Probab. Th. Rel. Fields 87, 253–274 (1990)

Probability Theory and Related Fields

Convergence rates in density estimation for data from infinite-order moving average processes

Peter Hall[1] and Jeffrey D. Hart[2,]*

[1] Department of Statistics, Faculty of Economics and Commerce,
Australian National University, GPO Box 4, Canberra ACT 2601, Australia
[2] Department of Statistics, Texas A & M University, College Station, TX 77843, USA

Received July 27, 1989; in revised form June 18, 1990

Summary. The effect of long-range dependence in nonparametric probability density estimation is investigated under the assumption that the observed data are a sample from a stationary, infinite-order moving average process. It is shown that to first order, the mean integrated squared error (MISE) of a kernel estimator for moving average data may be expanded as the sum of MISE of the kernel estimator for a same-size *random* sample, plus a term proportional to the variance of the moving average sample mean. The latter term does not depend on bandwidth, and so imposes a ceiling on the convergence rate of a kernel estimator regardless of how bandwidth is chosen. This ceiling can be quite significant in the case of long-range dependence. We show that *all* density estimators have the convergence rate ceiling possessed by kernel estimators.

1. Introduction

A tremendous amount of attention has been focused on the problem of nonparametric probability density estimation (see, e.g., [16] and references therein). Most of this attention has been directed to settings where the observations are independent random variables. However, starting with the work of Rosenblatt [14] there has also been considerable interest in estimating the marginal density of a stationary stochastic process. The latter problem is the subject of the current paper.

Suppose that one observes random variables $X_1, \ldots, X_n$ which are identically distributed, but not necessarily independent. Let each X_i have density f, which is to be estimated. By far the most popular estimator of $f(x)$ is the kernel estimator

$$\hat{f}(x)=\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right), \tag{1.1}$$

* The research of Dr. Hart was done while he was visiting the Australian National University, and was supported in part by ONR Contract N00014-85-K-0723

where K is usually taken to be a density function. In the context of dependent data, most of the results on estimating f have been of the following character: if the data satisfy an appropriate mixing condition (i.e., if dependence is sufficiently weak) then, as $n \to \infty$, the L_2 error of $\hat{f}(x)$ is asymptotically the same as when the observations are independent. Examples of this type of result may be found in Rosenblatt [14], Chanda [4], Hart [10], Robinson [13] and Castellana and Leadbetter [3]. A somewhat more pessimistic view on the effect of positive dependence in moderate-sized samples is given in Hart [10].

In contrast with previous work in this area, our primary interest is to describe in a precise way how long-range dependence can affect the convergence rates of kernel density estimators. A popular example of a process with long-range dependence is a stationary time series whose autocorrelation function decays very slowly to 0. In such a series, data which are far apart in time may still be highly correlated. Practical examples of processes exhibiting long-range dependence exist in hydrology [11] and atmosphere physics [8]. It seems impossible to give a concise description of convergence rates in the most general setting. We shall focus on the important class of stationary, infinite-order moving average processes. To be more precise, we shall assume that the data $X_1, \ldots, X_n$ are part of a process $\{X_j\}$ satisfying

$$X_j = \mu + \sum_{k=-\infty}^{\infty} a_k \xi_{k+j}, \qquad -\infty < j < \infty, \tag{1.2}$$

where μ is a constant, the ξ_j's are independent, identically distributed random variables with mean 0 and finite variance, and $\Sigma_k a_k^2 < \infty$. Our results also apply to any infinite-order autoregressive process which can be inverted and put in the form (1.2). (See Priestley [12, pp. 144–145] for invertibility conditions.)

Our main result is given in Sect. 2 and concerns the mean integrated squared error, $\int E(\hat{f}-f)^2$, of $\hat{f}$. Under mild regularity conditions on the characteristic function of ξ_0 and on the sequence $\{a_k\}$, we show that

$$\int E(\hat{f}-f)^2 \sim \int E(\hat{f}^0 - f)^2 + E(\bar{X}-\mu)^2 \int (f')^2, \tag{1.3}$$

where $\bar{X}$ is the usual sample mean and $\hat{f}^0$ is a kernel density estimator based on a *random* sample of size n from density f. This result makes it easy to understand how long-range dependence affects the rate of convergence of $\hat{f}$ to f. It is well known (see, e.g., Priestley [12]) that if the spectrum of the process $\{X_j\}$ is well-defined and finite at 0, then $\operatorname{var}(\bar{X}) = O(n^{-1})$ as $n \to \infty$. Since $\int E(\hat{f}^0 - f)^2$ converges to 0 at a slower rate than n^{-1}, it follows that when the spectrum of the process is bounded at 0 (and our other assumptions are met), the convergence rate of $\hat{f}$ is unaffected by dependence. On the other hand, if the spectrum is unbounded at 0, then the mean integrated squared error (MISE) of $\hat{f}$ can be dominated by the term $E(\bar{X}-\mu)^2 \int (f')^2$. In Sect. 2 we give examples of processes of the form (1.2) for which the convergence rate of $\hat{f}$ is slower than it is in the case of independent data.

Another important consequence of result (1.3) is that the bandwith h which is asymptotically optimal for minimising MISE of the estimator $\hat{f}^0$, also produces asymptotic minimisation of MISE for $\hat{f}$. This is true for both short- and long-range dependence, and is a consequence of the fact that the second term on the right-hand side of (1.3) does not depend on h.

In Sect. 3 we investigate the issue of how dependence affects the convergence rates of not just kernel, but all density estimators. We consider in some detail the parametric problem of estimating the marginal density of a stationary Gaussian process. This example illustrates that, under regularity conditions on the covariance function, all density estimators are subject to the same convergence rate ceiling encountered by kernel estimators. That is to say, none of them have a MISE which converges to zero at a faster rate than $E(\bar{X}-\mu)^2$.

Proofs of our results are given in Sect. 4.

2. Convergence rates for kernel estimators

We begin by defining the infinite order moving average process which generates our data. Let $\{\xi_j, -\infty<j<\infty\}$ be independent and identically distributed random variables with zero mean and variance σ^2, and let $\{a_j, -\infty<j<\infty\}$ be real numbers satisfying $\Sigma a_j^2<\infty$. Define

$$X_j=\mu+\sum_k a_k\,\xi_{k+j}, \qquad -\infty<j<\infty,$$

where μ is an arbitrary but fixed constant. Then $\{X_j\}$ is a stationary process with covariance function

$$r(j)=E(X_0 X_j)=\sigma^2\sum_k a_k\,a_{k+j},$$

and $r(j)\to 0$ as $j\to\infty$. To avoid trivialities we assume that an infinite number of the a_j's are nonzero; otherwise, the sequence $\{X_j\}$ is m-dependent, and then first-order properties of our density estimator are identical to those of an estimator based on an independent sample.

Let f denote the density of X_j. We wish to estimate f from the sample $\{X_1, \ldots, X_n\}$. To this end, define the kernel estimator $\hat{f}(x)$ as in (1.1), where K is a bounded, absolutely integrable function satisfying $\int K=1$. (Typically K is a probability density.) The quantity $h>0$ is termed the bandwidth. Our aim is to compare the performance of $\hat{f}$ with that of an equivalent kernel estimator $\hat{f}^0$ computed from a *random* sample $\{X_1^0, \ldots, X_n^0\}$ drawn from the distribution having density f, and employing the same bandwidth as $\hat{f}$:

$$\hat{f}^0(x)=(nh)^{-1}\sum_{j=1}^{n} K\{(x-X_j^0)/h\}.$$

Now, $\hat{f}$ and $\hat{f}^0$ have the same expected values, and so they have the same biases. Hence their mean integrated squared errors differ only in the variance components:

$$\int E(\hat{f}-f)^2=\int \operatorname{var}(\hat{f})+\int(E\hat{f}-f)^2, \tag{2.1}$$

$$\int E(\hat{f}^0-f)^2=\int \operatorname{var}(\hat{f}^0)+\int(E\hat{f}-f)^2. \tag{2.2}$$

Our first theorem compares the values of $\int \operatorname{var}(\hat{f})$ and $\int \operatorname{var}(\hat{f}^0)$.

