An elliptically symmetric angular Gaussian distribution

P. J. Paine[1] · S. P. Preston[2] · M. Tsagris[3] · Andrew T. A. Wood[2]

Abstract We define a distribution on the unit sphere $\mathbb{S}^{d-1}$ called the elliptically symmetric angular Gaussian distribution. This distribution, which to our knowledge has not been studied before, is a subfamily of the angular Gaussian distribution closely analogous to the Kent subfamily of the general Fisher–Bingham distribution. Like the Kent distribution, it has ellipse-like contours, enabling modelling of rotational asymmetry about the mean direction, but it has the additional advantages of being simple and fast to simulate from, and having a density and hence likelihood that is easy and very quick to compute exactly. These advantages are especially beneficial for computationally intensive statistical methods, one example of which is a parametric bootstrap procedure for inference for the directional mean that we describe.

Keywords Angular Gaussian · Bootstrap · Kent distribution · Spherical distribution

✉ S. P. Preston
simon.preston@nottingham.ac.uk

[1] School of Mathematics and Statistics, University of Sheffield, Hicks Building, Hounsfield Road, Sheffield S3 7RH, UK

[2] School of Mathematical Sciences, University of Nottingham, Nottingham NG7 2RD, UK

[3] Department of Computer Science, University of Crete, 70013 Heraklion, Crete, Greece

1 Introduction

A natural way to define a distribution on the unit sphere $\mathbb{S}^{d-1}$ is to embed $\mathbb{S}^{d-1}$ in $\mathbb{R}^d$, specify a distribution for a random variable $z \in \mathbb{R}^d$, then consider the distribution of z either conditioned to lie on, or projected onto, $\mathbb{S}^{d-1}$. The general Fisher–Bingham and angular Gaussian distributions, defined respectively in Mardia (1975) and Mardia and Jupp (2000) can both be constructed this way by taking z to be multivariate Gaussian in $\mathbb{R}^d$. Then the Fisher–Bingham distribution is the conditional distribution of z conditioned on $\|z\| = 1$, and the angular Gaussian is the distribution of the projection $z/\|z\|$. The choice of the mean, μ, and covariance matrix, V, of z controls the concentration and the shape of the contours of the induced probability density on $\mathbb{S}^{d-1}$.

It is usually not practical to work with the general Fisher–Bingham or angular Gaussian distributions, however, because they have too many free parameters to be identified well by data. This motivates working instead with subfamilies that have fewer free parameters and stronger symmetries.

In the spherical case, $d = 3$, the general distributions have 8 free parameters. Respective subfamilies with 3 free parameters are the Fisher and the isotropic angular Gaussian (IAG) distributions. Both are "isotropic" in the sense that they are rotationally symmetric about the mean direction, i.e., contours on the sphere are small circles centred on the mean direction. Respective subfamilies with 5 free parameters are the Bingham and the central angular Gaussian distributions, both of which are antipodally symmetric.

An important member of this Fisher–Bingham family is the Kent distribution (Kent, 1982). For $d = 3$, it has 5 free parameters, and it has ellipse-like contours on the sphere. This offers a level of complexity well suited to many applications, since the distribution is flexible enough to model anisotropic data yet its parameters can usually be estimated

well from data. To our knowledge, nobody to date has considered its analogue in the angular Gaussian family. The purpose of this paper is to introduce such an analogue, which we call the elliptically symmetric angular Gaussian (ESAG), and establish some of its basic properties.

The motivation for doing so is that in some ways the angular Gaussian family (and hence ESAG) is much easier to work with than the Fisher–Bingham family (and hence the Kent distribution). In particular, simulation is easy and fast, not requiring rejection methods (which are needed for the Fisher–Bingham family Kent et al. 2013), and the density is free of awkward normalising constants, so the likelihood can be computed quickly and exactly. Hence in many modern statistical settings the angular Gaussian family is the more natural choice; see for example Presnell et al. (1998) who use it in a frequentist approach for circular data, and Wang and Gelfand (2013) and Hernandez-Stumpfhauser et al. (2017) who use it in Bayesian approaches for circular and spherical data, respectively.

In the following section, we introduce ESAG, first for general d before specialising to the case $d = 3$.

2 The elliptically symmetric angular Gaussian distribution (ESAG)

2.1 The general angular Gaussian distribution

The angular Gaussian distribution is the marginal distribution of the directional component of the multivariate normal distribution. Let

$$\phi_d(z|\mu, V) = \frac{1}{(2\pi)^{d/2}|V|^{1/2}} \exp\left\{-(z-\mu)^\top V^{-1}(z-\mu)/2\right\} \quad (1)$$

denote the multivariate normal density in $\mathbb{R}^d$ with $d \times 1$ mean vector μ, $d \times d$ covariance matrix V, assumed non-singular, and where $|V|$ denotes the determinant of V. Then, writing $z = ry$, where $r = \|z\| = (z^\top z)^{1/2}$ and $y = z/\|z\| \in \mathcal{S}^{d-1}$, and using $dz = r^{d-1}dr\,dy$, where dy denotes Lebesgue, or geometric, measure on the unit sphere $\mathcal{S}^{d-1}$, and integrating r over $r > 0$, leads to

$$f_{AG}(y) = \frac{C_d}{|V|^{1/2}} \frac{1}{(y^\top V^{-1}y)^{d/2}}$$
$$\times \exp\left[\frac{1}{2}\left\{\frac{(y^\top V^{-1}\mu)^2}{y^\top V^{-1}y} - \mu^\top V^{-1}\mu\right\}\right]$$
$$\times \mathcal{M}_{d-1}\left\{\frac{y^\top V^{-1}\mu}{(y^\top V^{-1}y)^{1/2}}\right\}, \quad (2)$$

where $C_d = 1/(2\pi)^{(d-1)/2}$ and, for real α,

$$\mathcal{M}_{d-1}(\alpha) = \int_{u=0}^\infty u^{d-1}\frac{1}{(2\pi)^{1/2}}\exp\left\{-(u-\alpha)^2/2\right\}du. \quad (3)$$

Direct calculations show that

$$\mathcal{M}_0(\alpha) = \Phi(\alpha), \quad \mathcal{M}_1(\alpha) = \alpha\Phi(\alpha) + \phi(\alpha)$$
$$\mathcal{M}_2(\alpha) = (1+\alpha^2)\Phi(\alpha) + \alpha\phi(\alpha), \quad (4)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal probability density function and cumulative density function, respectively; more details about the $\mathcal{M}_d(\alpha)$ are given in Sect. A.1.

