The Annals of Probability
2013, Vol. 41, No. 5, 3658–3696
DOI: 10.1214/12-AOP798

STRONG APPROXIMATION RESULTS FOR THE EMPIRICAL PROCESS OF STATIONARY SEQUENCES

BY JÉRÔME DEDECKER, FLORENCE MERLEVÈDE AND EMMANUEL RIO

Université Paris Descartes, Université Paris-Est and Université de Versailles

We prove a strong approximation result for the empirical process associated to a stationary sequence of real-valued random variables, under dependence conditions involving only indicators of half lines. This strong approximation result also holds for the empirical process associated to iterates of expanding maps with a neutral fixed point at zero, as soon as the correlations decrease more rapidly than $n^{-1-\delta}$ for some positive δ. This shows that our conditions are in some sense optimal.

1. Introduction. Let $(X_i)_{i\in\mathbb{Z}}$ be a strictly stationary sequence of real-valued random variables with common distribution function F, and define the empirical process of $(X_i)_{i\in\mathbb{Z}}$ by

$$R_X(s,t)=\sum_{1\le k\le t}\bigl(\mathbf{1}_{X_k\le s}-F(s)\bigr),\qquad s\in\mathbb{R},t\in\mathbb{R}^+. \tag{1.1}$$

For independent identically distributed (i.i.d.) random variables X_i with the uniform distribution over $[0,1]$, Komlós, Major and Tusnády (1975) constructed a continuous centered Gaussian process K_X with covariance function

$$\mathbb{E}\bigl(K_X(s,t)K_X(s',t')\bigr)=\bigl(t\wedge t'\bigr)\bigl(s\wedge s'-ss'\bigr)$$

in such a way that

$$\sup_{s\in\mathbb{R},t\in[0,1]}\bigl|R_X\bigl(s,[nt]\bigr)-K_X\bigl(s,[nt]\bigr)\bigr|=O\bigl(\log^2 n\bigr)\qquad\text{almost surely} \tag{1.2}$$

[we refer also to Castelle and Laurent-Bonvalot (1998) for a detailed proof]. The rate of convergence given in (1.2) improves on the one obtained earlier by Kiefer (1972) and the two-parameter Gaussian process K_X is known in the literature as the Kiefer process.

Such a strong approximation allows not only to derive weak limit theorems, as Donsker's invariance principle for the empirical distribution function, but also almost sure results, as the functional form of the law of the iterated logarithm [see Finkelstein (1971)]. Moreover, from a statistical point of view, strong approximations with rates allow to construct many statistical procedures [we refer to the

Received October 2011.
MSC2010 subject classifications. 60F17, 60G10, 37E05.
Key words and phrases. Strong approximation, Kiefer process, stationary sequences, intermittent maps, weak dependence.

monograph of Shorack and Wellner (1986) which shows how the asymptotic behavior of the empirical process plays a crucial role in many important statistical applications].

In the dependent setting, the weak limiting behavior of the empirical process R_X has been studied by many authors in different cases. See, among many others, the following: Dehling and Taqqu (1989) for stationary Gaussian sequences, Giraitis and Surgailis (2002) for linear processes, Yu (1993) for associated sequences, Borovkova, Burton and Dehling (2001) for functions of absolutely regular sequences, Rio (2000) for strongly mixing sequences, Wu (2008) for functions of i.i.d. sequences and Dedecker (2010) for β-dependent sequences.

Strong approximations of type (1.2), for the empirical process with dependent data, have been less studied. Berkes and Philipp (1977) proved that, for functions of strongly mixing sequences satisfying $\alpha(n) = O(n^{-8})$ [where $\alpha(n)$ is the strong mixing coefficient of Rosenblatt (1956)], and if F is continuous, there exists a two-parameter continuous Gaussian process K_X such that

$$\begin{aligned}\sup_{s\in\mathbb{R},t\in[0,1]} &\bigl|R_X(s,[nt]) - K_X(s,[nt])\bigr| \\ &= O\bigl(\sqrt{n}(\ln(n))^{-\lambda}\bigr) \qquad \text{almost surely}\end{aligned} \tag{1.3}$$

for some $\lambda > 0$. The covariance function Γ_X of K_X is given by

$$\Gamma_X(s,s',t,t') = \min(t,t')\Lambda_X(s,s'),$$

where

$$\Lambda_X(s,s') = \sum_{k\geq 0}\mathrm{Cov}(\mathbf{1}_{X_0\leq s}, \mathbf{1}_{X_k\leq s'}) + \sum_{k>0}\mathrm{Cov}(\mathbf{1}_{X_0\leq s'}, \mathbf{1}_{X_k\leq s}). \tag{1.4}$$

As a corollary, Berkes and Philipp (1977) obtained that the sequence

$$\{(2n\ln\ln n)^{-1/2}R_X(s,[nt]), n\geq 3\}$$

of random functions on $\mathbb{R}\times[0,1]$ is with probability one relatively compact for the supremum norm, and that the set of limit points is the unit ball of the reproducing kernel Hilbert space (RKHS) associated with Γ_X. Their result generalizes the functional form of the Finkelstein's law of the iterated logarithm. Next, Yoshihara (1979) weakened the strong mixing condition required in Berkes and Philipp (1977) and proved the strong approximation (1.3) assuming $\alpha(n) = O(n^{-a})$ for some $a > 3$. However, this condition still appears to be too restrictive: indeed, Rio [(2000), Theorem 7.2, page 96] proved that the weak convergence of $n^{-1/2}R_X(s,n)$ to a Gaussian process holds in $D(\mathbb{R})$ under the weaker condition $\alpha(n) = O(n^{-a})$ for some $a > 1$. In view of this result, one may think that the strong approximation by a Kiefer process, as given in (1.3), holds as soon as the dependence coefficients are of the order of $O(n^{-a})$ for some $a > 1$.

Since the classical mixing coefficients have some limited applicability, many papers have been written in the last decade to derive limit theorems under various weak dependence measures [see, e.g., the monograph by Dedecker et al. (2007)]. Concerning the empirical process, Dedecker (2010) proved that the weak convergence of $n^{-1/2}R_X(s,n)$ to a Gaussian process holds in $D(\mathbb{R})$ under a dependence condition involving only indicators of a half line, whereas Wu (2008) obtained the same result under conditions on, what he called, the predictive dependent measures. These predictive dependence measures allow coupling by independent sequences and are well adapted to some functions of i.i.d. sequences. However, they seem to be less adequate for functionals of nonirreducible Markov chains or dynamical systems having some invariant probability. The recent paper by Berkes, Hörmann and Schauer (2009) deals with strong approximations as in (1.3) in the weak dependent setting by considering, what they called, S-mixing conditions. Actually, their S-mixing condition lies much closer to the predictive dependent measures considered by Wu (2008) and is also very well adapted to functions of i.i.d. sequences. Roughly speaking, they obtained (1.3) as soon as F is Lipschitz continuous, the sequence $(X_i)_{i\in\mathbb{Z}}$ can be approximated by a $2m$-dependent sequence, and one has a nice control of the deviation probability of the approximating error.

In this paper, we prove that the strong approximation (1.3) holds under a dependence condition involving only indicators of a half line, which is quite natural in this context [see the discussion at the beginning of Section 2 in Dedecker (2010)]. More precisely, if $\beta_{2,X}(n) = O(n^{-(1+\delta)})$ for some positive δ, where the coefficients $\beta_{2,X}(n)$ are defined in the next section, we prove that there exists a continuous (with respect to its natural metric) centered Gaussian process K_X with covariance function given by (1.4) such that

$$\sup_{s\in\mathbb{R},t\in[0,1]} \bigl|R_X(s,[nt]) - K_X(s,[nt])\bigr| = O\bigl(n^{1/2-\varepsilon}\bigr) \qquad \text{almost surely} \tag{1.5}$$

for some $\varepsilon > 0$. As consequences of (1.5), we obtain the functional form of Finkelstein's law of the iterated logarithm and we recover the empirical central limit theorem obtained in Dedecker (2010). Notice that our dependence condition cannot be directly compared to the one used in the paper by Berkes, Hörmann and Schauer (2009).

In Theorem 3.1 we show that (1.5) also holds for the empirical process associated to an expanding map T of the unit interval with a neutral fixed point at 0, as soon as the parameter γ belongs to $]0,1/2[$ (this parameter describes the behavior of T in the neighborhood of zero). Moreover, we shall prove that the functional law of the iterated logarithm cannot hold at the boundary $\gamma = 1/2$, which shows that our conditions are in some sense optimal (see Remark 3.2 for a detailed discussion about the optimality of the conditions).

Let us now give an outline of the methods used to prove the strong approximation (1.5). We consider the dyadic fluctuations $(R_X(s,2^{L+1}) - R_X(s,2^L))_{L\geq 0}$ of the empirical process on a grid with a number of points depending on L, let's say d_L. Our proof is mainly based on the existence of multidimensional Gaussian random variables in $\mathbb{R}^{d_L}$ that approximate, in a certain sense, the fluctuations of the empirical process on the grid. These multidimensional Gaussian random variables will be the skeleton of the approximating Kiefer process. To prove the existence of these Gaussian random variables, we apply a conditional version of the Kantorovich–Rubinstein theorem, as given in Rüschendorf (1985) (see our Section 4.1.1). The multidimensional Gaussian random variables are constructed in such a way that the error of approximation in $\mathbb{L}^1$ of the supremum norm between the fluctuations of the empirical process on the grid and the multidimensional Gaussian r.v.'s is exactly the expectation of the Wasserstein distance of order 1 (with the distance associated to the supremum norm) between the conditional law of the fluctuations of the empirical process on the grid and the corresponding multidimensional Gaussian law [see Definition 4.1 and equality (4.5)]. This error can be evaluated with the help of the Lindeberg method as done in Section 4.1.3 [a similar approach has been used recently by Merlevède and Rio (2012) for the partial sum process]. The oscillations of the empirical process, namely, the quantities involved in (4.21) and (4.22), are handled with the help of a suitable exponential inequality combined with the Rosenthal-type inequality proved by Dedecker (2010), Proposition 3.1. Moreover, it is possible to adapt the method of constructing the skeleton Kiefer process (by conditioning up to the future rather than to the past) to deal with the empirical process associated to intermittent maps.

The paper is organized as follows: in Section 2 (resp., Section 3) we state the strong approximation results for the empirical process associated to a class of stationary sequences (resp., to a class of intermittent maps). Section 4 is devoted to the proof of the main results, whereas some technical tools are stated and proved in the Appendix.

2. Strong approximation for the empirical process associated to a class of stationary sequences. Let $(X_i)_{i\in\mathbb{Z}}$ be a strictly stationary sequence of real-valued random variables defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Assume that $(\Omega, \mathcal{A}, \mathbb{P})$ is large enough to contain a sequence $(U_i)_{i\in\mathbb{Z}} = (\delta_i, \eta_i)_{i\in\mathbb{Z}}$ of i.i.d. random variables with uniform distribution over $[0, 1]^2$, independent of $(X_i)_{i\in\mathbb{Z}}$. Define the nondecreasing filtration $(\mathcal{F}_i)_{i\in\mathbb{Z}}$ by $\mathcal{F}_i = \sigma(X_k : k \le i)$. Let $\mathcal{F}_{-\infty} = \bigcap_{i\in\mathbb{Z}} \mathcal{F}_i$ and $\mathcal{F}_\infty = \bigvee_{i\in\mathbb{Z}} \mathcal{F}_i$. We shall denote by $\mathbb{E}_i$ the conditional expectation with respect to $\mathcal{F}_i$.

Let us now define the dependence coefficients that we consider in this paper.

DEFINITION 2.1. Let P be the law of X_0 and $P_{(X_i,X_j)}$ be the law of (X_i, X_j). Let $P_{X_k|X_0}$ be the conditional distribution of X_k given X_0, $P_{X_k|\mathcal{F}_\ell}$ be the conditional distribution of X_k given $\mathcal{F}_\ell$, and $P_{(X_i,X_j)|\mathcal{F}_\ell}$ be the conditional distribution of (X_i, X_j) given $\mathcal{F}_\ell$. Define the functions $f_t = \mathbf{1}_{]-\infty,t]}$, and $f_t^{(0)} = f_t - P(f_t)$.

Define the random variables

$$b(X_0,k)=\sup_{t\in\mathbb{R}}\bigl|P_{X_k|X_0}(f_t)-P(f_t)\bigr|,$$

$$b_1(\mathcal{F}_\ell,k)=\sup_{t\in\mathbb{R}}\bigl|P_{X_k|\mathcal{F}_\ell}(f_t)-P(f_t)\bigr|,$$

$$b_2(\mathcal{F}_\ell,i,j)=\sup_{(s,t)\in\mathbb{R}^2}\bigl|P_{(X_i,X_j)|\mathcal{F}_\ell}\bigl(f_t^{(0)}\otimes f_s^{(0)}\bigr)-P_{(X_i,X_j)}\bigl(f_t^{(0)}\otimes f_s^{(0)}\bigr)\bigr|.$$

Define now the coefficients

$$\beta\bigl(\sigma(X_0),X_k\bigr)=\mathbb{E}\bigl(b(X_0,k)\bigr),\qquad \beta_{1,X}(k)=\mathbb{E}\bigl(b_1(\mathcal{F}_0,k)\bigr)$$

and

$$\beta_{2,X}(k)=\max\Bigl\{\beta_1(k),\sup_{i>j\geq k}\mathbb{E}\bigl(\bigl(b_2(\mathcal{F}_0,i,j)\bigr)\bigr)\Bigr\}.$$

Define also

$$\alpha_{1,X}(k)=\sup_{t\in\mathbb{R}}\bigl\|P_{X_k|\mathcal{F}_0}(f_t)-P(f_t)\bigr\|_1$$

and note that $\alpha_{1,X}(k)\leq\beta_{1,X}(k)\leq\beta_{2,X}(k)$.

Examples of nonmixing sequences $(X_i)_{i\in\mathbb{Z}}$ in the sense of Rosenblatt (1956) for which the coefficients $\beta_{2,X}(n)$ can be computed may be found in the paper by Dedecker and Prieur (2007). Let us give a first elementary example. Let $X_i=\sum_{k\geq 0}a_k\varepsilon_{i-k}$, where $(\varepsilon_i)_{i\in\mathbb{Z}}$ is a sequence of i.i.d. random variables such that $\mathbb{E}(|\varepsilon_0|^\alpha)<\infty$ for some $\alpha>0$, and $a_i=O(\rho^i)$ for some $\rho\in\,]0,1[$. Let w be the modulus of continuity of F. If

$$w(x)\leq C\bigl|\ln(x)\bigr|^{-a}\qquad\text{in a neighborhood of 0, for some } a>1,$$

then $\beta_{2,X}(n)=O(n^{-a})$ [see Remark 2.3 in Dedecker (2010)]. We shall present another example in the next section.

Our main result is the following:

THEOREM 2.1. *Assume that $\beta_{2,X}(n)=O(n^{-1-\delta})$ for some $\delta>0$. Then*:

(1) *For all (s,s') in $\mathbb{R}^2$, the series $\Lambda_X(s,s')$ defined by* (1.4) *converges absolutely.*

(2) *For any $(s,s')\in\mathbb{R}^2$ and (t,t') in $\mathbb{R}^+\times\mathbb{R}^+$, let $\Gamma_X(s,s',t,t')=\min(t,t')\times\Lambda_X(s,s')$. There exists a centered Gaussian process K_X with covariance function Γ_X, whose sample paths are almost surely uniformly continuous with respect to the pseudometric*

$$d\bigl((s,t),(s',t')\bigr)=\bigl|F(s)-F(s')\bigr|+\bigl|t-t'\bigr|$$

and such that (1.5) *holds with $\varepsilon=\delta^2/(22(\delta+2)^2)$.*

Note that we do not make any assumption on the continuity of the distribution function F.

As in the paper of Berkes, Hörmann and Schauer (2009), we can formulate corollaries to Theorem 2.1. The first one is direct. Let $D(\mathbb{R} \times [0, 1])$ be the Skorohod space equipped with the Skorohod topology, as described in Bickel and Wichura (1971).

COROLLARY 2.1. *Assume that $\beta_{2,X}(n) = O(n^{-1-\delta})$ for some $\delta > 0$. Then the empirical process $\{n^{-1/2} R_X(s, [nt]), s \in \mathbb{R}, t \in [0, 1]\}$ converges in $D(\mathbb{R} \times [0, 1])$ to the Gaussian process K_X defined in item* (2) *of Theorem* 2.1.

To obtain the second one, we need to combine the strong approximation (1.5) with Theorem 2 in Lai (1974).

COROLLARY 2.2. *Assume that $\beta_{2,X}(n) = O(n^{-1-\delta})$ for some $\delta > 0$. Then, with probability one, the sequence $\{(2n \ln\ln n)^{-1/2} R_X(s, [nt]), n \geq 3\}$ of random functions on $\mathbb{R} \times [0, 1]$ is relatively compact for the supremum norm, and the set of limit points is the unit ball of the reproducing kernel Hilbert space* (*RKHS*) *associated with the covariance function Γ_X defined in Theorem* 2.1.

3. Strong approximation for the empirical process associated to a class of intermittent maps. In this section we consider the following class of intermittent maps, introduced in Dedecker, Gouëzel and Merlevède (2010):

DEFINITION 3.1. A map $T : [0, 1] \to [0, 1]$ is a generalized Pomeau–Manneville map (or GPM map) of parameter $\gamma \in]0, 1[$ if there exist $0 = y_0 < y_1 < \cdots < y_d = 1$ such that, writing $I_k =]y_k, y_{k+1}[$,

(1) The restriction of T to I_k admits a C^1 extension $T_{(k)}$ to $\overline{I}_k$.

(2) For $k \geq 1$, $T_{(k)}$ is C^2 on $\overline{I}_k$, and $\inf_{x \in \overline{I}_k} |T'_{(k)}(x)| > 1$.

