A Lower Bound for the Risk in Estimating the Value of a Probability Density

LAWRENCE D. BROWN and ROGER H. FARRELL*

This article considers estimation of $f(0)$ of a density function satisfying a Lipshitz condition in a neighborhood of 0. A nonstandard use of the Cramer–Rao inequality yields numerical lower bounds on the minimax squared error risk of any estimator. These bounds are then compared with the minimax risk of the asymptotically optimal kernel-type estimator. The asymptotic bounds obtained (as the sample size $n \to \infty$) are not quite as good as those in Donoho and Liu (in press a, b), but bounds are presented here also for finite values of n, and that paper contains no such bounds for this problem. The numerical results reported in Table 1 show that the asymptotically optimal kernel estimator performs within a factor of 3 of the minimax bound, even for sample size $n = 30$. As n increases the relative performance improves to its limiting value, although the convergence is fairly slow.

KEY WORDS: Cramer–Rao inequality; Density estimation; Kernel estimator.

1. INTRODUCTION

1.1 Overview

Nonparametric density estimation is one of several related statistical problems involving nonstandard asymptotic theory having rates of convergence other than the usual order of $1/\sqrt{n}$. This article examines one very specific density estimation problem from a new perspective. Numerical results of a nonasymptotic character are obtained for finite sample sizes. In addition, an asymptotic result is presented.

The problem under consideration has a special, easily manageable structure. The methods introduced in this article, however, are clearly applicable to a wide variety of nonstandard problems. It seems reasonable to conjecture that the general pattern of numerical results in many of those problems will be qualitatively similar to that observed here.

Kernel estimators for density estimation problems were introduced by Rosenblatt (1956) and Parzen (1962). Since then a variety of other methods have been suggested. Some of these, such as certain cross-validation techniques, are nonlinear methods constructed from (linear) kernel estimators (Hall 1983; Stone 1984); others, such as spline smoothing (Silverman 1984; Wahba 1975), possess different motivations.

A number of attempts has been made to examine absolute and relative properties of these various methods. Of particular note here are two general classes of studies. One type of study looks at asymptotic rates of convergence of particular estimators. These can then be compared with optimal rates as established, for example, in Farrell (1972, 1979). Another type of study compares finite sample properties of specific estimators. See, for example, Marron (1987) and Altman (1988).

The approach here is somewhat different. The goal is to compare the minimax risk of a specific kernel estimator at reasonable sample sizes with a lower bound derived from the Cramer–Rao inequality. This comparison provides a bound on the improvement in performance possible through use of other, more sophisticated procedures. In the case at hand the comparison shows that the optimum kernel estimator cannot be greatly improved in this minimax sense. This is particularly true for moderately large sample sizes. In the sense that this article looks at a bound for the actual value of the minimax risk, rather than just the asymptotic rate of convergence, it is closely related to Birge (1987a,b). However, his problem, his methods, and, to some extent, the qualitative pattern of his results differ from those in our article.

The numerical results are summarized in Table 1. In brief, even for moderate sample sizes the optimum kernel estimator usually performs within a factor of 3 of the minimax bound. As $n \to \infty$ it performs within a factor of 1.45 $= (.69)^{-1}$. (An even better asymptotic bound is possible. See the parenthetical comment in the next section.)

1.2 Asymptotic Results

This article considers estimation of $f(0)$ for a density function f satisfying the following boundedness and Lipshitz conditions:

$$f(x) \le a, \qquad \left|\frac{f(x) - f(y)}{x - y}\right| \le b \tag{1.1}$$

in a specified neighborhood of 0.

Sacks and Strawderman (1982) showed in a closely related problem that the optimum kernel-type estimator sequence is not asymptotically optimum in the appropriate minimax sense. Their methods also apply to the problem at hand (Mark Low, personal communication). Thus there exists another sequence of estimators for which the limiting ratio of maximum risk to that of the kernel estimator sequence is (slightly) less than 1. The present results show in spite of this fact that the optimum kernel-type estimator is not too bad in the asymptotic minimax sense by showing that in our problem such a ratio can never be less than .69.

* Lawrence D. Brown is Professor, Department of Mathematics, and Director of the Statistics Center, Cornell University, Ithaca, NY 14853. Roger H. Farrell is also Professor, Department of Mathematics, Cornell University. Research of both authors was supported in part by National Science Foundation Grants DMS-8506847 and 8809016.