2.2 An elliptically symmetric subfamily

The subfamily of (2) that we shall call the elliptically symmetric angular Gaussian distribution (ESAG) is defined by the two conditions

$$V\mu = \mu, \quad (5)$$
$$|V| = 1, \quad (6)$$

under which the angular Gaussian density (2) simplifies to

$$f_{ESAG}(y) = \frac{C_d}{(y^\top V^{-1}y)^{d/2}}\exp\left[\frac{1}{2}\left\{\frac{(y^\top\mu)^2}{y^\top V^{-1}y} - \mu^\top\mu\right\}\right]$$
$$\mathcal{M}_{d-1}\left\{\frac{y^\top\mu}{(y^\top V^{-1}y)^{1/2}}\right\}. \quad (7)$$

From (5), the positive definite matrix V has a unit eigenvalue. If the other eigenvalues are

$$0 < \rho_1 \leq \cdots \leq \rho_{d-1}, \quad (8)$$

then the inverse of V can be written

$$V^{-1} = \xi_d\xi_d^\top + \sum_{j=1}^{d-1}\xi_j\xi_j^\top/\rho_j, \quad (9)$$

where $\xi_1, \ldots, \xi_{d-1}$ and $\xi_d = \mu/\|\mu\|$ is a set of mutually orthogonal unit vectors. Moreover, constraint (6) means that

$$\prod_{j=1}^{d-1}\rho_j = 1. \quad (10)$$

Once the d parameters in μ are fixed, then from (5) and (6) there are $d - 2$ remaining degrees of freedom for the eigenvalues of V, and $d(d-1)/2 - (d-1)$ degrees of freedom for its unit eigenvectors. The total number of free parameters

is thus $(d-1)(d+2)/2$, the same as for the multivariate normal in a tangent space $\mathbb{R}^{d-1}$ to the sphere.

Condition (5) imposes symmetry about the eigenvectors of V. Without loss of generality, suppose that the eigenvectors are parallel to the coordinates axes; that is, each element of the vector ξ_j equals 0 except the jth which equals 1. Then if $y = (y_1, \ldots, y_d)^\top$,

$$y^\top \xi_d = y_d, \quad y^\top V^{-1} y = y_d^2 + \sum_{j=1}^{d-1} y_j^2/\rho_j.$$

In this case, the density (7) depends only on y_j through y_j^2 for $j = 1, \ldots, d-1$. Consequently, the density is invariant with respect to sign changes of the $y_1, \ldots, y_{d-1}$, that is,

$$f_{\mathrm{ESAG}}(\pm y_1, \ldots, \pm y_{d-1}, y_d) = f_{\mathrm{ESAG}}(y_1, \ldots, y_{d-1}, y_d),$$

which implies reflective symmetry about 0 along the axes defined by $\xi_1, \ldots, \xi_{d-1}$. This type of symmetry is implied by ellipse-like contours of constant density inscribed on the sphere, and such contours arise when the density (7) is unimodal. Whether the density is unimodal depends on the nature of the stationary point at $y = \mu/\|\mu\|$, which is characterised by the following proposition.

Proposition 1 *Write $\alpha = \|\mu\|$ and assume without loss of generality that (8) holds. Then (i) the ESAG always has a stationary point at $y = \xi_d = \mu/\alpha$, and (ii) the stationary point at $y = \xi_d$ is a local maximum if $\rho_{d-1} \leq \mathcal{H}_d(\alpha)$ and a local minimum if $\rho_{d-1} > \mathcal{H}_d(\alpha)$, where*

$$\mathcal{H}_d(\alpha) = 1 + \left\{ \alpha^2 + (d-1)\alpha \mathcal{M}_{d-2}(\alpha)/\mathcal{M}_{d-1}(\alpha) \right\}/d. \quad (11)$$

The proof of Proposition 1 is given in Appendix A. 2.

We conjecture that if the stationary point at $y = \xi_d$ is a local maximum, then it is also a global maximum, and that in this case the distribution is unimodal; and in the case that the stationary point is a local minimum, then the distribution is bimodal. A rigorous proof appears difficult, but the conjecture is strongly supported by some extensive numerical investigations that we have performed and describe as follows. For each of a wide variety of combinations of ESAG parameters with $d = 3$ — in particular, choosing for $(\alpha, \gamma_1, \gamma_2)^\top$ values on a $9 \times 9 \times 9$ rectangular lattice on $[0.2, 20] \times [-5, 5] \times [-5, 5]$ — we performed numerical maximisation to find

