Neo-Classical Minimax Problems, Thresholding, and Adaptation

David L. Donoho[1] Iain M. Johnstone[2]

Abstract

We study the problem of estimating θ from data $Y \sim N(\theta, \sigma^2)$ under squared-error loss. We define three new scalar minimax problems in which the risk is weighted by the size of θ. Simple thresholding gives asymptotically minimax estimates of all three problems. We indicate the relationships of the new problems to each other and to two other neo-classical problems: the problems of the bounded normal mean and of the risk-constrained normal mean.

Via the wavelet transform, these results have implications for adaptive function estimation, to: (1) estimating functions of unknown type and degree of smoothness in a global ℓ^2 norm; (2) estimating a function of unknown degree of local Hölder smoothness at a fixed point. In setting (2), the scalar minimax results imply: (a) that it is not possible to fully adapt to unknown degree of smoothness – adaptation imposes a performance cost; and (b) that simple thresholding of the empirical wavelet transform gives an estimate of a function at a fixed point which is, to within constants, optimally adaptive to unknown degree of smoothness.

Key words and phrases: Minimax Estimation, Adaptive Estimation, ℓ^p-balls, weak ℓ^p-balls.

AMS-MOS Subject Classifications: Primary 62G07, 62C20, Secondary 60G70, 41A25.

Acknowledgements: Presented at the Fifth Purdue Symposium on Statistical Decision Theory, June 20, 1992. Research partially supported by NSF DMS 92-09130.

---

[1] donoho@playfair.stanford.edu. [2] imj@playfair.stanford.edu.

1. Introduction

Suppose we have normally distributed scalar data $Y \sim N(\theta, \sigma^2)$ and we wish to estimate θ, measuring performance by squared-error loss $E(\hat{\theta}-\theta)^2$. The classical minimax theorem (Wolfowitz, 1950) says that, in the absence of prior information on θ, Y is optimal as an estimator of θ in a worst-case sense:

$$E\,(Y-\theta)^2 \quad = \quad \inf_{\hat{\theta}} \sup_{\theta \in \mathcal{R}} E\left(\hat{\theta}(Y)-\theta\right)^2 \quad . \tag{1.1}$$

This theorem, via Le Cam's theory of local asymptotic normality, lies at the heart of many developments in asymptotic minimaxity in parametric and nonparametric statistics.

Recently, we have entered a *neo-classical* period, where modifications of the classical minimax problem are studied, with applications to determining the precise constants in the minimax risk of various curve estimation problems. We mention two specific examples. In the first, the problem of *estimating a bounded normal mean*, one assumes that θ is known to lie in a finite interval $[-\tau, \tau]$. The study and evaluation of the minimax risk

$$\inf_{\hat{\theta}} \sup_{\theta \in [-\tau,\tau]} E\left(\hat{\theta}(Y)-\theta\right)^2 \tag{1.2}$$

was initiated and solved by Bickel (1981), Casella and Strawderman (1981) and Levit (1981). By the method of hardest one-dimensional subproblems and hardest Cartesian subproblems, this problem has been found to lie at the heart of several important asymptotic minimaxity and near-minimaxity results in nonparametric statistics: for example in minimax estimation of a linear functional (see e.g. Ibragimov and Has'minskii (1984), and Donoho and Liu (1991)), and in minimax estimation of the whole object (see Donoho, Liu, and MacGibbon(1990)).

In the second example, one searches for a minimax estimator *subject to constraints on the risk at a fixed point*. Let $\hat{\Theta}_0(\rho)$ denote the class of estimators with risk $\leq \rho$ at the origin:

$$\hat{\Theta}_0(\rho) \quad = \quad \{\hat{\theta} : E_0\hat{\theta}(Y)^2 \ \leq \ \rho\}, \tag{1.3}$$

and consider

$$K^0(\rho) = \inf_{\hat{\theta} \in \hat{\Theta}_0(\rho)} \sup_{\theta \in \mathcal{R}} E_\theta \left(\hat{\theta}(Y)-\theta\right)^2 \quad .$$

The study of this problem was initiated by Bickel (1983). Results in this area apply to problems of estimating sparse signals in Gaussian noise (Donoho, Johnstone, Hoch, and Stern, 1991), (Johnstone, 1992).

In this paper we will introduce three (apparently) new neo-classical minimax problems, derive asymptotically minimax rules, discuss their relations to the other neo-classical problems above,

and give applications to infinite-dimensional estimation problems such as adaptive estimation of objects of unknown smoothness.

1.1 Three Problems

The first problem, in the scalar case $Y \sim N(\theta, 1)$, is to obtain the minimax value

$$L^p(\delta) \quad = \quad \inf_{\hat{\theta}} \sup_{\theta \in \mathcal{R}} \quad \frac{E(\hat{\theta}(Y) - \theta)^2}{\delta + |\theta|^p} \tag{1.5}$$

where $\delta > 0$ and $p \in (0, 2)$.

THEOREM 1.

$$L^p(\delta) \sim \left(2 \log(\delta^{-1})\right)^{1-p/2}, \qquad \delta \to 0 . \tag{1.6}$$

An asymptotically minimax rule is the soft threshold rule

$$\hat{\theta}^*(Y) \quad = \quad \eta_t(Y) \quad \equiv \quad (|Y| - t)_+ \ sgn(Y),$$

with threshold

$$t \quad = \quad t(\delta) \quad = \quad \sqrt{2 \log(\delta^{-1})} . \tag{1.7}$$

The estimator $\eta_t(Y)$ is a simple nonlinear shrinker (also called limited-translation or Efron-Morris in other contexts).

The following strengthening of the L^p problem is of particular importance to us and provides the second new situation. Let $\tau_\delta = \sqrt{2 \log(\delta^{-1})}$ and let $m_\delta^p(\theta) = \tau_\delta^p \min((\theta/\tau_\delta)^2, 1)$. Define

$$M^p(\delta) \quad = \quad \inf_{\hat{\theta}} \sup_{\theta \in \mathcal{R}} \quad \frac{E(\hat{\theta}(Y) - \theta)^2}{\delta + m_\delta^p(\theta)}.$$

For θ near τ_δ, $m_\delta^p(\theta) \approx |\theta|^p$. However, away from τ_δ, $m_\delta^p(\theta)$ behaves differently. Since $m_\delta^p(\theta) \le |\theta|^p$, $M^p(\delta) \ge L^p(\delta)$.

THEOREM 2.

$$M^p(\delta) \ \sim \ \left(2 \log(\delta^{-1})\right)^{1-p/2}, \qquad \delta \to 0 . \tag{1.8}$$

An asymptotically minimax procedure is of the form $\hat{\theta}^*(Y) = \eta_t(Y)$ *with* $t = \sqrt{2 \log(\delta^{-1})}$.

The third new problem is the study of

$$K^p(\rho) \quad = \quad \inf_{\hat{\theta} \in \hat{\Theta}_0(\rho)} \quad \sup_{\tau \ge 1} \quad \tau^{-p} \quad \sup_{|\theta| \le \tau} \quad E\left(\hat{\theta}(Y) - \theta\right)^2 \tag{1.9}$$

where $\hat{\Theta}_0(\rho)$ is as in (1.3). Evidently, this is a mixture between the "subject to doing well at a point" and "Bounded Normal Mean" problems, with the additional twist of the "$\sup_{\tau \ge 1} \ \tau^{-p}$" weighting thrown in!

THEOREM 3.

$$K^p(\rho) \sim \left(2 \log(\rho^{-1})\right)^{1-p/2}, \qquad \rho \to 0 . \tag{1.10}$$

An asymptotically minimax procedure is of the form $\hat{\theta}^*(Y) = \eta_t(Y)$ *with* $t = \sqrt{2 \log(\rho^{-1})}$.

The proof of these Theorems, in sections 2-5 below, shows that these three results are closely connected. In a certain sense, K^p is smaller than L^p, which is smaller than M^p, so lower bounds on $K^p(\delta)$ and upper bounds on $M^p(\delta)$, both of size $(2 \log(\delta^{-1}))^{1-p/2} (1 + o(1))$, combine to prove both theorems simultaneously.

1.2 Two Phenomena

The above results expose two phenomena:

[UNI] If we calibrate δ and ρ appropriately, *a single estimator* $\hat{\theta}^*$ *is asymptotically minimax for all three problems, and the form of the estimator does not depend on* p. There is a single 'universal' kind of estimator for all these problems.

[LOG] *The minimax values* $K^p(\delta)$, $L^p(\delta)$ *and* $M^p(\delta)$ *are all asymptotically equivalent, and they behave as* $\left(2 \log(\delta^{-1})\right)^{1-p/2}$, *as* $\delta \to 0$.

These phenomena may at first appear to concern only problems of estimating a 1-dimensional parameter. Sections 6-8 of this paper will show that they cause similar phenomena for estimation of infinite-dimensional parameters.

1.3 Sequence Space

The significance of these scalar minimax problems comes, as usual, from applying the scalar results coordinatewise to multivariate problems. Section 6 below develops applications to two such problems. In both, we suppose that we have n noisy observations $y_i = \theta_i + \epsilon_n z_i$, $i = 1, ..., n$, with $z_i \overset{i.i.d.}{\sim} N(0,1)$, and that our goal is to estimate θ with small squared ℓ^2 risk: $E \left|\left|\hat{\theta}(y) - \theta\right|\right|^2_{\ell^2_n}$.

