Lower Bounds for the Asymptotic Minimax Risk with Spherical Data*

Jussi Klemelä

Ruprecht-Karls-Universität Heidelberg

*This work was supported in part by the Academy of Finland under Grant no. 1015998, by the University of Helsinki under a research grant, and by Deutsche Forschungsgemeinschaft under Grant MA 1026/6-2.

Abstract

Lower bounds for the asymptotic minimax risk are given when estimating the density function with spherical data. Both the pointwise risk and the integrated risk are studied. Results are derived by constructing a sequence of experiments that converge weakly to a Gaussian shift experiment. Lower bounds derived in this article complement some previously derived upper bounds. We give one upper bound for the pointwise risk which is only a factor of 1.18 larger than the corresponding lower bound. The lower bounds are formulated in a general setting so that also the Euclidean results follow as a special case.

AMS 2000 subject classifications

Primary: 62G07
Secondary: 62H11, 62G20

Key words and phrases

asymptotic minimax risk, density estimation, directional data, Gaussian shift experiment, local asymptotic normality, spherical data, weak convergence of experiments.

1 Introduction

We will derive asymptotic lower bounds for the minimax risk over parameter spaces consisting of density functions defined on the unit sphere of a Euclidean space. The sample space is the unit sphere of a Euclidean space when the observations are directional. Good introductions to the statistics of directional data are given by Mardia (1972), Watson (1983), Fisher, Lewis and Embleton (1987). As one application area one can mention Earth sciences, where the observations can be for example measurements of geomagnetic field, paleomagnetic measurements, or determinations of crystallographic axes for petrofabric studies. In some applications we have data on other manifolds: data on projective plane is closely related with spherical data.

The lower bounds given in this article complement the upper bounds studied previously by Hall, Watson, and Cabrera (1987), Hendricks (1990), and Klemelä (2000), showing rate optimality of the procedures of those articles. Hendricks (1990) considered orthogonal series estimation and other two papers studied kernel estimation. These articles did not formulate results on minimax setting. Minimax setting was considered in Klemelä (1999) and Efromovich (2000). These papers prove sharp asymptotic results for supremum error and for L_2 error. For the pointwise risk exact results are not known for the minimax setting but we give in Section 4.3.2 one upper bound for pointwise risk which is within of a factor 1.18 from the corresponding lower bound.

Sphere behaves locally as Euclidean space and thus we are able to give a unified treatment of the Euclidean and spherical cases. When estimating density function at one point and with non-parametric assumptions, we have to consider only a local neigbourhood of this one point of the sphere. Next, the study of the integrated risk is reduced to the study of the pointwise risk by partitioning the sphere to an increasing number of local neigbourhoods.

We will use weak concergence of experiments in constructing lower bounds for the risks. Previously known Theorem 2.1 shows that weak convergence of experiments implies a lower bound for the asymptotic minimax risk. Define,

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

where $f_0 : \mathbf{R} \to \mathbf{R}$ is a density and $h : \mathbf{R} \to \mathbf{R}$ a Borel function. The product experiment $(\mathbf{R}^n, \mathcal{B}(\mathbf{R})^n, f_{h,n}^{(n)} : h \in \mathbf{H}_c)$ converges weakly to a Gaussian shift experiment for a certain $\mathbf{H}_c \subset \{h : \mathbf{R} \to \mathbf{R} \,|\, \int h^2 < \infty\}$. This fact was used by Millar (1979) to study the estimation of distribution functions and by Ibragimov and Khasmin'skii (1991) in some functional estimation problems. 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, but we are interested from so large parameter spaces that $\sqrt{n}$-consistent estimators do not exist. 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 Theorem 3.1 we consider a general localisation of the form $f_{h,n} = f_0 + T_n h$, where T_n is a suitable operator. Conditions are given for the transformation T_n, which are sufficient for the weak convergence of the corresponding product experiment to a Gaussian shift experiment. Theorem 3.1 covers both the Euclidean case and the spherical case and it generalizes Theorem 2 in Low (1992).

In Section 4 a lower bound is given for the asymptotic minimax risk, when estimating density at a point, which covers both the Euclidean and the spherical cases. The lower bound is given by the same method which was used by Low (1992). This method for deriving lower bounds could be called the method of local parametrization. The lower bound covers the case of estimating derivatives and Laplacians.

Previously, Donoho and Liu (1991) have calculated lower bounds for the asymptotic minimax risk, when estimating density at a point. They chosed a one-dimensional subexperiment, proved the convergence of this experiment to a Gaussian experiment, and finally optimized with respect to one-dimensional subfamilies. In Section 4.1.1 we consider the case of double-smooth densities and prove sharp lower bounds for general dimension, generalising the result of Donoho and Liu (1991), who considered one dimensional case. We will show in Section 4.1.1 that in the case of double-smooth densities, it is possible to get by the method of local parametrization as good lower bound as Donoho and Liu (1991) got by the method of one-dimensional subexperiment, in one dimensional case. In Section 4.2 we spesialize to the spherical case. In Section 4.2.1 we consider double-smooth densities and in Section 4.2.2 densities with bounded Laplacian. Section 4.2.2 considers estimation of Laplacian of a density. In Section 4.3 upper bounds are given for the pointwise risk, showing the sharpness of the lower bounds.

The advantage of the method of Donoho and Liu is that it is independent of the sample space and thus can be modified to the spherical case. However, the main motivation in this article for considering the method of local parameterization in the case of pointwise risk is that it can be modified to give lower bounds for the integrated risk. When studying estimation of a density at point, Euclidean and spherical cases have resemblances, due to the local equivalence of a sphere to a Euclidean space. When studying estimation of the whole density, we have to partition the sample space and use the weak convergence of experiments separately in each element of the partition.

In Section 5 a lower bound is given for the asymptotic minimax risk, when estimating density globally, which covers both the Euclidean and the spherical cases. In Section 5.1 an Euclidean example is given and in Section 5.2 an spherical example is given.

2 Basic Concepts

Let us recall the definitions of some concepts from the general theory of statistical decisions needed in this article. 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

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

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 $y \in \mathbf{R}$, $\{x \,|\, f(x) > y\}$ is open), f is bounded from below and $\lambda \in L(E)$, $\lambda \geq 0$, define $\beta(f,\lambda) = \sup\{\beta(g,\lambda) \,|\, g \in \mathcal{C}_b(\mathbf{D}),\ g \leq f\}$. 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 $\mathcal{B}(E,\mathbf{D}) = \{\beta_\rho \,|\, \rho \in \mathcal{R}(E,\mathbf{D})\}$. 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 theorem. This theorem is essentially based on the results in Le Cam (1972).

Theorem 2.1 *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 next give necessary facts about the unit sphere $S_d = \{x \in \mathbf{R}^{d+1} : \|x\| = 1\}$, $d \geq 1$. In this note we will be interested only about the cases $d \geq 2$. The Lebesgue measure of the sphere will be denoted by $\mu = \mu_d$. The Lebesgue measure of the sphere can be defined with the help of the Lebesgue measure of $\mathbf{R}^{d+1}$, see Rudin (1986, page 175). It is a Borel measure which is invariant with respect to rotations and with total mass $\omega_d \stackrel{def}{=} \mu(S_d) = 2\pi^{(d+1)/2}/\Gamma((d+1)/2)$.

