Abstract

We discuss extremal problems which arise in various nonparametric statistical settings with Hölder function classes. We establish a new property of the solution of an optimal recovery problem that leads to the exact constants for asymptotic minimax risks. [1]

1. Introduction

Consider a classic nonparametric model

$$y_{in} = f(i/n) + \xi_{in}, \quad i = 1, 2, \ldots, n; \quad n = 1, 2, \ldots \tag{1}$$

where y_{in} are observations, ξ_{in} are i.i.d. $(0, \sigma^2)$- Gaussian errors, f is an unknown function belonging to Hölder smoothness class $\Sigma(\beta, L)$:

$$\Sigma(\beta, L) \; = \; \{ f : \; |f^{(m)}(x) - f^{(m)}(x_1)| \le L|x - x_1|^{\alpha} \; \} \;\; , \tag{2}$$

here $m = \lfloor \beta \rfloor$ is an integer such that $0 < \alpha \le 1$, $\alpha = \beta - m$; β, L are given positive constants. As discovered recently, the exact minimax estimators in several set-ups with Hölder classes (2) are closely related to the solution g_β of the extremal problem

$$g(0) \to \sup, \quad \int_{-\infty}^{\infty} g^2(t)dt \le 1 \;, \quad g \in \Sigma(\beta, 1) \;, \tag{3}$$

which was introduced into the nonparametrics by D.Donoho (see, e.g. Donoho (1994)) and which is often referred to as "optimal recovery problem".
To illustrate results, define a supremum norm minimax risk of a function class Σ by

$$R_n(\Sigma) \; = \inf_{\tilde{f}_n} \; \sup_{f \in \Sigma} E_f [d(\tilde{f}_n - f)]^2 \;, \tag{4}$$

where $E_f(\cdot)$ is the expectation w.r.t. the true regression f, $\tilde{f}_n$ is an arbitrary estimator; $d(\cdot)$ is a uniform norm in $[0, 1]$,

$$d(g) = d_\infty(g) \; = \; \sup_{t \in [0,1]} |g(t)| \tag{5}$$

(for the sake of simplicity, we discuss only quadratic loss function which entails mean squared error risk). As proved by Korostelev (1993), $\beta \le 1$, and by Donoho (1994), $\beta > 1$,

$$\lim_{n \to \infty} R_n(\Sigma(\beta, L)) \; / \; \psi^2_{n,\beta} \; = \; C(\beta, L, \sigma^2) \; D(g_\beta) \;, \tag{6}$$

[1]The author is grateful to A.Korostelev for the helpful remarks.

where $\psi_{n,\beta}$ is a well-known rate of convergence for the sup-norm minimax risk,

$$\psi_{n,\beta} \;=\; (\log n/n)^{\beta/(2\beta+1)} \;, \tag{7}$$

$C(\beta, L, \sigma^2)$ is a constant explicitly depending on parameters β, L, σ^2; $D(g_\beta)$ is a functional explicitly depending on the solution of (3). What's important, the lower bound of the minimax risk $R_n(\Sigma(\beta, L))$ is attained at the kernel estimator

$$\hat{f}_n(t) = \frac{1}{n} \sum_i y_{in} \frac{1}{h} K(\frac{t - i/n}{h}) \;, \tag{8}$$

with certain modifications near the boundaries, where $h = h(n)$ is a proper bandwidth and $K(t) = K_\beta(t) = g_\beta(t)/(\int g_\beta(s)ds)$. Moreover, function g_β underlies exact asymptotic estimators in other nonparametric settings with Hölder classes: density estimation (Korostelev and Nussbaum (1995)), adaptive estimation (Lepski (1992), Lepski and Spokoiny (1995)), minimax estimation with risks based on large deviation probabilities (Korostelev (1996), Korostelev and Leonov (1996)).
It is easier to deal with the solution x_β of the extremal problem dual to (3):

$$\|x\| \to \inf \;,\quad x(0) = 1 \;,\quad x \in \Sigma(\beta, 1) \;, \tag{9}$$

where $\|x\| = (\int_{-\infty}^{\infty} x^2(t)dt)^{1/2}$. As shown by Donoho (1994), Section 2,

$$x_\beta(t) = a g_\beta(bt), \quad a = [g_\beta(0)]^{-1}, \quad b = [g_\beta(0)]^{1/\beta}, \tag{10}$$

$$g_\beta(t) = a_1 x_\beta(b_1 t), \quad a_1 = \|x_\beta\|^{\frac{-2\beta}{2\beta+1}}, \quad b_1 = \|x_\beta\|^{\frac{2}{2\beta+1}} \;. \tag{11}$$