In the first, we suppose the parameter θ belongs to an dimensional ℓ^p ball, defined by

$$\ell^p_n(C) = \left\{ \theta \in \mathcal{R}^n : \sum_{i=1}^n \left|\theta_i\right|^p \le C^p \right\}.$$

Consider the minimax problem

$$L^p_n(C_n, \epsilon_n) = \inf_{\hat{\theta}} \sup_{\ell^p_n(C_n)} E \left|\left|\hat{\theta}(y) - \theta\right|\right|^2_{\ell^2_n}; \tag{1.12}$$

and let $\eta_n = \eta^{-1/p} C_n / \epsilon_n$ denote the normalized radius; we assume that $\eta_n \to 0$ as $n \to \infty$. For a full discussion, see Donoho and Johnstone (1994). Let $\gamma_n \to 0$ be defined by $\gamma_n^{-1} = \log \eta_n^{-p}$,

and set $\delta_n = \gamma_n \eta_n^p$ and consider the the estimator $\hat{\theta}_n^* = (\epsilon_n \hat{\theta}^*(y_i/\epsilon_n))_i$, built up coordinatewise out of the estimator which is asymptotically minimax for $L^p(1/n)$, $0 < p < 2$.

THEOREM 4. *Let $0 < p < 2$ and assume that $\eta_n \to 0$ and $(\epsilon_n/C_n)^2 \log n\,(\epsilon_n/C_n)^p \to 0$. Then $\hat{\theta}_n^*$ is an asymptotically minimax procedure, and*

$$L_n^p\,(C_n, \epsilon_n) \;\sim\; L^p\,(\eta_n^p)\, C_n^p\, \epsilon_n^{2-p}\;, \qquad n \to \infty\;.$$

In the second problem we suppose again that we have observations y_i, but now the parameter θ belongs to an n-dimensional *weak-ℓ^p ball*, (Marcinkiewicz-ball) defined as follows. For a vector θ let $|\theta|_{(1)} \geq |\theta|_{(2)} \geq \ldots$ denote the ordered coordinate values, and put

$$\mathbf{m}_n^p\,(C) \;\;=\;\; \left\{\theta \in \mathcal{R}^n : |\theta|_{(i)} \leq C \cdot i^{-1/p} \quad \forall i\right\}.$$

(Note that $\ell_n^p(C) \subset \mathbf{m}_n^p(C)$). Consider the minimax problem

$$M_n^p(C, \epsilon_n) = \inf_{\hat{\theta}} \sup_{\mathbf{m}_n^p(C)} E\left|\left|\hat{\theta}(y) - \theta\right|\right|_{\ell_n^2}^2; \tag{1.13}$$

Consider again the estimator $\hat{\theta}_n^* = (\epsilon_n \theta^*(y_i/\epsilon_n))_i$, built coordinatewise out of estimators with $\delta_n = \gamma_n \eta_n^p$, which are asymptotically minimax for $M^p(\delta_n)$, $0 < p < 2$.

THEOREM 5. *Let $0 < p < 2$ and assume that $\eta_n \to 0$ and $\epsilon_n^2 C_n^{-2} \log n\, \epsilon_n^p C_n^{-p} = 0\left((\log n)^{-6/p}\right)$. Then $\hat{\theta}_n^*$ is an asymptotically minimax procedure, and*

$$M_n^p(C_n, \epsilon_n) \;\sim\; \frac{2}{2-p} \cdot M^p(\eta_n) \cdot C_n^p \cdot \epsilon_n^{2-p}, \qquad n \to \infty\;.$$

In Theorems 4 and 5, the threshold level of the asymptotic minimax estimator depends on the radius, noise level and shape of ball through η_n. In the discussion in Section 1.4 of estimation over certain function spaces, it is natural to fix $C_n = C$, and to adopt the calibration $\epsilon_n = \sigma n^{-1/2}$. In this case, a *single* choice of threshold, based on $\delta = 1/n$, comes within a constant factor of asymptotic minimaxity, irrespective of the radius or type of ball:

COROLLARY 5. Let $0 < p < 2$, $\epsilon_n = \sigma n^{-1/2}$ and $\delta = 1/n$. Consider the estimator $\hat{\theta}_n^* = (\epsilon_n\, \eta_{tn}(y_i/\epsilon_n))_i$. Then

$$\sup_{L_n^p(C)} E\left|\left|\hat{\theta}_n^* - \theta\right|\right|_{\ell_n^2}^2 \;\leq\; \left(\frac{2}{2-p}\right)^{1-p/2} L_r^p(C, \sigma n^{-1/2})\,(1 + o(1))$$

and a corresponding result holds with $\mathbf{m}_n^p$ and M_n^p replacing ℓ_n^p and L_n^p.

This is an instance of phenomenon [**UNI**].

1.4 Function Space

It is now well known that orthogonal transformations can be employed to turn statements about estimation over bodies in sequence space into statements about estimation over classes of smooth functions in noisy data. Efroimovich and Pinsker (1981, 1982) and Nussbaum(1985) established this point with reference to the Fourier transform and ellipsoids; here we employ instead the wavelet transform, and the ℓ^p and $\mathbf{m}^p$ balls, which lead to different applications. Background on the use of the wavelet transform in this way can be found in De Vore and Lucier (1992), Donoho (1992, 1993a,b), Donoho, Johnstone, Kerkyacharian, and Picard (1993), Johnstone(1992), Johnstone, Kerkyacharian and Picard (1992), and Kerkyacharian and Picard (1992).

1.4.1 Adapting to Unknown Type and Degree of Smoothness

Suppose we have nonparametric regression observations

$$d_i = f(t_i) + \sigma z_i, \qquad i = 1, \ldots, n, \tag{1.14}$$

where the t_i are equispaced on $[-1/2, 1/2]$ and the z_i , are i.i.d. $N(0,1)$. Here is a method for estimation of the whole object $(f(t_i), i = 1, \ldots, n)$, based on application of $\hat{\theta}^*$ coordinatewise in a sequence space. Given $n = 2^{j_1+1}$ numbers (d_i), (j_1 an integer), take a discrete wavelet transform (section 7 below), giving empirical wavelet coefficients (y_i) which we treat as if they have standard deviation $\epsilon_n = \sigma/\sqrt{n}$. Then apply the estimator $\hat{\theta}^*_n$ to these coefficients and let $(\hat{f}^*_n(t_i))_i$ denote the inverse wavelet transform of the vector $\hat{\theta}^*_n$.

This approach reduces problems of estimating the smooth function f to problems of estimating the sequence of wavelet coefficients. If we are working with an orthogonal empirical wavelet transform we have an isometry of squared errors:

$$\frac{1}{n}\sum_i (\hat{f}^*_n(t_i) - f(t_i))^2 = \sum_i (\hat{\theta}^*_{n,i} - \theta_i)^2, \tag{1.15}$$

In such a case, we also have that the empirical coefficients y_i satisfy

$$y_i = \theta_i + \epsilon_n z_i, i = 1, \ldots, n$$

with z_i i.i.d. $N(0,1)$. (If we are working instead with a wavelet transform which is simply quasi-orthogonal, then slightly weaker relations hold; for example, the transform is a quasi-isometry, where the two sides in (1.15) lie within fixed multiples of each other, independently of $n = 2^{j_1+1}$.)

The empirical wavelet transform turns minimax estimation over classes of smooth functions into minimax estimation over the coefficient bodies induced by the transformation of functions in the class. For example, suppose that $W^p_m(C)$ is a ball of functions having W^p_m Sobolev norm $\leq C$,

for some fixed m and p. Let Θ_n denote the class of wavelet coefficient sequences $(\theta_i(f))$ arising from functions $f \in W_m^p(C)$. Then

$$\inf_{\hat{f}} \sup_{W_m^p(C)} E n^{-1} \|\hat{f} - f\|_{\ell_n^2}^2 \quad = \quad \inf_{\hat{\theta}} \sup_{\Theta_n} E\|\hat{\theta} - \theta\|^2.$$

Hence problems of estimating smooth functions reduce to problems of minimax estimation over bodies Θ_n in sequence-space.

A remarkable fact about the wavelet transform and traditional smoothness spaces is that the bodies Θ_n are nearly $\mathbf{m}_n^p$ balls, for a certain p. By applying Donoho (1992), one can show that with $W_m^p(C)$ a W_m^p-ball, the corresponding body Θ_n obeys, with a constant A depending only on the wavelet transform,

$$\Theta_n \subset \mathbf{m}_n^p(A \cdot C), \qquad n > n_0, \tag{1.16}$$

with $p = 2/(2m+1)$. Consequently, we have from Theorems 5, 2 and Corollary 5 the following bound for the worst-case risk of the estimator θ^*:

$$\sup_{W_m^p(C)} E n^{-1} \|\hat{f}_n^* - f\|_{\ell_n^2}^2 \;\le\; \left(\frac{2}{(2-p)}\right)^{1+r} A^p \cdot \log(n)^r (\sigma/\sqrt{n})^{2r} C^{2(1-r)} \cdot (1 + o(1))$$

where $r = 1 - p/2 = 2m/(2m+1)$. It is known from a variety of simple lower bound arguments that

$$\inf_{\hat{f}} \sup_{W_m^p(C)} E n^{-1} \|\hat{f} - f\|_{\ell_n^2}^2 \ge (\sigma/\sqrt{n})^{2r} C^{2(1-r)}$$

and so the estimator $\hat{f}_n^*$ is within logarithmic factors of optimal. We can interpret this geometrically as saying that while the bodies Θ_n do not quite fill up the weak ℓ^p balls, there is not much gap; the near-minimaxity of $\hat{\theta}^*$ over the inscribed bodies implies a near-minimaxity over the inscribing bodies.

