ESTIMATION OF DENSITIES AND FUNCTIONALS OF DENSITIES WITH SPHERICAL DATA

JUSSI KLEMELÄ

Academic dissertation

To be presented, with the permission of the Faculty of Science of the University of Helsinki, for public criticism in Auditorium XII of the Main Building on March 1st, 1997, at 10 a.m.

UNIVERSITY OF HELSINKI
DEPARTMENT OF MATHEMATICS
1997

Estimation of Densities and Functionals of Densities with Spherical Data

Jussi Klemelä
Rolf Nevanlinna Institute

Research Reports A 16
February 1997

Rolf Nevanlinna Institute
P.O. Box 4 (Yliopistonkatu 5)
00014 University of Helsinki, Finland

ISBN 952-9528-36-1
ISSN 0787-8338
YLIOPISTOPAINO

Acknowledgments

I wish to thank Dr. Lasse Holmström for originally directing my interest to kernel density estimation. I am grateful for his advice and support during this work and thank him for comments that led to substantial improvements of the work.

I am also grateful to Professor Timo Mäkeläinen for introducing me to the statistics of directional data. The Rolf Nevanlinna Institute and the Department of Statistics of the University of Helsinki have provided me with excellent working conditions. The financial support of the Academy of Finland and University of Helsinki is gratefully acknowledged.

Helsinki, February 1997

Jussi Klemelä

Contents

Chapter 1

Introduction

Sometimes statistical data consists of directions. Directions in a Euclidean space can be represented as points of the unit sphere. 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. The classical spherical sample spaces are the earth and the celestial sphere. Spherical data arise in such disciplines as astrophysics, meteorology, geology and geography when, for example, investigating the origins of comets, analyzing wind directions or interpreting paleomagnetic currents. Good introductions to the statistics of spherical data are given by Mardia (1972), Watson (1983), Fisher, Lewis and Embleton (1987). A breakthrough in the subject was made by Fisher (1953). The literature from the period 1975–1988 is reviewed in Jupp and Mardia (1989), and history of the subject prior to 1975 is reviewed in Mardia (1972), (1975).

Quite often the distribution of the observed random vector having values in a Euclidean space is absolutely continuous either with respect to the Lebesgue measure of the Euclidean space or some counting measure (discrete random vector). In this study the observations are random vectors in a Euclidean space having the special property, that their distribution is absolutely continuous with respect to the Lebesgue measure of the sphere, to be denoted by $\mu = \mu_d$. We will use the term spherical data to refer to independent and identically distributed observations from a distribution which is absolutely continuous with respect to the Lebesgue measure of the sphere and the term Euclidean data to refer to independent and identically distributed observations from a distribution which is absolutely continuous with respect to the

Lebesgue measure of the Euclidean space. 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 which is invariant with respect to rotations and with total mass $\omega_d \stackrel{def}{=} \mu(S_d) = 2\pi^{(d+1)/2}/\Gamma((d+1)/2)$. It is also the Borel measure which extends the volume element of the sphere. For the definition of the volume element see e.g. Helgason (1962, page 291).

A widely-used family of distributions for modeling spherical data is that of the von Mises–Fisher (or Langevin) distributions. These have probability density functions

$$f(x; e, \kappa) = a(\kappa) \exp\{\kappa e' x\}, x \in S_d,$$

where $e \in S_d$ and $\kappa \geq 0$. For $\kappa = 0$ this is the uniform distribution. Beran (1979) has generalized the von Mises–Fisher family by replacing x in the exponent by higher polynomials $t(x)$ in x. We get a generalization in another direction by considering families with distributions having densities of the form $g(e'x)$, where g is a known function. Further examples of parametric families of densities and references are given in Jupp and Mardia (1989).

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 $\delta > 0$. The term approximate identity or delta sequence has been used for such sequences of functions (Butzer and Nessel 1971, page 31). This estimator, which could be called a delta sequence estimator, was studied by Watson and Leadbetter (1963). The delta sequence estimator is called a 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 kernel estimator was defined by Fix and Hodges (1951), (1989), Rosenblatt (1956) (the one dimensional case) and Cacoullos (1966) (the multivariate case).

The delta sequence estimator can be defined also in the case of spherical data. 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 $\mathbf{R}^{d+1} \to \mathbf{R}^{d+1}$ whose matrix is orthogonal with determinant one). Let $e \in S_d$ 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 e, as $\kappa \to \infty$. 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

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_\kappa(R_{X_i,e}(x)) \tag{1.1}$$

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

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_\kappa(R_{x,e}(X_i)). \tag{1.2}$$

For practical purposes, in the case of Euclidean data, we could restrict ourselves to delta sequences of the form $K_\kappa(x) = \kappa^d L(\|\kappa x\|^2)$ 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 e. In Watson (1970), Beran (1979), Hall, Watson and Cabrera (1987), Bai, Radhakrishna Rao and Zhao (1988) the form $K_\kappa(x) = c(\kappa)L(\kappa^2(1 - x'e))$ 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 $K_\kappa(x) = c(\kappa)J(\kappa^2 x'e)$ 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 $K_\kappa(x) = c(\kappa)L(\kappa^2(1 - x'e))$. For this case it holds that $K_\kappa(R_{X_i,e}(x)) = c(\kappa)L(\kappa^2(1 - x'X_i)) = K_\kappa(R_{x,e}(X_i))$, and thus the definitions (1.1) and (1.2) can in this case 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*

$$\hat{f}_n(x) = \hat{f}_n(x, \kappa, L) = \frac{c(\kappa)}{n} \sum_{i=1}^{n} L(\kappa^2(1 - x'X_i)),$$

where $x \in S_d$ *and*

$$c(\kappa)^{-1} = \int_{S_d} L(\kappa^2(1 - x'y)) d\mu(y). \tag{1.3}$$

By the law of cosines, $\|x - y\|^2 = 2(1 - x'y)$ for $x, y \in S_d$. Thus this definition of the kernel estimator with data in S_d is similar to the definition of the kernel estimator with data in $\mathbf{R}^{d+1}$, except that the normalization constant is $c(\kappa)$ instead of κ^{d+1}. In Lemma 2.3 it is shown that $c(\kappa) \sim \kappa^d \left(2^{(d-2)/2}\omega_{d-1} \int_0^\infty t^{(d-2)/2} L(t) dt\right)^{-1}$, 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 $L(t) = e^{-t} I_{[0,\infty[}(t)$, $L(t) = (1 - t) I_{[0,1]}(t)$ 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 $C_\kappa(x) = \{y \in S_d \,|\, x'y \geq 1 - \kappa^{-2}\}$. The naive estimator was studied by Ruymgaart (1989), Hendriks, Janssen and Ruymgaart (1993).

Preliminary lemmas about the Laplace operator and approximation by convolution are given in Chapter 2. In Section 2.1 it is shown that the concept of a second derivative for a function $S_d \to \mathbf{R}$, as defined in Hall, Watson and Cabrera (1987), is the same as the Laplace operator multiplied by a constant. An associated delta sequence, which is a useful tool when studying approximation by convolution, is defined in Section 2.3. The results of Chapter 2 lead to the formulas for the asymptotic risk of the kernel estimator, which are given in Chapter 3. In Section 3.3 asymptotics for the mean squared error and the mean integrated squared error are given. In Section 3.4 the asymptotic for the mean integrated absolute error is given. It is seen that similarly to the Euclidean case, the risk has the rate of convergence $n^{s/(2s+d)}$ for both the L_1 and L_2 error, where $s \geq 2$ is a suitably defined smoothness index. This has already been proved for the mean integrated squared error and the smoothness index $s = 2$ in Hall, Watson and Cabrera (1987). In Chapter 4 it is shown that the smoothing parameter of the kernel estimator can be chosen empirically in such a way that the risk is asymptotically optimal. This is shown for the mean squared error criterion in Section 4.1 and for the mean integrated squared error criterion in Section 4.2. These results have been proved in the Euclidean case by Woodroofe (1970) and Nadaraya (1974).

In Chapter 5 estimators are defined for linear combinations of derivatives, iterated Laplacians and some integral functionals of these quantities. Their estimation is motivated in Section 3.3 where it is shown that the asymptotically optimal deterministic smoothing parameter involves such functionals of a density. In Section 5.1 an estimator for a linear combination of derivatives is defined as a linear combination of kernel estimators. In Section 5.2 an estimator for an iterated Laplacian is defined as an iterated Laplacian of the kernel estimator. The rate of convergence of the mean squared error of these estimators is shown to be $n^{(s-r)/(2s+d)}$, where $s \geq 2$ is the smoothness index and $0 \leq r < s$ denotes the order of derivative or Laplacian to be estimated. In Section 5.3 estimators for certain integral functionals of derivatives and Laplacians are given. It is shown that the estimator for the inner product of Laplacians has the rate of convergence $n^{1/2}$ for the mean squared error, if the density is sufficiently smooth.

In Chapter 6 some lower bounds are given for the asymptotic risk of arbitrary estimators. The proofs are based on the theory of convergence of experiments. In Section 6.2 a lower bound for the pointwise risk is given, which generalizes the result of Low (1992) to the general sample space. From this lower bound it follows that the rate of convergence $n^{(s-r)/(2s+d)}$ cited above is optimal for the estimation of a density at a single point ($r = 0$) and for the estimation of the Laplacian of a density at a point ($r \geq 2$). In Section 6.3 a lower bound for the integrated risk is given which seems to be new also in the Euclidean case. From this lower bound it follows that the rate of convergence $n^{s/(2s+d)}$ is optimal for the estimation of a density with integrated risk. In Section 6.4 it is shown that the estimator for the inner product of Laplacians, defined in Section 5.3, is optimal in the local asymptotic minimax sense. This result has been proved in the Euclidean case by Bickel and Ritov (1988).

Chapter 2

Approximation by Convolution

For $L : [0, \infty[\to \mathbf{R}$, define $L_\kappa : [-1, 1] \to \mathbf{R}$ by $L_\kappa(t) = L(\kappa^2(1-t))$, $\kappa > 0$. Define the convolution of $f : S_d \to \mathbf{R}$ and L_κ by

$$\begin{aligned} f * L_\kappa(x) &= \int_{S_d} f(y) L_\kappa(x'y) d\mu(y) = \int_{S_d} f(y) L(\kappa^2(1 - x'y)) d\mu(y) \qquad (2.1) \\ &= \int_{S_d} f(y) L(\|\kappa(x-y)\|^2/2) d\mu(y),\ x \in S_d. \end{aligned}$$

The term convolution will always refer to this definition above. For the kernel estimator $\hat{f}_n$,

$$\mathrm{E}\left(\hat{f}_n(x)\right) = c(\kappa)\mathrm{E}\left(L(\kappa^2(1 - x'X_1))\right) = c(\kappa) \int_{S_d} f(y) L(\kappa^2(1 - x'y)) d\mu(y),$$

which explains why the study of convolutions is important for the study of the kernel estimator. Also the variance of the kernel estimator involves a convolution as will be seen in Section 3.2.

Section 2.1 starts with the definitions of three parameterizations of S_d. Then it is proved that the concept of second derivative for functions $S_d \to \mathbf{R}$ which was defined by Hall, Watson and Cabrera (1987) is in fact the same as the Laplace operator multiplied by a constant. In Section 2.2 lemmas are proved which state that a function can be approximated by convolution both pointwise and in the L_p–norm. These lemmas are needed to take care of the asymptotics of the variance of the kernel estimator. In Section 2.3 the concept of an associated delta sequence is defined. This concept is used in Section 2.4, where an expansion of convolution is given using Taylor's theorem. It will be seen in Section 3.1 that, unlike in the Euclidean case, the asymptotics of the bias of the kernel estimator involves a certain linear combination of derivatives. This linear combination is defined in Section 2.4. Then the convergence of the remainder term of the expansion is studied. First pointwise convergence of the remainder term is proved, then L_p convergence and convergence under a Lipschitz condition. These different modes of convergence are needed in Chapter 3 in the study of the pointwise risk and the L_p risk for $p = 1, 2$.

2.1 The Laplace Operator

Let us start by giving definitions of some parameterizations of S_d. Let $\eta \in S_d$, $d \geq 2$, 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

$$\phi_\eta^{-1}(\xi, \theta) = \eta \cos\theta + \xi \sin\theta. \tag{2.2}$$

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. Let us also use ϕ_η to denote the mapping $S_d \setminus \{\eta, -\eta\} \to T_\eta \times]-1, 1[$, defined by

$$\phi_\eta^{-1}(\xi, t) = \eta t + \xi (1 - t^2)^{1/2}.$$

A third parameterization can be deduced inductively from the first one. Choose such $\eta^1, \ldots, \eta^{d+1} \in S_d$, $d \geq 1$, that $\eta^i \perp \eta^j$, when $i \neq j$. Let $\phi_{\eta^1,\ldots,\eta^{d+1}} : S_d \setminus U_{\eta^2,\ldots,\eta^{d+1}} \to]0, 2\pi[\times]0, \pi[^{d-1}$ be defined by

$$\phi^{-1}_{\eta^1,\ldots,\eta^{d+1}}(\theta) = \sum_{i=1}^{d+1} \eta^i \cos\theta_{i-1} \sin\theta_i \cdots \sin\theta_d, \tag{2.3}$$

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

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

To each parameterization corresponds an integration formula. For $f : S_d \to \mathbf{R}$ it holds that

$$\int_{S_d} f(x) d\mu(x) \;=\; \int_0^\pi d\theta_d \sin^{d-1}\theta_d \int_{T_\eta} f\left(\phi_\eta^{-1}(\xi, \theta_d)\right) d\mu_{d-1}(\xi) \tag{2.4}$$

$$= \int_{-1}^{1} dt\,(1-t^2)^{(d-2)/2} \int_{T_\eta} f\left(\phi_\eta^{-1}(\xi,t)\right) d\mu_{d-1}(\xi)$$
$$= \int_0^{2\pi} d\theta_1 \int_0^{\pi} d\theta_2 \sin\theta_2 \cdots \int_0^{\pi} f\left(\phi^{-1}_{\eta^1,\ldots,\eta^{d+1}}(\theta)\right) \sin^{d-1}\theta_d d\theta_d.$$

Let us proceed to the definition of the Laplace operator. Let $f : S_d \to \mathbf{R}$ and let $\eta^1,\ldots,\eta^{d+1} \in S_d$ be orthonormal (as before). Let $\phi = \phi_{\eta^1,\ldots,\eta^{d+1}}$ be as defined in (2.3). Define the Laplace operator by

$$\begin{aligned}&\Delta f(x)\\ &= \sum_{k=1}^{d} (\sin\theta_{k+1}\cdot\ldots\cdot\sin\theta_d)^{-2}\left[\frac{\partial^2}{\partial\theta_k^2} + (k-1)\frac{\cos\theta_k}{\sin\theta_k}\frac{\partial}{\partial\theta_k}\right] f\left(\phi^{-1}(\theta)\right)\big|_{\theta=\phi(x)}.\end{aligned}$$

The Laplace operator can also be defined recursively by

$$\begin{aligned}\Delta f(x) &= \Delta_d f(x) \qquad (2.5)\\ &= \left[\frac{\partial^2}{\partial\theta_d^2} + (d-1)\frac{\cos\theta_d}{\sin\theta_d}\frac{\partial}{\partial\theta_d} + \frac{1}{\sin^2\theta_d}\Delta_{d-1}\right] f\left(\phi^{-1}_{\eta^{d+1}}(\xi,\theta_d)\right)\Big|_{(\xi,\theta_d)=\phi_{\eta^{d+1}}(x)}\\ &= \left[(1-t^2)\frac{\partial^2}{\partial t^2} - t d\frac{\partial}{\partial t} + \frac{1}{1-t^2}\Delta_{d-1}\right] f\left(\phi^{-1}_{\eta^{d+1}}(\xi,t)\right)\Big|_{(\xi,t)=\phi_{\eta^{d+1}}(x)}\end{aligned}$$

for $d \geq 2$ and

$$\Delta_1 f(x) = \frac{\partial^2}{\partial\theta_1^2} f\left(\phi^{-1}_{\eta^1,\eta^2}(\theta_1)\right)\Big|_{\theta_1=\phi_{\eta^1,\eta^2}(x)}.$$

Next, an alternative expression for the Laplace operator is given. 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 $s \geq 2$ an 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

$$D^s f(x) = \omega_{d-1}^{-1} \int_{T_x} D_\xi^s \bar{f}(x) d\mu_{d-1}(\xi), \qquad (2.6)$$

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).

Lemma 2.1 *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.

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

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

$i, j = 1, \dots, d+1$. It can be seen after some calculation 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 $\phi = \phi_{\eta^1, \dots, \eta^{d+1}} = (\phi_1, \dots, \phi_d)'$ was defined in equation (2.3). 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,

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

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

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. □

Note that in the above proof the formula

$$
\Delta f(x) = \sum_{i=1}^{d+1}\frac{\partial^2}{\partial x_i^2}\bar{f}(x),
$$

was established. In addition, it was 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 can be shown that there exist no constant C for which $D^4 = C\Delta^2$. The following lemma gives a useful formula for directional derivatives.

Lemma 2.2 *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 $\theta \in \mathbf{R}$. Then*

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

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 $s \geq 2$ an integer. Let us first prove that

$$D_\xi^s f(x) = D_\xi^s \bar{f}(x) \tag{2.7}$$

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 (2.7) 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 (2.7) has been proved.

Let us prove the assertion of this lemma for $D_\xi^s f(x)$, which in view of equation (2.7) 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

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

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

From the above lemma 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

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

2.2 Convergence of a Convolution

When $L : [0, \infty[\to \mathbf{R}$ and $\kappa > 0$ define

$$d_0(\kappa) = d_0(\kappa, L) = \kappa^d \int_{S_d} L(\kappa^2(1 - x'\eta)) d\mu(x), \tag{2.8}$$

which does not depend on $\eta \in S_d$ because of the rotation invariance of μ. Note that $d_0(\kappa) = \kappa^d c(\kappa)^{-1}$, where $c(\kappa)$ was defined in (1.3). By equation (2.4) and using the substitution $\kappa^2(1-t) = u$,

$$\begin{aligned} |d_0(\kappa, L)| &\leq \omega_{d-1}\kappa^d \int_{-1}^{1} \left|L(\kappa^2(1-t))\right| (1-t^2)^{(d-2)/2} dt \\ &= \omega_{d-1} \int_0^{2\kappa^2} |L(u)|(2-\kappa^{-2}u)^{(d-2)/2} u^{(d-2)/2} du \\ &\leq \omega_{d-1} 2^{(d-2)/2} \int_0^{\infty} u^{(d-2)/2} |L(u)| du, \end{aligned} \tag{2.9}$$

where $\omega_{d-1} = \mu_{d-1}(S_{d-1})$. Thus $d_0(\kappa, L)$ is defined if $\int_0^\infty t^{(d-2)/2}|L(t)|dt < \infty$. Of course, it is also defined if L is bounded.

The concept of approximate identity is defined in Butzer and Nessel (1971, page 31) for the one dimensional Euclidean case. In the spherical case the family $K_\kappa : S_d \to \mathbf{R}$, $\kappa > 0$, is called an approximate identity at $\eta \in S_d$ if $\lim_{\kappa\to\infty} \int_{S_d} K_\kappa d\mu = 1$ and for every $\delta > 0$, $\lim_{\kappa\to\infty} \int_{\|x-\eta\|>\delta} K_\kappa(x) d\mu(x) = 0$. It follows from the following lemma that the family $K_\kappa(x) = c(\kappa)L(\kappa^2(1-x'\eta))$ is an approximate identity. Define

$$\alpha_0 = \alpha_0(L) = \int_0^\infty t^{(d-2)/2} L(t) dt. \tag{2.10}$$

Lemma 2.3 *Let* $\alpha_0(|L|) < \infty$.

(i) *Then,*

$$d_0(\kappa, L) = \omega_{d-1} 2^{(d-2)/2} \alpha_0(L) + o(1),$$

when $\kappa \to \infty$.

(ii) *For every* $\delta > 0$, $\eta \in S_d$,

$$\kappa^d \int_{\|x-\eta\|>\delta} L(\kappa^2(1-x'\eta)) d\mu(x) = o(1),$$

when $\kappa \to \infty$.

Proof. In the same way as in equation (2.9),

$$\int_{S_d} L(\kappa^2(1-x'\eta)) d\mu(x) = \omega_{d-1} \int_0^{2\kappa^2} L(u)\kappa^{-d}(2-\kappa^{-2}u)^{(d-2)/2} u^{(d-2)/2} du.$$

The first assertion follows by the dominated convergence theorem. Similarly,

$$\int_{\|x-\eta\|>\delta} L(\kappa^2(1-x'\eta)) d\mu(x) = \omega_{d-1} \int_{\delta^2\kappa^2/2}^{2\kappa^2} L(u)\kappa^{-d}(2-\kappa^{-2}u)^{(d-2)/2} u^{(d-2)/2} du$$

and the second assertion also follows by the dominated convergence theorem. $\square$

The next lemma establishes the pointwise convergence of convolutions, when convolution was defined in (2.1). The corresponding lemma in the Euclidean case has been proved by Bochner (1955, Theorem 1.1.1, page 2). It states that if $f : \mathbf{R}^d \to \mathbf{R}$ is continuous at $x_0 \in \mathbf{R}^d$ and $K : \mathbf{R}^d \to \mathbf{R}$ is integrable, then

$$\kappa^d \int_{\mathbf{R}^d} f(y)K(\kappa(x_0 - y))dy = f(x_0) \int_{\mathbf{R}^d} K + o(1),$$

when $\kappa \to \infty$.

Lemma 2.4 *Let $f : S_d \to \mathbf{R}$ be continuous at $x_0 \in S_d$ and assume f is bounded. Let $L : [0, \infty[\to \mathbf{R}$ be such that $\alpha_0(|L|) < \infty$. Then*

$$\kappa^d f * L_\kappa(x_0) = d_0(\kappa, L)f(x_0) + o(1)$$

when $\kappa \to \infty$.

Proof. Let $\epsilon > 0$. By the continuity of f at x_0 and Lemma 2.3 applied to $|L|$, there exists such $\delta > 0$ and $\kappa_0 > 0$ that for $\kappa \geq \kappa_0$,

$$\sup_{\{y \,|\, \|y - x_0\| \leq \delta\}} |f(y) - f(x_0)| \leq \frac{\epsilon}{2} d_0(\kappa, |L|)^{-1}$$

and

$$\kappa^d \int_{\|y - x_0\| > \delta} |L(\kappa^2(1 - x_0'y))| d\mu(y) \leq \frac{\epsilon}{2} \left[2 \sup_{y \in S_d} |f(y)| \right]^{-1}.$$

Now for $\kappa \geq \kappa_0$,

$$\begin{aligned}
&\left| \kappa^d \int_{S_d} L(\kappa^2(1 - x_0'y)) f(y) d\mu(y) - d_0(\kappa) f(x_0) \right| \\
&= \left| \kappa^d \int_{S_d} L(\kappa^2(1 - x_0'y))(f(y) - f(x_0)) d\mu(y) \right| \\
&\leq \sup_{\{y \,|\, \|y - x_0\| \leq \delta\}} |f(y) - f(x_0)| \kappa^d \int_{S_d} |L(\kappa^2(1 - x_0'y))| d\mu(y) \\
&\quad + 2 \sup_{y \in S_d} |f(y)| \kappa^d \int_{\|y - x_0\| > \delta} |L(\kappa^2(1 - x_0'y))| d\mu(y) \\
&\leq \epsilon.
\end{aligned}$$

$\square$

In the previous lemma the condition of boundedness of f can be removed. The next lemma proves L_p convergence of convolutions. The corresponding lemma in the Euclidean case has been proved for example in Nikol'skii (1975, page 28).

Lemma 2.5 *Let $1 \le p < \infty$. Let $f : S_d \to \mathbf{R}$ be such that $\|f\|_p < \infty$. Let $L : [0, \infty[\to \mathbf{R}$ be such that $\alpha_0(|L|) < \infty$. Then*

$$\left\| \kappa^d f * L_\kappa - d_0(\kappa) f \right\|_p = o(1)$$

when $\kappa \to \infty$.

Proof. 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

$$\left\| \kappa^d f * L_\kappa - d_0(\kappa) f \right\|_p \le \left\| \kappa^d (f-g) * L_\kappa \right\|_p + d_0(\kappa) \left\| f - g \right\|_p + \left\| \kappa^d g * L_\kappa - d_0(\kappa) g \right\|_p .$$

Also, by the generalized Minkowski inequality (Hoffmann–Jørgensen 1994, page 240) and equation (2.9),

$$\begin{aligned} \|\kappa^d (f-g) * L_\kappa\|_p &\le \|f-g\|_p \|\kappa^d L_\kappa(\cdot' \eta))\|_1 = \|f-g\|_p d_0(\kappa, |L|) \\ &\le \|f-g\|_p \left\{ \omega_{d-1} 2^{(d-2)/2} \alpha_0(|L|) \right\} . \end{aligned}$$

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 $e \in S_d$, $\delta > 0$ and define

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

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

$$\begin{aligned} &\left\{ \int_{S_d} \left| \kappa^d \int_{S_d} L(\kappa^2(1 - x'y)) f(y) d\mu(y) - d_0(\kappa) f(x) \right|^p d\mu(x) \right\}^{1/p} \\ &= \left\{ \int_{S_d} \left| \kappa^d \int_{S_d} L(\kappa^2(1 - x'y))(f(y) - f(x)) d\mu(y) \right|^p d\mu(x) \right\}^{1/p} \end{aligned}$$

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

The assertion follows from lemma 2.3. □

2.3 Associated Delta Sequence

The smoother the function f, the faster the convolution $c(\kappa) f * L_\kappa$ converges to f. When studying this phenomenon, the concept of associated delta sequence is useful for expressing the remainder term. 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 $s \ge 1$, then the parameter s kernel associated with K is defined by

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

Definition 2.6 *Let $s \ge 0$ be an integer. Let $L : [0, \infty[\to \mathbf{R}$ be such that $\int_0^\infty t^{(s+d-2)/2} |L(t)| dt < \infty$. The parameter s delta sequence associated with the delta sequence $c(\kappa) L(\kappa^2(1 - x'y))$ is $\tilde{L}_\kappa^{(s)} : [0, \pi] \to \mathbf{R}$, defined by*

$$
\tilde{L}_\kappa^{(s)}(\theta) = \omega_{d-1} \kappa^d \frac{\theta^s}{(s-1)!} \int_1^{\pi/\theta} (t-1)^{s-1} L\left(\kappa^2(1 - \cos(t\theta))\right) \sin^{d-1}(t\theta) dt,
$$

for $s \ge 1$ and

$$
\tilde{L}_\kappa^{(0)}(\theta) = \omega_{d-1} \kappa^d L(\kappa^2(1 - \cos\theta)) \sin^{d-1}\theta.
$$

Define, generalizing definition (2.8),

$$
d_s(\kappa) = d_s(\kappa, L) = \omega_{d-1} \kappa^d \int_0^\pi \theta^s L(\kappa^2(1 - \cos\theta)) \sin^{d-1}\theta d\theta, \tag{2.11}
$$

where $s \geq 0$ is real. The parameter s delta sequence is well defined almost everywhere in $[0, \pi]$. This can be seen in the following way. Using Fubini's theorem and the substitution $t\theta = \eta$, for $s \geq 1$,

$$\begin{aligned}
&\int_0^\pi \left| \tilde{L}_\kappa^{(s)}(\theta) \right| d\theta \qquad (2.12)\\
&\leq \frac{\omega_{d-1}\kappa^d}{(s-1)!} \int_0^\pi d\theta\, \theta^s \int_1^{\pi/\theta} (t-1)^{s-1} \left| L(\kappa^2(1-\cos(t\theta))) \right| \sin^{d-1}(t\theta) dt \\
&= \frac{\omega_{d-1}\kappa^d}{(s-1)!} \int_1^\infty dt (t-1)^{s-1} \int_0^{\pi/t} \theta^s \left| L(\kappa^2(1-\cos(t\theta))) \right| \sin^{d-1}(t\theta) d\theta \\
&= \frac{\omega_{d-1}\kappa^d}{(s-1)!} \int_1^\infty (t-1)^{s-1} t^{-s-1} dt \int_0^\pi \eta^s \left| L(\kappa^2(1-\cos\eta)) \right| \sin^{d-1}\eta d\eta \\
&= \frac{1}{s!} d_s(\kappa, |L|).
\end{aligned}$$

For $s = 0$ this follows from the definition. Define, for $s \geq 0$, $g :]0, 2] \to \mathbf{R}$ by $g(t) = (2t)^{-s/2} \arccos^s(1-t)$. Now $\lim_{t \downarrow 0} g(t) = 1$. Thus $|g| \leq M < \infty$ for some M. Thus, by the substitution $\kappa^2(1 - \cos\eta) = u$,

$$\begin{aligned}
d_s(\kappa, |L|) &= \omega_{d-1}\kappa^d \int_0^\pi \eta^s |L(\kappa^2(1-\cos\eta))| \sin^{d-1}\eta d\eta \qquad (2.13)\\
&= \omega_{d-1} \int_0^{2\kappa^2} \arccos^s(1-\kappa^{-2}u) |L(u)| (2-\kappa^{-2}u)^{(d-2)/2} u^{(d-2)/2} du \\
&\leq \omega_{d-1} 2^{(s+d-2)/2} \kappa^{-s} M \int_0^\infty u^{(s+d-2)/2} |L(u)| du.
\end{aligned}$$

The following lemma states that the associated delta sequence is an approximate identity up to the constant of normalization.

Lemma 2.7 *Let $s \geq 0$ integer, $a \geq 0$ real. Let $\int_0^\infty t^{(s+a+d-2)/2} |L(t)| dt < \infty$.*

(i) *Then,*

$$\int_0^\pi \theta^a \tilde{L}_\kappa^{(s)}(\theta) d\theta = \frac{\Gamma(a+1)}{\Gamma(s+a+1)} d_{s+a}(\kappa, L) = O\left(\kappa^{-s-a}\right)$$

when $\kappa \to \infty$.

(ii) *For every $0 < \delta \leq \pi$,*

$$\int_\delta^\pi \theta^a \tilde{L}_\kappa^{(s)}(\theta) d\theta = o(\kappa^{-s-a})$$

when $\kappa \to \infty$.

Proof. By the same steps as in (2.12), for $s \geq 1$,

$$
\begin{aligned}
&\int_0^\pi \theta^a \tilde{L}_\kappa^{(s)}(\theta) d\theta \\
&\quad = \frac{\omega_{d-1}\kappa^d}{(s-1)!} \int_1^\infty (t-1)^{s-1} t^{-s-a-1} dt \int_0^\pi \eta^{s+a} L(\kappa^2(1-\cos\eta)) \sin^{d-1}\eta d\eta \\
&\quad = \frac{1}{(s-1)!} \frac{\Gamma(a+1)\Gamma(s)}{\Gamma(s+a+1)} d_{s+a}(\kappa) = \frac{\Gamma(a+1)}{\Gamma(s+a+1)} d_{s+a}(\kappa).
\end{aligned}
$$

This equation follows by Definition 2.6 when $s = 0$. Similarly as in (2.13), for $s + a > 0$,

$$
\begin{aligned}
d_{s+a}(\kappa) &= \omega_{d-1}\kappa^d \int_0^{2\kappa^2} \arccos^{s+a}(1-\kappa^{-2}u) L(u) \kappa^{-d} (2-\kappa^{-2}u)^{(d-2)/2} u^{(d-2)/2} du \\
&= O(\kappa^{-s-a}).
\end{aligned}
$$

Assertion (i) has been proved. By the substitution $t\theta = \eta$, when $0 < \delta \leq \pi$, $s \geq 1$,

$$
\begin{aligned}
&\int_\delta^\pi \theta^a \tilde{L}_\kappa^{(s)}(\theta) d\theta \\
&\quad = \frac{\omega_{d-1}\kappa^d}{(s-1)!} \int_\delta^\pi d\theta\, \theta^{s+a} \int_1^{\pi/\theta} (t-1)^{s-1} L\left(\kappa^2(1-\cos(t\theta))\right) \sin^{d-1}(t\theta) dt \\
&\quad = \frac{\omega_{d-1}\kappa^d}{(s-1)!} \int_1^{\pi/\delta} dt (t-1)^{s-1} \int_\delta^{\pi/t} \theta^{s+a} L\left(\kappa^2(1-\cos(t\theta))\right) \sin^{d-1}(t\theta) d\theta \\
&\quad = \frac{\omega_{d-1}\kappa^d}{(s-1)!} \int_1^{\pi/\delta} dt (t-1)^{s-1} t^{-s-a-1} \int_{t\delta}^\pi \eta^{s+a} L(\kappa^2(1-\cos\eta)) \sin^{d-1}\eta d\eta
\end{aligned}
$$

and for $s \geq 0$,

$$
\begin{aligned}
&\int_\delta^\pi \eta^{s+a} |L(\kappa^2(1-\cos\eta))| \sin^{d-1}\eta d\eta \\
&\quad = \int_{(1-\cos\delta)\kappa^2}^{2\kappa^2} \arccos^{s+a}(1-\kappa^{-2}u) |L(u)| \kappa^{-d} (2-\kappa^{-2}u)^{(d-2)/2} u^{(d-2)/2} du \\
&\quad = o(\kappa^{-d-s-a}).
\end{aligned}
$$

□

Define, generalizing Definition (2.10),

$$
\alpha_s = \alpha_s(L) = \int_0^\infty t^{(s+d-2)/2} L(t) dt, \tag{2.14}
$$

where $s \geq 0$ is real. Let us next give an expansion for $d_m(L)$. Observe that Lemma 2.3 (i) is a special case of the following lemma.

