Renormalizing Upper and Lower Bounds for Integrated Risk in the White Noise Model

By

Mark G. Low

Technical Report No. 275
October 1990
(revised April 1992)

Department of Statistics
University of California
Berkeley, California

Renormalizing Upper and Lower Bounds for Integrated Risk in the White Noise Model

By

Mark G. Low

1. Introduction:

Many functional estimation problems arising in density estimation and nonparametric regression are easier to analyse in the following white noise model

$$(1) \qquad dX_t = f(t)\,dt + \sigma\,dW_t \qquad 0 \le t \le 1, \qquad f \in \mathbf{F} \subseteq L_2[0,1]$$

where W_t is Brownian motion.

Many results which might be difficult to prove in the density estimation or nonparametric regression context take on a more transparent form in this white noise model. A sample size of n in the density estimation and nonparametric regression problems corresponds to $\sigma_n = \frac{\sigma}{\sqrt{n}}$ in (1) when σ is suitably chosen. In particular the tools of rescaling developed in Low [1989a] and Donoho and Low [1990] and the hardest one dimensional subfamily arguments of Donoho and Liu [1987, 1988] have yielded a fairly complete picture of how to estimate both bounded and unbounded linear functionals on the basis of observations generated by (1). A separate literature is developing to show how to replace density estimation and regression problems by the corresponding white noise problems. See for example Low [1989b], Brown and Low [1990] and Donoho and Low [1990].

In this paper we focus attention on estimating the entire function f on the basis of the observation scheme given by (1), using integrated squared error as a measure of loss. In particular we shall let $R(\mathbf{F},\sigma)$ denote the minimax risk under this loss function. That is

$$(2) \qquad R(\mathbf{F},\sigma) = \inf_{\delta} \sup_{f \in \mathbf{F}} E \int_0^1 (f(x) - \delta(x))^2\,dx$$

where the infimum is taken over all procedures δ.

For ellipsoidal parameter spaces such as $\mathbf{F} = \{f : \int_0^1 f'^2(x)\,dx \le 1,\ f(0) = f(1)\}$, a fairly complete analysis has already been given for the asymptotic minimax risk $R(\mathbf{F},\sigma)$ as $\sigma \downarrow 0$ by Pinsker [1980] and Efroimovich and Pinsker [1982].

In this paper we derive upper and lower bounds for the minimax risk $R(\mathbf{F},\sigma)$ for nonellipsoidal parameter spaces satisfying certain renormalizable properties. This work may therefore be viewed as an extension of the use of invariance ideas to global estimation problems, although in the present context the renormalizing structure is more involved. We use invariance in this paper to accomplish two goals. First we show how optimal rates of convergence for estimating an entire function can sometimes be derived just from the renormalizing structure of a parameter set $\mathbf{F}$. Second we use invariance to reduce the calculation of lower bounds for global estimation to a single hardest one dimensional subfamily argument similar to those analysed in detail by Donoho and Liu. In this way we can find lower bounds for the minimax risk involving constants and not just rates. Upper bounds for the minimax risk can be given in terms of the corresponding pointwise estimation problem. As an example we compare upper and lower bounds for a class of functions with a uniformly bounded derivative.

The results of this paper should also 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]. One contribution of this paper is to show how to connect local problems to global problems.

2. Rescaling Properties of F:

Throughout this paper we always assume that $\mathbf{F} \subseteq L_2[0,1]$. However we can also extend any $f \in \mathbf{F}$ to a function, which we shall also call f, with domain $(-\infty, \infty)$ by defining $f(x) = 0$ for $x \notin [0,1]$. Hence we shall allow function evaluations at points outside the closed interval $[0,1]$ and always take the value to be zero. In the assumptions and theorems which follow we write $[T]^-$ for the greatest integer less than or equal to T and $[T]^+$ for the smallest integer greater than or equal to T.

In a previous paper, Low [1989a] we showed how optimal rates of convergence for estimating a function at a point can be derived from invariance properties of the parameter set **F**. In particular we required the space **F** to be invariant under particular scale and dilation transformations. In other words we needed to assume, for appropriate choices of a and b, that the map $f(t) \to af(bt)$ is a bijection on **F**. For the problem of estimating the entire function the renormalizing structure we need is more involved.

