Available online at www.sciencedirect.com

SCIENCE DIRECT

ELSEVIER Stochastic Processes and their Applications 115 (2005) 155–177

stochastic processes and their applications

www.elsevier.com/locate/spa

Nonparametric regression estimation for dependent functional data: asymptotic normality ☆

Elias Masry*

Department of Electrical and Computer Engineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0407, USA

Received 7 January 2003; received in revised form 25 May 2004; accepted 12 July 2004
Available online 12 August 2004

Abstract

We consider the estimation of a regression functional where the explanatory variables take values in some abstract function space. The principal aim of the paper is to establish the asymptotic normality of such estimates for dependent functional data.

MSC: primary 62G07, secondary 62M09

Keywords: Functional dependent data; Nonparametric regression estimation; Asymptotic normality

1. Introduction

Regression function estimation is an important problem in data analysis with a wide range of applications in filtering and prediction in communications and control systems, pattern recognition and classification, and econometrics (Györfi et al., [8], Härdle [10], Fukunaga [6], and Tjostheim [19]). There is an extensive literature on regression estimation for i.i.d. data (see, for example, Rosenblatt [16], Schuster [18],

☆This work was supported by the Center for Wireless Communications at UCSD.
*Tel.: +858-5343815; fax: +858-8221247.
E-mail address: masry@ece.ucsd.edu (E. Masry).

0304-4149/$ - see front matter
doi:10.1016/j.spa.2004.07.006

Mack and Silverman [11] and the references therein) as well as for dependent data (see, for example, Robinson [15], Collomb and Härdle [2], Roussas [17], Tran [20] and the references therein).

In this paper we consider the case of functional data. There has been an increasing interest in this area in recent years. For an introduction to this field, the reader is directed to the books of Ramsay and Silverman [13,14]. The latter provides some basic methods of analysis along with diverse case studies in several areas including criminology, economics, archaeology, and neurophysiology. It should be noted that the extension of probability theory to random variables taking values in normed spaces (e.g. Banach and Hilbert spaces), including extensions of certain classical asymptotic limit theorems predates the recent literature on functional data; the reader is referred to the books Araujo et al. [1] and Vakashnia et al. [21].

Gasser et al. [7] considers density and mode estimation for data taking values in a normed vector space. The paper highlights the issue of the curse of dimensionality for functional data and suggests methods to mitigate the problem. We shall revisit this issue in Section 4. In the context of regression estimation nonparametric models were considered by Ferraty and Vieu [4,5].

Let $\{Y_i, X_i\}_{i=1}^{\infty}$ be random processes where Y_i is real-valued and X_i takes values in some abstract space $\mathscr{H}$. While $\mathscr{H}$ can be assumed to be a semi-metric vector space with semi metric $d(\cdot,\cdot)$, in most practical applications, $\mathscr{H}$ is a normed space (e.g. Hilbert or a Banach space) with norm $\|.\|$ so that $d(u,v) = \|u-v\|$. Assume that $E|Y_i| < \infty$ and define the regression functional as

$$r(u) := E[Y_i | X_i = u]; \quad u \in \mathscr{H},$$

which is assumed to be independent of i. We do not assume that the processes $\{Y_i, X_i\}_{i=1}^{\infty}$ are necessarily strictly stationary; it suffices to assume second-order stationarity.

A Nadaraya–Watson type estimator for r was introduced in Ferraty and Vieu [5] taking the form

$$\hat{r}(x) = \frac{\sum_{i=1}^{n} Y_i K(d(x, X_i)/h)}{\sum_{i=1}^{n} K(d(x, X_i)/h)}, \tag{1.1}$$

where $K(\cdot)$ is a real-valued kernel function and $h = h_n$ is the bandwidth parameter. Rates of almost sure convergence were established in Ferraty and Vieu [5] for strongly mixing processes.

The purpose of this paper is to establish the asymptotic normality of the estimate $\hat{r}(x)$ for strongly mixing processes. It should be noted that even for i.i.d. functional data, no asymptotic normality has so far been established. We remark that establishing central limit theorems utilizes different methods of analysis than those used to obtain rates of a.s. convergence. For dependent data (functional or not) the usual method for establishing rates of a.s. convergence for nonparametric function(al) estimates employs upper bounds on moments of the estimate, an exponential inequality, and the use of the Borel–Cantelli lemma (see for example Masry [12]). This is indeed the method of analysis used in Ferraty and Vieu [5]. Establishing central limit theorems for nonparametric function estimates for dependent data utilizes appropriate form of Bernstein's blocking argument and a reduction analysis leading to the Lindeberg–Feller conditions for independent variables (see for example Fan and Masry [3]); while the overall approach is the same, the technical details of the proofs for functional data are more involved as can be seen from the derivations in Section 5.

The organization of the paper is as follows. Section 2 establishes the quadratic-mean convergence for certain estimates appearing in (1.1) along with convergence in probability (with rates) of $\hat{r}(x)$ (Theorems 2, 3 and Corollary 1). These are subsequently used in Section 3 to establish the asymptotic normality of $\hat{r}(x)$ (Theorems 4 and 5). Section 4 is devoted to the discussion of the results, including the issue of the curse of the dimensionality and an example. Derivations are presented in Section 5.

2. Quadratic-mean convergence

We first introduce a suitable decomposition of the estimation error that greatly facilitates the analysis. Set

$$\Delta_i(x) := K(d(x, X_i)/h)$$

and define

$$\hat{r}_1(x) := \frac{1}{nE[\Delta_1(x)]} \sum_{i=1}^{n} \Delta_i(x) \tag{2.1}$$

and

$$\hat{r}_2(x) := \frac{1}{nE[\Delta_1(x)]} \sum_{i=1}^{n} Y_i \Delta_i(x) \tag{2.2}$$

so that $\hat{r}(x) = \hat{r}_2(x)/\hat{r}_1(x)$. Now, define the "bias" term by

$$B_n(x) := \frac{E[\hat{r}_2(x)] - r(x)E[\hat{r}_1(x)]}{E[\hat{r}_1(x)]} \tag{2.3}$$

and a centered variate

$$Q_n(x) := (\hat{r}_2(x) - E[\hat{r}_2(x)]) - r(x)(\hat{r}_1(x) - E[\hat{r}_1(x)]). \tag{2.4}$$

Then it can be seen that

$$\hat{r}(x) - r(x) - B_n(x) = \frac{Q_n(x) - B_n(x)(\hat{r}_1(x) - E[\hat{r}_1(x)])}{\hat{r}_1(x)}. \tag{2.5}$$

The relationship (2.5) is fundamental to our goal of establishing a central limit theorem for $\hat{r}(x)$: under certain regularity conditions, we will show that $\hat{r}_1(x)$ converges in quadratic mean to 1 as $n \to \infty$. Moreover, the bias term $B_n(x) = o(1)$ as

$n \to \infty$. It then follows from (2.5) that

$$\hat{r}(x) - r(x) - B_n(x) = \frac{Q_n}{\hat{r}_1(x)}(1 + o_p(1)).$$

Thus, in order to obtain a central limit theorem for the left side of the above equation, it suffices to establish asymptotic normality for the variate $Q_n(x)$.

We make the following assumptions which are subsequently discussed in Remark 1.

Condition 1 (*Kernel*). $K(t)$ is nonnegative bounded kernel with support $[0,1]$ satisfying

$$0 < c_1 \leqslant K(t) \leqslant c_2 < \infty$$

for some constants c_1, c_2.

Condition 2 (*Smoothness*).

(i) $$|r(u) - r(v)| \leqslant c_3 \; d(u,v)^\beta$$

for all $u, v \in \mathscr{H}$ for some $\beta > 0$.

(ii) Let

$$g_2(u) := \mathrm{var}[Y_j | X_j = u], \;\; u \in \mathscr{H}.$$

$g_2(u)$ is independent of j and is continuous in some neighborhood of x:

$$\sup_{\{u: d(x,u) \leqslant h\}} |g_2(u) - g_2(x)| = o(1) \quad \text{as } h \to 0.$$

Assume $E|Y_i|^\nu < \infty$ for some $\nu > 2$. Assume

$$g_\nu(u) := E[|Y_i - r(x)|^\nu \,|\, X_i = u], \quad u \in \mathscr{H}$$

is continuous in some neighborhood of x.

(iii) Define

$$g(u,v;x) := E[(Y_i - r(x))(Y_j - r(x)) | X_i = u, X_j = v], \quad i \neq j, \;\; u, v \in \mathscr{H}.$$

Assume that $g(u,v;x)$ does not depend on i,j and is continuous in some neighborhood of (x,x).

