

**Lower Bounds for the Integrated Risk in
Nonparametric Density and Regression Estimation**

By

Mark G. Low

Technical Report No. 223
October 1989

Department of Statistics
University of California
Berkeley, California

**Lower Bounds for the Integrated Risk in
Nonparametric Density and Regression Estimation**

By

Mark G. Low

Abstract

Lower bounds for the minimax risk are found for density and regression functions satisfying a uniform Lipschitz condition. The measure of loss is integrated squared error or squared Hellinger distance. Ratios of known upper bounds to these lower bounds are shown to be as small as two in a specific example.

0. Introduction

Lower bounds for the asymptotic minimax risk under squared error loss have recently been found for a variety of nonparametric problems. The pointwise estimation of regression functions or densities has received particular attention. For a large class of models, ratios of the maximum risk of best linear estimators to the minimax risk has been shown to be close to one. (See Brown and Farrell (1987), Brown and Low (1988), Donoho and Liu (1987, 1988).)

When interest has focused on estimating an entire density or regression function loss functions typically considered are L_1 or integrated squared error. The L_1 approach to density estimation has been summarized by Devroye and Györfi (1985).

For an integrated squared error loss Stone (1982) has established best asymptotic rates for nonparametric regression and density estimation. For certain ellipsoidal parameter spaces, Efroimovich and Pinsker (1982) and Nussbaum (1985) found first order asymptotically optimal solutions.

In this paper we consider density or regression functions satisfying a uniform Lipschitz condition

$$(0.1) \quad |f(x) - f(y)| \leq M|x - y|.$$

We denote by $R(M)$ the set of functions defined on $[-\frac{1}{2}, \frac{1}{2}]$ which satisfy (0.1). The subset of all density functions which belong to $R(M)$ will be written $D(M)$. These parameter spaces are not ellipsoidal. In Section 1 we derive lower bounds for the minimax risk over the classes $D(M)$ and $R(M)$ for density estimation and nonparametric regression respectively. For density estimation we look at both integrated squared error loss and squared Hellinger distance losses. For nonparametric regression we only consider integrated squared error. We have restricted attention to densities and regression functions satisfying (0.1) so that we can compare these lower bounds to known upper bounds. This is done in Section 2. We should point out, however, that the method we use to construct lower bounds can be applied in a variety of other contexts even if sometimes more ingenuity is needed. Essential to our arguments is a knowledge of good lower bounds for the corresponding pointwise estimation problem. The proof of Theorem 1 then shows how to connect the pointwise estimation problem to the global estimation problem.

The results in this paper should therefore be understood as part of an ongoing effort to find general techniques for bounding the minimax risk in nonparametric problems. See for example Donoho and Johnstone (1989). The contribution of this paper is to show how to connect local problems to global problems.

1. Lower Bounds

a) Local

We consider two nonparametric statistical problems. In density estimation we observe i.i.d. random variables X_1, \dots, X_n each with density $f \in \mathcal{D}(M)$. In the regression problem observe $Y_i \sim N(r(y_i), 1)$, independent, $y_i = \frac{i}{n+1}$, $i = 1, \dots, n$, where $r \in \mathcal{R}(M)$. Estimators of f or r will be written \hat{f}_n or \hat{r}_n where the subscript n will indicate that \hat{f}_n and \hat{r}_n are functions of X_1, \dots, X_n and Y_1, \dots, Y_n .

Before establishing lower bounds for estimating an entire density or regression function we need corresponding results for the pointwise problem. This local problem has recently been addressed by Donoho and Liu (1987, 1988). In particular the proposition given below is essentially contained in these papers. For this reason we only give a brief outline of a proof here.

In both the pointwise and global estimation problems the lower bounds are expressed in terms of the minimax risk for the bounded normal mean problem. Let $\rho(d, \sigma^2)$ be the minimax risk of estimating θ from one observation for the family $\{N(\theta, \sigma^2), |\theta| \leq d\}$.