Assumptions

1) **Lower bounds**

For lower bounds we assume that we have a collection of parameter spaces $\mathbf{F}_T$ such that for each $T \in [1, \infty)$

a) If $f \in \mathbf{F}_T$ and if $x \notin (0, \frac{1}{T})$ then $f(x) = 0$ where $(0, \frac{1}{T})$ denotes the open interval $\{x : 0 < x < \frac{1}{T}\}$

b) If $f \in \mathbf{F}_T$ and if $|\theta| \le 1$ then $\theta f \in \mathbf{F}_T$

c) If $f_i \in \mathbf{F}_T$ then $g(t) + \sum_{i=0}^{[T]^- -1} \theta_i f_i(t - \frac{i}{T}) \in \mathbf{F}$ where g is some fixed function not depending on the choice of f_i, $|\theta_i| \le 1$, $1 \le i \le n$

d) $\phi : [1,\infty) \to (0,\infty)$ is a function such that if $T \in [1,\infty)$ then the mapping $f(t) \to \frac{f(Tt)}{\phi(T)}$ is 1-1 and onto from $\mathbf{F}_1$ to $\mathbf{F}_T$

Remark:

Assumptions a), b) and c) taken together allow us to give a lower bound for estimating $f \in \mathbf{F}$ in terms of a lower bound for estimating a single $f \in \mathbf{F}_T$. Assumption d) captures the renormalizing structure, needed to replace the problem of estimating $f \in \mathbf{F}_T$ by the problem of estimating $f \in \mathbf{F}_1$ but with a different value of σ. Details are found in Lemma 1 and Theorem 1 given below.

2) **Upper bounds**

For upper bounds we assume that we have a collection of parameter spaces $\mathbf{F}^T$, $T \in [1,\infty)$ such that the support of any function $f \in \mathbf{F}'$ is contained in the interval $[0,1]$ and

a) $\mathbf{F} \subseteq \mathbf{F}^1$

b) $\psi: [1,\infty) \to (0,\infty)$ is a function such that if $T \in [1,\infty)$ then the mapping $f(t) \to \dfrac{f(Tt)}{\psi(T)}$ is 1-1 and onto from $\mathbf{F}^1$ to $\mathbf{F}^T$. It follows that if $f \in \mathbf{F}^T$ then $f(x) = 0$ for $x \notin [0, \frac{1}{T}]$

c) If $f \in \mathbf{F}^1$ then for $i = 0,1,\ldots,[T]^- - 1$ there is an $f^T \in \mathbf{F}^T$ such that $f^T(t) = f(\frac{i}{T} + t)$, $0 \le t \le \frac{1}{T}$ and if $f \in \mathbf{F}^1$ then there is an $f^T \in \mathbf{F}^T$ such that $f^T(t) = f(1 - \frac{1}{T} + t)$, $0 \le t \le \frac{1}{T}$

Remark: If the functions ϕ in 1d) and ψ in 2b) are the same, the upper and lower bounds derived in the next section are of the same order and yield optimal rates of convergence. Compare theorem 1 and theorem 2 in section 3.

Example 1: Write $f^j(x)$ for the j^{th} derivative of f. Let

(3) $\mathbf{F}(k,M) = \{f : |f^{k-1}(x) - f^{k-1}(y)| \le M|x-y|,\ f^j(0) = f^j(1),\ j = 0,\ldots,k-1\}$.

Take

(4) $$\mathbf{F}_1(k,M) = \mathbf{F}(k,M) \cap \{f \mid f^j(0) = f^j(1) = 0, \quad j = 0,\ldots,k-1\}$$

and take

(5) $$\mathbf{F}^1(k,M) = \{f : |f^{k-1}(x) - f^{k-1}(y)| \le M|x-y|\}.$$

Let ϕ: $[1,\infty] \to (0,\infty)$ and $\psi: [1,\infty] \to (0,\infty)$ be defined by $\psi(t) = \phi(t) = t^k$ and take $g \equiv 0$. Define $\mathbf{F}_T$ and $\mathbf{F}^T$ by

(6) $$\mathbf{F}_T = \left\{ \frac{f(Tt)}{\phi(T)} : f \in \mathbf{F}_1 \right\}$$

and

$$\mathbf{F}^T = \left\{ \frac{f(Tt)}{\psi(T)} : f \in \mathbf{F}^1 \right\}. \tag{7}$$

Then assumptions 1d) and 2c) are by construction satisfied. Once we note that

$$\frac{d^k}{dt^k} \frac{f(Tt)}{\phi(T)} = \frac{T^k}{T^k} f^k(Tt) = \frac{1}{T} f^{k-1}(Tt)$$

it is easy to check the remaining assumptions given in 1 and 2. We leave the details to the reader. We shall return to this example at the end of section 3.

