Minimax estimation of norms of a probability density: II. Rate-optimal estimation procedures
A Goldenshluger, Oleg V Lepski

To cite this version:
A Goldenshluger, Oleg V Lepski. Minimax estimation of norms of a probability density: II. Rate-optimal estimation procedures. Bernoulli, In press. hal-03570742

HAL Id: hal-03570742
https://hal.science/hal-03570742
Submitted on 13 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Minimax estimation of norms of a probability density: II. Rate-optimal estimation procedures

A. Goldenshluger * and O. V. Lepski †

Department of Statistics
University of Haifa
Mount Carmel
Haifa 31905, Israel
e-mail: goldensh@stat.haifa.ac.il

Institut de Mathématique de Marseille
Aix-Marseille Université
39, rue F. Joliot-Curie
13453 Marseille, France
e-mail: oleg.lepski@univ-amu.fr

Abstract: In this paper we develop rate–optimal estimation procedures in the problem of estimating the $L_p$–norm, $p \in (1, \infty)$ of a probability density from independent observations. The density is assumed to be defined on $\mathbb{R}^d$, $d \geq 1$ and to belong to a ball in the anisotropic Nikol’skii space. We adopt the minimax approach and construct rate–optimal estimators in the case of integer $p \geq 2$. We demonstrate that, depending on the parameters of the Nikol’skii class and the norm index $p$, the minimax rates of convergence may vary from inconsistency to the parametric $\sqrt{n}$–estimation. The results in this paper complement the minimax lower bounds derived in the companion paper Goldenshluger and Lepski (2020).

AMS 2000 subject classifications: 62G05, 62G20.

Keywords and phrases: density estimation, minimax risk, $L_p$-norm, $U$-statistics, anisotropic Nikol’skii class.

1. Introduction

Suppose that we observe i.i.d. random vectors $X_i \in \mathbb{R}^d$, $i = 1, \ldots, n$, with common probability density $f$. Let $p > 1$ be a given number. We are interested in estimating the $L_p$-norm of $f$,

$$
\|f\|_p := \left[ \int_{\mathbb{R}^d} |f(x)|^p dx \right]^{1/p},
$$

from observation $X^{(n)} = (X_1, \ldots, X_n)$. By estimator of $\|f\|_p$ we mean any $X^{(n)}$-measurable map $\tilde{N} : \mathbb{R}^n \rightarrow \mathbb{R}$. Accuracy of an estimator $\tilde{N}$ is measured by the quadratic risk

$$
\mathcal{R}_n[\tilde{N}, f] := \left( \mathbb{E}_f [\tilde{N} - \|f\|_p]^2 \right)^{1/2},
$$

where $\mathbb{E}_f$ denotes the expectation with respect to the probability measure $P_f$ of observations $X^{(n)} = (X_1, \ldots, X_n)$.

We adopt the minimax approach to the outlined estimation problem. The maximal risk of an estimator $\tilde{N}$ on a set $F$ of probability densities is defined by

$$
\mathcal{R}_n[\tilde{N}, F] := \sup_{f \in F} \mathcal{R}_n[\tilde{N}, f],
$$

*Supported by the ISF grant No. 361/15.
†This work has been carried out in the framework of the Labex Archimède (ANR-11-LABX-0033) and of the A*MIDEX project (ANR-11-IDEX-0001-02), funded by the "Investissements d'Avenir" French Government program managed by the French National Research Agency (ANR).
and the minimax risk is
\[ \mathcal{R}_n[\varphi] := \inf_{\hat{N}} \mathcal{R}_n[\hat{N}; \varphi], \]
where \( \inf \) is taken over all possible estimators.

In the companion paper Goldenshluger and Lepski (2020) (referred to hereafter as Part I), we derived lower bounds on the minimax risk over functional classes \( \mathcal{F} = \mathcal{N}_{\vec{r},d}(\vec{\beta}, \vec{L}) \cap \mathbb{B}_q(Q) \), where \( \mathcal{N}_{\vec{r},d}(\vec{\beta}, \vec{L}) \) stands for the anisotropic Nikol’skii class [see Definition 1 below], and \( \mathbb{B}_q(Q) := \{ f : \| f \|_q \leq Q \} \) is the ball of radius \( Q \) in \( L_q(\mathbb{R}^d) \). Specifically, we found the sequence \( \phi_n \) completely determined by parameters \( \vec{\beta}, \vec{L}, \vec{r}, q, p \) and \( n \) such that
\[ \liminf_{n \to \infty} \phi_n^{-1} \mathcal{R}_n[\mathcal{N}_{\vec{r},d}(\vec{\beta}, \vec{L}) \cap \mathbb{B}_q(Q)] \geq c > 0. \tag{1.1} \]

The goal of the present paper is to develop a rate–optimal estimator, say, \( \hat{N} \), such that for any given \( \vec{\beta}, \vec{r}, \vec{L}, q, Q \)
\[ \limsup_{n \to \infty} \phi_n^{-1} \mathcal{R}_n[\mathcal{N}_{\vec{r},d}(\vec{\beta}, \vec{L}) \cap \mathbb{B}_q(Q)] < \infty. \]

We provide an explicit construction of the rate–optimal estimator for integer values of \( p \geq 2 \).

The problem of estimating nonlinear functionals of a probability density has been studied in the literature: we refer to Part I for background and pointers to the relevant literature. Here we restrict ourselves with a brief reminder of the main definitions and results obtained in Part I.

We begin with the definition of the anisotropic Nikol’skii classes. Let \( (e_1, \ldots, e_d) \) denote the canonical basis of \( \mathbb{R}^d \). For a function \( G : \mathbb{R}^d \to \mathbb{R} \) and real number \( u \in \mathbb{R} \) the first order difference operator with step size \( u \) in the direction of the variable \( x_j \) is defined by \( \Delta_{u,j}G(x) = G(x + ue_j) - G(x) \), \( j = 1, \ldots, d \). By induction, the \( k \)-th order difference operator is
\[ \Delta_{u,j}^k G(x) = \Delta_{u,j} \Delta_{u,j}^{k-1} G(x) = \sum_{l=1}^k (-1)^{l+k} \binom{k}{l} \Delta_{u,l,j} G(x). \]

**Definition 1.** For given vectors \( \vec{\beta} = (\beta_1, \ldots, \beta_d) \in (0, \infty)^d \), \( \vec{r} = (r_1, \ldots, r_d) \in [1, \infty]^d \), and \( \vec{L} = (L_1, \ldots, L_d) \in (0, \infty)^d \) a function \( G : \mathbb{R}^d \to \mathbb{R} \) is said to belong to anisotropic Nikol’skii’s class \( \mathcal{N}_{\vec{r},d}(\vec{\beta}, \vec{L}) \) if \( \| G \|_{r_j} \leq L_j \) for all \( j = 1, \ldots, d \), and there exist natural numbers \( k_j > \beta_j \) such that
\[ \| \Delta_{u,j}^k G \|_{r_j} \leq L_j |u|^{|\beta_j|}, \quad \forall u \in \mathbb{R}, \quad \forall j = 1, \ldots, d. \]

The important quantities that are related to Nikol’skii’s classes and determine asymptotics of the minimax risk are the following:
\[ \frac{1}{\beta} := \sum_{j=1}^d \frac{1}{\beta_j}, \quad \frac{1}{\omega} := \sum_{j=1}^d \frac{1}{\beta_j r_j}, \quad L_\beta := \prod_{j=1}^d L_j^{\frac{1}{\beta_j}}, \]
\[ \tau(s) := 1 - \frac{1}{\omega} + \frac{1}{\beta s}, \quad s \in [1, \infty]. \]

It is worth mentioning that \( \tau(\cdot) \) appears in embedding theorems for Nikol’skii’s spaces; see Section 5.1.3 for details.

In Part I we have established a lower bound of type (1.1) on the minimax risk with the
sequence $\phi_n$ defined as follows

$$\theta = \begin{cases} 
\frac{1}{\tau(1)}, & \tau(p) \geq 1; \\
\frac{1/p-1/q}{1-1/q-(1-1/p)\tau(q)}, & \tau(p) < 1, \tau(q) < 0; \\
\frac{\tau(p)}{\tau(1)}, & \tau(p) < 1, \tau(q) \geq 0;
\end{cases}$$

$$\phi_n = \phi_n(\tilde{\beta}, \tilde{r}, \bar{L}) =: L_{\tilde{\beta}(\tilde{r}, \bar{L})}^{1-1/p} n^{-\theta^*}, \quad \theta^* = 2^{-1} \land \theta; \tag{1.2}$$

see Theorem 1 of Part I. Our goal is to develop a rate–optimal estimator whose risk converges to zero at the rate $\phi_n$.

2. Estimator construction

Assuming that $p$ is an integer number let us first discuss the problem of estimating a closely related functional $\|f\|^p_p$.

Let $K : [-1, 1]^d \to \mathbb{R}$ be a given function (kernel), and let $h = (h_1, \ldots, h_d) \in (0, 1]^d$ be a given vector (bandwidth). Let

$$K_h(x) := (1/V_h)K(x/h), \quad V_h := \prod_{k=1}^d h_k;$$

here and in all what follows the division $y/x$ for vectors $x, y \in \mathbb{R}^d$ is understood in the coordinate–wise sense. Define

$$S_h(x) := \int K_h(x-y)f(y)dy, \quad B_h(x) := S_h(x) - f(x). \tag{2.1}$$

Obviously, $B_h(x)$ is the bias of the kernel density estimator of $f(x)$ associated with kernel $K$ and bandwidth $h$.

The construction of our estimator for $\|f\|^p_p$ is based on the following simple observation formulated below as Lemma 1.

**Lemma 1.** For any $p \in \mathbb{N}^*, p \geq 2$, $h \in (0, \infty)^d$ and any probability density $f$ one has

$$\|f\|^p_p = (1 - p) \int S_h^{p}(x)dx + p \int S_h^{p-1}(x)f(x)dx$$

$$+ \sum_{j=2}^{p} \binom{p}{j} (-1)^{j-1} \int [S_h(x)]^{p-j} B_h^j(x)dx. \tag{2.2}$$

The proof of the lemma is elementary and given in Appendix. It does not require any assumption on the kernel $K$ except of existence of the integrals on the right hand side of (2.2).