Let $B(x,h)$ be a ball centered at $x \in \mathscr{H}$ with radius h. Ferraty and Vieu [5] assume uniform upper and lower bounds on $P[X_j \in B(x,h)]$ of the form $0 < c_5 \, \phi(h) \leqslant P[X_j \in B(x,h)] \leqslant c_6 \, \phi(h)$. The uniformity was questioned by a referee. We adopt a different condition consistent with the assumptions made in Gasser et al. [7] in the context of density estimation for functional data: Let $D_i := d(x, X_i)$ so that D_i is a real-valued nonnegative random variable and denote its distribution by $F(u;x) = P[D_i \leqslant u]$. One is interested in the behavior of $F(u;x)$ as $u \to 0$. Gasser et al. [7] assume that $F(h;x) = \phi(h) f_1(x)$ as $h \to 0$ and refer to $f_1(x)$ as the probability density (functional). When $\mathscr{H} = \mathbb{R}^m$, then $F(h;x) = P[\|x - X_i\| \leqslant h]$ and it can be seen that in this case $\phi(h) = C(m)h^m$ ($C(m)$ is the volume of a unit ball in $\mathbb{R}^m$) and $f_1(x)$ is the probability density of the random variable X_1. Indeed, it can be shown directly that $\lim_{h\to 0}(1/h^m)F(h;x) = C(m)f_1(x)$. Motivated by the work of Gasser et al. and the above argument we make the following assumption:

Condition 3 (*Distributions*).

(i) For all $i \geqslant 1$,

$$0 < c_5\,\phi(h)f_1(x) \leqslant P[X_i \in B(x,h)] = F(h;x) \leqslant c_6\,\phi(h)f_1(x),$$

where $\phi(h) \to 0$ as $h \to 0$ and $f_1(x)$ is a nonnegative functional in $x \in \mathscr{H}$.

(ii)
$$\sup_{i \neq j} P[(X_i, X_j) \in B(x,h) \times B(x,h)] = \sup_{i \neq j} P[D_i \leqslant h, D_j \leqslant h] \leqslant \psi(h)f_2(x),$$

where $\psi(h) \to 0$ as $h \to 0$ and $f_2(x)$ is a nonnegative functional in $x \in \mathscr{H}$. We assume that the ratio $\psi(h)/\phi^2(h)$ is bounded.

Finally, we assume that the processes $\{Y_i, X_i\}$ are strongly mixing: Let $\mathscr{F}_a^b$ be the sigma algebra generated by the random variables $\{Y_i, X_i\}_{i=a}^b$. Set

$$\alpha(l) = \sup_t \sup_{\substack{A \in \mathscr{F}_{-\infty}^t \\ B \in \mathscr{F}_{t+l}^\infty}} |P[AB] - P[A]P[B]|.$$

We assume

Condition 4 (*Mixing*).

$$\sum_{l=1}^{\infty} l^\delta [\alpha(l)]^{1-2/v} < \infty$$

for some $v > 2$ and $\delta > 1 - 2/v$.

Remark 1. The above conditions are fairly mild. Condition 1 is standard except for $K(t)$ being bounded away from zero. This latter assumption can be dropped (see Condition 1' below). Condition 2 is a mild smoothness assumption on the regression functional $r(u)$ and continuity assumption on certain second-order moments. As was discussed earlier, Condition 3(i) is consistent with the assumptions made by Gasser et al. [7] in the context of density estimation for functional data. When $\mathscr{H}$ is a separable Hilbert space and is infinitely dimensional, $\phi(h)$ could decrease to zero as $h \to 0$ exponentially fast [7]. Similar argument applies to Condition 3(ii) which gives the behavior of the joint distribution of (D_i, D_j) near the origin. Condition 4 is a standard assumption on the decay of the strongly mixing coefficient $\alpha(l)$. We note that for Theorem 1, we can set $v = \infty$ and $\delta > 1$ since the kernel K is bounded; however, in Theorems 2 and 3, the random variables $\{Y_i\}$ are not necessarily bounded and there is a tradeoff between the decay of the mixing condition and the order v of the moment $E|Y_1|^v < \infty$.

An alternative to Condition 3 is the following in which Condition 3(i) is modified.

Condition 3′ (*Distributions*).

(i) $$F(u;x) = \phi(u)f_1(x) \quad \text{as } u \to 0,$$

where $\phi(0) = 0$ and $\phi(u)$ is absolutely continuous in a neighborhood of the origin.

(ii) $\sup_{i \neq j} P[D_i \leqslant u, D_j \leqslant u] \leqslant \psi(u) f_2(x)$ as $u \to 0$, where $\psi(u) \to 0$ as $u \to 0$. We assume that the ratio $\psi(h)/\phi^2(h)$ is bounded.

Before we state our first result, we remark on the asymptotic value of the integral

$$I_j(h) := \frac{1}{\phi(h)/h} \int_0^1 K^j(u)\phi'(hv)\,\mathrm{d}v, \quad j = 1, 2. \tag{2.6}$$

Note that if $K(t) = 1_{[0,1]}(t)$ then $I_j(h) = 1$ for every $h > 0$. If the kernel K satisfies $0 < c_1 \leqslant K(t) \leqslant c_2 < \infty$, then $c_1 \leqslant I_j(h) \leqslant c_2$, again for every $h > 0$. In order to obtain an expression of the asymptotic variance (rather than upper and lower bounds), one can drop the lower bound on K and modify Condition 1 as follows:

Condition 1′ (*Kernel and Approximation of the Identity*). $K(t)$ is a nonnegative bounded kernel with compact support $[0, 1]$ satisfying

(i) $K(t) \leqslant c_2 < \infty$.
(ii) $I_j(h) \to C_j$ as $h \to 0$, $j = 1, 2$, for some positive constant C_j.

Theorem 1. *Let* $n\phi(h_n) \to \infty$ *as* $n \to \infty$. *Under Conditions* 1, 3(i), *and* 4 (*or under Conditions* 1′, 3′(i), *and* 4),

$$\hat{r}_1(x) \xrightarrow{m.s.} 1$$

for each $x \in \mathscr{H}$ *as* $n \to \infty$.

Next we consider the variance of the centered variate $Q_n(x)$ defined in (2.4). Define

$$\mu_n(x) := E[(Y_i - r(x))\Delta_i(x)] \tag{2.7}$$

and

$$Z_{n,i}(x) := (Y_i - r(x))\Delta_i(x) - \mu_n. \tag{2.8}$$

Then,

$$Q_n(x) = \frac{1}{nE[\Delta_1(x)]} \sum_{i=1}^{n} Z_{n,i}. \tag{2.9}$$

Let

$$\sigma^2_{n,0}(x) := \frac{1}{E^2[\Delta_1(x)]} \mathrm{var}[Z_{n,1}]. \tag{2.10}$$

Theorem 2. *Let Conditions* 1–4 (*with* $v=\infty$ *and* $\delta>1$ *for Condition* 4) *hold. Let* $n\phi(h_n)\to\infty$ *as* $n\to\infty$. *Then, for large* n,

(a) $$c_8\frac{g_2(x)}{f_1(x)}\leqslant\phi(h_n)\sigma^2_{n,0}(x)\leqslant c_9\frac{g_2(x)}{f_1(x)}$$

for some positive constants c_8 *and* c_9 *whenever* $f_1(x)>0$.

(b) $$\frac{1}{nE^2[\Delta_1(x)]}\sum_{\substack{i=1\\|i-j|>0}}^{n}\sum_{j=1}^{n}\operatorname{cov}\{Z_{n,i}(x),Z_{n,j}(x)\}=o(\sigma^2_{n,0}(x)).$$

(c) $$\operatorname{var}[Q_n(x)]=\frac{1}{n}\sigma^2_{n,0}(x)(1+o(1)).$$

If we use Conditions 1′ and 3′ (instead of Conditions 1 and 3), then the rate of quadratic-mean convergence and the asymptotic variance can be specified:

Theorem 3. *Let Conditions* 1′, 2, 3′, 4 (*with* $v=\infty$ *and* $\delta>1$) *hold. Let* $n\phi(h_n)\to\infty$ *as* $n\to\infty$. *Then,*

(a) $$\phi(h_n)\sigma^2_{n,0}(x)\to\frac{C_2}{C_1^2}\frac{g_2(x)}{f_1(x)}=:\sigma^2(x)$$

whenever $f_1(x)>0$ *and the constants* C_1 *and* C_2 *are specified in Condition* 1′, *and* $g_2(x)$ *is defined in Condition* 2(ii).

