The Annals of Statistics
1990, Vol. 18, No. 3, 999–1010

ON DENSITY ESTIMATION IN THE VIEW OF KOLMOGOROV'S IDEAS IN APPROXIMATION THEORY

BY RAFAEL HASMINSKII[1] AND ILDAR IBRAGIMOV

Institute for Problems in Information Transmission and Leningrad Branch of the Steklov Mathematical Institute

The paper deals with upper and lower bounds for the quality of (probability) density estimation. Connections are established between these problems and the theory of approximation of functions. Particularly, it is demonstrated how some of Kolmogorov's concepts work.

1. Introduction. The aim of this paper is to present some ideas which can be used in nonparametric estimation problems. These ideas are connected with Kolmogorov's contribution to the theory of approximation of functions. It should be stressed that these ideas can be used for a wide class of statistical models, but we consider here only one example of them.

We assume that a statistician observes i.i.d. random vectors $X_1, X_2, \dots, X_n$ taking values in R^k and having density f with respect to the Lebesgue measure on R^k. The problem is to find an estimator, based on these data, for the density f, unknown to the statistician. We denote by $f_n(\cdot) = f_n(\cdot, X_1, \dots, X_n)$ any estimator for f, i.e., any real-valued Borel measurable function of all arguments. It is not assumed that $f_n(\cdot)$ is necessarily a density for the fixed data; it is not even necessarily nonnegative. The quality of estimation is measured by the loss function $\|f_n - f\|_p^r$, where $\|\cdot\|_p$ is the L_p norm on R^k and r is a positive number.

We assume that the unknown density function f is in some known set Σ, and the problem is considered as nonparametric if it is impossible to embed Σ in finite-dimensional space. A general scheme for nonparametric estimation is the following. The statistician chooses (in advance) some subset Φ of functions such that (a) the subset Φ is a sufficiently good approximation for Σ and (b) it is easy to find sufficiently good estimates for any $\phi \in \Phi$. In this way, the estimation of f is reduced to the estimation of a suitable element of Φ. This approach is popular not only for statistical problems, but also for some applied problems in physics [see Babenko (1979, 1985)]. So the estimation error consists of two parts: the error from approximating the element $f \in \Sigma$ by $\phi \in \Phi$ (this is bias in typical situations) and the error from estimating an element from Φ. Φ can be a finite set (then estimation is reduced to hypotheses testing), a finite-dimensional set (it leads to projection estimates) and so on. In general, Φ will be an element of a collection $\mathscr{F}$ of function classes. If Φ is chosen optimally, the first part of the error depends on the $\mathscr{F}$-diameter of Σ

Received September 1989; revised March 1990.

[1]Part of this work was completed while the author was visiting CWI, Amsterdam.

AMS 1980 *subject classification*. 62G05.

Key words and phrases. Nonparametric density estimation, best approximation in L_p norms, Fano's inequality.

[see Babenko (1979, 1985)], defined as

$$\inf_{\Phi \in \mathscr{F}} \sup_{f \in \Sigma} \inf_{\phi \in \Phi} \| f - \phi \|.$$

Here $\| \cdot \|$ is a certain norm (if $\mathscr{F}$ consists of N-dimensional linear manifolds, we have Kolmogorov's diameters [Kolmogorov (1936)]; if $\mathscr{F}$ consists of finite sets with a fixed number N of elements, we have something closely related to ε entropy [Kolmogorov and Tikhomirov (1959)]. Typically, we choose N depending on the sample size n in such a way to balance this part of the error with the other part.

We consider here the estimation problem for densities defined on R^k. It is well known in approximation theory that entire functions of finite exponential degree are convenient for approximation of such functions. Therefore we choose such function classes as the elements of $\mathscr{F}$.

Our main interest is the behavior of the value

$$\Delta_n(p, r) = \inf_{f_n} \sup_{f \in \Sigma} E_f \| f_n - f \|_p^r. \tag{1.1}$$

Here and further the infimum over f_n is taken over all estimators, and Σ is some subset of L_p, $1 \le p \le \infty$.

2. Upper bounds. An important part of the results presented here was published in Ibragimov and Hasminskii (1980, 1981) (henceforth we use the abbreviation IH for our names), but unfortunately some of these results are not sufficiently known in the West. We would like to mention, for example, that some of the results were reproved in Devroye and Györfi (1985) in a weakened form, and the problem they formulated in their book (page 133) as an open one, is solved by Theorem 4 in IH (1981).