Journal of the American Statistical Association
December 1990, Vol. 85, No. 412, Theory and Methods

Table 1. Values of $\{\rho^*, \rho_*\}$ for Selected Values of a, b, n

a	b	30	60	100	200	500	10^3	10^4	10^5	10^6	10^7	10^8	10^9
$\frac{1}{2}$	$\frac{1}{2}$	.71	.75	.78	.82	.86	.88	.94	.97	.98	.99	.99	.99
		.31	.33	.35	.36	.38	.44	.56	.63	.66	.68	.69	.69
1	$\frac{1}{2}$	.38	.47	.54	.62	.71	.76	.88	.94	.97	.98	.99	.99
		.21	.27	.31	.37	.43	.47	.56	.63	.66	.68	.69	.69
$\frac{1}{2}$	1	.75	.80	.82	.86	.89	.91	.95	.97	.98	.99	.99	.99
		.29	.32	.34	.36	.38	.39	.54	.62	.66	.68	.69	.69
1	1	.59	.65	.69	.75	.80	.84	.92	.96	.98	.99	.99	.99
		.33	.37	.39	.41	.44	.48	.57	.63	.66	.68	.69	.69
$\frac{1}{2}$	1	.73	.80	.84	.87	.90	.91	.96	.98	.99	.99	.99	.99
		.25	.30	.32	.35	.37	.38	.50	.61	.65	.68	.69	.69
1	2	.71	.75	.78	.82	.86	.88	.94	.97	.98	.99	.99	.99
		.31	.33	.35	.36	.38	.44	.56	.63	.66	.68	.69	.69

[At the same time as this article was being prepared Liu (1987), Donoho and Liu (in press a,b), and Donoho, Liu, and MacGibbon (1990) wrote a series of papers concerning problems with nonstandard asymptotics. These papers describe a number of new methods to derive lower bounds for the asymptotic minimax risk in such problems. Their best method yields an asymptotic bound that is even better than the value .69 mentioned here. The best bound comes from their "hardest linear subfamily" method combined with the tables of Casella and Strawderman (1981). This yields the better asymptotic bound of .85. However, the method used to derive this bound is based on asymptotic normality and does not seem suitable to derive non-asymptotic bounds like those given in our Table 1.]

1.3 One-Parameter Subfamilies

The Cramer–Rao method to be used here requires that for each given sample size, n, the problem be reduced to a suitably chosen one-parameter subproblem, say $\{f_{\theta,n} : \theta \in [-T_n, T_n]\}$. The subproblem chosen in Section 2 has only a mild dependence on n and on the constants a, b in (1.1). This is convenient, but it is not a logical necessity for application of the method.

There is a double intuition lying behind the choice of the family. One motivation is that the family should be chosen so that for each fixed value of $f_\theta(0) - f_\theta(0)$ the densities should be as close together as possible in some vaguely specified sense. The other motivation lies in the Fisher information itself. That information is locally of the form $I(\theta) = c|\theta|^d + 0(\theta)$. The object is to choose the family subject to (1.1) and to the constraint $f_\theta(0) = f_\theta(0) + \theta + o(\theta)$ so that d will be as large as possible ($d > 1$) and then c will be as small as possible. Such a family is locally least favorable with respect to Fisher information.

These intuitive considerations suggest the choice of a family that behaves for small $|\theta|$ like

$$f_\theta(x) = f_\theta(0) + \theta(1 - b|x|/|\theta|)^+ + o(\theta) \quad (1.2)$$

for x in the specified neighborhood of $x = 0$. The family defined in (2.5) has this property. It has been chosen, in addition, to balance somewhat the competing requirements that $f_\theta(0)$ be large and T_n be large. The desire to balance these two requirements as well as the need to guarantee that f_θ be a probability density explain the algebraic complexity of (2.4) and (2.5). If only an asymptotic result were requested, then it would suffice to choose virtually any family satisfying (1.2) and having $f_\theta(0) = a - \varepsilon$ with $\varepsilon > 0$ arbitrarily small.

[At the time this article was written we believed that this intuition would lead to the best asymptotic bound possible through use of the ordinary Cramer–Rao inequality. However, more recent results in Brown and Low (1990) for a different nonstandard problem suggest that this is not the case, although it does not appear that a different choice of the family would improve the resulting asymptotic bound by more than a few percent.]

One might contemplate choosing a multidimensional subfamily and then applying the multiparameter Cramer–Rao inequality. If an efficient method could be derived for applying this inequality, then such a procedure could in principle lead to somewhat improved bounds. The nature of the results here [and, even more, the results in Donoho and Liu (in press a,b)], however, make it clear that one can derive surprisingly accurate inequalities through the use of one-parameter subfamilies.

1.4 Cramer–Rao Inequality

Wald (1951) used the Cramer–Rao inequality via a worst case estimate of the size of the bias to show that maximum likelihood estimators were asymptotically minimax. In that sense the present methods can be considered a refinement of Wald's idea. Wald's problem, however, is of the standard type in that the asymptotically optimal sequence of maximum likelihood estimators is asymptotically unbiased and has a $1/\sqrt{n}$ rate of convergence. Hodges and Lehmann (1951) then used the Cramer–Rao inequality to prove minimaxity and admissibility under squared error loss in the case in which the information is constant. Their method involves dropping the $(\beta'(\theta))^2$ term from the inequality for the risk [see (2.6)] and then directly analyzing the remaining differential inequality. See also Gajek (1988). Brown and Gajek (1990) discussed the idea of dropping this same term, multiplying by a suitably chosen density function, integrating by parts, and completing the square. Farrell, in an earlier version of this article, considered dropping the $\beta^2(\theta)$ term to obtain a cruder lower bound. This crude lower bound is easy to compute and proved to be surprisingly close to the bound currently obtained.