Lemma 2.8 *Let $m \geq 0$, $r \geq 0$ be integers and suppose that*

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

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

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

where

$$\begin{aligned}
\gamma_{m,i} &= \sum_{j=0}^{i} \zeta_{m,j} \delta_{i-j}, \\
\delta_i &= \left(-\frac{1}{2}\right)^i \begin{pmatrix} (d-2)/2 \\ i \end{pmatrix}, \begin{pmatrix} a \\ i \end{pmatrix} = \frac{a(a-1)\cdots(a-i+1)}{i!}, \ \delta_0 = 1, \\
\zeta_{m,j} &= \sum_{\alpha_1 + \cdots + \alpha_m = j} \eta_{\alpha_1} \cdots \eta_{\alpha_m}, \ \zeta_{0,j} = 1, \\
\eta_i &= \frac{1 \cdot 3 \cdot 5 \cdots (2i-1)}{2^{2i}(2i+1)i!}, \ \eta_0 = 1.
\end{aligned}$$

Proof. By the substitution $\kappa^2(1 - \cos\theta) = t$,

$$\begin{aligned}
&\int_0^{\pi} \theta^m L(\kappa^2(1-\cos\theta)) \sin^{d-1}\theta d\theta \\
&\quad = \int_0^{2\kappa^2} \arccos^m(1 - \kappa^{-2}t) L(t) \kappa^{-d} (2 - \kappa^{-2}t)^{(d-2)/2} t^{(d-2)/2} dt.
\end{aligned}$$

By the formula for the binomial series and the series expansion for $\arccos(1-t)$ (Abramowitz and Stegun 1965, pages 14, 81),

$$\arccos^m(1-\kappa^{-2}t)(2-\kappa^{-2}t)^{(d-2)/2} = 2^{(m+d-2)/2}(\kappa^{-2}t)^{m/2} \sum_{i=0}^{r} \gamma_{m,i}(\kappa^{-2}t)^i + R_{r+1}(\kappa^{-2}t),$$

where $\lim_{\kappa\to\infty}(\kappa^{-2}t)^{-(m+2r)/2} R_{r+1}(\kappa^{-2}t) = 0$ and $|(\kappa^{-2}t)^{-(m+2r)/2} R_{r+1}(\kappa^{-2}t)| \leq M$, when $|t| \leq 2\kappa^2$, where $0 < M < \infty$. Thus

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

and the assertion follows from the dominated convergence theorem. □

2.4 Expansion of a Convolution

When $f : S_d \to \mathbf{R}$, define $\tilde{D}^s f : S_d \times \mathbf{R} \to \mathbf{R}$ by

$$\tilde{D}^s f(x,\theta) = \omega_{d-1}^{-1} \int_{T_x} D^s_{\phi_x^{-1}(\xi,\theta+\pi/2)} f\left(\phi_x^{-1}(\xi,\theta)\right) d\mu_{d-1}(\xi), \tag{2.15}$$

where $s \geq 0$ is an integer, $T_x = \{\xi \in S_d \,|\, \xi \perp x\}$ and ϕ_x was defined in (2.2). Note that $\phi_x^{-1}(\xi,\theta+\pi/2) \perp \phi_x^{-1}(\xi,\theta)$ and by lemma 2.2 (or by the 5th display of the proof of lemma 2.2),

$$D^s_{\phi_x^{-1}(\xi,\theta+\pi/2)} f\left(\phi_x^{-1}(\xi,\theta)\right) = \frac{\partial^s}{\partial\theta^s} f\left(\phi_x^{-1}(\xi,\theta)\right).$$

We shall next formulate a spherical counterpart of the following result given in Holmström and Klemelä (1992, Proposition 3). 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.

Lemma 2.9 *Let $x_0 \in S_d$ and $s \geq 0$ even. Assume that for all $\xi \in T_{x_0}$, $\partial^s/\partial\theta^s f(\phi_{x_0}^{-1}(\xi,\theta))$ is continuous as a function of $\theta \in \mathbf{R}$. Let $\alpha_i(|L|) < \infty$ for $i = 0, s$, where α_i was defined in (2.14). Then*

$$\kappa^d f * L_\kappa(x_0) = \sum_{i=0}^{s/2-1} \frac{1}{(2i)!} d_{2i}(\kappa, L) D^{2i} f(x_0) + \int_0^\pi \tilde{L}_\kappa^{(s)}(\theta) \tilde{D}^s f(x_0,\theta) d\theta,$$

where $d_{2i}(\kappa)$ were defined in (2.11).

Proof. The case $s = 0$ follows directly from formula (2.4). Let $s \geq 2$ be even, $\xi \in S_d$, $\xi \perp x_0$, and $\theta \in \mathbf{R}$. By Taylor's theorem,

$$\begin{aligned} f\left(\phi_{x_0}^{-1}(\xi,\theta)\right) &= f(x_0) + \sum_{i=1}^{s-1} \frac{\theta^i}{i!} \frac{\partial^i}{\partial\nu^i} f\left(\phi_{x_0}^{-1}(\xi,\nu)\right)\Big|_{\nu=0} \\ &\quad + \frac{\theta^s}{(s-1)!} \int_0^1 (1-t)^{s-1} \frac{\partial^s}{\partial\nu^s} f\left(\phi_{x_0}^{-1}(\xi,\nu)\right)\Big|_{\nu=t\theta} dt. \end{aligned}$$

Thus,

$$
\begin{aligned}
&\kappa^d \int_{S_d} L(\kappa^2(1-x_0'y))f(y)d\mu(y) \\
&= \kappa^d \int_0^\pi d\theta \sin^{d-1}\theta L(\kappa^2(1-\cos\theta)) \int_{T_{x_0}} f\left(\phi_{x_0}^{-1}(\xi,\theta)\right) d\mu_{d-1}(\xi) \\
&= \sum_{i=0}^{s-1} \frac{1}{i!} d_i(\kappa) D^i f(x_0) + \frac{\kappa^d}{(s-1)!} \int_0^\pi d\theta\, \theta^s L(\kappa^2(1-\cos\theta)) \sin^{d-1}\theta \\
&\quad \int_{T_{x_0}} d\mu_{d-1}(\xi) \int_0^1 (1-t)^{s-1} \left.\frac{\partial^s}{\partial \nu^s} f\left(\phi_{x_0}^{-1}(\xi,\nu)\right)\right|_{\nu=t\theta} dt,
\end{aligned}
$$

where Lemma 2.2 was used. By the comment after Lemma 2.2, $D^i f(x_0) = 0$ when i is odd. By the substitutions $t\theta = \eta$ and $t^{-1} = \tau$,

$$
\begin{aligned}
&\frac{\kappa^d}{(s-1)!} \int_0^\pi d\theta\, \theta^s L(\kappa^2(1-\cos\theta)) \sin^{d-1}\theta \\
&\quad \int_{T_{x_0}} d\mu_{d-1}(\xi) \int_0^1 (1-t)^{s-1} \left.\frac{\partial^s}{\partial \nu^s} f\left(\phi_{x_0}^{-1}(\xi,\nu)\right)\right|_{\nu=t\theta} dt \\
&= \frac{\kappa^d}{(s-1)!} \int_0^1 dt(1-t)^{s-1} \int_0^\pi \theta^s L(\kappa^2(1-\cos\theta)) \sin^{d-1}\theta \\
&\quad \int_{T_{x_0}} d\mu_{d-1}(\xi) D^s_{\phi_{x_0}^{-1}(\xi,t\theta+\pi/2)} f\left(\phi_{x_0}^{-1}(\xi,t\theta)\right) d\theta \\
&= \frac{\omega_{d-1}\kappa^d}{(s-1)!} \int_0^1 dt(1-t)^{s-1} \int_0^\pi \theta^s L(\kappa^2(1-\cos\theta)) \sin^{d-1}\theta\, \tilde{D}^s f(x_0,t\theta) d\theta \\
&= \frac{\omega_{d-1}\kappa^d}{(s-1)!} \int_0^1 dt(1-t)^{s-1} t^{-s-1} \int_0^{t\pi} \eta^s L\left(\kappa^2(1-\cos(\eta/t))\right) \sin^{d-1}(\eta/t) \tilde{D}^s f(x_0,\eta) d\eta \\
&= \frac{\omega_{d-1}\kappa^d}{(s-1)!} \int_0^\pi d\eta\, \eta^s \tilde{D}^s f(x_0,\eta) \int_{\eta/\pi}^1 (1-t)^{s-1} t^{-s-1} L\left(\kappa^2(1-\cos(\eta/t))\right) \sin^{d-1}(\eta/t) dt \\
&= \frac{\omega_{d-1}\kappa^d}{(s-1)!} \int_0^\pi d\eta\, \eta^s \tilde{D}^s f(x_0,\eta) \int_1^{\pi/\eta} (\tau-1)^{s-1} L\left(\kappa^2(1-\cos(\tau\eta))\right) \sin^{d-1}(\tau\eta) d\tau.
\end{aligned}
$$

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

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

$$
\mathcal{D}^s f = \sum_{i=1}^{s/2} \frac{2^i}{(2i)!} \gamma_{2i,s/2-i} D^{2i} f, \tag{2.16}
$$

where $\gamma_{2i,s/2-i}$ were defined in Lemma 2.8. Suppose that

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

for $i = 1, \ldots, s/2$. Then we have a useful equation,

$$\begin{aligned}\kappa^d f * L_\kappa(x_0) &= d_0(\kappa) f(x_0) \\ &+\omega_{d-1} 2^{(d-2)/2} \left[\sum_{i=1}^{s/2-1} \kappa^{-2i} \alpha_{2i} \mathcal{D}^{2i} f(x_0) + \kappa^{-s} \alpha_s \sum_{i=1}^{s/2-1} \frac{2^i}{(2i)!} \gamma_{2i,s/2-i} D^{2i} f(x_0) \right] \\ &+ \int_0^\pi \tilde{L}_\kappa^{(s)}(\theta) \tilde{D}^s f(x_0, \theta) d\theta + o(\kappa^{-s}) \sum_{i=1}^{s/2-1} D^{2i} f(x_0),\end{aligned} \tag{2.17}$$

where $o(\kappa^{-s})$ does not depend on f. In fact, by Lemma 2.8 ($m = 2i$, $r = s/2 - i$) and the substitution $i + j = l$,

$$\begin{aligned}&\sum_{i=1}^{s/2-1} \frac{1}{(2i)!} d_{2i}(\kappa) D^{2i} f(x_0) \\ &= \sum_{i=1}^{s/2-1} \frac{1}{(2i)!} D^{2i} f(x_0) \left\{ \omega_{d-1} 2^{(2i+d-2)/2} \sum_{j=0}^{s/2-i} \kappa^{-2i-2j} \gamma_{2i,j} \alpha_{2i+2j} + o(\kappa^{-s}) \right\} \\ &= \sum_{i=1}^{s/2-1} \frac{1}{(2i)!} D^{2i} f(x_0) \omega_{d-1} 2^{(2i+d-2)/2} \sum_{l=i}^{s/2} \kappa^{-2l} \gamma_{2i,l-i} \alpha_{2l} + o(\kappa^{-s}) \sum_{i=1}^{s/2-1} D^{2i} f(x_0) \\ &= \omega_{d-1} 2^{(d-2)/2} \sum_{l=1}^{s/2-1} \kappa^{-2l} \alpha_{2l} \sum_{i=1}^{l} \frac{2^i}{(2i)!} \gamma_{2i,l-i} D^{2i} f(x_0) \\ &\quad + \omega_{d-1} 2^{(d-2)/2} \kappa^{-s} \alpha_s \sum_{i=1}^{s/2-1} \frac{2^i}{(2i)!} \gamma_{2i,s/2-i} D^{2i} f(x_0) + o(\kappa^{-s}) \sum_{i=1}^{s/2-1} D^{2i} f(x_0).\end{aligned}$$

Thus equation (2.17) follows from Lemma 2.9 and Definition 2.16. The following modification of equation (2.17) is useful when estimating derivatives (Lemma 5.2). Define, for $s \geq 0$ even,

$$\mathcal{D}_0^s f = \sum_{i=0}^{s/2} \frac{2^i}{(2i)!} \gamma_{2i,s/2-i} D^{2i} f. \tag{2.18}$$

Suppose that

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

for $i = 0, \ldots, s/2$. Then from Equation (2.17) and Lemma 2.8 ($m = 0$, $r = s/2$) follows the equation

$$\begin{aligned} &\kappa^d f * L_\kappa(x_0) \qquad\qquad (2.19)\\ &= \omega_{d-1} 2^{(d-2)/2} \left[\sum_{i=0}^{s/2-1} \kappa^{-2i} \alpha_{2i} \mathcal{D}_0^{2i} f(x_0) + \kappa^{-s} \alpha_s \sum_{i=0}^{s/2-1} \frac{2^i}{(2i)!} \gamma_{2i,s/2-i} D^{2i} f(x_0) \right] \\ &\quad + \int_0^\pi \tilde{L}_\kappa^{(s)}(\theta) \tilde{D}^s f(x_0, \theta) d\theta + o(\kappa^{-s}) \sum_{i=0}^{s/2-1} D^{2i} f(x_0), \end{aligned}$$

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

To make the previous results useful, convergence of the remainder term of the expansions must be studied. It is started with the pointwise convergence. In the case $s = 0$, the following lemma combined with Lemma 2.9 gives the same formula as Lemma 2.4.

Lemma 2.10 *Let $s \geq 0$ be even and $x_0 \in S_d$. Let $\left|\tilde{D}^s f(x_0, \theta)\right|$ be bounded for $\theta \in [0, \pi]$ and $\lim_{\theta \to 0} \tilde{D}^s f(x_0, \theta) = D^s f(x_0)$. Let $\alpha_s(|L|) < \infty$. Then*

$$\int_0^\pi \tilde{L}_\kappa^{(s)}(\theta) \tilde{D}^s f(x_0, \theta) d\theta = \frac{1}{s!} d_s(\kappa, L) D^s f(x_0) + o(\kappa^{-s})$$

when $\kappa \to \infty$.

Proof. By lemma 2.7 (i),

$$\begin{aligned} &\left| \int_0^\pi \tilde{L}_\kappa^{(s)}(\theta) \tilde{D}^s f(x_0, \theta) d\theta - \frac{1}{s!} d_s(\kappa) D^s f(x_0) \right| \\ &= \left| \int_0^\pi \tilde{L}_\kappa^{(s)}(\theta) \left(\tilde{D}^s f(x_0, \theta) - D^s f(x_0) \right) d\theta \right| \leq \int_0^\pi \left| \tilde{L}_\kappa^{(s)}(\theta) \right| \left| \tilde{D}^s f(x_0, \theta) - D^s f(x_0) \right| d\theta. \end{aligned}$$

The assertion can be deduced from Lemma 2.7 (applied to $|L|$) by the same argument which was used in the proof of Lemma 2.4, because $\int_0^\pi \left| \tilde{L}_\kappa^{(s)} \right| \leq \int_0^\pi \widetilde{|L|}_\kappa^{(s)}$. □

Let us move on to the convergence in the L_p norm. In the second part of the following lemma, a certain Lipschitz condition is formulated.

Lemma 2.11 *Let $s \geq 0$ even and $1 \leq p \leq \infty$. Put $\|f\|_p^p = \int_{S_d} |f|^p d\mu$ when $p < \infty$ and $\|f\|_\infty$ equal to essential supremum of $|f|$. Assume that $f : S_d \to \mathbf{R}$ is such that $\|D^s f\|_p < \infty$ and $\left\| \tilde{D}^s f(\cdot, \theta) \right\|_p$ is bounded for $\theta \in [0, \pi]$. Let $\alpha_s(|L|) < \infty$.*

(i) *If* $\lim_{\theta\to 0}\left\|\tilde{D}^s f(\cdot,\theta) - D^s f\right\|_p = 0$*, then*

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

when $\kappa\to\infty$*.*

(ii) *If* $\theta^{-a}\left\|\tilde{D}^s f(\cdot,\theta) - D^s f\right\|_p$ *is bounded for* $\theta\in[0,\pi]$*, where* $0\le a<2$*, then*

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

Proof. By the generalized Minkowski's inequality,

$$\left\|\int_0^\pi \tilde{L}_\kappa^{(s)}(\theta)\tilde{D}^s f(\cdot,\theta)d\theta\right\|_p \le \int_0^\pi \left|\tilde{L}_\kappa^{(s)}(\theta)\right| \left\|\tilde{D}^s f(\cdot,\theta)\right\|_p d\theta < \infty$$

by the calculations in (2.12) and (2.13). By Lemma 2.7 (i) and the generalized Minkowski's inequality,

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

The assertion (i) can be inferred from Lemma 2.7 by the same argument that was used in the proof of Lemma 2.10. The assertion (ii) follows by Lemma 2.7. □

Chapter 3

Asymptotic Risk of the Kernel Estimator

In this chapter three measures of risk for kernel estimators will be studied. First, the mean squared error, then the mean integrated squared error and thirdly, the mean integrated absolute error.

Measuring loss by the integrated squared error is technically easier but the integrated absolute error is a more natural measure 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) and in Devroye (1987).

Section 3.1 begins with the definition of a class s kernel. After this, pointwise bias, integrated bias, and bias under a Lipschitz condition are studied. It is seen that unlike the Euclidean case, the asymptotics of the bias involves a certain linear combination of the derivatives of the density. In Section 3.2, pointwise variance, integrated variance, and integrated standard deviation are studied. In Section 3.3, the asymptotics of the mean squared error and the mean integrated squared error are studied. These results follow directly from the results of Sections 3.1 and 3.2. The asymptotically optimal smoothing parameter and the asymptotically optimal nonnegative kernel are given. Finally, in Section 3.4, the asymptotics of the mean integrated absolute error is given.

3.1 Bias

The expectation of the kernel estimator $\hat{f}_n(x)$ is

$$\mathrm{E}\left(\hat{f}_n(x)\right) = c(\kappa)\int_{S_d} L(\kappa^2(1-x'y))f(y)d\mu(y) = c(\kappa)f * L_\kappa(x).$$

Thus the results of the previous chapter concerning approximation by convolution are directly relevant for the asymptotics of the bias of the kernel estimator. As before, let us denote $\alpha_i(L) = \int_0^\infty t^{(i+d-2)/2}L(t)dt$, for $i \geq 0$.

Definition 3.1 *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) $\alpha_0(L) \neq 0$,

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

For a class s kernel, $\alpha_{2i}(L) = 0$ for $i = 1, \ldots, s/2-1$. When L has a compact support, condition (iii) is equivalent to $\alpha_{2i}(L) = 0$ for $i = 1, \ldots, s/2-1$. Thus, to construct a class s kernel, a polynomial on $[0,1]$ can be fitted to L, as in Devroye (1987, page 100). The definition of a class s kernel in the Euclidean case is given for example in Prakasa Rao (1983). The idea of reducing the bias by using a class s kernel was introduced in Parzen (1962) and Bartlett (1963). To see the similarity between the spherical case and the Euclidean case, note that when $K : \mathbf{R}^d \to \mathbf{R}$, $K(x) = L(\|x\|^2)$, then

$$\begin{aligned}\int_{\mathbf{R}^d} K &= \int_0^\infty dr\, r^{d-1}\int_{S_{d-1}} K(rx)d\mu_{d-1}(x) = \omega_{d-1}\int_0^\infty r^{d-1}L(r^2)dr \\ &= \omega_{d-1}\frac{1}{2}\int_0^\infty t^{(d-2)/2}L(t)dt = \omega_{d-1}\frac{1}{2}\alpha_0(L) \qquad (3.1)\end{aligned}$$

and

$$\begin{aligned}&\int_{\mathbf{R}^d} y_1^s K(y)dy = \int_0^\infty dr\, r^{d-1}\int_{S_{d-1}} (rx_1)^s K(rx)d\mu_{d-1}(x) \\ &= \int_{S_{d-1}} x_1^s d\mu_{d-1}(x)\int_0^\infty r^{s+d-1}L(r^2)dr = \int_{S_{d-1}} x_1^s d\mu_{d-1}(x)\frac{1}{2}\int_0^\infty t^{(s+d-2)/2}L(t)dt \\ &= \int_{S_{d-1}} x_1^s d\mu_{d-1}(x)\cdot\frac{1}{2}\alpha_s(L). \qquad (3.2)\end{aligned}$$

Let $s \geq 2$ be even and $x_0 \in S_d$. Let $\mathbf{F}_1(s, x_0)$ be the set of such functions $f : S_d \to \mathbf{R}$ that $D^i f(x_0)$ is defined for $i = 1, \ldots, s$, for all $\xi \in T_{x_0} = \{\xi \in S_d \,|\, \xi \perp x_0\}$, $\partial^s/\partial\theta^s f(\phi_{x_0}^{-1}(\xi, \theta))$ is continuous as a function of $\theta \in \mathbf{R}$, $\left|\tilde{D}^s f(x_0, \theta)\right|$ is bounded for $\theta \in \mathbf{R}$, and $\lim_{\theta \to 0} \tilde{D}^s f(x_0, \theta) = D^s f(x_0)$.

For example, a function $f : S_d \to \mathbf{R}$, $f(x) = g(x'\eta)$, where $\eta \in S_d$ and $g : [-1, 1] \to \mathbf{R}$, belongs to $\mathbf{F}_1(s, x_0)$ if g is s times continuously differentiable.

Lemma 3.2 *Assume that $f \in \mathbf{F}_1(s, x_0)$. If L is a class s kernel,*

$$c(\kappa) f * L_\kappa(x_0) = f(x_0) + \kappa^{-s} \alpha_0(L)^{-1} \alpha_s(L) \mathcal{D}^s f(x_0) + o(\kappa^{-s})$$

when $\kappa \to \infty$, where $\mathcal{D}^s f$ was defined in (2.16).

Proof. By equation (2.17) ($c(\kappa) = \kappa^d d_0(\kappa)^{-1}$),

$$\begin{aligned} c(\kappa) f * L_\kappa(x_0) &= f(x_0) + d_0(\kappa)^{-1} \omega_{d-1} 2^{(d-2)/2} \kappa^{-s} \alpha_s \sum_{i=1}^{s/2-1} \frac{2^i}{(2i)!} \gamma_{2i, s/2-i} D^{2i} f(x_0) \\ &\quad + d_0(\kappa)^{-1} \int_0^\pi \tilde{L}_\kappa^{(s)}(\theta) \tilde{D}^s f(x_0, \theta) d\theta + d_0(\kappa)^{-1} o(\kappa^{-s}). \end{aligned}$$

By Lemma 2.3 (i) (or by Lemma 2.8), $d_0(\kappa) = \omega_{d-1} 2^{(d-2)/2} \alpha_0 + o(1)$. By Lemma 2.10,

$$\int_0^\pi \tilde{L}_\kappa^{(s)}(\theta) \tilde{D}^s f(x_0, \theta) d\theta = \frac{1}{s!} d_s(\kappa) D^s f(x_0) + o(\kappa^{-s}).$$

By Lemma 2.8 (choose $m = s$, $r = 0$), $d_s(\kappa) = \omega_{d-1} 2^{(s+d-2)/2} \kappa^{-s} \alpha_s + o(\kappa^{-s})$. □

It is seen from the proof of the above lemma that the reason why the asymptotic bias involves the linear combination of derivatives (term $\mathcal{D}^s f$) is that the coefficients $d_{2i}(\kappa)$ appearing in the asymptotic expansion of the convolution given in Lemma 2.4 do not necessarily vanish with any choice of the kernel function. However, when using a class s kernel, they are of order $O(\kappa^{-s})$. Each $d_{2i}(\kappa)$ gives one term to the sum $\mathcal{D}^s f$.

Let $s \geq 2$ be even and $1 \leq p < \infty$. Let $\mathbf{F}_2(s, p)$ 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 $\xi \in T_x$, $\partial^s/\partial\theta^s f(\phi_x^{-1}(\xi, \theta))$ 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 $\lim_{\theta \to 0} \left\|\tilde{D}^s f(\cdot, \theta) - D^s f\right\|_p = 0$.

A function $f : S_d \to \mathbf{R}$, $f(x) = g(x'\eta)$, where $\eta \in S_d$ and $g : [-1, 1] \to \mathbf{R}$, also belongs to $\mathbf{F}_2(s, p)$ if g is s times continuously differentiable.

Lemma 3.3 *Assume that $f \in \mathbf{F}_2(s,p)$. Let L be a class s kernel. Then*

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

Proof. By equation (2.17) ($c(\kappa) = \kappa^d d_0(\kappa)^{-1}$),

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

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

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

By Lemma 2.3 (i) (or by Lemma 2.8), $\lim_{\kappa\to\infty} \left| d_0(\kappa)^{-1} \omega_{d-1} 2^{(d-2)/2} - \alpha_0^{-1} \right| = 0$. Also,

$$\begin{aligned} &\left| d_0(\kappa)^{-1} \kappa^s \int_0^\pi \tilde{L}_\kappa^{(s)}(\theta) \tilde{D}^s f(x,\theta) d\theta - \alpha_0^{-1} \alpha_s \frac{2^{s/2}}{s!} D^s f(x) \right| \\ &\quad \leq |d_0(\kappa)|^{-1} \kappa^s \left| \int_0^\pi \tilde{L}_\kappa^{(s)}(\theta) \tilde{D}^s f(x,\theta) d\theta - \frac{1}{s!} d_s(\kappa) D^s f(x) \right| \\ &\qquad + \left| d_0(\kappa)^{-1} \kappa^s d_s(\kappa) - \alpha_0^{-1} \alpha_s 2^{s/2} \right| \frac{1}{s!} |D^s f(x)|. \end{aligned}$$

The L_p norm of the first term on the right hand side converges to zero by Lemma 2.11 (i). The L_p norm of the second term on the right hand side converges to zero because, by Lemma 2.8, $d_0(\kappa) = \omega_{d-1}2^{(d-2)/2}\alpha_0 + o(1)$ and $d_s(\kappa) = \omega_{d-1}2^{(s+d-2)/2}\kappa^{-s}\alpha_s + o(\kappa^{-s})$ (choose $m = s$, $r = 0$). $\square$

In the following lemma, the convergence of the bias is analyzed under a Lipschitz condition. A similar type of analysis in the one dimensional Euclidean case has been given in Devroye (1987, page 107). This lemma is needed in Section 5.3.

Let $\sigma > 0$ and $1 \le p \le \infty$. Write $\sigma = s + a$, where $s \ge 0$ even and $0 \le a < 2$. Let $\mathbf{F}_3(\sigma, p)$ 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 $\xi \in T_x$, $\partial^s/\partial\theta^s f(\phi_x^{-1}(\xi, \theta))$ is continuous as a function of $\theta \in \mathbf{R}$, $\left\|\tilde{D}^s f(\cdot, \theta)\right\|_p$ and $\theta^{-a}\left\|\tilde{D}^s f(\cdot, \theta) - D^s f\right\|_p$ are bounded for $\theta \in [0, \pi]$.

Lemma 3.4 *Assume that $f \in \mathbf{F}_3(s + a, p)$, where $s \ge 0$ is even and $0 \le a < 2$. Let L be a class $s + 2$ kernel. Then*

$$\|c(\kappa) f * L_\kappa - f\|_p = O\left(\kappa^{-s-a}\right)$$

when $\kappa \to \infty$.

Proof. By Lemma 2.8, $d_{2i}(\kappa) = O(\kappa^{-s-2})$, for $i = 1, \ldots, s/2$ (choose $m = 2i$, $r = (s - 2i + 2)/2$). Thus, by Lemma 2.9 ($c(\kappa) = \kappa^d d_0(\kappa)^{-1}$),

$$\begin{aligned}
&\|c(\kappa) f * L_\kappa - f\|_p \\
&\quad = |d_0(\kappa)|^{-1}\left\|\sum_{i=1}^{s/2-1}\frac{1}{(2i)!}d_{2i}(\kappa)D^{2i}f + \int_0^\pi \tilde{L}_\kappa^{(s)}(\theta)\tilde{D}^s f(\cdot, \theta)d\theta\right\|_p \\
&\quad \le |d_0(\kappa)|^{-1}\left\|\int_0^\pi \tilde{L}_\kappa^{(s)}(\theta)\tilde{D}^s f(\cdot, \theta)d\theta\right\|_p + O\left(\kappa^{-s-2}\right) \\
&\quad \le |d_0(\kappa)|^{-1}\left\|\int_0^\pi \tilde{L}_\kappa^{(s)}(\theta)\tilde{D}^s f(\cdot, \theta)d\theta - \frac{1}{s!}d_s(\kappa)D^s f\right\|_p + \frac{d_s(\kappa)}{d_0(\kappa)s!}\|D^s f\|_p + o\left(\kappa^{-s-a}\right).
\end{aligned}$$

The assertion follows from Lemma 2.11 (ii). $\square$

3.2 Variance

Let $\hat{f}_n$ be the kernel estimator defined in Definition 1.1. Let $x \in S_d$. Now

$$\mathrm{Var}\left(\hat{f}_n(x)\right) = n^{-1}c(\kappa)^2\mathrm{Var}\left[L(\kappa^2(1 - x'X_1))\right]$$

$$= n^{-1}\kappa^{2d}d_0(\kappa,L)^{-2}\left\{\mathrm{E}L^2(\kappa^2(1-x'X_1)) - \left[\mathrm{E}L(\kappa^2(1-x'X_1))\right]^2\right\}$$
$$= n^{-1}\kappa^{2d}d_0(\kappa,L)^{-2}\left\{f*\left(L^2\right)_\kappa(x) - (f*L_\kappa(x))^2\right\},$$

where

$$d_0(\kappa,L) = \kappa^d c(\kappa,L)^{-1} = \kappa^d\int_{S_d} L(\kappa^2(1-x'y))d\mu(x).$$

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

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

and

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

To calculate the mean squared error of the kernel estimator, the following result is needed.

Lemma 3.5 *Let f be a bounded density which is continuous at $x_0 \in S_d$. Let $\alpha_0(|L|^i) < \infty$ for $i = 1, 2$. Then*

$$\mathrm{Var}\left(\hat{f}_n(x_0)\right) = \frac{\kappa^d}{n}\frac{\alpha_0(L^2)}{\omega_{d-1}2^{(d-2)/2}\alpha_0(L)^2}f(x_0) + \frac{o(\kappa^d)}{n}$$

when $\kappa \to \infty$.

Proof. By lemma 2.4,

$$\kappa^d\left(L^2\right)_\kappa * f(x_0) = d_0(\kappa,L^2)f(x_0) + o(1).$$

By lemma 2.3 (i),

$$d_0(\kappa,L^i) = \omega_{d-1}2^{(d-2)/2}\alpha_0(L^i) + o(1) \tag{3.3}$$

for $i = 1, 2$. By lemma 2.4 and equation (3.3),

$$\kappa^d\left(f*L_\kappa(x_0)\right)^2 = \kappa^{-d}\left(d_0(\kappa,L)f(x_0)+o(1)\right)^2 = O(\kappa^{-d}) = o(1).$$

Thus $n\kappa^{-d}R(x_0) = o(1)$. The assertion follows because by equation (3.3),

$$\left|V(x_0) - \frac{\kappa^d}{n}\frac{\alpha_0(L^2)}{\omega_{d-1}2^{(d-2)/2}\alpha_0(L)^2}f(x_0)\right| = \frac{o(\kappa^d)}{n}.$$

□

To calculate the mean integrated squared error of the kernel estimator the following result is needed.

Lemma 3.6 *Let* $\int_{S_d} f^2 d\mu < \infty$. *Let* $\alpha_0(|L|^i) < \infty$ *for* $i = 1, 2$. *Then*

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

when $\kappa \to \infty$.

Proof. By Lemma 2.5,

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

By Lemma 2.5 and equation (3.3),

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

Thus, $n\kappa^{-d} \int_{S_d} R d\mu = o(1)$. The assertion follows because by equation (3.3),

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

□

To calculate the mean integrated absolute error of the kernel estimator the following result is needed.

Lemma 3.7 *Let* $\alpha_0(|L|^i) < \infty$ *for* $i = 1, 2$. *Then*

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

when $\kappa \to \infty$.

Proof. 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 (3.3), and Lemma 2.5,

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

when $\kappa \to \infty$. The assertion follows because by equation (3.3),

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

□

3.3 Mean Squared Error and Mean Integrated Squared Error

Let us start with the mean squared error. From Lemma 3.2 and Lemma 3.5 it follows that

$$\begin{aligned}
\mathrm{E}\left(\hat{f}_n(x_0, \kappa) - f(x_0)\right)^2 &= \left(\mathrm{E}(\hat{f}_n(x_0, \kappa)) - f(x_0)\right)^2 + \mathrm{Var}\left(\hat{f}_n(x_0, \kappa)\right) \\
&= \kappa^{-2s} \frac{\alpha_s(L)^2}{\alpha_0(L)^2} \left(\mathcal{D}^s f(x_0)\right)^2 + \frac{\kappa^d}{n} \frac{\alpha_0(L^2)}{\omega_{d-1} 2^{(d-2)/2} \alpha_0(L)^2} f(x_0) + o(\kappa^{-2s}) + \frac{o(\kappa^d)}{n}.
\end{aligned}$$

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

$$\begin{aligned}
&\lim_{n\to\infty} n^{2s/(2s+d)} E\left(\hat{f}_n(x_0, \kappa) - f(x_0)\right)^2 \\
&\quad = C^{-2s} \frac{\alpha_s(L)^2}{\alpha_0(L)^2} \left(\mathcal{D}^s f(x_0)\right)^2 + C^d \frac{\alpha_0(L^2)}{\omega_{d-1} 2^{(d-2)/2} \alpha_0(L)^2} f(x_0).
\end{aligned} \tag{3.4}$$

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

$$C^* = \left[A(L) \frac{(\mathcal{D}^s f(x_0))^2}{f(x_0)} \right]^{1/(2s+d)}$$

and $A(L) = 2^{d/2} s \omega_{d-1} \alpha_s(L)^2 / (d\alpha_0(L^2))$. The expression (3.4) evaluated at C^* has the value

$$\begin{aligned}
&(\mathcal{D}^s f(x_0))^{2d/(2s+d)} f(x_0)^{2s/(2s+d)} \alpha_0(L)^{-2} \alpha_s(L)^{2d/(2s+d)} \alpha_0(L^2)^{2s/(2s+d)} \\
&\quad \times \left(\omega_{d-1} 2^{(d-2)/2}\right)^{-2s/(2s+d)} (2s/d)^{(d-2s)/(2s+d)}.
\end{aligned} \tag{3.5}$$

Let us move on to mean integrated squared error. From Lemma 3.3 and Lemma 3.6 it follows that

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

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

$$\begin{aligned}
&\lim_{n\to\infty} n^{2s/(2s+d)}\mathrm{E}\int_{S_d} \left(\hat{f}_n(\cdot,\kappa) - f\right)^2 d\mu \\
&\quad = C^{-2s}\frac{\alpha_s(L)^2}{\alpha_0(L)^2}\int_{S_d} (\mathcal{D}^s f)^2\, d\mu + C^d\frac{\alpha_0(L^2)}{\omega_{d-1}2^{(d-2)/2}\alpha_0(L)^2}.
\end{aligned} \tag{3.6}$$

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

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

and $A(L)$ is as before. The expression (3.6) evaluated at C^* has the value

$$\begin{aligned}
&\left[\int_{S_d} (\mathcal{D}^s f)^2\right]^{d/(2s+d)} \alpha_0(L)^{-2}\alpha_s(L)^{2d/(2s+d)}\alpha_0(L^2)^{2s/(2s+d)} \\
&\qquad \times \left(\omega_{d-1}2^{(d-2)/2}\right)^{-2s/(2s+d)} (2s/d)^{(d-2s)/(2s+d)}.
\end{aligned} \tag{3.7}$$

It can be seen that the formulas (3.5) and (3.7) both depend on L through

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

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

$$L_0(t) = (1 - t^{s/2})I_{[0,1]}(t)$$

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 $K : \mathbf{R}^d \to \mathbf{R}$, $K(x) = L(\|x\|^2)$.

