Inverse Batschelet distributions

M.C. Jones

The Open University, Milton Keynes, UK

and Arthur Pewsey

University of Extremadura, Cáceres, Spain

Summary. We provide four-parameter families of distributions on the circle which are unimodal and display the widest ranges of both skewness and peakedness. Our approach is to transform the scale of a generating distribution, such as the von Mises, in various extensions of an approach first brought to general attention in Batschelet's (1981) book. The key is to employ inverses of Batschelet-type transformations in certain ways. The skewness transformation is especially appealing as it has no effect on the normalising constant. As well as a variety of interesting theoretical properties, when likelihood inference is explored these distributions display orthogonality between elements of a pairing of parameters into (location, skewness) and (concentration, peakedness). Further, the location parameter can be made approximately orthogonal to all the other parameters. Profile likelihoods come to the fore in practice. Two illustrative applications are investigated.

Keywords: Circular statistics; Flat-topped; Parameter orthogonality; Skew distributions; Transformation of scale; Unimodality; Von Mises distribution.

Address for correspondence: M.C. Jones, Department of Mathematics & Statistics, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK.
E-mail: m.c.jones@open.ac.uk

1. Introduction

Let $f(\theta)$, $-\pi < \theta \leq \pi$, be the density of a symmetric circular distribution i.e. $f(\theta) = f(-\theta)$ and $f(\theta + 2k\pi) = f(\theta)$, $k = 0, \pm 1, ...$, which in addition is unimodal. Without loss of generality, set the mode of f to 0, and its antimode to π. In practice, one typically works with $f(\theta - \mu)$ where $-\pi < \mu \leq \pi$ and f also includes a concentration parameter $\kappa > 0$, but these will not figure in our notation except when it is necessary to do so. Popular examples of unimodal symmetric circular distributions include the von Mises, wrapped normal, wrapped Cauchy and cardioid distributions; see, for example, Mardia & Jupp (1999) and Jammalamadaka & SenGupta (2001) for book length treatments of these and many related topics.

How can one introduce skewness into these distributions, controlled by a single parameter ν, say, in such a way that unimodality is retained? There are many good answers to this question when the random variable has support the real line $\mathbb{R}$ but surprisingly few of them transfer successfully to the circle $\mathbb{S}^1$, as seen at the end of this introduction. Our starting point in this paper is a fairly old idea, which has seen renewed interest very recently, on 'transformation of scale' i.e. unimodal skew circular distributions with densities of the form

$$g_\tau(\theta) \propto f(\tau(\theta)) \tag{1}$$

for some appropriate monotone 'skewing function' τ. Note that this is not the same as transforming the underlying circular random variable. In particular, some interesting examples of this approach are presented in Section 15.6 of Batschelet (1981) utilising the transformation

$$\tau(\theta) = \tau_\nu(\theta) = \theta + \nu \cos\theta. \tag{2}$$

The first appearance of this idea is in Papakonstantinou (1979) in the special case where f is the density of the cardioid distribution; Batschelet (1981) also applies the idea to the von Mises distribution. The choice (2) ensures that g_τ is a bona fide circular density because $\tau_\nu(\theta + 2k\pi) = \tau_\nu(\theta) + 2k\pi$ and g_τ is unimodal if the 'skewing parameter' ν satisfies $-1 \leq \nu \leq 1$, a condition we always apply to ν from here on. The normalising constant associated with g_τ is, however, not immediately available in general and this presages relative disadvantages of this approach in terms of tractability.

In this paper, we present an alternative version of (1) in which τ is taken to be the inverse of a slightly extended version of (2). Remarkably, this choice

turns out to have no effect whatsoever on the normalising constant associated with g_τ i.e. the constant of proportionality in (1) is 1. A wide variety of appealing properties of this 'inverse Batschelet' approach to unimodal skew circular distributions follow and are explored in Section 2.

If, instead of τ_ν given by (2), the function τ in (1) is given by

$$\tau(\theta) = \tau_\lambda(\theta) = \theta + \lambda \sin\theta, \tag{3}$$

a family of symmetric circular distributions ensues. These too are bona fide circular densities and are unimodal if $-1 \leq \lambda \leq 1$, which we also assume from here on. The idea is again due to Papakonstantinou (1979) and Batschelet (1981, Section 15.7); more details on the particular special cases with which they worked can be found in Abe et al. (2009) and Pewsey et al. (2010).

In this paper, we also present an advantageous extension of this idea; the inverse of Batschelet's transformation function, here τ_λ, again plays a leading role. The effect on the normalising constant is no longer null — we suspect this is not possible for symmetric distributions — but the normalising constant remains a simple integral comparable to those associated with direct Batschelet approaches. Instead, the main advantage in the symmetric case is that our new families of distributions display the widest range of shapes ranging from the flat-topped, through f, to the sharply peaked. Details are in Section 3.

While we stress that our skewing approach can be applied to any family of symmetric distributions (e.g. that of Jones and Pewsey, 2005), it is nonetheless most appealing to apply it to the families of distributions generated in Section 3, and this is done in Section 4. We believe the result to display easily the most wide-ranging slate of skew and symmetric shapes of four-parameter unimodal distributions on the circle. The reader is invited to look ahead to Figures 2, 6 and 7 for illustrations of this claim when the underlying f is the von Mises density.

Our new families of distributions display inferential advantages under maximum likelihood too. First, the parameters 'pair off' into two pairs of mutually orthogonal parameters, namely the location and skewing parameters $\{\mu, \nu\}$ and the concentration and peakedness parameters $\{\kappa, \lambda\}$. This kind of behaviour is rare in four-parameter families on the line or circle; see Jones and Anaya-Izquierdo (2010). Second, a further approximate orthogonalisation of the location and skewing parameters is derived which results in the location parameter (assumed to be the parameter of main interest, be

it μ or the mode of the distribution) being either exactly or approximately orthogonal to all the other parameters. See Section 5 for details.

Two examples of the fitting of our new families of distributions to data are presented in Section 6. Profile likelihoods play a major role there. The examples illustrate the breadth of shapes available within our family by focussing on cases where sharply peaked and wide bodied distributions arise, respectively. Of course, our family accommodates the many situations where 'less extreme' density shapes are required also. The paper concludes with a brief discussion in Section 7.

Our desire to retain unimodality is because we see these distributions as components which can be combined in interpretable ways to model bi- or multi-modal distributions by mixing or perhaps by the kind of 'multiplicative mixing' underlying the generalised von Mises distributions (Maksimov, 1967, Gatto & Jammalamadaka, 2007). And most alternative approaches fail to retain unimodality; examples include the families of Kato & Jones (2010) – which has advantages in terms of interaction with the Möbius transformation — and the adaptation of skew distributions of Azzalini (1985) type (Umbach & Jammalamadaka, 2009, Abe & Pewsey, 2011). Approaches involving distribution functions in their densities are not really well-defined in the circular case, while the differential rescaling normally involved in skew two-piece distributions is simply not available. The most direct, unimodal, competitors are indeed the 'direct Batschelet' distributions already mentioned, but we regard the distributions of the current paper as considerably superior implementations of essentially those ideas. Intriguingly, transformation of scale distributions do have analogues on the real line, but they are as yet relatively obscure (Baker, 2008, Jones, 2009) and arguably not as useful as alternatives there. Finally, wrapping skew distributions does readily retain unimodality but at the expense, usually, of tractable density functions and likelihoods: the wrapped stable distributions (Mardia, 1972, p.57, Pewsey, 2008) are a good example.

2. Skewing by inverse Batschelet-type transformation of scale

2.1. Definition and basic properties

Return to the technique of skewing circular distributions by employing the transformation of scale (2) as τ in formula (1). We propose to improve on this approach by changing the transformation τ_ν by adding ν, changing the

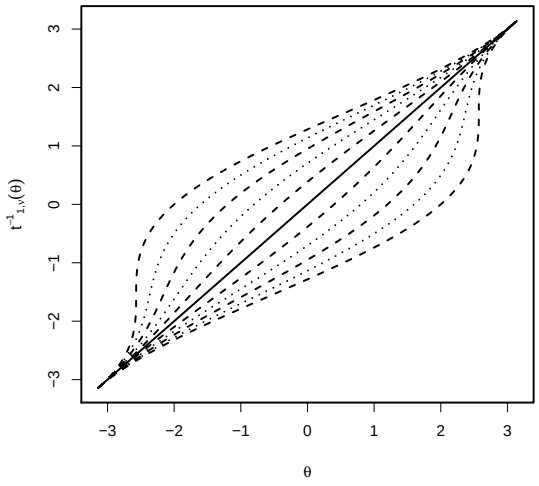

Figure 1: The function $t^{-1}_{1,\nu}(\theta)$ for $\theta \in (-\pi, \pi]$ and, in order of *increasing* height at $\theta = 0$, $\nu = -1(0.2)1$. The curves for the different ν-values have been drawn using alternating dashed and dotted lines to aid visualisation. The solid line corresponds to $\nu = 0$.

sign of ν and then employing the inverse of the transformation. That is, we define

$$t_\nu(\theta) = t^{-1}_{1,\nu}(\theta) \quad \text{where} \quad t_{1,\nu}(\theta) = \theta - \nu - \nu\cos\theta, \qquad -1 \le \nu \le 1. \tag{4}$$

Note that t_ν and $t_{1,\nu}$ are bijections from $\mathbb{S}^1$ to $\mathbb{S}^1$. A range of t_ν functions, the lack of an explicit formula for which is but a minor inconvenience, are shown in Figure 1.

