 distribution, which we will continue to denote as GvM, is flexible enough for many practical situations. For these reasons, we will mainly focus on the GvM_2 distribution.

In Section 2 we provide some important properties and characterizations of the GvM_k distributions in connection with the exponential family of distributions, with the entropy and, for the case $k = 2$, with the bivariate normal distribution. We also analyze maximum likelihood estimators (MLE) for the parameters paying special attention to the GvM (i.e. GvM_2) model and to its submodels. In Section 3 we provide a numerical illustration with recent data from meteorology in order to show the practical importance and the effectiveness of the GvM model.

2. Important results and characterization properties

In Section 2.1 we emphasize that the GvM_k distribution admits the canonical exponential family form. In Section 2.2 we give an important characterization property relating a GvM_k distribution to a maximum of the entropy, and in Section 2.3 we give another important characterization property relating a GvM_2 distribution to a conditional offset normal distribution. Finally in Section 2.4 we discuss in general the MLE under the canonical exponential family form, and in some greater detail for some submodels.

2.1. Member of the exponential family

We consider the re-parameterization of the GvM_k density (7)

$$\lambda_1 = \kappa_1 \cos\mu_1, \quad \lambda_2 = \kappa_1 \sin\mu_1, \quad \lambda_3 = \kappa_2 \cos 2\mu_2 \quad \text{and} \quad \lambda_4 = \kappa_2 \sin 2\mu_2, \ldots, \lambda_{2k-1} = \kappa_k \cos k\mu_k, \quad \lambda_{2k} = \kappa_k \sin k\mu_k. \quad (8)$$

By expanding the cosines in (7) and by defining $\underline{\lambda} = (\lambda_1, \ldots, \lambda_{2k})^{\mathrm{T}} \in \mathrm{R}^{2k}$ and $T(\theta) = (\cos\theta, \sin\theta, \cos 2\theta, \sin 2\theta, \ldots, \cos k\theta, \sin k\theta)^{\mathrm{T}}$, we can re-express the GvM_k density as

$$f^*(\theta \mid \underline{\lambda}) = \exp\{\underline{\lambda}^{\mathrm{T}} T(\theta) - K(\underline{\lambda})\}. \quad (9)$$

This re-parameterization of the GvM_k density corresponds to the $2k$-parameters canonical exponential family. $T(\theta)$ is a sufficient and complete statistic for $\underline{\lambda}$ and the normalizing constant is

$$K(\underline{\lambda}) = \log(2\pi) + \log G_0^{(k)}(\delta_1, \ldots, \delta_{k-1}, \|\underline{\lambda}^{(1)}\|, \ldots, \|\underline{\lambda}^{(k)}\|),$$

where $\|\cdot\|$ is the Euclidean norm, $\underline{\lambda}^{(1)} = (\lambda_1, \lambda_2)^{\mathrm{T}}, \underline{\lambda}^{(2)} = (\lambda_3, \lambda_4)^{\mathrm{T}}, \ldots, \underline{\lambda}^{(k)} = (\lambda_{2k-1}, \lambda_{2k})^{\mathrm{T}}$, and $\delta_1 = (\arg \underline{\lambda}^{(1)} - \arg \underline{\lambda}^{(2)}/2) \bmod \pi$, $\delta_2 = (\arg \underline{\lambda}^{(1)} - \arg \underline{\lambda}^{(3)}/3) \bmod (2\pi/3), \ldots, \delta_{k-1} = (\arg \underline{\lambda}^{(1)} - \arg \underline{\lambda}^{(k)}/k) \bmod (2\pi/k)$. For the practical case of $k = 2$, this constant can be evaluated with (6).

As pointed out by a referee, it is also directly seen that the original Maksimov distribution with k summands (1) belongs to the $2k$-parameters canonical exponential family: it has the form (9) with

$$\underline{\lambda} = (\lambda_1, \ldots, \lambda_k, \lambda_{k+1}, \ldots, \lambda_{2k}) = (a_1, \ldots, a_k, b_1, \ldots, b_k)$$

and

$$T(\theta) = (\cos\theta, \ldots, \cos k\theta, \sin\theta, \ldots, \sin k\theta).$$

However, the canonical re-parameterization (9) is not as intuitive as e.g. the GvM form (2). For example, the hypothesis $\mathrm{H}_0 : \delta = 0$ can be directly seen under the GvM parameterization as circular symmetry in both frequency components of order 1 and 2.