$$y^{\mathrm{max}} = \mathrm{argmax}_y \, f_{\mathrm{ESAG}}(y).$$

Using the Manopt implementation (Boumal et al., 2014) of the trust region approach of Absil et al. (2007), for each $(\alpha, \gamma_1, \gamma_2)^\top$, we performed the optimisation multiple times from distinct starting points with the rationale that if the distribution is indeed multimodal, then optimisations from different starting points will converge to different local optima. We chose to use 42 different starting points since this enabled the points to be exactly equi-spaced on $\mathbb{S}^2$ using the method of Teanby (2006). For each $(\alpha, \gamma_1, \gamma_2)^\top$, we hence computed $y_1^{\mathrm{max}}, \ldots, y_{42}^{\mathrm{max}}$. Instances that converged to the same mode had values of y^{max} that were not quite identical, owing to the finite tolerance of the numerical optimisation. To account for this, we identified the number of modes according to clustering of the $\{y_i^{\mathrm{max}}\}$, designating the distribution unimodal if and only if $\|y_i^{\mathrm{max}} - \bar{y}^{\mathrm{max}}\| < 10^{-6}$ for all i, where $\bar{y}^{\mathrm{max}} = (1/42) \sum_i y_i^{\mathrm{max}}$. In cases identified as non-unimodal by this criterion, we used k-means clustering to identify $k = 2$ clusters; in each such case, every y_i^{max} was within a distance 10^{-6} of its cluster centre indicating bimodality. In agreement with the conjecture, amongst the $9^3 = 729$ parameter cases we considered, in every 553 cases with $\rho_{d-1} \leq \mathcal{H}_d(\alpha)$, the foregoing procedure identified the distribution to be unimodal, and in every 176 cases with $\rho_{d-1} > \mathcal{H}_d(\alpha)$, it identified the distribution to be bimodal.

The next proposition concerns the limiting distribution of a sequence of unimodal ESAG distributions as the sequence becomes more highly concentrated. Without loss of generality, we fix $\xi_d = (0, \ldots, 0, 1)^\top$, take $\xi_1, \ldots, \xi_{d-1}$ to be the other coordinate axes and define

$$y = \left\{ y_1, \ldots, y_{d-1}, (1 - y_1^2 - \cdots - y_{d-1}^2)^{1/2} \right\}. \quad (12)$$

Neglecting the hemisphere defined by y_d negative is no drawback when considering $\alpha = \|\mu\| \to \infty$ as follows, because in this limit the distribution becomes increasingly concentrated about $y = \xi_d$.

Proposition 2 *Assume the conditions (5) and (6), and therefore (10), hold, where the ρ_j are assumed to be fixed, and suppose that $\alpha = \|\mu\| \to \infty$. Then*

$$\alpha \tilde{y} \xrightarrow{d} N_{d-1}(0_{d-1}, \Omega), \quad (13)$$

where $\tilde{y} = (y_1, \ldots, y_{d-1})^T$ and
$\Omega = diag\{\rho_1^{-1}, \ldots, \rho_{d-1}^{-1}\} = \sum_{j=1}^{d-1} \rho_j^{-1} \xi_j \xi_j^T$.

Remark 1 In the general case, we replace the coordinate vectors $\xi_1, \ldots, \xi_d$ by an arbitrary orthonormal basis, and then the limit distribution lies in the vector subspace spanned by $\xi_1, \ldots, \xi_{d-1}$.

Remark 2 Proposition 2 is noteworthy because it is atypical for high-concentration limits within the angular Gaussian family to be Gaussian.

2.3 A parameterisation of ESAG for $d = 3$

An important practical question is how to specify a convenient parameterisation for the matrix V so that it satisfies the constraints (5) and (6). With $d = 3$, such a V has two free parameters.

We first define a pair of unit vectors $\tilde{\xi}_1$ and $\tilde{\xi}_2$ which are orthogonal to each other and to the mean direction $\xi_3 = \mu/\|\mu\|$:

$$\tilde{\xi}_1 = \left(-\mu_0^2, \mu_1\mu_2, \mu_1\mu_3\right)^\top /(\mu_0\|\mu\|) \quad \text{and}$$
$$\tilde{\xi}_2 = (0, -\mu_3, \mu_2)^\top /\mu_0, \tag{14}$$

where $\mu = (\mu_1, \mu_2, \mu_3)^\top$ and $\mu_0 = (\mu_2^2 + \mu_3^2)^{1/2}$; then $\tilde{\xi}_1$ and $\tilde{\xi}_2$ in (14) are smooth functions of μ except at $\mu_2 = \mu_3 = 0$, where there is indeterminacy. To enable the axes of symmetry, ξ_1 and ξ_2, to be an arbitrary rotation of $\tilde{\xi}_1$ and $\tilde{\xi}_2$, we define

$$\xi_1 = \cos\psi\, \tilde{\xi}_1 + \sin\psi\, \tilde{\xi}_2 \quad \text{and}$$
$$\xi_2 = -\sin\psi\, \tilde{\xi}_1 + \cos\psi\, \tilde{\xi}_2, \tag{15}$$

where $\psi \in (0, \pi]$ is the angle of rotation. Substituting ξ_1 and ξ_2 from (15) into (9), and putting $\rho_1 = \rho$ and $\rho_2 = 1/\rho$ where $\rho \in (0, 1]$, gives the parameterisation

$$V^{-1} = \left(\rho^{-1}\cos^2\psi + \rho\sin^2\psi\right)\tilde{\xi}_1\tilde{\xi}_1^\top$$
$$+ \left(\rho^{-1}\sin^2\psi + \rho\cos^2\psi\right)\tilde{\xi}_2\tilde{\xi}_2^\top$$
$$+ \tfrac{1}{2}(\rho^{-1} - \rho)\sin 2\psi \left(\tilde{\xi}_1\tilde{\xi}_2^\top + \tilde{\xi}_2\tilde{\xi}_1^\top\right)$$
$$+ \xi_3\xi_3^\top. \tag{16}$$

The disadvantage that ρ and ψ are restricted can be resolved by writing (16) in terms of unrestricted parameters γ_1 and γ_2 as follows.

