HAL
open science

Central limit theorem and almost sure results for bivariate empirical W 1 distances

Philippe Berthet, Jérôme Dedecker, Florence Merlevède

To cite this version:

Philippe Berthet, Jérôme Dedecker, Florence Merlevède. Central limit theorem and almost sure results for bivariate empirical W 1 distances. 2020. hal-02881842v1

HAL Id: hal-02881842

https://hal.science/hal-02881842v1

Preprint submitted on 26 Jun 2020 (v1), last revised 10 Mar 2022 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Central limit theorem and almost sure results for bivariate empirical W_1 distances

Philippe Berthet[*], Jérôme Dedecker[†], Florence Merlevède [‡]

Abstract

In this paper we study the behavior of the Wasserstein distance of order 1 (also called Kantorovich distance) between the two marginal empirical measures of a stationary sequence of bivariate random variables. We give sufficient conditions for the central limit theorem, the compact law of the iterated logarithm and the Maricinkiewicz-Zygmund strong law.

Keywords. Kantorovich distance, empirical measure, central limit theorem, law of the iterated logarithm, stationary sequences, Banach spaces

Mathematics Subject Classification (2010). 60F05, 60F15, 60G10, 60B12

1 Introduction

Let $(X_i, Y_i)_{i\in\mathbb{Z}}$ be a stationary and ergodic sequence of $\mathbb{R}^2$-valued random variables. Let F be the cumulative distribution function (cdf) of the X_i's, and let G be the cdf of the Y_i's. Let also, for any $t \in \mathbb{R}$,

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i\le t} \quad \text{and} \quad G_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{Y_i\le t}\,.$$

In this paper, we study the behavior of

$$W_1(F_n, G_n) - W_1(F, G)\,, \tag{1.1}$$

where $W_1(F, G)$ is the Wasserstein distance of order 1 (or Kantorovich distance) between the probabilities with cdfs F and G. Recall that W_1 is a minimal distance, defined by

$$W_1(F, G) = \inf_{\pi\in M(F,G)} \int |x - y|\pi(dx, dy)\,, \tag{1.2}$$

where $M(F, G)$ is the set of probability measures on $\mathbb{R}^2$ with marginal cdfs F and G. It is well known that $W_1(F, G)$ can also be written as

$$W_1(F, G) = \int_0^1 |F^{-1}(u) - G^{-1}(u)|du = \int_{\mathbb{R}} |F(t) - G(t)|dt\,, \tag{1.3}$$

where F^{-1} is the generalized inverse of F. For $p > 1$, the distance $W_p(F, G)$ (defined with the cost $|\cdot|^p$ instead of $|\cdot|$) is equal to the $\mathbb{L}^p([0, 1])$-distance between F^{-1} and G^{-1}, which generalizes the first equality in (1.3).

[*]Philippe Berthet, Université de Toulouse, Institut de Mathématiques, UMR 5219 CNRS, UPS IMT, F-31062 Toulouse Cedex 9, France

[†]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.

For independent and identically distributed (i.i.d.) random vectors (X_i, Y_i), the central limit theorem for $W_p^p(F_n, G_n) - W_p^p(F, G)$ (and other similar quantities, for a large class of cost functions) has been studied in the two recent papers [3] and [2], starting from the exact expression involving F^{-1}, G^{-1}, F_n^{-1} and G_n^{-1}. As a matter of fact, for the special case of W_1 it is easier to start from the second equality in (1.3), which gives an expression in terms of F_n, G_n, F and G (see (2.8) below). We shall see that, using a first order Taylor expansion (see relation (2.5) below), we are back to the study of partial sums in the (cotype 2) Banach space $\mathbb{L}_1(dt)$, up to some negligible residual terms. Following this strategy, we also prove a compact law of the iterated logarithm and a Marcinkiewicz-Zygmund strong law for the quantity (1.1). Moreover, following [9] and [7] we are able to extend all these results to the context of α-dependent sequences (as defined for instance in [11]) under quite sharp conditions.

Let us quote that all the results of this paper are new. Only the central limit theorem for i.i.d. random variables is considered in [2], but under stronger conditions than ours (see Remark 2.2 below). Moreover, all the results of Section 2 are consequences of the corresponding results of Section 3 (dependent case); for the sake of clarity, we prefer to give the complete proofs in the i.i.d case, and then show how they can be adapted to a dependent context.

To conclude this introduction, we wish to emphasize that this work is located at the intersection of two of Denis Bosq's important research fields: "Statistical methods for stochastic processes" (see for instance [4]), and "Stochastic processes with values in Banach spaces" (see for instance [5]).

2 The case of i.i.d. random variables

In this section, we assume that $(X_i, Y_i)_{1 \le i \le n}$ is a sequence of i.i.d. random variables with values in $\mathbb{R}^2$.