We write f_θ^n and s_θ^n for densities and regression functions on the interval $[-\frac{1}{2}, \frac{1}{2}]$ defined by

$$\begin{aligned} f_\theta^n(x) &= \left[1 + g_\theta^n(x)\right] (C_n(\theta))^{-1} \\ s_\theta^n(x) &= g_\theta^n(x) \end{aligned}$$

where

$$(1.1) \quad g_\theta^n(x) = \theta \left[1 - \frac{|x|}{Dn^{-1/3}}\right]_+.$$

$$(1.2) \quad C_n(\theta) = (1 + \int g_\theta^n(x) dx) = 1 + \theta Dn^{-1/3}.$$

Let

$$(1.3) \quad d = \left[\frac{2}{3}\right]^{1/2} M D^{3/2}$$

Proposition

a) If X_1, \dots, X_n are i.i.d. random variables with common density f_θ^n , then

$$(1.4) \quad \limsup_{n \rightarrow \infty} n^{2/3} \inf_{\hat{\theta}_n} \sup_{|\theta| \leq M D n^{-1/3}} E_{f_\theta^n} (\theta - \hat{\theta}_n)^2 = M^{2/3} \left[\frac{3}{2}\right]^{2/3} d^{-2/3} \rho(d, 1).$$

b) If Y_1, \dots, Y_n are independent $N(s_0^n(y_i), 1)$ then

$$(1.5) \quad \limsup_{n \rightarrow \infty} n^{2/3} \inf_{\hat{\theta}_n} \sup_{|\theta| \leq MDn^{-1/3}} E_{s_0^n} (\theta - \hat{\theta}_n)^2 = M^{2/3} \left(\frac{3}{2} \right)^{2/3} d^{-2/3} \rho(d, 1).$$

Proof

Let $\psi = \left(\frac{2}{3} \right)^{1/2} D^{1/2} n^{1/3} \theta$ and P_ψ^n be the probability with density $\prod_1^n f_\theta^n(x_i)$ with respect to Lebesgue measure on R^n . Straightforward calculations show that the experiments $\{P_\psi^n : |\psi| \leq \left(\frac{2}{3} \right)^{1/2} MD^{3/2}\}$ converge to the experiment $\{N(\psi, 1) : |\psi| \leq \left(\frac{2}{3} \right)^{1/2} MD^{3/2}\}$. Similar calculations may be found in Donoho and Liu (1988). Hence

$$(1.6) \quad \lim_{n \rightarrow \infty} n^{2/3} \inf_{\hat{\theta}_n} \sup_{|\theta| \leq MDn^{-1/3}} E_{f_\theta^n} (\theta - \hat{\theta}_n)^2 = \left(\frac{3}{2} \right) D^{-1} \rho \left(\left(\frac{2}{3} \right)^{1/2} MD^{3/2}, 1 \right).$$

$$\text{Since } d = \left(\frac{2}{3} \right)^{1/2} MD^{3/2}$$

$$(1.7) \quad D^{-1} = \left(\frac{2}{3} \right)^{1/3} M^{2/3} d^{-2/3}$$

(1.6) and (1.7) taken together yield (1.4). (1.5) is established in a similar way.

b) Global

To obtain our global lower bounds we construct perturbations of a fixed density f_0 and regression s_0 . We choose

$$f_0 : [0, 1] \rightarrow R, \quad f_0(x) \equiv 1$$

and

$$s_0 : [0, 1] \rightarrow R, \quad s_0(x) \equiv 0.$$

Let $D > 0$ and $\beta_n(D) = [n^{1/3}/2D]$ where $[\]$ denotes the greatest integer less than. Our subfamilies of interest are

$$f_{\underline{\theta}^n}(x) = \left[f_0(x) + \sum_1^{\beta_n} g_{\theta_i^n}(x - (2i-1)Dn^{-1/3}) \right] (C^{(n)}(\underline{\theta}^n))^{-1}$$

$$\underline{\theta}^n = (\theta_1, \dots, \theta_{\beta_n}), \quad |\theta_i| \leq MDn^{-1/3}, \quad i = 1, \dots, \beta_n$$

$$C^{(n)}(\underline{\theta}^n) = \sum_{i=1}^{\beta_n} C_n(\theta_i)$$

$$s_{\underline{\theta}^n}(x) = s_0(x) + \sum_{i=1}^{\beta_n} g_{\theta_i}^n(x - (2i-1)Dn^{-1/3})$$

where g_{θ}^n and $C_n(\theta)$ are defined in (1.1) and (1.2).