Further on we shall discuss the properties of x_β while corresponding properties of g_β will be obtained as a consequence of (10) and (11).
So far, the solution of (9) in the closed form is known only for 2 special cases:
(a) $\beta \le 1$ - the solution is straightforward (Korostelev (1993)): $x_\beta(t) = (1 - |t|^\beta)_+$.
(b) $\beta = 2$ - the solution was given by Fuller (1960), see also Zhao (1997).
For the general case, it was proved by Leonov (1995) that x_β has a compact support for any $\beta > 1$. The goal of the present paper is to prove a new identity for the solutions of (9) which establishes a relation between L_2-norm of x_β and its integral, $\int x_\beta(s)ds$, and which is given in Theorem 1.
The paper is organized as follows. In Section 2 we discuss how optimal recovery

problems (3), (9) are related to other deterministic optimization problems traditionally arising in nonparametric settings with Hölder classes. Main results are presented in Section 3. Section 4 contains the comparison of various kernels for Hölder class with $\beta = 2$.

2. Preliminaries: bias-variance balance and optimization problems.

Asymptotic inference for deriving upper bounds in nonparametric regression (as well as in density estimation) is based on the analysis of the bias-variance balance of estimators. Here we compare the standard approach exploiting the Taylor expansion for the bias term, with renormalization techniques used by Donoho (1994). Let $t \in (Th, 1 - Th)$, let kernel K be continuous in its support $[-T, T]$, and integrate to 1:

$$\int_{-T}^{T} K(t)dt \;=\; 1, \quad T > 0. \tag{12}$$

To introduce optimization problems, it is more convenient to consider a continuous analogue of model (1), namely "white noise" model

$$dY(t) \;=\; f(t)dt \;+\; \sigma/(\sqrt{n})dW(t), \quad t \in [0,1]\ , \tag{13}$$

where $W(t)$ is a standard Wiener process. In that case a kernel estimator is defined as

$$\tilde{f}_n(t) \;=\; (1/h) \int K((t-u)/h)dY(u)\ . \tag{14}$$

(13) and (14) entail a standard decomposition of $[\tilde{f}_n(t) - f(t)]$ into the sum of a bias s_1 and a stochastic term s_2, such that

$$E_f[\tilde{f}_n(t) - f(t)]^2 \;=\; s_1^2(t; f; h) \;+\; var(s_2(t; n, h))\ , \tag{15}$$

where

$$s_1(t; f; h) \;=\; \int_{-T}^{T} K(u)[f(t - uh) - f(t)]du\ , \tag{16}$$

$$var(s_2(t; n, h)) \;=\; \frac{\sigma^2}{hn} V(K)\ , \quad V(K) \;=\; \int_{-T}^{T} K^2(t)dt. \tag{17}$$

First, recall the standard approach for treating (15). Let β be an integer, $\beta = m+1$. If $f \in \tilde{\Sigma}(\beta, L) \;\cap\; \{f^{(\beta)}$ is continuous $\}$, where

$$\tilde{\Sigma}(\beta, L) \;=\; \{f : \;\; |f^{(\beta)}(t)| \le L\}\ ,$$

and if K satisfies orthogonality conditions

$$\int_{-T}^{T} K(t)t^j dt = 0,\ j = 1, \ldots, m; \quad \int_{-T}^{T} K(t)t^\beta dt \ \neq\ 0\ , \tag{18}$$

then the Taylor expansion and (18) imply

$$\sup_{f\in\Sigma(\beta,L)} |s_1(t;f;h)| \ \leq\ \frac{Lh^\beta}{\beta!}\tilde{B}(K) \ \text{ for all } t \in [Th, 1-Th]\ , \tag{19}$$

with $\tilde{B}(K) = |\int_{-T}^{T} K(t)t^\beta dt\ |$. If we switch for a while to the minimax risk at a fixed point $t \in (0,1)$, i.e. if we substitute the uniform norm (5) in (4) by a distance $d_t(\tilde{f} - f) = |\tilde{f}(t) - f(t)|$, then the balance equation between the squared bias and the variance leads to the following rate of the bandwidth in this setting:

$$\tilde{h}_n = \tilde{a}\tilde{\psi}_{n,\beta}^{1/\beta}\ , \quad \tilde{\psi}_{n,\beta} = n^{-\beta/(2\beta+1)}\ , \tag{20}$$

with some positive $\tilde{a}$, so putting $\tilde{h}_n$ into (15) and optimizing it via $\tilde{a}$ give

$$\lim_{n\to\infty} E_f[\tilde{f}_n(t) - f(t)]^2/\tilde{\psi}_{n,\beta}^2 \ =\ C_1(\beta, L, \sigma^2)\tilde{D}_\beta^{2/(2\beta+1)}(K)\ , \tag{21}$$

where $C_1(\beta, L, \sigma^2)$ is a positive constant explicitly depending on β, L, σ^2,

$$\tilde{D}_\beta(K) \ =\ \tilde{B}(K)V^\beta(K)\ ,$$

see Gasser and Müller (1979). Thus, to minimize the right-hand side of (21), one needs to solve the extremal problem

$$\tilde{D}_\beta(K) \ \to\ \min\ , \quad \text{subject to (18),(12).} \tag{22}$$

This problem is thoroughly studied by Legostaeva and Shiryaev (1971), Gasser and Müller (1979), Gasser et al. (1985). Note that though $\tilde{\psi}_{n,\beta}$ is indeed the optimal rate of convergence for estimation at a fixed point (see Ibragimov and Khasminskii (1981)), the exact constants for this problem are not known so far, and non-linear estimators perform better than linear ones in this set-up, see Sacks and Strawderman (1982).

