Bernoulli **11**(5), 2005, 933–948

On invariant distribution function estimation for continuous-time stationary processes

DOMINIQUE DEHAY

Laboratoire de Statistique, UFR Sciences Sociales, Université de Rennes 2 Haute Bretagne, Campus de Villejean, 35043 Rennes Cedex, France. E-mail: dominique.dehay@uhb.fr

This paper is concerned with the asymptotic behaviour of the empirical distribution function for a large class of continuous-time weakly dependent stationary processes. Under mild mixing conditions the empirical distribution function is an unbiased consistent estimator of the marginal distribution function of the process. For strongly mixing processes this estimator is asymptotically normal. We propose a consistent estimator of the asymptotic variance, and then study the functional central limit theorem for the empirical distribution function.

Keywords: asymptotic normality; central limit theorem; consistency; continuous time; empirical distribution function; mixing condition; stationary process; weak convergence

1. Introduction

The development of theory on the empirical distribution function for continuous-time stationary processes has received relatively little attention in the literature, in contrast to the discrete-time case, and particularly for sequences of independent and identically distributed (i.i.d.) random variables. The aim of the present paper is to investigate the asymptotic behaviour of the empirical distribution function for a large class of continuous-time weakly dependent stationary processes. This includes exponentially ergodic Markov processes, ergodic diffusion processes, Markov jump processes as well as nonlinear (and linear) transformations of Brownian motion.

For a sequence of real-valued independent variables $(Y_n)_{n\geqslant 1}$ with common distribution function $F(\cdot)$, the empirical distribution function

$$\hat{F}_n(x) := \frac{1}{n}\sum_{k=1}^{n} \mathbb{I}_{\{Y_k<x\}}$$

is a consistent estimator of $F(x)$ and

$$\hat{G}_n(x) := \sqrt{n}\big\{\hat{F}_n(x) - F(x)\big\} \to \mathcal{N}\big(0, \sigma^2(x)\big) \qquad \text{in law,}$$

as $n \to \infty$, where $\mathcal{N}(0, \sigma^2(x))$ is the zero-mean normal distribution with variance $\sigma^2(x) = F(x)(1 - F(x))$. In this situation $\hat{\sigma}^2_n(x) := \hat{F}_n(x)\big(1 - \hat{F}_n(x)\big)$ is a consistent estimator of $\sigma^2(x)$. Furthermore, the empirical process $\hat{G}_n := \big\{\hat{G}_n(x), x \in \mathbb{R}\big\}$ is weakly convergent to a generalized Brownian bridge in the space $\mathbb{D}(\mathbb{R})$ endowed with the uniform topology (see

1350–7265

Dudley 1984; Shorak and Wellner 1986; Pollard 1990), $\mathbb{D}(\mathbb{R})$ denoting the set of right-continuous real-valued functions defined on $\mathbb{R}$ with limits from the left.

These results are still applicable under weak dependence conditions. For instance, under the φ-mixing (uniformly mixing) condition $\sum n^2\varphi_n^{1/2} < \infty$, Billingsley (1968) proved the weak convergence in the space $\mathbb{D}(\mathbb{R})$ of the empirical process $\hat{G}_n$ to a stationary Gaussian process G with covariance

$$\operatorname{cov}\{G(x),\, G(y)\} = F(x \wedge y) - F(x)F(y) + \sum_{k>0}\{F_k(x,\, y) + F_k(y,\, x) - 2F(x)F(y)\},$$

where $F_k(x,\, y) := \mathrm{P}\{X_0 \leqslant x,\, X_k \leqslant y\}$, the real-valued series being absolutely convergent. For α-mixing (strongly mixing) sequences with $\alpha_n = \mathcal{O}(n^{-\gamma})$ as $n \to \infty$ for some $\gamma > 1$, Rio (2000) stated the same result if the distribution function $F(\cdot)$ is continuous on $\mathbb{R}$. Recently Dehling and Philipp (in Dehling *et al.* 2002) presented a survey of the results and techniques for dependent discrete-time data (see also the other contributions in Dehling *et al.* 2002).

For continuous-time ergodic stationary processes, Davydov (2001) studied the empirical measure. He stated that the existence of a local time is a necessary and sufficient condition for the almost sure convergence in variation of the empirical measure to the stationary marginal law of the process. Here we do not consider this point of view and we refer to Davydov's paper for further details on this subject.

The empirical distribution function of an ergodic diffusion process is a consistent and asymptotically normal estimator of the stationary marginal distribution function (see Negri 1998; Kutoyants 2003). In the following we improve and extend these results on the empirical distribution function

$$\hat{F}_T(x) = \frac{1}{T}\int_0^T \mathbb{I}_{\{X_s \leqslant x\}}\, \mathrm{d}s$$

for a wider class of continuous-time weakly dependent stationary processes $X := \{X_t,\, t \geqslant 0\}$. No continuity condition is assumed on the stationary marginal distribution function $F(\cdot)$.

The paper is organized as follows. In Section 2 we present the weak-dependence notions used in the rest of the paper. Then in Section 3 we recall standard results on the convergence of $\hat{F}(x)$ for each $x \in \mathbb{R}$. If the process X satisfies mild mixing conditions then $\hat{F}_T(x)$ is an unbiased consistent estimator of $F(x)$. We study the rate of convergence in quadratic mean. Hence we derive the rate of almost sure convergence and the uniform strong consistency.

Under strong-mixing conditions, $\hat{G}_T(x) := \sqrt{T}\big(\hat{F}_T(x) - F(x)\big)$ is asymptotically normal. As the variance $\sigma^2(x)$ of the asymptotic normal distribution depends on the distribution of the process X, an estimate of this asymptotic variance is often needed for statistical purposes. In Section 4 we construct a consistent estimator $\hat{\sigma}^2(x)$ of $\sigma^2(x)$ which is more tractable than the specific one presented by Dehay and Kutoyants (2004) for ergodic diffusion processes. Furthermore, we evaluate the rate of convergence in quadratic mean and the rate of almost sure convergence of $\hat{\sigma}^2(x)$ to $\sigma^2(x)$.

Then in Section 5 we investigate the weak convergence of the empirical process

$\hat{G}_T := \{\hat{G}_T(x), x \in \mathbb{R}\}$ in a convenient functional space. In contrast to the usual framework (see Kutoyants 2003), we do not assume that $F(\cdot)$ is continuous. However, following Pollard (1990) and Rio (2000), we consider the semi-metric space $\mathbb{X} := (\mathbb{R}, \rho)$ where $\rho(x, y) := |F(x) - F(y)|$. We state the weak convergence of $\hat{G}_T$ in the space $\mathbb{D}(\mathbb{X})$ endowed with the uniform topology, $\mathbb{D}(\mathbb{X})$ denoting the set of right-continuous functions from $\mathbb{X}$ into $\mathbb{R}$ with limits from the left. Then we deduce the more standard statement of weak convergence in the space $\mathbb{D}(\mathbb{R})$ endowed with the uniform topology. Section 6 is devoted to the proof of the weak convergence.