The unit sphere can be parameterized in the following way. Choose such vectors $\eta^1, \ldots, \eta^{d+1} \in S_d$, $d \geq 1$, that $\eta^i \perp \eta^j$, when $i \neq j$. Let $\phi = \phi_{\eta^1,\ldots,\eta^{d+1}} : S_d \backslash U_{\eta^2,\ldots,\eta^{d+1}} \to]0, 2\pi[\, \times \,]0, \pi[^{d-1}$ be defined by

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

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

$$\int_{S_d} f\, d\mu = \int_0^{2\pi} d\theta_1 \int_0^\pi d\theta_2 \sin\theta_2 \cdots \int_0^\pi f\left(\phi^{-1}_{\eta^1,\ldots,\eta^{d+1}}(\theta)\right) \sin^{d-1}\theta_d\, d\theta_d. \tag{2}$$

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

$$\begin{aligned} &\Delta f(x) \qquad (3)\\ &= \sum_{k=1}^{d} (\sin\theta_{k+1}\cdot\ldots\cdot\sin\theta_d)^{-2}\left[\frac{\partial^2}{\partial\theta_k^2} + (k-1)\frac{\cos\theta_k}{\sin\theta_k}\frac{\partial}{\partial\theta_k}\right] f\left(\phi^{-1}(\theta)\right)\Big|_{\theta=\phi(x)}. \end{aligned}$$

3 Convergence to a Gaussian Shift Experiment

Firstly, let us define the Gaussian shift experiment, which will be our limit experiment. 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 $\|h\|_{\mathbf{H}}^2 =< h, h >_{\mathbf{H}}$, 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 $(\Omega, \mathcal{A}, \mathrm{P})$. The process $X(h)$ is such that its finite dimensional distributions are Gaussian, $\int X(h)d\mathrm{P} = 0$, and $\int X(h_1)X(h_2)d\mathrm{P} =< h_1, h_2 >_{\mathbf{H}}$. Let

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

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

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

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 us define the sequence of local experiments. Let $\mathcal{X}$ be a set carrying a sigma–field $\mathcal{F}$ with a sigma–finite measure μ. Let F be the set of measurable functions $\mathcal{X} \to \mathbf{R}$. Let $f_0 \in \mathrm{F}$ be a density with respect to μ. Let $\mathbf{H}_c \subset \mathbf{H}$, where $0 \in \mathbf{H}_c$, and let $T_n : \bar{\mathbf{H}}_c \to \mathrm{F}$ 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

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

be the direct n-fold product of the experiment $(\mathcal{X}, \mathcal{F}, P_{h,n} : h \in \mathbf{H}_c)$, where $P_{h,n}$ is probability measure corresponding to density $f_{h,n}$. 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).

Theorem 3.1 *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 $P_{h_1,n}^n(A_n) \to 0$, then $P_{h_2,n}^n(A_n) \to 0$.*

(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 $J \subset \mathbf{H}_c$,

$$\mathcal{L}\left(\left(\frac{dP_{h,n}^n}{dP_0^n}\right)_{h \in J} \middle| \; P_0^n\right) \longrightarrow \mathcal{L}\left(\left(\frac{dG_h}{dG_0}\right)_{h \in J} \middle| \; G_0\right).$$

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

$$\mathcal{L}\left(\left(\log \frac{f_{h,n}^{(n)}}{f_0^{(n)}}\right)_{h \in J} \middle| \; P_0^n\right) \longrightarrow N(m, \Sigma),$$

where $N(m, \Sigma)$ is the normal law with expectation $m = (-\frac{1}{2}\|h\|_{\mathbf{H}}^2)_{h \in J}$ and covariance matrix $\Sigma = (< h_1, h_2 >_{\mathbf{H}})_{h_1, h_2 \in J}$. Let $X^{(n)} = (X_1, \ldots, X_n) \sim P_0^n$ and define $X_{hni} = (T_n h)(X_i)/f_0(X_i)$, $i = 1, \ldots, n$. Now for $h \in \mathbf{H}_c$,

$$\log \frac{f_{h,n}^{(n)}(X^{(n)})}{f_0^{(n)}(X^{(n)})} = \sum_{i=1}^{n} \log(1 + X_{hni}) = Q_{hn} - \frac{1}{2} R_{hn} + S_{hn},$$

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 $\bar{X}_{hni} = X_{hni}I(|X_{hni}| \leq \delta)$ and $\bar{R}_{hn} = \sum_{i=1}^{n} \bar{X}_{hni}^2$. Now

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

Firstly,

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

by assumption (iv). Secondly,

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

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, $|S_{hn}| \leq R_{hn} \max_{i=1,\ldots,n} |r(X_{hni})|$, 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,

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

by assumption (iv). Let $J = \{h_1, \ldots, h_k\} \subset \mathbf{H}_c$ and $a = (a_1, \ldots, a_k)' \in \mathbf{R}^k$. 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 $j = 1, \ldots n$,

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

and Lindeberg's condition is satisfied. □

This proof shows in fact that the LAN condition, as formulated by Ibragimov and Khas'minskii (1991), is satisfied.

4 Pointwise Risk

In this section a lower bound is given for the local asymptotic minimax risk in the case of the pointwise risk. Then some examples are presented, where bounds for the global asymptotic minimax risk are given, both in the Euclidean case and in the spherical case. Finally examples from upper bounds are given.

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

Assume that

$$E_n(\mathbf{H}_c) \xrightarrow{w} E(\mathbf{H}_c), \tag{4}$$

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 $D_{\mathbf{F}} : \mathbf{F} \to \mathrm{F}$ and $D_{\mathbf{G}} : \mathbf{G} \to \mathrm{G}$ that for a fixed $x_0 \in \mathcal{X}$,

$$D_{\mathbf{F}}(T_n h)(x_0) = \gamma_n^{-1} \Phi(f_0) \left(D_{\mathbf{G}} h(y_0) + \delta_n(h) \right), \tag{5}$$

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.

Theorem 4.1 *Let $p > 0$, $W_f : \mathbf{R} \to \mathbf{R}$, $W_f(a) = |D_{\mathbf{F}} f(x_0) - a|^p$, where $f \in \mathbf{F}$, $\tilde{W}_h : \mathbf{R} \to \mathbf{R}$, $\tilde{W}_h(a) = |D_{\mathbf{G}} h(y_0) - a|^p$, where $h \in \mathbf{G}$. Then, under the assumptions (4) and (5),*

$$\liminf_{n\to\infty} \gamma_n^p \inf_{\rho_n \in \mathcal{R}(E_n(\mathbf{H}_c), \mathbf{R})} \sup_{h \in \mathbf{H}_c} \beta_{\rho_n} \left(W_{f_{h,n}}, P_{h,n}^n \right) \geq |\Phi(f_0)|^p \inf_{\rho \in \mathcal{R}(E(\mathbf{H}_c), \mathbf{R})} \sup_{h \in \mathbf{H}_c} \beta_\rho \left(\tilde{W}_h, G_h \right).$$

Proof. By the assumption (5),

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

Thus,

$$
\begin{aligned}
&\inf_{\rho_n\in\mathcal{R}(E_n(\mathbf{H}_c),\mathbf{R})}\sup_{h\in\mathbf{H}_c}\beta_{\rho_n}\left(W_{f_{h,n}},P^n_{h,n}\right)\\
&= \quad |\gamma_n^{-1}\Phi(f_0)|^p\inf_{\rho_n\in\mathcal{R}(E_n(\mathbf{H}_c),\mathbf{R})}\sup_{h\in\mathbf{H}_c}\beta_{\rho_n}\left\{\tilde{W}_h\left(\gamma_n\Phi(f_0)^{-1}(\cdot - D_{\mathbf{F}}f_0(x_0))+\delta_n(h)\right),P^n_{h,n}\right\}\\
&= \quad |\gamma_n^{-1}\Phi(f_0)|^p\inf_{\rho_n\in\mathcal{R}(E_n(\mathbf{H}_c),\mathbf{R})}\sup_{h\in\mathbf{H}_c}\beta_{\rho_n}\left(\tilde{W}_h(\cdot-\delta_n(h)),P^n_{h,n}\right).
\end{aligned}
$$

Now $\mathcal{R}(E_n(\mathbf{H}_c),\mathbf{R})\subset\mathcal{B}(E_n(\mathbf{H}_c),\mathbf{R})$ and $\mathcal{B}(E(\mathbf{H}_c),\mathbf{R})=\{\beta_\rho\mid\rho\in\mathcal{R}(E(\mathbf{H}_c),\mathbf{R})\}$ 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 (4) and Lemma 2.1. $\square$

The minimax risk in the Gaussian shift experiment, which appears in the lower bound in Theorem 4.1, has a lower bound that will be given in the next lemma.

