Ann. Inst. Statist. Math.
Vol. 52, No. 4, 599–611 (2000)

NONPARAMETRIC DENSITY ESTIMATION FOR A LONG-RANGE DEPENDENT LINEAR PROCESS

TOSHIO HONDA

Institute of Social Sciences, University of Tsukuba Tsukuba, Ibaraki 305-8571, JAPAN

(Received February 17, 1999; revised August 12, 1999)

Abstract. We estimate the marginal density function of a long-range dependent linear process by the kernel estimator. We assume the innovations are i.i.d. Then it is known that the term of the sample mean is dominant in the MISE of the kernel density estimator when the dependence is beyond some level which depends on the bandwidth and that the MISE has asymptotically the same form as for i.i.d. observations when the dependence is below the level. We call the latter the case where the dependence is not very strong and focus on it in this paper. We show that the asymptotic distribution of the kernel density estimator is the same as for i.i.d. observations and the effect of long-range dependence does not appear. In addition we describe some results for weakly dependent linear processes.

Key words and phrases: Kernel density estimator, long-range dependence, linear process, bandwidth, asymptotic normality.

1. Introduction

Nonparametric density estimators have been studied extensively. However, most of attention was directed to i.i.d. and weakly dependent observations before 1990. The asymptotic distribution of the kernel estimator is the same for both of them. See Silverman (1986) for i.i.d. observations, Györfi *et al.* (1989) for weakly dependent observations, Hallin and Tran (1996) for weakly dependent linear processes, and Castellana and Leadbetter (1986) and Liebscher (1996) for strongly mixing processes.

Statistical inference for long-range dependent observations has gained a lot of attention for recent years because many time series data exhibit long-range dependence, especially in hydrology and economics, an the theories for weakly dependent observations do not apply to long-range dependent observations. See Beran (1994) for surveys of this subject.

Although it is hard to study the problems of statistical inference for long-range dependent observations, asymptotic theories for empirical processes of long-range dependent linear processes and robust estimation in a linear model with errors from a long-range dependent linear process have been developed by several authors, e.g. Ho and Hsing (1996, 1997), Giraitis *et al.* (1996). Their results are useful for the purpose here. Particularly we often cite Giraitis *et al.* (1996), which we refer to as GKS in this paper.

As for kernel estimation of the marginal density function of long-range dependent observations, Robinson (1991), Cheng and Robinson (1991), and Ho (1996) examined the cases of nonlinear transformations of long-range dependent Gaussian processes. Ho (1996) showed that when the dependence is not very strong, the asymptotic distribution is the same as for i.i.d. observations. His method is based on proving convergence of moments and the proof is rather complicated. He also proved that the kernel estimator behaves as the sample mean when the dependence is very strong. Hall *et al.* (1995) considered the selection of bandwidth.

Hall and Hart (1990) gave the asymptotic representation of the MISE of the kernel density estimator for long-range dependent linear processes. In the case of very strong dependence, Ho and Hsing (1996) showed in Theorem 3.1 that the effect of the long-range dependence appears and the kernel density estimator behaves as the sample mean. Some theorems on the asymptotic distribution of the kernel density estimator for long-range dependent linear processes are also given in Hidalgo (1997). It does not seem that his use of Appell polynomial expansions are validated by the assumptions there.

We consider estimation of the marginal density function of a long-range dependent linear process with i.i.d. innovations. Since the case of very strong dependence is already considered in Ho and Hsing (1996) and so on, we focus on the case where the dependence is not very strong. We show the asymptotic distribution of the kernel estimator is the same as for i.i.d. observations. Theorem 2.1 here corresponds to Theorems 1 and 3 in Ho (1996), which deals with nonlinear transformations of long-range dependent Gaussian processes. Since the method here also applies to the cases of weakly dependent linear processes which are not covered in Hallin and Tran (1996), we describe some results for them in Theorem 2.2 below.

Note that the results here are out of the scope of those on strongly mixing processes because the well-known sufficient conditions for linear processes to be strongly mixing require summability of the coefficients $\{a_i\}$ in (1.1) below and it seems impossible to verify the asymptotic normality with slowly decaying mixing coefficients. Tran *et al.* (1996) also discussed some limitations of asymptotic theories based on the properties of strongly mixing processes and considered another nonparametric estimation problem on linear processes. See Doukhan (1994) for the definition and the properties of strongly mixing processes.

In this paper we deal with a real-valued long-range dependent linear process

$$X_i = \sum_{j=0}^{\infty} a_j \epsilon_{i-j}, \tag{1.1}$$

where $a_0 = 1$ and $a_j = j^{-(1+\theta)/2}L(j)(0 < \theta < 1)$, $j = 1, 2, \ldots$, and L is a slowly varying function. Let $\{\epsilon_i\}$ be an i.i.d. process with mean 0 and variance 1. It is well known that $E\{X_0 X_j\} \asymp j^{-\theta}L^2(j)$, where $b_n \asymp c_n$ means $b_n = O(c_n)$ and $c_n = O(b_n)$. Denoting the characteristic function of ϵ_i by ϕ, we assume

$$|\phi(u)| \le C(1 + |u|)^{-\delta}, \tag{1.2}$$

where $\delta > 0$ and $C > 0$. Besides in order to avoid inessential complications, we assume the existence of the moments of ϵ_i of sufficiently high order. Hereafter C and C_i stand for generic positive constants and the values differ from place to place.

