The Annals of Statistics
1996, Vol. 24, No. 6, 2524–2535

A CONSTRAINED RISK INEQUALITY WITH APPLICATIONS TO NONPARAMETRIC FUNCTIONAL ESTIMATION

BY LAWRENCE D. BROWN[1] AND MARK G. LOW[2]

University of Pennsylvania

A general constrained minimum risk inequality is derived. Given two densities f_θ and f_0 we find a lower bound for the risk at the point θ given an upper bound for the risk at the point 0. The inequality sheds new light on superefficient estimators in the normal location problem and also on an adaptive estimation problem arising in nonparametric functional estimation.

1. Introduction. The problems of estimating a function at a point under squared error loss and the whole function under integrated squared error loss have held a central position in the nonparametric functional estimation literature. In particular, each has received a fairly detailed analysis in density estimation, nonparametric regression and white noise models. Progress to date can be contrasted as follows.

In both problems asymptotic rates of convergence are typically slower than $\sqrt{n}$. For integrated squared error loss, asymptotically minimax procedures have been found when the parameter space is a Sobolev space. For a given space these procedures may be chosen to be linear. For the pointwise estimation problem, typically there do not exist linear procedures which are asymptotically minimax. However, under mild regularity conditions appropriately chosen linear procedures have maximum risk within a small constant multiple of the minimax risk. See, for example, Ibragimov and Hasminskii (1984) or Donoho and Liu (1991). In particular, minimax rates of convergence can be achieved by linear procedures.

One of the most important results for the global estimation problem was the construction of adaptive estimators which are simultaneously asymptotically minimax over a large number of Sobolev spaces; see Efromovich and Pinsker (1984), Efromovich (1985) and Golubev (1987). Such adaptive procedures have not been found for the pointwise estimation problem.

Received February 1992; revised February 1995.

[1]Work supported in part by NSF Grants DMS-91-07842 and DMS-93-10228.

[2]Work supported by an NSF Postdoctoral Research Fellowship.

AMS 1991 *subject classifications*. Primary 62G99; secondary 62F12, 62F35, 62M99.

Key words and phrases. Adaptive estimation, superefficient estimators, nonparametric functional estimation, minimum risk inequalities, white noise model, density estimation, nonparametric regression.

Recently Lepskii (1990) has shown that for a white noise model it is not possible to find adaptive estimators for the pointwise problem which preserve minimaxity over a range of Lipschitz classes. Furthermore Lepskii showed that if an estimator is asymptotically rate minimax over one Lipschitz class, it must inflate the maximum risk over the other Lipschitz class by at least a logarithmic factor of the sample size. Lepskii even showed that these bounds can be attained. See also Lepskii (1991, 1992), and see Donoho and Johnstone (1992, 1994) and Efromovich and Low (1994) for some recent, related results.

In this paper we develop a two point inequality which is particularly well suited to providing lower bounds for adaptive estimation problems. The inequality gives, in a general setting, a lower bound for the squared error risk at one parameter point subject to having a small risk at another parameter point. The relationship to adaptation is spelled out in Section 4. Such an inequality is also related to the study of ε-minimax procedures and to the notion of superefficient estimation. The notion of ε-minimax procedures (also called almost subminimax procedures) was introduced by Robbins (1951) and Frank and Kiefer (1951) and further studied by Hodges and Lehmann (1952) and Bickel (1983, 1984). An example of superefficient estimation was found in Hodges (1952) and many further results were obtained by Le Cam (1953). We apply our inequality to this topic in Section 3A.

2. A constrained minimum risk inequality. Let Z have distribution with density f_{θ_1} or f_{θ_2} with respect to a measure μ. For any estimator δ based on Z, its risk is defined by

$$R(\varphi, \delta) = E(\varphi - \delta)^2 = \int (\varphi - \delta(x))^2 f_\varphi(x) \mu(dx). \tag{2.1}$$

The following theorem gives a lower bound for $R(\theta_2, \delta)$ given that $R(\theta_1, \delta) \le \varepsilon^2$. In the theorem and subsequently let $q(x) = f_{\theta_2}(x)/f_{\theta_1}(x)$ and

$$I = I(\theta_1, \theta_2) = E_{\theta_1}(q^2(X)). \tag{2.2}$$

[$q(x) = \infty$ for some x is possible, with the obvious interpretation $q(x)f_{\theta_1}(x) = f_{\theta_2}(x)$.]