(b) $$\frac{1}{nE^2[\Delta_1(x)]}\sum_{\substack{i=1\\|i-j|>0}}^{n}\sum_{j=1}^{n}\operatorname{cov}\{Z_{n,i}(x),Z_{n,j}(x)\}=o(\sigma^2_{n,0}(x)).$$

(c) $$n\phi(h_n)\ \operatorname{var}[Q_n(x)]\to\sigma^2(x)$$

whenever $f_1(x)>0$.

Remark 2. When $K(t)=1_{[0,1]}(t)$, Condition 1′ is automatically satisfied in which case

$$\phi(h_n)\sigma^2_{n,0}(x)\to\frac{g_2(x)}{f_1(x)}$$

whenever $f_1(x)>0$.

In Section 4 we need the asymptotic expression of the variance when $\mathscr{H}=\mathbb{R}^m$: in this case it is easy to see that $f_1(x)$ is the probability density of the random variable X_1, $d(x,u)=\|x-u\|$ and $\phi(h_n)=C(m)h_n^m$. Then it can be shown directly that

$$h_n^m\sigma^2_{n,0}(x)\to\frac{g_2(x)}{f_1(x)}\frac{\int_{\|v\|\leqslant 1}K^2(\|v\|)\,\mathrm{d}v}{\left[\int_{\|v\|\leqslant 1}K(\|v\|)\,\mathrm{d}v\right]^2}. \tag{2.11}$$

Simple algebra shows that

$$\int_{\|v\|\leqslant 1} K^j(\|v\|)\,\mathrm{d}v = C(m)\,C_j = C(m)\left(m\int_0^1 K^j(u)u^{m-1}\,\mathrm{d}u\right), \quad j = 1, 2.$$

Theorems 1 and 2 (respectively 3) imply the convergence in probability of the estimate $\hat{r}(x)$.

Corollary 1. *Let Conditions* 1–4 (*or Conditions* 1′, 2, 3′, *and* 4) *hold and* $n\phi(h_n)/\log_2 n \to \infty$ *as* $n \to \infty$. *Then,*

$$\left(\frac{n\phi(h_n)}{\log_2 n}\right)^{1/2} [\hat{r}(x) - r(x) - B_n(x)] \xrightarrow{P} 0 \quad \text{as } n \to \infty$$

where the "bias" term $B_n(x) = O(h_n^{\beta})$ *and* $\log_2 n := \log\log n$.

3. Asymptotic normality

In this section we establish the asymptotic normality of the regression estimate $\hat{r}(x)$ of (1.1). We first state the asymptotic normality of $Q_n(x)$ of (2.8). We remark that by Theorem 2 we have for large n

$$\frac{c_8}{f_1(x)\phi(h_n)} \leqslant \sigma_{n,0}^2(x) \leqslant \frac{c_9}{f_1(x)\phi(h_n)}. \tag{3.1}$$

This result is sufficient to establish a normalized central limit theorem of the form

$$n^{1/2}\frac{Q_n(x)}{\sigma_{n,0}(x)} \xrightarrow{L} N(0,1) \tag{3.2}$$

when the response variable Y_i is bounded. Such an assumption may be viewed by some as being restrictive (even though the bound A_1 in $|Y_i| \leqslant A_1 < \infty$ can be arbitrarily large). We therefore proceed to establish in this section a central limit theorem that avoids this restriction by utilizing the result of Theorem 3 which states that

$$\phi(h_n)\sigma_{n,0}^2(x) \to \frac{C_2}{C_1^2}\frac{g_2(x)}{f_1(x)} =: \sigma^2(x), \quad x \in \mathscr{H} \tag{3.3}$$

whenever $f_1(x) > 0$.

Condition 5 below is an assumption on the rate of decay of the mixing coefficient $\alpha(j)$.

Condition 5. Let $h_n \to 0$ and $n\phi(h_n) \to \infty$ as $n \to \infty$. Let $\{v_n\}$ be a sequence of positive integers satisfying $v_n \to \infty$ such that $v_n = o((n\phi(h_n))^{1/2})$ and $(n/\phi(h_n))^{1/2}\alpha(v_n) \to 0$ as $n \to \infty$.

Theorem 4. *Under Conditions* 1′, 2, 3′, 4, *and* 5 *we have as* $n \to \infty$,

$$(n\phi(h_n))^{1/2}\ Q_n(x) \xrightarrow{L} N(0, \sigma^2(x)),$$

where $\sigma^2(x)$ *is defined in* (3.3). *Finally we establish the asymptotic normality of* $\hat{r}(x)$.

Theorem 5. *Under Conditions* 1′, 2, 3′, 4, *and* 5, *we have as* $n \to \infty$,

$$(n\phi(h_n))^{1/2}\ (\hat{r}(x) - r(x) - B_n(x)) \xrightarrow{L} N(0, \sigma^2(x)).$$

If one imposes a stronger assumption on the bandwidth h_n, then one can remove the bias term $B_n(x)$ from Theorem 5.

Corollary 2. *If in addition to the assumptions of Theorem* 5, *the bandwidth parameter* h_n *satisfies* $nh_n^{2\beta}\phi(h_n) \to 0$ *as* $n \to \infty$, *then*

$$(n\phi(h_n))^{1/2}\ (\hat{r}(x) - r(x)) \xrightarrow{L} N(0, \sigma^2(x)).$$

Remark 3. We remark on the conditions imposed on the mixing coefficient $\alpha(j)$ under which Theorem 5 holds. These are Conditions 4 and 5. Let $\alpha(j) = O(j^{-a})$ for some $a > 0$. Then Condition 4 is satisfied if $a > (2 - 2/v)/(1 - 2/v)$. Now select the small block $v_n = (n\phi(h_n))^{1/2}/\log n$. Since $n\phi(h_n) \to \infty$, suppose $\phi(h_n) = n^{-c}$ for some $0 < c < 1$. Then Condition 5 is satisfied provided $a > (2/c) - 1$. Thus, the strongly mixing coefficient must satisfy

$$\alpha(j) = O(j^{-a}); \quad a > \max\left\{\frac{2}{c} - 1, \frac{2 - 2/v}{1 - 2/v}\right\}.$$

Note that there is a tradeoff between the moment order v in the assumption $E|Y_i|^v < \infty$ and the decay rate of the mixing coefficient $\alpha(j)$: the larger v is, the weaker the decay of $\alpha(j)$. Also note that if $\alpha(j)$ decays exponentially fast, $\alpha(j) = e^{-aj}$, then Conditions 4 and 5 are automatically satisfied.

4. Discussion

We discuss in this section the ramifications of the results of this paper.

When the data $\{Y_i, X_i\}$ is i.i.d., Conditions 2(iii), 3(ii) and 3′(ii), 4, and 5 are clearly not needed and can be dropped.

Central limit theorems are normally used to establish confidence intervals for the estimates. In the context of nonparametric estimation (covariance, spectral density, probability density, regression), the asymptotic variance $\sigma^2(x)$ in the central limit theorem depends on certain functions possibly including the one being estimated; see for example (3.3) where $g_2(x)$ and $f_1(x)$ are unknown a priori and have to be estimated in practice (see also (2.11) for the case $\mathscr{H} = \mathbb{R}^m$). This situation is classical regardless of whether the data is i.i.d or dependent. As a consequence, only approximate confidence intervals can be obtained in practice even when $\sigma^2(x)$ is functionally specified. The usage is as follows: by Corollary 2, we have for every $\varepsilon > 0$ and large n

$$P[|\hat{r}(x) - r(x)| \leqslant \varepsilon] \sim 2\Phi\left(\frac{\varepsilon\sqrt{n\phi(h_n)}}{\sigma(x)}\right) - 1.$$

Let $\hat{\sigma}^2(x)$ be any consistent estimate of $\sigma^2(x)$. Then approximate confidence intervals can be obtained from

$$P[|\hat{r}(x) - r(x)| \leqslant \varepsilon] \sim 2\Phi\left(\frac{\varepsilon\sqrt{n\phi(h_n)}}{\hat{\sigma}(x)}\right) - 1.$$

Equivalently, as was pointed out by a referee, one can use the normalized central limit theorem (3.2) to also obtain an approximate confidence interval: Let $\hat{\sigma}^2_{n,0}$ be a consistent estimate of $\sigma^2_{n,0} = \mathrm{var}[Z_{n,1}]/E^2[\Delta_1(x)]$. Then

$$P[|\hat{r}(x) - r(x)| \leqslant \varepsilon] \sim 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\hat{\sigma}_{n,0}(x)}\right) - 1.$$

Thus, in practice, the computation of confidence intervals requires the estimation of the asymptotic variance regardless of whether the structure of $\sigma^2(x)$ is specified or not.

