THE GENERALIZED VON MISES DISTRIBUTION

Riccardo Gatto†
Sreenivasa Rao Jammalamadaka‡

January 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, has shown up in different contexts. We study this distribution in some detail here and discuss its many features, some inferential and computational aspects, and we provide some important results including characterization properties for this distribution.

Key words and phrases

Circular distribution, entropy, exponential family, Fourier expansion, maximum likelihood, normal distribution, offset distribution, symmetry and asymmetry, unimodality and multimodality.

---

† Research partially supported by a grant of the Swiss National Science Foundation.

Address: Institute of Mathematical Statistics and Actuarial Science, University of Bern, Sidlerstrasse 5, 3012 Bern, Switzerland.

Internet: gatto@stat.unibe.ch; http://www.stat.unibe.ch/∼gatto.

‡ Address: Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106, USA.

1 Introduction and significance of the model

Maksimov (1967) introduces a wide class of absolute continuous circular distributions which admits nontrivial sufficient statistics for the parameters and with continuous densities of the form

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

for $\theta \in [0, 2\pi)$ and for some constants $a_1, a_2, \ldots, b_1, b_2, \ldots \in$ R. We consider the important case where $k = 2$ which leads to an important extension of the circular normal or von Mises (vM) density below, which we re-express as

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

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

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

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 (1967), there have been mentions of this distribution in Yfantis and Borgman (1982), who focus on numerical illustrations, and in Kato and Shimizu (2004), 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 interrupting the sum in the exponent of (1) at $k = 1$, which yields

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

for $\theta \in [0, 2\pi)$, $\mu \in [0, 2\pi)$, $\kappa > 0$ and where $I_r(z) = (2\pi)^{-1} \int_0^{2\pi} \cos r\theta \, \exp\{z \cos\theta\} d\theta$, $z \in$ C, is the modified Bessel function I of integer order r (see Abramowitz and Stegun, 1965). 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 a fixed rate of decay from the center.

Another wide and important class of absolute 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 WαS density admits the Fourier series

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

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 WαS distributions and inference, refer to Gatto and Jammalamadaka (2003) and Jammalamadaka and SenGupta (2001). However, in comparison with GvM densities, WαS densities cannot be bimodal and do not share the same theoretical properties (see Section 2).

There are of course many other ways of obtaining asymmetric or bimodal densities, as for example by linear combinations of two vM densities. However, in contrast with such mixture models, both the GvM and the WαS models posses important theoretical properties and lead to simpler inference.

As previously stated, the GvM densities can be symmetric, asymmetric, unimodal or bimodal. We now give some basic results in relation with the possible shapes of the densities together with some graphical illustrations. Let us first consider the hypothesis $\mathrm{H}_0 : \mu_2 = \mu_1 \mathrm{mod}\pi$, 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