THEOREM 1. *Let $\theta = \theta_2 - \theta_1$ and assume $0 < \varepsilon < |\theta|/\sqrt{I}$ and $R(\theta_1, \delta) \le \varepsilon^2$. Then*

$$R(\theta_2, \delta) \ge \left(|\theta| - \varepsilon\sqrt{I}\right)^2. \tag{2.3}$$

Hence, also,

$$R(\theta_2, \delta) \ge \theta^2\left(1 - \frac{2\varepsilon\sqrt{I}}{|\theta|}\right). \tag{2.4}$$

PROOF. By translation invariance we need only prove the theorem for $\theta_1 = 0$, $\theta_2 = \theta > 0$. We need to find δ which minimizes $R_\theta = \int(\delta(x) - \theta)^2 \times f_\theta(x)\mu(dx)$ subject to $R_0 = \int \delta^2(x) f_0(x)\mu(dx) \le \varepsilon^2$. By Lagrange multipliers the minimizing δ satisfies

$$2(\delta(x) - \theta) f_\theta(x) + 2\lambda\delta(x) f_0(x) = 0$$

for some $\lambda \ge 0$. If $\lambda = 0$, then $\delta(x) = \theta$ and it follows that $\theta^2 \le \varepsilon^2$, which contradicts the assumption that $\varepsilon \le \theta/\sqrt{I}$ since $I \ge 1$. Hence $\lambda > 0$ and

$$\delta(x) = \frac{\theta f_\theta(x)}{f_\theta(x) + \lambda f_0(x)} = \frac{\rho\theta q(x)}{1 + \rho q(x)}, \tag{2.5}$$

where $\rho = 1/\lambda$. [If $q(x) = \infty$, then $\delta(x) = \theta$.] Furthermore, this δ satisfies $R_0 = \varepsilon^2$, so that

$$\varepsilon^2 = \theta^2 \int \left(\frac{\rho q(x)}{1 + \rho q(x)} \right)^2 f_0(x)\mu(dx). \tag{2.6}$$

Then, by Cauchy-Schwarz,

$$\begin{aligned} \varepsilon\sqrt{I} &= \theta \left(\int \left(\frac{\rho q(x)}{1 + \rho q(x)} \right)^2 f_0(x)\mu(dx) \right)^{1/2} \\ &\quad \times \left(\int q^2(x) f_0(x)\mu(dx) \right)^{1/2} \\ &\ge \theta \int \left(\frac{\rho q(x)}{1 + \rho q(x)} \right) f_\theta(x)\mu(dx) \end{aligned} \tag{2.7}$$

since $q(x)f_0(x) = f_\theta(x)$. Hence

$$\begin{aligned} (\theta - \varepsilon\sqrt{I})^2 &\le \theta^2 \left(1 - \int \frac{\rho q(x)}{1 + \rho q(x)} f_\theta(x)\mu(dx) \right)^2 \\ &= \theta^2 \left(\int \frac{f_\theta(x)}{1 + \rho q(x)} \mu(dx) \right)^2 \\ &\le \theta^2 \int \left(\frac{1}{1 + \rho q(x)} \right)^2 f_\theta(x)\mu(dx) \\ &= R_\theta \end{aligned} \tag{2.8}$$

since $(\delta(x) - \theta)^2 = \theta^2/(1 + \rho q(x))^2$ by (2.5). [In the second inequality of (2.8) we have again used Cauchy–Schwarz.] This proves (2.3). Equation (2.4) follows because $(a - b)^2 \ge a^2(1 - 2b/a)$ for $a, b > 0$. □

It is of interest to note that the first bound in the theorem, (2.3), is sharp. As before, let $\theta_1 = 0$, $\theta = \theta_2 - \theta_1 > 0$.

