A REMEDY TO REGRESSION ESTIMATORS AND NONPARAMETRIC MINIMAX EFFICIENCY

Jianqing Fan
Department of Statistics
University of North Carolina
Chapel Hill, N.C. 27514

May 7, 1990

Abstract

It is known that both Watson-Nadaraya and Gasser-Müller types of regression estimators have some disadvantages. A smooth version of local polynomial regression estimators are proposed to remedy the disadvantages. The mean squared error and mean integrated squared errors are computed explicitly. It turns out that by suitably selecting a kernel and a bandwidth, the proposed estimator has at least asymptotic *minimax efficiency* 89.6%—proposed estimator is *efficient in rates* and *nearly efficient in constant factors*!

In nonparametric regression context, the asymptotic minimax lower bound is developed via the heuristic of the "hardest 1-dimensional subproblem". The explicit connections of minimax risks with modulus of continuity are made. Normal submodels are used to avoid the technical difficulty of Le Cam's theory of convergence of experiments. The lower bound is applicable for estimating conditional mean (regression) and conditional quantiles (including median) for both fixed design and random design regression problems.

Abbreviated title. Minimax nonparametric regression.

AMS 1980 subject classification. Primary 62G20. Secondary 62G05, 62F35.

Key words and phrases. Hardest 1-dimensional subproblem; Local polynomials; Minimax risk; Modulus of continuity; Nonparametric regression; Optimal rates of convergence.

1 Introduction

Nonparametric regression provides a useful tool for data analysis. A useful model is to think of estimating a regression function

$$m_f(x_0) = E_f(Y|X = x_0), \tag{1.1}$$

based on a random sample of data $(X_1, Y_1), \cdots, (X_n, Y_n)$ from an unknown joint density $f(\cdot, \cdot)$. For convenience of notation, we will suppress the dependence of the regression function $m_f(\cdot)$ on f. Several approaches have been proposed in the literature. Let $K(\cdot)$ denote a kernel function, and h_n be a bandwidth.

1. Watson-Nadaraya estimator:

$$\hat{m}_{WN}(x_0) = \left[\sum_1^n K\left(\frac{x_0 - X_j}{h_n}\right) Y_j\right] \left[\sum_1^n K\left(\frac{x_0 - X_j}{h_n}\right)\right]^{-1} \tag{1.2}$$

2. Gasser-Müller estimator:

$$\hat{m}_{GM}(x_0) = \sum_1^n Y_j' \int_{t_{j-1}}^{t_j} \frac{1}{h_n} K\left(\frac{x_0 - t}{h_n}\right) dt, \tag{1.3}$$

   where $\{(X_j', Y_j')\}$ are ordered sample according to X_j's, $t_0 = -\infty$, $t_n = \infty$, and $t_j = \frac{X_{j+1}' + X_j'}{2}$.

3. Stone's local polynomial. To formulate the idea, let's assume that $m(x)$ has two derivatives. Stone (1977, 80) suggests to use the estimate $\hat{m}_S(x_0)$, which minimizes

$$\sum_{\{|x_0 - X_j| \leq h_n\}} (Y_j - m_S - b(x_0 - X_j))^2, \tag{1.4}$$

   with respect to m_S, and b.

All of these estimators have some advantages and disadvantages in terms of asymptotic biases and variances. In particular, unless the marginal density f_X is uniformly distributed, the asymptotic bias of a Watson-Nadaraya estimator depends on the derivatives of the marginal density f_X, while the asymptotic variance of a Gasser-Müller estimator is 1.5

times as large as that of a Watson-Nadaraya estimator. Similar comments apply to a fixed design regression problem. See Chu and Marron (1990) for a detailed discussion. In terms of asymptotic minimax theory, we will see *none of these types* of estimates can be a candidate of a nearly asymptotic minimax estimator, unless in a very special circumstance.

It is our purpose to propose a new class of estimates, which take advantages of the above estimators. Theoretical advantages of new estimators involve getting rid of the dependence of biases on the derivatives of the marginal density, while keeping variances the same as Watson-Nadaraya estimates. Comparing with the new class of estimators, *Gasser-Müller estimators are not admissible!* It is worthwhile to point out that the dependence of biases on the derivatives of the marginal density f_X is *not any intrinsic part* of nonparametric regression, but *an artifact* of Watson-Nadaraya kernel method, hence the *asymptotic minimax efficiency of Watson-Nadaraya estimators can be arbitrarily small.* See Remark 2. In contrast, we will show that the best estimator in the new class can be as high as 89.6% asymptotic minimax efficiency.

It is emphasized that the new class of estimators have practical advantages, too. The MSE of the new class of estimators does not depend on the derivatives of the marginal density $f_X(\cdot)$. Thus, data-driven bandwidth selection will not involve the effort for estimating the derivative of the marginal density. This makes the data-driven bandwidth much easier.

Another important motivation of introducing the new class of estimators is to make (nearly) precise minimax risk in the regression setup. With an optimal choice of kernel and bandwidth, the estimator provides a good upper bound for minimax risk. The lower bound is derived by using the heuristic of the "hardest 1-dimensional subproblem". In particular, a geometric quantity—modulus of continuity is involved in both lower and upper bounds. We show that the minimax lower bound is *nearly sharp* for some cases:

- bounded-two-derivative constraints (see (3.3)).
- bounded Lipschitz constraints (see Example in section 5.1).

We would expect, but have not shown yet that such a lower bound is nearly sharp for other constraints.

We decompose the difficulty of nonparametric regression into two parts: the constraints on the regression function itself, and the constraints on the marginal density and conditional variance. It turns out that the upper bound of conditional variance and the lower bound of marginal density in the constraint class of joint densities are strongly related to minimax risks. An important application of our lower bound is that one could use it to determine how efficient an estimator is for a regression problem (see section 5.1). Even though our attentions are focused on randomly designed problems whose marginal densities are also unknown, the lower bound is applicable for both fixed design problems, and random design problems whose marginal distributions are known (e.g. uniform distributions). See Remark 3.

Our approach is related to other work in the literature, and in particular the work in white noise models and density estimation models. See section 5.2 for further references. What seems innovative in our approach is the use of normal submodel to avoid the technical difficulty of the theory of convergence of experiments (Le Cam (1985)).

The paper is organized as follows. Section 2 introduces a class of smooth local polynomial estimators, whose mean squared error and mean integrated squared error are computed in section 3. We use risks of this class of regression estimators as upper bounds of the minimax risks. The minimax problems are studied in section 4, where attentions on the lower bound are particularly focused. Potential applications of lower bound are discussed in section 5, where closely related work in minimax theory is cited. Proof is deferred in section 6.

