Behavior of the empirical Wasserstein distance in $\mathbb{R}^d$ under moment conditions

Jérôme Dedecker[a] and Florence Merlevède[b]

Abstract

We establish some deviation inequalities, moment bounds and almost sure results for the Wasserstein distance of order $p \in [1, \infty)$ between the empirical measure of independent and identically distributed $\mathbb{R}^d$-valued random variables and the common distribution of the variables. We only assume the existence of a (strong or weak) moment of order rp for some $r > 1$, and we discuss the optimality of the bounds.

Mathematics subject classification. 60B10, 60F10, 60F15, 60E15.

Keywords. Empirical measure, Wasserstein distance, independent and identically distributed random variables, deviation inequalities, moment inequalities, almost sure rates of convergence.

[a] Université Paris Descartes, Sorbonne Paris Cité, Laboratoire MAP5 (UMR 8145).
Email: jerome.dedecker@parisdescartes.fr

[b] Université Paris-Est, LAMA (UMR 8050), UPEM, CNRS, UPEC.
Email: florence.merlevede@u-pem.fr

1 Introduction and notations

We begin with some notations, that will be used all along the paper. Let $X_1, \dots, X_n$ be n independent and identically distributed (i.i.d.) random variables with values in $\mathbb{R}^d$, with common distribution μ. Let μ_n be the empirical distribution of the X_i's, that is

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

Let X denote a random variable with distribution μ. For any $x \in \mathbb{R}^d$, let $|x| = \max\{|x_1|, \dots, |x_d|\}$. Define then the tail of the distribution μ by

$$H(t) = \mathbb{P}(|X| > t) = \mu(\{x \in \mathbb{R}^d \text{ such that } |x| > t\}) .$$

As usual, for any $q \geq 1$, the weak moment of order q of the random variable X is defined by

$$\|X\|_{q,w}^q := \sup_{t \in \mathbb{R}} t^q H(t)\,,$$

and the strong moment of order $q \geq 1$ is defined by

$$\|X\|_q^q = \mathbb{E}\left(|X|^q\right) = q \int_0^\infty t^{q-1} H(t) dt\,.$$

For $p \geq 1$, the Wasserstein distance between two probability measures ν_1, ν_2 on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is defined by

$$W_p^p(\nu_1, \nu_2) = \inf_{\pi \in \Pi(\nu_1, \nu_2)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|_2^p \; \pi(dx, dy)\,,$$

where $|\cdot|_2$ is the euclidean norm on $\mathbb{R}^d$ and $\Pi(\nu_1, \nu_2)$ is the set of probability measures on the product space $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^d))$ with margins ν_1 and ν_2.

In this paper, we prove deviation inequalities, moment inequalities and almost sure results for the quantity $W_p(\mu_n, \mu)$, when X has a weak or strong moment of order rp for $r > 1$. As in [13], the upper bounds will be different according as $p > d\min\{(r-1)/r, 1/2\}$ (small dimension case) or $p < d\min\{(r-1)/r, 1/2\}$ (large dimension case). Most of the proofs are based on Lemma 6 in [13] (see the inequality (6.4) in Section 6), which may be seen as an extension of Èbralidze's inequality [12] to the case $d > 1$. Hence we shall use the same approach as in [8], where we combined Èbralidze's inequality with truncation arguments to get moment bounds for $W_p(\mu_n, \mu)$ when $d = 1$.

There are many ways to see that the upper bounds obtained in the present paper are optimal in some sense, by considering the special cases $d = 1$, $p = 1$, $p = 2$, or by following the general discussion in [13], and we shall make some comments about this question all along the paper. However, the optimality for large d is only a kind of minimax optimality: one can see that the rates are exact for compactly supported measures which are not singular with respect to the Lebesgue measure on $\mathbb{R}^d$ (by using, for instance, Theorem 2 in [9]).

In fact, since the rates depend on the dimension d, it is easy to see that they cannot be optimal for all measures: for instance the rates will be faster as announced if the measure μ is supported on a linear subspace of $\mathbb{R}^d$ with dimension strictly less than d. This is of course not the end of the story, and the problem can be formulated in the general context of metric spaces (X, δ). For instance, for compactly supported measures, Boissard and Le Gouic [7] proved that the rates of convergence depend on the behavior of the metric entropy of the support of μ (with an extension to non-compact support in their Corollary 1.3). In the same context, Bach and Weed [2] obtain sharper results by generalizing some ideas going back to Dudley ([11], case $p = 1$). They introduce the notion of Wasserstein dimension $d_p^*(\mu)$ of the measure μ, and prove that $n^{p/s}\mathbb{E}(W_p^p(\mu_n, \mu))$ converges to 0 for any $s > d_p^*(\mu)$ (with sharp lower bounds in most cases).

Note that our context and that of Bach and Weed are clearly distincts: we consider measures on $\mathbb{R}^d$ having only a finite moment of order rp for $r > 1$, while they consider measures on compact metric spaces. However, the Wasserstein dimension is well defined for any probability measure (thanks to

Prohorov's theorem), and some arguments in [2] are common with [9] and [13]. A reasonable question is then: in the case of a singular measures on $\mathbb{R}^d$, are the results of the present paper still valid if we replace the dimension d by any $d' \in (d_p^*(\mu), d]$?

The paper is organized as follows: in Section 2 we state some deviations inequalities for $W_p(\mu_n, \mu)$ under weak moment assumptions. In Section 3 we bound up the probability of large and moderate deviations. In Section 4 we present some almost sure results, and in Section 5 we give some upper bounds for the moments of order r of $W_p(\mu_n, \mu)$ (von Bahr-Esseen and Rosenthal type bounds) under strong moment assumptions. The proofs are given in Section 6.

All along the paper, we shall use the notation $f(n, \mu, x) \ll g(n, \mu, x)$, which means that there exists a positive constant C, not depending on n, μ, x such that $f(n, \mu, x) \leq C g(n, \mu, x)$ for all positive integer n and all positive real x.

2 Deviation inequalities under weak moments conditions

In this section, we give some upper bound for the quantity $\mathbb{P}(W_p^p(\mu_n, \mu) > x)$ when the random variables X_i have a weak moment of order rp for some $r > 1$. We first consider the case where $r \in (1, 2)$.

Theorem 2.1. *If $\|X\|_{rp,w} < \infty$ for some $r \in (1, 2)$, then for any $x > 0$,*

$$\mathbb{P}(W_p^p(\mu_n, \mu) > x) \ll \begin{cases} \dfrac{\|X\|_{rp,w}^{rp}}{x^r n^{r-1}} & \text{if } p > d(r-1)/r \\ \dfrac{\|X\|_{rp,w}^{rp} (\log n)^r}{x^r n^{r-1}} \left(1 + \log_+ \left(\dfrac{x^{1/p} n^{r/(dr-d))}}{\|X\|_{rp,w}}\right)\right)^r & \text{if } p = d(r-1)/r \\ \dfrac{\|X\|_{rp,w}^{rp}}{x^r n^{rp/d}} & \text{if } p \in [1, d(r-1)/r) \end{cases}$$

where $\log_+(x) = \max\{0, \log x\}$.

Remark 2.1. As will be clear from the proof, the upper bounds of Theorem 2.1 still hold if the quantity $\mathbb{P}(W_p^p(\mu_n, \mu) > x)$ is replaced by its maximal version

$$\mathbb{P}\left(\max_{1 \leq k \leq n} k W_p^p(\mu_k, \mu) > nx\right).$$

Since $\|W_p^p(\mu_n, \mu)\|_1 \leq (r/(r-1)) \|W_p^p(\mu_n, \mu)\|_{r,w}$, according to the discussion after Theorem 1 in [13], if $p \neq d(r-1)/r$, one can always find some measure μ for which the rates of Theorem 2.1 are reached (see example (e) in [13] for $p > d(r-1)/r$ and example (c) in [13] for $p < d(r-1)/r$).

We now consider the case where $r > 2$. We follow the approach of Fournier and Guillin [13], but we use a different upper bound for the quantity controlled in their Lemma 13 (see the proof of Theorem 2.2 for more details).

Theorem 2.2. *If $\|X\|_{rp,w} < \infty$ for some $r \in (2,\infty)$, then, for any $x > 0$ and any $q > r$,*

$$\mathbb{P}(W_p^p(\mu_n,\mu) > x) \ll a\left(n, \frac{x}{\|X\|_{rp,w}^p}\right) + \frac{\|X\|_{rp,w}^{rp}}{x^r n^{r-1}} + \frac{1}{x^q n^{q/2}} \left(\int_0^\infty t^{p-1}\sqrt{H(t)}dt\right)^q ,$$

where

$$a(n,x) = C \begin{cases} \exp(-cnx^2)\mathbf{1}_{x\le A} & \text{if } p > d/2 \\ \exp(-cn(x/\log(2+x^{-1}))^2)\mathbf{1}_{x\le A} & \text{if } p = d/2 \\ \exp(-cnx^{d/p})\mathbf{1}_{x\le A} & \text{if } p \in [1, d/2) \end{cases}$$

for some positive constants C, c depending only on p, d, and a positive constant A depending only on p, d, r.

Remark 2.2. Let us compare our inequality with that of Theorem 2 of Fournier and Guillin [13] (under the moment condition (3) in [13]). We first note that the inequality in [13] is stated under a strong moment of order rp for $r > 2$, but their proof works also under a weak moment of order rp. Hence, under the assumptions of our Theorem 2.2, Fournier and Guillin obtained the bound (we assume here that $\|X\|_{rp,w} = 1$ for the sake of simplicity):

$$\mathbb{P}(W_p^p(\mu_n,\mu) > x) \ll a\,(n,x) + \frac{n}{(nx)^{(rp-\varepsilon)/p}}\,, \tag{2.1}$$

for any $\varepsilon > 0$ (the constant implicitly involved in the inequality depending on ε). In particular, one cannot infer from (2.1) that

$$\limsup_{n\to\infty} n^{r-1}\mathbb{P}\left(W_p^p(\mu_n,\mu) > x\right) \ll \frac{\|X\|_{rp,w}^{rp}}{x^r}\,,$$

which follows from our Theorem 2.2.

3 Large and moderate deviations

We consider here the probability of moderate deviations, that is

$$\mathbb{P}\left(W_p^p(\mu_n,\mu) > \frac{x}{n^{1-\alpha}}\right),$$

for $\alpha \le 1$ in a certain range. As usual, the case $\alpha = 1$ is the probability of large deviations.

As for partial sums, we shall establish two type of results, under weak moment conditions or under strong moment conditions. If the random variables have a weak moment of order rp for some $r > 1$, the results of Subsection 3.1 are immediate corollaries of the theorems of the preceding section. On the contrary, the Baum-Katz type results of Subsection 3.2 cannot be derived from the results of Section 2 and will be proved in Subsection 6.4.

3.1 Weak moments

As a consequence of Theorem 2.1, we obtain the following corollary.

Corollary 3.1. *If $\|X\|_{rp,w} < \infty$ for some $r \in (1,2)$, then for any $x > 0$,*

- *If $p > d(r-1)/r$ and $1/r \leq \alpha \leq 1$,*

$$\limsup_{n\to\infty} n^{\alpha r-1}\mathbb{P}\left(W_p^p(\mu_n,\mu) > \frac{x}{n^{1-\alpha}}\right) \ll \frac{\|X\|_{rp,w}^{rp}}{x^r}\,.$$

- *If $p = d(r-1)/r$ and $1/r < \alpha \leq 1$,*

$$\limsup_{n\to\infty} \frac{n^{\alpha r-1}}{(\log n)^{2r}}\mathbb{P}\left(W_p^p(\mu_n,\mu) > \frac{x}{n^{1-\alpha}}\right) \ll \frac{\|X\|_{rp,w}^{rp}}{x^r}\,.$$

- *If $p \in [1, d(r-1)/r)$ and $(d-p)/d \leq \alpha \leq 1$,*

$$\limsup_{n\to\infty} n^{(pr-(1-\alpha)rd)/d}\,\mathbb{P}\left(W_p^p(\mu_n,\mu) > \frac{x}{n^{1-\alpha}}\right) \ll \frac{\|X\|_{rp,w}^{rp}}{x^r}\,.$$

Remark 3.1. Let us comment on the case $p = 1, d = 1$. In that case, del Barrio et al. [4] proved that, for $\beta \in (1,2)$, $n^{(\beta-1)/\beta}W_1(\mu_n,\mu)$ is stochastically bounded if and only if $\|X\|_{\beta,w} < \infty$ (see their Theorem 2.2). This is consistent with the first inequality of Corollary 3.1 applied with $r = \beta$ and $\alpha = 1/r$.

Remark 3.2. Let us now comment on the case $p = 2, d = 1$. In that case del Barrio et al. [5] proved that, if the distribution function F of X is twice differentiable and if $F' \circ F^{-1}$ is a regularly varying function in the neighborhood of 0 and 1, then there exists a sequence of positive numbers v_n tending to ∞ as $n \to \infty$, such that $v_n W_2^2(\mu_n,\mu)$ converges in distribution to a non degenerate distribution. For instance, it follow from their Theorem 4.7 that, if X is a positive random variable, F is twice differentiable and $F(t) = (1 - t^{-\beta})$ for any $t > t_0$ and some $\beta > 2$, then $n^{(\beta-2)/\beta}W_2^2(\mu_n,\mu)$ converges in distribution to a non degenerate distribution. In that case, there is a weak moment of order β, and, for $\beta \in (2,4)$, the first inequality of Corollary 3.1 applied with $r = \beta/2$ and $\alpha = 1/r$ gives

$$\limsup_{n\to\infty} \mathbb{P}\left(n^{(\beta-2)/\beta}W_2^2(\mu_n,\mu) > x\right) \ll \frac{\|X\|_{\beta,w}^{\beta}}{x^{\beta/2}}\,.$$

Hence, in the case where $\beta \in (2,4)$, our result is consistent with that given in [5], and holds without assuming any regularity on F.

As a consequence of Theorem 2.2, we obtain the following corollary.