2. Background and hypotheses

From now on, we consider a real-valued stationary process $X = \{X_t, t \geqslant 0\}$ defined on a probability space $(\Omega, \mathcal{A}, \mathrm{P})$ with stationary distribution function $F(\cdot)$. The marginal stationary law is denoted by $\mu : F(x) = \mu(-\infty, x] = \mathrm{P}\{X_t \leqslant x\}$. It is assumed that the process X is measurable with respect to the σ-algebra $\mathcal{B}((0, \infty]) \otimes \mathcal{A}$, and that the probability space $(\Omega, \mathcal{A}, \mathrm{P})$ is complete. If the process is also ergodic, we have

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T g(X_s)\,\mathrm{d}s = \int_{\mathbb{R}} g(x)\,\mathrm{d}\mu(x), \qquad \text{almost surely,}$$

for any real-valued bounded Borel function $g : \mathbb{R} \to \mathbb{R}$. Let $\mu_t^{(2)}$ be the two-dimensional law of (X_s, X_{s+t}) which does not depend on s since the process is stationary. We require the following weak-dependence condition:

(AIW) $m_T(A) \to \mu \otimes \mu(A)$ as $T \to \infty$, for any $A \in \mathcal{B}(\mathbb{R}^2)$. The probability measure m_T is defined on $\mathbb{R}^2$ by

$$m_T(A) := \frac{1}{T}\int_0^T \mu_t^{(2)}(A)\,\mathrm{d}t.$$

As the process is measurable, the function $t \mapsto \mu_t^{(2)}(A)$ is measurable for any $A \in \mathcal{B}(\mathbb{R}^2)$, and the previous integral is well defined. Under condition (AIW) we have

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \mathrm{cov}\{g_1(X_0), g_2(X_t)\}\mathrm{d}t = 0,$$

for all real-valued bounded Borel functions $g_1, g_2 : \mathbb{R} \to \mathbb{R}$.

To obtain the asymptotic variance and covariance (Lemma 3.2), we need more asymptotic independence. Based on the condition given by Castellana and Leadbetter (1986) in the density estimation context, we introduce the following condition:

(CL*) $\Gamma_T(A) \to \Gamma(A)$ as $T \to \infty$, for any $A \in \mathcal{B}(\mathbb{R}^2)$ and for some signed measure Γ on $\mathbb{R}^2$. The signed measure Γ_T is defined on $\mathbb{R}^2$ by

$$\Gamma_T(A) := \int_0^T \left\{\mu_t^{(2)}(A) - \mu \otimes \mu(A)\right\}\mathrm{d}t. \tag{2.1}$$

This *Castellana–Leadbetter type condition* implies that

$$\int_0^\infty \operatorname{cov}\{g_1(X_0),\ g_2(X_t)\}\mathrm{d}t := \lim_{T\to\infty}\int_0^T \operatorname{cov}\{g_1(X_0),\ g_2(X_t)\}\mathrm{d}t = \iint_{\mathbb{R}^2} g_1(x)g_2(y)\,\mathrm{d}\Gamma(x,\ y)$$

for all real-valued bounded Borel functions $g_1,\ g_2 : \mathbb{R}\to\mathbb{R}$.

Finally, we recall the strong-mixing (α-mixing) condition (see Doukhan 1984):

(SM) Consider the σ-fields $\sigma^s(X) := \sigma\{X(u) : u \leqslant s\}$ and $\sigma_s(X) := \sigma\{X(u) :\ u \geqslant s\}$ for any $s \geqslant 0$, and define the strong-mixing coefficient function $\alpha : [0,\ \infty)\to[0,\ \infty)$ of the process X by

$$\alpha(t) := \sup\{|\mathrm{P}(A\cap B) - \mathrm{P}(A)\,\mathrm{P}(B)| : A\in\sigma^s(X),\ B\in\sigma_{s+t}(X) \text{ and } s\in[0,\ \infty)\}.$$

The process X is said to be *strongly mixing* when $\alpha(t)\to 0$ as $t\to\infty$.

Remark 1. Setwise convergence of sequences of measures in conditions (AWI) and (CL*) is stronger than weak convergence. However, it is weaker than convergence in variation used by Davydov (2001).

Remark 2. Assume that the two-dimensional law $\mu_t^{(2)}$ has a density function $f_t(\cdot,\ \cdot)$ with respect to Lebesgue measure on $\mathbb{R}^2$ such that $(t,\ x,\ y)\mapsto f_t(x,\ y)$ is measurable and

$$\int_0^\infty |f_t(x,\ y) - f(x)f(y)|\,\mathrm{d}t \leqslant c(x,\ y),$$

where $\int\int_{\mathbb{R}^2} c(x,\ y)\,\mathrm{d}x\,\mathrm{d}y < \infty$ for some non-negative measurable function $c(\cdot,\ \cdot)$, $f(\cdot)$ being the density function of the marginal stationary law. Then condition (CL*) is satisfied. Furthermore, the measure Γ is absolutely continuous with respect to Lebesgue measure in $\mathbb{R}^2$ with density function

$$\int_0^\infty \{f_t(x,\ y) - f(x)f(y)\}\mathrm{d}t.$$

Note that the condition given by Castellana and Leadbetter (1986) entails that $\Gamma_T(A)\to\Gamma(A)$ as $T\to\infty$ for any bounded Borel $A\in\mathcal{B}(\mathbb{R}^2)$. In this case the results stated below in Sections 3.1 and 3.2 hold if we replace $\hat{F}_T(x)$ by $\hat{F}_T(x) - \hat{F}_T(y)$ and $F(x)$ by $F(x) - F(y)$ for $x > y$. In this paper, for simplicity of exposition, we do not develop this point of view.

Remark 3. Assume that the stationary process X is a Markov process with transition kernel P^t, $t\geqslant 0$. Thus $\mu_t^{(2)}(\mathrm{d}x,\ \mathrm{d}y) = P^t(x,\ \mathrm{d}y)\mu(\mathrm{d}x)$. If the Markov process X is 1-ergodic in the sense that $\lim_{t\to\infty}\|P^t(x,\ \cdot) - \mu\|_1 = 0$, then condition (AIW) is satisfied. Here $\|\nu\|_1$ denotes the total variation of the signed measure ν. If X is 1-exponentially ergodic, that is $\|P^t(x,\ \cdot) - \mu\|_1 \leqslant M(x)\rho^t$ with some $0<\rho<1$ and some $M(\cdot)\in\mathrm{L}^1(\mu)$, then

$$|\operatorname{cov}\{g_1(X_s),\ g_2(X_{s+t})\}| \leqslant \sup_{x\in\mathbb{R}}|g_1(x)|\sup_{x\in\mathbb{R}}|g_2(x)|\int_{\mathbb{R}} M(x)\mu(\mathrm{d}x)\,\rho^t$$

for all $s, t \geqslant 0$ and for all real-valued bounded Borel functions $g_1, g_2 : \mathbb{R} \to \mathbb{R}$. Hence condition (CL*) is satisfied. Moreover, the asymptotic distribution Γ is defined as

$$\Gamma(A) = \int_0^\infty \left(\iint_A \{P^t(x, \mathrm{d}y) - \mu(\mathrm{d}y)\}\mu(\mathrm{d}x) \right) \mathrm{d}t, \qquad A \in \mathcal{B}(\mathbb{R}^2).$$

Actually we can establish that any 1-exponentially ergodic process is strongly mixing with $\alpha(\cdot) \in \mathrm{L}^1[0, \infty)$. Meyn and Tweedie (1993a; 1993b) stated Lyapunov type conditions on the infinitesimal generator of the Markov process X which ensure 1-ergodicity and 1-exponential ergodicity.

