The Annals of Statistics
1996, Vol. 24, No. 3, 1075–1083

A MINIMAXITY CRITERION IN NONPARAMETRIC REGRESSION BASED ON LARGE-DEVIATIONS PROBABILITIES

BY ALEXANDER KOROSTELEV

Wayne State University

A large-deviations criterion is proposed for optimality of nonparametric regression estimators. The criterion is one of minimaxity of the large-deviations probabilities. We study the case where the underlying class of regression functions is either Lipschitz or Hölder, and when the loss function involves estimation at a point or in supremum norm. Exact minimax asymptotics are found in the Gaussian case.

1. Introduction. Consider observations

$$Y_{in} = f(i/n) + \xi_{in}, \qquad i = \ldots, -1, 0, 1, \ldots, n = 1, 2, \ldots, \tag{1}$$

of a regression function $f(t)$, $t \in R^1$, in an additive random noise ξ_{in}. For each n the random variables ξ_{in}'s are i.i.d. $(0, \sigma^2)$-Gaussian. The problem is to estimate the regression function f at a fixed point, say, at the origin. This problem of estimating $f(0)$ from observations (1) has been studied in various aspects [see Ibragimov and Khas'minskii (1981) and Stone (1982)].

Let $\hat{f}_n$ be an arbitrary estimator of $f(0)$; that is, $\hat{f}_n$ is a measurable function of Y_{in}, $|i| = 0, 1, \ldots$. In this paper we introduce a minimax risk associated with the logarithm of $P_f(|\hat{f}_n - f(0)| > c)$, where $P_f = P_f^{(n)}$ is the probability of the observations Y_{in} in (1) for a fixed "true" regression f; $c > 0$ is a given constant independent of n. The superscript n is omitted in the notation P_f for the sake of brevity. The random event $\{|\hat{f}_n - f(0)| > c\}$ relates to the large deviations since for each $c > 0$ and for any consistent estimator $\hat{f}_n$ its probability is vanishing as $n \to \infty$.

In Section 2 we assume that regression f belongs a priori to a class $\Sigma(L)$ of the Lipschitz functions:

$$\Sigma(L) = \big\{f : |f(t_1) - f(t_2)| \le L|t_1 - t_2|, t_1, t_2 \in R^1\big\},$$

where L is given, $L > 0$. Introduce a minimax risk

$$\beta_n(c) = \inf_{\hat{f}_n} \sup_{f \in \Sigma(L)} \frac{1}{n} \log P_f\big(|\hat{f}_n - f(0)| > c\big), \tag{2}$$

Received April 1993; revised September 1995.
AMS 1991 *subject classifications*. Primary 62G07, 62G20.
Key words and phrases. Nonparametric regression, Gaussian noise, large-deviations probabilities, minimax risk, exact asymptotics.

where the infimum is taken over all possible estimators $\hat{f}_n$. It turns out that the asymptotics of $\beta_n(c)$ as $n \to \infty$ can be found explicitly, unlike the case of the traditional risks, for example, the quadratic risk

$$\inf_{\hat{f}_n} \sup_{f \in \Sigma(L)} n^{2/3} E_f\big(\hat{f}_n - f(0)\big)^2, \tag{3}$$

where E_f is the expectation with respect to P_f [cf. Ibragimov and Khas'minskii (1981), Chapter 7, and Stone (1982)].

The minimaxity criterion (2) has a strong resemblance to the Bahadur-type risks in parametric problems [Bahadur (1960, 1971), Ibragimov and Khas'minskii (1981), Section 1.9, and Fu (1982)]. We study the problem for a fixed value of threshold c which makes the considerations global. From this point of view our results are closer in the parametric case to Sievers (1978). In the nonparametric setup the asymptotic analysis of (2) is far from being trivial even in the Gaussian case.

In Section 3 we give an extension to the case of the Hölder regression functions, applying the optimal recovery theory [Micchelli and Rivin (1977) and Donoho (1994)]. Section 4 presents a generalization of the estimation problem on the sup-norm. The asymptotic minimaxity in the sup-norm is shown to be the same as that at a point. The results are discussed in brief in Section 5.

2. Asymptotic minimaxity over Lipschitz classes.

THEOREM 1. *Under the assumptions on the observations Y_{in} in* (1), *the equality*

$$\lim_{n \to \infty} \beta_n(c) = -\frac{c^3}{3L\sigma^2} \tag{4}$$

holds.

