The Annals of Statistics
1995, Vol. 23, No. 4, 1376–1392

SEQUENTIAL NONPARAMETRIC ESTIMATION WITH ASSIGNED RISK[1]

BY SAM EFROMOVICH

University of New Mexico

The problem is to estimate sequentially a nonparametric function known to belong to an α-th-order Sobolev subspace ($\alpha > \frac{1}{2}$) with a minimax mean stopping time subject to an assigned maximum mean integrated squared error. For the case of a given α there exists a sharp estimator which has a minimal constant and a rate of minimax mean stopping time increasing as the assigned risk decreases. The situation changes drastically if α is unknown: a necessary and sufficient condition for sharp estimation is that $\gamma < \alpha \le 2\gamma$ for some given $\gamma \ge \frac{1}{2}$.

1. Introduction. Substantial research has been devoted to sharp estimation of a function known to belong to a Sobolev subspace of order α when the sample size n is fixed [see the review in Golubev and Nussbaum (1990)]. In particular, for statistical models of filtering, density estimation and nonparametric regression it is shown that

(1.1)
$$\inf \sup_{f \in \mathscr{F}(\alpha, Q)} E_f \left\{ \int_0^1 \left(\hat{f}_n(x) - f(x) \right)^2 dx \right\} = P n^{-2\alpha/(2\alpha+1)} (1 + o_n(1)),$$

where the infimum is over all possible estimators based on both data and the parameters α and Q. Hereafter, $o_n(1) \to 0$ as $n \to \infty$ and $\mathscr{F}(\alpha, Q)$ is the α-th-order Sobolev subspace of periodic functions supported on $[0, 1]$,

$$\mathscr{F}(\alpha, Q) = \left\{ f\colon f(x) = \sum_{i=0}^{\infty} \theta_i \varphi_i(x);\ \theta_0^2 + \sum_{j=1}^{\infty} \left[1 + (2\pi j)^{2\alpha}\right] \left[\theta_{2j-1}^2 + \theta_{2j}^2\right] \le Q \right\}, \tag{1.2}$$

where

$$\left\{ \varphi_0(x) = 1,\ \varphi_{2j-1}(x) = \sqrt{2}\sin(2\pi jx) \text{ and } \varphi_{2j}(x) = \sqrt{2}\cos(2\pi jx),\ j = 1, 2, \dots \right\}$$

is the classical Fourier trigonometric basis on $[0, 1]$, $\langle f, g \rangle = \int_0^1 f(x) g(x)\, dx$

Received February 1992; revised December 1994.
[1]Supported in part by NSF Grant DMS-91-23956 and Sandia National Laboratories Grant AE-1679.
AMS 1991 *subject classifications*. Primary 62G05, 62G07; secondary 62E20, 62F12, 62J02.
Key words and phrases. Sequential estimation, minimax, stopping time, curve fitting.

is the inner product in $L_2(0,1)$, $\theta_j = \langle f, \varphi_j \rangle$,

$$P = Q^{1/(2\alpha+1)} \left(\frac{2\alpha}{2\pi(\alpha+1)} \right)^{2\alpha/(2\alpha+1)} (2\alpha+1)^{1/(2\alpha+1)}$$

is Pinsker's constant [see Pinsker (1980)].

In addition, if $\alpha > \frac{1}{2}$, then there exists an adaptive estimate based only on data (but not on α or Q) such that its mean integrated squared error (MISE) does not exceed the right-hand side of (1.1) for all $f \in \mathscr{F}(\alpha, Q)$ [see Efromovich and Pinsker (1984)]. Moreover, (1.1) holds for all sequential estimators $\mathscr{E}(\{\hat{f}_m, m = 1, 2, \dots,\}, \tau)$ with restricted moment of a stopping time τ, that is, for sequential estimators with a stopping time τ subject to restriction $E_f\{(\tau/n)^\beta\} \le 1$, $\beta \ge 1$ [see Efromovich and Pinsker (1989a, b) and Efromovich (1989)].

The main focus of this paper is to solve the inverse problem of minimizing the expected stopping time subject to assigned maximum risk. To shed light on the problem, let us denote the right-hand side of (1.1) by ε. Then the sample size

$$n^*(\alpha, Q, \varepsilon) = \left\lfloor (\varepsilon P^{-1})^{-(2\alpha+1)/2\alpha)} \right\rfloor + 1 \tag{1.3}$$

is asymptotically sufficient for estimating $f \in \mathscr{F}(\alpha, Q)$ with MISE of at most $\varepsilon(1 + o_\varepsilon(1))$. Hereafter $\varepsilon > 0$, $\lfloor x \rfloor$ is the integer part of x and $o_\varepsilon(1) \to 0$ as $\varepsilon \to 0$. A natural question is the possibility of decreasing the average sample size in comparison with $n^*(\alpha, Q, \varepsilon)$ by implementing a sequential approach [see discussion in Prakasa Rao (1983) on a sequential approach in nonparametric curve estimation theory].

The answer is twofold. On the one hand, if α and Q are given, then the sequential approach does not decrease the average sample size and, due to Efromovich and Pinsker (1989a, b) and Efromovich (1989),

$$\inf \sup_{f \in \mathscr{F}(\alpha, Q)} E_f\{\tau\} = n^*(\alpha, Q, \varepsilon)(1 + o_\varepsilon(1)), \tag{1.4}$$

where the infimum is over all possible sequential estimators with the MISE of at most ε.

On the other hand, in contrast to estimating with a minimal risk based on a fixed sample size, there is no sharp adaptive sequential estimate which attains the right-hand side of (1.4) uniformly over all $\alpha > \frac{1}{2}$ and $0 < Q < \infty$. More precisely, let $\alpha > \frac{1}{2}$ be unknown but fixed, and let Q be also unknown and $Q \in [Q_\varepsilon^{-1}, Q_\varepsilon]$, where the function Q_ε increases to infinity as slowly as desired when $\varepsilon \to \infty$ and $Q_\varepsilon > 1$. We shall refer to this setting as *adaptive*. We claim that for the adaptive setting a sharp estimate exists iff $\alpha \in (\gamma, 2\gamma]$ for some given $\gamma \ge \frac{1}{2}$, that is, there is a known constant γ such that

$$\tfrac{1}{2} \le \gamma < \alpha \le 2\gamma. \tag{1.5}$$

In other words, the necessity of (1.5) for sharp estimation means that if $\alpha \in [\gamma, 2\gamma]$ and $Q \in [Q_\varepsilon^{-1}, Q_\varepsilon]$, where $Q_\varepsilon \to \infty$ as slowly as desired when $\varepsilon \to 0$, then there is no sharp adaptive estimate. The sufficiency of (1.5) means that there exists a sharp estimate which is based only on data and γ.

Some remarks on terminology. The phrase "sharp estimation" means that MISE is to be at most $\varepsilon(1 + o_\varepsilon(1))$ and $\sup_{f \in \mathscr{F}(\alpha, Q)} E_f\{\tau\} = n^*(\alpha, Q, \varepsilon)(1 + o_\varepsilon(1))$. The phrase "adaptive estimate" means that the estimate is based only on data and given γ.

To make our setting more specific, from now on only the problem of filtering is considered; that is, we observe a process $Y(t)$ which is defined by the stochastic equation $dY(t) = f(t)\,dt + dW(t)$, $0 \le t < n$, where $W(t)$ is a standard Wiener process and $f(x)$ is an unknown periodic function with the period one from $\mathscr{F}(\alpha, Q)$. We refer to n as the sample size and $Z_j = \{Y(x + j - 1), 0 \le x < 1\}$ as the j-th observation.

