Integrated Risk of the Kernel Estimator with Spherical Data*

Jussi Klemelä

Rolf Nevanlinna Institute
University of Helsinki

*This work was supported in part by the Academy of Finland under Grant no. 1015998 and by the University of Helsinki under a research grant.

Proposed running head: Integrated Risk of the Kernel Estimator

Correspondence author:

Jussi Klemelä
Rolf Nevanlinna Institute
P.O. Box 4
SF-00014 UNIVERSITY OF HELSINKI
FINLAND

Abstract

Estimating the density function of a random vector taking values on the d-dimensional unit sphere is considered. The article extends some results of Hall, Watson and Cabrera (1987) about the asymptotics of the kernel estimator. First, the definition of the kernel estimator is motivated using the concept of rotation. Then fast convergence rate theory is developed with both the L_1 error and the L_2 error. It is proved that the concept of second derivative for functions $S_d \to \mathbf{R}$, appearing in the asymptotic bias of the kernel estimator is the same as the Laplace operator multiplied by a constant. Finally, it is proved that asymptotically the plug–in method is as good as using the asymptotically optimal deterministic smoothing parameter sequence.

Key words and phrases

kernel estimator, Laplace operator, L_p error, nonparametric density estimation, plug–in method, spherical data.

AMS 1991 subject classifications

Primary: 62G07
Secondary: 62H11

1 Introduction

When statistical data consists of directions, it can be represented as points of the unit sphere in the Euclidean space. The notation $S_d = \{x \in \mathbf{R}^{d+1} : \|x\| = 1\}$, $d \geq 1$, is used but only the case $d \geq 2$ will be considered. From a practical point of view, the case $d = 2$ is the most important, but see Diaconis (1988, page 100) for the case where $d \geq 3$. Good introductions to the statistics of spherical data are given by Mardia (1972), Watson (1983) and Fisher, Lewis and Embleton (1987).

It is assumed to have n independent identically distributed observations from a distribution which is absolutely continuous with respect to the Lebesgue measure of the sphere, which will be denoted by $\mu = \mu_d$. The Lebesgue measure of the sphere can be defined with the help of the Lebesgue measure of $\mathbf{R}^{d+1}$, see Rudin (1986, page 175). It is a Borel measure, invariant with respect to rotations and $\omega_d \overset{def}{=} \mu(S_d) = 2\pi^{(d+1)/2}/\Gamma((d+1)/2)$.

If $X_1, \ldots, X_n$ are independent identically distributed random variables with values in $\mathbf{R}^d$, $d \geq 1$, an estimator for their common density is

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_\kappa(x - X_i),$$

where $x \in \mathbf{R}^d$ and $K_\kappa : \mathbf{R}^d \to \mathbf{R}$ is a function which depends on a positive parameter κ and whose mass concentrates more and more in the vicinity of the origin as $\kappa \to \infty$. By this we mean that $\lim_{\kappa\to\infty} \int_{\|y\|>\delta} |K_\kappa(y)| dy = 0$ for every $\delta > 0$. If also $\lim_{\kappa\to\infty} \int_{\mathbf{R}^d} K_\kappa = 1$, then the term approximate identity or delta sequence has been used for such sequences of functions (Butzer and Nessel, 1971). This estimator, which could be called a delta sequence estimator, was studied by Watson and Leadbetter (1963). The delta sequence estimator is called kernel estimator if $K_\kappa(x) = \kappa^d K(\kappa x)$, where $K : \mathbf{R}^d \to \mathbf{R}$ is a kernel function (typically a density function) and $\kappa > 0$ is a smoothing parameter.

The delta sequence estimator can be defined also in the case of spherical data. Let $\eta \in S_d$ be fixed and assume that we have constructed such functions $K_\kappa : S_d \to \mathbf{R}$ for $\kappa > 0$ that the mass of K_κ concentrates more and more in the vicinity of η, when $\kappa \to \infty$. For $x, y \in S_d$, let $R_{x,y} : S_d \to S_d$ be such a rotation that $R_{x,y}(x) = y$ (rotation is a restriction to S_d of a linear map whose matrix is orthogonal with determinant one). If $X_1, \ldots, X_n$ are i.i.d random variables with values on S_d, the delta sequence estimator of their common density can be defined as

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_\kappa(R_{X_i,\eta}(x)), \tag{1}$$

where $x \in S_d$. This definition replaces translation in Euclidean space by rotation on the sphere. The function K_κ is "centered" on each observation in turn and the average over the observations is taken. An essentially equivalent way of defining the delta sequence estimator in the case of Euclidean data is

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_\kappa(X_i - x).$$

In the case of spherical data we could have defined

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_\kappa(R_{x,\eta}(X_i)). \tag{2}$$

In the case of Euclidean data, sometimes a restriction is made to consider only delta sequences of the form $K_\kappa(x) = \kappa^d L(\|\kappa x\|)$, where $x \in \mathbf{R}^d$ and $L : [0,\infty[\to \mathbf{R}$. These functions depend only on the distance between the argument and the origin. Similarly in the case of spherical data we get a large class of delta sequences when restricting ourselves to functions which depend only on the angle between the argument and the vector η. In Watson (1970), Beran (1979), Hall, Watson and Cabreira (1987), Bai, Radhakrishna Rao and Zhao (1988) the form $K_\kappa(x) = c(\kappa)L(\kappa^2(1 - x'\eta))$ for the delta sequence has been considered, where $L : [0,\infty[\to \mathbf{R}$ and $c(\kappa)$ is the normalization constant. In Hall, Watson and Cabrera (1987), also the form $K_\kappa(x) = c(\kappa)J(\kappa^2 x'\eta)$ has been considered, where $J : \mathbf{R} \to \mathbf{R}$. It was shown that asymptotically the latter form is a special case of the first form. This study will be restricted to the case of $K_\kappa(x) = c(\kappa)L(\kappa \arccos(x'\eta))$. For this case it holds that $K_\kappa(R_{X_i,\eta}(x)) = c(\kappa)L(\kappa \arccos(x'X_i)) = K_\kappa(R_{x,\eta}(X_i))$ and thus in this case the definitions (1) and (2) can be simplified to the following.

Definition 1.1 *Kernel density estimator with the kernel function $L : [0,\infty[\to \mathbf{R}$ and the smoothing parameter $\kappa > 0$ is*

$$\hat{f}_n(x) = \hat{f}_n(x,\kappa,L) = \frac{c(\kappa)}{n}\sum_{i=1}^{n} L(\kappa \arccos(x'X_i)),$$

where $x \in S_d$ *and*

$$c(\kappa)^{-1} = \int_{S_d} L(\kappa \arccos(x'y))d\mu(y). \tag{3}$$

From equation (11) follows that $c(\kappa) \sim \kappa^d \left(\omega_{d-1} \int_0^\infty t^{d-1}L(t)dt\right)^{-1}$, where $a_\kappa \sim b_\kappa$ means $\lim_{\kappa\to\infty}(a_\kappa/b_\kappa) = 1$.

Reasonable choices for the kernel function are for example $L(t) = e^{-t^2} I_{[0,\infty[}(t)$, $L(t) = (1-t^2)I_{[0,1]}(t)$ and $L(t) = I_{[0,1]}(t)$. The choice $L(t) = I_{[0,1]}(t)$ leads to the so called naive estimator, $\hat{f}_n(x) = P_n(C_\kappa(x))/\mu(C_\kappa(x))$, where $P_n(C_\kappa(x)) = n^{-1}\sum_{i=1}^n I_{C_\kappa(x)}(X_i)$ and $C_\kappa(x) = \{y \in S_d \mid x'y \le \cos(\kappa^{-1})\}$. The naive estimator was studied by Ruymgaart (1989), Hendriks, Janssen and Ruymgaart (1993).

In section 2 the bias of the kernel estimator is studied. The expectation of the kernel estimator is a convolution and thus section 2 is concerned with the study of approximation by convolution. In Lemma 2.2 an expansion of convolution is given. The remainder term of this expansion can be written with the help of the so–called associated delta sequence, which is defined in Definition 2.1. The expansion of convolution in the spherical case has a different property from the corresponding expansion in the Euclidean case. In the spherical case lower order terms of the expansion do not necessarily vanish with any choice of the kernel function. They can, however, be made to approach zero at a sufficiently fast rate of convergence. For this, the concept of a class s kernel is needed, which is defined in Definition 2.4. This leads to the result that, unlike in the Euclidean case, the asymptotics of the bias involves a linear combination of derivatives of the density. Hall, Watson and Cabrera (1987) gave the asymptotics of the bias assuming smoothness index $s = 2$. In Lemma 2.5 the asymptotics of the bias is given assuming a general smoothness index.

In Section 3 it is proved that the concept of second derivative for functions $S_d \to \mathbf{R}$, defined by Hall, Watson and Cabrera (1987), is the same as the Laplace operator multiplied by a constant. In Section 4 the results about the asymptotics of the mean integrated squared error and the mean integrated absolute error are given.

With spherical data, selection of the smoothing parameter of the kernel estimator with the plug–in method has not been studied previously. In Section 5 it is proved that the smoothing parameter can be chosen in such a way that asymptotic risk is the same as when using the asymptotically optimal deterministic smoothing parameter.

The result differs from the result in the case of real line by Nadaraya (1974) in that he considered only a specific initial estimator (derivative of kernel estimator). Here sufficient conditions are formulated for any initial estimator, to be used in the plug–in method, to satisfy. The proofs and some technical details are given in the remainng sections.

Let us give the definitions of two parameterizations of S_d. Let $\eta \in S_d$ and $T_\eta = \{\xi \in S_d \,|\, \xi \perp \eta\}$. Let $\phi_\eta : S_d \setminus \{\eta, -\eta\} \to T_\eta \times]0,\pi[$ be a parameterization of S_d defined by

$$\phi_\eta^{-1}(\xi,\theta) = \eta\cos\theta + \xi\sin\theta. \tag{4}$$

Note that $\phi_\eta^{-1}(\xi,\theta)$ is well defined for all $\theta \in \mathbf{R}$, although there does not exist any function for which $\phi_\eta^{-1} : T_\eta \times \mathbf{R} \to S_d$ would be an inverse function. A second parameterization can be deduced from the first one. Let $\eta^1, \ldots, \eta^{d+1} \in S_d$ be such that $\eta^i \perp \eta^j$, when $i \neq j$. Let $\phi_{\eta^1,\ldots,\eta^{d+1}} : S_d \setminus U_{\eta^1,\ldots,\eta^{d+1}} \to]0,2\pi[\times]0,\pi[^{d-1}$ be defined by

$$\phi^{-1}_{\eta^1,\ldots,\eta^{d+1}}(\theta) = \sum_{i=1}^{d+1} \eta^i \cos\theta_{i-1}\sin\theta_i\cdots\sin\theta_d, \tag{5}$$

where $\theta = (\theta_1,\ldots,\theta_d)'$, $\cos\theta_0 = 1$, and

$$U_{\eta^1,\ldots,\eta^{d+1}} = \left\{ \sum_{i=2}^{d+1} \eta^i \cos\theta_{i-1}\sin\theta_i\cdots\sin\theta_d \;\middle|\; (\theta_2,\ldots,\theta_d) \in [0,\pi]^{d-1},\ \theta_1 = 0 \right\}.$$

To the first parameterization corresponds the integration formula

$$\int_{S_d} f(x)d\mu(x) = \int_0^\pi d\theta \sin^{d-1}\theta \int_{T_\eta} f\left(\phi_\eta^{-1}(\xi,\theta)\right) d\mu_{d-1}(\xi). \tag{6}$$

2 The Bias of the Kernel Estimator

For $L : [0,\infty[\to \mathbf{R}$, define $L_\kappa : [-1,1] \to \mathbf{R}$ by $L_\kappa(t) = L(\kappa \arccos t)$, $\kappa > 0$. Define the convolution of $f : S_d \to \mathbf{R}$ and L_κ by

$$f * L_\kappa(x) = \int_{S_d} f(y)L_\kappa(x'y)d\mu(y) = \int_{S_d} f(y)L(\kappa\arccos x'y)d\mu(y),\ x \in S_d.$$

For the kernel estimator $\hat{f}_n$,

$$\mathrm{E}\left(\hat{f}_n(x)\right) = c(\kappa)\mathrm{E}\left(L(\kappa\arccos(x'X_1)\right) = c(\kappa) f * L_\kappa(x),$$

which explains why the study of convolutions is important for the study of the kernel estimator. Also the variance of the kernel estimator involves convolution as will be seen in Section 4. Later, in Lemma 2.2, an expansion of convolution will be given. The remainder term of this expansion is convenient to write in terms of an associated delta sequence, which will be defined next.