The proof of (4) is based on the following two lemmas. First, to show that $\beta_n(c)$ does not exceed asymptotically $-c^3/(3L\sigma^2)$, we define a kernel estimator

$$f_n^* = N^{-1} \sum_{|j| \le N} Y_{jn}\left(1 - \frac{|j|}{N}\right), \tag{5}$$

where $N = [cn/L]$. We call f_n^* asymptotically minimax in the sense of risk (2) since it satisfies the following inequality.

LEMMA 1. *The inequality*

$$\limsup_{n \to \infty} \frac{1}{n} \log P_f\big(|f_n^* - f(0)| > c\big) \le -\frac{c^3}{3L\sigma^2},$$

holds uniformly in $f \in \Sigma(L)$.

PROOF. Define the bias term $b_n(f) = E_f(f_n^*) - f(0)$. Put $K_{jn} = (1 - |j|/N)^+$ and note that

$$f_n^* - f(0) = N^{-1} \sum_{|j| \le N} K_{jn} \xi_{jn} + b_n(f). \tag{6}$$

Further,

$$\begin{aligned} \sup_{f \in \Sigma(L)} b_n(f) &\le N^{-1} \sum_{|j|\le N} K_{jn} \left| f\left(\frac{j}{n}\right) - f(0) \right| \le L\left(\frac{N}{n}\right) \sum_{|j| \le N} \frac{K_{jn}|j|}{N} \\ &= c(1 + o(1)) \int_{-1}^{1} (1 - |t|)|t|\, dt = \frac{c}{3}(1 + o(1)), \end{aligned} \tag{7}$$

where $o(1) \to 0$ as $n \to \infty$, the vanishing term $o(1)$ being independent of f.

It is easily seen from (6) that

$$P_f(|f_n^* - f(0)| > c) \le P\left(\left| N^{-1} \sum_{|j| \le N} K_{jn} \xi_{jn} \right| > c - \sup_{f \in \Sigma(L)} b_n(f) \right). \tag{8}$$

The random variable $N^{-1} \sum_{|j| \le N} K_{jn} \xi_{jn}$ is zero-mean Gaussian and has variance

$$V_n = \sigma^2 N^{-2} \sum_{|j| \le N} K_{jn}^2 = \frac{2}{3}\left(\frac{L\sigma^2}{c}\right) n^{-1}(1 + o(1)). \tag{9}$$

Applying (7) and (9) to (8), we obtain the inequality

$$\limsup_{n \to \infty} \frac{1}{n} \log P_f(|f_n^* - f(0)| > c) \le \lim_{n \to \infty} \frac{1}{n}\left(- \frac{(c - c/3)^2}{2V_n} \right) = -\frac{c^3}{3L\sigma^2},$$

uniformly in $f \in \Sigma(L)$. □

To prove the reverse inequality, introduce a one-parameter family $\Sigma_0(L)$ of the regression functions $f = f(t, \theta)$. Choose $\varepsilon > 0$ small, take an arbitrary θ, $|\theta| \le c + \varepsilon$, and define

$$f(t, \theta) = \begin{cases} \theta(1 - L|t|/(c + \varepsilon)), & \text{if } |t| \le (c + \varepsilon)/L, \\ 0, & \text{otherwise.} \end{cases}$$

Use the notation P_θ for $P_{f(\cdot, \theta)}$. Note that θ coincides with the value $f(0)$ if $f \in \Sigma_0(L)$. Hence, for f known to belong to this parametric subfamily, the problem of regression estimation reduces to the problem of estimation of θ.

LEMMA 2. *For an arbitrary estimator $\hat{\theta}_n$ of θ obtained from the observations Y_{in} in* (1) *with $f = f(t, \theta)$, we have, for each $\varepsilon > 0$, the inequality*

$$\liminf_{n \to \infty} \sup_{|\theta| \le c + \varepsilon} \frac{1}{n} \log P_\theta(|\hat{\theta}_n - \theta| > c) \ge -\frac{(c + \varepsilon)^3}{3L\sigma^2} - \varepsilon.$$

PROOF. The observations Y_{in} in (1) are equivalent to a single Gaussian observation y_n which has mean value θ and variance

$$D_n^2 = (1 + o(1)) L\sigma^2 / ((2/3) n(c + \varepsilon)),$$

where $o(1) \to 0$ as $n \to \infty$. Indeed, in the model of observations $Y_{in} = a_{in}\theta + \xi_{in}$ with $a_{in} = 1 - (L/(c + \varepsilon))|i/n|$, the Fisher information with respect to θ is equal to $\sum_i a_{in}^2/\sigma^2$. Thus, the variance D_n^2 of the equivalent Gaussian experiment equals the inverse value of the Fisher information, the term $o(1)$ being independent of θ. For an arbitrary estimator $\hat{\theta}_n(y_n)$ define the random events

$$A_1 = \left\{|\hat{\theta}_n(y_n) - (c + \varepsilon)| > c\right\} \quad \text{and} \quad A_2 = \left\{|\hat{\theta}_n(y_n) + (c + \varepsilon)| > c\right\}.$$