(3) $T_{(0)}$ is C^2 on $]0, y_1]$, with $T'_{(0)}(x) > 1$ for $x \in (0, y_1]$, $T'_{(0)}(0) = 1$ and $T''_{(0)}(x) \sim c x^{\gamma - 1}$ when $x \to 0$, for some $c > 0$.

(4) T is topologically transitive, that is, there exists some x in $]0, 1[$ such that $\{T^n(x) : n \in \mathbb{N}\}$ is a dense subset of $]0, 1[$.

The third condition ensures that 0 is a neutral fixed point of T, with $T(x) = x + c' x^{1+\gamma}(1 + o(1))$ when $x \to 0$. The fourth condition is necessary to avoid situations where there are several absolutely continuous invariant measures or where the neutral fixed point does not belong to the support of the absolutely continuous invariant measure. As a well-known example of a GPM map, let us cite the Liverani, Saussol and Vaienti (1999) map (LSV map) defined by

$$T(x) = \begin{cases} x(1 + 2^\gamma x^\gamma), & \text{if } x \in [0, 1/2], \\ 2x - 1, & \text{if } x \in (1/2, 1]. \end{cases}$$

Theorem 1 in Zweimüller (1998) shows that a GPM map T admits a unique absolutely continuous invariant probability measure ν, with density h_ν. Moreover, it is ergodic, has full support, and $h_\nu(x)/x^{-\gamma}$ is bounded from above and below.

Let Q be the Perron–Frobenius operator of T with respect to ν, defined by

$$\nu(f \cdot g \circ T) = \nu\big(Q(f)g\big) \tag{3.1}$$

for any bounded measurable functions f and g. Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary Markov chain with invariant measure ν and transition Kernel Q. Dedecker and Prieur [(2009), Theorem 3.1] have proved that

$$\beta_{2,X}(n) = O\big(n^{-a}\big) \qquad \text{for any } a < (1-\gamma)/\gamma \tag{3.2}$$

[this upper bound was stated for the Liverani–Saussol–Vaienti map only, but is also valid in our context: see the last paragraph of the introduction in Dedecker and Prieur (2009)]. As a consequence, if $\gamma < 1/2$, the stationary sequence $(X_i)_{i\in\mathbb{Z}}$ satisfies all the assumptions of Theorem 2.1.

Now $(T, T^2, \dots, T^n)$ is distributed as $(X_n, X_{n-1}, \dots, X_1)$ on $([0,1], \nu)$ [see, e.g., Lemma XI.3 in Hennion and Hervé (2001)]. Hence, any information on the law of the sums $\sum_{i=1}^n (f \circ T^i - \nu(f))$ can be obtained by studying the law of $\sum_{i=1}^n (f(X_i) - \nu(f))$. However, the reverse time property cannot be used directly to transfer the almost sure results for $\sum_{i=1}^n (f(X_i) - \nu(f))$ to the sum $\sum_{i=1}^n (f \circ T^i - \nu(f))$.

For any $s \in [0,1]$ and $t \in \mathbb{R}$, let us consider the empirical process associated to the dynamical system T:

$$R_T(s,t) = \sum_{1\le i\le t} \big(\mathbf{1}_{T^i \le s} - F_\nu(s)\big) \qquad \text{where } F_\nu(s) = \nu\big([0,s]\big). \tag{3.3}$$

For any ν-integrable function g, let $g^{(0)} = g - \nu(g)$ and recall that $f_s = \mathbf{1}_{]-\infty,s]}$. Our main result is the following:

THEOREM 3.1. *Let T be a GPM map with parameter $\gamma \in]0,1/2[$. Then*:

(1) *For all $(s,s') \in [0,1]^2$, the following series converges absolutely*:

$$\Lambda_T(s,s') = \sum_{k\ge 0} \nu\big(f_s^{(0)} \cdot f_{s'}^{(0)} \circ T^k\big) + \sum_{k>0} \nu\big(f_{s'}^{(0)} \cdot f_s^{(0)} \circ T^k\big). \tag{3.4}$$

(2) *For any $(s,s') \in [0,1]^2$ and any $(t,t') \in \mathbb{R}^+ \times \mathbb{R}^+$, let $\Gamma_T(s,s',t,t') = \min(t,t')\Lambda_T(s,s')$. There exists a continuous centered Gaussian process K_T^* with covariance function Γ_T such that for some $\varepsilon > 0$*,

$$\sup_{(s,t)\in[0,1]^2} \big|R_T(s,[nt]) - K_T^*(s,[nt])\big| = O\big(n^{1/2-\varepsilon}\big) \qquad \textit{almost surely}.$$

REMARK 3.1. According to the proof of Theorem 3.1, item (2) holds for any ε in $]0, (1-2\gamma)^2/22[$.

REMARK 3.2. In the case $\gamma = 1/2$, Dedecker [(2010), Proposition 4.1] proved that, for the LSV map with $\gamma = 1/2$, the finite-dimensional marginals of the process $\{(n \ln n)^{-1/2} R_T(\cdot, n)\}$ converge in distribution to those of the degenerated Gaussian process G defined by

$$\text{for any } t \in [0, 1] \qquad G(t) = \sqrt{h_\nu(1/2)}\big(1 - F_\nu(t)\big)\mathbf{1}_{t \neq 0} Z,$$

where Z is a standard normal. This shows that an approximation by a Kiefer process as in Theorem 3.1 cannot hold at the boundary $\gamma = 1/2$.

For the same reason, when $\gamma = 1/2$, the conclusion of Theorem 2.1 does not apply to the stationary Markov chain $(X_i)_{i \in \mathbb{Z}}$ with invariant measure ν and transition kernel Q given in (3.1). In fact, it follows from Theorem 3.1 in Dedecker and Prieur (2009) that $\beta_{2,X}(k) > C/k$ for some positive constant C, so that the Markov chain $(X_i)_{i \in \mathbb{Z}}$ does not satisfy the assumptions of Theorem 2.1.

In the case $\gamma = 1/2$, with the same proof as that of Theorem 1.7 of Dedecker, Gouëzel and Merlevède (2010), we see that, for any $(s, t) \in [0, 1]^2$ and $b > 1/2$,

$$\lim_{n \to \infty} \frac{1}{\sqrt{n}(\ln n)^b} R_T\big(s, [nt]\big) = 0 \qquad \text{almost everywhere.}$$

This almost sure result is of the same flavor as in the corresponding i.i.d. case, when the random variables have exactly a weak moment of order 2, so that the normalization in the central limit theorem is $(n \ln n)^{-1/2}$: see the discussion in Dedecker, Gouëzel and Merlevède (2010), last paragraph of Section 1.2.

4. Proofs. In this section we shall sometimes use the notation $a_n \ll b_n$ to mean that there exists a numerical constant C not depending on n such that $a_n \leq C b_n$, for all positive integers n.

4.1. *Proof of Theorem* 2.1. Notice first that for any $(s, s') \in \mathbb{R}^2$,

$$\big|\mathrm{Cov}(\mathbf{1}_{X_0 \leq s}, \mathbf{1}_{X_k \leq s'})\big| \leq \big\|\mathbb{E}_0\big(\mathbf{1}_{X_k \leq s'} - F(s')\big)\mathbf{1}_{X_0 \leq s}\big\|_1 \leq \mathbb{E}\big(b(X_0, k)\big) \leq \beta_{1,X}(k).$$

Since $\sum_{k \geq 0} \beta_{1,X}(k) < \infty$, item (1) of Theorem 2.1 follows.

To prove item (2), we first introduce another probability on Ω. Let $\mathbb{P}_0^*$ be the probability on Ω whose density with respect to $\mathbb{P}$ is

$$C(\beta)^{-1}\left(1 + 4\sum_{k=1}^{\infty} b(X_0, k)\right) \qquad \text{with } C(\beta) = 1 + 4\sum_{k=1}^{\infty} \beta\big(\sigma(X_0), X_k\big). \tag{4.1}$$

Recall that P is the distribution of X_0. Then the image measure P^* of $\mathbb{P}_0^*$ by X_0 is absolutely continuous with respect to P with density

$$C(\beta)^{-1}\left(1 + 4\sum_{k=1}^{\infty} b(x, k)\right). \tag{4.2}$$

Let F_{P^*} be the distribution function of P^*, and let $F_{P^*}(x-0) = \sup_{z<x} F_{P^*}(z)$. Recall that the sequence $(\eta_i)_{i\in\mathbb{Z}}$ of i.i.d. random variables with uniform distribution over [0, 1] has been introduced at the beginning of Section 2. Define then the random variables

$$Y_i = F_{P^*}(X_i - 0) + \eta_i\big(F_{P^*}(X_i) - F_{P^*}(X_i - 0)\big). \tag{4.3}$$

Let P_Y be the distribution of Y_0 and F_Y be the distribution function of Y_0. Some properties of the sequence $(Y_i)_{i\in\mathbb{Z}}$ are given in Lemma A.1 of the Appendix. In particular, it follows from Lemma A.1 that $X_i = F_{P^*}^{-1}(Y_i)$ almost surely, where $F_{P^*}^{-1}$ is the generalized inverse of the cadlag function F_{P^*}. Hence, $R_X(\cdot,\cdot) = R_Y(F_{P^*}(\cdot),\cdot)$ almost surely, where

$$R_Y(s,t) = \sum_{1\le k\le t} \big(\mathbf{1}_{Y_k\le s} - F_Y(s)\big), \qquad s\in[0,1], t\in\mathbb{R}^+.$$

We now prove that, if $\beta_{2,X}(n) = O(n^{-1-\delta})$ for some $\delta > 0$, then the conclusion of Theorem 2.1 holds for the stationary sequence $(Y_i)_{i\in\mathbb{Z}}$ and the associated continuous Gaussian process K_Y with covariance function $\Gamma_Y(s,s',t,t') = \min(t,t')\Lambda_Y(s,s')$, where

$$\Lambda_Y(s,s') = \sum_{k\ge 0} \mathrm{Cov}(\mathbf{1}_{Y_0\le s}, \mathbf{1}_{Y_k\le s'}) + \sum_{k>0} \mathrm{Cov}(\mathbf{1}_{Y_0\le s'}, \mathbf{1}_{Y_k\le s}). \tag{4.4}$$

This implies Theorem 2.1, since $\Gamma_X(s,s',t,t') = \Gamma_Y(F_{P^*}(s), F_{P^*}(s'), t, t')$.

The proof is divided in two steps: the construction of the Kiefer process with the help of a conditional version of the Kantorovich–Rubinstein theorem and a probabilistic upper bound for the error of approximation.

4.1.1. *Construction of the Kiefer process.* For $L\in\mathbb{N}$, let $m(L)\in\mathbb{N}$ and $r(L)\in\mathbb{N}^*$ be such that $m(L)\le L$ and $4r(L)\le m(L)$. For j in $\{1,\dots,2^{r(L)}-1\}$, let $s_j = j2^{-r(L)}$ and define for any $\ell\in\{1,\dots,2^{L-m(L)}\}$,

$$I_{L,\ell} = \left]2^L + (\ell-1)2^{m(L)}, 2^L + \ell 2^{m(L)}\right] \cap \mathbb{N}$$

and

$$U_{L,\ell}^{(j)} = \sum_{i\in I_{L,\ell}} \big(\mathbf{1}_{Y_i\le s_j} - F_Y(s_j)\big).$$

The associated column vectors $U_{L,\ell}$ are then defined in $\mathbb{R}^{2^{r(L)}-1}$ by

$$U_{L,\ell} = \big(U_{L,\ell}^{(1)},\dots,U_{L,\ell}^{(2^{r(L)}-1)}\big)'.$$

Let us now introduce some definitions.

DEFINITION 4.1. Let m be a positive integer. Let P_1 and P_2 be two probabilities on $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$. Let d be a distance on $\mathbb{R}^m$ associated to a norm. The Wasserstein distance of order 1 between P_1 and P_2 with respect to the distance d is defined by

$$\begin{aligned} W_d(P_1, P_2) &= \inf\{\mathbb{E}(d(X, Y)), (X, Y) \text{ such that } X \sim P_1, Y \sim P_2\} \\ &= \sup_{f \in \mathrm{Lip}(d)} (P_1(f) - P_2(f)), \end{aligned}$$

where $\mathrm{Lip}(d)$ is the set of functions from $\mathbb{R}^m$ into $\mathbb{R}$ that are 1-Lipschitz with respect to d; namely, for any x and y of $\mathbb{R}^m$, $|f(x) - f(y)| \le d(x, y)$.

DEFINITION 4.2. Let r be a positive integer. For any points $x = (x^{(1)}, \dots, x^{(2^r-1)})'$ and $y = (y^{(1)}, \dots, y^{(2^r-1)})'$, we set

$$d_r(x, y) = \sup_{j \in \{1,\dots,2^r-1\}} |x^{(j)} - y^{(j)}|.$$

Let $L \in \mathbb{N}$ and $\ell \in \{1, \dots, 2^{L-m(L)}\}$. Let

$$\Lambda_{Y,L} = (\Lambda_Y(s_j, s_{j'}))_{j,j'=1,\dots,2^{r(L)}-1},$$

where the $\Lambda_Y(s_j, s_{j'})$ are defined in (4.4). Let $G_{2^{m(L)}\Lambda_{Y,L}}$ denote the $\mathcal{N}(0, 2^{m(L)}\Lambda_{Y,L})$-law and $P_{U_{L,\ell}|\mathcal{F}_{2^L+(\ell-1)2^{m(L)}}}$ be the conditional distribution of $U_{L,\ell}$ given $\mathcal{F}_{2^L+(\ell-1)2^{m(L)}}$.

According to Rüschendorf (1985) [see also Theorem 2 in Dedecker, Prieur and Raynaud De Fitte (2006)], there exists a random variable $V_{L,\ell} = (V_{L,\ell}^{(1)}, \dots, V_{L,\ell}^{(2^{r(L)}-1)})'$ with law $G_{2^{m(L)}\Lambda_{Y,L}}$, measurable with respect to $\sigma(\delta_{2^L+\ell 2^{m(L)}}) \vee \sigma(U_{L,\ell}) \vee \mathcal{F}_{2^L+(\ell-1)2^{m(L)}}$, independent of $\mathcal{F}_{2^L+(\ell-1)2^{m(L)}}$ and such that

$$\begin{aligned} &\mathbb{E}(d_{r(L)}(U_{L,\ell}, V_{L,\ell})) \\ &\quad = \mathbb{E}(W_{d_{r(L)}}(P_{U_{L,\ell}|\mathcal{F}_{2^L+(\ell-1)2^{m(L)}}}, G_{2^{m(L)}\Lambda_L})) \\ &\quad = \mathbb{E} \sup_{f \in \mathrm{Lip}(d_{r(L)})} (\mathbb{E}(f(U_{L,\ell})|\mathcal{F}_{2^L+(\ell-1)2^{m(L)}}) - \mathbb{E}(f(V_{L,\ell}))). \end{aligned} \tag{4.5}$$

By induction on ℓ, the random variables $(V_{L,\ell})_{\ell=1,\dots,2^{L-m(L)}}$ are mutually independent, independent of $\mathcal{F}_{2^L}$ and with law $\mathcal{N}(0, 2^{m(L)}\Lambda_{Y,L})$. Hence, we have constructed Gaussian random variables $(V_{L,\ell})_{L\in\mathbb{N},\ell=1,\dots,2^{L-m(L)}}$ that are mutually independent. In addition, according to Lemma 2.11 of Dudley and Philipp (1983), there exists a Kiefer process K_Y with covariance function Γ_Y such that for any $L \in \mathbb{N}$, any $\ell \in \{1, \dots, 2^{L-m(L)}\}$ and any $j \in \{1, \dots, 2^{r(L)-1}\}$,

$$V_{L,\ell}^{(j)} = K_Y(s_j, 2^L + \ell 2^{m(L)}) - K_Y(s_j, 2^L + (\ell-1)2^{m(L)}). \tag{4.6}$$

Our construction is now complete.

In Proposition 4.1 proved in Section 4.1.3, we shall give some upper bounds for the quantities $\mathbb{E}(d_{r(L)}(U_{L,\ell}, V_{L,\ell}))$ for $L \in \mathbb{N}$ and $\ell \in \{1, \dots, 2^{L-m(L)}\}$, showing that under our condition on the dependence coefficients there exists a positive constant C such that

$$\mathbb{E}\big(d_{r(L)}(U_{L,\ell}, V_{L,\ell})\big) \le C 2^{(m(L)+2r(L))/((2+\delta)\wedge 3)} L^2. \tag{4.7}$$

In Section 4.1.2 below, starting from (4.7), we bound up the error of approximation between the empirical process and the Kiefer process.