Definition 2.1 *Let $s \geq 0$ be integer. Let $L : [0,\infty[\to \mathbf{R}$ be such that $\int_0^\infty t^{s+d-1}|L(t)|dt < \infty$. The parameter s delta sequence associated with the delta sequence $c(\kappa)L(\kappa \arccos x'y)$ is $\tilde{L}_\kappa^{(s)} : [0,\pi] \to \mathbf{R}$, defined by*

$$\tilde{L}_\kappa^{(s)}(\theta) = \omega_{d-1}\kappa^d \frac{\theta^s}{(s-1)!}\int_1^{\pi/\theta}(t-1)^{s-1}L(\kappa t\theta)\sin^{d-1}(t\theta)dt,$$

for $s \geq 1$ and

$$\tilde{L}_\kappa^{(0)}(\theta) = \omega_{d-1}\kappa^d L(\kappa\theta)\sin^{d-1}\theta.$$

The concept of an associated delta sequence is related to the concept of an associated kernel which is used in the Euclidean case. The associated kernel was defined for the one dimensional case in Bretagnolle and Huber (1979) and for the multivariate case in Holmström and Klemelä (1992). If $K : \mathbf{R} \to \mathbf{R}$ satisfies $\int_{\mathbf{R}} |x^s K(x)|dx < \infty$, for $s \geq 1$, then the parameter s kernel associated with K is defined by

$$\tilde{K}(x) = \frac{(-1)^s x^s}{(s-1)!}\int_1^\infty (t-1)^{s-1}K(tx)dt,\ x \in \mathbf{R}.$$

Some properties of the associated delta sequence are given in Section 6.

When $g : \mathbf{R}^{d+1} \to \mathbf{R}$ and $x, \xi \in \mathbf{R}^{d+1}$, define the derivative of g at x in the direction of ξ to be $D_\xi g(x) = \lim_{h\to 0} h^{-1}[g(x+h\xi) - g(x)]$ and $D_\xi^s g = D_\xi D_\xi^{s-1} g$, for $s \geq 2$ integer. When $f : S_d \to \mathbf{R}$, define $\bar{f} : \mathbf{R}^{d+1} \to \mathbf{R}$ by $\bar{f}(x) = f(x/\|x\|)$. Define $D^s f : S_d \to \mathbf{R}$ by

$$D^s f(x) = \omega_{d-1}^{-1}\int_{T_x} D_\xi^s \bar{f}(x) d\mu_{d-1}(\xi), \tag{7}$$

where $d \geq 2$, $T_x = \{\xi \in S_d : \xi \perp x\}$, and $\omega_{d-1} = \mu_{d-1}(S_{d-1})$. This concept was first defined by Hall, Watson and Cabrera (1987). Its relation to the Laplace operator is discussed in the next section. When $f : S_d \to \mathbf{R}$, define $\tilde{D}^s f : S_d \times \mathbf{R} \to \mathbf{R}$ by

$$\tilde{D}^s f(x,\theta) = \omega_{d-1}^{-1}\int_{T_x} D^s_{\phi_x^{-1}(\xi,\theta+\pi/2)}\bar{f}\left(\phi_x^{-1}(\xi,\theta)\right)d\mu_{d-1}(\xi), \tag{8}$$

where $s \geq 0$ is integer and ϕ_x was defined in (4). Note that $\phi_x^{-1}(\xi,\theta+\pi/2) \perp \phi_x^{-1}(\xi,\theta)$. The following lemma gives an expansion of convolution.

Lemma 2.2 *Let $x \in S_d$ and $s \geq 0$ even. Assume that for all $\xi \in T_x$, $\partial^s/\partial\theta^s f(\phi_x^{-1}(\xi,\theta))$ is continuous as a function of $\theta \in \mathbf{R}$. Let $\int_0^\infty t^{i+d-1}|L(t)|dt < \infty$ for $i = 0, s$. Then*

$$\kappa^d f * L_\kappa(x) = \sum_{i=0}^{s/2-1} \frac{1}{(2i)!} d_{2i}(\kappa, L) D^{2i} f(x) + \int_0^\pi \tilde{L}_\kappa^{(s)}(\theta) \tilde{D}^s f(x,\theta) d\theta,$$

where

$$d_{2i}(\kappa, L) = \omega_{d-1}\kappa^d \int_0^\pi \theta^{2i} L(\kappa\theta) \sin^{d-1}\theta d\theta. \tag{9}$$

The proof of Lemma 2.2 is given in Section 6. In the Euclidean case we have the following expansion of the convolution. If $f : \mathbf{R} \to \mathbf{R}$ is s times continuously differentiable and $K : \mathbf{R} \to \mathbf{R}$ is such that $\int_{\mathbf{R}} |x^s K(x)| dx < \infty$, where $s \geq 1$ is an integer, then

$$\kappa \int_{\mathbf{R}} f(y) K(\kappa(x-y)) dy = \sum_{i=0}^{s-1} \frac{\kappa^{-i}}{i!} \int_{\mathbf{R}} y^i K(y) dy f^{(i)}(x) + \kappa^{1-s} \int_{\mathbf{R}} f^{(s)}(y) \tilde{K}(\kappa(x-y)) dy,$$

when $\kappa \to \infty$, where $\tilde{K}$ is the parameter s kernel associated with K. For the proof of this expansion, see e. g. Holmström and Klemelä (1992).

The next step is to expand the terms $d_{2i}(\kappa, L)$ appearing in the expansion of convolution. Define

$$\alpha_s = \alpha_s(L) = \int_0^\infty t^{s+d-1} L(t) dt, \tag{10}$$

where $s \geq 0$. The next lemma gives an expansion of $d_m(\kappa, L)$ in terms of α_i.

Lemma 2.3 *Let $m \geq 0$, $r \geq 0$ be integers and suppose that*

$$\int_0^\kappa t^{m+2i+d-1} L(t) dt = \alpha_{m+2i}(L) + o(\kappa^{2i-2r}),$$

for $i = 0, \ldots, r$. Then

$$d_m(\kappa, L) = \omega_{d-1} \sum_{i=0}^{r} \kappa^{-m-2i} \gamma_i \alpha_{m+2i}(L) + o(\kappa^{-m-2r}),$$

where

$$\gamma_i = \sum_{\alpha_1 + \cdots + \alpha_{d-1} = i} \frac{(-1)^{\alpha_1}}{(2\alpha_1 + 1)!} \cdots \frac{(-1)^{\alpha_{d-1}}}{(2\alpha_{d-1} + 1)!},\ \gamma_0 = 1.$$

The proof of Lemma 2.3 is given in Section 6. As a special case of Lemma 2.3 we have that when $\alpha_0(|L|) < \infty$, then

$$d_0(\kappa, L) = \kappa^d c(\kappa)^{-1} = \kappa^d \int_{S_d} L(\kappa \arccos x'\eta) d\mu(x) = \omega_{d-1}\alpha_0(L) + o(1), \tag{11}$$

when $\kappa \to \infty$, where $c(\kappa)$ was defined in (3).

Define, for $f : S_d \to \mathbf{R}$ and $s \geq 2$ even,

$$\mathcal{D}^s f = \sum_{i=1}^{s/2} \frac{1}{(2i)!} \gamma_{s/2-i} D^{2i} f, \tag{12}$$

where $\gamma_{s/2-i}$ were defined in Lemma 2.3. Suppose that

$$\int_0^\pi \tilde{L}_\kappa^{(s)}(\theta) \tilde{D}^s f(x, \theta) d\theta = \frac{1}{s!} d_s(\kappa, L) D^s f(x) + o(\kappa^{-s})$$

and

$$\int_0^\kappa t^{2i+d-1} L(t) dt = \alpha_{2i}(\kappa, L) + o(\kappa^{2i-s})$$

for $i = 1, \ldots, s/2$. Then from Lemmas 2.2 and 2.3 it would follow that

$$c(\kappa) f * L_\kappa(x) = \alpha_0(L)^{-1} \sum_{i=0}^{s/2} \kappa^{-2i} \alpha_{2i}(L) \mathcal{D}^{2i} f(x) + o(\kappa^{-s}).$$

This equation is not proved here but it motivates the definition of a class s kernel. A class s kernel is such that the lower order terms of the previous expansion vanish.

Definition 2.4 *Let $s \geq 0$ be even. A class s kernel is a measurable function $L : [0, \infty[\to \mathbf{R}$ which satisfies*

(i) $\alpha_i(|L|) < \infty$ *for* $i = 0, s$,

(ii) $\int_0^\kappa t^{2i+d-1} L(t) dt = \alpha_{2i}(L) + o(\kappa^{2i-s})$ *for* $i = 1, \ldots, s/2 - 1$.

(iii) $\alpha_0(L) \neq 0$, $\alpha_{2i}(L) = 0$ *for* $i = 1, \ldots, s/2 - 1$.

When L has a compact support, condition (ii) is satisfied. Thus, to construct a class s kernel, L can be fit into a polynomial model on $[0, 1]$. The definition of a class s kernel in the Euclidean case is given for example in Prakasa Rao (1983). The idea for reducing the bias by using a class s kernel was introduced in Parzen (1962), Bartlett (1963). To see the similarity between the spherical case and the Euclidean case note that when $K : \mathbf{R}^d \to \mathbf{R}$, $K(x) = L(\|x\|)$, then

$$\begin{aligned} \int_{\mathbf{R}^d} K &= \int_0^\infty dr\, r^{d-1} \int_{S_{d-1}} K(rx) d\mu_{d-1}(x) = \omega_{d-1} \int_0^\infty r^{d-1} L(r) dr \\ &= \omega_{d-1} \alpha_0(L) \end{aligned} \tag{13}$$

and

$$\begin{aligned} &\int_{\mathbf{R}^d} y_1^s K(y) dy = \int_0^\infty dr\, r^{d-1} \int_{S_{d-1}} (rx_1)^s K(rx) d\mu_{d-1}(x) \qquad (14) \\ &= \int_{S_{d-1}} x_1^s d\mu_{d-1}(x) \int_0^\infty r^{s+d-1} L(r) dr = \int_{S_{d-1}} x_1^s d\mu_{d-1}(x) \alpha_s(L). \end{aligned}$$

Let us define the smoothness class used in this study. Let $s \geq 2$ be even and $1 \leq p < \infty$. Let $\mathbf{F}(s,p)$ be the set of such functions $f : S_d \to \mathbf{R}$ that $\|D^i f\|_p < \infty$ for $i = 0, \ldots, s$, for all $x \in S_d$ and for all $\xi \in T_x = \{\xi \in S_d \mid \xi \perp x\}$, $\partial^s / \partial\theta^s f(\phi_x^{-1}(\xi, \theta))$ is continuous as a function of $\theta \in \mathbf{R}$, $\left\| \tilde{D}^s f(\cdot, \theta) \right\|_p$ is bounded for $\theta \in [0, \pi]$, and $\lim_{\theta \to 0} \left\| \tilde{D}^s f(\cdot, \theta) - D^s f \right\|_p = 0$.

Now we are ready to give a result about the L_p convergence of convolutions. The first part of the next lemma is needed when studying the variance of the kernel estimator. The corresponding lemma in the Euclidean case has been proved for example in Nikol'skii (1975, page 28). The second part assumes a smoothness condition and it is used when studying the bias of the kernel estimator.

Theorem 2.5 *Let $f : S_d \to \mathbf{R}$ and $L : [0, \infty[\to \mathbf{R}$. Let $1 \leq p < \infty$.*

(i) *Assume $\|f\|_p < \infty$ and let L be a class 0 kernel. Then*

$$\lim_{\kappa \to \infty} \|c(\kappa) f * L_\kappa - f\|_p = 0.$$

(ii) *Assume $f \in \mathbf{F}(s,p)$ and let L be a class s kernel, where $s \geq 2$ is even. Then*

$$\lim_{\kappa \to \infty} \left\| \kappa^s \left| c(\kappa) f * L_\kappa - f \right| - \left| \alpha_0(L)^{-1} \alpha_s(L) \mathcal{D}^s f \right| \right\|_p = 0.$$

The proof of Theorem 2.5 is given in Section 6.

