Behavior of the Wasserstein distance between the empirical and the marginal distributions of stationary α-dependent sequences

Jérôme Dedecker
Université Paris-Descartes

Joint work with F. Merlevède

CIRM, February 2016

1. Introduction

Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary sequence of integrable real-valued random variables, with common marginal distribution μ.
Let μ_n be the empirical measure of $\{X_1, \ldots, X_n\}$, that is

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

The Wasserstein distance of order 1 between μ_n and μ is

$$W_1(\mu_n, \mu) = \inf_{\pi\in M(\mu_n,\mu)} \int |x-y|\pi(dx, dy) \, , \qquad (1)$$

where $M(\mu_n, \mu)$ is the set of probability measures on $\mathbb{R}^2$ with marginal distributions μ_n and μ.

The quantity $W_1(\mu_n, \mu)$ appears very frequently in statistics, and can be understood from many points of view :

- The well known dual representation of W_1 implies that

$$W_1(\mu_n, \mu) = \sup_{f \in \Lambda_1} \left| \frac{1}{n} \sum_{k=1}^{n} (f(X_k) - \mu(f)) \right| , \qquad (2)$$

  where Λ_1 is the set of functions f such that $|f(x) - f(y)| \leq |x - y|$.
- In the one dimensional setting the minimization problem (1) can be explicitely solved, and leads to the expression

$$W_1(\mu_n, \mu) = \int_0^1 |F_n^{-1}(t) - F^{-1}(t)| dt \, , \qquad (3)$$

  where F_n and F are the distribution functions of μ_n and μ, and F_n^{-1} and F^{-1} are their usual generalized inverses.
- Starting from (3), it follows immediately that

$$W_1(\mu_n, \mu) = \int_{\mathbb{R}} |F_n(t) - F(t)| dt \, . \qquad (4)$$

Assume now that the sequence $(X_i)_{i\in\mathbb{Z}}$ is ergodic. Since μ has a finite first moment, it is well known that

$$\lim_{n\to\infty} W_1(\mu_n,\mu)=0 \ \ a.s. \quad \text{and} \quad \lim_{n\to\infty} \mathbb{E}(W_1(\mu_n,\mu))=0\,.$$

However, without additional assumptions on μ the rate of convergence can be arbitrarily slow.

In the i.i.d. case, del Barrio et al. (1999) proved that, if

$$\int_0^\infty \sqrt{H(t)}dt<\infty \quad \text{where} \quad H(t)=\mathbb{P}(|X_1|>t)\,, \tag{5}$$

then $\sqrt{n}W_1(\mu_n,\mu)$ converges in distribution to the random variable $\int |G(t)|\,dt$, where G is a Gaussian random variable in $\mathbb{L}^1(dt)$.

In fact (5) is necessary and sufficient for the stochastic boundedness of $\sqrt{n}W_1(\mu_n,\mu)$.

2. α-dependent sequences

Definition

For the stationary sequence $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$, let

$$\alpha_{1,\mathbf{X}}(n) = \sup_{x\in\mathbb{R}} \left\| \mathbb{E}\left(\mathbf{1}_{X_n\le x} | \mathcal{F}_0\right) - F(x) \right\|_1 , \tag{6}$$

where F is the distribution function of μ and $\mathcal{F}_0 = \sigma(X_i, i \le 0)$. Let also

$$\alpha_{2,\mathbf{X}}(n) = \sup_{x,y\in\mathbb{R}} \sup_{n\le i\le j} \Big\| \mathbb{E}\left((\mathbf{1}_{X_i\le x} - F(x)) \left(\mathbf{1}_{X_j\le y} - F(y)\right) \Big| \mathcal{F}_0\right) - \mathbb{E}\left((\mathbf{1}_{X_i\le x} - F(x)) \left(\mathbf{1}_{X_j\le y} - F(y)\right)\right) \Big\|_1 . \tag{7}$$

These coefficients are weaker than the α-mixing coefficients of Rosenblatt (1956).

3. Moments of order 1 and 2

For any $t \geq 0$, let

$$S_{\alpha,n}(t) = \sum_{k=0}^{n} \min\left\{\alpha_{1,\mathbf{X}}(k), H(t)\right\} . \quad (8)$$

The following upper bounds hold :

$$\mathbb{E}(W_1(\mu_n, \mu)) \leq 4 \int_0^{\infty} \sqrt{\min\left\{\left(H(t)\right)^2, \frac{S_{\alpha,n}(t)}{n}\right\}}\, dt\,, \quad (9)$$

and

$$\|W_1(\mu_n, \mu)\|_2 \leq \frac{2\sqrt{2}}{\sqrt{n}} \int_0^{\infty} \sqrt{S_{\alpha,n}(t)}\, dt\,. \quad (10)$$

For i.i.d. sequences (9) is the same (up to the numerical constant) as in Bobkov and Ledoux (2014).