2.2. *Maximum of entropy*

The concept of entropy arose with the work of Shannon [14] in attempting to create a theoretical model for the transmission of information. The (differential) entropy of a circular distribution with density $g > 0$ over $[0, 2\pi)$ is given by

$$H(g) = -\int_0^{2\pi} \log g(\theta) g(\theta) \mathrm{d}\theta, \tag{10}$$

and it is an appropriate measure of the uncertainty carried by g. For a recent introduction to this concept see e.g. [7]. Distributions which maximize the entropy often have important properties. The exponential maximizes the entropy among distributions with positive domain and a given mean, and it is memoryless. The normal maximizes the entropy among distributions with a given variance, and covariance 0 is equivalent to independence. It is known (see e.g. [6], p. 39) that the vM density (4) maximizes the entropy (10) among all densities g having a fixed first trigonometric moment $\varphi_1 = \gamma_1 + i\sigma_1$, i.e. among all g satisfying

$$\int_0^{2\pi} \mathrm{e}^{\mathrm{i}\theta} g(\theta) \mathrm{d}\theta = \varphi_1 \Leftrightarrow \int_0^{2\pi} \cos\theta g(\theta) \mathrm{d}\theta + \mathrm{i} \int_0^{2\pi} \sin\theta g(\theta) \mathrm{d}\theta = \gamma_1 + \mathrm{i}\sigma_1,$$

for φ_1 fixed, i.e. for γ_1 and σ_1 fixed. Obviously $|\varphi_1| \leq 1$ and $-1 \leq \gamma_1, \sigma_1 \leq 1$. In this case, the parameters μ and κ are the solutions of the equations

$$\cos\mu A(\kappa) = \gamma_1, \quad \text{and} \quad \sin\mu A(\kappa) = \sigma_1,$$

where $A(\kappa) = I_1(\kappa)/I_0(\kappa)$. As other maximum entropy distributions, the vM distribution has several important properties; see e.g. Section 2.3.

If, in addition to this, we fix the second trigonometric moment as well, then we obtain an analogue characterization for the GvM distribution. If, in general, we fix the k first trigonometric moments, then we have a characterization for the GvM_k distribution.

Characterization 1. The circular density $g > 0$ over $[0, 2\pi)$ which maximizes the entropy $H(g)$ subject to

$$\int_0^{2\pi} e^{ir\theta} g(\theta) d\theta = \varphi_r \Leftrightarrow \int_0^{2\pi} \cos r\theta g(\theta) d\theta + i \int_0^{2\pi} \sin r\theta g(\theta) d\theta = \gamma_r + i\sigma_r, \qquad (11)$$

for some determined $\gamma_r, \sigma_r \in [-1, 1]$, $r = 1, \ldots, k$, is the $\text{GvM}_k(\mu_1, \ldots, \mu_k, \kappa_1, \ldots, \kappa_k)$ density (7).

Proof. It follows from Kagan et al. [8, Theorem 13.2.1, p. 409] that the maximum of the entropy subject to the constraints (11) is attained by densities of the form of

$$g(\theta) \propto \exp\{\lambda_1 \cos\theta + \lambda_2 \sin\theta + \lambda_3 \cos 2\theta + \lambda_4 \sin 2\theta + \cdots + \lambda_{2k-1} \cos k\theta + \lambda_{2k} \sin k\theta\},$$

provided that the parameters $\lambda_1, \ldots, \lambda_{2k}$ satisfying (11) exist. But this form can be re-expressed in the GvM_k form (7). Let us define

$$G_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k) = \frac{1}{2\pi} \int_0^{2\pi} \cos r\theta \, \exp\{\kappa_1 \cos\theta + \kappa_2 \cos 2(\theta + \delta_1) + \cdots + \kappa_k \cos k(\theta + \delta_{k-1})\} \, d\theta,$$

$$H_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k) = \frac{1}{2\pi} \int_0^{2\pi} \sin r\theta \, \exp\{\kappa_1 \cos\theta + \kappa_2 \cos 2(\theta + \delta_1) + \cdots + \kappa_k \cos k(\theta + \delta_{k-1})\} \, d\theta,$$

