

A Risk Bound in Sobolev Class Regression

By

GRIGORI K. GOLUBEV

Institute for Information Transmission, Moscow

and

MICHAEL NUSSBAUM

Karl Weierstrass Institute, Berlin

AMS 1980 subject classifications. Primary 62G20, 62G05; secondary 62C20

Technical Report No. 227

November 1989

Department of Statistics
University of California
Berkeley California

A RISK BOUND IN SOBOLEV CLASS REGRESSION

By GRIGORI K. GOLUBEV AND MICHAEL NUSSBAUM

Institute for Information Transmission, Moscow and

Karl Weierstrass Institute, Berlin

Summary

For nonparametric regression estimation, when the unknown function belongs to a Sobolev smoothness class, sharp risk bounds for integrated mean square error have been found recently which improve on optimal rates of convergence results. The key to these has been the fact that under normality of the errors, the minimax linear estimator is asymptotically minimax in the class of all estimators. We extend this result to the nonnormal case, when the noise distribution is unknown. The pertaining lower asymptotic risk bound is established, based on an analogy with a location model in the independent identically distributed case. Attainment of the bound and its relation to adaptive optimal smoothing are discussed.

Running head

SOBOLEV CLASS REGRESSION

AMS 1980 subject classifications. Primary 62G20, 62G05; secondary 62C20

Key words and phrases. Nonparametric regression, asymptotic minimax L_2 -risk, smoothness ellipsoid, location model, shrinking Hellinger neighborhoods, adaptive bandwidth choice, experimental design, robust smoothing.

1. Introduction and main result. In the area of nonparametric curve estimation, some attention has recently been devoted to asymptotically minimax estimation for integrated mean square error. In a class of problems, it has been possible to improve the results on best obtainable rates of convergence, by finding the exact asymptotic value of the minimax risk in the class of all estimators. The constant involved represents the analog of Fisher's bound for asymptotic variances, for those "ill-posed" curve estimation problems where \sqrt{n} -consistency does not obtain. The key original result is due to Pinsker (1980); it concerned a filtering problem over ellipsoids in Hilbert space. The notion of ellipsoid is important in this context since Sobolev smoothness classes can be described in this way.

Consider observations

$$(1.1) \quad y_{in} = f(x_{in}) + \xi_i, \quad x_{in} \in [0, 1], i = 1, \dots, n$$

where $\{\xi_i\}$ are independent random variables with zero expectation, and the function f is to be estimated. The nonrandom design points x_{in} are assumed to be generated by a density g on $[0, 1]$ such that

$$(1.2) \quad \int_0^{x_{in}} g(t) dt = i/n$$

where g is assumed to be continuous and positive on $[0, 1]$. Let $L_2 = L_2(0, 1)$ be the Hilbert space of square integrable functions on $[0, 1]$, and let $\|\cdot\|$ denote the usual norm therein. Let, for natural m and $f \in L_2$, $D^m f$ denote the derivative of order m in the distributional sense, and let

$$W_2^m = \{f \in L_2; D^m f \in L_2\}$$

be the corresponding Sobolev space on the unit interval. The nonparametric class of functions to which f is assumed to belong is

$$W_2^m(P) = \{f \in W_2^m; \|D^m f\|^2 \leq P\}$$

for given m and $P > 0$. We are interested in the limiting minimax risk

$$(1.3) \quad \Delta = \liminf_n \sup_{\hat{f}} \inf_f n^{2m/(2m+1)} E_{f,n} \|\hat{f} - f\|^2$$

(sup over $f \in W_2^m(P)$, inf over all estimators \hat{f}). In the paper of Nussbaum (1985) the case of normal ξ_i with variance σ^2 and uniform design ($g \equiv 1$) was studied. The result was

$$(1.4) \quad \Delta = \gamma(m) \sigma^{2(1-r)} P^r$$

where

$$(1.5) \quad \gamma(m) = (2m+1)^{1/(2m+1)} (m/\pi(m+1))^{2m/(2m+1)}$$

is Pinsker's constant. The method of proof was to show that with the help of some spline smoothing theory, the regression problem can be reduced to the original filtering problem. Normality of the errors was essential there. For some closely related results see Speckman (1985).

The present paper addresses the problem of a risk bound for unknown error distribution. For the heuristics it is helpful to consider an analogy with mean estimation. The sample mean of independent identically distributed observations with mean ϑ

$$(1.6) \quad y_i = \vartheta + \xi_i, \quad i = 1, \dots, n$$

is an asymptotically efficient estimator of ϑ when (A) the errors ξ_i and $N(0, \sigma^2)$, (B) when, loosely speaking, the distribution of the errors is unknown. The result (B) is due to the fact that the sample mean is a linear functional of the empirical distribution function, see Levit (1975). It will be instructive first to formulate the risk bound for the mean in the "semiparametric" form, where the distribution of the errors ξ_i appears as an infinite dimensional nuisance parameter, varying in a shrinking Hellinger neighborhood of some "central" measure Q_0 . Let for distributions Q_0, Q

$$H(Q_0, Q) = (\int ((dQ_0)^{1/2} - (dQ)^{1/2})^2)^{1/2}$$

be the Hellinger distance. Consider a sequence τ_n such that

$$\tau_n \rightarrow 0, \quad \tau_n n^{1/2} \rightarrow \infty \quad \text{as } n \rightarrow \infty.$$

Introduce the set of probability measures on the real line

$$(1.7) \quad \mathcal{Q}_n^H = \{Q; H(Q_0, Q) \leq \tau_n, \quad E_Q \xi = 0\}.$$

The central measure is assumed to have zero expectation, finite second moment, and to fulfill the following regularity condition: if Q_{0t} denotes the shifted measure $Q_{0t}(\cdot) = Q_0(\cdot + t)$ then

$$(1.8) \quad H(Q_{0t}, Q_0) = o(t) \quad \text{as } t \rightarrow 0.$$

We can now formulate a lower asymptotic risk bound, where the infimum is taken over all estimators $\hat{\vartheta}$ of the mean ϑ at sample size n .

PROPOSITION 1. Assume that in model (1.6), the ξ_i are independent with distribution $Q \in \mathcal{Q}_n^H$, where the central measure Q_0 has zero expectation, second moment σ^2 and fulfills (1.8). Then for all ϑ_0 we have

$$\liminf_n \inf_{\hat{\vartheta}} \sup_{|\vartheta - \vartheta_0| \leq \tau_n, Q \in \mathcal{Q}_n^H} n E_{\vartheta, Q, n}(\hat{\vartheta} - \vartheta)^2 \geq \sigma^2.$$

The sample mean \bar{y}_n will indeed attain this bound when the appropriate uniform convergence of its variance is ensured, e.g. by a moment condition. Suppose that both Q_0 and Q are in the set

$$(1.9) \quad \mathcal{Q}_c^M = \{Q; E_Q \xi^4 < c\}$$

for some $c > 0$. Then we have (compare relation (3.1) below)

$$E_{\vartheta, Q, n}(\bar{y}_n - \vartheta)^2 \rightarrow \sigma^2 \quad \text{as } n \rightarrow \infty$$

uniformly over $(\vartheta, Q): |\vartheta - \vartheta_0| \leq \tau_n, Q \in \mathcal{Q}_n^H \cap \mathcal{Q}_c^M$. This means that the risk bound of proposition 1 is sharp and that the sample mean is asymptotically efficient, provided that the lower bound holds also on the narrowed parameter set.

PROPOSITION 2. If, in addition, Q_0 is in a class \mathcal{Q}_c^M for some $c > 0$ then

$$\liminf_n \inf_{\hat{\vartheta}} \sup_{|\vartheta - \vartheta_0| \leq \tau_n, Q \in \mathcal{Q}_n^H \cap \mathcal{Q}_c^M} n E_{\vartheta, Q, n}(\hat{\vartheta} - \vartheta)^2 = \sigma^2.$$

As the bound is attained by \bar{y}_n , proposition 2 holds relative to the class of estimators $\hat{\vartheta}$ which do not depend on Q_0 . The shrinking Hellinger ball

model is appropriate when investigating the sample mean as an estimator of the mean functional of a distribution (Levit (1975), see also Ibragimov, Khasminski (1981), chap. 4.1). Proposition 2 is in fact a reformulation of these results for the “parameter + noise” model (1.6) (note the condition $E_Q \xi = 0$ in (1.7)). This is a convenient way of describing the efficiency of the sample mean when the error distribution is unknown, in analogy to the case of normal errors. Proposition 2 can be extended to parametric linear regression, stating efficiency of the Gauss-Markov linear estimator. However from studies in the context of robustness (e.g. Beran (1982)) one particular feature has emanated: the model giving meaningful results here is one of nonidentically distributed errors. The distributions of ξ_i will still vary in a small neighborhood of some (unknown) central measure Q_0 , but will in general be different.

