The Annals of Statistics
1991, Vol. 19, No. 2, 633–667

GEOMETRIZING RATES OF CONVERGENCE, II[1]

BY DAVID L. DONOHO AND RICHARD C. LIU

University of California, Berkeley

Consider estimating a functional $T(F)$ of an unknown distribution $F \in \mathbf{F}$ from data $X_1, \ldots, X_n$ i.i.d. F. Let $\omega(\varepsilon)$ denote the modulus of continuity of the functional T over $\mathbf{F}$, computed with respect to Hellinger distance. For well-behaved loss functions $l(t)$, we show that $\inf_{T_n} \sup_{\mathbf{F}} E_F l(T_n - T(F))$ is equivalent to $l(\omega(n^{-1/2}))$ to within constants, whenever T is linear and $\mathbf{F}$ is convex. The same conclusion holds in three nonlinear cases: estimating the rate of decay of a density, estimating the mode and robust nonparametric regression.

We study the difficulty of testing between the composite, infinite dimensional hypotheses H_0: $T(F) \leq t$ and H_1: $T(F) \geq t + \Delta$. Our results hold, in the cases studied, because the difficulty of the full infinite-dimensional composite testing problem is comparable to the difficulty of the hardest simple two-point testing subproblem.

1. Introduction. Let $T(F)$ be a functional of an unknown distribution F and let $X_1, \ldots, X_n$ be i.i.d. F. As in Donoho and Liu (1987, 1988) (hereafter [GR I] and [GR III]), we are interested in estimating $T(F)$. For example, $T(F)$ might be the linear functional $f(0)$, the density of F at zero, or the nonlinear functional $\int f^2$, the squared L_2-norm of the density f. Such functionals arise in nonparametric estimation and have the general property that they cannot usually be estimated at a root-n rate. In fact if all that is known is that $F \in \mathbf{F}$, where $\mathbf{F}$ is a given class of smooth densities, it may turn out that no estimator $T_n = T_n(X_1, \ldots, X_n)$ can converge to $T(F)$ at rate faster than $n^{-r/2}$ for some $r < 1$.

In [GR I], this phenomenon was discussed and a new way of establishing it was introduced. Given the modulus of continuity of T over the class $\mathbf{F}$ with respect to Hellinger distance.

$$\omega(\varepsilon) = \sup\{|T(F_1) - T(F_0)| : H(F_1, F_0) \leq \varepsilon, F_i \in \mathbf{F}\}, \tag{1.1}$$

it was shown that no estimator can converge to $T(F)$ faster than $\omega(n^{-1/2})$ uniformly over $\mathbf{F}$. This bound is valid for all functionals and it was shown in [GR I] that the bound is at least as strong as rate bounds due to Farrell, Stone and Has'minskii.

Received April 1988; revised May 1990.

[1]Supported by NSF Grant DMS-84-51753 and by donations from Schlumberger Computer Aided Systems and from Sun Microsystems, Inc.

AMS 1980 *subject classifications*. Primary 62G20; secondary 62G05, 62F35.

Key words and phrases. Density estimation, estimating the mode, estimating the rate of tail decay, robust nonparametric regression, modulus of continuity, Hellinger distance, minimax tests, monotone likelihood ratio.

In this paper we discuss the *attainability* of this bound as regards *rate*. Since the $\omega(n^{-1/2})$ bound subsumes several existing nonparametric, parametric and semiparametric bounds, we know, of course, from the extensive work on nonparametrics [e.g., Farrell (1972), Wahba (1975), Stone (1980), ...] that the bound is often attainable. We show in this paper that for *linear* functionals, the rate is attainable in great generality.

SOME TERMINOLOGY. The loss function $l(t)$ is *well-behaved* if it is a symmetric increasing function of $|t|$ and if $l(\frac{3}{2}t) \le al(t)$ for all t. Thus t^2 and $|t|$ are well-behaved, with $a = \frac{9}{4}$ and $a = \frac{3}{2}$, respectively. We write $f(n) \asymp g(n)$ if the ratio of the two terms is bounded away from zero and infinity as $n \to \infty$. Combining Theorems 2.1, 2.4 and 3.1, we get:

COROLLARY. *Let T be linear and $\mathbf{F}$ be convex. If T is bounded on $\mathbf{F}$, so that*

$$\sup_{F \in \mathbf{F}} |T(F)| < \infty,$$

and if $\omega(\varepsilon)$ is Hölderian with exponent r, so that

$$\omega(\varepsilon) = C\varepsilon^r + o(\varepsilon^r),$$

then the optimal rate of convergence is $\omega(n^{-1/2})$:

$$\inf_{T_n} \sup_{\mathbf{F}} E_F l(T_n - T) \asymp l\big(\omega(n^{-1/2})\big)$$

for any well-behaved loss function l.

Thus, for linear functionals—the density at a point, the derivative of a density at a point, the density of a convolution factor at a point—the optimal rate of convergence is $r/2$, where r is the exponent in the modulus of continuity. In short, the rate of convergence (a statistical quantity) is determined by the modulus of continuity—a quantity deriving from the geometry of the graph of T over the regularity class $\mathbf{F}$.

We establish this result by directing attention *away* from the modulus of continuity and focusing instead on (another) new bound on the rate of convergence. In Section 2 we derive a new bound from a measure of the difficulty of testing the composite hypothesis H_0: $T(F) \le t$ against the composite hypothesis H_1: $T(F) \ge t + \Delta$. While in general, this new bound is much more difficult to compute than the modulus bound, it appears to be the right thing to be computed. Indeed, under a certain hypothesis on the asymptotic behavior of the new bound [see (2.9)], it is *always attainable* to within constant factors, whatever be the functional, linear or nonlinear. This is, to our knowledge, the first lower bound on estimation of general functionals which comes equipped with a (near-) attainability result.

In Section 3, we show that, in the linear T, convex $\mathbf{F}$ case, the modulus bound and the new bound agree to within constants. When the modulus is Holderian, the hypothesis (2.9) holds and so the attainability of the new bound to within constants implies that of the modulus bound to within constants.

In Section 4, we show that the modulus bound and the new bound are equivalent if and only if a certain minimax identity holds, at least approximately. That is, the testing difficulty of the hardest simple subproblem H_0: F_0 versus H_1: F_1, with $T(F_0) \le t$ and $T(F_1) \ge t + \Delta$, should be roughly the same as the difficulty of the full composite problem H_0: $T(F) \le t$ versus H_1: $T(F) \ge t + \Delta$. Thus, the modulus bound works in the case of T linear, $\mathbf{F}$ convex, because the difficulty of the hardest 2-point subproblem is comparable to the difficulty of the full problem.

In Section 5 we discuss some examples of nonlinear functionals. The first is the rate of tail decay [Du Mouchel (1983), Hall and Welsh (1984)]. For this functional, the minimax identity holds precisely. Actually, in this case, the minimax test of H_0: $T(F) \le t$ versus H_1: $T(F) > t + \Delta$ can be worked out in detail; it turns out to have a certain monotonicity in t within shows that the new bound can be attained to within a factor 4. In the second example, robust nonparametric regression, a combination of the minimax identity, translation invariance and reflection invariance, show that the new bound can be attained within a factor 2. In the third example, estimating the mode, the minimax identity does not hold, but the hardest 2-point subproblem has a difficulty that is again comparable to the full problem, and so the modulus is again attainable.

One should not always suppose the modulus bound to be attainable in the nonlinear setting. As one can infer from recent results of Ritov and Bickel (1990) and, in a related problem, of Ibragimov, Nemirovskii and Has'minskii (1987), attainability of the modulus bound can fail already for quadratic functionals. See Section 6.

An interesting feature of our approach is the use of notation and techniques due to Le Cam (1973, 1975, 1986) and Birgé (1983). In brief, the idea is that the difficulty of an estimation problem ought to be determined by the difficulty of a corresponding testing problem. As Le Cam has shown how to bound the difficulty of certain testing problems in terms of Hellinger affinity and has developed certain useful tools for computing Hellinger affinity, his machinery is well-suited for our work, which seeks to relate the Hellinger modulus to the difficulty of certain tests. In particular, Le Cam's little-known result, given as Lemma 3.4, is fundamental. Also, a technique of Birgé (1983) allows us to translate exponential bounds on testing errors [such as (2.10)] into bounds on expectations of well-behaved loss functions [such as (2.12)]. Finally, we utilize a connection between estimates and families of tests with a certain monotonicity property [see Huber (1982)] to develop estimates of certain nonlinear functionals.

These results should be compared with those of Birgé (1983). He found that for the problem of estimating the entire density (and not just a single functional of it), the geometry of the problem, expressed in terms of certain dimension numbers, determines the optimal rate. In this paper, we show that for estimating a linear functional, the geometry, expressed in terms of the modulus of continuity, determines the optimal rate. We note that the problem of recovering the entire density is like recovering a whole collection of linear

functionals and so is in some sense a linear problem. Thus, our work and Birgé's both say that for *linear problems the optimal rate derives from the geometry of the problem*.

2. An attainable bound. As in Section 1, let T be a functional of interest and let $\mathbf{F}$ be the regularity class in which F is known to lie. Let $\mathbf{F}_{\leq t}$ and $\mathbf{F}_{\geq t+\Delta}$ denote the subsets of $\mathbf{F}$, where T takes values less than or equal to t and greater than or equal to $t + \Delta$, respectively. Let $\mathbf{F}_{\leq t}^{(n)}$ denote the set of product measures of $X_1, \ldots, X_n$ i.i.d. F, $F \in \mathbf{F}_{\leq t}$, and similarly for $\mathbf{F}_{\geq t+\Delta}^{(n)}$. Denote by $\text{conv}(\mathbf{F}_{\leq t}^{(n)})$ the set of all measures on R^n which can be gotten as convex combinations of the product measures in $\mathbf{F}_{\leq t}^{(n)}$. Such a measure corresponds to the following: a random device is used to select an element $F \in \mathbf{F}_{\leq t}$, and then n observations are taken from this realized F. In words, $\text{conv}(\mathbf{F}_{\leq t}^{(n)})$ represents all the joint distribution of data $X_1, \ldots, X_n$ which can be obtained by Bayesians under a scheme in which $(X_1, \ldots, X_n)$ and F are random, with $X_1, \ldots, X_n$ conditionally i.i.d. F, and where F is a random element taking values in $\mathbf{F}_{\leq t}$.

Let P and Q be probability distributions on a common space. Then the *testing affinity* [Le Cam (1973, 1986)] is

$$\pi(P, Q) = \inf_{\substack{0 \leq \phi \leq 1 \\ \phi\text{-measurable}}} E_P\phi + E_Q(1 - \phi); \tag{2.1}$$

it is the minimal sum of type I and type II errors of any test between P and Q, and a natural measure of the difficulty of distinguishing P from Q. If $\mathbf{P}$ and $\mathbf{Q}$ are sets of measures, let $\pi(\mathbf{P}, \mathbf{Q})$ denote the largest testing affinity $\pi(P, Q)$ between any pair (P, Q) with $P \in \mathbf{P}$ and $Q \in \mathbf{Q}$—the difficulty of the hardest two-point testing subproblem. Symbolically,

$$\pi(\mathbf{P}, \mathbf{Q}) = \sup_{\substack{P \in \mathbf{P} \\ Q \in \mathbf{Q}}} \inf_{\substack{0 \leq \phi \leq 1 \\ \phi\text{-measurable}}} E_P\phi + E_Q(1 - \phi).$$

This is not, however, a measure of the difficulty of distinguishing $\mathbf{P}$ from $\mathbf{Q}$. We note, following Le Cam (1973, 1986) that if we view $\mathbf{P}$ and $\mathbf{Q}$ as composite hypotheses, the best sum of the two types of errors,

$$\inf_{\substack{0 \leq \phi \leq 1 \\ \phi\text{-measurable}}} \sup_{\substack{P \in \mathbf{P} \\ Q \in \mathbf{Q}}} E_P\phi + E_Q(1 - \phi)$$

is $\pi(\text{conv}(\mathbf{P}), \text{conv}(\mathbf{Q}))$. Unless $\mathbf{P}$ and $\mathbf{Q}$ are convex, this minimax difficulty is usually unequal to the difficulty $\pi(\mathbf{P}, \mathbf{Q})$ of the hardest two-point problem. We note that $\pi(P, Q) = 1 - \frac{1}{2}L_1(P, Q)$, where $L_1(P, Q) = \int |dP - dQ|$ denotes the L_1 distance, so computing the minimax risk amounts to finding the L_1 distance between the convex hulls of $\mathbf{P}$ and $\mathbf{Q}$. Note that $0 \leq \pi \leq 1$.

2.1. *The lower bound.* Our two main definitions are as follows. The *upper affinity* $\alpha_A(n, \Delta)$ of the estimation problem is

$$\alpha_A(n, \Delta) = \sup_t \pi\big(\mathrm{conv}(\mathbf{F}^{(n)}_{\le t}), \mathrm{conv}(\mathbf{F}^{(n)}_{\ge t+\Delta})\big). \tag{2.2}$$

This is the minimax risk of the hardest problem of distinguishing H_0: $\mathbf{F}_{\le t}$ and H_1: $\mathbf{F}_{\ge t+\Delta}$ at sample size n. Next, we let $\Delta_A(n, \alpha)$ be the function inverse to α_A:

$$\Delta_A(n, \alpha) = \sup\{\Delta: \alpha_A(n, \Delta) \ge \alpha\}.$$

In words, $\Delta_A(n, \alpha)$ measures the largest Δ at which, in a sample of size n, one cannot test hypotheses H_0: $\mathbf{F}_{\le t}$ and H_1: $\mathbf{F}_{\ge t+\Delta}$ with sum of errors less than α for all t. Δ_A places certain limits on how well T can be estimated. Essentially, this is because any estimator T_n of T gives rise to a test: decide H_0 if $T_n \le T + \Delta/2$, decide H_1 if $T_n > T + \Delta/2$.