Our family of densities then becomes

$$g_\nu(\theta) = f(t_\nu(\theta)). \tag{5}$$

Notice the first major advantage of this approach: there is no change to the normalising constant from that of f. To prove this, use the substitution $\omega = t_\nu(\theta)$ to find that

$$\int_{-\pi}^{\pi} f\{t_\nu(\theta)\}d\theta = \int_{-\pi}^{\pi} t'_{1,\nu}(\omega)f(\omega)d\omega = \int_{-\pi}^{\pi} (1 + \nu\sin\omega)f(\omega)d\omega = 1$$

by the symmetry of f. The transformation of scale t_ν is special in allowing this remarkable property to hold.

Periodicity of g_ν is assured by the following argument. First, note that $t_{1,\nu}(\theta+2k\pi) = t_{1,\nu}(\theta)+2k\pi$, $k = 0, \pm 1, \ldots$. Thus, $\theta+2k\pi = t_\nu\{t_{1,\nu}(\theta)+2k\pi\}$, but it also equals $t_\nu\{t_{1,\nu}(\theta)\} + 2k\pi$ by definition. Setting $\phi = t_{1,\nu}(\theta)$, with $-\pi \leq \phi \leq \pi$ by the bijective nature of $t_{1,\nu}$, shows that $t_\nu(\phi + 2k\pi) = t_\nu(\phi) + 2k\pi$ and hence that

$$g_\nu(\phi + 2k\pi) = f\{t_\nu(\phi + 2k\pi)\} = f\{t_\nu(\phi) + 2k\pi\} = f\{t_\nu(\phi)\} = g_\nu(\phi)$$

as required.

Unimodality of g defined by (1), and hence (5), is assured for any monotone increasing transformation τ. This is because the values of g increase and then decrease, taking on the values of f, except at a different rate governed by the transformation of scale. The mode of (1) is, therefore, at $\tau^{-1}(0)$. This value is not available explicitly for existing Batschelet distributions but it is explicitly available for our version since $\tau^{-1}(0) = t_{1,\nu}(0)$. That is, the mode of (5) is at -2ν. Also, the antimode remains at π. For differentiable f, a simple calculus proof of the unimodality of g_ν is also available; in general, g_ν retains the degree of differentiability of f.

It is also the case that the density associated with $-\nu$ is the reflection in zero of the density associated with ν. To see this, note that $t_{1,-\nu}(\theta) = -t_{1,\nu}(-\theta)$ so that $t_{-\nu}(\theta) = -t_\nu(-\theta)$ and hence

$$g_\nu(\theta; -\nu) = f\{t_{-\nu}(\theta)\} = f\{-t_\nu(-\theta)\} = f\{t_\nu(-\theta)\} = g_\nu(-\theta; \nu). \qquad (6)$$

2.2. A skew von Mises distribution

A particularly attractive asymmetric family of distributions (which has three parameters when the location μ is included) are those with density

$$g_{\nu;VM}(\theta) = \{2\pi I_0(\kappa)\}^{-1} \exp[\kappa \cos\{t_{1,\nu}^{-1}(\theta)\}] \qquad (7)$$

where $I_0(\kappa)$ is the modified Bessel function of the first kind and order zero i.e. density (5) when f is the von Mises density. Some of these densities are graphed in Figure 2. The maximum amount of skewness available is very satisfactory, as are the shapes of all the densities. We display only versions of our densities with $\nu \geq 0$; by virtue of (6), all the corresponding reflected densities, with $\nu \leq 0$, are available, too, of course.

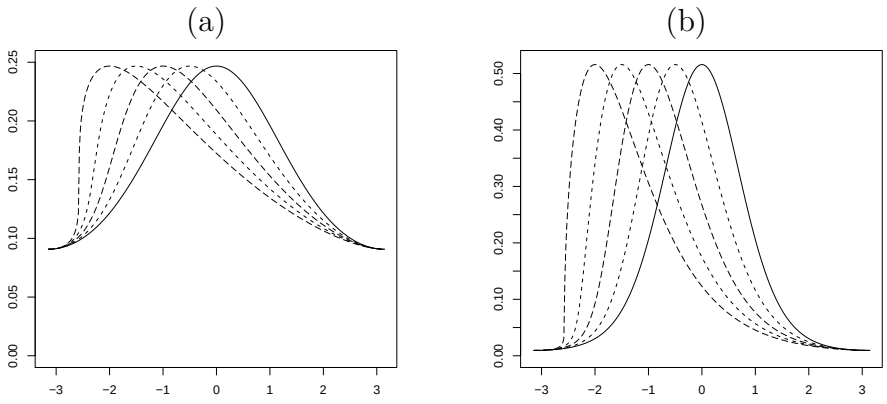

Figure 2: Skew von Mises densities of the form (7) with $\mu = 0$ and (a) $\kappa = 1/2$, (b) $\kappa = 2$. In each frame, reading from right to left, the densities correspond to $\nu = 0, 0.25, 0.5, 0.75, 1$. The curves for the different ν-values have been drawn using alternating short-dashed and long-dashed lines to aid visualisation.

2.3. Circular skewness

Moments of the new distributions are not especially tractable, being at essentially the same level of tractability as those of direct Batschelet distributions. They can, however, be readily calculated numerically. This allows us to consider, in particular, the circular skewness of Batschelet (1965) defined as $E[\sin\{2(\Theta - \mu_0)\}]$ where μ_0 is the circular mean direction. The circular skewness is plotted in Figure 3(a) for $g_{\nu,\psi}$ for $\kappa = 1$, $\nu > 0$ and all ψ. Here, $g_{\nu,\psi}$ is the version of (5) based on the family of symmetric distributions proposed by Jones and Pewsey (2005) which have densities

$$f_\psi(\theta) \propto \{1 + \tanh(\kappa\psi)\cos\theta\}^{1/\psi}.$$

We utilise this symmetric family here in order to display properties and allow comparisons for certain well-known distributions; in particular, $\psi = 0$ corresponds to the von Mises distribution (so $g_{\nu,0} = g_{\nu;VM}$) while $\psi = 1$ and -1 correspond to the cardioid and wrapped Cauchy distributions, respectively. The circular skewness is zero for $\nu = 0$, negative for $\nu > 0$, and decreasing as a function of ν. Those steeped in linear statistics should

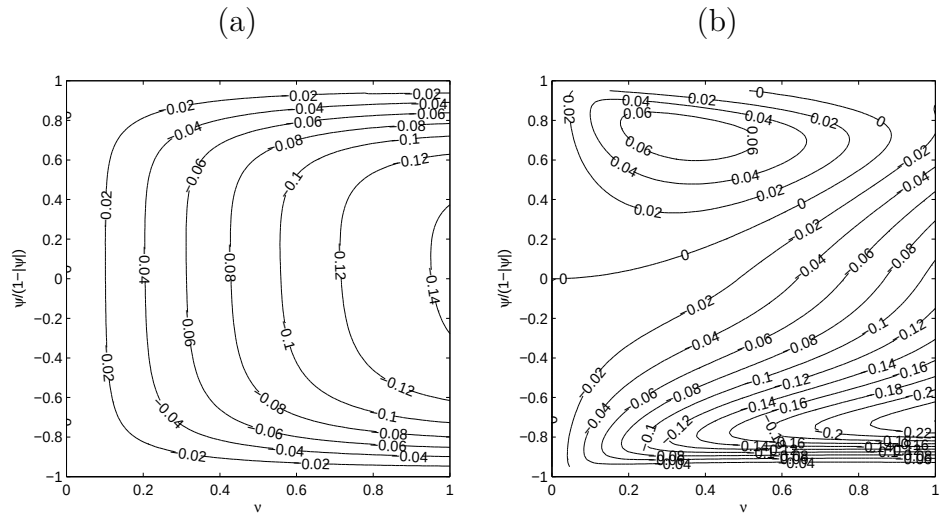

Figure 3: The circular skewness of the distributions with density (a) $g_{\nu,\psi}$ and (b) $h_{\nu,\psi}$, plotted as a function of $\nu > 0$ and $\psi/(1-|\psi|)$. In each case, $\kappa = 1$.

note that skewness on the circle is naturally defined to have the opposite sign from the skewness of its usual linear representation. Figure 3(a) shows that the greatest absolute values of circular skewness with the skew Jones-Pewsey family are achieved around $\psi = 0$. However, circular skewness is also considerable for $\psi \in (-1, 1)$, disappearing quite rapidly for the more extreme $g_{\nu,\psi}$ distributions with $|\psi| > 1$, as the underlying Jones & Pewsey distributions tend to the uniform.

