Ann. Inst. Statist. Math.
Vol. 48, No. 3, 429–449 (1996)

KERNEL DENSITY ESTIMATION FOR LINEAR PROCESSES: ASYMPTOTIC NORMALITY AND OPTIMAL BANDWIDTH DERIVATION

MARC HALLIN[1] AND LANH TAT TRAN[2]*

[1] *Institut de Statistique, Université Libre de Bruxelles, Campus de la Plaine CP 210, Boulevard du Triomphe, B-1050 Bruxelles, Belgium*
[2] *Department of Mathematics, College of Arts and Sciences, Indiana University, Rawles Hall, Bloomington, IN 47405-5701, U.S.A.*

(Received January 4, 1994; revised September 21, 1995)

Abstract. The problem of estimating the marginal density of a linear process by kernel methods is considered. Under general conditions, kernel density estimators are shown to be asymptotically normal. Their limiting covariance matrix is computed. We also find the optimal bandwidth in the sense that it asymptotically minimizes the mean square error of the estimators. The assumptions involved are easily verifiable.

Key words and phrases: Density estimation, linear process, kernel, bandwidth, mean square error.

1. Introduction

The literature dealing with density estimation when the observations are independent random variables (r.v.'s) is extensive. Density estimation for dependent r.v.'s has recently received increasing attention. This paper is concerned with density estimation when the observations come from a linear process. Many important time series models such as the autogressive processes and the mixed autoregressive moving average time series models are linear processes. Long memory fractional processes also belong to that class. Since parameter estimation in time series analysis is generally carried out under the Gaussian assumption or, at least, from Gaussian likelihood methods, it may be useful to check whether or not the density of a time series is Gaussian or not. Recently, density estimation for time series has been employed in the problem of testing serial dependence (see Chan and Tran (1991)). For further motivation and background material, the reader is referred to Chanda (1983), Robinson (1983), (1986), (1987), Yakowitz (1985), (1987), Masry (1986, 1987), Masry and Györfi (1987), Ioannides and Roussas

* Supported in part by NSF grant DMS-9403718.

(1987), Roussas (1988), Gyorfi *et al.* (1989), and Tran (1989, 1992). Chanda (1983) has investigated the asymptotic normality and consistency of kernel estimators of the marginal density for linear processes. Recently, Tran (1992) has shown that such estimators can achieve sharp rates of convergence of L_∞ norm on compact sets.

The general setting is the following: $X_1, \ldots, X_n$ are n consecutive observations of a linear process $X_t = \mu + \sum_{r=0}^{\infty} a_r Z_{t-r}$, where μ is a constant and $\{Z_t\}$ is an innovation process consisting of independent and identically distributed r.v.'s with mean zero and finite variance. Assume that X_1 has a probability density f, which we wish to estimate. As an estimator of f, we will consider the nonparametric kernel estimate (see Rosenblatt (1956) and Parzen (1962)) given by

$$f_n(x) = (nb_n)^{-1} \sum_{i=1}^{n} K((x - X_i)/b_n),$$

where K is a kernel function and $\{b_n\}$ is a sequence of bandwidths with b_n tending to zero as n tends to infinity.

In Section 2, we provide some preliminaries which are crucial for the proofs of our results in Sections 3, 4, and 5. Section 3 studies the asymptotic distribution of $(f_n(x_1), \ldots, f_n(x_k))'$ for distinct points $x_1, \ldots, x_k$ of R. For linear processes with $|a_r| = O(r^{-(4+\delta)})$ for some $\delta > 0$, Theorem 3.1 gives general conditions under which the limiting distribution of $(nb_n)^{1/2}(f_n(x_1) - Ef_n(x_1), \ldots, f_n(x_k) - Ef_n(x_k))'$ is normal. The conditions are stated in Assumptions 1–3. Assumption 1 involves some standard conditions imposed on the kernel K. Assumption 2 concerns the linear process X_t. Assumption 3 points out an important difference between density estimation for i.i.d. r.v.'s and for linear processes. There exists a delicate relationship between the rates at which the bandwidths and the coefficients of the linear process tend to zero. As can be expected, the bandwidth has to tend to zero more slowly at small values of δ than at large values of δ. When δ is large, the condition imposed on b_n of Assumption 3 is marginally close to the usual condition that $nb_n \to \infty$ normally seen in the independent case. In the case $k = 1$, Chanda (1983) has investigated the asymptotic normality of f_n under different regularity conditions than ours. Theorem 2.1 of Chanda (1983) shows that $n^{1/2}(f_n(x) - Ef_n(x))$ is asymptotically normal under general conditions. The only condition imposed on the bandwidth is that $nb_n \to \infty$, which is identical to the usual condition imposed in the independent case. This result appears to be invalid. There are some gaps in his arguments. Condition (iii) of Lemma 2.3 in Chanda (1983) is inconsistent with Theorem 2.1 in the same paper.

In Section 4, we consider the asymptotic normality of f_n in the particular case when a_r tends to zero at an exponential rate specified in Assumption $2'$. We sense a practical need for this case since Assumption $2'$ is satisfied by most causal-invertible autoregressive moving average time series models. Assumption $3'$ imposes a very weak condition on the bandwidth for the asymptotic normality of $(nb_n)^{1/2}(f_n(x_1) - Ef_n(x_1), \ldots, f_n(x_k) - Ef_n(x_k))'$.

We compute the limiting covariance matrix of $(f_n(x_1), \ldots, f_n(x_k))'$ in Section 5 for the general case. This result is then used to find the optimal bandwidth which asymptotically minimizes the mean square error.

Most of the results on density estimation for time series hold under some assumptions on the type of dependence of the relevant processes, for example, strong mixing and absolute regularity (see, for example, Roussas (1988) and Tran (1989)). The results of the present paper cannot be obtained by applying known theorems for mixing or absolutely regular r.v.'s. The discussion on this point is rather technical and will be given in Section 2.

Our method of proof is based on two different truncations of X_t. The truncation in the definition of $\hat{X}_t$ in the beginning of Section 2 is somewhat standard. The point of truncation $(q-1)$ varies only with n. However, the truncation in the definition of X^*_{1+v} in Lemma 2.4 varies with the index of the variable and is rather unusual.

Here and throughout the paper x denotes a fixed point of the real line (R). We use $[a]$ at times to indicate the integer part of a number a. All limits are taken as n tends to infinity unless otherwise indicated. For sequences of numbers $\{a_n\}$ and $\{b_n\}$, we write $a_n \sim b_n$ to indicate that $a_n/b_n \to 1$.

