Adaptive estimation of linear functionals under different performance measures

T. TONY CAI* and MARK G. LOW**

Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA. E-mail: *tcai@wharton.upenn.edu; **pw@wharton.upenn.edu

Lower bounds are given for probabilistic error subject to a mean squared error constraint. Consequences for the expected length of variable length confidence intervals centred on adaptive estimators are given. It is shown that in many contexts centring confidence intervals on adaptive estimators must lead either to poor coverage probability or unnecessarily long intervals.

Keywords: adaptive estimation; confidence intervals; coverage probability; expected length; modulus of continuity; white noise model

1. Introduction

Minimax theory for estimating linear functionals in nonparametric function estimation is now well developed, particularly in Gaussian settings:

$$\mathrm{d}Y(t) = f(t)\mathrm{d}t + n^{-1/2}\,\mathrm{d}W(t), \qquad -\tfrac{1}{2} \leqslant t \leqslant \tfrac{1}{2}, \tag{1}$$

where $W(t)$ is standard Brownian motion and

$$Y(i) = f(i) + n^{-1/2}z_i, \qquad i \in \mathcal{M}, \tag{2}$$

z_i being independent and identically distributed standard normal random variables and $\mathcal{M}$ a finite or countably infinite index set.

In particular, for these models minimax theory for mean squared error, confidence intervals and probabilistic error can all be precisely characterized by a modulus of continuity introduced by Donoho and Liu (1991). More specifically, for any linear functional T and convex parameter space $\mathcal{F}$ the minimax mean squared error is of order $\omega^2(1/\sqrt{n}, \mathcal{F})$ where the modulus $\omega(\epsilon, \mathcal{F})$ is defined by

$$\omega(\epsilon, \mathcal{F}) = \sup\{|Tg - Tf| : \|g - f\|_2 \leqslant \epsilon;\ f,\ g \in \mathcal{F}\}, \tag{3}$$

$\|\cdot\|_2$ being the $L_2(-\tfrac{1}{2}, \tfrac{1}{2})$ function norm in the white noise model (1) and the ℓ_2 sequence norm over the index set $\mathcal{M}$ in the Gaussian model (2). Moreover, linear procedures can be constructed which have maximum risk within a small constant factor of the minimax risk. See Ibragimov and Hasminskii (1984), Donoho and Liu (1991) and Donoho (1994) for precise versions of these results.

Optimal rates of convergence have also been given in terms of the probability that the

estimator is close to the unknown value of the functional; see, for example, Weiss and Wolfowitz (1967) and Farrell (1972). From this point of view $\omega(1/\sqrt{n}, \mathcal{F})$ is the optimal rate for estimating the linear functional Tf over a convex parameter space $\mathcal{F}$. More precisely, results in Donoho (1994) and Cai and Low (2003) show that, for any $\alpha > 0$, there exists a linear estimator $\hat{T}$ such that

$$\sup_{f \in \mathcal{F}} P_f \left(|\hat{T} - Tf| > \frac{3}{2} \omega \left(\frac{z_{\alpha/2}}{\sqrt{n}}, \mathcal{F} \right) \right) \leq \alpha \tag{4}$$

and also that, for any procedure $\hat{T}$ and any $\alpha < \frac{1}{2}$,

$$\sup_{f \in \mathcal{F}} P_f \left(|\hat{T} - Tf| > \frac{1}{2} \omega \left(\frac{z_\alpha}{\sqrt{n}}, \mathcal{F} \right) \right) > \alpha. \tag{5}$$

Note that these bounds on probabilistic error have direct consequences for the construction of fixed length confidence intervals and, in particular, show that for any given coverage probability the shortest fixed length interval has length of order $\omega(1/\sqrt{n}, \mathcal{F})$. See Donoho (1994) for further analysis of fixed length intervals.

An important goal in nonparametric function estimation is the construction of estimators which perform well according to a number of different criteria. Typically, a loss function is first fixed and the goal is to construct adaptive procedures which perform well over a collection of parameter spaces. In this paper attention is focused on criteria which connect the problem of adaptive estimation under mean squared error with that of the construction of confidence intervals. In nonparametric function estimation it is common to construct confidence intervals centred on adaptive estimators. We examine whether such practice can yield good confidence intervals.

To address such questions we first focus on the multiple goal of finding estimators which have both good mean squared error performance and also optimal probabilistic error performance. In Section 2 we quantify the penalty that must be paid for probabilistic error over one parameter space given that the estimator performs well under mean squared error over another parameter space. In another direction, if an estimator performs well under probabilistic error we quantify the penalty that must be paid on mean squared error.

In Section 3 we turn to the analysis of confidence intervals centred on adaptive mean squared error estimators. We show that one consequence of the results given in Section 2 is that centring confidence intervals on adaptive mean squared error estimators in general yields suboptimal confidence procedures. Either the resulting interval has poor coverage probability or it is unnecessarily long. The results are illustrated by examples. The proofs of the main results are postponed to Section 4.

2. Constrained error bounds

In nonparametric function estimation it is often of interest to find an estimator that is near optimal from a number of different points of view. In this section known results about adaptive estimation are first reviewed. Then the problem of constructing estimators which

have both good mean squared error performance and optimal probabilistic performance is considered. It is shown that sometimes these goals compete with each other in such a way that both cannot be simultaneously realized. More specifically, a probabilistic lower bound on the actual error subject to an upper bound on the mean squared error is given. Implications for the expected length of confidence intervals centred at adaptive estimators are given in Section 3.

2.1. Adaptive estimation

Much attention in the nonparametric function estimation literature has focused on the construction of adaptive estimators, those which are within a small factor of the minimax performance simultaneously over a collection of convex parameter spaces. In this subsection some known results are recalled which will place into context the main results given later in the paper.

It is well known that adaptive estimators often exist for problems of estimating the whole function under integrated squared error; see, for example, Efromovich and Pinsker (1984). In contrast, for estimating a function at a point, Lepski (1990) showed that any estimator must have maximum mean squared error over one Lipschitz class inflated by a logarithmic factor whenever it is minimax rate-optimal over another Lipschitz class.

A more general theory for adaptive estimation of linear functionals under mean squared error has recently been given in Cai and Low (2002). Geometric quantities, the ordered and between-class moduli of continuity, were introduced. For a linear functional T and parameter spaces $\mathcal{F}$ and $\mathcal{G}$, the between-class modulus $\omega_+(\epsilon, \mathcal{F}, \mathcal{G})$ is defined by

$$\omega_+(\epsilon, \mathcal{F}, \mathcal{G}) = \sup\left\{|Tg - Tf| : \|g - f\|_2 \leqslant \epsilon; f \in \mathcal{F}, g \in \mathcal{G}\right\}, \tag{6}$$

where once again $\|\cdot\|_2$ is the $L_2(-\frac{1}{2}, \frac{1}{2})$ function norm in the white noise model (1) and the ℓ_2 sequence norm over the index set $\mathcal{M}$ in the Gaussian model (2). When $\mathcal{G} = \mathcal{F}$, $\omega_+(\epsilon, \mathcal{F}, \mathcal{F})$ is the modulus of continuity over $\mathcal{F}$ given in (3). The between-class modulus of continuity is used in Cai and Low (2002) to quantify precisely the degree of adaptability for estimating a linear functional under mean squared error.

