ROCKY MOUNTAIN
JOURNAL OF MATHEMATICS
Volume 45, Number 5, 2015

OSCILLATIONS AND MODULI OF CONTINUITY OF KERNEL DENSITY ESTIMATORS UNDER DEPENDENCE

XIAOYONG XIAO AND HONGWEI YIN

ABSTRACT. We provide an almost sure bound for oscillation rates of kernel density estimators for stationary processes, under the predictive dependence measures which are directly related to the data-generating mechanisms of the underlying processes. We also discuss moduli of continuity of the kernel density estimators.

1. Introduction. In this paper, we consider a general class of stationary and causal sequences represented in the form

$$X_n = J(\ldots, \varepsilon_{n-1}, \varepsilon_n), \tag{1.1}$$

where J is a measurable function and $\{\varepsilon_n\}_{n\in\mathbb{Z}}$ are independent and identically distributed (iid) random variables defined on the same probability space $(\Omega, \mathcal{A}, P)$, see e.g., **[4, 7, 9, 13, 16]** among others. As is explained in **[16]**, (1.1) can be taken as a physical system with the input $\mathcal{F}_n = (\ldots, \varepsilon_{n-1}, \varepsilon_n)$, a filter J and the output X_n. Then the dependence can be taken as the degree of dependence of the output X_n on the input $\mathcal{F}_n$, which is a sequence of innovations that drive the system.

Given a stationary sequence $\{X_k\}_{1\le k\le n}$, the empirical distribution function is known as

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i\le x\}}, \quad x \in \mathbb{R}.$$

2010 AMS *Mathematics subject classification.* Primary 60G10, 60F15, Secondary 60G42.

Keywords and phrases. Kernel density estimators, almost sure convergence, predictive dependence measure, moduli of continuity.

This work was supported by the Natural Science Foundation of Jiangxi Province of China (grant No. 20151BAB201021) and the National Natural Science Foundation of China (grant No. 61563033).

Received by the editors on December 13, 2011, and in revised form on August 17, 2013.

DOI:10.1216/RMJ-2015-45-5-1659

We denote by F the cumulative distribution function of X_k and by f the density. For the moduli of continuity for the function

$$\widetilde{G}_n(x) = \sqrt{n}[F_n(x) - F(x)],$$
$$\widetilde{\Delta}_n(b_n) = \sup_{|x-y|\le b_n} |\widetilde{G}_n(x) - \widetilde{G}_n(y)|, \tag{1.2}$$

where the sequence b_n of positive numbers satisfies $b_n \to 0$ and $nb_n \to \infty$ as $n \to \infty$. Plenty of literature exists on the asymptotic behavior of $\widetilde{\Delta}_n(b_n)$, under the condition that $\{X_k\}_{1\le k\le n}$ are iid, see e.g., **[5, 10, 12]**. However, the behavior of $\widetilde{\Delta}_n(b_n)$ has been much less studied under dependence. The commonly adopted framework for dependence is the strong mixing condition. However, Wu **[15]** implemented the new dependence measures proposed in **[14]** and provided an almost sure bound for $\widetilde{\Delta}_n(b_n)$ under short-range dependence.

Another way of studying the distribution of X_k is to consider the kernel density estimator of f. From **[8]**, given the data $X_1, \dots, X_n$, the kernel density estimator of $f(x)$ is defined by

$$f_n(x) = \frac{1}{nb_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{b_n}\right) = \frac{1}{n} \sum_{i=1}^{n} K_{b_n}(x - X_i), \tag{1.3}$$

where the kernel K and the bandwidth b_n satisfy the conditions:

$$\int_{\mathbb{R}} K(u)\, du = 1, \qquad K_{b_n}(x) = \frac{1}{b_n} K\left(\frac{x}{b_n}\right), \tag{1.4}$$
$$b_n \to 0 \quad \text{and} \quad nb_n \to \infty.$$

When $\{X_k\}_{1\le k\le n}$ are iid, Bickel and Rosenblatt **[1]** provided the asymptotic distribution of $\sup_{0\le x\le 1} |f_n(x) - E[f_n(x)]|$. Neumann **[6]** generalized their results to geometrically β-mixing processes. Wu **[14]** obtained asymptotic normality of f_n under short-range dependence.

In this paper, we are interested in the oscillatory behavior of the kernel density estimators $f_n(x) - f(x)$. Moreover, we shall obtain an almost sure bound for the moduli of continuity for the functions

$$G_n(x) = \sqrt{nb_n}\Big[f_n(x) - E[f_n(x)]\Big], \tag{1.5}$$

$$\Delta_n(\delta_n) = \sup_{|x-y|\le \delta_n} |G_n(x) - G_n(y)|, \tag{1.6}$$

under dependence.

2. Conditions and notation. First we introduce some notation. We write $X \in \mathcal{L}^p$ with $p > 0$ for a random variable X, if $\|X\|_p := [E(|X|^p)]^{1/p} < \infty$, and $\|\cdot\| := \|\cdot\|_2$. We say a function g is Hölder continuous on the set A with index $0 < \tau \le 1$ and write $g \in C^\tau(A)$, if there exists a constant $C_g < \infty$ such that $|g(x) - g(y)| \le C_g|x - y|^\tau$ for all $x, y \in A$. We denote by C a constant whose value varies from line to line.

For $i \in \mathbb{Z}$ and $k \in \mathbb{N}$, we define the conditional cumulative distribution function $F_k(x|\mathcal{F}_i) = P(X_{i+k} \le x|\mathcal{F}_i)$, and $f_k(x|\mathcal{F}_i) = \frac{d}{dx}F_k(x|\mathcal{F}_i)$ as the conditional density. Moreover, let the conditional characteristic function

$$\varphi_k(t|\mathcal{F}_i) = E(e^{\sqrt{-1}tX_{i+k}}|\mathcal{F}_i) = \int_{\mathbb{R}} e^{\sqrt{-1}tx} f_k(x|\mathcal{F}_i)\,dx,$$

where $\sqrt{-1}$ is the imaginary unit.

Let $\{\varepsilon_i'\}$ defined on $(\Omega, \mathcal{A}, P)$ be an iid copy of $\{\varepsilon_i\}$. We denote by

$$\mathcal{F}_i^* = (\dots, \varepsilon_{-2}, \varepsilon_{-1}, \varepsilon_0', \varepsilon_1, \dots, \varepsilon_i) \quad \text{for } i \ge 0,$$

$\mathcal{F}_i^* = \mathcal{F}_i$ for $i < 0$, and $X_i^* = J(\mathcal{F}_i^*)$. So $\mathcal{F}_i^*$ (respectively, X_i^*) is a coupled process of $\mathcal{F}_i$ (respectively, X_i) with ε_0 replaced by an iid copy ε_0'. Next, for $p > 1$ and $k \ge 0$, we introduce the $\mathcal{L}^p$ integral distance

$$\overline{\theta}_p(k) = \left[\int_{\mathbb{R}} \|f_{1+k}(x|\mathcal{F}_0) - f_{1+k}(x|\mathcal{F}_0^*)\|_p^p\,dx\right]^{1/p}, \tag{2.1}$$

and the sup-distance

$$\theta_p(k) = \sup_{x \in \mathbb{R}} \|f_{1+k}(x|\mathcal{F}_0) - f_{1+k}(x|\mathcal{F}_0^*)\|_p. \tag{2.2}$$