2 Smooth local polynomials

Let's extend the idea of local polynomial. A similar idea can be found in Stone (1977), and Cleveland (1979), Müller (1987). Assume that we know that the second derivative of $m(x)$ exists. Our proposal is to construct a smooth version of local polynomial: finding a, b to minimize

$$\sum_{1}^{n}(Y_j - a - b(x_0 - X_j))^2 K\left(\frac{x_0 - X_j}{h_n}\right). \tag{2.1}$$

Let $\hat{a}, \hat{b}$ be the solution to the weighted least square problem (2.1). We propose to use $\hat{m}(x_0) = \hat{a}$ to estimate the regression function $m(x_0)$. Simple calculation yields that

$$\hat{m}(x_0) = \sum_{1}^{n} w_j y_j / \sum_{1}^{n} w_j, \tag{2.2}$$

with

$$w_j \equiv K\left(\frac{x_0 - X_j}{h_n}\right)[s_{n,2} - (x_0 - X_j)s_{n,1}], \tag{2.3}$$

where

$$s_{n,l} = \sum_{1}^{n} K\left(\frac{x_0 - X_j}{h_n}\right)(x - X_j)^l. \tag{2.4}$$

A nice feature of the estimate (2.2) is that the weight function satisfies (which is a property of least square estimate, and can easily be checked)

$$\sum_{1}^{n}(x_0 - X_j)w_j = 0. \tag{2.5}$$

This property ensures that the bias of the estimate does not involve the derivatives of the marginal density. To see this, we note that by (2.5)

$$E\hat{m}(x_0) = m(x_0) + E\sum_{1}^{n}(m(X_j) - m(x_0) + m'(x_0)(X_j - x_0))\, w_j / \sum_{1}^{n} w_j.$$

If we do Taylor expansions for $m(X_j)$ at point x_0, the second term would be of order $O(h_n^2)$, as effective design points have order $(X_j - x_0)^2 = O(h_n^2)$. Thus no derivatives of $f_X(\cdot)$ are involved in the above calculation (rigorous proof can be found in the proof of Theorem 1).

Let's briefly mention how the idea above can be extended to the case where $m(x)$ has a bounded kth derivative. The idea is exactly the same, except replacing the linear polynomial in (2.1) to a $k-1$ order polynomial. In particular, when $m(x)$ has one derivative, one finds a minimizer of

$$\sum_{1}^{n}(Y_j - a)^2 K\left(\frac{x_0 - X_j}{h_n}\right), \tag{2.6}$$

and resulting estimate is exactly the same as Watson-Nadaraya estimator (1.2). In other words, we use this class of estimate only when the constraint on the unknown regression function m has a bounded derivative.

3 Asymptotic properties

Let's discuss the asymptotic properties of the estimator (2.2). Assumptions are as follows.

Condition 1.

i. The regression function $m(\cdot)$ has a bounded second derivative.

ii. The marginal density $f_X(\cdot)$ of X is bounded and continuous with $f_X(x_0) > 0$.

iii. The conditional variance of $\sigma^2(x) = \mathrm{Var}(Z|X = x)$ is bounded and continuous.

iv. The kernel $K(\cdot)$ is bounded and continuous, satisfying

$$\int_{-\infty}^{\infty} K(y)dy = 1, \int_{-\infty}^{\infty} yK(y)\,dy = 0, \int_{-\infty}^{\infty} y^2K(y)\,dy \neq 0,$$
$$\int_{-\infty}^{\infty} y^3K(y) < \infty, \sup_k |y^3K(y)| < \infty.$$

Theorem 1. Under condition 1, if $h_n \to 0$, and $nh_n \to \infty$, then the estimator (2.2) has the mean-squared error:

$$\begin{aligned} E(\hat{m}(x_0) - m(x_0))^2 &= \frac{1}{4}\left(m''(x_0)\int_{-\infty}^{\infty} u^2K(u)du\right)^2 h_n^4 \\ &\quad + \frac{1}{nh_n} f^{-1}(x_0)\sigma^2(x_0)\int_{-\infty}^{\infty} K^2du + o(h_n^4 + \frac{1}{nh_n}) \end{aligned} \quad (3.1)$$

Let $w(\cdot)$ be a bounded weight function with a compact support $[a, b]$. Then the Mean Integrated Squared Error (MISE) can be computed as follows.

Theorem 2. Under condition 1, if $f_X(\cdot)$ is bounded away from zero in the interval $[a, b]$, then the MISE is given by

$$\begin{aligned} E\int_{-\infty}^{\infty}(\hat{m}(x) - m(x))^2w(x)dx &= \frac{1}{4}[\int_{-\infty}^{\infty} u^2K(u)du]^2\int_{-\infty}^{\infty}(m''(x))^2 w(x)dxh_n^4 \\ &\quad + \frac{1}{nh_n}\int_{-\infty}^{\infty}\frac{\sigma^2(x)}{f_X(x)}w(x)dx\int_{-\infty}^{\infty} K^2du + o(h_n^4 + \frac{1}{nh_n}). \end{aligned} \quad (3.2)$$

Simple algebra yields the optimal bandwidth for MISE (3.2) is

$$h_{\text{opt}} = \left(\frac{\int_{-\infty}^{\infty} f^{-1}(x)\sigma^2(x)w(x)dx \int_{-\infty}^{\infty} K^2 du}{[\int_{-\infty}^{\infty} u^2 K(u)du]^2 \int_{-\infty}^{\infty} (m''(x))^2\, w(x)dx} \right)^{1/5} n^{-1/5}.$$

Remark 1. The equation (3.1) holds uniformly in the class of joint densities

$$\mathcal{C}_2 = \{f(\cdot,\cdot) : |m''(x)| \le C\} \cap \{f(\cdot,\cdot) : \sigma^2(x) \le B, f_X(x_0) \ge b, |f_X(x) - f_X(y)| \le c|x-y|^\alpha\}, \tag{3.3}$$

where C, B, b, c, and α are positive constants. Note that Lipschitz's condition is imposed *only for the equicontinuity* of $f_X(\cdot)$ in the class of joint densities. The constants α, c will not involved in our further developments. The uniform convergence will be used in the asymptotic minimax theory. A similar remark applies to MISE given by (3.2) with an appropriate change of conditions.

Remark 2. The bias of estimator (2.2) does not depend on the derivatives of the marginal density f_X, while a corresponding Watson-Nadaraya estimator does. Since the bias of a Watson-Nadaraya estimator depends on the derivatives of f_X, its maximum risk over $\mathcal{C}_2$, say, blows up as the derivatives are unbounded in this class, hence its asymptotic minimax efficiency is arbitrarily small. However, if f_X is known to be *uniform*, then Watson-Nadaraya estimator (1.2) is equivalent to estimator (2.2). Note that the variance of estimator (2.2) is only about two thirds of a corresponding Gasser-Müller estimator (1.3), while the bias is the same. Thus, Gasser-Müller estimators are not admissible!