THEOREM 2.1 (lower bound).

$$\inf_{T_n} \sup_{\mathbf{F}} P_F\{|T_n - T(F)| \ge \Delta_A(n, \alpha)/2\} \ge \alpha/2. \tag{2.3}$$

PROOF. Without loss of generality, let the supremum over t in the definition of α_A be attained, at t_0, and the supremum over Δ in the definition of Δ_A be attained; otherwise an ε_1 and ε_2 would have to be added in several places below and later picked arbitrarily close to zero.

The minimax difficulty for testing between H_0: $\mathbf{F}_{\le t_0}$ and H_1: $\mathbf{F}_{\ge t_0+\Delta}$ is α. It follows that for any test statistic,

$$\alpha \le \sup_{\substack{F_0 \in \mathbf{F}_{\le t_0} \\ F_1 \in \mathbf{F}_{\ge t_0+\Delta}}} P_{F_0}\{\text{reject } H_0\} + P_{F_1}\{\text{accept } H_0\},$$

so

$$\alpha/2 \le \sup_{\substack{F_0 \in \mathbf{F}_{\le t_0} \\ F_1 \in \mathbf{F}_{\ge t_0+\Delta}}} \max\big(P_{F_0}\{\text{reject } H_0\}, P_{F_1}\{\text{accept } H_0\}\big).$$

This implies that the test mentioned earlier, based on T_n, has at least the indicated maximum of type I and type II errors. Now

$$\begin{aligned} P_{F_0}\{|T_n - T(F)| \ge \Delta/2\} &\ge P_{F_0}\{T_n - T(F) > \Delta/2\} \\ &= P_{F_0}\{\text{reject } H_0\} \end{aligned}$$

and similarly

$$P_{F_1}\{|T_n - T(F)| \ge \Delta/2\} \ge P_{F_1}\{\text{accept } H_0\}.$$

Combining the last 3 displays gives

$$\sup_{\substack{F_0 \in \mathbf{F}_{\le t_0} \\ F_1 \in \mathbf{F}_{\ge t_0 + \Delta}}} \max_{F_0, F_1} P_{F_i}\{|T_n - T(F)| \ge \Delta/2\} \ge \alpha/2 \tag{2.4}$$

as $F_0, F_1 \in \mathbf{F}$ and T_n was arbitrary, (2.3) is proved. □

COROLLARY. *Fix α in $(0, 1)$. $\Delta_A(n, \alpha)$ is a bound on the rate of convergence: For any symmetric increasing loss function $l(t)$,*

$$\inf_{T_n} \sup_{\mathbf{F}} E_F l(T_n - T(F)) \ge l(\Delta_A(n, \alpha)/2)\alpha/2 \tag{2.5}$$

for all n.

The reader should note that Δ_A is (nearly) the *best* lower bound derivable by a testing argument. Indeed, for each t, Δ and n, there exists a test between $\mathbf{F}_{\le t}$ and $\mathbf{F}_{\ge t+\Delta}$ which attains the lower bound (2.4) within a factor 2. Thus the key inequality (2.4) cannot be improved by more than a factor 2. In the form that we have stated here, the lower bound is original. However, there is some relation with a bound on the size of confidence sets, due to Meyer (1977). The only examples the authors know of where an attempt is made to calculate something resembling this bound are Hall and Marron (1987) and Ritov and Bickel (1990). In both examples, the authors are attempting to lower bound an estimation error by the Bayes risk in testing between highly composite finite hypotheses. While they do not explicitly define any of the quantities we will deal with in this paper, a sympathetic reader may agree that their efforts are in the same direction.

2.2. *An estimator derived from minimax tests.* It is reasonable to guess that because the bound (2.3) cannot be substantially improved by a testing argument, it might be nearly attainable. Let us consider, then, constructing an estimator using the minimax tests which come close to attaining the key inequality (2.4). This estimator is not intended to be implemented on a computer, say; but its finite, concrete character allows us to demonstrate that the bound $\Delta_A(n, \alpha)$ can be (near) attained in great generality.

THE BINARY SEARCH ESTIMATOR. The estimator we propose requires that T be bounded on $\mathbf{F}$: $M = \sup_{\mathbf{F}}|T(F)| < \infty$. The estimator has a tuning constant Δ, which will depend in a prescribed way on sample size. At a given sample size n, Δ is fixed and we proceed as follows. Let $N = N(M, \Delta)$ be the smallest integer such that $(\frac{3}{2})^N\Delta > 2M$. Let $l_N = -(\frac{3}{2})^N\Delta/2$ and $h_N = +(\frac{3}{2})^N\Delta/2$. Then the interval $[l_N, h_N]$ contains $[-M, M]$. At this point we proceed as follows. Given data $X_1, \ldots, X_n$, we perform a minimax test between the upper third of $[l_N, h_N]$ and the lower third, that is, we test $\mathbf{F}_{\le -M/3}$ against $\mathbf{F}_{\ge M/3}$. We then form a new interval $[l_{N-1}, h_{N-1}]$ by deleting from the current one whichever third (upper or lower) is rejected by the test. After testing the lower third of the new interval $[l_{N-1}, h_{N-1}]$ against the upper third, we form the

interval $[l_{N-2}, h_{N-2}]$ by deleting from $[l_{N-1}, h_{N-1}]$ whichever third was rejected. Continuing in this way, we get a sequence of intervals, each one $\frac{2}{3}$ as long as the previous one; we arrive after N stages at an interval $[l_0, h_0]$ of length Δ and we pick as our estimate T_n the midpoint of this interval. The key result about this procedure is:

LEMMA 2.2. *Apply the binary search estimator with parameters* Δ, M *and* N. *Set* $\eta_0 = \Delta/2$ *and* $\eta_k = (\frac{3}{2})^k\Delta$ *for* $k \geq 1$. *Set* $d_k = \frac{1}{2}(\frac{3}{2})^k\Delta$ *for* $k = 0, 1, \ldots$. *Then*

$$\sup_{\mathbf{F}} P_F\{|T_n - T(F)| > \eta_k\} \leq \sum_{i=k}^{N-1} \alpha_A(n, d_i). \tag{2.6}$$

In particular,

$$P_F\{|T_n - T(F)| > \Delta/2\} \leq \sum_{k=0}^{N-1} \alpha_A\left(n, \tfrac{1}{2}\left(\tfrac{3}{2}\right)^k\Delta\right),$$

which makes an interesting comparison with (2.3). Later we will see that under some conditions, the terms in the sum decrease rapidly with k and this upper bound is comparable with the lower bound (2.3).

PROOF. We first give a formal description of the algorithm.

Algorithm Estimate (Δ, N):
$k := N$
$l_N := -\frac{1}{2}(\frac{3}{2})^N\Delta$
$h_N := \frac{1}{2}(\frac{3}{2})^N\Delta$
while $k > 0$ **do**
  $a_k := l_k + \frac{1}{3}(h_k - l_k)$
  $b_k := l_k + \frac{2}{3}(h_k - l_k)$
  Test H_0: $\mathbf{F}_{\leq a_k}$ *against* H_1: $\mathbf{F}_{\geq b_k}$
  if *Accept* H_0, **then** /* new interval is (l, b) */
    $l_{k-1} := l_k$; $h_{k-1} := b_k$
  if *Reject* H_0, **then** /* new interval is (a, h) */
    $l_{k-1} := a_k$; $h_{k-1} := h_k$
  $k := k - 1$
end while
$T_n = (l_0 + h_0)/2$
end Algorithm

Suppose that in place of the *test* step in the algorithm, we could substitute an oracle that always answered correctly. Running such an ideal algorithm would produce sequences $\{(l_k^*, h_k^*), k = 0, \ldots, N\}$ and $\{(a_k^*, b_k^*), k = 1, \ldots, N\}$, all functionals of F.

Consider now the tests $\xi_1, \ldots, \xi_N$, with ξ_k minimax for testing

$$H_0: \mathbf{F}_{\le a_k^*} \quad \text{versus} \quad H_1: \mathbf{F}_{\ge b_k^*}.$$

When $T(F)$ belongs to the middle third of the interval $[l_k, h_k]$, the test by definition decides correctly, because $T(F)$ will certainly be included in the interval $[l_{k-1}, h_{k-1}]$. When $T(F)$ does not lie in the middle third of the interval $[l_k, h_k]$, the probability that ξ_k decides incorrectly is bounded above by

$$\pi\left(\operatorname{conv}\left(\mathbf{F}^{(n)}_{\le a_k^*}\right), \operatorname{conv}\left(\mathbf{F}^{(n)}_{\ge b_k^*}\right)\right) \le \alpha_A(n, b_k^* - a_k^*) \equiv \alpha_A(n, d_{k-1}). \tag{2.7}$$

Consider now (2.6), and let $k > 0$. If the tests ξ_i all decide correctly for $i = k+1, \ldots, N$, then $T_n \in (l_k^*, h_k^*)$ and so $|T_n - T(F)| \le h_k^* - l_k^* \equiv \eta_k$. Therefore,

$$\begin{aligned} P\{|T_n - T(F)| > \eta_k\} &\le P\left(\bigcup_{i=k+1}^{N} \{\xi_i \text{ decides incorrectly}\}\right) \\ &\le \sum_{i=k+1}^{N} P\{\xi_i \text{ decides incorrectly}\} \\ &\le \sum_{i=k}^{N-1} \alpha_A(n, d_i); \end{aligned} \tag{2.8}$$

the last step uses (2.7). The argument in the case $k = 0$ is similar. □

While the sum $\sum_{i=k}^{N-1} \alpha_A(n, d_i)$ may look difficult to work with, a simple hypothesis on $\Delta_A(n, \alpha)$ affords a useful bound.

THEOREM 2.3. *Let $\alpha \in (0,1)$ be fixed. Suppose there exist $r > 0$ and $0 < A_0 \le A_1 < \infty$ so that*

$$A_0\left(\frac{|\log \alpha|}{n}\right)^{r/2} \le \Delta_A(n, \alpha) \le A_1\left(\frac{|\log \alpha|}{n}\right)^{r/2} \tag{2.9}$$

for $|\log \alpha|/n < \varepsilon_0$. Pick n_0 so that

$$|\log \alpha|/n_0 < \varepsilon_0 \quad \text{and} \quad \alpha_0 \equiv \alpha_A\left(n_0, A_0^2/A_1\varepsilon_0^{r/2}\right) < 1.$$

Define $\beta \equiv \frac{1}{12}|\log(\alpha(2-\alpha))|(A_0/A_1)^{r/2}$ and $\gamma = (1/4n_0)|\log(\alpha_0(2-\alpha_0))|$. Pick C so large that $CA_0 > 2A_1$. Then with $\Delta \equiv C\Delta_A(n, \alpha)$ and d_i as in Lemma 2.2,

$$\sum_{i=0}^{N-1} \alpha_A(n, d_i) \le \frac{2\theta}{1-\theta^2} + e_n, \tag{2.10a}$$

$$\sum_{i=k}^{N-1} \alpha_A(n, d_i) \le \frac{\theta^{2k}}{1-\theta^2} + e_n \tag{2.10b}$$

for $n > 2n_0$, *where*

$$\theta \equiv \exp(-C^{2/r}\beta)$$

$$e_n \equiv \frac{\log(3M)}{\log(A_0\varepsilon_0^{r/2})}\exp(-n\gamma). \tag{2.11}$$

The proof is given in Section 7. In view of (2.6), these bounds imply that for the binary search estimator with parameters Δ, M and N, we can have $P\{|T_n - T| > K\Delta_A(n, \alpha)\}$ as near zero as we like, by choosing C large and K still larger. Thus $\Delta_A(n, \alpha)$ is the optimal rate of convergence [compare (2.5)]. A more precise statement is possible for well-behaved loss functions (recall the definition in the introduction).

THEOREM 2.4. *Suppose that* $l(t)$ *is well-behaved with constant* a *and that* (2.9) *holds. Pick* C *so large that* $\theta^2 a < 1$. *Then for the binary search estimator with parameter* $\Delta = C\Delta_A(n, \alpha)$, *we have*

$$E_F l(T_n - T) \le Al(\Delta_A(n, \alpha)) \qquad n > n_1 \tag{2.12}$$

for every $F \in \mathbf{F}$, *where*

$$A = \frac{2\theta a}{1 - \theta^2}\left(2 + \frac{\theta a}{1 - \theta^2 a}\right)a^{\lceil \log C \log 1.5 \rceil}. \tag{2.13}$$

The proof is in Section 7. Combining (2.12) with (2.5) gives:

COROLLARY. *Under the assumptions of Theorem* 2.4,

$$\inf_{T_n} \sup_{\mathbf{F}} E_F l(T_n - T) \asymp l(\Delta_A(n, \alpha)).$$

In words, the minimax risk has the same asymptotic behavior as $l(\Delta_A(n, \alpha))$, to within constants.