The present method involves rewriting the full inequality for the squared error risk as an ordinary differential inequality. It is then observed that this inequality has a solution only if the corresponding equality also has a solution. That ordinary differential equality, (2.8), can be easily solved by standard numerical methods. It contains a putative value, r, for the minimax risk. The least value of r for which this equation has a solution on all of $\theta \in [-T_n, T_n]$ is a lower bound for the minimax risk.

For the given parametric family the method outlined above gives the best possible result available from the Cramer–Rao inequality. In the current instance a small degree of precision is sacrificed to the convenient but not logically necessary symmetrization at (2.7). This imprecision vanishes as $n \to \infty$, so the asymptotic bound is the best possible (using the Cramer–Rao inequality and the given parametric subfamily, or any other subfamily asymptotically yielding the same Fisher information near $\theta = 0$).

2. MAIN RESULTS

2.1 Constraints

Fix $a > 0$, $b > 0$. Let $\mathfrak{F}_{a,b}$ denote the class of density functions satisfying

$$f(x) \leq a \ \forall \ |x| \leq \tfrac{1}{8} \tag{2.1}$$

and

$$\left|\frac{f(x) - f(y)}{x - y}\right| \leq b \ \forall \ |x| \leq \tfrac{1}{8},\ |y| \leq \tfrac{1}{8},\ x \neq y. \tag{2.2}$$

Let $X = X_1, \ldots, X_n$ be a random sample from $f \in \mathfrak{F}_{a,b}$, and let $\delta(x_1, \ldots, x_n) \in \mathbf{R}$ be an estimator of $f(0)$. Then the risk of δ is

$$R_n = R_n(f, \delta) = E_f(\delta(X_1, \ldots, X_n) - f(0))^2.$$

The value $\frac{1}{8}$ in this formulation is an ad hoc choice, which is convenient in the numerical cases to be examined in Table 1. It could be replaced by any reasonable sequence of constants $k(n)$ such that $k^{-1}(n) = o(n^{1/3})$ as $n \to \infty$. Doing so would not alter the asymptotic results to be presented, but could have some effect on the comparisons for smaller values of n.

2.2 The Lower Bound

Given a, b let

$$\begin{aligned} m &= \frac{b}{a}\left(1 - \sqrt{1 - a^2/b}\right) && \text{if } a^2 < b \\ &= \frac{b}{a} && \text{if } a^2 \geq b. \end{aligned} \tag{2.3}$$

Then let

$$T = T_n = \min\left(b^{1/2}, m, b/2a, 4.5\left(\frac{ab}{n\sqrt{3}}\right)^{1/3}\right) \tag{2.4}$$

and

$$\xi = \xi_n = \frac{a - T_n}{1 - T_n^2/b}.$$

Now define the family of densities $f_\theta = f_{\theta,n}$ for $|\theta| \leq T_n$ by

$$\begin{aligned} f_\theta(x) &= \xi(1 - (\operatorname{sgn}\theta)\theta^2/b) + (\operatorname{sgn}\theta)\max(|\theta| - b|x|, 0) \\ &\qquad \text{if } |x| < \tfrac{1}{2}\xi \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{2.5}$$

The definition of f_θ, including the restriction $|\theta| \leq T_n$, guarantees that $f_\theta \in \mathfrak{F}_{a,b}$. In fact, each f_θ satisfies the inequalities in (2.1) and (2.2) for every x and y. Note that $f_0(x) = \xi 1_{(-1/2\xi, 1/2\xi)}(x)$. In addition, $f_{T_n}(0) = a$.

Remark 2.1. The restriction $|\theta| \leq b^{1/2}$ is obviously needed for (2.5) to define a density. This explains why the term $b^{1/2}$ appears in (2.4). When $a^2 \geq b$, $\min(b^{1/2}, m) = b^{1/2}$. If $a^2 < b$, then m is the smallest solution to

$$\frac{a - m}{1 - m^2/b}\left(1 + \frac{m^2}{b}\right) - m = 0.$$

Thus if $a^2 < b$ and $\theta < -m$ the expression in (2.5) would not be nonnegative when $x = 0$. This explains why the term m appears in (2.4). It can now easily be seen that $m < a$ when $a^2 \leq b$, and hence $\min(b^{1/2}, m) \leq a$. For (2.5) to define a density it is also necessary for the truncation point $\frac{1}{2}\xi$ to satisfy $(\frac{1}{2}\xi) \geq (|\theta|/b)$. The term $b/2a$ that appears in (2.4) guarantees this to be the case, since $T < \min(b^{1/2}, b/a)$ implies that $(a - T)/(1 - T^2/b) < a$, which then implies that $(\frac{1}{2}\xi) > \frac{1}{2}a \geq |\theta|/b$ for $|\theta| \leq T_n \leq b/2a$. The factor $((ab)/n)^{1/3}$ that appears in the last part of (2.4) is a natural scaling factor, which can be derived from the asymptotic theory described subsequently. The numerical constant $4.5/3^{1/6}$ that multiplies that factor has been chosen by trial and error to yield good lower bounds in Table 1. [Large choices make $f_0(0)$ small, which tends to decrease the lower bound, but small values reduce the range, $[-T_n, T_n]$, of θ, which also tends to decrease the lower bound.] As $n \to \infty$, it is possible to replace this constant by some $c(n) \to \infty$ with $c(n) = o(n^{1/3})$. Although this should slightly improve the lower bound when $n \to \infty$, it turns out that it will not affect the first two decimals (.69) as given in Table 1.