Remark 4. For stationary ergodic diffusion processes, Veretennikov (1988) presented sufficient conditions on the drift and the diffusion coefficients for the strong-mixing property. Veretennikov (1999) also gave relationships between the strong-mixing property and his conditions (CL1)–(CL2). We refer to these papers for more details.

We readily obtain the following relationships between these notions of asymptotic independence.

Lemma 2.1. (SM) $\Rightarrow$ (AIW), (CL*) $\Rightarrow$ (AIW), *and* $\alpha(\cdot) \in \mathrm{L}^1[0, \infty) \Rightarrow$ (CL*).

In general, the reverse implications are not valid.

3. Consistency

In this section, we study the convergence of $\hat{F}_T(x)$ as $T \to \infty$, for a stationary process $X := \{X_t, t \geqslant 0\}$. As the process X is stationary, $\hat{F}_T(x)$ is an unbiased estimator of $F(x)$: $\mathrm{E}\,\hat{F}_T(x) = F(x)$ for any $x \in \mathbb{R}$.

3.1. Convergence in quadratic mean

For any $x \in \mathbb{R}$, we have

$$\begin{aligned} \operatorname{var} \hat{F}_T(x) &= \frac{2}{T} \int_0^T \left(1 - \frac{t}{T}\right) \left(F_t(x, x) - F(x)^2\right) \mathrm{d}t \\ &= \frac{2}{T} \int_0^T \left(1 - \frac{t}{T}\right) \Gamma_t\{(-\infty, x] \times (-\infty, y]\} \mathrm{d}t, \end{aligned}$$

where $F_t(x, y) := \mathrm{P}\{X_0 \leqslant x, X_t \leqslant y\}$ and the measure Γ_t is defined by (2.1) for each $t \geqslant 0$. Thus we can readily check the convergence in quadratic mean of the estimator $\hat{F}_T(x)$.

Theorem 3.1. *Let X be a stationary process which satisfies* (AIW). *Then*

$$\lim_{T\to\infty} \hat{F}_T(x) = F(x) \qquad \textit{in quadratic mean},$$

for any $x \in \mathbb{R}$.

Hence for any stationary process satisfying one of the two conditions of Theorem 3.3, and for any probability measure ν on $\mathbb{R}$, the ν-integrated quadratic error converges to 0:

$$\lim_{T\to\infty} \int_{\mathbb{R}} \mathrm{E}\Big\{ \big(\hat{F}_T(x) - F(x)\big)^2 \Big\} \mathrm{d}\nu(x) = 0.$$

We can also evaluate the asymptotic covariance between the two estimators $\hat{F}_T(x)$ and $\hat{F}_T(y)$.

Lemma 3.2. *If the stationary process* X *satisfies* (CL*)*, then*

$$\begin{aligned} C(x, y) := \lim_{T\to\infty} T \operatorname{cov}\big\{\hat{F}_T(x), \hat{F}_T(y)\big\} &= \int_0^\infty \{F_t(x, y) + F_t(y, x) - 2F(x)F(y)\} \mathrm{d}t \\ &= \Gamma\{(-\infty, x] \times (-\infty, y]\} + \Gamma\{(-\infty, y] \times (-\infty, x]\} \end{aligned}$$

for all $x, y \in \mathbb{R}$. *The integral is well defined.*

Thus under (CL*) we deduce the asymptotic variance

$$\sigma^2(x) := \lim_{T\to\infty} T \operatorname{var} \hat{F}_t(x) = 2\int_0^\infty \big\{F_t(x, x) - F(x)^2\big\} \mathrm{d}t = 2\,\Gamma\{(-\infty, x] \times (-\infty, x]\}. \quad (3.1)$$

If the process X is strongly mixing with $\alpha(t) = \mathrm{o}(\mathrm{t}^{-\gamma})$ as $t \to \infty$ for some $0 \leqslant \gamma < 1$, then we have

$$\lim_{T\to\infty} T^\gamma \operatorname{var} \hat{F}_T(x) = 0.$$

3.2. Almost sure convergence

The Borel–Cantelli lemma yields the ergodicity of the process X (see also Davydov 2001). Next we specify the rate of convergence.

Theorem 3.3. *Assume that the stationary process* X *satisfies one of the two following conditions:* (i) $\alpha(t) = \mathcal{O}(t^{-\gamma})$ *as* $t \to \infty$ *for some* $0 < \gamma < 1$*;* (ii) (CL*). *Then*

$$\lim_{T\to\infty} T^\delta |\hat{F}_T(x) - F(x)| = 0 \qquad \textit{almost surely},$$

for any $x \in \mathbb{R}$ *and for any* $\delta < \gamma/3$*; in case (ii) we take* $\gamma = 1$.

Proof. The proof of the almost sure convergence under condition (i) is quite similar to but requires slightly more cumbersome calculus than under condition (ii). Thus for simplicity of exposition we only present the proof under condition (ii).

Assume that condition (ii) is satisfied. Let $a > 1/(1-2\delta)$. Thanks to the Bienaymé–Chebyshev inequality and convergence (3.1), we have

$$\sum_{n=1}^{\infty} \mathrm{P}\left\{ n^{a\delta} |\hat{F}_{n^a}(x) - F(x)| > \epsilon \right\} \leqslant 2\,\epsilon^{-2} \sum_{n=1}^{\infty} n^{a(2\delta-1)} < \infty$$

for every $\epsilon > 0$. Applying the Borel–Cantelli lemma, we readily obtain that

$$\lim_{n\to\infty} n^{a\delta}\left\{ \hat{F}_{n^a}(x) - F(x) \right\} = 0 \qquad \text{almost surely.}$$

For $n^a \leqslant T < (n+1)^a$, we have

$$|\hat{F}_{n^a}(x) - \hat{F}_T(x)| \leqslant \frac{T-n^a}{Tn^a}\int_0^{n^a} \mathbb{I}_{\{X_t \leqslant x\}}\,\mathrm{d}t + \frac{1}{T}\int_{n^a}^{T} \mathbb{I}_{\{X_t \leqslant x\}}\,\mathrm{d}t \leqslant 2\left\{ \left(1+\frac{1}{n}\right)^a - 1 \right\}.$$

We deduce that

$$\lim_{n\to\infty} \sup_{n^a \leqslant T < (n+1)^a} |n^{a\delta}\left\{ \hat{F}_{n^a}(x) - F(x) \right\} - T^{\delta}\left\{ \hat{F}_T(x) - F(x) \right\}| = 0 \qquad \text{almost surely,}$$

for any $a < 1/\delta$. Now we can easily complete the proof of the theorem. □

Now, by ergodicity and the Glivenko–Cantelli theorem (see Billingsley 1995, Theorem 20.6), we can deduce the uniform almost sure convergence.

