RESEARCH

Open Access

CrossMark

The asymptotic normality of internal estimator for nonparametric regression

Penghua Li[1], Xiaoqin Li[2]* and Liping Chen[3]

*Correspondence:
lixiaoqin1983@163.com
[2]School of Mathematical Sciences,
Anhui University, Hefei, China
Full list of author information is
available at the end of the article

Abstract

In this paper, we aim to study the asymptotic properties of internal estimator of nonparametric regression with independent and dependent data. Under some weak conditions, we present some results on asymptotic normality of the estimator. Our results extend some corresponding ones.

MSC: 62G08; 62G20

Keywords: Asymptotic normality; Nonparametric regression; Internal estimator; Dependent data

1 Introduction

In this paper, we consider the nonparametric regression model

$$Y_i = m(X_i) + U_i, \quad 1 \le i \le n, n \ge 1,$$

where $(X_i, Y_i) \in R^d \times R, d \ge 1$, and U_i are random variables satisfying $E(U_i|X_i) = 0, 1 \le i \le n, n \ge 1$. So we have

$$E(Y_i|X_i = x) = m(x), \quad i \ge 1.$$

Let $K(x)$ be a kernel function. Define $K_h(x) = h^{-d}K(x/h)$, where $h = h_n$ is a sequence of positive bandwidths tending to zero as $n \to \infty$. Kernel-type estimators of the regression function are widely used in various situations because of their flexibility and efficiency in the dependent and independent data. For the independent data, Nadaraya [1] and Watson [2] gave the most popular nonparametric estimator of the unknown function $m(x)$ named the Nadaraya–Watson estimator $\widehat{m}_{NW}(x)$:

$$\widehat{m}_{NW}(x) = \frac{\sum_{i=1}^n Y_i K_h(x - X_i)}{\sum_{i=1}^n K_h(x - X_i)}. \tag{1.1}$$

Jones et al. [3] considered various versions of kernel-type regression estimators such as the Nadaraya–Watson estimator (1.1) and the local linear estimator. They also investigated the internal estimator

$$\widehat{m}_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{Y_i K_h(x - X_i)}{f(X_i)} \tag{1.2}$$

for a known density $f(\cdot)$. Here the factor $\frac{1}{f(X_i)}$ is internal to the summation, whereas the estimator $\widehat{m}_{\mathrm{NW}}(x)$ has the factor $\frac{1}{\widehat{f}(x)} = \frac{1}{n^{-1}\sum_{i=1}^{n} K_h(x-X_i)}$ externally to the summation.

The internal estimator was first proposed by Mack and Müller [4]. Jones et al. [3] studied various kernel-type regression estimators, including the introduced internal estimator (1.2). Linton and Nielsen [5] introduced an integration method based on direct integration of initial pilot estimator (1.2). Linton and Jacho-Chávez [6] studied the other internal estimator

$$\widetilde{m}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i K_h(x - X_i)}{\widehat{f}(X_i)}, \tag{1.3}$$

where $\widehat{f}(X_i) = \frac{1}{n}\sum_{j=1}^{n} L_b(X_i - X_j)$ and $L_b(\cdot) = L(\cdot/b)/b^d$. Here $L(\cdot)$ is a kernel function, b is the bandwidth, and the density $f(\cdot)$ is unknown. Under the independent data, Linton and Jacho-Chávez [6] obtained the asymptotic normality of the internal estimator $\widetilde{m}_n(x)$ in (1.3). Shen and Xie [7] obtained the complete convergence and uniform complete convergence of internal estimator $\widehat{m}_n(x)$ in (1.2) under the geometrical α-mixing (or strong mixing) data. Li et al. [8] weakened the conditions of Shen and Xie [7] and obtained the convergence rate and uniform convergence rate for the estimator $\widehat{m}_n(x)$ in probability.

As far as we know, there are no results on asymptotic normality of the internal estimator $\widehat{m}_n(x)$. Similarly to Linton and Jacho-Chávez [6], we investigate the asymptotic normality of the internal estimator $\widehat{m}_n(x)$ with independent data and φ-mixing data, respectively. Asymptotic normality results are presented in Sect. 3.

Denote $\mathcal{F}_n^m = \sigma(X_i, n \leq i \leq m)$ and define the coefficients

$$\varphi(n) = \sup_{m \geq 1} \sup_{A \in \mathcal{F}_1^m, B \in \mathcal{F}_{m+n}^{\infty}, P(A) \neq 0} \left| P(B|A) - P(B) \right|.$$

If $\varphi(n) \downarrow 0$ as $n \to \infty$, then $\{X_n\}_{n \geq 1}$ is said to be a φ-mixing sequence.

The concept of φ-mixing is introduced by Dobrushin [9], and many properties of φ-mixing are presented in Chap. 4 of Billingsley [10]. If the coefficient of the process is geometrically decreasing, then the autoregressive moving average (ARMA) process can construct a geometric φ-mixing sequence. Györfi et al. [11, 12] gave more examples and applications to nonparametric estimation. We can also refer to Fan and Yao [13] and Bosq and Blanke [14] for the works on nonparametric regression under independent and dependent data.