Then by equation (3.1), $\int_{\mathbf{R}^d} K = \omega_{d-1}\frac{1}{2}\alpha_0(L)$ and similarly, $\int_{\mathbf{R}^d} K^2 = \omega_{d-1}\frac{1}{2}\alpha_0(L^2)$. Also, by equation (3.2), $\int_{\mathbf{R}^d} y_1^s K(y)dy = \int_{S_{d-1}} x_1^s d\mu_{d-1}(x)\frac{1}{2}\alpha_s(L)$. Thus

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

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) and Epanechnikov (1969).

3.4 Mean Integrated Absolute Error

In this section 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), Holmström and Klemelä (1992), and in a different form in Hall and Wand (1988). Define

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

where

$$\gamma(u) = \sqrt{2/\pi}\left(u\int_0^u e^{-t^2/2}dt + e^{-u^2/2}\right), u \geq 0.$$

The function A appears also in the Euclidean case. Recall that the definition of the smoothness class $\mathbf{F}_2(s,p)$ is given before Lemma 3.3.

Theorem 3.8 *Let $s \geq 2$ be even. Assume that f is a density in $\mathbf{F}_2(s,1)$. 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

$$z = \alpha_0(L)^{-1}\alpha_s(L)\mathcal{D}^s f,$$

$$w = \left[\frac{\alpha_0(L^2)}{\omega_{d-1}2^{(d-2)/2}\alpha_0(L)^2} f\right]^{1/2}.$$

Proof. Put $B_n = f - \mathrm{E}(\hat{f}_n)$ and $\sigma_n = \sqrt{\mathrm{Var}(\hat{f}_n)}$. Firstly, by an argument similar to that in Devroye and Györfi (1985, Lemma 9, page 91),

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

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, Lemma 12, page 93),

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

by Lemma 3.3 and Lemma 3.7. □

It is seen that by choosing $\kappa = C n^{1/(2s+d)}$,

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

Compare this formula to formula (3.6). 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).

Chapter 4

Empirical Choice of the Smoothing Parameter

In Section 3.3 the asymptotically optimal smoothing parameters were derived. 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 asymptotics for the risk. In this chapter such data–driven devices for choosing the smoothing parameter are presented. The choice will be different depending on whether the risk is measured with the mean squared error or with the mean integrated squared error.

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) in the estimation of density functions. There are many other approaches, for example methods based on cross-validation and bootstrap. These are explained for instance in Cao, Cuevas, and González–Manteiga (1994). 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 empirical smoothing parameter as a statistic $\hat{\kappa}$ which minimizes $\mathrm{E}\left(\mathrm{MISE}(\hat{\kappa})\right)$, where $\mathrm{MISE}(\kappa) = \mathrm{E}\int\left(\hat{f}_n(\cdot,\kappa) - f\right)^2$ and minimum is taken over all random variables. However, in Section 4.2 the optimal empirical 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 4.1 the optimal empirical smoothing parameter is taken to be a random variable $\hat{\kappa}$ which minimizes

$$\mathrm{E} \left(\hat{f}_n(x_0, \hat{\kappa}) - f(x_0) \right)^2 .$$

4.1 Mean Squared Error Criterion

Selection of the smoothing parameter when estimating density at a point has received much less attention compared to smoothing parameter selection when estimating the whole density. In the Euclidean case the problem has been studied at least by Woodroofe (1970), Krieger and Pickands (1981), Sheather (1983), (1986), Hall (1993).

In Section 3.3 it was shown that if f has smoothness index s, the asymptotically optimal smoothing parameter sequence for estimating $f(x_0) > 0$ in the mean squared error sense is

$$\kappa_n^* = C^* n^{1/(2s+d)}, \tag{4.1}$$

where

$$C^* = \left[A(L) \frac{(\mathcal{D}^s f(x_0))^2}{f(x_0)} \right]^{1/(2s+d)}$$

and $A(L) = 2^{d/2} s \omega_{d-1} \alpha_s(L)^2 / (d \alpha_0(L^2))$. The unknown quantities in the formula of C^* will have to be estimated.

Let us assume that there is an initial estimator $\hat{\theta}_{0,n}$ for $f(x_0) \stackrel{def}{=} \mathcal{D}^0 f(x_0)$ and an estimator $\hat{\theta}_{s,n}$ for $\mathcal{D}^s f(x_0)$ such that for $k = 0, s$,

$$\mathrm{E}(\hat{\theta}_{k,n}) - \mathcal{D}^k f(x_0) = o(1) \tag{4.2}$$

and

$$\mathrm{E} \left| \hat{\theta}_{k,n} - \mathrm{E}(\hat{\theta}_{k,n}) \right|^m = o(n^{-ms/(2s+d)}), \tag{4.3}$$

for all $m \in \{1, 2, \ldots\}$. Furthermore, it is assumed that for some $0 \leq \beta < 1$

$$0 \leq \hat{\theta}_{0,n} \leq n^{\beta} \tag{4.4}$$

with probability one when n is large enough. An estimator $\hat{\theta}_{0,n}$ satisfying these conditions will be constructed in Lemma 4.4 and the estimator $\hat{\theta}_{s,n}$ will be constructed in Lemma 5.6. Let

$$\hat{C}_n = \left(A(L) \frac{\hat{\theta}_{s,n}^2 + b_n}{\hat{\theta}_{0,n} + b_n} \right)^{1/(2s+d)} \wedge n^{\gamma},$$

where $b_n = o(1)$, $b_n n^{1-\beta} \to \infty$, and $\gamma > 0$ is arbitrary. The data–driven smoothing parameter is defined as

$$\hat{\kappa}_n = \hat{C}_n n^{1/(2s+d)}. \tag{4.5}$$

Recall that the definition of the smoothness class $\mathbf{F}_1(s, x_0)$ is given before Lemma 3.2.

Theorem 4.1 *Let $s \geq 2$ be even and $x_0 \in S_d$. Let $f \in \mathbf{F}_1(s, x_0)$ be a bounded density which is continuous at x_0 and for which $f(x_0) > 0$ and $\mathcal{D}^s f(x_0) \neq 0$. Assume that estimators $\hat{\theta}_{0,n}$ and $\hat{\theta}_{s,n}$ satisfying conditions (4.2), (4.3) and (4.4) exist. 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*

$$\mathrm{E} \left(\hat{f}_n(x_0, \hat{\kappa}_n) - f(x_0) \right)^2 \sim \mathrm{E} \left(\hat{f}_n(x_0, \kappa_n^*) - f(x_0) \right)^2,$$

where $a_n \sim b_n$ means $\lim_{n \to \infty}(a_n / b_n) = 1$ and κ_n^* was defined in (4.1).

Proof. The structure of the argument is similar to that in Woodroofe (1970). Let us denote $\mu_{0,n} = \mathrm{E}(\hat{\theta}_{0,n})$ and $\mu_{s,n} = \mathrm{E}(\hat{\theta}_{s,n})$. Put

$$C_n = \left(A(L)(\mu_{s,n}^2 + b_n)/(\mu_{0,n} + b_n) \right)^{1/(2s+d)} \wedge n^{\gamma}$$

and $\lambda_n = C_n n^{1/(2s+d)}$. From assumption (4.2) it follows that $\lim_{n \to \infty} C_n = C^*$. Thus, from equation (3.4) it is seen that

$$\mathrm{E} \left(\hat{f}_n(x_0, \kappa_n^*) - f(x_0) \right)^2 \sim \mathrm{E} \left(\hat{f}_n(x_0, \lambda_n) - f(x_0) \right)^2$$

and it remains to prove that

$$\mathrm{E}\left(\hat{f}_n(x_0,\hat{\kappa}_n) - \hat{f}_n(x_0,\lambda_n)\right)^2 = o\left(n^{-2s/(2s+d)}\right).$$

Now

$$\begin{aligned}\mathrm{E}\left(\hat{f}_n(x_0,\hat{\kappa}_n) - \hat{f}_n(x_0,\lambda_n)\right)^2 &= \mathrm{E}\left[\frac{\partial}{\partial\kappa}\hat{f}_n(x_0,\kappa)\bigg|_{\kappa=\xi_n}(\hat{\kappa}_n-\lambda_n)\right]^2 \\ &\le \mathrm{E}^{1/2}\left[n^{1/(2s+d)}\,\frac{\partial}{\partial\kappa}\hat{f}_n(x_0,\kappa)\bigg|_{\kappa=\xi_n}\right]^4 \mathrm{E}^{1/2}(\hat{C}_n - C_n)^4,\end{aligned}$$

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

Next it is proved that

$$\mathrm{E}^{1/2}\left(\hat{C}_n - C_n\right)^4 = o\left(n^{-2s/(2s+d)}\right). \tag{4.6}$$

If $\mathrm{E}|X_n|^m = o(a_n^m)$ for $m = 1, 2, \ldots$, then it is denoted $X_n = o_E(a_n)$. Similarly, 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 assertions in Woodroofe (1970).

Lemma 4.2 **(i)** *If $X_{ni} - Y_{ni} = o_E(a_n)$ and $X_{ni} = O_E(b_n)$ for $i = 1, 2$, then $X_{n1}X_{n2} - Y_{n1}Y_{n2} = o_E(a_n(a_n \vee b_n))$.*

(ii) *If $X_n - Y_n = o_E(a_n)$ and $X_n = O_E(b_n)$, then $X_n^k - Y_n^k = o_E\left(a_n(a_n \vee b_n)^{(k-1)}\right)$ for $k \ge 1$ rational.*

(iii) *Let $X_n \ge b_n > 0$ and $Y_n \ge \delta > 0$ both 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 $k \le 1$.*

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

$$\hat{C}_n - C_n = o_E(a_n). \tag{4.7}$$

Equation (4.6) follows from this.

From assumption (4.4) it follows that $\hat{\theta}_{0,n} + b_n \ge b_n$ for sufficiently large n with probability one. From assumption (4.2) it follows that $\mu_{0,n} + b_n \ge \delta$ for sufficiently large n and for a certain $\delta > 0$. Assumption (4.3) for $k = 0$ says that $\hat{\theta}_{0,n} - \mu_{0,n} =$

$o_E(a_n)$. It holds also that $a_n^r = o(b_n)$, when $r \geq (2s+d)(1-\beta)/s$. Thus, by Lemma 4.2 (iii) it holds that

$$\left(\hat{\theta}_{0,n} + b_n\right)^{-1} - (\mu_{0,n} + b_n)^{-1} = o_E(a_n). \tag{4.8}$$

Assumption (4.3) for $k = s$ says that $\hat{\theta}_{s,n} - \mu_{s,n} = o_E(a_n)$. Because $\mu_{s,n} = O_E(1)$ by assumption (4.2), we have by Lemma 4.2 (ii) that

$$\hat{\theta}_{s,n}^2 - \mu_{s,n}^2 = o_E(a_n). \tag{4.9}$$

By assumption (4.2) for $k = 0$, $(\mu_{0,n} + b_n)^{-1} = O_E(1)$ and by assumption (4.2) for $k = s$, $\mu_{s,n}^2 + b_n = O_E(1)$. Thus we have by equations (4.8), (4.9) and Lemma 4.2 (i) that

$$\frac{\hat{\theta}_{s,n}^2 + b_n}{\hat{\theta}_{0,n} + b_n} - \frac{\mu_{s,n}^2 + b_n}{\mu_{0,n} + b_n} = o_E(a_n). \tag{4.10}$$

By assumption (4.4), $(\hat{\theta}_{s,n}^2 + b_n)(\hat{\theta}_{0,n} + b_n)^{-1} \geq b_n(n^\beta + b_n)^{-1}$ for sufficiently large n with probability one. By assumption (4.2), $(\mu_{s,n}^2 + b_n)(\mu_{0,n} + b_n)^{-1} \geq \delta$ for a certain $\delta > 0$ for sufficiently large n. Now $a_n^r = o(b_n(n^\beta + b_n)^{-1})$, when $r \geq (2s+d)/s$. Thus, by equation (4.10) and Lemma 4.2 (iii),

$$\left(\frac{\hat{\theta}_{s,n}^2 + b_n}{\hat{\theta}_{0,n} + b_n}\right)^{1/(2s+d)} - \left(\frac{\mu_{s,n}^2 + b_n}{\mu_{0,n} + b_n}\right)^{1/(2s+d)} = o_E(a_n).$$

From this we have equation (4.7).

It remains to prove that

$$\mathrm{E}\left[n^{1/(2s+d)} \frac{\partial}{\partial \kappa} \hat{f}_n(x_0, \kappa)\Big|_{\kappa=\xi_n}\right]^4 = O(1), \tag{4.11}$$

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 $Q(L) = -2^{2-d/2}\omega_{d-1}^{-1}\alpha_2(L')\alpha_0(L)^{-2}$ and $c(\kappa)^{-1} = \int_{S_d} L(\kappa^2(1 - x'y))d\mu(y)$. This holds, because by Lemma 2.3, $c(\kappa) \sim \kappa^d R(L)$, where $R(L) = \left[\omega_{d-1}2^{(d-2)/2}\alpha_0(L)\right]^{-1}$ and

$$c'(\kappa) = -c^2(\kappa)\frac{\partial}{\partial \kappa}\left[\int_{S_d} L(\kappa^2(1 - x'y))d\mu(y)\right],$$

where

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

It holds that

$$
\frac{\partial}{\partial\kappa}\hat{f}_n(x_0,\kappa) = \frac{1}{n}\sum_{i=1}^n\left[c'(\kappa)L(\kappa^2(1-x_0'X_i)) + c(\kappa)2\kappa(1-x_0'X_i)L'(\kappa^2(1-x_0'X_i))\right].
$$

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

$$
\begin{aligned}
&\left|\frac{\partial}{\partial\kappa}\hat{f}_n(x_0,\kappa)\bigg|_{\kappa=\xi_n}\right|\\
&\le \left(|c'(\xi_n)| + 2|c(\xi_n)\xi_n^{-1}|\right)\frac{1}{n}\sum_{i=1}^n J(\xi_n^2(1-x_0'X_i))\\
&\le \left(|Q(L)| + 2|R(L)|\right)(1+\epsilon)\frac{\xi_n^{d-1}}{n}\sum_{i=1}^n J(\xi_n^2(1-x_0'X_i))\\
&\le \left(|Q(L)| + 2|R(L)|\right)(1+\epsilon)\frac{s_{2n}^{d-1}}{n}\sum_{i=1}^n J(s_{1n}^2(1-x_0'X_i))\\
&\le \left(|Q(L)| + 2|R(L)|\right)(1+\epsilon)(C^*)^{-1}2^{2d-1}n^{-1/(2s+d)}\frac{s_{1n}^d}{n}\sum_{i=1}^n J(s_{1n}^2(1-x_0'X_i)).
\end{aligned}
$$

Thus, for sufficiently large n,

$$
\begin{aligned}
&\mathrm{E}\left[I_{A_n}n^{1/(2s+d)}\ \frac{\partial}{\partial\kappa}\hat{f}_n(x_0,\kappa)\bigg|_{\kappa=\xi_n}\right]^4\\
&\le \left[\left(|Q(L)| + 2|R(L)|\right)(1+\epsilon)(C^*)^{-1}2^{2d-1}\right]^4\mathrm{E}\left[\frac{s_{1n}^d}{n}\sum_{i=1}^n J(s_{1n}^2(1-x_0'X_i))\right]^4.
\end{aligned}
$$

By Lemma 4.3, which is given after this proof (choose $Z_{ni} = s_{1n}^d J(s_{1n}^2(1-x_0'X_i))$, so that $\mathrm{Var}Z_{ni} = O(s_{1n}^d)$ and $\|Z_{ni}\|_\infty = O(s_{1n}^d)$), and by Lemma 2.4,

$$
\mathrm{E}\left[\frac{s_{1n}^d}{n}\sum_{i=1}^n J(s_{1n}^2(1-x_0'X_i))\right]^4
$$

$$\leq 2^3\mathrm{E}\left[\frac{s_{1n}^d}{n}\sum_{i=1}^{n}J(s_{1n}^2(1-x_0'X_i)) - s_{1n}^d\mathrm{E}J(s_{1n}^2(1-x_0'X_1))\right]^4$$
$$+2^3\left[s_{1n}^d\mathrm{E}J(s_{1n}^2(1-x_0'X_1))\right]^4 = O\left(\left(s_{1n}^d n^{-1}\right)^2\right) + O(1) = O(1).$$

Thus,

$$\mathrm{E}\left[I_{A_n} n^{1/(2s+d)} \frac{\partial}{\partial\kappa}\hat{f}_n(x_0,\kappa)\Big|_{\kappa=\xi_n}\right]^4 = O(1). \tag{4.12}$$

With probability one, for sufficiently large n,

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

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 $|J|$,

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

Thus, for sufficiently large n,

$$\mathrm{E}\left[I_{A_n^c} n^{1/(2s+d)} \frac{\partial}{\partial\kappa}\hat{f}_n(x_0,\kappa)\Big|_{\kappa=\xi_n}\right]^4 \tag{4.13}$$
$$\leq \left(M(|Q(L)|+2|R(L)|)(1+\epsilon)n^{\gamma(d-1)}n^{d/(2s+d)}\right)^4 \mathrm{P}(A_n^c) = o(1),$$

because, for sufficiently large n,

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

for $m \in \{1,2,\ldots\}$ by equation (4.7). Equation (4.11) follows from equations (4.12) and (4.13). □

To construct estimators satisfying the condition (4.3) the following result is needed.

Lemma 4.3 *Let $Z_{n1}, \ldots, Z_{nn}$ be i.i.d,* $\mathrm{Var} Z_{ni} = O(a_n)$*, and* $\|Z_{ni}\|_\infty = O(b_n)$*, where* $\lim_{n\to\infty} a_n = \infty$*,* $\lim_{n\to\infty} b_n = \infty$*, and* $a_n^{-1/2} b_n = o(n^{1/2})$*. Then for $m \geq 4$ even,*

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

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

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

where

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

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

$$|\gamma_{nj}| \leq j^j \mathrm{E}|Y_{n1}|^j \leq j^j \left(\frac{2\|Z_{n1}\|_\infty}{(\mathrm{Var}Z_{n1})^{1/2}}\right)^{j-2} \mathrm{E}|Y_{n1}|^2 = j^j \left(\frac{2\|Z_{n1}\|_\infty}{(\mathrm{Var}Z_{n1})^{1/2}}\right)^{j-2}.$$

Thus,

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

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

$$\begin{aligned}
\mathrm{E}\left|\frac{1}{n}\sum_{i=1}^{n} Z_{ni} - \mathrm{E}Z_{n1}\right|^m &= \left(\frac{\mathrm{Var}Z_{n1}}{n}\right)^{m/2} \mathrm{E}\left|n^{-1/2}\sum_{i=1}^{n} Y_{ni}\right|^m \\
&\leq \sum_{j=0}^{m-2} C(j,m)(\mathrm{Var}Z_{n1})^{(m-j)/2} n^{-(m+j)/2} \|Z_{n1}\|_\infty^j \\
&= O\left(\left(\frac{a_n}{n}\right)^{m/2}\right) \sum_{j=0}^{m-2} O\left(n^{-j/2} a_n^{-j/2} b_n^j\right).
\end{aligned}$$

□

Construction of estimators for $\mathcal{D}^s f(x_0)$ is postponed to the next chapter. An initial estimator for $f(x_0)$ is given in the next lemma.

Lemma 4.4 *Let the density be bounded and continuous at x_0. A kernel estimator $\hat{\theta}_{0,n} = \hat{f}_n(x_0, \kappa_n, L)$ satisfies the conditions (4.2), (4.3) and (4.4) if $L \geq 0$, L is bounded, $\alpha_0(|L|^i) < \infty$ for $i = 1, 2$, and $\kappa_n = n^c$, where $0 < c < 1/(2s+d)$.*

Proof. The consistency condition (4.2) is satisfied by Lemma 2.4. Lemma 4.3 can be applied with $Z_{ni} = c(\kappa_n)L(\kappa_n^2(1 - x_0'X_i))$, so that $\mathrm{Var}Z_{ni} = O(\kappa_n^d)$, $\|Z_{ni}\|_\infty = O(\kappa_n^d)$, and $O((\kappa_n^d n^{-1})^{m/2}) = o(n^{-ms/(2s+d)})$. Thus condition (4.3) is satisfied. Condition (4.4) is satisfied for $d/(2s+d) < \beta < 1$, because $0 \leq \hat{\theta}_{0,n} \leq c(\kappa_n)M$, where M is a bound for L, and $c(\kappa_n) \sim \kappa_n^d R(L)$. □

Condition (4.3) for $k = 0$, $m = 2$, says that $\mathrm{Var}(\hat{\theta}_{0,n}) = o(n^{-2s/(2s+d)})$. The kernel estimator which is asymptotically optimal in the mean squared error sense satisfies $\mathrm{Var}(\hat{f}_n(x_0)) = O(n^{-2s/(2s+d)})$. This explains the condition $c < 1/(2s+d)$ in the previous lemma, because the asymptotically optimal smoothing parameter in the mean squared error sense is $\kappa_n^* = C^* n^{1/(2s+d)}$.

4.2 Mean Integrated Squared Error Criterion

The plug–in method with integrated squared error loss was apparently first introduced by Nadaraya (1974). Later developments are due to for example Scott, Tapia and Thompson (1977), Park and Marron (1990), Sheather and Jones (1991), Hall, Sheather, Jones and Marron (1991), Wand and Jones (1994), Engel, Herrmann and Gasser (1995).

In Section 3.3 it was shown that if f has smoothness index s, the asymptotically optimal smoothing parameter sequence in the mean integrated squared error sense is

$$\kappa_n^* = C^* n^{1/(2s+d)}, \tag{4.14}$$

where

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

and $A(L) = 2^{d/2}\omega_{d-1} s \alpha_s(L)^2/(d\alpha_0(L^2))$.

Let us assume that there is such estimator $\hat{\theta}_{s,n}$ for $\int_{S_d} (\mathcal{D}^s f)^2 d\mu$ that,

$$\mathrm{E}(\hat{\theta}_{s,n}) - \int_{S_d} (\mathcal{D}^s f)^2 d\mu = o(1) \tag{4.15}$$

and

$$\mathrm{E}\left|\hat{\theta}_{s,n} - E(\hat{\theta}_{s,n})\right|^m = o\left(n^{-ms/(2s+d)}\right), \tag{4.16}$$

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

$$\hat{\theta}_{s,n} \geq 0 \tag{4.17}$$

with probability one when n is large enough. Such an estimator is constructed in Lemma 5.12. 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 4.5 *Let $s \geq 2$ be even. Let $f : S_d \to \mathbf{R}$ be a continuous density, $f \in \mathbf{F}_2(s, 2)$, and $\int_{S_d} (\mathcal{D}^s f)^2 \, d\mu > 0$. (The definition of $\mathbf{F}_2$ was given before Lemma 3.3.) Assume that an estimator satisfying conditions (4.15), (4.16) and (4.17) 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 (4.14).

Proof. This theorem can be proved in quite a similar way as Theorem 4.1 and thus only a sketch of proof will be given. Let us denote $\mu_{s,n} = \mathrm{E}(\hat{\theta}_{s,n})$. Put

$$C_n = (A(L)(\mu_{s,n} + b_n))^{1/(2s+d)} \wedge n^\gamma \tag{4.18}$$

and $\lambda_n = C_n n^{1/(2s+d)}$. It has to be proved that

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

Now

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

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

It holds that $\hat{C}_n - C_n = o_E(a_n)$, where $a_n = n^{-s/(2s+d)}$. Thus

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

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

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

because by Jensen's inequality,

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

by a similar application of Lemma 4.3 as in the proof of Theorem 4.1. (Choose $Z_{ni} = Z_{ni}(x) = s_{1n}^d J(s_{1n}^2(1-x'X_i))$, and note that

$$\begin{aligned}
&\mathrm{Var}(Z_{ni}) \leq \mathrm{E}(Z_{ni}^2) = s_{1n}^{2d}\int_{S_d} L^2(s_{1n}^2(1-x'y))f(y)d\mu(y)\\
&= s_{1n}^{2d}\int_{-1}^{1} dt(1-t^2)^{(d-2)/2}L^2(s_{1n}^2(1-t))\int_{T_x} f\left(\phi_x^{-1}(\xi,t)\right)d\mu_{d-1}(\xi)\\
&= s_{1n}^{d}\int_{0}^{2s_{1n}^2} du(2-s_{1n}^{-2}u)^{(d-2)/2}u^{(d-2)/2}L^2(u)\int_{T_x} f\left(\phi_x^{-1}(\xi,1-s_{1n}^{-2}u)\right)d\mu_{d-1}(\xi)\\
&\leq s_{1n}^d\omega_{d-1}2^{(d-2)/2}\alpha_0(L^2)\|f\|_\infty
\end{aligned}$$

and $\|Z_{ni}\|_\infty \le s_{1n}^d M$, where M is a bound for $|J|$, so that the convergence is uniform with respect to $x \in S_d$.) Also, for sufficiently large n,

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

because $P(A_n^c) = o(n^{-ms/(2s+d)})$ for $m \in \{1, 2, \ldots\}$, as shown in the proof of Theorem 4.1. □

The proof of Nadaraya (1974) differs from the proof given here in that he defines the constant C_n in equation (4.18) 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.

Chapter 5

Estimation of Derivatives

In this chapter, two estimators for the Laplacian of a density are constructed. More generally, an estimator for a linear combination of higher order derivatives of a density is constructed, the concept of a derivative being that defined in (2.6). Also, an estimator for an iterated Laplacian of a density is constructed and an estimator for some integral functionals of these quantities will be defined. It has been shown in Section 3.3 that the formula for the asymptotically optimal smoothing parameter in the mean squared error and the mean integrated error sense involves some such functionals of a density.

In Section 5.1 an estimator will be constructed for a linear combination of derivatives of a density. The estimator is a linear combination of kernel estimators. The definition of a class s kernel is generalized to develop a fast convergence rate theory for this estimator. The rate of convergence for the mean squared error and the asymptotically optimal smoothing parameter are given. It is also shown that this estimator can be made to satisfy the conditions of Section 4.1 for the estimator used in the plug–in method.

In Section 5.2 an estimator will be constructed for an iterated Laplacian of a density. The estimator is an iterated Laplacian of the kernel estimator. It should be noted that this estimator is not a kernel estimator. The situation differs from the Euclidean case where the derivative of a kernel estimator is also a kernel estimator. The rate of convergence for the mean squared error and the asymptotically optimal smoothing parameter are given. As a special case, the estimators of Section 5.1 and

5.2 are two different estimators for the Laplacian of a density.

In Section 5.3 an estimator for the inner product of linear combinations of derivatives of a density is constructed. It is shown that this estimator can be made to satisfy the conditions of Section 4.2 for the estimator used in the plug–in method. Secondly, an estimator for the inner product of iterated Laplacians of a density is constructed. It is shown that this estimator achieves $\sqrt{n}$–rate of convergence, when the density is sufficiently smooth. In particular, for the integral of the squared Laplacian of a density, two different estimators will be defined.

5.1 The First Estimator

In this section an estimator for a linear combination of derivatives of a density is constructed when the concept of derivative is defined as in (2.6). Estimation of derivatives with Euclidean data has been studied by Bhattacharya (1967), Singh (1976), (1979), (1987), Sheather (1983), and Donoho and Liu (1991).

The estimator to be defined is a linear combination of kernel estimators. Let $r \geq 0$ be an even integer. Let $a = (a_0, \ldots, a_{r/2}) \in \mathbf{R}^{r/2+1}$ and let $L_i : [0, \infty[\to \mathbf{R}$, $i = 0, \ldots, r/2$, be such that $\alpha_{2i}(|L_i|) < \infty$, $\alpha_{2i}(L_i) \neq 0$. Define

$$\hat{h}_{n,a}(x) = \frac{\kappa^d}{n} \sum_{i=1}^{n} M_{\kappa,a}(x' X_i)$$

where

$$M_{\kappa,a}(t) = \sum_{i=0}^{r/2} a_i C_i(L_i) \kappa^{2i} L_i(\kappa^2(1-t)),\ t \in [-1, 1]$$

$C_i(L_i) = \left[2^{(d-2)/2} \omega_{d-1} \alpha_{2i}(L_i)\right]^{-1}$. When $r = 0$ and $a_0 = 1$, estimator $\hat{h}_{n,a}$ is a kernel estimator up to the constant of normalization. It will be seen that $\hat{h}_{n,a}$ is an estimator for $\sum_{i=0}^{r/2} a_i \mathcal{D}_0^{2i} f$, where $\mathcal{D}_0^{2i} f$ was defined in equation (2.18). If we, however, want to estimate $\sum_{i=0}^{r/2} b_i D^{2i} f$, we have to choose $a_0, \ldots, a_{r/2}$ in such a way that

$$\frac{2^i}{(2i)!} \sum_{j=i}^{r/2} \gamma_{2i,j-i} a_j = b_i,\ i = 0, \ldots, r/2.$$

For example, for the estimation of $D^2 f$ one has to choose $(a_0, a_1) = (\gamma_{0,1}/\gamma_{0,0}, 1/\gamma_{2,0}) = ((d-6)/4, 1)$. The following definition generalizes Definition 3.1.

Definition 5.1 *Let $0 \le r \le s$ be even. A class (r,s) kernel is a measurable function $L : [0,\infty[\to \mathbf{R}$ which satisfies*

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

(ii) $\int_0^{\kappa^2} t^{(r+d-2)/2} L(t)dt = \alpha_r(L) + o(\kappa^{r-s}), \quad \alpha_r(L) \neq 0,$

(iii) $\int_0^{\kappa^2} t^{(2i+d-2)/2} L(t)dt = o(\kappa^{2i-s})$ *for $i = 0, \dots, r/2-1, r/2+1, \dots, s/2-1$.*

A class s kernel in the sense of Definition 3.1 for which $\int_0^{\kappa^2} t^{(d-2)/2} L(t)dt = \alpha_0(L) + o(\kappa^{-s})$ is a class $(0,s)$ kernel in the sense of Definition 5.1. For a class (r,s) kernel, $\alpha_{2i}(L) = 0$ for $i = 0, \dots, r/2-1, r/2+1, \dots, s/2-1$. When L has a compact support, condition (iii) is equivalent to $\alpha_{2i}(L) = 0$ for $i = 0, \dots, r/2-1, r/2+1, \dots, s/2-1$. Also, when L has a compact support, a class $(r, r+2)$ kernel is the same as a class (r,r) kernel. To construct a class (r,s) kernel, a polynomial on $[0,1]$ can be fitted to L, as in Devroye (1987, page 100). Part (i) of the next lemma states the consistency of the estimator $\hat{h}_{n,a}(x_0)$. In part (ii) more smoothness is assumed and thus the rate of convergence is faster.

Lemma 5.2 *Let $0 \le r \le s$ be even and $a = (a_0, \dots, a_{r/2}) \in \mathbf{R}^{r/2+1}$. Let $x_0 \in S_d$. If $s \ge 2$, assume that $f \in \mathbf{F}_1(s, x_0)$. (The definition of $\mathbf{F}_1$ was given before Lemma 3.2.) If $s = 0$, assume that f is bounded and continuous at x_0. Let L_i be a class $(2i, s)$ kernel for $i = 0, \dots, r/2$.*

(i) *If $r = s$, then*

$$\mathrm{E}\left(\hat{h}_{n,a}(x_0)\right) = \sum_{i=0}^{r/2} a_i \mathcal{D}_0^{2i} f(x_0) + o(1)$$

when $\kappa \to \infty$.

(ii) *If $s > r$, then*

$$\mathrm{E}\left(\hat{h}_{n,a}(x_0)\right) = \sum_{i=0}^{r/2} a_i \mathcal{D}_0^{2i} f(x_0) + \kappa^{r-s} a_{r/2} \alpha_r(L_{r/2})^{-1} \alpha_s(L_{r/2}) \mathcal{D}_0^s f(x_0) + o(\kappa^{r-s})$$

when $\kappa \to \infty$.