The Sobolev class model can be regarded as an extended or nearly linear regression model. Define $r = 1/(2m + 1)$; then the normalizing factor of the risk in (1.3) is n^{1-r} . The shrinking rate of the distribution neighborhoods to define will be tied to this factor. Let τ_n be a sequence such that

$$(1.10) \quad \tau_n \rightarrow 0, \quad \tau_n n^{(1-r)/2} \rightarrow \infty \quad \text{as } n \rightarrow \infty.$$

Consider a central measure Q_0 as above and a neighborhood \mathcal{Q}_n^H , defined in terms of the new τ_n (see (1.7)). We will also consider a “moment neighborhood” \mathcal{Q}_c^M containing Q_0 . Denote the distribution of (ξ_1, \dots, ξ_n) in model (1.1) by Π , and define a set of product measures

$$\mathcal{Q}_n^* = \{\otimes_{i=1}^n Q_i; Q_i \in \mathcal{Q}_n^H \cap \mathcal{Q}_c^M, \quad i = 1, \dots, n\}.$$

The distribution model for the noise in (1.1) will be “ $\Pi \in \mathcal{Q}_n^*$ ”. We study the asymptotic minimax risk

$$(1.11) \quad \Delta = \liminf_n \inf_{\hat{f}} \sup_{f, \Pi} n^{1-r} E_{f, \Pi, n} \|\hat{f} - f\|^2.$$

Here the supremum is taken over $(f, \Pi) \in W_2^m(P) \times \mathcal{Q}_n^*$, while the infimum is taken over all estimators \hat{f} at sample size n which may depend on m , P and Q_0 . Our main result is as follows.

THEOREM 1. Suppose that in the model (1.1) the design points are generated according to (1.2), and the central measure defining the neighborhood

Q_n^* fulfills the conditions of proposition 2. Then

$$\Delta \geq \gamma(m)(\sigma^2 d)^{1-r} P^r,$$

where $\sigma^2 = E_{Q_0} \xi^2$, $d = \int_0^1 g^{-1}(x) dx$.

This represents the desired extension of the result (1.4) to the case of unknown error distribution. We also claim that this risk bound is sharp, and we will provide evidence on the basis of a first two moments argument for linear estimators (section 3).

An extension to the case of weighted L_2 -loss can be given as follows. Let w be a continuous and positive function on $[0, 1]$, and consider a loss given by

$$(1.12) \quad \int_0^1 w(x)(\hat{f}(x) - f(x))^2 dx.$$

Such a loss arises naturally when one considers the “design loss” $n^{-1} \sum_{i=1}^n (\hat{f}(x_i) - f(x_i))^2$ which may be viewed as a discrete approximation to (1.12) for $w = g$.

THEOREM 2. Let Δ_w be the analog of (1.11) when the loss (1.12) is substituted for $\|\hat{f} - f\|^2$. Then, under the conditions of theorem 1,

$$\Delta_w \geq \gamma(m)(\sigma^2 d)^{1-r} P^r,$$

where $d = \int_0^1 g^{-1}(x) w^{1+1/2m}(x) dx$.

We note the following implications for experimental design and robustness.

REMARK 1. Optimal designs of nonparametric regression experiments have been studied for a variety of settings and criteria. For the asymptotic L_2 -risk we mention Agarwal, Studden (1980), Müller (1984); for a result involving Sobolev classes see Spruill (1984). As the present bound is sharp for given design, it is of interest to try to minimize it further. For given w , we obtain with $\alpha = (2m + 1)/4m$ from Jensen’s inequality

$$d = \int (g(x) w^{-\alpha}(x))^{-1} w^{\alpha}(x) dx \geq \left(\int w^{\alpha}(x) dx \right)^2$$

so that $g = w^{\alpha} / \int w^{\alpha}$ is optimal. In particular, for L_2 -loss ($w \equiv 1$) the uniform design is best. On the other hand, when g and w are tied by $w = g$

(“design loss”) then $d = 0$ is achieved in the limit by taking all x_i equal, which is in agreement with intuition since the rate of convergence then changes.

REMARK 2. The Hellinger neighborhood model for the noise distribution adopted here resembles the light contamination neighborhoods occurring in the robustness study of Beran (1981). The crucial difference is the additional moment restriction (1.9) which ensures robustness of the sample mean (when robustness is given the asymptotic minimax definition). The analogy with the location model exploited here quite naturally suggests an asymptotic minimax theory for robust smoothing, based on infinitesimal distribution neighborhoods expressing heavier contamination (cp. Millar (1983)).

The problem of best possible estimation in terms of optimal rates of convergence has been extensively investigated (Ibragimov, Khasminski (1982), Stone (1982), Birgé (1983)). In our study on the level of constants a global error criterion is adopted (L_2 -loss); for comparable recent results on functionals (like the value of f at a point) see Ibragimov, Khasminski (1984), Donoho and Liu (1988).

In section 2, we review the background of the risk evaluation (1.4) in the normal case. In section 3 we argue that our new bounds are attainable, and discuss some recent results indicating that this should be possible adaptively. Refined bounds are the topic of section 4, and proofs are in section 5. An appendix contains a short proof of an auxiliary result related to the Hajek-Le Cam asymptotic minimax theorem.

The following notations are adopted. $\int f$ means integral with respect to Lebesgue measure; $a \sim b$ means $a = b(1 + o(1))$.

2. Some background on L_2 -optimal smoothing. For additional insight, we try to elucidate why, under normality, the minimax linear estimator is asymptotically overall minimax (Pinsker (1980)). This task is facilitated by a related minimax identity due to Pilz (1986). Suppose an n -dimensional observed random vector η has expectation ϑ and covariance matrix Σ , where $\vartheta \in \Theta$, and Θ is known to be a compact subset of \mathbb{R}^n which is symmetric about the origin. Consider the class of linear estimators of ϑ : $\hat{\vartheta}_B = B\eta$, where B is a matrix of fixed coefficients. Their risk under squared Euclidean loss is

$$(2.1) \quad E_{\vartheta} \|\hat{\vartheta}_B - \vartheta\|^2 = \text{tr}[(I - B)\vartheta\vartheta'(I - B')] + \text{tr}[B\Sigma B'] =: R(B, \vartheta\vartheta').$$

Along with “minimax” or “Bayesian” we shall employ the terms “minimax (or Bayesian) linear”, meaning the respective extremal property within this special class of estimators. Let ν be an arbitrary prior distribution on Θ , and consider the mixed risk of $\hat{\vartheta}_B$. It can be expressed as

$$(2.2) \quad E_{\nu} R(B, \vartheta\vartheta') = R(B, M_{\nu}), \quad M_{\nu} = E_{\nu} \vartheta\vartheta'.$$

Let \mathcal{M} be the set of all second moment matrices M_{ν} when ν is concentrated on Θ . Clearly (2.2) implies

$$\sup_{\vartheta \in \Theta} R(B, \vartheta\vartheta') = \sup_{M \in \mathcal{M}} R(B, M).$$

According to the result of Pilz (1986) there is a saddle point (B^*, M^*) such that

$$R(B^*, M^*) = \sup_{\vartheta \in \Theta} R(B^*, \vartheta\vartheta') = \inf_B R(B, M^*).$$

Hence $\hat{\vartheta}_{B^*}$ is minimax linear, and it is Bayesian linear for a prior on Θ having second moment matrix M^* (a least favorable prior). If $\hat{\vartheta}_{B^*}$ were also Bayesian with respect to such a prior it would be minimax. But if η is normal then $\hat{\vartheta}_{B^*}$ is Bayesian with respect to a normal prior $N(O_n, M^*)$ on \mathbb{R}^n . This prior is not concentrated on Θ but if in some asymptotic setting it tends to concentrate on Θ then $\hat{\vartheta}_{B^*}$ can be expected to be nearly minimax.

In the ellipsoid framework of Pinsker (1980) Θ is e.g. a set

$$(2.3a) \quad \Theta^m(P) = \{\vartheta \in \mathbb{R}^n; \sum_{j=1}^n a_j \vartheta_j^2 \leq P\}, \quad a_j = (\pi j)^{2m}, \quad j = 1, \dots, n$$

while η has a structure

$$(2.3b) \quad \eta_j = \vartheta_j + n^{-1/2}\xi_j, \quad j = 1, \dots, n,$$

ξ_j being independent normal with variance σ^2 . In the saddle point pair (B^*, M^*) both matrices are diagonal with respective diagonal elements

$$(2.4a) \quad b_j^* = b(jt), \quad m_j^* = n^{-1}\sigma^2\beta(jt), \quad j = 1, \dots, n$$

where the functions b, β are defined on $(0, \infty)$ by

$$(2.4b) \quad b(x) = (1 - (\pi x)^m)_+, \quad \beta(x) = b(x)(1 - b(x))_1$$

and $t > 0$ is chosen such that $\sum_{j=1}^n a_j m_j^* = P$. This latter identity implies that for $n \rightarrow \infty$, $N(O_n, M^*)$ is asymptotically concentrated on $\Theta^m(P')$ for any $P' > P$. Then $R(B^*, M^*)$ is asymptotic to the minimax risk over $\Theta^m(P)$. From (2.1) and (2.4) we obtain

$$(2.5) \quad R(B^*, M^*) = n^{-1}\sigma^2 \sum_{j=1}^n b_j^*.$$

The above choice of t implies

$$(2.6) \quad t \sim n^{-r}(\sigma^2/P)^r \mu, \quad \mu^{2m+1} = \int b(1-b).$$

We then obtain from (2.5) and (1.5)

$$(2.7) \quad R(B^*, M^*) \sim (\sigma^2/n)^{1-r} P^r \mu^{-1} \int b = (\sigma^2/n)^{1-r} P^r \gamma(m).$$

For recent results on more general sets Θ and an interesting geometric perspective see Donoho, Macgibbon and Liu (1988).