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

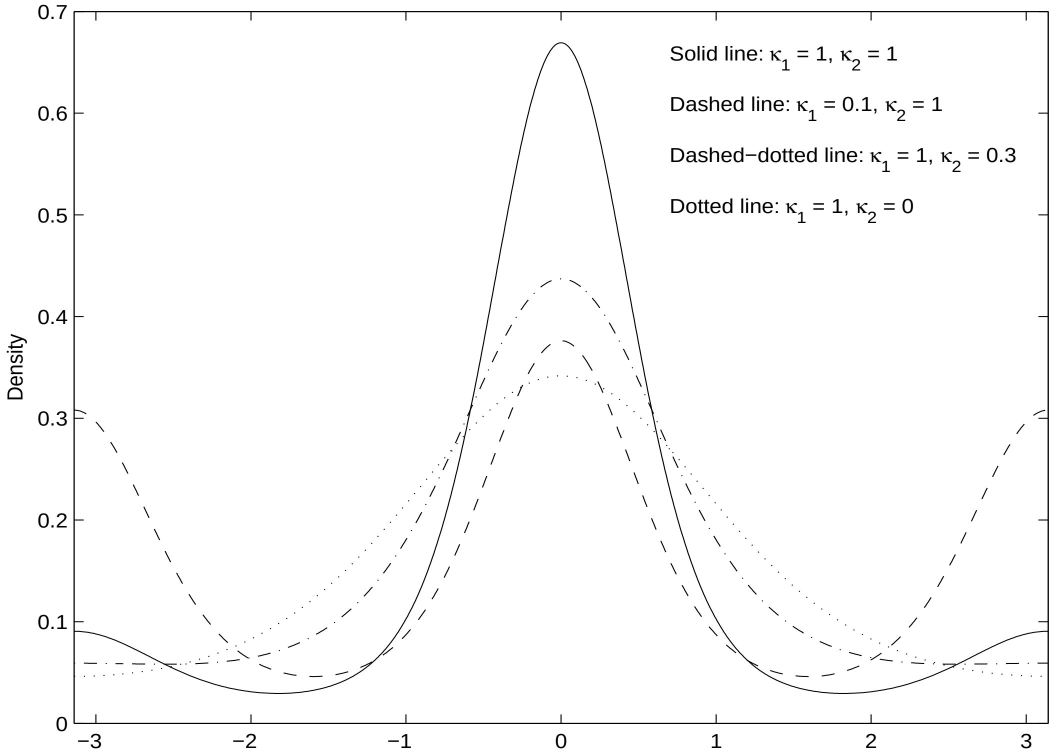
If $\kappa_1 < 4\kappa_2$, then there are two trivial and two non-trivial critical points in $[-\pi, \pi)$. All these critical points are displayed in Table 1 for a general μ_1, together with their natures and the corresponding values of the GvM density. (As usually, $\arccos : [-1, 1] \to [0, \pi]$.) If the above inequality is not satisfied, then there remain only the two trivial critical points. In Figure 1 we can see various GvM densities for $\mu_1 = 0$ and

Table 1: *critical points of the GvM density under $H_0 : \mu_2 = \mu_1 \mathrm{mod}\pi$ 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	$\{2\pi G_0(0, \kappa_1, \kappa_2)\}^{-1} \exp\{-\kappa_2 - \frac{\kappa_1^2}{8\kappa_2}\}$

under $\mathrm{H}_0 : \delta = 0$. 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 to a unimodal heavy-tailed one, although there are two minimums in the tails (as $1 < 4 \cdot 0.3$).

Figure 1: *some GvM densities with $\mu_1 = 0$ under $H_0 : \mu_2 = \mu_1 \mathrm{mod} \pi$*



Let us now consider a general δ and, again without loss of generality, $\mu_1 = 0$. It is direct to see that the extrema are the solutions in θ of the equation

$$(1 - 2\sin^2\delta)\sin\theta\cos\theta - 2\sin\delta\cos\delta\sin^2\theta + \frac{\kappa_1}{4\kappa_2}\sin\theta + \sin\delta\cos\delta = 0. \quad (5)$$

In terms of $x = \sin\theta$, these extrema are the solutions in $x \in [-1, 1]$ of the equation

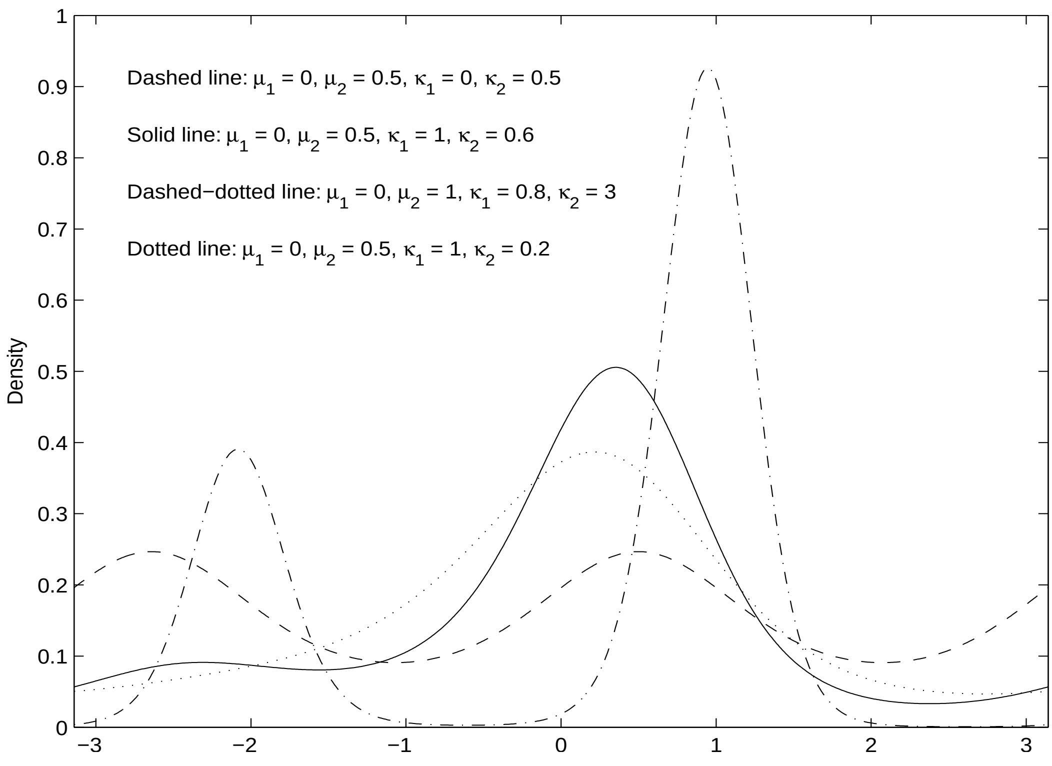
$$\pm\left(1 - 2\,\sin^2\delta\right)x\sqrt{1 - x^2} - 2\,\sin\delta\cos\delta x^2 + \frac{\kappa_1}{4\kappa_2}x + \sin\delta\cos\delta = 0.$$