Any estimation procedure gives some upper bound for the quality of estimation. The literature on this theme is very rich, and we do not try to systematize it. It seems that Centsov (1962, 1972) proposed the following elegant reasoning. Let $f \in L_2(R^k) = L_2$ and let $\Phi_1, \Phi_2, \dots$ be a sequence of subspaces of L_2, with $\dim \Phi_N = N$. Let $\phi_1^{(N)}, \dots, \phi_N^{(N)}$ be an orthobasis of Φ_N. The equality

$$f = \sum_{i=1}^{N} C_i^{(N)} \phi_i^{(N)} + R_N(f) \tag{2.1}$$

($C_i^{(N)}$ are Fourier coefficients) generates the projection estimator

$$f_{n,N}(x) = \sum_{i=1}^{N} C_{in}^{(N)} \phi_i^{(N)}(x). \tag{2.2}$$

Here, $C_{in}^{(N)}$ is defined by

$$C_{in}^{(N)} = \frac{1}{n} \sum_{j=1}^{n} \phi_i^{(N)}(X_j) = \int \phi_i^{(N)}(x) F_n(dx), \tag{2.3}$$

where F_n is the sample distribution function.

We have that $E_f C_{in}^{(N)} = C_i^{(N)}$ and moreover the inequality

$$E_f \left| C_{in}^{(N)} - C_i^{(N)} \right|^2 \le A/n \tag{2.4}$$

is often true. But then, the L_2 risk of the estimator $f_{n,N}$ for an optimal choice of Φ_N (from the class $\mathscr{F}$ of N-dimensional subspaces) does not exceed the value

$$\begin{aligned} E_f \| f_{n,N} - f \|_2^2 &\le \frac{AN}{n} + \| R_N(f) \|_2^2 \\ &\le \frac{AN}{n} + d_N^2(\Sigma). \end{aligned} \tag{2.5}$$

Here A is a constant and $d_N(\Sigma)$ is the N-dimensional Kolmogorov diameter of Σ. It is possible to choose $N = N(n)$ in an optimal way such that the best order for the rate of convergence to zero of the risk is obtained.

Modifications of this simple idea allow the use of similar estimators for other loss functions. It is also important that it is possible to fix the sequence $\Phi_1, \Phi_2, \dots,$ because variability of this family is not suitable for applications.

The classical results of approximation theory yield the best order for the rate of convergence as $N \to \infty$, when approximating functions $f \in \Sigma \subset L_p$, $1 \le p \le \infty$, by elements from the standard families in $\mathscr{F}$. For example, let f have a known parallelepiped $\Pi_k \subset R^k$ as support and let its periodical extension have some smoothness β. The natural $\mathscr{F}$ for this case is families of trigonometric polynomials. Another example is functions f with support Π_k without periodicity condition. Then the elements of $\mathscr{F}$ can be chosen as families of splines. Finally, for functions f with support R^k, it is convenient to choose as the elements of $\mathscr{F}$ the families of functions having analytical extension as entire functions of some exponential degree. For a suitably chosen sequence $\Phi_1, \Phi_2, \dots,$ the rates of convergence to zero as $N \to \infty$ of the best approximations by functions from Φ_N for these examples, have the order $d_N(\Sigma)$ in L_p for any $p \in [1, \infty]$ and for a wide class of sets Σ. We consider here for brevity only the last example. [This class is apparently the most natural for density estimation, but for other problems, for instance, regression estimation, two other families $\mathscr{F}$ are more natural; see IH (1980b and Nussbaum (1985)].

Let $\mathbb{C}^k$ be the k-dimensional complex space. Recall that an entire function $g(z) = g(z_1, \dots, z_k)$ is of the exponential type $\nu = (\nu_1, \dots, \nu_k)$ if for any $\varepsilon > 0$ the inequality

$$|g(z)| < A_\varepsilon \exp\left\{ \sum_{i=1}^{k} (\nu_i + \varepsilon)|z_i| \right\}$$

holds for all $z \in \mathbb{C}^k$ and some constant A_ε. Denote by $\mathscr{M}_{\nu,p}(R^k)$ the set of such functions which also belong to $L_p(R^k)$ as functions of real variables, and let $\mathscr{E}_\nu^{(p)}(\phi)$ be the value of the best approximation in L_p norm of ϕ by the functions from $\mathscr{M}_{\nu,p}(R^k)$:

$$\mathscr{E}_\nu^{(p)}(\phi) = \inf_{g \in \mathscr{M}_{\nu,p}(R^k)} \| g - \phi \|_p.$$

Furthermore, denote by $V_\nu(x)$ the kernel of Vallée Poussin type [see Nikolskii (1969)] for $\nu = (\nu_1, \nu_2, \dots, \nu_k)$:

$$V_\nu(x) = \prod_{j=1}^{k} \frac{\cos \nu_j x_j - \cos 2\nu_j x_j}{\pi \nu_j x_j^2}.$$

It is shown in IH (1980, 1981) that the best order for the rate of convergence for estimates in L_p norms for a wide class of sets Σ is reached for kernel estimators with kernel $V_\nu(x)$, provided $\nu = \nu(n, \Sigma)$ is suitably chosen. The estimator has the form

$$\tilde{f}_{n,\nu}(x) = \frac{1}{n} \sum_{j=1}^{n} V_\nu(x - X_j) = \int_{R^k} V_\nu(x-y) F_n(dy). \tag{2.6}$$

The following theorem plays the most important role in obtaining upper bounds for risk of $\tilde{f}_{n,\nu}$ (A_i are some constants).