Regarding notation, for $x = (x_1, \ldots, x_d) \in R^d$, set $\|x\| = \max(|x_1|, \ldots, |x_d|)$. Throughout the paper, $c, c_1, c_2, c_3, \ldots, d, B_0, B_1$ denote some positive constants not depending on n, which may be different in various places, $\lfloor x \rfloor$ denotes the largest integer not exceeding x, $\to$ means to take the limit as $n \to \infty$, and $c_n \sim d_n$ means that $\frac{c_n}{d_n} \to 1$, $\xrightarrow{\mathscr{D}}$ means the convergence in distribution, and $X \overset{\mathscr{D}}{=} Y$ means that random variables X and Y have the same distribution. A sequence $\{X_i, i \geq 1\}$ is said to be second-order stationary if $(X_1, X_{1+k}) \overset{\mathscr{D}}{=} (X_i, X_{i+k})$ for $i \geq 1, k \geq 1$.

2 Some assumptions
In this section, we list some assumptions.

Assumption 2.1 There exist two positive constants $\bar{K} > 0$ and $\mu > 0$ such that

$$\sup_{u \in R^d} |K(u)| \leq \bar{K} \quad \text{and} \quad \int_{R^d} |K(u)| \, du = \mu. \tag{2.1}$$

Assumption 2.2 Let S_f denote the compact support of known density $f(\cdot)$ of X_1. For $x \in S_f$, the function $m(x)$ is twice differentiable, and there exists a positive constant b such that

$$\left| \frac{\partial^2 m(x)}{\partial x_i \, \partial x_j} \right| \leq b, \quad \forall i, j = 1, 2, \ldots, d.$$

The kernel density function is symmetric and satisfies

$$\int_{R^d} |v_i| |v_j| K(v) \, dv < \infty, \quad \forall i, j = 1, 2, \ldots, d.$$

Assumption 2.3 We assume the data observed $\{(X_i, Y_i), i \geq 1\}$ is an independent and identically distributed stochastic sequence with values in $R^d \times R$. The known density $f(\cdot)$ of X_1 is upon its compact support S_f and such that $\inf_{x \in S_f} f(x) > 0$. For $0 < \delta \leq 1$, we suppose that

$$E|Y_1|^{2+\delta} < \infty \tag{2.2}$$

and

$$\sup_{x \in S_f} E\left(|Y_1|^{2+\delta} | X_1 = x \right) f(x) \leq B_0 < \infty. \tag{2.3}$$

Assumption 2.3* We assume that the data observed $\{(X_i, Y_i), i \geq 1\}$ is a second-order stationary stochastic sequence with values in $R^d \times R$. The sequence $\{(X_i, Y_i), i \geq 1\}$ is also assumed to be φ-mixing with $\sum_{n=1}^{\infty} \varphi^{1/2}(n) < \infty$. The known density $f(\cdot)$ of X_1 is upon its compact support S_f and such that $\inf_{x \in S_f} f(x) > 0$. Let (2.2) and (2.3) be fulfilled. Moreover, for all $j \geq 1$, we have

$$\sup_{x_1 \in S_f, x_{j+1} \in S_f} E\left(|Y_1 Y_{j+1}| | X_1 = x_1, X_{j+1} = x_{j+1} \right) f_j(x_1, x_{j+1}) \leq B_1 < \infty, \tag{2.4}$$

where $f_j(x_1, x_{j+1})$ denotes the joint density of (X_1, X_{j+1}).

Remark 2.1 Assumption 2.1 is a usual condition on the kernel function, and Assumption 2.2 is used to get the convergence rate of $|E\widehat{m}_n(x) - m(x)|$. Assumptions 2.3 and 2.3* are the conditions of independent and dependent data $\{(X_i, Y_i), i \geq 1\}$, respectively. Similarly to Hansen [15], conditions (2.2) and (2.3) are used to control the tail behavior of the conditional expectation $E(|Y_1|^{2+\delta}|X_1 = x)$, and (2.4) is used to estimate the covariance $\text{Cov}(Y_1, Y_{j+1})$.

3 Asymptotic normality of internal estimator $\widehat{m}_n(x)$ with independent and dependent data

In this section, we show some results on asymptotic normality of the internal estimator of a nonparametric regression model with independent and dependent data. Theorem 3.1 is for independent data, and Theorem 3.2 is for φ-mixing data.

Theorem 3.1 *Let Assumptions* 2.1–2.3 *hold, and let* $\lim_{\|u\|\to\infty} \|u\|^d K(u) = 0$. *Suppose that* $\frac{E(Y_1^2|X_1=x)}{f(x)}$ *is positive and continuous at point* $x \in S_f$. *If* $0 < h^d \to 0$, $nh^d \to \infty$, *and* $nh^{d+4} \to 0$ *as* $n \to \infty$, *then*

$$\sqrt{nh^d}\big[\widehat{m}_n(x) - m(x)\big] \xrightarrow{\mathscr{D}} N\big(0, \sigma^2(x)\big), \tag{3.1}$$

where $\sigma^2(x) = \frac{E(Y_1^2|X_1=x)}{f(x)} \int_{R^d} K^2(u)\,du$.

Theorem 3.2 *Let the conditions of Theorem* 3.1 *be fulfilled, where Assumption* 2.3 *is replaced by Assumption* 2.3*. Then* (3.1) *holds.*

Remark 3.1 The choice of a positive bandwidth h is easy to design. For example, with $d \geq 1$, if $h = n^{-\beta}$ and $\beta \in (\frac{1}{d+4}, \frac{1}{d})$, then the conditions $0 < h^d \to 0$, $nh^d \to \infty$, and $nh^{d+4} \to 0$ are satisfied as $n \to \infty$.

