Title

Predictive inference for directional data

Permalink

Authors

Jammalamadaka, S Rao
SenGupta, Ashis

Predictive inference for directional data

S. Rao Jammalamadaka[a,*,1], Ashis SenGupta[a,b]

[a]*Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106, USA*
[b]*Indian Statistical Institute, Calcutta, West Bengal, India*

Received March 1998

Abstract

Predictive density for a future observation is derived when the given data comes from circular or spherical distributions. Model robustness of these predictive densities for the circular case is exhibited. Predictive intervals corresponding to the highest density regions are derived and an example is given. © 1998 Elsevier Science B.V. All rights reserved

Keywords: Bessel functions; Circular normal distribution; Predictive density; von Mises–Fisher distribution; Symmetric wrapped stable family

1. Introduction and summary

If $Y = (Y_1, \ldots, Y_n)$ is the "past" and has density $p_\theta(y)$, what can we say about the distribution of the "future" (a single value, a set of values or a function of these), say z? It is generally agreed that the single most important aim of statistics is to predict the nature of a future occurrence based on empirical evidence, i.e. the observed past.

There are several approaches to finding predictive likelihoods and predictive densities. Excellent reviews can be found in Butler (1986) and Bjornstad (1990). Main approaches include (i) methods based on conditioning through sufficiency, (ii) Bayes predictive densities and (iii) maximum or profile likelihood methods. Specifically, let $Y = (Y_1, \ldots, Y_n) \sim p_\theta(y)$ and let $Y' = (Y_1', \ldots, Y_m')$ denote future observations from the same model. The quantity of interest Z is Y' itself or is a function of Y'. Let $p_\theta(y, z)$ denote the p.d.f. of (Y, Z). It can be argued that this is the likelihood of the unknowns (Z, θ) given y i.e.

$$L_y(z, \theta) = p_\theta(y, z).$$

The goal is to eliminate the nuisance parameter(s) θ and obtain the predictive density of Z given Y. Three basic approaches include:

 (i) *Sufficiency approach.* Suppose $T(Y, Z)$ is a minimal sufficient statistic for θ based on both Y and Z.

Then

$$p^{\mathrm{cond}}(z \mid y) \propto \frac{p(y, z, \theta)}{p(T(y, z), \theta)}$$

$$= p(y, z \mid T(y, z)),$$

where factorization criterion makes θ drop out of the last expression.

(ii) *Bayes approach.*

$$p^{\mathrm{Bayes}}(z \mid y) = \frac{\int p_{\theta}(z, y) \pi(\theta) \, \mathrm{d}\theta}{\int p_{\theta}(y) \pi(\theta) \, \mathrm{d}\theta},$$

where $\pi(\theta)$ is a prior for θ.

(iii) *Maximization.*

$$p^{\mathrm{max}}(z \mid y) = \sup_{\theta} p_{\theta}(y, z) = L_y(z, \hat{\theta}(y, z)),$$

where $\hat{\theta}(y, z)$ is the MLE based on both y and z.

In many natural and physical sciences, one measures directions, for instance, the direction of wind (two-dimensional), the direction of the earth's magnetic pole (three-dimensional). Also, many periodic/cyclical phenomena like biorhythms can be represented as measurements on a circle of given circumference. Indeed, prediction is clearly as important in the context of directional data as it is for linear measurements and we apply the above ideas in this situation.

We study some of these approaches for obtaining the predictive density of a future observation, given n past observations from the same population. We use the Bayesian approach in Section 5 while the rest of the paper uses the sufficiency approach. For the circular case, we derive the predictive densities assuming that the underlying distribution is the von Mises or circular normal (CN) while for the spherical data, we assume that the parent population is the von Mises–Fisher model. Some generalizations that include the multimodal as well as axial distributions, are also considered. We will show that conditional sufficiency argument of Fisher (1956) leads to exact solutions. We also observe that the Bayesian approach which often becomes intractable, also leads to an elegant and computationally appealing representation, in large samples.

For CN data, we find that the predictive density is symmetric around the mean of the past data and is nearly CN although less concentrated. We also demonstrate through exact numerical computations as well as simulations that this predictive density derived under the assumption of CN population is robust against a large class of other symmetric unimodal circular distributions, called the symmetric wrapped stable family, introduced by Levy (1939). We compute the highest density regions (HDRs) and illustrate the results with an example from a real-life data set.