Consider now the Sobolev class regression model (1.1) with $g \equiv 1$ and normal noise ξ_j with variance σ^2 . In Speckman (1985) and Nussbaum (1985) it was shown how to use an orthogonal transformation in \mathbb{R}^n (a spline analog of the Fourier transform on $[0, 1]$) to reduce the model to one of (essentially) the type (2.3). The risk bound (1.4) is then equivalent to Pinsker's (1980) result.

For the nonnormal errors case, the basic reasoning is that a smooth function can be well approximated by one which is constant on small intervals. The

problem would be then to estimate a “local” mean, in the presence of random noise ξ_i . When the ξ_i are independently distributed with given, possibly non-normal distribution Q_0 one can apply maximum likelihood theory to find a risk bound which involves the Fisher information of Q_0 in the location problem. Such a result was obtained in Golubev (1984). However our present goal is to emulate the efficiency of the sample mean as described by proposition 2. We establish that the same risk bound as in the normal case is valid for a large class of distributions Q_0 , when a small Hellinger neighborhood around Q_0 is taken into account.

3. Attainability. A complete proof is beyond the scope of this paper, but we provide theoretical backing for our claim that the bounds are indeed attainable.

3.1. Consider first the regression model (1.1) with $g \equiv 1$ and normal noise ξ_j with variance σ^2 . From the previous section it is clear that the risk bound (1.4) is attained by the minimax linear estimator, given in the frequency domain by coefficients b_j^* (see (2.4a)). In the “time domain” (1.1) this corresponds to a certain linear spline smoothing procedure. In (2.4), the function b can be interpreted as a filter shape, while t serves as a smoothing parameter. The relation (2.6) gives the appropriate choice of t , in dependence on P and σ^2 .

3.2. In the nonnormal case, when the noise in (1.1) is uncorrelated with zero expectation and variance σ^2 , the risk behaviour of the minimax linear smoothing spline estimator of 3.1 remains unchanged. Indeed, the risk of linear estimators under quadratic loss depends only on the first two moments of the observations, cp. (2.1). Now, the actual noise distribution model in theorem 1 ensures that $\text{var } \xi_i \sim \sigma^2$. Indeed for $Q \in \mathcal{Q}_n^H \cap \mathcal{Q}_c^M$ we have

$$(3.1) \quad |E_Q x^2 - \sigma^2|^2 = \left| \int x^2 d(Q - Q_0) \right|^2 \\ \leq \left(\int x^4 ((dQ)^{1/2} + (dQ_0)^{1/2})^2 \right) H^2(Q_0, Q) \leq 4cH^2(Q_0, Q) = o(1).$$

Thus it is obvious that the bound of theorem 1 is attainable, for $g \equiv 1$ and known P, σ^2 .

3.3. Speckman (1985) established that the case of general design density g in (1.2) can be treated as in 3.2, if the a_j defining the ellipsoid are properly adjusted. As a result, we obtain attainability in theorem 2 for $w = g$, still on the basis of the minimax linear smoothing spline. The general case of theorem 2, with w, P, σ^2 known, can also be covered by linear estimators, but we invoke here the nonlinear (adaptive) smoother of point 3.6 below.

3.4. Up to now σ^2 , i.e. the variance of the “central measure” Q_0 , has been assumed known. But the basic motivation of the present paper is to give a risk bound for unknown noise distribution. As (2.6) shows, σ^2 enters in the smoothing (or bandwidth) parameter of the optimal procedure, along with P . Thus an unknown σ^2 leads to a similar problem as an unknown P , namely

adaptive (or automatic) selection of the smoothing parameter based on the sample. However, when P is known, the plug-in type procedure based on an estimate of σ^2 is relatively easy to treat theoretically. In the present model σ^2 can be estimated with parametric convergence rate, see Rice (1984), Li (1985).

3.5. In the problem of adaptive smoothing parameter selection there has been much progress recently; for a survey see Marron (1988). In the present context one could ask for estimators which attain the bound of theorem 1 without depending on P and σ^2 . In fact any combination of the filter shape b (see (2.4)) with a known optimal bandwidth selector such as cross-validation, empirical risk minimization or plug-in (estimating σ^2 and $\|D^m f\|^2$) could be considered. Note that the decision-theoretic risk and the minimax aspect are not at the center of many of the recent investigations (Rice (1984), Härdle and Marron (1985), Li (1986), Marron (1987)). Earlier results on risk performance of the plug-in method are due to Woodrofe (1970), Nadaraya (1974) (for density estimation, without the minimax aspect). Speckman (1985) came close to proving minimax risk optimality of the appropriate smoothing spline estimator with bandwidth chosen by generalized cross-validation (in the setting of 3.1).

3.6. For our attainment question, on the adaptive level, the most relevant result is in Golubev (1987). For a Gaussian model similar to (2.3), with known σ^2 and m but unknown P , it is proved that the bound (2.7) is attainable by an adaptive smoother with plug-in type bandwidth selection. Actually the estimator is a refinement based on the following idea. Return to the “time domain”, i.e. to the regression model (1.1) on $[0, 1]$. Let $\{A\} = \mathbb{A}$ be a partition of $[0, 1]$ into intervals A of equal length. When $f \in W_2^m(P)$ then the restriction of f to any $A \in \mathbb{A}$ is in a Sobolev class on that interval, i.e.

$$(3.2) \quad \int_A (D^m f)^2 \leq P_A, \quad A \in \mathbb{A}, \quad \sum P_A = P.$$

Here the P_A are unknown even when P is known. Now, on each A use an adaptive estimator rescaled to that interval. The resulting estimator on $[0, 1]$ will then be adaptive also with respect to P . Furthermore, when the length of the A 's tends to zero sufficiently slowly this estimator will also be risk optimal with respect to weighted L_2 -loss (1.12), even though it does not depend on w . As this result holds under normality, the above arguments 3.2,

3.3 suggest that the bound of theorem 2 is attained by adaptive estimators, where at most an additional moment assumption for the noise would come into play.

The “locally adaptive” procedure described is optimal in an even stronger sense, see section 4.2 below. The idea of a locally varying adaptive bandwidth choice is also developed by Müller and Stadtmüller (1987).

3.7. The question of adaptivity with respect to the degree of smoothness m is also of interest. For minimax *rate* optimality, the problem was raised by Stone (1982) and solved by Härdle and Marron (1985). Simultaneous choice of kernel order and bandwidth by cross-validation was treated by Hall and Marron (1988). We briefly review here the method of adaptive estimation which has been developed by Efroimovich and Pinsker (1985) and independently by Rudzkis (1985). In the ellipsoid model (2.3) one could ask for the linear estimator $\hat{\vartheta}_B$ which at a particular $\vartheta \in \Theta^m(P)$ minimizes the risk $R(B, \vartheta\vartheta')$; call its coefficient matrix $B(\vartheta)$. In what follows it suffices to consider only matrices B of diagonal kind, i.e. given by a set of coefficients b_j , $j = 1, \dots, n$. Then $B(\vartheta)$ is given by

$$(3.3) \quad b_j(\vartheta) = \vartheta^2 / (n^{-1}\sigma^2 + \vartheta_j^2), \quad j = 1, \dots, n.$$

If the unknown $b_j(\vartheta)$ could be determined from the data, the resulting estimator might asymptotically dominate any linear estimator, and hence attain the minimax bound. Plugging in the η_j for ϑ_j in (3.3) does not yield the desired result. Consider now a restriction on the set of coefficients and the corresponding minimizer $\tilde{B}(\vartheta)$ of $R(B, \vartheta\vartheta')$, such that (A) the set is wide enough so that $R(\tilde{B}(\vartheta), \vartheta\vartheta') \sim R(B(\vartheta), \vartheta\vartheta')$ as $n \rightarrow \infty$, (B) it is narrow enough to ensure that $\tilde{B}(\vartheta)$ is estimable. A solution is to require that b_j as a function of j is constant between indices k^2 , $k = 1, 2, \dots$. The resulting estimator of ϑ is shown to be asymptotically minimax over any ellipsoid Θ from a large class; in particular in the Sobolev class model it is adaptive with respect to m and P . For further results on this type of smoothers in density estimation see Efroimovich (1985), Kazbaras (1986). Clearly the method is applicable in principle also in the present regression model.

4. Localized bounds. In theorems 1 and 2 the supremum with respect to the regression function f is taken with respect to the whole Sobolev class $W_2^m(P)$. It is compelling to consider some shrinking neighborhood setting also here, in analogy to the noise distribution model adopted. A localization can be achieved in two ways.