Lemma 4.2 *Let $\mathbf{H}_c\subset\mathbf{H}$ be convex. Let $T:\mathbf{H}_c\to\mathbf{R}$ be linear. Define*

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

and

$$\psi_p(\tau,\sigma)=\inf_{\hat{T}}\sup_{|\theta|\le\tau}\int_{\mathbf{R}}\left|\hat{T}-\theta\right|^p dN(\theta,\sigma^2), \tag{6}$$

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

$$\inf_{\rho\in\mathcal{R}(E(\mathbf{H}_c),\mathbf{R})}\sup_{h\in\mathbf{H}_c}\beta_\rho\left(\tilde{W}_h,G_h\right)\ge\sup_{\epsilon>0}\left(\frac{\omega_{\mathbf{H}_c}(\epsilon)}{\epsilon}\right)^p\psi_p\left(\frac{\epsilon}{2},1\right)$$

where $\tilde{W}_h:\mathbf{R}\to\mathbf{R}$, $\tilde{W}_h(a)=|T(h)-a|^p$.

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 $\psi_p(\tau,\sigma)=\sigma^p\psi_p(\tau/\sigma,1)$. $\square$

Information and references concerning the function $\psi_2(\tau,\sigma)$ is given in Donoho, Liu and MacGibbon (1990). Note that when $\omega_{\mathbf{H}_c}(\epsilon)=A\epsilon^r$, $0<A<\infty$, $0<r\le1$, the lower bound of Lemma 4.2 can be written as

$$\left(\frac{\omega_{\mathbf{H}_c}(2)}{2}\right)^p\xi_p(r), \tag{7}$$

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

4.1 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 $f_0 : \mathbf{R}^d \to \mathbf{R}$ be a density function, $f_0(x_0) > 0$, and let f_0 be continuous at x_0. For $h : \mathbf{R}^d \to \mathbf{R}$, define

$$T_n h = f_0 \left[U_n h - \int_{\mathcal{X}} f_0 (U_n h) d\mu \right] \tag{8}$$

where

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

$\lim_{n\to\infty} \kappa_n = \infty$, and $\lim_{n\to\infty} \kappa_n^d n^{-1} = 0$. 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 $\int h^2 < \infty$ and $|h| \le c$. From Theorem 3.1 it is seen that $E_n(\mathbf{H}_c) \xrightarrow{w} E(\mathbf{H}_c)$.

In the application of Theorem 4.1 we choose $\gamma_n = n^{1/2} \kappa_n^{-r-d/2}$, $\Phi(f_0) = f_0(x_0)^{1/2}$, $D_{\mathbf{F}} = D_{\mathbf{G}} = D^\alpha$, $\delta_n(h) = \int f_0(x) h(\kappa_n(x - x_0)) dx$ when $r = 0$, and $\delta_n(h) = 0$ when $r \ge 1$ and f_0 is constant in some neigborhood of x_0.

4.1.1 Double-smooth densities

Let us study carefully the estimation of $f(x_0)$ in the case f is a 2-smooth density. We define the Sacks–Ylvisaker class of densities by

$$\mathbf{SY}(\mathcal{X}, M, \delta, C, x_0) = \left\{ f : \mathcal{X} \to \mathbf{R} \;\middle|\; 0 \le f \le M, \int_{\mathcal{X}} f d\mu = 1 \right\} \cap \mathbf{PP}(\mathcal{X}, \delta, C, x_0)$$

where $\mathcal{X} = \mathbf{R}^d$ or $\mathcal{X} = S_d$, and

$$\begin{aligned} \mathbf{PP}(\mathcal{X}, \delta, C, x_0) = \{ f : \mathcal{X} \to \mathbf{R} \quad | \quad & f(x) = f(x_0) + (x - x_0)' \nabla f(x_0) + r(x_0, x), \\ & \left. |r(x_0, x)| \le \frac{C}{2} \|x - x_0\|^2, \|x - x_0\| \le \delta \right\}. \end{aligned}$$

For $f : \mathbf{R}^d \to \mathbf{R}$ the notation $\nabla f(x)$ means the gradient of f at x, and for $f : S_d \to \mathbf{R}$ it means the gradient of $f(x/\|x\|)$.

In this section $\mathcal{X} = \mathbf{R}^d$. This case was discussed in Donoho and Liu (1991) with $d = 1$, $p = 2$, $C = 1$, and $x_0 = 0$. We will see that the lower bound obtained from Theorem 4.1 and Lemma 4.2 will give the same lower bound as was given by Donoho and Liu (1991, Theorem 9), for the case $d = 1$, $p = 2$.

Let $f_0 : \mathbf{R}^d \to \mathbf{R}$ be a density function, $|f_0| < M$, and $f_0(x) = f_0(x_0) > 0$, when $\|x - x_0\| \leq \delta$. Let $\mathbf{H}_c$, $0 < c < \infty$, be the set of such functions in $\mathbf{PP}(\mathbf{R}^d, \delta, f_0(x_0)^{-1/2}C, 0)$ that $\int h^2 < \infty$ and $|h| \leq c$. By choosing $\kappa_n = n^{1/(d+4)}$ we obtain $\{f_{h,n} \mid h \in \mathbf{H}_c\} \subset \mathbf{SY}(\mathbf{R}^d, M, \delta, C, x_0)$ ultimately. Thus, using Theorem 4.1,

$$\begin{aligned} &\liminf_{n\to\infty} \gamma_n^p \inf_{\rho_n \in \mathcal{R}(F_n, \mathbf{R})} \sup_{f \in \mathbf{SY}(\mathbf{R}^d, M, \delta, C, x_0)} \beta_{\rho_n}\left(W_f, P_f^n\right) \\ &\geq \; f_0(x_0)^{p/2} \inf_{\rho \in \mathcal{R}(E(\mathbf{H}_c), \mathbf{R})} \sup_{h \in \mathbf{H}_c} \beta_\rho\left(\tilde{W}_h, G_h\right) \qquad (9) \end{aligned}$$

where $F_n = \left(\mathcal{X}^n, \mathcal{F}^n, P_f^n : f \in \mathbf{SY}(\mathbf{R}^d, M, \delta, C, x_0)\right)$, $W_f : \mathbf{R} \to \mathbf{R}$, $W_f(a) = |D^\alpha f(x_0) - a|^p$, $\tilde{W}_h : \mathbf{R} \to \mathbf{R}$, $\tilde{W}_h(a) = |D^\alpha h(0) - a|^p$, and

$$\gamma_n = n^{1/2} \kappa_n^{-d/2} = n^{2/(d+4)}.$$

It can be calculated similarily as in Donoho and Liu (1991, Section 4.1), that

$$\begin{aligned} \omega_{\mathbf{H}_c}(\epsilon) &= \omega_{\mathbf{PP}(\mathbf{R}^d, \delta, f_0(x_0)^{-1/2}C, x_0)}(\epsilon) \qquad (10) \\ &= \epsilon^{4/(d+4)} \left(f_0(x_0)^{-1/2} C\right)^{d/(d+4)} \left(\frac{(d+4)(d+2)d}{8\omega_{d-1}}\right)^{2/(d+4)} \end{aligned}$$

and thus from (7) we get a lower bound for (9) as

$$M^{2p/(d+4)} C^{pd/(d+4)} \left(\frac{(d+4)(d+2)d}{8\omega_{d-1}}\right)^{2p/(d+4)} 2^{-pd/(d+4)} \xi(4/(d+4)).$$

Because $\xi_2(4/5) \geq 0.488$ (Donoho and Liu, 1991, Table 1) and $\omega_0 = 2$, then in the case $d = 1$, $p = 2$, we get the lower bound $M^{4/5} C^{2/5} 15^{-1/5} 3/5$, as given in Donoho and Liu (1991, Theorem 9). We have $\xi_2(2/3) \geq 0.449$ (Donoho and Liu, 1991, Table 1) and $\omega_1 = 2\pi$, and then in the case $d = 2$, $p = 2$, we get the lower bound $(3MC/(2\pi))^{2/3} 0.499$.