4.1.2. *Upper bound for the approximation error.* Let $\{K_Y(s,t), s \in [0,1], t \ge 0\}$ be the Gaussian process constructed as in step 1 with the following choice of $r(L)$ and $m(L)$. For $\varepsilon < 1/10$, let

$$r(L) = \big([L/5] \wedge \big[2\varepsilon L + 5\log_2(L)\big]\big) \vee 1 \quad \text{and} \quad m(L) = L - r(L), \tag{4.8}$$

so that, for L large enough,

$$\begin{gathered} 2^{2\varepsilon L-1} L^5 \le 2^{r(L)} \le 2^{2\varepsilon L} L^5 \quad \text{and} \\ 2^{L(1-2\varepsilon)} L^{-5} \le 2^{m(L)} \le 2^{1+L(1-2\varepsilon)} L^{-5}. \end{gathered} \tag{4.9}$$

Let $N \in \mathbb{N}^*$ and let $k \in]1, 2^{N+1}]$. To shorten the notation, let $K_Y = K$ and $R_Y = R$. We first notice that

$$\sup_{1\le k\le 2^{N+1}} \sup_{s\in[0,1]} \big|R(s,k) - K(s,k)\big| \le \sup_{s\in[0,1]} \big|R(s,1) - K(s,1)\big| + \sum_{L=0}^{N} D_L, \tag{4.10}$$

where

$$D_L := \sup_{2^L<\ell\le 2^{L+1}} \sup_{s\in[0,1]} \big|\big(R(s,\ell) - R\big(s,2^L\big)\big) - \big(K(s,\ell) - K\big(s,2^L\big)\big)\big|. \tag{4.11}$$

Notice first that $\sup_{s\in[0,1]} |R(s,1) - K(s,1)| \le 1 + \sup_{s\in[0,1]} |K(s,1)|$. Dedecker (2010) (see the beginning of the proof of his Theorem 2.1) has proved that, for u and v in $[0,1]$ and any positive integer n,

$$\operatorname{Var}\big(K(u,n) - K(v,n)\big) \le C(\beta) n |u - v|. \tag{4.12}$$

Therefore, according to Theorem 11.17 in Ledoux and Talagrand (1991), $\mathbb{E}(\sup_{s\in[0,1]} |K(s,1)|) = O(1)$. It follows that for any $\varepsilon \in]0, 1/2[$,

$$\sup_{s\in[0,1]} \big|R(s,1) - K(s,1)\big| = O\big(2^{N(1/2-\varepsilon)}\big) \qquad \text{a.s.} \tag{4.13}$$

To prove Theorem 2.1, it then suffices to prove that for any $L \in \{0, \dots, N\}$,

$$D_L = O\big(2^{L(1/2-\varepsilon)}\big) \qquad \text{a.s. for } \varepsilon = \delta^2/(22(\delta+2)^2). \tag{4.14}$$

With this aim, we decompose D_L with the help of several quantities. For any $K \in \mathbb{N}$ and any $s \in [0,1]$, let $\Pi_K(s) = 2^{-K}[2^K s]$. Notice that the following decomposition is valid: for any $L \in \mathbb{N}$,

$$D_L \le D_{L,1} + D_{L,2} + D_{L,3}, \tag{4.15}$$

where

$$\begin{aligned}
D_{L,1} = \sup_{2^L < \ell \le 2^{L+1}} \sup_{s \in [0,1]} \big| \big(R(s,\ell) - R\big(\Pi_{r(L)}(s), \ell\big)\big) \\
- \big(R\big(s, 2^L\big) - R\big(\Pi_{r(L)}(s), 2^L\big)\big)\big|, \\
D_{L,2} = \sup_{2^L < \ell \le 2^{L+1}} \sup_{s \in [0,1]} \big| \big(K(s,\ell) - K\big(\Pi_{r(L)}(s), \ell\big)\big) \\
- \big(K\big(s, 2^L\big) - K\big(\Pi_{r(L)}(s), 2^L\big)\big)\big|, \\
D_{L,3} = \sup_{2^L < \ell \le 2^{L+1}} \sup_{s \in [0,1]} \big| \big(R\big(\Pi_{r(L)}(s), \ell\big) - R\big(\Pi_{r(L)}(s), 2^L\big)\big) \\
- \big(K\big(\Pi_{r(L)}(s), \ell\big) - K\big(\Pi_{r(L)}(s), 2^L\big)\big)\big|.
\end{aligned}$$

In addition,

$$D_{L,3} \le A_{L,3} + B_{L,3} + C_{L,3}, \tag{4.16}$$

where

$$\begin{aligned}
A_{L,3} &= \sup_{j \in \{1,\dots,2^{r(L)}-1\}} \sup_{k \le 2^{L-m(L)}} \left| \sum_{\ell=1}^{k} \big(U_{L,\ell}^{(j)} - V_{L,\ell}^{(j)}\big) \right|, \\
B_{L,3} &= \sup_{j \in \{1,\dots,2^{r(L)}-1\}} \sup_{k \le 2^{L-m(L)}} \sup_{\ell \in I_{L,k}} \big| R(s_j, \ell) - R\big(s_j, 2^L + (k-1)2^{m(L)}\big) \big|, \\
C_{L,3} &= \sup_{j \in \{1,\dots,2^{r(L)}-1\}} \sup_{k \le 2^{L-m(L)}} \sup_{\ell \in I_{L,k}} \big| K(s_j, \ell) - K\big(s_j, 2^L + (k-1)2^{m(L)}\big) \big|
\end{aligned}$$

with $s_j = j2^{-r(L)}$.

Let us first deal with the terms $D_{L,2}$ and $C_{L,3}$ involving only the approximating Kiefer process. For any positive λ,

$$\begin{aligned}
&\mathbb{P}\big(|D_{L,2}| \ge \lambda\big) \\
&\quad \le \sum_{j=1}^{2^{r(L)}} \mathbb{P}\Big(\sup_{2^L < \ell \le 2^{L+1}} \sup_{s_{j-1} \le s \le s_j} \big| \big(K(s,\ell) - K\big(s, 2^L\big)\big) \\
&\qquad\qquad - \big(K(s_j, \ell) - K\big(s_j, 2^L\big)\big)\big| \ge \lambda \Big).
\end{aligned}$$

Setting

$$\begin{aligned}
X(u,v) = {} & \big(K\big(s_j + u(s_{j+1} - s_j), 2^L + v2^L\big) - K\big(s_j + u(s_{j+1} - s_j), 2^L\big)\big) \\
& - \big(K\big(s_j, 2^L + v2^L\big) - K\big(s_j, 2^L\big)\big),
\end{aligned}$$

we have

$$\mathbb{P}(D_{L,2} \geq \lambda) \leq \sum_{j=1}^{2^{r(L)}} \mathbb{P}\Big(\sup_{(u,v)\in[0,1]^2} |X(u,v)| \geq \lambda\Big).$$

Using (4.12), we infer that

$$\mathbb{E}|X(u,v) - X(u',v')|^2 \ll 2^{L-r(L)}(|u-u'| + |v-v'|)$$

and

$$\sup_{(u,v)\in[0,1]^2} \mathbb{E}|X(u,v)|^2 \ll 2^{L-r(L)}.$$

Next, using Lemma 2 in Lai (1974), as done in Lemma 6.2 in Berkes and Philipp (1977), and taking into account (4.9), we infer that there exists a positive constant c such that, for L large enough,

$$\mathbb{P}(|D_{L,2}| \geq c2^{L(1/2-\varepsilon)}) \ll 2^{r(L)} \exp(-L^5/2).$$

Therefore,

$$\sum_{L>0} \mathbb{P}(D_{L,2} \geq c2^{L(1/2-\varepsilon)}) < \infty. \tag{4.17}$$

Consider now the term $C_{L,3}$. For any positive λ,

$$\mathbb{P}(C_{L,3} \geq \lambda) \leq \sum_{k=1}^{2^{L-m(L)}} \mathbb{P}\Big(\sup_{s\in[0,1]} \sup_{\ell\in I_{L,k}} |K(s,\ell) - K(s, 2^L + (k-1)2^{m(L)})| \geq \lambda\Big).$$

Setting $X(s,u) = K(s, 2^L + (k-1)2^{m(L)} + u2^{m(L)}) - K(s, 2^L + (k-1)2^{m(L)} + u2^{m(L)})$ and using (4.12), we have that

$$\mathbb{E}|X(s,u) - X(s',u')|^2 \ll 2^{m(L)}(|s-s'| + |u-u'|)$$

and

$$\sup_{(s,u)\in[0,1]^2} \mathbb{E}|X(s,u)|^2 \ll 2^{m(L)}.$$

Therefore, by using once again Lemma 2 in Lai (1974), as done in Lemma 6.3 in Berkes and Philipp (1977), and taking into account (4.9), we infer that there exists a positive constant c such that, for L large enough,

$$\mathbb{P}\Big(\sup_{s\in[0,1]} \sup_{\ell\in I_{L,k}} |K(s,\ell) - K(s, 2^L + (k-1)2^{m(L)})| \geq c2^{L(1/2-\varepsilon)}\Big) \ll \exp(-L^5/2).$$

Therefore,

$$\sum_{L>0} \mathbb{P}(C_{L,3} \geq c2^{L(1/2-\varepsilon)}) < \infty. \tag{4.18}$$

We now prove that

$$\sum_{L>0} \mathbb{P}\big(A_{L,3} \ge 2^{L(1/2-\varepsilon)}\big) < \infty. \tag{4.19}$$

From the stationarity of the sequence $((U_{L,\ell}, V_{L,\ell}))_{\ell=1,\dots,2^{L-m(L)}}$,

$$\mathbb{P}\big(A_{L,3} \ge 2^{L(1/2-\varepsilon)}\big) \le 2^{L-m(L)} 2^{L(\varepsilon-1/2)} \mathbb{E}\big(d_{r(L)}(U_{L,1}, V_{L,1})\big).$$

Therefore, by using (4.7), we get that

$$\mathbb{P}\big(A_{L,3} \ge 2^{L(1/2-\varepsilon)}\big) \ll 2^{L(\varepsilon-1/2)} 2^{L-m(L)} 2^{m(L)+2r(L)/((2+\delta)\wedge 3)} L^2,$$

which together with (4.9) proves (4.19), provided that

$$\varepsilon < \frac{\delta \wedge 1}{2(8+3(\delta\wedge 1))}. \tag{4.20}$$

We now show that

$$\sum_{L>0} \mathbb{P}\big(B_{L,3} \ge C 2^{L(1/2-\varepsilon)}\big) < \infty. \tag{4.21}$$

By stationarity, for any positive λ,

$$\mathbb{P}(B_{L,3} \ge \lambda) \le 2^{L-m(L)} \sum_{j=1}^{2^{r(L)}} \mathbb{P}\bigg(\sup_{\ell \le 2^{m(L)}} \bigg|\sum_{i=1}^{\ell} \big(\mathbf{1}_{Y_i \le j2^{-r(L)}} - F_Y\big(j2^{-r(L)}\big)\big)\bigg| \ge \lambda\bigg).$$

By Lemma A.1, $|\mathrm{Cov}(\mathbf{1}_{Y_0 \le j2^{-r(L)}}, \mathbf{1}_{Y_i \le j2^{-r(L)}})| \le \mathbb{E}(b(X_0, i)) = \beta(\sigma(X_0), X_i)$ and, consequently,

$$\sum_{i\in\mathbb{Z}} \big|\mathrm{Cov}(\mathbf{1}_{Y_0 \le j2^{-r(L)}}, \mathbf{1}_{Y_i \le j2^{-r(L)}})\big| \le C(\beta).$$

Applying Theorem 1 in Dedecker and Merlevède (2010), we get that for any $v \ge 1$,

$$\begin{aligned}\mathbb{P}\bigg(\sup_{\ell \le 2^{m(L)}} \bigg|\sum_{i=1}^{\ell} \bigg(\mathbf{1}_{Y_i \le j2^{-r(L)}} - F_Y\bigg(\frac{j}{2^{r(L)}}\bigg)\bigg)\bigg| \ge 4\lambda\bigg) \\ \ll \bigg(1 + \frac{\lambda^2}{2^{m(L)} v C(\beta)}\bigg)^{-v/4} + \bigg(\frac{2^{m(L)}}{\lambda} + \frac{\lambda}{v}\bigg)\beta_{2,X}\bigg(\bigg[\frac{\lambda}{v}\bigg]\bigg).\end{aligned}$$

Applying this inequality with $4\lambda = 2^{L(1/2-\varepsilon)}$ and $v = L^5/C(\beta)$ and taking into account (4.9) together with our condition on the dependence coefficients, we derive that for L large enough,

$$\begin{aligned}\mathbb{P}\bigg(\sup_{\ell \le 2^{m(L)}} \bigg|\sum_{i=1}^{\ell} \big(\mathbf{1}_{Y_i \le j2^{-r(L)}} - F_Y\big(j2^{-r(L)}\big)\big)\bigg| \ge 2^{L(1/2-\varepsilon)}\bigg) \\ \ll \exp(-c_1 L^5) + L^{5\delta} 2^{-L(1/2-\varepsilon)\delta}.\end{aligned}$$

Therefore, (4.21) holds provided that $\varepsilon < \delta/(8+2\delta)$, which holds under (4.20).

Taking into account (4.17), (4.18), (4.19) and (4.21) together with the decompositions (4.15) and (4.16), the proof of (4.14) will be complete if we prove that, for some positive constant A to be chosen later,

$$\sum_{L>0} \mathbb{P}\big(D_{L,1} \geq \sqrt{AC(\beta)}2^{L(1/2-\varepsilon)}\big) < \infty. \tag{4.22}$$

To shorten the notation, we set, for $\ell > m \geq 0$,

$$\mu_{\ell,m}(s) = R(s,\ell) - R(s,m) \quad \text{and} \quad Z_{\ell,m} = d\mu_{\ell,m}.$$

We start from the elementary decomposition

$$\begin{aligned}
&\mu_{\ell,2^L}(s) - \mu_{\ell,2^L}\big(\Pi_{r(L)}(s)\big) \\
&\quad = \sum_{K=r(L)+1}^{L} \big(\mu_{\ell,2^L}\big(\Pi_K(s)\big) - \mu_{\ell,2^L}\big(\Pi_{K-1}(s)\big)\big) + \mu_{\ell,2^L}(s) - \mu_{\ell,2^L}\big(\Pi_L(s)\big).
\end{aligned}$$

Consequently,

$$\sup_{s\in[0,1]} \big|\mu_{\ell,2^L}(s) - \mu_{\ell,2^L}\big(\Pi_{r(L)}(s)\big)\big| \leq \sum_{K=r(L)+1}^{L} \Delta_{K,\ell,2^L} + \Delta^*_{L,\ell,2^L}, \tag{4.23}$$

where

$$\Delta_{K,\ell,m} = \sup_{1\leq i\leq 2^K} \big|Z_{\ell,m}\big(\big](i-1)2^{-K}, i2^{-K}\big]\big)\big|$$

and

$$\Delta^*_{L,\ell,m} = \sup_{s\in[0,1]} \big|Z_{\ell,m}\big(\big]\Pi_L(s), s\big]\big)\big|.$$

Note that

$$-(\ell - 2^L)\mathbb{P}\big(\Pi_L(s) < Y_0 \leq \Pi_L(s) + 2^{-L}\big) \leq Z_{\ell,2^L}\big(\big]\Pi_L(s), s\big]\big) \tag{4.24}$$

and

$$\begin{aligned}
Z_{\ell,2^L}\big(\big]\Pi_L(s), s\big]\big) &\leq Z_{\ell,2^L}\big(\big]\Pi_L(s), \Pi_L(s) + 2^{-L}\big]\big) \\
&\quad + (\ell - 2^L)\mathbb{P}\big(\Pi_L(s) < Y_0 \leq \Pi_L(s) + 2^{-L}\big).
\end{aligned} \tag{4.25}$$

Applying Lemma A.1,

$$\begin{aligned}
\mathbb{P}\big(\Pi_L(s) < Y_0 \leq \Pi_L(s) + 2^{-L}\big) &\leq C(\beta)\mathbb{P}^*_0\big(\Pi_L(s) < Y_0 \leq \Pi_L(s) + 2^{-L}\big) \\
&= C(\beta)2^{-L}.
\end{aligned} \tag{4.26}$$

From (4.24), (4.25) and (4.26), we infer that $\Delta^*_{L,\ell,2^L} \le \Delta_{L,\ell,2^L} + C(\beta)$. Hence, it follows from (4.23) that

$$\sup_{s\in[0,1]} \left|\mu_{\ell,2^L}(s) - \mu_{\ell,2^L}\big(\Pi_{r(L)}(s)\big)\right| \le C(\beta) + 2\sum_{K=r(L)+1}^{L} \Delta_{K,\ell,2^L}.$$

Therefore,

$$\begin{aligned}\sup_{2^L<\ell\le 2^{L+1}} \sup_{s\in[0,1]} &\left|\mu_{\ell,2^L}(s) - \mu_{\ell,2^L}\big(\Pi_{r(L)}(s)\big)\right| \\ &\le C(\beta) + 2\sum_{K=r(L)+1}^{L} \sup_{2^L<\ell\le 2^{L+1}} \Delta_{K,\ell,2^L}.\end{aligned}$$

Hence, to prove (4.22), it suffices to show that

$$\sum_{L>0} \mathbb{P}\Bigg(\sum_{K=r(L)+1}^{L} \sup_{2^L<\ell\le 2^{L+1}} \Delta_{K,\ell,2^L} > \sqrt{AC(\beta)}2^{L(1/2-\varepsilon)-2}\Bigg) < \infty. \tag{4.27}$$

Let $c_K = (K(K+1))^{-1}$. Clearly, using the stationarity, (4.27) is true provided that

$$\sum_{L>0}\sum_{K=r(L)+1}^{L} \mathbb{P}\Big(\sup_{0<\ell\le 2^L} \Delta_{K,\ell,0} > \sqrt{AC(\beta)}c_K 2^{L(1/2-\varepsilon)-2}\Big) < \infty. \tag{4.28}$$

We now give two upper bounds for the quantity

$$\mathbb{P}\Big(\sup_{0<\ell\le 2^L} \Delta_{K,\ell,0} > \sqrt{AC(\beta)}c_K 2^{L(1/2-\varepsilon)-2}\Big).$$

Choose $p \in]2,3]$ such that $p < 2(1+\delta)$. Applying Markov's inequality at order p, we have

$$\mathbb{P}\Big(\sup_{0<\ell\le 2^L} \Delta_{K,\ell,0} > \sqrt{AC(\beta)}c_K 2^{L(1/2-\varepsilon)-2}\Big) \ll c_K^{-p} 2^{L(\varepsilon p - p/2)} \Big\|\sup_{0<\ell\le 2^L} \Delta_{K,\ell,0}\Big\|_p^p.$$

Applying inequality (7) of Proposition 1 in Wu (2007) to the stationary sequence $(T^{(j)}_{K,i})_{j\in\mathbb{Z}}$ defined by $T^{(j)}_{K,i} = \mathbf{1}_{(i-1)2^{-K}<Y_j\le i2^{-K}}$, we have

$$\Big\|\sup_{0<\ell\le 2^L} \Delta_{K,\ell,0}\Big\|_p \le 2^{L/p}\sum_{j=0}^{L} 2^{-j/p}\|\Delta_{K,2^j,0}\|_p.$$