4 Conclusion

Linton and Jacho-Chávez [6] obtained some asymptotic normality results of the internal estimator $\widetilde{m}_n(x)$ under independent data. Comparing Theorem 1 and Corollary 1 of Linton and Jacho-Chávez [6], our asymptotic normality results on the internal estimator $\widehat{m}_n(x)$ in Theorems 3.1 and 3.2 are relatively simple. Meanwhile, we use the method of Bernstein's big-block and small-block and the inequalities of φ-mixing random variables to investigate the asymptotic normality of the internal estimator $\widehat{m}_n(x)$ for $m(x)$, and we also obtain the asymptotic normality result of (3.1). Obviously, α-mixing is weaker than φ-mixing, but some moment inequalities of α-mixing are more complicated than those of φ-mixing [16, 17]. For simplicity, we study the asymptotic normality of internal estimator $\widehat{m}_n(x)$ under φ-mixing and obtain the asymptotic normality result of Theorem 3.2.

5 Some lemmas and the proofs of main results

Lemma 5.1 (Liptser and Shiryayev [18], Theorem 9 in Sect. 5) *Let* $(\xi_{nk}, \mathscr{H}_k^n)_{k\geq 1}$ *be martingale differences (i.e.* $\mathscr{H}_0^n = \{\emptyset, \Omega\}$, $\mathscr{H}_k^n \subset \mathscr{H}_{k+1}^n$, ξ_{nk} *is an* $\mathscr{H}_k^n$-*measurable random variable,* $E(\xi_{nk}|\mathscr{H}_{k-1}^n) = 0$ *a.s., for all* $k \geq 1$ *and* $n \geq 1$) *with* $E\xi_{nk}^2 < \infty$ *for all* $k \geq 1$ *and* $n \geq 1$. *Let* $(\gamma_n)_{n\geq 1}$ *be a sequence of Markov times with respect to* $(\mathscr{H}_k^n)_{k\geq 0}$, *taking values in the set* $\{0, 1, 2, \ldots\}$. *If*

$$\sum_{k=1}^{\gamma_n} E\big(\xi_{nk}^2 I\big(|\xi_{nk}| > \delta\big)|\mathscr{H}_{k-1}^n\big) \xrightarrow{P} 0, \quad \forall \delta \in (0, 1],$$

$$\sum_{k=1}^{\gamma_n} E\big(\xi_{nk}^2|\mathscr{H}_{k-1}^n\big) \xrightarrow{P} \sigma^2,$$

then

$$\sum_{k=1}^{\gamma_n} \xi_{nk} \xrightarrow{\mathscr{D}} N\big(0, \sigma^2\big).$$

Lemma 5.2 (Billingsley [10], Lemma 1) *If* ξ *is measurable with respect to* $\mathscr{M}_{-\infty}^k$ *and* η *is measurable with respect to* $\mathscr{M}_{k+n}^\infty$ ($n \geq 0$), *then*

$$E|\xi|^r < \infty, \qquad E|\eta|^s < \infty, \quad r, s > 1, r^{-1} + s^{-1} = 1,$$

implies

$$\left|E(\xi\eta) - E(\xi)E(\eta)\right| \leq 2\varphi^{\frac{1}{r}}(n)\left(E|\xi|^r\right)^{\frac{1}{r}}\left(E|\eta|^s\right)^{\frac{1}{s}}.$$

Lemma 5.3 (Yang [16], Lemma 2) *Let $p \geq 2$, and let $\{X_n\}_{n\geq 1}$ be a φ-mixing sequence with $\sum_{n=1}^{\infty}\varphi^{1/2}(n) < \infty$. If $EX_n = 0$ and $E|X_n|^p < \infty$ for all $n \geq 1$, then*

$$E\left|\sum_{i=1}^{n}X_i\right|^p \leq C\left(\sum_{i=1}^{n}E|X_i|^p + \left(\sum_{i=1}^{n}EX_i^2\right)^{p/2}\right),$$

where C is a positive constant depending only on $\varphi(\cdot)$.

Lemma 5.4 (Fan and Yao [13], Proposition 2.6) *Let $\mathscr{F}_i^j$ and $\alpha(\cdot)$ be the same as in (2.57) of Fan and Yao [13]. Let $\xi_1, \xi_2, \ldots, \xi_k$ be complex-valued random variables measurable with respect to the σ-algebras $\mathscr{F}_{i_1}^{j_1}, \ldots, \mathscr{F}_{i_k}^{j_k}$, respectively. Suppose $i_{l+1} - j_l \geq n$ for $l = 1, \ldots, k-1$ and $j_l \geq i_l$ and $P(|\xi_l| \leq 1) = 1$ for $l = 1, 2, \ldots, k$. Then*

$$\left|E(\xi_1 \cdots \xi_k) - E(\xi_1) \cdots E(\xi_k)\right| \leq 16(k-1)\alpha(n).$$

Proof of Theorem 3.1 It is easy to see that

$$\sqrt{nh^d}\left(\widehat{m}_n(x) - m(x)\right) = \sqrt{nh^d}\left(\left[\widehat{m}_n(x) - E\widehat{m}_n(x)\right] + \left[E\widehat{m}_n(x) - m(x)\right]\right). \tag{5.1}$$