Note that, if $f_{1+k}(x|\mathcal{F}_0)$ does not depend on ε_0, then $\overline{\theta}_p(k) = 0$ and $\theta_p(k) = 0$. So $\overline{\theta}_p(k)$ and $\theta_p(k)$ measure the contribution of the innovation ε_0 in predicting the future output X_{1+k} given $\mathcal{F}_0$ by perturbing the input via coupling. If $f_1(\cdot|\mathcal{F}_0) \in \mathcal{C}^2$, then we define the

$\mathcal{L}^p$ integral distance $\overline{\psi}_p(k)$ and $\overline{\phi}_p(k)$ on derivatives:

$$\overline{\psi}_p(k) = \left[\int_{\mathbb{R}} \|f'_{1+k}(x|\mathcal{F}_0) - f'_{1+k}(x|\mathcal{F}_0^*)\|_p^p \, dx \right]^{1/p},$$

$$\overline{\phi}_p(k) = \left[\int_{\mathbb{R}} \|f''_{1+k}(x|\mathcal{F}_0) - f''_{1+k}(x|\mathcal{F}_0^*)\|_p^p \, dx \right]^{1/p}. \tag{2.3}$$

These quantities are directly related to the data-generating mechanism of X_k, thus play an important role in the study of asymptotic properties of f_n.

For a real sequence $a = \{a_k\}_{k\in\mathbb{Z}}$, we define

$$S_p(n;a) = \sum_{j\in\mathbb{Z}} \left(\sum_{k=1-j}^{n-j} |a_k| \right)^{\hat{p}}, \tag{2.4}$$

where $\widehat{p} = \min(2,p)$ and $p > 1$. Let $\overline{\theta}_p = \{\overline{\theta}_p(k)\}_{k\in\mathbb{Z}}$, and $\overline{\theta}_p(k) \equiv 0$ for $k < 0$. We similarly define θ_p, $\overline{\psi}_p$ and $\overline{\phi}_p$. Let

$$\overline{\Theta}_p(n) = S_p(n;\overline{\theta}_p), \qquad \Theta_p(n) = S_p(n;\theta_p), \tag{2.5}$$
$$\overline{\Psi}_p(n) = S_p(n;\overline{\psi}_p), \qquad \overline{\Phi}_p(n) = S_p(n;\overline{\phi}_p).$$

These quantities can be used to interpret the degree of dependence in **[16]**.

Note that, for a nonnegative sequence (a_j), let

$$g(u) = \sum_{j\in\mathbb{Z}} a_j e^{\sqrt{-1}ju}, \quad u \in \mathbb{R},$$

be its Fourier transform. Then, by Parseval's identity, we have the Fejér kernel representation

$$\begin{aligned} 2\pi S_2(n;a) &= \int_0^{2\pi} \Big| g(u) \sum_{k=1}^{n} e^{-\sqrt{-1}ku} \Big|^2 du \\ &= \int_0^{2\pi} |g(u)|^2 \frac{\sin^2(nu/2)}{\sin^2(u/2)} \, du. \end{aligned} \tag{2.6}$$

Condition 2.1.

(i) There exists a positive constant $C_0 < \infty$, such that

$$\sup_{x\in\mathbb{R}} f_1(x|\mathcal{F}_0) \le C_0 \quad \text{almost surely.}$$

(ii) For some $\tau > 0$, assume that $K \in \mathcal{C}^\tau(A)$ with bounded support A, and $X_k \in \mathcal{L}^a$ for some $a > 0$.

(iii) Suppose that $\int_{\mathbb{R}} uK(u)\,du = 0$, $f \in \mathcal{C}^2$, and $\sup_x |f''(x)| < \infty$.

Condition (i) implies that X_k has a density f such that $f(x) = E[f_1(x|\mathcal{F}_0)] \le C_0$. From Condition (ii), we see that K is bounded, and $\int_{\mathbb{R}} |u|^\beta |K(u)|\,du < \infty$ for each constant $\beta > 0$.

3. Main results.

3.1. Oscillations of kernel density estimators.

Theorem 3.1. *Assuming conditions* (i)–(iii), *let* $\widetilde{l}(n) = \sqrt{(\log n)}\log\log n$ *and* $\log n = o(nb_n)$.

(i) *If* $\overline{\Theta}_2(n) + \overline{\Psi}_2(n) = O(n^\alpha l(n))$, *where* $1 \le \alpha \le 2$ *and* l *is a slowly varying function, then*

$$\sup_{x\in\mathbb{R}} |f_n(x) - f(x)| = O_{\text{a.s.}}\left(\sqrt{\frac{\log n}{nb_n}} + b_n^2\right) + o_{\text{a.s.}}(n^{\alpha/2-1} l^{1/2}(n)\widetilde{l}(n))\mathbf{1}_{\{\alpha\ge(6/5)\}}. \tag{3.1}$$

(ii) *If* $\overline{\Theta}_2(n) + \overline{\Psi}_2(n) = O(n)$, *then*

$$\sup_{x\in\mathbb{R}} |f_n(x) - f(x)| = O_{\text{a.s.}}\left(\sqrt{\frac{\log n}{nb_n}} + b_n^2\right). \tag{3.2}$$

Note that Wu et al. **[16]** showed that

$$\sup_{x\in\mathbb{R}} |f_n(x) - E[f_n(x)]| = O_{\text{a.s.}}\left(\sqrt{\frac{\log n}{nb_n}} + n^{\alpha/2-1}\widehat{l}(n)\right),$$

where $\widehat{l}(n) = (\log n)^{1/2+\epsilon} l(n)$ if $\alpha > 1$, and $\widehat{l}(n) = (\log n)^{1/2+\epsilon} \sum_{j:2^j\le n} l^{1/2}(2^j)$ if $\alpha = 1$ for $\epsilon > 0$.

Remark 3.2. The order of $\overline{\Theta}_2(n)+\overline{\Psi}_2(n)$ may be a little difficult to distinguish, while, from **[16,** Lemma 1], we see that a sufficient condition of $\overline{\Theta}_2(n)+\overline{\Psi}_2(n)=O(n^\alpha l(n))$ with $1<\alpha<2$ is a condition of long-rang dependence

$$\overline{\theta}_2(k)+\overline{\psi}_2(k)=O(|k|^{-(3-\alpha)/2}l^{1/2}(|k|)), \tag{3.3}$$

or, by Parseval's identity, we have an equivalent condition

$$\left[\int_{\mathbb{R}}(1+t^2)\|\varphi_{1+k}(t|\mathcal{F}_0)-\varphi_{1+k}(t|\mathcal{F}_0^*)\|^2dt\right]^{1/2}=O(|k|^{-(3-\alpha)/2}l^{1/2}(|k|)), \tag{3.4}$$

since

$$\int_{\mathbb{R}}|\varphi_{1+k}(t|\mathcal{F}_0)-\varphi_{1+k}(t|\mathcal{F}_0^*)|^2dt=2\pi\int_{\mathbb{R}}|f_{1+k}(x|\mathcal{F}_0)-f_{1+k}(x|\mathcal{F}_0^*)|^2dx,$$
$$\int_{\mathbb{R}}t^2|\varphi_{1+k}(t|\mathcal{F}_0)-\varphi_{1+k}(t|\mathcal{F}_0^*)|^2dt=2\pi\int_{\mathbb{R}}|f'_{1+k}(x|\mathcal{F}_0)-f'_{1+k}(x|\mathcal{F}_0^*)|^2dx.$$