Let $0 < \eta < (p-2)/2$. Dedecker (2010) [see the displayed inequality after (2.19) in his paper] proved that

$$\|\Delta_{K,2^j,0}\|_p^p \ll 2^{jp/2}\big(2^{-K(p-2)/2} + 2^{-j\eta(2(1+\delta)-p)/2} + 2^{j\eta - j(p-2)/2}\big).$$

Therefore,

$$\Big\| \sup_{0<\ell\le 2^L} \Delta_{K,\ell,0} \Big\|_p^p \ll 2^{Lp/2}\big(2^{-K(p-2)/2} + 2^{-\eta L(2(1+\delta)-p)/2} + 2^{\eta L - L(p-2)/2}\big). \tag{4.29}$$

On the other hand,

$$\mathbb{P}\Big(\sup_{0<\ell\le 2^L} \Delta_{K,\ell,0} > \sqrt{AC(\beta)}c_K 2^{L(1/2-\varepsilon)-2}\Big)$$
$$\le \sum_{i=1}^{2^K} \mathbb{P}\Big(\sup_{0<\ell\le 2^L} \big|Z_{\ell,0}\big(\big](i-1)2^{-K}, i2^{-K}\big]\big)\big| > \sqrt{AC(\beta)}c_K 2^{L(1/2-\varepsilon)-2}\Big).$$

We now apply Theorem 1 in Dedecker and Merlevède (2010), taking into account the stationarity: for any $x>0$, $v\ge 1$, and $s_L^2 \ge 2^L \sum_{j=0}^{2^L} |\mathrm{Cov}(T_{K,i}^{(0)}, T_{K,i}^{(j)})|$,

$$\mathbb{P}\Big(\sup_{0<\ell\le 2^L} \big|Z_{\ell,0}\big(\big](i-1)2^{-K}, i2^{-K}\big]\big)\big| > 4x\Big)$$
$$\ll \bigg(\Big(1+\frac{x^2}{vs_L^2}\Big)^{-v/4} + 2^L\Big(\frac{1}{x}+\frac{2x}{vs_L^2}\Big)\beta_{2,X}\Big(\Big[\frac{x}{v}\Big]\Big)\bigg).$$

Applying Lemma A.1, we have $|\mathrm{Cov}(T_{K,i}^{(0)}, T_{K,i}^{(j)})| \le 2\mathbb{E}(T_{K,i}^{(0)} b(X_0,j))$. Hence,

$$\sum_{j=0}^{\infty} \big|\mathrm{Cov}\big(T_{K,i}^{(0)}, T_{K,i}^{(j)}\big)\big| \le C(\beta)\mathbb{P}_0^*\big((i-1)2^{-K} < Y_0 \le i2^{-K}\big) = C(\beta)2^{-K}. \tag{4.30}$$

It follows that, for $K \ge r(L)$,

$$\sum_{j=0}^{\infty} \big|\mathrm{Cov}\big(T_{K,i}^{(0)}, T_{K,i}^{(j)}\big)\big| \le C(\beta)2^{-r(L)}.$$

For $L\ge 2$, let $x = x_{K,L} = \sqrt{AC(\beta)}c_K 2^{L(1/2-\varepsilon)-4}$, $s_L^2 = C(\beta)2^{L-r(L)}$ and $v = v_L = 4L$. Taking into account (4.9) and noting that $c_K \ge (L(L+1))^{-1}$ for $K\le L$, we obtain for L large enough and $K \le L$,

$$\Big(1+\frac{x^2}{vs_L^2}\Big)^{-v/4} \le \Big(1+\frac{A2^{L(1-2\varepsilon)}}{2^{10}L^3(L+1)^2 2^{L-r(L)}}\Big)^{-L} \le 3^{-L},$$

the last bound being true provided A is large enough. Hence, for L large enough and $r(L) \le K \le L$,

$$\mathbb{P}\Big(\sup_{0<\ell\le 2^L} \big|Z_{\ell,0}\big(\big](i-1)2^{-K}, i2^{-K}\big]\big)\big| > 4x_{K,L}\Big) \ll \Big(\frac{1}{3^L}+\frac{L^{5+3\delta}2^{L\varepsilon(2+\delta)}}{2^{L\delta/2}}\Big). \tag{4.31}$$

From (4.29) and (4.31), we then get that for L large enough and any $\kappa \le 1$,

$$\begin{aligned}\sum_{K=r(L)+1}^{L} &\mathbb{P}\Big(\sup_{0<\ell\le 2^L} \Delta_{K,\ell,0} > \sqrt{AC(\beta)}c_K 2^{L(1/2-\varepsilon)-2}\Big)\\ &\ll \sum_{K=r(L)+1}^{[\kappa L]} 2^K\Big(\frac{1}{3^L} + \frac{L^{5+3\delta}2^{L\varepsilon(2+\delta)}}{2^{L\delta/2}}\Big)\\ &\quad + 2^{\varepsilon L p}L^{2p}\sum_{K=[\kappa L]+1}^{L}\big(2^{-K(p-2)/2} + 2^{-\eta L(2(1+\delta)-p)/2} + 2^{-L(p-2)/2+\eta L}\big).\end{aligned}$$

Take $\kappa = \kappa(\varepsilon) = 1 \wedge 2\varepsilon(p+1)/(p-2)$. It follows that (4.27) [and then (4.22)] holds provided that the following constraints on ε are satisfied:

$$\varepsilon < \frac{p-2}{2(p+1)}, \qquad \varepsilon\Big(2+\delta+\frac{2(p+1)}{p-2}\Big) < \delta/2, \qquad \varepsilon p < \frac{p-2}{2} - \eta$$

and

$$\varepsilon p < \eta(1+\delta-p/2).$$

Let us take

$$\eta = \frac{p-2}{4+2\delta-p} \quad \text{and} \quad p = 3 \wedge (2+\delta/2).$$

Both the above constraints on ε and (4.20) are satisfied for $\varepsilon = \delta^2/(22(\delta+2)^2)$. Therefore, (4.22) holds, and Theorem 2.1 follows.

4.1.3. *Gaussian approximation.*

PROPOSITION 4.1. *For $L \in \mathbb{N}$, let $m(L) \in \mathbb{N}$ and $r(L) \in \mathbb{N}^*$ be such that $m(L) \le L$ and $4r(L) \le m(L)$. Under the assumptions of Theorem 2.1 and the notation of Section 4.1.1, the following inequality holds: there exists a positive constant C not depending on L such that, for any $\ell \in \{1, \dots, 2^{L-m(L)}\}$,*

$$\mathbb{E}\big(d_{r(L)}(U_{L,\ell}, V_{L,\ell})\big) \le C2^{(m(L)+2r(L))/((2+\delta)\wedge 3)}L^2.$$

PROOF. From the stationarity of the sequence $((U_{L,\ell}, V_{L,\ell}))_{\ell=1,\dots,2^{L-m(L)}}$, it suffices to prove the proposition for $\ell = 1$. Let $L \in \mathbb{N}$ and $K \in \{0, \dots, r(L)-1\}$. To shorten the notation, let us define the following set of integers:

$$\mathcal{E}(L,K) = \{1, \dots, 2^{r(L)-K} - 1\} \cap (2\mathbb{N}+1),$$

meaning that if $k \in \mathcal{E}(L,K)$, then k is an odd integer in $[1, 2^{r(L)-K} - 1]$.

For $K \in \{0, \dots, r(L)-1\}$ and $k \in \mathcal{E}(L, K)$, define

$$B_{K,k} = \left]\frac{(k-1)2^K}{2^{r(L)}}, \frac{k2^K}{2^{r(L)}}\right] \quad \text{and} \quad Z_L^{(K,k)} = \sum_{i \in I_{L,1}} \big(\mathbf{1}_{Y_i \in B_{K,k}} - P_Y(B_{K,k})\big).$$

The associated column vector Z_L in $\mathbb{R}^{2^{r(L)}-1}$ is then defined by

$$Z_L = \big((Z_L^{(i,k_i)}, k_i \in \mathcal{E}(L, i))_{i=0,\dots,r(L)-1}\big)'.$$

Notice that for any $j \in \{1, \dots, 2^{r(L)}-1\}$,

$$U_{L,1}^{(j)} = \sum_{K=0}^{r(L)-1} \sum_{k_K \in \mathcal{E}(L,K)} b_{K,k_K}(j) Z_L^{(K,k_K)} \tag{4.32}$$

with $b_{K,k_K}(j) = 0$ or 1. This representation is unique in the sense that, for j fixed, there exists only one vector $(b_{(K,k_K)}(j), k_K \in \mathcal{E}(L, K))_{K=0,\dots,r(L)-1}$ satisfying (4.32). In addition, for any K in $\{0, \dots, r(L)-1\}$, $\sum_{k \in \mathcal{E}(L,K)} b_{K,k}(j) \le 1$. Let the column vector $b(j, L)$ and the matrix $\mathbf{P}_L$ be defined by

$$b(j, L) = \big((b_{K,k_K}(j), k_K \in \mathcal{E}(L, K))_{K=0,\dots,r(L)-1}\big)'$$

and

$$\mathbf{P}_L = \big(b(1, L), b(2, L), \dots, b(2^{r(L)}-1, L)\big)'.$$

$\mathbf{P}_L$ has the following property: it is a square matrix of $\mathbb{R}^{2^{r(L)}-1}$ with determinant equal to 1. Let us denote by $\mathbf{P}_L^{-1}$ its inverse. With this notation, we then notice that

$$Z_L = \mathbf{P}_L^{-1} U_{L,1}. \tag{4.33}$$

Let now a^2 be a positive real and $V = (V^{(1)}, \dots, V^{(2^{r(L)}-1)})'$ be a random variable with law $\mathcal{N}(0, a^2 \mathbf{P}_L \mathbf{P}_L^T)$. According to the coupling relation (4.5), we have that

$$\begin{aligned}
\mathbb{E}\big(d_{r(L)}(U_{L,1}, V_{L,1})\big) &= \mathbb{E}\big(W_{d_{r(L)}}(P_{U_{L,1}|\mathcal{F}_{2^L}}, G_{2^{m(L)}\Lambda_L})\big) \\
&\le \mathbb{E}\big(W_{d_{r(L)}}(P_{U_{L,1}|\mathcal{F}_{2^L}} * P_V, G_{2^{m(L)}\Lambda_L} * P_V)\big) \\
&\quad + 2\mathbb{E}\big(d_{r(L)}(V, 0)\big),
\end{aligned} \tag{4.34}$$

where $*$ stands for the usual convolution product. Since $V^{(j)}$ is a centered real Gaussian random variable with variance $v_j^2 = a^2 \sum_{K=0}^{r(L)-1} \sum_{k \in \mathcal{E}(L,K)} b_{K,k}(j)$, according to inequality (3.6) in Ledoux and Talagrand (1991), we derive that

$$\begin{aligned}
\mathbb{E}\big(d_{r(L)}(V, 0)\big) &= \mathbb{E}\Big(\max_{j \in \{1,\dots,2^{r(L)}-1\}} |V^{(j)}|\Big) \\
&\le \big(2 + 3(\log(2^{r(L)}-1))^{1/2}\big) \max_{j \in \{1,\dots,2^{r(L)}-1\}} v_j.
\end{aligned}$$

Since $v_j^2 \le a^2 r(L) \le a^2 L$, we then get that

$$\mathbb{E}\big(d_{r(L)}(V,0)\big) \le 5aL. \tag{4.35}$$

Let us now give an upper bound for the quantity $\mathbb{E}(W_{d_{r(L)}}(P_{U_{L,1}|\mathcal{F}_{2L}} * P_V, G_{2^{m(L)}\Lambda_L} * P_V))$ in (4.34). Let $(N_{i,L})_{i\in\mathbb{Z}}$ be a sequence of independent random variables with normal distribution $\mathcal{N}(0,\Lambda_L)$. Suppose, furthermore, that the sequence $(N_{i,L})_{i\in\mathbb{Z}}$ is independent of $\mathcal{F}_\infty \vee \sigma(\eta_i, i\in\mathbb{Z})$. Denote by $\mathrm{I}_{2^{r(L)}-1}$ the identity matrix on $\mathbb{R}^{2^{r(L)}-1}$ and let N be a $\mathcal{N}(0, a^2\mathrm{I}_{2^{r(L)}-1})$-distributed random variable, independent of $\mathcal{F}_\infty \vee \sigma(N_{i,L}, i\in\mathbb{Z}) \vee \sigma(\eta_i, i\in\mathbb{Z})$. Set $\widetilde{N}_L = N_{1,L} + N_{2,L} + \cdots + N_{2^{m(L)},L}$. We first notice that

$$\begin{aligned}
&\mathbb{E}\big(W_{d_{r(L)}}(P_{U_{L,1}|\mathcal{F}_{2L}} * P_V, G_{2^{m(L)}\Lambda_L} * P_V)\big)\\
&\quad = \mathbb{E} \sup_{f\in \mathrm{Lip}(d_{r(L)})} \big(\mathbb{E}\big(f(U_{L,1} + \mathbf{P}_L N)|\mathcal{F}_{2L}\big) - \mathbb{E}\big(f(\widetilde{N}_L + \mathbf{P}_L N)\big)\big).
\end{aligned} \tag{4.36}$$

Introduce now the following definition:

DEFINITION 4.3. For two column vectors

$$x = \big((x^{(i,k_i)}, k_i \in \mathcal{E}(L,i))_{i=0,\dots,r(L)-1}\big)'$$

and $y = ((y^{(i,k_i)}, k_i \in \mathcal{E}(L,i))_{i=0,\dots,r(L)-1})'$ of $\mathbb{R}^{2^{r(L)}-1}$, let $d^*_{r(L)}$ be the following distance:

$$d^*_{r(L)}(x,y) = \sum_{K=0}^{r(L)-1} \sup_{k\in\mathcal{E}(L,K)} \big|x^{(K,k)} - y^{(K,k)}\big|.$$

Let also $\mathrm{Lip}(d^*_{r(L)})$ be the set of functions from $\mathbb{R}^{2^{r(L)}-1}$ into $\mathbb{R}$ that are Lipschitz with respect to $d^*_{r(L)}$, namely, $|f(x) - f(y)| \le \sum_{K=0}^{r(L)-1} \sup_{k\in\mathcal{E}(L,K)} |x^{(K,k)} - y^{(K,k)}|$.

Let $x = (x^{(1)},\dots,x^{(2^{r(L)}-1)})'$ and $y = (y^{(1)},\dots,y^{(2^{r(L)}-1)})'$ be two column vectors of $\mathbb{R}^{2^{r(L)}-1}$. Let now $u = \mathbf{P}_L^{-1}x$ and $v = \mathbf{P}_L^{-1}y$. The vectors u and v of $\mathbb{R}^{2^{r(L)}-1}$ can be rewritten $u = ((u^{(i,k_i)}, k_i \in \mathcal{E}(L,i))_{i=0,\dots,r(L)-1})'$ and $v = ((v^{(i,k_i)}, k_i\in\mathcal{E}(L,i))_{i=0,\dots,r(L)-1})'$. Notice now that if $f \in \mathrm{Lip}(d_{r(L)})$, then

$$\begin{aligned}
&\big|f(x) - f(y)\big|\\
&\quad \le d_{r(L)}(x,y) = \sup_{j\in\{1,\dots,2^{r(L)}-1\}} \big|b(j,L)'u - b(j,L)'v\big|\\
&\quad \le \sup_{j\in\{1,\dots,2^{r(L)}-1\}} \sum_{K=0}^{r(L)-1} \sum_{k_K\in\mathcal{E}(L,K)} b_{K,k_K}(j)\big|u^{(K,k_K)} - v^{(K,k_K)}\big|\\
&\quad \le \sup_{j\in\{1,\dots,2^{r(L)}-1\}} \sum_{K=0}^{r(L)-1} \sum_{k_K\in\mathcal{E}(L,K)} b_{K,k_K}(j) \sup_{i\in\mathcal{E}(L,K)} \big|u^{(K,i)} - v^{(K,i)}\big|.
\end{aligned}$$

Since for any $K \in \{0, \dots, r(L)-1\}$ and any $j \in \{0, \dots, 2^{r(L)}-1\}$,

$$\sum_{k \in \mathcal{E}(L,K)} b_{K,k}(j) \le 1,$$

it follows that if $f \in \mathrm{Lip}(d_{r(L)})$,

$$\begin{aligned}|f(x) - f(y)| = |f \circ \mathbf{P}_L(u) - f \circ \mathbf{P}_L(v)| &\le \sum_{K=0}^{r(L)-1} \sup_{k \in \mathcal{E}(L,K)} |u^{(K,k)} - v^{(K,k)}| \\ &= d^*_{r(L)}(u, v).\end{aligned}$$

Therefore, starting from (4.36) and taking into account (4.33), we get

$$\begin{aligned}&\mathbb{E}\big(W_{d_{r(L)}}(P_{U_{L,1}|\mathcal{F}_{2^L}} * P_V, G_{2^{m(L)}\Lambda_L} * P_V)\big) \\ &\quad \le \mathbb{E} \sup_{f \in \mathrm{Lip}(d^*_{r(L)})} \big(\mathbb{E}(f(Z_L + N)|\mathcal{F}_{2^L}) - \mathbb{E}\big(f(\mathbf{P}_L^{-1}\widetilde{N}_L + N)\big)\big).\end{aligned} \tag{4.37}$$

Let $\mathrm{Lip}(d^*_{r(L)}, \mathcal{F}_{2^L})$ be the set of measurable functions $g : \mathbb{R}^{2^{r(L)}-1} \times \Omega \to \mathbb{R}$ wrt the σ-fields $\mathcal{B}(\mathbb{R}^{2^{r(L)}-1}) \otimes \mathcal{F}_{2^L}$ and $\mathcal{B}(\mathbb{R})$, such that $g(\cdot, \omega) \in \mathrm{Lip}(d^*_{r(L)})$ and $g(0, \omega) = 0$ for any $\omega \in \Omega$. For the sake of brevity, we shall write $g(x)$ in place of $g(x, \omega)$. From Point 2 of Theorem 1 in Dedecker, Prieur and Raynaud De Fitte (2006), the following inequality holds:

$$\begin{aligned}&\mathbb{E} \sup_{f \in \mathrm{Lip}(d^*_{r(L)})} \big(\mathbb{E}(f(Z_L + N)|\mathcal{F}_{2^L}) - \mathbb{E}\big(f(\mathbf{P}_L^{-1}\widetilde{N}_L + N)\big)\big) \\ &\quad = \sup_{g \in \mathrm{Lip}(d^*_{r(L)}, \mathcal{F}_{2^L})} \mathbb{E}\big(g(Z_L + N)\big) - \mathbb{E}\big(g(\mathbf{P}_L^{-1}\widetilde{N}_L + N)\big).\end{aligned} \tag{4.38}$$

We shall prove that if $a \in [L, L2^{m(L)}]$, there exists a positive constant C not depending on (L, a), such that

$$\begin{aligned}&\sup_{g \in \mathrm{Lip}(d^*_{r(L)}, \mathcal{F}_{2^L})} \mathbb{E}\big(g(Z_L + N)\big) - \mathbb{E}\big(g(\mathbf{P}_L^{-1}\widetilde{N}_L + N)\big) \\ &\quad \le C a^{-3} L^{5/2} 2^{m(L)} \\ &\qquad + C L^{-1} 2^{2r(L)} + C a^{-1-\delta} L^{\delta} 2^{2r(L)+m(L)} \\ &\qquad + C a^{-2} L^2 2^{2r(L)+m(L)} + C a^{-1} L^2 2^{r(L)}.\end{aligned} \tag{4.39}$$

Gathering (4.39), (4.38), (4.37), (4.34) and (4.35), and taking

$$a = L 2^{(m(L)+2r(L))/((2+\delta)\wedge 3)},$$

Proposition 4.1 will follow.