$$A_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k) = \frac{G_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k)}{G_0^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k)} \qquad (12)$$

and

$$B_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k) = \frac{H_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k)}{G_0^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k)} \qquad (13)$$

for $r = 1, \ldots, k$, where $\delta_1 \in [0, \pi)$, $\delta_2 \in [0, 2\pi/3), \ldots, \delta_{k-1} \in [0, 2\pi/k)$, and $\kappa_1, \ldots, \kappa_k > 0$. Then we have that μ_1, $\mu_2 = (\mu_1 - \delta_1) \bmod \pi, \ldots, \mu_k = (\mu_1 - \delta_{k-1}) \bmod (2\pi/k)$ and $\kappa_1, \ldots, \kappa_k$ are the simultaneous solutions of

$$e^{ir\mu_1}\{A_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k) + iB_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k)\} = \varphi_r,$$

for $r = 1, \ldots, k$. Equivalently, they are the simultaneous solutions of

$$\cos r\mu_1 A_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k) - \sin r\mu_1 B_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k) = \gamma_r$$

and

$$\cos r\mu_1 B_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k) + \sin r\mu_1 A_r^{(k)}(\delta_1, \ldots, \delta_{k-1}, \kappa_1, \ldots, \kappa_k) = \sigma_r,$$

for $r = 1, \ldots, k$. •

This characterization is highly relevant in e.g. Bayesian statistics. Sometimes partial prior information is available and it is desired to determine the most noninformative prior distribution which satisfies the known partial prior information. An optimal prior is the solution of the constrained maximization of the entropy (see e.g. [3]). When the k first trigonometric moments represent this partial prior information, and when amongst prior distributions having these k trigonometric moments the most noninformative is sought, then the distribution maximizing the entropy under these trigonometric moment conditions turns out to be the GvM_k one. Other information theoretic properties of the GvM_k distribution can be found in [4].

2.3. Conditional offset distribution

An offset distribution is the marginal distribution of the directional component of a multivariate distribution. Equivalently, it is the radial projection of a p-variate distribution, so it is also called a projected distribution on S^{p-1}. A conditional offset distribution is the conditional distribution of this directional component given a fixed length from the origin. In what follows, we show that the GvM distribution is the conditional offset distribution of the bivariate normal distribution.

Consider first a p-variate normal random vector $\underline{X}$ with expectation $\underline{\nu}$ and covariance matrix $\sigma^2 I$, i.e. $\underline{X} \sim \mathcal{N}(\underline{\nu}, \sigma^2 I)$, I denoting the identity matrix of order p. The density of $\underline{X} \mid \|\underline{X}\| = 1$ is

$$N_p(\underline{\nu}, \sigma) \exp\left\{-\frac{1}{2\sigma^2}(\underline{x} - \underline{\nu})^{\mathrm{T}}(\underline{x} - \underline{\nu})\right\} = C_p(\kappa) \exp\{\kappa \underline{\bar{\nu}}^{\mathrm{T}} \underline{x}\},$$

$\forall \underline{x}$ such that $\|\underline{x}\| = 1$, $\underline{\bar{\nu}} = \underline{\nu}/\|\underline{\nu}\|$, $\kappa = \|\underline{\nu}\|/\sigma^2 > 0$, and for some normalizing constants $N_p(\sigma)$ and $C_p(\kappa) > 0$, depending on σ and κ only, respectively. This directional density takes values on S^{p-1} and was introduced by Langevin [10]. For $p = 2$ and expressed in terms of the angular difference between $\underline{x}$ and $\underline{\nu}$, it is the vM density $C_2(\kappa) \exp\{\kappa \cos(\theta - \mu)\}$, where $\theta = \arg\{\underline{x}\}$, $\mu = \arg\{\underline{\nu}\} \in [0, 2\pi)$ and $C_2(\kappa) = \{2\pi I_0(\kappa)\}^{-1}$. Still for $k = 2$ but now allowing for a general covariance matrix $\Sigma = \operatorname{var}(\underline{X})$, we have the following result.

