ONE-DIMENSIONAL EMPIRICAL MEASURES, ORDER STATISTICS, AND KANTOROVICH TRANSPORT DISTANCES

Sergey Bobkov and Michel Ledoux
University of Minnesota* and University of Toulouse†

May 3, 2020

Abstract. This work is devoted to the study of rates of convergence of the empirical measures $\mu_n = \frac{1}{n}\sum_{k=1}^n \delta_{X_k}$, $n \geq 1$, over a sample $(X_k)_{k\geq 1}$ of independent identically distributed real-valued random variables towards the common distribution μ in Kantorovich transport distances W_p. The focus is on finite range bounds on the expected Kantorovich distances $\mathbb{E}(W_p(\mu_n,\mu))$ or $\big[\mathbb{E}(W_p^p(\mu_n,\mu))\big]^{1/p}$ in terms of moments and analytic conditions on the measure μ and its distribution function. The study describes a variety of rates, from the standard one $\frac{1}{\sqrt{n}}$ to slower rates, and both lower and upper-bounds on $\mathbb{E}(W_p(\mu_n,\mu))$ for fixed n in various instances. Order statistics, reduction to uniform samples and analysis of beta distributions, inverse distribution functions, log-concavity are main tools in the investigation. Two detailed appendices collect classical and some new facts on inverse distribution functions and beta distributions and their densities necessary to the investigation.

Mathematics Subject Classification (2010). Primary 60B10, 60F99, 60G57, 62G30, 60B12; Secondary 62G20.

*School of Mathematics, University of Minnesota, Minneapolis, MN 55455 USA, bobkov@math.umn.edu

†Institut de Mathématiques de Toulouse, Université de Toulouse, F-31062 Toulouse, France, and Institut Universitaire de France, ledoux@math.univ-toulouse.fr

Contents

1 Introduction

This work is devoted to an in-depth investigation of orders of growth of Kantorovich transport distances for one-dimensional empirical measures.

Let X be a real-valued random variable on some probability space $(\Omega, \Sigma, \mathbb{P})$, with law (distribution) μ (which defines a Borel probability measure on $\mathbb{R}$) and distribution function

$$F(x) = \mu\big((-\infty, x]\big), \qquad x \in \mathbb{R}.$$

Consider a sequence $(X_k)_{k\geq 1}$ of independent copies of X thus with the same distribution μ, and, for each $n \geq 1$, the (random) empirical measure

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

where δ_x is Dirac mass at the point $x \in \mathbb{R}$. Denote by F_n the distribution function of μ_n,

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{(-\infty,x]}(X_k), \qquad x \in \mathbb{R}.$$

The classical limit theorems by Glivenko-Cantelli and Donsker ensure respectively that, almost surely,

$$\sup_{x\in\mathbb{R}} \big|F_n(x) - F(x)\big| \to 0$$

and, weakly in the Skorokhod topology,

$$\sqrt{n}\,\big(F_n(x) - F(x)\big) \to W^o\big(F(x)\big), \qquad x \in \mathbb{R},$$

where W^o is a Brownian bridge (on $[0,1]$).

This work is concerned with rates of convergence in the Kantorovich[1] distances W_p, $p \geq 1$, of the empirical measures μ_n towards the theoretical distribution μ. The Kantorovich transport distance $W_p(\mu_n, \mu)$, $p \geq 1$, between μ_n and μ is defined by

$$W_p^p(\mu_n, \mu) = \inf_{\pi} \int_{\mathbb{R}}\int_{\mathbb{R}} |x-y|^p \, d\pi(x,y),$$

[1]In the literature, the distance W_p is also called the Monge-Kantorovich, or Kantorovich-Rubinshtein, or Wasserstein transport distance, as well as the Fréchet distance (in case $p = 2$), or a minimal distance. Recently, Vershik [Ve1] wrote an interesting historic essay explaining why it is more fair to fix the name "Kantorovich distance" for all metrics like W_p (calling them Kantorovich power metrics) according to the original reference [Ka1]. Some general topological properties of W_1 were studied in 1970 by Dobrushin [Do], who re-introduced this metric with reference to [Vas]; apparently, that is why the name "Wasserstein distance" has become rather traditional. As Vershik writes, "Leonid Vasershtein is a famous mathematician specializing in algebraic K-theory and other areas of algebra and analysis, and ... he is absolutely not guilty of this distortion of terminology, which occurs primarily in Western literature". It should be noted that the notation W for the quantities like W_p is the one used by Kantorovich in [Ka1], keeping therefore a balance with the nowadays terminology!

where the infimum is taken over all probability measures π on the product space $\mathbb{R} \times \mathbb{R}$ with respective marginals μ_n and μ.

More precisely, we focus in this work on the possible behaviour of the expected Kantorovich distance $\mathbb{E}(W_p(\mu_n, \mu))$ as a function of n, where $p \geq 1$ is given. Note that this distance is finite as long as

$$\int_{-\infty}^{\infty} |x|^p d\mu(x) \,=\, \mathbb{E}\big(|X|^p\big) \,<\, \infty,$$

in which case it will be shown below that $W_p(\mu_n, \mu) \to 0$ with probability one. The rates at which $\mu_n \to \mu$ in W_p depends on a variety of hypotheses and properties on the underlying distribution μ discussed here as completely as possible.

As such, these questions were only partially studied in the literature (as far as we can tell). The asymptotic behaviour of $W_p(\mu_n, \mu)$ for $p = 1$ and 2 has been investigated previously in papers by del Barrio, Giné, Matrán [B-G-M] and del Barrio, Giné, Utzet [B-G-U], providing in particular necessary and sufficient conditions for the weak convergence of $W_p(\mu_n, \mu)$ (for these values of p) towards integrals of the Brownian bridge under some regularity conditions on μ. The purpose of the present work is rather the study of finite range bounds (that is, for $n \geq 1$ large but fixed), both upper and lower-bounds, on the expected Kantorovich distances $\mathbb{E}(W_p(\mu_n, \mu))$ or $\mathbb{E}(W_p^p(\mu_n, \mu))$ for all $p \geq 1$ and under fairly general assumptions on the distribution μ. The functional central limit theorem $\sqrt{n}\,(F_n(x) - F(x)) \to W^o(F(x))$ already indicates that under proper assumptions the value of $\mathbb{E}(W_p(\mu_n, \mu))$ should have the rate of order $\frac{1}{\sqrt{n}}$ (which is in general best possible). Therefore, we will be in particular interested in conditions that ensure this "standard" rate. We next present the various parts and summarize some of the main conclusions obtained here.

The first section (Section 2) collects a number of standard results on the Kantorovich transport distances W_p and the topology that they generate. Quantile representations of W_p on the real line are also addressed there. The last paragraph gathers some basic facts on the convergence of empirical measures in W_p over a sample of independent identically distributed random variables towards the common distribution.

Section 3 is devoted to the Kantorovich distance $W_1(\mu_n, \mu)$. It is shown in particular that if $\mathbb{E}(|X|) < \infty$, then $\mathbb{E}(W_1(\mu_n, \mu)) \to 0$, but the convergence may actually hold at an arbitrarily slow rate. On the other hand, the convergence rate cannot be better than $\frac{1}{\sqrt{n}}$. This standard rate is reached under the moment condition $\mathbb{E}(|X|^{2+\delta}) < \infty$ for some $\delta > 0$. In fact, a necessary and sufficient condition for the standard rate is that

$$J_1(\mu) \,=\, \int_{-\infty}^{\infty} \sqrt{F(x)(1 - F(x))}\, dx \,<\, \infty.$$

Morever, explicit two-sided bounds, depending on n, for $\mathbb{E}(W_1(\mu_n, \mu))$ in terms of the distribution function F may be provided. Connections with functional limit theorems are also addressed.

Section 4 investigates general order statistics and quantile representations of $W_p(\mu_n, \mu)$ which will be useful in the case $p > 1$. The classical reduction to the uniform distribution via inverse distribution functions is presented, leading in particular to representations of $W_p(\mu_n, \mu)$ in terms of beta distributions. On this basis, a complete description of the rates for $\mathbb{E}(W_p^p(\mu_n, \mu))$ when μ is uniform may be obtained.

The next Section 5 describes some main results. In particular, it will be proved that for the property (standard rate)

$$\big[\,\mathbb{E}\big(W_p^p(\mu_n, \mu)\big)\big]^{1/p} \,\leq\, \frac{c}{\sqrt{n}}$$

to hold with some constant $c > 0$, it is necessary and sufficient that μ be supported and have an almost everywhere (a.e.) positive density f on some interval of the real line (for the absolutely continuous component of μ), with finite integral

$$J_p(\mu) \,=\, \int_{-\infty}^{\infty} \frac{[F(x)(1-F(x))]^{p/2}}{f(x)^{p-1}}\,dx \,=\, \int_0^1 \bigg(\frac{\sqrt{t(1-t)}}{I(t)}\bigg)^p\,dt.$$

In this case, the (generalized) inverse function F^{-1} of F has to be absolutely continuous on $(0,1)$ in the local sense, and then

$$I(t) \,=\, I_F(t) \,=\, \frac{1}{(F^{-1})'(t))}\,, \qquad 0 < t < 1,$$

with $(F^{-1})'$ being understood as a Radon-Nikodym derivative. Such I-functions will play an important role in the analysis throughout this work. In fact $I(t) = f\big(F^{-1}(t)\big)$ a.e. (where the density f should be specially defined when μ is not absolutely continuous with respect to Lebesgue measure).

To reach this result, we first bound from above $\mathbb{E}(W_p^p(\mu_n, \mu))$ by the functional $J_p(\mu)$. We then present analytic assumptions on the I-function $I_F(t)$, $0 < t < 1$, in order for the latter to be finite. For example, if $I(t) \geq c\sqrt{t(1-t)}$ for some constant $c > 0$, or equivalently, if μ represents a Lipschitz transform of the beta distribution with parameters $\alpha = \beta = 2$, then $J_p(\mu) < \infty$, for all $p \geq 1$. If $I(t) \geq ct(1-t)$ for some constant $c > 0$, i.e. μ has a positive Cheeger's constant, or equivalently, if μ is a Lipschitz transform of the two-sided exponential distribution, then $J_p(\mu) < \infty$, for all $1 \leq p < 2$.

To show the necessity of the condition $J_p(\mu) < \infty$ for the standard rate to hold, we first establish the connectedness of μ and the absolute continuity of the inverse distribution function F^{-1}. A further study of beta distributions (which is postponed to Appendix B) will allow us to reach a lower bound

$$\liminf_{n\to\infty} \sqrt{n}\;\mathbb{E}\big(W_p(\mu_n, \mu)\big) \,\geq\, c\,J_p^{1/p}(\mu)$$

holding with some absolute constant $c > 0$. The section is concluded by a first study of the standard rate for the distance W_∞.

Quite a bit of work is devoted to the family of log-concave measures μ as presented in Section 6. This study relies in particular on precise bounds on variances of order statistics in particular via the associated I-functions. As a sample of results, it will be established in this case that the value $\mathbb{E}(W_p^p(\mu_n,\mu))$ is approximately given by

$$\frac{1}{n^{p/2}} \int_{1/(n+1)}^{n/(n+1)} \bigg(\frac{\sqrt{t(1-t)}}{I(t)}\bigg)^p\,dt.$$

Bounds in terms of the variance of μ may also be achieved. They will imply, for instance, that $\big[\mathbb{E}(W_p^p(\mu_n,\mu))\big]^{1/p} = O(\frac{1}{\sqrt{n}})$ whenever $1 \leq p < 2$, and $\big[\mathbb{E}(W_p^p(\mu_n,\mu))\big]^{1/p} = O(\frac{1}{n^{1/p}})$ for $p > 2$. In addition,

$$\big[\mathbb{E}\big(W_2^2(\mu_n,\mu)\big)\big]^{1/2} \,=\, O\bigg(\sqrt{\frac{\log n}{n}}\,\bigg).$$

Nevertheless, $\big[\mathbb{E}(W_2^2(\mu_n,\mu))\big]^{1/2} = O(\frac{1}{\sqrt{n}})$ for compactly supported log-concave μ. A variety of examples, from Gaussian to beta distributions, illustrate the conclusions.

Section 7 collects miscellaneous bounds and results, supplementing the preceding conclusions. If μ satisfies a Poincaré-type inequality, one can control deviations of $W_p(\mu_n,\mu)$ from the mean $\mathbb{E}(W_p(\mu_n,\mu))$. In particular in this case, $\big[\mathbb{E}(W_p^p(\mu_n,\mu))\big]^{1/p}$ and $\mathbb{E}\,W_p(\mu_n,\mu)$ are of the same order, whenever $1 \leq p \leq 2$. They are of the same order also for $p > 2$ if μ is log-concave.

If μ is compactly supported on an interval of length c, then for any $p \geq 1$,

$$\big[\,\mathbb{E}\big(W_p^p(\mu_n,\mu)\big)\big]^{1/p} \,\leq\, \frac{c}{n^{1/2p}}$$

(which may not be improved in this class). If the support of μ is not an interval, then, for any $p > 1$, there is the lower-bound

$$\mathbb{E}\big(W_p(\mu_n,\mu)\big) \,\geq\, \frac{c}{n^{1/2p}}\,,$$

where the constant does not depend on n. These results indicate in particular that the standard rate $\big[\,\mathbb{E}(W_p^p(\mu_n,\mu))\big]^{1/p} = O(\frac{1}{\sqrt{n}})$ cannot be obtained under moment-type conditions. On the other hand, by developing ideas of Ebralidze [Eb] providing moment bounds on the Kantorovich distance W_p, it may be shown that, for every $p \geq 1$,

$$\mathbb{E}\big(W_p^p(\mu_n,\mu)\big) \,=\, O\bigg(\frac{1}{\sqrt{n}}\bigg)$$

as soon as $\mathbb{E}(|X|^s) < \infty$ for some $s > 2p$.

The particular case $p = \infty$ deserves some special investigation. In addition to the conclusions in Section 5, it may be proved that $\mathbb{E}(W_\infty(\mu_n,\mu)) \to 0$ if and only if the support of μ is a finite closed interval. The same property characterizes the convergence in probability and almost surely. For the standard rate to hold, it is necessary and sufficient that μ be supported and have a density on some finite interval, which is separated from zero.

The results developed in this work require a somewhat in-depth analysis of distribution functions and their inverses, as well as of several regularity properties. Another main tool is provided by order statistics, and, after reduction to the uniform distribution, to refined properties of beta distributions and their densities. The two appendices aims at fully supplementing all the analytical results necessary to the investigation and results presented in the core of the text. Appendix A is devoted to inverse functions, and the associated I-functions. Appendix B on the other hand is concerned with various aspects of beta distributions, from their log-concavity and spectral gap properties to refined bounds on their densities. While a number of results contained in these appendices are classical, some of them are new. Besides, we found it convenient to collect all these conclusions in a coherent way towards the results developed in the core of the text.

It should be mentioned that we do not address in this work the corresponding analysis for samples of vector-valued random variables $(X_k)_{k\geq 1}$ which leads to delicate questions and methods of different nature, as illustrated for examples in the papers [T-Y], [T1], [B-G-V], [B-B], [B-M1], [B-M2]... Nevertheless, the recent works [D-S-S] and [F-G] achieve moment estimates in higher dimension suitably extending the conclusions of sub-Section 7.5. Besides, the articles [B-M1], [B-M2] investigate some of these questions in the context of partial transport for which one-dimensional results might be transferred to higher dimension. Note that there is also an intensive literature devoted to the study of empirical measures for dependent data. For example, the papers [Ki], [Se], [Y] focus on the approximation problems of μ_n by μ under mixing conditions on the sequence $(X_k)_{k\geq 1}$, while [B-G] develops an investigation under analytic conditions on the distribution of the sample. Another fruitful modern direction deals with spectral empirical measures. They correspond to dependent observations $(X_k)_{k\geq 1}$ that appear as spectra of large random matrices. See e.g. [A-G-Z], [P-S], [B-G-T], [C-L], [Dal], [M-M].

The standard probabilistic data of this work, common to most sections, will be a real-valued random variable X on some probability space $(\Omega, \Sigma, \mathbb{P})$ with (law) distribution μ, and distribution function $F(x) = \mu((-\infty, x])$, $x \in \mathbb{R}$. The probability measure μ is said to be degenerate if it is a Dirac mass. If X has a second moment, the variance $\mathbb{E}(X^2) - (\mathbb{E}(X))^2$ is denoted by $\mathrm{Var}(X)$. A median m of X (or μ) is a real number such that $\mathbb{P}\{X \leq m\} \geq \frac{1}{2}$ and $\mathbb{P}\{X \geq m\} \geq \frac{1}{2}$.

Given a sequence $(X_k)_{k\geq 1}$, of independent copies of X, μ_n is the (random) empirical measure

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

with associated (random) distribution function

$$F_n(x) \,=\, \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{(-\infty,x]}(X_k), \qquad x \in \mathbb{R}.$$

We usually denote by U a uniform random variable on $(0,1)$, with associated sample $(U_k)_{k\geq 1}$.

Given a Borel probability measure μ with distribution function F, several analytic statements will involve the inverse distribution function

$$F^{-1}(t) \,=\, \inf\big\{x \in \mathbb{R} : F(x) \geq t\big\}, \qquad 0 < t < 1,$$

as well as the associated I-functions

$$I(t) \,=\, I_F(t) \,=\, \frac{1}{(F^{-1})'(t)}\,, \qquad 0 < t < 1.$$

The function I_F is well-defined a.e. as long as F^{-1} is absolutely continuous on $(0,1)$, and then $(F^{-1})'$ denotes the corresponding Radon-Nikodym derivative. We refer to Appendix A for complete details in this regard.

As is common, we abbreviate "almost everywhere" in "a.e." (with respect to Lebesgue measure on $\mathbb{R}$) and "almost surely" in "a.s.".

Acknowledgement. We thank A. Guillin, E. Rio, J. Wellner for pointing out several related references. This work started in June 2011 during the visit of the first author at the University of Toulouse, France, and he is grateful for the hospitality. The work was continued in the Fall 2013 when both authors stayed at the Simons Institute, Berkeley, USA. The authors would like to thank the Institute for this opportunity. The work of the first author was also supported by the Simons Foundation and NSF grant DMS-1855575.