We are now in a position to explain the underlying idea of condition (1.5). The necessity of (1.5) follows from the theory of hypotheses testing. Consider the familiar problem of minimax testing of the simple hypothesis of absence of a signal (i.e., $f = f_0 = 0$) versus the alternative composite hypothesis that a signal has at least some minimal power $C_\varepsilon \varepsilon$ and belongs to a subspace $\mathscr{F}(\alpha, Q)$, that is, that $f \in \mathscr{F}(\alpha, Q) \cap \{f\colon \int_0^1 f^2(x)\,dx > C_\varepsilon \varepsilon\}$. The minimax error is defined as $e(n, \alpha, Q) = \inf\max\{E_{f_0}\{\phi_n\}, \sup E_f\{1 - \phi_n\}\}$, where the infimum is over all possible critical functions ϕ_n based on n observations, α and Q; the supremum is over the signals under the alternative hypothesis. Hereafter $C_\varepsilon \to \infty$ as $\varepsilon \to 0$ and the function C_ε can be either known or unknown. A test is called consistent iff the error tends to zero as $n \to \infty$. Ingster (1982, 1988) shows that if C_ε is not given, then boundedness below from zero of the ratio $n/n^*(2\alpha, Q, \varepsilon)$ is a necessary and sufficient condition for consistent testing. A similar result, with some additional factor depending on C_ε for the ratio, holds for the case of known C_ε (see details in Section 3).

This result explains the necessity of (1.5), for if a rate-optimal adaptive estimate exists and (1.5) does not hold, then the adaptive estimate may be employed for constructing a consistent test whose properties contradict the above mentioned criterion of consistent hypotheses testing.

The sufficiency of (1.5) for sharp estimation is much more involved. Here we discuss only one aspect of the problem. It is well known that both the hypotheses-testing problem and sequential estimation with assigned risk are closely related with that of estimating with assigned risk the quadratic functional $I(f) = \int_0^1 f^2(x)\,dx$ of small signals. More precisely, $I(f)$ is to be estimated with an assigned maximum risk $\sup E_f^{1/2}\{(\hat{I}_n - I(f))^2\} \le \varepsilon$, where the supremum is over $f \in \mathscr{F}(\alpha, Q) \cap \{f\colon I(f) \le C_\varepsilon \varepsilon\}$; here the estimate $\hat{I}_n$ is based on n observations, α and Q, and C_ε does not increase too fast as $\varepsilon \to 0$. It follows from Efromovich (1994) that for this problem again the sample size proportional to $n^*(2\alpha, Q, \varepsilon)$ is the smallest when the problem can be solved. This sheds light on the sufficiency of (1.5). However, the interested reader is referred to details in the following sections because the sufficiency is not a plain issue. It is also worthwhile to note that Bickel and Ritov (1988), Donoho and Liu (1991), Fan (1991) and Brown and Low (1992) give an interesting insight into the related problems of functional estimation.

In Section 2 the results are formulated and a sharp adaptive sequential estimate is constructed. Note that only asymptotic approach when $\varepsilon \to 0$ is investigated and therefore estimation with risk of at most $\varepsilon(1 + o_\varepsilon(1))$ is considered. The reader who is interested in a nonasymptotic approach is referred to Efromovich and Pinsker (1989a), where the problem of estimating with risk of at most ε is considered for arbitrary $\varepsilon > 0$. However, the adaptive estimate is suggested only for the case $\gamma < \alpha \leq \frac{5}{4}\gamma$. It might be possible to implement the machinery developed in that paper to expand the interval to that given in (1.5); we leave this to the interested reader. All proofs are given in Section 3.

Finally, note that the formulated results hold for density and nonparametric regression models as well. The interested reader is referred to Efromovich (1992).

2. Filtering with assigned risk. Let us recall the known results on nonadaptive estimation when $f \in \mathscr{F}(\alpha, Q)$ and both α and Q are given. Using the notation and terminology of the Introduction, set

$$\hat{\theta}_j(n) = n^{-1} \sum_{l=1}^{n} \int_0^1 \varphi_j(x)\, dY(x + l - 1), \tag{2.1}$$

$$\begin{aligned} \hat{f}_n(x, \alpha, Q) = \hat{\theta}_0(n) + \sum_{j=1}^{J} \left(1 - \left(\frac{j}{J}\right)^{\alpha}\right) \\ \times \left[\hat{\theta}_{2j-1}(n)\varphi_{2j-1}(x) + \hat{\theta}_{2j}(n)\varphi_{2j}(x)\right], \end{aligned} \tag{2.2}$$

where $J = J(n, \alpha, Q) = \lfloor [n(2\alpha + 1)(\alpha + 1)Q/(2\alpha(2\pi)^{2\alpha})]^{1/(2\alpha+1)} \rfloor + 1$. Here $\hat{\theta}_j(n)$ is an estimate of the Fourier coefficient θ_j and $\hat{f}_n(x, \alpha, Q)$ is a nonadaptive orthogonal series estimate of the signal.

Efromovich and Pinsker (1989a) show that the following statement holds.

THEOREM 2.1. *Let $F \in \mathscr{F}(\alpha, Q)$ and assume that α and Q are given. Then the sequential estimate $\mathscr{E} = (\{\hat{f}_m(x, \alpha, Q),\ m = 1, 2, \dots\}, \tau)$ with the orthogonal series estimate* (2.2) *and the fixed stopping time $\tau = n^*(\alpha, Q, \varepsilon)$ is sharp.*

The following assertion holds for the adaptive setting.

THEOREM 2.2. *For the adaptive setting, condition* (1.5) *is necessary and sufficient for sharp estimation.*

Theorem 2.2 will be proved in Section 3. The rest of the section is devoted to one particular example of sharp estimation under condition (1.5).

To introduce the sharp estimate, in the first place we recall the known adaptive minimax estimate based on given n observations. This estimate will play a key role in the sharp estimate. Set $d(0) = 0$, $d(k) = d(k - 1) + 2k$, $T(0) = \{0\}$, $T(k) = \{d(k - 1) + 1, d(k - 1) + 2, \dots, d(k)\}$, $S = \lfloor n^{1/4} \rfloor$, $\hat{\Theta}(0, n)$

$= \hat{\theta}_0^2(n) - n^{-1}$, $\hat{L}(0,n) = 1$,

$$(2.3)\qquad \hat{\Theta}(k,n) = (2k)^{-1} \sum_{j \in T(k)} \left(\hat{\theta}_j^2(n) - n^{-1}\right),$$

$$(2.4)\qquad \hat{L}(k,n) = \hat{\Theta}(k,n)\left(\hat{\Theta}(k,n) + n^{-1}\right)^{-1} \chi\left(\hat{\Theta}(k,n) - \ln^{-1}(k+3)n^{-1}\right)$$

where $k = 1, 2, \dots, S$ and $\chi(x) = 1$ if $x > 0$ and $\chi(x) = 0$ if $x \le 0$. Efromovich and Pinsker (1984) suggest the following adaptively smoothed orthogonal series estimate:

$$(2.5)\qquad \hat{f}_a(x,n,S) = \sum_{k=0}^{S} \hat{L}(k,n) \sum_{j \in T(k)} \hat{\theta}_j(n)\varphi_j(x),$$

which is asymptotically minimax, that is, its MISE does not exceed the right-hand side of (1.1) for all $f \in \mathscr{F}(\alpha, Q)$.

Moreover, Efromovich and Pinsker (1984) show that the MISE of the adaptive estimate (2.5) is equal to $R_n(1 + o(1))$ whenever $R_n n \to \infty$ as $n \to \infty$, where

$$(2.6)\qquad \begin{aligned} R_n &= n^{-1} \sum_{k=1}^{S'} (2k)\frac{\Theta_k}{\Theta_k + n^{-1}} + \sum_{k=S'+1}^{S''} (2k)\Theta_k, \\ \sup_{f \in \mathscr{F}(\alpha, Q)} R_n &= P n^{-2\alpha/(2\alpha+1)}(1 + o_n(1)), \end{aligned}$$

$\Theta_k = (2k)^{-1}\sum_{j \in T_k}\theta_j^2$. Here $S' = S'_n > (2J)^{1/2} + 1$ and $S'' = S''_n$ are some sequences of natural numbers such that $S' < S''$ and $J = o_n(1)(S'')^2$.