Corollary 3.2. *If $\|X\|_{rp,w} < \infty$ for some $r \in (2,\infty)$, then, for any $x > 0$ and any*

$$\alpha \in \left(\max\left(\frac{1}{2}, \frac{d-p}{d}\right), 1\right]\,,$$

$$\limsup_{n\to\infty} n^{\alpha r-1}\mathbb{P}\left(W_p^p(\mu_n,\mu) > \frac{x}{n^{1-\alpha}}\right) \ll \frac{\|X\|_{rp,w}^{rp}}{x^r}\,.$$

3.2 Baum-Katz type results

We first consider the case where the variables have a strong moment of order rp for $r \in (1,2)$.

Theorem 3.3. *If $\|X\|_{rp} < \infty$ for some $r \in (1,2)$, then for any $x > 0$,*

- *If $p > d(r-1)/r$ and $1/r \leq \alpha \leq 1$,*

$$\sum_{n=1}^{\infty} n^{\alpha r - 2} \mathbb{P}\left(\max_{1 \leq k \leq n} k W_p^p(\mu_k, \mu) > n^{\alpha} x \right) < \infty \, .$$

- *If $p \in [1, d(r-1)/r)$ and $\alpha \in ((d-p)/d, 1]$,*

$$\sum_{n=1}^{\infty} n^{(pr-(1-\alpha)rd-d)/d} \, \mathbb{P}\left(\max_{1 \leq k \leq n} k W_p^p(\mu_k, \mu) > n^{\alpha} x \right) < \infty \, .$$

- *If $p \in [1, d(r-1)/r)$,*

$$\sum_{n=1}^{\infty} \frac{1}{n} \, \mathbb{P}\left(\max_{1 \leq k \leq n} k W_p^p(\mu_k, \mu) > n^{(d-p)/d} (\log n)^{1/r} x \right) < \infty \, .$$

Remark 3.3. Our proof does not allow to deal with the case where $p = d(r-1)/r$. As an interesting consequence of Theorem 3.3, we shall obtain almost sure convergence rates for the sequence $W_p^p(\mu_n, \mu)$ (see Corollary 4.1 of the next section).

We now consider the case where the variables have a strong moment of order rp for $r > 2$.

Theorem 3.4. *If $\|X\|_{rp} < \infty$ for some $r \in (2, \infty)$, then, for any*

$$\alpha \in \left(\max\left(\frac{1}{2}, \frac{d-p}{d} \right), 1 \right] ,$$

$$\sum_{n=1}^{\infty} n^{\alpha r - 2} \mathbb{P}\left(\max_{1 \leq k \leq n} k W_p^p(\mu_k, \mu) > n^{\alpha} x \right) < \infty \, .$$

4 Almost sure results

Using well known arguments, we derive from Theorem 3.3 the following almost sure rates of convergence for the sequence $W_p^p(\mu_n, \mu)$ (taking $\alpha = 1/r$ in the case where $p > d(r-1)/r$, and applying the third item in the case where $p > d(r-1)/r$).

Corollary 4.1. *If $\|X\|_{rp} < \infty$ for some $r \in (1,2)$, then*

- *If $p > d(r-1)/r$,*

$$\lim_{n \to \infty} n^{(r-1)/r} W_p^p(\mu_n, \mu) = 0 \ \ a.s.$$

- *If* $p \in [1, d(r-1)/r)$,

$$\lim_{n\to\infty} \frac{n^{p/d}}{(\log n)^{1/r}} W_p^p(\mu_n, \mu) = 0 \ \ a.s.$$

Remark 4.1. Let us comment on these almost sure results in the case where $p = 1$ and $d < r/(r-1)$. Starting from the dual definition of $W_1(\mu_n, \mu)$, we get that

$$W_1(\mu_n, \mu) \geq \frac{1}{n} \left| \sum_{k=1}^{n} \left(|X_k|_2 - \mathbb{E}(|X_k|_2) \right) \right| . \tag{4.1}$$

Now, by the classical Marcinkiewicz-Zygmund theorem (see [15]) for i.i.d. random variables, we know that

$$\lim_{n\to\infty} \frac{n^{(r-1)/r}}{n} \left| \sum_{k=1}^{n} \left(|X_k|_2 - \mathbb{E}(|X_k|_2) \right) \right| = 0 \ \text{ a.s.}$$

if and only if $\|X\|_r < \infty$. It follows that, for $p = 1$, the rates of Corollary 4.1 are optimal is the case where $d < r/(r-1)$.

We now give some almost sure rates of convergence in the case where $\int_0^\infty t^{p-1}\sqrt{H(t)}dt < \infty$. Note that this condition is a bit more restrictive than $\|X\|_{2p} < \infty$ (but is satisfied, for instance, if $\|X\|_{rp} < \infty$ for some $r > 2$).

Theorem 4.2. *Assume that* $\int_0^\infty t^{p-1}\sqrt{H(t)}dt < \infty$.

- *If* $p > d/2$, *there exists an universal positive constant* C *depending only on* (p, d) *such that*

$$\limsup_{n\to\infty} \sqrt{\frac{n}{\log\log n}} W_p^p(\mu_n, \mu) \leq C \int_0^\infty t^{p-1}\sqrt{H(t)}dt \quad a.s.$$

- *If* $p \in [1, d/2)$, *there exists an universal positive constant* C *depending only on* (p, d) *such that*

$$\limsup_{n\to\infty} \left(\frac{n}{\log\log n} \right)^{p/d} W_p^p(\mu_n, \mu) \leq C \int_0^\infty t^{p-1}\sqrt{H(t)}dt \quad a.s.$$

Remark 4.2. In the case $p = 1, d = 1$, it follows from the central limit theorem for $W_1(\mu_n, \mu)$ (see [4]) and from Theorem 10.12 in [14] that the sequence $(\sqrt{n/\log\log n}\ W_1(\mu_n, \mu))_{n\geq 0}$ is almost surely relatively compact if $\int_0^\infty \sqrt{H(t)}dt < \infty$, which is consistent with the first item of Theorem 4.2.

Remark 4.3. For $p = 1$, concerning the rate of Corollary 4.1 when $d > r/(r-1)$ or the rate Theorem 4.2 when $d > 2$, the situation is not as clear as in the small dimension case. According to Talagrand [17], if $d > 2$ and μ is the uniform measure on $[0, 1]^d$, $W_1(\mu_n, \mu)$ is exactly of order $n^{-1/d}$ almost surely. More generally, let us recall a famous result by Dobrić and Yukich [10]: if $d > 2$ and μ is compactly supported, then

$$\lim_{n\to\infty} n^{1/d} W_1(\mu_n, \mu) = c(d) \int (f_\mu(x))^{(d-1)/d} \ \text{ a.s.} \tag{4.2}$$

where $c(d)$ depends only on d, and f_μ is the density of the absolutely continuous part of μ (hence the limit is zero if μ is singular with respect to the Lebesgue measure on $\mathbb{R}^d$). Corollary 4.1 or Theorem 4.2 are difficult to compare with (4.2), because the results do not hold under the same assumptions on d and r. Some reasonable questions are : can we prove (4.2) if $\|X\|_r < \infty$ for some $r \in (1,2)$ and $d > r/(r-1)$? can we prove (4.2) if $\int_0^\infty \sqrt{H(t)}dt < \infty$ and $d > 2$?

5 Moment inequalities

In this section, we give some upper bounds for the moments $\|W_p^p(\mu_n,\mu)\|_r^r$ when the variables have a strong moment of order rp.

As will be clear from the proofs, the maximal versions of these inequalities hold, meaning that the quantity $\|W_p^p(\mu_n,\mu)\|_r$ can be replaced by

$$\frac{1}{n^r}\left\|\max_{1\le k\le n} kW_p^p(\mu_k,\mu)\right\|_r^r$$

in all the statements of this section.

5.1 Moment of order 1 and 2

Theorem 5.1. *Let $q \in (1,2]$. If $\|X\|_p < \infty$, then, for any $M > 0$,*

$$\left\|W_p^p(\mu_n,\mu)\right\|_1 \ll \begin{cases} \displaystyle\int_0^\infty t^{p-1}H(t)\mathbf{1}_{t>M}\,dt + \frac{1}{n^{(q-1)/q}}\int_0^\infty t^{p-1}(H(t))^{1/q}\mathbf{1}_{t\le M}\,dt & \text{if } p > d(q-1)/q \\ \displaystyle\int_0^\infty t^{p-1}H(t)\mathbf{1}_{t>M}\,dt + \frac{\log n}{n^{p/d}}\int_0^\infty t^{p-1}(H(t))^{(d-p)/d}\mathbf{1}_{t\le M}\,dt & \text{if } p = d(q-1)/q \\ \displaystyle\int_0^\infty t^{p-1}H(t)\mathbf{1}_{t>M}\,dt + \frac{1}{n^{p/d}}\int_0^\infty t^{p-1}(H(t))^{(d-p)/d}\mathbf{1}_{t\le M}\,dt & \text{if } p \in [1, d(q-1)/q) \end{cases}$$

where the constant implicitly involved does not depend on M.

Remark 5.1. In particular, if $H(t) \le Ct^{-p}(\log(1+t))^{-a}$ for some $C > 0, a > 1$, then

$$\left\|W_p^p(\mu_n,\mu)\right\|_1 = O\left(\frac{1}{(\log n)^{a-1}}\right).$$

Remark 5.2. If $\|X\|_{rp,w} < \infty$ for $r \in (1,2)$ and $p \ne d(r-1)/r$, we easily infer from Theorem 5.1 that

$$\left\|W_p^p(\mu_n,\mu)\right\|_1 \ll \begin{cases} \dfrac{\|X\|_{rp,w}^p}{n^{(r-1)/r}} & \text{if } p > d(r-1)/r \\ \dfrac{\|X\|_{rp,w}^p}{n^{p/d}} & \text{if } p \in [1, d(r-1)/r) \end{cases}$$

which can also be deduced from Theorem 2.1. If $p = d(r-1)/r$, we get

$$\left\|W_p^p(\mu_n,\mu)\right\|_1 \ll \frac{\|X\|_{rp,w}^p(\log n)^2}{n^{p/d}}.$$

Now, if $\|X\|_{2p,w} < \infty$, we get from Theorem 5.1 that

$$\left\|W_p^p(\mu_n,\mu)\right\|_1 \ll \begin{cases} \dfrac{\|X\|_{2p,w}^p \log n}{\sqrt{n}} & \text{if } p > d/2 \\ \dfrac{\|X\|_{2p,w}^p (\log n)^2}{\sqrt{n}} & \text{if } p = d/2 \\ \dfrac{\|X\|_{2p,w}^p}{n^{p/d}} & \text{if } p \in [1, d/2) \end{cases}$$

Finally, if $\int_0^\infty t^{p-1}\sqrt{H(t)}\,dt < \infty$, the rates in the cases $p > d/2$ and $p = d/2$ can be slightly improved (taking $q = 2$ and $M = \infty$ in Theorem 5.1); this can be directly deduced from Theorem 5.2 below. Note that all those bounds are consistent with that given in Theorem 1 of [13], and slightly more precise in terms of the moment conditions. Hence, the discussion on the optimality of the rates in [13] is also valid for our Theorem 5.1 (see Remark 5.3 below). For $p > d/2$ and $\|X\|_q < \infty$ for some $q > dp/(d-p)$, it follows from Theorem 2(ii) in [9] that $\liminf_{n\to\infty} n^{p/d}\left\|W_p^p(\mu_n,\mu)\right\|_1 > 0$ if μ has a non degenerate absolutely continuous part with respect to the Lebesgue measure, and that $\limsup_{n\to\infty} n^{p/d}\left\|W_p^p(\mu_n,\mu)\right\|_1 = 0$ if μ is singular. Still for $p < d/2$, we refer to the paper [2], which shows that, for compactly supported singular measures, the rates of convergence of $\|W_p^p(\mu_n,\mu)\|_1$ can be much faster than $n^{-p/d}$.

Theorem 5.2. *If $\int_0^\infty t^{p-1}\sqrt{H(t)}\,dt < \infty$, then*

$$\left\|W_p^p(\mu_n,\mu)\right\|_2^2 \ll \begin{cases} \dfrac{1}{n}\left(\displaystyle\int_0^\infty t^{p-1}\sqrt{H(t)}\,dt\right)^2 & \text{if } p > d/2 \\ \dfrac{(\log(n))^2}{n}\left(\displaystyle\int_0^\infty t^{p-1}\sqrt{H(t)}\,dt\right)^2 & \text{if } p = d/2 \\ \dfrac{1}{n^{2p/d}}\left(\displaystyle\int_0^\infty t^{p-1}\sqrt{H(t)}\,dt\right)^2 & \text{if } p \in [1, d/2) \end{cases}$$

Remark 5.3. According to the discussion after Theorem 1 in [13], if $p \neq d/2$, one can always find some measure μ for which the rate of Theorem 5.2 is reached (see example (a) and (b) in [13] for $p > d/2$ and example (c) in [13] for $p < d/2$). Note also, that, in the case $d = 1, p = 1$, del Barrio et al. [4] proved that $\sqrt{n}W_1(\mu_n,\mu)$ is stochastically bounded if and only if $\int_0^\infty \sqrt{H(t)}\,dt < \infty$ (see their Theorem 2.1(b)), which is consistent with the first inequality of Theorem 5.2. For $d = 1, p > 1$, we refer to the paper by Bobkov and Ledoux [6] for some conditions on μ ensuring faster rates of convergence. Finally, when $p = 1, d = 2$ and μ is the uniform measure over $[0,1]^2$, Ajtai et al. [1] proved that $\mathbb{E}(W_1(\mu_n,\mu))$ is exactly of order $(\log n/n)^{1/2}$, while we get a rate of order $\log n/\sqrt{n}$, which is therefore suboptimal in that particular case.

5.2 von Bahr-Esseen type inequalities

We now consider the case where $r \in (1,2)$.