Theorem 3.4. *Under the hypotheses of Theorem 3.3 the stationary process X is ergodic. Moreover, we have*

$$\mathrm{P}\left\{ \lim_{T\to\infty} \sup_x |\hat{F}_T(x) - F(x)| = 0 \right\} = 1.$$

4. Asymptotic normality

The proof of the asymptotic normality of the empirical distribution function requires more independence for the process X than condition (CL*). Thus from now on we assume that the process is strongly mixing with $\alpha(\cdot) \in \mathrm{L}^1[0, \infty)$.

4.1. Central limit theorem

Theorem 4.1. *Assume that the stationary process X is strongly mixing with $\alpha(\cdot) \in \mathrm{L}^1[0, \infty)$. Then*

$$\hat{G}_T(x) := \sqrt{T}\left\{ \hat{F}_T(x) - F(x) \right\} \to \mathcal{N}\left(0, \sigma^2(x)\right) \qquad \textit{in law,}$$

as $T \to \infty$, for any $x \in \mathbb{R}$. The asymptotic variance $\sigma^2(x)$ is defined in (3.1).

Proof. Let $x \in \mathbb{R}$ be fixed. Ibragimov and Linnik's version of the central limit theorem for mixing stationary sequences of bounded random variables (Ibragimov and Linnik 1971, Theorem 18.5.4) can be applied to the strongly mixing stationary sequence $\{Y_n(x)\}_{n\in\mathbb{N}}$ where

$$Y_n(x) := \int_{n-1}^{n} \mathbb{I}_{\{X_t \leqslant x\}}\,\mathrm{d}t - F(x).$$

Then we deduce that

$$\sqrt{n}\{\hat{F}_n(x) - F(x)\} \to \mathcal{N}\big(0, \sigma^2(x)\big) \qquad \text{in law,}$$

as $n \to \infty$. To complete the proof, it suffices to note that

$$|\sqrt{T}\{\hat{F}_T(x) - F(x)\} - \sqrt{[T]}\{\hat{F}_{[T]}(x) - F(x)\}| \leqslant 3\,[T]^{-1/2}, \tag{4.1}$$

for any $T > 0$. Here $[T]$ denotes the integer value of the real number T. □

4.2. Estimation of the asymptotic variance $\sigma^2(x)$

Here we construct a consistent estimator of the variance of the asymptotic law in Theorem 4.1. Let $x \in \mathbb{R}$. Into the expression

$$\sigma^2(x) = 2\int_0^\infty \{F_t(x, x) - F(x)^2\}\mathrm{d}t$$

we plug estimators of each of the two terms under the integral symbol. That is, we define an estimator of $\sigma^2(x)$ by

$$\hat{\sigma}_T^2(x) := 2\int_0^{T^\eta} \{\hat{F}_{t,T}(x) - \hat{F}_T(x)^2\}\mathrm{d}t$$

for some fixed $\eta > 0$, where

$$\hat{F}_{t,T}(x) := \frac{1}{T}\int_0^T \mathbb{I}_{\{X_s \leqslant x\,, X_{s+t} \leqslant x\}}\,\mathrm{d}s.$$

Proposition 4.2. *Let X be a strongly mixing stationary process with $\alpha(\cdot) \in \mathrm{L}^1[0, \infty)$. For each $0 < \eta < 1/3$, we have*

$$\lim_{T\to\infty} \hat{\sigma}_T^2(x) = \sigma^2(x) \quad \textit{in quadratic mean.}$$

Proof. The stationarity of the process X ensures that the estimators $\hat{F}_T(x)$ and $\hat{F}_{t,T}(x)$ are unbiased: $\mathrm{E}\,\hat{F}_T(x) = F(x)$ and $\mathrm{E}\,\hat{F}_{t,T}(x) = F_t(x, x)$. The mean of $\hat{\sigma}_T^2(x)$ can be expressed in the following way:

$$\mathrm{E}\,\hat{\sigma}_T^2(x) = 2\int_0^{T^\eta} \{F_t(x, x) - F(x)^2\}\mathrm{d}t - 2\,T^\eta \operatorname{var}\hat{F}_T(x). \tag{4.2}$$

As $\alpha(\cdot) \in \mathrm{L}^1[0, \infty)$ and $0 < \eta < 1/3$, the last term on the right-hand side of (4.2) converges to 0 as $T \to 0$. Thus

$$\lim_{T\to\infty} \mathrm{E}\,\hat{\sigma}_T^2(x) = 2\int_0^\infty \left\{F_t(x, x) - F(x)^2\right\}\mathrm{d}t$$

and the estimator $\hat{\sigma}_T^2(x)$ is asymptotically unbiased.

We now establish the convergence in quadratic mean. For this purpose, consider the decomposition

$$\hat{\sigma}_T^2(x) - \mathrm{E}\,\hat{\sigma}_T^2(x) = 2\int_0^{T^\eta} \left\{\hat{F}_{t,T}(x) - F_t(x, x)\right\}\mathrm{d}t$$
$$- 2\,T^\eta\left\{\hat{F}_T(x)^2 - F(x)^2\right\} - 2\,T^\eta \operatorname{var} \hat{F}_T(x).$$

We know that the last term converges to 0. As the random variable $\hat{F}_T(x) + F(x)$ is bounded, we obtain the convergence

$$\lim_{T\to\infty} T^\eta\left\{\hat{F}_T(x)^2 - F(x)^2\right\} = 0 \qquad \text{in quadratic mean.}$$

The Cauchy–Schwarz inequality gives the bound

$$\mathrm{E}\left\{\left(\int_0^{T^\eta} \{\hat{F}_{t,T}(x) - F_t(x, x)\}\mathrm{d}t\right)^2\right\} \leqslant T^\eta \int_0^{T^\eta} \operatorname{var} \hat{F}_{t,T}(x)\,\mathrm{d}t.$$

Furthermore, we have

$$\operatorname{var} \hat{F}_{t,T}(x) = \frac{2}{T^2}\int_0^T\int_0^{T-u} \operatorname{cov}\{g(X_s)g(X_{s+t}),\ g(X_{s+u})g(X_{s+u+t})\}\mathrm{d}s\mathrm{d}u$$
$$\leqslant c\left(\frac{2t}{T} + \frac{2}{T}\int_0^T \alpha(s)\,\mathrm{d}s\right),$$

for some $c > 0$. Then, as $\alpha(\cdot) \in \mathrm{L}^1[0, \infty)$ and $0 < \eta < 1/3$, we deduce that

$$\lim_{T\to\infty} \mathrm{E}\left\{\left(\int_0^{T^\eta} \left\{\hat{F}_{t,T}(x) - F_t(x, x)\right\}\mathrm{d}t\right)^2\right\} = 0.$$

This completes the proof of the lemma. □

Similarly we may state the rate of convergence in quadratic mean for the estimator $\hat{\sigma}_T^2(x)$.