Consider two function classes $\mathcal{F}_1$ and $\mathcal{F}_2$ with $\mathcal{F}_1 \cap \mathcal{F}_2 \neq \emptyset$. Let T be a linear functional and suppose that

$$\sup_{f \in \mathcal{F}_1} \mathrm{E}_f(\hat{T} - Tf)^2 \leqslant \gamma_n^{-2}\omega_+^2\left(\frac{1}{\sqrt{n}}, \mathcal{F}_1, \mathcal{F}_2\right), \tag{7}$$

for some $\gamma_n > e$. Then Cai and Low (2002) show that

$$\sup_{f \in \mathcal{F}_2} \mathrm{E}_f(\hat{T} - Tf)^2 \geqslant \left(\omega_+\left(\sqrt{\frac{\ln \gamma_n^2}{n}}, \mathcal{F}_1, \mathcal{F}_2\right) - \omega_+\left(\frac{1}{\sqrt{n}}, \mathcal{F}_1, \mathcal{F}_2\right)\right)^2. \tag{8}$$

This bound can easily be applied under the following commonly occurring mild regularity conditions on the moduli which, for example, hold when estimating the value of a function at a point over Lipschitz classes, Sobolev spaces or Besov spaces.

Condition A.

$$\lim_{\epsilon \to 0} \frac{\omega_+(\epsilon, \mathcal{F}_1, \mathcal{F}_2)}{\omega(\epsilon, \mathcal{F}_1)} = \infty, \tag{9}$$

$$\overline{\lim_{\epsilon \to 0}} \frac{\omega(\epsilon, \mathcal{F}_2)}{\omega_+(\epsilon, \mathcal{F}_1, \mathcal{F}_2)} < \infty, \tag{10}$$

$$\overline{\lim_{\epsilon \to 0}} \frac{\omega_+(\epsilon, \mathcal{F}_1, \mathcal{F}_2)}{\omega_+(D\epsilon, \mathcal{F}_1, \mathcal{F}_2)} < 1, \qquad \text{for some } D > 1. \tag{11}$$

Under Condition A if an estimator is rate-optimal in mean squared error over $\mathcal{F}_1$ it must satisfy (7) with

$$\gamma_n^{-2} = C \frac{\omega^2(1/\sqrt{n}, \mathcal{F}_1),}{\omega_+^2(1/\sqrt{n}, \mathcal{F}_1, \mathcal{F}_2)}$$

for some $C > 0$, and hence, by (9), $\gamma_n^2 \to \infty$. It then follows that a penalty in mean squared error must be paid over $\mathcal{F}_2$ since by (8), (10) and (11),

$$\sup_{f \in \mathcal{F}_2} \mathrm{E}_f(\hat{T} - Tf)^2 \geqslant \omega_+^2 \left(\sqrt{\frac{\ln \gamma_n^2}{n}}, \mathcal{F}_1, \mathcal{F}_2 \right) (1 + o(1)) \gg \omega^2 \left(\frac{1}{\sqrt{n}}, \mathcal{F}_2 \right). \tag{12}$$

Moreover, adaptive estimators are constructed in Cai and Low (2002) which are minimax rate-optimal over $\mathcal{F}_1$ and which also have maximum risk over $\mathcal{F}_2$ within a constant factor of the lower bound given in (12).

The theory of adaptation under probabilistic error is quite different from this mean squared error theory. In fact fully rate-optimal adaptive estimation is possible under probabilistic error. Let $\mathcal{F}_1 \subset \mathcal{F}_2$ be two nested convex parameter spaces. Then Cai and Low (2003) show that for any linear functional T there is a procedure $\hat{T}$ and a constant $C > 0$ such that

$$\sup_{f \in \mathcal{F}_1} P_f \left(|\hat{T} - Tf| > C\omega \left(\frac{z_{\alpha/2}}{\sqrt{n}}, \mathcal{F}_1 \right) \right) \leqslant \alpha \tag{13}$$

and

$$\sup_{f \in \mathcal{F}_2} P_f \left(|\hat{T} - Tf| > C\omega \left(\frac{z_{\alpha/2}}{\sqrt{n}}, \mathcal{F}_2 \right) \right) \leqslant \alpha. \tag{14}$$

Hence, for probabilistic error the minimax rate can always be achieved simultaneously over both $\mathcal{F}_1$ and $\mathcal{F}_2$.

2.2. Bounds on probabilistic error under a mean squared error constraint

In this section attention is focused on providing a general lower bound on the probabilistic error under a constraint on the mean squared error in terms of the between-class modulus

$\omega_+(\epsilon, \mathcal{F}_1, \mathcal{F}_2)$ defined in (6). This bound is given in Theorem 1. The corollaries that follow provide more specific results under stronger, but commonly occurring, assumptions on the moduli of continuity. Examples illustrating the results are given in Section 2.3.

Theorem 1. *Consider two parameter spaces $\mathcal{F}_1$ and $\mathcal{F}_2$ with $\mathcal{F}_1 \cap \mathcal{F}_2 \neq \varnothing$. Let T be a linear functional and suppose that*

$$\sup_{f \in \mathcal{F}_1} E_f(\hat{T} - Tf)^2 \leq \gamma_n^{-2} \omega_+^2 \left(\frac{1}{\sqrt{n}}, \mathcal{F}_1, \mathcal{F}_2 \right), \tag{15}$$

for some $\gamma_n > 1$. Then, for any $0 < \rho \leq 1$ and $0 < \lambda < 1$,

$$\sup_{f \in \mathcal{F}_2} P_f \left(|\hat{T} - Tf| \geq \lambda \omega_+ \left(\sqrt{\frac{\rho \ln \gamma_n^2}{n}}, \mathcal{F}_1, \mathcal{F}_2 \right) \right) \geq 1 - \frac{\omega_+(1/\sqrt{n}, \mathcal{F}_1, \mathcal{F}_2)}{(1-\lambda)\gamma_n^{1-\rho} \omega_+(\sqrt{\rho(\ln \gamma_n^2)/n}, \mathcal{F}_1, \mathcal{F}_2)}. \tag{16}$$

An interesting consequence of Theorem 1 is given in the following corollary for estimators which are mean squared error rate-optimal over $\mathcal{F}_1$.

