Geometrizing Rates of Convergence, II

David L. Donoho

Richard C. Liu

Technical Report No. 120
October 1987
First Revision March 1988
Second Revision October 1988

Department of Statistics
University of California
Berkeley, California
Consider estimating a functional $T(F)$ of an unknown distribution $F \sim F$ from data $X_1, \cdots, X_n$ i.i.d. $F$. A companion paper introduced a bound on the rate of convergence of estimates $T_n$ of $T$ as a function of $n$. The bound involved the modulus of continuity $b(e)$ of the functional $T$ over $F$. The bound says that the estimation error $T_n - T$ cannot converge to zero faster than $b(n^{-1/2})$ uniformly over $F$. This rate bound was shown to be at least as strong as some earlier bounds on rates of convergence.

In this paper we show that the "modulus of continuity" bound is attainable, to within constants, whenever $T$ is linear and $F$ is convex. In two nonlinear cases -- estimating the rate of decay of a density, and estimating the mode -- the bound is also attainable to within constants.

We do this by introducing a new bound on the rate of convergence and showing that this new bound is always attainable (to within constants). The new bound is based on the difficulty of testing between the composite, infinite dimensional hypotheses $H_0: T(F) \leq t$ and $H_1: T(F) \geq t + \Delta$.

The modulus bound and the new bound are comparable -- and hence the modulus bound is attainable to within constants -- whenever the difficulty of the hardest simple two-point testing subproblem is comparable to the difficulty of the full infinite-dimensional composite problem. This property holds whenever $T$ is linear and $F$ is convex, and also in the cases of the tail rate functional and the mode discussed above.
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 (1987a,b) (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 $F \in F$ where $F$ 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 $n^{-q/2}$ for some $q < 1$.

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 $F$, with respect to Hellinger distance,

$$b(\varepsilon) = \sup \{|T(F_1) - T(F_0)| : H(F_1, F_0) \leq \varepsilon, F_1 \in F\}$$

(1.1)

it was shown that no estimator can converge to $T(F)$ faster than $b(n^{-1/2})$ uniformly over $F$. 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 Hasminskii.

In this paper we discuss the attainability of this bound as regards rate. Since the $b(n^{-1/2})$ 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 $l(\frac{3}{2}t) \leq a l(t)$ for all $t$. Thus $t^2$ and $|t|$ are well-behaved, with $a = 9/4$ and $a = 3/2$ 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 below, we get

Corollary. Let $T$ be linear and $F$ be convex. If $T$ is bounded on $F$, so that

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

and if $b(\varepsilon)$ is Holderian with exponent $q$, so that
then the optimal rate of convergence is \( b(n^{-1/2}) \):

\[
\inf_T \sup_{\mathcal{F}} \mathbb{E}_F l(T_n - T) \asymp l(b(n^{-1/2})).
\]

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 \( r = q/2 \), where \( q \) 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 \( \mathcal{F} \).

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 \( H_0: T(F) \leq t \) against the composite hypothesis \( H_1: T(F) \geq t + \Delta \). 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 functionals which comes equipped with a (near-) attainability result.

In section 3, we show that, in the linear \( T \), convex \( \mathcal{F} \) case, the modulus bound and the new bound agree to within constants. 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 \( T(F_0) \leq t \) and \( T(F_1) \geq t + \Delta \), should be roughly the same as the difficulty of the full composite problem \( H_0:T(F) \leq t \) versus \( H_1:T(F) \geq t + \Delta \). Thus, the modulus bound "works" in the case of \( T \) linear, \( \mathcal{F} \) 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 (1983)). For this functional, the minimax identity holds precisely. Actually, in this case, the minimax test of $H_0: T(F) \leq t$ versus $H_1: T(F) > t + \Delta$ can be worked out in detail; it turns out to have a certain monotonicity in $t$ which shows that the new bound can be attained to within a factor 2. In the second 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 (1987) and, in a related problem, of Ibragimov, Nemirovskii, and Hasminskii (1987), attainability of the modulus bound can fail already for quadratic functionals. Our calculations, which we plan to present in another paper, give examples where the modulus bound and the Farrell/Stone/Hasminskii bounds fail to give the optimal rate, but our new bound can be computed, satisfies the hypothesis (2.9), and so gives the right rate.

An interesting feature of our approach is the use of notation and techniques due to Le Cam (1973, 1975, 1985) 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 this paper, which seeks to relate the Hellinger modulus to the difficulty of certain tests. In particular, Le Cam's little known result, given below 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)).

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 recover-
ing 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 $F$ be the regularity class in which $F$ is known to lie. Let $F_{sl}$ and $F_{sl+\Delta}$ denote the subsets of $F$ where $T$ takes values $\leq t$ and $\geq t + \Delta$, respectively. Let $F_{s_1}^{(s)}$ denote the set of product measures of $X_1, \ldots, X_n$ iid $F$, $F \in F_{sl}$, and similarly for $F_{s_1 + \Delta}^{(s)}$. Denote by $\text{conv} (F_{s_1}^{(s)})$ the set of all measures on $R^n$ which can be gotten as convex combinations of the product measures in $F_{s_1}^{(s)}$. Such a measure corresponds to the following: a random device is used to select an element $F \in F_{sl}$, and then $n$ observations are taken from this realized $F$. In words, $\text{conv} (F_{s_1}^{(s)})$ represents all the joint distributions 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 $F_{sl}$.

Let $P$ and $Q$ be probability distributions on a common space. Then the testing affinity (LeCam (1973), (1985)) is

$$\pi(P, Q) = \inf_{0 \leq \phi \leq 1} E_P \phi + E_Q (1 - \phi)$$

it is the sum of errors of the best test between $P$ and $Q$. If $P$ and $Q$ are sets of measures, let $\pi(P, Q)$ denote the largest testing affinity $\pi(P, Q)$ between any pair $(P, Q)$ with $P \in P$ and $Q \in Q$—the difficulty of the hardest "two-point" testing problem. We note, following Le Cam (1973, 1985) that if we view $P$ and $Q$ as composite hypotheses, the minimax risk, i.e. the risk of the best test for separating $P$ and $Q$ is $\pi(\text{conv} (P), \text{conv} (Q))$. (Unless $P$ and $Q$ are convex, this minimax risk is usually unequal to the risk $\pi(P, Q)$ of the hardest 2-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 $P$ and $Q$. 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

$$\alpha_A (n, \Delta) = \sup_t \pi(\text{conv} (F_{s_1}^{(s)}), \text{conv} (F_{s_1 + \Delta}^{(s)})).$$

This is the minimax risk of the hardest problem of distinguishing $H_0: F_{sl}$ and $H_1: F_{sl+\Delta}$ at sample
size n. Next, we let \( \Delta_A (n, \alpha) \) be the function inverse to \( \alpha_A \):\[
\Delta_A (n, \alpha) = \sup \{ \Delta : \alpha_A (n, \Delta) \geq \alpha \}.
\]
In words, \( \Delta_A (n, \alpha) \) measures the largest \( \Delta \) at which, in a sample of size \( n \), one cannot test hypotheses \( H_0 : F_{s t} \) and \( H_1 : F_{s t + \Delta} \) with sum of errors less than \( \alpha \). As one might expect based on the exposition in [GR I], \( \Delta_A \) 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 \( T_n \leq T + \Delta/2 \), decide \( H_1 \) if \( T_n > T + \Delta/2 \).

**Theorem 2.1 (Lower Bound).**

\[
\inf_{T_n} \sup_{F \in [F_{s t}, F_{s t + \Delta}]} P_{F} (|T_n - T (F)| \geq \Delta_A (n, \alpha)/2) \geq \alpha/2 \tag{2.3}
\]

**Proof.** Without loss of generality let the supremum over \( t \) in the definition of \( \alpha_A \) be attained, at \( t_0 \), and the supremum over \( \Delta \) in the definition of \( \Delta_A \) be attained; otherwise \( \epsilon_1 \) and \( \epsilon_2 \) would have to be added in several places below, and later picked arbitrarily close to zero.

The minimax risk for testing between \( H_0 : F_{s t_0} \) and \( H_1 : F_{s t_0 + \Delta} \) is \( \alpha \). It follows that for any test statistic

\[
\alpha \leq \sup_{F_0 \in [F_{s t_0}, F_{s t_0 + \Delta}]} P_{F_0} (\text{reject } H_0) + P_{F_1} (\text{accept } H_0)
\]

so

\[
\alpha/2 \leq \sup_{F_0 \in [F_{s t_0}, F_{s t_0 + \Delta}]} \max (P_{F_0} (\text{reject } H_0), P_{F_1} (\text{accept } H_0)).
\]

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

\[
P_{F_0} (|T_n - T (F)| \geq \Delta/2) \geq P_{F_0} (T_n - T (F) > \Delta/2) = P_{F_0} (\text{reject } H_0)
\]

and similarly

\[
P_{F_1} (|T_n - T (F)| \geq \Delta/2) \geq P_{F_1} (\text{accept } H_0).
\]

Combining the last 3 displays gives
as $F_0, F_1 \in \mathbf{F}$ and $T_\alpha$ was arbitrary, (2.3) is proved. 

**Corollary.** For each $\alpha$ in $(0, 1)$, $\Delta_\alpha (n, \alpha)$ is a bound on the rate of convergence: for any symmetric increasing loss function $l(t)$,

$$
\inf_{T_\alpha} \sup_{F \in \mathbf{F}} E_F l(T_\alpha - T(F)) \geq l(\Delta_\alpha (n, \alpha)/2) \cdot \alpha/2
$$

for all $n$.

The reader should note that $\Delta_\alpha$ is (nearly) the best lower bound derivable by a testing argument. Indeed, for each $t$, $\Delta$, and $n$, there exists a test between $F_{\leq t}$ and $F_{\geq t + \Delta}$ 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 we have stated it 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 (1987). 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 don't 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). The minimax test for a given $n$, $t$, and $\Delta$ may be thought of as follows: It defines an *acceptance region*, a measurable set $A = A(t, n, \Delta) \subset R^n$, such that if the sample $X_1, X_2, \ldots, X_n$ falls in $A$, we accept $H_0; F_{\leq t}$; otherwise we reject $H_0$. The existence of such tests allows us to "construct" an estimator. This estimator is not intended to be implemented on a computer; but its finite, concrete character allows us to demonstrate that the bound $\Delta_\alpha (n, \alpha)$ can be (nearly) attained in great generality. In order to guarantee that a
minimax test has the indicated form, it is convenient to assume all the elements of \( F \) are absolutely continuous with respect to a fixed measure.

The Binary Search Estimator

The estimator we propose requires that \( T \) be bounded on \( F \): \( M = \sup_F |T(F)| < \infty \). The estimator has a "tuning constant" \( \Delta \), which will depend in a prescribed way with sample size. At a given sample size \( n \), \( \Delta \) 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, i.e. we test \( F_{\leq M/3} \) against \( F_{M/3} \). 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 \( 2/3 \) as long as the previous one; we arrive after \( N \) stages at an interval \([l_0, h_0]\) of length \( \Delta \), and we pick as our estimate \( T_n \) the midpoint of this interval. The key result about the behavior of this procedure is

**Lemma 2.2.** Apply the binary search estimator with parameters \( \Delta, M, \) and \( N \). Set \( \eta_0 = \frac{\Delta}{2} \), and

\[
\eta_k = (\frac{3}{2})^k \Delta \text{ for } k \geq 1. \text{ Set } d_k = \frac{1}{2} (\frac{3}{2})^k \Delta \text{ for } k = 0, 1, \ldots. \text{ Then }
\]

\[
\sup_F P_F \{ |T_n - T(F)| > \eta_k \} \leq \sum_{i=k}^{N-1} \alpha_A(n, d_i).
\]

(2.6)

In particular,

\[
P_F \{ |T_n - T(F)| > \frac{\Delta}{2} \} \leq \sum_{k=0}^{N-1} \alpha_A(n, \frac{1}{2} (\frac{3}{2})^k \Delta)
\]

which makes an interesting comparison with (2.3). Below we will see that under (2.9), \( \alpha_A(n, \frac{1}{2} (\frac{3}{2})^k \Delta) \) decreases 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 $(\Delta, N)$:

\begin{align*}
k &:= N \\
l_N &:= -\frac{1}{2} \left( \frac{3}{2} \right)^N \Delta \\
h_N &:= \frac{1}{2} \left( \frac{3}{2} \right)^N \Delta
\end{align*}

while $k > 0$ do

\begin{align*}
a_k &:= l_k + \frac{1}{3} (h_k - l_k) \\
b_k &:= l_k + \frac{2}{3} (h_k - l_k)
\end{align*}

Test $H_0: F_{l_a}$ against $H_1: F_{l_b}$

if $Accept$ $H_0$ then /* new interval is $(l, b)$ */

\hspace{1cm} l_{k-1} := l_k ; \quad h_{k-1} := b_k

if $Reject$ $H_0$ then /* new interval is $(a, h)$ */

\hspace{1cm} l_{k-1} := a_k ; \quad h_{k-1} := h_k

$k := k-1$

end while

$T_\alpha = (l_0 + h_0)/2$

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 $\{(l^*_k, h^*_k), k = 0, \cdots, N\}$ and $\{(a^*_k, b^*_k), k = 1, \cdots, N\}$, all functionals of $F$.