This near-minimaxity is for a single estimator, defined without regard for p or C, and valid for a range of m. The phenomenon **[UNI]**, which implies the universality of $\hat{\theta}^*$ as an asymptotically minimax estimator for weak-ℓ^p balls of arbitrary radius and arbitrary $p \in (0,2)$, therefore means that there is a kind of universal near-minimax estimator for estimating functions of unknown type and degree of smoothness. In fact this holds much more generally than for Sobolev balls; it holds for balls in the whole range of Besov and Triebel-Lizorkin spaces, which includes Hölder classes; and it holds also for balls of functions of total variation. Compare also Donoho(1993b) and Donoho, Johnstone, Kerkyacharian, and Picard (1993).

1.4.2 Spatial Adaptation

As an alternative to point-samples, we might instead assume area samples $d_i = Ave\{f(t) : t \in [t_{i-1}, t_i]\} + \sigma z_i$, $i = 1, \ldots, n$, and choose, as our goal, the de-noising of the area-samples, yielding

risk

$$R_n(\hat{f}, f) = n^{-1} \sum_i (Ave\{\hat{f}|[t_{i-1}, t_i]\} - Ave\{f|[t_{i-1}, t_i]\})^2$$

We can apply the empirical wavelet transform to these samples and de-noise them according to the thresholding recipe given above. This has the following application. R. A. DeVore and collaborators, in a series of papers, have proposed the use of certain special Approximation Spaces A_p^m to model the process of spatial adaptation. These spaces are defined as collections of functions with the property that spatially-adaptive methods, such as splines with n optimally-chosen free knots, or best rational approximation of degree n, give error which decays like n^{-m}, $p = 2/(2m+1)$. More precisely, if $e_n(f)$ denotes the L^2 error of best approximation by such a spatially adaptive scheme, let $A_p^m(C)$ denote the ball of those functions where $\sum_n (e_n n^m)^p n^{-1}$ is less than or equal to C^p. Roughly speaking, the members in such a ball are all those which are equally easy to approximate using spatially adaptive methods.

A remarkable fact about these balls from approximation-space theory is that, although they are non-convex and seemingly rather exotic, they are equivalent to known objects in the space of theoretical wavelet coefficient sequences, and those balls are ℓ^p balls. This is a result of DeVore and Popov (1988) and DeVore, Jawerth, and Popov (1990). By using this fact, and arguments in Donoho (1993a), one can show that the empirical wavelet coefficients obey, with constants A_i depending on the wavelet transform alone

$$\ell_n^p(A_0 C) \subset \Theta_n \subset \ell_n^p(A_1 C), \qquad n > n_0. \tag{1.17}$$

Now the estimator $\hat{\theta}_n^*$ is asymptotically near-minimax for $\ell_n^p(A_i C)$, $i = 0, 1$; so we conclude that $\hat{f}_n^*$ is nearly minimax for $A_p^m(C)$. In fact the minimax risk differs from the worst case risk of A_p^m at most by the factor $[2/(2-p)]^r (A/A_0)^p$.

As wavelet shrinkage achieves the optimal rate performance over this class of functions - a class defined by spatially adaptive methods - one can say that wavelet shrinkage is a kind of optimally spatially adaptive method.

1.5 Adaptation at a Point

In Section 7 below, we trace two implications of phenomenon **[LOG]**. First, that in attempting to estimate a function at a point it is not possible to estimate as well when the degree of smoothness is unknown as one could estimate if the degree of smoothness were known; and second, that one can adapt as well as it is possible to do by a simple estimator - a coordinatewise application, in the wavelet transform domain, of the soft threshold estimator $\hat{\theta}_n^*$.

1.5.1 Impossibility of Adaptation "for free"

Suppose that we again have observations (1.14), and we are interested in estimating the single value $f(0)$. We suspect that f obeys a Hölder (-Zygmund) smoothness condition $f \in \Lambda(\alpha, C)$, where

$$\Lambda(\alpha, C) = \{f : |f^{(m)}(s) - f^{(m)}(t)| \le C|s-t|^{\delta}\},$$

with $m = \lceil \alpha \rceil - 1$ and $\delta = \alpha - m$. However, we are not sure of α and C.

If we did know α and C, then we could construct a linear minimax estimator $\hat{f}_n^{(\alpha,C)} = \sum_i c_i y_i$ where the (c_i) are the solution of a quadratic programming problem depending on C, α, σ, and n (compare Donoho (1994)). This estimator has worst-case risk

$$\sup_{\Lambda(\alpha,C)} E(\hat{f}_n^{(\alpha,C)} - f(0))^2 \sim A(\alpha)(C^2)^{1-r}(\frac{\sigma^2}{n})^r, \qquad n \to \infty,$$

where $A(\alpha)$ is the value of a certain optimization problem, and the rate exponent satisfies

$$r = \frac{2\alpha}{2\alpha + 1} \tag{1.17}$$

(Donoho and Low, 1992). This behavior is optimal among linear procedures and within a factor $5/4$ of the minimax risk over all measurable procedures.

Unfortunately, if α and C are actually unknown and we misspecify the degree α of the Hölder condition, the resulting estimator will achieve a worse rate of convergence than the rate which would be optimal for a correctly specified condition.

Can we develop an estimator which does not require knowledge of α and C and yet performs essentially as well as $\hat{f}_n^{(\alpha,C)}$? Lepskii (1991) and Brown and Low (1992) show that the answer is no, even if we know that the correct Hölder class is one of two specific classes. Hence for $0 < \alpha_0 < \alpha_1 < \infty$ and $0 < C_0, C_1 < \infty$,

$$\inf_{\hat{f}_n} \max_{i=0,1} C_i^{2(r_i-1)} n^{r_i} \sigma^{-2r_i} \sup_{\Lambda(\alpha_i, C_i)} E(\hat{f}_n - f(0))^2 \ge const \cdot \log(n)^{r_0}.$$

In short, we must gain an increase in risk in estimating a smooth function of unknown degree of smoothness; when the risk attainable in estimating at a point (smoothness known) is n^{-r}, the risk one must pay with smoothness unknown is at least $(\log(n)/n)^r$.

Section 7.1 below shows that this phenomenon can be traced to the asymptotics (1.9) for $K^p(\rho)$. More specifically, we have the lower bound

THEOREM 6. *Let $p_i = 2(1 - r_i)$. Then with explicitly computable constants A_i,*

$$\inf_{\hat{f}_n} \max_{i=0,1} C_i^{2(r_i-1)} n^{r_i} \sigma^{-2r_i} \sup_{\Lambda(\alpha_i, C_i)} E(\hat{f}_n - f(0))^2 \ge A_0 \cdot K^{p_0}(A_1/n), \tag{1.18}$$

The lower bound is based on a special 1-dimensional subfamily argument. Phenomenon **[LOG]** shows that this lower bound is (ignoring constants) best possible among 1-dimensional subfamily arguments.

1.5.2 Adaptation with minimal cost.

Section 7.2 studies the estimator $(\hat{f}_n^*(t_i))$ induced by soft thresholding in the wavelet domain. The estimator $\hat{f}_n^*(0)$ makes no assumptions about the smoothness or lack of smoothness of f.

THEOREM 7. *Suppose we use a wavelet transform with wavelets having $D > 1$ vanishing moments. For each Hölder class $\Lambda(\alpha, C)$ with $0 < \alpha < D$, we have*

$$\sup_{\Lambda(\alpha,C)} E(\hat{f}_n^*(0) - f(0))^2 \le M^p(1/n) \cdot A(\alpha) \cdot (C^2)^{1-r} \cdot (\frac{\sigma^2}{n})^r \cdot (1 + o(1)) \ . \tag{1.19}$$

Here $r = 2\alpha/(2\alpha + 1)$ is as in (1.17), and with $\delta(\alpha) = \min(\alpha, 1/2)$,

$$A(\alpha) = 2^r \cdot (C_5^2)^{1-r} \cdot (2C_3C_4)^2 \cdot \left(\frac{2}{1 - 2^{-\delta}}\right)^2 , \tag{1.20}$$

the C_i being constants associated to the Wavelet Transform (see [W1]-[W5] below).

Hence $\hat{f}_n^*(0)$ achieves, within a logarithmic factor, the minimax risk for every Hölder class in a broad range.

The logarithmic factor cannot be further reduced, because of Theorem 6. We have closed the circle: Because the lower bound of Theorem 6 derives from the K^p-problem, and the upper bound of Theorem 7 derives from the M^p-problem, the fact that the K^p problem and the M^p-problem have identical asymptotics (i.e. **[LOG]**), means that no estimator can improve on this one except perhaps at the level of constants.

The constructions of Lepskii (1991) and Efromovich and Low (1992) show that there exist estimators attaining this level of performance over restricted collections of smoothness classes; but the present estimator is simpler in construction and application, and the results seem stronger. The present estimator is also, as we have seen above, near-minimax for estimation in global risk over a wide range of smoothness assumptions.