Theorem 5.3. *If $\|X\|_{rp} < \infty$ for some $r \in (1,2)$, then*

$$\left\|W_p^p(\mu_n,\mu)\right\|_r^r \ll \begin{cases} \dfrac{\|X\|_{rp}^{rp}}{n^{r-1}} & \text{if } p > d(r-1)/r \\ \dfrac{\|X\|_{rp}^{rp}}{n^{rp/d}} & \text{if } p \in [1, d(r-1)/r) \end{cases}$$

Remark 5.4. For $d = 1$, the first inequality of Theorem 5.3 has been proved in [8]. Our proof does not allow to deal with the case where $p = d(r-1)/r$. However, in that case, it is easy to see that

$$\left\|W_p^p(\mu_n,\mu)\right\|_r^r \le \frac{(\log n)^r}{n^{r-1}} \left(\int_0^\infty t^{p-1}(H(t))^{1/r}\,dt\right)^r$$

(same proof as the second inequality of Theorem 5.2). For $p = 1$ and $d < r/(r-1)$, using the dual definition of $W_1(\mu_n,\mu)$, we get the upper bound

$$\left\| \sup_{f\in\Lambda_1} \left| \frac{1}{n}\sum_{k=1}^n (f(X_k) - \mu(f)) \right| \right\|_r^r \ll \frac{\|X\|_r^r}{n^{r-1}}\,, \tag{5.1}$$

where Λ_1 is the the set of functions f such that $|f(x) - f(y)| \le |x - y|$. Note that (5.1) may be seen as a uniform version of the von Bahr-Esseen inequality (see [3]) over the class Λ_1.

5.3 Rosenthal type inequalities

We consider now the case where $r > 2$.

Theorem 5.4. *If $\|X\|_{rp} < \infty$ for some $r > 2$, then*

$$\|W_p^p(\mu_n,\mu)\|_r^r \ll \begin{cases} \dfrac{1}{n^{r/2}}\left(\displaystyle\int_0^\infty t^{p-1}\sqrt{H(t)}\,dt\right)^r + \dfrac{\|X\|_{rp}^{rp}}{n^{r-1}} & \text{if } p > d(r-1)/r \\ \dfrac{1}{n^{r/2}}\left(\displaystyle\int_0^\infty t^{p-1}\sqrt{H(t)}\,dt\right)^r + \dfrac{n^\gamma}{n^{pr/d}}\|X\|_{rp}^{rp} & \text{if } d/2 < p \le d(r-1)/r \\ \dfrac{(\log n)^r}{n^{r/2}}\left(\displaystyle\int_0^\infty t^{d/2-1}\sqrt{H(t)}dt\right)^r + \dfrac{(\log n)^2}{n^{r/2}}\|X\|_{rp}^{rp} & \text{if } p = d/2 \\ \dfrac{\|X\|_{rp}^{rp}}{n^{rp/d}} & \text{if } p \in [1, d/2) \end{cases}$$

where, for the second inequality, γ can be taken as $\gamma = \frac{\varepsilon(2p-d)}{d(r-2+\varepsilon)}$ for any $\varepsilon > 0$ (and the constants implicitely involved in the inequality depend on ε).

Remark 5.5. For $d = 1$, the first inequality of Theorem 5.4 has been proved in [8]. As a consequence of the two first inequalities of Theorem 5.4, we obtain that, if $p > d/2$,

$$\limsup_{n\to\infty} \sqrt{n}\|W_p^p(\mu_n,\mu)\|_r \ll \int_0^\infty t^{p-1}\sqrt{H(t)}\,dt\,.$$

As a consequence of the third inequality of Theorem 5.4, we obtain that, if $p = d/2$,

$$\limsup_{n\to\infty} \frac{\sqrt{n}}{\log n}\|W_p^p(\mu_n,\mu)\|_r \ll \int_0^\infty t^{p-1}\sqrt{H(t)}\,dt\,.$$

6 Proofs

The starting point of the proofs is Lemmas 5 and 6 in [13], which we recall below.

For $\ell \geq 0$, let $\mathcal{P}_\ell$ be the natural partition of $(-1,1]^d$ into $2^{d\ell}$ translations of $(-2^{-\ell}, 2^\ell]^d$. Let also $B_0 = (-1,1]^d$ and for any integer $m \geq 1$, $B_m = (-2^m, 2^m]^d \setminus (-2^{m-1}, 2^{m-1}]^d$. For a set $F \subset \mathbb{R}^d$ and $a > 0$, we use the standard notation $aF = \{ax : x \in F\}$. For a probability measure ν on $\mathbb{R}^d$ and $m \geq 0$, let $\mathcal{R}_{B_m}\nu$ be the probability measure on $(-1,1]^d$ defined as the image of $\nu|_{B_m}/\nu(B_m)$ by the map $x \mapsto x/2^m$. For two probability measures μ and ν on $\mathbb{R}^d$, by Lemma 5 in [13], there exists a positive constant $\kappa_{p,d}$ depending only on p and d such that

$$W_p^p(\mu,\nu) \leq \kappa_{p,d} \mathcal{D}_p(\mu,\nu)\,, \tag{6.1}$$

where

$$\mathcal{D}_p(\mu,\nu) := \sum_{m\geq 0} 2^{pm} |\mu(B_m) - \nu(B_m)| + \sum_{m\geq 0} 2^{pm} (\mu(B_m) \wedge \nu(B_m)) \mathcal{D}_p(\mathcal{R}_{B_m}\mu, \mathcal{R}_{B_m}\nu)\,, \tag{6.2}$$

with

$$\mathcal{D}_p(\mathcal{R}_{B_m}\mu, \mathcal{R}_{B_m}\nu) = \frac{2^p - 1}{2} \sum_{\ell \geq 1} 2^{-p\ell} \sum_{F \in \mathcal{P}_\ell} \left| \frac{\mu(2^m F \cap B_m)}{\mu(B_m)} - \frac{\nu(2^m F \cap B_m)}{\nu(B_m)} \right|. \tag{6.3}$$

In addition, by Lemma 5 in [13],

$$\mathcal{D}_p(\mu,\nu) \leq \left(\frac{3}{2} \vee \frac{2^p - 1}{2} \right) \Delta_p(\mu_k, \mu)$$

where

$$\Delta_p(\mu,\nu) = \sum_{m\geq 0} 2^{pm} \sum_{\ell \geq 0} 2^{-p\ell} \sum_{F \in \mathcal{P}_\ell} |\mu(2^m F \cap B_m) - \nu(2^m F \cap B_m)|\,.$$

From the considerations above, there exists a constant C depending only on p and d such that

$$W_p^p(\mu_k, \mu) \leq C \Delta_p(\mu_k, \mu)\,, \tag{6.4}$$

where $\mu_k = \frac{1}{k}\sum_{i=1}^k \delta_{X_i}$. This inequality may be seen as an extension to the case $d > 1$ of Èbralidze's inequality [12], which we used in [8] to obtain moment bounds for $W_p^p(\mu_n, \mu)$ when $d = 1$.

As in [8] we shall use truncation arguments. For a positive real M, let $\mathcal{C}_M = [-M, M]^d$,

$$A_{p,M}(\mu_k, \mu) = \sum_{m\geq 0} 2^{pm} \sum_{\ell \geq 0} 2^{-p\ell} \sum_{F \in \mathcal{P}_\ell} |\mu_k(2^m F \cap B_m \cap \mathcal{C}_M) - \mu(2^m F \cap B_m \cap \mathcal{C}_M)|$$

and

$$B_{p,M}(\mu_k, \mu) = \sum_{m\geq 0} 2^{pm} \sum_{\ell \geq 0} 2^{-p\ell} \sum_{F \in \mathcal{P}_\ell} |\mu_k(2^m F \cap B_m \cap \mathcal{C}_M^c) - \mu(2^m F \cap B_m \cap \mathcal{C}_M^c)|\,.$$

With these notations, it follows from (6.4) that

$$\Delta_p(\mu_k, \mu) \leq A_{p,M}(\mu_k, \mu) + B_{p,M}(\mu_k, \mu)\,. \tag{6.5}$$

For the proofs, we shall follow the order of the theorems, except for Theorem 5.3 whose proof comes naturally after that of Theorems 2.1 and 2.2.

6.1 Proof of Theorem 2.1

Starting from (6.5), we get that

$$\mathbb{P}\left(\max_{1\le k\le n} kW_p^p(\mu_k,\mu) > nx\right) \le \mathbb{P}\left(\max_{1\le k\le n} kA_{p,M}(\mu_k,\mu) > (nx/2C)\right) + \mathbb{P}\left(\max_{1\le k\le n} kB_{p,M}(\mu_k,\mu) > (nx/2C)\right). \quad (6.6)$$

Let $y = x/2C$. By Markov's inequality at order $q \in (r,2)$ and $s \in [1,r)$,

$$\mathbb{P}\left(\max_{1\le k\le n} kA_{p,M}(\mu_k,\mu) > ny\right) \le \frac{\|\max_{1\le k\le n} kA_{p,M}(\mu_k,\mu)\|_q^q}{n^q y^q}, \quad (6.7)$$

$$\mathbb{P}\left(\max_{1\le k\le n} kB_{p,M}(\mu_k,\mu) > ny\right) \le \frac{\|\max_{1\le k\le n} kB_{p,M}(\mu_k,\mu)\|_s}{n^s y^s}. \quad (6.8)$$

To deal with (6.7), we first note that

$$\left\|\max_{1\le k\le n} kA_{p,M}(\mu_k,\mu)\right\|_q \ll \sum_{m\ge 0} 2^{pm} \sum_{\ell\ge 0} 2^{-p\ell} \left\|\max_{1\le k\le n} \sum_{F\in\mathcal{P}_\ell} |k\mu_k(2^m F\cap B_m\cap\mathcal{C}_M) - k\mu(2^m F\cap B_m\cap\mathcal{C}_M)|\right\|_q.$$

Now, clearly

$$\left\|\max_{1\le k\le n} \sum_{F\in\mathcal{P}_\ell} |k\mu_k(2^m F\cap B_m\cap\mathcal{C}_M) - k\mu(2^m F\cap B_m\cap\mathcal{C}_M)|\right\|_q \le \left\|\max_{1\le k\le n} (k\mu_k(B_m\cap\mathcal{C}_M) + k\mu(B_m\cap\mathcal{C}_M))\right\|_q \le 2n\left(\mu(B_m\cap\mathcal{C}_M)\right)^{1/q}. \quad (6.9)$$

On the other hand, by the (maximal version of) von Bahr-Essen inequality (see [3]),

$$\left\|\max_{1\le k\le n} |k\mu_k(2^m F\cap B_m\cap\mathcal{C}_M) - k\mu(2^m F\cap B_m\cap\mathcal{C}_M)|\right\|_q^q \ll n\mu(2^m F\cap B_m\cap\mathcal{C}_M),$$

so that, by using Hölder's inequality and the fact that $|\mathcal{P}_\ell| = 2^{\ell d}$,

$$\sum_{F\in\mathcal{P}_\ell} \left\|\max_{1\le k\le n} |k\mu_k(2^m F\cap B_m\cap\mathcal{C}_M) - k\mu(2^m F\cap B_m\cap\mathcal{C}_M)|\right\|_q \ll 2^{\ell d(q-1)/q} n^{1/q} \left(\mu(B_m\cap\mathcal{C}_M)\right)^{1/q}. \quad (6.10)$$

Combining (6.7), (6.9) and (6.10), we obtain that

$$\mathbb{P}\left(\max_{1\le k\le n} kA_{p,M}(\mu_k,\mu) > ny\right) \ll \frac{1}{y^q}\left(\sum_{m\ge 0} 2^{pm}\left(\mu(B_m\cap\mathcal{C}_M)\right)^{1/q} \sum_{\ell\ge 0} \frac{1}{2^{p\ell}} \min\left(1, n^{-(q-1)/q} 2^{\ell d(q-1)/q}\right)\right)^q. \quad (6.11)$$

In the same way, for the term (6.8), we obtain the upper bound

$$\mathbb{P}\left(\max_{1\le k\le n} kB_{p,M}(\mu_k,\mu) > ny\right) \ll \frac{1}{y^s}\left(\sum_{m\ge 0} 2^{pm}\left(\mu(B_m\cap \mathcal{C}_M^c)\right)^{1/s} \sum_{\ell\ge 0}\frac{1}{2^{p\ell}}\min\left(1, n^{-(s-1)/s}2^{\ell d(s-1)/s}\right)\right)^s. \tag{6.12}$$

From (6.11) and (6.12), we see that three cases arise:

• If $p > d(r-1)/r$, then, taking $q > r$ such that $p > d(q-1)/q$ and $s = 1$, we get the upper bounds

$$\mathbb{P}\left(\max_{1\le k\le n} kA_{p,M}(\mu_k,\mu) > ny\right) \ll \frac{1}{n^{q-1}y^q}\left(\sum_{m\ge 0} 2^{pm}\left(\mu(B_m\cap \mathcal{C}_M)\right)^{1/q}\right)^q \ll \frac{1}{n^{q-1}y^q}\left(\int_0^\infty t^{p-1}(H(t))^{1/q}\mathbf{1}_{t\le M}dt\right)^q, \tag{6.13}$$

and

$$\mathbb{P}\left(\max_{1\le k\le n} kB_{p,M}(\mu_k,\mu) > ny\right) \ll \frac{1}{y}\int_0^\infty t^{p-1}H(t)\mathbf{1}_{t>M}dt\,. \tag{6.14}$$

Using that $H(t) \le \|X\|_{rp,w}^{rp} t^{-rp}$ for $r \in (1,2)$, we infer from (6.6), (6.13) and (6.14) that

$$\mathbb{P}\left(\max_{1\le k\le n} kW_p^p(\mu_k,\mu) > nx\right) \ll \|X\|_{rp,w}^{rp}\left(\frac{1}{xM^{p(r-1)}} + \frac{M^{p(q-r)}}{n^{q-1}x^q}\right).$$

Taking $M = (nx)^{1/p}$, we obtain the desired result when $p > d(r-1)/r$.

• If $p = d(r-1)/r$, then, taking $q = r$, we get the upper bound

$$\mathbb{P}\left(\max_{1\le k\le n} kA_{p,M}(\mu_k,\mu) > ny\right) \ll \frac{(\log(n))^r}{n^{r-1}y^r}\left(\sum_{m\ge 0} 2^{pm}\left(\mu(B_m\cap \mathcal{C}_M)\right)^{1/r}\right)^r \ll \frac{(\log n)^r}{n^{r-1}y^r}\left(\int_0^\infty t^{p-1}(H(t))^{1/r}\mathbf{1}_{t\le M}dt\right)^r. \tag{6.15}$$

Using that $H(t) \le \|X\|_{rp,w}^{rp} t^{-rp}$ for $r \in (1,2)$, we infer from (6.6), (6.14) and (6.15) that

$$\mathbb{P}\left(\max_{1\le k\le n} kW_p^p(\mu_k,\mu) > nx\right) \ll \|X\|_{rp,w}^{rp}\left(\frac{1}{xM^{p(r-1)}} + \frac{(\log n)^r}{n^{r-1}x^r}\left(1+\log_+\left(\frac{M}{\|X\|_{rp,w}}\right)\right)^r\right).$$

Taking $M = (nx)^{1/p}$, we obtain the desired result when $p = d(r-1)/r$.