PROPOSITION 1. *Fix* $B > 1$ *and* $0 < \varepsilon \le \theta/\sqrt{B}$. *Let* f_0 *be the uniform distribution on* $(0, 1)$ *and let* f_θ *be the uniform distribution on* $(0, B^{-1})$. *Then* $I = B$ *and*

$$\min\{R(\theta, \delta): R(0, \delta) \le \varepsilon^2\} = (\theta - \varepsilon\sqrt{I})^2. \tag{2.9}$$

Hence the bound (2.3) *is attained.*

PROOF. The best estimate of θ subject to $R(0, \delta) \le \varepsilon^2$ is

$$\delta(x) = \begin{cases} 0, & \text{if } B^{-1} \le x \le 1, \\ \varepsilon\sqrt{B}, & \text{if } 0 < x < b^{-1}. \end{cases}$$

[(2.5) shows that the best δ is constant on the region $0 < x < B^{-1}$ and 0 on the region $B^{-1} < x < 1$. Then, $R(0, \delta) = \varepsilon^2$ yields (2.10).] This estimator satisfies $R(\theta, \delta) = (\theta - \varepsilon\sqrt{I})^2$ and verifies (2.9) in view of Theorem 1. □

REMARKS. It can be seen that apart from arbitrary measurable transformations of the sample space, the example in the preceding proof is the only one in which the minimum in (2.9) is the same as (2.3). In other words, the bound in (2.3) is attained if and only if the likelihood ratio, $q(x)$, takes only the two values 0 and I.

In many applications f_{θ_1} and f_{θ_2} will be densities of independent identically distributed random variables $X_1, \dots, X_n$, each with density $f_{\theta_1}^{(1)}$. Letting $q^{(1)}(x) = f_{\theta_2}^{(1)}(x)/f_{\theta_1}^{(1)}(x)$ and $I^{(1)} = E_{\theta_1}((q^{(1)}(X))^2)$, we see that the information measure I satisfies

$$I = (I^{(1)})^n. \tag{2.10}$$

Furthermore, the bound in Theorem 1 is still sharp because letting f_{θ_1} be uniform on $(0, 1)$ and f_{θ_2} be uniform on $(0, I^{-1/n})$ yields

$$\min\{R(\theta, \delta): R(0, \delta) \le \varepsilon^2\} = (\theta - \varepsilon\sqrt{I})^2$$

for $\varepsilon\sqrt{I} < \theta$, as in Proposition 1.

3. Applications.

A. Superefficiency in the normal location model. As a first and simple example, the inequality given in the last section can be used to yield the following result about superefficient estimates in the standard normal location model.

Let $X_1, \dots, X_n$ be i.i.d. $N(\theta, 1)$ with $\theta \in \Theta_n$. Write $R(\theta, \delta_n)$ for the risk of an estimator δ_n [based on $(X_1, \dots, X_n)$]. Thus

$$R(\theta, \delta_n) = E(\theta - \delta_n)^2. \tag{3.1}$$

THEOREM 2. *Let* $\varepsilon_n \to \infty$ *and let* $(\ln \varepsilon_n / n)^{1/2} \in \Theta_n$. *If*

$$\limsup_{n\to\infty} n\varepsilon_n R(0, \delta_n) < \infty, \tag{3.2}$$

then

$$\limsup_{n\to\infty}\left(\frac{n}{\ln \varepsilon_n}\right)\sup_{\theta\in\Theta_n} R(\theta, \delta_n) > 0. \tag{3.3}$$

REMARK. (3.3) with $\ln \varepsilon_n$ deleted, $\Theta_n = (-1, 1)$ and the right-hand side replaced by ∞ is of course implied by (3.3) and is well known. See, for example, pages 407–408 in Lehmann (1983).

PROOF OF THEOREM 2. Since $\bar{X}$ is sufficient, we need only consider estimates δ_n which are functions of $\bar{X}$. Now if (3.2) holds, then there exists N and $M < \infty$ such that for all $n \geq N$,

$$R(0, \delta_n) \leq \frac{M}{n\varepsilon_n}.$$

Note that $\bar{X} \sim N(\theta, 1/n)$ and we may apply Theorem 1 with $f_{\theta,\sigma}$, the density of a normal distribution with mean θ and variance $\sigma^2 = 1/n$. Take $\theta_2 = \theta = (\ln \varepsilon_n / n)^{1/2}$ and $\theta_1 = 0$. Then

$$I = \int_{-\infty}^{\infty} \frac{f_{\theta,\sigma}^2(x)}{f_{0,\sigma}}\, dx = \exp\left(\frac{\theta^2}{\sigma^2}\right) = \exp\left(\frac{(\ln \varepsilon_n)/n}{1/n}\right) = \varepsilon_n. \tag{3.4}$$