2. Preliminaries

It will be clear from the proofs of the paper that we can without loss of generality assume $\mu = 0$ and $a_0 \neq 0$. Define

$$\hat{X}_t = \sum_{r=0}^{q-1} a_r Z_{t-r},$$

where $q = q(n) \uparrow \infty$ is a sequence of positive integers to be specified later on in (2.13). Let

$$\tilde{X}_t = \hat{X}_t + \Gamma_t,$$

where the Γ_t's are independent r.v.'s with

$$\Gamma_t \sim \sum_{r=q}^{\infty} a_r Z_{t-r}. \tag{2.1}$$

We also require that $\{Z_t\}$ and $\{\Gamma_t\}$ be independent sequences. Note that both the processes $\{\hat{X}_t\}$ and $\{\tilde{X}_t\}$ are q-dependent. Consider the kernel estimator $g_n(x)$ defined by

$$g_n(x) = (nb_n)^{-1} \sum_{i=1}^{n} K((x - \tilde{X}_i)/b_n). \tag{2.2}$$

Define the average kernel $K_n(x)$ by

$$K_n(x) = \frac{1}{b_n} K(x/b_n). \tag{2.3}$$

Then

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_n(x - X_i)$$

and

$$g_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_n(x - \tilde{X}_i). \tag{2.4}$$

Define

$$\Delta_i(x) = K_n(x - X_i) - \mu_n, \qquad \tilde{\Delta}_i(x) = K_n(x - \tilde{X}_i) - \tilde{\mu}_n \tag{2.5}$$

where

$$\mu_n = EK_n(x - X_i), \qquad \tilde{\mu}_n = EK_n(x - \tilde{X}_i). \tag{2.6}$$

Clearly, $\mu_n = \tilde{\mu}_n$ since X_i and $\tilde{X}_i$ have the same distribution.

ASSUMPTION 1. The kernel function K is a density function with an integrable radial majorant $Q(x)$, that is, $Q(x) \equiv \sup\{|K(y)| : |y| \geq |x|\}$ is integrable. Assume in addition that K satisfies the following Lipschitz condition

$$|K(x) - K(y)| \leq C|x - y|.$$

ASSUMPTION 2. The coefficients of the linear process X_t tend to zero sufficiently fast that $|a_r| = O(r^{-(4+\delta)})$ for some $\delta > 0$ as $r \to \infty$. In addition, Z_1 has mean zero and finite variance and an absolutely integrable characteristic function ϕ.

ASSUMPTION 3. The bandwidth b_n tends to zero sufficiently slow so that

$$nb_n^{(13+2\delta)/(3+2\delta)}(\log\log n)^{-1} \to \infty.$$

A linear process is trong mixing or absolutely regular if it satisfies a variety of conditions (see Gorodetskii (1977) and Withers (1981)). One condition is that $\sum_{r=0}^{\infty} a_r z^r$ is not equal to zero for $|z| \leq 1$ (no roots inside the unit circle). This condition is often impossible to check.

There are numerous linear processes satisfying Assumption 2 without being strong mixing. If a_r is the coefficient of the power series expansion of $a(z) = (1-z)^p$ where $p > 5$ is a noninteger, then the condition of "no roots in the unit circle" is violated. In this case $|a_r| = O(r^{1-p})$ and Assumption 2 is satisfied since $p > 5$. Indeed, if Z_i's are normally distributed with mean zero and variance 1, then $\{X_i\}$ has a spectral density equal to $|a(e^{i\theta})|^2$ (see Gorodetskii (1977)) and it follows from the Helson-Sarason theorem (see Ibragimov and Rozanov (1978), Helson and Sarason (1967) and Sarason (1972)) that X_t does not satisfy the strong mixing condition.

LEMMA 2.1. *If Assumption* 1 *holds, then*

$$\int_{-\infty}^{\infty} K_n(x-u)f(u)du \to f(x), \tag{2.7}$$

$$\int_{-\infty}^{\infty} [K((x-u)/b_n)]^2 f(u)du \to f(x)\int_{-\infty}^{\infty}[K(u)]^2du. \tag{2.8}$$

PROOF. Relation (2.7) follows from the Lebesgue Density Theorem (see Devroye and Györfi (1985)). Relation (2.8) also follows from the same theorem by noting that $\int_{-\infty}^{\infty}[K(u)]^2 f(u)du < \infty$. □

We will tacitly assume that Assumptions 1–2 hold throughout this section.

LEMMA 2.2. *Let $f_{\tilde{X}_j,\tilde{X}_k}$ be the joint density of $(\tilde{X}_j, \tilde{X}_k)$. Then*

$$\sup_{j\neq k}\ \sup_{(x,y)\in R\times R} |f_{\tilde{X}_j,\tilde{X}_k}(x,y) - f(x)f(y)| \le C$$

for some constant C independent of n.

PROOF. The characteristic function $\phi_{\tilde{X}_j,\tilde{X}_k}$ of $(\tilde{X}_j,\tilde{X}_k)$ is given by

$$\phi_{\tilde{X}_j,\tilde{X}_k}(u,v) = E\exp(iu\tilde{X}_j + iv\tilde{X}_k).$$

Without loss of generality, assume $j<k$. We need to consider two cases.

Case (i): $j \ge -q+k+1$. We decompose $\tilde{X}_k$ and $\tilde{X}_j$ as follows:

$$\tilde{X}_k = \sum_{r=0}^{q-1-k+j} a_{k-j+r}Z_{j-r} + \sum_{r=0}^{k-j-1} a_r Z_{k-r} + \Gamma_k.$$

$$\tilde{X}_j = \sum_{r=0}^{q-1-k+j} a_r Z_{j-r} + \sum_{r=q-k+j}^{q-1} a_r Z_{j-r} + \Gamma_j.$$

Thus

$$\begin{aligned} iu\tilde{X}_j + iv\tilde{X}_k = {} & iu\Gamma_j + iv\Gamma_k \\ & + iu\sum_{r=q-k+j}^{q-1} a_r Z_{j-r} + i\sum_{r=0}^{q-1-k+j}(ua_r + va_{k-j+r})Z_{j-r} \\ & + iv\sum_{r=0}^{k-j-1} a_r Z_{k-r}. \end{aligned}$$

The three summands on the right hand side are measurable with, respectively, the σ-fields generated by $Z_{-q+j+1},\dots,Z_{-q+k},Z_{-q+k+1},\dots,Z_j$, and $Z_{j+1},\dots,Z_k$. The summands are thus independent r.v.'s. Therefore

$$\begin{aligned}\phi_{\tilde X_j,\tilde X_k}(u,v) &= E\exp(iu\tilde X_j+iv\tilde X_k)\\ &= \phi_{\Gamma_j}(u)\phi_{\Gamma_k}(v)\prod_{r=q-k+j}^{q-1}\phi(ua_r)\\ &\quad\cdot\prod_{r=0}^{q-1-k+j}\phi(ua_r+va_{k-j+r})\prod_{r=0}^{k-j+1}\phi(va_r).\end{aligned}$$

Using Fourier inversion formula,

$$\begin{aligned}\sup_{(x,y)\in R\times R}|f_{\tilde X_j,\tilde X_k}(x,y)| &\le \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|\phi_{\tilde X_j,\tilde X_k}(u,v)|dudv\\ &\le \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|\phi(ua_0+va_{k-j}||\phi(va_0)|dudv.\end{aligned}$$