4. Central limit theorem

Assume that the sequence is ergodic, and that

$$\int_0^\infty \sqrt{\sum_{k=0}^{\infty} \min\{\alpha_{1,\mathbf{X}}(k), H(t)\}}\; dt < \infty\,. \tag{11}$$

Then $\sqrt{n}W_1(\mu_n, \mu)$ converges in distribution to the random variable $\int |G(t)|\, dt$, where G is a Gaussian random variable in $\mathbb{L}^1(dt)$ whose covariance function may be described as follows :

for any f, g in $\mathbb{L}^\infty(\mu)$,

$$\mathrm{Cov}\left(\int f(t)G(t)dt, \int g(t)G(t)dt\right) = \sum_{k\in\mathbb{Z}} \mathbb{E}\left(\iint f(t)g(s)(\mathbf{1}_{X_0\le t} - F(t))(\mathbf{1}_{X_k\le s} - F(s))\, dtds\right)\,. \tag{12}$$

5. Moments of order $p \in (1,2)$

For $p \in (1,2)$, the following inequality holds

$$\|W_1(\mu_n,\mu)\|_p^p \leq \frac{C_p}{n^{p-1}} \sum_{k=0}^{n} \frac{1}{(k+1)^{2-p}} \int_0^{\alpha_{1,\mathbf{X}}(k)} Q^p(u)du\,. \qquad (13)$$

where Q is the generalized inverse of H.
In the i.i.d. case, Inequality (13) becomes

$$\|W_1(\mu_n,\mu)\|_p^p \leq \frac{C_p}{n^{p-1}} \|X_0\|_p^p\,. \qquad (14)$$

This last inequality seems to be new.

(14) is the same as the moment bound of order p for partial sums of i.i.d. random variables (cf. von Bahr and Esseen (1965)).

6. Moments of order $p > 2$

For $p > 2$, the following inequality holds :

$$\|W_1(\mu_n,\mu)\|_p^p \le C_p \left(\frac{s_{\alpha,n}^p}{n^{p/2}} + \frac{1}{n^{p-1}} \sum_{k=0}^{n} (k+1)^{p-2} \int_0^{\alpha_{2,\mathbf{X}}(k)} Q^p(u)du \right) . \tag{15}$$

where

$$s_{\alpha,n} = \int_0^\infty \sqrt{S_{\alpha,n}(t)}dt \quad \text{with } S_{\alpha,n} \text{ defined in (8).}$$

In the i.i.d. case, Inequality (15) becomes

$$\|W_1(\mu_n,\mu)\|_p^p \le C_p \left(\frac{1}{n^{p/2}} \left(\int_0^\infty \sqrt{H(t)}dt \right)^p + \frac{1}{n^{p-1}} \|X_0\|_p^p \right) .$$

This last inequality seems to be new.

Compared to the usual Rosenthal bound for sums of i.i.d. random variables, the variance term is replaced by the integral involving H.

7. Application to GPM maps

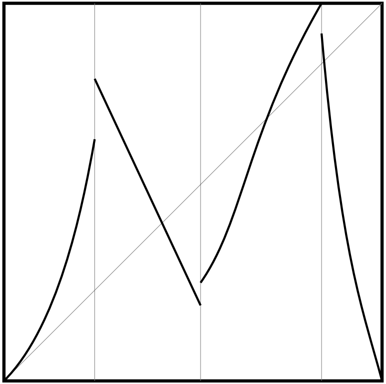

FIGURE: The graph of a GPM map, with $d = 4$

The map θ is uniformly expanding, except at 0, where there is a neutral fixed point, with $\theta(x) = x + cx^{1+\gamma}(1 + o(1))$ when $x \to 0$, for $\gamma \in (0, 1)$.