2.4. A transformation link

Umbach & Jammalamadaka (2009), as part of a wider proposal, and Abe & Pewsey (2011), more specifically, consider the family of 'sine-skewed' circular distributions with densities

$$h_\nu(\phi) = (1 + \nu \sin \phi) f(\phi), \qquad -1 \leq \nu \leq 1, \tag{8}$$

for f symmetric about zero and $\phi \in \mathbb{S}^1$. Densities (8) and (5) are related through transformation of random variables thus: if $\Theta \sim g_\nu(\theta)$ then $\Phi = t_\nu(\Theta) \sim h_\nu(\phi)$. This is implicit in the proof that the normalising constant is unity in Section 2.1.

It is natural to compare (8) and (5), and we contend that (5) is more advantageous. This is on two main grounds. First, as discussed in Abe & Pewsey (2011), the h_ν densities are not necessarily unimodal if f is; this even happens for some values of κ and ν for the von Mises distribution and to a greater extent for the cardioid distribution. Second, the skewness properties of h_ν are not as attractive as those of g_ν. Figure 3(b) displays the circular skewness of h_ν when $f = f_\psi$, referred to as $h_{\nu,\psi}$, and can be directly compared with Figure 3(a). Figure 3(b) indicates a good range of skewness for $\psi << 0$, including the sine-skewed wrapped Cauchy distribution, rather less skewness for $\psi \approx 0$, around the sine-skewed von Mises distribution, and low levels of skewness including skewness in the opposite direction for $\psi >> 0$, including the sine-skewed cardioid distribution. This reflects well the appearance of the $h_{\nu,\psi}$ densities shown in Figure 1 of Abe & Pewsey (2011).

2.5. Random variate generation

The above transformation link yields a neat method for random variate generation from g if one has available a random variate Ψ generated from f and an independent $U(0,1)$ random variable U. Because h_ν can be considered a weighted density, proportional to wf with weight $w(\phi) = \frac{1}{2}(1 + \nu \sin\phi)$ satisfying $w(\phi) + w(-\phi) = 1$, a random variate Φ follows density h_ν if

$$\Phi = \begin{cases} \Psi & \text{if } U \leq \frac{1}{2}(1 + \nu \sin\Psi), \\ -\Psi & \text{otherwise.} \end{cases}$$

Applying the transformation completes the algorithm for generating $\Theta \sim g_\nu$:

(1) Generate $\Psi \sim f$ and, independently, $U \sim U(0,1)$;

(2) Set

$$\Theta = \begin{cases} \Psi - \nu - \nu\cos\Psi & \text{if } U \leq \frac{1}{2}(1 + \nu \sin\Psi), \\ -\Psi - \nu - \nu\cos\Psi & \text{otherwise.} \end{cases}$$

Note that there is no rejection in this algorithm.

2.6. Density-based asymmetry

A beauty of transformation-of-scale approaches is that they lend themselves to tractable forms of the density-based asymmetry function(s) mentioned by O'Hagan (1994, Section 2.6) and developed by Avérous et al.

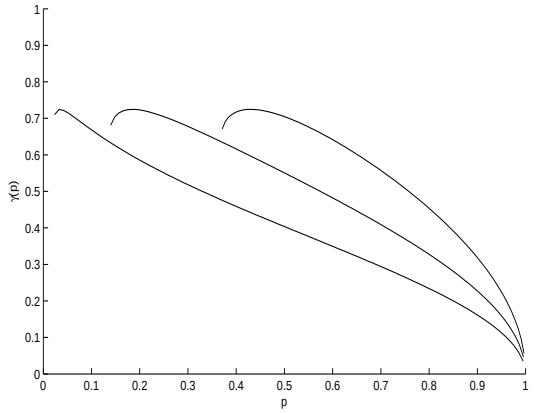

Figure 4: The density-based asymmetry function for distribution (5) when $\nu = 1$ and f is the von Mises density. From right to left, the asymmetry functions correspond to $\kappa = 0.5$, 1 and 2, respectively.

(1996), Boshnakov (2007) and Critchley & Jones (2008); see Jones (2009). This approach to skewness of unimodal distributions is density and mode-based. We need x_R such that $x_R = g_R^{-1}(pg(x_0))$ where the 'R' subscript means "the part of the density to the right of the mode", x_0 is the mode, and $0 \leq p_{\min} < p < 1$. For density (5), $x_R = t_{1,\nu}(f_R^{-1}[pf\{t_\nu(x_0)\}]) = t_{1,\nu}(c_p)$ where $c_p = f_R^{-1}\{pf(0)\}$. Similarly, $x_L = t_{1,\nu}(-c_p)$. Also, $p_{\min} = f(\pi)/f(x_0)$ is typically greater than 0. It follows that the asymmetry function associated with (5) is

$$\gamma(p) = \frac{x_R - 2x_0 + x_L}{x_R - x_L} = \frac{t_{1,\nu}(c_p) - 2x_0 + t_{1,\nu}(-c_p)}{t_{1,\nu}(c_p) - t_{1,\nu}(-c_p)} = \nu\frac{(1 - \cos c_p)}{c_p}. \quad (9)$$

First, (9) says that skewness has the same sign as ν, for all p. Second, the function $C(x) = (1 - \cos x)/x$ increases steadily from 0 to a maximum at P, say, between $x = \pi/2$ and $x = \pi$ and then decreases again such that $C(\pi) = C(\pi/2) = 2/\pi$; P satisfies $P \sin P + \cos P = 1$. Third, c_p is a decreasing function of p with $c_{p_{\min}} = \pi$ and $c_1 = 0$. Thus, $C(c_p)$ is an increasing function of p from $C(c_{p_{\min}}) = 2/\pi$ until a maximum is reached at p_m such that $c_{p_m} = P$ and then decreases again until $C(c_1) = 0$. Figure 4 shows $\gamma(p)$ for $\nu = 1$ and f being the von Mises distribution for which $c_p = \cos^{-1}(1 + \kappa^{-1} \log p)$, $e^{-2\kappa} < p < 1$ and therefore $C(c_p) = -\log p/(\kappa c_p)$; reduce the values of $\gamma(p)$ proportionately for other ν. Compared with the types of asymmetry

function shown in Figures 2 and 3 of Critchley & Jones (2008) for some standard distributions on the positive half-line, these asymmetry functions differ principally in their relatively large values continuing for p near one which suggests a greater skewness in the "main body" of the distribution compared with more familiar shapes of density. This is consistent with the graphs of densities shown in Figure 2. The initial increase in the asymmetry functions is only slight and fairly inconsequential. Note that this asymmetry measure depends on the concentration parameter but, in shape, only to a small extent.

Perhaps most importantly, the fact that (9) is directly proportional to ν means that any scalar measure of skewness based on (9) will also be directly proportional to ν, and this strongly reinforces the notion that ν itself is an excellent skewness measure for these distributions.

3. Wider symmetric families by further inverse Batschelet-type transformation of scale

3.1. Direct and inverse Batschelet-type transformations of scale

As described in Section 1, the original direct Batschelet formulation is to consider the family of unimodal symmetric distributions of the form $g_{1,\lambda}(\theta) \propto f(\tau_\lambda(\theta))$ where $\tau_\lambda(\theta)$, $-1 \leq \lambda \leq 1$, is given by (3). Figure 1 of Pewsey et al. (2010) shows a wide variety of density shapes arising when f is the von Mises distribution, including a number of wide bodied ones associated with $\lambda < 0$.

In the spirit of Section 2, an alternative approach that suggests itself is the inverse Batschelet transformation of scale resulting in

$$g_{2,\lambda}(\theta) = \{1 + \lambda\alpha_f\}^{-1} f(\tau_\lambda^{-1}(\theta))$$

where $\alpha_f = E_f(\cos\Psi)$. Note the relative, but not 'complete', simplicity of the normalising constant associated with these densities. They, too, display a wide variety of density shapes, including a number of sharply peaked ones associated with $\lambda < 0$, but without the wide bodied cases mentioned above.

