ON ESTIMATION OF NON-SMOOTH FUNCTIONALS

LEPSKI, O., NEMIROVSKI, A., AND SPOKOINY, V.

Humboldt University, SFB 373, Spandauer Str. 1, 10178 Berlin

Technion–Israel Institute of Technology, Haifa

Weierstrass Institute for Applied Analysis and Stochastics,
Mohrenstr. 39, 10117 Berlin

December 1996

1991 *Mathematics Subject Classification.* 62G07; Secondary 62G20.

Key words and phrases. non-smooth functionals, integral norm, rate of estimation.

This research was partially supported by INTAS grant N93/1585.

ABSTRACT. Let a function f be observed with noise. In the present paper we concern the problem of nonparametric estimation of some non-smooth functionals of f, more precisely, L_r-norm $\|f\|_r$ of f. Existing in the literature results on estimation of functionals deal mostly with two extreme cases: estimation of a smooth (differentiable in L_2) functional or estimation of a singular functional like the value of f at a certain point or the maximum of f. In the first case, the rate of estimation is typically $n^{-1/2}$, n being the number of observations. In the second case, the rate of functional estimation coincides with the nonparametric rate of estimation of the whole function f in the corresponding norm.

We show that the case of estimation of $\|f\|_r$ is in some sense intermediate between the above extreme two. The optimal rate of estimation is worse than $n^{-1/2}$ but better than the usual nonparametric rate. The results depend on the value of r. For r even integer, the rate occurs to be $n^{-\beta/(2\beta+1-1/r)}$ where β is the degree of smoothness. If r is not even integer, then the nonparametric rate $n^{-\beta/(2\beta+1)}$ can be improved only by some logarithmic factor.

1. Introduction

The problem of estimation of a functional is one of the basic problems in statistical inference. Below we consider this problem in the nonparametric set-up. Let a function f be observed with noise. The goal is to estimate by observed data some real functional $F(f)$. Clearly the quality of estimation depends heavily on smoothness properties of the functional F. The theory for estimating linear functionals is in some sense the most developed. Hardest one dimensional subfamily arguments yield both linear estimators with smallest maximum risk among linear estimators and also show that this maximum risk is only a small multiple of the minimax risk, see Levit (1974, 1975), Koshevnik and Levit (1976), Ibragimov and Khasminski (1981, 1987) and Donoho and Liu (1991).

Another well studied situation concerns the case of "smooth" functionals. This is typically understood in the sense that F is differentiable in L_2. It was shown in Levit (1978), Hasminski and Ibragimov (1979), Ibragimov, Nemirovski and Khasminski (1986) that if F is smooth and the underlying function f is also smooth enough then $F(f)$ can be estimated with the parametric rate $n^{-1/2}$, see also Ibragimov and Khasminski (1991), Birgé and Massart (1995). The problem of estimation of quadratic functionals is studied in details in Hall and Marron (1987), Bickel and Ritov (1988), Donoho and Nussbaum (1990), Fan (1991), Efroimovich and Low (1996), Laurent (1996) among others. Estimation of functionals of the type $\int f^3$ is discussed in Kerkyacharian and Picard (1996).

The problem of estimation of non-smooth functionals is not well developed so far and there are very few results of this sort in the literature. Ibragimov and Khasminski (1980) established the rate of estimation of the maximum of f, Korostelev (1990) studied the problem of estimating L_1-norm of f. Korostelev and Tsybakov (1994) considered some functional estimation problems relying on the image model like estimation of the length of the image boundary or estimation of the size of image.

In the present paper we are focusing on the problem of estimating L_r-norm $\|f\|_r$ with some $r \geq 1$. It is worth to mention that at least three cases with $r = 1, 2$

and ∞ have very natural interpretation. The case with $r=\infty$ corresponds to estimation of the maximum of f. Ibragimov and Khasminski (1980) shown that the rate of estimation of $F(f)=\|f\|_\infty$ coincides with the rate of global estimation of the function f and one may therefore use plug-in estimator $\hat{F}=\|\hat{f}\|_\infty$ where $\hat{f}$ is a rate-optimal estimator of f.

Korostelev (1990) claimed the similar qualitative result for estimating of L_1-norm $\|f\|_1=\int|f(t)|$: plug-in estimator $\int|\hat{f}(t)|$ provides with the optimal rate $n^{-\beta/(2\beta+1)}$. However, the inspection of the proof shown a gap in establishing the lower bound. By more detailed analysis it was found out that the result itself is not correct: some improvement of the nonparametric rate is possible. Note meanwhile that this improvement is only by some log-factor and it is therefore unessential for practical application.

Another interesting phenomenon is met in estimating L_r-norm for $r>1$. It turns out that both the results and the methods differ essentially between the cases with r even integer and the remaining cases. In the first case the rate of estimation can be substantially improved compared with the nonparametric one, it is about $n^{-\beta/(2\beta+1-1/r)}$, for the remaining situations the result is not better than for L_1-norm.

The final remark concerns the question of correspondence between the problems of estimating L_r-norm and the nonparametric hypothesis testing problem when the distance between the null hypothesis and the alternative set is measured in L_r-norm, see Ingster (1982, 1993), Lepski and Spokoiny (1995) or Spokoiny (1996). The origin of this question is very natural. When considering the testing problem, one may first estimate the corresponding L_r-norm and then use the estimate as test statistic. Particularly, this recipe is correct for $r=2$. However, by comparison the above mentioned results one can see that the case with $r=2$ is the only one when this recipe "works". For other cases, the rates in testing and estimation problems are different.

The paper is organized as follows. In Section 2 we state the results separately for r even integer and for the remaining cases. The estimation procedures for $r=1$ and for even integer r are presented in Section 3. The proofs are deferred to Section 4.

2. Problem and main results

We begin by formulating the problem. Throughout the paper we consider the idealized "signal + white noise" model. Suppose we are given data $X(t)$, $t\in[0,1]$ obeying the stochastic differential equation

$$dX(t)=f(t)dt+n^{-1/2}dW(t) \tag{2.1}$$

where f is the unknown function, $W=(W(t),t\in[0,1])$ is the standard Wiener process, and the parameter n is taken by analogy with more realistic statistical models like regression or distribution density models where n is the number of observations. We consider further the asymptotic set-up when the parameter n tends to infinity. The function f is assumed to possess some smoothness properties. Namely, we suppose that f belongs to the Hölder class $\Sigma(\beta,L)$ with known parameters β, L. Define m as the maximal integer number less than β. Then $\Sigma(\beta, L)$ is the set of functions f such that

$$|f^{(m)}(t) - f^{(m)}(s)| \le L|t-s|^{\beta-m}, \qquad t,s \in R^1.$$

Here $f^{(m)}$ means the m-th derivative of f. For technical reason, we assume also that our function f is uniformly bounded by some constant $\varrho < 1$,

$$f \in \Sigma_\varrho(\beta, L) = \{f \in \Sigma(\beta, L) : \|f\|_\infty \le \varrho\}.$$

Given $r \ge 1$, we are interested to estimate L_r-norm of f,

$$\|f\|_r = \left[\int_0^1 |f(t)|^r dt\right]^{1/r}.$$

For an estimate $\hat{f}_n$ of $\|f\|_r$, let

$$\mathcal{R}(\hat{f}_n) = \sup_{f \in \Sigma_\varrho(\beta,L)} E\, l\left(\hat{f}_n - \|f\|_r\right)$$

where $l(\cdot)$ is a loss function. For our results, it is enough to require that l is a homogeneous function satisfying the standard conditions, see e.g. Ibragimov and Khasminski (1981, Section 2.3). However, to simplify our exposition, we prefer to be more definitive and suppose that $l(z) = |z|$. Therefore,

$$\mathcal{R}(\hat{f}_n) = \sup_{f \in \Sigma_\varrho(\beta,L)} E\left|\hat{f}_n - \|f\|_r\right|.$$

Set also

$$\mathcal{R}^*(n) = \inf_{\hat{f}_n} \sup_{f \in \Sigma_\varrho(\beta,L)} E\left|\hat{f}_n - \|f\|_r\right|$$

where inf is taken over the class of all measurable functions of the observation X.

We are about to state our results starting from the case with $r = 1$.

Theorem 2.1. *Let $r = 1$. There exist estimators $\hat{f}_n$ and a positive constant $C > 0$ which depend on β only such that for all large enough values of n, one has*

$$\mathcal{R}(\hat{f}_n) \le C L^{1/(2\beta+1)} (n \log n)^{-\beta/(2\beta+1)}. \tag{2.2}$$

This result shows that L_1-norm can be estimated with a better rate than the nonparametric rate $n^{-\beta/(2\beta+1)}$ but the improvement is only by some log-factor. The next result claims that more substantial improvement is impossible. This lower bound is valid for an arbitrary norm L_r when r is not even integer.