We denote the density and distribution function of X_i by f and F, respectively. Then we estimate $f(x)$ for a fixed $x \in R$ by the kernel estimator

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right), \tag{1.3}$$

where K is the symmetric kernel function with compact support and of bounded variation and h is the bandwidth tending to 0. Note that K must satisfy $\int_R K(u)du = 1$.

The representation of the MISE of the kernel density estimator in Hall and Hart (1990) implies that the effect of the long-range dependence is not seen when $hn^{1-\theta} L^2(n) \to 0$. On the other hand the term of the sample mean is dominant when $hn^{1-\theta}L^2(n) \to \infty$. We concentrate on the case of $hn^{1-\theta}L^2(n) \to 0$, which we call the case where the dependence is not very strong.

Since the method here applies to the cases of $\theta \geq 1$ in (1.1), we formulate the assumptions on θ and h as follows:

$$\theta < 1 : hn^{1-\theta}L^2(n) \to 0 \quad \text{and} \quad \frac{L^2(n)}{n^\theta h^2} \to 0, \tag{1.4}$$

$$\theta = 1 : C_1 n^{-\delta_1} < h < C_2 n^{-\delta_2} \quad \text{for some} \quad 0 < \delta_2 < \delta_1 < \frac{1}{2}, \tag{1.5}$$

$$\theta > 1 : h \to 0 \quad \text{and} \quad \frac{1}{nh^2} \to 0. \tag{1.6}$$

The latter of (1.4) is added to prove Lemma 2.1. It is not restrictive at all. For example, we usually take $h = Cn^{-1/5}$ for second order kernels and this is optimal. Then $\theta \in (4/5, 1)$ satisfies the former of (1.4). The latter is satisfied, too.

When $\theta < 4/5$, the term of the sample mean is dominant in the representation of the MISE of Hall and Hart (1990) and the asymptotic distribution is considered in Ho and Hsing (1996). Then $\hat{f}(x) - E\hat{f}(x)$ is asymptotically equivalent to $-f'(x)\bar{X}$.

In Section 2, the main theorems and the proofs are given. Two technical lemmas are deferred to Section 3.

2. Asymptotic normality of the estimator

We take l distinct points $x_1, x_2, \ldots, x_l \in R$ and estimate the value of the density function f at each point. Then the asymptotic simultaneous distribution of the kernel density estimators under (1.1) with $\theta < 1$ is given in the following theorem. The asymptotic distribution is the same as for i.i.d. observations.

THEOREM 2.1. *We assume* (1.1), (1.2), *and* (1.4) *and we estimate* $f(x)$ *by* (1.3). *Then* $(nh)^{1/2}(\hat{f}(x_1) - E\hat{f}(x_1), \hat{f}(x_2) - E\hat{f}(x_2), \ldots, \hat{f}(x_l) - E\hat{f}(x_l))'$

$$\xrightarrow{D} N\left(0, \int_R K^2(u)du \,\mathrm{diag}\{f(x_1), f(x_2), \ldots, f(x_l)\}\right),$$

where $\xrightarrow{D}$ *and* $N(\cdot,\cdot)$ *stand for convergence in distribution and the normal distribution, respectively and* $\mathrm{diag}\{d_1, \ldots, d_l\}$ *means an* $l \times l$ *diagonal matrix having* d_i *as the* (i,i) *element.*

The following theorem deals with the cases of $\theta \geq 1$ in (1.1). We can prove it in almost the same way as Theorem 2.1. The necessary modifications are given at the end of this section.

THEOREM 2.2. *We assume* (1.1) *with* $\theta \geq 1$ *and* (1.2). *In addition* (1.5) *or* (1.6) *is assumed according to the value of* θ. *Then we have the same result as Theorem* 2.1.

Remark 2.1. From (1.2) and the argument of Lemma 1 of GKS, f is sufficiently smooth. Hence we have

$$(nh)^{1/2}(E\hat{f}(x) - f(x)) \to \frac{C^2 f''(x)}{2} \int_R u^2 K(u)du \quad \text{for} \quad h = Cn^{-1/5}.$$

Remark 2.2. We can establish the same theorems as Theorems 2.1–2.2 when we estimate the marginal density function of $\{\tilde{X}_i\}$, where $\tilde{X}_i = G(X_i)$. We assume that G is monotone increasing and continuously differentiable around $G^{-1}(x_i)$. In addition $G'(G^{-1}(x_i)) > 0$. Note that G^{-1} stands for the inverse function of G. We describe the necessary modifications in the proof after we introduce several notations.