2.1 Central limit Theorem

Proposition 2.1. *Assume that*

$$\int_{\mathbb{R}} \sqrt{\mathrm{Var}(\mathbf{1}_{X_1 \le t} - \mathbf{1}_{Y_1 \le t})}\ dt < \infty\,. \tag{2.1}$$

Then

$$\sqrt{n}(W_1(F_n, G_n) - W_1(F, G)) \underset{n \to +\infty}{\xrightarrow{\mathcal{L}}} \int_{F>G} B(t)\ dt - \int_{F<G} B(t)\ dt + \int_{F=G} |B(t)|\ dt\,, \tag{2.2}$$

where B is a Gaussian random variable with values in $\mathbb{L}_1(dt)$ and covariance function defined as follows: for any $f, g \in \mathbb{L}_\infty(dt)$,

$$\begin{aligned}\Gamma(f, g) = \mathrm{Cov}\left(\int f(t)B(t)\ dt, \int g(t)B(t)\ dt\right)\\ = \iint f(t)g(s)\mathrm{Cov}(\mathbf{1}_{X_1 \le t} - \mathbf{1}_{Y_1 \le t}, \mathbf{1}_{X_1 \le s} - \mathbf{1}_{Y_1 \le s})\ ds\, dt\,.\end{aligned} \tag{2.3}$$

Remark 2.1. Note that (2.1) is satisfied if both

$$\int_0^\infty \sqrt{\mathbb{P}(|X_1| > t)}\ dt < \infty \quad \text{and} \quad \int_0^\infty \sqrt{\mathbb{P}(|Y_1| > t)}\ dt < \infty\,. \tag{2.4}$$

Note also that (2.1) and (2.4) are in fact equivalent if X_1 is independent of Y_1.

Now, the first condition in (2.4) implies the central limit theorem for $\sqrt{n}W_1(F_n, F)$ as proved in [1, Theorem 2.1]. In the same theorem, it is also proved that the sequence $\sqrt{n}W_1(F_n, F)$ is stochastically bounded if and only if the first condition in (2.4) is satisfied.

Remark 2.2. In Corollary 13 of [2], a central limit theorem is proved for $\sqrt{n}(W_1(F_n, G_n) - W_1(F, G))$ but under stronger conditions than (2.4). In particular, the assumptions in [2] imply that X_1 and Y_1 both have a positive density on $\mathbb{R}$, and that there exists a positive constant C such that, for any $t > 0$,

$$\mathbb{P}(|X_1| > t) \leq \frac{C}{t^6} \quad \text{and} \quad \mathbb{P}(|Y_1| > t) \leq \frac{C}{t^6}\,.$$

Proof of Proposition 2.1. The proof is based on the central limit theorem for random variables with values in $\mathbb{L}_1$ proved by Jain [12]. It follows from this theorem that

$$\sqrt{n}\,((F_n - G_n) - (F - G))$$

converges in distribution in $\mathbb{L}_1(dt)$ to B if and only if (2.1) holds.

We also need a preliminary decomposition:

$$\begin{aligned}|x+h| - |x| = h\mathbf{1}_{x+h\geq 0, x>0} - h\mathbf{1}_{x+h<0,x<0} + |h|\mathbf{1}_{x=0} \\ + (|x+h| - |x|)(\mathbf{1}_{x+h\geq 0,x<0} + \mathbf{1}_{x+h<0,x>0}),\end{aligned}$$

which implies that

$$\begin{aligned}|x+h| - |x| = h\mathbf{1}_{x>0} - h\mathbf{1}_{x<0} + |h|\mathbf{1}_{x=0} \\ - h\mathbf{1}_{x+h<0,x>0} + h\mathbf{1}_{x+h\geq 0,x<0} + (|x+h| - |x|)(\mathbf{1}_{x+h\geq 0,x<0} + \mathbf{1}_{x+h<0,x>0})\,,\end{aligned}$$

and finally

$$|x+h| - |x| = h\mathbf{1}_{x>0} - h\mathbf{1}_{x<0} + |h|\mathbf{1}_{x=0} + 2R(h,x)\,, \tag{2.5}$$

where $|R(h,x)| \leq |h|(\mathbf{1}_{x+h\geq 0,x<0} + \mathbf{1}_{x+h<0,x>0})$.

From (2.5) applied with $x = F - G$ and $h = (F_n - G_n) - (F - G)$, we deduce that

$$\begin{aligned}\sqrt{n}\,(|F_n - G_n| - |F - G|) = \text{sign}\{F - G\} \times \sqrt{n}\,((F_n - G_n) - (F - G))\,\mathbf{1}_{F\neq G} \\ + \sqrt{n}\,|(F_n - G_n) - (F - G)|\,\mathbf{1}_{F=G} + 2R_n\,,\end{aligned} \tag{2.6}$$

where

$$|R_n| \leq \sqrt{n}\,|(F_n - G_n) - (F - G)|\,(\mathbf{1}_{F>G,F_n<G_n} + \mathbf{1}_{F<G,F_n\geq G_n})\,. \tag{2.7}$$

Now, from (1.3) we have

$$\sqrt{n}(W_1(F_n, G_n) - W_1(F, G)) = \sqrt{n}\int_{\mathbb{R}} (|F_n(x) - G_n(x)| - |F(x) - G(x)|)\,dx\,. \tag{2.8}$$

Combining (2.8), (2.6) and Jain's result, we infer that (2.2) holds as soon as

$$\int_{\mathbb{R}} |R_n(t)|\,dt \quad \text{converges in probability to 0 as } n \to \infty. \tag{2.9}$$

To prove this, we first note that

$$\left\|\int_{\mathbb{R}} |R_n(t)|\,dt\right\|_1 = \int \|R_n(t)\|_1\,dt\,,$$

and

$$\|R_n(t)\|_1 \leq \|R_n(t)\|_2 \leq \sqrt{\text{Var}(\mathbf{1}_{X_1\leq t} - \mathbf{1}_{Y_1\leq t})}\,. \tag{2.10}$$

We infer from (2.1), (2.10), and the dominated convergence theorem that, if for any $t \in \mathbb{R}$, $\|R_n(t)\|_1$ converges to 0 as $n \to \infty$, then

$$\lim_{n\to\infty} \int \|R_n(t)\|_1\,dt = 0\,,$$

which implies (2.9).