Changing variables by setting

$$\alpha = ua_0+va_{k-j},\qquad \beta=va_0,$$
$$\sup_{(x,y)\in R\times R}|f_{\tilde X_j,\tilde X_k}(x,y)| \le \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{a_0^2}|\phi(\alpha)||\phi(\beta)|d\alpha d\beta\le C.$$

Case (ii): $j<-q+k+1$. In this case $\tilde X_k$ and $\tilde X_j$ are independent and the proof follows immediately since $\tilde X_k$ and $\tilde X_j$ have density f. The joint density $f(x)f(y)$ is then also bounded. □

LEMMA 2.3. *Let $\tilde\Delta_i$ be as defined above. Then*

$$\sup_{i\neq j}\sup_{(x,y)\in R\times R}|\operatorname{Cov}\{\tilde\Delta_i(x),\tilde\Delta_j(y)\}|\le C.$$

PROOF. Using Assumption 1 and Lemma 2.2,

$$\begin{aligned}&|\operatorname{Cov}\{K_n(x-\tilde X_i),K_n(y-\tilde X_j)\}|\\ &\le \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K_n(x-u)K_n(y-v)|f_{\tilde X_i,\tilde X_j}(u,v)-f(u)f(v)|dudv\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K(u)K(v)|f_{\tilde X_i,\tilde X_j}(x-b_nu,y-b_nv)\\ &\qquad\qquad -f(x-b_nu)f(y-b_nv)|dudv\\ &\le C\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K(u)K(v)dudv=C.\end{aligned}$$ □

LEMMA 2.4. *For all* $1 \le v \le n-1$ *and all* x *and* y *in* R,

$$|\operatorname{Cov}\{\tilde{\Delta}_1(x), \tilde{\Delta}_{1+v}(y)\}| \le C b_n^{-3} \sum_{r=v}^{\infty} |a_r|.$$

PROOF. The proof of this lemma is similar to the proof of Lemma 2.2 in Tran (1992). For $v \ge 1$, define

$$X_{1+v}^* = \sum_{r=0}^{v-1} a_r Z_{1+v-r},$$

and

$$\begin{aligned} R_v &= \tilde{X}_{1+v} - X_{1+v}^* \\ &= \sum_{r=0}^{q-1} a_r Z_{1+v-r} + \Gamma_{1+v} - \sum_{r=0}^{v-1} a_r Z_{1+v-r} = \sum_{r=v}^{q-1} a_r Z_{1+v-r} + \Gamma_{1+v}. \end{aligned}$$

Thus

$$|\operatorname{Cov}\{\tilde{\Delta}_1(x), \tilde{\Delta}_{1+v}(y)\}| = |E[K_n(y - R_v - X_{1+v}^*)\{K_n(x - \tilde{X}_1) - \mu_n\}]|.$$

Clearly,

$$\begin{aligned} &\operatorname{Cov}\{\tilde{\Delta}_1(x), \tilde{\Delta}_{1+v}(y)\} \\ &\quad = E[\{K_n(x - \tilde{X}_1) - \mu_n\}\{K_n(y - R_v - X_{1+v}^*) - K_n(y - X_{1+v}^*)\}] \\ &\quad\quad + E[\{K_n(x - \tilde{X}_1) - \mu_n\}K_n(y - X_{1+v}^*)]. \end{aligned}$$

The last term equals zero by the independence of $\tilde{X}_1$ and X_{1+v}^*. Using (2.3), the Lipschitz condition satisfied by K and the boundedness of K,

$$\begin{aligned} |\operatorname{Cov}\{\tilde{\Delta}_1(x), \tilde{\Delta}_{1+v}(y)\}| &\le C b_n^{-1} E|K_n(y - R_v - X_{1+v}^*) - K_n(y - X_{1+v}^*)| \\ &\le C b_n^{-3} E|R_v| \le C b_n^{-3} \sum_{r=v}^{\infty} |a_r|. \qquad \square \end{aligned}$$

LEMMA 2.5. *Let* g_n *be the kernel density estimator defined in* (2.2). *Then*

$$\lim_{n \to \infty} n b_n \operatorname{Var}[g_n(x)] = f(x) \int_{-\infty}^{\infty} K^2(y)\,dy.$$

PROOF. Using (2.7) and (2.8) and noting that $b_n \to 0$,

$$\begin{aligned} b_n \operatorname{Var}[\tilde{\Delta}_i(x)] &= \int_{-\infty}^{\infty} \frac{1}{b_n}[K((x-y)/b_n)]^2 f(y)\,dy \\ &\quad - b_n \left[\int_{-\infty}^{\infty} \frac{1}{b_n} K((x-y)/b_n) f(y)\,dy\right]^2 \\ &\to f(x) \int_{-\infty}^{\infty} K^2(y)\,dy, \end{aligned} \tag{2.9}$$

where $\tilde{\Delta}_i(x)$ is defined in (2.5). By (2.4), (2.5) and (2.6),

$$g_n(x) - Eg_n(x) = g_n(x) - \mu_n = \frac{1}{n}\sum_{i=1}^{n}[K_n(x - \tilde{X}_i) - \mu_n] = \frac{1}{n}\sum_{i=1}^{n}\tilde{\Delta}_i(x).$$

Thus

$$nb_n \operatorname{Var}[g_n(x)] = b_n \operatorname{Var}[\tilde{\Delta}_1(x)] + 2(b_n/n) \sum_{1\le i<j\le n} \operatorname{Cov}\{\tilde{\Delta}_i(x), \tilde{\Delta}_j(x)\}.$$

By (2.9), to complete the proof of the lemma, it is sufficient to show that the second term tends to zero. Choose an arbitrary number θ with $\delta/(2+\delta) < \theta < 1$. Clearly,

$$(1-\theta)(2+\delta) > 2. \tag{2.10}$$

Define

$$m = m(n) = [b_n^{\theta-1}]. \tag{2.11}$$

Note that $m \to \infty$ since $b_n \to 0$. Using Lemmas 2.3–2.4, (2.11), (2.10) and Assumption 2,

$$\begin{aligned} &|(b_n/n) \sum_{1\le i<j\le n} \operatorname{Cov}\{\tilde{\Delta}_i(x), \tilde{\Delta}_j(x)\}| \\ &\le b_n \sum_{v=1}^{m-1} |\operatorname{Cov}\{\tilde{\Delta}_1(x), \tilde{\Delta}_{1+v}(x)\}| \\ &\quad + b_n \sum_{v=m}^{\infty} |\operatorname{Cov}\{\tilde{\Delta}_1(x), \tilde{\Delta}_{1+v}(x)\}| \\ &\le Cb_n^{\theta} + Cb_n \sum_{v=m}^{\infty} b_n^{-3} \sum_{r=v}^{\infty} |a_r| \le o(1) + Cb_n^{-2} \sum_{v=m}^{\infty} v^{-(3+\delta)} \\ &\le o(1) + Cb_n^{-2} m^{-(2+\delta)} \le o(1) + Cb_n^{-2+(1-\theta)(2+\delta)}, \end{aligned} \tag{2.12}$$

which tends to zero since $b_n \to 0$. □

The following lemma due to Masry (1986) is needed in the sequel.