Let $\hat{T}_{1,h}$ and $\hat{T}_{2,h}$ be estimators of

$$T_{p}^{(1)}(f) := \int S_h^{p}(x)dx, \quad T_{p}^{(2)}(f) := \int S_h^{p-1}(x)f(x)dx,$$

respectively. Then we estimate $\|f\|^p_p$ by

$$\hat{T}_h = (1 - p)\hat{T}_{1,h} + p \hat{T}_{2,h}.$$
Note that if $\hat{T}_{1,h}$ and $\hat{T}_{2,h}$ are unbiased estimators of $T_p^{(1)}(f)$ and $T_p^{(2)}(f)$ then, in view of Lemma 1, the bias of $\hat{T}_h$ in estimation of $\|f\|_p$ is

$$\sum_{j=2}^{p} \binom{p}{j} (-1)^j \int |S_h(x)|^{p-j} B_j^h(x) dx.$$ 

The last quantity can be efficiently bounded from above via norms of the bias $B_h(\cdot)$ and the underlying density $f$.

The natural unbiased estimators for $T_p^{(1)}(f)$ and $T_p^{(2)}(f)$ are based on the U-statistics:

$$\hat{T}_{1,h} := \frac{1}{\binom{n}{p}} \sum_{i_1, \ldots, i_p} U_h^{(1)}(X_{i_1}, \ldots, X_{i_p}),$$

$$\hat{T}_{2,h} := \frac{1}{\binom{n}{p}} \sum_{i_1, \ldots, i_p} U_h^{(2)}(X_{i_1}, \ldots, X_{i_p}),$$

where the summations are taken over all possible combinations of $p$ distinct elements $\{i_1, \ldots, i_p\}$ of $\{1, \ldots, n\}$, and

$$U_h^{(1)}(x_1, \ldots, x_p) := \int K_h(y - x_1) \cdots K_h(y - x_p) dy,$$

$$U_h^{(2)}(x_1, \ldots, x_p) := \frac{1}{p} \prod_{i=1}^{p} K_h(x_i - x_i).$$

It is worth mentioning that not only there is the explicit formula for the bias of $\hat{T}_h$, but also its variance admits a rather simple analytical bound. The following result states an upper bound on the variance of $\hat{T}_h$; the proof is given in Appendix.

**Lemma 2.** Let $K$ be symmetric bounded function supported on $[-1, 1]^d$. Then for all $p \in \mathbb{N}^*, p \geq 2$, $h \in (0, \infty)^d$ and any probability density $f$ one has

$$\text{var}_f[\hat{T}_h] \leq C \|K\|_\infty^2 \sum_{k=1}^{p} \|f\|^{2p-k}_{2p-k} (n^{k} V_{h}^{k-1})^{-1},$$

where $C$ is a constant depending on $p$ and $d$ only.

If $\hat{T}_h$ is a "reasonable" estimator of $\|f\|_p$ then it is natural to define

$$\hat{N}_h := |\hat{T}_h|^{1/p}$$

as an estimator for $\|f\|_p$.

### 3. Main result

In this section we demonstrate that the estimator $\hat{N}_h$ with properly chosen bandwidth $h$ is a rate-optimal estimator for $\|f\|_p$, provided that

$$p \in [1,p]^d \cup [p, \infty)^d, \quad p \in \mathbb{N}^*, \quad p \geq 2,$$
i.e., \( r_j \leq p \) for all \( j = 1, \ldots, d \), or \( r_j \geq p \) for all \( j = 1, \ldots, d \). The proof is based on the derivation of tight uniform upper bounds on the bias and the variance of \( \hat{T}_h \) over anisotropic Nikolskii’s classes. To get such bounds for the bias term we use a special construction of kernel \( K \) [see, e.g., Goldenshluger and Lepski (2014)].

Let \( K : \mathbb{R} \to \mathbb{R} \) be a function supported on \([-1, 1]\), symmetric and satisfying \( \int_{\mathbb{R}} K(y)dy = 1 \), and \( \|K\|_\infty < \infty \). For a given positive integer number \( \ell \in \mathbb{N}^* \) define

\[
K_\ell(y) = \sum_{i=1}^{\ell} \binom{\ell}{i} (-1)^{i+1} i^{-1} K(y/i).
\]

**Assumption 1.** The kernel \( K \) involved in the construction of the estimator \( \hat{N}_h \) is given by

\[
K(x) = \prod_{j=1}^{d} K_\ell(x_j), \quad \forall x \in \mathbb{R}^d,
\]

with \( \ell > \max_{j=1, \ldots, d} \beta_j \).

Let us introduce the following notation. For \( j = 1, \ldots, d \) let

\[
x_j := \begin{cases} \frac{\beta_j \tau(p)}{\beta_j}, & r_j \leq p, \tau(q) > 0; \\ \beta_j, & \text{otherwise}, \end{cases}
\]

\[
p_j := \begin{cases} \frac{2(1-p)}{1-p/r_j}, & r_j \geq p; \\ 2, & r_j \leq p, \tau(q) > 0; \\ \frac{2(1-p/q)}{1/p_j - 1/q}, & r_j < p, \tau(q) \leq 0,
\end{cases}
\]

and let

\[
\frac{1}{\upsilon} := \sum_{j=1}^{d} \frac{1}{p_j x_j}.
\]

Define \( \mathfrak{h} = (h_1, \ldots, h_d) \) by

\[
h_j := L_j^{-1/\upsilon_j} \left( \frac{2}{\upsilon_{\ell\beta_j}(\mathfrak{h})} \frac{2}{p_j x_j + \frac{2}{1/p_j - 1/q}} \right),
\]

where \( \mathfrak{L} \) is a constant that is completely determined by the class parameters \( \bar{\beta}, \bar{L}, \bar{r} \) and \( p \) (a cumbersome but explicit expression for \( \mathfrak{L} \) is given in Section 5.3.3).

The main result of this paper is given in the next theorem.

**Theorem.** Let \( p \in \mathbb{N}^*, p \geq 2, Q > 0, \bar{\beta} \in (0, \infty)^d, \bar{L} \in (0, \infty)^d, \bar{r} \in [1, p]^d \cup [p, \infty]^d \) and \( q \geq 2p - 1 \) be fixed, and let Assumption 1 hold. Let \( \hat{N}_h \) be the estimator (2.3) associated with bandwidth \( \mathfrak{h} \) defined in (3.2); then

\[
\limsup_{n \to \infty} \phi_n^{-1} \mathcal{R}_n \left[ \hat{N}_h, N_{\mathfrak{r} \bar{r}}(\bar{\beta}, \bar{L}) \cap B_q(Q) \right] < \infty,
\]

where \( \phi_n \) is given in (1.2).

**Remark 1.**

(i) Combining the result of this theorem with that of Theorem 1 in Part I we conclude that the suggested estimator \( \hat{N}_h \) is rate-optimal. Thus, the problem of constructing a rate–optimal estimator is solved completely if \( r_j \leq p \) for all \( j = 1, \ldots, d \), or \( r_j \geq p \) for all \( j = 1, \ldots, d \). In particular the obtained result is definitive in the one–dimensional case or, more generally, in the semi-isotropic case when \( r_j = r, \forall j = 1, \ldots, d, \) for some \( r \in [1, \infty] \).
(ii) It is surprising that the $L_p$-norm can be estimated with parametric rate $n^{-1/2}$. To the best of our knowledge this phenomenon has not been observed in the literature. Note that the parametric regime is possible only if $\tau(q) > 0$. Indeed,

$$\frac{1/p - 1/q}{1 - 1/q - (1 - 1/p)\tau(q)} \geq \frac{1}{2} \iff \frac{2}{p} - 1 - \frac{1}{q} \geq -(1 - \frac{1}{p})\tau(q),$$

and the last inequality is impossible if $\tau(q) \leq 0$ because $p \geq 2$.

(iii) A particularly simple description of the minimax rate of convergence $\phi_n$ is obtained in the specific case $p = 2$ and $q = \infty$. Here $\theta^* = (\max\{\tau(1), 2\})^{-1}$ when $\vec{r} \in [2, \infty]^d$, and

$$\theta^* = \begin{cases} \frac{1}{1+\frac{1}{\omega}}, & \omega > 1; \\ \frac{1}{\omega}, & \omega \leq 1, \end{cases}$$

when $\vec{r} \in [1, 2]^d$. As we see, the regime corresponding to the rate of convergence $n^{-\tau(p)/\tau(1)}$ does not appear when $p = 2, q = \infty$.

4. Discussion

In this section we compare and contrast our results with existing results in the literature. In the first two paragraphs we discuss the problem of estimating the $L_p$–norm with integer $p$. The fundamental differences between our results and other results are mainly due to the following two reasons.

• Statistical model. We argue that when estimating the $L_p$–norm estimation in the density model, much better accuracy can be achieved than in the regression/Gaussian white noise model.

• Domain of observations. We explain why estimation accuracy is completely different for compactly supported densities $f$, and densities $f$ supported on the entire space $\mathbb{R}^d$.

We also discuss estimation of the $L_p$-norm for non–integer values of $p$. The goal here is to understand how optimal are the risk bounds of Theorems 2 and 3 established in the companion paper Goldenshluger and Lepski (2020). We will show that for non–integer values of $p$ in some cases straightforward plug-in constructions lead to nearly rate–optimal adaptive estimators of the $L_p$-norms. We finish this section with several remarks about adaptive estimation.

4.1. Norm estimation in density and Gaussian white noise models

To the best of our knowledge, in the context of the Gaussian white noise model, estimation of the $L_p$-norm was studied only in the one–dimensional case, $d = 1$. So we focus on the one–dimensional Nikolskii class that is denoted $N_{r,1}(\beta, L)$, $\beta > 0$, $L > 0$, and $r \in [1, \infty]$.

