Bernoulli **9**(5), 2003, 809–831

Empirical processes of long-memory sequences

WEI BIAO WU

Department of Statistics, University of Chicago, 5734 S. University Avenue, Chicago, IL 60637, USA. E-mail: wbwu@galton.uchicago.edu

Asymptotic expansions of long-memory sequences indexed by piecewise differentiable functionals are investigated, and upper bounds of outer expectations of these functionals are given. These results differ strikingly from the classical theories of empirical processes of independent random variables. Our results go beyond earlier ones by allowing wider classes of function as well as by presenting sharper bounds, and thus provide a more versatile approach for related statistical inferences. A complete characterization of empirical processes for the class of indicator functions is presented, and an application to M-estimation is discussed.

Keywords: Long- and short-range dependence; linear process; martingale central limit theorem.

1. Introduction

Motivated by many practical examples, long-memory processes have been extensively investigated by the statistical community over the past several decades; see Beran (1994). A distinctive feature of such processes is that their correlations decay fairly slowly as the time-lag increases. An important model is the linear process $X_n = \sum_{i=0}^{\infty} a_i \varepsilon_{n-i}$, where the $\{\varepsilon_i, i \in \mathbb{Z}\}$ are independent and identically distributed (i.i.d.) random variables with zero mean and finite variance, and the coefficients a_i satisfy $\sum_{i=0}^{\infty} a_i^2 < \infty$. Many important time-series models, such as the autoregressive moving average and fractional autoregressive integrated moving average, take this form. If a_n decays to 0 at a sufficiently slow rate, then the covariances of X_n are not summable and thus the process exhibits long-range dependence. It is clearly necessary to consider the theoretical properties of $S_n(K) = \sum_{i=1}^{n} K(X_i)$ for statistical inferences of such processes. In the paper we will investigate the uniform asymptotic behaviour of $S_n(K)$, where K belongs to certain classes $\mathcal{K}$, and will present a complete characterization of $S_n(K)$ when $\mathcal{K} = \mathcal{I} = \{\mathbf{1}_{x \leqslant s}, s \in \mathbb{R}\}$, the class of indicator functions.

The theory of empirical processes for independent random variables is well developed; see the extensive treatment by van der Vaart and Wellner (1996). Among the results there are Vapnik–Chervonenkis and bracketing theories. Under certain conditions on bracketing numbers on the class $\mathcal{K}$, the abstract Donsker theorem asserts uniform central limit theorems and the limiting distribution is the so-called abstract Brownian bridge. Results of this sort have many applications in statistics. A large number of examples are given by van der Vaart and Wellner (1996).

1350–7265

However, the problem of uniform convergence becomes much harder when dependence is present. The dependence structure itself is often of interest in time series analysis. For example, the estimation of Hurst's index is of critical importance in the study of long-memory processes. Much previous work has been concerned with very weakly dependent processes; see, for example, Doukhan *et al*. (1995) and Rio (1998) on β-mixing sequences. Andrews and Pollard (1994) and Arcones and Yu (1994) provided surveys of empirical processes for mixing processes. Under suitable mixing rates, results of this sort usually assert that empirical processes behave as if the observations were i.i.d. For other dependent processes, Bae and Levental (1995) considered a uniform central limit theorem for Markov chains; Dehling and Taqqu (1989) and Arcones and Yu (1994) discussed functionals of long-range dependent Gaussian processes and Gaussian random fields. Ho and Hsing (1996) raised the problem of uniform convergence for linear processes X_t which may not necessarily be Gaussian. For the indicator function class $\mathcal{I}$, Ho and Hsing (1996) successfully derived uniform asymptotic expansions. See their paper and the recent review by Koul and Surgailis (2002) for further references and some important historical developments.

For long-memory linear processes we are able to establish uniform limiting distributions of $S_n(K)$ when the class $\mathcal{K}$ consists of piecewise differentiable functions. In particular, $\mathcal{K}$ contains the Huber-type functions $H_s(x) = \min[\max(x - s, -1), 1]$, $s \in \mathbb{R}$, which frequently appear in robust inference. Our treatment is similar to Arcones' (1996) work in which weak convergence properties of stochastic processes indexed by smooth functions were discussed. The empirical processes behave significantly differently from those of independent random variables in that the limiting distributions are often degenerate. While we impose weaker conditions, sharper upper bounds are obtained for the special function class consisting of indicators. Our results could be possibly extended and applied to other problems related to linear processes.

The paper is organized as follows. Our main results are presented in Section 2 and proved in Section 4. Section 3 contains an application to M-estimation theory.

2. Main results

Denote the measure $w_\lambda(\mathrm{d}t) = (1 + |t|)^\lambda \mathrm{d}t$. For $\gamma \geqslant 0$, define the class

$$\mathcal{K}(\gamma) = \left\{ K(x) = \int_0^x g(t)\mathrm{d}t \colon \int_{\mathbb{R}} |g(t)|^2 w_{-\gamma}(\mathrm{d}t) \leqslant 1 \right\}.$$

For $K \in \mathcal{K}$, we have

$$K^2(s) \leqslant \int_0^s |g(t)|^2 w_{-\gamma}(\mathrm{d}t) \int_0^s w_\gamma(\mathrm{d}t) \leqslant |s|(1 + |s|)^\gamma \leqslant 2^\gamma (1 + |s|^{\gamma+1}) \tag{1}$$

by Cauchy's inequality, which gives a growth rate for K. Let

$$\mathcal{K}(\gamma;\, I) = \left\{ K(x) = \sum_{i=1}^{I+1} \mathbf{1}_{[\lambda_{i-1}\lambda_i)}(x) K_i(x)\colon\ K_i \in \mathcal{K}(\gamma),\ \ |K_i(s)| \leqslant (1+|s|)^{\gamma/2}, \right.$$

$$\left. -\infty = \lambda_0 < \lambda_1 < \ldots < \lambda_I < \lambda_{I+1} = \infty \right\}. \tag{2}$$

So $\mathcal{K}(\gamma;\, I)$ contains piecewise differentiable functions. Denote by $\mathcal{C}^p = \mathcal{C}^p(\mathbb{R})$ the class of functions having derivatives up to pth order. For a measurable function K, let $K_\infty(x) = \mathrm{E}[K(X_1 + x)]$ if it exists. If $K_\infty \in \mathcal{C}^p$, then, as in Ho and Hsing (1997), let

$$S_n(K;\, p) = \sum_{i=1}^{n} \left[K(X_i) - \sum_{j=0}^{p} K_\infty^{(j)}(0) U_{i,j} \right], \qquad U_{n,r} = \sum_{0 \leqslant j_1 < \ldots < j_r} \prod_{s=1}^{r} a_{j_s} \varepsilon_{n-j_s}, \qquad U_{n,0} = 1. \tag{3}$$

We are interested in the uniform upper bound $\sup_{K \in \mathcal{K}(\gamma;I)} |S_n(K;\, p)|$, which may not be a *bona fide* random variable since the class $\mathcal{K}$ is not countable. So the notion of *outer expectation* $\mathrm{E}^*\xi = \inf\{\mathrm{E}\tau\colon \tau \text{ is a random variable and } \tau \geqslant \xi,\, \mathrm{E}\,\tau \text{ exists}\}$ (van der Vaart, 1998) is used.

Let F_k and $F = F_\infty$ be the distribution functions of $\sum_{i=0}^{k-1} a_i \varepsilon_{-i}$ and $X_0 = \sum_{i=0}^{\infty} a_i \varepsilon_{-i}$, respectively; let $F_k^{(r)}$ and $F^{(r)}$ be the corresponding rth derivatives if they exist; let $\widetilde{\mathbf{X}}_n = (\ldots,\, \varepsilon_{n-1},\, \varepsilon_n)$ be the one-sided shift process. Write $f = F'$ and $f_k = F'_k$ for the first-order derivatives. Define

$$S_n(y;\, p) = \sum_{i=1}^{n} L(\widetilde{\mathbf{X}}_i,\, y), \qquad L(\widetilde{\mathbf{X}}_n,\, y) = \mathbf{1}(X_n \leqslant y) - \sum_{i=0}^{p} (-1)^i F^{(i)}(y) U_{n,i}. \tag{4}$$

Let $A_n(k) = \sum_{i=n}^{\infty} |a_i|^k$, $\theta_{n,p} = |a_{n-1}|[|a_{n-1}| + A_n^{1/2}(4) + A_n^{p/2}(2)]$, $\Theta_{n,p} = \sum_{k=1}^{n} \theta_{k,p}$ and

$$\Xi_{n,p} = n\Theta_{n,p}^2 + \sum_{i=1}^{\infty} (\Theta_{n+i,p} - \Theta_{i,p})^2.$$

Clearly, for $k \geqslant 2$, $A_n(k) \downarrow 0$ as $n \to \infty$. In Theorem 1 we do not require a_n to take special forms such as $n^{-\beta}\ell(n)$, where throughout the paper ℓ stands for slowly varying functions. Without loss of generality, we assume henceforth that $a_0 = 1$ and that there are infinitely many i such that $a_i \neq 0$. The latter requirement is imposed to avoid the degenerate case in which X_n is reduced to m-dependent processes (Hoeffding and Robbins, 1948).

Theorem 1. *Assume that* $\mathrm{E}(|\varepsilon_1|^{4+\gamma}) < \infty$ *for some* $\gamma \geqslant 0$, *and that* $f_\kappa \in \mathcal{C}^p$ *for some integers* $\kappa > 0$ *and* $p \geqslant 0$. *Furthermore, assume that*

$$\sum_{r=0}^{p} \int_{\mathbb{R}} |f_\kappa^{(r)}(x)|^2 w_\gamma(\mathrm{d}x) < \infty. \tag{5}$$

Then

$$\mathrm{E}^*\left[\sup_{K\in\mathcal{K}(\gamma)}|S_n(K;\ p)|^2\right]=O(\Xi_{n,p}). \tag{6}$$

Relation (6) expresses the *uniform reduction principle* (Taqqu 1975; Dehling and Taqqu 1989; Ho and Hsing 1997) which says that $S_n(K)$ can be approximated by linear combinations of $\sum_{i=1}^n U_{i,j}$, $j=1,\ldots,p$. Theorem 2 provides a uniform upper bound for the special class $\mathcal{I}$ consisting of indicator functions $\mathbf{1}_s(\cdot)=\mathbf{1}(\cdot\leqslant s)$.