A fuller range of density shapes including both wide bodied and sharply peaked densities is available through a different transformation of scale, $t_\lambda(\theta)$, which includes the most wide bodied direct transformation $\theta - \sin\theta$ when $\lambda = -1$ and the most sharply peaked transformation $(\theta - \sin\theta)^{-1}$ when

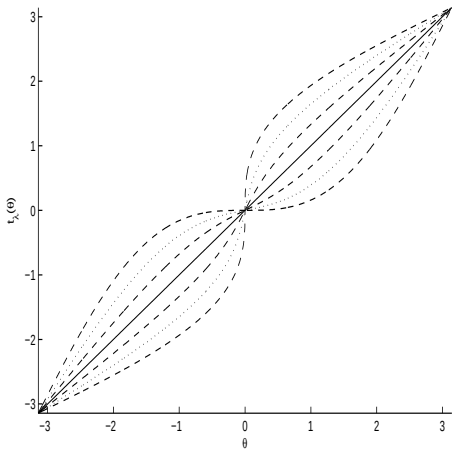

Figure 5: The function $t_\lambda(\theta)$ for $\theta \in (-\pi, \pi]$ and, in order of *increasing* height at $\theta = \pi/2$, $\lambda = -1(\frac{1}{3})1$. The curves for the different λ-values have been drawn using alternating dashed and dotted lines to aid visualisation. The solid line corresponds to $\lambda = 0$.

$\lambda = 1$ (as well as θ when $\lambda = 0$), and fills in between them differently. This transformation is given by

$$t_\lambda(\theta) = \frac{1-\lambda}{1+\lambda}\theta + \frac{2\lambda}{1+\lambda}s_\lambda^{-1}(\theta), \quad -1 < \lambda \leq 1, \tag{10}$$

where

$$s_\lambda(t) = t - \tfrac{1}{2}(1+\lambda)\sin t \tag{11}$$

((11) being a subset of (3)). The definition is completed by allowing $t_{-1}(\theta) = \lim_{\lambda\to-1} t_\lambda(\theta)$ which is readily shown to equal $\theta - \sin\theta$.

The transformation (10) is shown in Figure 5. It is clear that $t_\lambda : S^1 \to S^1$ has the basic properties required of it, namely that it is one-to-one increasing such that $t_\lambda(0) = 0$ and $t_\lambda(\pm\pi) = \pm\pi$, and that $t_\lambda(\theta + 2k\pi) = t_\lambda(\theta) + 2k\pi$. It is also the case that

$$t_\lambda^{-1}(\theta) = t_{-\lambda}(\theta). \tag{12}$$

The key to proving this is that, as is readily seen,

$$t_\lambda(s_\lambda(\theta)) = s_{-\lambda}(\theta). \tag{13}$$

Then, setting λ to $-\lambda$ in (13) leads to $s_{-\lambda}(\theta) = t_{-\lambda}^{-1}(s_\lambda(\theta))$ which, when compared with (13), yields (12).

3.2. The new symmetric family

The new symmetric family has density

$$g_\lambda(\theta) = K_{\kappa,\lambda}^{-1} f(t_\lambda(\theta)). \tag{14}$$

This family is, of course, unimodal with mode at 0 and antimode at π. Judicious substitutions (for which see Section 4.2) show that

$$K_{\kappa,\lambda} = \int_{-\pi}^{\pi} \{1 - \tfrac{1}{2}(1+\lambda)\cos\theta\} f\{\theta - \tfrac{1}{2}(1-\lambda)\sin\theta\} d\theta \tag{15}$$

which equals

$$\frac{1+\lambda}{1-\lambda} - \frac{2\lambda}{(1-\lambda)} \int_{-\pi}^{\pi} f\{\theta - \tfrac{1}{2}(1-\lambda)\sin\theta\} d\theta \tag{16}$$

for $\lambda < 1$. This normalising constant is on a numerical par with normalising constants for direct Batschelet distributions, being single integrals of straightforward functions involving no implicit inverses. Indeed, representation (16) shows that $K_{\kappa,\lambda}$ is a simple linear function of the normalising constant associated with $g_{1,-\lambda}$ above. Note also that $K_{\kappa,1}^{-1} = 1 - \alpha_f$.

Again, it is useful to look at the special case where f is the von Mises distribution: a variety of densities (14) are plotted for this case in Figure 6. The extremely wide range of (symmetric) shapes ranging from the very wide bodied to the very sharply peaked are illustrated in this figure.

3.3. Random variate generation

In this case, if $\Theta \sim g_\lambda(\theta)$ then $\Phi = t_\lambda(\Theta) \sim \iota_\lambda(\phi)$ where

$$\iota_\lambda(\phi) = K_{\kappa,\lambda}^{-1}(t_{-\lambda})'(\phi) f(\phi) = K_{\kappa,\lambda}^{-1} \left[\frac{1 - \tfrac{1}{2}(1+\lambda)\cos\{S_{-\lambda}^{-1}(\phi)\}}{1 - \tfrac{1}{2}(1-\lambda)\cos\{S_{-\lambda}^{-1}(\phi)\}} \right] f(\phi).$$

The first of these equalities arises because of property (12).

This transformation link allows random generation of $\Theta \sim g_\lambda$ from $\Psi \sim f$ via $\Phi \sim \iota_\lambda$. First, for completeness, when $\lambda = 0$, $\Phi = \Psi$. Second, when $\lambda = 1$

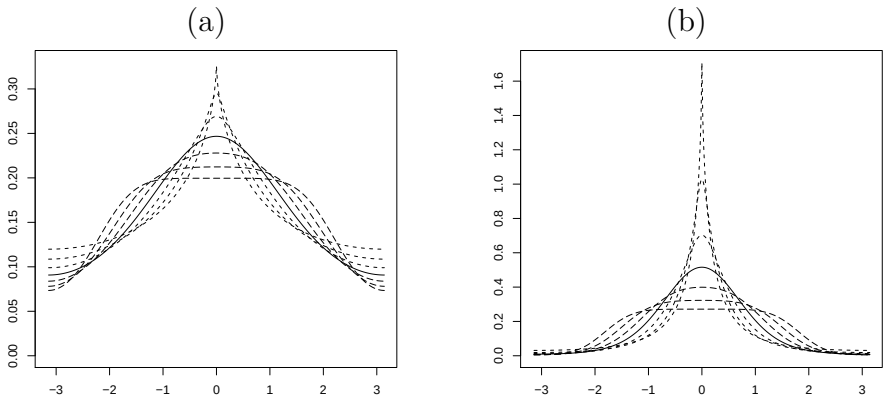

Figure 6: Symmetric von Mises-based densities of the form (14) with $\mu = 0$ and (a) $\kappa = 1/2$, (b) $\kappa = 2$. In each frame, reading from bottom to top at $\theta = 0$, the densities correspond to $\lambda = -1(\frac{1}{3})1$. The curves have been drawn using short-dashed lines for $\lambda > 0$, a solid line for $\lambda = 0$ and long-dashed lines for $\lambda < 0$ to aid visualisation.

we can take the weight function associated with ι_1 to be $\omega(\phi) = \frac{1}{2}(1 - \cos\phi)$ and note that $\omega(\phi) + \omega(\phi + \pi) = 1$. It follows that we can take

$$\Phi = \begin{cases} \Psi & \text{if } U \leq \frac{1}{2}(1 - \cos\Psi), \\ \Psi + \pi & \text{otherwise,} \end{cases}$$

where U is an independent $U(0, 1)$ random variable.

However, unlike in the skewness case, the form of the weighting function in other cases is such that we cannot 'reuse' the rejected random variables. We have slightly different rejection algorithms for the cases $\lambda > 0$ and $\lambda < 0$. First, when $\lambda > 0$ the weight function in ι_λ is maximised when $\phi = \pi$, so we use

$$w_{\lambda_+}(\phi) = \left(\frac{3 - \lambda}{3 + \lambda}\right) \left[\frac{1 - \frac{1}{2}(1 + \lambda)\cos\{S_{-\lambda}^{-1}(\phi)\}}{1 - \frac{1}{2}(1 - \lambda)\cos\{S_{-\lambda}^{-1}(\phi)\}}\right].$$

When $\lambda < 0$ the weight function is maximised when $\phi = 0$, so we use

$$w_{\lambda_-}(\phi) = \left(\frac{1 + \lambda}{1 - \lambda}\right) \left[\frac{1 - \frac{1}{2}(1 + \lambda)\cos\{S_{-\lambda}^{-1}(\phi)\}}{1 - \frac{1}{2}(1 - \lambda)\cos\{S_{-\lambda}^{-1}(\phi)\}}\right].$$

In either case, we then take

$$\Phi = \begin{cases} \Psi & \text{if } U \leq w(\Psi), \\ \text{reject and try again} & \text{otherwise} \end{cases}$$

where $w(\Psi)$ is $w_{\lambda_\pm}(\Psi)$ as appropriate. This algorithm works but is not especially efficient for larger $|\lambda|$. Very roughly, rejection probabilities greater than one half tend to hold sway when $\lambda \notin (-0.5, 0.5)$.

Finally, set $\Theta = t_{-\lambda}(\Phi)$.

4. Skew-symmetric families by inverse Batschelet-type transformation of scale

4.1. Combining (5) and (14)

While we emphasise that the particularly attractive skewing mechanism (5) can be applied to any symmetric circular distributions, it is especially appealing to apply it to family (14) to try to provide the widest possible range of skew and symmetric unimodal density shapes. The resulting densities are of the form

$$g_{\nu,\lambda}(\theta) = K_{\kappa,\lambda}^{-1} f(t_\lambda(t_\nu(\theta))). \tag{17}$$

In the case of f being the von Mises density, Figure 7 complements the symmetric members of family (17) in Figure 6 by showing a range of skew members of the family. The very wide range of symmetric and skew, wide bodied to sharply peaked, densities in family (17) is clear from these plots.