Let $W$ be the standard Wiener process and assume that we observe the trajectory of the process

$$X_n(t) = \int_0^t f(u)du + n^{-1/2}W(t), \quad t \in [0, 1].$$

(4.1)

The estimation of $\|f\|_p$, $1 \leq p < \infty$ in the Gaussian white noise model (4.1) was initiated in Lepski et al. (1999) under assumption that $f$ belongs to Hölder’s functional class, i.e., $f \in N_{\infty,1}(\beta, L)$. 
The recent paper Han et al. (2020) considers the same problem for Nikolskii’s classes with $r \in [p, \infty)$. In the Gaussian white noise model (4.1) in the case of integer $p$ there is a significant difference in the minimax rates of convergence obtained for odd and even values of $p$. Such phenomenon does not occur in the density model because density is a positive function. Thus in the subsequent discussion of the results in the Gaussian white noise model we restrict ourselves with the case of even $p$.

According to the results obtained in Lepski et al. (1999) and Han et al. (2020) in the case of even $p$, the minimax rate of convergence (expressed in our notation) is given by $\varphi_n = n^{-\frac{1}{2+2r}}$. Note that in both papers $r \geq p$ and, therefore,

$$\varphi_n = n^{-\frac{1}{2+2r}} \gg n^{-\frac{1}{p}} = \phi_n,$$

where $\phi_n$ is defined in (1.2). Also, since $\tau(1) > 1$ for any $r \neq 1$ we conclude that the parametric regime is impossible in the model (4.1). Before providing the explanations why the results discussed above are so different, let us discuss another result.

Lepski and Spokoiny (1999) studied a hypothesis testing problem in the model (4.1) when the set of alternatives consists of functions from $\mathbb{N}_{r,1}(\beta, L)$, $r \in [1, 2)$ separated away from zero in the $L_2$-norm. It is well known that this problem is equivalent to the problem of estimating the $L_2$-norm, and the minimax rate of testing coincides with the rate of estimation. The minimax rate found in Lepski and Spokoiny (1999) is $\varphi_n = n^{-\frac{1}{p+2}}$ under assumption $\tau(\infty) > 0$. Noting that $\tau(\infty) > 0$ implies $q = \infty$ and comparing this result with the result obtained in this paper for the case $p = 2$, $r \in [1, 2)$ and $\tau(\infty) > 0$ we conclude that

$$\varphi_n = n^{-\frac{1}{2+2r}} \gg n^{-1/2} = \phi_n.$$

The last inequality follows from the fact that $\tau(1) > \tau(2)$. Thus we again conclude that the parametric regime is impossible in the model (4.1). Another interesting feature should be mentioned: the approach used in Lepski and Spokoiny (1999) is based on a pointwise bandwidth selection scheme while our estimation procedure is based on a fixed bandwidth.

The explanation why estimation accuracy in the density model is better than in the Gaussian white noise model (4.1) is rather simple. In fact, the maximal value of the risk in the density model is attained on the densities with small $L_p$-norm. Analysis of our estimation strategy shows that in the density model the variance of the corresponding $U$-statistics is proportional to the $\mathbb{L}_p$-norm of the density (see Lemma 2 and Proposition 4 for details). The smaller this norm, the smaller the stochastic error of the estimation procedure, and careful bandwidth selection employed in our procedure takes into account this fact and improves considerably the estimation accuracy. In contrast, in the Gaussian white noise model (4.1) the stochastic error of the estimation procedure is independent of the signal $f$ and of the value of its $\mathbb{L}_p$-norm.

### 4.2. Norm estimation for compactly and non-compactly supported densities

We start with the following simple observation. If a probability density $f$ is defined on a compact set $\mathcal{I} \subset \mathbb{R}^d$ then necessarily $\|f\|_p \geq |\mathcal{I}|^{1-1/p} > 0$ for any $p \in (1, \infty]$. This fact reduces estimation of $\|f\|_p$ for compactly supported densities to the problem of estimating $\|f\|_p^p$, which is a much smoother functional. Indeed, let $\tilde{N}$ be an estimator of $\|f\|_p^p$. Then, for any density $f$ and any $p \in \mathbb{N}^*, p \geq 2$

$$\|\tilde{N}^{1/p} - \|f\|_p\| \leq \|f\|_p^{-(p-1)}\|\tilde{N} - \|f\|_p^p\| \leq |\mathcal{I}|^{1-1/p} |\tilde{N} - \|f\|_p|.$$
The problem of estimating \( \|f\|_p \) has been considered by many authors starting from the seminal work of Bickel and Ritov (1988); see, for instance, Birgé and Massart (1995), Kerkyacharian and Picard (1996), Laurent (1997), Cai and Low (2005), Tchetgen et al. (2008) among others. It is worth noting that majority of the papers deal with compactly supported densities from a semi–isotropic functional class, that is \( r_l = r \) for any \( l = 1, \ldots, d \). Below we discuss these results and present them in a unified way.

Let \( p \in \mathbb{N}^*, p \geq 2 \) be fixed, and let \( \mathbb{N}_{p,d}(\tilde{\beta}, \tilde{L}, \mathcal{I}) \) denote a semi-isotropic Nikol’skii class of densities supported on \( \mathcal{I} \). Assume that \( r \geq p \). Then the minimax rate of convergence in estimating \( \|f\|_p \) on this functional class, and, therefore, of \( \|f\|_p \) as well, is given by

\[
\varphi_n = n^{-\left(\frac{1}{2} + \frac{1}{r} \wedge \frac{1}{2}\right)}.
\]

Hence, taking into account that \( \tau(1) = 1 - 1/(r\beta) + 1/\beta \leq 1 + 1/(2\beta) \) for any \( r \geq p \geq 2 \) we conclude that

\[
\varphi_n = n^{-\left(\frac{1}{2} + \frac{1}{r} \wedge \frac{1}{2}\right)} \ll n^{-\left(\frac{1}{2} + \frac{1}{2r} \wedge \frac{1}{2}\right)} = \phi_n.
\]

In particular, the parametric regime in estimating compactly supported densities is possible if and only if \( \beta \geq 1/4 \) which is a less restrictive condition than \( \tau(1) \leq 2 \).

### 4.3.Norm estimation for non–integer values of \( p \)

As it was mentioned above, for non–integer values of \( p \) in some cases straightforward plug–in constructions lead to nearly rate–optimal adaptive estimators of \( \|f\|_p \)-norms. We discuss two such specific cases below.

\(^{1}\). Assume that density \( f \) is uniformly bounded, i.e., \( q = \infty \), and let \( p \notin \mathbb{N}^*, p > 1 \) be fixed. Consider the following sets of parameters:

\[
\mathcal{D}_1 = \{ (\tilde{\beta}, \tilde{r}) : \tau(p) > 2(1 - 1/p) \}; \\
\mathcal{D}_2 = \{ (\tilde{\beta}, \tilde{r}) : \tau(p) < 1 - 2/p, \tau(\infty) < 0 \}.
\]

Let \( \ell > 0 \) be an arbitrary a priori chosen integer number, and let

\[
\varphi_n := \left\{ \begin{array}{ll}
(L \ln(n)/n)^{\frac{1-1/p}{d}} \ln(n)^{d-1}, & (\tilde{\beta}, \tilde{r}) \in \mathcal{D}_1; \\
(L \ln(n)/n)^{\omega/p}, & (\tilde{\beta}, \tilde{r}) \in \mathcal{D}_2.
\end{array} \right.
\]

Let \( \hat{f}(x), x \in \mathbb{R}^d \) be the estimator of \( f(x) \) built in Theorem 1 of Lepski and Willer (2019) in the case \( \alpha = 0 \) [see also Goldenshluger and Lepski (2014)], and consider the plug–in estimator of the \( \|f\|_p \)-norm, \( \hat{F} := \|\hat{f}\|_p \).

**Proposition 1.** For any \( Q > 0, L_0 > 0, \ell \geq 1, \tilde{L} \in [L_0, \infty)^d \) and any \( \tilde{\beta} \in (0, \ell)^d, \tilde{r} \in (1, \infty)^d \) belonging to \( \mathcal{D}_1 \cup \mathcal{D}_2 \) there exists \( C < \infty \), independent of \( \tilde{L} \), such that

\[
\limsup_{n \to \infty} \varphi_n^{-1} \mathcal{R}_n[\hat{F}; \mathbb{N}_{p,d}(\tilde{\beta}, \tilde{L}) \cap B_\infty(Q)] \leq C.
\]

The proof of this proposition is trivial. By the triangle inequality \( \|\hat{F} - f\|_p \leq \|\hat{f} - f\|_p \), so that the problem of estimating \( \|f\|_p \) can be reduced to the problem of adaptive estimation of \( f \) under the \( \|\cdot\|_p \)-loss. The stated upper bound follows from the results of Theorem 3 in Lepski and Willer (2019) corresponding to what is called in that paper tail zone and sparse zone 1. Combining bounds of Theorem 3 in Goldenshluger and Lepski (2020) and Proposition 1 we come to the following statement.
Corollary 1. For any $Q > 0$, $L_0 > 0$, $\ell > 0$, $\tilde{L} \in [L_0, \infty)^d$ and any $\tilde{\beta} \in (0, \ell]^d$, $\tilde{r} \in (1, \infty)^d$ belonging to $\mathcal{D}_1 \cup \mathcal{D}_2$ one has for all $n$ large enough

$$[\ln(n)]^{-\upsilon} \lesssim \left(\frac{\ln(n)}{n}\right)^{\theta} [N_{\mathcal{F}, d}(\tilde{\beta}, \tilde{L}) \cap B_\infty(Q)] \lesssim [\ln(n)]^{\gamma},$$

where $\gamma = d - 1$ on $\mathcal{D}_1$ and $\gamma = 0$ on $\mathcal{D}_2$, $\upsilon = 3\theta$ on $\mathcal{D}_1$ and $\upsilon = 2\theta$ on $\mathcal{D}_2$, and parameter $\theta$ is defined in Theorem 2 in Goldenshluger and Lepski (2020).

