Biometrika (201*), **10***, *, *pp*. 1–19

Printed in Great Britain

Advance Access publication on ** **** 201*

Supplementary Material for "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

1. PROOF OF THEOREM 1 (§2 IN ARTICLE)

Assuming that ϕ_Θ is a valid set of trigonometric moments, the associated density can be obtained by inversion. Using $\sum_{p=1}^{\infty} a^p = a/(1-a)$, where $|a| < 1$, we have

$$\begin{aligned} g(\theta) &= \frac{1}{2\pi} \sum_{p=-\infty}^{\infty} \phi_\Theta(p)\, e^{-ip\theta} = \frac{1}{2\pi}\left(1 + 2\,\mathrm{Re}\left[\beta e^{i\alpha} \sum_{p=1}^{\infty} \{\rho e^{i(\eta-\theta)}\}^p\right]\right) \\ &= \frac{1}{2\pi}\left[1 + 2\,\mathrm{Re}\left\{\beta e^{i\alpha} \frac{\rho e^{i(\eta-\theta)}}{1-\rho e^{i(\eta-\theta)}}\right\}\right] \\ &= \frac{1}{2\pi}\left\{1 + 2\beta\rho\, \frac{\cos(\theta-\eta-\alpha) - \rho\cos\alpha}{1+\rho^2-2\rho\cos(\theta-\eta)}\right\}, \end{aligned} \tag{S1}$$

which gives (3) upon using the reparametrization leading to (2).

It remains to show that, under appropriate parameter restrictions, ϕ_Θ is a valid set of trigonometric moments; this is more easily done by showing that g is a valid circular density function. Integrability of g is easily shown provided $0 \le \rho < 1$. It remains to consider further conditions on γ, $\bar\alpha_2$ and $\bar\beta_2$ such that $g(\theta) \ge 0$ for all $-\pi \le \theta < \pi$. We work with the density in the form (4) which can be rewritten as

$$g(\theta) = \frac{1}{2\pi} \frac{\gamma^2 + \bar\alpha_2^2 + \bar\beta_2^2 - 2\gamma^2\bar\alpha_2 + 2\gamma\{(\gamma^2 - \bar\alpha_2)\cos(\theta-\mu) - \bar\beta_2\sin(\theta-\mu)\}}{\gamma^2 + \bar\alpha_2^2 + \bar\beta_2^2 - 2\gamma\{\bar\alpha_2\cos(\theta-\mu) + \bar\beta_2\sin(\theta-\mu)\}}.$$

The denominator of $g(\theta)$ is positive and, using the fact that $a\cos(\theta-\mu) + b\sin(\theta-\mu)$ has minimum value $-(a^2+b^2)^{1/2}$ whatever the signs of a and b, the numerator of $g(\theta)$ has minimum value

$$\gamma^2 + \bar\alpha_2^2 + \bar\beta_2^2 - 2\gamma^2\bar\alpha_2 - 2\gamma\{(\bar\alpha_2 - \gamma^2)^2 + \bar\beta_2^2\}^{1/2} = (Q-\gamma)^2 - \gamma^4,$$

say, where $Q = \{(\bar\alpha_2 - \gamma^2)^2 + \bar\beta_2^2\}^{1/2}$. The numerator is therefore nonnegative whenever either $Q \ge \gamma(1+\gamma)$, which is impossible because $\bar\alpha_2^2 + \bar\beta_2^2 = \rho^2\gamma^2 < \gamma^2$, or $Q \le \gamma(1-\gamma)$, which is equivalent to condition (5).

2. DISTRIBUTION FUNCTION FORMULAE (§2 IN ARTICLE)

When defined to start from $\mu - \pi$, and in parametrization (3), the distribution function associated with the full density can readily be checked to be

$$G_\Theta(\theta) = \frac{1}{2\pi}\left(\left(1 - \frac{\gamma\cos\lambda}{\rho}\right)(\theta - \mu + \pi) + \frac{\gamma\sin\lambda}{\rho}\log\left\{\frac{1+\rho^2+2\rho\cos\lambda}{1+\rho^2-2\rho\cos(\theta-\mu-\lambda)}\right\}\right.$$
$$+\frac{2\gamma\cos\lambda}{\rho}\left[\arctan\left\{\frac{1+\rho}{1-\rho}\tan\left(\frac{\theta-\mu-\lambda}{2}\right)\right\}\right.$$
$$\left.\left. - \arctan\left\{\frac{1+\rho}{(1-\rho)\tan(\lambda/2)}\right\} + \pi I\left\{\tan\left(\frac{\theta-\mu-\lambda}{2}\right) < \frac{1}{\tan(\lambda/2)}\right\}\right]\right),$$

$\mu - \pi \le \theta \le \mu + \pi$, provided $\rho \neq 0,\ \lambda \neq -\pi, 0$, where I is an indicator function.

When $\rho = 0$, the model reduces to the cardioid distribution whose density is given in §4.1 of the article and whose distribution function is simply

$$G_C(\theta) = \frac{1}{2\pi}\left\{\theta - \mu + \pi + 2\gamma\sin(\theta - \mu)\right\}, \quad \mu - \pi \le \theta \le \mu + \pi.$$

When $\lambda = -\pi, 0\ \rho \neq 0$, the distribution function is essentially that of the three-parameter symmetric submodel described in §4.2 of the article, and is given by

$$G(\theta) = \frac{1}{2\pi}\left(\left(1 - \frac{\gamma}{\tilde\rho}\right)(\theta - \mu + \pi) + \frac{\gamma\pi}{\tilde\rho} + \frac{2\gamma}{\tilde\rho}\left[\arctan\left\{\frac{1+\tilde\rho}{1-\tilde\rho}\tan\left(\frac{\theta-\mu}{2}\right)\right\}\right]\right),$$

$\mu - \pi \le \theta \le \mu + \pi$, where $\tilde\rho = \rho\cos\lambda$. It is interesting to note how simple an extension this is of the wrapped Cauchy distribution function which results when $\gamma = \rho$.

3. PROOF OF UNIMODALITY AND EXPRESSIONS FOR MODE AND ANTIMODE (§2 IN ARTICLE)

Proof of unimodality follows immediately from the representation of $g(\theta)$ given in equation (S1) above, which can be recognized as $g(\theta) = [1 + 2\mathrm{Re}\{\mathcal{M}(\theta)\}]/(2\pi)$ where $\mathcal{M}(\theta)$ is a Möbius transformation on $-\pi \le \theta < \pi$. Since the Möbius transformation with $\beta, \rho \neq 0$ is a continuous bijective function mapping $[-\pi, \pi)$ onto a circle in the complex plane, the real part of $\mathcal{M}(\theta)$ is monotonically increasing from θ_m to θ_M and monotonically decreasing from θ_M to $\theta_m + 2\pi$, where $\theta_M = \mathrm{argmax}_\theta \mathrm{Re}\{\mathcal{M}(\theta)\}$ is the mode and $\theta_m = \mathrm{argmin}_\theta \mathrm{Re}\{\mathcal{M}(\theta)\}$ is the antimode of g (Rudin, 1987, Chapter 14). Therefore, unimodality holds for model (3). In addition, explicit expressions for the mode and antimode of the distribution when $\gamma > 0$ are:

$$\theta_M = \begin{cases} v_1 + v_2, & \sin\lambda > 0, \\ \mu, & \sin\lambda = 0, \\ v_1 + v_2 + \pi, & \sin\lambda < 0, \end{cases} \qquad \theta_m = \begin{cases} v_1 - v_2, & \sin\lambda > 0, \\ \mu + \pi, & \sin\lambda = 0, \\ v_1 - v_2 + \pi, & \sin\lambda < 0, \end{cases}$$

where

$$v_1 = \mu + \lambda + \arctan\left\{\frac{1-\rho^2}{(1+\rho^2)\tan\lambda}\right\}, \quad v_2 = -\arccos\left[\frac{2\rho\sin\lambda}{\{(1+\rho^2)^2 - 4\rho^2\cos^2\lambda\}^{1/2}}\right].$$

4. SOME EXTRA ASPECTS OF SHAPE (§2 IN ARTICLE)

In addition to the properties discussed in §3, the following properties regarding the shapes of the density also hold for the proposed model.

(i) Write $g(\theta) = \{1 + 2\gamma q(\theta; \mu, \rho, \lambda\}/(2\pi)$ where

$$q(\theta; \mu, \rho, \lambda) = \frac{\cos(\theta - \mu) - \rho\cos\lambda}{1 + \rho^2 - 2\rho\cos(\theta - \mu - \lambda)}, \quad -\pi \le \theta < \pi.$$

Then

$$q(\theta + \pi; \mu, \rho, \lambda + \pi) = -q(\theta; \mu, \rho, \lambda) \quad \text{for any } -\pi \le \theta < \pi.$$

Therefore the density g associated with $\lambda + \pi$ equals the reflection in the circular uniform density $1/(2\pi)$ of the density associated with λ and the shift of location $\theta \to \theta + \pi$.

(ii) The following hold for the density:

$$g(\theta) \begin{cases} > 1/(2\pi),\ \mu - w < \theta < \mu + w, \\ = 1/(2\pi),\ \theta = \mu \pm w, \\ < 1/(2\pi),\ \mu + w < \theta < \mu - w + 2\pi, \end{cases}$$

where $w = \arccos(\rho\cos\lambda) = \arccos(\bar{\alpha}_2/\gamma)$.