Theorem 2.2. *Let $r \ne 2k$, $k = 1, 2, \ldots$. Then for n large enough*

$$L^{-1/(2\beta+1)} (n \log n)^{\beta/(2\beta+1)} \mathcal{R}_n^* \ge c/\log n$$

with some positive $c > 0$ depending only on β.

Finally we present the result concerning the estimation of the norm L_r when r an is even integer.

Theorem 2.3. *Let $r = 2k$, $k = 1, 2, \ldots$. Then there are positive constants c, C depending possibly on β and such that for n large enough,*

$$c \le L^{-(1-1/r)/(2\beta+1-1/r)} n^{\beta/(2\beta+1-1/r)} \mathcal{R}^*(n) \le C.$$

3. Estimation procedures

In this section we present two estimation procedures: one for estimation of L_1 and another one for estimation of L_r-norm with r even integer.

We begin with the case of $r = 1$. First we explain the idea behind the construction. The function $|t|$ is not smooth because of the irregularity at the point $t = 0$. However, this function can be approximated by Fourier series $\sum_{k=1}^{N} c_k \cos(\pi k t)$ with the accuracy about N^{-1}. Therefore, our functional $\int |f(t)| dt$ can be approximated by the sum

$$\sum_{k=1}^{N} c_k \int_0^1 \cos(\pi k f(t)) dt$$

and each term in this sum is already a smooth functional estimated with the rate $n^{-1/2}$. To do this, one may use the method proposed in Ibragimov, Nemirovski and Khasminski (1986). Let $\tilde{f}(t)$ be a proper nonparametric estimator of $f(t)$, e.g. a kernel estimator, with the variance λ. Then the estimator $\hat{F}_k$ of $\int_0^1 \cos(\pi k f(t)) dt$ can be taken in the form

$$\hat{F}_k = E_\xi \int_0^1 \cos(\pi k (\tilde{f}(t) + i\lambda\xi)) dt = \int_0^1 \cos(\pi k \tilde{f}(t)) \exp\{\pi^2 k^2 \lambda^2 / 2\} dt.$$

Here ξ means a standard normal random variable independent of our observation X and E_ξ is the expectation w.r.t. ξ. It remains to select the number N in the Fourier expansion in an optimal way to balance the error of approximation and the stochastic error.

Our estimation procedure just follows this program. Let $m = \lfloor \beta \rfloor$ and let K be a compactly supported kernel of order m i.e. K is a continuous function satisfying the conditions

$(K.1)$ $\quad K(t) = 0$ for $|t| > 1$;
$(K.2)$ $\quad \int K(t) dt = 1$;
$(K.3)$ $\quad \int t^i K(t) = 0$ for $i = 1, \ldots, m$.

By $\|K\|$ we denote L_2-norm of K,

$$\|K\|^2 = \int K^2(t) dt. \tag{3.1}$$

Let also h be a bandwidth, $h \in (0,1)$. We make more precise the choice of h a bit later. Define a standard kernel estimation $\tilde{f}_f$ of f by

$$\tilde{f}_h(t) = \frac{1}{h} \int_0^1 K\left(\frac{t-u}{h}\right) dX(u).$$

As usual in kernel estimation, the kernel K is to be corrected near edge-points $0, 1$. With the aim to make our exposition more readable, we use the same notation for the original kernel K and for the boundary corrected one. The necessary changes in the exposition are obvious and we omit them everywhere.

Due to (2.1), the estimate $\tilde{f}_h(t)$ admits the standard decomposition into deterministic and stochastic components,

$$\tilde{f}_h(t) = f_h(t) + \lambda_h \xi_h(t), \tag{3.2}$$

where

$$f_h(t) = \frac{1}{h}\int_0^1 K\left(\frac{t-u}{h}\right) f(u)du,$$
$$\lambda_h = \frac{\|K\|}{\sqrt{nh}},$$
$$\xi_h(t) = \frac{1}{\|K\|\sqrt{h}}\int_0^1 K\left(\frac{t-u}{h}\right) dW(u).$$

Obviously $\xi_h(t)$ is standard normal and hence

$$E\,\tilde{f}_h(t) = f_h(t),$$
$$\mathrm{Var}\,\tilde{f}_h(t) = E\left(\tilde{f}_h(t) - f_h(t)\right)^2 = \lambda_h^2.$$

Let now

$$h = \left(L^2 n \log n\right)^{-1/(2\beta+1)}, \tag{3.3}$$
$$N = \theta L^{-1/(2\beta+1)}(n\log n)^{\beta/(2\beta+1)} \tag{3.4}$$

where

$$\theta = \frac{1}{\pi\|K\|\sqrt{2\beta+1}}.$$

Without loss of generality we will suppose that N is an integer number.

For all $k = 1, 2, \ldots, N$ and $\lambda > 0$, define functions $\nu_{k,\lambda}(\cdot)$ by

$$\nu_{k,\lambda}(t) = \cos(\pi k t)\exp\{\pi^2 k^2 \lambda^2/2\}. \tag{3.5}$$

Set now

$$Q_{N,\lambda}(t) = c_0 + \sum_{k=1}^{N} c_k \nu_{k,\lambda}(t) \tag{3.6}$$

where c_k are the Fourier coefficients of the function $\mu(t) = |t|$,

$$c_k = 2\int_0^1 t\cos(\pi k t)dt = \begin{cases} 1 & k = 0, \\ 0 & k = 2, 4, 6, \ldots, \\ 4(\pi k)^{-2} & k = 1, 3, 5, \ldots. \end{cases} \tag{3.7}$$

Finally we define the estimator $\hat{F}$ of $\|f\|_1$ as follows.

$$\hat{F}_n = \int_0^1 Q_{N,\lambda_h}(\tilde{f}_h(t))dt = c_0 + \int_0^1 \sum_{k=1}^{N} c_k \nu_{k,\lambda_h}(\tilde{f}_h(t))dt.$$

3.1. Estimation of $\|f\|_r$ for an even integer r

The difference between this case and the above considered is based on the trivial observation that the function $|t|^r$ is analytical only for even integer r. This fact will be essentially used in the construction.

Let us consider first the functional $\Phi_r(f) = F_r^r(f)$:

$$\Phi_r(f) = \|f\|_r^r = \int_0^1 f^r(t)dt.$$

This functional is smooth and it can be estimated (under some mild conditions) by observations X with the rate $n^{-1/2}$.

Let $\tilde{f}_h(t)$ be the kernel estimator of f from the above. Applying the method from Ibragimov, Nemirovski and Khasminski (1986), we arrive at the following estimator $\hat{\Phi}_n$ of $\Phi_r(f)$:

$$\hat{\Phi}_n = E_\xi \int_0^1 \left(\tilde{f}_h(t) + i\lambda_h \xi\right)^r dt = \int_0^1 \sum_{j=0}^{r/2} b_{2j} \lambda_h^{2j} |\tilde{f}_h(t)|^{r-2j} dt. \tag{3.8}$$

Here $i = \sqrt{-1}$, ξ means a standard Gaussian random variable independent of observations X, and E_ξ is the expectation w.r.t. ξ, so that

$$b_{2j} = (-1)^j E_\xi \xi^{2j}. \tag{3.9}$$

We specify

$$h = (L^2 n)^{-\frac{1}{2\beta+1-1/r}} \tag{3.10}$$

and define the estimator $\hat{F}_n$ of $\|f\|_r$ by

$$\hat{F}_n = (\max\{0, \hat{\Phi}_n\})^{1/r}.$$

4. Proofs

Below we present detailed proofs of Theorem 2.1 through 2.3. Everywhere $\varkappa$ with indices and not denote appropriate positive quantities depending on r only.

4.1. Proof of the upper bound in Theorem 2.1

We begin with some technical lemmas. Let the functions $\nu_{k,\lambda}$ be defined by (3.5), $\nu_{k,\lambda}(t) = \cos(\pi k t) \exp\{\pi^2 k^2 \lambda^2/2\}$, $k \ge 1$.