Comparison of (1.1) with equations (2.6) sheds light on the underlying idea of the sharp estimate. First, we sequentially estimate R_τ and stop the procedure as soon as the estimate is less than ε. Then the adaptive estimate (2.5) is implemented.

To simplify the proofs, it is convenient to split observations into two groups. The first group of observations is used for finding a stopping time, and the second one for constructing estimate (2.5). Recall that we observe $Z_l = \{Y(x + l - 1), 0 \le x < 1\}$, $l = 1, 2, \dots,$ and we denote the first group as $Z1 = (Z_1, \dots, Z_\nu)$ and the second as $Z2 = (Z_{\nu+1}, \dots, Z_\tau)$, $\nu < \tau$. It is also convenient to denote the observations within these groups as $\{Z1_k,\ k = 1, 2, \dots\}$ and $\{Z2_k,\ k = 1, 2, \dots\}$, where $Z1_k = Z_k$ and $Z2_k = Z_{\nu+k}$. We also append 1 or 2 to all statistics and estimates based on these groups of observations. For instances, $\hat{\theta}2_j(n)$ means an estimate $\hat{\theta}_j(n)$ which is defined in (2.1) and based on n observations $Z2_1, Z2_2, \dots, Z2_n$.

We need some additional notation. Recall that the constant $\gamma \ge \frac{1}{2}$ is given, and we assume that (1.5) holds. Set $g = |\ln(\varepsilon^{-1})| + 3$, $S(a,\varepsilon) = \lfloor g^{1/4}\varepsilon^{-1/4a} \rfloor + 1$, $n(a,\varepsilon) = \lfloor \varepsilon^{-1-1/2a}/\ln(g) \rfloor + 2$, $m(a,\varepsilon) = \lfloor n(a,\varepsilon)/\ln(g) \rfloor + 1$, $I = i(\varepsilon) = \lfloor \gamma(g^2 - g^{3/2}) \rfloor$, $a(i) = 2\gamma - ig^{-2}$, $i = 0, 1, \dots, I$. Hereafter we always assume that ε is sufficiently small and therefore all these functions are well defined.

Notice that $n(a(i+1), \varepsilon)/n(a(i), \varepsilon) = 1 + o_\varepsilon(1)$ and $\gamma + \frac{1}{2}g^{-1/2} < a(I) < \gamma + g^{-1/2}$ for sufficiently small ε. Therefore there exists an integer-valued function $i^*(\varepsilon) \in \{0, 1, \dots, I\}$ such that $n(a(i^*(\varepsilon)), \varepsilon) = n^*(\alpha, Q, \varepsilon)(1 + o_\varepsilon(1))$. Thus, our first step is a sequential estimation of $i^*(\varepsilon)$.

Set

$$\hat{\Lambda}1(k, a) = \hat{\Theta}1(k, m(a, \varepsilon))\left[\hat{\Theta}1(k, m(a, \varepsilon)) + n^{-1}(a, \varepsilon)\right]^{-1} \times \chi\left(\hat{\Theta}1(k, m(a, \varepsilon)) - \ln^{-1}(k+3)n^{-1}(a, \varepsilon)\right),$$

and set

$$\hat{R}1(a, \varepsilon) = n^{-1}(a, \varepsilon) \sum_{k=1}^{S(a, \varepsilon)} (2k)\hat{\Lambda}1(k, a) + \sum_{k=S(a, \varepsilon)+1}^{S(a(I), \varepsilon)} (2k)\hat{\Theta}1(k, m(a, \varepsilon))$$

if $a > a(I)$ and $\hat{R}1(a, \varepsilon) = \varepsilon$ otherwise. Note that statistic $\hat{R}1$ obviously mimics R_n defined in (2.6). A sequential estimation is implemented via comparison of $\hat{R}1(a(i), \varepsilon)$ with ε. Recall that $\hat{R}1(a(i), \varepsilon)$ is based on $m(a(i), \varepsilon)$ observations, where $m(a(i), \varepsilon)$ increases as i increases. Thus, we can define the following sequential procedure of estimating $i^*(\varepsilon)$:

$$\hat{i} = \min\{i : \hat{R}1(a(i), \varepsilon) \le \varepsilon, i = 0, 1, \dots, I\} \tag{2.7}$$

with a stopping time $\nu = m(a(\hat{i}), \varepsilon)$, where ν takes at most $I + 1$ possible values. Notice that the stopping time ν is a classical example of the hitting time of a set; ν defines the sample size of the first group $Z1$ of observations.

We are now in a position to define a stopping time for the sharp estimate as $\tau = m(a(\hat{i}), \varepsilon) + n(a(\hat{i}), \varepsilon)$, where the additional $n(a(\hat{i}), \varepsilon)$ observations $Z2$ will be used for constructing the adaptive estimate (2.5). It is convenient to denote $\hat{a}1 = a(\hat{i})$ because this stresses the fact that the stopping time τ is based only on the first $m(a(\hat{i}), \varepsilon)$ observations, that is, τ is a statistic which is based only on $Z1$. Finally, the estimate (2.5), based only on additional $n(\hat{a}1, \varepsilon)$ observations Z2, is implemented. Thus, one can write down the sharp adaptive estimate as

$$\mathscr{E}_a = \left(\hat{f}_a 2(x, n(\hat{a}1, \varepsilon), S(\hat{a}1, \varepsilon)); \tau\right), \qquad \tau = m(\hat{a}1, \varepsilon) + n(\hat{a}1, \varepsilon). \tag{2.8}$$

This estimate has the following properties:

$$\sup_{f \in \mathscr{F}(\alpha, Q)} E_f\left\{\int_0^1 \left(\hat{f}_a 2(x, n(\hat{a}1, \varepsilon), S(\hat{a}1, \varepsilon)) - f(x)\right)^2 dx\right\} = \varepsilon(1 + o_\varepsilon(1)); \tag{2.9}$$

$$\sup_{f \in \mathscr{F}(\alpha, Q)} E_f\{\tau\} = n^*(\alpha, Q, \varepsilon)(1 + o_\varepsilon(1)); \tag{2.10}$$

and (2.9) and (2.10) hold uniformly over $\alpha \in [\gamma + Q_\varepsilon^{-1/2}, 2\gamma]$ and $Q \in [Q_\varepsilon^{-1}, Q_\varepsilon]$, where here $Q_\varepsilon = \ln(3 + |\ln(\varepsilon^{-1})|)$.

Notice that the estimate (2.8) is relatively simple because the sequential procedure (2.7) is based on at most I trials and then the familiar nonsequential adaptive estimate (2.5) is implemented.

3. Proofs.

PROOF OF THEOREM 2.2. To prove the necessity of (1.5) for sharp estimation suppose that this assertion is wrong, that is, there exists an adaptive sequential estimate $\mathscr{E} = (\{\tilde{f}_m, m = 1, 2, \dots\}, \tau')$ such that, for some increasing function $Q_\varepsilon \to \infty$ as $\varepsilon \to \infty$, $Q_\varepsilon > 1$, the following holds:

$$(3.1)\qquad \begin{aligned} \sup E_f\left\{\int_0^1 \left[\tilde{f}_{\tau'}(x) - f(x)\right]^2 dx\right\} &\le \varepsilon(1 + o_\varepsilon(1)), \\ \sup E_f\{\tau'/n^*(\alpha, Q, \varepsilon)\} &\le 1 + o_\varepsilon(1), \end{aligned}$$

where the supremum is taken over $f \in \mathscr{F}(\alpha, Q)$ and then over $\alpha \in [\gamma, 2\gamma]$ and $Q \in [Q_\varepsilon^{-1}, Q_\varepsilon]$.