4 Asymptotic minimax theory

It is well known that estimators (2.2) are optimal in terms of rates of convergence (see Stone (1980)). More precisely, it is not possible to improve the rate $n^{-4/5}$ *uniformly* in $\mathcal{C}_2$ [defined by (3.3)]:

$$R(n, \mathcal{C}_2) \equiv \inf_{\hat{T}_n} \sup_{f \in \mathcal{C}_2} E_f(\hat{T}_n - m(x_0))^2 \asymp n^{-4/5}, \tag{4.1}$$

where "$\asymp$" means that both sides have the same order. In other words, one knows *only the rate* of the asymptotic minimax risk. Naturally, one would ask how far the constant factor of our estimate away from optimal. In this section, we are going to show that by a suitable choice of a bandwidth and a kernel function, the estimator (2.2) is *nearly optimal* (in asymptotic minimax sense) in *constant factors* as well. Such a type of results seems to be new in nonparametric regression context. Indeed, without introducing the new class of estimators, it is not possible to give a precise evaluation of the minimax risk $R(n, \mathcal{C}_2)$ (see Remark 2).

4.1 An upper bound of minimax risk

Note that by Theorem 1 and Remark 1, we have an obvious upper bound:

$$R(n,\mathcal{C}_2) \le \frac{1}{4}C^2\left(\int_{-\infty}^{\infty} u^2 K(u)du\right)^2 h_n^4 + \frac{B}{nh_nb}\int_{-\infty}^{\infty} K^2 du + o(h_n^4 + \frac{1}{nh_n}). \qquad (4.2)$$

Minimizing the right hand side of (4.2) yields an optimal choice of the bandwidth and kernel function:

$$h_n^{(1)} = (\frac{15B}{bC^2n})^{1/5}, \quad K_0(x) = \frac{3}{4}\left[1-x^2\right]_+. \qquad (4.3)$$

(Note that K_0 is a version of Epanechinkov's kernel). Substituting them into (4.2) yields a minimax upper bound given by

$$R(n,\mathcal{C}_2) \le \frac{3}{4}15^{-1/5}C^{2/5}\left(\frac{B}{bn}\right)^{4/5}(1+o(1)). \qquad (4.4)$$

Let $\hat{m}^*(x_0)$ be the estimate (2.2) with bandwidth and kernel given by (4.3).

Theorem 3. *An upper bound of the asymptotic minimax risk is given by (4.4). Moreover, the estimator $\hat{m}^*(x_0)$ has asymptotic minimax efficiency at least 89.6%:*

$$\frac{R(n,\mathcal{C}_2)}{\sup_{f\in\mathcal{C}_2} E_f\left(\hat{m}^*(x_0) - m_f(x)\right)^2} > 0.896^2 + o(1).$$

The last statement will be verified in following sections, where a more general theory for lower bound is devoted. Combining the two statements in Theorem 4 yields the minimax risk:

$$0.896^2 + o(1) \le \frac{R(n,\mathcal{C}_2)}{\frac{3}{4}15^{-1/5}C^{2/5}\left(\frac{B}{bn}\right)^{4/5}} \le 1 + o(1).$$

4.2 Modulus of continuity

The connections of modulus continuity with both upper and lower bounds for nonparametric density models and Gaussian white models have been extensively studied in the literature. See Donoho (1990), Donoho and Liu (1988), Donoho and Neusbaum (1990), among others. However, in nonparametric regression context, the connections appear to be new.

Let assume more generally that we want to estimate $m_f(x_0) = E_f(Y|X = x_0)$ with a nonparametric constraint $f \in \mathcal{F}$. For convenience of discussion, assume that $\mathcal{F} = \mathcal{F}_m \cap \mathcal{F}_{b,B}$ [compare (3.3)], where $\mathcal{F}_m$ are constraints on m, and $\mathcal{F}_{b,B}$ are constraints on marginal densities and conditional variance:

$$\mathcal{F}_{b,B} = \left\{ f(\cdot,\cdot) : f_X(x_0) \geq b, \sigma^2(x) \leq B, |f_X(x) - f_X(y)| \leq c|x-y|^\alpha \right\}, \qquad (4.5)$$

and $\sigma^2(x)$ is the conditional variance of Y given $X = x$. Note that the condition $|f_X(x) - f_X(y)| \leq c|x-y|^\alpha$ is only used to guarantee the equicontinuity of marginal densities f_X, and hence the constants c, and α are not related with both upper and lower bounds. Indeed, in the lower bound development below, this condition will not be used.

Define the modulus of continuity at a point x_0 over $\mathcal{F}_m$ by

$$\omega_{\mathcal{F}_m}(\varepsilon) = \sup\{|m_1(x_0) - m_2(x_0)| : m_j \in \mathcal{F}_m, \|m_1 - m_2\| \leq \varepsilon\}, \qquad (4.6)$$

where $\|\cdot\|$ is the usual L_2-norm on $L_2(-\infty, \infty)$. In nonparametric applications, one typically has

$$\omega_{\mathcal{F}_m}(\varepsilon) = A\varepsilon^p(1 + o(1)), \text{ as } \varepsilon \to 0, p \in (0,1) \qquad (4.7)$$

and the extremal pair is attained at $m_1(\cdot)$, and $m_2(\cdot)$, which has form

$$m_1(x) - m_2(x) = \varepsilon^p H\left(\frac{x_0 - x}{\varepsilon^{2q}}\right)(1 + o(1)), (\text{uniformly in } x \text{ as } \varepsilon \to 0), \qquad (4.8)$$

where $q = 1 - p$, and $H(\cdot)$ is a bounded continuous function.

Definition: A functional $m_f(x_0)$ is regular on $\mathcal{F}_m$ with exponent p, if the extremal pair of the modulus of continuity (4.6) exists, and has form (4.8).