We shall impose the following conditions on the distribution of ξ_j and on the sequence $\{a_j\}$: $E(\xi_j^4)<\infty$, and the characteristic function χ of ξ_j satisfies either

(2.3) $$C_1(1+|t|)^{-c_1} \leqq |\chi(t)| \leqq C_2(1+|t|)^{-c_2}, \quad \text{all } t,$$

for positive constants C_1, C_2, c_1, c_2, or

(2.4) $$-|t|^{-c}\log|\chi(t)| \to C \quad \text{as } |t|\to\infty, \quad \text{and} \quad |\chi(t)|>0 \quad \text{for all } t,$$

for positive constants C, c; and as $n\to\infty$,

(2.5) $$\sum_{j=1}^{n}\{(\sup_k |a_k a_{j+k}| + \sup_{|k|\geqq j/2} a_k^2) + (\sum_k |a_k a_{k+j}| + \sum_{|k|\geqq j/2} a_k^2)^2\}$$
$$= o\left\{r(0)+2\sum_{j=1}^{n}(1-n^{-1}j)\,r(j)\right\} + O(1).$$

Condition (2.3) is satisfied if, for example, the ξ_j's have an exponential or gamma distribution, while (2.4) holds if the ξ_j's have a normal distribution. To interpret the right-hand side of (2.5), note that

(2.6) $$n^{-1}\operatorname{var}\left(\sum_{j=1}^{n} X_j\right) = r(0)+2\sum_{j=1}^{n}(1-n^{-1}j)\,r(j).$$

Condition (2.5) holds in a wide range of circumstances; see Remark 2.6.

Theorem 2.1. *Assume that as $n\to\infty$, $h=h(n)\to 0$ and $nh\to\infty$. If the conditions $E(\xi_j^4)<\infty$, either (2.3) or (2.4), and (2.5) hold, then*

(2.7) $$\int \operatorname{var}(\hat f) \sim \int \operatorname{var}(\hat f^0) + \operatorname{var}\left(n^{-1}\sum_{j=1}^{n} X_j\right)\int (f')^2$$

as $n\to\infty$.

Condition (2.3) or (2.4), and the fact that an infinite number of a_j's are nonzero, imply that f is differentiable and $\int (f')^2<\infty$.

Remark 2.1. The variance of $\hat f^0$ is of size $(nh)^{-1}$, and in fact

$$\int \operatorname{var}(\hat f^0) \sim (nh)^{-1}\int K^2$$

(e.g. Silverman [16, p. 40]). From this relation and formula (2.6) we see that (2.7) may be written equivalently as

(2.8) $$\int \operatorname{var}(\hat f) \sim (nh)^{-1}\int K^2 + n^{-1}\left\{r(0)+2\sum_{j=1}^{n}(1-n^{-1}j)\,r(j)\right\}\int (f')^2$$

as $n\to\infty$ and $h\to 0$.

Remark 2.2. Results (2.1), (2.2) and (2.7) together imply that

$$\int E(\hat{f}-f)^2 \sim \int E(\hat{f}^0-f)^2 + E(\bar{X}-\mu)^2 \int (f')^2, \tag{2.9}$$

where $\bar{X}=n^{-1} \sum_{1\leq j\leq n} X_j$ denotes the mean of the sample from which $\hat{f}$ was computed. Thus, we have the following property: *the L^2 rate of convergence of $\hat{f}$ to f equals the worst of the L^2 rates of convergence of $\hat{f}^0$ to f and of $\bar{X}$ to μ*. The presence of the term in $E(\bar{X}-\mu)^2$ in (2.9) implies a ceiling to the rate at which $\hat{f}$ can converge, irrespective of how the bandwidth h is chosen.

Remark 2.3. It is of interest to know whether the convergence rate ceiling of $E(\bar{X}-\mu)^2$ is optimal in any sense – that is, whether a different choice of estimator (e.g., a histogram estimator or an orthogonal series estimator) might be able to do better than a kernel estimator in the case of very-long-range dependence. We shall show in Sect. 3 that the answer to this query is, essentially, "no" – *any* estimator experiences the same ceiling to convergence rates.

Remark 2.4. To more easily appreciate the size of the convergence rate ceiling it is helpful to develop a simpler expression for the size of $E(\bar{X}-\mu)^2$. To this end, define

$$R_n = r(0) + 2\sum_{j=1}^{n} r(j),$$

and assume that for some $C>0$ and all sufficiently large n,

$$0 \leq C^{-1} \max_{1\leq j\leq n} (R_j + C) \leq R_{2n} + C \leq C \min_{n\leq j\leq 2n} (R_j + C).$$

(These inequalities will certainly hold if $r(j)$ is ultimately decreasing.) Then as $n\to\infty$,

$$n + \operatorname{var}\left(\sum_{j=1}^{n} X_j\right) \asymp n(1+|R_n|),$$

where $b_n \asymp c_n$ means that b_n/c_n and c_n/b_n are both bounded as $n\to\infty$. Therefore the convergence rate ceiling is $n^{-1}(1+|R_n|)$. However, we already know that the convergence rate must be slower than n^{-1}, owing to the presence of the term $(nh)^{-1}$ in (2.8). Therefore the convergence rate ceiling is actually $n^{-1}|R_n|$. If the series $\Sigma r(j)$ converges then this term is insignificant relative to $(nh)^{-1}$, and therefore the ceiling has no impact on the rate of convergence of $\hat{f}$. However if $\Sigma r(j)$ diverges then the impact can be significant. Divergence of the series $\Sigma r(j)$ is sometimes taken as the definition of long-range dependence; see Taqqu [18], Cox [5] and Beran [2].

For example, if

$$r(j) = Cj^{-\alpha}, \quad j\geq 1,$$

where $0<\alpha<1$, then we know that no matter how the bandwidth h is chosen (or indeed, no matter how the estimator is chosen, be it a kernel estimator or a histogram estimator or any other type – see Sect. 3), the L^2 convergence rate cannot be faster than $n^{-\alpha}$.

Remark 2.5. If $E(\bar{X}-\mu)^2$ is of smaller order than the minimum MISE for an *independent* sample, i.e. of smaller order than

$$\inf_h \int E(\hat{f}^0 - f)^2, \tag{2.10}$$

then the bandwidth which produces the minimum in (2.10) also produces asymptotic minimisation of MISE for the *dependent* sample, and

$$\inf_h \int E(\hat{f} - f)^2 \sim \inf_h \int E(\hat{f}^0 - f)^2, \tag{2.11}$$

as $n \to \infty$. These results are immediate from (2.9). To give a specific example, suppose K is a symmetric probability density function. Then if f has two continuous, square-integrable derivatives,

$$\inf_h \int E(\hat{f}^0 - f)^2 \sim C_1 n^{-4/5}$$

as $n \to \infty$, where $C_1 > 0$ [16, p. 41]. Hence, if the dependence is sufficiently weak to ensure that $E(\bar{X}-\mu)^2 = o(n^{-4/5})$ (in particular, if $\alpha > 4/5$ in the example in Remark 2.4, or if $\beta > 9/10$ in the example in Remark 2.6 below) then (2.11) holds, and the same bandwidth produces asymptotic minimisation of MISE for both $\hat{f}$ and $\hat{f}^0$. This bandwidth is the traditional one of size $n^{-1/5}$; see [16, pp. 40–41] for a detailed discussion of minimising MISE in the case of independent data.

Remark 2.6. Condition (2.5) holds for a wide variety of different choices of the weight sequence $\{a_j\}$. For example it holds if $a_j = 0$ for $j < 0$ and a_j is regularly varying at infinity as $j \to \infty$; and it holds if $a_j = a_{-j}$ and a_j is regularly varying at $+\infty$. To illustrate why, let us assume for simplicity that $a_j = 0$ for $j < 0$ and $a_j \sim C_1 j^{-\beta}$ as $j \to \infty$, where $\beta > \frac{1}{2}$ and $C_1 \neq 0$. Then

$$r(j) \sim \sum_k |a_k a_{k+j}| \sim \begin{cases} C_2 j^{1-2\beta} & \text{for } \frac{1}{2} < \beta < 1 \\ C_2 j^{-1} \log j & \text{for } \beta = 1 \\ C_2 j^{-\beta} & \text{for } \beta > 1, \end{cases}$$

$$\sup_k |a_k a_{k+j}| + \sup_{|k| \geq j/2} a_k^2 = O(j^{-\beta}), \qquad \sum_{|k| \geq j/2} a_k^2 = O(j^{1-2\beta}),$$

$$\sum_{j=1}^{n} (1 - n^{-1} j)\, r(j) \begin{cases} \sim C_3 n^{2(1-\beta)} & \text{for } \frac{1}{2} < \beta < 1 \\ \sim C_3 (\log n)^2 & \text{for } \beta = 1 \\ = O(1) & \text{for } \beta > 1, \end{cases}$$

and the left-hand side of (2.5) equals $O(n^{1-\beta})$ for $\frac{1}{2} < \beta < 1$, $O(\log n)$ for $\beta = 1$, and $O(1)$ for $\beta > 1$. Condition (2.5) now follows immediately.

