The Annals of Statistics
1996, Vol. 24, No. 1, 232–254

AN EMPIRICAL BAYES APPROACH TO DIRECTIONAL DATA AND EFFICIENT COMPUTATION ON THE SPHERE[1]

BY DENNIS M. HEALY, JR. AND PETER T. KIM

Dartmouth College and University of Guelph

This paper proposes a consistent nonparametric empirical Bayes estimator of the prior density for directional data. The methodology is to use Fourier analysis on S^2 to adapt Euclidean techniques to this non-Euclidean environment. General consistency results are obtained. In addition, a discussion of efficient numerical computation of Fourier transforms on S^2 is given, and their applications to the methods suggested in this paper are sketched.

1. Introduction. The statistical study of directional data concerns observations from a sample space taken as the $p - 1$ dimensional unit sphere, S^{p-1}, where $p \geq 2$. The field goes back some time, with some of the earliest developments provided by Fisher (1953). Over the years, several monographs have appeared: Mardia (1972), Watson (1983) and Fisher, Lewis and Embleton (1987). Jupp and Mardia (1989) provided a comprehensive in depth review of the directional statistics literature over the period 1975–1988. The period prior to 1975 is reviewed in Mardia (1975).

The usual statistical approach to directional data is a frequentist one. To date, a very limited number of Bayesian techniques have been formulated. Indeed, we are aware of only Mardia and El-Atoum (1976), Lo and Cabrera (1987), Guttorp and Lockhart (1988) and Bagchi and Guttman (1988). Part of the difficulty is that Bayesian techniques for directional data do not often work out as in the Euclidean setting. Indeed, even with parametric assumptions, it is often the case that closed form solutions do not exist. One way to partly alleviate this difficulty is to implement conjugate priors, which is the case in most of the above-mentioned works; in fact, all of the above-mentioned works with exception of Lo and Cabrera (1987).

Although the use of conjugate priors mathematically simplifies the calculations, the resulting loss of freedom in the choice of the prior may not be desirable. In fact, prior selection in Bayesian analysis can be problematic, particularly if the analysis is sensitive to the prior density. Even when it is

Received December 1993; revised September 1995.
[1] Research supported in part by ARPA, as administered by the AFOSR under contracts AFOSR-90-0292 and DOD F4960-93-1-0567, and by the Natural Sciences and Engineering Research Council of Canada Grant OGP46204.
AMS 1991 *subject classifications*. Primary 62G05; secondary 58G25.
Key words and phrases. Consistency, fast Fourier transform, Legendre transform, spherical harmonics.

agreed that the sampling distribution is of some particular parametric form, there may be disagreement as to what the "correct" prior is. In such a setting, one could adopt a nonparametric Bayesian approach as in Ferguson (1973) and Brunner and Lo (1989); alternatively, one could make the prior data dependent. In this paper we will adopt the latter strategy and take the empirical Bayes approach as suggested by Robbins (1955).

In a nonparametric Euclidean setting, the empirical Bayes approach to reconstructing the prior can be realized using Fourier methods; see Fan (1991). In general, there is a well developed literature available for nonparametric empirical Bayes estimation in the Euclidean case; see, for example, Singh (1992) as well as the references contained therein. This is not true in the non-Euclidean situation: while there has been some work on density estimation on S^{p-1} [see Beran (1979), Hall, Watson and Cabrera (1987), Bai, Rao and Zhao (1988) and Hendriks (1990)], there has been no attempt to formulate nonparametric empirical Bayes estimation of the prior density in the spherical setting. Our goal is to provide such a technique.

In this paper we pay special attention to the case of the two-sphere. Although, in principle, we can formally perform the analysis on any S^{p-1}, $p \geq 3$, in so doing, we would have to substantially increase the technical sophistication. Consequently, our focus will be on S^2 with some remarks on how one would extend to higher dimensions. Note that this restriction is also motivated by the low dimension of many of the applications, in particular, those coming from geophysics.

We now provide a summary of what is to follow.

Our approach is to use Fourier analysis on $L^2(S^2)$, in analogy to techniques used in the Euclidean setting. This means that this approach largely comes down to the study of special functions known as spherical harmonics. Therefore, in Section 2 we provide a brief summary of some of the notions of Fourier analysis on S^2 and of spherical harmonics, as they will play an integral part in the reconstruction of the prior.

In Section 3, we demonstrate how rotationally invariant densities on S^2 can be thought of as the spherical analogue of the location type parametrization in Euclidean space. The effect is that when the sampling density is integrated with respect to a prior on S^2, the resulting unconditional marginal density is a convolution of functions on S^2. Applying a spherical Fourier transform breaks apart the convolution similar to the Euclidean setting. Consequently, the spherical transform of the prior can be formulated in terms of the spherical transform of the marginal and the spherical transform of the sampling densities. At this point we bring in the empirical Bayes methodology. Indeed, unconditionally, the data are a random sample from the marginal distribution. We can thus form an empirical version of the spherical transform of the marginal, which then produces an empirical version of the spherical transform of the prior. The final detail is to invert and smooth the transform, which produces the nonparametric prior density estimator. Under minimal conditions, we obtain consistency. If an additional smoothness condition is imposed, asymptotic L^2 rate of convergence can be obtained.

In Section 4, we illustrate how this can be used to obtain empirical Bayes point estimates of the location parameter. In addition we work out the situation when the sampling distribution is the Fisher–von Mises distribution. It turns out that in order to implement our procedure, the concentration parameter has to be sufficiently large. In other words, the sampling density needs to be significantly concentrated if reconstruction of the prior is to be successful. We quantify how large the concentration parameter has to be.

In Section 5 we provide a brief discussion on efficient computation on S^2 as developed in Driscoll and Healy (1994). In particular, for bandlimited functions Driscoll and Healy show that for computing spherical harmonic expansions, one can significantly decrease the number of computations required by a naive implementation. This amounts to a fast Fourier transform on S^2 which is applicable to the problem at hand.

All proofs of consistency are presented in Section 6. The relationship between the rotation group $SO(3)$ and the sphere S^2 is included in the Appendix as well as some technical proofs.

Finally, we would like to add that although this paper is written in a Bayesian context, we could have equally written this paper in a mixture framework. Indeed, most of the results of this paper are applicable to nonparametric estimation of the mixture density in the directional setting. This would provide a spherical analogue to the Euclidean version; see Zhang (1990).

2. Fourier analysis and synthesis on the sphere. We provide a brief discussion of the Fourier transform and its inverse for functions defined on the sphere. This amounts to an orthogonal decomposition of any reasonable function as a linear combination of special functions known as spherical harmonics. These are adapted to the symmetries of the sphere, provided by the usual rotations of space. A more thorough summary with references may be found in Driscoll and Healy (1994); a comprehensive treatment can be found in Courant and Hilbert (1953). Other statistical works that use expansions in spherical harmonics include Giné (1975), Wahba (1981) and Hendriks (1990).

The Hilbert space $L^2(S^2)$ is defined in the usual way with the inner product $\langle f, h \rangle = \int_{S^2} f\overline{h}\, d\omega$, where the overbar denotes complex conjugation. Here we use the (essentially) unique rotation-invariant area element on the sphere,

$$\int_{\omega \in S^2} f(\omega)\, d\omega = \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} f(\omega(\theta, \phi)) \sin\theta \, d\theta \, d\phi, \tag{1}$$

in the usual coordinates. Its rotation invariance, that is,

$$\int_{\omega \in S^2} f(g\omega)\, d\omega = \int_{\omega \in S^2} f(\omega)\, d\omega,$$

where $g \in SO(3)$ is the group of 3×3 real orthogonal matrices of determinant 1, follows from the observation that this is the angular part of the polar coordinate decomposition of the (rotation-invariant) Lebesgue measure on $\mathbf{R}^3$.

For a generic point $\omega \in S^2$, we can write it as $\omega(\theta, \phi) = (\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta)^t$, where $0 < \theta \leq \pi$, $0 < \phi \leq 2\pi$ and superscript t denotes transposition. The conventional expressions for the spherical harmonics in these coordinates are

$$Y_q^l(\omega) = (-1)^q \sqrt{\frac{(2l+1)(l-q)!}{4\pi(l+q)!}} P_q^l(\cos\theta) e^{iq\phi}, \tag{2}$$

where $-l \leq q \leq l$, $l = 0, 1, \dots$ and P_q^l are the Legendre functions. These functions are defined in the following way.

