Efficient Non-parametric Estimation of Probability Density Functions

Ömer Eğecioğlu and *Ashok Srinivasan*

Department of Computer Science
University of California Santa Barbara, CA 93106

Abstract

Accurate and fast estimation of probability density functions is crucial for satisfactory computational performance in many scientific problems. When the type of density is known a priori, then the problem becomes statistical estimation of parameters from the observed values. In the non-parametric case, usual estimators make use of kernel functions. If $X_j,\ j = 1, 2, \ldots, n$ is a sequence of i.i.d. random variables with estimated probability density function f_n, in the kernel method the computation of the values $f_n(X_1), f_n(X_2), \ldots, f_n(X_n)$ requires $O(n^2)$ operations, since each kernel needs to be evaluated at every X_j. We propose a sequence of special weight functions for the non-parametric estimation of f which requires almost linear time: if m is a slowly growing function that increases without bound with n, our method requires only $O(m^2 n)$ arithmetic operations. We derive conditions for convergence under a number of metrics, which turn out to be similar to those required for the convergence of kernel based methods. We also discuss experiments on different distributions and compare the efficiency and the accuracy of our computations with kernel based estimators for various values of n and m.

Key words: Probability density, non-parametric estimation, convergence, kernel method, efficient algorithm.

AMS Subject Classifications: 65U05, 62G05, 62G07, 65D15, 65Y20.

1 Introduction

Non-parametric 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 come to from a given family, and the parameters are then estimated by various statistical methods. Early contributors to the theory of non-parametric estimation include N.V. Smirnov [21], M. Rosenblatt [18], E. Parzen [17], and N.N. Chentsov [3]. Extensive descriptions of various approaches to non-parametric estimation along with a comprehensive bibliography can be found in relatively recent books by B.W. Silverman [23], and E.A. Nadaraya [16]. Results of the experimental comparison of some widely used methods appear in [7, 24].

Along with applications in discriminant analysis and cluster analysis, an important application of non-parametric 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 by 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 [15, 6, 13]. 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. Furthermore in some applications, for example in problems involving directional data, the samples lie on the unit circle S^1, or along the surface of the unit sphere S^2.

Various methods have been proposed for non-parametric density estimation in mathematical statistics, such as the kernel [17, 2, 25] and the orthogonal series methods [19, 9]. 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

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

where 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 using a series expansion of the kernel [4], this approach is problematic for narrow window widths.

We propose a cosine based weight function estimator $c_m(x)$ for non-parametric density estimation which is similar to the kernel estimator, but has the ease of evaluation of a series expansion. The role of the window size parameter h of the kernel method is replaced by a parameter m in our method, and f_n is now of the form

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

The definition of c_m allows for the efficient computation of the n values $f_n(X_1), f_n(X_2), \ldots, f_n(X_n)$ using only $O(m^2 n)$ arithmetic operations, 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. Furthermore after the computation of $f_n(X_1), f_n(X_2), \ldots, f_n(X_n)$, our method requires only $O(m^2)$ operations for each subsequent evaluation $f_n(x)$ for an arbitrary x. This is in contrast to the $O(n)$ operations required by the kernel method. The factor m^2 in this operation count is replaced by m^{1+d} when the sample space is the d-dimensional unit sphere S^d.

We derive conditions under which the sequence of estimated density functions f_n constructed in this fashion converges to the unknown density f, and experimentally verify the accuracy and the efficiency of our method in practical test cases. The experiments also indicate how the optimal m should be chosen as a function of n while keeping the error small.

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 1). The conditions guarantee that

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

as $n \to \infty$. These conditions also apply when f is defined on $\mathbb{R}$. In section (3) we give the conditions for the almost sure convergence of the f_n to f in the uniform metric. The conditions guarantee that

$$\sup |f_n(x) - f(x)| \to 0$$

almost surely as $n \to \infty$ (Theorem 2). In section (4) we derive conditions for the almost sure convergence of derivatives of f_n in the uniform metric (Theorem 3). Section (5) presents convergence results in the MISE criterion for S^2 (Theorem 4). In section (6) a scheme for efficient computation of these estimates on $\mathbb{R}$, S^1, and S^2 are presented. In sections (7) and (8), we describe experimental results that give accurate estimates for a number of distributions encountered in practice, for relatively small values of m. Our experiments imply a net savings on the number of operations performed over kernel methods for engineering applications, and also provide a an empirical formula for the optimal value of m. Main conclusions are presented in section (9).

The elements of the probability theory we make use of can be found in Loève's book [14].

2 Convergence of MISE on S^1

Let $C^1[-\pi, \pi]$ denote continuously differentiable functions on $[-\pi, \pi]$. Assume X_j, $j = 1, 2, \ldots, n$ is a sequence of i.i.d. random variables (observations) on $[-\pi, \pi]$ with probability density function $f \in C^1[-\pi, \pi]$. We impose the additional condition that $f(-\pi) = f(\pi)$. In this way f and the random variables X_j can be viewed as being defined on the unit circle S^1 and $C^1[-\pi, \pi]$ can be identified with $C^1(S^1)$. Also, for ease of analysis on $\mathbb{R}$, we extend f periodically to all of the real line by defining $f(x + 2\pi L) = f(x)$, $L \in \mathbb{Z}$.

As an estimator of the density $f(x)$, $x \in [-\pi, \pi]$, we consider a non-parametric estimator of the form

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

for m approaching infinity, to be determined as a function of n, where

$$c_m(x) \;=\; \frac{1}{A_m} \cos^{2m}(\frac{x}{2}) \tag{2}$$

on $[-\pi, \pi]$. The normalization factor A_m is given by

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

so that the integral of $c_m(x)$ is 1 on $[-\pi, \pi]$. 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).

Making use of a table of integrals such as Gradshteyn/Ryzhik [5] and by using Stirling's formula, it can be shown that

$$A_m \;=\; \frac{\pi}{2^{2m-1}} \binom{2m}{m} \;\sim\; \frac{2\sqrt{\pi}}{\sqrt{m}}\,. \tag{3}$$

We wish to find sufficient conditions under which the sequence of estimators f_n converges to f in the Mean Integrated Square Error (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.

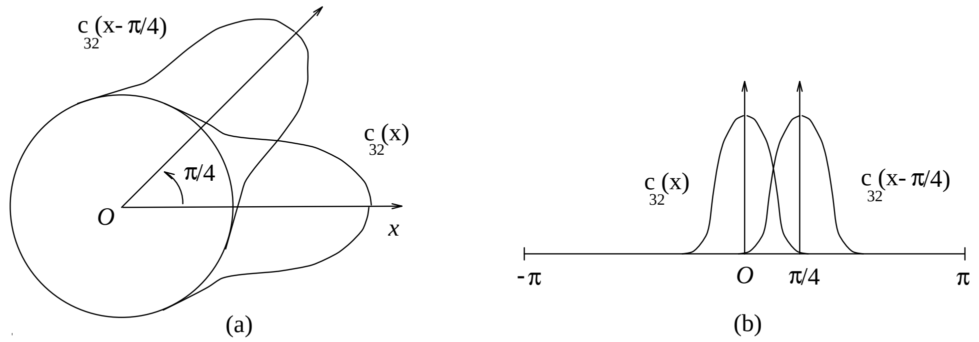

Figure 1: *The functions $c_{32}(x)$ and $c_{32}(x-\frac{\pi}{4})$ on S^1 and on $[-\pi,\pi]$.*

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

Lemma 1 *Suppose $f \in C^1[-\pi,\pi]$ and let $f_n(x)$ be as given in (1). Then $Ef_n(x) \to f(x)$ as $m \to \infty$ uniformly, independently of n.*

Proof

$$Ef_n(x) = \int_{-\pi}^{\pi} c_m(x-s)f(s)ds \tag{4}$$

as shown in Silverman [20], and Whittle [26]. By a change of variable

$$\int_{-\pi}^{\pi} c_m(x-s)f(s)ds = \int_{x-\pi}^{x+\pi} c_m(-y)f(x+y)dy = \int_{x-\pi}^{x+\pi} c_m(y)f(x+y)dy \tag{5}$$

since $c_m(y) = c_m(-y)$. By using the periodicity of c_m and f along with (4), (5), and the mean value theorem,

