

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

Geometrizing Rates of Convergence, II

David L. Donoho

Richard C. Liu

Department of Statistics
University of California at Berkeley

ABSTRACT

Consider estimating a functional $T(F)$ of an unknown distribution $F \in \mathbf{F}$ from data X_1, \dots, 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 $b(\epsilon)$ of the functional T over \mathbf{F} . The bound says that the estimation error $T_n - T$ cannot converge to zero faster than $b(n^{-1/2})$ uniformly over \mathbf{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 \mathbf{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 \mathbf{F} is convex, and also in the cases of the tail rate functional and the mode discussed above.

Key words and Phrases: Density Estimation, Estimating the Mode, Estimating the Rate of Tail Decay, Rates of Convergence of estimators, Modulus of Continuity, Hellinger Distance, Minimax Tests, Monotone Likelihood Ratio.

AMS-MOS Subject Classification: Primary 62G20. Secondary 62G05, 62F35.

Abbreviated Title: Geometrizing Rates, II

Acknowledgement: This work was supported by NSF Grant DMS-84-51753 and by donations from Schlumberger Computer Aided Systems and from Sun Microsystems, Inc. As in the other papers of this series, the authors would like to thank Lucien Le Cam for many helpful discussions. Jianqing Fan provided valuable comments on an earlier draft.

1. Introduction

Let $T(F)$ be a functional of an unknown distribution F and let X_1, \dots, 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 $F \in \mathbf{F}$ where \mathbf{F} is a given class of smooth densities, it may turn out that no estimator $T_n = T_n(X_1, \dots, X_n)$ 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 \mathbf{F} , with respect to Hellinger distance,

$$b(\epsilon) = \sup \{|T(F_i) - T(F_0)| : H(F_1, F_0) \leq \epsilon, F_i \in \mathbf{F}\} \quad (1.1)$$

it was shown that no estimator can converge to $T(F)$ faster than $b(n^{-1/2})$ uniformly over \mathbf{F} . This bound is valid for all functionals, and it was shown in [GR I] that the bound is at least as strong as rate bounds due to Farrell, Stone, and 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 $n \rightarrow \infty$.

Combining Theorems 2.1, 2.4, and 3.1 below, we get

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

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

and if $b(\epsilon)$ is Holderian with exponent q , so that

$$b(\epsilon) = C\epsilon^q + o(\epsilon^q),$$

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

$$\inf_{T_n} \sup_F E_F l(T_n - T) \asymp l(b(n^{-1/2})).$$

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 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 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, 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 $F_{\leq t}$ and $F_{\geq t+\Delta}$ denote the subsets of F where T takes values $\leq t$ and $\geq t+\Delta$, respectively. Let $F_{\leq t}^{(n)}$ denote the set of product measures of X_1, \dots, X_n iid F , $F \in F_{\leq t}$, and similarly for $F_{\geq t+\Delta}^{(n)}$. Denote by $\text{conv}(F_{\leq t}^{(n)})$ the set of all measures on R^n which can be gotten as convex combinations of the product measures in $F_{\leq t}^{(n)}$. Such a measure corresponds to the following: a random device is used to select an element $F \in F_{\leq t}$, and then n observations are taken from this realized F . In words, $\text{conv}(F_{\leq t}^{(n)})$ represents all the joint distributions of data X_1, \dots, X_n which can be obtained by Bayesians under a scheme in which (X_1, \dots, X_n) and F are random, with X_1, \dots, X_n conditionally i.i.d. F , and where F is a random element taking values in $F_{\leq t}$.

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

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

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

$$\alpha_A(n, \Delta) = \sup_t \pi(\text{conv}(F_{\leq t}^{(n)}), \text{conv}(F_{\geq t+\Delta}^{(n)})). \quad (2.2)$$

This is the minimax risk of the hardest problem of distinguishing $H_0: F_{\leq t}$ and $H_1: F_{\geq t+\Delta}$ at sample

size n . Next, we let $\Delta_A(n, \alpha)$ be the function inverse to α_A :

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

In words, $\Delta_A(n, \alpha)$ measures the largest Δ at which, in a sample of size n , one cannot test hypotheses $H_0: F_{\leq t}$ and $H_1: F_{\geq t + \Delta}$ with sum of errors less than α . As one might expect based on the exposition in [GR I], Δ_A places certain limits on how well T can be estimated. Essentially, this is because any estimator T_n of T gives rise to a test: decide H_0 if $T_n \leq T + \Delta/2$, decide H_1 if $T_n > T + \Delta/2$.

Theorem 2.1 (Lower Bound).

$$\inf_{T_n} \sup_F P_F \{ |T_n - T(F)| \geq \Delta_A(n, \alpha)/2 \} \geq \alpha/2 \quad (2.3)$$

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

The minimax risk for testing between $H_0: F_{\leq t_0}$ and $H_1: F_{\geq t_0 + \Delta}$ is α . It follows that for any test statistic

$$\alpha \leq \sup_{\substack{F_0 \in F_{\leq t_0} \\ F_1 \in F_{\geq t_0 + \Delta}}} P_{F_0} \{ \text{reject } H_0 \} + P_{F_1} \{ \text{accept } H_0 \}$$

so

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

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

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

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

$$\sup_{\substack{F_0 \in \mathcal{F}_{\leq t_0} \\ F_1 \in \mathcal{F}_{\geq t_0 + \Delta}}} \max_{F_0, F_1} P_{F_i} \{ |T_n - T(F)| \geq \Delta/2 \} \geq \alpha/2 \quad (2.4)$$

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

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

$$\inf_{T_n} \sup_{\mathcal{F}} E_F l(T_n - T(F)) \geq l(\Delta_A(n, \alpha)/2) \cdot \alpha/2 \quad (2.5)$$

for all n .

The reader should note that Δ_A is (nearly) the *best* lower bound derivable by a testing argument. Indeed, for each t , Δ , and n , there exists a test between $\mathcal{F}_{\leq t}$ and $\mathcal{F}_{\geq t + \Delta}$ 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 Δ may be thought of as follows: It defines an *acceptance region*, a measurable set $A = A(t, n, \Delta) \subset \mathcal{R}^n$, such that if the sample X_1, X_2, \dots, X_n falls in A , we accept $H_0: \mathcal{F}_{\leq t}$; 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 $\Delta_A(n, \alpha)$ 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" Δ , which will depend in a prescribed way with sample size. At a given sample size n , Δ is fixed and we proceed as follows. Let $N = N(M, \Delta)$ be the smallest integer such that $(\frac{3}{2})^N \Delta > 2M$. Let $l_N = -(\frac{3}{2})^N \Delta/2$ and $h_N = +(\frac{3}{2})^N \Delta/2$. Then the interval $[l_N, h_N]$ contains $[-M, M]$. At this point we proceed as follows. Given data X_1, \dots, 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 $F_{\geq M/3}$. 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 Δ , 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 Δ , M , and N . Set $\eta_0 = \frac{\Delta}{2}$, and $\eta_k = (\frac{3}{2})^k \Delta$ for $k \geq 1$. Set $d_k = \frac{1}{2}(\frac{3}{2})^k \Delta$ for $k = 0, 1, \dots$. Then*

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

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 (Δ , N):

```

 $k := N$ 
 $l_N := \frac{-1}{2} (\frac{3}{2})^N \Delta$ 
 $h_N := \frac{1}{2} (\frac{3}{2})^N \Delta$ 

while  $k > 0$  do
   $a_k := l_k + \frac{1}{3}(h_k - l_k)$ 
   $b_k := l_k + \frac{2}{3}(h_k - l_k)$ 

  Test  $H_0: F_{\leq a_k}$  against  $H_1: F_{\geq b_k}$ 

  if Accept  $H_0$  then /* new interval is (l,b) */
     $l_{k-1} := l_k$  ;  $h_{k-1} := b_k$ 
  if Reject  $H_0$  then /* new interval is (a,h) */
     $l_{k-1} := a_k$  ;  $h_{k-1} := h_k$ 

   $k := k-1$ 
end while

 $T_n = (l_0 + h_0)/2$ 

end Algorithm

```

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

Consider now the tests ξ_1, \dots, ξ_N , with ξ_k minimax for testing

$$H_0: F_{\leq a_k^*} \text{ versus } H_1: F_{\geq b_k^*}.$$

The probability that ξ_k decides incorrectly is

$$\pi(\text{conv}(F_{\leq a_k^*}^{(n)}), \text{conv}(F_{\geq b_k^*}^{(n)})) \leq \alpha_A(n, b_k^* - a_k^*) \equiv \alpha_A(n, d_{k-1}) \quad (2.7)$$

Consider now (2.6), and let $k > 0$. If the tests ξ_i all decide correctly for $i = k+1, \dots, N$, then

$T_n \in (l_k^*, h_k^*)$ and so $|T_n - T(F)| \leq h_k^* - l_k^* \equiv \eta_k$. Therefore,

$$\begin{aligned} P\{|T_n - T(F)| > \eta_k\} &\leq P\left(\bigcup_{i=k+1}^N \{\xi_i \text{ decides incorrectly}\}\right) \\ &\leq \sum_{i=k+1}^N P\{\xi_i \text{ decides incorrectly}\} \end{aligned} \quad (2.8)$$

$$\leq \sum_{i=k}^{N-1} \alpha_A(n, d_i);$$

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

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*

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

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

$$\beta \equiv \frac{1}{4} |\log(\alpha(2-\alpha))| \left(\frac{A_1}{A_0} \right)^{q/2} \text{ and } \gamma = \frac{1}{4n_0} |\log(\alpha_0(2-\alpha_0))|. \text{ Then with } \Delta \equiv C \Delta_A(n, \alpha) \text{ and } d_i \text{ as in}$$

Lemma 2.2

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

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

for $n > 2n_0$, where

$$\begin{aligned} \theta &\equiv \exp(-C^{2/q} \beta) \\ r_n &\equiv \frac{\log(3M)}{\log(A_0 \varepsilon_0^{q/2})} \cdot \exp(-n \gamma) \end{aligned} \quad (2.11)$$

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

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

$$E_F l(T_n - T) \leq A \cdot l(\Delta_A(n, \alpha)) \quad n > n_1 \quad (2.12)$$

for every $F \in \mathcal{F}$, where

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

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

Corollary. *Under the assumptions of Theorem 2.4,*

$$\inf_{T_n} \sup_F E_F l(T_n - T) \asymp l(\Delta_A(n, \alpha))$$

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

This use of minimax tests to construct estimators is inspired by work of Le Cam (1973), (1975), (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 Δ_A is not easy to calculate. In the important special case where T is linear, it may be bounded using the modulus of continuity, as we show in this section.