Proof. Firstly,

$$\begin{aligned}\mathrm{E}\left(\hat{h}_{n,a}(x_0)\right) &= \kappa^d \mathrm{E}\left(M_{\kappa,a}(x_0'X_1)\right) = \kappa^d \int_{S_d} M_{\kappa,a}(x_0'y) f(y) d\mu(y) \\ &= \sum_{i=0}^{r/2} a_i C_i(L_i) \kappa^{2i+d} \int_{S_d} L_i(\kappa^2(1-x_0'y)) f(y) d\mu(y).\end{aligned} \tag{5.1}$$

By equation (2.19), for a class $(2i, s)$ kernel L,

$$\begin{aligned}&\kappa^d \int_{S_d} L(\kappa^2(1-x_0'y)) f(y) d\mu(y) \\ &= \omega_{d-1} 2^{(d-2)/2} \left[\sum_{j=0}^{s/2-1} \kappa^{-2j} \alpha_{2j}(L) \mathcal{D}_0^{2j} f(x_0) + \kappa^{-s} \alpha_s(L) \sum_{j=0}^{s/2-1} \frac{2^j}{(2j)!} \gamma_{2j,s/2-j} D^{2j} f(x_0)\right] \\ &\quad + \int_0^{\pi} \tilde{L}_\kappa^{(s)}(\theta) \tilde{D}^s f(x_0,\theta) d\theta + o(\kappa^{-s}) \\ &= \omega_{d-1} 2^{(d-2)/2} \sum_{j=0}^{s/2} \kappa^{-2j} \alpha_{2j}(L) \mathcal{D}_0^{2j} f(x_0) + o(\kappa^{-s}) \\ &= \omega_{d-1} 2^{(d-2)/2} \left[\kappa^{-2i} \alpha_{2i}(L) \mathcal{D}_0^{2i} f(x_0) + \kappa^{-s} \alpha_s(L) \mathcal{D}_0^{s} f(x_0)\right] + o(\kappa^{-s}),\end{aligned}$$

because, by Lemma 2.10 and Lemma 2.8,

$$\begin{aligned}\int_0^{\pi} \tilde{L}_\kappa^{(s)}(\theta) \tilde{D}^s f(x_0,\theta) d\theta &= \frac{1}{s!} d_s(\kappa) D^s f(x_0) + o(\kappa^{-s}) \\ &= \frac{1}{s!}\left[\omega_{d-1} 2^{(s+d-2)/2} \kappa^{-s} \alpha_s(L) + o(\kappa^{-s})\right] D^s f(x_0) + o(\kappa^{-s}) \\ &= \omega_{d-1} 2^{(d-2)/2} \kappa^{-s} \alpha_s(L) \frac{2^{s/2}}{s!} D^s f(x_0) + o(\kappa^{-s}).\end{aligned}$$

The assertions follow from equation (5.1). □

Let us move on to study the variance of the estimator. The following lemma states that $\kappa^{d-mr} M_{\kappa,a}^m$ is an approximate identity up to the constant of normalization.

Lemma 5.3 *Let $r \geq 0$ be even, $a = (a_0, \ldots, a_{r/2}) \in \mathbf{R}^{r/2+1}$ and $m \geq 1$ integer. Let $L_i : [0,\infty[\to \mathbf{R}$, $i = 0, \ldots, r/2$, be such that*

$$\alpha_{2i}(|L_i|) < \infty, \ \ \alpha_{2i}(L_i) \neq 0, \ \ \alpha_0\left(|L_i|^m\right) < \infty. \tag{5.2}$$

(i) *Then,*

$$\kappa^d \int_{S_d} M^m_{\kappa,a}(x'y)d\mu(y) = \kappa^{mr} a^m_{r/2} \left(\omega_{d-1} 2^{(d-2)/2}\right)^{1-m} \alpha_r(L_{r/2})^{-m} \alpha_0(L^m_{r/2}) + o(\kappa^{mr})$$

when $\kappa \to \infty$.

(ii) *For every* $\delta > 0$,

$$\kappa^d \int_{\|y-x\|>\delta} M^m_{\kappa,a}(x'y)d\mu(y) = o(\kappa^{mr})$$

when $\kappa \to \infty$.

Proof. The proof is similar to the proof of Lemma 2.3. By equation (2.4) and the substitution $t = 1 - \kappa^{-2}u$,

$$\begin{aligned}
\kappa^d \int_{S_d} M^m_{\kappa,a}(x'y)d\mu(y) &= \omega_{d-1}\kappa^d \int_{-1}^{1} M^m_{\kappa,a}(t)(1-t^2)^{(d-2)/2}dt \\
&= \omega_{d-1} \int_0^{2\kappa^2} M^m_{\kappa,a}(1-\kappa^{-2}u)u^{(d-2)/2}(2-\kappa^{-2}u)^{(d-2)/2}du \\
&= \omega_{d-1} \int_0^{2\kappa^2} \left[\sum_{i=0}^{r/2} a_i C_i(L_i)\kappa^{2i} L_i(u)\right]^m u^{(d-2)/2}(2-\kappa^{-2}u)^{(d-2)/2}du.
\end{aligned}$$

The assertion follows from the dominated convergence theorem. The second assertion follows similarly, because

$$\kappa^d \int_{\|y-x\|>\delta} M^m_{\kappa,a}(x'y)d\mu(y) = \omega_{d-1} \int_{\delta^2\kappa^2/2}^{2\kappa^2} M^m_{\kappa,a}(1-\kappa^{-2}u)u^{(d-2)/2}(2-\kappa^{-2}u)^{(d-2)/2}du.$$

□

The following lemma describes how the convolution of $M^m_{\kappa,a}$ and f approximates f.

Lemma 5.4 *Let* $f : S_d \to \mathbf{R}$ *be continuous at* $x_0 \in S_d$ *and bounded. Let* $r \geq 0$ *be an even integer and* $a = (a_0, \ldots, a_{r/2}) \in \mathbf{R}^{r/2+1}$. *Let* $L_i : [0, \infty[\to \mathbf{R}$, $i = 0, \ldots, r/2$, *be such that condition (5.2) is satisfied. Then,*

$$\begin{aligned}
&\kappa^d \int_{S_d} M^m_{\kappa,a}(x_0'y)f(y)d\mu(y) \\
&= \kappa^{mr} a^m_{r/2} \left(\omega_{d-1} 2^{(d-2)/2}\right)^{1-m} \alpha_r(L_{r/2})^{-m} \alpha_0(L^m_{r/2}) f(x_0) + o(\kappa^{mr}),
\end{aligned}$$

when $\kappa \to \infty$.

Proof. In view of Lemma 5.3, the proof is similar to the proof of Lemma 2.4. □

Now we are ready to give the asymptotic variance of $\hat{h}_{n,a}$.

Lemma 5.5 *Let $f : S_d \to \mathbf{R}$ be a bounded density which is continuous at $x_0 \in S_d$. Let $r \geq 0$ be even, $a = (a_0, \ldots, a_{r/2}) \in \mathbf{R}^{r/2+1}$. Let $L_i : [0, \infty[\to \mathbf{R}$, $i = 0, \ldots, r/2$ be such that condition (5.2) is satisfied for $m = 1, 2$. Then*

$$\operatorname{Var}\left(\hat{h}_{n,a}(x_0)\right) = \frac{\kappa^{2r+d}}{n} \frac{a_{r/2}^2 \alpha_0(L_{r/2}^2)}{\omega_{d-1} 2^{(d-2)/2} \alpha_r(L_{r/2})^2} f(x_0) + \frac{o(\kappa^{2r+d})}{n}$$

when $\kappa \to \infty$.

Proof. Firstly,

$$\begin{aligned} \operatorname{Var}\left(\hat{h}_{n,a}(x_0)\right) &= n^{-1}\kappa^{2d}\operatorname{Var}(M_{\kappa,a}(x_0'X_1)) \\ &= n^{-1}\kappa^{2d}\left\{\mathrm{E}\, M_{\kappa,a}^2(x_0'X_1) - \left[\mathrm{E}\, M_{\kappa,a}(x_0'X_1)\right]^2\right\}. \end{aligned}$$

By Lemma 5.4,

$$\begin{aligned} \kappa^d \mathrm{E}\, M_{\kappa,a}^2(x_0'X_1) &= \kappa^d \int_{S_d} M_{\kappa,a}^2(x_0'y) f(y) d\mu(y) \\ &= \kappa^{2r} a_{r/2}^2 \left(\omega_{d-1} 2^{(d-2)/2} \alpha_r(L_{r/2})^2\right)^{-1} \alpha_0(L_{r/2}^2) f(x_0) + o(\kappa^{2r}), \end{aligned}$$

and

$$\kappa^{2d}\left[\mathrm{E} M_{\kappa,a}(x_0'X_1)\right]^2 = \left[\kappa^d \int_{S_d} M_{\kappa,a}(x_0'y) f(y) d\mu(y)\right]^2 = \left[O(\kappa^r)\right]^2 = o(\kappa^{2r+d}).$$

□

From Lemma 5.2 (ii) and Lemma 5.5 it follows that

$$\begin{aligned} &\mathrm{E}\left(\hat{h}_{n,a}(x_0,\kappa) - \sum_{i=0}^{r/2} a_i \mathcal{D}_0^{2i} f(x_0)\right)^2 \\ &= \left(\mathrm{E}(\hat{h}_{n,a}(x_0,\kappa)) - \sum_{i=0}^{r/2} a_i \mathcal{D}_0^{2i} f(x_0)\right)^2 + \operatorname{Var}\left(\hat{h}_{n,a}(x_0,\kappa)\right) \\ &= \kappa^{2(r-s)} \frac{a_{r/2}^2 \alpha_s(L_{r/2})^2}{\alpha_r(L_{r/2})^2} \left(\mathcal{D}_0^s f(x_0)\right)^2 + \frac{\kappa^{2r+d}}{n} \frac{a_{r/2}^2 \alpha_0(L_{r/2}^2)}{\omega_{d-1} 2^{(d-2)/2} \alpha_r(L_{r/2})^2} f(x_0) \\ &\quad + o(\kappa^{2(r-s)}) + \frac{o(\kappa^{2r+d})}{n}. \end{aligned}$$

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

$$\lim_{n\to\infty} n^{2(s-r)/(2s+d)} E\left(\hat{h}_{n,a}(x_0,\kappa) - \sum_{i=0}^{r/2} a_i \mathcal{D}_0^{2i} f(x_0)\right)^2$$
$$= C^{2(r-s)} \frac{a_{r/2}^2 \alpha_s(L_{r/2})^2}{\alpha_r(L_{r/2})^2} \left(\mathcal{D}_0^s f(x_0)\right)^2 + C^{2r+d} \frac{a_{r/2}^2 \alpha_0(L_{r/2}^2)}{\omega_{d-1} 2^{(d-2)/2} \alpha_r(L_{r/2})^2} f(x_0).$$

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

$$C^* = \left[A(L) \frac{\left(\mathcal{D}_0^s f(x_0)\right)^2}{f(x_0)}\right]^{1/(2s+d)}$$

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

$$a_{r/2}^2 \left(\mathcal{D}_0^s f(x_0)\right)^{2(2r+d)/(2s+d)} f(x_0)^{2(s-r)/(2s+d)} \alpha_r(L_{r/2})^{-2} \alpha_s(L_{r/2})^{2(2r+d)/(2s+d)}$$
$$\times \alpha_0(L_{r/2}^2)^{2(s-r)/(2s+d)} \left(\omega_{d-1} 2^{(d-2)/2}\right)^{2(r-s)/(2s+d)} \left(\frac{2(s-r)}{2r+d}\right)^{(4r-2s+d)/(2s+d)}.$$

Let us note that the estimator $\hat{h}_{n,a}(x_0)$ can be made to satisfy the conditions (4.2) and (4.3) for an estimator of $\mathcal{D}^s f(x_0)$, $s \geq 2$ even.

Lemma 5.6 *Let the density be bounded, continuous at x_0, and belong to $\mathbf{F}_1(s, x_0)$. Let $\hat{\theta}_{s,n} = \hat{h}_{n,a}(x_0)$, where $a = (a_0, \ldots, a_{s/2}) \in \mathbf{R}^{s/2+1}$, $a_0 = -\gamma_{0,s/2}$ is defined in Lemma 2.8, $a_i = 0$ for $i = 1, \ldots, s/2-1$, and $a_{s/2} = 1$. Let $L_i : [0,\infty[\to \mathbf{R}$, $i = 0, s/2$, be bounded class $(2i, s)$ kernels and such that condition (5.2) is satisfied for $m = 1, 2$. Let $\kappa_n = n^c$, where $0 < c < d/(2s+d)^2$. Then conditions (4.2) and (4.3) are satisfied.*

Proof. The consistency condition (4.2) is satisfied by Lemma 5.2 (i). To prove the satisfaction of condition (4.3), apply Lemma 4.3 with $Z_{ni} = \kappa_n^d M_{\kappa_n,a}(x_0' X_i)$, $i = 1, \ldots, n$. Now $\mathrm{Var} Z_{ni} = O(\kappa_n^{2s+d})$, $\|Z_{ni}\|_\infty = O(\kappa_n^{s+d})$, and $O((\kappa_n^{2s+d} n^{-1})^{m/2}) = o(n^{-ms/(2s+d)})$. Thus condition (4.3) is satisfied. □

Assume that the density f belongs to $\mathbf{F}_1(2, x_0)$ and $\mathbf{F}_1(s, x_0)$, where $s \geq 4$ is even. If it is decided to use a non-negative kernel in the kernel estimator, then

$D^2 f(x_0)$ needs to be estimated in order to apply the plug–in method. On page 54 it was shown that in the mean squared error sense asymptotically optimal smoothing parameter of the estimator $\hat{h}_{n,a}$ has the form $Cn^{1/(2s+d)}$. However, Lemma 5.6 says that condition (4.3) is satisfied if one uses $\tilde{C}n^c$, where $0 < c < d/(d+4)^2$. According to this, one has to choose $c < 1/(2s+d)$, when $s = 4$ or $s = 6$, $d \leq 4$, or $s = 8$, $d = 2$. The choice $c = 1/(2s+d)$ can be made only when $s = 6$, $d > 4$ or $s = 8$, $d > 2$, or $s \geq 10$.

5.2 The Second Estimator

In this section an estimator for $\Delta^{r/2} f(x_0)$, $r \geq 0$ even, will be constructed. By Lemma 2.1, $D^2 f = d^{-1} \Delta f$ for functions f whose radial extensions and their partial derivatives are differentiable. Thus $D^2 f$ will be estimated again, now with a different estimator than in the previous section.

The estimator to be defined is the Laplacian of the kernel estimator. Let $r \geq 0$ be even. Let $L : [0, \infty[\to \mathbf{R}$ be such that $\alpha_0(|L|) < \infty$ and L has r derivatives. Define

$$\hat{g}_n(x) = \Delta^{r/2} \left(\hat{f}_n(x, \kappa, L) \right) = \frac{c(\kappa)}{n} \sum_{i=1}^{n} L^{(r,\kappa)}(x' X_i),$$

where $L^{(r,\kappa)} : [-1, 1] \to \mathbf{R}$ is defined by $L^{(0,\kappa)}(t) = L_\kappa(t) = L(\kappa^2(1-t))$ and

$$L^{(r,\kappa)}(t) = \left[(1-t^2) \frac{\partial^2}{\partial t^2} - dt \frac{\partial}{\partial t} \right] L^{(r-2,\kappa)}(t)$$

if $r \geq 2$. When $r = 0$, $\hat{g}_n$ is the usual kernel estimator of a density. Part (i) of the next lemma states the consistency of the estimator $\hat{g}_n(x_0)$. In part (ii) more smoothness is assumed and thus the rate of convergence is faster.

Lemma 5.7 *Let $r \geq 0$ be even and let L be a class 0 kernel which has r derivatives.*

(i) *Let $\Delta^{r/2} f$ be continuous at $x_0 \in S_d$ and assume $\Delta^{r/2} f$ is bounded. Then,*

$$\mathrm{E}\left(\hat{g}_n(x_0) \right) = \Delta^{r/2} f(x_0) + o(1)$$

when $\kappa \to \infty$.

(ii) *Let $s > r$ be even. Assume that $\Delta^{r/2} f \in \mathbf{F}_1(s-r, x_0)$. (The definition of $\mathbf{F}_1$ was given before Lemma 3.2.) Let L be a class $s-r$ kernel. Then*

$$\mathrm{E}\left(\hat{g}_n(x_0)\right) = \Delta^{r/2} f(x_0) + \kappa^{r-s} \alpha_0(L)^{-1} \alpha_{s-r}(L) \mathcal{D}^{s-r} \Delta^{r/2} f(x_0) + o(\kappa^{r-s})$$

when $\kappa \to \infty$.

Proof. Because the Laplace operator is symmetric,

$$\begin{aligned}\mathrm{E}\left(\hat{g}_n(x_0)\right) &= c(\kappa)\mathrm{E}(L^{(r,\kappa)}(x_0' X_1)) = c(\kappa) \int_{S_d} L^{(r,\kappa)}(x_0' y) f(y) d\mu(y) \\ &= c(\kappa) \int_{S_d} L(\kappa^2(1 - x_0' y)) \Delta^{r/2} f(y) d\mu(y).\end{aligned}$$

Assertion (i) follows from Lemma 2.4. Assertion (ii) follows from Lemma 3.2. □

Let us move on to study the variance of the estimator. The next lemma gives a formula for the iterated Laplacian of a symmetric function.

Lemma 5.8 *Let $s \geq 2$ be even, $\eta \in S_d$ and $g : [-1, 1] \to \mathbf{R}$. Then*

$$\Delta^{s/2}\left[g(x'\eta)\right] = \sum_{i=1}^{s} P_{s,i}(x'\eta) g^{(i)}(x'\eta),$$

where $P_{s,i}$ is a polynomial of degree i. Furthermore, for $i = s/2, \ldots, s$,

$$P_{s,i}(t) = (1 - t^2)^{i - s/2} Q_{s,i}(t),$$

where $Q_{s,i}$ is a polynomial of degree $s - i$.

Proof. The proof is by induction. Assume that the assertion holds for s. By equation (2.5),

$$\begin{aligned}\Delta^{s/2+1}\left[g(x'\eta)\right] &= \Delta\Delta^{s/2}\left[g(x'\eta)\right] = \Delta\left[\sum_{i=1}^{s} P_{s,i}(x'\eta) g^{(i)}(x'\eta)\right] \\ &= \left[(1-t^2)\frac{\partial^2}{\partial t^2} - dt\frac{\partial}{\partial t}\right]\left[\sum_{i=1}^{s} P_{s,i}(t) g^{(i)}(t)\right]\Bigg|_{t=x'\eta}.\end{aligned}$$

Put $\phi(t) = 1 - t^2$ and $\psi(t) = -dt$. Now

$$\begin{aligned}
&\left[\phi(t)\frac{\partial^2}{\partial t^2} + \psi(t)\frac{\partial}{\partial t}\right]\left[\sum_{i=1}^{s} P_{s,i}(t)g^{(i)}(t)\right] \\
&= \sum_{i=1}^{s}\Big[\phi(t)\left(P_{s,i}(t)g^{(i+2)}(t) + 2P_{s,i}^{(1)}(t)g^{(i+1)}(t) + P_{s,i}^{(2)}(t)g^{(i)}(t)\right) \\
&\quad + \psi(t)\left(P_{s,i}(t)g^{(i+1)}(t) + P_{s,i}^{(1)}(t)g^{(i)}(t)\right)\Big] \\
&= \phi(t)P_{s,s}(t)g^{(s+2)}(t) + \left[\phi(t)\left(P_{s,s-1}(t) + 2P_{s,s}^{(1)}(t)\right) + \psi(t)P_{s,s}(t)\right]g^{(s+1)}(t) \\
&\quad + \sum_{i=1}^{s}\left[\phi(t)\left(P_{s,i-2}(t) + 2P_{s,i-1}^{(1)} + P_{s,i}^{(2)}(t)\right) + \psi(t)\left(P_{s,i-1}(t) + P_{s,i}^{(1)}(t)\right)\right]g^{(i)}(t),
\end{aligned}$$

where $P_{s,0} = 0 = P_{s,-1}$. The first assertion has been proved. Let us examine the polynomials $P_{s+2,i}$, $i = s/2+2, \ldots, s+2$. Firstly,

$$P_{s+2,s+2} = \phi P_{s,s} = \phi\phi^{s/2}Q_{s,s} = \phi^{s/2+1}Q_{s+2,s+2}$$

where $Q_{s+2,s+2} = Q_{s,s} = 1$ is a polynomial of degree 0. Secondly, $\phi P_{s,s-1} = \phi\phi^{s/2-1}Q_{s,s-1} = \phi^{s/2}Q_{s,s-1}$,

$$\begin{aligned}
2\phi P_{s,s}^{(1)} &= 2\phi\left(\phi^{s/2}Q_{s,s}\right)^{(1)} = 2\phi\left(\phi^{s/2}Q_{s,s}^{(1)} + s/2\phi^{s/2-1}\phi^{(1)}Q_{s,s}\right) \\
&= \phi^{s/2}2\left(\phi Q_{s,s}^{(1)} + s/2\phi^{(1)}Q_{s,s}\right)
\end{aligned}$$

and $\psi P_{s,s} = \phi^{s/2}\psi Q_{s,s}$. Thus $P_{s+2,s+1} = \phi^{s/2}Q_{s+2,s+1}$, where

$$Q_{s+2,s+1} = Q_{s,s-1} + 2\left(\phi Q_{s,s}^{(1)} + s/2\phi^{(1)}Q_{s,s}\right) + \psi Q_{s,s}$$

is a polynomial of degree 1. Thirdly, for $i \in \{s/2+2, \ldots, s\}$,

$$\phi P_{s,i-2} = \phi\phi^{i-2-s/2}Q_{s,i-2} = \phi^{i-1-s/2}Q_{s,i-2},$$

$$\begin{aligned}
2\phi P_{s,i-1}^{(1)} &= 2\phi\left(\phi^{i-1-s/2}Q_{s,i-1}\right)^{(1)} \\
&= 2\phi\left(\phi^{i-1-s/2}Q_{s,i-1}^{(1)} + (i-1-s/2)\phi^{i-2-s/2}\phi^{(1)}Q_{s,i-1}\right) \\
&= \phi^{i-1-s/2}2\left(\phi Q_{s,i-1}^{(1)} + (i-1-s/2)\phi^{(1)}Q_{s,i-1}\right),
\end{aligned}$$

$$\begin{aligned}
\phi P_{s,i}^{(2)} &= \phi\left(\phi^{i-s/2}Q_{s,i}\right)^{(2)} = \phi\left(\phi^{i-s/2}Q_{s,i}^{(2)} + 2(\phi^{i-s/2})^{(1)}Q_{s,i}^{(1)} + (\phi^{i-s/2})^{(2)}Q_{s,i}\right)\\
&= \phi\Big[\phi^{i-s/2}Q_{s,i}^{(2)} + 2(i-s/2)\phi^{i-1-s/2}\phi^{(1)}Q_{s,i}^{(1)}\\
&\quad + (i-s/2)\phi^{i-2-s/2}\left((i-1-s/2)(\phi^{(1)})^2 + \phi\phi^{(2)}\right)Q_{s,i}\Big]\\
&= \phi^{i-1-s/2}\Big[\phi^2 Q_{s,i}^{(2)} + 2(i-s/2)\phi\phi^{(1)}Q_{s,i}^{(1)}\\
&\quad + (i-s/2)\left((i-1-s/2)(\phi^{(1)})^2 + \phi\phi^{(2)}\right)Q_{s,i}\Big],
\end{aligned}$$

$\psi P_{s,i-1} = \phi^{i-1-s/2}\psi Q_{s,i-1}$ and

$$\begin{aligned}
\psi P_{s,i}^{(1)} &= \psi\left(\phi^{i-s/2}Q_{s,i}\right)^{(1)} = \psi\left(\phi^{i-s/2}Q_{s,i}^{(1)} + (i-s/2)\phi^{i-1-s/2}\phi^{(1)}Q_{s,i}\right)\\
&= \phi^{i-1-s/2}\left(\psi\phi Q_{s,i}^{(1)} + (i-s/2)\psi\phi^{(1)}Q_{s,i}\right).
\end{aligned}$$

Thus, for $i \in \{s/2+2, \ldots, s\}$, $P_{s+2,i} = \phi^{i-1-s/2}Q_{s+2,i}$, where

$$\begin{aligned}
Q_{s+2,i} &= Q_{s,i-2} + 2\left(\phi Q_{s,i-1}^{(1)} + (i-1-s/2)\phi^{(1)}Q_{s,i-1}\right) + \phi^2 Q_{s,i}^{(2)}\\
&\quad + 2(i-s/2)\phi\phi^{(1)}Q_{s,i}^{(1)} + (i-s/2)\left((i-1-s/2)(\phi^{(1)})^2 + \phi\phi^{(2)}\right)Q_{s,i}\\
&\quad + \psi Q_{s,i-1} + \psi\phi Q_{s,i}^{(1)} + (i-s/2)\psi\phi^{(1)}Q_{s,i}
\end{aligned}$$

is a polynomial of degree $s+2-i$. □

The following lemma states that $\kappa^d\left(d_0(\kappa)^{-1}L^{(r,\kappa)}\right)^m$ is an approximate identity up to the constant of normalization.

Lemma 5.9 *Let $r \geq 0$ be even and $m \geq 1$ an integer. Let $L : [0,\infty[\to \mathbf{R}$ be such that*

$$\begin{aligned}
\alpha_0(|L|) &< \infty,\\
\alpha_0\left(|L^{(i)}|^m\right) &< \infty,\ \text{for } i = 1, \ldots, r/2,\\
\alpha_{2m(i-r/2)}\left(|L^{(i)}|^m\right) &< \infty,\ \text{for } i = r/2+1, \ldots, r.
\end{aligned} \tag{5.3}$$

(i) *Then,*

$$\begin{aligned}
&\kappa^d \int_{S_d}\left(d_0(\kappa)^{-1}L^{(r,\kappa)}(x'y)\right)^m d\mu(y)\\
&= \kappa^{mr}\left(\omega_{d-1}2^{(d-2)/2}\right)^{1-m}\alpha_0(L)^{-m}\beta_{r,m}(L) + o(\kappa^{mr})
\end{aligned}$$

when $\kappa \to \infty$, where

$$\beta_{r,m}(L) = \int_0^\infty t^{(d-2)/2} \left\{ \sum_{i=r/2}^{r} Q_{r,i}(1)(-1)^i (2t)^{i-r/2} L^{(i)}(t) \right\}^m dt.$$

(ii) *For every $\delta > 0$,*

$$\kappa^d \int_{\|y-x\|>\delta} \left(d_0(\kappa)^{-1} L^{(r,\kappa)}(x'y)\right)^m d\mu(y) = o(\kappa^{mr})$$

when $\kappa \to \infty$.

Proof. By equation (2.4) and the substitution $t = 1 - \kappa^{-2}u$,

$$\begin{aligned}\kappa^d \int_{S_d} \left(L^{(r,\kappa)}(x'y)\right)^m d\mu(y) &= \kappa^d \omega_{d-1} \int_{-1}^{1} \left(L^{(r,\kappa)}(t)\right)^m (1-t^2)^{(d-2)/2} dt \\ &= \omega_{d-1} \int_0^{2\kappa^2} \left(L^{(r,\kappa)}(1-\kappa^{-2}u)\right)^m u^{(d-2)/2} (2-\kappa^{-2}u)^{(d-2)/2} du.\end{aligned}$$

By Lemma 5.8,

$$L^{(r,\kappa)}(t) = \sum_{i=1}^{r} P_{r,i}(t)(-1)^i \kappa^{2i} L^{(i)}(\kappa^2(1-t))$$

$$= \sum_{i=1}^{r/2-1} P_{r,i}(t)(-1)^i \kappa^{2i} L_i^{(i)}(\kappa^2(1-t)) + \sum_{i=r/2}^{r} (1-t^2)^{i-r/2} Q_{r,i}(t)(-1)^i \kappa^{2i} L_i^{(i)}(\kappa^2(1-t)).$$

Thus,

$$\begin{aligned}\left(L^{(r,\kappa)}(1-\kappa^{-2}u)\right)^m &= \kappa^{mr} \left\{ \sum_{i=1}^{r/2-1} P_{r,i}(1-\kappa^{-2}u)(-1)^i \kappa^{2i-r} L^{(i)}(u) \right. \\ &\left. + \sum_{i=r/2}^{r} u^{i-r/2} (2-\kappa^{-2}u)^{i-r/2} Q_{r,i}(1-\kappa^{-2}u)(-1)^i L^{(i)}(u) \right\}^m .\end{aligned}$$

The first assertion follows from the dominated convergence theorem and the fact that $d_0(\kappa) \sim \omega_{d-1} 2^{(d-2)/2} \alpha_0(L)$ (by Lemma 2.3). The second assertion follows similarly, because

$$\begin{aligned}&\kappa^d \int_{\|y-x\|>\delta} \left(L^{(r,\kappa)}(x'y)\right)^m d\mu(y) \\ &= \omega_{d-1} \int_{\delta^2\kappa^2/2}^{2\kappa^2} \left(L^{(r,\kappa)}(1-\kappa^{-2}u)\right)^m u^{(d-2)/2} (2-\kappa^{-2}u)^{(d-2)/2} du.\end{aligned}$$

□

The following lemma describes how the convolution of $(L^{(r,\kappa)})^m$ and f approximates f.

Lemma 5.10 *Let $f : S_d \to \mathbf{R}$ be continuous at $x_0 \in S_d$ and assume f is bounded. Let $r \geq 0$ be even. Let $L : [0, \infty[\to \mathbf{R}$ be such that condition (5.3) is satisfied. Then,*

$$\begin{aligned}&\kappa^d \int_{S_d} \left(d_0(\kappa)^{-1} L^{(r,\kappa)}(x_0' y)\right)^m f(y) d\mu(y) \\ = \;& \kappa^{mr} \left(\omega_{d-1} 2^{(d-2)/2}\right)^{1-m} \alpha_0(L)^{-m} \beta_{r,m}(L) f(x_0) + o(\kappa^{mr}),\end{aligned}$$

when $\kappa \to \infty$, where $\beta_{r,m}(L)$ was defined in Lemma 5.9.

Proof. In view of Lemma 5.9, the proof is similar to the proof of Lemma 2.4. □

Now we are ready to give the asymptotic variance of $\hat{g}_n$.

Lemma 5.11 *Let $f : S_d \to \mathbf{R}$ be a bounded density which is continuous at $x_0 \in S_d$. Let $r \geq 0$ be even. Let L be such that the condition (5.3) is satisfied for $m = 1, 2$. Then*

$$\mathrm{Var}\left(\hat{g}_n(x_0)\right) = \frac{\kappa^{2r+d}}{n} \frac{\beta_{r,2}(L)}{\omega_{d-1} 2^{(d-2)/2} \alpha_0(L)^2} f(x_0) + \frac{o(\kappa^{2r+d})}{n}$$

when $\kappa \to \infty$, where $\beta_{r,m}(L)$ was defined in Lemma 5.9.

Proof. The proof follows from Lemma 5.10 in the same way as Lemma 5.5 was proved. □

From Lemma 5.7 (ii) and Lemma 5.11 it follows that

$$\begin{aligned}&\mathrm{E}\left(\hat{g}_n(x_0, \kappa) - \Delta^{r/2} f(x_0)\right)^2 = \left(\mathrm{E}(\hat{g}_n(x_0, \kappa)) - \Delta^{r/2} f(x_0)\right)^2 + \mathrm{Var}\left(\hat{g}_n(x_0, \kappa)\right) \\ = \;& \kappa^{2(r-s)} \frac{\alpha_{s-r}(L)^2}{\alpha_0(L)^2} \left(\mathcal{D}^{s-r} \Delta^{r/2} f(x_0)\right)^2 + \frac{\kappa^{2r+d}}{n} \frac{\beta_{r,2}(L)}{\omega_{d-1} 2^{(d-2)/2} \alpha_0(L)^2} f(x_0) \\ & + o(\kappa^{2(r-s)}) + \frac{o(\kappa^{2r+d})}{n}.\end{aligned}$$

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

$$\begin{aligned}&\lim_{n\to\infty} n^{2(s-r)/(2s+d)} E\left(\hat{g}_n(x_0, \kappa) - \Delta^{r/2} f(x_0)\right)^2 \\ &= C^{2(r-s)} \frac{\alpha_{s-r}(L)^2}{\alpha_0(L)^2} \left(\mathcal{D}^{s-r} \Delta^{r/2} f(x_0)\right)^2 + C^{2r+d} \frac{\beta_{r,2}(L)}{\omega_{d-1} 2^{(d-2)/2} \alpha_0(L)^2} f(x_0).\end{aligned}$$

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

$$C^* = \left[A(L) \frac{\left(\mathcal{D}^{s-r}\Delta^{r/2} f(x_0)\right)^2}{f(x_0)} \right]^{1/(2s+d)}$$

and $A(L) = 2^{d/2}(s-r)\omega_{d-1}\alpha_{s-r}(L)^2/((2r+d)\beta_{r,2}(L^2))$. The expression evaluated at C^* has the value

$$\left(\mathcal{D}^{s-r}\Delta^{r/2} f(x_0)\right)^{2(2r+d)/(2s+d)} f(x_0)^{2(s-r)/(2s+d)} \alpha_0(L)^{-2} \alpha_{s-r}(L)^{2(2r+d)/(2s+d)}$$
$$\times \beta_{r,2}(L)^{2(s-r)/(2s+d)} \left(\omega_{d-1} 2^{(d-2)/2}\right)^{2(r-s)/(2s+d)} \left(\frac{2(s-r)}{2r+d}\right)^{(4r-2s+d)/(2s+d)} .$$

5.3 Estimation of Some Integral Functionals of a Density

The asymptotically optimal smoothing parameter in the mean integrated squared error sense was given in (4.14). It is seen that this smoothing parameter depends on the unknown density f through

$$\theta_s(f) = \int_{S_d} (\mathcal{D}^s f)^2 d\mu,$$

where $\mathcal{D}^s f$ is as defined in equation (2.16). An estimator will be constructed for a functional,

$$\theta_a(f) = \int_{S_d} \left(\sum_{i=0}^{s/2} a_i \mathcal{D}_0^{2i} f \right)^2 d\mu,$$

where $s \geq 0$ is even, $a = (a_0, \ldots, a_{s/2}) \in \mathbf{R}^{s/2+1}$, and $\mathcal{D}_0 f$ was defined in equation (2.18). For a suitable choice of a we have the equality $\theta_a(f) = \theta_s(f)$. Estimation of integrated squared derivatives of a density with Euclidean data has been studied by Hall and Marron (1987), Bickel and Ritov (1988), Jones and Sheather (1991). More general integral functionals of derivatives of a density have been studied in the Euclidean case by Laurent (1993), and Birgé and Massart (1995).