4.1.2 Densities with Bounded Derivatives

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

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

Let f_0 be a density which is constant (> 0) in some neigborhood of x_0 and let $\mathbf{H}_c$, $0 < c < \infty$ be the set of measurable functions $h : \mathbf{R}^d \to \mathbf{R}$ for which $\int h^2 < \infty$, $|h| \le c$, $|D^\alpha h(0)| \le c$ for $|\alpha| = r$, and $|D^\beta h(0)| \le f_0(x_0)^{-1/2} C$ for $|\beta| = s$. By choosing $\kappa_n = n^{1/(2s+d)}$ we obtain $\{f_{h,n} \,|\, h \in \mathbf{H}_c\} \subset \mathbf{F}(s, C, x_0)$ ultimately. Thus, using Theorem 4.1,

$$\begin{aligned}
&\liminf_{n\to\infty} \gamma_n^p \inf_{\rho_n \in \mathcal{R}(F_n, \mathbf{R})} \sup_{f \in \mathbf{F}(s,C,x_0)} \beta_{\rho_n}\left(W_f, P_f^n\right) \\
&\quad \ge \liminf_{n\to\infty} \gamma_n^p \inf_{\rho_n \in \mathcal{R}(E_n(\mathbf{H}_c), \mathbf{R})} \sup_{h \in \mathbf{H}_c} \beta_{\rho_n}\left(W_{f_{h,n}}, P_{h,n}^n\right) \\
&\quad \ge f_0(x_0)^{p/2} \inf_{\rho \in \mathcal{R}(E(\mathbf{H}_c), \mathbf{R})} \sup_{h \in \mathbf{H}_c} \beta_\rho\left(W_h, G_h\right)
\end{aligned}$$

where $F_n = \left(\mathcal{X}^n, \mathcal{F}^n, P_f^n : f \in \mathbf{F}(s, C, x_0)\right)$, $W_f : \mathbf{R} \to \mathbf{R}$, $W_f(a) = |D^\alpha f(x_0) - a|^p$, and

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

Analogous results have been proved by Farrell (1972), Meyer (1977), Stone (1980), and Low (1992).

4.2 Spherical case

Let $\mathcal{X} = S_d$, let μ be the Lebesgue measure of S_d, and $\mathcal{F}$ the Borel sigma–algebra of S_d. Without loss of generality, let $x_0 = (0, \ldots, 0, 1) \in S_d$. For a density $f : S_d \to \mathbf{R}$, the estimation of $\Delta^{r/2} f(x_0)$ will be discussed, where $r \ge 0$ is even and Δ is the Laplace operator of the sphere.

Let $f_0 : S_d \to \mathbf{R}$ be a density function, $f_0(x_0) > 0$, and let f_0 be continuous at x_0. Let $T_n h$ be defined as in (8) but

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

where $A_n = \{x \in UH \,|\, \kappa_n \phi(x) \in B(\sqrt{2})\}$, $UH = \{x \in S_d \,|\, x_{d+1} \ge 0\}$ (upper hemisphere), $B(\sqrt{2}) = \{y \in \mathbf{R}^d \,|\, \|y\| \le \sqrt{2}\}$, $\phi : UH \to B(\sqrt{2})$ is the equal area projection, defined by $\phi(x) = \sqrt{2}(1 + x_{d+1})^{-1/2}(x_1, \ldots, x_d)$, $\lim_{n\to\infty} \kappa_n = \infty$, and $\lim_{n\to\infty} \kappa_n^d n^{-1} = 0$.

Let $\mathcal{Y} = \mathbf{R}^d$, let ν be the Lebesgue measure of $\mathbf{R}^d$, and $\mathcal{G}$ the Borel sigma–algebra of $\mathbf{R}^d$. Let $\mathbf{H}_c$, $0 < c < \infty$, be the set of measurable functions $h : \mathbf{R}^d \to \mathbf{R}$ for which

h has support on $B(\sqrt{2})$ and $|h| < c$. From Lemma 3.1 it is seen that

$$E_n(\mathbf{H}_c) \xrightarrow{w} E(\mathbf{H}_c) \tag{11}$$

Indeed,

$$\begin{aligned} & n \int_{S_d} (T_n h_1)\,(T_n h_2)\, f_0^{-1} d\mu \\ &= n \int_{S_d} (U_n h_1)\,(U_n h_2)\, f_0 d\mu - n \int_{S_d} (U_n h_1)\, f_0 d\mu \int_{S_d} (U_n h_2)\, f_0 d\mu \end{aligned}$$

and

$$\begin{aligned} & n \int_{S_d} (U_n h_1)\,(U_n h_2)\, f_0 d\mu \\ &= f_0(x_0)^{-1} \kappa_n^d \int_{A_n} h_1\left(\kappa_n \phi(x)\right) h_2\left(\kappa_n \phi(x)\right) f_0(x) d\mu(x) \\ &= f_0(x_0)^{-1} \kappa_n^d \int_{\kappa_n^{-1} B(\sqrt{2})} h_1(\kappa_n y) h_2(\kappa_n y) \left(1 - \frac{1}{4}\|y\|^2\right)^{(d-2)/2} f_0(\phi^{-1}(y)) dy \\ &= f_0(x_0)^{-1} \int_{B(\sqrt{2})} h_1(y) h_2(y) \left(1 - \frac{1}{4\kappa_n^2}\|y\|^2\right)^{(d-2)/2} f_0(\phi^{-1}(\kappa_n^{-1} y)) dy \\ &= \int_{\mathcal{Y}} h_1 h_2 d\nu + o(1). \end{aligned}$$

Also, $\int_{S_d} (U_n h_i)\, f_0 d\mu = O(\kappa_n^{-d/2})$, for $i = 1, 2$.

In the case $r = 0$, (5) holds in the form

$$(T_n h)(x_0) = f_0(x_0)^{1/2} n^{-1/2} \kappa_n^{d/2} \left[h(0) - \int_{A_n} f_0(x) h(\kappa_n \phi(x)) dx \right]$$

and Lemma 4.1 can be applied with $\gamma_n = n^{1/2}\kappa_n^{-d/2}$, $\Phi(f_0) = f_0(x_0)^{1/2}$, $\delta_n(h) = \int_{A_n} f_0(x) h(\kappa_n \phi(x)) d\mu(x)$, and $D_{\mathbf{F}}$ and $D_{\mathbf{G}}$ as identity operators. In the case $r \geq 1$ we will for simplicity restrict ourselves to those functions $h \in \mathbf{H}_c$ which have the form $h(x) = g(\|x\|^2)$, where $g : [0, \infty[\to \mathbf{R}$. If f_0 is constant in some neigborhood of x_0, then for $h(x) = g(\|x\|^2)$,

$$\begin{aligned} & \Delta^{r/2} (T_n h)(x_0) \\ &= f_0(x_0)^{1/2} n^{-1/2} \kappa_n^{r+d/2} \left[(-1)^{r/2} c_r g^{(r/2)}(0) + \sum_{i=1}^{r/2-1} (-1)^i c_{r,i} \kappa_n^{2i-r} g^{(i)}(0) \right] \end{aligned} \tag{12}$$

where c_r and $c_{r,i}$ are constants (for details, see Klemelä 1997, Lemma 5.8). Lemma 4.1 can be applied with $\gamma_n = n^{1/2}\kappa_n^{-r-d/2}$, $\Phi(f_0) = f_0(x_0)^{1/2}(-1)^{r/2}c_r$,

$$\delta_n(h) = \sum_{i=1}^{r/2-1} (-1)^{i-r/2} c_{r,i} c_r^{-1} \kappa_n^{2i-r} g^{(i)}(0),$$

$D_{\mathbf{F}} = \Delta^{r/2}$, and $D_{\mathbf{G}} = \partial^r/\partial t^r$.