The Legendre polynomial is defined by

$$P_l(x) = \frac{1}{2^l l!} \frac{d^l}{dx^l} (x^2 - 1)^l$$

for $l \geq 0$ and $x \in [-1, 1]$. Define Legendre functions by $P_0^l(x) = P_l(x)$ and

$$P_q^l(x) = (1 - x^2)^{q/2} \frac{d^q}{dx^q} P_0^l(x),$$

where $0 \leq q \leq l$, $l \geq 0$ and $x \in [-1, 1]$. For $-l \leq q \leq 0$, define them through the equation

$$P_{-q}^l(x) = (-1)^q \frac{(l-q)!}{(l+q)!} P_q^l(x)$$

for $x \in [-1, 1]$. The effect of this choice is that

$$\overline{Y_q^l} = (-1)^q Y_{-q}^l,$$

where $0 \leq q \leq l$ and $l \geq 0$.

The Legendre functions satisfy the recurrence relation

$$(l - q + 1) P_q^{l+1}(x) - (2l + 1) x P_q^l(x) + (l + q) P_q^{l-1}(x) = 0.$$

This recurrence is known to provide a numerically stable method of computing the values of the Legendre functions and it leads to efficient algorithms for the computation of Fourier coefficients, as in Driscoll and Healy (1994).

We noted that $\{Y_q^l\colon -l \leq q \leq l,\ l = 0, 1, \dots\}$ form a complete orthonormal basis over $L^2(S^2)$ adapted to the rotational symmetries of the sphere. Looking at the infinitesimal generators of these rotation operators shows that the spherical harmonics are the eigenfunctions of the Laplace–Beltrami operator on smooth functions on S^2,

$$\Delta^* = \frac{1}{\sin\theta} \frac{\partial}{\partial\theta} \left(\sin\theta \frac{\partial}{\partial\theta} + \frac{1}{\sin\theta} \frac{\partial^2}{\partial\phi^2} \right),$$

and for each $l \geq 0$, Y_q^l has corresponding eigenvalue $l(l+1)$ for all $-l \leq q \leq l$. In many treatments, this is used as the defining property of the spherical harmonics; see Giné (1975), Wahba (1981) and Hendriks (1990).

Let $f \in L^2(S^2)$. We define the *spherical Fourier transform* of f as the collection of its projections onto the elements of the spherical harmonic basis:

$$\hat{f}_q^l = \int_{S^2} f(\omega) \overline{Y_q^l(\omega)}\, d\omega \tag{3}$$

for $-l \leq q \leq l$ and $l \geq 0$. The *spherical inversion* can be obtained by

$$f(\omega) = \sum_{l \geq 0} \sum_{q=-l}^{l} \hat{f}_q^l Y_q^l(\omega) \tag{4}$$

for $\omega \in S^2$. We note that (4) should be interpreted as in the L^2 sense, although it can hold pointwise with additional smoothness conditions.

If one wants to generalize to arbitrary S^{p-1}, the way to proceed is to realize that the $(p-1)$ sphere is the quotient space of $SO(p)$ modulo $SO(p-1)$, where $SO(q)$ is the group of $q \times q$ real orthogonal matrices of determinant 1. Now $SO(p)$ is a compact Lie group. Consequently, a complete orthonormal basis of $L^2(SO(p))$ can be obtained through the irreducible representations of $SO(p)$; see the Appendix for the $p = 3$ case. By transforming these irreducible representations by the quotient map $SO(p) \to SO(p)/SO(p-1)$, one can recover the higher dimensional spherical harmonics which will satisfy (3) and (4) in the appropriate dimension. The case of $p = 4$ is illustrated in Chapter 4 of Talman (1968). The abstract case is illustrated in Chapter 4 of Helgason (1984). Discussions on the theory of group representations can be found in Helgason (1984) and Diaconis (1988). Generalizations of the algorithms of Driscoll and Healy to higher dimensional spheres may be found in Maslen (1993).

3. Estimation. We begin the discussion by assuming rotational invariance,

$$f(x \mid \mu) = f(\mu^t x) \tag{5}$$

for $x, \mu \in S^2$; see Watson (1983). Let $\pi(\cdot)$ be the prior density on S^2. Then the marginal density is

$$m(x) = \int_{S^2} \pi(\mu) f(\mu^t x)\, d\mu, \tag{6}$$

$x \in S^2$.

Now (6) represents convolution in the spherical setting and if we take the spherical transform, then a simplification similar to that which occurs in the familiar Euclidean case is obtained. We have the following convolution lemma, whose proof is deferred to the Appendix.

LEMMA 3.1. *Suppose $f, \pi \in L^2(S^2)$. Then*

$$\hat{m}_q^l = 2\pi \sqrt{\frac{4\pi}{2l+1}}\, \hat{\pi}_q^l \hat{f}_0^l,$$

where

$$\hat{f}_0^l = \int_{S^2} f(\eta^t \omega) \overline{Y_0^l(\omega)}\, d\omega,$$

$\eta = (0, 0, 1)^t$ *for* $-l \leq q \leq l$ *and* $l \geq 0$.

The consequence of Lemma 3.1 is that if $\hat{f}_0^l \neq 0$ for $l \geq 0$, then

$$\hat{\pi}_q^l = \left[\frac{(2l+1)}{16\pi^3}\right]^{1/2} \frac{\hat{m}_q^l}{\hat{f}_0^l} \tag{7}$$

for $-l \leq q \leq l$ and $l \geq 0$.

Let us assume $f(\mu^t x)$ is known. Consequently, $\hat{f}_0^l$ is known for $l \geq 0$. The statistical analysis comes in with respect to prior uncertainty, that is, an unknown $\pi(\cdot)$, which of course implies an unknown $m(\cdot)$ as defined in (6). We note, however, that if a random sample $X_1, \dots, X_n$ is observed, then from a Bayesian point of view, we can regard this random sample as unconditionally coming from (6). This of course can then be used to construct an unbiased estimator of $\hat{m}^l$ for $l \geq 0$. Indeed, define

$$\hat{m}_q^{n,l} = \frac{1}{n} \sum_{j=1}^{n} \overline{Y_q^l(X_j)}$$

for $-l \leq q \leq l$ and $l \geq 0$. Assuming that $\hat{f}_0^l \neq 0$ for $l \geq 0$, a logical estimator for (7) would be

$$\hat{\pi}_q^{n,l} = \left[\frac{(2l+1)}{16\pi^3}\right]^{1/2} \frac{\hat{m}_q^{n,l}}{\hat{f}_0^l} \tag{8}$$

for $-l \leq q \leq l$ and $l \geq 0$. Finally, use spherical inversion (4) to define an estimator of $\pi(\cdot)$, namely,

$$\pi^n(\omega) = \sum_{l=0}^{1/b} \sum_{q=-l}^{l} \hat{\pi}_q^{n,l} Y_q^l(\omega), \tag{9}$$

where $\omega \in S^2$ and $b = b(n) \to 0$ as $n \to \infty$. The latter is acting as a damping factor which controls the accumulation of the higher order frequencies.

We have the following results, whose proofs will be deferred to Section 6. For two sequences $\{a_n\}$ and $\{c_n\}$, $a_n = O(c_n)$ as $n \to \infty$ will be symbolized by $a_n \ll c_n$ as $n \to \infty$. Furthermore, $\|\cdot\|_2$ will denote the $L^2(S^2)$ norm.

THEOREM 3.2. *For $b = b(n) \to 0$ as $n \to \infty$, suppose $|\hat{f}_0^l| \neq 0$ for $l = 0, 1, \dots, 1/b$. Furthermore, suppose $f(\cdot)$ is continuous and $\pi(\cdot)$ can be represented pointwise by its Fourier series. If*

$$\frac{1}{n} \sum_{l=0}^{1/b} \frac{(2l+1)^2}{|\hat{f}_0^l|^2} \to 0$$

as $n \to \infty$, then

$$E|\pi^n(\omega) - \pi(\omega)|^2 \to 0$$

as $n \to \infty$ for all $\omega \in S^2$.

Under stronger conditions we can determine the rate of convergence in the $L^2(S^2)$ norm.

