

EFFICIENT NONPARAMETRIC DENSITY ESTIMATION ON THE SPHERE WITH APPLICATIONS IN FLUID MECHANICS*

ÖMER EĞECIOĞLU[†] AND ASHOK SRINIVASAN[‡]

Abstract. The application of nonparametric probability density function estimation for the purpose of data analysis is well established. More recently, such methods have been applied to fluid flow calculations since the density of the fluid plays a crucial role in determining the flow. Furthermore, when the calculations involve directional or axial data, the domain of interest falls on the surface of the sphere. Accurate and fast estimation of probability density functions is crucial for these calculations since the density estimation is performed at each iteration during the computation. In particular the values $f_n(X_1), f_n(X_2), \dots, f_n(X_n)$ of the density estimate at the sampled points X_i are needed to evolve the system. Usual nonparametric estimators make use of kernel functions to construct f_n . We propose a special sequence of weight functions for nonparametric density estimation that is especially suitable for such applications. The resulting method has a computational advantage over kernel methods in certain situations and also parallelizes easily. Conditions for convergence turn out to be similar to those required for kernel-based methods. We also discuss experiments on different distributions and compare the computational efficiency of our method with kernel based estimators.

Key words. probability density, nonparametric estimation, fluid mechanics, convergence, kernel method, efficient algorithm

AMS subject classifications. 65U05, 62G05, 62G07, 65D15, 65Y20

PII. S1064827595290462

1. Introduction. Nonparametric density estimation is the problem of the estimation of the values of a probability density function, given samples from the associated distribution. No assumption is made about the type of the distribution from which the samples are drawn. This is in contrast to parametric estimation in which the density is assumed to come from a given family, and the parameters are then estimated by various statistical methods. Early contributors to the theory of nonparametric estimation include Smirnov [21], Rosenblatt [16], Parzen [15], and Chentsov [3]. Extensive descriptions of various approaches to nonparametric estimation along with a comprehensive bibliography can be found in books by Silverman [23] and Nadaraya [14]. More recent developments are presented in books by Scott [18] and Wand and Jones [27]. Results of the experimental comparison of some widely used methods appear in [10, 25].

In addition to data analysis, an important application of nonparametric density estimation is in computational fluid mechanics. When the flow calculations are performed in a Lagrangian framework, a set of points in space are evolved through time using the governing equations. In time, points that were initially close can move apart, leading to mesh distortion and numerical difficulties. Problems with mesh distortion can be eliminated to a certain extent by the use of smoothed particle hydrodynamics (SPH) techniques [2, 13, 9, 12]. SPH treats the points being tracked as samples coming from an unknown probability distribution. These calculations often require the computation of the values of not only the unknown density, but its gradient as well.

*Received by the editors August 16, 1995; accepted for publication (in revised form) August 3, 1999; published electronically June 13, 2000.

<http://www.siam.org/journals/sisc/22-1/29046.html>

[†]Department of Computer Science, University of California, Santa Barbara, CA 93106 (omer@cs.ucsb.edu).

[‡]Department of Mathematics, Indian Institute of Technology, Bombay, India (ashok@math.iitb.ernet.in).

In contrast to applications concerned with the display of the density, where it is sufficient to estimate the density on some grid, in these fluid flow calculations the density estimate is required at each sample point. Another difference in these two types of applications is that when dealing with data analysis, one is usually concerned with the optimal accuracy one can get for a given sample size. In fluid flow calculations, where additional “data” can be obtained with increased discretization, one is usually more concerned with the optimal variation of the computational effort as a function of error.

In some applications, for example, in problems involving directional data [24], the samples lie on the unit circle S^1 or along the surface of the unit sphere S^2 . A special case of directional data is axial data, in which the density is symmetric about the center of the circle or the sphere, that is, $f(\vec{x}) = f(-\vec{x})$.

Various methods have been proposed for nonparametric density estimation in mathematical statistics, such as the kernel [15, 1, 28] and the orthogonal series methods [17, 11]. The kernel method has been extensively studied, and it is probably the most popular scheme in applications such as SPH. In this method, the value of the density at the point x is estimated as

$$(1) \quad f_n(x) = \frac{1}{nA_h} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right),$$

where f_n is the estimate of the density given a sample of size n , X_i are the positions of the samples drawn from a probability distribution with an unknown density function f , K is a kernel function, h is the window width, and A_h is a normalization factor to make f_n into a probability density. One of the drawbacks of the kernel method is the computational cost involved. Even though it is possible to reduce the cost in the one-dimensional case using the expansion of a polynomial kernel and an updating strategy [19], this strategy cannot be easily extended to higher dimensions [5]. Binning methods [5] can be used in any dimension. However, since the density in this case is evaluated on a uniform grid, this method is not suitable for the fluid flow calculations in which we are interested, where an estimate is required at each sample point.

We propose a cosine-based weight function estimator $c_m(x)$ for nonparametric density estimation, which is a special case of the class of estimators that form a δ sequence [26, 28]. This estimator is similar to the kernel estimator but has the ease of evaluation of a series expansion. The role of the window width parameter h of the kernel method is replaced by a smoothing parameter m in our method, and f_n is now of the form

$$(2) \quad f_n(x) = \frac{1}{n} \sum_{i=1}^n c_m(x - X_i).$$

Our choice of c_m is particularly suitable for applications in fluid flow calculations where the values $f_n(X_1), f_n(X_2), \dots, f_n(X_n)$ at the sampled directions themselves are required at each point in each time step in the flow simulation. We show that with this estimator the required n values can be computed efficiently using only $O(m^{1+d}n)$ arithmetic operations for directional data and $O(m^d n)$ arithmetic operations for axial data in d dimensions, where m need not be large as long as it increases without bound with n . This is in contrast to the $O(n^2)$ operations required by the kernel method for this computation in the worst case and an expected complexity of $O(h^d n^2)$ with kernels having bounded support. However, in the special case of $d = 1$, the complexity of the kernel method can be reduced to linear after an initial sorting step.

We derive conditions under which the sequence of estimated density functions f_n constructed in this fashion converge to the unknown density f , and experimentally verify the accuracy and the efficiency of our method in practical test cases. Experiments and theoretical analyses also indicate how m should vary with n for optimal accuracy.

The paper is organized as follows. In section 2 we define the weight function estimator c_m and give the conditions for the convergence of the mean integrated square error (MISE) when the sample space is S^1 (Theorem 3). The conditions guarantee that

$$\int E(f_n(x) - f(x))^2 dx \rightarrow 0$$

as $n \rightarrow \infty$. We also present corresponding results for S^2 . In section 3 schemes for efficient computation of these estimates on S^1 and S^2 are presented. In sections 4 and 5, we describe experimental results with our estimator and compare it with the kernel method for some distributions encountered in practice. Our experiments imply a net savings on the number of operations performed over kernel methods in certain situations and also verify the formula found for the optimal choice of m . The results show that the kernel method and our estimator perform well in different settings, and thus complement each other. The main conclusions are presented in section 6. The appendix contains additional test results.

2. The cosine estimator and the convergence of MISE. In this section, we first mention some related work done on spherical data; then we define our estimator and derive conditions for its convergence for directional data on the circle, and give corresponding results for directional and axial data on the sphere and axial data on the circle.

The kernel method for nonparametric density estimation for directional and axial data is discussed in [6, 8]. While dealing with directional data, Fisher, Lewis, and Embleton [6] recommend using the following kernel:

$$(3) \quad W_n(P, P_i) = \left[\frac{C_n}{4\pi \sinh(C_n)} \right] \exp [C_n(x^T X_i)].$$

For axial data they recommend the kernel

$$(4) \quad W_n(P, P_i) = A(C_n) \exp [C_n(x^T X_i)],$$

where $A(C_n)$ normalizes W to a probability density function, and C_n is the reciprocal of h used in the definition of kernel estimators. x and X_i are the Cartesian representation of points P and P_i , respectively, and $x^T X_i$ is the inner product of these two vectors. $W_n(P, P_i)$ plays the role of $K(x - X_i)$ of (1). Hall, Watson, and Cabrera [8] analyze estimators for directional data with the $x - X_i$ term in (1) replaced by $1 - x^T X_i$. Observe that the term $x^T X_i$ is the cosine of the angle between the points P and P_i , and therefore $1 - x^T X_i$ is a measure of the distance along the surface of the sphere between points P and P_i . Inner product plays a crucial role in these estimators. We consider an estimator that can be written in terms of powers of the inner product, the power playing the role of the smoothing parameter. This enables us to expand the estimator in a series and facilitates fast computation.

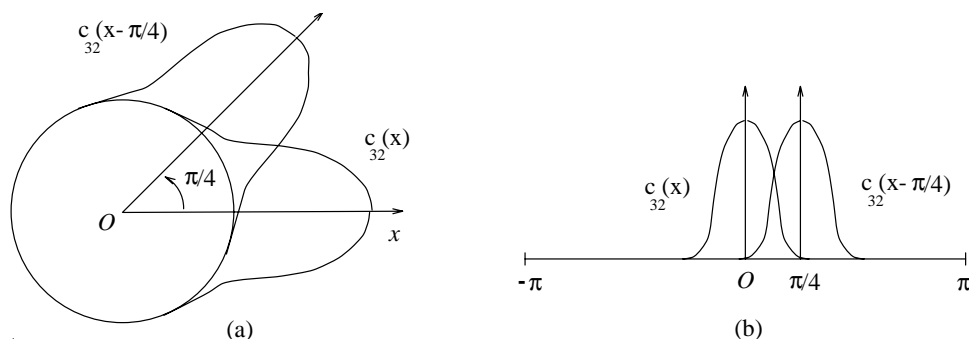


FIG. 1. The functions $c_{32}(x)$ and $c_{32}(x - \frac{\pi}{4})$ on S^1 and on $[-\pi, \pi]$.

2.1. The case of S^1 . We first define our estimator on S^1 . Assume X_j , $j = 1, 2, \dots, n$, is a sequence of independently and identically distributed (i.i.d.) random variables (observations) for directional data on $[-\pi, \pi]$ with probability density function $f \in C^2[-\pi, \pi]$. We impose the additional condition that $f(-\pi) = f(\pi)$ since the random variables X_j are defined on the unit circle S^1 .