2 Generalities on Kantorovich transport distances

This section introduces the basic Kantorovich transport distances W_p and their first properties. The latter are first discussed in the general framework of metric spaces. Specific representations on the real line are emphasized next via inverse distribution functions (or quantiles). In the last paragraph, convergence of empirical measures over a sample of independent identically distributed random variables in Kantorovich distances are presented.

General references on Kantorovich transport distances mostly covering the material summarized here are the books [R-R] by Rachev and Rüschendorf and [Vi1], [Vi2] by Villani, to which we refer for more details. See also [Du1] and [S-W], [C-H] in the context of empirical measures and quantile processes.

2.1 Kantorovich transport distance W_p

Given a separable metric space (E, d), let $Z(E, d)$ denote the space of all Borel probability measures μ on E. For $p \geq 1$, denote by $Z_p(E, d)$, or just $Z_p(E)$ when the underlying metric d is clear from the context, the collection of all probability measures μ in $Z(E, d)$ such that

$$\int_E d(x, x_0)^p \, d\mu(x) \, < \, \infty$$

for some, or equivalently all, $x_0 \in E$. The space $Z_p(E, d)$ may be described as the family of all Borel probability measures μ on E with a finite p-th moment.

Definition 2.1 (Kantorovich transport distance W_p)**.** For $\mu, \nu \in Z(E, d)$, define the quantity $W_p(\mu, \nu) \geq 0$ by

$$W_p^p(\mu, \nu) \, = \, \inf_\pi \int_E \int_E d(x, y)^p \, d\pi(x, y),$$

where the infimum is taken over all probability measures π on the product space $E \times E$ (equipped with the product σ-algebra) with marginals μ and ν in the sense that

$$\pi(A \times E) \, = \, \mu(A), \qquad \pi(E \times B) \, = \, \nu(B),$$

for all Borel sets $A, B \subset E$. The value $W_p(\mu, \nu)$ will be called the Kantorovich (transport) distance of order p between μ and ν.

Most important particular orders are $p = 1$ and $p = 2$. The value W_2 may be referred to as the quadratic Kantorovich distance. Of a certain interest is also the limit case

$$W_\infty(\mu, \nu) \, = \, \lim_{p \to \infty} W_p(\mu, \nu) \, = \, \sup_{p \geq 1} W_p(\mu, \nu),$$

which is natural to be considered in the space $Z_\infty(E,d)$ of all Borel probability measures on E with bounded support.

The number $W_p^p(\mu,\nu)$ is often interpreted as the minimal cost needed to transport the measure μ to ν, provided that the cost for transportation of a "particle" $x \in E$ to any "particle" $y \in E$ is equal to $d(x,y)^p$. Definition 2.1 is usually applied to the measures in $Z_p(E,d)$, which guarantees that $W_p(\mu,\nu)$ is finite. Note that if one of these measures has a finite p-th moment, and $W_p(\mu,\nu)$ is finite, then necessarily the other measure must have a finite p-th moment.

It is known that W_p is indeed a metric in the space $Z_p(E,d)$, at least when (E,d) is separable and complete (Polish). The proof of the triangle inequality for W_p is based on the so-called glueing lemma, which is discussed in detail in [Du2], pp. 7-10, cf. also [Vi1], pp. 207-208.

Example 2.2. For $x,y \in E$, consider the mass points or delta-measures $\mu = \delta_x$, $\nu = \delta_y$. Then it is easy to see that

$$W_p(\delta_x,\delta_y) \ = \ d(x,y).$$

Hence, (E,d) is isometrically embedded in each $Z_p(E,d)$ via the map $x \to \delta_x$.

Example 2.3. For $x,y \in E$ and $0 \leq a,b \leq 1$, consider the Bernoulli measures

$$\mu \ = \ a\delta_x + (1-a)\delta_y, \qquad \nu \ = \ b\delta_x + (1-b)\delta_y.$$

Then

$$W_p(\mu,\nu) \ = \ |a-b|^{1/p}\, d(x,y).$$

This formula is of course consistent with the previous example of delta-measures in the particular case $a = 1$, $b = 0$. To prove the above equality in the general case, assume that $x \neq y$. Let π be a probability measure on $E \times E$ with marginals μ and ν. Necessarily, it is concentrated on the 4 point set $\{(x,x),(x,y),(y,x),(y,y)\}$, assigning some probabilities, say, π_1, π_2, π_3, π_4, respectively. The requirement that π has marginals μ and ν is equivalent to the relations

$$\begin{aligned} \pi_1 + \pi_2 \ &= \ \mu\big(\{x\}\big) \ = \ a, \\ \pi_1 + \pi_3 \ &= \ \nu\big(\{x\}\big) \ = \ b, \end{aligned}$$

under which we need to minimize the integral

$$\int_E \int_E d(x,y)^p\, d\pi(x,y) \ = \ (\pi_2 + \pi_3)\, d(x,y)^p.$$

If $a \geq b$, the minimum is obviously attained for $\pi_3 = 0$, $\pi_1 = b$ and $\pi_2 = a - b$, and in this case it is equal to $(a-b)\, d(x,y)^p$.

Remark 2.4. It also natural to consider more general transport functionals, such as

$$W(\mu,\nu) \;=\; \inf_{\pi} \int_E \int_E c(x,y)\, d\pi(x,y)$$

where c is a given non-negative (cost) function on $E \times E$, and the infimum is taken over all probability measures π on $E \times E$ with respective marginals μ and ν, as before. In connection with the problem of mass transportation, such functionals were introduced in 1942 by Kantorovich in [Ka1], cf. also [Ka2], [K-A]. Assuming that (E,d) is compact and $c = d$ (and using the same notation W), he proved that the above infimum is attained at some probability measure π_0, which is characterized by the following properties: There exists a function U on E (called a potential) such that

a) $|U(x) - U(y)| \le d(x,y)$ for all $x, y \in E$;

b) $U(x) - U(y) = d(x,y)$ for all x, y in the support of π_0.

As a direct consequence,

$$W(\mu,\nu) \;=\; \int_E U\, d\mu - \int_E U\, d\nu \;=\; \max \left| \int_E u\, d\mu - \int_E u\, d\nu \right|,$$

where the maximum is over all functions u obeying the property a).

In particular, the functional W gives rise to a metric. As Kantorovich remarks in [Ka1], this way to introduce the metric in the space of distributions looks most natural. In a bit more general setting, part of his theorem may be stated as follows.

Theorem 2.5 (Kantorovich duality theorem)**.** *Given a separable metric space* (E,d)*, for all* $\mu, \nu \in Z_1(E,d)$*,*

$$W_1(\mu,\nu) \;=\; \sup_{\|u\|_{\mathrm{Lip}} \le 1} \left| \int_E u\, d\mu - \int_E u\, d\nu \right|$$

where the supremum is taken over all Lipschitz functions $u : E \to \mathbb{R}$ *with Lipschitz semi-norm* $\|u\|_{\mathrm{Lip}} \le 1$*.*

After the work [K-R] in 1958, also dealing with a compact space setting, Theorem 2.5 is now referred to as the Kantorovich-Rubinstein theorem. See [Ra], [Du1] for the proof and the history of the duality problems, and [B-K] for a recent survey on the state of the art of the research connected to the Monge-Kantorovich problem.

2.2 Topology generated by W_p

Once the transport distances are introduced, a natural question is “What topology do they generate?” In many practical situations, we actually deal with the topology of the weak convergence of probability measures.

Recall that a sequence $(\mu_n)_{n\in\mathbb{N}}$ in $Z(E,d)$ is weakly convergent to μ in $Z(E,d)$ if

$$\int_E u\,d\mu_n \to \int_E u\,d\mu \qquad \text{as } \ n\to\infty$$

for any bounded continuous function u on E. This convergence is metrized, for example, by the Lévy-Prokhorov metric, defined for all μ and ν in $Z(E,d)$ by

$$\rho(\mu,\nu) \;=\; \inf\big\{h>0 : \mu(A)\le \nu(A_h)+h \ \text{ for all Borel sets } A\subset E\big\}.$$

Here $A_h=\{x\in E : d(x,A)<h\}$ denotes the open h-neighbourhood of A with respect to the metric d.

Equivalent definitions and basic general results on the weak convergence and the associated weak topology on $Z(E,d)$ may be found in the book of Billingsley [Bi]. To clarify the meaning of the convergence with respect to W_p, we quote a theorem from [Vi1], p. 212, cf. also [Do] and [S-W].

Theorem 2.6 (Convergence in W_p)**.** *Let $1\le p<\infty$. Given $\mu\in Z_p(E,d)$ and a sequence $(\mu_n)_{n\in\mathbb{N}}\in Z_p(E,d)$ on a Polish space (E,d), the following properties are equivalent:*

a) $W_p(\mu_n,\mu)\to 0$ as $n\to\infty$;

b) $\mu_n\to\mu$ weakly and for some, or equivalently all, $x_0\in E$,

$$\int_E d(x,x_0)^p\,d\mu_n(x) \;\to\; \int_E d(x,x_0)^p\,d\mu(x);$$

c) For any continuous function $u : E\to\mathbb{R}$, satisfying the growth condition $|u(x)|\le C(1+d(x,x_0))^p$, $x\in E$, with some point $x_0\in E$ and a constant $C\ge 0$,

$$\int_E u(x)\,d\mu_n(x) \;\to\; \int_E u(x)\,d\mu(x).$$

This description shows that the convergence in (Z_p,W_p) is stronger than the standard weak convergence. Alternatively, one may use an elementary relation between the Lévy-Prokhorov and Kantorovich transport distances (which in case $p=1$ was emphasized in [Do]).

Theorem 2.7. *For all $\mu,\nu\in Z(E,d)$ and $p\ge 1$,*

$$\rho(\mu,\nu) \;\le\; \big(W_p(\mu,\nu)\big)^{p/(p+1)}.$$

In particular $\rho(\mu,\nu)\le W_\infty(\mu,\nu)$.

Proof. Suppose that $\mu(A) \geq \nu(A_h) + h$ for some $h > 0$ and some Borel set A in E. Then, for any probability measure π on $E \times E$ with marginals μ and ν,

$$\begin{aligned}\int_E \int_E d(x,y)^p \, d\pi(x,y) &\geq \int_A \int_{E \setminus A_h} d(x,y)^p \, d\pi(x,y) \\ &\geq h^p \, \pi\big(A \times E \cap E \times (E \setminus A_h)\big) \\ &\geq h^p \, \big(\mu(A) - \nu(A_h)\big) \, \geq \, h^{p+1}.\end{aligned}$$

Hence $W_p^p(\mu,\nu) \geq h^{p+1}(\mu,\nu)$ as announced. □

If the metric d is bounded, Z_p consists of all Borel probability measures on E, and each W_p metrizes the topology of the weak convergence according to Theorem 2.6. Otherwise, the topology in (Z_p, W_p) is strictly stronger. Indeed, if x_n, $n \geq 1$, and x_0 are elements of E, and if $r_n = d(x_n, x_0) \to \infty$ as $n \to \infty$, let

$$\mu_n \, = \, (1 - r_n^{-p}) \, \delta_{x_0} + r_n^{-p} \, \delta_{x_n}, \qquad \mu \, = \, \delta_{x_0}.$$

Then $\mu_n \to \mu$ weakly while $W_p(\mu_n, \mu) = 1$ for all $n \geq 1$ according to Example 2.3 so that there is no convergence with respect to W_p.

Nevertheless, by Theorem 2.6, the topology generated by W_p and the weak topology do coincide on any subset K of $Z_p(E,d)$ possessing the property

$$\lim_{R \to \infty} \sup_{\mu \in K} \int_{\{d(x,x_0) > R\}} d(x,x_0)^p \, d\mu(x) \, = \, 0$$

with $x_0 \in E$ fixed. For example, given $C > 0$ and $p_0 > p$, one may consider the set K of all Borel probability measures μ on (E,d) such that

$$\int_E d(x,x_0)^{p_0} \, d\mu(x) \, \leq \, C.$$

In particular, the metric W_1, when it is used for the set K of all probability distributions μ on $E = \mathbb{R}$ with $\int_{-\infty}^{\infty} x^2 \, d\mu(x) = 1$, generates the topology of weak convergence in K. Moreover, in this case, by Prokhorov's compactness criterion, this set will be compact for this topology.

The topology generated by W_∞ is however much stronger than the weak topology, even if they are compared on compactly supported measures. As noticed in [Ve2], this metric was in essence considered by Strassen [Str]. Using his results on nearby variables with nearby laws, one can give a description of W_∞ which is formally rather close to the Lévy-Prokhorov metric.

Theorem 2.8 (Topology generated by W_∞)**.** *For all $\mu, \nu \in Z(E,d)$ on a Polish space (E,d),*

$$W_\infty(\mu,\nu) \, = \, \inf\big\{h > 0 : \mu(A) \leq \nu(A_h) \ \ \textit{for all Borel sets } A \subset E\big\}.$$

Proof. From Definition 2.1,

$$W_\infty(\mu,\nu) \,=\, \sup_{p\geq 1} W_p(\mu,\nu) \,=\, \sup_{p\geq 1}\inf_\pi \|d\|_{L^p(\pi)} \,\leq\, \inf_\pi \|d\|_{L^\infty(\pi)} \tag{2.1}$$

where the infimum is taken over the set $K(\mu,\nu)$ of all probability measures π on $E\times E$ with marginals μ and ν (couplings), and where $\|d\|_{L^p(\pi)}$ and $\|d\|_{L^\infty(\pi)} = \operatorname{ess\,sup} d(x,y)$ are the usual L^p and L^∞ norms with respect to π.

The inequality (2.1) may actually be reversed. Since the space (E,d) is Polish, the set $K(\mu,\nu)$ is compact in $Z(E,d)$. Indeed, all finite Borel measures on E are Radon, so, given $\varepsilon>0$, one can choose a compact set $K_\varepsilon\subset E$ with $\mu(K_\varepsilon)>1-\varepsilon$ and $\nu(K_\varepsilon)>1-\varepsilon$. It then readily follows that $\pi(K_\varepsilon\times K_\varepsilon)>1-2\varepsilon$ for any $\pi\in K(\mu,\nu)$. It remains to apply Prokhorov's compactness criterion and use the property that $K(\mu,\nu)$ is a closed subset in $Z(E,d)$.

Now, assume that $W_\infty(\mu,\nu)<h$. For any $n\geq 1$, there is $\pi_n\in K(\mu,\nu)$ such that $\|d\|_{L^n(\pi_n)}<h$. In particular, for all $1\leq p\leq n$,

$$\int_0^\infty \pi_n(d>r^{1/p})\,dr \,=\, \int_E\int_E d(x,y)^p\,d\pi_n(x,y) \,<\, h^p.$$

Assume that $\pi_n\to\pi$ weakly in $K(\mu,\nu)$ (otherwise take a convergent subsequence). Since the sets $U(r)=\{(x,y): d(x,y)>r^{1/p}\}$, $r>0$, are open in $E\times E$, we have $\pi(U(r))\leq\liminf_{n\to\infty}\pi_n(U(r))$, so, by Fatou's lemma,

$$\int_E\int_E d(x,y)^p\,d\pi(x,y) \,\leq\, h^p.$$

Hence $\|d\|_{L^p(\pi)}\leq h$ for all $p\geq 1$, and thus $\|d\|_{L^p(\pi)}\leq W_\infty(\mu,\nu)$. Recalling (2.1), this shows that $W_\infty(\mu,\nu)=\inf_\pi\,\|d\|_{L^\infty(\pi)}$.

To conclude the proof, it is time to involve Strassen's theorem which we quote from [Du1], p. 319. Put $A^h=\{x\in E: d(x,A)\leq h\}$. If the metric space (E,d) is separable, and $h\geq 0$, $\beta\geq 0$ are fixed, then the following two properties are equivalent:

a) For any closed set A in E, $\mu(A)\leq\nu(A^h)+\beta$;
b) For any $\alpha>h$, there is $\pi\in K(\mu,\nu)$ such that $\pi\{d>\alpha\}\leq\beta$.

If (E,d) is Polish, the latter property with $\beta=0$ reads $\|d\|_{L^\infty(\pi)}\leq h$. Note also that $A_h=(\operatorname{clos} A)_h$ and $A_h\subset A^h\subset A_{h+\varepsilon}$ for any $\varepsilon>0$. The proof is complete. □

2.3 Representations for W_p on the real line

There are several results concerning various representations for the transport distances W_p. In the case $p=1$, the representation in the Kantorovich-Rubinstein theorem when it is specialized to the real line $E=\mathbb{R}$ may considerably be simplified and stated explicitly in terms of the distribution functions $F(x) = \mu((-\infty, x])$, $x \in \mathbb{R}$, associated to probability measures μ. More precisely, Theorem 2.5 easily yields the following description of W_1.

Theorem 2.9 (Representation for W_1)**.** *Let μ and ν be probability measures in $Z_1(\mathbb{R})$ with respective distribution functions F and G. Then*

$$W_1(\mu,\nu) \,=\, \int_{-\infty}^{\infty} \big|F(x) - G(x)\big|\, dx.$$

This formula was apparently first obtained by Dall'Aglio ([Da], cf. also [Val]).

There is also a representation for W_p of a similar nature in case $p \geq 1$. It is however given not like the L^p-distance between F and G, but involves the inverse distribution functions

$$F^{-1}(t) \,=\, \inf\big\{x \in \mathbb{R} : F(x) \geq t\big\}, \qquad 0 < t < 1. \tag{2.2}$$

The next fundamental identity can be found in [C-S-S] or [Ru], Theorem 2.

Theorem 2.10 (Representation for W_p)**.** *Let μ and ν be probability measures in $Z_p(\mathbb{R})$, $p \geq 1$, with respective distribution functions F and G. Then*

$$W_p^p(\mu,\nu) \,=\, \int_0^1 \big|F^{-1}(t) - G^{-1}(t)\big|^p\, dt.$$

In particular,

$$\int_{-\infty}^{\infty} \big|F(x) - G(x)\big|\, dx \,=\, \int_0^1 \big|F^{-1}(t) - G^{-1}(t)\big|\, dt.$$

This equality was already emphasized by Prokhorov [P].

The inverse distribution functions will be discussed in more details in Appendix A (where a simple argument leading to Theorem 2.10 will be described). At this point, let us only mention that any such function is non-decreasing and left-continuous. In particular, the infimum in the definition (2.2) of F^{-1} is always attained.