2. Upper Bound on $M^p(\delta)$.

Let $t = \sqrt{2 \log(\delta^{-1})}$. We will argue that for this specific choice of t, η_t has risk $R_\delta(\theta) = E\left(\eta_t(Y) - \theta\right)^2$ satisfying

$$M_p^*(\delta) \ = \ \sup_\theta \ \frac{R_\delta(\theta)}{\delta + m_\delta^p(\theta)} \ \le \ \left(2 \log(\delta^{-1})\right)^{1-p/2} \ (1 + o(1)) \ . \tag{2.1}$$

This furnishes the upper bounds $L^p(\delta) \le M^p(\delta) \le \left(2 \log(\delta^{-1})\right)^{1-p/2} \ (1 + o(1)) \ .$

LEMMA 1.

$$R_\delta(\theta) \;\le\; R_\delta(0) + \theta^2 \tag{2.2}$$

$$R_\delta(\theta) \;\le\; t^2 + 1 \tag{2.3}$$

$$R_\delta(0) \;\le\; \phi(t)\left[\frac{4}{t^3} + \frac{6}{t^5}\right] \;. \tag{2.4}$$

PROOF. All three relations derive from the identity

$$\begin{aligned} R_\delta(\theta) \;&=\; 1 + t^2 + (\theta^2 - t^2 - 1)\;[\Phi(t-\theta) - \Phi(-t-\theta)] \\ &\quad -\; (t-\theta)\,\phi(t+\theta) - (t+\theta)\,\phi(t-\theta)\;. \end{aligned} \tag{2.5}$$

For details, see Lemma 1 of Donoho and Johnstone (1992b). ■

We argue separately on the ranges $[0,t)$ and $[t,\infty)$. On $[t,\infty)$, $m_\delta^p(\theta) = t^p$ and we use bound (2.3):

$$\frac{R(\theta)}{\delta + t^p} \;\le\; \frac{t^2+1}{\delta + t^p} \;\le\; t^{2-p} + t^{-p} \sim t^{2-p}\;.$$

On $[0,t]$, we have $m_\delta^p(\theta) = t^{p-2}\theta^2$, and from bound (2.2)

$$\frac{R(\theta)}{\delta + t^{p-2}\theta^2} \;\le\; t^{2-p}\left(\frac{R(0)+\theta^2}{t^{2-p}\delta + \theta^2}\right)\;. \tag{2.6}$$

¿From bound (2.4) and the identity $\phi(t) = \delta\phi(0)$, we notice that

$$R(0) \;\le\; \delta\phi(0)t^{-3}(4 + 6t^{-2}) \;\le\; t^{2-p}\delta$$

for all δ sufficiently small.

Thus (2.6) is bounded by t^{2-p}, and so combining results from the two subintervals gives the following strengthening of bound (2.1), valid for δ sufficiently small:

$$M_p^*(\delta) \;\le\; t^{2-p}(1 + t^{-2})\;. \tag{2.7}$$

3. Relation Between $K^p(\delta)$ and $L^p(\delta)$.

Make temporarily the calibration $\rho = L^p(\delta)\cdot\delta$. An estimator $\hat\theta$ attaining $L^p(\delta)$ has

$$R(\hat\theta,\,0) \;\le\; L^p(\delta)\,(\delta + 0^p) \;=\; \rho\;.$$

Hence it satisfies $\hat{\theta} \in \hat{\Theta}_0(\rho)$, and is eligible for competition the problem associated with $K^p(\rho)$. In that problem its risk satisfies

$$\begin{aligned} \sup_{\tau \geq 1} \tau^{-p} \sup_{|\theta| \leq \tau} R(\hat{\theta}, \theta) &\leq \sup_{\tau \geq 1} \tau^{-p} L^p(\delta) \ (\delta + \tau^p) \\ &= L^p(\delta) \cdot \sup_{\tau \geq 1} \left(\delta \tau^{-p} + 1 \right) \\ &= L^p(\delta) \cdot (\delta + 1) \ . \end{aligned}$$

In short

$$K^p(\rho) \leq L^p(\delta) \ (1 + \delta),$$

so L^p is "larger" than K^p.

4. Lower Bound on $K^p(\rho)$.

In this section we will establish the lower bound

$$K^p(\rho) \geq \left(2 \ \log(\rho^{-1})\right)^{1-p/2} \ \left(1 + o(1)\right), \qquad \rho \to 0 \ . \tag{4.1}$$

Together with the results of the last two sections, this will establish (1.6) and (1.9).

Our approach will be by an argument on Bayes Risks which shows that a certain 3-point prior is asymptotically least favorable. Let $\nu_{\varepsilon,\mu}$ denote the 3-point symmetric prior distribution

$$\nu_{\varepsilon,\mu} = (1 - \varepsilon) \ \nu_0 + \frac{\varepsilon}{2} \ \nu_\mu + \frac{\varepsilon}{2} \ \nu_{-\mu} \tag{4.2}$$

where ν_x denotes Dirac mass at x. This family of priors has been used extensively in the study of $K^0(\rho)$, by Bickel (1983), Donoho and Johnstone (1992a, 1994).

We make a particular choice of ε and μ as follows. Let $\varepsilon > 0$ be small and let $a > 0$. Define $\mu(\varepsilon, a)$ as the solution to

$$\mu^2 + 2a\mu = -2 \ \log\left(\frac{\varepsilon/2}{1 - \varepsilon}\right) \ . \tag{4.3}$$

and define $a(\varepsilon)$ as the solution to

$$\sqrt{\pi} \ \mu \ \phi(a + \mu) = \varepsilon \ , \qquad \mu = \mu(\varepsilon, a) \ . \tag{4.4}$$

Then set $\tilde{\mu}(\varepsilon) \equiv \mu(\varepsilon, a \ (\varepsilon))$ and

$$\pi_\varepsilon = \nu_{\varepsilon,\tilde{\mu}(\varepsilon)} \ . \tag{4.5}$$

In the appendix we prove the following Lemma.

LEMMA 2.

$$\tilde{\mu}(\varepsilon) \sim \sqrt{2\ \log(\varepsilon^{-1})}, \qquad \varepsilon \to 0\ . \tag{4.6}$$

Let $B(\pi_\varepsilon)$ *denote the Bayes risk of the prior* π_ε *in the problem* $Y \sim N(\theta, 1)$, $\theta \sim \pi_\varepsilon$. *Then*

$$B(\pi_\varepsilon) \sim \varepsilon\tilde{\mu}\ (\varepsilon)^2 \qquad \varepsilon \to 0\ . \tag{4.7}$$

Let $\hat{\theta}_\varepsilon$ *denote the Bayes rule for prior* π_ε .

$$R\left(\hat{\theta}_\varepsilon,\ \tilde{\mu}(\varepsilon)\right) \sim \tilde{\mu}(\varepsilon)^2 \qquad \varepsilon \to 0 \tag{4.8}$$

$$R\left(\hat{\theta}_\varepsilon,\ 0\right) \sim \varepsilon \qquad \varepsilon \to 0\ . \tag{4.9}$$

It follows that there is a parametrization $\varepsilon = \varepsilon(\rho)$ such that

$$\varepsilon(\rho)\ \sim\ \rho \qquad \rho \to 0 \tag{4.10}$$

$$R\left(\hat{\theta}_{\varepsilon(p)},\ 0\right)\ =\ \rho\ , \qquad 0\ <\ \rho\ <\ \rho_0\ .$$

Let $\tilde{\theta}_\rho \equiv \hat{\theta}_{\varepsilon(\rho)}$. Then $\tilde{\theta}_\rho$ is an estimator in $\hat{\Theta}_0(\rho)$. Moreover, if we define $\tilde{\pi}_\rho = \pi_{\varepsilon(\rho)}$ then

$$B\left(\tilde{\pi}_\rho\right)\ =\ \rho \cdot \left(2\ \log(\rho^{-1})\right)\ (1 + o\ (1)) \qquad \rho \to 0\ .$$

Now let $\hat{\theta}$ be any estimator satisfying $R(\hat{\theta}, 0) \le \rho$. By the definition of Bayes Risk, $E_{\tilde{\pi}}\ R(\hat{\theta},\ \theta) \ge B\ (\tilde{\pi}_\rho)$. Setting $\tilde{\mu}_\rho = \tilde{\mu}(\varepsilon(\rho))$ we may rewrite this as

$$\begin{aligned}(1-\varepsilon)\ R(\hat{\theta}, 0)\ +\ \frac{\varepsilon}{2}\ R(\hat{\theta},\ \tilde{\mu}_\rho)\ +\ \frac{\varepsilon}{2}\ R(\hat{\theta},\ -\tilde{\mu}_\rho)\ &\ge\ (1-\varepsilon)\ R(\tilde{\theta}_\rho, 0) + \varepsilon R(\tilde{\theta}_\rho,\ \tilde{\mu}_\rho)\ \ , \\ &=\ \ (1-\varepsilon)\rho\ +\ \varepsilon R(\tilde{\theta}_\rho,\ \tilde{\mu}_\rho)\ .\end{aligned}$$

Hence,

$$\begin{aligned}\mathrm{Ave}\left\{R(\hat{\theta},\ \tilde{\mu}_\rho),\ \ R(\hat{\theta},\ -\tilde{\mu}_\rho\ )\right\}\ &\ge\ R(\tilde{\theta}_\rho,\ \tilde{\mu}_\rho)\ +\ \frac{1-\varepsilon}{\varepsilon}\ \ \left(\rho - R(\hat{\theta},\ 0)\right) \\ &\ge\ R(\tilde{\theta}_\rho,\ \tilde{\mu}_\rho) \qquad\qquad \left(\text{as}\ \ \tilde{\theta}_\rho\ \in\ \hat{\Theta}_0(\rho)\right)\ .\end{aligned}$$