As an illustration, let's consider the constraint $\mathcal{C}_2$. A similar computation can be found in Donoho and Liu (1988). In this case, $\mathcal{C}_2 = \mathcal{D}_2 \cap \mathcal{F}_{b,B}$, where

$$\mathcal{D}_2 = \{m(\cdot) : |m''(\cdot)| \le C\}. \tag{4.9}$$

Let's determine the modulus function for the class $\mathcal{D}_2$:

$$\omega_{\mathcal{D}_2}(\varepsilon) = \sup\{|m_1(x_0) - m_2(x_0)| : \|m_1 - m_2\| \le \varepsilon, m_1 \in \mathcal{D}_2, m_2 \in \mathcal{D}_2\}.$$

First, by Lemma 7 of Donoho and Liu (1988), the extremal pair can be chosen of form: $m_1 = m_0^*$, and $m_2 = -m_0^*$. Thus,

$$\omega_{\mathcal{D}_2}(\varepsilon) = 2\sup\{|m(x_0)| : \|m(\cdot)\| \le \varepsilon/2, m \in \mathcal{D}_2\}.$$

It follows that $\omega_{\mathcal{D}_2}$ is the inverse function of

$$\varepsilon(\omega) = 2\inf\{\|m(\cdot)\| : |m(x_0)| = \omega/2, m \in \mathcal{D}_2\}.$$

A solution to the last problem is obviously the function $m_0(\cdot)$ which is equal to $\omega/2$ at x_0 and which descends to 0 as rapidly as possible, and which stays constantly at 0 once it reaches 0, subject to the constraint that $m_0 \in \mathcal{D}_2$. In other words,

$$m_0(x) = \left[\omega - C(x - x_0)^2\right]_+ /2(1 + o(1)),$$

where $o(1)$ is used to modify the function $[\omega - C(x - x_0)^2]_+ /2$ near x_0 so that $m_0(x)$ is in $\mathcal{D}_2$. Now,

$$\int_{-\infty}^{\infty} m_0^2(x) = \frac{4}{15} C^{-1/2} \omega^{5/2} (1 + o(1)),$$

which implies

$$\varepsilon(\omega) = 2\sqrt{\frac{4}{15} C^{-1/2} \omega^{5/2}} (1 + o(1)).$$

Hence,

$$\omega_{\mathcal{D}_2}(\varepsilon) = (15/16)^{2/5} C^{1/5} \varepsilon^{4/5} (1 + o(1)). \tag{4.10}$$

The extremal pair is attained at $m_1 = -m_0$, and $m_2 = m_0$, where

$$\begin{aligned} m_0(x) &= 2^{-1}\left[(15/16)^{2/5}C^{1/5}\varepsilon^{4/5} - C(x-x_0)^2\right]_+ (1+o(1)) \\ &= 2^{-1}C^{1/5}\varepsilon^{4/5}\left[(15/16)^{2/5} - C^{4/5}\left(\frac{x-x_0}{\varepsilon^{2/5}}\right)^2\right]_+ (1+o(1)). \end{aligned}$$

Thus, the condition (4.8) is satisfied with $p = 4/5$ and

$$H(x) = C^{1/5}\left[(15/16)^{2/5} - C^{4/5}x^2\right]_+. \tag{4.11}$$

In terms of modulus of continuity, the upper bound (4.4) can be expressed as

$$\begin{aligned} R(n, C_2) &\le \frac{q}{4}\omega^2_{\mathcal{D}_2}\left(2\sqrt{\frac{pB}{nbq}}\right)(1+o(1)) \\ &= \frac{p^p q^q}{4}\omega^2_{\mathcal{D}_2}\left(2\sqrt{\frac{B}{nb}}\right)(1+o(1)). \end{aligned} \tag{4.12}$$

where $p = 4/5$, the exponent of the modulus of continuity, and $q = 1-p$.

4.3 Heuristics of the hardest 1-dimensional subproblem

Let turn attention away from the specific constraint C_2 towards a general constraint $\mathcal{F} = \mathcal{F}_m \cap \mathcal{F}_{b,B}$. Assume that $m(x_0)$ (suppress the dependence on f) is regular on $\mathcal{F}_m$ with exponent p. Consider the nonparametric minimax risk:

$$R(n, \mathcal{F}) = \inf_{\hat{T}_n \text{ measurable}} \sup_{f\in\mathcal{F}} E_f\left(\hat{T}_n - m(x_0)\right)^2.$$

Assume that $\mathcal{F}_m$ is convex, so that

$$m_\theta(x) = (1-\theta)m_0(x) + \theta m_1(x) \in \mathcal{F}_m, \quad \theta \in [0,1], \tag{4.13}$$

where m_0, and m_1 is an extremal pair of the modulus of continuity $\omega_{\mathcal{F}_m}\left(2\sqrt{pB/nbq}\right)$ [compare (4.12)]. Thus, there exists a family of joint densities $\mathcal{F}_\theta \equiv \{f_\theta : \theta \in [0,1]\}$ such that

$$E_{f_\theta}[Y|X = x] = m_\theta(x).$$

An obvious lower bound of $R(n,\mathcal{F})$ is

$$\begin{aligned} R(n,\mathcal{F}) &\geq R(n,\mathcal{F}_\theta) \\ &= |m_0(x_0)-m_1(x_0)|^2 \inf_{\hat{T}_n \text{ measurable}} \sup_{0\leq\theta\leq 1} E(\hat{T}_n-\theta)^2. \\ &= \omega^2_{\mathcal{F}_m}\left(2\sqrt{\frac{pB}{nbq}}\right) \inf_{\hat{T}_n} \sup_{|\theta|\leq 1/2} E(\hat{T}_n-\theta)^2(1+o(1)), \end{aligned} \tag{4.14}$$

as m_0, and m_1 are the extremal pair of the modulus. Thus, we reduce the full nonparametric problem to a 1-dimensional subproblem, and the connection with modulus is made.

Relevant information to the second factor of (4.14) is estimating a bounded normal mean from normal model. See Bickel (1981), Ibragimov and Khas'minskii (1984), Donoho *et al* (1989), among others. Consider observing the real-valued random variable $Y \sim N(\theta,\sigma^2)$; the objective is to estimate θ, and it is known that θ is bounded: $|\theta| \leq \tau$. The minimax risk for this problem is denoted by

$$\rho_N(\tau,\sigma) = \inf_{\hat{T} \text{ measurable}} \sup_{|\theta|\leq\tau} E(\hat{T}_n(Y)-\theta)^2, \tag{4.15}$$

which has a simple relation

$$\rho_N(\tau,\sigma) = \sigma^2\rho(\tau/\sigma,1).$$

Similarly, minimax affine risk is

$$\rho_A(\tau,\sigma) \equiv \inf_{a,b} \sup_{|\theta|\leq\tau} (a+bY-\theta)^2 = \frac{\tau^2\sigma^2}{\tau^2+\sigma^2}.$$

There is no close form for ρ_N. However, a simple inequality is available:

$$0.8 \leq \eta_\varepsilon \stackrel{\text{def}}{=} \rho_N(1/2,\varepsilon)/\rho_A(1/2,\varepsilon) \leq 1. \tag{4.16}$$

(By Donoho *et al.* (1989))

We would expect that the second factor of (4.14) is (see section 4.4 for detail)

$$\inf_{\hat{T}_n} \sup_{|\theta|\leq 1/2} E(\hat{T}_n(X_1,Y_1,\cdots,X_n,Y_n)-\theta)^2 \approx \rho_N\left(1/2,\sqrt{\frac{q}{4p}}\right).$$

