Almost sure invariance principle for the Kantorovich distance between the empirical and the marginal distributions of strong mixing sequences

Jérôme Dedecker*, Florence Merlevède †

November 6, 2020

Abstract

We prove a strong invariance principle for the Kantorovich distance between the empirical distribution and the marginal distribution of stationary α-mixing sequences.

Running head. ASIP for the empirical W_1 distance.

Keywords. Empirical process, Wasserstein distance, Almost sure invariance principle, Compact law of the iterated logarithm, Bounded law of the iterated logarithm, Conditional Value at Risk

Mathematics Subject Classification (2010). 60F15, 60G10, 60B12

1 Introduction and notations

Let $(X_i)_{i\in\mathbb{Z}}$ be a strictly stationary sequence of real-valued random variables. Define the two σ-algebras $\mathcal{F}_0 = \sigma(X_i, i \leq 0)$ and $\mathcal{G}_k = \sigma(X_i, i \geq k)$, and recall that the strong mixing coefficients $(\alpha(k))_{k\geq 0}$ of Rosenblatt [12] are defined by

$$\alpha(k) = \sup_{A\in\mathcal{F}_0, B\in\mathcal{G}_k} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \, . \tag{1.1}$$

Let μ be the common distribution of the X_i's, and let

$$\mu_n = \frac{1}{n}\sum_{k=1}^{n} \delta_{X_k}$$

be the empirical measure based on $X_1, \ldots, X_n$. In this paper, we prove a strong invariance principle for the Kantorovich distance $W_1(\mu_n, \mu)$ between μ_n and μ under a condition on the mixing coefficients $\alpha(k)$. Recall that the Kantorovich distance (also called Wasserstein distance of order 1) between two probability measures μ and ν is defined by

$$W_1(\mu, \nu) = \inf_{\pi\in M(\mu,\nu)} \int |x - y|\pi(dx, dy) \, ,$$

where $M(\mu, \nu)$ is the set of probability measures on $\mathbb{R}^2$ with marginals μ and ν. We shall use the following well known representation for probabilities on the real line:

$$W_1(\mu, \nu) = \int |F_\mu(x) - F_\nu(x)| dx \, , \tag{1.2}$$

*Jérôme Dedecker, Université de Paris, CNRS, MAP5, UMR 8145, 45 rue des Saints-Pères, F-75006 Paris, France.
†Florence Merlevède, Université Gustave Eiffel, LAMA, UMR 8050 CNRS, F-77454 Marne-La-Vallée, France.

where F_μ is the cumulative distribution function of μ.

Let $H : t \to \mathbb{P}([X_0| > t)$ be the tail function of $|X_0|$. In the case where $(X_i)_{i\in\mathbb{Z}}$ is a sequence of independent and identically distributed (i.i.d.) random variables, del Barrio et al. [2] used the representation (1.2) and a general result of Jain [7] for Banach-valued random variables to prove a central limit theorem for $\sqrt{n}W_1(\mu_n, \mu)$. More precisely, they showed that $\sqrt{n}W_1(\mu_n, \mu)$ converges in distribution to the $\mathbb{L}_1(dt)$ norm of an $\mathbb{L}_1(dt)$-valued Gaussian random variable, provided that

$$\int_0^\infty \sqrt{H(t)}\, dt < \infty\,. \tag{1.3}$$

They also proved that $\sqrt{n}W_1(\mu_n, \mu)$ is stochastically bounded iff (1.3) holds, proving that this condition is necessary and sufficient for the weak convergence of $\sqrt{n}W_1(\mu_n, \mu)$.

Still in the i.i.d. case, we easily deduce from Chapters 8 and 10 in Ledoux and Talagrand [8] that: if (1.3) holds, then the sequence

$$\frac{\sqrt{n}}{\sqrt{2\log\log n}} W_1(\mu_n, \mu) \tag{1.4}$$

satisfies a compact law of the iterated logarithm.

For strongly mixing sequences in the sense of Rosenblatt [12], we proved in [6] the central limit theorem for $\sqrt{n}W_1(\mu_n, \mu)$ under the condition

$$\int_0^\infty \sqrt{\sum_{k=0}^{\infty} (\alpha(k) \wedge H(t))}\, dt < \infty \tag{1.5}$$

(where $a \wedge b$ means the minimum between two reals a and b), and we give sufficient conditions for (1.5) to hold. Note that, in [6], we used a weaker version of the α-mixing coefficients, that enables to deal with a large class of non-mixing processes in the sense of Rosenblatt [12].

In Section 2 of this paper, we prove a strong invariance principle for $W_1(\mu_n, \mu)$ under the condition (1.5). The compact law of the iterated logarithm for (1.4) easily follows from this strong invariance principle. In Section 3, we apply our general result to derive the almost sure rate of convergence of the empirical estimator of the Conditional Value at Risk ($CVaR$) for stationary α-mixing sequences.