Now by (2.4), if $n > N$,

$$\begin{aligned} R\left(\left(\frac{\ln \varepsilon_n}{n}\right)^{1/2}, \delta\right) &\geq \frac{\ln \varepsilon_n}{n}\left(1 - \frac{2(M/n\varepsilon_n)^{1/2}\varepsilon_n^{1/2}}{(\ln \varepsilon_n/n)^{1/2}}\right) \\ &= \frac{\ln \varepsilon_n}{n}\left(1 - \frac{2M^{1/2}}{(\ln \varepsilon_n)^{1/2}}\right). \end{aligned} \tag{3.5}$$

The theorem is proved by taking limits since $2M^{1/2}/(\ln \varepsilon_n)^{1/2} \to 0$ as $n \to \infty$. □

B. Superefficiency in the white noise model. We now turn to our main application of Theorem 1. Consider the following white noise model

$$dX_t = f(t)\, dt + \frac{1}{\sqrt{n}}\, dW_t, \qquad -\frac{1}{2} \leq t \leq \frac{1}{2}, \qquad f \in \mathscr{F}, \tag{3.6}$$

where W_t is Brownian motion on $[-1/2, 1/2]$.

This model has been studied extensively as a prototypical model for many other functional estimation problems such as nonparametric regression and density estimation. See, for example, Ibragimov and Hasminskii (1981), Low (1992), Donoho and Low (1992), Brown and Low (1996) and Nussbaum (1996).

We shall focus on the following class of parameter spaces. Write $f^{(k)}(x)$ for the kth derivative of f and let

$$\mathscr{F}(k) = \{f \in L^2(-1/2, 1/2): |f^{(k)}(x)| \le M \ \forall x\}. \tag{3.7}$$

Minimax rates of convergence, as $n \to \infty$, for estimating $f(0)$ are well known. For estimators δ_n based on the signal (3.6), write $E_f(f(0) - \delta_n)^2$ for the mean squared error for estimating $f(0)$. Then

$$0 < \limsup_{n\to\infty} n^{2k/(2k+1)} \inf_{\delta_n} \sup_{f\in\mathscr{F}(k)} E_f\big(f(0) - \delta_n\big)^2 < \infty. \tag{3.8}$$

THEOREM 3. *Let $\varepsilon_n \nearrow \infty$, $n/\ln \varepsilon_n \nearrow \infty$ and let δ_n be estimators based on* (3.6). *If*

$$\sup_x |f_0^{(k)}(x)| = m < M \tag{3.9}$$

and

$$\limsup_{n\to\infty} n^{2k/(2k+1)} \varepsilon_n E_{f_0}\big(f_0(0) - \delta_n\big)^2 < \infty, \tag{3.10}$$

then

$$\liminf_{n\to\infty} \left(\frac{n}{\ln \varepsilon_n}\right)^{2k/(2k+1)} \sup_{f\in\mathscr{F}(k)} E_f\big(f(0) - \delta_n\big)^2 > 0. \tag{3.11}$$

[As noted in the introduction, a similar statement, but with a different proof, appears in Lepskii (1990).]

PROOF. Let $g: R \to R$ be a k times differentiable function such that:

(i) $g(0) > 0$;
(ii) g has compact support;
(iii) for some $M > m$, $|g^{(k)}(x)| \le M - m \ \forall x$;
(iv) $\int_{-\infty}^{\infty} g^2(x)\,dx = 1$.

Such a function is easy to construct. Set

$$\alpha_n = \left(\frac{n}{\ln \varepsilon_n}\right)^{k/(2k+1)}, \tag{3.12}$$

$$\beta_n = \left(\frac{n}{\ln \varepsilon_n}\right)^{1/(2k+1)}. \tag{3.13}$$

Then

$$\alpha_n^2 \beta_n = \frac{n}{\ln \varepsilon_n} \tag{3.14}$$

and

$$\frac{\beta_n^k}{\alpha_n} = 1. \tag{3.15}$$

Let $f_n: [-1/2, 1/2] \to R$ be defined by

$$f_n(t) = f_0(t) + \frac{g(\beta_n t)}{\alpha_n}. \tag{3.16}$$

By (3.11), assumption (iii) for g and (3.15), $f_n \in \mathscr{F}(k)\ \forall n$. Write P_g^n for the probability measure associated with the process

$$dX_t = f_n(t) + \frac{1}{\sqrt{n}}\, dW_t, \qquad -\frac{1}{2} \le t \le \frac{1}{2}.$$