Proposition 4.3. *Assume that* $\alpha(t) = \mathcal{O}(t^{-\gamma})$ *as* $t \to \infty$, *for some* $\gamma > 1$. *Then*

$$\lim_{T\to\infty} T^\delta\left\{\hat{\sigma}_T^2(x) - \sigma^2(x)\right\} = 0 \quad \text{in quadratic mean,}$$

for any $0 < \eta < 1/3$ *and any* $\delta < \min\{(1 - 3\eta)/2,\ \eta(\gamma - 1)\}$.

We also obtain almost sure convergence of $\hat{\sigma}_T^2(x)$.

Proposition 4.4. *Assume that* $\alpha(t) = \mathcal{O}(t^{-\gamma})$ *as* $t \to \infty$, *for some* $\gamma > 3/2$. *Then*

$$\lim_{T\to\infty} T^{\delta}\big\{\hat{\sigma}_T^2(x) - \sigma^2(x)\big\} = 0 \qquad \textit{almost surely,}$$

for any $0 < \eta < 1/4$ *and any* $\delta < \min\{(1-4\eta)/3, \eta(2\gamma-3)/3\}$.

The asymptotic covariance $C(x, y)$ can be estimated by

$$\hat{C}_T(x, y) := \int_0^{T^\eta} \big\{\hat{F}_{t,T}(x, y) + \hat{F}_{t,T}(y, x) - 2\,\hat{F}_T(x)\hat{F}_T(y)\big\}\mathrm{d}t,$$

where

$$\hat{F}_{t,T}(x, y) := \frac{1}{T}\int_0^T \mathbb{I}_{\{X_s \leqslant x, X_{s+t} \leqslant y\}}\,\mathrm{d}s.$$

Propositions 4.2, 4.3 and 4.4 are still valid if $\hat{\sigma}_T^2(x)$ is replaced by $\hat{C}_T(x, y)$. We also deduce the following result which can be applied for the construction of confidence intervals of $F(x)$:

Corollary 4.5. *If the stationary process is strongly mixing with* $\alpha(\cdot) \in \mathrm{L}^1[0, \infty)$, *then*

$$\frac{\sqrt{T}}{\hat{\sigma}_T(x)}\big\{\hat{F}_T(x) - F(x)\big\} \to \mathcal{N}(0, 1) \qquad \textit{in law,}$$

as $T \to \infty$, *for any* $x \in \mathbb{R}$ *such that* $\sigma^2(x) \neq 0$.

5. Weak convergence

The goal of this section is to establish a functional central limit result for the process $\hat{G}_T := \big\{\hat{G}_T(x) := \sqrt{T}\big(\hat{F}_T(x) - F(x)\big), x \in \mathbb{R}\big\}$ as $T \to \infty$. By the Cramér–Wold device we obtain the following multivariate version of Theorem 4.1.

Corollary 5.1. *Let* X *be a strongly mixing stationary process with* $\alpha(\cdot) \in \mathrm{L}^1[0, \infty)$. *Then*

$$\big(\hat{G}_T(x_1), \ldots, \hat{G}_T(x_n)\big) \to \mathcal{N}_n(0, \Sigma(x_1, \ldots, x_n)) \qquad \textit{in law,}$$

for all $n \in \mathbb{N}^*$ *and all* $x_1, x_2, \ldots, x_n \in \mathbb{R}$. *Here* $\mathcal{N}_n(0, \Sigma(x_1, \ldots, x_n))$ *designates the multivariate zero-mean normal law with covariance matrix* $\Sigma(x_1, \ldots, x_n)$ *given by* $\big(C(x_i, x_j)\big)_{1\leqslant i,j\leqslant n}$.

Thanks to Prohorov's theorem (see Billingsley 1968) it remains to prove the tightness property in a separable complete metric functional space. Notice that almost all paths of the processes $\hat{G}_T$, $T > 0$, are bounded functions from $\mathbb{R}$ into $\mathbb{R}$ which are right-continuous with limits from the left. The tightness conditions are less cumbersome to prove when almost all paths of the asymptotic process belong to a separable space of continuous functions. Thus we will consider such a space.

Let $\mathbb{X}$ be the semi-metric space $(\mathbb{R}, \rho)$ where $\rho(x, y) := |F(x) - F(y)|$. Following Pollard

(1990), we consider the space $\mathbb{B}(\mathbb{X})$ of bounded real-valued Borel functions defined on $\mathbb{X}$, endowed with the uniform metric $d(f, g) := \sup_{x\in\mathbb{X}}|f(x) - g(x)|$. We will prove that the asymptotic law of $\hat{G}_T$, as $T \to \infty$, is concentrated on the space $\mathbb{U}(\mathbb{X}) := \{g \in \mathbb{B}(\mathbb{X}) : g : \mathbb{X} \to \mathbb{R} \ \text{uniformly continuous}\}$ which is separable and complete, the space $\mathbb{X}$ being totally bounded. We adapt to our setting the chaining method as described by Rio (2000, Theorem 7.2) for the weak convergence of the empirical process for strongly mixing stationary sequences of random variables. Hence we obtain the following weak convergence of $\hat{G}_T$ in $\mathbb{B}(\mathbb{X})$:

Theorem 5.2. *Assume that $\alpha(t) = \mathcal{O}(t^{-\gamma})$ as $t \to \infty$ for some $\gamma > 1$. Then the process $\hat{G}_T$ converges in law in $\mathbb{B}(\mathbb{X})$, as $T \to \infty$, to a Gaussian process $G := \{G(x), x \in \mathbb{R}\}$ with covariance function $\operatorname{cov}\{G(x), G(y)\} = C(x, y)$. Furthermore, the asymptotic Gaussian process can be chosen such that almost all its paths are uniformly continuous from $\mathbb{X}$ into $\mathbb{R}$.*

Under the hypotheses of Theorem 5.2, almost all paths of $\hat{G}_T$ are bounded elements of $\mathbb{D}(\mathbb{R})$. Furthermore, the space $\mathbb{U}(\mathbb{X})$ is a separable closed subset of $\mathbb{D}(\mathbb{R})$ endowed with the uniform metric. Thus we deduce the following more classical result.