Let then $a \in [L, L2^{m(L)}]$ and continue the proof by proving (4.39). For any $i \geq 1$, let $Y_{i,L}$ be the column vector defined by $Y_{i,L} = (Y_{i,L}^{(1)}, \dots, Y_{i,L}^{(2^{r(L)}-1)})'$, where $Y_{i,L}^{(j)} = \mathbf{1}_{Y_{i+2^L} \leq s_j} - F_Y(s_j)$. Notice then that

$$Z_L = \sum_{i=1}^{2^{m(L)}} Z_{i,L} \qquad \text{where } Z_{i,L} = \mathbf{P}_L^{-1} Y_{i,L}.$$

Therefore,

$$Z_{i,L} = \bigl(\bigl(Z_{i,L}^{(K,k_K)}, k_K \in \mathcal{E}(L,K)\bigr)_{K=0,\dots,r(L)-1}\bigr)',$$

where $Z_{i,L}^{(K,k)} = \mathbf{1}_{Y_{i+2^L} \in B_{K,k}} - P_Y(B_{K,k})$.

NOTATION 4.1. Let φ_a be the density of N and let for $x = ((x^{(i,k_i)}, k_i \in \mathcal{E}(L,K))_{i=0,\dots,r(L)-1})'$,

$$g * \varphi_a(x, \omega) = \int g(x + y, \omega) \varphi_a(y)\, dy.$$

For the sake of brevity, we shall write $g * \varphi_a(x)$ instead of $g * \varphi_a(x, \omega)$ (the partial derivatives will be taken wrt x). Let also

$$S_{0,L} = 0 \quad \text{and} \quad \text{for } j > 0, \qquad S_{j,L} = \sum_{i=1}^{j} Z_{i,L}.$$

We now use the Lindeberg method to prove (4.39). We first write that

$$\begin{aligned} &\mathbb{E}\bigl(g(Z_L + N) - g\bigl(\mathbf{P}_L^{-1}\widetilde{N}_L + N\bigr)\bigr) \\ &\quad = \sum_{i=1}^{2^{m(L)}} \mathbb{E}\Biggl(g\Biggl(S_{i-1,L} + Z_{i,L} + \sum_{j=i+1}^{2^{m(L)}} \mathbf{P}_L^{-1} N_{j,L} + N\Biggr) \\ &\qquad\qquad - g\Biggl(S_{i-1,L} + \mathbf{P}_L^{-1} N_{i,L} + \sum_{j=i+1}^{2^{m(L)}} \mathbf{P}_L^{-1} N_{j,L} + N\Biggr)\Biggr) \qquad (4.40) \\ &\quad \leq \sum_{i=1}^{2^{m(L)}} \sup_{g \in \mathrm{Lip}(d^*_{r(L)}, \mathcal{F}_{2^L})} \mathbb{E}\bigl(g(S_{i-1,L} + Z_{i,L} + N) \\ &\qquad\qquad - g\bigl(S_{i-1,L} + \mathbf{P}_L^{-1} N_{i,L} + N\bigr)\bigr). \end{aligned}$$

Let us introduce some notation and definitions.

DEFINITION 4.4. For two positive integers m and n, let $\mathcal{M}_{m,n}(\mathbb{R})$ be the set of real matrices with m lines and n columns. The Kronecker product (or Tensor product) of $A = [a_{i,j}] \in \mathcal{M}_{m,n}(\mathbb{R})$ and $B = [b_{i,j}] \in \mathcal{M}_{p,q}(\mathbb{R})$ is denoted by $A \otimes B$ and is defined to be the block matrix

$$A \otimes B = \begin{pmatrix} a_{1,1}B & \cdots & a_{1,n}B \\ \vdots & & \vdots \\ a_{m,1}B & \cdots & a_{m,n}B \end{pmatrix} \in \mathcal{M}_{mp,nq}(\mathbb{R}).$$

For any positive integer k, the kth Kronecker power $A^{\otimes k}$ is defined inductively by $A^{\otimes 1} = A$ and $A^{\otimes k} = A \otimes A^{\otimes (k-1)}$.

If ∇ denotes the differentiation operator given by $\nabla = (\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_m})'$ acting on the differentiable functions $f : \mathbb{R}^m \to \mathbb{R}$, we define

$$\nabla \otimes \nabla = \left(\frac{\partial}{\partial x_1} \circ \nabla, \dots, \frac{\partial}{\partial x_m} \circ \nabla \right)'$$

and $\nabla^{\otimes k}$ by $\nabla^{\otimes 1} = \nabla$ and $\nabla^{\otimes k} = \nabla \otimes \nabla^{\otimes (k-1)}$. If $f : \mathbb{R}^m \to \mathbb{R}$ is k-times differentiable, for any $x \in \mathbb{R}^m$, let $D^k f(x) = \nabla^{\otimes k} f(x)$, and for any vector A of $\mathbb{R}^m$, we define $D^k f(x).A^{\otimes k}$ as the usual scalar product in $\mathbb{R}^{m^k}$ between $D^k f(x)$ and $A^{\otimes k}$.

For any $i \in \{1, \dots, 2^{m(L)}\}$, let $G_{i,L} = \mathbf{P}_L^{-1} N_{i,L}$,

$$\Delta_{1,i,L}(g) = g * \varphi_a(S_{i-1,L} + Z_{i,L}) - g * \varphi_a(S_{i-1,L}) - \tfrac{1}{2} D^2 g * \varphi_a(S_{i-1,L}).G_{i,L}^{\otimes 2}$$

and

$$\Delta_{2,i,L}(g) = g * \varphi_a(S_{i-1,L} + G_{i,L}) - g * \varphi_a(S_{i-1,L}) - \tfrac{1}{2} D^2 g * \varphi_a(S_{i-1,L}).G_{i,L}^{\otimes 2}.$$

With this notation,

$$\begin{aligned} &\mathbb{E}\big(g(S_{i-1,L} + Z_{i,L} + N) - g\big(S_{i-1,L} + \mathbf{P}_L^{-1} N_{i,L} + N\big)\big) \\ &\qquad = \mathbb{E}\big(\Delta_{1,i,L}(g)\big) - \mathbb{E}\big(\Delta_{2,i,L}(g)\big). \end{aligned} \tag{4.41}$$

By the Taylor integral formula, noticing that $\mathbb{E}(G_{i,L}^{\otimes 3}) = 0$, we get

$$\big|\mathbb{E}\big(\Delta_{2,i,L}(g)\big)\big| \le \frac{1}{6} \left| \mathbb{E} \int_0^1 D^4 g * \varphi_a(S_{i-1,L} + t G_{i,L}).G_{i,L}^{\otimes 4}\, dt \right|.$$

Applying Lemma A.5, we then derive that

$$\begin{aligned} &\big|\mathbb{E}\big(\Delta_{2,i,L}(g)\big)\big| \\ &\quad \ll a^{-3} \mathbb{E}\left(\left(\sum_{K=0}^{r(L)-1} \sup_{k \in \mathcal{E}(L,K)} \big|G_{1,L}^{(K,k)}\big| \right) \left(\sum_{K=0}^{r(L)-1} \sum_{k_K \in \mathcal{E}(L,K)} (G_{1,L}^{(K,k_K)})^2 \right)^{3/2} \right) \\ &\quad \ll a^{-3} \left(\mathbb{E}\left(\sum_{K=0}^{r(L)-1} \sup_{k \in \mathcal{E}(L,K)} \big|G_{1,L}^{(K,k)}\big| \right)^4 \right)^{1/4} \\ &\qquad \times \left(\mathbb{E}\left(\sum_{K=0}^{r(L)-1} \sum_{k_K \in \mathcal{E}(L,K)} (G_{1,L}^{(K,k_K)})^2 \right)^2 \right)^{3/4}. \end{aligned} \tag{4.42}$$

Notice that

$$\sum_{K=0}^{r(L)-1} \sup_{k\in\mathcal{E}(L,K)} \big|G_{1,L}^{(K,k)}\big| \le \sum_{K=0}^{r(L)-1}\Big(\sum_{k_K\in\mathcal{E}(L,K)} \big(G_{1,L}^{(K,k_K)}\big)^2\Big)^{1/2}$$
$$\le \sqrt{r(L)}\Big(\sum_{K=0}^{r(L)-1}\sum_{k_K\in\mathcal{E}(L,K)} \big(G_{1,L}^{(K,k_K)}\big)^2\Big)^{1/2}. \tag{4.43}$$

Moreover,

$$\mathbb{E}\Big(\sum_{K=0}^{r(L)-1}\sum_{k_K\in\mathcal{E}(L,K)} \big(G_{1,L}^{(K,k_K)}\big)^2\Big)^2 \le \Big(\sum_{K=0}^{r(L)-1}\sum_{k_K\in\mathcal{E}(L,K)} \big(\mathbb{E}\big(G_{1,L}^{(K,k_K)}\big)^4\big)^{1/2}\Big)^2$$
$$\le 3\Big(\sum_{K=0}^{r(L)-1}\sum_{k_K\in\mathcal{E}(L,K)} \mathbb{E}\big(\big(G_{1,L}^{(K,k_K)}\big)^2\big)\Big)^2$$

and

$$\sum_{k\in\mathcal{E}(L,K)} \mathbb{E}\big(\big(G_{1,L}^{(K,k)}\big)^2\big) = \sum_{k\in\mathcal{E}(L,K)}\Big(\mathrm{Var}\big(Z_{1,L}^{(K,k)}\big) + 2\sum_{i>0}\mathrm{Cov}\big(Z_{1,L}^{(K,k)}, Z_{i+1,L}^{(K,k)}\big)\Big).$$

Arguing as to get (4.30), we then obtain that

$$\sum_{k\in\mathcal{E}(L,K)} \mathbb{E}\big(\big(G_{1,L}^{(K,k)}\big)^2\big) \le C(\beta)\sum_{k\in\mathcal{E}(L,K)} 2^{K-r(L)} \le C(\beta).$$

From the above computations, it follows that

$$\mathbb{E}\Big(\sum_{K=0}^{r(L)-1}\sum_{k_K\in\mathcal{E}(L,K)} \big(G_{1,L}^{(K,k_K)}\big)^2\Big)^2 \le 3\big(C(\beta)r(L)\big)^2. \tag{4.44}$$

Therefore, starting from (4.42), taking into account (4.43), (4.44) and the fact that $r(L) \le L$, we then derive that

$$\big|\mathbb{E}\big(\Delta_{2,i,L}(g)\big)\big| \ll a^{-3}L^{5/2}. \tag{4.45}$$

Let now

$$R_{1,i,L}(g) = g*\varphi_a(S_{i-1,L}+Z_{i,L}) - g*\varphi_a(S_{i-1,L}) - Dg*\varphi_a(S_{i-1,L}).Z_{i,L}$$
$$- \tfrac{1}{2}D^2g*\varphi_a(S_{i-1,L}).Z_{i,L}^{\otimes 2}$$

and

$$D_{1,i,L}(g) = Dg*\varphi_a(S_{i-1,L}).Z_{i,L} + \tfrac{1}{2}D^2(g*\varphi_a)(S_{i-1,L}).Z_{i,L}^{\otimes 2}$$
$$- \tfrac{1}{2}D^2g*\varphi_a(S_{i-1,L}).\mathbb{E}\big(G_{i,L}^{\otimes 2}\big).$$

With this notation,

$$\mathbb{E}\big(\Delta_{1,i,L}(g)\big) = \mathbb{E}\big(R_{1,i,L}(g)\big) + \mathbb{E}\big(D_{1,i,L}(g)\big). \tag{4.46}$$

By the Taylor integral formula,

$$\big|\mathbb{E}\big(R_{1,i,L}(g)\big)\big| \le \left|\mathbb{E}\int_0^1 \frac{(1-t)^2}{2} D^3 g * \varphi_a(S_{i-1,L} + tZ_{i,L}).Z_{i,L}^{\otimes 3}\right|.$$

Applying Lemma A.5 and using the fact that $\sup_{k\in\mathcal{E}(L,K)} |Z_{i,L}^{(K,k)}| \le 2$ and $\sum_{k\in\mathcal{E}(L,K)} (Z_{i,L}^{(K,k)})^2 \le 2$, we get that

$$\big|\mathbb{E}\big(R_{1,i,L}(g)\big)\big| \ll a^{-2}\big(r(L)\big)^2 \ll a^{-2}L^2. \tag{4.47}$$

Let

$$\Delta(i,j)(g) = D^2 g * \varphi_a(S_{i-j,L}) - D^2 g * \varphi_a(S_{i-j-1,L}) \tag{4.48}$$

and

$$u_L = [aL^{-1}]. \tag{4.49}$$

Clearly, with the notation $X^{(0)} = X - \mathbb{E}(X)$,

$$\begin{aligned} D^2 g * \varphi_a(S_{i-1,L}).\big(Z_{i,L}^{\otimes 2}\big)^{(0)} = &\sum_{j=1}^{(u_L\wedge i)-1} \Delta(i,j)(g).\big(Z_{i,L}^{\otimes 2}\big)^{(0)} \\ &+ D^2 g * \varphi_a(S_{i-(u_L\wedge i),L}).\big(Z_{i,L}^{\otimes 2}\big)^{(0)}. \end{aligned} \tag{4.50}$$

For any $j \le (u_L \wedge i) - 1$, write

$$\mathbb{E}\big(\Delta(i,j)(g).\big(Z_{i,L}^{\otimes 2}\big)^{(0)}\big) = \mathbb{E}\big(\Delta(i,j)(g).\mathbb{E}_{i-j+2^L}\big(\big(Z_{i,L}^{\otimes 2}\big)^{(0)}\big)\big)$$

and notice that, by Lemma A.6,

$$\begin{aligned} &\mathbb{E}\big(\Delta(i,j)(g).\mathbb{E}_{i-j+2^L}\big(Z_{i,L}^{\otimes 2}\big)^{(0)}\big) \\ &\quad\le \sup_{t\in[0,1]} \big|\mathbb{E}\big(D^3 g * \varphi_a(S_{i-j-1,L} + tZ_{i-j,L}).\big(Z_{i-j,L} \otimes \mathbb{E}_{i-j+2^L}\big(Z_{i,L}^{\otimes 2}\big)^{(0)}\big)\big)\big| \\ &\quad\ll a^{-2} \sum_{K_1,k_{K_1}} \sum_{K_2,k_{K_2}} \sum_{K_3,k_{K_3}} \mathbb{E}\big(\big|Z_{i-j,L}^{K_1,k_{K_1}}\big|\big|\mathbb{E}_{i-j+2^L}\big(Z_{i,L}^{K_2,k_{K_2}} Z_{i,L}^{K_3,k_{K_3}} \\ &\qquad\qquad - \mathbb{E}\big(Z_{i,L}^{K_2,k_{K_2}} Z_{i,L}^{K_3,k_{K_3}}\big)\big)\big|\big), \end{aligned}$$

where $K_i \in \{0,\dots,r(L)-1\}$ and $k_{K_i} \in \mathcal{E}(L,K_i)$, for any $i \in \{1,2,3\}$. Applying Lemma A.1, we infer that

$$\big|\mathbb{E}_{i-j+2^L}\big(Z_{i,L}^{K_2,k_{K_2}} Z_{i,L}^{K_3,k_{K_3}} - \mathbb{E}\big(Z_{i,L}^{K_2,k_{K_2}} Z_{i,L}^{K_3,k_{K_3}}\big)\big)\big| \le 4b_1\big(\mathcal{F}_{i-j+2^L}, i+2^L\big).$$

Since $\sum_{K_1=0}^{r(L)-1}\sum_{k_{K_1}\in\mathcal{E}(L,K_1)}|Z_{i-j,L}^{K_1,k_{K_1}}| \le 2r(L)$ and $\mathbb{E}(b_1(\mathcal{F}_{i-j+2^L}, i+2^L)) \le \beta_{1,X}(j)$, we then derive that

$$\mathbb{E}\big(\Delta(i,j)(g).\big(Z_{i,L}^{\otimes 2}\big)^{(0)}\big) \ll a^{-2}r(L)2^{2r(L)}\beta_{1,X}(j). \tag{4.51}$$