4.1. Let f_0 be some function serving as a “center of localization”. The bound of theorem 2 remains valid when the supremum with respect to f is taken or

$$(4.1) \quad \{f; f - f_0 \in W_2^m(P), \|f - f_0\| \leq \tau_n\}$$

where τ_n fulfills (1.10). As usual f_0 may be assumed known for the lower risk bound. The proof is continued in section 5.4. Attainment over a set (4.1), with f_0 unknown, can be shown if f_0 is of higher smoothness than f , e.g. if $f_0 \in W_2^{m+1}$. To see this, consider the analogous problem in the ellipsoid model (2.3). Suppose that instead of (2.3b) we have

$$\eta_j = \vartheta_{0j} + \vartheta_j + n^{-1/2}\xi_j, \quad j = 1, \dots, n, \quad \sum_{j=1}^n j^{2(m+1)}\vartheta_{0j}^2 < \infty.$$

In the optimal filter (2.4a), replace the first $[n^r / \log n]$ coefficients b_j^* by 1. In this way the influence of the ϑ_{0j} in the worst case asymptotic risk is made negligible.

4.2. Another possibility consists in narrowing the class $W_2^m(P)$ as follows. Observe that the prior distribution on f constructed in section 5.3 is not only asymptotically concentrated on $W_2^m(P)$ but, more specifically, on the ellipsoidal shell $\{f; \delta P \leq \|D^m f\|^2 \leq P\}$, for some $\delta < 1$. One might now pass to subintervals A of $[0, 1]$ and ellipsoidal shells on each of them, possibly with different radii P_A (compare relation (3.2)). Refinement of the partition leads to a priori sets for f which prescribe a given approximate mass distribution of the squared m -th derivative on $[0, 1]$. Let v be a continuous positive function, and τ_n^* be a sequence: $\tau_n^* \rightarrow 0$, $\tau_n^* n^{r/2} \rightarrow \infty$. Consider a class

$$\mathcal{B}_n(v) = \{f \in W_2^m; \sup_{x \in [0,1]} |\int_0^x ((D^m f)^2 - v)| \leq \tau_n^*\}.$$

Let $\Delta_{w,v}$ be the analog of Δ_w when $W_2^m(P)$ is substituted by $\mathcal{B}_n(v)$; then

$$(4.2) \quad \Delta_{w,v} \geq \gamma(m)\sigma^{2(1-r)} \int wv^r g^{r-1}.$$

The proof of sketched in section 5.4. For Gaussian noise and continuous observations, this bound and its attainability for unknown w and v have been established by Golubev (1987). The estimator employed is described in section 3.6.

5. Proofs.

5.1. Analytic preliminaries. For establishing the lower risk bound it is convenient to restrict the parameter space by boundary conditions on the unknown f . Consider the Sobolev space \dot{W}_2^m with boundary conditions on $[0, 1]$:

$$\dot{W}_2^m = \{f \in W_2^m; (D^k f)(0) = (D^k f)(1) = 0, \quad k = 0, \dots, m-1\};$$

it is a Hilbert subspace of W_2^m with respect to the norm $(\|f\|^2 + \|D^m f\|^2)^{1/2}$. We will make use of the results on the spectral theory of differential operators, see e.g. Agmon (1968). There exists a basis φ_j , $j = 1, 2, \dots$ in \dot{W}_2^m such that, if (\cdot, \cdot) denotes the inner product in $L_2(0, 1)$,

$$(\varphi_i, \varphi_j) = \delta_{ij}, \quad (D^m \varphi_i, D^m \varphi_j) = \lambda_j \delta_{ij}, \quad i, j = 1, 2, \dots$$

where

$$0 < \lambda_1 < \lambda_2 < \dots$$

and the asymptotics of the eigenvalues λ_j is given by

$$(5.1) \quad \lambda_j \sim (\pi j)^{2m}, \quad j \rightarrow \infty.$$

The boundary conditions ensure that, when the functions φ_j are continued by zero outside $[0, 1]$, these functions belong to the Sobolev space of order m on any interval containing $[0, 1]$. Furthermore, this property allows the construction of another orthogonal system in \dot{W}_2^m which is obtained by a change of scale. Fix a natural number q ; later we will let q tend to infinity with n . Define functions

$$(5.2) \quad \varphi_{jkq}(x) = q^{1/2} \varphi_j(qx - k + 1), \quad k = 1, \dots, q, \quad j = 1, 2, \dots$$

Each function φ_{jkq} is in \dot{W}_2^m , has support $[(k-1)q^{-1}, kq^{-1}]$, and

$$(5.3) \quad (\varphi_{ikq}, \varphi_{jkq}) = \delta_{ij}, \quad (D^m \varphi_{ikq}, D^m \varphi_{jkq}) = q^{2m} \lambda_j \delta_{ij}.$$

Furthermore, fix a natural s , and define $W(q, s, P)$ as the intersection of the linear span of φ_{jkq} , $j = 1, \dots, s$, $k = 1, \dots, q$ with $W_2^m(P)$. From (5.3) we obtain that for $f \in W(q, s, P)$

$$(5.4) \quad \|D^m f\|^2 = \sum_{j=1}^s \sum_{k=1}^q q^{2m} \lambda_j (\varphi_{jkq}, f)^2$$

and obviously $W(q, s, P)$ is nonempty. Restricting f to this set, we reduce the problem to the one of estimating the “local” Fourier coefficients $f_{jkq} = (\varphi_{jkq}, f)$. The indices q and n will frequently be dropped from notation in the sequel.

The functions φ_{jk} are orthonormal in $L_2(0, 1)$; we have to take into account that our observation model is discrete. Observe that, under the assumptions made on the regression design $\{x_j\}$, the Kolmogorov distance between the distribution function G having density g and its “empirical” counterpart G_n (assigning mass n^{-1} to x_j) is $O(n^{-1})$. The following statement then can be proved in the same manner as lemma 4.2. (i) of Cox (1984).

LEMMA 1. Let f_1, f_2 be functions from W_2^m . Then

$$|\int f_1 f_2 d(G_n - G)| < C n^{-1} (\|f_1\| + \|D^m f_1\|) (\|f_2\| + \|D^m f_2\|)$$

where C does not depend on f_1, f_2, n .

Define $g_k = g(kq^{-1})$, $k = 1, \dots, q$. In the following result concerning the functions φ_{jk} , $j \leq s$, $k \leq q$, the number s will remain fixed until the last step in the proof of theorem 1.

LEMMA 2. Suppose that $q \rightarrow \infty$, $q^{2m}/n \rightarrow 0$. Then

$$g_k^{-1} \int \varphi_{ik} \varphi_{jk} dG_n = \delta_{ij} + o(1)$$

uniformly over $i, j \leq s$, $k \leq q$.

PROOF. From (5.3) it follows that

$$\|\varphi_{jk}\| + \|D^m \varphi_{jk}\| = 1 + q^m \lambda_j^{1/2}.$$

Furthermore, the assumptions on g imply that

$$g_k^{-1} (\int \varphi_{ik} \varphi_{jk} g) = (\varphi_{ik}, \varphi_{jk}) + o(1)$$

uniformly over $k \leq q$ and all i, j . The result follows now from lemma 1.

5.2. Local regression models. By restricting f to the subset $W(q, s, P)$ of the Sobolev class $W_2^m(P)$, we achieve that the observations y_i have a structure

$$(5.5) \quad y_i = \sum_{j=1}^s \varphi_{jk}(x_i) f_{jk} + \xi_i, \quad i = 1, \dots, n$$

where k above is uniquely defined by $i \in \mathfrak{S}(k) := \{i; x_i \in q^{-1}(k-1, k]\}$. This may be construed as a collection of q linear regression models, each accounting for observations in the interval $q^{-1}(k-1, k]$ and having s parameters. The parameters f_{jk} satisfy (cp. (5.4))

$$\sum_{j=1}^s \sum_{k=1}^q q^{2m} \lambda_j f_{jk}^2 \leq P$$

while the risk can now be bounded by

$$(5.6) \quad E\|\hat{f} - f\|^2 \geq E \sum_{k=1}^q \sum_{j=1}^s (\hat{f}_{jk} - f_{jk})^2.$$

At this point, let us specify q by

$$q = [Kn^r],$$

where K , assumed fixed as well as s , will be selected later. Let us rescale the parameter vector in each local model by the proper normalizing factor which in view of lemma 2 is $(ng_k)^{1/2}$. Define vectors

$$h_k = (ng_k)^{1/2} (f_{jk})_{j=1, \dots, s},$$

$$\bar{\varphi}_i = (ng_k)^{-1/2} (\varphi_{jk}(x_i))_{j=1, \dots, s}, \quad i \in \mathfrak{S}(k).$$