If we show that

$$\liminf_{n\to\infty} \inf_{\hat{T}_n} \sup_{|\theta|\leq 1/2} E(\hat{T}_n(X_1,Y_1,\cdots,X_n,Y_n)-\theta)^2 \geq \rho_N\left(1/2,\sqrt{\frac{q}{4p}}\right), \tag{4.17}$$

then by (4.14) a lower bound would be

$$\begin{aligned} R(n,\mathcal{F}) &\geq \rho_N\left(1/2, \sqrt{\frac{q}{4p}}\right)\omega^2_{\mathcal{F}_m}\left(2\sqrt{\frac{pB}{qbn}}\right)(1+o(1)) \\ &= \xi_p \frac{p^p q^q}{4}\omega^2_{\mathcal{F}_m}\left(2\sqrt{\frac{B}{nb}}\right)(1+o(1)). \end{aligned}$$

where (4.16) is used in the last expression, and $\xi_p = \eta_{\sqrt{q/4p}}$.

Comparing the last display with (4.12), we give a nearly precise evaluation of asymptotic minimax risk for the class of constraints $\mathcal{C}_2$. In this case, $p = 4/5$, and a better evaluation is available: $\xi_{4/5} \geq 1/1.243$ (see Table 1 of Donoho and Liu (1988)). Therefore, the second conclusion of Theorem 3 is proved, if we verify (4.17).

4.4 Modulus continuity and minimax lower bound

To validate (4.17), we consider a normal submodel:

$$f_\theta(x,y) = \frac{1}{\sqrt{2\pi B}}\exp(-(y - m_\theta(x))^2/B)g(x), \tag{4.18}$$

where $g(x)$ is a marginal density, and m_θ is defined by (4.13). We make an assumption on the richness of $\mathcal{F}$.

Richness of joint densities:

There exists a bounded density g with $g(x_0) = b$ such that the normal submodel (4.18) is in the class of constraint $\mathcal{F} = \mathcal{F}_m \cap \mathcal{F}_{b,B}$.

Based on the normal submodel (4.18), a sufficient statistics for θ would be

$$\hat{\delta}_n \equiv \sum_1^n (Y_j - m_0(X_j))(m_1(X_j) - m_0(X_j)), \quad \hat{\sigma}_n^2 \equiv \sum_1^n (m_0(X_j) - m_1(X_j))^2. \tag{4.19}$$

Thus, considering statistics based on $\hat{\delta}_n$ and $\hat{\sigma}_n^2$ would be good enough for estimating the unknown parameter θ. Note that conditioning on $X_1, \cdots, X_n$,

$$\hat{\delta}_n/\hat{\sigma}_n^2 \sim N\left(\theta, B\hat{\sigma}_n^{-2}\right).$$

Thus,

$$\inf_{\hat{T}_n} \sup_{|\theta|\leq 1/2} E\left((\hat{T}_n - \theta)^2 | X_1, \cdots, X_n\right) = \rho_N\left(1/2, \sqrt{B}/\hat{\sigma}_n\right). \tag{4.20}$$

Since m_0, m_1 is the extremal pair of $\omega_{\mathcal{F}_m}(\varepsilon_n)$ with $\varepsilon_n = 2\sqrt{pB/nbq}$, and by assumption (4.8), we have

$$m_0(x) - m_1(x) = \varepsilon_n^p H\left(\frac{x_0 - x}{\varepsilon_n^{2q}}\right)(1 + o(1)), \quad \int_{-\infty}^{\infty}(m_0 - m_1)^2 dx = \varepsilon_n^2, \qquad (4.21)$$

which implies $\int_{-\infty}^{\infty} H^2(x)dx = 1$. Note that by (4.21),

$$\begin{aligned} E\hat{\sigma}_n^2 &= n\int_{-\infty}^{\infty}(m_0(x) - m_1(x))^2 g(x)dx \\ &= ng(x_0)\varepsilon_n^2 \int_{-\infty}^{\infty} H^2 dx(1 + o(1)) \\ &= \frac{4pB}{q} + o(1) \end{aligned}$$

and

$$\begin{aligned} \mathrm{Var}(\hat{\sigma}_n^2) &\leq n\int_{-\infty}^{\infty}(m_0(x) - m_1(x))^4 g(x)dx \\ &\leq \varepsilon_n^{2p} \sup_x H^2(x)\left[n\int_{-\infty}^{\infty}(m_0(x) - m_1(x))^2 g(x)dx\right] \\ &\longrightarrow 0, \quad \text{as } n \to \infty. \end{aligned}$$

Consequently,

$$\hat{\sigma}_n \xrightarrow{\text{P}} \sqrt{\frac{4pB}{q}}, \qquad (4.22)$$

Heuristically, by (4.20)

$$\inf_{\hat{T}_n} \sup_{|\theta| \leq 1/2} E(\hat{T}_n - \theta)^2 \approx E\rho_N\left(1/2, \sqrt{B}/\hat{\sigma}_n\right) \approx \rho_N\left(1/2, \sqrt{\frac{q}{4p}}\right),$$

By (4.20), we validate heuristically (4.17).

Theorem 4. *Let $\mathcal{F}_m$ be convex, and $\mathcal{F} = \mathcal{F}_m \bigcap \mathcal{F}_{b,B}$ be rich. If $m_f(x_0)$ is regular on $\mathcal{F}_m$ with exponent p, then a minimax lower bound is given by*

$$R(n, \mathcal{F}) \geq \xi_p \frac{p^p q^q}{4} \omega_{\mathcal{F}_m}^2 \left(2\sqrt{\frac{B}{nb}}\right)(1 + o(1))$$

where $m(\cdot) = E(Y|X = \cdot)$, and $\xi_p = \eta_{\sqrt{q/4p}}$ is defined by (4.16).

Remark 3. The result of theorem 4 holds also for a randomly designed problem whose marginal density is known to be $g(\cdot)$ (e.g uniform [0,1]). The reason is that in our lower bound development, the marginal is fixed all the time. Our lower bound is also applicable to a fixed design problem, where the design points are $x_i = G(\frac{i}{n})$ with $G' = g$ (specifically uniform design), since our arguments above are conditioned on covariates $X_1, \cdots, X_n$.

