Biometrika (2015), **102**, 1, *pp.* 181–190
Printed in Great Britain
doi: 10.1093/biomet/asu059
Advance Access publication 7 December 2014

A tractable and interpretable four-parameter family of unimodal distributions on the circle

By SHOGO KATO
Institute of Statistical Mathematics, Tachikawa, Tokyo 190–8562, Japan
skato@ism.ac.jp

AND M. C. JONES
Department of Mathematics and Statistics, The Open University, Milton Keynes MK7 6AA, U.K.
m.c.jones@open.ac.uk

Summary

This article presents a class of four-parameter distributions for circular data that are unimodal, possess simple characteristic and density functions and a tractable distribution function, can be interpretably parameterized directly in terms of their trigonometric moments, afford a very wide range of skewness and kurtosis, envelop numerous interesting submodels including the wrapped Cauchy and cardioid distributions, allow straightforward parameter estimation by both method of moments and maximum likelihood, and are closed under convolution. This class of distributions exhibits the widest range of attractive properties yet available while retaining unimodality.

Some key words: Directional data; Skewness; Trigonometric moment; Unimodality; Wrapped Cauchy distribution.

1. Introduction

On the circle, as on the line, families of unimodal distributions with parameters controlling location, scale or concentration, skewness and, in some appropriate sense, kurtosis are very useful for robust modelling. We stress unimodality because bi- and multi-modal distributions can most interpretably be modelled by mixtures of unimodal components or perhaps, on the circle, by multiplicative mixtures of, for example, generalized von Mises form (Yfantis & Borgman, 1982; Gatto & Jammalamadaka, 2007; Gatto, 2009). Although numerous such families now exist on the line, fewer exist on the circle; other natural proposals and/or ones with interesting properties have been made, mostly at the expense of maintaining unimodality for every parameter combination. We will show a number of ways in which our new proposal compares favourably with what we perceive to be the best of the current unimodal families on the circle: the wrapped stable distribution (Pewsey, 2008) and distributions generated by a skewing device (Batschelet, 1981), both in direct (Abe et al., 2013) and inverse (Jones & Pewsey, 2012) forms, applied to a broad symmetric family such as that of Jones & Pewsey (2005).

As well as unimodality, our four-parameter family of circular distributions has the following benefits: (i) simple trigonometric moments, equivalently a simple characteristic function; (ii) tractable density and distribution functions; (iii) from (i) and (ii), straightforward parameter estimation by both method of moments and maximum likelihood; (iv) interpretable parameters, as the distribution can be directly parameterized in terms of parameters individually measuring location, concentration, skewness and kurtosis; (v) a very wide range of skewness and kurtosis;

(vi) a number of attractive submodels, including wrapped Cauchy and cardioid distributions; and (vii) closure under convolution and multiplication by certain constants. Disadvantages include the lack of a natural probabilistic construction for the new model and, to some, its close relationship with the wrapped Cauchy rather than the von Mises distribution.

Proofs and further details can be found in the Supplementary Material.

2. BASIC FORMULATION AND PARAMETERIZATIONS

The wrapped Cauchy distribution (Mardia & Jupp, 1999, pp. 51–2; Jammalamadaka & SenGupta, 2001, § 2.2.7) is an important two-parameter symmetric unimodal distribution on the circle with simple trigonometric moments and density function. Let Θ_c follow the wrapped Cauchy distribution with location parameter $-\pi \leqslant \mu < \pi$, concentration parameter $0 \leqslant \rho < 1$ and trigonometric moments $\phi_{\Theta_c}(p) = \left(\rho e^{i\mu}\right)^p$ $(p = 1, 2, \ldots)$. Trigonometric moments for $p \leqslant 0$ can be obtained from $\phi(0) = 1$ and $\phi(-p) = \overline{\phi(p)}$, where $\bar{z}$ is the complex conjugate of z. The probability density function of Θ_c is

$$g_{\mathrm{WC}}(\theta) = \frac{1}{2\pi} \frac{1-\rho^2}{1+\rho^2 - 2\rho\cos(\theta-\mu)}, \quad -\pi \leqslant \theta < \pi. \tag{1}$$

Along with the book treatments mentioned above, see Kent & Tyler (1988) and McCullagh (1996) for further insights into the wrapped Cauchy distribution.

We propose to consider the distributions arising by extending the trigonometric moments of the wrapped Cauchy distribution by multiplying them by a complex constant: $\phi_\Theta(p) = \beta e^{i\alpha}\left(\rho e^{i\eta}\right)^p$ $(p = 1, 2, \ldots)$ where $-\pi \leqslant \eta, \alpha < \pi$, $0 \leqslant \rho < 1$ and $\beta \geqslant 0$ satisfy conditions discussed below. These trigonometric moments reduce to those of the wrapped Cauchy distribution when $\alpha = 0$ and $\beta = 1$. For convenience, we reparameterize to $\mu = \eta + \alpha$, $\lambda = -\alpha$ and $\gamma = \rho\beta$. This gives

$$\phi_\Theta(p) = \gamma\left(\rho e^{i\lambda}\right)^{-1}\left\{\rho e^{i(\mu+\lambda)}\right\}^p, \quad p = 1, 2, \ldots. \tag{2}$$

