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).


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

\[ |f(x) - f(y)| \leq M|x - y|. \] (0.1)

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 $X_1, \ldots, X_n$ each with density $f \in \mathcal{D}(M)$. In the regression problem observe $Y_i \sim N(r(y_i), 1)$, independent, $y_i = \frac{i}{n+1}$, $i = 1, \ldots, n$, where $r \in \mathbb{R}(M)$. 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 $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$.

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 $\theta$ from one observation for the family $\{N(\theta, \sigma^2), |\theta| \leq d\}$.

We write $f_\theta^0$ and $s_\theta^0$ for densities and regression functions on the interval $[-\frac{1}{2}, \frac{1}{2}]$ defined by

$$f_\theta^0(x) = \left[1 + g_\theta^0(x)\right] (C_n(\theta))^{-1}$$

$$s_\theta^0(x) = g_\theta^0(x)$$

where

$$g_\theta^0(x) = \theta \left[1 - \frac{|x|}{D_n^{-1/3}}\right]_+$$

$$C_n(\theta) = (1 + \int g_\theta^0(x) \, dx) = 1 + \theta D_n^{-1/3}.$$  

Let

$$d = \left[\frac{2}{3}\right]^{1/2} MD_n^{3/2}$$

Proposition

a) If $X_1, \ldots, X_n$ are i.i.d. random variables with common density $f_\theta^0$, then

$$\lim_{n \to \infty} \sup_{\theta_0} \frac{1}{n^{2/3}} \inf_{\theta, \theta_0} \frac{1}{d} E_{f_\theta^0}[(\theta - \hat{\theta}_n)^2] = M^{2/3} \left[\frac{3}{2}\right]^{2/3} d^{-2/3} \rho(d, 1).$$
b) If $Y_1, \ldots, Y_n$ are independent $N(s_i^n(y_i), 1)$ then

$$\limsup_{n \to \infty} n^{2/3} \inf_{\delta_n} \sup_{1 \leq M \leq D_n^{-1/3}} \mathbb{E}_{s_n^n}(\theta - \hat{\theta}_n)^2 = M^{2/3} \left( \frac{3}{2} \right)^{2/3} d^{-2/3} \rho(d, 1).$$

**Proof**

Let $\psi = \left( \frac{2}{3} \right)^{1/2} D^{1/2} n^{1/3} \theta$ and $P^n_{\psi}$ be the probability with density $\prod_{i=1}^{n} f_0^n(x_i)$ with respect to Lebesgue measure on $\mathbb{R}^n$. Straightforward calculations show that the experiments $\{P^n_{\psi}: |\psi| \leq \left( \frac{2}{3} \right)^{1/2} M D^{3/2}\}$ converge to the experiment $\{N(\psi, 1): |\psi| \leq \left( \frac{2}{3} \right)^{1/2} M D^{3/2}\}$. Similar calculations may be found in Donoho and Liu (1988). Hence

$$\lim n^{2/3} \inf_{\delta_n} \sup_{1 \leq M \leq D_n^{-1/3}} \mathbb{E}_{s_n^n}(\theta - \hat{\theta}_n)^2 = \left[ \frac{3}{2} \right] D^{-1} \rho \left[ \left( \frac{2}{3} \right)^{1/2} M D^{3/2}, 1 \right].$$

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

$$D^{-1} = \left[ \frac{2}{3} \right]^{1/3} M^{2/3} d^{-2/3}$$

(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

$$f_0: [0, 1] \to \mathbb{R}, \quad f_0(x) = 1$$

and

$$s_0: [0, 1] \to \mathbb{R}, \quad s_0(x) = 0.$$ 

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

$$f_{\theta^n}(x) = \left[ f_0(x) + \sum_{i=1}^{\beta_n} g^n_{\theta_i}(x - (2i - 1) D n^{-1/3}) \right] (C^n(\theta^n))^{-1}$$

$$\theta^n = (\theta_1, \ldots, \theta_{\beta_n}), \quad |\theta_i| \leq MDn^{-1/3}, \quad i = 1, \ldots, \beta_n$$
\[ C^n(\Theta^n) = \sum_{i=1}^{\beta_n} C_n(\theta_i) \]

\[ s_0(x) = s_0(x) + \sum_{i=1}^{\beta_n} g^n_i(x - (2i - 1) D n^{-1/3}) \]

where \( g^n_i \) and \( C_n(\theta) \) are defined in (1.1) and (1.2).