As an estimator of the density of directional data $f(x)$, $x \in [-\pi, \pi]$, we consider a nonparametric estimator of the form given by (2) with

$$(5) \quad c_m(x) = \frac{1}{A_m} \cos^{2m} \left(\frac{x}{2} \right)$$

on $[-\pi, \pi]$. The normalization factor A_m given below makes $c_m(x)$ integrate to 1 on $[-\pi, \pi]$:

$$A_m = \int_{-\pi}^{\pi} \cos^{2m} \left(\frac{x}{2} \right) dx.$$

Making use of a table of integrals such as Gradshteyn and Ryzhik [7] and by using Stirling's formula, it can be shown that

$$(6) \quad A_m = \frac{\pi}{2^{2m-1}} \binom{2m}{m} \sim \frac{2\sqrt{\pi}}{\sqrt{m}}.$$

As examples, the functions $c_m(x)$ and $c_m(x - \frac{\pi}{4})$ for $m = 32$ are shown on S^1 in Figure 1(a) and on the interval $[-\pi, \pi]$ in Figure 1(b).

We wish to find sufficient conditions under which the sequence of estimators f_n converges to f in the MISE sense. In order to do this, we first show the convergence of the bias and then derive the conditions under which the variance converges to 0. We shall then use these results to prove convergence of MISE on S^1 .

First we show that as $m \rightarrow \infty$, the expected value of the estimate $f_n(x)$ approaches the actual density $f(x)$ uniformly for any given n .

LEMMA 1. Suppose $f \in C^2[-\pi, \pi]$ and let $f_n(x)$ be as given in (2). Then $E f_n(x) \rightarrow f(x)$ as $m \rightarrow \infty$ uniformly, independently of n .

Proof.

$$(7) \quad E f_n(x) = \int_{-\pi}^{\pi} c_m(x - s) f(s) ds,$$

as shown in Silverman [20] and Whittle [29]. By a change of variable

$$(8) \quad \int_{-\pi}^{\pi} c_m(x-s)f(s)ds = \int_{x-\pi}^{x+\pi} c_m(-y)f(x+y)dy = \int_{x-\pi}^{x+\pi} c_m(y)f(x+y)dy$$

since $c_m(y) = c_m(-y)$. By using the periodicity of c_m and f , along with (7), (8), and the mean value theorem,

$$Ef_n(x) = \int_{-\pi}^{\pi} c_m(y)f(x+y)dy = \int_{-\pi}^{\pi} c_m(y)(f(x) + f'(x)y + f''(\xi_x(y))y^2/2)dy,$$

where $\xi_x(y)$ is some point between x and y . Therefore

$$Ef_n(x) = f(x) \int_{-\pi}^{\pi} c_m(y)dy + \int_{-\pi}^{\pi} c_m(y)f'(x)ydy + \int_{-\pi}^{\pi} c_m(y)f''(\xi_x(y))y^2/2dy.$$

From (6), the first integral evaluates to 1, and since $yc_m(y)$ is an odd function, the second integral evaluates to 0. Let $2M' = \max_{x \in [-\pi, \pi]} |f''(x)|$. We then have the following estimate for the bias:

$$\begin{aligned} |Ef_n(x) - f(x)| &\leq \left| \int_{-\pi}^{\pi} c_m(y)f''(\xi_x(y))y^2dy \right| \\ &\leq M' \int_{-\pi}^{\pi} c_m(y)y^2dy = M' \int_{-\pi}^{\pi} \frac{1}{A_m} \cos^{2m}(y/2)y^2dy. \end{aligned}$$

For any δ such that $0 < \delta \leq \pi$,

$$(9) \quad \begin{aligned} |Ef_n(x) - f(x)| &\leq M' \int_{|y| < \delta} \frac{1}{A_m} \cos^{2m}(y/2)y^2dy + M' \int_{|y| \geq \delta} \frac{1}{A_m} \cos^{2m}(y/2)y^2dy \\ &\leq M'\delta^2 \int_{|y| < \delta} \frac{1}{A_m} \cos^{2m}(y/2)dy + \frac{2\pi^3 M'}{3} \frac{1}{A_m} \cos^{2m}(\delta/2) \end{aligned}$$

since $\cos y$ decreases as $|y|$ increases on the interval under consideration. Furthermore, the integral in (9) is bounded above by 1. Therefore,

$$(10) \quad \begin{aligned} |Ef_n(x) - f(x)| &\leq M'\delta^2 + \frac{2\pi^3 M'}{3} \frac{\cos^{2m}(\delta/2)}{A_m} \\ &\leq M'\delta^2 + \frac{2\pi^3 M'}{3} \frac{(1 - \delta^2/8 + \delta^4/384)^{2m}}{A_m} \\ &= M'\delta^2 + \frac{2\pi^3 M'}{3} \frac{(1 - \delta^2(1 - \delta^2/48)/8)^{2m}}{A_m}. \end{aligned}$$

In order to get a bound, we will choose δ as a function of m . If we take $\delta \rightarrow 0$ as $m \rightarrow \infty$, then $(1 - \delta^2(1 - \delta^2/48)/8)^{2m} \rightarrow \exp(-m\delta^2(1 - \delta^2/48)/4)$. Observe that if $m\delta^2 \rightarrow \infty$ as $m \rightarrow \infty$, then this term decays exponentially. The second expression in (10) is the product of this term and $1/A_m = O(\sqrt{m})$, and thus the product approaches 0 since the exponential decay dominates. In order to get a good bound on the first term of (10), we wish to choose δ satisfying the condition that $m\delta^2 \rightarrow \infty$ such that the δ^2 is as small as possible. We can choose $\delta = 1/m^{\frac{1}{2}-\epsilon}$, where $\epsilon > 0$ is arbitrarily small. Thus M'/m is an asymptotic bound on the bias. Furthermore, the bound is independent of x ; hence, the convergence is uniform. \square

LEMMA 2. Suppose $f \in C^2[-\pi, \pi]$ and let $f_n(x)$ be as given in (2). Then $\text{var } f_n(x) \rightarrow 0$ uniformly as $n \rightarrow \infty$, provided $m \rightarrow \infty$ as $n \rightarrow \infty$, and $m = o(n^2)$.

Proof.

$$\text{var } f_n(x) = \frac{1}{n} \int_{-\pi}^{\pi} c_m(x-s)^2 f(s) ds - \frac{1}{n} \left\{ \int_{-\pi}^{\pi} c_m(x-s) f(s) ds \right\}^2$$

as shown in Whittle [29]. As a consequence of Lemma 1, the second integral approaches $f(x)$ asymptotically, and hence, the second term approach 0 since f is bounded. It thus suffices to show the convergence of the first integral to 0.

Let $M = \max_{x \in [-\pi, \pi]} |f(x)|$. As in Lemma 1, making a change of variable $y = x - s$ and using periodicity as in (8), we get

$$\frac{1}{n} \int_{-\pi}^{\pi} c_m(y)^2 f(x+y) dy \leq \frac{MA_{2m}}{nA_m^2} \rightarrow \frac{M}{\sqrt{8\pi}} \frac{\sqrt{m}}{n},$$

where the expression on the right-hand side is a consequence of the asymptotic expression for A_m in (6). Since $m/n^2 \rightarrow 0$ as $n \rightarrow \infty$, the above integral converges to 0. Since m is independent of x , the convergence is uniform. Therefore the variance of $f_n(x)$ converges uniformly to 0 under the conditions of the lemma. \square

Note that the bound on the bias for the cosine method given by Lemma 1 is of the form

$$(11) \quad |Ef_n(x) - f(x)| \leq \frac{1}{2} \max |f''(x)|/m,$$

and the bound for the variance given by Lemma 2 is

$$(12) \quad \text{var } f_n(x) \leq \frac{1}{\sqrt{8\pi}} \max |f(x)| \sqrt{m}/n.$$

Therefore, the role played by m for the cosine method is the same as h^{-2} of the kernel based methods, where h is the window width of the kernel estimator. In other words when $m \sim h^{-2}$, the bounds on the bias and the variance of the cosine estimator are in accordance with the asymptotic behavior of the kernel method found in Silverman [23]. Such similarity of rates of convergence is to be expected since the cosine estimator is essentially like the kernel estimator, though the forms of the functions differ. It will be shown later that the main advantage of the cosine estimator lies in its computational efficiency.

THEOREM 3. Suppose $f \in C^2[-\pi, \pi]$ and $f_n(x)$ is as given in (2). If $m \rightarrow \infty$ as $n \rightarrow \infty$, and $m = o(n^2)$, then

$$\text{MISE} = \int_{-\pi}^{\pi} E(f_n(x) - f(x))^2 dx \rightarrow 0$$

as $n \rightarrow \infty$.

Proof.

$$\int E(f_n(x) - f(x))^2 dx = \int (Ef_n(x) - f(x))^2 dx + \int \text{var } f_n(x) dx$$

as shown in Whittle [29]. From Lemmas 1 and 2, each of the integrals approaches 0. Hence, the MISE converges to 0. \square

In fact, the MISE is of the form

$$(13) \quad \text{MISE} = O(1/m^2) + O(\sqrt{m}/n),$$

where the asymptotic bounds on the constants are as given in (11) and (12). However, as we shall explain later, the exact asymptotic constants are not all that important for practical computations.

Conditions for the convergence of the estimates and its derivatives on the real line instead of S^1 can be found in [4]. Next we consider the case when the directional data lie along the surface of the unit sphere.

2.2. The case of S^2 . Let X_j , $j = 1, 2, \dots, n$, be a sequence of i.i.d. random variables (observations) with values on the surface of the unit sphere S^2 centered at the origin in \mathbb{R}^3 . Suppose that the probability density function $f(x)$ of the X_j has bounded second derivatives. We consider a nonparametric estimator of the form

$$(14) \quad f_n(x) = \frac{1}{n} \sum_{i=1}^n c_m(x, X_i)$$

for some m to be determined as a function of n . The c_m are defined in this case as follows. If α_{xX} denotes the angle between points x and X , then

$$(15) \quad c_m(x, X) = \frac{1}{A_m} \cos^{2m} \left(\frac{\alpha_{xX}}{2} \right).$$

The normalizing factor A_m is given below:

$$A_m = \frac{4\pi}{m+1}.$$

Through a derivation along the lines of the case of the circle, the following theorem can be proved for the convergence of the estimators.

THEOREM 4. *Suppose $f \in C^2(S^2)$ and let $f_n(x)$ be as given in (14). If $m \rightarrow \infty$ as $n \rightarrow \infty$ and $m = o(n)$, then*

$$\text{MISE} = \int E(f_n(x) - f(x))^2 dx \rightarrow 0.$$

Analogous to (13), the form of the MISE is found to be

$$(16) \quad \text{MISE} = O\left(\frac{1}{m^2}\right) + O\left(\frac{m}{n}\right).$$

From this expression for MISE we see that as in the case of S^1 , $m \sim h^{-2}$, where h is the window width of the kernel estimator.