The triangular inequality guarantees that $I(A_1) + I(A_2) \geq 1$ for any $\varepsilon \to 0$. Further, the following inequalities are true:

$$\begin{aligned}
&\sup_{|\theta| < c + \varepsilon} P_\theta\big(|\hat{\theta}_n(y_n) - \theta| > c\big) \\
&\quad \geq \frac{1}{2} P_{c+\varepsilon}(A_1) + \frac{1}{2} P_{c+\varepsilon}(A_2) \\
&\quad = \frac{1}{2} E_0\left[\frac{dP_{c+\varepsilon}}{dP_0} I(A_1) + \frac{dP_{-c-\varepsilon}}{dP_0} I(A_2)\right] \\
&\quad \geq \frac{1}{2}\exp\left(-\frac{(c+\varepsilon)^2}{2D_n^2}\right) E_0\left[\exp\left(\frac{y_n(c+\varepsilon)}{2D_n^2}\right) I(A_1)\right. \\
&\qquad\qquad \left. + \exp\left(-\frac{y_n(c+\varepsilon)}{2D_n^2}\right) I(A_2)\right] \\
&\quad \geq \frac{1}{2}\exp\left(-\frac{(c+\varepsilon)^2}{2D_n^2} - n\varepsilon\right) P_0\left(\frac{|y_n|(c+\varepsilon)}{2D_n^2} \leq n\varepsilon\right) \\
&\quad \geq \left(\frac{1}{2} - \varepsilon\right)\exp\left(-\frac{(c+\varepsilon)^2}{2D_n^2} - n\varepsilon\right)
\end{aligned}$$

for n large enough. This proves the lemma since

$$\liminf_{n\to\infty} \sup_{|\theta| < c+\varepsilon} \frac{1}{n} \log P_\theta\big(|\hat{\theta} - \theta| > c\big) \geq \lim_{n\to\infty} \frac{1}{n}\frac{(c+\varepsilon)^2}{2D_n^2} - \varepsilon = \frac{(c+\varepsilon)^3}{3L\sigma^2} - \varepsilon.$$

□

PROOF OF THEOREM 1. Lemma 1 implies that $\limsup_{n\to\infty} \beta_n(c) \leq -c^3/(3L\sigma^2)$, while Lemma 2 guarantees inequality $\liminf_{n\to\infty} \beta_n(c) \geq -(c + \varepsilon)^3/(3L\sigma^2) - \varepsilon$. This proves the theorem since $\varepsilon > 0$ is arbitrarily small. □

3. Asymptotic minimaxity over Hölder classes. In this section we assume that $f \in \Sigma(\gamma, L, B)$, where

$$\Sigma(\gamma, L, B) = \big\{f: |f(t_1)| \le B; |f^{\lfloor\gamma\rfloor}(t_1) - f^{\lfloor\gamma\rfloor}(t_2)| \le L|t_1 - t_2|^{\gamma - \lfloor\gamma\rfloor}, t_1, t_2 \in R^1\big\},$$

with given positive constants γ, L and B. Here $\lfloor\gamma\rfloor$ denotes the greatest integer strictly less than γ. Note that the class of Lipschitz functions in the previous section $\Sigma(L) = \Sigma(1, L, \infty)$. For γ integer, $\gamma \ge 1$, we have $\lfloor\gamma\rfloor = \gamma - 1$, and the class $\Sigma(\gamma, L, B)$ is a set of bounded regression functions having $(\gamma - 1)$th Lipschitz derivative.

The definition of $\beta_n(c)$ in this case also must be modified:

$$\beta_n(c) = \inf_{\hat{f}_n} \sup_{f \in \Sigma(\gamma, L, B)} \frac{1}{n} \log P_f\big(|\hat{f}_n - f(0)| > c\big), \qquad 0 < c < B. \tag{10}$$

Let a function $\psi_*(t)$, $t \in R^1$, be a solution of the extremal problem

$$\psi(0) \to \max,$$

under the restrictions

$$\|\psi\|_2^2 \le 1 \quad \text{and} \quad \psi \in \Sigma(\gamma, 1, \infty),$$

where $\|\psi\|_2^2 = \int_{-\infty}^{+\infty} (\psi(t))^2\, dt$.