LEMMA 2.6. *For distinct points x and y,*

$$q_{1n} \equiv \frac{1}{b_n}\int_{-\infty}^{\infty} K((x-u)/b_n)K((y-u)/b_n)f(u)du \to 0.$$

LEMMA 2.7. *Let $\tilde{\Delta}_i$ be as defined in* (2.5). *Then*

$$q_{2n} \equiv \frac{b_n}{n} \sum_{1 \le i \ne j \le n} |\operatorname{Cov}\{\tilde{\Delta}_i(x), \tilde{\Delta}_j(y)\}| \to 0.$$

PROOF. Let S_1 and S_2 be defined by

$$S_1 = \{(i,j) \mid i,j \in \{1,\dots,n\}, 1 \le j - i \le m\},$$
$$S_2 = \{(i,j) \mid i,j \in \{1,\dots,n\}, m+1 \le j - i \le n-1\}.$$

Then

$$\sum_{1 \le i < j \le n} |\operatorname{Cov}\{\tilde{\Delta}_i(x), \tilde{\Delta}_j(y)\}| = \sum_{S_1} |\operatorname{Cov}\{\tilde{\Delta}_i(x), \tilde{\Delta}_j(y)\}| + \sum_{S_2} |\operatorname{Cov}\{\tilde{\Delta}_i(x), \tilde{\Delta}_j(y)\}|.$$

Using the value of m in (2.11),

$$n^{-1} b_n \sum_{S_1} |\operatorname{Cov}\{\tilde{\Delta}_i(x), \tilde{\Delta}_j(y)\}| \le C n^{-1} b_n n b_n^{\theta-1} \le C b_n^{\theta} \to 0.$$

Next, by Lemma 2.4

$$\begin{aligned} n^{-1} b_n \sum_{S_2} |\operatorname{Cov}\{\tilde{\Delta}_i(x), \tilde{\Delta}_j(y)\}| &\le C n^{-1} b_n n \sum_{v=m}^{n-1} |\operatorname{Cov}\{\tilde{\Delta}_1(x), \tilde{\Delta}_{1+v}(y)\}| \\ &\le C b_n \sum_{v=m}^{\infty} |\operatorname{Cov}\{\tilde{\Delta}_1(x), \tilde{\Delta}_{1+v}(y)\}| \end{aligned}$$

which converges to zero as shown in (2.12). □

LEMMA 2.8. *Let x and y be distinct points of R. Then*

$$\lim_{n \to \infty} n b_n \operatorname{Cov}\{g_n(x), g_n(y)\} = 0.$$

PROOF. By (2.4) and (2.3),

$$\begin{aligned} &n b_n \operatorname{Cov}\{g_n(x), g_n(y)\} \\ &\quad = b_n \operatorname{Cov}\{K_n((x - \tilde{X}_1)/b_n), K_n((y - \tilde{X}_1)/b_n)\} + q_{2n} \\ &\quad = q_{1n} - b_n E K_n((x - \tilde{X}_1)/b_n) E K_n((y - \tilde{X}_1)/b_n) + q_{2n}. \end{aligned}$$

The lemma then follows by (2.7), Lemma 2.6 and Lemma 2.7. □

We will show that under certain assumptions, f_n and g_n have the same asymptotic distribution.

LEMMA 2.9. *For any $\epsilon > 0$,*

$$P[(nb_n)^{1/2}|f_n(x) - g_n(x)| > \epsilon] \to 0.$$

PROOF. Using the Lipschitz property of K stated in Assumption 1, we have

$$\begin{aligned}
&(nb_n)^{1/2}|f_n(x) - g_n(x)| \\
&\quad \le (nb_n)^{1/2}(nb_n)^{-1}\sum_{i=1}^{n}\left|K\left(\frac{x-X_i}{b_n}\right) - K\left(\frac{x-\tilde{X}_i}{b_n}\right)\right| \\
&\quad \le C(nb_n)^{-1/2}\sum_{i=1}^{n}\frac{|X_i - \tilde{X}_i|}{b_n} \\
&\quad \le Cn^{-1/2}b_n^{-3/2}\sum_{i=1}^{n}\left|\sum_{r=q}^{\infty} a_r Z_{i-r} - \Gamma_i\right|.
\end{aligned}$$

Choose

$$q = [(\log\log n)^{(3+2\delta)/(28+8\delta)}(n^2 b_n^{-3})^{1/(7+2\delta)}]. \tag{2.13}$$

Using the Chebyshev's inequality, independence, Assumption 2 and (2.1),

$$\begin{aligned}
&P[(nb_n)^{1/2}|f_n(x) - g_n(x)| > \epsilon] \\
&\quad \le P\left[\sum_{i=1}^{n}\left|\sum_{r=q}^{\infty} a_r z_{i-r} - \Gamma_i\right| > C^{-1}\epsilon n^{1/2} b_n^{3/2}\right] \\
&\quad \le \sum_{i=1}^{n} P\left[\left|\sum_{r=q}^{\infty} a_r z_{i-r} - \Gamma_i\right| > C^{-1}\epsilon n^{-1/2} b_n^{3/2}\right] \\
&\quad \le C^2\epsilon^{-2}n^2 b_n^{-3}\,\mathrm{Var}\left(\sum_{r=q}^{\infty} a_r z_{1-r} - \Gamma_1\right) \\
&\quad = 2C^2\epsilon^{-2}n^2 b_n^{-3} E Z_1^2 \sum_{r=q}^{\infty} a_r^2 \\
&\quad = O(n^2 b_n^{-3} q^{-(7+2\delta)}) = O((\log\log n)^{-(3+2\delta)/4}) = o(1).
\end{aligned} \tag{2.14}$$

□

3. Asymptotic normality: the polynomial case

We will first establish asymptotic normality of

$$(nb_n)^{1/2}(g_n(x_1) - Eg_n(x_1), \ldots, g_n(x_k) - Eg_n(x_k))',$$

for arbitrary k and arbitrary fixed points $x_1, \ldots, x_k$ in R. This result is stated in Lemma 3.3. It will be clear that it is sufficient to consider $k = 2$. To avoid using subscripts, we refer to x_1 and x_2 as x and y. By the Cramér-Wold device, it suffices to prove asymptotic normality for $c\xi_n(x) + d\xi_n(y)$ for arbitrary constants c and d, where

$$\xi_n(x) = (nb_n)^{1/2}(g_n(x) - Eg_n(x)), \qquad \xi_n(y) = (nb_n)^{1/2}(g_n(y) - Eg_n(y)). \tag{3.1}$$

