Minimax estimation of norms of a probability density: I. Lower bounds

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: The paper deals with the problem of nonparametric estimating the $L^p$-norm, $p \in (1, \infty)$, of a probability density on $\mathbb{R}^d$, $d \geq 1$ from independent observations. The unknown density is assumed to belong to a ball in the anisotropic Nikolskii’s space. We adopt the minimax approach, and derive lower bounds on the minimax risk. In particular, we demonstrate that accuracy of estimation procedures essentially depends on whether $p$ is integer or not. Moreover, we develop a general technique for derivation of lower bounds on the minimax risk in the problems of estimating nonlinear functionals. The proposed technique is applicable for a broad class of nonlinear functionals, and it is used for derivation of the lower bounds in the $L^p$-norm estimation.

AMS 2000 subject classifications: 62G05, 62G20.
Keywords and phrases: estimation of nonlinear functionals, minimax estimation, minimax risk, anisotropic Nikolskii’s class, best approximation.

1. Introduction

Suppose that we observe i.i.d. vectors $X_i \in \mathbb{R}^d$, $i = 1, \ldots, n$, with common probability density $f$. Let $p > 1$ be a given real number. We want to estimate the $L^p$-norm of $f$,

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

using observations $X^{(n)} = (X_1, \ldots, X_n)$. By estimator we mean any $X^{(n)}$-measurable map $\tilde{F} : \mathbb{R}^n \to \mathbb{R}$, and accuracy of an estimator $\tilde{F}$ is measured by the quadratic risk

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

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

*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).
We adopt the minimax approach to the outlined estimation problem. With any estimator $\tilde{F}$ and any set $\mathcal{F}$ of probability densities on $\mathbb{R}^d$ we associate the maximal risk of $\tilde{F}$ on $\mathcal{F}$:

$$R_n[\tilde{F}, \mathcal{F}] := \sup_{f \in \mathcal{F}} R_n[\tilde{F}, f].$$

The minimax risk is

$$R_n[\mathcal{F}] := \inf_{\tilde{F}} R_n[\tilde{F}, \mathcal{F}],$$

where $\inf$ is taken over all possible estimators. An estimator $\tilde{F}_*$ is called optimal in order or rate–optimal if

$$R_n[\tilde{F}_*; \mathcal{F}] \asymp R_n[\mathcal{F}], \quad n \to \infty.$$

The rate at which $R_n[\mathcal{F}]$ converges to zero as $n$ tends to infinity is referred to as the minimax rate of convergence.

The problems of minimax nonparametric estimation of density functionals have been extensively studied in the literature. The case of linear functionals is particularly well understood: here a complete optimality theory under rather general assumptions has been developed [see, e.g., Ibragimov and Khasminskii (1986), Donoho and Liu (1991), Cai and Low (2004) and Juditsky and Nemirovski (2020)]. As for nonlinear functionals, the situation is completely different: even in the problem of estimating quadratic functionals of a density rate–optimal estimators are known only for very specific functional classes. For representative publications dealing with estimation of quadratic and closely related integral functionals of a probability density we refer to Bickel and Ritov (1988), Birgé and Massart (1995), Kerkyacharian and Picard (1996), Laurent (1996, 1997), Giné and Nickl (2008) and Tchetgen et al. (2008). The problems of estimating non-linear functionals were also considered in the framework of the Gaussian white noise model; e.g., Ibragimov et al. (1986), Nemirovskii (1990) [see also (Nemirovski 2000, Chapters 7 and 8)], Donoho and Nussbaum (1990), Cai and Low (2005). The contribution of this paper is closely related to the works Lepski et al. (1999), Cai and Low (2011) and Han et al. (2020), where the problem of estimation of norms of a signal observed in the Gaussian white noise was studied. Additional pointers to relevant work and discussion of relations between our results and the existing literature are provided in Section 5.5.

This paper deals with the problem of estimating the $L^p$–norm of a probability density and derives lower bounds on asymptotics of the minimax risk over anisotropic Nikolskii’s classes $N_{\vec{r},d}(\vec{\beta}, \vec{L})$ (precise definition of the functional class is given below). In the companion paper Goldenshluger and Lepski (2020a) we develop the corresponding rate–optimal estimators demonstrating that the derived lower bounds are tight. We also study how boundedness of the underlying density $f$ in some integral norm influences estimation accuracy by considering the minimax risk over the functional class $\mathcal{F} = N_{\vec{r},d}(\vec{\beta}, \vec{L}) \cap B_q(Q)$, where

$$B_q(Q) := \{ f : \mathbb{R}^d \to \mathbb{R} : \|f\|_q \leq Q \}, \quad q > 1, \quad Q > 0.$$

The contribution of this paper is two–fold. First, we derive lower bounds on the minimax risk on the class $\mathcal{F}$ in the problem of estimating the $L^p$–norm, $p \in (1, \infty)$, of a probability density. It turns out that the behavior of the minimax risk is completely different in the cases of integer and non–integer values of $p$. The companion paper Goldenshluger and Lepski (2020a) presents matching upper bounds on the minimax risk for the integer values of $p$. Second, we develop general machinery for derivation of lower bounds on the minimax risk in the problems of estimating nonlinear functionals of the type

$$\Psi(f) = G \left( \int_{\mathbb{R}^d} H(f(x)) \, dx \right),$$

(1.1)
where \( G : \mathbb{R} \rightarrow \mathbb{R} \) and \( H : \mathbb{R}^d \rightarrow \mathbb{R} \) are fixed functions. The developed machinery is applied to the derivation of the lower bounds in the problem of estimating \( \|f\|_p \). In order to demonstrate broad applicability of the proposed technique we also provide lower bounds in problems of estimation of other nonlinear functionals of interest.

The rest of this paper is structured as follows. Section 2 presents lower bounds on the minimax risk in the problem of estimating the \( L_p \)–norms of \( f \). Section 3 develops a general technique for derivation of lower bounds in the problems of estimating nonlinear functionals of type (1.1). The main results of these two sections are proved in Sections 5 and 6 respectively. Auxiliary results are presented in the supplementary material Goldenshluger and Lepski (2020b).

2. Lower bounds for estimation of the \( L_p \)–norm

We start with the definition of the anisotropic Nikolskii functional classes. Let \( (L_{2,p}) \). Lower bounds for estimation of the \( \|f\|_p \) are presented in the supplementary material Goldenshluger and Lepski (2020b).

The main results of these two sections are proved in Sections 5 and 6 respectively. Auxiliary results are presented in the supplementary material Goldenshluger and Lepski (2020b).

Definition 1. For given vectors \( \tilde{\beta} = (\beta_1, \ldots, \beta_d) \in (0, \infty)^d \), \( \tilde{r} = (r_1, \ldots, r_d) \in [1, \infty]^d \), and \( \tilde{L} = (L_1, \ldots, L_d) \in (0, \infty)^d \) we say that function \( G : \mathbb{R}^d \rightarrow \mathbb{R} \) belongs to anisotropic Nikolskii’s class \( \mathcal{N}_{r,d}(\tilde{\beta}, \tilde{L}) \) if \( \|G\|_{r_j} \leq L_j \) for all \( j = 1, \ldots, d \), and there exist natural numbers \( k_j \geq \beta_j \) such that

\[
\|\Delta_{r_j}^k G\|_{r_j} \leq L_j |u|^\beta_j, \quad \forall u \in \mathbb{R}, \quad \forall j = 1, \ldots, d.
\]

In addition to constraint \( f \in \mathcal{N}_{r,d}(\tilde{\beta}, \tilde{L}) \) we also assume that \( f \in \mathbb{B}_q(Q) \). By definition of Nikolskii’s class, \( f \in \mathcal{N}_{r,d}(\tilde{\beta}, \tilde{L}) \) implies \( f \in \mathbb{B}_{r^*}(\max_{l=1,\ldots,d} L_l) \), where \( r^* := \max_{l=1,\ldots,d} r_l \). Since we are interested in estimating \( \|f\|_p \), it is necessary to suppose that this norm is bounded. Therefore in all what follows we assume that \( q \geq p \lor r^* \).

Asymptotic behavior of the minimax risks on anisotropic Nikolskii’s classes is conveniently expressed in terms of the following parameters:

\[
\begin{align*}
\frac{1}{\beta_j} &:= \sum_{j=1}^d \frac{1}{\beta_j}, \quad \frac{1}{\omega} := \sum_{j=1}^d \frac{1}{\beta_j r_j}, \quad L := \prod_{j=1}^d L_j^{1/\beta_j}, \\
\tau(s) &:= 1 - \frac{1}{\omega} + \frac{1}{\beta s}, \quad s \in [1, \infty).
\end{align*}
\]

It is worth mentioning that quantities \( \tau() \) appear in embedding theorems for Nikolskii’s classes; for details see Nikolskii (1977).

Now we are ready to state lower bounds on the minimax risk in the problem of estimating the \( L_p \)–norm \( \|f\|_p \). We consider the cases of integer and non–integer \( p \) separately.
2.1. The case of integer \( p \geq 2 \)

Define

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

and let \( \phi_n := L^{\frac{1}{(p-1)/p}} n^{-\theta} \).

**Theorem 1.** Assume that \( p \in \mathbb{N}^*, \; p \geq 2 \). For any \( \vec{\beta} \in (0, \infty)^d \), \( \vec{L} \in (0, \infty)^d \), \( r \in [1, \infty]^d \), and \( q \geq p \vee r^* \) there exists \( c > 0 \) independent of \( \vec{L} \) such that

\[
\liminf_{n \to \infty} \phi_n^{-1} \mathcal{R}_n^* [\mathcal{H}_{r,d}(\vec{\beta}, \vec{L}) \cap \mathbb{B}_q(Q)] \geq c.
\]

**Remark 1.** In the companion paper Goldenshluger and Lepski (2020a) we demonstrate that the rates of convergence of the minimax risk established in Theorem 1 are minimax, that is, they are attained by explicitly constructed estimation procedures.

The lower bounds on the minimax rates of convergence of Theorem 1 exhibit rather unusual features as compared to the results on estimating the \( L_p \)-norm of a signal in the Gaussian white noise model [see Lepski et al. (1999) and Han et al. (2020)].

1. It is quite surprising that the obtained asymptotics of the minimax risk does not depend on \( p \) and \( q \) if \( \tau(p) \geq 1 \). Perhaps it is even more surprising that in some cases the \( L_p \)-norm of a probability density can be estimated with the parametric rate! On the other hand, it is easily seen that \( \theta < 1/2 \) if \( \tau(p) < 1, \; \tau(q) < 0 \); therefore, the parametric rate is not achievable in this regime.

2. If \( r^* = \max_{i=1, \ldots, d} r_i \leq p \) and \( q = p \) then uniformly consistent estimators over anisotropic Nikol’skii’s classes do not exist when \( \tau(p) \leq 0 \). This together with Remark 1 implies that condition \( \tau(p) > 0 \) is necessary and sufficient for existence of uniformly consistent estimators of the \( L_p \)-norm.

3. Taking together the previous remarks, we see that in the considered estimation problem the full spectrum of asymptotic behavior for the minimax risk is possible: from parametric rate of convergence to inconsistency. To the best of our knowledge this phenomenon has not been observed before.

2.2. The case of non-integer \( p > 1 \)

The next two statements present lower bounds for the settings \( q < \infty \) and \( q = \infty \). In the former case all coordinates of the vector \( r \) are finite \((r^* < \infty)\). On the other hand, if at least one of the coordinates of \( r \) equals infinity \((r^* = \infty)\) then necessarily \( q = \infty \).

Define

\[
\delta_n := \begin{cases} 
\frac{\ln^3 \ln(n)}{n \ln^3(n)}, & \tau(p) \geq 1 - 1/p; \\
\frac{1}{n \ln^3(n)}, & \tau(p) < 1 - 1/p, \; \tau(q) < 0; \\
\frac{1}{n}, & \tau(p) < 1 - 1/p, \; \tau(q) \geq 0,
\end{cases}
\]

and let

\[
\vartheta := \begin{cases} 
\frac{1-1/p}{\tau(1)}, & \tau(p) \geq 1 - 1/p; \\
\frac{1/p-1/q}{\tau(q-\tau(q))}, & \tau(p) < 1 - 1/p, \; \tau(q) < 0; \\
\frac{\tau(q)}{\tau(1)}, & \tau(p) < 1 - 1/p, \; \tau(q) \geq 0.
\end{cases}
\]
Theorem 2. Let \( p \notin \mathbb{N}^*, p > 1 \), and let \( \phi_n := \left[ L^{1 - 1/p} \delta_n^q \right] \lor n^{-1/2} \). For any \( \beta \in (0, \infty)^d \), \( \tilde{L} \in (0, \infty)^d, \tilde{r} \in [1, \infty)^d, Q > 0 \) and any \( r^* \lor p \leq q < \infty \) there exists \( c > 0 \) independent of \( \tilde{L} \) such that

\[
\liminf_{n \to \infty} \phi_n^{-1} \mathcal{R}_n^* \left[ N_{r,d}(\beta, \tilde{L}) \cap B_q(Q) \right] \geq c.
\]

The lower bounds in the case \( q = \infty \) are given in the next theorem. Define

\[
\rho_n := \begin{cases} 
\frac{\ln(n)}{\ln(\ln(n))}, & \tau(p) \geq 1 - 1/p; \\
\ln(n), & \tau(p) < 1 - 1/p, \; \tau(q) < 0; \\
1, & \tau(p) < 1 - 1/p, \; \tau(q) \geq 0,
\end{cases}
\]

\[
\varphi_n := \left[ L^{1 - 1/p} \delta_n^q \rho_n^{2(d-1)} \right] \lor n^{-1/2}.
\]

Theorem 3. Let \( p \notin \mathbb{N}^*, p > 1 \), and assume that \( q = \infty \). For any \( Q > 0 \), \( \beta \in (0, \infty)^d \), \( \tilde{L} \in (0, \infty)^d, \tilde{r} \in [1, \infty)^d \) there exists \( c > 0 \) independent of \( \tilde{L} \) such that

\[
\liminf_{n \to \infty} \varphi_n^{-1} \mathcal{R}_n^* \left[ N_{r,d}(\beta, \tilde{L}) \cap B_\infty(Q) \right] \geq c.
\]

1. Note that the rates of convergence established in Theorem 1 and in Theorems 2, 3 are different, except the case \( \tau(p) < 1 - 1/p, \; \tau(q) \geq 0 \). As we can see, the estimation accuracy for integer values of \( p \) is much better than for the non-integer ones. For the first time this phenomenon has been observed by Lepski et al. (1999) in the problem of estimating the \( L_p \)-norm of a signal in the univariate Gaussian white noise model.

2. Theorem 2 shows that if \( q = p \) and \( \tau(p) = 0 \) then there no uniformly consistent estimators exist. If \( q = p \) and \( \tau(p) < 0 \) then the lower bound becomes log-logarithmic in \( n \) if \( q < \infty \). We conjecture that in this case there are no uniformly consistent estimators as well. If our conjecture is true, the proof of lower bounds will require more elaborate considerations.

3. It is not difficult to check that the rate of convergence corresponding to the zone \( \tau(p) \leq 1 - 1/p \) is slower than the one corresponding to \( \tau(p) > 1 - 1/p \) independently of the value of \( q \).

4. As it was mentioned above, in this paper we do not discuss estimation procedures; we refer to Goldenshluger and Lepski (2020a) for construction of rate-optimal estimators of \( \|f\|_p \) for integer values of \( p \). However, for non-integer values of \( p \) in some cases straightforward plug-in constructions lead to nearly rate-optimal adaptive estimators of \( L_p \)-norms. We discuss such specific cases in the companion paper Goldenshluger and Lepski (2020a).

3. Lower bounds for estimation of general non-linear functionals

The results of Theorems 1–3 follow from general machinery for derivation of lower bounds on minimax risks in the density model. In this section we develop this technique in full generality for a broad class of nonlinear functionals to be estimated.