(iii) Let g_1 and g_2 be the densities of $\mathrm{G}(\mu, \rho, \gamma_1, \lambda)$ and $\mathrm{G}(\mu, \rho, \gamma_2, \lambda)$, respectively, where $\gamma_1, \gamma_2 \neq 0$. Then the shapes of g_1 and g_2, which differ only in concentration, are related in the sense that

$$g_2(\theta) = \frac{\gamma_2}{\gamma_1} g_1(\theta) + \left(1 - \frac{\gamma_2}{\gamma_1}\right) \frac{1}{2\pi}.$$

In particular, if $\gamma_2 < \gamma_1$ then the density with the smaller concentration, g_2, is a mixture of the density with the larger concentration, g_1, and the uniform density, the probability associated with g_1 being γ_2/γ_1. See Figure 1(a) in the article for plots of such densities.

5. GRAPHS OF MINIMUM KURTOSIS AND MAXIMUM SKEWNESS DENSITIES (§3.1 IN ARTICLE)

Set $\mu = 0$. Then, the minimum kurtosis density $g_{\mathrm{MK}}(\theta)$, which corresponds to $\gamma = 1/4$, $\bar{\alpha}_2 = -1/8$ and $\bar{\beta}_2 = 0$ and is given in §3.2 of the article, is shown in Figure S1(a); the maximum skewness density, $g_{\mathrm{M}}(\theta)$ with $M = 1$, which corresponds to $\gamma = 1/2$, $\bar{\alpha}_2 = \bar{\beta}_2 = 1/4$, is shown in Figure S1(b).

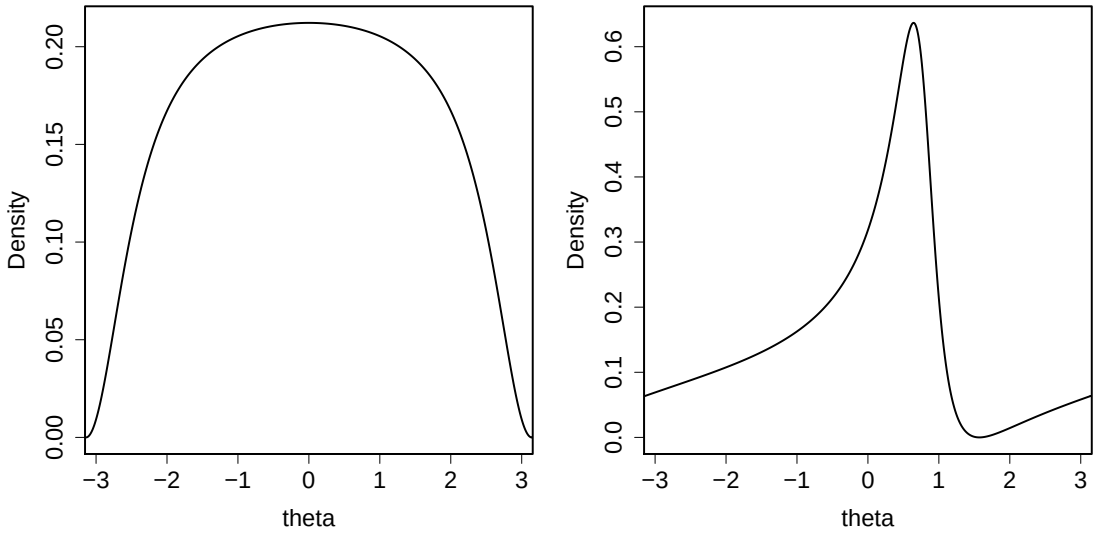


Fig. S1. (a) minimum kurtosis density g_{MK}; (a) maximum skewness density g_{M} with $M = 1$.

6. RANDOM VARIATE GENERATION ALGORITHM AND ACCEPTANCE PROBABILITY (§3.2 IN ARTICLE)

6·1. *The algorithm*

If $\Theta \sim g(\theta)$, then the density of $\Phi = \Theta - \mu - \lambda$ can be written

$$\begin{aligned} g_\Phi(\phi) &= \frac{1}{2\pi}\left[1 + \frac{2\gamma\{\cos(\phi+\lambda) - \rho\cos\lambda\}}{1+\rho^2-2\rho\cos\phi}\right] \\ &= \frac{1+\rho^2-2\bar\alpha_2 + 2\{(\bar\alpha_2-\rho^2)\cos\phi - \bar\beta_2\sin\phi\}/\rho}{2\pi\,(1+\rho^2-2\rho\cos\phi)} \\ &= \frac{1+\rho^2-2\bar\alpha_2 + 2R\cos(\phi+A)/\rho}{2\pi\,(1+\rho^2-2\rho\cos\phi)} \end{aligned}$$

where

$$R = \{(\bar\alpha_2-\rho^2)^2 + \bar\beta_2^2\}^{1/2} \text{ and } \cos A = (\bar\alpha_2-\rho^2)/R, \quad \sin A = \bar\beta_2/R.$$

Now let W be a random variable generated from the wrapped Cauchy distribution with density g_{WC} given at (1) in the article with $\mu = 0$. It is easy to see that the maximal value of $g_\Phi(\phi)/g_{\mathrm{WC}}(\phi)$ is $(1+\rho^2-2\bar\alpha_2+2R/\rho)/(1-\rho^2)$. It follows that Φ can be generated by an acceptance/rejection algorithm whereby $W = w$ is accepted as a value for Φ with probability

$$\frac{1+\rho^2-2\bar\alpha_2+2R\cos(w+A)/\rho}{1+\rho^2-2\bar\alpha_2+2R/\rho}$$

and rejected otherwise.

6·2. *Acceptance probability*

Using basic properties of the wrapped Cauchy distribution it is easy to see that the unconditional acceptance probability of the above algorithm is

$$P = \frac{1-\rho^2}{1+\rho^2-2\bar{\alpha}_2+2R/\rho}.$$

To get a feel for how good the acceptance probability is for symmetric distributions, consider the cases with $\bar{\beta}_2 = 0$, $\bar{\alpha}_2 = \pm\gamma\rho$. Then, when $\bar{\alpha}_2 = \gamma\rho$,

$$P = \frac{1-\rho^2}{1+\rho^2-2\gamma\rho+2|\gamma-\rho|}.$$

When $\gamma > \rho$,

$$P = \frac{1+\rho}{1+2\gamma-\rho} \geq \frac{1+\rho}{2} \geq \frac{1}{2},$$

using $2\gamma \leq 1+\rho$. When $\gamma < \rho$,

$$P = \frac{1-\rho}{1-2\gamma+\rho} \geq \frac{1-\rho}{1+\rho};$$

the lower bound is greater than or equal to $1/2$ whenever $\rho \leq 1/3$. However, when $\bar{\alpha}_2 = -\gamma\rho$,

$$P = \frac{1-\rho^2}{1+\rho^2+2\gamma\rho+2(\gamma+\rho)} = \frac{1-\rho}{1+2\gamma+\rho} \geq \frac{1-\rho}{2(1+\rho)},$$

using $2\gamma \leq 1+\rho$ again.

The special case in which $\bar{\alpha}_2 = \rho^2$ is one in which no rejection is necessary, as will be explained in §8 of this document. Another special case where we can look at the acceptance probability is for wide-bodied, but not symmetric, distributions for which $\bar{\alpha}_2 = 0$, $\bar{\beta}_2 = \pm\gamma\rho$. Then,

$$P = \frac{1-\rho^2}{1+\rho^2+2(\gamma^2+\rho^2)^{1/2}}.$$

Again using $\gamma \leq (1+\rho)/2$,

$$P \geq \frac{1-\rho^2}{1+\rho^2+(1+2\rho+5\rho^2)^{1/2}}.$$

Interestingly, compared with similar bounds above, it can be checked that

$$\frac{1-\rho}{2(1+\rho)} \leq \frac{1-\rho^2}{1+\rho^2+(1+2\rho+5\rho^2)^{1/2}} \leq \frac{1-\rho}{1+\rho}.$$

7. HIGH CONCENTRATION LIMIT OF DENSITY (§4.1 IN ARTICLE)

Theorem S1. *Let $\Theta \sim \mathrm{G}(\mu,\gamma,\rho,\lambda)$. Then, for $\gamma \simeq 1$ so that by dint of $\rho > 2\gamma-1$, $\rho \simeq 1$, the density g_Φ of the distribution of $\Phi = \rho^{1/2}(1-\rho)^{-1}(\Theta-\mu-\lambda)$ can be approximated around $\phi \simeq 0$ by the Cauchy density:*

$$g_\Phi(\phi) \simeq \frac{1}{\pi}\frac{1}{1+\phi^2}.$$

Proof. From (3) in the paper, the density of Θ can be expressed as

$$g(\theta) = \frac{1}{2\pi}\frac{1+\rho^2-2\rho\gamma\cos\lambda+2(\gamma\cos\lambda-\rho)\cos\tilde{\theta}-2\gamma\sin\lambda\sin\tilde{\theta}}{1+\rho^2-2\rho\cos\tilde{\theta}},$$

where $\tilde{\theta} = \theta-\mu-\lambda$. If follows from $\rho \simeq 1$ and $\phi \simeq 0$ that $\cos\tilde{\theta} \simeq 1-\tilde{\theta}^2/2$ and $\sin\tilde{\theta} \simeq \tilde{\theta}$. Then

$$\begin{aligned} g(\theta) &\simeq \frac{1}{2\pi}\frac{(1-\rho)^2+2(1-\rho)\gamma\cos\lambda+(\rho-\gamma\cos\lambda)\tilde{\theta}^2-2\gamma\sin\lambda\,\tilde{\theta}}{(1-\rho)^2+\rho\tilde{\theta}^2} \\ &= \left\{1+\frac{2\gamma\cos\lambda}{1-\rho}+\left(1-\frac{\gamma\cos\lambda}{\rho}\right)\left(\frac{\rho^{1/2}}{1-\rho}\,\tilde{\theta}\right)^2\right. \\ &\qquad \left.-\frac{2\gamma\sin\lambda}{\rho^{1/2}(1-\rho)}\left(\frac{\rho^{1/2}}{1-\rho}\tilde{\theta}\right)\right\}\Bigg/\left[2\pi\left\{1+\left(\frac{\rho^{1/2}}{1-\rho}\,\tilde{\theta}\right)^2\right\}\right], \quad \theta \simeq \mu+\lambda. \end{aligned}$$

Then the density of Φ can be approximated around $\phi \simeq 0$ as

$$g_\Phi(\phi) \simeq \frac{1}{2\pi}\frac{1-\rho+2\gamma\cos\lambda+(\rho-\gamma\cos\lambda)(1-\rho)\rho^{-1}\phi^2-2\gamma\sin\lambda\rho^{-1/2}\phi}{1+\phi^2}\cdot\frac{1}{\rho^{1/2}}.$$

Theorem 1 of the paper also implies that, as $\rho \to 1$, $\lambda \to 0$ and, in particular, Remark 1 of the paper shows that

$$2\gamma\cos\lambda-\rho \geq (2\gamma-1)/\rho \simeq 1.$$

It follows that

$$g_\Phi(\phi) \simeq \frac{1}{2\pi}\frac{2}{1+\phi^2} = \frac{1}{\pi}\frac{1}{1+\phi^2}, \quad \phi \simeq 0,$$

as claimed.