This use of minimax tests to construct estimators is inspired by work of Le Cam (1973, 1975, 1986) and by Birgé (1983). The Birgé–Le Cam approach was developed for the problem of estimating an entire density, not just a single functional of it. It is based on covering the space $\mathbf{F}$ by Hellinger balls and then testing between balls to see in which ball the true F lies. Our approach differs, in that we are testing between level sets of the functionals in question. As far as the authors can see, testing between balls could not give the results we are looking for.

3. Attainability and linearity. The reader may suppose, rightly, that Δ_A is not easy to calculate. In the important special case where T is linear, it may be bounded using the modulus of continuity, as we show in this section.

3.1. *The main result.*

THEOREM 3.1 *Suppose T is linear and $\mathbf{F}$ is convex. Fix $\varepsilon_0 \in (0,1)$ and $\alpha_0 \in (0,1)$. There exist universal constants c, C such that for $\alpha \le \alpha_0$ and $|\log \alpha|/n < \varepsilon_0$*

$$\omega\left(c\sqrt{\frac{|\log \alpha|}{n}}\right) \le \Delta_A(n,\alpha) \le \omega\left(C\sqrt{\frac{|\log \alpha|}{n}}\right). \tag{3.1}$$

We may take $C = \sqrt{2}$ and $c = \frac{1}{2}$ for α_0, ε_0 small enough.

If ω is Hölderian, (3.1) establishes assumption (2.9). Invoking now the corollaries of Theorems 2.1 and 2.4, we get the corollary cited in the introduction.

We should emphasize that an inequality of this sort should not be expected for every functional—the modulus bound is simply not attainable in general. The lower bound can always be established; it is the upper bound that may fail.

3.2. *The best two-point testing bound.* To clarify matters somewhat, let us introduce yet another lower bound on the rate of convergence. The two-point testing bound $\Delta_2(n,\alpha)$ is defined as follows. Let

$$\alpha_2(n,\Delta) = \sup_t \pi\left(\mathbf{F}^{(n)}_{\le t}, \mathbf{F}^{(n)}_{\ge t+\Delta}\right). \tag{3.2}$$

Note the *omission of the convex hull* operation in comparison with the definition (2.2) of α_A. Similarly, let $\Delta_2(n,\alpha)$ be the inverse function of α_2. We can also write

$$\Delta_2(n,\alpha) = \sup\left\{|T(F_1) - T(F_0)| : \pi\left(F_1^{(n)}, F_0^{(n)}\right) \ge \alpha\right\}. \tag{3.3}$$

This *is* a lower bound on the rate of convergence. Indeed, as $\alpha_2 \le \alpha_A$, we have

$$\Delta_2(n,\alpha) \le \Delta_A(n,\alpha); \tag{3.4}$$

as Δ_A has the lower bound property (2.3), it follows that Δ_2 is a lower bound as well. Thus, (2.3) holds with Δ_2 in place of Δ_A. One could also argue directly (compare Theorem 2.1 of [GR I]).

One can say more; Δ_2 is (nearly) the *best possible* two-point testing bound. Thus, for a given n and α, the largest δ for which there exists a pair (F_0, F_1) with $T(F_1) - T(F_0) \ge \delta$ which cannot be distinguished by the best test with sum of errors better than α, is precisely $\Delta_2(n,\alpha)$. No two-point bound on the maximum probability of error can exceed α, while this bound guarantees at least $\alpha/2$.

The two-point bound is closely related to the modulus. Indeed, we have:

LEMMA 3.2. *Fix* $\varepsilon_0 \in (0,1)$ *and* $\alpha_0 \in (0,1)$. *There exist constants* c *and* C *so that for* $\alpha < \alpha_0 < 1$ *and* $|\log \alpha|/n < \varepsilon_0$,

$$\omega\left(c\sqrt{\frac{|\log \alpha|}{n}}\right) \le \Delta_2(n,\alpha) \le \omega\left(C\sqrt{\frac{|\log \alpha|}{n}}\right). \tag{3.5}$$

We may take $C = \sqrt{2}$ *and*

$$c^2 = \frac{(1-e^{-\varepsilon_0})}{\varepsilon_0}\,\frac{\log(2-\alpha_0)\alpha_0}{\log \alpha_0}.$$

Before giving the proof, we need some facts about Hellinger distance. We use conventions similar to Beran (1977), Donoho and Liu (1988), Ibragimov and Has'minskii (1981) and Pitman (1979). These conventions enforce, in a few cases, different normalizations from Le Cam [(1973, 1986), Chapter 4]. First, recall the Hellinger affinity

$$\rho(P,Q) = \int \sqrt{p}\,\sqrt{q}\,d\mu, \tag{3.6}$$

where p and q denote densities with respect to a measure μ which dominates P and Q (e.g., $\mu = P + Q$). We have the inequalities

$$\pi(P,Q) \le \rho(P,Q), \qquad \rho^2 \le \pi(2-\pi), \tag{3.7}$$

where π is the testing affinity, and the identity

$$\rho(P,Q) = \tfrac{1}{2}(2 - H^2(P,Q)), \tag{3.8}$$

where H denotes Hellinger distance. We also have the elementary, but very useful, formula

$$\rho\big(P^{(n)},Q^{(n)}\big) = \rho(P,Q)^n, \tag{3.9}$$

where $P^{(n)}$ and $Q^{(n)}$ denote n-fold product measures with marginals P and Q. Armed with these, we can proceed.

PROOF. Define

$$h_0(n,\alpha) = \inf\big\{H(F_1,F_0)\colon \pi\big(F_1^{(n)},F_0^{(n)}\big) \le \alpha\big\}$$

and

$$h_1(n,\alpha) = \sup\big\{H(F_1,F_0)\colon \pi\big(F_1^{(n)},F_0^{(n)}\big) \ge \alpha\big\}.$$

Using (3.6)–(3.9), we have the easy inequalities

$$h_0^2(n,\alpha) \ge 2\big(1 - (\alpha(2-\alpha))^{1/2n}\big), \tag{3.10}$$

$$h_1^2(n,\alpha) \le 2(1-\alpha^{1/n}). \tag{3.11}$$

Combining these, for each $\delta > 0$,

$$\omega\big(h_0(n,\alpha) - \delta\big) \le \Delta_2(n,\alpha) \le \omega\big(h_1(n,\alpha) + \delta\big). \tag{3.12}$$

The result then follows by (3.13) and (3.14). □

LEMMA 3.3. *For* $\alpha \in (0,1)$,

$$(1 - \alpha^{1/n}) < \frac{|\log \alpha|}{n}. \tag{3.13}$$

Fix $\alpha_0 < 1$, $\varepsilon_0 > 0$. *There exists a finite positive constant* c *so that for* $0 < \alpha < \alpha_0$, $|\log \alpha|/n < \varepsilon_0$, *we have*

$$\left(1 - (\alpha(2-\alpha))^{1/2n}\right) > \frac{c^2}{2}\frac{|\log \alpha|}{n}. \tag{3.14}$$

We may take

$$c^2 = \frac{1 - e^{-\varepsilon_0}}{\varepsilon_0}\frac{\log(2-\alpha_0)\alpha_0}{\log \alpha_0}.$$

This result is proved in the technical report [GR II].

In particular, if $\omega(\varepsilon)$ is Hölderian, then $\omega(n^{-1/2})$ is equivalent, to within constants, with $\Delta_2(n,\alpha)$. And so the question of the attainability, as regards rate, of $\omega(n^{-1/2})$ is equivalent to the attainability of the best two-point, testing bound. Compare also Section 6 of [GR I].

The reader will note that (3.4)–(3.5) together establish the lower bound of (3.1)—without any hypotheses on T or $\mathbf{F}$.

3.3. *Establishing the upper bound.* Le Cam has established a fact which seems, at first, quite similar to (3.9) but is in fact far deeper.

LEMMA 3.4 [Le Cam (1986), Chapter 16, page 477]. *Let* $\mathbf{P}$ *and* $\mathbf{Q}$ *denote sets of probabilities and* $\mathbf{P}^{(n)}$, $\mathbf{Q}^{(n)}$ *the sets of corresponding product measures. Then*

$$\rho(\operatorname{conv}\mathbf{P}^{(n)}, \operatorname{conv}\mathbf{Q}^{(n)}) \le \rho(\operatorname{conv}\mathbf{P}, \operatorname{conv}\mathbf{Q})^n. \tag{3.15}$$

We remark that this is *not* an obvious consequence of the identity $\rho(P^{(n)}, Q^{(n)}) = \rho(P,Q)^n$. Combining (3.7), (3.15) and the definition of α_A, we have:

COROLLARY.

$$\alpha_A(n,\Delta) \le \sup_t \rho\left(\operatorname{conv}(\mathbf{F}_{\le t}), \operatorname{conv}(\mathbf{F}_{\ge t+\Delta})\right)^n. \tag{3.16}$$

Thus the Hellinger distance between the convex hulls of $\mathbf{F}_{\le t}$ and $\mathbf{F}_{\ge t+\Delta}$ may be used to bound α_A. The upper bound in (3.1) follows more or less directly from this. To see how, notice that for each $\delta > 0$,

$$\varepsilon \le \inf_t H\left(\mathbf{F}_{\le t}, \mathbf{F}_{\ge t+\omega(\varepsilon)+\delta}\right). \tag{3.17}$$

Combining this with (3.8) we have

$$\rho(\mathbf{F}_{\le t}, \mathbf{F}_{\ge t+\omega(\varepsilon)+\delta}) \le 1 - \varepsilon^2/2. \tag{3.18}$$

Now, and this is the key observation, if T is a linear functional and if $\mathbf{F}$ is convex, then $\mathbf{F}_{\le t}$ and $\mathbf{F}_{\ge t+\Delta}$ are *both convex* for all t and all Δ. Thus $\mathbf{F}_{\le t} = \operatorname{conv} \mathbf{F}_{\le t}$ and $\mathbf{F}_{\ge t+\omega(\varepsilon)+\delta} = \operatorname{conv} \mathbf{F}_{\ge t+\omega(\varepsilon)+\delta}$; combining (3.16) and (3.18),

$$\alpha_A(n, \omega(\varepsilon) + \delta) \le (1 - \varepsilon^2/2)^n. \tag{3.19}$$

It follows that

$$\Delta_A(n, \alpha) \le \omega\left(\sqrt{2\left(1 - (\alpha^*)^{1/n}\right)}\right) + \delta$$

for every $\alpha^* > \alpha$. Since (3.13) holds with strict inequality, we get the upper bound in (3.1). This completes the proof of Theorem 3.1.

4. Attainability and the minimax identity. In general, a relation such as (3.1) between $\omega(n^{-1/2})$ and $\Delta_A(n, \alpha)$ is not to be expected. It requires essentially that the hardest two-point subproblem of testing $\mathbf{F}_{\le t}$ versus $\mathbf{F}_{\ge t+\Delta}$ be roughly as hard as the full problem. Let us see how.

4.1. *The minimax identity.* The two-point testing bound and the attainable bound have an interesting connection. As (3.4) shows, the two-point bound is always smaller; as (3.1) and (3.5) make plain, when T is linear and $\mathbf{F}$ is convex,

$$\Delta_A(n, \alpha) \le C\Delta_2(n, \alpha) \tag{4.1}$$

for an appropriate constant C, for small α and large n.

It seems natural to ask if the two-point and the attainable bounds can ever agree, that is, if we can have $C = 1$ in (4.1). Chasing a few definitions, this leads in turn to the question of whether we can have

$$\pi\left(\operatorname{conv}\left(\mathbf{F}^{(n)}_{\le t}\right), \operatorname{conv}\left(\mathbf{F}^{(n)}_{\ge t+\Delta}\right)\right) = \pi\left(\mathbf{F}^{(n)}_{\le t}, \mathbf{F}^{(n)}_{\ge t+\Delta}\right), \tag{4.2}$$

Indeed, the quantity on the left-hand side is the main ingredient in the definition of Δ_A, while that on the left is the main ingredient in Δ_2. Now, from the definition of π, the quantity on the left is

$$\inf_{\substack{\text{tests} \\ \zeta}} \sup_{\substack{F_0 \in \mathbf{F}_{\le t} \\ F_1 \in F_{\ge t+\Delta}}} D_n(\zeta, (F_0, F_1)),$$

where $D_n(\zeta, (F_0, F_1))$ is the difficulty $E_{F_0^{(n)}}\zeta + E_{F_1^{(n)}}(1 - \zeta)$ representing the sum of errors of the test ζ. This is the minimax difficulty for the problem of testing the composite hypotheses $\mathbf{F}_{\le t}$ versus $\mathbf{F}_{\ge t+\Delta}$. On the other hand, the quantity on the right of (4.2) is

$$\sup_{\substack{F_0 \in \mathbf{F}_{\le t} \\ F_1 \in \mathbf{F}_{\ge t+\Delta}}} \inf_{\substack{\text{tests} \\ \zeta}} D_n(\zeta, (F_0, F_1)).$$

This is the difficulty of the hardest two-point testing problem. Consequently, the identity (4.2) is equivalent to the *minimax identity*

$$\inf_{\zeta} \sup_{(F_1, F_0)} D_n(\zeta, (F_0, F_1)) = \sup_{(F_1, F_0)} \inf_{\zeta} D_n(\zeta, (F_0, F_1)). \tag{4.3}$$

This identity says, in words, that the difficulty in testing between the infinite-dimensional composite hypotheses $\mathbf{F}_{\leq t}$ and $\mathbf{F}_{\geq t+\Delta}$ is precisely the difficulty of the hardest two-point testing problem.