It may be shown that if a_j vanishes on one side of the origin or is an even function of j, and if a_j is regularly varying on the side where it does not vanish, then the condition $E(\xi_j^4)$ in Theorem 2.1 may be replaced by $E(|\xi_j|^{2+\varepsilon}) < \infty$ for arbitrary $\varepsilon > 0$.

Remark 2.7. Entirely analogous results may be developed for estimators of derivatives of densities. For example, writing $\hat{f}^{(r)}$, $\hat{f}^{0(r)}$, $f^{(r)}$ for the r'th derivatives of $\hat{f}$, $\hat{f}^0$ and f respectively, we may obtain the following generalisation of (2.9):

$$\int E(\hat{f}^{(r)} - f^{(r)})^2 \sim \int E(\hat{f}^{0(r)} - f^{(r)})^2 + E(\bar{X} - \mu)^2 \int (f^{(r+1)})^2,$$

for each $r \geqq 0$. The conditions of Theorem 2.1 are sufficient for this result, provided K has r continuous derivatives.

3. Issues of optimality

In classical studies of optimality for density estimators, one examines the performance of general density estimators over classes of densities which have only limited smoothness properties. For example, attention usually focusses on the performance of estimators when the true density has only k bounded derivatives, for some fixed k [6, 7, 17]. This approach is of limited scope in the context of infinite moving average processes and infinite autoregressions, since the unknown density f will very often be exceptionally smooth. To more fully appreciate this point, let us return to the model in Sect. 2, where

$$X_j = \mu + \sum_k a_k \xi_{k+j}, \qquad -\infty < j < \infty,$$

and $\{\xi_j, -\infty < j < \infty\}$ is an independent and identically distributed sequence with zero mean. Even if the distribution of ξ_k is very non-smooth, the distribution of X_j can be exceptionally smooth. For example, suppose ξ_j has the discrete two-point distribution given by $P(\xi_j = \pm 1) = \frac{1}{2}$. If a_j is regularly varying in j as $j \to \infty$ then X_j is absolutely continuous and its density has an infinite number of bounded derivatives. (See the end of Sect. 4 for a proof.) Less strikingly, if ξ_j has a density which admits only a Lipschitz condition (without having so much as a single derivative) then the density of X_j has an infinite number of bounded derivatives. In view of these properties it seems pointless, in the case of data generated by infinite moving average processes, to assess optimality the traditional way.

One might argue that it is best to use an estimator which takes full advantage of the smoothness possessed by the underlying density. If the density has an infinite number of derivatives, one might wish to use, for example, a kernel estimator with $K(y) = \sin y/(\pi y)$. However, practitioners usually insist on using estimators (as in Remark 2.5) which are guaranteed to be densities, even though such estimators may not achieve the optimal convergence rate.

We have already noted in Remark 2.4 that, provided the dependence of the moving average process is not *too* long-range, the convergence rate of a kernel estimator applied to moving average data is identical to the rate obtained for a sample of the same size from an i.i.d. process having the same marginal distribution as the moving average. Our remaining problem is to determine the way in which very-long-range dependence affects a *general* estimator, not just a kernel estimator. It turns out that in the case of very-long-range dependence, the convergence rate ceiling for kernel density estimators is virtually identical to a ceiling which arises for general parametric or nonparametric density estimation in a *Gaussian* moving average. This relationship is the subject of Theorems 3.1 and 3.2 below, which treat the special case where the underlying process is Gaussian.

To state Theorem 3.1, let $r(j)$, $j \geqq 0$, be a sequence of real numbers with the property that for some $m_0 \geqq 1$,

$$(3.1) \qquad r(0)+2\sum_{j=1}^{m} r(j)\cos\{2\pi jk/(2m+1)\} \geqq 0, \qquad \text{all } 0 \leqq k \leqq 2m \text{ and } m \geqq m_0.$$

Define $r(-j)=r(j)$. Then there exists a zero mean, stationary Gaussian process $\{Y_j, -\infty<j<\infty\}$ with covariance function r:

$$r(j)=E(Y_k Y_{k+j}), \qquad \text{all } j, k.$$

Let μ be a real number and put

$$(3.2) \qquad X_j=\mu+Y_j, \qquad -\infty<j<\infty.$$

Then X_j has the normal $N\{\mu, r(0)\}$ distribution, with density f_μ. We assume that the sequence $\{r(j)\}$ is known, so that μ is the only unknown quantity. Let $\tilde{f}$ denote any estimator of f_μ. (For example, we might have $\tilde{f}=f_{\hat{\mu}}$, where $\hat{\mu}$ is the maximum likelihood estimator of μ.) Let $\mathscr{I}$ denote a nondegenerate interval within which μ might lie. Theorem 3.1 shows that no matter how $\tilde{f}$ is chosen, the mean integrated squared error of $\tilde{f}$ may converge to zero no more rapidly than

$$n^{-1}\left\{r(0)+2\sum_{j=1}^{n} r(j)\right\}$$

as $n\to\infty$, uniformly in $\mu\in\mathscr{I}$. We note that the results of Grenander and Rosenblatt [9, pp. 88–90] are similar in spirit to Theorem 3.1, but do not yield the rate which we provide.

Theorem 3.1. *Assume that the function r satisfies* (3.1). *Then*

$$(3.3) \qquad \liminf_{n\to\infty} n\{\sum_{|j|\leqq n} r(j)\}^{-1} \sup_{\mu\in\mathscr{I}} \int E(\tilde{f}-f_\mu)^2>0.$$

Theorem 3.2 will provide a lower bound to the convergence rate without assuming (3.1).

Remark 3.1. In most cases of interest the quantity $n^{-1}\{r(0)+2\sum_{1\leqq j\leqq n} r(j)\}$ converges to zero at the same rate as the "extra term" in the variance expansion (2.7), whenever this term dominates the right-hand side of (2.7); see Remark 2.4. Therefore Theorem 3.1 shows that the convergence rate ceiling for kernel estimators, arising from the extra term, is in fact an intrinsic part of the problem of density estimation for dependent data, and is not an artifact of kernel estimators.

Remark 3.2. It is particularly interesting that the convergence rate ceiling exists even in the context of parametric density estimation, where the density f is completely specified except for the single unknown parameter μ. This demonstrates that the ceiling arises because of considerations which are of a parametric, rather than a nonparametric nature.

Remark 3.3. The argument leading to (3.3) may be used to show that if $\tilde{\mu}$ is any estimator of μ then

$$\liminf_{n\to\infty} n\{\sum_{|j|\leqq n} r(j)\}^{-1} \sup_{\mu\in\mathscr{I}} E(\tilde{\mu}-\mu)^2>0. \tag{3.4}$$

Indeed, in both (3.3) and (3.4) we may allow $\mathscr{I}$ to shrink to a point μ_0 as n increases, without upsetting either result. We may even replace $\mathscr{I}$ by the two-point set $\{\mu_0+c_1\delta, \mu_0+c_2\delta\}$, where μ_0 and $c_1 \neq c_2$ are any fixed numbers and $\delta=\{n^{-1}\sum_{|j|\leqq n} r(j)\}^{1/2}$.

Remark 3.4. If $\hat{f}$ denotes the kernel estimator discussed in Sect. 2 then $\int \mathrm{var}(\hat{f})$, and also $\int E(\hat{f}-f)^2$, may be described very easily when the data are Gaussian. Indeed, making use of Lemma 4.1 from the next section we may prove that

$$\begin{aligned}\int \mathrm{var}(\hat{f}) &= (2\pi n)^{-1}\int|\omega(ht)|^2(1-e^{-t^2})\,dt\\ &\quad+(\pi n)^{-1}\sum_{j=1}^{n}(1-n^{-1}j)\int|\omega(ht)|^2 e^{-t^2}(e^{r(j)t^2}-1)\,dt\\ &=(nh)^{-1}\int K^2+2n^{-1}\left\{\sum_{j=1}^{n}(1-n^{-1}j)\,r(j)\right\}\int(f')^2\\ &\quad+O\left[n^{-1}\sum_{j=1}^{n}\{r(j)^2+h^2|r(j)|\}\right],\end{aligned}$$

where ω denotes the Fourier transform of K. Compare this formula for $\int \mathrm{var}(\hat{f})$ with (2.8).