Theorem 2. *Assume that* $\mathrm{E}(|\varepsilon_1|^{4+\gamma})<\infty$ *for some* $\gamma\geqslant 0$, *that* $f_\kappa\in\mathcal{C}^{p+1}$ *for some integers* $\kappa>0$ *and* $p\geqslant 0$, *and that*

$$\sum_{r=0}^{p+1}\int_{\mathbb{R}}|f_\kappa^{(r)}(x)|^2 w_\gamma(\mathrm{d}x)<\infty. \tag{7}$$

Then

$$\mathrm{E}\left[\sup_{t\in\mathbb{R}}(1+|t|)^\gamma|S_n(t;\ p)|^2\right]=O\big(n\log^2 n+\Xi_{n,p}\big). \tag{8}$$

Corollary 1. *Let the assumptions of Theorem 2 be satisfied and, in addition, let* $a_n=n^{-\beta}\ell(n)$, $n\geqslant 1$, *where* $\beta\in(\frac{1}{2},1)$ *and* ℓ *is a slowly varying function. Then*

$$\mathrm{E}^*\left[\sup_{K\in\mathcal{K}(\gamma;I)}|S_n(K;\ p)|^2\right]=O[n\log^2 n+\Xi_{n,p}], \tag{9}$$

where $\Xi_{n,p}=O(n)$, $O[n^{2-(p+1)(2\beta-1)}\ell^{2(p+1)}(n)]$ *or* $O(n)\big[\sum_{i=1}^n|\ell^{p+1}(i)|/i\big]^2$ *if* $(p+1)(2\beta-1)>1$, $(p+1)(2\beta-1)<1$ *or* $(p+1)(2\beta-1)=1$, *respectively.*

Let $\{\mathbb{B}(u),\,u\in\mathbb{R}\}$ be a standard two-sided Brownian motion, $\mathcal{S}=\{(u_1,\ldots,u_r)\in\mathbb{R}^r\colon -\infty<u_1<\ldots<u_r<1\}$ and define the multiple Wiener–Itô integral (Major 1981)

$$Z_{r,\beta}=\xi(r,\beta)\int_{\mathcal{S}}\left\{\int_0^1\prod_{i=1}^r[\max(v-u_i,0)]^{-\beta}\mathrm{d}v\right\}\mathrm{d}\mathbb{B}(u_1)\ldots\mathrm{d}\mathbb{B}(u_r), \tag{10}$$

where the norming constant ξ ensures that $\mathrm{E}(Z_{r,\beta}^2)=1$. Let

$$\sigma_{n,r}^2=n^{2-r(2\beta-1)}\ell^{2r}(n). \tag{11}$$

Theorem 3. *Assume that* $a_n=n^{-\beta}\ell(n)$ *for* $n\geqslant 1$, $\mathrm{E}(|\varepsilon_1|^4)<\infty$, $f_\kappa\in\mathcal{C}^{p+2}$ *for some integers* $\kappa>0$ *and* $p\geqslant 0$, *and*

$$\sum_{r=0}^{p+2}\int_{\mathbb{R}}|f_\kappa^{(r)}(x)|^2\mathrm{d}x<\infty. \tag{12}$$

(i) *If* $(p+1)(2\beta-1)>1$ *or* $(p+1)(2\beta-1)=1$ *and* $\sum_{n=1}^\infty|\ell^{p+1}(n)|/n<\infty$, *then the weak convergence*

$$\frac{1}{\sqrt{n}} S_n(s;\ p) \Rightarrow W(s) \tag{13}$$

holds in the Skorokhod space $\mathcal{D}(\mathbb{R})$, where $W(s)$ is a Gaussian process.

(ii) *If $(p+1)(2\beta-1) < 1$, then*

$$\frac{1}{\sigma_{n,p+1}} S_n(s;\ p) \Rightarrow (-1)^{p+1}\{f^{(p)}(s),\ s \in \mathbb{R}\} Z_{p+1,\beta}. \tag{14}$$

If we view $S_n(s;\ p)$ as the remainder of the 'Taylor' expansion of $S_n(\mathbf{1}_s)$, then (13) and (14) describe the interesting phenomena that the remainder has a degenerate distribution for low-order expansions and a non-degenerate Gaussian limit for high-order ones.

Corollary 2. *Let the conditions of Theorem 3 be satisfied with $p = 0$ and $\beta = 1$.*

(i) *If $\sum_{n=1}^{\infty} |\ell(n)|/n < \infty$, then we have* (13).

(ii) *If $\sum_{n=1}^{\infty} |\ell(n)|/n = \infty$, then*

$$\frac{1}{\tilde{\sigma}_n} S_n(s;\ 0) \Rightarrow \{f(s),\ s \in \mathbb{R}\} Z, \tag{15}$$

where Z is standard normal and $\tilde{\sigma}_n = \|\sum_{i=1}^{n} X_i\| \sim c\sqrt{n}|\sum_{i=1}^{n} \ell(i)/i|$ for some $c > 0$.

Interestingly, Corollary 2 gives a complete characterization of the limiting behaviour of $S_n(s;\ 0) = \sum_{i=1}^{n} \mathbf{1}(X_i \leqslant s) - nF(s)$ on the boundary $\beta = 1$. It is well known that the process X_t is long- (or short-)range dependent if $\beta < 1$ (or $\beta > 1$). On the boundary $\beta = 1$ it depends on the finiteness of $\sum_{i=1}^{\infty} |\ell(i)|/i$. This result in some sense suggests the power of our approach. It is unclear whether similar characterizations exist on other boundaries $\beta = (2+p)/(2+2p)$, where $p \geqslant 1$ is an integer.

Remark 1. Theorem 2 and Corollary 1 improve and generalize the earlier important results of Ho and Hsing (1996) in several aspects. Consider the special case in which $\gamma = 0$. The latter paper requires that F_1, the distribution function of ε_1, is $p+3$ times differentiable with bounded, continuous and integrable derivatives. Our assumption (5) is clearly weaker. Next, Corollary 1 allows a wider class $\mathcal{K}(0,\ 1) \supset \mathcal{I}$. Furthermore, if a_n adopts the special form $n^{-\beta}\ell(n)$, $n \geqslant 1$, then for $b > 0$, (9) gives a sharper upper bound via Markov's inequality:

$$P\left[\sup_{t \in \mathbb{R}} |S_n(\mathbf{1}_t;\ p)| > b\right] = b^{-2} O[n \log^2 n + \Xi_{n,p}];$$

see Lemma 5 for upper bounds of $\Xi_{n,p}$ and Theorem 2.1 in Ho and Hsing (1996) for a comparison. Consequently, applications derived in the latter paper which are based on inequalities of this type can be correspondingly improved. We do not pursue this matter here.

Remark 2. The quantity $\int_{\mathbb{R}} |f_\kappa^{(r)}(x)|^2 \mathrm{d}x$ in condition (5) with $\gamma = 0$ is interestingly related to many aspects in statistics, such as Wilcoxon's rank test, optimal bandwidth selection and projection pursuit. The estimation problem has been widely studied; see Wu (1995) for further references.

Remark 3. For Gaussian random fields, Arcones and Yu (1994) obtained weak convergence of empirical processes under the bracket condition $\int_0^\infty \{N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|)\}^{1/2} \mathrm{d}\epsilon < \infty$, where $\mathcal{F}$ is the index set and the bracketing number $N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|)$ is the minimum number of ϵ-brackets needed to cover $\mathcal{F}$ under the $\mathcal{L}^2$ norm. This bracket condition excludes the class of indicator functions $\mathcal{I}$ since $N_{[]}(\epsilon, \mathcal{I}, \|\cdot\|)$ has order $1/\epsilon^2$ as $\epsilon \downarrow 0$.

Remark 4. Recently Giraitis and Surgailis (2002) considered the uniform upper bound $\sup_{s\in\mathbb{R}}|S_n(s;\, p)|$ for two-sided linear processes with $p = 1$. A reduction principle is derived. It seems that our approach cannot be directly applied to two-sided processes.

We say that K has *power rank* p if $K_\infty^{(i)}(0)$ exist and vanish for $1 \leqslant i < p$ and $K_\infty^{(p)}(0) \neq 0$ (Ho and Hsing 1997). Power rank is reduced to Hermite's rank if X_1 is standard normal. Define the class $\mathcal{K}_p = \{K \in \mathcal{K}(\gamma;\, I) : \ K_\infty \in \mathcal{C}^p,\, K_\infty^{(i)}(0) = 0,\ 1 \leqslant i < p\}$, which contains functions with power rank at least p. Corollary 1, together with $\sum_{k=1}^n Y_{n,p}/\sigma_{n,p} \Rightarrow Z_{p,\beta}$ (Surgailis, 1982), immediately yields the following result:

Corollary 3. *Let* $1 \leqslant p < 1/(2\beta - 1)$ *and the conditions of Corollary 1 be satisfied. Then*

$$\frac{1}{\sigma_{n,p}}\{S_n(K) - nK_\infty(0),\ K \in \mathcal{K}_p\} \Rightarrow \{K_\infty^{(p)}(0),\ K \in \mathcal{K}_p\} Z_{p,\beta}. \tag{16}$$

The limiting distribution in (16) is degenerate in the sense that it forms a line of multiples of $Z_{p,\beta}$. In contrast, the empirical processes for i.i.d. samples take abstract Brownian bridges as limits. We conjecture that if $p(2\beta - 1) > 1$, then the limiting distributions are non-degenerate Gaussian processes.

3. *M*-estimators

For i.i.d. observations, van der Vaart and Wellner (1996) presented a detailed account of various statistical applications based on convergence properties of empirical processes. Regarding long-memory processes, our theory can likewise provide a basis for inference, particularly in the study of certain functionals of such processes with unknown parameters for which estimates are plugged in. To fix this idea, let $\mathcal{M} \subset \mathbb{R}^d$, $d \geqslant 1$, be the parameter space and $\mathbf{m}_0 \in \mathcal{M}$ be the unknown parameter to be estimated; let $\mathbf{H}(x, \mathbf{m}) = (H^1(x, \mathbf{m}), \ldots, H^d(x, \mathbf{m}))$, where H^j, $1 \leqslant j \leqslant d$, are measurable functions defined on the space $\mathbb{R} \times \mathbb{R}^d$. Then the functional $\mathcal{A}_n(\mathbf{m}_0) = \sum_{j=1}^n \mathbf{H}(X_j, \mathbf{m}_0)$ which contains the unknown parameter $\mathbf{m}_0$ is often studied via $\mathcal{A}_n(\mathbf{m}_n)$, where $\mathbf{m}_n$ is an estimator of $\mathbf{m}_0$.

An estimator $\mathbf{m}_n = \mathbf{m}_n(X_1, \ldots, X_n)$ of $\mathbf{m}_0$ is generically called an M-estimator if it satisfies $A_n(\mathbf{m}_n) \approx 0$. In this section we shall establish asymptotic distributions of

M-estimators. Let $\mathbf{H}_\infty(x, \mathbf{m}) = \mathrm{E}\mathbf{H}(X_1 + x, \mathbf{m})$ and $\mathcal{M}(\delta) = \{\mathbf{m}\colon |\mathbf{m} - \mathbf{m}_0| \leqslant \delta\} \cap \mathcal{M}$, where $|\cdot|$ denotes Euclidean distance.

Assumption 1. *There exist $\delta_0 > 0$ and an integer $p \geqslant 1$ such that for all $\mathbf{m} \in \mathcal{M}(\delta_0)$, $\mathbf{H}_\infty(\cdot, \mathbf{m})$ is p times differentiable at $x = 0$. Let $c_i(\mathbf{m}) = \partial^i \mathbf{H}_\infty(x, \mathbf{m})/\partial x^i|_{x=0}$ and assume that $c_p(\cdot)$ is continuous at $\mathbf{m}_0$, $c_p(\mathbf{m}_0) \neq 0$ and $c_i(\mathbf{m}) = 0$ for all $1 \leqslant i < p$ and all $\mathbf{m} \in \mathcal{M}(\delta_0)$.*

Assumption 2. *For all $1 \leqslant j \leqslant d$, $H^j_\infty(0, \cdot)$ is Fréchet differentiable at $\mathbf{m} = \mathbf{m}_0$. That is, there exists a matrix $\Sigma(\mathbf{m}_0) = \left(\partial H^j_\infty(0, \mathbf{m})/\partial m^i\right)^{d,d}_{i,j=1}|_{\mathbf{m}=\mathbf{m}_0}$ such that $|\mathbf{H}(0, \mathbf{m}) - \mathbf{H}(0, \mathbf{m}_0) - (\mathbf{m} - \mathbf{m}_0)\Sigma(\mathbf{m}_0)| = o(|\mathbf{m} - \mathbf{m}_0|)$. Suppose that the matrix $\Sigma(\mathbf{m}_0)$ is non-singular.*

Assumption 3. *The estimator $\mathbf{m}_n \to \mathbf{m}_0$ in probability and $A_n(\mathbf{m}_n) = o_P(\sigma_{n,p})$.*

Remark 5. In Assumption 1, since $c_p(\cdot)$ is continuous at $\mathbf{m}_0$, there exists ϵ_0 such that $c_p(\mathbf{m}) \neq 0$ for all $|\mathbf{m} - \mathbf{m}_0| \leqslant \epsilon_0$. Hence we can substitute δ_0 by $\min(\epsilon_0, \delta_0)$. Assumptions 2 and 3 are standard in M-estimation theory (see Section 3.3 in van der Vaart and Wellner, 1996).