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

**Theorem 1** Let \( X_1, \ldots, X_n \) be i.i.d. random variables with density \( f \in D(M) \). Then

\[ \limsup_{n \to \infty} n^{2/3} \inf_{\hat{f} \in D(M)} \sup_{f \in D(M)} \mathbb{E}_f \left[ (f(x) - \hat{f}_n(x))^2 \right] \geq \left[ \frac{1}{12} \right]^{1/3} M^{2/3} \sup d^{-2/3} p(d, 1) \]

\[ \approx \left[ \frac{1}{12} \right]^{1/3} M^{2/3} 0.45 \]

b)

\[ \limsup_{n \to \infty} n^{2/3} \inf_{\hat{f} \in D(M)} \sup_{f \in D(M)} \mathbb{E}_f \left[ (\sqrt{\hat{f}} - \sqrt{\hat{f}_n(x)})^2 \right] \geq \frac{1}{4} \left[ \frac{1}{12} \right]^{1/3} M^{2/3} \sup d^{-2/3} p(d, 1) \]

**Proof**

a) Let \( R_i(\Theta^n, \hat{f}) = \int_{(2i-2)D n^{-1/3}}^{(2i-1)D n^{-1/3}} (f_{\Theta^n}(x) - \hat{f}(x))^2 dx \) and

\( \Theta^n(\Theta^n) = (0, \ldots, 0, \theta_i, 0, \ldots, 0) \). Then

\[ \inf_{\hat{f} \in D(M)} \sup_{f \in D(M)} \mathbb{E}_f \left[ (f_n(x) - \hat{f}(x))^2 \right] \geq \inf_{\hat{f} \in D(M)} \sup_{f \in D(M)} \mathbb{E}_f \left[ (f_{\Theta^n}(x) - \hat{f}_n(x))^2 \right] \]

\[ \geq \sum_{i=1}^{\beta_n} \inf_{\hat{f} \in D(M)} \mathbb{E}_f \left[ (f_{\Theta^n}(x) - \hat{f}_n(x))^2 \right] \]

\[ \geq \beta_n \inf_{\hat{f} \in D(M)} \mathbb{E}_f \left[ (f_{\Theta^n}(x) - \hat{f}_n(x))^2 \right] \]

since \( |C^n(\Theta^n) - 1| \to 0 \) as \( n \to \infty \), uniformly for all \( \Theta^n \) s.t. \( |\theta_i| \leq MDn^{-1/3} \). Hence
\[
\inf_{\hat{f}_n} \sup_{f \in D(M)} E_n \left[ \int_0^1 (f(x) - \hat{f}_n(x))^2 \, dx \right] \geq \beta_n \int \left[ \inf_{\hat{f}_n} \sup_{\theta \in \Theta} E_{\theta, \hat{f}_n} \left( g_{\theta, \hat{f}_n}(x) - \hat{f}_n(x) \right)^2 \right] \, dx
\]

\[
= \beta_n \inf_{\theta} \sup_{\theta \in \Theta} \int \left[ 1 - \frac{|x|}{D_n^{-1/3}} \right]^2 E_{\theta, \hat{f}_n} (\theta_1 - \hat{\theta}(x))^2 \, dx
\]

\[
= \frac{n^{1/3}}{2D} \cdot \frac{2}{3} Dn^{-1/3} M^{2/3} \left[ \frac{3}{2} \right]^{2/3} d^{-2/3} \rho(d, 1)(1 + o(1)).
\]

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

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

**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

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

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

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

The upper bounds given in (2.1) are quite conservative. For the class of densities $E_P(B) = \{ f: [0, 1] \rightarrow \mathbb{R}, f \geq 0, \int f = 1, \int f^2 \leq B^2 \}$. Efroimovich and Pinsker (1982) found the asymptotic minimax risk. Nussbaum (1985) did likewise for the regression problem with $N(M) = \{ f: [0, 1] \rightarrow \mathbb{R}, \int f^2 \leq B^2 \}$.

Since $D(M, 1) \subseteq E_P(M)$ 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

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

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

**Conjecture**

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

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

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

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

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

Then we conjecture

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

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

\[ \frac{D^2}{63} + \frac{2}{3D} \]

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

\[ \frac{(21)^{-1/3}}{(\frac{1}{12})^{1/3} \cdot 0.45} \equiv 1.84. \]

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


Department of Statistics
University of California
Berkeley, California