THEOREM 2. *For any* c, $0 < c < B$, *the minimax risk* (10) *satisfies*

$$\lim_{n\to\infty} \beta_n(c) = -(2\sigma^2)^{-1} L^{-1/\gamma} \left(\frac{c}{\psi_*(0)}\right)^{2+1/\gamma}.$$

PROOF. Introduce a kernel function $K(t) = \psi_*(t)/\int_{-\infty}^{+\infty} \psi_*(t)\, dt$ and a bandwidth $h = (c/(L\psi_*(0)))^{1/\gamma}$. The correctness of these definitions follows from the optimal recovery theory which we apply after Donoho (1994). Introduce the kernel estimator obtained from the observations Y_{in} in (1):

$$f_n^* = (hn)^{-1} \sum_{j=-\infty}^{+\infty} K\left(\frac{j}{hn}\right) Y_{jn}. \tag{11}$$

Similarly to Lemma 1, the following inequality holds uniformly in $f \in \Sigma(\gamma, L)$:

$$\begin{aligned} &P_f(|f_n^* - f(0)| > c) \\ &\quad \le P\left(\left|(hn)^{-1} \sum_j K\left(\frac{j}{hn}\right) \xi_{jn}\right| > c - \sup_{f \in \Sigma(\gamma, L)} b_n(f)\right), \end{aligned} \tag{12}$$

with the bias term

$$b_n(f) = (hn)^{-1} \sum_j K\left(\frac{j}{hn}\right) f\left(\frac{j}{n}\right) - f(0)$$
$$= (1 + o(1)) h^{-1} \int_{-\infty}^{\infty} K\left(\frac{t}{h}\right)(f(t) - f(0))\, dt,$$

where $o(1) \to 0$ uniformly in $f \in \Sigma(\gamma, L, B)$ as $n \to \infty$ since the functions in this class are bounded and uniformly continuous over any finite interval. The following renormalization relation is true as shown in Donoho (1994):

$$\sup_{f \in \Sigma(\gamma, L, B)} h^{-1} \int_{-\infty}^{\infty} K\left(\frac{t}{h}\right)(f(t) - f(0))\, dt$$
$$\leq \sup_{f \in \Sigma(\gamma, L, \infty)} h^{-1} \int_{-\infty}^{\infty} K\left(\frac{t}{h}\right)(f(t) - f(0))\, dt$$
$$= Lh^{\gamma} \sup_{f \in \Sigma(\gamma, 1, \infty)} \int_{-\infty}^{\infty} K(t)(f(t) - f(0))\, dt$$
$$= \frac{c}{\psi_*(0)} \sup_{f \in \Sigma(\gamma, 1\, \infty)} \int_{-\infty}^{\infty} K(t)(f(t) - f(0))\, dt = \frac{c}{\psi * (0)} b_{\gamma,1},$$

where

$$b_{\gamma,1} = \sup_{f \in \Sigma(\gamma, 1)} \int_{-\infty}^{\infty} K(t)(f(t) - f(0))\, dt.$$

The random variable $(hn)^{-1} \sum_j K(j/(hn)) \xi_{jn}$ is zero-mean Gaussian with variance $V_n = (1 + o(1))(hn)^{-1} \sigma^2 \|K\|_2^2$, where $o(1) \to 0$ as $n \to \infty$. Note that here $o(1)$ is independent of $f \in \Sigma(\gamma, L, B)$. As in Lemma 1, one gets from (12) that

$$\limsup_{n \to \infty} \frac{1}{n} \log P\left(\left|(hn)^{-1} \sum_j K\left(\frac{j}{hn}\right) \xi_{jn}\right| > c - \sup_{f \in \Sigma(\gamma, L, B)} b_n(f)\right)$$
$$\leq \limsup_{n \to \infty} \left[- \frac{\left(c - (c/\psi * (0)) b_{\gamma,1}\right)^2}{2nV_n} \right]$$
$$= - \frac{h\left(c - (c/\psi * (0)) b_{\gamma,1}\right)^2}{2\sigma^2 \|K\|_2^2}$$
$$= -(2\sigma^2)^{-1} \left(\frac{c}{L\psi_*(0)}\right)^{1/\gamma} c^2 \left(\|K\|_2^{-1}\left(1 - \frac{b_{\gamma,1}}{\psi_*(0)}\right)\right)^2.$$

It is proven in the optimal recovery theory that $\psi_*(0) = b_{\gamma,1} + \|K\|_2$. This relation implies that $1 - b_{\gamma,1}/\psi_*(0) = \|K\|_2/\psi_*(0)$. Finally, we obtain

$$\limsup_{n\to\infty} \frac{1}{n}\log P_f(|f_n^* - f(0)| > c) \leq -(2\sigma^2)^{-1}\left(\frac{c}{L\psi_*(0)}\right)^{1/\gamma}\left(\frac{c}{\psi_*(0)}\right)^2.$$

This proves the upper bound for $\beta_n(c)$.