8. THE SINE-SKEWED WRAPPED CAUCHY DISTRIBUTION SUBMODEL (§4.1 IN ARTICLE)

In this subsection, we consider the sine-skewed wrapped Cauchy distribution (Umbach & Jammalamadaka, 2009, Abe & Pewsey, 2011) as a three-parameter submodel of our model. In the original parametrization of density (3), take $\rho = \gamma\cos\lambda$, $-\pi/2 \leq \lambda \leq \pi/2$, so that $0 \leq \rho \leq \gamma \leq (1+\rho^2)/2$ and $\bar{\alpha}_2 = \rho^2$. This results in

$$g_{\mathrm{SSWC}}(\theta) = \left\{1+\check{\lambda}\sin(\theta-\mu-\lambda)\right\}\frac{1}{2\pi}\frac{1-\rho^2}{1+\rho^2-2\rho\cos(\theta-\mu-\lambda)}, \quad -\pi \leq \theta < \pi,$$

where $-1 \leq \check{\lambda} = -2\,\mathrm{sign}(\lambda)(\gamma^2-\rho^2)^{1/2}/(1-\rho^2) \leq 1$. This distribution, with arbitrary centring, has been proposed as an alternative three-parameter skew extension of the wrapped Cauchy distribution, $\check{\lambda}$ being its skewness parameter. It is unimodal, a property far from assured for sine-skewed distributions in general (Abe & Pewsey, 2011).

The skewness associated with this distribution is $\bar{\beta}_2 = \mathrm{sign}(\lambda)\rho(\gamma^2-\rho^2)^{1/2}$. When $\gamma \leq 1/2$, the parameters of the sine-skewed wrapped Cauchy distribution are such that

$$(\bar{\alpha}_2-\gamma^2/2)^2+\bar{\beta}_2^2 = \gamma^4/4;$$

inter alia, the maximum absolute skewness is $\gamma^2/2$, which is very small compared with $\gamma(1-\gamma)$. When $\gamma > 1/2$, the additional constraint $(\bar{\alpha}_2-\gamma^2)^2+\bar{\beta}_2^2 \leq \gamma^2(1-\gamma)^2$ comes into play: the

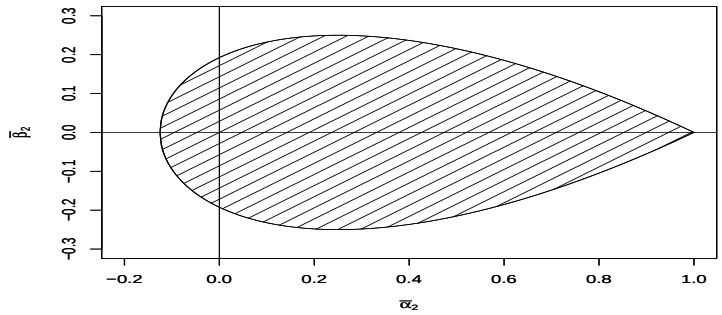


Fig. S2. $\cup D_\gamma$, $0 \le \gamma < 1$.

maximum absolute skewness remains $\gamma^2/2$ for $\gamma \le 2 - 2^{1/2}$ and then reduces to $(1-\gamma)(2\gamma - 1)^{1/2}$. The maximal skewness of this submodel does not nearly approach the maximal skewness of the full model.

As with any sine-skewed circular distribution, this special case of the full model allows random variate generation without rejection. The specific algorithm for g_{SSWC} centred at 0 is to set Φ equal to a wrapped Cauchy random variate $W = w$ with probability

$$\frac{1 + \check{\lambda} \sin w}{1 + |\check{\lambda}|}$$

and to set $\Phi = -w$ otherwise.

9. Mixture interpretation of g_S (§4.2 in article)

The three-parameter symmetric submodel with density g_S given by (6) in the article can be written as a mixture of the uniform distribution on the circle and another symmetric circular distribution, namely,

$$g_S(\theta) = (1-p)\frac{1}{2\pi} + p\,\frac{1}{2\pi}\frac{(1-\rho)\{1+\cos(\theta-\mu)\}}{1+\rho^2-2\rho\cos(\theta-\mu)}$$

where $p = 2\gamma/(1+\rho)$ and $\rho = \bar{\alpha}_2/\gamma$. Here, the non-uniform component is the rather obscure "unit-cosine-expanded-symmetric" wrapped Cauchy distribution with density $(1 \pm \cos\theta) g_{\mathrm{WC}}(\theta)/(1 \pm \rho)$, which has mean resultant length $(1 \pm \rho)/2$. In the mixture distribution, positive kurtosis is therefore increased partly by decreasing the weight of the uniform component and partly by increasing the concentration of the non-uniform component.

10. Union of parameter domains D_γ, $0 \le \gamma < 1$ (§4.3 in article)

As mentioned in §4.3 of the article, the union of the parameter domains D_γ for $0 \le \gamma < 1$ is the set

$$|\bar{\beta}_2| \le (1 + 8\bar{\alpha}_2)^{1/2}(1 - \bar{\alpha}_2)/27^{1/2}.$$

This exhibits the teardrop shape shown in Figure S2.

11. ASYMPTOTIC JOINT DISTRIBUTION OF METHOD OF MOMENTS ESTIMATORS (§5.1 IN ARTICLE)

The following theorem follows from manipulations based on Theorem 1 of Pewsey (2004).

Theorem S2. *Let* $\overline{D}_\gamma = \{(x,y) \in \mathbb{R}^2\,;\,(x-\gamma^2)^2 + y^2 = \gamma^2(1-\gamma)^2\}$. *If* $(\bar\alpha_2, \bar\beta_2) \in D_\gamma \setminus \overline{D}_\gamma$, *the large-sample asymptotic joint distribution of* $(\hat\mu, \hat\gamma, \hat\alpha_2, \hat\beta_2)^T$ *has the four-dimensional normal distribution with mean vector* ξ *and covariance matrix* Σ *where, to* $O(n^{-3/2})$,

$$\xi = \left(\mu - \frac{\bar\beta_2}{2n\gamma^2},\ \gamma + \frac{1-\bar\alpha_2}{4n\gamma},\ \bar\alpha_2 + \frac{1}{n}\left(1 - \frac{\bar\alpha_2}{\gamma^2}\right),\ \bar\beta_2 + \frac{\bar\beta_2}{n\gamma^2}\left(-1 - 2\bar\alpha_2 + \frac{2\bar\alpha_2}{\gamma^2}\right)\right)^T,$$

and $2n\gamma^2\Sigma = (v_{jk}),\ j,k = 1,\ldots,4$, *where*

$$\begin{aligned}
&v_{11} = 1 - \bar\alpha_2, \quad v_{12} = v_{21} = \gamma\bar\beta_2, \quad v_{13} = v_{31} = 2\bar\beta_2, \quad v_{14} = v_{41} = \gamma^2 - 2\bar\alpha_2 + \bar\alpha_2^2 + \bar\beta_2^2,\\
&v_{22} = \gamma^2(1 - 2\gamma^2 + \bar\alpha_2), \quad v_{23} = v_{32} = \gamma^3(1 - 2\bar\alpha_2) + \gamma(\bar\alpha_2^2 + \bar\beta_2^2), \quad v_{24} = v_{42} = -2\gamma^3\bar\beta_2,\\
&v_{33} = \gamma^2(1 - 2\bar\alpha_2^2) + \bar\alpha_2^3 + \bar\alpha_2\bar\beta_2^2 + 4\bar\beta_2^2,\\
&v_{34} = v_{43} = 2\gamma^2(1-\bar\alpha_2)\bar\beta_2 + \bar\beta_2(-4\bar\alpha_2 + \bar\alpha_2^2 + \bar\beta_2^2),\\
&v_{44} = \gamma^2(1 - 4\bar\alpha_2 - 2\bar\beta_2^2) + \bar\alpha_2(4\bar\alpha_2 - \bar\alpha_2^2 - \bar\beta_2^2).
\end{aligned}$$

If $(\bar\alpha_2, \bar\beta_2) \in \overline{D}_\gamma$, *the joint distribution of* $(\hat\mu, \hat\gamma, \hat\alpha_2, \hat\beta_2)^T$ *converges to the distribution of* N_B *with probability* p *and* N_I *otherwise. Here,* $p = \int_\Omega \phi(x_1,x_2,x_3,x_4)dx_1dx_2dx_3dx_4$, $\Omega = \{(x_1,x_2,x_3,x_4) \in \mathbb{R}^4\,;\,\bar\alpha_2 x_3 + \bar\beta_2 x_4 - \bar\alpha_2^2 - \bar\beta_2^2 > 0\}$, N_B *is a vector defined by* $N_B = (Y_1, Y_2, \bar\alpha_2, \bar\beta_2)^T$, *where* (Y_1, Y_2) *has the bivariate normal distribution with mean vector* (μ,γ) *and covariance matrix* $\tilde\Sigma = (v_{jk}/(2n\gamma^2)),\ j,k = 1,2$, *and* N_I *follows the four-dimensional normal distribution with mean vector* ξ *and covariance matrix* Σ *given above truncated to* $\mathbb{R}^4 \setminus \Omega$.