By Lemmas 2.5 and 2.8,

$$\operatorname{Var}(c\xi_n(x) + d\xi_n(y)) \to (c^2 f(x) + d^2 f(y)) \int_{-\infty}^{\infty} K^2(u)du \equiv \tau^2. \tag{3.2}$$

We now proceed to prove

$$c\xi_n(x) + d\xi_n(y) \xrightarrow{d} N(0, \tau^2). \tag{3.3}$$

Let

$$Z_i(x) = b_n^{1/2}\tilde{\Delta}_i(x), \qquad Z_i(y) = b_n^{1/2}\tilde{\Delta}_i(y).$$

From (3.1), (2.4) and (2.5)

$$c\xi_n(x) + d\xi_n(y) = n^{-1/2} b_n^{1/2} \sum_{i=1}^{n} (c\tilde{\Delta}_i(x) + d\tilde{\Delta}_i(y)).$$

Choose

$$p = p(n) = [(\log\log n)^{-(3+2\delta)/(28+8\delta)}(nb_n)^{1/2}]. \tag{3.4}$$

Using the value of q in (2.13),

$$\frac{p}{q} \sim ((\log\log n)^{-1} n b_n^{(13+2\delta)/(3+2\delta)})^{(3+2\delta)/(14+4\delta)}.$$

By Assumption 3,

$$\frac{p}{q} \to \infty. \tag{3.5}$$

From (3.4),

$$\frac{p}{n} \le \frac{(\log\log n)^{-(3+2\delta)/(28+8\delta)} b_n^{1/2}}{n^{1/2}} \to 0. \tag{3.6}$$

We now set the r.v.'s $cZ_i(x) + dZ_i(y)$ into alternate large blocks of size p and small blocks of size q. Denote

$$\gamma = [n/(p+q)],$$

$$(3.7)\quad U_m = \sum_{i=\gamma_m}^{\gamma_m+p-1} (cZ_i(x) + dZ_i(y)) = b_n^{1/2} \sum_{i=\gamma_m}^{\gamma_m+p-1} (c\tilde{\Delta}_i(x) + d\tilde{\Delta}_i(y)),$$

$$U'_m = \sum_{i=\ell_m}^{\ell_m+q-1} (cZ_i(x) + dZ_i(y)) = b_n^{1/2} \sum_{i=\ell_m}^{\ell_m+q-1} (c\tilde{\Delta}_i(x) + d\tilde{\Delta}_i(y)),$$

$$U'_\gamma = \sum_{i=\gamma(p+q)+1}^{n} (cZ_i(x) + dZ_i(y)) = b_n^{1/2} \sum_{i=\gamma(p+q)+1}^{n} (c\tilde{\Delta}_i(x) + d\tilde{\Delta}_i(y)),$$

where $\gamma_m = (m-1)(p+q)+1$, $\ell_m = (m-1)(p+q)+p+1$ and $m = 1, \ldots, \gamma$. Also, set

$$(3,8)\qquad S_n = n^{1/2}[c\xi_n(x) + d\xi_n(y)] = b_n^{1/2} \sum_{i=1}^{n} (c\tilde{\Delta}_i(x) + d\tilde{\Delta}_i(y)),$$

$$(3.9)\qquad S'_n = \sum_{m=1}^{\gamma} U_m, \qquad S''_n = \sum_{m=1}^{\gamma} U'_m, \qquad S'''_n = U'_\gamma. \qquad \square$$

Clearly,

$$(3.10)\qquad S_n = S'_n + S''_n + S'''_n.$$

Here S'_n is the sum of r.v.'s in large blocks of size p, S''_n is the sum in small blocks of size q and S'''_n is the sum of left-over r.v.'s.

LEMMA 3.1. *Let S''_n and S'''_n be as defined above. Then*

$$\frac{1}{n}[E(S''_n)^2 + E(S'''_n)^2] \to 0.$$

PROOF. We will show that $n^{-1}E(S''_n)^2$ tends to zero. The proof that $n^{-1}E(S'''_n)^2 \to 0$ is similar and is omitted. By Minkowski's inequality,

$$(3.11)\qquad \left[\frac{E(S''_n)^2}{n}\right]^{1/2} \le |c| \left[\frac{1}{n} E\left(\sum_{m=1}^{\gamma} \sum_{i=\ell_m}^{\ell_m+q-1} b_n^{1/2} \tilde{\Delta}_i(x)\right)^2\right]^{1/2} + |d| \left[\frac{1}{n} E\left(\sum_{m=1}^{\gamma} \sum_{i=\ell_m}^{\ell_m+q-1} b_n^{1/2} \tilde{\Delta}_i(y)\right)^2\right]^{1/2}.$$

It is sufficient to show that the first term on the right hand side of (3.11) tends to zero; the proof that the second term tends to zero is similar. Since $\sum_{i=\ell_m}^{\ell_m+q-1} b_n^{1/2}\tilde{\Delta}_i(x)$, $m=1,\ldots,\gamma$, are independent,

$$\frac{1}{n}E\left(\sum_{m=1}^{\gamma}\sum_{i=\ell_m}^{\ell_m+q-1} b_n^{1/2}\tilde{\Delta}_i(x)\right)^2 = \frac{1}{n}\sum_{m=1}^{\gamma}\operatorname{Var}\left(\sum_{i=\ell_m}^{\ell_m+q-1} b_n^{1/2}\tilde{\Delta}_i(x)\right). \tag{3.12}$$

By Lemma 2.1,

$$\begin{aligned}\operatorname{Var}(b_n^{1/2}\tilde{\Delta}_i(x)) &= \operatorname{Var}(b_n^{1/2}\Delta_i(x)) \\ &= b_n E[K_n(x-X_i)]^2 \\ &= \int_{-\infty}^{\infty}\frac{1}{b_n}[K((x-u)/b_n)]^2 f(u)du \\ &\to f(x)\int_{\infty}^{\infty}[K(u)]^2du.\end{aligned} \tag{3.13}$$

Employing (3.13), the right hand side of (3.12) equals

$$\begin{aligned}&\frac{\gamma q}{n}\operatorname{Var}(b_n^{1/2}\tilde{\Delta}_1(x)) + \frac{2}{n}\sum_{m=1}^{\gamma}\sum_{\ell_m\le i<j\le \ell_m+q-1}\operatorname{Cov}\{b_n^{1/2}\tilde{\Delta}_i(x), b_n^{1/2}\tilde{\Delta}_j(x)\} \\ &\le Cq\gamma/n + \frac{b_n}{n}\sum_{1\le i<j\le n}|\operatorname{Cov}\{\tilde{\Delta}_i(x),\tilde{\Delta}_j(x)\}|.\end{aligned} \tag{3.14}$$

Using (3.5),

$$\frac{q\gamma}{n} = \frac{qp\gamma}{pn} \le \frac{q}{p} \to 0.$$

The last term of (3.14) tends to zero by (2.12). It is now easy to deduce from (3.11) to (3.14) that $n^{-1}E(S_n'')^2$ tends to zero. □

LEMMA 3.2. *Let*

$$s_n^2 \equiv \sum_{m=1}^{\gamma}\operatorname{Var}(n^{-1/2}U_m).$$

Then $s_n^2 \to \tau^2$.