Example 2: We now give an example where we do not take $g \equiv 0$ in 1c). Let

$$\mathbf{F}(M) = \{f : 0 \le f'(x) \le M\}. \tag{8}$$

Take

$$\mathbf{F}_1(M) = \{f : -\frac{M}{2} \le f'(x) \le \frac{M}{2},\ f(0) = f(1) = 0\} \tag{9}$$

and take

$$\mathbf{F}^1(M) = \mathbf{F}(M). \tag{10}$$

Then if we let $\phi(t) = \psi(t) = t$, take $g(t) = \frac{M}{2} t$ and define $\mathbf{F}_T$ and $\mathbf{F}^T$ by (6) and (7) it is easy to check that assumptions 1) and 2) are once again satisfied.

3. Upper and Lower Bounds:

Conditions 1a), 1b) and 1c) given in the previous section enable us to give lower bounds for the minimax risk $R(\mathbf{F}, \sigma)$ in terms of the minimax risk for a single bounded normal mean problem. The analysis combines invariance ideas with hardest one-dimensional subfamily arguments due to Donoho and Liu [1987, 1988]. Let us denote by $\rho(d, \sigma)$ the minimax risk for estimating θ on the basis of $X \sim N(\theta, \sigma^2)$ where $|\theta| \le d$. Then

$$\rho(d, \sigma) = \inf_{\delta} \sup_{|\theta| \le d} E(\theta - \delta(x))^2. \tag{11}$$

Explicit values of $\rho(d, \sigma)$ were first given by Casella and Strawderman (1981) for $\frac{d}{\sigma} \le 1.01$ where it was also shown that

$$\rho(d, \sigma) = \sigma^2 \rho(\frac{d}{\sigma}, 1). \tag{12}$$

Extensive tables for $\frac{d}{\sigma} \le 5$ can now be found in Brown and Feldman [1990], and Donoho, McGibbon and Liu [1988]. In the following lemmas and theorems when we refer to the white noise process we shall always be refering to the process given by equation (1). We also write $\|f\|_2$ for the L_2 norm of a function f, $\|f\|_2^2 = \int f^2(t)\,dt$.

Lower Bounds

Lemma 1: Suppose we observe the white noise process and that the parameter spaces $\mathbf{F}_T$ satisfy assumptions 1a), 1b) and 1c) then

$$(13)\qquad R(\mathbf{F},\sigma) \ \geq\ \sup_T [T]^- R(\mathbf{F}_T,\sigma)$$

where

$$(14)\qquad R(\mathbf{F}_T,\sigma) \ \geq\ \sup_{f\in \mathbf{F}_T} \sigma^2 \rho\left(\frac{\|f\|_2}{\sigma}, 1\right)$$

and hence

$$(15)\qquad R(\mathbf{F},\sigma) \ \geq\ \sup_T \sup_{f\in \mathbf{F}_T} [T]^- \sigma^2 \rho\left(\frac{\|f\|_2}{\sigma}, 1\right)$$

Proof of Lemma 1: Let $\mathbf{F}_g = \{f_g : f_g = f + g, f \in \mathbf{F}\}$. Then if X_t satisfies (1), it follows that $Y_t = X_t + \int_0^t g(s)\,ds$ satisfies

$$(16)\qquad dY_t \ =\ f_g(t)\,dt \ +\ \sigma dw_t .$$

Hence

$$(17)\qquad R(\mathbf{F}_g,\sigma) \ =\ R(\mathbf{F},\sigma) .$$

We may thus without loss of generality assume in condition 1c) that $g \equiv 0$. Now fix $T \in [1,\infty)$ and suppose we observe

$$(18)\qquad dX_t \ =\ \sum_{i=0}^{[T]^- -1} f_i(t - \frac{i}{T})\,dt \ +\ \sigma dW_t$$

where $f_i \in \mathbf{F}_T$ for $i = 0, \ldots, [T]^- - 1$. Then, since $\sum_{i=0}^{[T]^- -1} f_i(t - \frac{i}{T}) \in \mathbf{F}$ it follows that

$$(19)\qquad R(\mathbf{F},\sigma) \ \geq\ \inf_{\hat f} \sup_{f_i\in \mathbf{F}_T} E\int \Big(\sum_{i=0}^{[T]^- -1} f_i(t - \frac{i}{T}) - \hat f(t)\Big)^2 dt$$

$$(20)\qquad =\ \inf_{\hat f} \sup_{f_i\in \mathbf{F}_T} \sum_{i=0}^{[T]^- -1} E \int_{i/T}^{\frac{i+1}{T}} \Big(f_i(t - \frac{i}{T}) - \hat f(t)\Big)^2 dt .$$

Now for a prior v on F_T write $R(F_T, \sigma, v)$ for the Bayes risk in estimating f under loss $\int_0^{1/T} (f(t) - \hat{f}(t))^2 dt$ based on

(21) $$dX_t = f(t)\,dt + \sigma\,dw_t$$

where $f \in F_T$.