THEOREM 1. *There exists an absolutely continuous distribution on the circle whose trigonometric moments are given by* (2) *if and only if* $-\pi \leqslant \mu < \pi$, $0 \leqslant \gamma < 1$, *and* $0 \leqslant \rho < 1$ *and* $-\pi \leqslant \lambda < \pi$ *satisfy* $(\rho\cos\lambda - \gamma)^2 + (\rho\sin\lambda)^2 \leqslant (1-\gamma)^2$. *Its probability density function is almost everywhere equal to*

$$g(\theta) = \frac{1}{2\pi}\left\{1 + 2\gamma\frac{\cos(\theta-\mu) - \rho\cos\lambda}{1+\rho^2 - 2\rho\cos(\theta-\mu-\lambda)}\right\}, \quad -\pi \leqslant \theta < \pi. \tag{3}$$

Remark 1. The conditions for the parameters can also be expressed as $-\pi \leqslant \mu < \pi$, $0 \leqslant \rho < 1$, $0 \leqslant \gamma \leqslant (1+\rho)/2$, and $-\pi \leqslant \lambda < \pi$ satisfies $\rho\gamma\cos\lambda \geqslant (\rho^2 + 2\gamma - 1)/2$.

Write $\Theta \sim \mathrm{G}(\mu, \gamma, \rho, \lambda)$ if a circular random variable follows density (3). The distribution function associated with this density is also explicitly available.

Standard trigonometric moment-based measures of circular location, concentration, skewness and kurtosis can be read off from (2). They have a very simple form.

LEMMA 1. *Suppose* $\Theta \sim \mathrm{G}(\mu, \gamma, \rho, \lambda)$. *Then the following hold for the distribution of* Θ*:* (i) *the mean direction is* $\mu_1 \equiv \arg\{\phi_\Theta(1)\} = \mu$; (ii) *the mean resultant length is* $\gamma_1 \equiv |\phi_\Theta(1)| = \gamma$; (iii) *the circular kurtosis of Batschelet* (1981, § 2.6) *is* $\bar{\alpha}_2 \equiv \mathrm{Re}\{\phi_\Theta(2)\, e^{-2i\mu_1}\} = \gamma\rho\cos\lambda$; *and* (iv) *the circular skewness of Batschelet* (1981, § 2.6) *is* $\bar{\beta}_2 \equiv \mathrm{Im}\{\phi_\Theta(2)\, e^{-2i\mu_1}\} = \gamma\rho\sin\lambda$.

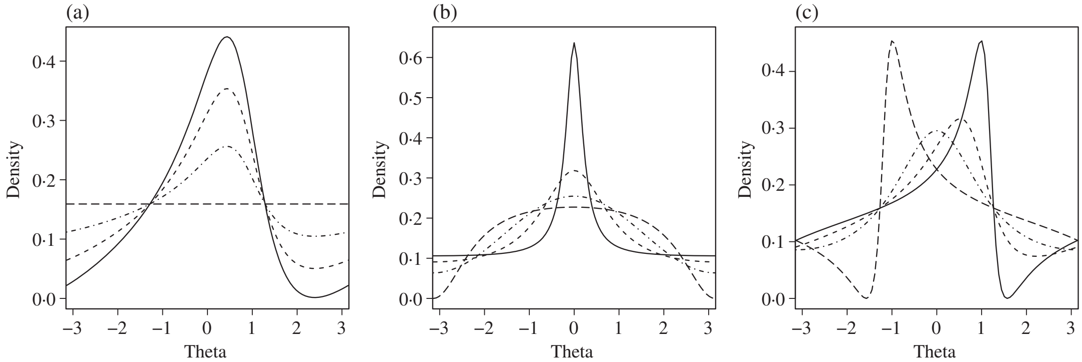

Fig. 1. Density (4) for: (a) $\mu = 0$, $\bar{\alpha}_2 = 0{\cdot}4\gamma\cos(\pi/4)$, $\bar{\beta}_2 = 0{\cdot}4\gamma\sin(\pi/4)$, and $\gamma = 0$ (long-dashed), $\gamma = 0{\cdot}2$ (dot-dashed), $\gamma = 0{\cdot}4$ (dashed), and $\gamma = 0{\cdot}58$ (solid); (b) $\mu = 0$, $\gamma = 0{\cdot}3$, $\bar{\beta}_2 = 0$ and $\bar{\alpha}_2 = -0{\cdot}12$ (long-dashed), $\bar{\alpha}_2 = 0$ (a cardioid density; dot-dashed), $\bar{\alpha}_2 = 0{\cdot}12$ (dashed), and $\bar{\alpha}_2 = 0{\cdot}24$ (solid); (c) $\mu = 0$, $\gamma = 0{\cdot}3$, $\bar{\alpha}_2 = 0{\cdot}09$, and $\bar{\beta}_2 = -0{\cdot}21$ (long-dashed), $\bar{\beta}_2 = 0$ (dot-dashed), $\bar{\beta}_2 = 0{\cdot}165$ (dashed), and $\bar{\beta}_2 = 0{\cdot}21$ (solid).

The circular kurtosis and skewness of Mardia (1972, § 3.7) can also be expressed in closed form.