Corollary 1. *Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be two convex parameter spaces with $\mathcal{F}_1 \cap \mathcal{F}_2 \neq \varnothing$ and let T be a linear functional. Suppose that $\hat{T}$ is minimax rate-optimal over $\mathcal{F}_1$, that is,*

$$\sup_{f \in \mathcal{F}_1} E_f(\hat{T} - Tf)^2 \leq C \omega^2 \left(\frac{1}{\sqrt{n}}, \mathcal{F}_1 \right), \tag{17}$$

and suppose that

$$\xi_n = \frac{\omega_+(1/\sqrt{n}, \mathcal{F}_1, \mathcal{F}_2)}{\omega(1/\sqrt{n}, \mathcal{F}_1)} \to \infty.$$

Then

$$\lim_{n \to \infty} \sup_{f \in \mathcal{F}_2} P_f \left(|\hat{T} - Tf| \geq \frac{1}{2} \omega_+ \left(\sqrt{\frac{\ln \xi_n^2}{n}}, \mathcal{F}_1, \mathcal{F}_2 \right) \right) = 1. \tag{18}$$

Remark. The conditions in Corollary 1 are commonly satisfied. For example, suppose that $\mathcal{F}_1$ and $\mathcal{F}_2$ are symmetric, namely that $-f \in \mathcal{F}_i$ whenever $f \in \mathcal{F}_i$, that they are nested $\mathcal{F}_1 \subset \mathcal{F}_2$ and that minimax rates of convergence differ on $\mathcal{F}_1$ and $\mathcal{F}_2$, that is,

$$\frac{\omega(1/\sqrt{n}, \mathcal{F}_2)}{\omega(1/\sqrt{n}, \mathcal{F}_1)} \to \infty.$$

Then the condition $\xi_n \to \infty$ always holds. If the algebraic part of the rates over $\mathcal{F}_1$ and $\mathcal{F}_2$ also differ then Condition A holds and there is a constant $d > 0$ such that

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}_2} P_f\left(|\hat{T} - Tf| \geqslant d\omega\left(\sqrt{\frac{\ln\xi_n}{n}}, \mathcal{F}_2\right)\right) = 1. \tag{19}$$

This result should also be contrasted with adaptation results under probabilistic error given in (13) and (14) where the minimax rate can always be achieved simultaneously over both $\mathcal{F}_1$ and $\mathcal{F}_2$, whereas under Condition A if an estimator is mean squared error rate-optimal over $\mathcal{F}_1$ it must pay a penalty for probabilistic error over $\mathcal{F}_2$.

Theorem 1 also yields a lower bound on the mean squared error for estimators with good probabilistic performance, as given in the following corollary.

Corollary 2. *Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be convex parameter spaces with $\mathcal{F}_1 \cap \mathcal{F}_2 \neq \varnothing$ and let T be a linear functional. Suppose that Condition A holds and that $\hat{T}$ satisfies*

$$\sup_{f\in\mathcal{F}_2} P_f\left(|\hat{T} - Tf| > \lambda\omega\left(\frac{z_{\alpha/2}}{\sqrt{n}}, \mathcal{F}_2\right)\right) \leqslant \alpha, \tag{20}$$

for some constant $\lambda > 0$. Then the maximum mean squared error of $\hat{T}$ over $\mathcal{F}_1$ satisfies

$$\sup_{f\in\mathcal{F}_1} E_f(\hat{T} - Tf)^2 \geqslant c\omega^2\left(\frac{1}{\sqrt{n}}, \mathcal{F}_2\right) \tag{21}$$

for some constant $c > 0$ and hence is far from minimax rate-optimal over $\mathcal{F}_1$.

Further clarifications can be made in the common case when the parameter spaces are convex and the moduli are Hölderian, that is,

$$\omega_+(\epsilon, \mathcal{F}_i, \mathcal{F}_j) = C\epsilon^q(1 + o(1)),$$

for some constants $0 < q \leqslant 1$ and $C > 0$. In such cases write $q(\mathcal{F}_i, \mathcal{F}_j)$ for the exponent q in this formula.

Corollary 3. *Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be convex parameter spaces with $\mathcal{F}_1 \cap \mathcal{F}_2 \neq \varnothing$. Suppose that the moduli $\omega_+(\epsilon, \mathcal{F}_i, \mathcal{F}_j)$ are Hölderian with exponent $q_{i,j} = q(\mathcal{F}_i, \mathcal{F}_j)$, for $1 \leqslant i, j \leqslant 2$. If $q_{1,2} = q_{2,2} < q_{1,1}$ or $q_{1,2} < q_{2,2} \leqslant q_{1,1}$, then, for any estimator $\hat{T}$ that is minimax rate-optimal under mean squared error over $\mathcal{F}_1$, there exists a constant $d > 0$ such that*

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}_2} P_f\left(|\hat{T} - Tf| \geqslant d\left(\frac{\ln n}{n}\right)^{q_{1,2}/2}\right) = 1, \tag{22}$$

and hence $\hat{T}$ is not minimax rate-optimal under probabilistic error over $\mathcal{F}_2$.

Remark. The conditions in Corollary 3 are satisfied in many common estimation problems such as estimating the function at a point over Hölder spaces or Besov spaces.

2.3. Example

We illustrate the above adaptation results by contrasting them with adaptation under probabilistic error and adaptation under mean squared error as discussed in Section 2.1. For example, consider the white noise model (1) and the special case of estimating a function at a point, say $f(0)$, over two Lipschitz classes $\mathcal{F}_1 = F(\beta_1, M)$ and $\mathcal{F}_2 = F(\beta_2, M)$, with $0 < \beta_2 < \beta_1 \leq 1$, where the Lipschitz function classes $F(\beta_i, M)$ are defined as

$$F(\beta_i, M) = \{f : [-\tfrac{1}{2}, \tfrac{1}{2}] \to \mathbb{R}, |f(x) - f(y)| \leq M|x - y|^{\beta_i}\}. \tag{23}$$

Standard calculations as in Donoho and Liu (1987) and Cai and Low (2002) yield

$$\omega(\epsilon, \mathcal{F}_1) \asymp \epsilon^{2\beta_1/(1+2\beta_1)} \quad \text{and} \quad \omega_+(\epsilon, \mathcal{F}_1, \mathcal{F}_2) \asymp \omega(\epsilon, \mathcal{F}_2) \asymp \epsilon^{2\beta_2/(1+2\beta_2)}. \tag{24}$$

In this case it follows from (7) and (8) that if $\hat{f}(0)$ is a rate-optimal estimator over $\mathcal{F}_1$ under mean squared error, that is,

$$\sup_{f \in \mathcal{F}_1} \mathrm{E}_f(\hat{f}(0) - f(0))^2 \leq C_1 n^{-2\beta_1/(1+2\beta_1)} \tag{25}$$

for some constant $C_1 > 0$, then

$$\sup_{f \in \mathcal{F}_2} \mathrm{E}_f(\hat{f}(0) - f(0))^2 \geq C_2 \left(\frac{\ln n}{n}\right)^{2\beta_2/(1+2\beta_2)} \tag{26}$$

for another constant $C_2 > 0$, recovering the result of Lepski (1990). See also Brown and Low (1996), Efromovich and Low (1994) and Lepski and Spokoiny (1997) for further developments.

The general results of Cai and Low (2003) as given in (13) and (14) show that there is an estimator $\tilde{f}(0)$, not depending on α, satisfying

$$\lim_{n \to \infty} \sup_{f \in \mathcal{F}_i} P_f\left(|\tilde{f}(0) - f(0)| \leq C(\alpha) n^{-\beta_i/(1+2\beta_i)}\right) \geq 1 - \alpha, \tag{27}$$

for both $i = 1$ and $i = 2$, where the constant $C(\alpha) > 0$ depends only on α. Hence, from (5) under probabilistic error it is possible to attain the exact minimax rate simultaneously over $\mathcal{F}_1$ and $\mathcal{F}_2$.