2. Distribution on a circle

We first consider the most commonly used circular distribution – the von Mises distribution (also called the circular normal distribution), $CN(\mu_0, \kappa)$, with the p.d.f.,

$$f(\theta; \mu_0, \kappa) = [2\pi I_0(\kappa)]^{-1} \exp\{\kappa \cos(\theta - \mu_0)\}, \quad 0 \leqslant \theta, \ \mu_0 < 2\pi, \ \kappa > 0,$$

where

$$I_0(\kappa) = \sum_{r=0}^{\infty} \frac{1}{(r!)^2} \left(\frac{\kappa}{2}\right)^{2r}$$

is the modified Bessel function of the first kind and order zero.

Theorem 1. *Under the sufficiency approach,*
(1) *The predictive density of θ_{n+1} is given by*

$$g(\theta_{n+1} \mid \theta_1, \dots, \theta_n) \propto \frac{1}{\phi_{n+1}(r_{n+1}^2)},$$ (2.1)

where

$$\phi_n(r^2) = \int_0^{\infty} u J_0(ru) J_0^n(u) \, du,$$ (2.2)

$J_0(x)$ *being the standard Bessel function of zeroth order.*
(2) $g(\cdot)$ *above is a symmetric, unimodal density with its mode at $\bar{\theta}_n$, the MLE of μ_0 based on the past n observations, $\theta_1, \theta_2, \dots, \theta_n$.*
(3) *The predictive density, $g(\cdot)$, is proportional to the von Mises model, $f(\theta_{n+1}; \bar{\theta}_n, 2\bar{R})$ for large n.*

Proof. In this model, $T_n = (C_n, S_n)$ with $S_n = \sum_i \sin \theta_i$ and $C_n = \sum_i \cos \theta_i$ can be seen to be the minimal sufficient statistic whose distribution is given by

$$p(S_n, C_n) = [2\pi I_0^n(\kappa)]^{-1} \exp\{\kappa(\nu S_n + \mu C_n)\} \phi_n(S_n^2 + C_n^2),$$

where $\nu = \sin \mu_0, \mu = \cos \mu_0$ and $\phi_n(\cdot)$ is defined in (2.2) above (see, for instance, Mardia, 1972, p. 97).
(1) Following the conditioning approach with the minimal sufficient statistic (C_{n+1}, S_{n+1}),

$$p^{\text{cond}}(\theta_{n+1} \mid \theta_1, \dots, \theta_n) = \frac{p(\theta_1, \dots, \theta_{n+1})}{p(C_{n+1}, S_{n+1})}$$

$$= \left(\frac{1}{2\pi}\right)^n \frac{1}{\phi_{n+1}(C_{n+1}^2 + S_{n+1}^2)}$$ (2.3)

after some simplification, which proves (1).
(2) Noting $C_{n+1} = C_n + \cos\theta_{n+1}$, $S_{n+1} = S_n + \sin\theta_{n+1}$ and $C_n = R_n \cos\bar{\theta}_n, S_n = R_n \sin\bar{\theta}_n$ where $\bar{\theta}_n$ is the direction of the resultant vector (C_n, S_n), the expression (2.3) is equal to

$$[(2\pi)^n \phi_{n+1}(R_n^2 + 1 + 2R_n \cos(\theta_{n+1} - \bar{\theta}_n))]^{-1}$$

and the statement made in (2) follows from this.
(3) From Rayleigh's approximation, it follows that

$$\phi_n(r^2) \approx \frac{2}{n} \exp\left(-\frac{r^2}{n}\right)$$

for large n. Using this, it can be seen that the predictive density for large n is proportional to

$$\exp\left\{\left(-\frac{2r_n}{n+1}\right) \cos(\theta_{n+1} - \bar{\theta}_n)\right\}$$

which is essentially a von Mises distribution with center at $\bar{\theta}_n$, the estimate of μ based on the past. Indeed, $\hat{\kappa} = 2r_n/(n+1)$ can be seen to be the approximate MLE of the concentration parameter κ for smaller values of κ. The predictive density is less concentrated around $\bar{\theta}_n$ than the "plug-in" (but incorrect) method would suggest, namely a $CN(\hat{\theta}_n, \hat{\kappa}_{MLE})$. $\square$