4.2.1 Spherical double-smooth densities

Estimation of $f(x_0)$ will be discussed, when f is a density known to lie in the spherical Sacks–Ylvisaker class of densities $\mathbf{SY}(S_d, M, \delta, C, x_0)$, where $d = 2$.

Theorem 4.3 *Let $d = 2$, $F_n = \left(\mathcal{X}^n, \mathcal{F}^n, P_f^n : f \in \mathbf{SY}(S_d, M, \delta, C, x_0)\right)$, $p > 0$, $W_f : \mathbf{R} \to \mathbf{R}$, $W_f(a) = |f(x_0) - a|^p$. Then*

$$\liminf_{n\to\infty} n^{p/3} \inf_{\rho_n \in \mathcal{R}(F_n, \mathbf{R})} \sup_{f \in \mathbf{SY}(S_d, M, \delta, C, x_0)} \beta_{\rho_n}\left(W_f, P_f^n\right) \geq \left(\frac{3MC}{2\pi}\right)^{p/3} 0.449.$$

Proof. Let $f_0 : S_d \to \mathbf{R}$ be a density function, $|f_0| < M$, and $f_0(x) = f_0(x_0) > 0$, when $\|x - x_0\| \leq \delta$. Let us define the following modification of the parallelipiped,

$$\begin{aligned}
&\mathbf{PP}(\mathbf{R}^d, \delta, C, y_0, \zeta) \\
&= \Big\{ f : \mathbf{R}^d \to \mathbf{R} \,|\, f(y) = f(y_0) + \zeta(\|y - y_0\|)(y - y_0)'\nabla f(y_0) + r(y_0, y), \\
&\qquad |r(y_0, y)| \leq \frac{C}{2}\|y - y_0\|^2, \|y - y_0\| \leq \delta \Big\}
\end{aligned}$$

where $\zeta : [0, \delta] \to [0, 1]$, $\zeta(0) = 1$. Let $\mathbf{H}_c$, $0 < c < \infty$, be the set of such functions in $\mathbf{PP}(\mathbf{R}^d, \delta, f_0(x_0)^{-1/2}C, 0, \zeta)$, $\zeta(r) = (1 - r^2/4)^{1/2}$, that h has support on $B(\sqrt{2}) = \{y \in \mathbf{R}^d \,|\, \|y\| \leq \sqrt{2}\}$ and $|h| \leq c$.

Choosing $\kappa_n = n^{1/(d+4)}$ it follows that $\{f_{h,n} \,|\, h \in \mathbf{H}_c\} \subset \mathbf{SY}(S_d, M, \delta, C, x_0)$ ultimately. Indeed, putting $\tilde{x} = (x_1, \ldots x_d)$,

$$\begin{aligned}
&|f_{h,n}(x) - f_{h,n}(x_0) - (x - x_0)'\nabla f_{h,n}(x_0)| \\
&\leq f_0(x_0)\,|U_n h(x) - U_n h(x_0) - (x - x_0)'\nabla(U_n h)(x_0)| \\
&= f_0(x_0)^{1/2}\kappa_n^{-2}\,|h(\kappa_n\phi(x)) - h(0) - \kappa_n\tilde{x}'\nabla h(0)| \\
&\leq C\|\phi(x)\|^2/2 = C\|x - x_0\|^2/2
\end{aligned}$$

because $(x-x_0)'\nabla(U_n h)(x_0) = \kappa_n \tilde{x}'\nabla h(0)$ by the fact that

$$\frac{\partial}{\partial x_i}\phi_k\left(\frac{x}{\|x\|}\right)\Big|_{x=x_0} = \begin{cases} 0 & \text{if } i \neq k, k \in \{1,\ldots,d\}, i \in \{1,\ldots,d+1\} \\ 1 & \text{if } i = k \in \{1,\ldots,d\} \end{cases}$$

and $\tilde{x}'\nabla h(0) = \zeta(\|\phi(x)\|)\phi(x)'\nabla h(0)$.

From (11) we have the satisfaction of (4) and from (12) we have the satisfaction of (5). Thus Theorem 4.1 can be applied with

$$\gamma_n = n^{1/2}\kappa_n^{-d/2} = n^{2/(d+4)},$$

$\Phi(f_0) = f_0(x_0)^{1/2}$, $\delta_n(h) = \int f_0(x)h(\kappa_n\phi(x))d\mu(x)$, and $D_{\mathbf{F}}$ and $D_{\mathbf{G}}$ as identity operators. Thus

$$\begin{aligned}\liminf_{n\to\infty} n^{p/3} \inf_{\rho_n\in\mathcal{R}(F_n,\mathbf{R})} \sup_{f\in\mathbf{SY}(S_d,M,\delta,C,x_0)} \beta_{\rho_n}\left(W_f, P_f^n\right) \\ \geq f_0(x_0)^{p/2} \inf_{\rho\in\mathcal{R}(E(\mathbf{H}_c),\mathbf{R})} \sup_{h\in\mathbf{H}_c} \beta_\rho\left(\tilde{W}_h, G_h\right). \qquad (13)\end{aligned}$$

The modulus of continuity of $\mathbf{PP}(\mathbf{R}^d,\delta,C,x_0,\zeta)$ does not depend on ζ. Thus from (7) and (10) we get the lower bound for (13) as

$$f_0(x_0)^{p/2}\left(f_0(x_0)^{-1/2}C\right)^{p/3}\left(\frac{3}{\pi}\right)^{p/3} 2^{-p/3}\xi(2/3).$$

Because $\|f_0\|_\infty = M-\epsilon$, where $0<\epsilon<M$ was chosen arbitrarily, we get a better lower bound as

$$\left(\frac{3MC}{2\pi}\right)^{p/3}\xi(2/3).$$

By Donoho and Liu (1991, Table 1) $\xi(2/3) \geq 0.449$. □

4.2.2 Spherical Densities with Bounded Laplacians

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

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

Let $f_0 : S_d \to \mathbf{R}$ be a density function, which is constant (>0) in some neigborhood of x_0. Let $\mathbf{H}_{0\infty}$ be the set of measurable functions $h : \mathbf{R}^d \to \mathbf{R}$ for which h has support on $B(\sqrt{2})$, $|h| < \infty$, $h(x) = g(\|x\|^2)$, $g : [0,\infty[\to \mathbf{R}$, $g^{(s)}$ exists on $[0,\infty[$, and $|c_s g^{(s/2)}(0)| \leq f(x_0)^{-1/2}C$, where c_s is the constant from equation (12)

Theorem 4.4 *Let* $F_n = \left(\mathcal{X}^n, \mathcal{F}^n, P_f^n : f \in \mathbf{F}(s, C, x_0)\right)$*,* $p > 0$*,* $W_f : \mathbf{R} \to \mathbf{R}$*,* $W_f(a) = |\Delta^{r/2} f(x_0) - a|^p$*,* $\tilde{W}_h : \mathbf{R} \to \mathbf{R}$*,* $\tilde{W}_h(a) = |g^{(r/2)}(0) - a|^p$*, where* $h(x) = g(\|x\|^2)$*. Then*

$$\begin{aligned}
&\liminf_{n\to\infty} n^{p(s-r)/(2s+d)} \inf_{\rho_n \in \mathcal{R}(F_n, \mathbf{R})} \sup_{f \in \mathbf{F}(s,C,x_0)} \beta_{\rho_n}\left(W_f, P_f^n\right) \\
&\geq \quad |\Phi(f_0)|^p \inf_{\rho \in \mathcal{R}(E(\mathbf{H}_{0\infty}), \mathbf{R})} \sup_{h \in \mathbf{H}_{0\infty}} \beta_\rho\left(\tilde{W}_h, G_h\right).
\end{aligned}$$