Let us show that assumption (3.1) contradicts the criterion of consistent testing mentioned in Section 1. Set $C_\varepsilon = Q_\varepsilon^{1/(4\gamma+1)}$ (recall that C_ε is used in the definition of the alternative hypothesis) and define for the hypotheses testing problem formulated in Section 1 the following critical function:

$$\psi_m = \begin{cases} 0, & \text{if } \tau' \le m \text{ and } \int_0^1 \left[\tilde{f}_{\tau'}(x) - f_0(x)\right]^2 dx \le C_\varepsilon^{1/2}\varepsilon, \\ 1, & \text{otherwise,} \end{cases}$$

where $m = n^*(2\gamma, 1, \varepsilon)$. Recall that $f_0(x) \equiv 0$. Note that the test is based on at most m first observations (we can always assume that it is based on exactly m observations) and that it is consistent. Indeed, to estimate the probability of type I error, we write

$$E_{f_0}\{\psi_m\} \le E_{f_0}\left\{\chi(\tau' - m) + \chi\left(\int_0^1 \left[\tilde{f}_\tau(x) - f_0(x)\right]^2 dx - C_\varepsilon^{1/2}\varepsilon\right)\right\}$$

and recall that $\chi(z) = 1$ if $z > 0$ and $\chi(z) = 0$ if $z \le 0$. Using the Chebyshev inequality, we obtain that

$$E_{f_0}\{\psi_m\} \le m^{-1}E_{f_0}\{\tau'\} + \left(C_\varepsilon^{1/2}\varepsilon\right)^{-1} E_{f_0}\left\{\int_0^1 \left[\tilde{f}_{\tau'}(x) - f_0(x)\right]^2 dx\right\}.$$

Note that $f_0 \in \mathscr{F}(2\gamma, Q_\varepsilon^{-1})$ and therefore due to (3.1) we obtain $E_{f_0}\{\tau'\} \le (1 + o_\varepsilon(1))n^*(2\gamma, Q_\varepsilon^{-1}, \varepsilon) = o_\varepsilon(1)m$. At the same time, (3.1) and the Chebyshev inequality yield that the second term in the right-hand side of the last line is asymptotically not greater than $C_\varepsilon^{-1/2}$. Thus, the probability of committing type I error tends to zero as $\varepsilon \to 0$.

To estimate the probability of type II error, we note that, under the alternative hypothesis and for sufficiently small ε, the probability of the event $\{\int_0^1[\tilde{f}_{\tau'}(x) - f_0(x)]^2\, dx \le C_\varepsilon^{1/2}\varepsilon\}$ is not greater than the probability of the event $\{\int_0^1[\tilde{f}_{\tau'}(x) - f(x)]^2\, dx \ge \frac{1}{2}C_\varepsilon \varepsilon\}$. Therefore, (3.1) and the Chebyshev inequality yield that the probability of type II error satisfies

$$\sup E_f\{1-\psi_m\} \le \sup \frac{E_f\left\{\int_0^1 \left(\tilde{f}_{\tau'}(x) - f(x)\right)^2 dx\right\}}{(1/2)C_\varepsilon \varepsilon} = o_\varepsilon(1),$$

where the supremum is over $f \in F_\varepsilon = \mathscr{F}(\gamma, Q_\varepsilon) \cap \{f\colon \int_0^1 f^2(x)\,dx \ge C_\varepsilon \varepsilon\}$.

Thus, we have shown that the suggested test is consistent under assumption (3.1) and that it is based only on $m = n^*(2\gamma, 1, \varepsilon)$ observations. The latter contradicts Ingster (1982, 1988), where it is shown that the necessary sample size for consistent testing of such hypotheses increases faster than $Q_\varepsilon^{1/4\gamma} m$ as $\varepsilon \to \infty$. The contradiction proves the necessity of (1.5).

To show the sufficiency of (1.5) for sharp estimation we prove that the sequential estimate (2.8) is sharp whenever (1.5) holds. First we prove (2.9), that is, that MISE of the estimate is at most $\varepsilon(1 + o_\varepsilon(1))$. Note that (2.9) [as well as (2.10)] is to be verified for unknown but fixed $\alpha \in (\gamma, 2\gamma]$ and $Q \in [Q_\varepsilon^{-1}, Q_\varepsilon]$, where $Q_\varepsilon \to \infty$ arbitrarily slowly as $\varepsilon \to 0$. Thus, it suffices to show that all the following relations are valid uniformly over $f \in \mathscr{F}(\alpha, Q)$, where $\alpha \in [\gamma + |\ln(g)|^{-1/2}, 2\gamma]$ and $Q \in [(1 + |\ln(g)|)^{-1}, 1 + |\ln(g)|]$; in this case we say that an assertion is valid uniformly.

To simplify the following formulas, we use notation $\hat{a} = \hat{a}1$, $\hat{n} = n(\hat{a}1, \varepsilon)$, $\hat{m} = m(\hat{a}1, \varepsilon)$, $\hat{S} = S(\hat{a}1, \varepsilon)$, $n^* = n^*(\alpha, Q, \varepsilon)$ and $K = S(a(I), \varepsilon)$; here $\hat{a}$, $\hat{n}$, $\hat{m}$ and $\hat{S}$ are based only on $Z1$. It is also assumed that $o_\varepsilon(1)$ do not depend on f, α or Q. Statistics $Z1$ and $Z2$ are independent and therefore it follows from Efromovich and Pinsker (1984) that

$$\begin{aligned}
&\sup_{f \in \mathscr{F}(\alpha, Q)} E_f\left\{\int_0^1 \left(\hat{f}_a 2(x, \hat{n}, \hat{S}) - f(x)\right)^2 dx\right\} \\
&\quad < (1 + o_\varepsilon(1)) \\
&\qquad \times \sup_{f \in \mathscr{F}(\alpha, Q)} E_f\left\{\hat{n}^{-1} \sum_{k=1}^{\hat{S}} (2k)\Theta_k\left(\Theta_k + \hat{n}^{-1}\right)^{-1} + \sum_{k=\hat{S}}^{K} (2k)\Theta_k\right\} \\
&\qquad + \sup_{f \in \mathscr{F}(\alpha, Q)} \sum_{k > K} (2k)\Theta_k.
\end{aligned} \tag{3.2}$$

Straightforward calculation shows that the second addend in the right-hand side of (3.2) is less than $o_\varepsilon(1)\varepsilon$. Hence, (2.9) will follow from inequality

$$\begin{aligned}
&\sup_{f \in \mathscr{F}(\alpha, Q)} E_f\left\{\hat{n}^{-1} \sum_{k=1}^{\hat{S}} (2k)\left[\Theta_k\left(\Theta_k + \hat{n}^{-1}\right)^{-1}\right] + \sum_{k=\hat{S}+1}^{K} (2k)\Theta_k\right\} \\
&\quad \le (1 + o_\varepsilon(1))\varepsilon.
\end{aligned} \tag{3.3}$$

Due to (2.7) and inequality $\hat{S} > N(n^*, \alpha, Q)$, which is valid whenever $\hat{n} \ge n^*$, we can restrict our attention to the case

$$\hat{n} < n^* \quad \text{and} \quad \hat{a} > a(I). \tag{3.4}$$

Under (3.4) and due to the definition of $\hat{a}$, the following inequality holds:

$$\hat{R}1(\hat{a}, \varepsilon) \le \varepsilon, \tag{3.5}$$

and therefore

$$\begin{aligned} &E_f\left\{\hat{n}^{-1} \sum_{k=1}^{\hat{S}} (2k)\Theta_k\left(\Theta_k + \hat{n}^{-1}\right)^{-1} + \sum_{k=\hat{S}+1}^{K} (2k)\Theta_k\right\} \\ &\quad \le E_f\left\{\hat{n}^{-1} \sum_{k=1}^{\hat{S}} (2k)|\Theta_k\left(\Theta_k + \hat{n}^{-1}\right)^{-1} - \hat{\Lambda}1(k, \hat{a})|\right\} \\ &\qquad + E_f\left\{\sum_{k=\hat{S}+1}^{K} (2k)\left(\Theta_k - \hat{\Theta}1(k, \hat{m})\right)\right\} + \varepsilon \\ &\quad \triangleq w_1 + w_2 + \varepsilon. \end{aligned} \tag{3.6}$$