It follows that

$$\begin{aligned}\inf_{\hat{\theta} \in \hat{\Theta}_0(\rho)}\ \ \sup_{|\theta| \le \tilde{\mu}_\rho}\ \ R(\hat{\theta}, \theta)\ &\ge\ R(\tilde{\theta}_\rho,\ \tilde{\mu}_\rho) \\ &=\ R\left(\hat{\theta}_{\varepsilon(\rho)},\ \tilde{\mu}(\varepsilon(\rho))\right) \\ &=\ \tilde{\mu}\left(\varepsilon(\rho)\right)^2\ \ (1 + o(1)) \qquad\qquad \rho \to 0\ \ (\text{by}\ \ 4.7)\ .\end{aligned} \tag{4.11}$$

(This is a lower bound for a hybrid minimax risk problem: worst case risk over a bounded interval, subject to doing well at a point.) Apply this to give (4.1):

$$\begin{aligned}
& \inf_{\hat\theta \in \hat\Theta_0(\rho)} \sup_{\tau \geq 1} \tau^{-p} \sup_{|\theta| \leq \tau} R(\hat\theta, \theta) \\
\geq\ & \inf_{\hat\theta \in \hat\Theta_0(\rho)} \tilde\mu_\rho^{-p} \sup_{|\theta| \leq \tilde\mu_\rho} R(\hat\theta, \theta) && \text{as } \tilde\mu_\rho \geq 1 \\
\geq\ & \tilde\mu^{-p}(\varepsilon(\rho))\ \tilde\mu^2(\varepsilon(\rho))\ (1 + o(1)) && \rho \to 0 \ \text{(by (4.11))} \\
\sim\ & \tilde\mu(\varepsilon(\rho))^{2-p} \sim \left(2\ \log(\varepsilon(\rho)^{-1})\right)^{1-p/2} \\
\sim\ & \left(2\ \log(\rho^{-1})\right)^{1-p/2} && \text{as } \varepsilon \to 0 \ \text{(by (4.10))}.
\end{aligned}$$

5. Attainability of $K^p(\rho)$

To complete the proof of Theorem 3 we now show that the estimator $\hat\theta^*$ defined there is asymptotically minimax. First, using (2.5),

$$R(\hat\theta^*, 0) = (1 + t^2)\, 2\Phi(-t) + 2t\phi(t).$$

The asymptotic expansion

$$\Phi(-u) = \varphi(u)\left\{u^{-1} - u^{-3} + 3u^{-5} + O(u^{-7})\right\}, \qquad u \geq 1$$

gives $R(\hat\theta^*, 0) \sim 4t^{-3}\ \varphi(t)$ as $\rho \to 0$. Hence $\rho^{-1}\ R(\hat\theta^*, 0) \sim \frac{4t^{-3}}{\sqrt{2\pi}}$ as $\rho \to 0$.
We conclude that for all sufficiently small $\rho > 0$, $R(\hat\theta^*, 0) \leq \rho$, and that $\hat\theta^* \in \hat\Theta_0(\rho)$.

On the other hand, under the calibration $\rho = \delta$, we apply the analysis of section 1; $\hat\theta^*$ is asymptotically minimax for L^p as $\delta \to 0$, so

$$R(\hat\theta^*, \theta) \leq L^*(\delta)\,(\delta + \theta^p) \qquad \forall \theta \in \mathcal{R}$$

where L^* is a factor satisfying $L^*(\delta) = L^p(\delta)\ (1 + o(1))$. Thus

$$\begin{aligned}
\sup_{\tau \geq 1} \tau^{-p} \sup_{|\theta| \leq \tau} R(\hat\theta^*, \theta) &\leq \sup_{\tau \geq 1} \tau^{-p} \cdot L^*(\delta)\ (\delta + \tau^p) \\
&= L^*(\delta)\,(1 + \delta) \\
&= L^p(\delta)\,(1 + \delta)\ (1 + o(1)) \\
&= K^p(\rho)\ (1 + o(1))
\end{aligned}$$

under the calibration $\rho = \delta$. We conclude that $\hat{\theta}^*$ is asymptotically minimax for $K^p(\rho)$.

6. Sequence Space.

Turn now to the sequence space described in the introduction. We suppose that we have n noisy observations $y_i = \theta_i + \epsilon_n z_i$, $i = 1, ..., n$, with $z_i \overset{i.i.d.}{\sim} N(0,1)$, where $\epsilon_n \to 0$, and that our goal is to estimate θ with small squared ℓ^2 risk: $E\left|\left|\hat{\theta}(y) - \theta\right|\right|^2_{\ell^2_n}$.

6.1. Minimax Risk over ℓ^p Balls

We recall that

$$E(\hat{\theta}^*(Y) - \theta)^2 \le L^*(\delta_n)\{\delta_n + |\theta|^p\},$$

where $L^*(\delta_n) \sim L^p(\delta_n)$, $n \to \infty$. Now if $\theta \in \ell^p_n(C)$,

$$\begin{aligned} E\left|\left|\hat{\theta}^*_n(y) - \theta\right|\right|^2_{\ell^2_n} &\le L^*(\delta_n)\,\epsilon_n^2 \sum_{i=1}^n \left(\delta_n + \left|\frac{\theta_i}{\epsilon_n}\right|^p\right) \\ &\le L^*(\delta_n)\left(n\epsilon_n^2\delta_n + \epsilon_n^{2-p}C_n^p\right) . \end{aligned}$$

By choice of $\delta_n = \gamma_n \eta_n^p$, we have $n\epsilon_n^2\delta_n = \gamma_n \epsilon_n^{2-p} C_n^p = o(\epsilon_n^{2-p}C_n^p)$, and so

$$\sup_{\ell^p_n(C_n)} E\left|\left|\hat{\theta}^*_n(y) - \theta\right|\right|^2_{\ell^2_n} \le L^p(\delta_n) C_n^p \epsilon_n^{2-p}\,(1 + o(1)) .$$

We conclude from lower bounds in Donoho and Johnstone (1994) that this behavior is optimal; Theorem 4 follows.

6.2. Minimax Risk over weak ℓ^p Balls

We recall that

$$E(\hat{\theta}^*(Y) - \theta)^2 \le M^*(\delta_n)\{1/n + m^p_{1/n}(\theta)\},$$

where $M^*(\delta_n) \sim M^p(\delta_n)$, $n \to \infty$. Now if $\theta \in \mathbf{m}^p_n(C)$,

$$E\left|\left|\hat{\theta}^*_n(y) - \theta\right|\right|^2_{\ell^2_n} \le M^*(\delta_n)\,\epsilon_n^2 \sum_1^n \left(\delta_n + m^p_{\delta_n}\left(\frac{\theta_i}{\epsilon_n}\right)\right) \tag{6.1}$$

Now a side calculation (reproduced in the Appendix) gives

$$\sup\left\{\sum_{i=1}^n m^p_{\delta_n}\left(\frac{\theta_i}{\epsilon_n}\right) : \theta \in \mathbf{m}^p_n(C_n)\right\} \le \frac{2}{2-p}\,\epsilon_n^{-p} C_n^p \tag{6.2}$$

which leads to a bound for the right side of (6.2) as

$$M^*(\delta_n)\left(n\epsilon_n^2\delta_n + \frac{2}{2-p}\,\epsilon_n^{2-p}C_n^p\right) .$$

¿From $M^*(\delta_n) \sim M^p(\delta_n)$, and $n\epsilon_n^2\delta_n = o(\epsilon_n^{2-p}C_n^p)$,

$$\sup_{\mathbf{m}_n^p(C_n)} E\left|\left|\hat{\theta}_n^*(y) - \theta\right|\right|_{\ell_n^2}^2 \;\leq\; \frac{2}{2-p} M^p(\delta_n) C_n^p \epsilon_n^{2-p}\,(1+o(1)) \;.$$

We conclude from lower bounds in Johnstone (1993) that this behavior is optimal; Theorem 5 follows.

7. Adaptation to Unknown Smoothness: Pointwise Risk

7.1 The Cost of Adaptation

In demonstrating Theorem 6, rather than work with the sampled white-noise data model (1.14), we begin with the continuous white noise model

$$Y(dt) = f(t) + \epsilon W(dt), \qquad t \in \mathcal{R}.$$

We will show that if an estimator $\hat{f}(0)$ satisfies

$$\sup_{\Lambda(\alpha_1, C_1)} E_f(\hat{f}(0) - f(0))^2 \leq B \cdot (\epsilon^2)^{r_1} (C_1^2)^{1-r_1} \tag{7.1}$$

then that same estimator must necessarily satisfy

$$\sup_{C \geq C_0} \left((\epsilon^2)^{r_0}(C^2)^{(1-r_0)}\right)^{-1} \sup_{\Lambda(\alpha_0, C)} E_f(\hat{f}(0) - f(0))^2 \;\geq\; A_0 K^{p_o} \left(A_1 \epsilon^{2r_1 - 2r_0}\right) \tag{7.2}$$

Theorem 6 follows by two comments we make after we prove (7.2).