Lower bounds for the global problem will be written in terms of $\sup_d d^{-2/3} p(d, 1)$. This is facilitated by a recent study of $m(q) = \sup_d d^{2q-2} p(d, 1)$ by Donoho and Liu (1987). They found $m\left[\frac{2}{3}\right] = \sup_d d^{-2/3} p(d, 1) = 0.450$

Theorem 1 Let X_1, \dots, X_n be i.i.d. random variables with density $f \in \mathbf{D}(M)$. Then

a)

$$(1.6) \quad \limsup_{n \rightarrow \infty} n^{2/3} \inf_{\hat{f}_n} \sup_{f \in \mathbf{D}(M)} E_f \left[\int_0^1 (f(x) - \hat{f}_n(x))^2 dx \right] \geq \left[\frac{1}{12} \right]^{1/3} M^{2/3} \sup_d d^{-2/3} p(d, 1) \\ \approx \left[\frac{1}{12} \right]^{1/3} M^{2/3} 0.45$$

b)

$$(1.7) \quad \limsup_{n \rightarrow \infty} n^{2/3} \inf_{\hat{f}_n} \sup_{f \in \mathbf{D}(M)} E_f \left[\int_0^1 \left[\sqrt{f} - \sqrt{\hat{f}_n(x)} \right]^2 dx \right] \geq \frac{1}{4} \left[\frac{1}{12} \right]^{1/3} M^{2/3} \sup_d d^{-2/3} p(d, 1).$$

Proof

$$a) \quad \text{Let } R_i(\underline{\theta}^n, \hat{f}) = \int_{(2i-2)Dn^{-1/3}}^{2iDn^{-1/3}} (f_{\theta_i^n}(x) - \hat{f}(x))^2 dx \text{ and}$$

$\theta^i(\underline{\theta}^n) = (0, \dots, 0, \theta_i, 0, \dots, 0)$. Then

$$\inf_{\hat{f}_n} \sup_{f \in \mathbf{D}(M)} E_f \left[\int_0^1 (f_n(x) - \hat{f}(x))^2 dx \right] \geq \inf_{\hat{f}_n} \sup_{\underline{\theta}^n} E_f \sum_{i=1}^{\beta_n} R_i(\underline{\theta}^n, \hat{f}_n) \\ \geq \sum_{i=1}^{\beta_n} \inf_{\hat{f}_n} \sup_{\theta_i \neq \theta_i} R_i(\underline{\theta}^n, \hat{f}_n) \\ \sim \beta_n \inf_{\hat{f}_n} \sup_{\theta_i} E_{\theta^i(\underline{\theta}^n)} \left[\int_0^1 (g_{\theta_i}^n(x) - \hat{f}_n(x))^2 dx \right]$$

since $|C^{(n)}(\underline{\theta}^n) - 1| \rightarrow 0$ as $n \rightarrow \infty$, uniformly for all $\underline{\theta}^n$ s.t. $|\theta_i| \leq MDn^{-1/3}$. Hence