When dealing with axial data, we can consider the following axial estimator for spherical data:

$$c_m(x, X) = \frac{1}{A_{2m}} \cos^{2m}(\alpha_{xX}).$$

We can also define a corresponding estimator on the circle, where we take the cosine of the arc length between two points, instead of the cosine of half the arc length as in the case of directional data. The relationship between the asymptotic MISE, m and n , is the same as in (13) and (16) for the cases of the circle and the sphere, respectively.

3. Efficient evaluation of the density estimator. In this section, we shall describe an efficient algorithm for the computation of the density estimates $f_n(x)$ evaluated at a set of n observed points X_1, X_2, \dots, X_n on the unit circle S^1 ($d = 1$ case) and on the unit sphere S^2 ($d = 2$ case). We also show that if the value of f_n at some arbitrary x is desired, then this is also easily accomplished once $f_n(X_1), f_n(X_2), \dots, f_n(X_n)$ have been computed. The efficiency of our method is based on the fact that $f_n(x)$ is defined in terms of the functions $c_m(x)$ as given in (5) or (15).

Suppose we represent the positions of the observed points X_1, X_2, \dots, X_n by their Cartesian coordinates. We show that for any x , $f_n(x)$ can be expressed as a polynomial of total degree m in the coordinates of x . The coefficients of this polynomial can be determined in turn from the coordinates of the X_i . Moreover, the coefficients are the sums of the contributions due to each X_i independently.

First we consider the case of directional data on S^1 . From (2), (5), and the half-angle formula for cosine we get

$$(17) \quad f_n(x) = \frac{1}{nA_m} \sum_{i=1}^n \left(\frac{1 + \cos(x - X_i)}{2} \right)^m.$$

Denote the points on S^1 corresponding to the angles x and X_i for $i = 1, 2, \dots, n$, by $\vec{x} = (x_1, x_2)$ and $\vec{X}_i = (X_{i1}, X_{i2})$ in Cartesian coordinates and let $\langle \cdot, \cdot \rangle$ represent the standard inner product on \mathbb{R}^2 . Then $\cos(x - X_i) = \langle \vec{x}, \vec{X}_i \rangle$. Substituting this into (17) we get

$$(18) \quad f_n(x) = \frac{1}{n2^m A_m} \sum_{i=1}^n \left(1 + \langle \vec{x}, \vec{X}_i \rangle \right)^m = \frac{1}{n2^m A_m} \sum_{i=1}^n (1 + x_1 X_{i1} + x_2 X_{i2})^m.$$

The expression (18) is a polynomial of degree m in x_1 and x_2 . For a fixed m , we can compute the coefficients by adding the contribution of each X_i as follows. Using the multinomial theorem and (18)

$$(19) \quad f_n(x) = \frac{1}{n2^m A_m} \sum_{i=1}^n \sum_{r+s \leq m} \binom{m}{r, s} x_1^r x_2^s X_{i1}^r X_{i2}^s,$$

where the inner summation is over all $r + s \leq m$ with $r, s \geq 0$. Changing the order of summation

$$(20) \quad f_n(x) = \frac{1}{2^m A_m} \sum_{r+s \leq m} \binom{m}{r, s} M(r, s) x_1^r x_2^s,$$

where

$$M(r, s) = \frac{1}{n} \sum_{i=1}^n X_{i1}^r X_{i2}^s,$$

and A_m is as given by (6). If we use the asymptotic expression for A_m in (6) for computational ease, then (20) simplifies to

$$(21) \quad f_n(x) = \frac{\sqrt{m}}{2^{m+1} \sqrt{\pi}} \sum_{r+s \leq m} \binom{m}{r, s} M(r, s) x_1^r x_2^s$$

TABLE 1
Computational complexity of the cosine estimator.

	Circle ($d = 1$)	Sphere ($d = 2$)
Axial data	mn	m^2n
Directional data	m^2n	m^3n

for large m . Now we consider the number of operations required for the evaluation of $f_n(x)$ given the observations X_1, X_2, \dots, X_n . The powers X_{i1}^r and X_{i2}^r for a fixed i and $r = 1, 2, \dots, m$ can be computed with $O(m)$ multiplications. Doing this for $i = 1, 2, \dots, n$ requires $O(mn)$ multiplications. After the conclusion of this step, each of the $O(m^2)$ averages $M(r, s)$ for a given r and s can be computed with an additional $O(n)$ operations. Since there are a total of $O(m^2)$ terms in $f_n(x)$ corresponding to the pairs r, s with $0 \leq r, s, \leq m$, this means that the coefficients of the polynomial in (20) or (21) can be computed with a total of $O(m^2n)$ arithmetic operations.

Once the coefficients of $f_n(x)$ have been computed, to evaluate $f_n(x)$ with $\vec{x} = (x_1, x_2)$ we calculate the powers x_1^r and x_2^r for $r = 1, 2, \dots, m$ in $O(m)$ operations. Since the coefficients are already available, the remaining computation in (20) requires only an additional $O(m^2)$ multiplications and additions. The results for different cases are summarized in Table 1.

Remark. For our MISE convergence condition for S^1 (Theorem 3) to be satisfied, m must increase without bound with n . Theoretically, we can take m to be as slowly increasing as we like. Then the above result implies that the computation of the density at all of the sample points can be accomplished using only about $O(n)$ arithmetic operations if the magnitude of m gives acceptable accuracy for $f_n(x)$. The problem with taking m to be too slowly growing is that the magnitude of m controls the error terms in our convergence proofs.

An efficient algorithm for the evaluation of $f_n(\vec{x})$ for directional data on S^2 is constructed similarly. When $\vec{X}_1, \vec{X}_2, \dots, \vec{X}_n$ are observations on S^2 drawn from an unknown density, it can be shown [4] that

$$(22) \quad f_n(x) = \frac{m+1}{\pi 2^{m+2}} \sum_{r+s+t \leq m} \binom{m}{r, s, t} M(r, s, t) x_1^r x_2^s x_3^t,$$

where the summation is over all $r + s + t \leq m$ with $r, s, t \geq 0$, and

$$M(r, s, t) = \frac{1}{n} \sum_{i=1}^n X_{i1}^r X_{i2}^s X_{i3}^t.$$

This time the coefficients of the polynomial in (22) can be computed with a total of $O(m^3n)$ arithmetic operations. After this preprocessing, each evaluation of $f_n(x)$ for $x \in S^2$ requires only an additional $O(m^3)$ operations.

Corresponding results can be derived for axial data as summarized in Table 1.

It should also be noted that we needed the Cartesian representation of the data. If the data are given in spherical coordinates, then there will be an additional overhead for obtaining the Cartesian representation. However, this overhead takes only linear time and so will be negligible for sufficiently large data. Furthermore, it has been shown [24] that for an important class of applications, Cartesian coordinates are preferable to spherical coordinates, as the latter system is not numerically stable for solving the differential equations that arise.

In the subsequent parts of this section we shall compare the computational efficiency of our scheme with that of the kernel method.

3.1. Parallelization. One of the advantages of the computational strategy described above is the ease of parallelization. Parallelization is required in many fluid flow calculations due to the large sizes of the systems. The kernel method is somewhat difficult to parallelize. If we use an efficient kernel implementation that performs kernel evaluations only for those points which are within a cut-off distance h of the given sample, then an efficient implementation of the parallelization requires periodic load balancing and domain decomposition so that points that are close by remain on the same processor, and so that each processor has roughly the same load in terms of the computational effort. Also, the communication pattern for the kernel method is not very regular. In contrast, parallelization for the cosine estimator can easily be accomplished by a global reduction operation, for which efficient implementations are usually available. This method requires the same computational effort for each point, and so the load is easily balanced by having the same number of points in each processor. Domain decomposition does not play an important role in this scheme, since the points can be on any processor.

3.2. Theoretical comparison of the kernel and the cosine estimators.

Now we analyze the computational efficiency of the kernel and the cosine density estimation methods. An important measure of the efficiency of the algorithms is not just the convergence rate of the error with sample size n , but the relationship of the computational effort C required as a function of the error E . For the kernel estimator, we can write the asymptotic MISE E as

$$(23) \quad E = O\left(h^4 + \frac{1}{h^d n}\right),$$

where h is the smoothing parameter, $d = 1, 2$ is the dimension, and n is the sample size. The computational effort required for this nonparametric density estimation can be expressed as

$$(24) \quad C = O(h^\beta n^\gamma),$$

where $0 \leq \beta \leq d$ and $\gamma = 1$ or 2 depending on the details of the algorithm used. For a given sample size, (23) gives the optimal h as $h \sim n^{-1/(d+4)}$. However, since the equation for the computational complexity also depends on h , we need to consider the possibility that a value of h smaller than this optimal value may actually result in a lower computational effort.

Let us consider a variation of h with n of the form

$$(25) \quad h = O(n^{-\alpha}), \quad 1/(d+4) \leq \alpha < 1/d.$$

From (25), (24), and (23) we obtain

$$C = n^{-\alpha\beta+\gamma} \Rightarrow n = C^{\frac{1}{\gamma-\beta\alpha}} \Rightarrow E = C^{\frac{-4\alpha}{\gamma-\beta\alpha}} + C^{\frac{d\alpha-1}{\gamma-\beta\alpha}}.$$

For minimum error, the exponent of both the terms on the right should be the same, otherwise the error due to the higher term will dominate. This leads to $\alpha = 1/(d+4)$ which is the same as the value of optimal α for a given n . If we let h_{optn} represent the optimal h minimizing the MISE for a given n , and h_{optC} represent the optimal

h minimizing the computational effort as a function of the error, then the expression derived above for α does not necessarily imply that $h_{optn} = h_{optC}$ since the relation $h_{optn} = kh_{optC}$ would still satisfy the expression for α for some constant k . If $k < 1$, then we can choose a suboptimal value of h in order to improve the speed of the algorithm.

The optimal variation of error with computational complexity using this value of α is given by

$$C = E^{\frac{\beta}{4} - \frac{(d+4)\gamma}{4}}.$$

Let us now consider the cosine estimator. We can write the asymptotic MISE as follows:

$$E = O\left(m^{-2} + \frac{m^{d/2}}{n}\right)$$

where E is the MISE, m is the smoothing parameter, d is the dimension ($d = 1$ for the circle and $d = 2$ for the sphere), and n is the sample size. The computational effort required for this estimator can be expressed as $C = O(m^\xi n)$, where ξ can be determined from Table 1. The expression for C above is the same as (24) with $\gamma = 1$, and $\beta = -2\xi$, recalling that m behaves as h^{-2} . By an analysis similar to the previous case we can show that the optimal variation of error with computational complexity is given by

$$C = E^{-\frac{(d/2+\xi)}{2}-1}.$$

As examples, for the cosine estimator on the circle ($d = 1$) with axial data ($\xi = 1$) and directional data ($\xi = 2$), the computational complexity and the error are related by $C = E^{-1.75}$ and $C = E^{-2.25}$, respectively.