Similarly, a sufficient condition of $\overline{\Theta}_2(n)+\overline{\Psi}_2(n)=O(n)$ is a condition of short-rang dependence (the cumulative contribution of the input ε_0 in predicting future values $\{X_k\}_{k\geq 1}$ is finite):

$$\sum_{k=0}^{\infty}\left[\overline{\theta}_2(k)+\overline{\psi}_2(k)\right]<\infty, \tag{3.5}$$

or an equivalent condition

$$\sum_{k=0}^{\infty}\left[\int_{\mathbb{R}}(1+t^2)\|\varphi_{1+k}(t|\mathcal{F}_0)-\varphi_{1+k}(t|\mathcal{F}_0^*)\|^2dt\right]^{1/2}<\infty. \tag{3.6}$$

Since

$$\int_{\mathbb{R}}t^4|\varphi_{1+k}(t|\mathcal{F}_0)-\varphi_{1+k}(t|\mathcal{F}_0^*)|^2dt = 2\pi\int_{\mathbb{R}}|f''_{1+k}(x|\mathcal{F}_0)-f''_{1+k}(x|\mathcal{F}_0^*)|^2dx,$$

we have similar comments for $\overline{\Psi}_2(n)+\overline{\Phi}_2(n)$, and, for $\overline{\psi}_2(k)+\overline{\phi}_2(k)$, we only need to replace the term $(1+t^2)$ in (3.4) and (3.6) with $t^2(1+t^2)$.

Note that $E[f_1(x|\mathcal{F}_k)|\mathcal{F}_0] = f_{1+k}(x|\mathcal{F}_0)$ for $k \geq 0$. Then, by Jensen's inequality, we have

$$\begin{aligned}\|f_{1+k}(x|\mathcal{F}_0) - f_{1+k}(x|\mathcal{F}_0^*)\|_p &= \|E[f_1(x|\mathcal{F}_k) - f_1(x|\mathcal{F}_k^*)|\mathcal{F}_0, \mathcal{F}_0^*]\|_p \\ &\leq \|f_1(x|\mathcal{F}_k) - f_1(x|\mathcal{F}_k^*)\|_p.\end{aligned}$$

Similarly, $E[\varphi_1(t|\mathcal{F}_k)|\mathcal{F}_0] = \varphi_{1+k}(t|\mathcal{F}_0)$ for $k \geq 0$, and

$$\|\varphi_{1+k}(t|\mathcal{F}_0) - \varphi_{1+k}(t|\mathcal{F}_0^*)\|_p \leq \|\varphi_1(t|\mathcal{F}_k) - \varphi_1(t|\mathcal{F}_k^*)\|_p.$$

Then the term $\|f_{1+k}(x|\mathcal{F}_0) - f_{1+k}(x|\mathcal{F}_0^*)\|^2$ under the integral of $\overline{\theta}_2(k)$ and $\overline{\psi}_2(k)$ in equations (3.3) and (3.5) can be replaced by $\|f_1(x|\mathcal{F}_k) - f_1(x|\mathcal{F}_k^*)\|^2$, and the term $\|\varphi_{1+k}(t|\mathcal{F}_0) - \varphi_{1+k}(t|\mathcal{F}_0^*)\|^2$ in equations (3.4) and (3.6) can be replaced by $\|\varphi_1(t|\mathcal{F}_k) - \varphi_1(t|\mathcal{F}_k^*)\|^2$.

Remark 3.3. Theorem 3.1 shows the interesting dichotomous phenomenon as follows.

For case (i), given a small b_n, if $\alpha > 6/5$ and

$$b_n = o(n^{(1-\alpha)} l^{-1}(n)(\log\log n)^{-2}),$$

or $1 \leq \alpha < 6/5$ and $b_n = o((\log n/n)^{1/5})$, then the first term $\sqrt{\log n/(nb_n)}$ dominates equation (3.1); however, for a large bandwidth b_n, if $\alpha > 6/5$ and $(n^{-(2-\alpha)/4} l^{1/4}(n)\widetilde{l}^{1/2}(n)) = o(b_n)$, or $1 \leq \alpha < 6/5$ and $((\log n/n)^{1/5}) = o(b_n)$, then the second term b_n^2 dominates, when $\alpha = 6/5$, the behavior of the first two terms depend on the representation of $l(n)$; for a mild bandwidth b_n, if $(n^{-(\alpha-1)} l^{-1}(n)(\log\log n)^{-2}) = o(b_n)$ and $b_n = o(n^{-(2-\alpha)/4} l^{1/4}(n)\widetilde{l}^{1/2}(n))$ for $\alpha \geq 6/5$, then the third term $n^{\alpha/2-1} l^{1/2}(n)\widetilde{l}(n)$ dominates.

For case (ii), if $b_n = o((\log n/n)^{1/5})$, then the first term $\sqrt{\log n/(nb_n)}$ dominates equation (3.2); however, if $((\log n/n)^{1/5}) = o(b_n)$, then the second term b_n^2 dominates. In a word, the overall bound depends on the interplay between the bandwidth b_n and the dependence parameter α.

From the analysis above, we briefly discuss how to choose an appropriate b_n. If we take $b_n \asymp (n^{-1}\log n)^{1/5}$ for $\alpha < 6/5$, then it holds that

$$\sup_{\mathbb{R}} |f_n(x) - f(x)| = O_{\text{a.s.}}(n^{-1}\log n)^{2/5}.$$

This gives the optimal convergence rate $(n^{-1}\log n)^{2/5}$, since Stute **[11]** showed that $(n\log n)^{2/5}\sup_{|x|\le c}|f_n(x)-f(x)|/f(x)$ converges almost surely to a non-zero constant if $\inf_{|x|\le c} f(x)>0$ for iid random variables X_k. For $\alpha\ge 6/5$, if we take a mild bandwidth b_n such that $(n^{-(\alpha-1)}l^{-1}(n)(\log\log n)^{-2})=o(b_n)$ and $b_n = o(n^{-(2-\alpha)/4}l^{1/4}(n)\widetilde{l}^{1/2}(n))$, then we have

$$\sup_{\mathbb{R}}|f_n(x)-f(x)| = o_{\text{a.s.}}(n^{-(1-\alpha/2)}l^{1/2}(n)\widetilde{l}(n)).$$

This gives the infinitesimal order of the lower bound on the right hand side of (3.1).