Since the inverse functions might happen to be less convenient in applications in comparison with the usual distribution functions, one may wonder whether one can give an explicit formula in the spirit of Theorem 2.9. Let us state one such formula for the important case $p = 2$, which was kindly communicated to us by E. del Barrio. We adopt the standard notations $x \wedge y = \min\{x, y\}$, $x \vee y = \max\{x, y\}$ and $x^+ = x \vee 0$ for real numbers x, y.

Theorem 2.11 (Representation for W_2)**.** *Let μ and ν be probability measures in $Z_2(\mathbb{R})$ with respective distribution functions F and G. Then*

$$\begin{aligned} W_2^2(\mu,\nu) \;&=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \big(F(x\wedge y) - G(x\vee y)\big)^+\, dx\, dy \\ &\quad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \big(G(x\wedge y) - F(x\vee y)\big)^+\, dx\, dy. \end{aligned}$$

Both integrands on the right-hand side are symmetric under the reflection $(x,y) \mapsto (y,x)$, so, one may also write a more compact expression

$$W_2^2(\mu,\nu) \,=\, 2\iint_{x\le y} \Big[\big(F(x)-G(y)\big)^+ + \big(G(x)-F(y)\big)^+\Big]\,dx\,dy.$$

Proof. Let us explain how to derive this formula from Theorem 2.10, i.e.

$$W_2^2(\mu,\nu) \,=\, \int_0^1 \big(F^{-1}(t)-G^{-1}(t)\big)^2\,dt.$$

Using a simple approximation argument (to handle the convergence in the space $Z_2(\mathbb{R})$), it may be assumed that μ and ν are regular in the sense that they are supported and have positive continuous densities f and g on some intervals, say, (a,b) and (c,d), which may be bounded or not. The inverse functions $F^{-1}:(0,1)\to(a,b)$ and $G^{-1}:(0,1)\to(c,d)$ are then well-defined in the usual sense and are continuously differentiable with positive derivatives. Moreover, in terms of the function $I(t)=f(F^{-1}(t))$, $0<t<1$,

$$F^{-1}(t)-G^{-1}(t) \,=\, F^{-1}(t)-F^{-1}(F(G^{-1}(t))) \,=\, \int_{F(G^{-1}(t))}^t \frac{du}{I(u)}$$

so that

$$(F^{-1}(t)-G^{-1}(t))^2 \,=\, \int_{F(G^{-1}(t))}^t\int_{F(G^{-1}(t))}^t \frac{du\,dv}{I(u)\,I(v)}\,.$$

Now, integrate this equality over $t\in(0,1)$, keeping the values u and v fixed. Put $R(t)=G(F^{-1}(t))$ and distinguish between the case $F(G^{-1}(t))<t$ and $F(G^{-1}(t))>t$.

Case 1: $F(G^{-1}(t))<u,v<t$. These inequalities are solved as $t<R(u\wedge v)$ and $t>u\vee v$, which represents an interval of length $(R(u\wedge v)-u\vee v)^+$ which may be empty or not.

Case 2: $t<u,v<F(G^{-1}(t))$. These inequalities are solved as $t<u\wedge v$ and $t>R(u\vee v)$, which represents and interval of length $(u\wedge v-R(u\vee v))^+$ which may also be empty.

Therefore, collecting the two cases together, we have after integration that

$$\begin{aligned}\int_0^1 \big(F^{-1}(t)-G^{-1}(t)\big)^2\,dt \;&=\; \int_0^1\int_0^1 \big(R(u\wedge v)-u\vee v\big)^+\,\frac{du\,dv}{I(u)\,I(v)}\\ &\quad+\int_0^1\int_0^1 \big(u\wedge v-R(u\vee v)\big)^+\,\frac{du\,dv}{I(u)\,I(v)}\,.\end{aligned}$$

Now, make the substitution $u=F(x)$, $v=F(y)$. Since $R(u)=G(x)$, $R(v)=G(y)$, and $I(u)=f(x)$, $I(v)=f(y)$, and thus $\frac{du}{I(u)}=dx$, $\frac{dv}{I(v)}=dy$, we get

$$\begin{aligned}\int_0^1 \big(F^{-1}(t)-G^{-1}(t)\big)^2\,dt \;&=\; \int_a^b\int_a^b \big(G(x\wedge y)-F(x\vee y)\big)^+\,dx\,dy\\ &\quad+\int_a^b\int_a^b \big(F(x\wedge y)-G(x\vee y)\big)^+\,dx\,dy\end{aligned}$$

which is the announced claim. □

For another explicit example, let $p \to \infty$ in Theorem 2.10 so to obtain the representation

$$W_\infty(\mu,\nu) \ = \ \sup_{0<t<1} \left|F^{-1}(t) - G^{-1}(t)\right|. \tag{2.3}$$

Changing the variables, we arrive at the following refinement of Theorem 2.8. Recall that $Z_\infty(\mathbb{R})$ denotes the collection of all compactly supported probability measures on $\mathbb{R}$.

Theorem 2.12 (Representation for W_∞). *Let μ and ν be probability measures in $Z_\infty(\mathbb{R})$ with respective distribution functions F and G. Then, $W_\infty(\mu,\nu)$ is the infimum over all $h \geq 0$ such that*

$$G(x-h) \ \leq \ F(x) \leq G(x+h) \qquad \text{for all } x \in \mathbb{R}.$$

The last description resembles the Lévy distance $L(\mu,\nu)$, which is defined as the infimum over all $h \geq 0$ such that

$$G(x-h) - h \ \leq \ F(x) \leq G(x+h) + h \qquad \text{for all } \ x \in \mathbb{R}.$$

This metric metrizes the weak topology in the whole space $Z(\mathbb{R})$. In view of the obvious relation $L \leq W_\infty$, the metric W_∞ is stronger. In fact, even being restricted to probability measures supported on a common finite interval, the topology generated by W_∞ is strictly stronger than the weak topology. Let, for example, μ_n, $n \geq 1$, be a sequence of probability measures on $[0,1]$ with distribution functions $F_n(x) = x^{1/n}$ $(0 \leq x \leq 1)$. Then $\mu_n \to \mu = \delta_0$ weakly, and so $L(\mu_n,\mu) \to 0$, while $W_\infty(\mu_n,\mu) = 1$ for all $n \geq 1$.

Proof of Theorem 2.12. The statement remains to hold for arbitrary probability measures μ and ν on the real line $\mathbb{R}$. To see this, one can apply the representation (2.3) for $W_\infty(\mu,\nu)$.

By the properties of inverse functions (cf. Lemma A.3 of Appendix A), $t \leq F(x)$ if and only if $F^{-1}(t) \leq x$ for all $t \in (0,1]$ and $x \in \mathbb{R}$ (with the convention that $F^{-1}(1) = F^{-1}(1-)$). In particular, $F((F^{-1}(t)) \geq t$ and $F^{-1}(F(x)) \leq x$ in case $F(x) > 0$.

Assume first that $\sup_{0<t<1} |F^{-1}(t) - G^{-1}(t)| \leq h$ for a finite value $h \geq 0$ and let us show that

$$F(x) \ \leq \ G(x+h).$$

If $F(x) = 0$, there is nothing to prove, so let $F(x) > 0$. Since $G^{-1}(t) \leq F^{-1}(t) + h$, we get $t \leq G(F^{-1}(t) + h)$. For $t = F(x)$, this gives $F(x) \leq G(x+h)$. By a similar argument, $G(x) \leq F(x+h)$, so $W_\infty(\mu,\nu) \leq h$.

Conversely, assume that $G(x-h) \leq F(x) \leq G(x+h)$ for all x. The second inequality yields $G^{-1}(F(x)) \leq x + h$ provided that $G(x+h) > 0$. This condition is fulfilled for $x = F^{-1}(t)$ with $t \in (0,1)$, since then $G(x+h) \geq F(x) \geq t$. Hence, $G^{-1}(t) \leq F^{-1}(t) + h$. Similarly, $F^{-1}(t) \leq G^{-1}(t) + h$. □

2.4 Empirical measures

In this section, we recall the basic convergence of empirical measures over a sample of independent and identically distributed random variables towards the common distribution and address the question of bounds in the Kantorovich distances W_p. We refer to [S-W], [Du1], [C-H] for standard references on the topic of convergence of empirical measures, in particular for real-valued samples and their interplay with the quantile processes.

Let X be a random element in the metric space (E, d) with law μ, and let $(X_k)_{k\geq 1}$ be a sequence of independent copies of X. Consider the empirical measures

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

which therefore define random probability measures on (E, d).

A well-known theorem of Varadarajan asserts that, if (E, d) is separable, then, with probability one, $\mu_n \to \mu$ weakly (cf. [Var] or [Du1], p. 313). Equivalently, with probability one, $\rho(\mu_n, \mu) \to 0$ as $n \to \infty$, where for ρ one may take, for example, the Lévy-Prokhorov metric in the space of all Borel probability measures on E. As a full analogue, the following assertion is also valid.

Theorem 2.13 (Convergence of empirical measures)**.** *Assume that the metric space (E, d) is separable. If $\mu \in Z_p(E, d)$ for $p \geq 1$, then, with probability one, $W_p(\mu_n, \mu) \to 0$ as $n \to \infty$.*

Indeed, using Varadarajan's theorem and Theorem 2.6, it is enough to verify that, with probability one, and any $x_0 \in E$,

$$\int_E d(x, x_0)^p \, d\mu_n(x) = \frac{1}{n}\sum_{k=1}^{n} d(X_k, x_0)^p \to \int_E d(x, x_0)^p \, d\mu(x) = \mathbb{E}\big(d(X, x_0)^p\big).$$

But the latter does hold by the strong law of large numbers completing therefore the proof of Theorem 2.13.

On the basis of Theorem 2.13, a general question of interest investigated here is the rates of convergence of the empirical measures μ_n to the limit μ with respect to Kantorovich distances W_p, with probability one or in distribution. In fact, a main focus here will be to explore bounds on the mean distances $\mathbb{E}(W_p(\mu_n, \mu))$ or $\mathbb{E}(W_p^p(\mu_n, \mu))$. These notes will be furthermore restricted to the particular, but important, scenario where E is the real line $\mathbb{R}$.

For a first natural step, it would already be good to see that these average distances do tend to zero. And indeed, Theorem 2.13 may be complemented with the following statement which, however, is not completely immediate.

Theorem 2.14 (Convergence of empirical measures in W_p). *For any $\mu \in Z_p(\mathbb{R})$ and $1 \le p < \infty$, $\mathbb{E}(W_p^p(\mu_n, \mu)) \to 0$ as $n \to \infty$.*

In particular, $\mathbb{E}(W_p(\mu_n, \mu)) \to 0$. As we will see later on, this property, as well as Theorem 2.13 are no longer true for arbitrary $\mu \in Z_\infty(\mathbb{R})$ in the limit case $p = \infty$.

To address the proof of Theorem 2.14, consider the distribution function $F(x) = \mu\big((-\infty, x]\big) = \mathbb{P}\{X \le x\}$, $x \in \mathbb{R}$, of μ (or X) and the associated empirical distribution functions on a sample $(X_k)_{k\ge1}$ of independent copies of X,

$$F_n(x) \;=\; \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\{X_k \le x\}}, \qquad x \in \mathbb{R},\; n \ge 1.$$

Proof. To begin with, we present a proof of the simpler case $p = 1$ for which we describe an additional argument separately from the general case. This argument is based on the representation

$$W_1(\mu_n, \mu) \;=\; \int_{-\infty}^{\infty} \big|F_n(x) - F(x)\big|\, dx$$

from Theorem 2.9.

Denote by $(\Omega, \Sigma, \mathbb{P})$ the probability space where all random variables X_n, $n \ge 1$, are defined. Let λ stand for the Lebesgue measure on the real line and let $\nu = \lambda \otimes \mathbb{P}$ be the product measure on the product space $\mathbb{R} \times \Omega$. Since

$$\mathbb{E}\bigg(\int_{-\infty}^{\infty} \big(F_n(x) - F(x)\big)\, dx\bigg) \;=\; \int_{-\infty}^{\infty}\int_{\Omega} \big(F_n(x,\omega) - F(x)\big)\, d\nu(x,\omega) \;=\; 0,$$

for any $a \in \mathbb{R}$,

$$\begin{aligned}
\mathbb{E}\big(W_1(\mu_n,\mu)\big) &= 2\int_{-\infty}^{\infty}\int_{\Omega} \big(F_n(x,\omega) - F(x)\big)^+\, d\nu(x,\omega) \\
&\le 2\int_{-\infty}^{a}\int_{\Omega} F_n(x,\omega)\, d\nu(x,\omega) + 2\int_{a}^{\infty}\int_{\Omega} \big(F_n(x,\omega) - F(x)\big)^+\, d\nu(x,\omega) \\
&= 2\int_{-\infty}^{a} F(x)\, dx + 2\int_{a}^{\infty}\int_{\Omega} \big(F_n(x,\omega) - F(x)\big)^+\, d\nu(x,\omega),
\end{aligned}$$

where we used that $\mathbb{E}(F_n(x)) = F(x)$ in the last step.

Now, fix $\varepsilon > 0$ and choose $a \in \mathbb{R}$ such that $\int_{-\infty}^{a} F(x)\, dx < \varepsilon$. As for the second integral over $x > a$, one may use the bound $(F_n(x,\omega) - F(x))^+ \le 1 - F(x)$. The latter function is integrable on (a, ∞) and serves as an integrable majorant on $(a, \infty) \times \Omega$ for the sequence $(F_n(x,\omega) - F(x))^+$. Hence, by the Lebesgue dominated convergence theorem,

$$\int_{a}^{\infty}\int_{\Omega} \big(F_n(x,\omega) - F(x)\big)^+\, d\nu(x,\omega) \;\to\; 0$$

as $n \to \infty$. Therefore, there exists n_0 such that for all $n \geq n_0$ all such integrals are smaller than ε. As a result, $\mathbb{E}(W_1(\mu_n, \mu)) < 4\varepsilon$ for all $n \geq n_0$. The claim follows.

Note that a different proof in this case $p = 1$ also follows from Theorem 3.5 in the next section.

We next address the general case $p \geq 1$. We start this time from the representation in Theorem 2.10 which gives, for any $n \geq 1$,

$$\mathbb{E}\big(W_p^p(\mu_n, \mu)\big) \, = \, \int_0^1 \mathbb{E}\big(\big|F_n^{-1}(t) - F^{-1}(t)\big|^p\big)\, dt.$$

First we show that, for any fixed $t_0 \in (0, \frac{1}{2})$,

$$\int_{t_0}^{1-t_0} \mathbb{E}\big(\big|F_n^{-1}(t) - F^{-1}(t)\big|^p\big)\, dt \, \to \, 0 \tag{2.4}$$

as $n \to \infty$. To this end, using the Lebesgue dominated convergence theorem and the property that $F_n^{-1}(t) \to F^{-1}(t)$ for any point t of continuity of F^{-1} (cf. Lemma A.5), it will be sufficient to see that the random variables

$$M(t) \, = \, \sup_{n \geq 1} \big|F_n^{-1}(t)\big|, \qquad t_0 \leq t \leq 1 - t_0,$$

have $L^p(\Omega, \mathbb{P})$-norms bounded by a quantity which is independent of t.

Choose $x_0 > 0$ such that $\mu([x_0, \infty)) + \mu((-\infty, -x_0]) < \frac{t_0}{2}$. Then, for all $x \geq x_0$ and $t \in [t_0, 1 - t_0]$,

$$\begin{aligned} F_n^{-1}(t) > x \; &\Longrightarrow \; 1 - F_n(x) > t \; \Longrightarrow \; F(x) - F_n(x) > \frac{t_0}{2}\,, \\ -F_n^{-1}(t) \geq x \; &\Longrightarrow \; F_n(-x) \geq t \; \Longrightarrow \; F_n(-x) - F(-x) > \frac{t_0}{2}\,. \end{aligned}$$

Hence,

$$\begin{aligned} \mathbb{P}\Big\{\sup_{n \geq 1} F_n^{-1}(t) > x\Big\} \; &\leq \; \mathbb{P}\Big\{\sup_{n \geq 1} \big(F(x) - F_n(x)\big) > \frac{t_0}{2}\Big\}, \\ \mathbb{P}\Big\{\sup_{n \geq 1} \big(-F_n^{-1}(t)\big) > x\Big\} \; &\leq \; \mathbb{P}\Big\{\sup_{n \geq 1} \big(F_n(-x) - F(-x)\big) > \frac{t_0}{2}\Big\}. \end{aligned}$$

These two bounds imply that

$$\mathbb{P}\big\{M(t) > x\big\} \, \leq \, \mathbb{P}\Big\{\sup_{n \geq 1} \big|F_n(x) - F(x)\big| > \frac{t_0}{2}\Big\} + \mathbb{P}\Big\{\sup_{n \geq 1} \big|F_n(-x) - F(-x)\big| > \frac{t_0}{2}\Big\}.$$

In order to estimate the latter two probabilities, write

$$F_n(x) - F(x) \, = \, \frac{S_n}{n}, \qquad S_n \, = \, \xi_1 + \cdots + \xi_n,$$

where $\xi_i = 1\!\!1_{\{X_i \leq x\}} - F(x)$, $i \geq 1$. One can now involve a generalized form of the Kolmogorov maximal inequality due to Hájek and Rényi,

$$\mathbb{P}\Big\{ \sup_{n \geq 1} \frac{|S_n|}{n} > r \Big\} \ \leq \ \frac{1}{r^2} \sum_{n=1}^\infty \frac{\mathbb{E}(\xi_n^2)}{n^2}\,, \qquad r > 0,$$

which holds for all independent centered random variables ξ_n, $n \geq 1$ (cf. [H-K], [Et]). In the present Bernoulli case, this maximal inequality yields

$$\mathbb{P}\big\{M(t) > x\big\} \ \leq \ \frac{2\pi^2}{3t_0^2}\,\Big[F(x)\big(1 - F(x)\big) + F(-x)\big(1 - F(-x)\big)\Big], \qquad x \geq x_0,$$

so that

$$\mathbb{E}\big(M(t)^p\big) \ = \ p \int_0^\infty x^{p-1}\,\mathbb{P}\big\{M(t) > x\big\}\,dx \ \leq \ c_0 + c_1\,\mathbb{E}\big(|X_1|^p\big)$$

with some constants independent on t. This proves the desired convergence (2.4) of integrals over the intervals $(t_0, 1 - t_0)$.