In the rest of the paper, we shall use the following notation: for two sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ of positive reals, $a_n \ll b_n$ means there exists a positive constant C not depending on n such that $a_n \leq Cb_n$ for any $n \geq 1$.

2 Main result

Our main result is the following strong invariance principle for $W_1(\mu_n, \mu)$.

Theorem 2.1. *Assume that* (1.5) *is satisfied. Then, enlarging the probability space if necessary, there exists a sequence of i.i.d.* $\mathbb{L}_1(dt)$*-valued centered Gaussian random variables* $(Z_i)_{i\geq 1}$ *with covariance function defined as follows: for any* $f, g \in \mathbb{L}_\infty(dt)$*,*

$$\Gamma(f, g) = \mathrm{Cov}\left(\int f(t)Z_1(t)\, dt, \int g(t)Z_1(t)\, dt\right) = \sum_{k\in\mathbb{Z}} \iint f(t)g(s)\mathrm{Cov}(\mathbf{1}_{X_0\leq t}, \mathbf{1}_{X_k\leq s})\, ds\, dt\,, \tag{2.1}$$

and such that

$$nW_1(\mu_n, \mu) - \int \left|\sum_{k=1}^{n} Z_k(t)\right| dt = o(\sqrt{n \log\log n}) \quad \textit{almost surely.}$$

Remark 2.2. In [4], Cuny proved a strong invariance principle for $W_1(\mu_n,\mu)$. under the condition

$$\sum_{k=0}^{\infty} \frac{1}{\sqrt{k+1}} \int_0^{\infty} \sqrt{\alpha(k) \wedge H(t)}\; dt < \infty \tag{2.2}$$

(in fact, he proved the result for a weaker version of the α-mixing coefficient, the same as that used in [6] for the central limit theorem). It follows from Section 5 of [6], that the condition (1.5) is always less restrictive than (2.2).

As a consequence of Theorem 2.1, we get the compact law of the iterated logarithm. Let K be the unit ball of the reproducing kernel Hilbert space (RKHS) associated with Γ, and C be the image of K by the $\mathbb{L}_1(dt)$ norm. The following corollary holds:

Corollary 2.1. *Assume that* (1.5) *is satisfied. Then the sequence*

$$\frac{\sqrt{n}}{\sqrt{2\log\log n}} W_1(\mu_n,\mu)$$

is almost surely relatively compact, with limit set C.

The proof of Theorem 2.1 is based on two ingredients: a martingale approximation in $\mathbb{L}_1(dt)$, as in [6], and the following version of the bounded law of the iterated logarithm, which has an interest in itself.

Proposition 2.1. *Assume that* (1.5) *holds, and let*

$$V = \int_0^{\infty} \sqrt{\sum_{k=0}^{\infty} (\alpha(k) \wedge H(t))}\; dt\,. \tag{2.3}$$

Then, there exists a universal constant η *such that for any* $\varepsilon > 0$,

$$\sum_{n\geq 2} \frac{1}{n} \mathbb{P}\left(\max_{1\leq k\leq n} kW_1(\mu_k,\mu) > (\eta V + \varepsilon)\sqrt{n\log\log n}\right) < \infty\,. \tag{2.4}$$

Remark 2.3. *(The bivariate case).* Let $(X_i,Y_i)_{i\in\mathbb{Z}}$ be a stationary sequence of $\mathbb{R}^2$-valued random variables, and define the coefficients $\alpha(k)$ as in (1.1), with the two σ-algebras $\mathcal{F}_0 = \sigma(X_i,Y_i, i\leq 0)$ and $\mathcal{G}_k = \sigma(X_i,Y_i,i\geq k)$. Let μ_X (resp. μ_Y) be the common distribution of the X_i's (resp. the Y_i's), and let

$$\mu_{n,X} = \frac{1}{n}\sum_{k=1}^{n} \delta_{X_k} \quad \text{and} \quad \mu_{n,Y} = \frac{1}{n}\sum_{k=1}^{n} \delta_{Y_k}\,.$$

Combining the arguments in [3] and the proof of Theorem 2.1, one can prove the following strong invariance principle for $n\,(W_1(\mu_{n,X},\mu_{n,Y}) - W_1(\mu_X,\mu_Y))$.

Let φ be the continuous function from $\mathbb{L}_1(dt)$ to $\mathbb{R}$ defined by

$$\varphi(x) = \int \left(\text{sign}\{F_X(t)-F_Y(t)\}\, x(t)\mathbf{1}_{F_X(t)\neq F_Y(t)} + |x(t)|\mathbf{1}_{F_X(t)=F_Y(t)}\right) dt\,,$$

where F_X (resp. F_Y) is the cumulative distribution function of μ_X (resp. μ_Y). Assume that

$$\int_0^{\infty} \sqrt{\sum_{k=0}^{\infty} (\alpha(k) \wedge H_X(t))}\; dt < \infty \quad \text{and} \quad \int_0^{\infty} \sqrt{\sum_{k=0}^{\infty} (\alpha(k) \wedge H_Y(t))}\; dt < \infty\,.$$