Lemma 1 *Define* $\gamma = (\gamma_1, \gamma_2)^\top$ *by*

$$\gamma_1 = 2^{-1}(\rho^{-1} - \rho)\cos 2\psi \quad \text{and}$$
$$\gamma_2 = 2^{-1}(\rho^{-1} - \rho)\sin 2\psi, \tag{17}$$

then V^{-1} *in (16) is*

$$V^{-1} = I_3 + \gamma_1\left(\tilde{\xi}_1\tilde{\xi}_1^\top - \tilde{\xi}_2\tilde{\xi}_2^\top\right) + \gamma_2\left(\tilde{\xi}_1\tilde{\xi}_2^\top + \tilde{\xi}_2\tilde{\xi}_1^\top\right)$$
$$+ \left\{(\gamma_1^2 + \gamma_2^2 + 1)^{1/2} - 1\right\}\left(\tilde{\xi}_1\tilde{\xi}_1^\top + \tilde{\xi}_2\tilde{\xi}_2^\top\right). \tag{18}$$

Henceforth, we will use this parameterisation. For a random variable y with ESAG distribution, we will write $y \sim$
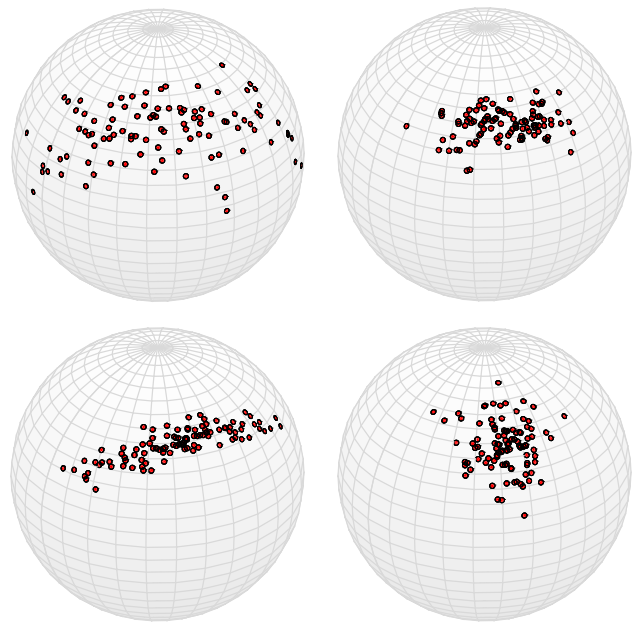

Fig. 1 Samples of 100 observations from some example ESAG distributions with parameters (clock-wise from top left): $\mu = (-1, -2, 2)^\top$, $\gamma = (-1, 1)^\top$; $\mu = (-2, -4, 4)^\top$, $\gamma = (-1, 1)^\top$; $\mu = (-2, -4, 4)^\top$, $\gamma = (0, 0)^\top$; $\mu = (-2, -4, 4)^\top$, $\gamma = (-3, 1)^\top$

ESAG(μ, γ). The rotationally symmetric isotropic angular Gaussian corresponds to

$$V = I_3 \Leftrightarrow \gamma = (0, 0)^\top = 0$$

hence we will write ESAG$(\mu, 0) \equiv$ IAG(μ).

Remark 3 (Simulation.) To simulate $y \sim$ ESAG(μ, γ), simulate $z \sim N(\mu, V)$ where $V = V(\mu, \gamma)$ is defined in (18) then set $y = z/\|z\|$.

Figure 1 shows some examples of samples from the ESAG distribution for various values of the parameters μ and γ.

Remark 4 (A test for rotational symmetry.) For a sample of observations $y_1, \ldots, y_n$ assumed independent and identically distributed, a standard large-sample likelihood ratio test can be used to test $H_0 : y_i \sim$ IAG(μ) vs $H_1 : y_i \sim$ ESAG(μ, γ). Let $\hat{l}_0$ and $\hat{l}_1$ be the values of the maximised log-likelihoods under H_0 and H_1, respectively. The models are nested and differ by two degrees of freedom, and by Wilks' theorem, when n is large the statistic $T = 2(\hat{l}_1 - \hat{l}_0)$ has approximately a χ_2^2 distribution if H_0 is true, and H_0 is rejected for large values of T. The null distribution can alternatively be approximated using simulation by the parametric bootstrap, that is, by simulating a sample of size n from the null model H_0 at the maximum likelihood estimate of the parameters, computing the test statistic, T, and then repeating this a large number, say B, times. The empirical distribution of the resulting bootstrapped statistics $T_1^*, \ldots, T_B^*$ approximates the null distribution of T.
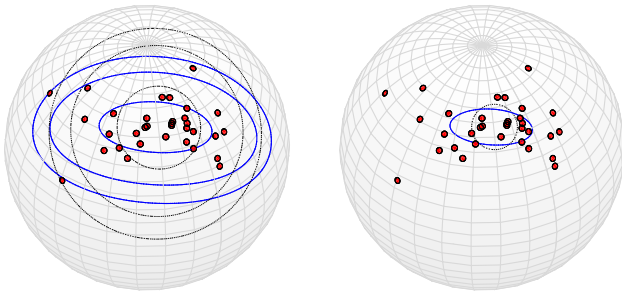
Fig. 2 Estimates of the Earth's historic magnetic pole position; see Sect. 2.5 for a description of the data. (Left) shows contours of constant probability density for fitted models corresponding to 99, 95 and 50% coverage and (Right) shows 95% confidence regions for the mean direction calculated as described in Sect. 2.4. ESAG is shown with solid blue lines and the IAG with black dotted lines

Remark 5 (Parameter orthogonality.) The parameter vectors μ and γ are orthogonal, in the sense that

$$\mathcal{I}_{\mu,\gamma} = \mathcal{I}_{\gamma,\mu}^\top = 0_{3,2},$$

where $\mathcal{I}_{\mu,\gamma} = -\mathrm{E}_{\mu,\gamma}\left[\partial^2 \log f_1/\partial\mu\partial\gamma^\top\right]$. Moreover, the directional and magnitudinal components μ and γ, that is, $\|\mu\|$, $\mu/\|\mu\|$, $\|\gamma\|$ and $\gamma/\|\gamma\|$, are all mutually orthogonal. The proofs follow easily from the symmetries of f_{ESAG} and are omitted.