Consider now the tests $\xi_1, \cdots, \xi_N$, with $\xi_k$ minimax for testing

$H_0: F_{l_a}$ versus $H_1: F_{l_b}$.

The probability that $\xi_k$ decides incorrectly is

$$
\pi(\text{conv}(F_{\xi_k}), \text{conv}(F_{\xi_k}^*) ) \leq \alpha_A(n, b^*_k - a^*_k) = \alpha_A(n, a_{k-1})
$$

Consider now (2.6), and let $k > 0$. If the tests $\xi_i$ all decide correctly for $i = k+1, \cdots, N$, then $T_\alpha \in (l^*_k, h^*_k)$ and so $|T_\alpha - T(F)| \leq h^*_k - l^*_k = \eta_k$. Therefore,

$$
P(|T_\alpha - T(F)| > \eta_k) \leq P(\bigcup_{i=k+1}^N \{\xi_i \text{ decides incorrectly}\}) \leq \sum_{i=k+1}^N P(\xi_i \text{ decides incorrectly})
$$
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 \( q > 0 \) and \( 0 < A_0 \leq A_1 < \infty \) so that

\[
A_0 \left( \frac{\log \alpha}{n} \right)^{q/2} \leq \Delta_A(n, \alpha) \leq A_1 \left( \frac{\log \alpha}{n} \right)^{q/2} \tag{2.9}
\]

for \( |\log \alpha|/n < \varepsilon_0 \). Pick \( n_0 \) so that \( |\log \alpha|/n_0 < \varepsilon_0 \) and \( \alpha_0 = \alpha_A(n_0, A_0^2/A_1 \varepsilon_0^2) < 1 \). Define

\[
\beta = \frac{1}{4} \left( \log(2-\alpha) \right) \left( \frac{A_1}{A_0} \right)^{q/2} \quad \text{and} \quad \gamma = \frac{1}{4n_0} \left( \log(2-\alpha) \right).
\]

Then with \( \Delta = C \Delta_A(n, \alpha) \) and \( d_i \) as in Lemma 2.2

\[
\begin{align*}
\sum_{i=0}^{N-1} \alpha_A(n, d_i) & \leq \frac{2 \theta}{1-\theta^2} + r_n \quad (2.10a) \\
\sum_{i=k}^{N-1} \alpha_A(n, d_i) & \leq \frac{\theta^{2k}}{1-\theta^2} + r_n \quad (2.10b)
\end{align*}
\]

for \( n > 2n_0 \), where

\[
\theta = \exp(-C^2 \beta) \quad \text{and} \quad r_n = \frac{\log(3M)}{\log(A_0 \varepsilon_0^2)} \exp(-n \gamma) \quad (2.11)
\]

The proof is given in section 7. In view of (2.6), these bounds imply that for the binary search estimator with parameters \( \Delta, M, \) and \( N \), we can have \( \mathbb{P}(|T^n_n - T| > K \Delta_A(n, \alpha)) \) 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

\[
E_F l(T^n_n - T) \leq A \cdot l(\Delta_A(n, \alpha)) \quad n > n_1 \quad (2.12)
\]

for every \( F \in \mathcal{F} \), where
\[ A = \frac{28a}{1-\theta^2} (2 + 8a) (a \log C / \log 1.5) \]  

(2.13)

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

Corollary. Under the assumptions of Theorem 2.4,

\[
\inf_{T_a} \sup_{F} E_F l(T_a - T) \asymp l(\Delta_A(n, \alpha))
\]

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), (1985) and by Birgé (1983). The Le Cam-Birgé 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 \( F \) 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 $\Delta_A$ 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 $F$ is convex. Fix $\varepsilon_0 \in (0,1)$ and $\alpha_0 \in (0,1)$. Then for $\alpha \leq \alpha_0$ and $|\log \alpha|/n < \varepsilon_0$ there exist universal constants $c, C$ with

$$b(c \sqrt{\frac{|\log \alpha|}{n}}) \leq \Delta_A(n, \alpha) \leq b(C \sqrt{\frac{|\log \alpha|}{n}}).$$

(3.1)

We may take $C = \sqrt{2}$ and $c = 1/2$, for $\alpha_0, \varepsilon_0$ small enough.

If $b$ is Holderian, (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 2-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

$$\alpha_2(n, \Delta) = \sup_i \pi(F^{(n)}_{x_i}, F^{(n)}_{x_i + \Delta}).$$

(3.2)

Note the omission of the convex hull operation in comparison with the definition (2.2) of $\alpha_A$. Similarly, let $\Delta_2(n, \alpha)$ be the inverse function of $\alpha_2$. We can also write

$$\Delta_2(n, \alpha) = \sup \{|T(F, \alpha) - T(F_0)| : \pi(F^{(n)}_{x_i}, F^{(n)}_{x_0}) \geq \alpha\}.$$ 

(3.3)

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

$$\Delta_2(n, \alpha) \leq \Delta_A(n, \alpha);$$

(3.4)

as $\Delta_A$ has the lower bound property (2.3), it follows that $\Delta_2$ is a lower bound as well. Thus, (2.3) holds with $\Delta_2$ in place of $\Delta_A$. One could also argue directly -- compare Theorem 2.1 of [GR I].
One can say more; \( \Delta_2 \) is (nearly) the best possible two-point testing bound. Thus, for a given \( n \) and \( \alpha \), the largest \( \delta \) for which there exists a pair \((F_0, F_1)\) with \( T(F_1) - T(F_0) \geq \delta \), and which cannot be distinguished by the best test with sum of errors better than \( \alpha \), is precisely \( \Delta_2(n, \alpha) \). No 2-point bound on the maximum probability of error can exceed \( \alpha \), 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 \( |\log \alpha|/n < \varepsilon_0 \),

\[
\frac{c}{\log \alpha} \leq \Delta_2(n, \alpha) \leq \frac{C}{\log \alpha}
\]

We may take \( C = \sqrt{2} \) and \( c = 2 \epsilon_0 \cdot \frac{\log (2-\alpha_0)}{\log \alpha_0} \).

Before giving the proof, we need some facts from Le Cam (1973), (1985, Chapter 4). First, recall the Hellinger Affinity

\[
\rho(P, Q) = \int \sqrt{p} \sqrt{q} \, d\mu
\]

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

\[
\pi(P, Q) \leq \rho(P, Q), \quad \rho^2 \leq \pi(2-\pi)
\]

where \( \pi \) is the testing affinity, and the identity

\[
\rho(P, Q) = \frac{1}{2} (2 - H^2(P, Q))
\]

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

\[
\rho(P^{(n)}, Q^{(n)}) = \rho(P, Q)^n
\]

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

**Proof.** Define

\[
h_0(n, \alpha) = \inf \{ H(F_1, F_0) : \pi(F_1^{(n)}, F_2^{(n)}) \leq \alpha \}
\]

and
Using (3.6)-(3.9), we have the easy inequalities

\[ h_0^2 (n, \alpha) \geq 2 \left( 1 - (\alpha(2-\alpha))^{1/2n} \right) \]
\[ h_1^2 (n, \alpha) \leq 2 \left( 1 - \alpha^{1/n} \right) \tag{3.10} \]

Combining these with the definition of \( b(\varepsilon) \), we have

\[ b(h_0(n, \alpha)) \leq \Delta_2(n, \alpha) \leq b(h_1(n, \alpha)). \tag{3.12} \]

The result then follows by (3.13) and (3.14) below. \( \square \)

Lemma 3.3.

\[ (1 - \alpha^{1/n}) \leq \frac{|\log \alpha|}{n}, \tag{3.13} \]

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

\[ (1 - (2\alpha)^{2n}) \geq c^{2/2} \frac{|\log \alpha|}{n} \tag{3.14} \]

We may take \( c^{2/2} = \frac{(1-\varepsilon_0)}{\varepsilon_0} \cdot \frac{\log (2-\alpha_0) \alpha_0}{|\log \alpha_0|} \). This result is proved in the appendix, section 7.

In particular, if \( b(\varepsilon) \) is Holderian, then \( b(n^{-1/2}) \) is equivalent, to within constants, with \( \Delta_2(n, \alpha) \). And so the question of the attainability, as regards rate, of \( b(n^{-1/2}) \) is equivalent to the attainability of the best 2-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 \( F \).

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 (LeCam, 1985, Chapter 16, page 477). Let \( P \) and \( Q \) denote sets of probabilities and \( P^{(\alpha)} \), \( Q^{(\alpha)} \) the sets of corresponding product measures. Then

\[ \rho(\text{conv } P^{(\alpha)}, \text{conv } Q^{(\alpha)}) \leq \rho(\text{conv } P, \text{conv } Q)^n. \tag{3.15} \]

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

\[
\alpha_A(n, \Delta) \leq \sup_i \rho(\text{conv}(F_{s_t}), \text{conv}(F_{s_t+\Delta}))^a.
\]  

(3.16)

Thus the Hellinger Distance between the convex hulls of \(F_{s_t}\) and \(F_{s_t+\Delta}\) may be used to bound \(\alpha_A\).

The upper bound in (3.1) follows more or less directly from this. To see how, notice that

\[
\varepsilon = \inf_i H(F_{s_t}, F_{s_t+b(e)}).
\]  

(3.17)

Combining this with (3.8) we have

\[
\rho(F_{s_t}, F_{s_t+b(e)}) \leq 1 - \varepsilon^2/2.
\]  

(3.18)

Now, and this is the key observation, if \(T\) is a linear functional, and if \(F\) is convex, then \(F_{s_t}\) and \(F_{s_t+\Delta}\) are both convex, for all \(t\) and all \(\Delta\). Thus \(F_{s_t} = \text{conv} F_{s_t}\) and \(F_{s_t+b(e)} = \text{conv} F_{s_t+b(e)}\); combining (3.16) and (3.18),

\[
\alpha_A(n, b(e)) \leq (1 - \varepsilon^2/2)^a
\]  

(3.19)

and so

\[
\Delta_A(n, \alpha) \leq b(\sqrt{2(1 - \alpha^{1/\alpha})}).
\]  

(3.20)

At this point we invoke again Lemma 3.3. Equation (3.13), combined with (3.20), gives 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 \( b(n^{-1/2}) \) and \( \Delta_A(n, \alpha) \) is not to be expected. It requires essentially that the hardest two-point subproblem of testing \( F_{<t} \) versus \( F_{<t+\Delta} \) be roughly as hard as the full problem. Let us see how.

4.1. The Minimax Identity

The 2-point testing bound and the attainable bound have an interesting connection. As (3.4) shows, the 2-point bound is always smaller; as (3.1) and (3.5) make plain, when \( T \) is linear and \( F \) is convex

\[
\Delta_A(n, \alpha) \leq C \Delta_2(n, \alpha)
\]

for an appropriate constant \( C \), for small \( \alpha \) and large \( n \).

It seems natural to ask if the 2-point and the attainable bounds can ever agree, i.e. 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

\[
\pi(\text{conv}(F_{<t}^{(n)}), \text{conv}(F_{<t+\Delta}^{(n)})) = \pi(F_{<t}^{(n)}, F_{<t+\Delta}^{(n)});
\]

Indeed, the quantity on the left hand side is the main ingredient in the definition of \( \Delta_A \), while that on the left is the main ingredient in \( \Delta_2 \). Now if we return to the definition of \( \pi \) as a measure of the difficulty of testing, we see that the quantity on the left is

\[
\inf_{\zeta} \sup_{F_0 \in F_{<t}} \inf_{F_1 \in F_{<t+\Delta}} R_A(\zeta(F_0, F_1))
\]

where \( R_A(\zeta(F_0, F_1)) \) is the “risk” \( E_{F_0}(\zeta) + E_{F_1}(1 - \zeta) \) representing the sum of errors of the test \( \zeta \).

This is the minimax risk for the problem of testing the composite hypotheses \( F_{<t} \) versus \( F_{<t+\Delta} \). On the other hand, the quantity on the right of (4.2) is

\[
\sup_{F_0 \in F_{<t}} \inf_{\zeta} R_A(\zeta(F_0, F_1)) \inf_{F_1 \in F_{<t+\Delta}}
\]

This is the risk of the hardest 2-point testing problem. Consequently, the identity (4.2) is equivalent to the minimax identity.
\[
\inf_{\zeta} \sup_{(F_0,F_1)} R_n (\zeta, (F_0,F_1)) = \sup_{(F_0,F_1)} \inf_{\zeta} R_n (\zeta, (F_0,F_1)). \tag{4.3}
\]
This identity says, in words, that the minimax risk in testing between the infinite dimensional composite hypotheses \( F_{\pm} \) and \( F_{\pm + \Delta} \) is precisely the risk of the hardest 2-point testing problem.