Now if we return to the sup-norm minimax risk, it's worth noting that an extra logarithmic term appears in the rate to suppress the large deviations probabilities (see Stone (1982), Härdle (1990), Korostelev (1993)), so that $h_{n,\beta} = a\psi_{n,\beta}^{1/\beta}$, with $\psi_{n,\beta}$ defined by (7). But the optimization problems for multiplier a and for kernel K remain the same as above for the minimax risk at a fixed point. However, the solution of (22) do not generate kernels leading to the exact constants for the sup-norm minimax risk.

To obtain the exact constants, a more accurate estimate for the bias term (16) should be used. This is done by exploiting renormalization ideas (Donoho and Low (1992), Donoho (1994)) for any $\beta > 0$: if $f \in \Sigma(\beta, 1)$, then $f_1(t) = Lh^\beta f(t/h) \in \Sigma(\beta, L)$, and

$$\sup_{f\in\Sigma(\beta,L)} s_1(t; f; h) \;=\; Lh^\beta B_\beta(K) \;,\; \text{for all } t \in [Th, 1-Th] \;, \tag{23}$$

with

$$B_\beta(K) \;=\; \sup_{f\in\Sigma(\beta,1)} \int_{-T}^{T} K(u)[f(u) - f(0)]du.$$

Note that bandwidth h enters the right-hand sides of (19) and (23) in the same way, but with different multipliers. Therefore, the renormalization leads:

(a) to the same rates of convergence (7) and (20) for the sup-norm minimax risk and minimax risk at a fixed point, respectively,

(b) to the analogue of (21) where functional $\tilde{D}_\beta(K)$ is substituted by $D_\beta(K)$,

$$D_\beta(K) \;=\; B_\beta(K)V^\beta(K).$$

Thus, to optimize kernels, the following extremal problem should be solved:

$$D_\beta(K) \to \inf, \;\; \text{subject to (12)} \;. \tag{24}$$

Finally, it turns out that (24) is equivalent to (9) and that the solution of (9) generates optimal kernels and the exact constants for the sup-norm minimax risk (Donoho (1994)).

3. Main results.

We begin the section with the technical lemmas used for establishing the result of Theorem 1. Introduce the following notation.

Let $[-T_\beta, T_\beta]$ be the support of x_β, i.e. $x_\beta = 0$ for $|t| \geq T_\beta$. Further on we shall omit the limits of integration if it is over the interval $[-T_\beta, T_\beta]$. Let t_β be the support of g_β. Put

$$I_{1\beta} = \int x_\beta(t)dt, \quad I_{2\beta} = \int x_\beta^2(t)dt, \quad \tilde{I}_{1\beta} = \int_{-t_\beta}^{t_\beta} g_\beta(t)dt,$$

$$Q(f) = \int x_\beta(t)(f(t) - f(0))dt, \quad Q_\beta = \sup_{f \in \Sigma(\beta,1)} Q(f) \ ,$$

$$K_\beta(t) = x_\beta(t)/I_{1\beta} \ , \quad \tilde{K}_\beta(t) = g_\beta(t)/\tilde{I}_{1\beta} \ .$$

Remark that (11) leads to

$$\tilde{K}_\beta(t) = b_1 K_\beta(b_1 t) \ . \tag{25}$$

The first two lemmas are a reformulation of the results by Donoho (1994), Sections 2 and 4; see also Lepski and Tsybakov (1996), Lemma 1.

Lemma 1. If

$$z(\beta, b) = \sup_{f \in \Sigma(\beta,1)} \int_{-\infty}^{\infty} x_\beta(bt)[f(t) - f(0)]dt,$$

then $z(\beta, b) \ = \ b^{-(\beta+1)} Q_\beta$ for any $b > 0$.

