The Annals of Statistics
1997, Vol. 25, No. 6, 2607–2625

SUPEREFFICIENCY IN NONPARAMETRIC FUNCTION ESTIMATION

BY LAWRENCE D. BROWN,[1] MARK G. LOW[2] AND LINDA H. ZHAO

University of Pennsylvania

Fixed parameter asymptotic statements are often used in the context of nonparametric curve estimation problems (e.g., nonparametric density or regression estimation). In this context several forms of superefficiency can occur. In contrast to what can happen in regular parametric problems, here *every* parameter point (e.g., unknown density or regression function) can be a point of superefficiency.

We begin with an example which shows how fixed parameter asymptotic statements have often appeared in the study of adaptive kernel estimators, and how superefficiency can occur in this context. We then carry out a more systematic study of such fixed parameter statements. It is shown in four general settings how the degree of superefficiency attainable depends on the structural assumptions in each case.

1. Introduction. Asymptotic analysis as the sample size $n \to \infty$ is an important tool for developing and analyzing statistical procedures. Asymptotic statements can generally be classified into one of two varieties: those in which the limit is taken as the (unknown) parameter is held fixed, and those in which the limit is uniform over all of the parameter space, or at least over some significant subset of it. Asymptotic minimax and local minimax statements are examples of the second variety.

Fixed parameter statements are generally easier to formulate and prove than are uniform ones. However, it has long been recognized that they can involve misleading conclusions via the phenomenon of "superefficiency" in regular parametric problems. Le Cam (1953) presents Hodges superefficient estimator and the first comprehensive treatment of this topic. Further study has solidified understanding of superefficiency in regular parametric problems and has made it relatively straightforward to avoid this pitfall. [See, e.g., Huber (1966), Weiss and Wolfowitz (1966) and Hájek (1972).]

Fixed parameter asymptotic statements are also commonly used in the context of nonparametric curve estimation (e.g., in nonparametric density or regression problems). In spite of this it has not been widely understood that superefficiency can occur in several forms in these problems and can be much more dramatic than in regular parametric contexts. In parametric problems

Received August 1995; revised March 1997.
[1]Supported in part by NSF Grant DMS-96-26118.
[2]Supported in part by NSA Grant MDA 904-96-1-0028.
AMS 1991 *subject classifications.* Primary 62G07; secondary 62G20, 62B15, 62M05.
Key words and phrases. Superefficiency, nonparametric function estimation, asymptotics.

the set of superefficiency has (Lebesgue) measure zero. By contrast we show that in nonparametric problems *every* parameter point can be a point of superefficiency. This means that considerable caution is needed before interpreting such fixed parameter asymptotic statements as recommendations of asymptotic desirability or optimality.

Abramson (1982) presents a nonparametric example showing that superefficiency can hold everywhere in the parameter space. Our Section 2 discusses the way in which fixed parameter asymptotics often appear in nonparametric density estimation and presents a simplified form of Abramson's superefficiency phenomenon. Section 3 provides some notation and background needed for a more general treatment of the question. Section 4 reviews what is known in the regular parametric case, in order to provide further background.

Sections 5 through 8 present examples of superefficiency. The situation in Section 5 has a character similar to that in Section 2, but it is also shown that the degree of superefficiency is somewhat limited in this context. (See Theorem 5.2, whose proof is deferred to the Appendix.) Section 6 presents a different estimator which is even more superefficient than that in Section 5. Furthermore, this estimator has acceptable minimax behavior. This nonparametric example thus contrasts with the regular parametric situation in which superefficient estimators must behave poorly with respect to uniform criteria such as local minimaxity.

Section 7 shows there are nonparametric problems in which superefficient estimators do not exist to the dramatic degree present in Sections 5 and 6. Section 8, on the other hand, exhibits an even more extreme degree of superefficiency than that displayed in the earlier examples.

2. A generic example: adaptive estimation of a density at a point. The first aim of this example is to focus attention on fixed parameter asymptotics by sketching how they have often appeared in the literature. The example then continues with a further development which suggests that these asymptotic statements can be misleading.

The following presentation is modelled on that in Woodroofe (1970), which contains the earliest adaptive result of this type. Since then a number of other structurally similar treatments have appeared. We will later examine more carefully one by Abramson (1982). For other instances see, for example, Hall (1993) and for the analogous nonparametric regression problem see, for example, Brockmann, Gasser and Herrman (1993) or Schucany (1995).

Let $X_1, \dots, X_n$ be a sample of size n from a population with density f. About f assume basically only that $f \in C^r_+$, the set of all densities with r continuous derivatives and with $f(0) > 0$, $f^{(r)}(0) \neq 0$. (Additional technical assumptions on f may be required in some treatments.) The goal is to estimate $f(0)$ under normalized squared error loss, $L_n(f, \hat{f}) = n^{2r/(2r+1)}(\hat{f} - f(0))^2$, where $\hat{f}$ denotes the estimated value of $f(0)$. Let $\mathscr{H}^s$ denote the family of sth-order kernels with support in $[-1, 1]$; that is, those which satisfy

$$\int_{-1}^{1} x^j K(x)\,dx \begin{cases} = 1, & \text{if } j = 0, \\ = 0, & \text{if } 1 \le j \le s - 1, \\ \ne 0, & \text{if } j \ge s, \end{cases}$$

and $K(x) = 0$ if $|x| > 1$.

The kernel estimator with bandwidth b_n is defined as

$$\hat{f} = \frac{1}{b_n} \sum_{i=1}^{n} K\left(\frac{X_i}{b_n}\right). \tag{2.1}$$

Let $R_n(f, \hat{f}) = E_f(L_n(f, \hat{f}))$. Then R_n can be decomposed as a sum of variance and squared-bias terms. *If one further assumes* that $s = r$, then this decomposition yields that

$$0 < R_{\text{opt}} \doteq \inf_{b_n} \lim_{n \to \infty} R_n(f, \hat{f}) < \infty \tag{2.2}$$

is attained by

$$b_{\text{opt}} = \left[k_K \frac{f(0)}{n f^{(r)}(0)} \right]^{1/(2r+1)} \tag{2.3}$$

for an appropriate constant k_K depending only on K (and r).

For "adaptive estimation" one often constructs "pilot" estimates $\hat{f}$ and $\hat{f}^{(r)}$ of $f(0)$ and $f^{(r)}(0)$. These are then plugged into (2.3). The resulting adaptive bandwidth $\tilde{b}$, say, is substituted in (2.1) to get an estimator $\tilde{f}$. A statement of the following type is frequently cited as a major support in such a study.