Let \( G : \mathbb{R} \to \mathbb{R} \) and \( H : \mathbb{R}_+ \to \mathbb{R} \) be fixed functions. Consider the problem of estimating the functional

\[
\Psi(f) = G \left( \int_{\mathbb{R}^d} H(f(x))dx \right)
\]

from observation \( X^{(n)} = (X_1, \ldots, X_n) \). Let \( F \) be a class of functions defined on \( \mathbb{R}^d \), and let

\[
\mathcal{R}_n[F] := \inf_{\Psi \in F} \left( \mathbb{E}_f [\Psi - \Psi(f)]^2 \right)^{1/2},
\]
where the infimum is taken over all possible estimators of $\Psi$. Our goal is to derive an explicit lower bound on the minimax risk under mild condition on functions $G$ and $H$ and functional class $\mathcal{F}$.

The class of functionals in (3.1) is rather broad and includes many problem instances of interest. Let us give some examples.

1. Let $G(y) = y^{1/p}$ and $H(y) = y^p$ for some $p \in (1, \infty)$; then $\Psi(f)$ is the $L_p$-norm of $f$, and estimation of this functional is the subject of the present paper.

2. The choice $G(y) = ay$ leads to the estimation of the integral-type functionals. The following particular cases have been considered in the literature.

   (a) if $a = 1$ and $H(y) = y^p$ with $p \in \mathbb{N}^*$, $p \geq 2$, then the corresponding functional is $\Psi(f) = \|f\|_p^p$; see for instance, Bickel and Ritov (1988), Kerkyacharian and Picard (1996), Laurent (1996, 1997), Tchetgen et al. (2008);

   (b) the case $a = 1$ and $H(y) = -y \ln(y)$ corresponds to the differential entropy, $\Psi(f) = -\int f(x) \ln f(x) dx$; see, e.g., Kozachenko and Leonenko (1987);

   (c) if $a = (p - 1)^{-1}$ and $H(y) = y - y^p$ with $p \neq 1$ then $\Psi(f)$ is the Tsallis entropy, $\Psi(f) = (p - 1)^{-1} \int |f(x)|^p dx$; see Tsallis (1988), Leonenko et al. (2008).

3. Let $G(y) = (1 - p)^{-1} \ln(y)$ and $H(y) = y^p$ with $p \neq 1$; then the corresponding functional is the Rényi entropy, $\Psi(f) = (1 - p)^{-1} \ln \left( \int |f(x)|^p dx \right)$; see Rényi (1961), Leonenko et al. (2008).

The technique for derivation of lower bounds relies upon construction of a parameterized family of functions equipped with a pair of prior probability measures on it. Below we discuss these construction ingredients in succession.

### 3.1. Parameterized family of functions

Let $\Lambda : \mathbb{R}^d \to \mathbb{R}_+$ be a function satisfying the following conditions:

$$\Lambda(x) = 0, \quad \forall x \notin [-1,1]^d, \quad \int_{\mathbb{R}^d} \Lambda(x) dx = 1. \tag{3.2}$$

Let $| \cdot |_\infty$ denote the $\ell_\infty$-norm on $\mathbb{R}^d$, and let $\mathcal{M}$ be a given finite set of indices of cardinality $M = \text{card}(\mathcal{M})$. Let $\{x_m \in \mathbb{R}^d, m \in \mathcal{M}\}$ be a finite set of points in $\mathbb{R}^d$ satisfying

$$|x_k - x_m|_\infty \geq 2, \quad \forall k \neq m, k, m \in \mathcal{M}.$$  

Fix vector $\vec{\sigma} = (\sigma_1, \ldots, \sigma_d) \in (0,1)^d$ and constant $A > 0$ and define for any $m \in \mathcal{M}$

$$\Lambda_m(x) = A \Lambda\left(\frac{|x - x_m|}{\vec{\sigma}}\right), \quad \Pi_m = \left\{ x \in \mathbb{R}^d : \frac{|x - x_m|}{\vec{\sigma}_\infty} \leq 1 \right\}, \tag{3.3}$$

where the division is understood in the coordinate-wise sense. In words, $\Pi_m$ is a rectangle in $\mathbb{R}^d$ centered at $x_m$ with edges of half-lengths $\sigma_1, \ldots, \sigma_d$ that are parallel to the coordinate axes. It is obvious that $\Lambda_m$ is supported on $\Pi_m$ for any $m \in \mathcal{M}$, and $\Pi_m$ are disjoint:

$$\Pi_m \cap \Pi_k = \emptyset, \quad \forall k \neq m, k, m \in \mathcal{M}. \tag{3.4}$$

Let $\Pi_0 \setminus \bigcup_{m \in \mathcal{M}} \Pi_m$, $\sigma := \prod_{i=1}^d \sigma_i$, and

$$q_w(z) := \sum_{m \in \mathcal{M}} (w_m)^z, \quad w \in [0,1]^M, \quad z > 0.$$
Let $f_0$ be a probability density supported on $\Pi_0$. Define the family of functions:

$$f_w(x) := [1 - A\sigma w(1)] f_0(x) + A \sum_{m \in M} w_m A_m(x), \quad w \in [0, 1]^M. \tag{3.5}$$

The family $\{f_w, w \in [0, 1]^M\}$ involves tuning parameters $A, \sigma$ and $M$ that will be specified in the sequel. The most important element of our approach consists in equipping $[0, 1]^M$ with two product probability measures, thus assuming that $w$ is a random vector distributed in accordance with one of them. Then functions $f_w$ become random, and they are not necessarily density functions and/or functions from the functional class $F$ for all realizations of $w$. With conditions introduced below we ensure that $f_w$ is a probability density, and $f_w$ belongs to $F$ with large enough probability.

### 3.2. Prior probability measures

Let $\mathcal{P}[0, 1]$ be the set of all probability measures with total mass on $[0, 1]$. For $\pi \in \mathcal{P}[0, 1]$, $z \geq 0$ we define the $z$-th moment of $\pi$ by

$$e_\pi(z) := \int_0^1 x^z \pi(dx).$$

Let $\zeta := (\zeta_m, m \in M)$ be a family of independent identically distributed random variables, and $\zeta_m$ is distributed according to $\pi \in \mathcal{P}[0, 1]$. The law of $\zeta$ and the corresponding expectation will be denoted by $P_\pi$ and $E_\pi$ respectively. Define

$$p_\zeta(x) := \prod_{i=1}^n f_\zeta(x_i), \quad x \in \mathbb{R}^d;$$

here and from now on we regard $x = (x_1, \ldots, x_n)$, $x_i \in \mathbb{R}^d$ as an element of $\mathbb{R}^d$.

### 3.3. Assumptions on the family of functions and prior measures

Now we introduce general assumptions that relate properties of parameterized family $\{f_w, w \in [0, 1]^M\}$ and prior measures on $[0, 1]^M$.

**Assumption 1.** There exist $\varepsilon \in (0, 1)$ and two probability measures $\mu, \nu \in \mathcal{P}[0, 1]$ satisfying $e_\mu(1) = e_\nu(1)$ such that

$$\mathbb{P}_\pi\{f_\zeta \in F\} \geq 1 - \varepsilon, \quad \pi \in \{\mu, \nu\},$$

where $f_\zeta$ is defined in (3.5).

Assumption 1 stipulates that under prior probability measures $\mu$ and $\nu$ the random function $f_\zeta$ belongs to functional class $F$ with probability at least $1 - \varepsilon$. Note that this assumption does not guarantee that $f_\zeta$ is a probability density; by construction, only assumption $\int f_\zeta = 1$ is fulfilled for all realizations of $\zeta$.

We also need conditions that relate parameters $A, \sigma, M$ of the family of functions with the sample size $n$. 
Assumption 2. For sufficiently small \( \varepsilon_0 > 0 \) and sufficiently large \( v \geq 1 \) for \( \pi \in \{\mu, \nu\} \)

\[
\begin{align*}
\text{either } (i) & \quad A\sigma \sqrt{nM} \leq \varepsilon_0, \\
\text{or } (ii) & \quad nA\sigma \leq \varepsilon_0 \ln (1 + e^{-1}(2)); \\
& \quad M \geq 36v^2; \\
& \quad A\sigma M \sqrt{e}z(2) \leq \varepsilon_0.
\end{align*}
\]

Conditions (3.6) and (3.8) imply that \( f_x \) is a probability density with high probability. Indeed, by construction \( \int f_x(x) dx = 1 \), and, in view of (3.8), \( f_x \geq 0 \) for all realizations of \( \zeta \) satisfying \( |\zeta(1) - Me_{\pi}(1)| \leq \sqrt{Me_{\pi}(2)} \). Moreover, condition (3.6) allows us to construct the product form upper and lower bounds on the Bayesian likelihood ratio \( E_{\mu}[p_{\pi}(\cdot)]/E_{\nu}[p_{\pi}(\cdot)] \) which is an essential step in the proposed derivation technique.

### 3.4. Main results

To state lower bounds on the minimax risk for estimating functional \( \Psi(f) \) we require notation that involves functions \( H \) and \( G \) appearing in (3.1).

Define functions \( S_0 : [0, 1] \to \mathbb{R} \) and \( S : \mathbb{R}_+ \to \mathbb{R} \) by

\[
S_0(z) := \int_{1-z}^1 H((1-z)f_0(x)) dx, \quad S(z) := \int_{[-1,1]^d} H(z\Lambda(x)) dx.
\]  

(3.9)

For \( \pi \in \mathcal{P}[0, 1] \) let

\[
E_{\pi}(A) := \int_0^1 S(Ay) \pi(dy), \quad V_{\pi}(A) := \left[ \int_0^1 S^2(Ay) \pi(dy) \right]^{1/2}.
\]  

(3.10)

We tacitly assume that \( E_{\pi}(A) \) and \( V_{\pi}(A) \) are finite for all \( A > 0 \) and for all considered measures \( \pi \in \mathcal{P}[0, 1] \); fulfillment of this assumption should be verified in every concrete problem instance.

To clarify the notation introduced in (3.9) and (3.10) we observe that for the family of functions \( \{f_w, w \in [0, 1]^M\} \) one has

\[
\int_{\mathbb{R}^d} H(f_x(x)) dx = S_0(A\sigma \zeta(1)) + \sigma \sum_{m \in M} S(A\zeta_m) \\
\approx S_0(A\sigma Me_{\pi}(1)) + \sigma ME_{\pi}(A),
\]

(3.11)

where the approximate equality in the second line designates that the sums of independent random variables \( \zeta(1) \) and \( \sum_{m \in M} S(A\zeta_m) \) concentrate properly around their expectations \( Me_{\pi}(1) \) and \( ME_{\pi}(A) \) respectively. In addition, the lower bound derivation requires analysis of discrepancy between the values of the functional \( \Psi(f_x) \) when \( \zeta \) is distributed according to prior measures \( \mu, \nu \in \mathcal{P}[0, 1] \). This fact along with (3.11) motivates the following notation.

Let \( \mu, \nu \in \mathcal{P}[0, 1] \); for \( \pi \in \{\mu, \nu\} \) define

\[
\begin{align*}
H_{\pi}^* & := S_0(A\sigma Me_{\pi}(1)) + \sigma ME_{\pi}(A), \\
\alpha_{\pi} & := \eta S_0(A\sigma Me_{\pi}(1); A\sigma \sqrt{vMe_{\pi}(2)}) + \sigma \sqrt{vM} V_{\pi}(A), \\
J_{\pi} & := \left[ \inf_{|\alpha| \leq \alpha_{\pi}} G(H_{\pi}^* + \alpha), \sup_{|\alpha| \leq \alpha_{\pi}} G(H_{\pi}^* + \alpha) \right],
\end{align*}
\]

(3.12) - (3.14)
where \( \eta_{S_0}(x; \delta) \) stands for the local modulus of continuity of function \( S_0 \),
\[
\eta_{S_0}(x; \delta) := \sup_{y: |y-x| \leq \delta} |S_0(x) - S_0(y)|, \ x, y \in [0, 1], \ \delta > 0.
\]

Define also
\[
\Delta(\mu, \nu) := \min \{|x - x'| : x \in J_\mu, x' \in J_\nu\};
\]
clearly, \( \Delta(\mu, \nu) \) is the Hausdorff distance between the intervals \( J_\mu \) and \( J_\nu \).

Finally we let
\[
n_m(x) := \sum_{i=1}^{n} 1_{n_m}(x_i), \ n_0(x) := n - \sum_{m \in \mathcal{M}} n_m(x), \ x \in \mathbb{R}^d,
\]
where sets \( \Pi_m, m \in \mathcal{M} \) are defined in (3.3). The quantities \( n_m(x) \) and \( n_0(x) \) have evident probabilistic interpretation: if \( X^{(n)} = (X_1, \ldots, X_n) \) is a sample then \( n_m(X^{(n)}) \) and \( n_0(X^{(n)}) \) are the numbers of observations in the sets \( \Pi_m \) and \( \Pi_0 \) respectively. Furthermore, for a pair of measures \( \mu, \nu \) satisfying \( e_\mu(1) = e_\nu(1) \) we define
\[
\Upsilon(x) := \prod_{m \in \mathcal{M}} \frac{\gamma_{m,\mu}(x)}{\gamma_{m,\nu}(x)}, \ x \in \mathbb{R}^d,
\]
\[
\gamma_{m,\pi}(x) := \int_0^1 y^{n_m(x)} e^{-D_{m,n}(y)} \pi(dy), \ D := \frac{A \sigma}{1 - \lambda \sigma M e_\nu(1)}.
\]

Observe that \( D \) does not depend on \( \pi \in \{\mu, \nu\} \) because \( e_\mu(1) = e_\nu(1) \).

Now we are in a position to formulate the main result of this section.

**Theorem 4.** Let \( \mu, \nu \in \mathcal{P}[0, 1] \) satisfy \( e_\mu(1) = e_\nu(1) \) and suppose that Assumptions 1 and 2 are fulfilled. If \( \Delta(\mu, \nu) > 0 \) then
\[
36e[\Delta(\mu, \nu)]^{-2} \mathcal{R}^2_n[\pi] \geq \mathbb{E}_\nu \left[ \mathbb{P}_\mu \left\{ \Upsilon(X^{(n)}) \geq \frac{1}{2} \right\} \right] - v^{-1} - 4\sqrt{2(2v^{-1} + \epsilon)}. \tag{3.16}
\]

In order to apply general lower bound (3.16) in concrete problem instances we need to compute or bound from below the quantity \( \Delta(\mu, \nu) \) and to show that the right hand side is strictly positive. The following corollary of Theorem 4 derives lower bounds on the right hand side of (3.16) under additional conditions on the prior measures and parameters of the family \( \{f_w, w \in [0, 1]^M\} \).

**Definition 2.** For \( \mu, \nu \in \mathcal{P}[0, 1] \) we write \( \mu \asymp \nu \) if
1. \( e_\mu(k) = e_\nu(k) \) for all \( k = 1, \ldots, t - 1, t \in \mathbb{N}^* \);
2. \( e_\mu(k) = e_\nu(k) \) for all \( k = 1, \ldots, 2[t], t \notin \mathbb{N}^* \).

**Corollary 1.** Let \( t > 1 \) be a fixed integer number, or a non–integer number that can depend on the sample size \( n \). Let \( \mu, \nu \in \mathcal{P}[0, 1] \) satisfy \( \mu \asymp \nu \) and suppose that Assumptions 1 and 2 are fulfilled. Moreover, assume that for sufficiently small \( \varepsilon_1 > 0 \) independent of \( n \) one has
\[
n A \sigma t^{-1} M^{1/t} \leq \varepsilon_1. \tag{3.17}
\]
If \( \Delta(\mu, \nu) > 0 \) then \( \mathcal{R}_n[\pi] \geq C_* \Delta(\mu, \nu) \), where \( C_* = (36e)^{-1/2} \left[ \frac{1}{3} - v^{-1} - \sqrt{2(\varepsilon + 2/v)} \right]^{1/2}. \)
Theorem 4 along with Corollary 1 provides general guidelines for derivation of lower bounds in problems of estimating nonlinear functionals $\Psi(f)$ of type (3.1). Given the family of functions \{\(f_w, w \in [0, 1]^M\)\} defined in (3.5) we need to construct a pair of prior probability measures $\mu, \nu$ on \([0, 1]\) such that Assumptions 1, 2 and condition (3.17) hold. Then the minimax risk in estimating $\Psi(f)$ is bounded below by the quantity $\Delta(\mu, \nu)$. The moment matching property of the probability measures, $\mu \sim \nu$, is essential for maximization of the value of $\Delta(\mu, \nu)$, and the parameter $t$ should chosen differently in different settings. In Propositions 1 and 4 of Section 4 we present two constructions of prior measures $\mu$ and $\nu$ satisfying the moment matching property $\mu \sim \nu$ and possessing some additional properties.