Theorem 4. *Let Assumptions 1, 2 and 3 be satisfied. Suppose that there exist $C > 0$ and $\gamma \geqslant 0$ such that $H^q(\cdot, \mathbf{m})/C \in \mathcal{K}(\gamma; I)$, for all $1 \leqslant q \leqslant d$ and all $\mathbf{m} \in \mathcal{M}(\delta_0)$. If $p(2\beta - 1) < 1$, then*

$$\frac{n}{\sigma_{n,p}}(\mathbf{m}_n - \mathbf{m}_0) \Rightarrow c_p(\mathbf{m}_0)\Sigma^{-1}(\mathbf{m}_0) Z_{\beta,p}. \tag{17}$$

Koul and Surgailis (1997) considered the one-dimensional location estimation with $H(x, m) = \psi(x - m)$ in which one observes $Z_t = X_t + m$. Beran (1991) discussed M-estimation of location parameters for long-memory Gaussian processes. Arcones and Yu (1994) treated $H(X_\cdot, m) = h[G(X_\cdot), m)]$ where $X_\cdot$ is a Gaussian random field. Theorem 4 can be applied to the location estimation problem in the non-Gaussian and nonlinear model $Z_t = g(X_t) + m$ by letting $H(x, m) = \psi(g(x) - m)$, where ψ is a non-decreasing function.

4. Proofs

Let $\underline{X}_{n,i} = \sum_{j=-\infty}^{i} a_{n-j}\varepsilon_j$ and $\overline{X}_{n,i} = \sum_{j=i}^{n} a_{n-j}\varepsilon_j$ be truncated processes; let $\{\varepsilon'_n, n \in \mathbb{Z}\}$ be an i.i.d. copy of $\{\varepsilon_n, n \in \mathbb{Z}\}$ and $X'_n = \sum_{i=0}^{\infty} a_i \varepsilon'_{n-i}$. Define $\underline{X}'_{n,i}$ and $\overline{X}'_{n,i}$ similarly. For a random variable ξ, denote its $\mathcal{L}^\rho$ norm ($\rho \geqslant 1$) by $\|\xi\|_\rho = [\mathrm{E}(|\xi|^\rho)]^{1/\rho}$, and $\mathcal{L}^2$ norm $\|\xi\| = \|\xi\|_2$. Define the projection operators $\mathcal{P}_j\xi = \mathrm{E}[\xi|\widetilde{\mathbf{X}}_j] - \mathrm{E}[\xi|\widetilde{\mathbf{X}}_{j-1}]$.

Lemma 1. *Suppose* $\mathrm{E}(\varepsilon_1) = 0$, *and* $\mathrm{E}[|\varepsilon_1|^\tau] < \infty$ *for some* $\tau \geqslant 2$. *Then there exists a $B_\tau > 0$ such that $\mathrm{E}[|\sum_{i=1}^n b_i\varepsilon_i|^\varrho] \leqslant B_\tau(\sum_{i=1}^n b_i^2)^{\varrho/2}$ holds for all real numbers $b_1, \ldots, b_n$ and all ϱ for which $0 < \varrho \leqslant \tau$.*

This lemma is an easy consequence of the Rosenthal inequalities (see Theorem 1.5.11 in de la Peña and Giné, 1999).

Lemma 2. *Let* $H(t,\delta,\eta) = g(t+\delta+\eta) - \sum_{i=0}^{q} g^{(i)}(t+\eta)\delta^i/i!$, *where* $g \in \mathcal{C}^{q+1}$, $q \geqslant -1$. *Then*

$$\int_{\mathbb{R}} |H(t,\delta,\eta)|^2 w_\gamma(\mathrm{d}t) \leqslant \frac{|\delta|^{2q+2}(1+|\delta|)^\gamma (1+|\eta|)^\gamma}{[(q+1)!]^2} \int_{\mathbb{R}} |g^{(q+1)}(t)|^2 w_\gamma(\mathrm{d}t). \tag{18}$$

Proof. Let $t' = t+\eta$. Then it suffices to show (18) with $\eta = 0$ since $1+|t| \leqslant (1+|t'|)(1+|\eta|)$. We make extensive use of this simple inequality. Using the convention $\sum_{i=0}^{-1} = 0$, (18) trivially holds when $q=-1$. Assume without loss of generality that $\delta > 0$ and $q = 1$, since general cases follow similarly. Note that $g(t+\delta) - g(t) - \delta g'(t) = \int_0^\delta \int_0^u g''(t+v)\mathrm{d}v\,\mathrm{d}u$. By Cauchy's inequality, the left-hand side of (18) is no greater than

$$\int_{\mathbb{R}} \left[\int_0^\delta \int_0^u \mathrm{d}v\,\mathrm{d}u\right] \times \left[\int_0^\delta \int_0^u |g''(t+v)|^2 \mathrm{d}v\,\mathrm{d}u\right] w_\gamma(\mathrm{d}t)$$

$$\leqslant \frac{\delta^2}{2} \int_0^\delta \int_0^u \int_{\mathbb{R}} |g''(t)|^2 (1+|t-v|)^\gamma \mathrm{d}t\,\mathrm{d}v\,\mathrm{d}u, \tag{19}$$

which yields (18) again by the elementary inequality $1+|t-v| \leqslant (1+|t|)(1+|v|)$. □

Lemma 3. *Let* $\{\xi_n\}_{n\in\mathbb{Z}}$ *be a stationary and ergodic Markov chain and* h *be a measurable function on the state space of the chain such that* $h(\xi_i)$ *has mean zero and finite variance. Define* $S_n(h) = \sum_{i=1}^n h(\xi_i)$ *and* $\alpha_n = \|\mathrm{E}[h(\xi_n)|\xi_1] - \mathrm{E}[h(\xi_n)|\xi_0]\|$ *for* $n \geqslant 1$. *Then*

$$\sum_{n=1}^\infty \alpha_n < \infty \tag{20}$$

entails $S_n(h)/\sqrt{n} \Rightarrow N(0,\sigma_h^2)$ *for some* $\sigma_h^2 < \infty$.

Proof. The central limit theorem here is essentially an easy consequence of Woodroofe (1992) which asserts that $\{S_n(h) - \mathrm{E}[S_n(h)|\xi_0]\}/\sqrt{n} \Rightarrow N(0,\sigma_h^2)$ if condition (20) is satisfied. For $j \geqslant 0$, $\|\mathrm{E}[S_n(h)|\xi_{-j}] - \mathrm{E}[S_n(h)|\xi_{-j-1}]\| \leqslant \sum_{i=1}^n \alpha_{i+j+1}$. Thus by (20) and since $\mathrm{E}[S_n(h)|\xi_{-j}] - \mathrm{E}[S_n(h)|\xi_{-j-1}]$, $j \geqslant 0$, are orthogonal,

$$\|\mathrm{E}[S_n(h)|\xi_0]\|^2 = \sum_{j=0}^\infty \|\mathrm{E}[S_n(h)|\xi_{-j}] - \mathrm{E}[S_n(h)|\xi_{-j-1}]\|^2 = O\left(\sum_{j=0}^\infty \sum_{i=1}^n \alpha_{i+j+1}\right) = o(n)$$

yields the lemma. □

Lemma 4. *Let* $H \in \mathcal{C}^1$ *and* $\delta > 0$. *Then*

$$\sup_{t\leqslant s\leqslant t+\delta} H^2(s) \leqslant 2\delta^{-1}\int_t^{t+\delta} H^2(u)\mathrm{d}u + 2\delta\int_t^{t+\delta} H'^2(u)\mathrm{d}u$$

and, for $\gamma \geqslant 0$,

$$\sup_{s\in\mathbb{R}}[(1+|s|)^\gamma H^2(s)] \leqslant 2^{1+2\gamma}\int_{\mathbb{R}} H^2(t)w_\gamma(\mathrm{d}t) + 2^{1+2\gamma}\int_{\mathbb{R}}[H'(t)]^2 w_\gamma(\mathrm{d}t).$$

Proof. For $x, y \in [t, t+\delta]$, $|H(x) - H(y)| \leqslant \int_t^{t+\delta}|H'(u)|\mathrm{d}u$. The inequality $[H(x) - 2H(y)]^2 \geqslant 0$ implies $0 \leqslant 2|H(x) - H(y)|^2 + 2H^2(y) - H^2(x)$. Integrating the latter inequality over $[t, t+\delta]$ gives $\int_t^{t+\delta}[2|H(x) - H(y)|^2 + 2H^2(y)]\mathrm{d}y \geqslant \delta H^2(x)$, which results in the first inequality in the lemma by Cauchy's inequality. For the second, let $\delta = 1$. Observe that if $k \leqslant s \leqslant k+1$, then $1+|s| \leqslant 2(1+|k|) \leqslant 4(1+|s|)$. So

$$\begin{aligned}
\sup_{s\in\mathbb{R}}[(1+|s|)^\gamma H^2(s)] &\leqslant \sum_{k\in\mathbb{Z}}\sup_{k\leqslant s\leqslant k+1}[(1+|s|)^\gamma H^2(s)] \\
&\leqslant \sum_{k\in\mathbb{Z}} 2^{1+\gamma}(1+|k|)^\gamma \int_k^{k+1}[H^2(u) + H'^2(u)]\mathrm{d}u \\
&\leqslant \sum_{k\in\mathbb{Z}} 2^{1+2\gamma}\int_k^{k+1}[H^2(u) + H'^2(u)]w_\gamma(\mathrm{d}u) \\
&= 2^{1+2\gamma}\int_{\mathbb{R}}[H^2(u) + H'^2(u)]w_\gamma(\mathrm{d}u).
\end{aligned}$$

□

Lemma 5. *Let* $\ell(n)$ *be a slowly varying function,* $\beta > \frac{1}{2}$ *and* $|a_n| = n^{-\beta}\ell(n)$, $n \geqslant 1$.

(i) *If* $(p+1)(2\beta-1) > 1$, *then* $\Xi_{n,p} = O(n)$.
(ii) *If* $(p+1)(2\beta-1) < 1$, *then* $\Xi_{n,p} = O[n^{2-(p+1)(2\beta-1)}\ell^{2(p+1)}(n)]$.
(iii) *If* $(p+1)(2\beta-1) = 1$, *then* $\Xi_{n,p} = O(n)\left[\sum_{i=1}^n|\ell^{p+1}(i)|/i\right]^2$.

Proof. By Karamata's theorem, $A_n(i) = O[n^{1-i\beta}\ell^i(n)]$ for $i \geqslant 2$. Since ℓ is a slowly varying function, it is easily seen that, for $i \geqslant n$, $\Theta_{n+i,p} - \Theta_{i,p} = O(n\theta_{i,p})$. Therefore,

$$\Xi_{n,p} \leqslant n\Theta_{n,p}^2 + \sum_{i=1}^n \Theta_{n+i,p}^2 + \sum_{i=n+1}^\infty(\Theta_{n+i,p} - \Theta_{i,p})^2 = O(n\Theta_{2n,p}^2) + O(n^3\theta_{n,p}^2).$$

where another application of Karamata's theorem is used for $\sum_{i=n+1}^\infty \theta_{i,p}^2$.

(i) In this case $\theta_{n,p}$ is summable over n and hence $\Xi_{n,p} = O(n)$ easily follows.