Next, we introduce

$$H_n(x)=\sum_{i=1}^n[f_1(x|\mathcal{F}_{i-1})-f(x)], \tag{3.7}$$

and rewrite

$$n\{f_n(x)-E[f_n(x)]\}=P_n(x)+Q_n(x),\qquad R_n(x)=E[f_n(x)]-f(x), \tag{3.8}$$

where

$$P_n(x)=\sum_{i=1}^n\{K_{b_n}(x-X_i)-E[K_{b_n}(x-X_i)|\mathcal{F}_{i-1}]\}, \tag{3.9}$$

$$\begin{aligned} Q_n(x) &= \sum_{i=1}^n\{E[K_{b_n}(x-X_i)|\mathcal{F}_{i-1}]-E[K_{b_n}(x-X_i)]\} \\ &= \int_{\mathbb{R}} K_{b_n}(x-u)H_n(u)\,du = \int_{\mathbb{R}} K(u)H_n(x-b_nu)\,du. \end{aligned} \tag{3.10}$$

Lemma 3.4.

(i) *Assume conditions* (i)–(ii) *and* $\log n = o(nb_n)$. *Then*

$$\sup_{x\in\mathbb{R}}|P_n(x)| = O_{\text{a.s.}}(\sqrt{n\log n/b_n}).$$

(ii) *If* $\overline{\Theta}_2(n)+\overline{\Psi}_2(n) = O(n^\alpha l(n))$, *where* $1\le\alpha\le 2$ *and* l *is a slowly varying function, then* $\sup_{x\in\mathbb{R}}|Q_n(x)| = o_{\text{a.s.}}(n^{\alpha/2}l^{1/2}(n)\widetilde{l}(n))$.

(iii) *If* $\overline{\Theta}_2(n)+\overline{\Psi}_2(n) = O(n)$, *then* $\sup_{x\in\mathbb{R}}|Q_n(x)| = o_{\text{a.s.}}(\sqrt{n}\widetilde{l}(n))$.

Proof. Case (i) easily follows from the proof of Theorem 2 in Wu et al. **[16]**. As for case (ii), we define projection operators $\mathcal{P}_k$, by

$$\mathcal{P}_kZ = E(Z|\mathcal{F}_k) - E(Z|\mathcal{F}_{k-1}), \quad Z\in\mathcal{L}^1 \quad \text{and} \quad k\in\mathbb{Z}.$$

Note that $\mathcal{P}_kH_n(x)$ for $k\le n$ are martingale differences and

$$H_n(x) = \sum_{k=-\infty}^{n}\mathcal{P}_kH_n(x).$$

Let $\theta_i(x) = \|f_{1+i}(x|\mathcal{F}_0) - f_{1+i}(x|\mathcal{F}_0^*)\|$. Then $\overline{\theta}_2(i) = \sqrt{(\int_{\mathbb{R}}\theta_i^2(x)\,dx)}$, and we have $\|\mathcal{P}_0f_1(x|\mathcal{F}_i)\|\le\theta_i(x)$. In fact, $\theta_i(x)=0$ for $i\le -1$, and $\|\mathcal{P}_0f_1(x|\mathcal{F}_i)\| = 0$ for $i\le -1$ by their definitions. Next, for $i\ge 0$, we have

$$\begin{aligned}
\|\mathcal{P}_0f_1(x|\mathcal{F}_i)\| &= \|E[f_1(x|\mathcal{F}_i)|\mathcal{F}_0] - E[f_1(x|\mathcal{F}_i)|\mathcal{F}_{-1}]\|\\
&= \|f_{1+i}(x|\mathcal{F}_0) - E\{E[f_1(x|\mathcal{F}_i)|\mathcal{F}_0]|\mathcal{F}_{-1}\}\|\\
&= \|f_{1+i}(x|\mathcal{F}_0) - E[f_{1+i}(x|\mathcal{F}_0)|\mathcal{F}_{-1}]\|\\
&= \|E[f_{1+i}(x|\mathcal{F}_0)|\mathcal{F}_0] - E[f_{1+i}(x|\mathcal{F}_0^*)|\mathcal{F}_0]\|\\
&\le \|f_{1+i}(x|\mathcal{F}_0) - f_{1+i}(x|\mathcal{F}_0^*)\|,
\end{aligned}$$

where we have used the fact that

$$E[f_k(x|\mathcal{F}_0)|\mathcal{F}_{-1}] = E[f_k(x|\mathcal{F}_0^*)|\mathcal{F}_0], \quad \text{for } k\ge 1,$$

and Jensen's inequality.

By Doob's inequality, Burkholder's inequality and then Minkowski's inequality, we have

$$\begin{aligned}
E\Big[\max_{l\le n} H_l^2(x)\Big] &\le 4\|H_n(x)\|^2 \le C_1 \sum_{k=-\infty}^{n} \|\mathcal{P}_k H_n(x)\|^2 \\
&= C_1 \sum_{k=-\infty}^{n} \Big\| \sum_{i=1}^{n} \mathcal{P}_k f_1(x|\mathcal{F}_i)\Big\|^2 \\
&\le C_1 \sum_{k=-\infty}^{n} \Big(\sum_{i=1}^{n} \|\mathcal{P}_k f_1(x|\mathcal{F}_i)\|\Big)^2 \\
&\le C_1 \sum_{k=-\infty}^{n} \Big(\sum_{i=1}^{n} \theta_{i-k}(x)\Big)^2 \\
&= C_1 \sum_{k=-\infty}^{n} \Big(\sum_{i=1-k}^{n-k} \theta_{i}(x)\Big)^2.
\end{aligned}$$

On the other hand, by Hölder's inequality, it holds that

$$\int_{\mathbb{R}} \Big(\sum_{i=1-k}^{n-k} \theta_i(x)\Big)^2 dx \le \int_{\mathbb{R}} \sum_{i=1-k}^{n-k} \frac{\theta_i^2(x)}{\overline{\theta}_2(i)} \Big(\sum_{i=1-k}^{n-k} \overline{\theta}_2(i)\Big)\, dx.$$

Therefore,

$$\begin{aligned}
\int_{\mathbb{R}} E[\max_{l\le n} H_l^2(x)]\, dx &\le C_1 \sum_{k=-\infty}^{n} \Big(\sum_{i=1-k}^{n-k} \overline{\theta}_2(i)\Big) \sum_{i=1-k}^{n-k} \frac{\int_{\mathbb{R}} \theta_i^2(x)\, dx}{\overline{\theta}_2(i)} \\
&= C_1 \sum_{k=-\infty}^{n} \Big(\sum_{i=1-k}^{n-k} \overline{\theta}_2(i)\Big)^2 = O(\overline{\Theta}_2(n)).
\end{aligned}$$

Similarly, we have

$$\int_{\mathbb{R}} E\Big[\max_{l\le n} (H_l'(x))^2\Big]\, dx = O(\overline{\Psi}_2(n)).$$

By [**15,** Lemma 1], we have

$$\sup_{x\in\mathbb{R}} H_l^2(x) \le \int_{\mathbb{R}} H_l^2(x)\, dx + \int_{\mathbb{R}} (H_l'(x))^2 dx.$$

