The Annals of Statistics
1988, Vol. 16, No. 3, 1180–1187

A LOWER BOUND ON THE ERROR IN NONPARAMETRIC REGRESSION TYPE PROBLEMS[1]

BY YANNIS G. YATRACOS[2]

Université du Québec à Montréal and Columbia University

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a sample, denote the conditional density of $Y_i|X_i = x_i$ as $f(y|x_i, \theta(x_i))$ and θ an element of a metric space (Θ, d). A lower bound is provided for the d-error in estimating θ. The order of the bound depends on the local behavior of the Kullback information of the conditional density. As an application, we consider the case where Θ is the space of q-smooth functions on $[0,1]^d$ metrized with the L_r distance, $1 \le r < \infty$.

1. Introduction. In the classical nonparametric regression problem, we consider a sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ where $X_1, \ldots, X_n$ are R^d valued measurements that might be random or nonrandom, $Y_1, \ldots, Y_n$ are the corresponding responses such that $E(Y_i|X_i = x_i) = \theta(x_i)$ with θ in an infinite dimensional space Θ. Conditionally on $X_1 = x_1, \ldots, X_n = x_n$, the responses are independent with distributions of the same form $f(y|x, \theta(x))\,dy := P_{\theta(x)}(dy)$, but with parameters depending on the measurements x_i, $i = 1, \ldots, n$. Under this setup, Stone (1980, 1982) and Ibragimov and Khas'minskii (1980) have constructed optimal estimators $\hat{\theta}_n$ of θ in L_r, $1 \le r$, when Θ consists of q-smooth functions on $[0,1]^d$. Ibragimov and Khas'minskii proved that their estimators are almost minimax modulo a constant, that is, there are constants C_L, C_U such that $\sup\{E_\theta \|\hat{\theta}_n - \theta\|_r; \theta \in \Theta\} \le C_U n^{-\gamma}$ and $\inf\{\sup\{E_\theta \|\hat{T}_n - \theta\|_r; \theta \in \Theta\}; \hat{T}_n\} \ge C_L n^{-\gamma}$, $\gamma > 0$, $n \in N$. Stone has considered other definitions of optimality using bounds in probability for the loss $\|\hat{\theta}_n - \theta\|_r$ as described at the end of the paper.

We will relax the condition that $\theta(x)$ is a conditional mean as in the classical regression problem. We will only assume that $\theta(x)$ is a parameter of the conditional density and we will call the problem of estimating θ a regression type problem. When θ is an element of a metric space (Θ, d), we will provide for the regression type problem a lower bound on the d-minimax risk (Theorem 1). This theorem can be used as a tool to provide lower bounds for different choices of (Θ, d). We will apply the theorem and evaluate the lower bound in the case where Θ is a family of smooth functions and d is the L_r distance, $1 \le r < \infty$. In a remark at the end of the paper, we provide with the same technique, lower bounds for the d-loss in probability. A minimum distance estimate for the

Received November 1985; revised November 1987.

[1]Research partially supported by a grant from the Office of Research and Technology of Greece.

[2]On leave from Columbia University during 1988–1989.

AMS 1980 *subject classification.* Primary 62G20.

Key words and phrases. Nonparametric regression, lower bound on minimax risk, lower bound of loss in probability, optimal rates of convergence, Kullback information.

regression type problem proposed in Yatracos (1985a) shows achievability of the lower bound. The error of this estimate depends on the entropy of Kolmogorov of (Θ, d) as in the density estimation problem [Yatracos (1985b)].