Lemma 4.1. *Let $z \in [-1, 1]$, $\lambda > 0$ and let ξ be a standard Gaussian random variable. Then for all $k \ge 1$,*

$$E\, \nu_{k,\lambda}(z + \lambda\xi) = \cos(\pi k z). \tag{4.1}$$

If $\sigma_{k,\lambda}(t)$ is defined by

$$\sigma_{k,\lambda}^2(t) \equiv \operatorname{Var} \nu_{k,\lambda} = E \left|\nu_{k,\lambda}(z + \lambda\xi) - \cos(\pi k z)\right|^2,$$

then

$$\sigma_{k,\lambda}(t) \le \pi k \lambda \exp\{\pi^2 k^2 \lambda^2/2\}.$$

Proof. Let $\varphi(x)=(2\pi)^{-1/2}\exp\{-x^2/2\}$ be the standard normal density. Then

$$\begin{aligned} E\,\nu_{k,\lambda}(z+\lambda\xi) &= \int \nu_{k,\lambda}(z+\lambda x)\varphi(x)dx \\ &= \exp\{\pi^2k^2\lambda^2/2\}\int \cos(\pi k(z+\lambda x))\varphi(x)dx \\ &= (2\pi)^{-1/2}\mathrm{Re}\left(\int \exp\{\pi^2k^2\lambda^2/2+i\pi k(z+\lambda x)-x^2/2\}dx\right) \\ &= \mathrm{Re}\left(\exp\{i\pi kz\}\,(2\pi)^{-1/2}\int \exp\{-(x-i\pi k\lambda)^2/2\}dx\right) \\ &= \cos(\pi kz) \end{aligned}$$

and (4.1) follows.

Next, proceeding as above we obtain

$$\begin{aligned} \sigma^2_{k,\lambda}(t) &\equiv \int (\nu_k(z+\lambda x)-\cos(\pi kz))^2\varphi(x)dx \\ &= \int \nu_k^2(z+\lambda x)\varphi(x)dx-\cos^2(\pi kz) \\ &= \exp\{\pi^2k^2\lambda^2\}\int 0.5\{1+\cos(2\pi kz+2\pi k\lambda x)\}\varphi(x)dx-\cos^2(\pi kz) \\ &= 0.5\exp\{\pi^2k^2\lambda^2\}\left[1+\cos(2\pi kz)\exp\{-2\pi^2k^2\lambda^2\}\right]-0.5\,[1+\cos(2\pi kz)] \\ &= 0.5\left[\exp\{\pi^2k^2\lambda^2\}-\cos(2\pi kz)\right]\cdot[1-\exp\{-\pi^2k^2\lambda^2\}] \\ &\le \pi^2k^2\lambda^2\exp\{\pi^2k^2\lambda^2\}, \end{aligned}$$

as required. □

Lemma 4.2. *Let $\lambda>0$ be fixed and let $Q_{N,\lambda}$ be defined by (3.6). Then for every $z\in[-1,1]$*

$$\begin{aligned} E\,Q_{N,\lambda}(z+\lambda\xi) &= c_0+\sum_{k=1}^N c_k\cos(\pi kz), \\ \mathrm{Var}\,Q_{N,\lambda}(z+\lambda\xi) &\le \varkappa_1^2\lambda^2\exp\{\pi^2N^2\lambda^2\}\log^2(N+1). \end{aligned}$$

with $\varkappa_1\le 2/\pi$.

Proof. The first statement follows directly from the definition of $Q_{N,\lambda}$ and Lemma 4.1. Next, clearly

$$[\mathrm{Var}\,Q_{N,\lambda}(z+\lambda\xi)]^{1/2}\le\sum_{k=1}^N c_k\,[\mathrm{Var}\,\nu_{k,\lambda}(z+\lambda\xi)]^{1/2}$$

and we get by application of Lemma 4.1

$$\begin{aligned}[\mathrm{Var}\, Q_{N,\lambda}(z+\lambda\xi)]^{1/2} &\le \sum_{k=1}^{N} c_k \pi k\lambda \exp\{\pi^2 k^2\lambda^2/2\} \\ &\le \pi\lambda \exp\{\pi^2 N^2\lambda^2/2\} \sum_{k=1}^{N} k c_k \\ &\le 2\lambda\pi^{-1}\exp\{\pi^2 N^2 \lambda^2/2\}\log(N+1)\end{aligned}$$

and the assertion follows. □

Lemma 4.3. *Let* $c_k, k=0,1,\ldots$ *be due to (3.7). Then, for each* $N \ge 1$ *and all* $z \in [-1,1]$,

$$\left| |z| - c_0 - \sum_{k=1}^{N} c_k \cos(\pi k z)\right| \le \varkappa_2 N^{-1}$$

with $\varkappa_2 = 2\pi^{-2}$.

Proof. One has by definition of c_k

$$|z| = c_0 + \sum_{k=1}^{\infty} c_k \cos(\pi k z)$$

and therefore

$$\left| |z| - c_0 - \sum_{k=1}^{N} c_k \cos(\pi k z)\right| \le \sum_{k=N+1}^{\infty} c_k \le \frac{1}{2}\sum_{k=N+1}^{\infty} \frac{4}{(\pi k)^2} \le 2\pi^{-2}N^{-1}$$

as required. □

Now we turn directly to the proof of the result. We use the decomposition (3.2) of the kernel estimate $\tilde{f}_h(t)$. Note first that the Hölder constraint $f \in \Sigma(\beta, L)$ implies in a usual way, see e.g. Ibragimov and Khasminski (1981), that

$$|f_h(t) - f(t)| \le \varkappa_3 L h^{\beta} \tag{4.2}$$

where $\varkappa_3$ depends on β and the kernel K only. Along with the constraint $\|f\|_\infty \le \varrho$, this provides for n large enough and hence h small enough that $|f_h(t)| \le 1$. This allows to apply Lemmas 4.1 and 4.2 with $z = f_h(t)$ and $\lambda = \lambda_h$.

Denote

$$\gamma_n(t) = Q_{N,\lambda_h}(\tilde{f}_h(t))$$

so that $\hat{F}_n = \int_0^1 \gamma_n(t)dt$. Then, in view of the decomposition (3.2) and by Lemma 4.2

$$E\,\gamma_n(t) = c_0 + \sum_{k=1}^{N} c_k \cos(\pi k f_h(t)).$$

Using also Lemma 4.3 and (4.2), we get

$$|E\,\gamma_n(t) - f(t)| \le |E\,\gamma_n(t) - f_h(t)| + |f_h(t) - f(t)| \le \varkappa_2 N^{-1} + \varkappa_3 L h^{\beta}$$

and hence

$$\left| E \int_0^1 \gamma_n(t)dt - \|f\|_1 \right| \le \int_0^1 |E\,\gamma_n(t) - f(t)| \le \varkappa_2 N^{-1} + \varkappa_3 L h^\beta.$$

Next we estimate the variance of our estimator $\hat{F}_n$.

The definition of $\tilde{f}_h(t)$ and the condition $(K.1)$ yield that $\tilde{f}_h(t)$ and $\tilde{f}_h(t')$ are independent random variables when $|t-t'| \ge 2h$. Let $\mathrm{Cov}\,\xi\xi'$ means the covariance $E(\xi - E\,\xi)(\xi' - E\xi')$ between two random variables ξ, ξ'. Using the Cauchy-Schwarz inequality and Lemma 4.2 and we have

$$\begin{aligned} \mathrm{Cov}(\gamma_n(t), \gamma_n(t')) &\le [\mathrm{Var}\,\gamma_n(t)\,\mathrm{Var}\,\gamma_n(t')]^{1/2}\,1(|t-t'| \le 2h) \\ &\le 0.5\,(\mathrm{Var}\,\gamma_n(t) + \mathrm{Var}\,\gamma_n(t'))\,1(|t-t'| \le 2h). \end{aligned}$$

This gives

$$\begin{aligned} \mathrm{Var}\,\hat{F}_n &= \mathrm{Var}\left(\int_0^1 \gamma_n(t)dt\right) = \int_0^1\int_0^1 \mathrm{Cov}(\gamma_n(t), \gamma_n(t'))dt\,dt' \\ &\le 0.5 \int_0^1\int_0^1 (\mathrm{Var}\,\gamma_n(t) + \mathrm{Var}\,\gamma_n(t'))\,1(|t-t'| \le 2h)\,dt\,dt' \\ &\le 4h \int_0^1 \mathrm{Var}\,\gamma_n(t)dt. \end{aligned}$$

Using Lemma 4.2 we get

$$\mathrm{Var}\,\hat{F}_n \le \varkappa_1^2 4\|K\|^2 n^{-1} \exp\{\pi^2 N^2 \|K\|^2/(nh)\} \log^2(N+1).$$

Now

$$\begin{aligned} E\left|\hat{F}_n - \|f\|_1\right| &\le E\left|E\,\hat{F}_n - \|f\|_1\right| + E\left|\hat{F}_n - E\,\hat{F}_n\right| \\ &\le E\left|E\,\hat{F}_n - \|f\|_1\right| + \left[\mathrm{Var}\,\hat{F}_n\right]^{1/2} \\ &\le \varkappa_2 N^{-1} + \varkappa_3 L h^\beta + 2\varkappa_1 \|K\| n^{-1/2} \log(N+1) \exp\left\{\frac{\pi^2 N^2 \|K\|^2}{2nh}\right\}. \qquad (4.3) \end{aligned}$$

By substituting h, N from (3.3), (3.4) respectively, we find out that

$$\theta N^{-1} = L h^\beta = L^{1/(2\beta+1)} (n \log n)^{-\beta/(2\beta+1)}$$

and

$$\exp\{\pi^2 N^2 \|K\|^2/(2nh)\} = \exp\{0.5\pi^2 \|K\|^2 \theta^2 \log n\} = n^{1/(4\beta+2)}.$$

Summing up all these estimates we arrive at (2.2).