We will see below 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 \( \Delta_0 = \Delta_2 \) for large \( n \), and \( b(n^{-1/2}) \) represents the optimal rate of convergence of an estimate \( T_n \) to \( T \).

Indeed, the conclusion that \( \Delta_0 = \Delta_2 \) follows from the definition of these quantities, and the conclusion that \( b(n^{-1/2}) \) is the optimal rate follows from (3.5), and Theorem 3.1.

It does happen that (4.2) holds in interesting examples.

**Theorem 4.2.** Let \( T(F) = f(0) \) and let \( F \) be the Sacks-Ylvisaker (1981) class

\[
SY = \{ f : f(x) = f(0) + x f'(0) + r(x), \quad f(0) \leq M, \quad \int f = 1, \quad f \geq 0, \quad |r(x)| \leq x^2/2 \}.
\]
(Here we must have \( \frac{4\sqrt{2}}{3} M^{3/2} < 1 \). Then for every \( t \) and \( n \) and every \( \Delta \) small enough, the minimax identity (4.2) holds, and so \( \Delta_0 = \Delta_2 \) for large \( n \).

It is known that in general, one cannot expect (4.2) to hold. One case when (4.2) does hold is when the sets \( F_{\pm} \) and \( F_{\pm + \Delta} \) 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 above, says that convexity alone is enough to guarantee that a certain approximate minimax identity holds.

**Lemma 4.3.** If \( F_{\pm} \) and \( F_{\pm + \Delta} \) are both convex,

\[
\rho(\text{conv} (F_{\pm}^{(n)}), \text{conv} (F_{\pm + \Delta}^{(n)})) = \rho(F_{\pm}^{(n)}, F_{\pm + \Delta}^{(n)}). \tag{4.4}
\]
This says that, although (4.2) may not hold when just convexity is assumed, its analog, with \( n \) replaced by \( \rho \), does hold.
Proof. We have

$$
\rho(F_{t}, F_{2t+\rho}) \geq \rho(\text{conv}(F_{t}^{(n)}), \text{conv}(F_{2t+\rho}^{(n)}))
$$

$$
\geq \rho(F_{t}^{(n)}, F_{2t+\rho}^{(n)})
$$

$$
= \rho(F_{t}, F_{2t+\rho})^n,
$$

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

$$
\pi \leq \rho, \quad \rho^2 \leq \pi (2-\pi)
$$

(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

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

(4.5)

Thus, at every $n$ and $\Delta$ for which $\alpha_A(n, \Delta) \leq \alpha_0 < 1$, we can bound the discrepancy between $\alpha_2$ and $\alpha_A$.

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 $F$ is convex.

4.2. A near-equivalence

In the case where $T$ is linear and $F$ convex, we have seen that $\Delta_A \leq C \Delta_2$ and also that $||\log \alpha_2|| \leq M ||\log \alpha_A|| + D$. In general, whatever be $T$ and $F$, 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 $b(n^{-1/2})$ 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 two formal results; they are proved in section 7.

Theorem 4.4. Suppose that $b(\varepsilon)$ is Holderian with exponent $q \in (0,1]$. Then there are constants $\alpha_0 \in (0,1/2)$ and $\varepsilon_0 \in (0,1)$ with the following property. If there exists a finite positive $M$ such that

$$
\frac{||\log \alpha_2(n, \Delta)||}{||\log \alpha_A(n, \Delta)||} \leq M \quad \alpha_A(n, \Delta) \leq \alpha_0,
$$

(4.6)

then there exists a finite positive $C$ such that
\[ \Delta_A(n, \alpha) \leq C \Delta_2(n, \alpha) \quad \alpha \leq \alpha_0, \quad |\log \alpha|/n < \varepsilon_0. \]

**Theorem 4.5.** Suppose that \( b(\varepsilon) \) is Holderian with exponent \( q \in (0,1) \). Then there are constants \( \alpha_0 \in (0,1/2) \) and \( \varepsilon_0 \in (0,1) \) with the following property. If there exists a finite positive \( C \) such that
\[ \Delta_A(n, \alpha) \leq C \Delta_2(n, \alpha) \quad \alpha < \alpha_0 < 1, \quad |\log \alpha|/n < \varepsilon_0. \]
then there exists a finite positive \( M \) such that
\[ \frac{|\log \alpha_2(n, \Delta)|}{|\log \alpha_A(n, \Delta)|} \leq M \quad \alpha_A(n, \Delta) \leq \alpha_0. \quad (4.7) \]

In the sequel [GR III] and in Donoho and Liu (1988c) we give examples where a minimax identity is key to attainability of the modulus at the level of *constants*. Compare also Ibragimov and Hasminskii (1984); this paper, although it does not use the modulus of continuity, shows a connection between a minimax theorem and precise evaluation of constants in certain nonparametric estimation problems.

**5. Attainability in two nonlinear cases**

In this section we study two nonlinear functionals in order to see how the ideas of the preceding sections carry over. In the first example, the minimax identity (4.2) holds, and everything flows automatically. In the second example, (4.2) fails, and we must work hard with our bare hands.

**5.1. Estimating Tail Rates**

While it is most intuitive 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 \( F \) be the set of distributions supported on \((0,\infty)\) with densities \( f \) satisfying
\[ f(x) = Cx^t(1+r(x)) \quad 0 \leq x \leq \delta \quad (5.1) \]
with
\[ 0 < t_0 \leq t \leq t_1 < \infty \quad (5.2a) \]
and
\[ 0 < C_- \leq C \leq C_+ \leq \infty \] (5.2b)

and

\[ |r(x)| \leq c_2 x^p \] (5.2c)

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

Consider now the problem of testing \( F_0 \) against \( F_{2t+\Delta} \). In [GR I] we have shown that the closest pair in a Hellinger sense has the form

\[
\begin{align*}
    f_0^*(x) &= C_- x^t (1 - c_2 x^p) & x \leq a_1(t, \Delta) \\
    f_1^*(x) &= C_+ x^t \Delta (1 + c_2 x^p) & x \leq a_1(t, \Delta)
\end{align*}
\] (5.3)

and

\[
\frac{f_0^*(x)}{f_1^*(x)} = \frac{f_0^*(a_1)}{f_1^*(a_1)} \quad x > a_1
\] (5.5)

As we will see, this closest Hellinger pair represents the 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 pair \((F_0^*, F_1^*)\) described above, we have

\[
\pi(\text{conv } (F_0^*), \text{conv } (F_{2t+\Delta}^*)) = \pi(F_{2t}^*, F_{2t+\Delta}^*) = \pi((F_0^*)^a, (F_1^*)^a)
\]

and the minimax test between \( F_{2t}^* \) and \( F_{2t+\Delta}^* \) 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

\[
L_{t, \Delta}(x) = \begin{cases} 
    C_+ x^\Delta (1 + c_2 x^p) & 0 < x < a_1 \\
    C_- x^\Delta (1 - c_2 x^p) & x > a_1
\end{cases}
\] (5.7)

This is a non-decreasing function of \( x \).

Among all distributions in \( F_{2t}^* \), \( F_0^* \) is the stochastically largest. Similarly, among all distributions in \( F_{2t+\Delta}^* \), \( F_1^* \) is the stochastically smallest. 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 iid \( F \). Consider the likelihood ratio statistic

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

Under \( H_0 ; F_0 \Delta \), this statistic is then stochastically largest at \( F = F_0^* \), etc. Therefore, if we consider accepting \( H_0 \) when \( L_{n, t, \Delta} \leq 1 \) and rejecting when \( L_{n, t, \Delta} > 1 \), we have

\[
\sup_{F \in F_0^*} P_F \{ \text{Reject } H_0 \} = P_{F_0^*} \{ \text{Reject } H_0 \}
\]

\[
\sup_{F \in F_1^*} P_F \{ \text{Accept } H_0 \} = P_{F_1^*} \{ \text{Accept } H_0 \}
\]

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. \( \square \)

As we show in [GR I], the modulus is in this case not Holderian, so that Lemma 4.1 in this case does not apply. However, we can use the minimax identity to show attainability in a different way. An extra level of structure in the minimax tests of section 2 may exist which we have not previously considered: monotonicity in \( t \). We can state this in terms of acceptance regions as

\[
A(t, n, \Delta) = A(t + h, n, \Delta) \quad \forall h > 0. \tag{5.9}
\]

The following result is proved in section 7.

**Theorem 5.2.** For all sufficiently small \( \Delta \), 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

\[
A(t, n, \Delta) = \{(X_i)_{i=1}^n : \prod_{i=1}^{n} L_{t, \Delta}(X_i) \leq 1 \}
\]

with the monotonicity property (5.9).

Consider what we call the likelihood ratio estimator

\[
T_{n, \Delta}^* = \frac{\Delta}{2} + \sup \{ t : \prod_{i=1}^{n} L_{t, \Delta}(X_i) \geq 1, \quad t \in [t_0, t_1 - \Delta] \}. \tag{5.10}
\]

By the monotonicity established in Theorem 5.3, \( T_{n, \Delta}^* \) is always uniquely defined.
Theorem 5.3. Suppose that \( L_{t,A}(x) \) is monotone decreasing in \( t \) for each fixed \( x \). Then

\[
\sup_F \left\{ \left| T_{n,\Delta}^* - T(F) \right| > \frac{\Delta}{2} \right\} \leq 2 \alpha_A(n, \Delta)
\]  

(5.11)

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

Proof. By the monotonicity in \( t \) of \( L_{t,A} \),

\[
T_{n,\Delta}^* - T(F) > \Delta/2
\]

happens if and only if the minimax test between \( H_0 : F_{tT(F)} \) and \( H_1 : F_{tT(F)+\Delta} \) 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 \( |T_{n,\Delta}^* - T(F)| > \frac{\Delta}{2} \) is the union of these two events, (5.11) follows. \( \square \)

Thus, in this case, the lower bound \( \Delta_A (=\Delta_2) \) is achievable within a factor 4. We consider it likely that the factor 2 on the right hand side of (5.11) can be dropped (asymptotically).

5.2. Estimating the Mode

Now let \( F \) be the class of distributions with unimodal densities \( f \), that are uniformly bounded:

\[
f(x) \leq M
\]

(5.12)

and have quadratic maxima:

\[
f(\text{mode}) - c_x x^2 \leq f(x) \leq f(\text{mode}) - c_x x^2 \quad |x - \text{mode}| < \delta.
\]

(5.13)

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

In [GR I] the modulus was computed for this problem, and so the closest pair in Hellinger distance was derived; it has the form (for \( \Delta \) small enough)

\[
\begin{align*}
  f_0^* (x) &= M - c_x (x-t)^2 \quad x \in (t-a_2(\Delta), t+a_3(\Delta)) \\
  f_0^* (x) &= M - c_x (x-t)^2 \quad x \in (t+a_3(\Delta), t+\Delta-a_3(\Delta)) \\
  f_0^* (x) &= M - c_x (x-t)^2 \quad x \in (t+\Delta-a_3(\Delta), t+\Delta+a_2(\Delta)). \\
  f_1^* (x) &= f_0(2(t+\Delta/2)-x) \\
  f_2^* (x) &= f_1^*(x) \quad x \in (t-a_2(\Delta), t+\Delta+a_2(\Delta))
\end{align*}
\]

(5.14a)

(5.14b)

(5.14c)

(5.15)

(5.16)
This closest Hellinger pair probably represents the hardest 2-point testing problem. A proof based on stochastic minorization, as in Theorems 4.2 and 5.1, will not quite work here, however. On the other hand, one can show, using the convergence of experiments approach of [GR III], section 7, that this pair is asymptotically hardest. That is, if we set $\Delta = cn^{-1/5}$, then for large $n$, we will have

$$\pi((F_0^*)^{(a)}, (F_1^*)^{(a)}) = \pi(F_{2a}^{(a)}, F_{2a+\Delta}^{(a)}) \cdot (1 + o(1)).$$

However, the minimax identity (4.2) definitely does not hold in this case. Consider using the likelihood ratio of this pair to test $F_0$ against $F_1$, where $F_0$ has its mode at $t-2\delta$ and $F_1$ has its mode at $t+2\delta$, and both $f_0$ and $f_1$ are equal on the interval $(t - \frac{\Delta}{2}, t+\Delta)$. Then $F_0 \in F_{2a}$ and $F_1 \in F_{2a+\Delta}$, but the likelihood ratio statistic based on $f_1^*(x)/f_0^*(x)$ has the same distribution under $F_0$ as under $F_1$. Consequently, the worst sum of type I and type II errors of this test is 1.