• If $p < d(r-1)/r$, then, taking $q > r$ and $s \in (1,r)$ such that $p < d(s-1)/s$, we get the upper bounds

$$\mathbb{P}\left(\max_{1\le k\le n} kA_{p,M}(\mu_k,\mu) > ny\right) \ll \frac{1}{n^{qp/d}y^q}\left(\sum_{m\ge 0} 2^{pm}\left(\mu(B_m\cap \mathcal{C}_M)\right)^{1/q}\right)^q \ll \frac{1}{n^{qp/d}y^q}\left(\int_0^\infty t^{p-1}(H(t))^{1/q}\mathbf{1}_{t\le M}dt\right)^r, \tag{6.16}$$

and

$$\mathbb{P}\left(\max_{1\le k\le n} kB_{p,M}(\mu_k,\mu) > ny\right)$$
$$\ll \frac{1}{n^{sp/d}y^s}\left(\sum_{m\ge 0} 2^{pm}\left(\mu(B_m\cap \mathcal{C}_M^c)\right)^{1/s}\right)^s \ll \frac{1}{n^{sp/d}y^s}\left(\int_0^\infty t^{p-1}(H(t))^{1/s}\mathbf{1}_{t>M}dt\right)^s , \quad (6.17)$$

Using that $H(t)\le \|X\|_{rp,w}^{rp}t^{-rp}$ for $r\in(1,2)$, we infer from (6.6), (6.16) and (6.17) that

$$\mathbb{P}\left(\max_{1\le k\le n} kW_p^p(\mu_k,\mu) > nx\right) \ll \|X\|_{rp,w}^{rp}\left(\frac{1}{n^{sp/d}x^sM^{p(r-s)}} + \frac{M^{p(q-r)}}{n^{qp/d}x^q}\right).$$

Taking $M = n^{1/d}x^{1/p}$, we obtain the desired result when $p < d(r-1)/r$.

6.2 Proof of Theorem 2.2

Let $r > 2$. Note first that, by homogeneity, the general inequality may be deduced from the case where $\|X\|_{rp,w} = 1$ by considering the variables $X_i/\|X\|_{rp,w}$. Hence, from now, we shall assume that $\|X\|_{rp,w} = 1$.

According to the beginning of the proof of Theorem 2 in [13], we get that

$$W_p^p(\mu_n,\mu) \le C\sum_{m\ge 0} 2^{pm}\left|\mu_n(B_m)-\mu(B_m)\right| + CV_n^p , \quad (6.18)$$

for some positive constant $C = C_{p,d}$, where the random variable V_n^p is such that

$$\mathbb{P}(V_n^p \ge x/(2C)) \le a(n,x). \quad (6.19)$$

Consequently, it remains to bound up the quantity

$$\mathbb{P}\left(\sum_{m\ge 0} 2^{pm}\left|\mu_n(B_m)-\mu(B_m)\right| \ge x/(2C)\right).$$

For a positive real M, let $\mathcal{C}_M = [-M,M]^d$,

$$A_{p,M}^*(\mu_k,\mu) = \sum_{m\ge 0} 2^{pm}|\mu_k(B_m\cap\mathcal{C}_M) - \mu(B_m\cap\mathcal{C}_M)|$$

and

$$B_{p,M}^*(\mu_k,\mu) = \sum_{m\ge 0} 2^{pm}|\mu_k(B_m\cap\mathcal{C}_M^c) - \mu(B_m\cap\mathcal{C}_M^c)|.$$

With these notations,

$$\mathbb{P}\left(\sum_{m\ge 0} 2^{pm}\left|\mu_n(B_m)-\mu(B_m)\right| \ge x/(2C)\right) \le \mathbb{P}\left(A_{p,M}^*(\mu_n,\mu) > x/(4C)\right)$$
$$+\mathbb{P}\left(B_{p,M}^*(\mu_n,\mu) > x/(4C)\right). \quad (6.20)$$

Let $y = x/4C$. By Markov's inequality at order $q > 2$ and 1,

$$\mathbb{P}\left(A^*_{p,M}(\mu_n,\mu) > y\right) \le \frac{\left\|A^*_{p,M}(\mu_n,\mu)\right\|_q^q}{y^q}\,, \tag{6.21}$$

$$\mathbb{P}\left(B^*_{p,M}(\mu_n,\mu) > y\right) \le \frac{\left\|B^*_{p,M}(\mu_n,\mu)\right\|_1}{y}\,. \tag{6.22}$$

Applying Rosenthal's inequality, we get

$$\begin{aligned}\|A^*_{p,M}(\mu_n,\mu)\|_q &\ll \frac{1}{\sqrt{n}}\sum_{m\ge 0} 2^{pm}\mu^{1/2}(B_m\cap\mathcal{C}_M) + \frac{1}{n^{(q-1)/q}}\sum_{m\ge 0}2^{pm}\mu^{1/q}(B_m\cap\mathcal{C}_M)\\ &\ll \frac{1}{\sqrt{n}}\int_0^\infty t^{p-1}\sqrt{H(t)}\mathbf{1}_{t\le M}dt + \frac{1}{n^{(q-1)/q}}\int_0^\infty t^{p-1}H^{1/q}(t)\mathbf{1}_{t\le M}dt\,.\end{aligned}$$

Choosing $q > r$, it follows that

$$\begin{aligned}\|A^*_{p,M}(\mu_n,\mu)\|_q &\ll \frac{1}{\sqrt{n}}\int_0^\infty t^{p-1}\sqrt{H(t)}dt + \frac{M^{p(q-r)/q}}{n^{(q-1)/q}}\left(\sup_{t>0} t^{rp}H(t)\right)^{1/q}\\ &\ll \frac{1}{\sqrt{n}}\int_0^\infty t^{p-1}\sqrt{H(t)}dt + \frac{M^{p(q-r)/q}}{n^{(q-1)/q}}\,.\end{aligned} \tag{6.23}$$

the last inequality being true since we assumed that $\sup_{t>0} t^{rp}H(t) = 1$.

On another hand,

$$\begin{aligned}\|B^*_{p,M}(\mu_n,\mu)\|_1 &\ll \sum_{n\ge 0} 2^{pn}\mu(B_n\cap\mathcal{C}_M^c)\\ &\ll \int_0^\infty t^{p-1}H(t)\mathbf{1}_{t>M}dt \ll M^{p(1-r)}\sup_{t>0} t^{rp}H(t) \ll M^{p(1-r)}\,.\end{aligned} \tag{6.24}$$

Gathering (6.20) - (6.24), we get that for any $q > r$,

$$\mathbb{P}\left(\sum_{m\ge 0} 2^{pm}\left|\mu_n(B_m) - \mu(B_m)\right| > x/(2C)\right) \ll \frac{1}{x^q n^{q/2}}\left(\int_0^\infty t^{p-1}\sqrt{H(t)}dt\right)^q + \frac{M^{p(q-r)}}{x^q n^{q-1}} + \frac{M^{p(1-r)}}{x}\,, \tag{6.25}$$

Hence choosing $M = n^{1/p}x^{1/p}$, we infer from (6.18), (6.19) and (6.25) that for any $q > r$,

$$\mathbb{P}\left(W_p^p(\mu_n,\mu) > x\right) \ll a(n,x) + \frac{1}{x^r n^{r-1}} + \frac{1}{x^q n^{q/2}}\left(\int_0^\infty t^{p-1}\sqrt{H(t)}dt\right)^q\,, \tag{6.26}$$

which is the desired inequality when $\sup_{t>0} t^{rp}H(t) = \|X\|_{rp,w}^{rp} = 1$.

6.3 Proof of Theorem 5.3

We start from the elementary equality

$$\frac{1}{n^r}\left\|\max_{1\le k\le n} kW_p^p(\mu_k,\mu)\right\|_r^r = r\int_0^\infty x^{r-1}\mathbb{P}\left(\max_{1\le k\le n} kW_p^p(\mu_k,\mu) > nx\right)dx\,. \tag{6.27}$$

Then, we use the upper bounds (6.6), (6.11) and (6.12). We consider two cases:

• If $p > d(r-1)/r$, let $q > r$ such that $p > d(q-1)/q$, and let $M = (nx)^{1/p}$. From (6.6), (6.13) and (6.14) we get the upper bound

$$\begin{aligned}&\int x^{r-1}\mathbb{P}\left(\max_{1\le k\le n} kW_p^p(\mu_k,\mu) > nx\right)dx\\ &\ll \int_0^\infty x^{r-2}\left(\int_0^\infty t^{p-1}H(t)\mathbf{1}_{t>M}dt\right)dx + \frac{1}{n^{q-1}}\int_0^\infty x^{r-1-q}\left(\int_0^\infty t^{p-1}(H(t))^{1/q}\mathbf{1}_{t\le M}dt\right)^q dx\,. \quad (6.28)\end{aligned}$$

Note that

$$\int_0^\infty x^{r-2}\left(\int_0^\infty t^{p-1}H(t)\mathbf{1}_{t>(nx)^{1/p}}dt\right)dx \ll \frac{1}{n^{r-1}}\int_0^\infty t^{rp-1}H(t)dt \ll \frac{\|X\|_{rp}^{rp}}{n^{r-1}}\,. \quad (6.29)$$

Let $\beta < (q-1)/q$. Applying Hölder's inequality, we obtain

$$\begin{aligned}&\frac{1}{n^{q-1}}\int_0^\infty x^{r-1-q}\left(\int_0^\infty t^{p-1}(H(t))^{1/q}\mathbf{1}_{t\le(nx)^{1/p}}dt\right)^q dx\\ &\qquad\ll \frac{1}{n^{q-1}}\int_0^\infty x^{r-1-q}(nx)^{(q-1-q\beta)/p}\left(\int_0^\infty t^{q(p-1+\beta)}H(t)\mathbf{1}_{t\le(nx)^{1/p}}dt\right)dx\\ &\qquad\ll \frac{n^{(q-1-q\beta)/p}}{n^{q-1}}\int_0^\infty t^{q(p-1+\beta)}H(t)\left(\int_0^\infty x^{r-1-q+(q-1-q\beta)/p}\mathbf{1}_{x\ge t^p/n}dx\right)dt\,. \quad (6.30)\end{aligned}$$

Taking β close enough to $(q-1)/q$ in such a way that $q+1-r-(q-1-q\beta)/p > 1$, we get

$$\frac{1}{n^{q-1}}\int_0^\infty x^{r-1-q}\left(\int_0^\infty t^{p-1}(H(t))^{1/q}\mathbf{1}_{t\le(nx)^{1/p}}dt\right)^q dx \ll \frac{\|X\|_{rp}^{rp}}{n^{r-1}}\,. \quad (6.31)$$

Gathering (6.28), (6.29) and (6.31), we obtain the desired result.

• If $p < d(r-1)/r$, let $q > r$, $s \in (1,r)$ such that $p < d(s-1)/s$, and let $M = n^{1/d}x^{1/p}$. From (6.6), (6.16) (6.17) we get the upper bound

$$\begin{aligned}\int x^{r-1}\mathbb{P}\left(\max_{1\le k\le n} kW_p^p(\mu_k,\mu) > nx\right)dx &\ll \frac{1}{n^{sp/d}}\int_0^\infty x^{r-1-s}\left(\int_0^\infty t^{p-1}(H(t))^{1/s}\mathbf{1}_{t>M}dt\right)^s dx\\ &\quad + \frac{1}{n^{qp/d}}\int_0^\infty x^{r-1-q}\left(\int_0^\infty t^{p-1}(H(t))^{1/q}\mathbf{1}_{t\le M}dt\right)^q dx\,. \quad (6.32)\end{aligned}$$

Proceeding exactly as for (6.30)-(6.31), with the choice $M = n^{1/d}x^{1/p}$, we get

$$\frac{1}{n^{pq/d}}\int_0^\infty x^{r-1-q}\left(\int_0^\infty t^{p-1}(H(t))^{1/q}\mathbf{1}_{t\le n^{1/d}x^{1/p}}dt\right)^q dx \ll \frac{1}{n^{pr/d}}\int_0^\infty t^{rp-1}H(t)dt \ll \frac{\|X\|_{rp}^{rp}}{n^{pr/d}}\,. \quad (6.33)$$

In the same way, we get

$$\frac{1}{n^{ps/d}}\int_0^\infty x^{r-1-s}\left(\int_0^\infty t^{p-1}(H(t))^{1/s}\mathbf{1}_{t>n^{1/d}x^{1/p}}dt\right)^s dx \ll \frac{1}{n^{pr/d}}\int_0^\infty t^{rp-1}H(t)dt \ll \frac{\|X\|_{rp}^{rp}}{n^{pr/d}}\,. \quad (6.34)$$

Gathering (6.32), (6.33) and (6.34), we obtain the desired result.

6.4 Proof of Theorem 3.3

Let $r \in (1,2)$. We start from the upper bounds (6.6), (6.12) and (6.11).