Next we examine the issue of the curse of dimensionality. It was noted earlier that when $\mathscr{H} = \mathbb{R}^m$, then $\phi(h) = C(m)h^m$ and the central limit theorem has the form given in Theorem 5 with convergence rate $(nh_n^m)^{1/2}$. When $\mathscr{H}$ is infinite dimensional, $\phi(h)$ could decrease to zero as $h \to 0$ exponentially fast and the convergence rate becomes effectively $(n\phi(h_n))^{1/2}$. How does one mitigate the curse of dimensionality? This issue was addressed by Gasser et al. [7], in the context of density estimation in functional space, who suggested employing finite dimensional approximations. We adopt their suggestion: Let $\mathscr{H}$ be a separable Hilbert space and let $\{e_j\}$ be an orthonormal basis. Approximate X_i and x via the expansions

$$\tilde{X}_i = \sum_{j=1}^{m} X_{i,j}e_j, \quad \tilde{x} = \sum_{j=1}^{m} x_j e_j,$$

where

$$X_{i,j} := (X_i, e_j), \quad x_j := (x, e_j).$$

It is then clear that the problem becomes finite dimensional in $\mathbb{R}^m$ with $\tilde{\mathbf{X}}_i = (X_{i,1}, \ldots, X_{i,m}), \tilde{\mathbf{x}} = (x_1, \ldots, x_m)$ and Theorem 5 is applicable with $\sigma^2(x)$ given by (2.11).

We finally provide an application in the context of prediction of real-valued continuous-time stationary processes: Let $\{X(t), t \geqslant 0\}$ be a zero mean stationary mean-square continuous random process. Let $\mathscr{H} = L_2[0,1]$. Define

$$X_i := \{X(i+t), 0 \leqslant t \leqslant 1\}, \quad i = 0, \ldots, n.$$

Define for $0<\tau\leqslant 1$

$$Y_i := X((i+1)+\tau).$$

Then by stationarity

$$r(x) := E[Y_i|X_i = x] = E[X(1+\tau)|X(t) = x(t),\ 0\leqslant t\leqslant 1],$$

which is the predictor of $X(1+\tau)$ from $\{X(t), 0\leqslant t\leqslant 1\}$. Note that

$$d(x, X_i) = \left\{\int_0^1 [x(t) - X(i+t)]^2\, \mathrm{d}t\right\}^{1/2}.$$

We now estimate $r(x)$ for each $x \in L_2[0,1]$ by (1.1) and the results of the paper are then applicable. The problem is clearly infinitely dimensional. We now consider a suitable reduction to a vector valued problem: let $\{e_j(t)\}$ be an orthonormal basis in $L_2[0,1]$; for example, the eigenfunctions satisfying

$$\lambda_j e_j(t) = \int_0^1 R(t-s)e_j(s)\, \mathrm{d}s,$$

where R is the covariance matrix of the process and the eigenvalues λ_j are arranged to be non-increasing. Then $\{e_j(t)\}$ are orthonormal in $L_2[0,1]$; they are also complete if R is positive definite [the Karhunen–Loeve expansion]. Since $E\int_0^1 X^2(t)\,\mathrm{d}t = \sum_{j=1}^{\infty}\lambda_j$, one can retain the first m largest eigenfunctions leading to the finite approximation

$$\tilde{X}(t) = \sum_{j=1}^{m} X_j e_j(t), \quad 0\leqslant t\leqslant 1,$$

where

$$X_j = \int_0^1 X(t)e_j(t)\, \mathrm{d}t.$$

5. Derivations

Proof of Theorem 1.

$$\hat{r}_1(x) = \frac{1}{nE[\Delta_1(x)]}\sum_{i=1}^{n}\Delta_i(x); \quad \Delta_i(x) = K(d(x, X_i)/h).$$

By stationarity of order one of the X_i's,

$$E[\hat{r}_1(x)] = 1. \tag{5.1}$$

Next consider

$$
\begin{aligned}
\operatorname{var}[\hat{r}_1(x)] &= \frac{1}{n^2E^2[\varDelta_1(x)]}\sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{cov}\{\varDelta_i(x),\varDelta_j(x)\} \\
&= \frac{1}{nE^2[\varDelta_1(x)]}\operatorname{var}[\varDelta_1(x)] + \frac{1}{n^2E^2[\varDelta_1(x)]}\sum_{\substack{i=1 \\ |i-j|>0}}^{n}\sum_{j=1}^{n}\operatorname{cov}\{\varDelta_i(x),\varDelta_j(x)\} \\
&=: J_1 + J_2.
\end{aligned}
\tag{5.2}
$$

Now,

$$
J_1 = \frac{1}{n}\frac{E[\varDelta_1^2(x)]}{E^2[\varDelta_1(x)]} - \frac{1}{n}
$$

and by Conditions 1 and 3(i),

$$
\frac{c_1^2 c_5 f_1(x)\phi(h_n)}{nE^2[\varDelta_1(x)]} - \frac{1}{n} \leqslant J_1 \leqslant \frac{c_2^2 c_6 f_1(x)\phi(h_n)}{n E^2[\varDelta_1(x)]} - \frac{1}{n}.
$$

Also, by Conditions 1 and 3(i),

$$
c_1^j c_5 f_1(x)\phi(h_n) \leqslant E[\varDelta_1^j(x)] \leqslant c_2^j c_6 f_1(x)\phi(h_n), \quad j=1,2 \tag{5.3}
$$

so that

$$
\frac{(c_1^2 c_5)/(c_2 c_6)^2}{nf_1(x)\phi(h_n)} - \frac{1}{n} \leqslant J_1 \leqslant \frac{(c_2^2 c_6)/(c_1 c_5)^2}{nf_1(x)\phi(h_n)} - \frac{1}{n},
$$

whenever $f_1(x) > 0$. Since $\phi(h_n) \to 0$ as $n \to \infty$, we have for large n,

$$
\frac{const1}{n\, f_1(x)\phi(h_n)} \leqslant J_1 \leqslant \frac{const2}{n\, f_1(x)\phi(h_n)}. \tag{5.4}
$$

Alternatively, under Condition 1′ and 3′(i), we have as $n \to \infty$,

$$
\begin{aligned}
\frac{1}{\phi(h_n)}E[\varDelta_1^j(x)] &= \frac{1}{\phi(h_n)}\int_0^{h_n} K^j(u/h_n)F(\mathrm{d}u;x) \\
&\sim \frac{f_1(x)}{\phi(h_n)/h_n}\int_0^1 K^j(u)\phi'(h_n u)\,\mathrm{d}u \to f_1(x)C_j, \quad j=1,2.
\end{aligned}
\tag{5.5}
$$

It follows that

$$
n\phi(h_n)J_1 \to \frac{C_2}{C_1^2 f_1(x)} \quad \text{as } n\to\infty \tag{5.4a}
$$

whenever $f_1(x) > 0$. For J_2 decompose the sum in (5.2)

$$J_2 = \frac{1}{n^2 E^2[\Delta_1(x)]} \left\{ \sum_{\substack{i=1 \\ 0<|i-j|\leqslant a_n}}^{n} \sum_{j=1}^{n} \operatorname{cov}\{\Delta_i(x), \Delta_j(x)\} + \sum_{\substack{i=1 \\ |i-j|>a_n}}^{n} \sum_{j=1}^{n} \operatorname{cov}\{\Delta_i(x), \Delta_j(x)\} \right\}$$
$$=: J_{2,1} + J_{2,2}, \tag{5.6}$$

where $a_n = o(n)$ at a rate specified below. For $J_{2,1}$ we have by Conditions 1 and 3(ii) (or Conditions 1′(i) and 3′(ii))

$$\begin{aligned}\operatorname{cov}\{\Delta_i(x), \Delta_j(x)\} &= E[\Delta_i(x)\Delta_j(x)] - E^2[\Delta_1(x)] \\ &\leqslant c_2^2 \sup_{i \neq j} P[(X_i, X_j) \in B(x,h) \times B(x,h)] + E^2[\Delta_1(x)] \\ &\leqslant c_2^2 f_2(x)\psi(h_n) + E^2[\Delta_1(x)].\end{aligned}$$