The complexity of the kernel estimator is the same for axial and directional data. However, several different possibilities exist depending on how efficient the implementation of the estimator is. If we consider estimators of the form given by (3) and (4), then we have $d = 1$, $\beta = 0$, $\gamma = 2$, and therefore

$$(26) \quad C = E^{-2.5}.$$

However, if we consider a kernel with bounded support, and use an efficient implementation of the algorithm that computes the kernel only for those points that have a nonzero contribution, then the *expected* value of β is 1 and we get $C = E^{-2.25}$ for data on the circle. Note that the worst case complexity remains as in (26). For the $d = 1$ case, we can consider an efficient algorithm using polynomial kernels and updating [19], which uses a linear time after an initial $O(n \log n)$ sorting step. In this case $\beta = 0$, $\gamma = 1$, and $C = E^{-1.25}$, which means that the kernel method has a better asymptotic complexity than the cosine kernel. However there appears to be no natural generalization of this update strategy to higher dimensions [5].

Results for the different cases can be determined in the manner demonstrated above and are presented in Table 2. We wish to mention that the exact asymptotic constants in Theorem 3 are not quite that important (compared with the exponent on E), since asymptotically the slowdown incurred by the cache misses dominates the overall running time. We can expect that the simpler memory access pattern of our estimator will make it advantageous over the kernel method in the asymptotic case.

TABLE 2

Variation of the optimal computational effort versus MISE. The numbers in the table represent η , where the relationship between the computational effort C and $E = \text{MISE}$ is given by $C = E^{-\eta}$. *Does not take into account an initial sorting step.

Estimator	Circle ($d = 1$)	Sphere ($d = 2$)
Cosine, axial data	1.75 ($\xi = 1$)	2.5 ($\xi = 2$)
Cosine, directional data	2.25 ($\xi = 2$)	3.0 ($\xi = 3$)
Kernel, worst case	1.25 * ($\beta = 0, \gamma = 1$)	3.0 ($\beta = 0, \gamma = 2$)
Kernel, expected case	1.25 * ($\beta = 0, \gamma = 1$)	2.5 ($\beta = 2, \gamma = 2$)

Since the worst case complexities of the kernel method and the cosine estimator for directional data on the sphere have the same order, the relative efficiencies of the methods can be tested only through experiments. Similarly, since the worst case complexity of the cosine estimator for axial data on the sphere is the same as the expected case for an efficient implementation of the kernel estimator, we need to perform experiments to test the relative merits of the two estimators.

4. Experimental results. We performed numerical experiments for axial and directional data on the circle and sphere in order to test the effectiveness of our estimator. We first plot estimates for known distributions and then demonstrate that the MISE follows expected trends for certain distributions. We finally compare the computational efficiency of our estimator with that of kernel methods. More empirical results are presented in the appendix.

We consider the function $\psi(\phi, \theta) = \exp(US \cos^2(\theta))/A$ where A normalizes the function to be a density on the surface of a sphere and S is a known function of U . The angles ϕ and θ are the azimuth and the elevation in spherical coordinates. This is the solution to a particular problem in fluid mechanics. In Figure 2(a) we present a typical estimate for the $d = 1$ version of the above density function where ψ was taken to be $\psi(x) = \exp(US \cos^2(x))/A$. In this figure, we take the data to be directional. However, since this function is symmetric with respect to the center of the circle, we can consider the data as axial and use the axial estimator. We can see from Figure 2(b) that this requires a much smaller value of m .

In Figure 3(a) the MISE is compared versus m and n for the one-dimensional ψ distribution using the axial cosine estimator. We also compare with one case of the directional estimator in order to show the benefit of using the axial estimator. In Figure 3(b) the MISE is compared versus m and n for the two-dimensional version of the ψ distribution on the surface of a sphere using the directional estimator.

We next present results for experiments comparing the speed of the cosine and the kernel estimators. We consider the optimal variation of the computational effort with MISE. In order to get the optimal computational effort for a given MISE, we allow for the possibility that we may require different sample sizes for the kernel and the cosine estimators. This is justified because in iterative calculations one can easily change the “sample” size by changing the discretization of the system. We have performed these comparisons only for spherical data. The case of data on the circle was not considered because of the asymptotic analyses of the previous section which clearly indicate that the linear kernel algorithm in the one-dimensional case will outperform the cosine estimator. However, in a parallel implementation, the sorting step for the linear kernel algorithm may be slow, and then one may wish to consider the cosine estimator.

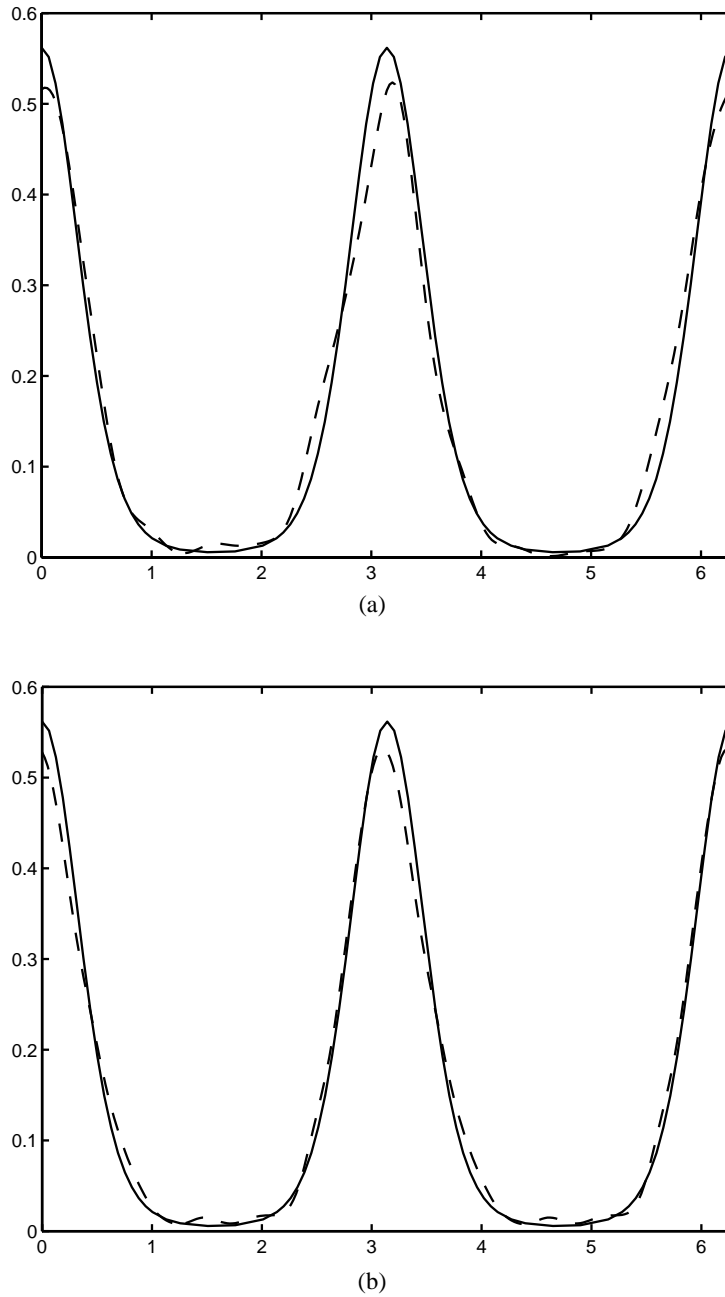


FIG. 2. Cosine estimates for $\psi(x) = \exp(US \cos^2(x))/A$, the one-dimensional case of the ψ function defined above. The solid line represents the true density. (a) The dashed line represents the directional estimate with $m = 100$, $n = 500$. (b) The dashed line represent the axial estimate with $m = 40$, $n = 500$.

The following kernel was chosen for the comparisons:

$$K(x) = \begin{cases} \frac{1}{A}(1 - 1.5x^2 + 0.75x^3), & x \in [0, 1], \\ \frac{1}{A}0.25(2 - x)^3, & x \in [1, 2], \\ 0 & \text{otherwise,} \end{cases}$$

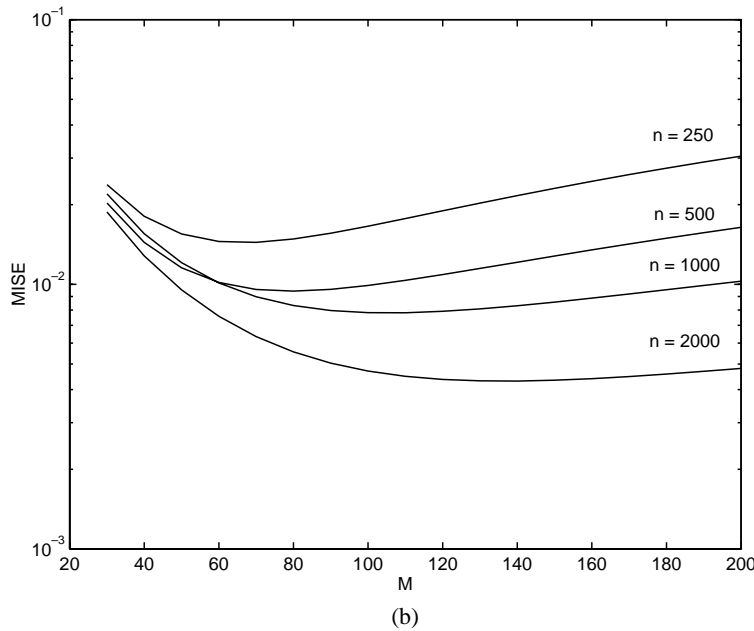
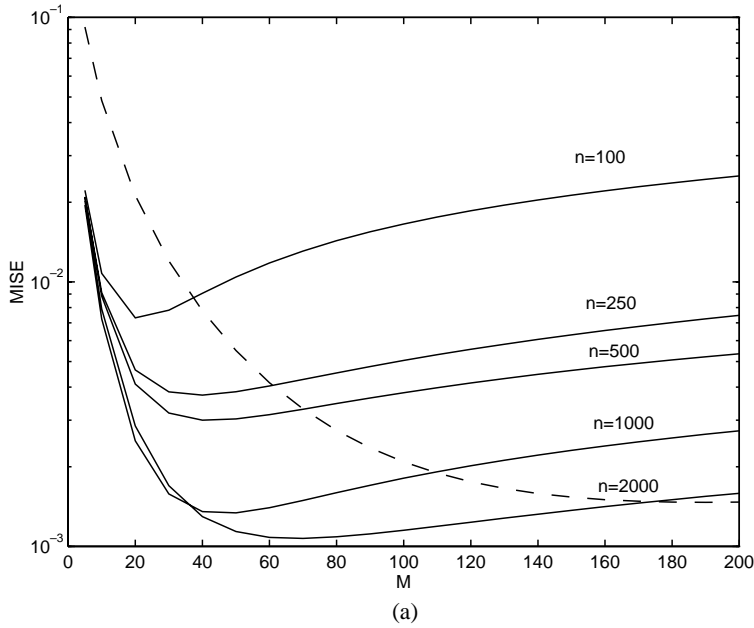