3.1. The Main Result

Theorem 3.1. *Suppose T is linear and 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*

$$b(c \sqrt{\frac{|\log \alpha|}{n}}) \leq \Delta_A(n, \alpha) \leq b(C \sqrt{\frac{|\log \alpha|}{n}}). \quad (3.1)$$

We may take $C = \sqrt{2}$ and $c=1/2$, for α_0, ε_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

$$\alpha_2(n, \Delta) = \sup_i \pi(F_{\leq i}^{(n)}, F_{\geq i+\Delta}^{(n)}) \quad (3.2)$$

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

$$\Delta_2(n, \alpha) = \sup \{ |T(F_1) - T(F_0)| : \pi(F_1^{(n)}, F_0^{(n)}) \geq \alpha \}. \quad (3.3)$$

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

$$\Delta_2(n, \alpha) \leq \Delta_A(n, \alpha); \quad (3.4)$$

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

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

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

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

$$b \left[c \sqrt{\frac{|\log \alpha|}{n}} \right] \leq \Delta_2(n, \alpha) \leq b \left[C \sqrt{\frac{|\log \alpha|}{n}} \right] \quad (3.5)$$

We may take $C = \sqrt{2}$ and $c^2/2 = \frac{(1 - e^{-\varepsilon_0})}{\varepsilon_0} \cdot \frac{\log(2 - \alpha_0) \alpha_0}{\log \alpha_0}$.

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

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

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

$$\pi(P, Q) \leq \rho(P, Q), \quad \rho^2 \leq \pi(2 - \pi) \quad (3.7)$$

where π is the testing affinity, and the identity

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

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

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

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

Proof. Define

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

and

$$h_1(n, \alpha) = \sup \{ H(F_1, F_0) : \pi(F_1^{(n)}, F_0^{(n)}) \geq \alpha \}.$$

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

$$h_0^2(n, \alpha) \geq 2(1 - (\alpha(2-\alpha))^{1/2n}) \quad (3.10)$$

$$h_1^2(n, \alpha) \leq 2(1 - \alpha^{1/n}) \quad (3.11)$$

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

$$b(h_0(n, \alpha)) \leq \Delta_2(n, \alpha) \leq b(h_1(n, \alpha)). \quad (3.12)$$

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

Lemma 3.3.

$$(1 - \alpha^{1/n}) \leq \frac{|\log \alpha|}{n}, \quad (3.13)$$

Fix $\alpha_0 < 1$, $\epsilon_0 > 0$. There exists a finite positive constant c so that for $\alpha < \alpha_0$, $|\log \alpha|/n < \epsilon_0$ we have

$$(1 - (2\alpha)^{\frac{1}{2n}}) \geq c^{2/2} \frac{|\log \alpha|}{n} \quad (3.14)$$

$$\text{We may take } c^{2/2} = \frac{(1 - e^{-\epsilon_0})}{\epsilon_0} \cdot \frac{\log(2 - \alpha_0) \alpha_0}{\log \alpha_0}.$$

This result is proved in the appendix, section 7.

In particular, If $b(\epsilon)$ 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 $P^{(n)}$, $Q^{(n)}$ the sets of corresponding product measures. Then

$$\rho(\text{conv } P^{(n)}, \text{conv } Q^{(n)}) \leq \rho(\text{conv } P, \text{conv } Q)^n. \quad (3.15)$$

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

Corollary.

$$\alpha_A(n, \Delta) \leq \sup_t \rho(\text{conv}(\mathbf{F}_{\leq t}), \text{conv}(\mathbf{F}_{\geq t + \Delta}))^n. \quad (3.16)$$

Thus the Hellinger Distance between the convex hulls of $\mathbf{F}_{\leq t}$ and $\mathbf{F}_{\geq t + \Delta}$ may be used to bound α_A .

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

$$\varepsilon = \inf_t H(\mathbf{F}_{\leq t}, \mathbf{F}_{\geq t + b(\varepsilon)}). \quad (3.17)$$

Combining this with (3.8) we have

$$\rho(\mathbf{F}_{\leq t}, \mathbf{F}_{\geq t + b(\varepsilon)}) \leq 1 - \varepsilon^2/2. \quad (3.18)$$

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

$$\alpha_A(n, b(\varepsilon)) \leq (1 - \varepsilon^2/2)^n \quad (3.19)$$

and so

$$\Delta_A(n, \alpha) \leq b(\sqrt{2(1 - \alpha^{1/n})}). \quad (3.20)$$

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 $F_{\leq t}$ versus $F_{\geq t+\Delta}$ 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

$$\Delta_A(n, \alpha) \leq C \Delta_2(n, \alpha) \quad (4.1)$$

for an appropriate constant C , for small α 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

$$\pi(\text{conv}(F_{\leq t}^{(n)}), \text{conv}(F_{\geq t+\Delta}^{(n)})) = \pi(F_{\leq t}^{(n)}, F_{\geq t+\Delta}^{(n)}); \quad (4.2)$$

Indeed, the quantity on the left hand side is the main ingredient in the definition of Δ_A , while that on the left is the main ingredient in Δ_2 . Now if we return to the definition of π as a measure of the difficulty of testing, we see that the quantity on the left is

$$\inf_{\zeta} \sup_{\substack{F_0 \in F_{\leq t} \\ F_1 \in F_{\geq t+\Delta}}} R_n(\zeta, (F_0, F_1))$$

where $R_n(\zeta, (F_0, F_1))$ is the "risk" $E_{F_0} \zeta + E_{F_1} (1 - \zeta)$ representing the sum of errors of the test ζ .

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

$$\sup_{\substack{F_0 \in F_{\leq t} \\ F_1 \in F_{\geq t+\Delta}}} \inf_{\zeta} R_n(\zeta, (F_0, F_1)).$$

This is the risk of the hardest 2-point testing problem. Consequently, the identity (4.2) is equivalent to the *minimax identity*

$$\inf_{\zeta} \sup_{(F_1, F_0)} R_n(\zeta, (F_0, F_1)) = \sup_{(\zeta, F_0)} \inf_{(F_1, F_0)} R_n(\zeta, (F_0, F_1)). \quad (4.3)$$

This identity says, in words, that the minimax risk in testing between the infinite dimensional composite hypotheses $F_{\leq t}$ and $F_{\geq t+\Delta}$ 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 $\Delta_A = \Delta_2$ 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 $\Delta_A = \Delta_2$ 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*

$$\begin{aligned} \text{SY} = \{ & f : f(x) = f(0) + x f'(0) + r(x), \\ & f(0) \leq M, \quad \int f = 1, \quad f \geq 0 \\ & |r(x)| \leq x^2/2 \}. \end{aligned}$$

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

It is known that in general, one cannot expect (4.2) to hold. One case when (4.2) does hold is when the sets $F_{\leq t}$ and $F_{\geq t+\Delta}$ are generated by capacities -- see Huber and Strassen (1973), Bednarski (1982). This is much stronger than simple convexity of the two sets. However, Le Cam's result, as recorded in Lemma 3.4 above, says that convexity alone is enough to guarantee that a certain approximate minimax identity holds.

Lemma 4.3. *If $F_{\leq t}$ and $F_{\geq t+\Delta}$ are both convex,*

$$\rho(\text{conv}(F_{\leq t}^{(n)}), \text{conv}(F_{\geq t+\Delta}^{(n)})) = \rho(F_{\leq t}^{(n)}, F_{\geq t+\Delta}^{(n)}). \quad (4.4)$$

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

Proof. We have

$$\begin{aligned} \rho(\mathbf{F}_{\leq t}, \mathbf{F}_{\geq t+\Delta})^n &\geq \rho(\text{conv}(\mathbf{F}_{\leq t}^{(n)}), \text{conv}(\mathbf{F}_{\geq t+\Delta}^{(n)})) \\ &\geq \rho(\mathbf{F}_{\leq t}^{(n)}, \mathbf{F}_{\geq t+\Delta}^{(n)}) \\ &= \rho(\mathbf{F}_{\leq t}, \mathbf{F}_{\geq t+\Delta})^n, \end{aligned}$$

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

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

$$1 \leq \frac{|\log \alpha_2(n, \Delta)|}{|\log \alpha_A(n, \Delta)|} \leq 2(1 + 1/|\log \alpha_A(n, \Delta)|). \quad (4.5)$$

Thus, at every n and Δ for which $\alpha_A(n, \Delta) \leq \alpha_0 < 1$, we can bound the discrepancy between α_2 and α_A .

In this sense, Le Cam's Lemma 3.4, which underlies (4.4), is an approximate minimax theorem. And one could say that Theorem 3.1 holds because (4.2) "almost" holds when T is linear and \mathbf{F} is convex.

4.2. A near-equivalence

In the case where T is linear and \mathbf{F} convex, we have seen that $\Delta_A \leq C \Delta_2$ and also that $|\log \alpha_2| \leq M |\log \alpha_A| + D$. In general, whatever be T and \mathbf{F} , these two accompany each other, so that if one holds, so does the other. This gives a clue to the general attainability issue; attainability of $b(n^{-1/2})$ 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*

$$\frac{|\log \alpha_2(n, \Delta)|}{|\log \alpha_A(n, \Delta)|} \leq M \quad \alpha_A(n, \Delta) \leq \alpha_0, \quad (4.6)$$

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 $x \rightarrow \infty$ (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 $x \rightarrow 0^+$ (compare Hall and Welsh (1984)). We adopt this point of view here. Accordingly, let \mathbf{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

$$0 < C_- \leq C \leq C_+ < \infty \quad (5.2b)$$

and

$$|r(x)| \leq c_2 x^p \quad (5.2c)$$

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

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

$$f_0^*(x) = C_- x^t (1 - c_2 x^p) \quad x \leq a_1(t, \Delta) \quad (5.3)$$

$$f_1^*(x) = C_+ x^{t+\Delta} (1 + c_2 x^p) \quad x \leq a_1(t, \Delta) \quad (5.4)$$

and

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

and

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

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*

$$\pi(\text{conv}(\mathbb{F}_{\leq t}^*), \text{conv}(\mathbb{F}_{\geq t+\Delta}^*)) = \pi(\mathbb{F}_{\leq t}^*, \mathbb{F}_{\geq t+\Delta}^*) = \pi((F_0^*)^{(*)}, (F_1^*)^{(*)})$$

and the minimax test between $\mathbb{F}_{\leq t}$ and $\mathbb{F}_{\geq t+\Delta}$ is the likelihood ratio test between F_0^ and F_1^* .*

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

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

This is a non-decreasing function of x .

Among all distributions in $\mathbb{F}_{\leq t}$, F_0^* is the stochastically largest. Similarly, among all distributions in $\mathbb{F}_{\geq t+\Delta}$, 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, \dots, X_n be iid F . Consider the likelihood ratio statistic

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

Under $H_0: F \leq F_0^*$ 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

$$\begin{aligned} \sup_{F \in \mathbb{F}_{\leq}} P_F \{ \text{Reject } H_0 \} &= P_{F_0^*} \{ \text{Reject } H_0 \} \\ \sup_{F \in \mathbb{F}_{\geq t+\Delta}} P_F \{ \text{Accept } H_0 \} &= P_{F_1^*} \{ \text{Accept } H_0 \} \end{aligned}$$

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

$$A(t, n, \Delta) \subset A(t + h, n, \Delta) \quad \forall h > 0. \quad (5.9)$$

The following result is proved in section 7.

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

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

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

with the monotonicity property (5.9).

Consider what we call the likelihood ratio estimator

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

By the monotonicity established in Theorem 5.3, $T_{n,\Delta}^*$ is always uniquely defined.