Then, enlarging the probability space if necessary, there exists a sequence of i.i.d. $\mathbb{L}_1(dt)$-valued centered Gaussian random variables $(Z_i)_{i\geq 1}$ with covariance function given by: for any $f,g \in \mathbb{L}_\infty(dt)$,

$$\begin{aligned}\widetilde{\Gamma}(f,g) = \text{Cov}&\left(\int f(t)Z_1(t)\;dt, \int g(t)Z_1(t)\;dt\right)\\ &= \sum_{k\in\mathbb{Z}} \iint f(t)g(s)\text{Cov}(\mathbf{1}_{X_0\leq t} - \mathbf{1}_{Y_0\leq t}, \mathbf{1}_{X_k\leq s} - \mathbf{1}_{Y_k\leq s})\; ds\,dt\,,\end{aligned}$$

and such that

$$n\left(W_1(\mu_{n,X},\mu_{n,Y})-W_1(\mu_X,\mu_Y)\right)-\varphi\left(\sum_{k=1}^n Z_k\right)=o(\sqrt{n\log\log n})\quad\text{almost surely.}$$

3 Rates of convergence of the empirical estimator of the Conditional Value at Risk

The Conditional Value at Risk at level $u\in(0,1]$ of a real-valued integrable random variable X ($CVaR_u(X)$) is a "risk measure" (according to the definition of Acerbi and Tasche [1]), which is widely used in mathematical finance. It is sometimes called Expected Shortfall of Average Value at Risk. We refer to the paper [1] for a clear definition of that indicator, and for its relation with other well known measures, such as the Value at Risk, the Worst Conditional Expectation, the Tail Conditional Expectation, and so on. According to Acerbi and Tasche [1], $CVaR_u(X)$ can be expressed as

$$CVaR_u(X)=-\frac{1}{u}\int_0^u F_X^{-1}(x)dx\,,$$

where F_X is the cumulative distribution function of the variable X, and F_X^{-1} is its usual cadlag inverse: $F_X^{-1}(u)=\inf\{x\in\mathbb{R}:F_X(x)\geq u\}$.

Concerning the difference between the Conditional Value at Risk of two random variables X and Y, the following elementary inequality holds (see for instance [11]):

$$|CVaR_u(X)-CVaR_u(Y)|\leq\frac{1}{u}\int_0^1|F_X^{-1}(x)-F_Y^{-1}(x)|dx=\frac{1}{u}W_1(\mu_X,\mu_Y)\,,\tag{3.1}$$

where μ_X (resp. μ_Y) is the distribution of X (resp. Y).

Consider now the problem of estimating $CVaR_u(X)$ from the random variables $X_1,...,X_n$, where $(X_i)_{i\in\mathbb{Z}}$ is a stationary sequence of α-mixing random variables with common distribution $\mu=\mu_X$. A natural estimator is then

$$\widehat{CVaR}_{u,n}=-\frac{1}{u}\int_0^u F_n^{-1}(x)dx\,,$$

where F_n is the empirical distribution function based on $X_1,\ldots,X_n$. From (3.1), we get the upper bound

$$\left|CVaR_u(X)-\widehat{CVaR}_{u,n}\right|\leq\frac{1}{u}\int_0^1|F_X^{-1}(x)-F_n^{-1}(x)|dx=\frac{1}{u}W_1(\mu_n,\mu)\,,$$

From Corollary 2.1, we obtain the almost sure rate of convergence of $\widehat{CVaR}_{u,n}$: if (1.5) holds, then

$$\limsup_{n\to\infty}\frac{\sqrt{n}}{\sqrt{2\log\log n}}\left|CVaR_u(X)-\widehat{CVaR}_{u,n}\right|\leq\frac{\kappa(\Gamma)}{u}\quad\text{almost surely,}$$

where $\kappa(\Gamma)$ is the largest value of the compact set C of Corollary 2.1 (recall that the covariance function Γ is defined in (2.1)). It is well known (see for instance Section 8 in [8]) that the constant $\kappa(\Gamma)$ can be expressed as

$$\kappa(\Gamma)=\sup_{f:\|f\|_\infty\leq 1}\left(\mathrm{Var}\left(\int f(t)Z(t)dt\right)\right)^{1/2}\leq\left\|\int|Z(t)|dt\right\|_2\,,$$

where Z is an $\mathbb{L}_1(dt)$-valued centered random variable with covariance function Γ.

4 Proofs

4.1 Proof of Theorem 2.1

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be the underlying probability space. By a standard argument, one may assume that $X_i = X_0 \circ T$, where $T : \Omega \mapsto \Omega$ is a bijective, bi-measurable transformation, preserving the probability $\mathbb{P}$. Let also $\mathcal{F}_i = \sigma(X_k, k \leq i)$.