Proof. If $f \in \Sigma(\beta, 1)$, then $g(t) = b^{-\beta} f(bt) \in \Sigma(\beta, 1)$ as well, thus by changing variables for the integration,

$$z(\beta, b) \ = \ b^{-\beta} \sup_{f \in \Sigma(\beta,1)} \int_{-\infty}^{\infty} x_\beta(bt)[f(bt) - f(0)]dt \ = \ b^{-(\beta+1)} Q_\beta.$$

Lemma 2. $Q_\beta = Q(1 - x_\beta) = I_{1\beta} - I_{2\beta}, \ Q_\beta \geq 0.$

Proof. To begin with, remark that $Q_\beta \geq Q(0) = 0$. Next, note that $Q(f) = Q(f + C)$ for any constant C. If $f \in \Sigma(\beta, 1)$, then $g = -f \in \Sigma(\beta, 1)$ as well, thus

$$Q_\beta = \sup_{f \in \Sigma(\beta,1)} Q(-f) \ = \ \sup_{f \in \Sigma(\beta,1), \ f(0)=1} \int x_\beta(t)[-f(t)]dt \ + \ \int x_\beta(t)dt \ =$$

$$= \ I_{1\beta} \ - \ \inf_{f \in \Sigma(\beta,1), \ f(0)=1} \int x_\beta(t) f(t) dt \ .$$

Hence, to prove the lemma, it is sufficient to establish that

$$\inf_{f \in \Sigma(\beta,1), \ f(0)=1} \int x_\beta(t) f(t) dt \ = \ I_{2\beta}. \tag{26}$$

Let g be an arbitrary function from $\Sigma(\beta, 1)$ with g(0)=1. Similar to Lemma 2 from Leonov (1995), it is proved that there exists a finite constant $T, T \geq T_\beta$, and $\tilde{g} \in \Sigma(\beta, 1)$, such that

$$\tilde{g}(t) = g(t) \text{ for } |t| \leq T_\beta; \quad \tilde{g}(t) = 0 \text{ for } |t| \geq T,$$

and

$$\|\tilde{g} - x_\beta\|^2 \;=\; \int [g(t) - x_\beta(t)]^2 dt \;+\; \int_{T_\beta \le |t| \le T} \tilde{g}^2(t) dt \;<\; \infty. \tag{27}$$

Let $g_\lambda(t) = \lambda \tilde{g}(t) + (1 - \lambda) x_\beta(t)$ for $0 \le \lambda \le 1$. Note that $g_\lambda \in \Sigma(\beta, 1)$ and $g_\lambda(0) = 1$, thus, $\|g_\lambda\|^2 \ge \|x_\beta\|^2$, or

$$\lambda[\lambda \|\tilde{g} - x_\beta\|^2 + 2(\int x_\beta(t) g(t) dt \;-\; \|x_\beta\|^2)] \;\ge\; 0 \,.$$

Due to (27) the last inequality is valid for all $\lambda, 0 \le \lambda \le 1$, if and only if $\int x_\beta(t) g(t) dt \ge \|x_\beta\|^2$. This proves the lemma since the infimum in (26) is attained at $f = x_\beta$.□

Corollary 1. $B_\beta(K_\beta) = 1 - I_{2\beta}/I_{1\beta}$.

Corollary 2. $B_\beta(\tilde{K}_\beta) = g_\beta(0) - 1/\tilde{I}_{1\beta}$.

Proof follows immediately from Lemmas 1,2 and from (10), (11).□

Let K satisfy (12) and be continuous in $[-T, T]$. Introduce an equivalence class of kernels

$$\mathcal{F}(K) \;=\; \{\tilde{K} : \;\; \tilde{K}(t) = K_b(t) = bK(bt), \;\; b > 0\}.$$

Lemma 3. Functional $D_\beta(\tilde{K})$ is invariant for $\tilde{K} \in \mathcal{F}(K)$.

Proof is straightforward and is similar to Lemma 1 from above and to Lemma 1 from Gasser and Müller (1979) which establishes the invariance of functional $\tilde{D}_\beta(\tilde{K})$ for $\tilde{K} \in \mathcal{F}(K)$.□

Let $\mathcal{F}_\beta \;=\; \{f : \; f$ is a solution of (24)$\}$.

Lemma 4. $\mathcal{F}_\beta = \mathcal{F}(K_\beta)$.

Proof. Let K be an arbitrary function such that

$$\|K\| = \|K_\beta\|. \tag{28}$$

To prove the lemma, due to Lemma 3 it is sufficient to validate that K_β is a unique solution of the following problem

$$B_\beta(K) \to \inf, \quad \text{subject to (12), (28).}$$

Let K be a continuous function satisfying (12),(28), where without loss of generality $T \ge T_\beta$. If $r(t) = K(t) - K_\beta(t)$, then (12) and (28) imply

$$\int_{-T}^{T} r(t) dt = 0, \quad \int_{-T}^{T} r^2(t) dt = -2 \int K_\beta(t) r(t) dt. \tag{29}$$

Let $g(t) = 1 - x_\beta(t)$. Then Lemma 2 entails:

$$B_\beta(K) \;\geq\; \int_{-T}^{T} K(t)[g(t) - g(0)]dt \;=\; \int_{-T}^{T} [K_\beta(t) + r(t)][g(t) - g(0)]dt \;=$$

$$= \; B_\beta(K_\beta) - \int x_\beta(t) r(t) dt \;=\; B_\beta(K_\beta) + \int_{-T}^{T} r^2(t)dt/2$$

due to (29), thus $B_\beta(K) \geq B_\beta(K_\beta)$, and this inequality is a strict one if $K \neq K_\beta$ which proves the lemma. □

Theorem 1. $I_{2\beta}/I_{1\beta} = 2\beta/(2\beta+1)$ for any $\beta > 0$.