Combining Assumption 2.2 with the proof of Lemma 2 of Shen and Xie [7], we obtain that

$$\left|E\widehat{m}_n(x) - m(x)\right| = O\left(h^2\right), \quad x \in S_f.$$

Then, it follows from $nh^{d+4} \to 0$ that

$$\sqrt{nh^d}\left[E\widehat{m}_n(x) - m(x)\right] = O\left(\sqrt{nh^{d+4}}\right) \to 0, \quad x \in S_f. \tag{5.2}$$

For $x \in S_f$, let $Z_i := \sqrt{h^d}\frac{Y_i K_h(x - X_i)}{f(X_i)}$, $1 \leq i \leq n$. Denote

$$\sqrt{nh^d}\left[\widehat{m}_n(x) - E\widehat{m}_n(x)\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sqrt{h^d}\left[\frac{Y_i K_h(x - X_i)}{f(X_i)} - E\frac{Y_i K_h(x - X_i)}{f(X_i)}\right]$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_i - EZ_i). \tag{5.3}$$

To prove (3.1), we apply (5.1)–(5.3) and have to show that

$$\sqrt{nh^d}\left[\widehat{m}_n(x) - E\widehat{m}_n(x)\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_i - EZ_i) \xrightarrow{\mathscr{D}} N\left(0, \sigma^2(x)\right), \tag{5.4}$$

where $\sigma^2(x)$ is defined by (3.1).

Combining the independent and identically distributed stochastic sequence of $\{(X_i, Y_i), i \geq 1\}$ with Lemma 5.1, to prove (5.4), we have to show that

$$\frac{1}{n} \sum_{i=1}^{n} E(Z_i - EZ_i)^2 = \mathrm{Var}(Z_1) \to \sigma^2(x) \tag{5.5}$$

and, for all $\lambda \in (0, 1]$,

$$\frac{1}{n} \sum_{i=1}^{n} E\left((Z_i - EZ_i)^2 I\left(\frac{|Z_i - EZ_i|}{\sqrt{n}} > \lambda\right)\right) \to 0. \tag{5.6}$$

Obviously, for any $1 \leq r \leq 2 + \delta$ $(0 < \delta \leq 1)$, by (2.1) and (2.3) we have

$$
\begin{aligned}
h^{d(r-1)} &E\left|\frac{K_h(x - X_1)Y_1}{f(X_1)}\right|^r \\
&= h^{d(r-1)} E\left(\frac{|K_h(x - X_1)|^r}{f^r(X_1)} E\left(|Y_1|^r | X_1\right)\right) \\
&= \int_{S_f} \left|K\left(\frac{x - u}{h}\right)\right|^r E\left(|Y_1|^r | X_1 = u\right) \frac{1}{h^d} \frac{f(u)}{f^r(u)} \, du \\
&\leq \int_{S_f} \left|K\left(\frac{x - u}{h}\right)\right|^r \left(E\left(|Y_1|^{2+\delta} | X_1 = u\right)f(u)\right)^{\frac{r}{2+\delta}} \frac{1}{h^d} \frac{1}{f^{\frac{(3+\delta)r}{2+\delta} - 1}(u)} \, du \\
&\leq \frac{(B_0)^{\frac{r}{2+\delta}} \bar{K}^{r-1} \mu}{(\inf_{x \in S_f} f(x))^{\frac{(3+\delta)r}{2+\delta} - 1}} := \bar{\mu}(r) < \infty.
\end{aligned}
\tag{5.7}
$$

By (5.7) with $r = 1$ this yields

$$(EZ_1)^2 = h^d \left(E \frac{K_h(x - X_1)Y_1}{f(X_1)}\right)^2 \leq ch^d \to 0. \tag{5.8}$$

Define

$$
g(x) = \begin{cases} \frac{E(Y_1^2 | X_1 = x)}{f(x)} & \text{if } x \in S_f, \\ 0 & \text{otherwise.} \end{cases}
$$

In view of condition (2.3), we have

$$
\begin{aligned}
\int_{R^d} g(x) \, dx &= \int_{S_f} \frac{E(Y_1^2 | X_1 = x)}{f(x)} \, dx = \int_{S_f} \frac{E(Y_1^2 | X_1 = x) f^{\frac{2}{2+\delta}}(x)}{f^{\frac{4+\delta}{2+\delta}}(x)} \, dx \\
&\leq \int_{S_f} \frac{(E(|Y_1|^{2+\delta} | X_1 = x))^{\frac{2}{2+\delta}} f^{\frac{2}{2+\delta}}(x)}{f^{\frac{4+\delta}{2+\delta}}(x)} \, dx \\
&\leq \frac{B_0^{\frac{2}{2+\delta}}}{(\inf_{x \in S_f} f(x))^{\frac{6+2\delta}{2+\delta}}} \int_{R^d} f(x) \, dx = \frac{B_0^{\frac{2}{2+\delta}}}{(\inf_{x \in S_f} f(x))^{\frac{6+2\delta}{2+\delta}}} < \infty.
\end{aligned}
$$