THEOREM 3.3. *Suppose there is some $C > 0$ and $u \in \mathbf{R}$ such that $|\hat{f}_0^l|^2 \geq Cl^{-u}$ for all $l = 0, 1, \ldots$. Furthermore, suppose $f(\cdot)$ is continuous and $\pi(\cdot)$ is s times differentiable with square integrable derivatives. Then*

$$E\|\pi^n - \pi\|_2^2 \ll n^{-2s/(2s+3+u)}$$

for $s \geq 1$ and $u > -2s - 3$ as $n \to \infty$.

4. Applications. In this section we discuss two applications. The first application is in the empirical Bayes estimation of the mean direction, whereas the second application is when we assume the sampling distribution follows the Fisher–von Mises distribution.

4.1. *Empirical Bayes estimation of the mean direction.* Suppose we wish to make inference about μ based on the observation X. We note that in terms of squared error loss, if μ^* is an estimator of μ, then

$$L(\mu, \mu^*) = (1 - \mu^t\mu^*)$$

for $\mu \in S^2$. Consequently, if $\pi(\cdot)$ is the prior density, then the Bayes risk of μ^* is

$$r(\mu^*) = 1 - \int_{S^2\times S^2} \mu^t\mu^* f(\mu^t x)\pi(\mu)\,dx\,d\mu.$$

Now in terms of the usual Fubini argument, we have

$$\begin{aligned}\int_{\mu\in S^2}\int_{x\in S^2} \mu^t\mu^* f(x^t\mu)\pi(\mu)\,dx\,d\mu &= \int_{x\in S^2}\int_{\mu\in S^2}\mu^t\mu^*\pi(\mu|x)m(x)\,d\mu\,dx\\ &= \int_{x\in S^2}\mu^{*t}\left\{\int_{\mu\in S^2}\mu\pi(\mu|x)\,d\mu\right\}m(x)\,dx\\ &= \int_{x\in S^2}\mu^{*t}\{E^{\pi(\mu|x)}\mu\}m(x)\,dx,\end{aligned}$$

where $\pi(\mu|x)$ is the posterior density. Thus for each $x \in S^2$, the solution to

$$\max_{\mu^*\in S^2} \mu^{*t}E^{\pi(\mu|x)}\mu$$

is the Bayes estimator

$$\mu_b = \frac{E^{\pi(\mu|x)}\mu}{\|E^{\pi(u|x)}\mu\|} = \frac{\int_{S^2}\mu f(x^t\mu)\pi(\mu)\,d\mu}{\|\int_{S^2}\mu f(x^t\mu)\pi(\mu)\,d\mu\|}. \tag{10}$$

We are assuming that the prior density $\pi(\cdot)$ is unknown. However, suppose we have a random sample $X_1, \ldots, X_{n+1}$. Let $X = X_{n+1}$ and use $X_1, \ldots, X_n$ to form a consistent estimator of $\pi(\cdot)$ as in (9). An empirical Bayes estimator of μ can be formulated by

$$\mu_{\text{eb}} = \frac{\int_{S^2}\mu f(x^t\mu)\pi^n(\mu)\,d\mu}{\|\int_{S^2}\mu f(x^t\mu)\pi^n(\mu)\,d\mu\|}. \tag{11}$$

In terms of convergence, by Theorem 3.2 and the continuous mapping theorem,

$$\mu_{\rm eb}(x) \to \mu_b(x)$$

as $n \to \infty$ for almost all $x \in S^2$. By the dominated convergence theorem, we have

$$r(\mu_{\rm eb}) \to r(\mu_b)$$

as $n \to \infty$.

4.2. *Fisher–von Mises distribution.* By far the most popular parametric density on S^2 is the Fisher–von Mises distribution. The density takes on the parametric form

$$f(x|\mu, \kappa) = c(\kappa)\exp\{\kappa\mu^t x\}, \tag{12}$$

where $x, \mu \in S^2$, $\kappa > 0$ and $c(\kappa) = \kappa/\sinh \kappa$. In this section we will outline the methodology when (12) is taken as the sampling distribution.

The fact that we need $\kappa > 0$ to be large has the following intuitive explanation. Notice in (12) that as $\kappa \to \infty$, the density is concentrating on the mean direction $\mu \in S^2$. Since we are trying to reconstruct the prior $\pi(\cdot)$ from "noisy" data, the sampling distribution has to be sufficiently sharp in order to effectively do this reconstruction, where the sharpness is parametrized by $\kappa > 0$. Therefore, we need to quantify how sharp things must be in order for the reconstruction of the prior to take place.

Assume $\kappa > 0$ is known. As in Lemma 3.1, by the rotational invariance and the definition of the Legendre polynomials P_l, one can show that

$$\hat{f}_0^l \propto \frac{1}{\sinh \kappa} \sum_{j=0}^{l} \left(\frac{-1}{\kappa}\right)^j \left[e^{\kappa} P_l^{(j)}(1) - e^{-\kappa} P_l^{(j)}(-1)\right], \tag{13}$$

where $\kappa > 0$, $l \geq 0$, the proportionality constant is $\sqrt{(2l+1)/4\pi}$ and $P_l^{(j)}$ is in reference to the jth derivative for $j = 0, 1, \dots, l$.

In order for us to reconstruct the prior density by (9), we have to be able to bound (13) away from 0 for all l between 0 and $1/b$. In general, one can show that this condition cannot be satisfied for all $l \geq 0$. However, if the concentration parameter $\kappa \geq 0$ is sufficiently large, we can keep (13) positive for $l = 0, 1, \dots, 1/b$. Define,

$$C(\kappa, b) = \max\left\{\frac{1}{2}\log\left(1 + 2\frac{1}{b}\left[\frac{1}{b}\right]!(2\kappa)^{1/b}\right), \frac{1}{b^2}\right\} \tag{14}$$

for $\kappa, b > 0$. Notice that for $b > 0$ fixed, $C(\kappa, b) = o(\kappa)$ as $\kappa \to \infty$. If $\kappa > 0$ is large so that

$$\kappa > C(\kappa, b) \tag{15}$$

for $b > 0$ fixed, then $\hat{f}_0^l > 0$ for all $l = 0, 1, 2, \dots, 1/b$. The rest of this section will be devoted to proving this statement. Figure 1 demonstrates these facts pictorally.