Let us now turn to the missing intervals $(0, t_0)$ and $(1 - t_0, 1)$. By Fatou's lemma,

$$\int_0^{t_0} \mathbb{E}\big(\big|F^{-1}(t)\big|^p\big)\,dt \ \leq \ \liminf_{n \to \infty} \int_0^{t_0} \mathbb{E}\big(\big|F_n^{-1}(t)\big|^p\big)\,dt,$$

and similarly for the integrals over $(1 - t_0, 1)$. Hence, it remains to show that uniformly over all large n the integrals

$$\int_0^{t_0} \mathbb{E}\big(\big|F_n^{-1}(t)\big|^p\big)\,dt, \qquad \int_{1-t_0}^1 \mathbb{E}\big(\big|F_n^{-1}(t)\big|^p\big)\,dt$$

can be made as small, as we wish, by choosing an appropriate small value of t_0. Indeed, changing the variable $t = F_n(x)$, write

$$\int_0^{t_0} \big|F_n^{-1}(t)\big|^p\,dt \ = \ \int_{\{F_n(x) \leq t_0\}} |x|^p\,dF_n(x) \ = \ \frac{1}{n} \sum_{k=1}^n |X_k|^p\,1\!\!1_{\{F_n(X_k) \leq t_0\}}.$$

The distribution of $(X_k, F_n(X_k))$ is the same for all k, and therefore

$$\int_0^{t_0} \mathbb{E}\big(\big|F_n^{-1}(t)\big|^p\big)dt \ = \ \mathbb{E}\big(|X_1|^p\,1\!\!1_{\{F_n(X_1) \leq t_0\}}\big).$$

By the Glivenko-Cantelli theorem, with probability one, $\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \to 0$ as $n \to \infty$. So, applying the Lebesgue dominated convergence theorem, we get

$$\mathbb{E}\big(|X|^p\,1\!\!1_{\{F_n(X_1) \leq t_0\}}\big) \ \to \ \mathbb{E}\big(|X|^p\,1\!\!1_{\{F(X_1) \leq t_0\}}\big) \ = \ \int_{\{F(x) \leq t_0\}} |x|^p\,dF(x).$$

Given $\varepsilon > 0$, the last integral is smaller than ε for sufficiently small $t_0 > 0$, and then

$$\int_0^{t_0} \mathbb{E}(|F_n^{-1}(t)|^p)dt < 2\varepsilon$$

for all $n \geq n_0$ with sufficiently large n_0.

By a similar argument, if t_0 is small enough,

$$\int_{1-t_0}^{1} \mathbb{E}\big(\big|F_n^{-1}(t)\big|^p\big)\, dt \;<\; 2\varepsilon \qquad \text{for all } n \geq n_0.$$

The proof of Theorem 2.14 is therefore complete. □

3 The Kantorovich distance $W_1(\mu_n, \mu)$

This section is devoted to the investigation of the Kantorovich transport distance $W_1(\mu_n, \mu)$ along the sequence of empirical measures

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

of a sample $(X_k)_{k \geq 1}$ of independent copies of a real-valued random random variable X with law μ and distribution function F. More precisely, we study the rates of the expected Kantorovich distance $\mathbb{E}(W_1(\mu_n, \mu))$, with a special emphasis for the standard rate $\frac{1}{\sqrt{n}}$. The first paragraph describes the best and worst rates. We then characterize the standard rate in terms of the functional

$$J_1(\mu) = \int_{-\infty}^{\infty} \sqrt{F(x)(1 - F(x))}\, dx. \tag{3.1}$$

In fact, this will be achieved on the basis of two-sided bounds at fixed n on $\mathbb{E}(W_1(\mu_n, \mu))$. The last paragraph compares these conclusions with former results in the context of functional limit theorems.

3.1 Best and worst rates for the means $\mathbb{E}(W_1(\mu_n, \mu))$

According to Theorem 2.14, $\mathbb{E}(W_1(\mu_n, \mu)) \to 0$ as $n \to \infty$. We will be interested here at the rate at which this convergence takes place. Before turning to this natural question, we first emphasize that the standard rate $\frac{1}{\sqrt{n}}$ is best possible (unless μ is degenerate). As usual, X denotes a random variable with law μ.

Theorem 3.1 (Best rate for $\mathbb{E}(W_1(\mu_n, \mu))$). *Under the first moment assumption, for every $n \geq 1$,*

$$\mathbb{E}\big(W_1(\mu_n, \mu)\big) \geq \frac{c}{\sqrt{n}}\, \mathbb{E}\big(|X - m|\big) \tag{3.2}$$

where m is a median of X and $c > 0$ is an absolute constant ($c = \frac{1}{2\sqrt{2}}$ is an admissible value).

The proof of Theorem 3.1 is postponed to the end of the section. Before, let us observe that the inequality (3.2) cannot be reversed in general. On the other hand, a simple sufficient condition insuring the validity of the rate $\frac{1}{\sqrt{n}}$ may be given in terms of the finiteness of the integral (3.1), which, as will be discussed below, is a stronger condition than just existence of the first moment.

Theorem 3.2 (Upper-bound on $\mathbb{E}(W_1(\mu_n, \mu))$). *For every $n \geq 1$,*

$$\mathbb{E}\big(W_1(\mu_n, \mu)\big) \leq \frac{1}{\sqrt{n}}\, J_1(\mu). \tag{3.3}$$

This bound is elementary, since by Fubini's theorem,

$$\begin{aligned}\mathbb{E}\big(W_1(\mu_n,\mu)\big) \ &= \ \int_{-\infty}^{\infty} \mathbb{E}\big(\big|F_n(x) - F(x)\big|\big)dx \\ &\le \ \int_{-\infty}^{\infty} \sqrt{\mathrm{Var}(F_n(x))}\,dx \ = \ \frac{1}{\sqrt{n}}\ J_1(\mu).\end{aligned}$$

In fact, due to the triangle inequality in the space L^2, there is a stronger bound

$$\big[\,\mathbb{E}\big(W_1^2(\mu_n,\mu)\big)\big]^{1/2} \ \le \ \frac{1}{\sqrt{n}}\ J_1(\mu).$$

Furthermore, using a Khinchine-type inequality for sums of independent Bernoulli summands, one may further extend it to L^p norms, $p \ge 1$, as

$$\big[\,\mathbb{E}\big(W_1^p(\mu_n,\mu)\big)\big]^{1/p} \ \le \ \frac{c\sqrt{p}}{\sqrt{n}}\ J_1(\mu),$$

where c is a positive absolute constant.

Returning to the basic L^1-bound of Theorem 3.2, here are a few remarks. If X is a random variable on a probability space $(\Omega, \Sigma, \mathbb{P})$ with distribution function F associated to the law μ on the real line, the finiteness of the integral $J_1(\mu)$ is equivalent to the finiteness of the functional

$$\Lambda_{2,1}(X) \ = \ \int_0^{\infty} \sqrt{\mathbb{P}\big\{|X| > x\big\}}\,dx. \tag{3.4}$$

In turn, this functional is equivalent to a norm defining the Lorentz Banach space $L^{2,1} = L^{2,1}(\Omega, \Sigma, \mathbb{P})$, dual to the weak-$L^2$ space $L^{2,\infty}$. In general, one has the inclusions $L^{2+\delta} \subset L^{2,1} \subset L^2$, where $L^p = L^p(\Omega, \Sigma, \mathbb{P})$ denote the usual Lebesgue spaces with norms $\|X\|_p = (\mathbb{E}(|X|^p))^{1/p}$. The preceding inclusions are strict. For example, the random variable X with law μ and distribution function

$$F(x) \ = \ 1 - \frac{1}{(1+x)^2\,\log^2(e+x)}\,, \qquad x \ge 0,$$

has a finite second moment, but $J_1(\mu) = \infty$ (equivalently $\Lambda_{2,1}(X) = \infty$). On the other hand, the condition $\mathbb{E}(|X|^{2+\delta}) < \infty$ with $\delta > 0$ is sufficient for the finiteness of $J_1(\mu)$.

When $J_1(\mu)$ is infinite, the means $\mathbb{E}(W_1(\mu_n,\mu))$ may decay at an arbitrary slow rate (of course at least $\frac{1}{\sqrt{n}}$ by Theorem 3.1). The following statement will be obtained as a consequence of the more general Theorem 3.5 as will be developed in the next section.

Theorem 3.3 (Worst rate for $\mathbb{E}(W_1(\mu_n,\mu))$)**.** *For any sequence of numbers $\varepsilon_n \to 0$, there exists $\mu \in Z_1(\mathbb{R})$ such that*

$$\mathbb{E}\big(W_1(\mu_n,\mu)\big) \ \ge \ \varepsilon_n$$

for all n large enough.

We conclude this sub-section with the proof of Theorem 3.1. It is based on the following classical lemma.

Lemma 3.4. *Given independent mean zero random variables $\xi_1, \dots, \xi_n$,*

$$\mathbb{E}\bigg(\Big|\sum_{k=1}^n \xi_k\Big|\bigg) \geq c\,\mathbb{E}\bigg(\Big(\sum_{k=1}^n \xi_k^2\Big)^{1/2}\bigg)$$

where $c > 0$ is an absolute constant. (One may take $c = \frac{1}{2\sqrt{2}}$.)

The lower-bound of Lemma 3.4 is standard and represents a particular case of a more general two-sided inequality due to Marcinkiewicz and Zygmund for p-th moments of sums of independent mean zero random variables, cf. [M-Z]. To obtain an explicit value of the constant c in the particular case $p = 1$, one may use a symmetrization argument. Let $\varepsilon_1, \dots, \varepsilon_n$ be independent random variables with a symmetric Bernoulli distribution, that is, $\mathbb{P}\{\varepsilon_k = \pm 1\} = \frac{1}{2}$, $k = 1, \dots, n$. Let all ε_k be also independent of all ξ_j. If $(\xi_1', \dots, \xi_n')$ is an independent copy of $(\xi_1, \dots, \xi_n)$, which is independent of the ε_k's, by the triangle inequality and symmetry,

$$\begin{aligned}\mathbb{E}\bigg(\Big|\sum_{k=1}^n \xi_k\Big|\bigg) &\geq \frac{1}{2}\,\mathbb{E}\bigg(\Big|\sum_{k=1}^n (\xi_k - \xi_k')\Big|\bigg) \\ &= \frac{1}{2}\,\mathbb{E}\bigg(\Big|\sum_{k=1}^n \varepsilon_k(\xi_k - \xi_k')\Big|\bigg) \geq \frac{1}{2}\,\mathbb{E}\bigg(\Big|\sum_{k=1}^n \varepsilon_k \xi_k\Big|\bigg)\end{aligned}$$

where the last step follows from Jensen's inequality in the centered ξ_k' variables. On the other hand, by Khinchine's inequality with optimal constant (due to Haagerup [Ha]), for all scalars $a_1, \dots, a_n$,

$$\mathbb{E}\bigg(\Big|\sum_{k=1}^n \varepsilon_k a_k\Big|\bigg) \geq \frac{1}{\sqrt{2}}\Big(\sum_{k=1}^n a_k^2\Big)^{1/2}.$$

Combining the two inequalities yields Lemma 3.4 with $c = \frac{1}{2\sqrt{2}}$.

On the basis of Lemma 3.4, we address the proof of Theorem 3.1.

Proof of Theorem 3.1. From Theorem 2.9, for every $n \geq 1$,

$$\mathbb{E}\big(W_1(\mu_n, \mu)\big) = \int_{-\infty}^{\infty} \mathbb{E}\big(|F_n(x) - F(x)|\big)dx.$$

By Lemma 3.4 applied to $\xi_k = \mathbb{1}_{\{X_k \leq x\}} - F(x)$, $k = 1, \dots, n$, we have

$$\mathbb{E}\big(|F_n(x) - F(x)|\big) \geq \frac{c}{n}\,\mathbb{E}\bigg(\Big(\sum_{k=1}^n \xi_k^2\Big)^{1/2}\bigg).$$

But
$$\mathbb{E}\bigg(\Big(\sum_{k=1}^n \xi_k^2\Big)^{1/2}\bigg) \, \geq \, \Big(\sum_{k=1}^n \big(\mathbb{E}(|\xi_k|)\big)^2\Big)^{1/2} \, = \, 2\sqrt{n}\, F(x)\big((1-F(x)\big),$$
so that
$$\mathbb{E}\big(\big|F_n(x) - F(x)\big|\big) \, \geq \, \frac{1}{\sqrt{2n}}\, F(x)\big((1-F(x)\big).$$
After integration over all $x \in \mathbb{R}$, we arrive at
$$\mathbb{E}\big(W_1(\mu_n,\mu)\big) \, \geq \, \frac{1}{\sqrt{2n}} \int_{-\infty}^{\infty} F(x)\big(1-F(x)\big)\,dx.$$

To further estimate the last integral, it remains to recall the general identity
$$\int_{-\infty}^{\infty} F(x)\big(1-F(x)\big)\,dx \, = \, \frac{1}{2}\,\mathbb{E}\big(|X_1 - X_1'|\big)$$
where X_1' is an independent copy of X_1. Note also that the expectation $\mathbb{E}(|X_1 - a|)$ is minimized for (any) median $a = m(X_1)$ of a given random variable X_1 with distribution function F. Theorem 3.1 follows. □

3.2 Two-sided bounds on $\mathbb{E}(W_1(\mu_n,\mu))$

Now, we turn to the more delicate question on how to bound explicitly, both from above and below, the mean distance $\mathbb{E}(W_1(\mu_n,\mu))$ in terms of the distribution function F of μ for each fixed $n \geq 1$. The next statement refines both Theorem 3.1 and Theorem 3.2, and may be used to obtain a variety of possible rates.

Theorem 3.5 (Two-sided bounds on $\mathbb{E}(W_1(\mu_n,\mu))$)**.** *There is an absolute constant $c > 0$ such that for any $\mu \in Z_1(\mathbb{R})$, for every $n \geq 1$,*
$$c\,(A_n + B_n) \, \leq \, \mathbb{E}\big(W_1(\mu_n,\mu)\big) \leq A_n + B_n$$
where
$$A_n \;=\; 2\int_{\{4nF(x)(1-F(x)) \leq 1\}} F(x)\big(1-F(x)\big)\,dx,$$
$$B_n \;=\; \frac{1}{\sqrt{n}} \int_{\{4nF(x)(1-F(x)) > 1\}} \sqrt{F(x)(1-F(x))}\,dx.$$
One may take $c = \frac{1}{2}\,5^{-4}$.

Since $\sqrt{n}\,B_n \to J_1(\mu)$ as $n \to \infty$, together with Theorem 3.2, the preceding result yields the following characterization.

Corollary 3.6 (Characterization of the standard rate for $\mathbb{E}(W_1(\mu_n,\mu))$). *Given $\mu \in Z_1(\mathbb{R})$, $\mathbb{E}(W_1(\mu_n,\mu)) = O(\frac{1}{\sqrt{n}})$ as $n \to \infty$ if and only if the integral*

$$J_1(\mu) = \int_{-\infty}^{\infty} \sqrt{F(x)(1-F(x))}\, dx$$

is finite.

Before turning to the proof of Theorem 3.5, let us look at the possible behavior of $\mathbb{E}(W_1(\mu_n,\mu))$ for some classes of underlying distributions μ with finite first absolute moment and such that the integral $J_1(\mu)$ is infinite. Consider for instance the modified Pareto distributions μ on the real line that are symmetric about the origin and have distribution functions F such that

$$4F(x)\big(1-F(x)\big) = x^{-\beta}, \qquad x \geq 1,$$

with parameter $1 < \beta < 2$. In this case,

$$A_n = 2\int_{\{4nF(x)(1-F(x))\leq 1\}} F(x)\big(1-F(x)\big)dx = c_\beta\, n^{-1/\beta^*}$$

where $\beta^* = \frac{\beta}{\beta-1}$ is the conjugate exponent. In addition,

$$B_n = \frac{1}{\sqrt{n}} \int_{\{4nF(x)(1-F(x))>1\}} \sqrt{F(x)(1-F(x))}\, dx = c_\beta'\, n^{-1/\beta^*} - c_\beta''\, n^{-1/2}.$$

Since $\beta^* > 2$, it follows that

$$A_n + B_n \sim c_\beta\, n^{-1/\beta^*}$$

with some constant c_β depending on β. Therefore, by Theorem 3.5, $\mathbb{E}(W_1(\mu_n,\mu))$ can be of order $n^{-\alpha}$ with any prescribed value of α such that $0 < \alpha < \frac{1}{2}$. A similar conclusion can also be made about standard Pareto distributions.

We next address the proof of Theorem 3.3 of the last section on the same basis. Indeed, given a rate $\varepsilon_n \to 0$, it is sufficient, by Theorem 3.5, to construct a distribution function F with finite first absolute moment such that

$$\frac{1}{2}\, A_n = \int_{\{4nF(x)(1-F(x))\leq 1\}} F(x)\big(1-F(x)\big)\, dx \geq \varepsilon_n$$

for all n large enough. Restricting ourselves to the case where F is continuous, (strictly) increasing and symmetric about the origin, one may formulate this task in terms of the function $u(x) = 4F(x)(1-F(x))$ considered on $[0,\infty)$. Indeed, u may be an arbitrary continuous, (strictly) decreasing, integrable function in $x \geq 0$, such that $u(0) = 1$, and we need to have the additional property that

$$\int_{\{u(x)\leq 1/n\}} u(x)\, dx \geq \varepsilon_n$$

for all n large enough. But this is a rather obvious statement justifying therefore Theorem 3.3.

The proof of Theorem 3.5 requires some preparation. First, we need to refine the bound of Lemma 3.4 in the case of Bernoulli random variables. The next statement is a preliminary step.

Lemma 3.7. *Given independent random variables $\xi_1, \dots, \xi_n$ such that $|\xi_k| \leq 1$ a.s., $k = 1, \dots, n$, we have*

$$\mathrm{Var}\big(\|\xi\|\big) \,\leq\, 1$$

where $\|\xi\| = \big(\sum_{k=1}^n \xi_k^2\big)^{1/2}$ denotes the Euclidean norm of the vector $\xi = (\xi_1, \dots, \xi_n)$.

Proof. For any random variable $R \geq 0$ with finite 4-th moment and such that $\mathbb{E}(R^2) > 0$, there is a general upper-bound for its variance,

$$\mathrm{Var}(R) \,\leq\, \frac{\mathrm{Var}(R^2)}{\mathbb{E}(R^2)}\,. \tag{3.5}$$

Indeed, putting $a = \sqrt{\mathbb{E}(R^2)}$,

$$\mathrm{Var}(R^2) \,=\, \mathbb{E}\big((R^2 - a^2)^2\big) \,=\, \mathbb{E}\big((R-a)^2 (R+a)^2\big) \,\geq\, a^2\, \mathbb{E}\big((R-a)^2\big) \,\geq\, a^2\, \mathrm{Var}(R)$$

which is exactly the desired bound.