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

where $\xi_x(y)$ is some point between 0 and y. Therefore

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

From (3) the first integral evaluates to 1. Let $M' = \max_{x\in[-\pi,\pi]}|f'(x)|$. We then have the following bound for the bias

$$|Ef_n(x) - f(x)| \le \int_{-\pi}^{\pi} |c_m(y)f'(\xi_x(y))y|dy \ \le \ M'\int_{-\pi}^{\pi} c_m(y)|y|dy = M'\int_{-\pi}^{\pi} \frac{1}{A_m}\cos^{2m}(y/2)|y|dy \quad .$$

Splitting the integral on the right into two parts, for any $\delta > 0$,

$$\begin{aligned} |Ef_n(x) - f(x)| \ &\le \ M'\int_{|y|<\delta} \frac{1}{A_m}\cos^{2m}(y/2)|y|dy \ + \ M'\int_{|y|\ge\delta} \frac{1}{A_m}\cos^{2m}(y/2)|y|dy \\ &\le \ M'\delta\int_{|y|<\delta} \frac{1}{A_m}\cos^{2m}(y/2)dy \ + \ 2\pi M'\frac{1}{A_m}\cos^{2m}(\delta/2) \end{aligned} \tag{6}$$

The second inequality is obtained by noting that $\cos y$ decreases as $|y|$ increases on the interval under consideration. Furthermore the the integral in (6) is bounded above by 1. Therefore

$$|Ef_n(x) - f(x)| \ \leq \ M'\delta \ + \ 2\pi M' \ \frac{\cos^{2m}(\delta/2)}{A_m} \quad . \tag{7}$$

Let $\eta > 0$ be arbitrarily small. By choosing δ sufficiently small the first summand on the r.h.s. of (7) can be made smaller than $\frac{\eta}{2}$. For a fixed δ one can select m so large that the second term on the r.h.s. of (7) is also smaller than $\frac{\eta}{2}$. This is due to the fact that $\cos^{2m}(\delta/2)$ decreases exponentially, while $1/A_m$ increases only as $\sqrt{m}$ by the asymptotic expression in (3) for A_m.

Thus $|Ef_n(x) - f(x)| \to 0$ as $m \to \infty$. Furthermore, the choice of δ and hence m is independent of x, proving uniform convergence. □

Lemma 2 *Suppose $f \in C^1[-\pi, \pi]$ and let $f_n(x)$ be as given in* (1). *Then $var\ f_n(x) \to 0$ uniformly as $n \to \infty$ provided $m \to \infty$ as $n \to \infty$, and $m = o(n^2)$.*

Proof

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

as shown in Whittle [26]. As a consequence of lemma (1)

$$\int_{-\pi}^{\pi} c_m(x-s)f(s)ds \to f(x)\ ,\ \text{as } m \to \infty$$

uniformly. It follows that

$$\frac{1}{n}\left\{\int_{-\pi}^{\pi} c_m(x-s)f(s)ds\right\}^2 \to 0 \text{ as } n \to 0$$

since f is bounded. To prove the lemma it suffices to show that

$$\frac{1}{n}\int_{-\pi}^{\pi} c_m(x-s)^2 f(s)ds \to 0,\ as\ n \to \infty\ .$$

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 (5), we get

$$\frac{1}{n}\int_{-\pi}^{\pi} c_m(y)^2 f(x+y)dy \ \leq \ \frac{M}{n}\frac{A_{2m}}{A_m^2}\int_{-\pi}^{\pi} c_{2m}(y)dy \ \leq \ \frac{MA_{2m}}{nA_m^2} \ \to \ \frac{Mk_1 m}{n\sqrt{m}} \ = \ Mk_1\frac{\sqrt{m}}{n}\ .$$

Above and in what follows, let k_i, $i \geq 1$ denote a positive constant. The last inequality is on account of the asymptotic expression for A_m in (3). Since $m/n^2 \to 0$ as $n \to \infty$, 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. □

Theorem 1 *Suppose $f \in C^1[-\pi, \pi]$ and $f_n(x)$ is as given in* (1). *If $m \to \infty$ as $n \to \infty$, and $m = o(n^2)$, then*

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

as $n \to \infty$.

Proof

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

as shown in Whittle [26]. From lemma (1) and lemma (2), each of the integrals approaches 0. Hence the MISE converges to 0. □

3 Almost Sure Convergence on $\mathbb{R}$

Let $X_j,\ j = 1, 2, \ldots$ be a sequence of i.i.d. random variables (observations) with values in $\mathbb{R}$. Suppose their common distribution function $F(x)$ is absolutely continuous with density $f(x)$. As an estimator of the density $f(x)$, we again consider a non-parametric estimator of the form

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} c_m(x - X_i) \tag{8}$$

for some m to be determined as a function of n. Here

$$c_m(x) \;=\; \begin{cases} \frac{1}{A_m} \cos^{2m}(\frac{x}{2})\ , & x \in [-\pi, \pi] \\ 0 & otherwise \end{cases} \tag{9}$$

where the normalization factor A_m is explicitly given in (3). Let

$$F_n(x) = \frac{\#\ X_1, X_2, \ldots, X_n \leq x}{n} \tag{10}$$

be the empirical sample distribution function. We will show that $f_n(x)$ approaches to $f(x)$ almost surely (a.s.) under the uniform metric.

Theorem 2 *Suppose $f(x)$ is uniformly continuous on $\mathbb{R}$, and $f_n(x)$ is as given in (8). If $m \to \infty$ as $n \to \infty$, and $m = o(\frac{n}{\log n})$, then*

$$\sup_{x \in \mathbb{R}} |f_n(x) - f(x)| \to 0 \qquad a.s.$$

Proof We proceed along the lines of the proof of Theorem 2.1.1 of Nadaraya [[16], p.42]. The proof is split up into two parts. First we show that

$$\sup_{x \in \mathbb{R}} |f_n(x) - Ef_n(x)| \to 0 \qquad a.s., \tag{11}$$

and then

$$\sup_{x \in \mathbb{R}} |Ef_n(x) - f(x)| \to 0\ . \tag{12}$$

For the proof of (11), we use the integral expression for the expectation from Whittle [26] and integration by parts to get

$$\begin{aligned} \sup_{x \in \mathbb{R}} |f_n(x) - Ef_n(x)| &= \sup_{x \in \mathbb{R}} \left| \int c_m(x-u) dF_n(u) - \int c_m(x-u) dF(u) \right| \\ &\leq \sup_{x \in \mathbb{R}} \int |F_n(u) - F(u)| |dc_m(x-u)| \;\leq\; \sup_{u \in \mathbb{R}} |F_n(u) - F(u)| \cdot k_2 \sqrt{m} \end{aligned}$$

where the last inequality uses the fact that the function $c_m(x)$ is of bounded variation with total variation $2\,c_m(0)$, and the expression for $c_m(0)$ from (3). By using Smirnov's general formula, it can be shown that [[16], p. 43]

$$P\,[\,\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| > \lambda\,] \;\leq\; k_3 e^{-2n\lambda^2}.$$

Therefore

$$P\,[\,\sup_{x\in\mathbb{R}}|f_n(x)-Ef_n(x)|>\epsilon\,]\ \le\ P\,[\,\sup_{x\in\mathbb{R}}|F_n(x)-F(x)|>\epsilon/k_2\sqrt{m}\,]\ \le\ k_3e^{-\beta\frac{n}{m}} \tag{13}$$

where $\beta=2(\frac{\epsilon}{k_2})^2$. By applying the Borel-Cantelli lemma to the inequality (13), we can assert that $\sup_{x\in\mathbb{R}}|f_n(x)-Ef_n(x)|$ approaches zero a.s. as $n\to\infty$ provided the series

$$\sum_{n=1}^{\infty}e^{-\beta\frac{n}{m}} \tag{14}$$

converges for any $\beta>0$. However, (14) is a special case of a Dirichlet series $\sum_{n=1}^{\infty}a_ne^{-\lambda_nz}$ of type $\lambda_n=n/m$ with $a_n=1$. From Markushevich [[11], p.31] or Nadaraya, [[16], p.44], the series (14) converges if and only if $m\frac{\log n}{n}\to0$ as $n\to0$. This ends the proof of (11).