We will see later two concrete examples where this minimax identity holds. For clarity, we summarize some implications the identity would have:

LEMMA 4.1. *If* (4.2) *holds for every* t *and* n *and all* $\Delta < \Delta_0$, *then* $\Delta_A = \Delta_2$ *for large,* n *and so* $\omega(n^{-1/2})$ *represents the optimal rate of convergence of an estimate* T_n *to* T.

Indeed, the conclusion that $\Delta_A = \Delta_2$ follows from the definition of these quantities and the conclusion that $\omega(n^{-1/2})$ is the optimal rate follows from (3.5) and Theorem 3.1.

It does happen that (4.2) holds in interesting examples. The following result is proved in the technical report [GR II].

THEOREM 4.2. *Let* $T(F) = f(0)$ *and let* $\mathbf{F}$ *be the Sacks–Ylvisaker* (1981) *class*

$$\mathbf{SY} = \Big\{ f\colon f(x) = f(0) + xf'(0) + h(x),\ f(0) \leq M,\ \int f = 1,\ f \geq 0, |h(x)| \leq x^2/2 \Big\}.$$

(*Here we must have* $(4\sqrt{2}/3)M^{3/2} < 1$.) *Then for every* t *and* n *and every* Δ *small enough, the minimax identity* (4.2) *holds and so* $\Delta_A = \Delta_2$ *for large* n.

For other examples, see Section 5. In general, one cannot expect (4.2) to hold. One case when (4.2) does hold is when the sets $\mathbf{F}_{\leq t}$ and $\mathbf{F}_{\geq t+\Delta}$ are generated by capacities [see Huber and Strassen (1973), Bednarski (1982)]. This is much stronger than simple convexity of the two sets. However, Le Cam's result, as recorded in Lemma 3.4, may be used to show that convexity *alone* is enough to guarantee that a certain *approximate* minimax identity holds.

LEMMA 4.3. *If* $\mathbf{F}_{\leq t}$ *and* $\mathbf{F}_{\geq t+\Delta}$ *are both convex,*

$$\rho\big(\mathrm{conv}(\mathbf{F}^{(n)}_{\leq t}), \mathrm{conv}(\mathbf{F}^{(n)}_{\geq t+\Delta})\big) = \rho(\mathbf{F}^{(n)}_{\leq t}, \mathbf{F}^{(n)}_{\geq t+\Delta}). \tag{4.4}$$

This says that, although (4.2) may not hold when just convexity is assumed, its analog, with π replaced by ρ, does hold.

PROOF. We have

$$\begin{aligned}\rho(\mathbf{F}_{\le t}, \mathbf{F}_{\ge t+\Delta})^n &\ge \rho\big(\mathrm{conv}(\mathbf{F}^{(n)}_{\le t}), \mathrm{conv}(\mathbf{F}^{(n)}_{\ge t+\Delta})\big) \\ &\ge \rho\big(\mathbf{F}^{(n)}_{\le t}, \mathbf{F}^{(n)}_{\ge t+\Delta}\big) \\ &= \rho(\mathbf{F}_{\le t}, \mathbf{F}_{\ge t+\Delta})^n,\end{aligned}$$

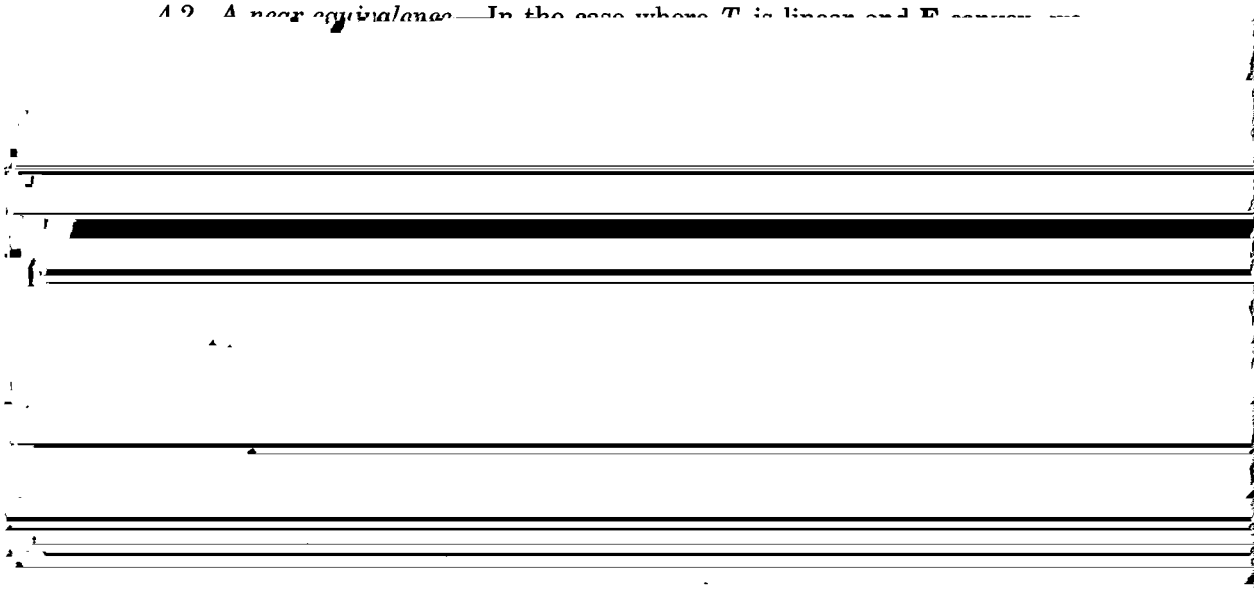
the first line following from Lemma 3.4 and the assumed convexity; the second from the obvious inclusion relation; and the third from the formula (3.9) for affinity of product measures. As the first and last quantities are the same, it follows that the middle inequality is actually an equality. Hence, (4.4). □

Because of the inequalities (3.7), (4.4) places definite limits on how different the two sides of (4.2) can be for large n. In fact, we get for the ratio of logarithms that

$$1 \le \frac{|\log \alpha_2(n,\Delta)|}{|\log \alpha_A(n,\Delta)|} \le 2\left(1 + \frac{1}{|\log \alpha_A(n,\Delta)|}\right). \tag{4.5}$$

Thus, at every n and Δ for which $\alpha_A(n,\Delta) \le \alpha_0 < 1$, we can bound the discrepancy between α_2 and α_A. In this sense, Le Cam's Lemma 3.4, which underlies (4.4), is an approximate minimax theorem. And one could say that Theorem 3.1 holds because (4.2) almost holds when T is linear and $\mathbf{F}$ is convex.

4.2. *A near equivalence.* In the case where T is linear and $\mathbf{F}$ convex, we



have seen that $\Delta_A \le C\Delta_2$ and also that $|\log \alpha_2| \le M|\log \alpha_A| + D$. In general, whatever be T and $\mathbf{F}$, these two accompany each other, so that if one holds, so does the other. This gives a clue to the general attainability issue; attainability of $\omega(n^{-1/2})$ really does imply that the two sides of (4.2) are close, but only in the sense that an inequality on logarithms such as (4.5) holds. We state a formal result, whose proof is sketched in Section 7.

THEOREM 4.4. *Suppose that $\omega(\varepsilon)$ is Hölderian with exponent $r \in (0,1]$. Then there are constants $\alpha_0 \in (0, \frac{1}{2})$ and $\varepsilon_0 \in (0,1)$ so that the following two statements are equivalent.*

(i) *There exists a finite positive M such that*

Has'minskii (1984); in our own work, this theme appears in the sequel [GR III], in Donoho, Liu and MacGibbon (1990) and in Donoho (1989).

5. Attainability in nonlinear cases. In this section we study three nonlinear functionals in order to see how the ideas of the preceding sections carry over.

5.1. *Estimating tail rates.* While it is most natural to consider estimating the rate at which the tail of a density approaches 0 as $x \to \infty$ [compare Du Mouchel (1983)], a transformation of the problem (to observations $Y_i = 1/X_i$) leads one to consider estimating the rate at which a density, known to be zero at the origin, approaches this limit as $x \to 0^+$ [compare Hall and Welsh (1984)]. We adopt this point of view here. Accordingly, let $\mathbf{F} = \mathbf{Tails}(C_-, C_+, \delta, t_0, t_1, \gamma, p)$, the set of distributions supported on $[0, \infty)$ with densities f satisfying

$$f(x) = Cx^t(1 + h(x)) \qquad 0 \le x \le \delta < 1, \tag{5.1}$$

with

$$0 < t_0 \le t \le t_1 < \infty \tag{5.2a}$$

and

$$0 < C_- \le C \le C_+ < \infty \tag{5.2b}$$

and

$$|h(x)| \le \gamma x^p. \tag{5.2c}$$

For such an $F \in \mathbf{F}$, let $T(F) = t$, where t is the exponent in (5.1). This functional is nonlinear.

Consider now the problem of testing $\mathbf{F}_{\le t}$ against $\mathbf{F}_{\ge t+\Delta}$. In [GR I] we have shown that the closest pair in a Hellinger sense has the form

$$f_0^*(x) = C_- x^t(1 - \gamma x^p), \qquad x \le a_1(t, \Delta), \tag{5.3}$$

$$f_1^*(x) = C_+ x^{t+\Delta}(1 + \gamma x^p), \qquad x \le a_1(t, \Delta), \tag{5.4}$$

and

$$\frac{f_0^*(x)}{f_1^*(x)} = \frac{f_0^*(a_1)}{f_1^*(a_1)}, \qquad x > a_1, \tag{5.5}$$

and

$$\frac{f_0^*(a_1)}{f_1^*(a_1)} = \frac{1 - \int_0^{a_1} f_0^*(v)\, dv}{1 - \int_0^{a_1} f_1^*(v)\, dv}. \tag{5.6}$$

As we will see, this closest Hellinger pair represents a hardest two-point testing problem. From the properties of this pair, we can show that the minimax identity (4.2) holds in this case.

THEOREM 5.1. *For the previously described pair* (F_0^*, F_1^*), *we have*

$$\pi\big(\mathrm{conv}\big(\mathbf{F}_{\le t}^{(n)}\big), \mathrm{conv}\big(\mathbf{F}_{\ge t+\Delta}^{(n)}\big)\big) = \pi\big(\mathbf{F}_{\le t}^{(n)}, \mathbf{F}_{\ge t+\Delta}^{(n)}\big) = \pi\big((F_0^*)^{(n)}, (F_1^*)^{(n)}\big)$$

and the minimax test between $\mathbf{F}_{\le t}$ *and* $\mathbf{F}_{\ge t+\Delta}$ *is the likelihood ratio test between* F_0^* *and* F_1^*.

PROOF. The likelihood ratio $L_{t,\Delta}(x) = f_1^*(x)/f_0^*(x)$ has, according to (5.3)–(5.5), the form

$$L_{t,\Delta}(x) = \begin{cases} \dfrac{C_+}{C_-} x^{\Delta} \dfrac{1+\gamma x^p}{1-\gamma x^p}, & 0 < x < a_1, \\ \dfrac{C_+}{C_-} a_1^{\Delta} \dfrac{1+\gamma a_1^p}{1-\gamma a_1^p}, & x \ge a_1. \end{cases} \tag{5.7}$$

This is a nondecreasing function of x.

Among all distributions in $\mathbf{F}_{\le t}$, F_0^* is the stochastically largest on $[0, a_1]$. Similarly, among all distributions in $\mathbf{F}_{\ge t+\Delta}$, F_1^* is the stochastically smallest on $[0, a_1]$. This implies that the distribution of $L_{t,\Delta}(X)$, where X is distributed F, is stochastically largest under the null hypothesis at $F = F_0^*$ and stochastically smallest under the alternative hypothesis at $F = F_1^*$. Now let $X_1, \ldots, X_n$ be i.i.d. F. Consider the likelihood ratio statistic

$$L_{n,t,\Delta} = \prod_{i=1}^{n} L_{t,\Delta}(X_i). \tag{5.8}$$

Under H_0: $F_{\le t}$ this statistic is then stochastically largest at $F = F_0^*$, and so on. Therefore, if we consider accepting H_0 when $L_{n,t,\Delta} \le 1$ and rejecting when $L_{n,t,\Delta} > 1$, we have

$$\sup_{F \in \mathbf{F}_{\le t}} P_F\{\text{Reject } H_0\} = P_{F_0^*}\{\text{Reject } H_0\},$$

$$\sup_{F \in \mathbf{F}_{\ge t+\Delta}} P_F\{\text{Accept } H_0\} = P_{F_1^*}\{\text{Accept } H_0\}.$$

It follows that the worst sum of type I and type II errors of our test occurs at (F_0^*, F_1^*). But the likelihood ratio test is optimal for that pair, and hence it is minimax. □

We now use the minimax identity to show that the lower bound of Theorem 2.1 may be nearly attained. An extra level of structure in the minimax tests of Section 2 may exist which we have not previously considered: *monotonicity in* t. Suppose that we are in a situation like the present one, where the minimax test can be taken to be nonrandomized. If $A(t, n, \Delta)$ is the region in which such a test accepts H_0: $\mathbf{F}_{\le t}$ rather than H_1: $\mathbf{F}_{\ge t+\Delta}$, we say that the acceptance regions are monotone in t if

$$A(t, n, \Delta) \subset A(t + h, n, \Delta) \qquad \forall\, h > 0. \tag{5.9}$$

The following result is proved in Section 7.

THEOREM 5.2. *For all sufficiently small* Δ, *the likelihood ratio* $L_{t,\Delta}(x)$ *is monotone decreasing in* t *for each fixed* x.