THEOREM 1. (i) *Let* $f \in L_p(R^k)$, $2 \le p < \infty$. *Then*

$$\begin{aligned} E_f \| \tilde{f}_{n,\nu}(\cdot) - f(\cdot) \|_p^r &\le A_1 (1 + \| f \|_p)^{(r+1)/2} \\ &\quad \times \left[\left(\mathscr{E}_\nu^{(p)}(f) \right)^r + \left(\frac{\nu_1 \cdots \nu_k}{n} \right)^{r/2} \right]. \end{aligned} \tag{2.7}$$

(ii) *Let* $f \in L_\infty(R^k)$. *Then*

$$\begin{aligned} E_f \| \tilde{f}_{n,\nu}(\cdot) - f(\cdot) \|_\infty^r &\le A_2 (1 + \| f \|_\infty)^{(r+1)/2} \\ &\quad \times \left[\left(\mathscr{E}_\nu^{(\infty)}(f) \right)^r + \left(\frac{\nu_1 \cdots \nu_k \ln(\nu_1 \cdots \nu_k)}{n} \right)^{r/2} \right]. \end{aligned} \tag{2.8}$$

(iii) *Let* $f \in L_p(R^k)$, $1 \le p < 2$. *Then*

$$\begin{aligned} E_f \| \tilde{f}_{n,\nu}(\cdot) - f(\cdot) \|_p^r &\le A_3 (1 + \| f \|_p)^{(r+1)/2} \\ &\quad \times \left[\left(\mathscr{E}_\nu^{(p)}(f) \right)^r + \left(\frac{\nu_1 \cdots \nu_k}{n} \right)^{(p-1)r/p} \right]. \end{aligned} \tag{2.9}$$

PROOF. The proof of the assertions (i) and (ii) is given in IH (1980, 1981), so we prove (iii) only and restrict ourselves to the case $r = p$ [the consideration of arbitrary r is also analogous to IH (1980, 1981)].

The equality

$$E_f \tilde{f}_{n,\nu}(x) = \int V_\nu(x-y) f(y)\, dy$$

is the immediate consequence of (2.6). So $E_f \tilde{f}_{n,\nu}(x) \in \mathscr{M}_{2\nu, p}(R^k)$ and

$$\left\| E_f \tilde{f}_{n,\nu} - f \right\|_p < C \mathscr{E}_\nu^{(p)}(f).$$

[See IH (1980) for details.]

Therefore

$$E_f\|\tilde{f}_{n,\nu} - f\|_p^p \le C\Big\{\big(\mathscr{E}_\nu^{(p)}(f)\big)^p + E\|\tilde{f}_{n,\nu} - E\tilde{f}_{n,\nu}\|_p^p\Big\}.$$

Furthermore, we have

$$E\|\tilde{f}_{n,\nu} - E\tilde{f}_{n,\nu}\|_p^p = \int_{R^k} E_f\left|\frac{1}{n}\sum_{j=1}^n \big(V_\nu(x - X_j) - E_f V_\nu(x - X_j)\big)\right|^p dx$$

$$= \int_{R^k} E_f\left|\frac{1}{n}\sum_{j=1}^n \xi_j(x)\right|^p dx.$$

It is known that for independent random variables $\xi_1, \dots, \xi_n$ with $E\xi_i = 0$ the inequality [see von Bahr and Essen (1965)]

$$E\left|\sum_{i=1}^n \xi_i\right|^p \le 2\sum_{i=1}^n E|\xi_i|^p$$

is true. Therefore

$$E_f\|\tilde{f}_{n,\nu} - E_f\tilde{f}_{n,\nu}\|_p^p \le c_1 n^{1-p}\int_{R^k} f(y)\,dy \int_{R^k}|V_\nu(x - y)|^p\,dx.$$

This relation and inequality

$$\int_{R^k}|V_\nu(x - y)|^p\,dx = \prod_{j=1}^k \int_{R^1}\left|\frac{\cos\nu_j x_j - \cos 2\nu_j x_j}{\pi\nu_j x_j^2}\right|^p dx_j \le c_2(\nu_1\nu_2 \cdots \nu_k)^{p-1}$$

give assertion (iii) for $r = p$. □

Theorem 1 and the known upper bounds for $\mathscr{E}_\nu^{(p)}(f)$ with $f \in \Sigma$ imply upper bounds for risks of the estimator $\tilde{f}_{n,\nu}$. It allows us to choose $\nu = \nu(n, \Sigma)$ optimally. Consider two examples.