Hence by (5.6),

$$J_{2,1} \leqslant \frac{const.\, f_2(x)\psi(h_n) + E^2[\Delta_1(x)]}{n^2 E^2[\Delta_1(x)]}\, n a_n = \frac{const.\, f_2(x)\psi(h_n) a_n}{n E^2[\Delta_1(x)]} + \frac{a_n}{n},$$

where *const*. is a generic finite positive constant. Using either the lower bound (5.3), or the asymptotic value (5.5), of $E[\Delta_1(x)]$,

$$J_{2,1} \leqslant \frac{const.\, f_2(x)\psi(h_n) a_n}{n f_1^2(x)\phi^2(h_n)} + \frac{a_n}{n}. \tag{5.7}$$

Using the lower bound on J_1 in (5.4), or its asymptotic value in (5.4a), we obtain

$$\frac{J_{2,1}}{J_1} \leqslant const. \frac{f_2(x)}{f_1(x)} \frac{\psi(h_n) a_n}{\phi(h_n)} + a_n \phi(h_n) f_1(x). \tag{5.8}$$

We shall subsequently select a_n such that the above bound tends to zero as $n \to \infty$. Now consider $J_{2,2}$. By Davydov's lemma [9, Corollary A.2],

$$\operatorname{cov}\{\Delta_i(x), \Delta_j(x)\} \leqslant 8 \left\{ E|\Delta_i(x)|^v \right\}^{2/v} [\alpha(|i-j|)]^{1-2/v}.$$

Now by Conditions 1 and 3(i), or Conditions 1′(i) and 3′(i), $E|\Delta_i(x)|^v \leqslant const.\, f_1(x)\phi(h_n)$. Thus,

$$\operatorname{cov}\{\Delta_i(x), \Delta_j(x)\} \leqslant const.\, f_1^{2/v}(x)\{\phi(h_n)\}^{2/v}[\alpha(|i-j|)]^{1-2/v}.$$

It follows that

$$J_{2,2} \leqslant \frac{const. f_1^{2/v}(x)[\phi(h_n)]^{2/v}}{n^2 E^2[\Delta_1(x)]} \sum_{\substack{i=1 \\ |i-j|>a_n}}^{n} \sum_{j=1}^{n} [\alpha(|i-j|)]^{1-2/v}.$$

Using the lower bound (5.3) or the asymptotic value (5.5) for $E[\Delta_1(x)]$ and reducing the double sum above to a single sum, we find that

$$J_{2,2} \leqslant \frac{const.}{na_n^\delta[\phi(h_n)]^{2(1-1/v)} f_1^{2(1-1/v)}(x)} \sum_{l=a_n+1}^{\infty} l^\delta[\alpha(l)]^{1-2/v}.$$

Using the lower bound on J_1 in (5.4), or its asymptotic value in (5.4a),

$$\frac{J_{2,2}}{J_1} \leqslant \frac{const.}{a_n^\delta[\phi(h_n)]^{(1-2/v)} f_1^{1-2/v}(x)} \sum_{l=a_n+1}^{\infty} l^\delta[\alpha(l)]^{1-2/v}. \tag{5.9}$$

Now select a_n as $a_n := 1/[\phi(h_n)]^{(1-2/v)/\delta}$. Then by Condition 4,

$$\frac{J_{2,2}}{J_1} \to 0 \text{ as } n \to \infty. \tag{5.10}$$

With this choice of a_n, Eq. (5.8) becomes

$$\frac{J_{2,1}}{J_1} \leqslant const.\frac{f_2(x)}{f_1(x)} \frac{\psi(h_n)}{\phi^2(h_n)} \phi(h_n)a_n + [\phi(h_n)]^{1-(1-2/v)/\delta}.$$

The first term on the right side tends to zero since $\psi(h)/\phi^2(h)$ is assumed bounded and $\phi(h_n)a_n \to 0$ since $(1-2/v)/\delta < 1$. The second term tends to zero since $(1-2/v)/\delta < 1$. □

Proof of Theorem 2. We first obtain a bound on the rate of convergence of μ_n of (2.6).

$$\mu_n(x) := E[(Y_i - r(x))\Delta_i(x)]. \tag{5.11}$$

Conditioning on X_i,

$$\mu_n(x) = E[(r(X_i) - r(x))\Delta_i(x)]$$

and using Condition 2(i),

$$\mu_n(x) \leqslant \sup_{u \in B(x,h)} |r(u) - r(x)| E[\Delta_1(x)] \leqslant const.\ h^\beta |E[\Delta_1(x)]. \tag{5.12}$$

Now

$$\begin{aligned} n\,\mathrm{var}[Q_n(x)] &= \frac{1}{E^2[\Delta_1(x)]} \mathrm{var}[Z_{n,1}(x)] \\ &\quad + \frac{1}{nE^2[\Delta_1(x)]} \sum_{\substack{i=1 \\ |i-j|>0}}^{n} \sum_{j=1}^{n} \mathrm{cov}\{Z_{n,i}(x), Z_{n,j}(x)\} \\ &=: I_1 + I_2 \end{aligned} \tag{5.13}$$

and note that $I_1 = \sigma_{n,0}^2(x)$. By (5.12), we have

$$\sigma_{n,0}^2(x) = \frac{1}{E^2[\Delta_1(x)]} E[(Y_1 - r(x))^2 \Delta_1^2(x)] + O(h_n^{2\beta}).$$

Conditioning on X_1,

$$\sigma_{n,0}^2(x) = \frac{1}{E^2[\Delta_1(x)]} E[g_2(X_1)\Delta_1^2(x)] + \frac{E[(r(X_1) - r(x))^2 \Delta_1^2(x)]}{E^2[\Delta_1(x)]} + O(h_n^{2\beta}).$$

Using Condition 2(i) the second term is $O(h_n^{2\beta})$. We now establish upper and lower bounds on $\sigma_{n,0}^2(x)$. Write

$$\begin{aligned} E[g_2(X)\Delta_1^2(x)] &= g_2(x)E[\Delta_1^2(x)] + E[(g_2(X_1) - g_2(x))\Delta_1^2(x)] \\ &=: I_{1,1} + I_{1,2}. \end{aligned} \tag{5.14}$$

By Condition 2(ii)

$$|I_{1,2}| \leqslant \sup_{\{u:d(x,u)\leqslant h\}} |g_2(u) - g_2(x)| E[\Delta_1^2(x)] = o(1)E[\Delta_1^2(x)],$$

whereas $I_{1,1} = g_2(x)E[\Delta_1^2(x)]$. Thus, $E[g_2(X)\Delta_1^2(x)] = g_2(x)(1 + o(1))E[\Delta_1^2(x)]$. It follows that

$$\sigma_{n,0}^2(x) = g_2(x)(1 + o(1)) \frac{E[\Delta_1^2(x)]}{E^2[\Delta_1(x)]} + O(h_n^{2\beta}).$$

By (5.3), there exist positive constants c_8 and c_9 such that

$$c_8 \frac{g_2(x)}{f_1(x)} + O(\phi(h_n)h_n^{2\beta}) \leqslant \phi(h_n)\sigma_{n,0}^2(x) \leqslant c_9 \frac{g_2(x)}{f_1(x)} + O(\phi(h_n)h_n^{2\beta}), \tag{5.15}$$

which proves Part (a) of the theorem. To prove Parts (b) and (c) we consider next the contribution of the term I_2 defined in (5.13). Split the sum as follows:

$$\begin{aligned} I_2 = \frac{1}{nE^2[\Delta_1(x)]} &\left\{ \sum_{\substack{i=1 \\ 1\leqslant|i-j|\leqslant a_n}}^{n} \sum_{j=1}^{n} \operatorname{cov}\{Z_{n,i}(x), Z_{n,j}(x)\} \right. \\ &\left. + \sum_{\substack{i=1 \\ |i-j|>a_n}}^{n} \sum_{j=1}^{n} \operatorname{cov}\{Z_{n,i}(x), Z_{n,j}(x)\} \right\} \\ =: I_{2,1} + I_{2,2}, \end{aligned} \tag{5.16}$$

where $a_n = o(n)$ at a rate specified in the sequel. For $I_{2,1}$,

$$\operatorname{cov}\{Z_{n,i}(x), Z_{n,j}(x)\} = E\big[(Y_i - r(x))(Y_j - r(x))\Delta_i(x)\Delta_j(x)\big] - \mu_n^2.$$

Conditioning on (X_i, X_j) and using Condition 2(iii),

$$\operatorname{cov}\{Z_{n,i}(x), Z_{n,j}(x)\} = E[g(X_i, X_j, ; x)\Delta_i(x)\Delta_j(x)] - \mu_n^2.$$

By Condition 1 (upper bound) and Condition 2(iii), there exists a finite constant such that

$$|\operatorname{cov}\{Z_{n,i}(x), Z_{n,j}(x)\}| \leqslant const. \sup_{i\neq j} P[(X_i, X_j) \in B(x,h) \times B(x,h)] + \mu_n^2.$$