• If $p > d(r-1)/r$, let $q \in (r,2]$ such that $p > d(q-1)/q$ and let $M = n^{\alpha/p}$. From (6.6), (6.13) and (6.14) we get the upper bound

$$\mathbb{P}\left(\max_{1\le k\le n} kW_p^p(\mu_k,\mu) > n^\alpha x\right) \ll \frac{n^{1-\alpha}}{x}\int_0^\infty t^{p-1}H(t)\mathbf{1}_{t>n^{\alpha/p}}dt + \frac{n^{1-q\alpha}}{x^q}\left(\int_0^\infty t^{p-1}(H(t))^{1/q}\mathbf{1}_{t\le n^{\alpha/p}}dt\right)^q.$$

Hence, it remains to prove that

$$\sum_{n=1}^\infty n^{\alpha(r-1)-1}\int_0^\infty t^{p-1}H(t)\mathbf{1}_{t^{p/\alpha}>n}dt < \infty \quad \text{and} \quad \sum_{n=1}^\infty n^{\alpha(r-q)-1}\left(\int_0^\infty t^{p-1}(H(t))^{1/q}\mathbf{1}_{t^{p/\alpha}\le n}dt\right)^q < \infty\,. \tag{6.35}$$

Interverting the sum and the integral, we easily get that

$$\sum_{n=1}^\infty n^{\alpha(r-1)-1}\int_0^\infty t^{p-1}H(t)\mathbf{1}_{t^{p/\alpha}>n}dt \ll \int_0^\infty t^{pr-1}H(t)dt \ll \|X\|_{pr}^{pr} < \infty\,. \tag{6.36}$$

Arguing as in (6.30) with $\beta < (q-1)/q$, we get

$$\left(\int_0^\infty t^{p-1}(H(t))^{1/q}\mathbf{1}_{t^{p/\alpha}\le n}dt\right)^q \ll n^{\alpha(q-1-q\beta)/p}\int_0^\infty t^{q(p-1+\beta)}H(t)\mathbf{1}_{t^{p/\alpha}\le n}dt\,.$$

Hence, the second series in (6.35) will be summable provided

$$\sum_{n=1}^\infty n^{\alpha(r-q)+\alpha(q-1-q\beta)/p-1}\int_0^\infty t^{q(p-1+\beta)}H(t)\mathbf{1}_{t^{p/\alpha}\le n}dt < \infty \tag{6.37}$$

Taking β close enough to $(q-1)/q$ so that $\alpha(r-q)+\alpha(q-1-q\beta)/p < 0$ and interverting the sum and the integral, we get that

$$\sum_{n=1}^\infty n^{\alpha(r-q)+\alpha(q-1-q\beta)/p-1}\int_0^\infty t^{q(p-1+\beta)}H(t)\mathbf{1}_{t^{p/\alpha}\le n}dt \ll \int_0^\infty t^{pr-1}H(t)dt \ll \|X\|_{pr}^{pr} < \infty\,. \tag{6.38}$$

The result follows from (6.35), (6.36), (6.37) and (6.38).

• If $p < d(r-1)/r$, let $q \in (r,2]$, $s \in (1,r)$ such that $p < d(s-1)/s$, and let $M = n^{(p-d(1-\alpha))/(dp)}$. From (6.6), (6.16) and (6.17), we get the upper bound

$$\begin{aligned}\mathbb{P}\left(\max_{1\le k\le n} kW_p^p(\mu_k,\mu) > n^\alpha x\right) \ll{}& \frac{n^{s-s\alpha}}{n^{sp/d}x^s}\left(\int_0^\infty t^{p-1}(H(t))^{1/s}\mathbf{1}_{t>n^{(p-d(1-\alpha))/(dp)}}dt\right)^s \\ &+ \frac{n^{q-q\alpha}}{n^{qp/d}x^q}\left(\int_0^\infty t^{p-1}(H(t))^{1/q}\mathbf{1}_{t\le n^{(p-d(1-\alpha))/(dp)}}dt\right)^q.\end{aligned} \tag{6.39}$$

Proceeding as in (6.37) (taking the quantity $\beta < (q-1)/q$ close enough to $(q-1)/q$ in such a way that $(p-d(1-\alpha))((r-q)+(q-1-\beta q)/p) < 0$), we get that

$$\sum_{n=1}^\infty n^{(pr-(1-\alpha)rd-d)/d}\frac{n^{q-q\alpha}}{n^{qp/d}}\left(\int_0^\infty t^{p-1}(H(t))^{1/q}\mathbf{1}_{t\le n^{(p-d(1-\alpha))/(dp)}}dt\right)^q \ll \|X\|_{pr}^{pr} < \infty\,. \tag{6.40}$$

In the same, we get

$$\sum_{n=1}^{\infty} n^{(pr-(1-\alpha)rd-d)/d} \frac{n^{s-s\alpha}}{n^{sp/d}} \left(\int_0^{\infty} t^{p-1} (H(t))^{1/s} \mathbf{1}_{t>n^{(p-d(1-\alpha))/(dp)}} dt \right)^s \ll \|X\|_{pr}^{pr} < \infty \,. \tag{6.41}$$

The result follows from (6.39), (6.40) and (6.41).

• If $p < d(r-1)/r$, let $q \in (r, 2]$, $s \in (1, r)$ such that $p < d(s-1)/s$, and let $M = (\log n)^{1/pr}$. From (6.6), (6.16) and (6.17), we get the upper bound

$$\mathbb{P}\left(\max_{1\le k\le n} kW_p^p(\mu_k, \mu) > n^{(d-p)/d} (\log n)^{1/r} x \right) \ll \frac{1}{(\log n)^{s/r} x^s} \left(\int_0^{\infty} t^{p-1} (H(t))^{1/s} \mathbf{1}_{t>(\log n)^{1/pr}} dt \right)^s + \frac{1}{(\log n)^{q/r} x^q} \left(\int_0^{\infty} t^{p-1} (H(t))^{1/q} \mathbf{1}_{t\le(\log n)^{1/pr}} dt \right)^q . \tag{6.42}$$

Proceeding as in (6.37) (taking the quantity $\beta < (q-1)/q$ close enough to $(q-1)/q$ in such a way that $(q/r) + (q-1-\beta q)/(pr) > 1$), we get that

$$\sum_{n=1}^{\infty} \frac{1}{n(\log n)^{q/r}} \left(\int_0^{\infty} t^{p-1} (H(t))^{1/q} \mathbf{1}_{t\le(\log n)^{1/pr}} dt \right)^q \ll \|X\|_{pr}^{pr} < \infty \,. \tag{6.43}$$

In the same, we get

$$\sum_{n=1}^{\infty} \frac{1}{n(\log n)^{s/r}} \left(\int_0^{\infty} t^{p-1} (H(t))^{1/s} \mathbf{1}_{t>(\log n)^{1/pr}} dt \right)^s \ll \|X\|_{pr}^{pr} < \infty \,. \tag{6.44}$$

The result follows from (6.42), (6.43) and (6.44).

6.5 Proof of Theorem 3.4

Let $r > 2$. As in the proof of Theorem 2.2, we assume without loss of generality that $\|X\|_{rp,w} = 1$; hence, we can use directly some of the upper bounds given in the proof of Theorem 2.2.

From (6.18), we see that

$$\max_{1\le k\le n} kW_p^p(\mu_k, \mu) \le C \sum_{m\ge 0} 2^{pm} \max_{1\le k\le n} |k\mu_k(B_m) - k\mu(B_m)| + C \max_{1\le k\le n} kV_k^p \,. \tag{6.45}$$

Now, for any $x > 0$

$$\mathbb{P}\left(\max_{1\le k\le n} kV_k^p > x/(2C) \right) \le \sum_{k=1}^{n} \mathbb{P}\left(kV_k^p > x/(2C) \right) \le n \max_{1\le k\le n} \mathbb{P}\left(kV_k^p > x/(2C) \right) .$$

By (6.19), it follows that, for any $x > 0$,

$$\mathbb{P}\left(\max_{1\le k\le n} kV_k^p > x/(2C) \right) \le n \max_{1\le k\le n} a(k, x/k) \le n a(n, x/n) \,, \tag{6.46}$$

the last inequality being true because $k \to a(k, x/k)$ is increasing. Now, by definition of $a(n,x)$, we infer that, for any

$$\alpha \in \left(\max\left(\frac{1}{2}, \frac{d-p}{d}\right), 1\right],$$

$$\sum_{n=1}^{\infty} n^{\alpha r-2} \mathbb{P}\left(\max_{1\le k\le n} kV_k^p > n^\alpha x/(2C)\right) < \infty\,. \tag{6.47}$$

Hence, it remains to prove that

$$\sum_{n=1}^{\infty} n^{\alpha r-2} \mathbb{P}\left(\sum_{m\ge 0} 2^{pm} \max_{1\le k\le n} |k\mu_k(B_m) - k\mu(B_m)| \ge n^\alpha x/(2C)\right) < \infty\,.$$

Arguing as in the proof of Theorem 2.2, and using a maximal version of Rosenthal's inequality (see for instance [16]), we get that, for any $q > r$ and $M > 0$,

$$\mathbb{P}\left(\sum_{m\ge 0} 2^{pm} \max_{1\le k\le n} |k\mu_k(B_m) - k\mu(B_m)| \ge n^\alpha x/(2C)\right) \ll \frac{n^{(1-\alpha)q}}{x^q n^{q/2}} \left(\int_0^\infty t^{p-1}\sqrt{H(t)}dt\right)^q$$
$$+ \frac{n^{1-q\alpha}}{x^q}\left(\int_0^\infty t^{p-1} H^{1/q}(t)\mathbf{1}_{t\le M}dt\right)^q + \frac{n^{1-2\alpha}}{x^2}\left(\int_0^\infty t^{p-1}\sqrt{H(t)}\mathbf{1}_{t>M}dt\right)^2. \tag{6.48}$$

Clearly, since $\alpha \in (1/2, 1]$, taking q large enough, we get that

$$\sum_{n=1}^{\infty} n^{\alpha r-2}\frac{n^{(1-\alpha)q}}{n^{q/2}} \left(\int_0^\infty t^{p-1}\sqrt{H(t)}dt\right)^q < \infty\,. \tag{6.49}$$

Let $M = n^{\alpha/p}$. Arguing as in (6.30) with $\beta > 1/2$, we get

$$\left(\int_0^\infty t^{p-1}\sqrt{H(t)}\mathbf{1}_{t^{p/\alpha}>n}dt\right)^2 \ll n^{\alpha(1-2\beta)/p}\int_0^\infty t^{2(p-1+\beta)}H(t)\mathbf{1}_{t^{p/\alpha}>n}dt\,.$$

Hence, the sum over n of the last term in (6.48) will be finite provided

$$\sum_{n=1}^{\infty} n^{\alpha(r-2)+\alpha(1-2\beta)/p-1}\int_0^\infty t^{2(p-1+\beta)}H(t)\mathbf{1}_{t^{p/\alpha}>n}dt < \infty\,. \tag{6.50}$$

Taking β close enough to $1/2$ so that $\alpha(r-2)+\alpha(1-2\beta)/p > 0$ and interverting the sum and the integral, we get that

$$\sum_{n=1}^{\infty} n^{\alpha(r-2)+\alpha(1-2\beta)/p-1}\int_0^\infty t^{2(p-1+\beta)}H(t)\mathbf{1}_{t^{p/\alpha}>n}dt \ll \int_0^\infty t^{pr-1}H(t)dt \ll \|X\|_{pr}^{pr} < \infty\,. \tag{6.51}$$

Arguing as in (6.30) with $\beta < (q-1)/q$, we get

$$\left(\int_0^\infty t^{p-1}H^{1/q}(t)\mathbf{1}_{t^{p/\alpha}\le n}dt\right)^q \ll n^{\alpha(q-1-\beta q)/p}\int_0^\infty t^{q(p-1+\beta)}H(t)\mathbf{1}_{t^{p/\alpha}\le n}dt\,.$$

Hence, the sum over n of the second term in (6.48) will be finite provided

$$\sum_{n=1}^{\infty} n^{\alpha(r-q)+\alpha(q-1-\beta q)/p-1} \int_0^{\infty} t^{q(p-1+\beta)} H(t) \mathbf{1}_{t^{p/\alpha} \le n} dt < \infty \,. \tag{6.52}$$

Taking β close enough to $(q-1)/q$ so that $\alpha(r-q)+\alpha(q-1-\beta q)/p < 0$ and interverting the sum and the integral, we get that

$$\sum_{n=1}^{\infty} n^{\alpha(r-q)+\alpha(q-1-\beta q)/p-1} \int_0^{\infty} t^{q(p-1+\beta)} H(t) \mathbf{1}_{t^{p/\alpha} \le n} dt \ll \int_0^{\infty} t^{pr-1} H(t) dt \ll \|X\|_{pr}^{pr} < \infty \,. \tag{6.53}$$

6.6 Proof of Theorem 4.2

Recall that

$$\mathcal{D}_p(\mathcal{R}_{B_m}\mu_n, \mathcal{R}_{B_m}\mu) = \frac{2^p-1}{2} \sum_{\ell \ge 1} 2^{-p\ell} \sum_{F \in \mathcal{P}_\ell} \left| \frac{\mu_n(\tilde{F}_m)}{\mu_n(B_m)} - \frac{\mu(\tilde{F}_m)}{\mu(B_m)} \right| ,$$

where $\tilde{F}_m = 2^m F \cap B_m$. Let $\alpha \in (0,1)$ and

$$n_k = [e^{k^{1-\alpha}}] \text{ and } m_k = n_{k+1} - n_k \,.$$

Note that $m_k \sim (1-\alpha) k^{-\alpha} n_k$, as $k \to \infty$. Setting

$$\mu_{n_k,n} = \frac{1}{n-n_k} \sum_{i=n_k+1}^{n} \delta_{X_i} \,,$$

we infer that, for $n_k + 1 \le n < n_{k+1}$,

$$\begin{aligned} &\frac{\mu_n(\tilde{F}_m)}{\mu_n(B_m)} - \frac{\mu(\tilde{F}_m)}{\mu(B_m)} \\ &\qquad = \frac{n_k(\mu_{n_k}(\tilde{F}_m) - \mu(\tilde{F}_m))}{n\mu_n(B_m)} + \frac{(n-n_k)(\mu_{n_k,n}(\tilde{F}_m) - \mu(\tilde{F}_m))}{n\mu_n(B_m)} + \frac{\mu(B_m) - \mu_n(B_m)}{\mu_n(B_m)\mu(B_m)} \mu(\tilde{F}_m) \,. \end{aligned}$$

Taking into account that, for any positive measure ν,

$$\sum_{\ell \ge 1} 2^{-p\ell} \sum_{F \in \mathcal{P}_\ell} \nu(\tilde{F}_m) \le (2^p-1)^{-1} \nu(B_m) \,,$$

simple algebras lead to the following inequality: for $n_k + 1 \le n < n_{k+1}$,

$$\begin{aligned} \sum_{\ell \ge 1} 2^{-p\ell} \sum_{F \in \mathcal{P}_\ell} \frac{n_k |\mu_{n_k}(\tilde{F}_m) - \mu(\tilde{F}_m)|}{n\mu_n(B_m)} &\le \sum_{\ell \ge 1} 2^{-p\ell} \sum_{F \in \mathcal{P}_\ell} \left| \frac{\mu_{n_k}(\tilde{F}_m)}{\mu_{n_k}(B_m)} - \frac{\mu(\tilde{F}_m)}{\mu(B_m)} \right| \\ &\quad + \frac{2}{2^p-1} \frac{|\mu_{n_k}(B_m) - \mu_n(B_m)|}{\mu_n(B_m)} + \frac{1}{2^p-1} \frac{|\mu_{n_k}(B_m) - \mu(B_m)|}{\mu_n(B_m)} \,. \end{aligned}$$