Although the likelihood ratio between $F_0^*$ and $F_1^*$ cannot furnish us with a useful test, it does suggest a useful estimator. Roughly speaking, for small $\Delta$, the likelihood ratio test decides in favor of $F_0^*$ if

$$\sum K_n(X_i - t) \geq \sum K_n(X_i - (t+\Delta))$$

and in favor of $F_1^*$ if

$$\sum K_n(X_i - t) < \sum K_n(X_i - (t+\Delta)),$$

where

$$K_n(u) = \log \left[ \frac{M - c_n u^2}{M - c_n a_3(\Delta)^2} \right] \quad |u| < a_3(\Delta),$$

$$= 0 \quad \text{else}.$$  

This suggests the estimator

$$T^*_{n,\Delta} = \arg \max \frac{1}{n} \sum K_n(X_i - t).$$

(Note that the maximum need not be unique). This estimator is closely related to estimating a density using Epanechnikov’s Kernel with bandwidth $a_3(\Delta)$ and setting $T^*_{n,\Delta}$ to be (any) maximizer of the estimated density. Let us analyze the behavior of any such Kernel estimate of the mode. Our rate result applies to any kernel satisfying
Assumption (K). $K$ is a positive, even function of compact support, bounded, square integrable, and absolutely continuous, with

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

and

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

where the norms of $K'$ are defined distributionally, and so represent the smallest constants $C_2$ and $C_\infty$ for which

$$\|K(-\delta) - K(-\delta)\|_2 \leq C_2 \delta$$

$$\|K'(-\delta) - K'(-\delta)\|_\infty \leq C_\infty \delta$$

are valid.

**Theorem 5.4.** Let $T$ be the mode, and $F$ be as in (5.12)-(5.13). Then $b(t)$ is Holderian with exponent $2/5$, and so no estimator can achieve faster than an $n^{-1/5}$ rate of convergence uniformly over $F$:

$$\lim \inf \inf \sup_{n \to \infty} P_F \{ |T_n - T| > b(n^{-1/2}) \} \geq e^{-1/2}/2$$

Let $K$ satisfy the assumption (K) above and let $h_n = cn^{-1/5}$. Let $T_{n}^{(k)}$ be any maximizer of

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

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

$$\lim \inf \lim \sup_{C \to \infty} \sup_{n \to \infty} P_F \{ |T_n^{(k)} - T| > Cb(n^{-1/2}) \} = 0$$

This is proved in section 7.

It follows from this theorem that the estimator $T_{n,\Delta}$, properly tuned, achieves the optimal rate of convergence, although its actual performance is not measured by $\pi(F_{x_1}, F_{x_2})$. This may be understood as follows. Roughly speaking, $\pi(F_{x_1}, F_{x_2})$ is comparable to

$$\sup_{u > \Delta} \sup_{F} P_F \{ \frac{1}{n} \sum_{i=1}^{n} K_n(X_i - T(F)) < \frac{1}{n} \sum_{i=1}^{n} K_n(X_i - (T(F) + u)) \}$$

(5.17)

whereas $\pi(\text{conv}(F_{x_1}), \text{conv}(F_{x_2}))$ is comparable to

$$\sup_{u > \Delta} \sup_{F} P_F \{ \frac{1}{n} \sum_{i=1}^{n} K_n(X_i - T(F)) < \sup_{u > \Delta} \frac{1}{n} \sum_{i=1}^{n} K_n(X_i - (T(F) + u)) \}$$

(5.18)

Underlying the proof of Theorem 5.4 is the idea that although (5.18) is much larger than (5.17), they
are still comparable.

Hasminskii (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. Hasminskii 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 the Lemmas 7.1 and 7.2 proved below, and some facts about \(\|f_n - Ef_n\|_\infty\) 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 \(\mathcal{F}\). However, to show that \(n^{-1/5}\) is the optimal rate in probability seems genuinely harder; here we do this by using Bernstein's inequality and a chaining argument. Thus, Theorem 5.4 verifies Hasminskii's claim and shows that \(n^{-1/5}\) is the optimal rate for estimation of the mode over the class \(\mathcal{F}\).

6. An interesting example

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

In a very penetrating analysis, Ritov and Bickel (1987) 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 \(\mathcal{F}\) as above,

\[
\alpha_A(n, \Delta) = 1 \quad \text{for all } n \text{ and } \Delta \in (0,(M-1)/2).
\]

\[
\Delta_A(n, \alpha) \geq (M-1)/2 > 0 \quad \text{for all } n.
\]

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

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

but
\[ \Delta_A(n, \alpha) \to 0; \]
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 2-point and composite bounds are not compar-
able.
7. Proofs

Proof of Theorem 2.3

Before proving (2.10a-b), we first establish some exponential bounds on $\alpha_A(n,d_k)$. We consider two cases, depending on $k$. For $k$ small,

\[
\left(\frac{3}{2}\right)^k C A_1\left(\frac{\log \alpha}{n}\right)^{q/2} \leq A_0 \varepsilon^{l/2} \quad \text{(Case 1)}
\]

In this case, there exists an integer $m$ satisfying

\[
n \left( \frac{A_0}{A_1(\frac{3}{2})^k C} \right)^{2q} \geq m,
\]

and $|\log \alpha/m| \leq \varepsilon_0$. Then a calculation reveals that

\[
A_0\left(\frac{|\log \alpha|}{m}\right)^{q/2} \geq \left(\frac{3}{2}\right)^k C A_1\left(\frac{\log \alpha}{n}\right)^{q/2}
\]

and, as $|\log \alpha/m| \leq \varepsilon_0$, (2.9) implies

\[
\Delta_A(m,\alpha) \geq d_k
\]

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

\[
\pi_{jm} \leq (\sqrt{\pi_m}(2-\pi_m))^{j/2}
\]

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

\[
\alpha_A(j \cdot m,d_k) \leq (\sqrt{2-\alpha})^{j/2}.
\]

Putting $j = \lfloor n/m \rfloor$ and using monotonicity of $\alpha_A$ in $n$ we get

\[
\alpha_A(n,d_k) \leq (\sqrt{2-\alpha})^{n-1}
\]

\[
= \exp(-\frac{1}{2}|\log(2-\alpha)|)(\frac{n}{m} - 1)
\]

\[
\leq \exp\left\{ -\frac{1}{2}|\log(2-\alpha)| \left[ \left( \frac{3}{2} \right)^k C A_1 \right]^{2q} \right\}
\]

\[
= \exp\left\{ -2\beta C^{2q} \left( \frac{3}{2} \right)^{2k/q} + \frac{1}{2}|\log(2-\alpha)| \right\}
\]

The hypothesis $\frac{CA_1}{A_0} > 2^{l/2}$ implies
\[2\beta C^{2q} \left( \frac{3}{2} \right)^{2k/q} - \frac{1}{2} \log(\alpha(2 - \alpha)) | > \beta C^{2q} \left( \frac{3}{2} \right)^{2k/q}\]

for \(k = 0, 1, 2, \cdots\) and so we have, in Case 1

\[\alpha_A(n, d_k) \leq \exp(-\beta C^{2q} \left( \frac{3}{2} \right)^{2k/q}) \tag{7.1}\]

In Case 2, \(k\) is so large that condition (Case 1) does not hold. It follows that

\[d_k > \frac{A_0}{A_1} e^{\Phi/2}; \tag{Case 2}\]

using the definition of \(n_0\) and arguing as above,

\[\alpha_A(n, d_k) \leq \left( \sqrt{\alpha_0(2 - \alpha_0)} \right)^{n_0} = \exp(-2\gamma n + \frac{1}{2} \log(\alpha_0(2 - \alpha_0)))\]

Now as \(2 \left( \frac{n}{n_0} - 1 \right) > \frac{n}{n_0}\) for \(n \geq 2n_0\), we get \(2\gamma n - \frac{1}{2} \log(\alpha_0(2 - \alpha_0)) > \gamma n\), and so

\[\alpha_A(n, d_k) \leq \exp(-n\gamma), \quad n > 2n_0. \tag{7.2}\]

Consider now (2.10a). Let \(K\) be the number of \(d_k\) satisfying (Case 2). Formally,

\[K = \# \{ k : A_0 e^{\Phi/2} \leq \left( \frac{3}{2} \right)^k A_1 \left( \frac{1}{n} \log \alpha \right)^{1/2} \leq \left( \frac{3}{2} \right)^N \Delta \} \]

As \(\left( \frac{3}{2} \right)^N \Delta \leq 3M, K \leq \log(3M)/\log(A_0 e^{\Phi/2});\) thus \(K\) is bounded independently of \(n\) and \(N\). Now by (7.1) and (7.2), if \(n > 2n_0\)

\[\sum_{k=0}^{N-1} \alpha_A(n, d_k) = \sum_{k=0}^{N-K-1} + \sum_{k=N-K}^{N-1} \leq \sum_{k=0}^{N-K-1} \exp(-\beta C^{2q} \left( \frac{3}{2} \right)^{2k/q}) + K \exp(-\gamma n)\]

\[\leq \sum_{k=0}^{N-K-1} \exp(-\beta C^{2q} \left( \frac{3}{2} \right)^{2k/q}) + r_n \tag{7.3}\]

If \(l = 0\), then since \(\left( \frac{3}{2} \right)^{2k/q} > 2k\) for \(k = 1, 2, \cdots\),

\[\sum_{k=0}^{N-K-1} \exp(-\beta C^{2q} \left( \frac{3}{2} \right)^{2k/q}) \leq \exp(-\beta C^{2q}) + \sum_{k=1}^{N-K} \exp(-\beta C^{2q} 2k) = 0 + \frac{\theta^2}{1-\theta^2}\]

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

\[\sum_{k=0}^{N-K-1} \exp(-\beta C^{2q} \left( \frac{3}{2} \right)^{2k/q}) \leq \sum_{k=0}^{N-K} \exp(-\beta C^{2q} 2k) = \frac{\theta^2}{1-\theta^2};\]

which, with (7.3) gives (2.10b). □
Proof of Theorem 2.4

\[
E_{\beta} \mathbb{P} (T_{n} - T) \leq \sum_{k=0}^{N-1} \mathbb{P} \{ |T_{n} - T| > \eta_k \} l(\eta_{k+1})
\]

\[
\leq \sum_{k=0}^{N-1} \mathbb{P} \{ |T_{n} - T| > \eta_k \} l(\eta_{k+1})
\]

\[
\leq l(\eta_{1}) \frac{2\theta}{1-\theta^{2}} + \sum_{k=1}^{N-1} l(\eta_{k+1}) \frac{\theta^{2k}}{1-\theta^{2}} + \sum_{0}^{N-1} l(\eta_{k+1}) \cdot r_{n}
\]

Now as \( l \) is well-behaved, \( l(\eta_{k}) = l((3/2)^{k} \Delta) \leq a^{k} l(\Delta) \), so

\[
E_{\beta} l(T_{n} - T) \leq l(\Delta) \left[ \frac{2\theta}{1-\theta^{2}} a + \sum_{k=1}^{N-1} \frac{\theta^{2k} a^{k+1}}{1-\theta^{2}} + \frac{r_{n} a^{N-1}}{a-1} \right]
\]

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

\[
= l(\Delta)(A_{2} + A_{3,n}),
\]

say. Now as \( N \leq \log (3M)/\log (\Delta) \leq \log (3M)/2 (\log(n) - \log|\log(\alpha)|) - \log(\Delta) \),

\[
a^{N} r_{n} = \exp(-n \gamma + \log(a)N) \leq \exp(-n \gamma + A_{4} \log(n) + A_{5}) \to 0, n \to \infty.
\]

Therefore, \( A_{3,n} \to 0 \). For large enough \( n \), \( A_{3,n} \leq A_{2} \), and so \( l(\Delta)(A_{2} + A_{3,n}) \leq 2A_{2} \). Then, as

\[
l(\Delta) = l(C \Delta_{A}(n, \alpha)) \leq a^{[\log C \log 1.5]} l(\Delta_{A}(n, \alpha)),
\]

we have (2.11) with \( A = 2A_{2} a^{[\log C \log 1.5]} \). \( \square \)

Proof of Lemma 3.3

First, we prove (3.13). As \( (1 - \alpha^{1/n}) = 1 - \exp \left[ - \frac{\log(\alpha)}{n} \right] \), (3.13) is equivalent to \( 1 - e^{-\varepsilon} \leq \varepsilon \).

As \( e^{x} \) is a strictly convex function it lies above its tangent line at \( x = 0 \), so

\[
e^{-\varepsilon} \geq 1 - \varepsilon
\]

which is precisely what we need.