The method for computing the lower bound comes from Le Cam's idea in hypothesis testing [Le Cam (1986) or Kraft (1955)] that you cannot test (and so estimate) θ_0 versus $\Theta - \{\theta_0\}$ if θ_0 is in the convex hull of $\Theta - \{\theta_0\}$. So it will be difficult to test θ_0 versus $\Theta_n \subset \Theta - \{\theta_0\}$ when Θ_n consists of functions close to θ_0, the difficulty being reflected in the lower bound of the minimax or Bayes risk. This idea has already been used by Bretagnolle and Huber (1979) to obtain lower bounds for the risk in the nonparametric density estimation problem. A similar approach, using Fano's lemma, has been considered to obtain lower bounds for minimax risks by Khas'minskii (1978) and Birgé (1983) in density estimation and by Ibragimov and Khas'minskii (1981) in classical regression with equidistant design. An observation that a regression problem is almost a density estimation problem leads to the use of Fano's lemma and a lower bound for an arbitrary metric space (Θ, d). An elegant result of Birgé helps to obtain the best lower bound when Θ is the space of (q, L) smooth functions on $[0,1]^d$ metrized with the L_r distance, $1 \leq r < \infty$ (i.e., Θ consists of p times differentiable functions in $[0,1]^d$, uniformly bounded in sup-norm with the pth derivative satisfying a Lipschitz condition with parameters (L, α), $q = p + \alpha$, $0 \leq p$, $0 < \alpha \leq 1$). Note that Fano's lemma involves the Kullback information $K(P_{\theta_1(x)}, P_{\theta_2(x)})$, so we will have to evaluate it or find an upper bound for it. It is easy to see that for the case considered by Stone, $K(P_{\theta_1(x)}, P_{\theta_2(x)}) \leq C(\theta_1(x) - \theta_2(x))^2$. It is this condition that makes the estimators of Ibragimov and Khas'minskii and Stone asymptotically optimal and not the nature of $\theta(x)$ in the conditional density. It is the behavior of the Kullback information $K(P_{\theta_1(x)}, P_{\theta_2(x)})$ locally that will determine the lower bound on the risk and the lower bound of the loss in probability.

For sample size n, we will compute a lower bound on the $\sup\{E_\theta\, d(\hat{T}_n, \theta);\ \theta \in \Theta\}$ by considering a bound on $\sup\{E_\theta\, d(\hat{T}_n, \theta);\ \theta \in \Theta_n\}$ with Θ_n an appropriate subset of Θ according to Le Cam's idea. It turns out that when Θ is the set of q-smooth functions on $[0,1]^d$, one can use a set Θ_n similar to the one used by Kolmogorov and Tihomirov (1959) to compute a lower bound for the entropy of smooth functions on $[0,1]^d$. We should also mention, at this point, the work by Boyd and Steele (1978) and Assouad (1983). The former proved that in the nonparametric density estimation problem, considering all densities with squared error loss, the minimax risk cannot be better than $O(n^{-1})$. The latter provided a lower bound on risks for any loss and related the $O(n^{-1/2})$ minimax risk with dimensionality properties of the space of probability measures under consideration.

Khas'minskii (1978) provides lower bounds on the risks of nonparametric estimates of densities in the uniform metric. Devroye (1986) computes minimax bounds on the L_1 loss for the class of kernel estimates. For a detailed study on lower bounds on minimax risks, the reader could consult Devroye and Gyorfi (1985) and Devroye (1987).

2. Notation. Definitions. The results. Let $(\mathscr{X}, \mathscr{A})$, $(\mathscr{Y}_1, \mathscr{B}_1), \dots, (\mathscr{Y}_n, \mathscr{B}_n)$ be spaces with their σ-fields. Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a sample with X_i taking values in $\mathscr{X}$, $i = 1, \dots, n$, Y_i taking values in $\mathscr{Y}_i$, $i = 1, \dots, n$.

DEFINITION. For any two functions f, g on $(\mathscr{X}, \mathscr{A})$, $L_r(\lambda)$ integrable, $1 \le r < \infty$, their L_r distance is

$$\|f - g\|_r := \left(\int_{\mathscr{X}} |f(x) - g(x)|^r \lambda(dx) \right)^{1/r}.$$

DEFINITION. For any two probability measures P, Q on $(\mathscr{Y}, \mathscr{B})$, their Kullback information $K(P, Q) = E_p \log(dP/dQ)$ if P is absolutely continuous with respect to Q; otherwise, $K(P, Q) = +\infty$.

In the case of product measures $K(P_1 \times P_2 \times \cdots \times P_n, Q_1 \times Q_2 \times \cdots \times Q_n) = \sum_{i=1}^n K(P_i, Q_i)$.