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

$$\sup_{\mathbf{F}} P_{\mathbf{F}} \{ |T_{n,\Delta}^* - T(F)| > \frac{\Delta}{2} \} \leq 2\alpha_A(n, \Delta) \quad (5.11)$$

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

Proof. By the monotonicity in t of $L_{t,\Delta}$,

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

happens if and only if the minimax test between $H_0: \mathbf{F}_{\leq T(F)}$ and $H_1: \mathbf{F}_{\geq T(F)+\Delta}$ would reject H_0 . The probability of this event is smaller than $\alpha_A(n, \Delta)$ by definition. Similarly, the probability of the event

$$T_{n,\Delta}^* - T(F) < -\Delta/2$$

is also less than $\alpha_A(n, \Delta)$. As $|T_{n,\Delta}^* - T(F)| > \frac{\Delta}{2}$ 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 \mathbf{F} be the class of distributions with *unimodal* densities f , that are uniformly bounded:

$$f(x) \leq M \quad (5.12)$$

and have quadratic maxima:

$$f(\text{mode}) - c_- x^2 \leq f(x) \leq f(\text{mode}) - c_+ x^2 \quad |x - \text{mode}| < \delta. \quad (5.13)$$

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 Δ small enough)

$$f_0^*(x) = M - c_+(x-t)^2 \quad x \in (t-a_2(\Delta), t+a_3(\Delta)) \quad (5.14a)$$

$$f_0^*(x) = M - c_+ a_3(\Delta)^2 \quad x \in (t+a_3(\Delta), t+\Delta-a_3(\Delta)) \quad (5.14b)$$

$$f_0^*(x) = M - c_-(x-t)^2 \quad x \in (t+\Delta-a_3(\Delta), t+\Delta+a_2(\Delta)). \quad (5.14c)$$

$$f_1^*(x) = f_0(2(t+\frac{\Delta}{2})-x) \quad (5.15)$$

$$f_0^*(x) = f_1^*(x) \quad x \in (t-a_2(\Delta), t+\Delta+a_2(\Delta)) \quad (5.16)$$

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

$$\pi((F_0^*)^{(n)}, (F_1^*)^{(n)}) = \pi(F_{\leq t}^{(n)}, F_{\geq t+\Delta}^{(n)}) \cdot (1 + o(1)).$$

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 $F_0 \in \mathcal{F}_{\leq t}$ and $F_1 \in \mathcal{F}_{\geq t+\Delta}$, 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 Δ , 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

$$\begin{aligned} K_n(u) &= \log \left[\frac{M - c_+ u^2}{M - c_+ a_3(\Delta)^2} \right] \quad |u| < a_3(\Delta) \\ &= 0 \quad \text{else.} \end{aligned}$$

This suggests the estimator

$$T_{n,\Delta}^+ = \arg \max_t \frac{1}{n} \sum K_n(X_i - t).$$

(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_∞ for which

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

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

are valid.

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

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

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

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

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

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

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 $\pi(\mathbf{F}_{\leq t}^{(n)}, \mathbf{F}_{\geq t+\Delta}^{(n)})$. This may be understood as follows. Roughly speaking, $\pi(\mathbf{F}_{\leq t}^{(n)}, \mathbf{F}_{\geq t+\Delta}^{(n)})$ is comparable to

$$\sup_{\mathbf{F}} \sup_{u > \Delta} P_F \left\{ \frac{1}{n} \sum_{i=1}^n K_n(X_i - T(F)) < \frac{1}{n} \sum_{i=1}^n K_n(X_i - (T(F) + u)) \right\} \quad (5.17)$$

whereas $\pi(\text{conv}(\mathbf{F}_{\leq t}^{(n)}), \text{conv}(\mathbf{F}_{\geq t+\Delta}^{(n)}))$ is comparable to

$$\sup_{\mathbf{F}} P_F \left\{ \frac{1}{n} \sum_{i=1}^n K_n(X_i - T(F)) < \sup_{u > \Delta} \frac{1}{n} \sum_{i=1}^n K_n(X_i - (T(F) + u)) \right\} \quad (5.18)$$

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 $\|\hat{f}_n - Ef_n\|_\infty$ 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 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 F .

6. An interesting example

Consider now the nonlinear functional $T(F) = \int f^2$. Let 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 F as above,*

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

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

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

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

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 comparable.

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,

$$\left(\frac{3}{2}\right)^k C A_1 \left(\frac{|\log \alpha|}{n}\right)^{q/2} \leq A_0 \varepsilon_0^{q/2} \quad (\text{Case 1})$$

In this case, there exists an integer m satisfying

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

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

$$\pi_{jm} \leq (\sqrt{\pi_m(2-\pi_m)})^j$$

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

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

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

$$\begin{aligned} \alpha_A(n, d_k) &\leq (\sqrt{\alpha(2-\alpha)})^{\frac{n}{m}-1} \\ &= \exp\left(-\frac{1}{2} |\log(\alpha(2-\alpha))| \left(\frac{n}{m} - 1\right)\right) \\ &\leq \exp\left\{-\frac{1}{2} |\log(\alpha(2-\alpha))| \left[\left[\frac{\left(\frac{3}{2}\right)^k C A_1}{A_0}\right]^{2/q} - 1\right]\right\} \\ &= \exp\left[-2\beta C^{2/q} \left(\frac{3}{2}\right)^{2k/q} + \frac{1}{2} |\log(\alpha(2-\alpha))| \right] \end{aligned}$$

The hypothesis $\frac{C A_1}{A_0} > 2^{q/2}$ implies

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

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

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

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

$$d_k > \frac{A_0^2}{A_1} \varepsilon_0^{q/2}; \quad (\text{Case 2})$$

using the definition of n_0 and arguing as above,

$$\begin{aligned} \alpha_A(n, d_k) &\leq (\sqrt{\alpha_0(2-\alpha_0)})^{\frac{n}{n_0}-1} \\ &= \exp(-2\gamma n + \frac{1}{2} |\log(\alpha_0(2-\alpha_0))|) \end{aligned}$$

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

$$\alpha_A(n, d_k) \leq \exp(-n\gamma), \quad n > 2n_0. \quad (7.2)$$

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

$K = \#\{k: A_0 \varepsilon_0^{q/2} \leq (\frac{3}{2})^k A_1 (\frac{|\log \alpha|}{n})^{q/2} \leq (\frac{3}{2})^N \Delta\}$. As $(\frac{3}{2})^N \Delta \leq 3M$, $K \leq \log(3M)/\log(A_0 \varepsilon_0^{q/2})$; thus K is

bounded independently of n and N . Now by (7.1) and (7.2), if $n > 2n_0$

$$\begin{aligned} \sum_{k=l}^{N-1} \alpha_A(n, d_k) &= \sum_{k=l}^{N-K-1} + \sum_{k=N-K}^{N-1} \\ &\leq \sum_{k=l}^{N-K-1} \exp(-\beta C^{2/q} \left(\frac{3}{2}\right)^{2k/q}) + K \exp(-n\gamma) \\ &\leq \sum_{k=l}^{\infty} \exp(-\beta C^{2/q} \left(\frac{3}{2}\right)^{2k/q}) + r_n \end{aligned} \quad (7.3)$$

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

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

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

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

which, with (7.3) gives (2.10b). \square

Proof of Theorem 2.4

$$\begin{aligned}
 E_F l(T_n - T) &\leq \sum_{k=0}^{N-1} P\{\eta_{k+1} \geq |T_n - T| > \eta_k\} l(\eta_{k+1}) \\
 &\leq \sum_{k=0}^{N-1} P\{|T_n - T| > \eta_k\} l(\eta_{k+1}) \\
 &\leq l(\eta_1) \frac{2\theta}{1-\theta^2} + \sum_{k=1}^{N-1} l(\eta_{k+1}) \frac{\theta^{2k}}{1-\theta^2} + \sum_0^{N-1} l(\eta_{k+1}) \cdot r_n
 \end{aligned}$$

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

$$\begin{aligned}
 E_F l(T_n - T) &\leq l(\Delta) \left[\frac{2\theta}{1-\theta^2} a + \sum_{k=1}^{\infty} \frac{\theta^{2k} a^{k+1}}{1-\theta^2} + r_n \frac{a^N - 1}{a - 1} \right] \\
 &= l(\Delta) \left[\frac{2a\theta}{1-\theta^2} + \frac{1}{1-\theta^2} \frac{a^2 \theta^2}{1-a\theta^2} + r_n \frac{a^N - 1}{a - 1} \right] \\
 &= l(\Delta)(A_2 + A_{3,n}),
 \end{aligned}$$

say. Now as $N \leq \log(3M)/\log(\Delta) \leq \log(3M)(\frac{Q}{2}(\log(n) - \log|\log \alpha|) - \log(A_0))$,

$$a^N r_n = \exp(-n\gamma + \log(a)N) \leq \exp(-n\gamma + A_4 \log(n) + A_5) \rightarrow 0, n \rightarrow \infty.$$

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

$$\begin{aligned} g(\varepsilon) &= 1 - \frac{\varepsilon}{2!} + \frac{\varepsilon^2}{3!} - \frac{\varepsilon^3}{4!} + \frac{\varepsilon^4}{5!} - \dots \\ &= 1 - \frac{\varepsilon}{2!} \left(1 - \frac{\varepsilon}{3}\right) - \frac{\varepsilon^3}{4!} \left(1 - \frac{\varepsilon}{5}\right) - \dots \end{aligned}$$

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).

$$\begin{aligned} \left[1 - (\alpha(2 - \alpha))^{1/2n}\right] &= \left[1 - \exp\left[-\frac{|\log(\alpha(2 - \alpha))|}{n}\right]\right] \\ &\geq g(\varepsilon_0) \frac{|\log(\alpha(2 - \alpha))|}{n} \\ &\geq g(\varepsilon_0) \frac{\log(\alpha_0(2 - \alpha_0))}{\log(\alpha_0)} \cdot \frac{|\log(\alpha)|}{n} \\ &\geq c^2/2 \frac{|\log(\alpha)|}{n} \end{aligned}$$

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 $F_0^* \in \mathbf{F}_{\leq t}$ and $F_1^* \in \mathbf{F}_{\geq t+\Delta}$

$$H(F_1^*, F_0^*) = \min_{\substack{F_1 \in \mathbf{F}_{\geq t+\Delta} \\ F_0 \in \mathbf{F}_{\leq t}}} H(F_1, F_0).$$

This closest pair in Hellinger distance has the form

$$f_1^*(x) = t + \Delta - x^2/2 \quad |x| \leq s_1(\Delta)$$

$$f_0^* = t + x^2 / 2 \quad |x| \leq s_1(\Delta)$$

where $s_1(\Delta) = \sqrt{\Delta} (1 + o(1))$ as $\Delta \rightarrow 0$, 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 $H_0: F \leq t$, $\min(|X|, s_1(\Delta))$ 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 ε_0 small enough, then by (3.5) there exist constants C_- , C_+ so that

$$C_- \left(\frac{|\log \alpha|}{n} \right)^{q/2} \leq \Delta_2(n, \alpha) \leq C_+ \left(\frac{|\log \alpha|}{n} \right)^{q/2} \quad (7.5)$$

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

$$\frac{|\log(\alpha_A(m, n, \Delta))|}{|\log(\alpha_A(n, \Delta))|} \geq m/2 + 1 . \quad (7.6)$$

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

$$|\log \alpha_A(m, n, \Delta)| \geq m [.5 |\log \alpha_A(n, \Delta)| + \log 2] .$$

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

Combining (7.6) with hypothesis (4.6) gives

$$|\log \alpha_A(m \cdot n, \Delta)| \geq |\log \alpha_2(n, \Delta)| .$$

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)

$$\Delta_2(n, \alpha) \leq C_+ \left[\frac{|\log \alpha|}{n} \right]^{q/2}$$

and

$$C_- \left[\frac{|\log \alpha|}{m \cdot n} \right]^{q/2} \leq \Delta_2(m \cdot n, \alpha).$$

Combining these,

$$\Delta_2(n, \alpha) \leq \left[\frac{C_+ m^{q/2}}{C_-} \right] \Delta_2(m \cdot n, \alpha) . \quad (7.7)$$

Hence, for every $k = m \cdot n$,

$$\Delta_A(k, \alpha) \leq C_0 \Delta_2(k, \alpha) \quad (7.8)$$

where $C_0 = \left[\frac{C_+ m^{q/2}}{C_-} \right]$.