So we have $g(x) \in L_1$. Since that $\frac{E(Y_1^2|X_1=x)}{f(x)}$ is positive and continuous at a point $x \in S_f$ and $\lim_{\|u\| \to \infty} \|u\|^d K(u) = 0$, we obtain by Bochner lemma [14] that

$$E(Z_1^2) = h^d E\left(\frac{K_h(x-X_1)Y_1}{f(X_1)}\right)^2 = \int_{S_f} K^2\left(\frac{x-u}{h}\right)E(Y_1^2|X_1=u)\frac{1}{h^d}\frac{1}{f(u)}\,du$$

$$= \int_{R^d} K^2\left(\frac{x-u}{h}\right)\frac{1}{h^d}g(u)\,du \to \frac{E(Y_1^2|X_1=x)}{f(x)}\int_{R^d} K^2(u)\,du. \tag{5.9}$$

Then, it follows from (5.8) and (5.9) that, for $x \in S_f$,

$$\text{Var}(Z_1) \to \frac{E(Y_1^2|X_1=x)}{f(x)}\int_{R^d} K^2(u)\,du = \sigma^2(x), \tag{5.10}$$

which implies (5.6). Meanwhile, for some $\delta \in (0,1]$ and any $\lambda \in (0,1]$, by C_r inequality and (5.7) we get that

$$\frac{1}{n}\sum_{i=1}^{n} E\left((Z_i - EZ_i)^2 I\left(\frac{|Z_i - EZ_i|}{\sqrt{n}} > \lambda\right)\right) = E\left((Z_1 - EZ_1)^2 I\left(\frac{|Z_1 - EZ_1|}{\sqrt{n}} > \lambda\right)\right)$$

$$\leq \frac{1}{n^{\frac{\delta}{2}}\lambda^\delta}E|Z_1 - EZ_1|^{2+\delta} \leq \frac{c_1}{n^{\frac{\delta}{2}}\lambda^\delta}E|Z_1|^{2+\delta}$$

$$\leq \frac{c_2}{(nh^d)^{\frac{\delta}{2}}} \to 0, \tag{5.11}$$

since $nh^d \to \infty$. Thus, (5.6) follows from (5.11). Consequently, the proof of the theorem is completed. □

Proof of Theorem 3.2 We use the same notation as in the proof of Theorem 3.1. Under the conditions of Theorem 3.2, by (5.1), (5.2), and (5.3), to prove (3.1), we need to show that

$$\sqrt{nh^d}\left[\widehat{m}_n(x) - E\widehat{m}_n(x)\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_i - EZ_i) \xrightarrow{\mathscr{D}} N(0, \sigma^2(x)), \tag{5.12}$$

where $\sigma^2(x)$ is defined by (3.1). By the second-order stationarity, $\{(X_i, Y_i), i \geq 1\}$ are identically distributed. Then, for $1 \leq i \leq n$, we have by (5.8) and (5.9) that

$$\text{Var}(Z_i - EZ_i) = \text{Var}(Z_1) \to \frac{E(Y_1^2|X_1=x)}{f(x)}\int_{R^d} K^2(u)\,du = \sigma^2(x). \tag{5.13}$$

For $j \geq 1$, in view of (2.4), we have

$$E\left|\frac{K_h(x-X_1)K_h(x-X_{j+1})Y_1Y_{j+1}}{f(X_1)f(X_{j+1})}\right|$$

$$= E\left(\frac{|K_h(x-X_1)K_h(x-X_{j+1})|}{f(X_1)f(X_{j+1})}E(|Y_1Y_{j+1}||X_1, X_{j+1})\right)$$

$$= \int_{S_f}\int_{S_f}\left|K\left(\frac{x-u_1}{h}\right)K\left(\frac{x-u_{j+1}}{h}\right)\right|E(|Y_1Y_{j+1}||X_1=u_1, X_j=u_{j+1})$$

$$\times \frac{1}{h^{2d}}\frac{1}{f(u_1)f(u_{j+1})}f_j(u_1, u_{j+1})\,du_1\,du_{j+1}$$

$$\leq \frac{B_1}{(\inf_{x \in S_f} f(x))^2} \int_{R^d} \int_{R^d} \left| K\left(\frac{x - u_1}{h}\right) K\left(\frac{x - u_{j+1}}{h}\right) \right| \frac{1}{h^{2d}} \, du_1 \, du_{j+1}$$

$$\leq \frac{B_1 \mu^2}{(\inf_{x \in S_f} f(x))^2} < \infty. \tag{5.14}$$

So it follows from (5.7) and (5.14) that

$$\left| \mathrm{Cov}(Z_1, Z_j) \right| \leq E|Z_1 Z_j| + \left(E|Z_1| \right)^2 \leq c_1 h^d, \quad j > 1. \tag{5.15}$$

Obviously, by the stationarity we establish that

$$\frac{1}{n} \mathrm{Var}\left(\sum_{i=1}^n (Z_i - EZ_i) \right) = \frac{1}{n} \mathrm{Var}\left(\sum_{i=1}^n Z_i \right) = \mathrm{Var}(Z_1) + \frac{2}{n} \sum_{1 \leq i < j \leq n} \mathrm{Cov}(Z_i, Z_j)$$

$$= \mathrm{Var}(Z_1) + \frac{2}{n} \left\{ \left[\sum_{\substack{1 \leq i < j \leq n \\ 1 \leq j - i \leq r_n}} + \sum_{\substack{1 \leq i < j \leq n \\ j - i > r_n}} \right] \mathrm{Cov}(Z_i, Z_j) \right\}. \tag{5.16}$$