By Condition 3(ii) and (5.12),

$$|\mathrm{cov}\{Z_{n,i}(x), Z_{n,j}(x)\}| \leqslant const. f_2(x)\,\psi(h_n) + O(h_n^{2\beta})E^2[\Delta_1(x)].$$

Thus, using (5.16),

$$\begin{aligned}|I_{2,1}| &\leqslant \frac{const. f_2(x)\,\psi(h_n) + O(h_n^{2\beta})E^2[\Delta_1(x)]}{nE^2[\Delta_1(x)]} \sum_{\substack{i=1 \\ 1\leqslant|i-j|\leqslant a_n}}^{n} \sum_{j=1}^{n} 1 \\ &\leqslant \frac{const. f_2(x)\,\psi(h_n)a_n}{E^2[\Delta_1(x)]} + O(h^{2\beta})a_n.\end{aligned}$$

Finally, using the lower bound in (5.3),

$$|I_{2,1}| \leqslant \frac{const. f_2(x)\,\psi(h_n)a_n}{f_1^2(x)\phi^2(h_n)} + O(h_n^{2\beta})a_n. \tag{5.17}$$

It now follows from the lower bound on $\sigma_{n,0}^2$ in (5.15) that

$$\frac{|I_{2,1}(x)|}{\sigma_{n,0}^2(x)} \leqslant const. \frac{f_2(x)}{f_1(x)g_2(x)} \frac{\psi(h_n)a_n}{\phi(h_n)} + const. \frac{f_1(x)}{g_2(x)} O(h_n^{2\beta})\phi(h_n)a_n. \tag{5.18}$$

We shall subsequently select a_n to make the right side of (5.18) tend to zero as $n \to \infty$. Now consider the contribution of $I_{2,2}$ of (5.16). By Davydov's lemma (Hall and Heyde [9], Corollary A.2),

$$|\mathrm{cov}\{Z_{n,i}(x), Z_{n,j}(x)\}| \leqslant 8\{E|(Y_i - r(x))\Delta_i(x)|^v\}^{2/v}[\alpha(|i-j|)]^{1-2/v}.$$

By Condition 1 (upper bound) and the continuity of g_v in Condition 2(ii),

$$E|(Y_i - r(x))\Delta_i(x)|^v = E|g_v(X_i)\Delta_i(x)|^v \leqslant const.\, P[X_i \in B(x,h)]$$

and by Condition 3(i) (upper bound),

$$|\mathrm{cov}\{Z_{n,i}(x), Z_{n,j}(x)\}| \leqslant const.\, \{f_1(x)\phi(h_n)\}^{2/v}[\alpha(|i-j|)]^{1-2/v}.$$

It then follows from (5.16) that

$$I_{2,2} \leqslant \frac{const. f_1^{2/v}(x)[\phi(h_n)]^{2/v}}{nE^2[\Delta_1(x)]} \sum_{\substack{i=1 \\ |i-j|>a_n}}^{n} \sum_{j=1}^{n} [\alpha(|i-j|)]^{1-2/v}.$$

Using the lower bound (5.3) for $E[\Delta_1(x)]$ and reducing the double sum above into a single sum, we find that

$$I_{2,2} \leqslant \frac{const.}{a_n^\delta f_1^{2(1-1/v)}(x)[\phi(h_n)]^{2(1-1/v)}} \sum_{l=a_n+1}^{\infty} l^\delta[\alpha(l)]^{1-2/v}. \tag{5.19}$$

Now using the lower bound (5.15) on $\sigma_{n,0}^2$, we obtain

$$\frac{I_{2,2}}{\sigma_{n,0}^2(x)} \leqslant \frac{const.}{a_n^\delta g_2(x) f_1^{1-2/v}(x)[\phi(h_n)]^{(1-2/v)}} \sum_{l=a_n+1}^{\infty} l^\delta[\alpha(l)]^{1-2/v}. \tag{5.20}$$

Now select a_n as $a_n := \frac{1}{[\phi(h_n)]^{(1-2/v)/\delta}}$. Then by Condition 4,

$$\frac{I_{2,2}}{\sigma_{n,0}^2(x)} \to 0 \quad \text{as } n \to \infty. \tag{5.21}$$

Now Eq. (5.18) can be written as

$$\frac{I_{2,1}}{\sigma_{n,0}^2} \leqslant const. \frac{f_2(x)}{f_1(x)g_2(x)} \frac{\psi(h_n)}{\phi^2(h_n)} \phi(h_n)a_n + \frac{f_1(x)}{g_2(x)} O(h_n^{2\beta})\phi(h_n)a_n.$$

The first term on the right side tends to zero since $\psi(h)/\phi^2(h)$ is assumed bounded and $\phi(h_n)a_n \to 0$ with the choice of a_n above. The second term clearly tends to zero as $n \to \infty$. $\square$

Proof of Theorem 3. It is seen from the proof of Theorem 2 that the dominant term for $\sigma_{n,0}^2(x)$ is given by

$$g_2(x) \frac{E[\Delta_1^2(x)]}{E^2[\Delta_1(x)]}.$$

Let $F(u;x) = P[D_i \leqslant u]$. Under Conditions 1' and 3'(i), we have for large n (small h_n)

$$\frac{1}{\phi(h_n)} E[\Delta_1^j(x)] = \frac{1}{\phi(h_n)} \int_0^{h_n} K^j(u/h_n)\, \mathrm{d}F(u;x)$$
$$\sim f_1(x) \frac{1}{\phi(h_n)} \int_0^{h_n} K^j(u/h_n)\phi'(u)\, \mathrm{d}u \to C_j f_1(x), \quad j = 1, 2.$$

It follows that

$$\phi(h_n)\sigma_{n,0}^2(x) \to \frac{C_2}{C_1^2} \frac{g_2(x)}{f_1(x)},$$

which specifies the structure of the asymptotic variance and proves part (a) of the theorem. The proof of part (b) follows the same steps as in the proof of Theorem 2 except that we use the asymptotic value of $\sigma_{n,0}^2(x)$ given above instead of its lower bound. $\square$

Proof of Corollary 1. By (5.1) $E[\hat{r}_1(x)] = 1$ so that by (2.3),

$$B_n(x) = E[\hat{r}_2(x)] - r(x) = \frac{E[Y_1\Delta_1(x)]}{E[\Delta_1(x)]} - r(x) = \frac{\mu_n(x)}{E[\Delta_1(x)]}$$

by (5.11). Thus, by (5.12), $B_n(x) \leqslant const.\ h_n^\beta$. It follows that

$$\hat{r}(x) - r(x) - B_n(x) = \frac{Q_n(x)}{\hat{r}_1(x)} (1 + o_p(1)). \tag{5.22}$$

Now by Theorem 1, $\hat{r}_1(x) \xrightarrow{m.s.} 1$ and by Theorem 2 or Theorem 3, $n\phi(h_n)\ \mathrm{var}[Q_n(x)] \leqslant const.g_2(x)/f_1(x)$. Thus,

$$\left(\frac{n\phi(h_n)}{\log_2 n}\right)^{1/2} [\hat{r}(x) - r(x) - B_n(x)] \xrightarrow{P} 0 \quad \text{as } n \to \infty. \quad \square$$

Proof of Theorem 4. In view of (5.22) and Theorem 1, it suffices to establish the asymptotic normality of $Q_n(x)$. We normalize $Z_{n,i}$ of (2.8) by

$$\tilde{Z}_{n,i}(x) := \frac{Z_{n,i}(x)\phi^{1/2}(h_n)}{E[\Delta_1(x)]}; \ S_n(x) := \sum_{i=1}^{n} \tilde{Z}_{n,i}(x) \tag{5.23}$$

so that

$$\mathrm{var}[\tilde{Z}_{n,i}(x)] = \sigma_{n,0}^2(x)\phi(h_n) \to \sigma^2(x) \text{ as } n \to \infty \tag{5.24}$$

by Theorem 3. Also by Theorem 3,

$$\sum_{i=1}^{n} \sum_{\substack{j=1 \\ |i-j|>0}}^{n} \mathrm{cov}\{\tilde{Z}_{n,i}(x), \tilde{Z}_{n,j}(x)\} = o(n). \tag{5.25}$$