Let

$$\hat{\theta}_{a,n} = \frac{1}{n(n-1)} \sum_{i,j=1}^{n} V_{a,\kappa}(X_i, X_j),$$

where

$$V_{a,\kappa}(x,y) = \kappa^{2d} \int_{S_d} M_{\kappa,a}(x'z) M_{\kappa,a}(y'z) d\mu(z),$$

with $M_{\kappa,a}(t) = \sum_{i=0}^{s/2} a_i C_i(L_i) \kappa^{2i} L_i(\kappa^2(1-t))$, $t \in [0,1]$, and $C_i(L_i) = \left[2^{(d-2)/2}\omega_{d-1}\alpha_{2i}(L_i)\right]^{-1}$, as defined in Section 5.1. Now $\hat{\theta}_{a,n}$ equals $\frac{n}{n-1}\int_{S_d}(\hat{h}_{a,n})^2 d\mu$, where $\hat{h}_{a,n}$ was defined in Section 5.1. Let us prove that $\hat{\theta}_{a,n}$ can be made to satisfy the conditions (4.15), (4.16) and (4.17) for the plug–in method.

Lemma 5.12 *Let $s \geq 2$ be an even integer. Let $a \in \mathbf{R}^{s/2+1}$ be such that $\theta_a(f) = \theta_s(f)$. Let $f \in \mathbf{F}_2(s,2)$ be bounded. (The definition of $\mathbf{F}_2$ was given before Lemma 3.3.) Let L_i be a class $(2i,s)$ kernel with support on $[0,1]$ such that $\alpha_0(L_i^2) < \infty$, $i = 0,\ldots,s/2$. Let $\kappa = n^c$, where $0 < c < d/((2s+d)(4s+d))$. Then $\hat{\theta}_{a,n}$ satisfies the conditions (4.15), (4.16), and (4.17).*

Proof. Condition (4.17) is satisfied because $\hat{\theta}_{a,n}$ equals $\frac{n}{n-1}\int_{S_d}(\hat{h}_{a,n})^2 d\mu$, where $\hat{h}_{a,n}$ was defined in Section 5.1.

It can be proved, in a similar manner as in Lemma 5.2 (i) but this time applying Lemma 2.11 (i), that

$$\left\| \kappa^d f * M_{\kappa,a} - \sum_{i=0}^{s/2} a_i \mathcal{D}_0^{2i} f \right\|_2 = o(1),$$

when $\kappa \to \infty$, where $f * M_{\kappa,a}(z) = \int_{S_d} f(y) M_{\kappa,a}(y'z) d\mu(y)$. Thus,

$$\begin{aligned}
\mathrm{E}\,\hat{\theta}_{a,n} &= \mathrm{E}\,V_{a,\kappa}(X_1,X_2) + \frac{1}{n-1} V_{a,\kappa}(x,x) \\
&= \int_{S_d}\int_{S_d} V_{a,\kappa}(x,y) f(x) f(y) d\mu(x) d\mu(y) + o(1) \\
&= \int_{S_d} \left[\kappa^d f * M_{\kappa,a}(z)\right]^2 d\mu(z) + o(1) = \theta_a(f) + o(1),
\end{aligned}$$

where $\frac{1}{n-1} V_{a,\kappa}(x,x) = \frac{1}{n-1}\kappa^{2d}\int_{S_d} M_{\kappa,a}^2(x'z) d\mu(z) = O(\kappa^d n^{-1}) = o(1)$. Thus $\hat{\theta}_{a,n}$ satisfies condition (4.15).

Let us move on to condition (4.16). If the support of $L : [0,\infty[\to \mathbf{R}$ is a subset of $[0,1]$, then $L(\kappa^2(1-x'e)) \neq 0$ implies that $\kappa^2(1-x'e) \leq 1$. Thus $|V_{a,\kappa}(X_1,X_2)| > 0$

implies that $\{x \in S_d \,|\, \kappa^2(1 - x'X_1) \le 1,\, \kappa^2(1 - x'X_2) \le 1\} \ne \emptyset$. If x_0 is in this set, then $\|X_i - x_0\|^2 = 2(1 - X_i'x_0) \le 2\kappa^{-2}$ for $i = 1, 2$ and

$$X_1'X_2 = 1 - \frac{1}{2}\|X_1 - X_2\|^2 \ge 1 - \|X_1 - x_0\|^2 - \|X_2 - x_0\|^2 \ge 1 - 4\kappa^{-2}.$$

Thus, for $j \ge 1$ an integer,

$$\begin{aligned} \mathrm{E}\,|V_{a,\kappa}(X_1, X_2)|^j &\le \|V_{a,\kappa}(X_1, X_2)\|_\infty^j \mathrm{P}(|V_{a,\kappa}(X_1, X_2)| > 0) \\ &\le \|V_{a,\kappa}(X_1, X_2)\|_\infty^j \mathrm{P}(X_1'X_2 \ge 1 - 4\kappa^{-2}) \qquad (5.4) \\ &= O\left(\kappa^{j(2s+d)}\right) O\left(\kappa^{-d}\right) = O\left(\kappa^{2js+(j-1)d}\right), \end{aligned}$$

because, by Lemma 5.3 (i),

$$\begin{aligned} \|V_{a,\kappa}(X_1, X_2)\|_\infty &\le \kappa^{2d} \left\{ \int_{S_d} M_{\kappa,a}^2(x'X_1) d\mu(x) \int_{S_d} M_{\kappa,a}^2(x'X_2) d\mu(x) \right\}^{1/2} \\ &= \kappa^{2d} O(\kappa^{2s-d}) = O(\kappa^{2s+d}) \qquad (5.5) \end{aligned}$$

and by Lemma 2.4

$$\begin{aligned} \mathrm{P}\left(X_1'X_2 \ge 1 - 4\kappa^{-2}\right) &= \int_{S_d} d\mu(y) f(y) \int_{S_d} f(x) I_{\{(x,y)\,|\,x'y \ge 1-4\kappa^{-2}\}}(x, y) d\mu(x) \\ &\le \omega_d \|f\|_\infty \int_{S_d} I_{[0,4]}(\kappa^2(1 - x'y)) d\mu(x) = O\left(\kappa^{-d}\right). \end{aligned}$$

Define $\tilde\theta_{a,n}$ otherwise similarly as $\hat\theta_{a,n}$ but with the diagonal terms deleted,

$$\tilde\theta_{a,n} = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} V_{a,\kappa}(X_i, X_j) = \hat\theta_{a,n} - \frac{1}{n-1} V_{a,\kappa}(x, x).$$

Because $\mathrm{E}\left|\hat\theta_{a,n} - \mathrm{E}\,\hat\theta_{a,n}\right|^m = \mathrm{E}\left|\tilde\theta_{a,n} - \mathrm{E}\,\tilde\theta_{a,n}\right|^m$, it suffices to prove the condition (4.16) for the estimator $\tilde\theta_{a,n}$. The estimator $\tilde\theta_{a,n}$ is a U–statistic and $V_{a,\kappa}(x_1, x_2) = V_{a,\kappa}(x_2, x_1)$. Thus we can write as in Serfling (1980, page 180),

$$\tilde\theta_{a,n} = \frac{1}{n!} \sum_p W(X_{i_1}, \ldots, X_{i_n}),$$

where the summation is over all $n!$ permutations $(i_1, \ldots, i_n)$ of $(1, \ldots, n)$,

$$W(x_1, \ldots, x_n) = \frac{1}{[n/2]} \left(V_{a,\kappa}(x_1, x_2) + \cdots + V_{a,\kappa}(x_{2[n/2]-1}, x_{2[n/2]})\right)$$

and $[n/2]$ is the greatest integer $\leq n/2$. Denote $Z_{ni} = V_{a,\kappa}(X_{2i-1}, X_{2i})$, $i = 1, \ldots, [n/2]$. Then,

$$\begin{aligned}
\mathrm{E}\left|\tilde{\theta}_{a,n} - \mathrm{E}\,\tilde{\theta}_{a,n}\right|^m &\leq \mathrm{E}\left|W(X_1, \ldots, X_n) - \mathrm{E}\,W(X_1, \ldots, X_n)\right|^m \\
&= \mathrm{E}\left|\frac{1}{[n/2]}\sum_{i=1}^{[n/2]} Z_{ni} - \mathrm{E}Z_{n1}\right|^m .
\end{aligned}$$

From equation (5.4) it follows that $\mathrm{Var}Z_{n1} = \mathrm{Var}V_{a,\kappa}(X_1, X_2) = O\left(\kappa^{4s+d}\right)$ and from equation (5.5) it follows that $\|Z_{ni}\|_\infty = O(\kappa^{2s+d})$. Lemma 4.3 can be applied. Because $O((\kappa_n^{4s+d} n^{-1})^{m/2}) = o(n^{-ms/(2s+d)})$, the condition (4.16) is satisfied. □

The statistic $\hat{\theta}_{a,n}$ is an estimator for $\theta_2(f) = \int_{S_d}(D^2 f)^2 d\mu$, when $a = (a_0, a_1) = (\gamma_{0,1}/\gamma_{0,0}, 1/\gamma_{2,0}) = ((d-6)/4, 1)$, However, it is possible to construct an estimator for this quantity which has even $\sqrt{n}$–convergence when f is sufficiently smooth. Let us discuss, a bit more generally, the estimation of

$$\theta_{q,r}(f) = \int_{S_d} (\Delta^{q/2} f)(\Delta^{r/2} f) d\mu,$$

where $q, r \geq 0$ are even. The ideas of Bickel and Ritov (1988) are applicable when constructing an estimator for $\theta_{q,r}(f)$.

Let us split the sample in two parts of size $n_1 = n_1(n)$ and $n_2 = n_2(n)$ respectively. Assume that $\lim_{n\to\infty} n_1/n_2 = 1$. Let $\hat{F}_1$ and $\hat{F}_2$ be the empirical distribution function, based on the first and the second part of the sample, respectively. Let $\hat{f}_1$ and $\hat{f}_2$ be kernel estimators, based on the first and the second part of the sample, respectively. Note that $\hat{f}_i(x) = c(\kappa)\int_{S_d} L(\kappa^2(1 - x'y))d\hat{F}_i(y)$ for $i = 1, 2$. Define

$$\hat{\theta}_{q,r} = \frac{n_1}{n}\hat{\theta}_{q,r,1} + \frac{n_2}{n}\hat{\theta}_{q,r,2},$$

where

$$\begin{aligned}
\hat{\theta}_{q,r,1} &= 2c(\kappa)\int L^{(q+r,\kappa)}(x'y)d\hat{F}_1(x)d\hat{F}_2(y) \qquad (5.6)\\
&\quad -2[n_2(n_2-1)]^{-1}c(\kappa)^2 \sum_{n_1+1\leq i<j\leq n} \int_{S_d} L^{(q,\kappa)}(x'X_i)L^{(r,\kappa)}(x'X_j)d\mu(x),
\end{aligned}$$

$L^{(r,\kappa)}(x'e) = \Delta^{r/2}\left[L(\kappa^2(1 - x'e))\right]$, and $\hat{\theta}_{q,r,2}$ is defined similarly. This definition can be motivated by noting that a natural estimator for $\theta_{q,r}(f)$ would be a one step estimator

$$\tilde{\theta}_{q,r,1} = \theta_{q,r}(\hat{f}_2) + \frac{2}{n_1}\sum_{i=1}^{n_1}\left[\Delta^{(q+r)/2}\hat{f}_2(X_i) - \theta_{q,r}(\hat{f}_2)\right]$$

(see equation (6.7) in page 93; Bickel 1982, Schick 1986). Now

$$\frac{2}{n_1}\sum_{i=1}^{n_1}\Delta^{(q+r)/2}\hat{f}_2(X_i) = \frac{2c(\kappa)}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n} L^{(q+r,\kappa)}(X_i'X_j),$$

which is the first term on the right hand side of equation (5.6). Also,

$$\begin{aligned}\theta_{q,r}(\hat{f}_2) &= \int_{S_d}(\Delta^{q/2}\hat{f}_2)(\Delta^{r/2}\hat{f}_2)d\mu \\ &= \frac{c(\kappa)^2}{n_2^2}\sum_{i,j=n_1+1}^{n}\int_{S_d} L^{(q,\kappa)}(x'X_i)L^{(r,\kappa)}(x'X_j)d\mu(x),\end{aligned}$$

which leads to the second term on the right hand side of equation (5.6) when the diagonal terms are deleted and n_2^{-2} is changed to $[n_2(n_2-1)]^{-1}$. Note that the diagonal terms do not depend on the data.

Theorem 5.13 *Let $q, r \geq 0$ be even. Let $\sigma > q + r + d/4$. Let $f : S_d \to \mathbf{R}$ be a bounded density for which $\Delta^{i/2}f$ are bounded with $\Delta^{i/2}f \in \mathbf{F}_3(\sigma_i, p)$ for $i = q, r$ and $p = 2, \infty$ where $\sigma_i \geq \sigma - i$. (The definition of $\mathbf{F}_3$ was given before Lemma 3.4.) Let also $\Delta^{(q+r)/2}f \in \mathbf{F}_3(\sigma_{q,r}, 2)$, where $\sigma_{q,r} \geq \sigma - q - r$. Write $\max\{\sigma_q, \sigma_r, \sigma_{q,r}\} = s + a$, where $s \geq 0$ is even and $0 \leq a < 2$. Let $L : [0, \infty[\to \mathbf{R}$ be a class $s + 2$ kernel with support on $[0, 1]$, for which $\alpha_0(|L|) < \infty$, $\alpha_0(|L^{(i)}|^p) < \infty$ for $i = 1, \ldots, l$, where $p = 1$, $l = \max\{q/2, r/2\}$ or $p = 2$, $l = (q + r)/2$, and $\alpha_{2p(i-l/2)}(|L^{(i)}|^p) < \infty$ for $i = l/2 + 1, \ldots, l$, where $p = 1, 2$, $l = q, r$ or $p = 2$, $l = (q + r)/2$. Let $\kappa = n^{2/(4\sigma+d)}$. Then*

$$\lim_{n\to\infty} n\mathrm{E}\left(\hat{\theta}_{q,r} - \theta_{q,r}(f)\right)^2 = C_{q,r}(f)$$

and

$$\sqrt{n}\left(\hat{\theta}_{q,r} - \theta_{q,r}(f)\right) \longrightarrow N(0, C_{q,r}(f)),$$

where

$$C_{q,r}(f) = 4\mathrm{Var}\left(\Delta^{(q+r)/2}f(X_1)\right) = 4\left\{\int_{S_d}\left(\Delta^{(q+r)/2}f\right)^2 f d\mu - \theta_{q,r}(f)^2\right\}.$$

Proof. Denote

$$\hat{Q}_{q,r,1} = \hat{\theta}_{q,r,1} - \theta_{q,r}(f) - \frac{2}{n_1}\sum_{i=1}^{n_1}\left\{\Delta^{(q+r)/2} f(X_i) - \theta_{q,r}(f)\right\}.$$

It suffices to prove that $\lim_{n\to\infty} n\mathrm{E}\left(\hat{Q}_{q,r,1}\right)^2 = 0$. Write

$$\mathrm{E}\left(\hat{Q}_{q,r,1}\right)^2 = \left(\mathrm{E}\,\hat{Q}_{q,r,1}\right)^2 + \mathrm{E}\,\mathrm{Var}\left(\hat{Q}_{q,r,1}\,\middle|\,\hat{F}_2\right) + \mathrm{Var}\,\mathrm{E}\left(\hat{Q}_{q,r,1}\,\middle|\,\hat{F}_2\right).$$

Let us prove first that

$$\left(\mathrm{E}\,\hat{Q}_{q,r,1}\right)^2 = \left(\mathrm{E}\,\hat{\theta}_{q,r,1} - \theta_{q,r}(f)\right)^2 = o\left(n^{-1}\right).$$

In fact,

$$\begin{aligned}
\mathrm{E}&\left[c(\kappa)\int L^{(q+r,\kappa)}(x'y)d\hat{F}_1(x)d\hat{F}_2(y)\,\middle|\,\hat{F}_2\right] = \mathrm{E}\left[\frac{1}{n_1}\sum_{i=1}^{n_1}\Delta^{(q+r)/2}\hat{f}_2(X_i)\,\middle|\,\hat{F}_2\right]\\
&= \int_{S_d}\left(\Delta^{(q+r)/2}\hat{f}_2\right) f d\mu = \int_{S_d}\left(\Delta^{q/2}\hat{f}_2\right)\left(\Delta^{r/2} f\right) d\mu\\
&= \frac{c(\kappa)}{n_2}\sum_{i=n_1+1}^{n}\int_{S_d} L^{(q,\kappa)}(x'X_i)\Delta^{r/2} f(x)d\mu(x)\\
&= [n_2(n_2-1)]^{-1}\sum_{i=n_1+1}^{n}\sum_{j=n_1+1}^{i-1}\\
&\quad\left\{c(\kappa)\int_{S_d} L^{(q,\kappa)}(x'X_i)\Delta^{r/2} f(x)d\mu(x) + c(\kappa)\int_{S_d} L^{(q,\kappa)}(x'X_j)\Delta^{r/2} f(x)d\mu(x)\right\}.
\end{aligned}$$

Thus,

$$\mathrm{E}\left(\hat{\theta}_{q,r,1}\,\middle|\,\hat{F}_2\right) - \theta_{q,r}(f) = \frac{2}{n_2(n_2-1)}\sum_{i=n_1+1}^{n}\sum_{j=n_1+1}^{i-1} U_{q,r,\kappa}(X_i, X_j), \tag{5.7}$$

where

$$U_{q,r,\kappa}(x_i, x_j) = -\int_{S_d}\left\{c(\kappa)L^{(q,\kappa)}(x'x_i) - \Delta^{q/2} f(x)\right\}\left\{c(\kappa)L^{(r,\kappa)}(x'x_j) - \Delta^{r/2} f(x)\right\} d\mu(x).$$

Now

$$c(\kappa)\int_{S_d} L^{(q,\kappa)}(x'y)f(y)d\mu(y) = c(\kappa)\int_{S_d}\Delta^{q/2}\left[L(\kappa^2(1-x'y))\right] f(y)d\mu(y)$$

$$= c(\kappa) \int_{S_d} L(\kappa^2(1 - x'y))\Delta^{q/2} f(y) d\mu(y) = c(\kappa) \left(\Delta^{q/2} f\right) * L_\kappa(x).$$

Thus, from equation (5.7) it follows that

$$\begin{aligned}
&\left(\mathrm{E}\hat{\theta}_{q,r,1} - \theta_{q,r}(f)\right)^2 \\
&= \left\{\int_{S_d} \left(c(\kappa)\left(\Delta^{q/2} f\right) * L_\kappa - \Delta^{q/2} f\right)\left(c(\kappa)\left(\Delta^{r/2} f\right) * L_\kappa - \Delta^{r/2} f\right) d\mu\right\}^2 \\
&\leq \left\|c(\kappa)\left(\Delta^{q/2} f\right) * L_\kappa - \Delta^{q/2} f\right\|_2^2 \left\|c(\kappa)\left(\Delta^{r/2} f\right) * L_\kappa - \Delta^{r/2} f\right\|_2^2 \\
&= O\left(\kappa^{-2(\sigma_q + \sigma_r)}\right) = O\left(\kappa^{2(q+r-2\sigma)}\right) = o\left(n^{-1}\right)
\end{aligned}$$

by Lemma 3.4.

Secondly, let us prove that

$$\operatorname{Var} \mathrm{E}\left(\hat{Q}_{q,r,1} \,\middle|\, \hat{F}_2\right) = \operatorname{Var} \mathrm{E}\left(\hat{\theta}_{q,r,1} \,\middle|\, \hat{F}_2\right) = o(n^{-1}).$$

In equation (5.7) $\mathrm{E}\left(\hat{\theta}_{q,r,1} \,\middle|\, \hat{F}_2\right) - \theta_{q,r}(f)$ was written as a U–statistic. Because the Laplace operator is symmetric, $U_{q,r,\kappa}(x_1, x_2) = U_{q,r,\kappa}(x_2, x_1)$. From Serfling (1980, page 183) it is seen that

$$\begin{aligned}
\operatorname{Var} \mathrm{E}\left(\hat{\theta}_{q,r,1} \,\middle|\, \hat{F}_2\right) &= \operatorname{Var}\left(\mathrm{E}\left(\hat{\theta}_{q,r,1} \,\middle|\, \hat{F}_2\right) - \theta_{q,r}(f)\right) \\
&= \frac{4(n_2 - 2)}{n_2(n_2 - 1)}\xi_1 + \frac{2}{n_2(n_2 - 1)}\xi_2,
\end{aligned} \tag{5.8}$$

where $\xi_1 = \operatorname{Var}(h_1(X_1))$, $h_1(x_1) = \mathrm{E}(U_{q,r,\kappa}(x_1, X_2))$ and $\xi_2 = \operatorname{Var}(U_{q,r,\kappa}(X_1, X_2))$. Now

$$\begin{aligned}
h_1(x_1) &= -\mathrm{E} \int_{S_d} \left\{c(\kappa) L^{(q,\kappa)}(x'x_1) - \Delta^{q/2} f(x)\right\}\left\{c(\kappa) L^{(r,\kappa)}(x'X_2) - \Delta^{r/2} f(x)\right\} d\mu(x) \\
&= -\int_{S_d} \left\{c(\kappa) L^{(q,\kappa)}(x'x_1) - \Delta^{q/2} f(x)\right\}\left\{c(\kappa)\left(\Delta^{r/2} f\right) * L_\kappa(x) - \Delta^{r/2} f(x)\right\} d\mu(x) \\
&= -\int_{S_d} \left\{c(\kappa) L^{(q,\kappa)}(x'x_1) - \Delta^{q/2} f(x)\right\}\delta_{r,\kappa}(x) d\mu(x),
\end{aligned}$$

where $\delta_{r,\kappa} \stackrel{def}{=} c(\kappa)\left(\Delta^{r/2} f\right) * L_\kappa - \Delta^{r/2} f$. Thus, by Lemmas 3.4 and 5.9 (i),

$$\begin{aligned}
\xi_1 &= \operatorname{Var}(h_1(X_1)) \leq \mathrm{E}(h_1^2(X_1)) \\
&= \mathrm{E}\left[\int_{S_d} \left\{c(\kappa) L^{(q,\kappa)}(x'X_1) - \Delta^{q/2} f(x)\right\}\delta_{r,\kappa}(x) d\mu(x)\right]^2
\end{aligned} \tag{5.9}$$

$$
\begin{aligned}
&\leq \|\delta_{r,\kappa}\|_\infty^2 \mathrm{E}\left[\int_{S_d} \left|c(\kappa)L^{(q,\kappa)}(x'X_1)\right| d\mu(x) + \int_{S_d} |\Delta^{q/2} f| d\mu\right]^2 \\
&= O\left(\kappa^{-2\sigma_r}\right) O\left(\kappa^{2q}\right) = O\left(\kappa^{2(q+r-\sigma)}\right).
\end{aligned}
$$

Also,

$$
\begin{aligned}
\xi_2 &= \mathrm{Var}\left(U_{q,r,\kappa}(X_1, X_2)\right) \qquad (5.10) \\
&= \mathrm{Var}\left\{c(\kappa)^2 \int_{S_d} L^{(q,\kappa)}(x'X_1) L^{(r,\kappa)}(x'X_2) d\mu(x)\right. \\
&\quad \left. - c(\kappa)\int_{S_d} L^{(q,\kappa)}(x'X_1)\Delta^{r/2} f(x) d\mu(x) - c(\kappa)\int_{S_d} L^{(r,\kappa)}(x'X_2)\Delta^{q/2} f(x) d\mu(x)\right\} \\
&\leq 3\mathrm{E}\left\{c(\kappa)^2 \int_{S_d} L^{(q,\kappa)}(x'X_1) L^{(r,\kappa)}(x'X_2) d\mu(x)\right\}^2 \\
&\quad + 6\mathrm{E}\left\{c(\kappa)\int_{S_d} L^{(q,\kappa)}(x'X_1)\Delta^{r/2} f(x) d\mu(x)\right\}^2 = O\left(\kappa^{2(q+r)+d}\right)
\end{aligned}
$$

because we get, in a similar manner as equation (5.4) was proved, but this time applying Lemma 5.9 (i), that

$$
\mathrm{E}\left\{c(\kappa)^2 \int_{S_d} L^{(q,\kappa)}(x'X_1) L^{(r,\kappa)}(x'X_2) d\mu(x)\right\}^2 = O\left(\kappa^{2(q+r)+d}\right)
$$

and by Lemma 5.9 (i),

$$
\begin{aligned}
&\mathrm{E}\left\{c(\kappa)\int_{S_d} L^{(q,\kappa)}(x'X_1)\Delta^{r/2} f(x) d\mu(x)\right\}^2 \\
&\leq \|\Delta^{r/2} f\|_\infty^2 \mathrm{E}\left\{c(\kappa)\int_{S_d} L^{(q,\kappa)}(x'X_1) d\mu(x)\right\}^2 = O\left(\kappa^{2q}\right).
\end{aligned}
$$

From equations (5.8), (5.9) and (5.10) it follows that

$$
\mathrm{Var}\,\mathrm{E}\left(\hat{\theta}_{q,r,1}\middle|\hat{F}_2\right) = O\left(n^{-1}\kappa^{2(q+r-\sigma)} + n^{-2}\kappa^{2(q+r)+d}\right) = o\left(n^{-1}\right). \qquad (5.11)
$$

It remains to prove $\mathrm{E}\,\mathrm{Var}\left(\hat{Q}_{q,r,1}\middle|\hat{F}_2\right) = o\left(n^{-1}\right)$. Now

$$
\begin{aligned}
&\mathrm{Var}\left(\hat{Q}_{q,r,1}\middle|\hat{F}_2\right) \\
&= 4\mathrm{Var}\left(c(\kappa)\int L^{(q+r,\kappa)}(x'y) d\hat{F}_1(x) d\hat{F}_2(y) - \int \Delta^{(q+r)/2} f(x) d\hat{F}_1(x)\middle|\hat{F}_2\right)
\end{aligned}
$$

$$
\begin{aligned}
&= 4\mathrm{Var}\left(n_1^{-1} \sum_{i=1}^{n_1} \left[\Delta^{(q+r)/2} \hat{f}_2(X_i) - \Delta^{(q+r)/2} f(X_i) \right] \middle| \hat{F}_2 \right) \\
&= 4n_1^{-1}\mathrm{Var}\left(\Delta^{(q+r)/2} \hat{f}_2(X_1) - \Delta^{(q+r)/2} f(X_1) \middle| \hat{F}_2 \right) \\
&\leq 4n_1^{-1}\mathrm{E}\left(\left[\Delta^{(q+r)/2} \hat{f}_2(X_1) - \Delta^{(q+r)/2} f(X_1) \right]^2 \middle| \hat{F}_2 \right) \\
&= 4n_1^{-1} \int_{S_d} \left[\Delta^{(q+r)/2} \hat{f}_2(y) - \Delta^{(q+r)/2} f(y) \right]^2 f(y) d\mu(y).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\mathrm{E}\,\mathrm{Var}\left(\hat{Q}_{q,r,1} \middle| \hat{F}_2 \right) &\leq 4n_1^{-1} \int_{S_d} \mathrm{E}\left[\Delta^{(q+r)/2} \hat{f}_2(y) - \Delta^{(q+r)/2} f(y) \right]^2 f(y) d\mu(y) \\
&= 4n_1^{-1} \int_{S_d} \left[\mathrm{E}\,\Delta^{(q+r)/2} \hat{f}_2(y) - \Delta^{(q+r)/2} f(y) \right]^2 f(y) d\mu(y) \\
&\quad + 4n_1^{-1} \int_{S_d} \mathrm{Var}\left[\Delta^{(q+r)/2} \hat{f}_2(y) \right] f(y) d\mu(y) \\
&= O\left(n^{-1} \kappa^{2(q+r-\sigma)} \right) + O\left(n^{-2} \kappa^{2(q+r)+d} \right) = o\left(n^{-1} \right),
\end{aligned}
$$

because, by Lemma 3.4,

$$
\begin{aligned}
&\int_{S_d} \left[\mathrm{E}\,\Delta^{(q+r)/2} \hat{f}_2(y) - \Delta^{(q+r)/2} f(y) \right]^2 f(y) d\mu(y) \\
&= \int_{S_d} \left[c(\kappa) \left(\Delta^{(q+r)/2} f \right) * L_\kappa(y) - \Delta^{(q+r)/2} f(y) \right]^2 f(y) d\mu(y) \\
&= O\left(\kappa^{2\sigma_{q,r}} \right) = O\left(\kappa^{2(q+r-\sigma)} \right)
\end{aligned}
$$

and, by Lemma 5.9 (i),

$$
\begin{aligned}
&\int_{S_d} \mathrm{Var}\left[\Delta^{(q+r)/2} \hat{f}_2(y) \right] f(y) d\mu(y) = n_2^{-1} \int_{S_d} \mathrm{Var}\left[c(\kappa) L^{(q+r,\kappa)}(X_1' y) \right] f(y) d\mu(y) \\
&\leq n_2^{-1} c(\kappa)^2 \int_{S_d} d\mu(y) f(y) \int_{S_d} \left(L^{(q+r,\kappa)}(x'y) \right)^2 f(x) d\mu(x) = n_2^{-1} O\left(\kappa^{2(q+r)+d} \right).
\end{aligned}
$$

□

A lower bound related to this theorem is given in Section 6.4.

Chapter 6

Lower Bounds

In Sections 3.3 and 3.4 it was shown that the risk of the kernel estimator for smoothness index s has the rate of convergence $n^{s/(2s+d)}$. In Section 5.1 an estimator for the rth order iterated Laplacian was given whose risk has the rate of convergence $n^{(s-r)/(2s+d)}$. In this chapter it is shown that these rates are the best possible in the minimax sense. Also, it is shown that the estimator for the inner product of Laplacians, defined in Section 5.3, is optimal in the sense of local asymptotic minimaxity.

The method of proof is based on the theory of convergence of experiments. Define, for example,

$$f_{h,n} = f_0 + n^{-1/2}\sqrt{f_0}\,h,$$

where $f_0 : \mathbf{R} \to \mathbf{R}$ is a density and $h : \mathbf{R} \to \mathbf{R}$ a Borel function. Now the product experiment $(\mathbf{R}^n, \mathcal{B}(\mathbf{R})^n, f_{h,n}^{(n)} : h \in \mathbf{H}_c)$ converges to a Gaussian shift experiment for a certain $\mathbf{H}_c \subset \{h : \mathbf{R} \to \mathbf{R} \mid \int h^2 < \infty\}$. This fact was used by Millar (1979) to study the estimation of distribution functions. More generally, L_2–differentiable parameterizations of a neighborhood of f_0 lead to convergence to a Gaussian shift experiment (Strasser 1985, Chapter 12). Localization by the factor $n^{-1/2}$ can be used when there exist $\sqrt{n}$–consistent estimators. However, also the product of the experiments consisting of the densities of the form

$$f_{h,n}(x) = f_0(x) + n^{-1/2}\sqrt{f_0(0)}\kappa_n^{1/2}h(\kappa_n x),\ x \in \mathbf{R},$$

where $\kappa_n \to \infty$, $n^{-1}\kappa_n \to 0$, converges to a Gaussian shift experiment. This kind of localization was used by Low (1992) to study problems where there exists only estimators with slower than $\sqrt{n}$–convergence.

In Section 6.1 sufficient conditions are formulated for a product of experiments consisting of densities of the form $f_{h,n} = f_0 + T_n h$, with T_n a suitable operator, to converge to a Gaussian shift experiment. The result is formulated for the general sample space to cover both the Euclidean and the spherical cases. In Section 6.2 a general lower bound is given for the pointwise risk. Then it is shown how the well–known Euclidean lower bounds for the estimation of a density at a point and for the estimation of derivatives of a density at a point fit in the present framework. Then the corresponding spherical lower bound is given. In Section 6.3 a general lower bound is given for the L_p estimation of a density. Then a Euclidean example and two spherical examples are given. In Section 6.4 it is shown that the estimator for the inner product of Laplacians, defined in Section 5.3, is optimal in the local asymptotic minimax sense.

6.1 Convergence of Experiments

Let us start with the definitions of some concepts needed in this chapter. These definitions are given in Le Cam (1986) and Strasser (1985). Let $E = (\Omega, \mathcal{A}, P_\theta : \theta \in \Theta)$ be an experiment, where Ω is a set with sigma–field $\mathcal{A}$ and P_θ are probability measures indexed by the set Θ. The *L–space* of the experiment E is

$$L(E) = \{\lambda \in ca(\Omega, \mathcal{A}) \,|\, \lambda \perp \sigma \text{ for all } \sigma \in ca(\Omega, \mathcal{A}) \text{ such that } \sigma \perp P_\theta \text{ for all } \theta \in \Theta\}\,.$$

where $ca(\Omega, \mathcal{A})$ is the set of finite signed measures on $(\Omega, \mathcal{A})$. For $\lambda \in L(E)$, let $\|\lambda\| = \sup_{A \in \mathcal{A}} |\lambda(A)|$. Let $\mathbf{D}$ be a topological space and $\mathcal{C}_b(\mathbf{D})$ the set of bounded continuous functions $\mathbf{D} \to \mathbf{R}$. For $g \in \mathcal{C}_b(\mathbf{D})$, let $\|g\|_\infty = \sup_{a \in \mathbf{D}} |g(a)|$. A *generalized decision function* for E and $\mathbf{D}$ is a bilinear function $\beta : \mathcal{C}_b(\mathbf{D}) \times L(E) \to \mathbf{R}$ satisfying the following conditions:

(i) $|\beta(g, \lambda)| \leq \|g\|_\infty \|\lambda\|$, if $g \in \mathcal{C}_b(\mathbf{D})$, $\lambda \in L(E)$.

(ii) $\beta(g, \mu) \geq 0$, if $g \geq 0$, $\mu \geq 0$.

(iii) $\beta(1, \lambda) = \lambda(\Omega)$, if $\lambda \in L(E)$.