### 3.5. Discussion

In this section we discuss applicability of Theorem 4 and Corollary 1, and main ideas that underlie the proofs of these results.

**More general statistical experiments** The proofs of Theorem 4 and Corollary 1 do not use the fact that density $f$ is defined on $\mathbb{R}^d$. In fact, after minor changes and modifications our construction is applicable in an arbitrary density model.

Let $(X, \mathcal{B}, \lambda)$ be a measurable space, and let $X$ be an $X$-valued random variable whose law has the density $f$ with respect to measure $\lambda$. Assume that we observe $X^{(n)} = (X_1, \ldots, X_n)$, where $X_i, i = 1, \ldots, n$, are independent copies of $X$. The goal is to estimate the functional

$$
\Psi(f) = G\left(\int_X H(f(x))\lambda(dx)\right),
$$

where as before $G : \mathbb{R} \to \mathbb{R}$ and $H : \mathbb{R}_+ \to \mathbb{R}$ are fixed functions.

Let $\mathcal{M}$ be a finite set of indices with cardinality $M$, possibly dependent on $n$. Let $f_0, \Lambda_m : X \to \mathbb{R}_+$ and $\Pi_m \in \mathcal{B}, m \in \mathcal{M}$ be collections of measurable functions and sets satisfying the following conditions.

(a) $\Pi_m \cap \Pi_k = \emptyset$ for any $m \neq k, m, k \in \mathcal{M}$;
(b) $\lambda(\Pi_m) = \sigma > 0$ for any $m \in \mathcal{M}$;
(c) $\Lambda_m(x) = 0, x \notin \Pi_m$ for any $m \in \mathcal{M}$;
(d) $\int_{\Pi_m} \Lambda_m(x)\lambda(dx) = 1$ for any $m \in \mathcal{M}$;
(e) $\int_X f_0(x)\lambda(dx) = 1$ and $f_0(x) = 0$ for any $x \in \bigcup_{m \in \mathcal{M}} \Pi_m$.

Under these conditions some evident minor modifications in definitions should be made; for instance, function $S$ should be defined as

$$
S(z) = M^{-1} \sum_{m \in \mathcal{M}} \int_X H(z\Lambda_m(x))\lambda(dx).
$$

With these changes the results of Theorem 4 and Corollary 1 remain valid.

**Method of proof** The following main ideas lie at the core of the proof of Theorem 4 and Corollary 1.

The first idea goes back to the paper Lepski et al. (1999). It reduces the original estimation problem to a problem of testing two composite hypotheses for mixture distributions which are
obtained by imposing prior probability measures with intersecting supports on parameters of a
functional family. In Tsybakov (2009) this technique is called the method of two fuzzy hypotheses.
The choice of the prior measures is based on the moment matching technique; see also Cai and
Low (2011), Wu and Yang (2016) and Han et al. (2020), where further references can be found.
We clarify the moment matching technique in Propositions 1–4 of Section 4. Detailed proofs of
these statements are given in the supplementary paper Goldenshluger and Lepski (2020b).

The second idea is related to construction of a specific parameterized family of densities on
which the lower bound on the minimax risks is established. Here we use a construction that is
similar to the one proposed in Goldenshluger and Lepski (2014).

The third main idea is related to the analysis of the so-called Bayesian likelihood ratio. This
analysis, being common in problems of estimating different nonlinear functionals, depends on
the considered statistical model. The multivariate density model on $\mathbb{R}^d$ requires development of
the original technique because, in contrast to the Gaussian white noise or regression models, the
Bayesian likelihood ratio is not a product of independent random variables. As a consequence, the
standard methods based on computation of the Kullback-Leibler, Hellinger or other divergences
between distributions are not directly applicable in the analysis of the Bayesian likelihood ratio.
That is why our proof of Theorem 4 develops two–sided product–form bounds on this quantity.

Remark 2. It worth mentioning that another technique for derivation of lower bounds in the
problem of estimating the entropy has been developed in Wu and Yang (2016), and Han et al.
(2020). This technique is based on Poissonization and reduction by sufficiency, and it allows one
to reduce the model to the scheme with independent observations. Then this setting can be handled
by methods developed in the literature. For more details we refer the reader to the aforementioned
papers.

3.6. Additional results

In this section we discuss implications of our results and the developed technique for other
problems of estimating nonlinear functionals. In all examples below we consider functionals $\Psi(f)$
of type (3.1) with $G(y) \equiv y$. Denote also

$$
F_\Psi := \{ f : |\Psi(f)| < \infty \}.
$$

Estimation of $\| f \|_p^p$, $p \in \mathbb{N}^*$

In this example $H(y) = y^p$, $p \in \mathbb{N}^*$. Apparently, the case $p = 2$ is the most well studied setting.
Many authors, starting from the seminal paper of Bickel and Ritov (1988), made fundamental
contributions to the minimax and minimax adaptive estimation of quadratic functionals of prob-
ability density; see, for instance, Birgé and Massart (1995), Laurent (1996, 1997) among many
others. Kerkyacharian and Picard (1996) studied the case $p = 3$ and Tchetgen et al. (2008)
considered the setting with arbitrary integer $p$. It is worth noting that all aforementioned pa-
pers consider either univariate or compactly supported densities belonging to a semi–isotropic
functional class, that is $r_l = r$ for all $l = 1, \ldots, d$.

Let us consider the case $p = 2$ and recall one of the most well known results. Assume that
the underlying density $f$ is compactly supported and belongs to the anisotropic H"older class
$\mathbb{N}_{\infty, d}^r(\vec{r}, \vec{L})$, that is $r_l = \infty$ for all $l = 1, \ldots, d$. It is well known that in this setting the minimax
rate of convergence in estimating $\| f \|_2^2$ is given by

$$(1/n) \frac{4d}{d+1} \frac{1}{r_1};$$
see, e.g., Bickel and Ritov (1988) for one-dimensional case, and Tchetgen et al. (2008) for the multivariate one. In particular, the parametric regime is possible if and only if \( \beta \geq 1/4 \). On the other hand, close inspection of the proof of Theorem 1 shows that the lower bound on the minimax risk in estimating of \( \|f\|_2^2 \) is simply the squared rate found in this theorem in the "nonparametric" regime. In particular, if \( r_l = \infty \) for all \( l = 1, \ldots, d \) then Theorem 1 yields the rate \((1/n)^{2\beta + 1}/2\). We do not know whether this rate is the minimax rate of convergence, but we can assert that the parametric rate is not possible if \( \beta \leq 1/3 \). This shows that problems of estimating \( \|f\|_2^2 \) for compactly supported densities, and densities supported on the entire space \( \mathbb{R}^d \) are completely different.

Another interesting feature is that if \( q = p, r^* \leq p \) and \( \tau(p) \leq 0 \) then uniformly consistent estimators of \( \|f\|_2^2 \) over anisotropic Nikolskii’s class do not exist. This phenomenon is again due to the fact that the underlying density is supported on the entire space \( \mathbb{R}^d \).

**Estimation of the differential entropy**

This setting corresponds to \( H(y) = -y \ln(y) \). Applying the same reasoning as in the proof of Theorem 2 in conjunction with Proposition 3 we are able to prove the following statement.

**Theorem 5.** There exists \( c > 0 \) such that for any \( \vec{\beta} \in (0, \infty)^d, \vec{L} \in (0, \infty)^d, \vec{r} \in [1, \infty]^d \)

\[
\liminf_{n \to \infty} \ln n \inf_{\tilde{F}} \sup_{F \in \Psi \cap \mathcal{N}_{r,d}(\vec{\beta}, \vec{L})} \left( \mathbb{E}_f [\tilde{F} - \Psi(f)]^2 \right)^{1/2} \geq c.
\]

**Estimation of** \( \|f\|_p^p, p \in (0, 1) \)

Let \( H(y) = y^p, p \in (0, 1) \). The corresponding functional coincides up to a constant with the Tsallis entropy with index \( p \in (0, 1) \).

**Theorem 6.** For any \( p \in (0, 1) \) there exists \( c > 0 \) such that for any \( \vec{\beta} \in (0, \infty)^d, \vec{L} \in (0, \infty)^d \) and \( \vec{r} \in [1, \infty]^d \)

\[
\liminf_{n \to \infty} \ln n^{2p} \inf_{\tilde{F}} \sup_{F \in \Psi \cap \mathcal{N}_{r,d}(\vec{\beta}, \vec{L})} \left( \mathbb{E}_f [\tilde{F} - \Psi(f)]^2 \right)^{1/2} \geq c.
\]

The rates of convergence established in Theorems 5 and 6 are very slow and do not depend on the parameters of the functional class. In particular, these results demonstrate that smoothness alone is not sufficient in order to guarantee a "reasonably good" accuracy of estimation. For related results on non–existence of uniformly convergent estimators in the problem of estimating the entropy of discrete distributions on large alphabets we refer to Wu and Yang (2016). On the other hand, the recent paper Han et al. (2020) dealing with estimation of the differential entropy of a probability density shows that the minimax risk converges to zero at the polynomial (in \( n \)) rate provided that the underlying density satisfies moment conditions. It is also clear that the polynomial rate of convergence in estimating of the considered functionals is possible for smooth compactly supported densities. The proof of Theorems 5 and 6 goes along the same lines as the proof of Theorems 2–3, and it is omitted.
4. Construction of prior probability measures with prescribed properties

As we have seen, one of the crucial elements of our approach consists in finding prior probabilities measures belonging to $\mathcal{P}[0,1]$ and satisfying Definition 2. In some cases these measures should possess additional properties. In particular, the following condition is required in the proof of Theorems 1–3:

$$\sqrt{e_\pi(2z)} \leq 2e_\pi(z), \quad \forall z \in \{r_1, \ldots, r_d, p, q\}, \quad \pi \in \{\mu, \nu\}. \tag{4.1}$$

Recall that $r_1, \ldots, r_d$ are the coordinates of the vector $r$ used in the definition of the Nikol’skii class. By convention, here and from now on we put $|e_\pi(z)|^{1/2} = 1$ for $z = \infty$.

In this section we present results that establish existence and provide explicit construction of probability measures with required properties. The proofs of these results are given in the supplementary paper Goldenshluger and Lepski (2020b).

**Proposition 1.** For any $s \in \mathbb{N}^*$, $s > 1$ one can construct a pair of probability measures $\mu, \nu \in \mathcal{P}[0,1]$ such that $e_\mu(z), e_\nu(z) \geq 1/2$ for any $z > 0$, $\mu \sim \nu$ and

$$e_\mu(s) - e_\nu(s) \geq C_s := \sqrt{2s - 1}[(s - 1)!][2s - 1]!^{-1}.$$

We remark that the measures $\mu, \nu$ constructed in Proposition 1 obviously satisfy the requirement (4.1).

Let $\Psi = \{\psi_0, \psi_1, \ldots, \psi_s\}$, $\psi_0 \equiv 1$, be a family of continuous functions defined on bounded interval $I \subset \mathbb{R}$. For given continuous function $S : I \to \mathbb{R}$ let

$$\varpi_s(S, I, \Psi) := \inf_{(c_0, \ldots, c_s) \in \mathbb{R}^{s+1}} \sup_{x \in I} \left| S(x) - \sum_{k=0}^{s} c_k \psi_k(x) \right|.$$

**Proposition 2.** Suppose that functions from $\Psi$ are linearly independent on $I$, and let $\varpi_s(S, I, \Psi) > 0$. Then one can construct two probabilities measures $\mu$ and $\nu$ with total mass on $I$ having the following properties:

$$\int_I \psi_k(x) \mu(dx) = \int_I \psi_k(x) \nu(dx), \quad k = 1, \ldots, s;$$

$$\int_I S(x) \mu(dx) - \int_I S(x) \nu(dx) = 2\varpi_s(S, I, \Psi).$$

Proposition 2 reduces construction of probability measures with required properties to the choice of approximation system $\Psi$, and computation of the accuracy of the best approximation of some function $S$ by the span of $\Psi$. In particular, if $\Psi^* = \{1, x, x^2, \ldots, x^s\}$ then $\varpi_s(S, I, \Psi^*)$ is the accuracy of best approximation of $S$ on $I$ by algebraic polynomials of degree $s$.

Lower and upper bounds for the accuracy of best polynomial approximation $\varpi_s(S, [0,1])$ are known for many continuous functions $S$. In particular, combining Proposition 2 with lower bounds on the accuracy of the best polynomial approximation in Timan (1963), §7.1.41 and §7.5.4 we come to the following statements.

**Proposition 3.**

1. For any $p > 0$, $p \notin \mathbb{N}^*$, there exists constant $C_p > 0$ and probability measures $\mu, \nu \in \mathcal{P}[0,1]$ such that $\mu \ast^{s+1} \nu$, $e_\mu(z), e_\nu(z) \geq 1/2$ for any $z \geq 1$, and

$$e_\mu(p) - e_\nu(p) \geq C_p s^{-2p}, \quad \forall s \in \mathbb{N}^*.$$
2. There exist \( C > 0 \) and \( \mu, \nu \in \mathcal{P}[0, 1] \) such that \( \mu \overset{s+1}{\sim} \nu \), and
\[
\int_0^1 x \ln(x) \mu(dx) - \int_0^1 x \ln(x) \nu(dx) \geq C s^{-2}, \quad \forall s \in \mathbb{N}^*.
\]

Propositions 1-3 are applicable for derivation of lower bounds in many problems of estimating nonlinear functionals. However, sometimes it is desirable to have prior probability measures with properly small moments [see the proofs of Theorems 2 and 3]. In this setting we will use the following result.

**Proposition 4.** For any \( p \notin \mathbb{N}^* \), \( p > 1 \), \( s \in \mathbb{N}^* \) and \( \ell \in \mathbb{N}^* \), \( \ell > 0 + 1 \), one can construct probability measures \( \mu, \nu \in \mathcal{P}[0, 1] \) such that \( \mu \overset{s+\ell+1}{\sim} \nu \) and
\[
e_{\mu}(p) - e_{\nu}(p) \geq 2 x^\lambda R s^{-2p},
\]
\[
\frac{1}{2} x^\lambda s^{-2z} \leq e_{\pi}(z) \leq x^\lambda s^{-2z}, \quad \forall z \in [1, \ell], \quad \pi \in \{\mu, \nu\},
\]
where \( \lambda > 0 \) and \( R > 0 \) are constants independent of \( s \) and completely determined by \( p \) and \( \ell \).

Obviously, probability measures \( \mu, \nu \) constructed in Proposition 4 satisfy the requirement (4.1).

The main step in the proposition proof is the derivation of a non-trivial lower bound on the accuracy of the best approximation of function \( x \mapsto x^{p-\ell} \) by rational functions, established in Goldenshluger and Lepski (2020b).

5. Proofs of Theorems 1–3

We will prove only "nonparametric rates". A lower bound corresponding to the parametric rate of convergence \( n^{-1/2} \) can be easily derived by reduction of the considered problem to parameter estimation in a regular statistical model.

5.1. Preliminary remarks

The proofs of Theorems 1–3 are based on application of Corollary 1, and contain many common elements. In particular, in all proofs we consider parameterized family of functions \( \{f_w, w \in [0, 1]^M\} \) defined in (3.2)–(3.5), and we choose the sets \( \Pi_0 \) and \( \Pi_m \), \( m \in \mathcal{M} \) so that \( \Pi_0 \subset (-\infty, 0]^d \) and \( \Pi_m \subset [0, \infty)^d \) for all \( m \in \mathcal{M} \). We equip the parameter set \([0, 1]^M\) with a pair of probability measures \( \mu, \nu \) satisfying conditions of Definition 2. Along with conditions of Definitions 2 in the proofs of Theorems 1–3 we require that the specified probability measures \( \mu \) and \( \nu \) possess the property (4.1). Once measures satisfying conditions of Definition 2 are constructed, they can be easily modified to satisfy (4.1); for details we refer to the proofs of Propositions 1 and 4.