Then (5.5) transforms to

$$(5.7) \quad y_i = \bar{\varphi}_i' h_k + \xi_i, \quad i \in \mathfrak{S}(k)$$

for $k = 1, \dots, q$. Here the disturbance distributions are assumed to be in $\mathcal{Q}_n^H \cap \mathcal{Q}_c^M$ and are as yet unspecified. We will now select them in accordance with the method of least favorable parametric subfamilies. Consider a bounded function ψ on \mathcal{R} such that, if u is the identity map in \mathcal{R} ,

$$\int \psi dQ_0 = 0, \quad \int u \psi DQ_0 = 1.$$

For $h \in \mathbf{R}^s$, let $Q_i(h)$ be the measure defined by

$$dQ_i(h) = (1 + h' \bar{\varphi}_i \psi) dQ_0.$$

For the vector $\bar{\varphi}_i$ we find the bound

$$(5.8) \quad \|\bar{\varphi}_i\|^2 = O(n^{-1} q \sup_{j \leq k} \sup_x |\varphi_j(x)|^2) = O(n^{r-1}).$$

Thus, when τ_n satisfies (1.10), we infer that for $\|h\|^2 \leq \tau_n^2 n^{1-r}$ and sufficiently large n all $Q_i(h)$ are probability measures. Let $Q_i^*(h)$ be the shifted measure

$$Q_i^*(h)(\cdot) = Q_i(h)(\cdot + \bar{\varphi}_i' h).$$

LEMMA 3. Let τ_n be the sequence occurring in the definition of \mathcal{Q}_n^* , and t_n be such that $t_n \rightarrow \infty$, $t_n = o(\tau_n n^{(1-r)/2})$ as $n \rightarrow \infty$. Then for sufficiently large n , the set of measures $\{Q_i^*(h); \|h\| \leq t_n, i \in \{1, \dots, n\}\}$ is contained in $\mathcal{Q}_n^H \cap \mathcal{Q}_c^M$.

PROOF. For the expectation we have

$$\int u dQ_i^*(h) = \int u dQ_i(h) - \bar{\varphi}_i' h = 0.$$

Let $Q_i^{**}(h)$ be the shifted measure $Q_0(\cdot + \bar{\varphi}_i' h)$. Then for the Hellinger distance we have

$$(5.9) \quad H(Q_i^*(h), Q_0) \leq H(Q_i^*(h), Q_i^{**}(h)) + H(Q_i^{**}(h), Q_0).$$

Here the first term on the right hand side equals $H(Q_i(h), Q_0)$ and can be bounded by

$$(5.10) \quad O(\bar{\varphi}_i' h) = O(t_n n^{(r-1)/2}) = o(\tau_n)$$

in view of (5.8). The second term on the right hand side of (5.9) can be bounded similarly in view of condition (1.8). Hence all $Q_i^*(h)$ are in \mathcal{Q}_n^H , for n sufficiently large, $\|h\| \leq t_n$. For the fourth moment we find

$$\begin{aligned} \int u^4 dQ_i^*(h) &= \int (u - \bar{\varphi}_i' h)^4 (1 + \bar{\varphi}_i' h \psi) dQ_0 \\ &= \int u^4 dQ_0 + O(\bar{\varphi}_i' h) \end{aligned}$$

so that all $Q_i^*(h)$ are in \mathcal{Q}_c^M for sufficiently large n . \square

Now, in (5.7), assume that $\|h_k\| \leq t_n$ and that $\text{distr}(\xi_i) = Q_i^*(h_k)$. Lemma 3 guarantees that this is compatible with the initial errors distribution model “ $\Pi \in \mathcal{Q}_n^*$ ”. It is equivalent to the model

$$(5.11) \quad \text{distr}(y_i) = Q_i(h_k), \quad i \in \mathfrak{I}(k)$$

for $k = 1, \dots, q$, where the parameters $h_k = (h_{jk})_{j=1, \dots, s}$ are now restricted by

$$(5.12) \quad \sup_{k \leq q} \|h_k\| \leq t_n, \quad \sum_{j=1}^s \sum_{k=1}^q q^{2m} \lambda_j n^{-1} g_k^{-1} h_{jk}^2 \leq P.$$

Our next goal is to establish that each of the k distributional models (5.11) converges to a normal shift model (local asymptotic normality). To achieve uniformity, we let $k(n)$ be an arbitrary sequence $1 \leq k(n) \leq q$ and consider the logarithmic likelihood ratio in the $k(n)$ -th model of (5.11) (for hypothesis $h = 0$)

$$\Lambda(h) = \sum_{i \in \mathfrak{I}(k(n))} \log(1 + \bar{\varphi}_i' h \psi(\xi_i)),$$

where ξ_i are independent with distribution Q_0 . In the same setting, define σ_*^2 and an \mathbb{R}^2 -valued random variable L by

$$\sigma_*^2 = (E\psi^2(\xi_1))^{-1}, \quad L = \sum_{i \in \mathfrak{I}(k(n))} \bar{\varphi}_i \psi(\xi_i).$$

LEMMA 4. The random vector L converges in distribution to a multivariate normal $N(O_s, \sigma_*^{-2} I_s)$, and for each $h \in \mathbb{R}^s$ we have

$$\Lambda(h) - h' L = -\|h\|^2 \sigma_*^{-2} / 2 + o_P(1).$$

PROOF. First note that lemma 2 and (5.10) imply

$$\sum_{i \in \mathfrak{I}(k(n))} (\bar{\varphi}_i' h)^2 \rightarrow \|h\|^2, \quad \sup_{i \in \mathfrak{I}(k(n))} (\bar{\varphi}_i' h)^2 = o(1).$$

The proof is concluded via the expansion

$$\log(1 + t) = t - t^2/2 + o(t^2)$$

and the Lindeberg-Feller theorem. \square

Note that the function $\psi(x)$ can be selected to approximate x/σ^2 in the norm of $L_2(Q_0)$; then σ_*^2 approximates σ^2 . The above lemma means that each model (5.11) converges to $\{N(h, \sigma_*^2 I_s), h \in \mathbb{R}^s\}$ through an arbitrary sequence $k = k(n)$.

5.3. Main argument of proof. We shall introduce a prior distribution on the parameter in the collection of “local” models (5.11). The h_k will be independent identically distributed random variables such that the prior measure tends to concentrate on the space given by the restrictions (5.12). Since the models (5.11) are asymptotically normal and independent, we can evaluate the posterior risk by the general result proved in the appendix.

Let \mathcal{R} be the set in \mathbb{R}^{qs} defined by the inequalities (5.12).

LEMMA 5. Let ν be a measure on \mathbb{R}^s with bounded support fulfilling

$$(5.13) \quad \int \sum_{j=1}^s \lambda_j x_j^2 d\nu(x) < P/K^{1/r}d.$$

Let $\nu^q = \nu \otimes \dots \otimes \nu$ (q -fold). Then

$$\nu^q(\mathcal{R}) \rightarrow 1, \quad n \rightarrow \infty.$$

PROOF. The first inequality of (5.12) is ensured by $t_n \rightarrow \infty$ and the bounded support of ν . For the second, note that

$$q^{2m}n^{-1} \sim q^{-1}K^{1/r}, \quad q^{-1} \sum_{k=1}^q g_k^{-1} \sim \int g^{-1} = d.$$

Hence the right hand side has expectation bounded by δP , $\delta < 1$ for n large enough, while its variance tends to zero as $n \rightarrow \infty$. \square

In the collection of models (5.11) the parameter is (h_1, \dots, h_q) ; call it now \underline{h} . Consider a loss for an estimate $\hat{\underline{h}}$

$$|\hat{\underline{h}} - \underline{h}|^2 := \sum_{k=1}^q \|\hat{h}_k - h_k\|^2 g_k^{-1}.$$

The arguments connected with (5.6) and (5.7) imply for the asymptotic minimax risk

$$(5.14) \quad \Delta \geq \liminf_n \inf_{\underline{h}} \sup_{\underline{h} \in \mathcal{R}} n^{-r} E_{\underline{h}} |\hat{\underline{h}} - \underline{h}|^2.$$

LEMMA 6. Let ν be a measure as in lemma 5. Then

$$n^{-r} \sup_{\underline{g} \in \mathcal{R}} \int_{\mathcal{R}^c} |\underline{g} - \underline{h}|^2 d\nu^q(\underline{h}) \rightarrow 0, \quad n \rightarrow \infty.$$

PROOF. For $\underline{g} \in \mathcal{R}$ we have

$$|\underline{g} - \underline{h}|^2 \leq 2|\underline{g}|^2 + 2|\underline{h}|^2,$$

$$n^{-r} |\underline{g}|^2 \leq (q^{2m} n^{r-1} \lambda_1)^{-1} P = O(1), \quad n \rightarrow \infty.$$

Hence it suffices to prove

$$\int_{\mathcal{R}^c} (1 + q^{-1} |\underline{h}|^2) d\nu^q(\underline{h}) \rightarrow 0, \quad n \rightarrow \infty.$$

This however follows immediately from $g_k^{-1} = O(1)$ and lemma 5.