Given Lemma 1, it can be advantageous in terms of interpretation to reparameterize ρ and λ via $\rho = \gamma^{-1}(\bar{\alpha}_2^2 + \bar{\beta}_2^2)^{1/2}$ and $\lambda = \arg\left(\bar{\alpha}_2 + i\bar{\beta}_2\right)$. Clearly there is one-to-one correspondence between $(\mu, \gamma, \rho, \lambda)$ and $(\mu, \gamma, \bar{\alpha}_2, \bar{\beta}_2)$. Then, density (3) can be written as

$$g(\theta) = \frac{1}{2\pi}\left[1 + 2\gamma^2 \frac{\gamma\cos(\theta-\mu) - \bar{\alpha}_2}{\gamma^2 + \bar{\alpha}_2^2 + \bar{\beta}_2^2 - 2\gamma\{\bar{\alpha}_2\cos(\theta-\mu) + \bar{\beta}_2\sin(\theta-\mu)\}}\right], \quad -\pi \leqslant \theta < \pi, \tag{4}$$

where $-\pi \leqslant \mu < \pi$, $0 \leqslant \gamma < 1$, and $(\bar{\alpha}_2, \bar{\beta}_2) \neq (\gamma, 0)$ satisfy

$$(\bar{\alpha}_2 - \gamma^2)^2 + \bar{\beta}_2^2 \leqslant \gamma^2(1-\gamma)^2. \tag{5}$$

With the density in form (4), the parameters $\mu, \gamma, \bar{\alpha}_2$ and $\bar{\beta}_2$ are interpreted directly as location, concentration as measured by the mean resultant length, kurtosis and skewness parameters, respectively. Write $\Theta \sim \mathrm{G}(\mu, \gamma, \bar{\alpha}_2, \bar{\beta}_2)$ when the distribution is in this parameterization.

Density (3) or (4) is unimodal if $\gamma > 0$ and uniform if $\gamma = 0$. It is symmetric if and only if $\bar{\beta}_2 = 0$. Figure 1 plots the proposed densities using the parametrization in (4) for selected values of the parameters. Panel (a) illustrates the interpretation of γ: as γ increases, the concentration of the density increases as shown. The function of the parameter $\bar{\alpha}_2$ is clear from panel (b). It appears that the proposed model can provide a wide range of kurtosis even for a fixed value of γ. If $\bar{\alpha}_2$ is small, the density is flat-topped; the greater the value of $\bar{\alpha}_2$, the sharper the peakedness. Finally, the interpretation of the parameter $\bar{\beta}_2$ as controlling the skewness of the distribution is exemplified by panel (c). As this panel suggests, the range of skewness of our model also seems wide. The density associated with $-\lambda$ is the reflection in μ of the density associated with λ.

3. Skewness, kurtosis and other properties

3·1. *Extent of kurtosis and skewness*

Conditions (5) lead directly to bounds on the circular kurtosis and circular skewness.

THEOREM 2. *For fixed* γ $(=\gamma_1)$*:*

(i) *for the circular kurtosis of* Batschelet (1981)*,* $\bar{\alpha}_2 = \bar{\alpha}_2(\rho, \lambda)$*,*

$$\sup_{\rho,\lambda} \bar{\alpha}_2(\rho, \lambda) = \lim_{\rho\to 1} \bar{\alpha}_2(\rho, 0) = \gamma, \qquad \min_{\rho,\lambda} \bar{\alpha}_2(\rho, \lambda) = \bar{\alpha}_2(\rho_K, \lambda_K) = \gamma(2\gamma - 1),$$

respectively, where $\rho_K = |2\gamma - 1|$ *and* $\lambda_K = 0$ *if* $\gamma \geqslant 0{\cdot}5$ *and* $\lambda_K = \pi$ *if* $\gamma < 0{\cdot}5$*. The extremal values of kurtosis are associated with symmetric distributions;*

(ii) *for the circular skewness of* Batschelet (1981)*,* $\bar{\beta}_2 = \bar{\beta}_2(\rho, \lambda)$*,*

$$\max_{\rho,\lambda} \bar{\beta}_2(\rho, \lambda) = \bar{\beta}_2(\rho_S, \lambda_S) = \gamma(1 - \gamma), \qquad \min_{\rho,\lambda} \bar{\beta}_2(\rho, \lambda) = \bar{\beta}_2(\rho_S, -\lambda_S) = -\gamma(1 - \gamma),$$

respectively, where $\rho_S = (2\gamma^2 - 2\gamma + 1)^{1/2}$ *and* $\lambda_S = \arccos(\gamma/\rho_S)$*. The extremal values of skewness correspond to distributions where* $\bar{\alpha}_2 = \gamma^2$*.*

For $\gamma = 0{\cdot}3$, the minimal kurtosis density is that with $\bar{\alpha}_2 = -0{\cdot}12$ shown in Fig. 1(b). When $\bar{\alpha}_2$ is close to its supremum, the density shows a spike around $\theta = \mu$. For the same value of γ, the minimum and maximum skewness densities are those shown when $\bar{\beta}_2 = \pm 0{\cdot}21$ in Fig. 1(c).

Overall extremal kurtosis and skewness densities, optimizing over γ too, are readily found.

COROLLARY 1.