FANO'S LEMMA [Birgé (1983) or Ibragimov and Khas'minskii (1981)]. *Let $(\mathscr{Y}, \mathscr{B})$ be a space with a σ-field, $P_1, \dots, P_m$ probability measures on $\mathscr{B}$ and δ an estimator of the measures defined on $\mathscr{Y}$. Then*

$$\frac{1}{m} \sum_{i=1}^{m} P_i[\delta(y) \neq P_i] \ge 1 - \frac{m^{-2}\sum_{i,j} K(P_i, P_j) + \log 2}{\log(m-1)}.$$

DEFINITION. Let d be a distance on a subset $\mathscr{P}$ of the $L_1(\lambda)$ functions on $(\mathscr{X}, \mathscr{A})$, λ a probability measure on $\mathscr{A}$, Φ a function, Φ: $R^+ \to R^+$. The function $\Phi \circ d$ is called superadditive if for every finite partition $\{A_i;\ 1 \le i \le l\}$ of $\mathscr{X}$, we have for f, g in $\mathscr{P}$

$$\Phi(d(f, g)) = \sum_{i=1}^{l} \Phi\left[d(fI_{A_i}, gI_{A_i}) \right].$$

This property has been introduced by Bretagnolle and Huber (1979) and is satisfied by $\|f - g\|_r^r$ on $L_r(\lambda)$, $r \ge 1$.

BIRGÉ'S THEOREM [Birgé (1983), Proposition 3.8]. *Let $\{A_i;\ 1 \le i \le l\}$ be a partition of $\mathscr{X}$, and f, g_i and g_i' be elements of $L_1(\lambda)$ with support on A_i. Let $\Theta = \{f + \sum_{i=1}^l \lambda_i | \lambda_i = g_i$ or $g_i'\}$ and assume that for all i, $d(f + g_i, f + g_i') \ge a$ and that d^r is superadditive for some $r \ge 1$. Then there is a subset Θ^* of Θ such that $d(f^*, g^*) \ge a(0.125l)^{1/r}$ for $f^* \neq g^*$ elements of Θ^* and* $\log(\text{card}\,\Theta^* - 1) > 0.316l$ *for any $l \ge 8$.*

THEOREM 1. *Under the regression setup of the Introduction for the sample $(X_1, Y_1), \dots, (X_n, Y_n)$, for Θ_n a subset of Θ with finite cardinality, d a distance* *on* Θ *and* $\hat{T}_n$ *an estimator of the regression type function* θ,

$$\sup\{E_\theta\, d(\hat{T}_n, \theta);\ \theta \in \Theta\}$$
$$\geq \frac{1}{2}\inf\{d(\theta_1, \theta_2);\ \theta_1 \neq \theta_2, (\theta_1, \theta_2) \in \Theta_n^2\}$$
$$\times E\left[1 - \frac{(\operatorname{card}\Theta_n)^{-2}\sum_{(\theta_1,\theta_2)\in\Theta_n^2}\sum_{i=1}^n K\big(P_{\theta_1(X_i)}, P_{\theta_2(X_i)}\big) + \log 2}{\log(\operatorname{card}\Theta_n - 1)}\right],$$

the last expectation taken with respect to the probability measure of $(X_1, \dots, X_n)$.

PROOF.

$$\sup\{E_\theta\, d(\hat{T}_n, \theta);\ \theta \in \Theta\} \geq \sup\{E_\theta\, d(\hat{T}_n, \theta);\ \theta \in \Theta_n\}$$
$$\geq \frac{1}{\operatorname{card}\Theta_n}\sum_{\theta\in\Theta_n} E_\theta\, d(\hat{T}_n, \theta)$$
$$= E\left[\frac{1}{\operatorname{card}\Theta_n}\sum_{\theta\in\Theta_n} E_\theta\big(d(\hat{T}_n, \theta)|X_1, \dots, X_n\big)\right].$$