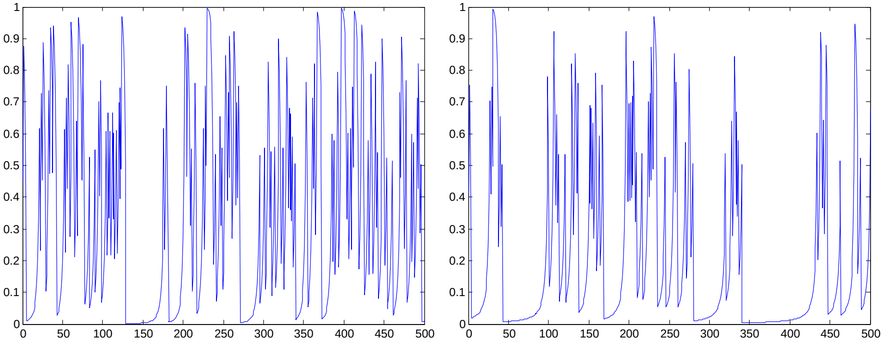

FIGURE: Time series of 500 iterations of a GPM map with $\gamma = 0.5$ (left) and $\gamma = 0.9$ (right).

The associated Markov Chain

- There exists an unique θ-invariant absolutely continuous probability measure ν.
- Define then the Perron-Frobenius operator K with respect to ν : for any f, g in $L^2(\nu)$

$$\nu(f \circ \theta \cdot g) = \nu(f \cdot K(g)),$$

  which means that $\mathbb{E}(g|\theta) = K(g)(\theta)$, so K is a transition kernel.
- Define then the Markov chain (X_i) with invariant measure ν and kernel K : on $([0,1], \nu)$ the n-tuple $(\theta, \theta^2, \ldots, \theta^n)$ is distributed as $(X_n, X_{n-1}, \ldots, X_1)$.
- From a previous work with S. Gouëzel and F. Merlevède (2010) : there exist $C > 0$ and $D > 0$ such that,

$$\frac{D}{n^{(1-\gamma)/\gamma}} \leq \alpha_{2,\mathbf{X}}(n) \leq \frac{C}{n^{(1-\gamma)/\gamma}} .$$

Central limit theorem

We shall illustrate each result by controlling, on the probability space $([0,1],\nu)$, the quantity $W_1(\tilde{\mu}_n,\mu)$, where

$$\tilde{\mu}_n = \frac{1}{n}\sum_{k=1}^{n} \delta_{g\circ\theta^k} \,,$$

θ is a GPM map, g is a non increasing function from $(0,1)$ to $\mathbb{R}$, and μ is the distribution of g.

For the CLT : assume that $\gamma \in (0,1/2)$. If g is positive and non increasing on (0, 1), with

$$g(x) \leq \frac{C}{x^{(1-2\gamma)/2}|\ln(x)|^b} \quad \text{near 0, for some } C>0 \text{ and } b>1,$$

then $\sqrt{n}W_1(\tilde{\mu}_n,\mu)$ converges in distribution to $\int |G(t)|dt$.

Rem. The usual CLT for $\sum_{k=1}^{n}(g\circ\theta^k - \nu(g))$ requires $b>1/2$.

Moments of order $p \in (1,2)$

Let $p \in (1,2)$, and let g be positive and non increasing on (0, 1), with

$$g(x) \leq \frac{C}{x^b} \quad \text{near 0, for some } C > 0 \text{ and } b \in [0, (1-\gamma)/p).$$

For $\gamma \in (0, 1/p]$, the following upper bounds hold.

$$\|W_1(\tilde{\mu}_n, \mu))\|_p \ll \begin{cases} n^{(1-p)/p} & \text{if } b < (1-p\gamma)/p \\ (n^{(1-p)} \ln(n))^{1/p} & \text{if } b = (1-p\gamma)/p \\ n^{(pb+\gamma-1)/p\gamma} & \text{if } b > (1-p\gamma)/p. \end{cases}$$

Moreover, if $b = (1-p\gamma)/p$,

$$\mathbb{P}\left(W_1(\mu_n, \mu) \geq x\right) \ll \frac{1}{n^{p-1}x^p}\,.$$

For $\gamma \in (1/p, 1)$, $\|W_1(\tilde{\mu}_n, \mu))\|_p \ll n^{(pb+\gamma-1)/p\gamma}$.

Moments of order $p > 2$

Let $p > 2$, and let g be positive and non increasing on (0, 1), with

$$g(x) \leq \frac{C}{x^b} \quad \text{near 0, for some } C > 0 \text{ and } b \in [0, (1-\gamma)/p).$$

For $\gamma \in (0, 1/2)$, the following upper bounds hold.

$$\|W_1(\tilde{\mu}_n, \mu))\|_p \ll \begin{cases} n^{-1/2} & \text{if } b \leq (2-\gamma(p+2))/2p \\ n^{(pb+\gamma-1)/p\gamma} & \text{if } b > (2-\gamma(p+2))/2p. \end{cases}$$

For $\gamma \in [1/2, 1)$,

$$\|W_1(\tilde{\mu}_n, \mu))\|_p \ll n^{(pb+\gamma-1)/p\gamma} .$$

Rem. In the bounded case ($b = 0$) all these bounds are optimal, see Gouëzel and Melbourne (2014) and a recent work with H. Dehling and M. Taqqu (2015).