EXAMPLE 1. Let

$$H_p^\beta L, \qquad \beta = (\beta_1, \dots, \beta_k), \qquad \beta_i = r_i + \alpha_i,\ 0 < \alpha_i \le 1,\ i = \overline{1, k},$$

be the set of functions having Sobolev's derivative with respect to x_i of order r_i and suppose

$$\left\|\Delta_{x_i}\frac{\partial^{r_i}\phi}{\partial x_i^{r_i}}\right\|_p \le L|\Delta x_i|^{\alpha_i}$$

(here $\Delta_{x_i} g$ is the partial increment of g over x_i).

It is well known [Nikolskii (1969)] that for all $p \in [1, \infty]$,

$$\sup_{f \in H_p^\beta L} \mathscr{E}_\nu^{(p)}(f) \le c\sum_{j=1}^k \nu_j^{-\beta_j}. \tag{2.10}$$

The substitution of this bound in the right part of relations (2.7)–(2.9) and optimization over $\nu_1, \dots, \nu_k$ imply the results about upper bounds which are written in the first and second lines of Table 1.

EXAMPLE 2. Let $A_p^\lambda L$ be the family of functions g in R^k which admit analytical extension in the set $|\text{Im } z_i| < \lambda$, $i = \overline{1, k}$, and suppose

$$\|g(\cdot + i\lambda)\|_p < L.$$

It is known that for this class, for $\nu_1 = \nu_2 = \cdots = \nu_k = \nu$ and for all $p \in [1, \infty]$, the relation

$$\sup_{f \in A_p^\lambda L} \mathscr{E}_\nu^{(p)}(f) \le L e^{-\lambda \nu}$$

holds.

In a similar way as in Example 1 we obtain results (concerning upper bounds) which are written in the third and fourth lines in Table 1.

3. Lower bounds. Establishing lower bounds, in the minimax sense of (1.1), is a more complicated problem. A very important step was made by Farrell (1972). He has obtained lower bounds for the estimation quality in a point by reducing this problem to discrimination between two close hypotheses. IH (1979) have proposed for the same problem another approach, which is based on reducing it to an estimation problem. IH (1979) also proposed (for another situation) an approach to obtaining lower bounds in L_p norms, $2 \le p \le \infty$. The method here is to reduce the problem to discrimination between an increasing number of hypotheses and to use an information theoretical approach to the latter. Independently, a different approach to establishing lower bounds in L_p norms, $p < \infty$, for the classes $H_p^\beta L$ and some other ones, have been proposed by Bretagnolle and Huber (1979).

Here we present briefly the main ideas of our approach. Its application to the density estimation problem was published in IH (1980, 1981). Let ρ be some metric in $\Sigma \subset L_p$. Let $l(x)$, $x > 0$, be a nondecreasing function with $l(0) = 0$. Assume that there are $N(\delta)$ densities $f_{i\delta} \in \Sigma$ such that $\rho(f_{i\delta}, f_{j\delta}) > \delta$. Then the inequalities

$$\begin{aligned}\sup_{f\in\Sigma} E_f l\left(\frac{\rho(f_n, f)}{\delta}\right) &\ge \sup_{i=\overline{1, N(\delta)}} E_{f_{i\delta}} l\left(\frac{\rho(f_n, f_{i\delta})}{\delta}\right) \\ &\ge \frac{l(1/2)}{N(\delta)} \sum_{i=1}^{N(\delta)} P_{f_{i\delta}}\left\{\rho(f_{i\delta}, f_n) > \frac{\delta}{2}\right\} = l\left(\frac{1}{2}\right) P_e\end{aligned} \tag{3.1}$$

are evident and the problem is reduced to obtaining the lower bound of the average probability error P_e for the discrimination problem of $N(\delta)$ equidistributed hypotheses, i.e., $P\{\eta = f_{i\delta}\} = N_\delta^{-1}$, $i = \overline{1, N_\delta}$, on base of the sample $X_1, \dots, X_n$. The value P_e can be estimated with help of Fano's lemma in information theory [see, for example, IH (1979)]. As result we have the bound

$$\begin{aligned}P_e &\ge 1 - (\ln N(\delta))^{-1} I(\eta; (X_1, \dots, X_n)) \\ &\ge 1 - (\ln N(\delta))^{-1} n I(\eta, X_1).\end{aligned} \tag{3.2}$$

The desired result will be obtained if we can find a sufficiently good upper bound for Shannon's information $I(\eta, X_1)$.