The parameters \( A, \tilde{\sigma}, M \) of the family \( \{f_w, w \in [0, 1]^M\} \) are specified to guarantee that under imposed prior measures random functions \( f_\zeta \) satisfy required smoothness conditions with the probability controlled by parameter \( \nu \). Under these circumstances, parameter \( \nu \) of Assumption 2 should be such that constant \( C_* \) in Corollary 1 is strictly positive. For instance, the choice \( \nu = 64(d + 1) \) is sufficient and assumed throughout the proof.

In what follows \( C_1, C_2, \ldots \), and \( c_1, c_2, \ldots \), denote constants that may depend on \( \tilde{\beta}, \tilde{r}, \tilde{L}, q, Q \) and \( \Lambda \), but they are independent of \( n \).
5.2. Verification of Assumption 1

Let \( C := \int_{-1}^{1} e^{-\frac{|x|}{2}} dx \), and
\[
U(x) := C^{-d} e^{-\sum_{j=1}^{d} \frac{1}{2} |x_j|^2} 1_{[-1,1]^d}(x), \quad x = (x_1, \ldots, x_d) \in \mathbb{R}^d.
\]

For \( N > 0 \) and \( a > 0 \) define
\[
\tilde{f}_{0,N}(x) := (N)^{-d} \int_{\mathbb{R}^d} U(y-x) 1_{[-N-1,-1]^d}(y)dy, \quad f_{0,N}(x) := a^d \tilde{f}_{0,N}(xa).
\]

**Lemma 1.** The following statements hold.

(a) For any \( N \) and \( a \), \( f_{0,N} \) is a probability density. For any \( \vec{\beta}, \vec{L} \in (0, \infty)^d \) and \( \vec{r} \in (0, \infty)^d \) there exists \( a > 0 \) such that
\[
f_{0,N} \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \frac{1}{2} \vec{L}), \quad \forall N > 0.
\]

(b) For any \( Q > 0 \) and \( q \in (r^*, \infty] \) there exists \( N(q,Q) > 0 \) such that
\[
f_{0,N} \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \frac{1}{2} \vec{L}) \cap \mathbb{B}_q(\frac{1}{2} Q), \quad \forall N \geq N(q,Q).
\]

The proof of Lemma 1 is trivial; it is omitted.

For \( N \geq N(q,Q) \) let
\[
f_0 := f_{0,N}.
\] (5.1)

The statements (a) and (b) of Lemma 1 hold for \( f_0 \) defined in (5.1). For a pair of probability measures \( \mu, \nu \in \mathcal{P}[0,1] \) and real number \( z > 0 \) we write \( e^*(z) := \max[e_\mu(z), e_\nu(z)] \).

**Lemma 2.** Let \( f_0 \) be given in (5.1), \( f_w \) be defined in (3.5) with \( \Lambda \in \mathcal{C}_C(\mathbb{R}^d) \) satisfying (3.2), and \( \tilde{\sigma} \in (0,1]^d \). Let \( \mu, \nu \in \mathcal{P}[0,1] \) satisfy (4.1) and assume that
\[
A \|A\|_q[\sigma Me^*(q)]^{1/d} \leq \frac{1}{4} Q, \tag{5.2}
\]
\[
A \sigma^{-\beta_l}_{l} \|A\|_{r_l}[\sigma Me^*(r_l)]^{1/r_l} \leq C l L_l, \quad \forall l = 1, \ldots, d. \tag{5.3}
\]

Then Assumption 1 is fulfilled for \( \mathcal{F} = \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}) \cap \mathbb{B}_q(Q) \) with \( \varepsilon = 2^{-6} \).

**Proof of Lemma 2.** Note that the construction of the set of functions \( \{f_w, w \in \{0,1\}^M\} \) almost coincides with the construction proposed in Goldenshluger and Lepski (2014) in the proof of Theorem 3. Therefore, denoting \( F_w = \sum_{m \in \mathcal{M}} w_{m} \Lambda_{m} \) and repeating computations in the cited paper we can verify that assumption
\[
A \sigma^{-\beta_l}_{l} \|A\|_{r_l}[\sigma w(r_l)\sigma]^{1/r_l} \leq C l L_l, \quad \forall l = 1, \ldots, d, \tag{5.4}
\]

with \( \tilde{\sigma} \in (0,1]^d \) guarantees \( F_w \in \mathbb{N}_{\vec{r},d}(\vec{\beta}, \frac{1}{2} \vec{L}) \), \( w \in \{0,1\}^M \). It is important to realize that the only conditions used in the proof of (5.4) is (3.2) and \( \Lambda \in \mathcal{C}_C(\mathbb{R}^d) \) which are the same as in Goldenshluger and Lepski (2014). Define for \( z \geq 0 \)
\[
\mathcal{W}_{\pi,z} := \{ w \in \{0,1\}^M : |\sigma w(z) - Me_{\pi}(z)| \leq \sqrt{vMe_{\pi}(2z)} \}.
\]

First we note that if for some \( z \geq 1 \) the event \( \{ \zeta \in \mathcal{W}_{\pi,z} \} \) is realized then
\[
\sigma_{\zeta}(z) \leq Me_{\pi}(z) + \sqrt{vMe_{\pi}(2z)} = e_\pi(z)[M + \sqrt{2vM}] \leq 2Me_{\pi}(z) \leq 2Me^*(z), \tag{5.5}
\]
because \(e_\epsilon(z) \leq 1\), and \(M \geq 36 \epsilon^2\) in view of (3.7). Also we have used (4.1). Next, it follows from (5.5) that if \(\{\zeta \in W_{\pi, z}, z \in [1, \infty)\} \) is realized then
\[
\|F_\zeta\|_z = \|\Lambda\|_z (\sigma \varphi_\zeta(z))^{1/z} \leq 2 \|\Lambda\|_z (\sigma M e^* (z))^{1/z}.
\]
It yields in particular, on the event \(\{\zeta \in W_{\pi, q}\}\)
\[
\|f_\zeta\|_q = \left[1 - A\sigma \varphi_\zeta(1)\right] \|f_0\|_q + A \|F_\zeta\|_q \leq \frac{1}{2} Q + 2A \|\Lambda\|_q (\sigma M e^*(q))^{1/q} \leq Q.
\]
To get the last inequality we have used (5.2) and the fact that \(\varphi_\zeta\) because
\[
A \|\Lambda\|_{r_i} (\sigma \varphi_\zeta(r_i))^{1/r_i} \leq 2A \|\Lambda\|_{r_i} (\sigma M e^*(r_i))^{1/r_i} \leq 2c_1 L_i.
\]
For the indices \(l \in \{1, \ldots, d\}\) for which \(r_l = \infty\), the last inequality obviously hold for all realizations of \(\zeta\).

It follows from (5.4) with \(C_1 = 2c_1\) that \(F_\zeta \in \mathcal{N}_{r,d}(\beta, \frac{1}{2} L)\); hence \(f_\zeta \in \mathcal{N}_{r,d}(\beta, L)\) because \(f_\zeta = [1 - A\sigma \varphi_\zeta(1)] f_0 + F_\zeta\) and \(f_0 \in \mathcal{N}_{r,d}(\beta, \frac{1}{2} L)\) in view of the second statement of Lemma 1. Thus we have shown that

\[
\left\{\zeta \in W_{\pi, z}\right\} \subset \left\{f_\zeta \in \mathcal{N}_{r,d}(\beta, L) \cap \mathcal{B}_q(Q)\right\}.
\]

Hence
\[
\mathbb{P}_z\{f_\zeta \in \mathcal{N}_{r,d}(\beta, L) \cap \mathcal{B}_q(Q)\} \geq 1 - \sum_{z \in \{r_1, \ldots, r_d, q\}} \mathbb{P}_z\{\zeta \notin W_{\pi, z}\} \geq 1 - \frac{1}{2} (d + 1) = \frac{d+1}{2d}.
\]
The last inequality follows from (4.1). Therefore
\[
H^*_x = [1 - A\sigma Me^*(1)]^p \|f_0\|^p_p + A^p \sigma^p \|\Lambda\|^p_p M e^*(p)\int.
\]
Since the choice of parameters \(A, \sigma\) and \(M\) satisfies (3.8), \(1 - A\sigma Me^*(1) \geq 1 - \epsilon_0 > 1/2\). By definition of \(f_0\), \(\|f_0\|^p_p = c_1 N^{-d(p-1)}\), and \(N\) can be chosen arbitrarily large. In particular, denoting \(e_\epsilon(p) := \min\{e_\epsilon(p), e_\nu(p)\}\) and choosing \(N\) so that
\[
N^{-d(p-1)} = c_2 A^p \|\Lambda\|^p_p e_\epsilon(p) \sqrt{vM}
\]
Finally, using the elementary inequality
\[ A^p \sigma \| \mathcal{A}^p M e_x(p) \| \leq H^*_\pi \leq A^p \sigma \| \mathcal{A}^p M e_x(p) + A^p \sigma \| \Lambda^p e_x(p) \sqrt{v M}. \] (5.6)

2\textsuperscript{nd}. Note that the conditions (3.6) and (3.8) imply
\[ n(A^\sigma)^2 M e_x(2) \leq \alpha_0^2. \] (5.7)
Indeed, (3.6)(i) implies (5.7) because \( e_x(2) \leq 1 \). On the other hand, in view of (3.6)(ii), (3.8) and (4.1)
\[ n(A^\sigma)^2 M e_x(2) \leq \alpha_0^2 \sqrt{e_x(2)} \ln (1 + e_x^{-1}(2)) < \alpha_0^2 \]
because \( \sqrt{\ln (1 + 1/x)} < 1 \) for \( x \in [0, 1] \). Furthermore,
\[ |S_0(z) - S_0(z')| = |f_0|^p |(1 - z)^p - (1 - z')^p| \leq p |f_0|^p |z - z'|, \quad \forall z, z' \in [0, 1], \]
so that
\[ \eta_{\text{c}}(A^\sigma M e_x(1); A^\sigma \sqrt{v M} e_x(2)) \leq p |f_0|^p A^\sigma \sqrt{v M} e_x(2) \leq c_3 N^{-d(p-1)} A^\sigma \sqrt{v M}. \]
Taking into account that \( N^{-d(p-1)} = c_2 A^p \sigma \| \Lambda^p e_x(p) \sqrt{v M} \) we obtain
\[
\begin{align*}
\alpha_\pi & \leq c_4 A^{p+1} \sigma^2 \| \Lambda^p e_x(p) v M + 2 A^p \sigma \| \Lambda^p e_x(p) \sqrt{v M} \\
& \leq 2 A^p \sigma \| \Lambda^p e_x(p) \sqrt{v M} (1 + (c_4/2) A^\sigma \sqrt{v M}) \leq 3 A^p \sigma \| \Lambda^p e_x(p) \sqrt{v M},
\end{align*}
\]
where in the last inequality we have used (5.7), and the fact that \( \alpha_0 \) is sufficiently small. Therefore (5.6) and (3.7) imply
\[
H^*_\pi - \alpha_\pi \geq A^p \sigma \| \mathcal{A}^p M e_x(p) (1 - 3 \sqrt{v M}) \geq A^p \sigma \| \mathcal{A}^p M e_x(p) (1 - \frac{1}{2 \sqrt{v}}),
\]
\[
H^*_\pi + \alpha_\pi \leq A^p \sigma \| \mathcal{A}^p M e_x(p) (1 + 4 \sqrt{v M}) \leq A^p \sigma \| \mathcal{A}^p M e_x(p) (1 + \frac{2}{3 \sqrt{v}}).
\]

4\textsuperscript{th}. Since \( G(x) = x^{1/p} \) we can assert that
\[ \mathcal{J}_\pi \subseteq \left[ A(M\sigma)^{1/p} \| \mathcal{A}_p [e_x(p)]^{1/p} \right] \frac{1}{1 - \frac{1}{2 \sqrt{v}}} \right]^{1/p}, \quad A(M\sigma)^{1/p} \| \mathcal{A}_p [e_x(p)]^{1/p} \left( 1 + \frac{2}{3 \sqrt{v}} \right)^{1/p}, \]
and denoting \( [y]_+ = \max[0, y] \) we get
\[ \Delta(\mu, \nu) \geq A(M\sigma)^{1/p} \| \mathcal{A}_p [e^*(p)]^{1/p} \left( 1 - \frac{1}{2 \sqrt{v}} \right)^{1/p} - [e_x(p)]^{1/p} \left( 1 + \frac{2}{3 \sqrt{v}} \right)^{1/p} \]
Therefore condition
\[ |e_{\mu}(p) - e_x(p)| > \frac{\epsilon^*(p)}{\sqrt{v}}, \] (5.8)
we can guarantees \( \Delta(\mu, \nu) > 0 \), and
\[ \Delta(\mu, \nu) \geq A(M\sigma)^{1/p} \| \mathcal{A}_p [e^*(p)]^{1/p} \left( 1 - \frac{1}{2 \sqrt{v}} \right)^{1/p} - [e_x(p)]^{1/p} \left( 1 + \frac{2}{3 \sqrt{v}} \right)^{1/p}. \]
Finally, using the elementary inequality
\[ \frac{1}{p} (a \lor b)^{-1+1/p} |a - b| \leq |a|^{1/p} - b^{1/p} \quad \forall a, b > 0, \quad p \geq 1 \]
with \( a = e^*(p)(1 - \frac{1}{2 \sqrt{v}}) \) and \( b = e_x(p)(1 + \frac{2}{3 \sqrt{v}}) \) we obtain
\[ \Delta(\mu, \nu) \geq c_5 A(M\sigma)^{1/p} [e_{\mu}(p) - e_x(p)] [e^*(p)]^{1/p-1}. \] (5.9)
5.4. Choice of probability measures

In the subsequent proofs depending on the setting we make three different choices of probability measures \( \mu \) and \( \nu \).

1. **The case of integer \( p \):** For \( p \in \mathbb{N}^* \) we always use the measures constructed in Proposition 1 with \( s = p \).
2. **The case of non-integer \( p \):** For \( p \notin \mathbb{N}^* \) we use one of the following two choices:
   a. \( \mu \) and \( \nu \) are constructed in Proposition 3 with \( s = 2|p| \);
   b. \( \mu \) and \( \nu \) are constructed in Proposition 4 with \( t = p \) and \( \ell = \lfloor |p| + 2 \rfloor \) if \( q < \infty \), and \( \ell = |p| + 2 \) if \( q = \infty \). The parameter \( s \geq 2|p| \) is specified later.

Note that in the cases (1) or (2a) one has \( |e_\mu(p) - e_\nu(p)| = c_1 \) so that (5.8) holds for sufficiently large \( v > 0 \). In addition, it follows from (5.9) that in these both cases

\[
\Delta(\mu, \nu) \geq c_2 A(M\sigma)^{1/p}.
\]  

In the case (2b) for any \( s \in \mathbb{N}^* \)

\[
|e_\mu(p) - e_\nu(p)| \geq 2 \varphi'(s)^{-2p}, \quad e^*(p) \leq \varphi^p s^{-2p}
\]

and, therefore, (5.8) is verified for any \( s \in \mathbb{N}^* \) if \( v > 2\varphi^{-p} \). Additionally we deduce from (5.9) that in this case

\[
\Delta(\mu, \nu) \geq c_3 A(M\sigma)^{1/p} s^{-2p}, \quad \forall s \in \mathbb{N}^*.
\]

5.5. Generic choice of parameters

In this part we specify parameters \( A \) and \( \bar{\sigma} \) in terms of of \( M, t, n \) so that they satisfy conditions (3.6)–(3.8), (5.2)–(5.3), and (3.17).