Alternatively, these extrema can be found by computing the roots of the fourth degree polynomial

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

where $\rho = \kappa_1/(4\kappa_2)$. These roots can be found numerically. Then we transform these roots back to $\theta = \arcsin x$, $\pi - \arcsin x$ and retain only the values θ which satisfy (5). (As usually, $\arcsin : [-1, 1] \to [-\pi/2, \pi/2]$.) 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 $\theta = 0$, π, $\pi - \arcsin \pm\sqrt{1-\rho^2}$, 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 Figure 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.

Figure 2: *some asymmetric GvM densities*



In order to compute a GvM density we must evaluate the constant G_0 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

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

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

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

Then (6) follows directly by noting that

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

In Section 2 we provide some important properties and characterizations in connection with the exponential family of distributions, with the entropy and with the bivariate normal distribution in the plane. Thenstudy maximum likelihood estimators (MLE) for the parameters of the GvM model and of some submodels.

2 Important results and characterization properties

In Subsection 2.1 we give the most central property of the GvM distribution, namely that it admits the canonical exponential family form. In Subsection 2.2 we give an important characterization property relating the GvM distribution to a maximum of the entropy, and in Subsection 2.3 we give another important characterization property relating the GvM distribution to a conditional offset normal distribution. Finally in Subsection 2.4 we discuss in some detail the MLE under the canonical exponential family form, its computation and the likelihood ratio test. We also discuss the MLE for some GvM submodels.

2.1 Member of the exponential family

We consider the re-parameterization

$$\lambda_1=\kappa_1\cos\mu_1,\ \lambda_2=\kappa_1\sin\mu_1,\ \lambda_3=\kappa_2\cos 2\mu_2 \text{ and } \lambda_4=\kappa_2\sin 2\mu_2. \tag{7}$$

By expanding both cosines in (2) and by defining $\underline{\lambda} = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)^{\mathrm{T}} \in \mathrm{R}^4$ and $t(\theta) = (\cos\theta, \sin\theta, \cos 2\theta, \sin 2\theta)^{\mathrm{T}}$, we can re-express the GvM density as

$$f^*(\theta \mid \underline{\lambda}) = \exp\{\underline{\lambda}^{\mathrm{T}} t(\theta) - k(\underline{\lambda})\}. \tag{8}$$

This re-parameterization of the GvM density belongs to the 4 parameters canonical exponential family. The normalizing constant is

$$k(\underline{\lambda}) = \log(2\pi) + \log G_0(\delta, ||\underline{\lambda}^{(1)}||, ||\underline{\lambda}^{(2)}||),$$

where $\|\cdot\|$ is the Euclidean norm, $\underline{\lambda}^{(1)} = (\lambda_1, \lambda_2)^{\mathrm{T}}$, $\underline{\lambda}^{(2)} = (\lambda_3, \lambda_4)^{\mathrm{T}}$ and $\delta = (\arg\underline{\lambda}^{(1)} - \arg\underline{\lambda}^{(2)}/2) \bmod \pi$. This constant can be evaluated with (6). Note that the canonical reparameterization (8) is not as intuitive as the original form (2). For example, $\mathrm{H}_0 : \delta = 0$ can be directly seen under the original parameterization as symmetry in both frequency components.