If $\overline{\Theta}_2(n) + \overline{\Psi}_2(n) = O(n^\alpha l(n))$, then

$$\begin{aligned}
&\sum_{d=1}^{\infty} \frac{E[\max_{l\le 2^d} \sup_{x\in\mathbb{R}} H_l^2(x)]}{2^{d\alpha} l(2^d)\widetilde{l}^2(2^d)} \\
&\le \sum_{d=1}^{\infty} \frac{E[\max_{l\le 2^d} \int_{\mathbb{R}} H_l^2(x) + (H_l'(x))^2 dx]}{2^{d\alpha} l(2^d)\widetilde{l}^2(2^d)} \\
&\le \sum_{d=1}^{\infty} \frac{E[\int_{\mathbb{R}} \max_{l\le 2^d} H_l^2(x) + \max_{l\le 2^d}(H_l'(x))^2 dx)]}{2^{d\alpha} l(2^d)\widetilde{l}^2(2^d)} \\
&= \sum_{d=1}^{\infty} \frac{O(2^{d\alpha} l(2^d))}{2^{d\alpha} l(2^d)\widetilde{l}^2(2^d)} = \sum_{d=1}^{\infty} O\left(\frac{1}{d\log^2 d}\right) < \infty.
\end{aligned}$$

By the Borel-Cantelli lemma, as $d \to \infty$,

$$\max_{l\le 2^d} \sup_{x\in\mathbb{R}} |H_l(x)| = o_{\mathrm{a.s.}}(2^{d\alpha/2} l^{1/2}(2^d)\widetilde{l}(2^d)).$$

For any $n \ge 2$, there is a $d \in \mathbb{N}$ such that $2^{d-1} < n \le 2^d$. Since $\max_{l\le n} |H_l(x)| \le \max_{l\le 2^d} |H_l(x)|$, then we are done. Case (iii) easily follows from the proof of case (ii), and we only need to replace $n^\alpha l(n)$ with n. □

From the proof of case (ii) in Lemma 3.4, we easily obtain the uniform bounds of empirical functions under long-rang dependence.

Corollary 3.5. *Assuming condition* (i), *if* $\overline{\Theta}_2(n)+\overline{\Psi}_2(n) = O(n^\alpha l(n))$, *where* $1 \le \alpha \le 2$ *and* l *is a slowly varying function, then*

$$\widetilde{\Delta}_n(b_n) = O_{\mathrm{a.s.}}(\sqrt{b_n \log n}) + o_{\mathrm{a.s.}}(b_n n^{(\alpha-1)/2} l^{1/2}(n)\widetilde{l}(n)), \tag{3.11}$$

where $\widetilde{l}(n) = \sqrt{(\log n)} \log\log n$.

Proof. We have the decomposition

$$\begin{aligned}
\widetilde{G}_n(x) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\mathbf{1}_{\{X_i\le x\}} - F_1(x|\mathcal{F}_{i-1})] + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [F_1(x|\mathcal{F}_{i-1}) - F(x)] \\
&=: \widetilde{G}_{n,1}(x) + \widetilde{G}_{n,2}(x).
\end{aligned}$$

From [**15,** Lemma 3], it holds that

$$\sup_{|x-y|\le b_n} |\widetilde{G}_{n,1}(x) - \widetilde{G}_{n,1}(y)| = O_{\text{a.s.}}(\sqrt{b_n \log n}).$$

On the other hand, when $\overline{\Theta}_2(n) + \overline{\Psi}_2(n) = O(n^\alpha l(n))$, we have

$$\begin{aligned}\sup_{x\in\mathbb{R}} |H_n(x)| &= \sup_{x\in\mathbb{R}} \left| \sum_{i=1}^n [f_1(x|\mathcal{F}_{i-1}) - f(x)] \right| \\ &= o_{\text{a.s.}}(n^{\alpha/2} l^{1/2}(n)\widetilde{l}(n));\end{aligned}$$

therefore,

$$\begin{aligned}\sup_{|x-y|\le b_n} |\widetilde{G}_{n,2}(x) - \widetilde{G}_{n,2}(y)| &\le \frac{b_n}{\sqrt{n}} \sup_{x\in\mathbb{R}} |H_n(x)| \\ &= o_{\text{a.s.}}(b_n n^{(\alpha-1)/2} l^{1/2}(n)\widetilde{l}(n)).\end{aligned}$$

Now we are done, since

$$\widetilde{\Delta}_n(b_n) \le \sup_{|x-y|\le b_n} |\widetilde{G}_{n,1}(x) - \widetilde{G}_{n,1}(y)| + \sup_{|x-y|\le b_n} |\widetilde{G}_{n,2}(x) - \widetilde{G}_{n,2}(y)|. \quad \square$$

Assume that X_k are iid standard uniform random variables, and

$$\log n = o(nb_n), \qquad \log\log n = o(\log b_n^{-1}),$$

Stute [**12**] showed that

$$\lim_{n\to\infty} \frac{\widetilde{\Delta}_n(b_n)}{\sqrt{b_n \log b_n^{-1}}} = \sqrt{2} \quad \text{almost surely.} \tag{3.12}$$

If $b_n + (nb_n)^{-1} = O(n^{-\gamma})$ for some $\gamma > \alpha - 1$, then the bound in (3.11) turns out to be $\sqrt{b_n \log n}$, which is the optimal bound, since it has the same order of magnitude as $\sqrt{b_n \log b_n^{-1}}$.

Lemma 3.6.

(i) *Assume that* $\int_{\mathbb{R}} |uK(u)|\, du < \infty$, $f \in \mathcal{C}^1$ *and* $\sup_{x\in\mathbb{R}} |f'(x)| < \infty$. *Then* $\sup_{x\in\mathbb{R}} |R_n(x)| = O(b_n)$.

(ii) *If condition* (iii) *holds, then* $\sup_{x\in\mathbb{R}} |R_n(x)| = O(b_n^2)$.

Proof. This result is well known, and it easily follows from Taylor's expansion, see also [**16,** Corollary 2]. $\square$

Now we go back to the proof of Theorem 3.1. Since

$$\sqrt{\frac{\log n}{nb_n}} + b_n^2 \geq \frac{5}{4^{4/5}}\left(\frac{\log n}{n}\right)^{2/5},$$

then

$$n^{\alpha/2-1}l^{1/2}(n)\widetilde{l}(n)\mathbf{1}_{\{\alpha<(6/5)\}} = o\left(\sqrt{\frac{\log n}{nb_n}} + b_n^2\right).$$

Moreover,

$$n^{-1/2}\widetilde{l}(n) = o\left(\sqrt{\frac{\log n}{nb_n}} + b_n^2\right).$$

Hence, from Lemmas 3.4 and 3.6, the results of Theorem 3.1 are obvious.

Note that if Theorem 3.1 (iii) is replaced by condition (i) in Lemma 3.6, then the term b_n^2 in (3.1) and (3.2) should be replaced by b_n. Since

$$\sqrt{\frac{\log n}{nb_n}} + b_n \geq \frac{3}{2^{2/3}}\left(\frac{\log n}{n}\right)^{1/3},$$

then the indicator $\mathbf{1}_{\{\alpha\geq(6/5)\}}$ in (3.1) should be replaced by $\mathbf{1}_{\{\alpha\geq(4/3)\}}$.