Let $f_{0\delta}$ be an arbitrary density in R^k. Then

$$\begin{aligned} I(\eta, X_i) &= E\left\{\ln \frac{dP_{X_1/\eta}}{dP_{X_1}}(X_1, \eta)\right\} \\ &= \frac{1}{N(\delta)} \sum_{i=1}^{N(\delta)} \int_{R^k} \ln\left(f_{i\delta}(x) \Big/ \frac{1}{N(\delta)} \sum_{j=1}^{N(\delta)} f_{j\delta}(x)\right) f_{i\delta}(x)\, dx \\ &\le \frac{1}{N(\delta)} \sum_{i=1}^{N(\delta)} \int_{R^k} f_{i\delta}(x) \ln \frac{f_{i\delta}(x)}{f_{0\delta}(x)}\, dx \\ &\le \frac{1}{N(\delta)} \sum_{i=1}^{N(\delta)} \int_{R^k} (f_{i\delta}(x) - f_{0\delta}(x)) \ln \frac{f_{i\delta}(x)}{f_{0\delta}(x)}\, dx \\ &\le \frac{1}{N(\delta)} \sum_{i=1}^{N(\delta)} \int_{R^k} \frac{(f_{i\delta}(x) - f_{0\delta}(x))^2}{f_{0\delta}(x)}\, dx \\ &\le \max_{i=\overline{1, N(\delta)}} \left\| \frac{f_{i\delta} - f_{0\delta}}{\sqrt{f_{0\delta}}} \right\|_2^2 . \end{aligned}$$

So

$$(3.3)\quad \sup_{f\in\Sigma} E_f l\left(\frac{\rho(f_n, f)}{\delta}\right) \ge l\left(\frac{1}{2}\right)\left(1 - \frac{n}{\ln N(\delta)} \max_{i=\overline{1,N(\delta)}} \left\| \frac{f_{i\delta} - f_{0\delta}}{\sqrt{f_{0\delta}}} \right\|_2^2\right).$$

Now, let us define the number $\delta(n, \Sigma)$ by the formula

$$(3.4)\quad \delta(n, \Sigma) = \sup\left\{\delta \colon (\ln N(\delta))^{-1} \max_{i=\overline{1,N(\delta)}} \left\| \frac{f_{i\delta} - f_{0\delta}}{\sqrt{f_{0\delta}}} \right\|_2^2 \le \frac{1}{2n}\right\}.$$

Relations (3.3) and (3.4) imply the inequality

$$(3.5)\quad \sup_{f\in\Sigma} E_f l(\rho(f_n, f)/\delta(n, \Sigma)) \ge \tfrac{1}{2} l(\tfrac{1}{2}).$$

So the following theorem, which is a more strong and precise version of Theorem 8 in IH (1980) is established.

THEOREM 3.1. *For any $\delta > 0$, let there be densities $f_{i\delta} \in \Sigma$, $i = \overline{1, N(\delta)}$, such that $\rho(f_{i\delta}, f_{j\delta}) \ge \delta$, $i, j = \overline{1, N(\delta)}$, $i \ne j$, and let the value $\delta(n, \Sigma)$ be defined according to* (3.4) *for an arbitrary family of densities $f_{0,\delta}$, $\delta > 0$. Then the lower bound* (3.5) *is true for any nondecreasing function $l(x)$.*

This theorem is connected conceptually with Kolmogorov's ε capacity $C_\varepsilon(\Sigma)$ of the set Σ in the metric ρ [see Kolmogorov and Tikhomirov (1959)].

TABLE 1

	$1 \le p < 2$	$2 \le p < \infty$	$p = \infty$
$\Sigma = H_p^\beta L \cap \{\|f\|_p \le M\}$			
$\Delta_n(p,r)$	$\asymp n^{-\beta r/(q\beta+1)}$	$\asymp n^{-\beta r/(2\beta+1)}$	$\asymp \left(\frac{n}{\ln n}\right)^{-\beta r/(2\beta+1)}$
ν_i	$n^{\beta/(\beta_i(q\beta+1))}$	$n^{\beta/(\beta_i(2\beta+1))}$	$\left(\frac{n}{\ln n}\right)^{\beta/(\beta_i(2\beta+1))}$
$\Sigma = A_p^\lambda L$			
$\Delta(p,r)$	$\asymp \left(\frac{(\ln n)^k}{n}\right)^{r/q}$	$\asymp \left(\frac{(\ln n)^k}{n}\right)^{r/2}$	$\asymp \left(\frac{(\ln n)^k \ln\ln n}{n}\right)^{r/2}$
ν_i	$\frac{\ln n - k\ln\ln n}{q\lambda}$	$\frac{\ln n - k\ln\ln n}{2\lambda}$	$\frac{\ln n - k\ln\ln\ln n - \ln\ln\ln n}{2\lambda}$