Remark 2.2. If T_n is small compared with $(4.5) \cdot [ab/n\sqrt{3}]^{1/3}$, then the method below should not be expected to yield a good lower bound. In such a case it could be considerably advantageous to begin with a family different from (2.5), such as one with densities having the same triangular shape as those in (2.5) but beginning from $f_0(x) = (a/2)1_{(-a/4, a/4)}(x)$. [Here one would need $a \geq \frac{1}{2}$ for the density to satisfy (2.2).]

Fix n. Let

$$I(\theta) = E_{f_\theta}\left[\left(\frac{\partial}{\partial t}\ln f_t(X)|_{t=\theta}\right)^2\right]$$

denote the Fisher information of the family $\{f_\theta\}$. Given an estimator $\delta_n \in [0, \infty)$, based on the sample of size n, let $\beta = \beta_n = E_\theta(\delta_n) - f_\theta(0)$ denote the bias of δ_n for estimating $f_\theta(0)$. Let

$$r = r_n = \sup\{R_n(f_\theta, \delta_n) : 0 \leq |\theta| \leq T_n\}.$$

The Cramer–Rao inequality yields

$$r \geq \frac{\left(1 - \frac{2\xi|\theta|}{b} + \beta'(\theta)\right)^2}{nI(\theta)} + \beta^2(\theta),$$
$$0 < |\theta| \leq T_n, \quad (2.6)$$

since

$$\frac{\partial}{\partial\theta} f_\theta(0) = 1 - \frac{2\xi|\theta|}{b}$$

for $0 < |\theta| < T_n$. [No regularity conditions on δ_n are needed here since the family $\{f_\theta\}$ satisfies condition (A6) of Brown and Gajek (1990).] Let

$$c(\theta) = (\beta(\theta) - \beta(-\theta))/2,$$

and let

$$J(\theta) = \max(I(\theta), I(-\theta)), \qquad 0 \leq \theta \leq T_n.$$

Note that $c(0) = 0$ since β is continuous. Jensen's inequality applied to (2.6) gives the result

$$r \geq \frac{\left(1 - \frac{2\xi|\theta|}{b} + c'(\theta)\right)^2}{nJ(\theta)} + c^2(\theta), \qquad 0 \leq \theta \leq T_n, \quad (2.7)$$

and $c(0) = 0$.

Theorem A.1 (see the Appendix) establishes that the inequality (2.7) has a solution under the initial condition $c(0) = 0$ iff (2.7) has a solution as an equality. As an equality (2.7) can be rewritten as

$$c'(\theta) = \sqrt{nJ(\theta)(r - c^2(\theta))} - 1 + 2\xi\theta/b,$$
$$0 \leq \theta \leq T_n;\ c(0) = 0. \quad (2.8)$$

This is an ordinary differential equation that can easily be solved numerically to a high degree of accuracy to ascertain whether a solution exists on the entire interval $0 \leq \theta \leq T_n$. It is also shown in Theorem A.1 (see the Appendix) that the set of r for which solutions of (2.8) exist is a half line.

Let $r_* = r_*(a, b, n)$ denote the (numerically determined) lower bound for values of r such that a solution to (2.8) exists. Then

$$r_*(a, b, n) \leq \sup\{R_n(f_\theta, \delta_n) : 0 \leq |\theta| \leq T_n\}$$
$$\leq \sup\{R_n(f, \delta_n) : f \in \mathfrak{F}_{a,b}\} \quad (2.9)$$

for all δ_n. Hence r_* is a lower bound for the minimax risk over $\mathfrak{F}_{a,b}$. For reasons that will appear in following paragraphs it is convenient to table the normalized values

$$\rho_*(a, b, n) = \left(\frac{3n^2}{a^2b^2}\right)^{1/3} r_*(a, b, n). \quad (2.10)$$

Selected values of ρ_* are presented in Table 1.

2.3 The Asymptotic Lower Bound

Note that

$$\frac{\partial}{\partial\theta} f_\theta(x) = \operatorname{sgn}\theta + 0(\theta) \qquad \text{if } |x| < |\theta|/b$$
$$= 0(\theta) \qquad \text{otherwise,}$$

with the error term being uniform in x and ξ. Hence

$$I(\theta) = \int_{-|\theta|/b}^{|\theta|/b} \frac{(\operatorname{sgn}\theta)^2}{\xi}\, dx + o(\theta) = \frac{2|\theta|}{b\xi} + o(\theta).$$

Now let $n \to \infty$, so that $T_n \to 0$ and $\xi_n \to a$. The Cramer–Rao statement (2.6) becomes: For any $\varepsilon > 0$, there is an $n(\varepsilon)$ such that

$$r_n + \varepsilon \geq \frac{(1 + \beta'(\theta))^2}{n\left(\frac{2|\theta|}{ab}\right)(1 + \varepsilon)} + \beta^2(\theta), \qquad 0 \leq |\theta| \leq T_n, \quad (2.11)$$

whenever $n \geq n(\varepsilon)$.