Thus, estimator $\hat{F}$ is nearly rate–optimal adaptive over a scale of Nikolskii’s classes with parameters belonging $\mathcal{D}_1 \cup \mathcal{D}_2$.

Proposition 2. Let $p \in (1, 2)$ be fixed and assume that $\tilde{r} = (p, \ldots, p)$. For any $L_0 > 0, \ell > 0$, $\tilde{L} \in [L_0, \infty)^d$ and $\tilde{\beta} \in (0, \ell]^d$ there exists $C < \infty$, independent of $\tilde{L}$, such that

$$\limsup_{n \to \infty} \left[ L^{-1} n \frac{[N_{\mathcal{F}, d}(\tilde{\beta}, \tilde{L})]}{[N_{\mathcal{F}, d}(\tilde{\beta}, \tilde{L})]} \right] \leq C.$$

The proposition follows from Theorem 4 in Goldenshluger and Lepski (2011) and the triangle inequality. Note that under conditions of Theorem 2 we have $\tau(p) = 1 + 1/p$. Also in the case $p \in (1, 2)$ the parametric rate is impossible, and in Theorem 2 of $\vartheta = (1 - 1/p)/\tau(1)$. Therefore comparing the results of Theorems and 2 we conclude that the lower and upper bounds differ by factor $(\frac{\ln\ln(n)}{\ln^2(n)})^{\vartheta}$.

4.4. Open problems

We close this section with several remarks about adaptive estimation and other open problems.

The case of integer $p$. Minimax estimation. Recall that the proposed estimator is rate-optimal under additional assumption $\tilde{r} \in [1, p]^d \cup [p, \infty]^d$. Construction of rate–optimal estimators in the general setting when coordinates of vector $\tilde{r}$ can be arbitrary with respect to $p$ remains an open problem. It is not difficult to derive an explicit upper bound on the risk of our estimator in the general setting whenever $\tilde{r} \in [1, \infty]^d$. However the obtained bound does not match the lower bound on the minimax risk found in Theorem 1 in Part I. We conjecture that the lower bound is tight, and in the general setting a rate–optimal estimator does not belong to the family of estimators $\{\hat{N}_h, h \in \mathbb{R}^d\}$, and a different estimation procedure has to be developed. Another interesting question is whether condition $q \geq 2p - 1$ is necessary in order to guarantee the obtained estimation accuracy.

The case of integer $p$. Adaptive estimation. The bandwidth choice employed in the proposed estimation procedure requires prior knowledge of the class parameters $\tilde{\beta}, \tilde{r}$ and $\tilde{L}$. A natural question of adaptive estimation arises:

Is it possible to construct a single estimator for $\|f\|_p$ whose maximal risk is asymptotically proportional to $\phi_n(\tilde{\beta}, \tilde{r}, \tilde{L})$ simultaneously for all values of $\tilde{\beta}, \tilde{r}$ and $\tilde{L}$?

Recall that such optimally-adaptive estimators do not always exist; see Lepskii (1992). We conjecture that optimally-adaptive estimators of the $\mathbb{L}_p$-norm of a density do not exist, and there
is a price to pay for adaptation. We will not present here the exact definition of optimality but conjecture that the attainable family of risk normalizations is given by
\[ \Phi = \{ \phi_{\ln(n)}(\beta, \vec{r}, \vec{L}) \}_{\beta, \vec{r}, \vec{L}}, \]
where sequence \( \{\phi_n\} \) is defined in (1.2). We also conjecture that in the one-dimensional case \( d = 1 \) the estimator that achieves \( \Phi \) can be constructed using selection from the family \( \{\hat{N}_h\}_h \) by the original method due to Lepskii (1991). The important technical tools here are the concentration inequalities for \( U \)-statistics, see e.g. Houdré and Reynaud-Bouret (2003), Major (2013). The situation changes considerably in the multivariate case when adaptation over a collection of anisotropic Nikolskii classes is studied. Here it is not clear if one of the existing approaches to adaptive estimation is applicable. The construction of an adaptive procedure in estimating of \( L_p \)-norm of a density possessing anisotropic and inhomogeneous smoothness is the most challenging problem in this subject.

The case of non-integer \( p \). The main open problem in estimating of the \( L_p \)-norm when for non–integer values of \( p \not\in \mathbb{N}^* \) consists in confirming the asymptotics of the minimax risk found in Theorems 2 and 3 of Part I. We believe that the approach developed recently in Han et al. (2020) in the framework of Gaussian white noise model can be adapted to the density model.

5. Proofs

5.1. Preliminaries

In this section we collect known facts from functional analysis as well as some recent results related to anisotropic Nikolskii’s spaces. These results will be used in the subsequent proofs.

5.1.1. Strong maximal operator

For locally integrable function \( f \) on \( \mathbb{R}^d \) the strong Hardy–Littlewood maximal operator of \( f \) is defined by
\[ M[f](x) = \sup \left\{ \frac{1}{|I|} \int_I f(y)dy : x \in I \right\}, \]
where the supremum is taken over all rectangles \( I \) with edges parallel to the coordinate axes and containing point \( x \); here \( |\cdot| \) stands for the Lebesgue measure. In view of the Lebesgue differentiation theorem
\[ f(x) \leq M[f](x) \quad \text{a.e.} \quad (5.1) \]
Moreover, if \( f \in L_p(\mathbb{R}^d) \) then
\[ \|M[f]\|_p \leq c_0\|f\|_p, \quad 1 < p \leq \infty \quad (5.2) \]
with constant \( c_0 \) depending on \( p \) and \( d \) only; see, e.g., Guzman (1975).

5.1.2. Some useful inequalities

For the ease of reference we recall some well known inequalities that are routinely used in the sequel. These results can be found, e.g., in Folland (1999).
• **Interpolation inequality.** Let $1 \leq s_0 < s < s_1 \leq \infty$. If $f \in L_{s_0}(\mathbb{R}^d) \cap L_{s_1}(\mathbb{R}^d)$ then $f \in L_s(\mathbb{R}^d)$ and
\[
\|f\|_s \leq (\|f\|_{s_0})^{(r_1-s_1)/(r_1-s_0)} (\|f\|_{s_1})^{(r-s_1)/(r-s_0)}.
\]

• **Young’s inequality (general form).** Let Assumption 1 hold. Then $\|f\|_p \leq \|f\|_r \leq \|f\|_q$.

#### 5.1.3. Some facts related to Nikolskii’s classes

Let $s > 1$, and let the functional class $\mathbb{N}_{r,d}(\vec{\beta}, \vec{L})$ be fixed. Define for any $j = 1, \ldots, d$
\[
\gamma_j(s) := \begin{cases} 
\beta_j \tau(s)/\tau(r_j), & r_j < s; \\
\beta_j, & r_j \geq s,
\end{cases}
\]
and let
\[
s^* := s \vee \left[ \max_{j=1,\ldots,d} r_j \right], \quad \vec{s} := (r_1 \vee s, \ldots, r_d \vee s).
\]
Recall that the bias $B_h(\cdot) = B_h(\cdot, f)$ of a kernel density estimator associated with kernel $K$ is defined in (2.1). The following result has been proved in Goldenshluger and Lepski (2014).

**Lemma 3.** Let Assumption 1 hold. Then $B_h(x)$ admits for any $x \in \mathbb{R}^d$ the representation $B_h(x) = \sum_{j=1}^d B_h^{(j)}(x)$ with functions $B_h^{(j)}$ satisfying the following inequalities. There exist $C_1 > 0$ and $C_2 > 0$ independent of $\vec{L}$ such that for any $f \in \mathbb{N}_{r,d}(\vec{\beta}, \vec{L})$ and $h \in (0, \infty)^d$
\[
\|B_h^{(j)}\|_{r_j} \leq C_1 L_j h_j^{\gamma_j(s)} , \; \forall j = 1, \ldots, d. \tag{5.3}
\]

Moreover, for any $s > 1$ satisfying $\tau(s^*) > 0$ one has
\[
\|B_h^{(j)}\|_{s^*} \leq C_2 L_j h_j^{\gamma_j(s)} , \; \forall j = 1, \ldots, d. \tag{5.4}
\]

Finally, for any $r \in [1, \infty]$ and $R > 0$ there exists $C_3 > 0$ such that for any $f \in \mathbb{B}_r(R)$ and any $h \in (0, \infty)^d$
\[
\|B_h^{(j)}\|_r \leq C_3, \; \forall j = 1, \ldots, d. \tag{5.5}
\]

The next lemma presents an inequality between different norms of a function belonging to anisotropic Nikolskii’s class. The proof of this lemma is given in Appendix.