To complete the proof of the theorem, we need to show that $\sup_{x\in\mathbb{R}}|Ef_n(x)-f(x)|\to0$. Let $\delta>0$, and set $M=\max_{x\in\mathbb{R}}|f(x)|$. Then

$$\begin{aligned}\sup_{x\in\mathbb{R}}|Ef_n(x)-f(x)| &\le \sup_{x\in\mathbb{R}}\int_{|u|\le\delta}|f(x-u)-f(x)|c_m(u)du+\sup_{x\in\mathbb{R}}\int_{|u|>\delta}|f(x-u)-f(x)|c_m(u)du\\ &\le \sup_{x\in\mathbb{R}}\sup_{|u|\le\delta}|f(x-u)-f(x)|+2M\int_{|u|>\delta}c_m(u)du\ .\end{aligned} \tag{15}$$

Let $\eta>0$ be arbitrarily small. By choosing δ sufficiently small the first summand on the r.h.s. of (15) can be made smaller than $\frac{\eta}{2}$ since $f(x)$ is uniformly continuous. For a fixed δ one can select m (and therefore n) so large that the second term on the r.h.s. of (15) is also smaller than $\frac{\eta}{2}$. Therefore

$$\sup_{x\in\mathbb{R}}|Ef_n(x)-f(x)|<\eta\ .$$

Now the theorem follows by (11), (12), and the triangle inequality. □

4 Almost sure convergence of derivatives on $\mathbb{R}$

In this section we prove the almost sure convergence of the derivatives of the estimator $f_n(x)$ defined by (8) to the derivatives of the unknown density function $f(x)$. As before, let $X_j,\ j=1,2,\ldots$ be a sequence of i.i.d. random variables (observations) with values in $\mathbb{R}$. Suppose their common distribution function $F(x)$ is absolutely continuous with density $f(x)$. Assume that $f\in C^r(\mathbb{R})$. We denote the r-th derivative of $f(x)$ by $f^{(r)}(x)$. As an estimator of $f^{(r)}(x)$ we take the r-th derivative $f_n^{(r)}(x)$ of $f_n(x)$. Thus

$$f_n^{(r)}(x)=\frac{1}{n}\sum_{i=1}^{n}c_m^{(r)}(x-X_i) \tag{16}$$

where c_m is as defined in (9) and m is to be determined as a function of n. The results of this section are generalizations of the case $r=0$ in which $f^{(0)}(x)=f(x)$ and $f_n^{(0)}(x)=f_n(x)$. We show that $f_n^{(r)}(x)$ approaches to $f^{(r)}(x)$ a.s. under the uniform metric.

Theorem 3 *Suppose $f(x)\in C^r(\mathbb{R})$, $f^{(r)}(x)$ is uniformly continuous, and $f_n^{(r)}(x)$ is as given in (16). If $m\to\infty$ as $n\to\infty$, and $m^{2r+1}=o(\frac{n}{\log n})$, then*

$$\sup_{x\in\mathbb{R}}|f_n^{(r)}(x)-f^{(r)}(x)|\to0\qquad a.s.$$

Proof Nadaraya [[16], p.42]. The proof is again split up into two parts. First we show that

$$\sup_{x\in\mathbb{R}} |f_n^{(r)}(x) - Ef_n^{(r)}(x)| \to 0 \qquad a.s., \tag{17}$$

and then

$$\sup_{x\in\mathbb{R}} |Ef_n^{(r)}(x) - f^{(r)}(x)| \to 0\ . \tag{18}$$

Let $F_n(x)$ be the empirical sample distribution function defined in (10). For the proof of (17), we use the integral expression for the expectation from Whittle ([26]) and integration by parts to get

$$\begin{aligned}\sup_{x\in\mathbb{R}} |f_n^{(r)}(x) - Ef_n^{(r)}(x)| &= \sup_{x\in\mathbb{R}} |\int c_m^{(r)}(x-u)dF_n(u) - \int c_m^{(r)}(x-u)dF(u)| \\ &\leq \sup_{x\in\mathbb{R}} \int |F_n(u) - F(u)||dc_m^{(r)}(x-u)| \ \leq\ \sup_{u\in\mathbb{R}} |F_n(u) - F(u)| \cdot k_2(r) m^r \sqrt{m}\end{aligned}$$

where the last inequality is obtained by differentiation of c_m, and $k_2(r)$ is a constant that depends only on r. By Smirnov's formula,

$$P\,[\,\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| > \lambda\,] \ \leq\ k_3 e^{-2n\lambda^2}.$$

Therefore

$$P\,[\,\sup_{x\in\mathbb{R}} |f_n^{(r)}(x) - Ef_n^{(r)}(x)| > \epsilon\,] \ \leq\ P\,[\,\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| > \epsilon/(k_2(r) m^r \sqrt{m})\,] \ \leq\ k_3 e^{-\beta \frac{n}{m^{2r+1}}} \tag{19}$$

where $\beta = 2(\frac{\epsilon}{k_2(r)})^2$. Now we can apply the Borel-Cantelli lemma to the inequality (19), and conclude that $\sup_{x\in\mathbb{R}} |f_n^{(r)}(x) - Ef_n^{(r)}(x)|$ approaches zero a.s. as $n \to \infty$ provided

$$\sum_{n=1}^{\infty} e^{-\beta \frac{n}{m^{2r+1}}} \tag{20}$$

converges for any $\beta > 0$. Again, (20) is a Dirichlet series of type $\lambda_n = n/m^{2r+1}$ with $a_n = 1$. It converges if and only if $m^{2r+1}(\frac{\log n}{n}) \to 0$ as $n \to 0$. This ends the proof of (17).

To complete the proof of the theorem, we need to prove (18). The trick we use is a generalization of an identity of [1]: If $\phi, \psi \in C^r(\mathbb{R})$ have uniformly continuous r-th derivatives then it can be shown that

$$\phi^{(r)}(x-u)\psi(u) = \phi(x-u)\psi^{(r)}(u) - \frac{d}{du}\sum_{s=0}^{r-1} \phi^{(s)}(x-u)\psi^{(r-s-1)}(u) \tag{21}$$

where the superscripts refer to derivatives with respect to x. Using (21) with $\phi(x) = c_m(x)$ and $\psi(x) = f(x)$ we obtain

$$Ef_n^{(r)}(x) = \int c_m^{(r)}(x-u) f(u) du = \int c_m(x-u) f^{(r)}(u) du \tag{22}$$

The second equality in (22) is a consequence of the fact that terms within the summation in (21) integrate to zero as c_m has bounded support. Let $\delta > 0$, and set $M = \max_{x\in\mathbb{R}} |f^{(r)}(x)|$. Using (22) we get

$$\begin{aligned}\sup_{x\in\mathbb{R}} |Ef_n^{(r)}(x) - f^{(r)}(x)| &\le \sup_{x\in\mathbb{R}} \int_{|u|\le\delta} |f^{(r)}(x-u) - f^{(r)}(x)| c_m(u)du \\ &\quad + \sup_{x\in\mathbb{R}} \int_{|u|>\delta} |f^{(r)}(x-u) - f^{(r)}(x)| c_m(u)du \quad (23)\\ &\le \sup_{x\in\mathbb{R}} \sup_{|u|\le\delta} |f^{(r)}(x-u) - f^{(r)}(x)| + 2M\int_{|u|>\delta} c_m(u)du \ . \quad (24)\end{aligned}$$

Let $\eta > 0$ be arbitrarily small. By choosing δ sufficiently small the first summand on the r.h.s. of (23) can be made smaller than $\frac{\eta}{2}$ since $f^{(r)}(x)$ is uniformly continuous. For a fixed δ one can select m (and therefore n) so large that the second term on the r.h.s. of (23) is smaller than $\frac{\eta}{2}$. Therefore

$$\sup_{x\in\mathbb{R}} |Ef_n^{(r)}(x) - f^{(r)}(x)| < \eta \ .$$

Now the theorem follows by (17), (18), and the triangle inequality. □

5 Convergence of MISE on S^2

Now we assume that $X_j,\ j = 1, 2, \ldots\ n$ is a sequence of i.i.d. random variables (observations) with values on the unit sphere S^2, which we assume to be centered at the origin in $\mathbb{R}^3$. The position of each X_j can be represented in spherical coordinates by the vector (ϕ_j, θ_j) where $\phi_j \in [0, 2\pi),\ \theta_j \in [0, \pi]$. As shown in Figure 2, ϕ_j is the positive angle between the projection of X_j on the x-y plane and the x-axis, and θ_j is the angle between X_j and the z-axis. Suppose that the probability density function