Consider the problem of estimating $T(f) = f(0)$ when f is known to lie in $\Lambda(\alpha_0, 1)$ and the noise level $\epsilon = 1$. There is a hardest 1-dimensional subfamily for linear estimates, i.e. a segment $\{\xi\psi_{1,1} : \xi \in [-1,1]\}$ with the property that the problem of estimating $T(f)$ over this segment is equally as hard as estimating $T(f)$ over all of $\Lambda(\alpha_0, 1)$. We are really interested in the hardest 1-dimensional subfamily for estimating $T(f)$ over a particular class $\Lambda(\alpha_0, \gamma)$ at noise level ϵ, where $\gamma = C_0\|\psi_{1,1}\|_2$. By a renormalization argument, (Donoho and Low, 1992), defining

$$\psi_{\epsilon,\gamma}(t) = a\psi_{1,1}(bt), \qquad ab^{\alpha_0} = \gamma, \quad ab^{-1/2} = \epsilon \tag{7.3}$$

gives the hardest 1-dimensional subfamily in the form $\{\xi\psi_{\epsilon,\gamma} : \xi \in [-1,1]\}$. Morever, the minimax linear risk for estimating $T(f)$ over this family at noise level ϵ is $A(\alpha_0)(\gamma^2)^{1-r_0} \cdot (\epsilon^2)^{r_0}$. Define

$$\theta = \xi\|\psi_{1,1}\|_2, \qquad f_\theta = \xi\psi_{\epsilon,\gamma}, \qquad \theta \in \mathcal{R}.$$

Consider now the problem of estimating $f(0)$, for f in the unbounded 1-dimensional subfamily $\{f_\theta\}_{\theta\in\mathcal{R}}$. Note that

$$f_\theta \in \Lambda(\alpha_0, \xi\gamma)$$

so that this is a one-dimensional subproblem of the problem of estimating f which is known to belong to some $\Lambda(\alpha_0, C)$, with C unknown. Moreover, note that

$$f_0 \in \Lambda(\alpha_1, C_1).$$

The one-dimensionality of the family $\{f_\theta\}$ means that the unit-variance statistic

$$y = (\int \psi_{\epsilon,\gamma} Y(dt))/(\epsilon \|\psi_{\epsilon,\gamma}\|_2)$$

is a sufficient statistic for θ and hence for $T(f_\theta)$. Hence, applying the Rao-Blackwell process if necessary, we have a correspondence between (non-randomized) estimation $\hat{f}(0)$ of $f(0)$ based on Y and estimation of θ, given by $\hat{\theta}(y) = (\left|\left|\psi_{1,1}\right|\right|_2 / \psi_{\epsilon,\gamma}(0)) E\{\hat{f}(0)|y\}$. Under this correspondence, we have

$$E_{f_\theta}\left(\hat{f}(0) - f(0)\right)^2 \;\geq\; \left(\frac{\psi_{\epsilon,\gamma}(0)}{\left|\left|\psi_{1,1}\right|\right|_2}\right)^2 E_\theta(\hat{\theta}(y) - \theta)^2.$$

Moreover, the renormalization principle (7.3) gives $\psi^2_{\epsilon,\gamma}(0) = (\epsilon^2)^{r_0}(\gamma^2)^{1-r_0}\psi^2_{1,1}(0)$, and hence the lower bound

$$E_{f_\theta}\left(\hat{f}(0) - f(0)\right)^2 \;\geq\; \left(\frac{\psi_{1,1}(0)}{\left|\left|\psi_{1,1}\right|\right|_2}\right)^2 \cdot (\epsilon^2)^{r_0}(\gamma^2)^{1-r_0} \cdot E_\theta(\hat{\theta}(y) - \theta)^2. \tag{7.4}$$

Now suppose we have an estimator $\hat{f}(0)$ satisfying (7.1); then applying the above correspondence, we obtain an estimator $\hat{\theta}$ satisfying $E_{\theta=0}(\hat{\theta}(y) - \theta)^2 \leq \rho$, where

$$\rho = B(\epsilon^2)^{r_1}(C_1^2)^{1-r_1}\frac{\left|\left|\psi_{1,1}\right|\right|_2^2}{\psi^2_{\epsilon,\gamma}(0)} \;\;=\;\; A_1 \cdot \epsilon^{2(r_1 - r_0)},$$

say. Using (7.4) and the observation that $f_\theta \in \Lambda(\alpha_0, C)$ entails $|\theta| \leq \tau(C) \equiv \left|\left|\psi_{1,1}\right|\right|_2 C/\gamma$, we may bound the left side of (7.2) from below via

$$\frac{\psi^2_{1,1}(0)}{\left|\left|\psi_{1,1}\right|\right|_2^2} \sup_{C \geq C_0} \left(\frac{\gamma}{C}\right)^{2(1-r_0)} \sup_{|\theta| \leq \tau(C)} E(\hat{\theta} - \theta)^2 \;.$$

¿From the choice $\gamma = C_0\left|\left|\psi_{1,1}\right|\right|_2$, and the formula for $\tau(C)$, we have $\gamma/C = \left|\left|\psi_{1,1}\right|\right|_2/\tau$. Setting $A_0 = \psi^2_{1,1}(0)/\left|\left|\psi_{1,1}\right|\right|_2^{2r_0}$, the previous expression becomes

$$A_0 \sup_{\tau \geq 1} \tau^{-p_0} \sup_{|\theta| \leq \tau} E(\hat{\theta} - \theta)^2 \geq A_0\, K^{p_0}(\rho) \;\;=\;\; A_0\, K^{p_0}\left(A_1 \epsilon^{2r_1 - 2r_0}\right) \;.$$

This proves (7.2). To get Theorem 6, we use the following two remarks. (a) *Allowing B to vary.* Instead of assuming that B were constant, we can retrace the above steps under the

condition $B = B(\epsilon)$ in (7.1), where $B(\epsilon) \le B_0\, K^{p_0}\left(B_1 \epsilon^{2r_1 - 2r_0}\right) = O((\log \epsilon^{-1})^{r_0}),$. The formula for ρ changes slightly; everything else remains the same, and the conclusion is

$$\sup_{C \ge C_0} \left((\epsilon^2)^{r_0}(C^2)^{(1-r_0)}\right)^{-1} E_f(\hat{f}(0) - f(0))^2 \ge A' K^{p_0}(A'' \log(\epsilon^{-1})^{r_1}(\epsilon^2)^{r_1 - r_0}).$$

Owing to the logarithmic growth of K^{p_0}, the extra logarithmic term inside of K^{p_0} in this display does not affect the leading asymptotics, which therefore turn out the same as those of the right-hand-side of (7.2), so allowance for this extra logarithmic factor leads to the same type of bound. In sum,

$$\max_{i=0,1} \sup_{C \ge C_i} \left((\epsilon^2)^{r_i}(C^2)^{(1-r_i)}\right)^{-1} \sup_{\Lambda(\alpha_i, C)} E_f(\hat{f}(0) - f(0))^2 \ge A' K^{p_0}(A''(\epsilon^2)^{r_1 - r_0})) \cdot (1 + o(1)) \quad (7.5)$$

(b) *Discretization.* With data (1.14), calibrate $\epsilon = \sigma/\sqrt{n}$ and use the subfamily $\{f_\theta\}$ constructed in the proof of (7.2). Then introduce the sufficient statistic

$$y_n = \sum_i \psi_{\epsilon,\gamma}(t_i) y_i / (Var \sum_i \psi_{\epsilon,\gamma}(t_i) y_i)^{1/2}$$

and argue as above, bounding various approximation errors. One gets a conclusion just like (7.5), only with $\sigma/\sqrt{n}$ in place of ϵ.

7.2 Adapting with Minimal Cost

For Theorem 7, the wavelet transform will be based on *periodized orthogonal wavelets*; although we could also use the boundary-corrected orthogonal wavelets of Meyer (1991) or of Cohen, Daubechies, Jawerth and Vial (1992). We fix D, the number of vanishing moments, and j_0, a "low-resolution cutoff". To make the proof simpler, we suppose that the low resolution cutoff is chosen finely enough that *the boundary does not interact with zero*; more precisely so that any wavelet which is nonvanishing at zero is vanishing in a neighborhood of the boundary $\{-1/2, 1/2\}$.

These choices determine a wavelet transform which takes $n = 2^{j_1+1}$ numbers (d_i), viewed as samples at equispaced points $t_i \in (-1/2, 1/2]$, and delivers n wavelet coefficients $(v_{j,k})$. The coefficients yield the reconstruction formula

$$y_i = \sum_{j,k} v_{j,k} w_{j,k}(i)$$

where the vectors $w_{j,k}$ are "wavelets". Here we use a double indexing scheme, where j refers to scale, and k refers to position. j runs from j_0 to j_1; the $v_{j_0,k}$, $0 \le k < 2^{j_0+1}$ are low frequency terms. For $j > j_0$, the $v_{j,k}$, $0 \le k < 2^j$, represent terms of resolution $\approx 2^{-j}$ and position $\approx k/2^j - 1/2$.

This transform has a number of useful properties. The inequalities below hold with finite positive constants C_i that are independent of $n = 2^{j_1+1}$ as soon as $j_1 > J$.

[W1] *Orthogonality.* For all $(y_i) \in \mathcal{R}^n$, $n^{-1}\|(y_i)\|_2^2 = \|(v_{j,k})\|_2^2$.

[W2] *Noise.* Let (z_i) be i.i.d. $N(0,\sigma^2)$. Then the corresponding wavelet coefficients $(z_{j,k})$ have a joint Normal distribution with $Var(z_{j,k}) = \sigma^2 n^{-1}$.

[W3] *Height of Wavelets.* $\|w_{j,k}\|_\infty \le C_3 \cdot 2^{j/2}$, $j \ge j_0$.

[W4] *Width of Wavelets.* With $Q(j,k) = supp(w_{j,k})$, $n^{-1}|Q(j,k)| \le C_4 \cdot 2^{-j}$, $j \ge j_0$.