where $\Phi(f_0) = f_0(x_0)^{1/2}(-1)^{r/2} c_r$*.*

Proof. Let $\mathbf{H}_{0c}$, $0 < c < \infty$, be the set of those functions in $\mathbf{H}_{0\infty}$ for which $|h| \leq c$, $|g^{(i)}(0)| \leq c$ for $i = 1, \ldots, s/2 - 1$, and $(1 + c^{-1})|c_s g^{(s/2)}(0)| \leq f(x_0)^{-1/2} C$. Choosing $\kappa_n = n^{1/(2s+d)}$ it follows that $\{f_{h,n} \,|\, h \in \mathbf{H}_{0c}\} \subset \mathbf{F}(s, C, x_0)$ ultimately. In fact, now $n^{-1/2}\kappa_n^{d/2} = \kappa_n^{-s}$ and for $h \in \mathbf{H}_{0c}$,

$$\begin{aligned}
&|\Delta^{s/2} f_{h,n}(x_0)| = \left|\Delta^{s/2} f_0(x_0) + \Delta^{s/2}(T_n h)(x_0)\right| \\
&\leq \quad |\Delta^{s/2} f_0(x_0)| + f_0(x_0)^{1/2}\left|c_s g^{(s/2)}(0)\right| + f_0(x_0)^{1/2} \kappa_n^{-s} \sum_{i=1}^{s/2-1} \kappa_n^{2i} \left|c_{s,i} g^{(i)}(0)\right| \leq C
\end{aligned}$$

for sufficiently large n.

From (11) we have the satisfaction of (4) and from (12) we have the satisfaction of (5) and thus Theorem 4.1 can be applied with $\gamma_n = \kappa_n^{s-r} = n^{(s-r)/(2s+d)}$. Let $\Pi_0(\mathbf{H}_{0c})$ be the set of probability measures on $\mathbf{H}_{0c}$ with finite support. Because $0 < c < \infty$ was chosen arbitrarily, we have in fact the following lower bound.

$$\begin{aligned}
&|\Phi(f_0)|^p \sup_{c>0} \inf_{\rho \in \mathcal{R}(E(\mathbf{H}_{0c}), \mathbf{R})} \sup_{h \in \mathbf{H}_{0c}} \beta_\rho\left(\tilde{W}_h, G_h\right) \\
&= \quad |\Phi(f_0)|^p \sup_{c>0} \sup_{\pi \in \Pi_0(\mathbf{H}_{0c})} \inf_{\rho \in \mathcal{R}(E(\mathbf{H}_{0c}), \mathbf{R})} \int_{\mathbf{H}_{0c}} \beta_\rho\left(\tilde{W}_h, G_h\right) d\pi(h) \\
&= \quad |\Phi(f_0)|^p \sup_{\pi \in \Pi_0(\mathbf{H}_{0\infty})} \inf_{\rho \in \mathcal{R}(E(\mathbf{H}_{0\infty}), \mathbf{R})} \int_{\mathbf{H}_{0\infty}} \beta_\rho\left(\tilde{W}_h, G_h\right) d\pi(h) \\
&= \quad |\Phi(f_0)|^p \inf_{\rho \in \mathcal{R}(E(\mathbf{H}_{0\infty}), \mathbf{R})} \sup_{h \in \mathbf{H}_{0\infty}} \beta_\rho\left(\tilde{W}_h, G_h\right),
\end{aligned}$$

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

4.3 Upper bounds for pointwise risk

Define global minimax risk by

$$\inf_{\hat{T}} \sup_{f \in \mathbf{F}_0} E_{f^n} \left| D_{\mathbf{F}} f(x_0) - \hat{T} \right|^p .$$

If $\{f_{h,n} : h \in \mathbf{H}_c\} \subset \mathbf{F}_0$ ultimately, then Theorem 4.1 gives an asymptotic lower bound for the global minimax risk. We may derive upper bound for the global minimax risk by the reasoning of Donoho and Liu (1991, Lemma 4, Theorem 2). Indeed, if $p = 2$, $\sup_{f \in \mathbf{F}_0} \|f\|_\infty \le M$, $\mathbf{F}_0$ is convex, and $\omega_{\mathbf{F}_0}(\epsilon) \to 0$, when $\epsilon \to 0$, then an upper bound for the global minimax risk is

$$\sup_{\epsilon > 0} \left(\frac{\omega_{\mathbf{F}_0}(\epsilon)}{\epsilon} \right)^2 \psi_{A,2} \left(\frac{\epsilon}{2}, \sqrt{\frac{M}{n}} \right)$$

where $\psi_{A,2}$ is otherwise similaly defined as ψ_2 in (6), but infimum is taken only with respect to affine estimators $\mathbf{R} \to \mathbf{R}$. By Donoho and Liu (1991, Corollary 3), if $\omega_{\mathbf{F}_0}(\epsilon) = A\epsilon^r$, then the previous expression is equal to

$$\left(\frac{M}{n} \right)^r \left(\frac{\omega_{\mathbf{F}_0}(2)}{2} \right)^2 r^r (1-r)^{1-r}. \tag{14}$$

4.3.1 Double-smooth densities in Euclidean case

Consider $\mathbf{F}_0 = \mathbf{SY}(\mathbf{R}^d, M, \delta, C, x_0)$. We have $\omega_{\mathbf{F}_0}(\epsilon) = \omega_{\mathbf{SY}(\mathbf{R}^d, M, \delta, C, x_0)}(\epsilon)$. Thus equation (10) and (14) give the upper bound, when $p = 2$,

$$\begin{aligned} &\liminf_{n \to \infty} \gamma_n^2 \inf_{\rho_n \in \mathcal{R}(F_n, \mathbf{R})} \sup_{f \in \mathbf{SY}(\mathbf{R}^d, M, \delta, C, x_0)} \beta_{\rho_n} \left(W_f, P_f^n \right) \\ &\le M^{4/(d+4)} C^{2d/(d+4)} \left(\frac{(d+4)(d+2)d}{8\omega_{d-1}} \right)^{4/(d+4)} 2^{-2d/(d+4)} \\ &\quad \times \left(\frac{4}{d+4} \right)^{4/(d+4)} \left(\frac{d}{d+4} \right)^{d/(d+4)} \end{aligned}$$

where $\gamma_n = n^{2/(d+4)}$. This generalises the result of Donoho and Liu (1991) to multivariate case. When $d = 1$ we get the upper bound $M^{4/5} C^{2/5} 15^{-1/5} 3/4$, which is larger than the lower bound of Section 4.1.1 by a factor of 1.25. When $d = 2$, we get the upper bound $(MC/\pi)^{2/3} (1/3)^{1/3}$, which is larger than the lower bound of Section 4.1.1 by a factor of 1.179.

4.3.2 Double-smooth densities in the spherical case

When $d = 2$ the modulus of continuity of $\mathbf{PP}(S_d, \delta, C, x_0)$ is the same as the modulus of continuity of $\mathbf{PP}(\mathbf{R}^d, \delta, C, x_0)$, and then we get from (14) and (10) an upper bound, when $p = 2$,

$$\inf_{\rho_n \in \mathcal{R}(F_n, \mathbf{R})} \sup_{f \in \mathbf{SY}(S_d, M, \delta, C, x_0)} \beta_{\rho_n} \left(W_f, P_f^n\right) \leq \left(\frac{MC}{n\pi}\right)^{2/3} \left(\frac{1}{3}\right)^{1/3}.$$

The upper bound is the same as in Euclidean case and it is larger than the lower bound of Theorem 4.3 by a factor of 1.179.

5 Integrated Risk

In this section a lower bound is given for the local asymptotic minimax risk in the case of the integrated risk. Then examples are presented, where lower bounds for the global asymptotic minimax risk are given in the Euclidean case and in the spherical case.