Statement. With a suitable (and, often, explicitly given) definition of $\hat{f}$ and $\hat{f}^{(r)}$

$$E_f\left(\left(\frac{\tilde{f}}{\tilde{f}^{(r)}} - \frac{f(0)}{f^{(r)}(0)}\right)^2\right) \to 0 \qquad \forall f \in C_+^r, \tag{2.4}$$

and (possibly) under additional conditions on f,

$$\lim_{n \to \infty} \frac{R_n(f, \tilde{f})}{R_{\text{opt}}(f)} = 1 \qquad \forall f \in C_+^r. \tag{2.5}$$

Expression (2.5) is an example of fixed parameter asymptotics. It is a statement in which the limit is computed for fixed values of f. This contrasts with uniform limits, like those in the minimax statements (3.3) in the next section.

The statement (2.5) looks like a strong endorsement of $\tilde{f}$. Indeed, it has undoubtedly been interpreted by some as claiming that $\tilde{f}$ is asymptotically optimal.

However, it is not hard to construct an estimator which does much better than $\tilde{f}$ in the sense of (2.5). Let $K \in \mathscr{K}^s$ be any bounded kernel, with $s > r$. Let

$$\check{b}_\gamma = \left[k_K \frac{\hat{f}}{n\gamma \hat{f}^{(r)}} \right]^{1/(2r+1)}$$

and let $\check{f}_\gamma$ denote the corresponding estimator.

THEOREM 2.1. *Let* $1 > \varepsilon > 0$. *Assume* (2.4) *holds. Then there is* $\gamma > 0$ *such that*

$$\lim_{n\to\infty} \frac{R_n(f, \check{f}_\gamma)}{R_{\text{opt}}} < \varepsilon < 1 = \lim_{n\to\infty} \frac{R_n(f, \tilde{f})}{R_{\text{opt}}} \qquad \forall f \in C_+^r. \tag{2.6}$$

SKETCH OF PROOF. Since $f \in C_+^{(r)}$ it can be written as

$$f(x) = f(0) + P_r(x) + o(x^r),$$

where $P_r(x)$ is a polynomial of degree r with no constant term. From this and (2.4) it is easy to see that the asymptotic bias B_∞ is

$$\begin{aligned} B_\infty &\doteq \lim_{n\to\infty} \left\{ n^{r/(2r+1)} \left(E\left(\check{f}_\gamma\right) - f(0) \right) \right\} \\ &= o(1) \end{aligned}$$

since $K \in \mathscr{K}^s$ with $s > r$.

Calculation and comparison of the asymptotic variance terms reveals that for γ sufficiently small the asymptotic variance of $\check{f}_\gamma$ is less than ε times that of $\hat{f}$, since its bandwidth is larger by the factor $(1/\gamma)^{1/(2r+1)} + o_p(1)$. Statement (2.4) supplies an additional bound needed for the formal argument. An explicit formula for suitable γ is feasible in terms of r, K; however, we omit this formula since we do not want to suggest that $\check{f}_\gamma$ is a valuable practical alternative to $\tilde{f}$. □

One presumably could conclude from (2.6) as compared to (2.5) that when γ is small and n is sufficiently large $\check{f}_\gamma$ is a more desirable estimator than $\tilde{f}$. However, we suggest the appropriate conclusion is that fixed parameter asymptotics may be misleading, and that (2.5) itself is not a reliable argument that $\tilde{f}$ is desirable for sufficiently large n. [Correspondingly, (2.6) is also not convincing evidence that $\check{f}_\gamma$ is preferable to $\tilde{f}$ as a statistical procedure for large n.]

We remark that there are other ways to produce an estimator with the property in (2.6). In particular, for the case $r = 2$, Abramson (1982) began with an arbitrary, bounded kernel $K \in \mathscr{K}^2$. He then defined the estimator

$$\check{f}_{A,n} = \frac{1}{n} \sum_{i=1}^{n} \frac{f^{1/2}(X_i)}{\check{b}_\gamma} K\left(\frac{f^{1/2}(X_i)}{\check{b}_\gamma} X_i \right),$$

and showed that $\check{f}_{A,n}$ also satisfies (2.6) for suitable $\hat{f}, \hat{f}^{(r)}$ in the definition of $\check{b}_\gamma$. An interesting feature of $\check{f}_{A,n}$ is that it can be interpreted as an adaptive version of a procedure involving a $\mathscr{K}^2$ kernel, and thus may appeal to those who allow adaptation but resist the use of kernels which are not nonnegative.

The remainder of this paper is devoted to an investigation of the forms of behavior (and misbehavior) of fixed parameter limits. We hope that the additional understanding this provides, along with (much) further research, will lead to a better understanding of such limiting statements. Perhaps there is some simple supplementary information [other than a minimax limit such as (3.3)] which might convert them into reliable measurements of the statistical value of procedures for sufficiently large n.

3. A canonical model. As a general setting for further study consider a canonical model in which one observes $\{Z_i^{(n)}: i = 1, \dots\}$ with $Z_i^{(n)} \sim N(\theta_i, 1/n)$, independent, $i = 1, \dots$. The index n plays the role of sample size. Denote the "nonparametric" parameter space by Θ. Let $\|\cdot\|_*$ be a nonnegative functional —usually a norm—on Θ with the following properties. Let $\Theta(B) = \{\theta \in \Theta: \|\theta\|_* \le B\}$ and then require that $B_1 < B_2$ implies $\Theta_{B_1} \subset \Theta_{B_2}$, and also that $\Theta = \bigcup_{B<\infty}\Theta(B)$.

In Sections 5 and 6, Θ will be the Sobolev space

$$\Theta = S_r = \left\{\theta = \{\theta_i\}: \|\theta\|_*{}^2 = \sum i^{2r}\theta_i^2 < \infty\right\} \tag{3.1}$$

and then

$$\Theta(B) = \left\{\theta: \|\theta\|_*{}^2 \le B^2\right\}. \tag{3.2}$$

For notational convenience let $\Theta(\infty) = \Theta = \bigcup_{B<\infty}\Theta(B)$. Let $r > \frac{1}{2}$ to avoid technical complications in part of what follows. When necessary to display the dependence on r we will write $\Theta_r(B)$.

The goal to be pursued in Sections 5 and 6 is to estimate θ under normalized squared-error loss

$$\begin{aligned} L_n(\theta, \tilde{\theta}) &= n^{2r/(2r+1)} \sum \left(\tilde{\theta}_i - \theta_i\right)^2 \\ &\doteq n^{2r/(2r+1)} \|\theta - \theta\|^2. \end{aligned}$$

Let $R_n(\theta, \tilde{\theta})$ denote the corresponding risk function.