Now, take $R = \|\xi\|$. Then, by the independence of the ξ_k's, and using that $|\xi_k| \leq 1$,

$$\mathrm{Var}(R^2) \,=\, \sum_{k=1}^n \mathrm{Var}(\xi_k^2) \,\leq\, \sum_{k=1}^n \mathbb{E}(\xi_k^4) \,\leq\, \sum_{k=1}^n \mathbb{E}(\xi_k^2) \,=\, \mathbb{E}(R^2).$$

Hence $\mathrm{Var}(R) \leq 1$ by (3.5), and the lemma follows. □

Observe that the bound of Lemma 3.7 represents a special case of the concentration of measure phenomenon for product measures on the cube $[-1,1]^n$, stated for the specific convex Lipschitz function $u(x) = \|x\|$. See [T2], [L1-3], [Bob1].

The next crucial lemma for sums of independent Bernoulli random variables is perhaps classical, but we could not find an appropriate specific reference.

Lemma 3.8. *Let $S_n = \eta_1 + \dots + \eta_n$ be the sum of n independent Bernoulli random variables η_k, $k = 1, \dots, n$, with $\mathbb{P}\{\eta_k = 1\} = p$ and $\mathbb{P}\{\eta_k = 0\} = q = 1 - p$ where $p \in (0,1)$. Then*

$$c\, \min\{2npq, \sqrt{npq}\,\} \,\leq\, \mathbb{E}\big(|S_n - np|\big) \,\leq\, \min\{2npq, \sqrt{npq}\,\},$$

where $c > 0$ is an absolute constant. One may take $c = \frac{1}{2}\, 5^{-4}$.

Proof. The upper-bound is elementary. On the one hand,

$$\mathbb{E}\big(|S_n - np|\big) \ \le \ \sqrt{\mathrm{Var}(S_n)} \ = \ \sqrt{npq}.$$

On the other hand, by the triangle inequality,

$$\mathbb{E}\big(|S_n - np|\big) \ \le \ \sum_{k=1}^n \mathbb{E}\big(|\eta_k - p|\big) \ = \ 2npq.$$

The two estimates imply the upper-bound of the lemma.

To derive a lower-bound, we apply Lemma 3.7 to the random vector $\xi = (\xi_1, \dots, \xi_n)$ in $\mathbb{R}^n$ where $\xi_k = \eta_k - p$, for $k = 1, \dots, n$. It gives

$$\big(\mathbb{E}\big(\|\xi\|\big)\big)^2 \ \ge \ \mathbb{E}\big(\|\xi\|^2\big) - 1 \ = \ npq - 1 \ \ge \ \frac{1}{2}\, npq,$$

where the last inequality holds whenever $npq \ge 2$. In this case, we get therefore $\mathbb{E}(\|\xi\|) \ge \frac{1}{\sqrt{2}}\sqrt{npq}$, and by Lemma 3.4,

$$\mathbb{E}\big(|S_n - np|\big) \ = \ \mathbb{E}\bigg(\bigg|\sum_{k=1}^n \xi_k\bigg|\bigg) \ \ge \ \frac{1}{4}\sqrt{npq}.$$

It yields the lower bound of the lemma (with the better constant $c = \frac{1}{4}$).

Now, assume that $npq \le 2$ and, without loss of generality, that $p \ge \frac{1}{2}$. Hence, $q \le \frac{4}{n}$, and moreover,

$$q \ \le \ q_n \ = \ \min\Big\{\frac{1}{2}, \frac{4}{n}\Big\}.$$

We have

$$\begin{aligned} \mathbb{E}\big(|S_n - np|\big) \ &\ge \ \mathbb{E}\big(|S_n - np|\, 1\!\!1_{\{\eta_1 = \dots = \eta_n = 1\}}\big) \\ &= \ nq\, \mathbb{P}\{\eta_1 = \dots = \eta_n = 1\} \\ &= \ p^{n-1}\, npq. \end{aligned} \tag{3.6}$$

If $n \le 4$, just use $p^{n-1} \ge \frac{1}{8}$. For the remaining values $n \ge 5$, we have

$$p^{n-1} \ge (1 - q_n)^{n-1} \ge \Big(1 - \frac{4}{n}\Big)^{n-1} \ge 5^{-4}. \tag{3.7}$$

(Here, the last inequality may be explained by the fact that the function $(1 - \frac{4}{x})^{x-1}$ is increasing in $x \ge 5$.) Combining the two lower estimates (3.6) and (3.7), we finally get

$$\mathbb{E}\big(|S_n - np|\big) \ \ge \ c \cdot 2npq \ \ge \ c\, \min\{2npq, \sqrt{npq}\}$$

with constant $c = \frac{1}{2}\, 5^{-4}$. $\square$

Proof of Theorem 3.5. Start again from Theorem 2.9 to get

$$\mathbb{E}\big(W_1(\mu_n,\mu)\big) \ = \ \int_{-\infty}^{\infty} \mathbb{E}\big(\big|F_n(x) - F(x)\big|\big)\, dx.$$

By Lemma 3.8 applied to $\eta_k = 1\!\!1_{\{X_k \leq x\}}$, $k = 1,\dots,n$, and $p = F(x)$, $q = 1 - F(x)$, we obtain

$$\mathbb{E}\big(\big|F_n(x) - F(x)\big|\big) \ \leq \ \frac{1}{n}\, \min\Big\{2nF(x)\big(1-F(x)\big),\, \sqrt{nF(x)(1-F(x))}\,\Big\}.$$

Hence,

$$\begin{aligned}\mathbb{E}\big(W_1(\mu_n,\mu)\big) \ &\leq \ 2\int_{\{4nF(x)(1-F(x))\leq 1\}} F(x)\big(1-F(x)\big)\, dx \\ &\quad + \frac{1}{\sqrt{n}}\int_{\{4nF(x)(1-F(x))>1\}} \sqrt{F(x)(1-F(x))}\, dx \ = \ A_n + B_n.\end{aligned}$$

The lower-bound of the same lemma yields the reverse bound $\mathbb{E}(W_1(\mu_n,\mu)) \geq c(A_n + B_n)$ with constant $c = \frac{1}{2}\,5^{-4}$. The proof is complete. $\square$

3.3 Functional limit theorems

The condition $J_1(\mu) < \infty$ as in Corollary 3.6 appears naturally in functional central limit theorems. Namely, suppose that Y is a random element in the Banach space $L^1(\mathbb{R})$, and let $(Y_k)_{k\geq 1}$ be independent copies of Y. By the assumption, the integral

$$\|Y\|_{L^1(\mathbb{R})} \ = \ \int_{-\infty}^{\infty} |Y(x)|\, dx$$

is almost surely finite. As a stronger condition, one may assume the finiteness of the first L^1-norm moment

$$\mathbb{E}\big(\|Y\|_{L^1(\mathbb{R})}\big) \ = \ \int_{-\infty}^{\infty} \mathbb{E}\big(|Y(x)|\big)\, dx.$$

Define $Z_n = \frac{1}{\sqrt{n}}\,(Y_1 + \cdots + Y_n)$, $n \geq 1$. Assuming in addition that $\mathbb{E}(Y(x)) = 0$, for all $x \in \mathbb{R}$, by a theorem due to Giné and Zinn [G-Z], the sequence $(Z_n)_{n\geq 1}$ is convergent weakly in distribution in $L^1(\mathbb{R})$ to a Gaussian limit γ if and only if

$$\int_{-\infty}^{\infty} \big[\,\mathbb{E}\big(|Y(x)|^2\big)\big]^{1/2} dx \ < \ \infty. \tag{3.8}$$

In particular, this theorem can be applied to the random function

$$Y(x) \ = \ 1\!\!1_X(x) - F(x), \qquad x \in \mathbb{R},$$

which belong to $L^1(\mathbb{R})$ as long as $\mathbb{E}(|X|) < \infty$. For this choice of Y, the condition (3.8) exactly amounts to the finiteness of $J_1(\mu)$. Therefore

$$Z_n = \sqrt{n}\,\big(F_n(x) - F(x)\big) \to \gamma \qquad \text{weakly in } L^1(\mathbb{R})$$

if and only if $J_1(\mu) < \infty$.

To describe the Gaussian limit γ, recall the classical fact (known as Donsker's theorem) that when F is a uniform distribution on the unit interval $[0,1]$,

$$\sqrt{n}\,\big(F_n(t) - t\big) \to W^o(t).$$

weakly in the Skorokhod space $D([0,1])$. That is, γ is the distribution of the standard Brownian bridge $W^o(t) = W(t) - tW(1)$, $0 \leq t \leq 1$. Recall that $D([0,1])$ is equipped with a metric generating the topology such that all balls with respect to the uniform metric are Borel measurable, cf. [Bi], [S-W], [C-H].

Using the change of the variable $t = F(x)$, this result extends to

$$\sqrt{n}\,\big(F_n(x) - F(x)\big) \to W^o\big(F(x)\big)$$

where the weak convergence is understood in the space $D_0(-\infty,\infty)$ of all functions $u = u(x)$ such that the limits $u(x-)$ and $u(x+)$ exist and are finite for all points x of the real line, and in addition $u(x) \to 0$ as $|x| \to \infty$.

Note that no moment assumption is needed for such weak convergence. But the Gaussian process $G(x) = W^o(F(x))$, $x \in \mathbb{R}$, has trajectories in $L^1(\mathbb{R})$, that is,

$$\|G\|_{L^1(\mathbb{R})} = \int_{-\infty}^{\infty} \big|W^o(F(x))\big|\,dx < \infty \qquad \text{a.s.}$$

if and only if

$$\mathbb{E}\big(\|G\|_{L^1(\mathbb{R})}\big) = \int_{-\infty}^{\infty} \mathbb{E}\big(\big|W^o(F(x))\big|\big)\,dx = \sqrt{\frac{2}{\pi}}\,J_1(\mu) < \infty.$$

The following statement summarizes the previous conclusions and complements Corollary 3.6.

Corollary 3.9 (Functional limit theorem for $W_1(\mu_n,\mu)$)**.** *For any $\mu \in Z_1(\mathbb{R})$ with distribution function F, the following properties are equivalent:*

a) $\sqrt{n}\,\mathbb{E}(W_1(\mu_n,\mu)) = O(1)$ *as* $n \to \infty$*;*

b) $\sqrt{n}\,(F_n(x) - F(x)) \to W^o(F(x))$ *weakly in* $L^1(\mathbb{R})$*;*

c) $J_1(\mu) < \infty$.

Under one of these equivalent conditions,

$$\sqrt{n}\,W_1(\mu_n,\mu) \to \int_{-\infty}^{\infty} \big|W^o(F(x))\big|\,dx$$

weakly in distribution on $\mathbb{R}$.

The equivalence of b) and c) was first emphasized in the work of del Barrio, Giné and Matrán [B-G-M], cf. Theorem 2.1 therein. It is also stated there that, under the condition c), the sequence $\sqrt{n}\, W_1(\mu_n, \mu)$ is stochastically bounded, i.e.

$$\lim_{t \to \infty} \sup_{n \geq 1} \mathbb{P}\big\{\sqrt{n}\, W_1(\mu_n, \mu) > t\big\} \, = \, 0.$$

Moreover, it is shown in [B-G-M] (Theorem 2.4) that, if $J_1(\mu) < \infty$, and if additionally μ has a finite absolute moment of order $p \geq 2$, then there is a finite limit of moments

$$\lim_{n \to \infty} \mathbb{E}\Big(\big(\sqrt{n}\, W_1(\mu_n, \mu)\big)^r\Big) \, = \, \mathbb{E}\left(\Big(\int_{-\infty}^{\infty} \big|W^o(F(x))\big|\, dx\Big)^r\right)$$

for any $0 < r \leq p$.

In case $J_1(\mu) = \infty$ with μ being a stable law, an asymptotic behavior of distributions of $W_1(\mu_n, \mu)$ was also studied in [B-G-M]. In particular, Proposition 4.3 and Corollary 4.4 therein indicate that, if μ is in the domain of normal attraction with a normalizing sequence b_n and with $J_1(\mu) = \infty$, then

$$\lim_{n \to \infty} \frac{W_1(\mu_n, \mu)}{b_n} \, = \, \lim_{n \to \infty} \frac{W_1(\mu_n, \mu)}{\mathbb{E}(W_1(\mu_n, \mu))} \, = \, 1 \qquad \text{in probability.}$$

This may be viewed as a variant of the law of large numbers.

4 Order statistics representations of $W_p(\mu_n, \mu)$

With this section, we start the investigation of the rate as $n \to \infty$ of $\mathbb{E}\big(W_p(\mu_n, \mu)\big)$ for some $p > 1$. As will be clear in the next sections, it actually turns out that the behaviour of $\mathbb{E}\big(W_p(\mu_n, \mu)\big)$ for $p > 1$ can be very different than in the case $p = 1$ (for which the preceding section provided the universal rate $\frac{1}{\sqrt{n}}$ under mild conditions on μ). The investigation of the case $p > 1$ thus requires different tools and methods. This section introduces new tools in this investigation, namely the use of order statistics and related beta distributions after reduction to the uniform distribution, which are more suited to the case $p > 1$. Exact rates for the uniform distribution are provided.

As in the previous section, we deal with a Borel probability measure μ on $\mathbb{R}$ with distribution function F, law of a random variable X, and with the sequence $\mu_n = \frac{1}{n}\sum_{k=1}^n \delta_{X_k}$, $n \geq 1$, of empirical measures of a sample $(X_k)_{k\geq 1}$ of independent copies of X.

4.1 Optimal transport, order statistics and inverse functions

We first develop a description of the Kantorovich distances in terms of order statistics. First, let us recall a few basic facts from the field of transportation of measure. We refer to [R-R], [Vi1], [Vi2] for more complete accounts.

Any collection of real numbers $x_1, \dots, x_n$ may be arranged in increasing order $x_1^* \leq \cdots \leq x_n^*$. In particular,

$$x_1^* = \min_{1\leq k\leq n} x_k \qquad \text{and} \qquad x_n^* = \max_{1\leq k\leq n} x_k.$$

A similar notation is applied to random variables $X_1, \dots, X_n$, in which case X_k^* is called the k-th order statistic.

The following lemma is classical.

Lemma 4.1. *Let $V : \mathbb{R} \to \mathbb{R}$ be even and convex. For any two collections of real numbers $x_1, \dots, x_n$ and $y_1, \dots, y_n$,*

$$\inf_\sigma \sum_{k=1}^n V(x_k - y_{\sigma(k)}) = \sum_{k=1}^n V(x_k^* - y_k^*)$$

where the infimum is taken over all permutations σ of $\{1, \dots, n\}$.

With every collection of real numbers $x_1, \dots, x_n$, we associate an "empirical" measure

$$\mu = \frac{1}{n}\sum_{k=1}^n \delta_{x_k}.$$

The next lemma specializes the representation of W_p in terms of the inverse distribution functions (Theorem 2.10) to the class of such measures. For completeness, we include a proof on the basis of Lemma 4.1.

Lemma 4.2. *Given two collections of real numbers* $x_1, \dots, x_n$ *and* $y_1, \dots, y_n$, *let* μ *and* ν *be the corresponding empirical measures. Then, for any* $p \geq 1$,

$$W_p^p(\mu, \nu) \;=\; \frac{1}{n} \sum_{k=1}^n |x_k^* - y_k^*|^p.$$

In particular,

$$W_\infty(\mu, \nu) \;=\; \max_{1 \leq k \leq n} |x_k^* - y_k^*|.$$

Proof. By the very definition of the Kantorovich distance $W_p(\mu, \nu)$ between μ and ν,

$$W_p^p(\mu, \nu) \;=\; \inf_\pi \int_{\mathbb{R}} \int_{\mathbb{R}} |x - y|^p \, d\pi(x, y) \;=\; \inf_\pi \sum_{i=1}^n \sum_{j=1}^n |x_i - y_j|^p \, \pi_{ij}$$

where the infimum is taken over all probability measures π on the plane $\mathbb{R} \times \mathbb{R}$ with marginals μ and ν, and where we put $\pi_{ij} = \pi\{(x_i, y_j)\}$ (necessarily, π is supported on the points (x_i, y_j), $1 \leq i, j \leq n$). Thus, the second infimum is taken over the set M_n of all $n \times n$ matrices (π_{ij}) with non-negative entries such that, for any $i = 1, \dots, n$ and any $j = 1, \dots, n$,

$$\sum_{j=1}^n \pi_{ij} \;=\; \sum_{i=1}^n \pi_{ij} \;=\; \frac{1}{n}\,.$$

Note that M_n represents a convex compact subset of $\mathbb{R}^{n^2}$, and the functional

$$\pi \;\to\; \sum_{i=1}^n \sum_{j=1}^n |x_i - y_j|^p \, \pi_{ij}$$

is affine on it. Therefore, this functional attains minimum at one of the extreme points of M_n. But, by the well-known Birkhoff theorem, any such point has the form $\pi_{ij} = \frac{1}{n}\, 1_{\{j = \sigma(i)\}}$, where $\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$ is an arbitrary permutation. Hence,

$$W_p^p(\mu, \nu) \;=\; \frac{1}{n} \inf_\sigma \sum_{i=1}^n |x_i - y_{\sigma(i)}|^p$$

where the infimum is taken over all permutations $\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$. Finally, to specify the last infimum, it remains to apply Lemma 4.1 with the convex function $V(x) = |x|^p$. $\square$

Let us now apply Lemma 4.2 to arbitrary collections of random variables. We denote by $\mathcal{L}(\xi)$ the distribution (the law) of a (real) random variable ξ.