To estimate w_1, we note that

$$\begin{aligned} w_1 < E_f\Bigg\{&\hat{n}^{-1} \sum_{k=1}^{\hat{S}} (2k)|\Theta_k\left(\Theta_k + \hat{n}\right)^{-1} - \hat{\Theta}1(k, \hat{m})\left(\hat{\Theta}1(k, \hat{m}) + \hat{n}^{-1}\right)^{-1}| \\ &\times \chi\left(\hat{\Theta}1(k, \hat{m}) - \ln^{-1}(k+3)\hat{n}^{-1}\right)\Bigg\} \\ &+ E_f\left\{\hat{n}^{-1} \sum_{k=1}^{\hat{S}} (2k)\Theta_k\left(\Theta_k + \hat{n}^{-1}\right)^{-1} \chi\left(\ln^{-1}(k+3)\hat{n}^{-1} - \hat{\Theta}1(k, \hat{m})\right)\right\}. \end{aligned}$$

Line (3.3) in Efromovich (1994) yields that

$$\sup_{f \in \mathscr{F}(\alpha, Q)} E_f\left\{|\hat{\Theta}1(k, n) - \Theta_k|^4\right\} \le 128(2k)^{-2} n^{-2}\left(\Theta_k + n^{-1}\right)^{-2}, \tag{3.7}$$

and we use (3.7) and the Chebyshev inequality to estimate w_1. Write $n_i = n(a(i), \varepsilon)$, $m_i = m(a(i), \varepsilon)$, $S(i) = S(a(i), \varepsilon)$ and generic positive constants as C. We get, similar to (A.9) in Efromovich and Pinsker (1989a), that

$$\begin{aligned} w_1 \le \sum_{i=0}^{I-1} E_f\Bigg\{&n_i^{-1} \sum_{k=1}^{S(i)} (2k)\left|\frac{\Theta_k}{\Theta_k + n_i^{-1}} - \frac{\hat{\Theta}1(k, m_i)}{\hat{\Theta}1(k, m_i) + n_i^{-1}}\right| \\ &\times \chi\left(\hat{\Theta}1(k, m_i) - \ln^{-1}(k+3)n_i^{-1}\right)\Bigg\} \\ &+ E_f\left\{\hat{n}^{-1} \sum_{k=1}^{\hat{S}} (2k)2\ln^{-1}(k+3)\chi\left(2\ln^{-1}(k+3)\hat{n}^{-1} - \Theta_k\right)\right\} \\ &+ \sum_{i=0}^{I-1} E_f\Bigg\{n_i^{-1} \sum_{k=1}^{S(i)} (2k)\Theta_k\left(\Theta_k + n_i^{-1}\right)^{-1} \chi\left(\Theta_k - \hat{\Theta}1(k, m_i) - \frac{1}{2}\Theta_k\right) \\ &\times \chi\left(\Theta_k - 2\ln^{-1}(k+3)n_i^{-1}\right)\Bigg\} \end{aligned}$$

$$< C \sum_{i=0}^{I-1} n_i^{-1} \sum_{k=1}^{S(i)} (2k)(2k)^{-1/2} m_i^{-1/2} \big(\Theta_k + m_i^{-1}\big)^{1/2} \big(\Theta_k + n_i^{-1}\big)^{-1}$$

$$+ E_f \left\{ \hat{n}^{-1} \left[2 \sum_{k=1}^{\lfloor \varepsilon^{-1/10\gamma} \rfloor + 1} (2k) \right] + \left[2 \ln^{-1}(\varepsilon^{-1/10\gamma} + 3) \hat{n}^{-1} \hat{S}^2 \right] \right\}$$

$$+ C \sum_{i=0}^{I-1} n_i^{-1} \sum_{k=1}^{S(i)} (2k) \Theta_k \big[\Theta_k + n_i^{-1}\big]^{-1}$$

$$\times \left[(2k)^{-1/2} m_i^{-1/2} \big(\Theta_k + m_i^{-1}\big)^{1/2} \Theta_k^{-1} \right] \chi\big(\Theta_k - 2 \ln^{-1}(k+3) n_i^{-1}\big).$$

Under the stated assumptions, the inequality $\hat{n}^{-1}\hat{S}^2 < C \ln(g) g^{1/2} \varepsilon$ holds, and therefore straightforward simplifications show that

$$\begin{aligned} w_1 &< C \sum_{i=0}^{I-1} n_i^{-1} \sum_{k=1}^{S(i)} (2k)^{1/2} \ln^2(g) \\ &\quad + C\big[n_0^{-1} \varepsilon^{-1/5\gamma} + \ln(g) g^{-1/2} \varepsilon \big] \\ &< C \ln^2(g) \left[\sum_{i=0}^{I-1} n_i^{-1} S^{3/2}(i) + \varepsilon^{1/20\gamma} \varepsilon + g^{-1/2} \varepsilon \right] = o_\varepsilon(1) \varepsilon. \end{aligned} \tag{3.8}$$

To estimate w_2, we use the following corollary of equality (3.2) in Efromovich (1994):

$$\begin{aligned} &E_f \left\{ \left[\sum_{k=S(i)+1}^{K} (2k) \big(\hat{\Theta}1(k, m_i) - \Theta_k \big) \right]^2 \right\} \\ &\quad < 4 m_i^{-1} \left[\sum_{k=S(i)+1}^{K} (2k) \big(\Theta_k + m_i^{-1}\big) \right]. \end{aligned} \tag{3.9}$$

The Chebyshev inequality, (3.9), inequality $\sum_{k=\hat{S}+1}^{K} (2k) \hat{\Theta}1(k, \hat{m}) \le \varepsilon$, which follows from (3.5), and elementary $K^2 m_0^{-1} < C\varepsilon$ yield that

$$\begin{aligned} w_2 \le E_f \Bigg\{ & \left| \sum_{k=\hat{S}+1}^{K} (2k) \big(\Theta_k - \hat{\Theta}1(k, \hat{m}) \big) \right| \\ & \times \left[\chi\left(\sum_{k=\hat{S}+1}^{K} (2k) \Theta_k - 2\varepsilon \right) \right. \\ & \times \chi\left(\left[\sum_{k=\hat{S}+1}^{K} (2k) \big(\Theta_k - \hat{\Theta}1(k, \hat{m}) \big) \right] - \tfrac{1}{2} \sum_{k=\hat{S}+1}^{K} (2k) \Theta_k \right) \\ & \left. + \chi\left(2\varepsilon - \sum_{k=\hat{S}+1}^{K} (2k) \Theta_k \right) \right] \Bigg\} \end{aligned}$$

$$
\begin{aligned}
&\le C\sum_{i=0}^{I-1} m_i^{-1}\left(\sum_{k=S(i)+1}^{K}(2k)\Theta_k + K^2 m_i^{-1}\right)\left(\sum_{k=S(i)+1}^{K}(2k)\Theta_k\right)^{-1}\\
&\quad\times\chi\left(\sum_{k=S(i)+1}^{K}(2k)\Theta_k - 2\varepsilon\right)\\
&\quad + C\sum_{i=0}^{I-1} m_i^{-1/2}\left(\sum_{k=S(i)+1}^{K}(2k)\Theta_k + K^2 m_i^{-1}\right)^{1/2}\\
&\quad\times\chi\left(2\varepsilon - \sum_{k=S(i)+1}^{K}(2k)\Theta_k\right)\\
&< CIKm_0^{-1}\left(1 + Km_0^{-1}\varepsilon^{-1}\right) + CIm_0^{-1/2}\varepsilon^{1/2}.
\end{aligned}
$$

The second addend in the right-hand side of the last inequality is equal to $o_\varepsilon(1)\varepsilon$; to estimate the first one, we note that $Km_0^{-1}\varepsilon^{-1} = o_\varepsilon(1)$ and that

$$
IKm_0^{-1} < C\varepsilon g^3\varepsilon^{1/4\gamma - 1/[4(2\gamma - Ig^{-2})]} < C\varepsilon g^3\varepsilon^{g^{-1/2}/5\gamma} = o_\varepsilon(1)\varepsilon.
$$

Thus, $w_2 = o_\varepsilon(1)\varepsilon$ and this, together with (3.6) and (3.8), yields (3.3). Assertion (2.9) is proved.