(ii) This is an easy consequence of $\Theta_{2n,p} = O[n^{-(p+1)(2\beta-1)}\ell^{2(p+1)}(n)]$ by a third application of Karamata's theorem.

(iii) Since $(p+1)(2\beta-1) = 1$,

$$\Xi_{n,p} = O(n)\left[\sum_{i=1}^{2n} \frac{|\ell^{p+1}(i)|}{i}\right]^2 + O(n\ell^{2(p+1)}(n)).$$

We now argue that $\hat{\ell}(n) = \sum_{i=1}^{n} |\ell^{p+1}(i)|/i$ is also a slowly varying function. Note that $\hat{\ell}$ is non-increasing, and it suffices to verify $\lim_{n\to\infty} \hat{\ell}(2n)/\hat{\ell}(n) = 1$. For any $G > 1$, by properties of slowly varying functions,

$$\lim_{m\to\infty} \frac{\sum_{i=m}^{mG} |\ell^{p+1}(i)|/i}{|\ell^{p+1}(m)|} = \log G.$$

Thus

$$\limsup_{n\to\infty} \frac{\hat{\ell}(2n) - \hat{\ell}(n)}{\hat{\ell}(n)} \leqslant \limsup_{n\to\infty} \frac{\sum_{i=1+n}^{2n} |\ell^{p+1}(i)|/i}{\sum_{i=n/G}^{n} |\ell^{p+1}(i)|/i} = \frac{\log 2}{\log G}$$

implies that $\hat{\ell}$ is slowly varying by taking $G \to \infty$ and (iii) follows. □

The next three lemmas consider the existence of K_∞ and F and their higher-order derivatives. In particular, Lemma 6 imposes conditions such that the expectation and differentiation operators can be exchanged; Lemma 7 provides expressions for $F^{(r)}$ and $F_n^{(r)}$; and Lemma 8 gives sufficient conditions for $K_\infty \in \mathcal{C}^p$ so that the expansion (3) is meaningful.

Lemma 6. *Let X and Y be two independent random variables such that X has density $f_X \in \mathcal{C}^p$ and $\mathrm{E}(|Y|^\gamma) < \infty$ for some $\gamma \geqslant 0$. Assume that*

$$\sum_{r=0}^{p} \int_{\mathbb{R}} |f_X^{(r)}(t)|^2 w_\gamma(\mathrm{d}t) < \infty. \tag{21}$$

Then F_Z, the distribution function of $Z = X + Y$, is also in $\mathcal{C}^p$ and

$$F_Z^{(r)}(z) = \mathrm{E} F_X^{(r)}(z - Y), \qquad 0 \leqslant r \leqslant p. \tag{22}$$

Moreover, for $C = \mathrm{E}[(1 + |Y|)^\gamma]$, we have

$$\begin{aligned}&\int_{\mathbb{R}} |F_Z^{(r)}(u + \delta + \eta) - F_Z^{(r)}(u + \eta)|^2 w_\gamma(\mathrm{d}u)\\ &\quad \leqslant C\delta^2 (1 + |\delta|)^\gamma (1 + |\eta|)^\gamma \int_{\mathbb{R}} |f_X^{(r)}(u)|^2 w_\gamma(\mathrm{d}u), \qquad 0 \leqslant r \leqslant p,\end{aligned} \tag{23}$$

$$\int_{\mathbb{R}} |F_Z^{(r-1)}(u+\delta+\eta) - F_Z^{(r-1)}(u+\eta) - \delta F_Z^{(r)}(u+\eta)|^2 w_\gamma(\mathrm{d}u)$$

$$\leqslant C\delta^4(1+|\delta|)^\gamma(1+|\eta|)^\gamma \int_{\mathbb{R}} |f_X^{(r)}(u)|^2 w_\gamma(\mathrm{d}u), \qquad 1 \leqslant r \leqslant p, \tag{24}$$

and

$$\int_{\mathbb{R}} |F_Z^{(r)}(u)|^2 w_\gamma(\mathrm{d}u) \leqslant C \int_{\mathbb{R}} |F_X^{(r)}(u)|^2 w_\gamma(\mathrm{d}u), \qquad 1 \leqslant r \leqslant p. \tag{25}$$

Proof. By Lemma 4, $\sum_{i=0}^{p-1} \sup_s |f_X^{(i)}(s)| < \infty$. Using conditioning, $F_Z(z) = \mathrm{E}F_X(z-Y)$. Then the Lebesgue dominated convergence theorem asserts that $F_Z'(z) = \mathrm{E}F_X'(z-Y)$ by letting $\delta \to 0$ in

$$\frac{F_Z(z+\delta) - F_Z(z)}{\delta} = \mathrm{E}\frac{F_X(z-Y+\delta) - F_X(z-Y)}{\delta}.$$

Higher-order derivatives follows similarly in a recursive way and hence (22) holds. To establish (24), by (22) and Cauchy's inequality,

$$|F_Z^{(r-1)}(u+\delta+\eta) - F_Z^{(r-1)}(u+\eta) - \delta F_Z^{(r)}(u+\eta)|^2$$
$$\leqslant \mathrm{E}\Big[|F_X^{(r-1)}(u-Y+\delta+\eta) - F_X^{(r-1)}(u-Y+\eta) - \delta F_X^{(r)}(u-Y+\eta)|^2\Big].$$

So (24) results from Lemma 2 with $q = 1$. A similar argument yields (23) and (25) via (22). $\square$

Lemma 7. *Assume* (5) *and* $\mathrm{E}[|\varepsilon_l|^{\max(\gamma,2)}] < \infty$. *Then for all* $m \geqslant \kappa$, $n \geqslant 0$ *and* $0 \leqslant r \leqslant p$,

$$F_{m+n}^{(r)}(z) = \mathrm{E}F_m^{(r)}\left(z - \sum_{l=m}^{m+n-1} a_l \varepsilon_{\kappa-l}\right), \tag{26}$$

$$F^{(r)}(z) = \mathrm{E}F_m^{(r)}\left(z - \sum_{l=m}^{\infty} a_l \varepsilon_{\kappa-l}\right). \tag{27}$$

Moreover, there exists a $C > 0$ *such that for all* n, $\kappa \leqslant n \leqslant \infty$,

$$\int_{\mathbb{R}} |F_n^{(r)}(u+\delta+\eta) - F_n^{(r)}(u+\eta)|^2 w_\gamma(\mathrm{d}u)$$

$$\leqslant C\delta^2(1+|\delta|)^\gamma(1+|\eta|)^\gamma \int_{\mathbb{R}} |f_\kappa^{(r)}(u)|^2 w_\gamma(\mathrm{d}u), \qquad 0 \leqslant r \leqslant p, \tag{28}$$

$$\int_{\mathbb{R}} |F_n^{(r-1)}(u+\delta+\eta) - F_n^{(r-1)}(u+\eta) - \delta F_n^{(r)}(u+\eta)|^2 w_\gamma(\mathrm{d}u)$$

$$\leqslant C\delta^4(1+|\delta|)^\gamma(1+|\eta|)^\gamma \int_{\mathbb{R}} |f_\kappa^{(r)}(u)|^2 w_\gamma(\mathrm{d}u), \qquad 1 \leqslant r \leqslant p, \tag{29}$$

and

$$\int_{\mathbb{R}} |F_n^{(r)}(u)|^2 w_\gamma(\mathrm{d}u) \leqslant C \int_{\mathbb{R}} |F_\kappa^{(r)}(u)|^2 w_\gamma(\mathrm{d}u), \qquad 1 \leqslant r \leqslant p. \tag{30}$$

Proof. Let X in Lemma 6 be $\overline{X}_{\kappa,1} = \sum_{l=0}^{\kappa-1} a_l \varepsilon_{\kappa-l}$. By (22), for $m \geqslant \kappa$ and $n \geqslant 0$, $F_{m+n}^{(r)}(z) = \mathrm{E}F_\kappa^{(r)}(z - \sum_{l=\kappa}^{m+n-1} a_l \varepsilon_{\kappa-l})$ and $F_m^{(r)}(u) = \mathrm{E}F_\kappa^{(r)}(u - \sum_{l=\kappa}^{m-1} a_l \varepsilon_{\kappa-l})$. Then (26) follows by letting $u = z - \sum_{l=m}^{m+n-1} a_l \varepsilon_{\kappa-l}$ in the latter identity and a smoothing argument. Letting $n = \infty$, (27) is obtained. By Lemma 1, $C = \sup_{n\geqslant 0} \mathrm{E}[(1 + |\sum_{l=n}^{\infty} a_l \varepsilon_{-l}|)^\gamma] < \infty$. Thus (29), (28) or (30) follows from (24), (23) or (25), respectively. □

Lemma 8. *Assume* (5) *and* $\mathrm{E}[|\varepsilon_l|^{\max(1+\gamma,2)}] < \infty$, *and that* $K \in \mathcal{K}(\gamma)$ *has the representation* $K(x) = \int_0^x g(t)\mathrm{d}t$. *Then*

$$K_\infty(x) = \int_{\mathbb{R}} g(t)[\mathbf{1}(0 \leqslant t) - F(t-x)]\mathrm{d}t \tag{31}$$

and $K_\infty^{(r)}(x) = -(-1)^r \int_{\mathbb{R}} g(t) F^{(r)}(t-x)\mathrm{d}t$, $r = 1, \ldots, p$.

Proof. Recall that $K_\infty(x) = \mathrm{E}[K(X_1 + x)]$. Write $K(x) = \int_{\mathbb{R}} g(t)[\mathbf{1}(0 \leqslant t) - \mathbf{1}(x \leqslant t)]\mathrm{d}t$. To prove (31), by Fubini's theorem it suffices to verify that

$$\int_{\mathbb{R}} |g(t)| \mathrm{E}[|\mathbf{1}(0 \leqslant t) - \mathbf{1}(x + X_1 \leqslant t)|]\mathrm{d}t$$

$$= \int_{-\infty}^0 |g(t)| F(t-x)\mathrm{d}t + \int_0^\infty |g(t)|[1 - F(t-x)]\mathrm{d}t < \infty.$$

Using Cauchy's inequality, $0 \leqslant F \leqslant 1$ and $K \in \mathcal{K}(\gamma)$ (hence $\int_{\mathbb{R}} g^2(t) w_{-\gamma}(\mathrm{d}t) \leqslant 1$),

$$\left[\int_{-\infty}^0 |g(t)| F(t-x)\mathrm{d}t\right]^2 \leqslant \int_{-\infty}^0 g^2(t) w_{-\gamma}(\mathrm{d}t) \int_{-\infty}^0 F(t-x) w_\gamma(\mathrm{d}t)$$

$$\leqslant \int_{-\infty}^{-x} \int_{y+x}^0 w_\gamma(\mathrm{d}t) f(y)\mathrm{d}y$$

$$\leqslant \int_{\mathbb{R}} \frac{(1+|y+x|)^{1+\gamma}}{1+\gamma} f(y)\mathrm{d}y$$

$$\leqslant (1+|x|)^{\gamma+1} \mathrm{E}(1+|X_1|)^{1+\gamma} < \infty.$$

The finiteness of the second integral follows in a similar way. Next we compute the derivatives of K_∞. Let $k(x;\,\epsilon) = [K_\infty(x+\epsilon) - K_\infty(x)]/\epsilon$ and $f(x;\,\epsilon) = [F(x) - F(x-\epsilon)]/\epsilon$. By Cauchy's inequality, (29) and (31),

$$\left|k(x;\,\epsilon) - \int_{\mathbb{R}} g(t) f(t-x)\mathrm{d}t\right|^2 \leqslant \int_{\mathbb{R}} g^2(t) w_{-\gamma}(\mathrm{d}t) \int_{\mathbb{R}} [f(t-x;\,\epsilon) - f(t-x)]^2 w_\gamma(\mathrm{d}t)$$

$$\leqslant C\epsilon^2(1+|\epsilon|)^\gamma(1+|x|)^\gamma \int_{\mathbb{R}} |f_\kappa^{(r)}(u)|^2 w_\gamma(\mathrm{d}u) = O(\epsilon^2).$$