3. Generalized von Mises-type distributions

3.1. ℓ-modal von Mises distribution

A multi (ℓ)-modal von Mises density $CNM(\mu_0, \kappa, \ell)$ can be obtained (see Mardia, 1972, p. 73) from the von Mises density and has the p.d.f.

$$p_M(\theta) = \{2\pi I_0(\kappa)\}^{-1} \exp\{\kappa \cos \ell(\theta - \mu_0)\}, \quad 0 \leqslant \theta < 2\pi, \; 0 \leqslant \mu_0 < 2\pi/\ell. \tag{3.1}$$

For known ℓ, a sufficient statistic for (μ_0, κ) is (S_ℓ, C_ℓ) where

$$S_\ell = \sum_i \sin \ell\theta_i, \qquad C_\ell = \sum_i \cos \ell\theta_i.$$

Let

$$R_\ell^2 = S_\ell^2 + C_\ell^2.$$

We may thus apply the above approach here also, provided the joint density of (S_ℓ, C_ℓ) can be derived. Towards this, we have

Lemma 1. *The joint density of* (S_ℓ, C_ℓ), *where* $\alpha_1, \ldots, \alpha_n$ *are i.i.d. observations from the* $CNM(\mu_0, \kappa, \ell)$ *distribution is*:

$$g(s_\ell, c_\ell) = [2\pi I_0^n(\kappa)]^{-1} \exp(\kappa v s_\ell + \kappa \mu c_\ell) \mathscr{L}_n(r_\ell^2), \tag{3.2}$$

where $v = \sin \mu_0, \mu = \cos \mu_0$ *and*

$$\mathscr{L}_n(r_\ell^2) = \ell \int_0^\infty J_0(u r_\ell) J_0^n(u) u \, \mathrm{d}u = \ell \phi_n(r_\ell^2), \tag{3.3}$$

$$J_0(u) = (2\pi)^{-1} \int_0^{2\pi} \exp\{(iu) \cos \ell\theta\} \, \mathrm{d}\theta$$

$$= (2\pi)^{-1} \int_0^{2\pi} \exp\{(iu) \cos \theta\} \, \mathrm{d}\theta.$$

Proof. Let $x_\ell = r_\ell \cos \ell\theta, y_\ell = r_\ell \sin \ell\theta$ be the rectangular coordinate representation of an arbitrary two-dimensional variable. If $\psi(t_1, t_2)$ represents its characteristic function, then by the inversion theorem, the p.d.f. of (x_ℓ, y_ℓ) is given by,

$$f(x_\ell, y_\ell) = (2\pi)^{-2} \int_{-\infty}^\infty \int_{-\infty}^\infty \exp(-it_1 x - it_2 y) \psi(t_1, t_2) \, \mathrm{d}t_1 \, \mathrm{d}t_2. \tag{3.4}$$

Writing $t_1 = \rho \cos \ell\phi, t_2 = \rho \sin \ell\phi$,

$$\psi(t_1, t_2) = E[\exp\{i\rho r_\ell \cos \ell(\theta - \phi)\}] \equiv \bar{\psi}_\ell(\rho, \phi),$$

say. The joint density of r_ℓ and θ is,

$$f_1(r_\ell, \theta) = (2\pi)^{-2} r_\ell \ell \int_0^{2\pi} \int_0^\infty \exp\{-i\rho \ r_\ell \cos\ell(\theta - \phi)\} \rho \ \bar\psi_\ell(\rho, \phi) \, d\rho \, d\phi. \tag{3.5}$$

Then invoking change of variables $(x_\ell, y_\ell) \to (r_\ell, \theta)$, using (3.5), integrating out θ and interchanging the order of integrations, the density of r_ℓ is given by

$$f_1(r_\ell) = (2\pi)^{-2} r_\ell \ell \int_0^\infty \int_0^{2\pi} \left[\int_0^{2\pi} \exp\{-i\rho r_\ell \cos\ell(\theta - \phi)\} d\theta \right] \rho \ \bar\psi_\ell(\rho, \phi) \, d\phi \, d\rho$$

$$= (2\pi)^{-1} r_\ell \ell \int_0^\infty \int_0^{2\pi} \rho J_0(\rho r_\ell) \ \bar\psi_\ell(\rho, \phi) \, d\phi \, d\rho, \tag{3.6}$$

where $J_0(u)$ is as defined in (3.3) above.