FIG. 3. *MISE versus m and n for the ψ density. (a) One-dimensional (on the circle): $\psi(x) = \exp(US \cos^2(x))/A$; the solid line shows the results for the axial cosine estimator and the dashed line for the directional estimator (with $n = 2000$). (b) Two-dimensional (on the sphere): $\psi(\phi, \theta) = \exp(US \cos^2(\theta))/A$ as defined above; *MISE* for the directional cosine estimator.*

where A is the normalization constant, and the ratio of the distance between two points along the surface of the sphere to h is given as the argument to the kernel function. The use of this kernel for comparisons can be justified by its popular use

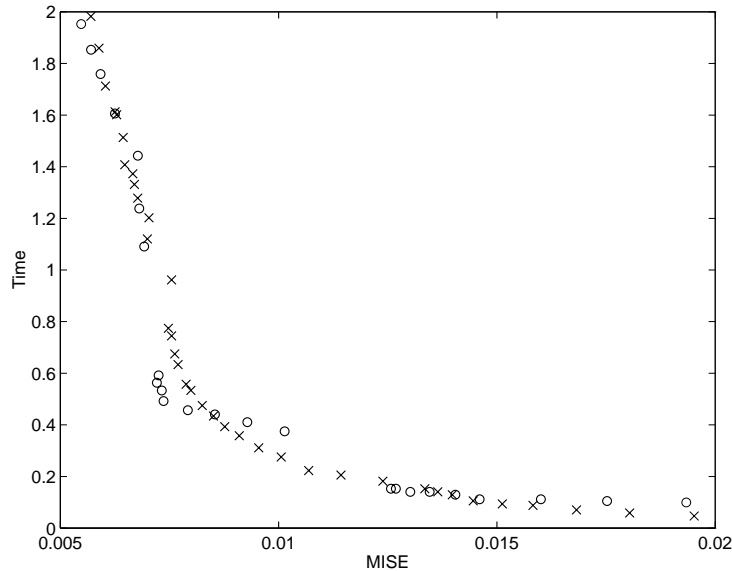


FIG. 4. Comparison of time (in seconds) versus the MISE for the cosine and kernel estimation of axial data sampled from $\psi(\phi, \theta) = \exp(US \cos^2(\theta))/A$ as defined above. The points marked in o represent the kernel estimate. The points marked in x represent the cosine estimate.

in fluid mechanics calculations [12]. Furthermore, we cannot expect any other kernel to give a significantly better performance, for the following reasons: (i) It is well known that most kernels are equally good [23, Table 3.1] with respect to “efficiency,” as defined in [23]. (ii) Given that the efficiency is about the same, the only other consideration is the computational effort involved. Our kernel takes between 6 and 10 floating point operations for a nonzero evaluation (including the cost of computing the square of the distance). Any other reasonable kernel would require at least 6 floating point operations. Apart from this, the memory access times and zero-evaluations would add the same constant to all kernels.

Figures 4 and 5 compare the computational effort required for the cosine weight function estimator and the kernel estimator for certain densities. We obtained the data using the following procedure. We performed estimates for various values of n , m , and h and obtained the MISE and the time for the calculations. For the cosine and the kernel estimates, we separately plotted the data for the time required for the calculations versus the error. We chose the lower envelope of the data as the curve for that particular estimator since the values of m, n for the data on the lower envelope give the best obtainable speed for a given MISE.

In the implementation of the kernel estimator, we divided the sphere into cells such that the sides of the cells had length at least $2h$ (since the kernel defined above has window width $2h$, rather than h). We placed each sample in the appropriate cell. When computing the density for a point in a particular cell, we need to search over points in only a few cells. The expected complexity of this scheme is $O(n^2h^2)$.

We first consider data estimated using the axial cosine estimator and an axial variant of the kernel estimator. Figure 4 shows the results for the two-dimensional ψ distribution. This is an example of a highly nonuniform distribution. We can see that the kernel and the cosine estimators are about equally fast.

We next consider a more uniform distribution given by $f(\phi, \theta) \propto \cos(\theta) + 1/(8\pi)$, where ϕ is the azimuth and θ is the elevation. The results presented in Figure 5(a) show that the cosine estimator outperforms the kernel estimator by more than an order of magnitude when the axial forms of both the estimators are used.

After this we considered the two densities mentioned above but treated the data as directional and estimated them using the directional variants of the kernel and the cosine estimators. For the ψ distribution, the cosine estimator performed very poorly in computational efficiency, and we do not present results for this case. Figure 5(b) presents the results for the distribution given by $f(\phi, \theta) \propto \cos(\theta) + 1/(8\pi)$. We can see that the cosine estimator still outperforms the kernel estimator, though only slightly.

5. Discussion. The comparison of the estimates with the true density indicate that the cosine estimator produces accurate results for the distributions tested. Plots of MISE versus m and n follow the expected trends. As the sample size increases, the error for the optimal m decreases. Besides, the optimal value of m increases as the sample size increases. It can also be seen that as the number of points increases, the range of m over which the estimate performs well also increases. We can use this to our advantage by choosing a suboptimal value of m which decreases the computational effort significantly but increases the error only slightly.

The experiments comparing the computational efficiencies show that the cosine estimator outperforms the kernel estimator for axial data when the distribution is moderately uniform. If the distribution is highly nonuniform, then the two estimators give comparable performance for axial data. The cosine estimator outperforms the kernel estimator slightly for directional data when the distribution is moderately uniform. However, the timing results for the cosine estimator are poor for highly nonuniform directional data. In general, when the data is not very nonuniform, smoother weight functions are used. This gives a low value of m which implies a fast evaluation using the cosine estimator. However, this leads to a higher h for the kernel estimator, which implies that more samples contribute to the kernel evaluation of each sample point and, hence, this leads to more computational effort. Conversely, when the distribution is highly nonuniform, especially for directional data, the kernel method is to be preferred. The empirical test results presented in the appendix further demonstrate this point.

We also analyzed our experimental data to estimate an optimum variation of m with n . Using the results of our experiments, we can perform a least squares analysis and approximate m as $kn^{1/2.5}$ for one-dimensional density estimation which is the same as that expected based on the asymptotic expression for the MISE. Choosing $m = kn^{1/3}$ appears to give reasonable estimates for density on the surface of a sphere. This result is also consistent with the theoretical predictions. Here, the magnitude of k depends on the complexity of the density function. It varies between 1 and 10 for the distributions considered here.

We also noted the values of m , n , and h , which gave the optimal computational effort for a given MISE, and compared the results for the kernel and the cosine estimators. We observed that the values of h were close to the values which gave the minimum MISE for the given sample size. However, the values of m were significantly lower than the values which gave the minimum MISE for the given sample size, though the error involved itself was not much higher than the minimum MISE. Thus, it appears that we can choose a suboptimal smoothing parameter in order to increase the speed in the case of the cosine estimator.

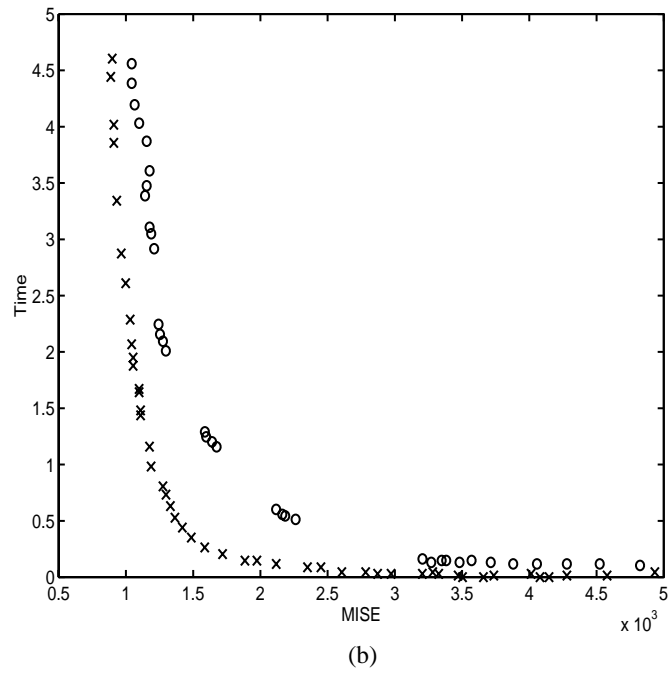
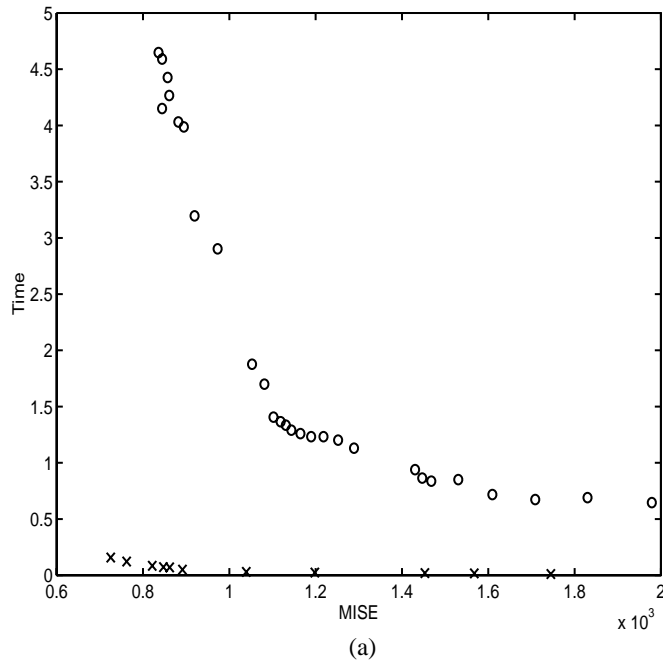


FIG. 5. Plot of time (in seconds) versus the MISE for the cosine and the kernel estimation of data sampled from $f(\phi, \theta) \propto \cos(\theta) + 1/(8\pi)$. The points marked in o represent the kernel estimate. The points marked in x represent the cosine estimate. (a) Data treated as axial. (b) Data treated as directional.