For h^d, we can choose r_n satisfying that $r_n \to \infty$ and $h^d r_n \to 0$ as $n \to \infty$. So, by (5.15),

$$\frac{2}{n} \sum_{\substack{1 \leq i < j \leq n \\ 1 \leq j - i \leq r_n}} \left| \mathrm{Cov}(Z_i, Z_j) \right| \leq c h^d r_n \to 0. \tag{5.17}$$

By Lemma 5.2 with $s = r = 2$, the condition $\sum_{n=1}^\infty \varphi^{1/2}(n) < \infty$, and (5.9), we can show that

$$\frac{2}{n} \sum_{\substack{1 \leq i < j \leq n \\ j - i > r_n}} \left| \mathrm{Cov}(Z_i, Z_j) \right| \leq \frac{c_1}{n} \sum_{\substack{1 \leq i < j \leq n \\ j - i > r_n}} \varphi^{1/2}(j - i) \leq c_2 \sum_{k > r_n} \varphi^{1/2}(k) \to 0. \tag{5.18}$$

Therefore, by (5.13), (5.16), (5.17), and (5.18), we get that

$$\frac{1}{n} \mathrm{Var}\left(\sum_{i=1}^n Z_i \right) = \sigma^2(x)\left(1 + o(1)\right).$$

Next, we employ Bernstein's big-block and small-block procedure (see Fan and Yao [13] and Masry [19]). Partition the set $\{1, 2, \ldots, n\}$ into $2k_n + 1$ subsets with large block of size $\mu = \mu_n$ and small block of size $\nu = \nu_n$ and set

$$k = k_n = \left\lfloor \frac{n}{\mu_n + \nu_n} \right\rfloor. \tag{5.19}$$

Define $\mu = \mu_n = \lfloor \sqrt{\frac{n}{h^d}} \rfloor$ and $\nu = \nu_n = \lfloor \sqrt{nh^d} \rfloor$. So we have by $h^d \to 0$ and $nh^d \to \infty$ that

$$\mu_n \to \infty, \quad \nu_n \to \infty, \quad \frac{\mu_n}{n} \to 0,$$

$$\frac{\nu_n}{n} \to 0, \quad \frac{\nu_n}{\mu_n} \to 0, \quad k_n = O\left(\sqrt{nh^d}\right). \tag{5.20}$$

Define η_j, ξ_j, and ζ_j as follows:

$$\eta_j := \sum_{i=j(\mu+v)+1}^{j(\mu+v)+\mu} (Z_i - EZ_i), \quad 0 \le j \le k-1, \tag{5.21}$$

$$\xi_j := \sum_{i=j(\mu+v)+\mu+1}^{(j+1)(\mu+v)} (Z_i - EZ_i), \quad 0 \le j \le k-1, \tag{5.22}$$

$$\zeta_k := \sum_{i=k(\mu+v)+1}^{n} (Z_i - EZ_i). \tag{5.23}$$

In view of

$$S_n := \sum_{i=1}^{n}(Z_i - EZ_i) = \sum_{j=0}^{k-1} \eta_j + \sum_{j=0}^{k-1} \xi_j + \zeta_k := S_n' + S_n'' + S_n''', \tag{5.24}$$

we have to show that

$$\frac{1}{n}E(S_n'')^2 \to 0, \qquad \frac{1}{n}E(S_n''')^2 \to 0, \tag{5.25}$$

$$\left| E\left(\exp\left(itn^{-1/2}S_n'\right)\right) - \prod_{j=0}^{k-1} E\left(\exp\left(itn^{-1/2}\eta_j\right)\right) \right| \to 0, \tag{5.26}$$

$$\frac{1}{n}\sum_{j=0}^{k-1} E\left(\eta_j^2\right) \to \sigma^2(x), \tag{5.27}$$

$$\frac{1}{n}\sum_{j=0}^{k-1} E\left(\eta_j^2 I\left(|\eta_j| > \varepsilon\sigma(x)\sqrt{n}\right)\right) \to 0, \quad \forall \varepsilon > 0. \tag{5.28}$$

Relation (5.25) implies that $\frac{S_n''}{\sqrt{n}}$ and $\frac{S_n'''}{\sqrt{n}}$ are asymptotically negligible, (5.26) shows that the summands $\{\eta_j\}$ in S_n' are asymptotically independent, and (5.27)–(5.28) are the standard Lindeberg–Feller conditions for the asymptotic normality of S_n' under independence.

First, we prove (5.25). By (5.22) and (5.24) we have

$$E(S_n'')^2 = \mathrm{Var}\left(\sum_{j=0}^{k-1} \xi_j\right) = \sum_{j=0}^{k-1} \mathrm{Var}(\xi_j) + 2\sum_{0 \le i < j \le k-1} \mathrm{Cov}(\xi_i, \xi_j) := F_1 + F_2. \tag{5.29}$$

By the stationarity and (5.10), similarly to the proof of (5.17) and (5.18), for $0 \le j \le k-1$, we have

$$\mathrm{Var}(\xi_j) = v_n \mathrm{Var}(Z_1) + 2\sum_{1 \le i < j \le v_n} \mathrm{Cov}(Z_i, Z_j) = v_n \sigma^2(x)\left(1 + o(1)\right). \tag{5.30}$$

Thus it follows from (5.19) and (5.20) that

$$F_1 = k_n v_n \sigma^2(x)\left(1 + o(1)\right) \sim \frac{nv_n}{\mu_n + v_n} \sim \frac{nv_n}{\mu_n} = o(n). \tag{5.31}$$