Now we prove (2.10). The proof is based on the following inequality:

$$
\begin{aligned}
&E_f\left\{\left(\frac{\hat{n}}{n^*}\right)\chi(\hat{n} - n^*)\right\}\\
&\qquad\le (1 + o_\varepsilon(1))E_f\left\{\left(\frac{\hat{n}}{n^*}\right)^{1/(2\alpha+1)}\chi(\hat{n} - n^*)\right\} + o_\varepsilon(1),
\end{aligned}
\tag{3.10}
$$

which is to be valid uniformly. Let us show that (3.10) implies (2.10) and then prove (3.10). Equation (2.10) straightforwardly follows from the elementary inequality $(\hat{n}/n^*)\chi(n^* - \hat{n}) \le (\hat{n}/n^*)^{1/(2\alpha+1)}\chi(n^* - \hat{n})$, Jensen's inequality and (3.10). Indeed, we see that, for sufficiently small ε uniformly,

$$
\begin{aligned}
E_f\left\{\frac{\tau}{n^*}\right\} &\le (1 + \ln^{-1}(g))E_f\left\{\frac{\hat{n}}{n^*}\right\}\\
&\le (1 + o_\varepsilon(1))E_f\left\{\left(\frac{\hat{n}}{n^*}\right)^{1/(2\alpha+1)}\right\} + o_\varepsilon(1)\\
&\le (1 + o_\varepsilon(1))\left[E_f\left\{\frac{\hat{n}}{n^*}\right\}\right]^{1/(2\alpha+1)} + o_\varepsilon(1)\\
&\le (1 + o_\varepsilon(1))\left[E_f\left\{\frac{\tau}{n^*}\right\}\right]^{1/(2\alpha+1)} + o_\varepsilon(1).
\end{aligned}
$$

Thus, to complete the proof, we need to verify (3.10). Notice that all statistics considered from now on are based only on $Z1$. Write $\hat{R}(i) =$

$\hat{R}1(a(i), \varepsilon)$ and $R(i) = n_i^{-1}\sum_{k=1}^{S(i)}(2k)\Theta_k(\Theta_k + n_i^{-1})^{-1} + \sum_{S(i)+l}^{K}(2k)\Theta_k$, and note that due to (2.7) and definition of n_i the inequality $\hat{R}(\hat{i} - 1) > \varepsilon$ holds for sufficiently small ε whenever $\hat{n} > n^*$. Using this remark, we write

$$\begin{aligned} E_f\left\{\left(\frac{\hat{n}}{n^*}\right)\chi(\hat{n} - n^*)\right\} &\leq E_f\left\{\left[\hat{n}(n^*\varepsilon)^{-1}\left(R(\hat{i} - 1) + \hat{R}(\hat{i} - 1) - R(\hat{i} - 1)\right)\right]\chi(\hat{n} - n^*)\right\} \\ &= E_f\left\{\hat{n}(n^*\varepsilon)^{-1}R(\hat{i} - 1)\chi(\hat{n} - n^*)\right\} \\ &\quad + E_f\left\{\hat{n}(n^*\varepsilon)^{-1}\left[\hat{n}^{-1}\sum_{k=1}^{S(\hat{i}-1)}(2k)\right.\right. \\ &\qquad \left.\times\left[\hat{\Lambda}1\left(k, a(\hat{i} - 1)\right) - \Theta_k\left(\Theta_k + n_{\hat{i}-1}^{-1}\right)^{-1}\right]\right] \\ &\qquad \left.\times\chi(\hat{n} - n^*)\right\} \\ &\quad + E_f\left\{\hat{n}(n^*\varepsilon)^{-1}\sum_{k=S(\hat{i}-1)+1}^{K}(2k)\left(\hat{\Theta}1(k, m_{\hat{i}-1}) - \Theta_k\right)\chi(\hat{n} - n^*)\right\} \\ &\triangleq V_1 + V_2 + V_3. \end{aligned} \tag{3.11}$$

To estimate V_1, we recall that $\hat{S} = \lfloor g^{1/4}\varepsilon^{-1/4\hat{a}}\rfloor + 1$. Therefore, for $\hat{n} > n^*$ and sufficiently small ε, the inequality $S(\hat{a}, \varepsilon) > g^{1/4}\hat{n}^{1/2(2\alpha+1)}$ holds and, together with (2.7), yields that the inequality $\sup_{f\in\mathscr{F}(\alpha,Q)} R(\hat{i} - 1) \leq (1 + o_\varepsilon(1))P\hat{n}^{-2\alpha/(2\alpha+1)}$ holds uniformly. Now note that, due to (1.3),

$$\begin{aligned} \hat{n}P\hat{n}^{-2\alpha/(2\alpha+1)}(n^*\varepsilon)^{-1} &= P\hat{n}^{1/(2\alpha+1)}(Pn^*)^{-1/(2\alpha+1)}(1 + o_\varepsilon(1)) \\ &= \left(\frac{\hat{n}}{n^*}\right)^{1/(2\alpha+1)}(1 + o_\varepsilon(1)). \end{aligned}$$

Therefore, the following inequality holds uniformly:

$$V_1 \leq (1 + o_\varepsilon(1))E_f\left\{\left(\frac{\hat{n}}{n^*}\right)^{1/(2\alpha+1)}\right\}. \tag{3.12}$$

To estimate V_2 we define $\hat{S}_\alpha = \min\{S_\alpha, S(\hat{i} - 1)\}$, where $S_\alpha = \lfloor g^{1/2}\varepsilon^{-1/4\alpha}\rfloor + 1$. Then

$$\begin{aligned} V_2 = E_f\left\{\hat{n}(n^*\varepsilon)^{-1}\left[\hat{n}^{-1}\sum_{k=1}^{S(\hat{i}-1)}(2k)\left[\hat{\Lambda}1\left(k, a(\hat{i} - 1)\right)\right.\right.\right. \\ \left.\left.\left. -\Theta_k\left(\Theta_k + n_{\hat{i}-1}^{-1}\right)^{-1}\right]\right]\chi(\hat{n} - n^*)\right\} \end{aligned}$$

$$
(3.13)\quad \begin{aligned}
&= E_f\left\{\hat{n}(n^*\varepsilon)^{-1}\left[\hat{n}^{-1}\sum_{k=1}^{\hat{S}_\alpha}(2k)\left[\hat{\Lambda}1(k,a(\hat{\imath}-1))\right.\right.\right.\\
&\qquad\qquad\left.\left.\left.-\Theta_k\left(\Theta_k+n_{\hat{\imath}-1}^{-1}\right)^{-1}\right]\right]\chi(\hat{n}-n^*)\right\}\\
&\quad+E_f\left\{\chi\left(S(\hat{\imath}-1)-S_\alpha\right)\hat{n}(n^*\varepsilon)^{-1}\right.\\
&\qquad\times\left[\hat{n}^{-1}\sum_{k=\hat{S}_\alpha+1}^{S(\hat{\imath}-1)}(2k)\left[\hat{\Lambda}1(k,a(\hat{\imath}-1))\right.\right.\\
&\qquad\qquad\left.\left.\left.-\Theta_k\left(\Theta_k+n_{\hat{\imath}-1}^{-1}\right)^{-1}\right]\right]\chi(\hat{n}-n^*)\right\}\\
&\triangleq V_{21}+V_{22}.
\end{aligned}
$$