Now, we prove (3.14). Note first that if we put

\[
g(\varepsilon) = (1 - e^{-\varepsilon}) / \varepsilon
\]

then for all \( \varepsilon < \varepsilon_{0} < 1.5 \) we have

\[
(1 - e^{-\varepsilon}) > g(\varepsilon_{0}) \varepsilon. \quad (7.4)
\]

Indeed, \( g(\varepsilon) \) is a monotone decreasing function at least on \([0,1.5]\) as may easily be seen from its power
Indeed, \( g(\varepsilon) \) is a monotone decreasing function at least on \([0, 1.5)\) as may easily be seen from its power series

\[
g(\varepsilon) = 1 - \frac{\varepsilon}{2!} + \frac{\varepsilon^2}{3!} - \frac{\varepsilon^3}{4!} + \frac{\varepsilon^4}{5!} - \ldots
\]

\[
= 1 - \frac{\varepsilon}{2!} (1 - \frac{\varepsilon}{3}) - \frac{\varepsilon^3}{4!} (1 - \frac{\varepsilon}{5}) - \ldots
\]

absolutely convergent for all \( \varepsilon > 0 \). So \( g(\varepsilon) > g(\varepsilon_0) \) if \( 0 < \varepsilon < \varepsilon_0 < 1.5 \); but this is just (7.4).

Now by monotonicity of \( \alpha (2 - \alpha) \) on \( 0 < \alpha < 1 \),

\[
\inf_{0 < \alpha < \alpha_0} \frac{\log(\alpha (2 - \alpha))}{\log(\alpha)} = \frac{\log(\alpha_0 (2 - \alpha_0))}{\log(\alpha_0)}.
\]

Consider now (3.14).

\[
\left[ 1 - (\alpha (2 - \alpha))^{1/2n} \right] = \left[ 1 - \exp \left( -\frac{\|\log(\alpha (2 - \alpha))\|}{n} \right) \right]
\]

\[
\geq g(\varepsilon_0) \frac{\|\log(\alpha (2 - \alpha))\|}{n}
\]

\[
\geq g(\varepsilon_0) \frac{\|\log(\alpha_0 (2 - \alpha_0))\|}{\log(\alpha_0)} \cdot \frac{\|\log(\alpha)\|}{n}
\]

\[
\geq c^{2/2} \frac{\|\log(\alpha)\|}{n}
\]

if we put

\[
c^{2/2} = g(\varepsilon_0) \frac{\|\log(\alpha_0 (2 - \alpha_0))\|}{\log(\alpha_0)}
\]

as claimed.

**Proof of Theorem 4.2**

[GR III] gives the pair \((F_1^*, F_0^*)\) with \(F_0^* \in \mathcal{F}_{S\downarrow}\) and \(F_1^* \in \mathcal{F}_{2\uparrow\Delta}\)

\[
H(F_1^*, F_0^*) = \min_{F_0 \in \mathcal{F}_{s\downarrow}, \ F_1 \in \mathcal{F}_{2\uparrow\Delta}} H(F_1, F_0).
\]

This closest pair in Hellinger distance has the form

\[
f_1^*(x) = t + \Delta - x^2 / 2 \quad \text{for } |x| \leq s_1(\Delta)
\]
\[ f_0^* = t + x^2 / 2 \quad |x| \leq s_1(\Delta) \]

where \( s_1(\Delta) = \sqrt{\Delta} (1 + o(1)) \) as \( \Delta \to 0 \), and

\[ \frac{f_1^*(x)}{f_0^*(x)} = \frac{f_1^*(s_1)}{f_0^*(s_1)} \quad |x| > s_1(\Delta). \]

Define the likelihood ratio

\[ L_{t,\Delta}(x) = \frac{f_1^*(x)}{f_0^*(x)}. \]

This is a monotone decreasing function of \( |x| \). Consider the random variable \( L_{t,\Delta}(X) \) where \( X \) has distribution \( F \). By definition of the Sacks Ylvisaker class, under the null hypothesis \( H_0 : F_{S_1} \), min \( (|X|, s_1(\Delta)) \) is stochastically largest at \( F = F_0^* \). Similarly, under the alternative, min \( (|X|, s_1(\Delta)) \) is stochastically smallest at \( F = F_1^* \). Hence \( L_{t,\Delta}(X) \) is stochastically largest under the null at \( F = F_0^* \) and stochastically smallest under the alternative at \( F = F_1^* \). From this point on the proof is the same as the proof of Theorem 5.1.

**Proof of Theorem 4.4**

As \( b \) is Holderian, if we pick \( \epsilon_0 \) small enough, then by (3.5) there exist constants \( C_-, C_+ \) so that

\[ C_-(\frac{|\log \alpha|}{n})^{g/2} \leq \Delta_2(n, \alpha) \leq C_+(\frac{|\log \alpha|}{n})^{g/2} \]

(7.5)

for \( \alpha < \alpha_0 \) and \( |\log \alpha|/n < \epsilon_0 \). Let \( m \) be an integer so that \( m/2 + 1 \geq M \). Now we claim that

\[ \frac{|\log(\alpha A (m, n, \Delta))|}{|\log(\alpha A (n, \Delta))|} \geq m/2 + 1. \]

(7.6)

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

\[ \alpha A (m, n, \Delta) \leq \rho_A (m, n, \Delta) \leq \rho_A (n, \Delta)^m \leq (2 \alpha A (n, \Delta)^{1/2})^m \]

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

\[ |\log(\alpha A (m, n, \Delta))| \geq m [\log(2^{1/2}) |\log(\alpha A (n, \Delta))| + \log 2]. \]

Now, for \( \alpha < 1/2 \), \( \log 2/|\log(\alpha A)| \geq 1 \), and so the last display proves (7.6).

Combining (7.6) with hypothesis (4.6) gives
It then follows that, with \( \alpha = \alpha_2(n, \Delta) \)
\[
\Delta_A(m \cdot n, \alpha) \leq \Delta_2(n, \alpha).
\]
Now by (7.5)
\[
\Delta_2(n, \alpha) \leq C_+ \left( \frac{\log n}{n} \right)^{q'/2},
\]
and
\[
C_- \left( \frac{\log n}{m \cdot n} \right)^{q/2} \leq \Delta_2(m \cdot n, \alpha).
\]
Combining these,
\[
\Delta_2(n, \alpha) \leq \left[ \frac{C_+ \cdot m^{q/2}}{C_-} \right] \Delta_2(m \cdot n, \alpha).
\] (7.7)

Hence, for every \( k = m \cdot n \),
\[
\Delta_A(k, \alpha) \leq C_0 \cdot \Delta_2(k, \alpha)
\] (7.8)
where \( C_0 = \left[ \frac{C_+ \cdot m^{q/2}}{C_-} \right] \).

We now consider \( l \) which is not divisible by \( m \): \( l = n \cdot m + r \) for \( r < m \). As \( \Delta_A(n, \alpha) \) is monotone in \( n \) for fixed \( \alpha \)
\[
\Delta_A(l, \alpha) \leq \Delta_A(m \cdot n, \alpha)
\]
and
\[
\Delta_2((n + 1) \cdot m, \alpha) \leq \Delta_2(l, \alpha)
\].

Now by a repeat of the reasoning behind (7.7),
\[
\Delta_2(n \cdot m, \alpha) \leq \left[ \frac{C_+ \cdot ((n + 1)/n)^{q/2}}{C_-} \right] \Delta_2((n + 1) \cdot m, \alpha)
\] (7.9)

Combining these relations, and noting that \((n + 1)/n \leq 2\),
\[
\Delta_A(l, \alpha) \leq \Delta_A(m \cdot n, \alpha) \leq \left[ \frac{C_+ \cdot m^{q/2}}{C_-} \right] \Delta_2(m \cdot n, \alpha) \quad \text{(by (7.8))}
\]
\[
\leq \left[ \frac{C_+}{C_-} \right]^2 m^{q/2} 2^{q/2} A_2((n+1)m, \alpha) \quad \text{(by (7.9))}
\]
\[
\leq C A_2(l, \alpha)
\]

where \( C = \left[ \frac{C_+}{C_-} \right]^2 m^{q/2} 2^{q/2} \).

**Proof of Theorem 4.5**

As in the last proof, (7.4) holds by hypothesis. Pick an integer \( m \) so that
\[
\frac{C_-}{C_+} m^{q/2} \geq C.
\]
Then
\[
\Delta_A(m, n, \alpha) \leq C A_2(m, n, \alpha)
\]
\[
\leq C C_2\left(\frac{\log |\alpha|}{m n}\right)^{q/2} \leq C \frac{C_+}{m^{q/2} C_-} C_2\left(\frac{\log |\alpha|}{n}\right)^{q/2}
\]
\[
\leq C_2\left(\frac{\log |\alpha|}{n}\right)^{q/2} \leq \Delta_2(n, \alpha)
\]

Hence, with \( \Delta = \Delta_A(m, n, \alpha) \)
\[
\alpha_2(n, \Delta) \geq \alpha.
\]
Defining \( \rho_2(n, \Delta) \) in a fashion analogous to \( \Delta_2(n, \Delta) \), only using \( \rho \) in place of testing affinity, we have that
\[
\rho_2(n, \Delta) \geq \alpha
\]
and also
\[
\rho_2(m, n, \Delta) \geq \alpha^n.
\]
Then from \( \pi \geq \frac{1}{2} \rho^2 \),
\[
\alpha_2(m, n, \alpha) \geq \frac{1}{2} \alpha^{2m}
\]
And so, if \( \alpha < 1/2 \),
\[
\frac{|\log \alpha_2(m, n, \Delta)|}{|\log \alpha_A(m, n, \Delta)|} \leq 2m + 1.
\] (7.10)
It follows that, for $\alpha_A < 1/2$, and $k$ of the form $m \cdot n$, we have (4.7) with $M = 2m+1$. Consider now $l = m \cdot n + r$, $0 < r \leq m$. By monotonicity of $\alpha_2$ and $\alpha_A$ in $n$,

$$\frac{\log(\alpha_2 (l, \Delta))}{\log(\alpha_A (l, \Delta))} \leq \frac{\log(\alpha_2 (m \cdot n + 1, \Delta))}{\log(\alpha_A (m \cdot n, \Delta))}$$

and

$$\frac{\log(\alpha_2 (m \cdot n + 1, \Delta))}{\log(\alpha_2 (m \cdot n, \Delta))} \leq \frac{\log(\alpha_2 (2m \cdot n, \Delta))}{\log(\alpha_2 (m \cdot n, \Delta))}$$

Then, from (3.7) again,

$$\alpha_2(2m \cdot n, \Delta) \geq .5 \rho_2(2m \cdot n, \Delta)^2 = .5 \rho_2(m \cdot n, \Delta)^4 \geq .5 \alpha_2(m \cdot n, \Delta)^4$$

and so, combining the last three displays,

$$\frac{\log(\alpha_2 (l, \Delta))}{\log(\alpha_A (l, \Delta))} \leq 10m + 5$$

which gives (4.7) in the general case with $M = 10m + 5$.

**Proof of Theorem 5.3**

(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

$$\lambda(x) = r(x, t)$$

where

$$\lambda(x) = \frac{C_+}{C_-} x^\Delta \frac{(1 + C_2 x^p)}{(1 - C_2 x^p)}$$

and

$$r(x, t) = \frac{1 - \int_0 ^{x_0} f_1(\nu) \, d\nu}{1 - \int_0 ^{x_0} f_0(\nu) \, d\nu}$$

Now $\lambda$ is monotone increasing in $x$. We claim that for sufficiently small $x_0$, $x \in (0, x_0)$, $0 < x_0 < 1$, $r(x, t)$ is monotone decreasing in $t$. Then, at any $(t, \Delta)$ pair at which the solution $a_1(t, \Delta)$ of (7.11) 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, i.e. to show that \( r(x, t) \) is monotone decreasing in \( t \). Put

\[
r(x, t) = \frac{1 - \beta(t)}{1 - \alpha(t)}
\]

where

\[
\alpha(t) = C_+ \left( \frac{x_{t+1}}{t+1} \right) \left[ 1 + C_2 x^p \frac{t + \Delta + 1}{t + \Delta + p + 1} \right] \\
\beta(t) = C_- \left( \frac{x_{t+1}}{t+1} \right) \left[ 1 - C_2 x^p \frac{t + 1}{t + p + 1} \right] .
\]

Now one can easily verify that

\[
\alpha(t), \beta(t) \quad \text{are decreasing in } t \quad \text{(7.12)}
\]

Then monotonicity of \( r(x, t) \) follows from

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

In fact, we can show that for \( x_0 \) small enough

\[
\frac{1 - \beta(t)}{1 - \alpha(t)} < 2 < \frac{\beta'(t)}{\alpha'(t)} \quad \text{(7.13)}
\]

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

Let us first establish the left hand inequality. This can be rewritten as

\[
1 > 2 \beta(t) - \alpha(t)
\]

and as \( 2 \beta(t) + \alpha(t) > 2 \beta(t) - \alpha(t) \) it is implied by

\[
1 > \max_t 2 \beta(t) + \alpha(t) .
\]

By (7.12), this reduces to

\[
1 > 2 \beta(0) + \alpha(0) . \quad \text{(7.14)}
\]

Now pick \( x_1 \) so that

\[
1 > 2 C_+ x_1 \left( 1 + C_2 \frac{x_1^p}{p + 1} \right) + C_+ \frac{x_1^{p+1}}{\Delta + 1} \left( 1 + C_2 x_1^p \frac{\Delta + 1}{\Delta + p + 1} \right) .
\]
Then for $x \in (0, x_1)$

$$2 \beta(0) + \alpha(0) = 2 C_+ x \left( 1 - C_2 \frac{x^p}{p + 1} \right) + C_+ \frac{x^\Delta + 1}{\Delta + 1} \left( 1 + C_2 x^p \frac{\Delta + 1}{\Delta + p + 1} \right)$$

$$\leq 2 C_+ x_1 \left( 1 + C_2 \frac{x_1^p}{p + 1} \right) + C_+ \frac{x_1^\Delta + 1}{\Delta + 1} \left( 1 + C_2 x_1^p \frac{\Delta + 1}{\Delta + p + 1} \right)$$

$$< 1$$

and so (7.14) follows. Thus the left hand side of (7.13) is established for $x_0 \leq x_1$.