Define $\hat{T}_n^*$ taking values in Θ_n such that $d(\hat{T}_n, \hat{T}_n^*) = \inf\{d(\hat{T}_n, \theta);\ \theta \in \Theta_n\}$. Then we have for $\theta \in \Theta_n$, $d(\hat{T}_n^*, \theta) \leq d(\hat{T}_n^*, \hat{T}_n) + d(\hat{T}_n, \theta) \leq 2d(\hat{T}_n, \theta)$. So

$$\frac{1}{\operatorname{card}\Theta_n}\sum_{\theta\in\Theta_n} E_\theta\big[d(\hat{T}_n, \theta)|X_1, \dots, X_n\big]$$
$$\geq \frac{1}{2}\inf\{d(\theta_1, \theta_2);\ \theta_1 \neq \theta_2, (\theta_1, \theta_2) \in \Theta_n^2\}$$
$$\times \frac{1}{\operatorname{card}\Theta_n}\sum_{\theta\in\Theta_n} P_\theta\big[\hat{T}_n^* \neq \theta|X_1, \dots, X_n\big]$$
$$\geq \frac{1}{2}\inf\{d(\theta_1, \theta_2);\ \theta_1 \neq \theta_2, (\theta_1, \theta_2) \in \Theta_n^2\}$$
$$\times\left[1 - \frac{(\operatorname{card}\Theta_n)^{-2}\sum_{(\theta_1,\theta_2)\in\Theta_n^2}\sum_{i=1}^n K\big(P_{\theta_1(x_i)}, P_{\theta_2(x_i)}\big) + \log 2}{\log(\operatorname{card}\Theta_n - 1)}\right],$$

by applying Fano's lemma to the measures $P_{\theta(x_1)} \times \cdots \times P_{\theta(x_n)}$, $\theta \in \Theta_n$, on the product space $\mathscr{Y}_1 \times \mathscr{Y}_2 \times \cdots \times \mathscr{Y}_n$. □

REMARK 1. The selection of an appropriate (almost least favorable) subset Θ_n of Θ (as in Corollary 2) will lead to an almost minimax lower bound.

COROLLARY 1. *If* $K(P_{\theta_1(x_i)}, P_{\theta_2(x_i)}) \le c_n$ *and* $\inf\{d(\theta_1, \theta_2);\ \theta_1 \ne \theta_2, (\theta_1, \theta_2) \in \Theta_n^2\} \ge a_n$, *then*

$$\sup\{E_\theta\, d(\hat{T}_n, \theta);\ \theta \in \Theta\} \ge \frac{1}{2} a_n \left[1 - \frac{nc_n + \log 2}{\log(\operatorname{card}\Theta_n - 1)}\right].$$

PROPOSITION 1. *Assume for the conditional density* $f(y|x, t)$ *that*:

(i) $\int_{\mathcal{Y}} f'(y|x, t)\mu(dy) = 0.$

(ii) *If* $l(y|x, t) = \log f(y|x, t)$, *there are positive constants* ε_0 *and* K_1 *and a function* $M(y|x, t)$ *such that* $|l''(y|x, t + \varepsilon)| \le M(y|x, t)$ *for* $|\varepsilon| \le \varepsilon_0$ *and*

$$\int_{\mathcal{Y}} M(y|x, t) f(y|x, t)\mu(dy) \le K_1.$$

(f', l'' *are partial derivatives with respect to t, the parameter of the conditional density.*) *Then* $K(P_t, P_s) \le C(t - s)^2$. [*Conditions* (i) *and* (ii) *were used in Stone* (1982).]

PROOF. When $|t - s|$ is small, making a Taylor expansion, we have

$$\begin{aligned} K(P_t, P_s) &= \int_{\mathcal{Y}} f(y|x, t)\log\frac{f(y|x, t)}{f(y|x, s)}\mu(dy) \\ &= -\int_{\mathcal{Y}} f(y|x, t)\left[(s - t)\frac{f'(y|x, t)}{f(y|x, t)} + \frac{(t - s)^2}{2} l''(y|x, c)\right]\mu(dy) \\ &\le \frac{K_1}{2}(t - s)^2, \end{aligned}$$

where c is in the open interval determined by t and s. □

Let Θ be the space of (q, L) smooth functions on $[0, 1]$. We introduce a family Θ_n, but we will use a subset Θ_n^* of it to apply Theorem 1.