Characterization 2. If $\underline{X}$ is a bivariate normal vector with expectation $\underline{\nu} = (\nu_1, \nu_2)^{\mathrm{T}}$ and covariance matrix $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, then the density of $\arg\{\underline{X}\} \mid \|\underline{X}\| = 1$ is given by

$$f(\theta \mid \mu_1, \mu_2, \kappa_1, \kappa_2) = \frac{1}{2\pi G_0(\delta, \kappa_1, \kappa_2)} \exp\{\kappa_1 \cos(\theta - \mu_1) + \kappa_2 \cos 2(\theta - \mu_2)\},$$

where $\delta = (\mu_1 - \mu_2) \bmod \pi$, and where $\mu_1 \in [0, 2\pi)$, $\mu_2 \in [0, \pi)$ and $\kappa_1, \kappa_2 > 0$ are the solutions of

$$\kappa_1 \cos \mu_1 = -\frac{1}{1 - \rho^2}\left(\frac{\rho\nu_2}{\sigma_1\sigma_2} - \frac{\nu_1}{\sigma_1^2}\right), \quad \kappa_1 \sin \mu_1 = -\frac{1}{1 - \rho^2}\left(\frac{\rho\nu_1}{\sigma_1\sigma_2} - \frac{\nu_2}{\sigma_2^2}\right),$$
$$\kappa_2 \cos 2\mu_2 = -\frac{1}{4(1 - \rho^2)}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right), \quad \text{and} \quad \kappa_2 \sin 2\mu_2 = \frac{\rho}{2(1 - \rho^2)\sigma_1\sigma_2}.$$

Proof. The logarithm of the density of $\underline{X}$ is

$$\begin{aligned} &c_1 - \frac{1}{2}(\underline{x} - \underline{\nu})^{\mathrm{T}} \Sigma^{-1} (\underline{x} - \underline{\nu}) \\ &\quad = c_1 - \frac{1}{2}\frac{1}{1 - \rho^2}\left\{\frac{(x_1 - \nu_1)^2}{\sigma_1^2} + \frac{(x_2 - \nu_2)^2}{\sigma_2^2} - 2\rho\frac{x_1 - \nu_1}{\sigma_1}\frac{x_2 - \nu_2}{\sigma_2}\right\}, \end{aligned}$$

for some constant c_1 depending on σ_1, σ_2 and ρ only. With the change of variables $r\cos\theta = x_1$ and $r\sin\theta = x_2$ (and by noting that $\cos^2\theta = (1+\cos 2\theta)/2$, $\sin^2\theta = (1-\cos 2\theta)/2$ and $\cos\theta\sin\theta = (\sin 2\theta)/2$), this logarithmic density becomes

$$c_2 - \frac{1}{2}\frac{r}{1-\rho^2}\left\{2\left(\frac{\rho\nu_2}{\sigma_1\sigma_2} - \frac{\nu_1}{\sigma_1^2}\right)\cos\theta + 2\left(\frac{\rho\nu_1}{\sigma_1\sigma_2} - \frac{\nu_2}{\sigma_2^2}\right)\sin\theta \right.$$
$$\left. + \frac{r}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right)\cos 2\theta - r\frac{\rho}{\sigma_1\sigma_2}\sin 2\theta\right\},$$

for some other constant c_2 depending on $\nu_1, \nu_2, \sigma_1, \sigma_2$ and ρ only. This last expression evaluated at $r = 1$, together with the re-parameterization (8), shows that the GvM density $f(\theta \mid \mu_1, \mu_2, \kappa_1, \kappa_2)$ is indeed the density of $\arg\{\underline{X}\} \mid \|\underline{X}\| = 1$. •

Applying this result to $\rho = 0$, $\sigma_1 = \sigma_2$ leads to $\kappa_2 = 0$, $\mu_1 = \arg\{\underline{\nu}\}$ and $\kappa_1 = \|\underline{\nu}\|/\sigma_1^2$, in accordance with the discussion preceding Characterization 2.

Hence, in two dimensions, the conditional offset normal distribution is the GvM distribution, and it has a substantially simpler form than the offset normal, given e.g. in [12, p. 52]. For $p > 2$ dimensions, the offset normal distribution was given in the form of an infinite series by Bingham. It is given e.g. in Watson [15, pp. 226–231]. It is interesting to ask whether there are other distributions on R^2 whose offset or conditional offset distributions are GvM_k distributions with $k > 2$. Clearly, a GvM_k does not come by conditioning a k-dimensional multivariate normal distribution for $\|\underline{X}\| = 1$, since this would lead to a distribution on S^{k-1}.