Before proving Theorem 2.1, we introduce several notations. Letting $g(u)$ be some bounded function of bounded variation, we have by intergation by parts

$$\int_R K\left(\frac{u-x}{h}\right) dg(u) = -\int_R g(u) dK\left(\frac{u-x}{h}\right). \tag{2.1}$$

For notational convenience, we often write the right-hand side of (2.1) as

$$\int_R g(u) dK_u.$$

The domain of integration is often omitted. In addition we prove the theorem with $l = 1$ for simplicity of presentation. We show in the proof

$$(nh)^{1/2} Z(x) = (nh)^{1/2}(\hat{f}(x) - E\hat{f}(x)) \xrightarrow{\mathcal{D}} N\left(0, f(x)\int K^2(u)du\right). \tag{2.2}$$

When $l > 1$, we should replace $K((u-x)/h)$ and $K((v-x)/h)$ in the proof with $\sum_{i=1}^{l} \alpha_i K((u-x_i)/h)$ and $\sum_{i=1}^{l} \alpha_i K((v-x_i)/h)$.

By integration by parts, we can rewrite $Z(x)$ in (2.2) as

$$Z(x) = -\frac{1}{nh}\int \sum_{i=1}^{n} \xi_i(u) dK\left(\frac{u-x}{h}\right) = -\frac{1}{nh}\int \sum_{i=1}^{n} \xi_i(u) dK_u, \tag{2.3}$$

where $\xi_i(u) = I(X_i \le u) - F(u)$.

Following Ho and Hsing (1996), we define the martingale decomposition of $\xi_i(u)$. Let $\mathcal{S}_i$ be the σ-field generated by $\{\epsilon_i, \epsilon_{i-1}, \ldots\}$. Then we have

$$\xi_i(u) = \sum_{j=1}^{\infty} \xi_{i,j}(u), \tag{2.4}$$

where $\xi_{i,j+1} = E\{\xi_i(u) \mid \mathcal{S}_{i-j}\} - E\{\xi_i(u) \mid \mathcal{S}_{i-j-1}\}$, $j = 0, 1, 2, \ldots$.

Writing F_k for the distribution function of $\sum_{j=0}^{k-1} a_j \epsilon_{i-j}$, we can represent $\xi_{i,j}(u)$ as

$$\xi_{i,j}(u) = F_{j-1}(u - \bar{X}_{i,j-1}) - F_j(u - \bar{X}_{i,j}),$$

where $\bar{X}_{i,j} = \sum_{l=j}^{\infty} a_l \epsilon_{i-l}$. Using the above expression, we have

$$\sum_{j=1}^{k} \xi_{i,j}(u) = I(X_i \le u) - F_k(u - \bar{X}_{i,k}) \tag{2.5}$$

$$\xi_i(u) - \sum_{j=1}^{k} \xi_{i,j}(u) = F_k(u - \bar{X}_{i,k}) - F(u). \tag{2.6}$$

For sufficiently large k, F_k has the bounded density. See Lemma 1 of GKS.

Here we describe the modifications for Remark 2.2. Since

$$I(\tilde{X}_i \leq u) = I(X_i \leq G^{-1}(u)),$$

we have only to make the following replacements in the proof of Theorem 2.1: In $F(\cdot)$, $F_k(\cdot)$, $\xi_{i,j}(\cdot)$, and $I(\cdot)$,

$$u \to G^{-1}(u), \qquad v \to G^{-1}(v), \quad \text{and} \quad x \to G^{-1}(x).$$

Note that u, v, and x in $K((u-x)/h)$ and $K((v-x)/h)$ remain unchanged.

PROOF OF THEOREM 2.1. We prove (2.2). Employing (2.3)–(2.6), we can rewrite $Z(x)$ as

$$\begin{aligned}
(2.7) \qquad -Z(x) &= \frac{1}{nh}\int \sum_{i=1}^{n}\sum_{j=1}^{k} \xi_{i,j}(u)dK\left(\frac{u-x}{h}\right) \\
&\quad + \frac{1}{nh}\int \sum_{i=1}^{n}\sum_{j=k+1}^{\infty} \xi_{i,j}(u)dK\left(\frac{u-x}{h}\right) \\
(2.8) \qquad &= \frac{1}{nh}\int \sum_{i=1}^{n}(I(X_i \leq u) - F_k(u - \bar{X}_{i,k}))dK\left(\frac{u-x}{h}\right) \\
&\quad + \frac{1}{nh}\int \sum_{i=1}^{n}(F_k(u - \bar{X}_{i,k}) - F(u))dK\left(\frac{u-x}{h}\right) \\
(2.9) \qquad &= Z_k(x) + R_k(x),
\end{aligned}$$

where $Z_k(x)$ and $R_k(x)$ are clearly defined.

If (2.10) and (2.11) below are proved for some fixed sufficiently large k, (2.2) follows from (2.9).

$$(2.10) \qquad (nh)^{1/2} Z_k(x) \xrightarrow{D} N\left(0, f(x)\int K^2(u)du\right).$$

$$(2.11) \qquad \lim_{n\to\infty} E\{nhR_k^2(x)\} = 0.$$

We prove (2.11) in Lemma 2.3 below. Note again that k is sufficiently large and fixed. We omit x of $Z_k(x)$ and $R_k(x)$ hereafter.

We verify (2.10) using the martingale central limit theorem. See Theorem 9.5.2 of Chow and Teicher (1988). Modifying Z_k by addition and elimination of some negligible terms and writing $\tilde{Z}_k$ for the modified Z_k, we have from (2.7)

$$(2.12) \quad (nh)^{1/2}\tilde{Z}_k = (nh)^{-1/2}\sum_{i=1}^{n}\int(\xi_{i+1,1}(u)+\xi_{i+2,2}(u)+\cdots+\xi_{i+k,k}(u))dK\left(\frac{u-x}{h}\right).$$

By the definition of $\xi_{i,j}(u)$, we can apply the martingale central limit theorem.