Let

$$\phi_{i,n}(x) = \begin{cases} ab_n^q\left[1 - \left(\dfrac{x - 0.5(2i - 1)b_n}{0.5b_n}\right)^2\right]^q, & \text{if } (i - 1)b_n \le x \le ib_n, \\ 0, & \text{otherwise}, \end{cases}$$

$i = 1, 2, \ldots, b_n^{-1}$, where $a > 0$ can be chosen appropriately to make the constant of the Lipschitz condition less than or equal to L. The set Θ_n will consist of functions $\theta(x)$ of the form $\sum_{i=1}^{b_n^{-1}} \gamma_i \phi_{i,n}(x)$, $\gamma_i = 0$ or 1.

Note that the L_r distance between functions of Θ_n will be greater than or equal to

$$ab_n^q\left[\int_0^{b_n}\left[1-\left(\frac{x-0.5b_n}{0.5b_n}\right)^2\right]^{qr}dx\right]^{1/r}$$
$$=\frac{a}{2^{1/r}}b_n^{q+(1/r)}\left[\int_{-1}^{1}(1-y^2)^{qr}\,dy\right]^{1/r}=C_{q,r,a}b_n^{q+(1/r)}.$$

It is also easy to see that $I_r=\int_{-1}^1(1-y^2)^r\,dy=2r/(2r+1)I_{r-1}$ for $r\geq 1$, $I_0=2$, and that $|\theta_1(x)-\theta_2(x)|\leq ab_n^q$ for all θ_1, θ_2 in Θ_n and for all $x\in[0,1]$.

In the case $\mathscr{X}=[0,1]^d$, we consider functions $\phi_{j_1,\dots,j_d,n}$ of the form

$$\phi_{j_1,j_2,\dots,j_d,n}(x_1,\dots,x_d)$$
$$=\begin{cases} ab_n^q\prod_{i=1}^d\left[1-\left(\dfrac{x_i-0.5(2j_i-1)b_n}{0.5b_n}\right)^2\right]^q, \\ \qquad \text{if } (j_i-1)b_n\leq x_i\leq j_ib_n,\ i=1,\dots,d, \\ 0, \qquad \text{otherwise,}\end{cases}$$

$j_i=1,2,\dots,b_n^{-1}$.

Note that there are b_n^{-d} such rectangles in $[0,1]^d$. Let us enumerate them as $I_{1,n},\dots,I_{b_n^{-d},n}$. So, we can write $\phi_{I_{1,n}},\dots,\phi_{I_{b_n^{-d},n}}$ instead of $\phi_{1,1,\dots,1,n},\dots,\phi_{b_n^{-1},\dots,b_n^{-1},n}$. So, the functions $\theta(x_1,\dots,x_d)$ of Θ_n will be of the form $\sum_{i=1}^{b_n^{-d}}\gamma_i\phi_{I_{i,n}}(x_1,\dots,x_d)$, $\gamma_i=0$ or 1. The lower bound on the L_r distance between functions of Θ_n will be greater than or equal to $C_{q,r,a,d}b_n^{q+(d/r)}$.

Corollary 2. *If Θ is the set of q-smooth functions on $[0,1]^d$ for conditional densities such that $K(P_{\theta_1(x)},P_{\theta_2(x)})\leq C|\theta_1(x)-\theta_2(x)|^K$ for some $K>0$, the L_r minimax risk is greater than or equal to $C^*n^{-q/(Kq+d)}$.*

Proof. Consider the preceding set of functions Θ_n of the form $\sum_{i=1}^{b_n^{-d}}\gamma_i\phi_{I_{i,n}}(x_1,\dots,x_d)$. By Birgé's theorem, there exists a subset Θ_n^* of Θ_n such that $\|\theta_1-\theta_2\|_r\geq(0.125b_n^{-d})^{1/r}C_{q,r,a,d}b_n^{q+(d/r)}$ for all $\theta_1\neq\theta_2$ in Θ_n^* and $\log(\operatorname{card}\Theta_n^*-1)>0.316b_n^{-d}$. By assumption,

$$K\left(P_{\theta_1(x)},P_{\theta_2(x)}\right)\leq C|\theta_1(x)-\theta_2(x)|^K\leq Cb_n^{Kq}\quad\text{for all }\theta_1,\theta_2\text{ in }\Theta_n.$$