Remark concerning other models. Note that this model is equivalent to observing the white-noise model $dZ(t) = f(t)\,dt + (1/\sqrt{n})\,dB(t)$, $t \in (0,1)$, where $B(t)$ denotes ordinary Brownian motion, with parameter space $\{f: \int_0^1 [f^{(r)}(t)]^2\,dt < \infty\}$ and with loss $n^{2r/(2r+1)}\int(\tilde{f}(t) - f(t))^2\,dt$. The model is also asymptotically equivalent to similar Sobolev versions in nonparametric regression and in density estimation. [See Donoho and Liu (1991a, b), Brown and Low (1996a) and Nussbaum (1996).] Consequently, there are direct analogues in those nonparametric settings to the statements in Sections 5–8 concerning the canonical model. We presume that there are also analogues in other, less standard nonparametric estimation problems.

In Section 2 the asymptotic standard R_{opt} was defined via optimal kernel estimation with a particular rth-order kernel. For the problems to follow it is better to use a differently defined, more general standard. For any $B < \infty$

define the minimax value

$$M(B) = \limsup_{n\to\infty} \inf_{\{\delta_n\}} \sup_{\theta\in\Theta(B)} R_n(\theta, \delta_n). \tag{3.3}$$

This type of limiting statement is uniform [over $\Theta(B)$], as opposed to the fixed parameter statements seen in Section 2 and again in (3.5) and later. Then denote

$$m(\theta) = M(\|\theta\|_*) \tag{3.4}$$

and investigate whether there exist procedures δ_n such that for some $0 \le l < u \le \infty$

$$\limsup_{n\to\infty} \frac{R_n(\theta, \delta_n)}{m(\theta)} < 1 \qquad \forall\, \theta \ni l < \|\theta\|_* < u. \tag{3.5}$$

[Typically, $\|\cdot\|_*$ will have been chosen so that $\|\theta\|_* \neq 0$ implies $m(\theta) > 0$.] An estimator which satisfies (3.5) will be called superefficient on $\{\theta\colon l < \|\theta\|_* < u\}$.

4. Comments on standard, parametric models. Superefficiency is a well-studied anomaly which occurs when using fixed parameter asymptotics. To see how this relates to the formulation in Section 3 and the results in succeeding sections, consider briefly a smooth, regular statistical problem with an open parameter space $\Theta \subset \mathbf{R}^p$. For this finite dimensional problem let $L_n(\theta, \tilde\theta) = n\|\tilde\theta - \theta\|^2$ and let R_n denote the corresponding risk function. Then a uniform (minimax) limiting process yields

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} \sup_{\|\theta' - \theta\| < \varepsilon} R_n(\theta', \delta_n) \ge \frac{1}{\operatorname{tr}(I(\theta))} \doteq m(\theta), \tag{4.1}$$

say, for all $\{\delta_n\}$, where $0 < I(\theta) < \infty$ denotes the usual Fisher information matrix. Here $m(\theta)$ is the quantity defined in (3.4) relative to the present risk function, and with $\|\cdot\|_* = \|\cdot\|$ on R^p. [Lehmann (1983) is a good general reference for this result though statements like this of course appear much earlier in the literature, such as in Le Cam (1953) and Hájek (1972).]

It is well known that there exist "superefficient" estimators $\{\delta_n\}$ such that the fixed parameter limit satisfies

$$\limsup_{n\to\infty} \frac{R_n(\theta, \delta_n)}{m(\theta)} \le 1 \tag{4.2}$$

with strict inequality for some $\theta \in \Theta$. However, it is also well known that the set on which inequality holds in (4.2) must have measure 0. Also, when $p = 1$, δ_n must behave poorly somewhere in a neighborhood of any point of superefficiency. [See Brown and Low (1996b) for a recently obtained result of this type.] It follows that in these parametric models

$$\sup_{l<\|\theta\|<w} \limsup_{n\to\infty} \frac{R_n(\theta, \delta_n)}{m(\theta)} \ge 1 \tag{4.3}$$

since the set of superefficiency has measure zero. Thus the answer here to the question at (3.5) is always negative, and there cannot exist estimators $\{\delta_n\}$

which exhibit superefficiencies like those to be displayed in Sections 5, 6 and 8.

5. Global superefficiency: linear estimation over Sobolev parameter space. Consider the canonical Sobolev setting described in Section 3. The situation is very similar to that described in Section 2, as follows.

THEOREM 5.1. *Let $\varepsilon > 0$. Then there is a sequence of linear estimators $\{\delta_n\}$ such that*

$$\sup_{\|\theta\|_* \geq \varepsilon} \limsup_{n\to\infty} \frac{R_n(\theta, \delta_n)}{m(\theta)} \leq \varepsilon. \tag{5.1}$$

PROOF. The minimax value, defined in (3.3), is

$$M(B) = C_r B^{2/(2r+1)} \tag{5.2}$$

for an appropriate $C_r > 0$. See, for example, Efromovich and Pinsker (1984) or Donoho, Liu and MacGibbon (1990). Let $\alpha > 0$, $\beta(n) = n^{1/(2r-1)}$ and $\gamma^{-1}(n) = n^{2r/(2r+1)}$, and define the linear estimator via truncation as

$$(\delta_n)_i = \begin{cases} z_i, & \text{if } i \leq [\alpha\beta(n)], \\ 0, & \text{if } i > [\alpha\beta(n)]. \end{cases} \tag{5.3}$$

Then the usual variance plus bias-squared decomposition yields

$$\begin{aligned} \gamma^{-1}(n) E_\theta(\|\delta - \theta\|^2) &= \frac{[\alpha\beta(n)]}{\beta(n)} + \gamma^{-1}(n) \sum_{[\alpha\beta(n)]+1}^{\infty} \theta_i^2 \\ &\leq \alpha + \frac{\gamma^{-1}(n)}{(\alpha\beta(n))^{2r}} \sum_{[\alpha\beta(n)]+1} i^{2r}\theta_i^2 \\ &\to \alpha \end{aligned} \tag{5.4}$$

since $\gamma^{-1}(n)/(\beta(n))^{2r} = 1$ and since $\sum_{[\alpha\beta(n)]+1} i^{2r}\theta_i^2 \to 0$ as $n \to \infty$, for any fixed $\alpha > 0$. The choice $\alpha = C_r \varepsilon^{2/(2r+1)}$ then yields (5.1), since $m(\theta) \geq M(\|\varepsilon\|)$ for $\|\theta\|_* \geq \varepsilon$. □

Here is a converse to Theorem 5.1 which shows that (5.1) is the strongest possible statement.