Similarly, for $n_k+1\le n<n_{k+1}$,

$$\sum_{\ell\ge1}2^{-p\ell}\sum_{F\in\mathcal{P}_\ell}\frac{(n-n_k)|\mu_{n_k,n}(\tilde F_m)-\mu(\tilde F_m)|}{n\mu_n(B_m)}\le\frac{(n-n_k)}{n}\sum_{\ell\ge1}2^{-p\ell}\sum_{F\in\mathcal{P}_\ell}\left|\frac{\mu_{n_k,n}(\tilde F_m)}{\mu_{n_k,n}(B_m)}-\frac{\mu(\tilde F_m)}{\mu(B_m)}\right|$$
$$+\frac{2(n-n_k)}{(2^p-1)n}\frac{|\mu_{n_k,n}(B_m)-\mu_n(B_m)|}{\mu_n(B_m)}+\frac{(n-n_k)}{(2^p-1)n}\frac{|\mu_{n_k,n}(B_m)-\mu(B_m)|}{\mu_n(B_m)}\,.$$

So overall, for $n_k+1\le n<n_{k+1}$,

$$\mathcal{D}_p(\mathcal{R}_{B_m}\mu_n,\mathcal{R}_{B_m}\mu)\le\mathcal{D}_p(\mathcal{R}_{B_m}\mu_{n_k},\mathcal{R}_{B_m}\mu)+\frac{(n-n_k)}{n}\mathcal{D}_p(\mathcal{R}_{B_m}\mu_{n_k,n},\mathcal{R}_{B_m}\mu)$$
$$+\frac{|\mu_{n_k}(B_m)-\mu_n(B_m)|}{\mu_n(B_m)}+\frac12\frac{|\mu_{n_k}(B_m)-\mu(B_m)|}{\mu_n(B_m)}+\frac12\frac{|\mu_n(B_m)-\mu(B_m)|}{\mu_n(B_m)}$$
$$+\frac{(n-n_k)}{n}\frac{|\mu_{n_k,n}(B_m)-\mu_n(B_m)|}{\mu_n(B_m)}+\frac{(n-n_k)}{2n}\frac{|\mu_{n_k,n}(B_m)-\mu(B_m)|}{\mu_n(B_m)}\,.\quad(6.54)$$

• If $p>d/2$, let

$$v_n=\sqrt{\frac{\log\log n}{n}}\quad\text{and}\quad V=\int_0^\infty t^{p-1}\sqrt{H(t)}dt\,.$$

Starting from (6.2) and (6.54), it follows that

$$\max_{n_k+1\le n\le n_{k+1}}\frac{\mathcal{D}_p(\mu_n,\mu)}{v_n}\le\sum_{m\ge0}2^{pm}\mu(B_m)\frac{\mathcal{D}_p(\mathcal{R}_{B_m}\mu_{n_k},\mathcal{R}_{B_m}\mu)}{v_{n_{k+1}}}$$
$$+\max_{n_k+1\le n\le n_{k+1}}\frac{(n-n_k)}{n}\sum_{m\ge0}2^{pm}\mu(B_m)\frac{\mathcal{D}_p(\mathcal{R}_{B_m}\mu_{n_k,n},\mathcal{R}_{B_m}\mu)}{v_{n_{k+1}}}$$
$$+\max_{n_k\le n\le n_{k+1}}\Big(3+\frac{(n-n_k)}{n}\Big)\sum_{m\ge0}2^{pm}\frac{|\mu_n(B_m)-\mu(B_m)|}{v_{n_{k+1}}}$$
$$+\max_{n_k+1\le n\le n_{k+1}}\frac{3(n-n_k)}{2n}\sum_{m\ge0}2^{pm}\frac{|\mu_{n_k,n}(B_m)-\mu(B_m)|}{v_{n_{k+1}}}\,.$$

Using the Rosenthal inequality (with the constants given in (4.2) of Theorem 4.1 in [16]), as in the proof of Theorem 2.2, we get that there exist positive universal constants c_1, c_2 and c_3 such that for any $q>2$, any $M>0$ and $\lambda>0$, and any $\varepsilon>0$,

$$\mathbb{P}\left(\max_{n_k\le n\le n_{k+1}}\sum_{m\ge0}2^{pm}|\mu_n(B_m)-\mu(B_m)|\ge(\lambda+\varepsilon)Vv_{n_{k+1}}\right)$$
$$\ll\left(\frac{c_1}{\lambda Vn_kv_{n_{k+1}}}\right)^q q^{q/2}n_{k+1}^{q/2}V^q+\left(\frac{c_2}{\lambda Vn_kv_{n_{k+1}}}\right)^q q^q n_{k+1}\left(\int_0^\infty t^{p-1}(H(t))^{1/q}\mathbf{1}_{t\le M}dt\right)^q$$
$$+\frac{c_3n_{k+1}}{\varepsilon Vn_kv_{n_{k+1}}}\int_0^\infty t^{p-1}H(t)\mathbf{1}_{t>M}dt\,.$$

Let $\kappa > 1$ and choose $q = q_k = \gamma \log\log n_k$ with γ such that

$$\frac{c_1\sqrt{\gamma}}{\lambda} = 1/\kappa \text{ and } \gamma \log \kappa > 1\,.$$

With this choice of q_k, it follows that

$$\sum_{k\geq K_0} q_k \left(\frac{c_1}{\lambda n_k v_{n_{k+1}}}\right)^{q_k} q_k^{q_k/2} n_{k+1}^{q_k/2} < \infty\,.$$

On the other hand, by Hölder's inequality, setting $\beta = p - 2p/q$,

$$\left(\int_0^\infty t^{p-1}(H(t))^{1/q}\mathbf{1}_{t\leq M}dt\right)^q \leq p^{-q}2^{q-1} + 2^{q-1}M^{pq-2p}\beta^{1-q}\int_0^\infty t^{2p-1}H(t)dt\,.$$

Choose now

$$M = M_k = a\left(\frac{n_k}{\log\log n_k}\right)^{1/(2p)}$$

with a such that

$$\frac{4c_2 a^p \gamma}{\varepsilon V p} = 1/\kappa\,.$$

Let K_1 be such that $q_{K_1} \geq 4$. It follows that

$$\sum_{k\geq K_1}\left(\frac{c_2}{\varepsilon V n_k v_{n_{k+1}}}\right)^q q^q n_{k+1}\left(\int_0^\infty t^{p-1}(H(t))^{1/q}\mathbf{1}_{t\leq M}dt\right)^q$$
$$\ll \sum_{k\geq K_1}\left(\frac{4c_2a^p\gamma}{\varepsilon V p}\right)^{q_k} n_{k+1}\frac{\log\log n_k}{n_k} \ll \sum_{k\geq K_1}\kappa^{-q_k}\log\log n_k < \infty\,.$$

With similar arguments, one can prove that, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\max_{n_k+1\leq n\leq n_{k+1}}\frac{n-n_k}{n}\sum_{m\geq 0}2^{pm}|\mu_{n_k,n}(B_m) - \mu(B_m)| \geq \varepsilon v_{n_{k+1}}\right) < \infty\,.$$

By the direct part of the Borel-Cantelli lemma, it follows that, almost surely,

$$\limsup_{k\to\infty}\max_{n_k+1\leq n\leq n_{k+1}}\frac{\mathcal{D}_p(\mu_n,\mu)}{v_n} \leq \lambda V + \limsup_{k\to\infty}\sum_{m\geq 0}2^{pm}\mu(B_m)\frac{\mathcal{D}_p(\mathcal{R}_{B_m}\mu_{n_k},\mathcal{R}_{B_m}\mu)}{v_{n_{k+1}}}$$
$$+\limsup_{k\to\infty}\max_{n_k+1\leq n\leq n_{k+1}}\frac{(n-n_k)}{n}\sum_{m\geq 0}2^{pm}\mu(B_m)\frac{\mathcal{D}_p(\mathcal{R}_{B_m}\mu_{n_k,n},\mathcal{R}_{B_m}\mu)}{v_{n_{k+1}}}\,. \quad (6.55)$$

Let now

$$s_k = \left\lceil \frac{k^{1-\alpha}}{p\ln 2}\right\rceil\,.$$

Note that

$$\sum_{m\geq s_k+2}2^{pm}\mu(B_m)\mathcal{D}_p(\mathcal{R}_{B_m}\mu_{n_k},\mathcal{R}_{B_m}\mu) \leq \sum_{m\geq s_k+2}2^{pm}\mu(B_m)$$
$$\leq \tilde{C}_p\int_{2^{s_k}}^\infty t^{p-1}H(t)dt \leq \tilde{C}_p 2^{-ps_k}\int_0^\infty t^{2p-1}H(t)dt\,. \quad (6.56)$$

It follows that

$$\lim_{k\to\infty} \sum_{m\geq s_k+1} 2^{pm}\mu(B_m)\frac{\mathcal{D}_p(\mathcal{R}_{B_m}\mu_{n_k},\mathcal{R}_{B_m}\mu)}{v_{n_{k+1}}} = 0 \quad a.s. \tag{6.57}$$

Next, let

$$b_m = \int_{2^{m-2}}^{2^{m-1}} t^{p-1}\sqrt{H(t)}dt\mathbf{1}_{m\geq 2} + \int_0^1 t^{p-1}\sqrt{H(t)}dt\mathbf{1}_{m\leq 2} \text{ and } B = \sum_{m\geq 0} b_m = V + \int_0^1 t^{p-1}\sqrt{H(t)}dt\,.$$

Note that

$$\mathbb{P}\left(\sum_{m=0}^{s_k+1} 2^{pm}\mu(B_m)\mathcal{D}_p(\mathcal{R}_{B_m}\mu_{n_k},\mathcal{R}_{B_m}\mu) \geq CBv_{n_{k+1}}\right)$$
$$\leq \sum_{m=0}^{s_k+1} \mathbb{P}\left(2^{pm}\mu(B_m)\mathcal{D}_p(\mathcal{R}_{B_m}\mu_{n_k},\mathcal{R}_{B_m}\mu) \geq Cb_m v_{n_{k+1}}\right)\,.$$

Proceeding as in the proof of Theorem 2 in [13] (case $p > d/2$), and noting that

$$\mu(B_m) \leq \mathbb{P}(|X| > 2^{m-1}) \leq \left(\frac{1}{2^{m-2}}\int_{2^{m-2}}^{2^{m-1}} \sqrt{H(t)}dt\right)^2$$
$$\leq \left(\frac{1}{2^{(m-2)p}}\int_{2^{m-2}}^{2^{m-1}} t^{p-1}\sqrt{H(t)}dt\right)^2 = 2^{4p}2^{-2mp}b_m^2\,, \tag{6.58}$$

we derive that

$$\mathbb{P}\left(\sum_{m=0}^{s_k+1} 2^{pm}\mu(B_m)\mathcal{D}_p(\mathcal{R}_{B_m}\mu_{n_k},\mathcal{R}_{B_m}\mu) \geq CBv_{n_{k+1}}\right)$$
$$\leq \sum_{m=0}^{s_k+1} \exp\left(-aC^2b_m^2 n_k v_{n_k}^2/(2^{2pm}\mu(B_m))\right)$$
$$\leq (s_k+2)\exp\left(-\frac{aC^2}{2^{4p}}\log\log n_k\right) \ll \frac{k^{1-\alpha}}{k^{(1-\alpha)aC^2/2^{4p}}} < \infty\,, \tag{6.59}$$

for C large enough. Starting from (6.65) and (6.57), it follows that

$$\limsup_{k\to\infty} \sum_{m\geq 0} 2^{pm}\mu(B_m)\frac{\mathcal{D}_p(\mathcal{R}_{B_m}\mu_{n_k},\mathcal{R}_{B_m}\mu)}{v_{n_{k+1}}} \leq CB \quad a.s. \tag{6.60}$$

On another hand, using (6.56), we get that

$$\limsup_{k\to\infty} \max_{n_k+1\leq n\leq n_{k+1}} \frac{(n-n_k)}{n} \sum_{m\geq s_k+2} 2^{pm}\mu(B_m)\frac{\mathcal{D}_p(\mathcal{R}_{B_m}\mu_{n_k,n},\mathcal{R}_{B_m}\mu)}{v_{n_{k+1}}} = 0 \quad a.s. \tag{6.61}$$

In addition, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\max_{n_k+1\le n\le n_{k+1}} \frac{(n-n_k)}{n}\sum_{m=0}^{s_k+1} 2^{pm}\mu(B_m)\mathcal{D}_p(\mathcal{R}_{B_m}\mu_{n_k,n},\mathcal{R}_{B_m}\mu) \ge \varepsilon B v_{n_{k+1}}\right)$$
$$\le \sum_{n=n_k+1}^{n_{k+1}}\sum_{m=0}^{s_k+1}\mathbb{P}\left(\frac{(n-n_k)}{n}2^{pm}\mu(B_m)\mathcal{D}_p(\mathcal{R}_{B_m}\mu_{n_k},\mathcal{R}_{B_m}\mu) \ge \varepsilon b_m v_{n_{k+1}}\right)$$
$$\le 2e(1-\alpha)k^{-\alpha}n_k\sum_{m=0}^{s_k+1}\mathbb{P}\left(2e(1-\alpha)k^{-\alpha}2^{pm}\mu(B_m)\mathcal{D}_p(\mathcal{R}_{B_m}\mu_{n_k},\mathcal{R}_{B_m}\mu) \ge \varepsilon b_m v_{n_{k+1}}\right).$$

Proceeding as before and setting $\kappa = 2^{-4p}aC^2/(2e(1-\alpha))^2$, we get that, for any $\varepsilon > 0$,

$$\sum_{k\ge 4}\mathbb{P}\left(\max_{n_k+1\le n\le n_{k+1}} \frac{(n-n_k)}{n}\sum_{m=0}^{s_k+1} 2^{pm}\mu(B_m)\mathcal{D}_p(\mathcal{R}_{B_m}\mu_{n_k,n},\mathcal{R}_{B_m}\mu) \ge \varepsilon B v_{n_{k+1}}\right)$$
$$\ll \sum_{k\ge 4} k^{-\alpha}n_k\sum_{m=0}^{s_k+1}\exp\left(-aC^2k^{2\alpha}b_m^2 n_k v_{n_k}^2/(2^{2pm}\mu(B_m))\right)$$
$$\ll \sum_{k\ge 4} k^{-\alpha}k^{1-\alpha}e^{k^{1-\alpha}}\exp\left(-\kappa k^{2\alpha}\log\log n_k\right) < \infty\,, \quad (6.62)$$

provided we choose α such that $1-\alpha \le 2\alpha$. This proves that, almost surely,

$$\lim_{k\to\infty}\max_{n_k+1\le n\le n_{k+1}}\frac{n-n_k}{n}\sum_{m=0}^{s_k+1}2^{pm}\mu(B_m)\frac{\mathcal{D}_p(\mathcal{R}_{B_m}\mu_{n_k,n},\mathcal{R}_{B_m}\mu)}{v_{n_{k+1}}} = 0\,. \quad (6.63)$$

Starting from (6.55) and taking into account (6.60), (6.61) and (6.63), it follows that there exists an universal constant C_p depending on p such that

$$\limsup_{n\to\infty}\frac{\mathcal{D}_p(\mu_n,\mu)}{v_n} \le C_p V \quad \text{a.s.} \quad (6.64)$$

To conclude the case $p > d/2$, it suffices to use inequality (6.1).