Denote \( \epsilon = [e^*(q)]^{1/q}, q = 2 \lor q \) and recall that \( \epsilon = 1 \) if \( q = \infty \). For any \( t \in (1, \infty] \) and \( \varsigma < 1 \) we let

\[
\sigma_1 = c_1^{1/\beta} L_t^{-1/\beta} L^{1/\beta} \mathcal{M}^{1/\beta} M^{1/\gamma} \mathcal{N}^{1/\gamma} \tau^{(r_1)/(r_1 + 1)},
\]

\[
B = c_2 \epsilon^{-1} L_t^{-1/\gamma} \mathcal{M}^{1/\gamma} M^{1/\gamma} \tau^{(r_1)/(r_1 + 1)} \mathcal{N}^{1/\gamma},
\]

Recall that \( \beta, \tau(\cdot), \mathcal{L} \) and \( \omega \) are the quantities that are defined via parameters of the functional class \( N_{\mathcal{F},d}(\bar{\beta}, \bar{\mathcal{L}}) \) in (2.1). These formulas imply

\[
\sigma = \prod_{l=1}^d \sigma_1 = c_1^{1/\beta} L_t^{-1/\beta} \mathcal{M}^{1/\beta} M^{1/\gamma} \tau^{(r_1)/(r_1 + 1)} \mathcal{N}^{1/\gamma}, \quad A = c_2 \epsilon^{-1} c_1^{1/\beta} \varsigma M^{-1/t}.
\]

In addition,

\[
A(\sigma M)^{1/p} = c_3 \epsilon^{-1} L_t^{1/\gamma} \mathcal{M}^{1/\gamma} M^{1/\gamma} \tau^{(r_1)/(r_1 + 1)} (\varsigma M^{-1/t})^{1/\gamma}.
\]  

Note that the lower bound on \( \delta(\mu, \nu) \) is expressed in terms of the quantity \( A(\sigma M)^{1/p} \) [see (5.10) and (5.11)].
1. Assume that constant $c_1 > 0$ is chosen to satisfy
\[
\max_{l=1,\ldots,d} c_1^{1/\beta} L_l^{-1/\beta} L_l^{1/\beta - 1/(\beta r_l)} = 1,
\] (5.13)
while the constant $c_2$ is sufficiently small in order to guarantee
\[
c_2 c_1^{1/(3r^*) - 1} \leq C_1, \quad c_2 c_1^{1/\beta} \leq \kappa_0^2 < 1, \quad c_2 c_1^{1/(\beta q)} \leq \frac{1}{4} L^{1/\alpha_1} Q, \]
(5.14)
where $C_1$ is a constant appearing in (5.3). Simple algebra shows that (5.3) holds if $c_2 c_1^{1/(3r^*) - 1} \leq C_1$ because \(\lceil c^*(r_l) \rceil_{r_l} \leq c\) for all $l = 1, \ldots, d$, and $q \geq r^*$. Moreover, conditions (3.6) and (3.8) are reduced respectively to
\[
\text{(i)} \quad nA^2 \sigma^2 M = c_2 c_1^{2/\beta} M^{1-2/\beta} \kappa^2 n \leq \kappa_0^2 \quad \Rightarrow \quad M^{1-2/\beta} \kappa^2 n \leq 1;
\]
\[
\text{(ii)} \quad nA \sigma = c_2 c_1^{1/\beta} n \kappa M^{-1/t} \leq \kappa_0^2 \ln(1 + 1/e) \quad \Rightarrow \quad n \kappa M^{-1/t} \leq \ln(1 + 1/e);
\]
\[
A \sigma M \sqrt{e_{\pi}(2)} \leq c_2 c_1^{1/\beta} \kappa c M^{1-1/t} \leq \kappa_0 \quad \Rightarrow \quad \kappa c M^{1-1/t} \leq 1;
\]
while (5.2) becomes
\[
A \ell_{\sigma M}^{1/q} = c_2 c_1^{1/(\beta q)} L^{1-1/q} \left[ \kappa_1^{r(q)} \right] M^{1/q - 1/(1/q + \beta q)} \leq \frac{1}{4} Q
\]
and, therefore, it is reduced to
\[
\left[ \kappa_1^{r(q)} \right] M^{1/q - 1/(1/q + \beta q)} \leq 1.
\]
(5.17)

2. Our current goal now is to show that under (5.15)–(5.17) one has
\[
\sigma_l \leq 1, \quad \forall l = 1, \ldots, d.
\]
(5.18)
This condition is required in construction of the family \(\{f_w\}\). We will distinguish between two possible cases.

(a) Let $\tau(q) \geq 0$; then $\tau(r_l) \geq 0$ for all $l = 1, \ldots, d$ because $q \geq r^*$, and in view of (5.13)
\[
\sigma_l \leq \left[ \kappa_c M^{1-1/t} \right] \frac{\tau(q)}{\beta q} M^{1/q - \frac{1}{1/q + (1/q + \beta q)}} \leq M^{1/r_l - 1} \leq 1.
\]
The first inequality follows from (5.16), while the second one is a consequence of $M \geq 1$ and $r_l \geq 1$.

(b) Now let $\tau(q) < 0$. First we note that in this case in view of (5.17)
\[
\kappa_c \leq M^{-1/q - 1/(1/q + \beta q)} \quad \Leftrightarrow \quad \kappa_c M^{1-1/t} \leq M^{-1/q - 1/(1/q + \beta q)}.
\]
On the other hand
\[
\sigma_l \leq \left[ \kappa_1^{r(q)} \right] \frac{\tau(q)}{\beta q} \left[ \kappa_1^{r(q)} \right] M^{1/r_l - 1/q} \frac{1}{1 - \frac{1}{1/q + \beta q} - \frac{1}{1/q + \beta q}} M^{1/r_l - 1/q} \frac{1}{1 - \frac{1}{1/q + \beta q} - \frac{1}{1/q + \beta q}} \leq M^{1/r_l - 1/q} \left[ 1 - \frac{1}{1/q + \beta q} \right] \leq M^{1/r_l - 1/q} \frac{1}{1/q + \beta q} \leq 1,
\]
since $q \geq r_1$ for any $l = 1, \ldots, d$, $M > 1$ and $\tau(q) < 0$. Thus, (5.18) is proved.

3°. Now let us turn to the condition (3.17) of Corollary 1; it will be verified separately for two different cases.

(a). Choose $t = p$ in (3.17). In view of the second equality in (5.14) condition (3.17) is reduced to

$$
eq 2 c_1^{1/\beta} n \varsigma M^{1/p-1/t} \leq p \varsigma^1 \Leftarrow n \varsigma M^{1/p-1/t} \leq 1,$$

if $c_2 > 0$ is sufficiently small. Letting $t = p$ and $\varsigma = n^{-1}$ we guarantee the verification of (3.17).

It remains to note that with this choice of $t$ and $\varsigma$ inequalities (5.15)(i) and (5.16) are reduced to

$$M \leq n^{\varsigma^{-1}}, \quad (5.19)$$

while (5.17) becomes

$$M^{1/q-1+[(1-1/p)\tau(q)]} \leq (n/\epsilon)^{\tau(q)/\tau(q)}. \quad (5.20)$$

To get (5.19) we have also used that $\epsilon \leq 1$.

Assume now that the measures $\mu$ and $\nu$ are constructed in accordance with the choice (1) or (2a) as described in Section 5.4. Here we have $\epsilon = c_4$. Therefore we deduce from (5.10), (5.12) and Corollary 1 that under (5.19), (5.20) and in view of $M \geq 36 c^2$ [see (3.7)] one has

$$\mathcal{R}_n [N_{\tilde{P},d} (\tilde{\beta}, \tilde{L}) \cap B_q(Q)] \geq c_4 L^{\tau(q)/\tau(q)} n^{-\lfloor \tau(q)/2 \rfloor} M^{1-1/p(1/\nu(\beta)-1/\omega)}. \quad (5.21)$$

(b). Assume now that measures $\mu$ and $\nu$ are constructed in accordance with the choice (2b) of Section 5.4. In (3.17) we let $t = \tau/2$ for some $s \geq 2[p]$, and let $t = \infty$. First, note that in this case in view of Proposition 4 and because $\ell \geq q$ as defined in setting (2b) we have $\epsilon = c_5 s^{-2}$ if $q < \infty$, and $\epsilon = 1$ otherwise. Next, (5.15), (5.16), (5.17) and (3.17) take the form

(i) $M \varsigma^2 n \leq 1$;

(ii) $n \varsigma \leq \ln (1 + c_5 s^2)$, $M \geq n$;

$$\varsigma \varsigma M \leq 1; \quad (5.23)$$

$$[\varsigma]^\tau(q)/\tau(q) M^{1/q+\tau(q)} \leq 1; \quad (5.24)$$

$$n \varsigma M^{2/s} \leq 2^{-1} s \varsigma^1. \quad (5.25)$$

Hence we deduce from (5.11), (5.12) and Corollary 1 that under (5.22)–(5.25) and $M \geq 36 c^2$

$$\mathcal{R}_n [N_{\tilde{P},d} (\tilde{\beta}, \tilde{L}) \cap B_q(Q)] \geq c_6 L^{1-1/p} [\varsigma]^\tau(q)/\tau(q) M^{1/q+1/\nu(\beta)-1/\omega} [s^2 \varsigma]^{-1}. \quad (5.26)$$

The results of Theorems 1–3 are derived from lower bounds (5.21) and (5.26).

### 5.6. Lower bounds corresponding to the regime $n^{-\tau(p)/\tau(1)}$

The regime corresponding to the rate of convergence $n^{-\tau(p)/\tau(1)}$ appears in Theorems 1–3, and the corresponding lower bounds are proved in a unified way. This regime takes place if $\tau(q) \geq 0$. Here we choose $M = 36 c^2$ and remark that (5.19) holds for all $n$ large enough. Since $\tau(q) \geq 0$ and $q > 1$, (5.20) holds as well for large $n$. The statemets of Theorems 1–3 in the cases $\tau(q) \geq 0$, $\tau(p) \leq 1$ if $p \in \mathbb{N}^*$, and $\tau(q) \geq 0$ $\tau(p) \leq 1 - 1/p$ if $p \notin \mathbb{N}^*$ follow now from (5.21).
5.7. Proof of Theorem 1

Here \( p \) is integer, and we use the lower bound (5.21) together with the preceding restrictions on the choice of parameter \( M \). We consider two cases.

(a) Assume first that \( 1/(\beta p) - 1/\omega \geq 0 \) which is equivalent to \( \tau(p) \geq 1 \). Let \( M := c_7 n^{2/p} \); then (5.19) holds if \( c_7 > 0 \) is sufficiently small, and \( M \geq 36 \omega^2 \) for all large \( n \). Moreover,

\[
[n/e]^{-\tau(q)/\tau(\omega)} M^{1/q-1/(1+p)\tau(q)} = e^{\tau(q)/c_7} M^{1/q-1/(1+p)\tau(q)} n^{1/(1+p)\tau(q)} \leq 1
\]

for all \( n \) large enough because \( q > 1 \). Hence (5.20) holds as well, and we conclude from (5.21)

\[
\mathcal{R}_n [N_{\mathbb{R},d}((\tilde{\beta}, \tilde{L}) \cap B_q(Q)] \geq c_8 L^{1/(1+p)} n^{-\tau(q)/\tau(q) - 1}. \]

(b) Now let us assume that \( 1/(\beta p) - 1/\omega < 0 \), i.e. \( \tau(p) < 1 \), and let \( \tau(q) < 0 \). Here we choose

\[
M := [\tau n]^{1/(1+p)\tau(q)} n^{-\tau(q)/\tau(q)}. \]

With this choice (5.20) holds, and \( M \geq 36 \omega^2 \) for all large \( n \). Moreover,

\[
\frac{\tau(q)}{1 - 1/p} (1 - 1/q) < 0
\]

and, therefore, \( M < n^{\frac{2}{\tau(q)}} \) for all \( n \) large enough. Thus, (5.19) is checked and we conclude from (5.21) that

\[
\mathcal{R}_n [N_{\mathbb{R},d}((\tilde{\beta}, \tilde{L}) \cap B_q(Q)] \geq c_8 L^{1/(1+p)} (1/n)^{1/(1+p)\tau(q) - 1}. \]

The theorem is proved.

5.8. Proof of Theorems 2 and 3

Here \( p \) is non–integer, and we derive the lower bounds using inequality (5.26) and selecting parameters of \( \varsigma \), \( M \) and \( s \) properly.

(a) First we assume first that \( 1/p + 1/(\beta p) - 1/\omega \geq 0 \) which is equivalent to \( \tau(p) \geq 1 - 1/p \). Here we choose

\[
\varsigma = \frac{C_1 \ln(n)}{n}, \quad M = \frac{n}{\varsigma \ln(n)} , \quad s = \frac{C_1 \ln(n)}{\ln(n)} + 1.
\]

Note that (5.22)(ii) and (5.23) are fulfilled for all \( n \) provided that \( C_1 \) is sufficiently small and \( C_2 \) is sufficiently large. It is also obvious that \( M \geq 36 \omega^2 \) for all \( n \) large enough. Next,

\[
[\varsigma e]^{\tau(q)/\tau(\omega)} M^{1/q-1/(1+p)\tau(q)} = e^{\tau(q)/c_7} M^{1/q-1/(1+p)\tau(q)} n^{1/(1+p)\tau(q)} \leq 1
\]

for all \( n \) large enough and, therefore, (5.24) is fulfilled. Finally, (5.25) is equivalent to

\[
\frac{2 \ln(M)}{s} < \ln(s) - \ln \ln(n) - \ln(2C_1/\varsigma_1).\]

Taking into account that \( \epsilon^{-1} \leq c_8 \varsigma^2 \leq c_8 \ln^2(n) \) and that

\[
\frac{2 \ln(M)}{s} \leq (2/C_2) \ln \ln(n) + \frac{\ln(1/\varsigma)}{\ln(n)}, \quad \ln(s) \geq \ln \ln(n) - \ln \ln(n) + \ln(C_2)
\]
and choosing $C_2$ sufficiently large we can assert that and (5.25) holds for all large $n$. Thus, we conclude from (5.26) that

$$R_n[\mathcal{N}_{\vec{r},d}(\vec{b}, \vec{L}) \cap B_q(Q)] \geq c_8 L^{1-1/p} \left( \frac{\ln \ln(n)}{n} \right)^{1-1/p} \left( \frac{s^2 \epsilon}{\tau(1)} \right).$$

If $q < \infty$ then the established inequality yields the corresponding statement of Theorem 2; if $q = \infty$ then we come to the statement of Theorem 3.

(b). Let $1/p + 1/(\beta p) - 1/\omega < 0$, $\tau(q) < 0$ and choose

$$\varsigma = \frac{\ln(n)}{n}, \quad M = \left( \frac{n}{\ln(n)} \right)^{-1/q - \frac{\tau(q)}{1-1/q}}, \quad s = \left[ C_3 \ln(n) \right] + 1.$$

With this choice (5.24) is fulfilled. Moreover $M \geq 36 \upsilon^2$ for all $n$ large enough because $\tau(q) < 0$ and $q > 1$. Note also that

$$\tau(q) \left( \frac{1}{1-1/q} - \frac{1}{1-1/q - \tau(q)} \right) = 1 - \frac{1 - 1/q}{1 - 1/q - \tau(q)} \in (0,1),$$

and, therefore, $M \leq n^\alpha$ for some $\alpha \in (0,1)$. Hence (5.22)(i), and (5.23) are verified for large $n$. Finally, (5.25) is reduced to

$$\frac{2 \ln(M)}{s} \leq \ln \left( C_3/[2\upsilon]\right),$$

and this inequality holds obviously for all $n$ if $C_3$ is sufficiently large. Thus, we conclude from (5.26) that

$$R_n[\mathcal{N}_{\vec{r},d}(\vec{b}, \vec{L}) \cap B_q(Q)] \geq C_7 L^{1-1/p} \left( \ln(n)/n \right)^{1-1/q - \frac{\tau(q)}{1-1/q - \tau(q)}}.$$ 

This inequality yield the corresponding statements of Theorems 2 and 3 in the cases $q < \infty$ and $q = \infty$ respectively.