On the other hand, by using Lemma A.6, we infer that

$$\begin{aligned}
&\mathbb{E}\big(D^2 g*\varphi_a(S_{i-(u_L\wedge i),L}).\big(Z_{i,L}^{\otimes 2}\big)^{(0)}\big)\\
&\quad=\mathbb{E}\big(D^2 g*\varphi_a(S_{i-(u_L\wedge i),L}).\mathbb{E}_{i-(u_L\wedge i)+2^L}\big(Z_{i,L}^{\otimes 2}\big)^{(0)}\big)\\
&\quad\ll a^{-1}\sum_{K_1,k_{K_1}}\sum_{K_2,k_{K_2}}\mathbb{E}\big(\big|\mathbb{E}_{i-(u_L\wedge i)+2^L}\big(Z_{i,L}^{K_1,k_{K_1}}Z_{i,L}^{K_1,k_{K_1}}\\
&\qquad\qquad\qquad\qquad -\mathbb{E}\big(Z_{i,L}^{K_1,k_{K_1}}Z_{i,L}^{K_1,k_{K_1}}\big)\big)\big|\big).
\end{aligned}$$

Using the same arguments as to get (4.51), we obtain that

$$\mathbb{E}\big(D^2 g*\varphi_a(S_{i-(u_L\wedge i),L}).\big(Z_{i,L}^{\otimes 2}\big)^{(0)}\big) \ll a^{-1}2^{2r(L)}\beta_{1,X}(u_L\wedge i). \tag{4.52}$$

Starting from (4.50) and taking into account (4.51), (4.52), the choice of u_L and the condition on the β-dependence coefficients, we then derive that

$$\begin{aligned}
&\sum_{i=1}^{2^{m(L)}}\mathbb{E}\big(D^2 g*\varphi_a(S_{i-1,L}).\big(Z_{i,L}^{\otimes 2}\big)^{(0)}\big)\\
&\qquad\ll 2^{2r(L)}a^{-1}\bigg(\frac{2^{m(L)}L^{1+\delta}}{a^{1+\delta}}+2^{m(L)}\frac{L}{a}\bigg).
\end{aligned}\tag{4.53}$$

To give now an estimate of the expectation of $Dg*\varphi_a(S_{i-1,L}).Z_{i,L}$, we write

$$Dg*\varphi_a(S_{i-1,L}) = Dg*\varphi_a(0)+\sum_{j=1}^{i-1}\big(Dg*\varphi_a(S_{i-j,L})-Dg*\varphi_a(S_{i-j-1,L})\big).$$

Hence,

$$\begin{aligned}
&\mathbb{E}\big(Dg*\varphi_a(S_{i-1,L}).Z_{i,L}\big)\\
&\quad=\mathbb{E}\big(Dg*\varphi_a(0).Z_{i,L}\big)\\
&\qquad+\sum_{j=1}^{i-1}\mathbb{E}\big(\big(Dg*\varphi_a(S_{i-j,L})-Dg*\varphi_a(S_{i-j-1,L})\big).Z_{i,L}\big).
\end{aligned}\tag{4.54}$$

Applying Lemma A.1,

$$\begin{aligned}
\big|\mathbb{E}\big(Dg*\varphi_a(0).Z_{i,L}\big)\big| &= \big|\mathbb{E}\big(Dg*\varphi_a(0).\mathbb{E}_{2^L}(Z_{i,L})\big)\big|\\
&\le \mathbb{E}\Bigg(\sum_{K=0}^{r(L)-1}\sum_{k_K\in\mathcal{E}(L,K)}\bigg|\frac{\partial g*\varphi_a}{\partial x^{(K,k_K)}}(0)\bigg|b_1\big(\mathcal{F}_{2^L}, i+2^L\big)\Bigg).
\end{aligned}$$

Notice now that by inequality (A.3), for any K in $\{0,\dots,r(L)-1\}$, the random variable

$$\sum_{k\in\mathcal{E}(L,K)}\left|\frac{\partial g*\varphi_a}{\partial x^{(K,k)}}(0)\right|$$

is a $\mathcal{F}_{2^L}$-measurable random variable with infinite norm less than one. Therefore,

$$\bigl|\mathbb{E}\bigl(Dg*\varphi_a(0).Z_{i,L}\bigr)\bigr|\ll r(L)\beta_{1,X}(i). \tag{4.55}$$

We give now an estimate of $\sum_{j=1}^{i-1}\mathbb{E}((Dg*\varphi_a(S_{i-j,L})-Dg*\varphi_a(S_{i-j-1,L})).Z_{i,L})$. By Lemmas A.6 and A.1, for any $i\ge j+1$,

$$\begin{aligned}
&\bigl|\mathbb{E}\bigl((Dg*\varphi_a(S_{i-j,L})-Dg*\varphi_a(S_{i-j-1,L})).Z_{i,L}\bigr)\bigr|\\
&\quad=\bigl|\mathbb{E}\bigl((Dg*\varphi_a(S_{i-j,L})-Dg*\varphi_a(S_{i-j-1,L})).\mathbb{E}_{i-j+2^L}(Z_{i,L})\bigr)\bigr|\\
&\quad\le\sup_{t\in[0,1]}\bigl|\mathbb{E}\bigl(D^2g*\varphi_a(S_{i-j-1,L}+tZ_{i,L}).\bigl(Z_{i-j,L}\otimes\mathbb{E}_{i-j+2^L}(Z_{i,L})\bigr)\bigr)\bigr|\\
&\quad\ll a^{-1}\sum_{K_1=0}^{r(L)-1}\sum_{k_{K_1}\in\mathcal{E}(L,K_1)}\sum_{K_2=0}^{r(L)-1}\sum_{k_{K_2}\in\mathcal{E}(L,K_2)}\mathbb{E}\bigl(|Z_{i-j,L}^{K_1,k_{K_1}}|b_1(\mathcal{F}_{i-j+2^L},i+2^L)\bigr).
\end{aligned}$$

We then infer that for any $i\ge j+1$,

$$\begin{aligned}
&\bigl|\mathbb{E}\bigl((Dg*\varphi_a(S_{i-j,L})-Dg*\varphi_a(S_{i-j-1,L})).Z_{i,L}\bigr)\bigr|\\
&\qquad\ll a^{-1}r(L)2^{r(L)}\beta_{1,X}(j).
\end{aligned}\tag{4.56}$$

From now on, we assume that $j<i\wedge u_L$. Notice that

$$\begin{aligned}
&\bigl(Dg*\varphi_a(S_{i-j,L})-Dg*\varphi_a(S_{i-j-1,L})\bigr).Z_{i,L}\\
&\quad-D^2g*\varphi_a(S_{i-j-1,L}).(Z_{i-j,L}\otimes Z_{i,L})\\
&\qquad=\int_0^1(1-t)D^3g*\varphi_a(S_{i-j-1,L}+tZ_{i-j,L}).\bigl(Z_{i-j,L}^{\otimes 2}\otimes Z_{i,L}\bigr)\,dt.
\end{aligned}$$

By using Lemmas A.6 and A.1, we infer that

$$\begin{aligned}
&\left|\mathbb{E}\left(\int_0^1(1-t)D^3g*\varphi_a(S_{i-j-1,L}+tZ_{i-j,L}).\bigl(Z_{i-j,L}^{\otimes 2}\otimes Z_{i,L}\bigr)\,dt\right)\right|\\
&\quad\ll a^{-2}\sum_{K_1=0}^{r(L)-1}\sum_{k_{K_1}\in\mathcal{E}(L,K_1)}\sum_{K_2=0}^{r(L)-1}\sum_{k_{K_2}\in\mathcal{E}(L,K_2)}\sum_{K_3=0}^{r(L)-1}\sum_{k_{K_3}\in\mathcal{E}(L,K_3)}\\
&\qquad\qquad\mathbb{E}\bigl(|Z_{i-j,L}^{K_1,k_{K_1}}||Z_{i-j,L}^{K_2,k_{K_2}}|b_1(\mathcal{F}_{i-j+2^L},i+2^L)\bigr).
\end{aligned}$$

Therefore,

$$
\begin{aligned}
(4.57)\qquad &\Big|\mathbb{E}\Big(\int_0^1 (1-t)D^3 g * \varphi_a(S_{i-j-1,L} + tZ_{i-j,L}) . (Z_{i-j,L}^{\otimes 2} \otimes Z_{i,L})\,dt\Big)\Big| \\
&\quad \ll a^{-2}\big(r(L)\big)^2 2^{r(L)} \beta_{1,X}(j).
\end{aligned}
$$

In order to estimate the term $\mathbb{E}(D^2 g * \varphi_a(S_{i-j-1,L}) . (Z_{i-j,L} \otimes Z_{i,L}))$, we use the following decomposition:

$$
\begin{aligned}
&D^2 g * \varphi_a(S_{i-j-1,L}) \\
&\quad = \sum_{l=1}^{(j-1)\wedge(i-j-1)} \big(D^2 g * \varphi_a(S_{i-j-l,L}) - D^2 g * \varphi_a(S_{i-j-l-1,L})\big) \\
&\qquad + D^2 g * \varphi_a(S_{(i-2j)\vee 0,L}).
\end{aligned}
$$

For any $l \in \{1, \dots, (j-1) \wedge (i-j-1)\}$, using the same arguments as to get (4.57), we obtain that

$$
\begin{aligned}
(4.58)\qquad &\big|\mathbb{E}\big(\big(D^2 g * \varphi_a(S_{i-j-l,L}) - D^2 g * \varphi_a(S_{i-j-l-1,L})\big) . (Z_{i-j,L} \otimes Z_{i,L})\big)\big| \\
&\quad \ll a^{-2}\big(r(L)\big)^2 2^{r(L)} \beta_{1,X}(j).
\end{aligned}
$$

As a second step, we bound up $|\mathbb{E}(D^2 g * \varphi_a(S_{(i-2j)\vee 0,L}) . (Z_{i-j,L} \otimes Z_{i,L})^{(0)})|$. Assume first that $j \le [i/2]$. Clearly, using the notation (4.48),

$$
D^2 g * \varphi_a(S_{i-2j,L}) = \sum_{l=j}^{(u_L-1)\wedge(i-j-1)} \Delta(i, l+j)(g) + D^2 g * \varphi_a(S_{(i-j-u_L)\vee 0,L}).
$$

Now for any $l \in \{j, \dots, (u_L - 1) \wedge (i-j-1)\}$, by using Lemma A.6, we get that

$$
\begin{aligned}
&\big|\mathbb{E}\big(\Delta(i, l+j) . (Z_{i-j,L} \otimes Z_{i,L})^{(0)}\big)\big| \\
&\quad \ll a^{-2} \sum_{K_1,k_{K_1}} \sum_{K_2,k_{K_2}} \sum_{K_3,k_{K_3}} \mathbb{E}\big|Z_{i-j-l,L}^{K_1,k_{K_1}} \mathbb{E}_{i-j-l+2^L}\big(Z_{i-j,L}^{K_2,k_{K_2}} Z_{i,L}^{K_3,k_{K_3}} \\
&\qquad\qquad - \mathbb{E}\big(Z_{i-j,L}^{K_2,k_{K_2}} Z_{i,L}^{K_3,k_{K_3}}\big)\big)\big|.
\end{aligned}
$$

Applying Lemma A.1, we infer that

$$
\begin{aligned}
&\big|\mathbb{E}_{i-j-l+2^L}\big(Z_{i-j,L}^{K_2,k_{K_2}} Z_{i,L}^{K_3,k_{K_3}} - \mathbb{E}\big(Z_{i-j,L}^{K_2,k_{K_2}} Z_{i,L}^{K_3,k_{K_3}}\big)\big)\big| \\
&\quad \le 4b_2\big(\mathcal{F}_{i-j-l+2^L}, i-j+2^L, i+2^L\big).
\end{aligned}
$$

Therefore,

$$
(4.59)\qquad \big|\mathbb{E}\big(\Delta(i, l+j) . (Z_{i-j,L} \otimes Z_{i,L})^{(0)}\big)\big| \ll a^{-2} r(L) 2^{2r(L)} \beta_{2,X}(l).
$$

If $j \le i - u_L$, with similar arguments,

$$|\mathbb{E}(D^2 g * \varphi_a(S_{i-j-u_L,L}).(Z_{i-j,L} \otimes Z_{i,L})^{(0)})| \ll a^{-1} 2^{2r(L)} \beta_{2,X}(u_L). \tag{4.60}$$

Now if $j > i - u_L$, we infer that

$$|\mathbb{E}((D^2 g * \varphi_a(0)).(Z_{i-j,L} \otimes Z_{i,L})^{(0)})| \ll a^{-1} 2^{2r(L)} \beta_{2,X}([i/2]) \tag{4.61}$$

by using also the fact that, since $j \le [i/2]$, $\beta_{2,X}(i-j) \le \beta_{2,X}([i/2])$. Assume now that $j \ge [i/2]+1$. For any $j \le i$, we get

$$|\mathbb{E}((D^2 g * \varphi_a(0)).Z_{i-j,L} \otimes Z_{i,L})| \ll a^{-1} r(L) 2^{r(L)} \beta_{1,X}([i/2]). \tag{4.62}$$

Starting from (4.54), adding inequalities (4.55)–(4.62) and summing on j and l, we then obtain

$$\begin{aligned}
&\Biggl|\mathbb{E}(Dg * \varphi_a(S_{i-1,L}).Z_{i,L}) \\
&\quad - \sum_{j=1}^{u_L-1} \mathbb{E}(D^2 g * \varphi_a(S_{i-2j,L})).\mathbb{E}(Z_{i-j,L} \otimes Z_{i,L}) \mathbf{1}_{j \le [i/2]}\Biggr| \\
&\quad \ll r(L)\beta_{1,X}(i) + a^{-1} L 2^{r(L)} \sum_{j=u_L}^{i} \beta_{1,X}(j) + a^{-1} 2^{2r(L)} u_L \beta_{2,X}(u_L) \\
&\qquad + a^{-1} 2^{2r(L)} u_L \beta_{2,X}([i/2]) + a^{-2} L 2^{2r(L)} \sum_{j=1}^{u_L} j \beta_{2,X}(j).
\end{aligned}$$

Next, summing on i and taking into account the condition on the β-dependence coefficients and the choice of u_L, we get that

$$\begin{aligned}
&\sum_{i=1}^{2^{m(L)}} \Biggl|\mathbb{E}(Dg * \varphi_a(S_{i-1,L}).Z_{i,L}) \\
&\quad - \sum_{j=1}^{u_L-1} \mathbb{E}(D^2 g * \varphi_a(S_{i-2j,L})).\mathbb{E}(Z_{i-j,L} \otimes Z_{i,L}) \mathbf{1}_{j \le [i/2]}\Biggr| \\
&\quad \ll L^{-1} 2^{2r(L)} + a^{-1-\delta} L^{\delta} 2^{2r(L)+m(L)} + a^{-2} L^2 2^{2r(L)+m(L)}.
\end{aligned} \tag{4.63}$$

It remains to bound up

$$\begin{aligned}
A_i := \Biggl|&\sum_{j=1}^{u_L-1} \mathbb{E}(D^2 g * \varphi_a(S_{i-2j})).\mathbb{E}(Z_{i-j,L} \otimes Z_{i,L}) \mathbf{1}_{j \le [i/2]} \\
&- \sum_{j=1}^{\infty} \mathbb{E}(D^2 g * \varphi_a(S_{i-1})).\mathbb{E}(Z_{i-j,L} \otimes Z_{i,L})\Biggr|.
\end{aligned}$$

We first notice that by Lemma A.6, for any positive integer j,

$$\begin{aligned}&\big|\mathbb{E}\big(D^2 g * \varphi_a(S_{i-1})\big).\mathbb{E}(Z_{i-j,L} \otimes Z_{i,L})\big| \\ &\quad \ll a^{-1} \sum_{K_1=0}^{r(L)-1} \sum_{k_{K_1} \in \mathcal{E}(L,K_1)} \sum_{K_2=0}^{r(L)-1} \sum_{k_{K_2} \in \mathcal{E}(L,K_2)} \big|\mathbb{E}\big(Z_{i-j,L}^{K_1,k_{K_1}} \mathbb{E}_{i-j+2^L}\big(Z_{i,L}^{K_2,k_{K_2}}\big)\big)\big|.\end{aligned}$$

Therefore,

$$\big|\mathbb{E}\big(D^2 g * \varphi_a(S_{i-1})\big).\mathbb{E}(Z_{i-j,L} \otimes Z_{i,L})\big| \ll a^{-1} r(L) 2^{r(L)} \beta_{1,X}(j). \tag{4.64}$$

On an other hand, applying Lemma A.6, we obtain for any $i \ge 2$ and any $j \in \{1, \dots, [i/2]\}$,

$$\begin{aligned}&\big|\mathbb{E}\big(D^2 g * \varphi_a(S_{i-1}) - D^2 g * \varphi_a(S_{i-2j})\big).\mathbb{E}(Z_{i-j,L} \otimes Z_{i,L})\big| \\ &\quad \ll a^{-2} \sum_{K_1=0}^{r(L)-1} \sum_{k_{K_1} \in \mathcal{E}(L,K_1)} \sum_{K_2=0}^{r(L)-1} \sum_{k_{K_2} \in \mathcal{E}(L,K_2)} \sum_{K_3=0}^{r(L)-1} \sum_{k_{K_3} \in \mathcal{E}(L,K_3)} \sum_{\ell=1}^{2j-1} \\ &\qquad\qquad \big(\mathbb{E}\big|Z_{i-\ell,L}^{K_1,k_{K_1}}\big|\big) \big|\mathbb{E}\big(Z_{i-j,L}^{K_2,k_{K_2}} \mathbb{E}_{i-j+2^L}\big(Z_{i,L}^{K_3,k_{K_3}}\big)\big)\big|,\end{aligned}$$

which implies that

$$\begin{aligned}&\sum_{j=1}^{u_L-1} \big|\mathbb{E}\big(D^2 g * \varphi_a(S_{i-1}) - D^2 g * \varphi_a(S_{i-2j})\big).\mathbb{E}(Z_{i-j,L} \otimes Z_{i,L})\big| \mathbf{1}_{j \le [i/2]} \\ &\qquad \ll a^{-2} \big(r(L)\big)^2 2^{r(L)} \sum_{j=1}^{u_L} j \beta_{1,X}(j).\end{aligned} \tag{4.65}$$

Therefore, (4.64) together with (4.65), the choice of u_L and the condition on the β-dependence coefficients entail that

$$\sum_{i=1}^{2^{m(L)}} A_i \ll a^{-1} L^2 2^{r(L)} + a^{-2} L^3 2^{r(L)+m(L)} + a^{-1-\delta} L^{1+\delta} 2^{r(L)+m(L)}. \tag{4.66}$$

Taking into account (4.40)–(4.47), (4.53), (4.63) and (4.66), the bound (4.39) follows. □

4.2. *Proof of Theorem* 3.1. Let $(X_i)_{i \in \mathbb{Z}}$ be a stationary Markov chain with transition Kernel Q defined in (3.1). Notice that for all $(s, s') \in [0, 1]^2$,

$$\nu\big(f_s^{(0)} \cdot f_{s'}^{(0)} \circ T^k\big) = \operatorname{Cov}(\mathbf{1}_{X_k \le s}, \mathbf{1}_{X_0 \le s'}).$$

Since $\beta_{2,X}(k)$ satisfies (3.2), according to the proof of item (1) of Theorem 2.1, it follows that item (1) of Theorem 3.1 holds true.