3 The Laplace Operator

In equation (7) a concept of a derivative for functions $S_d \to \mathbf{R}$ was defined. In the next theorem the relation of this concept to the concept of Laplace operator is given. Let $f : S_d \to \mathbf{R}$ and let $\eta^1, \ldots, \eta^{d+1} \in S_d$ be orthogonal. Let $\phi = \phi_{\eta^1,\ldots,\eta^{d+1}}$ be as defined in (5). Define the Laplace operator by

$$\Delta f(x) = \sum_{k=1}^{d} \sin^{-2}\theta_{k+1} \cdots \sin^{-2}\theta_d \left[\frac{\partial^2}{\partial \theta_k^2} + (k-1)\frac{\cos\theta_k}{\sin\theta_k}\frac{\partial}{\partial\theta_k} \right] f\left(\phi^{-1}(\theta)\right)\Big|_{\theta=\phi(x)}.$$

Theorem 3.1 *For $f : S_d \to \mathbf{R}$, $d \geq 2$,*

$$D^2 f = d^{-1}\Delta f,$$

if $\bar{f}$ and its partial derivatives are differentiable, where $\bar{f}(x) = f(x/\|x\|)$.

The proof of Theorem 3.1 is given in Section 7. The proof uses the fact that for the Laplace operator of S_d, $\Delta f(x) = \sum_{i=1}^{d+1} \frac{\partial^2}{\partial x_i^2}\bar{f}(x)$. Additionally it is proved that $x'\mathrm{Hess}\bar{f}(x)x = 0$ for all $x \in S_d$. By considering a function of form $f : S_d \to \mathbf{R}$, $f(x) = g(x'\eta)$, where $\eta \in S_d$, $g : [-1,1] \to \mathbf{R}$, it is seen that there exist no constant C for which $D^4 = C\Delta^2$.

4 Variance and Asymptotic Risk of the Kernel Estimator

Two measures of risk will be studied. First the mean integrated squared error and then the mean integrated absolute error. Measuring loss by the integrated squared error is technically easier but the integrated absolute error is more natural for several reasons. Firstly, it is defined for all densities. Secondly, it is invariant under scale changes. Thirdly, it is proportional to the total variation metric, that is $\int_{S_d} |f - g| d\mu = 2\sup_B |\int_B f d\mu - \int_B g d\mu|$. Extensive theory of the L_1 error in the Euclidean case has been developed in Devroye and Györfi (1985), Devroye (1987).

Let us start with studying the variance of the kernel estimator. Part (i) of the following lemma is needed when calculating the mean integrated squared error of the kernel estimator and part (ii) is needed when calculating the mean integrated absolute error.

Lemma 4.1 *Let $f : S_d \to \mathbf{R}$ be a density function. Let $\alpha_0(|L|^i) < \infty$ for $i = 1, 2$.*

(i) *Let $\|f\|_2 < \infty$. Then,*

$$\int_{S_d} \operatorname{Var}\left(\hat{f}_n\right) d\mu = \frac{\kappa^d}{n} \frac{\alpha_0(L^2)}{\omega_{d-1}\alpha_0(L)^2} + \frac{o(\kappa^d)}{n},$$

when $\kappa \to \infty$.

(ii) *Secondly,*

$$\int_{S_d} \left| \sqrt{\operatorname{Var}(\hat{f}_n)} - \sqrt{\frac{\kappa^d}{n} \frac{\alpha_0(L^2)}{\omega_{d-1}\alpha_0(L)^2} f} \right| d\mu = \frac{o(\kappa^{d/2})}{n^{1/2}},$$

when $\kappa \to \infty$.

The proof of Lemma 4.1 is given in Section 8. Now it is possible to give the asymptotic of the mean integrated squared error. From Lemma 2.5 (ii) and Lemma 4.1 (i) it follows that

$$\begin{aligned} \mathrm{E} \int_{S_d} \left(\hat{f}_n(\cdot,\kappa) - f\right)^2 d\mu &= \int_{S_d} \left(\mathrm{E}(\hat{f}_n(\cdot,\kappa)) - f\right)^2 d\mu + \int_{S_d} \operatorname{Var}\left(\hat{f}_n(\cdot,\kappa)\right) d\mu \\ &= \kappa^{-2s} \frac{\alpha_s(L)^2}{\alpha_0(L)^2} \int_{S_d} (\mathcal{D}^s f)^2 \, d\mu + \frac{\kappa^d}{n} \frac{\alpha_0(L^2)}{\omega_{d-1}\alpha_0(L)^2} + o(\kappa^{-2s}) + \frac{o(\kappa^d)}{n}. \end{aligned}$$

The optimal rate of convergence is achieved by choosing $\kappa = Cn^{1/(2s+d)}$ and then it is true that

$$\begin{aligned} &\lim_{n\to\infty} n^{2s/(2s+d)} \mathrm{E} \int_{S_d} \left(\hat{f}_n(\cdot,\kappa) - f\right)^2 d\mu \\ &= C^{-2s} \frac{\alpha_s(L)^2}{\alpha_0(L)^2} \int_{S_d} (\mathcal{D}^s f)^2 \, d\mu + C^d \frac{\alpha_0(L^2)}{\omega_{d-1}\alpha_0(L)^2}. \end{aligned} \tag{15}$$

This expression is minimized with respect to C, when $C = C^*$, where

$$C^* = \left[A(L) \int_{S_d} (\mathcal{D}^s f)^2 d\mu\right]^{1/(2s+d)} \tag{16}$$

and $A(L) = 2s\omega_{d-1}\alpha_s(L)^2/(d\alpha_0(L^2))$. The expression (15) evaluated at C^* has the value

$$\left[\int_{S_d} (\mathcal{D}^s f)^2\right]^{d/(2s+d)} M(L) \left(\omega_{d-1}\right)^{-2s/(2s+d)} (2s/d)^{(d-2s)/(2s+d)} \tag{17}$$

where

$$M(L) = \left[\left(\frac{\alpha_s(L)}{\alpha_0(L)} \right)^{2d} \left(\frac{\alpha_0(L^2)}{\alpha_0(L)^2} \right)^{2s} \right]^{1/(2s+d)}.$$

Minimum of $M(L)$ with respect to functions $L : [0, \infty[\to \mathbf{R}$ is achieved by

$$L_0(t) = (1 - t^s) I_{[0,1]}(t),$$

where $s \geq 2$ is an even integer. For $s = 2$ this was noted by Hall, Watson and Cabrera (1987). Note that L_0 is a class s kernel only when $s = 2$. The following argument can be used to prove that L_0 minimizes $M(L)$. Let $K : \mathbf{R}^d \to \mathbf{R}$, $K(x) = L(\|x\|)$. Then by equation (13), $\int_{\mathbf{R}^d} K = \omega_{d-1}\alpha_0(L)$ and similarly, $\int_{\mathbf{R}^d} K^2 = \omega_{d-1}\alpha_0(L^2)$. Also, by equation (14), $\int_{\mathbf{R}^d} y_1^s K(y) dy = \int_{S_{d-1}} x_1^s d\mu_{d-1}(x) \alpha_s(L)$. Thus

$$M(L) = C \left[\left(\frac{\int_{\mathbf{R}^d} y_1^s K(y) dy}{\int_{\mathbf{R}^d} K} \right)^{2d} \left(\frac{\int_{\mathbf{R}^d} K^2}{(\int_{\mathbf{R}^d} K)^2} \right)^{2s} \right]^{1/(2s+d)},$$

where $0 < C < \infty$ is a constant. Thus, the expression to be minimized is the same as in the Euclidean case and the same argument can be used as in Devroye (1987, page 120). Note that the optimal kernel is the same as in the Euclidean case considered by Bartlett (1963), Epanechnikov (1969).

Next, the asymptotic of risk is calculated when the loss is the L_1 error. In the Euclidean case, the corresponding theorem has been given in Devroye and Györfi (1985, Theorem 1, Chapter 5), Hall and Wand (1988), and Holmström and Klemelä (1992). Define

$$A(t, u) = \begin{cases} u\gamma(t/u), & t \geq 0,\ u > 0 \\ 0, & t \geq 0,\ u = 0, \end{cases}$$

where

$$\gamma(u) = E|Z - u| = \sqrt{2/\pi} \left(u \int_0^u e^{-t^2/2} dt + e^{-u^2/2} \right), \text{ where } Z \sim N(0,1),\ u \geq 0.$$

The function A appears also in the Euclidean case. Recall that the definition of smoothness class $\mathbf{F}(s, p)$ is given before Lemma 2.5.

Theorem 4.2 *Let $s \geq 2$ be even. Assume that the density $f \in \mathbf{F}(s, 1)$. Let L be a bounded class s kernel. Let $\{\kappa_n\}$ be such a sequence that $\lim_{n\to\infty} \kappa_n = \infty$ and*

$\lim_{n\to\infty} \kappa_n^d n^{-1} = 0$. *Then*

$$\mathrm{E}\int_{S_d} \left|\hat{f}_n - f\right| d\mu = \int_{S_d} A\left(\kappa_n^{-s}|z|, \kappa_n^{d/2} n^{-1/2} w\right) d\mu + o(\kappa_n^{-s}) + o(\kappa_n^{d/2} n^{-1/2}),$$

where

$$\begin{aligned} z &= \alpha_0(L)^{-1}\alpha_s(L)\mathcal{D}^s f, \\ w &= \left[\frac{\alpha_0(L^2)}{\omega_{d-1}\alpha_0(L)^2} f\right]^{1/2}. \end{aligned}$$

Proof. Put $B_n = f - \mathrm{E}(\hat{f}_n)$ and $\sigma_n = \sqrt{\mathrm{Var}(\hat{f}_n)}$. Firstly, by a similar argument as in Devroye and Györfi (1985, page 91),

$$\int_{S_d} \left|\mathrm{E}|\hat{f}_n - f| - A(|B_n|, \sigma_n)\right| d\mu \le C\, c(\kappa_n) n^{-1} = o(\kappa_n^{d/2} n^{-1/2}),$$

where $0 < C < \infty$ is a constant. Secondly, because $|A(x,y) - A(t,u)| \le |x - t| + \sqrt{2/\pi}|y - u|$ by Devroye and Györfi (1985, page 93),

$$\begin{aligned} &\int_{S_d} \left|A(|B_n|, \sigma_n) - A\left(\kappa_n^{-s}|z|, \kappa_n^{d/2} n^{-1/2} w\right)\right| d\mu \\ &\quad \le \int_{S_d} \left||B_n| - \kappa_n^{-s}|z|\right| d\mu + \sqrt{\frac{2}{\pi}} \int_{S_d} \left|\sigma_n - \kappa_n^{d/2} n^{-1/2} w\right| d\mu = o(\kappa_n^{-s}) + o(\kappa_n^{d/2} n^{-1/2}) \end{aligned}$$

by Lemma 2.5 and Lemma 4.1 (ii). □

The optimal rate of convergence is achieved by choosing $\kappa = Cn^{1/(2s+d)}$ and then it is true that

$$\lim_{n\to\infty} n^{s/(2s+d)} \mathrm{E}\int_{S_d} |\hat{f}_n - f| d\mu = \int_{S_d} A\left(C^{-s}|z|, C^{d/2} w\right) d\mu.$$

Compare this formula to formula (15). Numerical minimization of this expression with respect to C has been considered in Hall and Wand (1988), in the Euclidean case. On the other hand, because $A(t,u) \le t + \sqrt{2/\pi} u$ by Devroye and Györfi (1985, page 77), we get an upper bound

$$\int_{S_d} A\left(C^{-s}|z|, C^{d/2} w\right) d\mu \le C^{-s} \int_{S_d} |z| d\mu + C^{d/2}\sqrt{2/\pi} \int_{S_d} w d\mu$$

which can be minimized explicitly with respect to C. For the Euclidean case, see also Holmström and Klemelä (1992, page 257).

5 Empirical Choice of the Smoothing Parameter

In the previous section the smoothing parameter was given, which is asymptotically optimal in the mean integrated squared error sense. The question arises whether there exists an empirical choice of the smoothing parameter which is equally good in the sense that it gives the same asymptotic for the risk. In this section such data–driven device for choosing the smoothing parameter is presented.

The so–called plug–in method will be used. This method is based on plugging in estimates of the unknown quantities appearing in the formula for the asymptotically optimal smoothing parameter. This method was apparently first introduced by Woodroofe (1970), with mean squared error criterion. The plug–in method with the mean integrated squared error was apparently first introduced by Nadaraya (1974). Later developments have been made for example by Scott, Tapia and Thompson (1977), Park and Marron (1990), Sheather and Jones (1991), Wand and Jones (1994), Engel, Herrmann and Gasser (1995).