By Corollary 1 for any estimator $\hat{T}_n$,

$$\sup\left\{E_\theta\|\hat{T}_n-\theta\|_r;\theta\in\Theta\right\}\geq\frac{1}{2}C_{q,r,a,d}(0.125)^{1/r}b_n^q\left[1-\frac{C'nb_n^{Kq}+\log 2}{0.316b_n^{-d}}\right].$$

For $[1-(C'nb_n^{Kq}+\log 2)/0.316b_n^{-d}]$ to be greater than a positive number, it is enough to take $b_n\sim n^{-1/(Kq+d)}$. The minimax risk cannot be better than $C^*n^{-q/(Kq+d)}$. □

COROLLARY 3. *When Θ is the space of all regression functions on $[0,1]^d$ and the conditional densities have the property $K(P_{\theta_1(x)}, P_{\theta_2(x)}) \leq C|\theta_1(x) - \theta_2(x)|^K$, $K > 0$, the L_r minimax risk, $1 \leq r < \infty$, is greater than or equal to $n^{-1/K}$.*

EXAMPLE 1. In the case of conditional densities $f(y|x, \theta(x))$ that are one of the following, Bernoulli $(\theta(x))$, binomial $(N, \theta(x))$, geometric $(\theta(x))$ and exponential $(\theta(x))$, we see that $K(P_{\theta_1(x)}, P_{\theta_2(x)}) \leq C(\theta_1(x) - \theta_2(x))^2$ so the lower bound for the L_r-minimax risk is of the order $n^{-q/(2q+d)}$. The same holds for the normal $(\theta_1(x), \theta_2^2(x))$ when we are interested in $\theta_1(x)$ and $\theta_2(x)$ is bounded away from 0 or when we are interested in $\theta_2(x)$ and it is bounded away from 0 and ∞.

EXAMPLE 2. In the case the conditional density is either uniform $(\theta(x))$ or has the form $e^{\theta(x)-y}$, $K(P_{\theta_1(x)}, P_{\theta_2(x)}) \leq C|\theta_1(x) - \theta_2(x)|$, so the lower bound of the L_r-minimax risk is $n^{-q/(q+d)}$. One could also derive lower bounds in the case $K(P_{\theta_1(x)}, P_{\theta_2(x)}) \leq g(|\theta_1(x) - \theta_2(x)|)$.

REMARK 2. When the X's are nonrandom and equidistant, the set Θ_n considered is a least favorable set since at the design points $b_n, 2b_n, \dots,$ the value of θ is 0.

REMARK 3. One can define d-optimality for a sequence $\{\hat{\theta}_n\}$ of estimators of θ in a regression type problem using bounds in probability for the loss $d(\hat{\theta}_n, \theta)$ [Stone (1980)]. We say $\{\hat{\theta}_n\}$ is d-optimal in probability for θ if there is a sequence a_n, $n = 1, 2, \dots,$ decreasing to zero such that

$$\lim_{C \to 0} \liminf_{n \to \infty} \inf_{\hat{T}_n} \sup_{\theta \in \Theta} P_\theta\left[d\left(\hat{T}_n, \theta\right) > Ca_n\right] = 1$$

and

$$\lim_{C \to \infty} \limsup_{n \to \infty} \sup_{\theta \in \Theta} P_\theta\left[d\left(\hat{\theta}_n, \theta\right) > Ca_n\right] = 0.$$