12. FISHER INFORMATION (§5.2 IN ARTICLE)

12·1. *Observed Fisher information*

To express the elements of the observed Fisher information matrix in a unified manner, write density (3) in the article as

$$f(\theta) = \frac{1}{2\pi}\frac{a + b\cos(\theta - \eta)}{1 + \rho^2 - 2\rho\cos(\theta - \lambda)},$$

where $a = 1 + \gamma^{-2}(\bar\alpha_2^2 + \bar\beta_2^2) - 2\bar\alpha_2$, $b = 2\{\gamma^2 + \gamma^{-2}(\bar\alpha_2 + \bar\beta_2) - 2\bar\alpha_2\}^{1/2}$, $\rho = \gamma^{-1}(\bar\alpha_2^2 + \bar\beta_2^2)^{1/2}$, $\eta = \mu + \arg(\bar\alpha_2 - \gamma^2 + i\bar\beta_2)$, and $\lambda = \mu + \arg(\bar\alpha_2 + i\bar\beta_2)$. Elements of the observed Fisher information matrix are minus the second derivatives of ℓ. Denote each element of the observed information matrix by $J_{xy} = -\sum_{j=1}^n \partial^2 \log f(\theta_j)/\partial x \partial y$ for (x,y) taken from $(\mu, \gamma, \bar\alpha_2, \bar\beta_2)$. Then J_{xy} has the expression

$$\begin{aligned}
J_{xy} = \sum_{j=1}^n \Bigg(&\frac{F_0 + F_{c1}\cos(\theta_j - \eta) + F_{s1}\sin(\theta_j - \eta) + F_{c2}\cos\{2(\theta_j - \eta)\} + F_{s2}\sin\{2(\theta_j - \eta)\}}{\{a + b\cos(\theta_j - \eta)\}^2}\\
&+\big[A_0(\rho,\lambda) + A_{c1}(\rho,\lambda)\cos(\theta_j - \lambda) + A_{s1}(\rho,\lambda)\sin(\theta_j - \lambda) + A_{c2}(\rho,\lambda)\cos\{2(\theta_j - \lambda)\}\\
&+A_{s2}(\rho,\lambda)\sin\{2(\theta_j - \lambda)\}\big] \big/ \left\{1 + \rho^2 - 2\rho\cos(\theta_j - \lambda)\right\}^2\Bigg),
\end{aligned}$$

where

$$
\begin{aligned}
F_0 &= aa_{xy} - a_x a_y + (bb_{xy} - b_x b_y)/2 - b^2\eta_x\eta_y,\\
F_{c1} &= a_{xy}b + ab_{xy} - a_x b_y - a_y b_x - ab\eta_x\eta_y,\\
F_{s1} &= a(b_x\eta_y + b_y\eta_x + b\eta_{xy}) - b(a_x\eta_y + a_y\eta_x)\\
F_{c2} &= (bb_{xy} - b_x b_y)/2, \quad F_{s2} = b^2\eta_{xy}/2,\\
A_0(\rho,\lambda) &= \rho^2(\rho\rho_{xy} - \rho_x\rho_y) + 2\rho(\rho_{xy} - \lambda_x\lambda_y), \qquad \text{(S2)}\\
A_{c1}(\rho,\lambda) &= (1+\rho^2)(\rho\lambda_x\lambda_y - \rho_{xy}) + 2\rho(\rho_x\rho_y - \rho\rho_{xy}), \qquad \text{(S3)}\\
A_{s1}(\rho,\lambda) &= -(1-\rho^2)(\rho_x\lambda_y + \lambda_x\rho_y + \rho\lambda_{xy}) - 2\rho^3\lambda_{xy}, \qquad \text{(S4)}\\
A_{c2}(\rho,\lambda) &= \rho\rho_{xy} - \rho_x\rho_y, \quad A_{s2}(\rho,\lambda) = \rho^2\lambda_{xy}, \qquad \text{(S5)}
\end{aligned}
$$

and the first and second derivatives of a, b, η, ρ and λ, denoted by $a_\mu, a_{\mu\gamma}$, and so on, are

$$
\begin{aligned}
a_\mu &= 0, \quad a_\gamma = -2\gamma^{-3}(\bar\alpha_2^2 + \bar\beta_2^2), \quad a_{\bar\alpha_2} = 2\gamma^{-2}\bar\alpha_2 - 2, \quad a_{\bar\beta_2} = 2\gamma^{-2}\bar\beta_2,\\
a_{\mu\mu} &= a_{\mu\gamma} = a_{\mu\bar\alpha_2} = a_{\mu\bar\beta_2} = 0, \quad a_{\gamma\gamma} = 6\gamma^{-4}(\bar\alpha_2^2 + \bar\beta_2^2), \quad a_{\gamma\bar\alpha_2} = -4\gamma^{-3}\bar\alpha_2,\\
a_{\gamma\bar\beta_2} &= -4\gamma^{-3}\bar\beta_2, \quad a_{\bar\alpha_2\bar\alpha_2} = 2\gamma^{-2}, \quad a_{\bar\alpha_2\bar\beta_2} = 0, \quad a_{\bar\beta_2\bar\beta_2} = 2\gamma^{-2},\\
b_\mu &= 0, \quad b_\gamma = 4b^{-1}\gamma^{-3}(\gamma^4 - \bar\alpha_2^2 - \bar\beta_2^2), \quad b_{\bar\alpha_2} = 4b^{-1}(\gamma^{-2}\bar\alpha_2 - 1), \quad b_{\bar\beta_2} = 4b^{-1}\gamma^{-2}\beta,\\
b_{\mu\mu} &= b_{\mu\gamma} = b_{\mu\bar\alpha_2} = b_{\mu\bar\beta_2} = 0,\\
b_{\gamma\gamma} &= 8\sqrt{2}b^{-3/2}\gamma^{-6}\left\{(\bar\alpha_2^2 + \bar\beta_2^2)^2 - 3\gamma^2\bar\alpha_2(\bar\alpha_2^2 + \bar\beta_2^2) + 3\gamma^4(\bar\alpha_2^2 + \bar\beta_2^2) - \gamma^6\bar\alpha_2\right\},\\
b_{\gamma\bar\alpha_2} &= -4\sqrt{2}b^{-3/2}\gamma^{-5}\left\{\bar\alpha_2(\bar\alpha_2^2 + \bar\beta_2^2) - 3\gamma^2\bar\alpha_2^2 + \gamma^2\bar\beta_2^2 + 3\gamma^4\bar\alpha_2 - \gamma^6\right\},\\
b_{\gamma\bar\beta_2} &= -4\sqrt{2}b^{-3/2}\gamma^{-5}\bar\beta_2\left\{\bar\alpha_2^2 + \bar\beta_2^2 - 4\gamma^2\bar\alpha_2 + 3\gamma^4\right\}, \quad b_{\bar\alpha_2\bar\alpha_2} = 4\sqrt{2}b^{-3/2}\gamma^{-4}\bar\beta_2^2,\\
b_{\bar\alpha_2\bar\beta_2} &= -4\sqrt{2}b^{-3/2}\gamma^{-4}\bar\beta_2(\bar\alpha_2 - \gamma^2), \quad b_{\bar\beta_2\bar\beta_2} = 4\sqrt{2}b^{-3/2}\gamma^{-4}(\bar\alpha_2 - \gamma^2)^2,
\end{aligned}
$$

$$
\begin{aligned}
\eta_\mu &= 1, \quad \eta_\gamma = \frac{2\gamma\bar\beta_2}{(\bar\alpha_2 - \gamma^2)^2 + \bar\beta_2^2}, \quad \eta_{\bar\alpha_2} = -\frac{\bar\beta_2}{(\bar\alpha_2 - \gamma^2)^2 + \bar\beta_2^2},\\
\eta_{\bar\beta_2} &= \frac{\bar\alpha_2 - \gamma^2}{(\bar\alpha_2 - \gamma^2)^2 + \bar\beta_2^2}, \quad \eta_{\mu\mu} = \eta_{\mu\gamma} = \eta_{\mu\bar\alpha_2} = \eta_{\mu\bar\beta_2} = 0,\\
\eta_{\gamma\gamma} &= \frac{2\bar\beta_2(\bar\alpha_2^2 + \bar\beta_2^2 + 2\gamma^2\bar\alpha_2 - 3\gamma^4)}{\{(\bar\alpha_2 - \gamma^2)^2 + \bar\beta_2^2\}^2}, \quad \eta_{\gamma\bar\alpha_2} = -\frac{4\gamma\beta(\bar\alpha_2 - \gamma^2)}{\{(\bar\alpha_2 - \gamma^2)^2 + \bar\beta_2^2\}^2},\\
\eta_{\gamma\bar\beta_2} &= \frac{2\gamma\{(\bar\alpha_2 - \gamma^2)^2 - \bar\beta_2^2\}}{\{(\bar\alpha_2 - \gamma^2)^2 + \bar\beta_2^2\}^2}, \quad \eta_{\bar\alpha_2\bar\alpha_2} = \frac{2\bar\beta_2(\bar\alpha_2 - \gamma^2)}{\{(\bar\alpha_2 - \gamma^2)^2 + \bar\beta_2^2\}^2},\\
\eta_{\bar\alpha_2\bar\beta_2} &= -\frac{(\bar\alpha_2 - \gamma^2)^2 - \bar\beta_2^2}{\{(\bar\alpha_2 - \gamma^2)^2 + \bar\beta_2^2\}}, \quad \eta_{\bar\beta_2\bar\beta_2} = -\frac{2\bar\beta_2(\bar\alpha_2 - \gamma^2)}{\{(\bar\alpha_2 - \gamma^2)^2 + \bar\beta_2^2\}^2},
\end{aligned}
$$