[W5] *Analysis of Hölder Classes.* Suppose that f satisfies the Hölder condition $f \in \Lambda(\alpha, C)$, $\alpha < D$. Denote the wavelet coefficients $(\theta_{j,k})$ of the samples $(f(t_i))$; let $\mathcal{I}$ denote the collection of indices of wavelets $w_{j,k}$ which are nonzero at 0. Then

$$|\theta_{j,k}| \le C_5 \cdot C \cdot 2^{-j(\alpha+1/2)}, \quad (j,k) \in \mathcal{I}.$$

Before proceeding with the proof of Theorem 7, we make a few remarks about the M^p-problem. Define

$$\mu_n^p(\xi,\epsilon) = m_{1/n}^p(\xi/\epsilon) \cdot \epsilon^2.$$

Setting $r = (1 - p/2)$ we have $\mu_n^p(\xi,\epsilon) \le |\xi|^{2(1-r)} \cdot \epsilon^{2r}$; in fact, with $\tau_n = \sqrt{2\log(n)}$ we have

$$\xi^{2(r-1)} \cdot m_{1/n}^p(\xi) \quad = \quad \min((\frac{\xi}{\tau_n})^{2r}, (\frac{\tau_n}{\xi})^{2(1-r)}), \qquad \xi > 0. \tag{7.6}$$

Also, recalling definition and inequality (2.1) in the noise-level one problem, for an estimate $\hat\theta_{j,k}^*$ we have

$$E(\hat\theta_{j,k}^* - \theta_{j,k})^2 \le M_p^*(1/n)(\epsilon^2/n + \mu_n^p(\theta_{j,k},\epsilon)).$$

where, of course $M_p^*(1/n) \sim M^p(1/n)$ as $n \to \infty$.

Our proof begins with a decomposition of $\hat f_n^*(0) - f(0)$. With $Q(j,k)$ the support of $w_{j,k}$ and, again, $\mathcal{I} = \{(j,k) : 0 \in Q(j,k)\}$, then

$$\hat f_n^*(0) - f(0) = \sum_{\mathcal{I}} (\hat\theta_{j,k}^* - \theta_{j,k}) w_{j,k}(0) = \sum_{\mathcal{I}} X_{j,k},$$

say. Applying our M^p-analysis of $\hat\theta_{j,k}^*$ with $\epsilon^2 = \sigma^2/n$, $p = 2(1-r)$, and using [W3] we get

$$EX_{j,k}^2 \le M_p^*(1/n) C_3^2 \cdot 2^j (\epsilon^2/n + \mu_n^p(\theta_{j,k},\epsilon)),$$

Now if $f \in \Lambda(\alpha, C)$ then by [W5] $|\theta_{j,k}| \le C_5 \cdot C \cdot 2^{-j(\alpha+1/2)}$ for those coefficients having influence at 0, and so for such coefficients

$$EX_{j,k}^2 \le M_p^*(1/n) \cdot C_3^2 \cdot (2^j \epsilon^2/n + 2^j \cdot \mu_n^p(C_5 \cdot C \cdot 2^{-j(\alpha+1/2)}, \epsilon)) \quad (j,k) \in \mathcal{I}.$$

In the appendix, we prove

Lemma 3. *For $\alpha > 0$ and $j \geq 0$,*

$$2^j \cdot \mu_n^p(C_5 \cdot C \cdot 2^{-j(\alpha+1/2)}, \epsilon) \;\leq\; R_n 2^{-2\delta|j-j_n|}, \tag{7.7}$$

where $\delta = min(\alpha, 1/2)$, $r = r(\alpha) = 2\alpha/(2\alpha+1)$, $p = 2(1-r)$,

$$j_n = (\log_2(CC_5) - \log_2(\epsilon\tau n))/(\alpha + 1/2),$$

$$R_n \equiv (\epsilon^2)^r \cdot (C_5^2 C^2)^{1-r}. \tag{7.8}$$

This gives immediately the decisive inequality

$$EX_{j,k}^2 \leq M_p^*(1/n) \cdot C_3^2 \cdot (2^j\epsilon^2/n + R_n 2^{-2\delta|j-j_n|}). \tag{7.9}$$

We plan to use

$$E(\sum_{\mathcal{I}} X_{j,k})^2 \leq (\sum_{\mathcal{I}} \sqrt{EX_{j,k}^2})^2 \tag{7.10}$$

and $(a+b)^{1/2} \leq (a^{1/2} + b^{1/2})$. Inequality (7.9) gives

$$\sum_{\mathcal{I}} \sqrt{EX_{j,k}^2} \leq M_p^*(1/n)^{1/2} \cdot C_3 \cdot \sum_{\mathcal{I}} (2^{j/2}\epsilon/\sqrt{n} + R_n^{1/2} 2^{-\delta|j-j_n|}).$$

Now at each level j there are at most $2 \cdot C_4$ different k such that $(j,k) \in \mathcal{I}$; hence

$$\sum_{\mathcal{I}} \sqrt{EX_{j,k}^2} \leq M_p^*(1/n)^{1/2} \cdot 2 \cdot C_3 C_4 \cdot \left[\sum_{j=j_0}^{j_1} 2^{j/2}\epsilon/\sqrt{n} + R_n^{1/2} \sum_{j=j_0}^{j_1} 2^{-\delta|j-j_n|} \right]$$

Now use $n^{-1/2} \sum_{j=j_0}^{j_1} 2^{j/2} \leq 2^{3/2}$ and $\sum_{j=-\infty}^{\infty} 2^{-\delta|j-j_n|} \leq 2/(1-2^{-\delta})$.

$$\sum_{\mathcal{I}} \sqrt{EX_{j,k}^2} \leq M_p^*(1/n)^{1/2} \cdot 2 \cdot C_3 C_4 \cdot \left[2^{3/2}\epsilon + R_n^{1/2} \frac{2}{1-2^{-\delta}} \right],$$

Square both sides and apply (7.10):

$$E(\sum_{\mathcal{I}} X_{j,k})^2 \leq M_p^*(1/n) \cdot R_n \cdot (2 \cdot C_3 C_4)^2 \cdot \left[\frac{2}{1-2^{-\delta}} + 2^{3/2}\epsilon/R_n^{1/2} \right]^2$$

Comparing this with (1.19), taking into account inequality (2.7), the definition (7.8) of R_n, the definition (1.20) of $A(\alpha)$, leads to the following strengthening of the conclusion of Theorem 7, valid for $n \geq n_0$,

$$\sup_{\Lambda(\alpha,C)} E\left(\hat{f}_n^*(0) - f(0)\right)^2 \leq (2\log n)^r A(\alpha)(C^2)^{1-r}(\epsilon^2)^r \left(1 + (2\log n)^{-1}\right) \left(1 + 2^{3/2} C_\delta \left(\frac{\epsilon}{C_5 C}\right)^{1-r}\right)^2$$

(where $C_\delta = 2/(1-2^{-\delta})$).

Appendix

PROOF OF LEMMA 2 (4.6) - (4.9)

Note that (4.3) may be written in the form

$$(1-\varepsilon)\ \phi(a+\mu) \quad = \quad \frac{\varepsilon}{2}\ \phi(a)\ .$$

Combining this with (4.4) shows that $a(\varepsilon) \sim \sqrt{2\ \log\ \mu(a(\varepsilon),\varepsilon)}$, and in particular, $a = o(\mu)$, so that (4.6) follows directly from (4.3). Since

$$B(\pi_\varepsilon) \quad = \quad (1-\varepsilon)\ R(\hat{\theta}_\varepsilon,\ 0)\ +\ \frac{\varepsilon}{2}\left[R\left(\hat{\theta}_\varepsilon, -\tilde{\mu}(\varepsilon)\right)\ +\ R\left(\hat{\theta}_\varepsilon, -\tilde{\mu}(\varepsilon)\right)\right]\ ,$$

(4.7) follows from (4.8) and (4.9), symmetry of $R(\hat{\theta}_\varepsilon, \mu)$ about 0 and (4.6).

Let $p(\theta|y)$ denote the posterior distribution of θ under prior π_ε. Abbreviating $\hat{\mu}(\epsilon)$ by μ, we have that the Bayes rule $\hat{\theta}_\varepsilon = \mu\ q_\varepsilon(y)$, where $q_\varepsilon(y) = p\,(\mu|y) - p\,(-\mu|y)$. Thus

$$R\left(\hat{\theta}_\varepsilon,\ \tilde{\mu}(\varepsilon)\right) \quad = \quad \tilde{\mu}^2(\varepsilon)\ E_\mu\ \left[1 - q_\varepsilon\ (\mu+z)\right]^2$$

where $z \sim N(0,\ 1)$. We have

$$q_\varepsilon(y) \quad = \quad \frac{\frac{\varepsilon}{2}\left[e^{y\mu-\mu^2/2}\ -\ e^{-y\mu-\mu^2/2}\right]}{1-\varepsilon+\frac{\varepsilon}{2}\left[e^{-y\mu-\mu^2/2}+e^{-y\mu-\mu^2/2}\right]}\ .$$

Using (4.3), we obtain

$$q_e(\mu+z) \quad = \quad \frac{e^{\mu(z-a)}-e^{-2\mu^2-\mu(z+a)}}{1+e^{\mu(z-a)}+e^{-2\mu^2-\mu(z+a)}} \qquad (A.1)$$

Since $a(\epsilon) \uparrow \infty$ it follows that $q_\varepsilon(\mu+z) \quad \stackrel{P}{\to} \quad 0$, as $\varepsilon \to 0$ which establishes (4.8) since $|q_\varepsilon| \le 1$.