Theorem 3.1 reduces obtaining of lower bounds for risks to the construction of the "richest" family $f_{i\delta}$ with the required properties. This construction is realized in IH (1980, 1981) for Examples 1 and 2 and $p \ge 2$. As a result we have that the upper bounds of Section 2 for these examples coincide with the lower bounds in the sense of rate of convergence (see Table 1). The construction of a suitable family for the case $1 \le p \le 2$ is the content of Section 4. Table 1 is the final result of the considerations in Sections 2–4.

We use the notations

$$q = p/(p-1), \qquad \beta = \left(\sum_{i=1}^{k} \beta_i^{-1}\right)^{-1}.$$

The notation $a_n \asymp b_n$ means that $0 < \liminf(a_n/b_n) \le \limsup(a_n/b_n) < \infty$ and $\Delta_n(p,r)$ is the quantity which is defined in (1.1). The first and third lines give rough (in the sense $\asymp$) asymptotics for the risk; the second and fourth lines present the values $\nu = (\nu_1, \dots, \nu_k)$ in the estimator $\tilde{f}_{n,\nu}$, for which this order for the rate of convergence is reached.

4. An illustration. Let us demonstrate the construction of the family $f_{i\delta}$, satisfying the conditions of Theorem 3.1, for the case $p \in [1, 2[$. We restrict ourselves to the case $k = 1$ for simplicity [the generalization for arbitrary k can be made analogously; see IH (1981)].

Put $f_{0\delta}(x) = \pi^{-1}\delta^q/(x^2\delta^{2q} + 1)$ and consider the family of densities $f_{a\delta}$, depending on the M-dimensional vector $a = (a_1, \dots, a_M)$, where a_i is 0 or 1:

$$f_{a\delta}(x) = f_{0\delta}(x) + \gamma \sum_{m=1}^{M} a_m \phi_{m\delta}(x). \tag{4.1}$$

Here $M = \lceil \delta^{-q-\beta^{-1}} \rceil$, $\phi_{m\delta}(x) = \delta^q \phi(x\delta^{-\beta^{-1}} - m)$ and the function ϕ [cf. IH (1980)] is infinitely differentiable, has the support $]-\frac{1}{2}, \frac{1}{2}[$ and satisfies

$\int \phi(x)\,dx = 0$. It is easy to verify that the conditions

$$f_{a\delta} \ge 0, \qquad f_{a\delta} \in H_p^\beta L, \qquad \int f_{a\delta}(x)\,dx = 1$$

are fulfilled, if γ is sufficiently small. Furthermore

$$\| f_{a\delta} - f_{a'\delta} \|_p = \gamma \left(\sum_{m=1}^{M} |a_m - a'_m| \right)^{1/p} \| \phi_{m\delta} \|_p$$

$$= \gamma \| \phi \| \delta^{q+(\beta p)^{-1}} \left(\sum_1^M |a_m - a'_m| \right)^{1/p}.$$

This equality and the relation $M \sim \delta^{-q-\beta^{-1}}$ guarantee the fulfilment of all conditions of Theorem 3.1, provided $a \in A$, where A is the largest set of a for which

(4.2) $$\sum_1^M |a_m - a'_m| \ge M/4, \qquad a, a' \in A,\ a \ne a'.$$

It is known [for instance, it follows from the Varshamov–Gilbert bound; see IH (1980, 1981)] that

(4.3) $$\operatorname{card} A > \exp(M/2).$$

Finally, a very simple verification gives the inequality

$$\left\| \frac{f_{a\delta} - f_{0\delta}}{\sqrt{f_{0\delta}}} \right\|_2^2 < C \quad \text{for } \delta \le \delta_0.$$

The last inequality and the relation $\ln N(\delta) \asymp M\delta^{-q-\beta^{-1}}$ [this follows from (4.3)] give the equality

$$\delta(n, \Sigma) = cn^{-\beta/(q\beta+1)}.$$

So we have established the lower bound in the first line of the table for the situation $1 \le p < 2$.