Then, since observing (18) is equivalent to observing $[T]^-$ independent experiments of the form (21) we have

(*) $$R(F, \sigma) \geq [T]^{-R}(F_T, \sigma, v).$$

Now since the minimax risk is the supremum of the Bayes risks we have $\sup_v R(F_T, \sigma, v) = R(F_T, \sigma)$ and hence

(**) $$R(F, \sigma) \geq [T]^- R(F_T, \sigma)$$

(22) $$R(\mathbf{F}, \sigma) \geq [T]^- R(\mathbf{F}_T, \sigma)$$

(13) is established by taking $\sup_T$ in (22). Now fix $f \in \mathbf{F}_T$. By assumption 1b), $\theta f \in \mathbf{F}_T$ for all $|\theta| \leq 1$. Hence

(23) $$R(\mathbf{F}_T, \sigma) \geq \inf_{\hat{f}} \sup_{\theta} E \int_0^{1/T} (\theta f(t) - \hat{f}(t))^2 dt$$

Now for each $\hat{f}(t)$ we may define $\hat{\theta}(t)$ by

(24) $$\hat{\theta}(t) f(t) = \hat{f}(t).$$

It then follows from (23) that

(25) $$R(\mathbf{F}_T, \sigma) \geq \inf_{\hat{\theta}} \sup_{\theta} E\left(\int_0^{1/T} (\theta f(t) - \hat{\theta}(t) f(t))^2 dt\right).$$

Let $\tilde{\theta} = \dfrac{\int \hat{\theta}(t) f^2(t)\,dt}{\int f^2(t)\,dt}$. Then $\int_0^{1/T} (\theta f(t) - \hat{\theta}(t) f(t))^2 dt = \int_0^{1/T} f^2(t)(\theta - \tilde{\theta} + \tilde{\theta} - \hat{\theta}(t))^2 dt$ $= \int_0^{1/T} f^2(t)((\theta - \tilde{\theta})^2 + (\tilde{\theta} - \hat{\theta}(t))^2)\,dt$. Hence

(26) $$\int_0^{1/T} (\theta f(t) - \hat{\theta}(t) f(t))^2 dt \geq \int_0^{1/T} (\theta f(t) - \tilde{\theta} f(t))^2 dt.$$

We can thus replace the infimum in (25) by an infimum over $\tilde{\theta}$ which yields

(27) $$R(\mathbf{F}_T, \sigma) \geq \|f\|_2^2 \inf_{\tilde{\theta}} \sup_{\theta} E(\theta - \tilde{\theta})^2.$$

Now note that $\hat{\delta} = \int f(t) X(dt)$ is sufficient for θ and $\dfrac{\hat{\delta}}{\|f\|_2^2} \sim N(\theta, \dfrac{\sigma^2}{\|f\|_2^2})$. It then follows by (11) and (12) that

$$(28)\qquad \inf_{\tilde\theta}\sup_{\theta} E(\theta-\tilde\theta)^2 = \rho\left[1, \frac{\sigma}{\|f\|_2}\right] = \frac{\sigma^2}{\|f\|_2^2}\rho\left[\frac{\|f\|_2}{\sigma}, 1\right]$$

and combining (27) and (28) yields

$$(29)\qquad R(\mathbf{F}_T, \sigma) \geq \sigma^2 \rho\left[\frac{\|f\|_2}{\sigma}, 1\right].$$

Now take $\sup_{f\in \mathbf{F}_T}$ to yield (14). (15) follows immediately from (13) and (14).

If in addition to the assumptions imposed in Lemma 1 we add 1d) then bounds on the minimax risk $R(\mathbf{F},\sigma)$ can be given in an even more convenient form which is especially useful for asymptotic results as $\sigma \downarrow 0$. An example of such an application is given at the end of this section.