$$\begin{aligned}
 \inf_{\hat{f}_n} \sup_{f \in D(M)} E_f \left[\int_0^1 (f(x) - \hat{f}_n(x))^2 dx \right] &\geq \beta_n \int \left[\inf_{\hat{f}_n} \sup_{\theta_1} E_{\theta^1(\theta^n)} (g_{\theta_1}^n(x) - \hat{f}_n(x))^2 \right] dx \\
 &= \beta_n \inf_{\hat{\theta}(x)} \sup_{\theta_1} \int \left[1 - \frac{|x|}{Dn^{-1/3}} \right]_+^2 E_{\theta^1(\theta^n)} (\theta_1 - \hat{\theta}(x))^2 dx \\
 &= \frac{n^{1/3}}{2D} \cdot \frac{2}{3} Dn^{-1/3} M^{2/3} \left[\frac{3}{2} \right]^{2/3} d^{-2/3} \rho(d, 1)(1 + o(1)).
 \end{aligned}$$

where d is defined in (1.3). Now (1.6) follows on taking \sup_d .

(1.7) Follows from the observation $\sqrt{1 + \theta} = 1 + \frac{1}{2} \theta + o(\theta^2)$ and the analysis given in a).

Theorem 2 Let $Y_i = r(y_i) + e_i$ where e_i are i.i.d. $N(0, 1)$ random variables, $r \in R(M)$ and y_i are equally spaced in the interval $[0, 1]$. Then

$$\lim_{n \rightarrow \infty} n^{2/3} \inf_{\hat{f}_n} \sup_{r \in R(M)} E_r \left[\int_0^1 (\hat{f}_n(x) - r(x))^2 dx \right] \geq \left[\frac{1}{12} \right]^{1/3} M^{2/3} \sup_d d^{-2/3} \rho(d, 1)$$

Proof

Replace f by r and D by R in the proof of Theorem 1.

2. Upper Bounds

Throughout this section minimax risk refers to an integrated squared error loss. An obvious but crude upper bound for the minimax risk can be given in terms of the asymptotic linear minimax risk for the pointwise estimation problem under squared error loss. Let $D(M, 1)$ be the class of densities such that $f \in D(M)$ and $f(x_0) \leq 1$. Then clearly the asymptotic minimax risk of the best linear estimator for estimating $f(x_0)$ when f is assumed to belong to $D(M, 1)$ is an upper bound for the minimax risk under integrated squared loss over the class $D(M)$. Similarly in the regression context the asymptotic linear minimax risk over the class $R(M)$ for the pointwise problem is an upper bound for the minimax risk of the global problem.

For the classes $D(M, 1)$ and $R(M)$ the asymptotic best linear estimators for the pointwise estimation of density and regression functions can be found in Sacks and Ylvisaker (1978, 1981). Alternatively the hardest linear subfamily methodology of Donoho and Liu (1987, 1988) also yields these best linear estimators.

The asymptotic minimax risk for these best linear estimators over the classes $D(M, 1)$ and $R(M)$ for the density and regression problems respectively is the same and is given by

$$(2.1) \quad BL(M) n^{-2/3}$$

where $BL(M) = M^{2/3} \left[\frac{1}{3} \right]^{1/3}.$

The ratio of these upper bounds to the lower bounds given in Theorem 1 and Theorem 2 is given by $\frac{M^{2/3} (\frac{1}{3})^{1/3}}{M^{2/3} (\frac{1}{12})^{1/3} 0.45} \approx 3.5.$

The upper bounds given in (2.1) are quite conservative. For the class of densities

$$(2.3) \quad EP(B) = \{f: [0, 1] \rightarrow R, f \geq 0, \int f = 1, \int f'^2 \leq B^2\}.$$

Efroimovich and Pinsker (1982) found the asymptotic minimax risk. Nussbaum (1985) did likewise for the regression problem with

$$N(M) = \{f: [0, 1] \rightarrow R, \int f'^2 \leq B^2\}.$$

Since $D(M, 1) \subseteq EP(M)$ and $R(M, 1) \subseteq N(M)$, these numbers are also upper bounds for the classes $D(M)$ and $R(M)$. The ratios of these upper bounds to our lower bounds are the same and equal to

$$\frac{\left[\frac{3}{4} \pi^2 \right]^{1/3}}{\left[\frac{1}{12} \right]^{1/3} \cdot 0.45} \approx 2.2.$$

Moreover, we believe the upper bounds derived from Efroimovich and Pinsker are an overestimate of the minimax risk.