**Lemma 4.** Let $1 \leq p < \infty$ be fixed. For any $s \in (1, \infty]$ satisfying $s \geq p \vee \max_{j=1,\ldots,d} r_j$ and $\tau(s) > 0$ there exists constant $C > 0$ independent of $\vec{L}$ such that for any $f \in \mathbb{N}_{r,d}(\vec{\beta}, \vec{L})$
\[
\|f\|_s \leq C (L_{\gamma(s)})^{\left(\frac{1}{\tau(s)} - \frac{1}{\tau(p)}\right) \frac{s}{p}} (\|f\|_p)^{s} \|f\|_p^{\frac{s}{p}}, \; L_{\gamma(s)} := \prod_{j=1}^d L_j^{\gamma_j(s)}.
\]
5.2. Reduction to the risk of \( \hat{T}_h \)

We will analyze the risk of the proposed estimator \( \hat{N}_h \) using two different upper bounds that relate the risk of \( \hat{N}_h \) to the risk of \( \hat{T}_h \) via elementary inequalities:

\[
\mathbb{E}_f |\hat{N}_h - \|f\|_p|^2 = \mathbb{E}_f |\hat{T}_h|^{1/p} - \|f\|_p|^2 \leq \mathbb{E}_f |\hat{T}_h - \|f\|_p^p|^{2/p} \\
\leq \left[ \mathbb{E}_f (\hat{T}_h - \|f\|_p^p)^2 \right]^{1/p} ; \\
\mathbb{E}_f |\hat{N}_h - \|f\|_p|^2 = \mathbb{E}_f \left( \frac{\hat{N}_h^p - \|f\|_p^p}{\sum_{i=0}^{p-1} \hat{N}_h^i \|f\|_p^{p-i-1}} \right)^2 \leq \frac{\mathbb{E}_f (\hat{T}_h - \|f\|_p^p)^2}{\|f\|_p^p}.
\]

Thus, for any underlying density \( f \) and any \( p \in \mathbb{N}^*, p \geq 2 \)

\[
\mathbb{E}_f |\hat{N}_h - \|f\|_p|^2 \leq \min \left\{ \left[ \mathbb{E}_f (\hat{T}_h - \|f\|_p^p)^2 \right]^{1/p} , \frac{\mathbb{E}_f (\hat{T}_h - \|f\|_p^p)^2}{\|f\|_p^p} \right\}.
\] (5.6)

5.3. Proof of Theorem

The proof is divided in three steps. First we establish upper bounds (uniform over \( N_{r,d}(\vec{\beta}, \vec{L}) \cap B_q(Q) \)) on the bias and the variance of the estimator \( \hat{T}_h \) for any value of \( h \). Then using the derived upper bounds on the risk of \( \hat{T}_h \) and the risk reduction argument given in Section 5.2, we complete the proof.

5.3.1. Bound for the bias

The next proposition states the upper bound on the bias of the estimator \( \hat{T}_h \).

**Proposition 3.** Let \( \vec{r} \in [1, p]^d \cup [p, \infty]^d \). Then there exists \( C > 0 \) independent of \( \vec{L} \) such that for any \( f \in N_{r,d}(\vec{\beta}, \vec{L}) \cap B_q(Q) \) and any \( h \in (0, \infty)^d \)

\[
\mathbb{E}_f |\hat{T}_h - \|f\|_p| \leq C \|f\|_p^{p-2} \sum_{j=1}^d L_j^{p_j} b_j^{p_j \kappa_j},
\]

where \( p_j \) and \( \kappa_j \) are defined in (3.1).

**Proof** Our first objective is to prove that

\[
|\mathbb{E}_f (\hat{T}_h) - \|f\|_p| \leq c_1 \|f\|_p^{p-2} \|B_h\|_p^2, \quad \forall h \in (0, \infty)^d.
\] (5.7)

If \( p = 2 \) then, by Lemma 1, (5.7) holds as equality with \( c_1 = 1 \). If \( p \geq 3 \) then we have in view of Lemma 1

\[
|\mathbb{E}_f (\hat{T}_h) - \|f\|_p| \leq c_2 \sum_{k=2}^p \int |B_h(x)|^k |S_h(x)|^{p-k} dx.
\] (5.8)

Note that for any \( h \in (0, \infty)^d \) in view of (5.1)

\[
|B_h(x)| \leq |S_h(x)| + f(x) = K \|f\|_1 \mathcal{M}[f](x) + f(x) \leq c_4 \mathcal{M}[f](x) \text{ a.e.}
\]
It yields together with (5.8)

$$\|E_f(\hat{T}_h) - \|f\|^p\| \leq c_4 \int |B_h(x)|^2 (\mathcal{W}[f](x))^{p-2} \, dx.$$ 

Hence, applying Hölder’s inequality with exponents $p/2$ and $p/(p - 2)$ and (5.2) we obtain

$$\|E_f(\hat{T}_h) - \|f\|^p\| \leq c_4 \|\mathcal{W}[f]\|_p^{p-2} \|B_h\|_p^2 \leq c_5 \|f\|^{p-2} \|B_h\|_p^2.$$ 

This completes the proof of (5.7). We deduce from (5.7) and Lemma 3

$$\|E_f(\hat{T}_h) - \|f\|^p\| \leq c_6 \|f\|^{p-2} \sum_{j=1}^d \|B_h^{(j)}\|_p^2, \quad \forall h \in (0, \infty)^d. \quad (5.9)$$

Consider now separately three cases. Let $\vec{r} \in [p, \infty]^d$. Then, for any $j = 1, \ldots, d$, applying the interpolation inequality with $s_0 = 1$, $s = p$ and $s_1 = r_j$, and (5.3) and (5.5) of Lemma 3 we obtain

$$\|B_h^{(j)}\|_p^2 \leq c_7 \|B_h^{(j)}\|_{r_j}^{2(1-(1/p)r_j)} \leq c_9 (L_j h_j^{(p)}) = c_8 L_j h_j^{(p)} h_j^{\kappa(p)}; \quad (5.10)$$

for any $f \in \mathcal{N}_{r,d}(\vec{r}, \vec{L})$.

Let now $\vec{r} \in [1, p]^d$ and $\tau(q) > 0$, and recall that in this case $\tau(p) > 0$ because $q > p$. Applying (5.4) of Lemma 3 with $s = p$ and $s_1 = p$, $j = 1, \ldots, d$, we obtain

$$\|B_h^{(j)}\|_p^2 \leq c_9 (L_j h_j^{(p)})^2 = c_9 L_j h_j^{\kappa(p)}, \quad (5.11)$$

for any $f \in \mathcal{N}_{r,d}(\vec{r}, \vec{L})$.

It remains to consider the case $\vec{r} \in [1, p]^d$ and $\tau(q) \leq 0$. Applying the interpolation inequality with $s_0 = r_j$, $s = p$ and $s_1 = q$ we get for any $j = 1, \ldots, d$ and any $f \in \mathcal{N}_{r,d}(\vec{r}, \vec{L}) \cap \mathcal{B}_q(Q)$

$$\|B_h^{(j)}\|_p^2 \leq \|B_h^{(j)}\|_{r_j}^{2(1-(1/p)r_j)} \|B_h^{(j)}\|_{q}^{2(1-(1/p)q)} \leq c_{10} \|B_h^{(j)}\|_{r_j}^p;$$

To get the last inequality we have used (5.5) of Lemma 3 with $r = q$ and $R = Q$. It yields together with (5.3) of Lemma 3 that

$$\|B_h^{(j)}\|_p^2 \leq c_{11} L_j h_j^{\kappa(p)}, \quad (5.12)$$

The required result follows from (5.9), (5.10), (5.11) and (5.12).

### 5.3.2. Bound for the variance

Now we derive the upper bound on the variance of $\hat{T}_h$. Recall that $q \geq 2p - 1$ and define

$$\mathcal{L} := \left\{ \max_{k=1, \ldots, p} \left( L_{(2p-k)} \right)^{\frac{(1-k/p)h(2p-k)}{(r_k/p)}}, \quad \tau(q) > 0; \right\}$$

$$\left\{ 1, \quad \tau(q) \leq 0, \right\}$$

**Proposition 4.** Let $\vec{r}, \vec{L}, r, q > 2p - 1$ and $Q > 0$ be fixed. Then there exists $C > 0$ independent of $L$ such that for any $f \in \mathcal{N}_{r,d}(\vec{r}, \vec{L}) \cap \mathcal{B}_q(Q)$

$$\text{var}_f[\hat{T}_h] \leq C \mathcal{L} \left\{ \sum_{k=1}^p \left( \|f\|_p \right)^{\frac{(2-k)h(2p-k)}{(r_k/p)}} (n^{k} V_k^{-1})^{\frac{1}{1}}, \quad \tau(q) > 0; \right\}$$

$$\left\{ \sum_{k=1}^p \left( \|f\|_p \right)^{\frac{(q-2p+k)p}{q-p}} (n^{k} V_k^{-1})^{\frac{1}{1}}, \quad \tau(q) \leq 0, \right\}$$
Proof  In view of Lemma 2

$$\text{var}_f [\hat{T}_h] \leq c_1 \sum_{k=1}^{p} \| f \|_{2p-k}^{2p-k} (n^k V_h^{k-1})^{-1},$$

Assume first that $\tau(q) > 0$; this also implies that $\tau(r) > 0$ for any $r \leq q$. Applying Lemma 4 with $s = 2p - k$ we get for any $k \in \{1, \ldots, p\}$

$$\| f \|_{2p-k}^{2p-k} \leq (L_{\gamma(2p-k)})^{(1-k/p)(2p-k)/p} (\| f \|_p)^{(2p-k)(2p-k)/p} \leq \mathcal{L} (\| f \|_p)^{(2p-k)(2p-k)/p}. $$

Now assume that $\tau(q) \leq 0$. Applying the interpolation inequality with $s_0 = p$, $s = 2p - k$ and $s_1 = q$ we get for any $k \in \{1, \ldots, p\}$

$$\| f \|_{2p-k}^{2p-k} \leq (\| f \|_p)^{(2p-k)/q-p} (\| f \|_q)^{(2p-k)/q} \leq c_2 (\| f \|_p)^{(2p-k)/q-p}.$$

This completes the proof.

5.3.3. Completion of the theorem proof

Now we are in a position to complete the proof of the theorem.