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 α

$$\Delta_A(l, \alpha) \leq \Delta_A(m \cdot n, \alpha) .$$

and

$$\Delta_2((n+1) \cdot m, \alpha) \leq \Delta_2(l, \alpha) .$$

Now by a repeat of the reasoning behind (7.7),

$$\Delta_2(n \cdot m, \alpha) \leq \left[\frac{C_+ ((n+1)/n)^{q/2}}{C_-} \right] \Delta_2((n+1) \cdot m, \alpha) . \quad (7.9)$$

Combining these relations, and noting that $(n+1)/n \leq 2$,

$$\begin{aligned} \Delta_A(l, \alpha) &\leq \Delta_A(m \cdot n, \alpha) \\ &\leq \left[\frac{C_+ m^{q/2}}{C_-} \right] \Delta_2(m \cdot n, \alpha) \quad (\text{by (7.8)}) \end{aligned}$$

$$\begin{aligned} &\leq \left(\frac{C_+}{C_-} \right)^2 m^{q/2} 2^{q/2} \Delta_2((n+1)m, \alpha) \quad (\text{by (7.9)}) \\ &\leq C \Delta_2(l, \alpha) \end{aligned}$$

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

$$\frac{C_-}{C_+} m^{q/2} \geq C \quad .$$

Then

$$\begin{aligned} \Delta_A(m, n, \alpha) &\leq C \Delta_2(m, n, \alpha) \\ &\leq C C_+ \left(\frac{|\log \alpha|}{m, n} \right)^{q/2} \leq C \frac{C_+}{m^{q/2} C_-} C_- \left(\frac{|\log \alpha|}{n} \right)^{q/2} \\ &\leq C_- \left(\frac{|\log \alpha|}{n} \right)^{q/2} \leq \Delta_2(n, \alpha) \end{aligned}$$

Hence, with $\Delta = \Delta_A(m, n, \alpha)$

$$\alpha_2(n, \Delta) \geq \alpha \quad .$$

Defining $\rho_2(n, \Delta)$ in a fashion analogous to $\Delta_2(n, \Delta)$, only using ρ in place of testing affinity, we have that

$$\rho_2(n, \Delta) \geq \alpha$$

and also

$$\rho_2(m, n, \Delta) \geq \alpha^m \quad .$$

Then from $\pi \geq \frac{1}{2} \rho^2$,

$$\alpha_2(m, n, \alpha) \geq \frac{1}{2} \alpha^{2m}$$

And so, if $\alpha < 1/2$,

$$\frac{|\log \alpha_2(m, n, \Delta)|}{|\log \alpha_A(m, n, \Delta)|} \leq 2m + 1 \quad . \quad (7.10)$$

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 α_2 and α_A in n ,

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

and

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

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

$$\lambda(x) = r(x, t) \tag{7.11}$$

where

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

and

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

Now λ is monotone increasing in x . We claim that for sufficiently small x_0 , $x \in (0, x_0)$, $0 < x_0 < 1$,

$r(x, t)$ is monotone decreasing in t . Then, at any (t, Δ) 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

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

Now one can easily verify that

$$\alpha(t), \beta(t) \quad \text{are decreasing in } t. \quad (7.12)$$

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

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

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

$$1 > 2\beta(0) + \alpha(0). \quad (7.14)$$

Now pick x_1 so that

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

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

$$\begin{aligned}
 & 2 \beta(0) + \alpha(0) \\
 &= 2 C_- x (1 - C_2 \frac{x^p}{p+1}) + C_+ \frac{x^{\Delta+1}}{\Delta+1} (1 + C_2 x^p \frac{\Delta+1}{\Delta+p+1}) \\
 &\leq 2 C_- x_1 (1 + C_2 \frac{x_1^p}{p+1}) + C_+ \frac{x_1^{\Delta+1}}{\Delta+1} (1 + C_2 x_1^p \frac{\Delta+1}{\Delta+p+1}) \\
 &< 1
 \end{aligned}$$

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

$$\begin{aligned}
 \beta'(t) &= \Psi(x, t) [B_1 - B_2 + B_3] \\
 \alpha'(t) &= \Psi(x, t) [A_1 - A_2 + A_3]
 \end{aligned}$$

where

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

and

$$\begin{aligned}
 B_1 &= -C_- \left[1 - \frac{C_2 (t+1) x^p}{t+p+1} \right] \\
 B_2 &= C_- \left[1 - \frac{C_2 (t+1) x^p}{t+p+1} \right] \cdot \frac{1}{(t+1) |\log(x)|} \\
 B_3 &= C_- \left[\frac{C_2 (t+1) x^p}{(t+p+1)^2} - \frac{C_2 x^p}{t+p+1} \right] \cdot \frac{1}{|\log(x)|} \\
 A_1 &= -C_+ \left[1 + \frac{C_2 (t+\Delta+1) x^p}{t+p+\Delta+1} \right] x^\Delta \frac{t+1}{t+\Delta+1} \\
 A_2 &= C_+ \left[1 + \frac{C_2 (t+\Delta+1) x^p}{t+p+\Delta+1} \right] x^\Delta \frac{t+1}{(t+\Delta+1)^2} \frac{1}{|\log(x)|} \\
 A_3 &= C_+ \left[\frac{C_2 x^p}{t+p+\Delta+1} - \frac{C_2 (t+\Delta+1) x^p}{(t+p+\Delta+1)^2} \right] x^\Delta \frac{t+1}{(t+\Delta+1)^2} \frac{1}{|\log(x)|} .
 \end{aligned}$$

The desired inequality is then equivalent to

$$B_1 - B_2 + B_3 < 2 (A_1 - A_2 - A_3) \quad (7.15)$$

for all t and all $x < x_0$.

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

$$B_1 \leq \lim_{t \rightarrow \infty} B_1 = -C_- (1 - C_2 x^p) = B_4(x);$$

similarly A_1 is decreasing in t and

$$A_1 \geq \lim_{t \rightarrow \infty} A_1 = -C_+ x^\Delta (1 + C_2 x^p) = A_4(x).$$

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 ,

$$B_4(x) < 2 A_4(x) - \varepsilon \quad (7.16)$$

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

$$|B_2| + |B_3| + 2 |A_2| + 2 |A_3| < \varepsilon \quad (7.17)$$

for $x < x_3$. Then we have

$$\begin{aligned} B_1 - B_2 + B_3 &< B_4 + B_2 - B_3 \\ &\leq 2 A_4 + B_2 - B_3 - \varepsilon \quad [\text{by (7.16)}] \\ &= 2 A_4 - 2 A_2 + 2 A_3 + B_2 - B_3 + 2 A_2 - 2 A_3 - \varepsilon \\ &\leq 2 (A_1 - A_2 + A_3) + (|B_2| + |B_3| + 2 |A_2| + 2 |A_3|) - \varepsilon \quad [\text{by (7.17)}] \\ &\leq 2 (A_1 - A_2 + A_3) \end{aligned}$$

and so (7.15) follows, for $x_0 \leq \min (x_2, x_3)$.

Note the inequalities

$$\begin{aligned} |B_2| &\leq C_- (1 + C_2 x^p) / |\log(x)| \\ |B_3| &\leq C_- (2 C_2 x^p) / |\log(x)| \\ |A_2| &\leq C_+ x^\Delta (1 + C_2 x^p) / |\log(x)| \\ |A_3| &\leq C_+ x^\Delta (2 C_2 x^p) / |\log(x)| \end{aligned} \quad (7.18)$$

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 II]. 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 $T_n^{(k)} - T(F) = O_p(n^{-1/5})$ 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 $h_n = c n^{-1/5}$ 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 δ 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 Δ is so large that $n^{-1/5}\Delta > (r+1)h_n s$, we must have

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

$$+ P_F \{ \sup_{t > t_n + c} \hat{f}_n(t) > \hat{f}_n(t_n) \}.$$

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

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

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

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

By Lemma 7.3,

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

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

$$\lim_{\Delta \rightarrow \infty} \limsup_n \sup_F P_F \{ Z_n(t_n) < -\frac{1}{2} \gamma \Delta^2 \} = 0.$$

It follows that if we can also show that

$$\lim_{\Delta \rightarrow \infty} \limsup_n \sup_F P_F \{ Z_n(t) > \frac{1}{2} \gamma (t - t_n)^2 \text{ for some } t > t_n + n^{-1/5} \Delta \} \quad (7.22)$$

then

$$\lim_{\Delta \rightarrow \infty} \limsup_n \sup_F P_F \{ \hat{t}_n - t_n > n^{-1/5} \Delta \} = 0.$$

By the obvious symmetry in the problem, a similar relation would hold for $\hat{t}_n - t_n < -n^{-1/5} \Delta$, and so we would have $(\hat{t}_n - t_n) = O_p(n^{-1/5})$ uniformly in \mathbf{F} and the proof would be done. Consider then

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

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

Now, for $\delta > 0$, and for $i = 0, 1, \dots$, 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 $\Delta_0 = \Delta$).

$$P_F \{ Z_n(t) > n^{2/5} \frac{1}{2} \gamma t^2 \text{ for some } t > n^{-1/5} \Delta \} \quad (7.23)$$

$$\leq P_F \{ Z_n(t_i) > \frac{1}{4} \gamma \Delta_i^2 \text{ for some } i \geq 0 \} + P_F \left\{ \sup_{t_i \leq t \leq t_{i+1}} Z_n(t) > \frac{1}{4} \gamma \Delta_i^2 \text{ for some } i \geq 0 \right\}.$$

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

$$\lim_{\Delta \rightarrow \infty} \limsup_n \sup_{F \leq 0} P_F \{ Z_n(t_i) > \frac{1}{4} \gamma \Delta_i^2 \text{ for some } i \geq 0 \} = 0 \quad (7.24)$$

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

$$\lim_{\Delta \rightarrow \infty} \limsup_n \sup_F P_F \left\{ \sup_{t_i \leq t \leq t_{i+1}} Z_n(t) > \frac{1}{4} \gamma \Delta_i^2 \text{ for some } i \geq 0 \right\} = 0. \quad (7.25)$$

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} , $f(t_0+u) \geq f(T(F)) - c_+(h_n s + u)^2$ while $f(t_0+u+v) \leq f(T(F)) - c_-(h_n s + u+v)^2$. 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 $F \in \mathbf{F}$,

$$f_n(t_n) - f_n(t) > \gamma(t - t_n)^2 \quad t - t_n \in (r h_n s, c - t_n)$$

for $c < \delta$.