THEOREM 5.2. *Let $\{\delta_n\}$ be any sequence of linear estimators. Then, for every $0 \leq l < u \leq \infty$,*

$$\sup_{l<\|\theta\|_* < u} \limsup_{n\to\infty} \frac{R_n(\theta, \delta_n)}{m(\theta)} > 0. \tag{5.5}$$

PROOF. The proof begins by assuming $\{\delta_n\}$ is a sequence for which (5.5) is false and then constructing a particular θ for which

$$\limsup_{n\to\infty} \frac{R_n(\theta, \delta_n)}{m(\theta)} > 0,$$

a contradiction. Details of the argument appear in the Appendix. □

6. Global superefficiency: general estimation over Sobolev space. Estimators in the previous section were required to be linear. If that restriction is lifted, then a somewhat stronger form of global superefficiency can occur, as demonstrated in the following theorem.

THEOREM 6.1. *For the Sobolev setting of Section* 3 *there is a sequence of estimators* $\{\delta_n\}$ *such that*

$$\lim_{n\to\infty} \frac{R_n(\theta, \delta_n)}{m(\theta)} = 0 \tag{6.1}$$

for all $\theta \neq 0$.

PROOF. We will explicitly construct an adaptive sequence of estimators with the desired property. We use this particular sequence because it is easy to define and its direct explicit definition leads to a simple verification of (6.1). However, we note that many other adaptive estimators would also yield (6.1). In particular, a referee has correctly pointed out that we could instead have used the original Efromovich and Pinsker (1984) adaptive estimator to obtain (6.1) in a natural fashion.

Our estimator is based on a suggestion made to us by D. Donoho in a different but related context. There may also be some connection of the following to ideas in Stein (1966).

Partition $z = (z_1, \dots)$ as $y = (y_{(1)}, \dots)$, where $y_{(j)} = (z_{2^{j-1}}, \dots, z_{2^j-1})$; $\theta = (\theta_1, \dots)$ can be similarly partitioned as $\tilde{\theta} = (\theta_{(1)}, \dots)$ so that $\theta_{(j)} = E(Y_{(j)})$. Write $\|y - \tilde{\theta}\|^2 = \|z - \theta\|^2$ in the natural manner. Note that $y_{(j)}$ is 2^{j-1}-dimensional. Let

$$\delta_{n(j)}(y_{(j)}) = \left(1 - \frac{(3/2)(2^{j-1} - 2)^+}{n\|y_{(j)}\|^2}\right)^+ y_{(j)}.$$

Then $\delta_n = (\delta_{n(1)}, \dots)$ is an estimate of $\tilde{\theta}$ and thus yields an estimate of θ in the obvious manner. [Note that the customary positive-part James–Stein formula would use 1 in place of (3/2) in the above expression.] We then need the following lemma.

LEMMA 6.1. *Let* $Y \sim N_p(\mu, \sigma^2)$, $\mu \in \mathbf{R}^p$, *be a p-dimensional normal variable with* $p \geq 3$. *There is a* $\zeta > 0$ (*independent of* $p \geq 3$) *such that*

$$\delta(y) = \left(1 - \frac{(3/2)(p-2)}{\|y\|^2/\sigma^2}\right)^+ y \tag{6.2}$$

satisfies

$$E_\mu(\|\delta-\mu\|^2) \le (\|\mu\|^2 \wedge \sigma^2 p) + \sigma^2\zeta^{-1}e^{-\zeta p}. \tag{6.3}$$

PROOF. It suffices to consider the case $\sigma^2 = 1$. Then, by Stein's unbiased estimate of risk [see, e.g., Berger (1985), page 361],

$$E_\mu(\|\delta-\mu\|^2) \le E_\mu\Big((\|Y\|^2-p)I_{\|y\|^2\le 3(p-2)/2}(Y) + pI_{\|y\|^2>3(p-2)/2}(Y)\Big).$$

The expectation on the right-hand is bounded by p. Hence (as is well known) $E_\mu(\|\delta-\mu\|^2) \le p$.

It remains only to consider the case $\|\mu\|^2 \le p$. In that case standard large deviation results [see, e.g., Brown (1986), page 211] yield the existence of an $\varepsilon > 0$ such that

$$P_\mu\big(\{\|Y\|^2 > 3(p-2)/2\}\big) < e^{-\varepsilon p} \qquad \forall\, p \ge 3.$$

[Here are some details of this result. Note that the probability is maximized when $\|\mu\|^2 = p$, and then $\|y\|^2$ has the same distribution as $\sum_{i=1}^p z_i^2$, where $z_i \sim N(1,1)$, independent. Write $w_i = z_i^2 - 1$ and consider the exponential family generated by the distribution of W. This is a regular exponential family. Hence, for $p \ge 9$, $P(p^{-1}\sum_{i=1}^p W_i > 1/6 \ge p^{-1}(3(p-2)/2) - 1) < e^{-\varepsilon' p}$ for suitable $\varepsilon' > 0$. Finally, for $3 \le p \le 8$, $P_{\|\mu\|^2=p}(\{\|Y\|^2 > 3(p-2)/2\}) < 1$.] Also

$$\begin{aligned}E_\mu\big((\|Y\|^2-p)I_{\|y\|^2\le 3(p-2)/2}\big) &< E_\mu(\|Y\|^2-p)P_\mu\big(\{\|Y\|^2 \le 3(p-2)/2\}\big)\\ &< E_\mu(\|Y\|^2-p) - \|\mu\|^2.\end{aligned}$$

Hence

$$\begin{aligned}E_\mu(\|\delta-\mu\|^2) &\le (\|\mu\|^2 + pe^{-\varepsilon p})I_{\|\mu\|^2<p}(\mu) + pI_{\|\mu\|^2\ge p}(\mu)\\ &\le (\|\mu\|^2 \wedge p) + \zeta^{-1}e^{-\zeta p}\end{aligned}$$

for $\zeta = \varepsilon/2$. □

Then, by Lemma 6.1 (with β, γ as in Theorem 5.1),

$$\begin{aligned}R_n(\theta,\delta_n) &= \gamma^{-1}(n)\sum_{j=1}^\infty E(\|\delta_{n(j)} - \theta_{(j)}\|^2)\\ &\le \gamma^{-1}(n)\left\{\frac{3}{n} + \sum_{j=3}^\infty\left[\left(\|\theta_{(j)}\|^2 \wedge \frac{2^{j-1}}{n}\right) + \frac{\zeta^{-1}\exp(-2^{j-1}\zeta)}{n}\right]\right\}\\ &\le \gamma^{-1}(n)\left\{\sum_{j=1}^{l}\frac{2^{j-1}}{n} + \sum_{l+1}^\infty\|\theta_{(j)}\|^2 + O\left(\frac{1}{n}\right)\right\}\\ &\le \gamma^{-1}(n)\left\{\frac{2^l-1}{n} + \sum_{i=2^l}^\infty\theta_i^2 + O\left(\frac{1}{n}\right)\right\}\end{aligned} \tag{6.4}$$

for any integer $l = l(n)$. Now, choose $\varepsilon > 0$ and choose l so that $\varepsilon\beta(n) \le 2^l - 1 \le 2\varepsilon\beta(n)$. Then

$$\begin{aligned} R_n(\theta, \delta_n) &\le \gamma^{-1}(n)\left\{\frac{2\varepsilon\beta(n)}{n} + \sum_{\varepsilon\beta(n)}^{\infty} \theta_i^2\right\} + o(1) \\ &\le 2\varepsilon + \gamma^{-1}(n)\frac{1}{[2\varepsilon\beta(n)]^{2r}} \sum_{\varepsilon\beta(n)}^{\infty} i^{2r}\theta_i^2 + o(1) \\ &= 2\varepsilon + o(1). \end{aligned} \tag{6.5}$$

This completes the proof of the theorem since $\varepsilon > 0$ is arbitrary. □

Note that the estimator sequence defined by (6.2) does not depend on r. It also can easily be shown to be adaptive in rate to both B and r in a minimax sense.