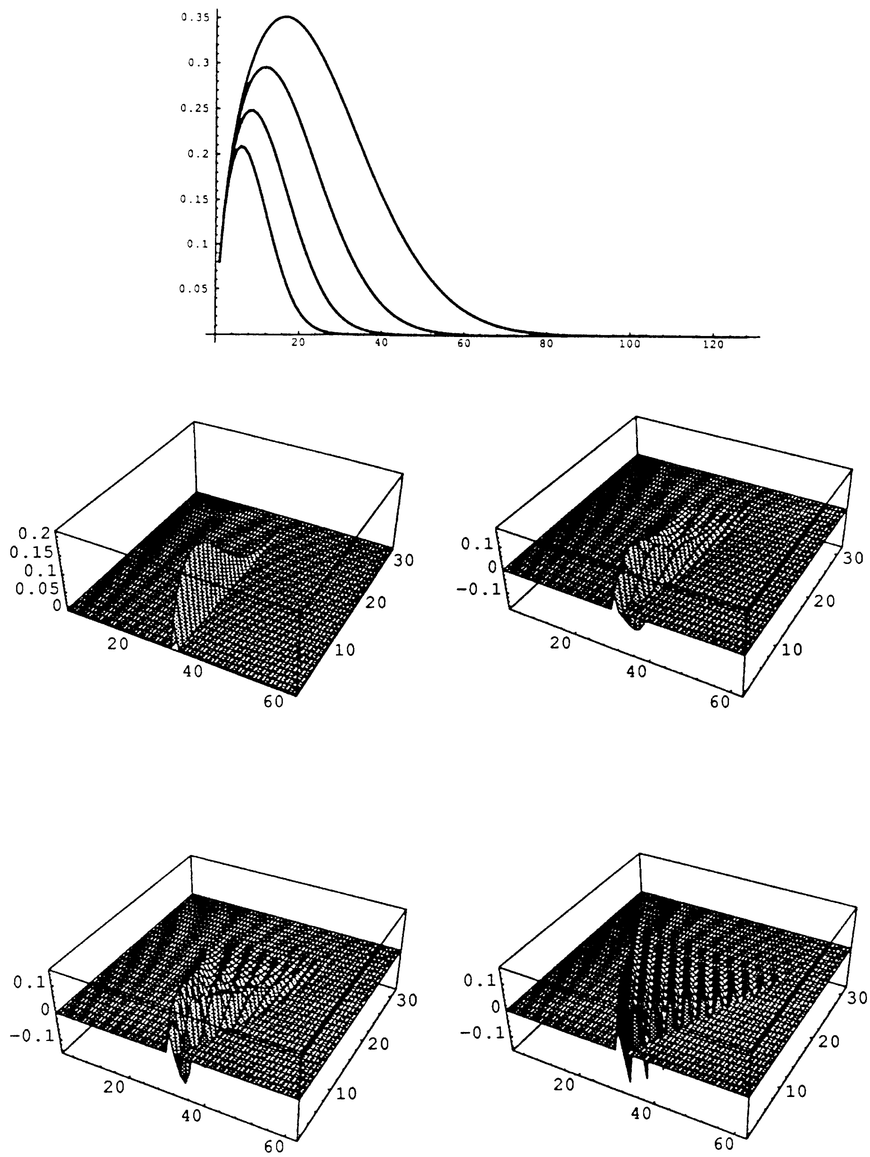

FIG. 1. *Top*: *Legendre coefficients of Fisher–von Mises density* (*centered at the north pole*) *with* κ *increasing from* 64 *to* 512 *by powers of* 2. *Below*: *Four frames showing the effect of changing the position of the Fisher–von Mises density* ($\kappa = 64$) *in the spherical Fourier domain. The colatitude of the center position, clockwise from the upper frame is* 0, $\pi/8$, $\pi/2$ *and* $3\pi/8$.

Using the Rodrigues formula for the Legendre polynomials, we have

$$P_l^{(j)}(1) = \frac{1}{2^l l!} D^{l+j}(x^2-1)^l \Big|_{x=1}.$$

To evaluate this, use the Leibnitz rule:

$$= \frac{1}{2^l l!} \sum_{\substack{\mathbf{s}\in \mathbf{N}^l \\ s_1+\cdots+s_l=l+j}} \mu_{\mathbf{s}}^{l+j} \prod_{i=1}^{l} D^{s_i}(x^2-1) \Bigg|_{x=1},$$

where the multinomial coefficients μ_s^{l+j} are

$$\mu_s^{l+j} = \frac{(l+j)!}{s_1! s_2! \cdots s_l!}.$$

Now observe that $D^s(x^2-1)|_{x=1} = 0$ unless $s = 1$ or 2. In both of these cases the result is 2. Therefore, for each j between 0 and l the only nonvanishing terms in the sum are those corresponding to vectors $\mathbf{s}$ having j components with value 2, and the remaining components all 1. For any of these $\binom{l}{j}$ vectors, the product always has the same value of 2^l. Consequently,

$$P_l^{(j)}(1) = \frac{(l+j)!}{(l-j)!j!2^j}. \tag{16}$$

Note in particular that all the derivatives are positive. A direct consequence of (16) is

$$\begin{aligned} &\frac{P_l^{(j)}(1)}{\kappa^j} - \frac{P_l^{(j+1)}(1)}{\kappa^{j+1}} \\ &\quad = \frac{1}{\kappa^j 2^{j+1}(j+1)!}\left[2(j+1)\kappa - (l+j+1)(l-j)\right]\frac{(l+j)!}{(l-j)!} \end{aligned} \tag{17}$$

for any $j = 0, 1, \dots, l-1$. Note that a sufficient condition for $2(j+1)\kappa > (l+j+1)(l-j)$ for $j = 0, 1, \dots, l-1$, $l = 0, 1, \dots, 1/b$, is for $\kappa > 1/b^2$.

For evaluating (13), we note that

$$P_l^{(j)}(-1) = (-1)^{l-j} P_l^{(j)}(1).$$

Thus when $l > 0$ is odd,

$$\begin{aligned} \hat{f}_0^l &\propto \left(\frac{\cosh\kappa}{\sinh\kappa} P_l(1) - \frac{P_l^{(1)}(1)}{\kappa}\right) + \cdots + \left(\frac{\cosh\kappa}{\sinh\kappa}\frac{P_l^{(l-1)}(1)}{\kappa^{l-1}} - \frac{P_l^{(l)}(1)}{\kappa^l}\right) \\ &\geq \left(P_l(1) - \frac{P_l^{(1)}(1)}{\kappa}\right) + \cdots + \left(\frac{P_l^{(l-1)}(1)}{\kappa^{l-1}} - \frac{P_l^{(l)}(1)}{\kappa^l}\right) \\ &> 0, \end{aligned}$$

provided $\kappa > 1/b^2$.

For the case $l \geq 0$ is even,

$$
\begin{aligned}
\hat{f}_0^l \propto & \left(P_l(1) - \frac{\cosh \kappa}{\sinh \kappa} \frac{P_l^{(1)}(1)}{\kappa} \right) + \cdots \\
& + \left(\frac{P_l^{(l-2)}(1)}{\kappa^{l-2}} - \frac{\cosh \kappa}{\sinh \kappa} \frac{P_l^{(l-1)}(1)}{\kappa^{l-1}} \right) + \frac{P_l^{(l)}(1)}{\kappa^l} \\
= & \left(P_l(1) - \frac{P_l^{(1)}(1)}{\kappa} \right) + \cdots + \left(\frac{P_l^{(l-2)}(1)}{\kappa^{l-2}} - \frac{P_l^{(l-1)}(1)}{\kappa^{l-1}} \right) + \frac{P_l^{(l)}(1)}{\kappa^l} \\
& - \frac{2}{e^{2\kappa} - 1} \left[\frac{P_l^{(1)}(1)}{\kappa} + \cdots + \frac{P_l^{(l-1)}(1)}{\kappa^{l-1}} \right] \\
> & \frac{1}{\kappa^l} \frac{(2l)!}{2^l l!} - \frac{1}{\kappa} \frac{2(2l)!(l-1)}{e^{2\kappa} - 1},
\end{aligned}
$$

provided $\kappa > 1/b^2$. Consequently, this term will be positive provided that the last term is positive, which is guaranteed if (15) is satisfied.

5. Computational considerations of Fourier transforms on S^2. In order to efficiently apply the methods discussed in this paper, we need an effective method for computing the Fourier coefficients as defined in (3) and for computing the Fourier synthesis formulas needed for the reconstruction of the estimates of the prior, as in (9). In general, these integrals and sums cannot be computed in closed form. Furthermore, the use of computers to implement numerical techniques for the approximate computation of Fourier transform integrals forces us to create algorithms which use only the values, or “samples,” of a function taken on a prescribed finite set. Note that this use of the word “sample” is in conflict with that in the term “random sample.” Unfortunately, the former usage is as entrenched in numerical analysis as the latter is in statistics. We will try to keep the two usages clear.

Although there has been a growing effort already devoted to the issue of “quadrature” on the sphere in the applied mathematics literature, this issue remains relatively dormant in the statistical literature. In particular, although Giné (1975) and Wahba (1981) discuss computational aspects for spherical data, there has been little discussion of efficient algorithms which are required to keep computational costs reasonable in problems requiring high resolution.

In cases where we require even a moderate resolution, we may need thousands of function values, or samples, in order to compute a similar number of that function’s Fourier coefficients. A straightforward application of the standard quadrature rules to this problem produces a computational problem of very high cost, typically involving a number of calculations on the order of the square of the number of desired Fourier coefficients. A less costly approach is often needed.

A final consideration in these matters concerns the numerical reliability of the calculations when performed on a standard computer using finite precision arithmetic. Many seemingly reasonable algorithms have been found to be useless when implemented due to the cumulative effects of finite precision rounding.

In this section, we review the well known and very useful methods in the Euclidean case which have successfully addressed these issues. We review a recently developed analogue for the case of Fourier transforms on the sphere. Finally we apply this new method to the problem considered in the present paper.

5.1. *Euclidean fast Fourier transform.* In Euclidean data analysis, the foregoing considerations are well known, and there is already a well developed and widely used approach to numerical Fourier analysis. A particular application of this toolbox to density estimation on the circle may be found in Silverman (1986).

In general, one begins with the well known Shannon sampling theorem, which shows that for periodic bandlimited functions having no Fourier components above, say, the Bth cutoff frequency, the nonvanishing Fourier coefficients may be computed exactly in exact arithmetic from B uniform samples of the function on the circle. The quadrature formula in this setting comes down to a simple trapezoid rule on the uniform samples for each of the B integrals.