Since the summand is bounded and $nh \to \infty$, we have only to show that the conditional variance tends to $f(x)\int K^2(u)du$ in probability. The conditional variance of (2.12) is written as

$$(2.13) \qquad \frac{1}{nh}\sum_{i=1}^{n}\iint\sum_{j=1}^{k}\sum_{l=1}^{k} E\{\xi_{i+j,j}(u)\xi_{i+l,l}(v) \mid \mathcal{S}_i\}dK\left(\frac{u-x}{h}\right)dK\left(\frac{v-x}{h}\right).$$

The proof of the convergence in probability of (2.13) consists of (2.14)–(2.16) below. The proof of (2.14) is deferred to Lemma 2.1.

$$(2.14)\quad (2.13) - \frac{1}{h}\iint \sum_{j=1}^{k}\sum_{l=1}^{k} E\{\xi_{1+j,j}(u)\xi_{1+l,l}(v)\} \times dK\left(\frac{u-x}{h}\right) dK\left(\frac{v-x}{h}\right) \\ \to 0 \text{ in probability.}$$

Since (2.12) is negligibly different from the martingale decomposition of $(nh)^{1/2}Z_k$, we have

$$(2.15)\quad \frac{1}{h}\iint \sum_{j=1}^{k}\sum_{l=1}^{k} E\{\xi_{1+j,j}(u)\xi_{1+l,l}(v)\} dK\left(\frac{u-x}{h}\right) dK\left(\frac{v-x}{h}\right) \\ = E\{nhZ_k^2\} + o(1).$$

Thus the convergence of the conditional variance follows from (2.14) and

$$(2.16)\quad E\{nhZ_k^2\} \to f(x)\int K^2(u)du.$$

(2.16) follows from (2.9), (2.11), and Lemma 2.2 below. Thus (2.10) is verified and the proof of Theorem 2.1 is complete. □

LEMMA 2.1.

$$(2.17)\quad \frac{1}{nh}\sum_{i=1}^{n}\iint \sum_{j=1}^{k}\sum_{l=1}^{k}[E\{\xi_{i+j,j}(u)\xi_{i+l,l}(v) \mid \mathcal{S}_i\} - E\{\xi_{i+j,j}(u)\xi_{i+l,l}(v)\}] \\ \cdot dK\left(\frac{u-x}{h}\right) dK\left(\frac{v-x}{h}\right) \to 0 \quad \textit{in probability.}$$

PROOF. We prove the lemma by evaluating the variance. Notice that

$$(2.18)\quad E\{\xi_{i+j,j}(u)\xi_{i+l,l}(v) \mid \mathcal{S}_i\} - E\{\xi_{i+j,j}(u)\xi_{i+l,l}(v)\}$$

is the function of $\sum_{p=j}^{\infty} a_p\epsilon_{i+j-p}$ and $\sum_{q=l}^{\infty} a_q\epsilon_{i+l-q}$. Put these to Y_j and Z_j in Lemma 3.1 with j replacing i. The assumptions of the lemma hold by the definition of $\{a_i\}$. Write $H_{j,l}(Y_i, Z_i, u, v)$ for (2.18) and note that $H_{j,l}$ are uniformly bounded. If $j = l$, Lemma 2 of GKS is employed. The lemma deals with the case of only Y_j and Y_{j+t} in Lemma 3.1 and gives the same estimate of p_t.

Consider the case of $|i-m| > t_d$, where t_d comes from Lemma 3.1. When $|i-m| \le t_d$, (2.19) below is uniformly bounded.

Putting $t = i - m$ and exploiting Lemmas 3.1–3.2, we have

$$(2.19)\quad E\{H_{j,l}(Y_i, Z_i, u_1, v_1)H_{j,l}(Y_m, Z_m, u_2, v_2)\} \\ = \iiiint H_{j,l}(y_0, z_0, u_1, v_1)H_{j,l}(y_t, z_t, u_2, v_2)g_t(y_0, z_0, y_t, z_t)dy_0dz_0dy_tdz_t \\ = \iiiint H_{j,l}(y_0, z_0, u_1, v_1)H_{j,l}(y_t, z_t, u_2, v_2)p_t(y_0, z_0, y_t, z_t)dy_0dz_0dy_tdz_t \\ + O(|t|^{-\theta}L^2(|t|)).$$

Therefore for $|t| > t_d$,

$$|(2.19)| \leq \iiiint |p_t| dy_0 dz_0 dy_t dz_t + O(|t|^{-\theta} L^2(|t|)). \tag{2.20}$$

Following GKS, we evaluate the integral of the right-hand side of (2.20). Take $0 < \rho < 1$ and $\eta > 0$, where ρ is specified later and η is arbitrary, and define

$$A_- = \{(y_0, z_0, y_t, z_t)' \mid |p_t|^\rho < \{(1+|y_0|)(1+|z_0|)(1+|y_t|)(1+|z_t|)\}^{-(1+\eta)}\},$$
$$A_+ = R^4 - A_-.$$