Properties of densities (17) follow from those of densities (5) and (14). In brief, as well as the normalising constant $K_{\kappa,\lambda}$ being given by (15) or (16), densities (17) remain periodic, unimodal with mode at -2ν, and satisfying the reflection property (6). Random variate generation from (17) arises by following the algorithm of Section 3.3 by that of Section 2.5.

4.2. Moment integrals

Let Θ follow density (17) and consider $E(H(\Theta))$ for an arbitrary function H; examples of H include certain sine and cosine functions, and hence trigonometric moments. The following formula shows that $E(H(\Theta))$ is available as a one-dimensional integral whose integrand involves no implicit inverse functions and which can, therefore, be calculated numerically with ease.

$$E(H(\Theta)) = K_{\kappa,\lambda}^{-1} I_{\kappa,\lambda,\nu}$$

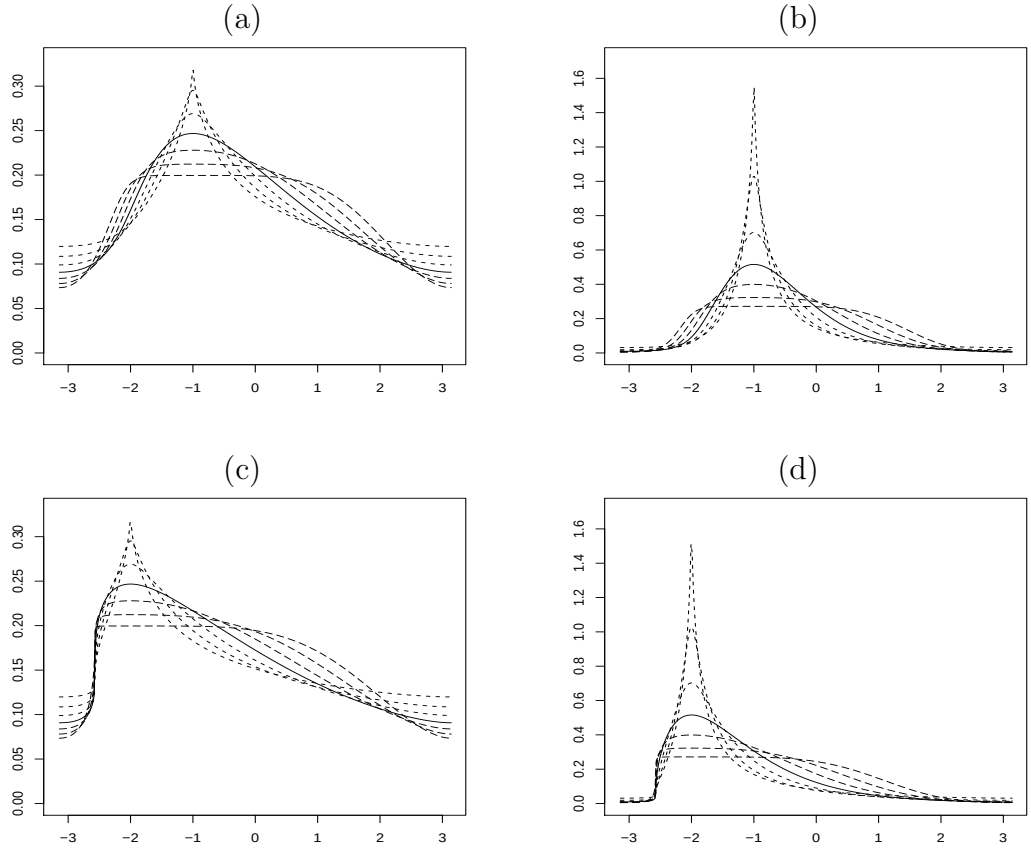

Figure 7: Skew von Mises-based densities of the form (17) with $\mu = 0$ and (a) $\kappa = 1/2, \nu = 1/2$, (b) $\kappa = 2, \nu = 1/2$, (c) $\kappa = 1/2, \nu = 1$, (d) $\kappa = 2, \nu = 1$. In each frame, reading from bottom to top at $\theta = -2\nu$, the densities correspond to $\lambda = -1(\frac{1}{3})1$. The curves have been drawn using short-dashed lines for $\lambda > 0$, a solid line for $\lambda = 0$ and long-dashed lines for $\lambda < 0$ to aid visualisation.

where

$$I_{\kappa,\lambda,\nu} = \int_{-\pi}^{\pi} H\{s_\lambda(\theta)-\nu-\nu\cos s_\lambda(\theta)\}\{1+\nu\sin s_\lambda(\theta)\}\{1-\tfrac{1}{2}(1+\lambda)\cos\theta\} f\{s_{-\lambda}(\theta)\}d\theta \tag{18}$$

and $s_\lambda(\theta)$ is given by (11). To see this, start from

$$I_{\kappa,\lambda,\nu} = \int_{-\pi}^{\pi} H(\omega) f(t_\lambda(t_\nu(\omega)))d\omega$$

and make the substitution $\theta = s_{-\lambda}^{-1}(t_\lambda(t_\nu(\omega))) = s_\lambda^{-1}(t_\nu(\omega))$ by (13).

5. Properties of maximum likelihood estimation

5.1. Asymptotic properties

Consider distribution (5) in the four-parameter form

$$g(\theta) = f\{t_\nu(\theta-\mu);\kappa,\psi\}, \tag{19}$$

that is, reintroduce location parameter $-\pi < \mu \leq \pi$, concentration parameter $\kappa > 0$ and a further symmetric shape parameter ψ. The main points we wish to make are consequences of the skewness transformation of scale (4) and hence apply to (5) whatever the family of symmetric distributions which are skewed; of course, family (17) will constitute a particularly important special case of this (in which case $\psi = \lambda$). For notational convenience, however, we drop explicit mention of κ and ψ from f together with arguments and variables of integration in the next paragraph. We use primes to denote differentiation with respect to x, and superscripts ν, κ and ψ to denote differentiation with respect to those parameters. It is useful to note that

$$t'_{1,\nu}(\theta) = 1+\nu\sin\theta, \quad t''_{1,\nu}(\theta) = \nu\cos\theta, \quad t^{\nu}_{1,\nu}(\theta) = -(1+\cos\theta), \quad t'^{\nu}_{1,\nu}(\theta) = \sin\theta.$$

Let $\iota_{\eta,\xi}$ denote $-E(\partial^2\ell/\partial\eta\partial\xi)$, for (η,ξ) taken from (μ,ν,κ,ψ), these being n times the elements of the expected information matrix. They turn out to be

$$\iota_{\mu\mu} = \int f' t''_{1,\nu}/(t'_{1,\nu})^2 - \int (\log f)'' f/t'_{1,\nu};$$

$$\iota_{\mu\nu} = \int f' t^{\nu}_{1,\nu} t''_{1,\nu}/(t'_{1,\nu})^2 - \int f' t'^{\nu}_{1,\nu}/t'_{1,\nu} - \int (\log f)'' t^{\nu}_{1,\nu} f/t'_{1,\nu};$$

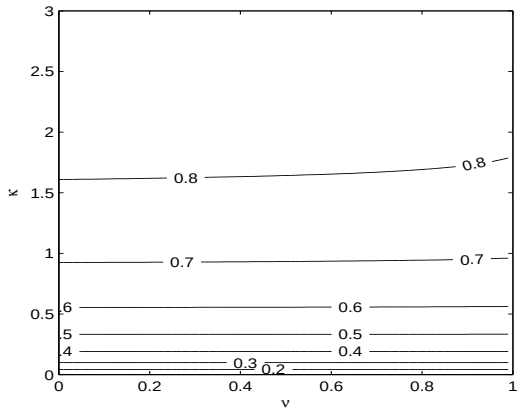

Figure 8: The asymptotic correlation between $\hat{\mu}$ and $\hat{\nu}$ for distribution (4) when f is the von Mises distribution, plotted as a function of ν and κ.

$$\iota_{\mu\kappa} = 0 = \iota_{\mu\psi};$$

$$\iota_{\nu\nu} = \int f'(t^{\nu}_{1,\nu})^2 t''_{1,\nu}/(t'_{1,\nu})^2 - \int f' t'^{\nu}_{1,\nu} t^{\nu}_{1,\nu}/t'_{1,\nu} - \int (\log f)''(t^{\nu}_{1,\nu})^2 f/t'_{1,\nu};$$

$$\iota_{\nu\kappa} = 0 = \iota_{\nu\psi};$$

$$\iota_{\kappa\kappa} = \int (f^{\kappa})^2/f; \quad \iota_{\kappa\psi} = \int f^{\kappa} f^{\psi}/f; \quad \iota_{\psi\psi} = \int (f^{\psi})^2/f.$$

None of these quantities depends on μ.