Proof. We shall prove the theorem via the analysis of the minimax risk based on large deviations probabilities (see Korostelev (1996), Korostelev and Leonov (1996)). Let $\hat{f}_n(t)$ be a kernel estimator (8) with continuous kernel K satisfying (12) and with a bandwidth h, such that $t \in (Th, 1 - Th)$ and for some $b > 0$

$$Lh^\beta = bc, \quad c \text{ is a small positive.} \tag{30}$$

Let f be uniformly bounded,

$$f \in \Sigma(\beta, L; A) = \Sigma(\beta, L) \cap \{|f(t)| \leq A\},$$

$A > 0$ is a sufficiently large constant. Then $\hat{f}_n(t) - f(t) = S_1(t; f; n, h) + S_2(t; n, h)$, where

$$S_1(t; f; n, h) = s_1(t; f, h) \; + \; o_1(1) \;, \tag{31}$$

$$var(S_2(t; n, h)) \;=\; \frac{\sigma^2}{hn} V(K) \; + \; o_2(1) \;, \tag{32}$$

$s_1(t; f, h)$ and $V(K)$ are defined by (16) and (17), respectively, and $o_i(1) \to 0$ uniformly in $f \in \Sigma(\beta, L; A)$ as $n \to \infty$, $i = 1, 2$. Then, following the proof of Theorem 2 in Korostelev (1996), we get

$$P_f^{(n)}(|\hat{f}_n(t) - f(t)| \geq c) \;\leq\; P_f^{(n)}(|S_2(t; n, h)| \geq c - \sup_{f \in \Sigma(\beta, L; A)} S_1(t; f; n, h)) \;,$$

here $P_f^{(n)}$ is the probability of observations y_{in} for a fixed regression f. (17), (23), (30)-(32) entail

$$\sup_{f \in \Sigma(\beta, L; A)} \limsup_{n \to \infty} \frac{1}{n} \log P_f^{(n)}(|\hat{f}_n(t) - f(t)| \geq c) \;\leq\; - \frac{c^{2+1/\beta}}{2\sigma^2 L^{1/\beta}} M(K, b) \;,$$

where $M(K,b) = [1 - bB_\beta(K)]^2 b^{1/\beta}/V(K)$. Maximizing $M(K,b)$ via b leads to $b_K = [B_\beta(K)(1+2\beta)]^{-1}$, and

$$M(K,b_K) = (2\beta)^2(2\beta+1)^{-(2+1/\beta)}[B_\beta(K)V^\beta(K)]^{-1/\beta} \ . \tag{33}$$

As proved by Korostelev (1996), kernel $\tilde{K}_\beta$ with a specific choice of $b = b_{\tilde{K}_\beta}$ maximizes (33),

$$M(\tilde{K}_\beta, b_{\tilde{K}_\beta}) \ = \ [g_\beta(0)]^{-(2+1/\beta)} \ ,$$

which due to (11), (25), (33), and Lemma 4, is equivalent to

$$(2\beta)^2(2\beta+1)^{-(2+1/\beta)}[B_\beta(K_\beta)V^\beta(K_\beta)]^{-1/\beta} \ = \ I_{2\beta} \ . \tag{34}$$

Put $z = I_{2\beta}/I_{1\beta}$. Lemma 2 and Corollary 1 imply:

$$0 < z \leq 1, \quad B_\beta(K_\beta) = 1 - z, \quad V(K_\beta) = z^2 I_{2\beta}^{-1},$$

and (34) entails:

$$(2\beta)^2(2\beta+1)^{-(2+1/\beta)}(1-z)^{-1/\beta}z^{-2}I_{2\beta} \ = \ I_{2\beta}.$$

Taking β-power of both sides leads to

$$r^r/(r+1)^{r+1} \ = \ z^r - z^{r+1}, \text{ where } r = 2\beta \ . \tag{35}$$

A function $v(z) = z^r - z^{r+1}$ in the interval $[0,1]$ has a unique maximum at a point $z_r = r/(r+1)$, moreover $v(z_r) = r^r/(r+1)^{r+1}$, which together with (35) means that

$$I_{2\beta}/I_{1\beta} \ = \ r/(r+1) \ = \ 2\beta/(2\beta+1) \ .\square$$

Corollary 3. $g_\beta(0)\int g_\beta(t)dt = (2\beta+1)/(2\beta)$.
Proof follows directly from Theorem 1 and (10).