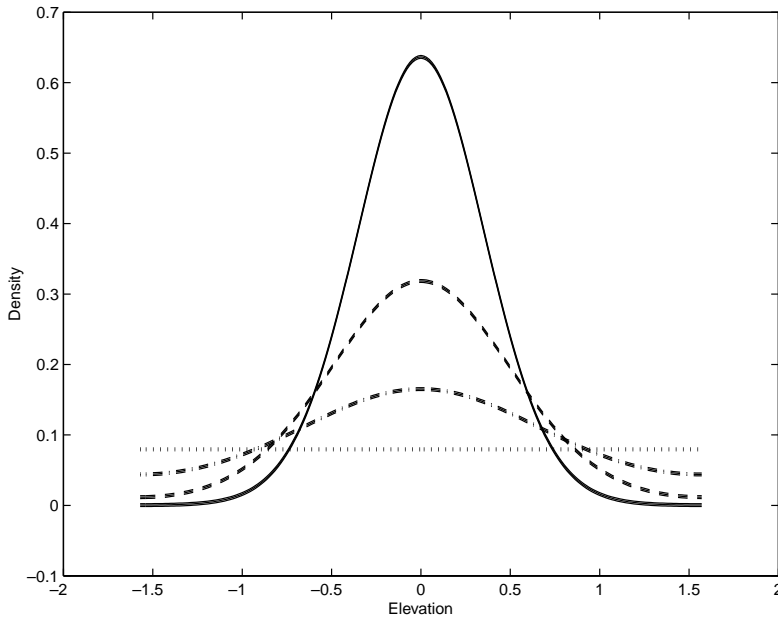


FIG. 6. Plot of density functions $g(\phi; s) = k \cosh(s \sin \phi)$ for different values of s , where ϕ is the elevation. The solid line denotes $s = 8.0$, the dashed line $s = 4.0$, the dash-dotted line $s = 2.0$, and the dotted line $s = 0.0$.

6. Conclusions. In this paper, we have described a weight function estimator for nonparametric estimation of probability density functions based on cosines, and we provided conditions under which the estimate and its derivatives converge to the actual functions. We have developed a scheme for the efficient computation of the density and presented experimental results to check the performance of the estimator for practical problems. These results are particularly relevant to certain fluid mechanics calculations and, in general, to situations where the sample size can be controlled, for example, through refinement of the discretization. We have also given an empirical formula for choosing the weight function exponent parameter of the estimator. Our experimental results suggest that the cosine estimator outperforms the kernel estimator for both directional and axial data that are moderately uniform. It gives performance comparable to the kernel estimator for highly nonuniform axial data, while the kernel method is preferable for highly nonuniform directional data. There is potential for further theoretical study of our estimator.

Appendix. Further test results. We present more test results in this section in order to study the relative efficiencies of the cosine and the kernel techniques, as the density function is varied systematically from being relatively uniform to being sharply peaked on the unit sphere.

For these tests, we chose density functions $g(\phi; s) = k \cosh(s \sin \phi)$, where s is a constant that governs the sharpness of the density function, ϕ is the elevation, and $k = s/(4\pi \sinh s)$ normalizes this to a probability density function. Figure 6 shows the density as a function of the elevation alone for different values of the parameter s . This function is symmetric about the center of the sphere, and thus we can use the axial estimators, as in section 4. We can also ignore our knowledge of this symmetry and use the general directional estimators.

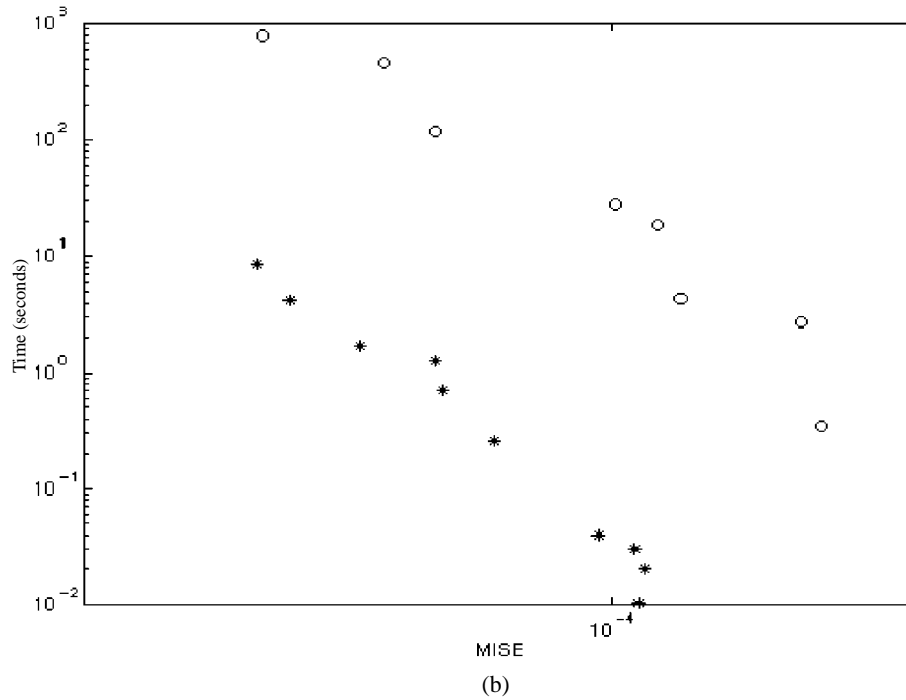
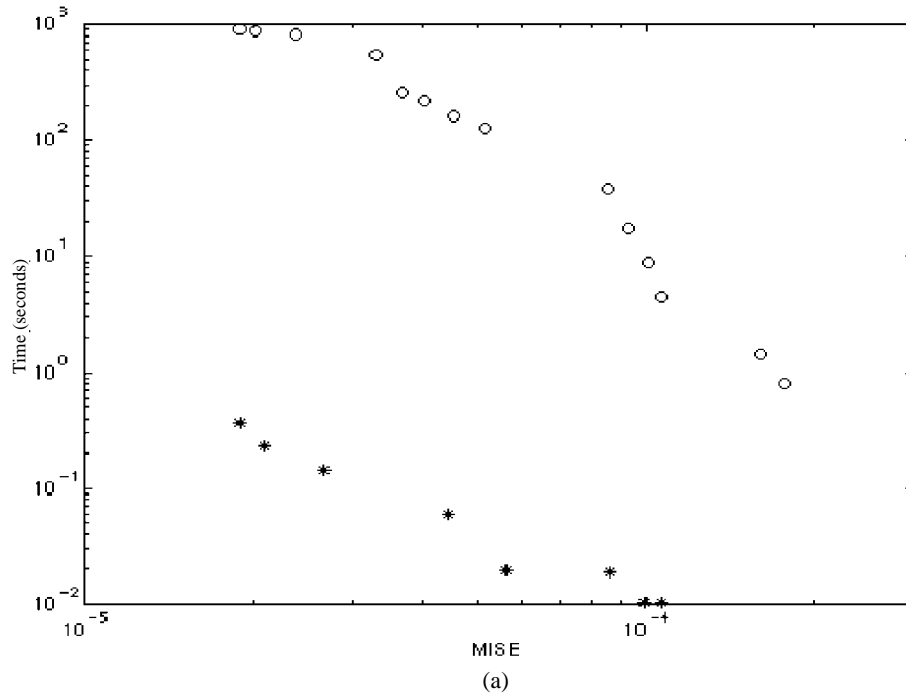


FIG. 7. Plot of MISE versus time for the density function $g(\phi; 0.5) = k \cosh(0.5 \sin \phi)$. The points marked by * are for the cosine estimator, and the points marked by o are for the kernel estimator. (a) Axial estimator. (b) Directional estimator.

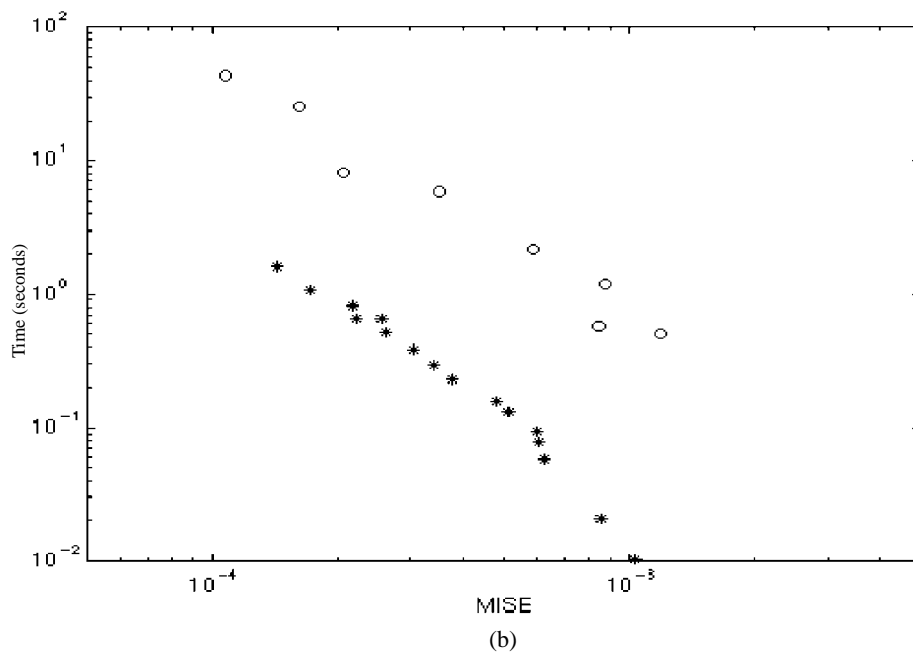
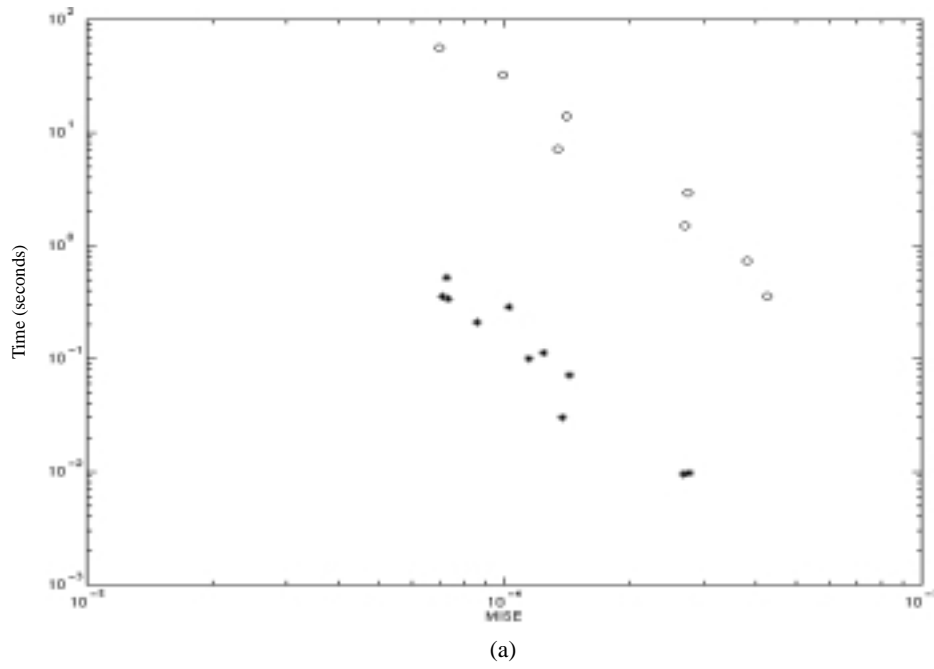