As at the beginning of the proof of Theorem 2.1, we start by considering the probability P_ν^* whose density with respect to ν is given by (4.2). Let F_ν^* be the distribution function of P^*_ν (F^*_ν is continuous since ν is absolutely continuous with respect to the Lebesgue measure). Let now $\widetilde{T}_i = F^*_\nu(T^i)$ and $Y_i = F^*_\nu(X_i)$. Let F_Y be the distribution function of Y_0. Clearly, $R_T(\cdot,\cdot) = R_{\widetilde{T}}(F^*_\nu(\cdot),\cdot)$ almost surely, where

$$R_{\widetilde{T}}(s,t) = \sum_{1\le k\le t} \big(\mathbf{1}_{\widetilde{T}_k \le s} - F_Y(s)\big), \qquad s \in [0,1], t \in \mathbb{R}^+.$$

Theorem 3.1 will then follow if we can prove that there exists a two-parameter Gaussian process $K^*_{\widetilde{T}}$ with covariance function $\Gamma_{\widetilde{T}}$ given by $\Gamma_{\widetilde{T}}(s,s',t,t') = \min(t,t')\Lambda_{\widetilde{T}}(s,s')$, where

$$\Lambda_{\widetilde{T}}(s,s') = \sum_{k\ge 0} \nu\big(f_s^{(0)} \cdot f_{s'}^{(0)} \circ F^*_\nu(T^k)\big) + \sum_{k>0} \nu\big(f_{s'}^{(0)} \cdot f_s^{(0)} \circ F^*_\nu(T^k)\big). \tag{4.67}$$

For $L \in \mathbb{N}$, let $m(L)$ and $r(L)$ be the two sequences of integers defined by (4.8). For any integer j, let $s_j = j2^{-r(L)}$. As for the proof of Theorem 2.1, we start by constructing the approximating Kiefer process $K^*_{\widetilde{T}}$ with covariance function $\Gamma_{\widetilde{T}}$. With this aim, we first define for any $\ell \in \{1,\dots,2^{L-m(L)}\}$,

$$I_{L,\ell} =]2^L + (\ell-1)2^{m(L)}, 2^L + \ell 2^{m(L)}] \cap \mathbb{N}$$

and

$$U_{L,\ell}^{*(j)} = \sum_{i\in I_{L,\ell}} \big(\mathbf{1}_{\widetilde{T}_i \le s_j} - F_Y(s_j)\big).$$

The associated column vectors $U^*_{L,\ell}$ are then defined in $\mathbb{R}^{2^{r(L)}-1}$ by the equality $U^*_{L,\ell} = (U_{L,\ell}^{*(1)},\dots,U_{L,\ell}^{*(2^{r(L)}-1)})'$. Let

$$\Lambda_{\widetilde{T},L} = \big(\Lambda_{\widetilde{T}}(s_j,s_{j'})\big)_{j,j'=1,\dots,2^{r(L)}-1},$$

where the $\Lambda_{\widetilde{T}}(s_j,s_{j'})$ are defined in (4.67). Let $G_{2^{m(L)}\Lambda_{\widetilde{T},L}}$ denote the $\mathcal{N}(0, 2^{m(L)}\Lambda_{\widetilde{T},L})$-law, and for any $\ell \in \{1,\dots,2^{L-m(L)}\}$, let $P_{U^*_{L,\ell}|\mathcal{G}_{2^L+\ell 2^{m(L)}+1}}$ be the conditional law of $U^*_{L,\ell}$ given $\mathcal{G}_{2^L+\ell 2^{m(L)}+1}$, where $\mathcal{G}_m = \sigma(T^i, i\ge m)$. By the Markov property, the following equality holds: $P_{U^*_{L,\ell}|\mathcal{G}_{2^L+\ell 2^{m(L)}+1}} = P_{U^*_{L,\ell}|T^{2^L+\ell 2^{m(L)}+1}}$.

According to Rüschendorf (1985), there exists $V^*_{L,\ell} = (V_{L,\ell}^{*(1)},\dots,V_{L,\ell}^{*(2^{r(L)}-1)})'$ with law $G_{2^{m(L)}\Lambda_{\widetilde{T},L}}$, measurable with respect to $\sigma(\delta_{2^L+\ell 2^{m(L)}}) \vee \sigma(U^*_{L,\ell}) \vee \mathcal{G}_{2^L+\ell 2^{m(L)}+1}$, independent of $\mathcal{G}_{2^L+\ell 2^{m(L)}+1}$, and such that, with the notation of Section 4.1.1,

$$\mathbb{E}\big(d_{r(L)}(U^*_{L,\ell}, V^*_{L,\ell})\big) = \mathbb{E}\big(W_{d_{r(L)}}(P_{U^*_{L,\ell}|\mathcal{G}^*_{2^L+\ell 2^{m(L)}+1}}, G_{2^{m(L)}\Lambda_{\widetilde{T},L}})\big). \tag{4.68}$$

By induction on ℓ, the random variables $(V^*_{L,\ell})_{\ell=1,\dots,2^{L-m(L)}}$ are mutually independent, independent of $\mathcal{G}_{2^{L+1}+1}$ and with law $\mathcal{N}(0, 2^{m(L)}\Lambda_{\widetilde{T},L})$. Hence, we have constructed Gaussian random variables $(V^*_{L,\ell})_{L\in\mathbb{N},\ell=1,\dots,2^{L-m(L)}}$ that are mutually independent. In addition, according to Lemma 2.11 of Dudley and Philipp (1983), there exists a Kiefer process $K^*_{\widetilde{T}}$ with covariance function $\Gamma_{\widetilde{T}}$ such that for any $L\in\mathbb{N}$, any $\ell\in\{1,\dots,2^{L-m(L)}\}$ and any $j\in\{1,\dots,2^{r(L)-1}\}$,

$$V^{*(j)}_{L,\ell}=K^*_{\widetilde{T}}\big(s_j,2^L+\ell 2^{m(L)}\big)-K^*_{\widetilde{T}}\big(s_j,2^L+(\ell-1)2^{m(L)}\big). \tag{4.69}$$

Thus, our construction is now complete.

Notice now that, by stationarity, for any $\ell\in\{1,\dots,2^{L-m(L)}\}$,

$$\mathbb{E}\big(d_{r(L)}\big(U^*_{L,\ell},V^*_{L,\ell}\big)\big)=\mathbb{E}\big(d_{r(L)}\big(U^*_{L,1},V^*_{L,1}\big)\big).$$

In addition, on the probability space $([0,1],\nu)$, the random variable $(T^{2^L+1}, T^{2^L+2},\dots,T^{2^{L+1}})$ is distributed as $(X_{2^{L+1}},X_{2^{L+1}-1},\dots,X_{2^L+1})$. Let $U^{(j)}_{L,\ell}=\sum_{i\in I_{L,\ell}}(\mathbf{1}_{Y_i\le s_j}-F_Y(s_j))$, and let $U_{L,\ell}$ be the associated column vectors in $\mathbb{R}^{2^{r(L)}-1}$ defined by $U_{L,\ell}=(U^{(1)}_{L,\ell},\dots,U^{(2^{r(L)}-1)}_{L,\ell})'$. According to the coupling relation (4.5), we get that

$$\begin{aligned}
&\mathbb{E}\big(W_{d_{r(L)}}(P_{U^*_{L,1}|\mathcal{G}_{2^L+2^{m(L)}+1}},G_{2^{m(L)}\Lambda_{T,L}})\big)\\
&\quad=\mathbb{E}\sup_{f\in\mathrm{Lip}(d_{r(L)})}\big(\mathbb{E}\big(f\big(U^*_{L,1}\big)|T^{2^L+\ell 2^{m(L)}+1}\big)-\mathbb{E}\big(f\big(V^*_{L,1}\big)\big)\big)\\
&\quad=\mathbb{E}\sup_{f\in\mathrm{Lip}(d_{r(L)})}\big(\mathbb{E}\big(f(U_{L,2^{L-m(L)}})|X_{2^{L+1}-2^{m(L)}}\big)-\mathbb{E}\big(f\big(V^*_{L,1}\big)\big)\big).
\end{aligned}\tag{4.70}$$

Let us construct the Gaussian random variables $V_{L,\ell}$ associated to the $U_{L,\ell}$ as in Section 4.1.1. Notice that since the covariance function $\Lambda_{\widetilde{T}}$ is the same as the covariance function Λ_Y defined by (4.4), for any measurable function f, $\mathbb{E}(f(V^*_{L,1}))=\mathbb{E}(f(V_{L,2^{L-m(L)}}))$. Therefore, starting from (4.68) and taking into account (4.70) together with (4.5), we get that

$$\begin{aligned}
&\mathbb{E}\big(d_{r(L)}\big(U^*_{L,1},V^*_{L,1}\big)\big)\\
&\quad=\mathbb{E}\sup_{f\in\mathrm{Lip}(d_{r(L)})}\big(\mathbb{E}\big(f(U_{L,2^{L-m(L)}})|\mathcal{F}_{2^{L+1}-2^{m(L)}}\big)-\mathbb{E}\big(f(V_{L,2^{L-m(L)}})\big)\big)\\
&\quad=\mathbb{E}\big(d_{r(L)}(U_{L,2^{L-m(L)}},V_{L,2^{L-m(L)}})\big).
\end{aligned}\tag{4.71}$$

Setting $\Pi_{r(L)}(s)=2^{-r(L)}[s2^{r(L)}]$ and mimicking the notation of Section 4.1.2, let now

$$\begin{aligned}
D^*_{L,1}={}&\sup_{2^L<\ell\le 2^{L+1}}\sup_{s\in[0,1]}\big|\big(R_{\widetilde{T}}(s,\ell)-R_{\widetilde{T}}\big(\Pi_{r(L)}(s),\ell\big)\big)\\
&\qquad-\big(R_{\widetilde{T}}\big(s,2^L\big)-R_{\widetilde{T}}\big(\Pi_{r(L)}(s),2^L\big)\big)\big|,\\
B^*_{L,3}={}&\sup_{j\in\{1,\dots,2^{r(L)}-1\}}\sup_{1\le k\le 2^{L-m(L)}}\sup_{\ell\in I_{L,k}}\big|R_{\widetilde{T}}(s_j,\ell)-R_{\widetilde{T}}\big(s_j,2^L+(k-1)2^{m(L)}\big)\big|,
\end{aligned}$$

and let $D_{L,1}$ and $B_{L,3}$ be the same quantities with R_Y replacing $R_{\widetilde{T}}$. Using once again that, on $([0,1],\nu)$, the random variable $(T^{2^L+1}, T^{2^L+2}, \dots, T^{2^{L+1}})$ is distributed as the random variable $(X_{2^{L+1}}, X_{2^{L+1}-1}, \dots, X_{2^L+1})$, we infer that for any positive λ,

$$\mathbb{P}(D^*_{L,1} \ge \lambda) \le \mathbb{P}(2D_{L,1} \ge \lambda) \quad \text{and} \quad \mathbb{P}(B^*_{L,3} \ge \lambda) \le \mathbb{P}(2B_{L,3} \ge \lambda). \tag{4.72}$$

Proceeding as in Section 4.1.2 of the proof of Theorem 2.1, using the fact that the covariance function $\Gamma_{\widetilde{T}}$ is the same as the covariance function Γ_Y defined by (4.4) (so that all the quantities involving only the Kiefer process $K^*_{\widetilde{T}}$ can be computed as in Section 4.1.2) and taking into account (4.71), (4.72) and the fact that the Markov chain $(X_i)_{i\in\mathbb{Z}}$ satisfies the assumptions of Theorem 2.1, Theorem 3.1 follows.

APPENDIX

A.1. Properties of the random variables Y_i. For the next lemma, we keep the same notation as that of Definition 2.1 and of the beginning of Section 4.1. Recall that the random variables Y_i have been defined in (4.3).

LEMMA A.1. *The following assertions hold*:

(1) *The image measure of* $\mathbb{P}^*_0$ *by the variable* Y_0 *is the uniform distribution over* $[0,1]$.

(2) *The equality* $F^{-1}_{P^*}(Y_i) = X_i$ *holds* $\mathbb{P}$-*almost surely. Moreover*, $\mathbb{P}$-*almost surely*,

$$\begin{aligned} b(X_0,k) &\ge \sup_{t\in\mathbb{R}} \bigl|P_{Y_k|X_0}(f_t) - P_Y(f_t)\bigr|, \\ b_1(\mathcal{F}_\ell,k) &\ge \sup_{t\in\mathbb{R}} \bigl|P_{Y_k|\mathcal{F}_\ell}(f_t) - P_Y(f_t)\bigr|, \\ b_2(\mathcal{F}_\ell,i,j) &\ge \sup_{(s,t)\in\mathbb{R}^2} \bigl|P_{(Y_i,Y_j)|\mathcal{F}_\ell}\bigl(f_t^{(0)} \otimes f_s^{(0)}\bigr) - P_{(Y_i,Y_j)}\bigl(f_t^{(0)} \otimes f_s^{(0)}\bigr)\bigr|. \end{aligned}$$

PROOF. As in Definition 2.1, define

$$b(X_i,k) = \sup_{t\in\mathbb{R}} \bigl|P_{X_k|X_i}(f_t) - P(f_t)\bigr|.$$

On Ω, we introduce the probability $\mathbb{P}^*_i$ whose density with respect to $\mathbb{P}$ is

$$\begin{aligned} &C(\beta)^{-1}\Biggl(1 + 4\sum_{k=i+1}^{\infty} b(X_i,k)\Biggr) \\ &\qquad \text{with } C(\beta) = 1 + 4\sum_{k=1}^{\infty} \beta\bigl(\sigma(X_0), X_k\bigr). \end{aligned} \tag{A.1}$$

By stationarity of $(X_i)_{i\in\mathbb{Z}}$, the image measure of $\mathbb{P}_i^*$ by X_i is again P^*. It follows from Lemma F.1, page 161, in Rio (2000) that the image measure of $\mathbb{P}_i^*$ by the variable Y_i is the uniform distribution over [0, 1] [proving item (1)], and that the equality $F_{P^*}^{-1}(Y_i) = X_i$ holds $\mathbb{P}_i^*$-almost surely. Since the probabilities $\mathbb{P}$ and $\mathbb{P}_i^*$ are equivalent, it follows that the equality $F_{P^*}^{-1}(Y_i) = X_i$ holds $\mathbb{P}$-almost surely, proving the first point of item (2).

Now, note that $Y_i = g(X_i, \eta_i)$, where the function $x \to g(x, u)$ is nondecreasing for any $u \in [0, 1]$. Since (X_0, X_k) is independant of η_k,

$$\begin{aligned}&\left|P_{Y_k|X_0}(f_t) - P_Y(f_t)\right| \\ &\quad = \left|\int_0^1 \{\mathbb{E}\big(f_t\big(g(X_k, u)\big)|X_0\big) - \mathbb{E}\big(f_t\big(g(X_k, u)\big)\big)\}\,du\right| \qquad \text{almost surely.}\end{aligned}$$

The function $x \to g(x, u)$ being nondecreasing, we infer that

$$\left|\mathbb{E}\big(f_t\big(g(X_k, u)\big)|X_0\big) - \mathbb{E}\big(f_t\big(g(X_k, u)\big)\big)\right| \le b(X_0, k) \qquad \text{almost surely,}$$

in such a way that

$$\left|P_{Y_k|X_0}(f_t) - P_Y(f_t)\right| \le b(X_0, k) \qquad \text{almost surely.}$$

The two last inequalities of item (2) may be proved in the same way. $\square$

A.2. Some upper bounds for partial derivatives. Let x and y be two column vectors of $\mathbb{R}^{2^{r(L)}-1}$ with coordinates

$$x = \big(\big(x^{(i,k_i)}, k_i \in \mathcal{E}(L, i)\big)_{i=0,\dots,r(L)-1}\big)'$$

and

$$y = \big(\big(y^{(i,k_i)}, k_i \in \mathcal{E}(L, i)\big)_{i=0,\dots,r(L)-1}\big)',$$

where $\mathcal{E}(L, i) = \{1, \dots, 2^{r(L)-i} - 1\} \cap (2\mathbb{N} + 1)$. Let $f \in \mathrm{Lip}(d^*_{r(L)})$, meaning that

$$\left|f(x) - f(y)\right| \le \sum_{K=0}^{r(L)-1} \sup_{k\in\mathcal{E}(L,K)} \left|x^{(K,k)} - y^{(K,k)}\right|$$

[the distance $d^*_{r(L)}$ is defined in Definition 4.3]. Let $a > 0$ and φ_a be the density of a centered Gaussian law of $\mathbb{R}^{2^{r(L)}-1}$ with covariance $a^2\mathrm{I}_{2^{r(L)}-1}$ ($\mathrm{I}_{2^{r(L)}-1}$ being the identity matrix on $\mathbb{R}^{2^{r(L)}-1}$). Let also

$$\|x\|_{\infty,L} = \sum_{K=0}^{r(L)-1} \sup_{k\in\mathcal{E}(L,K)} \left|x^{(K,k)}\right|$$

and

$$\|x\|_{2,L} = \left(\sum_{K=0}^{r(L)-1} \sum_{k_K\in\mathcal{E}(L,K)} \big(x^{(K,k_K)}\big)^2\right)^{1/2}.$$

For the statements of the lemmas, we refer to Notation 4.4.