We mention that Wu et al. **[16]** also showed the $\mathcal{L}^p$ bounds for $f_n(x) - E[f_n(x)]$ under condition (i). Suppose that $\int_{\mathbb{R}} |K(u)|\, du < \infty$ and $\sup_u |K(u)| < \infty$. Then

$$\sup_x \|f_n(x) - f(x)\|_p = O\left(\frac{1}{\sqrt{nb_n}} + \Theta_p^{1/p'}(n)/n\right),$$

where $p' = \min(2, p)$. Moreover, assume condition (iii). We provide the rate of convergence in probability, by Lemma 3.6, that

$$\sup_x |f_n(x) - f(x)| = O_p\left(\frac{1}{\sqrt{nb_n}} + b_n^2 + \Theta_2^{1/2}(n)/n\right). \tag{3.13}$$

At last, according to **[16,** Theorem 3], assuming conditions (i)–(iii), if the kernel K is symmetric and $b_n\Theta_2(n) = o(n)$, then for a fixed point x_0, it holds that

$$\sqrt{nb_n}\{f_n(x_0) - E[f_n(x_0)]\} \xrightarrow{d} N\Big[0, f(x_0)\kappa\Big],$$

where $\xrightarrow{d}$ means convergence in distribution and $\kappa = \int_{\mathbb{R}} K^2(u)\,du$. Moreover, if $\sqrt{n}b_n^{5/2} = o(1)$, due to $\sup_{x\in\mathbb{R}} |R_n(x)| = O(b_n^2)$, then we have

$$\sqrt{nb_n}\{f_n(x_0) - f(x_0)\} \xrightarrow{d} N\Big[0, f(x_0)\kappa\Big].$$

In addition, from (3.13), it holds that

$$\frac{\sqrt{nb_n}\{f_n(x_0) - f(x_0)\}}{\sqrt{f_n(x_0)\kappa}} \xrightarrow{d} Z, \tag{3.14}$$

where Z is the standard normal random variable. Therefore, A confidence interval for the unknown $f(x_0)$ with confidence level $1-\alpha$ becomes

$$\left(f_n(x_0) - \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{f_n(x_0)\kappa}{nb_n}}, f_n(x_0) + \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{f_n(x_0)\kappa}{nb_n}}\right),$$

where $\Phi(\cdot)$ is the standard normal distribution function and n is large enough.

3.2. Moduli of continuity. Wu [**14**] showed asymptotic normality of $f_n(x) - E[f_n(x)]$ for each $x \in \mathbb{R}$. Now we consider moduli of continuity of functions $G_n(x)$ when $\delta_n < C$ is a sequence of positive numbers for a certain constant C.

Theorem 3.7. *Assume conditions* (i)–(ii), *and*

$$\infty \leftarrow (n\delta_n)^\tau \log n = o((nb_n)^{\tau+1/2}). \tag{3.15}$$

Let $\widetilde{l}(n) = \sqrt{(\log n)} \log\log n$.

(i) *If* $\overline{\Psi}_2(n) + \overline{\Phi}_2(n) = O(n^\alpha l(n))$, *where* $1 \le \alpha \le 2$ *and* l *is a slowly varying function, then*

$$\Delta_n(\delta_n) = O_{\text{a.s.}}\left(\left(\frac{\delta_n}{b_n}\right)^\tau \frac{\log n}{\sqrt{nb_n}}\right) + o_{\text{a.s.}}(\delta_n b_n^{1/2} n^{(\alpha-1)/2} l^{1/2}(n)\widetilde{l}(n)). \tag{3.16}$$

(ii) *If* $\overline{\Psi}_2(n) + \overline{\Phi}_2(n) = O(n)$, *then*

$$\Delta_n(\delta_n) = O_{\text{a.s.}}\left(\left(\frac{\delta_n}{b_n}\right)^\tau \frac{\log n}{\sqrt{nb_n}}\right) + o_{\text{a.s.}}(\delta_n b_n^{1/2}\widetilde{l}(n)). \tag{3.17}$$

Proof.

(i) Letting $T_n = \sup_{\{|x-y|\le\delta_n,|y|\ge n^{5/a}\}} |G_n(x) - G_n(y)|$, since $K \in \mathcal{C}^\tau(A)$ with a bounded support A and $X_1 \in \mathcal{L}^a$, we have, by Markov's inequality, that

$$\sqrt{\frac{b_n}{n}} E(|T_n|) \le 4E\left[\sup_{|x|\ge n^{5/a}/2}\left|K\left(\frac{x-X_1}{b_n}\right)\right|\right] = O(1)P\left(|X_1| \ge \frac{n^{5/a}}{4}\right) = \frac{O(1)}{n^5}.$$

By the Borel-Cantelli lemma, we have

$$T_n = o_{\text{a.s.}}(n^{-3}b_n^{-1/2}) = o_{\text{a.s.}}\left(\left(\frac{\delta_n}{b_n}\right)^\tau \sqrt{\frac{\log n}{nb_n}}\right).$$

Write $n\{f_n(x) - E[f_n(x)]\} = P_n(x) + Q_n(x)$ as (3.8). By a similar proof of Lemma 3.4 (ii), we have $\sup_x |Q_n'(x)| = o_{\text{a.s.}}(n^{\alpha/2} l^{1/2}(n)\widetilde{l}(n))$. Then

$$\sup_{|x-y|\le\delta_n} \sqrt{nb_n}|[Q_n(x) - Q_n(y)]/n| = o_{\text{a.s.}}(\delta_n\sqrt{b_n}n^{(\alpha-1)/2}l^{1/2}(n)\widetilde{l}(n)). \tag{3.18}$$

Next, we only have to consider the behavior of $P_n(x) - P_n(y)$ over $y \in [-n^{5/a}, n^{5/a}]$. We define $I_{x,y} = \mathbf{1}_{\{(x,y):|x-y|\le\delta_n\}}$ and

$$Z_i(x,y) = K_{b_n}(x - X_i) - K_{b_n}(y - X_i) - E[K_{b_n}(x - X_i) - K_{b_n}(y - X_i)|\mathcal{F}_{i-1}] \tag{3.19}$$

by the summands of $P_n(x) - P_n(y)$. Let $l = \lfloor n^{1+5/a+1/\tau}\rfloor$ and $\lfloor x\rfloor_l = \lfloor xl\rfloor/l$. Note that $|Z_i I_{x,y}| \le 2C_K\delta_n^\tau b_n^{-(1+\tau)}$, and

$$E[Z_i^2 I_{x,y}|\mathcal{F}_{i-1}] \le E\{[K_{b_n}(x - X_i) - K_{b_n}(y - X_i)]^2|\mathcal{F}_{i-1}\} \le C_K^2\delta_n^{2\tau}b_n^{-2(1+\tau)}.$$

Take $\rho_n = \delta_n^\tau b_n^{-(1+\tau)} \log n$ and $\sqrt{\lambda} = 16C_K(1 + 5/a + 1/\tau)$. By the inequality of Freedman [**3**], we have

$$\begin{aligned} P(|P_n(x) - P_n(y)|I_{x,y} \geq \sqrt{\lambda}\rho_n) \\ &\leq 2 \exp\left(\frac{-\lambda\rho_n^2}{4C_K\delta_n^\tau b_n^{-(1+\tau)}\sqrt{\lambda}\rho_n + 2C_K^2\delta_n^{2\tau}b_n^{-2(1+\tau)}}\right) \\ &= O\Big(n^{-\sqrt{\lambda}/(4C_K)}\Big); \end{aligned}$$

therefore,

$$\begin{aligned} P\left(\max_{|y|\leq n^{5/a}} |P_n(\lfloor x\rfloor_l) - P_n(\lfloor y\rfloor_l)|I_{x,y} > \sqrt{\lambda}\rho_n\right) &= O\left(n^{10/a}l^2 n^{-\sqrt{\lambda}/(4C_K)}\right) \\ &= o(n^{-2}). \end{aligned}$$