That is,

$$\sup_{B>0,\, r>0} \ \limsup_{n\to\infty} \ \sup_{\theta\in\Theta_r(B)} \frac{R_n(\theta, \delta_n)}{M(B)} \le 2.5. \tag{6.6}$$

[Choose ε to depend on B, r in an optimal fashion in the proof of Theorem 6.1 and use Donoho, Liu and MacGibbon (1990).]

Property (6.6) is moderately good albeit not the best possible since the Efromovich–Pinsker (1984) adaptive estimator does even better by obtaining the optimal bound of 1. This shows that estimators which are globally superefficient in the sense of (6.1) need not behave badly with respect to uniform limits. This contrasts with the standard parametric superefficiency discussed in Section 4.

Theorem 6.1 leaves open the logical possibility that there exists an $h_n \to 0$ and a sequence of estimators such that

$$\sup_{l<\|\theta\|_*<u} \ \limsup_{n\to\infty} h^{-1}(n) R_n(\theta, \delta_n) < \infty. \tag{6.7}$$

In fact we do not know whether there exist any $h(n) \to 0$ and $\{\delta_n\}$ for which (6.7) is valid. However, Theorem 6.2, below, does show that if any such $h(n)$ exists, then it must converge to zero very slowly.

THEOREM 6.2. *Suppose* (6.7) *holds for some* $\{\delta_n\}$ *and some* $h(n) \to 0$. *Then, for any* $\eta > 0$,

$$\lim_{n\to\infty} h(n)(\ln n)^{(1+\eta)/(2r+1)} = \infty.$$

See the Appendix for the proof.

7. An example where the fixed parameter rate exists. Here is an example where we can show that there cannot be superefficiency as extreme as that in Sections 5 and 6. In fact we do not know here whether there is an estimator sequence which is superefficient on all of Θ.

Let $\{Z_i^{(n)}\}$ be as in Section 3, but define

$$\Theta(B) = \{\theta: \theta_i^2 \le B^2/i^{2r}, i = 1, \dots\}.$$

and $\Theta = \bigcup \Theta(B)$. For this example define $R_n(\theta, \delta) = n^{(2r-1)/2r}E(\|\theta - \delta\|^2)$ since $n^{-(2r-1)/2r}$ is now the correct minimax rate. With this definition M and m can be defined via the formal expressions (3.3) and (3.4), and it can be shown that

$$M(B) = C_r^1 B^{1/r} \tag{7.1}$$

so that $0 < m(\theta) < \infty$ for $\theta \neq 0$. See Donoho, Liu and MacGibbon (1990).

THEOREM 7.1. *In the above setting there is a* $\gamma > 0$ *such that*

$$\inf_{\{\delta_n\}} \sup_{\theta \neq 0} \limsup_{n \to \infty} \frac{R_n(\theta, \delta_n)}{m(\theta)} > \gamma. \tag{7.2}$$

NOTE. The left-hand side of (7.2) is necessarily bounded above by 1, but we do not know whether 1 is a sharp bound. If it were, that would mean that there is no estimator sequence which is superefficient everywhere. A minor modification of the proof would show that

$$\inf_{\{\delta_n\}} \sup_{l \le m(\theta) \le u} \limsup_{n \to \infty} \frac{R_n(\theta, \delta_n)}{m(\theta)} > \gamma.$$

PROOF. In the following lemma let $Y \sim N(\mu, \sigma^2)$. Suppose σ^2 is known and $|\mu| \le \tau$. Define the truncated loss

$$L_C(\mu, d) = (\mu - d)^2 \wedge C, \qquad d \in \mathbf{R}. \tag{7.3}$$

Let g_τ denote the Huber–Bickel prior

$$g_\tau(\mu) = \tau^{-1} I_{\{\mu \le \tau\}}(\mu) \cos^2(\mu/\tau) \tag{7.4}$$

[see Huber (1981) and Bickel (1981)], and let G_τ denote the corresponding probability distribution. Let R_{L_C} denote the risk function corresponding to L_C.

LEMMA 7.1. *There is an* $\varepsilon > 0$ *such that for every* $C > 10\sigma^2$ *the risk* R_{L_C} *satisfies*

$$\inf_\delta \int R_{L_C}(\mu, \delta) g_\tau(\mu)\, d\mu \ge \varepsilon(\tau^2 \wedge \sigma^2). \tag{7.5}$$

PROOF. This result follows from combining Brown [(1992), Theorem 2.1] with Brown and Gajek [(1990), Example 3.3]. □

Now let G^* denote the product distribution Θ, $G_B^*(d\theta) = \prod_{i=1}^{\infty} G_{B_i^{-r}}(d\theta_i)$. Then

$$\int E_\theta\left(n^{(2r-1)/2r}\|\delta_n - \theta\|^2 \wedge 10\right)G_B^*(d\theta)$$

$$\begin{aligned}(7.6)\qquad &\geq \sum_{i=1}^{B^{1/r}n^{1/2r}} \int E_\theta\left(n^{(2r-1)/2r}\left[(\delta_{ni} - \theta_i)^2 \wedge \frac{10}{n}\right]\right)G_B^*(d\theta)\\ &\geq \sum_{i=1}^{B^{1/r}n^{1/2r}} n^{(2r-1)/2r}\varepsilon(B^2 i^{-2r} \wedge n^{-1})\\ &= \varepsilon B^{1/r}\end{aligned}$$

since $n^{-1} \leq B^2 i^{-2r}$ for $i \leq B^{1/r}n^{1/2r}$.