Remark 1. For $0 < \beta \leq 1$ and $\beta = 2$ the result of Theorem 1 can be verified directly via the explicit solutions of (9).
Remark 2. Theorem 1 and Corollary 1 provide an explicit formula for the bias term of the optimal kernel estimator,

$$B_\beta(K_\beta) = 1/(2\beta+1) \ \text{ for any } \beta > 0.$$

4. Some examples for class $\Sigma(2,1)$.

The solution of (9) for $\beta = 2$ (Fuller (1960)) remained unknown to the statistical community until recently. Therefore, it seems interesting to compare the performance of various kernels with respect to values of functional $D_2(K)$. To accomplish it, we prove the following lemma.

Lemma 5. $B_2(K) = \int_0^\infty K(t)t^2 dt$ for any symmetric non-negative kernel K.

Proof. Similar to the proof of Lemma 2, note that for any $\beta > 0$

$$B_\beta(K) = \sup_{f \in \Sigma(\beta,1),\ f(0)=0} \int_{-\infty}^{\infty} K(t)f(t)dt \ . \tag{36}$$

Next, if $f(0) = 0$, then the definition of $\Sigma(2,1)$ entails

$$f(t) = tf'(0) + F_f(t), \quad |F_f(t)| \le t^2/2 \ , \tag{37}$$

thus,

$$\int_{-\infty}^{\infty} K(t)f(t)dt = \int_0^\infty K(t)[f(t) + f(-t)]dt = \int_0^\infty K(t)[F_f(t) + F_f(-t)]dt \ .$$

The lemma follows from (36) and (37) since for $\beta = 2$ the supremum in (36) is attained at $f(t) = t^2/2$.□

Table 1 presents values of B_2, V, D_2 for several popular kernels. Theorem 1 is applied for computing B_2 for the optimal kernel K_2 while Lemma 5 is used for all other non-negative kernels. It is worthy to note that

(K1) x_2 is a symmetric quadratic spline having an infinite number of knots in the finite interval. Local extrema of x_2 form a geometric series with parameter $-q$, $q = (3 + \sqrt{33} - \sqrt{26 + 6\sqrt{33}})^2/16 \approx 0.0586$, so that

$$x_2(t) = \begin{cases} 1 - t^2/2, & \text{if } |t| \le t_1;\ \ t_1 = \sqrt{1+q} \approx 1.03;, \\ -q + (|t| - t_2)^2/2, & \text{if } t_1 \le |t| \le t_3;\ t_2 = 2t_1,\ t_3 = \sqrt{1+q}(2+\sqrt{q}) \approx 2.31, \\ q^2 - (|t| - t_4)^2/2, & \text{if } t_3 \le |t| \le t_4;\ \ t_4 = 2\sqrt{1+q}(1+\sqrt{q}) \approx 2.56, \end{cases}$$

and for $t > t_4, |x_2(t)| < q^2 \approx 0.0034$. The support of x_2 is $[-T_2, T_2]$, $T_2 = 2\sqrt{1+q}/(1-\sqrt{q}) \approx 2.715$, and

$$\|x_2\|^2 = 2(23q^2 - 14q + 23)\sqrt{1+q}/(30(1 - q^{5/2})) \approx 1.528,$$

(see Fuller (1960), or Leonov (1995)).

(K2) $x_{opt}(t)$, for $t > 0$, is the solution of time-optimal problem

$$T \to \min, \quad |x^{(2)}(t)| \le 1, \; x(0) = 1, x'(0) = 0; \quad x(T) = x'(T) = 0 \; :$$

$$x_{opt}(t) = \begin{cases} 1 - t^2/2, & \text{if } 0 \le |t| \le 1, \\ (|t| - 2)^2/2, & \text{if } 1 \le |t| \le 2, \\ 0, & \text{if } |t| > 2. \end{cases}$$

x_{opt} is "nearly" optimal for (9) since $\|x_2\|^2/\|x_{opt}\|^2 \approx 0.9965$ (see Fuller (1960)).

(K4) Triangular kernel is the solution of (9) for $\beta = 1$.