We now consider the right hand inequality of (7.13). Now

$$\beta'(t) = \Psi(x, t) \left[ B_1 - B_2 + B_3 \right] \quad \text{and}$$

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

where

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

and

$$B_1 = - C_- \left( 1 - \frac{C_2 (t + 1) x^p}{t + p + 1} \right)$$

$$B_2 = C_- \left( 1 - \frac{C_2 (t + 1) x^p}{t + p + 1} \right) \cdot \frac{1}{(t + 1) \|\log(x)\|}$$

$$B_3 = C_- \left( \frac{C_2 (t + 1) x^p}{(t + p + 1)^2} - \frac{C_2 x^p}{t + p + 1} \right) \cdot \frac{1}{\|\log(x)\|}$$

$$A_1 = - C_+ \left[ 1 + \frac{C_2 (t + \Delta + 1) x^p}{t + p + \Delta + 1} \right] x^\Delta \frac{t + 1}{t + \Delta + 1}$$

$$A_2 = C_+ \left[ 1 + \frac{C_2 (t + \Delta + 1) x^p}{t + p + \Delta + 1} \right] x^\Delta \frac{t + 1}{(t + \Delta + 1)^2} \frac{1}{\|\log(x)\|}$$

$$A_3 = C_+ \left[ \frac{C_2 x^p}{t + p + \Delta + 1} - \frac{C_2 (t + \Delta + 1) x^p}{(t + p + \Delta + 1)^2} \right] x^\Delta \frac{t + 1}{(t + \Delta + 1)^2} \frac{1}{\|\log(x)\|}$$

The desired inequality is then equivalent to

$$B_1 - B_2 + B_3 < 2 (A_1 - A_2 - A_3)$$

(7.15)

for all $t$ and all $x < x_0$.

Note that for $x \in (0, 1)$, $B_1$ is increasing in $t$. Thus
\[
B_1 \leq \lim_{t \to \infty} B_1 = -C_-(1 - C_2 x^p) = B_4(x);
\]
similarly \(A_1\) is decreasing in \(t\) and
\[
A_1 \geq \lim_{t \to \infty} A_1 = -C_+ x^\delta (1 + C_2 x^p) = A_4(x).
\]

Pick \(\varepsilon > 0\). For \(x_2\) small enough
\[
B_4(x_2) < 2 A_4(x_2) - \varepsilon
\]
and so by the obvious monotonicities in \(x\),
\[
B_4(x) < 2 A_4(x) - \varepsilon \tag{7.16}
\]
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
\[
|B_2| + |B_3| + 2 |A_2| + 2 |A_3| < \varepsilon \tag{7.17}
\]
for \(x < x_3\). Then we have
\[
B_1 - B_2 + B_3 < B_4 + B_2 - B_3
\]
\[
\leq 2 A_4 + B_2 - B_3 - \varepsilon \quad \text{[by (7.16)]}
\]
\[
= 2 A_4 - 2 A_2 + 2 A_3 + B_2 - B_3 + 2 A_2 - 2 A_3 - \varepsilon
\]
\[
\leq 2 (A_1 - A_2 + A_3) + (|B_2| + |B_3| + 2 |A_2| + 2 |A_3|) - \varepsilon \quad \text{[by (7.17)]}
\]
\[
\leq 2 (A_1 - A_2 + A_3)
\]
and so (7.15) follows, for \(x_0 \leq \min(x_2, x_3)\).

Note the inequalities
\[
|B_2| \leq C_- (1 + C_2 x^p) / |\log(x)| \tag{7.18}
\]
\[
|B_3| \leq C_- (2 C_2 x^p) / |\log(x)|
\]
\[
|A_2| \leq C_+ x^\delta (1 + C_2 x^p) / |\log(x)|
\]
\[
|A_3| \leq C_+ x^\delta (2 C_2 x^p) / |\log(x)|
\]
valid for \(t \geq 0, 0 < x < 1\). Pick \(x_3\) so small that the sum of the upper bounds in (7.18) is less than \(\frac{\varepsilon}{2}\).

Then we have
\[
|B_2| + |B_3| + 2 |A_2| + 2 |A_3| < \varepsilon
\]
and this holds for all \(x < x_3\) by the monotonicity in \(x\) of the upper bounds in (7.18). Hence, (7.17).

Putting \(x_0 = \min(x_1, x_2, x_3)\) we see that both sides of (7.13) hold for all \(t \geq 0\) and all \(x \in (0, x_0)\).
Proof of Theorem 5.4

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 $b(n^{-1/2})$ is asymptotic to $A n^{-1/5}$ for an appropriate constant $A$, it turns out that to establish the final claim, it is sufficient to show that $T_{\Delta}^{(k)} - T(F) = O_p(n^{-1/5})$ uniformly in $F$.

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

Now we have

$$\hat{f}_n(\hat{t}_n) \geq \hat{f}_n(t_n) \quad (7.19)$$

and so

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

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

$$f_n(t_n) - f_n(t) \geq \gamma (t - t_n)^2 \quad (7.20)$$

uniformly in $F$, if $t \in (T(F) + r \cdot h_n \cdot s, T(F) + c)$, for some $r > 0$ and any $c$ smaller than the constant $\delta$ used in defining the class $F$.

It follows that if $\hat{t}_n - T(F) > r \cdot h_n \cdot s$ it must satisfy

$$Z_n(\hat{t}_n) - Z_n(t_n) \geq \gamma n^{2/5} (\hat{t}_n - t_n)^2 \quad (7.21)$$

where $Z_n$ is the stochastic process

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

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

$$P_F \{ \hat{t}_n - t_n > n^{-1/5} \Delta \} \leq P_F \{ Z_n(t) - Z_n(t_n) > \frac{1}{2} \gamma n^{2/5} (t - t_n)^2 \text{ for some } t \in (t_n + n^{-1/5} \Delta, t_n + c) \}.$$
+ \Pr \{ \sup_{t > t_n + e} \hat{f}_n(t) > \hat{f}_n(t_n) \}.

By unimodality of $F$ and (7.21), we have that

$$\Pr \{ \sup_{t > t_n + e} \hat{f}_n(t) > \hat{f}_n(t_n) \} \leq 2 \Pr \{ \sup_{t} |\hat{f}_n(t) - f_n(t)| > \frac{1}{2} \gamma e^2 \}$$

Using results of Silverman (1978) and the fact that $f(t) \leq M$ for every $t$ and every $F \in \mathcal{F}$, we can show that the last expression tends to zero uniformly in $F$. Thus,

$$\Pr (\hat{t}_n - t_n > n^{-1/5} \Delta) \leq \Pr (Z_n(t) > \frac{1}{2} \gamma n^{25} (t - t_n)^2 \text{ for some } t > t_n + n^{-1/5} \Delta) + o(1).$$

By Lemma 7.3,

$$\Pr (Z_n(t_n) < - \frac{1}{2} \gamma \Delta^2) \leq \exp\left(-\frac{n^{1/5} h_n \gamma^2 \Delta^4}{8 M \|K\|^2 + \max(M, h_n \|K\|) n^{-25} \Delta^2}\right)$$

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

$$\lim_{\Delta \to 0} \limsup_{n} \sup_{F} \Pr (Z_n(t_n) < - \frac{1}{2} \gamma \Delta^2) = 0.$$

It follows that if we can also show that

$$\lim_{\Delta \to 0} \limsup_{n} \sup_{F} \Pr (Z_n(t) > \frac{1}{2} \gamma (t - t_n)^2 \text{ for some } t > t_n + n^{-1/5} \Delta)$$

then

$$\lim_{\Delta \to 0} \limsup_{n} \sup_{F} \Pr (\hat{t}_n - t_n > n^{-1/5} \Delta) = 0.$$ 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 $(\hat{t}_n - t_n) = O_P(n^{-5})$ uniformly in $\mathcal{F}$ and the proof would be done. Consider then

$$\Pr (Z_n(t) > \frac{1}{2} \gamma (t - t_n)^2 \text{ for some } t > t_n + n^{-1/5} \Delta)$$

Note, as the class $\mathcal{F}$ is closed under translation, we can always assume $F$ is such that $t_n = 0$.

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 $\Delta_0 = \Delta$).

$$\ell_P (Z_n(t) > n^{25} \frac{1}{2} \gamma i^2 \text{ for some } t > n^{-1/5} \Delta)$$

$$\leq P_F \{ Z_n(t_i) > \frac{1}{4} \gamma \Delta_i^2 \text{ for some } i \geq 0 \} + P_F \{ \sup_{t_i < t < t_{i+1}} Z_n(t) > \frac{1}{4} \gamma \Delta_i^2 \text{ for some } i \geq 0 \}.$$
By Lemma 7.4, with \( \alpha = \frac{1}{4} \)

\[
\lim_{\Delta \to -\infty} \limsup_{n \to \infty} P_{F_{\Delta}} \left( Z_n(t_i) > \frac{1}{4} \gamma \Delta_i^2 \text{ for some } i \geq 0 \right) = 0 \tag{7.24}
\]

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

\[
\lim_{\Delta \to -\infty} \limsup_{n \to \infty} P_{F} \left( \sup_{t_i \leq t \leq t_{i+1}} Z_n(t) > \frac{1}{4} \gamma \Delta_i^2 \text{ for some } i \geq 0 \right) = 0 . \tag{7.25}
\]

So (7.22) is established and the proof is completed.

**Lemma 7.1.** Let \( K \) be positive and of support \([-s,s]\). Then for every \( F \) in \( \mathbb{F} \)

\[
|t_n - T(F)| \leq h_n s
\]

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

**Proof.** We have

\[
f_n(t) - f_n(t + v) = \int_{-h_n s}^{h_n s} K_n(u) (f(t + u) - f(t + u + v)) \, du
\]

If \( t \geq T(F) + h_n s \) then, \( t + u \) and \( t + u + v \) are on the same side of \( T(F) \), for every value \( u \) in the range of integration. By the unimodality of \( F \), it follows that \( f(t + u) > f(t + u + v) \). Therefore, \( f_n(t) - f_n(t + v) \) is an integral of nonnegative quantities, and so is nonnegative. Thus \( f_n(t) \) is monotone decreasing on \( t > T(F) + h_n s \). Similarly, \( f_n(t) \) is monotone increasing on \((-\infty, T(F) - h_n s)\). Thus a maximum of \( f_n(t) \) occurs in \([T(F) - h_n s, T(F) + h_n s]\). We now argue that no maximum of \( f_n \) occurs outside this interval. This is equivalent to saying that \( t_0 = T(F) + h_n s \) is a point of strict decrease of \( f_n \). Consider then \( f(t_0 + u) - f(t_0 + u + v) \), viewed as a function of \( u \). Unless this is zero a.e., the above display shows that \( f_n(t) \) is strictly bigger than \( f_n(t + v) \). Now by definition of \( \mathbb{F} \), \( f(t_0 + u) \geq f(T(F)) - c_s(h_n s + u)^2 \)

while \( f(t_0 + u + v) \leq f(T(F)) - c_s(h_n s + u + v)^2 \). It follows that the difference of these two quantities is bounded below by a quadratic function which is strictly positive at \( u = -h_n s \). Therefore the difference is not 0 a.e. in the range of the integrand, and \( t_0 \) is a point of strict decrease of \( f_n \).