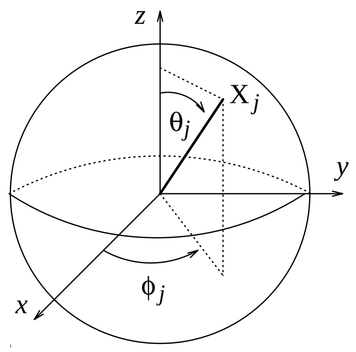

Figure 2: *Spherical coordinates.*

$f(x)$ of the X_j is continuously differentiable. We shall express this condition by writing $f \in C^1(S^2)$. Let $M = \max_{x\in S^2} \|f(x)\|$ and $M' = \max_{x\in S^2} \|\nabla f(x)\|$, where $\nabla f(x) = (\partial f/\partial\phi, \partial f(x)/\partial\theta)$. As an estimator of the density $f(x)$, we consider a non-parametric estimator of the form

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n} c_m(x - X_i) \quad (25)$$

for some m to be determined as a function of n. The c_m are defined in this case as follows. If α_x denotes the angle between x and the point with spherical coordinates $(0, 0)$, then

$$c_m(x) = \frac{1}{A_m}\cos^{2m}(\frac{\alpha_x}{2}) \quad (26)$$

for $x \in S^2$. The normalizing factor A_m is given by

$$A_m = \int_{S^2} \cos^{2m}(\frac{\alpha_x}{2})\, dx \tag{27}$$

so that $c_m(x)$ integrates to 1 over S^2. By a direct computation from (27)

$$A_m \;=\; \frac{4\pi}{m+1} \tag{28}$$

It can be seen that $c_m(x - X_i)$ is independent of the coordinate system (i.e., independent of the point taken as $(0,0)$) since it is only a function of the angle between the points x and X_i ; i.e. the geodesic defined by the great circle through x and X_i. As in section (2), we wish to derive 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. These results are then combined to prove convergence of MISE on S^2, analogous to the proof of Theorem 1 of section (2). First we show that as $m \to \infty$, the expected value of the estimate $f_n(x)$ approaches the actual value $f(x)$ of the density function uniformly for any given n.

Lemma 3 *Suppose $f \in C^1(S^2)$ and let $f_n(x)$ be as given in (25) with $x \in S^2$. Then $Ef_n(x) \to f(x)$ uniformly as $m \to \infty$, independently of n.*

Proof

$$E\, f_n(x) = E\, \frac{1}{n} \sum_{i=1}^{n} c_m(x - X_i) = \int_{S^2} c_m(x-s) f(s) ds \tag{29}$$

where ds is the element of area on the sphere. To simplify the analysis we introduce new coordinate system to represent the points. Let the new coordinate of a point be obtained by translation by $-x$. Thus, In the new coordinate system x becomes $(0,0)$. Let f_n^x represent the function f_n in the new coordinates, and let f^x be the corresponding representation of f. Note that $c_m(x - s)$ is translation invariant. In the new coordinate system (29) can be written as

$$Ef_n^x(0) = \int_{S^2} c_m(y) f^x(y) dy = \int_{S^2} c_m(y)(f^x(0) + \nabla f^x(\xi_x(y)) \cdot y dy$$

by the mean value theorem, where $\xi_x(y)$ is some point along the geodesic between $(0,0)$ and y in the new coordinate system. The existence of such a point $\xi_x(y)$ on S^2 is guaranteed because of the locally Euclidean nature of S^2. Therefore

$$Ef^x(0) = f^x(0) \int_{S^2} c_m(y) dy + \int_{S^2} c_m(y) \nabla f^x(\xi_y(y)) \cdot y dy\ .$$

By a derivation similar to that in lemma (1) one can show that this implies

$$\|Ef_n^x(0) - f^x(0)\| \;\leq\; M\delta + 4\pi^2 M' \,\frac{\cos^{2m}(\delta/2)}{A_m} \tag{30}$$

in the new coordinates. Let $\eta > 0$ be arbitrarily small. By choosing δ sufficiently small the first summand on the r.h.s. of (30) can be made smaller than $\frac{\eta}{2}$. For a fixed δ one can select m so large that the second term on the r.h.s. of (30) is also smaller than $\frac{\eta}{2}$. This is due to the fact that $\cos^{2m}(\delta/2)$ decreases exponentially, while $1/A_m$ increases only as m in view of (28).

We have shown that $\|Ef^x(0) - f^x(0)\| \to 0$ as $m \to \infty$. Consequently in the original coordinates $\|Ef(x) - f(x)\| \to 0$. Since the choice of δ and hence of m is independent of x, the convergence is uniform, and the lemma is proved. □

Lemma 4 *Suppose $f \in C^1(S^2)$ and let $f_n(x)$ be as given in* (25). *Then $var\ f_n(x) \to 0$ uniformly as $n \to \infty$ provided $m \to \infty$ as $n \to \infty$, and $m = o(n)$.*

Proof

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

as shown in Whittle ([26]). We have already shown in lemma (3) that

$$\int_S c_m(x-s)f(s)ds \to f(x)\ , \quad \text{as } m \to \infty$$

uniformly. Thus

$$\frac{1}{n}\left\{\int_S c_m(x-s)f(s)ds\right\}^2 \to 0 \text{ as } n \to 0$$

since f is bounded. To prove the lemma it suffices to show that

$$\frac{1}{n}\int_S c_m(x-s)^2 f(s)ds \to 0,\ as\ n \to \infty\ .$$

Using a change of coordinates as in lemma (3) and by a derivation similar to that of lemma (2) it can be shown that

$$\frac{1}{n}\int_S c_m(x-s)^2 f(s)ds \ \leq\ \frac{MA_{2m}}{nA_m^2} \ \to\ \frac{Mk_5m^2}{nm} \ =\ Mk_5\frac{m}{n}$$

By assumption $m/n \to 0$ as $n \to \infty$, and therefore the above integral converges to 0. Since the values of m and n are independent of x, the convergence is uniform. □

Theorem 4 *Suppose $f \in C^1(S^2)$ and let $f_n(x)$ be as given in* (1). *If $m \to \infty$ as $n \to \infty$ and $m = o(n)$, then*

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

Proof

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

as shown in Whittle ([26]). From lemma (3) and lemma (4), each of the integrals on the right approaches 0. Hence the MISE converges to 0. □

6 Efficient computation of density

In this section, we shall describe an efficient algorithm for the computation of the values of the density estimator $f_n(x)$ at a set of n observed points $X_1, X_2, \ldots, X_n$ on the unit circle S^1 (1-D case) and on the unit sphere S^2 (2-D case). The idea of the 1-D case easily extends to $\mathbb{R}$. We also show that if the value of f_n at some arbitrary x is desired, then this is also easily accomplished once $f_n(X_1), f_n(X_2), \ldots, f_n(X_n)$ 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 (2) or (26). Suppose we represent the positions of the observed points $X_1, X_2, \ldots, X_n$ 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 S^1. From (1) and (25) we have

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n} c_m(x - X_i) \quad \text{with} \quad c_m(x) \;=\; \frac{1}{A_m}\cos^{2m}(\frac{x}{2}) \tag{31}$$

and A_m is as given by (3). By using the half angle formula for cosine,

$$f_n(x) = \frac{1}{nA_m}\sum_{i=1}^{n}\left(\frac{1+\cos(x - X_i)}{2}\right)^m \tag{32}$$

Let $x = (x_1, x_2)$ and $X_i = (X_{i1}, X_{i2})$ be the Cartesian coordinates of x and X_i, $i = 1, 2, \ldots, n$ on S^1 and let $<,>$ represent the standard inner product on $\mathbb{R}^2$. Then $\cos(x - X_i) =< x, X_i >$. Substituting this into (32) we get

$$f_n(x) = \frac{1}{n2^m A_m}\sum_{i=1}^{n}(1+ < x, X_i >)^m = \frac{1}{n2^m A_m}\sum_{i=1}^{n}(1 + x_1 X_{i1} + x_2 X_{i2})^m \tag{33}$$