6. Proofs of Theorem 4 and Corollary 1

6.1. Proof of Theorem 4

We break the proof into several steps.

1\textsuperscript{st}. Product form bounds for $p_\zeta(x)$

Our first step is to develop tight bounds on $p_\zeta(x)$ for all $x \in \mathbb{R}^{dn}$ possessing a product form structure with respect to the coordinates of $\zeta$. Recall that $p_\zeta(x) = \prod_{i=1} f_\zeta(x_i)$ where $f_\zeta$ is defined in (3.5).

For $\pi \in \{\mu, \nu\}$ define the set

$$W_\pi := \left\{ w \in [0,1]^M : |g_w(1) - Me_\pi(1)| \leq \sqrt{vMe_\pi(2)} \right\}.$$  (6.1)

We begin by showing that if $\zeta \in W_\pi$ then $f_\zeta$ is a probability density. By construction $\int f_\zeta = 1$ for any realization of $\zeta$, and we need to show only that $1 - A\sigma g_\zeta(1) > 0$. Recall that $E_\pi g_\zeta(1) = Me_\pi(1)$; then on the event $\{\zeta \in W_\pi\}$ in view of (3.7)

$$|g_\zeta(1) - Me_\pi(1)| \leq \sqrt{vMe_\pi(2)} \leq M \sqrt{\epsilon_\pi(2)},$$  (6.2)
and therefore by (3.8)

$$A \sigma \varrho_\xi(1) \leq A \sigma M e_x(1) + A \sigma M \sqrt{e_x(2)} \leq 2 \kappa_0 < 1.$$ 

Let $\Lambda_0(\cdot) := f_0(\cdot)\{1 - A \sigma \varrho_\xi(1)\}$, and suppose that $\zeta \in W_\pi$. Because $\Pi_0 \cap \Pi_m = \emptyset$ for all $m \in \mathcal{M}$, and $\Pi_j \cap \Pi_m = \emptyset$ for all $m, j \in \mathcal{M}, m \neq j$ we have the following representation of function $f_\zeta(\cdot)$: for any $y \in \mathbb{R}^d$

$$f_\zeta(y) = \Lambda_0(y)1_{\Pi_0}(y) + A \sum_{m \in \mathcal{M}} \zeta_m \Lambda_m(y)1_{\Pi_m}(y) = \Lambda_0(y)1_{\Pi_0}(y) \prod_{m \in \mathcal{M}} [A \zeta_m \Lambda_m(y)]^{1_{\Pi_m}(y)}.$$

Therefore

$$p_\zeta(x) = \prod_{i=1}^n f_\zeta(x_i) = \left[ \prod_{i=1}^n \Lambda_0(x_i)1_{\Pi_0(x_i)} \right] \left[ \prod_{m \in \mathcal{M}} \prod_{i=1}^n [A \zeta_m \Lambda_m(x_i)]^{1_{\Pi_m}(x_i)} \right].$$

If we put

$$T(x) := \prod_{i=1}^n \left\{ [f_0(x_i)]^{1_{\Pi_0(x_i)}} \prod_{m \in \mathcal{M}} \left[ A \Lambda_m(x_i) \right]^{1_{\Pi_m}} \right\}$$

then

$$p_\zeta(x) = T(x)\left[1 - A \sigma \varrho_\xi(1)\right]^{n_\Pi(x)} \prod_{m \in \mathcal{M}} \left[ \zeta_m \right]^{n_m(x)}.$$ (6.3)

Our current goal is to derive bounds on $1 - A \sigma \varrho_\xi(1)$. Assume that $\zeta \in W_\pi$ and denote for brevity

$$b := A \sigma, \quad u := E_{\sigma} \varrho_\xi(1) = Me_x(1), \quad D := b/(1 - bu),$$

where $D$ is given in (3.15). In view of (3.7) and (6.2), $bu = A \sigma Me_x(1) \leq \kappa_0, b[\varrho_\xi(1) - u] \leq \kappa_0$, and also $D \leq (1 - \kappa_0)^{-1}b$. We have

$$1 - b \varrho_\xi(1) = 1 - bu - b[\varrho_\xi(1) - u] = (1 - bu)(1 - D[\varrho_\xi(1) - u]).$$ (6.5)

Using elementary inequality $1 - t \leq e^{-t}$ we obtain from (6.5)

$$1 - b \varrho_\xi(1) \leq (1 - bu)e^{Du}\exp\{-D \varrho_\xi(1)\}.\quad (6.6)$$

On the other hand, taking into account that, by smallness of $\kappa_0$, $D[\varrho_\xi(1) - u] \leq 1$ on the event $\{\zeta \in W_\pi\}$ and applying inequality $1 - t \geq e^{-t} - \frac{1}{2}t^2$, $\forall t \geq -1$ we get

$$1 - D[\varrho_\xi(1) - u] \geq e^{Du}\exp\{-D \varrho_\xi(1)\}\left(1 - \frac{1}{2}D^2[\varrho_\xi(1) - u]e^{D[\varrho_\xi(1) - u]}\right) \geq e^{Du}\exp\{-D \varrho_\xi(1)\}\left(1 - 2b^2[\varrho_\xi(1) - u]e^{D[\varrho_\xi(1) - u]}\right) \geq e^{Du}\exp\{-D \varrho_\xi(1)\}\left(1 - 2eb^2Me_x(2)\right) \geq e^{Du}\exp\{-D \varrho_\xi(1)\}(1 - 2e\kappa_0^2/n),$$ (6.7)

where the second inequality follows from $D \leq (1 - \kappa_0)^{-1}b \leq 2b$, the third inequality is a consequence of (6.2), and the last inequality follows from condition (5.7) which holds under (3.6). This together with (6.5) yields

$$1 - b \varrho_\xi(1) \geq (1 - bu)e^{Du}\exp\{-D \varrho_\xi(1)\}(1 - 1/n).$$ (6.8)
Combining (6.6) and (6.8) with (6.4) and \( n_0(x) \leq n \) we get 
\[
e^{-1}p_\zeta^*(x) \leq p_\zeta(x) \leq p_\zeta^*(x), \ \forall x \in \mathbb{R}^d,
\]
where 
\[
p_\zeta^*(x) := T(x)(1 - bu)^{n_0(x)}e^{D_{n_0}(x)} \prod_{m \in \mathcal{M}} e^{-D_{n_0}(x)c_m|\zeta_m|^{n_m(x)}}.
\]
Thus we showed that under conditions (3.6)(i) and (3.8) 
\[
\{ \zeta \in \mathcal{W}_\pi \} \subseteq \{ e^{-1}p_\zeta^*(x) \leq p_\zeta(x) \leq p_\zeta^*(x) \}, \ \forall x \in \mathbb{R}^d.
\]
(6.9)

Since \( \zeta_m, m \in \mathcal{M} \) are independent random variables we get 
\[
\mathbb{E}_{\pi} \left\{ p_\zeta^*(x) \right\} = T(x)(1 - bu)^{n_0(x)}e^{D_{n_0}(x)} \prod_{m \in \mathcal{M}} \mathbb{E}_{\pi} \left\{ e^{-D_{n_0}(x)c_m|\zeta_m|^{n_m(x)}} \right\},
\]
It remains to note that since \( \epsilon_{\mu}(1) = \epsilon_{\nu}(1) \), values of \( D \) and \( u \) do not depend on \( \pi \in \{\mu, \nu\} \); therefore 
\[
\Upsilon(x) := \frac{\mathbb{E}_{\mu} \left\{ p_\zeta^*(x) \right\}}{\mathbb{E}_{\nu} \left\{ p_\zeta^*(x) \right\}} = \prod_{m \in \mathcal{M}} \gamma_{m, \mu}(x) \gamma_{m, \nu}(x), \ \forall x \in \mathbb{R}^d,
\]
where \( \gamma_{m, \pi} \) is given in (3.15).

2\textsuperscript{nd}. Derivation of lower bound (3.16)

(a). According to (3.11), 
\[
\int_{\mathbb{R}^d} H(f_\zeta(x)) \, dx = S_0(A\sigma g_\zeta(1)) + \sigma \sum_{m \in \mathcal{M}} S(A\zeta_m).
\]

For \( \pi \in \{\mu, \nu\} \) define 
\[
\mathcal{W}_{\pi, S} := \left\{ w \in [0,1]^M : \left| \sum_{m \in \mathcal{M}} S(Aw_m) - ME_\pi(A) \right| \leq \sqrt{\nu} MV_\pi(A) \right\}.
\]
The events \( \{ \zeta \in \mathcal{W}_\pi \} \) [see (6.1)] and \( \{ \zeta \in \mathcal{W}_{\pi, S} \} \) control deviations of sums of independent random variables from their expectations, where the thresholds on the right hand side in the definitions of \( \mathcal{W}_\pi \) and \( \mathcal{W}_{\pi, S} \) are the upper bounds on the standard deviations of the sum inflated by a factor \( \sqrt{\nu} \). This fact allows to assert that by Chebyshev’s inequality 
\[
\mathbb{P}_{\pi} \{ \zeta \notin \mathcal{W}_\pi \} \leq 1/\nu, \ \mathbb{P}_{\pi} \{ \zeta \notin \mathcal{W}_{\pi, S} \} \leq 1/\nu.
\]

Assume that \( \zeta \in \mathcal{W}_\pi \cap \mathcal{W}_{\pi, S} \); then 
\[
\left| \int_{\mathbb{R}^d} H(f_\zeta(x)) \, dx - H_\pi^* \right| \\
\leq \left| S_0(A\sigma g_\zeta(1)) - S_0(A\sigma ME_\pi(1)) \right| + \sigma \sum_{m \in \mathcal{M}} S(A\zeta_m) - ME_\pi(A),
\]
where \( H_\pi^* \) is defined in (3.12). We have 
\[
A\sigma |g_\zeta(1) - ME_\pi(1)| \leq A\sigma \sqrt{\nu ME_\pi(2)} \text{ in view of } \zeta \in \mathcal{W}_\pi \text{ and (3.6); therefore}
\]
\[
\left| S_0(A\sigma g_\zeta(1)) - S_0(A\sigma ME_\pi(1)) \right| \leq \eta_{S_0}(A\sigma ME_\pi(1); A\sigma \sqrt{\nu ME_\pi(2)}).
\]
If $\zeta \in \mathcal{W}_\pi \cap \mathcal{W}_{\pi,S}$ is realized then
\[
\left| \int_{\mathbb{R}^d} H(f_\zeta(x))dx - H^*_\pi \right| \leq \eta_{S_\alpha}(A\sigma Me_\pi(1); A\sigma \sqrt{vM e_\pi(2)}) + \sigma \sqrt{vM} V_e(A) = \alpha_\pi,
\]
where $\alpha_\pi$ is defined in (3.13). Therefore we have shown that
\[
\{\zeta \in \mathcal{W}_\pi\} \cap \{\zeta \in \mathcal{W}_{\pi,S}\} \subseteq \{\Psi(f_\zeta) \in \mathcal{J}_\pi\},
\]
where $\mathcal{J}_\pi$ is defined in (3.14).

(b). For $\pi \in \{\mu, \nu\}$ define $C_\pi := \{\zeta \in \mathcal{W}_\pi\} \cap \{\zeta \in \mathcal{W}_{\pi,S}\} \cap \{f_\zeta \in F\}$. For the sake of brevity in the subsequent proof we write $\Delta := \Delta(\mu, \nu)$. For arbitrary estimator $\hat{\Psi}$ of $\Psi(f)$ we have
\[
2 \sup_{f \in \mathcal{F} \cap \mathcal{G}} \mathbb{P}_f\left\{ |\hat{\Psi} - \Psi(f)| \geq \frac{1}{2} \Delta \right\} \geq \mathbb{E}_\pi \left[ 1\{C_\pi\} \mathbb{P}_{f_\zeta}\left\{ |\hat{\Psi} - \Psi(f_\zeta)| \geq \frac{1}{2} \Delta \right\} \right] + \mathbb{E}_\nu \left[ 1\{C_\pi\} \mathbb{P}_{f_\zeta}\left\{ |\hat{\Psi} - \Psi(f_\zeta)| \geq \frac{1}{2} \Delta \right\} \right].
\]
Let $a_\pi := \inf_{\alpha_1, \alpha_2} G(\hat{H}_\pi + \alpha)$, $b_\pi := \sup_{\alpha_1, \alpha_2} G(\hat{H}_\pi + \alpha)$ so that $\mathcal{J}_\pi = [a_\pi, b_\pi]$. Therefore letting $\mathcal{I}_\pi(\Delta) := [a_\pi - \frac{1}{2} \Delta, b_\pi + \frac{1}{2} \Delta]$ we have from (6.10)
\[
C_\pi \cap \left\{ |\hat{\Psi} - \Psi(f_\zeta)| \geq \frac{1}{2} \Delta \right\} \supseteq C_\pi \cap \{\hat{\Psi} \notin \mathcal{I}_\pi(\Delta)\}.
\]
This implies
\[
J_\pi := \mathbb{E}_\pi \left[ 1\{C_\pi\} \mathbb{P}_{f_\zeta}\left\{ |\hat{\Psi} - \Psi(f_\zeta)| \geq \frac{1}{2} \Delta \right\} \right] \geq \mathbb{E}_\pi \left[ 1\{C_\pi\} \mathbb{P}_{f_\zeta}\left\{ \hat{\Psi} \notin \mathcal{I}_\pi(\Delta) \right\} \right]
\]
\[
= \mathbb{E}_\pi \left[ 1\{C_\pi\} \int_{\mathbb{R}^{dn}} 1\{\hat{\Psi}(x) \notin \mathcal{I}_\pi(\Delta)\} p_\pi(x)dx \right]
\]
\[
\geq e^{-1} \mathbb{E}_\pi \left[ 1\{C_\pi\} \int_{\mathbb{R}^{dn}} 1\{\hat{\Psi}(x) \notin \mathcal{I}_\pi(\Delta)\} \right] p_\pi(x)dx \right]
\]
\[
\geq e^{-1} \int_{\mathbb{R}^{dn}} 1\{\hat{\Psi}(x) \notin \mathcal{I}_\pi(\Delta)\} \mathbb{E}_\pi [p_\pi(x)]dx - e^{-1} \mathbb{E}_\pi \left[ 1\{\bar{C}_\pi\} \right] p_\pi(x),
\]
where $\bar{C}_\pi$ is the event complementary to $C_\pi$, and $p_\pi(x) := \int_{\mathbb{R}^{dn}} p_\pi(x)dx$. In the third line we have used that $p_\pi(x) \geq e^{-1} p_\pi^*(x)$ for all $x \in \mathbb{R}^{dn}$ on the event $\{\zeta \in \mathcal{W}_\pi\}$. Note that by Chebyshev’s inequality and in view of Assumption 1
\[
\mathbb{P}_\pi\{\bar{C}_\pi\} \leq \mathbb{P}_\pi\{\zeta \notin \mathcal{W}_\pi\} + \mathbb{P}_\pi\{\zeta \notin \mathcal{W}_{\pi,S}\} + \mathbb{P}_\pi\{\zeta \notin F\} \leq 2v^{-1} + \varepsilon.
\]
Then by the Cauchy–Schwarz inequality
\[
\mathbb{E}_\pi \left[ 1\{\bar{C}_\pi\} \right] p_\pi^2 \leq \sqrt{(2v^{-1} + \varepsilon)} \max_{\pi \in \{\mu, \nu\}} \left\{ \mathbb{E}_\pi [p_\pi(x)]^2 \right\}^{1/2} =: R
\]
which leads to
\[
J_\pi \geq e^{-1} \int_{\mathbb{R}^{dn}} 1\{\hat{\Psi}(x) \notin \mathcal{I}_\pi(\Delta)\} \mathbb{E}_\pi [p_\pi(x)]dx - e^{-1} R.
\]
Furthermore, we note that
\[
J_\mu + e^{-1} R \geq e^{-1} \int_{\mathbb{R}^{dn}} 1\{\hat{\Psi}(x) \notin \mathcal{I}_\mu(\Delta)\} \mathbb{E}_\nu [p_\pi(x)]dx
\]
\[
\geq (2e)^{-1} \int_{\mathbb{R}^{dn}} 1\{\hat{\Psi}(x) \notin \mathcal{I}_\mu(\Delta)\} \{T(x) \geq \frac{1}{2}\} \mathbb{E}_\nu [p_\pi(x)]dx,
\]
and that for all \( x \in \mathbb{R}^d \)
\[
1 \{ \tilde{\Psi}(x) \not\in I_\mu(\Delta) \} + 1 \{ \tilde{\Psi}(x) \not\in I_\nu(\Delta) \} \geq 1. \tag{6.15}
\]