There are many other approaches to smoothing parameter selection, for example methods based on cross–validation and bootstrap. The plug–in method has had one of the best performances in simulation studies with Euclidean data (see Park and Marron 1990, Cao, Cuevas, and González–Manteiga 1994). With spherical data, least squares cross–validation and likelihood cross–validation methods were considered by Hall, Watson and Cabrera (1987).

It has been common to define the optimal smoothing parameter as a statistic $\hat{\kappa}$ which minimizes $\mathrm{E}\,(\mathrm{MISE}(\hat{\kappa}))$, where $\mathrm{MISE}(\kappa) = \mathrm{E} \int \left(\hat{f}_n(\cdot,\kappa) - f\right)^2$. However, in this study the optimal smoothing parameter is taken to be a statistic $\hat{\kappa}$ which minimizes $\mathrm{E} \int \left(\hat{f}_n(\cdot,\hat{\kappa}) - f\right)^2$. A discussion of the differences between these approaches is given by Grund, Hall, and Marron (1994).

In Section 4 it was shown that if f has smoothness index s, the asymptotically optimal smoothing parameter sequence in the mean integrated squared error sense is

$$\kappa_n^* = C^* n^{1/(2s+d)}, \tag{18}$$

where C^* was defined in (16). The constant C^* depends on the unknown density through $\int_{S_d} (\mathcal{D}^s f)^2 d\mu$. Let us assume that there is such estimator $\hat{\theta}_{s,n}$ for $\int_{S_d} (\mathcal{D}^s f)^2 d\mu$

that,

$$\mathrm{E}(\hat{\theta}_{s,n}) - \int_{S_d} (\mathcal{D}^s f)^2 d\mu = o(1) \tag{19}$$

and

$$\mathrm{E}\left|\hat{\theta}_{s,n} - E(\hat{\theta}_{s,n})\right|^m = o\left(n^{-ms/(2s+d)}\right), \tag{20}$$

for all $m \in \{1, 2, \ldots\}$. Furthermore it is assumed that

$$\hat{\theta}_{s,n} \geq 0 \tag{21}$$

with probability one when n is large enough. For the construction of such estimator, see Klemelä (1997). Let

$$\hat{C}_n = \left(A(L)(\hat{\theta}_{s,n} + b_n)\right)^{1/(2s+d)} \wedge n^\gamma,$$

where $b_n = o(1)$, $nb_n \to \infty$, and $\gamma > 0$ is arbitrary. The data–driven smoothing parameter is defined as

$$\hat{\kappa}_n = \hat{C}_n n^{1/(2s+d)}.$$

Theorem 5.1 *Let $s \geq 2$ be even. Let $f : S_d \to \mathbf{R}$ be a continuous density, $f \in \mathbf{F}(s, 2)$, and $\int_{S_d} (\mathcal{D}^s f)^2 \, d\mu > 0$. (The definition of $\mathbf{F}$ was given before Lemma 2.5.) Assume that an estimator satisfying conditions (19), (20) and (21) exists. Let L be a class s kernel for which $\alpha_0(|L|^2) < \infty$ and $\alpha_2(|L'|) < \infty$. Put $J_1(t) = tL'(t)$ and assume that $|L| + |J_1| \leq J$, where $J : [0, \infty[\to \mathbf{R}$ is monotonically decreasing, bounded, and $\alpha_0(|J|^i) < \infty$ for $i = 1, 2$. Then*

$$\mathrm{E} \int_{S_d} \left(\hat{f}_n(\cdot, \hat{\kappa}_n) - f\right)^2 d\mu \sim \mathrm{E} \int_{S_d} \left(\hat{f}_n(\cdot, \kappa_n^*) - f\right)^2 d\mu,$$

where $a_n \sim b_n$ means $\lim_{n\to\infty}(a_n/b_n) = 1$ and κ_n^* was defined in equation (18).

The proof of Theorem 5.1 is given in Section 9.

6 The Proofs for Section 2

Let us first give some facts about the parameter s delta sequence $\tilde{L}_\kappa^{(s)}(\theta)$, defined in Definition 2.1. Let us prove that the parameter s delta sequence $\tilde{L}_\kappa^{(s)}(\theta)$ is well defined a. e. in $[0, \pi]$. This can be seen in the following way. Using Fubini's theorem and the substitution $t\theta = \eta$, for $s \geq 1$,

$$
\begin{aligned}
& \int_0^\pi \left| \tilde{L}_\kappa^{(s)}(\theta) \right| d\theta \qquad (22) \\
\leq \ & \frac{\omega_{d-1}\kappa^d}{(s-1)!} \int_0^\pi d\theta \, \theta^s \int_1^{\pi/\theta} (t-1)^{s-1} \left| L(\kappa t\theta) \right| \sin^{d-1}(t\theta) dt \\
= \ & \frac{\omega_{d-1}\kappa^d}{(s-1)!} \int_1^\infty dt (t-1)^{s-1} \int_0^{\pi/t} \theta^s \left| L(\kappa t\theta) \right| \sin^{d-1}(t\theta) d\theta \\
= \ & \frac{\omega_{d-1}\kappa^d}{(s-1)!} \int_1^\infty (t-1)^{s-1} t^{-s-1} dt \int_0^\pi \eta^s \left| L(\kappa\eta) \right| \sin^{d-1} \eta d\eta \\
= \ & \frac{1}{s!} d_s(\kappa, |L|).
\end{aligned}
$$

For $s = 0$ this follows from the definition. Define $g :]0, \pi] \to \mathbf{R}$ by $g(t) = t^{1-d} \sin^{d-1} t$. Now $\lim_{t \downarrow 0} g(t) = 1$. Thus there is such $0 < M < \infty$ that $|g| \leq M$. Thus, by the substitution $\kappa\eta = u$,

$$
\begin{aligned}
d_s(\kappa, |L|) &= \omega_{d-1}\kappa^d \int_0^\pi \eta^s \left| L(\kappa\eta) \right| \sin^{d-1} \eta d\eta \\
&= \omega_{d-1}\kappa^d \int_0^{\pi\kappa} (\kappa^{-1}u)^s |L(u)| \sin^{d-1}(\kappa^{-1}u) \kappa^{-1} du \\
&\leq \omega_{d-1}\kappa^{-s} M \alpha_s(|L|).
\end{aligned}
$$

If $\alpha_s(|L|) < \infty$, for $s \geq 0$ integer, then by similar steps as in (22),

$$
\int_0^\pi \tilde{L}_\kappa^{(s)}(\theta) d\theta = \frac{1}{s!} d_s(\kappa, L) \qquad (23)
$$

when $\kappa \to \infty$. Also, for every $0 < \delta \leq \pi$,

$$
\int_\delta^\pi \left| \tilde{L}_\kappa^{(s)}(\theta) \right| d\theta = o(\kappa^{-s}) \qquad (24)
$$

when $\kappa \to \infty$.

The following lemma gives an expression for directional derivatives, which will be needed in the proof of Lemma 2.2.

Lemma 6.1 *Let $f : S_d \to \mathbf{R}$, $x, \xi \in S_d$, $\xi \perp x$. Put $\phi_x^{-1}(\xi, \theta) = x \cos\theta + \xi \sin\theta$ where $\theta \in \mathbf{R}$. Now*

$$
D_\xi^s \bar{f}(x) = \left. \frac{\partial^s}{\partial \theta^s} f\left(\phi_x^{-1}(\xi, \theta)\right) \right|_{\theta=0}
$$

where $s \geq 0$ is an integer and $\bar{f}(x) = f(x/\|x\|)$.

Proof. When $\xi \in S_d$, define $D_\xi f : S_d \to \mathbf{R}$ by $D_\xi f = (D_\xi \bar{f})\big|_{S_d}$. Define $D_\xi^s f : S_d \to \mathbf{R}$ by $D_\xi^s f = (D_\xi \overline{D_\xi^{s-1} f})\big|_{S_d}$, for $s \geq 2$ integer. Let us first prove that

$$D_\xi^s f(x) = D_\xi^s \bar{f}(x) \tag{25}$$

for $x \in S_d$. This holds by definition for $s = 1$. When $x, \xi \in S_d$, $x \perp \xi$, we can write $x + h\xi = (1+h^2)^{1/2} \phi_x^{-1}(\xi, \arctan h)$. It holds also that $D_\xi^s \bar{f}(rx) = D_\xi^s \bar{f}(x)/r^s$, for $0 < r < \infty$. Thus, assuming (25) holds for s,

$$\begin{aligned} D_\xi^s \bar{f}(x + h\xi) &= D_\xi^s \bar{f}\left((1+h^2)^{1/2} \phi_x^{-1}(\xi, \arctan h)\right) \\ &= (1+h^2)^{-s/2} D_\xi^s \bar{f}\left(\phi_x^{-1}(\xi, \arctan h)\right) \\ &= (1+h^2)^{-s/2} D_\xi^s f\left(\phi_x^{-1}(\xi, \arctan h)\right) = (1+h^2)^{-s/2} \overline{D_\xi^s f}(x + h\xi) \end{aligned}$$

and

$$\begin{aligned} D_\xi^{s+1} \bar{f}(x) &= \lim_{h\to 0} h^{-1} \left[D_\xi^s \bar{f}(x + h\xi) - D_\xi^s \bar{f}(x)\right] \\ &= \lim_{h\to 0} h^{-1} \left[(1+h^2)^{-s/2} \overline{D_\xi^s f}(x + h\xi) - \overline{D_\xi^s f}(x)\right] = D_\xi^{s+1} f(x), \end{aligned}$$

because $\lim_{h\to 0} h^{-1}((1+h^2)^{-s/2} - 1) = 0$. Equation (25) has been proved.

Let us prove the assertion of this lemma for $D_\xi^s f(x)$, which in view of equation (25) proves it for $D_\xi^s \bar{f}(x)$. It holds that

$$\begin{aligned} D^s_{\phi_x^{-1}(\xi,\theta)} f(x) &= \lim_{h\to 0} h^{-1} \left[\overline{D^{s-1}_{\phi_x^{-1}(\xi,\theta)} f}(x + h\phi_x^{-1}(\xi, \theta)) - \overline{D^{s-1}_{\phi_x^{-1}(\xi,\theta)} f}(x)\right] \\ &= \lim_{h\to 0} h^{-1} \left[\overline{D^{s-1}_{\phi_x^{-1}(\xi,\theta)} f}\left(x + \frac{h \sin\theta}{1 + h\cos\theta}\xi\right) - \overline{D^{s-1}_{\phi_x^{-1}(\xi,\theta)} f}(x)\right] \\ &= \sin\theta \, D_\xi \overline{D^{s-1}_{\phi_x^{-1}(\xi,\theta)} f}(x) = \sin^s\theta \, D_\xi^s f(x). \end{aligned}$$

Put $z = \phi_x^{-1}(\xi, h)$ and $\eta = \phi_x^{-1}(\xi, h + \pi/2)$. Now $\phi_x^{-1}(\xi, h + \theta) = \phi_z^{-1}(\eta, \theta)$. Thus, assuming induction hypothesis,

$$\frac{\partial^s}{\partial\theta^s} f(\phi_x^{-1}(\xi, h + \theta))\bigg|_{\theta=0} = \frac{\partial^s}{\partial\theta^s} f(\phi_z^{-1}(\eta, \theta))\bigg|_{\theta=0} = D_\eta^s f(z) = D^s_{\phi_x^{-1}(\xi, h+\pi/2)} f(\phi_x^{-1}(\xi, h)).$$