Hence it remains to prove that $\|R_n(t)\|_1$ converges to 0 as $n \to \infty$. Let $M > 0$ and let $T_n(t) = \sqrt{n}\,|(F_n(t) - G_n(t)) - (F(t) - G(t))|$. By (2.7),

$$\|R_n(t)\|_1 \leq \mathbb{E}(T_n(t)\mathbf{1}_{T_n(t)>M}) + M\mathbb{P}\left(\{F(t) > G(t)\}, \{F_n(t) < G_n(t)\}\right) + M\mathbb{P}\left(\{F(t) < G(t)\}, \{F_n(t) \geq G_n(t)\}\right). \quad (2.11)$$

Since $(F_n(t), G_n(t))$ converges almost surely to $(F(t), G(t))$, we infer that

$$\lim_{n\to\infty} \mathbb{P}\left(\{F(t) > G(t)\}, \{F_n(t) < G_n(t)\}\right) + \mathbb{P}\left(\{F(t) < G(t)\}, \{F_n(t) \geq G_n(t)\}\right) = 0\,. \quad (2.12)$$

Moreover, by standard computations,

$$\mathbb{E}(T_n(t)\mathbf{1}_{T_n(t)>M}) \leq \frac{\|T_n(t)\|_2^2}{M} \leq \frac{1}{M}\,, \quad (2.13)$$

and consequently

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{E}(T_n(t)\mathbf{1}_{T_n(t)>M}) = 0\,. \quad (2.14)$$

From (2.11), (2.12) and (2.14), we infer that $\|R_n(t)\|_1$ converges to 0 as $n \to \infty$, which concludes the proof.

2.2 Compact law of the iterated logarithm

Under (2.1), one can also describe the almost sure behavior of

$$\frac{\sqrt{n}}{\sqrt{2\log\log n}}(W_1(F_n, G_n) - W_1(F, G)). \quad (2.15)$$

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

$$\varphi(x) = \int \left(\mathrm{sign}\{F(t) - G(t)\}\, x(t)\mathbf{1}_{F(t)\neq G(t)} + |x(t)|\mathbf{1}_{F(t)=G(t)}\right) dt\,. \quad (2.16)$$

From Sections 8 and 10 in [13] (see Theorem 10.12 in [13], since $\mathbb{L}_1(dt)$ is of cotype 2), we know that, under (2.1),

$$\frac{\sqrt{n}}{\sqrt{2\log\log n}}\left((F_n - G_n) - (F - G)\right)$$

satisfies the compact law of the iterated logarithm (LIL) in $\mathbb{L}_1(dt)$, with compact set K being the unit ball of the reproducing kernel Hilbert space (RKHS) associated with the covariance operator Γ defined in (2.3). Hence, starting from (2.6) and (2.8), one can prove the following result

Proposition 2.2. *Assume that* (2.1) *holds. Then the sequence defined in* (2.15) *is almost surely relatively compact, with limit set* $\varphi(K)$.

Remark 2.3. In fact, since the function φ satisfies $|\varphi(x) - \varphi(y)| \leq \|x - y\|_{\mathbb{L}_1}$, we also have a strong invariance principle, by applying a general result in [7]: enlarging the probability space if necessary, there exists a sequence of i.i.d. $\mathbb{L}_1(dt)$-valued Gaussian random variables $(Z_i)_{i\geq 1}$ with covariance function Γ such that

$$n\left(W_1(F_n, G_n) - W_1(F, G)\right) - \varphi\left(\sum_{k=1}^{n} Z_k\right) = o(\sqrt{n\log\log n}) \quad \text{almost surely.}$$

This implies the compact law of the iterated logarithm of Proposition 2.2. The same remark applies to Subsection 3.2.

Proof of Proposition 2.2. From the above considerations, it suffices to prove that

$$\limsup_{n\to\infty} \frac{1}{\sqrt{\log\log n}} \int_{\mathbb{R}} |R_n(t)|\ dt = 0 \text{ a.s.} \tag{2.17}$$

Let $\varepsilon > 0$, and note that

$$\begin{aligned}
&\frac{1}{\sqrt{\log\log n}} \int_{\mathbb{R}} |R_n(t)|\ dt \\
&\qquad \le \frac{\sqrt{n}}{\sqrt{\log\log n}} \int_{G<F\le G+\varepsilon} |(F_n(t)-G_n(t))-(F(t)-G(t))|\ dt \\
&\qquad\quad + \frac{\sqrt{n}}{\sqrt{\log\log n}} \int_{F<G\le F+\varepsilon} |(F_n(t)-G_n(t))-(F(t)-G(t))|\ dt \\
&\qquad\quad + \frac{\sqrt{n}}{\sqrt{\log\log n}} \int_{G+\varepsilon<F,F_n<G_n} |(F_n(t)-G_n(t))-(F(t)-G(t))|\ dt \\
&\qquad\quad + \frac{\sqrt{n}}{\sqrt{\log\log n}} \int_{F+\varepsilon<G,G_n\le F_n} |(F_n(t)-G_n(t))-(F(t)-G(t))|\ dt\,.
\end{aligned}$$

By the Glivenko-Cantelli Lemma, for almost all ω, F_n (resp. G_n) converges uniformly to F (resp. G). Hence, the two last terms on right hand are exactly 0 for almost all ω and $n \ge N(\varepsilon,\omega)$.