On the other hand, since $K \in \mathcal{C}^\tau(A)$, $(n\delta_n)^\tau \log n \to \infty$ and

$$\sup_x |P_n(x) - P_n(\lfloor x\rfloor_l)| = O(nb_n^{-1}[n^{5/a}/(lb_n)]^\tau) = O(\rho_n),$$

then by the Borel-Cantelli lemma, we have

$$\max_{|y|\leq n^{5/a}} |P_n(x) - P_n(y)|I_{x,y} = O_{\text{a.s.}}(\rho_n).$$

Now (3.16) is proved and (3.17) is obvious. □

4. Applications. Next, we apply Theorems 3.1 and 3.7 to the following processes (X_k) with the structure

$$X_k = \varepsilon_k + Y_{k-1}, \tag{4.1}$$

where Y_{k-1} is $\mathcal{F}_{k-1}$ measurable. It is a large class of processes. The widely used linear processes

$$X_k = \sum_{i=0}^{\infty} a_i\varepsilon_{k-i}, \tag{4.2}$$

chains of the form

$$X_k = G(X_{k-1}, X_{k-2}, \ldots) + \varepsilon_k \tag{4.3}$$

with infinite memory, and nonlinear processes of the form

$$X_k = m(X_{k-1}) + \varepsilon_k, \tag{4.4}$$

with a stationary solution all fall within the framework of (4.1). Let $\varphi(t) = E(e^{\sqrt{-1}t\varepsilon_1})$ be the characteristic function of (ε_k).

Theorem 4.1. *Let* $0 < \beta \leq 2$. *For* $X_k = \varepsilon_k + Y_{k-1}$, *assume that*

$$\int_{\mathbb{R}} |\varphi(t)|^2(1+t^2)|t|^\beta dt < \infty. \tag{4.5}$$

(i) *For some* $1 < \alpha < 2$, *if*

$$\|X_k - X_k^*\|_\beta^{\beta/2} = O\big(k^{-(3-\alpha)/2} l^{1/2}(k)\big), \quad \textit{for } k \geq 0, \tag{4.6}$$

then (3.11) *holds. Moreover, assume conditions* (ii)–(iii). *Then equation* (3.1) *holds.*

(ii) *Assume conditions* (ii)–(iii). *If*

$$\sum_{k=0}^{\infty} \|X_k - X_k^*\|_\beta^{\beta/2} < \infty, \tag{4.7}$$

then equation (3.2) *is satisfied.*

Note that Wu [**15**] has studied this process (4.1) under short-rang dependence for empirical processes, and the results above can be obtained by similar analysis. Condition (4.5) is not overly restrictive. Obviously, it is satisfied if $|\varphi(t)| = O(|t|^{-\gamma})$ as $|t| \to \infty$, where $\gamma > (3+\beta)/2$. It is also satisfied for the important symmetric-α-stable distributions with heavy tails. Recall that X_k^* and X_k are identically distributed, and X_k^* is a coupled version of X_k with ε_0 replaced by ε_0'. Then the quantity $\|X_k - X_k^*\|_\beta$ measures the degree of dependence of $J(\dots, \varepsilon_{k-1}, \varepsilon_k)$ on ε_0. In many applications, condition (4.6) or (4.7) is easily verifiable since it is directly related to the data-generating mechanism and the calculation of $\|X_k - X_k^*\|_\beta$ is easy [**14**]. From Remark 3.2, (4.6) is a sufficient condition of $\overline{\Theta}_2(n)+\overline{\Psi}_2(n) = O(n^\alpha l(n))$ with $1 < \alpha < 2$ and a slowly varying function $l(n)$, which suggests long-rang dependence. However, (4.7) is a sufficient condition of $\overline{\Theta}_2(n) + \overline{\Psi}_2(n) = O(n)$, suggesting short-range dependence.

As for the moduli of continuity of functions $G_n(x)$ for the model (4.1), we have the following results whose proofs are similar to Theorem 4.1.

Corollary 4.2. *Let $0 < \beta \le 2$. For $X_k = \varepsilon_k + Y_{k-1}$, assume condition* (ii) *and* (4.5) *are replaced by a stronger condition*:

$$\int_{\mathbb{R}} |\varphi(t)|^2(1+t^2)|t|^{2+\beta}dt < \infty. \tag{4.8}$$

(i) *For some $1 < \alpha < 2$, if* (3.15) *and* (4.6) *hold, then equation* (3.16) *is satisfied.*
(ii) *If* (3.15) *and* (4.7) *hold, then equation* (3.17) *is satisfied.*

Next, we analyze the linear process $X_k = \sum_{i=0}^{\infty} a_i \varepsilon_{k-i}$, where ε_i are iid random variables with density f_ε. Important special cases of linear process include ARMA and fractional ARIMA models. Assume that $\varepsilon_k \in \mathcal{L}^q$ for some $q > 0$, and for $p > 1$,

$$C_2 := \int_{\mathbb{R}} \left[|f_\varepsilon(x)|^p + |f'_\varepsilon(x)|^p + |f''_\varepsilon(x)|^p\right] dx < \infty. \tag{4.9}$$

Let $a_0 = 1$, and $Y_k = X_{k+1} - \varepsilon_{k+1}$ for $k \ge 0$. Then $f_1(x|\mathcal{F}_k) = f_\varepsilon(x - Y_k)$. By Hölder's inequality, since $f_\varepsilon(x+t) - f_\varepsilon(x) = \int_0^t f'_\varepsilon(x+u)\,du$, we have

$$\begin{aligned}\int_{\mathbb{R}} |f_\varepsilon(x+t) - f_\varepsilon(x)|^p dx &\le \int_{\mathbb{R}} \left| |t|^{p-1} \int_0^t |f'_\varepsilon(x+u)|^p du \right| dx \\ &\le C_2 |t|^p.\end{aligned}$$

On the other hand, the above integral is also bounded by $2^p C_2$. By Jensen's inequality, we have

$$\begin{aligned}\|f_{1+k}(x|\mathcal{F}_0) - f_{1+k}(x|\mathcal{F}_0^*)\|_p &= \|E[f_1(x|\mathcal{F}_k)|\mathcal{F}_0] - E[f_1(x|\mathcal{F}_k^*)|\mathcal{F}_0^*]\|_p \\ &\le \|f_\varepsilon(x - Y_k) - f_\varepsilon(x - Y_k^*)\|_p.\end{aligned}$$

Then

$$\begin{aligned}\overline{\theta}_p(k) &\le \left(E \int_{\mathbb{R}} |f_\varepsilon(x - Y_k) - f_\varepsilon(x - Y_k^*)|^p dx\right)^{1/p} \\ &\le \{E[\min(2^p C_2, C_2|a_{k+1}(\varepsilon_0 - \varepsilon_0')|^p)]\}^{1/p} \qquad (4.10) \\ &\le 2(C_2)^{1/p}\{E[|a_{k+1}(\varepsilon_0 - \varepsilon_0')|^{\min(q,p)}]\}^{1/p} \\ &= O[|a_{k+1}|^{\min(1,q/p)}].\end{aligned}$$