4.2. Proof of the upper bound in Theorem 2.3

First we study the behavior of the estimator $\hat{\Phi}_n$ of $\Phi_r(f)$, see (3.8).

Lemma 4.4. *Let $f_h(t)$ be due to (3.3). Then*

$$E\,\hat{\Phi}_n = \int_0^1 f_h^r(t)dt = \|f_h\|_r^r,$$

$$\mathrm{Var}\,\hat{\Phi}_n \le \varkappa_4 n^{-1} \max\{\lambda_h^{2r-2}, \|f_h\|_{2r-2}^{2r-2}\}$$

where $\varkappa_4$ depends only on r and the kernel K.

Proof. We begin by observing that for every two independent standard normal random variables ξ and ξ' and for all $j>0$, one has

$$E(\xi+i\xi')^j=0.$$

(Here $i=\sqrt{-1}$.) This implies also for each numbers z,λ

$$E(z+\lambda\xi+i\lambda\xi')^j=z^j.$$

Now using the decomposition (3.2) of the kernel estimator $\tilde{f}_t(t)$, we have clearly

$$E\,\hat{\Phi}_n=E\int_0^1 E_\xi\left(f_h(t)+\lambda_h\xi_h(t)+i\lambda_h\xi\right)^r dt=\int_0^1 f_h^r(t)dt.$$

Set

$$\gamma_n(t)=E_\xi(\tilde{f}_h(t)+i\lambda_h\xi)^r.$$

Then $E\gamma_n(t)=|f_h(t)|^r$. Using again the decomposition (3.2) we may write

$$\begin{aligned}\gamma_n(t)-E\,\gamma_n(t) &= E_\xi\left(f_h(t)+\lambda_h\xi_h(t)+i\lambda_h\xi\right)^r-|f_h(t)|^r\\ &= \sum_{j=1}^r \mathrm{C}_r^j f_h^{r-j}(t)\lambda_h^j E_\xi\left(\lambda_h\xi_h(t)+i\lambda_h\xi\right)^j.\end{aligned}$$

This yields

$$\operatorname{Var}\gamma_n(t)\le\lambda_h^2\sum_{j=1}^r a_j\lambda_h^{2j-2}|f_h(t)|^{2r-2j}$$

with some positive numbers a_j depending only on r. Now using (4.3) and Jensen's inequality we obtain

$$\begin{aligned}\operatorname{Var}\hat{\Phi}_n &\le 4h\int_0^1\operatorname{Var}\gamma_n(t)dt\\ &\le 4h\lambda_h^2\sum_{j=1}^r\int_0^1 a_j\lambda_h^{2j-2}|f_h(t)|^{2r-2j}dt\\ &\le 4\|K\|^2n^{-1}\sum_{j=1}^r a_j\lambda_h^{2j-2}\|f_h\|_{2r-2}^{2r-2j}\end{aligned}$$

and the assertion follows in an obvious way. □

Lemma 4.5. *There exists a constant $\varkappa_5$ depending only on r and the kernel K such that*

$$\|f_h\|_{2r-2}^{2r-2}\le\varkappa_5 h^{-1+1/r}\|f\|_r^{r-1}\|f_h\|_r^{r-1}.$$

Proof. By application of Minkovski's inequality one get

$$\begin{aligned} |f_h(t)|^{r-1} &= \left| \int f(u) h^{-1} K\left(\frac{t-u}{h}\right) du \right|^{r-1} \\ &\le \left[\left(\int |f(u)|^r du \right)^{1/r} \left(h^{-r/(r-1)} \int |K((t-u)/h)|^{r/(r-1)} du \right)^{(r-1)/r} \right]^{r-1} \\ &= \varkappa_5 h^{-1+1/r} \|f\|_r^{r-1} \end{aligned}$$

where $\varkappa_5 = \int |K(u)|^{r/(r-1)} du$. Now with the help of Jensen's inequality we derive

$$\begin{aligned} \|f\|_{2r-2}^{2r-2} &= \int_0^1 |f_h(t)|^{2r-2} dt \\ &\le \varkappa_5 h^{-1+1/r} \|f\|_r^{r-1} \int_0^1 |f_h(t)|^{r-1} dt \\ &\le \varkappa_5 h^{-1+1/r} \|f\|_r^{r-1} \|f_h\|_r^{r-1} \end{aligned}$$

as required. □

Now we are ready to complete the proof of the theorem. Denote

$$\varrho_n = L^{\frac{1-1/r}{2\beta+1-1/r}} n^{-\frac{\beta}{2\beta+1-1/r}}. \tag{4.4}$$

Then ϱ_n is exactly the rate shown in the theorem and it is easy to check that $\varrho_n = Lh^\beta$ for h from (3.10).

First we recall that the Hölder smoothness constraint implies the bound

$$\|f - f_h\|_r \le \varkappa_3 L h^\beta = \varkappa_3 \varrho_n \tag{4.5}$$

and particularly $\|f_h\|_r \le \|f\|_r + \varkappa_3 \varrho_n$. Below we separate between two cases with $\|f\|_r \le 2\varrho_n$ and $\|f\|_r > 2\varrho_n$. If $\|f\|_r \le 2\varrho_n$, then

$$\begin{aligned} E|\hat{F}_n - \|f\|_r| &\le E|\hat{F}_n| + 2\varrho_n \\ &\le (E\,\hat{\Phi}_n^2)^{1/(2r)} + 2\varrho_n \\ &\le [\mathrm{Var}\,\hat{\Phi}_n + (E\,\hat{\Phi}_n)^2]^{1/(2r)} + 2\varrho_n \\ &\le (\mathrm{Var}\,\hat{\Phi}_n)^{1/(2r)} + (E\,\hat{\Phi}_n)^{1/r} + 2\varrho_n. \end{aligned}$$

It is easily seen that $\varrho_n < \lambda_h^2 = \|K\|^2/(nh)$ at least for n large enough and using the results of Lemma 4.4 we may bound

$$E|\hat{F}_n - \|f\|_r| \le (\varkappa_4 n^{-1} \lambda_h^{2r-2})^{1/(2r)} + \|f_h\|_r + 2\varrho_n.$$

By substituting $\lambda_h = (nh)^{-1/2}$ and h from (3.10) and using the bound (4.5), we get the assertion of the theorem for the considered case.

For the case with $\|f\|_r > 2\varrho_n$, one has also $\|f_h\|_r \ge \|f\|_r - \varrho_n \ge \varrho_n$ and

$$\begin{aligned} E|\hat{F}_n - \|f\|_r| &\le E|\hat{F}_n - \|f_h\|_r| + \varkappa_3\varrho_n \\ &\le \frac{E|\hat{F}_n^r - \|f_h\|_r^r|}{\|f_h\|_r^{r-1}} + \varkappa_3\varrho_n \\ &\le \frac{E|\hat{\Phi}_n - E\,\hat{\Phi}_n|}{\|f_h\|_r^{r-1}} + \varkappa_3\varrho_n \\ &\le \frac{(\operatorname{Var}\hat{\Phi}_n)^{1/2}}{\|f_h\|_r^{r-1}} + \varkappa_3\varrho_n. \end{aligned}$$

The result of Lemma 4.4 and (4.5) allow to bound

$$\operatorname{Var}\hat{\Phi}_n)^{1/2} \le \varkappa_6 n^{-1/2}(\lambda_h^{r-1} + h^{-(r-1)/(2r)}\|f_h\|_r^{r-1})$$

and we end up with

$$E|\hat{F}_n - \|f\|_r| \le \varkappa_6 n^{-1/2}(\lambda_h^{r-1}\varrho_n^{-r+1} + h^{-(r-1)/(2r)}) + \varkappa_3\varrho_n$$

and the theorem follows by straightforward calculation.

4.3. Proof of the lower bound in Theorem 2.3

To get the lower bound announced in Theorem 2.3, we change the original non-parametric set by a high-dimensional parametric subset. Let g be a function from the set $\Sigma_\varrho(\beta, 1)$ vanishing outside the interval $[0,1]$ and with $\|g\|^2 = \int g^2 > 0$. Let some positive number $h < 1$ be fixed such that $N = h^{-1}$ is an integer. We make more precise the choice of h later on. Note that by standard renormalization argument, each function of the form $g_{a,b}(t) = b^{-\beta}g(a+bt)$ also belongs to $\Sigma_\varrho(\beta, L)$ for all a and all positive b.