2.4. Maximum likelihood inference

In this subsection we give several important facts regarding maximum likelihood estimation and inference for the GvM_k model, and with emphasis on the GvM_2 and its submodels.

As seen in Section 2.1, the GvM_k density re-parameterized by (8) takes the canonical exponential family form (9). Under this parameterization, a sample of independent angles $\theta_1, \ldots, \theta_n$ from the GvM_k distribution has the logarithmic likelihood function

$$l(\underline{\lambda} \mid \theta_1, \ldots, \theta_n) = \sum_{i=1}^{n} \log f^*(\theta_i \mid \underline{\lambda}) = \underline{\lambda}^{\mathrm{T}} \sum_{i=1}^{n} T(\theta_i) - nK(\underline{\lambda}).$$

From the classical literature on the exponential family (cf. e.g. [2]) we obtain the following properties:

- $l(\underline{\lambda} \mid \theta_1, \ldots, \theta_n)$ is strictly concave in $\underline{\lambda}$, $\forall \theta_1, \ldots, \theta_n \in [0, 2\pi)$;
- $\partial/(\partial\underline{\lambda})K(\underline{\lambda}) = \mathrm{E}[T(\theta_1)]$;
- $\partial^2/(\partial\underline{\lambda}^{\mathrm{T}}\partial\underline{\lambda})K(\underline{\lambda}) = \mathrm{var}(T(\theta_1))$;
- if the MLE $\hat{\underline{\lambda}}$ of $\underline{\lambda}$ exists then it is unique;
- with probability 1, $\exists n_0$ such that $\forall n \geq n_0$, $\hat{\underline{\lambda}}$ exists;
- the MLE exists iff $n^{-1}\sum_{i=1}^{n} T(\theta_i) \in \mathrm{int\,conv}\{T(\theta) \mid \theta \in [0, 2\pi)\}$, where $\mathrm{conv}\{S\}$ denotes the convex hull of S;
- the MLE exists iff $\mathrm{E}[T(\theta_1)] = n^{-1}\sum_{i=1}^{n} T(\theta_i)$ has a solution in $\underline{\lambda}$, and when there is one it is the unique MLE; and
- $\sqrt{n}(\hat{\underline{\lambda}} - \underline{\lambda}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I^{-1}(\underline{\lambda}))$, where $I(\underline{\lambda}) = \partial^2/(\partial\underline{\lambda}^{\mathrm{T}}\partial\underline{\lambda})K(\underline{\lambda})$ is the Fisher information matrix.

In the above points, all expectations and covariances are taken with respect to the GvM_k distribution.

From transformation invariance, the MLE of $\mu_1, \ldots, \mu_k$ and $\kappa_1, \ldots, \kappa_k$ under the original GvM_k parameterization (7) are the transformation of $\hat{\underline{\lambda}}$, the MLE under the canonical exponential family.

Suppose we want to test the null hypothesis $\mathrm{H}_0 : \underline{\lambda} \in \Lambda_0$ against $\mathrm{H}_1 : \underline{\lambda} \notin \Lambda_0$, where Λ_0 is a subset of R^{2k}. Then, the scaled likelihood ratio test statistic for this problem is

$$Q_n = 2 \left\{ l(\hat{\underline{\lambda}} \mid \theta_1, \ldots, \theta_n) - \sup_{\lambda \in \Lambda_0} l(\underline{\lambda} \mid \theta_1, \ldots, \theta_n) \right\}.$$

When Λ_0 is determined by $q \leq 2k$ restrictions of the type $r_1(\underline{\lambda}) = 0, \ldots, r_q(\underline{\lambda}) = 0$, then large sample testing is based on $Q_n \xrightarrow{\mathcal{D}} \chi_q^2$ (as $n \to \infty$).