On the other hand, if $\hat{f}(0)$ is a rate-optimal estimator over $\mathcal{F}_1$ under mean squared error satisfying (25), then it follows from (22) that

$$\lim_{n \to \infty} \sup_{f \in \mathcal{F}_2} P_f\left(|\hat{f}(0) - f(0)| \geq d\left(\frac{\ln n}{n}\right)^{\beta_2/(1+2\beta_2)}\right) = 1, \tag{28}$$

for some constant $d > 0$. Comparing (28) with (27), it is clear that $\hat{f}(0)$ is not optimal probabilistically.

Note that equation (28) also directly yields the previously known bound on mean squared error given in (26). Finally, if $\hat{f}(0)$ is a rate-optimal estimator under probabilistic error over $\mathcal{F}_2$, then it follows from Corollary 2 that the mean squared error of $\hat{f}(0)$ over $\mathcal{F}_1$ must satisfy

$$\sup_{f \in \mathcal{F}_1} \mathrm{E}_f (\hat{f}(0) - f(0))^2 \geqslant C n^{-2\beta_2/(1+2\beta_2)} \gg C n^{-2\beta_1/(1+2\beta_1)}.$$

The above results show that questions of adaptability can depend strongly on how such procedures are evaluated. In particular, probabilistic adaptation can be achieved whereas fully mean squared error adaptation is typically unattainable. In the next section we shall connect the problem of adaptive estimation with that of the construction of confidence intervals.

3. Nonparametric confidence intervals

The construction of honest confidence intervals in nonparametric function estimation is an important but challenging problem. The theory is straightforward for a given convex parameter space $\mathcal{F}$ and linear functional T. In such a case optimal fixed length intervals can be centred on minimax rate-optimal mean squared error estimators, in which case the $(1 - \alpha)$-level confidence intervals for Tf over $\mathcal{F}$ have length of order $\omega(z_{\alpha/2}/\sqrt{n}, \mathcal{F})$. In fact these confidence intervals can be centred on linear estimators with have standard deviation and maximum bias of this same order.

An adaptation theory for the construction of confidence intervals has been developed in Cai and Low (2004). For two nested convex parameter spaces $\mathcal{F}_1$ and $\mathcal{F}_2$ with $\mathcal{F}_1 \subset \mathcal{F}_2$, the theory can be summarized as follows. Denote by $L(CI)$ the length of a confidence interval CI and let $\mathcal{I}(\alpha)$ be the collection of confidence intervals with coverage probability of at least $1 - \alpha$ over $\mathcal{F}_2$. Cai and Low (2004) shows that there is a constant $C > 0$ such that, for any confidence interval $CI \in \mathcal{I}(\alpha)$,

$$\sup_{f \in \mathcal{F}_i} \mathrm{E}_f L(CI) \geqslant C \omega_+ \left(\frac{z_{\alpha/2}}{\sqrt{n}}, \mathcal{F}_i, \mathcal{F}_2 \right). \tag{29}$$

Moreover, a confidence interval interval $CI \in \mathcal{I}(\alpha)$ is constructed which satisfies

$$\sup_{f \in \mathcal{F}_i} \mathrm{E}_f L(CI) \leqslant D \omega_+ \left(\frac{z_{\alpha/2}}{\sqrt{n}}, \mathcal{F}_i, \mathcal{F}_2 \right) \tag{30}$$

for some constant $D > 0$, for both $i = 1$ and $i = 2$. In particular, such a confidence interval has a rate-optimal expected length over both $\mathcal{F}_1$ and $\mathcal{F}_2$ given that it has guaranteed coverage probability of at least $1 - \alpha$ over $\mathcal{F}_2$. In this sense the confidence interval can be called adaptively rate-optimal. It should be stressed that the maximum expected length over $\mathcal{F}_1$ also depends on the parameter space $\mathcal{F}_2$ through the between-class modulus.

3.1. Confidence intervals centred on adaptive estimators

As mentioned in the Introduction, it is a common practice in nonparametric function estimation to centre confidence intervals on adaptive estimators. In this section the consequences of such an approach are examined. In particular, the lower bounds on

probabilistic error given in Section 2 have immediate implications for the expected length of confidence intervals which have guaranteed coverage probability.

Theorem 2. *Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be convex parameter spaces with $\mathcal{F}_1 \cap \mathcal{F}_2 \neq \varnothing$ and let T be a linear functional. Let $\hat{T}$ be an estimator which satisfies*

$$\sup_{f \in \mathcal{F}_1} \mathrm{E}_f(\hat{T} - Tf)^2 \leqslant \gamma_n^{-2} \omega_+^2 \left(\frac{1}{\sqrt{n}}, \mathcal{F}_1, \mathcal{F}_2 \right), \tag{31}$$

where $\gamma_n \to \infty$. Let CI be any confidence interval for Tf centred at $\hat{T}$ with coverage probability of at least $1 - \alpha$ over $\mathcal{F}_2$. Then

$$\lim_{n \to \infty} \sup_{f \in \mathcal{F}_2} P_f \left(L(CI) \geqslant \omega_+ \left(\sqrt{\frac{\ln \gamma_n^2}{n}}, \mathcal{F}_1, \mathcal{F}_2 \right) \right) \geqslant 1 - \alpha \tag{32}$$

and consequently, for some constant $C > 0$,

$$\sup_{f \in \mathcal{F}_2} \mathrm{E}_f L(CI) \geqslant C \omega_+ \left(\sqrt{\frac{\ln \gamma_n^2}{n}}, \mathcal{F}_1, \mathcal{F}_2 \right). \tag{33}$$

Remark. Note that under Condition A,

$$\omega_+ \left(\sqrt{\frac{\ln \gamma_n^2}{n}}, \mathcal{F}_1, \mathcal{F}_2 \right) \gg \omega \left(\frac{1}{\sqrt{n}}, \mathcal{F}_2 \right).$$

Hence, it follows from Theorem 2 that any confidence interval centred at an estimator which has maximum mean squared error over $\mathcal{F}_1$ converging at a rate faster than the minimax mean squared error over $\mathcal{F}_2$ must either have maximum expected length over $\mathcal{F}_2$ larger than necessary or have poor coverage probability.

It is also interesting to consider the special case of Hölderian moduli as is summarized in the following corollary.