To verify the former relation, we can use as a tool a variant of Theorem 1 under the additional assumption that $d(\theta_1, \theta_2) > 2Ca_n$ for all $(\theta_1, \theta_2) \in \Theta_n^2$, $\theta_1 \neq \theta_2$. For every estimator $\hat{T}_n$, we have then

$$\sup_{\theta \in \Theta} P_\theta\left[d\left(\hat{T}_n, \theta\right) > Ca_n\right]$$
$$\geq E\left[1 - \frac{(\operatorname{card} \Theta_n)^{-2} \Sigma_{(\theta_1, \theta_2) \in \Theta_n^2} K\left(P_{\theta_1(X_i)}, P_{\theta_2(X_i)}\right) + \log 2}{\log(\operatorname{card} \Theta_n - 1)}\right].$$

Under the previous setup, when Θ is the set of q-smooth functions on $[0,1]^d$, $\Theta_n = \Theta_n^*$, $C = 0.5(0.125)^{1/r} C_{q, r, a, d}$, $a_n = b_n^q$, $K(P_{\theta_1(x)}, P_{\theta_2(x)}) \leq \tilde{C}|\theta_1(x) - \theta_2(x)|^K$ and d is the L_r distance, $1 \leq r < \infty$, we have that for every estimate $\hat{T}_n$ and $\varepsilon > 0$, $\sup_{\theta \in \Theta} P_\theta[\|\hat{T}_n - \theta\|_r > Cb_n^q] > 1 - \varepsilon$ for $n \geq n(\varepsilon)$ if $b_n \sim$

$(\varepsilon/n)^{1/(Kq+d)}$. So

$$\lim_{C \to 0} \liminf_{n \to \infty} \inf_{\hat{T}_n} \sup_{\theta \in \Theta} P_\theta\left[\|\hat{T}_n - \theta\|_r > Cn^{-q/(Kq+d)}\right] = 1.$$

REFERENCES

ASSOUAD, P. (1983). Deux remarques sur l'estimation. *C. R. Acad. Sci. Paris Sér. I* **296** 1021–1025.

BIRGÉ, L. (1983). Approximation dans les espaces metriques et theorie de l'estimation. *Z. Wahrsch. verw. Gebiete* **65** 181–237.

BOYD, D. W. and STEELE, J. M. (1978). Lower bounds for nonparametric density estimation rates. *Ann. Statist.* **6** 932–934.

BRETAGNOLLE, J. and HUBER, C. (1979). Estimation des densites: Risque minimax. *Z. Wahrsch. verw. Gebiete* **47** 119–137.

DEVROYE, L. (1986). A universal lower bound for the kernel estimate. Technical Report, School of Computer Science, McGill Univ.

DEVROYE, L. (1987). *A Course in Density Estimation.* Birkhäuser, Boston.

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

IBRAGIMOV, I. A. and KHAS'MINSKII, R. Z. (1980). On nonparametric estimation of regression. *Soviet Math. Dokl.* **21** 810–815.

IBRAGIMOV, I. A. and KHAS'MINSKII, R. Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer, New York.

KHAS'MINSKII, R. Z. (1978). A lower bound on the risks of nonparametric estimates of densities in the uniform metric. *Theory Probab. Appl.* **23** 794–798.

KOLMOGOROV, A. N. and TIHOMIROV, V. M. (1959). ε-entropy and ε-capacity of sets in function spaces. *Uspehi Mat. Nauk.* **14** 3–86 (Russian). [English translation in *Amer. Math. Soc. Transl.* **2** (1961) 277–364.]

KRAFT, C. (1955). Some conditions for consistency and uniform consistency of statistical procedures. *Univ. Calif. Publ. Statist.* **2** 125–142.

LE CAM, L. M. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer, New York.

STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

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

YATRACOS, Y. G. (1985a). A nonparametric regression type problem. Technical Report 85-001, Dept. Statistics, Rutgers Univ.

YATRACOS, Y. G. (1985b). Rates of convergence of minimum distance estimators and Kolmogorov's entropy. *Ann. Statist.* **13** 768–774.

DEPARTEMENT DE MATHÉMATIQUES
ET INFORMATIQUE
UNIVERSITÉ DU QUÉBEC À MONTRÉAL
BP 8888, SUCCURSALE A
MONTRÉAL, QUÉBEC H3C 3P8
CANADA