Hence $K'_\infty(x) = \int_{\mathbb{R}} g(t) f(t-x)\mathrm{d}t$. A simple induction yields higher-order derivatives. $\square$

Lemma 9. *Assume* (5) *and* $\mathrm{E}(|\varepsilon_1|^{4+\gamma}) < \infty$. *Then*

$$\int_{\mathbb{R}} \|\mathcal{P}_1 L(\widetilde{\mathbf{X}}_n,\, t)\|^2 w_\gamma(\mathrm{d}t) = O(\theta_{n,p}^2). \tag{32}$$

Proof. For notational convenience we write θ_n for $\theta_{n,p}$. We shall first show that (32) holds for $1 \leqslant n \leqslant \kappa$. If $\theta_n = 0$, then $a_{n-1} = 0$ and hence $\mathcal{P}_1 L(\widetilde{\mathbf{X}}_n,\, t) = 0$. Thus it suffices to verify that $\int_{\mathbb{R}} \|\mathcal{P}_1 L(\widetilde{\mathbf{X}}_n,\, t)\|^2 w_\gamma(\mathrm{d}t) \leqslant \int_{\mathbb{R}} \|L(\widetilde{\mathbf{X}}_n,\, t)\|^2 w_\gamma(\mathrm{d}t) = O(1)$, which follows from (30) asserting that $\int_{\mathbb{R}} |F^{(r)}(u)|^2 w_\gamma(\mathrm{d}u) < \infty$ for $1 \leqslant r \leqslant p$ and $\int_{\mathbb{R}} \|\mathbf{1}(X_n \leqslant u) - F(u)\|^2 w_\gamma(\mathrm{d}u) < \infty$, an easy consequence of $\mathrm{E}[(1+|X_1|)^{1+\gamma}] < \infty$.

From now on we assume $n \geqslant \kappa + 1$. Set $\delta = -a_{n-1}\varepsilon_1$ and $\eta = -\underline{X}_{n,0}$. Since δ and η are independent, by Lemma 1, $\mathrm{E}[|\delta|^4(1+|\delta|)^\gamma(1+|\eta|)^\gamma] = O(a_{n-1}^4)$. So inequality (29) in Lemma 7 yields that, for $1 \leqslant \alpha \leqslant p$,

$$\int_{\mathbb{R}} \|F_n^{(\alpha-1)}(t - \underline{X}_{n,1}) - F_n^{(\alpha-1)}(t - \underline{X}_{n,0}) + F_n^{(\alpha)}(t - \underline{X}_{n,0}) a_{n-1}\varepsilon_1\|^2 w_\gamma(\mathrm{d}t) = O(a_{n-1}^4). \tag{33}$$

By (26), $F_n^{(\alpha-1)}(y) = \mathrm{E}[F_{n-1}^{(\alpha-1)}(y - a_{n-1}\varepsilon'_1) - a_{n-1}\varepsilon'_1 F_{n-1}^{(\alpha)}(y)]$. Thus by Cauchy's inequality,

$$\|F_{n-1}^{(\alpha-1)}(y) - F_n^{(\alpha-1)}(y)\| \leqslant \|F_{n-1}^{(\alpha-1)}(y) - F_{n-1}^{(\alpha-1)}(y - a_{n-1}\varepsilon'_1) + a_{n-1}\varepsilon'_1 F_{n-1}^{(\alpha)}(y)\|.$$

Again by (29) in Lemma 7,

$$\int_{\mathbb{R}} \|F_{n-1}^{(\alpha-1)}(t - \underline{X}_{n,1}) - F_n^{(\alpha-1)}(t - \underline{X}_{n,1})\|^2 w_\gamma(\mathrm{d}t) = O(a_{n-1}^4). \tag{34}$$

Combining (33) and (34),

$$\int_{\mathbb{R}} \|F_{n-1}^{(\alpha-1)}(t - \underline{X}_{n,1}) - F_n^{(\alpha-1)}(t - \underline{X}_{n,0}) + F_n^{(\alpha)}(t - \underline{X}_{n,0}) a_{n-1}\varepsilon_1\|^2 w_\gamma(\mathrm{d}t) = O(a_{n-1}^4). \tag{35}$$

Define

$$M_n^{(r)}(\widetilde{\mathbf{X}}_0,\, y) = F_n^{(r)}(y - \underline{X}_{n,0}) + \sum_{i=r}^{p} (-1)^{i+r+1} F^{(i)}(y) \mathrm{E}[U_{n,i-r} | \widetilde{\mathbf{X}}_0]. \tag{36}$$

Next we use the method of induction to establish that, for $0 \leqslant r \leqslant p$,

$$\int_{\mathbb{R}} \|M_n^{(r)}(\widetilde{\mathbf{X}}_0,\ t)\|^2 w_\gamma(\mathrm{d}t) = O[A_n(4) + A_n^{p-r+1}(2)]. \tag{37}$$

When $r = p$, $M_n^{(r)}(\widetilde{\mathbf{X}}_0,\ t) = F_n^{(p)}(t - \underline{X}_{n,0}) - F^{(p)}(t)$. By (27), $F^{(p)}(t) = \mathrm{E}F_n^{(p)}(t - \underline{X}'_{n,0})$. So

$$\|M_n^{(p)}(\widetilde{\mathbf{X}}_0,\ t)\| \leqslant \|F_n^{(p)}(t - \underline{X}_{n,0}) - F_n^{(p)}(t - \underline{X}'_{n,0})\| \leqslant 2\|F_n^{(p)}(t - \underline{X}_{n,0}) - F_n^{(p)}(t)\|$$

and, by Lemma 1 and (28) in Lemma 7,

$$\frac{1}{4}\int_{\mathbb{R}} \|M_n^{(p)}(\widetilde{\mathbf{X}}_0,\ t)\|^2 w_\gamma(\mathrm{d}t) \leqslant \int_{\mathbb{R}} \|F_n^{(p)}(t - \underline{X}_{n,0}) - F_n^{(p)}(t)\|^2 w_\gamma(\mathrm{d}t)$$

$$= O\{\mathrm{E}[|\underline{X}_{n,0}|^2(1 + |\underline{X}_{n,0}|)^\gamma]\} = O[A_n(2)].$$

Now suppose that (37) holds for $1 \leqslant r = \alpha \leqslant p$. To complete the induction it suffices to consider $r = \alpha - 1$. To this end, observing that the projection operators $\mathcal{P}_{-j}$ are orthogonal, we have

$$\frac{1}{2}\int_{\mathbb{R}} \|M_n^{(\alpha-1)}(\widetilde{\mathbf{X}}_0,\ t)\|^2 w_\gamma(\mathrm{d}t) = \frac{1}{2}\sum_{j=0}^{\infty}\int_{\mathbb{R}} \|\mathcal{P}_{-j}M_n^{(\alpha-1)}(\widetilde{\mathbf{X}}_0,\ t)\|^2 w_\gamma(\mathrm{d}t) \leqslant I_n + J_n,$$

where

$$I_n = \sum_{j=0}^{\infty}\int_{\mathbb{R}} \|\mathcal{P}_{-j}F_n^{(\alpha-1)}(t - \underline{X}_{n,0}) + F_{n+j+1}^{(\alpha)}(t - \underline{X}_{n,-j-1})a_{n+j}\varepsilon_{-j}\|^2 w_\gamma(\mathrm{d}t)$$

and

$$J_n = \sum_{j=0}^{\infty}\int_{\mathbb{R}} \left\| F_{n+j+1}^{(\alpha)}(t - \underline{X}_{n,-j-1})a_{n+j}\varepsilon_{-j} - \sum_{i=\alpha-1}^{p}(-1)^{i+\alpha}F^{(i)}(y)\mathcal{P}_{-j}\mathrm{E}[U_{n,i-\alpha+1}|\widetilde{\mathbf{X}}_0]\right\|^2 w_\gamma(\mathrm{d}t).$$

Observe that by (26) in Lemma 7,

$$\mathcal{P}_{-j}F_n^{(\alpha-1)}(t - \underline{X}_{n,0}) = F_{n+j}^{(\alpha-1)}(t - \underline{X}_{n,-j}) - F_{n+j+1}^{(\alpha-1)}(t - \underline{X}_{n,-j-1}).$$

Thus (35) ensures that $I_n = O(\sum_{j=0}^{\infty} a_{n+j}^4) = O[A_n(4)]$. Since $\mathcal{P}_{-j}\mathrm{E}[U_{n,i-\alpha+1}|\widetilde{\mathbf{X}}_0] = a_{n+j}\varepsilon_{-j}\mathrm{E}[U_{n,i-\alpha}|\widetilde{\mathbf{X}}_{-j-1}]$ if $i \geqslant \alpha$ and vanishes if $i = \alpha - 1$, the induction is now completed since by the induction hypothesis

$$J_n = \sum_{j=0}^{\infty}|a_{n+j}|^2\int_{\mathbb{R}} \|M_{n+j+1}^{(\alpha)}(\widetilde{\mathbf{X}}_0,\ t)\|^2 w_\gamma(\mathrm{d}t) = \sum_{j=0}^{\infty} a_{n+j}^2 O[A_{n+j+1}(4) + A_{n+j+1}^{p-\alpha+1}(2)]$$

$$= O[A_n(4)] + A_n(2)O[A_{n+1}(4) + A_{n+1}^{p-\alpha+1}(2)] = O[A_n(4) + A_{n+1}^{p-\alpha+2}(2)].$$

Let $R_n(t) = \mathcal{P}_1\mathbf{1}(X_n \leqslant t) + a_{n-1}\varepsilon_1 F'_n(t - \underline{X}_{n,0})$. Since $\mathcal{P}_1\mathbf{1}(X_n \leqslant t) = F_{n-1}(t - \underline{X}_{n,1}) - F_n(t - \underline{X}_{n,0})$, by (35), $\int_{\mathbb{R}}\|R_n(t)\|^2 w_\gamma(\mathrm{d}t) = O(a_{n-1}^4)$. Observe that $R_n(t) = \mathcal{P}_1 L(\widetilde{\mathbf{X}}_n,\ t) + a_{n-1}\varepsilon_1 M_n^{(1)}(\widetilde{\mathbf{X}}_0,\ t)$. Then by (37) with $r = 1$, we have

$$\int_{\mathbb{R}} \|\mathcal{P}_1 L(\widetilde{\mathbf{X}}_n,\ t)\|^2 w_\gamma(\mathrm{d}t) \leqslant 2\int_{\mathbb{R}} [\|R_n(t)\|^2 + \|a_{n-1}\varepsilon_1 M_n^{(1)}(\widetilde{\mathbf{X}}_0,\ t)\|^2] w_\gamma(\mathrm{d}t) = O(\theta_n^2),$$

completing the proof. □

Lemma 10. *Assume that* $\mathrm{E}(\varepsilon_1^4) < \infty$ *and that, for all* $0 \leqslant i \leqslant p$, $\sup_{s\in\mathbb{R}}|f_1^{(i)}(s)| < \infty$. *Then for all* s, $\|\mathcal{P}_1 L(\widetilde{\mathbf{X}}_n, s)\| = O(\theta_n)$.