PROOF. From equation (3.9),

$$\sum_{m=1}^{\gamma}\operatorname{Var}(n^{-1/2}U_m) = \frac{1}{n}E(S_n')^2 \tag{3.15}$$

because $U_1,\ldots,U_\gamma$ are independent r.v.'s. It is easily seen that,

$$\frac{E(S_n')^2}{n} = \frac{ES_n^2}{n} + \frac{E(S_n''+S_n''')^2}{n} - \frac{2E[S_n(S_n''+S_n''')]}{n}. \tag{3.16}$$

Lemma 3.1 implies

$$\left[\frac{E(S_n'' + S_n''')^2}{n}\right]^{1/2} \le \left[\frac{E(S_n'')^2}{n}\right]^{1/2} + \left[\frac{E(S_n''')^2}{n}\right]^{1/2} \to 0. \tag{3.17}$$

By (3.2) and (3.8),

$$\frac{ES_n^2}{n} \to \tau^2. \tag{3.18}$$

The lemma follows from (3.15), (3.16), (3.17) and (3.18). □

LEMMA 3.3. *Let g_n be as defined in* (2.2). *Then*

$$(nb_n)^{1/2}(g_n(x_1) - Eg_n(x_1), \ldots, g_n(x_k) - Eg_n(x_k))' \xrightarrow{d} N(\mathbf{0}, \boldsymbol{C}),$$

where $\boldsymbol{C}$ is a diagonal matrix with diagonal elements $C_{ii} = f(x_i)\int_{-\infty}^{\infty} K^2(u)du$, $i = 1, \ldots, k$.

PROOF. By the definition of s_n^2 in Lemma 3.2,

$$\sum_{m=1}^{\gamma} \operatorname{Var}\left(n^{-1/2}\frac{U_m}{s_n}\right) = 1.$$

The r.v.'s $\{U_m : 1 \le m \le \gamma\}$ are independent since the process $\{\tilde{\Delta}_i\}$ is q-dependent. By the Lindeberg Central Limit Theorem,

$$\sum_{m=1}^{\gamma} n^{-1/2}\frac{U_m}{s_n} \xrightarrow{d} N(0,1) \tag{3.19}$$

if for every $\epsilon > 0$

$$\sum_{m=1}^{\gamma} \int_{(|x| \ge \epsilon)} x^2 dF_m \to 0, \tag{3.20}$$

where F_m is the distribution function of $n^{-1/2}U_m/s_n$.

From (3.7)

$$|U_m| \le Cpb_n^{-1/2} \qquad \text{a.s.},$$

since K is bounded. Therefore

$$\frac{n^{-1/2}|U_m|}{s_n} \le \frac{Cp}{s_n\sqrt{nb_n}} = \frac{C}{s_n}\left(\frac{p^2}{nb_n}\right)^{1/2} \qquad \text{a.s.}$$

A simple computation using the definition of p in (3.4) and Lemma 3.2 shows that

$$P[n^{-1/2}|U_m|/s_n > \epsilon] = 0$$

for sufficiently large n. The left hand side of (3.20) is thus zero for large n and (3.19) follows. By (3.9) and (3.10),

$$\frac{S_n - (S_n'' + S_n''')}{s_n\sqrt{n}} \xrightarrow{d} N(0,1). \tag{3.21}$$

Employing Lemmas 3.1 and 3.2 together with the definition of τ in (3.2),

$$\frac{S_n'' + S_n'''}{s_n\sqrt{n}} \xrightarrow{P} 0. \tag{3.22}$$

The proof of (3.3) is completed by (3.21), (3.22), (3.8) and Lemma 3.2. □

THEOREM 3.1. *Suppose Assumptions* 1–3 *hold and* $x_1, \ldots, x_k$ *are* k *distinct points of* R. *Then*

$$(nb_n)^{1/2}(f_n(x_1) - Ef_n(x_1), \ldots, f_n(x_k) - Ef_n(x_k))' \xrightarrow{d} N(\mathbf{0}, \boldsymbol{C}).$$

PROOF. Define

$$\eta_n(x) = (nb_n)^{1/2}(f_n(x) - Ef_n(x)), \qquad \eta_n(y) = (nb_n)^{1/2}(f_n(y) - Ef_n(y)).$$

Then

$$\begin{aligned} c\eta_n(x) + d\eta_n(y) = c\xi_n(x) + d\xi_n(y) &+ c(nb_n)^{1/2}(f_n(x) - g_n(x)) \\ &+ d(nb_n)^{1/2}(g_n(y) - g_n(y)). \end{aligned}$$

By Lemma 2.9,

$$c(nb_n)^{1/2}(f_n(x) - g_n(x)) \xrightarrow{P} 0, \qquad c(nb_n)^{1/2}(f_n(y) - g_n(y)) \xrightarrow{P} 0.$$

Employing (3.3),

$$c\eta_n(x) + d\eta_n(y) \xrightarrow{d} N(0, \tau^2).$$

The theorem follows by the Cramér-Wold device. □

ASSUMPTION 4. For some $C > 0$ and any $x, y \in R$,

$$|f(x) - f(y)| \le C|x - y|.$$

ASSUMPTION 5. The bandwidth b_n tends to zero slowly enough that

$$nb_n^3 \to \infty.$$

THEOREM 3.2. *If Assumptions* 1–5 *hold and in addition,*

$$\int_{-\infty}^{\infty} |x|K(x)dx < \infty,$$

then for any k and any distinct points $x_1, \dots, x_k$,

$$(nb_n)^{1/2}(f_n(x_1) - f(x_1), \dots, f_n(x_k) - f(x_k))' \xrightarrow{d} N(\mathbf{0}, \boldsymbol{C}).$$

PROOF. Again, assume $k = 2$. We have

$$\begin{aligned}(3.23)\quad & c(nb_n)^{1/2}(f_n(x) - f(y)) + d(nb_n)^{1/2}(f_n(y) - f(y)) \\ & \qquad - [c(nb_n)^{1/2}(f_n(x) - Ef_n(x)) + d(nb_n)^{1/2}(f_n(y) - Ef_n(y))] \\ & \quad = (nb_n)^{1/2}[c(Ef_n(x) - f(x)) + d(Ef_n(y) - f(y))].\end{aligned}$$