Let $Y_0(t) = \mathbf{1}_{X_0 \leq t} - F(t)$, and $Y_k(t) = Y_0(t) \circ T^k = \mathbf{1}_{X_k \leq t} - F(t)$. With these notations and the representation (1.2) one has that

$$nW_1(\mu_n, \mu) = \int \left| \sum_{k=1}^{n} Y_k(t) \right| dt \,. \tag{4.1}$$

From Section 4 in [6], we know that, if (1.5) holds, then

$$Y_0(t) = D_0(t) + A(t) - A(t) \circ T, \tag{4.2}$$

where D_0 is such that $\mathbb{E}(D_0(t)|\mathcal{F}_{-1}) = 0$ almost surely and $\int \|D_0(t)\|_2 \, dt < \infty$, and A is such that $\int \|A(t)\|_1 \, dt < \infty$. Moreover, the covariance operator of D_0 is exactly Γ: for any $f, g \in \mathbb{L}_\infty(dt)$,

$$\Gamma(f, g) = \mathrm{Cov}\left(\int f(t) D_0(t) \; dt, \int g(t) D_0(t) \; dt \right) \,. \tag{4.3}$$

Let $D_k(t) = D_0 \circ T^k$. From (4.2), it follows that

$$\sum_{k=1}^{n} Y_k = \sum_{k=1}^{n} D_k + A \circ T - A \circ T^n \,. \tag{4.4}$$

From [4, Proposition 3.3], we know that, enlarging the probability space if necessary, there exists a sequence of i.i.d. $\mathbb{L}_1(dt)$-valued centered Gaussian random variables $(Z_i)_{i \geq 1}$ with covariance function Γ such that

$$\int \left| \sum_{k=1}^{n} D_k(t) - \sum_{k=1}^{n} Z_k(t) \right| dt = o\left(\sqrt{n \log \log n} \right) \quad \text{almost surely.} \tag{4.5}$$

Hence, the result will follow from (4.1), (4.4) and (4.5) if we can prove that

$$\lim_{n \to \infty} \frac{1}{\sqrt{n \log \log n}} \int |A(t) \circ T^n| \; dt = 0 \quad \text{almost surely.} \tag{4.6}$$

To prove (4.6), we start by considering the integral over $[-M, M]^c$, for $M > 0$. Applying again [4, Proposition 3.3], we infer that

$$\limsup_{n \to \infty} \frac{1}{\sqrt{2n \log \log n}} \int_{[-M,M]^c} \left| \sum_{k=1}^{n} D_k(t) \right| dt \leq \int_{[-M,M]^c} \|D_0(t)\|_2 \, dt \quad \text{almost surely.} \tag{4.7}$$

Now, as will be clear from the proof, Proposition 2.1 also holds on the space $\mathbb{L}^1([-M, M]^c, dt)$, and implies that there exists a universal constant η such that, for any positive ε,

$$\limsup_{n \to \infty} \frac{1}{\sqrt{n \log \log n}} \int_{[-M,M]^c} \left| \sum_{k=1}^{n} Y_k(t) \right| dt \leq \varepsilon + \eta \int_{M}^{\infty} \sqrt{\sum_{k=0}^{\infty} \min\{\alpha(k), H(t)\}} \, dt \quad \text{almost surely.} \tag{4.8}$$

From (4.7) and (4.8), we infer that

$$\lim_{M \to \infty} \limsup_{n \to \infty} \frac{1}{\sqrt{n \log \log n}} \int_{[-M,M]^c} |A(t) \circ T^n| \; dt = 0 \quad \text{almost surely.}$$

Hence the proof of (4.6) will be complete if we prove that, for any $M > 0$,

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n \log\log n}} \int_{-M}^{M} |A(t) \circ T^n| \, dt = 0 \quad \text{almost surely.} \tag{4.9}$$

To prove (4.9), we work in the space $\mathbb{H} = \mathbb{L}_2([-M, M], dt)$, and we denote by $\|\cdot\|_{\mathbb{H}}$ and $\langle \cdot, \cdot \rangle$ the usual norm and scalar product on $\mathbb{H}$. Since $\mathbb{E}(\|D_0\|_{\mathbb{H}}^2) < \infty$, we know from [4] that $\sum_{k=1}^n D_k$ satisfies the compact law of the iterated logarithm in $\mathbb{H}$. Since $\sum_{k\geq 0} \alpha(k) < \infty$ and Y_0 is bounded in $\mathbb{H}$, we infer from [5] that $\sum_{k=1}^n Y_k$ satisfies also the compact law of the iterated logarithm in $\mathbb{H}$.