The set of generalized decision functions is denoted by $\mathcal{B}(E, \mathbf{D})$. The concept of generalized decision function generalizes the concept of risk. If $\beta \in \mathcal{B}(E, \mathbf{D})$, $f : \mathbf{D} \to$

$]-\infty,\infty]$ is lower semicontinuous (for all $y \in \mathbf{R}$, $\{x \mid f(x) > y\}$ is open), f is bounded from below and $\lambda \in L(E)$, $\lambda \geq 0$, define $\beta(f,\lambda) = \sup\{\beta(g,\lambda) \mid g \in \mathcal{C}_b(\mathbf{D}),\, g \leq f\}$. A *decision function* for E and $\mathbf{D}$ is a Markov kernel $\rho : \Omega \times \mathcal{B}_0(\mathbf{D}) \to [0,1]$, where $\mathcal{B}_0(\mathbf{D})$ is the Baire sigma–field of $\mathbf{D}$. Thus it holds that

(i) $\rho(\cdot, D)$ is $\mathcal{A}$–measurable for every $D \in \mathcal{B}_0(\mathbf{D})$.

(ii) $\rho(\omega, \cdot)$ is probability measure on $\mathcal{B}_0(\mathbf{D})$ for every $\omega \in \Omega$.

The set of decision functions is denoted by $\mathcal{R}(E,\mathbf{D})$. The set of measurable functions $(\Omega,\mathcal{A}) \to (\mathbf{D},\mathcal{B}_0(\mathbf{D}))$ is denoted by $\mathcal{R}_0(E,\mathbf{D})$. If $\hat{a} \in \mathcal{R}_0(E,\mathbf{D})$, define $\rho_{\hat{a}} \in \mathcal{R}(E,\mathbf{D})$ by $\rho_{\hat{a}}(\omega, D) = I_D(\hat{a}(\omega))$. If $\rho \in \mathcal{R}(E,\mathbf{D})$, define $\beta_\rho \in \mathcal{B}(E,\mathbf{D})$ by $\beta_\rho(g,\lambda) = \int_\Omega d\lambda(\omega) \int_{\mathbf{D}} g(a)\rho(\omega, da)$. With these identifications we can state that $\mathcal{R}_0(E,\mathbf{D}) \subset \mathcal{R}(E,\mathbf{D}) \subset \mathcal{B}(E,\mathbf{D})$. By Strasser (1985, Remark 55.6 (3)), if E is dominated and $\mathbf{D}$ is locally compact with countable base then $\mathcal{B}(E,\mathbf{D}) = \{\beta_\rho \mid \rho \in \mathcal{R}(E,\mathbf{D})\}$. This holds also if E is dominated and $\mathbf{D}$ is a Polish space (complete separable metric space), see Nussbaum (1995, Proposition 10.2).

Let $E_n = (\Omega_n, \mathcal{A}_n, P_{\theta,n} : \theta \in \Theta)$ be experiments and denote by $E_n \xrightarrow{w} E$ the weak convergence of experiments, which means that for every finite $J \subset \Theta$ and every $\eta \in J$,

$$\mathcal{L}\left(\left(\frac{dP_{\theta,n}}{dP_{\eta,n}}\right)_{\theta\in J} \,\middle|\, P_{\eta,n}\right) \longrightarrow \mathcal{L}\left(\left(\frac{dP_{\theta}}{dP_{\eta}}\right)_{\theta\in J} \,\middle|\, P_{\eta}\right).$$

A consequence of weak convergence of experiments is the lower bound for the asymptotic minimax risk as given in the next lemma. This lemma is essentially based on the results in Le Cam (1972).

Lemma 6.1 *Let* $E_n \xrightarrow{w} E$. *Let* $W_\theta : \mathbf{D} \to]-\infty,\infty]$, $\theta \in \Theta$, *be loss functions which are lower semicontinuous and bounded from below. Then*

$$\liminf_{n\to\infty} \inf_{\beta_n \in \mathcal{B}(E_n,\mathbf{D})} \sup_{\theta\in\Theta} \beta_n\left(W_\theta, P_{\theta,n}\right) \geq \inf_{\beta\in\mathcal{B}(E,\mathbf{D})} \sup_{\theta\in\Theta} \beta\left(W_\theta, P_\theta\right).$$

Proof. Let $\Pi_0(\Theta)$ be the set of probability measures with support on Θ of finite cardinality and suppose $\pi \in \Pi_0(\Theta)$. Let $V_\theta \in \mathcal{C}_b(\mathbf{D})$, $V_\theta \leq W_\theta$ for $\theta \in \Theta$. Then

$$\begin{aligned}\liminf_{n\to\infty} \inf_{\beta_n \in \mathcal{B}(E_n,\mathbf{D})} \sup_{\theta\in\Theta} \beta_n\left(W_\theta, P_{\theta,n}\right) &\geq \liminf_{n\to\infty} \inf_{\beta_n \in \mathcal{B}(E_n,\mathbf{D})} \int_\Theta \beta_n\left(V_\theta, P_{\theta,n}\right) d\pi(\theta) \\ &= \inf_{\beta\in\mathcal{B}(E,\mathbf{D})} \int_\Theta \beta\left(V_\theta, P_\theta\right) d\pi(\theta)\end{aligned}$$

by Strasser (1985, Theorem 60.3, Corollary 53.8, Theorem 49.7). Thus,

$$\begin{aligned}
&\liminf_{n\to\infty} \inf_{\beta_n\in\mathcal{B}(E_n,\mathbf{D})} \sup_{\theta\in\Theta} \beta_n\left(W_\theta, P_{\theta,n}\right) \\
&\geq \sup_{\pi\in\Pi_0(\Theta)} \sup_{V_\theta\in\mathcal{C}_b(\mathbf{D}),\, V_\theta\leq W_\theta} \inf_{\beta\in\mathcal{B}(E,\mathbf{D})} \int_\Theta \beta\left(V_\theta, P_\theta\right) d\pi(\theta) \\
&= \sup_{\pi\in\Pi_0(\Theta)} \inf_{\beta\in\mathcal{B}(E,\mathbf{D})} \int_\Theta \beta\left(W_\theta, P_\theta\right) d\pi(\theta) = \inf_{\beta\in\mathcal{B}(E,\mathbf{D})} \sup_{\theta\in\Theta} \beta\left(W_\theta, P_\theta\right)
\end{aligned}$$

by Strasser (1985, Corollary 47.4, Theorem 46.3). □

Let us define two particular experiments, which will be studied. Firstly, let $\mathcal{Y}$ be a set carrying a sigma–field $\mathcal{G}$ with a sigma–finite measure ν. Let $\mathbf{H} = L_2(\mathcal{Y}, \mathcal{G}, \nu)$ and denote $< h_1, h_2 >_{\mathbf{H}} = \int_{\mathcal{Y}} h_1 h_2 d\nu$, and $\|h\|^2_{\mathbf{H}} =< h, h >_{\mathbf{H}}$, for $h_1, h_2, h \in \mathbf{H}$. Assume given a standard Gaussian process $X(h)$, $h \in \mathbf{H}$, on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The process $X(h)$ is such that its finite dimensional distributions are Gaussian, $\int X(h) d\mathbb{P} = 0$ and $\int X(h_1)X(h_2) d\mathbb{P} =< h_1, h_2 >_{\mathbf{H}}$. Let

$$E(\mathbf{H}) = (\Omega, \mathcal{A}, G_h : h \in \mathbf{H})$$

be an experiment, where the measures G_h are mutually equivalent, $G_0 = \mathbb{P}$, and

$$\frac{dG_h}{dG_0} = \exp\left\{X(h) - \frac{1}{2}\|h\|^2_{\mathbf{H}}\right\}.$$

For any $\mathbf{H}_c \subset \mathbf{H}$, let $E(\mathbf{H}_c) = (\Omega, \mathcal{A}, G_h : h \in \mathbf{H}_c)$. Experiment $E(\mathbf{H})$ is a Gaussian shift experiment not only in the broader sense of Le Cam (1986, Definition 2, Chapter 9) but also in the narrower one of Strasser (1985, Definition 69.2), on account of the characterization in Strasser (1985, Theorem 69.4). The experiment $E(\mathbf{H})$ is what Le Cam (1986, Chapter 9) calls a standard Gaussian shift experiment. Strasser (1985, Examples 70.7) gives examples of standard Gaussian processes with $\mathcal{Y} = \mathbf{R}$, expressed in terms of the Wiener integral. A more general construction of a standard Gaussian process is given in Le Cam and Yang (1991, page 117).

Secondly, let $\mathcal{X}$ be a set carrying a sigma–field $\mathcal{F}$ with a sigma–finite measure μ. Let $\mathbb{F}$ be the set of measurable functions $\mathcal{X} \to \mathbf{R}$. Let $f_0 \in \mathbb{F}$ be a density with respect to μ. Let $\mathbf{H}_c \subset \mathbf{H}$, where $0 \in \mathbf{H}_c$, and let $T_n : \bar{\mathbf{H}}_c \to \mathbb{F}$ be mappings that possibly depend on f_0, where $\bar{\mathbf{H}}_c$ is the linear closure of $\mathbf{H}_c$. Let us denote

$$f_{h,n} = f_0 + T_n h.$$

It will be assumed that $\{f_{h,n} \mid h \in \mathbf{H}_c\}$ is ultimately a subset of densities. For $h \in \mathbf{H}_c$ it follows that $\int_{\mathcal{X}} (T_n h) d\mu = 0$ ultimately. Let

$$E_n(\mathbf{H}_c) = \left(\mathcal{X}^n, \mathcal{F}^n, P^n_{h,n} : h \in \mathbf{H}_c\right)$$

be the direct n-fold product of the experiment $(\mathcal{X}, \mathcal{F}, P_{h,n} : h \in \mathbf{H}_c)$, where $P_{h,n}$ corresponds to $f_{h,n}$. In the next lemma, sufficient conditions for the weak convergence of the experiment $E_n(\mathbf{H}_c)$ to the experiment $E(\mathbf{H}_c)$ are given. This lemma generalizes Theorem 2 in Low (1992).

Lemma 6.2 *Let P_0 be the probability measure corresponding to f_0. Assume that,*

(i) *$E_n(\mathbf{H}_c)$ is contiguous, that is, for every $h_1, h_2 \in \mathbf{H}_c$, if $\{A_n\}$, $A_n \in \mathcal{F}^n$, is such a sequence that $P^n_{h_1,n}(A_n) \to 0$, then $P^n_{h_2,n}(A_n) \to 0$.*

(ii) $\int_{\mathcal{X}} (T_n h) d\mu = 0$, $h \in \mathbf{H}_c$.

(iii) $n \int_{\mathcal{X}} (T_n h_1)(T_n h_2) f_0^{-2} dP_0 =< h_1, h_2 >_{\mathbf{H}} + o(1)$, $h_1, h_2 \in \mathbf{H}_c$.

(iv) *The mapping T_n is linear and for all h in the linear closure of $\mathbf{H}_c$, and for all $\epsilon > 0$,*

$$n \int_{(|(T_n h) f_0^{-1}| > \epsilon)} \left((T_n h) f_0^{-1}\right)^2 dP_0 = o(1).$$

Then $E_n(\mathbf{H}_c) \xrightarrow{w} E(\mathbf{H}_c)$, when $n \to \infty$.

Proof. From condition (i) it follows by Strasser (1985, Theorem 61.6) that it suffices to prove that for every finite $J \subset \mathbf{H}_c$,

$$\mathcal{L}\left(\left(\frac{dP^n_{h,n}}{dP^n_0}\right)_{h \in J} \,\middle|\, P^n_0\right) \longrightarrow \mathcal{L}\left(\left(\frac{dG_h}{dG_0}\right)_{h \in J} \,\middle|\, G_0\right).$$

Thus it suffices to prove that for every finite $J \subset \mathbf{H}_c$,

$$\mathcal{L}\left(\left(\log \frac{f^{(n)}_{h,n}}{f^{(n)}_0}\right)_{h \in J} \,\middle|\, P^n_0\right) \longrightarrow N(m, \Sigma),$$

where $N(m, \Sigma)$ is the normal law with expectation $m = (-\frac{1}{2}\|h\|^2_{\mathbf{H}})_{h \in J}$ and covariance matrix $\Sigma = (< h_1, h_2 >_{\mathbf{H}})_{h_1, h_2 \in J}$. Let $X^{(n)} = (X_1, \ldots, X_n) \sim P^n_0$ and define $X_{hni} = (T_n h)(X_i)/f_0(X_i)$, $i = 1, \ldots, n$. Now for $h \in \mathbf{H}_c$,

$$\log \frac{f^{(n)}_{h,n}(X^{(n)})}{f^{(n)}_0(X^{(n)})} = \sum_{i=1}^{n} \log(1 + X_{hni}) = Q_{hn} - \frac{1}{2} R_{hn} + S_{hn},$$

where

$$Q_{hn} = \sum_{i=1}^{n} X_{hni}, \quad R_{hn} = \sum_{i=1}^{n} X_{hni}^2, \quad S_{hn} = \sum_{i=1}^{n} X_{hni}^2 r\left(X_{hni}\right)$$

and $r(t) = (\log(1+t) - t + t^2/2)/t^2$.

Here $R_{hn} - \mathrm{E}R_{hn} = o_P(1)$. Indeed, let $\epsilon > 0$ and, for $\delta > 0$, put $\bar{X}_{hni} = X_{hni} I(|X_{hni}| \le \delta)$ and $\bar{R}_{hn} = \sum_{i=1}^{n} \bar{X}_{hni}^2$. Now

$$\begin{aligned} &\mathrm{P}(|R_{hn} - \mathrm{E}R_{hn}| > \epsilon) \\ &\le \mathrm{P}(|R_{hn} - \bar{R}_{hn}| > \epsilon/3) + \mathrm{P}(|\bar{R}_{hn} - \mathrm{E}\bar{R}_{hn}| > \epsilon/3) + \mathrm{P}(|\mathrm{E}\bar{R}_{hn} - \mathrm{E}R_{hn}| > \epsilon/3). \end{aligned}$$

Firstly,

$$\begin{aligned} \mathrm{P}(|R_{hn} - \bar{R}_{hn}| > \epsilon/3) &\le \mathrm{P}\left(\max_{i=1,\ldots,n} |X_{hni}| > \delta\right) \le \sum_{i=1}^{n} \mathrm{P}(|X_{hni}| > \delta) \\ &= n\mathrm{P}(|X_{hn1}| > \delta) \le \delta^{-2} n \int_{(|X_{hn1}|>\delta)} X_{hn1}^2 d\mathrm{P} = o(1) \end{aligned}$$

by assumption (iv). Secondly,

$$\begin{aligned} &\mathrm{P}(|\bar{R}_{hn} - \mathrm{E}\bar{R}_{hn}| > \epsilon/3) \\ &\le 3^2\epsilon^{-2}\mathrm{E}\left(\bar{R}_{hn} - \mathrm{E}\bar{R}_{hn}\right)^2 = 3^2\epsilon^{-2}\mathrm{E}\left(\sum_{i=1}^{n}(\bar{X}_{hni}^2 - \mathrm{E}\bar{X}_{hni}^2)\right)^2 \\ &= 3^2\epsilon^{-2}n\mathrm{E}\left(\bar{X}_{hn1}^2 - \mathrm{E}\bar{X}_{hn1}^2\right)^2 \le 3^2\epsilon^{-2}n\mathrm{E}\bar{X}_{hn1}^4 \le 3^2\epsilon^{-2}\delta^2 n\mathrm{E}\bar{X}_{hn1}^2 = \delta^2 O(1) \end{aligned}$$

by assumption (iii). Thirdly, $\mathrm{P}(|\mathrm{E}\bar{R}_{hn} - \mathrm{E}R_{hn}| > \epsilon/3) = o(1)$, because

$$\begin{aligned} |\mathrm{E}\bar{R}_{hn} - \mathrm{E}R_{hn}| &= \left|\sum_{i=1}^{n} \mathrm{E}(\bar{X}_{hni}^2 - X_{hni}^2)\right| = n\left|\mathrm{E}(\bar{X}_{hn1}^2 - X_{hn1}^2)\right| \\ &= n\int_{(|X_{hn1}|>\delta)} X_{hn1}^2 d\mathrm{P} = o(1) \end{aligned}$$

by assumption (iv). Because δ was chosen arbitrarily, $R_{hn} = \mathrm{E}R_{hn} + o_P(1)$. Also, $\mathrm{E}R_{hn} = n\mathrm{E}X_{hn1}^2 = \|h\|_{\mathbf{H}}^2 + o_P(1)$ by assumption (iii).

Further, $S_{hn} = o_P(1)$. Indeed, $|S_{hn}| \le R_{hn} \max_{i=1,\ldots,n} |r(X_{hni})|$, it was just proved that $R_{hn} = O_P(1)$, and it holds that $\max_{i=1,\ldots,n} |r(X_{hni})| = o_P(1)$. Indeed, because

$\lim_{t\to 0} r(t) = 0$, for given $\epsilon > 0$ there is such $\delta > 0$ that $(|r(X_{hn1})| > \epsilon) \subset (|X_{hn1}| > \delta)$. Thus,

$$\begin{aligned} \mathrm{P}\left(\max_{i=1,\dots,n} |r(X_{hni})| > \epsilon\right) &\leq n\mathrm{P}(|r(X_{hn1})| > \epsilon) \leq n\mathrm{P}(|X_{hn1}| > \delta) \\ &\leq n\delta^{-2} \int_{(|X_{hn1}|>\delta)} X_{hn1}^2 d\mathrm{P} = o(1) \end{aligned}$$

by assumption (iv). Let $J = \{h_1, \dots, h_k\} \subset \mathbf{H}_c$ and $a = (a_1, \dots, a_k)' \in \mathbf{R}^k$. Now

$$\begin{aligned} \sum_{i=1}^k a_i \log \frac{f_{h_i,n}^{(n)}(X^{(n)})}{f_0^{(n)}(X^{(n)})} &= \sum_{i=1}^k a_i \left(Q_{h_i,n} - \frac{1}{2} R_{h_i,n} + S_{h_i,n}\right) \\ &= \sum_{i=1}^k a_i Q_{h_i,n} + \sum_{i=1}^k a_i \left(-\frac{1}{2}\|h_i\|_{\mathbf{H}}^2\right) + o_P(1). \end{aligned}$$

It remains to prove that $\sum_{i=1}^k a_i Q_{h_i,n} \xrightarrow{d} N(0, a'\Sigma a)$. Note that $\sum_{i=1}^k a_i Q_{h_i,n} = \sum_{j=1}^n \sum_{i=1}^k a_i X_{h_i nj}$ is a sum of independent identically distributed random variables. Also,

$$\mathrm{E}\left[\sum_{i=1}^k a_i X_{h_i n1}\right] = \sum_{i=1}^k a_i \mathrm{E} X_{h_i n1} = 0$$

by assumption (ii), and

$$\begin{aligned} \mathrm{Var}\left[\sum_{i=1}^k a_i Q_{h_i,n}\right] &= n\mathrm{Var}\left[\sum_{i=1}^k a_i X_{h_i n1}\right] = n \sum_{i,j=1}^k a_i a_j \mathrm{E}\left[X_{h_i n1} X_{h_j n1}\right] \\ &= \sum_{i,j=1}^k a_i a_j < h_i, h_j >_{\mathbf{H}} + o(1) \end{aligned}$$

by assumption (iii). By assumption (iv), for $j = 1, \dots n$,

$$\sum_{i=1}^k a_i X_{h_i nj} = \sum_{i=1}^k a_i (T_n h_i)(X_j) f_0(X_j)^{-1} = T_n \left(\sum_{i=1}^k a_i h_i\right)(X_j) f_0(X_j)^{-1}$$

and Lindeberg's condition is satisfied. □

Note that in the above proof it was shown that the LAN condition, as formulated by Ibragimov and Khas'minskii (1991), is satisfied.

6.2 Pointwise Risk

Let $(\mathcal{X}, \mathcal{F}, \mu)$ and $(\mathcal{Y}, \mathcal{G}, \nu)$ be measurable spaces and let $\mathbb{F}$ and $\mathbb{G}$ be spaces of measurable functions $\mathcal{X} \to \mathbf{R}$ and $\mathcal{Y} \to \mathbf{R}$, respectively. Let $\mathbf{F} \subset \mathbb{F}$ and $\mathbf{G} \subset \mathbb{G}$ be linear subspaces. Assume that $\mathbf{F}$ contains a density f_0. Let $T_n : \mathbb{G} \to \mathbb{F}$ be a mapping that possibly depends on f_0 and such that $T_n\mathbf{G} \subset \mathbf{F}$. Let $\mathbf{H}_c \subset \mathbf{H} \cap \mathbf{G}$, where $0 \in \mathbf{H}_c$ and $\mathbf{H} = L_2(\mathcal{Y}, \mathcal{G}, \nu)$.

Assume that

$$E_n(\mathbf{H}_c) \xrightarrow{w} E(\mathbf{H}_c), \tag{6.1}$$

where $E_n(\mathbf{H}_c)$ and $E(\mathbf{H}_c)$ are as defined in the previous section. Assume secondly that there are such linear functions $D_{\mathbf{F}} : \mathbf{F} \to \mathbb{F}$ and $D_{\mathbf{G}} : \mathbf{G} \to \mathbb{G}$ that for a fixed $x_0 \in \mathcal{X}$,

$$D_{\mathbf{F}}(T_n h)(x_0) = \gamma_n^{-1}\Phi(f_0)\left(D_{\mathbf{G}}h(y_0) + \delta_n(h)\right), \tag{6.2}$$

where $\lim_{n\to\infty} \gamma_n = \infty$, $\Phi(f_0) \in \mathbf{R}$, $\Phi(f_0) \neq 0$, $y_0 \in \mathcal{Y}$, $\lim_{n\to\infty} \sup_{h\in\mathbf{H}_c} |\delta_n(h)| = 0$, and $\sup_{h\in\mathbf{H}_c} |D_{\mathbf{G}}h(y_0)| < \infty$. The operators $D_{\mathbf{F}}$ and $D_{\mathbf{G}}$ are introduced to take care of the estimation of derivatives.

Lemma 6.3 *Let $p > 0$. Under the assumptions (6.1) and (6.2),*

$$\liminf_{n\to\infty} \gamma_n^p \inf_{\rho_n \in \mathcal{R}(E_n(\mathbf{H}_c),\mathbf{R})} \sup_{h\in\mathbf{H}_c} \int_{\mathcal{X}^n} dP_{h,n}^n\left(x^{(n)}\right) \int_{\mathbf{R}} |D_{\mathbf{F}} f_{h,n}(x_0) - a|^p \rho_n\left(x^{(n)}, da\right)$$

$$\geq |\Phi(f_0)|^p \inf_{\rho\in\mathcal{R}(E(\mathbf{H}_c),\mathbf{R})} \sup_{h\in\mathbf{H}_c} \int_{\Omega} dG_h(\omega) \int_{\mathbf{R}} |D_{\mathbf{G}}h(y_0) - a|^p \rho(\omega, da).$$

Proof. By the assumption (6.2),

$$\begin{aligned}
|D_{\mathbf{F}} f_{h,n}(x_0) - a|^p &= |D_{\mathbf{F}} f_0(x_0) + D_{\mathbf{F}}(T_n h)(x_0) - a|^p \\
&= |D_{\mathbf{F}} f_0(x_0) + \gamma_n^{-1}\Phi(f_0)(D_{\mathbf{G}}h(y_0) + \delta_n(h)) - a|^p \\
&= |\gamma_n^{-1}\Phi(f_0)|^p |D_{\mathbf{G}}h(y_0) - \gamma_n\Phi(f_0)^{-1}(a - D_{\mathbf{F}} f_0(x_0)) + \delta_n(h)|^p.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\inf_{\rho_n} \sup_{h\in\mathbf{H}_c} \int_{\mathcal{X}^n} dP_{h,n}^n\left(x^{(n)}\right) \int_{\mathbf{R}} |D_{\mathbf{F}} f_{h,n}(x_0) - a|^p \rho_n\left(x^{(n)}, da\right) \\
&\quad = |\gamma_n^{-1}\Phi(f_0)|^p \inf_{\rho_n} \sup_{h\in\mathbf{H}_c} \int_{\mathcal{X}^n} dP_{h,n}^n\left(x^{(n)}\right) \int_{\mathbf{R}} |D_{\mathbf{G}}h(y_0) - a + \delta_n(h)|^p \rho_n\left(x^{(n)}, da\right).
\end{aligned}$$

Now $\mathcal{R}(E_n(\mathbf{H}_c), \mathbf{R}) \subset \mathcal{B}(E_n(\mathbf{H}_c), \mathbf{R})$ and $\mathcal{B}(E(\mathbf{H}_c), \mathbf{R}) = \{\beta_\rho \,|\, \rho \in \mathcal{R}(E(\mathbf{H}_c), \mathbf{R})\}$ by the facts that $E(\mathbf{H}_c)$ is dominated and the decision space $\mathbf{R}$ is locally compact with a countable base. Thus the assertion follows from (6.1) and Lemma 6.1. The loss function is $W_h : \mathbf{R} \to \mathbf{R}$, $W_h(a) = |(D_\mathbf{G} h)(y_0) - a|^p$, where $h \in \mathbf{H}_c$. □

An alternative method for deriving lower bounds when estimating the value of a linear functional is presented in Donoho and Liu (1991). Their method can produce, in some cases, nearly optimal constants and optimal linear estimators. The optimality of the lower bound of the Lemma 6.3 is not discussed here but note that the constant given in Lemma 6.3 has a lower bound that will be derived in the next lemma. The method of Lemma 6.3 has the advantage that it can be modified to give lower bounds in the case of an integrated risk, as will be seen in the next section.

Lemma 6.4 *Let $\mathbf{H}_c \subset \mathbf{H}$ be convex. Let $T : \mathbf{H}_c \to \mathbf{R}$ be linear. Define*

$$\omega(\epsilon) = \sup\{|T(h_1) - T(h_2)| \;\mid\; h_0, h_1 \in \mathbf{H}_c,\ \|h_1 - h_2\|_\mathbf{H} \le \epsilon\}$$

and

$$\psi(\tau, \sigma) = \inf_{\hat{T}} \sup_{|\theta| \le \tau} \int_\mathbf{R} \left| \hat{T} - \theta \right|^p dN(\theta, \sigma^2), \tag{6.3}$$

where the infimum is taken with respect to measurable functions $\mathbf{R} \to \mathbf{R}$. Then

$$\inf_{\rho \in \mathcal{R}(E(\mathbf{H}_c), \mathbf{R})} \sup_{h \in \mathbf{H}_c} \int_\Omega dG_h(\omega) \int_\mathbf{R} |T(h) - a|^p \rho(\omega, da) \ge \sup_{\epsilon > 0} \left(\frac{\omega(\epsilon)}{\epsilon} \right)^p \psi\left(\frac{\epsilon}{2}, 1 \right).$$

Proof. The argument (in the case $p = 2$) given by Donoho and Liu (1991) or by Ibragimov and Khas'minskii (1984) (for a centrosymmetric $\mathbf{H}_c$) is seen to apply here when it is noted that $\psi(\tau, \sigma) = \sigma^p \psi(\tau/\sigma, 1)$. □

Information and references concerning the function $\psi(\tau, \sigma)$ (in the case $p = 2$) is given in Donoho, Liu and MacGibbon (1990). Note that when $\omega(\epsilon) = A\epsilon^r$, $0 < A < \infty$, $0 < r \le 1$, the lower bound of Lemma 6.4 can be written as

$$\left(\frac{\omega(2)}{2} \right)^p \xi(r), \tag{6.4}$$

where $\xi(r) = \sup_{\epsilon > 0} \epsilon^{p(r-1)} \psi(\epsilon, 1)$.

6.2.1 Euclidean case

Let $\mathcal{X} = \mathbf{R}^d$, let μ be the Lebesgue measure of $\mathbf{R}^d$, and $\mathcal{F}$ the Borel sigma–algebra of $\mathbf{R}^d$. Let $x_0 \in \mathbf{R}^d$. For a density $f : \mathbf{R}^d \to \mathbf{R}$, the estimation of $D^\alpha f(x_0)$ will be discussed, where $\alpha = (\alpha_1, \ldots, \alpha_d)$, $|\alpha| = \alpha_1 + \cdots + \alpha_d = r$, $r \in \{0, 1, \ldots\}$, $D^\alpha f = D_1^{\alpha_1} \cdots D_d^{\alpha_d} f$, and D_i means the ith partial derivative.

Let $s \in \{r+1, r+2, \ldots\}$ and $0 < a < \infty$. Assume that the unknown density belongs to the set

$$\mathbf{F}(s, a, x_0) = \left\{ f : \mathbf{R}^d \to \mathbf{R} \,\middle|\, f \text{ density}, \ |D^\beta f(x_0)| \le a, \text{ for } |\beta| = s \right\}.$$

Let $f_0 : \mathbf{R}^d \to \mathbf{R}$ be a density function, $f_0(x_0) > 0$, let f_0 be continuous at x_0, and let $|D^\beta f_0(x_0)| < a$ for all $|\beta| = s$. Let $\mathcal{Y} = \mathcal{X} = \mathbf{R}^d$, $\nu = \mu$, and $\mathcal{G} = \mathcal{F}$. Let $\mathbf{H}_c$, $0 < c < \infty$, be the set of measurable functions $h : \mathcal{Y} \to \mathbf{R}$ for which h has support on A, where A is bounded with 0 as its interior point, $\int_{\mathcal{Y}} h d\nu = 0$, $\sup_{y \in \mathcal{Y}} |h(y)| \le c$, $|D^\alpha h(0)| \le c$ for $|\alpha| = r$, and $|D^\beta h(0)| \le (a - |D^\beta f_0(x_0)|)/\Phi(f_0)$ for $|\beta| = s$, where $\Phi(f_0) = \sqrt{f_0(x_0)}$. Let

$$(T_n h)(x) = \Phi(f_0) n^{-1/2} \kappa_n^{d/2} h(\kappa_n(x - x_0)) I_{A_n}(x),$$

where $\lim_{n\to\infty} \kappa_n = \infty$, $\lim_{n\to\infty} n^{-1}\kappa_n^d = 0$, and $A_n = \{x \in \mathbf{R}^d \,|\, \kappa_n(x - x_0) \in A\}$. From Lemma 6.2 it is seen that $E_n(\mathbf{H}_c) \xrightarrow{w} E(\mathbf{H}_c)$. By choosing $\kappa_n = n^{1/(2s+d)}$ we obtain $\{f_{h,n} \,|\, h \in \mathbf{H}_c\} \subset \mathbf{F}(s, a, x_0)$ ultimately. Finally, note that (6.2) is satisfied because $D^\alpha (T_n h)(x_0) = \Phi(f_0) n^{-1/2} \kappa_n^{r+d/2} D^\alpha h(0)$ and Lemma 6.3 can be applied with

$$\gamma_n = n^{1/2} \kappa_n^{-r-d/2} = n^{(s-r)/(2s+d)}$$

and $D_{\mathbf{F}} = D_{\mathbf{G}} = D^\alpha$. Analogous results have been proved by Farrell (1972), Meyer (1977), Stone (1980), and Low (1992).

6.2.2 Spherical case

Let $\mathcal{X} = S_d$, $d \ge 2$, let μ be the Lebesgue measure of S_d, and $\mathcal{F}$ the Borel sigma–algebra of S_d. Let $x_0 \in S_d$. For a density $f : S_d \to \mathbf{R}$, estimation of $\Delta^{r/2} f(x_0)$ will be discussed, where $r \ge 0$ is even and Δ is the Laplace operator of S_d.

Let $s > r$ be even and $0 < a < \infty$. Set

$$\mathbf{F}(s, a, x_0) = \left\{ f : S_d \to \mathbf{R} \,\middle|\, f \text{ density}, \ |\Delta^{s/2} f(x_0)| \le a \right\}.$$

Let $f_0 : S_d \to \mathbf{R}$ be a density function, $f_0 \geq \epsilon > 0$ in S_d, suppose f_0 is continuous at x_0 and that $|\Delta^{s/2} f_0(x_0)| < a$. Take $\mathcal{Y} = [0, \infty[$, let $\nu = t^{(d-2)/2} dt$, where dt is the Lebesgue measure of $\mathcal{Y}$, and let $\mathcal{G}$ be the Borel sigma–algebra of $\mathcal{Y}$. Let $\mathbf{H}_\infty$ be the set of measurable functions $h : \mathcal{Y} \to \mathbf{R}$ for which $\int_0^\infty |h| d\nu < \infty$, $\sup_{y \in \mathcal{Y}} |h(y)| < \infty$, $h^{(s)}$ exists on $[0, \infty[$, and $|Q_{s,s/2}(1) h^{(s/2)}(0)| \leq (a - |\Delta^{s/2} f_0(x_0)|)/\Psi(f_0)$, where $Q_{s,s/2}$ is the polynomial specified in Lemma 5.8, and $\Psi(f_0) = (2^{(d-2)/2} \omega_{d-1})^{-1/2} \sqrt{f_0(x_0)}$.

Theorem 6.5 *For* $p > 0$,

$$\liminf_{n \to \infty} n^{p(s-r)/(2s+d)} \inf_{\rho_n} \sup_{f \in \mathbf{F}(s,a,x_0)} \int_{S_d^n} dP_f^n \left(x^{(n)} \right) \int_{\mathbf{R}} |\Delta^{r/2} f(x_0) - a|^p \rho_n \left(x^{(n)}, da \right)$$

$$\geq |\Phi(f_0)|^p \inf_{\rho} \sup_{h \in \mathbf{H}_\infty} \int_\Omega dG_h(\omega) \int_{\mathbf{R}} |h^{(r/2)}(0) - a|^p \rho(\omega, da),$$

where $\Phi(f_0) = \Psi(f_0) Q_{r,r/2}(1)(-1)^{r/2}$.

Proof. Define, for $h : \mathcal{Y} \to \mathbf{R}$, $\int_0^\infty |h| d\nu < \infty$,

$$T_n h = V_n h - \omega_d^{-1} \int_{S_d} (V_n h) d\mu,$$

where

$$(V_n h)(x) = \Psi(f_0) n^{-1/2} \kappa_n^{d/2} h(\kappa_n^2 (1 - x' x_0)), \;\; x \in S_d,$$

where $\{\kappa_n\}$ is such a sequence that $\lim_{n \to \infty} \kappa_n = \infty$ and $\lim_{n \to \infty} n^{-1} \kappa_n^d = 0$. Let $\mathbf{H}_c$, $0 < c < \infty$, be the set of those functions in $\mathbf{H}_\infty$ for which $\sup_{y \in \mathcal{Y}} |h(y)| \leq c$, $|h^{(i)}(0)| \leq c$ for $i = 1, \ldots, s/2 - 1$, and $(1 + c^{-1}) |Q_{s,s/2}(1) h^{(s/2)}(0)| \leq (a - |\Delta^{s/2} f_0(x_0)|)/\Psi(f_0)$.