Theorem 1: Suppose we observe the white noise process and that the parameter spaces $\mathbf{F}_T$ satisfy the assumption given by 1) then

$$(30)\qquad R(\mathbf{F},\sigma) \geq \sup_T \sup_{f\in \mathbf{F}_1} [T]^- \sigma^2 \rho\left[\frac{\|f\|_2}{\sigma\sqrt{T}\phi(T)}, 1\right]$$

and

$$(31)\qquad R(\mathbf{F}, \frac{\sigma}{\sqrt{T}\phi(T)}) \geq \frac{[T]^-}{T}\cdot\frac{1}{\phi^2(T)} R(\mathbf{F}_1,\sigma).$$

Proof of Theorem 1: Consider the model

$$(33)\qquad dX_t = f(t)\,dt + \frac{\sigma}{\sqrt{T}\phi(T)} dW_t, \qquad f\in \mathbf{F}_T.$$

Write E^1 for expectations taken with respect to this model. Since $f(t) \to \frac{f(Tt)}{\phi(T)}$ is 1-1 and onto from $\mathbf{F}_1$ to $\mathbf{F}_T$, (33) can be replaced by the model

$$(34)\qquad dX_t = \frac{f(Tt)}{\phi(T)} dt + \frac{\sigma}{\sqrt{T}\phi(T)} dW_t, \qquad f\in \mathbf{F}_1.$$

Write E^2 for expectations taken with respect to this model. It then follows that

$$R(\mathbf{F}_T, \frac{\sigma}{\sqrt{T}\phi(T)}) = \inf_{\hat f}\sup_{f\in \mathbf{F}_T} E^1 \int (f(t)-\hat f(t))^2 dt = \inf_{\hat f}\sup_{f\in \mathbf{F}_1} E^2 \int \left[\frac{f(Tt)}{\phi(T)} - \frac{\hat f(Tt)}{\phi(T)}\right]^2 dt$$

$$= \inf_{\hat f} \sup_{f \in F_1} \frac{1}{T\phi^2(T)} E^2 \int (f(t) - \hat f(t))^2 \, dt .$$

Now in Low (1989) it was shown that the model given by (34) is equivalent as an experiment to

(35) $$dX_t = f(t)\,dt + \sigma dW_t \qquad f \in \mathbf{F}_1 .$$

In particular it follows that

$$\inf_{\hat f} \sup_{f \in F_1} E^2 (\int (f(t) - \hat f(t))^2 \, dt) = R(\mathbf{F}_1, \sigma)$$

and therefore

(36) $$R\left[\mathbf{F}_T, \frac{\sigma}{\sqrt{T}\phi(T)}\right] = \frac{1}{T\phi^2(T)} R(\mathbf{F}_1, \sigma) .$$

Finally lemma 1 showed that $R(\mathbf{F}, \sigma) \geq [T]^- R(\mathbf{F}_T, \sigma)$ and so

$$R\left[\mathbf{F}, \frac{\sigma}{\sqrt{T}\phi(T)}\right] \geq [T]^- \cdot \frac{1}{T\phi^2(T)} R(\mathbf{F}_1, \sigma)$$

and the proof of () is complete.

Now it follows from (31) that

$$R(F, \sigma) \geq \frac{[T]^-}{T} \frac{1}{\phi^2(T)} R(F_1, \sigma\sqrt{T}\phi(T))$$

and equation (14) of lemma 1 yields

$$R(F_1, \sigma\sqrt{T}\phi(T)) \geq \sup_{f \in F_1} \sigma^2 T \phi^2(T) \rho\left(\frac{\|f\|_2}{\sigma\sqrt{T}\phi(T)}, 1\right).$$

Hence we have

() $$R(F, \sigma) \geq \sup_{f \in F_1} [T]^- \sigma^2 \rho\left(\frac{\|f\|_2}{\sigma\sqrt{T}\phi(T)}, 1\right)$$

and (30) follows on taking $\sup_T$ in ().

Upper Bounds

Upper bounds for the minimax risk can be derived from invariance ideas similar to those used in Lemma 1 and Theorem 1.