Now, let $\theta_1, \ldots, \theta_n$ be i.i.d. random variables following the uniform distribution on $(0, 2\pi]$. Then the c.f. of $(\cos\ell\theta, \sin\ell\theta)$ is,

$$\bar\psi_\ell^*(\rho) = \bar\psi_\ell(\rho, \phi)|_{\phi, r_\ell = 1} \equiv J_0(\rho).$$

Then the joint c.f. of (S_ℓ, C_ℓ) is $[\bar\psi_\ell^*(\rho)]^n$ and inverting this as in (3.4) we get the p.d.f. of R_ℓ is,

$$p(r_\ell) = (2\pi)^{-1} \ell r_\ell \int_0^\infty \int_0^{2\pi} \rho J_0(-\rho r_\ell)[\bar\psi_\ell^*(\rho)]^n \, d\phi \, d\rho$$

$$= (2\pi)^{-1} \ell r_\ell \int_0^\infty \int_0^{2\pi} \rho J_0(-\rho r_\ell) J_0^n(\rho) \, d\phi \, d\rho$$

$$= r_\ell \mathscr{L}_n(r_\ell^2) \tag{3.7}$$

(see Section 4.5.5 of Mardia, 1972).

Let, $S_\ell = \sum_i \sin\ell\theta_i = R_\ell \sin\bar x_\ell, C_\ell = \sum_i \cos\ell\theta_i = R_\ell \cos\bar x_\ell$. From this discussion, it is seen that, for uniform samples, the distribution of $\bar x_\ell$ is uniform and $\bar x_\ell$ and R_ℓ are independent. Thus the joint p.d.f. of $(\bar x_\ell, R_\ell)$ is

$$r_\ell \mathscr{L}_n(r_\ell^2)/(2\pi), \quad 0 \leqslant \bar x_\ell < 2\pi, R_\ell > 0. \tag{3.8}$$

Thus, the joint p.d.f. of (S_ℓ, C_ℓ) on transforming, is

$$g_0(s_\ell, c_\ell) = (2\pi)^{-1} \mathscr{L}_n(r_\ell^2). \tag{3.9}$$

Now consider the general case where $\theta_1, \ldots, \theta_n$ are i.i.d. random variables following the multimodal von Mises distribution CNM (μ_0, κ, ℓ). The joint p.d.f. $g_\kappa(s_\ell, c_\ell)$ of (S_ℓ, C_ℓ) is then,

$$g_\kappa(s_\ell, c_\ell) = \{I_0(\kappa)\}^{-n} \exp\{\kappa(vs_\ell + \mu c_\ell)\} \int_{\{\theta : \Sigma_i \sin l\theta_i = s_\ell, \Sigma_i \cos l\theta_i = c_\ell\}} \prod_i (d\theta_i/2\pi). \tag{3.10}$$

But the integral represents the density of (C_ℓ, S_ℓ) when $\theta_1, \ldots, \theta_n$ is random sample from the circular uniform distribution on $(0, 2\pi]$, which is given in (3.9). On substituting (3.9) in (3.10), (3.2) follows. $\quad\square$

We thus have the following:

Theorem 2. *The predictive density of θ_{n+1} is given by*

$$g(\theta_{n+1}|\theta_1, \ldots, \theta_n) \propto \frac{1}{\mathscr{L}_{n+1}(r_{\ell, n+1}^2)}.$$

Proof. This follows exactly as in part (1) of Theorem 1 by virtue of the above Lemma. □

Remark. Arnold (1941) suggested an angular distribution of the von Mises type for a circular random variable restricted to only a semi-circular arc, i.e. $\theta \in (0, \pi]$. Its generalization to the t-angular case is given by

$$g(\theta^*) = [t/(2\pi I_0(\kappa))]\exp\{\kappa \cos t(\theta^* - \mu_0)\}, \quad 0 \leqslant \theta^*, \quad \mu_0 < 2\pi/t.$$

The predictive density of θ^* is equivalent to that in part (1) of Theorem 1 above as $\theta = t\theta^* \sim CN(\mu_0, \kappa)$, and t is known.