If $\Sigma = A_p^\lambda L$ we can consider the analogous family

$$f_{a\delta}(x) = f_{0\delta}(x) + \gamma\delta^q \sum_{m=1}^{M} a_m \phi_0(x \ln(1/\delta)/\lambda_0 - m).$$

Here $\phi_0(x) = \sin^5 x/x^4$, $f_{0\delta}(x)$ is the same as in (4.1), γ is sufficiently small, λ_0 is sufficiently large and $M = [\delta^{-q} \ln(1/\delta)]$. The analogous verification gives the results: For suitable γ and λ_0, the functions $f_{a\delta}$ are densities from $A_p^\lambda L$ and for $a, a' \in A$, $a \ne a'$,

$$\| f_{a\delta} - f_{a'\delta} \| \ge c_1 \delta \qquad (c_1 > 0); \qquad \left\| \frac{f_{a\delta} - f_{0\delta}}{\sqrt{f_{0\delta}}} \right\|_2^2 < C_2.$$

These inequalities and (3.4) imply the relation

$$n \asymp \ln N(\delta) \asymp (\delta(n, \Sigma))^{-q} \ln(1/\delta(n, \Sigma)).$$

So $\delta(n, \Sigma) \asymp (\ln n/n)^{1/q}$ and Theorem 3.1 gives the lower bound in the third line of Table 1 for the case $1 \leq p < 2$ and $k = 1$.

5. Further results. Let Σ_Λ be the set of densities which have analytic extension as entire function of exponential type Λ, where Λ is a symmetric convex compact set in R^k. In other words Σ_Λ is the set of densities, the characteristic function of which is equal to zero outside Λ. The set Σ_Λ is essentially infinite dimensional, but nevertheless it is proved in IH (1980, 1981, 1982) that for $p \geq 2$, the rate of convergence of the best estimator has the order $n^{1/2}$:

$$\inf_{f_n} \sup_{f \in \Sigma_\Lambda} E_f \| f_n - f \|_p^r \leq c_r n^{-r/2}.$$

Moreover for $p = r = 2$ the precise asymptotic bound is found in IH (1982):

$$\lim_{n \to \infty} \left\{ n \inf_{f_n} \sup_{f \in \Sigma_\Lambda} E_f \| f_n - f \|_2^2 \right\} = \frac{\text{meas } \Lambda}{(2\pi)^k}. \tag{5.1}$$

This result gave occasion to formulate the hypothesis [Devroye and Györfi (1985)] that the same order $n^{1/2}$ for the rate of convergence is preserved for $p < 2$ and in particular for $p = 1$. But in IH (1981) the following relations, refuting this hypothesis, are proved:

$$\inf_{f_n} \sup_{f \in \Sigma_\Lambda} E_f \| f_n - f \|_p \asymp n^{-1/q}, \qquad 1 < p < 2,$$

$$\liminf_{n} \inf_{f_n} \sup_{f \in \Sigma_\Lambda} E_f \| f_n - f \|_1 \geq \aleph > 0.$$

6. Remarks and problems.

1. The intrinsic reason for the existence of an estimator with property (5.1) is the fact that the densities $f \in \Sigma_\Lambda$ have a reproducing kernel. This means for $k = 1$, for example,

$$f(x) = \int_{R^1} \frac{\sin \Lambda(x - y)}{\pi(x - y)} f(y)\, dy.$$

So an estimator, which is asymptotically efficient in the sense of (5.1), has the form

$$f_n(x) = \int_{R^1} \frac{\sin \Lambda(x - y)}{\pi(x - y)} F_n(dy).$$

Similar estimators are possible in other situations, with sets $\tilde{\Sigma}$ say, provided that for each $f \in \tilde{\Sigma}$ the representation

$$f(x) = \int K(x, y) f(y)\, dy$$

holds with sufficiently good kernel K. If we restrict ourselves to the case

$K(x, y) = K_1(x - y)$, then the equality

$$E\| \hat{f}_n - f \|_2^2 = \frac{1}{n}\left(K_1(0) - \| f \|^2\right)$$

holds. We think that the following generalization of IH (1982) is true in the latter case:

$$\lim_{n \to \infty} \left\{ n \inf_{f_n} \sup_{f \in \Sigma} E_f \| f_n - f \|_2^2 \right\} = K_1(0).$$