Corollary 4. *Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be convex parameter spaces with $\mathcal{F}_1 \cap \mathcal{F}_2 \neq \varnothing$. Suppose that the moduli $\omega_+(\epsilon, \mathcal{F}_i, \mathcal{F}_j)$ are Hölderian with exponent $q_{i,j} = q(\mathcal{F}_i, \mathcal{F}_j)$, for $1 \leqslant i, j \leqslant 2$. If $q_{1,2} = q_{2,2} < q_{1,1}$ or $q_{1,2} < q_{2,2} \leqslant q_{1,1}$, then any confidence interval CI centred at a mean squared error adaptive estimator $\hat{T}$ with coverage probability of at least $1 - \alpha$ over $\mathcal{F}_1 \cup \mathcal{F}_2$ satisfies, for some constant $d > 0$,*

$$\lim_{n \to \infty} \sup_{f \in \mathcal{F}_2} P_f \left(L(CI) \geqslant d \left(\frac{\ln n}{n} \right)^{q_{1,2}/2} \right) \geqslant 1 - \alpha \tag{34}$$

and consequently, for some $C > 0$,

$$\sup_{f \in \mathcal{F}_2} \mathrm{E}_f L(CI) \geqslant C \left(\frac{\ln n}{n} \right)^{q_{1,2}/2}. \tag{35}$$

3.2. Examples

Now again consider the white noise model (1) and the problem of estimating a function at a point, say $f(0)$, over the two Lipschitz classes $\mathcal{F}_1 = F(\beta_1, M)$ and $\mathcal{F}_2 = F(\beta_2, M)$ with $0 < \beta_2 < \beta_1 \leqslant 1$, discussed in Section 2. Note that $\mathcal{F}_1 \subset \mathcal{F}_2$. In this example confidence interval results differ dramatically from those of mean squared error.

From the moduli given in (24) and the lower bound in (29), any confidence interval with coverage probability of at least $1 - \alpha$ over $\mathcal{F}_2$ must have expected length satisfying

$$\sup_{f \in \mathcal{F}_i} \mathrm{E}_f L(CI) \geqslant C n^{-\beta_2/(1+2\beta_2)}$$

for both $i = 1$ and $i = 2$. Hence, over $\mathcal{F}_1$ the expected length must be much longer than would be the case if an optimal confidence interval was constructed with the knowledge that $f \in \mathcal{F}_1$. It is easy to see that this bound can easily be attained by using optimal fixed length confidence intervals over $\mathcal{F}_2$.

Now suppose $\hat{f}(0)$ is an estimator which has maximum mean squared error over $\mathcal{F}_1$ which converges at a rate n^{-r}, where $r > 2\beta_2/(1 + 2\beta_2)$. Then for any confidence interval CI centred at $\hat{f}(0)$ with coverage probability of at least $1 - \alpha$ over $\mathcal{F}_2$, there exists some constant $d > 0$ such that

$$\varliminf_{n \to \infty} \sup_{f \in \mathcal{F}_2} P_f \left(L(CI) \geqslant d \left(\frac{\ln n}{n} \right)^{\beta_2/(1+2\beta_2)} \right) \geqslant 1 - \alpha \tag{36}$$

and consequently,

$$\sup_{f \in \mathcal{F}_2} \mathrm{E}_f L(CI) \geqslant d(1 - \alpha) \left(\frac{\ln n}{n} \right)^{\beta_2/(1+2\beta_2)}. \tag{37}$$

Hence confidence intervals centred on a mean squared error rate adaptive estimator must have a longer maximum expected length over $\mathcal{F}_2$.

It should be noted that the penalty in the expected length of the confidence interval need only be paid on $\mathcal{F}_2$. More precisely, we construct below a confidence interval which has a given coverage probability over $\mathcal{F}_2$ and which satisfies

$$\sup_{f \in \mathcal{F}_1} \mathrm{E}_f L(CI) \leqslant C n^{-\beta_2/(1+2\beta_2)} \tag{38}$$

and

$$\sup_{f \in \mathcal{F}_2} \mathrm{E}_f L(CI) \leqslant C \left(\frac{\ln n}{n} \right)^{\beta_2/(1+2\beta_2)}, \tag{39}$$

for some constant $C > 0$.

For $i = 1, 2$, let $h_{n,i} = n^{-1/(2\beta_i+1)}$ and let $h_{n,3} = (n^{-1} \ln n)^{1/(2\beta_2+1)}$. Let

$$\hat{T}_{n,i} = \frac{1}{2h_{n,i}} (Y(h_{n,i}) - Y(-h_{n,i})), \qquad i = 1, 2, 3.$$

It is then easy to check that, for $i = 1, 2$,

$$\sup_{f \in \mathcal{F}_i} \mathrm{E}_f(\hat{T}_{n,i} - f(0))^2 \leqslant (M^2 + 1)n^{-2\beta_i/(2\beta_i+1)} \tag{40}$$

and

$$\sup_{f \in \mathcal{F}_2} \mathrm{E}_f(\hat{T}_{n,3} - f(0))^2 \leqslant \left(M^2 + \frac{1}{2\ln n}\right)\left(\frac{\ln n}{n}\right)^{2\beta_2/(2\beta_2+1)}. \tag{41}$$

Finally, note that

$$\mathrm{var}(\hat{T}_{n,3}) = \frac{1}{2\ln n}\left(\frac{\ln n}{n}\right)^{2\beta_2/(2\beta_2+1)}.$$

Let

$$\gamma_n = (2M + 4)\left(\frac{\ln n}{n}\right)^{\beta_2/(2\beta_2+1)}.$$

Then as a centre of the confidence interval take

$$\hat{T}_n = \hat{T}_{n,1}\mathbf{1}(|\hat{T}_{n,1} - \hat{T}_{n,3}| \leqslant \gamma_n) + \hat{T}_{n,2}\mathbf{1}(|\hat{T}_{n,1} - \hat{T}_{n,3}| > \gamma_n). \tag{42}$$

First, note that

$$\mathrm{E}_f(\hat{T}_n - f(0))^2 \leqslant \mathrm{E}_f(\hat{T}_{n,1} - f(0))^2 + \mathrm{E}_f(\hat{T}_{n,2} - f(0))^2\mathbf{1}(|\hat{T}_{n,1} - \hat{T}_{n,3}| > \gamma_n). \tag{43}$$

Note also that, for $f \in \mathcal{F}_1$, $P_f(|\hat{T}_{n,1} - \hat{T}_{n,3}| > \gamma_n) \leqslant n^{-2}$. Hence, if $f \in \mathcal{F}_1$,

$$\mathrm{E}_f(\hat{T}_{n,2} - f(0))^2\mathbf{1}(|\hat{T}_{n,1} - \hat{T}_{n,3}| > \gamma_n) \leqslant (\mathrm{E}_f(\hat{T}_{n,2} - f(0))^4)^{1/2}(P_f(|\hat{T}_{n,1} - \hat{T}_{n,3}| > \gamma_n))^{1/2}$$

$$\leqslant Dn^{-1},$$

for some constant $D > 0$. Hence, for some constant $C > 0$,

$$\sup_{f \in \mathcal{F}_1} \mathrm{E}_f(\hat{T}_n - f(0))^2 \leqslant Cn^{-2\beta_1/(2\beta_1+1)} \tag{44}$$

and it follows that $\hat{T}_n$ is minimax mean squared error rate-optimal over $\mathcal{F}_1$. Similar calculations show that

$$\sup_{f \in \mathcal{F}_2} \mathrm{E}_f(\hat{T}_n - f(0))^2 \leqslant C\left(\frac{\ln n}{n}\right)^{2\beta_2/(2\beta_2+1)}. \tag{45}$$

Let $C(\alpha) = M + \sqrt{2}\, z_{\alpha/2}$. Then it is easy to see that the interval $\hat{T}_{n,2} \pm C(\alpha)n^{-\beta_2/(2\beta_2+1)}$ has coverage probability of at least $1 - \alpha$ over $\mathcal{F}_2$. Set the confidence interval

$$CI_n \equiv \hat{T}_n \pm (|\hat{T}_{n,2} - \hat{T}_{n,1}|\mathbf{1}(|\hat{T}_{n,1} - \hat{T}_{n,3}| \leqslant \gamma_n) + C(\alpha)n^{-\beta_2/(2\beta_2+1)}). \tag{46}$$

Since CI_n always contains the interval $\hat{T}_{n,2} \pm C(\alpha)n^{-\beta_2/(2\beta_2+1)}$ it follows that it also has at least $1 - \alpha$ coverage probability over $\mathcal{F}_2$.