Theorem 4.3 (Order statistics representation of $W_p(\mu_n, \mu)$)**.** *Given random vectors $(X_1, \dots, X_n)$ and $(Y_1, \dots, Y_n)$ in $\mathbb{R}^n$, let μ_n and ν_n be the corresponding empirical measures. Then, for any $p \geq 1$,*

$$\mathbb{E}\big(W_p^p(\mu_n, \nu_n)\big) \, = \, \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\big(|X_k^* - Y_k^*|^p\big). \tag{4.1}$$

Moreover, if $(Y_1, \dots, Y_n)$ is an independent copy of $(X_1, \dots, X_n)$ and $\mu = \mathbb{E}(\mu_n) = \frac{1}{n} \sum_{k=1}^{n} \mathcal{L}(X_k)$ is the mean marginal distribution,

$$\frac{2^{-p}}{n} \sum_{k=1}^{n} \mathbb{E}\big(\big|X_k^* - \mathbb{E}(X_k^*)\big|^p\big) \, \leq \, \mathbb{E}\big(W_p^p(\mu_n, \mu)\big) \, \leq \, \frac{2^p}{n} \sum_{k=1}^{n} \mathbb{E}\big(\big|X_k^* - \mathbb{E}(X_k^*)\big|^p\big). \tag{4.2}$$

In the case $p = 2$ similar bounds hold with better constants:

$$\frac{1}{2n} \sum_{k=1}^{n} \mathrm{Var}(X_k^*) \, \leq \, \mathbb{E}\big(W_2^2(\mu_n, \mu)\big) \leq \frac{2}{n} \sum_{k=1}^{n} \mathrm{Var}(X_k^*).$$

Proof. The first relation (4.1) immediately follows from Lemma 4.2. To derive the lower-bound for $\mathbb{E}(W_p^p(\mu_n, \mu))$ in (4.2), one may use the triangle inequality for the distance W_p to get that

$$W_p^p(\mu_n, \nu_n) \, \leq \, 2^{p-1}\big(W_p^p(\mu_n, \mu) + W_p^p(\nu_n, \mu)\big).$$

After taking the expectations,

$$\mathbb{E}\big(W_p^p(\mu_n, \nu_n)\big) \, \leq \, 2^{p-1}\Big(\mathbb{E}\big(W_p^p(\mu_n, \mu)\big) + \mathbb{E}\big(W_p^p(\nu_n, \mu)\big)\Big) \, = \, 2^p \, \mathbb{E}\big(W_p^p(\mu_n, \mu)\big)$$

so that

$$\mathbb{E}\big(W_p^p(\mu_n, \mu)\big) \, \geq \, \frac{2^{-p}}{n} \sum_{k=1}^{n} \mathbb{E}\, |X_k^* - Y_k^*|^p.$$

Also, by Jensen's inequality and independence,

$$\mathbb{E}\big(|X_k^* - Y_k^*|\big)^p \, \geq \, \mathbb{E}\big(\big|X_k^* - \mathbb{E}(Y_k^*)\big|^p\big) \, = \, \mathbb{E}\big(\big|X_k^* - \mathbb{E}(Y_k^*)\big|^p\big)$$

so the lower-bound in (4.2) follows. Note that $\mathbb{E}(|X_k^* - Y_k^*|^2) = 2\,\mathrm{Var}(X_k^*)$, which leads to the improvement in the constant when $p = 2$.

To derive the upper-bound in (4.2), we use the convexity of the functional $\nu \mapsto W_p^p(\mu, \nu)$. This property may actually be verified in the setting of an arbitrary metric space (E, d). Fix $t_1, t_2 \geq 0$, $t_1 + t_2 = 1$. Given $\mu, \nu_1, \nu_2 \in Z_p(E, d)$, let probability measures π_1 and

π_2 on $E \times E$ have marginals (μ, ν_1) and (μ, ν_2), respectively. Then $\pi = t_1\pi_1 + t_2\pi_2$ has marginals (μ, ν) where $\nu = t_1\nu_1 + t_2\nu_2$. Hence,

$$\begin{aligned} W_p^p(\mu,\nu) &\le \int_E\int_E |x-y|^p \, d\pi(x,y) \\ &= t_1 \int_E\int_E |x-y|^p \, d\pi_1(x,y) + t_2 \int_E\int_E |x-y|^p \, d\pi_2(x,y). \end{aligned}$$

Taking the infimum on the right-hand side over all admissible measures π_1 and π_2, we arrive at

$$W_p^p(\mu,\nu) \le t_1 W_p^p(\mu,\nu_1) + t_2 W_p^p(\mu,\nu_2)$$

which means exactly the convexity.

Since the functional $\nu \to W_p^p(\mu,\nu)$ is also continuous on $Z_p(E,d)$, the above Jensen inequality extends to infinite sums or integrals, at least, when (E,d) is separable. In particular, in the space $Z_p(\mathbb{R})$

$$W_p^p(\mu_n,\mu) = W_p^p\big(\mu_n, \mathbb{E}_Y(\nu_n)\big) \le \mathbb{E}_Y\big(W_p^p(\mu_n,\nu_n)\big).$$

After the next integration with respect to X this yields

$$\mathbb{E}\big(W_p^p(\mu_n,\mu)\big) \le \frac{1}{n}\sum_{k=1}^n \mathbb{E}\big(|X_k^* - Y_k^*|^p\big)$$

from which (4.2) immediately follows. If $p = 2$, the latter expectation is equal to $2\,\mathrm{Var}(X_k^*)$. Theorem 4.3 is proved. □

To conclude this section, we briefly investigate another possible approach to the study of the Kantorovich distances between probability distributions on the real line by means of inverse (distribution) functions. With every distribution function F, recall the associated inverse distribution function

$$F^{-1}(t) = \inf\big\{x \in \mathbb{R} : F(x) \ge t\big\}, \qquad 0 < t < 1$$

from (2.2). (We refer to Appendix A for a complete analytic investigation of inverse distribution functions.) The alternative description of Kantorovich distances is based on the general explicit representation

$$W_p^p(\mu,\nu) = \int_0^1 \big|F^{-1}(t) - G^{-1}(t)\big|^p \, dt,$$

from Theorem 2.10. Applying it to the empirical measures, one obtains the following alternative variant of Theorem 4.3 for $\mathbb{E}(W_p^p(\mu_n,\mu))$ in terms of the quantile process $t \to F_n^{-1}(t)$.

Theorem 4.4 (Quantile representation of $W_p^p(\mu_n,\mu)$)**.** *Let $(X_1,\dots,X_n)$ be a vector of random variables with finite p-th absolute moments ($p \geq 1$). Let μ_n be the corresponding empirical measure, and $\mu = \frac{1}{n}\sum_{k=1}^n \mathcal{L}(X_k)$ be the mean marginal distribution. Then*

$$\mathbb{E}\big(W_p^p(\mu_n,\mu)\big) \;=\; \int_0^1 \mathbb{E}\big(\big|F_n^{-1}(t) - F^{-1}(t)\big|^p\big)\, dt$$

where F_n and F are distribution functions associated with μ_n and μ respectively.

The representation of Theorem 4.4 leads to a different expression for $\mathbb{E}(W_p^p(\mu_n,\mu))$ in comparison with the bounds of Theorem 4.3. For definiteness, assume that $X_1^* < \dots < X_n^*$ a.s., so that μ_n assigns the mass $\frac{1}{n}$ to n distinct points (this assumption can be removed in the resulting representations). Then $F_n(X_k^*) = \frac{k}{n}$ and

$$F_n^{-1}(t) \;=\; X_k^* \quad \text{for} \quad \frac{k-1}{n} < t \leq \frac{k}{n}, \quad k = 1,\dots,n.$$

Hence, by Theorem 4.4,

$$\mathbb{E}\, W_p^p(\mu_n,\mu) \;=\; \sum_{k=1}^n \int_{(k-1)/n}^{k/n} \mathbb{E}\big(\big|X_k^* - F^{-1}(t)\big|^p\big) dt.$$

Under further regularity assumptions on F, one may change the variable $t = F(x)$, and then the above formula becomes

$$\mathbb{E}\big(W_p^p(\mu_n,\mu)\big) \;=\; \sum_{k=1}^n \int_{F^{-1}((k-1)/n)}^{F^{-1}(k/n)} \mathbb{E}\big(|X_k^* - x|^p\big) dF(x).$$

Both formulas can be used for computations in special cases. For example, for the basic exponent $p = 2$,

$$\mathbb{E}\big(\big|X_k^* - F^{-1}(t)\big|^2\big) \;=\; \mathrm{Var}(X_k^*) + \big|\mathbb{E}(X_k^*) - F^{-1}(t)\big|^2.$$

The following statement results.

Corollary 4.5. *In the setting of Theorem 4.4 with $p = 2$,*

$$\mathbb{E}\big(W_2^2(\mu_n,\mu)\big) \;=\; \frac{1}{n}\sum_{k=1}^n \mathrm{Var}(X_k^*) + \sum_{k=1}^n \int_{(k-1)/n}^{k/n} \big|\mathbb{E}(X_k^*) - F^{-1}(t)\big|^2\, dt.$$

4.2 Reduction to the uniform distribution

In the study of rates for the mean Kantorovich distances in the scheme of independent identically distributed random variables, the uniform distribution appears as the best possible example (in some sense). On the other hand, many questions about $\mathbb{E}(W_p^p(\mu_n, \mu))$ in the general case can be reduced to this example.

When a sample $(U_1, \dots, U_n)$ consists of independent random variables uniformly distributed in the unit interval $(0, 1)$, the order statistics $U_1^* \leq \cdots \leq U_n^*$ are well-studied in this case. Indeed, the k-th order statistic U_k^*, $k = 1, \dots, n$, has a beta distribution $B_{k,n-k+1}$ with parameters $(k, n - k + 1)$, i.e. it has the density

$$p_{k,n}(x) \,=\, n\, C_{n-1}^{k-1}\, x^{k-1}(1-x)^{n-k}, \qquad 0 < x < 1,$$

where $C_{n-1}^{k-1} = \frac{(n-1)!}{(k-1)!(n-k)!}$ are the usual binomial coefficients. We refer to Appendix B for a complete analysis of beta distributions and their densities.

Therefore, when the random variables $X_1, \dots, X_n$ are independent and distributed according to μ with distribution function F and associated inverse function F^{-1} (cf. (2.2)), the corresponding order statistics may be represented as

$$X_k^* \,=\, F^{-1}(U_k^*), \qquad k = 1, \dots, n$$

(cf. Proposition A.1). As a result, there is yet another alternative to the representation of $\mathbb{E}\big(W_p^p(\mu_n, \nu_n)\big)$ given in Theorem 4.3. As usual, $\mu_n = \frac{1}{n}\sum_{k=1}^n \delta_{X_k}$.

Theorem 4.6 (Beta representation of $W_p(\mu_n, \mu)$)**.** *If ν_n is an independent copy of μ_n, for all $p \geq 1$,*

$$\mathbb{E}\big(W_p^p(\mu_n, \nu_n)\big) \,=\, \frac{1}{n}\sum_{k=1}^n \int_0^1\!\!\int_0^1 \big|F^{-1}(t) - F^{-1}(s)\big|^p \, dB_{k,n-k+1}(t)\, dB_{k,n-k+1}(s). \tag{4.3}$$

Recall that

$$\mathbb{E}\big(W_p^p(\mu_n, \mu)\big) \,\leq\, \mathbb{E}\big(W_p^p(\mu_n, \nu_n)\big) \,\leq\, 2^p\, \mathbb{E}\big(W_p^p(\mu_n, \mu)\big)$$

so that the right-hand side of (4.3) with F the distribution function of μ describes in essence the behaviour of $\mathbb{E}(W_p^p(\mu_n, \mu))$. This approach will be used in Section 5 below.

At this point, let us illustrate how Theorem 4.3 may be used to derive an exact formula for the mean of the quadratic Kantorovich distances in the case of the uniform distribution.

We first recall from Appendix B the moment formulas for the beta distributions. For the sample $(U_1, \dots, U_n)$ drawn from the uniform distribution and the associated order statistics $U_1^* \leq \cdots \leq U_n^*$, we have $\alpha = k$ and $\beta = n - k + 1$ so that

$$\mathbb{E}(U_k^*) \,=\, \frac{k}{n+1} \qquad \text{and} \qquad \mathrm{Var}(U_k^*) \,=\, \frac{k(n-k+1)}{(n+1)^2\,(n+2)}\,.$$

It then easily follows that

$$\frac{1}{n}\sum_{k=1}^{n}\mathrm{Var}(U_k^*) \;=\; \frac{1}{6(n+1)}\,.$$

Using Theorem 4.3, we obtain an exact formula for $\mathbb{E}(W_2^2(\mu_n,\nu_n))$, where ν_n is an independent copy of μ_n. It also provides the two-sided bound

$$\frac{1}{12\,(n+1)} \le \mathbb{E}\big(W_2^2(\mu_n,\mu)\big) \le \frac{1}{3(n+1)}\,.$$

For an exact formula, one may appeal to Corollary 4.5 which gives

$$\begin{aligned}\mathbb{E}\big(W_2^2(\mu_n,\mu)\big) \;&=\; \frac{1}{n}\sum_{k=1}^{n}\mathrm{Var}(U_k^*) + \sum_{k=1}^{n}\int_{(k-1)/n}^{k/n}\big(\mathbb{E}(U_k^*)-t\big)^2\,dt \\ &=\; \frac{1}{6(n+1)} + \sum_{k=1}^{n}\int_{(k-1)/n}^{k/n}\Big(t-\frac{k}{n+1}\Big)^2\,dt \\ &=\; \frac{1}{6(n+1)} + \frac{1}{3\,(n(n+1))^3}\sum_{k=1}^{n}\big[k^3+(n-k+1)^3\big].\end{aligned}$$

Using that $\sum_{k=1}^{n}k^3 = \big(\frac{n(n+1)}{2}\big)^2$, we arrive at the following conclusion.

Theorem 4.7 (Exact rate of $\mathbb{E}(W_2^2(\mu_n,\mu))$ for the uniform distribution)**.** *If μ is the uniform distribution on $(0,1)$, for each $n\ge 1$,*

$$\mathbb{E}\big(W_2^2(\mu_n,\mu)\big) \;=\; \frac{1}{6n}\,.$$

In addition, if ν_n is an independent copy of μ_n,

$$\mathbb{E}\big(W_2^2(\mu_n,\nu_n)\big) \;=\; \frac{1}{3(n+1)}\,.$$

Similar asymptotic behaviours are valid for $\mathbb{E}(W_p^p(\mu_n,\mu))$ with arbitrary $p\ge 1$. Theorem 3.1 produces the lower-bound

$$\mathbb{E}\big(W_p(\mu_n,\mu)\big) \;\ge\; \mathbb{E}\big(W_1(\mu_n,\mu)\big) \;\ge\; \frac{c}{\sqrt{n}}\,\mathbb{E}\big(|X-m|\big) \tag{4.4}$$

where m is a median of X (with law μ) and $c=\frac{1}{2\sqrt{2}}$. To improve the factor $\mathbb{E}(|X-m|)$, recall that on the final step of derivation, we obtained a slightly better general bound

$$\mathbb{E}\big(W_1(\mu_n,\mu)\big) \;\ge\; \frac{1}{\sqrt{2n}}\int_{-\infty}^{\infty}F(x)\big(1-F(x)\big)\,dx,$$

which in case of the uniform distribution μ on $(0,1)$ becomes $\mathbb{E}(W_1(\mu_n,\mu)) \geq \frac{1}{6\sqrt{2n}}$.

For an upper-bound, one option is to apply a Khinchine-type inequality holding for random variables with log-concave probability distributions (and using the fact that the beta distributions are log-concave for $\alpha \geq 1$ and $\beta \geq 1$). To get a sharper estimate, one can also use the property that the U_k^*'s are uniformly sub-Gaussian. Indeed, as will be shown in Appendix B (cf. Proposition B.10), for all $k = 1,\dots,n$,

$$\mathbb{P}\big\{\big|U_k^* - \mathbb{E}(U_k^*)\big| \geq r\big\} \,\leq\, 2\,e^{-(n+1)r^2/8}, \qquad r \geq 0.$$

This readily provides bounds on the moments, for example

$$\mathbb{E}\big(\big|U_k^* - \mathbb{E}(U_k^*)\big|^p\big) \,\leq\, \Big(\frac{Cp}{n}\Big)^{p/2}$$

with some absolute constant $C > 0$. The following statement is then immediately obtained from Theorem 4.3.

Theorem 4.8 (Exact rate of $\mathbb{E}(W_p^p(\mu_n,\mu))$ for the uniform distribution)**.** *Let μ be the uniform distribution on $(0,1)$. Then, for any $p \geq 1$, and any $n \geq 1$, with some absolute constant $C > 0$,*

$$\big[\mathbb{E}\big(W_p^p(\mu_n,\mu)\big)\big]^{1/p} \,\leq\, C\,\sqrt{\frac{p}{n}}\,.$$

Note that from (4.4) and Theorem 4.8, $\mathbb{E}(W_p(\mu_n,\mu))$ is of order $\frac{1}{\sqrt{n}}$ for any $p \geq 1$.

However, the behaviour of $\mathbb{E}(W_p(\mu_n,\mu))$ and $\big[\mathbb{E}(W_p^p(\mu_n,\mu))\big]^{1/p}$ is different for growing p. Indeed, in the case of the uniform distribution μ on $(0,1)$, $F(x) = x$ for all $x \in [0,1]$. Let as usual F_n denote the distribution function associated to μ, which is thus constant outside $[0,1]$ (with probability one). Then, by Theorem 2.8 applied to μ, the distance $W_\infty(\mu_n,\mu)$ may be described as the infimum over all $h \geq 0$ such that $x - h \leq F_n(x) \leq x + h$, for all $x \in [0,1]$. That is, with probability one we have

$$W_\infty(\mu_n,\mu) \,=\, \|F_n - F\| \,=\, \sup_{x\in\mathbb{R}} \big|F_n(x) - F(x)\big|,$$

which represents the (uniform) Kolmogorov distance between F_n and F. In that case, by the Dvoretzky-Kiefer-Wolfowitz theorem [D-K-W], the random variables $\sqrt{n}\,\|F_n - F\|$ are uniformly sub-Gaussian, and more precisely,

$$\mathbb{P}\big\{\sqrt{n}\,\|F_n - F\| \geq t\big\} \,\leq\, 2\,e^{-2r^2}, \qquad r > 0,$$

cf. [Mas]. As a direct consequence, we obtain:

Theorem 4.9 (Exact rate of $\mathbb{E}(W_\infty(\mu_n,\mu))$ for the uniform distribution)**.** *Let μ be the uniform distribution on $(0,1)$. Then, with some absolute constant $C > 0$ (for example $C = \sqrt{\frac{\pi}{2}}$), for any $n \geq 1$,*

$$\mathbb{E}\big(W_\infty(\mu_n,\mu)\big) \,\leq\, \frac{C}{\sqrt{n}}\,.$$

On the other hand, the bound of Theorem 4.8 cannot be true with a p-independent constant, and the stated dependence in p is correct. Indeed,

$$\mathbb{E}\big(W_p^p(\mu_n,\mu)\big) \,=\, \mathbb{E}\bigg(\int_0^1 \big|F_n^{-1}(t)-t\big|^p\,dt\bigg) \,=\, \mathbb{E}\bigg(\int_0^1 \big|F_n(x)-x\big|^p\,dx\bigg).$$

By the central limit theorem, the random variables $\sqrt{n}\,(F_n(x)-F(x))$ are weakly convergent to the normal law $N(0,x(1-x))$. By convergence of moments, we get that

$$\lim_{n\to\infty}\sqrt{n}\;\mathbb{E}\big(W_p^p(\mu_n,\mu)\big)^{1/p} \,=\, \bigg[\mathbb{E}\big(|Z|^p\big)\int_0^1 \big(x(1-x)\big)^{p/2}\,dx\bigg]^{1/p}$$

where $Z\sim N(0,1)$. Here $[\mathbb{E}(|Z|^p)]^{1/p}$ is of order $\sqrt{p}$.