$$\rho_\mu = 0, \quad \rho_\gamma = -\gamma^{-2}(\bar\alpha_2+\bar\beta_2)^{1/2}, \quad \rho_{\bar\alpha_2} = \frac{\bar\alpha_2}{\gamma(\bar\alpha_2^2+\bar\beta_2^2)^{1/2}}, \quad \rho_{\bar\beta_2} = \frac{\bar\beta_2}{\gamma(\bar\alpha_2^2+\bar\beta_2^2)^{1/2}},$$
$$\rho_{\mu\mu} = \rho_{\mu\gamma} = \rho_{\mu\bar\alpha_2} = \rho_{\mu\bar\beta_2} = 0, \quad \rho_{\gamma\gamma} = 2\gamma^{-3}(\bar\alpha_2^2+\bar\beta_2^2)^{1/2}, \quad \rho_{\gamma\bar\alpha_2} = -\frac{\bar\alpha_2}{\gamma^2(\bar\alpha_2^2+\bar\beta_2^2)^{1/2}},$$
$$\rho_{\gamma\bar\beta_2} = -\frac{\bar\beta_2}{\gamma^2(\bar\alpha_2^2+\bar\beta_2^2)^{1/2}}, \quad \rho_{\bar\alpha_2\bar\alpha_2} = \frac{\bar\beta_2^2}{\gamma(\bar\alpha_2^2+\bar\beta_2^2)^{3/2}}, \quad \rho_{\bar\alpha_2\bar\beta_2} = -\frac{\bar\alpha_2\bar\beta_2}{\gamma(\bar\alpha_2^2+\bar\beta_2^2)^{3/2}},$$
$$\rho_{\bar\beta_2\bar\beta_2} = \frac{\bar\alpha_2^2}{\gamma(\bar\alpha_2^2+\bar\beta_2^2)^{3/2}}, \quad \lambda_\mu = 1, \quad \lambda_\gamma = 0, \quad \lambda_{\bar\alpha_2} = -\frac{\bar\beta_2}{\bar\alpha_2^2+\bar\beta_2^2}, \quad \lambda_{\bar\beta_2} = \frac{\bar\alpha_2}{\bar\alpha_2^2+\bar\beta_2^2},$$
$$\lambda_{\mu\mu} = \lambda_{\mu\gamma} = \lambda_{\mu\bar\alpha_2} = \lambda_{\mu\bar\beta_2} = \lambda_{\gamma\gamma} = \lambda_{\gamma\bar\alpha_2} = \lambda_{\gamma\bar\beta_2} = 0, \quad \lambda_{\bar\alpha_2\bar\alpha_2} = \frac{2\bar\alpha_2\bar\beta_2}{(\bar\alpha_2^2+\bar\beta_2^2)^2},$$
$$\lambda_{\bar\alpha_2\bar\beta_2} = -\frac{\bar\alpha_2^2-\bar\beta_2^2}{(\bar\alpha_2^2+\bar\beta_2^2)^2}, \quad \lambda_{\bar\beta_2\bar\beta_2} = -\frac{2\bar\alpha_2\bar\beta_2}{(\bar\alpha_2^2+\bar\beta_2^2)^2}.$$

12·2. *Expected Fisher information*

For the calculation of the expected Fisher information matrix, it is convenient to change the expression of the density to

$$f(\theta) = \frac{c}{2\pi}\frac{1+\xi^2-2\xi\cos(\theta-\eta)}{1+\rho^2-2\rho\cos(\theta-\lambda)}, \quad -\pi \le \theta < \pi,$$

where $c = \{a+(a^2-b^2)^{1/2}\}/2$ and $\xi = \{a-(a^2-b^2)^{1/2}\}/b$. It follows from $a \ge b \ge 0$ that $0 \le \xi < 1$.

With this representation of the density, the expected Fisher information can be expressed in closed form without using any special functions or infinite sums. Denote n^{-1} times the expected information matrix by $\iota = (\iota_{xy})$, where (x, y) are taken from $(\mu, \gamma, \bar\alpha_2, \bar\beta_2)$. Then the element of the expected information matrix is of the form

$$\iota_{xy} = -C(x,y) - \Xi(x,y) + Q(x,y),$$

where

$$C(x,y) = \frac{\partial^2}{\partial x\partial y}\log c = \frac{c_{xy}c - c_x c_y}{c^2},$$
$$\Xi(x,y) = E\left\{\frac{\partial^2}{\partial x\partial y}\log\{1+\xi^2-2\xi\cos(\Theta-\eta)\}\right\}$$
$$= 2c\left\{A_0(\xi,\eta)I_0 + A_{c1}(\xi,\eta)I_{c1} + A_{s1}(\xi,\eta)I_{s1} + A_{c2}(\xi,\eta)I_{c2} + A_{s2}(\xi,\eta)I_{s2}\right\},$$

$$Q(x,y) = E\left\{\frac{\partial^2}{\partial x\partial y}\log\{1+\rho^2-2\rho\cos(\Theta-\lambda)\}\right\}$$
$$= 2c\left\{B_0J_0 + B_1J_1 + B_2J_2 + B_3J_3\right\},$$

where A's with subscripts are as in (S2)–(S5) and

$$
\begin{aligned}
c_x &= \frac{\partial c}{\partial x} = \frac{1}{2}\left(a_x + \frac{aa_x - bb_x}{(a^2-b^2)^{1/2}}\right),\\
c_{xy} &= \frac{\partial^2 c}{\partial x \partial y} = \frac{1}{2}\left\{a_{xy} + \frac{(aa_{xy} - bb_{xy})(a^2-b^2) - (a_x b - ab_x)(a_y b - ab_y)}{(a^2-b^2)^{3/2}}\right\},\\
B_0 &= (1+\xi^2)A_0(\rho,\lambda) - \xi\cos(\eta-\lambda)A_{c1}(\rho,\lambda) - \xi\sin(\eta-\lambda)A_{s1}(\rho,\lambda),\\
B_1 &= -2\xi\cos(\eta-\lambda)A_0(\rho,\lambda) + (1+\xi^2)A_{c1}(\rho,\lambda) - \xi\cos(\eta-\lambda)A_{c2}(\rho,\lambda)\\
&\quad -\xi\sin(\eta-\lambda)A_{s2}(\rho,\lambda),\\
B_2 &= -\xi\cos(\eta-\lambda)A_{c1}(\rho,\lambda) + \xi\sin(\eta-\lambda)A_{s1}(\rho,\lambda) + (1+\xi^2)A_{c2}(\rho,\lambda),\\
B_3 &= -\xi\cos(\eta-\lambda)A_{c2}(\rho,\lambda) + \xi\sin(\eta-\lambda)A_{s2}(\rho,\lambda),\\
I_0 &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{1+\xi^2-2\xi\cos(\theta-\eta)}\frac{1}{1+\rho^2-2\rho\cos(\theta-\lambda)}\,d\theta\\
&= \frac{1-\xi^2\rho^2}{(1-\xi^2)(1-\rho^2)\{1+\xi^2\rho^2-2\xi\rho\cos(\lambda-\eta)\}},\\
I_{c1} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\cos(\theta-\eta)}{1+\xi^2-2\xi\cos(\theta-\eta)}\frac{1}{1+\rho^2-2\rho\cos(\theta-\lambda)}\,d\theta\\
&= \frac{(1-\rho^2)\xi + (1-\xi^2)\rho\cos(\lambda-\eta)}{(1-\xi^2)(1-\rho^2)\{1+\xi^2\rho^2-2\xi\rho\cos(\lambda-\eta)\}},\\
I_{s1} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\sin(\theta-\eta)}{1+\xi^2-2\xi\cos(\theta-\eta)}\frac{1}{1+\rho^2-2\rho\cos(\theta-\lambda)}\,d\theta\\
&= \frac{\rho\sin(\lambda-\eta)}{(1-\rho^2)\{1+\xi^2\rho^2-2\xi\rho\cos(\lambda-\eta)\}},\\
I_{c2} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\cos\{2(\theta-\eta)\}}{1+\xi^2-2\xi\cos(\theta-\eta)}\frac{1}{1+\rho^2-2\rho\cos(\theta-\lambda)}\,d\theta\\
&= \frac{(1-\rho^2)\xi^2 + (1-\xi^2)\rho^2\cos\{2(\lambda-\eta)\} + (1-\xi^2)(1-\rho^2)\xi\rho\cos(\lambda-\eta)}{(1-\xi^2)(1-\rho^2)\{1+\xi^2\rho^2-2\xi\rho\cos(\lambda-\eta)\}},\\
I_{s2} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\sin\{2(\theta-\eta)\}}{1+\xi^2-2\xi\cos(\theta-\eta)}\frac{1}{1+\rho^2-2\rho\cos(\theta-\lambda)}\,d\theta\\
&= \frac{\rho^2\sin\{2(\lambda-\eta)\} + (1-\rho^2)\xi\rho\sin(\lambda-\eta)}{(1-\rho^2)\{1+\xi^2\rho^2-2\xi\rho\cos(\lambda-\eta)\}},
\end{aligned}
$$

$$
\begin{aligned}
J_0 &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{\{1+\rho^2-2\rho\cos(\theta-\lambda)\}^3}\,d\theta = \frac{1+4\rho^2+\rho^4}{(1-\rho^2)^5},\\
J_1 &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\cos(\theta-\lambda)}{\{1+\rho^2-2\rho\cos(\theta-\lambda)\}^3}\,d\theta = \frac{3\rho(1+\rho^2)}{(1-\rho^2)^5},\\
J_2 &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\cos\{2(\theta-\lambda)\}}{\{1+\rho^2-2\rho\cos(\theta-\lambda)\}^3}\,d\theta = \frac{6\rho^2}{(1-\rho^2)^5},\\
J_3 &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\cos\{3(\theta-\lambda)\}}{\{1+\rho^2-2\rho\cos(\theta-\lambda)\}^3}\,d\theta = \frac{\rho^3(10-5\rho^2+\rho^4)}{(1-\rho^2)^5}.
\end{aligned}
$$

The results for I's with subscripts are obvious from Theorem 4 of Kato and Jones (2013). As for the integration for J's with subscripts, one can apply the residue theorem (e.g. Rudin, 1987, Theorem 10.42).