Remark 4. Suppose that we want to estimate a conditional quantile function $Q_p(x_0)$ defined by (see Truong (1989))

$$P\{Y \le Q_p(x_0) | X = x_0\} = p,$$

based on i.i.d. samples $(X_1, Y_1), \cdots, (X_n, Y_n)$. Then for the normal submodel (4.18),

$$Q_p(x_0) = m_\theta(x_0) + z_p \sqrt{B},$$

where $z_p = \Phi^{-1}(p)$, and $\Phi(\cdot)$ is the standard normal cdf. Thus, estimating $Q_p(x_0)$ in the normal submodel has the same difficulty as estimating $m_\theta(x_0)$. This yields a lower bound

$$\inf_{\hat{T}_n} \sup_{f \in \mathcal{F}} E_f \left(\hat{T}_n - Q_p(x_0)\right)^2 \ge \xi_p \frac{p^p q^q}{4} \omega^2_{\mathcal{F}_m} \left(2\sqrt{\frac{B}{nb}}\right) (1 + o(1)).$$

Specially, when $p = 0.5$, this is a lower bound for estimating a conditional median.

5 Discussion

The minimax lower bound is derived via the heuristic the hardest one dimensional subproblem. We have shown that such a bound is indeed *nearly sharp* for a two-bounded-derivative constraints. Analysis of minimax upper bounds for other constraints goes beyond the intent of this paper, but provides interesting topics for future research.

5.1 Nearly sharp lower bound

We have shown that a minimax risk lower bound is [$\xi_p \ge 0.8$, by (4.16)]

$$R(n, \mathcal{F}) \ge 0.894^2 \frac{p^p q^q}{4} \omega^2_{\mathcal{F}_m} \left(2\sqrt{\frac{B}{nb}}\right) (1 + o(1)). \tag{5.1}$$

Thus, if one can find an estimator such that its maximum risk is no larger than

$$\frac{p^p q^q}{4} \omega^2_{\mathcal{F}_m} \left(2\sqrt{\frac{B}{nb}} \right) (1 + o(1)), \tag{5.2}$$

then such an estimator has at least a minimax efficiency 89.4% in a similar sense of Theorem 3, and consequently the lower bound is *nearly sharp.* With such a sharp minimax lower bound, we can compute the efficiency of *any estimator* in the following way:

$$\text{Efficiency of an esimator} \geq \left(\frac{\text{Minimax lower bound (5.1)}}{\text{Maximum MSE of the estimator}} \right)^{1/2}. \tag{5.3}$$

Two-bounded-derivative constraints $\mathcal{C}_2$ are not the only examples that upper bound (5.2) holds. We conjecture that a general theory can be made if one makes connections with white noise models as in Donoho and Liu (1988) where minimax theory is devoted for density estimation. Let's give another example, where the minimax upper bound (5.2) holds.

Example. (Bounded Lipschitz constraints). Let $(X_1, Y_1) \cdots, (X_n, Y_n)$ be i.i.d from a joint density $f \in \mathcal{C}_1 = \mathcal{D}_1 \cap \mathcal{F}_{b,B}$ with

$$\mathcal{D}_1 = \{m(\cdot) : |m(x) - m(y)| \leq C|x - y|, \forall x, y \in \Re\},$$

A similarly machinery to (4.9)—(4.10) yields the modulus of continuity:

$$\omega_{\mathcal{D}_1} = 3^{1/3} C^{1/3} \varepsilon^{2/3}$$

and $m(x_0)$ is regular on $\mathcal{D}_1$. Thus, Theorem 4 holds:

$$R(n, \mathcal{C}_1) \geq \xi_{2/3} 3^{-1/3} \left(\frac{BC}{bn} \right)^{2/3},$$

where $\xi_{2/3} = 1/1.178 = 0.92^2$ by Donoho and Liu (1988). On the other hand, exhibiting the maximum risk of the estimator

$$\sum_1^n \left(1 - \left|\frac{x_0 - X_j}{h_n^{(2)}}\right|\right)_+ Y_j \Big/ \sum_1^n \left(1 - \left|\frac{x_0 - X_j}{h_n^{(2)}}\right|\right)_+ \tag{5.4}$$

with $h_n^{(2)} = (\frac{3B}{bC^2n})^{1/3}$ [corresponding to a local polynomial estimate (2.6) with $K(x) = (1-|x|)_+$] yields an upper bound:

$$R(n, C_1) \leq 3^{-1/3}\left(\frac{BC}{bn}\right)^{2/3}. \tag{5.5}$$

Thus, (5.2) holds. In summary,

Theorem 5. *Under the constraint C_1, the minimax risk is bounded by*

$$0.92^2 \times 3^{-1/3}\left(\frac{BC}{bn}\right)^{2/3} \leq R(n, C_1) \leq 3^{-1/3}\left(\frac{BC}{bn}\right)^{2/3}.$$

Moreover, the estimator (5.4) has asymptotic minimax efficiency at least 92%.

5.2 Relation to other work

Vast literature has been devoted in analyzing the behavior of Watson-Nadaraya and Gasser-Müller regression estimators. The drawbacks of these estimators are eliminated via introduction a new class of estimators.

Previous work on minimax regression problems has mainly focused (Stone (1980)) on determining optimal *rates* of convergence. The local polynomial regression estimators are used by Stone (1980) to determine the rates of convergence. To analyze the constant factors, we extend the idea of local polynomial regression estimators.

A closely related idea for minimax bounds is the work of Donoho and Liu (1988), Donoho (1990), where the density estimation and white noise models are focused. What seems innovative in our approach is the decomposition of nonparametric constraints into two parts $\mathcal{F}_m$ and $\mathcal{F}_{b,B}$, and the use of normal submodels to avoid the technical difficulty of Le Cam's theory of convergence of experiments (compare with Donoho and Liu (1988)).

Other efforts in finding minimax risks in regression setup include Nussbaum (1985), Low (1989), and Fan (1989), where the attentions are mostly focused on some special global problems. In density estimation setup, these include Efroimovich and Pinsker (1982), Sacks and Ylvisaker (1981), Sacks and Strawderman (1982), Birgé (1987), Donoho and Liu (1988). There is also a long history in finding minimax risks for Gaussian white noise models and other related problems. See Pinsker (1980), Ibragimov and Khasminskii (1984), Brown and Liu (1989), Donoho *et al* (1989), Donoho and Jonestone (1989), Donoho and Neusbaum (1990), among others.