The last inequality is an immediate consequence of definition of \( I_n(\Delta) \) and the fact that \( \Delta = \Delta(\mu, \nu) > 0 \). Therefore combining (6.15), (6.14), (6.13) and (6.11) we obtain
\[
\frac{1}{2}(J_\mu + J_\nu) \geq (4e)^{-1} \int_{\mathbb{R}^d} 1\{ T(x) \geq \frac{1}{2} \} \mathbb{E}_\nu[p_\zeta^*(x)] dx - e^{-1} R \geq (4e)^{-1} \int_{\mathbb{R}^d} 1\{ T(x) \geq \frac{1}{2} \} \mathbb{E}_\nu \left[ 1\{ \zeta \in W_\nu \} p_\zeta(x) \right] dx - e^{-1} R \geq (4e)^{-1} \mathbb{E}_\nu \left[ \mathbb{P}_\zeta \left\{ T(X(\zeta)) \geq \frac{1}{2} \right\} \right] - e^{-1} R,
\]
where in the second line we have used that \( p_\zeta^*(x) \geq p_\zeta(x) \) for all \( x \in \mathbb{R}^d \) on the event \( \{ \zeta \in W_\nu \} \), see (6.9). This together with (6.11) and Chebyshev’s inequality implies that
\[
[\Delta(\mu, \nu)]^{-2} \mathcal{R}_n^2(\mathcal{F}) \geq (36e)^{-1} \mathbb{E}_\nu \left[ \mathbb{P}_\zeta \left\{ T(X(\zeta)) \geq \frac{1}{2} \right\} \right] - (36ev)^{-1} - (9e)^{-1} R. \tag{6.16}
\]

3. Bounding the remainder in (6.16)

In order to complete the proof of the theorem, in view of (6.12), it remains to show that
\[
\mathbb{E}_x(p_\zeta^*)^2 \leq 2. \tag{6.17}
\]

Indeed, if (6.17) is established then the theorem statement follows from (6.16) and (6.12).

(a) First, we note that in view of (6.3) and by definition of \( n_0(x) \) and \( n_m(x) \)
\[
T(x)(1 - bu)^{n_0(x)} e^{Dn_0(x)} = \prod_{i=1}^n \left\{ e^{Dn(1 - bu) f_0(x_i)} \right\}^{n_0(x_i)} \prod_{m \in \mathcal{M}} \Lambda_m(x_i)^{n_m(x_i)},
\]
and
\[
\prod_{m \in \mathcal{M}} e^{-Dn_0(x) \zeta_m [\zeta_m]_{n_0(x)}} = \prod_{i=1}^d \left[ e^{-Dn(1)} \right]^{1 n_0(x_i)} \prod_{m \in \mathcal{M}} [\zeta_m]^{1 n_m(x_i)}.
\]

Therefore
\[
p_\zeta^*(x) = \prod_{i=1}^n \left\{ f_0(x_i)(1 - bu)e^{-D[n_0(1) - u]} \right\}^{1 n_0(x_i)} \prod_{m \in \mathcal{M}} \left( A \Lambda_m(x_i) \zeta_m \right)^{1 n_m(x_i)}
\]
\[
= \prod_{i=1}^n \left\{ [1 n_0(x_i) + f_0(x_i)(1 - bu)e^{-D[n_0(1) - u]}] \right\}^{1 n_0(x_i)} \prod_{m \in \mathcal{M}} A \Lambda_m(x_i)
\]
\[
= \prod_{i=1}^n \left[ f_0(x_i)(1 - bu)e^{-D[n_0(1) - u]} + \sum_{m \in \mathcal{M}} A \Lambda_m(x_i) \right],
\]
and taking into account that \( \int \Lambda_\alpha(x)dx = \sigma \) and \( b = \Lambda \sigma \) we obtain
\[
p^\alpha_\chi = \left[ (1 - bu) e^{-D|p_\alpha(1)|u} + b\rho_\chi(1) \right]^n = \left[ e^{-\frac{b\rho_\chi(1)}{1-bu}} (1 - bu) + b\rho_\chi(1) \right]^n.
\]
Denote \( \chi := \rho_\chi(1) - \mathbb{E_\pi} \{ \rho_\chi(1) \} \). Since \( \mathbb{E_\pi} \{ \rho_\chi(1) \} = u \) we have
\[
p^\alpha_\chi = \left[ e^{-\frac{b\chi}{1-bu}} (1 - bu) + bu + b\chi \right]^n. \tag{6.18}
\]
Note that \( e_\pi(1) \leq 1; \) hence \( 0 < bu \leq 1/2 \) in view of (3.8). Also since \( \rho_\chi(1) \) is a positive random variable
\[
\frac{b(\rho_\chi(1) - u)}{1-bu} \geq -\frac{bu}{1-bu} \geq -1.
\]
Since \( e^{-t} \leq 1 - t + \frac{1}{2} t^2 \) for all \( t \geq -1 \), we have \( e^{-\frac{b\chi}{1-bu}} (1 - bu) \leq 1 - bu - b\chi + 2b^2\chi^2 \) which together with (6.18) leads to
\[
p^\alpha_\chi \leq \left[ 1 + 2b^2\chi^2 \right]^n. \tag{6.19}
\]
Now we bound second moment of the random variable on the right hand side of (6.19).

(b). Since the random variable \( \chi \) takes values in \([0, M]\) we have
\[
\mathbb{E_\pi}(p^\alpha_\chi)^2 = \mathbb{E_\pi} \left\{ \left[ 1 + 2b^2\chi^2 \right]^{2n} \right\} = 1 + 8b^2n\int_0^M y(1 + 2b^2y^2)^{2n-1} \mathbb{P_\pi}(|\chi| \geq y) dy. \tag{6.20}
\]
Our current goal is to bound from above the integral on the right hand side of the previous formula under conditions (3.6)(i) and (3.6)(ii).

(i). First assume that (3.6)(i) holds. Note that \( \chi \) is a sum of i.i.d. centered random variables taking values in \([0, 1]\). Therefore applying Bernstein’s inequality we obtain that
\[
\int_0^M y(1 + 2b^2y^2)^{2n-1} \mathbb{P_\pi}(|\chi| \geq y) dy \leq 2 \int_0^M y \exp \left\{ 4nb^2y^2 - \frac{y^2}{2M + \frac{3}{4}y} \right\} dy \leq 2M \int_0^\infty y \exp \left\{ -y^2 \left( \frac{1}{4} - 4nb^2M \right) \right\} dy \leq 2M \int_0^\infty y \exp \left\{ -y^2 \left( \frac{1}{4} - 4\zeta_0^2 \right) \right\} dy,
\]
where in the last inequality we took into account that \( nb^2M \leq \zeta_0 \) by (3.6)(i). Then it follows from (6.20) and (3.6)(i) that
\[
\mathbb{E_\pi}(p^\alpha_\chi)^2 = \mathbb{E_\pi} \left\{ \left[ 1 + 2b^2\chi^2 \right]^{2n} \right\} \leq 1 + 16\zeta_0^2 \int_0^\infty y \exp \left\{ -y^2 \left( \frac{1}{4} - 4\zeta_0^2 \right) \right\} dy \leq 2,
\]
where the last inequality holds because \( \zeta_0 \) is sufficiently small. Thus (6.17) is proved under (3.6).

(ii). Now assume that (3.6)(ii) holds. Put for brevity \( v^2 = e_\pi(2) \) and remind that \( e_\pi(2) \leq e_\pi(1) \leq 1 \). We split the integral on the right hand of (6.20) as follows
\[
\mathbb{E_\pi}(p^\alpha_\chi)^2 = 1 + 8b^2n \left[ \int_0^{\vartheta^{1/2}M} y(1 + 2b^2y^2)^{2n-1} \mathbb{P_\pi}(|\chi| \geq y) dy + \int_{\vartheta^{1/2}M}^M y(1 + 2b^2y^2)^{2n-1} \mathbb{P_\pi}(|\chi| \geq y) dy \right] := 1 + 8b^2n [Q_1 + Q_2]. \tag{6.21}
\]
Our current goal is to bound integrals $Q_1$ and $Q_2$ on the right hand side of the above formula. We use the following convention: if the lower integration limit in the integral $Q_2$ is greater than the upper one, we put $Q_2 = 0$.

Let us begin with $Q_1$. Therefore applying Bernstein’s inequality we obtain

$$Q_1 \leq 2 \int_0^{9^{n/2}M} y(1 + 2b^2 y^2)^{2n-1} \exp \left\{ -\frac{y^2}{2Mv^2 + \frac{3}{2}y} \right\} dy$$

$$\leq 2 \int_0^{9^{n/2}M} y \exp \left\{ 4nb^2 y^2 - \frac{y^2}{8Mv^3/4} \right\} dy = 2Mv^{3/2} \int_0^{9^{n/2}M} z \exp \left\{ -z^2 \left( \frac{1}{8} - 4nb^2 Mv^{3/2} \right) \right\} dz$$

$$\leq 2Mv^{3/2} \int_0^{9^{n/2}M} z \exp \left\{ -z^2 \left( \frac{1}{8} - 2x_0 nb\sqrt{v} \right) \right\} dz \leq 2Mv^{3/2} \int_0^{\infty} z \exp \left\{ -z^2 \left( \frac{1}{8} - 2x_0^2 \right) \right\} dz,$$

where the penultimate inequality follows because $bMv = A\sigma M\sqrt{e_\pi(2)} \leq x_0$ [see (3.8)], and the last inequality is a consequence of $\sqrt{\pi} \ln(1 + 1/x) < 1$ for $0 \leq x \leq 1$. Then

$$8b^2 nQ_1 \leq 16b^2 nMv^{3/2} \int_0^{\infty} z \exp \left\{ -z^2 \left( \frac{1}{8} - 2x_0^2 \right) \right\} dz$$

$$\leq 16cx_0^2 \int_0^{\infty} z \exp \left\{ -z^2 \left( \frac{1}{8} - 2x_0^2 \right) \right\} dz \leq \frac{1}{2}, \quad (6.22)$$

where the last inequality holds because $x_0$ is sufficiently small.

Now we bound $Q_2$. For this purpose we use Bennett’s inequality:

$$Q_2 \leq 2 \int_0^{M} y(1 + 2b^2 y^2)^{2n-1} \exp \left\{ -Mv^2 \left( 1 + \frac{y}{Mv^2} \right) \ln \left( 1 + \frac{y}{Mv^2} \right) - \frac{y}{Mv^2} \right\} dy$$

$$\leq 2 \int_0^{M} y(1 + 2b^2 y^2)^{2n-1} \exp \left\{ -Mv^2 \left( 1 + \frac{y}{2Mv^2} \right) \ln \left( 1 + \frac{y}{2Mv^2} \right) \right\} dy$$

$$\leq 2(Mv^2)^2 e^{-Mv^2} \int_0^{1/v^2} z \left[ 1 + 2b^2 (Mv^2)^2 z^2 \right]^{2n-1} \exp \left\{ -\frac{1}{2} Mv^2 z \ln \left( 1 + 1/\sqrt{v} \right) \right\} dz$$

$$\leq 2(Mv^2)^2 e^{-Mv^2} \int_0^{1/v^2} z e^{\psi(z)} \exp \left\{ -\frac{1}{2} Mv^2 z \ln \left( 1 + 1/\sqrt{v} \right) \right\} dz,$$ \quad (6.23)

where the second line follows from the fact that $\ln(1 + (Mv^2)^{-1}y) \geq 2$ for $9Mv^{3/2} \leq y \leq M$, and in the last line we have denoted

$$\psi(z) := 2n \ln \left[ 1 + 2b^2 (Mv^2)^2 z^2 \right] - \frac{1}{2} Mv^2 z \ln \left( 1 + 1/\sqrt{v} \right).$$

Note that $\psi(0) = 0$. Our current goal is to demonstrate that $\psi$ is monotone decreasing on the positive real line. Indeed,

$$\psi'(z) = \frac{8nb^2 M^2 v^4 z}{1 + 2b^2 M^2 v^4 z^2} - \frac{1}{2} Mv^2 \ln(1 + 1/\sqrt{v}),$$

and inequality $\psi'(z) < 0$ for $z > 0$ is equivalent

$$\ln(1 + 1/\sqrt{v}) b^2 M^2 v^4 z^2 - 16ab^2 Mv^2 z + \frac{1}{2} \ln(1 + 1/\sqrt{v}) > 0, \quad \forall z > 0.$$
This inequality is fulfilled if \( nb \leq \frac{1}{\sqrt{25}} \ln (1 + 1/\sqrt{v}) \) which holds in view of (3.6)(ii). Hence \( \psi'(z) < 0 \) for all \( z > 0 \), so that \( \psi \) is negative for \( z > 0 \). Then it follows from (6.23) that

\[
Q_2 \leq 2(Mv^2) e^{-Mv^2} \int_{1/v}^{1/v^2} z \exp \left\{ - \frac{1}{2} Mv^2 z \ln (1 + 1/\sqrt{v}) \right\} \, dz
\]

\[
\leq 32 e^{-Mv^2} \left[ \ln (1 + 1/\sqrt{v}) \right]^{-2} \int_0^{\infty} te^{-t} \, dt = 32 e^{-Mv^2} \left[ \ln (1 + 1/\sqrt{v}) \right]^{-2},
\]

and therefore using (3.6)(ii) we obtain

\[
8b^2 n Q_2 \leq \frac{256 \kappa_0^2}{n} \leq \frac{1}{2} \text{ by smallness of } \kappa_0.
\]

Combining this bound with (6.22) and (6.21) we obtain that

\[
E \pi (\hat{p}^* \zeta)
\]

\[
\leq 2 \text{ under (3.6)(ii). The theorem is proved.}
\]

6.2. Proof of Corollary 1

The following well known inequality on the tail of binomial random variable, see Boucheron et al. (2013), will be exploited in the proof of the theorem.

**Lemma 3.** Let \( \xi \) be a binomial random variable with parameters \( n \) and \( p \). Then for \( pn \leq z \leq n \) one has

\[
P(\xi \geq z) \leq \left( \frac{pn}{z} \right)^z e^{z-pn}.
\]

Let us remark that (3.17) implies

\[
D_n = A \sigma n \leq \kappa_1 t M^{-1/t}.
\]

Without loss of generality we assume that \( e_\mu(t) \leq e_\nu(t) \), where we have put for brevity \( t = \lfloor t \rfloor \).

Define the random events

\[
\mathcal{D} := \bigcap_{m \in M} \{ \hat{n}_m \leq t - 1 \}, \quad \mathcal{E} := \bigcap_{k = 0}^{t-1} \{ \eta_k \leq M^{1-\gamma} \},
\]

where \( \hat{n}_m = n_m(X^{(n)}) \), and we have put \( \eta_k := \sum_{m \in M} 1 \{ \hat{n}_m = k \} \).