Remark 3.5. Condition (3.1) obviously holds if $\Sigma|r(j)|<\infty$ and $r(0)$ is sufficiently large. It also holds if $r(j)$ is ultimately convex in j (e.g. $r(j)=C_1|j|^{-\alpha}$ where $0<\alpha\leqq 1$) and $r(0)$ is sufficiently large. This follows from the oscillating character of the cosine function, and the fact that convexity entails

$$r(j)-r(j+k)-r(j+2k)+r(j+3k)\geqq 0.$$

Note that $r(j)=C_1|j|^{-\alpha}$ implies

$$n^{-1}\sum_{|j|\leqq n} r(j)\sim\begin{cases}C_3 n^{-\alpha} & \text{if } 0<\alpha<1\\ C_3 n^{-1}\log n & \text{if } \alpha=1.\end{cases} \tag{3.5}$$

The right-hand sides represent convergence rate ceilings for general density estimators when $r(j)=C_1|j|^{-\alpha}$. The same setting was discussed in Remark 2.4, although there with a different import.

Remark 3.6. The analysis in this section is vacuous if the population mean μ is known. It may happen that the availability of extra information (such as values of moments) will, for certain types of processes, lead to improved convergence rates of estimators which are capable of utilising the extra information. However, few of the standard nonparametric density estimators have provision for the incorporation of information which would improve the convergence rate.

Remark 3.7. In some instances, where properties of the spectrum are available, the following theorem can prove useful. Again, $\tilde{f}$ denotes a general estimator of the marginal density f_μ for the Gaussian process $\{X_j\}$ defined at (3.2), based on the sample $X_1, \ldots, X_n$. Let ρ be the spectrum.

$$\rho(\theta) = r(0) + 2 \sum_{j=1}^{\infty} r(j) \cos(j\theta), \qquad -\pi < \theta < \pi.$$

Theorem 3.2. *If r is a covariance function then*

$$\liminf_{n \to \infty} \left\{ \int_0^\pi \min(n^2, \theta^{-2}) \rho(\theta)^{-1} \, d\theta \right\} \sup_{\mu \in \mathscr{I}} \int E(\tilde{f} - f_\mu)^2 > 0.$$

Once again, the theorem may be strengthened and generalised as indicated in Remark 3.3. To illustrate the application of Theorem 3.2, suppose $r(j) \sim C_1 j^{-\alpha}$ as $j \to \infty$, where $0 < \alpha < 1$. Then as $\theta \downarrow 0$,

$$\rho(\theta) \sim \begin{cases} C_2 \theta^{\alpha - 1} & \text{if } 0 < \alpha < 1 \\ C_2 \log \theta^{-1} & \text{if } \alpha = 1, \end{cases}$$

and as $n \to \infty$

$$\int_0^\pi \min(n^2, \theta^{-2}) \rho(\theta)^{-1} \, d\theta \sim \begin{cases} C_3 n^\alpha & \text{if } 0 < \alpha < 1 \\ C_3 n / \log n & \text{if } \alpha = 1, \end{cases}$$

where $C_2, C_3 > 0$. Therefore $\int E(\tilde{f} - f)^2$ may converge to zero no more rapidly than $n^{-\alpha}$ (in the case $0 < \alpha < 1$) or $n^{-1} \log n$ (if $\alpha = 1$). This is also the conclusion reached in Remark 3.5, for the case $r(j) = C_1 j^{-\alpha}$; see (3.5). See Adenstedt [1] and Samarov and Taqqu [15] for related work on estimating the mean.

Remark 3.8. There exist versions of all these results in the context of density derivative estimation. In particular, we have the following analogue of Theorem 3.1: if the function r satisfies (3.1), and if $\tilde{f}^{(r)}$ denotes any estimator of $f^{(r)}$ then

$$\liminf_{n \to \infty} n \{ \sum_{|j| \leq n} r(j) \}^{-1} \sup_{\mu \in \mathscr{I}} \int E(\tilde{f}^{(r)} - f_\mu^{(r)})^2 > 0.$$

Compare Remark 2.7.

4. Proofs

Proof of Theorem 2.1. Our proof is by a sequence of three lemmas. Let *Re* denote "real part", and define

$$\omega(t) = \int e^{itx} K(x) \, dx, \qquad \varphi(t) = E(e^{itX}), \qquad \varphi_n(t) = n^{-1} \sum_{j=1}^{n} \exp(itX_j).$$

Lemma 4.1. *If* $\{X_j, -\infty<j<\infty\}$ *is stationary then*

$$2\pi \int_{-\infty}^{\infty} \operatorname{var}\{\hat{f}(x)\}\, dx = n^{-1} \int |\omega(ht)|^2 \{1-|\varphi(t)|^2\}\, dt$$

$$+2n^{-1} \sum_{j=1}^{n} (1-n^{-1}j) \int |\omega(ht)|^2$$

$$\cdot\, [Re E \exp\{it(X_0-X_j)\} - |\varphi(t)|^2]\, dt.$$

Proof. By Fourier inversion

$$\hat{f}(x) - E\hat{f}(x) = (2\pi)^{-1} \int e^{-itx} \omega(ht) \{\varphi_n(t) - \varphi(t)\}\, dt,$$

and so by Parseval's identity,

$$\int E\{\hat{f}(x) - E\hat{f}(x)\}^2\, dx = (2\pi)^{-1} \int |\omega(ht)|^2 E|\varphi_n(t) - \varphi(t)|^2\, dt.$$

Since

$$E|\varphi_n(t) - \varphi(t)|^2 = n^{-1}\{1-|\varphi(t)|^2\}$$

$$+2n^{-1} \sum_{j=1}^{n} (1-n^{-1}j) [Re E \exp\{it(X_0-X_j)\} - |\varphi(t)|^2]$$

then the lemma is immediate. □

Recall that χ is the characteristic function of ξ_j, and put $\psi = |\chi|^2$. Define

$$a = \sup_k |a_k|, \quad B_T^{-1} = \inf_{|t| \leq aT} |\chi(t)|, \quad \alpha(t) = \prod_k \psi(a_k t) = |E\{\exp(itX_j)\}|^2,$$

$$\alpha_j(t) = Re \prod_k \chi\{(a_k - a_{k+j}) t\} = Re E[\exp\{it(X_0 - X_j)\}].$$

Lemma 4.2. *There exist constants* A_1, $A_2>0$, *depending on* ψ *and the sequence* $\{a_k\}$ *but not on j, t or T, such that if*

$$B_T T^4 \{\sup_k |a_k a_{k+j}| + \sup_{|k| \geq j/2} a_k^2 + (\sum_k |a_k a_{k+j}| + \sum_{|k| \geq j/2} a_k^2)^2\} \leq A_1$$

then

$$|\alpha_j(t) - \alpha(t)\{1 + r(j) t^2\}| \leq A_2 t^4 \alpha(t) \{B_T (\sup_k |a_k a_{k+j}| + \sup_{|k| \geq j/2} a_k^2)$$

$$+ (\sum_k |a_k a_{k+j}| + \sum_{|k| \geq j/2} a_k^2)^2\}$$

uniformly in $0 \leq t \leq T$.

Proof. In the proof, C, C_1, C_2, ... denote generic positive constants not depending on j, k, t or T. The value of C differs from one appearance to another. We assume throughout that $0 \leq t \leq T$ and $j \geq 1$.