Conjecture

$$\text{Let } \alpha_i = (-1)^i D n^{-1/3}, N_n = 2 \left[\frac{n^{1/3}}{4D} \right]$$

$$\hat{f}_n(x) = \frac{1}{n D n^{-1/3}} \sum_{i=1}^n \left[1 - \frac{|x - X_i|}{D n^{-1/3}} \right]_+ \quad D n^{-1/3} \leq x \leq 1 - D n^{-1/3}$$

$$\hat{f}_n(x) = \hat{f}_n(D n^{-1/3}) \quad 0 \leq x \leq D n^{-1/3}$$

$$\hat{f}_n(x) = \hat{f}_n(1 - D n^{-1/3}) \quad 1 - D n^{-1/3} \leq x \leq 1$$

and

$$f_n(x) = 1 + \sum_{i=1}^{N_n} g_{\alpha_i}^n(x - (2i - 1)Dn^{-1/3}).$$

Then we conjecture

$$(2.4) \quad \lim_{n \rightarrow \infty} n^{2/3} \sup_{f \in D(I, 1)} E_f \left[\int_0^1 (\hat{f}_n(x) - f(x))^2 dx \right] = \lim_{n \rightarrow \infty} n^{2/3} E_{f_n} \left[\int_0^1 (\hat{f}_n(x) - f_n(x))^2 dx \right].$$

Calculation of the right hand side of (2.4) is straightforward and yields

$$(2.5) \quad \frac{D^2}{63} + \frac{2}{3D}$$

(2.5) takes its minimum value of $\frac{1}{(21)^{1/3}}$ when $D = (21)^{1/3}$. This gives a ratio

$$\frac{(21)^{-1/3}}{(\frac{1}{12})^{1/3} \cdot 0.45} \cong 1.84.$$

Acknowledgements. This work is based on part of the author's doctoral dissertation, which was written at Cornell University under the supervision of Professor L.D. Brown. His encouragement is gratefully acknowledged.

References

- Brown, L.D. and Farrell, R.H. (1987) A lower bound for the risk in estimating the value of a probability density. Cornell Statistics Center preprint series.
- Brown, L.D. and Low, M.G. (1988) Information inequality bounds on the minimax risk (with an application to nonparametric regression). Cornell Statistics Center preprint series.
- Casella, G. and Strawderman, W.E. (1981) Estimating a bounded normal mean. *Ann. Stat.* **9**, 870-878.
- Devroye, L. and Györfi, L. (1985) *Nonparametric density estimation: The L^1 view*. Wiley, New York.
- Donoho, D.L. and Johnstone, I. (1989). Minimax Risk over l_p -Balls. Technical Report, University of California, Berkeley.
- Donoho, D.L. and Liu, R.C. (1987) On the minimax estimation of linear functions. Technical Report, University of California, Berkeley.
- Donoho, D.L. and Liu, R.C. (1988) Geometrizing rates of convergence, III. Technical Report, University of California, Berkeley.
- Efroimovich, S.Y. and Pinsker, M.S. (1982) Estimation of square-integrable probability density of a random variable (1983), 175-189.
- Nussbaum, M. (1985) Spline smoothing in regression models and asymptotic efficiency in L_2 . *Ann. Stat.* **13**, 984-997.
- Sacks, J. and Ylvisaker, D. (1978) Linear estimation for approximately linear models. *Ann. Stat.* **6**, 1122-1137.
- Sacks, J. and Ylvisaker, D. (1981) Asymptotically optimum kernels for density estimation at a point. *Ann. Stat.* **9**, 334-346.
- Stone, C. (1982) Optimal global rates of convergence for nonparametric regression. *Ann. Stat.* **10**, **4**, 1040-1053.

Department of Statistics
University of California
Berkeley, California