• If $p \in [1, d/2)$, we proceed as for $p > d/2$, choosing now

$$v_n = \left(\frac{\log\log n}{n}\right)^{p/d}.$$

We start again from (6.54). All the terms can be handled similarly as in the previous case except for the terms studied in (6.65) and (6.62). For instance, proceeding again as in the proof of Theorem 2 in [13] (case $p < d/2$), we get

$$\mathbb{P}\left(\sum_{m=0}^{s_k+1}2^{pm}\mu(B_m)\mathcal{D}_p(\mathcal{R}_{B_m}\mu_{n_k},\mathcal{R}_{B_m}\mu) \ge CBv_{n_{k+1}}\right)$$
$$\le \sum_{m=0}^{s_k+1}\exp\left(-aC^{d/p}(\log\log n_k)\mu(B_m)\left(\frac{b_m}{2^{pm}\mu(B_m)}\right)^{d/p}\right).$$

The term studied in (6.62) can be handled similarly, and the result follows by taking into account that $(\mu(B_m))^{1-p/d} \le (\mu(B_m))^{1/2}$ and inequality (6.58).

6.7 Proof of Theorem 5.1

Let $q \in (1,2]$. From (6.4) and (6.5), we get the upper bound

$$\frac{1}{n}\left\|\max_{1\le k\le n} kW_p^p(\mu_k,\mu)\right\|_1 \le \frac{C}{n}\left(\left\|\max_{1\le k\le n} kB_{p,M}(\mu_k,\mu)\right\|_1 + \left\|\max_{1\le k\le n} kA_{p,M}(\mu_k,\mu)\right\|_q\right).$$

Using (6.9), (6.10) and (6.14), it follows that

$$\frac{1}{n}\left\|\max_{1\le k\le n} kW_p^p(\mu_k,\mu)\right\|_1 \ll \int_0^\infty t^{p-1}H(t)\mathbf{1}_{t>M}dt + \sum_{m\ge 0} 2^{pm}\left(\mu(B_m\cap\mathcal{C}_M)\right)^{1/q}\sum_{\ell\ge 0} 2^{-p\ell}\min\left(1, n^{-(q-1)/q}2^{\ell d(q-1)/q}\right). \quad (6.65)$$

Then we conclude as in Subsection 6.1 by considering the three cases $p > d(q-1)/q$, $p = d(q-1)/q$ and $p < d(q-1)/q$.

6.8 Proof of Theorem 5.2

From (6.4), we have that

$$\frac{1}{n}\left\|\max_{1\le k\le n} kW_p^p(\mu_k,\mu)\right\|_2 \le \frac{C}{n}\left\|\max_{1\le k\le n} k\Delta_p(\mu_k,\mu)\right\|_2$$

From (6.9) and (6.10) with $M=\infty$, we get the upper bound

$$\frac{1}{n}\left\|\max_{1\le k\le n} kW_p^p(\mu_k,\mu)\right\|_2 \ll \sum_{m\ge 0} 2^{pm}\left(\mu(B_m)\right)^{1/2}\sum_{\ell\ge 0} 2^{-p\ell}\min\left(1, 2^{\ell d/2}/\sqrt{n}\right).$$

Then we conclude as in Subsection 6.1 by considering the three cases $p > d/2$, $p = d/2$ and $p < d/2$.

6.9 Proof of Theorem 5.4

Let $r > 2$. Starting from (6.27), we infer that, for any positive constant v_n,

$$\frac{1}{n^r}\left\|\max_{1\le k\le n} kW_p^p(\mu_k,\mu)\right\|_r^r \le v_n^r + r\int_{v_n}^\infty x^{r-1}\mathbb{P}\left(\max_{1\le k\le n} kW_p^p(\mu_k,\mu) > nx\right)dx, \quad (6.66)$$

and we use the upper bound (6.6) to deal with the probability in (6.66). Let now $y = x/2C$. By Markov's inequality at order $q > r$ and 2,

$$\mathbb{P}\left(\max_{1\le k\le n} kA_{p,M}(\mu_k,\mu) > ny\right) \le \frac{\|\max_{1\le k\le n} kA_{p,M}(\mu_k,\mu)\|_q^q}{n^q y^q}, \quad (6.67)$$

$$\mathbb{P}\left(\max_{1\le k\le n} kB_{p,M}(\mu_k,\mu) > ny\right) \le \frac{\|\max_{1\le k\le n} kB_{p,M}(\mu_k,\mu)\|_2^2}{n^2 y^2}. \quad (6.68)$$

To deal with (6.68), we proceed as to get (6.11), and we obtain

$$\mathbb{P}\left(\max_{1\le k\le n} kB_{p,M}(\mu_k,\mu) > ny\right) \ll \frac{1}{y^2}\left(\sum_{m\ge 0} 2^{pm}\left(\mu(B_m\cap\mathcal{C}_M^c)\right)^{1/2}\sum_{\ell\ge 0} 2^{-p\ell}\min\left(1, 2^{\ell d/2}/\sqrt{n}\right)\right)^2 . \tag{6.69}$$

Let us now handle (6.67). We first note that

$$\left\|\max_{1\le k\le n} kA_{p,M}(\mu_k,\mu)\right\|_q \ll \sum_{m\ge 0} 2^{pm}\sum_{\ell\ge 0} 2^{-p\ell}\left\|\max_{1\le k\le n}\sum_{F\in\mathcal{P}_\ell}\left|k\mu_k(2^mF\cap B_m\cap\mathcal{C}_M) - k\mu(2^mF\cap B_m\cap\mathcal{C}_M)\right|\right\|_q .$$

Now, clearly

$$\left\|\max_{1\le k\le n}\sum_{F\in\mathcal{P}_\ell}\left|k\mu_k(2^mF\cap B_m\cap\mathcal{C}_M) - k\mu(2^mF\cap B_m\cap\mathcal{C}_M)\right|\right\|_q \le \left\|\max_{1\le k\le n}\left(k\mu_k(B_m\cap\mathcal{C}_M) + k\mu(B_m\cap\mathcal{C}_M)\right)\right\|_q \le 2n\left(\mu(B_m\cap\mathcal{C}_M)\right)^{1/q} . \tag{6.70}$$

On the other hand, by using a maximal version of Rosenthal's inequality (see for instance [16]),

$$\left\|\max_{1\le k\le n}\left|k\mu_k(2^mF\cap B_m\cap\mathcal{C}_M) - k\mu(2^mF\cap B_m\cap\mathcal{C}_M)\right|\right\|_q \ll \sqrt{n}\left(\mu(2^mF\cap B_m\cap\mathcal{C}_M)\right)^{1/2} + n^{1/q}\left(\mu(2^mF\cap B_m\cap\mathcal{C}_M)\right)^{1/q} ,$$

so that, by using Hölder's inequality (twice) and the fact that $|\mathcal{P}_\ell| = 2^{\ell d}$,

$$\sum_{F\in\mathcal{P}_\ell}\left\|\max_{1\le k\le n}\left|k\mu_k(B_m\cap\mathcal{C}_M) - k\mu(2^mF\cap B_m\cap\mathcal{C}_M)\right|\right\|_q \le 2^{\ell d/2}\sqrt{n}\left(\mu(B_m\cap\mathcal{C}_M)\right)^{1/2} + 2^{\ell d(q-1)/q}n^{1/q}\left(\mu(B_m\cap\mathcal{C}_M)\right)^{1/q} .$$

So, overall,

$$\left\|\max_{1\le k\le n} kA_{p,M}(\mu_k,\mu)\right\|_q \ll n(I_1+I_2) , \tag{6.71}$$

where

$$I_1 = \sum_{m\ge 0} 2^{pm}\sum_{\ell\ge 0} 2^{-p\ell}\min\left(\left(\mu(B_m\cap\mathcal{C}_M)\right)^{1/q}, n^{-1/2}2^{\ell d/2}\left(\mu(B_m\cap\mathcal{C}_M)\right)^{1/2}\right)$$

and

$$I_2 = \sum_{m\ge 0} 2^{pm}\sum_{\ell\ge 0} 2^{-p\ell}\left(\mu(B_m\cap\mathcal{C}_M)\right)^{1/q}\min\left(1, n^{-(q-1)/q}2^{\ell d(q-1)/q}\right) .$$

Combining (6.67), (6.70) and (6.71), we obtain that

$$\mathbb{P}\left(\max_{1\le k\le n} kA_{p,M}(\mu_k,\mu) > ny\right) \ll \frac{(I_1+I_2)^q}{y^q}\,. \tag{6.72}$$

From (6.72), we see that four cases arise:

• $p > d(r-1)/r$. In that case $p > d/2$, and

$$I_1 \le n^{-1/2}\sum_{m\ge 0} 2^{pm}\left(\mu(B_m\cap\mathcal{C}_M)\right)^{1/2} \ll n^{-1/2}\int_0^\infty t^{p-1}\sqrt{H(t)}\mathbf{1}_{t\le M}dt\,.$$

Consequently

$$\int_{v_n}^\infty x^{r-1-q}I_1^q dx \ll n^{-q/2}v_n^{r-q}\left(\int_0^\infty t^{p-1}\sqrt{H(t)}dt\right)^q\,.$$

Choosing $v_n = n^{-1/2}\int_0^\infty t^{p-1}\sqrt{H(t)}dt$, we get

$$\int_{v_n}^\infty x^{r-1-q}I_1^q dx \ll n^{-r/2}\left(\int_0^\infty t^{p-1}\sqrt{H(t)}dt\right)^r\,. \tag{6.73}$$

Let us now deal with the term involving I_2. First, we choose q close enough to r in such a way that $p > d(q-1)/q$. In that case

$$I_2 \le n^{-(q-1)/q}\sum_{m\ge 0} 2^{pm}\left(\mu(B_m\cap\mathcal{C}_M)\right)^{1/q} \ll n^{-(q-1)/q}\int_0^\infty t^{p-1}(H(t))^{1/q}\mathbf{1}_{t\le M}dt\,.$$

Let $M = (nx)^{1/p}$. Arguing as in (6.30) with $\beta < (q-1)/q$, we get

$$\int_0^\infty x^{r-1-q}I_2^q dx \ll \frac{n^{(q-1-\beta q)/p}}{n^{q-1}}\int_0^\infty t^{q(p-1+\beta)}H(t)\int_0^\infty x^{r-1-q}x^{(q-1-\beta q)/p}\mathbf{1}_{x\ge t^p/n}\,dx\,dt\,.$$

Taking β close enough to $(q-1)/q$ in such a way that $r-q+(q-1-\beta q)/p < 0$, we get that

$$\int_0^\infty x^{r-1-q}I_2^q dx \ll n^{-(r-1)}\int_0^\infty t^{rp-1}H(t)dt \ll n^{-(r-1)}\|X\|_{rp}^{rp}\,. \tag{6.74}$$

From (6.72), (6.73) and (6.74), we get that

$$\int_{v_n}^\infty x^{r-1}\mathbb{P}\left(\max_{1\le k\le n} kA_{p,M}(\mu_k,\mu) > nx/(2C)\right)dx \ll n^{-r/2}\left(\int_0^\infty t^{p-1}\sqrt{H(t)}dt\right)^r + n^{-(r-1)}\|X\|_{rp}^{rp}\,. \tag{6.75}$$

In the same way, since $p > d/2$, we infer from (6.69) that

$$\mathbb{P}\left(\max_{1\le k\le n} kB_{p,M}(\mu_k,\mu) > ny\right) \ll \frac{1}{nx^2}\left(\sum_{m\ge 0} 2^{pm}\left(\mu(B_m\cap\mathcal{C}_M^c)\right)^{1/2}\right)^2$$
$$\ll \frac{1}{nx^2}\left(\int_0^\infty t^{p-1}\sqrt{H(t)}\mathbf{1}_{t>M}dt\right)^2\,.$$

Proceeding again as in (6.30) with $\beta > 1/2$, we infer that

$$\int_{v_n}^{\infty} x^{r-1}\mathbb{P}\left(\max_{1\le k\le n} kB_{p,M}(\mu_k,\mu) > nx/(2C)\right)dx \\ \ll \frac{n^{(1-2\beta)/p}}{n}\int_0^{\infty} t^{2(p-1+\beta)}H(t)\int_0^{\infty} x^{r-3}x^{(1-2\beta)/p}\mathbf{1}_{x<t^p/n}\,dx\,dt\,.$$

Taking β close enough to $1/2$ in such a way that $(r-2)+(1-2\beta)/p > 0$, we get that

$$\int_{v_n}^{\infty} x^{r-1}\mathbb{P}\left(\max_{1\le k\le n} kB_{p,M}(\mu_k,\mu) > nx/(2C)\right)dx \ll n^{-(r-1)}\int_0^{\infty} t^{rp-1}H(t)dt \ll n^{-(r-1)}\|X\|_{rp}^{rp}\,. \quad (6.76)$$

Finally, starting from (6.66) with $v_n = n^{-1/2}\int_0^{\infty} t^{p-1}\sqrt{H(t)}dt$, and gathering (6.6), (6.75) and (6.76), the inequality is proved in the case where $p > d(r-1)/r$.