The expression (33) 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 (33)

$$f_n(x) = \frac{1}{n2^m A_m}\sum_{i=1}^{n}\sum_{r+s\le m}\binom{m}{r, s} x_1^r x_2^s\, X_{i1}^r X_{i2}^s \tag{34}$$

where the inner summation is over all $r + s \le m$ with $r, s \ge 0$, and

$$\binom{m}{r, s} \;=\; \frac{m!}{r!s!(m - r - s)!}\;.$$

Changing the order of summation

$$f_n(x) = \frac{1}{2^m A_m}\sum_{r+s\le m}\binom{m}{r, s} M(r, s)\, x_1^r x_2^s \tag{35}$$

where

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

and A_m is as given by (3). If we use the asymptotic expression for A_m in (3) for computational ease, then (35) simplifies to

$$f_n(x) = \frac{\sqrt{m}}{2^{m+1}\sqrt{\pi}}\sum_{r+s\le m}\binom{m}{r, s} M(r, s)\, x_1^r x_2^s\;. \tag{36}$$

for large m. Now we consider the number of operations required for the evaluation of $f_n(x)$ given the observations $X_1, X_2, \ldots, X_n$. The powers X_{i1}^r and X_{i2}^r for a fixed i and $r = 1, 2, \ldots, m$ can be computed with $O(m)$ multiplications. Doing this for $i = 1, 2, \ldots, n$ 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 (35) or (36) can be computed with a total of $O(m^2 n)$ arithmetic operations.

Once the coefficients of $f_n(x)$ have been computed, to evaluate $f_n(x)$ with $x = (x_1, x_2)$ we calculate the powers x_1^r and x_2^r for $r = 1, 2, \ldots, m$ in $O(m)$ operations. Since the coefficients are already available, the remaining computation in (35) requires only an additional $O(m^2)$ multiplications and additions. Thus we have proved

Theorem 5 *Suppose X_j, $j = 1, 2, \ldots, n$ is a sequence of i.i.d. observations with values on the unit circle S^1 and $f_n(x)$ is the non-parametric estimator to the density as defined in* (1). *Then the values*

$$f_n(X_1), f_n(X_2), \ldots, f_n(X_n) \tag{37}$$

of the approximate density can be evaluated in $O(m^2 n)$ arithmetic operations. After this preprocessing step, $f_n(x)$ can be computed for any $x \in S^1$ with only $O(m^2)$ additional operations.

Remark

Note that kernel based non-parametric estimators to the density require $O(n^2)$ arithmetic operations for the computation of the values in (37), since each $K(\frac{X_j - X_i}{h})$ needs to be evaluated for $1 \leq i, j \leq n$.. For our MISE convergence condition for S^1 (Theorem 1) to be satisfied, m must increase without bound with n. Theoretically, we can take m to be as slowly increasing as we like. Theorem (5) then implies that the computation of all of the n quantities in (37) can be accomplished with the estimator in (31) using only about $O(n)$ arithmetic operations, if the magnitude of m gives acceptable accuracy for $f_n(x)$. In fact in this case $f_n(x)$ can be evaluated at any n points on S^1 in $O(n)$ time since after the first $O(m^2 n)$ arithmetic operations required for preprocessing, each evaluation of $f_n(x)$ requires only $O(m^2)$ effort. The problem with taking m to be too slowly growing is that the magnitude of m controls the error terms in our convergence proofs. This is a crucial point which makes experimental results valuable in determining good values of m in practice.

The case of S^2 is similar. When $X_1, X_2, \ldots, X_n$ are observations on S^2 drawn from an unknown density, from (1) and (25) we have

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} c_m(x - X_i) \quad \text{with} \quad c_m(x) \;=\; \frac{1}{A_m} \cos^{2m}(\frac{\alpha_x}{2}) \;, \tag{38}$$

where A_m is as given by (28), and α_x denotes the angle between x and the point with spherical coordinates $(0, 0)$ on S^2. Switching to Cartesian coordinates, write $x = (x_1, x_2, x_3)$ and $X_i = (X_{i1}, X_{i2}, X_{i3})$ for $i = 1, 2, \ldots, n$. Proceeding as in the derivation of the expression (35) for $f_n(x)$, we obtain

$$f_n(x) = \frac{1}{n 2^m A_m} \sum_{i=1}^{n} (1 + < x, X_i >)^m = \frac{1}{n 2^m A_m} \sum_{i=1}^{n} (1 + x_1 X_{i1} + x_2 X_{i2} + x_3 X_{i3})^m \tag{39}$$

Using the multinomial theorem and the asymptotic expression (28) for the normalizing factor A_m, we obtain

$$f_n(x) = \frac{m+1}{\pi 2^{m+2}} \sum_{r+s+t \leq m} \binom{m}{r, s, t} M(r, s, t)\, x_1^r x_2^s x_3^t \tag{40}$$

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

$$M(r,s,t) = \frac{1}{n}\sum_{i=1}^{n} X_{i1}^r X_{i2}^s X_{i3}^t \ . \tag{41}$$

This time that the coefficients of the polynomial in (40) can be computed with a total of $O(m^3 n)$ arithmetic operations. After this preprocessing, each evaluation of $f_n(x)$ for $x \in S^2$ requires only an additional $O(m^3)$ operations. Therefore

Theorem 6 *Suppose X_j, $j = 1,2,\ldots,n$ is a sequence of i.i.d. observations with values on the unit sphere S^2 and $f_n(x)$ is the non-parametric estimator to the density as defined in* (25)*. Then the values*

$$f_n(X_1), f_n(X_2), \ldots, f_n(X_n) \tag{42}$$

of the approximate density can be evaluated in $O(m^3 n)$ arithmetic operations. After this preprocessing step, $f_n(x)$ can be computed for any $x \in S^2$ with only $O(m^3)$ additional operations.

Remark
As in the case of S^1, theorem (5) implies that the computation of all of the n values in (42) can be accomplished in almost linear time in n, since theorem (4) guarantees MISE convergence of the estimator (38) even for slowly growing m. If the values of m can be taken to be sufficiently small to get the accuracy we are interested in, then (38) estimator requires much fewer than $O(n^2)$ operations required by kernel estimators.

If we require estimation of derivatives in addition to that of the density, the results of section 4 give an upper bound on the acceptable increase in m with n in the 1-D case as $m^{2r+1} = o(\frac{n}{\log n})$. Since the computational complexity is $O(m^2 n)$, our estimator will asymptotically require less computational effort than a kernel estimator.

In the next section, we perform computations to test the range of m necessary to get accurate results for some distributions and compare the results with the kernel method.

7 Experimental results

Numerical experiments were carried out to determine the error as a function of m and n using various distributions. Comparisons were also made with the following kernel estimator [13]:

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

where A is the normalization constant. Since we were concerned only with the accuracy of the estimates in the experiments reported in this section, we did not use the scheme of section 6, but rather used a quadratic scheme to perform the estimation while using the cosine weight function.

The results for some of the experiments are presented below. The MISE was determined by numerical integration using 250 points in the case of S^1, and with 500 points in the case of S^2. In the determination of the MISE, trials were performed with 7 samples each time, and the mean is reported.

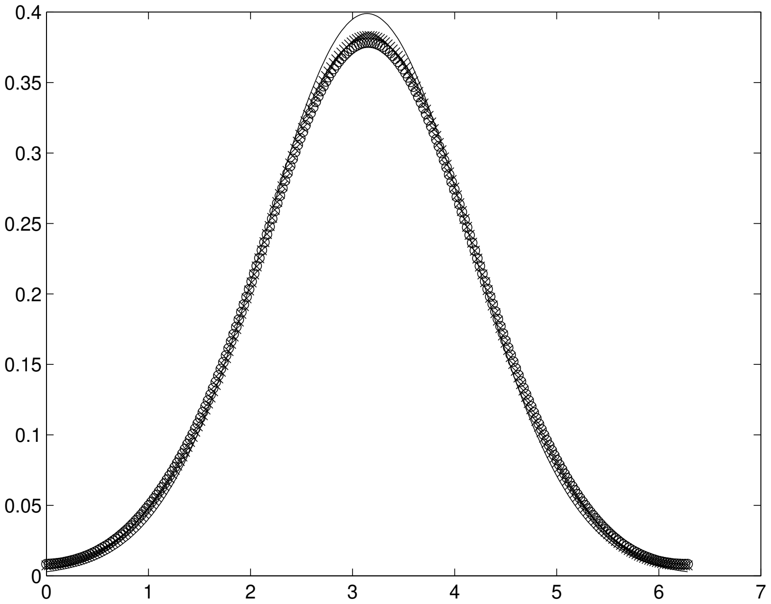

Figure 3: *Comparison of the cosine and kernel estimators for a normal density with mean* π *and standard deviation* 1*. The solid line represents the normal density. The points marked by* x *represent the kernel estimate with* $h = 0.55$*. The points marked by* o *represent the cosine estimate with* $m = 15$*. The estimates are based on the mean of* 50 *trials with samples of size* 250 *each.*

Figure (3) compares the cosine and the kernel estimates for a normal distribution with mean π and standard deviation 1, generated by using the function *randn* of *MATLAB*. The results given are the mean of 50 trials. Note that the two curves are virtually identical.