(i) *For the circular kurtosis of Batschelet* (1981)*,* $\bar{\alpha}_2 = \bar{\alpha}_2(\rho, \lambda, \gamma)$*,*

$$\sup_{\rho,\lambda,\gamma} \bar{\alpha}_2(\rho, \lambda, \gamma) = \lim_{\rho,\gamma\to 1} \bar{\alpha}_2(\rho, 0, \gamma) = 1,$$

$$\min_{\rho,\lambda,\gamma} \bar{\alpha}_2(\rho, \lambda, \gamma) = \bar{\alpha}_2(\tfrac{1}{2}, 0, \tfrac{1}{4}) = \bar{\alpha}_2(\tfrac{1}{2}, \pi, \tfrac{1}{4}) = -\tfrac{1}{8};$$

(ii) *for the circular skewness of Batschelet* (1981)*,* $\bar{\beta}_2 = \bar{\beta}_2(\rho, \lambda, \gamma)$*,*

$$\max_{\rho,\lambda,\gamma} \bar{\beta}_2(\rho, \lambda, \gamma) = \bar{\beta}_2(\tfrac{1}{\sqrt{2}}, \tfrac{\pi}{4}, \tfrac{1}{2}) = \tfrac{1}{4}, \qquad \min_{\rho,\lambda,\gamma} \bar{\beta}_2(\rho, \lambda, \gamma) = \bar{\beta}_2(\tfrac{1}{\sqrt{2}}, -\tfrac{\pi}{4}, \tfrac{1}{2}) = -\tfrac{1}{4}.$$

Maximum kurtosis is achieved by a point distribution with singularity at $\Theta = \mu$; such a distribution arises as $\rho, \gamma \to 1$. Minimum kurtosis is associated with the distribution with density $g_{\mathrm{MK}}(\theta) = 3\{1 + \cos(\theta - \mu)\}/[\pi\{5 + 4\cos(\theta - \mu)\}]$, $-\pi \leqslant \theta < \pi$. We conjecture that the minimum kurtosis achievable by a unimodal circular distribution is that associated with a uniform distribution on $\pm 3\pi/4$ which is $-2/(3\pi) \simeq -0{\cdot}212$; the kurtosis of $-0{\cdot}125$ associated with g_{MK} is not much greater than this. Wrapped stable distributions do not encompass flat-topped densities; their kurtosis is bounded below by 0, the kurtosis of the circular uniform distribution. Maximum ($M = 1$) and minimum ($M = -1$) amounts of skewness are associated with the distributions with densities $g_{\mathrm{M}}(\theta) = \{1 - M\sin(\theta - \mu)\}/[\pi\{3 - 2M\sin(\theta - \mu) - 2\cos(\theta - \mu)\}]$, $-\pi \leqslant \theta < \pi$. We venture that the corresponding skewnesses, which are broadly in line with the extrema achieved by other families of unimodal circular distributions, are large given the unimodality constraint, but make no more specific result or conjecture.

3·2. *Some other properties*

In the general case, we have successfully generated random variables from (3) using a natural but perhaps simplistic acceptance/rejection algorithm based on a random variable generated from the wrapped Cauchy distribution. Some special cases can be generated without any rejection.

The wrapped Cauchy and cardioid distributions, which are submodels of (3) as described in § 4·1, possess closure under convolution and multiplication by certain constants. As the following theorem shows, our extended family maintains these properties. Its veracity is clear from the form of the trigonometric moments (2).

THEOREM 3. *The following hold for the proposed model* (3)*:*

(i) $\Theta_1 \sim \mathrm{G}(\mu_1, \gamma_1, \rho_1, \lambda_1)$, $\Theta_2 \sim \mathrm{G}(\mu_2, \gamma_2, \rho_2, \lambda_2)$, $\Theta_1 \perp\!\!\!\perp \Theta_2$ *implies* $\Theta_1 + \Theta_2 \sim \mathrm{G}(\mu_1 + \mu_2, \gamma_1\gamma_2, \rho_1\rho_2, \lambda_1 + \lambda_2)$;
(ii) $\Theta \sim \mathrm{G}(\mu, \gamma, \rho, \lambda)$ *implies* $n\Theta \pmod{2\pi} \sim \mathrm{G}\{n\mu + (n-1)\lambda, \gamma\rho^{n-1}, \rho^n, \lambda\}$, $n \in \mathbb{N}$;
(iii) $\Theta \sim \mathrm{G}(\mu, \gamma, \rho, \lambda)$ *implies* $-\Theta \sim \mathrm{G}(-\mu, \gamma, \rho, -\lambda)$.

In addition, infinite divisibility holds for the subset of the symmetric submodels of (3) with $\gamma \leqslant \rho$ and $\lambda = 0$.

4. SUBMODELS

4·1. *Some two- and three-parameter submodels*

First, if $\bar{\alpha}_2 = \gamma^2$ and $\bar{\beta}_2 = 0$, or equivalently $\rho = \gamma$ and $\lambda = 0$, the model reduces to the wrapped Cauchy distribution with density (1). Second, when $\bar{\alpha}_2 = \bar{\beta}_2 = 0$, or equivalently $\rho = 0$, the model becomes the cardioid distribution (Mardia & Jupp, 1999, § 3.5.5; Jammalamadaka & SenGupta, 2001, § 2.2.2) with density $g_{\mathrm{C}}(\theta) = \{1 + 2\gamma\cos(\theta - \mu)\}/(2\pi)$, $-\pi \leqslant \theta < \pi$; here, $0 \leqslant \gamma < 1/2$ as a consequence of (5) when $\bar{\alpha}_2 = 0$. The opportunity thereby arises for likelihood ratio testing of the appropriateness of these models within our much wider class. As noted in § 2, when $\gamma = 0$, the distribution reduces to the uniform distribution on the circle. Suitably normalized, the high concentration limit of our model is, as for the wrapped Cauchy distribution, the Cauchy distribution. Another special case is the three-parameter sine-skewed wrapped Cauchy distribution (Umbach & Jammalamadaka, 2009; Abe & Pewsey, 2011).