Now,

$$(n\phi(h_n))^{1/2} Q_n(x) = \frac{1}{\sqrt{n}} S_n. \tag{5.26}$$

We thus need to show that

$$\frac{1}{\sqrt{n}}\, S_n \xrightarrow{L} N(0, \sigma^2(x))\ . \tag{5.27}$$

We employ Bernstein's big-block and small-block procedure. Partition the set $\{1, \ldots, n\}$ into $2k_n + 1$ subsets with large blocks of size $u = u_n$ and small blocks of size $v = v_n$ and set

$$k = k_n := \left\lfloor \frac{n}{u_n + v_n} \right\rfloor. \tag{5.28}$$

Condition 5 implies that there exists a sequence of positive integers $\{q_n\}$, $q_n \to \infty$, such that

$$q_n v_n = o((n\phi(h_n))^{1/2}), \quad q_n(n/\phi(h_n))^{1/2}\alpha(v_n) \to 0 \quad \text{as } n \to \infty. \tag{5.29}$$

Now define the large block size as $u_n = \lfloor (n\phi(h_n))^{1/2}/q_n \rfloor$. Then using (5.29) and simple algebra shows that as $n \to \infty$,

$$\frac{v_n}{u_n} \to 0, \quad \frac{u_n}{n} \to 0, \quad \frac{u_n}{(n\phi(h_n))^{1/2}} \to 0, \quad \frac{n}{u_n}\,\alpha(v_n) \to 0. \tag{5.30}$$

Let η_j, ξ_j, ζ_j be defined as follows:

$$\eta_j := \sum_{i=j(u+v)+1}^{j(u+v)+u} \tilde{Z}_{n,i}, \quad 0 \leqslant j \leqslant k-1, \tag{5.31}$$

$$\xi_j := \sum_{i=j(u+v)+u+1}^{(j+1)(u+v)} \tilde{Z}_{n,i}, \quad 0 \leqslant j \leqslant k-1 \tag{5.32}$$

and

$$\zeta_k := \sum_{i=k(u+v)+1}^{n} \tilde{Z}_{n,i}. \tag{5.33}$$

Write

$$S_n = \sum_{j=0}^{k-1} \eta_j + \sum_{j=0}^{k-1} \xi_j + \zeta_k =: S_n' + S_n'' + S_n'''. \tag{5.34}$$

We show that as $n \to \infty$,

$$\frac{1}{n} E[S_n'']^2 \to 0, \qquad \frac{1}{n} E[S_n''']^2 \to 0, \tag{5.35a}$$

$$|E[\exp(\mathrm{i}tn^{-1/2} S_n')] - \prod_{j=0}^{k-1} E[\exp(\mathrm{i}tn^{-1/2}\eta_j)]| \to 0, \tag{5.35b}$$

$$\frac{1}{n} \sum_{j=0}^{k-1} E[\eta_j^2] \to \sigma^2(x), \tag{5.35c}$$

$$\frac{1}{n} \sum_{j=0}^{k-1} E[\eta_j^2 I\{|\eta_j| > \varepsilon\sigma(x)\sqrt{n}\}] \to 0, \tag{5.35d}$$

for every $\varepsilon > 0$. Relation (5.35a) implies that S_n'' and S_n''' are asymptotically negligible, (5.35b) shows that the summands $\{\eta_j\}$ in S_n' are asymptotically independent, and (5.35c)–(5.35d) are the standard Lindeberg–Feller conditions for asymptotic normality of S_n' under independence.

We first establish (5.35a).

$$E[S_n''']^2 = \mathrm{var}\left[\sum_{j=0}^{k-1} \xi_j\right] = \sum_{j=0}^{k-1} \mathrm{var}[\xi_j] + \sum_{\substack{i=0 \\ i \neq j}}^{k-1} \sum_{j=0}^{k-1} \mathrm{cov}\{\xi_i, \xi_j\} =: F_1 + F_2. \tag{5.36}$$

Now by second-order stationarity,

$$\mathrm{var}[\xi_j] = v_n \mathrm{var}[\tilde{Z}_{n,1}] + \sum_{\substack{i=1 \\ i \neq j}}^{v_n} \sum_{j=1}^{v_n} \mathrm{cov}\{\tilde{Z}_{n,i}, \tilde{Z}_{n,j}\} = v_n \sigma^2(x)(1 + o(1)) \tag{5.37}$$

by (5.24) and (5.25). Thus,

$$F_1 = k_n v_n \sigma^2(x)(1+o(1)) \sim \frac{n v_n}{u_n + v_n} \sim \frac{n v_n}{u_n} = o(n), \tag{5.38}$$

by (5.30). Next consider the term F_2 in (5.36). With $\lambda_j = j(u_n + v_n) + u_n$, we have

$$F_2 = \sum_{\substack{i=0 \\ i \neq j}}^{k-1} \sum_{j=0}^{k-1} \sum_{l_1=1}^{v_n} \sum_{l_2=1}^{v_n} \operatorname{cov}\{\tilde{Z}_{n,\lambda_i + l_1}, \tilde{Z}_{n,\lambda_j + l_2}\},$$

but since $i \neq j, |\lambda_i - \lambda_j + l_1 - l_2| \geqslant u_n$, it follows that

$$|F_2| \leqslant \sum_{\substack{i=1 \\ |i-j| \geqslant u_n}}^{n} \sum_{j=0}^{n} \operatorname{cov}\{\tilde{Z}_{n,i}, \tilde{Z}_{n,j}\} = o(n) \tag{5.39}$$

by (5.25). Hence, by (5.36), (5.38), and (5.39), we have

$$\frac{1}{n} E[S_n'']^2 \to 0 \quad \text{as } n \to \infty.$$

By a similar argument we find, using (5.24), (5.25), and (5.30),

$$\begin{aligned}
\frac{1}{n} E[S_n''']^2 &\leqslant \frac{1}{n}[n - k_n(u_n + v_n)] \operatorname{var}[\tilde{Z}_{n,0}] \\
&\quad + \frac{1}{n} \sum_{\substack{i=1 \\ |i-j|>0}}^{n-k_n(u_n+v_n)} \sum_{j=1}^{n-k_n(u_n+v_n)} \operatorname{cov}\{\tilde{Z}_{n,i}, \tilde{Z}_{n,j}\} \\
&\leqslant \frac{u_n + v_n}{n} \sigma^2(x) + o(1) \to 0 \quad \text{as } n \to \infty.
\end{aligned} \tag{5.40}$$

In order to establish (5.35b) we make use of the fact that the processes $\{Y_i, X_i\}$ are strongly mixing and of Volkonskii and Rozanov's lemma stated in the appendix. Note that η_a is $\mathscr{F}_{i_a}^{j_a}$-measurable with $i_a = a(u+v)+1$ and $j_a = a(u+v)+u$. Hence, with $V_j = \exp(\mathrm{i}tn^{-1/2}\eta_j)$, we have

$$\left| E[\exp(\mathrm{i}tn^{-1/2}S_n')] - \prod_{j=0}^{k-1} E[\exp(\mathrm{i}tn^{-1/2}\eta_j)] \right| \leqslant 16\, k_n \alpha(v_n + 1) \sim 16 \frac{n}{u_n} \alpha(v_n + 1),$$

which tends to zero by (5.30). Next we establish (5.35c). By (5.37), with u_n replacing v_n, we have

$$\operatorname{var}[\eta_j] = \operatorname{var}[\eta_0] = u_n \sigma^2(x)(1+o(1)),$$

so that

$$\frac{1}{n} \sum_{j=0}^{k_n - 1} E[\eta_j^2] = \frac{k_n u_n}{n} \sigma^2(x)(1+o(1)) \to \sigma^2(x) \quad \text{as } n \to \infty$$

since $k_n u_n / n \to 1$.