Now, from the bounded LIL in the space $\mathbb{L}_1(\{F < G \le F+\varepsilon\}, dt)$ (for instance, since the CLT holds, one can apply Theorem 8.11 in [13]), it follows that, almost surely

$$\begin{aligned}
\limsup_{n\to\infty} \frac{\sqrt{n}}{\sqrt{2\log\log n}} \int_{G<F\le G+\varepsilon} &|(F_n(t)-G_n(t))-(F(t)-G(t))|\ dt \\
&\le \int_{G<F\le G+\varepsilon} \sqrt{\mathrm{Var}(\mathbf{1}_{X_1\le t}-\mathbf{1}_{Y_1\le t})}\ dt\,.
\end{aligned} \tag{2.18}$$

Hence, since (2.1) holds, by the dominated convergence theorem,

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty} \frac{\sqrt{n}}{\sqrt{\log\log n}} \int_{G<F\le G+\varepsilon} |(F_n(t)-G_n(t))-(F(t)-G(t))|\ dt = 0 \quad \text{a.s.}$$

Of course, the same is true for the integral over the set $\{F < G \le F+\varepsilon\}$, and (2.17) follows.

2.3 Almost sure rates under lower order moments

We now consider the case where X_1 (or Y_1) is not square integrable, so that (2.4) does not hold. Starting from the elementary remark that

$$|W_1(F_n,G_n) - W_1(F,G)| \le W_1(F_n,F) + W_1(G_n,G), \tag{2.19}$$

it follows from Corollary 4.1 in [10] that:

Proposition 2.3. *Assume that $\|X_1\|_p < \infty$ and $\|Y_1\|_p < \infty$ for some $p \in [1,2)$. Then*

$$\lim_{n\to\infty} n^{(p-1)/p}\, |W_1(F_n,G_n) - W_1(F,G)| = 0, \textit{ almost surely.} \tag{2.20}$$

Remark 2.4. For $p = 1$, the result follows from the strong law of large number for integrable $\mathbb{L}_1(dt)$-valued random variables. For $p \in (1,2)$, it is proved in [10] that $n^{(p-1)/p} W_1(F_n,F)$ converges to 0 almost surely if and only if $\|X_1\|_p < \infty$. The only if part is a consequence of the classical Marcinkiewicz-Zygmund Theorem (see [14] and Remark 4.2 in [10]).

Remark 2.5. Alternatively, one can also apply the Marcinkiewicz-Zygmund strong law of large numbers for $\mathbb{L}_p$-valued martingales given in Proposition 3.2 of [6]. Since

$$|W_1(F_n, G_n) - W_1(F, G)| \leq \int_{\mathbb{R}} |(F_n - G_n)(t) - (F - G)(t)| \, dt \,,$$

it follows directly from Proposition 3.2 of [6] that (2.20) holds for $p \in (1, 2)$ as soon as

$$\int_{\mathbb{R}} \|(\mathbf{1}_{X_1 \leq t} - \mathbf{1}_{Y_1 \leq t}) - (F - G)(t)\|_p \, dt < \infty \,. \tag{2.21}$$

Note that (2.21) is the $\mathbb{L}_p$-version of (2.1), but it is not comparable to the conditions given in Proposition 2.3.

3 Extension to dependent sequences

Let $(X_i, Y_i)_{i \in \mathbb{Z}}$ be a stationary and ergodic sequence of $\mathbb{R}^2$-valued random variables. We use the same notations as before for the cdfs F_n, G_n, F and G. Let also $\mathcal{F}_0 = \sigma(X_i, Y_i, i \leq 0)$, $\alpha_{X,Y}(0) = 1$ and

$$\alpha_{X,Y}(k) = \sup_{t \in \mathbb{R}} \|\mathbb{E}(\mathbf{1}_{X_k \leq t} - \mathbf{1}_{Y_k \leq t} | \mathcal{F}_0) - (F(t) - G(t))\|_1 \text{ for any } k \geq 1.$$

Finally, set

$$V(t) = \mathrm{Var}(\mathbf{1}_{X_1 \leq t} - \mathbf{1}_{Y_1 \leq t}) \,.$$

3.1 Central limit Theorem

Proposition 3.1. *Assume that*

$$\int_0^\infty \sqrt{\sum_{k=0}^\infty (\alpha_{X,Y}(k) \wedge V(t))} \, dt < \infty \,. \tag{3.1}$$