PROOF OF THEOREM 1. Let $\tilde{\mu} > \mu$ be some number, where μ is from (2.6), and specify now K as

$$K^{-1} = (\sigma_*^2 d/P)^r s \tilde{\mu}.$$

We select the prior measure ν as a distribution on \mathbf{R}^s finite support, zero mean and diagonal covariance matrix M with diagonal elements $\sigma_*^2 \beta(j/s)$, $j = 1, \dots, s$, where the function β is from (2.4b). Let us demonstrate that the condition of lemma 5 is fulfilled if s is large enough. Indeed we have for $s \rightarrow \infty$, in view of the eigenvalue asymptotics (5.1)

$$\begin{aligned} \sum_{j=1}^s \lambda_j \sigma_*^2 \beta(j/s) &\sim \sigma_*^2 s^{2m+1} \int_0^\infty (\pi x)^{2m} \beta(x) dx \\ &= \sigma_*^2 s^{1/r} \int b(1-b) = \sigma_*^2 (s\mu)^{1/r}, \end{aligned}$$

where (2.6) has been used. On the other hand

$$P/K^{1/r}d = \sigma_*^2(s\tilde{\mu})^{1/r}$$

so that (5.13) is fulfilled.

Note that the right hand side of (5.14) is not changed if the infimum is taken only over estimators \hat{h} with values in \mathcal{R} , since \mathcal{R} is closed and convex. We then obtain from lemma 6

$$\Delta \geq \inf_{\hat{h}} n^{-r} \int E_{\underline{h}} |\hat{h} - \underline{h}|^2 d\nu^q(\underline{h}) - o(1), \quad n \rightarrow \infty.$$

The product structure of the model implies that the above Bayes risk is a sum of Bayes risks in the q submodels (5.11). We obtain

$$\Delta \geq (n^{-r} \sum_{k=1}^q g_k^{-1}) \min_{k \leq q} \inf_{\hat{h}} \int E_{h,k} \|\hat{h} - h\|^2 d\nu(h) + o(1)$$

where $E_{h,k}$ denotes expectation in the k -th model (5.11), for $h \in \mathbf{R}^s$. Take a sequence $k(n)$ where $\min_{k \leq q}$ is attained; invoke lemma 4 and theorem A1 in the appendix for obtaining

$$\begin{aligned} \Delta &\geq K d \sigma_*^2 \sum_{j=1}^s \beta(j/s) (1 + \beta(j/s))^{-1} \\ &\geq (\sigma_*^2 d)^{1-r} P^r \tilde{\mu}^{-1} s^{-1} \sum_{j=1}^s b(j/s). \end{aligned}$$

The proof of theorem 1 is now completed by letting $s \rightarrow \infty$, $\tilde{\mu} \rightarrow \mu$, $\sigma_*^2 \rightarrow \sigma^2$ and recalling $\gamma(m) = \mu^{-1} \int b$ (cp. (2.7)).

PROOF OF THEOREM 2. Let $a \in (0, 1)$, and consider the problem of estimating f from n observations (1.1) for a loss $\int_0^a (\hat{f} - f)^2$ and prior information $\int_0^a (D^m f)^2 \leq P$. Let Δ_a be the appropriate analog of (1.11). By a change of scale, a bound for Δ_a may be obtained from theorem 1 as follows. Define $F(x) = f(ax)$, $x \in (0, 1)$; then

$$\int_0^a (D^m f)^2 = a^{-2m+1} \|D^m F\|^2, \quad \int_0^a (\hat{f} - f)^2 = a \|\hat{F} - F\|^2.$$

The proof of theorem 1 shows that, for estimating F , observations outside $[0, 1]$ may be disregarded; hence the relevant observation number is $\tilde{n} \sim n \int_0^a g$. Note that for theorem 1 to be valid, the regression design need not satisfy (1.2) exactly but only the condition mentioned before lemma 1. Then the design density for estimating F is

$$\tilde{g}(x) = ag(ax) / \int_0^a g, \quad x \in [0, 1].$$

Now theorem 1 implies

$$(5.15) \quad \begin{aligned} \Delta_a &\geq \lim_n (n/\tilde{n})^{1-r} a \gamma(m) (\sigma^2 \int_0^1 \tilde{g}^{-1})^{1-r} (a^{2m-1} P)^r \\ &= \gamma(m) (\sigma^2 \int_0^a g^{-1})^{1-r} P^r. \end{aligned}$$

Let now a^{-1} be natural, $\{A\} = \mathbb{A}$ be a partition of $[0, 1]$ into intervals A of length a , $w_A = \inf_{x \in A} w(x)$ and P_A be positive numbers with $\sum P_A = P$. We have

$$\int_0^1 w(\hat{f} - f)^2 \geq \sum w_A \int_A (\hat{f} - f)^2.$$

Furthermore, to estimate Δ_w from below, we restrict f to the set of functions fulfilling $\int_A (D^m f)^2 \leq P_A$, all $A \in \mathbb{A}$. Analogously to (5.15) it can be shown that

$$\Delta_w \geq \gamma(m) \sum w_A (\sigma^2 \int_A g^{-1})^{1-r} P_A^r.$$

For $P_A = P d_A / \sum d_A$, $d_A = w_A^{1+1/2m} \int_A g^{-1}$ we obtain

$$\Delta_w \geq \gamma(m) (\sigma^2 \sum d_A)^{1-r} P^r.$$

For $a \rightarrow 0$ we have $\sum d_A \rightarrow d$.

5.4. The localized lower bounds. For the result 4.1, note that the set $W(q, s, P)$ defined in section 5.1 is contained in an L_2 -ball of radius $O(n^{-mr})$. Indeed for fixed s and $f \in W(q, s, P)$ we have in view of (5.4)

$$\|f\|^2 = \sum_{j=1}^s \sum_{k=1}^q (\varphi_{jkq}, f)^2 \leq \lambda_1^{-1} q^{-2m} \|D^m f\|^2 = O(n^{-2mr}).$$

For the bound (4.2), suppose first that the design density g is smooth, $w \equiv 1$ and use the prior of section 5.3, but with (5.13) valid as an equality. This prior in fact asymptotically concentrates on $\mathcal{B}_n(v)$ for $v = g^{-1}P/d$. (To deal with the supremum involved, use the methods for stochastic processes on $[0, 1]$.) The case of general v and g however requires a nonuniform scaling of the local basis functions φ_{jkq} in (5.2). Let g^* be the density proportional to $(gv)^r$, and J_{kq} , $k = 1, \dots, q$ be intervals such that

$$\int_{J_{kq}} g^* = q^{-1}, \quad k = 1, \dots, q.$$

Each φ_{jkq} in (5.2) is now scaled so that it has support J_{kq} . This allows a proof of the bound (4.2) with essentially the previous argument.

6. Appendix: A decision theoretic result. The Hájek-Le Cam bound which refers to the minimax risk in a weakly convergent sequence of experiments, cannot be utilized here. The reason is that one has to evaluate a proper Bayes risk rather than a minimax risk, in an asymptotically normal model (5.11). An appropriate argument has been given first by Efroimovich and Pinsker (1981). We propose a concise proof using abstract notions, within the framework of Le Cam's (1986) asymptotic decision theory. The facts we need are found in a particularly convenient form in Millar (1983), abbreviated (M) hereafter.

Suppose that for each ν from some index set \mathcal{N} , a sequence of experiments $\{E_{n,h,\nu}, h \in \mathbb{R}^s, \|h\| \leq t_n\}$ is given, where $t_n \rightarrow \infty$. Assume that, for some $\sigma^2 > 0$, all $c > 0$ and all $\nu \in \mathcal{N}$ the experiments $\{E_{n,h,\nu}, \|h\| \leq c\}$ converge weakly as $n \rightarrow \infty$ to a limit $\{E_{0,h}, \|h\| \leq c\}$, where $E_{0,h} = N(h, \sigma^2 I_s)$ is a normal measure on \mathbb{R}^s .

THEOREM A1. Let M be a symmetric positive definite $s \times s$ -matrix, and \mathcal{N} be the set of all probability measures on \mathbb{R}^s with finite support, zero mean and second moment matrix M . Then

$$(6.1) \quad \sup_{\nu \in \mathcal{N}} \liminf_n \inf_{\hat{h}} \int E_{n,h,\nu} \|\hat{h} - h\|^2 d\nu(h) \\ \geq \text{tr}[\sigma^2 M(\sigma^2 I_s + M)^{-1}]$$

(infimum over all measurable maps $\hat{h}: \mathbb{R}^s \rightarrow \mathbb{R}^s$).