It follows from this theorem that the minimax test for our problem has acceptance region

$$A(t,n,\Delta) = \left\{(X_i)_{i=1}^n : \prod_{i=1}^n L_{t,\Delta}(X_i) \le 1\right\}$$

with the monotonicity property (5.9). Consider what we call the *likelihood ratio estimator*

$$T^*_{n,\Delta} = \frac{\Delta}{2} + \sup\left\{t : \prod_{i=1}^n L_{t,\Delta}(X_i) \ge 1, t \in [t_0, t_1 - \Delta]\right\}. \tag{5.10}$$

By the monotonicity established in Theorem 5.2, $T^*_{n,\Delta}$ is always uniquely defined.

THEOREM 5.3. *Suppose that* $L_{t,\Delta}(x)$ *is monotone decreasing in* t *for each fixed* x. *Then*

$$\sup_{\mathbf{F}} P_F\{|T^*_{n,\Delta} - T(F)| > \Delta/2\} \le 2\alpha_A(n,\Delta). \tag{5.11}$$

This is to be compared with the lower bound (2.3); it is parallel in form; but in the lower bound the 2α is replaced by $\alpha/2$.

PROOF. By the monotonicity in t of $L_{t,\Delta}$, $\{T^*_{n,\Delta} - T(F) > \Delta/2\}$ happens if and only if the minimax test between H_0: $\mathbf{F}_{\le T(F)}$ and H_1: $\mathbf{F}_{\ge T(F)+\Delta}$ would reject H_0. The probability of this event is smaller than $\alpha_A(n,\Delta)$ by definition. Similarly, the probability of the event $\{T^*_{n,\Delta} - T(F) < -\Delta/2\}$ is also less than $\alpha_A(n,\Delta)$. As $\{|T^*_{n,\Delta} - T(F)| > \Delta/2\}$ is the union of these two events, (5.11) follows. □

Thus, in this case, the lower bound Δ_A $(= \Delta_2)$ is achievable within a factor 4.

5.2. *Robust nonparametric regression.* Let (u_i, y_i), $i = 1, \ldots, n$ be our observations and suppose that

$$y_i = \xi(u_i) + z_i, \tag{5.12}$$

where $\xi(u)$ is an unknown function, known only to be Lipschitz with constant $\le C$, and the z_i are supposed to be independent errors, z_i having the distribution G_i which is unknown, but which is supposed to lie in the gross-errors neighborhood

$$\mathbf{G}_\varepsilon = \{G : G = (1-\varepsilon)\Phi + \varepsilon H\} \tag{5.13}$$

with Φ the standard normal distribution and ε a known constant, $0 \le \varepsilon < 1$. We are interested in estimating $\xi(u_0)$. If $\varepsilon = 0$, then $\xi(u) = E\{y|u\}$ is the

regression function; thus with $\varepsilon > 0$ we get a problem of *robust* nonparametric regression.

To fit this in our framework, suppose (for definiteness) that the u_i are i.i.d. according to Lebesgue measure on $[0, 1]$ and let $\mathbf{Reg}(C, \varepsilon)$ denote the class of distributions $F(u, y)$ on $\mathbf{R}^2$ such that the u marginal is uniform and

$$F(y - \xi(u)|u) \in \mathbf{G}_\varepsilon, \tag{5.14}$$

where $\xi \in \mathrm{Lip}(C)$.

It is now evident that the problem fits in the framework of this paper. We are trying to estimate the nonlinear functional $T(F) = \xi(u_0)$ from data $X_i = (u_i, y_i)$ which are i.i.d. F, with F an unknown element of $\mathbf{F} = \mathbf{Reg}(C, \varepsilon)$.

Our solution depends on results about minimax testing in the gross-errors model; see Huber [(1982), Chapter 10]. Define the Huber score function $\psi_{\Delta,\varepsilon}(x) = \min(k, \max(-k, x))$, where $k = k(\varepsilon, \Delta)$ is the solution to

$$e^{-\Delta k}\Phi\left(\frac{\Delta}{2-k}\right) - \Phi\left(\frac{-\Delta}{2-k}\right) = \frac{\varepsilon}{1-\varepsilon}.$$

Define the censored likelihood ratio

$$\lambda(x; \Delta, \varepsilon) = \exp\{\psi_{\Delta,\varepsilon}(x)\}.$$

Let $\mathbf{G}_0$ be the class of all distributions $G(\cdot + \Delta/2)$ with G in $\mathbf{G}_\varepsilon$, and $\mathbf{G}_1$ be the class of all distributions $G(\cdot - \Delta/2)$ with G in $\mathbf{G}_\varepsilon$. Suppose we wish to test H_0: $\mathbf{G}_0$ against H_1: $\mathbf{G}_1$ from n i.i.d. observations x_i. Then $\Pi_{i=1}^n \lambda(x_i; \Delta, \varepsilon)$ furnishes the minimax test. Another significant fact is that $\lambda(x - t)$ is monotone decreasing in t. Huber uses these to construct location estimates with a certain optimality property.

Now return to our model. Define $\Xi(u, \Delta) = \inf\{|\xi_1(u) - \xi_0(u)|\colon\ \xi_1(0) - \xi_0(0) \geq \Delta\}$. Then by the Lipschitz condition $\Xi(u, \Delta) = (\Delta - 2C|u|)_+$. Define

$$L_{t,\Delta}(u, y) = \lambda(y - t; \Xi(u - u_0), \varepsilon) \tag{5.15}$$

and

$$L_{n,t,\Delta} = \prod_{i=1}^n L_{t,\Delta}(X_i). \tag{5.16}$$

We can use this to test between H_0: $\mathbf{F}_{\leq t}$ and H_1: $\mathbf{F}_{\geq t+\Delta}$ as follows: Reject H_0 if $L_{n,t,\Delta} > 1$, accept if $L_{n,t,\Delta} < 1$ and randomize with equal probability otherwise. This test is, in fact, minimax: $L_{t,\Delta}(u, \cdot)$ is the likelihood ratio between a pair $(F_0^*(\cdot|u), F_1^*(\cdot|u))$ which makes $L_{t,\Delta}(u, \cdot)$ stochastically largest among all $F(\cdot|u)$ obeying $\xi(u_0) \leq t - \Delta/2$ and stochastically smallest among all $F(\cdot|u)$ obeying $\xi(u_0) \geq t + \Delta/2$.

Moreover, one sees that $L_{t,\Delta}(u, y)$ is decreasing in t for each fixed u, y and Δ. Let

$$T_{n,-} = \inf\{t\colon L_{n,t,\Delta} \leq 1\},$$

$$T_{n,+} = \sup\{t\colon L_{n,t,\Delta} \geq 1\}.$$

Finally, let $T^*_{n,\Delta}$ be the randomized estimator taking values $T_{n,-}$ and $T_{n,+}$ with equal likelihood.

It follows essentially as in Theorem 5.3 that (5.11) holds for the present setting. But in the present case the symmetry of the problem under reflections $(t \to c - t)$ means that the two types of errors of the minimax test are each bounded by $\alpha/2$ and so we can do twice as well as (5.11).

THEOREM 5.4. *For the nonparametric regression model,*

$$P_F\{|T^*_{n,\Delta} - T(F)| \geq \Delta/2\} \leq \alpha_A(n, \Delta). \tag{5.17}$$

Thus the estimator $T^*_{n,\Delta}$ is within a factor 2 of minimax. Actually, if we define the difficulty (as opposed to risk)

$$D(T_n, F) = \max_{\pm} P_F\{\pm(T_n - T_F) \geq \Delta/2\},$$

then the symmetry of worst case type I and type II errors implies that $D(T^*_{n,\Delta}, F) \leq \alpha_A(n, \Delta)/2$ for every $F \in \mathbf{F}$. On the other hand, close inspection of the proof of Theorem 2.1 reveals that for any measurable estimator T_n, $\sup_{\mathbf{F}} D(T_n, F) \geq \alpha_A(n, \Delta)/2$. It follows that $T^*_{n,\Delta}$ is exactly minimax, in finite samples:

$$\inf_{T_n} \sup_{\mathbf{F}} D(T_n, F) = \frac{\alpha_A(n, \Delta)}{2} = \sup_{\mathbf{F}} D\left(T^*_{n,\Delta}, F\right). \tag{5.18}$$

The analogous fact about location estimation occurs in Huber [(1982), Theorem 7.1, page 285].

The setting just developed is rather general and works in other cases as well. We give details for estimating the conditional median. Retain the observation scheme (5.12), but suppose that the unknown error distribution G_i belongs to

$$\mathbf{H}_{m,\delta} = \{G: \mathrm{Med}(G) = 0, g(t) \geq m \text{ on } [-\delta, \delta]\}. \tag{5.19}$$

We are interested in estimating $T(F) = \xi(u_0) = \mathrm{Med}(y|u = u_0)$, the conditional median. Our class $\mathbf{F} = \mathbf{C}\,\mathbf{Med}(C, m, \delta)$ is defined by (5.14) in a fashion similar to $\mathbf{Reg}(C, \varepsilon)$, only with $\mathbf{H}_{m,\delta}$ in place of $\mathbf{G}_\varepsilon$.

Suppose that $\Delta/2 < \delta$, that $2m\delta < 1$ and define $\gamma = (1 - 2m\Delta)^{-1}$. Set

$$\lambda(x; \Delta, m, \delta) = \begin{cases} \gamma, & x > \Delta/2, \\ 1, & x \in [-\Delta/2, \Delta/2], \\ \gamma^{-1}, & x < -\Delta/2. \end{cases}$$

Using this ratio, we can form $L_{t,\Delta}(u, y) = \lambda(y - t; \Xi(u - u_0), m, \delta)$. By mimicking the prescription (5.15)–(5.16), we obtain tests for $\mathbf{F}_{\leq t}$ against $\mathbf{F}_{\geq t+\Delta}$ and an estimator $T^*_{n,\Delta}$ for T. In fact, the tests will be minimax and the conclusion (5.17) will hold for the resulting estimate.

Indeed, consider the problem of testing between H_0: $\{G_0(\cdot + \Delta/2): G_0 \in \mathbf{H}_{m,\delta}\}$ against H_1: $\{G_1(\cdot - \Delta/2): G_1 \in \mathbf{H}_{m,\delta}\}$ on the basis of n i.i.d. observations (x_i). Then λ is stochastically largest under H_0 and stochastically small-

est under H_1, at a pair for which $\prod_i \lambda(x_i)$ furnishes the Neyman–Pearson test. Hence λ furnishes the minimax test between H_0 and H_1. Moreover, λ is monotone. Finally, the maximal type I and type II errors are equal. Hence the key points used in the analysis for $\mathbf{Reg}(C, \varepsilon)$ carry through and (5.17) holds in the present setting. Of course, the minimaxity (5.18) for the difficulty measure D carries through as well.

These examples should reinforce the points (1) that nonparametric regression is a special case of density estimation; (2) that the lower bound of Theorem 2.1 is nearly attainable in some interesting cases; (3) that the optimal rate of convergence is $\omega(n^{-1/2})$ in some nonparametric regression problems.

5.3. *Estimating the mode.* Let $\mathbf{F} = \mathbf{Mode}(M, c_-, c_+, \delta)$, the class of distributions with *unimodal* densities f, that are uniformly bounded:

$$f(x) \le M, \tag{5.20}$$

and have quadratic maxima:

$$\begin{aligned} &f(\text{mode}) - c_+(x - \text{mode})^2 \\ &\quad \le f(x) \le f(\text{mode}) - c_-(x - \text{mode})^2, \qquad |x - \text{mode}| < \delta. \end{aligned} \tag{5.21}$$

Let $T(F) = \text{mode}(F)$.

T is nonlinear. In this case the modulus satisfies $\omega(\varepsilon) = A\varepsilon^{2/5}(1 + o(1))$ and, as we shall see, the rate $\omega(n^{-1/2})$ is attainable. However, attainability is not easy to demonstrate by the methods used so far. Instead, we turn to methods of analysis. For reasons of space, our discussion is abbreviated; for more information, see the technical report.

We study rates of convergence of kernel estimators of the mode. Our rate result applies to any kernel satisfying:

ASSUMPTION. *K is a positive even function of compact support, bounded, square integrable and absolutely continuous, with*

$$\|K\|_2 < \infty, \qquad \|K\|_\infty < \infty,$$

and

$$\|K'\|_2 < \infty, \qquad \|K'\|_\infty < \infty,$$

where the norms of K' are defined distributionally and so represent the smallest constants C_2 and C_∞ for which

$$\|K(\cdot) - K(\cdot - \delta)\|_2 \le C_2\delta,$$

$$\|K(\cdot) - K(\cdot - \delta)\|_\infty \le C_\infty\delta$$

are valid for all $\delta > 0$.

THEOREM 5.5. *Let T be the mode and $\mathbf{F}$ be as in* (5.20)–(5.21). *Then $\omega(\varepsilon)$ is Hölderian with exponent $\frac{2}{5}$ and so no estimator can achieve faster than an*

$n^{-1/5}$ *rate of convergence uniformly over* $\mathbf{F}$:

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{\mathbf{F}} P_F\{|T_n - T| > \omega(n^{-1/2})\} \geq \frac{e^{-1/2}}{2}.$$

Let K *satisfy Assumption* K *and let* $h_n = bn^{-1/5}$. *Let* $T_n^{(k)}$ *be any maximizer of*

$$\hat{f}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \frac{K((X_i - t)/h_n)}{h_n}.$$

Then $T_n^{(k)}$ *attains the* $n^{-1/5}$ *rate uniformly over* $\mathbf{F}$:

$$\lim_{C\to\infty} \limsup_{n\to\infty} \sup_{\mathbf{F}} P_F\{|T_n^{(k)} - T| > C\omega(n^{-1/2})\} = 0.$$

The proof is sketched in Section 7. It involves analysis of the supremum and fluctuations of a certain empirical process. While papers such as Bickel and Rosenblatt (1973), Silverman (1978), Révész (1978) and Rudzkis (1985, 1987) consider similar processes, we are unable to simply apply their results because we must get bounds uniform in $\mathbf{F}$ and such uniformity does not seem to have been directly addressed in empirical process research.