Then we have

$$\iiiint_{A_-} |p_t|^{1-\rho} |p_t|^\rho dy_0 dz_0 dy_t dz_t \leq C|t|^{-(\theta+d)(1-\rho)}. \tag{2.21}$$

Setting $m^* = (1+\eta)(1-\rho)/\rho$, we have on A_+,

$$|p_t|^{\rho-1} \leq \{(1+|y_0|)(1+|z_0|)(1+|y_t|)(1+|z_t|)\}^{m^*}. \tag{2.22}$$

By (2.22), we get

$$\begin{aligned} &\iiiint_{A_+} |p_t|^{1-\rho} |p_t| |p_t|^{\rho-1} dy_0 dz_0 dy_t dz_t \\ &\leq C t^{-(\theta+d)(1-\rho)} \iiiint |p_t| \\ &\quad \times \{(1+|y_0|)(1+|z_0|)(1+|y_t|)(1+|z_t|)\}^{m^*} dy_0 dz_0 dy_t dz_t. \end{aligned} \tag{2.23}$$

We evaluate the last expression of (2.23) by replacing p_t with the definition of p_t in Lemma 3.1. The contribution from $g_t(y_0, z_0, y_t, z_t)$ is bounded above by

$$\begin{aligned} &\iiiint |g_t| \{(1+|y_0|)(1+|z_0|)(1+|y_t|)(1+|z_t|)\}^{m^*} dy_0 dz_0 dy_t dz_t \\ &\leq \iint g(y_0, z_0) \{(1+|y_0|)(1+|z_0|)\}^{2m^*} dy_0 dz_0 < \infty. \end{aligned} \tag{2.24}$$

Here we used Lemma 3.2. The contributions from the other terms can be evaluated similarly. Thus

$$|(2.23)| \leq C|t|^{-(\theta+d)(1-\rho)}. \tag{2.25}$$

We choose d and ρ such that $(\theta+d)(1-\rho) > \theta$. Then combining (2.20), (2.21), (2.23), and (2.25), we obtain

$$|(2.19)| \leq C|t|^{-\theta} L^2(|t|). \tag{2.26}$$

Since

$$\iiiint d\left|K\left(\frac{u_1 - x}{h}\right)\right| d\left|K\left(\frac{v_1 - x}{h}\right)\right| d\left|K\left(\frac{u_2 - x}{h}\right)\right| d\left|K\left(\frac{v_2 - x}{h}\right)\right| < \infty,$$

we have from (1.4) and (2.26),

$$\text{Variance of (2.17)} \leq \frac{C}{n^2h^2}\left(n + \sum_{|i-m|>t_d} |i-m|^{-\theta}L^2(|i-m|)\right) \leq C\frac{L^2(n)}{n^\theta h^2} \to 0.$$

The proof of the lemma is complete. □

LEMMA 2.2.

$$nhE\{Z^2(x)\} \to f(x)\int K^2(u)du.$$

PROOF. Write f_t for the simultaneous density of X_1 and X_{1+t}. From Lemmas 3.1–3.2 and Lemma 2 of GKS, f_t exists for any t and they are uniformly bounded.

Since

$$\sqrt{nh}Z(x) = \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\left(K\left(\frac{X_i - x}{h}\right) - EK\left(\frac{X_i - x}{h}\right)\right),$$

we have

$$\begin{aligned}(2.27)\qquad & E\{nhZ^2(x)\} - \frac{1}{h}E\left\{\left(K\left(\frac{X_1-x}{h}\right) - EK\left(\frac{X_1-x}{h}\right)\right)^2\right\} \\ &= \frac{1}{nh}\sum_{1\leq i\neq j\leq n}\iint K\left(\frac{x_1-x}{h}\right)K\left(\frac{x_2-x}{h}\right) \\ &\quad \cdot (f_{i-j}(x_1,x_2) - f(x_1)f(x_2))dx_1dx_2.\end{aligned}$$

From Lemma 2 of GKS, when $|i-j| > M$ for some $M > 0$,

$$(2.28)\qquad |f_{i-j}(x_1,x_2) - f(x_1)f(x_2)| \leq C|i-j|^{-\theta}L^2(|i-j|) \text{ uniformly in } x_1 \text{ and } x_2.$$

When $0 < |i-j| \leq M$,

$$(2.29)\qquad |f_{i-j}(x_1,x_2) - f(x_1)f(x_2)| \leq C \text{ uniformly in } x_1 \text{ and } x_2.$$

By (2.27)–(2.29), we have

$$(2.30)\qquad (2.27) = O(h) + O(n^{1-\theta}hL^2(n)).$$

The lemma follows from (1.4) and (2.30). □

LEMMA 2.3. *For some large k, we have*

$$\lim_{n\to\infty} nhE\{R_k^2(x)\} \to 0.$$

PROOF. By the definition and integration by parts, we get

$$\begin{aligned}(2.31)\qquad & \sqrt{nh}R_k(x) \\ &= \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\int(F_k(u - \bar{X}_{i,k}) - F(u))dK\left(\frac{u-x}{h}\right) \\ &= \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\int(f(u) - f_k(u - \bar{X}_{i,k}))K\left(\frac{u-x}{h}\right)du.\end{aligned}$$