4. von Mises–Fisher distribution on a sphere

We consider the analogue of Theorem 1, when the directions are three-dimensional, i.e. the observations are taken on a sphere. The most commonly used unimodal distribution in such a case is called the von Mises–Fisher distribution $F(\kappa, \mu)$, with the p.d.f.

$$f_2(x; \kappa, \mu) = C(\kappa) \cdot \exp\{\kappa x' \mu\}, \quad x'x = \mu'\mu = 1, \quad x, \mu \in S_3.$$

with $C(\kappa) = \kappa/(4\pi \sinh \kappa)$. Then, we have:

Theorem 3. *The predictive density of* $x_{n+1} = (x_{1,n+1}, x_{2,n+1}, x_{3,n+1})'$ *is given by*

$$g(x_{n+1}|x_1, \ldots, x_n) \propto \frac{1}{h_{n+1}(r_{n+1})/r_{n+1}^2}, \tag{4.1}$$

where, R_n^2 *is the squared length of the vector resultant in three-dimensions defined by*

$$R_n^2 = X_{1n}^2 + X_{2n}^2 + X_{3n}^2, \quad X_{in} = \sum_{j=1}^{n} x_{ij}, \quad i = 1, 2, 3$$

and

$$h_n(r) = \frac{r}{(n-2)!2^{n-1}} \sum_{k=0}^{n} (-1)^k \binom{n}{k} (n - r - 2k)^{n-2}.$$

Proof. Observe trivially that $x_n = (x_{1n}', x_{2n}', x_{3n}')'$ is a sufficient statistic for (μ, κ). The joint density of X is (cf. (8.6.30) of Mardia, 1972):

$$h(x_1, x_2, x_3) = \{2C(\kappa)\}^n e^{\kappa x_3} h_0(x_1, x_2, x_3),$$

where $h_0(\cdot)$ is the joint p.d.f. of X under the isotropic case, i.e. when the corresponding angles (θ, ϕ) form a random sample from the uniform distribution:

$$g(\theta, \phi) = (4\pi)^{-1} \sin \theta, \quad 0 \leqslant \theta < \pi, \quad 0 \leqslant \phi < 2\pi.$$

Then proceeding exactly as in part (1) of Theorem 1, it suffices to compute $h_0(\cdot)$, which is given in (8.6.24) of Mardia as

$$h_0(r) = (4\pi)^{-1} h_n(r)/r^2.$$

The proof is then immediate. □

5. Bayesian predictive density

The Bayes predictive density of a future observation u_f given the past $u_1, \ldots, u_n$ is given by

$$g(u_f | u_1, \ldots, u_n) = \int_\kappa \int_\mu p(u_f | \mu, \kappa) p(\mu, \kappa | u_1, \ldots, u_n) \, d\mu \, d\kappa.$$

Let the priors be given by

$$p(\mu) = C_p(\kappa_0) \exp(\kappa_0 \mu' \mu_0)$$

and

$$p(\kappa) \propto I_0(\kappa)^{-c} \exp(\kappa \Delta).$$

Assume (μ, κ) and the observations $(u_1, \ldots, u_n, u_f)$ are independent. Defining

$$R = \left\| \sum u_i \right\|, \qquad u_0 = \sum_1^n u_i / R,$$

we have

$$g(\cdot) \propto \int_0^\infty \int_0^{2\pi} C_p(\kappa) \exp(\kappa \mu' u_f) C_p^n(\kappa) \exp(\kappa R \mu' u_0)$$

$$\times C_p(\kappa_0) \exp(\kappa_0 \mu'_0 \mu) I_0^{-C}(\kappa) \exp(\kappa \Delta) \, d\mu \, d\kappa,$$

$$= K \int_0^\infty C_p^{n+1}(\kappa) C_p^{-1}(r_2) I_0^{-C}(\kappa) \exp(\kappa \Delta) \, d\kappa,$$

where

$$R_2 = \| \kappa(u_f + R u_0) + \kappa_0 \mu_0 \|,$$

$$C_p(\kappa) = (2\pi)^{-p/2} \kappa^{p/2-1} I_{p/2-1}^{-1}(\kappa)$$

and K is the normalizing constant. This needs to be numerically evaluated. $g(u_f | u_1, \ldots, u_n)$ above may simplify somewhat if we take as prior

$$p(\mu_1, \kappa) \propto \kappa^{-1/2} I(\mu \in S_p)$$

based on non-informative (vague) prior for μ and Jeffrey's prior for κ (which is not a proper prior here).