Define

$$\mathcal{L} := (\mathcal{L}_1^{-1})^{1/(2/(1+1/p)/p)}, \quad \mathcal{L} := \mathcal{L}^{1/p} L_1^{-1/p}, \quad L_{\alpha_j} := \prod_{j=1}^{d} L_1^{1/\alpha_j},$$

and recall that $\mathcal{N} := (\mathcal{N}_1, \ldots, \mathcal{N}_d)$ is defined as follows.

$$\mathcal{N}_j := L_j^{-1/\alpha_j} (\mathcal{L}_1^{-1})^{\alpha_j (1+1/(1+1/p)/p)} = L_j^{-1/\alpha_j} \mathcal{L}_1^{-1/\alpha_j} N^{2/(\alpha_j)}.$$

1°. Several remarks are in order. Direct computations show that

$$N^{p-2} \sum_{j=1}^{d} L_j^{p_j} \mathcal{N}_j^{p_j \alpha_j} = dN^p; \quad \mathcal{L} N^{p} [n^p (V_h)^{p-1}]^{-1} = N^{2p}. \quad (5.14)$$

The following useful equality is deduced from (5.14):

$$[n^p V_h]^{-1} = [\mathcal{L}^{-1} n N^p]^{1/p} = [\mathcal{L}^{-1} \mathcal{L}_1]^{1/p} N^{1-2/p}. \quad (5.15)$$

Let $R_n(f)$ denote the quadratic risk of the estimator $\hat{T}_h$, and let

$$F := \mathbb{N}_{\tau,d}(\mathcal{L}) \cap \mathbb{B}_q(Q) \cap \mathbb{B}_p(N), \quad \bar{F} := [\mathbb{N}_{\tau,d}(\mathcal{L}) \cap \mathbb{B}_q(Q)] \setminus F$$

We obviously have

$$\mathcal{R}_n := \mathcal{R}_n [\hat{N}_h, \mathbb{N}_{\tau,d}(\mathcal{L}) \cap \mathbb{B}_q(Q)] = \max_{f \in \bar{F}} \mathcal{R}_n [\hat{N}_h, f], \sup_{f \in \bar{F}} \mathcal{R}_n [\hat{N}_h, f]$$

A. Goldenshluger and O. V. Lepski/Estimation of norms of a probability density

14
and, we deduce from (5.6) that

\[ R_n^2 \leq \max \left( \sup_{f \in \mathcal{F}} \left[ R_n(f) \right]^{1/p}, \sup_{f \in \mathcal{F}} \| f \|^{2-2p} R_n(f) \right) \]

2\text{d}. In view of Propositions 3 and 4, and by (5.13) and (5.15) we have that for any \( f \in \mathcal{F} \)

\[ E_f |\hat{T}_h - \| f \|_p^p| \leq c_1 n^p; \]

\[ \text{var}_f [\hat{T}_h] \leq \frac{C_1(\bar{L})}{n} \left\{ \sum_{k=1}^{\infty} N^{(2k^{-1})\tau(2k^{-1})+1-\frac{p}{p-1} \tau}} (k-1), \quad \tau(q) > 0 \right\}
\]
\[ + \frac{p}{q-p} \sum_{k=1}^{\infty} N^{(q-2p+1)\tau(q)+1-\frac{p}{p-1} \tau}} (k-1), \quad \tau(q) \leq 0. \]

Setting \( \theta = 1 - \frac{2}{pq} + b \),

\[ a = \left\{ \begin{array}{ll}
\frac{(2p-1)\tau(2p-1)}{\tau(p)}, & \tau(q) > 0; \\
\frac{(q-2p+1)p}{q-p}, & \tau(q) \leq 0,
\end{array} \right. \quad b = \left\{ \begin{array}{ll}
\frac{\tau(\infty)}{\tau(p)}, & \tau(q) > 0; \\
\frac{p}{q-p}, & \tau(q) \leq 0,
\end{array} \right. \]

we obtain for any \( f \in \mathcal{F} \)

\[ \text{var}_f [\hat{T}_h] \leq \frac{C_2(\bar{L})N^{\alpha}}{n} \sum_{i=0}^{\infty} N^{(1-\frac{p}{p-1} \tau}} + \theta(p-1)] \]
\[ + \frac{C_3(\bar{L})N^3}{n} = C_3(\bar{L})N^3, \]

where here and later \( y_- := \min(y, 0) \), and \( \beta := a + 1 + \frac{2(p-1)}{pq} + \left[ \theta(p-1) \right]_- \). It yields together with (5.16)

\[ \sup_{f \in \mathcal{F}} \left[ R_n(f) \right]^{1/p} \leq C_4(\bar{L}) \left[ N^2 + N^{\alpha/2} \right]. \]

3\text{d}. Let us compute the quantity \( 1/v \). If \( \bar{r} \in [p, \infty]^d \) we have

\[ \frac{1}{v} = \frac{p}{2(p-1)} \sum_{j=1}^{d} \frac{1}{\beta_j} = \frac{p(1/\beta - 1/\omega)}{2(p-1)}. \]

If \( \bar{r} \in [1, p]^d \) and \( \tau(q) > 0 \) we have

\[ \frac{1}{v} = \frac{1}{2\tau(p)} \sum_{j=1}^{d} \frac{\tau(r_j)}{\beta_j} = \frac{1}{2\tau(p)} \left( \frac{\tau(\infty)}{\beta} + \frac{1}{\omega} \right) = \frac{1}{2\beta \tau(p)}. \]

Finally, if \( \bar{r} \in [1, p]^d \) and \( \tau(q) \leq 0 \) we have

\[ \frac{1}{v} = \frac{pq}{2(q-p)} \sum_{j=1}^{d} \frac{1/r_j - 1/q}{\beta_j} = \frac{pq}{2(q-p)} \left( \frac{1}{\omega} - \frac{1}{\beta q} \right). \]

It yields, in particular,

\[ (p-1)\theta = \left\{ \begin{array}{ll}
\frac{p-1}{p\beta \tau(p)} - \frac{1}{\beta} + \frac{1}{\omega}, & \bar{r} \in [p, \infty]^d; \\
0, & \bar{r} \in [1, p]^d, \quad \tau(q) > 0; \\
\frac{(p-1)\eta(q)}{q-p}, & \bar{r} \in [1, p]^d, \quad \tau(q) \leq 0.
\end{array} \right. \]
Noting that \( \bar{r} \in [p, \infty)^d \) implies \( \tau(p) \geq 1 \) we assert
\[
\frac{p - 1}{p_\beta \tau(p)} = \frac{1}{\beta} + \frac{1}{\omega} \leq \frac{p - 1}{\tau(p)} - \frac{1}{\beta} + \frac{1}{\omega} = \frac{1}{\omega} - \frac{1}{p_\beta} = 1 - \tau(p) \leq 0.
\]
Hence in all cases \( (p - 1)\theta \leq 0 \) and we conclude that
\[
3 = a + 1 + \frac{2(p - 1)}{p_\nu} + \theta(p - 1) = a + b(p - 1) + p
\]
\[
= p + \frac{(2p - 1)\tau(\infty) + 1/\beta}{\tau(p)} - \frac{(p - 1)\tau(\infty)}{\tau(p)} = 2p.
\]
Hence we deduce from (5.17) that
\[
\sup_{f \in \mathcal{F}} \left[ R_n(f) \right]^{1/p} \leq C_4(\bar{L}) N^2. \tag{5.21}
\]

4\( ^a \). Note that for any \( f \in \bar{F} \) one has in view of Proposition 3 and (5.13)
\[
\|f\|_p^{2-p} (\mathbb{E}_f [\bar{L} - \|f\|_p])^2 \leq c_1 \|f\|_p^{-2} \left( \sum_{j=1}^d \varrho_{i,j} h_p^{p,\kappa_j} \right)^2 \leq c_2 N^2. \tag{5.22}
\]
Here we have used that \( N \) is small then \( n \) is large.
(a) If \( \tau(q) > 0 \) then for any \( k \in \{1, \ldots, p\} \)
\[
\frac{(2p - k)\tau(2p - k) - (2p - 2)\tau(p)}{\tau(p)} = \frac{-(k - 2)\tau(\infty) + (p - 2)/(p_\beta)]}{\tau(p)}
\]
\[
= -(k - 2)\tau(\infty) + 1/(p_\beta) - (p - k)/(p_\beta),
\]
\[
= -(k - 2)\tau(p) - (p - k)/(p_\beta).
\]
Since \( \tau(p) > 0 \), for any \( k \in \{2, \ldots, p\} \)
\[
\frac{(2p - k)\tau(2p - k)}{\tau(p)} - (2p - 2) = -(k - 2) - \frac{(p - k)}{p_\beta \tau(p)} \leq 0.
\]
Putting \( \alpha = 1 - \frac{(p - 1)}{p_\beta \tau(p)} \) we deduce from Proposition 4 for any \( f \in \bar{F} \)
\[
\text{var}_f [\bar{L}] \leq C \mathbb{E} n^{-1} \left[ N^{\alpha} + \sum_{k=2}^p N^{-(k-2)} \frac{(p-k)}{p_\beta \tau(p)} (n \varrho_k)^{-1} \right]
\]
\[
\leq C_4(\bar{L}) n^{-1} \left[ N^{\alpha} + \sum_{k=2}^p N^{-(k-2)} \frac{(p-k)}{p_\beta \tau(p)} + (1 - \frac{1}{p_\beta})(k-1) \right]
\]
\[
\leq C_5(\bar{L}) n^{-1} \left[ N^{\alpha} + N^{-\gamma + (p-2)\theta} \right],
\]
where we have put \( \gamma = 1 - \frac{2}{p_\nu} - \frac{p - 2}{p_\beta \tau(p)} \) and used that \( \theta \leq 0 \). The last bound can be rewritten as
\[
\text{var}_f [\bar{L}] \leq C_6(\bar{L}) \left[ n^{-1} N^{\alpha} + N^z \right], \tag{5.23}
\]
where \( z = \gamma + 1 + \frac{2(p - 1)}{p_\nu} + (p - 2)\theta \). We have
\[
z = p - \frac{p - 2}{p_\beta \tau(p)} + (p - 2)b = p - \frac{p - 2}{\tau(p)} \left[ \tau(\infty) + 1/(p_\beta) \right] = 2.
and (5.23) becomes

\[
\text{var} [\hat{h}] \leq C\gamma(\vec{L}) [n^{-1}N^{\alpha} + N^2], \quad \forall f \in \mathbb{R}.
\] (5.24)

Thus, if \( \tau(q) > 0 \) we conclude from (5.21), (5.22) and (5.24)

\[
\mathcal{R}_n^2 \leq C\gamma(\vec{L}) [n^{-1}N^{\alpha} + N^2].
\] (5.25)

Additionally we remark that if \( \alpha < 0 \) then

\[
n^{-1}N^{\alpha} = C_8(\vec{L})N^{\alpha+1+2(1-1/p)/\upsilon}.
\]