4·2. *Main three-parameter submodels*

Let $\bar{\beta}_2 = 0$ in (4). Then we obtain the family of symmetric models with density

$$g_{\mathrm{S}}(\theta) = \frac{1}{2\pi}\left\{1 + 2\gamma^2 \frac{\gamma\cos(\theta - \mu) - \bar{\alpha}_2}{\gamma^2 + \bar{\alpha}_2^2 - 2\gamma\bar{\alpha}_2\cos(\theta - \mu)}\right\}, \tag{6}$$

where $\gamma(2\gamma - 1) \leqslant \bar{\alpha}_2 \leqslant \gamma$. For any fixed γ, the range of kurtosis of this submodel is as wide as that of the four-parameter full model. So this is a three-parameter symmetric family of circular distributions, extending from a point distribution through to the minimum kurtosis distribution with density g_{MK}, containing both wrapped Cauchy and, when $\gamma < 1/2$, cardioid distributions as special cases. This submodel has all the closure properties given in Theorem 3 and is infinitely divisible for $\gamma \leqslant \rho$. Density (6) for $\gamma = 0{\cdot}3$ and selected values of $\bar{\alpha}_2$ is shown in Fig. 1(c).

We define a three-parameter asymmetric submodel by letting $\bar{\alpha}_2 = \gamma^2$ in (4). The density is

$$g_{\mathrm{A}}(\theta) = \frac{1}{2\pi}\left[1 + 2\gamma^3 \frac{\cos(\theta - \mu) - \gamma}{\gamma^4 + \gamma^2 + \bar{\beta}_2^2 - 2\gamma\{\gamma^2\cos(\theta - \mu) + \bar{\beta}_2\sin(\theta - \mu)\}}\right], \tag{7}$$

where $-\gamma(1 - \gamma) \leqslant \bar{\beta}_2 \leqslant \gamma(1 - \gamma)$. The range of skewness of this submodel is the same as that of the full family (4). The kurtosis of the submodel, $\bar{\alpha}_2 = \gamma^2$, is fixed to be that of its

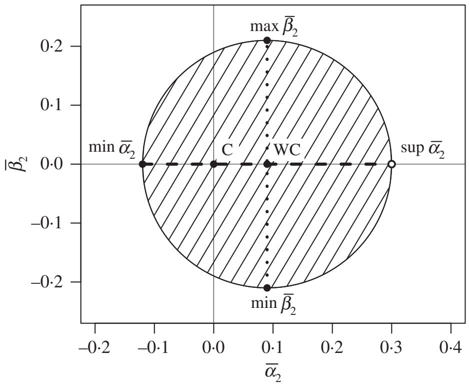

Fig. 2. Domain of $(\bar{\alpha}_2, \bar{\beta}_2)$ $(D_\gamma$; line-shaded) with $\gamma = 0{\cdot}3$. Positions of special cases indicated are of: WC, the wrapped Cauchy distribution; C, the cardioid distribution; min $\bar{\alpha}_2$, the submodel with minimum kurtosis; sup $\bar{\alpha}_2$, the limiting distribution with the supremum kurtosis; min $\bar{\beta}_2$, the submodel with minimum skewness; max $\bar{\beta}_2$, the submodel with maximum skewness; (6), the three-parameter symmetric submodel (horizontal dashed line); and (7), the three-parameter asymmetric submodel (vertical dotted line).

wrapped Cauchy special case. This model might therefore be called an asymmetric, or skew, wrapped Cauchy distribution. The distribution is unimodal and satisfies properties (ii) and (iii) of Theorem 3. Figure 1(d) plots density (7) for $\gamma = 0{\cdot}3$ and four selected values of $\bar{\beta}_2$.

4·3. *Submodels and domain of kurtosis and skewness*

For each fixed $0 \leqslant \gamma < 1$, conditions (5) give the domain of $(\bar{\alpha}_2, \bar{\beta}_2)$ to be the disc with centre $(\gamma^2, 0)$ and radius $\gamma(1-\gamma)$; call this D_γ. Figure 2 plots D_γ and the positions of submodels of density (4) for a fixed value of $\gamma < 1/2$. The centre of D_γ corresponds to the wrapped Cauchy distribution. The cardioid distribution corresponds to the origin of the plane, which is within D_γ for all $0 \leqslant \gamma \leqslant 1/2$. The three-parameter symmetric submodel of § 4·2 corresponds to all points on the horizontal diameter, while points on the vertical diameter map out the three-parameter asymmetric submodel of § 4·2. It can easily be shown that the union of D_γ as γ varies is the set $|\bar{\beta}_2| \leqslant (1+8\bar{\alpha}_2)^{1/2}(1-\bar{\alpha}_2)/\surd 27$.