6 Proof

6.1 Proof of Theorem 1 & 2

By conditional argument on covariates $X_j, j = 1, \cdots, n$, and by using mean and variance decomposition, we have

$$E(\hat{m}(x_0) - m(x_0))^2 = E\left(\frac{\sum_1^n [m(X_j) - m(x_0)]w_j}{\sum_1^n w_j}\right)^2 + E\frac{\sum_1^n \sigma^2(X_j)w_j^2}{(\sum_1^n w_j)^2}. \tag{6.1}$$

Denote $Z_n = O_r(a_n)$, if $E|Z_n|^r = O(a_n^r)$. A similar meaning extends to $o_r(a_n)$. Obvious operations include

$$O_r(a_n)O_r(b_n) = O_{r/2}(a_n b_n), \qquad \text{(Cauchy-Schwartz's inequality)}$$

and

$$Z_n = EZ_n + O_r\left([E|Z_n - EZ_n|^r]^{1/r}\right). \tag{6.2}$$

Then, it is easy to show that with $s_{n,l}$ defined by (2.4),

$$Es_{n,l} = nh^{l+1} f_X(x_0) s_l (1 + o(1)), \quad l = 0, 1, 2.$$

and [c.f. (6.2)]

$$\begin{aligned} \frac{1}{nh_n^{l+1}} s_{n,l} &= \frac{1}{nh_n^{l+1}} Es_{n,l} + O_8\left(\frac{1}{\sqrt{nh_n}}\right), \\ &= f_X(x_0)s_l + o_8(1) \quad l = 0, 1, 2. \end{aligned} \tag{6.3}$$

where $s_l = \int_{-\infty}^{\infty} u^l K(u)du$, and in particular, $s_0 = 1$, $s_1 = 0$. A directly consequence of (6.3) is that

$$\sum_1^n w_n = s_{n,0}s_{n,2} - (s_{n,1})^2 = n^2 h_n^4 s_2 f_X^2(x_0)\left(1 + o_4(1)\right). \tag{6.4}$$

Let $R(X_j) = m(X_j) - m(x_0) + m'(x_0)(x_0 - X_j)$. Then, by (2.5), we have the following expression for the "bias term" of (6.1):

$$\begin{aligned}\sum_1^n [m(X_j) - m(x_0)]w_j &= \sum_1^n R(X_j)w_j \\ &= \sum_1^n R(X_j)K\left(\frac{x_0 - X_j}{h_n}\right)s_{n,2} \\ &\quad - \sum_1^n R(X_j)(x_0 - X_j)K\left(\frac{x_0 - X_j}{h_n}\right)s_{n,1}. \end{aligned} \tag{6.5}$$

Note that by a standard argument [see (6.2)],

$$\begin{aligned}&\frac{1}{nh_n^3}\sum_1^n R(X_j)K\left(\frac{x_0 - X_j}{h_n}\right) \\ =\; & h_n^{-3}E[m(X) - m(x_0) + m'(x_0)(x_0 - X)]K\left(\frac{x_0 - X}{h_n}\right) + O_8(\frac{1}{\sqrt{nh_n}}) \\ =\; & \frac{1}{2}f_X(x_0)m''(x_0)s_2 + o_8(1),\end{aligned}$$

and similarly

$$\begin{aligned}&\frac{1}{nh_n^4}\sum_1^n R(X_j)(x_0 - X_j)K\left(\frac{x_0 - X_j}{h_n}\right) \\ =\; & \frac{1}{2}f_X(x_0)m''(x_0)s_3(1 + o(1)) + O_8(\frac{1}{\sqrt{nh_n}}) \\ =\; & O_8(1).\end{aligned}$$

Applying the last two displays to (6.5) and using (6.3), we have

$$\sum_1^n [m(X_j) - m(x_0)]w_j = \frac{n^2h_n^6}{2}f_X^2(x_0)s_2^2 m''(x_0)(1 + o_4(1)).$$

Thus, we conclude from (6.4) that

$$E\left(\frac{\sum_1^n [m(X_j) - m(x_0)]w_j}{\sum_1^n w_j}\right)^2 = \left(\frac{1}{2}s_2 m''(x_0)\right)^2 h_n^4(1 + o(1)). \tag{6.6}$$

To complete the proof, we need only to justify the second term of (6.1) has the right order. Note that

$$\sum_1^n \sigma^2(X_j)w_j^2 = \sum_1^n \sigma^2(X_j)K^2\left(\frac{x_0 - X_j}{h_n}\right)\left[s_{n,2}^2 - 2(x_0 - X_j)s_{n,2}s_{n,1} + (x_0 - X_j)^2 s_{n,1}^2\right], \tag{6.7}$$

which has the same structure as (6.5). A standard argument [c.f. (6.2)] yields that

$$\begin{aligned}&\frac{1}{nh_n^{l+1}}\sum_1^n \sigma^2(X_j)K^2\left(\frac{x_0-X_j}{h_n}\right)(x_0-X_j)^l\\ =\ &\sigma^2(x_0)f_X(x_0)\int_{-\infty}^{\infty}u^lK^2(u)du+o_4(1),\quad l=0,1,2.\end{aligned} \tag{6.8}$$

Since $s_1=0$, the dominant term of (6.7) is its first term:

$$\sum_1^n \sigma^2(X_j)K^2\left(\frac{x_0-X_j}{h_n}\right)s_{n,2}^2.$$

Consequently, by (6.3), (6.7) and (6.8), we have

$$\begin{aligned}&\sum_1^n \sigma^2(X_j)w_j^2\\ =\ &nh_n\left(\sigma^2(x_0)f_X(x_0)\int_{-\infty}^{\infty}K^2(u)du+o_4(1)\right)n^2h_n^6f_X^2(x_0)s_2^2(1+o_4(1))\\ =\ &n^3h_n^7\sigma^2(x_0)f_X^3(x_0)s_2^2\int_{-\infty}^{\infty}K^2(u)du(1+o_2(1)).\end{aligned}$$

Thus, by (6.4), we conclude that

$$\begin{aligned}E\frac{\sum_1^n\sigma^2(X_j)w_j^2}{(\sum_1^n w_j)^2} &= \frac{n^3h_n^7\sigma^2(x_0)f_X^3(x_0)s_2^2\int K^2}{n^4h_n^8s_2^2f_X^4(x_0)}E(1+o_1(1))\\ &= \frac{\sigma^2(x_0)}{nh_nf_X(x_0)}\int_{-\infty}^{\infty}K^2(u)du(1+o(1))\end{aligned}$$

This completes the proof of Theorem 1. Theorem 2 follows from Theorem 1.