If \( \vec{r} \in [p, \infty]^d \) then \( \tau(p) > 1 \), and we get from (5.18)

\[
\alpha + 1 + 2(1 - 1/p)/\upsilon = 2 - \frac{(p-1)}{\tau(p)p/\beta} + \frac{1}{\omega} \geq 2 + \frac{1}{p/\beta} - \frac{1}{\omega} \geq 2.
\]

If \( \vec{r} \in [1, p]^d \) then in view of (5.19)

\[
\alpha + 1 + 2(1 - 1/p)/\upsilon = 2 - \frac{(p-1)}{\tau(p)p/\beta} + \frac{(p-1)}{p/\beta\tau(p)} = 2.
\]

Thus, (5.25) is equivalent to

\[
\mathcal{R}_n^2 \leq C_9(\vec{L}) \left\{ \begin{array}{ll}
n^{-1} \lor N^2, & \alpha \geq 0; \\
N^2, & \alpha < 0.
\end{array} \right.
\] (5.26)

The definition of \( N \) implies that

\[
N^2 \leq n^{-1} \Leftrightarrow 2(1 - 1/p)/\upsilon \leq 1 \Leftrightarrow 1 \geq \left\{ \begin{array}{ll}
1/\beta - 1/\omega, & \vec{r} \in [p, \infty]^d; \\
\frac{(p-1)}{p/\beta\tau(p)}, & \vec{r} \in [1, p]^d.
\end{array} \right.
\]

In the case \( \vec{r} \in [1, p]^d \) it yields immediately that \( N^2 \leq n^{-1} \Leftrightarrow a \geq 0 \).

It remains to note that if \( \vec{r} \in [p, \infty]^d \) then

\[
\alpha \geq 0 \Leftrightarrow 1 - \frac{(p-1)}{p/\beta\tau(p)} = 1 - \frac{1}{\beta} + \frac{1}{\omega} + \left[ \frac{(p-1)}{p/\beta} - \frac{(p-1)}{p/\beta\tau(p)} \right] + \left[ \frac{1}{p/\beta} - \frac{1}{\omega} \right] \geq 0.
\]

Since in the considered case \( \tau(p) \geq 1 \Leftrightarrow 1/(p/\beta) > 1/\omega \) we assert that

\[
1 \geq 1/\beta - 1/\omega \Rightarrow a \geq 0.
\]

Thus, we deduce from (5.26)

\[
\mathcal{R}_n^2 \leq C_9(\vec{L}) \max \left[ n^{-1}, n^{-1+2/\upsilon(p/\beta)} \right]
\]

and the assertion of the theorem in the case \( \tau(q) > 0 \) follows from (5.18) and (5.19).

(b). Let \( \tau(q) \leq 0 \). For any \( k \in \{1, \ldots, p\} \) one has

\[
\frac{(q - 2p + k)p}{q - p} - 2(p - 1) = 2 - p + \frac{p(k - p)}{q - p} \leq 0
\]
and, therefore, we have in view of Proposition 4 and (5.15) for any $f \in \mathcal{F}$

\[
\text{var}_f[\tilde{T}_h] \leq C_{\mathcal{L}} N^{2-p} \sum_{k=1}^{p} N^{\frac{1-p}{q}} n^{\frac{1-p}{q}} (n^{\frac{1-p}{q}} (nV_h)^{-1})^{-1} \\
\leq C_{\mathcal{L}} N^{\frac{2p-p-q}{q+p}} n^{-1} \sum_{k=1}^{p} (N^{\frac{p}{q}} (nV_h)^{-1})^{k-1} \\
= C_{10}(\bar{L}) N^{2-\frac{(p-1)(q)}{q(p+q)}} \sum_{k=1}^{p} N^{\frac{q(q-k-1)}{q+p}} = C_{10}(\bar{L}) N^2.
\]

To get the penultimate equality we have used (5.20). It yields together with (5.22)

\[
\sup_{f \in \mathcal{F}} \|f\|_p^{-2} R_n(f) \leq C_{11}(\bar{L}) N^2.
\]

which together with (5.21) leads to

\[
\mathcal{R}_n^2 \leq C_{11}(\bar{L}) N^2 = C_{11}(\bar{L}) n^{-\frac{2p}{q+1}} = C_{11}(\bar{L}) n^{-\frac{2(1/p-1/(2q))}{1/q+1/(p/q)}}.
\]

This completes the theorem proof.

### 6. Appendix

**Proof of Lemma 1** Using the identity $a^p = \sum_{j=0}^{p} \binom{p}{j} b^{p-j} (a-b)^j$ with $a = f(x)$ and $b = S_h(x)$ we obtain for all $h \in (0, \infty)^d$ and $x \in \mathbb{R}^d$

\[
f^p(x) = \sum_{j=0}^{p} \binom{p}{j} [S_h(x)]^{p-j} [f(x) - S_h(x)]^j dx \\
= S_h^p(x) (1 - p) + p[S_h(x)]^{p-1} f(x) + \sum_{j=2}^{p} \binom{p}{j} (-1)^j [S_h(x)]^{p-j} B_j(x).
\]

Integrating the last equality, we come to the statement of the lemma.

**Proof of Lemma 2** For any $y \in \mathbb{R}^d$ let

\[
I_h(y) := \bigcap_{j=1}^{d} \{ x \in \mathbb{R}^d : |x_j - y_j| \leq h_j \}.
\]

1. We start with bounding the variance of $\tilde{T}_{1,h}$. Define

\[
g_k(x_1, \ldots, x_k) := \mathbb{E}_f \left[ T_h^{(1)}(x_1, \ldots, x_k, X_{k+1}, \ldots, X_p) \right], \quad k = 1, \ldots, p - 1,
\]

\[
g_p(x_1, \ldots, x_p) := U_h^{(1)}(x_1, \ldots, x_p).
\]

Then the variance of $\tilde{T}_{1,h}$ is given by the following well known formula [see, e.g., Serfling (1980)]:

\[
\text{var}_f[\tilde{T}_{1,h}] = \frac{1}{(p)^p} \sum_{k=1}^{p} \binom{p}{k} \frac{n-p}{p-k} \zeta_k, \quad \zeta_k := \text{var}_f[g_k(X_1, \ldots, X_k)].
\]
We note that \( g_k(x_1, \ldots, x_k), k = 1, \ldots, p \) are symmetric functions. Observe that for \( k = 1, \ldots, p-1 \)
\[
|g_k(x_1, \ldots, x_k)| = |E_f \left[ g(x_1, \ldots, x_k, X_{k+1}, \ldots, X_p) \right] | \\
\leq \int \prod_{i=1}^k |K_h(y - x_i)| \left[ \prod_{i=k+1}^p \int |K_h(y - x_i)| f(x_i) dx_i \right] dy \\
\leq \|K\|^p_{\infty} \int \prod_{i=1}^k |K_h(y - x_i)| (M[f](y))^{p-k} dy.
\]

Then
\[
\zeta_k \leq E_f \left[ g_k(X_1, \ldots, X_k) \right]^2 \\
\leq \|K\|^2_{\infty} (p-k) \int \int (M[f](y))^{p-k} (M[f](z))^{p-k} \\
\times \left[ \prod_{i=1}^k \int K_h(y - x_i) K_h(z - x_i) f(x_i) dx_i \right] dy dz \\
\leq \|K\|^2_{\infty} V_h^{p-k} \int \int (M[f](y))^{p-k} (M[f](z))^{p-k} 1\{y - x \in I_{2h}(0)\} \\
\times \left[ \prod_{i=1}^k \int K_h(z - x_i) f(x_i) dx_i \right] dy dz \\
\leq \|K\|^2_{\infty} V_h^{p-k} \int \int (M[f](y))^{p-k} (M[f](z))^{p} 1\{y - x \in I_{2h}(0)\} dz dy \\
\leq c_1 \|K\|^2_{\infty} V_h^{p-k+1} \int (M[f](y))^{p-k} M [M^p[f]](y) dy.
\]

Furthermore, by the Hölder inequality and (5.2)
\[
\int (M[f](y))^{p-k} M [M^p[f]](y) \leq \|M[f]\|_{2p-k} \|M [M^p[f]]\|_{(2p-k)/p} \\
\leq c_2 \|f\|_{2p-k} \|M^p[f]\|_{(2p-k)/p} = c_2 \|f\|_{2p-k} \|M[f]\|_{2p-k} \leq c_3 \|f\|_{2p-k}^{2p-k}. \tag{6.1}
\]

Therefore we obtain for any \( k = 1, \ldots, p-1 \)
\[
\zeta_k \leq c_2 \|K\|^2_{\infty} V_h^{p-k+1} \int |f(x)|^{2p-k} dx.
\]

In addition,
\[
E_f[g_p(X_1, \ldots, X_p)]^2 = \int \left[ \prod_{i=1}^p \int K_h(y - x_i) K_h(z - x_i) f(x_i) dx_i \right] dy dz \\
\leq \frac{\|K\|^2_{\infty}}{V_h^p} \int 1\{y - z \in I_{2h}(0)\} (M[f](y))^p dz dy \leq \frac{c_3 \|K\|^2_{\infty}}{V_h^{p-1}} \int |f(x)|^p dx.
\]

Thus we obtain
\[
\text{var}_f \left[ \hat{T}_{1, h} \right] \leq C_1 \|K\|^2_{\infty} \sum_{k=1}^p \frac{1}{n^k V_h^{k-1}} \int |f(x)|^{2p-k} dx. \tag{6.2}
\]
Bounding the variance of $\hat{T}_{2,h}$ goes along the same lines. Define

$$g_k(x_1, \ldots, x_k) := \mathbb{E}_f [r_h^{(2)}(x_1, \ldots, x_k, X_{k+1}, \ldots, X_p)], \quad k = 1, \ldots, p - 1,$$

$$g_p(x_1, \ldots, x_p) := U_h^{(2)}(x_1, \ldots, x_p).$$

We have for $k = 1, \ldots, p - 1$

$$g_k(x_1, \ldots, x_k) = \mathbb{E}_f [U_h^{(2)}(x_1, \ldots, x_k, X_{k+1}, \ldots, X_p)]$$

$$= \frac{1}{p} \sum_{i=1}^{k} \prod_{j=1, j \neq i}^{k} K_h(x_j - x_i) \left[ \prod_{j=k+1}^{p} K_h(x_j - x_i) f(x_j) dx_j \right]$$

$$+ \frac{1}{p} \sum_{i=k+1}^{p} \int \prod_{j=1}^{k} K_h(x_j - x_i) \left[ \prod_{j=k+1}^{p} K_h(x_j - x_i) f(x_j) dx_j \right] f(x_i) dx_i.$$