PROOF. Define truncated loss functions, for $c > 0$

$$L_{c,h}(x) = \min(\|x - h\|^2, c), \quad x, h \in \mathbb{R}^s.$$

We shall consider generalized procedures \hat{h} as bilinear forms according to (M), (II.1.4). Then the risk of \hat{h} for the (bounded continuous) loss function $L_{c,h}$ and for distribution $E_{n,h,\nu}$ is written $\hat{h}(E_{n,h,\nu}, L_{c,h})$. For $\nu \in \mathcal{N}$ define the mixed risk

$$\varrho_n(\hat{h}, \nu, c) = \int \hat{h}(E_{n,h,\nu}, L_{c,h}) d\nu(h), \quad n = 0, 1, 2, \dots$$

Now observe that relation (III.1.7) of (M), obtained in the course of proving the asymptotic minimax theorem, implies that for any $\nu \in \mathcal{N}$, $c > 0$

$$\liminf_n \inf_{\hat{h}} \varrho_n(\hat{h}, \nu, c) \geq \inf_{\hat{h}} \varrho_0(\hat{h}, \nu, c).$$

The map $h \rightarrow E_{0,h}$ is continuous in total variation norm, while $h \rightarrow L_{c,h}$ is continuous in the sup norm over \mathbf{R}^s . Since \hat{h} is a continuous bilinear form with norm 1, it follows that the family of functions $h \rightarrow \hat{h}(E_{0,h}, L_{c,h})$ is equicontinuous (and bounded by c) when \hat{h} runs through the procedures. Now select a sequence $\{\nu_k\} \subset \mathcal{N}$ such that $\nu_k \rightarrow \nu_0 = N(O_s, M)$ weakly, e.g. on the basis of the central limit theorem. By the uniform Helly-Bray theorem (see Parzen (1954))

$$\varrho_0(\hat{h}, \nu_k, c) \rightarrow \varrho_0(\hat{h}, \nu_0, c), \quad k \rightarrow \infty$$

uniformly in \hat{h} . Here the right hand side is continuous in \hat{h} for the weak topology, since all $\varrho_0(\hat{h}, \nu_k, c)$ are. It follows that if z is the left hand side of (6.1) then

$$z \geq \inf_{\hat{h}} \varrho_0(\hat{h}, \nu_0, c).$$

To evaluate this infimum, one may restrict oneself to procedures of Markov kernel type, since these are dense in the set of procedures. Standard reasoning involving Anderson's lemma (section VI.2 of (M)) shows that the infimum is attained for an estimator \hat{h} which does not depend on c (the posterior expectation of h), since $L_{c,h}$ is a subconvex loss function. Letting $c \rightarrow \infty$, we obtain as a lower bound for z the Bayes risk in $\{N(h, \sigma^2 I_s), h \in \mathbf{R}^s\}$ for a normal prior $N(O_s, M)$ and squared error loss, which is $\text{tr}[\sigma^2 M(\sigma^2 I_s + M)^{-1}]$.

REFERENCES

- AGARWAL, G. G. and STUDDEN, W. J. (1980). Asymptotic integrated mean square error using least squares and bias minimizing splines. *Ann. Statist.* **8** 1307-1325.
- AGMON, S. (1968). Asymptotic formulas with remainder estimates for eigenvalues of elliptic operators. *Arch. Rational Mech. Anal.* **28** 165-183.
- BERAN, R. (1981). Efficient robust estimates in parametric models. *Z. Wahrsch. verw. Gebiete* **57** 91-108.
- BERAN, R. (1982). Robust estimation in models for independent nonidentically distributed data. *Ann. Statist.* **10** 415-428.
- BIRGÉ, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. verw. Gebiete* **65** 181-237.
- COX, D. D. (1984). Multivariate smoothing spline functions. *SIAM J. Numer. Anal.* **21** 789-813.
- DONOHOO, D. L. and LIU, R. C. (1988). Geometrizing rates of convergence III. Technical Report No. 138, Department of Statistics, Univ. Cal. Berkeley.
- DONOHOO, D. L., MACGIBBON, B., and LIU, R. C. (1988). Minimax risk for Hyperrectangles. Technical Report No. 123, Department of Statistics, Univ. Cal. Berkeley.
- EFROIMOVICH, S. Yu and PINSKER, M. S. (1981). Estimating a square integrable spectral density from a sequence of observations (in Russian). *Problemy Peredači Informacii* **17** (3) 50-68. (English translation in *Problems Inform. Transmission* (1982) 182-196.
- EFROIMOVICH, S. Yu and PINSKER, M. S. (1984). A learning algorithm for nonparametric filtering (in Russian). *Avtomatika i Telemekhanika* (11) 58-65.
- EFROIMOVICH, S. Yu (1985). Nonparametric estimation of a density of unknown smoothness. *Theory Probab. Appl.* **30** 557-568.

- GOLUBEV, G. K. (1984). Experimental design for nonparametric estimation of a regression function (in Russian). In: Analysis of Complex Information Systems, part 2, 58-61, Inst. Inform. Transmission, Moscow.
- GOLUBEV, G. K. (1987). Adaptive asymptotically minimax estimates of smooth signals (in Russian). Problemy Peredači Informacii **23** (1) 57-67.
- HALL, P. and MARRON, J. S. (1988). Choice of kernel order in density estimation. *Ann. Statist.* **16** 161-173.
- HÄRDLE, W. and MARRON, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.* **13** 1465-1481.
- IBRAGIMOV, I. A. and KHASMINSKI, R. Z. (1981). Statistical Estimation: Asymptotic Theory. Springer, New York.
- IBRAGIMOV, I. A. and KHASMINSKI, R. Z. (1982). Bounds for the risk of nonparametric regression estimates. *Theory Probab. Appl.* **27** 84-99.
- IBRAGIMOV, I. A. and KHASMINSKI, R. Z. (1984). On nonparametric estimation of the value of a linear functional in Gaussian white noise. *Theory Probab. Appl.* **29** 18-32.
- KAZBARAS, A. (1986). An adaptive kernel-type estimator for a square integrable distribution density (in Russian). Litovski Mat. Sbornik **26** 673-683. English translation in *Lithuanian Math. J.* **26** 318-324.
- LEVIT, B. Ya. (1975). On the efficiency of a class of nonparametric estimates. *Theor. Probab. Appl.* **20** 723-740.
- LE CAM, L. (1986). Asymptotic Methods in Statistical Decision Theory. Springer, New York.
- LI, K. C. (1985). From Stein's unbiased risk estimates to the method of generalized cross validation. *Ann. Statist.* **13** 1352-1377.
- LI, K. C. (1986). Asymptotic optimality of C_L and generalized cross-validation in ridge regression. *Ann. Statist.* **14** 1101-1112.
- MARRON, J. S. (1987). A comparison of cross-validation techniques in density estimation. *Ann. Statist.* **15** 152-162.

- MARRON, J. S. (1988). Automatic smoothing parameter selection: a survey. *Empirical Economics* **13** 187-208.
- MILLAR, P. W. (1983). The minimax principle in asymptotic statistical theory. In: Ecole d'Eté de Probabilités de Saint Four XI (P. Hennequin, Ed.). Lecture Notes in Mathematics 976 75-365, Springer, New York.
- MÜLLER, H. G. (1984). Optimal designs for nonparametric kernel regression. *Statist. Probab. Letters* **2** 285-290.
- MÜLLER, H. G. and STADTMÜLLER (1987). Variable bandwidth kernel estimators of regression curves. *Ann. Statist.* **15** 182-201.
- NADARAYA, E. A. (1974). On the integral mean squared error of some nonparametric estimates for the density function. *Theor. Probab. Appl.* **19** 133-141.
- NUSSBAUM, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L_2 . *Ann. Statist.* **13** 984-997.
- PARZEN, E. (1954). On uniform convergence of families of sequences of random variables. Univ. California Publ. Statist. **2** 23-53.
- PILZ, J. (1986). Minimax linear regression estimation with symmetric parameter restrictions. *J. Statist. Plann. Inference* **13** 297-318.
- PINSKER, M. S. (1980). Optimal filtering of square integrable signals in Gaussian white noise (in Russian). Problemy Peredači Informacii **16** (2) 52-68. English translation in Problems Inform. Transmission (1980) 120-133.
- RICE, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12** 1215-1230.
- RUDZKIS, R. (1985). On an estimate of the spectral density (in Russian). Litovski Mat. Sbornik **25** (3) 163-174. English translation in Lithuanian Math. J. **25** 273-280.
- SPECKMAN, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Statist.* **13** 970-983.
- SPRUILL, M. C. (1984). Optimal designs for minimax extrapolation. *J. Multivar. Anal.* **15** 52-62.

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

WOODROFE, M. (1970). On choosing a delta sequence. *Ann. Math. Statist* **41** 1665-1671.