Corollary 5.3. *Assume $\alpha(t) = \mathcal{O}(t^{-\gamma})$ as $t \to \infty$, for some $\gamma > 1$. Then the process $\hat{G}_T$ converges in law in $\mathbb{D}(\mathbb{R})$ endowed with the uniform metric, as $T \to \infty$, to a Gaussian process.*

If the stationary distribution function $F(\cdot)$ is also continuous, then almost all paths of $\hat{G}_T$ are elements of the space $\mathcal{C}_b(\mathbb{R})$ of bounded continuous functions from $\mathbb{R}$ into $\mathbb{R}$. Moreover, the space $\mathbb{U}(\mathbb{X})$ is a separable closed subset of $\mathbb{U}(\mathbb{R}) \cap \mathcal{C}_b(\mathbb{R})$ endowed with the uniform metric. Hence we deduce the following result:

Corollary 5.4. *Assume that $F(\cdot)$ is continuous, and $\alpha(t) = \mathcal{O}(t^{-\gamma})$ as $t \to \infty$ for some $\gamma > 1$. Then the process $\hat{G}_T$ converges in law in $\mathcal{C}_b(\mathbb{R})$ endowed with the uniform convergence, as $T \to \infty$, to a Gaussian process such that almost all its paths are uniformly continuous on $\mathbb{R}$.*

6. Proof of Theorem 5.2

We have seen that we only need to verify the tightness of the sequence of processes G_T, $T > 0$. From (4.1) it suffices to consider the sequence $(G_N)_{N>0}$. For this purpose we shall verify that the following condition due to Pollard (1990, Theorem 10.2) is satisfied: for all $\epsilon > 0$ and $\eta > 0$ there is $\delta > 0$ such that

$$\limsup_{N\to\infty} \mathrm{P}^* \left\{ \sup_{\rho(x,y)<\delta} |\hat{G}_N(x) - \hat{G}_N(y)| > \eta \right\} < \epsilon, \tag{6.1}$$

where P^* is the outer probability associated with P (see also van der Vaart and Wellner 1996).

This condition and the finite-dimensional convergence (Corollary 5.1) imply that the sequence $\left(\hat{G}_N\right)_{N\in\mathbb{N}}$ weakly converges to a probability measure concentrated on $\mathbb{U}(\mathbb{X})$. Then Theorem 5.2 is proved.

We follow the proof given by Rio (2000, Theorem 7.2) in the discrete-time context. Here we replace the indicator function $\mathbb{I}_{\{X_i\in A\}}$ by the centred integrated indicator function $\Upsilon_i(A)$ defined by relation (6.3) below. Moreover, as we do not assume the continuity of $F(\cdot)$, we need to construct convenient intervals $I_{l,j}$ to control the suprema in (6.4).

Let Z_N be the stochastic measure defined on $\mathcal{B}(\mathbb{R})$ by

$$Z_N(A) := \sqrt{N}\left\{\frac{1}{N}\int_0^N \mathbb{I}_{\{X_s\in A\}}\,\mathrm{d}s - \mu(A)\right\} = \frac{1}{\sqrt{N}}\sum_{i=1}^N \Upsilon_i(A), \tag{6.2}$$

where μ is the marginal stationary law of the process $\{X_t,\ t \geqslant 0\}$, and the centred integrated indicator function is defined on $\mathcal{B}(\mathbb{R})$ by

$$\Upsilon_i(A) := \int_{i-1}^i \mathbb{I}_{\{X_s\in A\}}\,\mathrm{d}s - \mu(A). \tag{6.3}$$

As the stochastic process $\{X_t,\ t \geqslant 0\}$ is stationary with mixing coefficient function $\alpha(\cdot)$, for each $A\in\mathcal{B}(\mathbb{R})$ the stochastic sequence $(\Upsilon_i(A))_{i>0}$ is stationary with mixing coefficient $\alpha_i \leqslant \alpha(i-1)$ for $i \geqslant 1$. Notice that $\hat{G}_N(x) = Z_N(-\infty, x]$.

The generalized inverse function of $F(\cdot)$ is defined by $F^{-1}(u) := \inf\{x\in\mathbb{R} : F(x) \geqslant u\}$. This function $F^{-1} : [0, 1] \to [-\infty, \infty]$ is non-decreasing and left-continuous. Futhermore, we have $FF^{-1}(u) \geqslant u$, $F^{-1}F(x) \leqslant x$, $FF^{-1}F(x) = F(x)$ and $F^{-1}FF^{-1}(u) = F^{-1}(u)$, for any $u\in[0, 1]$ and any $x\in\mathbb{R}$. Now for $x\in\mathbb{R}$ and $k\in\mathbb{N}$, let $\pi_k(x) := F^{-1}s_k(x)$ where $s_k(x)$ is the kth partial sum of the dyadic expansion of $F(x)$, that is, $s_k(x) := 2^{-k}\left[2^kF(x)\right]$ where $\left[2^kF(x)\right]$ is the integer value of $2^kF(x)$.

Let $2^{n-1} < N \leqslant 2^n$ and $k < n$ be fixed. First of all, we estimate the supremum in (6.1) as

$$\begin{aligned}\sup_{\rho(x,y)\leqslant\delta}|\hat{G}_N(x) - \hat{G}_N(y)| \leqslant{} & 2\sup_{x\in\mathbb{R}}|Z_N(\pi_k(x), x]| \\ & + \sup_{\rho(x,y)\leqslant\delta}|\hat{G}_N(\pi_k(x)) - \hat{G}_N(\pi_k(y))|,\end{aligned} \tag{6.4}$$

for each $\delta > 0$. Moreover, we have

$$\sup_{x\in\mathbb{R}}|Z_N(\pi_k(x), x]| \leqslant \sum_{l=k+1}^n \sup_{x\in\mathbb{R}}|Z_N(\pi_{l-1}(x), \pi_l(x)]| + \sup_{x\in\mathbb{R}}|Z_N(\pi_k(x), x]|.$$

As $F(x) - 2^{-l} \leqslant F\pi_l(x) \leqslant F\pi_{l+1}(x) \leqslant F(x)$ for any $x\in\mathbb{R}$ and any $l\in\mathbb{N}$, the size of each jump of $F(\cdot)$ in the interval $(\pi_l(x), x]$ is smaller than 2^{-l}, and we can write

$$\sup_{x\in\mathbb{R}}|Z_N(\pi_k(x), x]| \leqslant \sum_{l=k+1}^n \Delta_l + \Delta_n^*$$

where

$$\Delta_n^* := \sup_{x\in\mathbb{R}} |Z_N(\pi_n(x), x]| \qquad \text{and} \qquad \Delta_l := \max_{j=1,\ldots,2^l} |Z_N(I_{l,j})|,$$

with

$$I_{l,j} := \begin{cases} \left(F^{-1}\left((j-1)2^{-l}\right), F^{-1}\left(j2^{-l}\right)\right] & \text{if } 0 \leqslant \rho\left\{F^{-1}\left(j2^{-l}\right), F^{-1}\left((j-1)2^{-l}\right)\right\} \leqslant 2^{-l}, \\ \left(F^{-1}\left((j-1)2^{-l}\right), F^{-1}\left(j2^{-l}\right)\right) & \text{if } \rho\left\{F^{-1}\left(j2^{-l}\right), F^{-1}\left((j-1)2^{-l}\right)\right\} > 2^{-l}, \end{cases}$$

for any $j = 1, \ldots, 2^l$. As

$$-\sqrt{N}\,\mu(\pi_n(x), x] \leqslant Z_N(\pi_n(x), x] = \frac{1}{\sqrt{N}} \int_0^N \mathbb{I}_{\{X_s \in (\pi_n(x), x]\}}\, \mathrm{d}t$$

for any $x \in \mathbb{R}$, we readily deduce that $\Delta_n^* \leqslant \Delta_n + \sqrt{N}\,2^{-n}$ and

$$\sup_{x\in\mathbb{R}} |Z_N(\pi_k(x), x]| \leqslant 2N^{-1/2} + 2 \sum_{l=k+1}^{n} \Delta_l. \tag{6.5}$$