Then

$$\sqrt{n}(W_1(F_n, G_n) - W_1(F, G)) \xrightarrow[n \to +\infty]{\mathcal{L}} \int_{F > G} B(t) \, dt - \int_{F < G} B(t) \, dt + \int_{F = G} |B(t)| \, dt \,,$$

for a Gaussian random variable B with values in $\mathbb{L}_1(dt)$ and covariance function: for any $f, g \in \mathbb{L}_\infty(dt)$,

$$\begin{aligned} \Gamma(f, g) &= \mathrm{Cov}\left(\int f(t) B(t) \, dt, \int g(t) B(t) \, dt \right) \\ &= \sum_{k \in \mathbb{Z}} \iint f(t) g(s) \mathrm{Cov}(\mathbf{1}_{X_0 \leq t} - \mathbf{1}_{Y_0 \leq t}, \mathbf{1}_{X_k \leq s} - \mathbf{1}_{Y_k \leq s}) \, ds \, dt \,. \end{aligned} \tag{3.2}$$

Remark 3.1. Note that (3.1) is satisfied if both

$$\int_0^\infty \sqrt{\sum_{k=0}^\infty (\alpha_{X,Y}(k) \wedge H_X(t))} \, dt < \infty \quad \text{and} \quad \int_0^\infty \sqrt{\sum_{k=0}^\infty (\alpha_{X,Y}(k) \wedge H_Y(t))} \, dt < \infty \,, \tag{3.3}$$

where

$$H_X(t) = \mathbb{P}(|X_1| > t) \quad \text{and} \quad H_Y(t) = \mathbb{P}(|Y_1| > t) \,.$$

Note that the first condition in (3.3) with α_X instead of $\alpha_{X,Y}$ (see (3.13) for the definition of α_X) implies the central limit Theorem for $\sqrt{n} W_1(F_n, F)$, as proved in [9]. We refer to Section 5 in [9] for sufficient conditions implying (3.3).

Remark 3.2. Note that the coefficient $\alpha_{X,Y}$ is weaker than the usual strong mixing coefficient (in the sense of Rosenblatt [17]) of the sequence $(X_i, Y_i)_{i\in\mathbb{Z}}$. It is also weaker than the α-dependence coefficient of $(X_i, Y_i)_{i\in\mathbb{Z}}$ defined in [11]. We then refer to the paper [11] for many examples of stationary processes for which $\alpha_{X,Y}$ can be computed.

Proof of Proposition 3.1. We follow the scheme of the proof of Proposition 2.1. To prove the convergence in distribution of $\sqrt{n}\,((F_n - G_n) - (F - G))$, we apply Corollary 6.1 in [9]. It suffices then to prove that

$$\int_0^\infty \sqrt{\sum_{k=0}^\infty \|\mathbb{E}\left(\mathbf{1}_{X_k\le t} - \mathbf{1}_{Y_k\le t}|\mathcal{F}_0\right) - (F(t) - G(t))\|_1}\; dt < \infty\,. \tag{3.4}$$

Let $S(t) = \mathbb{E}(|\mathbf{1}_{X_1\le t} - \mathbf{1}_{Y_1\le t} - (F(t) - G(t))|)$. By definition of $\alpha_{X,Y}$, we see that (3.1) is satisfied as soon as

$$\int_0^\infty \sqrt{\sum_{k=0}^\infty (\alpha_{X,Y}(k) \wedge S(t))}\; dt < \infty\,. \tag{3.5}$$

To see that (3.5) is equivalent to (3.1), one needs the following remark:

$$\mathbb{E}(|\mathbf{1}_{X_1\le t} - \mathbf{1}_{Y_1\le t} - (F(t) - G(t))|) \le 2V(t) \le 4\mathbb{E}(|\mathbf{1}_{X_1\le t} - \mathbf{1}_{Y_1\le t} - (F(t) - G(t))|)\,. \tag{3.6}$$

The first inequality in (3.6) is not completely obvious, and can be proved as follows. Let Z be a random variable taking values $0, 1$ or -1. Assume without loss of generality that $a = \mathbb{E}(Z) \ge 0$, and let $p_1 = \mathbb{P}(Z = 1)$ and $p_{-1} = \mathbb{P}(Z = -1)$. Clearly

$$\mathbb{E}(|Z - \mathbb{E}(Z)|) = p_1(1 - a) + (1 + a - 2p_1)a + (p_1 - a)(1 + a) = 2p_1(1 - a)\,,$$

and

$$\mathrm{Var}(Z) = p_1(1 - a)^2 + (1 + a - 2p_1)a^2 + (p_1 - a)(1 + a)^2 = 2p_1 - a(1 + a)\,.$$

Hence, it is enough to verify that $2p_1 - a(1 + a) \ge p_1(1 - a)$, which is clearly true since $p_1 \ge a = p_1 - p_{-1}$.

To see that the covariance function of the Gaussian process B can be expressed as in (3.2), it suffices to follow the proof of Proposition 2 in [9].

To conclude the proof, it remains to prove that $\|R_n(t)\|_1$ converges to 0 as $n \to \infty$ (recall that R_n has been defined in (2.6)). Clearly (2.11) and (2.12) hold, so that we only need to prove (2.14). By standard computations

$$\mathbb{E}(T_n(t)\mathbf{1}_{T_n(t)>M}) \le \frac{\|T_n(t)\|_2^2}{M} \le \frac{2}{M}\sum_{k=0}^\infty |\mathrm{Cov}(\mathbf{1}_{X_0\le t} - \mathbf{1}_{Y_0\le t}, \mathbf{1}_{X_k\le t} - \mathbf{1}_{Y_k\le t})|\,.$$