Table 1. Comparison of kernels for $\beta = 2$

#	Kernels	$B_2(K)$	$V(K)$	$D_2 = B_2 V^2$	D_*/D_2
1.	Optimal $K_2(t) = x_2(t)/\int x_2(s)ds$	0.2	0.4188	$D_* \approx 0.0351$	1
2.	Time-optimal $K_o(t) = x_{opt}/\int_{-2}^{2} x_{opt}(s)ds$	0.25	23/60	0.0367	0.955
3.	Epanechnikov $K_e(t) = 0.75(1-t^2)_+$	0.1	0.6	0.0360	0.974
4.	Triangular $K_1(t) = (1-|t|)_+$	1/12	2/3	$1/27 \approx 0.0374$	0.938
5.	Uniform $K_u(t) = 0.5\chi\{t \in [-1,1]\}$	1/6	0.5	$1/24 \approx 0.0417$	0.842
6.	Gaussian $K_g(t) = \exp(-t^2/2)/\sqrt{2\pi}$	0.5	$1/(2\sqrt{\pi})$	$\frac{1}{8\pi} \approx 0.0398$	0.881

The above results show that kernels (K1)-(K6) do not differ greatly with respect to values of $D_2(K)$. However, such results are not surprising for kernel comparison, see e.g. Epanechnikov (1969), Sacks and Ylvisaker (1981), Silverman (1986).
If kernels have compact support, then Lemma 3 allows to compare them visually in the interval $[-1, 1]$ (for a discussion on other methods of comparison see Marron and Nolan (1989)): in Fig. 1 kernels (K1) and (K2) are reduced to $[-1, 1]$, so that Fuller

kernel K_F and time-optimal kernel K_{t-opt} belong to corresponding equivalence classes and are defined as:

$$K_F(t) = T_2 K_2(T_2 t) = T_2 x_2(T_2 t) / \int x_2(s) ds; \quad K_F(0) \approx 1.42,$$

$$K_{t-opt}(t) = 2K_o(2t) = x_{opt}(2t).$$

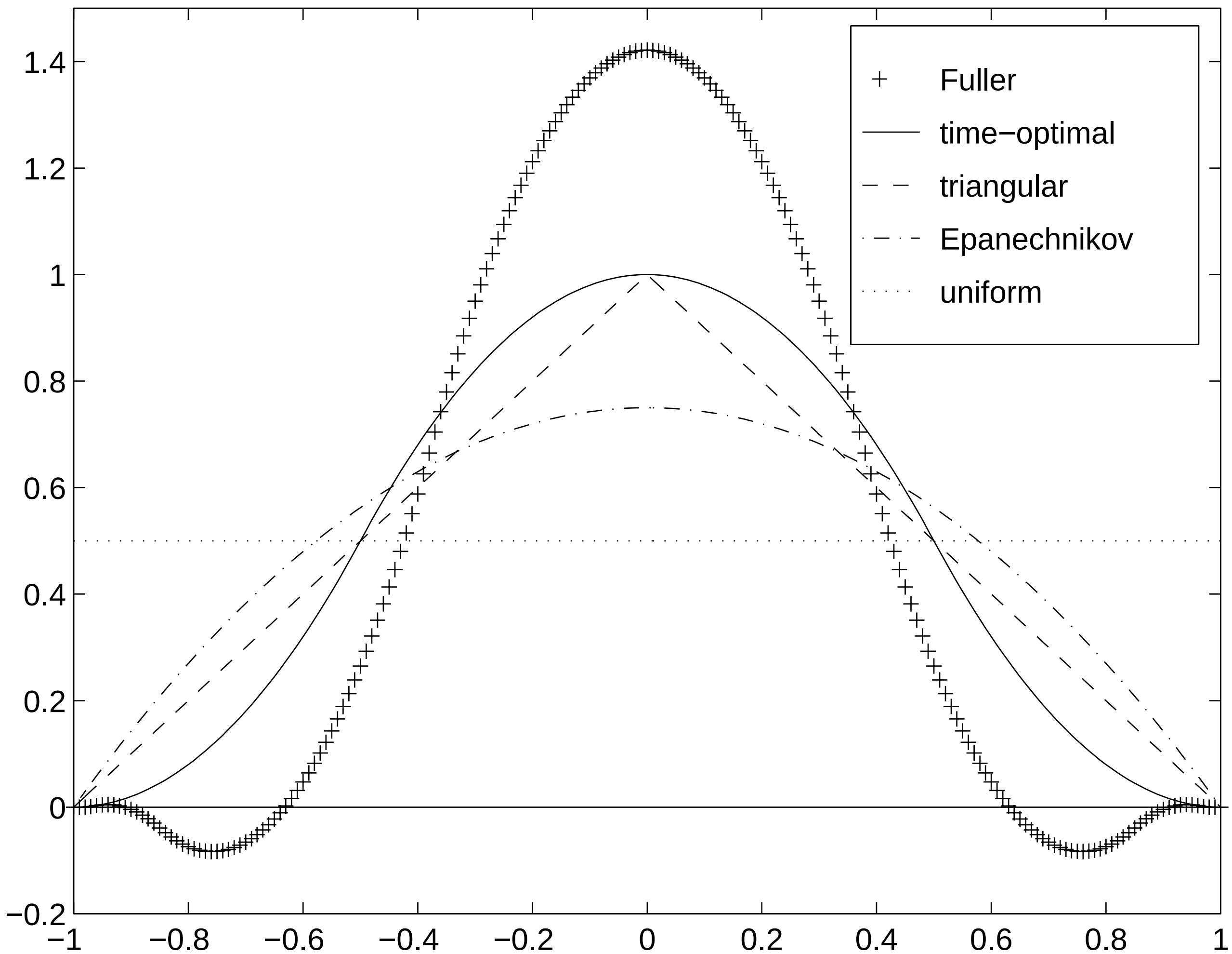

Figure 1: Kernels in [-1,1]