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

$$\begin{aligned} f_n(t_n) &\geq f(0) - c_+(2h_n s)^2 \\ f_n(t) &\leq f(0) - c_-(t - h_n s)^2 \end{aligned}$$

so that

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

If $t > r h_n s$ then, by the previous lemma

$$\begin{aligned} (t - h_n s) &\geq (r-1) h_n s \\ (t - t_n) &\geq (r-1) h_n s \\ (t - t_n) &\leq (r+1) h_n s \end{aligned}$$

and so

$$\begin{aligned} f_n(t_n) - f_n(t) &\geq c_-(t - t_n)^2 \left(1 - \frac{(r-1)^2 - (r+1)^2}{(r-1)^2} - 4 \frac{c_+}{c_-} \frac{1}{(r-1)^2}\right) \\ &\geq \frac{c_-}{2} (t - t_n)^2. \end{aligned}$$

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

Lemma 7.3

$$\sup_{\mathbf{F}} P_{\mathbf{F}}(Z_n(t) > \Delta) \leq \exp \left[\frac{-n^{1/5} h_n \Delta^2 / 2}{M \|K\|_2^2 + \max(M h_n, \|K\|_{\infty}) n^{-2/5} \Delta} \right]. \quad (7.26)$$

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

$$\begin{aligned} Z_n(t) &= n^{2/5} (\hat{f}_n(t) - E \hat{f}_n(t)) \\ &= n^{-3/5} \left[\sum_{i=1}^n \left[K \left(\frac{X_i - t}{h_n} \right) / h_n \right] - f_n(t) \right] \end{aligned}$$

defining $W_i = K \left[\frac{X_i - t}{h_n} \right] / h_n - f_n(t)$ we have

$$\begin{aligned} Z_n(t) &= n^{-3/5} \sum W_i \\ E W_i &= 0 \\ \text{Var } W_i &= \int K^2 \left[\frac{x - t}{h_n} \right] / h_n^2 f(x) dx - (f_n(t))^2 \\ &\leq \frac{M}{h_n} \int K^2(u) du - f_n^2(t) \\ &\leq \frac{M}{h_n} \|K\|_2^2. \end{aligned}$$

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

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

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

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

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

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

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

$$\sup_F \sum_{i=0}^{\infty} P_F \{Z_n(t_i) > \alpha \gamma \Delta_i^2\} \leq \frac{\exp(-\beta_0 \Delta_0^2)}{1 - \exp(-\beta_0 \Delta_0 \delta)} + \frac{\exp(-\beta_2 n h_n)}{1 - \exp(-\beta_1 \delta n^{2/5})} \quad (7.27)$$

where

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

Proof. Fix $N = 10$. Let m be the smallest integer such that

$$(\Delta_0 + m \delta) > N M \frac{\|K\|_2^2}{\|K\|_\infty} n^{2/5}. \quad (7.28)$$

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

$$\begin{aligned} & M \|K\|_2^2 + \max(M h_n, \|K\|_\infty) (\Delta_0 + i \delta) n^{-2/5} \\ & \leq (1 + \frac{1}{N}) \|K\|_\infty (\Delta_0 + i \delta) n^{-2/5}, \end{aligned}$$

while for $i < m$

$$\begin{aligned} & M \|K\|_2^2 + \max(M h_n, \|K\|_\infty) (\Delta_0 + i \delta) n^{-2/5} \\ & \leq (N + 1) \|K\|_2^2 M. \end{aligned}$$

Applying now the exponential bound from that lemma, we have

$$P_F \{Z_n(t_i) > \alpha \gamma n^{2/5} t_i^2\} \leq \begin{cases} e_1(i) & i < m \\ e_2(i) & i \geq m \end{cases}$$

where

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

and

$$e_2(i) = \exp(-\beta_1 (\Delta_0 + \delta i) n^{2/5}).$$

Thus

$$\sum_{i=0}^{\infty} P_F \{Z_n(t_i) > \alpha \gamma n^{2/5} t_i^2\} \leq \sum_{i=0}^{m-1} e_1(i) + \sum_{i=m}^{\infty} e_2(i). \quad (7.29)$$

Now

$$\begin{aligned} \sum_{i=0}^{m-1} e_1(i) & \leq \sum_{i=0}^{\infty} e_1(i) = \sum_{i=0}^{\infty} \exp(-\beta_0 (\Delta_0^2 + 2 \Delta_0 \delta i + \delta^2 i^2)) \\ & \leq \exp(-\beta_0 \Delta_0^2) \sum_{i=0}^{\infty} \exp(-2 \beta_0 \Delta_0 \delta i) \\ & = \frac{\exp(-\beta_0 \Delta_0^2)}{1 - \exp(-2 \beta_0 \Delta_0 \delta)} \end{aligned} \quad (7.30)$$

and

$$\sum_{i=m}^{\infty} e_2(i) = \exp(-\beta_1 n^{2/5} (\Delta_0 + \delta m)) \sum_{i=0}^{\infty} \exp(-\beta_1 n^{2/5} \delta i).$$

Now, using (7.28) we see that

$$\begin{aligned} \beta_1 n^{2/5} (\Delta_0 + m \delta) &> \beta_1 n^{4/5} N M \frac{\|K\|_2^2}{\|K\|_\infty} \\ &= \beta_2 n h_n . \end{aligned}$$

Thus

$$\sum_{i=m}^{\infty} e_2(i) \leq \frac{\exp(-\beta_2 n h_n)}{1 - \exp(-\beta_1 n^{2/5} \delta)} . \quad (7.31)$$

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

Lemma 7.5 *Suppose that the kernel K has*

$$\begin{aligned} \|K(\cdot) - K(\cdot + \delta)\|_2^2 &\leq \delta^2 \|K'\|_2^2 \\ \|K(\cdot) - K(\cdot + \delta)\|_\infty &\leq \delta \|K'\|_\infty \end{aligned}$$

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

$$\sup_{F \in \mathbf{F}_{\leq 0}} P_F \{n^{2/5} (\hat{f}_n(t + \delta h_n) - \hat{f}_n(t)) > \eta \delta\} \leq \exp \left[\frac{-n^{1/5} h_n \eta^2}{\|K'\|_2^2 M + \|K'\|_\infty n^{-2/5} \eta / 3} \right] . \quad (7.32)$$

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 $F \in \mathbf{F}_{\leq 0}$ 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 $F \in \mathbf{F}$,

$$\begin{aligned} \text{Var } W_i &\leq \frac{M}{h_n} \int (K(x + \delta) - K(x))^2 dx \\ &\leq \frac{M}{h_n} \|K'\|_2^2 \delta^2 . \end{aligned}$$

Also

$$\|W_i\|_\infty \leq \|K(\cdot + \delta) - K(\cdot)\|_\infty \cdot \frac{1}{h_n}$$

$$\leq \|K'\|_{\infty} \cdot \delta \cdot \frac{1}{h_n}.$$

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 $\eta \cdot \delta = n^{-1/10} \lambda$,

$$\begin{aligned} & P\{n^{-3/5} \sum W_i > \eta \delta\} \\ & \leq \exp\left[\frac{-n^{1/5} \eta^2 \delta^2}{\delta^2 \|K'\|_2^2 M \frac{1}{h_n} + \|K'\|_{\infty} \cdot \delta \cdot \frac{1}{h_n} \cdot \frac{\delta \eta n^{1/10}}{3 \sqrt{n}}}\right] \\ & = \exp\left[\frac{-n^{1/5} h_n \eta^2}{\|K'\|_2^2 M + \|K'\|_{\infty} n^{-2/5} \cdot \frac{\eta}{3}}\right], \end{aligned}$$

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 Δ_0 large enough (K, δ fixed).

$$\begin{aligned} & \sup_{F \leq 0} \sum_{i=0}^{\infty} P_F\left\{\sup_{t_i \leq t \leq t_{i+1}} n^{2/5} (\hat{f}_n(t) - \hat{f}_n(t_i)) > \alpha \gamma \Delta_i^2\right\} \\ & \leq \frac{\beta_5(\Delta_0) \exp(-\beta_7(\Delta_0))}{1 - \exp(-\beta_7(\Delta_0))} + \frac{\beta_8(\Delta_0) \exp(-\beta_{10}(\Delta_0))}{1 - \exp(-\beta_{10}(\Delta_0))} \end{aligned} \quad (7.33)$$

where

$$\begin{aligned} \beta_5 &= \left[1 - \exp\left[-\beta_3 (\alpha \gamma)^4 \gamma \Delta_0^3 \delta \frac{216}{\pi^4}\right]\right]^{-1} \\ \beta_7 &= \beta_3 (\alpha \gamma)^2 \Delta_0^4 \beta_6 - \log(2) \\ \beta_8 &= \left[1 - \exp\left[-\beta_4 \alpha \gamma \Delta_0 \delta \frac{32}{3 \pi^2} n^{2/5}\right]\right]^{-1} \\ \beta_{10} &= \beta_4 \alpha \gamma \Delta_0^2 \beta_9 n^{2/5} - \log(2) \end{aligned}$$

and $\beta_3, \beta_4, \beta_6$, and β_9 are defined below, and do not depend on δ , Δ_0 , or on n , provided $n^{1/5} \cdot h_n$ 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 ρ 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 ρ is a binary rational, in which case only a finite number of terms are not zero.

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

$$Y_{k,l,i} = n^{2/5} (\hat{f}_n(t_i + l 2^{-k}) - \hat{f}_n(t_i + (l-1) 2^{-k})) .$$

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 ρ a binary rational. Hence

$$P_F \left\{ \sup_{t_i \leq t \leq t_{i+1}} n^{2/5} (\hat{f}_n(t) - \hat{f}_n(t_i)) > \alpha \gamma \Delta_i^2 \right\} = P_F \left\{ \sup_{\rho \text{ rational}} \sum b_k Y_{k, 2^k \rho_{k,i}} > \alpha \gamma \Delta_i^2 \right\}$$

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

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

This would imply that for every binary rational ρ

$$\begin{aligned} \sum_k b_k Y_{k, 2^k \rho_{k,i}} &\leq \sum_k \sup_l Y_{k,l,i} \\ &< \sum_k \varepsilon_k \alpha \Delta_i^2 \\ &= \alpha \Delta_i^2 . \end{aligned}$$

It follows that

$$\begin{aligned} P_F \left\{ \sup_{\rho} \sum b_k Y_{k, 2^k \rho_{k,i}} \geq \alpha \gamma \Delta_i^2 \right\} &\leq \sum_k P_F \left\{ \sup_l Y_{k,l,i} \geq \varepsilon_k \alpha \gamma \Delta_i^2 \right\} \\ &\leq \sum_k 2^k P_F \left\{ Y_{k,1,i} \geq \varepsilon_k \alpha \gamma \Delta_i^2 \right\} . \end{aligned}$$

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 $n^{2/5}(\hat{f}_n(t) - \hat{f}_n(t_i))$ for $t \in (t_i, t_{i+1})$, where $i \geq 0$. In particular it applies to every $Y_{k,i}$.