Then a sufficient statistic for the family of measures $\{P_0^n, P_g^n\}$ is given by

$$T_n = \ln \frac{dP_g^n}{dP_0^n}. \tag{3.17}$$

Set $\gamma_n = n\int_{-1/2}^{1/2}(g^2(\beta_n(t))/\alpha_n^2\, dt$. Then

$$\text{under } P_0^n, \qquad T_n \sim N\left(-\frac{\gamma_n}{2}, \gamma_n\right),$$

$$\text{under } P_g^n, \qquad T_n \sim N\left(\frac{\gamma_n}{2}, \gamma_n\right).$$

Now since $\beta_n \nearrow \infty$, there exists an N_1 such that for all $n \ge N_1$, $g(\beta_n t) = 0$ if $|t| > 1/2$. Then by assumption (iv) for g and (3.14) it follows that for $n \ge N_1$, $\gamma_n = \ln \varepsilon_n$.

If assumption (3.10) of Theorem 3 holds, there exists an $R < \infty$ and $N_2 > N_1$ such that for all $n \ge N_2$,

$$E_{f_0}(f(0) - \delta_n)^2 \le \frac{R}{n^{2k/(2k+1)}\varepsilon_n}. \tag{3.18}$$

Since T_n is sufficient for $\{P_0^n, P_g^n\}$ we may apply Theorem 1 with $\theta_1 = f_0(0)$, $\theta_2 = f_n(0) = f_0(0) + (g(0)/\alpha_n)$, p_{θ_1} the density of T_n under P_0^n and p_{θ_2} the density of T_n under P_g^n, where we have used p_φ instead of f_φ for the densities in Theorem 1. By (3.4),

$$I(\theta) = \int \frac{p_{\theta_2}^2(x)}{p_{\theta_1}(x)}\, dx = \exp\left(\frac{\gamma_n^2}{\gamma_n}\right) = \exp(\gamma_n) = \varepsilon_n$$

for $n \ge N_2$.

Theorem 1, (2.4), yields for $n \ge N_2$,

$$E_{f_n}(f_n(0) - \delta_n)^2 \ge \frac{g^2(0)}{\alpha_n^2}(1 - \lambda_n), \tag{3.19}$$

where

$$\lambda_n = \frac{2\varepsilon_n^{1/2}R^{1/2}}{(g(0)/\alpha_n)n^{k/(2k+1)}\varepsilon_n^{1/2}} = \frac{2R^{1/2}}{g(0)(\ln \varepsilon_n)^{k/(2k+1)}}. \tag{3.20}$$

Then $\lambda_n \to 0$ as $n \to \infty$ and hence by (3.12) and (3.19) it follows that

$$\liminf_{n\to\infty}\left(\frac{n}{\ln \varepsilon_n}\right)^{2k/(2k+1)} E_{f_n}(f_n(0)-\delta_n)^2 \ge g^2(0) > 0$$

and the theorem is proved. □

REMARK. One may also also consider Sobolev parameter spaces in place of the Lipschitz spaces defined in (3.7). Thus, suppose the parameter space is

$$\mathscr{F}_S(k) = \left\{ f\in L^2(-1/2,1/2): f^{(j)}(-1/2) = f^{(j)}(1/2) = 0\ \forall j = 0,\dots,k-1, \right.$$
$$\left. \int_{-1/2}^{1/2} \left(f^{(k)}(x)\right)^2 dx \le M^2 \right\}.$$

In this case the optimal rate for estimating $f(0)$ is $n^{(2k-1)/2k}$ in place of the rate $n^{2k/(2k+1)}$ of Theorem 3. See, for example, Donoho and Low (1992). *For this case, Theorem* 3 *remains valid with the obvious modification to* (3.9) *and with this optimal rate substituted into* (3.10) *and* (3.11). The proof is very similar to that of Theorem 3. The principal difference is the definitions of α_n, β_n and g in (3.16) in order to establish a suitably unfavorable two point problem, as is done above (3.18). These α_n, β_n, g are given via the hardest linear subproblem algorithm, and are explicitly described in Donoho and Low (1992). Note that here g is really g_n, that is, its form as well as its scaling depend on n.