References

Donoho, D.L. and M.G.Low (1992), Renormalization exponents and optimal pointwise rates of convergence, *Ann. Statist.*, **20**, 2, 944-970.

Donoho, D.L. (1994), Asymptotic minimax risk for sup-norm loss: Solution via optimal recovery, *Probab. Theory Rel. Fields*, **99**, 145-170.

Epanechnikov, V.A. (1969), Nonparametric estimation of a multivariate probability density, *Theor. Probab. Appl.*, **14**, 153-158.

Fuller, A.T. (1960), Relay control systems optimized for various performance criteria, In: *Proceedings of I-st International IFAC Congress, Moscow*, v.2, 584-607 (in Russian). - Translated in: J.F. Coales et al., eds., *Automation and Remote Control*, v.1, Butterworths, London, 510-519.

Gasser, T. and H.-G. Müller (1979), Kernel estimation of regression functions. In: T. Gasser and M. Rosenblatt, eds. *Smoothing Techniques for Curve Estimation*, Lecture Notes in Mathematics, v.757, Springer, 23-68.

Gasser, T., H.-G. Müller, and V. Mammitzsch (1985), Kernels for nonparametric curve estimation of regression functions, *J. Roy. Stat. Soc. B*, **47**, 2, 238-252.

Härdle, W. (1990), *Applied Nonparametric Regression*, Cambridge Univ. Press, Cambridge.

Ibragimov, I.A. and R.Z. Khasminskii (1981), *Statistical Estimation: Asymptotic Theory* , Springer, New York.

Korostelev, A.P. (1993), Exact asymptotically minimax estimator for nonparametric regression in the uniform norm, *Theory Probab. Appl.*, **38**, 4, 875-882.

Korostelev, A.P. (1996), A minimaxity criterion in nonparametric regression based on large-deviations probabilities, *Ann. Statist.*, **24**, 3, 1075-1083.

Korostelev, A.P. and S.L. Leonov (1996), Minimax Bahadur efficiency for small confidence levels, *Probl. Inform. Transm.*, **32**, 4, 303-313.

Korostelev, A.P. and M. Nussbaum (1995), Density estimation in the uniform norm and white noise approximation, Weierstrass Institute for Applied Analysis and Stochastics, Preprint 153, Berlin.

Legostaeva, I.L. and A.N. Shiryaev (1971), Minimax weights in trend detection problem of a random process, *Theory Probab. Appl.*, **16**, 2, 344-349.

Leonov, S.L. (1995), On the solution of an optimal recovery problem and its applications in nonparametric statistics, Weierstrass Institute for Applied Analysis and Stochastics, Preprint 202, Berlin (to appear in *Mathematical Methods of Statistics*).

Lepski, O.V. (1992), On problems of adaptive estimation in white Gaussian noise, In: R.Z. Khasminskii, ed., *Topics in Nonparametric Estimation. Advances in Soviet Math.*, v.12, AMS, Providence, R.I., 87-106.

Lepski, O.V. and V.G. Spokoiny (1995), Optimal pointwise adaptive methods in nonparametric estimation, Humboldt University, Disc. Paper 22, SFB 373, Berlin.

Lepski, O.V. and A.B. Tsybakov (1996), Asymtotically exact nonparametric hypothesis testing in sup-norm and at a fixed point, Humboldt University, Disc.Paper 91, SFB 373, Berlin.

Marron, J.S. and D.Nolan (1989), Canonical kernels for density estimation, *Statist. Probab. Lett.*, **7**, 195-199.

Sacks, J. and W. Strawderman (1982), Improvements on linear minimax estimates, In: S. S. Gupta and J.O. Berger, eds., *Statistical Decision Theory and Related Topics III*, v.2, Academic Press, New York, 287-304.

Sacks, J. and D. Ylvisaker (1981), Asymptotically optimal kernels for density estimation at a point, *Ann. Statist.*, **9**,2, 334-346.

Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London.

Stone, C.J. (1982), Optimal global rates of convergence for nonparametric regression, *Ann. Statist.*, **10**, 4, 1040-1053.

Zhao, L.H. (1997), Minimax linear estimation in a white noise problem, *Ann. Statist.*, **25**, 2, 745-755.