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

$$P_F \{Y_{k,i} > \varepsilon \propto \Delta_i^2\} \leq \exp \left[\frac{-n^{1/5} h_n (2^k \varepsilon_k \propto \gamma \Delta_i^2)^2 / 2}{\|K'\|_2^2 M + \|K'\|_\infty n^{-2/5} (2^k \varepsilon_k \propto \gamma \Delta_i^2) / 3} \right] \quad (7.34)$$

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

$$2^m \varepsilon_m \propto \Delta_i^2 > 3 N M \frac{\|K'\|_2^2}{\|K'\|_\infty} n^{2/5}.$$

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

$$(1 + \frac{1}{N}) \|K'\|_\infty n^{-2/5} (2^k \varepsilon_k \propto \Delta_i^2) / 3$$

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

$$(N + 1) \|K'\|_2^2 M.$$

Hence,

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

Here

$$\begin{aligned} e_3(k,i) &= \exp(-\beta_3 (\alpha \gamma)^2 \Delta_i^4 2^{2k} \varepsilon_k^2) \\ e_4(k,i) &= \exp(-\beta_4 \alpha \gamma \Delta_i^2 2^k \varepsilon_k n^{2/5}) \end{aligned}$$

and

$$\begin{aligned} \beta_3 &= \frac{n^{1/5} h_n}{2 (N + 1) \|K'\|_2^2 M} \\ \beta_4 &= \frac{n^{1/5} h_n}{2 (1 + \frac{1}{N}) \|K'\|_\infty / 3}. \end{aligned}$$

As these inequalities depend on F only through the assumption that f is monotone decreasing on $[h_n s, \infty]$ and $f(t) \leq M$,

$$\sup_{F \leq 0} \sum_{i=0}^{\infty} P_F \left\{ \sup_{t_i \leq t \leq t_{i+1}} n^{2/5} (\hat{f}_n(t) - \hat{f}_n(t_i)) \geq \alpha \gamma \Delta_i^2 \right\} \leq \sum_{i=0}^{\infty} \left(\sum_{k=1}^{m-1} 2^k e_3(k, i) + \sum_{k=m}^{\infty} 2^k e_4(k, i) \right).$$

Now

$$\begin{aligned} \sum_{i=0}^{\infty} \sum_{k=1}^{m-1} 2^k e_3(k, i) &\leq \sum_{i=0}^{\infty} \sum_{k=1}^{\infty} 2^k e_3(k, i) \\ &= \sum_{k=1}^{\infty} 2^k \sum_{i=0}^{\infty} e_3(k, i). \end{aligned}$$

And

$$\begin{aligned} \sum_{i=0}^{\infty} e_3(k, i) &\leq \exp(-\beta_3 (\alpha \gamma)^2 \Delta_0^4 2^{2k} \varepsilon_k^2) \sum_{i=0}^{\infty} \exp(-\beta_3 (\alpha \gamma)^2 3 \Delta_0^3 \delta i 2^{2k} \varepsilon_k^2) \\ &= \frac{\exp(-\beta_3 (\alpha \gamma)^2 \Delta_0^4 2^{2k} \varepsilon_k^2)}{1 - \exp(-\beta_3 (\alpha \gamma)^2 3 \Delta_0^3 \delta 2^{2k} \varepsilon_k^2)}. \end{aligned}$$

Note that $\min_{k \geq 1} 2^{2k} \varepsilon_k^2 = \frac{72}{\pi^4}$. Using this in the denominator of the last expression, we get

$$\sum_{i=0}^{\infty} e_3(k, i) \leq \beta_5 \exp(-\beta_3 (\alpha \gamma)^2 \Delta_0^4 2^{2k} \varepsilon_k^2),$$

and so

$$\sum_{k=0}^{\infty} 2^k \sum_{i=0}^{\infty} e_3(k, i) \leq \beta_5 \sum_{k=0}^{\infty} 2^k \exp(-\beta_3 (\alpha \gamma)^2 \Delta_0^4 2^{2k} \varepsilon_k^2).$$

Now we note that

$$\min_{k \geq 1} \frac{2^{2k} \varepsilon_k^2}{k} = \min_{k \geq 1} \left[\frac{2^{2k}}{k^5} \right] \cdot \frac{6^2}{\pi^4} = \beta_6 > 0,$$

say, so that $2^{2k} \varepsilon_k^2 \geq \beta_6 k$. Then we have

$$2^k \exp(-\beta_3 (\alpha \gamma)^2 \Delta_0^4 2^{2k} \varepsilon_k^2) \leq \exp(-\beta_7 k).$$

and so, supposing Δ_0 is large enough that $\beta_7 > 0$,

$$\sum_{k=1}^{\infty} 2^k \sum_{i=0}^{\infty} e_3(k, i) \leq \beta_5 \frac{\exp(-\beta_7 (\Delta_0))}{1 - \exp(-\beta_7 (\Delta_0))}.$$

This is the first half of the lemma. Consider now the second term.

$$\sum_{i=0}^{\infty} \sum_{k=m(i)}^{\infty} 2^k e_4(k, i) \leq \sum_{k=1}^{\infty} 2^k \sum_{i=0}^{\infty} e_4(k, i)$$

and by an argument similar to the one for $e_3(k, i)$

$$\sum_{i=0}^{\infty} e_4(k, i) \leq \frac{\exp(-\beta_4 \alpha \gamma \Delta_0^2 2^k \varepsilon_k n^{2/5})}{1 - \exp(-\beta_4 2 \alpha \gamma \Delta_0 \delta 2^k \varepsilon_k n^{2/5})}.$$

Then, noting that

$$\min_{k \geq 1} 2^k \varepsilon_k = \frac{8}{9} \cdot \frac{6}{\pi^2}$$

we have

$$\sum_{i=0}^{\infty} e_4(k, i) \leq \beta_8 \exp(-\beta_4 \alpha \gamma \Delta_0^2 2^k \varepsilon_k n^{2/5}).$$

And, putting

$$\min_{k \geq 1} \frac{2^k \varepsilon_k}{k} = \beta_9 > 0,$$

we have $2^k \varepsilon_k \geq \beta_9 k$ and so, if we put

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

then

$$2^k \exp(-\beta_4 \alpha \gamma \Delta_0^2 2^k \varepsilon_k n^{2/5}) \leq \exp(-\beta_{10} \cdot k).$$

And, supposing Δ_0 is so large that $\beta_{10} > 0$,

$$\sum_{k=1}^{\infty} 2^k \sum_{i=0}^{\infty} e_4(k, i) \leq \beta_8 \frac{\exp(-\beta_{10}(\Delta_0))}{1 - \exp(-\beta_{10}(\Delta_0))}$$

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

8. References

- Bednarski, T. (1982) Binary experiments, minimax tests, and 2-alternating capacities. *Ann. Statist.* 10, 226-232.
- Birgé, L. (1983) Approximation dans les espaces métriques et théorie de l'estimation. *Zeit. Wahr.* 65, 181-237.
- Donoho, D.L., and Liu, R.C. (1988a) Geometrizing Rates of Convergence, I. Technical Report 137, Department of Statistics, University of California, Berkeley.
- Donoho, D.L., and Liu, R.C. (1988b) Geometrizing Rates of Convergence, III. Technical Report 138, Department of Statistics, University of California, Berkeley.
- Donoho, D. L. and Liu, R. C. (1988c) Hardest One-Dimensional Subfamilies. Technical Report ???, Department of Statistics, University of California, Berkeley.
- Donoho, D.L., MacGibbon, B., and Liu, R. C. (1988) Minimax risk for hyperrectangles. Technical Report 123, Department of Statistics, University of California, Berkeley.
- Du Mouchel, W. H. (1983) Estimating the stable index α in order to measure tail thickness. *Ann. Statist.* 11, 1019-1031.
- Farrell, R. H. (1972) "On the best obtainable asymptotic rates of convergence in estimation of a density function at a point." *Ann. Math. Stat.*, 43, #1, 170-180.
- Hall, P. and Marron, S. (1987) On the amount of noise inherent in bandwidth selection. *Ann. Stat.* 15
- Hall, P. and Welsh, A. H. (1984) Best attainable rates of convergence for parameters of regular variation. *Ann. Stat.* 12, 1079-1084.
- Ibragimov, I.A., and Hasminskii, R.Z. (1984) Nonparametric estimation of the values of a linear functional in Gaussian white noise. *Theory of Probability and its applications.* 28. — —
- Ibragimov, I.A., Nemirovskii, A.S., and Hasminskii, R.Z. (1987) Some problems on nonparametric estimation in Gaussian white noise. *Theory of Probability and its applications.* 31, 391-406
- Hasminskii, R. Z. (1979) "Lower bound for the risks of nonparametric estimates of the mode." in *Contribution to Statistics (J. Hajek memorial volume)*, Academia, Prague 1979, 91-97.
- Huber, P.J. and Strassen, V. (1973) Minimax tests and the Neyman-Pearson lemma for capacities. *Ann. Statist.* 1, 251-263.
- Le Cam, L. (1973) "Convergence of estimates under dimensionality restrictions." *Ann. Stat.*, 1, 38-53.

- Le Cam, L. (1975) "On local and global properties in the theory of asymptotic normality of experiments." in *Stochastic Processes and Related Topics*, Vol. 1.
- Le Cam, L. (1985) *Asymptotic Methods in Statistical Decision Theory*. Springer Verlag.
- Liu, R.C. (1987) *Geometry in Robustness and Nonparametrics*. Thesis, University of California, Berkeley.
- Meyer, T.G. (1977) On fixed or scaled radii confidence sets: the fixed sample size case. *Ann. Stat.* 5 65-78.
- Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer.
- Ritov, Y. and Bickel, P.J. (1987) Unachievable Information Bounds in Semiparametric and Nonparametric models. Technical Report 116, Department of Statistics, Univ. of Calif., Berkeley.
- Sacks, J. and Ylvisaker, N.D. (1981) Asymptotically optimum kernels for density estimation at a point. *Ann. Statist.* 9, 334-346.
- Shorack, G.R. and Wellner, J.A. (1986) *Empirical Processes with Applications in Statistics*. J. Wiley and Sons.
- Stone, C. J. (1980) "Optimal rates of convergence for nonparametric estimators." *Ann. Stat.*, 8,6, 1348-1360.
- Venter, J. (1967) On estimation of the mode. *Ann. Math. Statist.* 38, 1446-1455
- Wahba, G. (1975) Optimal convergence properties of variable knot, kernel, and orthogonal series methods for density estimation. *Ann. Stat.* 3, 15-29.