By the previous equations

$$\begin{aligned} \frac{\partial^s}{\partial\theta^s} f(\phi_x^{-1}(\xi, \theta))\bigg|_{\theta=h} &= \frac{\partial^s}{\partial\theta^s} f(\phi_x^{-1}(\xi, h + \theta))\bigg|_{\theta=0} = D^s_{\phi_x^{-1}(\xi, h+\pi/2)} f(\phi_x^{-1}(\xi, h)) \\ &= \sin^{-s}(\pi/2 - h) D_\xi^s f(\phi_x^{-1}(\xi, h)) = \cos^{-s} h D_\xi^s f(\phi_x^{-1}(\xi, h)), \end{aligned}$$

when $h \neq \pi/2, 3\pi/2$. This leads to

$$\begin{aligned}
\left.\frac{\partial^{s+1}}{\partial\theta^{s+1}} f(\phi_x^{-1}(\xi,\theta))\right|_{\theta=0} &= \lim_{h\to 0} h^{-1}\left[\left.\frac{\partial^s}{\partial\theta^s} f(\phi_x^{-1}(\xi,\theta))\right|_{\theta=h} - \left.\frac{\partial^s}{\partial\theta^s} f(\phi_x^{-1}(\xi,\theta))\right|_{\theta=0}\right] \\
&= \lim_{h\to 0} h^{-1}\left[\cos^{-s} h D_\xi^s f(\phi_x^{-1}(\xi,h)) - D_\xi^s f(x)\right] \\
&= \lim_{h\to 0} h^{-1}\left[D_\xi^s f(\phi_x^{-1}(\xi,h)) - D_\xi^s f(x)\right] \\
&= \lim_{h\to 0} h^{-1}\left[\overline{D_\xi^s f}\left(x + \frac{\sin h}{\cos h}\xi\right) - \overline{D_\xi^s f}(x)\right] \\
&= \lim_{h\to 0} h^{-1}\left[\overline{D_\xi^s f}\,(x+h\xi) - \overline{D_\xi^s f}(x)\right] = D_\xi^{s+1} f(x).
\end{aligned}$$

Thus the assertion has been proved for $s+1$. □

From Lemma 6.1 it follows that when $s \geq 1$ is odd, $D^s f(x) = 0$ for all $x \in S_d$. This is proved by first noting that $\phi_x^{-1}(-\xi,\theta) = \phi_x^{-1}(\xi,-\theta)$ and thus

$$D_{-\xi}^s f(x) = \left.\frac{\partial^s}{\partial\theta^s} f(\phi_x^{-1}(-\xi,\theta))\right|_{\theta=0} = (-1)^s \left.\frac{\partial^s}{\partial\theta^s} f(\phi_x^{-1}(\xi,\theta))\right|_{\theta=0} = (-1)^s D_\xi^s f(x). \tag{26}$$

Proof of Lemma 2.2. The case $s=0$ follows directly from formula (6). Let $s \geq 2$ be even, $\xi \in S_d$, $\xi \perp x$, and $\theta \in \mathbf{R}$. By Taylor's theorem,

$$\begin{aligned}
f\left(\phi_x^{-1}(\xi,\theta)\right) &= f(x) + \sum_{i=1}^{s-1} \frac{\theta^i}{i!} \left.\frac{\partial^i}{\partial\nu^i} f\left(\phi_x^{-1}(\xi,\nu)\right)\right|_{\nu=0} \\
&\quad + \frac{\theta^s}{(s-1)!}\int_0^1 (1-t)^{s-1} \left.\frac{\partial^s}{\partial\nu^s} f\left(\phi_x^{-1}(\xi,\nu)\right)\right|_{\nu=t\theta} dt.
\end{aligned}$$

Thus,

$$\begin{aligned}
\kappa^d f * L_\kappa(x) &= \kappa^d \int_0^\pi d\theta \sin^{d-1}\theta L(\kappa\theta) \int_{T_x} f\left(\phi_x^{-1}(\xi,\theta)\right) d\mu_{d-1}(\xi) \\
&= \sum_{i=0}^{s-1} \frac{1}{i!} d_i(\kappa) D^i f(x) + \frac{\kappa^d}{(s-1)!}\int_0^\pi d\theta\, \theta^s L(\kappa\theta)\sin^{d-1}\theta \\
&\quad \int_{T_x} d\mu_{d-1}(\xi) \int_0^1 (1-t)^{s-1} \left.\frac{\partial^s}{\partial\nu^s} f\left(\phi_x^{-1}(\xi,\nu)\right)\right|_{\nu=t\theta} dt,
\end{aligned}$$

where Lemma 6.1 was used. By equation (26), $D^i f(x) = 0$ when i is odd. By the substitutions $t\theta = \eta$ and $t^{-1} = \tau$,

$$\frac{\kappa^d}{(s-1)!}\int_0^\pi d\theta\, \theta^s L(\kappa\theta)\sin^{d-1}\theta$$

$$
\begin{aligned}
& \int_{T_x} d\mu_{d-1}(\xi) \int_0^1 (1-t)^{s-1} \left. \frac{\partial^s}{\partial \nu^s} f\left(\phi_x^{-1}(\xi,\nu)\right)\right|_{\nu=t\theta} dt \\
= \;& \frac{\kappa^d}{(s-1)!} \int_0^1 dt(1-t)^{s-1} \int_0^\pi \theta^s L(\kappa\theta)\sin^{d-1}\theta \\
& \int_{T_x} d\mu_{d-1}(\xi) D^s_{\phi_x^{-1}(\xi,t\theta+\pi/2)} f\left(\phi_x^{-1}(\xi,t\theta)\right) d\theta \\
= \;& \frac{\omega_{d-1}\kappa^d}{(s-1)!} \int_0^1 dt(1-t)^{s-1} \int_0^\pi \theta^s L(\kappa\theta)\sin^{d-1}\theta\, \tilde{D}^s f(x,t\theta) d\theta \\
= \;& \frac{\omega_{d-1}\kappa^d}{(s-1)!} \int_0^1 dt(1-t)^{s-1} t^{-s-1} \int_0^{t\pi} \eta^s L(\kappa\eta/t)\sin^{d-1}(\eta/t) \tilde{D}^s f(x,\eta) d\eta \\
= \;& \frac{\omega_{d-1}\kappa^d}{(s-1)!} \int_0^\pi d\eta\, \eta^s \tilde{D}^s f(x,\eta) \int_{\pi/\eta}^1 (1-t)^{s-1} t^{-s-1} L(\kappa\eta/t)\sin^{d-1}(\eta/t) dt \\
= \;& \frac{\omega_{d-1}\kappa^d}{(s-1)!} \int_0^\pi d\eta\, \eta^s \tilde{D}^s f(x,\eta) \int_1^{\pi/\eta} (\tau-1)^{s-1} L(\kappa\tau\eta)\sin^{d-1}(\tau\eta) d\tau.
\end{aligned}
$$

The assertion follows from the definition of $\tilde{L}^{(s)}_\kappa$. □

Proof of Lemma 2.3. We have that

$$
\frac{\sin x}{x} = \sum_{i=0}^\infty \frac{(-1)^i}{(2i+1)!} x^{2i}.
$$

Thus

$$
\left(\frac{\sin \kappa^{-1} t}{\kappa^{-1} t}\right)^{d-1} = \sum_{i=0}^r \gamma_i (\kappa^{-1}t)^{2i} + R_{r+1}(\kappa^{-1}t),
$$

where $\lim_{\kappa\to\infty}(\kappa^{-1}t)^{-2r} R_{r+1}(\kappa^{-1}t) = 0$ and $|(\kappa^{-1}t)^{-2r} R_{r+1}(\kappa^{-1}t)| \le M$, when $|t| \le \pi\kappa$, where $0 < M < \infty$. Thus, by the substitution $\kappa\theta = t$,

$$
\begin{aligned}
& \kappa^d \int_0^\pi \theta^m L(\kappa\theta)\sin^{d-1}\theta d\theta \\
& \quad = \kappa^d \int_0^{\pi\kappa} (\kappa^{-1}t)^m L(t)\sin^{d-1}(\kappa^{-1}t)\kappa^{-1} dt \\
& \quad = \sum_{i=0}^r \kappa^{-m-2i}\gamma_i \int_0^{\pi\kappa} t^{m+2i+d-1} L(t) dt \\
& \quad\quad + \kappa^{-m-2r} \int_0^{\pi\kappa} (\kappa^{-1}t)^{-2r} R_{r+1}(\kappa^{-1}t) t^{m+2r+d-1} L(t) dt
\end{aligned}
$$

and the assertion follows from the dominated convergence theorem. □

Proof of Theorem 2.5. **(i)** The statement needs to be proved only for continuous functions, because if f is not continuous, for every $\epsilon > 0$ there is such continuous

g that $\|f-g\|_p \le \epsilon$ (Rudin, 1986, Theorem 3.14) and

$$\|c(\kappa)f * L_\kappa - f\|_p \le \|c(\kappa)(f-g) * L_\kappa\|_p + \|f-g\|_p + \|c(\kappa)g * L_\kappa - g\|_p .$$

Also, by the generalized Minkowski inequality (Hoffmann–Jørgensen, 1994, page 240) and Lemma 2.3,

$$\begin{aligned}\|c(\kappa)(f-g) * L_\kappa\|_p &\le \|f-g\|_p c(\kappa)\|L_\kappa(\cdot'\eta))\|_1 = \|f-g\|_p c(\kappa, L)c(\kappa, |L|)^{-1} \\ &= \|f-g\|_p(\alpha_0(L)^{-1}\alpha_0(|L|) + o(1)).\end{aligned}$$

From now on, let f be continuous. For $x, y \in S_d$, let $R_{x,y}$ be such a rotation that $R_{x,y}x = y$. Let $\eta \in S_d$, $\delta > 0$ and define

$$\Lambda_p(\delta) = \sup_{\{z \,|\, \|z-\eta\| \le \delta\}} \left\{ \int_{S_d} |f(R_{\eta,x}z) - f(x)|^p d\mu(x) \right\}^{1/p} .$$

It holds that $\lim_{\delta\to 0} \Lambda_p(\delta) = 0$, because $\lim_{z\to e} \|R_{\eta,x}z - x\| = 0$. By the transformation $y = R_{\eta,x}z$ (so that $x'y = x'(R_{\eta,x}z) = z'R_{x,\eta}x = z'e$) and by the generalized Minkowski inequality,

$$\begin{aligned}&\left\{ \int_{S_d} \left| c(\kappa) \int_{S_d} L(\kappa^2(1-x'y))f(y)d\mu(y) - f(x) \right|^p d\mu(x) \right\}^{1/p} \\ &= \left\{ \int_{S_d} \left| c(\kappa) \int_{S_d} L(\kappa^2(1-x'y))(f(y)-f(x))d\mu(y) \right|^p d\mu(x) \right\}^{1/p} \\ &= \left\{ \int_{S_d} \left| c(\kappa) \int_{S_d} L(\kappa^2(1-z'\eta))(f(R_{\eta,x}z)-f(x))d\mu(z) \right|^p d\mu(x) \right\}^{1/p} \\ &\le c(\kappa) \int_{S_d} \left| L(\kappa^2(1-z'\eta)) \right| \left\{ \int_{S_d} |f(R_{\eta,x}z) - f(x)|^p \, d\mu(x) \right\}^{1/p} d\mu(z) \\ &\le \Lambda_p(\delta) d_0(\kappa, |L|) + 2 \sup_{x\in S_d} |f(x)| \, \omega_d^{1/p} c(\kappa) \int_{\{z \,|\, \|z-\eta\| > \delta\}} |L(\kappa^2(1-z'\eta))| d\mu(z).\end{aligned}$$

The assertion follows from (11) and the fact that for every $\delta > 0$, $x \in S_d$,

$$\lim_{\kappa\to\infty} \kappa^d \int_{\|y-x\|>\delta} L(\kappa^2(1-x'y))d\mu(y) = 0.$$

(ii) By Lemma 2.3 ($m = 2i$, $r = s/2 - i$) and the substitution $i + j = l$,

$$\sum_{i=1}^{s/2-1} \frac{1}{(2i)!} d_{2i}(\kappa) D^{2i} f(x)$$

$$
\begin{aligned}
&= \sum_{i=1}^{s/2-1} \frac{1}{(2i)!} D^{2i} f(x) \left\{ \omega_{d-1} \sum_{j=0}^{s/2-i} \kappa^{-2i-2j} \gamma_j \alpha_{2i+2j} + o(\kappa^{-s}) \right\} \\
&= \sum_{i=1}^{s/2-1} \frac{1}{(2i)!} D^{2i} f(x) \omega_{d-1} \sum_{l=i}^{s/2} \kappa^{-2l} \gamma_{l-i} \alpha_{2l} + o(\kappa^{-s}) \sum_{i=1}^{s/2-1} D^{2i} f(x) \\
&= \omega_{d-1} \sum_{l=1}^{s/2-1} \kappa^{-2l} \alpha_{2l} \sum_{i=1}^{l} \frac{1}{(2i)!} \gamma_{l-i} D^{2i} f(x) \\
&\quad + \omega_{d-1} \kappa^{-s} \alpha_s \sum_{i=1}^{s/2-1} \frac{1}{(2i)!} \gamma_{s/2-i} D^{2i} f(x) + o(\kappa^{-s}) \sum_{i=1}^{s/2-1} D^{2i} f(x).
\end{aligned}
$$