Has'minskii (1979) established a lower bound for estimation of the mode also of the order $n^{-1/5}$, although his results do not quite cover our class. Has'minskii claims in this article that the $n^{-1/5}$ rate is attainable and that the results of Venter (1967) show this. However, Venter's work only establishes individual (rather than uniform) rates and only almost sure (rather than in-probability) rates. Using Lemmas 7.1 and 7.2 and some facts about $\|\hat{f}_n - E\hat{f}_n\|_\infty$ due to Silverman (1978), it is possible to show that the almost sure rate suggested by Venter's result, $\log n/n^{1/5}$, does indeed hold uniformly over $\mathbf{F}$. However, to show that $n^{-1/5}$ is the optimal rate in probability seems genuinely harder; our approach uses Bernstein's inequality and a chaining argument. Thus, Theorem 5.4 verifies Has'minskii's claim and shows that $n^{-1/5}$ is the optimal rate for estimation of the mode over the class $\mathbf{F}$.

6. An interesting example. Consider now the nonlinear functional $T(F) = \int f^2$. Let $\mathbf{F}$ be the family of distributions supported in $[0, 1]$ with densities bounded by M. Then, it follows from Section 5 of [GR I] that $\omega(\varepsilon) \leq 4M\varepsilon$. This suggests that the rate $n^{-1/2}$ might be attainable in estimating this functional.

In a very penetrating analysis, Ritov and Bickel (1990) have shown this guess to be very far from true. Translating their results into the language of this paper, we have:

THEOREM (Ritov–Bickel). *With* T *and* $\mathbf{F}$ *as before,*

$$\alpha_A(n, \Delta) = 1 \quad \textit{for all } n \textit{ and } \Delta \in (0, (M-1)/2), \tag{6.1}$$

$$\Delta_A(n, \alpha) = (M-1)/2 > 0 \quad \textit{for all } n. \tag{6.2}$$

In short, no rate of any kind is available under these conditions. As $\omega(\varepsilon) = O(\varepsilon)$, we thus have an example where

$$\Delta_2(n, \alpha) = O(n^{-1/2})$$

but

$$\Delta_A(n, \alpha) \nrightarrow 0;$$

the two lower bounds behave as differently as it is reasonable to expect. In view of this result, there may be a large and interesting class of cases where the two-point and composite bounds are not comparable [see also Donoho and Nussbaum (1990)].

7. Proofs.

PROOF OF THEOREM 2.3. Before proving (2.10a, b), we first establish some exponential bounds on $\alpha_A(n, d_k)$. Let $\varepsilon_1 = \min((A_1/CA_0)^{2/r}\varepsilon_0, |\log \alpha|/5)$. We consider two cases, depending on k. For k small,

$$\text{(Case 1)} \qquad \left(\frac{3}{2}\right)^k CA_1\left(\frac{|\log \alpha|}{n}\right)^{r/2} \le A_0\varepsilon_1^{r/2}.$$

In this case, by our constraint on ε_1, there exists an integer m_k satisfying $n/2 \ge m_k \ge 5$ and

$$n\left(\frac{A_1}{A_0(3/2)^k C}\right)^{2/r} < m_k,$$

$$3\left\lfloor \frac{n}{m_k} \right\rfloor \ge \left(\left(\frac{3}{2}\right)^k \frac{CA_0}{A_1}\right)^{2/r} - 1$$

and $|\log \alpha|/m_k \le \varepsilon_0$. Then a calculation reveals that

$$A_1\left(\frac{|\log \alpha|}{m_k}\right)^{r/2} < \left(\frac{3}{2}\right)^k CA_0\left(\frac{|\log \alpha|}{n}\right)^{r/2}$$

and as $|\log \alpha|/m_k \le \varepsilon_0$, (2.9) implies

$$\Delta_A(m_k, \alpha) < d_k$$

and thus $\alpha_A(m_k, d_k) \le \alpha$. Le Cam (1973) gives the formula

$$\pi_{jm} \le \left(\sqrt{\pi_m(2 - \pi_m)}\right)^j,$$

where $\pi_{jm} = \pi(\text{conv}(\mathbf{P}^{(jm)}), \text{conv}(\mathbf{Q}^{(jm)}))$ and $\pi_m = \pi(\text{conv}(\mathbf{P}^{(m)}), \text{conv}(\mathbf{Q}^{(m)}))$. We conclude

$$\alpha_A(jm_k, d_k) \le \left(\sqrt{\alpha(2 - \alpha)}\right)^j.$$

Using monotonicity of α_A in n, we get

$$\alpha_A(n, d_k) \le \left(\sqrt{\alpha(2-\alpha)}\right)^{\lfloor n/m_k \rfloor}$$

$$\le \exp\left\{-\frac{1}{6}|\log(\alpha(2-\alpha))|\left[\left(\frac{(3/2)^k CA_0}{A_1}\right)^{2/r} - 1\right]\right\}$$

$$= \exp\left(-2\beta C^{2/r}\left(\frac{3}{2}\right)^{2k/r} + \frac{1}{6}|\log(\alpha(2-\alpha))|\right).$$

The hypothesis $CA_0 > 2A_1$ implies

$$2\beta C^{2/r}\left(\tfrac{3}{2}\right)^{2k/r} - \tfrac{1}{6}|\log(\alpha(2-\alpha))| > \beta C^{2/r}\left(\tfrac{3}{2}\right)^{2k/r}$$

for $k = 0, 1, 2, \ldots$ and so we have, in Case 1,

$$\alpha_A(n, d_k) \le \exp\left(-\beta C^{2/r}\left(\tfrac{3}{2}\right)^{2k/r}\right). \tag{7.1}$$

In Case 2, k is so large that condition (Case 1) does not hold. It follows that $d_k > (A_0^2/A_1)\varepsilon_1^{r/2}$; using the definition of n_0 and arguing as before,

$$\alpha_A(n, d_k) \le \left(\sqrt{\alpha_0(2-\alpha_0)}\right)^{(n/n_0)-1} = \exp\left(-2\gamma n + \tfrac{1}{2}|\log(\alpha_0(2-\alpha_0))|\right).$$

Now as $2((n/n_0) - 1) > n/n_0$ for $n > 2n_0$, we get

$$2\gamma n - \tfrac{1}{2}|\log(\alpha_0(2-\alpha_0))| > \gamma n$$

and so

$$\alpha_A(n, d_k) \le \exp(-n\gamma), \qquad n > 2n_0. \tag{7.2}$$

With our bounds established, we now consider (2.10a). Let K be the number of d_k covered by Case 2. Formally,

$$K = \#\left\{k: A_0\varepsilon_1^{r/2} \le \left(\frac{3}{2}\right)^k A_1\left(\frac{|\log\alpha|}{n}\right)^{r/2} \le \left(\frac{3}{2}\right)^N \Delta\right\}.$$

As $(\frac{3}{2})^N\Delta \le 3M$, $K \le \log(3M)/\log(A_0\varepsilon_1^{r/2})$; thus K is bounded independently of n and N. Now by (7.1) and (7.2), if $n > 2n_0$,

$$\sum_{k=l}^{N-1} \alpha_A(n, d_k) = \sum_{k=l}^{N-K-1} + \sum_{N-K}^{N-1}$$

$$\le \sum_{k=l}^{N-K-1} \exp\left(-\beta C^{2/r}\left(\tfrac{3}{2}\right)^{2k/r}\right) + K\exp(-n\gamma) \tag{7.3}$$

$$\le \sum_{k=l}^{\infty} \exp\left(-\beta C^{2/r}\left(\tfrac{3}{2}\right)^{2k/r}\right) + e_n.$$

If $l = 0$, then since $(\frac{3}{2})^{2k/r} > 2k$ for $k = 1, 2, \ldots,$

$$\sum_{k=0}^{\infty} \exp\left(-\beta C^{2/r}\left(\frac{3}{2}\right)^{2k/r}\right) \le \exp(-\beta C^{2/r}) + \sum_{k=1}^{\infty} \exp(-\beta C^{2/r} 2k)$$

$$= \theta + \frac{\theta^2}{1-\theta^2}.$$

Combining this with (7.3) gives (2.10a). If $l > 0$, we have

$$\sum_{k=l}^{\infty} \exp\left(-\beta C^{2/r}\left(\frac{3}{2}\right)^{2k/r}\right) \le \sum_{k=l}^{\infty} \exp(-\beta C^{2/r} 2k) = \frac{\theta^{2l}}{1-\theta^2}.$$

which, with (7.3), gives (2.10b). □

PROOF OF THEOREM 2.4.

$$E_F l(T_n - T) \le \sum_{k=0}^{N-1} P\{\eta_{k+1} \ge |T_n - T| > \eta_k\} l(\eta_{k+1})$$

$$\le \sum_{k=0}^{N-1} P\{|T_n - T| > \eta_k\} l(\eta_{k+1})$$

$$\le l(\eta_1)\frac{2\theta}{1-\theta^2} + \sum_{k=1}^{N-1} l(\eta_{k+1})\frac{\theta^{2k}}{1-\theta^2} + \sum_{0}^{N-1} l(\eta_{k+1}) e_n.$$

Now as l is well-behaved, $l(\eta_k) \equiv l((\frac{3}{2})^k \Delta) \le a^k l(\Delta)$, so

$$E_F l(T_n - T) \le l(\Delta)\left[\frac{2\theta}{1-\theta^2} a + \sum_{k=1}^{\infty} \frac{\theta^{2k} a^{k+1}}{1-\theta^2} + e_n \frac{a^N - 1}{a-1}\right]$$

$$= l(\Delta)\left[\frac{2a\theta}{1-\theta^2} + \frac{1}{1-\theta^2}\frac{a^2\theta^2}{1-a\theta^2} + e_n \frac{a^N - 1}{a-1}\right]$$

$$= l(\Delta)(A_2 + A_{3,n}),$$

say. Now as $N \le \log(3M)/\log(\Delta) \le \log(3M)((r/2)(\log(n) - \log|\log \alpha|) - \log(A_0))$,

$$a^N e_n = \exp(-n\gamma + \log(a) N)$$

$$\le \exp(-n\gamma + A_4 \log(n) + A_5) \to 0, \qquad n \to \infty.$$

Therefore, $A_{3,n} \to 0$. For large enough n, $A_{3,n} \le A_2$ and so $l(\Delta)(A_2 + A_{3,n}) \le 2A_2 l(\Delta)$. Then, as $l(\Delta) \equiv l(C\Delta_A(n, \alpha)) \le a^{\lceil \log C/\log 1.5 \rceil} l(\Delta_A(n, \alpha))$, we have (2.11) with $A = 2A_2 a^{\lceil \log C/\log 1.5 \rceil}$. □

PROOF OF THEOREM 4.4. [Proof that (i) implies (ii).] As ω is Hölderian, for ε_0 small enough, then by (3.5) there exist constants A_0, A_1 so that

$$A_0\left(\frac{|\log \alpha|}{n}\right)^{r/2} \le \Delta_2(n,\alpha) \le A_1\left(\frac{|\log \alpha|}{n}\right)^{r/2} \tag{7.4}$$

for $\alpha < \alpha_0$ and $|\log \alpha|/n < \varepsilon_0$. Let m be an integer so that $m/2 + 1 > M$. Now we claim that

$$\frac{|\log(\alpha_A(mn,\Delta))|}{|\log(\alpha_A(n,\Delta))|} \ge \frac{m}{2} + 1. \tag{7.5}$$

Let us see why. Let $\rho_A(n,\Delta)$ denote a quantity similar to $\alpha_A(n,\Delta)$, only defined using Hellinger affinity rather than testing affinity. Then

$$\alpha_A(mn,\Delta) \le \rho_A(mn,\Delta) \le \rho_A(n,\Delta)^m \le \left(2\alpha_A(n,\Delta)^{1/2}\right)^m,$$

where the second inequality follows from (3.15) and the third from (3.7). Thus

$$|\log \alpha_A(mn,\Delta)| \ge m\left[\tfrac{1}{2}|\log \alpha_A(n,\Delta)| + \log 2\right].$$

Now, for $\alpha_A < \frac{1}{2}$, $\log 2/|\log \alpha_A| \ge 1$ and so the last display proves (7.5).

Combining (7.5) with hypothesis (4.6) gives

$$|\log \alpha_A(mn,\Delta)| > |\log \alpha_2(n,\Delta)|.$$

It then follows that, with $\alpha = \alpha_2(n,\Delta)$, $\Delta_A(mn,\alpha) \le \Delta_2(n,\alpha)$. Now by (7.4),

$$\Delta_2(n,\alpha) \le A_1\left(\frac{|\log \alpha|}{n}\right)^{r/2}$$

and

$$A_0\left(\frac{|\log \alpha|}{mn}\right)^{r/2} \le \Delta_2(mn,\alpha).$$

Combining these,

$$\Delta_2(n,\alpha) \le \frac{A_1 m^{r/2}}{A_0}\Delta_2(mn,\alpha). \tag{7.6}$$

Hence, for every $k = mn$,

$$\Delta_A(k,\alpha) \le C_0\Delta_2(k,\alpha), \tag{7.7}$$

where $C_0 = A_1 m^{r/2}/A_0$. The argument extends to other k with the larger constant $C = (A_1/A_0)^2(2m)^{r/2}$; details are given in the technical report.