Note that the expected length of the confidence interval CI_n depends only on $E_f(|\hat{T}_{n,2} - \hat{T}_{n,1}|1(|\hat{T}_{n,1} - \hat{T}_{n,3}| \leq \gamma_n))$. For $f \in \mathcal{F}_1$, a simple calculation yields

$$E_f L(CI_n) \leq 2C(\alpha)n^{-\beta_2/(2\beta_2+1)} + 2E_f|\hat{T}_{n,2} - \hat{T}_{n,1}| \tag{47}$$

and, hence,

$$E_f L(CI_n) \leq 2C(\alpha)n^{-\beta_2/(2\beta_2+1)} + 2(E_f(\hat{T}_{n,2} - f(0))^2)^{1/2} + 2(E_f(\hat{T}_{n,1} - f(0))^2)^{1/2}, \tag{48}$$

and it follows that

$$\sup_{f \in \mathcal{F}_1} E_f L(CI_n) \leq 2C(\alpha)n^{-\beta_2/(2\beta_2+1)} + Dn^{-\beta_2/(2\beta_2+1)}, \tag{49}$$

for some $D > 0$. Now suppose that $f \in \mathcal{F}_2$. Then

$$E_f L(CI_n) \leq 2C(\alpha)n^{-\beta_2/(2\beta_2+1)} + 2E_f(|\hat{T}_{n,2} - \hat{T}_{n,1}|1(|\hat{T}_{n,1} - \hat{T}_{n,3}| \leq \gamma_n)). \tag{50}$$

Since

$$E_f(|\hat{T}_{n,2} - \hat{T}_{n,1}|1(|\hat{T}_{n,1} - \hat{T}_{n,3}| \leq \gamma_n)) \leq E_f|\hat{T}_{n,2} - \hat{T}_{n,3}| + E_f(|\hat{T}_{n,3} - \hat{T}_{n,1}|1(|\hat{T}_{n,1} - \hat{T}_{n,3}| \leq \gamma_n)),$$

it follows that

$$\sup_{f \in \mathcal{F}_2} E_f L(CI_n) \leq 2C(\alpha)n^{-\beta_2/(2\beta_2+1)} + 2\gamma_n + 2E_f|\hat{T}_{n,2} - \hat{T}_{n,3}| \leq (2C(\alpha) + D)\left(\frac{\ln n}{n}\right)^{\beta_2/(2\beta_2+1)}, \tag{51}$$

for another constant $D > 0$.

Equations (44) and (45) show that the centre of the confidence interval CI_n is a mean squared error rate adaptive estimator which attains the minimax mean squared error rate of convergence on $\mathcal{F}_1$ and only pays the minimal necessary logarithmic penalty over $\mathcal{F}_2$. Moreover, (49) and (51) show that the confidence interval CI_n has rate-optimal maximum expected length over $\mathcal{F}_1$ for all confidence intervals which have coverage probability of at least $1 - \alpha$ over $\mathcal{F}_2$, and, subject to the mean squared error constraint on the centre, it also minimizes the expected length over $\mathcal{F}_2$.

3.3. Mean squared error for the center of confidence intervals

Theorem 2 also yields an interesting result on the performance, in terms of mean squared error, of the centre of any $(1 - \alpha)$-level confidence interval with rate-optimal expected length. The result below shows the interplay between the mean squared error property of the centre and the maximum expected length of the confidence interval.

Corollary 5. *Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be convex parameter spaces with $\mathcal{F}_1 \cap \mathcal{F}_2 \neq \varnothing$ and let T be a linear functional. Suppose that Condition A holds. Let $CI = [\hat{T} - h_n, \hat{T} + h_n]$ be a $(1 - \alpha)$-level confidence interval over $\mathcal{F}_2$ which has rate-optimal expected length over $\mathcal{F}_2$, that is,*

$$\sup_{f \in \mathcal{F}_2} \mathrm{E}_f L(CI) \leqslant C\omega\left(\frac{1}{\sqrt{n}}, \mathcal{F}_2\right), \tag{52}$$

for some constant $C > 0$. Then the mean squared error of the centre of the confidence interval, $\hat{T}$, must satisfy

$$\sup_{f \in \mathcal{F}_1} \mathrm{E}_f(\hat{T} - Tf)^2 \geqslant \lambda\omega_+^2\left(\frac{1}{\sqrt{n}}, \mathcal{F}_1, \mathcal{F}_2\right), \tag{53}$$

for some constant $\lambda > 0$.

Corollary 5 has an interesting interpretation. Let $\mathcal{F}_2$ be a convex parameter space and let $CI = [\hat{T} - h_n, \hat{T} + h_n]$ be a confidence interval which has coverage probability of at least $1 - \alpha$ over $\mathcal{F}_2$ and which has minimax rate-optimal expected length over $\mathcal{F}_2$, namely that it satisfies (52). Let $\mathcal{F}_1$ be any convex subset of $\mathcal{F}_2$ for which Condition A holds. Then the mean squared error of $\hat{T}$ must satisfy

$$\sup_{f \in \mathcal{F}_1} \mathrm{E}_f(\hat{T} - Tf)^2 \geqslant \lambda\omega^2\left(\frac{1}{\sqrt{n}}, \mathcal{F}_2\right).$$

In particular, if Condition A holds for $\mathcal{F}_1 = \{f\}$, where $f \in \mathcal{F}_2$, then

$$\mathrm{E}_f(\hat{T} - Tf)^2 \geqslant \lambda\omega^2\left(\frac{1}{\sqrt{n}}, \mathcal{F}_2\right).$$

Example. Once again consider the white noise model (1) and the problem of estimating $f(0)$, and let $\mathcal{F}_2 = F(\beta, M)$ with $0 < \beta < 1$. For any $f \in F(\beta', M')$ where either $\beta < \beta' \leqslant 1$ or $M' < M$, Condition A always holds for $\mathcal{F}_1 = \{f\}$. In this case if $CI = [\hat{T} - h_n, \hat{T} + h_n]$ is a confidence interval with coverage probability of at least $1 - \alpha$ and minimax rate-optimal expected length over $\mathcal{F}_2$, then

$$\mathrm{E}_f(\hat{T} - Tf)^2 \geqslant \lambda\omega^2\left(\frac{1}{\sqrt{n}}, \mathcal{F}_2\right) \geqslant Cn^{-2\beta/(1+2\beta)},$$

for some $C > 0$. Hence $\hat{T}$ must have the same rate of convergence $n^{-2\beta/(1+2\beta)}$ under mean squared error at every point in the interior of $F(\beta, M)$.