Proof. The argument in the proof of Lemma 9 can be easily transplanted here with the integral $\int_{\mathbb{R}} H(t) w_\gamma(\mathrm{d}t)$ (say) replaced by $H(t)$. For example, since $\sum_{i=0}^{p} \sup_{s\in\mathbb{R}}|f_1^{(i)}(s)| < \infty$, (33) now becomes

$$\|F_n^{(\alpha-1)}(t - \underline{X}_{n,1}) - F_n^{(\alpha-1)}(t - \underline{X}_{n,0}) + F_n^{(\alpha)}(t - \underline{X}_{n,0}) a_{n-1}\varepsilon_1\|^2 = O(a_{n-1}^4);$$

(35) becomes

$$\|F_{n-1}^{(\alpha-1)}(t - \underline{X}_{n,1}) - F_n^{(\alpha-1)}(t - \underline{X}_{n,0}) + F_n^{(\alpha)}(t - \underline{X}_{n,0}) a_{n-1}\varepsilon_1\|^2 = O(a_{n-1}^4);$$

and (37) becomes

$$\|M_n^{(r)}(\widetilde{\mathbf{X}}_0, t)\|^2 = O[A_n(4) + A_n^{p-r+1}(2)].$$

It is easily seen that the induction in the proof of Lemma 9 still holds here. Thus Lemma 10 follows in a similar way. □

Lemma 11. *Under the conditions of Theorem 1,*

$$\int_{\mathbb{R}} \|S_n(t;\, p)\|^2 w_\gamma(\mathrm{d}t) = O(\Xi_{n,p}). \tag{38}$$

Proof. Let $\lambda_n^2 = \int_{\mathbb{R}} \|\mathcal{P}_1 L(\widetilde{\mathbf{X}}_n, t)\|^2 w_\gamma(\mathrm{d}t)$ and $a \vee b = \max(a, b)$. Note that $\mathcal{P}_j S_n(t;\, p)$, $-\infty < j \leqslant n$, are orthogonal and $\mathcal{P}_j L(\widetilde{\mathbf{X}}_l;\, t) = 0$ when $l < j$. Thus

$$\int_{\mathbb{R}} \|S_n(t;\, p)\|^2 w_\gamma(\mathrm{d}t) = \sum_{j=-\infty}^{n} \int_{\mathbb{R}} \|\mathcal{P}_j S_n(t;\, p)\|^2 w_\gamma(\mathrm{d}t)$$

$$\leqslant \sum_{j=-\infty}^{n} \mathrm{E}\int_{\mathbb{R}} \left\{ \sum_{l=1\vee j}^{n} \frac{[\mathcal{P}_j L(\widetilde{\mathbf{X}}_l;\, t)]^2}{\lambda_{l-j+1}} \right\} \left\{ \sum_{l=1\vee j}^{n} \lambda_{l-j+1} \right\} w_\gamma(\mathrm{d}t) = \sum_{j=-\infty}^{n} \left[\sum_{l=1\vee j}^{n} \lambda_{l-j+1} \right]^2$$

entails (38) since $\lambda_n = O(\theta_n)$ by Lemma 9. □

Lemma 12. *Let* $W_n(y;\, p) = \sum_{m=1}^{n} J(\widetilde{\mathbf{X}}_m, y)$, *where*

$$J(\widetilde{\mathbf{X}}_m, y) = F_\kappa(y - \underline{X}_{m,m-\kappa}) - \sum_{r=0}^{p} (-1)^r F^{(r)}(y) \sum_{\kappa\leqslant i_1<\ldots<i_r} \prod_{q=1}^{r} a_{i_q}\varepsilon_{m-i_q}.$$

Then, under the conditions of Theorem 2, we have

$$\int_{\mathbb{R}} \|\mathcal{P}_1 J(\widetilde{\mathbf{X}}_n, t)\|^2 w_\gamma(\mathrm{d}t) + \int_{\mathbb{R}} \|\mathcal{P}_1 \partial J(\widetilde{\mathbf{X}}_n, t)/\partial t\|^2 w_\gamma(\mathrm{d}t) = O(\theta_{n,p}^2) \tag{39}$$

and

$$\int_{\mathbb{R}} \|W_n(y;\, p)\|^2 w_\gamma(\mathrm{d}y) + \int_{\mathbb{R}} \|\partial W_n(y;\, p)/\partial y\|^2 w_\gamma(\mathrm{d}y) = O(\Xi_{n,p}), \tag{40}$$

while under the conditions of Theorem 3 we have

$$\int_{\mathbb{R}} \|\mathcal{P}_1 \partial J(\widetilde{\mathbf{X}}_n,\, t)/\partial t\|^2 \mathrm{d}t + \int_{\mathbb{R}} \|\mathcal{P}_1 \partial^2 J(\widetilde{\mathbf{X}}_n,\, t)/\partial t^2\|^2 \mathrm{d}t = O(\theta_{n,p}^2) \tag{41}$$

and

$$\int_{\mathbb{R}} \|\partial W_n(y;\, p)/\partial y\|^2 \mathrm{d}y + \int_{\mathbb{R}} \|\partial^2 W_n(y;\, p)/\partial y^2\|^2 \mathrm{d}y = O(\Xi_{n,p}). \tag{42}$$

Proof. The same argument in the proof of Lemma 9 yields that (7) and the moment condition $\mathrm{E}(|\varepsilon_1|^{4+\gamma}) < \infty$ imply (39), which leads to (40) in a similar manner to Lemma 11. The proof for (41) and (42) proceeds in a similar way. $\square$

Let $d_i(s) = \mathbf{1}(X_i \leqslant s) - \mathrm{E}[\mathbf{1}(X_i \leqslant s)|\widetilde{\mathbf{X}}_{i-1}]$, $D_n(s) = \sum_{i=1}^n d_i(s)$ and $G_n(s) = D_n(s)/\sqrt{n}$. A chain-type argument is used in the proof of Lemma 13. Lemma 14 concerns a functional central limit theorem for $G_n(s)$ in the Skorokhod space $\mathcal{D}(\mathbb{R})$. It is assumed that f_1 exists in these two lemmas. In Lemma 15 we consider the general case in which there exists a $\kappa \in \mathbb{N}$ such that the density f_κ exists.

Lemma 13. *Assume that* $\int_{\mathbb{R}} f_1^2(t) w_\gamma(\mathrm{d}t) < \infty$ *and* $\mathrm{E}(|X_1|^{1+\gamma}) < \infty$. *Then*

$$\mathrm{E}\left[\sup_{s\in\mathbb{R}} (1+|s|)^\gamma |G_n(s)|^2\right] = O(\log^2 n).$$

Proof. For $k \in \mathbb{Z}$, let $p_k(t) = \lfloor 2^k t\rfloor/2^k$ and $q_k(t) = \lfloor 2^k t + 1\rfloor/2^k$, where $\lfloor u\rfloor = \max\{k \in \mathbb{Z}\colon k \leqslant u\}$. Set $N = \lfloor 2\log_2 n\rfloor$. Then by the triangle and Cauchy's inequalities,

$$\begin{aligned}
|G_n(t)|^2 &\leqslant \left[|G_n(p_0(t))| + \sum_{k=1}^N |G_n(p_k(t)) - G_n(p_{k-1}(t))| + |G_n(t) - G_n(p_N(t))|\right]^2 \\
&\leqslant (N+2)\Bigg[|G_n(p_0(t))|^2 + \sum_{k=1}^N |G_n(p_k(t)) - G_n(p_{k-1}(t))|^2 \\
&\quad + |G_n(t) - G_n(p_N(t))|^2\Bigg].
\end{aligned} \tag{43}$$

For the first two terms in the preceding display, observe that

$$\sup_{t\in\mathbb{R}}(1+|t|)^\gamma |G_n(p_0(t))|^2 \leqslant \sum_{l\in\mathbb{Z}}(2+|l|)^\gamma |G_n(l)|^2, \tag{44}$$

$$\sup_{t\in\mathbb{R}}(1+|t|)^\gamma |G_n(p_k(t)) - G_n(p_{k-1}(t))|^2 \leqslant \sum_{l\in\mathbb{Z}}\left(1+\frac{|l|+1}{2^k}\right)^\gamma \left|G_n\left(\frac{l}{2^k}\right) - G_n\left(\frac{l-1}{2^k}\right)\right|^2. \tag{45}$$

After elementary manipulations, expectations of both terms are of order $O(1)$ by using the martingale structure in G_n, $\mathrm{E}[|G_n(l)|^2] \leqslant F(l)(1-F(l))$ and $\mathrm{E}[|G_n(x)-G_n(y)|^2] \leqslant |F(x)-F(y)|$, together with the moment condition $\mathrm{E}(|X_1|^{1+\gamma}) < \infty$. As to the third term, we have $|F_1(y)-F_1(x)| \leqslant \int_x^y f_1^2(u) w_\gamma(\mathrm{d}u) \int_x^y w_{-\gamma}(\mathrm{d}u)$ by Cauchy's inequality. Note that $\sup_{t\in\mathbb{R}}(1+|t|)^\gamma \int_{p_N(t)}^{q_N(t)} w_{-\gamma}(\mathrm{d}u) = O(2^{-N})$. Again by Cauchy's inequality, for all $t\in\mathbb{R}$,

$$\begin{aligned}
&(1+|t|)^\gamma \left\{\sum_{i=1}^n \mathrm{E}[\mathbf{1}(p_N(t) < X_i \leqslant q_N(t))|\widetilde{\mathbf{X}}_{i-1}]\right\}^2 \\
&\leqslant (1+|t|)^\gamma \left\{\sum_{i=1}^n \int_{p_N(t)}^{q_N(t)} f_1(v - \underline{X}_{i,i-1})\mathrm{d}v\right\}^2 \\
&= \frac{O(n)}{2^N}\sum_{i=1}^n \int_{p_N(t)}^{q_N(t)} f_1^2(v - \underline{X}_{i,i-1}) w_\gamma(\mathrm{d}v) \\
&\leqslant \frac{O(n)}{2^N}\sum_{i=1}^n (1+|\underline{X}_{i,i-1}|)^\gamma \int_{\mathbb{R}} f_1^2(v) w_\gamma(\mathrm{d}v)
\end{aligned}$$

which entails $\mathrm{E}[\sup_{t\in\mathbb{R}}(1+|t|)^\gamma |G_n(t) - G_n(p_N(t))|^2] = O(1)$ by (45) and

$$\begin{aligned}
&-\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathrm{E}[\mathbf{1}(p_N(t) < X_i \leqslant q_N(t))|\widetilde{\mathbf{X}}_{i-1}] \leqslant G_n(t) - G_n(p_N(t)) \\
&\leqslant G_n(q_N(t)) - G_n(p_N(t)) + \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathrm{E}[\mathbf{1}(p_N(t) < X_i \leqslant q_N(t))|\widetilde{\mathbf{X}}_{i-1}].
\end{aligned}$$

Therefore the lemma follows from (43), (44) and (45). □

Lemma 14. *Assume* $\int_{\mathbb{R}}[f_1^2(s) + |f_1'(s)|^2]\mathrm{d}s < \infty$. *Then* $G_n(s) \Rightarrow G(s)$ *in the space* $\mathcal{D}(\mathbb{R})$, *where* $G(s)$ *is a Gaussian process with mean* 0 *and covariance function* $\mathrm{E}[G(s)G(t)] = \mathrm{E}[d_1(s)d_1(t)]$.