4. Adaptation. Theorem 3 of the last section sheds new light on adaptation in the white noise model. For a sequence of estimates to be adaptive over two classes $\mathscr{F}(k_1)$ and $\mathscr{F}(k_2)$ given by (3.7), where k_1, k_2 are integers with $k_1 < k_2$, we would require

$$\limsup_{n\to\infty} n^{2k_1/(2k_1+1)} \sup_{f\in\mathscr{F}(k_1)} E_f(f(0)-\delta_n)^2 < \infty \tag{4.1}$$

and

$$\limsup_{n\to\infty} n^{2k_2/(2k_2+1)} \sup_{f\in\mathscr{F}(k_2)} E_f(f(0)-\delta_n)^2 < \infty. \tag{4.2}$$

In particular (4.2) must hold (with sup deleted) for some fixed $f_0 \in \mathscr{F}(k_2)$. Now since $\mathscr{F}(k_2) \subseteq \mathscr{F}(k_1)$, $f_0 \in \mathscr{F}(k_1)$, and it follows that

$$\limsup_{n\to\infty} n^{2k_1/(2k_1+1)} n^{(2(k_2-k_1))/((2k_2+1)(2k_1+1))} E_{f_0}(f_0(0)-\delta_n)^2 < \infty. \tag{4.3}$$

Theorem 3 then yields that

$$\limsup_{n\to\infty}\left(\frac{n}{\ln n}\right)^{2k_1/(2k_1+1)} \sup_{f\in\mathscr{F}(k)} E_f(f(0)-\delta_n)^2 > 0 \tag{4.4}$$

and it follows that (4.1) cannot hold if (4.2) holds. *Hence, adaptive estimation over any two classes* $\mathscr{F}(k_1)$ and $\mathscr{F}(k_2)$, $k_1 \neq k_2$, *is impossible.* (A similar conclusion is also valid for the Sobolev situation discussed in the preceding remark.)

In fact, a similar result can be proved under fairly general conditions using Theorem 1 and hardest one-dimensional subfamily arguments found in Donoho and Liu (1991).

In particular, suppose that T is a linear functional and Φ_1 and Φ_2 are convex and symmetric subsets of L_2 with optimal rates of convergence given by

$$0 < \liminf_{n \to \infty} n^{q_i} \inf_{\hat{T}_n} \sup_{\Phi_i} E\big(\hat{T}_n - T(\theta)\big)^2 < \infty,$$

where $0 < q_1 < q_2 < 1$.

Then it follows from essentially the same arguments used to prove the above that if

$$\liminf_{n \to \infty} n^{q_2} \sup_{\Phi_2} E\big(\hat{T}_n - T(\theta)\big)^2 < \infty,$$

then

$$\liminf_{n \to \infty} \left(\frac{n}{\ln n} \right)^{q_1} \sup_{\Phi_1} E\big(\hat{T}_n - T(\theta)\big)^2 > 0.$$

5. Nonparametric regression and density estimation models. In this section we show, with only minor modifications in the proofs, that Theorem 3 also holds in the nonparametric regression and density estimation models. This naturally yields analogs for the lack of adaptivity statement in Section 4.

The nonparametric regression model is given as follows. For each n, set $t_{in} = -\frac{1}{2} + (i/n)$, $i = 0, \dots, n$, and let

$$\begin{aligned} y_i &= f(t_{ni}) + e_i, \quad i = 1, \dots, n \\ f &\in \mathscr{F}(k), \qquad e_i \text{ i.i.d. } N(0,1), \end{aligned} \tag{5.1}$$

where $\mathscr{F}(k)$ is given by (3.7).

For estimates δ_n based on (5.1) the minimax rate of convergence is the same as in the white noise model and is given by (3.8). Moreover Theorem 3 also holds for estimators δ_n based on (5.1). The proof is essentially the same as in Section 3. Define g, c, α_n, β_n as before. Write P_g^n for the probability measure on $\mathbb{R}^n$ generating (5.1). Then $T_n = \ln(dP_g^n / dP_0^n)$ is sufficient for $\{P_0^n, P_g^n\}$. Set

$$\gamma_n' = \sum_{i=1}^{n} \frac{g^2(\beta_n t_{ni})}{\alpha_n^2}.$$

Then

$$\text{under } P_0^n, \qquad T_n \sim N\left(-\frac{\gamma_n'}{2}, \gamma_n'\right),$$

$$\text{under } P_g^n, \qquad T_n \sim N\left(\frac{\gamma_n'}{2}, \gamma_n'\right).$$