13. SIMULATION STUDY (§5.3 IN ARTICLE)

We compare the performance of the method of moments and maximum likelihood estimators for finite sample sizes via a Monte Carlo simulation study. We generated random samples of sizes $n = 10$, 25, 50, 200 and 1000 from the proposed family (4) with some selected values of the parameters. For each combination of sample size and parameter, $r = 2000$ samples were generated using the acceptance-rejection algorithm of §6 of the Supplementary Material.

For overall comparison of the two estimators, we discuss the generalized mean squared error which is defined by $\det(\Sigma)$, where $\Sigma = E\{(\hat{\xi} - \xi)(\hat{\xi} - \xi)^T\}$ and $\hat{\xi}$ is an estimator of $\xi = (\mu, \gamma, \bar{\alpha}_2, \bar{\beta}_2)^T$. An estimate of the generalized mean squared error is given by replacing Σ by its sample analogue $\hat{\Sigma} = r^{-1} \sum_{j=1}^{r} (\hat{\xi}_j - \xi)(\hat{\xi}_j - \xi)^T$, where the $\hat{\xi}_j$'s, $j = 1, \ldots, r$, are the estimates from the r simulation samples. We then consider the estimated relative generalized mean squared error of the method of moments estimator with respect to the maximum likelihood estimator, which is defined as

$$\widehat{\text{RGMSE}} = \frac{\det(\hat{\Sigma}_{MM})}{\det(\hat{\Sigma}_{ML})},$$

where $\hat{\Sigma}_{MM}$ and $\hat{\Sigma}_{ML}$ are sample estimates of the mean squared error matrices of the method of moments estimator and maximum likelihood estimator, respectively. Table S1 shows $\widehat{\text{RGMSE}}$ for selected combinations of sample size and parameters. The values for $n = \infty$ are the asymptotic relative generalized mean squared error defined by $\text{RGMSE} = \lim_{n\to\infty} \det(\Sigma_{MM}) / \det(\Sigma_{ML})$, where Σ_{MM} and Σ_{ML} are the theoretical mean squared error matrices of the method of moments estimator and maximum likelihood estimator, respectively. The values of RGMSE were calculated using the expected Fisher information matrix given in §12·2 of this Supplementary Material.

The main message from Table S1 is that the method of moments estimator is preferable for combinations of small to moderate n and values of each of γ, $\bar{\alpha}_2$ and $\bar{\beta}_2$ which are far from the boundary of the parameter space. Otherwise, the maximum likelihood estimator shows better performance than the method of moments estimator, as it must of course, for large n.

In a little more detail, Table S1(a) shows results for the proposed model with similar shapes of the densities. Here, for relatively small n, the performance of the method of moments estimator is generally better than that of the maximum likelihood estimator except for the case with low concentration, $\gamma = 0.14$. Table S1(b) exhibits estimates of the RGMSE of three-parameter symmetric submodels. The maximum likelihood estimator outperforms the method of moments estimator for large $\bar{\alpha}_2$, that is, peaked densities, for all n. For larger n, the the maximum likelihood estimator also exhibits considerably smaller generalised mean squared error than the method of moments estimator for low $\bar{\alpha}_2$, that is, wide-bodied densities. Estimates of the RGMSE for the three-parameter asymmetric case of §5.4 of the article are displayed in Table S1(c). Again, as well as for larger n, it is for more extreme parameter values, here skewness $\bar{\beta}_2$, that the method of moment estimator particularly loses out to the maximum likelihood estimator.

As seen in Table S1, the values of the relative generalized mean squared error for $n = \infty$, as well as for finite n, are quite different depending on the values of the parameters. In this section we consider how the parameters influence the values of the asymptotic RGMSE through some

Table S1. *Relative generalized mean squared error of the method of moments estimator with respect to the maximum likelihood estimator as estimated from a Monte Carlo study of 2000 samples of size n from the family (4) with $\mu = 0$ and (a) $\bar{\alpha}_2 = 0.4\gamma\cos(\pi/4)$, $\bar{\beta}_2 = 0.4\gamma\sin(\pi/4)$ and four selected values of γ, (b) $\gamma = 0.3$, $\bar{\beta}_2 = 0$ and four selected values of $\bar{\alpha}_2$, and (c) $\gamma = 0.3$, $\bar{\alpha}_2 = 0.09$ and four selected values of $\bar{\beta}_2$.*

(a)

γ	$n=10$	$n=25$	$n=50$	$n=200$	$n=1000$	$n=\infty$
0.14	2.43	1.34	1.33	1.11	1.45	2.98
0.28	0.805	0.506	0.483	1.10	2.14	2.47
0.42	0.288	0.307	0.474	1.62	2.22	2.34
0.56	0.111	0.458	1.31	2.70	4.06	4.64

(b)

$\bar{\alpha}_2$	$n=10$	$n=25$	$n=50$	$n=200$	$n=1000$	$n=\infty$
-0.1	0.285	0.266	0.570	6.28	14.71	23.93
0.02	0.442	0.385	0.331	0.559	0.954	1.08
0.14	1.40	0.633	0.470	1.24	2.31	2.51
0.26	6.28	3.23	14.66	560.62	1128.91	1544.03

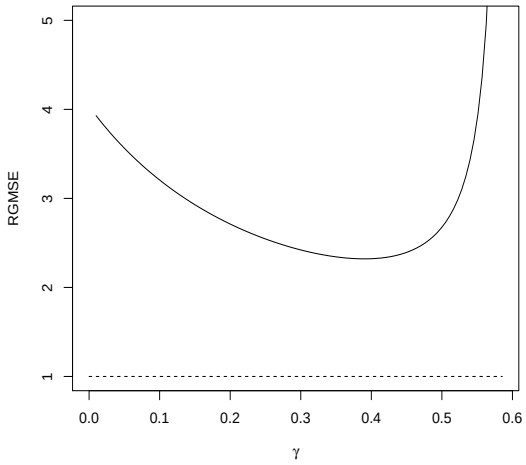
(c)

$\bar{\beta}_2$	$n=10$	$n=25$	$n=50$	$n=200$	$n=1000$	$n=\infty$
0.01	0.768	0.550	0.402	0.699	1.13	1.22
0.07	0.796	0.475	0.409	0.992	1.86	1.98
0.13	0.804	0.529	0.668	3.43	7.24	8.03
0.19	0.785	1.06	3.69	66.33	149.36	222.72

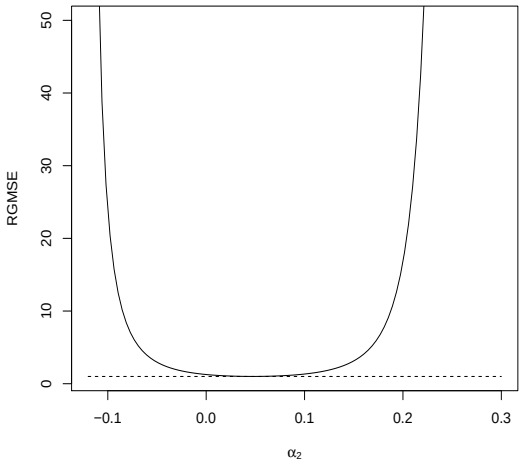
pictures of those values. The asymptotic RGMSE is shown as a function of γ, $\bar{\alpha}_2$ or $\bar{\beta}_2$ for fixed values of the other parameters in Figure S3.

All three solid curves in Figure S3 appear to be concave functions. Another common tendency is that the values of the asymptotic RGMSE are large around the boundary of the parameter space. The curve in panel (a) takes a minimum value around $\gamma = 0.39$. The minimum value of the curve in panel (b) seems to be at around $\bar{\alpha}_2 = 0.046$, but the values of the curve remain less than 1.1 when $\bar{\alpha}_2$ is in the interval between about 0.02 and 0.07. Panel (c) suggests that the RGMSE takes its minimum value at $\bar{\beta}_2 = 0$ and is monotonically increasing with respect to $\bar{\beta}_2 (> 0)$. Panels (b) and (c) imply that, as the parameter approaches the boundary of the parameter space, the value of the RGMSE tends to infinity.