Naively, these B sums require a total of $O(B^2)$ calculations. However, one may apply any of a number of related schemes for reorganizing the calculations to speed up the computation of Fourier coefficients. For bandlimited functions on the circle these tricks, known collectively as *fast Fourier transform* (FFT) algorithms, compute in $O(B \log B)$ time the B Fourier coefficients of such a function from its samples at B distinct points. This turns out to be a huge win for problems which occur in practice. As one may guess from the widespread practical application of FFT algorithms in applied mathematics, computer science, engineering and statistics, the FFT is numerically benign with respect to the perturbations of computational roundoff, or imprecisions in the data or the complex exponentials appearing in the computations.

So we see that in the Euclidean case we may reduce the needed Fourier coefficient integrals to discrete calculations that a computer may handle, and that these calculations may be done efficiently and reliably. These properties of FFT and sampling algorithms make Euclidean Fourier analysis a powerful and widely used tool in data analysis. This provides the gold standard for any numerical approach to Fourier analysis on the sphere.

5.2. *Fast Fourier transform on* S^2. Recently, a step has been made toward efficient, reliable numerical Fourier analysis for the sphere. The paper by Driscoll and Healy (1994) presented an algorithm for the efficient numerical computation of Fourier coefficients and for the Fourier inversion. This algorithm is exact (in exact arithmetic) for bandlimited functions on S^2. This leads to algorithms for the efficient computation of the convolution of two such functions on S^2.

The paper cited above presents an $O(B(\log B)^2)$ algorithm that, given a data structure of size $O(N \log N)$, computes the spherical harmonic expansion of a discrete function on the sphere defined on an equiangular grid of $B = 2^k \le N$ points. This improves the naive $O(B^2)$ bound, which is the best previously known. In addition Healy, Moore and Rockmore (1993) described methods for inverting this transform in the same time, and using these transforms, an $O(B(\log B)^2)$ time algorithm to convolve bandlimited functions on the sphere.

The remainder of this section gives a brief overview of the issues of sampling, fast computation and experimental performance results of actual implementations.

Sampling. We consider bandlimited functions of degree less than some bandlimit B. In this paper, the bandlimited assumption is necessarily met due to the finite sample size. In fact, in terms of (9), we can take $B = 1/b$. If $f \in L^2(S^2)$ has bandlimit B, the equiangular grid given by sample points $\{(\theta_j, \phi_k) | 0 \le j, k < 2B\}$ for $\theta_j = (\pi(j + 1/2))/(2B)$ and $\phi_k = (\pi(k + 1/2))/B$ give a quadrature rule for computing the Fourier coefficients of f. We have the following theorem, which is Theorem 3 in Driscoll and Healy (1994).

THEOREM 5.1. *Let $f(\theta, \phi)$ be a bandlimited function on S^2 such that $\hat{f}_q^l = 0$ for all $l \ge B$. Then*

$$\hat{f}_q^l = \frac{\sqrt{2\pi}}{2B} \sum_{j=0}^{2B-1} \sum_{i=0}^{2B-1} a_j^{(B)} f(\theta_j, \phi_k) \overline{Y_q^l}(\theta_j \phi_k) \tag{18}$$

for $l < B$ and $|q| \le l$.

The weights $a_j^{(B)}$ are a discrete analogue of the usual $\sin(\theta)$ in the invariant integral on the sphere (1). A closed form expression for these may be found in Driscoll and Healy (1994).

The number of sample points is asymptotically optimal in B, from the point of view of linear algebra. That is, if f is determined by its Fourier coefficient $\hat{f}_q^l$ for $0 \le l < B$, $|q| \le l$, then by simple linear algebra, at least that many $(B^2 - 2B + 2)$ samples are needed. Instead the grid $\{(\theta_i, \phi_j)\}$ uses $4B^2$ points, so $O(B^2)$ points as well.

This grid also provides algorithmic advantages, as we see in the following description of an efficient algorithm for the Fourier transform on S^2.

Efficient algorithms. The sampling theorem reduces the computation of the Fourier coefficients to discrete calculations (18). For bandlimited functions with bandlimit B, we have $N = O(B^2)$ function samples to compute the $O(N)$ sums required to give us the Fourier coefficients. A naive operation count suggests that this requires $O(N^2)$ time. However, we have the following result, proved in Driscoll and Healy (1994).

THEOREM 5.2. *If $f(\theta, \phi)$ is in the span of $\{Y_q^l | l < B, |q| \le l\}$, then the Fourier coefficients $\hat{f}_q^l$ for $l < B$, $|q| \le l$ can be computed in $O(M(\log M)^2)$ operations from the $M = 4B^2$, $2^k < N$, sampled values $f(k\pi/2B, j\pi/B)$, $0 \le j$, $k < 2B - 1$, using a preprocessed data structure of size $O(N \log^2 N)$.*

The transform of the function f may be viewed as the application of a matrix of sampled spherical harmonics to a vector of sampled values of f. The fast transform algorithm utilizes the recurrences satisfied by the harmonics to effect a factorization of this matrix into sparse, structured matrices. These may be applied with a net reduction in complexity over the naive approach.

Transposition of this matrix factorization permits the efficient inversion of the Legendre transforms and hence an $O(N \log^2 N)$ inversion of the spherical Fourier transform sampled at the N points of the spherical sampling grid described above. Details may be found in Healy, Moore and Rockmore (1993).

Stability and experimental results. The transform algorithms have been implemented in C and are undergoing tests. This efficient algorithm demonstrates significant reduction in running time over naive implementation, even at relatively moderate problem size. This is illustrated in Figure 2.

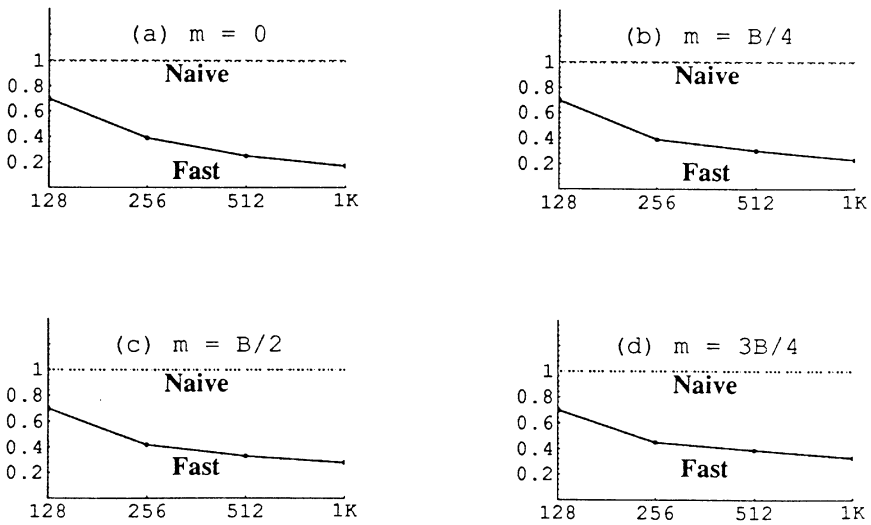

FIG. 2. *Experimental results on the speed of the fast spherical transform algorithm as a function of problem size (bandwidth B). Time is given as a fraction of that required by the naive implementation. Each of the four graphs shows the timing results for the associated Legendre transform with a given choice of the order parameter. More precisely, at a given bandwidth on a particular graph, we show the time for projection onto the family P_q^l: with the order q fixed at a given fraction of the bandwidth and the degree l varying over its appropriate range between q and B. Clockwise from top left: $q = 0$, $q = B/4$, $q = B/2$ and $q = 3B/4$.*

We have both analytical and experimental evidence for the numerical stability of the algorithms. In the case of the zonal Fourier transform on the sphere, a priori bounds have been given in Driscoll and Healy (1994). These calculations show that the zonal transform is reliable for a wide range of problem sizes. The same paper show some experimental results suggesting the same thing.

It is not so clear how analytical stability for the $q \neq 0$ case can be achieved. However, we can experimentally demonstrate numerical stability. See Healy, Moore and Rockmore (1993) for some of these results.