2. The precise asymptotics for the quadratic risk in L_2 of the type (5.1) were first obtained in Efroimovich and Pinsker (1982) for the case where Σ is the set of ellipsoids in L_2. Other references can be found in this paper.
3. The problem of constructing upper and lower bounds is interesting for other Σ too. Nemirovskii (1985), for instance, found the true asymptotics for the regression estimation problem if $\Sigma = \{\| f^{(K)} \|_p < L\}$ and the loss function is the L_{p_1} norm of the difference $\tilde{f}_n - f$. Here, $p, p_1 > 1$ are arbitrary. It would be interesting to obtain the corresponding results for the density estimation problem. Other interesting sets Σ are considered by Bretagnolle and Huber (1979), Bentkus and Kazbaras (1982) and Devroye and Györfi (1985).
4. The upper bound of Section 2 can be extended to more general classes of loss functions. It is possible to consider the loss function $l(\| f_n - f \|_p)$; see IH (1981) for details.
5. The present approach is also suitable for obtaining bounds in L_p norms for the estimation of the derivatives of a density. An interesting and difficult problem is obtaining the *precise* asymptotics for more general Σ and $p \neq 2$. For instance, it is interesting to consider Σ, which are determined by a condition of the type

$$\int \phi(t) |\hat{f}(t)|^2 \, dt \leq 1,$$

where $\phi \geq 0$ and $\hat{f}$ is the characteristic function. Interesting results concerning this problem were presented recently by G. K. Golubev at the Fifth Vilnius Conference on Probability Theory and Statistics.

Acknowledgments. We would like to thank W. R. van Zwet for his very pleasant and honourable suggestion to submit this paper to the *Annals of Statistics* and S. van de Geer for her kind help. We are indebted to R. Gill for helpful advice and criticism.

REFERENCES

BABENKO, K. I. (1979). *Theoretical Base and Construction of Numerical Algorithms for Physics Problems* 295. Nauka, Moscow.

BABENKO, K. I. (1985). On some problems of approximation theory and numerical problems. *Russian Math. Survey* **40** 3–27.

BAHR, B. VON and ESSEEN, C. (1965). The inequalities for the absolute moment for a sum of random variables. *Ann. Math. Statist.* **36** 299–303.

BENTKUS, R. and KAZBARAS, A. (1982). Optimal statistical estimates of a distribution density under a prior information. *Lietuvos Mathematikos Rinbeinus* **22** 29–39.

BRETAGNOLLE, J. and HUBER, C. (1979). Estimation de densities: risque minimax. *Z. Wahrsch. Verw. Gebiete* **47** 119–137.

CENTSOV, N. (1962). Estimation of the unknown density function. *Dokl. Akad. Nauk SSSR* **147** 45–48.

CENTSOV, N. (1972). *Statistical Decision Rules and Optimal Inference*. AMS, Providence, R.I.

DEVROYE, L. and GYÖRFI, L. (1985). *Nonparametrical Density Estimation: The L_1 View*. Wiley, New York.

EFROIMOVICH, S. and PINSKER, M. S. (1982). The estimation of the square integrable density function. *Problems Inform. Transmission* **18** 19–38.

FARRELL, R. (1972). On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. *Ann. Math. Statist.* **43** 170–180.

IBRAGIMOV, I. and HASMINSKII, R. (1979). Statistical estimation. *Asymptotic Theory*. Springer, New York.

IBRAGIMOV, I. and HASMINSKII, R. (1980a). On estimation of the density function. *Zap. Nauch. Sem. LOMI AN SSSR, 1980* **98** 61–85.

IBRAGIMOV, I. and HASMINSKII, R. (1980b). On the asymptotic rates of convergence in nonparametric regression estimation in L_p. *Zap. Nauch. Sem. LOMI AN SSSR, 1980* **97** 88–101.

IBRAGIMOV, I. and HASMINSKII, R. (1981). Further results nonparametric density estimation. *Zap. Nauch. Sem. LOMI AN SSSR, 1981* **108** 73–89.

IBRAGIMOV, I. and HASMINSKII, R. (1982). On a density estimation within a class of entire functions. *Probab. Theory Appl.* **27** 514–524.

KOLMOGOROV, A. (1936). Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse. *Ann. Mat.* **37** 107–110.

KOLMOGOROV, A. and TIKHOMIROV, V. (1959). ε-entropy and ε capacity of sets in the functional spaces. *Russian Math. Survey* **14** 3–86.

NEMIROVSKII, A. (1985). On parametrical estimation of the smooth regression functions. *Tech. Kibernetica* **1985** 50–60.

NIKOLSKII, S. (1969). The approximation of the functions of several variables and imbedding theorems. Nauka, Moscow.

NUSSBAUM, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L_2. *Ann. Statist.* **13** 984–997.

INSTITUTE FOR PROBLEMS IN
INFORMATION TRANSMISSION
UL. ERMOLOVOY 19
MOSCOW 101447
UNION OF SOVIET SOCIALIST REPUBLICS

LENINGRAD BRANCH
STEKLOV MATHEMATICAL INSTITUTE
27 FOTANKA QUAY, LENINGRAD, 191011
UNION OF SOVIET SOCIALIST REPUBLICS