We consider the term F_2 in (5.29). With $\lambda_j = j(\mu_n + \nu_n) + \mu_n$,

$$F_2 = 2 \sum_{0 \le i < j \le k-1} \sum_{l_1=1}^{\nu_n} \sum_{l_2=1}^{\nu_n} \mathrm{Cov}(Z_{\lambda_i + l_1}, Z_{\lambda_j + l_2}),$$

but since $i \ne j$, $|\lambda_i - \lambda_j + l_1 - l_2| \ge \mu_n$ for $0 \le i < j \le k-1$, $1 \le l_1 \le \nu_n$, and $1 \le l_2 \le \nu_n$, similarly to the proof of (5.18), it follows that

$$|F_2| \le 2 \sum_{\substack{1 \le i < j \le n \\ j - i \ge \mu_n}} \left| \mathrm{Cov}(Z_i, Z_j) \right| = o(n). \tag{5.32}$$

Hence by (5.29), (5.31), and (5.32) we have

$$\frac{1}{n} E(S_n'')^2 \to 0.$$

By (5.13), (5.20), and (5.23), similarly to the proofs of (5.17) and (5.18), we can find that

$$\frac{1}{n} E(S_n''')^2 \le \frac{1}{n}\left(n - k_n(\mu_n + \nu_n)\right) \mathrm{Var}(Z_1) + \frac{2}{n} \sum_{1 \le i < j \le n - k_n(\mu_n + \nu_n)} \left| \mathrm{Cov}(Z_i, Z_j) \right|$$

$$\le C \frac{\mu_n + \nu_n}{n} \sigma^2(x) + o(1) \to 0.$$

Thus

$$\frac{1}{\sqrt{n}} S_n = \frac{1}{\sqrt{n}}\left(S_n' + S_n'' + S_n'''\right) = \frac{1}{\sqrt{n}} S_n' + o_p(1). \tag{5.33}$$

Second, it is easy to see that $\varphi^{1/2}(n) = o(\frac{1}{n})$ by $\varphi(n) \downarrow 0$ and $\sum_{n=1}^{\infty} \varphi^{1/2}(n) < \infty$. Note that η_a is $\mathcal{M}_{i_a}^{j_a}$-measurable with $i_a = a(\mu + \nu) + 1$ and $j_a = a(\mu + \nu) + \mu$. Since φ-mixing random variables are strong mixing random variables and $\alpha(n) \le \varphi(n)$, letting $V_j = \exp(itn^{-1/2}\eta_j)$, by Lemma 5.4 we have

$$\left| E\left(\exp\left(itn^{-1/2} S_n'\right)\right) - \prod_{j=0}^{k-1} E\left(\exp\left(itn^{-1/2}\eta_j\right)\right) \right|$$

$$\le c k_n \varphi(\nu_n + 1) \le c \frac{n}{\mu_n + \nu_n} \frac{1}{\nu_n^2} \le \frac{c}{\sqrt{nh^d}} \to 0$$

by (5.19), (5.20), and the conditions $h_n \to 0$ and $nh^d \to \infty$ as $n \to \infty$.

Third, we show (5.27), where η_j is defined in (5.21). By the stationarity and (5.30) with μ_n replacing ν_n, we have

$$E(\eta_j^2) = \mathrm{Var}(\eta_j) = \mathrm{Var}(\eta_0) = \mu_n \sigma^2(x)\left(1 + o(1)\right), \quad 0 \le j \le k-1, \tag{5.34}$$

so that

$$\frac{1}{n} \sum_{j=0}^{k_n-1} E(\eta_j^2) = \frac{k_n \mu_n}{n} \sigma^2(x)\left(1 + o(1)\right) \to \sigma^2(x),$$

since $k_n \mu_n / n \to 1$.

Fourth, it is time to establish (5.28). Obviously, by (5.7) we obtain that

$$EZ_i^2 = EZ_1^2 \leq c_1 \quad \text{and} \quad E|Z_i|^{2+\delta} = E|Z_1|^{2+\delta} \leq c_2 \big(h^d\big)^{-\frac{\delta}{2}}, \quad 1 \leq i \leq n.$$

We can see that $\frac{\frac{1}{h^d}}{\mu_n} \leq \frac{c}{h^d \sqrt{\frac{n}{h^d}}} = \frac{c}{(nh^d)^{\frac{1}{2}}} \to 0$, since $nh^d \to \infty$ as $n \to \infty$. Therefore, by Lemma 5.3 with $\sum_{n=1}^{\infty} \varphi^{1/2}(n) < \infty$ we have that

$$E\left|\sum_{i=1}^{\mu_n}(Z_i - EZ_i)\right|^{2+\delta} \leq c_1\left(\sum_{i=1}^{\mu_n} E|Z_i|^{2+\delta} + \left(\sum_{i=1}^{\mu_n} EZ_i^2\right)^{\frac{2+\delta}{2}}\right)$$

$$\leq c_2\left(\mu_n \frac{1}{(h^d)^{\frac{\delta}{2}}} + \mu_n^{1+\frac{\delta}{2}}\right) \leq c_3 \mu_n^{1+\frac{\delta}{2}}. \tag{5.35}$$