5.3. *Computing the prior density*. An application of the algorithms described above uses the fast inverse transform to synthesize and estimate the prior, $\pi(\cdot)$, on an equiangular grid on the sphere from the estimates of its Fourier transform, as in (9). Use of the fast inverse transform allows us to evaluate the estimator

$$\pi^n(\omega) = \sum_{l=0}^{B(n)} \sum_{q=-l}^{l} \hat{\pi}_q^{n,l} Y_q^l(\omega) \tag{19}$$

for all $\omega = \omega(\theta, \phi)$ in the grid $\theta = k\pi/2B$, $\phi = j\pi/B$, $0 \leq j,\ k < 2B - 1$, in $O(B \log^2 B)$ calculations. This can amount to quite a savings over the naive $O(B^2)$ inversion for even moderate resolutions, particularly if many of these syntheses must be calculated.

Likewise, the forward direction fast transform may be used in a straightforward manner to obtain the Fourier coefficients of the sampling density $f(\cdot)$. In particular, should this function be taken from a parametric family without closed form Fourier transforms, then the algorithm can be used to efficiently obtain the required Fourier coefficients. This will be especially important if the exact parameter value is not known, and some experimentation needs to be done. Likewise, the fast forward algorithm should be very useful in the case that the sampling density is known only by their values on the equiangular grid.

Somewhat more work must be done to apply the fast Fourier algorithm to the calculation of the empirical estimate of the Fourier coefficients of the marginal density. Here one is required to take the Fourier transform of a sum of point masses associated to the observed directions. As the point mass is not bandlimited, we must replace each point mass by an essentially bandlimited approximation. The bandlimited function must then be sampled on the equiangular grid before the transform algorithm may be applied.

A description of a simple procedure for density estimation on the circle using the Euclidean FFT can be found in Silverman (1986). A key consideration of the algorithm described there is the act of replacing each observation with a bandlimited approximation to the point mass at the observation point. One then sum these up over all observations, samples the resulting function on the usual uniform grid and then passes to the Fourier domain with the Euclidean FFT. The obvious bandlimited approximation of a point mass is its projection into the space of bandlimited functions, namely, the appropriate Dirichlet kernel. Unfortunately, this function has many nonzero values on the usual sampling grid unless it happens to be centered at one of the points of that sampling grid. In general, this choice of approximate point mass will lead to high computational costs in summing up over the observations, as it is unlikely that one may line up a uniform grid with most of the observations. The cost will be close to the number of observations times the size of the grid.

To save computational steps, Silverman proposes instead that when an observation falls between two grid points, it should be replaced with a function whose values vanish at all grid points but the two bracketing grid points nearest it. The value at each of these grid points is taken proportional to the distance between the observation and the other grid point. In effect, this is an appropriate linear combination of the two Dirichlet kernels centered on each of the grid points adjacent to the observation. Upon summing over all observations, one forms the samples of a bandlimited function associated with the data in a time proportional to the number of observations. The result is an approximation to the sum of Dirichlet kernels centered at the true observation positions. This approximation will generally be acceptable in practice.

The situation in the case of the sphere is somewhat complicated by the nonuniform mesh of the equiangular grid. We must adjust the scheme for replacing an observation with a bandlimited approximate point mass according to where the observation is located. Nevertheless, we may suggest one fairly simple scheme at this time.

We assume that a random sample $X_1, \dots, X_n$ is observed, and we wish to construct an approximation of an unbiased estimator of $\hat{m}^l$ for $l \geq 0$. Indeed, define

$$\hat{m}_q^{n,l} = \frac{1}{n} \sum_{j=1}^{n} \overline{Y_q^l(X_j)}.$$

If n, the number of observations is large, then these coefficients must be obtained for a large bandwidth $B(n)$; that is, for all $0 \leq l \leq B(n)$ and for each fixed l in that range, for all q with $|q| \leq l$. The resulting calculation by evaluation of the spherical harmonics at the observation points is $O(nB(n)^2)$, which could be quite onerous.

Instead, we suggest that one replace each observation with a rotationally symmetric bump function approximating the point mass and which vanishes outside a small, constant distance of the observation point and is normalized to have unit mass. One method of construction we have studied begins with a bump function, $b(r)$, $r \in \mathbf{R}$, which is symmetric about the origin. This is lifted to a bump on the sphere by forming $b(1 - \mathbf{X} \cdot \omega)$ which is simply a function of the distance $d(\mathbf{X}, \omega) = \arccos \mathbf{X} \cdot \omega$ between the observation location $\mathbf{X}$ and the evaluation point ω, considered as unit vectors. Adding up the various bumps associated with all the observations amounts to smoothing those observations with the bump function.

By this device, we hope to build a function that is well concentrated near the observation, is effectively bandlimited and which nevertheless is a good approximation of the point mass subject to the bandlimit. We have considered simple constructions based on the use of B-splines for the bump function. A bump function based on a quintic spline is shown in Figure 3. Another approach might be based upon the use of several of the functions proposed by Grunbaum, Longhi and Perlstadt (1982), which obtain simultaneously good concentration on the sphere and in the spherical Fourier transform domain. At any rate, we continue to look for the best approach to building these bump functions.

The use of splines to smooth the individual observations involves polynomial evaluation at no more than $O(B)$ grid points in term of the known observation vector and the known nearby grid points. The worst case occurs for observation points near the north pole, due to the bunching up of the equiangular grid near the poles; see Figure 3. Note, however, that in the very common case that the data largely avoid at least one axis through the sphere, that a proper choice of the coordinate system insures that there will be few of these costly steps to worry about.

We now have a worst case of $O(nB(n))$ steps to evaluate the sum of the approximate point masses corresponding to the entire set of observations, with a constant depending on the order of the splines used and the radius chosen for the spline bumps. An additional $O(B(n)\log^2 B(n))$ gives the estimate of the Fourier transform of the density, so the total cost is dominated by the evaluation of the point masses at no more than $O(nB(n))$. If the data are clustered, the speed will be considerably better than this.

6. Consistency proofs.

Define

$$K_n(\nu, \omega) = \sum_{l=0}^{1/b} \sum_{q=-l}^{l} \left[\frac{(2l+1)}{16\pi^3} \right]^{1/2} \frac{1}{\hat{f}_0^l} \overline{Y}_q^l(\nu) Y_q^l(\omega), \tag{20}$$

where $\nu, \omega \in S^2$. Note that we can write

$$\pi^n(\omega) = \frac{1}{n} \sum_{j=1}^{n} K_n(X_j, \omega),$$

where $\omega \in S^2$. Let X be a generic random element. Then because $E\hat{\pi}_q^{n,l} = \hat{\pi}_q^l$ for all $|q| \le l$ and $l = 0, 1, \dots, 1/b$,

$$E\pi^n(\omega) = EK_n(X, \omega) \to \sum_{l=0}^{\infty} \sum_{q=-l}^{l} \hat{\pi}_q^l Y_q^l(\omega) = \pi(\omega) \tag{21}$$

as $n \to \infty$ for $\omega \in S^2$ if $\pi(\cdot)$ can be represented by its Fourier series.