5. Parameter estimation

5·1. *Method of moments estimation*

Throughout this section we assume that $\Theta_1, \ldots, \Theta_n$ is an independent and identically distributed sample from $\mathrm{G}(\mu, \gamma, \bar{\alpha}_2, \bar{\beta}_2)$. Equating the first trigonometric moment of Θ with its sample analogue, $S_{1n} = n^{-1}\sum_{j=1}^n e^{i\Theta_j}$, moment estimators of μ and γ are given by the sample mean direction and mean resultant length, $\hat{\mu} = \arg(S_{1n})$ and $\hat{\gamma} = |S_{1n}|$, respectively. Clearly these estimators are always within the range of μ and γ and hence well-defined. We derive moment estimators of $\bar{\alpha}_2$ and $\bar{\beta}_2$ from the sample kurtosis $\bar{a}_2 = n^{-1}\sum_j \cos\{2(\Theta_j - \hat{\mu})\}$ and sample skewness $\bar{b}_2 = n^{-1}\sum_j \sin\{2(\Theta_j - \hat{\mu})\}$. For given γ, the domain of $\bar{\alpha}_2$ and $\bar{\beta}_2$, D_γ, given by (5) and shown when $\gamma = 0.3$ in Fig. 2, is a sub-disc of the unit disc within which $(\bar{a}_2, \bar{b}_2)$ lies. We therefore define $(\hat{\alpha}_2, \hat{\beta}_2)$ to equal $(\bar{a}_2, \bar{b}_2)$ if $(\bar{a}_2, \bar{b}_2)$ is within $D_{\hat{\gamma}}$ and to be at the intersection of the boundary of $D_{\hat{\gamma}}$ and the line connecting $(\bar{a}_2, \bar{b}_2)$ with $(\hat{\gamma}^2, 0)$ if $(\bar{a}_2, \bar{b}_2)$ is not within $D_{\hat{\gamma}}$, that is, $\hat{\alpha}_2 =$

$\hat{\gamma}^2 + \hat{\gamma}(1-\hat{\gamma})(\bar{a}_2 - \hat{\gamma}^2)/\{(\bar{a}_2 - \hat{\gamma}^2)^2 + \bar{b}_2^2\}^{1/2}$, $\hat{\beta}_2 = \hat{\gamma}(1-\hat{\gamma})\bar{b}_2/\{(\bar{a}_2 - \hat{\gamma}^2)^2 + \bar{b}_2^2\}^{1/2}$, when $(\bar{a}_2, \bar{b}_2) \notin D_{\hat{\gamma}}$. It follows from Theorem 1 of Pewsey (2004) that the asymptotic distribution of $(\hat{\mu}, \hat{\gamma}, \bar{a}_2, \bar{b}_2)$ is a multivariate normal distribution.

5·2. *Maximum likelihood estimation*

Next we consider maximum likelihood estimation of the parameters of the proposed model. As usual, the maximum likelihood estimates of $(\mu, \gamma, \bar{\alpha}_2, \bar{\beta}_2)$ must be obtained numerically but also as usual this causes no great difficulty. The method of moments estimates provide useful starting values, but it is advisable to utilize multiple starting values to ensure that the global maximum is identified, particularly for data close to uniformly distributed. In order to use optimization methods such as nlminb in R (R Development Core Team, 2014), it is required that the range of each parameter does not depend on other parameters. It is possible to satisfy this requirement by reparameterizing $\bar{\alpha}_2$ and $\bar{\beta}_2$ as $\bar{\alpha}_2 = \gamma^2 + \gamma(1-\gamma)\omega\cos\phi$, $\bar{\beta}_2 = \gamma(1-\gamma)\omega\sin\phi$, where $0 \leqslant \omega \leqslant 1$, $-\pi \leqslant \phi < \pi$, and $(\omega, \phi) \neq (1, 0)$. In our work, we use the PORT routine in nlminb. Asymptotic normality of the maximum likelihood estimator follows from standard theory. Extensions to regression where one or more parameters, most usually μ, depend on covariates are immediate in the maximum likelihood setting, and will not be pursued explicitly.

5·3. *Simulations and example*

We compared the performance of the method of moments and maximum likelihood estimators for finite sample sizes via a Monte Carlo simulation study. We found that the method of moments estimator is preferable for small to moderate n and parameter values far from the boundary of the parameter space. Otherwise, the maximum likelihood estimator shows better performance, as it must, of course, for large n.

As an illustrative example, we consider a dataset recording the $n = 15\,831$ times of gun crimes committed in Pittsburgh, U.S.A., in the period 1987–98 (Gill & Hangartner, 2010). The data were recorded in intervals of an hour. Figure 3 shows a linear histogram of the data converted from 24 hours to angles in $[-\pi, \pi)$; for clarity, $-\pi$ corresponds to midday, 0 to midnight, etc. The data are asymmetric and peaked around 10 p.m. to 1 a.m. Table 1 shows the maximum likelihood estimates of the parameters, the maximized loglikelihood, and information criteria values for the full model (4), the three-parameter symmetric submodel (6), the three-parameter asymmetric submodel (7) and the wrapped Cauchy and cardioid distributions. Both Table 1 and Fig. 3 make it clear that the three symmetric submodels considered are not as good as the two asymmetric models at fitting these data. The full model (4) and the three-parameter asymmetric submodel (7) fit the data much better. The loglikelihood ratio test of model (4) against (7) yields a miniscule p-value, clearly rejecting model (7), which Fig. 3 suggests is less satisfactory around both the mode and antimode of the histogram of the dataset. It is possible to model these data in a visually similar manner by other four-parameter asymmetric distributions; for example, the maximized loglikelihood of the inverse Batschelet distribution (Jones & Pewsey, 2012) based on Jones & Pewsey's (2005) symmetric family increases the maximum loglikelihood by 1·78. Because of the very large sample size, even these models are rejected by chi-squared goodness-of-fit tests, but Fig. 3 suggests this level of modelling to be adequate for most practical purposes. And model (4) wins in interpretability terms; for example, $\hat{\beta}_2$ and $\hat{\alpha}_2$ immediately show estimated levels of skewness and kurtosis. In addition, for grouped data as in this example, maximum likelihood estimation, as well as method of moments, is very much faster for model (4) than for its competitors because of the explicit formula for the distribution function.