2.2 Maximum of entropy

The concept of entropy arose with the work of Shannon (1948) in attempting to create a theoretical model for the transmission of information. The entropy of a circular distribution with density $f > 0$ over $[0, 2\pi)$ is given by

$$H(f) = -\int_0^{2\pi} \log f(\theta) f(\theta) d\theta, \tag{9}$$

and it is an appropriate measure of the uncertainty carried by f. It is known (see e.g. Jammalamadaka and SenGupta, 2001, p. 39) that the vM density (4) maximizes the entropy (9) among all densities f having a fixed first trigonometric moment $\varphi_1 = \gamma_1 + i\sigma_1$, i.e. among all f satisfying

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

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 $A(\kappa) = I_1(\kappa)/I_0(\kappa)$. If, in addition to this, we fix the second trigonometric moment as well, then we obtain an analogue characterization for the GvM$(\mu_1, \mu_2, \kappa_1, \kappa_2)$

density.

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

$$\int_0^{2\pi} e^{ik\theta} f(\theta) d\theta = \varphi_k \Leftrightarrow \int_0^{2\pi} \cos k\theta f(\theta) d\theta + i \int_0^{2\pi} \sin k\theta f(\theta) d\theta = \gamma_k + i\sigma_k, \quad (10)$$

for $k = 1, 2$ and for some determined $\gamma_1, \sigma_1, \gamma_2, \sigma_2 \in [-1, 1]$, is the GvM$(\mu_1, \mu_2, \kappa_1, \kappa_2)$ density

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

where $\delta = (\mu_1 - \mu_2) \bmod \pi$.

Proof It follows from Kagan et al. (1973, Theorem 13.2.1, p. 409) that the maximum of the entropy subject to the constraints (10) is attained by densities of the form of

$$c \exp\{\lambda_1 \cos\theta + \lambda_2 \sin\theta + \lambda_3 \cos 2\theta + \lambda_4 \sin 2\theta\},$$

for some constant $c > 0$, provided that the parameters $\lambda_1, \lambda_2, \lambda_3$ and λ_4 satisfying (10) exist. But this comes indeed from GvM densities after the reparameterization (7) of Section 2.1. Let us define

$$H_r(\delta, \kappa_1, \kappa_2) = \frac{1}{2\pi} \int_0^{2\pi} \sin r\theta \, \exp\{\kappa_1 \cos\theta + \kappa_2 \cos 2(\theta + \delta)\} \, d\theta,$$

$$A_k(\delta, \kappa_1, \kappa_2) = \frac{G_k(\delta, \kappa_1, \kappa_2)}{G_0(\delta, \kappa_1, \kappa_2)} \quad (11)$$

and

$$B_k(\delta, \kappa_1, \kappa_2) = \frac{H_k(\delta, \kappa_1, \kappa_2)}{G_0(\delta, \kappa_1, \kappa_2)}, \quad (12)$$

for $k = 1, 2$, where $\delta \in [0, \pi)$ and $\kappa_1, \kappa_2 > 0$. Then we can see that μ_1, $\mu_2 = (\mu_1 - \delta) \bmod \pi$, κ_1 and κ_2 are the simultaneous solutions of

$$e^{ik\mu_1}\{A_k(\delta, \kappa_1, \kappa_2) + iB_k(\delta, \kappa_1, \kappa_2)\} = \varphi_k,$$

for $k = 1, 2$, or, equivalently, that they are the simultaneous solutions of

$$\cos k\mu_1 A_k(\delta, \kappa_1, \kappa_2) - \sin k\mu_1 B_k(\delta, \kappa_1, \kappa_2) = \gamma_k$$

and

$$\cos k\mu_1 B_k(\delta, \kappa_1, \kappa_2) + \sin k\mu_1 A_k(\delta, \kappa_1, \kappa_2) = \sigma_k,$$

for $k = 1, 2$. •

This characterization is highly relevant in Bayesian statistics, for example. Often in Bayesian statistics partial prior information is available and it is desired to determine a prior distribution which is as noninformative as possible, while satisfying the prior information. When the two first trigonometric moments represent this partial prior information and are hence specified, and when amongst prior distributions with these two trigonometric moments the most noninformative is sought, then the distribution maximizing the entropy under these two trigonometric moment conditions provides an optimal solution, and this solution turns out to be the GvM$(\mu_1, \mu_2, \kappa_1, \kappa_2)$ distribution.