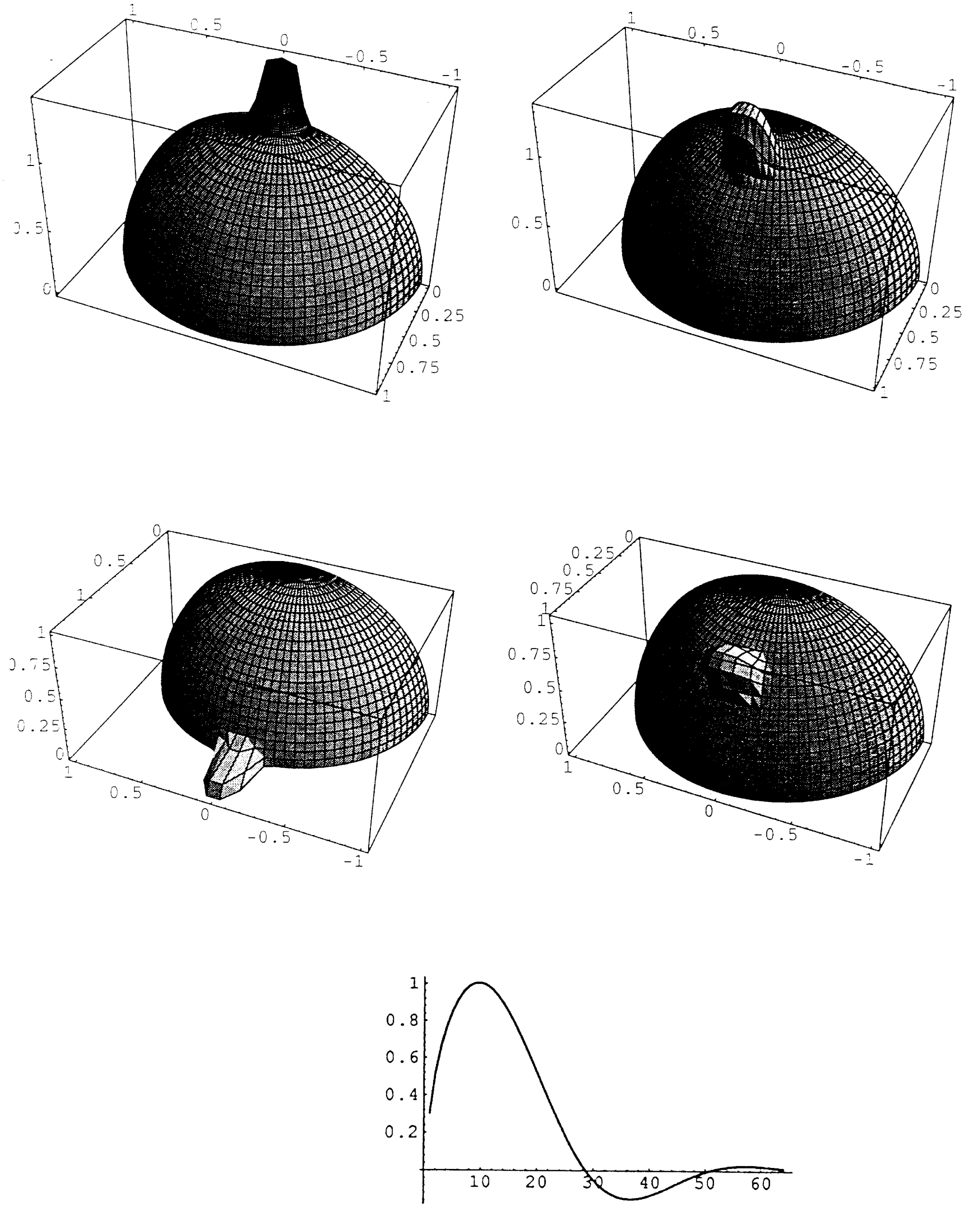

FIG. 3. *Top: Four frames showing a quintic spline bump function approximation to the point mass, at various positions. Clockwise from top left frame: Colatitudes* 0, $\pi/8$, $\pi/2$ *and* $3\pi/8$. *Note that many more samples are required to represent a bump function near the north pole than a bump at other locations. Below: Legendre coefficients of a quintic spline bump function located at the north pole.*

Note that

$$
\begin{aligned}
\int_{S^2} |K_n(x,\omega)|^2\,dx &\le \sum_{l=0}^{1/b} \sum_{q=-l}^{l} (2l+1) \frac{1}{|\hat{f}_0^l|^2} \overline{Y}_q^l(\omega) Y_q^l(\omega) \\
&= \sum_{l=0}^{1/b} (2l+1) \frac{1}{|\hat{f}_l^l|^2} \sum_{q=-l}^{l} \overline{Y}_q^l(\omega) Y_q^l(\omega) \qquad (22) \\
&\le \sum_{l=0}^{1/b} (2l+1)^2 \frac{1}{|\hat{f}_0^1|^2}
\end{aligned}
$$

for $\omega \in S^2$. In the above, we use the orthonormality property of Y_q^l for $|q| \le l$ and $l \ge 0$ along with the addition formula

$$
\sum_{q=-l}^{l} \overline{Y}_q^l(\omega) Y_q^l(\nu) = \frac{2l+1}{4\pi} P_l(\cos\gamma(\omega,\nu)),
$$

where $\gamma(\omega,\nu)$ represents the angle between $\nu, \omega \in S^2$ and $P_l(1) = 1$ for all $l \ge 0$.

For two sequences $\{a_n\}$ and $\{c_n\}$, denote the property $a_n/c_n \to 1$ as $n \to \infty$ by $a_n \sim c_n$. Furthermore, if $f(\cdot)$ is continuous, then $m(\cdot)$ is continuous, hence bounded since S^2 is compact. By using (21) and (22),

$$
\begin{aligned}
n\,\mathrm{Var}(\pi^n(\omega)) &\sim EK_n(X,\omega)\overline{K}_n(X,\omega) \\
&= \int_{S^2} K_n(x,\omega)\overline{K}_n(x,\omega) m(x)\,dx \\
&\le \sup m(z) \int_{S^2} K_n(x,\omega)\overline{K}_n(x,\omega)\,dx \\
&\ll \sum_{l=0}^{1/b} (2l+1)^2 \frac{1}{|\hat{f}_0^l|^2}
\end{aligned}
$$

as $n \to \infty$ for $\omega \in S^2$. Furthermore, since S^2 is compact, a uniform bound can be obtained so that

$$
\sup_{\omega \in S^2} \mathrm{Var}(\pi^n(\omega)) \ll \frac{1}{n} \sum_{l=0}^{1/b} (2l+1)^2 \frac{1}{|\hat{f}_0^l|^2} \qquad (23)
$$

as $n \to \infty$.

Proof of Theorem 3.2. By the assumption that $\pi(\cdot)$ can be represented by its Fourier series, we have (21). By the assumption that

$$
\frac{1}{n} \sum_{l=0}^{1/b} (2l+1)^2 \frac{1}{|\hat{f}_0^l|^2} \to 0
$$

as $n \to \infty$, we have the $\mathrm{Var}(\pi^n(\omega)) \to 0$ as $n \to \infty$, for all $\omega \in S^2$. □

For a differentiable function g on S^2, denote by $g^{(j)}$, the jth derivative for $j = 0, 1, 2, \dots,$ where $g^{(0)} = g$. Now suppose $g^{(j)} \in L^2(S^2)$ for $j = 1, \dots, s$. Then by Lemma 4.1 [Hendriks (1990)], we can write

$$\int_{S^2} |g^{(s)}(\omega)|^2 \, d\omega = \sum_{l \geq 0} \sum_{q=-l}^{l} (l(l+1))^s \, |\hat{g}_q^l|^2, \tag{24}$$

where $\hat{g}_q^l = \int g \overline{Y}_q^l$. Consequently, we can write

$$\|\pi - E\pi^n\|_2^2 \leq \|\pi^{(s)}\|_2^2 b^{2s} \tag{25}$$

as $n \to \infty$. □

PROOF OF THEOREM 3.3. Under the conditions placed on $\hat{f}_0^l$, we have that

$$\frac{1}{n} \sum_{l=0}^{1/b} (2l+1)^2 \frac{1}{|\hat{f}_0^l|^2} \ll \frac{1}{nb^{3+u}}$$

as $n \to \infty$. Therefore, by (23) and (24), we have

$$E\|\pi - \pi^n\|_2^2 \ll \frac{1}{nb^{3+u}} + b^{2s}$$

as $n \to \infty$. This rate is optimized by setting $b = n^{-1/(2s+3+u)}$ which gives the desired result. □

APPENDIX

Let

$$u(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad a(\theta) = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix},$$

where $\phi \in [0, 2\pi)$, $\theta \in [0, \pi)$. The well known *Euler angle decomposition* says that any $g \in SO(3)$ can almost surely be uniquely written as

$$g = u(\phi)a(\theta)u(\psi),$$

where $\phi \in [0, 2\pi)$, $\theta \in [0, \pi)$, $\psi \in [0, 2\pi)$ and are otherwise known as the *Euler angles*. Consider the function

$$D_{q_1 q_2}^l(u(\phi)a(\theta)u(\psi)) = \exp(-iq_1\phi) d_{q_1 q_2}^l(\cos\theta) \exp(-iq_2\psi),$$

where d^l are related to the Jacobi polynomials; see Vilenkin (1968).

The function $D_{q_1 q_2}^l$ can be thought of as matrix entries of a $(2l+1) \times (2l+1)$ matrix. Hence, define

$$D^l(g) = \left[D_{q_1 q_2}^l(g)\right],$$

where $-l \leq q_1, q_2 \leq l$ and $g \in SO(3)$. We note that $\{\sqrt{2l+1} D_{q_1 q_2}^l \colon -l \leq q_1, q_2 \leq l, l = 0, 1, \dots\}$ is an orthonormal basis for $L^2(SO(3))$ and is sometimes referred to as the *rotational harmonics*; see Lo and Eshelman (1979). Furthermore, $\{D^l \colon l = 0, 1, \dots\}$ are the irreducible representations of $SO(3)$.