Often in applications, there is particular interest in the directional mean, $m = \mu/\|\mu\|$. A parametric bootstrap procedure to construct confidence regions for m, which exploits both the ease of simulation and the parameter orthogonality, is as follows.

2.4 Parametric bootstrap confidence regions for $m = \mu/\|\mu\|$

Let $l(\mu, \gamma)$ denote the log-likelihood for a sample $y_1, \ldots, y_n$ each of which is assumed to be an independent ESAG(μ, γ) random variable and define

$$\Sigma(\mu, \gamma) = \left(-\frac{\partial^2 l(\mu, \gamma)}{\partial\mu\partial\mu^\top}\right)^{-1}.$$

Denote the maximum likelihood estimate of (μ, γ) by $(\hat{\mu}, \hat{\gamma})$, let $\hat{m} = \hat{\mu}/\|\hat{\mu}\|$, $\hat{\Sigma} = \Sigma(\hat{\mu}, \hat{\gamma})$ and let $\hat{\xi}_1$ and $\hat{\xi}_2$ denote the maximum likelihood estimates of the axes of symmetry (15), such that $\hat{m}$, $\hat{\xi}_1$ and $\hat{\xi}_2$ are mutually orthogonal unit vectors. Define the matrix $\hat{\xi} = (\hat{\xi}_1, \hat{\xi}_2)^\top$, then test statistic

$$T(m) = m^\top \hat{\xi}^\top \left(\frac{1}{\|\hat{\mu}\|^2}\hat{\xi}\hat{\Sigma}\hat{\xi}^\top\right)^{-1}\hat{\xi}m, \tag{19}$$
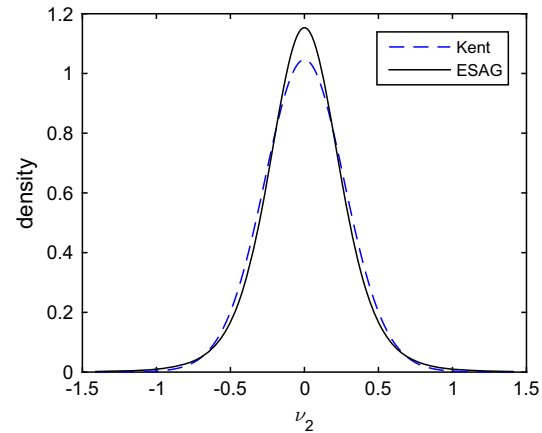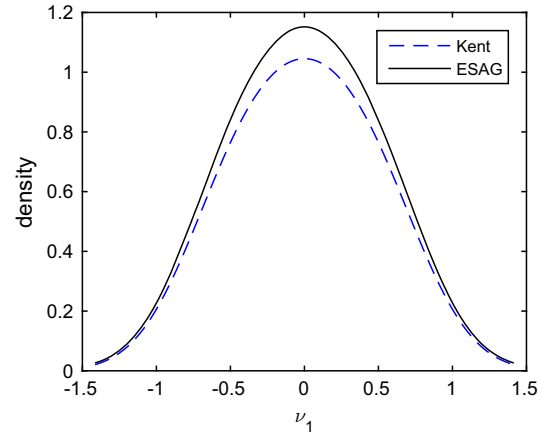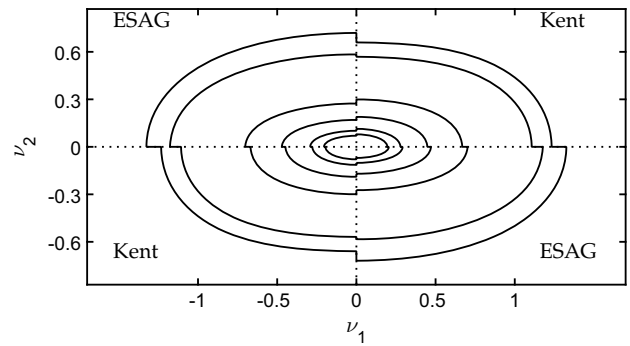

Fig. 3 A comparison of ESAG and Kent densities with parameters matched by fitting each distribution to the same dataset; see Sect. 3 for details. (Top) shows a tangent plane projection, in terms of tangent plane coordinates ν_1 and ν_2, of contours of constant probability density for coverage levels 95, 90, 50, 25, 10, 5%. (Centre) and (Bottom) show transects of the two probability density functions with $\nu_2 = 0$ and $\nu_1 = 0$, respectively

is suitable for defining confidence regions for m as $\{m \in \mathbb{S}^2 : T(m) \leq c\}$; see Fisher et al. (1996) for discussion of test statistics of this form in the context of non-parametric bootstrap procedures. For a given significance level, α, the constant c can be determined as follows: simulate B bootstrap samples each of size n from ESAG$(\hat{\mu}, \hat{\gamma})$, and hence with $m = \hat{m}$, and for each sample compute the test statis-

tic (19), with $\hat{\xi}, \hat{\mu}, \hat{\Sigma}$ replaced by corresponding quantities calculated from the bootstrap sample; then c is the $(1 - \alpha)$ quantile of the resulting statistics $T_1^*(\hat{m}), \ldots, T_B^*(\hat{m})$. Examples of confidence regions calculated by this algorithm are shown in Fig. 2 (right).