Now, arguing exactly as in the end of the proof of [5, Theorem 4], one has: for any f in $\mathbb{H}$

$$\lim_{n\to\infty} \frac{\langle f, A \circ T^n \rangle}{\sqrt{n \log\log n}} = 0 \quad \text{almost surely.} \tag{4.10}$$

Let $(e_i)_{i\geq 1}$ be a complete orthonormal basis of $\mathbb{H}$ and $P_N(f) = \sum_{k=1}^N \langle f, e_k \rangle e_k$ be the projection of f on the space spanned by the first N elements of the basis. From (4.10), we get that

$$\lim_{n\to\infty} \frac{P_N(A \circ T^n)}{\sqrt{n \log\log n}} = 0 \quad \text{almost surely.} \tag{4.11}$$

On another hand, applying again [4, Proposition 3.3] (as done in (4.7)), we get

$$\lim_{N\to\infty} \limsup_{n\to\infty} \frac{1}{\sqrt{n \log\log n}} \left\| (I - P_N) \left(\sum_{k=1}^n D_k \right) \right\|_{\mathbb{H}} = 0 \quad \text{almost surely,} \tag{4.12}$$

and applying [5, Theorem 4],

$$\lim_{N\to\infty} \limsup_{n\to\infty} \frac{1}{\sqrt{n \log\log n}} \left\| (I - P_N) \left(\sum_{k=1}^n Y_k \right) \right\|_{\mathbb{H}} = 0 \quad \text{almost surely.} \tag{4.13}$$

From (4.4), (4.12) and (4.13), we infer that

$$\lim_{N\to\infty} \limsup_{n\to\infty} \frac{\|(I - P_N) A \circ T^n\|_{\mathbb{H}}}{\sqrt{n \log\log n}} = 0 \quad \text{almost surely,}$$

which, together with (4.11), implies (4.9). The proof of Theorem 2.1 is complete. $\diamond$

4.2 Proof of Proposition 2.1

For any $n \in \mathbb{N}$, let us introduce the following notations:

$$R(u) = \min\{q \in \mathbb{N}^* : \alpha(q) \leq u\} Q(u) \quad \text{and} \quad R^{-1}(x) = \inf\{u \in [0, 1] : R(u) \leq x\}.$$

For a positive real a that will be specified later, let

$$m_n = a \sqrt{\frac{n}{\log \log n}}, \quad v_n = R^{-1}(m_n), \quad M_n = Q(v_n). \tag{4.14}$$

For any $M > 0$, let $g_M(y) = (y \wedge M) \vee (-M)$. For any integer i, define

$$X_i' = g_{M_n}(X_i) \text{ and } X_i'' = X_i - X_i'. \tag{4.15}$$

We first recall that, by the dual expression of $W_1(\mu_n, \mu)$,

$$n W_1(\mu_n, \mu) = \sup_{f \in \Lambda_1} \sum_{i=1}^n \left(f(X_i) - \mathbb{E}(f(X_i)) \right).$$

where Λ_1 is the set of Lipschitz functions such that $|f(x)-f(y)|\le |x-y|$. Hence,

$$nW_1(\mu_n,\mu)\le \sup_{f\in\Lambda_1}\sum_{i=1}^n \big(f(X_i')-\mathbb{E}(f(X_i'))\big)+\sup_{f\in\Lambda_1}\sum_{i=1}^n \big(f(X_i)-f(X_i')-\mathbb{E}(f(X_i)-f(X_i'))\big)\,.$$

Therefore, setting,

$$F_n'(t)=\frac{1}{n}\sum_{k=1}^n \mathbf{1}_{\{X_k'\le t\}}\quad\text{and}\quad F'(t)=\mathbb{P}(X_1'\le t)\,,$$

and noticing that

$$k\|F_k'-F'\|_1=\sup_{f\in\Lambda_1}\sum_{i=1}^k \big(f(X_i')-\mathbb{E}(f(X_i'))\big)\,,$$

we get

$$\max_{1\le k\le n} kW_1(\mu_k,\mu)\le \max_{1\le k\le n} k\|F_k'-F'\|_1+\sum_{i=1}^n (|X_i''|+\mathbb{E}(|X_i''|)\,. \tag{4.16}$$

Now, note that

$$\begin{aligned}\sum_{n\ge 2}\frac{1}{\sqrt{n\log\log n}}\mathbb{E}(|X_n''|)&\le \sum_{n\ge 2}\frac{1}{\sqrt{n\log\log n}}\int_0^{+\infty}\mathbb{P}\big(|X_0|\mathbf{1}_{|X_0|>Q(v_n)}>t\big)dt\\ &\le \sum_{n\ge 2}\frac{1}{\sqrt{n\log\log n}}\int_{Q(v_n)}^{+\infty}H(t)dt\le \sum_{n\ge 2}\frac{1}{\sqrt{n\log\log n}}\int_0^{v_n}Q(u)du\\ &\le \sum_{n\ge 2}\frac{1}{\sqrt{n\log\log n}}\int_0^1 Q(u)\mathbf{1}_{m_n\le R(u)}du\ll \int_0^1 R(u)Q(u)du\,.\end{aligned}$$

But, according to Propositions 5.1 and 5.2 in [6], condition (1.5) implies that

$$\int_0^1 R(u)Q(u)du<\infty\,. \tag{4.17}$$

Hence, to prove (2.4) it suffices to show that there exists an universal constant η such that for any $\varepsilon>0$,

$$\sum_{n\ge 2}\frac{1}{n}\mathbb{P}\left(\max_{1\le k\le n} k\|F_k'-F'\|_1>\eta V\sqrt{n\log\log n}\right)<\infty\,. \tag{4.18}$$