Next we study the MLE for some important submodels of the GvM model and use the following notation: $C_{1n} = \sum_{i=1}^n \cos\theta_i$, $S_{1n} = \sum_{i=1}^n \sin\theta_i$, $R_{1n} = (C_{1n}^2 + S_{1n}^2)^{1/2}$, $\bar{\theta}_1 = \arg\{C_{1n}, S_{1n}\}$, $C_{2n} = \sum_{i=1}^n \cos 2\theta_i$, $S_{2n} = \sum_{i=1}^n \sin 2\theta_i$, $R_{2n} = (C_{2n}^2 + S_{2n}^2)^{1/2}$, and $\bar{\theta}_2 = \arg\{C_{2n}, S_{2n}\}$, where $\theta_1, \ldots, \theta_n$ is a sample from a common GvM distribution.

The first submodel we consider is the $\mathrm{GvM}(\mu_1, \mu_2, \kappa_1, \kappa_2)$ under the hypothesis $\mathrm{H}_0 : \kappa_2 = 0$, which is the well-known $\mathrm{vM}(\mu_1, \kappa_1)$ model with $\kappa_1 > 0$ implicitely assumed. In this case the MLE of μ_1 and κ_1 satisfy the two equations

$$\sum_{i=1}^n \sin(\theta_i - \mu_1) = 0 \quad \text{and}$$

$$\sum_{i=1}^n \{\cos(\theta_i - \mu_1) - A(\kappa_1)\} = 0,$$

where $A(\kappa) = I_1(\kappa)/I_0(\kappa)$. The first equation leads to $S_{1n} \cos\mu_1 = C_{1n} \sin\mu_1$, and the only solution with negative second derivative in μ_1 of the log-likelihood is

$$\hat{\mu}_1 = \bar{\theta}_1.$$

With this the second equation yields

$$\hat{\kappa}_1 = A^{(-1)} \left(\frac{R_{1n}}{n} \right);$$

see e.g. [6, pp. 85–88] or Mardia and Jupp [13, pp. 85–86] for more details. Note that because R_{1n} is a measure of concentration and A is a monotone increasing function, $\hat{\kappa}_1$ is indeed a measure of concentration.

The second submodel we consider is the $\mathrm{GvM}(\mu_1, \mu_2, \kappa_1, \kappa_2)$ under the hypothesis $\mathrm{H}_0 : \kappa_1 = 0$. Now we have a bimodal density with two points of symmetry, one at μ_2 and the other at $\mu_2 + \pi$. The MLE has the same form as before, under the vM model, provided that we replace θ_i by $2\theta_i$, $i = 1, \ldots, n$, and μ_1 by $2\mu_2$. That is, the previous estimating equations lead to the MLE of μ_2 and κ_2:

$$\hat{\mu}_2 = \frac{\bar{\theta}_2}{2} \quad \text{and}$$

$$\hat{\kappa}_2 = A^{(-1)} \left(\frac{R_{2n}}{n} \right).$$

Finally we consider the circular symmetric submodel with both frequency components, i.e. under $H_0 : \mu_2 = \mu_1 \mod \pi$, which can be re-expressed as $H_0 : \delta = 0$, with $\delta = (\mu_1 - \mu_2) \mod \pi$, where $\kappa_1, \kappa_2 > 0$ are implicitly meant. As seen in Table 1 in the introduction, this model can have up to four critical points, and exactly four when $\kappa_1 < 4\kappa_2$. An interesting situation arises when κ_1 and κ_2 are known. Then the MLE of μ_1, denoted $\hat{\mu}_1$, satisfies

$$\sum_{i=1}^{n} \kappa_1 \sin(\theta_i - \hat{\mu}_1) + 2\kappa_2 \sin 2(\theta_i - \hat{\mu}_1) = 0.$$

This equation can be compactly re-expressed as

$$\kappa_1 R_{1n} \sin(\bar{\theta}_1 - \hat{\mu}_1) + 2\kappa_2 R_{2n} \sin 2(\bar{\theta}_2 - \hat{\mu}_1) = 0, \tag{14}$$

which has two solutions, or also in terms of $x = \cos\hat{\mu}_1$ and $y = \sin\hat{\mu}_1 \in [-1, 1]$ as

$$4\kappa_2 R_{2n} \sin\bar{\theta}_2 \cos\bar{\theta}_2 + \kappa_1 R_{1n} \sin\bar{\theta}_1 x - \kappa_1 R_{1n} \cos\bar{\theta}_1 y - 8\kappa_2 R_{2n} \sin\bar{\theta}_2 \cos\bar{\theta}_2 y^2 - 4\kappa_2 R_{2n}(1 - 2\sin^2\bar{\theta}_2)xy = 0.$$

From (14) we can see that in general, even for fixed κ_1 and κ_2, there is no explicit solution for the MLE of μ_1, and this MLE is neither $\bar{\theta}_1$ nor $\bar{\theta}_2$.