Clearly,

$$\begin{aligned}&|\mathrm{Cov}(\mathbf{1}_{X_0\le t} - \mathbf{1}_{Y_0\le t}, \mathbf{1}_{X_k\le t} - \mathbf{1}_{Y_k\le t})| \\ &\qquad\qquad \le \|\mathbb{E}(\mathbf{1}_{X_k\le t} - \mathbf{1}_{Y_k\le t}|\mathcal{F}_0) - (F(t) - G(t))\|_1 \le \alpha_{X,Y}(k)\end{aligned}$$

in such a way that

$$\mathbb{E}(T_n(t)\mathbf{1}_{T_n(t)>M}) \le \frac{\|T_n(t)\|_2^2}{M} \le \frac{2}{M}\sum_{k=0}^\infty \alpha_{X,Y}(k)\,. \tag{3.7}$$

Since (3.1) implies that $\sum_{k=0}^\infty \alpha_{X,Y}(k) < \infty$, we infer from (3.7) that (2.14) holds. This completes the proof of Proposition 3.1.

3.2 Compact law of the iterated logarithm

Proposition 3.2. *Assume that*

$$\sum_{k=0}^{\infty} \frac{1}{\sqrt{k+1}} \int_0^{\infty} \sqrt{\alpha_{X,Y}(k) \wedge V(t)}\ dt < \infty\,. \tag{3.8}$$

Then

$$\frac{\sqrt{n}}{\sqrt{2\log\log n}}\left((F_n - G_n) - (F - G)\right)$$

satisfies the compact LIL in $\mathbb{L}_1(dt)$*, with compact set* K *being the unit ball of the RKHS associated with the covariance operator* Γ *defined in* (3.2)*. Moreover the sequence*

$$\frac{\sqrt{n}}{\sqrt{2\log\log n}}(W_1(F_n, G_n) - W_1(F, G))$$

is almost surely relatively compact, with limit set $\varphi(K)$ *(where* φ *is the function defined in* (2.16)*).*

Remark 3.3. As in Section 5 of [9], one can prove that the condition (3.8) is slightly more restrictive than (3.1). For instance, if $\|X_1\|_\infty < \infty$ and $\|Y_1\|_\infty < \infty$ (bounded case), then (3.8) is equivalent to $\sum_{k>0} \sqrt{\alpha_{X,Y}(k)/k} < \infty$, while (3.1) is equivalent to $\sum_{k>0} \alpha_{X,Y}(k) < \infty$. Hence, a reasonable question is: does the conclusion of Proposition 3.2 still hold under (3.1)?

Note that, in the bounded case and for strongly mixing sequences in the sense of Rosenblatt [17], it follows from [15] that the conclusion of Proposition 3.2 still holds under the condition $\sum_{k>0} \alpha(k) < \infty$ (where $\alpha(k)$ is the usual strong mixing coefficient of the sequence $(X_i, Y_i)_{i\in\mathbb{Z}}$). Indeed, in the bounded case, the compact law of the iterated logarithm in $\mathbb{L}_2([-M, M], dt)$ can be applied, since the $\mathbb{L}_1$ norm is a continuous function for the $\mathbb{L}_2([-M, M], dt)$ topology.

Proof of Proposition 3.2. To prove the first part of the proposition, we apply Theorem 1.1 in [7], with $p = 1$. It suffices to prove that

$$\sum_{k=0}^{\infty} \frac{1}{\sqrt{k+1}} \int_0^{\infty} \left\|\mathbb{E}\left(\mathbf{1}_{X_k\le t} - \mathbf{1}_{Y_k\le t}|\mathcal{F}_0\right) - (F(t) - G(t))\right\|_2\ dt < \infty\,. \tag{3.9}$$

Using that

$$\begin{aligned}&\left\|\mathbb{E}\left(\mathbf{1}_{X_k\le t} - \mathbf{1}_{Y_k\le t}|\mathcal{F}_0\right) - (F(t) - G(t))\right\|_2 \\ &\qquad\qquad \le \left(\left\|\mathbb{E}\left(\mathbf{1}_{X_k\le t} - \mathbf{1}_{Y_k\le t}|\mathcal{F}_0\right) - (F(t) - G(t))\right\|_1\right)^{1/2}\,,\end{aligned}$$

we easily infer from the proof of Proposition 3.1 that (3.8) implies (3.9).

For the second part of Proposition 3.2, we follow the proof of Proposition 2.2. It suffices to prove that

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{\sqrt{n}}{\sqrt{\log\log n}} \int_{G<F\le G+\varepsilon} |(F_n(t) - G_n(t)) - (F(t) - G(t))|\ dt = 0 \quad \text{a.s.} \tag{3.10}$$

(and the same for the integral over the set $\{F < G \le F + \varepsilon\}$). Applying again Theorem 1.1 in [7] to the the space $\mathbb{L}_1(\{F < G \le F + \varepsilon\}, dt)$, we get that, almost surely,

$$\begin{aligned}&\limsup_{n\to\infty} \frac{\sqrt{n}}{\sqrt{2\log\log n}} \int_{G<F\le G+\varepsilon} |(F_n(t) - G_n(t)) - (F(t) - G(t))|\ dt \\ &\qquad\qquad \le C \sum_{k=0}^{\infty} \frac{1}{\sqrt{k+1}} \int_{G<F\le G+\varepsilon} \sqrt{\alpha(k) \wedge V(t)}\ dt\,,\end{aligned} \tag{3.11}$$

for some universal constant $C > 0$. Since (3.8) holds, (3.10) follows from (3.11) and the dominated convergence theorem.