It is also of interest to compare Theorem 4.9 with the asymptotics (without any normalization)

$$\lim_{p\to\infty}\mathbb{E}\big(W_p^p(\mu_n,\nu_n)\big)^{1/p} \,=\, 1,$$

where ν_n is an independent copy of μ_n, which can be made on the basis of Theorem 4.6. Indeed, since both μ_n and ν_n are supported on $[0,1]$ (with probability one), we have $W_p(\mu_n,\nu_n)\leq 1$. On the other hand, from (4.3), for any $k=1,\dots,n$,

$$\big[\,\mathbb{E}\big(W_p^p(\mu_n,\nu_n)\big)\big]^{1/p} \,\geq\, \bigg[\frac{1}{n}\int_0^1\!\!\int_0^1 \big|t-s\big|^p\;dB_{k,n-k+1}(t)\,dB_{k,n-k+1}(s)\bigg]^{1/p} \,\to\, 1$$

as $p\to\infty$.

5 Standard rate for $\mathbb{E}(W_p^p(\mu_n,\mu))$

In this main section, we study the rates for the mean transport distances $\mathbb{E}(W_p(\mu_n,\mu))$ or $\mathbb{E}(W_p^p(\mu_n,\mu))^{1/p}$, with a special focus on bounds that provide the standard $\frac{1}{\sqrt{n}}$-rate. Here, as usual, μ is a Borel probability measure on $\mathbb{R}$ with distribution function F, and $\mu_n = \frac{1}{n}\sum_{k=1}^n \delta_{X_k}$, $n \geq 1$, are the empirical measures on a sample $(X_k)_{k\geq 1}$ of independent random variables with common law μ.

Unlike the case $p = 1$, it actually turns out that, even if μ is compactly supported, an extra condition on μ is needed in order to insure that $W_p(\mu_n,\mu)$ with $p > 1$ is of order $\frac{1}{\sqrt{n}}$. A complete characterization of the standard rate for $\mathbb{E}(W_p(\mu_n,\mu))$ will be provided in this part in terms of the functionals

$$J_p(\mu) \,=\, \int_{-\infty}^{\infty} \frac{[F(x)(1-F(x))]^{p/2}}{f(x)^{p-1}}\,dx, \tag{5.1}$$

where, as before, F is the distribution function associated with μ and f is the density of the absolutely continuous component of μ. A first step is to achieve upper-bounds with the help of the functional $J_p(\mu)$ (starting with the important example $p = 2$). In this process, a further analysis will involve the I-function

$$I(t) \,=\, I_F(t) \,=\, f\big(F^{-1}(t)\big), \qquad 0 < t < 1, \tag{5.2}$$

of the distribution function F leading to sufficient conditions in terms of Cheeger-type inequalities. Necessity and lower-bounds are examined next, based on the refined lower integral bounds of Appendix B. The section is completed with a study of the standard rate in the metric W_∞. We refer to Appendix A for a complete account on inverse functions and the associated I-functions and to Appendix B for material on beta distributions and their densities.

5.1 General upper-bounds on $\mathbb{E}(W_2^2(\mu_n,\mu))$

This first paragraph is concerned with the important case $p = 2$ and states one of the general results.

Given a probability meausure μ on the real line $\mathbb{R}$, recall from (5.1) the J_2-functional

$$J_2(\mu) \,=\, \int_{-\infty}^{\infty} \frac{F(x)(1-F(x))}{f(x)}\,dx. \tag{5.3}$$

We agree that $\frac{0}{0} = 0$, and if μ is a delta-measure, that $J_2(\mu) = 0$.

Theorem 5.1 (Upper-bound on $\mathbb{E}(W_2^2(\mu_n,\mu))$). *For every $n \geq 1$,*

$$\mathbb{E}\big(W_2^2(\mu_n,\mu)\big) \,\leq\, \frac{2}{n+1}\,J_2(\mu).$$

Moreover, if ν_n is an independent copy of μ_n,

$$\mathbb{E}\big(W_2^2(\mu_n,\nu_n)\big) \,\leq\, \frac{2}{n+1}\,J_2(\mu).$$

The theorem is applicable when the functional $J_2(\mu)$ is finite, which is a stronger property than having a finite second moment (like finiteness of $J_1(\mu)$ is stronger than finiteness of the first absolute moment). For example, if μ has density of the form $f(x) = c_\alpha\, e^{-|x|^\alpha}$, $x \in \mathbb{R}$, with a parameter $\alpha \geq 1$, then $J_2(\mu) < \infty$ for $\alpha > 2$, but $J_2(\mu) = \infty$ for all $1 \leq \alpha \leq 2$.

As we know from Appendix A, for the finiteness of $J_2(\mu)$, it is necessary that μ be supported on an interval Δ of the real line, finite or not, and f be a.e. positive on it (and then integration in the definition of J_2 should be restricted to Δ). An equivalent approach to this definition is to require that the inverse distribution function F^{-1} be absolutely continuous on $(0,1)$, in which case

$$J_2(\mu) \,=\, \int_0^1 \big[(F^{-1})'(t)\big]^2\, t(1-t)\, dt$$

where $(F^{-1})'$ is the Radon-Nikodym derivative (cf. Corollary A.22). One can also represent this derivative in terms of the associated I-function $I(t) = f(F^{-1}(t))$, $0 < t < 1$, of (5.2), so that

$$J_2(\mu) \,=\, \int_0^1 \frac{t(1-t)}{I(t)^2}\, dt.$$

For more details we refer to sub-Sections A.4 and A.5 in Appendix A.

Proof of Theorem 5.1. Assume that $J_2(\mu)$ is finite. In particular, the inverse function F^{-1} is absolutely continuous on $(0,1)$. First, we recall from sub-Section 4.1 that $\mathbb{E}(W_2^2(\mu_n,\mu)) \leq \mathbb{E}(W_2^2(\mu_n,\nu_n))$. To derive the second bound of the theorem, let us rewrite the representation of Theorem 4.6 in terms of the mean beta distribution B_n of order n,

$$B_n \,=\, \frac{1}{n}\sum_{k=1}^n B_{k,n-k+1} \otimes B_{k,n-k+1},$$

as

$$\mathbb{E}\big(W_2^2(\mu_n,\nu_n)\big) \,=\, \int_0^1\!\int_0^1 \big[F^{-1}(t) - F^{-1}(s)\big]^2\, dB_n(t,s).$$

It remains to apply Proposition B.13 with $u = F^{-1}$ which then yields

$$\int_0^1\!\int_0^1 \big[F^{-1}(t) - F^{-1}(s)\big]^2\, dB_n(t,s) \,\leq\, \frac{2}{n+1}\int_0^1 \big[F^{-1\prime}(t)\big]^2\, t(1-t)\, dt.$$

□

Let us now return to the canonical representation from Theorem 2.10,

$$W_2^2(\mu_n, \mu) \,=\, \int_0^1 \big[F_n^{-1}(t) - F^{-1}(t)\big]^2 \, dt \,=\, \frac{1}{n} \int_0^1 \xi_n(t)^2 \, dt$$

in terms of the empirical quantile processes

$$\xi_n(t) \,=\, \sqrt{n}\,\big(F_n^{-1}(t) - F^{-1}(t)\big), \qquad 0 < t < 1.$$

Here, as before, F_n^{-1} denotes the inverse of the empirical distribution function F_n associated to the sample $X_1, \dots, X_n$ drawn from the distribution μ with the distribution function F. Since F^{-1} is distributed according to μ under the Lebesgue measure on $(0,1)$, the trajectories of ξ_n belong to $L^2(0,1)$ if and only if μ has a finite second moment. Under this condition, one may therefore wonder whether or not, the distributions of ξ_n (as probability measures on this Hilbert space) have a non-degenerate weak limit in $L^2(0,1)$ in analogy with a similar property for the distance W_1. This question was solved in the work by del Barrio, Giné and Utzet [B-G-U] under the additional regularity assumption on the distribution of the sample, namely that the measure μ is supported on an interval (a,b), where it has a positive differentiable density f such that

$$\sup_{a<x<b} \frac{F(x)(1-F(x))}{f(x)^2}\, \big|f'(x)\big| \,<\, \infty. \tag{5.4}$$

This condition is going back to [C-R] (see also [S-W]). Note that in terms of the function $I(t) = f(F^{-1}(t))$, (5.4) may be rewritten as

$$\sup_{0<t<t} \frac{t(1-t)}{I(t)}\, \big|I'(t)\big| \,<\, \infty.$$

One of the main results of [B-G-U] may then be stated as follows. As before, denote by W^o the standard Brownian bridge.

Theorem 5.2. (B-G-U) *Under the regularity assumption* (5.4)*, if* $J_2(\mu) < \infty$*, the distributions of the quantile processes* ξ_n *are weakly convergent in* $L^2(0,1)$ *to the distribution of the random process* $W^o(t)/I(t)$*. Moreover, in this case*

$$n\, W_2^2(\mu_n, \mu) \,\to\, \int_0^1 \frac{W^o(t)^2}{I(t)^2}\, dt$$

as $n \to \infty$ *weakly in distribution on the real line.*

In the same work [B-G-U], del Barrio, Giné and Utzet have also studied an asymptotic behaviour of the empirical distributions and the distributions of $nW_2^2(\mu_n, \mu)$ in the case where $J_2(\mu)$ is infinite.

5.2 General upper-bounds on $\mathbb{E}(W_p^p(\mu_n, \mu))$

Theorem 5.1 admits a natural extension to the general case $p \geq 1$ with the J_p-functional of (5.1). The following result is a combination of Theorem 4.6 with Proposition B.13 from Appendix B similarly to the case $p = 2$.

Theorem 5.3 (Upper-bound on $\mathbb{E}(W_p^p(\mu_n, \mu))$). *For any $p \geq 1$, and any $n \geq 1$,*

$$\mathbb{E}\big(W_p^p(\mu_n, \mu)\big) \,\leq\, \Big(\frac{5p}{\sqrt{n+2}}\Big)^p J_p(\mu).$$

In particular,

$$\big\|W_p(\mu_n, \mu)\big\|_p \,\leq\, \frac{5p}{\sqrt{n}}\, J_p^{1/p}(\mu),$$

which is effective when the integral $J_p(\mu)$ is finite.

In the case $p = 2$, the preceding amounts to Theorem 5.1 although with a worse absolute constant. The same holds true for $p = 1$ in which case the statement is covered by Theorem 3.2 with a worse constant and with an additional assumption about the absolutely continuous component of μ. The latter may actually be removed by a simple approximation argument (since the density f does not appear explicitly in J_1).

One immediate consequence of Theorem 5.1 (and essentially of the Poincaré-type inequality of Proposition B.8) is that the finiteness of the integral $J_p(\mu)$ implies the finiteness of the p-th absolute moment of μ. Indeed, since $W_p^p(\mu_n, \mu) < \infty$ a.s. and since μ_n is compactly supported, necessarily $\mathbb{E}(|X|^p) < \infty$. This fact is not so obvious if we only look at the definition of J_p.

As in the case $p = 2$, if $p > 1$, for the finiteness of $J_p(\mu)$, the measure μ should be supported on an interval, and f should be a.e. positive on it. In this case, in terms of the I-function of (5.2), the definition of $J_p(\mu)$ becomes more natural

$$J_p(\mu) \,=\, \int_0^1 \bigg(\frac{\sqrt{t(1-t)}}{I(t)}\bigg)^p\, dt. \tag{5.5}$$

In particular, it may be seen that the quantities $J_p^{1/p}(\mu)$ grow with p.

Often, the last representation (5.5) is more convenient for determining whether or not $J_p(\mu)$ is finite. For example, if μ is standard normal, the function I is symmetric about $t = \frac{1}{2}$ and $I(t) \sim t\sqrt{2\log(1/t)}$ as $t \to 0$. Therefore, $J_p(\mu) < \infty$ if and only if $1 \leq p < 2$. The value $p = 2$ is indeed critical for the Gaussian measure, since as we will see, the relation $\mathbb{E}(W_2(\mu_n, \mu)) = O(\frac{1}{\sqrt{n}})$ does not hold true in this case.

Actually, given $\alpha > 1$, let μ have density $c_\alpha\, e^{-|x|^\alpha}$, $x \in \mathbb{R}$, where $c_\alpha > 0$ is a normalizing constant. For large $x > 0$, the associated distribution function F has tails

$$\begin{aligned} 1 - F(x) &= \frac{c_\alpha}{\alpha} \int_x^\infty \frac{1}{y^{\alpha-1}}\, d\left(-e^{-y^\alpha}\right) \\ &= \frac{c_\alpha}{\alpha} \left[\frac{1}{x^{\alpha-1}}\, e^{-x^\alpha} - (\alpha - 1) \int_x^\infty \frac{1}{y^\alpha}\, e^{-y^\alpha}\, dy \right] \\ &= \frac{c_\alpha}{\alpha}\, \frac{1}{x^{\alpha-1}}\, e^{-x^\alpha} \left(1 + O(x^{-1})\right). \end{aligned}$$

Hence, the I-function is symmetric about the point $t = \frac{1}{2}$ and has an asymptotic $I(t) \sim \mathrm{const} \cdot t\,(\log(1/t))^{1/\alpha^*}$ as $t \to 0$, where $\alpha^* = \frac{\alpha}{\alpha-1}$ is the conjugate exponent. Therefore, $J_p(\mu) < \infty$ if and only if $1 \le p < 2$. This conclusion applies, in particular, to Gaussian measures.

On the other hand, the property $J_p(\mu) < \infty$ with $1 \le p < 2$ is true for a large family of probability distributions μ on the line. This will be discussed in further details in the next section. At this point, let us describe one class of compactly supported distributions for which the standard rate is applicable regardless of the range of p.

Corollary 5.4. *If the I-function of μ satisfies*

$$I(t) \,\ge\, c\,\sqrt{t(1-t)}, \qquad 0 < t < 1, \tag{5.6}$$

for some constant $c > 0$, then, for any $p \ge 1$,

$$\left[\mathbb{E}\left(W_p^p(\mu_n, \mu)\right)\right]^{1/p} \,\le\, \frac{5p}{c\sqrt{n}}\,.$$

In particular, the uniform distribution on $[0, 1]$ belongs to this class, so we extend Theorem 4.8, although with a worse behaviour of the p-dependent constants.

Note that all probability measures such that $I(t) \ge c\sqrt{t(1-t)}$ have a compact support. Indeed, if (a, b) is a supporting interval of μ, then, by Corollary A.23,

$$b - a \,=\, \int_0^1 \frac{dt}{I(t)} \,\le\, \frac{1}{c} \int_0^1 \frac{dt}{\sqrt{t(1-t)}} \,=\, \frac{\pi}{c}\,.$$

As we know from sub-Section A.6 in Appendix A, these measures may be described as Lipschitz transforms (with Lipschitz seminorm not greater than $1/c$) of a special symmetric probability distribution ν whose associated I-function is exactly $I(t) = \sqrt{t(1-t)}$. Namely, ν is supported on the interval $[-\pi, \pi]$ and has there density

$$g(x) \,=\, \frac{1}{4}\, \cos\Big(\frac{x}{2}\Big), \qquad |x| < \pi.$$

To relate with other known distribution, let us note that the beta distribution with parameters $\alpha = \beta = 2$ has a very similar I-function. Indeed, the beta density and distribution function are given in this case by

$$f_{2,2}(x) \,=\, 6x(1-x), \qquad F_{2,2}(x) \,=\, x^2(3-2x), \qquad 0 \le x \le 1.$$

For $x \in [0, \frac{1}{2}]$, we have $3x \le f_{2,2}(x) \le 6x$ and $2x^2 \le F_{2,2}(x) \le 3x^2$. The latter implies $\sqrt{\frac{t}{3}} \le F_{2,2}^{-1}(t) \le \sqrt{\frac{t}{2}}$ for $t \in [0, \frac{1}{2}]$. Since $F_{2,2}^{-1}(t) \in [0, \frac{1}{2}]$, we get that

$$f_{2,2}\big(F_{2,2}^{-1}(t)\big) \ \le\ 6F_{2,2}^{-1}(t) \ \le\ 6\sqrt{\frac{t}{2}} \ \le\ 6\sqrt{t(1-t)}$$

and

$$f_{2,2}\big(F_{2,2}^{-1}(t)\big) \ \ge\ 3F_{2,2}^{-1}(t) \ \ge\ \sqrt{3t} \ \ge\ \sqrt{3t(1-t)}.$$

Since the function $I_{2,2}(t) = f_{2,2}(F_{2,2}^{-1}(t))$ is symmetric about the point $t = 1/2$, the final estimates remain to hold for $t \in [\frac{1}{2}, 1]$ as well. Hence,

$$\sqrt{3t(1-t)} \le I_{2,2}(t) \ \le\ 6\sqrt{t(1-t)}, \qquad 0 < t < 1,$$

Therefore, the probability measures satisfying (5.6) in Corollary 5.4 can also be obtained as Lipschitz transforms (with Lipschitz seminorm not greater than $6/c$) of the beta distribution with parameters $\alpha = \beta = 2$.