As in the proof of Lemma 2.2, we can show

$$(2.32)\quad nhE\{R_k^2(x)\} = \frac{1}{h}E\left\{\iint (f(u) - f_k(u - \bar{X}_{1,k}))(f(v) - f_k(v - \bar{X}_{1,k})) \times K\left(\frac{u-x}{h}\right)K\left(\frac{v-x}{h}\right)dudv\right\} + o(1).$$

Since f_k and f are bounded and

$$\frac{1}{h}\iint \left|K\left(\frac{u-x}{h}\right)K\left(\frac{v-x}{h}\right)\right|dudv = O(h),$$

we have

$$nhE\{R_k^2(x)\} = o(1). \qquad \square$$

PROOF OF THEOREM 2.2. We can show that for $\theta \geq 1$, the estimates of p_t in Lemma 2 of GKS and Lemma 3.1 here become $O(t^{-(1+\theta)/2+\eta})$, where η is an arbitrary positive number. When $\theta > 1$, we can take η such that $(1+\theta)/2 - \eta > 1$. Then in the proof of Lemma 2.1, we should replace $(\theta + d)$ with $(1+\theta)/2 - \eta$ and choose ρ satisfying $((1+\theta)/2 - \eta)(1-\rho) > 1$. In (2.26), $|t|^{-\theta}L^2(|t|)$ is replaced with $|t|^{-((1+\theta)/2-\eta)(1-\rho)}$. Then

$$\text{Variance of (2.17)} \leq C\frac{1}{nh^2}.$$

Since the bound of (2.28) becomes $|i-j|^{-(1+\theta)/2+\eta}$ in Lemma 2.2, we have (2.27) $= O(h)$ in (2.30). The case of $\theta = 1$ can be treated similarly. □

3. Technical lemmas

In this section we give two technical lemmas which are used in the proof of Theorem 2.1. Lemma 3.1 is an extension of Lemma 2 of GKS.

LEMMA 3.1. *Define*

$$Y_j = \sum_{k=0}^{\infty} b_k\epsilon_{j-k} = \sum_{k=-\infty}^{\infty} b_{j-k}\epsilon_k \quad \text{and} \quad Z_j = \sum_{k=0}^{\infty} c_k\epsilon_{j-k} = \sum_{k=-\infty}^{\infty} c_{j-k}\epsilon_k,$$

where $b_k \asymp k^{-(1+\theta)/2}L(k)$ *and* $c_k \asymp k^{-(1+\theta)/2}L(k)$, $k = 0, 1, 2, \ldots,$ *and* $b_k = c_k = 0$, $k < 0$. *Denote* $(b_k, c_k)'$ *by* d_k. *Write* g *and* g_t *for the density functions of* $(Y_j, Z_j)'$ *and* $(Y_j, Z_j, Y_{j+t}, Z_{j+t})'$, *respectively. Then we have*

$$r_Y(t) = E\{Y_jY_{j+t}\}, \quad r_Z(t) = E\{Z_jZ_{j+t}\}, \quad \text{and}$$
$$r_{YZ}(t) = E\{Y_jZ_{j+t}\} \asymp |t|^{-\theta}L^2(|t|).$$

Refer to Mielniczuk (1997) *for these estimates. Define* $p_t(y_0, z_0, y_t, z_t)$ *by*

$$\begin{aligned} p_t(y_0, z_0, y_t, z_t) &= g_t(y_0, z_0, y_t, z_t) - g(y_0, z_0)g(y_t, z_t) - r_Y(t)\frac{\partial g}{\partial y}(y_0, z_0)\frac{\partial g}{\partial y}(y_t, z_t) \\ &\quad - r_{YZ}(t)\frac{\partial g}{\partial y}(y_0, z_0)\frac{\partial g}{\partial z}(y_t, z_t) - r_{YZ}(-t)\frac{\partial g}{\partial z}(y_0, z_0)\frac{\partial g}{\partial y}(y_t, z_t) \\ &\quad - r_Z(t)\frac{\partial g}{\partial z}(y_0, z_0)\frac{\partial g}{\partial z}(y_t, z_t). \end{aligned}$$

We fix some $d\,(0 < d < ((1-\theta)/18) \wedge (\theta/9))$ *and assume that for some* m *satisfying* $\delta m > 5 + (\theta + d)/d$, *where* δ *comes from* (1.2), *we can choose* $\{\{j_s, k_s\}\}_{s=1}^{m}$ *such that*

$$\begin{vmatrix} b_{j_s} & c_{j_s} \\ b_{k_s} & c_{k_s} \end{vmatrix} \neq 0 \quad and \quad \{j_s, k_s\} \cap \{j_u, k_u\} = \phi \quad for \ \ s \neq u.$$

Then when $t > t_d$, *where* t_d *depends on* d, *we have*

$$p_t(y_0, z_0, y_t, z_t) = O(t^{-\theta-d}) \qquad uniformlyin \quad y_0, z_0, y_t, z_t.$$