FIG. 8. Plot of MISE versus time for the density function $g(\phi; 1.0) = k \cosh(1.0 \sin \phi)$. The points marked by * are for the cosine estimator, and the points marked by o are for the kernel estimator. (a) Axial estimator. (b) Directional estimator.

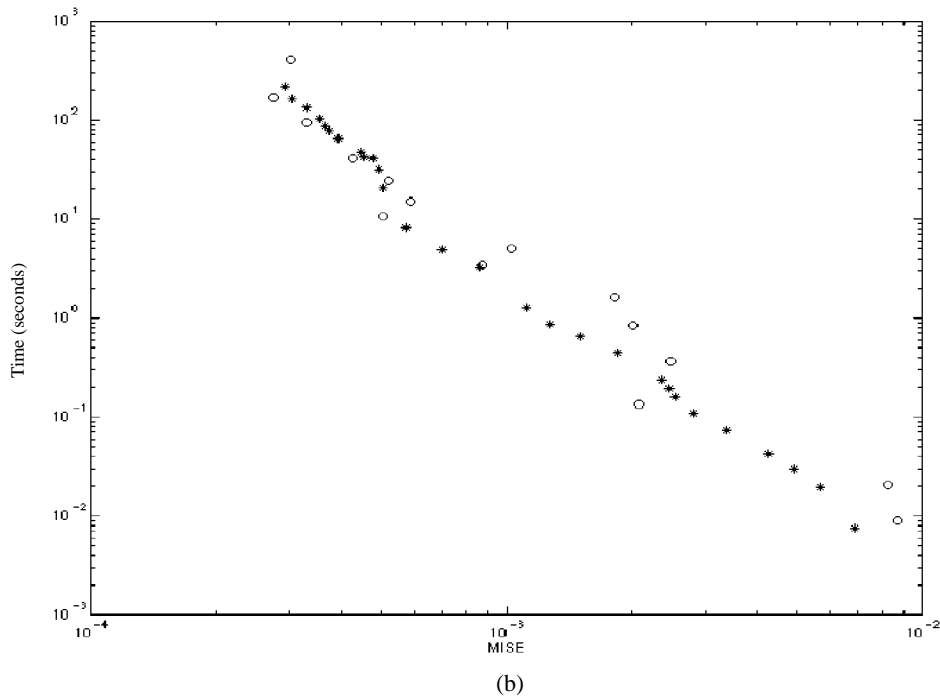
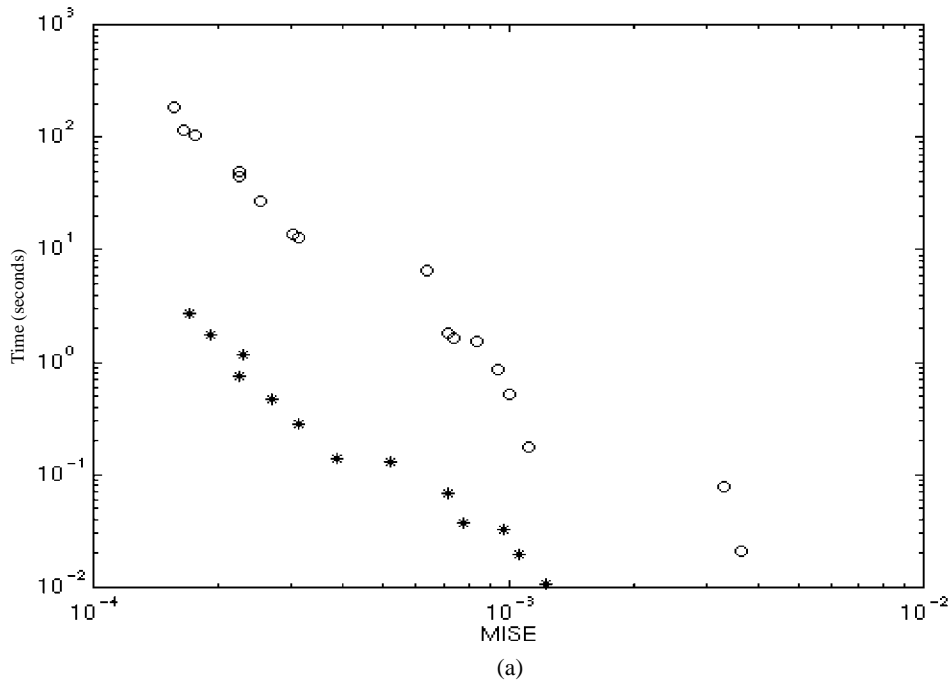


FIG. 9. Plot of MISE versus time for the density function $g(\phi; 2.0) = k \cosh(2.0 \sin \phi)$. The points marked by * are for the cosine estimator, and the points marked by o are for the kernel estimator. (a) Axial estimator. (b) Directional estimator.

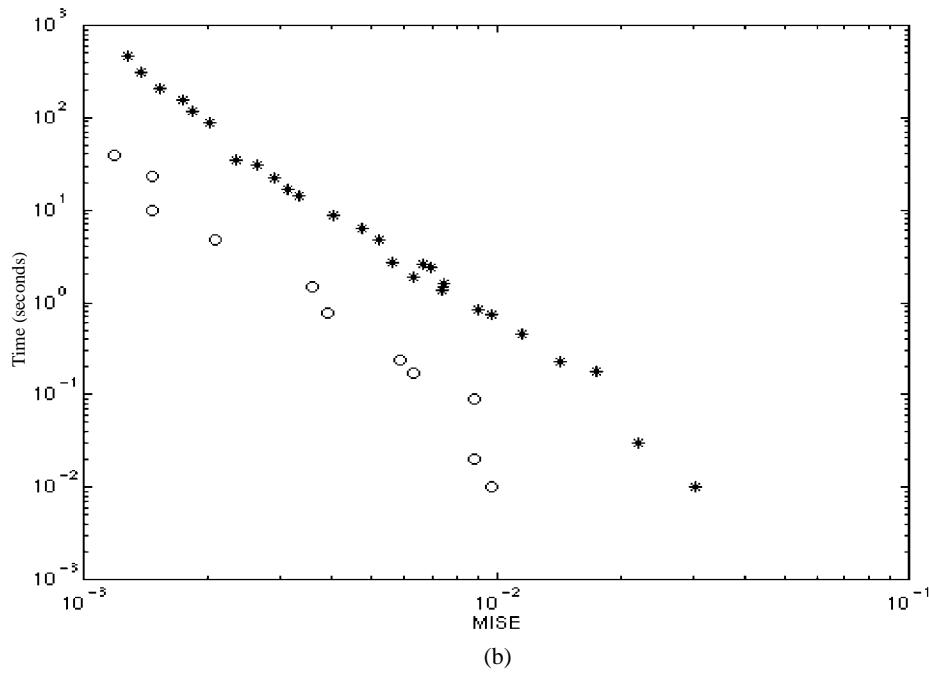
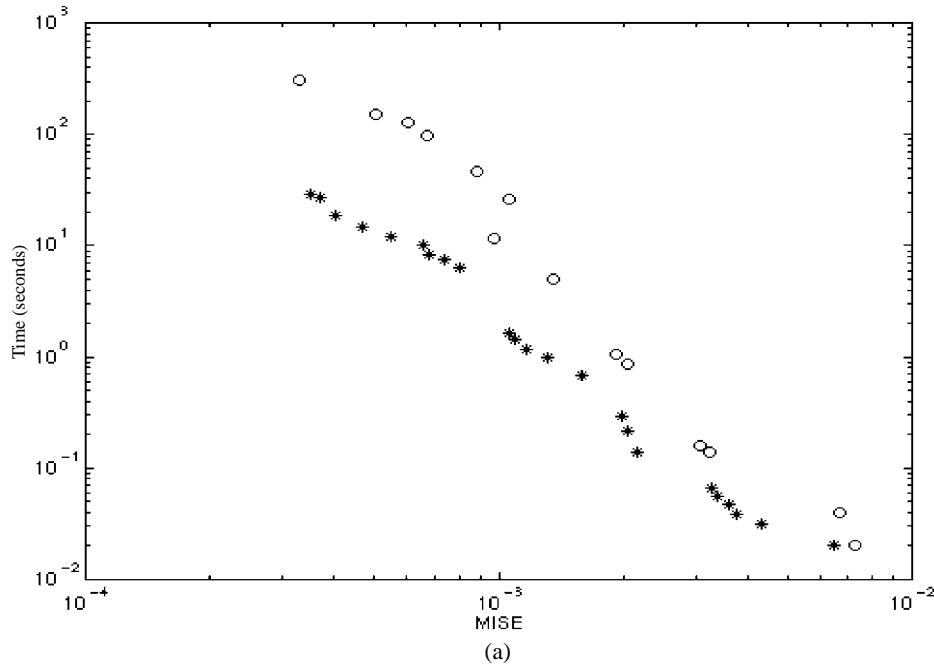


FIG. 10. Plot of MISE versus time for the density function $g(\phi; 4.0) = k \cosh(4.0 \sin \phi)$. The points marked by * are for the cosine estimator, and the points marked by o are for the kernel estimator. (a) Axial estimator. (b) Directional estimator.