5.1. Case of large κ

In case there is a prior reason to believe that only large values of κ need to be considered, some further simplifications result. These are somewhat analogous to the standard Laplace saddle-point-type approximations of the posterior.
Let

$$p(\kappa) \propto I_0^{-C}(\kappa) \exp(\kappa \Delta) I(\kappa > M), \quad M > 0, \text{large and known}.$$

For large κ, it is known that for all $v \geqslant 0$,

$$I_v(\kappa) \simeq (2\pi\kappa)^{-1/2} \exp(\kappa).$$

Using this approximation, we get

$$C_p(\kappa) \simeq \kappa^{(p-1)/2} \exp(-\kappa)(2\pi)^{-(p-1)/2}.$$

Then,

$$g(u_f \mid u_1, \ldots, u_n) \propto \int_M^\infty \kappa^{(n+1)(p-1)+c)/2} [\|\kappa(u_f + Ru_0) + \kappa_0\mu_0\|]^{-(p-1)/2}$$

$$\exp[\kappa\{(\Delta - n - c - 1) + \|\kappa(u_f + Ru_0) + \kappa_0\mu_0\|\}] \, d\kappa$$

Specializing to $p = 2$, and choosing $\kappa_0 = 0$, i.e. uniform prior for μ, we simplify to

$$g(u_f|u_1, \ldots, u_n) \propto \int_M^\infty \kappa^{(n+c+1)/2} [\kappa\|(u_f + Ru_0)\|]^{-1/2} \exp[\kappa\{(\Delta - n - c - 1) + \|u_f + Ru_0\|\}] \, d\kappa$$

$$= \|u_f + Ru_0\|^{-1/2} \int_M^\infty \kappa^{(n+c)/2} \exp[\kappa\{(\Delta - n - c - 1) + \|u_f + Ru_0\|\}] \, d\kappa, \qquad (5.1)$$

where B is the normalizing constant.

Note that, taking $C = 0$ and $\Delta = -1$ yields $p(\kappa) \propto \exp(-\kappa)$. For large n, (5.1) behaves as a usual incomplete Gamma integral, i.e.

$$g(u_f|u_1, \ldots, u_n) = B\|u_f + Ru_0\|^{-1/2} \Gamma_{(n+c+2)/2}\{M/(n + c + 1 - \Delta - \|u_f + Ru_0\|)\} \qquad (5.2)$$

where B is the normalizing constant, i.e.

$$B^{-1} = \int_{S_p} \|u_f + Ru_0\|^{-1/2} \Gamma_{(n+c+2)/2} \{M/(n + c + 1 - \Delta - \|u_f + Ru_0\|)\} \, du_f.$$

(5.2) is quite simple to calculate, especially for the CN population with $p = 2$. Such Bayes predictive densities can and have been numerically evaluated and provide encouraging results.

6. Robustness against symmetric wrapped stable family

Exact computation of the predictive densities is done through numerical integrations using Gauss–Laguerre 64-point quadrature in IMSL and are plotted in Figs. 1–4. The predictive density curve is superimposed on the true CN curve in Fig. 1 while in Figs. 2–4, we show what happens when the predictive density of Theorem 1, based on the CN distribution is used when the data is actually from various members of the symmetric wrapped stable family. All these results clearly indicate the encouraging performance and model-robustness of the predictive density as derived above. As can be expected, these results improve considerably for larger sample sizes.