To prove the lower bound, we have to revise Lemma 2. By analogy with this lemma, introduce a parametric subfamily of the regression functions

$$f(t,\theta) = \theta A\psi_*(t/T), \qquad |\theta| \leq c + \varepsilon,$$

where $\varepsilon > 0$ is arbitrarily small; $c + \varepsilon < B$. If we choose the constants $A = 1/\psi_*(0)$ and $T = ((c+\varepsilon)/(L\psi_*(0)))^{1/\gamma}$, then $f(t,\theta) \in \Sigma(\gamma, L, B)$ for any $\theta: |\theta| < c + \varepsilon$. The rest of the proof is the same as that in Lemma 2 with the final constant

$$(13) \quad \liminf_{n\to\infty} \sup_{|\theta|<c+\varepsilon} \frac{1}{n}\log P_\theta(|\hat{\theta} - \theta| > c) \geq -\frac{(c+\varepsilon)^2}{2\sigma^2\|A\psi_*(t/T)\|_2^2} - \varepsilon.$$

From the definition of $\psi_*(t)$ one has

$$\|A\psi_*(t/T)\|_2^2 = A^2T = (1/\psi_*(0))^2((c+\varepsilon)/(L\psi_*(0)))^{1/\gamma}.$$

Thus the right-hand side of (13) is estimated from below by the constant

$$-(2\sigma^2)^{-1}L^{-1/\gamma}((c+\varepsilon)/\psi_*(0))^{2+1/\gamma} - \varepsilon.$$

The theorem is proved. □

EXAMPLE. In the case $0 < \gamma \leq 1$ the explicit solution for $\psi_*(t)$ is known [see Donoho (1994)]: $\psi_*(0) = ((2\gamma+1)(\gamma+1)/(4\gamma^2))^{\gamma/(2\gamma+1)}$. This expression turns the equality in Theorem 2 into

$$\lim_{n\to\infty} \beta_n(c) = -(2\sigma^2)^{-1}L^{-1/\gamma}c^{2+1/\gamma}(4\gamma)^2/((2\gamma+1)(\gamma+1)).$$

In particular, for $\gamma = 1$ we arrive immediately at the result of Theorem 1.

4. Asymptotic minimaxity for sup-norm loss. Assume that we want to estimate a regression function $f \in \Sigma(\gamma, L, B)$ from observations (1) in the sup-norm; that is, the measure of discrepancy of $\hat{f}_n - f$ is defined as

$$\|\hat{f}_n - f\|_\infty = \sup_{0\leq t\leq 1} |\hat{f}_n(t) - f(t)|.$$

In this case the minimax risk (2) must be substituted by

$$(14) \quad \beta_n^{(\infty)}(c) = \inf_{\hat{f}_n} \sup_{f\in\Sigma(\gamma,L,B)} \frac{1}{n}\log P_f(\|\hat{f}_n - f\|_\infty > c), \qquad 0 < c < B.$$

Note that the estimator $\hat{f}_n$ is obtained from all the observations Y_{in} in (1) but not from the observations in the interval $0 \leq t \leq 1$. This helps to avoid the edge effects, which is important since these effects influence dramatically the final result.

THEOREM 3. *For any* c, $0 < c < B$, *the minimax risk* (14) *satisfies*

$$\lim_{n\to\infty} \beta_n^{(\infty)}(c) = -(2\sigma^2)^{-1} L^{-1/\gamma}\left(\frac{c}{\psi_*(0)}\right)^{2+1/\gamma},$$

where $\psi_*(0)$ *is the same as in Theorem* 2.