Let now $\mathcal{I} = \{I_i,\ i = 1, \dots, N\}$ be the partition of the interval $[0,1]$ into $N = h^{-1}$ subintervals of length h. By t_i we denote the left end-point of each subinterval I_i. For every point $\theta = (\theta_1, \dots, \theta_N)$ from N-dimensional cube $B_N = [-1,1]^N$, introduce a function $f_\theta(\cdot)$ by

$$f_\theta(t) = \sum_{i=1}^N \theta_i h^\beta g((t-t_i)/h)1(t \in I_i).$$

Then obviously $f_\theta \in \Sigma_\varrho(\beta, L)$ for N large enough and

$$\|f_\theta\|_r^r \;=\; h^{\beta r}\sum_{i=1}^N |\theta_i|^r \int_{I_i} \left| g\left(\frac{t-t_i}{h}\right)\right|^r dt = \left(\|g\|_r h^\beta F_r(\theta)\right)^r \tag{4.6}$$

where

$$F_r(\theta) = \left(\frac{1}{N}\sum_{i=1}^N |\theta_i|^r\right)^{1/r}. \tag{4.7}$$

Denote also for $i = 1, \dots, N$

$$Y_i = \frac{\sqrt{n}}{\|g\|\sqrt{h}} \int_{I_i} g\left(\frac{t-t_i}{h}\right) dX(t).$$

Using the model equation (2.1) one can write for $f = f_\theta$ from above

$$Y_i = \alpha(N)\theta_i + \xi_i, \qquad i = 1, \dots, N, \tag{4.8}$$

where

$$\alpha(N) = \|g\| n^{1/2} h^{\beta+1/2} = \|g\| n^{1/2} N^{-\beta-1/2},$$
$$\xi_i = \frac{1}{\|g\|\sqrt{h}} \int_{I_i} g\left(\frac{t-t_i}{h}\right) dW(t).$$

Clearly $\xi = (\xi_1, \dots, \xi_N)$ is a collection of independent standard normal random variables. It is also straightforward to see that the set of statistics $Y_i, i = 1, \dots, n$ is sufficient for the parametric submodel (with $f \in \{f_\theta, \theta \in B_N\}$). Therefore, when denoting $s_i = \alpha(N)\theta_i$, $i = 1, \dots, N$, the original "signal + white noise" model (2.1) is transferred into the "sequence space" model

$$Y_i = s_i + \xi_i, \qquad i = 1, \dots, N, \tag{4.9}$$

with $s = (s_1, \dots, s_N)$ from the cube $S_N = B_N^{\alpha(N)} = [-\alpha(N), \alpha(N)]^N$. By this transformation, the original estimation problem is reduced to estimating the quantity $F_r(s)$ due to (4.7) by observations Y. Let $\mathcal{R}_s(N)$ be the corresponding minimax risk:

$$\mathcal{R}_s(N) = \inf_{\hat F} \sup_{s \in S_N} E_s |\hat F - F_r(s)|,$$

the infimum being taken over all Borel functions $\hat F = \hat F(y)$ on R^N and E_s being the expectation under s. Then one gets from (4.6) and (4.8)

$$\mathcal{R}^*(n) \ge \|g\|_r h^\beta \alpha^{-1}(N) \mathcal{R}_s(N) = \varkappa_g \sqrt{N/n}\, \mathcal{R}_s(N) \tag{4.10}$$

where $\varkappa_g = \|g\|_r / \|g\|$.

Now we are going to establish the following

Proposition 4.1. *Let $\alpha(N) = N^{-1/(2r)}$. Then, for all large enough values of N,*

$$\mathcal{R}_s(N) \ge \varkappa_7 \alpha(N), \tag{4.11}$$

where $\varkappa_7 > 0$ depends on r only.

The proof of this assertion will be given below. Before doing this, we show how it implies the statement of the theorem. We set

$$N = (L^2 n)^{\frac{1}{2\beta+1-1/r}}.$$

Then (4.10) and (4.11) give

$$\mathcal{R}^*(n) \ge \varkappa_8 \sqrt{N/n} N^{-1/(2r)} = \varkappa_8 L^{\frac{1-1/r}{2\beta+1-1/r}} n^{-\frac{2\beta}{2\beta+1-1/r}}$$

as required.

Proof of Proposition 4.1 is based on the following idea. We introduce two prior measures $\mu_{N,0}$ and $\mu_{N,1}$ on the parameter set S_N and denote by $P_{N,0}$ and $P_{N,1}$ the corresponding Bayes measures on R^N,

$$P_{N,j} = \mu_{N,j} * \mathcal{L}(\xi), \qquad j = 0, 1.$$

Let also $\mathcal{K}(P_{N,0}, P_{N,1})$ be the Kullback information between $P_{N,0}$ and $P_{N,1}$

$$\mathcal{K}(P_{N,0}, P_{N,1}) = \int \log\left(\frac{dP_{N,1}}{dP_{N,0}}\right) dP_{N,1}.$$

We will estimate the minimax risk from below by the maximum of two risks under $P_{N,0}$ and $P_{N,1}$. For this we use the following technical assertion which can be deduced from more general Fano's lemma. However, we prefer to give a direct proof.

Lemma 4.6. *Let prior measures* $\mu_{N,0}$ *and* $\mu_{N,1}$ *be such that the Kullback information* $\mathcal{K}(P_{N,0}, P_{N,1})$ *satisfies the condition*

$$\mathcal{K}(P_{N,0}, P_{N,1}) \le \varkappa \tag{4.12}$$

with some positive $\varkappa$. *Let now* Φ *be some function on the parametric set* S_N *and let*

$$v_{N,j} = \int \Phi(s)\mu_{N,j}(ds), \tag{4.13}$$

$$d_{N,j}^2 = \int (\Phi(s) - v_{N,j})^2 \mu_{N,j}(ds), \tag{4.14}$$

for $j = 0, 1$, *then*

$$R(N) \equiv \inf_{\hat{\Phi}} \sup_{s\in S_N} E_s|\hat{\Phi} - \Phi(s)| \ge 0.5|v_{N,0} - v_{N,1}|e^{-\varkappa} - \max\{d_{N,0}, d_{N,1}\}.$$

Proof. First we note that, for an arbitrary prior measure μ and each estimator $\hat{\Phi}$ of $\Phi(s)$, one has

$$\begin{aligned}\sup_{s\in S_N} E_s|\hat{\Phi} - \Phi(s)| &\ge E_{N,\mu}|\hat{\Phi} - \Phi(s)| \\ &\ge E_{N,\mu}|\hat{\Phi} - E_{N,\mu}\Phi(s)| - E_{N,\mu}|\Phi(s) - E_{N,\mu}\Phi(s)| \\ &\ge E_{N,\mu}|\hat{\Phi} - E_{N,\mu}\Phi(s)| - d_{N,\mu}.\end{aligned}$$

Here $E_{N,\mu}$ means the expectation w.r.t. the Bayes measure $P_{N,\mu}$ corresponding to prior μ and $d_{N,\mu}$ is due to (4.14). This clearly implies

$$\begin{aligned}R(N) &\ge \inf_{\hat{\Phi}} \max\left\{E_{N,0}|\hat{\Phi} - v_{N,0}| - d_{N,0}, E_{N,1}|\hat{\Phi} - v_{N,1}| - d_{N,1}\right\} \\ &\ge \inf_{\hat{\Phi}} \max\left\{E_{N,0}|\hat{\Phi} - v_{N,0}|, E_{N,1}|\hat{\Phi} - v_{N,1}|\right\} \\ &\qquad - \max\{d_{N,0}, d_{N,1}\}.\end{aligned} \tag{4.15}$$

Next we use the fact that the maximum likelihood test $\hat{T}_N = \mathbf{1}(dP_{N,1}/dP_{N,0} > 1)$ is optimal for testing the hypothesis $H_0 : \mathcal{L}(Y) = P_{N,0}$ versus the alternative $H_1 : \mathcal{L}(Y) = P_{N,1}$, see Lehmann (1959): for an arbitrary test T_N,

$$\max\{P_{N,0}(T_N = 1), P_{N,1}(T_N = 0)\} \ge \max\left\{P_{N,0}(\hat{T}_N = 1), P_{N,1}(\hat{T}_N = 0)\right\}.$$