Change variables in (2.11) by letting

$$\zeta = \left(\frac{3n^2}{a^2b^2}\right)^{1/6} \theta,$$
$$d(\zeta) = \left(\frac{3n^2}{a^2b^2}\right)^{1/6} \beta\left(\left(\frac{3n^2}{a^2b^2}\right)^{-1/6} \zeta\right),$$
$$\text{and } \rho(a, b) = \liminf_{n\to\infty} \left(\frac{3n^2}{a^2b^2}\right)^{1/3} r_n. \quad (2.12)$$

Then (2.11) becomes

$$\rho(a, b) \geq \frac{3^{1/2}(1 + d'(\zeta))^2}{2|\zeta|(1 + \varepsilon)} + d^2(\zeta),$$
$$0 \leq |\zeta| \leq \left(\frac{3n^2}{a^2b^2}\right)^{1/6} T_n \to \infty. \quad (2.13)$$

This equation has the same general structure as (2.6) except that it is now exactly symmetric in ζ.

Let $\rho_*(a, b, \infty)$ be the minimal value for which the equation

$$\rho_*(a, b, \infty) = \frac{3^{1/2}(1 + d'(\zeta))^2}{2|\zeta|} + d^2(\zeta),$$
$$0 \leq |\zeta| \leq 4.5 \text{ (say)}, \quad (2.14)$$

has a solution subject to $d(0) = 0$. It follows that

$$\rho_*(a, b, \infty) \leq \rho(a, b)$$
$$\leq \liminf_{n\to\infty} \sup\{R_n(f, \delta_n) : f \in \mathfrak{F}_{a,b}\} \quad (2.15)$$

for all $\{\delta_n\}$, since $\varepsilon > 0$ can be chosen arbitrarily small.

[As previously remarked, the constant 4.5 that appears in (2.14) is arbitrary. Any larger choice is also reasonable, but numerical investigations have shown that a larger choice would not affect the two decimal places of the figures in Table 1.]

Table 1 exhibits the convergence of $\rho_*(a, b, n) \to \rho_*(a, b, \infty)$ that is indicated by the above.

2.4 Kernel Estimators: Asymptotic Theory

The general form of a kernel estimator is $\delta_n(x_1, \ldots, x_n) = n^{-1} \sum_{i=1}^n k_n(x_i)$, with $\int k_n(x)\, dx = 1$.

Sacks and Ylvisaker (1981) established that, if $\{\delta_n'\}$ is any sequence of kernel estimators, then

$$\liminf_{n\to\infty} \sup_{f\in\mathfrak{F}_{a,b}} n^{2/3} R_n(f, \delta_n') \geq \left[\frac{a^2b^2}{3}\right]^{1/3}. \quad (2.16)$$

Furthermore, if δ_n is the kernel estimator sequence defined by

$$k_n(x) = \alpha_n^{-1} k(\alpha_n^{-1} x), \quad (2.17)$$

where

$$k(x) = \max(1 - |x|, 0) \quad \text{and} \quad \alpha_n^{-1} = \beta n^{1/3} \quad (2.18)$$

with

$$\beta = \frac{(b \int |x| k(x)\, dx)^{2/3}}{((a/2) \int k^2(x)\, dx)^{1/3}} = \frac{b^{2/3}}{(3a)^{1/3}}, \quad (2.19)$$

then

$$\lim_{n\to\infty} \sup_{f\in\mathfrak{F}_{a,b}} n^{2/3} R_n(f, \delta_n) = \left[\frac{a^2b^2}{3}\right]^{1/3}. \quad (2.20)$$

Hence this sequence of estimators is asymptotically minimax within the class of kernel estimators over $\mathfrak{F}_{a,b}$.

2.5 Kernel Estimators: Numerical Results

Table 1 describes the performance of a minor variant of this estimator for various values of n, a, b. The constant β in (2.19) is replaced by

$$\beta' = \max(\beta, 8/n^{1/3}) = \max\left(\frac{b^{2/3}}{(3a)^{1/3}}, 8/n^{1/3}\right), \quad (2.21)$$

since the smoothness condition (2.2) is valid only on $[-\frac{1}{8}, \frac{1}{8}]$. [Note that use of a value $\beta < 8/n^{1/3}$ yields $\alpha_n = 1/\beta n^{1/3} > \frac{1}{8}$, which results in an unacceptably large bias if $f(x)$ is discontinuous at $x = \pm\frac{1}{8}$.]