3. Numerical illustration

The following example shows the effectiveness of the GvM model when applied to some data sets from a real and recent problem. The data sets are taken from ArcticRIMS (A Regional, Integrated Hydrological Monitoring System for the Pan Arctic Land Mass) at the WWW address http://rims.unh.edu. The main goal is to provide a regular monitoring of Pan Arctic water budgets and river discharge to the Arctic Ocean. The geography and dynamics of water across this region are important elements of the larger Earth system. This study is important given the growing evidence of the vulnerability of the Arctic climate and terrestrial biosphere to global change. Wind directions measured at several locations are among the important variables of this project. For our illustration, we selected the wind directions measured on four different sites at continental level: the Pan Arctic, the Europe, the Greenland and the North America basins. At each of these locations, wind directions were measured daily from January to December 2005. We fit the GvM model to these four data sets, of the four locations during 2005, by computing the MLE. This can be efficiently done with the help of *Matlab*'s routine `fminsearch`. We minimize minus the logarithmic likelihood. To avoid boundary problems, we minimize with respect to the logarithms of the concentrations. The default options of `fminsearch` are used and the starting values for all parameters are simply taken equal to zero. The selected data sets and the related *Matlab* programs are available at http://www.stat.unibe.ch/~gatto). The MLE for the four data sets are the following. For the Pan Arctic basins we have

$$\hat{\mu}_1 = 4.5055, \quad \hat{\mu}_2 = 0.9822, \quad \hat{\kappa}_1 = 0.8110 \quad \text{and} \quad \hat{\kappa}_2 = 1.9897.$$

For the Europe basins we have

$$\hat{\mu}_1 = 4.2330, \quad \hat{\mu}_2 = 0.8530, \quad \hat{\kappa}_1 = 0.2781 \quad \text{and} \quad \hat{\kappa}_2 = 1.6028.$$

For the Greenland basins we have

$$\hat{\mu}_1 = 4.7875, \quad \hat{\mu}_2 = 0.9523, \quad \hat{\kappa}_1 = 3.7756 \cdot 10^{-5} \quad \text{and} \quad \hat{\kappa}_2 = 1.2119.$$

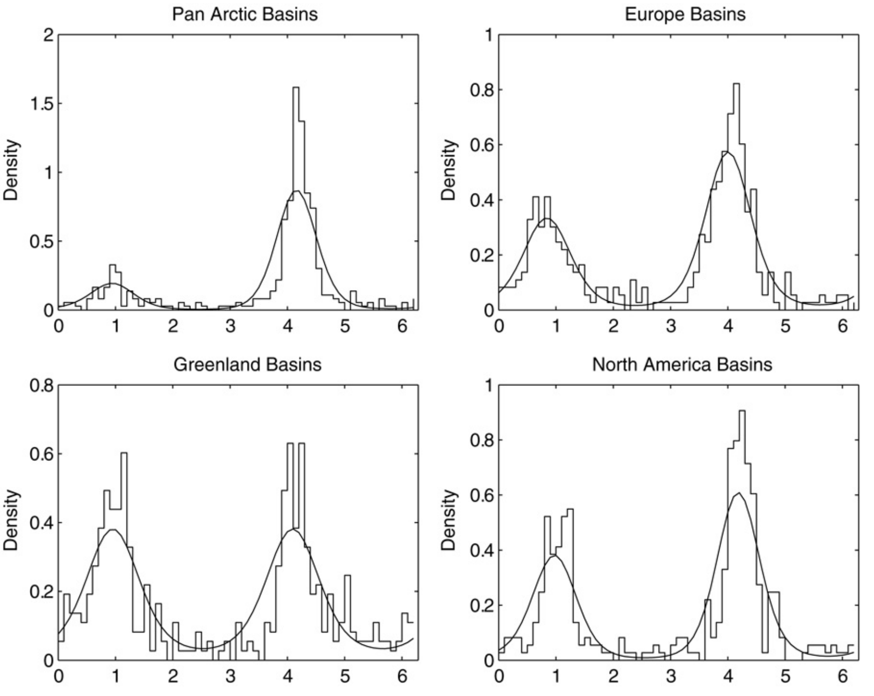

Fig. 3. Daily wind directions in four Arctic basins in the year 2005.