By a simple computation,

$$\begin{aligned}(3.24)\quad |Ef_n(x) - f(x)| &= \left| \int_{-\infty}^{\infty} K(z) f(x - b_n z) dz - \int_{-\infty}^{\infty} K(z) f(x) dz \right| \\ &\le Cb_n \int_{-\infty}^{\infty} |z| K(z) dz \le Cb_n.\end{aligned}$$

Similarly,

$$(3.25)\qquad |Ef_n(y) - f(y)| \le Cb_n.$$

By Theorem 3.1,

$$(3.26)\qquad c(nb_n)^{1/2}(f_n(x) - Ef_n(x)) + d(nb_n)^{1/2}(f_n(y) - Ef_n(y)) \xrightarrow{d} N(0, \tau^2).$$

Employing Assumption 5, we easily obtain from (3.23), (3.24) and (3.25),

$$c(nb_n)^{1/2}(f_n(x) - f(x)) + d(nb_n)^{1/2}(f_n(y) - f(y)) \xrightarrow{d} N(0, \tau^2),$$

since $nb_n^3 \to 0$. The theorem follows by the Cramér-Wold device. □

Remark 3.1. If $\delta > 1$ and Assumption 5 is satisfied, then $b_n > n^{-1/3}$ for large n. Then

$$\frac{nb_n^{(13+2\delta)/(3+2\delta)}}{\log \log n} = \frac{nb_n^3 b_n^{(4-4\delta)/(3+2\delta)}}{\log \log n} \ge \frac{Cnb_n^3 n^{(1/3)(4\delta-4)/(3+2\delta)}}{\log \log n} \to \infty.$$

Thus Assumption 5 implies Assumption 3 when $\delta > 1$.

If $\delta < 1$ and Assumption 3 is satisfied, then

$$nb_n^3 = \left(\frac{nb_n^{(13+2\delta)/(3+2\delta)}}{\log \log n} \right) \left(\frac{\log \log n}{b_n^{(4-4\delta)/(3+2\delta)}} \right) \to \infty.$$

Thus Assumption 3 implies Assumption 5 when $\delta < 1$.

4. Asymptotic normality: the exponential case

ASSUMPTION 2′. Suppose that the coefficients of the linear process X_t tend to zero sufficiently fast that $|a_r| = O(e^{-sr})$ for some $r > 0$ as $r \to \infty$. In addition, Z_1 has mean zero and finite variance and an absolutely integrable characteristic function ϕ.

ASSUMPTION 3′. Suppose $b_n \to 0$ in such a manner that

$$\frac{nb_n}{\log\log n(\log n)^2} \to \infty.$$

Note that Assumption 5 implies Assumption 3′ when $\delta < 1$.

THEOREM 4.1. *If Assumptions* 1, 2′, 3′ *hold, then*

$$(nb_n)^{1/2}(f_n(x_1) - Ef_n(x_1), \dots, f_n(x_k) - Ef_n(x_k))' \xrightarrow{d} N(\mathbf{0}, \boldsymbol{C})$$

for any k and any distinct points $x_1, \dots, x_k$ of R.

PROOF. We will obtain the theorem by making certain modifications to the proof of Theorem 3.2 where Assumptions 2 and 3 are employed.

Choose

$$q = 2\alpha \log n - 3\alpha \log b_n = \log(n^{2\alpha} b_n^{-3\alpha}) \tag{4.1}$$

for some $\alpha > 1/(2s)$. Turning now to (2.14), we have with the help of Assumption 2′

$$\begin{aligned} &P[(nb_n)^{1/2}|f_n(x) - g_n(x)| > \epsilon] \\ &\quad \le Cn^2 b_n^{-3} \sum_{r=q}^{\infty} a_r^2 \le Cn^2 b_n^{-3} e^{-2sq} \\ &\quad \le Cn^2 b_n^{-3} \exp(-2s\alpha \log(n^2 b_n^{-3})) \\ &\quad \le C(n^2 b_n^{-3})^{-2s\alpha+1} = o(1), \end{aligned} \tag{4.2}$$

since $n^2 b_n^{-3} \to \infty$ and $-2s\alpha + 1 < 0$.

Choose

$$p = (\log\log n)^{-1/2}(nb_n)^{1/2}.$$

Then

$$\frac{p}{q} = \alpha^{-1}(nb_n(\log\log n)^{-1}(\log(nb_n^{-3}))^{-2})^{1/2}.$$

Since $b_n \to 0$, we have $\log(nb_n^{-3}) > \log n$ for large n. Therefore Assumption 3′ warrants that

$$\frac{p}{q} \to \infty. \tag{4.3}$$

The proof of Theorem 4.1 can be obtained from the proof of Theorem 3.1 using (4.2) and (4.3) to make adequate changes. □

THEOREM 4.2. *If Assumptions* 1, 2′, 4, 5 *hold and in addition,*

$$\int_{-\infty}^{\infty} |x|K(x)dx < \infty,$$

then

$$(nb_n)^{1/2}(f_n(x_1) - f(x_1), \ldots, f_n(x_k) - f(x_k))' \xrightarrow{d} N(\mathbf{0}, \boldsymbol{C})$$

for any k and any distinct points $x_1, \ldots, x_k$ of R.

PROOF. The theorem follows easily by employing Theorem 4.1 and the proof of Theorem 3.2. Note, however that Assumption 3′ is not needed since it is implied by Assumption 5. To see this, let us suppose that Assumption 5 holds. Then $b_n > n^{-1/3}$ for large n and Assumption 3′ holds because $n^{2/3}$ goes to infinity at a faster rate than $\log\log n(\log n)^2$. □

5. Limiting covariance, MSE of f_n and optimal bandwidth

Under general assumptions, we next compute the limiting covariance matrix of $(f_n(x_1), \ldots, f_n(x_k))'$ for distinct points $x_1, \ldots, x_k$. This result is then employed to obtain the optimal bandwidth in the sense that it asymptotically minimizes the mean square error of f_n.

LEMMA 5.1. *Let f_{X_j,X_k} be the joint density of (X_j, X_k). Then*

$$\sup_{j\neq k} \sup_{(x,y)\in R\times R} |f_{X_j,X_k}(x,y) - f(x)f(y)| \leq C$$

for some constant C independent of n.