LEMMA A.2. *The partial derivatives of f exist almost everywhere and the following inequality holds*:

$$\sup_{y\in\mathbb{R}^{2^{r(L)}-1}}\ \sup_{u\in\mathbb{R}^{2^{r(L)}-1},\|u\|_{\infty,L}\le 1} |Df(y).u| \le 1. \tag{A.2}$$

In addition,

$$\sup_{K\in\{0,\dots,r(L)-1\}} \sum_{k_K\in\mathcal{E}(L,K)} \left|\frac{\partial f}{\partial x^{(K,k_K)}}(y)\right| \le 1. \tag{A.3}$$

PROOF. The first part of the lemma follows directly from the fact that f is Lipschitz with respect to the distance $d^*_{r(L)}$ together with the Rademacher theorem. We prove now (A.3). For any $K\in\{0,\dots,r(L)-1\}$, we consider the column vector $u_K=((u_K^{(i,k_i)},k_i\in\mathcal{E}(L,i))_{i=0,\dots,r(L)-1})'$ with coordinates given by

$$u_K^{(i,k_i)}=\operatorname{sign}\biggl(\frac{\partial f}{\partial x^{(i,k_i)}}(y)\biggr)\mathbf{1}_{i=K}.$$

Applying inequality (A.2) together with the fact that $\|u_K\|_{\infty,L}=1$, we get that

$$\sum_{k\in\mathcal{E}(L,K)}\left|\frac{\partial f}{\partial x^{(K,k)}}(y)\right| = |Df(y).u_K|\le 1$$

and (A.3) follows. □

LEMMA A.3. *Let X and Y be two random variables in $\mathbb{R}^{2^{r(L)}-1}$. For any positive integer m and any $t\in[0,1]$*,

$$|\mathbb{E}(D^m f*\varphi_a(Y+tX).X^{\otimes m})|\le \mathbb{E}(\|Df(\cdot).X\|_\infty \times \|D^{m-1}\varphi_a(\cdot).X^{\otimes m-1}\|_1).$$

PROOF. For any positive integer m and any $x,y\in\mathbb{R}^{2^{r(L)}-1}$, it follows, from the properties of the convolution product, that

$$D^m f*\varphi_a(y).x^{\otimes m}=(Df(\cdot).x)*(D^{m-1}\varphi_a(\cdot).x^{\otimes m-1})(y),$$

where $Df(\cdot).x : y\mapsto Df(y).x$ and $D^{m-1}\varphi_a(\cdot).x^{\otimes m-1} : y\mapsto D^{m-1}\varphi_a(y).x^{\otimes m-1}$. The lemma then follows immediately. □

LEMMA A.4. *Let X be a random variable in $\mathbb{R}^{2^{r(L)}-1}$. For any nonnegative integer m, there exists a positive constant c_m depending only on m such that*

$$\|D^m\varphi_a(\cdot).X^{\otimes m}\|_1\le c_m a^{-m}\|X\|^m_{2,L}. \tag{A.4}$$

PROOF. In order to simplify the proof, and to avoid the double indexes (K,k_K) for the coordinates of a column vector of $\mathbb{R}^{2^{r(L)}-1}$, we set $d=2^{r(L)}-1$

and we denote by $x = (x_1, \dots, x_d)'$ an element of $\mathbb{R}^d$. Proceeding by induction on m, we infer that for any u, x in $\mathbb{R}^d$ and any integer m,

$$\begin{aligned} D^m \varphi_a(u).x^{\otimes m} &= \frac{1}{(2\pi a^2)^{d/2}} \\ &\quad \times \exp\left(-\frac{1}{2a^2}\sum_{i=1}^d u_i^2\right)\sum_{\ell=0}^{[m/2]} \frac{c_{m,\ell}}{a^{2\ell}}\left(\sum_{i=1}^d x_i^2\right)^{\ell}\left(\sum_{i=1}^d \frac{u_i x_i}{a^2}\right)^{m-2\ell} \end{aligned} \tag{A.5}$$

with the following recurrence relations between the $c_{m,\ell}$:

$$c_{m,0} = (-1)^m \qquad \text{for any } m \geq 0, \qquad c_{2,1} = -1,$$

$$c_{m+1,\ell} = (m - 2\ell + 2)c_{m,\ell-1} - c_{m,\ell}$$

for $\ell \in \{1, \dots, [m/2]\}$ and $m \geq 2$,

$$c_{m+1,[(m+1)/2]} = c_{m,[m/2]} \qquad \text{if } m \text{ is odd},$$

$$c_{m+1,[(m+1)/2]} = c_{m+1,[m/2]} \qquad \text{if } m \text{ is even}.$$

Starting from (A.5) and setting $\|x\|_{2,d} = (\sum_{i=1}^d x_i^2)^{1/2}$, we get that for any integer m,

$$\begin{aligned} &\int_{\mathbb{R}^d} |D^m \varphi_a(u).x^{\otimes m}|\,du \\ &\quad \leq \frac{\|x\|_{2,d}^m}{a^m (2\pi a^2)^{d/2}} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2a^2}\sum_{i=1}^d u_i^2\right)\sum_{\ell=0}^{m}\left|c_{m,\ell}\left(\sum_{i=1}^d \frac{u_i x_i}{a\|x\|_{2,d}}\right)^{m-2\ell}\right|\prod_{i=1}^d du_i \\ &\quad \leq \frac{\|x\|_{2,d}^m}{a^m} \int_{\mathbb{R}^d} \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^d u_i^2\right)\sum_{\ell=0}^{m}\left|c_{m,\ell}\left(\sum_{i=1}^d \frac{u_i x_i}{\|x\|_{2,d}}\right)^{m-2\ell}\right|\prod_{i=1}^d du_i. \end{aligned}$$

Now, for any integer k, we have that

$$\frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\sum_{i=1}^d u_i^2\right)\left|\sum_{i=1}^d \frac{u_i x_i}{\|x\|_{2,d}}\right|^k \prod_{i=1}^d du_i = \mathbb{E}(|N|^k),$$

where $N \sim \mathcal{N}(0, 1)$. Therefore,

$$\int_{\mathbb{R}^d} |D^m \varphi_a(u).x^{\otimes m}|\,du \leq a^{-m}\|x\|_{2,d}^m \sum_{\ell=0}^{[m/2]} |c_{m,\ell}|\mathbb{E}(|N|^{m-2\ell}),$$

which completes the proof of (A.4). □

LEMMA A.5. *Let X and Y be two random variables with values in $\mathbb{R}^{2^{r(L)}-1}$. For any positive integer m and any $t \in [0, 1]$, there exists a positive constant c_{m-1} depending only on m such that*

$$|\mathbb{E}(D^m f * \varphi_a(Y + tX).X^{\otimes m})| \leq c_{m-1} a^{1-m} \mathbb{E}(\|X\|_{\infty,L} \times \|X\|_{2,L}^{m-1}).$$

PROOF. Applying Lemmas A.3 and A.4 and using the fact that, by (A.2),

$$\|Df(\cdot).X\|_\infty = \|X\|_{\infty,L} \sup_{y\in\mathbb{R}^{2^{r(L)}-1}} \left|Df(y).\frac{X}{\|X\|_{\infty,L}}\right| \le \|X\|_{\infty,L},$$

the result follows. $\square$

LEMMA A.6. *For any $y \in \mathbb{R}^{2^{r(L)}-1}$ and any integer $m \ge 1$, there exists a positive constant c_m depending only on m such that*

$$\sup_{(K_i,k_{K_i}),i=1,\dots,m} \left|\frac{\partial^m f * \varphi_a}{\prod_{i=1}^m \partial x^{(K_i,k_{K_i})}}(y)\right| \le c_m a^{1-m},$$

where the supremum is taken over all the indexes $K_i \in \{0,\dots,r(L)-1\}$ and $k_{K_i} \in \mathcal{E}(L,K_i)$ for any $i = 1,\dots,m$.

PROOF. Notice first that by the properties of the convolution product,

$$\frac{\partial^m f * \varphi_a}{\prod_{i=1}^m \partial x^{(K_i,k_{K_i})}}(y) = \left(\frac{\partial f}{\partial x^{(K_1,k_{K_1})}} * \frac{\partial^{m-1}\varphi_a}{\prod_{i=2}^m \partial x^{(K_i,k_{K_i})}}\right)(y).$$

Therefore, by using (A.3),

$$\begin{aligned}\left|\frac{\partial^m f * \varphi_a}{\prod_{i=1}^m \partial x^{(K_i,k_{K_i})}}(y)\right| &\le \left\|\frac{\partial f}{\partial x^{(K_1,k_{K_1})}}\right\|_\infty \left\|\frac{\partial^{m-1}\varphi_a}{\prod_{i=2}^m \partial x^{(K_i,k_{K_i})}}\right\|_1 \\ &\le \left\|\frac{\partial^{m-1}\varphi_a}{\prod_{i=2}^m \partial x^{(K_i,k_{K_i})}}\right\|_1.\end{aligned} \tag{A.6}$$

Let now h_a be the density of the $\mathcal{N}(0,a^2)$ distribution, and let

$$\mathcal{S}_m = \left\{(\ell_1,\dots,\ell_m) \in \{0,\dots,m\}^{\otimes m} \text{ such that } \sum_{i=1}^m \ell_i = m\right\}.$$

With this notation, we infer that

$$\left\|\frac{\partial^{m-1}\varphi_a}{\prod_{i=2}^m \partial x^{(K_i,k_{K_i})}}\right\|_1 \le \sup_{(\ell_1,\dots,\ell_{m-1})\in\mathcal{S}_{m-1}} \prod_{i=1}^{m-1} \|h_a^{(\ell_i)}\|_1,$$

where $h_a^{(\ell_i)}$ is the ℓ_ith derivative of h_a. Since for any real u, $h_a^{(\ell_i)}(u) = a^{-(\ell_i+1)}h_1^{(\ell_i)}(u/a)$, it follows that $\|h_a^{(\ell_i)}\|_1 = a^{-\ell_i}\|h_1^{(\ell_i)}\|_1$. Therefore,

$$\left\|\frac{\partial^{m-1}\varphi_a}{\prod_{i=2}^m \partial x^{(K_i,k_{K_i})}}\right\|_1 \le a^{1-m} \sup_{(\ell_1,\dots,\ell_{m-1})\in\mathcal{S}_{m-1}} \prod_{i=1}^{m-1} \|h_1^{(\ell_i)}\|_1. \tag{A.7}$$

Starting from (A.6) and using (A.7), the lemma is proved, with

$$c_m = \sup_{(\ell_1,\dots,\ell_{m-1})\in\mathcal{S}_{m-1}} \prod_{i=1}^{m-1} \|h_1^{(\ell_i)}\|_1.$$

$\square$

REFERENCES

BERKES, I., HÖRMANN, S. and SCHAUER, J. (2009). Asymptotic results for the empirical process of stationary sequences. *Stochastic Process. Appl.* **119** 1298–1324. MR2508575

BERKES, I. and PHILIPP, W. (1977). An almost sure invariance principle for the empirical distribution function of mixing random variables. *Z. Wahrsch. Verw. Gebiete* **41** 115–137. MR0464344

BICKEL, P. J. and WICHURA, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42** 1656–1670. MR0383482

BOROVKOVA, S., BURTON, R. and DEHLING, H. (2001). Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. *Trans. Amer. Math. Soc.* **353** 4261–4318. MR1851171

CASTELLE, N. and LAURENT-BONVALOT, F. (1998). Strong approximations of bivariate uniform empirical processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **34** 425–480. MR1632841

DEDECKER, J. (2010). An empirical central limit theorem for intermittent maps. *Probab. Theory Related Fields* **148** 177–195. MR2653226

DEDECKER, J., GOUËZEL, S. and MERLEVÈDE, F. (2010). Some almost sure results for unbounded functions of intermittent maps and their associated Markov chains. *Ann. Inst. Henri Poincaré Probab. Stat.* **46** 796–821. MR2682267

DEDECKER, J. and MERLEVÈDE, F. (2010). On the almost sure invariance principle for stationary sequences of Hilbert-valued random variables. In *Dependence in Probability, Analysis and Number Theory* 157–175. Kendrick Press, Heber City, UT. MR2731073

DEDECKER, J., PRIEUR, C. and RAYNAUD DE FITTE, P. (2006). Parametrized Kantorovich-Rubinštein theorem and application to the coupling of random variables. In *Dependence in Probability and Statistics. Lecture Notes in Statistics* **187** 105–121. Springer, New York. MR2283252

DEDECKER, J. and PRIEUR, C. (2007). An empirical central limit theorem for dependent sequences. *Stochastic Process. Appl.* **117** 121–142. MR2287106

DEDECKER, J. and PRIEUR, C. (2009). Some unbounded functions of intermittent maps for which the central limit theorem holds. *ALEA Lat. Am. J. Probab. Math. Stat.* **5** 29–45. MR2475605

DEDECKER, J., DOUKHAN, P., LANG, G., LEÓN R., J. R., LOUHICHI, S. and PRIEUR, C. (2007). *Weak Dependence: With Examples and Applications. Lecture Notes in Statistics* **190**. Springer, New York. MR2338725

DEHLING, H. and TAQQU, M. S. (1989). The empirical process of some long-range dependent sequences with an application to U-statistics. *Ann. Statist.* **17** 1767–1783. MR1026312

DUDLEY, R. M. and PHILIPP, W. (1983). Invariance principles for sums of Banach space valued random elements and empirical processes. *Z. Wahrsch. Verw. Gebiete* **62** 509–552. MR0690575

FINKELSTEIN, H. (1971). The law of the iterated logarithm for empirical distributions. *Ann. Math. Statist.* **42** 607–615. MR0287600

GIRAITIS, L. and SURGAILIS, D. (2002). The reduction principle for the empirical process of a long memory linear process. In *Empirical Process Techniques for Dependent Data* 241–255. Birkhäuser, Boston, MA. MR1958784

HENNION, H. and HERVÉ, L. (2001). *Limit Theorems for Markov Chains and Stochastic Properties of Dynamical Systems by Quasi-Compactness. Lecture Notes in Math.* **1766**. Springer, Berlin. MR1862393

KIEFER, J. (1972). Skorohod embedding of multivariate rv's, and the sample df. *Z. Wahrsch. Verw. Gebiete* **24** 1–35. MR0341636

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131. MR0375412

LAI, T. L. (1974). Reproducing kernel Hilbert spaces and the law of the iterated logarithm for Gaussian processes. *Z. Wahrsch. Verw. Gebiete* **29** 7–19. MR0368121

LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes. Ergebnisse der Mathematik und Ihrer Grenzgebiete* (3) [*Results in Mathematics and Related Areas* (3)] **23**. Springer, Berlin. MR1102015

LIVERANI, C., SAUSSOL, B. and VAIENTI, S. (1999). A probabilistic approach to intermittency. *Ergodic Theory Dynam. Systems* **19** 671–685. MR1695915

MERLEVÈDE, F. and RIO, E. (2012). Strong approximation of partial sums under dependence conditions with application to dynamical systems. *Stochastic Process. Appl.* **122** 386–417. MR2860454

RIO, E. (2000). *Théorie Asymptotique des Processus Aléatoires Faiblement Dépendants*. *Mathématiques & Applications* (*Berlin*) [*Mathematics & Applications*] **31**. Springer, Berlin. MR2117923

ROSENBLATT, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA* **42** 43–47. MR0074711

RÜSCHENDORF, L. (1985). The Wasserstein distance and approximation theorems. *Z. Wahrsch. Verw. Gebiete* **70** 117–129. MR0795791

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York. MR0838963

WU, W. B. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.* **35** 2294–2320. MR2353389

WU, W. B. (2008). Empirical processes of stationary sequences. *Statist. Sinica* **18** 313–333. MR2384990

YOSHIHARA, K.-I. (1979). Note on an almost sure invariance principle for some empirical processes. *Yokohama Math. J.* **27** 105–110. MR0560618

YU, H. (1993). A Glivenko–Cantelli lemma and weak convergence for empirical processes of associated sequences. *Probab. Theory Related Fields* **95** 357–370. MR1213196

ZWEIMÜLLER, R. (1998). Ergodic structure and invariant densities of non-Markovian interval maps with indifferent fixed points. *Nonlinearity* **11** 1263–1276. MR1644385

J. DEDECKER
LABORATOIRE MAP5
UNIVERSITÉ PARIS DESCARTES
SORBONNE PARIS CITÉ
UMR 8145 CNRS
45 RUE DES SAINTS-PÈRES
F-75270 PARIS CEDEX 06
FRANCE
E-MAIL: jerome.dedecker@parisdescartes.fr

F. MERLEVÈDE
UNIVERSITÉ PARIS-EST
LAMA (UMR 8050)
UPEMLV, CNRS, UPEC
F-77454 MARNE-LA-VALLÉE
FRANCE
E-MAIL: florence.merlevede@univ-mlv.fr

E. RIO
LABORATOIRE DE MATHÉMATIQUES
UNIVERSITÉ DE VERSAILLES
UMR 8100 CNRS
BÂTIMENT FERMAT
45 AVENUE DES ETATS-UNIS
F-78035 VERSAILLES
FRANCE
E-MAIL: emmanuel.rio@uvsq.fr