Set $Z_N = dP_{N,0}/dP_{N,1}$. Then $\hat{T}_N = \mathbf{1}(Z_N \le 1)$ and, since the function $\log(z)$ is concave, using Jensen's inequality we get

$$\begin{aligned}\log\max\left\{P_{N,0}(\hat{T}_N = 1), P_{N,1}(\hat{T}_N = 0)\right\} &\ge \log P_{N,0}(Z_N \le 1)\\ &= \log\int Z_N\mathbf{1}(Z_N \le 1)dP_{N,1}\\ &\ge \int \log(Z_N)\mathbf{1}(\log(Z_N) \le 0)dP_{N,1}\\ &\ge -\mathcal{K}(P_{N,0}, dP_{N,1}) \ge -\varkappa.\end{aligned}$$

Let now $\hat{\Phi}$ be an estimator of $\Phi(s)$. Consider the following test

$$T_N = \mathbf{1}(\hat{\Phi} - v_{\mu,0} > \Delta_N)$$

where

$$\Delta_N = (v_{N,1} - v_{N,0})/2.$$

(Here we assume that $v_{N,1} > v_{N,0}$.) By application of the above inequalities we obtain

$$\max\{P_{N,0}(T_N = 1), P_{N,1}(T_N = 0)\} \ge e^{-\varkappa}$$

or

$$\max\left\{P_{N,0}(\hat{\Phi} - v_{N,0} > \Delta_N), P_{N,1}(\hat{\Phi} - v_{N,1} < -\Delta_N)\right\} \ge e^{-\varkappa}.$$

We end up by use of (4.15) and of Chebyshev's inequality. □

We will apply this lemma with $\Phi(s) = N^{-1}(s_1^r + \ldots + s_N^r)$ to two prior measures $\mu_{N,0}$ and $\mu_{N,1}$ with product structure,

$$\begin{aligned}\mu_{N,0} &= \mu_0^N,\\ \mu_{N,1} &= \mu_1^N.\end{aligned}$$

We construct these measures in such a way that (4.12) holds with some fixed $\varkappa$ and the difference $|v_{N,1} - v_{N,0}|$ will be as large as possible.

First we note that, for $j = 0, 1$,

$$v_{N,j} = \frac{1}{N}\int\sum_{i=1}^N |s_i|^r\mu_{N,j}(ds) = \int |s|^r\mu_j(ds) = v_j$$

and similarly

$$d_{N,j}^2 = \frac{1}{N^2}\int\sum_{i=1}^N(|s_i|^{2r} - v_j^2)\mu_{N,j}(ds) = N^{-1}\int(|s|^{2r} - v_j^2)\mu_j(ds) = N^{-1}d_j^2$$

where

$$\begin{aligned}v_j &= \int |s|^r\mu_j(ds) \le \alpha^r(N)\\ d_j^2 &= \int |s|^{2r}\mu_j(ds) - v_j^2 \le \alpha^{2r}(N).\end{aligned}$$

In the same way we can estimate the Kullback distance between the Bayes measures $P_{N,0}$ and $P_{N,1}$. The product structure of the model (4.9) and of the priors $\mu_{N,0}, \mu_{N,1}$ implies that

$$\mathcal{K}(P_{N,0}, P_{N,1}) = N \int \log(p_{\mu_0}(y)/p_{\mu_1}(y)) p_{\mu_0}(y) dy \tag{4.16}$$

where, for each measure μ on $[0,1]$

$$p_\mu(y) = \int \varphi(y-t)\mu(dt),$$

$\varphi(y) = (2\pi)^{-1}\exp\{-y^2/2\}$ be the standard Gaussian density on the axis.

Now the application of Lemma 4.6 gives under condition (4.12) the following bound of the risk of an arbitrary estimate $\hat{\Phi}$ of $\Phi(s)$

$$\sup_{s\in S_N} E_s|\hat{\Phi} - \Phi(s)| \ge 0.5|v_1 - v_0|e^{-\varkappa} - \alpha^r(N)N^{-1/2}. \tag{4.17}$$

Next we observe what follows from this bound for the risk $\mathcal{R}_s(N)$ in estimating $F_r(s)$. If $\hat{F}$ is an estimate of $F_r(s)$, then $\hat{\Phi} = \hat{F}^r$ can be viewed as an estimate of $\Phi(s) = F_r^r(s)$. We may assume that $|\hat{F}| \le \alpha(N)$ and hence

$$E_s|\hat{\Phi} - \Phi(s)| = E_s|\hat{F}^r - F_r^r(s)| \le r\alpha^{r-1}(N)E_s|\hat{F} - F_r(s)|.$$

Now the bound (4.17) yields

$$\begin{aligned} \mathcal{R}_s(N) &\ge (r\alpha^{r-1}(N))^{-1}(0.5|v_1 - v_0|e^{-\varkappa} - \alpha^r(N)N^{-1/2}) \\ &= r^{-1}\alpha(N)(0.5\alpha^{-r}(N)|v_1 - v_0|e^{-\varkappa} - N^{-1/2}). \end{aligned} \tag{4.18}$$

Next we specify the choice of measures μ_0, μ_1 mentioned above.

Let δ be the distance (in the uniform norm on $[-1,1]$) from the function t^r to the space of polynomials of degree $\le r-2$. By the standard separation arguments, there exists a measure μ with variation 2 on $[-1,1]$ such that

$$\begin{aligned} \int t^l \mu(dt) &= 0, \qquad l = 0,1,...,r-2, \\ \int t^r \mu(dt) &= 2\delta. \end{aligned}$$

Note that if μ possesses the indicated properties, so is the "reflected" measure μ^* ($\mu^*(A) = \mu(-A)$) and hence the measure $(\mu+\mu^*)/2$; therefore μ may be assumed to be symmetric. Let $\mu_+, -\mu_-$ be the positive and the negative components of μ, respectively. Then μ_+ and μ_- are symmetric probability distributions on $[-1,1]$ such that $\mu = \mu_+ - \mu_-$ and

$$\begin{aligned} \int t^l \mu_+(dt) &= \int t^l \mu_-(dt), \ l = 0,1,...,r-2; \\ \int t^r \mu_+(dt) &= \int t^r \mu_-(dt) + 2\delta. \end{aligned} \tag{4.19}$$

We assign μ_0, μ_1 by rescaling the measures μ_+, μ_- respectively into the interval $[-\alpha(N), \alpha(N)]$,

$$\mu_0([a,b]) = \mu_+([a/\alpha(N), b/\alpha(N)]), \qquad a,b \in [-\alpha(N), \alpha(N)],$$

and similarly for μ_1. Obviously

$$v_0 - v_1 = \alpha^r(N) \int |t|^r \mu(dt) = 2\delta\alpha^r(N)$$

and now the bound (4.18) looks like

$$\mathcal{R}_s(N) \geq r^{-1}\alpha(N)(\delta e^{-\varkappa} - N^{-1/2}). \tag{4.20}$$

To complete the proof, we have only to verify (4.12).

Let us associate with a symmetric probability distribution ν on $[-1, 1]$ and a real α the distribution F_ν^α on the axis with the density

$$p_\nu(\alpha, y) = \int \varphi(y - \alpha t)\nu(dt) = \varphi(y) \int \mathrm{ch}(\alpha t x) \exp\{-\alpha^2 t^2/2\}\nu(dt).$$

Note that this relation defines a function $p_\nu(\alpha, y)$ for an arbitrary (not necessarily nonnegative) symmetric measure ν on $[-1, 1]$.

Set

$$\mathcal{K}(\alpha) = \int \log(p_{\mu_+}(\alpha, y)/p_{\mu_-}(\alpha, y))p_{\mu_+}(\alpha, y)dy$$

for the Kullback distance from $p_{\mu_+}(\alpha, \cdot)$ to $p_{\mu_-}(\alpha, \cdot)$.

Lemma 4.7. *The function $\mathcal{K}(\alpha)$ is infinitely differentiable and it has zero of order at least $2r$ at the point $\alpha = 0$.*

Proof. It is clearly seen that one may differentiate $\mathcal{K}(\alpha)$ arbitrarily many times and that

$$\mathcal{K}^{(l)}(\alpha) = \int \frac{\partial^l}{\partial \alpha^l} \left[\log \left(\frac{p_{\mu_+}(\alpha, y)}{p_{\mu_-}(\alpha, y)} \right) p_{\mu_+}(\alpha, y) \right] dy$$

for all l. Note that

$$p_{\mu_+}(\alpha, y) = p_{\mu_-}(\alpha, y) + p_\mu(\alpha, y).$$

To begin by, we show that for all x

$$\left. \frac{\partial^l p_\mu(\alpha, y)}{\partial \alpha^l} \right|_{\alpha=0} = 0, \qquad l = 0, 1, ..., r-1. \tag{4.21}$$

Indeed, one clearly has

$$\left. \frac{\partial^l p_\mu(\alpha, y)}{\partial \alpha^l} \right|_{\alpha=0} = \varphi(x) \int \left[\sum_{i=0}^{l} \mathrm{C}_l^i \left(\frac{\partial^i \exp\{-\alpha^2 t^2/2\}}{\partial \alpha^i} \right) \left(\frac{\partial^{l-i} \mathrm{ch}(\alpha t y)}{\partial \alpha^{l-i}} \right) \right] \mu(dt) \Bigg|_{\alpha=0}$$

$$= \int t^l (a_0 + a_1 y + \ldots + a_l y^l) \mu(t) = 0.$$

Here $a_0, \ldots, a_l$ are some numbers and we have used (4.19). This yields (4.21).