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 $\mathbf{F}_{n,i} = \{f I_{A_{n,i}} \mid f \in \mathbf{F}\}$, where $I_{A_{n,i}}$ is the indicator of the set $A_{n,i}$. Let $\mathbf{H}_c \subset \mathbf{H} \cap \mathbf{G}$, where $0 \in \mathbf{H}_c$ and $\mathbf{H} = L_2(\mathcal{Y}, \mathcal{G}, \nu)$. Let $f_0 \in \mathbf{F}$ be a density whose support includes all $A_{n,i}$. Let $T_{n,i} : \bar{\mathbf{H}}_c \to \mathbf{F}_{n,i}$, $i \in I_n$, 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 $\underline{h}_n = (h_i)_{i \in I_n} \in \mathbf{H}_c^{I_n}$. 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 $\{f_{\underline{h}_n, n} \mid \underline{h}_n \in \mathbf{H}_c^{I_n}\} \subset \mathbf{F}$.

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 $k(n) \in I_n$,

$$E_n(\mathbf{H}_c) \xrightarrow{w} E(\mathbf{H}_c), \tag{15}$$

where $E(\mathbf{H}_c)$ is a Gaussian shift experiment defined in Section 3. Assume secondly that for all $h \in \mathbf{H}_c$, $f \in \mathbf{F}_{n,i}$,

$$\int_{A_{n,i}} |T_{n,i}h - f|^p d\mu \geq \Phi_{n,i}^{(p)}(f_0) \int_{\mathcal{Y}} |h - U_{n,i}f|^p d\nu, \tag{16}$$

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 Theorem 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.

Theorem 5.1 *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 (15) and (16),*

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

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 (16), 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,

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

$$\geq \sup_{V_h \in \mathcal{C}_b(\mathbf{G}),\, V_h \leq W_h} \inf_{\beta_n \in \mathcal{B}(E_{n,i},\mathbf{G})} \int_{\mathbf{H}_c} \beta_n \left(V_h, P_{h,n,i}\right) d\pi(h)$$
$$= \inf_{\beta_n \in \mathcal{B}(E_{n,i},\mathbf{G})} \int_{\mathbf{H}_c} \beta_n \left(W_h, P_{h,n,i}\right) d\pi(h),$$

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

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

This identity is proved in Strasser (1985, Theorem 46.3). The assertion of this lemma follows from Lemma 2.1 and the assumption (15). □

The next lemma gives a lower bound for the lower bound of Theorem 5.1 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 5.2 *Let $\mathbf{H}_c \subset \mathbf{H}$ be convex. Then,*

$$\inf_{\rho \in \mathcal{R}(E(\mathbf{H}_c),\mathbf{G})} \sup_{h \in \mathbf{H}_c} \int_\Omega dG_h(\omega) \int_{\mathbf{G}} \rho(\omega, da) \int_{\mathcal{Y}} |h - a|^2 d\nu$$
$$\geq \sup_{h_1, h_2 \in \mathbf{H}_c \cap L_4(\mathcal{Y},\mathcal{G},\nu)} \psi_2 \left(\frac{1}{2}\|h_1 - h_2\|_{\mathbf{H}}, 1\right),$$

where ψ_2 was defined by equation (6).

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

5.1 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

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

where $0 < C < \infty$. 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. 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 $h : \mathcal{Y} \to \mathbf{R}$, 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 Theorem 3.1 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 $\{f_{\underline{h}_n,n} \,|\, \underline{h}_n \in \mathbf{H}_c^{I_n}\} \subset \mathbf{F}(s, C, p, f_0)$ ultimately. Theorem 5.1 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).

5.2 Spherical case

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 $f : S_d \to \mathbf{R}$ will be discussed.

It is assumed that the unknown density belongs to the set

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

where $0 < C < \infty$ and $p \ge 1$. 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 $[-a,a]^d$, $a = (2\sqrt{2})^{-1}$ (this choice is explained in the proof of the next theorem), $h = h^{<1>}h^{<2>}$, $h^{<1>} : [-a,a] \to \mathbf{R}$, $h^{<2>} : [-a,a]^{d-1} \to \mathbf{R}$, $\int_{-a}^{a} h^{<1>} = 0$, $\int_{[-a,a]^{d-1}} |h^{<2>}| < \infty$, $\sup_{y\in\mathcal{Y}} |h(y)| < \infty$, $\sum_{i=1}^d \int_{\mathcal{Y}} |\partial/\partial\theta_i h(\theta)|^p \, d\theta < \infty$, and

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

Theorem 5.3 *Let* $F_n = \left(\mathcal{X}^n, \mathcal{F}^n, P_f^n : f \in \mathbf{F}(C,p,f_0)\right)$, $\mathbf{F} = L_p(\mathcal{X},\mathcal{F},\mu)$, $\mathbf{G} = L_p(\mathcal{Y},\mathcal{G},\nu)$, $W_f(a) = \int_{S_d} |f-a|^p d\mu$, *and* $\tilde{W}_h(a) = \int_{\mathcal{Y}} |h-a|^p d\nu$. *Then*

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

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 (1) 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 $\sim D\kappa_n^d$, where $D > 0$ is a constant, and let $x_{n,i} \in S_d$, $i \in I_n$, be such that the sets $A_{n,i} = \{x \in S_d \,|\, \kappa_n \phi^{(n,i)}(x) \in [-a,a]^d\}$, $i \in I_n$, are 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 $\phi^{(n,i)}(A_{n,i}) = \kappa_n^{-1}[-a,a]^d$.

Let $f_0 = \omega_d^{-1} I_{S_d}$ and 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 $\Psi(f_0) = \sqrt{f_0(x_{n,i})} = \omega_d^{-1/2}$. 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

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

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 Theorem 3.1 are satisfied, and thus $E_n(\mathbf{H}_c) \xrightarrow{w} E(\mathbf{H}_c)$. Indeed,

$$
\begin{aligned}
&\int_{S_d} \left(T_{n,k(n)}h\right) d\mu \\
&= \Psi(f_0) n^{-1/2} \kappa_n^{d/2} \int_{A_{n,k(n)}} h\left(\kappa_n \phi^{(n,k(n))}(x)\right) d\mu(x) \\
&= \Psi(f_0) n^{-1/2} \kappa_n^{d/2} \int_{\phi^{-1}(\kappa_n^{-1}[-a,a]^d)} h(\kappa_n \phi(x)) d\mu(x) \\
&= \Psi(f_0) n^{-1/2} \kappa_n^{d/2} \int_{\kappa_n^{-1}[-a,a]^d} h(\kappa_n \theta) \sin^{d-1}(\theta_d + \pi/2) \cdots \sin(\theta_2 + \pi/2) d\theta \\
&= \Psi(f_0) n^{-1/2} \kappa_n^{-d/2} \int_{[-a,a]^d} h(\theta) \sin^{d-1}(\kappa_n^{-1}\theta_d + \pi/2) \cdots \sin(\kappa_n^{-1}\theta_2 + \pi/2) d\theta \\
&= \Psi(f_0) n^{-1/2} \kappa_n^{-d/2} \int_{-a}^{a} h^{<1>}(\theta_1) d\theta_1 \\
&\quad \times \int_{[-a,a]^{d-1}} h^{<2>}(\theta_2 \ldots, \theta_d) \sin^{d-1}(\kappa_n^{-1}\theta_d + \pi/2) \cdots \sin(\kappa_n^{-1}\theta_2 + \pi/2) d\theta_2 \cdots d\theta_d \\
&= 0.
\end{aligned}
$$

Also,

$$
\begin{aligned}
&n \int_{S_d} \left(T_{n,k(n)}h_1\right) \left(T_{n,k(n)}h_2\right) f_0^{-1} d\mu \\
&= \Psi(f_0)^2 \int_{[-a,a]^d} h_1(\theta) h_2(\theta) f_0 \left(R_{x_0, x_{n,k(n)}} \phi^{-1}(\kappa_n^{-1}\theta)\right)^{-1} \\
&\quad \times \sin^{d-1}(\kappa_n^{-1}\theta_d + \pi/2) \cdots \sin(\kappa_n^{-1}\theta_2 + \pi/2) d\theta \\
&= \int_{[-a,a]^d} h_1 h_2 d\nu + o(1).
\end{aligned}
$$