The mathematical relationship between $SO(3)$ and S^2 is a beautiful result in classical analysis. In terms of the Fourier basis, the relation can be described in terms of the Euler angles, where

$$Y_q^l(\theta, \phi) = \sqrt{\frac{(2l+1)}{4\pi}}\, \overline{D}_{q0}^l(u(\phi)a(\theta)u(\psi)),$$

$\phi \in [0, 2\pi)$, $\theta \in [0, \pi)$, $-l \le q \le l$and $l = 0, 1, \dots$.

PROOF OF LEMMA 3.1. We note that

$$\begin{aligned} m(x) &= \int_{S^2} \pi(\mu) f(\mu^t x)\, d\mu \\ &= \int_{SO(3)} \pi(g\eta) f(\eta^t g^{-1} x)\, dg. \end{aligned}$$

Therefore,

$$\begin{aligned} \hat{m}_q^l &= \int_{\omega \in S^2} \int_{g \in SO(3)} \pi(g\eta) f(\eta^t g^{-1}\omega) \overline{Y}_q^l(\omega)\, dg\, d\omega \\ &= \int_{SO(3)} \pi(g\eta) \int_{S^2} f(\eta^t g^{-1}\omega) \overline{Y}_q^l(\omega)\, d\omega\, dg \\ &= \int_{SO(3)} \pi(g\eta) \int_{S^2} f(\eta^t\omega) \overline{Y}_q^L(g\omega)\, d\omega\, dg \\ &= \int_{SO(3)} \pi(g\eta) \int_{S^2} f(\eta^t\omega) \overline{\sum_{|j| \le l} Y_j^l(\omega) D_{jq}^l(g^{-1})}\, d\omega\, dg \\ &= \sum_{|j| \le l} \int_{SO(3)} \pi(g\eta) \overline{D_{jq}^l(g^{-1})}\, dg \int_{S^2} f(\eta^t\omega) \overline{Y}_j^l(\omega)\, d\omega, \end{aligned}$$

where D_{jq}^l is defined in this for $-l \le j,\ q \le l,\ l \ge 0$.

Now note that $\overline{D_{jq}^l(g^{-1})} = D_{qj}^l(g)$ for all $g \in SO(3)$. Furthermore, it can be shown that

$$\begin{aligned} \int_{SO(3)} \pi(g\eta) \overline{D_{jq}^l(g^{-1})}\, dg &= \int_{SO(3)} \pi(g\eta) \overline{D_{0q}^l(g^{-1})}\, dg \\ &= \int_{SO(3)} \pi(g\eta) D_{q0}^l(g)\, dg \\ &= 2\pi \sqrt{\frac{4\pi}{2l+1}}\, \hat{\pi}_q^l \end{aligned}$$

for $|q| \le l$, $l \ge 0$.

Consequently

$$\hat{m}_q^l = 2\pi \sqrt{\frac{4\pi}{2l+1}}\, \hat{\pi}_q^l \int_{S^2} f(\eta^t\omega) \overline{Y_0^l}(\omega)\, d\omega,$$

for $|q| \le l$, $l \ge 0$ as required. □

Acknowledgments. The original version of this paper was written while the second author was visiting the Department of Statistics, University of Toronto. He would like to take this opportunity to thank them for their hospitality during that visit. In addition, both authors would like to thank the editors as well as an anonymous referee for their comments.

REFERENCES

BAGCHI, P. and GUTTMAN, I. (1988). Theoretical considerations of the multivariate von Mises–Fisher distribution. *Journal of Applied Statistics* **15** 149–169.

BAI, Z. D., RAO, C. R. and ZHAO, L. C. (1988). Kernel estimators of density function of directional data. *J. Multivariate Anal.* **27** 24–39.

BERAN, R. (1979). Exponential models for directional data. *Ann. Statist.* **7** 1162–1178.

BRUNNER, L. J. and LO, A. Y. (1989). Bayes methods for a symmetric unimodal density and its mode. *Ann. Statist.* **17** 1550–1566.

COURANT, R. and HILBERT, D. (1953). *Methods of Mathematical Physics* **1**. Interscience, New York.

DIACONIS, P. (1988). *Group Representations in Probability and Statistics*. IMS, Hayward, CA.

DRISCOLL, J. R. and HEALY, D. M., JR. (1994). Computing Fourier transforms and convolutions on the 2-sphere. *Adv. in Appl. Math.* **15** 202–250.

FAN, J. (1991). Global behavior of deconvolution kernel estimates. *Statist. Sinica* **1** 541–551.

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230.

FISHER, N. I., LEWIS, T. and EMBLETON, B. J. (1987). *Statistical Analysis of Spherical Data*. Cambridge Univ. Press.

FISHER, R. A. (1953). Dispersion on a sphere. *Proc. Roy. Soc. London Ser. A* **217** 295–305.

GINÉ, E. (1975). Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms. *Ann. Statist.* **3** 1243–1266.

GRUNBAUM, F. A., LONGHI, L. and PERLSTADT, M. (1982). Differential operators commuting with finite convolution integral operators: some non-abelian examples. *SIAM J. Appl. Math.* **42** 941–955.

GUTTORP, P. and LOCKHART, R. A. (1988). Finding the location of a signal: A Bayesian analysis. *J. Amer. Statist. Assoc.* **83** 322–330.

HALL, P., WATSON, G. and CABRERA, J. (1987). Kernel density estimation with spherical data. *Biometrika* **74** 751–762.

HEALY, D., MOORE, S. and ROCKMORE, D. (1993). Efficiency and stability issues in the numerical computation of Fourier transforms and convolutions on the 2-sphere. Preprint, Dartmouth College.

HELGASON, S. (1984). *Groups and Geometric Analysis*. Academic Press, New York.

HENDRIKS, H. (1990). Nonparametric estimation of a probability density on a Riemannian manifold using Fourier expansions. *Ann. Statist.* **18** 832–849.

JUPP, P. E. and MARDIA, K. V. (1989). A unified view of the theory of directional statistics, 1975–1988. *Internat. Statist. Rev.* **57** 261–294.

LO, A. Y. and CABRERA, J. (1987). Bayes procedures for rotationally symmetric models on the sphere. *Ann. Statist.* **15** 1257–1268.

LO, J. T. and ESHELMAN, L. R. (1979). Exponential Fourier densities on $SO(3)$ and optimal estimation and detection for rotational processes. *SIAM J. Appl. Math.* **36** 73–82.

MARDIA, K. V. (1972). *Statistics of Directional Data*. Academic Press, New York.

MARDIA, K. V. (1975). Statistics of directional data (with discussion). *J. Roy. Statist. Soc. Ser. B* **37** 349–393.

MARDIA, K. V. and EL-ATOUM, S. A. M. (1976). Bayesian inference for the von Mises–Fisher distribution. *Biometrika* **63** 203–206.

MASLEN, D. (1993). Fast transforms and sampling for compact groups. Ph.D. Dissertation, Dept. Mathematics, Harvard Univ.

ROBBINS, H. (1955). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist.* **1** 157–164. Univ. California Press, Berkeley.

SILVERMAN, B. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

SINGH, R. S. (1992). Empirical Bayes with rates and best rates of convergence in $u(x)C(\theta)\exp(-x/\theta)$-family: Estimation case. *Ann. Inst. Statist. Math.* **44** 435–449.

TALMAN, J. D. (1968). *Special Functions: A Group Theoretic Approach*. W. A. Benjamin, New York.

VILENKIN, N. J. (1968). *Special Functions and the Theory of Group Representations*. Amer. Math. Soc., Providence, RI.

WAHBA, G. (1981). Spline interpolation and smoothing on the sphere. *SIAM J. Sci. Statist. Comput.* **2** 5–16.

WATSON, G. S. (1983). *Statistics on Spheres*. Wiley, New York.

ZHANG, C. H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18** 806–831.

DEPARTMENT OF MATHEMATICS
AND COMPUTER SCIENCE
DARTMOUTH COLLEGE
HANOVER, NEW HAMPSHIRE 03755

DEPARTMENT OF MATHEMATICS
AND STATISTICS
UNIVERSITY OF GUELPH
GUELPH, ONTARIO
CANADA N1G 2W1