A remarkable property of this expected information matrix is that the parameters split up into pairs, each member of which is orthogonal to each member of the other. Those pairs are the location and skewness parameters (μ, ν) and the concentration and symmetric shape parameters (κ, ψ). This is the same as what happens in the two-piece distribution on $\mathbb{R}$ (Jones and Anaya-Izquierdo, 2010) and, as far as the authors are aware, for no other four-parameter univariate or circular distribution.

The asymptotic correlation between $\hat{\mu}$ and $\hat{\nu}$, $\mathrm{acorr}(\hat{\mu}, \hat{\nu})$ where 'hats' are used to denote maximum likelihood estimators, depends only on ν and κ and not on ψ. In particular cases we have explored, this correlation is small to moderate for small to moderate values of κ, but becomes quite high for tightly concentrated distributions. See Figure 8 for this correlation when f is the von Mises distribution. (The plot is symmetric in ν so only the

correlation for $\nu \geq 0$ is shown.) There is a remarkably weak dependence of the asymptotic correlation on the value of ν.

On the other hand, the part of the information matrix involving κ and ψ is not affected by ν: it is precisely the same as the equivalent part of the information matrix for f itself. For example, the correlation between $\hat{\kappa}$ and $\hat{\psi}$ is displayed for the distributions of Jones & Pewsey (2005) in Figure 6 of Jones and Pewsey (2004). Happily, this correlation, which appears to be positive, is not too high for most values of κ and ψ.

5.2. Further approximate parameter orthogonality

When the main interest parameter is the location μ (which may depend on covariates etc), its orthogonality to all the other parameters in the model would be advantageous. In the current model it is automatically orthogonal to κ and ψ; in this section, we provide an approximately orthogonal parameter to replace ν when f is the von Mises distribution. The key to our approach is the weak dependence of $\text{acorr}(\hat{\mu}, \hat{\nu})$ on ν observed in Figure 8. This suggests obtaining a parameter ν_μ, say, which is orthogonal to μ when $\nu = 0$ in the hope that it remains approximately orthogonal to μ for $\nu \neq 0$.

It turns out that a parameter of the form $\nu_\mu = \nu - H_1(\kappa)\mu$ will do the trick where, by consideration of $\text{acorr}(\hat{\mu}, \hat{\nu}_\mu)$, $H_1(\kappa) = (\iota_{\mu\nu}/\iota_{\nu\nu}) \mid_{\nu=0}$. When $\nu = 0$ and f is the von Mises density, it can easily be shown that

$$H_1(\kappa) = \{\alpha_3(\kappa) + 3\alpha_2(\kappa) + 3\alpha_1(\kappa) + 1\}^{-1} 2\{\alpha_2(\kappa) + \alpha_1(\kappa)\}$$

where $\alpha_j(\kappa) = E_f(\cos j\Psi) = I_j(\kappa)/I_0(\kappa)$ and $I_j(\kappa)$ is the modified Bessel function of the first kind and order j. Plots of the resulting $\text{acorr}(\hat{\mu}, \hat{\nu}_\mu)$ when κ is fixed (not shown) strongly suggest the success of this venture in the sense that it remains very small for virtually all ν. (Plots of $\text{acorr}(\hat{\mu}, \hat{\nu}_\mu)$ when κ is estimated are functions of ν, κ and ψ and so are not provided.)

If users prefer to use the mode $\mu - 2\nu$ as their measure of location (and hence modal regression, for example), as we suggest they do in the presence of substantial skewness, the same considerations lead to the approximate reparameterisation $(\mu - 2\nu, \nu - H_2(\kappa)\mu)$ where

$$H_2(\kappa) = \{\alpha_3(\kappa) - \alpha_2(\kappa) - \alpha_1(\kappa) + 1\}^{-1} 2\{\alpha_2(\kappa) - \alpha_1(\kappa)\},$$

with the same degree of success. Of course, $\mu - 2\nu$ remains orthogonal to (κ, ψ).

5.3. Remarks on practical issues

Maximisation of the log-likelihood function must generally be conducted numerically, and we have found the direct search simplex algorithm of Nelder and Mead (1965) to perform successfully. When f is the von Mises density, the maximum likelihood estimates of μ and κ for an assumed von Mises distribution together with $\lambda = 0$ and $\nu = 0$ generally provide robust starting values for all but close to uniform samples. Moreover, we have not come across spurious solutions like those reported by Jones and Pewsey (2005) for their extension of the von Mises distribution. Optimisation can be relatively time consuming due to the two inverse operations implicit in (17). However, even when the somewhat slow, but reliable, bisection method is used to carry them out, the simplex rarely takes more than a minute to converge.

In other work of a somewhat similar nature (Jones and Pewsey, 2009), we found the asymptotics to be sufficiently good for general use down to small sample sizes. This is not so much the case here, where we have found the asymptotics to need sample sizes in the hundreds to be reliable. (Contributing to this is that, for small sample sizes at least, maximum likelihood estimates are often on the boundary of the (ν, λ) parameter space.) Instead, in the examples to follow, we base specific inference for parameter values on profile likelihoods, which are a particularly natural tool to use in situations with considerable parameter orthogonality. That said, a potential drawback to the use of profile likelihoods is the computational time taken to arrive at those inferences.

6. Examples

6.1. Example 1: Italian wind directions

As our first example, we present an analysis of the $n = 310$ wind direction measurements referred to by Agostinelli (2007). These data were recorded at Col de la Roa in the Italian Alps and represent the direction of the wind, measured clockwise from north to the nearest 1000th of a degree, at 3:00, 3:15, 3:30, 3:45 and 4:00 am on each day between 29th January and 31st March, 2001. Agostinelli (2007) used robust methods to analyse them whilst Wouters et al. (2009) ran various goodness-of-fit tests for the von Mises distribution on them. As in both references, we treat the data as being independent and thus make no allowance for any potential dependencies that might

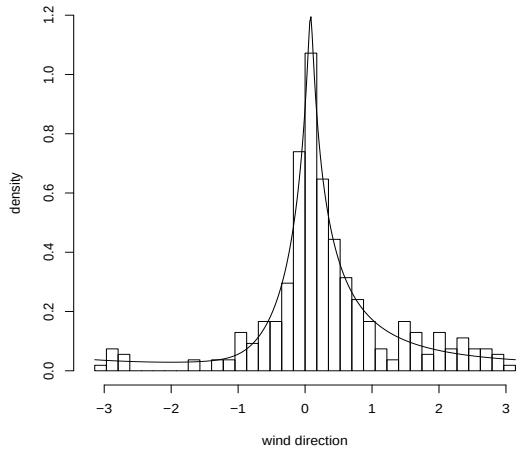

Figure 9: Linear histogram of the wind directions, measured clockwise from north (0) and represented in radians $\in (-\pi, \pi]$, with the fitted density for the full family superimposed. Each of the 36 class intervals has a width equivalent to 10°.

exist between the five observations recorded on the same day or, indeed, on consecutive days. Figure 9 presents a linear histogram of the data with the density of the best fitting member of the full family (17) with f the von Mises density superimposed upon it. The distribution of the data is sharply peaked and skewed towards directions running from north, through east, to south (0 to π radians). The fitted density models the data distribution around the mode well but appears to under-estimate (over-estimate) the frequencies of directions between 1.5 and 2.5 (-2.7 and -1.7) radians a little, i.e. easterlies to south-easterlies (south-westerlies). The underestimation of the easterlies to south-easterlies is perhaps suggestive of multimodality in the data, and the lack of south-westerlies may be a consequence of the geography of the region. Nevertheless, it would be hard to imagine a better unimodal fit to the data.

Results for the maximum likelihood fits of the full family (17) based on the von Mises density and its skew-von Mises ($\lambda = 0$), symmetric ($\nu = 0$) and von Mises ($\nu = 0, \lambda = 0$) sub-models are given in Table 1. There, all three likelihood based diagnostics identify the fit of the full family to be best, followed by that for its symmetric (and as sharply peaked as permitted)

	Model			
Parameter	von Mises	Symmetric	Skew-von Mises	Full family
μ	0.29	0.11	1.17	1.14
κ	1.77	1.90	1.94	1.86
ν	0	0	0.55	0.53
λ	0	1.00	0	0.85
ℓ_{max}	-417.1	-378.5	-399.0	-368.4
AIC	838.1	762.9	804.0	744.8
BIC	845.6	774.1	815.2	759.8
p-value	0.000	0.004	0.000	0.073

Table 1: Parameter estimates for the fits to the wind direction data of, reading from right to left, the full family (17) and its skew-von Mises ($\lambda = 0$), symmetric ($\nu = 0$) and von Mises ($\nu = 0, \lambda = 0$) sub-models. The maximised log–likelihood (ℓ_{max}), AIC and BIC values, and p-value for the chi–squared goodness-of-fit test, are included as fit diagnostics.

sub-family. The p-values of likelihood ratio tests for skew-vonMisesness, symmetry and vonMisesness, calculated using the usual asymptotic chi-squared approximation, are all zero to three decimal places. Also, the chi-squared goodness-of-fit test, calculated by combining, where necessary, the class intervals of the histogram in Figure 9 to obtain expected values of 5 or more, identifies the fit for the full family as being the only viable one of the four; the major contributions to any lack-of-fit result from the over- (under-) represented directions identified previously.