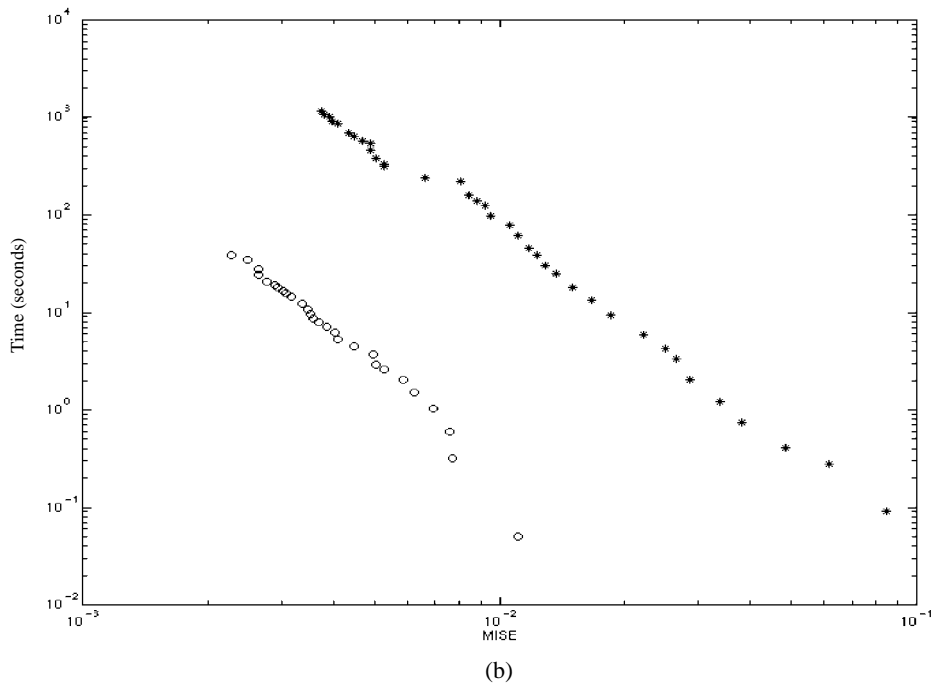
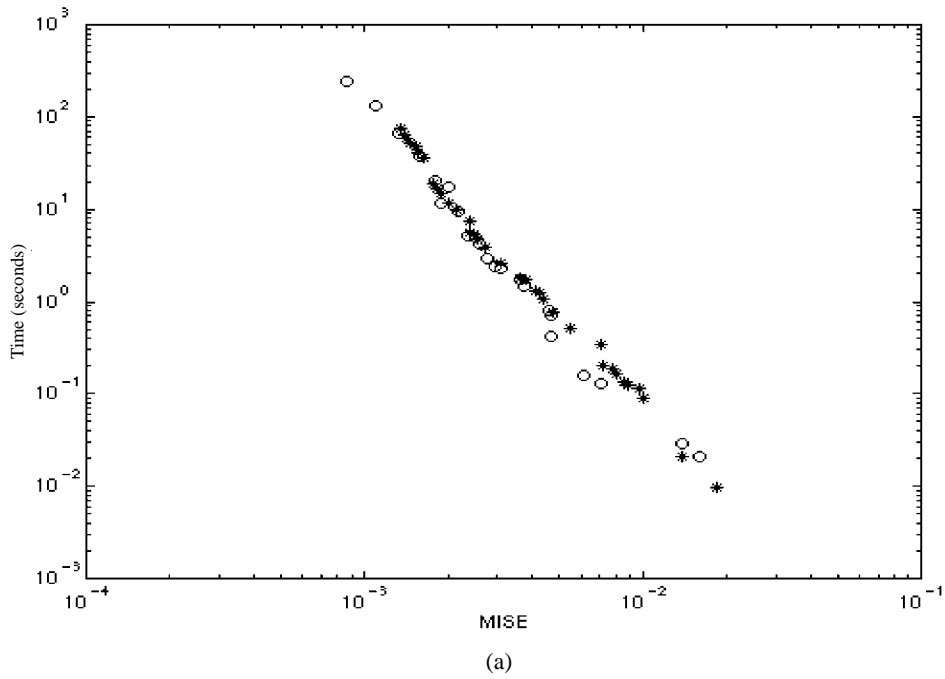


FIG. 11. Plot of MISE versus time for the density function $g(\phi; 8.0) = k \cosh(8.0 \sin \phi)$. The points marked by * are for the cosine estimator, and the points marked by o are for the kernel estimator. (a) Axial estimator. (b) Directional estimator.

We present the results of the experiments as plots of time versus MISE for the kernel estimator versus the cosine estimator, for both axial and directional data. The kernel estimator is the one used in the empirical tests in section 4. The tests were performed on an Intel Celeron 300MHz processor with 64 MB memory. The C code was compiled with the gcc compiler at optimization level $-O3$.

We can see from Figures 7, 8, 9, 10, and 11 that when the density function is not very sharp, the cosine estimator outperforms the kernel estimator for both axial and directional data. As the density becomes sharper, the kernel method starts outperforming the cosine estimator for directional data, though the latter is still better for axial data. When the density becomes extremely sharp, the kernel method becomes better for both types of data, though for axial data the two methods are still comparable to a certain extent in terms of speed. These results follow the theoretically predicted trends and demonstrate that these two methods complement each other for different types of data.

Appendix. We thank the referees for their detailed comments and advice, especially for directing our attention to the current literature.

REFERENCES

- [1] P. J. BICKEL AND M. ROSENBLATT, *On some global measures of the deviations of density function estimates*, Ann. Stat., 1 (1973), pp. 1071–1095.
- [2] C. CHAUBAL, Ö. EĞECIOĞLU, G. LEAL, AND A. SRINIVASAN, *Smoothed particle hydrodynamics techniques for the solution of kinetic theory problems. Part 1: Method*, J. Non-Newtonian Fluid Mech., 70 (1997), pp. 125–154.
- [3] N. N. CHENTSOV, *Estimation of unknown probability density based on observations*, Dokl. Akad. Nauk SSSR, 147 (1962), pp. 45–48 (in Russian).
- [4] Ö. EĞECIOĞLU AND A. SRINIVASAN, *Efficient Nonparametric Estimation of Probability Density Functions*, Technical Report TRCS95-21, University of California at Santa Barbara, 1995.
- [5] J. FAN AND J. S. MARRON, *Fast implementations of nonparametric curve estimators*, J. Comput. Graph. Statist., 3 (1994), pp. 35–56.
- [6] N. I. FISHER, T. LEWIS, AND B. J. J. EMBLETON, *Statistical Analysis of Spherical Data*, Cambridge University Press, Cambridge, 1993.
- [7] I. S. GRADSHTEYN AND I. M. RYZHIK, *Table of Integrals, Series, and Products*, Academic Press, New York, London, 1980, pp. 374.
- [8] P. HALL, G. S. WATSON, AND J. CABRERA, *Kernel density estimation with spherical data*, Biometrika, 74 (1987), pp. 751–762.
- [9] L. HERNQUIST AND N. KATZ, *TREESPH: A unification of SPH with the hierarchical tree method*, Astrophys. J. Suppl., 70 (1989), pp. 419–446.
- [10] J. HWANG, *Non-parametric multivariate density estimation: A comparative study*, IEEE Trans. Signal Process., 42 (1994), pp. 2795–2810.
- [11] R. KRONMAL AND M. TARTER, *The estimation of probability densities and cumulatives by Fourier series methods*, Amer. Statist., (1968), pp. 925–952.
- [12] J. J. MONAGHAN AND J. C. LATTANZIO, *A refined particle method for astrophysical problems*, Astron. Astrophys., 149 (1985), pp. 135–143.
- [13] J. J. MONAGHAN, *Smoothed particle hydrodynamics*, Annu. Rev. Astron. Astrophys., 30 (1992), pp. 543–574.
- [14] E. A. NADARAYA, *Non-parametric Estimation of Probability Densities and Regression Curves*, Mathematics and Applications (Soviet Series)1, Kluwer Academic Publishers, Boston, 1989.
- [15] E. PARZEN, *On estimation of a probability density function and mode*, Ann. Math. Statist., 33 (1962), pp. 1065–1076.
- [16] M. ROSENBLATT, *Remarks on some non-parametric estimates of a density function*, Ann. Math. Statist., 27 (1956), pp. 832–837.
- [17] S. C. SCHWARTZ, *Estimation of probability density by an orthogonal series*, Ann. Math. Statist., 38 (1967), pp. 1261–1265.
- [18] D. W. SCOTT, *Multivariate Density Estimation*, John Wiley, New York, 1992.

- [19] B. SEIFERT, M. BROCKMANN, J. ENGEL, AND T. GASSER, *Fast algorithms for nonparametric curve estimation*, J. Comput. Graph. Statist., 3 (1994), pp. 192–213.
- [20] B. W. SILVERMAN, *Kernel density estimation using the fast Fourier transform*, Appl. Statist., 31 (1982), pp. 93–99.
- [21] N. V. SMIRNOV, *On the approximation of probability densities of random variables*, Scholarly Notes of Moscow State Polytechnical Institute, 16 (1951), pp. 69–96 (in Russian).
- [22] G. D. SMITH, *Numerical Solution of Partial Differential Equations: Finite Difference Methods*, Oxford University Press, London, 1978.
- [23] B. W. SILVERMAN, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, 1986.
- [24] A. SZERI AND L. G. LEAL, *A new computational method for the solution of flow problems of microstructured fluids. Part 2. Inhomogeneous shear flow of a suspension*, J. Fluid Mech., 262 (1994), pp. 171–204.
- [25] R. VIO, G. FASANO, M. LAZZARIN, AND O. LESSI, *Probability density estimation in astronomy*, Astron. Astrophys., 289 (1994), pp. 640–648.
- [26] G. WALTER AND J. BLUM, *Probability density estimation using delta sequences*, Ann. Statist., 7 (1979), pp. 328–340.
- [27] M. P. WAND AND M. C. JONES, *Kernel Smoothing*, Chapman and Hall, London, 1995.
- [28] G. S. WATSON AND M. R. LEADBETTER, *On the estimation of the probability density*, I, Ann. Math. Statist., 34 (1963), pp. 480–491.
- [29] P. WHITTLE, *On the smoothing of probability density functions*, J. Roy. Statist. Soc. Ser. B, 20 (1958), pp. 334–343.