According to (4.21), $p_\mu(\alpha, y)$ can be represented in the form

$$p_\mu(\alpha, y) = \alpha^r w(\alpha, y)$$

with smooth function $w(\cdot, \cdot)$ (which, as it is easily seen, is a summable function of y). Since $\int p_\mu(\alpha, y)dy = 0$ for all α, so is also for $w(\alpha, y)$,

$$\int w(\alpha, y)dy = 0, \qquad \forall \alpha.$$

Now we have

$$\log\left(\frac{p_{\mu_-}(\alpha,y)}{p_{\mu_+}(\alpha,y)}\right) = \log\left(1-\frac{\alpha^r w(\alpha,y)}{p_{\mu_+}(\alpha,y)}\right) = -\frac{\alpha^q w(\alpha,y)}{p_{\mu_-}(\alpha,y)} - \alpha^{2r} v(\alpha,y),$$

v being a smooth function of y,α. Hence

$$\begin{aligned}\mathcal{K}(\alpha) &= -\int \log\left(\frac{p_{\mu_-}(\alpha,y)}{p_{\mu_+}(\alpha,y)}\right) p_{\mu_+}(\alpha,y)dy \\ &= \alpha^r \int w(\alpha,y)dy + \alpha^{2r}\int v(\alpha,y)p_{\mu_+}(\alpha,y)dy \\ &= \alpha^{2r}\int v(\alpha,y)p_{\mu_+}(\alpha,y)dy\end{aligned}$$

and the assertion follows. □

The result of this lemma means that, for α small, the following bound holds true

$$\mathcal{K}(\alpha) \le \varkappa\alpha^{2r}.$$

Particularly, by letting $\alpha(N) = N^{-1/(2r)}$, we get

$$\mathcal{K}(\alpha) \le \varkappa N^{-1}, \qquad \forall\alpha \le \alpha(N) \tag{4.22}$$

and the assertion (4.12) follows in view of (4.16).

4.4. Proof of the lower bound in Theorem 2.2

Now we establish the lower bound from the Theorem 2.2 for the case when r is not an even integer.

We follow the line of the proof of the similar result in Theorem 2.3. The only difference is in construction of two priors μ_0 and μ_1.

We begin by translation of the problem into the "sequence space" model (4.9). We apply now

$$N = L^{2/(2\beta+1)}(n\log n)^{\beta/(2\beta+1)}.$$

The bound (4.10) for $\mathcal{R}^*(n)$ is still valid and the statement of the theorem follows from this bound and the next proposition which delivers some information about accuracy of estimation of the functional $F_r(s) = (N^{-1}(s_1^r + \ldots + s_N^r))^{1/r}$ by observation Y from the sequence space model (4.9).

Proposition 4.2. *Let* $\alpha(N) = (100\log N)^{-1}$ *and* $S_N = [-\alpha(N),\alpha(N)]^N$. *Then for all large enough values of* N,

$$\mathcal{R}_s(N) \equiv \inf_{\hat F}\sup_{s\in S_N} E_s|\hat F - F_r(s)| \ge \varkappa_9(\log N)^{-r-1/2} \tag{4.23}$$

where $\varkappa_9 > 0$ *depends on* r *only.*

Proof. The most important step in the proof deals with constructing two measures μ_+ and μ_-. Denote by $\mathcal{P}_k$ the space of polynomials of degree k, and let $\delta(k)$ be the distance (in the uniform norm on $[-1,1]$) from the function $|t|^r$ to the space $\mathcal{P}_{2k}$. It is known (see, e.g., Timan A.F., *Theory of approximation of* *functions of real variable*, Moscow, 1960, p.430) that if k is a nonnegative integer, then

$$\delta(k) \geq \varkappa_{10} k^{-r}.$$

Given a positive integer $N > 3$, let us set

$$k(N) = \lfloor \log N \rfloor.$$

By the standard separation arguments, for a given N there exists a measure μ_N with variation 2 on $[-1,1]$ such that

$$\begin{aligned} \int t^l \mu_N(dt) &= 0, \ l = 0, 1, ..., 2k(N), \\ \int |t|^r \mu_N(dt) &= 2\delta(k(N)) \geq 2\varkappa_{10} k^{-r}(N). \end{aligned} \tag{4.24}$$

Arguing as in the proof of Proposition 4.1 we may assume from the beginning that the measure μ_N is symmetric and so are its positive and negative components μ_+ and μ_- (i.e. $\mu_N = \mu_+ - \mu_-$).

We define now measures μ_0 and μ_1 by rescaling μ_+ and μ_- into the interval $[-\alpha(N), \alpha(N)]$. Also we set $\mu_{N,0} = \mu_0^N$, $\mu_{N,1} = \mu_1^N$ and Bayes measures $P_{N,0}, P_{N,1}$ correspond to these priors. Following the arguments from the proof of Theorem 2.3 we arrive at the bound (4.20)

$$\mathcal{R}_s(N) \geq r^{-1} \alpha(N)(\delta(N) e^{-\varkappa} - N^{-1/2}). \tag{4.25}$$

under the condition

$$\mathcal{K}(P_{N,0}, P_{N,1}) \leq \varkappa. \tag{4.26}$$

If this condition holds true with some positive $\varkappa$ depending only on r, then the bound (4.25) yields the desirable assertion. Therefore, it remains to check (4.26).

Recall that the Kullback distance $\mathcal{K}(P_{N,0}, P_{N,1})$ satisfies

$$\mathcal{K}(P_{N,0}, P_{N,1}) = N\mathcal{K}(\alpha(N)) \tag{4.27}$$

where by definition, for $\alpha \in [-1,1]$ and a measure ν

$$\mathcal{K}(\alpha) = \int \log(p_{\mu_+}(\alpha, y)/p_{\mu_-}(\alpha, y)) p_{\mu_+}(\alpha, y) dy,$$

$$p_\nu(\alpha, y) = \int \varphi(y - \alpha t) \nu(dt) = \varphi(y) \int \operatorname{ch}(\alpha t x) \exp\{-\alpha^2 t^2/2\} \nu(dt).$$

Set for $T > 0$

$$K_T(\alpha) = \int_{|y| \leq T} \log(p_{\mu_+}(\alpha, y)/p_{\mu_-}(\alpha, y)) p_{\mu_+}(\alpha, y) dy. \tag{4.28}$$

Lemma 4.8. *For every* $T > 0$

$$\left. \frac{d^l K_T(\alpha)}{d\alpha^l} \right|_{\alpha=0} = 0, \ l = 0, ..., 2k(N).$$

Proof basing on (4.24) repeats the first part of the proof of Lemma 4.7.

Remark 4.1. Lemma 4.7 claims more strong assertion: if (4.24) holds for all $l \le 2k$, then $\mathcal{K}(\alpha)$ has zero of order $2 \cdot 2k$ at zero. For this we use except (4.24) also the property $\int p_\mu(\alpha, y)dy = 0$ which is not available when dealing with $K_T(\alpha)$.

The next lemma delivers more information about behavior of the function $K_T(\alpha)$.

Lemma 4.9. *For every $T \ge 20$ and all $\alpha \in [-1,1]$, one has*

$$\mathcal{K}(\alpha) \le \exp\{-(T-1)^2/2\} + K_T(\alpha). \tag{4.29}$$

The function $K_T(\alpha)$ can be extended analytically onto the circle $|\alpha| \le (10T)^{-1}$, and in this circle

$$|K_T(\alpha)| \le 2/3.$$

Proof. We clearly have

$$\begin{aligned} \mathcal{K}(\alpha) &= K_T(\alpha) + R_T, \\ R_T &= \int_{|y|>T} \log(p_{\mu_+}(\alpha,y)/p_{\mu_-}(\alpha,y))p_{\mu_+}(\alpha,y)dy. \end{aligned}$$

Now, R_T is a convex functional of the distributions μ_+, μ_-; therefore its supremum, over all (even non-symmetric) probability distributions on $[-1,1]$ is the same as its supremum over distributions on the same segment with singleton supports. For a distribution of this latter type, with μ_+ concentrated at a point t and μ_- concentrated at a point τ ($t, \tau \in [-1,1]$), we have

$$\begin{aligned} R_T &= \int_{|y|>T} \left[-\frac{(y-\alpha t)^2}{2} + \frac{(y-\alpha\tau)^2}{2} \right] \exp\{-\frac{(y-\alpha t)^2}{2}\} \frac{1}{\sqrt{2\pi}} dy \\ &= \int_{\{y \le -T-\alpha t\} \cup \{y \ge T - \alpha t\}} \left[\alpha(t-\tau)y + \alpha^2(t-\tau)^2/2 \right] \varphi(y)dy \\ &= \alpha(t-\tau)(2\pi)^{-1/2} \left[\exp\{-(T-\alpha t)^2/2\} - \exp\{-(T+\alpha t)^2/2\} \right] \\ &\quad + 2(2\pi)^{-1/2}\alpha^2(t-\tau)^2(T-1)^{-1}\exp\{-(T-1)^2/2\} \\ &\le (2\pi)^{-1/2}(2 + 8(T-1)^{-1})\exp\{-(T-1)^2/2\} \\ &\le \exp\{-(T-1)^2/2\} \end{aligned}$$

(we have taken into account that $T \ge 20$). Consequently, $R_T \le \exp\{-(T-1)^2/2\}$, and (4.29) follows.