4. Proofs

The proof of Theorem 1 makes use of the following general constrained error bound. Let X be a random variable having distribution P_θ with density f_θ with respect to a measure λ. The parameter $\theta \in \mathbb{R}$ takes on two possible values, θ_1 or θ_2. We wish to estimate θ based on X. Denote by $r(x) = f_{\theta_2}(x)/f_{\theta_1}(x)$ the ratio of the two density functions. ($r(x) = \infty$ for some x is possible, with the obvious interpretation $r(x)f_{\theta_1}(x) = f_{\theta_2}(x)$.) Set

$$\Delta = |\theta_2 - \theta_1| \tag{54}$$

and, as in Brown and Low (1996), let

$$I = I(\theta_1, \theta_2) = (E_{\theta_1}(r^2(X)))^{1/2}. \tag{55}$$

The following lemma gives a lower bound on the probabilistic error of any estimator δ under θ_2 subject to an upper bound on the mean squared error at θ_1.

Lemma 1. *If* $E_{\theta_1}(\delta - \theta_1)^2 \leq \epsilon^2$, *then for any* $0 < \lambda < 1$,

$$P_{\theta_2}(\delta - \theta_2 \leq -\lambda\Delta) \geq 1 - \frac{\epsilon I}{(1 - \lambda)\Delta}, \qquad \text{if } \theta_2 > \theta_1, \tag{56}$$

$$P_{\theta_2}(\delta - \theta_2 \geq \lambda\Delta) \geq 1 - \frac{\epsilon I}{(1 - \lambda)\Delta}, \qquad \text{if } \theta_2 < \theta_1 \tag{57}$$

and, consequently,

$$P_{\theta_2}(|\delta - \theta_2| \geq \lambda\Delta) \geq 1 - \frac{\epsilon I}{(1 - \lambda)\Delta}.$$

Proof. Without loss of generality, assume that $\theta_2 > \theta_1$ and that the estimator δ satisfies $\delta(X) \in [\theta_1, \theta_2]$. If this condition is not satisfied, replace δ by

$$\delta' = \begin{cases} \theta_1, & \text{if } \delta < \theta_1, \\ \delta & \text{if } \theta_1 \leq \delta \leq \theta_2, \\ \theta_2, & \text{if } \delta > \theta_2, \end{cases}$$

where δ' satisfies $E_{\theta_1}(\delta' - \theta_1)^2 \leq E_{\theta_1}(\delta - \theta_1)^2 \leq \epsilon^2$ and $P_{\theta_2}(\delta' - \theta_2 \leq -\lambda\Delta) = P_{\theta_2}(\delta - \theta_2 \leq -\lambda\Delta)$.

The Cauchy–Schwarz inequality now yields

$$E_{\theta_2}\delta - \theta_1 = E_{\theta_1}[(\delta(X) - \theta_1)r(X)] \leq \epsilon I.$$

On the other hand, since $\delta(X) \in [\theta_1, \theta_2]$,

$$E_{\theta_2}\delta - \theta_1 = E_{\theta_2}\delta 1(\delta - \theta_2 > -\lambda\Delta) + E_{\theta_2}\delta 1(\delta - \theta_2 \leq -\lambda\Delta) - \theta_1$$

$$\geq (\theta_2 - \lambda\Delta)P_{\theta_2}(\delta - \theta_2 > -\lambda\Delta) + \theta_1 P_{\theta_2}(\delta - \theta_2 \leq \lambda\Delta) - \theta_1$$

$$= (1 - \lambda)\Delta P_{\theta_2}(\delta - \theta_2 > -\lambda\Delta).$$

Therefore,

$$P_{\theta_2}(\delta - \theta_2 > -\lambda\Delta) \leq \frac{\epsilon I}{(1 - \lambda)\Delta},$$

and so

$$P_{\theta_2}(\delta - \theta_2 \leq -\lambda\Delta) \geq 1 - \frac{\epsilon I}{(1 - \lambda)\Delta}. \qquad \square$$

Proof of Theorem 1. We shall only consider the case where $\mathcal{F}_1$ and $\mathcal{F}_2$ are closed and norm bounded. The general case is proved by taking limits of this case.

For $0 < \rho \leqslant 1$, choose $f_{1,n} \in \mathcal{F}_1$ and $f_{2,n} \in \mathcal{F}_2$ such that

$$\|f_{1,n} - f_{2,n}\|_2 \leqslant \sqrt{\frac{\rho \ln \gamma_n^2}{n}}$$

and such that the modulus is attained at $\{f_{1,n}, f_{2,n}\}$:

$$|Tf_{2,n} - Tf_{1,n}| = \omega_+\left(\sqrt{\frac{\rho \ln \gamma_n^2}{n}}, \mathcal{F}_1, \mathcal{F}_2\right).$$

Let $\theta_1 = Tf_{1,n}$, $\theta_2 = Tf_{2,n}$, and let $\beta_n = n\|f_{1,n} - f_{2,n}\|_2^2$. Then $\beta_n \leqslant \rho \ln \gamma_n^2$.

Denote by $P_{i,n}$ the probability measure associated with the white noise process

$$\mathrm{d}Y(t) = f_{i,n}(t)\mathrm{d}t + \frac{1}{\sqrt{n}}\mathrm{d}W(t), \quad -\frac{1}{2} \leqslant t \leqslant \frac{1}{2}, \quad i = 1, 2.$$

Then a sufficient statistic for the family of measures $\{P_{i,n} : i = 1, 2\}$ is given by $S_n = \ln(\mathrm{d}P_{2,n}/\mathrm{d}P_{1,n})$ with

$$S_n \sim \begin{cases} N\left(-\dfrac{\beta_n}{2}, \beta_n\right) & \text{under } P_{1,n}, \\[2ex] N\left(\dfrac{\beta_n}{2}, \beta_n\right) & \text{under } P_{2,n}. \end{cases}$$

Denote by $\theta_1 = Tf_{1,n}$, $\theta_2 = Tf_{2,n}$, and s_{θ_i} the density of S_n under $P_{i,n}$ $(i = 1, 2)$. Then

$$I(\theta_1, \theta_2) = \left(\int \frac{s_{\theta_2}^2(x)}{s_{\theta_1}(x)}\,\mathrm{d}x\right)^{1/2} = \mathrm{e}^{\beta_n/2} \leqslant \gamma_n^\rho.$$

Applying Lemma 1 with $\Delta = |\theta_1 - \theta_2| = \omega_+(\sqrt{\rho(\ln \gamma_n^2)/n}, \mathcal{F}_1, \mathcal{F}_2)$, $I(\theta_1, \theta_2) \leqslant \gamma_n^\rho$, and $\epsilon \leqslant \gamma_n^{-1}\omega_+(1/\sqrt{n}, \mathcal{F}_1, \mathcal{F}_2)$, we have

$$P_{f_{2,n}}\left(|\hat{T} - Tf_{2,n}| \geqslant \lambda\omega_+\left(\sqrt{\frac{\rho\ln \gamma_n^2}{n}}, \mathcal{F}_1, \mathcal{F}_2\right)\right) \geqslant 1 - \frac{\omega_+(1/\sqrt{n}, \mathcal{F}_1, \mathcal{F}_2)}{(1-\lambda)\gamma_n^{1-\rho}\omega_+(\sqrt{\rho(\ln \gamma_n^2/n}, \mathcal{F}_1, \mathcal{F}_2)}.$$