In Figure (4) we compare the MISE versus m and n for the normal distribution. In Figure (5) the same comparison is made for the derivative of the density function.

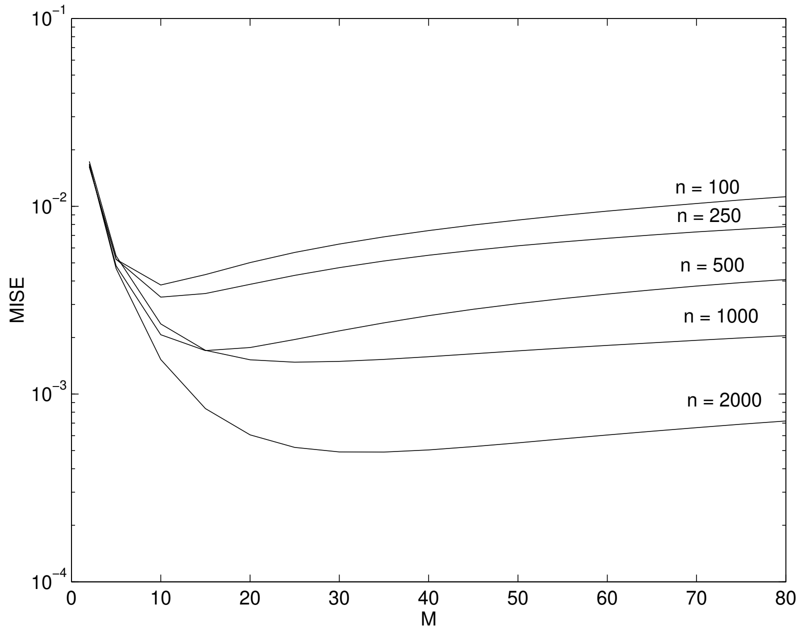

Figure 4: *MISE versus m and n for the normal density with mean π and standard deviation* 1.

As with most estimators, the estimation of bimodal distributions is more difficult. We give an example of a sum of β distributions in Figure (6). The distribution was taken to be

$$0.35\,\beta(6,2) + 0.65\,\beta(2,6)$$

and generated by the rejection-acceptance method [12]. The β function was scaled to take values in $[0, 2\pi]$. Compared with a unimodal distribution, more samples are required to get reasonable results.

We can modify our estimator if we know the density function to be symmetric by considering the estimator $c_m(x) = \cos^{2m}(x)/A_m$, where A_m is a normalization factor. We call this the *symmetric form* of the estimator. This results in a $O(m^d n)$ algorithm instead of an $O(m^{d+1} n)$ algorithm on the d-dimensional unit sphere S^d.

As an example, consider the function $\psi(\phi, \theta) = \exp{(US\theta^2)}/A$ 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. Since this function is symmetric with respect to the origin, we can consider the symmetric version of our estimator. This enables us to use lower values of m and to get improved accuracy for the same sample size. In Figure (7) we compare the cosine and kernel estimates for the 1-D version of the above density function where ψ was taken to be $\psi(x) = \exp{(USx^2)}/A$ using the general form of our

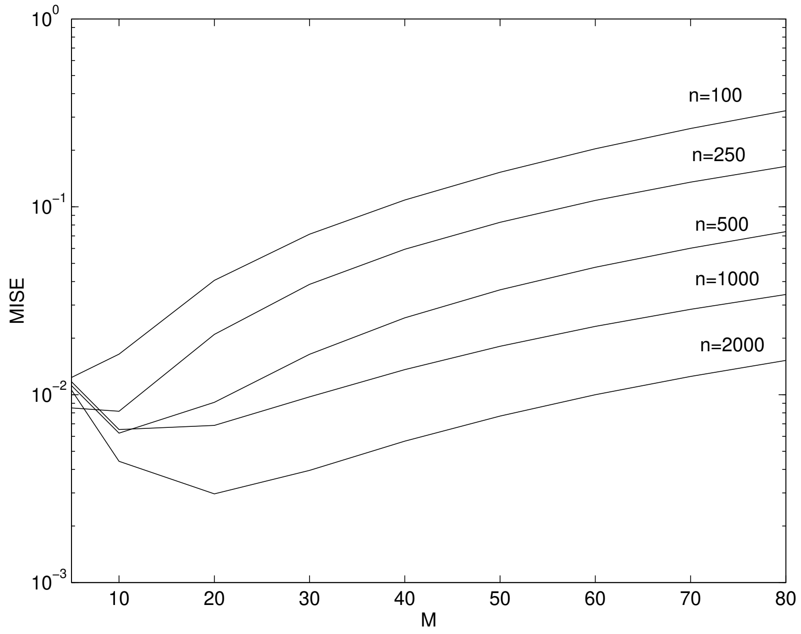

Figure 5: *MISE versus m and n for the derivative of the normal density with mean π and standard deviation* 1.

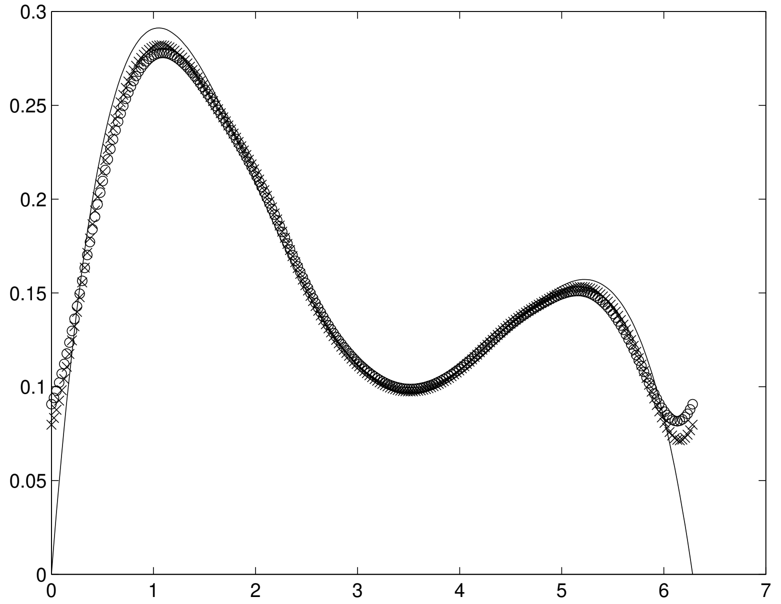

Figure 6: *Comparison of cosine and kernel estimators for a sum of β distributions on $[0, 2\pi]$. The solid line represents the true density. The points marked by* × *represent the kernel estimate with $h = 0.4$. The points marked by* o *represent the cosine estimate with $m = 25$. The estimates are based on the mean of* 50 *trials with samples of size* 500 *each.*

estimator. In Figure (8) we compare the estimates obtained using the symmetric form of the cosine estimator with the true density. The results for the derivative of the density using the symmetric form of the estimator are shown in Figure (9).