Thus, by Lemma 2.2 ($c(\kappa) = \kappa^d d_0(\kappa)^{-1}$),

$$
\begin{aligned}
c(\kappa) f * L_\kappa(x) &= f(x) + d_0(\kappa)^{-1} \Bigg\{ \omega_{d-1} \kappa^{-s} \alpha_s \sum_{i=1}^{s/2-1} \frac{1}{(2i)!} \gamma_{s/2-1} D^{2i} f(x) \\
&\quad + \int_0^\pi \tilde{L}_\kappa^{(s)}(\theta) \tilde{D}^s f(x,\theta) d\theta + o(\kappa^{-s}) \sum_{i=1}^{s/2-1} D^{2i} f(x) \Bigg\},
\end{aligned}
$$

where $o(\kappa^{-s})$ does not depend on f. Thus,

$$
\begin{aligned}
&\left| \kappa^s \left| c(\kappa) f * L_\kappa(x) - f(x) \right| - \left| \alpha_0^{-1} \alpha_s \mathcal{D}^s f(x) \right| \right| \\
&= \Bigg| |d_0(\kappa)|^{-1} \Bigg| \omega_{d-1} \alpha_s \sum_{i=1}^{s/2-1} \frac{1}{(2i)!} \gamma_{s/2-i} D^{2i} f(x) + \kappa^s \int_0^\pi \tilde{L}_\kappa^{(s)}(\theta) \tilde{D}^s f(x,\theta) d\theta \\
&\quad + o(1) \sum_{i=1}^{s/2-1} D^{2i} f(x) \Bigg| - \left| \alpha_0^{-1} \alpha_s \sum_{i=1}^{s/2} \frac{1}{(2i)!} \gamma_{s/2-i} D^{2i} f(x) \right| \Bigg| \\
&\le \left| d_0(\kappa)^{-1} \omega_{d-1} - \alpha_0^{-1} \right| \left| \alpha_s \sum_{i=1}^{s/2-1} \frac{1}{(2i)!} \gamma_{s/2-i} D^{2i} f(x) \right| \\
&\quad + \left| d_0(\kappa)^{-1} \kappa^s \int_0^\pi \tilde{L}_\kappa^{(s)}(\theta) \tilde{D}^s f(x,\theta) d\theta - \alpha_0^{-1} \alpha_s \frac{1}{s!} D^s f(x) \right| \\
&\quad + \left| d_0(\kappa)^{-1} o(1) \sum_{i=1}^{s/2-1} D^{2i} f(x) \right|.
\end{aligned}
$$

By equation (11), $\lim_{\kappa\to\infty} \left| d_0(\kappa)^{-1} \omega_{d-1} - \alpha_0^{-1} \right| = 0$. Also,

$$
\begin{aligned}
&\left| d_0(\kappa)^{-1} \kappa^s \int_0^\pi \tilde{L}_\kappa^{(s)}(\theta) \tilde{D}^s f(x,\theta) d\theta - \alpha_0^{-1} \alpha_s \frac{1}{s!} D^s f(x) \right| \\
&\le |d_0(\kappa)|^{-1} \kappa^s \left| \int_0^\pi \tilde{L}_\kappa^{(s)}(\theta) \tilde{D}^s f(x,\theta) d\theta - \frac{1}{s!} d_s(\kappa) D^s f(x) \right| \\
&\quad + \left| d_0(\kappa)^{-1} \kappa^s d_s(\kappa) - \alpha_0^{-1} \alpha_s \right| \frac{1}{s!} |D^s f(x)|.
\end{aligned}
$$

The L_p norm of the second term in the right hand side converges to zero because, by Lemma 2.3, $d_0(\kappa) = \omega_{d-1}\alpha_0 + o(1)$ and $d_s(\kappa) = \omega_{d-1}\kappa^{-s}\alpha_s + o(\kappa^{-s})$ (Choose $m = s$, $r = 0$). The L_p norm of the first term in the right hand side converges to zero because by equation (23) and the generalized Minkowski inequality,

$$\begin{aligned}\left\|\int_0^\pi \tilde{L}_\kappa^{(s)}(\theta)\tilde{D}^s f(\cdot,\theta)d\theta - \frac{1}{s!}d_s(\kappa)D^s f\right\|_p &= \left\|\int_0^\pi \tilde{L}_\kappa^{(s)}(\theta)\left(\tilde{D}^s f(\cdot,\theta) - D^s f\right)d\theta\right\|_p \\ &\leq \int_0^\pi \left|\tilde{L}_\kappa^{(s)}(\theta)\right| \left\|\tilde{D}^s f(\cdot,\theta) - D^s f\right\|_p d\theta.\end{aligned}$$

This upper bound converges to zero as is seen from equation (24). □

7 Proof of Theorem 3.1

When $x, \xi \in S_d$, $\xi \perp x$, $\xi = (\xi_1, \ldots, \xi_{d+1})'$, then

$$D_\xi^2 f(x) = \sum_{i,j=1}^{d+1} \xi_i\xi_j \frac{\partial^2}{\partial x_i \partial x_j}\bar{f}(x) = \xi' \mathrm{Hess}\bar{f}(x)\xi = \mathrm{tr}(\mathrm{Hess}\bar{f}(x)\xi\xi'),$$

where $\mathrm{Hess}\bar{f}(x)$ is a $(d+1)\times(d+1)$–matrix with elements

$$\left[\mathrm{Hess}\bar{f}(x)\right]_{i,j} = \frac{\partial^2}{\partial x_i \partial x_j}\bar{f}(x),$$

$i, j = 1, \ldots, d+1$. It can be seen after some calculations that

$$\int_{T_x} \xi\xi' d\mu_{d-1}(\xi) = \omega_{d-1} d^{-1}(I_{d+1} - xx').$$

Thus,

$$\begin{aligned} D^2 f(x) &= \mathrm{tr}\left(\mathrm{Hess}\bar{f}(x)\omega_{d-1}^{-1}\int_{T_x}\xi\xi' d\mu_{d-1}(x)\right) = \mathrm{tr}\left(\mathrm{Hess}\bar{f}(x)d^{-1}(I_{d+1} - xx')\right) \\ &= d^{-1}\left[\sum_{i=1}^{d+1}\frac{\partial^2}{\partial x_i^2}\bar{f}(x) - x'\mathrm{Hess}\bar{f}(x)x\right].\end{aligned}$$

Put $\check{f}(\theta) = f(\phi^{-1}(\theta))$, where $\theta \in]0, 2\pi[\times]0,\pi[^{d-1}$ and $\phi = \phi_{\eta^1,\ldots,\eta^{d+1}} = (\phi_1, \ldots, \phi_d)'$ was defined in equation (5). By the chain rule,

$$\begin{aligned}\frac{\partial^2}{\partial x_i \partial x_j}\bar{f}(x) &= \frac{\partial^2}{\partial x_i \partial x_j}\check{f}\left(\phi\left(\frac{x}{\|x\|}\right)\right) \\ &= \sum_{k,l=1}^d \frac{\partial^2}{\partial\theta_k\partial\theta_l}\check{f}(\theta)\bigg|_{\theta=\phi(x/\|x\|)} \frac{\partial}{\partial x_i}\phi_k\left(\frac{x}{\|x\|}\right)\frac{\partial}{\partial x_j}\phi_l\left(\frac{x}{\|x\|}\right) \\ &\quad + \sum_{k=1}^d \frac{\partial}{\partial\theta_k}\check{f}(\theta)\bigg|_{\theta=\phi(x/\|x\|)} \frac{\partial^2}{\partial x_i\partial x_j}\phi_k\left(\frac{x}{\|x\|}\right).\end{aligned}$$

Thus,

$$
\begin{aligned}
&\sum_{i=1}^{d+1} \frac{\partial^2}{\partial x_i^2} \bar{f}(x) \\
&= \sum_{k,l=1}^{d} \frac{\partial^2}{\partial\theta_k \partial\theta_l} \check{f}(\theta)\Big|_{\theta=\phi(x/\|x\|)} \sum_{i=1}^{d+1} \frac{\partial}{\partial x_i}\phi_k\left(\frac{x}{\|x\|}\right) \frac{\partial}{\partial x_i}\phi_l\left(\frac{x}{\|x\|}\right) \\
&\quad + \sum_{k=1}^{d} \frac{\partial}{\partial\theta_k} \check{f}(\theta)\Big|_{\theta=\phi(x/\|x\|)} \sum_{i=1}^{d+1} \frac{\partial^2}{\partial x_i^2}\phi_k\left(\frac{x}{\|x\|}\right) \\
&= \sum_{k=1}^{d} \left(\sum_{i=1}^{k+1} x_i^2\right)^{-1} \left[\frac{\partial^2}{\partial\theta_k^2} \check{f}(\theta)\Big|_{\theta=\phi(x/\|x\|)} + (k-1)x_{k+1}\left(\sum_{i=1}^{k} x_i^2\right)^{-1/2} \frac{\partial}{\partial\theta_k}\check{f}(\theta)\Big|_{\theta=\phi(x/\|x\|)}\right] \\
&= \sum_{k=1}^{d} \sin^{-2}\phi_{k+1}(x)\cdots\sin^{-2}\phi_d(x) \left[\frac{\partial^2}{\partial\theta_k^2} + (k-1)\frac{\cos\phi_k(x)}{\sin\phi_k(x)}\frac{\partial}{\partial\theta_k}\right] \check{f}(\theta)\Big|_{\theta=\phi(x/\|x\|)} \\
&= \Delta f(x),
\end{aligned}
$$

because

$$
\sum_{i=1}^{d+1} \frac{\partial}{\partial x_i}\phi_k\left(\frac{x}{\|x\|}\right) \frac{\partial}{\partial x_i}\phi_l\left(\frac{x}{\|x\|}\right) = \begin{cases} \left(\sum_{i=1}^{k+1} x_i^2\right)^{-1} & ,1 \le k = l \le d \\ 0 & ,1 \le k \ne l \le d \end{cases}
$$

and

$$
\sum_{i=1}^{d+1} \frac{\partial^2}{\partial x_i^2}\phi_k\left(\frac{x}{\|x\|}\right) = (k-1)x_{k+1}\left(\sum_{i=1}^{k} x_i^2\right)^{-1/2}\left(\sum_{i=1}^{k+1} x_i^2\right)^{-1},\ k = 1, \ldots, d,
$$

as can be seen after some calculations. Also,

$$
\begin{aligned}
x'\mathrm{Hess}\bar{f}(x)x &= \sum_{i,j=1}^{d+1} x_i x_j \frac{\partial^2}{\partial x_i \partial x_j}\bar{f}(x) \\
&= \sum_{k,l=1}^{d} \frac{\partial^2}{\partial\theta_k \partial\theta_l} \check{f}(\theta)\Big|_{\theta=\phi(x/\|x\|)} \sum_{i,j=1}^{d+1} x_i x_j \frac{\partial}{\partial x_i}\phi_k\left(\frac{x}{\|x\|}\right) \frac{\partial}{\partial x_j}\phi_l\left(\frac{x}{\|x\|}\right) \\
&\quad + \sum_{k=1}^{d} \frac{\partial}{\partial\theta_k} \check{f}(\theta)\Big|_{\theta=\phi(x/\|x\|)} \sum_{i,j=1}^{d+1} x_i x_j \frac{\partial^2}{\partial x_i \partial x_j}\phi_k\left(\frac{x}{\|x\|}\right) \\
&= 0,
\end{aligned}
$$

because

$$
\sum_{i=1}^{d+1} x_i x_j \frac{\partial}{\partial x_i}\phi_k\left(\frac{x}{\|x\|}\right) \frac{\partial}{\partial x_j}\phi_l\left(\frac{x}{\|x\|}\right) = \sum_{i,j=1}^{d+1} x_i x_j \frac{\partial^2}{\partial x_i \partial x_j}\phi_k\left(\frac{x}{\|x\|}\right) = 0,
$$

as can be seen after some calculations. □

8 Proof of Lemma 4.1

We have that

$$\begin{aligned}\operatorname{Var}\left(\hat{f}_n(x)\right) &= n^{-1}c(\kappa)^2\operatorname{Var}\left[L(\kappa\arccos(x'X_1))\right]\\ &= n^{-1}\kappa^{2d}d_0(\kappa,L)^{-2}\left\{\mathrm{E}L^2(\kappa\arccos(x'X_1)) - \left[\mathrm{E}L(\kappa\arccos(x'X_1))\right]^2\right\}\\ &= n^{-1}\kappa^{2d}d_0(\kappa,L)^{-2}\left\{f*\left(L^2\right)_\kappa(x) - (f*L_\kappa(x))^2\right\}.\end{aligned}$$