INSTITUTE FOR PROBLEMS OF
INFORMATION TRANSMISSION
ACADEMY OF SCIENCES OF THE USSR
ERMOLOVOY STR. 19
110 051 MOSCOW GSP-4
USSR

KARL WEIERSTRASS INSTITUTE
OF MATHEMATICS
ACADEMY OF SCIENCES OF THE GDR
MOHRENSTR, 39
DDR - 1086 BERLIN
GERMAN DEMOCRATIC REPUBLIC

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 Hydrosiences, (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. Statistical Science, 1988, Vol.3 No. 2 239-271.
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, 579-611.
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. Scandinavian J. Statist., 1988, 15, 1-23.
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. Biometrika, 1987, 74, 799-808.
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. (Nov 1986, revised Aug 1988). 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. J. Amer. Statist. Assoc., 1988, 83, 744-749.
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. SIAM J. Appl. Math., June, 1989.
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. Statistics a Guide to the Unknown, pp. 249-260 (Eds. J.M. Tanur et al.) Wadsworth, Pacific Grove.
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-Červonenkis 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. (Dec. 1987, revised June 1989). 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. CANCELLED
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. (Jan. 1988, revised Jan 1990). 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. (Jan. 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. Ann. Stat., December, 1988.
143. DALANG, R.C. (Feb. 1988, revised Nov. 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. (Sept. 1989). On min-max majority and deepest points.
150. SEILLIER, F. (March 1988). Sequential probability forecasts and the probability integral transform.
151. NOLAN, D. (Mar. 1988, revised Feb. 1990). Asymptotics for multivariate trimming.
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. and GASKO, M. (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, revised March 1990). Comparing automatic 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. (Aug. 1988, revised Jan. 1990). Maximum likelihood an introduction.
169. BREIMAN, L. (Aug.1988, revised Feb. 1989). Submodel selection and evaluation in regression I. The X-fixed 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. ABRAHAMS, 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 subproblems.
179. DONOHO, D.L. and STARK, P.B. (October 1988). Recovery of sparse signal when the low frequency information is missing.
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, revised Jan. 1989, Aug. '89). Models for variable 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.
183. BERAN, R. (November 1988). Calibrating prediction regions.
184. BARLOW, M.T., PITMAN, J. and YOR, M. (Feb. 1989). On Walsh's Brownian Motions.
185. DALANG, R.C. and WALSH, J.B. (Dec. 1988). Almost-equivalence of the germ-field Markov property and the sharp Markov property of the Brownian sheet.
186. HESSE, C.H. (Dec. 1988). Level-Crossing of integrated Ornstein-Uhlenbeck processes
187. NEVEU, J. and PITMAN, J.W. (Feb. 1989). Renewal property of the extrema and tree property of the excursion of a one-dimensional brownian motion.
188. NEVEU, J. and PITMAN, J.W. (Feb. 1989). The branching process in a brownian excursion.
189. PITMAN, J.W. and YOR, M. (Mar. 1989). Some extensions of the arcsine law.
190. STARK, P.B. (Dec. 1988). Duality and discretization in linear inverse problems.
191. LEHMANN, E.L. and SCHOLZ, F.W. (Jan. 1989). Ancillarity.
192. PEMANTLE, R. (Feb. 1989). A time-dependent version of Pólya's urn.
193. PEMANTLE, R. (Feb. 1989). Nonconvergence to unstable points in urn models and stochastic approximations.
194. PEMANTLE, R. (Feb. 1989, revised May 1989). When are touchpoints limits for generalized Pólya urns.
195. PEMANTLE, R. (Feb. 1989). Random walk in a random environment and first-passage percolation on trees.
196. BARLOW, M., PITMAN, J. and YOR, M. (Feb. 1989). Une extension multidimensionnelle de la loi de l'arc sinus.
197. BREIMAN, L. and SPECTOR, P. (Mar. 1989). Submodel selection and evaluation in regression — the X-random case.
198. BREIMAN, L., TSUR, Y. and ZEMEL, A. (Mar. 1989). A simple estimation procedure for censored regression models with known error distribution.
199. BRILLINGER, D.R. (Mar. 1989). Two papers on bilinear systems: a) A study of second- and third-order spectral procedures and maximum likelihood identification of a bilinear system. b) Some statistical aspects of NMR spectroscopy, *Actas del 2º congreso latinoamericano de probabilidad y estadística matemática*, Caracas, 1985.
200. BRILLINGER, D.R. (Mar. 1989). Two papers on higher-order spectra: a) Parameter estimation for nonGaussian processes via second and third order spectra with an application to some endocrine data. b) Some history of the study of higher-order moments and spectra.
201. DE LA PENA, V. and KLASS, M.J. (April 1989). L bounds for quadratic forms of independent random variables.
202. FREEDMAN, D.A. and NAVIDI, W.C. (April 1989). Testing the independence of competing risks.

203. TERDIK, G. (May 1989). Bilinear state space realization for polynomial stochastic systems.
204. DONOHO, D.L. and JOHNSTONE, I.M. (May 1989). Minimax risk over l_p -Balls.
205. PEMANTLE, R., PROPP, J. and ULLMAN, D. (May 1989). On tensor powers of integer programs.
206. MILASEVIC, P. and NOLAN, D. (May 1989). Estimation on the sphere: A geometric approach.
207. SPEED, T.P. and YU, B. (July 1989, rev. Oct. 1989). Stochastic complexity and model selection: normal regression.
208. DUBINS, L.E. (June 1989). A group decision device: Its pareto-like optimality.
209. BREIMAN, L. (July 1989). Fitting additive models to regression data.
210. PEMANTLE, R. (July 1989). Vertex-reinforced random walk.
211. LE CAM, L. (August 1989). On measurability and convergence in distribution.
212. FELDMAN, R.E. (July 1989). Autoregressive processes and first-hit probabilities for randomized random walks.
213. DONOHO, D.L., JOHNSTONE, I.M., HOCH, J.C. and STERN, A.S. (August 1989). Maximum entropy and the nearly black object.
214. DONOHO, D.L. (Aug. 1989, revised Sept., Nov. 1989). Statistical estimation and optimal recovery.
215. STONE, C. (August, 1989). Asymptotics for doubly-flexible logspline response models.
216. NOLAN, D. (August 1989). The excess mass ellipsoid.
217. FREEDMAN, D.A. (August 1989). Statistical models and shoe leather.
218. BICKEL, P.J., NAIR, V.N. and WANG, P.C.C. (August 1989). Nonparametric inference under biased sampling from a finite population.
219. DALANG, R.C. and WALSH, J.B. (September 1989). The sharp Markov property of the brownian sheet and related processes.
220. BURMAN, P. and NOLAN, D. (Oct. 1989). Location-adaptive density estimation and nearest-neighbor distance.
221. LOW, M. (Oct. 1989). Non-existence of an adaptive estimator for the value of an unknown probability density.
222. LOW, M.G. (Oct. 1989). Invariance and rescaling of infinite dimensional Gaussian shift experiments.
223. LOW, M.G. (Oct. 1989). Lower bounds for the integrated risk in nonparametric density and regression estimation.
224. BREIMAN, L. and CUTLER, A. (Oct. 1989). A deterministic algorithm for global optimization.
225. LOW, M.G. (Oct. 1989). Local convergence of nonparametric density estimation problems to Gaussian shift experiments on a Hilbert space.
226. LEHMANN, E.L. (Oct. 1989). Model specification: The views of Fisher and Neyman, and later developments.
227. GOLUBEV, G.K. and NUSSBAUM, M. (Nov. 1989). A risk bound in Sobolev class regression.
228. STARK, P.B. (Dec. 1989). Rigorous computer solutions of infinite-dimensional inverse problems.
229. EVANS, S.N. and PERKINS, E. (December 1989). Measure-valued Markov branching processes conditioned on non-extinction.
230. EVANS, S.N. (December 1989). The entrance space of a measure-valued Markov branching process conditioned on non-extinction.
231. BREIMAN, L. (December 1989). The Π -method for estimating multivariate functions from noisy data.
232. SHAFFER, J.P. (December 1989). Probability of directional errors with disordinal (qualitative) interaction.
233. BRODSKY, M. and PANAKHOV, E. (January 1990). Concerning a priori estimates of solution of the inverse logarithmic potential problem.
234. SHAFFER, J.P. (January 1990). The Gauss-Markov theorem and random regressors.
235. SPLAWA-NEYMAN, J. (January 1990). On the application of probability theory to agricultural experiments. Essay on principles by Jerzy Splawa-Neyman. (Roczniki Nauk. Bol. Tom. X (1922)). Translation from the Polish original of pp.29-42 by D.M. Dabrowska, edited by T.P. Speed.

- 236. DONOHO, D.L. and NUSSBAUM, M. (January 1990). Minimax quadratic estimation of a quadratic functional.
- 237. DONOHO, D.L., JOHNSTONE, I.M., STERN, A.S. and HOCH, J.C. (January 1990). Does the maximum entropy method improve sensitivity?
- 238. KOOPERBERG, C. and STONE, C.J. (February 1990). A study of logspline density estimation.
- 239. HESSE, C.H. (March 1990). Hitting-time densities of a two-dimensional Markov process.
- 240. EVANS, S.N. (March 1990). Trapping a measure-valued Markov branching process conditioned on non-extinction.
- 241. YU, B. and SPEED, T.P. (March 1990). Stochastic complexity and model selection II. Histograms.
- 242. BICKEL, P.J. and MILLAR, P.W. (March 1990). Uniform convergence of probability measures on classes of functions.
- 243. DALANG, R.C. and WALSH, J.B. (April 1990). The sharp Markov property of certain random surfaces.
- 244. BRILLINGER, D.R. (March 1990). Two reports on the analysis of spatially aggregate data:
 - a) Mapping aggregate birth data
 - b) Spatial-temporal modelling of spatially aggregate birth data.

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.