It remains to establish (5.35d). We need to employ a truncation argument since the response variable Y_i is not necessarily bounded. Let

$$a_L(y) = yI\{|y| \leqslant L\}, \tag{5.41}$$

where L is a fixed truncation point. Put

$$r_L(x) = E[a_L(Y_i)|X_i = x]. \tag{5.42}$$

Define

$$Z_{n,i}^L := (a_L(Y_i) - r_L(x))\Delta_i(x) - \mu_n^L, \tag{5.43}$$

where μ_n^L is the mean of the first term on the right side, and

$$\tilde{Z}_{n,i}^L := \frac{Z_{n,i}^L \phi^{1/2}(h_n)}{E[\Delta_1(x)]}, \qquad \sigma_{n,0,L}^2(x) := \frac{\operatorname{var}[Z_{n,i}^L]}{E^2[\Delta_1(x)]},$$

so that for each fixed $L > 0$, we have as in Theorem 3

$$\operatorname{var}[\tilde{Z}_{n,i}^L] = \sigma_{n,0,L}^2(x)\phi(h_n) \to \sigma_L^2(x) \quad \text{as } n \to \infty, \tag{5.44}$$

where

$$\sigma_L^2(x) = \frac{C_2}{C_1^2} \frac{g_{2,L}(x)}{f_1(x)}, \quad g_{2,L}(x) := \operatorname{var}[Y_1 I\{|Y_1| \leqslant L\}|X_1 = x] \tag{5.45}$$

(compare with $g_2(x)$ defined in Condition 2(i)). Finally, set

$$S_n^L := \sum_{i=1}^n \tilde{Z}_{n,i}^L \quad \text{and} \quad \bar{S}_n^L := \sum_{i=1}^n (\tilde{Z}_{n,i} - \tilde{Z}_{n,i}^L) \tag{5.46}$$

and let η_j^L be given by (5.31) with $\tilde{Z}_{n,i}$ replaced by $\tilde{Z}_{n,i}^L$. It is now seen from (5.43) and (5.3) that $\tilde{Z}_{n,i}^L$ is bounded by $|\tilde{Z}_{n,i}^L| \leqslant const./\phi^{1/2}(h_n)$. Thus by (5.31)

$$\max_{0 \leqslant j \leqslant k-1} |\eta_j^L|/\sqrt{n} \leqslant const. \frac{u_n}{(n\phi(h_n))^{1/2}} \to 0$$

by (5.30). Hence when n is large, the set $\{|\eta_j^L| \geqslant \varepsilon\sigma_L(x)\sqrt{n}\}$ becomes an empty set and thus (5.35d) holds. Consequently, (5.35a)–(5.35d) hold for S_n^L so that

$$\frac{1}{n^{1/2}} S_n^L \xrightarrow{L} N(0, \sigma_L^2(x)). \tag{5.47}$$

In order to complete the proof for the general case, it suffices to show

$$\frac{1}{n} \operatorname{var}[\bar{S}_n^L] \to 0 \quad \text{as first } n \to \infty \text{ and then } L \to \infty. \tag{5.48}$$

Indeed,

$$\begin{aligned}
&|E[\exp(itn^{-1/2}S_n)] - \exp(-t^2\sigma^2(x)/2)| \\
&\quad = |E[\exp(itn^{-1/2}(S_n^L + \bar{S}_n^L))] - \exp(-t^2\sigma_L^2(x)/2) \\
&\qquad + \exp(-t^2\sigma_L^2(x)/2) - \exp(-t^2\sigma^2(x)/2)|
\end{aligned}$$

$$\leqslant |E[\exp(\mathrm{i}tn^{-1/2}S_n^L)] - \exp(-t^2\sigma_L^2(x)/2)| + E|\exp(\mathrm{i}tn^{-1/2}\bar{S}_n^L) - 1|$$
$$+ |\exp(-t^2\sigma_L^2(x)/2) - \exp(-t^2\sigma^2(x)/2)|.$$

Letting $n \to \infty$, the first term goes to zero by (5.47) for every $L > 0$; the second term converges to zero by (5.48) as first $n \to \infty$ and then $L \to \infty$; the third term goes to zero as $L \to \infty$ since $\sigma_L^2(x) \to \sigma^2(x)$ as $L \to \infty$ (as $g_{2,L}(x) \to g_2(x)$ as $L \to \infty$; see (5.45)). Therefore, it remains to prove (5.48). Note that $\bar{S}_n^L$ has the same structure as S_n except that Y_i is replace by $Y_i I\{|Y_i| > L\}$. Hence by the argument of Theorem 3

$$\lim_{n\to\infty} \frac{1}{n} \operatorname{var}[\bar{S}_n^L] = \frac{C_2}{C_1^2} \frac{\bar{g}_{2,L}(x)}{f_1(x)},$$

where

$$\bar{g}_{2,L}(x) := \operatorname{var}[Y_1 I\{|Y_1| > L\}|X_1 = x]$$

(compare to $g_2(x)$ in Condition 2(ii)). By dominated convergence the right side converges to 0 as $L \to \infty$. This establishes (5.35d) for the general case and completes the proof of Theorem 4. □

Proof of Theorem 5. The result follows from (5.22), Theorems 4 and 1 noting that $\hat{r}_1(x) \xrightarrow{m.s.} 1$. □

Acknowledgements

The author wishes to express his gratitude to Professor Philippe Vieu, Laboratoire de Statistique et Probabilité,Université Paul Sabatier, Toulouse Cedex, FRANCE, for several helpful discussions on the subject of functional data estimation. The author is particularly grateful for the constructive comments and suggestions for improvements made by the referees and the Editor of the journal.

Appendix A

Lemma (*Volkonskii and Rozanov [22]*). *Let $V_1, \ldots, V_L$ be strongly mixing random variables measurable with respect to the σ-algebras $\mathscr{F}_{i_1}^{j_1}, \ldots, \mathscr{F}_{i_L}^{j_L}$ respectively with $1 \leqslant i_1 < j_1 < i_2 < \cdots < j_L \leqslant n$, $i_{l+1} - j_l \geqslant w \geqslant 1$ and $|V_j| \leqslant 1$ for $j = 1, \ldots, L$. Then*

$$\left| E\left(\prod_{j=1}^{L} V_j\right) - \prod_{j=1}^{L} E(V_j) \right| \leqslant 16(L-1)\alpha(w),$$

where $\alpha(w)$ is the strongly mixing coefficient.

References

[1] A. Araujo, E. Gine, The Central Limit Theorem for Real and Banach Valued Random Variables, Wiley, New York, 1980.

[2] G. Collomb, W. Härdle, Strong uniform convergence rates in robust nonparametric time series analysis and prediction: kernel regression estimation from dependent observations, Stochastic Process. Appl. 23 (1986) 77–89.

[3] J. Fan, E. Masry, Multivariate regression estimation with errors-in-variables: asymptotic normality for mixing processes, Multivariate Anal. 43 (1992) 237–271.

[4] F. Ferraty, P. Vieu, Dimension fractale et estimation de la régression dans des espaces vectoriela semi-normés, Compte Rendus Acad. Sci. Paris 330 (2000) 403–406.

[5] F. Ferraty, P. Vieu, Nonparametric models for functional data, with application in regression, time series prediction and curve discrimination, presented at the International Conference on Advances and Trends in Nonparametric Statistics, Crete, Greece, 2002.

[6] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, Boston, MA, 1990.

[7] T. Gasser, P. Hall, B. Presnell, Nonparametric estimation of the mode of a distribution of random curves, J. Roy. Statist. Soc. Ser. B 60 (1998) 681–691.

[8] L. Györfi, W. Härdle, P. Sarda, Ph. Vieu, Nonparametric Curve Estimation from Time Series, Springer, Berlin, 1989.

[9] P. Hall, C.C. Heyde, Martingale Limit Theory and its Applications, Academic Press, New York, 1980.

[10] W. Härdle, Applied Nonparametric Regression, Cambridge University Press, Boston, 1990.

[11] Y.P. Mack, B.W. Silverman, Weak and strong uniform consistency of kernel regression estimates, Z. Wahrsch. Verw. Gebiete 61 (1982) 405–415.

[12] E. Masry, Strong consistency and rates for deconvolution of multivariate densities of stationary processes, Stochastic Process. Appl. 47 (1993) 53–74.

[13] J. Ramsay, B. Silverman, Functional Data Analysis, Springer, New York, 1997.

[14] J. Ramsay, B. Silverman, Applied Functional Data Analysis: Methods and Case Studies, Springer, New York, 2002.

[15] P.M. Robinson, Nonparametric estimators for time series, J. Time Series Anal. 4 (1983) 185–297.

[16] M. Rosenblatt, Conditional probability density and regression estimates, in: P.R. Krishnaiah (Ed.), Multivariate Analysis II, Academic Press, 1969, pp. 25–31.

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

[18] E.F. Schuster, Joint asymptotic distribution of the estimated regression function at a finite number of distinct points, Ann. Math. Statist. 43 (1972) 84–88.

[19] D. Tjostheim, Non-linear time series: a selective review, Scandinavian J. Statist. 21 (1994) 97–130.

[20] L.T. Tran, Nonparametric function estimation for time series by local average estimators, Ann. Statist. 21 (1993) 1040–1057.

[21] N.N. Vakashnia, V.I. Tarieladze, S.A. Chobanyan, Probability Distributions on Banach Spaces, Boston: D. Reidel Publishing Co, Boston, 1987.

[22] V.A. Volkonskii, Yu.A. Rozanov, Some limit theorems for random functions, Theory Probab. Appl. 4 (1959) 178–197.