For this purpose, let

$$q_n=\min\{k\in\mathbb{N}^*\,:\,\alpha(k)\le v_n\}\wedge n\,. \tag{4.19}$$

Since R is right continuous, we have $R(R^{-1}(w))\le w$ for any w, hence

$$q_nM_n=R(v_n)=R(R^{-1}(m_n))\le m_n\,. \tag{4.20}$$

Assume first that $q_n=n$. Bounding $f(X_i')-\mathbb{E}(f(X_i'))$ by $2M_n$, we obtain

$$\max_{1\le k\le n} k\|F_k'-F'\|_1\le 2nM_n=2q_nM_n\le 2m_n\,. \tag{4.21}$$

Taking into account the definition of m_n, it follows that there exists n_0 depending on a, V and η, such that for any $n\ge n_0$, $8m_n\le \kappa V\sqrt{n\log\log n}$. This proves the proposition in the case where $q_n=n$.

From now on, we assume that $q_n<n$. Therefore $q_n=\min\{k\in\mathbb{N}^*\,:\,\alpha(k)\le v_n\}$ and then $\alpha(q_n)\le v_n$. For any integer i, define

$$U_i(t)=\sum_{k=(i-1)q_n+1}^{iq_n}\left(\mathbf{1}_{X_k'\le t}-\mathbb{E}\left(\mathbf{1}_{X_k'\le t}\right)\right)\,.$$

and notice that

$$\max_{1\le k\le n} k\|F'_k - F'\|_1 \le 2q_nM_n + \int_{-M_n}^{M_n} \max_{1\le j\le [n/q_n]} \Big|\sum_{i=1}^{j} U_i(t)\Big| \, dt\,.$$

Let $k_n = [n/q_n]$. For any t, applying Rio's coupling lemma (see [10, Lemma 5.2]) recursively, we can construct random variables $(U_i^*(t))_{1\le i\le k_n}$ such that

- $U_i^*(t)$ has the same distribution as U'_i for all $1\le i\le k_n$,

- the random variables $(U_{2i}^*(t))_{2\le 2i\le k_n}$ are independent, as well as the random variables $(U_{2i-1}^*(t))_{1\le 2i-1\le k_n}$,

- we can suitably control $\|U_i(t) - U_i^*(t)\|_1$ as follows: for any $i\ge 1$,

$$\|U_i(t) - U_i^*(t)\|_1 \le 4q_n\alpha(q_n)\,. \tag{4.22}$$

Substituting $U_i^*(t)$ to $U_i(t)$, we obtain

$$\max_{1\le k\le n} k\|F'_k - F'\|_1 \le 2q_nM_n + \max_{2\le 2j\le [n/q_n]} \Big|\sum_{i=1}^{j} U_{2i}^*(t)\Big| \\ + \max_{1\le 2j-1\le [n/q_n]} \Big|\sum_{i=1}^{j} U_{2i-1}^*(t)\Big| + \sum_{i=1}^{[n/q_n]} |U_i(t) - U_i^*(t)|\,. \tag{4.23}$$

Therefore, setting $\kappa = \eta/4$, for $n\ge n_0$,

$$\mathbb{P}\Big(\max_{1\le k\le n} k\|F'_k - F'\|_1 \ge 4V\kappa\sqrt{n\log\log n}\Big) \le I_1(n) + I_2(n) + I_3(n)\,, \tag{4.24}$$

where

$$I_1(n) = \mathbb{P}\Big(\int_{-M_n}^{M_n} \sum_{i=1}^{[n/q_n]} |U_i(t) - U_i^*(t)| \, dt \ge V\kappa\sqrt{n\log\log n}\Big)$$
$$I_2(n) = \mathbb{P}\Big(\int_{-M_n}^{M_n} \max_{2\le 2j\le [n/q_n]} \Big|\sum_{i=1}^{j} U_{2i}^*(t)\Big| \, dt \ge V\kappa\sqrt{n\log\log n}\Big)$$
$$I_3(n) = \mathbb{P}\Big(\int_{-M_n}^{M_n} \max_{1\le 2j-1\le [n/q_n]} \Big|\sum_{i=1}^{j} U_{2i-1}^*(t)\Big| \, dt \ge V\kappa\sqrt{n\log\log n}\Big)\,.$$

Using Markov's inequality and (4.22), we get

$$I_1(n) \ll \frac{n}{\sqrt{n\log\log n}} M_n\alpha(q_n) \ll \frac{n}{\sqrt{n\log\log n}} v_nQ(v_n) \ll \frac{n}{\sqrt{n\log\log n}} \int_0^{R^{-1}(m_n)} Q(u)du\,.$$