It is difficult to determine the precise values of $\sup\{R_n(f, \delta_n) : f \in \mathfrak{F}_{a,b}\}$ with δ_n as before. Hence the performance of δ_n is described in Table 1 by

$$\rho^* = \left(\frac{3n^2}{a^2b^2}\right)^{1/3} \sup\{R_n(f_\theta, \delta_n) : 0 \leq |\theta| \leq T_n\}, \quad (2.22)$$

where f_θ and T_n are defined by (2.4) and (2.5). Note that the lower bound ρ_* previously described is actually a lower bound for

$$\left(\frac{3n^2}{a^2b^2}\right)^{1/3} \sup\{R_n(f_\theta, \delta) : 0 \leq |\theta| \leq T_n\}$$

for any estimator δ. In that sense the values of ρ^* are directly comparable with those for this lower bound. The values of ρ^* are determined numerically. For the larger values of n in Table 1 the supremum occurs when $\theta = T_n$; but this is not the case for values of n from 30 through 200 and certain choices of a, b.

The asymptotic expression given in (2.20) is actually an upper bound for all n, as shown in the following lemma.

Lemma 2.1.

$$\sup_{f\in\mathfrak{F}_{a,b}} \left(\frac{3n^2}{a^2b^2}\right)^{1/3} R_n(f, \delta) \leq 1. \quad (2.23)$$

Proof. Note that for any kernel estimator

$$\begin{aligned} R_n(f, \delta) &= \frac{1}{n}\{\mathrm{var}_f(k)\} + (\mathrm{bias}_f(k))^2 \\ &= \frac{1}{n}\left\{\int k^2(x) f(x)\, dx - \left(\int k(x) f(x)\, dx\right)^2\right\} + (\mathrm{bias}_f(k))^2 \\ &\leq \frac{1}{n}\left\{\int k^2(x) f(x)\, dx\right\} + \mathrm{bias}_f(k))^2. \end{aligned} \quad (2.24)$$

It is easy to see that, for k_n as in (2.17), (2.18), and (2.21) and any $f \in \mathfrak{F}_{a,b}$,

$$(\mathrm{bias}_f(k))^2 \leq \left(b \int k(x)|x|\, dx\right)^2,$$

and this value is actually attained whenever $f(x) = a - b|x|$ on $|x| \leq 1/\beta n^{1/3}$. For the values a, b presented in Table 1 the choice f_{T_n} of (2.5) satisfies this condition (and, of course, $f_{T_n} \in \mathfrak{F}_{a,b}$). The other term on the right of (2.24) is obviously bounded for $f \in \mathfrak{F}_{a,b}$ by

$$\frac{1}{n}\int k^2(x) f(x)\, dx \leq \frac{1}{n}\int k^2(x) a\, dx.$$

Accordingly, for δ as before,

$$\begin{aligned} &\left(\frac{3n^2}{a^2b^2}\right)^{1/3} R_n(f, \delta) \\ &\leq \left(\frac{3n^2}{a^2b^2}\right)^{1/3} \left\{\frac{a}{n}\int k^2(x)\, dx + b^2\left(\int |x| k(x)\, dx\right)^2\right\} = 1. \end{aligned}$$

2.6 Description of the Table

Corresponding to each value of n, a, b in Table 1 are two numbers. The lower number is the lower bound $\rho_*(a, b, n)$ for the minimax risk over $\mathfrak{F}_{a,b}$ multiplied by $((3n^2)/(a^2b^2))^{1/3}$; see (2.9) and (2.10). The upper number is the value ρ^* that is $((3n^2)/(a^2b^2))^{1/3}$ times the risk attained by the asymptotically minimax kernel estimator at the least favorable density of the form $f_\theta \in \mathfrak{F}_{a,b}$ $(0 \leq |\theta| \leq T_n)$. The minimax value over all of $\mathfrak{F}_{a,b}$ is less than or equal to 1, as proven in Lemma 2.1. (This minimax value is probably closer to ρ^* than it is to the upper bound 1.)

The values in the table are the first two digits (unrounded) of the decimal expansion. In this way, the entries that appear for ρ_* are indeed lower bounds for the normalized minimax value.

The entries in this table provide a numerical basis for the conclusion, on the one hand, that the lower bound ρ_* cannot be dramatically improved, and, on the other hand, that the asymptotically minimax kernel estimators cannot be dramatically improved. Note that the convergence of the entries in Table 1 to their asymptotic value ($n = \infty$) is fairly slow, but the ratios ρ^*/ρ_* are all less than 3, and are mostly less than or equal to 2. Furthermore, for all (a, b) in Table 1, $\rho_* \geq (\frac{1}{5}) \cdot 1$, where 1 is the upper bound rigorously established in Lemma 2.1. [For values of (a, b) for which $\lambda \doteq T_n[(n\sqrt{3})/(ab)]^{1/3} \ll 4.5$ the values of ρ_* can be much smaller. For example, if $(a, b) = (2, \frac{1}{2})$ and $n = 30$, then $\rho_* = .04$, whereas $\rho^* = .60$. However, here $\lambda = .46$, and, as suggested in Remark 2.2, the Cramer–Rao method should really be applied to a family other than (2.5).]