PROOF. Decompose X_k as follows:

$$X_k = \sum_{r=0}^{\infty} a_{k-j+r} Z_{j-r} + \sum_{r=0}^{k-j-1} a_r Z_{k-r}.$$

Then

$$iuX_j + ivX_k = i\sum_{r=0}^{\infty}(ua_r + va_{k-j+r})Z_{j-r} + iv\sum_{r=0}^{k-j-1} a_r Z_{k-r}.$$

Using independence and the Fourier inversion formula, the proof of the lemma can be completed by the same argument as that of Lemma 2.2. □

LEMMA 5.2. *For all $1 \leq v \leq n-1$ and all x and y in R,*

$$|\operatorname{Cov}\{\Delta_1(x), \Delta_{1+v}(y)\}| \leq C b_n^{-3} \sum_{r=v}^{\infty} |a_r|.$$

PROOF. For $v \geq 1$, define

$$\tilde{R}_v = X_{1+v} - X^*_{1+v} = \sum_{r=v}^{\infty} a_r Z_{1+v-r}.$$

Then

$$|\operatorname{Cov}\{\Delta_1(x), \Delta_{1+v}(y)\}| = |E[K_n(y - \tilde{R}_v - X^*_{1+v})\{K_n(x - X_1) - \mu_n\}]|.$$

The proof of the lemma can now be completed using the same line of argument in Lemma 2.4. □

THEOREM 5.1. *Suppose Assumptions* 1, 2 *hold or Assumptions* 1, 2′ *hold. Let* $\boldsymbol{C}_n$ *be the covariance matrix of the random vector* $(f_n(x_1), \ldots, f_n(x_k))'$, *where* $x_1, \ldots, x_k$ *are* k *arbitrary points in* R. *Then* $\lim_{n\to\infty} nb_n \boldsymbol{C}_n = \boldsymbol{C}$.

PROOF. We will consider the polynomial case where a_r tends to zero at the rate $r^{-(4+\delta)}$. The proof of the exponential case under Assumptions 1, 2′ is similar. Without loss of generality, we assume $k = 2$. To complete the proof of the theorem, it is sufficient to show that

(i) $\lim_{n\to\infty} nb_n \operatorname{Var}[f_n(x)] = \int_{-\infty}^{\infty} K^2(y)dy$,

(ii) $\lim_{n\to\infty} nb_n \operatorname{Cov}\{f_n(x), f_n(y)\} = 0$.

The proof of (i) can be obtained by a slight variation of the proof of Lemma 2.5. Replace $\tilde{\Delta}_i$, g_n, $\tilde{X}_i$ by Δ_i, f_n, and X_i respectively and use Lemma 5.1 and Lemma 5.2 to make necessary changes.

The proof of (ii) follows by making some slight changes in the proof of Lemma 2.8 with the help of Lemma 5.1 and Lemma 5.2. □

Assume that K satisfies

$$\int_{-\infty}^{\infty} yK(y)dy = 0, \qquad \int_{-\infty}^{\infty} y^2|K(y)|dy < \infty \tag{5.1}$$

and $f''(x)$ exists. Following the same line of arguments as Parzen (1962),

$$\frac{Ef_n(x) - f(x)}{b_n^2} \to -\frac{1}{2}f''(x)\int_{-\infty}^{\infty} y^2K(y)dy.$$

Consequently,

$$E[f_n(x) - f(x)]^2 \sim \frac{f(x)}{nb_n}\int_{-\infty}^{\infty} K^2(y)dy + \frac{b_n^4}{4}\left[\int_{-\infty}^{\infty} y^2K(y)dy\right]^2 f''(x).$$

The value of b_n which asymptotically minimizes the mean square error is

$$b_n = \frac{f(x)\int_{-\infty}^{\infty} K^2(y)dy}{[n(f''(x)\int_{-\infty}^{\infty} y^2K(y)dy)^2]^{1/5}}. \tag{5.2}$$

Then, the mean square error tends to zero as $n^{-4/5}$, as in the independent case. Summarizing, we have

THEOREM 5.2. *Suppose K satisfies* (5.1) *and Assumptions* 1, 2 *or* 1, 2′ *hold. Then the bandwidth given in* (5.2) *asymptotically minimizes the mean square error.*

A natural method in bandwidth selection arising from (5.2) is to choose b_n with respect to some standard family of densities. For more details see Section 3.4.2 of Silverman (1986).

Acknowledgements

We would like to thank the referee for his useful comments.

REFERENCES

Chan, N. H. and Tran, L. T. (1991). Nonparametric tests for serial dependence, *J. Time Ser. Anal.*, **13**, 19–28.

Chanda, K. C. (1983). Density estimation for linear processes, *Ann. Inst. Statist. Math.*, **35**, 439–446.

Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation: The L_1 View*, Wiley, New York.

Gorodetskii, V. V. (1977). On the strong mixing properties for linear sequences, *Theory Probab. Appl.*, **22**, 411–413.

Györfi, L., Härdle, W. Sarda, P. and Vieu, P. (1989). Nonparametric curve estimation from time series, *Lecture Notes in Statist.*, **60**, Springer, New York.

Helson, H. and Sarason. D. (1990). Past and future, *Math. Scand.*, **21**, 5–16.

Ibragimov, I. A. and Rozanov, Yu. V. (1978). *Gaussian Random Processes*, Springer, New York.

Ioannides, D. and Roussas, G. G. (1987). Note on uniform convergence for mixing random variables, *Statist. Probab. Lett.*, **5**, 279–285.

Masry, E. (1986). Recursive probability density estimation for weakly dependent stationary processes, *IEEE Trans. Inform. Theory*, IT**18**, 254–267.

Masry, E. (1987). Almost sure convergence of recursive density estimators for stationary mixing processes, *Statist. Probab. Lett.*, **5**, 249–254.

Masry, E. and Györfi, L. (1987). Strong consistency and rates for recursive density estimators for stationary mixing processes, *J. Multivariate Anal.*, **22**, 79–93.

Parzen, E. (1962). On estimation of a probability density function and mode, *Ann. Math. Statist.*, **33**, 1065–1076.

Robinson, P. M. (1983). Nonparametric estimators for time series, *J. Time Ser. Anal.*, **4**, 185–297.

Robinson, P. M. (1986). On the consistency and finite sample properties of nonparametric kernel time series regression, autoregression and density estimators, *Ann. Inst. Statist. Math.*, **38**, 539–549.

Robinson, P. M. (1987). Time series residuals with application to probability density estimation, *J. Time Ser. Anal.*, **3**, 329–344.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function, *Ann. Math. Statist.*, **27**, 832–837.

Roussas, G. G. (1988). Nonparametric estimation in mixing sequences of random variables, *J. Statist. Plann. Inference*, **18**, 135–149.

Roussas, G. G. (1990). Nonparametric regression estimation under mixing conditions, *Stochastic Process. Appl.*, **36**, 107–116.

Sarason, D. (1972). An addendum to "Past and future", *Math. Scand.*, **30**, 62–64.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Tran, L. T. (1989). Recursive density estimation under dependence, *IEEE Trans. Inform. Theory*, **35**, 1103–1108.

Tran, L. T. (1992). Kernel density estimation for linear processes, *Stochastic Process. Appl.*, **41**, 281–296.

Withers, C. S. (1981). Conditions for linear processes to be strong mixing, *Z. Wahrsch. Verw. Gebiete*, **57**, 477–480.

Yakowitz, S. (1985). Nonparametric density estimation, prediction and regression for Markov sequences, *J. Amer. Statist. Assoc.*, **80**, 215–221.

Yakowitz, S. (1987). Nearest-neighbor methods for time series analysis, *J. Time Ser. Anal.*, **8**, 235–247.