TECHNICAL REPORTS
Statistics Department
University of California, Berkeley

1. BREIMAN, L. and FREEDMAN, D. (Nov. 1981, revised Feb. 1982). How many variables should be entered in a regression equation? Jour. Amer. Statist. Assoc., March 1983, 78, No. 381, 131-136.
2. BRILLINGER, D. R. (Jan. 1982). Some contrasting examples of the time and frequency domain approaches to time series analysis. Time Series Methods in Hydrosociences, (A. H. El-Shaarawi and S. R. Esterby, eds.) Elsevier Scientific Publishing Co., Amsterdam, 1982, pp. 1-15.
3. DOKSUM, K. A. (Jan. 1982). On the performance of estimates in proportional hazard and log-linear models. Survival Analysis, (John Crowley and Richard A. Johnson, eds.) IMS Lecture Notes - Monograph Series, (Shanti S. Gupta, series ed.) 1982, 74-84.
4. BICKEL, P. J. and BREIMAN, L. (Feb. 1982). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. Ann. Prob., Feb. 1982, 11, No. 1, 185-214.
5. BRILLINGER, D. R. and TUKEY, J. W. (March 1982). Spectrum estimation and system identification relying on a Fourier transform. The Collected Works of J. W. Tukey, vol. 2, Wadsworth, 1985, 1001-1141.
6. BERAN, R. (May 1982). Jackknife approximation to bootstrap estimates. Ann. Statist., March 1984, 12 No. 1, 101-118.
7. BICKEL, P. J. and FREEDMAN, D. A. (June 1982). Bootstrapping regression models with many parameters. Lehmann Festschrift, (P. J. Bickel, K. Doksum and J. L. Hodges, Jr., eds.) Wadsworth Press, Belmont, 1983, 28-48.
8. BICKEL, P. J. and COLLINS, J. (March 1982). Minimizing Fisher information over mixtures of distributions. Sankhyā, 1983, 45, Series A, Pt. 1, 1-19.
9. BREIMAN, L. and FRIEDMAN, J. (July 1982). Estimating optimal transformations for multiple regression and correlation.
10. FREEDMAN, D. A. and PETERS, S. (July 1982, revised Aug. 1983). Bootstrapping a regression equation: some empirical results. JASA, 1984, 79, 97-106.
11. EATON, M. L. and FREEDMAN, D. A. (Sept. 1982). A remark on adjusting for covariates in multiple regression.
12. BICKEL, P. J. (April 1982). Minimax estimation of the mean of a mean of a normal distribution subject to doing well at a point. Recent Advances in Statistics, Academic Press, 1983.
14. FREEDMAN, D. A., ROTHENBERG, T. and SUTCH, R. (Oct. 1982). A review of a residential energy end use model.
15. BRILLINGER, D. and PREISLER, H. (Nov. 1982). Maximum likelihood estimation in a latent variable problem. Studies in Econometrics, Time Series, and Multivariate Statistics, (eds. S. Karlin, T. Amemiya, L. A. Goodman). Academic Press, New York, 1983, pp. 31-65.
16. BICKEL, P. J. (Nov. 1982). Robust regression based on infinitesimal neighborhoods. Ann. Statist., Dec. 1984, 12, 1349-1368.
17. DRAPER, D. C. (Feb. 1983). Rank-based robust analysis of linear models. I. Exposition and review.
18. DRAPER, D. C. (Feb 1983). Rank-based robust inference in regression models with several observations per cell.
19. FREEDMAN, D. A. and FIENBERG, S. (Feb. 1983, revised April 1983). Statistics and the scientific method, Comments on and reactions to Freedman, A rejoinder to Fienberg's comments. Springer New York 1985 Cohort Analysis in Social Research, (W. M. Mason and S. E. Fienberg, eds.).
20. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Jan. 1984). Using the bootstrap to evaluate forecasting equations. J. of Forecasting, 1985, Vol. 4, 251-262.
21. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Aug. 1983). Bootstrapping an econometric model: some empirical results. JBES, 1985, 2, 150-158.
22. FREEDMAN, D. A. (March 1983). Structural-equation models: a case study.
23. DAGGETT, R. S. and FREEDMAN, D. (April 1983, revised Sept. 1983). Econometrics and the law: a case study in the proof of antitrust damages. Proc. of the Berkeley Conference, in honor of Jerzy Neyman and Jack Kiefer. Vol I pp. 123-172. (L. Le Cam, R. Olshen eds.) Wadsworth, 1985.

24. DOKSUM, K. and YANDELL, B. (April 1983). Tests for exponentiality. Handbook of Statistics, (P. R. Krishnaiah and P. K. Sen, eds.) 4, 1984.
25. FREEDMAN, D. A. (May 1983). Comments on a paper by Markus.
26. FREEDMAN, D. (Oct. 1983, revised March 1984). On bootstrapping two-stage least-squares estimates in stationary linear models. Ann. Statist., 1984, 12, 827-842.
27. DOKSUM, K. A. (Dec. 1983). An extension of partial likelihood methods for proportional hazard models to general transformation models. Ann. Statist., 1987, 15, 325-345.
28. BICKEL, P. J., GOETZE, F. and VAN ZWET, W. R. (Jan. 1984). A simple analysis of third order efficiency of estimate Proc. of the Neyman-Kiefer Conference, (L. Le Cam, ed.) Wadsworth, 1985.
29. BICKEL, P. J. and FREEDMAN, D. A. Asymptotic normality and the bootstrap in stratified sampling. Ann. Statist. 12 470-482.
30. FREEDMAN, D. A. (Jan. 1984). The mean vs. the median: a case study in 4-R Act litigation. JBES. 1985 Vol 3 pp. 1-13.
31. STONE, C. J. (Feb. 1984). An asymptotically optimal window selection rule for kernel density estimates. Ann. Statist., Dec. 1984, 12, 1285-1297.
32. BREIMAN, L. (May 1984). Nail finders, edifices, and Oz.
33. STONE, C. J. (Oct. 1984). Additive regression and other nonparametric models. Ann. Statist., 1985, 13, 689-705.
34. STONE, C. J. (June 1984). An asymptotically optimal histogram selection rule. Proc. of the Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer (L. Le Cam and R. A. Olshen, eds.), II, 513-520.
35. FREEDMAN, D. A. and NAVIDI, W. C. (Sept. 1984, revised Jan. 1985). Regression models for adjusting the 1980 Census. Statistical Science. Feb 1986, Vol. 1, No. 1, 3-39.
36. FREEDMAN, D. A. (Sept. 1984, revised Nov. 1984). De Finetti's theorem in continuous time.
37. DIACONIS, P. and FREEDMAN, D. (Oct. 1984). An elementary proof of Stirling's formula. Amer. Math Monthly. Feb 1986, Vol. 93, No. 2, 123-125.
38. LE CAM, L. (Nov. 1984). Sur l'approximation de familles de mesures par des familles Gaussiennes. Ann. Inst. Henri Poincaré, 1985, 21, 225-287.
39. DIACONIS, P. and FREEDMAN, D. A. (Nov. 1984). A note on weak star uniformities.
40. BREIMAN, L. and IHAKA, R. (Dec. 1984). Nonlinear discriminant analysis via SCALING and ACE.
41. STONE, C. J. (Jan. 1985). The dimensionality reduction principle for generalized additive models.
42. LE CAM, L. (Jan. 1985). On the normal approximation for sums of independent variables.
43. BICKEL, P. J. and YAHAV, J. A. (1985). On estimating the number of unseen species: how many executions were there?
44. BRILLINGER, D. R. (1985). The natural variability of vital rates and associated statistics. Biometrics, to appear.
45. BRILLINGER, D. R. (1985). Fourier inference: some methods for the analysis of array and nonGaussian series data. Water Resources Bulletin, 1985, 21, 743-756.
46. BREIMAN, L. and STONE, C. J. (1985). Broad spectrum estimates and confidence intervals for tail quantiles.
47. DABROWSKA, D. M. and DOKSUM, K. A. (1985, revised March 1987). Partial likelihood in transformation models with censored data.
48. HAYCOCK, K. A. and BRILLINGER, D. R. (November 1985). LIBDRB: A subroutine library for elementary time series analysis.
49. BRILLINGER, D. R. (October 1985). Fitting cosines: some procedures and some physical examples. Joshi Festschrift, 1986. D. Reidel.
50. BRILLINGER, D. R. (November 1985). What do seismology and neurophysiology have in common? - Statistics! Comptes Rendus Math. Rep. Acad. Sci. Canada. January, 1986.
51. COX, D. D. and O'SULLIVAN, F. (October 1985). Analysis of penalized likelihood-type estimators with application to generalized smoothing in Sobolev Spaces.

52. O'SULLIVAN, F. (November 1985). A practical perspective on ill-posed inverse problems: A review with some new developments. To appear in Journal of Statistical Science.
53. LE CAM, L. and YANG, G. L. (November 1985, revised March 1987). On the preservation of local asymptotic normality under information loss.
54. BLACKWELL, D. (November 1985). Approximate normality of large products.
55. FREEDMAN, D. A. (June 1987). As others see us: A case study in path analysis. Journal of Educational Statistics. 12, 101-128.
56. LE CAM, L. and YANG, G. L. (January 1986). Replaced by No. 68.
57. LE CAM, L. (February 1986). On the Bernstein - von Mises theorem.
58. O'SULLIVAN, F. (January 1986). Estimation of Densities and Hazards by the Method of Penalized likelihood.
59. ALDOUS, D. and DIACONIS, P. (February 1986). Strong Uniform Times and Finite Random Walks.
60. ALDOUS, D. (March 1986). On the Markov Chain simulation Method for Uniform Combinatorial Distributions and Simulated Annealing.
61. CHENG, C-S. (April 1986). An Optimization Problem with Applications to Optimal Design Theory.
62. CHENG, C-S., MAJUMDAR, D., STUFKEN, J. & TURE, T. E. (May 1986, revised Jan 1987). Optimal step type design for comparing test treatments with a control.
63. CHENG, C-S. (May 1986, revised Jan. 1987). An Application of the Kiefer-Wolfowitz Equivalence Theorem.
64. O'SULLIVAN, F. (May 1986). Nonparametric Estimation in the Cox Proportional Hazards Model.
65. ALDOUS, D. (JUNE 1986). Finite-Time Implications of Relaxation Times for Stochastically Monotone Processes.
66. PITMAN, J. (JULY 1986, revised November 1986). Stationary Excursions.
67. DABROWSKA, D. and DOKSUM, K. (July 1986, revised November 1986). Estimates and confidence intervals for median and mean life in the proportional hazard model with censored data.
68. LE CAM, L. and YANG, G.L. (July 1986). Distinguished Statistics, Loss of information and a theorem of Robert B. Davies (Fourth edition).
69. STONE, C.J. (July 1986). Asymptotic properties of logspline density estimation.
71. BICKEL, P.J. and YAHAV, J.A. (July 1986). Richardson Extrapolation and the Bootstrap.
72. LEHMANN, E.L. (July 1986). Statistics - an overview.
73. STONE, C.J. (August 1986). A nonparametric framework for statistical modelling.
74. BIANE, PH. and YOR, M. (August 1986). A relation between Lévy's stochastic area formula, Legendre polynomial, and some continued fractions of Gauss.
75. LEHMANN, E.L. (August 1986, revised July 1987). Comparing Location Experiments.
76. O'SULLIVAN, F. (September 1986). Relative risk estimation.
77. O'SULLIVAN, F. (September 1986). Deconvolution of episodic hormone data.
78. PITMAN, J. & YOR, M. (September 1987). Further asymptotic laws of planar Brownian motion.
79. FREEDMAN, D.A. & ZEISEL, H. (November 1986). From mouse to man: The quantitative assessment of cancer risks. To appear in Statistical Science.
80. BRILLINGER, D.R. (October 1986). Maximum likelihood analysis of spike trains of interacting nerve cells.
81. DABROWSKA, D.M. (November 1986). Nonparametric regression with censored survival time data.
82. DOKSUM, K.J. and LO, A.Y. (November 1986). Consistent and robust Bayes Procedures for Location based on Partial Information.
83. DABROWSKA, D.M., DOKSUM, K.A. and MIURA, R. (November 1986). Rank estimates in a class of semiparametric two-sample models.