6.2 Proof of Theorem 4

We need only to prove (4.17). Let $\pi(\theta)$ be a least favorable prior for problem (4.15) with $\tau=0.5$ and $\sigma=\sqrt{q/4p}$:

$$\rho_N(0.5,\sqrt{q/4p})=\inf_{\hat{T}}E_\theta E(\hat{T}(Y)-\theta)^2,\quad Y\sim N(\theta,\sqrt{q/4p}). \tag{6.9}$$

Denote the Bayes risk with the prior π for normal model $X\sim N(\theta,\sigma^2)$ by

$$B_\pi(\sigma)=\inf_{\hat{T}_n}E_\theta E_X\left(\hat{T}_n(X)-\theta\right)^2. \tag{6.10}$$

Then, (6.9) can be expressed as

$$B_\pi(\sqrt{q/4p}) = \rho_N(0.5, \sqrt{q/4p}). \tag{6.11}$$

Now, let's turn our attentions back to the problem (4.17) with n i.i.d. observations $\{(X_i, Y_i)\}$ from (4.18). By sufficiency,

$$\begin{aligned} \inf_{\hat{T}_n} \sup_{|\theta|\le 0.5} E\left(\hat{T}_n(X_1, Y_1, \cdots, X_n, Y_n) - \theta\right)^2 &= \inf_{\hat{T}_n^*} \sup_{|\theta|\le 0.5} E\left(\hat{T}_n^*(\hat{\delta}_n, \hat{\sigma}_n) - \theta\right)^2 \\ &\ge \inf_{\hat{T}_n^*} E_\theta E\left(\hat{T}_n^*(\hat{\delta}_n, \hat{\sigma}_n) - \theta\right)^2 \\ &\ge E_{\hat{\sigma}_n} \inf_{\hat{T}_n} E_\theta E_{\hat{\delta}_n}\left[\left(\hat{T}_n(\hat{\delta}_n, \hat{\sigma}_n) - \theta\right)^2 |\hat{\sigma}_n\right], \end{aligned} \tag{6.12}$$

where $\hat{\sigma}_n$ and $\hat{\delta}_n$ are defined by (4.19). Given $\hat{\sigma}_n$, $\hat{\delta}_n/\hat{\sigma}_n \sim N(\theta, B\hat{\sigma}_n^{-2})$. Thus, by (6.12),

$$\inf_{\hat{T}_n} \sup_{|\theta|\le 0.5} E(\hat{T}_n - \theta)^2 \ge E_{\hat{\sigma}_n} B_\pi(\sqrt{B}/\hat{\sigma}_n). \tag{6.13}$$

Note that $B_\pi(\cdot)$ is bounded by 1/4 (as $|\theta| \le 0.5$ in our discussion), and is continuous. By the dominated convergence theorem and (4.22), (6.11)

$$E_{\hat{\sigma}_n} B_\pi(\sqrt{B}/\hat{\sigma}_n) \longrightarrow B_\pi(\sqrt{q/4p}) = \rho_N(0.5, \sqrt{q/4p}).$$

Thus, by (6.13)

$$\liminf_{n\to\infty} \inf_{\hat{T}_n} \sup_{|\theta|\le 0.5} E(\hat{T}_n - \theta)^2 \ge \rho_N(0.5, \sqrt{q/4p}).$$

This completes the proof.

Acknowledgements

The author gratefully acknowledges Professors D. L. Donoho, S. J. Marron, Y. K. Truong for many useful discussions.

References

[1] Bickel, P.J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Statist.*, **9**, 1301-1309.

[2] Birgé, L. (1987). Estimating a density under the order restrictions: Nonasymptotic minimax risk. *Ann. Statist.*, **15**, 995-1012.

[3] Brown, L.D. and Liu, R.C. (1989). A sharpened inequality for the hardest affine subproblem. *Manuscripts.*

[4] Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Jour. Ameri. Statist. Assoc.*, **74**, 829-836.

[5] Chu, C. K. and Marron, J.S. (1990). Choosing a kernel regression estimator. *Manuscript.*

[6] Donoho, D. L. and Liu, R. C. (1988). Geometrizing rate of convergence III. *Tech. Report 138*, Dept. of Statist., University of California, Berkeley.

[7] Donoho, D. L., MacGibbon, B., and Liu, R.C. (1989). Minimax risk for hyperrectangles. *Tech. Report 123*, Dept. of Statist., University of California, Berkeley.

[8] Donoho, D. L. and Johnstone, I. (1989). Minimax risk over l_p-balls. *Tech. Report 204*, Dept. of Statist., University of California, Berkeley.

[9] Donoho, D. L. (1990). Statistical estimation and optimal recovery. *Tech. Report 214*, Dept. of Statist., University of California, Berkeley.

[10] Donoho, D. L. and Neusbaum, M. (1990). Minimax quadratic estimation of a quadratic functional. *Tech. Report 236*, Dept. of Statist., University of California, Berkeley.

[11] Efroimovich, S.Y. and Pinsker, M.S. (1982). Estimation of square-integrable probability density of a random variable. *Problems of Information Transmission*, 175-189.

[12] Fan, J. (1989). Adaptively local 1-dimensional subproblems. *Institute of Statistics Mimeo Series #2010*, Univ. of North Carolina, Chapel Hill.

[13] Gasser, T. and Müller, H.G. (1979). Kernel estimation of regression functions. In Smoothing techniques for curve estimation. Lectures Notes in Math. 757, 23–68, Springer–Verlag, New York.

[14] Ibragimov, I.A. and Khasminskii, R.Z. (1984). On nonparametric estimation of values of a linear functional in a Gaussian white noise (in Russian). *Teoria Verojatnostei i Primenenia*, **29**, 19-32.

[15] Le Cam, L. (1985). *Asymptotic methods in Statistical Decision Theory.* Springer-Verlag, New York-Berlin.

[16] Low, M. (1989). Lower bounds for the integrated risk in nonparametric density and regression estimation. *Tech. Report. 223*, Department of Statistics, University of California, Berkeley.

[17] Müller, H.G. (1987). Weighted local regression and kernel methods for nonparametric curve fitting. *Jour. Ameri. Statist. Assoc.*, **82**, 231-238.

[18] Nadaraya, E.A. (1964). On estimating regression. *Theory Probab Appli*, **9**, 141-142.

[19] Neusbaum, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L_2. *Ann. Statist.*, **13**, 984-997,

[20] Pinsker, M.S. (1980). Optimal filtering of square integrable signals in Gaussian white noise. *Problems Infor. Trans.*, **16**, 52-68.

[21] Sacks, J. and Strawderman, W. (1982). Improvements on linear minimax estimates. In *Statistical Decision Theory and Related Topics III*, **2**, (S. Gupta ed.), Academic, New York.

[22] Sacks, J. and Ylvisaker, D. (1981). Asymptotically optimum kernels for density estimation at a point. *Ann. Statist.*, **9**, 334-346.

[23] Stone, C.J. (1977). Consistent Nonparametric Regression. *Ann. Statist.*, **5**, 595-620.

[24] Stone, C.J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.*, **8**, 1348-1360.

[25] Watson, G.S. (1964). Smooth regression analysis. *Sankhyā*, Ser. A, **26**, 359-372.

[26] Truong, Y.K. (1989). Asymptotic properties of kernel estimators based on local medians. *Ann. Statist.*, **17**, 606-617.