If x and y are such that $\rho(x, y) \leqslant 2^{-k}$, then $s_k(x) = s_k(y)$ or $|s_k(y) - s_k(x)| = 2^{-k}$. Thus $\rho(\pi_k(x), \pi_k(y)) \leqslant 2^{-k}$. Then we have

$$\sup_{\rho(x,y)\leqslant 2^{-k}} |\hat{G}_N(\pi_k(x)) - \hat{G}_N(\pi_k(y))| \leqslant \max_{j=1,\ldots,2^k} |Z_N(I_{k,j})| = \Delta_k. \tag{6.6}$$

To prove (6.1) we need to study Δ_l in (6.5) and (6.6). Specifically, we will estimate $\mathrm{E}\,\Delta_l$. Remark that

$$\mathrm{E}\,\Delta_l = \int_0^\infty \mathrm{P}\{\Delta_l > z\}\mathrm{d}z.$$

To evaluate $\mathrm{P}\{\Delta_l > z\}$ we apply a symmetrization method (see Rio 2000). Let $(\epsilon_j)_{j>0}$ be a sequence of Rademacher random variables (i.i.d. symmetric random variables with values in $\{-1, 1\}$), independent of the process $\{X_t, t \geqslant 0\}$.

Let $z \geqslant 0$ and let J be a finite subset of $\left\{1, \ldots, 2^l\right\}$. Consider $\omega \in \Omega$ such that

$$\max_{j\in J} |Z_N(I_{l,j})(\omega)| \geqslant z.$$

Denote by $j_0 = j_0(\omega)$ the lowest integer j in J such that $|Z_N(I_{l,j})(\omega)| \geqslant z$. Then for any $\left(e_1, \ldots, e_{|J|}\right) \in \{-1, 1\}^{|J|}$, one of the two numbers

$$\sum_{j\in J\setminus\{j_0\}} e_j Z_N(I_{l,j})(\omega) + Z_N(I_{l,j_0})(\omega), \qquad \sum_{j\in J\setminus\{j_0\}} e_j Z_N(I_{l,j})(\omega) - Z_N(I_{l,j_0})(\omega)$$

is outside the interval $(-z, z)$. (Here $|J|$ denotes the number of elements in the set J.) Thus there are at least $2^{|J|-1}$ distinct values for $\left(e_1, \ldots, e_{|J|}\right)$ in $\{-1, 1\}^{|J|}$ such that

$$\left|\sum_{j\in J} e_j Z_N(I_{l,j})(\omega)\right| \geqslant z.$$

Then, thanks to the choice of the random variables ϵ_j, $j = 1, \ldots, 2^{|J|}$, we deduce the following maximal inequality:

$$\mathrm{P}\Big\{\max_{j\in J}|Z_N(I_{l,j})| \geqslant z\Big\} \leqslant 2\,\mathrm{P}\bigg\{\Big|\sum_{j\in J}\epsilon_j Z_N(I_{l,j})\Big| \geqslant z\bigg\}. \tag{6.7}$$

Let m be an integer in $\{1, \ldots, l\}$ whose value will be chosen later. For $k \in \{1, \ldots, 2^m\}$, let $J_k := \{(k-1)2^{l-m}+1, \ldots, k\,2^{l-m}\}$. Inequality (6.7) implies that

$$\mathrm{P}\{\Delta_l \geqslant z\} \leqslant \sum_{k=0}^{2^m}\mathrm{P}\Big\{\max_{j\in J_k}|Z_N(I_{l,j})| \geqslant z\Big\} \leqslant 2\sum_{k=0}^{2^m}\mathrm{P}\bigg\{\Big|\sum_{j\in J_k}\epsilon_j Z_N(I_{l,j})\Big| \geqslant z\bigg\}.$$

Then we fix the values of the random variables ϵ_i, $i = 1, \ldots, 2^{|J|}$, again. That is, let $(e_1, \ldots, e_{|J|}) \in \{-1, 1\}^{|J|}$. By (6.2) we have the decomposition

$$\sum_{j\in J_k} e_j Z_N(I_{l,j}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j\in J_k} e_j \Upsilon_i(I_{l,j}).$$

The stochastic sequence $(\sum_{j\in J_k} e_j \Upsilon_i(I_{l,j}))_{i>0}$ is stationary with strong-mixing coefficient $\alpha_i \leqslant \alpha(i-1) \leqslant c\,i^{-\gamma} \wedge 1$, for any $i \geqslant 1$ and for some $c > 0$. Then the extension of the Fuk–Nagaev inequality proved by Rio (2000, Theorem 6.2 and Relation 6.19b) applies. We obtain the bound

$$\mathrm{P}\bigg\{\sum_{i=1}^{N}\Big|\sum_{j\in J_k} e_j \Upsilon_i(I_{l,j})\Big| \geqslant 4\,z\sqrt{N}\bigg\} \leqslant 4\left(1+\frac{z^2}{rs_N^2}\right)^{-r/2} + 2c\,N r^{\gamma} z^{-\gamma-1}, \tag{6.8}$$

for all $r, z > 0$, where

$$s_N^2 := \sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\bigg|\mathrm{cov}\Big\{\sum_{j\in J_k} e_j \Upsilon_{i_1}(I_{l,j}),\, \sum_{j\in J_k} e_j \Upsilon_{i_2}(I_{l,j})\Big\}\bigg|.$$

By stationarity we have that

$$s_N^2 = N\,\mathrm{var}\Big\{\sum_{j\in J_k} e_j \Upsilon_1(I_{l,j})\Big\} + 2\sum_{i=2}^{n}(N-i+1)\mathrm{cov}\Big\{\sum_{j_1\in J_k} e_{j_1}\Upsilon_1(I_{l,j_1}),\, \sum_{j_2\in J_k} e_{j_2}\Upsilon_i(I_{l,j_2})\Big\}.$$

On the one hand, by construction of the intervals $I_{l,j}$ we have $\mu(I_{l,j}) \leqslant 2^{-l}$ for all l and j; thus we obtain that

$$\begin{aligned}\bigg|\mathrm{cov}\Big\{\sum_{j_1\in J_k} e_{j_1}\Upsilon_1(I_{l,j_1}),\, \sum_{j_2\in J_k} e_{j_2}\Upsilon_i(I_{l,j_2})\Big\}\bigg| &\leqslant \mathrm{E}\,\Upsilon_1\Big(\bigcup_{j\in J_k} I_{l,j}\Big) + \Big\{\mathrm{E}\,\Upsilon_1 \bigcup_{j\in J_k} I_{l,j}\Big\}^2 \\ &\leqslant 2\sum_{j\in J_k}\mu(I_{l,j}) \\ &\leqslant 2\times 2^{l-m}\times 2^{-l} = 2^{-m+1}.\end{aligned}$$