Write $\operatorname{Var}(\hat{f}_n) = V + R$, where

$$V = \frac{\kappa^d}{n}\frac{d_0(\kappa,L^2)}{d_0(\kappa,L)^2}f$$

and

$$R = \frac{\kappa^d}{n}\frac{1}{d_0(\kappa,L)}\left\{\kappa^d f*(L^2)_\kappa - d_0(\kappa,L^2)f - \kappa^d\left(f*L_\kappa\right)^2\right\}.$$

(i) By Lemma 2.5 (i),

$$\kappa^d\int_{S_d} f*\left(L^2\right)_\kappa d\mu = d_0(\kappa,L^2)\int_{S_d} f\,d\mu + o(1).$$

By equation (11),

$$d_0(\kappa,L^i) = \omega_{d-1}\alpha_0(L^i) + o(1) \tag{27}$$

for $i = 1,2$. By Lemma 2.5 (i) and equation (27),

$$\kappa^d\int_{S_d}(f*L_\kappa)^2 = \kappa^{-d}\left[d_0(\kappa,L)^2\int_{S_d} f^2 d\mu + o(1)\right] = o(1).$$

Thus, $n\kappa^{-d}\int_{S_d} R\,d\mu = o(1)$. The assertion follows, because by equation (27),

$$\left|\int_{S_d} V\,d\mu - \frac{\kappa^d}{n}\frac{\alpha_0(L^2)}{\omega_{d-1}\alpha_0(L)^2}\right| = \frac{o(\kappa^d)}{n}.$$

(ii) Denote $\sigma_n = \sqrt{\operatorname{Var}(\hat{f}_n)}$. Firstly, $\sigma_n = \sqrt{V+R} \le \sqrt{V+|R|} \le \sqrt{V} + \sqrt{|R|}$ and secondly, $\sqrt{V} \le \sqrt{V+R+|R|} \le \sqrt{V+R} + \sqrt{|R|} = \sigma_n + \sqrt{|R|}$. Thus $|\sigma_n - \sqrt{V}| \le \sqrt{|R|}$ and by the Schwarz inequality, equation (27), and Lemma 2.5,

$$\begin{aligned}&n^{1/2}\kappa^{-d/2}\int_{S_d}\sqrt{|R|}d\mu\\ &= |d_0(\kappa,L)|^{-1}\int_{S_d}\left|\kappa^d f*\left(L^2\right)_\kappa - d_0(\kappa,L^2)f - \kappa^d\left(f*L_\kappa\right)^2\right|^{1/2}d\mu\end{aligned}$$

$$
\begin{aligned}
&\leq \ |d_0(\kappa, L)|^{-1} \left\{ \int_{S_d} \left| \kappa^d f * \left(L^2\right)_\kappa - d_0(\kappa, L^2) f \right|^{1/2} d\mu + \kappa^{d/2} \int_{S_d} |f * L_\kappa| \, d\mu \right\} \\
&\leq \ |d_0(\kappa, L)|^{-1} \omega_d^{1/2} \left\{ \int_{S_d} \left| \kappa^d f * \left(L^2\right)_\kappa - d_0(\kappa, L^2) f \right| d\mu \right\}^{1/2} + O\left(\kappa^{-d/2}\right) \\
&= \ o(1),
\end{aligned}
$$

when $\kappa \to \infty$. The assertion follows, because by equation (27),

$$
\int_{S_d} \left| \sqrt{V} - \sqrt{\frac{\kappa^d}{n} \frac{\alpha_0(L^2)}{\omega_{d-1}\alpha_0(L)^2} f} \right| d\mu = \frac{o(\kappa^{d/2})}{n^{1/2}}.
$$

□

9 The Proofs for Section 5

Proof of Theorem 5.1. Let us denote $\mu_{s,n} = \mathrm{E}(\hat{\theta}_{s,n})$. Put

$$
C_n = (A(L)(\mu_{s,n} + b_n))^{1/(2s+d)} \wedge n^\gamma \tag{28}
$$

and $\lambda_n = C_n n^{1/(2s+d)}$. From assumption (19) it follows that $\lim_{n\to\infty} C_n = C^*$. Thus, from equation (15) it is seen that

$$
\mathrm{E} \int_{S_d} \left(\hat{f}_n(\cdot, \kappa_n^*) - f \right)^2 d\mu \sim \mathrm{E} \int_{S_d} \left(\hat{f}_n(\cdot, \lambda_n) - f \right)^2 d\mu
$$

and it remains to prove

$$
\mathrm{E} \int_{S_d} \left(\hat{f}_n(\cdot, \hat{\kappa}_n) - \hat{f}_n(\cdot, \lambda_n) \right)^2 d\mu = o\left(n^{-2s/(2s+d)}\right).
$$

Now

$$
\begin{aligned}
\mathrm{E} \int_{S_d} \left(\hat{f}_n(\cdot, \hat{\kappa}_n) - \hat{f}_n(\cdot, \lambda_n) \right)^2 d\mu &= \mathrm{E} \int_{S_d} \left[\frac{\partial}{\partial\kappa} \hat{f}_n(x, \kappa) \Big|_{\kappa=\xi_n} (\hat{\kappa}_n - \lambda_n) \right]^2 d\mu(x) \\
&\leq \mathrm{E}^{1/2} \left[\int_{S_d} \left[n^{1/(2s+d)} \frac{\partial}{\partial\kappa} \hat{f}_n(x, \kappa) \Big|_{\kappa=\xi_n} \right]^2 d\mu(x) \right]^2 \mathrm{E}^{1/2} \left(\hat{C}_n - C_n \right)^4,
\end{aligned}
$$

where ξ_n is between $\hat{\kappa}_n$ and λ_n with probability one.

Next it is proved that

$$
\mathrm{E}^{1/2} \left(\hat{C}_n - C_n \right)^4 = o\left(n^{-2s/(2s+d)}\right). \tag{29}
$$

If $\mathrm{E}|X_n|^m = o(a_n^m)$ for $m = 1, 2, \ldots$, then it is denoted $X_n = o_E(a_n)$. The following lemma is proved like the corresponding assertion in Woodroofe (1970).

Lemma 9.1 *Let $X_n \geq b_n > 0$ and $Y_n \geq \delta > 0$ with probability one for sufficiently large n, where $b_n = o(1)$. If $X_n - Y_n = o_E(a_n)$ and $a_n^r = o(b_n)$ for some $r > 0$, then $X_n^k - Y_n^k = o_E(a_n)$ for $k \leq 1$.*

Let us denote $a_n = n^{-s/(2s+d)}$. It is proved that

$$\hat{C}_n - C_n = o_E(a_n). \tag{30}$$

Equation (29) follows from this. From assumption (21) it follows that $\hat{\theta}_{s,n} + b_n \geq b_n$ for sufficiently large n with probability one. From assumption (19) it follows that $\mu_{s,n} + b_n \geq \delta$ for sufficiently large n and for a certain $\delta > 0$. Assumption (20) says that $\hat{\theta}_{s,n} - \mu_{s,n} = o_E(a_n)$. It holds also that $a_n^r = o(b_n)$, when $r \geq (2s+d)/s$. Thus, by Lemma 9.1 it holds that

$$\left(\hat{\theta}_{s,n} + b_n\right)^{1/(2s+d)} - (\mu_{s,n} + b_n)^{1/(2s+d)} = o_E(a_n)$$

and equation (30) follows.

It remains to prove

$$E\left\{\int_{S_d}\left[n^{1/(2s+d)}\, \frac{\partial}{\partial\kappa}\hat{f}_n(x,\kappa)\bigg|_{\kappa=\xi_n}\right]^2 d\mu(x)\right\}^2 = O(1), \tag{31}$$

where ξ_n is between $\hat{\kappa}_n$ and λ_n with probability one. First note that

$$c'(\kappa) \sim Q(L)\kappa^{d-1}$$

as $\kappa \to \infty$, where $Q(L) = -\omega_{d-1}^{-1}\alpha_2(L')\alpha_0(L)^{-2}$ and $c(\kappa)^{-1} = \int_{S_d} L(\kappa \arccos(x'y))d\mu(y)$. This holds because by equation (11), $c(\kappa) = \kappa^d d_0(\kappa, L)^{-1} \sim \kappa^d R(L)$, where $R(L) = [\omega_{d-1}\alpha_0(L)]^{-1}$ and

$$c'(\kappa) = -c^2(\kappa)\frac{\partial}{\partial\kappa}\left[\int_{S_d} L(\kappa \arccos(x'y))d\mu(y)\right],$$

where

$$\begin{aligned}
\frac{\partial}{\partial\kappa}\left[\int_{S_d} L(\kappa \arccos(x'y))d\mu(y)\right] &= \int_{S_d} \arccos(x'y)L'(\kappa \arccos(x'y))d\mu(y) \\
&= \omega_{d-1}\kappa^{-2}\int_0^{\pi\kappa} tL'(t)\sin^{d-1}(\kappa^{-1}t)dt \\
&\sim \omega_{d-1}\kappa^{-d-1}\int_0^{\infty} t^d L'(t)dt.
\end{aligned}$$

It holds that

$$\frac{\partial}{\partial \kappa}\hat{f}_n(x,\kappa) = \frac{1}{n}\sum_{i=1}^{n}\left[c'(\kappa)L(\kappa\arccos(x'X_i)) + c(\kappa)\arccos(x'X_i)L'(\kappa\arccos(x'X_i))\right].$$

Let us denote $A_n = (C^*/2 \le \hat{C}_n \le 2C^*)$, $s_{1n} = (C^*/2)n^{1/(d+4)}$ and $s_{2n} = 2C^* n^{1/(d+4)}$. Let $\epsilon > 0$. In A_n, for sufficiently large n,

$$\begin{aligned}
&\left|\frac{\partial}{\partial \kappa}\hat{f}_n(x,\kappa)\Big|_{\kappa=\xi_n}\right| \\
\le\ & \left(|c'(\xi_n)| + |c(\xi_n)\xi_n^{-1}|\right)\frac{1}{n}\sum_{i=1}^{n} J(\xi_n\arccos(x'X_i)) \\
\le\ & \left(|Q(L)| + |R(L)|\right)(1+\epsilon)\frac{\xi_n^{d-1}}{n}\sum_{i=1}^{n} J(\xi_n\arccos(x'X_i)) \\
\le\ & \left(|Q(L)| + |R(L)|\right)(1+\epsilon)\frac{s_{2n}^{d-1}}{n}\sum_{i=1}^{n} J(s_{1n}\arccos(x'X_i)) \\
\le\ & \left(|Q(L)| + |R(L)|\right)(1+\epsilon)(C^*)^{-1}2^{2d-1}n^{-1/(2s+d)}\frac{s_{1n}^{d}}{n}\sum_{i=1}^{n} J(s_{1n}\arccos(x'X_i)).
\end{aligned}$$

Thus, for sufficiently large n,

$$\begin{aligned}
&\mathrm{E}\left\{I_{A_n}\int_{S_d}\left[n^{1/(2s+d)}\ \frac{\partial}{\partial \kappa}\hat{f}_n(x,\kappa)\Big|_{\kappa=\xi_n}\right]^2 d\mu(x)\right\}^2 \\
\le\ & \left((|Q(L)| + |R(L)|)(1+\epsilon)2^{2d-1}(C^*)^{-1}\right)^4 \mathrm{E}\left\{\int_{S_d}\left[\frac{s_{1n}^d}{n}\sum_{i=1}^{n} J(s_{1n}\arccos(x'X_i))\right]^2 d\mu(x)\right\}^2.
\end{aligned}$$

By Jensen's inequality and by Lemma 9.2, which is given after this proof,

$$\begin{aligned}
&\mathrm{E}\left\{\int_{S_d}\left[\frac{s_{1n}^d}{n}\sum_{i=1}^{n} J(s_{1n}\arccos(x'X_i))\right]^2 d\mu(x)\right\}^2 \\
\le\ & \omega_d\int_{S_d}\mathrm{E}\left[\frac{s_{1n}^d}{n}\sum_{i=1}^{n} J(s_{1n}\arccos(x'X_i))\right]^4 d\mu(x) = O\left(\left(s_{1n}^d n^{-1}\right)^2\right) + O(1) = O(1)
\end{aligned}$$