PROOF. At first we consider the simultaneous density function of Y_j and Z_j. Set $u = (u_1, u_2)'$. The characteristic function of $(Y_j, Z_j)'$ is represented as

$$\hat{g}(u_1, u_2) = \prod_{j=-\infty}^{\infty} \phi(d_j' u).$$

From (1.2),

$$|\phi(d_i' u)\phi(d_j' u)| \leq C_1(1 + |d_i' u|)^{-\delta}(1 + |d_j' u|)^{-\delta}.$$

From elementary calculus, we have

$$(1 + |d_{j_s}' u|)(1 + |d_{k_s}' u|) \geq C_2[1 + \{(d_{j_s}' u)^2 + (d_{k_s}' u)^2\}^{1/2}] \geq \delta_s(1 + |u|),$$

where $|u| = (u'u)^{1/2}$ and δ_s depends on the matrix

$$\begin{pmatrix} b_{j_s} & c_{j_s} \\ b_{k_s} & c_{k_s} \end{pmatrix}. \tag{3.1}$$

Thus we get

$$|\phi(d_j' u)\phi(d_k' u)| \leq C_3 \delta_s^{-1}(1 + |u|)^{-\delta}. \tag{3.2}$$

From (3.2), we have

$$|\hat{g}(u_1, u_2)| \leq C_4(1 + |u|)^{-\delta m}. \tag{3.3}$$

(3.3) implies that the density function g and the derivatives of the first order exist and that they are bounded and continuous.

Next we consider the simultaneous density function of $(Y_j, Z_j, Y_{j+t}, Z_{j+t})'$. Set $v = (v_1, v_2)'$. The characteristic function is represented as

$$\hat{g}_t(u, v) = E\{e^{i(u_1 Y_j + u_2 Z_j + v_1 Y_{j+t} + v_2 Y_{j+t})}\} = \prod_{k=-\infty}^{\infty} \phi(u' d_{-k} + v' d_{t-k}).$$

For any $\{j_s, k_s\}$, there exists t_s such that when $t > t_s$, the minimum singular value of the matrix

$$\begin{pmatrix} b_{j_s} & c_{j_s} & b_{t+j_s} & c_{t+j_s} \\ b_{k_s} & c_{k_s} & b_{t+k_s} & c_{t+k_s} \\ b_{j_s-t} & c_{j_s-t} & b_{j_s} & c_{j_s} \\ b_{k_s-t} & c_{k_s-t} & b_{k_s} & c_{k_s} \end{pmatrix}$$

is bounded below by half of the minimum singular value of (3.1). Thus proceeding as before, we get

$$|\phi(u'd_{j_s}+v'd_{t+j_s})\phi(u'd_{k_s}+v'd_{t+k_s})\times\phi(u'd_{j_s-t}+v'd_{j_s})\phi(u'd_{k_s-t}+v'd_{k_s})| \\ \leq C\delta_s^{-2}\{1+(|u|^2+|v|^2)^{1/2}\}^{-\delta}.$$

Thus

$$|\hat{g}_t(u,v)|\leq C\{1+(|u|^2+|v|^2)^{1/2}\}^{-\delta m}. \tag{3.4}$$

(3.4) implies that the density function g_t exists and that it is bounded and continuous.

Finally we evaluate $p_t(y_0, z_0, y_t, z_t)$. Following GKS, we treat the inversion of the characteristic function. Then we divide the domain of the integration into $\{|u|^2+|v|^2\geq t^{2d}\}$ and $\{|u|^2+|v|^2<t^{2d}\}$.

Using (3.3) and (3.4), we obtain

$$\int_{|u|^2+|v|^2\geq t^{2d}} e^{-i(y_0u_1+z_0u_2+y_tv_1+z_tv_2)}\hat{p}_t du dv = O(t^{-\theta-d}).$$

By replacing u_1b_{-k} and u_2b_{t-k} with $u'd_{-k}$ and $v'd_{t-k}$, we can show

$$\sup_{|u|^2+|v|^2<t^{2d}}|\hat{p}_t(u,v)| = O(t^{-\theta-5d})$$

in the same way as in Lemma 2 of GKS. Then the integration over $\{|u|^2+|v|^2<t^{2d}\}$ is $O(t^{-\theta-d})$.

The proof of the lemma is complete. □

Lemma 3.2 is an extension of Lemma 1 of GKS. Although we assume that ϵ_i has the moments of sufficiently large order in this paper, we present the lemma in a rather general form.

LEMMA 3.2. *Consider Y_j and Z_j defined in Lemma 3.1 and replace $\delta m>5+(\theta+d)/d$ with $\delta m>2l+5$ there. In addition assume that (1.2) holds and that $E\{|\epsilon_1|^l\}<\infty$ for $l>2$. Then for any positive integers p and q such that $p+q=l$, we have the following inequalities:*

$$|g(x_1,x_2)|\leq C(1+|x_1|)^{-p}(1+|x_2|)^{-q+1}\wedge(1+|x_1|)^{-p+1}(1+|x_2|)^{-q}$$

and

$$\left|\frac{\partial g}{\partial x_i}(x_1,x_2)\right| = |g_j(x_1,x_2)|\leq C(1+|x_1|)^{-p}(1+|x_2|)^{-q}.$$

PROOF. We have to prove

$$\left|\frac{\partial^{s+t}}{\partial u_1^s\partial u_2^t}(u_j\hat{g}(u))\right|\leq C_1(1+|u_1|)^{-2}(1+|u_2|)^{-2},\qquad \text{for}\quad 0\leq s+t\leq l, \tag{3.5}$$

where $\hat{g}(u)$ is the characteristic function of $(Y_j, Z_j)'$.