Note that for large n,

$$\gamma_n' \approx \gamma_n = n \int_{-1/2}^{1/2} \frac{g^2(\beta_n t)}{\alpha_n^2} \, dt$$

and the distributions of T_n are close to those given in Section 3. Now define $\bar{g}$ by

$$\bar{g}(\beta_n t) = g(\beta_n t_{ni}), \qquad t_{n,i-1} \le t \le t_{n,i}, \, i = 1, \dots, n.$$

Then

$$\gamma_n' = n \int \frac{\bar{g}^2(\beta_n t)}{\alpha_n^2} \, dt$$

and

$$\begin{aligned} |\gamma_n - \gamma_n'| &= \left| n \int_{-1/2}^{1/2} \frac{g^2(\beta_n t)}{\alpha_n^2} \, dt - n \int_{-1/2}^{1/2} \frac{\bar{g}^2(\beta_n t)}{\alpha_n^2} \, dt \right| \\ &\le \frac{n}{\alpha_n^2} \int_{-1/2}^{1/2} |g^2(\beta_n t) - \bar{g}^2(\beta_n t)| \, dt \\ &= \frac{n}{\alpha_n^2} \sum_{i=1}^{n} \int_{t_{n,i-1}}^{t_{n,i}} |g^2(\beta_n t) - \bar{g}^2(\beta_n t)| \, dt. \end{aligned}$$

Note that since $|g^k(x)| \le M$ and g has compact support, it follows that for some M_1, M_2,

$$|g(x)| \le M_1 \qquad \forall x, \tag{5.2}$$

$$|g'(x)| \le M_1 \qquad \forall x \tag{5.3}$$

and hence for $t_{n,i-1} \le t \le t_{ni}$,

$$|g^2(\beta_n t) - \bar{g}^2(\beta_n t)| \le 2 M_1 M_2 \beta_n |t - t_{n,i-1}| + M_2^2 \beta_n^2 (t - t_{n,i-1})^2$$

and

$$\begin{aligned} |\gamma_n - \gamma_n'| &\le \frac{n}{\alpha_n^2} \sum_{i=1}^{n} \int_{t_{n,i-1}}^{t_{n,i}} 2 M_1 M_2 \beta_n |t - t_{n,i-1}| + M_2^2 \beta_n^2 (t - t_{n,i-1})^2 \, dt \\ &= \frac{n}{\alpha_n^2} n \left[2 M_1 M_2 \beta_n \frac{1}{2} \frac{1}{n^2} + M_2^2 \beta_n^2 \frac{1}{3} \frac{1}{n^3} \right] \\ &= \frac{M_1 M_2 \beta_n}{\alpha_n^2} + \frac{M_2^2}{3} \frac{\beta_n^2}{\alpha_n^2} \frac{1}{n}. \end{aligned}$$

Now by (3.12) and (3.13),

$$\frac{\beta_n}{\alpha_n^2} \to 0 \quad \text{and} \quad \frac{\beta_n^2}{\alpha_n^2}\frac{1}{n} \to 0 \quad \text{as } n \to \infty$$

and hence $\gamma_n' = \gamma_n + \mu_n$, where $\mu_n \to 0$ as $n \to \infty$. The rest of the proof follows that given in Section 3 except that for large n,

$$I = \varepsilon_n \exp(\mu_n)$$

and (3.19) becomes

$$E_{f_n}(f_n(0) - \delta_n)^2 \geq \frac{g^2(0)}{\alpha_n^2}\left(1 - \lambda_n \exp\left(\frac{\mu_n}{2}\right)\right), \tag{5.4}$$

where $\exp(\mu_n/2) \to 1$ as $n \to \infty$.