3.3 Almost sure rates of convergence

In this sub-section, we do not assume that the stationary sequence $(X_i, Y_i)_{i\in\mathbb{Z}}$ is ergodic. Let $\mathcal{F}_{0,X} = \sigma(X_k, k \le 0)$ and $\mathcal{F}_{0,Y} = \sigma(Y_k, k \le 0)$, and define

$$\alpha_X(k) = \sup_{t\in\mathbb{R}} \|\mathbb{E}(\mathbf{1}_{X_k\le t}|\mathcal{F}_{0,X}) - F(t)\|_1 \ , \tag{3.12}$$

$$\alpha_Y(k) = \sup_{t\in\mathbb{R}} \|\mathbb{E}(\mathbf{1}_{Y_k\le t}|\mathcal{F}_{0,Y}) - G(t)\|_1 \ . \tag{3.13}$$

Proposition 3.3. *Let $p \in (1, 2)$. Let Q_X (resp. Q_Y) be the cadlag inverse of $t \to \mathbb{P}(|X_1| > t)$ (resp. $t \to \mathbb{P}(|Y_1| > t)$). Assume that*

$$\sum_{k=0}^{\infty} \frac{1}{(k+1)^{2-p}} \int_0^{\alpha_X(k)} Q_X^p(u)\ du < \infty \ \textit{and} \ \sum_{k=0}^{\infty} \frac{1}{(k+1)^{2-p}} \int_0^{\alpha_Y(k)} Q_Y^p(u)\ du < \infty\,. \tag{3.14}$$

Then

$$\lim_{n\to\infty} n^{(p-1)/p}|W_1(F_n, G_n) - W_1(F, G)| = 0 \ \textit{almost surely.} \tag{3.15}$$

Remark 3.4. If $(X_i, Y_i)_{i\in\mathbb{Z}}$ is ergodic, or if $\alpha_X(n) \to 0$ and $\alpha_Y(n) \to 0$ as $n \to \infty$, then (3.15) is true for $p = 1$ by the ergodic theorem. The first condition in (3.17) implies that

$$\lim_{n\to\infty} n^{-1/p} \sum_{i=1}^{n} (X_i - \mathbb{E}(X_i)) = 0 \text{ a.s.} \tag{3.16}$$

as proved in [8]. As will be clear form the proof, it also implies that $n^{(p-1)/p}W_1(F_n, F)$ converges to 0 almost surely as $n \to \infty$, which may be seen as a uniform version of (3.16) over the class of Lipschitz functions (thanks to the dual formulation of $W_1(F_n, F)$). The optimality of this condition with respect to (3.16) is studied in [16] (case of strongly mixing sequences in the sense of Rosenblatt [17]) and in [8].

Remark 3.5. In the same way as for Remark 2.5 of Section 2, one can also apply the Marcinkiewicz-Zygmund strong law of large numbers for $\mathbb{L}_p$-valued random variables given in Proposition 4.5 of [6] (see also Section 8.2 in [6] for an application to the empirical distribution function). It follows that, if

$$\sum_{k=0}^{\infty} \frac{1}{(k+1)^{1/p}} \int_0^{\infty} (\alpha_{X,Y}(k) \wedge V(t))^{1/p}\ dt < \infty\,, \tag{3.17}$$

for some $p \in (1, 2)$, then (3.15) holds. Note that (3.17) is the $\mathbb{L}_p$-version of (3.8), but it is not comparable to the conditions given in Proposition 3.3.

Proof of Proposition 3.3. Starting from the inequality (2.19), it suffices to prove that

$$\lim_{n\to\infty} n^{(p-1)/p}W_1(F_n, F) = 0, \quad \text{and} \quad \lim_{n\to\infty} n^{(p-1)/p}W_1(G_n, G) = 0 \text{ almost surely.} \tag{3.18}$$

Hence, it suffices to prove that the first condition of Proposition 3.3 implies the first statement in (3.18). To do so, we need a maximal version of Proposition 5.1 in [9]: let

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

For any positive integer n, any $x > 0$, and any $\eta \in [1, 2)$, the following inequality holds

$$\mathbb{P}\left(\max_{1\le k\le n} kW_1(F_k, F) \ge 6x\right) \le c_1 \frac{n}{x} \int_0^{R^{-1}(x)} Q_X(u)du + c_2 \frac{n}{x^\eta} \int_{R^{-1}(x)}^{1} R^{\eta-1}(u)Q_X(u)du\,, \tag{3.19}$$

where $c_1 = 36$ and $c_2 = 128(2-\eta)^{-2}$. This inequality is stated in [9] for $\mathbb{P}\left(nW_1(F_n, F) \geq 6x\right)$ instead of $\mathbb{P}\left(\max_{1\leq k\leq n} kW_1(F_k, F) \geq 6x\right)$. The maximal version can be stated by following exactly the proof in [9].