Theorem 2: If the parameter spaces $\mathbf{F}^T$ satisfy assumptions 2a), 2b) and 2c) then

(37) $$R\left[\mathbf{F}, \frac{\sigma}{\sqrt{T}\psi(T)}\right] \leq \frac{[T]^+}{T\psi^2(T)} R(\mathbf{F}^1, \sigma) .$$

Proof: Let δ_T be the collection of estimators $\hat{f}(t)$ such that for $\frac{i}{T} \le t \le \frac{i+1}{T}$, $i = 0, \ldots, [T]^- - 1$, $\hat{f}(t)$ is a function only of X_t, $\frac{i}{T} \le t \le \frac{i+1}{T}$ and for $1 - \frac{1}{T} \le t \le 1$, $\hat{f}(t)$ is a function only of X_t, $1 - \frac{1}{T} \le t \le 1$. Then

$$R(\mathbf{F},\sigma) \le R(\mathbf{F}^1,\sigma) = \inf_{\hat{f}} \sup_{f \in \mathbf{F}^1} E \int_0^1 (f(t) - \hat{f}(t))^2 dt.$$

Now by restricting attention to estimators in the class δ_T it immediately follows that

(38) $$R(\mathbf{F}^1,\sigma) \le \inf_{\hat{f} \in \delta_T} \sup_{f \in \mathbf{F}^1} \Big(\sum_{i=0}^{[T]^- - 1} E \int_{i/T}^{\frac{i+1}{T}} (f(t) - \hat{f}(t))^2 dt + E \int_{1-1/T}^{1} (f(t) - \hat{f}(t))^2 dt \Big)$$

$$\le \sum_{i=0}^{[T]^- - 1} \inf_{\hat{f} \in \delta_T} \sup_{f \in \mathbf{F}^1} E \int_{i/T}^{\frac{i+1}{T}} (f(t) - \hat{f}(t))^2 dt + \inf_{\hat{f} \in \delta_T} \sup_{f \in \mathbf{F}^1} E \int_{1-1/T}^{1} (f(t) - \hat{f}(t))^2 dt.$$

Now by 2c) for each $i = 0,1, \ldots, [T]^- - 1$

$$\inf_{\hat{f} \in \delta_T} \sup_{f \in \mathbf{F}^1} E \int_{i/T}^{\frac{i+1}{T}} (f(t) - \hat{f}(t))^2 dt \le \inf_{\hat{f}} \sup_{f \in \mathbf{F}^T} E \int_0^{1/T} (f(t) - \hat{f}(t))^2 dt$$

and

$$\inf_{\hat{f} \in \delta_T} \sup_{f \in \mathbf{F}^1} E \int_{1-1/T}^{1} (f(t) - \hat{f}(t))^2 dt \le \inf_{\hat{f}} \sup_{f \in \mathbf{F}^T} E \int_0^{1/T} (f(t) - \hat{f}(t))^2 dt.$$

Hence since $[T]^- + 1 \le [T]^+$ it follows that

$$R(\mathbf{F},\sigma) \le [T]^+ \inf_{\hat{f} \in \delta_T} \sup_{f \in \mathbf{F}^T} E \int_0^{1/T} (f(t) - \hat{f}(t))^2 dt$$

$$= [T]^+ R(\mathbf{F}^T,\sigma).$$

Then

(39) $$R(\mathbf{F}, \frac{\sigma}{\sqrt{T}\psi(T)}) \le [T]^+ R(\mathbf{F}^T, \frac{\sigma}{\sqrt{T}\psi(T)}).$$

Now an argument essentially the same as that used to show (36) in the proof of theorem 1 yields

(40) $$R(\mathbf{F}^T, \frac{\sigma}{\sqrt{T}\psi(T)}) = \frac{1}{T\psi^2(T)} R(\mathbf{F}^1,\sigma).$$

The proof of theorem 2 immediately follows on combining (39) and (40).

Upper bounds can also be given in terms of corresponding results for the pointwise estimation problem. In the following theorem we write $R(\mathbf{F}, x, \sigma)$ for the minimax risk for estimating $f(x)$. That is

$$(41) \qquad R(\mathbf{F}, x, \sigma) = \inf_{\delta} \sup_{f \in \mathbf{F}} E(f(x) - \delta(x))^2$$

where the infimum is taken over all procedures δ based on the white noise model (1).

Theorem 3: Suppose we observe the white noise process (1) then

$$(42) \qquad R(\mathbf{F}, \sigma) \le \int_0^1 R(\mathbf{F}, x, \sigma)\, dx .$$

If in addition for each c, $0 \le c \le 1$ the map

$$(43) \qquad f(t) \rightarrow f((t + c) \bmod 1)$$

is a bijection on **F** then

$$(44) \qquad R(\mathbf{F}, \sigma) \le R(\mathbf{F}, x, \sigma) = R(\mathbf{F}, 0, \sigma).$$

Proof: Given $\varepsilon > 0$, let $\delta_\varepsilon(x)$ be an estimator such that for each x

$$\sup_f E(f(x) - \delta_\varepsilon(x))^2 \le R(\mathbf{F}, x, \sigma) + \varepsilon .$$