Set $\hat{D}(k,\hat{\imath}-1)=|\hat{\Theta}1(k,m_{\hat{\imath}-1})-\Theta_k|(\Theta_k+\hat{n}_{\hat{\imath}-1}^{-1})^{-1}$ and note that the following inequality always holds:

$$
(3.14)\quad |\hat{\Lambda}1\left(k,a(\hat{\imath}-1)\right)-\Theta_k\left(\Theta_k+n_{\hat{\imath}-1}^{-1}\right)^{-1}|\le\min\left\{1,\hat{D}(k,\hat{\imath}-1)\right\}.
$$

There are only two possible cases: either $\hat{D}(k,\hat{\imath}-1)\le(2k)^{-\lambda}$ or $\hat{D}(k,\hat{\imath}-1)>(2k)^{-\lambda}$, where $\lambda=\frac{1}{6}$. Using the obvious inequality $\hat{S}_\alpha\le S_\alpha$, we write

$$
(3.15)\quad \begin{aligned}
V_{21}&\le E_f\left\{\hat{n}(n^*\varepsilon)^{-1}\left[\hat{n}^{-1}\sum_{k=1}^{S_\alpha}\left[(2k)(2k)^{-\lambda}\right]\chi(\hat{n}-n^*)\right]\right\}\\
&\quad+E_f\left\{\hat{n}(n^*\varepsilon)^{-1}\left[\hat{n}^{-1}\sum_{k=1}^{S_\alpha}(2k)^{1+3\lambda}\left(\hat{D}(k,\hat{\imath}-1)\right)^4\right]\right..
\end{aligned}
$$

To estimate the first addend in the right-hand side of (3.15), we recall that, by (1.3), $\varepsilon=P(n^*)^{-2\alpha/(2\alpha+1)}(1+o_\varepsilon(1))$, and therefore straightforward calculation shows that

$$
\begin{aligned}
S_\alpha^{2-\lambda}&<Cg^{(2-\lambda)/2}\varepsilon^{-(2-\lambda)/4\alpha}<Cg(n^*)^{2\alpha(2-\lambda)/(2\alpha+1)(4\alpha)}\\
&=Cg(n^*)^{1/(2\alpha+1)}(n^*)^{-\lambda/2(2\alpha+1)}
\end{aligned}
$$

and, for sufficiently small ε, the following holds uniformly:

$$
E_f\left\{\hat{n}(n^*\varepsilon)^{-1}\left[\hat{n}^{-1}\sum_{k=1}^{S_\alpha}\left[(2k)(2k)^{-\lambda}\right]\chi(\hat{n}-n^*)\right]\right\}\le C(n^*\varepsilon)^{-1}S_\alpha^{2-\lambda}=o_\varepsilon(1).
$$

To estimate the second addend in the right-hand side of (3.15), we implement inequality (3.7) and then see that

$$\begin{aligned} &E_f\left\{\hat{n}(n^*\varepsilon)^{-1}\left[\hat{n}^{-1}\sum_{k=1}^{S_\alpha}(2k)^{1+3\lambda}|\hat{\Theta}1(k,m_{\hat{i}-1})-\Theta_k|^4(\Theta_k+n_{\hat{i}-1}^{-1})^{-4}\right]\right\} \\ &\quad\le C\sum_{i=0}^{I}(n^*\varepsilon)^{-1}\sum_{k=1}^{S_\alpha}(2k)^{-1+3\lambda}(\ln g)^4 \\ &\quad< CI(n^*\varepsilon)^{-1}S_\alpha^{3\lambda}(\ln g)^4 \\ &\quad< Cg^{2+3\lambda}(n^*)^{-1/(2\alpha+1)+3\lambda/4\alpha}=o_\varepsilon(1). \end{aligned}$$

Thus, $V_{21}=o_\varepsilon(1)$. To estimate V_{22}, we consider separately two different cases: $\Theta_k<(\frac{1}{2})\ln^{-1}(K)n_{\hat{i}-1}^{-1}$ and $\Theta_k\ge(\frac{1}{2})\ln^{-1}(K)n_{\hat{i}-1}^{-1}$. For the first case we write

$$\begin{aligned} &\left[\hat{\Lambda}1(k,a(\hat{i}-1))-\Theta_k(\Theta_k+n_{\hat{i}-1}^{-1})^{-1}\right]\chi\left((\tfrac{1}{2})\ln^{-1}(K)n_{\hat{i}-1}^{-1}-\Theta_k\right) \\ &\quad\le\left(\hat{\Theta}1(k,m_{\hat{i}-1})-\Theta_k\right)\left(\Theta_k+n_{\hat{i}-1}^{-1}\right)^{-1} \\ &\quad\quad\times\chi\left(\hat{\Theta}1(k,m_{\hat{i}-1})-\Theta_k-(\tfrac{1}{2})\ln^{-1}(K)n_{\hat{i}-1}^{-1}\right) \\ &\quad\quad\times\chi\left((\tfrac{1}{2})\ln^{-1}(K)n_{\hat{i}-1}^{-1}-\Theta_k)\right). \end{aligned}$$

To consider the second case, we note that for each positive natural l the inequality $\sup_{f\in\mathscr{F}(\alpha,Q)}\sum_{k=1}^{\infty}(2k)\Theta_k<CQl^{-4\alpha}$ holds, and therefore

$$\begin{aligned} &\sup_{f\in\mathscr{F}(\alpha,Q)}\sum_{k=S_\alpha+1}^{\infty}(2k)\chi\left(\Theta_k-(\tfrac{1}{2})\ln^{-1}(K)n_{\hat{i}-1}^{-1}\right) \\ &\quad< CQS_\alpha^{-4\alpha}\log(K)\hat{n} \\ &\quad< C\ln(g)g^{-2\alpha}\ln(K)\varepsilon\hat{n}=o_\varepsilon(1)\varepsilon\hat{n}. \end{aligned}$$

Using the last inequality and applying (3.7), we obtain that

$$\begin{aligned} V_{22}\le E_f\Bigg\{&\chi\left(S(\hat{i}-1)-S_\alpha\right)(n^*\varepsilon)^{-1} \\ &\times\sum_{k=S_\alpha+1}^{S(\hat{i}-1)}(2k)\Bigg[\left(\hat{\Theta}1(k,m_{\hat{i}-1})-\Theta_k\right)^4\chi\left(\left(\frac{1}{2}\right)\ln^{-1}(K)n_{\hat{i}-1}^{-1}-\Theta_k\right) \\ &\qquad\times\left(\left(\Theta_k+n_{\hat{i}-1}^{-1}\right)\left[\left(\frac{1}{2}\right)\ln^{-1}(K)n_{\hat{i}-1}^{-1}\right]^3\right)^{-1} \\ &\qquad+\chi\left(\Theta_k-\left(\frac{1}{2}\right)\ln^{-1}(K)n_{\hat{i}-1}^{-1}\right)\Bigg]\chi(\hat{n}-n^*)\Bigg\} \end{aligned}$$

$$\le C(n^*\varepsilon^{-1})^{-1}\sum_{i=0}^{I}\chi(S(i-1)-S_\alpha)\sum_{k=S_\alpha+1}^{S(i-1)}\ln^3(K)\ln^4(g)(2k)^{-1}$$
$$+ o_\varepsilon(1)E_f\{(n^*\varepsilon)^{-1}\varepsilon\hat{n}\chi(\hat{n}-n^*)\}$$
$$\le CI(n^*)^{-1/(2\alpha+1)}\ln^2(S(I))\ln(g) + o_\varepsilon(1)E_f\left\{\left(\frac{\hat{n}}{n^*}\right)\chi(\hat{n}-n^*)\right\}$$
$$\le o_\varepsilon(1)(1+E_f\left\{\left(\frac{\hat{n}}{n^*}\right)\chi(\hat{n}-n^*)\right\},$$

and therefore, uniformly,

$$V_2 = V_{21}+V_{22} = o_\varepsilon(1)\left(1+E_f\left\{\left(\frac{\hat{n}}{n^*}\right)^{1/(2\alpha+1)}\chi(\hat{n}-n^*)\right\}\right). \quad (3.16)$$