We now turn to the density estimation problem. Let $X_1, \dots, X_n$ be i.i.d. each with density $F \in \mathscr{F}(k)$. Then once again Theorem 3 holds for estimators based on $X_1, \dots, X_n$. For simplicity we shall give a proof for the case where the point of superefficiency is the uniform density. In other words, $f_0(t) \equiv 1$ on $[-1/2, 1/2]$. Once again define g, c, α_n and β_n as in Section 3, but with the added restriction that $\int g(x)\,dx = 0$. To apply Theorem 1 we need to find upper bounds for

$$I = \int \frac{f_n^2(x_1) \cdots f_n^2(x_n)}{f_0(x_1) \cdots f_0(x_n)}\, dx_1 \cdots dx_n. \tag{5.5}$$

Now since $f_0(t) \equiv 1$ on $[-1/2, 1/2]$,

$$\begin{aligned} I &= \left[\int_{-1/2}^{1/2} \left(1 + \frac{g(\beta_n t)}{\alpha_n}\right)^2 dt\right]^n \\ &= \left(1 + \frac{1}{\alpha_n^2 \beta_n}\int g^2(t)\right)^n \\ &= \left(1 + \frac{\ln \varepsilon_n}{n}\right)^n \quad [\text{by } (3.14)] \\ &\leq \exp(\ln \varepsilon_n) \quad \left[\text{since if } x > 0 \left(1 + \frac{x}{n}\right)^n \leq e^x\right] \\ &= \varepsilon_n. \end{aligned} \tag{5.6}$$

Then as in Section 3 for large n,

$$E_{f_n}(f_n(0) - \delta_n)^2 \geq \frac{g^2(0)}{\alpha_n^2}(1 - \lambda_n),$$

where λ_n is defined by (3.20). Now take limits and the theorem follows just as in Section 3.

REFERENCES

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

BICKEL, P. J. (1984). Parametric robustness, small biases can be worthwhile. *Ann. Statist.* **12** 864–879.

BROWN, L. D. and LOW, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** 2384–2398.

DONOHO, D. L. and JOHNSTONE, I. M. (1992). Adapting to unknown smoothness at a point via wavelet shrinkage. Unpublished manuscript.

DONOHO, D. L. and JOHNSTONE, I. M. (1994). Neo-classical minimax problems, threshholding, and adaptive function estimation. *Bernoulli* **2** 39–62.

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

DONOHO, D. L. and LOW, M. G. (1992). Renormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.* **20** 944–970.

EFROMOVICH, S. Y. (1985). Nonparametric estimation of a density of unknown smoothness. *Theory Probab. Appl.* **30** 557–568.

EFROMOVICH, S. Y. and LOW, M. G. (1994). Adaptive estimates of linear functionals. *Probab. Theory Related Fields* **98** 261–275.

EFROMOVICH, S. Y. and PINSKER, M. S. (1984). Self-learning algorithm of nonparametric filtration. *Automat. i Telemeh.* **11** 58–65. (In Russian.)

FRANK, P. and KIEFER, J. (1951). Almost subminimax and biased minimax procedures. *Ann. Math. Statist.* **22** 1–4.

GOLUBEV, G. K. (1987). Adaptive asymptotically minimax estimates of smooth signals. *Problemi Peredachi Informatsii* **23** 57–67. (In Russian.)

HODGES, J. L. and LEHMANN, E. L. (1952). The use of previous experience in reaching statistical decisions. *Ann. Math. Statist.* **23** 396–407.

HODGES, J. L. (1952). Communicated to Le Cam and Neyman.

IBRAGIMOV, I. A. and HASMINSKII, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*. Springer, Berlin.

IBRAGIMOV, I. A. and HASMINSKII, R. Z. (1984). Nonparametric estimation of the values of a linear functional in Gaussian white noise. *Theory Probab. Appl.* **29** 19–32.

LE CAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *Univ. Calif. Publ. Statist.* **1** 277–330.

LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.

LEPSKII, O. V. (1990). On a problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** 454–466.

LEPSKII, O. V. (1991). Asymptotically minimax adaptive estimation I: upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.* **36** 682–697.

LEPSKII, O. V. (1992). On problems of adaptive estimation in white Gaussian noise. In *Advances in Soviet Mathematics* **12**. Amer. Math. Soc., Providence, RI.

LOW, M. G. (1992). Renormalization and white noise approximation for nonparametric functional estimation problems. *Ann. Statist.* **20** 545–554.

NUSSBAUM, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430.

ROBBINS, H. (1951). Asymptotically subminimax solutions of compound statistical decision problems. *Proc. Second Berkeley Symp. Math. Statist. Probab.* 131–148. Univ. California Press, Berkeley.

DEPARTMENT OF STATISTICS
3000 STEINBERG–DIETRICH HALL
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104-6302