APPENDIX: A DIFFERENTIAL INEQUALITY

Here is the main result of this section.

Theorem A.1. Let $v_i : [0, T) \to [0, \infty)$ ($i = 1, 2$), with v_1 bounded and measurable and v_2 continuous and nondecreasing. Assume that $v_2(\theta) \geq v_1(\theta)$, $\theta \in [0, T)$. Let both $h : [0, T) \to [0, \infty)$ and $k : [0, T) \to (0, \infty)$ be continuous and bounded, with $h(\theta) > 0$ for $\theta > 0$. Suppose that $q_1 : [0, T) \to (-\infty, \infty)$ is absolutely continuous and satisfies

$$q_1'(\theta) = h(\theta)(v_1(\theta) - q_1^2(\theta))^{1/2} - k(\theta). \qquad \text{(A.1)}$$

Then there is a continuously differentiable function $q_2 : [0, T) \to (-\infty, \infty)$ such that $q_2(0) = q_1(0)$, and

$$q_2'(\theta) = h(\theta)(v_2(\theta) - q_2^2(\theta))^{1/2} - k(\theta) \qquad \text{(A.2)}$$

everywhere on $[0, T)$. Furthermore, $q_2(\theta) \geq q_1(\theta)$, $\theta \in [0, T)$. T).

Remark. In the application at (2.7)–(2.8), $h^2(\theta) = nJ(\theta)$, $k(\theta) = 1 - 2\xi|\theta|/b$, $q_1(\theta) = c(\theta)$, and

$$0 \leq v_1(\theta) = h^{-2}(\theta)(k(\theta) + q_1'(\theta))^2 + q_1^2(\theta) \leq r.$$

Hence (A.1) is satisfied. Then let $v_2(\theta) = r$, and the theorem establishes the desired solubility of the equation

$$r = h^{-2}(\theta)(k(\theta) + q_2'(\theta))^2 + q_2^2(\theta).$$

[This equation is an alternate form of (A.2) and is the same as (2.8) with $q_2(\theta) = c(\theta)$.]

Proof. Suppose that q_2 satisfies (A.2) on $[0, \tau)$, and suppose that $q_2^2(\tau) < v_2(\tau)$. Then for some $\varepsilon > 0$ the solution of (A.2) exists on $[0, \tau + \varepsilon)$ by Picard's theorem.

Now assume that, for some $\varepsilon > 0$, $v_2(\theta) > v_1(\theta) + \varepsilon$, $\theta \in (0, \tau)$, and q_2 satisfies (A.2) with $q_2(0) = q_1(0)$. Then

$$q_2'(\theta) - q_1'(\theta) = h(\theta)[(v_2(\theta) - q_2^2(\theta))^{1/2} - (v_1(\theta) - q_1^2(\theta))^{1/2}]$$

for $\theta \in (0, \tau)$. Hence $q_2'(\theta) - q_1'(\theta) > 0$ whenever $q_2^2(\theta) - q_1^2(\theta) < \varepsilon$, $\theta \in (0, \tau)$. It follows that $q_2(\theta) > q_1(\theta)$, $\theta \in (0, \tau)$.

Continue to assume that $v_2(\theta) > v_1(\theta) + \varepsilon$, $\theta \in (0, \tau)$, and q_2 satisfies (A.2) with $q_2(0) = q_1(0)$. If $0 < \tau' \leq \tau$, then $\lim_{\theta \to \tau'} q_2(\theta) = v_2^{1/2}(\tau')$ is impossible. [This equality would imply that $q_2'(\theta) < 0$ for all $\theta < \tau'$ sufficiently near τ' since v_2 is nondecreasing and k is positive and continuous. But then it would follow that $q_2(\tau') < q_2(\theta') \leq v_2^{1/2}(\theta') \leq v_2^{1/2}(\tau')$ for some $\theta' < \tau$; a contradiction.] It is also true that $q_1(\tau) \leq -v_2^{1/2}(\tau)$ is impossible. [That equality would imply that $v_1(\theta) - q_1^2(\theta) < 0$ for $\theta < \tau$ sufficiently close to τ, a contradiction of (A.1).] It follows that $-v_2^{1/2}(\tau) \leq q_1(\tau) < q_2(\tau) < v_2^{1/2}(\tau)$. Hence the solution of (A.2) exists on $[0, \tau + \varepsilon)$ by the first paragraph of the proof.

It follows that the conclusion of the theorem is valid if $v_2(\theta) > v_1(\theta) + \varepsilon$.