[Proof that (ii) implies (i).] As in the last proof, (7.4) holds by hypothesis. Pick an integer m so that $(A_0/A_1)m^{r/2} > C$. Then

$$\begin{aligned}\Delta_A(mn,\alpha) &\le C\Delta_2(mn,\alpha)\\ &\le CA_1\left(\frac{|\log \alpha|}{mn}\right)^{r/2} \le C\frac{A_1}{m^{r/2}A_0}A_0\left(\frac{|\log \alpha|}{n}\right)^{r/2}\\ &< A_0\left(\frac{|\log \alpha|}{n}\right)^{r/2} \le \Delta_2(n,\alpha).\end{aligned}$$

Hence, with $\Delta = \Delta_A(mn, \alpha)$, $\alpha_2(n, \Delta) \geq \alpha$. Defining $\rho_2(n, \Delta)$ in a fashion analogous to $\Delta_2(n, \Delta)$, only using ρ in place of testing affinity, we have that $\rho_2(n, \Delta) \geq \alpha$ and also $\rho_2(mn, \Delta) \geq \alpha^m$. Then from $\pi \geq \frac{1}{2}\rho^2$,

$$\alpha_2(mn, \Delta) \geq \tfrac{1}{2}\alpha^{2m}.$$

And so, if $\alpha < \frac{1}{2}$,

$$\frac{|\log \alpha_2(mn, \Delta)|}{|\log \alpha_A(mn, \Delta)|} \leq 2m + 1. \tag{7.8}$$

It follows that, for $\alpha_A < \frac{1}{2}$ and k of the form $m \cdot n$, we have (4.7) with $M = 2m + 1$. The argument extends to other k with the larger constant $M = 10m + 5$; see the technical report for details. □

PROOF OF THEOREM 5.2. (5.7) shows that $L_{t,\Delta}(x)$ is constant in t for $x \leq a_1(t, \Delta)$ and is a monotone increasing function of $a_1(t, \Delta)$ for $x > a_1(t, \Delta)$. Thus the proof requires showing that $a_1(t, \Delta)$ is monotone decreasing in t. Now $a_1(t, \Delta)$ is the value of x solving

$$\lambda(x) = \mu(x, t), \tag{7.9}$$

where

$$\lambda(x) = \frac{C_+}{C_-} x^{\Delta} \frac{1 + \gamma x^p}{1 - \gamma x^p}$$

and

$$\mu(x, t) = \frac{1 - \int_0^x f_1(v)\, dv}{1 - \int_0^x f_0(v)\, dv}.$$

Now λ is monotone increasing in x. We claim that for sufficiently small x_0, $x \in (0, x_0)$, $0 < x_0 < 1$, $\mu(x, t)$ is monotone decreasing in t. Then, at any (t, Δ) pair at which the solution $a_1(t, \Delta)$ of (7.9) falls in the interval $(0, x_0)$, the solution must be monotone decreasing in t. Finally, a little calculus will show that for a given x_0, there is a $\Delta_0 > 0$ so that $a_1(t_0, \Delta) < x_0$ for $\Delta < \Delta_0$, where t_0 is the constant used in (5.2a) defining the class **F**. Combining the last two sentences completes the proof.

It remains only to establish the claim, that is, to show that $\mu(x, t)$ is monotone decreasing in t. Put

$$\mu(x, t) = \frac{1 - \beta(t)}{1 - \alpha(t)},$$

where

$$\alpha(t) = C_+ \frac{x^{t+\Delta+1}}{t + \Delta + 1}\left(1 + \gamma x^p \frac{t + \Delta + 1}{t + \Delta + p + 1}\right),$$

$$\beta(t) = C_- \frac{x^{t+1}}{t + 1}\left(1 - \gamma x^p \frac{t + 1}{t + p + 1}\right).$$

Now one can easily verify that, if $x \le 1$ and also $\gamma x^p \le 1$,

$$\alpha(t), \beta(t) \quad \text{are decreasing in } t. \tag{7.10}$$

Then monotonicity of $\mu(x,t)$ follows from

$$\frac{1-\beta(t)}{1-\alpha(t)} < \frac{\beta'(t)}{\alpha'(t)}.$$

In fact, we will show that for x_0 small enough,

$$\frac{1-\beta(t)}{1-\alpha(t)} < 2 < \frac{\beta'(t)}{\alpha'(t)} \tag{7.11}$$

for all $t \ge 0$ and all $x \in (0, x_0)$.

Let us first establish the left-hand inequality. This can be rewritten as $1 > 2\alpha(t) - \beta(t)$ and as $2\alpha(t) + \beta(t) > 2\alpha(t) - \beta(t)$, it is implied by $1 > \max_t 2\alpha(t) + \beta(t)$. By (7.10), this reduces to

$$1 > 2\alpha(0) + \beta(0). \tag{7.12}$$

Now pick x_1 so that

$$1 > 2C_+ \frac{x_1^{\Delta+1}}{\Delta+1}\left(1 + \gamma x_1^p \frac{\Delta+1}{\Delta+p+1}\right) + C_- x_1\left(1 + \gamma \frac{x_1^p}{p+1}\right).$$

Then for $x \in (0, x_1)$,

$$\begin{aligned} &2\alpha(0) + \beta(0) \\ &\quad = 2C_- x\left(1 - \gamma \frac{x^p}{p+1}\right) + C_+ \frac{x^{\Delta+1}}{\Delta+1}\left(1 + \gamma x^p \frac{\Delta+1}{\Delta+p+1}\right) \\ &\quad \le 2C_- x_1\left(1 + \gamma \frac{x_1^p}{p+1}\right) + C_+ \frac{x_1^{\Delta+1}}{\Delta+1}\left(1 + \gamma x_1^p \frac{\Delta+1}{\Delta+p+1}\right) \\ &\quad < 1 \end{aligned}$$

and so (7.12) follows. Thus the left-hand side of (7.11) is established for $x_0 \le x_1$.

We now consider the right-hand inequality of (7.11). Now

$$\beta'(t) = \Psi(x,t)[B_1 - B_2 + B_3],$$

$$\alpha'(t) = \Psi(x,t)[A_1 - A_2 + A_3],$$

where $\Psi(x,t) = (x^{t+1}/(t+1))|\log(x)|$ and

$$B_1 = -C_-\left(1 - \frac{\gamma(t+1)x^p}{t+p+1}\right),$$

$$B_2 = C_-\left(1 - \frac{\gamma(t+1)x^p}{t+p+1}\right)\frac{1}{(t+1)|\log(x)|},$$

$$B_3 = C_-\left(\frac{\gamma(t+1)x^p}{(t+p+1)^2} - \frac{\gamma x^p}{t+p+1}\right)\frac{1}{|\log(x)|},$$

$$A_1 = -C_+\left(1 + \frac{\gamma(t+\Delta+1)x^p}{t+p+\Delta+1}\right)x^\Delta\frac{t+1}{t+\Delta+1},$$

$$A_2 = C_+\left(1 + \frac{\gamma(t+\Delta+1)x^p}{t+p+\Delta+1}\right)x^\Delta\frac{t+1}{(t+\Delta+1)^2}\frac{1}{|\log(x)|},$$

$$A_3 = C_+\left(\frac{\gamma x^p}{t+p+\Delta+1} - \frac{\gamma(t+\Delta+1)x^p}{(t+p+\Delta+1)^2}\right)x^\Delta\frac{t+1}{(t+\Delta+1)^2}\frac{1}{|\log(x)|}.$$

The desired inequality is then equivalent to

$$B_1 - B_2 + B_3 < 2(A_1 - A_2 - A_3) \tag{7.13}$$

for all t and all $x < x_0$.

Note that for $x \in (0,1)$, B_1 is increasing in t. Thus

$$B_1 \le \lim_{t\to\infty} B_1 = -C_-(1 - \gamma x^p) = B_4(x);$$

similarly A_1 is decreasing in t and

$$A_1 \ge \lim_{t\to\infty} A_1 = -C_+x^\Delta(1 + \gamma x^p) = A_4(x).$$

Pick $\varepsilon > 0$. For x_2 small enough, $B_4(x_2) < 2A_4(x_2) - \varepsilon$ and so by the obvious monotonicities in x, $B_4(x) < 2A_4(x) - \varepsilon$ for all $x \in (0, x_2)$. We will show below that B_2, B_3, A_2 and A_3 are negligible, in the sense that

$$|B_2| + |B_3| + 2|A_2| + 2|A_3| < \varepsilon \tag{7.14}$$

for $x < x_3$. Then (7.13) follows, for $x_0 \le \min(x_2, x_3)$.

Note the inequalities

$$\begin{aligned} |B_2| &\le C_-\frac{1+\gamma x^p}{|\log(x)|}, \qquad |B_3| \le C_-\frac{2\gamma x^p}{|\log(x)|}, \\ |A_2| &\le C_+x^\Delta\frac{1+\gamma x^p}{|\log(x)|}, \qquad |A_3| \le C_+x^\Delta\frac{2\gamma x^p}{|\log(x)|}, \end{aligned} \tag{7.15}$$

valid for $t \ge 0$, $0 < x < 1$. Pick x_3 so small that the sum of the upper bounds in (7.15) is less than $\varepsilon/2$. Then we have (7.14) for all $x < x_3$ by the monotonicity in x of the upper bounds in (7.15).

Putting $x_0 = \min(x_1, x_2, x_3)$, we see that both sides of (7.11) hold for all $t \geq 0$ and all $x \in (0, x_0)$. □

PROOF OF THEOREM 5.5. The claim about the modulus follows from Theorem 4.2 of [GR I]. The claim about the lower bound follows from Theorem 2.1 of [GR I]. As $\omega(n^{-1/2})$ is asymptotic to $An^{-1/5}$ for an appropriate constant A, the proof is completed by showing that $T_n^{(k)} - T(F) = O_p(n^{-1/5})$ uniformly in $\mathbf{F}$.

Suppose without loss of generality that K is a probability density: $\int K = 1$. Then $\hat{f}_n$ is an estimated density and $f_n(t) = E\hat{f}_n(t)$ is a density.

Let t_n be an maximizer of $f_n(t)$. Let $\hat{t}_n$ be any maximizer of $\hat{f}_n(t)$. By Lemma 7.1, the assumption $h_n = bn^{-1/5}$ guarantees that $t_n - T(F) = O(n^{-1/5})$ uniformly in $\mathbf{F}$. Thus the theorem is proved, if we can show that $\hat{t}_n - t_n = O_p(n^{-1/5})$ also uniformly in $\mathbf{F}$.

Now we have

$$\hat{f}_n(\hat{t}_n) \geq \hat{f}_n(t_n) \tag{7.16}$$

and so

$$\left(\hat{f}_n(\hat{t}_n) - f_n(\hat{t}_n)\right) - \left(\hat{f}_n(t_n) - f_n(t_n)\right) \geq f_n(t_n) - f_n(\hat{t}_n).$$

Now by Lemma 7.2, there is a constant $\gamma > 0$ so that

$$f_n(t_n) - f_n(t) \geq \gamma(t - t_n)^2 \tag{7.17}$$

uniformly in $\mathbf{F}$, if $t \in I_n \equiv (T(F) + qh_n s, T(F) + c)$, for a certain $q > 0$ defined in the lemma and any c smaller than the constant δ used in defining the class $\mathbf{F}$. It follows that if $\hat{t}_n \in I_n$, then

$$Z_n(\hat{t}_n) - Z_n(t_n) \geq \gamma n^{2/5}(\hat{t}_n - t_n)^2, \tag{7.18}$$

where Z_n is the stochastic process

$$Z_n(t) = n^{2/5}\left(\hat{f}_n(t) - f_n(t)\right).$$

Therefore, if Δ is so large that $n^{-1/5}\Delta > (q + 1)h_n s$, we must have

$$\begin{aligned} P_F\{\hat{t}_n - t_n > n^{-1/5}\Delta\} \leq\ & P_F\Big\{Z_n(t) - Z_n(t_n) > \gamma n^{2/5}(t - t_n)^2 \\ & \qquad \text{for some } t \in \left(t_n + n^{-1/5}\Delta, t_n + c\right)\Big\} \\ & + P_F\left\{\sup_{t > t_n + c} \hat{f}_n(t) > \hat{f}_n(t_n)\right\}. \end{aligned}$$

Now

$$\begin{aligned} P_F\left\{\sup_{t > t_n + c} \hat{f}_n(t) > \hat{f}_n(t_n)\right\} \leq\ & P_F\left\{\hat{f}_n(t_n) < f_n(t_n) - \frac{\gamma}{2}c^2\right\} \\ & + P_F\left\{\sup_{t > t_n + c} \hat{f}_n(t) > f_n(t_n) - \frac{\gamma}{2}c^2\right\}. \end{aligned}$$

Using Lemma 7.3, we get that the first probability on the right side may be bounded, for all $F \in \mathbf{F}$, by $\exp(-n^{4/5}\,\text{Const})$, while for each $\delta > 0$, an argument like that in Lemma 7.4 can be used to show that the second term on the right side is bounded above, for all $F \in \mathbf{F}$, by $P(n)\exp(-n^{4/5}\,\text{Const}(\delta)) + \delta$, for a polynomial $P(n)$. The conclusion is that the left-hand side of this display must tend to zero uniformly in $\mathbf{F}$.