Then $E_\theta(n^{(2r-1)/2r}\|\delta_n - \theta\|^2) \geq E_\theta(n^{(2r-1)/2r}\|\delta_n - \theta\|^2 \wedge 10)$ and

$$\sup_{\theta\in\Theta(B)} \limsup_{n\to\infty} E_\theta\left(n^{(2r-1)/2r}\|\delta_n - \theta\|^2 \wedge 10\right)$$

$$\begin{aligned}(7.7)\qquad &\geq \int \limsup_{n\to\infty} E_\theta\left(n^{(2r-1)/2r}\|\delta_n - \theta\|^2 \wedge 10\right)G^*(d\theta)\\ &\geq \limsup_{n\to\infty} \int E_\theta\left(n^{(2r-1)/2r}\|\delta_n - \theta\|^2 \wedge 10\right)G^*(d\theta)\\ &\qquad\qquad \text{(by the bounded convergence theorem)}\\ &\geq \varepsilon B^{1/r}\end{aligned}$$

by (7.6). In view of (7.1) this verifies (7.2) and completes the proof of Theorem 7.1. □

8. An example with global superefficiency in rate. In this example an even more dramatic form of global superefficiency holds. Here, for any function $h(n) \to \infty$ with $h(n) = o(n^{1/3})$ there is an estimator sequence δ_n for which

$$(8.1)\qquad \limsup_{n\to\infty} h(n)\frac{R_n(\theta, \delta_n)}{m(\theta)} = 0$$

with R_n the normalized risk having the appropriate normalization ($= n^{2/3}$) for which $0 < m(\theta) < \infty$, $\theta \neq 0$. This means that there is a sequence of estimators for which fixed parameter squared-error risk goes to zero faster than the rate $h^{-1}(n)n^{-2/3}$ even though the minimax mean squared-error risk goes to zero at the rate $n^{-2/3}$.

This example involves a somewhat artificial parameter space; we do not know whether examples of this behavior exist having more familiar parameter spaces.

It is most convenient to formulate this example in the white-noise setting, as mentioned in Section 3, where one observes

$$dZ(t) = f(t)\,dt + \frac{1}{\sqrt{n}}\,dB(t), \qquad t \in (0,1).$$

Let $c, d, r, m, \in \mathbf{R}$ and let $\Theta = \{f_{c,d,r,m}\colon m > 0\}$ where

$$f_{c,d,r,m}(t) = \begin{cases} (d - m|t - r|) \vee c, & \text{if } d \geq c, \\ (d + m|t - r|) \wedge c, & \text{if } d < c. \end{cases}$$

The functional $\|\cdot\|_*$ is defined by $\|f_{c,d,r,m}\|_* = |m|$ with $\|f_{c,d,r,m}\|_* = 0$ if $c = d$. (The values of c, d and r could be bounded without affecting the following results.) The goal is to estimate $f(0)$. For this example define $R_n(f, d) = n^{2/3}(d - f(0))^2$. It can be shown, using the methods of Donoho and Liu (1991a, b) or Brown and Farrell (1990), that $n^{-2/3}$ is the correct minimax rate, so that

(8.2) $\quad \theta < M(B) < \infty \quad$ for $B \neq 0$, and $\quad 0 < m(f) < \infty \quad$ for $\|f\|_* \neq 0$.

THEOREM 8.1. *Let $h(n) \to \infty$ and $h(n) = o(n^{1/3})$. Then there exists an estimator sequence such that* (8.1) *holds.*

PROOF. Let $K(t)$ be any bounded continuous function satisfying

$$K(t) = 0 \quad \text{if } |t| \geq 1, \quad K(t) = K(-t) \qquad \forall\, t, \tag{8.3}$$
$$\int K(t)\,dt = 1, \qquad \int |t| K(t)\,dt = 0.$$

Let δ_n be the linear estimator

$$\delta_n(Z(t)) = \int h(n) K(h(n)t)\,dZ(t). \tag{8.4}$$

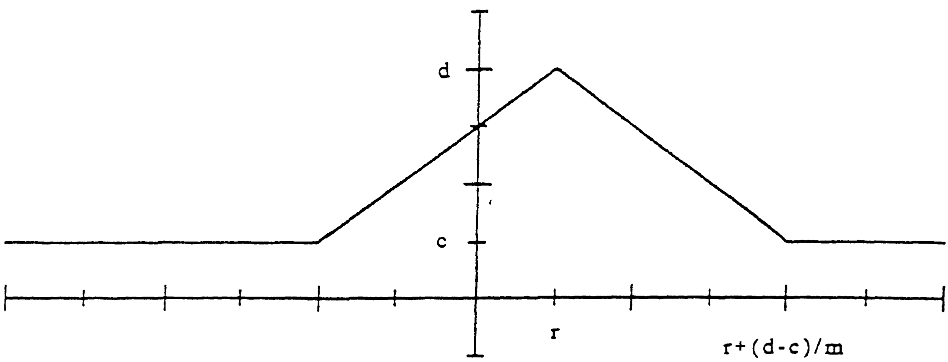

FIG. 1. $f_{c,d,r,m}$ *for $d > c$.*

A routine calculation yields

$$\mathrm{Var}_f(\delta_n) = n^{-1}h(n)\int K^2(t)\,dt. \tag{8.5}$$

If $r = 0$ and $h^{-1}(n) < (d - c) \le (d - c)/n$, then

$$\begin{aligned} E_f(\delta_n) &= \int h(n) f(T) K(h(n)t)\,dt \\ &= \int f(h^{-1}(n)v)K(v)\,dv \quad \text{(by a change of variables)} \\ &= \int (d - mh^{-1}(n)|v|)K(v)\,dv \\ &\qquad\qquad [\text{by the definition of } f \text{ and the bound on } g(n)] \\ &= d = f(0) \end{aligned}$$

by (8.3). Symmetrically, if $r = 0$ and $h^{-1}(n) < c - d$, then also $E_f(\delta_n) = f(0)$. When $d - c > 0$, $r + (b - a)/m = 0$ and $h^{-1}(n) < d - c$, then

$$E_f(\delta_n) = \int_{-1}^{0} (c - mh^{-1}(n)v)K(v)\,dv = c = f(0)$$

since (8.3) entails $\int_{-1}^{0} vK(v)\,dv = 0$. All other cases similarly yield

$$E_f(\delta_n) = f(0) \qquad \forall f \in \mathscr{F} \tag{8.6}$$

for all n sufficiently large (depending on f). Together (8.5) and (8.6) and the definition of g yields (8.1). □

There is a converse result to that in Theorem 8.1; if $\liminf n^{-1/3}h(n) > 0$, then (8.1) does not hold.