Let us now complete the proof of Proposition 3.3 with the help of Inequality (3.19). By the direct part of the Borel-Cantelli Lemma, it suffices to prove that, for any $\eta > 0$,

$$\sum_{n\geq 1} \mathbb{P}\left(\max_{1\leq k\leq 2^n} kW_1(F_k, F) \geq 2^{n/p}\eta\right) < \infty\,,$$

which is equivalent to: for any $\varepsilon > 0$,

$$\sum_{n\geq 1} \frac{1}{n}\mathbb{P}\left(\max_{1\leq k\leq n} kW_1(F_k, F) \geq 6n^{1/p}\varepsilon\right) < \infty\,. \tag{3.20}$$

Let $\eta \in (p, 2)$. Applying (3.19) with $x = n^{1/p}\varepsilon$, we get that

$$\mathbb{P}\left(\max_{1\leq k\leq n} kW_1(F_k, F) \geq 6n^{1/p}\varepsilon\right) \leq c_1 \frac{n^{(p-1)/p}}{\varepsilon}\int_0^{R^{-1}(n^{1/p}\varepsilon)} Q_X(u)du + c_2\frac{n^{(p-\eta)/p}}{\varepsilon^\eta}\int_{R^{-1}(n^{1/p}\varepsilon)}^1 R^{\eta-1}(u)Q_X(u)du\,.$$

Hence, it remains to prove that

$$\sum_{n\geq 1}\frac{1}{n^{1/p}}\int_0^{R^{-1}(n^{1/p}\varepsilon)} Q_X(u)du < \infty\,,\ \sum_{n\geq 1}\frac{1}{n^{\eta/p}}\int_{R^{-1}(n^{1/p}\varepsilon)}^1 R^{\eta-1}(u)Q_X(u)du < \infty\,. \tag{3.21}$$

By definition of $R^{-1}(u)$,

$$\int_0^{R^{-1}(n^{1/p}\varepsilon)} Q_X(u)du = \int_0^\infty \mathbf{1}_{n^{1/p}\varepsilon < R(u)} Q_X(u)du\,.$$

Hence, interverting the sum and the integral

$$\sum_{n\geq 1}\frac{1}{n^{1/p}}\int_0^{R^{-1}(n^{1/p}\varepsilon)} Q_X(u)du \leq \frac{K_1}{\varepsilon^{p-1}}\int_0^\infty R(u)^{p-1}Q_X(u)du = \frac{K_1}{\varepsilon^{p-1}}\int_0^\infty (\alpha^{-1}(u))^{p-1}Q_X^p(u)du\,, \tag{3.22}$$

for some positive constant K_1. The same kind of computations gives

$$\sum_{n\geq 1}\frac{1}{n^{\eta/p}}\int_{R_n^{-1}(n^{1/p}\varepsilon)}^1 R^{\eta-1}(u)Q_X(u)du \leq K_2\varepsilon^{\eta-p}\int_0^\infty (\alpha^{-1}(u))^{p-1}Q_X^p(u)du\,, \tag{3.23}$$

for some positive constant K_2. From (3.22) and (3.23), we infer that (3.21) is true provided that

$$\int_0^\infty (\alpha^{-1}(u))^{p-1}Q_X^p(u)\ du < \infty,$$

which is in fact equivalent to

$$\sum_{k=0}^\infty \frac{1}{(k+1)^{2-p}}\int_0^{\alpha_X(k)} Q_X^p(u)\ du < \infty\,,$$

as quoted for instance in [9]. This completes the proof of Proposition 3.3.

References

[1] 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.

[2] P. Berthet and J.-C. Fort, Weak convergence of empirical Wasserstein type distances. `https://hal.archives-ouvertes.fr/hal-01838700`

[3] P. Berthet, J.-C. Fort and T. Klein (2019), A central limit theorem for Wasserstein type distances between two different laws. To appear in *Ann. Institut H. Poincaré.*

[4] D. Bosq (1998), Nonparametric statistics for stochastic processes. Estimation and prediction. Second edition. *Lecture Notes in Statistics* **110**, Springer-Verlag, New York, xvi+210 pp.

[5] D. Bosq (2000), Linear processes in function spaces. Theory and applications. *Lecture Notes in Statistics* **149**, Springer-Verlag, New York, xiv+283 pp.

[6] C. Cuny (2014), Limit theorems under the Maxwell-Woodroofe condition in Banach spaces. `https://arxiv.org/pdf/1403.0772v1.pdf`

[7] C. Cuny (2017), Invariance principles under the Maxwell-Woodroofe condition in Banach spaces. *Ann. Probab.* **45**, 1578-1611.

[8] J. Dedecker and F. Merlevède (2007), Convergence rates in the law of large numbers for Banach-valued dependent variables. *Teor. Veroyatn. Primen.* **52**, 562-587

[9] 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.

[10] J. Dedecker and F. Merlevède (2019), Behavior of the empirical Wasserstein distance in $\mathbb{R}^d$ under moment conditions. *Electron. J. Probab.* **24** 32 pages.

[11] J. Dedecker and C. Prieur (2007), An empirical central limit theorem for dependent sequences, *Stochastic Process. Appl.* **117**, 121-142.

[12] 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.

[13] 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.

[14] J. Marcinkiewicz and A. Zygmund (1937), Sur les fonctions indépendantes. *Fund. Math.* **29**, 60-90.

[15] F. Merlevède (2008), On a maximal inequality for strongly mixing random variables in Hilbert spaces. Application to the compact law of the iterated logarithm. *Publ. Inst. Stat. Univ. Paris* **52** 47-60.

[16] E. Rio (1995), A maximal inequality and dependent Marcinkiewicz-Zygmund strong laws. *Ann. Probab* **23**, 918-937.

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