Figure 10 portrays nominally 90%, 95% and 99% confidence regions for (μ, ν), (μ, ν_μ), where $\nu_\mu = \nu - H_1(\kappa)\mu$, and (κ, λ), calculated from their joint profile log-likelihood functions together with the standard asymptotic χ^2_2 result. The highly elongated elliptical contours in its panel (a) reflect the strong positive dependence that exists between the estimates of μ and ν. Under the reparametrisation $\nu_\mu = \nu - H_1(\kappa)\mu$, the dependence between μ and ν_μ is considerably weaker. The dependence between the estimates of κ and λ appears low. Between them, the confidence regions confirm the positive skewness and sharp-peakedness apparent in the histogram. The maximum likelihood estimate of the mode $\mu - 2\nu$ is 0.08, and a 95% profile confidence interval for it is $(0.03, 0.13)$.

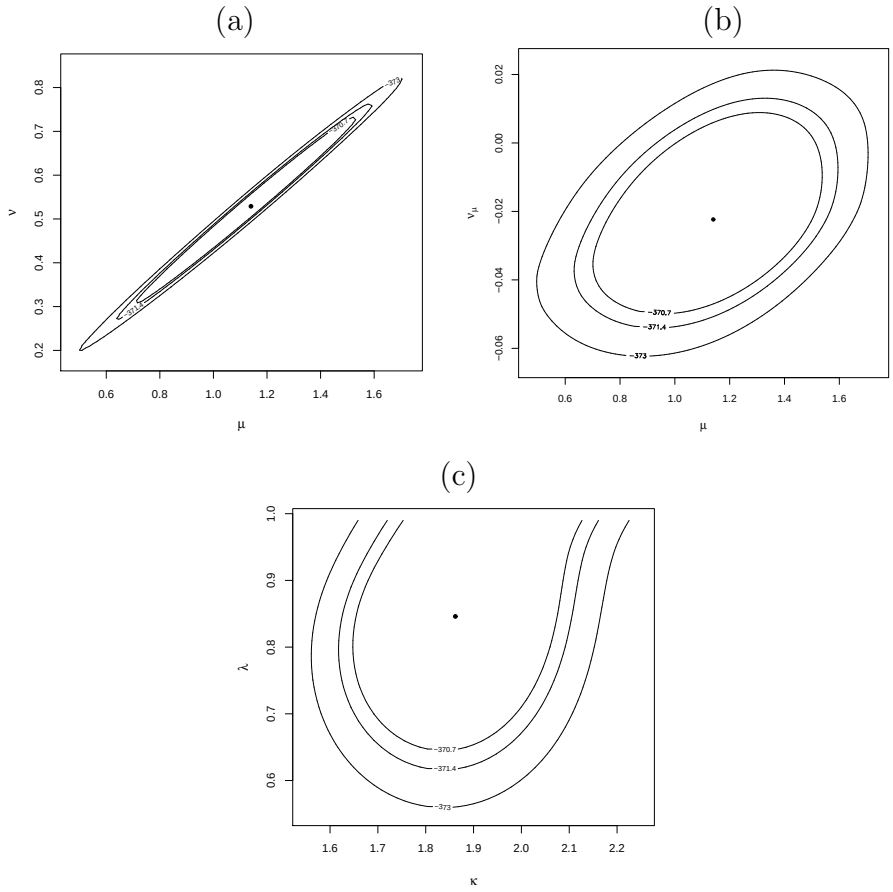

Figure 10: Nominally 90%, 95% and 99% profile log-likelihood based confidence regions of (a) (μ, ν), (b) (μ, ν_μ), where $\nu_\mu = \nu - H_1(\kappa)\mu$, and (c) (κ, λ) for the wind directions. The contours of the three regions lie at $\chi^2_2(0.1)/2 = 4.61/2 = 2.305$, $\chi^2_2(0.05)/2 = 5.99/2 = 2.995$ and $\chi^2_2(0.01)/2 = 9.21/2 = 4.605$, respectively, below the log-likelihood value for the maximum likelihood solution identified by the filled circle. In panel (c), the upper boundary corresponds to the maximum permitted value of $\lambda = 1$.

6.2. Example 2: SIDS data

In our second example we consider the sudden infant death syndrome (SIDS) data of Mooney et al. (2003), placing particular emphasis on the analysis of the data for 1986. Mooney et al. (2003) provide the combined monthly totals of SIDS deaths in England & Wales, Scotland and Northern Ireland for the years 1983-1998. This period covers the introduction of the 'Back to Sleep' campaign in the early 1990s which led to a considerable reduction in the number of SIDS deaths. For each of the 16 years, Mooney et al. (2003, 2006) explored the fits of different cyclical regression models to the monthly numbers of deaths and fitted a variety of circular distributions as models for the time of death. Here we present results for our new family as a model under the latter scenario. When fitting our model to the data for each year we allowed for grouping, different lengths of months and leap years using the multinomial form of the log-likelihood function for grouped data and so there was no need to apply the adjustment to standard 31-day months advocated by Mooney et al. (2003, 2006). A histogram of the $n = 1718$ times of SIDS deaths in 1986 appears in Figure 11. The superimposed densities correspond to the best fitting members of the full family (17) with f the von Mises density, and its symmetric sub-family. Both densities are relatively flat-topped, with any apparent lack-of-fit to the data being associated with the winter months appearing in the centre of the plot.

Table 2 presents analogous results to those presented in Table 1. For these data, the best fits for the full family (17) with f the von Mises density and its symmetric sub-family are judged by the AIC to be on a par. The densities for these two fits are those portrayed in Figure 11. The p-values of 0.026, 0.168 and 0.023 for likelihood ratio tests of skew-vonMisesness, symmetry and vonMisesness, respectively, confirm that the full family does not provide a significant improvement over the symmetric sub-family although it is superior to the other two sub-models. The p-values for the chi-squared goodness-of-fit test in Table 2 also indicate that the fits of these two models are almost equally good. Both are judged to provide adequate fits to the data, as is, marginally, the fit of the skew-von Mises model. The fitted von Mises distribution is rejected as a potential model at the 4% level.

Figure 12 provides profile log-likelihood based, nominally 90%, 95% and 99% confidence regions for (μ, ν) and (κ, λ). A similar plot for the reparametrisation (μ, ν_μ) is not included because it is essentially the same as the plot for (μ, ν) and hence appears to be of little benefit in this case. Figure 12

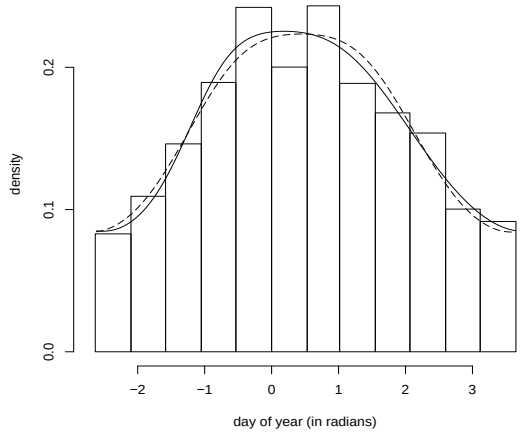

Figure 11: Linear histogram of the times of SIDS deaths in 1986 with the day of the year represented in radians lying between −2.634 (start of August) and 3.649 (end of July). Each bar represents a month, its width being proportional to the length of that month in days. The superimposed densities correspond to the best fitting member of the full family (17) (solid) and its symmetric sub-family (dashed).

	Model			
Parameter	von Mises	Symmetric	Skew-von Mises	Full family
μ	0.41	0.44	0.65	0.62
κ	0.47	0.49	0.47	0.49
ν	0	0	0.25	0.22
λ	0	−0.39	0	−0.36
ℓ_{max}	−4184.5	−4181.7	−4183.2	−4180.8
AIC	8373.0	8369.4	8372.4	8369.5
p-value	0.040	0.144	0.061	0.176

Table 2: Parameter estimates for the fits to the 1986 SIDS data of, reading from right to left, the full family (17) and its skew-von Mises ($\lambda = 0$), symmetric ($\nu = 0$) and von Mises ($\nu = 0, \lambda = 0$) sub-models. The maximised log–likelihood (ℓ_{max}) and AIC values, and p-value for the chi–squared goodness-of-fit test, are included as fit diagnostics. BIC values have not been included because the data are grouped.