Let $v_i(\theta) = v_2(\theta) + 1/i$ ($i = 3, \ldots$). It follows that for each $i = 3, \ldots$ a solution, q_i, exists to

$$q_i'(\theta) = h(\theta)(v_i(\theta) - q_i^2(\theta))^{1/2} - k(\theta), \qquad \text{(A.3)}$$

since $v_i(\theta) > v_1(\theta) + \varepsilon$ on $(0, T)$. Furthermore, $q_i \geq q_j \geq q_1$, $3 \leq i \leq j$. Let $q_2 = \lim_{j \to \infty} q_j \geq q_1$. Then

$$q_j'(\theta) = h(\theta)(v_j(\theta) - q_j^2(\theta))^{1/2} - k(\theta)$$
$$\to h(\theta)(v_2(\theta) - q_2^2(\theta))^{1/2} - k(\theta).$$

Hence $q_j'(\theta)$ converges. In addition,

$$|q_j'(\theta)| \leq B = \sup\{h(\theta)(v_2(\theta) + \tfrac{1}{3})^{1/2} + k(\theta) : \theta \in [0, T]\}.$$

It follows from the bounded convergence theorem that $\lim_{j \to \infty} q_j'(\theta) = q_2'(\theta)$ a.e., so q_2 satisfies (3.1(2)) a.e. on $(0, T)$. The right side of (A.2) is continuous. Hence q_2' has a continuous version that satisfies (A.2) everywhere.

[*Received January 1988. Revised April 1990.*]

REFERENCES

Altman, N. (1988), "Smoothing Data With Correlated Errors," in *Computing Science and Statistics; Proceedings of the 20th Symposium on the Interface*, eds. E. J. Wegman, D. T. Gantz, and J. J. Miller, Interface Foundation of North America, pp. 246–253.

Birge, L. (1987a), "Estimating a Density Under Order Restrictions: Nonasymptotic Minimax Risk," *The Annals of Statistics*, 15, 995–1012.

——— (1987b), "On the Risk of Histograms for Estimating Decreasing Densities," *The Annals of Statistics*, 15, 1013–1022.

Brown, L. D., and Gajek, L. (1990), "Information Inequalities for the Bayes Risk," *The Annals of Statistics*, 18.

Brown, L. D., and Low, M. (1990), "Information Inequality Bounds on the Minimax Risk (With Application to Nonparametric Regression)," *The Annals of Statistics*, 18.

Casella, G., and Strawderman, W. E. (1981), "Estimating a Bounded Normal Mean," *The Annals of Statistics*, 9, 870–878.

Donoho, D. L., and Liu, R. C. (in press a), "Geometrizing Rates of Convergence, II," *The Annals of Statistics*.

——— (in press b), "Geometrizing Rates of Convergence, III," *The Annals of Statistics*.

Donoho, D. L., Liu, R. C., and MacGibbon, B. (1990), "Minimax Risk for Hyperrectangles," *The Annals of Statistics*, 18, 1416–1437.

Farrell, R. H. (1972), "On the Best Obtainable Asymptotic Rates of Convergence in Estimation of a Density Function at a Point," *Annals of Mathematical Statistics*, 43, 170–180.

——— (1979), "Asymptotic Lower Bounds for the Risk of Estimators of the Value of a Spectral Density Function," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 49, 221–234.

Gajek, L. (1988), "On the Minimax Value in the Scale Model With Truncated Data," *The Annals of Statistics*, 16, 669–677.

Hall, P. (1983), "Large Sample Optimality of Least Squares Cross-Validation in Density Estimation," *The Annals of Statistics*, 11, 1156–1174.

Hodges, Jr., J. L., and Lehmann, E. L. (1951), "Some Applications of the Cramer–Rao Inequality," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA: University of California Press, pp. 13–22.

Liu, Richard C. (1987), "Geometry in Robustness and Nonparametrics," unpublished Ph.D. thesis, University of California, Berkeley.

Marron, J. S. (1987), "A Comparison of Cross-Validation Techniques in Density Estimation," *The Annals of Statistics*, 15, 152–162.

Parzen, E. (1962), "On the Estimation of a Probability Density Function and the Mode," *Annals of Mathematical Statistics*, 33, 1065–1076.

Rosenblatt, M. (1956), "Remarks on Some Nonparametric Estimates of a Density Function," *Annals of Mathematical Statistics*, 27, 832–837.

Sacks, J., and Strawderman, W. (1982), "Improvements on Linear Minimax Estimates," in *Statistical Decision Theory and Related Topics* III

(vol. 2), eds. S. S. Gupta and J. O. Berger, pp. 287–304.
Sacks, J., and Ylvisaker, D. (1981), "Asymptotically Optimum Kernels for Density Estimation at a Point," *The Annals of Statistics,* 9, 334–346.
Silverman, B. W. (1984), "Spline Smoothing: The Equivalent Variable Kernel Method," *The Annals of Statistics,* 12, 898–916.
Stone, C. J. (1984), "An Asymptotically Optimal Window Selection Rule for Kernel Density Estimates," *The Annals of Statistics,* 12, 1285–1297.
Wahba, G. (1975), "Optimal Convergence Properties of Variable Knot, Kernel, and Orthogonal Series Methods for Density Estimation," *The Annals of Statistics,* 3, 15–29.
Wald, A. (1951), "Asymptotic Minimax Solutions of Sequential Point Estimation Problems," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability,* Berkeley, CA: University of California Press, pp. 1–11.