PROOF. The lower bound for $\beta_n^{(\infty)}(c)$ follows from the lower bound at a fixed point since $\|\hat{f}_n - f_n\|_\infty \ge |\hat{f}_n(t_0) - f(t_0)|$ for any t_0, $0 \le t_0 \le 1$. To prove the upper bound, define the kernel estimator $f_n^*(t)$ at the points $t_k = k/n$, $k = 0,\dots,n$, by

$$f_n^* = (hn)^{-1} \sum_{j=-\infty}^{+\infty} K\left(\frac{j-k}{hn}\right) Y_{jn} \tag{15}$$

similarly to that in (11). Extend $f_n^*(t)$ to all the values of the continuous argument $t \in R^1$ as a piecewise constant function:

$$f_n^*(t) = f_n^*(k/n) \quad \text{if } (k-1)/n \le t < k/n.$$

Note that

$$P_f(\|f_n^* - f\|_\infty > c) \le \sum_{k=0}^{n} P_f\left(\left| f_n^*\left(\frac{k}{n}\right) - f\left(\frac{k}{n}\right)\right| > c - \frac{L}{n}\right).$$

Hence

$$\lim_{n\to\infty} \frac{1}{n} \log P_f\big(\|\hat{f}_n - f\|_\infty > c\big) \le \lim_{n\to\infty} \frac{1}{n} \log\big(nP_f\big(|\hat{f}_n(0) - f(0)| > c - L/n\big)\big)$$
$$= \lim_{n\to\infty} \frac{1}{n} \log P_f\big(|\hat{f}_n(0) - f(0)| > c\big).$$

Thus Theorem 2 applies and this proves the theorem. □

5. Discussion. If f is a constant observed in an additive $(0,\sigma^2)$-Gaussian noise, that is, if the observations are from the model $Y_i = f + \xi_i$, $i = 1,\dots,n$, ξ_i's are i.i.d. $(0,\sigma^2)$-Gaussian, then the Bahadur efficiency is determined by the constant $-c^2/(2\sigma^2)$ [Bahadur (1971) and Ibragimov and Khas'minskii (1981), Section 1.9]. Comparing this with the right-hand side of (4), we might consider $2cn/(3L)$ as an "equivalent number of observations" under the assumptions of Theorem 1. Under the assumptions of Theorem 2 this "equivalent number of observations" is $n(c/L)^{1/\gamma}(\psi_*(0))^{-2-1/\gamma}$. Note that this quantity loses its dependence on c and L as $\gamma \to \infty$.

Consider the sequences of estimators (5), (11) and (15) achieving their asymptotic minimax values in Sections 2, 3 and 4, respectively. Each of these sequences is inconsistent. Indeed, they are the kernel estimators with the finite bandwidths which are proportional to $c^{1/\gamma}$. Moreover, the traditional kernel estimators which are optimal with regard to, say, the minimax quadratic risk (3), turn out to be nonefficient in the sense of our minimax risks. If we study the estimating problem in the L_2-norm, then it can be shown that the large deviations of $\|\hat{f}_n - f\|_2$ are governed by the constant $-c^2/(2\sigma^2)$ as in the parametric case regardless of the value of γ.

In the non-Gaussian case when ξ_{in}'s in (1) have a "sufficiently good distribution" with the Fisher information I_ξ, for $\gamma = 1$ the following equality is true:

$$\lim_{c \to 0} \lim_{n \to \infty} c^{-3}\beta_n(c) = -I_\xi/(3L),$$

as shown recently in Korostelev and Leonov (1995).

REFERENCES

BAHADUR, R. R. (1960). On the asymptotic efficiency of tests and estimators. *Sankhyā Ser. A* **22** 229–252.

BAHADUR, R. R. (1971). *Some Limit Theorems in Statistics*. SIAM, Philadelphia.

DONOHO, D. L. (1994). Asymptotic minimax risk for sup-norm loss: solution via optimal recovery. *Probab. Theory Related Fields* **99** 145–170.

FU, J. C. (1982). Large sample point estimation: a large deviation theory approach. *Ann. Statist.* **10** 762–771.

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

KOROSTELEV, A. P. and LEONOV, S. L. (1995). Minimax Bahadur efficiency for small confidence intervals. *Problems Inform. Transmission*. To appear.

MICCHELLI, C. A. and RIVLIN, T. J. (1977). A survey of optimal recovery. In *Optimal Estimation in Approximation Theory* (C. A. Micchelli and T. J. Rivlin, eds.) 1–54. Plenum, New York.

SIEVERS, G. L. (1978). Estimation of location: a large deviation comparison. *Ann. Statist.* **6** 610–618.

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

DEPARTMENT OF MATHEMATICS
WAYNE STATE UNIVERSITY
DETROIT, MICHIGAN 48202
E-MAIL: apk@math.wayne.edu