THEOREM 8.2. *For any* $\{\delta_n\}$

$$\limsup_{|d-c|\to 0} \limsup_{n\to\infty} E_{f_{0,d,0,1}}\left(n\left(\delta_n - f(0)^2\right)\right) = \infty. \tag{8.7}$$

PROOF. For convenience, let $f_d = f_{0,d,0,1}$. Consider the one parameter subfamily of $\mathscr{F}$ defined by $\{f_d\colon |d| \le 1\}$. The Fisher information is $I(d) = 2d$, with $\lim_{d\to 0} I(d) = 0$. Consequently,

$$\begin{aligned} &\lim_{B\to\infty} \liminf_{n\to\infty} \int E_{f_d}\left(n(\delta_n - d)^2 \wedge B\right) g_L(d)\,dd \\ &\qquad \ge 1/2L = \min\{I^{-1}(d)\colon |d| \ge L\}. \end{aligned}$$

[This modification and extension of Lemma 7.1 also follows from Brown (1992) and Brown and Gajek (1990).] Choose B sufficiently large and use the bounded convergence theorem (as in the proof of Theorem 7.1) to conclude

$$\sup_{|d|\le L} \limsup_{n\to\infty} E_{f_d}\left(n(\delta_n - d)^2\right) \ge 1/4L.$$

This yields (8.7). □

APPENDIX

Two proofs.

PROOF OF THEOREM 5.2. It suffices to prove that, for any sequence $\{\delta_n\}$,

$$\sup_{\theta \in \Theta(B)} \limsup_{n \to \infty} R_n(\theta, \delta_n) > 0. \tag{A.1}$$

As before, let $\beta(n) = n^{1/(2r+1)}$ and $\gamma^{-1}(n) = n^{2r/(2r+1)}$. Let $\{\delta_n\}$ be any sequence of linear estimators. Write δ_n as $\delta_n = M_n z$, where M_n is $(\infty \times \infty)$. Let $\ell_{n,i}^2 = \Sigma_j (M_n)_{ij}^2$ and let

$$v(n) = \#\{i: \ell_{n,i}^2 \geq 1/4\}.$$

Note that Var $\delta_n \geq v(n)/4n$. Hence, either $\{\delta_n\}$ satisfies (A.1) or

$$\frac{v(n)}{\beta(n)} \to 0 \quad \text{as } n \to \infty, \tag{A.2}$$

since $\gamma^{-1}(n) = \beta(n)/n$. So, assume (A.2). Also assume $\ell_{n,i}^2 < \infty$ for all n sufficiently large, since otherwise (A.1) is satisfied by $\{\delta_n\}$. For convenience, assume with no loss of generality, that $\ell_{n,i}^2 < \infty$ for all n, i.

Let $w_j = w/j$ with $w > 0$ an appropriate, small constant [e.g., $w = (A^3 \Sigma_{j=1}^{\infty} 1/j^2)^{-1/2}$]. We construct inductively the coordinates of a point $\theta \in \Theta$ such that (A.1) is true. To begin, choose n_1 so that

$$\frac{\beta(n_1)}{v(n_1)} w_1 \geq 1, \qquad 2v(n_1) \leq n_1. \tag{A.3}$$

This is possible because of (A.2) and because $\beta(n) = o(n)$. Let

$$i_1 = \min\{i: v(n_1) \leq i \leq 2v(n_1) + 1, \ell_i^2 \leq 1/4\}; \tag{A.4}$$

i_1 exists because of the definition of v. Let $\theta_{i_1} = w_1/(v(n_1))^r$. Now choose $m_1 \geq i_1$ so that $\Sigma_{j=m_1}^{\infty} (M_{n_1})_{i_{1,j}}^2 / m_1^{2r} \leq \theta_{i_1}^2/8$. This is possible because $\ell_{n_1, i_1}^2 < \infty$. Then define $\theta = 0$ for $1 \leq i \leq m_1$, $i \neq i_1$.

Now proceed inductively. Given n_{k-1}, m_{k-1} and θ_i for $i = 1, \ldots, m_{k-1}$ choose n_k so that

$$\frac{\beta(n_k)}{v(n_k)} w_k \geq 1 m_{k-1} < v(n_k) \leq n_{k/2}. \tag{A.5}$$

Let

$$i_k = \min\{i: v(n_k) \leq i \leq 2v(n_k) + 1, \ell_i^2 \leq 1/4\}. \tag{A.6}$$

Let

$$(\text{A.7})\qquad \theta_{i_k} = \begin{cases} w_k/(v(n_k))^r, & \text{if } \sum_{j=1}^{m_{k-1}} (M_{n_k})_{i_k,j}\theta \le 0,\\ -w_k/(v(n_k))^r, & \text{otherwise.}\end{cases}$$

Finally, choose $m_k \ge i_k$ so that

$$(\text{A.8})\qquad \sum_{j=m_k}^{\infty} (M_{n_k})^2_{i_k,j}/m_k^2 \le \theta^2_{i_k}/8$$

and define

$$(\text{A.9})\qquad \theta_i = 0 \quad \text{if } m_{k-1} < i \le m_k \text{ and } i \ne i_k.$$

It remains to verify that $\theta \in \Theta$ and that (A.1) is valid. Toward the first goal, note for θ as defined above

$$(\text{A.10})\qquad \sum_{i=1}^{\infty} i^{2r}\theta_i^2 = \sum_{k=1}^{\infty} i_k^{2r}\theta_{i_k}^2 = \sum i_k^{2r} w_k^2/(v(n_k))^{2r} \le A^3 \sum w_k^2 \le 1.$$

Hence $\theta \in \Theta$.

Toward the second goal, note

$$(\text{A.11})\qquad \Big(\sum_j (M_{n_k})_{i_k,j}\theta_j\Big)^2 \le 2\Big(\sum_{j=1}^{m_k} (M_{n_k})_{i_k,j}\theta_j\Big)^2 + 2\Big(\sum_{j=m_k+1}^{\infty} (M_{n_k})_{i_k,j}\theta_j\Big)^2.$$

Now,

$$\begin{aligned}\Big|\sum_{j=1}^{m_k} (M_{n_k})_{i_k,j}\theta_j\Big| &\le \big|(M_{n_k})_{i_k,i_k}\theta_{i_k}\big| \quad \text{[by (A.7) and (A.9)]}\\ &\le \theta_{i_k}/2\end{aligned}$$

since $(M_{n_k})_{i_k,i_k} \le 1/2$ by (A.6). Also

$$\begin{aligned}\Big(\sum_{j=m_k+1}^{\infty} (M_{n_k})_{i_k,j}\theta_j\Big)^2 &\le \Big(\sum_{j=m_k+1}^{\infty} (M_{n_k})^2_{i_k,j}\Big)\Big(\sum_{j=m_k+1}^{\infty} \theta_j^2\Big)\\ &\le \Big(\sum_{j=m_k+1}^{\infty} (M_{n_k})^2_{i_k,j}\Big)\Big/(m_k+1)^{2r} \quad \text{[by (A.10)]}\\ &\le \theta^2_{i_k}/9\end{aligned}$$

by (A.8). Hence

$$(\text{A.12})\qquad \Big(\sum_{j=1}^{\infty} (M_{n_k})_{i_k,j}\theta_j\Big)^2 \le 3\theta^2_{i_k}/4.$$