84. BRILLINGER, D. (December 1986). Some statistical methods for random process data from seismology and neurophysiology.
85. DIACONIS, P. and FREEDMAN, D. (December 1986). A dozen de Finetti-style results in search of a theory. Ann. Inst. Henri Poincaré, 1987, 23, 397-423.
86. DABROWSKA, D.M. (January 1987). Uniform consistency of nearest neighbour and kernel conditional Kaplan - Meier estimates.
87. FREEDMAN, D.A., NAVIDI, W. and PETERS, S.C. (February 1987). On the impact of variable selection in fitting regression equations.
88. ALDOUS, D. (February 1987, revised April 1987). Hashing with linear probing, under non-uniform probabilities.
89. DABROWSKA, D.M. and DOKSUM, K.A. (March 1987, revised January 1988). Estimating and testing in a two sample generalized odds rate model.
90. DABROWSKA, D.M. (March 1987). Rank tests for matched pair experiments with censored data.
91. DIACONIS, P and FREEDMAN, D.A. (April 1988). Conditional limit theorems for exponential families and finite versions of de Finetti's theorem. To appear in the Journal of Applied Probability.
92. DABROWSKA, D.M. (April 1987, revised September 1987). Kaplan-Meier estimate on the plane.
- 92a. ALDOUS, D. (April 1987). The Harmonic mean formula for probabilities of Unions: Applications to sparse random graphs.
93. DABROWSKA, D.M. (June 1987, revised Feb 1988). Nonparametric quantile regression with censored data.
94. DONOHO, D.L. & STARK, P.B. (June 1987). Uncertainty principles and signal recovery.
95. CANCELLED
96. BRILLINGER, D.R. (June 1987). Some examples of the statistical analysis of seismological data. To appear in *Proceedings, Centennial Anniversary Symposium, Seismographic Stations, University of California, Berkeley*.
97. FREEDMAN, D.A. and NAVIDI, W. (June 1987). On the multi-stage model for carcinogenesis. To appear in *Environmental Health Perspectives*.
98. O'SULLIVAN, F. and WONG, T. (June 1987). Determining a function diffusion coefficient in the heat equation.
99. O'SULLIVAN, F. (June 1987). Constrained non-linear regularization with application to some system identification problems.
100. LE CAM, L. (July 1987, revised Nov 1987). On the standard asymptotic confidence ellipsoids of Wald.
101. DONOHO, D.L. and LIU, R.C. (July 1987). Pathologies of some minimum distance estimators. Annals of Statistics, June, 1988.
102. BRILLINGER, D.R., DOWNING, K.H. and GLAESER, R.M. (July 1987). Some statistical aspects of low-dose electron imaging of crystals.
103. LE CAM, L. (August 1987). Harald Cramér and sums of independent random variables.
104. DONOHO, A.W., DONOHO, D.L. and GASKO, M. (August 1987). Macspin: Dynamic graphics on a desktop computer. IEEE Computer Graphics and applications, June, 1988.
105. DONOHO, D.L. and LIU, R.C. (August 1987). On minimax estimation of linear functionals.
106. DABROWSKA, D.M. (August 1987). Kaplan-Meier estimate on the plane: weak convergence, LIL and the bootstrap.
107. CHENG, C-S. (Aug 1987, revised Oct 1988). Some orthogonal main-effect plans for asymmetrical factorials.
108. CHENG, C-S. and JACROUX, M. (August 1987). On the construction of trend-free run orders of two-level factorial designs.
109. KLASS, M.J. (August 1987). Maximizing $E \max_{1 \leq k \leq n} S_k^+ / ES_n^+$: A prophet inequality for sums of I.I.D. mean zero variates.
110. DONOHO, D.L. and LIU, R.C. (August 1987). The "automatic" robustness of minimum distance functionals. Annals of Statistics, June, 1988.
111. BICKEL, P.J. and GHOSH, J.K. (August 1987, revised June 1988). A decomposition for the likelihood ratio statistic and the Bartlett correction — a Bayesian argument.

112. BURDZY, K., PITMAN, J.W. and YOR, M. (September 1987). Some asymptotic laws for crossings and excursions.
113. ADHIKARI, A. and PITMAN, J. (September 1987). The shortest planar arc of width 1.
114. RITOV, Y. (September 1987). Estimation in a linear regression model with censored data.
115. BICKEL, P.J. and RITOV, Y. (Sept. 1987, revised Aug 1988). Large sample theory of estimation in biased sampling regression models I.
116. RITOV, Y. and BICKEL, P.J. (Sept.1987, revised Aug. 1988). Achieving information bounds in non and semiparametric models.
117. RITOV, Y. (October 1987). On the convergence of a maximal correlation algorithm with alternating projections.
118. ALDOUS, D.J. (October 1987). Meeting times for independent Markov chains.
119. HESSE, C.H. (October 1987). An asymptotic expansion for the mean of the passage-time distribution of integrated Brownian Motion.
120. DONOHO, D. and LIU, R. (Oct. 1987, revised Mar. 1988, Oct. 1988). Geometrizing rates of convergence, II.
121. BRILLINGER, D.R. (October 1987). Estimating the chances of large earthquakes by radiocarbon dating and statistical modelling. To appear in *Statistics a Guide to the Unknown*.
122. ALDOUS, D., FLANNERY, B. and PALACIOS, J.L. (November 1987). Two applications of urn processes: The fringe analysis of search trees and the simulation of quasi-stationary distributions of Markov chains.
123. DONOHO, D.L., MACGIBBON, B. and LIU, R.C. (Nov.1987, revised July 1988). Minimax risk for hyperrectangles.
124. ALDOUS, D. (November 1987). Stopping times and tightness II.
125. HESSE, C.H. (November 1987). The present state of a stochastic model for sedimentation.
126. DALANG, R.C. (December 1987, revised June 1988). Optimal stopping of two-parameter processes on nonstandard probability spaces.
127. Same as No. 133.
128. DONOHO, D. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean II.
129. SMITH, D.L. (December 1987). Exponential bounds in Vapnik-Cervonenkis classes of index 1.
130. STONE, C.J. (Nov.1987, revised Sept. 1988). Uniform error bounds involving logspline models.
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
133. DONOHO, D.L. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean, I.
134. DUBINS, L.E. and SCHWARZ, G. (December 1987). A sharp inequality for martingales and stopping-times.
135. FREEDMAN, D.A. and NAVIDI, W. (December 1987). On the risk of lung cancer for ex-smokers.
136. LE CAM, L. (January 1988). On some stochastic models of the effects of radiation on cell survival.
137. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On the uniform consistency of Bayes estimates for multinomial probabilities.
- 137a. DONOHO, D.L. and LIU, R.C. (1987). Geometrizing rates of convergence, I.
138. DONOHO, D.L. and LIU, R.C. (January 1988). Geometrizing rates of convergence, III.
139. BERAN, R. (January 1988). Refining simultaneous confidence sets.
140. HESSE, C.H. (December 1987). Numerical and statistical aspects of neural networks.
141. BRILLINGER, D.R. (January 1988). Two reports on trend analysis: a) An Elementary Trend Analysis of Rio Negro Levels at Manaus, 1903-1985 b) Consistent Detection of a Monotonic Trend Superposed on a Stationary Time Series
142. DONOHO, D.L. (Jan. 1985, revised Jan. 1988). One-sided inference about functionals of a density.

143. DALANG, R.C. (February 1988). Randomization in the two-armed bandit problem.
144. DABROWSKA, D.M., DOKSUM, K.A. and SONG, J.K. (February 1988). Graphical comparisons of cumulative hazards for two populations.
145. ALDOUS, D.J. (February 1988). Lower bounds for covering times for reversible Markov Chains and random walks on graphs.
146. BICKEL, P.J. and RITOV, Y. (Feb.1988, revised August 1988). Estimating integrated squared density derivatives.
147. STARK, P.B. (March 1988). Strict bounds and applications.
148. DONOHO, D.L. and STARK, P.B. (March 1988). Rearrangements and smoothing.
149. NOLAN, D. (March 1988). Asymptotics for a multivariate location estimator.
150. SEILLIER, F. (March 1988). Sequential probability forecasts and the probability integral transform.
151. NOLAN, D. (March 1988). Limit theorems for a random convex set.
152. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On a theorem of Kuchler and Lauritzen.
153. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On the problem of types.
154. DOKSUM, K.A. (May 1988). On the correspondence between models in binary regression analysis and survival analysis.
155. LEHMANN, E.L. (May 1988). Jerzy Neyman, 1894-1981.
156. ALDOUS, D.J. (May 1988). Stein's method in a two-dimensional coverage problem.
157. FAN, J. (June 1988). On the optimal rates of convergence for nonparametric deconvolution problem.
158. DABROWSKA, D. (June 1988). Signed-rank tests for censored matched pairs.
159. BERAN, R.J. and MILLAR, P.W. (June 1988). Multivariate symmetry models.
160. BERAN, R.J. and MILLAR, P.W. (June 1988). Tests of fit for logistic models.
161. BREIMAN, L. and PETERS, S. (June 1988). Comparing automatic bivariate smoothers (A public service enterprise).
162. FAN, J. (June 1988). Optimal global rates of convergence for nonparametric deconvolution problem.
163. DIACONIS, P. and FREEDMAN, D.A. (June 1988). A singular measure which is locally uniform. (Revised by Tech Report No. 180).
164. BICKEL, P.J. and KRIEGER, A.M. (July 1988). Confidence bands for a distribution function using the bootstrap.
165. HESSE, C.H. (July 1988). New methods in the analysis of economic time series I.
166. FAN, JIANQING (July 1988). Nonparametric estimation of quadratic functionals in Gaussian white noise.
167. BREIMAN, L., STONE, C.J. and KOOPERBERG, C. (August 1988). Confidence bounds for extreme quantiles.
168. LE CAM, L. (August 1988). Maximum likelihood an introduction.
169. BREIMAN, L. (August 1988). Submodel selection and evaluation in regression-The conditional case and little bootstrap.
170. LE CAM, L. (September 1988). On the Prokhorov distance between the empirical process and the associated Gaussian bridge.
171. STONE, C.J. (September 1988). Large-sample inference for logspline models.
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.
173. MILLAR, P.W. (October 1988). Optimal estimation in the non-parametric multiplicative intensity model.
174. YOR, M. (October 1988). Interwinings of Bessel processes.
175. ROJO, J. (October 1988). On the concept of tail-heaviness.
176. ABRAHAM, D.M. and RIZZARDI, F. (September 1988). BLSS - The Berkeley interactive statistical system: An overview.

- 177. MILLAR, P.W. (October 1988). Gamma-funnels in the domain of a probability, with statistical implications.
- 178. DONOHO, D.L. and LIU, R.C. (October 1988). Hardest one-dimensional subfamilies.
- 179. DONOHO, D.L. and STARK, P.B. (October 1988). Recovery of sparse signals from data missing low frequencies.
- 180. FREEDMAN, D.A. and PITMAN, J.A. (Nov. 1988). A measure which is singular and uniformly locally uniform. (Revision of Tech Report No. 163).
- 181. DOKSUM, K.A. and HOYLAND, ARNLJOT (Nov. 1988). A model for step-stress accelerated life testing experiments based on Wiener processes and the inverse Gaussian distribution.
- 182. DALANG, R.C., MORTON, A. and WILLINGER, W. (November 1988). Equivalent martingale measures and no-arbitrage in stochastic securities market models.

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.