Hence, by (4.17),

$$\sum_{n\ge 2} \frac{1}{n} I_1(n) \ll \sum_{n\ge 2} \frac{1}{\sqrt{n\log\log n}} \int_0^{R^{-1}(m_n)} Q(u)du \ll \int_0^1 R(u)Q(u)du < \infty\,.$$

To handle now the term $I_2(n)$ (as well as $I_3(n)$) in the decomposition (4.24), we shall use again Markov's inequality but this time at the order $p\ge 2$. Hence for $p\ge 2$, taking into account the stationarity, we get

$$I_2(n) \le \frac{1}{(V\kappa)^p (n\log\log n)^{p/2}} \Big(\int_{-Q(v_n)}^{Q(v_n)} \Big\|\max_{2\le 2j\le [n/q_n]} \Big|\sum_{i=1}^{j} \tilde U_{2i}(t)\Big|\Big\|_p \, dt\Big)^p\,.$$

Applying Rosenthal's inequality (see for instance [9, Theorem 4.1]) and taking into account the stationarity, there exist two positive universal constants c_1 and c_2 not depending on p such that

$$\Big\|\max_{2\le 2j\le [n/q_n]}\Big|\sum_{i=1}^{j} U_{2i}^*(t)\Big|\Big\|_p^p \le c_1^p p^{p/2}(n/q_n)^{p/2}\|U_2(t)\|_2^p + c_2^p p^p (n/q_n)\|U_2(t)\|_p^p := J_1(t)+J_2(t)\,. \tag{4.25}$$

Using similar arguments as to handle the quantity $I_2(n)$ in the proof of [6, Proposition 3.4], we have

$$\begin{aligned}\int_{-Q(v_n)}^{Q(v_n)} \|U_2(t)\|_2 dt &= \int_{-Q(v_n)}^{Q(v_n)} \Big(\mathrm{Var}\Big(\sum_{i=1}^{q_n}\mathbf{1}_{\{X_i'\le t\}}\Big)\Big)^{1/2} dt\\ &\le 2\sqrt{2}\sqrt{q_n}\int_0^{Q(v_n)}\Big(\sum_{k=0}^{q_n-1}\alpha(k)\wedge H(t)\Big)^{1/2} dt \le 2V\sqrt{2q_n}\,. \end{aligned}\tag{4.26}$$

Hence

$$\sum_{n\ge 2}\frac{1}{n(V\kappa)^p(n\log\log n)^{p/2}}\Big(\int_{-Q(v_n)}^{Q(v_n)} J_1(t)^{1/p}dt\Big)^p \le \sum_{n\ge 2}\frac{(2\sqrt{2}c_1\sqrt{p})^p}{n\kappa^p(\log\log n)^{p/2}}\,.$$

Let now

$$p = p_n = \max\{c\log\log n, 2\},$$

where c will be specified later. Set $n_1 = \min\{n\ge 2 : c\log\log n \ge 2\}$. It follows that

$$\sum_{n\ge n_1}\frac{1}{n(V\kappa)^p(n\log\log n)^{p/2}}\Big(\int_{-Q(v_n)}^{Q(v_n)} J_1(t)^{1/p}dt\Big)^p \le \sum_{n\ge n_1}\frac{1}{n}\Big(\frac{2c_1\sqrt{2c}}{\kappa}\Big)^{c\log\log n}\,,$$

which is finite provided we take κ such that $\frac{2c_1\sqrt{2c}}{\kappa} = \alpha^{-1}$ with $\alpha>1$ and $c > (\log\alpha)^{-1}$.

On another hand, proceeding as in (4.26), we deduce that, for any $t>0$,

$$\begin{aligned}\|U_2(t)\|_p^p = \Big\|\sum_{i=1}^{q_n}\Big(\mathbf{1}_{\{X_i'\le t\}} - \mathbb{P}(X_i'\le t)\Big)\Big\|_p^p &\le q_n^{p-2}\Big\|\sum_{i=1}^{q_n}\Big(\mathbf{1}_{\{X_i'\le t\}} - \mathbb{P}(X_i'\le t)\Big)\Big\|_2^2\\ &\le 2q_n^{p-1}\sum_{k=0}^{q_n-1}(\alpha(k)\wedge H(t))\,.\end{aligned}$$

In addition

$$\begin{aligned}\int_0^{Q(v_n)}\Big(\sum_{k=0}^{q_n-1}\alpha(k)\wedge H(t)\Big)^{1/p}dt &= \int_0^{Q(v_n)}\Big(\int_0^{H(t)}(\alpha^{-1}(u)\wedge q_n)du\Big)^{1/p}dt\\ &\le \int_0^{Q(v_n)}\Big(v_nq_n + \int_{v_n}^{H(t)}(\alpha^{-1}(u)\wedge q_n)du\Big)^{1/p}dt\,.\end{aligned}$$