This inequality implies

$$
\begin{aligned}
E_\theta\|\delta_{n_k} - \theta\|^2 &\geq E_\theta\Big(\big((\delta_{n_k})_{i_k} - \theta_{i_k}\big)^2\Big) \\
&\geq \Big[E_\theta\big((\delta_{n_k})_{i_k}\big) - \theta_{i_k}\Big]^2 \\
&\geq \left(1 - \frac{\sqrt{3}}{2}\right)^2 \theta_{i_k}^2 \quad [\text{by } (6.13)] \\
&> 0.01 \frac{w_k^2}{(v(n_k))^{2r}}.
\end{aligned} \tag{A.13}
$$

Recall that $\gamma^{-1}(n_k) = (\beta(n_k))^{2r}$. Hence

$$
E_\theta\big(\gamma^{-1}(n_k)\|\delta_{n_k} - \theta\|^2\big) \geq 0.01\left(\frac{\beta(n_k)}{v(n_k)}\right)^{2r} w_k^2 \geq 0.01
$$

by (A.5). This verifies (A.1) and completes the proof. □

Proof of Theorem 6.2. Fix $\{\delta_n\}$. It suffices to show that the choice

$$
h(n) = \ln^{-(1+\eta)/(2r+1)}(n) \tag{A.14}
$$

does not satisfy (6.7). Indeed, it is enough to show that, for arbitrary $B > 0$, and for h as in (A.14),

$$
\lim_{C\to\infty} \sup_{0\leq\|\theta\|_*\leq B} \limsup_{n\to\infty} h(n) n^{2r/(2r+1)} E_\theta\big(\|\delta_n - \theta\|^2\big) = \infty. \tag{A.15}
$$

[Note that a 0 replaces the lower bound ℓ in (6.7); but it is easy to show by translation that (A.15) implies (6.7) is not satisfied.]

To show (A.15), let

$$
M_i^2 = \frac{K}{i^{(2r+1)} \ln^{(1+\eta/2)}(i)}
$$

for an appropriate $K > 0$, depending only on B, r as specified in (A.16) below. Then

$$
\sum i^{2r} M_i^2 = K \sum \frac{1}{i \ln^{(1+\eta/2)}(i)} < B^2 \tag{A.16}
$$

for appropriate $K > 0$.

Let $G^* = \prod_{i=1}^{\infty} G_{M_i}$ with G_{M_i} as in (7.4). Then G^* is supported on $\{\theta: \|\theta\|_* \le B\}$ by (A.16). Reasoning as in (7.6), (7.7) then yields

$$
\begin{aligned}
&\sup_{\|\theta\|_* \le B} \limsup_{n\to\infty} \ln^{(1+\eta)/(2r+1)}(n) R_n(\theta, \delta_n) \\
&\quad \ge \limsup_{n\to\infty} \left\{ n^{2r/(2r+1)} \ln^{(1+\eta)/(2r+1)}(n) \sum \left(\frac{1}{n} \wedge M_i^2 \right) \right\} \\
&\quad \ge \limsup_{n\to\infty} \left\{ n^{2r/(2r+1)} \ln^{(1+\eta)/(2r+1)}(n) \right. \\
&\qquad \left. \times \frac{n^{1/(2r+1)}}{n(2r+1)^{(1+\eta/2)/(2r+1)} \ln^{(1+\eta/2)/(2r+1)}(n)} \right\}
\end{aligned} \tag{A.17}
$$

since

$$
\frac{1}{n} < M_i^2 \quad \text{for } i \le \frac{(2r+1)^{-(1+\eta/2)/(2r+1)} n^{1/(2r+1)}}{\ln^{(1+\eta/2)/(2r+1)}(n)}.
$$

This shows that (6.7) is not satisfied, since the right-hand side of (A.17) is infinite. □

REFERENCES

ABRAMSON, I. S. (1982). Bandwidth variation in kernel estimates—a square root law. *Ann. Statist.* **10** 1217–1223.

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer, New York.

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

BROCKMANN, M., GASSER, T. and HERRMAN, E. (1993). Locally adaptive bandwidth choice for kernel regression estimators, *J. Amer. Statist. Assoc.* **88** 1302–1309.

BROWN, L. D. (1986). Fundamentals of statistical exponential families with applications in statistical decision theory. IMS, Hayward, CA.

BROWN, L. D. (1992). An information inequality for the Bayes risk under truncated squared error loss. In *Multivariate Analysis: Future Directions* (C. R. Rao, ed.) 85–94. North-Holland, Amsterdam.

BROWN, L. D. and FARRELL, R. (1990). A lower bound for the risk in estimating the value of a probability density. *J. Amer. Statist. Assoc.* **85** 1147–1153.

BROWN, L. D. and GAJEK, L. (1990). Information inequalities for the Bayes risk. *Ann. Statist.* **18** 1578–1594.

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

BROWN, L. D. and LOW, M. G. (1996b). A constrained risk inequality with applications to nonparametric functional estimation. *Ann. Statist.* **24** 2524–2535.

DONOHO, D. L. and LIU, R. C. (1991a). Geometrizing rates of convergence II. *Ann. Statist.* **19** 633–667.

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

DONOHO, D. L., LIU, R. C. and MACGIBBON, B. (1990). Minimax risks over hyperrectangles, and implications. *Ann. Statist.* **18** 1416–1437.

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

HÁJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 175–194. Univ. California Press, Berkeley.

HALL, P. (1993). On plug-in rules for local smoothing of density estimates. *Ann. Statist.* **21** 694–710.

HUBER, P. J. (1966). Strict efficiency excludes superefficiency (abstract). *Ann. Math. Statist.* **37** 1425.

HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.

LE CAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *University of California Publications in Statistics* **1** 277–330.

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

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

SCHUCANY, W. R. (1995). Adaptive bandwidth choice for kernel regression. *J. Amer. Statist. Assoc.* **90** 535–540.

STEIN, C. (1966). An approach to the recovery of inter-block information in balanced incomplete block designs. *Research Papers in Statistics: Festschrift for Jerzy Neyman* (F. N. David, ed.) 351–366. Wiley, London.

WEISS, L. and WOLFOWITZ, J. (1966). Generalized maximum likelihood estimators. *Theory Probab. Appl.* **11** 58–81.

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

DEPARTMENT OF STATISTICS
THE WHARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104
E-MAIL: lbrown@compstat.wharton.upenn.edu
lowm@compstat.wharton.upenn.edu
lzhao@compstat.wharton.upenn.edu