Now let us look at the function K_T. Let y be a real with $|y| < T$, and let t be a real with $|t| \le 1$. The absolute value of the derivative of the function $g(\alpha) = \exp\{-\alpha^2 t^2/2\}\mathrm{ch}(\alpha t y)$ in the circle $|\alpha| \le z \le 1$ clearly does not exceed $(T+1)\exp\{zT + z^2/2\}$, and therefore $|g(\alpha) - 1| = |g(\alpha) - g(0)| \le (zT + z)\exp\{zT + z^2/2\}$ in this circle. It follows that in the circle $|\alpha| \le z \equiv (10T)^{-1}$ we have

$$\begin{aligned} &\left| \int \exp\{-\alpha^2 t^2/2\}\mathrm{ch}(\alpha t y)\nu(dt) - 1 \right| \\ &\qquad \le (zT + z)\exp\{zT + z^2/2\} \le 1/5\exp\{0.105\} \le 1/4, \end{aligned}$$

both for $\nu = \mu_+$ and for $\nu = \mu_-$. Consequently, for the indicated z and $|\alpha| \le z$ we have

$$\left| \frac{p_{\mu_+}(\alpha,y)}{p_{\mu_-}(\alpha,y)} - 1 \right| \le 1/3.$$

We see that if y is real with $|y| \le T$, then the function $\log(p_{\mu_+}(\alpha, y)/p_{\mu_-}(\alpha, y))$, regarded as a function of α, can be extended as an analytic function from the segment $|\alpha| \le z = (10T)^{-1}$ of the real axis onto the circle $|\alpha| \le z$ in the complex plane, and the absolute value of the extended function does not exceed in this circle the quantity

$$\sum_{m=1}^{\infty} \frac{1}{m} \left(\frac{1}{3}\right)^m = \log(3/2).$$

By the same reasons, for real y with $|y| \le T$ and α from the circle $|\alpha| \le z$ we have $|p_{\mu_+}(\alpha, y)| \le 5/4\varphi(y)$, and we see that indeed K_T is an analytic function in the circle $|\alpha| \le s_T$ with absolute value in the circle not exceeding $5/4 \log 3/2 \le 2/3$. □

According to results of Lemmas 4.7 and 4.8, $K_T(\alpha)$ is an analytic function of α in the circle $|\alpha| \le z = (10T)^{-1}$ which is bounded in absolute value in this circle by $2/3$ and has zero of order at least $2k(N)$ at the origin; since this function is even, the order of zero is at least $2k(N)+1$. Applying to the function $K_T(\alpha)z^{2k(N)+2}\alpha^{-2k(N)-2}$ the Maximum Principle, we come to

$$K_T(\alpha) \le \frac{2}{3} \frac{\alpha^{2k(N)+2}}{z^{2k(N)+2}}, \qquad -z \le \alpha \le z. \tag{4.30}$$

Now let us look what (4.30) implies for $\alpha(N) = (100\sqrt{\log N})^{-1}$. We have

$$\frac{\alpha(N)}{z} \le \frac{1+\sqrt{2}}{10} \le \exp\{-1\},$$

and (4.30) implies that

$$K_T(\alpha(N)) \le \exp\{-2k(N) - 2\} \le N^{-2}. \tag{4.31}$$

From this inequality using also (4.29) and (4.27) we conclude that the Kullback distance $\mathcal{K}(P_{N,0}, P_{N,1})$ does not exceed $N^{-1} + N\exp\{-(T(N)-1)^2/2\} = N^{-1} + 1$. This yields (4.26) and the assertion follows by (4.25).

References

[1] Birgé, L. and Massart, P. (1995). Estimation of integral functionals of a density. *Ann. Statist.* **23**, 11–29.

[2] Bickel, P.J. and Ritov, Y. (1988) Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhya*, **50**, 381–393.

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

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

[5] Efroimovich, S. and Low, M. (1996). On Bickel and Ritov's conjecture about adaptive estimation of the integral of the square of density derivative. *Ann. Statist.* **24**, 682–686.

[6] Fan, J. (1991). On the estimation of quadratic functionals. *Ann. Statist.* **19**, 1273–1294.

[7] Hall, P. and Marron, J.S. (1987). Estimation of integrated squared density derivatives. *Statist. and Prob. Letters*, **6**, 109–115.

[8] Hasminski, R.Z. and Ibragimov, I.A. (1979). On the nonparametric estimation of functionals. In *Proceedings of the Second Prague Symp. on Asymptotic Statistics* (P.Mandl and M.Huskova eds.) North-Holland, Amsterdam. 41–51.

[9] Hasminski, R.Z. and Ibragimov, I.A. (1980). Some estimation problems for stochastic differential equations. *Lecture Notes Control Inform. Sci.* Springer, New York. **25**, 1–12.

[10] Ibragimov, I.A. and Khasminski, R.Z. (1980). On the estimation of distribution density. *Zap. Nauch. Sem. LOMI*, **98**, 61–85.

[11] Ibragimov, I.A. and Khasminski, R.Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer. Berlin, Heidelberg, New York.

[12] Ibragimov, I.A. and Nemirovski and Khasminski, R.Z. (1986). Some problems on nonparametric estimation in Gaussian white noise. *Theory Probab. Appl.*, **31**, 391–406.

[13] Ibragimov, I.A. and Khasminski, R.Z. (1987). Estimation of linear functionals in Gaussian noise. *Theory Probab. Appl.*, **32**, 30–39.

[14] Ibragimov, I.A. and Khasminski, R.Z. (1991). Asymptotic normal families of distributions and effective estimation. *Ann. Statist.*, **19**, 1681–1724.

[15] Ingster, Yu.I. (1982). Minimax nonparametric detection of signals in white Gaussian noise. *Problems Inform. Transmission*, **18** 130 – 140.

[16] Ingster, Yu.I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I–III. *Math. Methods of Statist.* **2** (1993) 85 – 114, **3** (1993) 171 – 189, **4** (1993) 249 – 268.

[17] Kerkyacharian, J. and Picard, D. (1996). Estimating nonquadratic functionals of a density using Haar basis. *Ann: Statist.*, **24**, 485–508.

[18] Korostelev, A.P. (1990). On the accuracy of estimation of non-smooth functionals of regression. *Theory Probab. Appl.*, **35**, 768–770.

[19] Korostelev, A.P. and Tsybakov, A.B. (1994). *Minimax Theory of Image Reconstruction.* Lecture Notes in Statist. Springer, New York.

[20] Koshevnik, Yu. and Levit, B. Ya. (1976). On a nonparametric analogue of the information matrix. *Theory Probab. Appl.*, **21**, 738–753.

[21] Laurent, B. (1996) Efficient estimation of integral functionals of a density. *Ann: Statist.*, **24**, 659–682.

[22] Lehmann, E.L. (1959) *Testing Statistical Hypothesis* Wiley, New York.

[23] Lepski, O. and Spokoiny, V. (1995). Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. *Bernoulli*, to appear.

[24] Levit, B.Ya. (1974). On optimality of some statistical estimates. In *Proceedings of the Prague Symp. on Asymptotic Statistics* (J. Hajek, ed.) Univ. Karlova, Prague. **2**, 215–238.

[25] Levit, B.Ya. (1975). Efficiency of a class of nonparametric estimates. *Theory Probab. Appl.*, **20**, 738–754.

[26] Levit, B.Ya. (1978). Asymptotically efficient estimation of nonlinear functionals. *Problems info. Transmission*, **14**, 65–72.

[27] Spokoiny, V. (1996). Adaptive hypothesis testing using wavelets. *Annals of Statistics*, **26**.

Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstr. 39, 10117 Berlin, Germany