• $d/2 < p \le d(r-1)/r$. In that case we use the upper bound (6.73) without any changes. Let us now deal with the term involving I_2. Starting from the definition of I_2, and considering the two cases where $2^\ell < n^{1/d}$ and $2^\ell \ge n^{1/d}$, we infer that

$$I_2 \ll n^{-p/d}\sum_{m\ge 0} 2^{pm}\left(\mu(B_m\cap\mathcal{C}_M)\right)^{1/q} \ll n^{-p/d}\int_0^{\infty} t^{p-1}(H(t))^{1/q}\mathbf{1}_{t\le M}dt\,.$$

Let $M = (nx)^{1/p}/u_n$ for some sequence of positive numbers $(u_n)_{n>0}$. Arguing as in (6.74), we get

$$\int_0^{\infty} x^{r-1-q}I_2^q dx \ll n^{q-r-pq/d}u_n^{p(r-q)}\int_0^{\infty} t^{rp-1}H(t)dt \ll n^{q-r-pq/d}u_n^{p(r-q)}\|X\|_{rp}^{rp}\,. \quad (6.77)$$

In the same way, arguing as to get (6.76),

$$\int_{v_n}^{\infty} x^{r-1}\mathbb{P}\left(\max_{1\le k\le n} kB_{p,M}(\mu_k,\mu) > nx/(2C)\right)dx \\ \ll n^{-(r-1)}u_n^{p(r-2)}\int_0^{\infty} t^{rp-1}H(t)dt \ll n^{-(r-1)}u_n^{p(r-2)}\|X\|_{rp}^{rp}\,. \quad (6.78)$$

Now $n^{q-r-pq/d}u_n^{(r-q)p} = n^{-(r-1)}u_n^{p(r-2)}$ iff $u_n^p = n^{-1/(q-2)}n^{(1-p/d)q/(q-2)}$. With this choice of u_n and taking $q = r+\varepsilon$, we have

$$n^{q-r-pq/d}u_n^{(r-q)p} = n^{-rp/d}(n^{(d-p)/d}/u_n^p)^{q-r} = n^{-rp/d}n^{(2p-d)(q-r)/(d(q-2))} = n^{-rp/d}n^{\varepsilon(2p-d)/(d(r-2+\varepsilon))}\,.$$

Hence, with this choice of u_n, the upper bounds (6.73), (6.77) and (6.78) give the desired inequality for $d/2 < p \le d(r-1)/r$.

• $p < d/2$. Note first that, by homogeneity, the general inequality may be deduced from the case where $\|X\|_{rp} = 1$ by considering the variables $X_i/\|X\|_{rp}$. Hence, from now, we shall assume that $\|X\|_{rp} = 1$.

Let $M = (nx)^{1/p}/u_n$ for some sequence of positive numbers $(u_n)_{n>0}$. We first note that, since $q > d/(d-p)$, the upper bound (6.77) holds. Taking $u_n = n^{1/p}/n^{1/d}$, we get

$$\int_0^{\infty} x^{r-1-q}I_2^q dx \ll n^{-rp/d}\,. \quad (6.79)$$

Let us now deal with the term involving I_1. Starting from the definition of I_1, and considering the two cases where

$$2^\ell < n^{1/d}(\mu(B_m \cap \mathcal{C}_M))^{(2-q)/(dq)} \quad \text{and} \quad 2^\ell \geq n^{1/d}(\mu(B_m \cap \mathcal{C}_M))^{(2-q)/(dq)}\,,$$

we infer that

$$I_1 \ll n^{-p/d} \sum_{m\geq 0} 2^{pm} \left(\mu(B_m \cap \mathcal{C}_M)\right)^{(d+p(q-2))/(dq)} \ll n^{-p/d} \int_0^\infty t^{p-1} (H(t))^{(d+p(q-2))/(dq)} dt\,. \tag{6.80}$$

We choose now $q > r$ such that $d + p(q-2))/(dq) > 1/r$ (this is true whatever q if $p \geq d/r$, otherwise we need to choose $r < q < r(d-2p)/(d-rp)$). Since $\|X\|_{rp} = 1$, $H(t) \leq \min(1, t^{-rp})$, which together (6.80) and the choice of q implies that $I_1 \ll n^{-p/d}$. Consequently, taking $v_n = n^{-p/d}$,

$$\int_{v_n}^\infty x^{r-1-q} I_1^q \, dx \ll n^{-qp/d} v_n^{r-q} \ll n^{-rp/d}\,. \tag{6.81}$$

From (6.72), (6.79) and (6.81), we get that

$$\int_{v_n}^\infty x^{r-1} \mathbb{P}\left(\max_{1\leq k\leq n} k A_{p,M}(\mu_k, \mu) > nx/(2C)\right) dx \ll n^{-rp/d}\,. \tag{6.82}$$

On another hand, since $p < d/2$, we infer from (6.69) that

$$\mathbb{P}\left(\max_{1\leq k\leq n} k B_{p,M}(\mu_k, \mu) > ny\right) \ll \frac{1}{n^{2p/d} x^2} \left(\sum_{m\geq 0} 2^{pm} \left(\mu(B_m \cap \mathcal{C}_M^c)\right)^{1/2}\right)^2$$
$$\ll \frac{1}{n^{2p/d} x^2} \left(\int_0^\infty t^{p-1} \sqrt{H(t)} \mathbf{1}_{t>M} dt\right)^2.$$

Proceeding again as in (6.78), we get

$$\int_{v_n}^\infty x^{r-1} \mathbb{P}\left(\max_{1\leq k\leq n} k B_{p,M}(\mu_k, \mu) > nx/(2C)\right) dx \ll \frac{u_n^{p(r-2)}}{n^{((r-2)d+2p)/d}} \int_0^\infty t^{rp-1} H(t) dt \ll n^{-rp/d}\,, \tag{6.83}$$

the last inequality being true because $u_n = n^{1/p}/n^{1/d}$ and $\|X\|_{rp} = 1$.

Finally, starting from (6.66) with $v_n = n^{-p/d}$, and gathering (6.6), (6.82) and (6.83), the inequality is proved in the case where $p < d/2$ and $\|X\|_{rp} = 1$.

• $p = d/2$. Again, without loss of generality we can assume that $\|X\|_{rd/2} = 1$. We proceed as before to handle the term $\int_{v_n}^\infty x^{r-1} \mathbb{P}(\max_{1\leq k\leq n} k A_{p,M}(\mu_k, \mu) > nx) dx$. We take $q > r$ and use the Rosenthal inequality. We then have

$$I_1 \ll n^{-1/2} \log n \left(\int_0^\infty t^{d/2-1} \sqrt{H(t)} \mathbf{1}_{t\leq M} dt\right) + n^{-1/2} \left(\int_0^\infty t^{d/2-1} \sqrt{H(t)} \log(1/H(t)) \mathbf{1}_{t\leq M} dt\right).$$

Therefore, if we choose

$$v_n \geq n^{-1/2} \max\left(\log n \int_0^\infty t^{d/2-1} \sqrt{H(t)} dt, \int_0^\infty t^{d/2-1} \sqrt{H(t)} \log(1/H(t)) dt\right) =: v_n(1)\,, \tag{6.84}$$

we get

$$\int_{v_n}^{\infty} x^{r-1-q} I_1^q dx \ll n^{-r/2} (\log n)^r \left(\int_0^{\infty} t^{d/2-1} \sqrt{H(t)} dt \right)^r + n^{-r/2} \left(\int_0^{\infty} t^{d/2-1} \sqrt{H(t)} \log(1/H(t)) dt \right)^r .$$

Since $H(t) \leq \min(1, t^{-rd/2})$, it follows that

$$\int_{v_n}^{\infty} x^{r-1-q} I_1^q dx \ll n^{-r/2} (\log n)^r \left(\int_0^{\infty} t^{d/2-1} \sqrt{H(t)} dt \right)^r + n^{-r/2} . \tag{6.85}$$

On another hand, we have

$$\int_{v_n}^{\infty} x^{r-1-q} I_2^q dx \ll n^{-q/2} \int_{v_n}^{\infty} x^{r-1-q} \left(\int_0^{\infty} t^{d/2-1} H^{1/q}(t) \mathbf{1}_{t \leq M} dt \right)^q dx .$$

and choosing

$$M = (nx)^{2/d}/u_n \text{ with } u_n = n^{1/d} ,$$

we get, by taking into account previous computations, that

$$\int_{v_n}^{\infty} x^{r-1-q} I_2^q dx \ll n^{-r/2} \int_0^{\infty} t^{rp-1} H(t) dt = n^{-r/2} . \tag{6.86}$$

We handle now the quantity $\int_{v_n}^{\infty} x^{r-1} \mathbb{P}(\max_{1 \leq k \leq n} k B_{p,M}(\mu_k, \mu) > nx) dx$. We shall apply this time the Rosenthal inequality as we did to handle $\| \max_{1 \leq k \leq n} k B_{p,M}(\mu_k, \mu) \|_q$, but with $q \in (2, r)$. We obtain

$$\| B_{p,M}(\mu_n, \mu) \|_q \ll J_1 + J_2 + J_3 , \tag{6.87}$$

with

$$J_1 = n^{-1/2} \log n \int_0^{\infty} t^{d/2-1} \sqrt{H(t)} \mathbf{1}_{t > M} dt ,$$

$$J_2 = n^{-1/2} \int_0^{\infty} t^{d/2-1} \sqrt{H(t)} \log(1/H(t)) \mathbf{1}_{t > M} dt ,$$

and

$$J_3 = n^{-1/2} \int_0^{\infty} t^{d/2-1} H^{1/q}(t) \mathbf{1}_{t > M} dt .$$

Note that since $M = (nx)^{2/d}/u_n$ with $u_n = n^{1/d}$, applying of Hölder's inequality as in previous computations, we have

$$\int_{v_n}^{\infty} x^{r-1-q} J_3^q dx \ll n^{-r/2} \int_0^{\infty} t^{rd/2-1} H(t) dt . \tag{6.88}$$

On another hand, using that $H(t) \leq \min(1, t^{-rd/2})$, we have (since $r > 2$ and $M^{d/2} = x\sqrt{n}$),

$$\int_{v_n}^{\infty} x^{r-1-q} J_1^q dx \leq n^{-q/2} (\log n)^q \int_{v_n}^{\infty} x^{r-1-q} \left(\int_M^{\infty} t^{d/2-1} t^{-rd/4} dt \right)^q dx$$
$$\ll n^{-rq/4} (\log n)^q \int_{v_n}^{\infty} x^{r(1-q/2)-1} dx \ll n^{-rq/4} (\log n)^q v_n^r v_n^{-rq/2} .$$

Therefore if

$$v_n \geq n^{-1/2}(\log n)^{2/r} =: v_n(2)\,, \tag{6.89}$$

we get

$$\int_{v_n}^{\infty} x^{r-1-q} J_1^q dx \ll n^{-r/2}(\log n)^2\,. \tag{6.90}$$

We handle now the term involving J_2. We have

$$\int_{v_n}^{\infty} x^{r-1-q} J_2^q dx = n^{-q/2} \int_{v_n}^{\infty} x^{r-1-q} \left(\int_0^{\infty} t^{d/2-1} \sqrt{H(t)} \log(1/H(t)) \mathbf{1}_{t>M} dt \right)^q dx\,.$$

Using that $H(t) \leq \min(1, t^{-rd/2})$, simple computations lead to

$$\int_{v_n}^{\infty} x^{r-1-q} J_2^q dx \ll n^{-rq/4} v_n^{r(1-q/2)} (\log v_n)^q\,.$$

Therefore, if (6.89) holds, we get

$$\int_{v_n}^{\infty} x^{r-1-q} J_2^q dx \ll n^{-r/2}(\log n)^2\,. \tag{6.91}$$

So finally if we choose

$$v_n = \max(v_n(1), v_n(2))\,,$$

the constraints (6.84) and (6.89) are satisfied. Starting from (6.66), and gathering the bounds (6.6), (6.85), (6.86), (6.87), (6.88), (6.90), and (6.91), we get the desired inequality in the case $\|X\|_{rd/2} = 1$.

References

[1] M. Ajtai, J. Komlos and G. Tusnady, (1983). On optimal matching. *Combinatorica* **4** 259-264.

[2] F. Bach and J. Weed (2017). Sharp asymptotic and finite sample rates of convergence of empirical measures in Wasserstein distance. arXiv:1707.00087v1

[3] B. von Bahr and C.-G. Esseen, (1965). Inequalities for the rth absolute moment of a sum of random variables, $1 < r \leq 2$. *Ann. Math. Statist.* **36** 299-303.

[4] 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** (1999) 1009-1071.

[5] E. del Barrio, E. Giné and F. Utzet, (2005). Asymptotics for $\mathbb{L}^2$ functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli* **11** 131-189.

[6] S. Bobkov and M. Ledoux, (2016). One-dimensional empirical measures, order statistics and Kantorovich transport distances. To appear in the Memoirs of the Amer. Math. Soc.

[7] E. Boissard and T. Le Gouic, (2014). On the mean speed of convergence of empirical and occupation measures in Wasserstein distance. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 539563.

[8] 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.

[9] S. Dereich, M. Scheutzow and R. Schottstedt, (2013). Constructive quantization: approximation by empirical measures. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 1183-1203.

[10] V. Dobrić and J. E. Yukich, (1995). Asymptotics for transportation cost in high dimensions. *J. Theoret. Probab.* **8** 97-118.

[11] R. M. Dudley, (1968). The speed of mean Glivenko-Cantelli convergence. *Ann. Math. Statist.* **40** 40-50.

[12] S. S. Èbralidze, (1971). Inequalities for the probabilities of large deviations in terms of pseudomoments, (Russian). *Teor. Verojatnost. i Primenen* **16** 760-765.

[13] N. Fournier and A. Guillin, (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Relat. Fields* **162** 707-738.

[14] 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.

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

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

[17] M. Talagrand, (1992). Matching random samples in many dimensions. *Ann. Appl. Probab.* **2** 846-856.