On the other hand, the mixing property yields the inequality

$$\left|\operatorname{cov}\left\{\sum_{j_1\in J_k} e_{j_1}\Upsilon_1(I_{l,j_1}),\ \sum_{j_2\in J_k} e_{j_2}\Upsilon_i(I_{l,j_2})\right\}\right| \leqslant 2\,\alpha_{i-1}\left\|\Upsilon_1\left(\bigcup_{j\in J_k} I_{l,j}\right)\right\|_\infty^2 \leqslant 2\,\alpha_{i-1}.$$

As $\alpha_i = \mathcal{O}(i^{-\gamma})$ as $i\to\infty$, we deduce that

$$s_N^2 \leqslant 2^{-m+1} + 4\sum_{i=1}^N \alpha_{i-1}\wedge 2^{-m} \leqslant c\,2^{m(\gamma-1)/\gamma},$$

for some $c>0$. Then, using inequality (6.8), we write

$$\begin{aligned}\mathrm{P}\left\{\left|\sum_{j\in J_k} e_j Z_N(I_{L,j})\right| \geqslant 4\,z\right\} &\leqslant \mathrm{P}\left\{\left|\sum_{i=1}^n\sum_{j\in J_k} e_j\Upsilon_i(I_{l,j})\right| \geqslant 4\,z\sqrt{n}\right\}\\ &\leqslant 4\left(1+\frac{z^2N}{2^{m(1-\gamma)/\gamma}rN}\right)^{-r/2} + 2c\,N^{(1-\gamma)/2}r^\gamma z^{-\gamma-1},\end{aligned}$$

for all $r, z>0$. The independence properties of the sequence $(\epsilon_j)_j$ entail that

$$\mathrm{P}\left\{\left|\sum_{j\in J_k}\epsilon_j Z_N(I_{l,j})\right| \geqslant 4\,z\right\} \leqslant 4\times 2^{rm(1-\gamma)/(2\gamma)}r^{r/2}z^{-r} + 2\,c\,N^{(1-\gamma)/2}r^\gamma z^{-\gamma-1}.$$

Thus we deduce that

$$\mathrm{P}\{\Delta_L \geqslant 4\,z\} \leqslant 2^{3+m(2\gamma-\gamma r+r)/(2\gamma)}r^{r/2}z^{-r} + c\,2^{2+m}N^{(1-\gamma)/2}r^\gamma z^{-\gamma-1}.$$

Taking $r = 4\gamma/(\gamma-1)$, we obtain that

$$\mathrm{P}\{\Delta_L \geqslant 4\,z\} \leqslant c\left(2^{-m}z^{-r} + 2^m N^{(1-\gamma)/2}z^{-\gamma-1}\right),$$

for some $c>0$. Thus

$$\begin{aligned}\mathrm{E}\,\Delta_l &= 4\int_0^\infty \mathrm{P}\{\Delta_l \geqslant 4z\}\mathrm{d}z\\ &\leqslant 8\,c\left(2^{-m(\gamma-1)/(4\gamma)} + 2^{m/(\gamma+1)}N^{(1-\gamma)/(2\gamma+2)}\right).\end{aligned}$$

Following Rio (2000), we choose $m\in\{1,\ldots,l\}$ such that $\mathrm{E}\,\Delta_l$ is the general term of a convergent series. For instance, $m = [l(\gamma-1)/(4\gamma)] = [l/r]$. Indeed, in this situation, as $2^{l-1}\leqslant 2^{n-1} < N$, we have $2^{-m(\gamma-1)/(4\gamma)} \leqslant 2^{1/r}\times 2^{-l/r^2}$ and $2^{m/(\gamma+1)}N^{(1-\gamma)/(2\gamma+2)} \leqslant 2^{1/r}\times 2^{-l(2\gamma-1)/(r\gamma+r)}$. We deduce that

$$\mathrm{E}\,\Delta_l \leqslant c\,2^{5-l/r^2}.$$

Hence

$$\sum_{l=k+1}^n \mathrm{E}\,\Delta_l \leqslant 32\,c\,\frac{2^{-k/r^2}}{2^{1/r^2}-1}. \tag{6.9}$$

Then from (6.4)–(6.6) and (6.9) we can readily complete the proof of the theorem using Markov's inequality.

References

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

Billingsley, P. (1995) *Probability and Measure*, 3rd edition. New York: Wiley.

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

Davydov, Yu. (2001) On convergence of empirical measures. *Statist. Inference Stochastic Process.*, **4**(1), 1–15.

Dehay, D. and Kutoyants, Yu.A. (2004) On confidence intervals for distribution function and density of ergodic diffusion process. *J. Statist. Plann. Inference*, **124**(1), 63–73.

Dehling, H., Mikosch, T. and Sørensen, M (2002) *Empirical Process Techniques for Dependent Data*. Boston: Birkhäuser.

Doukhan, P. (1984) *Mixing: Properties and Examples*, Lecture Notes in Statist. 85. Berlin: Springer-Verlag.

Dudley, R.M. (1984) A course on empirical processes. In P.L. Hennequin (ed.), *Ecole d'Été de Probabilités de Saint Flour XII–1982*, Lecture Notes in Math. 1097, pp. 1–142. Berlin: Springer-Verlag.

Ibragimov, I.A. and Linnik, Yu.V. (1971) *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff.

Kutoyants, Yu.A. (2003) *Statistical Inference for Ergodic Diffusion Processes*. London: Springer-Verlag.

Meyn, S.P. and Tweedie, R.L. (1993a) Stability of Markovian processes II: Continuous-time processes and sampled chains. *Adv. Appl. Probab.*, **25**, 487–517.

Meyn, S.P. and Tweedie, R.L. (1993b) Stability of Markovian processes III: Foster–Lyapunov criteria for continuous-time processes. *Adv. Appl. Probab.*, **25**, 518–548.

Negri, I. (1998) Stationary distribution function estimation for ergodic diffusion process. *Statist. Inference Stochastic Process.*, **1**(1), 61–84.

Pollard, D. (1990) *Empirical Processes: Theory and Applications*, NSF-CBMS Reg. Conf. Ser. Probab. Statist. 2. Hayward, CA, and Alexandrias, VA: Institute of Mathematical Statistics and American Statistical Association.

Rio, E. (2000) *Théorie Asymptotique des Processus Aléatoires Faiblement Dépendants*, Math. Appl. 31. Berlin: Springer-Verlag.

Shorack, G.R. and Wellner, J.A. (1986) *Empirical Processes with Applications to Statistics*. New York: Wiley.

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

Veretennikov, A.Yu. (1988) Bounds for the mixing rate in the theory of stochastic equations. *Theory Probab. Appl.*, **32**(2), 273–281.

Veretennikov, A.Yu. (1999) On Castellana–Leadbetter's condition for diffusion density estimation. *Statist. Inference Stochastic Process.*, **2**(1), 1–9.

Received January 2004 and revised December 2004