2.5 An example: estimates of the historic position of Earth's magnetic pole

This dataset contains estimates of the position of the Earth's historic magnetic pole calculated from 33 different sites in Tasmania (Schmidt 1976). The data are shown in Fig. 2 with contours of fitted IAG and ESAG distributions. The maximum likelihood estimates of the parameters are respectively

$$\hat{\mu}^\top = (-2.18, 0.94, 3.06),$$

and

$$(\hat{\mu}^\top, \hat{\gamma}^\top) = (-2.33, 1.11, 3.34, 0.17, -0.78).$$

Twice the difference in maximised log-likelihoods equals 14.12, which when referred to a χ_2^2 distribution (see Remark 4) corresponds to a p-value of less than 10^{-3}, indicating strong evidence in favour of ESAG over the IAG.

3 A comparison of ESAG with the Kent distribution

Figure 3 shows contours and transects of the densities of ESAG and Kent distributions. The parameter values for each are computed by fitting the two models to a large sample of independent and identically distributed data from ESAG(μ, γ), with $\mu = (0, 0, 2.5)^\top$ and $\gamma = (0.75, 0)^\top$. For the inner contours ESAG is more anisotropic than the matched Kent distribution and appears slightly more peaked at the mean. Besides these small differences, the figure shows that ESAG and Kent distributions are very similar distributions in this example, as we have found them to be more generally. Indeed, preliminary results, not presented here, suggest that for typical sample sizes it is usually very difficult to distinguish between them using a statistical criterion. This warrants making the modelling choice between using the Kent distribution or the ESAG on grounds of practical convenience. The Kent distribution is a member of the exponential family, but its density involves a non-closed-form normalising constant, and simulation requires a rejection algorithm (Kent et al. 2013). The ESAG distribution has a density that is less tidy than the Kent density, hence less suited to computing moment estimators, etc., but this is not much of a drawback given that its density can be computed exactly so that the exact likelihood can be easily maximised. Moreover,

Table 1 Results of simulation study for fitting the ESAG and Kent distributions to both ESAG and Kent simulated data

	Data: ESAG $\mu = (0, 0, 1.18)^\top$ $\gamma = (0.29, 0)$	Data: Kent $\Gamma = \mathcal{I}_3,$ $\kappa = 2.16, \beta = 0.5.$
Sim. time (s)	2.1799	28.456
Fit: ESAG		
Fit time (s)	1431.8	1493.2
Error ($\hat{m}$)	0.1127	0.1290
Error ($\hat{\xi}_1$)	0.0988	0.3017
Error ($\hat{\xi}_2$)	0.1248	0.3051
Fit: Kent		
Fit time (s)	13896.7	13860.2
Error ($\hat{m}$)	0.1302	0.1289
Error ($\hat{\xi}_1$)	0.0992	0.2901
Error ($\hat{\xi}_2$)	0.1248	0.2938
	Data: ESAG $\mu = (0, 0, 2.6)^\top$ $\gamma = (0.53, 0)$	Data: Kent $\Gamma = \mathcal{I}_3,$ $\kappa = 7.38, \beta = 1.34.$
Sim. time (s)	3.2818	37.618
Fit: ESAG		
Fit time (s)	1867.4	1834.3
Error($\hat{m}$)	0.0584	0.0563
Error($\hat{\xi}_1$)	0.116	0.1487
Error($\hat{\xi}_2$)	0.122	0.1522
Fit Kent		
Fit time (s)	23922.1	24191.5
Error($\hat{m}$)	0.0587	0.0562
Error($\hat{\xi}_1$)	0.1248	0.1339
Error($\hat{\xi}_2$)	0.1261	0.1379
	Data: ESAG $\mu = (0, 0, 3.8)^\top$ $\gamma = (1.3, 0)$	Data: Kent $\Gamma = \mathcal{I}_3,$ $\kappa = 20.61, \beta = 8.9.$
Sim. time (s)	1.6446	39.4057
Fit: ESAG		
Fit time (s)	2222.3	2203.4
Error($\hat{m}$)	0.0428	0.0405
Error($\hat{\xi}_1$)	0.0453	0.0534
Error($\hat{\xi}_2$)	0.0581	0.0625
Fit: Kent		
Fit time (s)	32038.5	31541.3
Error($\hat{m}$)	0.0403	0.0415
Error($\hat{\xi}_1$)	0.0462	0.0522
Error($\hat{\xi}_2$)	0.0592	0.0624

The results are for three different cases, involving pairs of ESAG and Kent distributions with parameters "matched" in the sense that the were fitted to an initial common set of data. For each combination of model and parameter set, we simulated $b = 500$ Monte Carlo samples of $n = 100$ observations. The measures of error of $\hat{m}, \hat{\xi}_1$ and $\hat{\xi}_2$ are defined in the text. Simulation times, given in seconds, are cumulative over the $b = 500$ Monte Carlo runs

simulating from ESAG is particularly quick and easy (see Remark 3).

Table 1 shows the results of a simulation study, including computational timings, for fitting ESAG and Kent densities to ESAG and Kent simulated data. To approximate the Kent normalising constant when fitting the Kent distribution, we use a saddlepoint approximation method (see Kume et al. 2013), and for simulating from the Kent distribution, we use the rejection method of Kent et al. (2013). A notion of accuracy of the fitted model is how well the mean direction of the fitted model, $\hat{m}$, corresponds with the population mean direction, m. A measure we use for this is

$$\sqrt{2\{(1 - E(\hat{m}^\top m)\}}, \tag{20}$$

with the expectation approximated by Monte Carlo; hence in the tables, we report

$$\text{Error}(\hat{m}) = \sqrt{2\left\{1 - b^{-1} \sum_{i=1}^{b} \left(\hat{m}_{(i)}^\top m\right)\right\}}$$

where $\hat{m}_{(i)}$ is the mean direction of the fitted model for the ith run out of b Monte Carlo runs. We also consider accuracy of the major and minor axes of the fitted model. Since the signs of $\hat{\xi}_1$ and $\hat{\xi}_2$ are arbitrary, in this case we define

$$\text{Error}(\hat{\xi}_1) = \sqrt{2\left\{1 - b^{-1} \sum_{i=1}^{b} \left|\hat{\xi}_{1(i)}^\top \xi_1\right|\right\}},$$

and similar for $\hat{\xi}_2$.