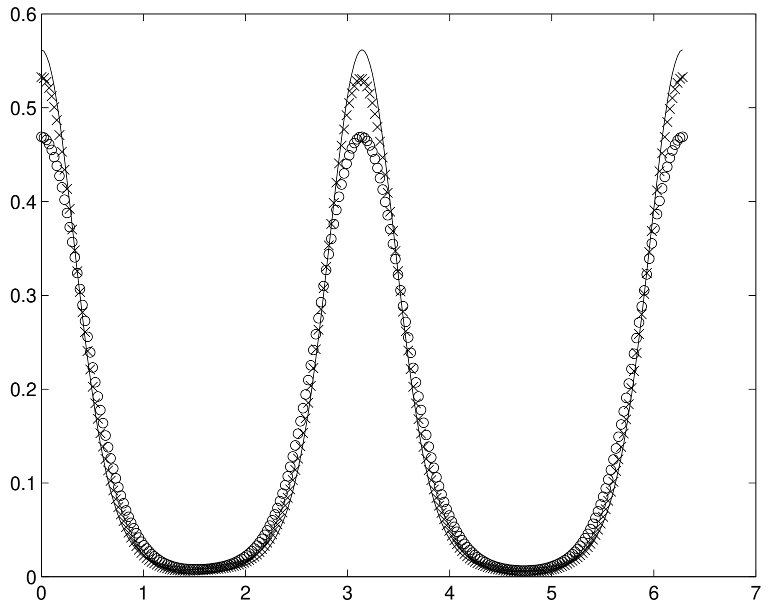

Figure 7: *Comparison of cosine and kernel estimates for $\psi(x)$ in 1-D. The solid line represents the true density. The points marked by* x *represent the kernel estimate with $h = 0.2$. The points marked by* o *represent the cosine estimate with $m = 40$. The estimates are the mean of 50 trials with samples of size 500 each.*

In Figure (10) the MISE is compared versus m and n for the 1-D ψ distribution. We also compare with one case of the general form of the estimator in order to show the benefit of using the symmetric form of the estimator. We perform the same calculations for the estimation of the derivative in Figure (11).

In Figure (12) the MISE is compared versus m and n for the 2-D version of the ψ distribution on the surface of a sphere using the symmetric form of the estimator.

8 Discussion

The comparison of the estimates with the true density indicate that this method produces good results for the distributions tested.

Plots of MISE versus m and n follow the expected trends. As the sample size increases, the error at the optimal value of m decreases. Besides, the optimal value of m increases as the sample size increases. The optimal value of m is smaller for the derivative than for the density estimation. It can also be seen that as the number of points increases, the range of m over which the estimate performs well also increases.

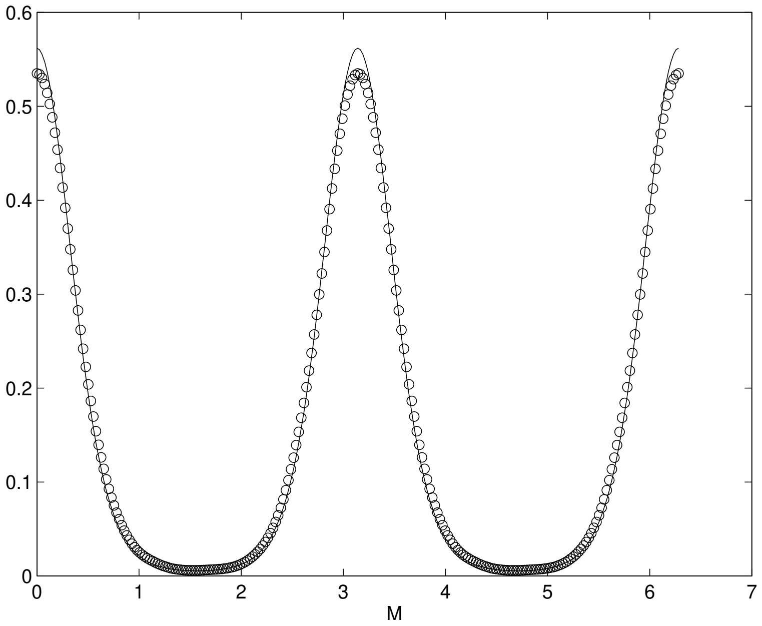

Figure 8: *Symmetric cosine estimate for $\psi(x)$ in 1-D. The solid line represents the true density. The points marked by* o *represent the cosine estimate with $m = 40$. The estimates are the mean of 50 trials with samples of size 500 each.*

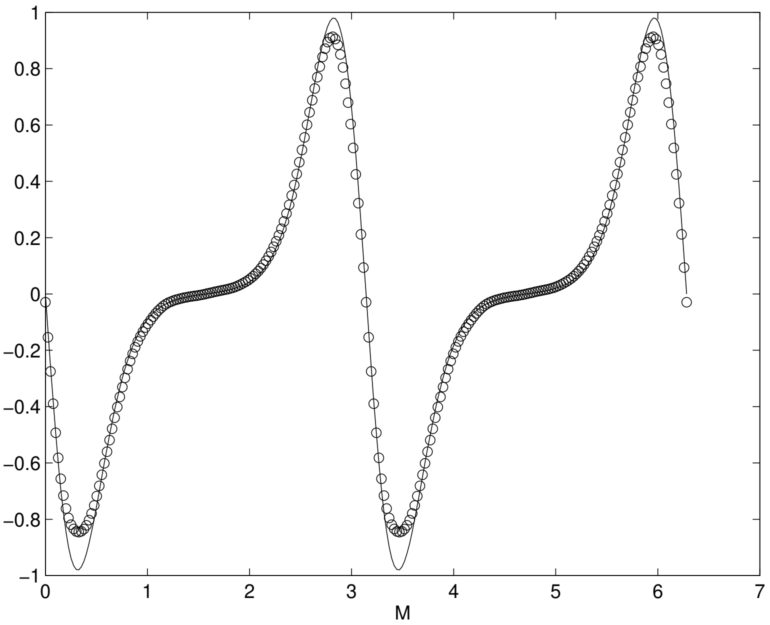

Figure 9: *Symmetric cosine estimate for the derivative of $\psi(x)$ in 1-D. The solid line represents the true derivative. The points marked by* o *represent the cosine estimate with $m = 40$. The estimates are the mean of 50 trials with samples of size 500 each.*

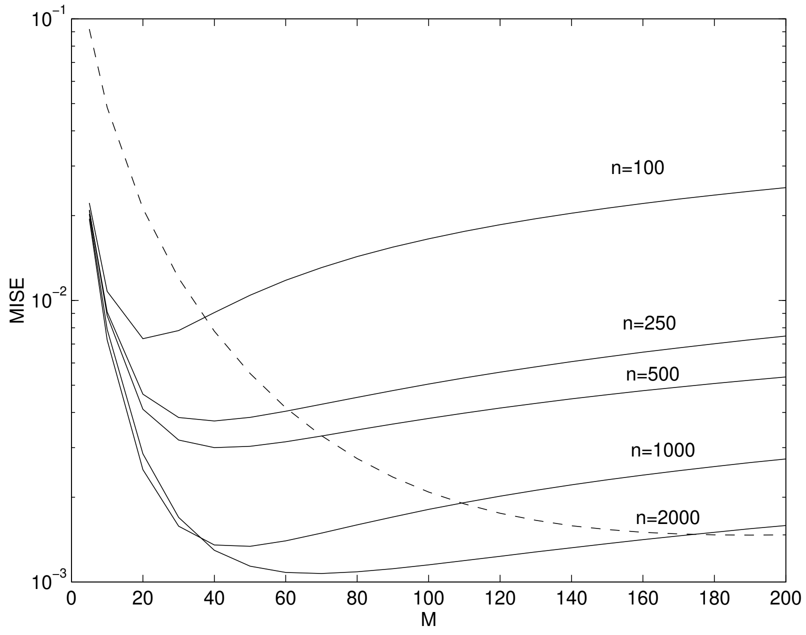

Figure 10: *MISE versus m and n for the 1-D ψ density. The solid line shows the results for the symmetric form of the estimator. The dashed line shows the results for the general form of the estimator using $n = 2000$.*