Proof. The martingale central limit theorem clearly entails the finite-dimensional convergence of $G_n(\cdot)$. For the tightness, we need to show that for each ϵ, $\eta > 0$, there is a $\delta > 0$ such that $P[\sup_{|s-t|<\delta}|G_n(s) - G_n(t)| > \epsilon] \leqslant \eta$ for large n. Since

$$\bigcup_{|s-t|<\delta}\{|G_n(s)-G_n(t)|>\epsilon\} \subseteq \bigcup_{k\in\mathbb{Z}}\left\{\sup_{k\delta\leqslant t\leqslant(k+1)\delta}|G_n(t)-G_n(k\delta)|>\frac{\epsilon}{3}\right\},$$

the tightness follows from the stronger statement

$$\sum_{k\in\mathbb{Z}} P\left[\sup_{k\delta\leqslant t\leqslant(k+1)\delta}|G_n(t)-G_n(k\delta)|>\epsilon\right]\leqslant\eta \tag{46}$$

for large n. To this end, we shall use the argument of Theorem 22.1 in Billingsley (1968). Let $c_0=\sup f_1(s)$, $\delta=\epsilon^4\eta$, $m=\lfloor 8\epsilon^3\eta c_0\sqrt{n}\rfloor+1$ and $p=\delta/m<\epsilon/(8c_0\sqrt{n})$; let $I_k=I_k(\delta)$ be the interval $[k\delta,(k+1)\delta]$. Observe that $\mathrm{E}[\mathbf{1}(s\leqslant X_i\leqslant s+p)|\tilde{\mathbf{X}}_{i-1}]\leqslant pc_0$. Then as in the proof of the inequality (22.17) in Billingsley (1968), we have, for $s\leqslant t\leqslant s+p$, that

$$|G_n(t)-G_n(s)|\leqslant|G_n(s+p)-G_n(s)|+pc_0\sqrt{n},$$

which implies, as in the proof of the inequality (22.18) in Billingsley (1968), that

$$\sup_{t\in I_k}|G_n(t)-G_n(k\delta)|\leqslant 3\max_{i\leqslant m}|G_n(k\delta+ip)-G_n(k\delta)|+pc_0\sqrt{n}. \tag{47}$$

For $y>x$, let $d_i=d_i(y)-d_i(x)$. Then $|d_i|\leqslant 1$ and $\mathrm{E}(d_i^2)\leqslant F(y)-F(x)$. In the rest of the proof, C stands for a constant which may vary from line to line. By the Burkholder inequality,

$$\begin{aligned}
\mathrm{E}[|G_n(y)-G_n(x)|^4] &\leqslant \frac{C}{n^2}\mathrm{E}[(d_1^2+\ldots+d_n^2)^2]\\
&\leqslant \frac{C}{n^2}\mathrm{E}\left\{\sum_{i=1}^n[d_i^2-\mathrm{E}(d_i^2|\mathbf{X}_{i-1})]\right\}^2+\frac{C}{n^2}\mathrm{E}\left[\sum_{i=1}^n\mathrm{E}(d_i^2|\mathbf{X}_{i-1})\right]^2\\
&\leqslant \frac{C}{n}\|d_1^2-\mathrm{E}(d_1^2|\mathbf{X}_0)\|^2+C\|\mathrm{E}(d_1^2|\mathbf{X}_0)\|^2\\
&\leqslant \frac{C}{n}[F(y)-F(x)]+C\mathrm{E}\{[F_1(y-\underline{X}_{1,0})-F_1(x-\underline{X}_{1,0})]^2\},
\end{aligned} \tag{48}$$

where in the third inequality we have used the orthogonality of the martingale differences $d_i^2-\mathrm{E}(d_i^2|\mathbf{X}_{i-1})$, $1\leqslant i\leqslant n$. Let $\alpha_k=\alpha_k(\delta)=\mathrm{E}[\sup_{z\in I_k}f_1^2(z-\underline{X}_{1,0})]$. By Lemma 4, and noting that $\int_{\mathbb{R}}f_1^2(u-\underline{X}_{1,0})\mathrm{d}u=\int_{\mathbb{R}}f_1^2(u)\mathrm{d}u$,

$$\begin{aligned}
\delta\sum_{k\in\mathbb{Z}}\alpha_k &\leqslant \delta\sum_{k\in\mathbb{Z}}\mathrm{E}\left[\frac{2}{\delta}\int_{I_k}f_1^2(u-\underline{X}_{1,0})\mathrm{d}u+2\delta\int_{I_k}f_1'^2(u-\underline{X}_{1,0})\mathrm{d}u\right]\\
&=2\int_{\mathbb{R}}f_1^2(u)\mathrm{d}u+2\delta^2\int_{\mathbb{R}}f_1'^2(u)\mathrm{d}u.
\end{aligned} \tag{49}$$

Note that for $x, y\in I_k$, $|F_1(y-\underline{X}_{1,0})-F_1(x-\underline{X}_{1,0})|\leqslant|y-x|\sup_{z\in I_k}f_1(z-\underline{X}_{1,0})$. Then

$$\mathrm{E}\{[F_1(y-\underline{X}_{1,0})-F_1(x-\underline{X}_{1,0})]^2\}\leqslant(y-x)^2\alpha_k.$$

For $1\leqslant i\leqslant m$, define

$$Z_i = Z_{i,k,n}(p, \delta) = G_n(ip + k\delta) - G_n((i-1)p + k\delta),$$

$$\Delta_i = \Delta_{i,k}(p, \delta) = P((i-1)p + k\delta \leqslant X_1 \leqslant ip + k\delta)$$

and $u_i = u_{i,k,n}(p, \delta) = \sqrt{\Delta_i/n} + p\sqrt{\alpha_k}$. Then, for $1 \leqslant i < j \leqslant m$,

$$\mathrm{E}[(Z_{i+1} + \ldots + Z_j)^4] \leqslant C(u_{i+1} + \ldots + u_j)^2 \tag{50}$$

by letting $x = ip + k\delta$ and $y = jp + k\delta$ in (48). Theorem 12.2 in Billingsley (1968) asserts that (50) implies

$$P\left[\max_{0\leqslant i\leqslant m}|Z_1 + \ldots + Z_i| \geqslant \frac{\epsilon}{8}\right] \leqslant \frac{C}{\epsilon^4}(u_1 + \ldots + u_m)^2 \leqslant \frac{C}{\epsilon^4}\left[\frac{m}{n}P(X_1 \in I_k) + m^2p^2\alpha_k\right]$$

since $\Delta_1 + \ldots + \Delta_m = P(X_1 \in I_k)$. Thus (46) follows from (47), (49) and

$$\sum_{k\in\mathbb{Z}} P\left[\sup_{t\in I_k}|G(t) - G(k\delta)| > \frac{\epsilon}{2}\right] \leqslant \frac{C}{\epsilon^4}\left[\frac{m}{n} + \delta^2\sum_{k\in\mathbb{Z}}\alpha_k\right] = C\eta\left[\frac{1}{np} + \delta\sum_{k\in\mathbb{Z}}\alpha_k\right]$$

by noticing that $\sum_{k\in\mathbb{Z}}P(X_1 \in I_k) = 1$, $np \to \infty$ and $pc_0\sqrt{n} \leqslant \epsilon/8$. □

Lemma 15. *Let* $G_n^*(s) = n^{-1/2}\sum_{m=1}^n[\mathbf{1}(X_m \leqslant s) - \mathrm{E}(\mathbf{1}(X_m \leqslant s)|\widetilde{\mathbf{X}}_{m-\kappa})]$.

(i) *Assume that* $\mathrm{E}(|X_1|^{1+\gamma}) < \infty$, f_κ *exists for some* $\kappa \in \mathbb{N}$ *and* $\int_\mathbb{R} f_\kappa^2(t)w_\gamma(\mathrm{d}t) < \infty$. *Then*

$$\mathrm{E}\left[\sup_{s\in\mathbb{R}}(1 + |s|)^\gamma|G_n^*(s)|^2\right] = O(\log^2 n).$$

(ii) *Assume that* $\int_\mathbb{R}|[f_\kappa^2(s) + |f'_\kappa(s)|^2]\mathrm{d}s < \infty$ *for some integer* $\kappa > 0$. *Then the process* $\{G_n^*(s),\ s \in \mathbb{R}\}$ *is tight and* $\sup_{s\in\mathbb{R}}|G_n^*(s)| = O_P(1)$.

Proof. For $1 \leqslant j \leqslant \kappa$, let

$$M_{n,j}^*(s) = \sum_{i=0}^{n-1}[\mathbf{1}(X_{i\kappa+j} \leqslant s) - \mathrm{E}(\mathbf{1}(X_{i\kappa+j} \leqslant s)|\widetilde{\mathbf{X}}_{(i-1)\kappa+j})].$$

(i) A similar argument as in the proof of Lemma 13 ensures that

$$\mathrm{E}\left[\sup_{s\in\mathbb{R}}(1 + |s|)^\gamma|M_{n,j}^*(s)|^2\right] = O(\log^2 n),$$

for each $1 \leqslant j \leqslant \kappa$.

(ii) Similarly, a careful examination of the proof of Lemma 14 reveals that the process $\{n^{-1/2}M_{n,j}^*(s),\ s \in \mathbb{R}\}$ is tight and converges to a Gaussian process with mean 0 and covariance function

$$\Gamma(s,\ t) = \mathrm{E}\{[\mathbf{1}(X_\kappa \leqslant s) - \mathrm{E}(\mathbf{1}(X_\kappa \leqslant s)|\widetilde{\mathbf{X}}_0)][\mathbf{1}(X_\kappa \leqslant t) - \mathrm{E}(\mathbf{1}(X_\kappa \leqslant t)|\widetilde{\mathbf{X}}_0)]\},$$

for each $1 \leqslant j \leqslant \kappa$. So the lemma follows in view of $G_{n\kappa}^*(s) = \sum_{j=1}^\kappa M_{n,j}^*(s)/\sqrt{n\kappa}$. □

With Lemmas 1–15 established, we can prove our main results.

Proof of Theorem 1. Let $K(x) = \int_0^x g_K(t)\mathrm{d}t$. By Lemma 8, under condition (5) we can write $S_n(K) = -\int_{\mathbb{R}} g_K(t) S_n(t;\, p)\mathrm{d}t$. Hence Cauchy's inequality gives

$$\mathrm{E}^*\left[\sup_{K\in\mathcal{K}(\gamma)} |S_n(K;\, p)|^2\right] \leqslant \mathrm{E}^*\left[\sup_{K\in\mathcal{K}(\gamma)} \int_{\mathbb{R}} g_K^2(t) w_{-\gamma}(\mathrm{d}t)\int_{\mathbb{R}} |S_n(t;\, p)|^2 w_\gamma(\mathrm{d}t)\right]$$
$$\leqslant \int_{\mathbb{R}} \|S_n(t;\, p)\|^2 w_\gamma(\mathrm{d}t)$$

which proves the theorem by (38) of Lemma 11. □

Proof of Theorem 2. Without loss of generality, let $\kappa = 1$. Define

$$V_{m,r} = \sum_{1\leqslant i_1<\ldots<i_r} \prod_{q=1}^{r} a_{i_q}\varepsilon_{m-i_q}.$$

Then $U_{m,r} - V_{m,r} = \varepsilon_m \sum_{1\leqslant i_2<\ldots<i_r} \prod_{q=2}^{r} a_{i_q}\varepsilon_{m-i_q}$ form stationary martingale differences and thus $\|\sum_{i=1}^n (U_{i,r} - V_{i,r})\|^2 = O(n)$. Now write

$$S_n(y;\, p) = \sqrt{n} G_n(y) + W_n(y;\, p) + \sum_{m=1}^{n}\sum_{r=1}^{p} (-1)^r F^{(r)}(y)(V_{m,r} - U_{m,r}), \tag{51}$$

where $W_n(y;\, p) = \sum_{m=1}^n J(\widetilde{\mathbf{X}}_m,\, y)$ and

$$J(\widetilde{\mathbf{X}}_i,\, y) = F_1(y - \underline{X}_{i,i-1}) - \sum_{r=0}^{p} (-1)^r F^{(r)}(y) V_{i,r}$$

is $\sigma(\mathbf{X}_{i-1})$ measurable. By Lemma 13, $\mathrm{E}\left[\sup_{s\in\mathbb{R}}(1+|s|)^\gamma |G_n(s)|^2\right] = O(\log^2 n)$. To complete the proof it suffices to verify that

$$\sup_{s\in\mathbb{R}}[(1+|s|)^\gamma |f^{(r)}(s)|^2] < \infty, \qquad 0 \leqslant r \leqslant p-1, \tag{52}$$

and

$$\mathrm{E}\left[\sup_{y\in\mathbb{R}}(1+|y|)^\gamma |W_n(y;\, p)|^2\right] = O(\Xi_{n,p}). \tag{53}$$