Then, for all $\varepsilon > 0$, by (5.34) and (5.35) it is easy to see that

$$E\big(\eta_0^2 I\big(|\eta_0| \geq \varepsilon\sigma(x)n^{1/2}\big)\big) \leq \frac{1}{(\varepsilon\sigma(x)n^{1/2})^{\delta}} E|\eta_0|^{2+\delta} I\big(|\eta_0| \geq \varepsilon\sigma(x)n^{1/2}\big)$$

$$\leq \frac{1}{(\varepsilon\sigma(x)n^{1/2})^{\delta}} E|\eta_0|^{2+\delta} \leq c_1 \frac{\mu_n^{1+\frac{\delta}{2}}}{n^{\frac{\delta}{2}}}.$$

Similarly, for $0 \leq j \leq k-1$, we get that

$$E\big(\eta_j^2 I\big(|\eta_j| \geq \varepsilon\sigma(x)n^{1/2}\big)\big) \leq c_2 \frac{\mu_n^{1+\frac{\delta}{2}}}{n^{\frac{\delta}{2}}}.$$

Therefore, since $0 < \delta \leq 1$ and $nh^d \to \infty$, we obtain that, for all $\varepsilon > 0$,

$$\frac{1}{n}\sum_{j=0}^{k-1} E\big(\eta_j^2 I\big(|\eta_j| \geq \varepsilon\sigma(x)n^{1/2}\big)\big) \leq \frac{ck\mu_n^{1+\frac{\delta}{2}}}{n^{1+\frac{\delta}{2}}} \leq \frac{c\mu_n^{1+\frac{\delta}{2}}\frac{n}{\mu_n+\nu_n}}{n^{1+\frac{\delta}{2}}} \leq \frac{c\mu_n^{\frac{\delta}{2}}}{n^{\frac{\delta}{2}}}$$

$$= c\left(\frac{\sqrt{\frac{n}{h^d}}}{n}\right)^{\frac{\delta}{2}} = \frac{c}{(nh^d)^{\frac{\delta}{4}}} \to 0.$$

Therefore, (5.26), (5.27), and (5.28) hold for S_n', so that

$$\frac{1}{\sqrt{n}} S_n' \xrightarrow{\mathcal{D}} N\big(0, \sigma^2(x)\big). \tag{5.36}$$

Consequently, (5.12) follows from (5.33) and (5.36). Finally, by (5.1), (5.2), and (5.12) we obtain (3.1). The proof of theorem is completed.　　　　　　　　　　　　　　□

Author details
¹Automotive Electronics Engineering Research Center, College of Automation, Chongqing University of Posts and Telecommunications, Chongqing, China. ²School of Mathematical Sciences, Anhui University, Hefei, China. ³School of Electrical Engineering and Automation, Hefei University of Technology, Hefei, China.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
 1. Nadaraya, E.A.: On estimating regression. Theory Probab. Appl. **9**, 141–142 (1964)
 2. Watson, G.S.: Smooth regression analysis. Sankhya, Ser. A **26**, 359–372 (1964)
 3. Jones, M.C., Davies, S.J., Park, B.U.: Versions of kernel-type regression estimators. J. Am. Stat. Assoc. **89**, 825–832 (1994)
 4. Mack, Y.P., Müller, H.G.: Derivative estimation in nonparametric regression with random predictor variable. Sankhya **51**, 59–72 (1989)
 5. Linton, O., Nielsen, J.: A kernel method of estimating structured nonparametric regression based on marginal integration. Biometrika **82**, 93–100 (1995)
 6. Linton, O., Jacho-Chávez, D.: On internally corrected and symmetrized kernel estimators for nonparametric regression. Test **19**, 166–186 (2010)
 7. Shen, J., Xie, Y.: Strong consistency of the internal estimator of nonparametric regression with dependent data. Stat. Probab. Lett. **83**, 1915–1925 (2013)
 8. Li, X.Q., Yang, W.Z., Hu, S.H.: Uniform convergence of estimator for nonparametric regression with dependent data. J. Inequal. Appl. **2016**, 142 (2016)
 9. Dobrushin, R.L.: The central limit theorem for non-stationary Markov chain. Theory Probab. Appl. **1**, 72–88 (1956)
10. Billingsley, P.: Convergence of Probability Measures. Wiley, New York (1968)
11. Györfi, L., Härdle, W., Sarda, P., Vieu, P.: Nonparametric Curve Estimation from Time Series. Springer, Berlin (1989)
12. Györfi, L., Kohler, M., Krzyżak, A., Walk, H.: A Distribution-Free Theory of Nonparametric Regression. Springer, New York (2002)
13. Fan, J.Q., Yao, Q.W.: Nonlinear Time Series: Nonparametric and Parametric Methods. Springer, New York (2003)
14. Bosq, D., Blanke, D.: Inference and Prediction in Large Dimensions. Wiley, Chichester (2007)
15. Hansen, B.E.: Uniform convergence rates for kernel estimation with dependent data. Econom. Theory **24**, 726–748 (2008)
16. Yang, S.C.: Almost sure convergence of weighted sums of mixing sequences. J. Syst. Sci. Math. Sci. **15**, 254–265 (1995)
17. Yang, S.C.: Maximal moment inequality for partial sums of strong mixing sequences and application. Acta Math. Sin. Engl. Ser. **23**, 1013–1024 (2007)
18. Liptser, R.S., Shiryayev, A.N.: Theory of Martingales. Kluwer Academic, Dordrecht (1989)
19. Masry, E.X.: Nonparametric regression estimation for dependent functional data: asymptotic normality. Stoch. Process. Appl. **115**, 155–177 (1989)