Note that $u<H(t) \iff t<Q(u)$. Consequently $u<H(t)$ implies that $Q^{-2}(u)<t^{-2}$. Hence

$$\begin{aligned}&\int_0^{Q(v_n)}\Big(\sum_{k=0}^{q_n-1}\alpha(k)\wedge H(t)\Big)^{1/p}dt\\ &\quad\le (v_nq_n)^{1/p}Q(v_n) + \int_0^{Q(v_n)}\Big(t^{-2}\int_{v_n}^{H(t)}(\alpha^{-1}(u)\wedge q_n)Q^2(u)du\Big)^{1/p}\\ &\quad\le (v_nq_n)^{1/p}Q(v_n) + \Big(\int_{v_n}^1 R(u)Q(u)du\Big)^{1/p}\int_0^{Q(v_n)} t^{-2/p}dt\\ &\quad\le (v_nq_n)^{1/p}Q(v_n) + \Big(\int_0^1 R(u)Q(u)du\Big)^{1/p} p(p-2)^{-1}Q(v_n)^{1-2/p}\,.\end{aligned}$$

Set $n_2 = \min\{n \geq 2 : c \log\log n \geq 4\}$. It follows that

$$\sum_{n \geq n_2} \frac{1}{n(V\kappa)^p (n \log\log n)^{p/2}} \left(\int_{-Q(v_n)}^{Q(v_n)} J_2(t)^{1/p} dt \right)^p$$
$$\leq 2 \sum_{n \geq n_2} \frac{(4c_2 p)^p}{(\kappa V)^p (n \log\log n)^{p/2}} q_n^{p-2} \left\{ v_n q_n Q^p(v_n) + 2^p Q(v_n)^{p-2} \int_0^1 R(u) Q(u) du \right\} .$$

Note that

$$v_n q_n Q^2(v_n) = v_n \alpha^{-1}(v_n) Q^2(v_n) \leq \int_0^1 R(u) Q(u) du \, .$$

Hence, since $q_n M_n \leq m_n$, we get

$$\sum_{n \geq n_2} \frac{1}{n(V\kappa)^p (n \log\log n)^{p/2}} \left(\int_{-Q(v_n)}^{Q(v_n)} J_2(t)^{1/p} dt \right)^p$$
$$\leq 4 \int_0^1 R(u) Q(u) du \sum_{n \geq n_2} \frac{(8c_2 p)^p}{(\kappa V)^p (n \log\log n)^{p/2}} m_n^{p-2}$$
$$\leq 4a^{-2} \int_0^1 R(u) Q(u) du \sum_{n \geq n_2} \left(\frac{8ac_2 c}{\kappa V} \right)^p \frac{\log\log n}{n} ,$$

which is finite by taking into account (4.17), and if we choose $a = (c_1 \kappa V)/(2c_2 \sqrt{2c})$. Indeed, in this case,

$$\frac{8ac_2 c}{\kappa V} = \frac{2c_1 \sqrt{2c}}{\kappa} \times \frac{2ac_2 \sqrt{2c}}{c_1 \kappa V} = \alpha^{-1} \, .$$

This ends the proof of the proposition. $\diamond$

References

[1] C. Acerbi and D. Tasche (2002), On the coherence of Expected Shortfall. *Journal of Banking and Finance* **26** 1487-1503.

[2] E. del Barrio, E. Giné and C. Matrán (1999), Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *Ann. Probab.* **27** 1009-1071.

[3] P. Berthet, J. Dedecker, and F. Merlevède, Central limit theorem and almost sure results for bivariate empirical W_1 distances. (2020) https://hal.archives-ouvertes.fr/hal-02881842

[4] C. Cuny (2017), Invariance principles under the Maxwell-Woodroofe condition in Banach spaces. *Ann. Probab.* **45** 15781611.

[5] J. Dedecker and F. Merlevède (2010), On the almost sure invariance principle for stationary sequences of Hilbert-valued random variables. *Dependence in probability, analysis and number theory*, 157175, Kendrick Press, Heber City, UT.

[6] J. Dedecker and F. Merlevède (2017), Behavior of the Wasserstein distance between the empirical and the marginal distributions of stationary α-dependent sequences. *Bernoulli* **23** 2083–2127.

[7] N. C. Jain (1977), Central limit theorems and related questions in Banach space. *Proceedings of Symposium in Pure and Applied Mathematics* **31** 55-65. Amer. Math. Soc. Providence, RI.

[8] M. Ledoux and M. Talagrand (1991), Probability in Banach spaces. Isoperimetry and processes. *Ergebnisse der Mathematik und ihrer Grenzgebiete* (3), **23** Springer-Verlag, Berlin, xii+ 480 pp.

[9] I. Pinelis (1994), Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.* **22** 1679–1706.

[10] E. Rio (2000), Théorie asymptotique des processus aléatoires faiblement dépendants. *Math. Appl.* **31** Berlin.

[11] E. Rio (2017), About the conditional value at risk of partial sums. *C. R. Math. Acad. Sci. Paris* **355** 1190-1195.

[12] M. Rosenblatt (1956), A central limit theorem and a strong mixing condition, *Proc. Nat. Acad. Sci. U.S.A.* **42** 43-47.