$\square$

Proof of Corollary 1. In this case set

$$\gamma_n = \frac{\omega_+(1/\sqrt{n}, \mathcal{F}_1, \mathcal{F}_2)}{C^{1/2}\omega(1/\sqrt{n}, \mathcal{F}_1)} = C^{-1/2}\xi_n$$

and choose $\lambda = \rho = \frac{1}{2}$. Since $\xi_n \to \infty$, for sufficiently large n, $\gamma_n \geqslant \xi_n^{1/2}$. It follows from Theorem 1 that

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}_2} P_f\left(|\hat{T} - Tf| \geq \frac{1}{2}\omega_+\left(\sqrt{\frac{\ln \xi_n^2}{n}}, \mathcal{F}_1, \mathcal{F}_2\right)\right)$$

$$\geq \lim_{n\to\infty} \sup_{f\in\mathcal{F}_2} P_f\left(|\hat{T} - Tf| \geq \frac{1}{2}\omega_+\left(\sqrt{\frac{\frac{1}{2}\ln \gamma_n^2}{n}}, \mathcal{F}_1, \mathcal{F}_2\right)\right)$$

$$\geq \lim_{n\to\infty}\left(1 - \frac{\omega_+(1/\sqrt{n}, \mathcal{F}_1, \mathcal{F}_2)}{\frac{1}{2}\gamma_n^{1/2}\omega_+(\sqrt{\frac{1}{2}(\ln \gamma_n^2)/n}, \mathcal{F}_1, \mathcal{F}_2)}\right)$$

$$= 1. \qquad\qquad\qquad \square$$

Proof of Corollary 2. Suppose that (21) does not hold. Then

$$\sup_{f\in\mathcal{F}_1} E_f(\hat{T} - Tf)^2 \leq B_n^2$$

with $B_n^2/\omega^2(1/\sqrt{n}, \mathcal{F}_2) \to 0$ (at least along a subsequence). It then follows from Condition A that

$$\gamma_n = \frac{\omega_+(1/\sqrt{n}, \mathcal{F}_1, \mathcal{F}_2)}{B_n} \to \infty.$$

An argument similar to that in the proof of Corollary 1 now yields that

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}_2} P_f\left(|\hat{T} - Tf| \geq \frac{1}{2}\omega_+\left(\sqrt{\frac{\ln \gamma_n^2}{n}}, \mathcal{F}_1, \mathcal{F}_2\right)\right) = 1,$$

which contradicts the assumption given in (20) since

$$\omega_+\left(\sqrt{\frac{\ln \gamma_n^2}{n}}, \mathcal{F}_1, \mathcal{F}_2\right) \gg \omega\left(\frac{z_{\alpha/2}}{\sqrt{n}}, \mathcal{F}_2\right) \qquad\qquad \square$$

Proof of Theorem 2. It follows from Theorem 1 that

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}_2} P_f\left(|\hat{T} - Tf| \geq \frac{1}{2}\omega_+\left(\sqrt{\frac{\ln \gamma_n^2}{n}}, \mathcal{F}_1, \mathcal{F}_2\right)\right) = 1,$$

which implies that, for any $\epsilon > 0$, there exists an $N > 0$ such that, for all $n \geq N$,

$$\sup_{f\in\mathcal{F}_2} P_f\left(|\hat{T} - Tf| \geq \frac{1}{2}\omega_+\left(\sqrt{\frac{\ln \gamma_n^2}{n}}, \mathcal{F}_1, \mathcal{F}_2\right)\right) \geq 1 - \frac{\epsilon}{2}.$$

Hence there exists a sequence $f_n \in \mathcal{F}_2$ such that, for $n \geq N$,

$$P_{f_n}\left(|\hat{T} - Tf_n| \geq \frac{1}{2}\omega_+\left(\sqrt{\frac{\ln \gamma_n^2}{n}}, \mathcal{F}_1, \mathcal{F}_2\right)\right) \geq 1 - \epsilon. \tag{58}$$

Now suppose *CI* is a confidence interval centred at $\hat{T}$ with coverage probability of at least $1 - \alpha$ over $\mathcal{F}_2$. Let $CI = [\hat{T} - h_n, \hat{T} + h_n]$. Then, for $n \geq N$,

$$1 - \alpha \leq P_{f_n}(Tf_n \in CI) = P_{f_n}(|\hat{T} - Tf_n| \leq h_n) \tag{59}$$

Equations (58) and (59) together yield

$$P_{f_n}\left(L(CI) \geq \omega_+\left(\sqrt{\frac{\ln \gamma_n^2}{n}}, \mathcal{F}_1, \mathcal{F}_2\right)\right) = P_{f_n}\left(h_n \geq \frac{1}{2}\omega_+\left(\sqrt{\frac{\ln \gamma_n^2}{n}}, \mathcal{F}_1, \mathcal{F}_2\right)\right) \geq 1 - \alpha - \epsilon.$$

Since the constant $\epsilon > 0$ can be chosen arbitrarily small, this implies (36). $\square$

Acknowledgement

References

Brown, L.D. and Low, M.G. (1996) A constrained risk inequality with applications to nonparametric functional estimation. *Ann. Statist.*, **24**, 2524–2435.

Cai, T. and Low, M. (2002) On adaptive estimation of linear functionals. *Ann. Statist*. To appear.

Cai, T. and Low, M. (2003) Adaptation under probabilistic error. *J. Multivariate Anal.* Tentatively accepted.

Cai, T. and Low, M. (2004) An adaptation theory for nonparametric confidence intervals. *Ann. Statist.*, **32**, 1805–1840.

Donoho, D.L. (1994) Statistical estimation and optimal recovery. *Ann. Statist.*, **22**, 238–270.

Donoho, D.L. and Liu, R.G. (1987) Geometrizing rates of convergence I. Technical Report 137, Department of Statistics, University of California, Berkeley.

Donoho, D.L. and Liu, R.G. (1991) Geometrizing rates of convergence III. *Ann. Statist.*, **19**, 668–701.

Efromovich, S. and Low, M.G. (1994) Adaptive estimates of linear functionals. *Probab. Theory Related Fields*, **98**, 261–275.

Efromovich, S.Y. and Pinsker, M.S. (1984) An adaptive algorithm of nonparametric filtering. *Autom. Remote Control*, **11**, 58–65.

Farrell, R.H. (1972) On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. *Ann. Math. Statist.*, **43**, 170–180.

Ibragimov, I.A. and Hasminskii, R.Z. (1984) Nonparametric estimation of the values of a linear functional in Gaussian white noise. *Theory Probab. Appl.*, **31**, 391–406.

Lepski, O.V. (1990) On a problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.*, **35**, 454–466.

Lepski, O.V. and Spokoiny, V.G. (1997) Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.*, **25**, 2512–2546.

Weiss, L. and Wolfowitz, J. (1967) Estimation of a density at a point. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **7**, 327–335.