Note that in interpreting the results in Table 1, the different simulation times of ESAG and Kent should be compared across columns, whereas the fitting times and accuracies should be compared across rows.

The results show, as expected, that the accuracy of $\hat{m}$, $\hat{\xi}_1$, and $\hat{\xi}_2$ is typically better when the data-generating model is fitted. However, the accuracy is not dramatically worse when the non-data-generating model is fitted, i.e., when ESAG is fitted to Kent data, or the Kent distribution is fitted to ESAG data. There is a very notable difference in computation times between ESAG and Kent: for both simulation and fitting, ESAG is typically more than an order of magnitude faster than Kent.

4 Conclusion

In the pre-computer days of statistical modelling, the Fisher–Bingham family was perhaps favoured over the angular Gaussian family on account of having a simpler density,

which makes it more amenable to constructing classical estimators such as moment estimators. However, in the era of computational statistics, the less simple form of the angular Gaussian density is hardly a barrier and is more than compensated by having a normalising constant that is trivial to evaluate. The likelihood can consequently be computed quickly and exactly, and maximised directly. Wang and Gelfand (2013) have recently argued in favour of the general angular Gaussian distribution as a model for Bayesian analysis of circular data. For spherical data, a major obstacle to using the general angular Gaussian distribution is that its parameters are poorly identified by the data. The ESAG subfamily overcomes this problem, and is a direct analogy of the Kent subfamily of the general Fisher–Bingham distribution. Besides having a tractable likelihood, the ease and speed with which ESAG can be simulated makes it especially well suited to methods of simulation-based inference. Natural wider applications of ESAG include using it as an error distribution for spherical regression models with anisotropic errors; for classification on the sphere (as a model for class-conditional densities); and for clustering spherical data (based on ESAG mixture models). Code written in MATLAB for performing calculations in this paper is available at the second author's web page.

A Proofs

A.1 Properties of the function $\mathcal{M}_d(\alpha)$

Integrating (3) by parts, we obtain

$$\begin{aligned}
\mathcal{M}_{d-1}(\alpha) &= \left[d^{-1} u^d \phi(u - \alpha) \right]_{u=0}^{\infty} \\
&\quad + d^{-1} \int_0^\infty u^d (u - \alpha) \phi(u - \alpha) du \\
&= d^{-1} \mathcal{M}_{d+1}(\alpha) - d^{-1} \alpha \mathcal{M}_d(\alpha),
\end{aligned}$$

which implies that

$$\mathcal{M}_{d+1}(\alpha) = \alpha \mathcal{M}_d(\alpha) + d \mathcal{M}_{d-1}(\alpha). \tag{21}$$

Moreover, differentiating (3) with respect to α, and exchanging the order of differentiation and integration on the RHS, we obtain

$$\mathcal{M}_d'(\alpha) = \mathcal{M}_{d+1} - \alpha \mathcal{M}_d(\alpha), \tag{22}$$

where a prime denotes differentiation; and, using (21) and (22) together, it is found that

$$\mathcal{M}_d'(\alpha) = d\mathcal{M}_{d-1}(\alpha). \tag{23}$$

Further points to note, which are easily established by induction, are that we can write

$$\mathcal{M}_d(\alpha) = \mathcal{P}_d(\alpha)\Phi(d)(\alpha) + \mathcal{Q}_d(\alpha)\phi(\alpha),$$

where $\mathcal{P}_d(\alpha)$ is a polynomial of order d and $\mathcal{Q}_d$ is a polynomial of order $d-1$; and $\mathcal{P}_d(\alpha)$ inherits the properties (21) and (23) of $\mathcal{M}_d(\alpha)$, while $\mathcal{Q}_d(\alpha)$ inherits (21) but not (23); instead of (23), we have $\mathcal{Q}_d'(\alpha) = \mathcal{Q}_{d+1}(\alpha) - \mathcal{P}_d(\alpha)$. One relevant property of $\mathcal{P}_d(\alpha)$ is that the coefficient of its leading term α^d is 1; this follows from either (21) or (23) applied repeatedly to $\mathcal{P}_d(\alpha)$. As an easy consequence, we obtain a result that is used below: for any fixed integer $d \geq 0$,

$$\mathcal{M}_d(\alpha) \sim \alpha^d \quad \text{as } \alpha \to \infty. \tag{24}$$

A.2 Proof of Proposition 1

From (8) it follows that $1/\rho_{d-1} \leq 1/\rho_j$ for $j = 1, \ldots, d-2$. As a consequence, we can restrict attention to y of the form $y = (0, \ldots, u, (1-u^2)^{1/2})$ where u is in a neighbourhood of 0. Substituting this form of y into (7) and differentiating, it is seen that

$$\left.\frac{\partial \log f_1}{\partial u}(u)\right|_{u=0} = 0,$$

and, using (22), it is found that

$$\left.\frac{\partial^2 \log f_1}{\partial u^2}(u)\right|_{u=0} = \delta d - \alpha^2 + \alpha^2 \delta$$
$$+ (d-1)(\delta-1)\alpha\mathcal{M}_{d-2}(\alpha)/\mathcal{M}_{d-1}(\alpha), \tag{25}$$

where $\delta = (\rho_{d-1} - 1)/\rho_{d-1}$. The point $u = 0$, which corresponds to $y = \xi_d$, is a maximum or minimum depending on whether (25) is negative or positive and, using this fact, Proposition 2.1 follows after some elementary further manipulations. $\quad\square$