2.3 Conditional offset distribution

An offset distribution is the marginal distribution of the directional component of a bivariate distribution on the plane. A conditional offset distribution is the conditional distribution of this same 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 a k-dimensional multinormal 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 the identity matrix of order k. The density of $\underline{X} \mid \|\underline{X}\| = 1$ is

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

$\forall \underline{x}$ such that $\|\underline{x}\| = 1$, μ such that $\|\underline{\nu}\| = 1$, $\kappa = \sigma^{-2} > 0$, and for some for some normalizing constants $N_k(\sigma)$ and $C_k(\kappa) > 0$ depending on σ and κ only, respectively. This directional density takes values on the surface of the unit sphere of R^k and it was introduced by Langevin (1905). For $k = 2$ and expressed in terms of the angle θ in between $\underline{x}$ and $\underline{\nu}$, this gives the vM density $C_2(\kappa) \exp\{\kappa \cos(\theta - \mu)\}$, where $\mu = \arg\{\underline{\nu}\} \in [0, 2\pi)$ and $C_2^{-1}(\kappa) = 2\pi I_0(\kappa)$. Still for $k = 2$ but now for a general covariance matrix Σ (not necessarily $\sigma^2 I$), 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 $\delta = (\mu_1 - \mu_2) \mathrm{mod} \pi$, and where $\mu_1 \in [0, 2\pi)$, $\mu_2 \in [0, \pi)$ and $\kappa_1, \kappa_2 > 0$ are the solutions of

$$\kappa_1 \cos \mu_1 = 2\left(\frac{\rho\nu_2}{\sigma_1\sigma_2} - \frac{\nu_1}{\sigma_1^2}\right), \quad \kappa_1 \sin \mu_1 = 2\left(\frac{\rho\nu_1}{\sigma_1\sigma_2} - \frac{\nu_2}{\sigma_2^2}\right),$$

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

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

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

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 $\nu_1, \nu_2, \sigma_1, \sigma_2, \rho$ and r only. X This last expression evaluated at $r = 1$, together with same the reparameterization as in (7), shows that the GvM density $f(\theta \mid \mu_1, \mu_2, \kappa_1, \kappa_2)$ is indeed the density of $\arg\{\underline{X}\} \mid \|\underline{X}\| = 1$. •

2.4 Maximum likelihood inference

In this paragraph we give several important facts regarding maximum likelihood estimation and inference for the GvM model and for some important submodels.

As seen in Section 2.1, the GvM density reparameterized by (7) takes the canonical exponential family form (8). Hence, given a sample of independent angles $\theta_1, \ldots, \theta_n$ from the GvM$(\mu_1, \mu_2, \kappa_1, \kappa_2)$ distribution, the logarithmic likelihood function is

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

From the classical literature on the exponential family (cf. e.g. Barndorff-Nielsen, 1973) 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 one, $\exists\, n_0$ 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 $\mathrm{E}[t(\theta_1)] = n^{-1}\sum_{i=1}^{n} t(\theta_i)$ have a solution in λ; 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 density (2).

From transformation invariance, the MLE of μ_1, μ_2, κ_1 and κ_2, i.e. under the original parameterization, is simply the transformation the MLE $\hat{\underline{\lambda}}$.

The derivatives of order one and two $k(\underline{\lambda})$ can be computed with two following steps. We can first compute the derivatives of the generic function

$$h\left(\arctan\frac{\lambda_2}{\lambda_1} - \arctan\frac{1}{2}\frac{\lambda_4}{\lambda_3}, \sqrt{\lambda_1^2 + \lambda_2^2}, \sqrt{\lambda_3^2 + \lambda_4^2}\right),$$

and, second, we can introduce the partial derivatives of order one and two of $h(x, y, z) = \log G_0(x, y, z)$, which can be obtained from the Fourier series (6). Both steps can be easily done with a software for symbolic computations (e.g. with *Maple* and the command `diff`).