8. About the proof of the CLT

Let $Y_k(t) = \mathbf{1}_{X_k \leq t} - F(t)$, and $S_n(t) = \sum_{k=1}^n Y_k(t)$. We want to prove that $\sqrt{n} S_n$ converges in $\mathbb{L}^1(dt)$ to a Gaussian random variable G.

We follow Gordin's approach (1971). Let $\mathbb{E}_i(\cdot)$ be the conditional expectation with respect to $\mathcal{F}_i = \sigma(X_j, j \leq i)$. Assume that

$$\sum_{k=1}^{\infty} \int \|\mathbb{E}_0(Y_k(t))\|_1 \, dt < \infty \,. \tag{16}$$

Let

$$D_i(t) = \sum_{k=i}^{\infty} \big(\mathbb{E}_i(Y_k(t)) - \mathbb{E}_{i-1}(Y_k(t))\big) \quad \text{and} \quad M_n(t) = \sum_{k=1}^{n} D_k(t) \,.$$

Then

$$\lim_{n \to \infty} \int \left\| \frac{S_n(t)}{\sqrt{n}} - \frac{M_n(t)}{\sqrt{n}} \right\|_1 dt = 0 \,. \tag{17}$$

It remains to prove the CLT in $\mathbb{L}^1(dt)$ for the martingale M_n. By de Acosta et al. (1978), it is enough to prove that

$$\int \|D_0(t)\|_2 \, dt < \infty \,. \tag{18}$$

Assume moreover that

$$C(t) = \liminf_{n\to\infty} \frac{1}{\sqrt{n}} \mathbb{E}(|S_n(t)|) < \infty \quad \text{and} \quad \int C(t) \, dt < \infty \,. \tag{19}$$

From (16), we also have that

$$\liminf_{n\to\infty} \frac{\|M_n(t)\|_1}{\sqrt{n}} = \liminf_{n\to\infty} \frac{\|S_n(t)\|_1}{\sqrt{n}} \,. \tag{20}$$

From (20) and (19), it follows that,

$$C(t) = \liminf_{n\to\infty} \frac{\|M_n(t)\|_1}{\sqrt{n}} < \infty \,.$$

Applying Theorem 1 in Esseen and Janson (1985), we deduce that,

$$\|D_0(t)\|_2 = \sqrt{\frac{\pi}{2}}C(t)\,. \tag{21}$$

From (21), we see that (19) implies (18), and the CLT for M_n follows.

It remains to check (16) and (19). The condition (16) follows easily from (11). Now,

$$C(t) \leq L(t) = \sqrt{\mathrm{Var}(Y_0(t)) + 2\sum_{k=1}^{\infty} |\mathrm{Cov}(Y_0(t), Y_k(t))|}\,,$$

and

$$L(t) \leq \sqrt{\sum_{k=0}^{\infty} 2\min\{\alpha_{1,\mathbf{X}}(k), B(t)\}}\,,$$

where $B(t) = F(t)(1 - F(t))$. This proves that (11) implies (19), and the proof is complete.

Bibliography

- A. de Acosta, A. Araujo, E. Giné (1978), On Poisson measures, Gaussian measures and the CLT in Banach spaces. Probability on Banach spaces.
- B. von Bahr, C.-G. Esseen (1965), Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$, Ann. Math. Statist.
- E. del Barrio, E. Giné, C. Matrán (1999), CLT for the Wasserstein distance between the empirical and the true distributions, Ann. Probab.
- S. Bobkov, M. Ledoux (2014), One-dimensional empirical measures, order statistics and Kantorovich transport distances. Preprint
- J. Dedecker, H. Dehling, M. S. Taqqu (2015), Weak conv. of the empirical process of intermittent maps in $\mathbb{L}^2$ under long-range dependence. S. Dyn.
- J. Dedecker, S. Gouëzel, F. Merlevède (2010), a.s. results for unbounded functions of intermittent maps and their associated Markov chains. AIHP.
- C.-G. Esseen, S. Janson (1985), On moment conditions for normed sums of independent variables and martingale differences, SPA.
- M. I. Gordin, International Conference on Probability Theory, Vilnius (1973).
- S. Gouëzel, I. Melbourne (2014), Moment bounds and concentration inequalities for slowly mixing dynamical systems. EJP.