Hence

$$\sup_f \int_0^1 E(f(x) - \delta_\varepsilon(x))^2 dx \le \int_0^1 [\sup_f E(f(x) - \delta_\varepsilon(x))^2]\, dx$$
$$= \int_0^1 R(\mathbf{F}, x, \sigma)\, dx + \varepsilon .$$

Since ε is arbitrary we have proved (42). Now if **F** satisfies the translation invariance condition given by (43) it immediately follows that

$$(45) \qquad R(\mathbf{F}, x, \sigma) = R(\mathbf{F}, 0, \sigma) \quad \forall x \in [0, 1]$$

(42) and (45) taken together yield (44).

Example 1 (continued).

As remarked earlier theorem 1 is especially useful for application to asymptotic problems as $\sigma \downarrow 0$. We now give a concrete example to show how this can be done.

Write $F(M)$, $F_1(M)$ and $F^1(M)$ for the class of functions denoted earlier by $F(1, M)$, $F_1(1, M)$ and $F^1(1, M)$ in (3). In other words

$$F(M) = \{f: [0,1] \to R: |f(x) - f(y)| \le M|x - y|,\ f(0) = f(1)\},$$

and

$$F_1(M) = F(M) \cap \{f: [0,1] \to R: f(0) = f(1) = 0\}.$$

and

$$F^1(M) = \{f: [0,1] \to R: |f(x) - f(y)| \le M|x - y|\}.$$

Furthermore if we define $F_T(M) = \left\{\frac{f(Tt)}{T} : f \in F_1(M)\right\}$ then the assumptions of theorem 1 are satisfied and yield

$$(46) \qquad R(F(M), \sigma) \ge \sup_T \sup_{f \in F_1(M)} [T]^- \sigma^2 \rho \left[\frac{\|f\|_2}{\sigma T^{3/2}}, 1\right].$$

Note that the function $\rho(x, 1)$ is an increasing and continuous function of x and hence the right hand side of (46) is equal to

$$\sup_T [T]^- \sigma^2 \rho \left[\frac{\sup_{f \in F_1(M)} \|f\|_2}{\sigma T^{3/2}}, 1\right].$$

Now the function g defined by

$$g(x) = \begin{cases} Mx & 0 \le x \le 1/2 \\ M(1-x) & 1/2 \le x \le 1 \end{cases}$$

belongs to $F_1(M)$. Moreover it is clear that for any $f \in F_1(M)$, $|f(x)| \le g(x)$ for all x. Hence

$$\sup_{f \in F_1(M)} \|f\|_2^2 = \int_0^1 g^2(x)\,dx = \frac{M^2}{12}.$$

We may thus replace (46) by

$$(47) \qquad R(F(M), \sigma) \ge \sup_T [T]^- \sigma^2 \rho \left[\frac{M}{2\sqrt{3}\sigma T^{3/2}}, 1\right].$$

If we put $\frac{d}{2} = \frac{M}{2\sqrt{3}\sigma T^{3/2}}$ then $T = \left[\frac{M}{3^{1/2} d\sigma}\right]^{2/3}$ and

$$(48) \qquad R(F(M), \sigma) \ge \sup_d \left[\frac{M^{2/3}}{3^{1/3}\sigma^{2/3} d^{2/3}}\right]^- \sigma^2 \rho\left(\frac{d}{2}, 1\right).$$

Analysis of (48) is made easy by an analysis of the functional $\sup_d d^\alpha \rho\left(\frac{d}{2}, 1\right)$ given in Donoho and Liu (1987). In our case $\alpha = -\frac{2}{3}$ and Donoho and Liu (1987) show that

(49) $$\sup_d d^{-2/3} \rho(\frac{d}{2}, 1) = 0.283 .$$

Let d^* be the value of d attaining the supremum in (49). Then $0 < d^* < \infty$ and hence for any $\tau > 0$.