Note. Observations from a wrapped stable distribution, $g(\theta)$,

$$g(\theta) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n=1}^\infty \rho^{n^\alpha} \cos(n\theta), \quad 0 \leqslant \theta < 2\pi, \ \rho \geqslant 0, \ 0 < \alpha \leqslant 2$$

are simulated by generating observations from the corresponding stable distribution and then taking mod 2π.
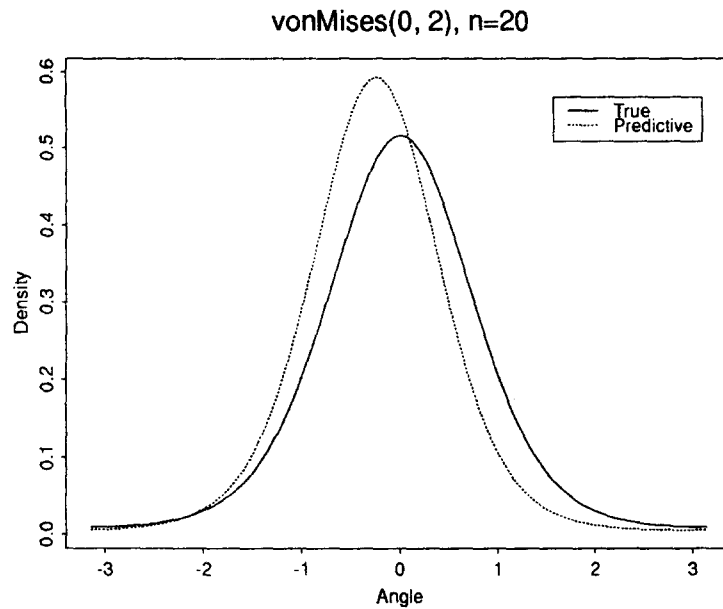
vonMises(0, 2), n=20

Fig. 1. Predictive density for vonMises (0, 2), $n = 20$.
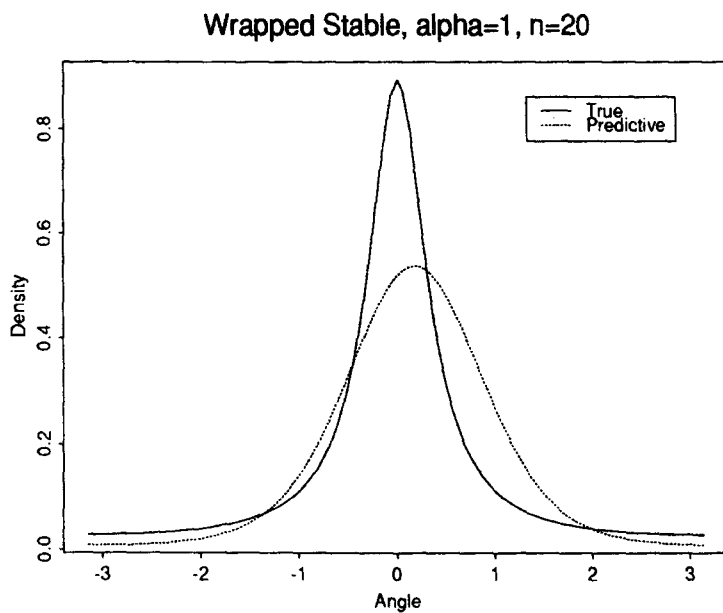
Wrapped Stable, alpha=1, n=20

Fig. 2. Predictive density for wrapped stable with $\alpha = 1.0$, $n = 20$.

7. HDR computation and an example

Consider the following example from Batschelet (1981) giving data on the time of day of major traffic accidents in a major city. The time (h) is converted to angle (η) (in degrees) by the formula $\eta = 15 \times h$. The $n = 21$ observations are given in Table 1.
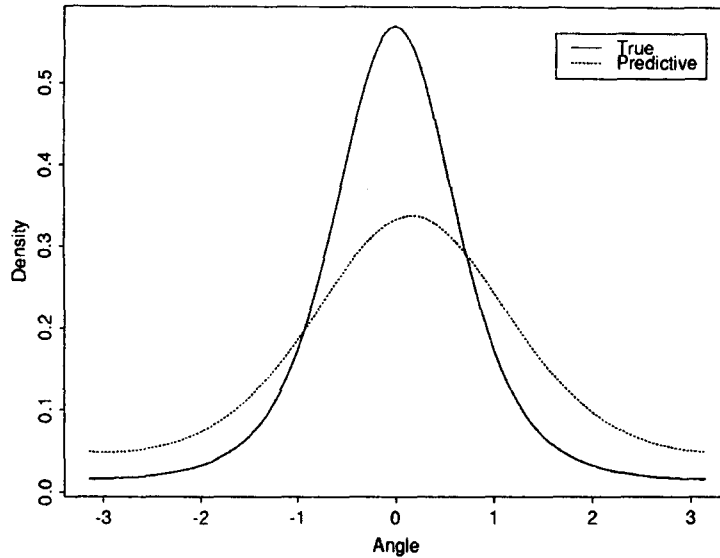
Wrapped Stable, alpha=1.5, n=20