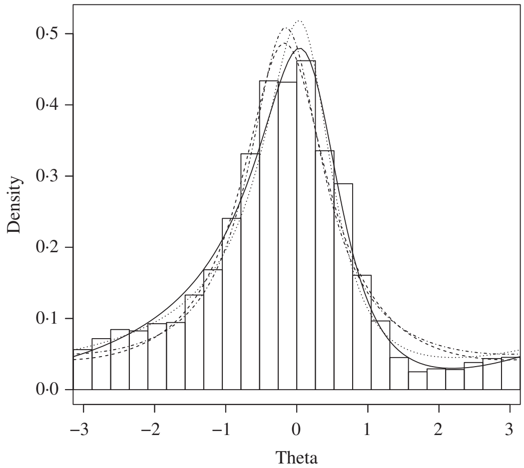

Fig. 3. Histogram of the gun crime data with maximum likelihood fits of the full model (4) (solid), the three-parameter symmetric submodel (6) (dashed), the three-parameter asymmetric submodel (7) (dotted), and the wrapped Cauchy submodel (dot-dashed).

Table 1. *Maximum likelihood estimates of the parameters, with standard errors in square brackets, the maximized loglikelihood, $\ell_{\max}$, and the values of Akaike Information Criterion, AIC, and Bayesian Information Criterion, BIC, for the full model* (4) *and four of its submodels fitted to the gun crime data*

Model	$\hat{\mu}$	$\hat{\gamma}$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\ell_{\max}$	AIC	BIC
Full model (4)	−0·384 [0·018]	0·547 [0·009]	0·233 [0·011]	0·115 [0·008]	−44741·00	89490·00	89520·68
Three-parameter symmetric (6)	−0·302 [0·018]	0·543 [0·009]	0·257 [0·010]	(0)	−44992·29	89990·58	90013·59
Three-parameter asymmetric (7)	−0·341 [0·017]	0·519 [0·008]	(0·269)	0·095 [0·008]	−44841·44	89688·88	89711·89
Wrapped Cauchy	−0·286 [0·017]	0·524 [0·008]	(0·274)	(0)	−45025·35	90054·70	90070·04
Cardioid	−0·510 [0·025]	0·432 [0·006]	(0)	(0)	−46059·42	92122·84	92138·18

6. Comparisons

In this final section, we make comparisons between the proposed model and the other unimodal four-parameter distributions explicitly mentioned in § 1, namely the wrapped stable distribution (Pewsey, 2008), direct Batschelet distributions (Abe et al., 2013) and inverse Batschelet distributions (Jones & Pewsey, 2012). We do so initially using the seven points listed in § 1 and then with respect to some other considerations. In the latter case, we mention some not-necessarily-unimodal competitors that have other advantages.

(i) The wrapped stable distribution shares with the new model the property of having simple trigonometric moments; this is not the case for direct or inverse Batschelet distributions.

(ii) The new model has the considerable advantage over the wrapped stable distribution of having tractable density and distribution functions. Direct and inverse Batschelet distributions often have tractable density functions, the inverse version having advantages in terms of simplicity of normalizing constant. Neither Batschelet distribution has a tractable distribution function.

(iii) From (i) and (ii), the wrapped stable distribution affords straightforward estimation by the method of moments, although also with a range constraint problem. Its lack of an explicit density function, in a form not including an infinite sum, makes maximum likelihood more problematic, but see Pewsey (2008, § 3.2) and, for the symmetric case, Gatto & Jammalamadaka (2003). Moment methods are not readily available for Batschelet distributions but maximum likelihood estimation is. Inverse Batschelet distributions display a large degree of parameter orthogonality unique amongst circular distributions and comparable only to two-piece distributions (Jones & Anaya-Izquierdo, 2011) on $\mathbb{R}$. In personal communication, Arthur Pewsey has indicated that the implicit inverse function associated with inverse Batschelet distributions slows computations sufficiently for him to maintain interest in direct Batschelet distributions.

(iv) The new model has fully interpretable parameters, individually measuring location, concentration, skewness and kurtosis in the classical trigonometric moment sense. Wrapped stable and Batschelet distributions have parameters fulfilling similar roles, though possibly in a less well understood sense, especially regarding kurtosis.

(v) All models being compared share a very wide range of skewness and kurtosis, except for the wrapped symmetric stable distribution and low kurtosis, as mentioned in § 3·1. It is, however, much harder to provide explicit bounds on skewness and/or kurtosis for the competing models.