Secondly, it is noted that (16) is satisfied because, by equation (2),

$$
\begin{aligned}
&\int_{A_{n,i}} |T_{n,i}h - f|^p d\mu \\
&= \left(\Psi(f_0)\kappa_n^{-2}\right)^p \int_{A_{n,i}} \left|h(\kappa_n \phi^{(n,i)}(x)) - \Psi(f_0)^{-1} \kappa_n^2 f(x)\right|^p d\mu(x) \\
&= \left(\Psi(f_0)\kappa_n^{-2}\right)^p \kappa_n^{-d} \int_{[-a,a]^d} \left|h(\theta) - \Psi(f_0)^{-1} \kappa_n^2 f\left(R_{x_0, x_{n,i}} \phi^{-1}(\kappa_n^{-1}\theta)\right)\right|^p \\
&\quad \times \sin^{d-1}(\kappa_n^{-1}\theta_d + \pi/2) \cdots \sin(\kappa_n^{-1}\theta_2 + \pi/2) d\theta \\
&\geq \delta_n \left(\Psi(f_0)\kappa_n^{-2}\right)^p \kappa_n^{-d} \int_{[-a,a]^d} \left|h(\theta) - \Psi(f_0)^{-1} \kappa_n^2 f\left(R_{x_0, x_{n,i}} \phi^{-1}(\kappa_n^{-1}\theta)\right)\right|^p d\theta,
\end{aligned}
$$

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 $\lim_{n\to\infty}\delta_n = 1$.

Thirdly, we have by choosing $\kappa_n = n^{1/(4+d)}$ that $\{f_{\underline{h}_n,n} \,|\, \underline{h}_n \in \mathbf{H}_c^{I_n}\} \subset \mathbf{F}(C,p)$ ultimately. In fact, now $n^{-1/2}\kappa_n^{d/2} = \kappa_n^{-2}$, and for $\underline{h}_n \in \mathbf{H}_c^{I_n}$, by definition (3),

$$\begin{aligned}
&\int_{S_d} |\Delta f_{\underline{h}_n,n} - |^p d\mu \\
&= \int_{S_d} \left|\sum_{i\in I_n} \Delta(T_{n,i}h_i)\right|^p d\mu = \sum_{i\in I_n}\int_{A_{n,i}} |\Delta(T_{n,i}h_i)|^p\, d\mu \\
&= \kappa_n^{-d}\sum_{i\in I_n}|\Psi(f_0)|^p \int_{[-b/2,b/2]^d}\left|\sum_{j=1}^d \eta_j(\kappa_n^{-1}\theta)\left[\left(\frac{\partial^2}{\partial\theta_j^2}h_i\right)(\theta) + \psi_j(\kappa_n^{-1}\theta)\kappa_n^{-1}\left(\frac{\partial}{\partial\theta_j}h_i\right)(\theta)\right]\right|^p \\
&\quad \times \sin^{d-1}(\kappa_n^{-1}\theta_d + \pi/2)\cdots\sin(\kappa_n^{-1}\theta_2 + \pi/2)d\theta \\
&\le C,
\end{aligned}$$

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 $j = 1,\ldots,d$ ($\eta_d(\theta) = 1$).

Lemma 5.1 can be applied with the identifications $\Phi_{n,i}^{(p)}(f_0) = \delta_n\left(\Psi(f_0)\kappa_n^{-2}\right)^p\kappa_n^{-d}$, $\Phi^{(p)}(f_0) = D\omega_d^{-p/2}$ 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 4.4.

Finally note that $x_{n,i}$, $i \in I_n$, can be chosen in such a way that $D \ge 1$. 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 $y_k^{(n,i_k)} = i_k 4a/\kappa_n$, $k \in \{1,\ldots,d\}$. Let $x_{n,i} = (y_{n,i}', 1 - \|y_{n,i}\|^2)'$. For this choice $D = 1$ and the sets $A_{n,i}$ are almost everywhere disjoint. This can be seen in the following way.

Let $A_{n,i}' = \{x \in S_d \,|\, \|\kappa_n(x - x_{n,i})\| \le 2a\}$. It can be shown that $A_{n,i} \subset A_{n,i}'$ ultimately. On the ather hand, clearly $A_{n,i}' \subset A_{n,i}''$, where $A_{n,i}'' = \{x \in S_d \,|\, (x_1,\ldots,x_d)' \in B_{n,i},\ x_{d+1} > 0\}$ and $B_{n,i} = y_{n,i} + \kappa_n^{-1}[-2a,2a]^d$ Furthermore, the sets $B_{n,i}$ are almost everywhere (μ) disjoint, and thus the sets $A_{n,i}''$ are almost everywhere (μ) disjoint. Because $A_{n,i} \subset A_{n,i}' \subset A_{n,i}''$ ultimately, the sets $A_{n,i}$ are ultimately almost everywhere (μ) disjoint. □

Acknowledgements

This article is part of the Doctoral dissertation of the author. I wish to thank my advisor Lasse Holmström for support and advise. I wish to thank the comments of two referees.

References

[1] Bretagnolle, J., and Huber, C. (1979). Estimation des densités: Risque minimax. *Z. Wahrsch. Verw. Gebiete* **47** 119–137.

[2] Cencov, N. N. (1972). *Statistical Decision Rules and Optimal Inference* Nauka, Moscow.

[3] Donoho, D. L. and Liu, R. C. (1991). Geometrizing rates of convergence III. *Ann. Statist.* **19** 668–701.

[4] Donoho, D. L., Liu, R. C., and MacGibbon, B. (1990). Minimax risk over hyperrectangles, and implications. *Ann. Statist.* **18** 1416–1437.

[5] Erfomovich, S. (2000) Sharp adaptive estimation of multivariate curves. *Math. Meth. Statist.* **9** 117-139.

[6] 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.

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

[8] Golubev, G. K. (1991). LAN in problems of nonparametric estimation of functions and lower bounds for quadratic risks. *Theory Probab. Appl.* **36** 152–157.

[9] Golubev, G. K. and Nussbaum, M. (1990). A risk bound in Sobolev class regression. *Ann. Statist.* **18** 758–778.

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

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

[12] 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.)

[13] Ibragimov, I. A. and Khas'minskii, R. Z. (1991). Asymptotically normal families of distributions and effective estimation. *Ann. Statist.* **19** 1681–1724.

[14] 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.

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

[16] Klemelä, J. (1999). Asymptotic Minimax Risk for the White Noise Model on the Sphere. *Scandinavian Journal of Statistics* **26** 465–473.

[17] Klemelä, J. (2000). Estimation of Densities and Functionals of Densities with Spherical Data. *J. Multivariate Anal.* **73** 18-40.

[18] Le Cam, L. (1972). Limits of experiments. *Proc. 6th Berkeley Symp. Math. Statist. Prob.* Vol. **1** 245–261. Univ. California Press, Berkeley.

[19] Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer, New York.

[20] Low, M. G. (1992). Renormalization and white noise approximation for non-parametric functional estimation problems. *Ann. Statist.* **20** 545–554.

[21] Low, M. G. (1993). Renormalization upper and lower bounds for integrated risk in the white noise model. *Ann. Statist.* **21** 577–589.

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

[23] Meyer, T. G. (1977). Bounds for estimation of density functions and their derivatives. *Ann. Statist.* **5** 136–142.

[24] Millar, P. (1979). Asymptotic minimax theorems for the sample distribution function. *Z. Wahrsch. Verw. Gebiete* **48** 233–252.

[25] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430.

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

[27] Samarov, A. M. (1976). Minimax bound on the risk of nonparametric density estimates. *Problems of Inform. Transmission* **12** 242–244.

[28] Stone, C. J. (1980). Optimum rate of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

[29] Strasser, H. (1985). *Mathematical Theory of Statistics.* de Gruyter, Amsterdam.

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