$\frac{M^{2/3}}{3^{1/3}\sigma^{2/3}d^{2/3}}$ is an integer for infinitely many values of σ satisfying $0 < \sigma < \tau$. For these values of σ

(50) $$R(F(M), \sigma) \geq \frac{M^{2/3}}{3^{1/3}} \sigma^{4/3} 0.283$$

and hence

(51) $$\underline{\lim}_{\sigma \downarrow 0} \sigma^{-4/3} R(F(M), \sigma) \geq 0.196 M^{2/3} .$$

It is also easy to see how theorem 2 can be used to find upper bounds for the rate of convergence. Set $\sigma = \frac{1}{T^{3/2}}$ then theorem 2 shows that

(52) $$R(F(M), \sigma) \leq \frac{[T]^+}{T \cdot T^2} R(F^1(M), 1) \leq \frac{(\frac{1}{\sigma})^{2/3} + 1}{(\frac{1}{\sigma})^2} R(F^1(M), 1) .$$

Hence

(53) $$\overline{\lim}_{\sigma \downarrow 0} \sigma^{-4/3} R(F(M), \sigma) \leq R(F^1(M), 1)$$

(51) and (53) taken together of course yield $\sigma^{-4/3}$ as an optimal rate since $R(F^1(M), 1) < \infty$. In this example theorem 3 can be used to give a more explicit bound since

$$R(F(M), \sigma) \leq R(F(M), 0, \sigma)$$

and we may bound $R(F(M), 0, \sigma)$ from above by using the optimal linear estimator for this pointwise problem, essentially given in Sacks and Ylvisaker [1981] and Donoho and Liu [1987], yielding for sufficiently small σ.

$$R(F(M), 0, \sigma) \leq \frac{M^{2/3} \sigma^{4/3}}{3^{1/3}} .$$

Hence

(54) $$0.196 M^{2/3} \leq \overline{\lim}_{\sigma \downarrow 0} \sigma^{-4/3} R(F(M), \sigma) \leq \frac{M^{2/3}}{3^{1/3}}$$

It is possible to improve on the upper bound in (54) by using an upper bound given for the minimax risk for an ellipsoidal parameter space considered by Pinsker [1980].

Let $P(M) = \{f: [0,1] \to R, \int_0^1 f'^2(x)\,dx \le M^2, f(0) = f(1)\}$. Then $F(M) \subseteq P(M)$ and Pinsker showed that

$$\lim_{\sigma \downarrow 0} \sigma^{-4/3} R(P(M), \sigma) = \frac{3^{1/3}}{(2\pi)^{2/3}} M^{2/3}$$
$$= 0.424 M^{2/3}.$$

Hence

$$0.196 M^{2/3} \le \overline{\lim_{\sigma \downarrow 0}}\, \sigma^{-4/3} R(F(M), \sigma) \le 0.424 M^{2/3}.$$

The ratio $\frac{0.424}{0.196} = 2.16$.

References

Brown, L.D. and Feldman, I. (1990). Manuscript.

Brown, L.D. and Low, M.G. (1990). Asymptotic equivalence of nonparametric regression and White Noise. Tech. Report.

Casella, G. and Strawderman, W.E. (1981). Estimating a bounded normal mean. *Ann. Stat.* **9**, 870-878.

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. (1988). Geometrizing rates of convergence III. Technical Report, University of California, Berkeley.

Donoho, D.L. and Liu, R.C. (1987). On the minimax estimation of linear functionals. Technical Report, University of California, Berkeley.

Donoho, D.L. and Low, M.G. (1990a). Renormalization exponents and optimal pointwise rates of convergence. Technical Report, University of California, Berkeley.

Donoho, D.L. and Low, M.G. (1990b). White noise approximation and minimax risk. Technical Report, University of California, Berkeley.

Donoho, D.L., McGibbon, B. and Liu, R.C. (1988). Unpublished manuscript.

Efroimovich, S.Y. and Pinsker, M.S. (1982). Estimation of square-integrable probability density of a random variable. Problems of Information Transmission. (1983) 175-189.

Ibragimov, I.A. and Has'minskii, R.Z. (1984). Nonparametric estimation of the value of a linear functional in a Gaussian white noise. Theory of Probability and its Applications 1-17.

Ibragimov, I.A. and Has'minskii, R.Z. (1981). Statistical Estimation. Springer Verlag.

Low, M.G. (1989a). Invariance and rescaling of infinite dimensional Gaussian shift experiments. Technical Report, University of California, Berkeley.

Low, M.G. (1989b). Local convergence of nonparametric density estimation problems to gaussian shift experimental, on a Hilbert Space. Technical Report, University of California, Berkeley.

Pinsker, M.S. (1980). Optimal filtration of square-integrable signals in Gaussian noise. Problems of Information Transmission. 16 (2), 52-68.

Sacks, J. and Ylvisaker, D. (1981). Asymptotically optimum kernels for density estimation at a point. *Ann. Stat.* 9 (2), 334-346.

Stone, C. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Stat.* **10, 4**, 1040-1053.