Let $g_r(s) = (1+|s|)^\gamma |f_1^{(r)}(s)|^2$. By Lemma 4, $\sup_{s\in\mathbb{R}} g_r(s) < \infty$ under (7). Hence by (27) and Cauchy's inequality, (52) follows from

$$\sup_{s\in\mathbb{R}}[(1+|s|)^\gamma |f^{(r)}(s)|^2] \leqslant \mathrm{E}\left\{\sup_{s\in\mathbb{R}}[(1+|s|)^\gamma |f_1^{(r)}(s-\underline{X}_{1,0})|^2]\right\}$$

$$\leqslant \mathrm{E}\left\{\sup_{s\in\mathbb{R}}[(1+|s-\underline{X}_{1,0}|)^\gamma |f_1^{(r)}(s-\underline{X}_{1,0})|^2](1+|\underline{X}_{1,0}|)^\gamma\right\}$$

$$\leqslant C\sup_{s\in\mathbb{R}} g_r(s)\mathrm{E}[(1+|\underline{X}_{1,0}|)^\gamma] < \infty.$$

For (53), again by Lemma 4,

$$\mathrm{E}\left[\sup_{y\in\mathbb{R}}(1+|y|)^\gamma \left|W_n(y;\ p)\right|^2\right] \leqslant 2^{1+2\gamma}\int_{\mathbb{R}}[\|W_n(y;\ p)\|^2 + \|\partial W_n(y;\ p)/\partial y\|^2]w_\gamma(\mathrm{d}y)$$

$$= O(\Xi_{n,p})$$

due to (40) in Lemma 12. □

Proof of Corollary 1. The case $I=0$ follows from Theorem 1 and Lemma 5. For $I>0$, to avoid non-essential complications we consider the special case $I=1$. Let $K(x, s) = K_1(x)\mathbf{1}(x \leqslant s) + K_2(x)\mathbf{1}(x > s)$, where $K_i \in \mathcal{K}(\gamma)$ and $|K_i(s)| \leqslant (1+|s|)^{\gamma/2}$; let $K_1^*(x, s) = K_1(x)\mathbf{1}(x \leqslant s) + K_1(s)\mathbf{1}(x > s)$. Then for all s, $K_1^*(\cdot, s) - K_1^*(0, s) \in \mathcal{K}(\gamma)$. By Theorem 1,

$$\mathrm{E}^*\left[\sup_{K_1\in\mathcal{K}(\gamma), s\in\mathbb{R}} |S_n(K_1^*(\cdot, s);\ p)|^2\right] = O(\Xi_{n,p}).$$

Since $|K_1(s)| \leqslant (1+|s|)^{\gamma/2}$, by Theorem 2,

$$\mathrm{E}\left[\sup_{t\in\mathbb{R}}(1+|t|)^\gamma |S_n(\mathbf{1}(\cdot > t);\ p)|^2\right] = O\big(n\log^2 n + \Xi_{n,p}\big),$$

which implies (9) by

$$\mathrm{E}^*\left[\sup_{K_1\in\mathcal{K}(\gamma), s\in\mathbb{R}} |S_n(K_1(\cdot)\mathbf{1}(\cdot > s);\ p)|^2\right] = O\big(n\log^2 n + \Xi_{n,p}\big)$$

in view of $S_n(L+M;\ p) = S_n(L;\ p) + S_n(M;\ p)$. □

Proof of Theorem 3. (i) Once again we assume without loss of generality that $\kappa = 1$. The finite-dimensional convergence easily results from Lemmas 3 and 10 by letting $\xi_i = \widetilde{\mathbf{X}}_i$ and $h(\xi_i) = L(\widetilde{\mathbf{X}}_i, y)$ since $\sum_{n=1}^\infty \theta_{n,p} < \infty$. For tightness, we shall use (51). By Lemma 4, $\sup_{s\in\mathbb{R}}|f_1^{(r)}(s)| < \infty$ for $r \leqslant p+1$. Hence by (27), $\sup_{s\in\mathbb{R}}|f^{(r)}(s)| < \infty$. Lemma 15 guarantees that $G_n(s)$ is tight. By Lemma 4 and (42) in Lemma 12,

$$\mathrm{E}\left[\sup_{y\in\mathbb{R}}\left|\frac{\partial W_n(s)}{\partial s}\right|^2\right] \leqslant 2\int_{\mathbb{R}}[\|\partial W_n(y;\ p)/\partial y\|^2 + \|\partial^2 W_n(y;\ p)/\partial y^2\|^2]\mathrm{d}y = O(\Xi_{n,p}).$$

Observe that $O(\Xi_{n,p}) = O(n)$ by parts (i) and (ii) of Lemma 5. Then $W_n(s)/\sqrt{n}$ is tight since $|W_n(s) - W_n(t)| \leqslant |t-s|\sup_{s\in\mathbb{R}} |\partial W_n(s)/\partial s|$.

(ii) By Corollary 1 with $\gamma = 0$, $\sup_s|S_n(s;\, p+1)|/\sigma_{n,p+1} = o_P(1)$ since $(p+1)(2\beta-1) < 1$. Hence (14) follows from $\sum_{k=1}^n U_{n,p+1}/\sigma_{n,p+1} \Rightarrow Z_{p+1,\beta}$ (Surgailis, 1982). $\square$

Proof of Corollary 2. Assume $\kappa = 1$.

(i) This trivially follows from part (i) of Theorem 3 with $p = 0$ and $\beta = 1$.

(ii) We shall use the decomposition (51). By Lemma 15, $\sup_{y\in\mathbb{R}}|G_n(y)| = O_P(1)$. Since $\sup_y f(y) < \infty$ and $V_{m,1} - U_{m,1} = -\varepsilon_m$, it then suffices to establish that $\sum_{m=1}^n J(\widetilde{\mathbf{X}}_i, s)/\tilde{\sigma}_n \Rightarrow f(s)Z$ in the space $\mathcal{D}(\mathbb{R})$. By Lemma 4, (40) in Lemma 12 and Lemma 5, (12) yields

$$\mathrm{E}\left\{\sup_{y\in\mathbb{R}}\left|\sum_{i=1}^n [F_1(y - \underline{X}_{i,i-1}) - F(y) + f(y)\underline{X}_{i,i-1}]\right|^2\right\} = O(\Xi_{n,1}) = O(n),$$

which completes the proof in view of $\sum_{i=1}^n \underline{X}_{i,i-1}/\tilde{\sigma}_n \Rightarrow Z$ (see, for example, Davydov 1970) and $\sqrt{n} = o(\tilde{\sigma}_n)$. $\square$

Proof of Theorem 4. Observe that the class $\{H^q(\cdot, \mathbf{m})/C\colon \mathbf{m} \in \mathcal{M}(\delta_0),\ 1 \leqslant q \leqslant d\}$ is a subset of $\mathcal{K}_p$ under Assumption 1 and conditions in Theorem 4. So Corollary 3 is applicable; and Theorem 4 then follows from the standard argument for asymptotic distributions of M-estimators (see Theorem 5.21 in van der Vaart 1998). $\square$

Acknowledgements

The author is very grateful to Jan Mielniczuk for his valuable suggestions. The comments of the referee lead to a significant improvement in the paper.

References

Andrews, D.W.K. and Pollard, D. (1994) An introduction to functional central limit theorems for dependent stochastic processes. *Internat. Statist. Rev.*, **62**, 119–132.

Arcones, M. (1996) Weak convergence of stochastic processes indexed by smooth functions. *Stochastic Process. Appl.*, **62**, 115–138.

Arcones, M. and Yu, B. (1994) Limit theorems for empirical processes under dependence, In C. Houdré and V. Pérez-Abreu (eds), *Proceedings in Chaos Expansions, Multiple Ito–Wiener integrals and Their Applications*, pp. 205–221. Boca Raton, FL: CRC Press.

Bae, J. and Levental, S. (1995) Uniform CLT for Markov chains and its invariance principle: A martingale approach. *J. Theoret. Probab.*, **8**, 549-570.

Beran, J. (1991) M-estimators of location for Gaussian and related processes with slowly decaying serial correlations. *J. Amer. Statist. Assoc.*, **86**, 704–708.

Beran, J. (1994) *Statistics for Long-Memory Processes*. London: Chapman & Hall.

Billingsley, P. (1968) *Convergence of Probability Measures*. New York: Wiley.

Davydov, Yu.A. (1970) The invariance principle for stationary processes. *Theory Probab. Appl.*, **15**, 487–498.

de la Peña, V. and Giné, E (1999) *Decoupling: From Dependence to Independence.* New York: Springer-Verlag.

Dehling, H. and Taqqu, M. (1989) The empirical process of some long-range dependent sequences with an application to U-statistics. *Ann. Statist.*, **17**, 1767–1783.

Doukhan, P., Massart, P. and Rio, E. (1995) Invariance principles for absolutely regular empirical processes. *Ann. Inst. H. Poincaré Probab. Statist.*, **31**, 393–427.

Giraitis, L. and Surgailis, D. (2002) The reduction principle for the empirical process of a long memory linear process. In H. Dehling, T. Mikosch and M. Sørensen (eds), *Empirical Process Techniques for Dependent Data*, pp. 241–255. Boston: Birkhäuser.

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

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

Hoeffding, W. and Robbins, H. (1948) The central limit theorem for dependent random variables. *Duke Math. J.*, **15**, 773–780.

Koul, H.L. and Surgailis, D. (1997) Asymptotic expansion of M-estimators with long-memory errors. *Ann. Statist.*, **25**, 818–850.

Koul, H.L. and Surgailis, D. (2002) Asymptotic expansion of the empirical process of long memory moving averages. In H. Dehling, T. Mikosch and M. Sørensen (eds), *Empirical Process Techniques for Dependent Data*, pp. 213–239. Boston: Birkhäuser.

Major, P. (1981) *Multiple Wiener–Ito integrals: with Applications to Limit Theorems*. Berlin: Springer-Verlag.

Rio, E. (1998) Absolutely regular empirical processes and universal entropy. *Probab. Theory Related Fields*, **111**, 585–608.

Surgailis, D. (1982) Zones of attraction of self-similar multiple integrals. *Lithuanian Math. J.*, **22**, 327–340.

Taqqu, M.S. (1975) Weak convergence to fractional Brownian-motion and to Rosenblatt process. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **31**, 287–302.

van der Vaart, A.W. (1998) *Asymptotic Statistics*. Cambridge: Cambridge University Press.

van der Vaart, A.W. and Wellner, J. (1996) *Weak Convergence and Empirical Processes. With Applications to Statistics.* New York: Springer-Verlag.

Woodroofe, M. (1992) A central limit theorem for functions of a Markov chain with applications to shifts. *Stochastic Process. Appl.*, **41**, 33–44.

Wu, T.J. (1995) Adaptive root *n* estimates of integrated squared density derivatives. *Ann. Statist.*, **23**, 1474–1495.

Received January 2002 and revised February 2003