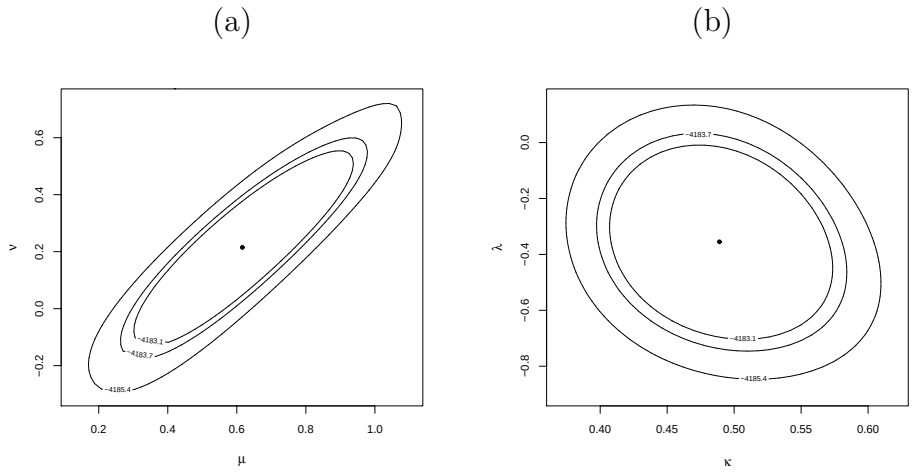

Figure 12: Nominally 90%, 95% and 99% profile log-likelihood based confidence regions of (a) (μ, ν) and (b) (κ, λ) for the 1986 SIDS data. The contours of the three regions lie at $\chi^2_2(0.1)/2 = 4.61/2 = 2.305$, $\chi^2_2(0.05)/2 = 5.99/2 = 2.995$ and $\chi^2_2(0.01)/2 = 9.21/2 = 4.605$, respectively, below the log-likelihood value for the maximum likelihood solution identified by the filled circle.

supports the symmetric, flat-topped fit as a parsimonious model for the data.

Mooney et al.'s (2003, 2006) starting point was that there is a generic asymmetric annual signature for monthly SIDS deaths, with a relatively sharp increase in deaths starting in September and a slower decrease over the spring and summer. The symmetric model identified by us for 1986 clearly does not have such a form. Moreover, when we fitted our family to the SIDS data for the other 15 years, a symmetric flat-topped distribution was also found to fit the data best in 1989. Symmetric sharply peaked distributions were found to fit the data for 1991 and 1998 best, and no gain in fit over the (symmetric) von Mises distribution was found for the 1988, 1990, 1992, 1993, 1995 and 1997 data. Only for 1987 was a skew model found to provide the best fit. Our family failed to provide adequate fits for the 1983–1985, 1994 and 1996 data. For these years, apart from 1983, Mooney et al. (2006) found evidence that two-component von Mises mixtures provided better fits. There would therefore appear to be evidence that other, relatively atypical, factors influenced the times of SIDS deaths in these years, and that just a minority

of years provide any real support for the presumption of asymmetric death patterns.

7. Discussion

We have presented a simple method of generating families of skew and symmetric unimodal distributions on the circle which have a very wide range of shapes and a number of other attractive properties. They are based on ideas of Batschelet (1981) but non-trivially extend them, most notably by involving inverses of Batschelet-type transformations of scale. Of course, as well as modelling single samples of circular data, such univariate families of distributions are valuable in many much more complex models as response (or other component) distributions affording automatically robust estimation of the structures on location (or other) parameters that form the all-important systematic parts of such models.

We would expect likelihood ratio tests of skewness and of the appropriateness of the density f (most usually, of the von Mises distribution) to be effective as omnibus tests of those hypotheses. Initial implementations were successful for large samples (for $n \geq 300$, say) but less so for smaller samples where the tests using the usual asymptotic χ^2 null distributions proved to be too liberal. It is clear that finite sample adjustments are required to the sampling distributions. Once obtained, we would expect them to provide strong competition to the test of symmetry based on circular skewness (Pewsey, 2002) and to tests of vonMisesness such as that of Barndorff-Nielsen and Cox (1979) and others in Wouters et al. (2009); exploration of this is a substantial project that is beyond the scope of this paper.

References

Abe, T. and Pewsey, A. (2011) Sine-skewed circular distributions. *Statist. Pap.*, to appear.

Abe, T., Pewsey, A. and Shimizu, K. (2009) On Papakonstantinou's extension of the cardioid distribution. *Statist. Probab. Lett.,* **79**, 2138–2147.

Agostinelli, C. (2007) Robust estimation for circular data. *Comput. Statist. Data Anal.,* **51**, 5867–5875.

Avérous, J., Fougères, A.L. and Meste, M. (1996) Tailweight with respect to the mode for unimodal distributions. *Statist. Probab. Lett.,* **28**, 367–373.

Azzalini, A. (1985) A class of distributions which includes the normal ones. *Scand. J. Statist.,* **12**, 171–178.

Baker, R. (2008) Probabilistic applications of the Schlömilch transformation. *Commun. Statist. Theory Meth.,* **37**, 2162–2176.

Barndorff-Nielsen, O. and Cox, D.R. (1979) Edgeworth and saddle-point approximations with statistical applications. *J. R. Statist. Soc. Ser. B,* **41**, 279–312.

Batschelet, E. (1965) *Statistical Methods for the Analysis of Problems in Animal Orientation and Certain Biological Rhythms*. Washington, D.C.: American Institute of Biological Sciences.

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

Boshnakov, G.N. (2007) Some measures for asymmetry of distributions. *Statist. Probab. Lett.,* **77**, 1111–1116.

Critchley, F. and Jones, M.C. (2008) Asymmetry and gradient asymmetry functions: density-based skewness and kurtosis. *Scand. J. Statist.,* **35**, 415–437.

Gatto, R. and Jammalamadaka, S.R. (2007) The generalized von Mises distribution. *Statist. Method.,* **4**, 341–353.

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

Jones, M.C. (2009) On distributions generated by transformation of scale. Technical Report 09/08, Statistics Group, Open University, available at `http://stats-www.open.ac.uk/TechnicalReports/TechnicalReportsIntro.htm`.

Jones, M.C. and Anaya-Izquierdo, K.A. (2010) On parameter orthogonality in symmetric and skew models. Revision submitted. Earlier version: Technical Report 09/02, Statistics Group, Open University, available at `http://stats-www.open.ac.uk/TechnicalReports/TechnicalReportsIntro.htm`.

Jones, M.C. and Pewsey, A. (2004) A family of symmetric distributions on the circle. Technical Report 04/07, Statistics Group, Open University, available at `http://stats-www.open.ac.uk/TechnicalReports/TechnicalReportsIntro.htm`.

Jones, M.C. and Pewsey, A. (2005) A family of symmetric distributions on the circle. *J. Amer. Statist. Assoc.,* **100**, 1422–1428.

Jones, M.C. and Pewsey, A. (2009) Sinh-arcsinh distributions. *Biometrika,* **96**, 761–780.

Kato, S. and Jones, M.C. (2010) A family of distributions on the circle with links to, and applications arising from, Möbius transformation. *J. Amer. Statist. Assoc.*, **105**, 249–262.

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

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

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

Mooney, J.A., Helms, P.J. and Jolliffe, I.T. (2003) Fitting mixtures of von Mises distributions: a case study involving sudden infant death syndrome. *Comput. Statist. Data Anal.,* **41**, 505–513.

Mooney, J.A., Jolliffe, I.T. and Helms, P.J. (2006) Modelling seasonally varying data: a case study for Sudden Infant Death Syndrome (SIDS). *J. Appl. Statist.,* **33**, 535–547.

Nelder, J.A. and Mead, R. (1965) A simplex method for function minimization. *Comput. J.,* **7**, 308–313.

O'Hagan, A. (1994) *Kendall's Advanced Theory of Statistics, Volume 2B: Bayesian Inference*. London: Edward Arnold.

Papakonstantinou, V. (1979) Beiträge zur zirkulären Statistik. Ph.D. dissertation, University of Zurich.

Pewsey, A. (2002) Testing circular symmetry. *Canad. J. Statist.*, **30**, 591–600.

Pewsey, A. (2008) The wrapped stable family of distributions as a flexible model for circular data. *Comput. Statist. Data Anal.*, **52**, 1516–1523.

Pewsey, A., Shimizu, K. and de la Cruz, R. (2010) On an extension of the von Mises distribution due to Batschelet. *J. Appl. Statist.,* to appear.

Umbach, D. and Jammalamadaka, S.R. (2009) Building asymmetry into circular distributions. *Statist. Probab. Lett.*, **79**, 659–663.

Wouters, H., Thas, O. and Ottoy, J.-P. (2009) Data-driven smooth tests and a diagnostic tool for lack-of-fit for circular data. *Austral. N. Z. J. Statist.*, **51**, 461–480.