First, note that the conditions of Lemma 6.2 are satisfied. Trivially $\int_{S_d} (T_n h) d\mu = 0$. Because $n^{1/2} \int_{S_d} (V_n h) d\mu = O(\kappa_n^{-d/2})$ and $n^{1/2} \int_{S_d} (V_n h) f_0^{-1} d\mu = O(\kappa_n^{-d/2})$,

$$\begin{aligned} n \int_{S_d} (T_n h_1)(T_n h_2) f_0^{-1} d\mu &= n \int_{S_d} (V_n h_1)(V_n h_2) f_0^{-1} d\mu + O\left(\kappa_n^{-d}\right) \\ &= \int_{\mathcal{Y}} t^{(d-2)/2} h_1(t) h_2(t) dt + o(1) \end{aligned}$$

by Lemma 2.4. Secondly, choosing $\kappa_n = n^{1/(2s+d)}$ it follows that $\{f_{h,n} \,|\, h \in \mathbf{H}_c\} \subset \mathbf{F}(s, a, x_0)$ ultimately. In fact, now $n^{-1/2} \kappa_n^{d/2} = \kappa_n^{-s}$ and by Lemma 5.8,

$$\Delta^{s/2} (T_n h)(x) \;\; = \;\; \Delta^{s/2} (V_n h)(x)$$

$$
\begin{aligned}
&= \Psi(f_0)\kappa_n^{-s}\sum_{i=1}^{s} P_{s,i}(x'x_0)(-1)^i\kappa_n^{2i}h^{(i)}(\kappa_n^2(1-x'x_0)) \\
&= \Psi(f_0)\kappa_n^{-s}\left[\sum_{i=1}^{s/2-1} P_{s,i}(x'x_0)(-1)^i\kappa_n^{2i}h^{(i)}(\kappa_n^2(1-x'x_0))\right. \\
&\quad \left. + \sum_{i=s/2}^{s}\left(1-(x'x_0)^2\right)^{i-s/2} Q_{s,i}(x'x_0)(-1)^i\kappa_n^{2i}h^{(i)}(\kappa_n^2(1-x'x_0))\right],
\end{aligned}
$$

where $P_{s,i}$ and $Q_{s,i}$ are polynomials. Thus, for $h \in \mathbf{H}_c$,

$$
\begin{aligned}
&|\Delta^{s/2}f_{h,n}(x_0)| = \left|\Delta^{s/2}f_0(x_0) + \Delta^{s/2}(V_n h)(x_0)\right| \\
&\leq |\Delta^{s/2}f_0(x_0)| + \Psi(f_0)\left|Q_{s,s/2}(1)h^{(s/2)}(0)\right| + \Psi(f_0)\kappa_n^{-s}\sum_{i=1}^{s/2-1}\kappa_n^{2i}\left|P_{s,i}(1)h^{(i)}(0)\right| \leq a
\end{aligned}
$$

for sufficiently large n. Thirdly, (6.2) is satisfied because

$$
\Delta^{r/2}(T_n h)(x_0) = \Psi(f_0)\kappa_n^{r-s}\left[Q_{r,r/2}(1)(-1)^{r/2}h^{(r/2)}(0) + \sum_{i=1}^{r/2-1} P_{r,i}(1)(-1)^i\kappa_n^{2i-r}h^{(i)}(0)\right].
$$

Let $\Pi_0(\mathbf{H}_c)$ be the set of probability measures on $\mathbf{H}_c$ with finite support. Because $c > 0$ was chosen arbitrarily, we have the following lower bound by applying Lemma 6.3 with $\gamma_n = \kappa_n^{s-r} = n^{(s-r)/(2s+d)}$,

$$
\begin{aligned}
&|\Phi(f_0)|^p \sup_{c>0}\inf_{\rho}\sup_{h\in\mathbf{H}_c}\int_{\Omega} dG_h(\omega)\int_{\mathbf{R}}|h^{(r/2)}(0)-a|^p\rho(\omega,da) \\
&= |\Phi(f_0)|^p \sup_{c>0}\sup_{\pi\in\Pi_0(\mathbf{H}_c)}\inf_{\rho}\int_{\mathbf{H}_c} d\pi(h)\int_{\Omega} dG_h(\omega)\int_{\mathbf{R}}|h^{(r/2)}(0)-a|^p\rho(\omega,da) \\
&= |\Phi(f_0)|^p \sup_{\pi\in\Pi_0(\mathbf{H}_\infty)}\inf_{\rho}\int_{\mathbf{H}_\infty} d\pi(h)\int_{\Omega} dG_h(\omega)\int_{\mathbf{R}}|h^{(r/2)}(0)-a|^p\rho(\omega,da) \\
&= |\Phi(f_0)|^p \inf_{\rho}\sup_{h\in\mathbf{H}_\infty}\int_{\Omega} dG_h(\omega)\int_{\mathbf{R}}|h^{(r/2)}(0)-a|^p\rho(\omega,da),
\end{aligned}
$$

where the order of infimum and supremum can be changed by Strasser's (1985) Theorem 46.3. □

6.3 Integrated Risk

Let $(\mathcal{X}, \mathcal{F}, \mu)$ and $(\mathcal{Y}, \mathcal{G}, \nu)$ be measurable spaces and denote $\mathbf{F} = L_p(\mathcal{X}, \mathcal{F}, \mu)$ and $\mathbf{G} = L_p(\mathcal{Y}, \mathcal{G}, \nu)$, where $p \geq 1$. Let I_n be a finite index set with cardinality tending to infinity as $n \to \infty$. Let $A_{n,i} \subset \mathcal{X}$, $i \in I_n$, be such that $\mu(A_{n,i} \cap A_{n,j}) = 0$ for $i \neq j$. Denote $\mathbf{F}_{n,i} = \{ f I_{A_{n,i}} \,|\, f \in \mathbf{F} \}$, where $I_{A_{n,i}}$ is the indicator of the set $A_{n,i}$. Let $f_0 \in \mathbf{F}$ be a density whose support includes all $A_{n,i}$. Let $T_{n,i} : \mathbf{G} \to \mathbf{F}_{n,i}$, $i \in I_n$, be mappings that possibly depend on f_0. Let us denote

$$f_{\underline{h}_n,n} = f_0 + \sum_{i \in I_n} T_{n,i} h_i,$$

where $\underline{h}_n = (h_i)_{i \in I_n} \in \mathbf{G}^{I_n}$. Let $\mathbf{H}_c \subset \mathbf{H} \cap \mathbf{G}$, where $0 \in \mathbf{H}_c$ and $\mathbf{H} = L_2(\mathcal{Y}, \mathcal{G}, \nu)$. It is assumed that ultimately $\int_{A_{n,i}} (T_{n,i} h) d\mu = 0$ for $h \in \mathbf{H}_c$, $f_{\underline{h}_n,n} \geq 0$ (a.e. μ) for $\underline{h}_n \in \mathbf{H}_c^{I_n}$, and $\{ f_{\underline{h}_n,n} \,|\, \underline{h}_n \in \mathbf{H}_c^{I_n} \} \subset \mathbf{F}$.

When $\{k(n)\}$ is a sequence with $k(n) \in I_n$, put $T_n = T_{n,k(n)}$, and $f_{h,n} = f_0 + T_n h$. Let

$$E_n(\mathbf{H}_c) = \left(\mathcal{X}^n, \mathcal{F}^n, P_{h,n}^n : h \in \mathbf{H}_c \right)$$

be the direct product of experiment $(\mathcal{X}, \mathcal{F}, P_{h,n} : h \in \mathbf{H}_c)$, where $P_{h,n}$ corresponds to $f_{h,n}$. Assume that for every sequence $\{k(n)\}$ with $k(n) \in I_n$,

$$E_n(\mathbf{H}_c) \xrightarrow{w} E(\mathbf{H}_c), \tag{6.5}$$

where $E(\mathbf{H}_c)$ is a Gaussian shift experiment defined in Section 6.1. Assume secondly that for all $h \in \mathbf{H}_c$, $f \in \mathbf{F}_{n,i}$,

$$\int_{A_{n,i}} |T_{n,i} h - f|^p d\mu \geq \Phi_{n,i}^{(p)}(f_0) \int_{\mathcal{Y}} |h - U_{n,i} f|^p d\nu, \tag{6.6}$$

where $U_{n,i} : \mathbf{F}_{n,i} \to \mathbf{G}$ is continuous, $\Phi_{n,i}^{(p)}(f_0) > 0$, and

$$\gamma_n^p \sum_{i \in I_n} \Phi_{n,i}^{(p)}(f_0) = \Phi^{(p)}(f_0) + o(1),$$

as $n \to \infty$, where $\Phi^{(p)}(f_0) > 0$ and $\lim_{n \to \infty} \gamma_n = \infty$. The proof of the following lemma is based on ideas similar to those used by Golubev and Nussbaum (1990) with a regression model and by Low (1993) with a white noise model. Also Golubev (1991) used the same kind of reasoning.

Lemma 6.6 *Let $p \geq 1$. Let $E_n^* = \left(\mathcal{X}^n, \mathcal{F}^n, P_{\underline{h}_n,n}^n : \underline{h}_n \in \mathbf{H}_c^{I_n} \right)$, where $P_{\underline{h}_n,n}$ corresponds to $f_{\underline{h}_n,n}$. Under the assumptions (6.5) and (6.6),*

$$\liminf_{n \to \infty} \gamma_n^p \inf_{\beta_n \in \mathcal{B}(E_n^*, \mathbf{F})} \sup_{\underline{h}_n \in \mathbf{H}_c^{I_n}} \beta_n \left(W_{\underline{h}_n}, P_{\underline{h}_n,n} \right) \geq \Phi^{(p)}(f_0) \inf_{\beta \in \mathcal{B}(E(\mathbf{H}_c), \mathbf{G})} \sup_{h \in \mathbf{H}_c} \beta(W_h, G_h),$$

where $W_{\underline{h}_n} : \mathbf{F} \to \mathbf{R}$, $W_{\underline{h}_n}(a) = \int_{\mathcal{X}} |f_{\underline{h}_n,n} - a|^p d\mu$ *and* $W_h : \mathbf{G} \to \mathbf{R}$, $W_h(a) = \int_{\mathcal{Y}} |h - a|^p d\nu$.

Proof. By the assumption (6.6), for $\underline{h}_n \in \mathbf{H}_c^{I_n}$ and $a \in \mathbf{F}$,

$$\begin{aligned} W_{\underline{h}_n}(a) &= \int_{\mathcal{X}} |f_{\underline{h}_n,n} - a|^p d\mu \geq \sum_{i \in I_n} \int_{A_{n,i}} |f_0 + T_{n,i}h_i - a|^p d\mu \\ &\geq \sum_{i \in I_n} \Phi_{n,i}^{(p)}(f_0) \int_{\mathcal{Y}} |h_i - U_{n,i}((a - f_0)I_{A_{n,i}})|^p d\nu \\ &= \sum_{i \in I_n} \Phi_{n,i}^{(p)}(f_0) W_{h_i}\left(U_{n,i}((a - f_0)I_{A_{n,i}})\right). \end{aligned}$$

Thus, for every $\underline{h}_n \in \mathbf{H}_c^{I_n}$ and $\beta_n \in \mathcal{B}(E_n^*, \mathbf{F})$,

$$\beta_n\left(W_{\underline{h}_n}, P_{\underline{h}_n,n}\right) \geq \sum_{i \in I_n} \Phi_{n,i}^{(p)}(f_0) \beta_n\left(W_{h_i}\left(U_{n,i}((\cdot - f_0)I_{A_{n,i}})\right), P_{\underline{h}_n,n}\right).$$

Let π be a probability measure on $\mathbf{H}_c$ with a support of finite cardinality and set $\pi_n = \prod_{i \in I_n} \pi$, so that π_n is a probability measure on $\mathbf{H}_c^{I_n}$. Now

$$\begin{aligned} &\inf_{\beta_n \in \mathcal{B}(E_n^*, \mathbf{F})} \sup_{\underline{h}_n \in \mathbf{H}_c^{I_n}} \beta_n\left(W_{\underline{h}_n}, P_{\underline{h}_n,n}\right) \\ &\geq \sum_{i \in I_n} \Phi_{n,i}^{(p)}(f_0) \inf_{\beta_n \in \mathcal{B}(E_n^*, \mathbf{F})} \int_{\mathbf{H}_c^{I_n}} \beta_n\left(W_{h_i}\left(U_{n,i}((\cdot - f_0)I_{A_{n,i}})\right), P_{\underline{h}_n,n}\right) d\pi_n(\underline{h}_n) \\ &\geq \sum_{i \in I_n} \Phi_{n,i}^{(p)}(f_0) \inf_{\beta_n \in \mathcal{B}(E_n^*, \mathbf{G})} \int_{\mathbf{H}_c^{I_n}} \beta_n\left(W_{h_i}, P_{\underline{h}_n,n}\right) d\pi_n(\underline{h}_n) \\ &\geq \sum_{i \in I_n} \Phi_{n,i}^{(p)}(f_0) \int_{\mathbf{H}_c^{I_n \setminus \{i\}}} \prod_{j \in I_n \setminus \{i\}} d\pi(h_j) \inf_{\beta_n \in \mathcal{B}(E_n^*, \mathbf{G})} \int_{\mathbf{H}_c} \beta_n\left(W_{h_i}, P_{\underline{h}_n,n}\right) d\pi(h_i). \end{aligned}$$

Let $i \in I_n$ and $\underline{h}_{n,i} = (h_j)_{j \in I_n \setminus \{i\}} \in \mathbf{H}_c^{I_n \setminus \{i\}}$ be fixed and denote $E_{n,i}^* = (\mathcal{X}^n, \mathcal{F}^n, f_{\underline{h}_n,n}^{(n)} : h_i \in \mathbf{H}_c)$. If we write $f_{\underline{h}_{n,i},n} = f_0 + \sum_{j \in I_n, j \neq i} T_{n,j}h_j$, then we can write $E_{n,i}^* = (\mathcal{X}^n, \mathcal{F}^n, (f_{\underline{h}_{n,i},n} + T_{n,i}h_i)^{(n)} : h_i \in \mathbf{H}_c)$. Because $L(E_{n,i}^*) = L_1(\mathcal{X}, \mathcal{F}, P_0) = L(E_n^*)$, then $\mathcal{B}(E_n^*, \mathbf{G}) = \mathcal{B}(E_{n,i}^*, \mathbf{G})$. Thus,

$$\inf_{\beta_n \in \mathcal{B}(E_n^*, \mathbf{G})} \int_{\mathbf{H}_c} \beta_n\left(W_{h_i}, P_{\underline{h}_n,n}\right) d\pi(h_i) = \inf_{\beta_n \in \mathcal{B}(E_{n,i}^*, \mathbf{G})} \int_{\mathbf{H}_c} \beta_n\left(W_{h_i}, P_{\underline{h}_n,n}\right) d\pi(h_i).$$

Let us denote $f_{h,n,i} = f_0 + T_{n,i}h$ and $E_{n,i} = (\mathcal{X}^n, \mathcal{F}^n, f_{h,n,i}^{(n)} : h \in \mathbf{H}_c)$. It holds that $E_{n,i}^* \subset E_{n,i}$ in the sense of Strasser (1985, Definition 49.5). Indeed, because the mapping $M : L_1(\mathcal{X}, \mathcal{F}, P_0) \to L_1(\mathcal{X}, \mathcal{F}, P_0)$, defined by

$$Mg = gI_{A_{n,i}} + \left[P_0(A_{n,i}^c)^{-1} \int_{A_{n,i}^c} g dP_0 \right] f_{\underline{h}_{n,i},n} f_0^{-1} I_{A_{n,i}^c},$$

is a stochastic operator for which $M(f_{h_i,n,i} f_0^{-1}) = f_{\underline{h}_n,n} f_0^{-1}$ for every $h_i \in \mathbf{H}_c$, we can apply Strasser (1985, Corollary 55.10). Thus,

$$\begin{aligned}
&\inf_{\beta_n \in \mathcal{B}(E_{n,i}^*, \mathbf{G})} \int_{\mathbf{H}_c} \beta_n \left(W_{h_i}, P_{\underline{h}_n,n} \right) d\pi(h_i) \\
&= \sup_{V_h \in \mathcal{C}_b(\mathbf{G}),\, V_h \le W_h} \inf_{\beta_n \in \mathcal{B}(E_{n,i}^*, \mathbf{G})} \int_{\mathbf{H}_c} \beta_n \left(V_{h_i}, P_{\underline{h}_n,n} \right) d\pi(h_i) \\
&\ge \sup_{V_h \in \mathcal{C}_b(\mathbf{G}),\, V_h \le W_h} \inf_{\beta_n \in \mathcal{B}(E_{n,i}, \mathbf{G})} \int_{\mathbf{H}_c} \beta_n \left(V_h, P_{h,n,i} \right) d\pi(h) \\
&= \inf_{\beta_n \in \mathcal{B}(E_{n,i}, \mathbf{G})} \int_{\mathbf{H}_c} \beta_n \left(W_h, P_{h,n,i} \right) d\pi(h),
\end{aligned}$$

by Strasser (1985, Corollary 47.4). Let $k(n) \in I_n$ be such a sequence that

$$\inf_{\beta_n \in \mathcal{B}(E_{n,k(n)}, \mathbf{G})} \int_{\mathbf{H}_c} \beta_n \left(W_h, P_{h,n,k(n)} \right) d\pi(h) = \min_{i \in I_n} \inf_{\beta_n \in \mathcal{B}(E_{n,i}, \mathbf{G})} \int_{\mathbf{H}_c} \beta_n \left(W_h, P_{h,n,i} \right) d\pi(h).$$

Denote $P_{h,n} = P_{h,n,k(n)}$, $E_n(\mathbf{H}_c) = E_{n,k(n)}$ and let $\Pi_0(\mathbf{H}_c)$ be the set of finite probability measures on $\mathbf{H}_c$. Because the probability measure π was chosen arbitrarily, we get the lower bound

$$\begin{aligned}
&\left(\sum_{i \in I_n} \Phi_{n,i}^{(p)}(f_0) \right) \sup_{\pi \in \Pi_0(\mathbf{H}_c)} \inf_{\beta_n \in \mathcal{B}(E_n(\mathbf{H}_c), \mathbf{G})} \int_{\mathbf{H}_c} \beta_n \left(W_h, P_{h,n} \right) d\pi(h) \\
&= \left(\sum_{i \in I_n} \Phi_{n,i}^{(p)}(f_0) \right) \inf_{\beta_n \in \mathcal{B}(E_n(\mathbf{H}_c), \mathbf{G})} \sup_{h \in \mathbf{H}_c} \beta_n \left(W_h, P_{h,n} \right).
\end{aligned}$$

This identity is proved in Strasser (1985, Theorem 46.3). The assertion of this lemma follows from Lemma 6.1 and the assumption (6.5). □

The next lemma gives a lower bound for the lower bound of Lemma 6.6 when $p = 2$ and when the set of generalized decision procedures coincides with the set of decision procedures. Because the experiment $E(\mathbf{H}_c)$ is dominated, this holds, for example, when the decision space is a Polish space. Because the decision space $\mathbf{G}$ is a L_p–space, it is always a complete metric space. In the Euclidean and spherical cases it is also separable.

Lemma 6.7 *Let* $\mathbf{H}_c \subset \mathbf{H}$ *be convex. Then,*

$$\inf_{\rho \in \mathcal{R}(E(\mathbf{H}_c),\mathbf{G})} \sup_{h \in \mathbf{H}_c} \int_\Omega dG_h(\omega) \int_{\mathbf{G}} \rho(\omega, da) \int_{\mathcal{Y}} |h-a|^2 d\nu$$
$$\geq \sup_{h_1,h_2 \in \mathbf{H}_c \cap L_4(\mathcal{Y},\mathcal{G},\nu)} \psi\left(\frac{1}{2}\|h_1-h_2\|_{\mathbf{H}}, 1\right),$$

where ψ *was defined by the equation (6.3).*

The proof of this lemma can be given using the reasoning in Low (1993, Lemma 1).

6.3.1 Euclidean case

Let $\mathcal{X} = \mathbf{R}^d$, let μ be the Lebesgue measure of $\mathbf{R}^d$ and $\mathcal{F}$ the Borel sigma–algebra of $\mathbf{R}^d$. Estimation of a density $f : \mathbf{R}^d \to \mathbf{R}$ will be discussed.

Let $f_0 : \mathcal{Z} \to \mathbf{R}$ be a density function, where $\mathcal{Z} \subset \mathbf{R}^d$ is bounded. Let $A \subset \mathcal{Z}$ be a rectangle with 0 as an interior point. Let $f_0 \geq \epsilon > 0$ in A, let f_0 be continuous in A and let $\int_{\mathcal{Z}} |D^\beta f_0|^p d\mu < \infty$, for $|\beta| = s \geq 1$. It is assumed that the unknown density belongs to the set

$$\mathbf{F}(s,a,p,f_0) = \left\{ f : \mathcal{Z} \to \mathbf{R} \;\middle|\; f \text{ density}, \int_{\mathcal{Z}} |D^\beta f - D^\beta f_0|^p d\mu \leq a, \text{ for } |\beta| = s \right\},$$

where $0 < a < \infty$. Let $\mathcal{Y} = \mathcal{X}$, $\nu = \mu$, and $\mathcal{G} = \mathcal{F}$. Let $\mathbf{H}_c$, $0 < c < \infty$, be the set of measurable functions $h : \mathcal{Y} \to \mathbf{R}$ for which h has support on A, $\int_{\mathcal{Y}} h d\nu = 0$, $\sup_{y \in \mathcal{Y}} |h(y)| \leq c$, and

$$(1+c^{-1})\mu(A)^{-1} \int_A f_0^{p/2} d\mu \int_{\mathcal{Y}} |D^\beta h|^p d\nu \leq a$$

for $|\beta| = s$.

Let $\{\kappa_n\}$ be a sequence with $\lim_{n\to\infty} \kappa_n = \infty$ and $\lim_{n\to\infty} n^{-1}\kappa_n^d = 0$. Let I_n be an index set with cardinality $\sim \kappa_n^d$. Let $x_{n,i} \in A$, $i \in I_n$, be such that $A_{n,i} = \{x \in \mathbf{R}^d \,|\, \kappa_n(x - x_{n,i}) \in A\}$ are contained in A, almost surely disjoint, and $\cup_{n=1}^\infty \cup_{i \in I_n} \{x_{n,i}\}$ is dense in A. For example, let $A = [-a_1, a_1] \times \cdots \times [-a_d, a_d]$. Let $x_{n,i} = \left(x_1^{(n,i_1)}, \ldots, x_d^{(n,i_d)}\right)' \in \mathbf{R}^d$, where $i = (i_1, \ldots, i_d)' \in I_n = \{-[(\kappa_n - 1)/2], \ldots, [(\kappa_n-1)/2]\}^d$, $[a]$ is the greatest integer $\leq a$, and $x_k^{(n,i_k)} = i_k 2a_k/\kappa_n$, $k \in$

$\{1,\ldots,d\}$. Then $A_{n,i} = A_1^{(n,i_1)} \times \cdots \times A_d^{(n,i_d)}$, where $A_k^{(n,i_k)} = [x_k^{(n,i_k)} - a_k/\kappa_n, x_k^{(n,i_k)} + a_k/\kappa_n]$, $k \in \{1,\ldots,d\}$.

For $h : \mathcal{Y} \to \mathbf{R}$, define

$$(T_{n,i}h)(x) = \Psi_{n,i}(f_0) n^{-1/2} \kappa_n^{d/2} h(\kappa_n(x - x_{n,i})) I_{A_{n,i}}(x),$$

where $i \in I_n$, and $\Psi_{n,i}(f_0) = \sqrt{f_0(x_{n,i})}$. From Lemma 6.2 it is seen that $E_n(\mathbf{H}_c) \xrightarrow{w} E(\mathbf{H}_c)$. By choosing $\kappa_n = n^{1/(2s+d)}$ it holds that $\{f_{\underline{h}_n,n} \,|\, \underline{h}_n \in \mathbf{H}_c^{I_n}\} \subset \mathbf{F}(s,a,p,f_0)$ ultimately. Finally note that (6.6) is satisfied because

$$\int_{A_{n,i}} |T_{n,i}h - f|^p d\mu = \left[\Psi_{n,i}(f_0) n^{-1/2} \kappa_n^{d/2}\right]^p \kappa_n^{-d} \int_{\mathcal{Y}} |h - T_{n,i}^{-1} f|^p d\nu$$

and Lemma 6.6 can be applied with $\Phi_{n,i}^{(p)}(f_0) = \left[\Psi_{n,i}(f_0) n^{-1/2} \kappa_n^{d/2}\right]^p \kappa_n^{-d}$, $\Phi^{(p)}(f_0) = \mu(A)^{-1} \int_A f_0^{p/2}$ and $\gamma_n = n^{1/2}\kappa_n^{-d/2} = n^{s/(2s+d)}$. The same kind of results have been proved by Cencov (1972), Samarov (1976), Bretagnolle and Huber (1979), Efroimovich and Pinsker (1982).

6.3.2 Spherical case, $d = 2$, $s \geq 2$

Let $\mathcal{X} = S_d$, $d = 2$, let μ be the Lebesgue measure of S_d and $\mathcal{F}$ the Borel sigma–algebra of S_d. Estimation of a density $S_d \to \mathbf{R}$ will be discussed. In this section the result is for $d = 2$ but for a general, even smoothness index $s \geq 2$. In the next section, general dimension $d \geq 2$ will be considered but only for the smoothness index $s = 2$.

Let $f_0 : S_d \to \mathbf{R}$ be a density function. Let $0 < b < 1$, $B = [-b,b]^d$, $A = \{x = (x_1,\ldots,x_{d+1})' \in S_d \,|\, (x_1,\ldots,x_d)' \in B,\, x_{d+1} > 0\}$, and $f_0(x) = M > 0$ for $x \in A$. Let also $\int_{S_d} |\Delta^{s/2} f_0|^p d\mu < \infty$. Define

$$\mathbf{F}(s,a,p,f_0) = \left\{ f : S_d \to \mathbf{R} \;\middle|\; f \text{ density}, \int_{S_d} |\Delta^{s/2} f - \Delta^{s/2} f_0|^p d\mu \leq a \right\},$$

where $s \geq 2$ is even, $0 < a < \infty$, and $p \geq 1$.

Let $\mathcal{Y} = [0,\infty[$, let ν be the Lebesgue measure of $\mathcal{Y}$ and $\mathcal{G}$ its Borel sigma–algebra. Let $\mathbf{H}_\infty$ be the set of measurable functions $h : \mathcal{Y} \to \mathbf{R}$ for which h has the support on $[0, b^2/2]$, $\int_0^\infty h d\nu = 0$, $\sup_{y\in\mathcal{Y}} |h(y)| < \infty$, $\int_0^\infty |h^{(i)}|^p d\nu < \infty$ for $i = 1, \ldots, s/2-1$, and

$$s^{p-1}(2\pi)^{1-p/2}M^{p/2}\sum_{i=s/2}^{s}\int_0^\infty (2t)^{i-s/2}\left|Q_{s,i}(1)h^{(i)}(t)\right|^p dt \le a,$$

where the polynomials $Q_{s,i}$ were defined in Lemma 5.8.

Theorem 6.8 *Let* $F_n = \left(\mathcal{X}^n, \mathcal{F}^n, P_f^n : f \in \mathbf{F}(s,a,p,f_0)\right)$, $\mathbf{F} = L_p(\mathcal{X}, \mathcal{F}, \mu)$, $\mathbf{G} = L_p(\mathcal{Y}, \mathcal{G}, \nu)$, $\tilde{W}_f(a) = \int_{S_d} |f - a|^p d\mu$, $W_h(a) = \int_{\mathcal{Y}} |h - a|^p d\nu$, *and* $\Phi^{(p)}(f_0) = (2\pi)^{1-p/2}M^{p/2}$. *Then,*

$$\liminf_{n\to\infty} n^{ps/(2s+2)} \inf_{\beta_n\in\mathcal{B}(F_n,\mathbf{F})} \sup_{f\in\mathbf{F}(s,a,p,f_0)} \beta_n\left(\tilde{W}_f, P_f^n\right)$$

$$\ge \Phi^{(p)}(f_0) \inf_{\beta\in\mathcal{B}(E(\mathbf{H}_\infty),\mathbf{G})} \sup_{h\in\mathbf{H}_\infty} \beta\left(W_h, G_h\right).$$

Proof. Let $\{\kappa_n\}$ be a sequence with $\lim_{n\to\infty}\kappa_n = \infty$ and $\lim_{n\to\infty} n^{-1}\kappa_n^d = 0$. Let I_n have cardinality $\sim \kappa_n^d$ and let $x_{n,i} \in A$, $i \in I_n$, be such that the sets $A_{n,i} = \{x \in S_d \,|\, \kappa_n^2(1 - x'x_{n,i}) \le b^2/2\}$ are contained in A and almost surely disjoint. For example, let $y_{n,i} = \left(y_1^{(n,i_1)}, \ldots, y_d^{(n,i_d)}\right)' \in \mathbf{R}^d$, where $i = (i_1, \ldots, i_d)' \in I_n = \{-[(\kappa_n - 1)/2], \ldots, [(\kappa_n - 1)/2]\}^d$, $[a]$ is the greatest integer $\le a$, and $y_k^{(n,i_k)} = i_k 2b/\kappa_n$, $k \in \{1, \ldots, d\}$. Let $x_{n,i} = (y_{n,i}', 1 - \|y_{n,i}\|^2)'$. Now the sets $A_{n,i}$ are almost everywhere disjoint. This is seen in the following way. Let $B_{n,i} = y_{n,i} + \kappa_n^{-1}B$. Because the sets $B_{n,i}$ are almost everywhere (ν) disjoint, the sets $A'_{n,i} = \{x \in S_d \,|\, (x_1, \ldots, x_d)' \in B_{n,i},\, x_{d+1} > 0\}$ are almost everywhere (μ) disjoint. Because $\kappa_n^2(1 - x'x_{n,i}) = \kappa_n^2\|x - x_{n,i}\|^2/2 \le b^2/2$ implies that $\max_{j=1,\ldots,d}|x_j - x_j^{(n,i_j)}| \le \kappa_n^{-1}b$, it is seen that $A_{n,i} \subset A'_{n,i}$.

For $h : \mathcal{Y} \to \mathbf{R}$, define

$$(T_{n,i}h)(x) = \Psi(f_0)n^{-1/2}\kappa_n^{d/2}h(\kappa_n^2(1 - x'x_{n,i}))I_{A_{n,i}}(x),$$

where $i \in I_n$ and $\Psi(f_0) = (\omega_{d-1}2^{(d-2)/2})^{-1/2}\sqrt{f_0(x_{n,i})} = \omega_{d-1}^{-1/2}M^{1/2}$. Let $\mathbf{H}_c$, $0 < c < \infty$, be the set of those functions in $\mathbf{H}_\infty$ for which $\sup_{y\in\mathcal{Y}}|h(y)| \le c$, $\int_0^\infty |h^{(i)}|^p d\nu \le c$ for $i = 1, \ldots, s/2-1$, and

$$(1 + c^{-1})s^{p-1}\omega_{d-1}^{1-p/2}M^{p/2}\sum_{i=s/2}^{s}\int_0^\infty (2t)^{i-s/2}\left|Q_{s,i}(1)h^{(i)}(t)\right|^p dt \le a.$$

It is seen that for every sequence $k(n) \in I_n$, $T_n = T_{n,k(n)}$, f_0 and $\mathbf{H}_c$ are such that conditions of Lemma 6.2 are satisfied and thus $E_n(\mathbf{H}_c) \xrightarrow{w} E(\mathbf{H}_c)$. Indeed, because $d = 2$ and by the substitution $\kappa_n^2(1-t) = u$,

$$\begin{aligned}
\int_{S_d} \left(T_{n,k(n)}h\right) d\mu &= \Psi(f_0)n^{-1/2}\kappa_n^{d/2} \int_{A_{n,k(n)}} h(\kappa_n^2(1 - x'x_{n,k(n)}))d\mu(x) \\
&= \Psi(f_0)n^{-1/2}\kappa_n^{d/2}\omega_{d-1} \int_{1-\kappa_n^{-2}b^2/2}^{1} h(\kappa_n^2(1-t))(1-t^2)^{(d-2)/2}dt \\
&= \Psi(f_0)n^{-1/2}\kappa_n^{-d/2}\omega_{d-1} \int_0^{b^2/2} h(u)du = 0.
\end{aligned}$$

Also,

$$\begin{aligned}
& n \int_{S_d} \left(T_{n,k(n)}h_1\right)\left(T_{n,k(n)}h_2\right) f_0^{-1} d\mu \\
&= \Psi(f_0)^2 \int_0^{b^2/2} dt h_1(t)h_2(t) \int_{T_{x_{n,k(n)}}} f_0\left(\phi^{-1}_{x_{n,k(n)}}(\xi, 1-\kappa_n^{-2}t)\right)^{-1} d\mu_{d-1}(\xi) \\
&= \int_{\mathcal{Y}} h_1(t)h_2(t)dt + o(1).
\end{aligned}$$

Secondly, it is noted that by choosing $\kappa_n = n^{1/(2s+d)}$ it holds that $\{f_{\underline{h}_n,n} \,|\, \underline{h}_n \in \mathbf{H}_c^{I_n}\} \subset \mathbf{F}(s,a,p,f_0)$ ultimately. In fact, now $n^{-1/2}\kappa_n^{d/2} = \kappa_n^{-s}$ and for $\underline{h}_n \in \mathbf{H}_c^{I_n}$,

$$\begin{aligned}
& \int_{S_d} |\Delta^{s/2} f_{\underline{h}_n,n} - \Delta^{s/2} f_0|^p d\mu \\
&= \int_{S_d} \left|\sum_{i\in I_n} \Delta^{s/2}(T_{n,i}h_i)\right|^p d\mu = \sum_{i\in I_n} \int_{A_{n,i}} \left|\Delta^{s/2}(T_{n,i}h_i)\right|^p d\mu \\
&= \sum_{i\in I_n} \int_{A_{n,i}} \left|\Psi(f_0)\kappa_n^{-s} h_i^{(s,\kappa_n)}(x'x_{n,i})\right|^p d\mu(x) \\
&= \omega_{d-1}\kappa_n^{-d} \sum_{i\in I_n} \int_0^{b^2/2} \left|\Psi(f_0)\kappa_n^{-s} h_i^{(s,\kappa_n)}(1-\kappa_n^{-2}u)\right|^p du \le a,
\end{aligned}$$

for sufficiently large n, where

$$\begin{aligned}
h_i^{(s,\kappa_n)}(1-\kappa_n^{-2}u) &= \kappa_n^s \Bigg[\sum_{j=1}^{s/2-1} P_{s,j}(1-\kappa_n^{-2}u)(-1)^j \kappa_n^{2j-s} h_i^{(j)}(u) \\
&\quad + \sum_{j=s/2}^{s} u^{j-s/2}(2-\kappa_n^{-2}u)^{j-s/2} Q_{s,j}(1-\kappa_n^{-2}u)(-1)^j h_i^{(j)}(u)\Bigg]
\end{aligned}$$

by Lemma 5.8.