Indeed if (3.5) is true, we have

$$\int e^{-ix'u} \frac{\partial^{s+t}}{\partial u_1^s \partial u_2^t}(u_j \hat{g}(u))du = (ix_1)^s (ix_2)^t \int e^{-ix'u} u_j \hat{g}(u)du$$
$$= C_{s,t} x_1^s x_2^t \frac{\partial g}{\partial x_j}(x),$$

where $x = (x_1, x_2)'$ and $C_{s,t}$ depends on s and t.

Since the derivatives of the first order are bounded and continuous, we have

$$\left|\frac{\partial g}{\partial x_j}(x)\right| \le C_2(1 + |x_1|^s |x_2|^t)^{-1}.$$

These inequalities yield the inequalities of the lemma.

We consider (3.5). We can represent $(Y_j, Z_j)'$ as

$$\begin{pmatrix} Y_j \\ Z_j \end{pmatrix} = \sum_{k=0}^{\infty} d_k \epsilon_{j-k} = \sum_{k=0}^{\infty} \begin{pmatrix} b_k \\ c_k \end{pmatrix} \epsilon_{j-k}.$$

We have as in Lemma 3.1,

$$\hat{g}(u) = \prod_{k=0}^{\infty} \phi(d_k' u),$$
$$\frac{\partial \hat{g}}{\partial u_1}(u) = \sum_{k=0}^{\infty} \phi'(d_k' u) b_k \prod_{j \ne k} \phi(d_j' u),$$

and

$$|\phi'(d_k' u)| \le C_3 |d_k' u|.$$

Hence estimating $\prod_{j\ne k} \phi(d_j' u)$ as in Lemma 3.1, we can prove (3.5) in the same way as in Lemma 1 of GKS. □

Acknowledgements

The author is grateful to the referees for their helpful comments. Especially one of the referees advised the author to consider the cases of $\theta \ge 1$ and suggested an expression which simplified the proof of Lemma 2.3.

References

Beran, J. (1994). *Statistics for Long-memory Processes*, Chapman & Hall, New York.

Castellana, J. V. and Leadbetter, M. R. (1986). On smoothed probability density estimation for stationary processes, *Stochastic Process. Appl.*, **21**, 179–193.

Cheng, B. and Robinson, P. M. (1991). Density estimation in strongly dependent nonlinear time series, *Statist. Sinica*, **1**, 335–359.

Chow, Y. S. and Teicher, H. (1988). *Probability Theory*, 2nd ed., Springer, New York.

Doukhan, P. (1994). *Mixing: Properties and Examples*, Springer, New York.

Giraitis, L., Koul, H. L., and Surgailis, D. (1996). Asymptotic normality of regression estimators with long memory errors, *Statist. Probab. Lett.*, **29**, 317–335.

Györfi, L., Härdle, W., Sarda, P. and Vieu, P. (1989). *Nonparametric Curve Estimation from Time Series*, Springer, New York.

Hall, P. and Hart, J. D. (1990). Convergence rates for density estimation for data from infinite order moving average sequences, *Probab. Theory Related Fields*, **87**, 253–274.

Hall, P., Lahiri, S. N. and Truong, Y. K. (1995). On bandwidth choice for density estimation with dependent data, *Ann. Statist.*, **23**, 2241–2263.

Hallin, M. and Tran, L. T. (1996). Kernel density estimation for linear processes: asymptotic normality and optimal bandwidth derivation, *Ann. Inst. Statist. Math.*, **48**, 429–449.

Hidalgo, J. (1997). Non-parametric estimation with strongly dependent multivariate time series, *J. Time Ser. Anal.*, **18**, 95–122.

Ho, H.-C. (1996). On central and non-central limit theorems in density estimation for sequences of long-range dependence, *Stochastic Processes Appl.*, **63**, 153–174.

Ho, H.-C. and Hsing, T. (1996). On the asymptotic expansion of the empirical process of long-memory moving averages, *Ann. Statist.*, **24**, 992–1024.

Ho, H.-C. and Hsing, T. (1997). Limit theorems for functionals of moving averages, *Ann. Probab.*, **25**, 1636–1669.

Liebscher, E. (1996). Strong convergence of sums of α-mixing random variables with applications to density estimation, *Stochastic Process. Appl.*, **65**, 69–80.

Mielniczuk, J. (1997). Long and short-range dependent sums of infinite-order moving averages and regression estimation, *Acta Sci. Math.*, **63**, 301–316.

Robinson, P. M. (1991). Nonparametric functional estimation for long-memory time series, *Nonparametric and Semiparametric Methods in Econometrics and Statistics* (eds. W. A. Barnett, J. Powell and G. Tauchen), 437–457, Cambridge University Press, Cambridge.

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

Tran, L., Roussas, G., Yakowitz, S., and Van, B. T. (1996) Fixed-design regression for linear time series, *Ann. Statist.*, **24**, 975–991.