By Lemma 7.3,

$$P_F\left\{Z_n(t_n) < -\frac{\gamma}{2}\Delta^2\right\} \le \exp\left\{-\frac{n^{1/5}h_n\gamma^2\Delta^4/8}{M\|K\|_2^2 + \max(Mh_n, \|K\|_\infty)n^{-2/5}\Delta^2}\right\}$$

for every $F \in \mathbf{F}$. Hence, if $n^{1/5}h_n = \text{constant}$,

$$0 = \lim_{\Delta\to\infty} \limsup_n \sup_{\mathbf{F}} P_F\left\{Z_n(t_n) < -\frac{\gamma}{2}\Delta^2\right\}.$$

If we can also show that

$$0 = \lim_{\Delta\to\infty} \limsup_n \sup_{\mathbf{F}} P_F\left\{Z_n(t) > \frac{\gamma}{2}n^{2/5}(t - t_n)^2 \text{ for some } t > t_n + n^{-1/5}\Delta\right\}, \tag{7.19}$$

then

$$\lim_{\Delta\to\infty} \limsup_n \sup_{\mathbf{F}} P_F\{\hat{t}_n - t_n > n^{-1/5}\Delta\} = 0.$$

By the obvious symmetry in the problem, a similar relation would hold for $\hat{t}_n - t_n < -n^{-1/5}\Delta$ and so we would have $(\hat{t}_n - t_n) = O_p(n^{-1/5})$ uniformly in $\mathbf{F}$ and the proof would be done. Consider then (7.19). As the class $\mathbf{F}$ is closed under translation, we can always assume F is such that $t_n(F) = 0$. Define $\mathbf{F}_{n,0} = \{F\colon F \in \mathbf{F} \text{ and } t_n(F) = 0\}$. Our aim is to show that the previous display tends to zero uniformly for $F \in \mathbf{F}_{n,0}$.

Now, for $\delta > 0$ and for $i = 0, 1, \ldots,$ put $t_i = n^{-1/5}(\Delta + \delta i)$; we will also refer to $\Delta_i = (\Delta + \delta i)$ so that $t_i = n^{-1/5}\Delta_i$ (and $\Delta_0 = \Delta$).

$$\begin{aligned} &P_F\left\{Z_n(t) > n^{2/5}\frac{\gamma}{2}t^2 \text{ for some } t > n^{-1/5}\Delta\right\} \\ &\le P_F\left\{Z_n(t_i) > \frac{\gamma}{4}\Delta_i^2 \text{ for some } i \ge 0\right\} \\ &\quad + P_F\left\{\sup_{t_i \le t \le t_{i+1}} Z_n(t) > \frac{\gamma}{4}\Delta_i^2 \text{ for some } i \ge 0\right\}. \end{aligned} \tag{7.20}$$

By Lemma 7.4, with $\alpha = \frac{1}{4}$,

$$\lim_{\Delta\to\infty} \limsup_n \sup_{\mathbf{F}_{n,0}} P_F\left\{Z_n(t_i) > \frac{\gamma}{4}\Delta_i^2 \text{ for some } i \ge 0\right\} = 0. \tag{7.21}$$

By Lemma 7.6, with $\alpha = \frac{1}{4}$,

$$(7.22)\quad \lim_{\Delta\to\infty} \limsup_{n} \sup_{\mathbf{F}_{n,0}} P_F\left\{ \sup_{t_i \le t \le t_{i+1}} Z_n(t) > \frac{\gamma}{4}\Delta_i^2 \text{ for some } i \ge 0 \right\} = 0.$$

So (7.19) is established and the proof is completed. The following lemmas are proven in the technical report. □

LEMMA 7.1. *Let K be positive and of support $[-s, s]$. Then for every F in $\mathbf{F}$,*

$$|t_n - T(F)| \le h_n s,$$

where t_n denotes the maximizer of $f_n(t)$.

LEMMA 7.2. *Let K be positive and of support $[-s, s]$. Let $q > 0$ be so large that*

$$\frac{1}{2} < \frac{(q-1)^2}{(q+1)^2} - 4\frac{C_+}{C_-}\frac{1}{(q-1)^2}.$$

Put $\gamma = C_-/2$. Then for every $F \in \mathbf{F}$,

$$f_n(t_n) - f_n(t) > \gamma(t - t_n)^2 \qquad t - t_n \in (qh_n s, c - t_n)$$

for $c < \delta$.

LEMMA 7.3.

$$\sup_{\mathbf{F}} P_F(Z_n(t) > \Delta) \le \exp\left(\frac{-n^{1/5}h_n\Delta^2/2}{M\|K\|_2^2 + \max(Mh_n, \|K\|_\infty)n^{-2/5}\Delta} \right).$$

This is an application of Bernstein's inequality.

LEMMA 7.4. *Let $t_i = n^{-1/5}(\Delta_0 + i\delta)$. Let $Mh_n < \|K\|_\infty$. Then for $\alpha \in (0, 1)$,*

$$\sup_{\mathbf{F}_{n,0}} \sum_{i=0}^{\infty} P_F\{Z_n(t_i) > \alpha\gamma\Delta_i^2\} \le \frac{\exp(-\beta_0\Delta_0^2)}{1 - \exp(-\beta_0\Delta_0\delta)} + \frac{\exp(-\beta_2 nh_n)}{1 - \exp(-\beta_1\delta n^{2/5})},$$

where

$$\beta_0 = \frac{n^{1/5}h_n\alpha^2\gamma^2}{22\|K\|_2^2 M}, \qquad \beta_1 = \frac{n^{1/5}h_n\alpha^2\gamma^2}{2.2\|K\|_\infty}, \qquad \beta_2 = \frac{\alpha^2\gamma^2\|K\|_2^2 M}{.22\|K\|_\infty^2}.$$

This is an application of the previous lemma and additional calculations.

LEMMA 7.5. *Suppose that the kernel K has*

$$\|K(\cdot) - K(\cdot + \xi)\|_2^2 \le \xi^2\|K'\|_2^2,$$
$$\|K(\cdot) - K(\cdot + \xi)\|_\infty \le \xi\|K'\|_\infty$$

for some constants $\|K'\|_2^2$, $\|K'\|_\infty$. *Let* $\eta > 0$. *Suppose that the kernel* K *is positive, supported in* $[-s, s]$ *and that* $2h_n s < t$. *Then*

$$\sup_{\mathbf{F}_{n,0}} P_F\{n^{2/5}(\hat{f}_n(t + \xi h_n) - \hat{f}_n(t)) > \eta\xi\} \le \exp\left(\frac{-n^{1/5}h_n\eta^2/2}{\|K'\|_2^2 M + \|K'\|_\infty n^{-2/5}\eta/3}\right).$$

This is an application of Bernstein's inequality.

LEMMA 7.6. *Suppose that* K *satisfies Assumption* K, *with support* $[-s, s]$. *Suppose that* $2h_n s < n^{-1/5}\Delta_0$. *Put* $t_i = n^{-1/5}(\Delta_0 + \delta i)$, $\Delta_i = \Delta_0 + \delta i$. *Then for* Δ_0 *large enough* (K, δ *fixed*).

$$\sup_{\mathbf{F}_{n,0}} \sum_{i=0}^{\infty} P_F\left\{\sup_{t_i \le t \le t_{i+1}} n^{2/5}(\hat{f}_n(t) - \hat{f}_n(t_i)) > \alpha\gamma\Delta_i^2\right\}$$
$$\le \frac{\beta_5(\Delta_0)\exp(-\beta_7(\Delta_0))}{1 - \exp(-\beta_7(\Delta_0))} + \frac{\beta_8(\Delta_0)\exp(-\beta_{10}(\Delta_0))}{1 - \exp(-\beta_{10}(\Delta_0))},$$

where

$$\beta_3 = \frac{b}{22\|K'\|_2^2 M},$$

$$\beta_4 = \frac{3b}{2.2\|K'\|_\infty},$$

$$\beta_5 = \frac{1}{1 - \exp\left(-\beta_3(\alpha\gamma)^2\Delta_0^3\delta b^2/2\right)}$$

$$\beta_6 = \min_{k \ge 1} 2^{2k}k^{-5}(36/\pi^4),$$

$$\beta_7 = \beta_3(\alpha\gamma b)^2\Delta_0^4\beta_6 - \log(2),$$

$$\beta_8 = \frac{1}{1 - \exp\left(-\beta_4\alpha\gamma b\Delta_0^2 n^{2/5}/2\right)},$$

$$\beta_9 = \min_{k \ge 1} 2^k k^{-3}(6/\pi^2),$$

$$\beta_{10} = \beta_4\alpha\gamma b\Delta_0^2\beta_9 n^{2/5} - \log(2),$$

provided $n^{1/5}h_n = b$ *independent of* n *and* $Mh_n < \|K'\|_\infty$. *The inequality is valid as soon as* $\beta_7 > 0$ *and* $\beta_{10} > 0$.

The proof uses a chaining argument. See the technical report.

Acknowledgments. As in other papers of this series, the authors would like to thank Lucien Le Cam for many helpful discussions. L. D. Brown, Jianqing Fan and a referee provided valuable comments on an earlier draft.

REFERENCES

BEDNARSKI, T. (1982). Binary experiments, minimax tests and 2-alternating capacities. *Ann. Statist.* **10** 226–232.

BERAN, R. J. (1977). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5** 445–463.

BICKEL, P. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095.

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

DONOHO, D. L. (1989). Statistical estimation and optimal recovery. Technical Report 214, Dept. Statistics, Univ. California, Berkeley.

DONOHO, D. L. and LIU, R. C. (1987). Geometrizing rates of convergence, I. Technical Report 137, Dept. Statistics, Univ. California, Berkeley.

DONOHO, D. L. and LIU, R. C. (1988). The "automatic" robustness of minimum distance functionals. *Ann. Statist.* **16** 552–586.

DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence, III. *Ann. Statist.* **19** 668–701.

DONOHO, D. L., LIU, R. C. and MACGIBBON, B. (1990). Minimax risk for hyperrectangles. *Ann. Statist.* **18** 1416–1437.

DONOHO, D. L. and NUSSBAUM, M. (1990). Minimax quadratic estimation of a quadratic functional. *J. Complexity* **6** 290–323.

DU MOUCHEL, W. H. (1983). Estimating the stable index α in order to measure tail thickness. *Ann. Statist.* **11** 1019–1031.

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

HALL, P. and MARRON, S. (1987). On the amount of noise inherent in bandwidth selection. *Ann. Statist.* **15** 163–181.

HALL, P. and WELSH, A. H. (1984). Best attainable rates of convergence for parameters of regular variation. *Ann. Statist.* **12** 1079–1084.

HAS'MINSKII, R. Z. (1979). Lower bound for the risks of nonparametric estimates of the mode. In *Contributions to Statistics: Jaroslav Hájek Memorial Volume* (J. Jurekova, ed.) 91–97. Academia, Prague.

HUBER, P. J. (1982). *Robust Statistics*. Wiley, New York.

HUBER, P. J. and STRASSEN, V. (1973). Minimax tests and the Neyman–Pearson lemma for capacities. *Ann. Statist.* **1** 251–263.

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

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1984). Nonparametric estimation of the value of a linear functional in Gaussian white noise. *Theory Probab. Appl.* **29** 1–32.

IBRAGIMOV, I. A., NEMIROVSKII, A. S. and HAS'MINSKII, R. Z. (1987). Some problems on nonparametric estimation in Gaussian white noise. *Theory Probab. Appl.* **31** 391–406.

LE CAM, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53.

LE CAM, L. (1975). On local and global properties in the theory of asymptotic normality of experiments. In *Stochastic Processes and Related Topics* (M. L. Puri, ed.) **1** 13–54. Academic, New York.

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

MEYER, T. G. (1977). On fixed or scaled radii confidence sets: The fixed sample size case. *Ann. Statist.* **5** 65–78.

PITMAN, E. J. G. (1979). *Some Basic Theory for Statistical Inference*. Wiley, New York.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer. New York.

RÉVÉSZ, P. (1978). A strong law for the empirical density function. *Period. Math. Hungar.* **9** 317–324.

RITOV, Y. and BICKEL, P. J. (1990). Achieving informations bounds in non and semiparametric models. *Ann. Statist.* **18** 925–938.

RUDZKIS, R. (1985). On the probability of large excursions of a nonstationary Gaussian process I. *Litovsk. Mat. Sb.* **25** 143–154.

RUDZKIS, R. (1987). On Gaussian process large excursion probability density II. *Litovsk. Mat. Sb.* **27** 731–746.

SACKS, J. and YLVISAKER, N. D. (1981). Asymptotically optimum kernels for density estimation at a point. *Ann. Statist.* **9** 334–346.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications in Statistics.* Wiley, New York.

SILVERMAN, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a probability density and its derivatives. *Ann. Statist.* **6** 177–184.

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

VENTER, J. (1967). On estimation of the mode. *Ann. Math. Statist.* **38** 1446–1455.

WAHBA, G. (1975). Optimal convergence properties of variable knot, kernel and orthogonal series methods for density estimation. *Ann. Statist.* **3** 15–29.

DEPARTMENT OF STATISTICS
STATISTICAL LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720