Thirdly, (6.6) is satisfied because

$$\begin{aligned}
&\int_{A_{n,i}} |T_{n,i}h - f|^p d\mu \\
&= \left(\Psi(f_0)\kappa_n^{-s}\right)^p \int_{A_{n,i}} \left|h(\kappa_n^2(1 - x'x_{n,i})) - \Psi(f_0)^{-1}\kappa_n^s f(x)\right|^p d\mu(x) \\
&= \left(\Psi(f_0)\kappa_n^{-s}\right)^p \kappa_n^{-d} \\
&\quad \times \int_{T_{x_{n,i}}} d\mu_{d-1}(\xi) \int_0^{b^2/2} \left|h(u) - \Psi(f_0)^{-1}\kappa_n^s f\left(\phi_{x_{n,i}}^{-1}(\xi, 1 - \kappa_n^{-2}u)\right)\right|^p du \\
&\geq \left(\Psi(f_0)\kappa_n^{-s}\right)^p \kappa_n^{-d}\omega_{d-1} \\
&\quad \times \int_0^{b^2/2} \left|h(u) - \omega_{d-1}^{-1}\Psi(f_0)^{-1}\kappa_n^s \int_{T_{x_{n,i}}} f\left(\phi_{x_{n,i}}^{-1}(\xi, 1 - \kappa_n^{-2}u)\right) d\mu_{d-1}(\xi)\right|^p du.
\end{aligned}$$

Lemma 6.6 can be applied with the identifications $\Phi_{n,i}^{(p)}(f_0) = \left(\Psi(f_0)\kappa_n^{-s}\right)^p \kappa_n^{-d}\omega_{d-1}$, $\Phi^{(p)}(f_0) = \omega_{d-1}^{1-p/2} M^{p/2}$ and $\gamma_n = \kappa_n^s = n^{s/(2s+d)}$. Transition from $\mathbf{H}_c$ to $\mathbf{H}_\infty$ can be made along the same lines as in the proof of Theorem 6.5. □

6.3.3 Spherical case, $d \geq 2$, $s = 2$

Let $\mathcal{X} = S_d$, let μ be the Lebesgue measure of S_d and $\mathcal{F}$ the Borel sigma–algebra of S_d. Estimation of a density $S_d \to \mathbf{R}$ will be discussed.

Let $f_0 : S_d \to \mathbf{R}$ be a density function, $f_0(x) = M > 0$ for $x \in A$, where $A = \{x = (x_1, \ldots, x_{d+1})' \in S_d \,|\, (x_1, \ldots, x_d)' \in B,\, x_{d+1} > 0\}$, $B = [-b, b]^d$, and $0 < b < 1$. Note that A is the same set as in the previous section. Let also $\int_{S_d} |\Delta f_0|^p d\mu < \infty$. Set

$$\mathbf{F}(a, p, f_0) = \left\{ f : S_d \to \mathbf{R} \,\middle|\, f \text{ density}, \int_{S_d} |\Delta f - \Delta f_0|^p d\mu \leq a \right\},$$

where $0 < a < \infty$ and $p \geq 1$.

Let $\mathcal{Y} = \mathbf{R}^d$, let ν be the Lebesgue measure of $\mathbf{R}^d$ and $\mathcal{G}$ its Borel sigma–algebra. Let $\mathbf{H}_\infty$ be the set of measurable functions $h : \mathcal{Y} \to \mathbf{R}$ for which h has the support on $[-b/2, b/2]^d$, $h = h^{<1>}h^{<2>}$, $h^{<1>} : [-b/2, b/2] \to \mathbf{R}$, $h^{<2>} : [-b/2, b/2]^{d-1} \to \mathbf{R}$, $\int_{-b/2}^{b/2} h^{<1>} = 0$, $\int_{[-b/2,b/2]^{d-1}} |h^{<2>}| < \infty$, $\sup_{y \in \mathcal{Y}} |h(y)| < \infty$,

$\sum_{i=1}^{d}\int_{\mathcal{Y}}|\partial/\partial\theta_i h(\theta)|^p\,d\theta<\infty$, and

$$(2d)^{p-1}M^{p/2}\sum_{i=1}^{d}\int_{\mathcal{Y}}\left|\frac{\partial^2}{\partial\theta_i^2}h(\theta)\right|^p d\theta\le a.$$

Theorem 6.9 *Let* $F_n=\left(\mathcal{X}^n,\mathcal{F}^n,P_f^n : f\in\mathbf{F}(a,p,f_0)\right)$, $\mathbf{F}=L_p(\mathcal{X},\mathcal{F},\mu)$, $\mathbf{G}=L_p(\mathcal{Y},\mathcal{G},\nu)$, $\tilde{W}_f(a)=\int_{S_d}|f-a|^p d\mu$, *and* $W_h(a)=\int_{\mathcal{Y}}|h-a|^p d\nu$. *Then*

$$\liminf_{n\to\infty} n^{2p/(4+d)}\inf_{\beta_n\in\mathcal{B}(F_n,\mathbf{F})}\sup_{f\in\mathbf{F}(a,p,f_0)}\beta_n\left(\tilde{W}_f,P_f^n\right)\ge M^{p/2}\inf_{\beta\in\mathcal{B}(E(\mathbf{H}_\infty),\mathbf{G})}\sup_{h\in\mathbf{H}_\infty}\beta\left(W_h,G_h\right).$$

Proof. Let $\phi : S_d\to[-\pi,\pi]\times[-\pi/2,\pi/2]^{d-1}$ be the parameterization of S_d obtained from the parameterization (2.3) by a translation. Let $\{\kappa_n\}$ be a sequence with $\lim_{n\to\infty}\kappa_n=\infty$ and $\lim_{n\to\infty}n^{-1}\kappa_n^d=0$. Let I_n have cardinality $\sim\kappa_n^d$, and let $x_{n,i}\in A$, $i\in I_n$, be such that the sets $A_{n,i}=\{x\in S_d\,|\,\kappa_n\phi^{(n,i)}(x)\in[-b/2,b/2]^d\}$, $i\in I_n$, are contained in A and almost surely disjoint, where $\phi^{(n,i)} : S_d\to[-\pi,\pi]\times[-\pi/2,\pi/2]^{d-1}$, $\phi^{(n,i)}(x)=\phi(R_{x_{n,i},x_0}x)$, $x_0=\phi^{-1}(0)$ and $R_{x_{n,i},x_0}$ is such a rotation that $R_{x_{n,i},x_0}x_{n,i}=x_0$. Note that $\phi^{(n,i)}(x_{n,i})=0$ and $\phi^{(n,i)}(A_{n,i})=\kappa_n^{-1}[-b/2,b/2]^d$. The vectors $x_{n,i}$ can be chosen in the same way as in the previous section, because $A_{n,i}\subset\{x\in S_d\,|\,\kappa_n^2(1-x'x_{n,i})\le b^2/2\}$.

For $h:\mathcal{Y}\to\mathbf{R}$, define

$$(T_{n,i}h)(x)=\Psi(f_0)n^{-1/2}\kappa_n^{d/2}h(\kappa_n\phi^{(n,i)}(x))I_{A_{n,i}}(x),$$

where $i\in I_n$ and $\Psi(f_0)=\sqrt{f_0(x_{n,i})}=M^{1/2}$. Let $\mathbf{H}_c$, $0<c<\infty$, be the set of those functions in $\mathbf{H}_\infty$ for which $\sup_{y\in\mathcal{Y}}|h(y)|\le c$, $\sum_{i=1}^{d}\int_{\mathcal{Y}}|\partial/\partial\theta_i h(\theta)|^p\,d\theta\le c$, and

$$(1+c^{-1})(2d)^{p-1}M^{p/2}\sum_{i=1}^{d}\int_{\mathcal{Y}}\left|\frac{\partial^2}{\partial\theta_i^2}h(\theta)\right|^p d\theta\le a.$$

It is seen that for every sequence $k(n)\in I_n$, $T_n=T_{n,k(n)}$, f_0 and $\mathbf{H}_c$ are such that conditions of Lemma 6.2 are satisfied, and thus $E_n(\mathbf{H}_c)\xrightarrow{w}E(\mathbf{H}_c)$. Indeed,

$$\begin{aligned}&\int_{S_d}\left(T_{n,k(n)}h\right)d\mu\\ &=\ \Psi(f_0)n^{-1/2}\kappa_n^{d/2}\int_{A_{n,k(n)}}h\left(\kappa_n\phi^{(n,k(n))}(x)\right)d\mu(x)\end{aligned}$$

$$
\begin{aligned}
&= \Psi(f_0) n^{-1/2} \kappa_n^{d/2} \int_{\phi^{-1}(\kappa_n^{-1}[-b/2,b/2]^d)} h(\kappa_n \phi(x)) d\mu(x) \\
&= \Psi(f_0) n^{-1/2} \kappa_n^{d/2} \int_{\kappa_n^{-1}[-b/2,b/2]^d} h(\kappa_n \theta) \sin^{d-1}(\theta_d + \pi/2) \cdots \sin(\theta_2 + \pi/2) d\theta \\
&= \Psi(f_0) n^{-1/2} \kappa_n^{-d/2} \int_{[-b/2,b/2]^d} h(\theta) \sin^{d-1}(\kappa_n^{-1}\theta_d + \pi/2) \cdots \sin(\kappa_n^{-1}\theta_2 + \pi/2) d\theta \\
&= \Psi(f_0) n^{-1/2} \kappa_n^{-d/2} \int_{-b/2}^{b/2} h^{<1>}(\theta_1) d\theta_1 \\
&\quad \times \int_{[-b/2,b/2]^{d-1}} h^{<2>}(\theta_2 \ldots, \theta_d) \sin^{d-1}(\kappa_n^{-1}\theta_d + \pi/2) \cdots \sin(\kappa_n^{-1}\theta_2 + \pi/2) d\theta_2 \cdots d\theta_d \\
&= 0.
\end{aligned}
$$

Also,

$$
\begin{aligned}
& n \int_{S_d} \left(T_{n,k(n)} h_1\right) \left(T_{n,k(n)} h_2\right) f_0^{-1} d\mu \\
&\quad = \Psi(f_0)^2 \int_{[-b/2,b/2]^d} h_1(\theta) h_2(\theta) f_0 \left(R_{x_0, x_{n,k(n)}} \phi^{-1}(\kappa_n^{-1}\theta)\right)^{-1} \\
&\quad\quad \times \sin^{d-1}(\kappa_n^{-1}\theta_d + \pi/2) \cdots \sin(\kappa_n^{-1}\theta_2 + \pi/2) d\theta \\
&\quad = \int_{[-b/2,b/2]^d} h_1 h_2 d\nu + o(1).
\end{aligned}
$$

Secondly, we have by choosing $\kappa_n = n^{1/(4+d)}$ that $\{f_{\underline{h}_n, n} \,|\, \underline{h}_n \in \mathbf{H}_c^{I_n}\} \subset \mathbf{F}(a, p, f_0)$ ultimately. In fact, now $n^{-1/2}\kappa_n^{d/2} = \kappa_n^{-2}$, and for $\underline{h}_n \in \mathbf{H}_c^{I_n}$,

$$
\begin{aligned}
& \int_{S_d} |\Delta f_{\underline{h}_n, n} - \Delta f_0|^p d\mu \\
&= \int_{S_d} \left| \sum_{i \in I_n} \Delta(T_{n,i} h_i) \right|^p d\mu = \sum_{i \in I_n} \int_{A_{n,i}} |\Delta(T_{n,i} h_i)|^p \, d\mu \\
&= \kappa_n^{-d} \sum_{i \in I_n} |\Psi(f_0)|^p \int_{[-b/2,b/2]^d} \left| \sum_{j=1}^d \eta_j(\kappa_n^{-1}\theta) \left[\left(\frac{\partial^2}{\partial \theta_j^2} h_i \right)(\theta) + \psi_j(\kappa_n^{-1}\theta) \kappa_n^{-1} \left(\frac{\partial}{\partial \theta_j} h_i \right)(\theta) \right] \right|^p \\
&\quad \times \sin^{d-1}(\kappa_n^{-1}\theta_d + \pi/2) \cdots \sin(\kappa_n^{-1}\theta_2 + \pi/2) d\theta \\
&\le a,
\end{aligned}
$$

for sufficiently large n, where $\eta_j(\theta) = \sin^{-2}(\theta_{j+1} + \pi/2) \cdots \sin^{-2}(\theta_d + \pi/2)$ and $\psi_j(\theta) = (j-1)\cos(\theta_j + \pi/2)/\sin(\theta_j + \pi/2)$, for $j = 1, \ldots, d$ ($\eta_d(\theta) = 1$).

Thirdly, it is noted that (6.6) is satisfied because

$$\begin{aligned}
&\int_{A_{n,i}} |T_{n,i}h - f|^p d\mu \\
&= \left(\Psi(f_0)\kappa_n^{-2}\right)^p \int_{A_{n,i}} \left|h(\kappa_n \phi^{(n,i)}(x)) - \Psi(f_0)^{-1}\kappa_n^2 f(x)\right|^p d\mu(x) \\
&= \left(\Psi(f_0)\kappa_n^{-2}\right)^p \kappa_n^{-d} \int_{[-b/2,b/2]^d} \left|h(\theta) - \Psi(f_0)^{-1}\kappa_n^2 f\left(R_{x0,x_{n,i}}\phi^{-1}(\kappa_n^{-1}\theta)\right)\right|^p \\
&\quad \times \sin^{d-1}(\kappa_n^{-1}\theta_d + \pi/2)\cdots\sin(\kappa_n^{-1}\theta_2 + \pi/2)d\theta \\
&\geq \delta_n \left(\Psi(f_0)\kappa_n^{-2}\right)^p \kappa_n^{-d} \int_{[-b/2,b/2]^d} \left|h(\theta) - \Psi(f_0)^{-1}\kappa_n^2 f\left(R_{x0,x_{n,i}}\phi^{-1}(\kappa_n^{-1}\theta)\right)\right|^p d\theta,
\end{aligned}$$

where $\delta_n = \inf_{\theta\in[-b/2,b/2]^d}\left[\sin^{d-1}(\kappa_n^{-1}\theta_d + \pi/2)\cdots\sin(\kappa_n^{-1}\theta_2 + \pi/2)\right]$ so that $\lim_{n\to\infty}\delta_n = 1$. Lemma 6.6 can be applied with the identifications $\Phi^{(p)}_{n,i}(f_0) = \delta_n\left(\Psi(f_0)\kappa_n^{-2}\right)^p \kappa_n^{-d}$, $\Phi^{(p)}(f_0) = M^{p/2}$ and $\gamma_n = \kappa_n^2 = n^{2/(4+d)}$. Transition from $\mathbf{H}_c$ to $\mathbf{H}_\infty$ can be made similarly as in the proof of Theorem 6.5. □

6.4 Integral Functionals

Let us prove that the constant $C_{q,r}(f)$ given in Theorem 5.13 is the best possible in the local asymptotic minimax sense. General infinite dimensional local asymptotic minimax results that could be applied here are given by Koshevnik and Levit (1976) and by Ibragimov and Khas'minskii (1991). However, an argument will be used which is based explicitly on the convergence of experiments. Let $f_0 : S_d \to \mathbf{R}$ be such a density that $C_{q,r}(f_0) < \infty$, $\|f_0\|_\infty < \infty$, $\|\Delta^{(q+r)/2}f_0\|_\infty < \infty$, and $\|\Delta^{(q+r)/2}(f_0\Delta^{(q+r)/2}f_0)\|_1 < \infty$. Let

$$\mathbf{H}_c = \left\{h : S_d \to \mathbf{R} \,\middle|\, \int_{S_d} h^2 d\mu < \infty,\ \int_{S_d} h\sqrt{f_0}d\mu = 0,\ \|h\|_\infty < c,\ |\theta_{q,r}(h\sqrt{f_0})| < c\right\},$$

where $0 < c \leq \infty$. Denote

$$f_h(\cdot,t) = f_0 + t\sqrt{f_0}\,h, \quad f_{h,n} = f_h(\cdot, n^{-1/2}).$$

Let $P_{h,n}$ be the probability measure corresponding to $f_{h,n}$, and $P^n_{h,n}$ the corresponding product measure on $(S_d^n, \mathcal{B}(S_d)^n)$, where $\mathcal{B}(S_d)$ is the Borel sigma–algebra of S_d. Denote

$$E_n(\mathbf{H}_c) = \left(S_d^n, \mathcal{B}(S_d)^n, P^n_{h,n} : h \in \mathbf{H}_c\right).$$

Theorem 6.10 *Let $q, r \geq 0$ be even. Let $\theta_{q,r}(f) = \int_{S_d} (\Delta^{q/2} f)(\Delta^{r/2} f) d\mu$. Then,*

$$\liminf_{n\to\infty} n \inf_{\rho_n \in \mathcal{R}(E_n(\mathbf{H}_\infty),\mathbf{R})} \sup_{h\in\mathbf{H}_\infty} \int_{S_d^n} dP_{h,n}^n\left(x^{(n)}\right) \int_{\mathbf{R}} \left(\theta_{q,r}(f_{h,n}) - a\right)^2 \rho_n\left(x^{(n)}, da\right)$$
$$\geq C_{q,r}(f_0) = 4\mathrm{Var}\left(\Delta^{(q+r)/2} f_0(X_1)\right).$$

Proof. It holds that

$$\lim_{t\to 0} \sup_{h\in\mathbf{H}_c} \left| t^{-1}\left(\theta_{q,r}(f_h(\cdot,t)) - \theta_{q,r}(f_0)\right) - < \dot{\theta}_{q,r}(f_0), h >_{\mathbf{H}} \right| = 0, \tag{6.7}$$

where $0 < c < \infty$, and $\dot{\theta}_{q,r}(f_0) : S_d \to \mathbf{R}$ is defined by

$$\dot{\theta}_{q,r}(f_0) = 2\left(\Delta^{(q+r)/2} f_0 - \theta_{q,r}(f_0)\right)\sqrt{f_0}.$$

By Lemma 6.2, the experiment $E_n(\mathbf{H}_c)$ converges weakly to a Gaussian shift experiment $E(\mathbf{H}_c) = (\Omega, \mathcal{A}, G_h : h \in \mathbf{H}_c)$, when $0 < c < \infty$. Because the decision space $\mathbf{R}$ is a Polish space and the experiments are dominated, the sets of decision procedures coincide with the sets of generalized decision procedures. The sets $\mathcal{R}(E_n(\mathbf{H}_c), \mathbf{R})$ are the same for all $c \in]0, \infty]$. The same holds for the sets $\mathcal{R}(E(\mathbf{H}_c), \mathbf{R})$. Thus, we can denote $\mathcal{R}(E_n(\mathbf{H}_c), \mathbf{R}) = \mathcal{R}(E_n, \mathbf{R})$ and $\mathcal{R}(E(\mathbf{H}_c), \mathbf{R}) = \mathcal{R}(E, \mathbf{R})$. Then, by equation (6.7) and Lemma 6.1,

$$\begin{aligned}
&\liminf_{n\to\infty} n \inf_{\rho_n \in \mathcal{R}(E_n,\mathbf{R})} \sup_{h\in\mathbf{H}_\infty} \int_{S_d^n} dP_{h,n}^n\left(x^{(n)}\right) \int_{\mathbf{R}} \left(\theta_{q,r}(f_{h,n}) - a\right)^2 \rho_n\left(x^{(n)}, da\right) \\
&\geq \liminf_{n\to\infty} \inf_{\rho_n \in \mathcal{R}(E_n,\mathbf{R})} \sup_{h\in\mathbf{H}_c} \int_{S_d^n} dP_{h,n}^n\left(x^{(n)}\right) \\
&\qquad \int_{\mathbf{R}} \left(< \dot{\theta}_{q,r}(f_0), h >_{\mathbf{H}} - \sqrt{n}\left(a - \theta_{q,r}(f_0)\right) + o(1)\right)^2 \rho_n\left(x^{(n)}, da\right) \\
&\geq \liminf_{n\to\infty} \inf_{\rho_n \in \mathcal{R}(E_n,\mathbf{R})} \sup_{h\in\mathbf{H}_c} \int_{S_d^n} dP_{h,n}^n\left(x^{(n)}\right) \int_{\mathbf{R}} \left(< \dot{\theta}_{q,r}(f_0), h >_{\mathbf{H}} - a\right)^2 \rho_n\left(x^{(n)}, da\right) \\
&\geq \inf_{\rho \in \mathcal{R}(E,\mathbf{R})} \sup_{h\in\mathbf{H}_c} \int_{\Omega} dG_h(\omega) \int_{\mathbf{R}} \left(< \dot{\theta}_{q,r}(f_0), h >_{\mathbf{H}} - a\right)^2 \rho(\omega, da) \\
&= \sup_{\pi\in\Pi_0(\mathbf{H}_c)} \inf_{\rho \in \mathcal{R}(E,\mathbf{R})} \int_{\mathbf{H}_c} d\pi(h) \int_{\Omega} dG_h(\omega) \int_{\mathbf{R}} \left(< \dot{\theta}_{q,r}(f_0), h >_{\mathbf{H}} - a\right)^2 \rho(\omega, da),
\end{aligned}$$

where $\Pi_0(\mathbf{H}_c)$ is the set of probability measures with finite support on $\mathbf{H}_c$. The last identity is proved in Strasser (1985, Theorem 46.3). Because c was chosen arbitrarily, we have the lower bound

$$\begin{aligned}
&\sup_{\pi\in\Pi_0(\mathbf{H}_\infty)} \inf_{\rho\in\mathcal{R}(E,\mathbf{R})} \int_{\mathbf{H}_\infty} d\pi(h) \int_\Omega dG_h(\omega) \int_{\mathbf{R}} \left(< \dot{\theta}_{q,r}(f_0), h >_{\mathbf{H}} - a\right)^2 \rho(\omega, da) \\
&= \inf_{\rho\in\mathcal{R}(E,\mathbf{R})} \sup_{h\in\mathbf{H}_\infty} \int_\Omega dG_h(\omega) \int_{\mathbf{R}} \left(< \dot{\theta}_{q,r}(f_0), h >_{\mathbf{H}} - a\right)^2 \rho(\omega, da) \\
&\geq \left\|\dot{\theta}_{q,r}(f_0)\right\|_{\mathbf{H}}^2 = C_{q,r}(f_0),
\end{aligned}$$

where the lower bound (6.4) was applied to $T(h) =< \dot{\theta}_{q,r}(f_0), h >_{\mathbf{H}}$ and the facts

$$\omega(\epsilon) = \sup\left\{ \left|< \dot{\theta}_{q,r}(f_0), h >_{\mathbf{H}}\right| \;\middle|\; h \in \mathbf{H}_\infty,\ \|h\|_{\mathbf{H}} \leq \epsilon \right\} = \epsilon \|\dot{\theta}_{q,r}(f_0)\|_{\mathbf{H}}$$

(by the fact that $\dot{\theta}_{q,r}(f_0) \in \mathbf{H}_\infty$) and $\xi(1) = 1$ were used. □

Chapter 7

Summary and Discussion

A large part of this thesis is concerned with the study of approximation by convolution. This is due to the fact that the expectation of the kernel estimator is a convolution. The basic result is the expansion of convolution, given in Lemma 2.9. The definition of the associated kernel in the Euclidean case is modified to get the definition of an associated delta sequence in the spherical case. The remainder term of the expansion of the convolution can be written with the help of the associated delta sequence. With this convenient form for the remainder term in hand different modes of convergence (pointwise convergence and L_p convergence) can be studied.

In previous studies, only a second order expansion has been given for the expectation of the kernel estimator, but in this study also a higher order expansion is given. Here a new phenomenon arises. There does not exist such kernel functions that the lower order terms of the expansion would vanish. The lower order terms can, however, be made to approach zero at a sufficiently fast rate of convergence. 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. The smoothness conditions assumed here are not so natural as those assumed in the Euclidean case (see Holmström and Klemelä 1992). It seems that further work might lead to more elegant smoothness conditions. Also, the interesting method using a mollifier to relax smoothness conditions, as was done in Devroye and Györfi (1985, page 77) in the case of the real line, has not been considered here.

In Lemma 2.1 it is proved that the concept of second derivative for functions

$S_d \to \mathbf{R}$, defined by Hall, Watson and Cabrera (1987), is proportional to the Laplace operator.

Hall, Watson and Cabrera (1987) gave asymptotic results for the mean integrated squared error assuming smoothness index $s = 2$. In this study a general smoothness index is assumed and results are given also for the mean squared error and the mean integrated absolute error.

With spherical data, selection of the smoothing parameter of the kernel estimator with the plug–in method has not been studied previously. In this study it is proved that the smoothing parameter can be chosen in such a way that the asymptotic risk is the same as when using the asymptotically optimal deterministic smoothing parameter. The results differ from the results in the case of the real line by Woodroofe (1970) and Nadaraya (1974) in that they considered only specific initial estimators (kernel estimators and derivatives of kernel estimators). Here sufficient conditions are formulated for any estimators, to be used in the plug–in method, to satisfy.

The estimation of derivatives, Laplacians, and their inner products has been studied previously only in the Euclidean case. The estimation of the Laplacian at a point and the integrated squared Laplacian is especially useful. Two different estimators are defined for the Laplacian at a point. For both estimators the asymptotic mean squared error is calculated. However, any comparison of the asymptotic mean squared errors is not made. An estimator is defined for the integral of the squared Laplacian using the ideas of Bickel and Ritov (1988). It is proved that this estimator has the $\sqrt{n}$–rate of convergence, if the density is sufficiently smooth. Furthermore, it is proved that this estimator is optimal in the local asymptotic minimax sense. Bickel and Ritov (1988) gave the rates of convergence also when the density is not so smooth that there could exist an estimator having the $\sqrt{n}$–rate of convergence, but this situation is not considered here.

In the Euclidean case there are well–known lower bound results for density estimation, giving the optimal rate of convergence as $n^{s/(2s+d)}$, where s is the smoothness index of the density and d is the dimension of the sample space. These results are now shown to extend to the spherical case. Estimating the density at a point, estimating an iterated Laplacian of the density at a point, and estimating the density with the L_p error are considered. General lower bounds are derived so that the Euclidean and the spherical cases are given a unified treatment. The lower bound given for the L_p error is not known in the Euclidean case, either. No minimax upper bounds corresponding to these lower bounds are given. In the Euclidean case the exact asymptotic minimax risk has been calculated by Efraimovich and Pinsker (1982), subject to an L_2 loss function. It is possible that their method could be adapted to the spherical case to give the exact asymptotic minimax risk for the L_2 error. These questions remain topics for future research.

Bibliography

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

[7] Bickel, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10** 647-671.

[8] Bickel, P. J. and Ritov, Y. (1988). Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhyā Ser. A* **50** 381-393.

[9] Birgé, L. and Massart, P. (1995). Estimation of integral functionals of a density. *Ann. Statist.* **23** 11-29.

[10] Bochner, S. (1955). *Harmonic Analysis and the Theory of Probability.* Univ. of California Press.

[11] Bretagnolle, J., and Huber, C. (1979). Estimation des densités: Risque minimax. *Z. Wahrsch. Verw. Gebiete* **47** 119-137.

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

[13] Cacoullos, T. (1966). Estimation of a multivariate density. *Ann. Inst. Statist. Math.* **18** 179-189.

[14] 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.

[15] Cencov, N. N. (1972). *Statistical Decision Rules and Optimal Inference* Nauka, Moscow.

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

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

[18] Donoho, D. L. and Liu, R. C. (1991). Geometrizing rates of convergence III. *Ann. Statist.* **19** 668-701.

[19] Donoho, D. L., Liu, R. C., and MacGibbon, B. (1990). Minimax risk over hyperrectangles, and implications. *Ann. Statist.* **18** 1416-1437.

[20] Efroimovich, S. Y. and Pinsker, M. S. (1982). Estimation of square integrable probability density of a random variable. *Problems Inform. Transmission* **18** 175-189.

[21] 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.

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

[23] Farrell, R.H. (1972). On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. *Ann. Statist.* **43** 170-180.

[24] Fisher, R. A. (1953). Dispersion on a sphere. *Proc. Roy. Soc. Lond. A* **217** 295-305.

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

[26] Fix, E. and Hodges, J. L. (1951). Discriminatory analysis – nonparametric discrimination: consistency properties. *Report No. 4, Project no. 21-29-004*, USAF School of Aviation Medicine, Randolph Field, Texas.

[27] Fix, E. and Hodges, J. L. (1989). Discriminatory analysis – nonparametric discrimination: consistency properties. *Internat. Statist. Rev.* **57** 238-247.

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

[29] Golubev, G. K. (1991). LAN in problems of nonparametric estimation of functions and lower bounds for quadratic risks. *Theory Probab. Appl.* **36** 152-157.

[30] Golubev, G. K. and Nussbaum, M. (1990). A risk bound in Sobolev class regression. *Ann. Statist.* **18** 758-778.

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

[32] Hall, P. (1993). On plug-in rules for local smoothing of density estimators. *Ann. Statist.* **21** 694-710.

[33] Hall, P. and Marron, J. S. (1987). Estimation of integrated squared density derivatives. *Statist. Probab. Lett.* **6** 109-115.

[34] 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.

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

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

[37] Helgason, S. (1962). *Differential Geometry and Symmetric Spaces.* Academic Press, New York.

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

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

[40] 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.

[41] Ibragimov, I. A. and Khas'minskii, R. Z. (1984). On nonparametric estimation of the value of a linear functional in a Gaussian white noise. *Teor. Verojatnost. i Primenen.* **29** 19-32. (In Russian.)

[42] Ibragimov, I. A. and Khas'minskii, R. Z. (1991). Asymptotically normal families of distributions and effective estimation. *Ann. Statist.* **19** 1681-1724.

[43] Jones, M. C. and Sheather, S. J. (1991). Using non-stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives. *Statist. Probab. Lett.* **11** 511-514.

[44] Jupp, P. E. and Mardia, K. V. (1989). A Unified view of the theory of directional statistics, 1975-1988. *Int. Statist. Rev.* **57** 261-294.

[45] Koshevnik, Yu. A. and Levit, B. Ya. (1976). On a non-parametric analogue of the information matrix. *Theory Probab. Appl.* **21** 738-753.

[46] Krieger, A. M. and Pickands, J., III (1981). Weak convergence and efficient density estimation at a point. *Ann. Statist.* **9** 1066-1078.

[47] Laurent, B. (1993). Estimation of integral functional of a density and its derivatives. Technical Report 93-57, Univ. Paris Sud.

[48] Le Cam, L. (1972). Limits of experiments. *Proc. 6th Berkeley Symp. Math. Stat. Prob.* Vol. **1** 245-261.

[49] Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer, New York.

[50] Le Cam, L. and Yang, G. L. (1990). *Asymptotics in Statistics.* Springer, New York.

[51] Low, M. G. (1992). Renormalization and white noise approximation for nonparametric functional estimation problems. *Ann. Statist.* **20** 545-554.

[52] Low, M. G. (1993). Renormalization upper and lower bounds for integrated risk in the white noise model. *Ann. Statist.* **21** 577-589.

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

[54] Mardia, K. V. (1975). Statistics of directional data. *J. Roy. Statist. Soc. B* **37** 349-371.

[55] Meyer, T. G. (1977). Bounds for estimation of density functions and their derivatives. *Ann. Statist.* **5** 136-142.

[56] Millar, P. (1979). Asymptotic minimax theorems for the sample distribution function. *Z. Wahrsch. Verw. Gebiete* **48** 233-252.

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

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

[59] Nussbaum, M. (1995). Asymptotic equivalence of density estimation and Gaussian white noise. Manuscript.

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

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

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

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

[64] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832-837.

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

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

[67] Samarov, A. M. (1976). Minimax bound on the risk of nonparametric density estimates. *Problems of Inform. Transmission* **12** 242-244.

[68] Schick, A. S. (1986). On asymptotically efficient estimation in semiparametric models. *Ann. Statist.* **14** 1139-1151.

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

[70] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

[71] Sheather, S. J. (1983). A data-based algorithm for choosing the window width when estimating the density at a point. *Comp. Statist. Anal.* **1** 229-238.

[72] Sheather, S. J. (1986). An improved data-based algorithm for choosing the window width when estimating the density at a point. *Comp. Statist. Anal.* **4** 61-65.

[73] 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.

[74] Singh, R. S. (1976). Nonparametric estimation of mixed partial derivatives of a multivariate density. *J. Multivariate Anal.* **6** 111-122.

[75] Singh, R. S. (1979). Mean squared errors of estimates of a density and its derivatives. *Biometrika* **66** 177-180.

[76] Singh, R. S. (1987). MISE of kernel estimates of a density and its derivatives. *Statist. Probab. Lett.* **5** 153-159.

[77] Stone, C. J. (1980). Optimum rate of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348-1360.

[78] Strasser, H. (1985). *Mathematical Theory of Statistics.* de Gruyter, Amsterdam.

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

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

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

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

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