Fig. 3. Predictive density for wrapped stable with $\alpha = 1.5$, $n = 20$.
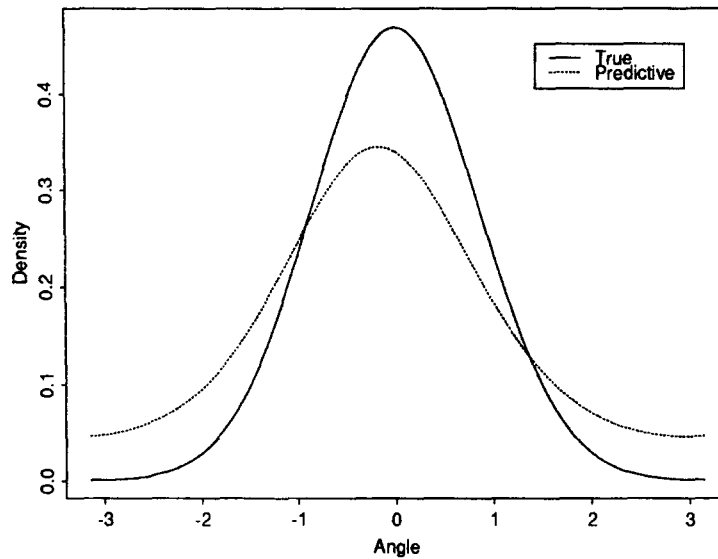
Wrapped Stable, alpha=2, n=20

Fig. 4. Predictive density for wrapped stable with $\alpha = 2.0$, $n = 20$.

Assuming a von Mises distribution of the data, we get $\hat{\mu} = 248.59°$ which translates to 16 h 34 min in time scale. Also, $\hat{\kappa} = 0.7021$ is the MLE of the concentration parameter. The 90% HDR is given by (104.50°, 32.68°), i.e. (6 h 58 min, 2 h 11 min) while the 75% HDR is given by (9 h 45 min, 23 h 24 min).

The large widths of the predictive intervals in this example can be explained by the fact that the data is not very concentrated and the sample size is not very large.

Table 1

h : min	Angle	h : min	Angle
00 : 56	14	16 : 44	251
03 : 08	47	17 : 04	256
04 : 52	73	17 : 20	260
07 : 16	109	17 : 24	261
08 : 08	122	18 : 08	272
10 : 00	150	18 : 16	274
11 : 24	171	18 : 56	284
12 : 08	182	19 : 32	293
13 : 28	202	20 : 52	313
14 : 16	214	22 : 08	332
16 : 20	245		

References

Arnold, K.J., 1941. On Spherical Probability Distributions, Ph.D. Thesis, MIT, Boston.

Batschelet, E., 1981. Circular Statistics in Biology. Academic Press, London.

Bjornstad, B.J.F., 1990. Predictive likelihood: a review. Statist. Sci. 5, 242–265.

Butler, R., 1986. Predictive likelihood inference with applications (with discussion). J. Roy. Statist. Soc., Ser. B 48, 1–38.

Fisher, R.A., 1956. Statistical Methods and Scientific Inference. Oliver and Boyd, London.

Lee, P.M., 1989. Bayesian Statistics: An Introduction. Oxford University Press, New York.

Levy, P., 1939. L'addition des variables ale'atories definis sur une circonference. Bull. Soc. Math., Paris 67, 1–41.

Mardia, K.V., 1972. Statistics of Directional Data. Academic Press, New York.

Zhong, J., 1991. Some contributions to the spherical regression model. Ph.D. Thesis, University of Kentucky.