1. If \( \mathcal{D} \) is realized then for any \( m \in M \) we have

\[
\gamma_{m,\pi}(X^{(n)}) = \sum_{k = 0}^{t-1} 1 \{ \hat{n}_m = k \} \int_0^1 y^k e^{-\hat{n}_0 D_y} \pi(dy) = \prod_{k = 0}^{t-1} \left[ \Sigma_k(\pi) \right]^{1(\hat{n}_m = k)},
\]

where \( \gamma_{m,\pi}(\cdot) \) is defined in (3.15), and we have denoted

\[
\Sigma_k(\pi) := \int_0^1 y^k e^{-\hat{n}_0 D_y} \pi(dy).
\]

Hence, if event \( \mathcal{D} \) is realized then

\[
\Upsilon(X^{(n)}) = \prod_{m \in M} \gamma_{m,\mu}(X^{(n)}) = \prod_{k = 0}^{t-1} \left[ \frac{\Sigma_k(\mu)}{\Sigma_k(\nu)} \right]^{\eta_k}.
\]
Let \( K = (t - 1)1_{\mathbb{N}^*}(t) + 2[\tau|\mathbb{N}^*](t) \) and for \( k = 0, 1, \ldots, t - 1 \) define
\[
U_k(\pi) := \sum_{j=0}^{K-k} \frac{(-1)^j}{j!} (D\hat{\nu}_0)^j e_\pi(j + k), \quad U_k(\pi) := \sum_{j=K-k+1}^{\infty} \frac{(-1)^j}{j!} (D\hat{\nu}_0)^j e_\pi(j + k).
\]

By the Taylor expansion, \( \mathcal{U}_k(\pi) = U_k(\pi) + \hat{U}_k(\pi) \). Moreover, since \( \mu \sim \nu \),
\[
u_k(\mu) = \nu_k(\nu), \quad \forall k = 0, 1, \ldots, t - 1.
\]

Next, in view of (6.24), for any \( k = 0, 1, \ldots, t - 1 \) and \( j \geq K - k + 1 \)
\[
\frac{(D\hat{\nu}_0)^j}{j!} \leq \left[ \frac{\epsilon Dn}{K-K+1} \right]^j \leq \left[ \frac{\epsilon \mathcal{M}^{1/4}}{K-K+1} \right]^j \leq \begin{cases} 
[\epsilon \mathcal{M}^2], & t \in \mathbb{N}^*; \\
[\epsilon \mathcal{M}], & t \notin \mathbb{N}^*. 
\end{cases}
\]

Recall that if \( t \in \mathbb{N}^* \) then \( t \) is a fixed number independent of \( n \). Hence, choosing \( \mathcal{M} \leq (2e)^{-1} \) if \( t \in \mathbb{N}^* \) and \( \mathcal{M} \leq (2e)^{-1} \) if \( t \notin \mathbb{N}^* \) we assert that
\[
|U_k(\pi)| \leq \left[ \frac{\epsilon Dn}{K-K+1} \right]^{K-k+1} \sum_{l=0}^{\infty} 2^{-l} e_\pi(l + K + 1), \quad k = 0, 1, \ldots, t - 1.
\]

Also for any \( k = 0, 1, \ldots, t - 1 \) one has
\[
\mathcal{U}_k(\pi) \geq e^{-Dn} e_\pi(k) \geq e^{-Dn} e_\pi(t).
\]

It follows from (6.26), (6.27) and (6.28) that on the event \( \mathcal{D} \) one has
\[
\frac{\mathcal{U}_k(\mu)}{\mathcal{U}_k(\nu)} \geq 1 - 4e^{Dn} \left[ \frac{\epsilon Dn}{K-K+1} \right]^{K-k+1}.
\]

Consider separately two cases. First assume that \( t \in \mathbb{N}^* \). Choosing \( \mathcal{M} \leq (e t)^{-1} \) and using (6.24) we have
\[
\frac{\mathcal{U}_k(\mu)}{\mathcal{U}_k(\nu)} \geq 1 - 4e^{\epsilon t \mathcal{M}^{1/4}} \left[ \frac{\epsilon \mathcal{M}^{1/4}}{t+2} \right]^{t-k} \geq 1 - 4e^{\epsilon t \mathcal{M}^{1/4}} \mathcal{M}^{k+1}. \quad (6.30)
\]

Assuming additionally that \( \mathcal{E} \) is realized we get from (6.25) and (6.30)
\[
\mathcal{T}(X^{(n)}) \geq \prod_{k=0}^{t-1} \left[ 1 - 4e^{\epsilon t \mathcal{M}^{1/4}} \mathcal{M}^{k} \right]^{M^{(t-k)/t}} \geq \left[ \frac{\inf_{z \geq 1} \left( 1 - 4e^{\epsilon t \mathcal{M}^{1/4}} \mathcal{M}^{k+1} \right)}{z} \right]^{1/t} > \frac{1}{2},
\]
provided that \( \mathcal{M} \) is sufficiently small, and \( t \) is independent of \( n \). Thus, we have proved that if \( t \in \mathbb{N}^* \)
\[
\mathcal{D} \cap \mathcal{E} \subseteq \{ \mathcal{T}(X^{(n)}) \geq \frac{1}{2} \}. \quad (6.31)
\]

Now assume that \( t \notin \mathbb{N}^* \), and let \( \mathcal{M} \leq 1/e \). In this case (6.29) becomes for any \( k = 0, 1, \ldots, t - 1 \)
\[
\frac{\mathcal{U}_k(\mu)}{\mathcal{U}_k(\nu)} \geq 1 - 4e^{Dn} \left[ \frac{\epsilon Dn}{2t+1} \right]^{2t-k+1} \geq 1 - 4e^{\epsilon t \mathcal{M}^{1/4}} \mathcal{M}^{2t-k+1} \geq 1 - 4e^{\epsilon t \mathcal{M}^{1/4}} \mathcal{M}^{-1}. \]
To get the last inequality we have used that $t = |t|$ and therefore $t + 2 \geq t$. It remains to note that $\eta_k \leq M$ by its definition and we get from (6.25)

$$
\mathcal{Y}(X^{(n)}) \geq \left(1 - 4e^{[\xi_1 + \ln(\xi_1)]t} M^{-1}\right)^{M_t} \geq \left(1 - 4e^{[\xi_1 + \ln(\xi_1)]t (M^{-1})^{M_t}}\right)^{M_t}.
$$

Choosing $\xi_1 > 0$ satisfying $a(\xi_1) := -\xi_1 - \ln(\xi_1) > 0$ and noting that $\sup_{y \geq 0} ye^{-ay} \leq 1/(ea)$ for any $a > 0$, we finally get

$$
\mathcal{Y}(X^{(n)}) \geq \inf_{z \geq 1} \left(1 - 2[a(\xi_1)z]^{-1}\right)^{z} > \frac{1}{2},
$$

provided that $\xi_1$ is sufficiently small. Thus, we have proved that if $t \notin N^*$

$$
\mathcal{D} \subseteq \{ \mathcal{Y}(X^{(n)}) \geq \frac{1}{2} \}. \quad (6.32)
$$

In view of (6.31) and (6.32) in order to complete the proof it suffices to show that

$$
\mathbb{E}_n[\mathbb{P}_{f_c} \{\mathcal{D} \cap \mathcal{E}\}] \geq 1/3, \quad t \in N^*; \quad \mathbb{E}_n[\mathbb{P}_{f_c} \{\mathcal{D}\}] \geq 1/3, \quad t \notin N^*. \quad (6.33) \quad (6.34)
$$

2°. Let $\mathcal{E}$ and $\mathcal{B}$ be the events complimentary to $\mathcal{E}$ and $\mathcal{B}$ respectively. (a). Let us prove that for any $t \in \mathbb{R}$, $t \geq 1$, and $\xi_1 \leq 1/(2e)$

$$
\mathbb{E}_n[\mathbb{P}_{f_c} \{\mathcal{D}\}] \leq 2e\xi_1. \quad (6.35)
$$

Indeed,

$$
\mathbb{P}_{f_c} \{\mathcal{D}\} \leq \sum_{m \in \mathcal{M}} \mathbb{P}_{f_c} \{\mathcal{F}_{m} \geq t\} = \sum_{m \in \mathcal{M}} \mathbb{P}_{f_c} \{\sum_{i=1}^{n} \Pi_m(X_i) \geq t\}.
$$

Note that under $f_c$ random variable $\mathcal{F}_{m}$ is binomial with parameters $n$ and

$$
p_n := \mathbb{P}_{f_c} (X_i \in \Pi_m) = \int_{\Pi_m} f_c(y)dy = A\sigma \zeta_m \leq A\sigma,
$$

$$
np_n \leq \xi_1 tM^{-1/t} \leq \xi_1 t \leq t \quad \text{in view of the condition (3.17). In the last inequality we took into account that } \zeta_m \leq 1 \text{ for all } m \in \mathcal{M}. \quad \text{Hence, applying Lemma 3 we obtain for any } m \in \mathcal{M}
$$

$$
\mathbb{P}_{f_c} (\mathcal{F}_{m} \geq t) \leq (enA\sigma t^{-1})^{t} \leq (2e\xi_1)^{t}M^{-1} \leq (2e\xi_1)^{t}M^{-1}.
$$

Thus, (6.35) is proved. It implies in particular (6.34) if one chooses $\xi_1 \leq 1/(3e)$ and the corollary statement in the case $t \notin N^*$ follows from from (6.34) and Theorem 4.

(b). Now we show that if $t \in N^*$ (recall that in this case $t = t$)

$$
\mathbb{E}_n[\mathbb{P}_{f_c} \{\mathcal{E}\}] \leq \xi_1 e^t. \quad (6.36)
$$

By Markov’s inequality

$$
\mathbb{P}_{f_c} \{\mathcal{E}\} \leq \sum_{k=1}^{t-1} \mathbb{P}_{f_c} \{\eta_k > M^{1-t} \} \leq \sum_{k=1}^{t-1} M^{t-1} \mathbb{E}_{f_c}(\eta_k),
$$
where in the first inequality we have used that \( \eta_0 \leq M \) by definition. Noting that

\[
\mathbb{E}_{f_{\xi}}(\eta_k) = \sum_{m \in \mathcal{M}} P_{f_{\xi}}\{\hat{n}_m = k\} = \sum_{m \in \mathcal{M}} P_{f_{\xi}}\{\sum_{i=1}^{n} 1_{\Pi_m}(X_i) = k\},
\]

we obtain

\[
\mathbb{E}_{f_{\xi}}(\eta_k) = \sum_{m \in \mathcal{M}} \binom{n}{k} \left[ \int_{\Pi_m} f_{\xi}(x) dx \right]^k \left[ 1 - \int_{\Pi_m} f_{\xi}(x) dx \right]^{n-k} \leq \frac{M(nA\sigma)^k}{k!}.
\]

Using the condition (3.17) we obtain \( \mathbb{E}_{f_{\xi}}(\eta_k) \leq \kappa_1 t^k M(t-k)^{1/2}(k!)^{-1} \), if \( \kappa \leq 1 \), and (6.36) follows.

This implies (6.36), and it follows from (6.36) and (6.35) that

\[
\mathbb{E}_{\nu}\left[ P_{f_{\xi}}(D \cap \mathcal{E}) \right] \geq 1 - \kappa_1 \left[ 2e + e^t \right] > 1/3,
\]

provided that \( \kappa_1 \) sufficiently small and \( t \) is independent of \( n \). The corollary statement in the case \( t \in \mathbb{N}^* \) follows now from (6.33) and Theorem 4.

Acknowledgments The authors are grateful to two anonymous referees and the Associate Editor for careful reading and useful remarks that led to significant improvements in the presentation.

Supplementary material The supplementary paper Goldenshuger and Lepski (2020b) contains the proofs of Propositions 1–4.

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.

Boucheron, S., Lugosi, G. and Massart, P. (2013). Concentration Inequalities: A Nonasymptotic Theory of Independence. Oxford University Press.

Cai, T. T. and Low, M. G. (2004) Minimax estimation of linear functionals over nonconvex parameter spaces. Ann. Statist. 32, 552–576.

Cai, T.T. and Low, M.G. (2005). Nonquadratic estimators of a quadratic functional. Ann. Statist. 33, 2930–2956.

Cai, T. T. and Low, M. G. (2011). Testing composite hypotheses, Hermite polynomials and optimal estimation of a nonsmooth functional. Ann. Statist. 39, 1012–1041.

Donoho, D. L. and Liu, R. C. (1991). Geometrizing rates of convergence. II, III. Ann. Statist. 19, 633–667, 668–701.

Donoho, D. and Nussbaum, M. (1990). Minimax quadratic estimation of a quadratic functional. J. Complexity 6, 290–323.

Giné, E. and Nickl, R. A simple adaptive estimator of the integrated square of a density. Bernoulli 14, 47–61.

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

Goldenshuger, A. and Lepski, O. (2020a). Minimax estimation of norms of a probability density: II. Rate–optimal estimation procedures. Manuscript.
Goldenshluger, A. and Lepski, O. (2020b). Supplementary material to “Minimax estimation of norms of a probability density: I. Lower bounds.” Manuscript.

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.

Han, Y., Jiao, J., Weissman, T. and Wu, Y. (2020). Optimal rates of entropy estimation over Lipschitz balls. *Ann. Statist.* **48**, 3228–3250.

Ibragimov, I. A. and Khasminskii, R. Z. (1986). An estimate for the value of a linear functional of the density of a distribution. (Russian) *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **153** (1986), 45-59; translation in *J. Soviet Math.* **44** (1989), no. 4, 454–465.

Ibragimov, I. A., Nemirovskii, A. S. and Khasminskii, R. Z. (1986). Some problems of nonparametric estimation in Gaussian white noise. (Russian) *Teor. Veroyatnost. i Primenen.* **31**, 451–466.

Juditsky, A. and Nemirovski, A. (2020) *Statistical Inference via Convex Optimization*. Princeton Series in Applied Mathematics, Princeton University Press.

Kerkyacharian, G. and Picard, D. (1996). Estimating nonquadratic functionals of a density using Haar wavelets. *Ann. Statist.* **24**, 485–507.

Kozachenko, L. F. and Leonenko, N. N. (1987). Sample estimate of the entropy of a random vector. *Probl. Inform. Transm.*, **23**, 95–101.

Laurent, B. (1996). Efficient estimation of integral functionals of a density. *Ann. Statist.* **24**, 659–681.

Laurent, B. (1997). Estimation of integral functionals of a density and its derivatives. *Bernoulli* **3**, 181–211.

Leonenko, N., Pronzato, L. and Savani, V. (2008). A class of Rényi information estimators for multidimensional densities. *Ann. Statist.* **36**, 2153–2182.

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

Natanson, I. P. (1964). *Constructive function theory. Vol. I. Uniform approximation*. Frederick Ungar Publishing Co., New York. The Russian original appeared in 1949 in one volume *Constructive theory of functions* (Russian), GITTL, Moscow, 1949.

Natanson, I. P. (1965). *Constructive function theory. Vol. II. Approximation in mean*. Frederick Ungar Publishing Co., New York. The Russian original appeared in 1949 in one volume *Constructive theory of functions* (Russian), GITTL, Moscow, 1949.

Nemirovskii, A. S. (1990). Necessary conditions for efficient estimation of functionals of a nonparametric signal observed in white noise. (Russian) *Teor. Veroyatnost. i Primenen.* **35**, 83–91; translation in *Theory Probab. Appl.* **35** (1990), 94–103 (1991).

Nemirovski, A. (2000). *Topics in non-parametric statistics*. Lectures on probability theory and statistics (Saint-Flour, 1998), 85–277, Lecture Notes in Math. **1738**, Springer, Berlin.

Nikol’skii, S. M. (1977). *Priblizhenie Funktsii Mnogikh Peremennykh i Teoremy Vlozheniya* 2nd ed. Nauka, Moscow.

Rényi, A. (1961). On measures of entropy and information. *Proc. 4th Berkeley Symp. Math. Statist. Probab. 1*, 547–561. Univ. California Press, Berkeley.

Suetin, P.K. (1979). *Classical Orthogonal Polynomials*. 2nd rev. ed., Nauka, Moscow.

Timan, A. F. (1963). *Theory of approximation of functions of a real variable*. Pergamon Press.

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.

Tsallis, C (1988). Possible generalization of Boltzmann-Gibbs statistics. *J. Statist. Phys.* **52**, 479–487.

Tsybakov, A. (2009). *Introduction to Nonparametric Estimation*. Springer Series in Statistics.
Springer, New York.

Wu, Y. and Yang, P. (2016). Minimax rates of entropy estimation on large alphabets via best polynomial approximation. *IEEE Trans. Information Theory* 62, 3702–3720.