Similarly, we have

$$\overline{\psi}_p(k) = O[|a_{k+1}|^{\min(1,q/p)}] \quad \text{and} \quad \overline{\phi}_p(k) = O[|a_{k+1}|^{\min(1,q/p)}].$$

Therefore, for the special case $p = q = 2$, we have $\|X_k - X_k^*\|_2 = |a_k|\|\varepsilon_0 - \varepsilon_0'\|_2$, and (4.6) is reduced to $|a_k| = O(k^{-(3-\alpha)/2} l^{1/2}(k))$, which implies $\overline{\Theta}_2(n) + \bar{\Psi}_2(n) + \overline{\Phi}_2(n) = O(n^\alpha l(n))$, and (4.7) is replaced by $\sum_{k=0}^\infty |a_k| < \infty$, which is a classical condition for linear processes to be short-range dependent and implies $\overline{\Theta}_2(n) + \overline{\Psi}_2(n) + \bar{\Phi}_2(n) = O(n)$.

Now we consider the model (4.3) for chains with infinite memory **[2, 16]**. We assume that G satisfies

$$|G(x_{-1}, x_{-2}, \ldots) - G(x'_{-1}, x'_{-2}, \ldots)| \le \sum_{j=1}^\infty \omega_j |x_{-j} - x'_{-j}|, \quad \text{where } \omega_j \ge 0. \tag{4.11}$$

For simplicity, we assume $\varepsilon_k \in \mathcal{L}^2$. Let $\rho_2(k) = \|X_k - X_k^*\|$. From (4.3) and (4.11), we have

$$\rho_2(k+1) \le \sum_{i=1}^{k+1} \omega_i \rho_2(k+1-i), \quad k \ge 0.$$

Define a sequence $(a_k)_{k\ge 0}$ by $a_0 = \rho_2(0)$ and

$$a_{k+1} = \sum_{i=1}^{k+1} \omega_i a_{k+1-i}.$$

Then $S_2(n; \rho_2(.)) \le S_2(n; a)$. Let $h(s) = \sum_{k=0}^\infty a_k s^k$ and $u(s) = \sum_{i=1}^\infty \omega_i s^i$, for $|s| \le 1$. By simple calculation, we have $h(s) = a_0(1 - u(s))^{-1}$. Suppose that, as $s \uparrow 1$, we have $1 - u(s) \sim (1 - s)^d$ with $d \in (0, 1/2)$ which implies $u(1) = 1$. As in (2.6), we obtain

$$2\pi S_2(n; a) = \int_0^{2\pi} \left|h(e^{\sqrt{-1}u})\right|^2 \frac{\sin^2(nu/2)}{\sin^2(u/2)}\, du = O(n^{1+2d}).$$

Suppose that the density function of ε_i satisfies (4.9). Let $Y_k = X_{k+1} - \varepsilon_{k+1}$. Following the calculation in (4.10), we get $\overline{\theta}_2(k) = O(\|Y_k - Y_k^*\|) = O(\rho_2(k+1))$. Thus, $\overline{\Theta}_2(n) = O(n^{1+2d})$. If $u(1) < 1$, then $h(1) < \infty$ and $\overline{\Theta}_2(n) = O(n)$. The other quantities $\overline{\Psi}_2(n)$ and

$\overline{\Phi}_2(n)$ can be handled similarly, therefore, Theorems 3.1 and 3.7 are applicable.

For the nonlinear processes $X_k = m(X_{k-1}) + \varepsilon_k$, an important example is the threshold autoregressive model **[13]**

$$X_k = a \max(X_{k-1}, 0) + b \min(X_{k-1}, 0) + \varepsilon_k,$$

where a and b are real parameters. If $\varepsilon_k \in \mathcal{L}^\beta$ and $\lambda = \sup_x |m'(x)| < 1$, then (4.4) has a stationary distribution and

$$\|X_k - X_k^*\|_\beta \le \lambda \|X_{k-1} - X_{k-1}^*\|_\beta \le \cdots \le \lambda^k \|\varepsilon_0 - \varepsilon_0'\|_\beta = O(\lambda^k),$$

(see also **[17]**); thus, (4.7) is satisfied.

Acknowledgments. The authors are grateful to two anonymous referees and the Editor for their constructive comments.

REFERENCES

1. P.J. Bickel and M. Rosenblatt, *On some global measures of the deviations of density function estimates*, Ann. Stat. **1** (1973), 1071–1095.

2. P. Doukhan and O. Wintenberger, *Weakly dependent chains with infinite memory*, Stoch. Proc. Appl. **118** (2008), 1997–2013.

3. D.A. Freedman, *On tail probabilities for martingales*, Ann. Prob. **3** (1975), 100–118.

4. G. Kallianpur, *Some ramifications of Wieners ideas on nonlinear prediction*, in *Norbert Wiener, collected works*, P. Masani, ed., MIT Press, Cambridge, 1981.

5. D.M. Mason, G.R. Shorack and J.A. Wellner, *Strong limit theorems for oscillation moduli of the uniform empirical process*, Prob. Theor. Rel. Fields **65** (1983), 83–97.

6. M.H. Neumann, *Strong approximation of density estimators from weakly dependent observations by density estimators from independent observations*, Ann. Stat. **26** (1998), 2014–2048.

7. M.B. Priestley, *Non-linear and non-stationary time series analysis*, Academic Press, New York, 1988.

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

9. ———, *A comment on a conjecture of N. Wiener*, Stat. Prob. Lett. **79** (2009), 347–348.

10. G.R. Shorack and J.A. Wellner, *Empirical processes with applications to statistics*, John Wiley, New York, 1986.

11. W. Stute, *A law of the logarithm for kernel density estimators*, Ann. Prob. **10** (1982), 414–422.

12. W. Stute, *The oscillation behavior of empirical processes*, Ann. Prob. **10** (1982), 86–107.

13. H. Tong, *Non-linear time series: A dynamical system approach*, Oxford University Press, Oxford, 1990.

14. W.B. Wu, *Nonlinear system theory*: *Another look at dependence*, Proc. Nat. Acad. Sci. **102** (2005), 14150–14154.

15. ______, *Oscillations of empirical distribution functions under dependence*, IMS Lect. Notes Mono. Ser.: High dimensional Probability **51** (2006), 53–61.

16. W.B. Wu, Y.X. Huang and Y.B. Huang, *Kernel estimation for time series*: *An asymptotic theory*, Stoch. Proc. Appl. **120** (2010), 2412–2431.

17. W.B. Wu and X. Shao, *Limit theorems for iterated random functions*, J. Appl. Prob. **41** (2004), 425–436.

Department of Mathematics, Nanchang University, 999 Xuefu Road, Nanchang, 330031, P.R. China
Email address: **xxy3290@126.com**

Department of Mathematics, Nanchang University, 999 Xuefu Road, Nanchang, 330031, P.R. China
Email address: **hongwei-yin@hotmail.com**