Finally, we estimate V_3. Write $B(k,i)=\sum_{k=S(\hat{i}-1)+1}^{K}(2k)(\hat{\Theta}1(k,m_{\hat{i}-1})-\Theta_k)$ and consider two cases when $B(k,\hat{i})$ is either less than or equal to $\lambda_\varepsilon\varepsilon$ or greater than $\lambda_\varepsilon\varepsilon$; here $\lambda_\varepsilon=\varepsilon^{g^{-1/2}/20\gamma^2}$. Using (3.9) we see that

$$V_3 = E_f\Big\{\hat{n}(n^*\varepsilon)^{-1}B(k,\hat{i})\chi(\hat{n}-n^*)\big[\chi(\lambda_\varepsilon\varepsilon - B(k,\hat{i}))$$
$$+\chi(B(k,\hat{i})-\lambda_\varepsilon\varepsilon)\big]\Big\}$$
$$\le \lambda_\varepsilon E_f\left\{\left(\frac{\hat{n}}{n^*}\right)\chi(\hat{n}-n^*)\right\} + E_f\{\hat{n}(n^*\varepsilon)^{-1}B^2(k,\hat{i})(\lambda_\varepsilon\varepsilon)^{-1}\chi(\hat{n}-n^*)\}$$
$$\le \lambda_\varepsilon E_f\left\{\left(\frac{\hat{n}}{n^*}\right)\chi(\hat{n}-n^*)\right\}$$
$$+(\lambda_\varepsilon\varepsilon)^{-1}\sum_{i=0}^{I}n_i(n^*\varepsilon)^{-1}Cn_i^{-1}\left[\sum_{k=S(i)+1}^{K}(2k)(\Theta_k+n_i^{-1})\right]\chi(n_i-n^*).$$

For $n_i > n^*$, the inequality $\alpha > a(i)$ is valid, and therefore

$$\sup_{f\in\mathscr{F}(\alpha,Q)}\chi(n_i-n^*)\sum_{k=S(i)+1}^{K}(2k)\Theta_k < CQ[S(\alpha,\varepsilon)]^{-4\alpha} < C\ln(g)g^{-\alpha}\varepsilon.$$

From the definition of K we also get

$$\sum_{k=S(0)+1}^{K}(2k) \le CK^2 < Cg^{1/2}\varepsilon^{-1/2a(I)}. \quad (3.17)$$

We have noted in Section 2 that $a(I) > \gamma + (1/2)g^{-1/2}$ for sufficiently small ε. Using this inequality for estimating the right-hand side of (3.17), we obtain that, for sufficiently small ε, the inequality

$$\sum_{k=S(0)+1}^{K}(2k) \le Cg^{1/2}\varepsilon^{-1/(2\gamma+g^{-1/2})} \le Cg^{1/2}\varepsilon^{-1/2\gamma}\varepsilon^{g^{-1/2}/5\gamma^2}$$

holds, and therefore

$$
\begin{aligned}
V_3 &\le \lambda_\varepsilon E_f\left\{\left(\frac{\hat n}{n^*}\right)\chi(\hat n - n^*)\right\}\\
&\quad + C(\lambda_\varepsilon \varepsilon)^{-1} I(n^*\varepsilon)^{-1}\Big[\ln(g) g^{-\alpha}\varepsilon + (n^*)^{-1} g^{1/2}\varepsilon^{-1/2\gamma}\varepsilon^{g^{-1/2}/5\gamma^2}\Big]\\
&\le \lambda_\varepsilon E_f\left\{\left(\frac{\hat n}{n^*}\right)\chi(\hat n - n^*)\right\}\\
&\quad + C\lambda_\varepsilon^{-1}\Big[(n^*\varepsilon)^{-1} g^{3-\alpha} + (n^*\varepsilon)^{-2} g^{5/2}\varepsilon^{-1/(2\gamma)+g^{-1/2}/(5\gamma^2)}\Big].
\end{aligned}
$$

It is easy to verify that $(n^*\varepsilon)^{-2} \le P^{-(2\alpha+1)/\alpha}\varepsilon^{1/\alpha}$, and therefore

$$
(n^*\varepsilon)^{-2} g^{5/2}\varepsilon^{-1/(2\gamma)+g^{-1/2}/(5\gamma^2)} \le C(\ln(g))^2 g^{5/2}\varepsilon^{g^{-1/2}/(5\gamma^2)} < C\varepsilon^{g^{-1/2}/(10\gamma^2)}.
$$

Thus we see that, uniformly,

$$
\begin{aligned}
V_3 &\le \lambda_\varepsilon E_f\left\{\left(\frac{\hat n}{n^*}\right)\chi(\hat n - n^*)\right\} + C\lambda_\varepsilon^{-1}\varepsilon^{g^{-1/2}/(10\gamma^2)}\\
&\le o_\varepsilon(1) E_f\left\{\left(\frac{\hat n}{n^*}\right)\chi(\hat n - n^*)\right\} + o_\varepsilon(1).
\end{aligned}
\tag{3.18}
$$

Substituting (3.12), (3.16) and (3.18) into the right-hand side of (3.11), we get (3.10). Theorem 2.2 is proved. □

Acknowledgments. The authors wishes to thank L. Brown, R. Khasminskii and M. Pinsker for useful discussion and communication. Thanks also go to the referee and an Associate Editor for their constructive suggestions.

REFERENCES

BICKEL, P. J. and RITOV, Y. (1988). Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhyā Ser. A* **50** 381–393.

BROWN, L. D. and LOW, M. G. (1992). Superefficiency and lack of adaptability in nonparametric functional estimation. Technical report, Cornell Univ.

DONOHO, D. and LIU, R. (1991). Geometrizing rates of convergence III. *Ann. Statist.* **19** 668–701.

EFROMOVICH, S. (1989). On sequential nonparametric estimation of a density. *Theory Probab. Appl.* **34** 228–239.

EFROMOVICH, S. (1992). On sequential nonparametric estimation with guaranteed precision (with an application to density estimation, filtering, and nonparametric regression). Technical report, Univ. New Mexico.

EFROMOVICH, S. (1994). On adaptive estimation of nonlinear functionals. *Statist. Probab. Lett.* **19** 57–63.

EFROMOVICH, S. and PINSKER, M. S. (1984). An adaptive algorithm of nonparametric filtering. *Automat. Remote Control* **45** 58–65.

EFROMOVICH, S. and PINSKER, M. S. (1989a). Detecting a signal with an assigned risk. *Automat. Remote Control* **50** 1383–1390.

EFROMOVICH, S. and PINSKER, M. S. (1989b). Sequential estimation of nonparametric regression. In *Abstracts of Fifth International Vilnius Conference on Probability Theory and Mathematical Statistics. Vilnius* **50** 127–128.

FAN, J. (1991). On the estimation of quadratic functionals. *Ann. Statist.* **19** 1273–1294.

GOLUBEV, G. K. and NUSSBAUM, M. (1990). A risk bound in Sobolev regression. *Ann. Statist.* **18** 758–778.

INGSTER, YU. V. (1982). Minimax nonparametric detection of signals in white Gaussian noise. *Problems Inform. Transmission* **18** 130–139.

INGSTER, YU. V. (1988). Minimax testing of a simple hypothesis against a weakly nonparametric set of alternatives. *Theory Probab. Appl.* **33** 552–556.

PINSKER, M. S. (1980). Optimal filtering of square integrable signals in gaussian white noise. *Problems Inform. Transmission* **16** 120–133.

PRAKASA RAO, B. L. S. (1983). *Nonparametric Functional Estimation.* Academic Press, New York.

DEPARTMENT OF MATHEMATICS
AND STATISTICS
UNIVERSITY OF NEW MEXICO
ALBUQUERQUE, NEW MEXICO 87131-1141