A.3 Proof of Proposition 2

First, consider the exponent in (7). Using (9) it follows after some elementary calculations that

$$\frac{1}{2}\left[\frac{(y^\top\mu)^2}{yV^{-1}y} - \mu^\top\mu\right] = -\frac{1}{2}\alpha^2\frac{\sum_{j=1}^{d-1}y_j^2/\rho_j}{1 + \sum_{j=1}^{d-1}y_j^2/\rho_j}. \tag{26}$$

Now define $u_j = \alpha y_j$, $j = 1, \ldots, d-1$ and suppose that $\alpha \to \infty$ while the u_j remain fixed. Then, since $\sum_{j=1}^d (y^\top\xi_j)^2 = 1$, it follows that as $\alpha \to \infty$, $y \to 1$. Consequently as the ρ_j are remaining fixed, it is easy to see that

$$y^\top V^{-1}y \to 1. \tag{27}$$

Moreover, from (12) it follows that $y^\top\mu \sim \alpha$ and therefore, from (24),

$$\mathcal{M}_{d-1}\left\{y^\top\mu/\left(y^\top V^{-1}y\right)^{1/2}\right\} \sim \alpha^{d-1}. \tag{28}$$

Letting $\alpha \to \infty$ and substituting (26), (27) and (28) into (7) we obtain pointwise convergence to the $(d-1)$-dimensional multivariate Gaussian density $\phi_{d-1}(u|0_{d-1}, \Omega)$ multiplied by a factor α^{d-1}; this factor is the Jacobian of the transformation from $(y_1, \ldots, y_{d-1})^\top$ to $(u_1, \ldots, u_{d-1})^\top$. $\quad\square$

A.4 Proof of Lemma 1

Using the standard trigonometric results $\cos^2\psi = (1 + \cos 2\psi)/2$ and $\sin^2\psi = (1 - \cos 2\psi)/2$, and making use of the fact that $\tilde{\xi}_1\tilde{\xi}_1^\top + \tilde{\xi}_2\tilde{\xi}_2^\top + \xi_3\xi_3^\top = I_3$ because $\tilde{\xi}_1$, $\tilde{\xi}_2$ and ξ_3 are mutually orthogonal unit vectors, (16) becomes

$$V^{-1} = I_3 + \tfrac{1}{2}(\rho^{-1} - \rho)\cos 2\psi \left(\tilde{\xi}_1\tilde{\xi}_1^\top - \tilde{\xi}_2\tilde{\xi}_2^\top\right)$$
$$+ \tfrac{1}{2}(\rho^{-1} - \rho)\sin 2\psi \left(\tilde{\xi}_1\tilde{\xi}_2^\top + \tilde{\xi}_2\tilde{\xi}_1^\top\right)$$
$$+ \tfrac{1}{2}(\rho^{-1} + \rho - 2)\left(\tilde{\xi}_1\tilde{\xi}_1^\top + \tilde{\xi}_2\tilde{\xi}_2^\top\right) \tag{29}$$

From (17), $(\gamma_1^2 + \gamma_2^2)^{1/2} = (\rho^{-1} - \rho)/2$, and we may solve for ρ to obtain $\rho = (1 + \gamma_1^2 + \gamma_2^2)^{1/2} - (\gamma_1^2 + \gamma_2^2)^{1/2}$, and with some elementary further calculation we see that

$$(\rho^{-1} + \rho - 2)/2 = (1 + \gamma_1^2 + \gamma_2^2)^{1/2} - 1. \tag{30}$$

Finally, substituting (17) and (30) into (29), we obtain (18).

References

Absil, P.A., Baker, C.G., Gallivan, K.A.: Trust-region methods on Riemannian manifolds. Found. Comput. Math. **7**(3), 303–330 (2007). doi:10.1007/s10208-005-0179-9

Boumal, N., Mishra, B., Absil, P.A., Sepulchre, R.: Manopt, a Matlab toolbox for optimization on manifolds. J. Mach. Learn. Res. vol. 15, pp. 1455–1459. http://www.manopt.org (2014)

Fisher, N., Hall, P., Jing, B.Y., Wood, A.T.A.: Improved pivotal methods for constructing confidence regions with directional data. J. Am. Stat. Assoc. **91**(435), 1062–1070 (1996)

Hernandez-Stumpfhauser, D., Breidt, F., van der Woerd, M.: The general projected normal distribution of arbitrary dimension: modeling and Bayesian inference. Bayesian Anal. **12**(1), 113–133 (2017)

Kent, J.T.: The Fisher-Bingham distribution on the sphere. J. R. Stat. Soc. Series B **44**, 71–80 (1982)

Kent, J.T., Ganeiber, A., Mardia, K.V.: A new method to simulate the Bingham and related distributions in directional data analysis with applications. arXiv preprint arXiv:1310.8110 (2013)

Kume, A., Preston, S.P., Wood, A.T.A.: Saddlepoint approximations for the normalizing constant of Fisher-Bingham distributions on products of spheres and Stiefel manifolds. Biometrika **100**(4), 971–984 (2013)

Mardia, K.V.: Statistics of directional data (with discussion). J. R. Stat. Soc. Series B **37**, 349–393 (1975)

Mardia, K.V., Jupp, P.E.: Directional Statistics. Wiley, Hoboken (2000)

Presnell, B., Morrison, S.P., Littell, R.C.: Projected multivariate linear models for directional data. J. Am. Stat. Assoc. **93**, 1068–1077 (1998)

Schmidt, P.: The non-uniqueness of the Australian Mesozoic palaeomagnetic pole position. Geophys. J. Int. **47**(2), 285–300 (1976)

Teanby, N.: An icosahedron-based method for even binning of globally distributed remote sensing data. Comput. Geosci. **32**(9), 1442–1450 (2006)

Wang, F., Gelfand, A.: Directional data analysis under the general projected normal distribution. Stat. Methodol. **10**(1), 113–127 (2013)