Thus,

$$\mathrm{E}\left\{I_{A_n}\int_{S_d}\left[n^{1/(2s+d)}\ \frac{\partial}{\partial \kappa}\hat{f}_n(x,\kappa)\Big|_{\kappa=\xi_n}\right]^2 d\mu(x)\right\}^2 = O(1). \tag{32}$$

With probability one, for sufficiently large n,

$$\hat{\kappa}_n = n^{1/(2s+d)}\hat{C}_n \geq n^{1/(2s+d)}\left((A(L)b_n)^{1/(2s+d)} \wedge n^\gamma\right) \to \infty$$

and thus $\xi_n \to \infty$ with probability one. Also, with probability one, for sufficiently large n, $\xi_n \leq n^\gamma n^{1/(2s+d)}$. Thus in A_n^c, for sufficiently large n, when M is a bound for $|J|$,

$$\begin{aligned}\left|\frac{\partial}{\partial\kappa}\hat{f}_n(x_0,\kappa)\Big|_{\kappa=\xi_n}\right| &\leq M\left(|c'(\xi_n)| + \left|c(\xi_n)\xi_n^{-1}\right|\right)\\ &\leq M\left(|Q(L)| + |R(L)|\right)(1+\epsilon)\xi_n^{d-1}\\ &\leq M\left(|Q(L)| + |R(L)|\right)(1+\epsilon)n^{\gamma(d-1)}n^{(d-1)/(2s+d)}.\end{aligned}$$

Thus, for sufficiently large n,

$$\begin{aligned}&E\left\{I_{A_n^c}\int_{S_d}\left[n^{1/(2s+d)}\,\frac{\partial}{\partial\kappa}\hat{f}_n(x,\kappa)\Big|_{\kappa=\xi_n}\right]^2 d\mu(x)\right\}^2 \qquad (33)\\ &\quad\leq \left(M(|Q(L)|+|R(L)|)(1+\epsilon)n^{\gamma(d-1)}n^{d/(2s+d)}\right)^4\omega_d^2 P(A_n^c) = o(1),\end{aligned}$$

because, for sufficiently large n,

$$\begin{aligned}P(A_n^c) &= P((\hat{C}_n < C^*/2)\cup(\hat{C}_n > 2C^*)) \leq P(|\hat{C}_n - C_n| > C^*/4)\\ &\leq (C^*/4)^{-m}\mathrm{E}|\hat{C}_n - C_n|^m = o\left(n^{-ms/(2s+d)}\right)\end{aligned}$$

for $m \in \{1,2,\ldots\}$ by equation (30). Equation (31) follows from equations (32) and (33). □

The proof of Nadaraya (1974) differs from the proof given here in that he defines the constant C_n in equation (28) by an Euclidean equivalent of $C_n^{2s+d} = A(L)\left[\int_{S_d}\mu_{s,n}^2 d\mu + b_n\right]$, where $\mu_{s,n} = \mathrm{E}\,\mathcal{D}^s\hat{f}_n$ and $\hat{f}_n$ is a kernel estimator. The following lemma was needed in the previous proof.

Lemma 9.2 *Let $X_1,\ldots,X_n$ be i.i.d with a bounded density. Let $L : [0,\infty[\to \mathbf{R}$ be bounded and let $\alpha_0(|L|^2) < \infty$. Let $\{\kappa_n\}$ be such a sequence that $\lim_{n\to\infty}\kappa_n = \infty$ and $\lim_{n\to\infty}\kappa_n^d n^{-1} = 0$. Put $Z_{ni} = Z_{ni}(x) = \kappa_n^d L(\kappa_n \arccos(x'X_i))$, $i = 1,\ldots,n$. Then for $m \geq 4$ even,*

$$\sup_{x\in S_d}\mathrm{E}\left|\frac{1}{n}\sum_{i=1}^n Z_{ni}(x) - \mathrm{E}Z_{n1}(x)\right|^m = O\left(\left(\frac{\kappa_n^d}{n}\right)^{m/2}\right).$$

Proof. Write

$$\mathrm{E}\left|\frac{1}{n}\sum_{i=1}^{n} Z_{ni} - \mathrm{E}Z_{n1}\right|^{m} = \left(\frac{\mathrm{Var}Z_{n1}}{n}\right)^{m/2} \mathrm{E}\left|n^{-1/2}\sum_{i=1}^{n}\frac{Z_{ni}-\mathrm{E}Z_{n1}}{(\mathrm{Var}Z_{n1})^{1/2}}\right|^{m}.$$

Define $Y_{ni} = (Z_{ni} - \mathrm{E}Z_{n1})/(\mathrm{Var}Z_{n1})^{1/2}$. Applying results of Bahr (1965) and Petrov (1962) as in Nadaraya (1974, formula (21)), we have for $m \geq 4$ even,

$$\mathrm{E}\left|n^{-1/2}\sum_{i=1}^{n} Y_{ni}\right|^{m} = \int u^m \phi(u)du + \sum_{j=1}^{m-2} n^{-j/2}\int u^m p_{nj}(-\phi)(u)du,$$

where

$$p_{nj}(-\phi)(u) = \phi(u) \sum_{k_1+2k_2+\cdots+jk_j=j} \frac{1}{k_1!\cdots k_j!}\left(\frac{\gamma_{n3}}{3!}\right)^{k_1}\cdots\left(\frac{\gamma_{n,j+2}}{(j+2)!}\right)^{k_j} H_{3k_1+\cdots+(j+2)k_j}(u),$$

where ϕ is the density of the standard normal distribution, H_k is Hermite's polynomial of degree k and γ_{nj} is the jth order cumulant of Y_{n1}. By Gnedenko and Kolmogorov (1954, page 66),

$$|\gamma_{nj}| \leq j^j \mathrm{E}|Y_{n1}|^j \leq j^j\left(\frac{2M\kappa_n^d}{(\mathrm{Var}Z_{n1})^{1/2}}\right)^{j-2}\mathrm{E}|Y_{n1}|^2 = j^j\left(\frac{2M\kappa_n^d}{(\mathrm{Var}Z_{n1})^{1/2}}\right)^{j-2},$$

where M is a bound for $|L|$. Thus

$$\mathrm{E}\left|n^{-1/2}\sum_{i=1}^{n} Y_{ni}\right|^{m} \leq \int u^m\phi(u)du + \sum_{j=1}^{m-2} n^{-j/2}C(j,m)\left(\frac{\kappa_n^d}{(\mathrm{Var}Z_{n1})^{1/2}}\right)^{j},$$

where $C(j,m)$ is a constant. The assertion follows because

$$\begin{aligned}\mathrm{Var}(Z_{n1}) &\leq \mathrm{E}(Z_{n1}^2) = \kappa_n^{2d}\int_{S_d} L^2(\kappa_n \arccos(x'y))f(y)d\mu(y)\\ &\leq \kappa_n^d d_0(\kappa_n, L^2)\|f\|_\infty \leq \kappa_n^d\omega_{d-1}\alpha_0(L^2)\|f\|_\infty.\end{aligned}$$

□

Acknowledgments

This paper is based on part of the author's doctoral dissertation, which was written at the University of Helsinki under the supervision of Dr. Lasse Holmström. His support is greatly acknowledged.

References

[1] Abramowitz, M. and Stegun, I. (eds.) (1965). *Handbook of Mathematical Functions.* Ninth Edition. Dover Publications, Inc., New York.

[2] von Bahr, B. (1965). On the convergence of moments in the central limit theorem. *Ann. Math. Statist.* **36** 808–818.

[3] Bai, Z. D., Radhakrishna Rao, C., and Zhao, L. C. (1988). Kernel estimators of density function of directional data. *J. Multivariate Anal.* **27** 24–39.

[4] Bartlett, M. S. (1963). Statistical estimation of density functions. *Sankhyā Ser. A* **25** 245–254.

[5] Beran, R. (1979). Exponential models for directional data. *Ann. of Statist.* **7** 1162–1178.

[6] Butzer, P. L. and Nessel, R. J. (1971). *Fourier Analysis and Approximation.* Birkhäuser, Basel.

[7] Cao, R., Cuevas, A., and González-Manteiga, W. (1994). A comparative study of several smoothing methods in density estimation. *Comp. Statist. Data Anal.* **17** 153–176.

[8] Devroye, L. (1987). *A Course in Density Estimation.* Birkhäuser, Boston.

[9] Devroye, L. and Györfi, L. (1985). *Density Estimation: The L^1 View.* Wiley, New York.

[10] Diaconis, P. (1988). *Group Representations in Probability and Statistics.* IMS Lecture Notes **11**

[11] Engel, J., Herrmann, E., and Gasser, T. (1995). An iterative band-width selector for kernel estimation of densities and their derivatives. *J. Nonparametric Statist.* **4** 21–34.

[12] Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory Probab. Appl.* **14** 153–158.

[13] Fisher, N. I., Lewis, T., and Embleton, B. J. J. (1987). *Statistical Analysis of Spherical Data.* Cambridge University Press.

[14] Gnedenko, B. V. and Kolmogorov, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley, Reading, Mass.

[15] Grund, B., Hall, P., and Marron, J. S. (1994). Loss and risk in smoothing parameter selection. *J. Nonparametric Statist.* **4** 107–132.

[16] Hall, P., Sheather, S. J., Jones, M. C., and Marron, J. S. (1991). On optimal data-based bandwidth selection in kernel density estimation. *Biometrika* **78** 263–269.

[17] Hall, P. and Wand, M. (1988). Minimizing L^1 distance in nonparametric density estimation. *J. Multivariate Anal.* **26** 59–88.

[18] Hall, P., Watson, G. S., and Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika* **74** 751–62.

[19] Hendriks, H., Janssen, J. H. M., and Ruymgaart, F. H. (1993). Strong uniform convergence of density estimators on compact Euclidian manifolds. *Statist. Probab. Lett.* **16** 305–311.

[20] Hoffmann-Jørgensen, J. (1994). *Probability with a View toward Statistics*, Vol. I. Chapman & Hall, New York.

[21] Holmström, L. and Klemelä, J. (1992). Asymptotic bounds for the expected L^1 error of a multivariate kernel density estimator. *J. Multivariate Anal.* **42** 245–266.

[22] Klemelä, J. (1997). *Estimation of Densities and Functionals of Densities with Spherical Data.* Rolf Nevanlinna Institute Research Reports A 16.

[23] Mardia, K. V. (1972). *Statistics of Directional Data.* Academic Press, London.

[24] Nadaraya, E. A. (1974). On the integral mean square error of some nonparametric estimates for the density function. *Theor. Probability Appl.* **19** 133–141.

[25] Nikol'skii, S.M. (1975). *Approximation of Functions of Several Variables and Imbedding Theorems.* Springer-Verlag, Berlin.

[26] Park, B. U. and Marron, J. S. (1990). Comparison of data-driven bandwidth selectors. *J. Amer. Statist. Assoc.* **85** 66–72.

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

[28] Petrov, V. V. (1962). On certain polynomials occuring in probability theory. *Vestnik Leningrad U.* **19** 150–153. (In Russian.)

[29] Prakasa Rao, B. L. S. (1983). *Nonparametric Functional Estimation.* Academic Press, New York.

[30] Rudin, W. (1986). *Real and Complex Analysis.* 3rd ed. McGraw-Hill, Singapore.

[31] Ruymgaart, F. H. (1989). Strong uniform convergence of density estimators on spheres. *J. Statistical Planning and Inference* **23** 45–52.

[32] Scott, D. W., Tapia, R. A., and Thompson, J. R. (1977). Kernel density estimation revisited. *Nonlinear Anal. Theory Meth. Applic.* **1** 339–372.

[33] Sheather, S. J. and Jones, M. C. (1991). A reliable data-based band-width selection method for kernel density estimation. *J. Roy. Statist. Soc.* Ser. B **53** 683–690.

[34] Wand, M. P. and Jones M. C. (1994). Multivariate plug-in bandwidth selection. *Comp. Statist.* **9** 97–116.

[35] Watson, G. S. (1970). Orientation statistics in the earth sciences. *Bull. Geol. Inst. Univ. Uppsala N.S.* **2** 73–89.

[36] Watson, G. S. (1983). *Statistics on Spheres,* University of Arkansas Lecture Notes in Mathematical Sciences 6. Wiley, New York.

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

[38] Woodroofe, M. (1970). On choosing a delta-sequence. *Ann. Math. Statist.* **41** 1665–1671.