(vi) Let us concentrate on the most familiar symmetric submodels of the competing distributions, since although each contains, for example, interesting three-parameter skewed submodels, these are as yet not well known. Our model includes wrapped Cauchy and cardioid distributions as special cases. Wrapped symmetric stable distributions include wrapped Cauchy and wrapped normal distributions. When used in skewing form, direct and inverse Batschelet distributions can be applied to classes of symmetric distributions such as that of Jones & Pewsey (2005); then, their familiar special cases include the wrapped Cauchy, cardioid and von Mises distributions. Inclusion of the von Mises distribution is attractive but not achieved by the model of this paper. However, symmetric submodels of our model approximate the von Mises distribution very closely for small and medium levels of concentration. Of course, all these models, and most others, include the circular uniform distribution.

(vii) Closure under convolution is shared with wrapped stable distributions but not the others. Families including the von Mises distribution do not possess this property in general.

All the distributions compared above are essentially mathematical constructs for use in empirical modelling of circular data. Families of distributions arising from physical mechanisms are appealing; a rare tractable example is that of Kato & Jones (2013) arising from the interaction of a planar Brownian particle with certain circles, but members of this family can be bimodal. It is also appealing if circular distributions interact kindly with Möbius transformations. Building on earlier specific cases, we explored such a family in Kato & Jones (2010), but again had to sacrifice universal unimodality. That family of distributions is, however, particularly attractive for use in circular–circular regression models. We have already mentioned the important class of generalized von Mises distributions (Yfantis & Borgman, 1982; Gatto & Jammalamadaka, 2007; Gatto, 2009), but they are especially attuned to coping with bi- and multi-modality. Some of the alternative distributions discussed above have natural bivariate circular extensions; we intend to pursue such extensions of the current model in future work, closure under convolution being helpful in one approach we have in mind. All in all, as a four-parameter family of unimodal circular distributions, we feel that the models with density (3)/(4) are an especially attractive addition to the genre.

Acknowledgement

We are very grateful to Jeff Gill for providing us with the dataset, to Arthur Pewsey and others for helpful comments and encouragement, and to the editorial team for further suggestions.

Supplementary material

Supplementary material available at *Biometrika* online relates to § 2, including distribution function, mode and antimode, § 3·1, including minimum kurtosis and extremal skewness densities, § 3·2, including random variate generation algorithm and analysis thereof, § 4·1, § 4·2, including (6) as a mixture distribution, § 4·3, including figure of $\cup D_\gamma$, § 5·1, § 5·2, including observed and expected information matrices, § 5·3, including a simulation study and another example, and § 6, including approximation to the von Mises distribution.

References

Abe, T. & Pewsey, A. (2011). Sine-skewed circular distributions. *Statist. Papers* **52**, 683–707.

Abe, T., Pewsey, A. & Shimizu, K. (2013). Extending circular distributions through transformation of argument. *Ann. Inst. Statist. Math.* **65**, 833–58.

Batschelet, E. (1981). *Circular Statistics in Biology.* London: Academic Press.

Gatto, R. (2009). Information theoretic results for circular distributions. *Statistics* **43**, 409–21.

Gatto, R. & Jammalamadaka, S. R. (2003). Inference for wrapped symmetric α-stable circular models. *Sankhyā* **65**, 333–55.

Gatto, R. & Jammalamadaka, S. R. (2007). The generalized von Mises distribution. *Statist. Meth.* **4**, 341–53.

Gill, J. & Hangartner, D. (2010). Circular data in political science and how to handle it. *Polit. Anal*. **18**, 316–36.

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

Jones, M. C. & Anaya-Izquierdo, K. (2011). On parameter orthogonality in symmetric and skew models. *J. Statist. Plan. Infer.* **141**, 758–70.

Jones, M. C. & Pewsey, A. (2005). A family of symmetric distributions on the circle. *J. Am. Statist. Assoc.* **100**, 1422–8.

Jones, M. C. & Pewsey, A. (2012). Inverse Batschelet distributions for circular data. *Biometrics* **68**, 183–93.

Kato, S. & Jones, M. C. (2010). A family of distributions on the circle with links to, and applications arising from, Möbius transformation. *J. Am. Statist. Assoc.* **105**, 249–62.

Kato, S. & Jones, M. C. (2013). An extended family of circular distributions related to wrapped Cauchy distributions via Brownian motion. *Bernoulli* **19**, 154–71.

Kent, J. T. & Tyler, D. E. (1988). Maximum likelihood estimation for the wrapped Cauchy distribution. *J. Appl. Statist.* **15**, 247–54.

Mardia, K. V. (1972). *Statistics of Directional Data.* London: Academic Press.

Mardia, K. V. & Jupp, P. E. (1999). *Directional Statistics.* Chichester: Wiley.

McCullagh, P. (1996). Möbius transformation and Cauchy parameter estimation. *Ann. Statist.* **24**, 787–808.

Pewsey, A. (2004). The large-sample joint distribution of key circular statistics. *Metrika* **60**, 25–32.

Pewsey, A. (2008). The wrapped stable family of distributions as a flexible model for circular data. *Comp. Statist. Data Anal.* **52**, 1516–23.

R Development Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0. http://www.R-project.org.

Umbach, D. & Jammalamadaka, S. R. (2009). Building asymmetry into circular distributions. *Statist. Prob. Lett.* **79**, 659–63.

Yfantis, E. A. & Borgman, L. E. (1982). An extension of the von Mises distribution. *Commun. Statist.* A **11**, 1695–706.

[*Received July* 2013. *Revised October* 2014]