Define $\chi_j(t) = \chi^{(j)}(t)/\chi(t)$. Then by Taylor expansion,

$$p_{1jk}(t) \equiv \chi\{(a_k - a_{k+j}) t\}/\chi(a_k t)$$
$$= 1 - a_{k+j} t \chi_1(a_k t) + \tfrac{1}{2}(a_{k+j} t)^2 \chi_2(a_k t) - \tfrac{1}{6}(a_{k+j} t)^3 \chi_3(a_k t) + R_4(t)$$

where $|R_4(t)| \leqq C|a_{k+j}t|^4 B_T$. Furthermore, writing $\beta = E(\xi^3)$ we have

$$|\chi_1(a_k t) + \sigma^2 a_k t + \tfrac{1}{2} i\beta a_k^2 t^2| \leqq C|a_k t|^3 B_T,$$
$$|\chi_2(a_k t) + \sigma^2 + i\beta a_k t| \leqq C|a_k t|^2 B_T,$$
$$|\chi_3(a_k t) + i\beta| \leqq C|a_k t| B_T.$$

Therefore

$$p_{1jk}(t) = 1 + \sigma^2 t^2 (a_k a_{k+j} - \tfrac{1}{2} a_{k+j}^2) + i\beta t^3 (\tfrac{1}{2} a_k^2 a_{k+j} - \tfrac{1}{2} a_k a_{k+j}^2 + \tfrac{1}{6} a_{k+j}^3) + R_2(t)$$

where

$$|R_2(t)| \leqq C_2 B_T t^4 \{(|a_{k+j}| + |a_k|)^4 - a_k^4\}.$$

Similarly,

$$p_{2jk}(t) \equiv \chi\{-(a_k - a_{k-j})t\}/\chi(-a_k t)$$
$$= 1 + \sigma^2 t^2 (a_k a_{k-j} - \tfrac{1}{2} a_{k-j}^2) - i\beta t^3 (\tfrac{1}{2} a_k^2 a_{k-j} - \tfrac{1}{2} a_k a_{k-j}^2 + \tfrac{1}{6} a_{k-j}^3) + R_3(t)$$

where

$$|R_3(t)| \leqq C_1 B_T t^4 \{(|a_{k-j}| + |a_k|)^4 - a_k^4\}.$$

Hence if j, t, T are such that

$$\sigma^2 t^2 (|a_k a_{k\pm j}| + a_{k\pm j}^2) + |\beta t^3| \{(|a_{k\pm j}| + |a_k|)^3 - |a_k|^3\} + C_1 B_T t^4 \{(|a_{k\pm j}| + |a_k|)^4 - a_k^4\} \leqq \tfrac{1}{2}$$

for both choices of the $+$, $-$ signs, and for all $|k| < j/2$; and if also

$$t^2 \sum_k |a_k a_{k+j}| + \sum_{|k| \geqq j/2} a_k^2 \leqq 1;$$

then

$$\prod_{|k|<j/2} p_{1jk}(t) p_{2jk}(t) = 1 + \sigma^2 t^2 \sum_{|k|<j/2} \{a_k(a_{k+j} + a_{k-j}) - \tfrac{1}{2}(a_{k+j}^2 + a_{k-j}^2)\}$$
$$+ i\beta t^3 \sum_{|k|<j/2} \{\tfrac{1}{2} a_k^2 (a_{k+j} - a_{k-j}) - \tfrac{1}{2} a_k (a_{k+j}^2 - a_{k-j}^2) - \tfrac{1}{6}(a_{k+j}^3 - a_{k-j}^3)\} + R_4(t)$$

where

$$|R_4(t)| \leqq C B_T t^4 \sum_{|k|<j/2} \{(|a_{k+j}| + |a_k|)^4 + (|a_{k-j}| + |a_k|)^4 - 2a_k^4\} + Ct^4 (\sum_k |a_k a_{k+j}| + \sum_{|k| \geqq j/2} a_k^2)^2.$$

Recall that $\psi = |\chi|^2$ and let $\mathscr{S}_j$ denote the set of indices k such that either $|k| < j/2$ or $|k+j| < j/2$. Then

$$\sum_{|k|<j/2} a_k(a_{k+j} + a_{k-j}) = \sum_{k \in \mathscr{S}_j} a_k a_{k+j},$$

$$\{\prod_{|k|<j/2} \psi(a_k t)\} \prod_{|k|<j/2} p_{1jk}(t) p_{2jk}(t) = \prod_{k \in \mathscr{S}_j} \chi\{(a_k - a_{k+j})t\}.$$

Therefore

$$(4.1)\ \prod_{k\in\mathscr{S}_j}\chi\{(a_k-a_{k+j})\,t\}=\{\prod_{|k|<j/2}\psi(a_k\,t)\}$$
$$\cdot\,[1+\sigma^2t^2\sum_{k\in\mathscr{S}_j}a_k\,a_{k+j}-\tfrac{1}{2}\sigma^2t^2\sum_{|k+j|<j/2\text{ or }|k-j|<j/2}a_k^2$$
$$+i\beta t^3\sum_{|k|<j/2}\{\tfrac{1}{2}a_k^2(a_{k+j}-a_{k-j})-\tfrac{1}{2}a_k(a_{k+j}^2-a_{k-j}^2)$$
$$-\tfrac{1}{6}(a_{k+j}^3-a_{k-j}^3)\}+R_4(t)].$$

Write $\mathscr{I}_j$ for the set of indices k such that $|k|\geqq j/2$ and $|k+j|\geqq j/2$. Then $\mathscr{I}_j$ is the set of all integers not in $\mathscr{S}_j$. Note that

$$|\chi(t)-(1-\tfrac{1}{2}\sigma^2t^2-\tfrac{1}{6}i\beta t^3)|\leqq C_2\,t^4.$$

If j, t are such that

$$\tfrac{1}{2}\sigma^2\{(a_k-a_{k+j})\,t\}^2+\tfrac{1}{6}|\beta\{(a_k-a_{k+j})\,t\}^3|+C_2\{(a_k-a_{k+j})\,t\}^4\leqq\tfrac{1}{2}$$

whenever $k\in\mathscr{I}_j$, then

$$\sum_{k\in\mathscr{I}_j}\log\chi\{(a_k-a_{k+j})\,t\}=-\tfrac{1}{2}\sigma^2t^2\sum_{k\in\mathscr{I}_j}(a_k-a_{k+j})^2-\tfrac{1}{6}i\beta t^3\sum_{k\in\mathscr{I}_j}(a_k-a_{k+j})^3+R_5(t)$$

where

$$|R_5(t)|\leqq Ct^4\sum_{k\in\mathscr{I}_j}(a_k-a_{k+j})^4.$$

If in addition

$$t^2\sum_{k\in\mathscr{I}_j}(a_k-a_{k+j})^2\leqq 1$$

then

$$\prod_{k\in\mathscr{I}_j}\chi\{(a_k-a_{k+j})\,t\}=1+\sigma^2t^2\sum_{k\in\mathscr{I}_j}a_k\,a_{k+j}$$
$$-\tfrac{1}{2}\sigma^2t^2\Big(\sum_{|k|\geqq j/2\text{ and }|k+j|\geqq j/2}+\sum_{|k|\geqq j/2\text{ and }|k-j|\geqq j/2}\Big)\,a_k^2$$
$$-\tfrac{1}{6}i\beta t^3\sum_{k\in\mathscr{I}_j}(a_k-a_{k+j})^3+R_6(t),$$

where

$$|R_6(t)|\leqq Ct^4\{\sum_{k\in\mathscr{I}_j}(a_k-a_{k+j})^2\}^2.$$

If j, t are such that

$$\tfrac{1}{2}\sigma^2(a_k\,t)^2+\tfrac{1}{6}|\beta(a_k\,t)^3|+C_2(a_k\,t)^4\leqq\tfrac{1}{2}$$

whenever $|k|\geqq j/2$, then

$$\sum_{|k|\geqq j/2}\log|\chi(a_k\,t)|^2=-\sigma^2t^2\sum_{|k|\geqq j/2}a_k^2+R_7(t)$$

where

$$|R_7(t)| \leqq Ct^4 \sum_{|k| \geqq j/2} a_k^4.$$

If in addition

$$t^2 \sum_{|k| \geqq j/2} a_k^2 \leqq 1$$

then

$$\{\prod_{|k| \geqq j/2} \psi(a_k t)\}^{-1} = 1 + \sigma^2 t^2 \sum_{|k| \geqq j/2} a_k^2 + R_8(t),$$

where

$$|R_8(t)| \leqq Ct^4 (\sum_{|k| \geqq j/2} a_k^2)^2.$$

Combining the results from (4.1) down, and noting that the summation operator

$$\sum_{|k+j| < j/2 \text{ or } |k-j| < j/2} + \sum_{|k| \geqq j/2 \text{ and } |k+j| \geqq j/2} + \sum_{|k| \geqq j/2 \text{ and } |k-j| \geqq j/2} - 2 \sum_{|k| \geqq j/2}$$

is identically zero, we see that if j, T are such that

$$B_T T^4 \{\sup_k |a_k a_{k+j}| + \sup_{|k| \geqq j/2} a_k^2 + (\sum_k |a_k a_{k+j}| + \sum_{|k| \geqq j/2} a_k^2)^2\} \leqq C_3$$

then

$$\prod_k \chi\{(a_k - a_{k+j})t\} = \{\prod_k \psi(a_k t)\}\{1 + r(j)t^2 + i\gamma_j t^3 + R(t)\},$$

where γ_j is a real number depending only on j and

$$|R(t)| \leqq Ct^4 [B_T \sum_k \{(|a_{k+j}| + |a_k|)^4 + (|a_{k-j}| + |a_k|)^4 - 4a_k^4\} + \sum_{|k| \geqq j/2} a_k^4 + (\sum_k |a_k a_{k+j}| + \sum_{|k| \geqq j/2} a_k^2)^2].$$