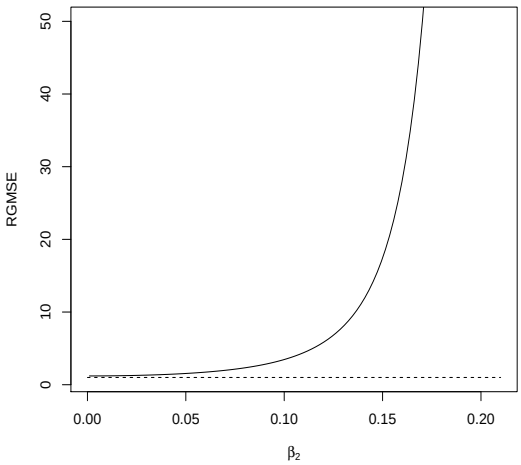
Table S2 seeks, in one particular quite high sample size case where maximum likelihood is generally to be preferred, to tease out whether there is any pattern to the way in which the method of moments tends to fail, by giving estimated mean squared errors, $r^{-1}\sum_{j=1}^{r}(\hat{\xi}_{kj} - \xi_k)^2$, for individual parameter estimators, where $\hat{\xi}_{kj}$ and ξ_k, $k = 1, \ldots, 4$, are the kth element of $\hat{\xi}_j$ and ξ, respectively. In most cases, of overall RGMSE preference for method of moments or relatively weakly for maximum likelihood, method of moments estimators of mean direction and mean resultant length remain slightly better than the maximum likelihood estimators. Where maximum likelihood 'takes over' is in its better estimation of the skewness and kurtosis parameters. For the more extreme parameter values, maximum likelihood also estimates the location and concentration better.



(a)



(b)



(c)

Fig. S3. Asymptotic relative generalized mean squared error (solid) of the method of moments estimators with respect to maximum likelihood estimators for the family in the article with $\mu = 0$ as a function of (a) γ for $\bar{\alpha}_2 = 0.4\gamma\cos(\pi/4)$ and $\bar{\beta}_2 = 0.4\gamma\sin(\pi/4)$, (b) $\bar{\alpha}_2$ for $\gamma = 0.3$ and $\bar{\beta}_2 = 0$, and (c) $\bar{\beta}_2 (> 0)$ for $\gamma = 0.3$ and $\bar{\alpha}_2 = 0.09$. The dashed lines represent the equation RGMSE $= 1$.

14. ANOTHER EXAMPLE (§5.3 IN ARTICLE)

As a second illustrative example, we consider a dataset recording the initial directions in which $n = 730$ ants moved in relation to a visual stimulus (Jander, 1957, Batschelet, 1981, Fig. 3.2.2(A)). The data were recorded in intervals of ten degrees with frequencies given to the nearest

Table S2. *Individual parameter mean squared errors, times $n = 200$, of the method of moments and maximum likelihood estimators as estimated from a Monte Carlo study of 10000 samples of size $n = 200$ from the family (4) with $\mu = 0$ and (a) $\bar{\alpha}_2 = 0.4\gamma\cos(\pi/4)$, $\bar{\beta}_2 = 0.4\gamma\sin(\pi/4)$ and four selected values of γ, (b) $\gamma = 0.3$, $\bar{\beta}_2 = 0$ and four selected values of $\bar{\alpha}_2$, and (c) $\gamma = 0.3$, $\bar{\alpha}_2 = 0.09$ and four selected values of $\bar{\beta}_2$.*

(a)

γ	RGMSE	$\hat{\mu}_{MM}$	$\hat{\mu}_{ML}$	$\hat{\gamma}_{MM}$	$\hat{\gamma}_{ML}$	$\hat{\alpha}_{2;MM}$	$\hat{\alpha}_{2;ML}$	$\hat{\beta}_{2;MM}$	$\hat{\beta}_{2;ML}$
0.14	1.11	31.04	36.35	0.481	0.565	0.559	0.386	0.508	0.484
0.28	1.10	6.33	6.54	0.450	0.510	0.635	0.567	0.478	0.451
0.42	1.62	2.50	2.50	0.405	0.406	0.651	0.554	0.391	0.302
0.56	2.70	1.33	1.27	0.254	0.231	0.548	0.453	0.274	0.181

(b)

$\bar{\alpha}_2$	RGMSE	$\hat{\mu}_{MM}$	$\hat{\mu}_{ML}$	$\hat{\gamma}_{MM}$	$\hat{\gamma}_{ML}$	$\hat{\alpha}_{2;MM}$	$\hat{\alpha}_{2;ML}$	$\hat{\beta}_{2;MM}$	$\hat{\beta}_{2;ML}$
−0.1	6.28	6.14	6.14	0.342	0.336	0.363	0.170	0.745	0.457
0.02	0.559	5.89	6.01	0.414	0.459	0.463	0.595	0.480	0.592
0.14	1.24	5.27	5.50	0.475	0.522	0.487	0.447	0.621	0.541
0.26	560.62	4.30	2.80	0.526	0.385	0.582	0.259	1.13	0.267

(c)

$\bar{\beta}_2$	RGMSE	$\hat{\mu}_{MM}$	$\hat{\mu}_{ML}$	$\hat{\gamma}_{MM}$	$\hat{\gamma}_{ML}$	$\hat{\alpha}_{2;MM}$	$\hat{\alpha}_{2;ML}$	$\hat{\beta}_{2;MM}$	$\hat{\beta}_{2;ML}$
0.01	0.699	5.50	5.62	0.455	0.502	0.491	0.553	0.505	0.567
0.07	0.992	5.24	5.42	0.455	0.514	0.594	0.581	0.479	0.445
0.13	3.43	5.22	5.23	0.415	0.418	0.797	0.503	0.449	0.296
0.19	66.33	5.39	4.11	0.461	0.356	1.09	0.430	0.398	0.091

five ants. Fig. S4 shows a linear histogram of the data converted to radians from 0 to 2π; the data are sharply peaked around the mode and flatly distributed over a wide range of the circle. Table S3 shows the maximum likelihood estimates of the parameters, the maximized log-likelihood, and information criteria values for the full model numbered (4) in the paper, the three-parameter symmetric submodel (6), the three-parameter asymmetric submodel (7) and the wrapped Cauchy and cardioid distributions. The last two, two-parameter, distributions, do not fit these data well, nor does the von Mises distribution; three- and four-parameter models are preferred for these data. The log-likelihood ratio test of model (4) against (6) yields a p-value of 0.041, marginally rejecting the null hypothesis that the skewness is zero; the full model is also preferred by AIC but not by BIC; χ^2 goodness-of-fit tests allow (4), p-value 0.111, and more marginally (6), p-value 0.053. We think this justifies modelling these data by (4); whether or not the reader agrees, the models with shape parameters afford principled discussion of the significance of the skewness, which is certainly small.

Table S3. *Maximum likelihood estimates of the parameters, the maximized log-likelihood, ℓ_{max}, and the values of Akaike Information Criterion, AIC, and Bayesian Information Criterion, BIC, for the full model and four of its submodels fitted to the ant data.*

Model	$\hat{\mu}$	$\hat{\gamma}$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\ell_{\max}$	AIC	BIC
Full model (4)	3.13	0.629	0.459	0.044	−2195.77	4399.54	4417.91
Three-parameter symmetric (6)	−3.10	0.627	0.461	(0)	−2197.85	4401.70	4415.48
Three-parameter asymmetric (7)	3.11	0.674	(0.454)	0.058	−2202.41	4410.82	4424.60
Wrapped Cauchy	−3.10	0.677	(0.458)	(0)	−2205.75	4415.50	4424.68
Cardioid	3.07	0.424	(0)	(0)	−2397.17	4798.34	4807.52

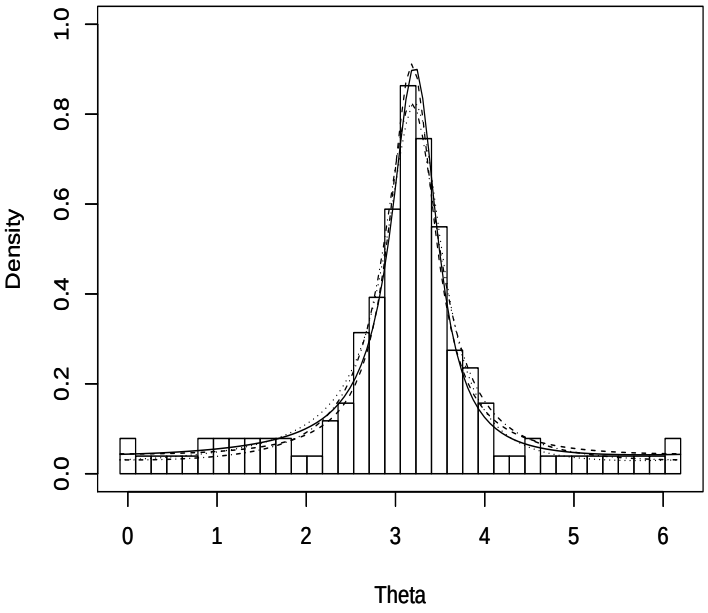


Fig. S4. Histogram of the ant data with maximum likelihood fits of the full model (solid), the three-parameter symmetric submodel (dashed), the three-parameter asymmetric submodel (dotted), and the wrapped Cauchy submodel (dot-dashed).

In this case, model (4) also fits the data better than the four-parameter direct Batschelet model, $\ell_{\max} = -2196.41$, indirect Batschelet model, $\ell_{\max} = -2197.14$, and wrapped stable model, $\ell_{\max} = -2199.30$. The fitted Batschelet distributions are skewed extensions of Jones & Pewsey's (2005) symmetric family. However, all fits are fairly similar and all but the last-mentioned competitor are marginally acceptable in goodness-of-fit terms. Model (4), as usual, wins in terms of interpretability and computational speed for grouped data, while if skewness is considered negligible, the mixture interpretation of (6) in §9 of this Supplementary Material yields an estimated proportion of $1 - \hat{p} = 1 - 2\hat{\gamma}^2/(\hat{\gamma} + \hat{\alpha}_2) \simeq 0.28$ of ants unaffected by the stimulus and setting off in a uniformly random direction.