For the North America basins we have

$$\hat{\mu}_1 = 4.9710, \quad \hat{\mu}_2 = 1.0082, \quad \hat{\kappa}_1 = 0.3440 \quad \text{and} \quad \hat{\kappa}_2 = 1.8601.$$

To assess the fit of the GvM model, for each data set we superpose the histogram of the 365 wind directions onto the GvM density with the parameters estimated by MLE. Fig. 3 shows that the GvM model provides very good summaries of these four real data sets. Obviously a simple vM density would not lead to accurate fits. Even if a mixture of two vM may also lead to good fits, such a mixture does not have the nice theoretical properties of the GvM and the estimation is presumably more complicated than with our GvM model. It is known that the likelihood of the mixture of vM(μ_1, κ_1) and vM(μ_2, κ_2) is unbounded: if for example μ_1 is equal to one of the observations and if κ_1 tends to infinity, then the likelihood will tend to infinity.

4. Conclusion

In this article we formally introduce the Generalized von Mises distribution which allows for much greater flexibility than the currently used von Mises or Circular Normal distribution, while retaining several important properties such as belonging to the exponential family, a relationship with the normal distribution, and while having maximum entropy. We study several important features of this distribution and establish some important properties and characterizations. We finally show the effectiveness of this model with recent data from meteorology.

Acknowledgements

The first author's research was supported by the Swiss National Science Foundation. The authors thank the editor and anonymous referees for thoughtful comments and suggestions which improved the quality of this article.

References

[1] M. Abramowitz, I.E. Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, in: Applied Mathematics Series, vol. 55, National Bureau of Standards, 1972.
[2] O. Barndorff-Nielsen, Exponential Families: Exact Theory, Aarhus University Monograph, 1973.
[3] J.O. Berger, Statistical Decision Theory and Bayesian Analysis, Springer-Verlag, 1980.
[4] R. Gatto, Some information theoretic aspects of circular distributions, Technical Report, University of Berne, Switzerland, 2006.
[5] R. Gatto, S.R. Jammalamadaka, Inference for wrapped symmetric α-stable circular models, Sankhyā, A 65 (2) (2003) 333–355.
[6] S.R. Jammalamadaka, A. SenGupta, Topics in Circular Statistics, World Scientific Press, 2001.
[7] O. Johnson, Information Theory and the Central Limit Theorem, Imperial College Press, 2004.
[8] A.M. Kagan, Y.V. Linnik, C.R. Rao, Characterization Problems in Mathematical Statistics, Wiley, 1973 (Translation of Russian original text).
[9] S. Kato, K. Shimizu, Dependent models on two tori, cylinders and unit discs, Technical Report, Keio University, Japan, 2004.
[10] P. Langevin, Magnétisme et théorie des éléctrons, Annales de Chimie et de Physique 5 (1905) 71–127.
[11] V.M. Maksimov, Necessary and sufficient statistics for the family of shifts of probability distributions on continuous bicompact groups, Theoria Verojatna 12 (1967) 307–321 (in Russian).
[12] K.V. Mardia, Statistics of Directional Data, Academic Press, 1972.
[13] K.V. Mardia, P.E. Jupp, Directional Statistics, Wiley, 2000.
[14] C.E. Shannon, The mathematical theory of communication, Bell Systems Technical Journal 27 (1948) 379–423, 623–656.
[15] G.S. Watson, Statistics on Spheres, in: University of Arkansas Lecture Notes in the Mathematical Sciences, vol. 6, 1983.
[16] E.A. Yfantis, L.E. Borgman, An extension of the von Mises distribution, Communications in Statistics, Theory and Methods 11 (1982) 1695–1706.