Now, for $l = 1$, 2 or 3,

$$\sum_k |a_k|^l |a_{k \pm j}|^{4-l} \leqq (\sup_k |a_k a_{k+j}|) \sum_k a_k^2.$$

Furthermore,

$$\sum_{|k| \geqq j/2} a_k^4 \leqq (\sup_{|k| \geqq j/2} a_k^2) \sum_k a_k^2.$$

Therefore

$$|R(t)| \leqq Ct^4 \{B_T (\sup_k |a_k a_{k+j}| + \sup_{|k| \geqq j/2} a_k^2) + (\sum_k |a_k a_{k+j}| + \sum_{|k| \geqq j/2} a_k^2)^2\}. \quad \square$$

Define

$$\Delta_j = \sup_k |a_k a_{k+j}| + \sup_{|k| \geqq j/2} a_k^2 + (\sum_k |a_k a_{k+j}| + \sum_{|k| \geqq j/2} a_k^2)^2.$$

Lemma 4.3. *Assume that either*

(4.2) $$C_1 (1+|t|)^{-c_1} \leqq \psi(t) \leqq C_2 (1+|t|)^{-c_2} \quad \text{all } t,$$

where $0 < C_1 < C_2 < \infty$ *and* $0 < c_2 < c_1 < \infty$; *or*

(4.3) $$-|t|^{-c} \log \psi(t) \to C \quad \text{as } |t| \to \infty \quad \text{and} \quad \psi(t) > 0 \quad \text{for all } t,$$

where $C, c > 0$. *Then*

$$\sum_{j=1}^{n} (1 - n^{-1} j) \int_0^\infty |\omega(ht)|^2 \{\alpha_j(t) - \alpha(t)\} \, dt$$

$$= \left\{ \sum_{j=1}^{n} (1 - n^{-1} j) \, r(j) \right\} \int_0^\infty |\omega(ht)|^2 t^2 \alpha(t) \, dt + O\left(\sum_{j=1}^{n} \Delta_j \right)$$

as $n \to \infty$.

Proof. We prove Lemma 4.3 by applying Lemma 4.2 with $T = t$. Throughout the proof, symbols $C_1, C_2, \ldots$ denote generic positive constants.

If (4.2) holds then since $a_k \neq 0$ for an infinite number of k's, and since $\alpha(t) = \Pi \psi(a_k t)$, given any $c_3 > 0$ we may find $C_3, j_0 > 0$ such that for all $j \geqq j_0$,

(4.4) $$\alpha(t) + |\alpha_j(t)| \leqq C_3 (1+|t|)^{-c_3}, \quad \text{all } t.$$

Since $B_t \leqq C_1^{-1} (1+t)^{c_1}$ then by Lemma 4.2,

(4.5) $$|\alpha_j(t) - \alpha(t) \{1 + r(j) t^2\}| \leqq A_2 C_1^{-1} C_3 (1+t)^{c_1 - c_3 + 4} \Delta_j$$

provided that $0 \leqq t \leqq T_j$ where T_j is the solution of

$$(1 + T_j)^{c_4} = A_1 C_1 \Delta_j^{-1}$$

and $c_1 + 4 \leqq c_4 \leqq c_3 - 3$. By (4.5), if $c_3 \geqq c_1 + 7$,

$$\int_0^{T_j} |\alpha_j(t) - \alpha(t) \{1 + r(j) t^2\}| \, dt \leqq C_4 \Delta_j,$$

and by (4.4),

$$\int_{T_j}^\infty \{|\alpha_j(t)| + \alpha(t) |1 + r(j) t^2|\} \, dt \leqq C_5 \Delta_j^{(c_3 - 3)/c_4}$$

$$\leqq C_6 \Delta_j.$$

Therefore

(4.6) $$\left| \int_0^\infty |\omega(ht)|^2 [\alpha_j(t) - \alpha(t) \{1 + r(j) t^2\}] \, dt \right| \leqq C_7 \Delta_j$$

for $j \geqq j_0$.

If (4.3) holds then for each $\eta > 0$,

$$C_1 \exp\{-(C+\eta)|at|^c\} \leqq \psi(t), \quad \text{all } t,$$

where $a = \sup|a_k|$ and $C_1 = C_1(\eta) > 0$. In place of (4.4) we have for all $j \geqq j_0$,

$$\alpha(t) + |\alpha_j(t)| \leqq C_2 \exp(-C_3|at|^c), \quad \text{all } t, \tag{4.7}$$

where $C_3 > C$. Since $B_t \leqq C_1^{-1} \exp\{(C+\eta)(at)^c\}$ for $t > 0$ then by Lemma 4.2,

$$|\alpha_j(t) - \alpha(t)\{1 + r(j)t^2\}| \leqq A_2 C_1^{-1} C_2 t^4 \exp\{(C+\eta-C_3)(at)^c\} \Delta_j \tag{4.8}$$

provided that $0 \leqq t \leqq T_j$, where T_j is the solution of

$$T_j^4 \exp\{(C+\eta)(aT_j)^c\} = A_1 C_1 \Delta_j^{-1}.$$

Take $\eta = (C_3 - C)/2$. Then by (4.8),

$$\int_0^{T_j} |\alpha_j(t) - \alpha(t)\{1 + r(j)t^2\}|\, dt \leqq C_4 \Delta_j,$$

and by (4.7),

$$\int_{T_j}^{\infty} \{|\alpha_j(t)| + \alpha(t)|1 + r(j)t^2|\}\, dt \leqq C_5 T_j^3 \exp\{-C_3(aT_j)^c\}$$
$$\leqq C_6 \Delta_j$$

for $j \geqq j_1$, say. Result (4.6) follows.

It is a trivial matter to establish (4.6) in the case $1 \leqq j \leqq \max(j_0, j_1)$. This completes the proof. □

Theorem 2.1 follows from Lemmas 4.1 and 4.3, on noting that

$$\int |\omega(ht)|^2 \{1 - |\varphi(t)|^2\}\, dt \sim \int |\omega(ht)|^2\, dt = 2\pi h^{-1} \int K^2,$$
$$\int |\omega(ht)|^2 t^2 \alpha(t)\, dt \sim \int t^2 \alpha(t)\, dt = 2\pi \int (f')^2. \quad \square$$

Proof of Theorem 3.1. We prove the stronger variant described in Remark 3.3, in which the interval $\mathscr{I}$ is replaced by $\{\mu_0 + c_1\delta, \mu_0 + c_2\delta\}$, and μ_0, $c_1 \neq c_2$ are constants. Without loss of generality we may take $\mu_0 = c_1 = 0$ and $c_2 = 1$. This reduces the problem to one of estimating θ in the model

$$X_j = \theta\delta + Y_j \qquad 1 \leqq j \leqq n, \tag{4.9}$$

where $\theta = 0$ or 1 and $\delta = \delta(n) \to 0$ as $n \to \infty$, with $\delta \geqq 0$. Write f_θ and P_θ for the probability density of X_j and the probability measure under model (4.9), respectively, when the parameter value is θ.

Our proof is in a sequence of four steps.