15. APPROXIMATION TO THE VON MISES DISTRIBUTION (§6 IN ARTICLE)

The von Mises distribution has density

$$g_{\mathrm{VM}}(\theta) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\theta - \mu)\}, \quad -\pi \leq \theta < \pi,$$

where $-\pi \leq \mu < \pi$, $\kappa \geq 0$, and $I_p(\cdot)$ denotes the modified Bessel function of the first kind and order p. In this section, we discuss, first using some figures and then in greater theoretical depth, whether and when our proposed model provides a good approximation to the von Mises distribution. We take each to have the same mean resultant length, $\gamma = I_1(\kappa)/I_0(\kappa)$, and choose

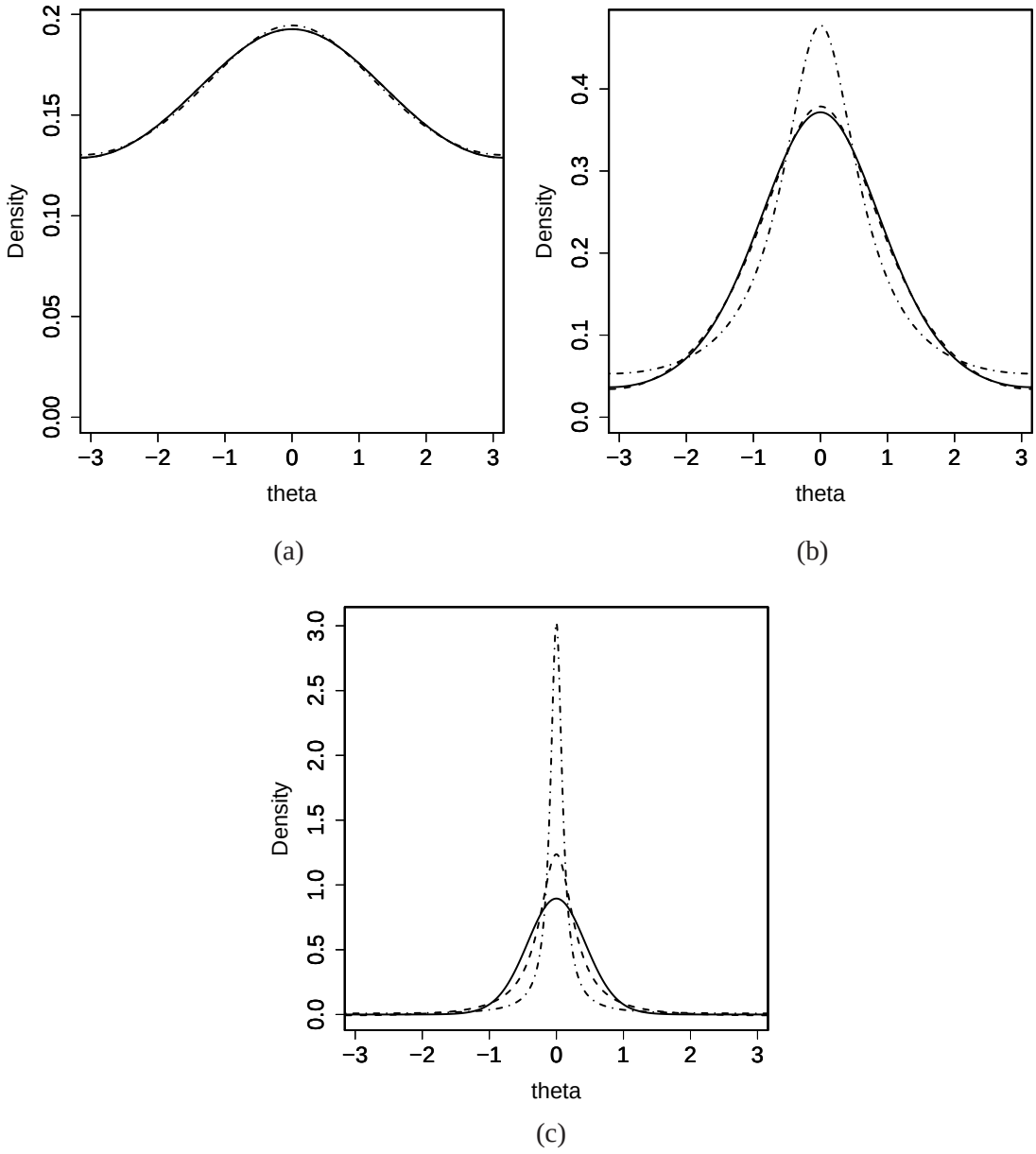


Fig. S5. Densities of the von Mises distribution (solid), the proposed model (dashed), and wrapped Cauchy distribution (dot-dashed), with mean direction 0 and mean resultant length: (a) 0.1, (b) 0.5 and (c) 0.9. The approximating member of the proposed model is chosen to have the same kurtosis as the von Mises distribution.

the approximating member of the full model family to be symmetric and have the same kurtosis as the von Mises distribution, $\bar{\alpha}_2 = I_2(\kappa)/I_0(\kappa)$.

Figure S5 exhibits the densities of the von Mises distribution, the proposed model as described above and, for further comparison, the wrapped Cauchy distribution with the same mean resultant length. Figure S5(a) suggests that, if the mean resultant length is small, both the proposed

family and its wrapped Cauchy submodel provide a good approximation to the von Mises distribution. It seems from Figure S5(b) that, when the mean resultant length is 0.5, the shape of the density of the appropriate special case of our model still looks very similar to that of the von Mises distribution, but that of the wrapped Cauchy distribution does not. For large mean resultant length, we do not provide good approximations to the von Mises distribution, at least not by moment matching as here.

Now we consider the theoretical background of the results of the figure. Note that the density of the von Mises distribution can be expressed using Fourier series expansion as

$$g_{\mathrm{VM}}(\theta) = \frac{1}{2\pi}\left[1 + 2\sum_{p=1}^{\infty}\frac{I_p(\kappa)}{I_0(\kappa)}\cos\{p(\theta-\mu)\}\right], \quad -\pi \le \theta < \pi.$$

For small κ, the Fourier coefficients of $g_{VM}(\theta)$ can be approximated as

$$\frac{I_p(\kappa)}{I_0(\kappa)} \sim \frac{1}{p!}\left(\frac{\kappa}{2}\right)^p + O(\kappa^{p+2})$$

which leads to the usual cardioid density approximation with mean resultant length $\kappa/2$:

$$g_{\mathrm{VM}}(\theta) \sim \frac{1}{2\pi}\{1 + \kappa\cos(\theta-\mu)\}$$

(Mardia & Jupp, 1999, (3.5.20)). Our model includes the cardioid distribution (§4.1 of the article), explaining its similarity to the von Mises distribution in this case.

Next consider the case in which κ is not large as exhibited in Figure S5(b). It is known that $I_p(\kappa)/I_0(\kappa)$ converges to 0 as p tends to infinity and the convergence is very fast. Therefore, if κ is not large, then $I_p(\kappa)/I_0(\kappa)$ with $p \ge 3$ takes a small value. This implies that our model, which has the same first and second Fourier coefficients as the von Mises distribution, still provides a fairly good approximation to the von Mises distribution. On the other hand, the wrapped Cauchy distribution gives a worse approximation because its second Fourier coefficient takes a different value from that of the von Mises distribution.

Finally we discuss the case in which the mean resultant lengths are large. In this case $I_p(\kappa)/I_0(\kappa)$ takes a fairly large value even for $p \ge 3$. To discuss the values of the densities, it is helpful to consider the ratio of the Fourier coefficient of the von Mises distribution to that of the proposed model and to that of the wrapped Cauchy distribution, which are given by

$$R_{\mathrm{G}}(p) = \frac{I_p(\kappa)}{I_0(\kappa)\gamma\rho^{p-1}} \quad \text{and} \quad R_{\mathrm{WC}}(p) = \frac{I_p(\kappa)}{I_0(\kappa)\gamma^p},$$

respectively. It follows from Stirling's formula and the definition of the modified Bessel function that

$$\lim_{p\to\infty} R_{\mathrm{G}}(p) = \lim_{p\to\infty} R_{\mathrm{WC}}(p) = 0.$$

Therefore the convergence of the Fourier coefficient of the von Mises is faster than that of both the proposed model and the wrapped Cauchy. This suggests that, for a large mean resultant length, the density of the von Mises behaves quite differently from the densities of the proposed model and wrapped Cauchy distribution. However, as can be seen in Figure S5(c), the proposed model provides a better approximation to the von Mises distribution than does the wrapped Cauchy distribution. This is partly because the second Fourier coefficient of the von Mises distribution is the same as that of the proposed model. In addition, the ratios of the Fourier coefficients

of the proposed model to those of the wrapped Cauchy distribution are less than one, i.e.,

$$\frac{\gamma\rho^{p-1}}{\gamma^p} = \left\{\frac{I_0(\kappa)I_2(\kappa)}{I_1^2(\kappa)}\right\}^{p-1} < 1 \quad \text{for } \kappa \geq 0, \quad p = 2, 3, ...$$

This is because the unpowered inequality is a Turán-type inequality for the modified Bessel function; see §1 of Baricz & Ponnusamy (2013) for a review. From these observations it appears that each Fourier coefficient of the von Mises distribution takes a closer value to that of the moment-matched proposed model than that of the wrapped Cauchy distribution.

REFERENCES

ABE, T. & PEWSEY, A. (2011). Sine-skewed circular distributions. *Statistical Papers*, **52**, 683–707.

BARICZ, A. & PONNUSAMY, S. (2013). On Turán type inequalities for modified Bessel functions. *Proceedings of the American Mathematical Society*, **141**, 523–32.

JANDER, R. (1957). Die optische richtungsorientierung der Roten Waldameise (Formica Rufa L.). *Z. Vergleichende Physiol.* **40**, 162–238.

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

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

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

RUDIN, W. (1987). *Real and Complex Analysis*, 3rd ed. New York: McGraw-Hill.

UMBACH, D. & JAMMALAMADAKA, S.R. (2009). Building asymmetry into circular distributions. *Statistics and Probability Letters*, **79**, 659–63.

[*Received July* 2013. *Revised*]