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^4. The scaled likelihood ratio test statistic for this problem is

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

When Λ_0 is determined by $q \leq 4$ restrictions of the type $r_1(\underline{\lambda}) = 0, \ldots, r_q(\underline{\lambda}) = 0$, then large sample testing is based on $Q_n \xrightarrow{\mathcal{D}} \chi_q^2$.

For the next paragraphs we study the MLE for some important submodels of the GvM model. We 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 GvM$(\mu_1, \kappa_1, \mu_2, \kappa_2)$ with the restriction or hypothesis $\mathrm{H}_0 : \kappa_2 = 0$, which is the well known vM(μ_1, κ_1) model. In this case the MLE of μ_1 and κ_1 satisfy the two equations

$$\begin{aligned}\sum_{i=1}^{n} \sin(\theta_i - \mu_1) &= 0 \text{ and} \\ \sum_{i=1}^{n} \{\cos(\theta_i - \mu_1) - A(\kappa_1)\} &= 0,\end{aligned}$$

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 with 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. Jammalamadaka and SenGupta (2001, p. 85-88) or Mardia and Jupp (2000, p.85-86) for detailed information. Note that because R_{1n} is a measure of concentration and A is a monotone increasing function, $\hat{\kappa}_1$ is indeed an estimator of concentration.

The second submodel we consider is the 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

$$\begin{aligned}\hat{\mu}_2 &= \frac{\bar{\theta}_2}{2} \text{ and} \\ \hat{\kappa}_2 &= A^{(-1)} \left(\frac{R_{2n}}{n} \right).\end{aligned}$$

The last submodel we consider here is the symmetric one with both frequency components, more precisely the one under $\mathrm{H}_0 : \mu_2 = \mu_1 \mathrm{mod} \pi$, which can be re-expressed as $\mathrm{H}_0 : \delta = 0$, with $\delta = (\mu_1 - \mu_2) \mathrm{mod} \pi$, 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

$$\kappa_1 R_{1n} \sin(\bar{\theta}_1 - \hat{\mu}_1) + 2\kappa_2 R_{2n} \sin 2(\bar{\theta}_2 - \hat{\mu}_1) = 0, \tag{13}$$

or also with the second order polynomial in $x = \cos\hat{\mu}_1$ and $y = \sin\hat{\mu}_1 \in [-1, 1]$

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

From (13) 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 Conclusion

In this article we studied some important features of the GvM distribution and we proved some important properties or characterizations. This distribution allows for greater flexibility than the currently used vM (i.e. circular normal) distribution, while maintaining some central properties such as the belonging to the exponential family the connections with the normal distribution and with the entropy. With the GvM distribution symbolic computation is a helpful tool for likelihood inference, where the amount of algebraic manipulations, although simple in nature, goes beyond the limits of the investigator.

References

Barndorff-Nielsen, O. (1973), *Exponential Families: Exact Theory*, Aarhus University Monograph.

Abramowitz, M., Stegun, I. E. (1972), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards, Applied Mathematics Series, 55.

Gatto, R., Jammalamadaka, S. R. (2003), "Inference for wrapped symmetric α-stable circular models", *Sankhyā, A*, 65, 2, 333-355.

Jammalamadaka, S. R., SenGupta, A. (2001), *Topics in Circular Statistics*, World Scientific Press.

Kagan, A. M., Linnik, Y. V., Rao, C. R. (1973), *Characterization Problems in Mathematical Statistics*, Wiley (translation of Russian original text).

Kato, S., Shimizu, K. (2004), "Dependent models on two tori, cylinders and unit discs", technical report, Keio University, Japan.

Langevin, P. (1905), "Magnétisme et théorie des éléctrons", *Annales de Chimie et de Physique*, 5, 71-127.

Maksimov, V. M. (1967), "Necessary and sufficient statistics for the family of shifts of probability distributions on continuous bicompact groups", *Theoria Verojatna*, 12, 307-321 (in Russian).

Shannon, C. E. (1948), "The mathematical theory of communication", *Bell Systems Technical Journal*, 27, 379-423, 623-656.

Yfantis, E. A., Borgman, L. E. (1982), "An extension of the von Mises distribution", *Communications in Statistics, Theory and Methods*, 11, 1695-1706.