**Lemma 7.2.** Let \( K \) be positive and of support \([-s,s]\). Let \( r \) be so large that
\[
\frac{1}{2} < (2 - \frac{(r+1)^2}{(r-1)^2}) - 4 \frac{C_+}{C_-} \frac{1}{(r-1)^2}
\]

Put \( \gamma = \frac{C_-}{2} \). Then for every \( F \in F \),

\[
f_n(t_n) - f_n(t) > \gamma(t-t_n)^2 \quad t-t_n \in (r \ h_n \ c - t_n)
\]

for \( c < \delta \).

Proof. Without loss of generality, put \( T(F) = 0 \). Now, by definition of the class \( F \), and using the previous lemma,

\[
f_n(t_n) \geq f(0) - c_+(2h_n s)^2 \]
\[
f_n(t) \leq f(0) - c_-(t-h_n s)^2
\]

so that

\[
f_n(t_n) - f_n(t) \geq c_-(t-t_n)^2 + c_-(t-h_n s)^2 - (t-t_n)^2 - c_+(2h_n s)
\]

If \( t > r h_n s \) then, by the previous lemma

\[
(t - h_n s) \geq (r-1) h_n s
\]
\[
(t - t_n) \geq (r-1) h_n s
\]
\[
(t - t_n) \leq (r+1) h_n s
\]

and so

\[
f_n(t_n) - f_n(t) \geq c_-(t-t_n)^2 (1 - \frac{(r-1)^2 - (r+1)^2}{(r-1)^2}) - 4 \frac{C_+}{C_-} \frac{1}{(r-1)^2}
\]
\[
\geq \frac{c_-}{2} (t-t_n)^2.
\]

where the inequality defining \( r \) has been used. The lemma follows.

Lemma 7.3

\[
\sup_{\mathcal{F}} P_p(Z_n(t) > \Delta) \leq \exp \left[ \frac{-n^{1/5} h_n \Delta^2 / 2}{M \|K\|^2 + \max(M h_n \|K\|, m) n^{-25} \Delta} \right].
\] (7.26)

Proof. This is an application of Bernstein's inequality, as follows.

\[
Z_n(t) = n^{-25} (\hat{f}_n(t) - E \hat{f}_n(t))
\]
\[
= n^{-35} \left[ \sum_{i=1}^{n} K \left( \frac{X_i - t}{h_n} \right) / h_n \right] - f_n(t)
\]
defining \( W_i = K \left[ \frac{X_i - t}{h_n} \right] / h_n - f_n(t) \) we have

\[
Z_n(t) = n^{-3/5} \sum W_i.
\]

\[ E W_i = 0 \]

\[
\text{Var } W_i = \int K^2 \left[ \frac{x - t}{h_n} \right] / h_n^2 f(x) \, dx - (f_n(t))^2
\]

\[
\leq \frac{M}{h_n} \int K^2(u) \, du - f_n^2(t)
\]

\[
\leq \frac{M}{h_n} \|K\|_2^2.
\]

\[
\|W_i\|_\infty = \max(M, \|K\|_\infty / h_n).
\]

Bernstein's inequality -- Shorack and Wellner (1986, page 855) or Pollard (1984, Appendix B) -- says

\[
P \left[ \frac{1}{\sqrt{n}} \sum W_i \geq \eta \right] \leq \exp \left[ -\frac{\eta^2}{2} \frac{\text{Var } W_i + \|W_i\|_\infty \cdot \eta}{(3n^{1/2})} \right]
\]

and so, putting \( \Delta = n^{-1/10} \eta \)

\[
P \left[ n^{-3/5} \sum W_i \geq \Delta \right] \leq \exp \left[ -\frac{n^{1/5} \Delta^2}{2} / \frac{M \|K\|_2^2 / h_n + \max(M, \|K\|_\infty / h_n) \cdot n^{-4} \Delta} \right]
\]

which, since nothing here depends on \( F \), except for \( \sup_t f(t) \leq M \), a condition which is true of all \( F \in \mathcal{F} \), we have (7.26).

**Lemma 7.4** Let \( t_i = n^{-1/5} (\Delta_0 + i \delta) \). Let \( M \cdot h_n \leq \|K\|_\infty \). Then for \( \alpha \in (0,1) \),

\[
\sup_{F} \sum_{i=0}^{\infty} P_F (Z_n(t_i) > \alpha \gamma \Delta_0^2) \leq \frac{\exp(-\beta_0 \Delta_0^2)}{1 - \exp(-\beta_0 \Delta_0 \delta)} + \frac{\exp(-\beta_2 n h_n)}{1 - \exp(-\beta_1 n h_n \Delta_0 \delta)} \tag{7.27}
\]

where

\[
\beta_0 = \frac{n^{1/5} h_n \alpha^2 \gamma^2}{22 \|K\|_2^2 M}
\]

\[
\beta_1 = \frac{n^{1/5} h_n \alpha^2 \gamma^2}{2.2 \|K\|_\infty}
\]

\[
\beta_2 = \frac{\alpha^2 \gamma^2 \|K\|_2^2}{22 \|K\|_\infty^2}.
\]

**Proof.** Fix \( N = 10 \). Let \( m \) be the smallest integer such that
Then for $i \geq m$, as $M \cdot h_n < \| K \|_\infty$ by assumption, the denominator in the exponential bound of the last lemma is bounded by

$$M \| K \|^2 + \max (M h_n, \| K \|_\infty) (\Delta_0 + i \delta) n^{-25}$$

$$\leq (1 + \frac{1}{N}) \| K \|_\infty (\Delta_0 + i \delta) n^{-25},$$

while for $i < m$

$$M \| K \|^2 + \max (M h_n, \| K \|_\infty) (\Delta_0 + i \delta) n^{-25}$$

$$\leq (N + 1) \| K \|^2 M .$$

Applying now the exponential bound from that lemma, we have

$$P_F \{ Z_n(t_i) > \alpha \gamma n^{25} t_i^2 \} \leq \begin{cases} e_1(i) & i < m \\ e_2(i) & i \geq m \end{cases}$$

where

$$e_1(i) = \exp(-\beta_0 (\Delta_0 + \delta i)^2)$$

and

$$e_2(i) = \exp(-\beta_1 (\Delta_0 + \delta i) n^{25}).$$

Thus

$$\sum_{i=0}^{m-1} P_F \{ Z_n(t_i) > \alpha \gamma n^{25} t_i^2 \} \leq \sum_{i=0}^{m-1} e_1(i) + \sum_{i=m}^{\infty} e_2(i).$$

(7.29)

Now

$$\sum_{i=0}^{m-1} e_1(i) \leq \sum_{i=0}^{\infty} e_1(i) = \sum_{i=0}^{\infty} \exp(-\beta_0 (\Delta_0^2 + 2 \Delta_0 \delta i + \delta^2 i^2))$$

$$\leq \exp(-\beta_0 \Delta_0^2) \sum_{i=0}^{\infty} \exp(-2 \beta_0 \Delta_0 \delta i)$$

$$= \frac{\exp(-\beta_0 \Delta_0^2)}{1 - \exp(-2 \beta_0 \Delta_0 \delta)}$$

(7.30)

and

$$\sum_{i=m}^{\infty} e_2(i) = \exp(-\beta_1 n^{25} (\Delta_0 + \delta m)) \sum_{i=0}^{\infty} \exp(-\beta_1 n^{25} \delta i).$$

Now, using (7.28) we see that
\[
\beta_1 n^{2/5} (\Delta_0 + \delta) > \beta_1 n^{4/5} N M \frac{\|K\|_2^2}{\|K\|_\infty}
\]

\[
= \beta_2 n h_n
\]

Thus

\[
\sum_{i=1}^n e_2(i) \leq \frac{\exp(-\beta_2 n h_n)}{1 - \exp(-\beta_1 n^{2/5} \delta)}
\]  

(7.31)

Note that all results depend on \( F \) only through the inequality \( \sup_t f(t) \leq M \), which is valid for all \( F \in \mathcal{F} \). Therefore, comparing (7.29) - (7.31) we have (7.27) and the lemma is proved.

**Lemma 7.5** Suppose that the kernel \( K \) has

\[
\|K(\cdot) - K(\cdot + \delta)\|_2^2 \leq \delta^2 \|K\|_2^2
\]

\[
\|K(\cdot) - K(\cdot + \delta)\|_\infty \leq \delta \|K\|_\infty
\]

for some constants \( \|K\|_2^2, \|K\|_\infty \). Let \( \delta > 0, \eta > 0 \). Suppose that the kernel \( K \) is positive, supported in \([-s,s]\), and that \( h_n s < t \). Then

\[
\sup_{h_n s_0} P_{F} \{ n^{2/5} (\hat{f}_n(t + h_n) - \hat{f}_n(t)) > \eta \delta \} \leq \exp \left[ \frac{-n^{1/5} h_n \eta^2}{\|K\|_2^2 M + \|K\|_\infty n^{-2/5} \eta / 3} \right].
\]  

(7.32)

**Proof.** Put

\[
W_i = K \left[ \frac{X_i - (t + \delta h_n)}{h_n} \right] / h_n - K \left[ \frac{X_i - t}{h_n} \right] / h_n
\]

Then, as \( f_n \) is monotone decreasing on \([h_n s, \infty)\) for any \( F \in \mathcal{F}_{20} \) we have

\[
E W_i = f_n(t+\delta) - f_n(t) \leq 0
\]

and as \( f(x) \leq M \) for all \( x \), for every \( F \in \mathcal{F} \),

\[
\text{Var} W_i \leq \frac{M}{h_n} \int (K(x + \delta) - K(x))^2 dx
\]

\[
\leq \frac{M}{h_n} \|K\|_2^2 \delta^2
\]

Also

\[
\|W_i\|_\infty \leq \|K(\cdot + \delta) - K(\cdot)\|_\infty \cdot \frac{1}{h_n}
\]
Now we apply Bernstein’s inequality.

\[ P \left( \frac{1}{\sqrt{n}} \sum (W_i - E W_i) \geq \lambda \right) \leq \exp \left( \frac{-\lambda^2 / 2}{\text{Var} W_i + \|W_i\|_\infty \cdot \frac{\lambda}{3 \sqrt{n}}} \right). \]

As \( E W_i \leq 0 \)

\[ P \left( \frac{1}{\sqrt{n}} \sum W_i \geq \lambda \right) \leq P \left( \frac{1}{\sqrt{n}} \sum (W_i - E W_i) \geq \lambda \right) \]

so putting \( \eta \cdot \delta = n^{-1/10} \lambda \),

\[ P \left( n^{-3/5} \sum W_i > \eta \delta \right) \leq \exp \left( \frac{-n^{1/5} \eta^2 \delta^2}{\delta^2 \|K\|_2^2 M \frac{1}{h_n} + \|K\|_\infty \cdot \delta \cdot \frac{1}{h_n} \cdot \frac{\delta n^{1/10}}{3 \sqrt{n}}} \right) \]

\[ = \exp \left( \frac{-n^{1/5} h_n \eta^2}{\|K\|_2^2 M + \|K\|_\infty n^{-2/5} \cdot \frac{n}{3}} \right), \]

which establishes the result.

**Lemma 7.6** Suppose that \( K \) satisfies assumption (K), with support \([-s, s]\). Suppose that \( h_n s < n^{-1/5} \Delta_0 \).

Put \( t_i = n^{-1/5} (\Delta_0 + \delta i) \), \( \Delta_i = \Delta_0 + \delta i \). Then for \( \Delta_0 \) large enough (\( K, \delta \) fixed).

\[ \sup_{F \leq 0} \sum_{i=0}^{n} P_F \left( \sup_{t_i \leq t \leq t_{i+1}} n^{2/5} (\hat{f}_n(t) - f_n(t_i)) > \alpha \gamma \Delta_i^2 \right) \]

\[ \leq \frac{\beta_5(\Delta_0) \exp(-\beta_7(\Delta_0))}{1 - \exp(-\beta_7(\Delta_0))} + \frac{\beta_8(\Delta_0) \exp(-\beta_9(\Delta_0))}{1 - \exp(-\beta_9(\Delta_0))} \]  

(7.33)

where

\[ \beta_5 = \left[ 1 - \exp \left( -\beta_3 (\alpha \gamma)^4 \Delta_0^3 \delta \frac{216}{\pi^4} \right) \right]^{-1} \]

\[ \beta_7 = \beta_3 (\alpha \gamma)^2 \Delta_0^4 \beta_6 - \log(2) \]

\[ \beta_8 = \left[ 1 - \exp \left( -\beta_4 \alpha \gamma \Delta_0 \delta \frac{32}{3 \pi^2} n^{2/5} \right) \right]^{-1} \]

\[ \beta_9 = \beta_4 \alpha \gamma \Delta_0^2 \beta_9 n^{2/5} - \log(2) \]
and $\beta_3, \beta_4, \beta_6,$ and $\beta_9$ are defined below, and do not depend on $\delta, \Delta_0,$ or on $n,$ provided $n^{1/2}k_n$ is constant independent of $n.$ The inequality is valid as soon as $\beta_7 > 0$ and $\beta_{10} > 0.$

**Proof.** We use a chaining argument. Let $t = t_i + \rho \delta, \rho \in [0,1].$ Now $\rho$ has a binary expansion $\rho = \sum_{j=1}^{\infty} b_j 2^{-j}$ where each $b_j = 0$ or $1.$ Letting $\rho_k = \sum_{j=1}^{k} b_j 2^{-j}$ we have the telescoping sum

$$\hat{f}_n(t) - \hat{f}_n(t_i) = \sum_{k=1}^{\infty} (\hat{f}_n(t_i + \rho_k \delta) - \hat{f}_n(t_i + \rho_{k-1} \delta)).$$

Note that $\rho_k - \rho_{k-1} = b_k 2^{-k} = 0$ or $2^{-k}.$ This formula is certainly rigorously valid if $\rho$ is a binary rational, in which case only a finite number of terms are not zero.