Step (i): A likelihood ratio rule. Put $\mathbf{X} = (X_1, \ldots, X_n)^T$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$, let $\mathbf{1} = (1, \ldots, 1)^T$ be a vector of length n, and let $\mathbf{V}$ denote the $n \times n$ variance matrix of $\mathbf{Y}$. The likelihood of θ under the model (4.9) is proportional to

$$L(\theta) = \exp\{-\tfrac{1}{2}(\mathbf{X} - \theta\delta\mathbf{1})^T \mathbf{V}^{-1} (\mathbf{X} - \theta\delta\mathbf{1})\},$$

the constant of proportionality not depending on θ. Hence if $\hat{\theta}$ is the likelihood ratio rule for discriminating between $\theta=0$ and $\theta=1$ then

$$\begin{aligned} P_0(\hat{\theta}=1) &= P\{L(0)/L(1)<1 \mid \theta=0\} \\ &= P(\mathbf{Y}^T\mathbf{V}^{-1}\mathbf{1}>\tfrac{1}{2}\delta\mathbf{1}^T\mathbf{V}^{-1}\mathbf{1}) \\ &= 1-\Phi\{\tfrac{1}{2}\delta(\mathbf{1}^T\mathbf{V}^{-1}\mathbf{1})^{1/2}\} = P_1(\hat{\theta}=0). \end{aligned}$$

If $\tilde{\theta}$ is any rule for discriminating between $\theta=0$ and $\theta=1$ then, by the Neyman-Pearson lemma

$$\begin{aligned} P_0(\tilde{\theta}=1)+P_1(\tilde{\theta}=0) &\geqq P_0(\hat{\theta}=1)+P_1(\hat{\theta}=0) \\ &= 2[1-\Phi\{\tfrac{1}{2}\delta(\mathbf{1}^T\mathbf{V}^{-1}\mathbf{1})^{1/2}\}]. \end{aligned} \tag{4.10}$$

Step (ii): Estimating truncated moments. Consider the problem of estimating the truncated moment

$$v(\theta)=\int_{-1}^{1} x f_\theta(x)\,dx=\int_{-1}^{1} x p(x-\delta\theta)\,dx,$$

where p denotes the $N\{0, r(0)\}$ density. Note that $v(0)=0$ and $v(1)\sim\delta C_1$ as $\delta\to 0$, where

$$C_1=-\int_{-1}^{1} x p'(x)\,dx>0.$$

Let $C_2>0$ be a constant such that $v(1)>C_2\,\delta$ for all $0\leqq\delta\leqq 1$. Put

$$\tilde{v}=\int_{-1}^{1} x\tilde{f}(x)\,dx,$$

and define $\tilde{\theta}$ to equal 0 or 1 according as $\tilde{v}$ is closer to $v(0)$ or $v(1)$. Now,

$$\begin{aligned} \{\tilde{v}-v(\theta)\}^2 &= \left[\int_{-1}^{1}\{\tilde{f}(x)-f_\theta(x)\}\,x\,dx\right]^2 \\ &\leqq \int_{-1}^{1}\{\tilde{f}(x)-f_\theta(x)\}^2|x|\,dx \\ &\leqq \int\{\tilde{f}(x)-f_\theta(x)\}^2\,dx\equiv I_\theta, \end{aligned}$$

say. If $\tilde{\theta}\neq\theta$ then

$$\{\tilde{v}-v(\theta)\}^2\geqq\tfrac{1}{4}\{v(0)-v(1)\}^2=\tfrac{1}{4}v(1)^2>C_3\,\delta^2,$$

where $C_3=C_2^2/4$. Therefore

$$\begin{aligned} \max_{\theta=0,1} P_\theta(I_\theta>C_3\,\delta^2) &\geqq \max_{\theta=0,1} P_\theta(\tilde{\theta}\neq\theta) \\ &\geqq \tfrac{1}{2}\{P_0(\tilde{\theta}=1)+P_1(\tilde{\theta}=0)\}. \end{aligned} \tag{4.11}$$

Combining (4.10) and (4.11) we deduce that

$$\max_{\theta=0,1} E(I_\theta) \geqq C_3\,\delta^2 \max_{\theta=0,1} P_\theta(I_\theta > C_3\,\delta^2) \tag{4.12}$$

$$\geqq C_3\,\delta^2[1-\Phi\{\tfrac{1}{2}\delta(\mathbf{1}^T\mathbf{V}^{-1}\mathbf{1})^{1/2}\}].$$

Step (iii): A circulant moving average process. Put $N=2n+1$ and

$$\bar{r}(j)=\begin{cases} r(\min\{|j|, N-|j|\}) & \text{for } 0\leqq|j|\leqq N-1 \\ \bar{r}(j \bmod N) & \text{for general } j. \end{cases}$$

Let $a_0, \ldots, a_{N-1}$ be constants, let $\xi_0, \ldots, \xi_{N-1}$ be independent standard normal random variables, and define the circulant moving average process

$$\eta_j = \sum_{k=0}^{N-1} a_k\,\xi_{k+j}, \qquad -\infty<j<\infty,$$

where the subscript of ξ_{k+j} is to be interpreted modulo N. In this step we show that under condition (3.1), there exist real numbers $a_0, \ldots, a_{N-1}$ which give the stationary Guassian process η_j the covariance function $\bar{r}(j)$:

$$E(\eta_j\,\eta_k)=\bar{r}(j-k), \qquad -\infty<j,\,k<\infty. \tag{4.13}$$

Observe that

$$E(\eta_j\,\eta_k)=\sum_{l=0}^{N-1} a_l\,a_{j-k+l},$$

where (here and below) the subscript of a_j is to be interpreted modulo N. Therefore we wish to choose the a_j's such that

$$\sum_{j=0}^{N-1} a_j\,a_{j+k}=\bar{r}(k), \qquad 0\leqq k\leqq N-1.$$

Multiply both sides by $e^{2\pi k l i/N}$, where $0\leqq l\leqq N-1$ and $i=(-1)^{1/2}$, and add over $0\leqq k\leqq N-1$:

$$\sum_{j=0}^{N-1} a_j\,e^{-2\pi j l i/N} \sum_{k=0}^{N-1} a_{j+k}\,e^{2\pi(j+k)li/N}=\sum_{k=0}^{N-1} \bar{r}(k)\,e^{2\pi k l i/N};$$

that is,

$$|A_1(2\pi l/N)|^2=A_2(2\pi l/N), \qquad 0\leqq l\leqq N-1, \tag{4.14}$$

where

$$A_1(\theta)=\sum_{k=0}^{N-1} a_k\,e^{ik\theta},$$

$$A_2(\theta)=r(0)+2\sum_{k=1}^{n} r(k)\cos(k\theta).$$

By hypothesis,

$$r(0)+2\sum_{k=1}^{n} r(k)\cos(2\pi kl/N)\geqq 0, \qquad 0\leqq l\leqq N-1.$$

Hence $A_2(2\pi l/N)\geqq 0$ for $0\leqq l\leqq N-1$. In this event we may solve the N simultaneous equations at (4.14) for the N *real* numbers $a_0, \ldots, a_{N-1}$. (Considerations of symmetry dictate that $a_j=a_{N-j}$ for $1\leqq j\leqq n$, so that there are in fact only $n+1$ a_j's to be determined, and

$$A_1(\theta)=a_0+2\sum_{j=1}^{n} a_j\cos(j\theta) \qquad \text{for } \theta=2\pi l/N, \qquad 0\leqq l\leqq n.)$$

The numbers a_j will, by construction, satisfy (4.13).

Step (iv): Inference for the circulant moving average process. Let $\{\eta_j, -\infty<j<\infty\}$ denote the circulant moving average process introduced in Step (iii), with $\bar r(0)\geqq 2B$. Let $\hat\theta_N$ denote the likelihood ratio rule for discriminating between $\theta=0$ and $\theta=1$ in the model

$$X_{Nj}=\theta\delta+\eta_j, \qquad 0\leqq j\leqq N-1.$$

Write $\mathbf{V}_N$ for the $N\times N$ variance matrix of $\eta=(\eta_0, \ldots, \eta_{N-1})^T$, and put $\mathbf{Y}_N=(Y_0, \ldots, Y_{N-1})^T$ and $\mathbf{1}_N=(1, \ldots, 1)^T$, both being column vectors of length N. The argument which formerly led to (4.10) now shows that

$$P_0(\hat\theta_N=1)+P_1(\hat\theta_N=0)=2[1-\Phi\{\tfrac{1}{2}\delta(\mathbf{1}_N^T\mathbf{V}_N^{-1}\mathbf{1}_N)^{1/2}\}].$$

Since the sequence $X_{N,0}, \ldots, X_{N,N-1}$ contains a subsequence having the same distribution as $X_1, \ldots, X_n$ then

$$P_0(\hat\theta=1)+P_1(\hat\theta=0)\geqq P_0(\hat\theta_N=1)+P_1(\hat\theta_N=0).$$

That is

$$1-\Phi\{\tfrac{1}{2}\delta(\mathbf{1}^T\mathbf{V}^{-1}\mathbf{1})^{1/2}\}\geqq 1-\Phi\{\tfrac{1}{2}\delta(\mathbf{1}_N^T\mathbf{V}_N^{-1}\mathbf{1}_N)^{1/2}\}. \tag{4.15}$$

The matrix $\mathbf{V}_N=(v_{ij})$ is a circulant, and has $\mathbf{1}_N$ as an eigenvector. The corresponding eigenvalue is

$$\lambda=\sum_{j=0}^{N-1} v_{ij}=\bar r(0)+\sum_{j=1}^{N-1}\bar r(j)=r(0)+2\sum_{j=1}^{n} r(j).$$