Therefore

$$|g_k(x_1, \ldots, x_k)| \leq \frac{\|K\|_{p-k}}{p} \sum_{i=1}^{k} (\mathfrak{M}[f](x_i))^{p-k} \prod_{j=1, j \neq i}^{k} |K_h(x_j - x_i)|$$

$$+ \frac{\|K\|_{p-k-1}}{p} \sum_{i=k+1}^{p} \int \prod_{j=1}^{k} |K_h(x_j - x_i)| (\mathfrak{M}[f](x_i))^{p-k-1} f(x_i) dx_i$$

$$:= g_k^{(1)}(x_1, \ldots, x_k) + g_k^{(2)}(x_1, \ldots, x_k).$$

For the first term on the right hand side we obtain

$$\mathbb{E}_f |g_k^{(1)}(X_1, \ldots, X_k)|^2 \leq c_1 \|K\|^{2(p-k)}_{\infty} \int (\mathfrak{M}[f](x_i))^{2(p-k)}$$

$$\times \left[ \prod_{j=1, j \neq i}^{k} f(x_j) dx_j \right] f(x_i) dx_i$$

$$\leq \frac{c_1 \|K\|^{2p-2}_{\infty}}{V_h^{k-1}} \int (\mathfrak{M}[f](x))^{2p-k-1} f(x) dx$$

$$\leq \frac{c_2 \|K\|^{2p-2}_{\infty}}{V_h^{k-1}} \int (f(x))^{2p-k} dx \leq \frac{c_2 \|K\|^{2p-2}_{\infty}}{V_h^{k-1}} \int (f(x))^{2p-k} dx.$$
The expectation of the squared second term is bounded as follows:

\[ \mathbb{E}_f |g_h^{(2)}(x_1, \ldots, x_k)|^2 \]

\[ \leq c_4 \|K\|^2_{\infty} \int \left\{ \prod_{j=1}^{k} \int K_h(x_j - y)K_h(x_j - z)f(x_j)dx_j \right\} \times (\mathcal{M}[f](y))^p \mathcal{M}[f](z))^p \, dydz \]

\[ \leq \frac{c_5 \|K\|_{\infty}^{2p-2}}{V_h^{p-1}} \int (\mathcal{M}[f](z))^p \mathcal{M}[\mathcal{M}[f]](z)dz \]

where we have used (6.1). Finally,

\[ \mathbb{E}_f |g_h(X_1, \ldots, X_p)|^2 \leq c_6 \int \left[ \prod_{j=2}^{p} \int K_h^2(x_j - x)f(x_j)dx_j \right] f(x)dx \]

\[ \leq \frac{c_6 \|K\|_{\infty}^{2p-2}}{V_h^{p-1}} \int (\mathcal{M}[f](x))^{p-1} f(x)dx \]

\[ \leq \frac{c_7 \|K\|_{\infty}^{2p-2}}{V_h^{p-1}} \int (\mathcal{M}[f](x))^p dx \leq \frac{c_7 \|K\|_{\infty}^{2p-2}}{V_h^{p-1}} \int [f(x)]^2 \, dx. \]

Thus, we obtain

\[ \text{var}_f \left[ \hat{T}_{2,h} \right] \leq C_2 \|K\|_{\infty}^{2p-2} \sum_{k=1}^{p} \frac{1}{n^k V_h^{k-1}} \int [f(x)]^2 \, dx. \quad (6.3) \]

The assertion of the lemma follows now from (6.2) and (6.3).

**Proof of Lemma 4** Let \( K \) be a kernel satisfying Assumption 1 with \( \ell \geq \max_{j=1, \ldots, d} \beta_j \). For any \( \eta = (\eta_1, \ldots, \eta_d) \), \( \eta_j > 0 \), \( j = 1, \ldots, d \) we have in view of Lemma 3

\[ \|f\|_s \leq \|B_0\|_s + \|S_0\|_s \leq \sum_{j=1}^{d} \|B^{(j)}_\eta\|_s + \|K_\eta * f\|_s. \]

By the general form of Young’s inequality with \( 1/q = 1 + 1/s - 1/p \)

\[ \|K_\eta * f\|_s \leq \|f\|_p \|K_\eta\|_q \leq c_1 (V_\eta)^{1/q-1} \|f\|_p = c_1 (V_\eta)^{1/s-1/p} \|f\|_p. \]

Furthermore, it follows from (5.4) of Lemma 3 with \( s^* = s \) and \( s_j = s \) that

\[ \sum_{j=1}^{d} \|B^{(j)}_\eta\|_s \leq c_2 \sum_{j=1}^{d} L_j \eta_j^{\gamma_j(s)}, \quad \gamma_j(s) = \frac{\beta_j \tau(s)}{\tau(r_j)}. \]
Therefore, for any $\eta_j > 0$, $j = 1, \ldots, d$
\[ \|f\|_s \leq c_1 (V_0)^{1/s-1/p} \|f\|_p + c_2 \sum_{j=1}^d L_j \eta_j^{\gamma_j(s)}. \]

Putting $1/\gamma = \sum_{j=1}^d 1/\gamma_j(s)$ and choosing $\eta_1, \ldots, \eta_d$ from the equality
\[ (V_\eta)^{1/s-1/p} \|f\|_p = \sum_{j=1}^d L_j \eta_j^{\gamma_j(s)} \]
we come to the following bound
\[ \|f\|_s \leq c_3 \left( \prod_{j=1}^d L_j^{\frac{1}{\gamma_j(s)}} \right)^{\frac{1/p - 1/s}{\gamma}} \|f\|_p^{\frac{1}{p-1/s}}. \]

Noting that
\[ 1 + \frac{1/p - 1/s}{\gamma} = 1 + \left[ \frac{1}{p} - \frac{1}{s} \right] \left[ \frac{\tau(\infty)}{\beta \tau(s)} + \frac{1}{\beta \omega \tau(s)} \right] = 1 + \frac{1/p - 1/s}{\beta \tau(s)} = \frac{\tau(p)}{\tau(s)} \]
we complete the proof.

References

BICKEL, R.J. and RITOV, Y. (1988). Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhya: The Indian Journal of Statistics* **50**, 381-393.

BIRGÉ, L. and MASSART, P. (1995). Estimation of the integral functionals of a density. *Ann. Statist.* **23**, 11–29.

CAI, T.T. and LOW, M.G. (2005). Nonquadratic estimators of a quadratic functional. *Ann. Statist.* **33**, 2930-2956.

FOLLAND, G. (1999). *Real Analysis: Modern Techniques and Their Applications*. Second edition. John Wiley & Sons, New York.

GOLDENSHLUGER, A. and LEPSKI, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39**, 1608-1632.

GOLDENSHLUGER, A. and LEPSKI, O.V. (2014). On adaptive minimax density estimation on $\mathbb{R}^d$. *Probab. Theory Related Fields* **159**, 479–543.

GOLDENSHLUGER, A. and LEPSKI, O.V. (2020). Minimax estimation of norms of a probability density: I. Lower bounds. *Manuscript*.

DE GUZMÁN, M. (1975). *Differentiation of Integrals in $R^n$*. Lecture Notes in Mathematics, Vol. 481. Springer-Verlag, Berlin-New York.

HAN, Y., JIAO, J. and MUKHERJEE, R. (2020). On estimation of $L_r$-norms in Gaussian white noise models. *Probab. Theory Related Fields* **177**, no. 3-4, 1243–1294.

HOUDRÉ, C. and REYNAUD-BOURET, P. (2003). Exponential Inequalities, with Constants, for U-statistics of Order Two. *Progress in Probability*, **56**, 55-69.

KERKYACHARIAN, G. and PICARD, D. (1996). Estimating nonquadratic functionals of a density using Haar wavelets. *Ann. Statist.* **24**, 485-507.

LAURENT, B. (1997). Estimation of integral functionals of a density and its derivatives. *Bernoulli* **3**, 181-211.
LEPSKII, O. V. (1991). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.* 36, 682–697.

LEPSKII, O. V. (1992). Asymptotically minimax adaptive estimation. II. Statistical model without optimal adaptation. Adaptive estimators. *Theory Probab. Appl.* 37, 468–481.

LEPSKI, O.V., NEMIROVSKI, A. and SPOKONY, V. (1999). On estimation of the $L_r$-norm of a regression function. *Probab. Theory and Related Fields* 113, 221-253.

LEPSKI, O.V. and SPOKONY, V. (1999). Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. *Bernoulli* 5, 333-358.

LEPSKI, O.V. and WILLER, T. (2019). Oracle inequalities and adaptive estimation in the convolution structure density model. *Ann. Statist.* 47, 233–287.

MAJOR, P. (2013) On the Estimation of Multiple Random Integrals and $U$-Statistics. *Lecture Notes in Mathematics* 2079, Springer-Verlag Berlin Heidelberg 2013.

NIKOL’SKII S.M. (1977). *Priblizienie Funktsii Mnogikh Peremennykh i Teoremy Vlozheniya.* (in Russian). [Approximation of functions of several variables and embedding theorems.] 2nd ed., revised and supplemented. Nauka, Moscow.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, Inc.

TCHETGEN, E., LI, L., ROBINS, J. and VAN DER VAART, A. (2008). Minimax estimation of the integral of a power of a density. *Stat. Probab. Letters* 78, 3307-3311.