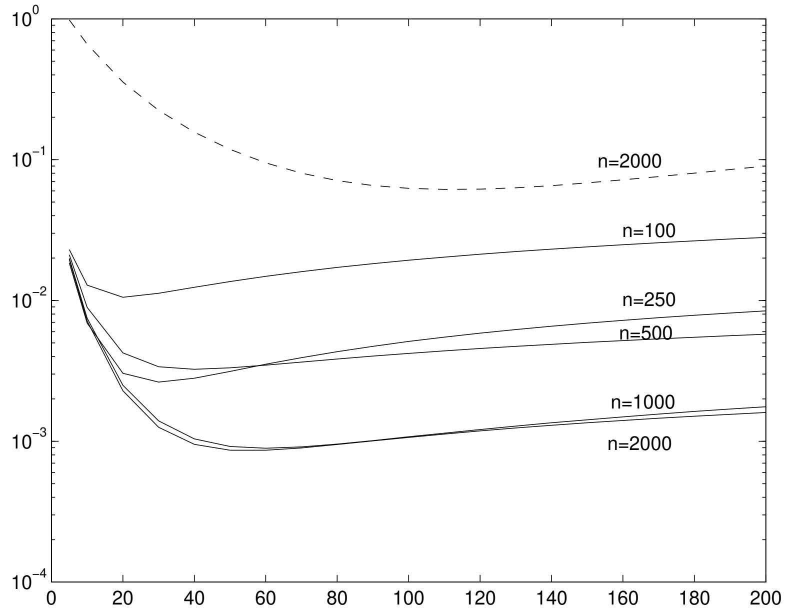

Figure 11: *MISE versus m and n for the derivative of 1-D ψ density. The solid line shows the results for the symmetric form of the estimator. The dashed line shows the results for the general form of the estimator using $n = 2000$.*

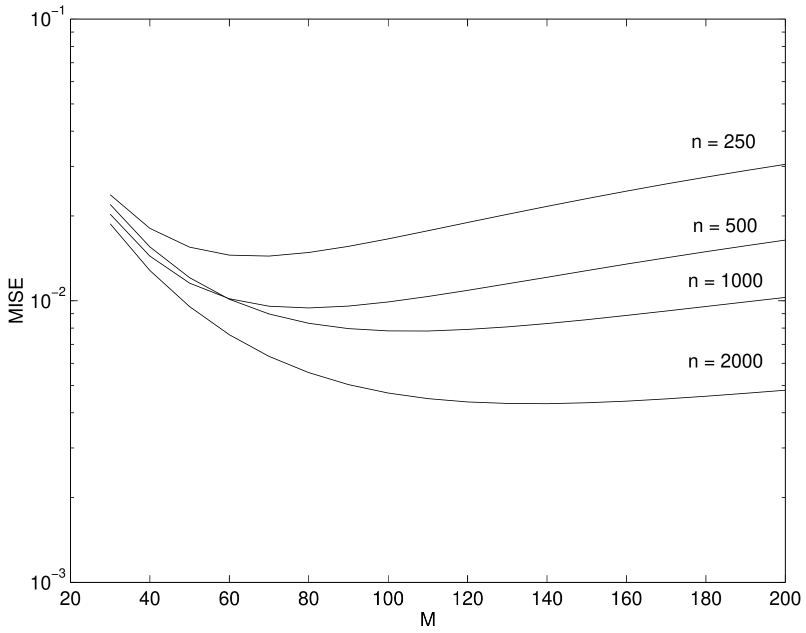

Figure 12: *MISE versus m and n for the ψ distribution on the surface of a sphere using the symmetric form of the estimator.*

The experiments show that the optimal values of m are sufficiently small for the distributions tested to make this scheme faster than the kernel estimate. However, as the distribution function gets more complicated (multimodal, for instance), we need higher values of m to get good estimates. Therefore such cases would require more computational effort than those with a more uniform distribution. If derivatives of the density are also required, than we are forced to increase m at a slower rate with respect to n to avoid sharp peaks, but then the cosine method becomes even more advantageous.

From experiments, we can perform a least squares analysis and approximate m as $kn^{1/2.5}$ for 1-D density estimation and as $kn^{1/(3+\delta)}$ for the derivative, for a constant k. δ can be any small number and is introduced in order to satisfy the upper bound for derivatives derived in section 4. Choosing $m = kn^{1/3}$ appears to give reasonable estimates for density on the surface of a sphere. If we need derivatives, we need to make m a more slowly increasing function of the density. Here, the magnitude of k depends on the complexity of the density function. It varies between 1 and 10 for the distributions considered here. If a more automatic choice of m is desired, it should be possible to make k a function of the variation. This can be done by choosing an initial value for m and and estimating the variation using the approximate density currently available. New density estimation computations can then be performed using the new value of m.

9 Conclusions

In this paper, we have described a weight function estimator for non-parametric estimation of probability density functions based on cosines, and provided conditions under which the estimate and its derivatives converge to the actual function. 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. We have also given an empirical formula for choosing the weight function exponent parameter of the estimator. There is a potential for further study of properties of this estimator and schemes for the automatic determination of the best exponent.

References

[1] Bhattacharya P.K. *Estimation of a probability density function and its derivatives*, Sankhyā Ser. A, 29 (1967) pp. 373–382.

[2] Bickel P.J. and M. Rosenblatt *On some global measures of the deviations of density function estimates*, Annals of Statistics, 1 (1973) pp. 1071–1095.

[3] Chentsov N.N. *Estimation of unknown probability density based on observations*, Dokl. Akad. Nauk SSSR, 147 (1962) pp. 45–48 (in Russian).

[4] Duda R.O. and P.E. Hart *Pattern classification and scene analysis*, John Wiley and sons, Inc., New York, 1973.

[5] Gradshteyn I.S. and I.M. Ryzhik *Table of Integrals Series, and Products*, Academic Press, Inc., Orlando, 1980, p. 374.

[6] Hernquist L. and Katz N. *TREESPH: A unification of SPH with the hierarchical tree method*, Astrophys. J. suppl., 70 (1989) pp. 419–446.

[7] Hwang J. *Non-parametric multivariate density estimation: A comparative study*, IEEE Trans. Signal Processing, 42 (1994) pp. 2795–2810.

[8] Jones M.C. and Lotwick H.W. *A remark on Algorithm AS 176. Kernel density estimation using the fast Fourier transform* , Appl. Statist., 33 (1984) pp. 120–122.

[9] Kronmal R. and Tarter M. *The estimation of probability densities and cumulatives by Fourier series methods*, American Statist. Association J., (1968) pp. 925–952.

[10] Knuth D.E. *The Art of Computer Programming, Vol. 2, Seminumerical Algorithms*, Addison-Wesley, Menlo Park, CA, 1968. pp. 41–52.

[11] Markushevich A.I. *Theory of Functions of a Complex Variable*, Chelsea Publishing Company, New York, 1985, Vol 2. pp. 26–34.

[12] Lavenberg S.S. *Computer Performance Modeling Handbook*, Academic Press Inc., New York, 1983.

[13] Monaghan J.J. and Lattanzio J.C. *A refined particle method for astrophysical problems*, Astron. Astrophys. 149 (1985) pp. 135–143.

[14] Loève M. *Probability Theory*, D. Van Nostrand Co. Inc., Princeton, NJ, 1963.

[15] Monaghan J.J. *Smoothed particle hydrodynamics*, Annu. Rev. Astron. Astrophys. 30 (1992) pp. 543–574.

[16] Nadaraya E.A. *Non-parametric Estimation of Probability Densities and Regression Curves*, Mathematics and Applications (Soviet Series)1, Kluwer Academic Publishers, Boston, 1989.

[17] Parzen E. *On estimation of a probability density function and mode*, Ann. Math. Statist., 33 (1962) pp. 1065–1076.

[18] Rosenblatt M. *Remarks on some non-parametric estimates of a density function*, Ann. Math. Statist., 27 (1956) pp. 832–837.

[19] Schwartz S.C. *Estimation of probability density by an orthogonal series*, Ann. Math. Statist., 38 (1967) pp. 1261–1265.

[20] Silverman B.W. *Kernel density estimation using the fast Fourier transform*, Appl. Statist., 31 (1982) pp. 93–99.

[21] Smirnov M.V *On the approximation of probability densities of random variables*, Scholarly Notes of Moscow State Polytechnical Institute, 16 (1951) pp. 69–96 (in Russian).

[22] Smith G. D. *Numerical solution of partial differential equations : finite difference methods*, Oxford University Press, New York, 1978.

[23] Silverman B.W. *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, 1993.

[24] Vio R., Fasano G., Lazzarin M., and Lessi O. *Probability density estimation in astronomy*, Astron. Astrophys., 289 (1994) pp. 640–648.

[25] Watson G.S. and Leadbetter M.R. *On the estimation of the probability density, I*, Ann. Math. Statist., 34 (1963) pp. 480–491.

[26] Whittle P. *On the smoothing of probability density functions*, J. Roy. Statist. Soc. B, 20 (1958) pp. 334–343.