Therefore

$$\mathbf{1}_N^T\mathbf{V}_N^{-1}\mathbf{1}_N=N\lambda^{-1}=N\left\{r(0)+2\sum_{j=1}^{n} r(j)\right\}^{-1}. \tag{4.16}$$

Combining (4.12), (4.15) and (4.16) we obtain

$$\max_{\theta=0,1} E(I_\theta)\geqq C_3\,\delta^2\left(1-\Phi\left[\frac{1}{2}\,\delta N^{1/2}\left\{r(0)+2\sum_{j=1}^{n} r(j)\right\}^{-1/2}\right]\right).$$

Taking $\delta=\{n^{-1}\sum_{|j|\leq n} r(j)\}^{1/2}$ in this inequality, and letting $n\to\infty$, we deduce that

$$\liminf_{n\to\infty}\left\{n\Big/\sum_{|j|\leq n} r(j)\right\}\max_{\theta=0,1} E(I_\theta)\geq C_3\left\{1-\Phi\left(\frac{1}{\sqrt{2}}\right)\right\}.$$

Theorem 3.1 is immediate. □

Proof of Theorem 3.2. The proof of this result is identical to that just given, except that we replace Steps (iii) and (iv) by a new Step (iii)′, below.

Step (iii)′: Making use of extra information. Suppose that instead of the process $X_j=\mu+Y_j$, $1\leq j\leq n$, we observe $X_j=v_j+Y_j$, $-\infty<j<\infty$, where $v_j=\mu$ for $1\leq j\leq n$ and $v_j=0$ otherwise. The class of all estimators $\tilde{J}$ based on the greater information in this extended sequence includes the class of estimators based on the subsequence $X_1, \ldots, X_n$. Therefore the argument which formerly led to (4.12) now produces the inequality

$$\max_{\theta=0,1} E(I_\theta)\geq C_3\,\delta^2[1-\Phi\{\tfrac{1}{2}\delta(\mathbf{u}^T\mathbf{U}^{-1}\mathbf{u})^{1/2}\}], \tag{4.17}$$

where $\mathbf{u}$ is the doubly infinite vector with 1 in positions $1, \ldots, n$ and zero elsewhere, and $\mathbf{U}=(r(j-k))$ is the doubly infinite variance matrix of the sequence $\{Y_j\}$.

If the sample size is odd, say $n=2m+1$, then we may write

$$\mathbf{u}^T\mathbf{U}^{-1}\mathbf{u}=(2\pi)^{-1}\int_{-\pi}^{\pi} u(t)^2\rho(t)^{-1}\,dt$$

where

$$\begin{aligned} u(t)&=\sum_{j=-m}^{m} e^{ijt}=1+2\sum_{j=1}^{m}\cos jt\\ &=\cos(mt)+(1-\cos t)^{-1}(\sin t)\sin(mt). \end{aligned}$$

Now, $|u(t)|\leq C_4\min(n, t^{-1})$, and so

$$\mathbf{u}^T\mathbf{U}^{-1}\mathbf{u}\leq(C_4^2/\pi)\int_0^{\pi}\min(n^2, t^{-2})\,\rho(t)^{-1}\,dt.$$

Hence by (4.17)

$$\max_{\theta=0,1} E(I_\theta)\geq C_3\,\delta^2\left(1-\Phi\left[C_5\,\delta\left\{\int_0^{\pi}\min(n^2, t^{-2})\,\rho(t)^{-1}\,dt\right\}^{1/2}\right]\right), \tag{4.18}$$

where $C_5=C_4/(2\pi^{1/2})$. The case of even n may be treated by considering the enlarged sample $X_1, \ldots, X_{n+1}$. Taking

$$\delta=\left\{\int_0^{\pi}\min(n^2, t^{-2})\,\rho(t)^{-1}\,dt\right\}^{-1/2}$$

in (4.18) we deduce that

$$\liminf_{n\to\infty}\left\{\int_0^{\pi} \min(n^2, t^{-2})\,\rho(t)^{-1}\,dt\right\} \max_{\theta=0,1} E(I_\theta) \geqq C_3\{1-\Phi(C_5)\}. \quad \square$$

Finally we verify a claim made in the first paragraph of Sect. 3, by showing that if the sequence a_j satisfies

$$(\log x)^{-1} x^2 \sum_{j:\,|a_j|\leqq x^{-1}} a_j^2 \to +\infty \tag{4.19}$$

as $x\to+\infty$, and if

$$X_0 = \mu + \sum_j a_j \xi_j$$

where the ξ_j's are independent and identically distributed with the distribution

$$P(\xi_j = \pm 1) = \tfrac{1}{2},$$

then X_0 has an absolutely continuous distribution with a density that admits an infinite number of uniformly bounded derivatives. Condition (4.19) is satisfied in a variety of circumstances, for example if a_j is a regularly varying function of j as $j\to\infty$.

Observe that X_0 has characteristic function $\beta(t) = e^{it\mu} \prod_j \chi(a_j t)$, where

$$\chi(t) = E\{\exp(it\xi_j)\} = \cos t.$$

Since $\cos\theta \leqq 1 - C_1\theta^2 \leqq \exp(-C_1\theta^2)$ for $|\theta|\leqq 1$ then

$$|\beta(t)| \leqq \prod_{j:\,|a_j t|\leqq 1} \cos(a_j t) \leqq \exp\Big\{-C_1 \sum_{j:\,|a_j t|\leqq 1} (a_j t)^2\Big\}.$$

Hence by (4.19), $|t^\lambda \beta(t)| \to 0$ as $|t|\to\infty$, for each $\lambda>0$. It follows that the density of X_0 is infinitely differentiable.

References

1. Adenstedt, R.K.: On large-sample estimation for the mean of a stationary random sequence. Ann. Statist. **2**, 1095–1107 (1974)
2. Beran, J.: Estimation, testing and prediction of self-similar and related processes. Ph.D. thesis, ETH Zürich, No. 8074 (1986)
3. Castellana, J.V., Leadbetter, M.R.: On smoothed probability density estimation for stationary processes. Stochastic Processes Appl. **21**, 179–193 (1986)
4. Chanda, K.C.: Density estimation for linear processes. Ann. Inst. Statist. Math. **35**, 439–446 (1983)
5. Cox, D.R.: Long-range dependence: a review. In: David, H.A., David, H.T. (eds.) Proceedings, 50th Anniversary Conference Iowa State Statistical Laboratory, pp. 55–74. Iowa State University Press 1984
6. Farrell, R.H.: On the lack of a uniformly consistent sequence of estimators of a density function in certain cases. Ann. Math. Statist. **38**, 471–474 (1967)
7. Farrell, R.H.: On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. Ann. Math. Statist. **43**, 170–180 (1972)

8. Graf, H.P.: Long-range correlations and estimation of the self-similarity parameter. Ph.D. thesis, ETH Zürich, No. 7357 (1983)
9. Grenander, U., Rosenblatt, M.: Statistical analysis of stationary time series. New York: Wiley 1957
10. Hart, J.D.: Efficiency of a kernel density estimator under an autoregressive dependence model. J. Am. Stat. Assoc. **79**, 110–117 (1984)
11. Mandelbrot, B.B., Wallis, J.R.: Some long-run properties of geophysical records. Water Resources Research **5**, 321–340 (1969)
12. Priestley, M.B.: Spectral analysis and time series. New York: Academic Press 1981
13. Robinson, P.M.: On the consistency and finite-sample properties of non-parametric kernel time series regression, autoregression and density estimators. Ann. Inst. Statist. Math. **38**, 539–549 (1986)
14. Rosenblatt, M.: Density estimates and Markov sequences. In: Puri, M. (ed.) Non-parametric techniques in statistical inference, pp. 199–210. London: Cambridge University Press 1970
15. Samorov, A., Taqqu, M.S.: On the efficiency of the sample mean in long-memory noise. J. Time Ser. Anal. **9**, 191–200 (1988)
16. Silverman, B.W.: Density estimation for statistics and data analysis. London: Chapman and Hall 1986
17. Stone, C.J.: Optimal global rates of convergence for nonparametric regression. Ann. Statist. **10**, 1040–1053 (1982)
18. Taqqu, M.S.: Weak convergence to fractional Brownian motion and to the Rosenblatt process. Z. Wahrscheinlichkeitstheor. Verw. Geb. **31**, 287–302 (1975)