For (4.9), we first write

$$R\left(\hat{\theta}_\varepsilon,\ 0\right) \quad = \quad 2\mu^2\ \int_0^\infty\ q_\varepsilon^2\ (y)\ \phi(y)\ dy\ . \qquad (A.2)$$

The identity (A.1) suggests that we consider first the above integral, I_ε say, with q_ε replaced by

$$\tilde{q}_\varepsilon(y) \quad = \quad \tilde{q}_\varepsilon(\mu+z) \quad = \quad \frac{e^{\mu(z-a)}}{\left[1+e^{\mu(z-a)}\right]}\ .$$

Laplace's method leads to the following Lemma.

LEMMA A. If

$$a = o(\mu), \qquad \widetilde{I}_\varepsilon = \int_{-\infty}^{\infty}\ \tilde{q}_\varepsilon^2(\mu+z)\ \phi(\mu+z)\ dz \sim \frac{\sqrt{\pi}}{2\mu}\ \phi(a+\mu), \qquad \mu \to \infty\ .$$

The proof of the Lemma makes it clear that $I_\varepsilon \sim \widetilde{I}_\varepsilon$, and hence by substitution of its conclusion into (A.2),

$$R\left(\hat{\theta}_\varepsilon,\ 0\right) \ \sim\ \sqrt{\pi}\ \mu\ \phi(a+\mu) \quad = \quad \varepsilon$$

by (4.4). This establishes (4.9).

PROOF OF LEMMA 3

Now $(2(1-r))^{-1} = (\alpha + 1/2)$. Hence $2^j = (2^{-j(\alpha+1/2)})^{2(r-1)}$ and so with $\xi_j = C_5 \cdot C \cdot 2^{-(\alpha+1/2)j}/\epsilon$,

$$2^j = R_n \cdot \frac{\xi_j^{2(r-1)}}{\epsilon^2}.$$

Hence, using $\mu_n^p(\theta,\epsilon) = \epsilon^2 \mu_n^p(\theta/\epsilon, 1)$,

$$2^j \mu_n^p(C_5 \cdot C \cdot 2^{-(\alpha+1/2)j}, \epsilon) = R_n \cdot \xi_j^{2(r-1)} \cdot m_{1/n}^p(\xi_j)\ .$$

The indicated value of j_n is the unique real solution of

$$C_5 \cdot C \cdot 2^{-(\alpha+1/2)j}/\epsilon = \tau_n.$$

Hence $\xi_j/\tau_n = 2^{-(j-j_n)(\alpha+1/2)}$; applying (7.6),

$$\xi_j^{2(r-1)} \cdot m_{1/n}^p(\xi_j) \ \le\ 2^{-2\,\delta|j-j_n|}$$

so that (7.7) follows and we are done.

VERIFICATION OF (6.2).

Set $\tau_n = \sqrt{2\log \delta_n^{-1}}$ and $\tau = t_n \epsilon_n C_n^{-1}$, and note from the definition of the weak ℓ_p-ball that

$$(6.2) \ \le\ \sup\left\{ \sum_1^n\ t_n^p\ \min\left(\theta_k^2 t_n^{-2}\epsilon_n^{-2},\ 1\right)\ :\ 0 \le \theta_k \le C_n k^{-1/p} \right\}$$

$$= \quad C_n^2 t_n^{p-2}\epsilon_n^{-2}\ \sum_{k=1}^n\ \min\left(k^{-2/p},\ \tau^2\right)\ .$$

Now setting $s_* = \tau^{-p}$, a simple calculation shows that

$$\sum_1^\infty\ k^{-2/p} \wedge \tau^2 \ \le\ \int_0^\infty s^{-2/p} \wedge \tau^2\, ds$$

$$= \quad s_*\tau^2 \ +\ \frac{p}{2-p}\ s_*^{1-2/p}\ .$$

$$= \quad \tau^{2-p}\left(\frac{1+p}{(2-p)}\right)$$

Hence

$$(6.2) \ \le\ C_n^2 t_n^{p-2}\epsilon_n^{-2}\ \cdot\ \frac{2}{2-p}\ \cdot\ \tau_n^{2-p} \quad = \quad \frac{2}{2-p}\ C_n^p \epsilon_n^{-p}\ .$$

References

Bickel, P.J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Statist.* **9**, 1301-1309.

Bickel, P. J. (1983). Minimax estimation of a normal mean subject to doing well at a point. In *Recent Advances in Statistics* (M. H. Rizvi, J. S. Rustagi, and D. Siegmund, eds.)., Academic Press, New York, 511–528.

Brown, L.D. and M.G. Low (1992). Superefficiency and lack of adaptability in functional estimation. Mss.

Casella, G. and Strawderman, W.E. (1981). Estimating a Bounded Normal Mean. *Ann. Statist.* **9** 870-878.

Cohen, A., I. Daubechies, B. Jawerth and P. Vial, Multiresolution analysis, wavelets, and fast algorithms on an interval, *Comptes Rendus Acad. Sci. Paris* (A). **316** (1992)., 417-421.

DeVore, R.A., Jawerth, B. and Popov V. (1990). Compression of Wavelet Decompositions.

DeVore, R.A. and B.J. Lucier (1992). Fast wavelet techniques for near-optimal image processing, *Proc. IEEE Mil. Commun. Conf.* Oct. 1992. IEEE Communications Society, NY, 1992.

DeVore, R.A. and V. Popov (1988). Interpolation of Besov spaces, *Trans. Amer. Math. Soc.* **1** 397-414.

Donoho, D.L. (1992). Interpolating Wavelet Transforms, To appear, *Applied and Computational Harmonic Analysis.*

Donoho, D.L. (1993a). Smooth Wavelet Decompositions with Blocky Coefficient Kernels, in *Recent Advances in Wavelet Analysis*, Ed. L.L. Schumaker and G. Webb, San Diego: Academic Press, 259-308.

Donoho, D.L. (1993b). Unconditional bases are optimal bases for data compression and for statistical estimation, *Applied and Computational Harmonic Analysis* **1**.

Donoho, D.L. (1994). Statistical Estimation and Optimal Recovery, To appear *Ann. Statist.*

Donoho, D.L. and I.M. Johnstone (1992a). Minimax Estimation via Wavelet Shrinkage. To appear *Ann. Statist.*

Donoho, D.L. and I.M. Johnstone (1992b). Ideal Spatial Adaptation via Wavelet Shrinkage. To appear *Biometrika.*

Donoho, D.L. and I.M. Johnstone (1994). Minimax Risk for ℓ^q losses over ℓ^p-balls. To appear *Prob. Thry. Rel. Flds.*

Donoho, D.L., Johnstone, I.M., Hoch, J.C. and Stern, A.S. (1992). Maximum Entropy and the Nearly Black Object. *J. Roy. Stat. Soc. Ser. B,* **54,** 41–81 (with discussion).

Donoho, D.L., I.M. Johnstone, G. Kerkyacharian and D. Picard (1993). Wavelet Shrinkage: Asymptopia?, To appear, *J. Roy. Stat. Soc. Ser. B,*.

Donoho, D.L., and R.C. Liu (1991). Geometrizing rates of convergence III. *Ann. Statist.,* **19,** June, 1991, 668-701.

Donoho, D.L., R.C. Liu and B. MacGibbon. (1990). Minimax risk over hyperrectangles, and implications. *Ann. Statist.,* **18,** 1416-1437.

Donoho, D.L., and M.G.Low (1992). Renormalization Exponents and optimal pointwise rates of convergence. *Ann. Statist.,* **20,** 944-970.

Efroimovich, S. Y. and M.S. Pinsker (1981). Estimation of square-integrable [spectral] density on the basis of a sequence of observations. *Problems Inform. Transmission* **17**, 50–68.

Efroimovich, S. Y. and M.S. Pinsker (1982). Estimation of square-integrable probability density of a random variable. *Problems Inform. Transmission* **18**, 175–182.

Efromovich, S.Y. and M.G. Low (1992). Adaptive Estimation of Linear Functionals. Manuscript.

Ibragimov, I.A. and R.Z. Has'minskii (1984). Nonparametric estimation of the value of a linear functional in a Gaussian white noise. *Theory of Probability and its Applications* **29**, 1-17.

Johnstone, I.M. (1993). Minimax Bayes, asymptotic minimax and sparse wavelet priors. To appear, *Proc. 5th Purdue International Symposium on Decision Theory and Related Topics* Springer. New York.

Johnstone, I.M. (1994). Minimax Estimation of a sparse normal mean vector. To appear, *Ann. Statist.*

Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1992). Estimation d'une densité de probabilité par méthode d'ondelettes. *Comptes Rendus Acad. Sciences Paris* (A). **315** 211-216.

Kerkyacharian, G. and Picard, D. (1992). Density estimation in Besov Spaces. *Statistics and Probability Letters* **13** 15-24

Levit, B.Y. (1981). On asymptotic minimax estimates of the second order. *Theory of Probability and its Applications* **25**, 552-568

Lepskii,O.V. (1991). On a problem of adaptive estimation in Gaussian White Noise. *Theory of Probability and its Applications* **35**, 3, 454-466

Meyer, Y., (1991). Ondelettes sur l'intervalle, *Revista Mat. Ibero-Americana* **7**, 115-134.

Nussbaum, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L_2. *Ann. Statist.* **13**, 984–997.

Wolfowitz, J. (1950). Minimax estimation of the mean of a normal distribution with known variance. *Ann. Math. Stat.* **21**, 218-230.