For $k = 1, 2, \ldots$ let $Y_{k,l,i}$ denote the random variable

$$Y_{k,l,i} = n^{25} (\hat{f}_n(t_i + l \ 2^{-k}) - \hat{f}_n(t_i + (l-1) \ 2^{-k})).$$

Now note that, because the kernel is continuous, $\hat{f}_n$ is a continuous function of $t.$ Therefore, the supremum of $\hat{f}_n(t) - \hat{f}_n(t_i)$ on the interval $(t_i, t_{i+1})$ is the same as the supremum at values of $t$ with $\rho$ a binary rational. Hence

$$P_F \left\{ \sup_{l \leq t \leq t_{i+1}} n^{25} (\hat{f}_n(t) - \hat{f}_n(t_i)) > \alpha \gamma \Delta_i^2 \right\} = P_F \left( \sup_{\rho \; \text{rational}} \sum_k b_k Y_{k,2^k \rho_k,i} > \alpha \gamma \Delta_i^2 \right)$$

Now let $e_k = k^{-2} \frac{6}{\pi^2}.$ Thus $\sum_{k=1}^{\infty} e_k = 1.$ Now suppose that, for each $k,$ it were true that

$$\sup_{l = 1, \ldots, 2^k} Y_{k,l,i} < e_k \alpha \Delta_i^2.$$

This would imply that for every binary rational $\rho$

$$\sum_k b_k Y_{k,2^k \rho_k,i} \leq \sum_k Y_{k,l,i},$$

$$\leq \sum_k e_k \alpha \Delta_i^2$$

$$= \alpha \Delta_i^2.$$

It follows that

$$P_F \left( \sup_{\rho} \sum_k b_k Y_{k,2^k \rho_k,i} \geq \alpha \gamma \Delta_i^2 \right) \leq \sum_k P_F \left( \sup_i Y_{k,l,i} \geq e_k \alpha \gamma \Delta_i^2 \right)$$

$$\leq \sum_k 2^k P_F \left( Y_{k,l,i} \geq e_k \alpha \gamma \Delta_i^2 \right).$$

Our strategy now is to use the exponential bound furnished by Lemma 7.5 to bound the final sum.
in this display. Let us first verify that the lemma applies. By hypothesis the kernel $K$ satisfies the assumptions of that lemma, and $h_n s \leq \Delta_0 n^{-1/5}$. Therefore that lemma applies to all increments $n^{2/5} (\hat{f}_n(i) - \hat{f}_n(i))$ for $i \in (t_i, t_{i+1})$, where $i \geq 0$. In particular it applies to every $Y_{k,i,d}$.

To apply Lemma 7.5, put $\eta 2^{-k} = \epsilon_k \alpha \gamma \Delta_i^2$; then by (7.32) for any $k \geq 1$

$$P_F \{Y_{k,i,d} > \epsilon_k \alpha \gamma \Delta_i^2\} \leq \exp \left( -\frac{n^{1/5} h_n (2^k \epsilon_k \alpha \gamma \Delta_i^2)^2 / 2}{\|K\|_M^2 + \|K\|_\infty n^{-2/5} (2^k \epsilon_k \alpha \gamma \Delta_i^2)^2 / 3} \right) \tag{7.34}$$

Fix $N (=10, \text{say})$. Let $m(i)$ be the least integer such that

$$2^m \epsilon_m \alpha \Delta_i^2 > 3 N M \frac{\|K\|_M^2}{\|K\|_\infty} n^{2/5}.$$  

Then for $k \geq m$, the denominator in (7.34) is smaller than

$$(1 + \frac{1}{N}) \|K\|_\infty n^{-2/5} (2^k \epsilon_k \alpha \gamma \Delta_i^2) / 3$$

while for $k < m$, the denominator is smaller than

$$(N + 1) \|K\|_M^2 M.$$  

Hence,

$$P_F \{Y_{k,i,d} > \epsilon_k \alpha \gamma \Delta_i^2\} \leq \begin{cases} e_3(k,i) & k < m(i) \\ e_4(k,i) & k \geq m(i) \end{cases}.$$

Here

$$e_3(k,i) = \exp(- \beta_3 (\alpha \gamma)^2 \Delta_i^4 2^{2k} \epsilon_k^2)$$

$$e_4(k,i) = \exp(- \beta_4 \alpha \gamma \Delta_i^2 2^k \epsilon_k n^{2/5})$$

and

$$\beta_3 = \frac{n^{1/5} h_n}{2 (N + 1) \|K\|_M^2 M}$$

$$\beta_4 = \frac{n^{1/5} h_n}{2 (1 + \frac{1}{N}) \|K\|_\infty / 3}.$$  

As these inequalities depend on $F$ only through the assumption that $f$ is monotone decreasing on $[h_n s, \infty]$ and $f(t) \leq M$.,
\[
\sup_{F \leq 0} \sum_{i=0}^{m-1} P\left(\sup_{t_i \leq t \leq t_{i+1}} n^{25} (\hat{f}_n(t) - \hat{f}_n(t_i)) \geq \alpha \gamma \Delta_i^2 \right) \leq \sum_{i=0}^{m-1} \left( \sum_{k=1}^{2^k} e_3(k,i) + \sum_{k=m+1}^{2^k} e_4(k,i) \right).
\]

Now
\[
\sum_{i=0}^{m-1} \sum_{k=1}^{2^k} e_3(k,i) \leq \sum_{i=0}^{m-1} \sum_{k=1}^{2^k} e_3(k,i) = \sum_{k=1}^{m-1} 2^k \sum_{i=0}^{m-1} e_3(k,i).
\]

And
\[
\sum_{i=0}^{m-1} e_3(k,i) \leq \exp(-\beta_3 (\alpha \gamma)^2 \Delta_0^4 2^{2k} \epsilon_k^2) \sum_{i=0}^{m-1} \exp(-\beta_3 (\alpha \gamma)^2 2^k \epsilon_k^2)
\]
\[
= \frac{\exp(-\beta_3 (\alpha \gamma)^2 \Delta_0^4 2^{2k} \epsilon_k^2)}{1 - \exp(-\beta_3 (\alpha \gamma)^2 2^k \epsilon_k^2)}.
\]

Note that \(\min_{k \geq 1} 2^{2k} \epsilon_k^2 = \frac{72}{\pi^2}\). Using this in the denominator of the last expression, we get
\[
\sum_{i=0}^{m-1} e_3(k,i) \leq \beta_5 \exp(-\beta_3 (\alpha \gamma)^2 \Delta_0^4 2^{2k} \epsilon_k^2),
\]
and so
\[
\sum_{k=0}^{m-1} 2^k \sum_{i=0}^{m-1} e_3(k,i) \leq \beta_5 \sum_{k=0}^{m-1} 2^k \exp(-\beta_3 (\alpha \gamma)^2 \Delta_0^4 2^{2k} \epsilon_k^2).
\]

Now we note that
\[
\min_{k \geq 1} \frac{2^{2k} \epsilon_k^2}{k} = \min_{k \geq 1} \left[ \frac{2^{2k}}{k^5} \right] \epsilon_k^2 = \beta_6 > 0,
\]
say, so that \(2^{2k} \epsilon_k^2 \geq \beta_6 k\). Then we have
\[
2^k \exp(-\beta_3 (\alpha \gamma)^2 \Delta_0^4 2^{2k} \epsilon_k^2) \leq \exp(-\beta_7 k).
\]
and so, supposing \(\Delta_0\) is large enough that \(\beta_7 > 0\),
\[
\sum_{k=0}^{m-1} 2^k \sum_{i=0}^{m-1} e_3(k,i) \leq \beta_5 \frac{\exp(-\beta_7 (\Delta_0))}{1 - \exp(-\beta_7 (\Delta_0))}.
\]

This is the first half of the lemma. Consider now the second term.
\[
\sum_{i=0}^{m} \sum_{k=m+1}^{2^k} e_4(k,i) \leq \sum_{k=0}^{m} 2^k \sum_{i=0}^{m} e_4(k,i)
\]
and by an argument similar to the one for \(e_3(k,i)\)
\[ \sum_{i=0}^\infty e_4(k,i) \leq \frac{\exp(-\beta_4 \alpha \gamma \Delta_0^2 2^k \epsilon_k n^{25})}{1 - \exp(-\beta_4 \alpha \gamma \Delta_0 2^{2k} \epsilon_k n^{25})} . \]

Then, noting that

\[ \min_{k \geq 1} 2^k \epsilon_k = \frac{8}{9} \frac{6}{\pi^2} \]

we have

\[ \sum_{i=0}^\infty e_4(k,i) \leq \beta_9 \exp(-\beta_4 \alpha \gamma \Delta_0^2 2^k \epsilon_k n^{25}) . \]

And, putting

\[ \min_{k \geq 1} \frac{2^k \epsilon_k}{k} = \beta_9 > 0 , \]

we have \( 2^k \epsilon_k \geq \beta_9 k \) and so, if we put

\[ \beta_{10}(\Delta_0) = \beta_4 \alpha \gamma \Delta_0^2 \beta_9 n^{25} - \log(2) \]

then

\[ 2^k \exp(-\beta_4 \alpha \gamma \Delta_0^2 2^k \epsilon_k n^{25}) \leq \exp(-\beta_{10} \cdot k) . \]

And, supposing \( \Delta_0 \) is so large that \( \beta_{10} > 0 \),

\[ \sum_{k=1}^\infty 2^k \sum_{i=0}^\infty e_4(k,i) \leq \beta_9 \frac{\exp(-\beta_{10}(\Delta_0))}{1 - \exp(-\beta_{10}(\Delta_0))} \]

by the same arguments as used for the \( e_3(k,i) \) sums. This gives the second bound and completes the proof.
8. References


17. DRAPER, D. C. (Feb 1983). Rank-based robust inference in regression models with several observations per cell.


32. BREIMAN, L. (May 1984). Nail finders, edifices, and Oz.


43. BICKEL, P. J. and YAHA0, J. A. (1985). On estimating the number of unseen species: how many executions were there?


56. LE CAM, L. and YANG, G. L. (January 1986). Replaced by No. 68.


64. O'SULLIVAN, F. (May 1986). Nonparametric Estimation in the Cox Proportional Hazards Model.


71. BICKEL, P.J. and YAHAV, J.A. (July 1986). Richardson Extrapolation and the Bootstrap.


77. O'SULLIVAN, F. (September 1986). Deconvolution of episodic hormone data.


82. DOKSUM, K.J. and LO, A.Y. (November 1986). Consistent and robust Bayes Procedures for Location based on Partial Information.


90. DABROWSKA, D.M. (March 1987). Rank tests for matched pair experiments with censored data.


95. CANCELLED


114. RITOV, Y. (September 1987). Estimation in a linear regression model with censored data.
118. ALDOUS, D.J. (October 1987). Meeting times for independent Markov chains.
127. Same as No. 133.
128. DONOHO, D. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean II.
131. Same as No. 140
132. HESSE, C.H. (December 1987). A Bahadur - Type representation for empirical quantiles of a large class of stationary, possibly infinite - variance, linear processes
166. FAN, JIANQING (July 1988). Nonparametric estimation of quadratic functionals in Gaussian white noise.
172. ADLER, R.J. and EPSTEIN, R. (September 1988). Intersection local times for infinite systems of planar brownian motions and for the brownian density process.


Copies of these Reports plus the most recent additions to the Technical Report series are available from the Statistics Department technical typist in room 379 Evans Hall or may be requested by mail from:
Department of Statistics
University of California
Berkeley, California 94720

Cost: $1 per copy.