In case of other beta distributions $B_{\alpha,\alpha}$ with general parameter $\alpha > 0$, the corresponding $I_{\alpha,\alpha}$-functions behave near zero like $t^{\alpha/(\alpha+1)}$ and therefore satisfy similar two-sided bounds

$$c_0\, \big(t(1-t)\big)^{\alpha/(\alpha+1)} \ \le\ I_{\alpha,\alpha}(t) \le c_1\, \big(t(1-t)\big)^{\alpha/(\alpha+1)}, \qquad 0 < t < 1,$$

up to some positive constants c_0 and c_1, depending on α.

5.3 Distributions with finite Cheeger constants

Theorem 5.3 may be applied to a large family of probability distributions μ, especially when $1 \le p \le 2$. For example, one may require that μ satisfies a Cheeger-type isoperimetric inequality, or, equivalently, a Sobolev-type inequality

$$h \int_{-\infty}^{\infty} \big|u(x) - m\big|\, d\mu(x) \ \le\ \int_{-\infty}^{\infty} \big|u'(x)\big|\, d\mu(x).$$

Here, u is an arbitrary absolutely continuous function with median $m = m(u)$ under μ, and $h \ge 0$ is a constant independent of u. An optimal value $h = h(\mu)$ in this inequality is called the Cheeger isoperimetric constant. It admits a simple explicit description

$$h(\mu) \ =\ \operatorname*{ess\,inf}_{x \in \mathbb{R}} \frac{f(x)}{\min\{F(x), 1 - F(x)\}}$$

where F is the distribution function associated to μ and f is the density of an absolutely continuous component of μ (cf. [B-H1]). In particular, for the property $h > 0$ it is necessary that f be supported and be a.e. positive on the supporting interval of μ.

For such measures, the inverse distribution function F^{-1} is absolutely continuous, and the associated function $I = 1/(F^{-1})'$ necessarily satisfies a lower-bound

$$I(t) \ \geq \ h \, \min\{t, 1-t\}, \qquad 0 < t < 1 \ \text{(a.e.)}$$

which readily implies the finiteness of $J_p(\mu) < \infty$ for the range $1 \leq p < 2$. Moreover, we have that

$$J_p(\mu) \ \leq \ 2h^{-p} \int_0^{1/2} t^{-p/2} \, dt < \frac{4}{2-p} \, .$$

Therefore, from Theorem 5.3, we obtain the following consequence.

Theorem 5.5 (Upper-bound on $\mathbb{E}((W_p^p(\mu_n, \mu))$ under Cheeger constant). *Let μ have a positive Cheeger constant h. Then, for any $1 \leq p < 2$ and any $n \geq 1$,*

$$\mathbb{E}\big(W_p^p(\mu_n, \mu)\big) \ \leq \ \frac{C}{2-p} \left(\frac{1}{h\sqrt{n}} \right)^p,$$

where $C > 0$ is an absolute constant.

Theorem 5.5 admits an alternative proof using the Lipschitz images as developed in sub-Section A.6 from Appendix A. Indeed, denote by ν the two-sided exponential distribution on the real line with density $\frac{1}{2}\, e^{-|x|}$. The canonical non-decreasing map $T : \mathbb{R} \to \mathbb{R}$ which pushes forward ν onto μ has the Lipschitz semi-norm $\|T\|_{\text{Lip}} \leq \frac{1}{h}$. Write the order statistics for the sample X_k as $X_k^* = T(U_k^*)$, where $U_1^*, \dots, U_n^*$ are order statistics in the sample taken from ν. Let $(V_1^*, \dots, V_n^*)$ be an independent copy of $(U_1^*, \dots, U_n^*)$ and put $Y_k^* = T(V_k^*)$. Then, by Theorem 4.3, and using the Lipschitz property of the map T, we get

$$\begin{aligned}
\mathbb{E}\big(W_p^p(\mu_n, \mu)\big) \ &\leq \ \frac{1}{n} \sum_{k=1}^n \mathbb{E}\big(|X_k^* - Y_k^*|^p\big) \\
&\leq \ \frac{1}{nh^p} \ \sum_{k=1}^n \mathbb{E}\big(|U_k^* - V_k^*|^p\big) \\
&\leq \ \frac{2^p}{h^p} \, \mathbb{E}\big(W_p^p(\nu_n, \nu)\big)
\end{aligned}$$

where ν_n is an empirical measure constructed for the sample drawn from ν. Thus, at the expense of a factor we are reduced in Theorem 5.5 to the particular case of the two-sided exponential distribution. In this case, estimates for $\mathbb{E}(|U_k^* - \mathbb{E}(U_k^*)|^p)$ can be explored directly.

5.4 Connectedness and absolute continuity

In this section, we turn to necessary conditions needed to get a standard rate for $\mathbb{E}(W_p(\mu_n, \mu))$. Being bounded by using $J_p(\mu)$, it has already been emphasized after

Theorem 5.1 for the case $p = 2$ (with reference to Appendix A) and after Theorem 5.3 for $p > 1$ that the finiteness of this functional includes the requirement that the distribution function F of μ has an absolutely continuous inverse function F^{-1}. In particular, the support of μ has to be an interval. What can one say therefore in this respect under the formally weaker assumption that $\mathbb{E}(W_p(\mu_n, \mu)) = O\big(\frac{1}{\sqrt{n}}\big)$?

As we will see later in Section 7, once μ has a disconnected support, the rate for $\mathbb{E}(W_p(\mu_n, \mu))$ cannot be asymptotically better than $n^{-1/2p}$. In case $p > 1$, this is of course worse than the standard rate. In fact, the validity of the standard rate requires more. The purpose of this section is to prove the following statement.

Theorem 5.6 (Necessary condition for the standard rate)**.** *Given $p > 1$, assume that*

$$\mathbb{E}\big(W_p(\mu_n, \mu)\big) = O\Big(\frac{1}{\sqrt{n}}\Big)$$

holds true along a subsequence $n = n_k \to \infty$ as $k \to \infty$. Then necessarily the inverse function F^{-1} is absolutely continuous on $(0,1)$.

This observation is partly based on the corresponding formulations of the lower-bounds appearing in Proposition B.18 of Appendix B. More precisely, the latter yields the following statement.

Theorem 5.7. *If ν_n is an independent copy of μ_n, then*

$$\mathbb{E}\big(W_p^p(\mu_n, \nu_n)\big) \, \geq \, c \int_{\{t(1-t) \geq \frac{4}{\sqrt{n+1}}\}} \Big[F^{-1}\Big(t + \frac{1}{2}\,\varepsilon_n(t)\Big) - F^{-1}\Big(t - \frac{1}{2}\,\varepsilon_n(t)\Big)\Big]^p \, dt$$

and

$$\big[\mathbb{E}\big(W_p(\mu_n, \nu_n)\big)\big]^p \, \geq \, c^p \int_{\{t(1-t) \geq \frac{4}{\sqrt{n+1}}\}} \Big[F^{-1}\Big(t + \frac{1}{6}\,\varepsilon_n(t)\Big) - F^{-1}\Big(t - \frac{1}{6}\,\varepsilon_n(t)\Big)\Big]^p \, dt$$

where $\varepsilon_n(t) = \sqrt{\frac{t(1-t)}{n+1}}$ and c is a positive numerical constant.

Proof. According to Theorem 4.3, $W_p^p(\mu_n, \nu_n) = \frac{1}{n}\sum_{k=1}^n |X_k^* - Y_k^*|^p$ where X_k^* is the k-th order statistic for the sample $X_1, \dots, X_n$ drawn from μ, and Y_k^* is an independent copy of X_k^*. Hence,

$$\mathbb{E}\big(W_p^p(\mu_n, \nu_n)\big) \; = \int_0^1 \int_0^1 \big|F^{-1}(x) - F^{-1}(y)\big|^p \, dB_n(x,y)$$

where B_n is the mean beta square distribution of order n. Therefore, the first inequality follows from the first bound of Proposition B.18.

In order to estimate $\mathbb{E}(W_p(\mu_n,\nu_n))$ from below, we use the following general inequality which is a variant of the triangle inequality. If $\xi_1,\dots,\xi_n$ are non-negative random variables, then for all $p \geq 1$,

$$\mathbb{E}\big((\xi_1^p+\dots+\xi_n^p)^{1/p}\big) \ \geq \ \big[\,\mathbb{E}(\xi_1)^p+\dots+\mathbb{E}(\xi_n)^p\big]^{1/p}.$$

This gives a lower-bound

$$\begin{aligned}\big[\,\mathbb{E}\big(W_p(\mu_n,\nu_n)\big)\big]^p \ &\geq \ \frac{1}{n}\sum_{k=1}^n \mathbb{E}\big(|X_k^*-Y_k^*|\big)^p \\ &= \ \frac{1}{n}\sum_{k=1}^n \mathbb{E}\big(\big|F^{-1}(U_k^*)-F^{-1}(V_k^*)\big|\big)^p.\end{aligned}$$

Here, U_k^* is the k-th order statistic for a sample of size n drawn from the uniform distribution on $(0,1)$ with its independent copy V_k^*. But U_k^* has the beta distribution $B_{k,n-k+1}$, so we may apply the second bound of Proposition B.18 with $u=F^{-1}$. $\square$

For the proof of Theorem 5.6, we need a general auxiliary lemma about integral moduli of continuity.

Lemma 5.8. *Let $p>1$, and let $u:(a,b)\to\mathbb{R}$ be a non-decreasing function such that*

$$\int_{a+h_n}^{b-h_n} \big[u(t+h_n)-u(t-h_n)\big]^p\,dt \ = \ O(h_n^p) \qquad (n\to\infty)$$

for some sequence $h_n\downarrow 0$. Then u has to be absolutely continuous on the interval (a,b).

Proof. Since the absolute continuity is understood in the local sense, one may assume that the interval (a,b) is finite, as well as the values $u(a+)$ and $u(b-)$. Then we need to show that, for any $\varepsilon>0$, there exists $\delta>0$ such that, for any sequence of non-overlapping intervals $(a_i,b_i)\subset[a,b]$,

$$\sum_i (b_i-a_i) \ = \ \delta \ \ \Longrightarrow \ \ \sum_i \big(u(b_i)-u(a_i)\big) \ \leq \ \varepsilon.$$

Without loss of generality, it suffices to require that there are finitely many of such intervals, say N, and all of them have equal length δ/N. Moreover, it suffices to consider such a property along any prescribed sequence $N=N_n\to\infty$ with sufficiently large n (this sequence may be chosen after ε is fixed).

Thus, fix $\varepsilon>0$ and put $\delta=(c\varepsilon)^q$, where $q=\frac{p}{p-1}$ is the conjugate power and $c>0$ is a constant to be chosen later on. Because of the constraint $\delta<b-a$, let ε be small enough. Moreover, put $N=[\frac{\delta}{h_n}]+1$.

Extending the function u as $u(t) = u(a+)$ for $t \le a$ and $u(t) = u(b-)$ for $t \ge b$ (and shrinking a little the interval (a,b) if necessary), the main hypothesis may be written as the inequality

$$\int_a^b \big[u(t+h_n) - u(t-h_n)\big]^p \, dt \, \le \, C^p h_n^p$$

holding with some constant C. Equivalently, for any measurable function $v \ge 0$ such that $\int_a^b v(t)^q \, dt \le 1$,

$$\int_a^b \big[u(t+h_n) - u(t-h_n)\big] v(t) \, dt \, \le \, C h_n.$$

We apply this inequality to a constant indicator function $v = \frac{1}{\delta^{1/q}} \, 1\!\!1_A$ with the set $A = \bigcup_{i=1}^N (a_i, b_i)$, in which case it becomes

$$\sum_{i=1}^N \int_{a_i}^{b_i} \big[u(t+h_n) - u(t-h_n)\big] \, dt \, \le \, C\delta^{1/q} \, h_n.$$

Note that $N = [\frac{\delta}{h_n}] + 1 > \frac{\delta}{h_n}$ implying $h_n > \frac{\delta}{N}$. Therefore, for any $t \in (a_i, b_i)$, we have $t - h_n < a_i < b_i < t + h_n$, so that by the monotonicity of u,

$$u(t+h_n) - u(t-h_n) \, \ge \, u(b_i) - u(a_i).$$

Hence, using again that $b_i - a_i = \delta/N$, we get

$$\frac{\delta}{N} \sum_{i=1}^N \big[u(b_i) - u(a_i)\big] \, \le \, C\delta^{1/q} \, h_n.$$

Finally, $\frac{\delta}{h_n} \ge 2$ for sufficiently large n, hence $N \le \frac{2\delta}{h_n}$ and thus $\frac{C\delta^{1/q} \, h_n}{\delta/N} \le 2C\delta^{1/q} = \varepsilon$ with $c = 1/(2C)$. The lemma is proved. $\square$

It would be interesting to know whether or not the statement of Lemma 5.8 continues to hold for arbitrary functions u (without monotonicity assumption). Let us describe an alternative argument for the class of functions of bounded variation, which however works for the case $p = 2$, only (and hence for all $p \ge 2$, since the statement is getting weaker when p grows).

Without loss of generality, let us start with the hypothesis $\int_{-\infty}^{\infty} (\Delta_h u(t))^2 \, dt \le C$ where, for $h > 0$,

$$\Delta_h u(t) \, = \, \frac{u(t+h) - u(t-h)}{2h},$$

assuming that u is a function of bounded variation with $u(-\infty) = u(\infty) = 0$. In particular, $\Delta_h u$ belongs to $L^2(\mathbb{R})$. Introduce the Fourier-Stieltjes transform

$$\hat u(x) \, = \, \int_{-\infty}^{\infty} e^{itx} \, du(t), \qquad x \in \mathbb{R}.$$

Integrating by parts, the Fourier transform of the function $\Delta_h u$ is given, for any $x \neq 0$, by

$$\begin{aligned}(\mathcal{F}\Delta_h u)(x) &= \int_{-\infty}^{\infty} e^{itx}\,\Delta_h u(t)\,dt \\ &\equiv \lim_{T\to\infty} \int_{-T}^{T} e^{itx}\,\Delta_h u(t)\,dt \ = \ \frac{\sin(hx)}{hx}\,\hat{u}(x).\end{aligned}$$

Hence, by Parseval's theorem,

$$\int_{-\infty}^{\infty} \Big(\frac{\sin(hx)}{hx}\Big)^2 \big|\hat{u}(x)\big|^2\,dx \ = \ 2\pi \int_{-\infty}^{\infty} \big(\Delta_h u(t)\big)^2\,dt.$$

Restricting the left integral to the interval $|xh| \leq 1$ and using $\sin(t) \geq t\sin(1)$, for $0 \leq t \leq 1$, we conclude that

$$\int_{-1/h}^{1/h} \big|\hat{u}(x)\big|^2\,dx \ \leq \ \frac{2\pi C}{\sin^2(1)}\,.$$

But, if this inequality holds true for all $h > 0$ small enough or even along some sequence $h = h_n \downarrow 0$, in the limit we get

$$\int_{-\infty}^{\infty} |\hat{u}(x)|^2\,dx \ \leq \ \frac{2\pi C}{\sin^2(1)}\,.$$

That is, $\hat{u}$ belongs to $L^2(\mathbb{R})$. But any function of bounded variation with a square integrable Fourier-Stieltjes transform is absolutely continuous and has a square integrable Radon-Nikodym derivative.

Proof of Theorem 5.6. Introducing an independent copy ν_n of μ_n, the hypothesis may be stated as the inequality

$$\mathbb{E}\big(W_p(\mu_n,\nu_n)\big) \ \leq \ \frac{C}{\sqrt{n}}, \qquad n = n_k,$$

holding with some constant C independent of k. Hence, under the hypothesis about the standard rate, we obtain with the help of the second bound of Theorem 5.7 that, for all $0 < t_0 < \frac{1}{2}$,

$$\int_{t_0}^{1-t_0} \big[F^{-1}(t+h_n) - F^{-1}\big(t-h_n\big)\big]^p\,dt \ = \ O(h_n^p) \qquad (n \to \infty)$$

with $h_n(t) = \frac{1}{6}\,\varepsilon_n(t_0)$. By Lemma 5.8, the function F^{-1} is absolutely continuous on $(t_0, 1-t_0)$ and thus it is absolutely continuous on the whole interval $(0,1)$. $\square$

5.5 Necessary and sufficient conditions

The inequality of Theorem 5.3 can be reversed in the following asymptotic forms.

Theorem 5.9. *For all $p \geq 1$,*

$$\liminf_{n\to\infty} \big[n^{p/2}\, \mathbb{E}\big(W_p^p(\mu_n,\mu)\big)\big] \,\geq\, cJ_p(\mu)$$

where $c > 0$ is an absolute constant. Moreover, for possibly another absolute constant $c > 0$,

$$\liminf_{n\to\infty} \big[\sqrt{n}\, \mathbb{E}\big(W_p(\mu_n,\mu)\big)\big]^p \,\geq\, c^p J_p(\mu).$$

Being combined, Theorems 5.3 and 5.9 provide the two-sided asymptotic bounds

$$\begin{aligned} c_p\, J_p(\mu) \ &\leq\ \liminf_{n\to\infty} \big[n^{p/2}\, \mathbb{E}\big(W_p^p(\mu_n,\mu)\big)\big] \\ &\leq\ \limsup_{n\to\infty} \big[n^{p/2}\, \mathbb{E}\big(W_p^p(\mu_n,\mu)\big)\big] \ \leq\ c_p'\, J_p(\mu) \end{aligned}$$

with some p-dependent constants. As a result, we obtain the characterization of all probability distributions μ on $\mathbb{R}$ to which μ_n are convergent in W_p at the standard rate.

Corollary 5.10 (Necessary and sufficient condition for the standard rate)**.** *Given $p > 1$, the following properties are equivalent:*

a) $\big[\mathbb{E}(W_p^p(\mu_n,\mu))\big]^{1/p} = O(\frac{1}{\sqrt{n}})$ *as* $n \to \infty$*;*
b) $\mathbb{E}(W_p(\mu_n,\mu)) = O(\frac{1}{\sqrt{n}})$ *as* $n \to \infty$*;*
c) The latter property holds true along a subsequence $n = n_k \to \infty$ *as* $k \to \infty$*;*
d) $J_p(\mu) < \infty$.

Recall that, according to Corollary 3.6, we also have

$$\mathbb{E}\big(W_1(\mu_n,\mu)\big) \,=\, O\Big(\frac{1}{\sqrt{n}}\Big) \iff J_1(\mu) \,<\, \infty.$$

A similar characterization about W_1 is also true when one considers the convergence at the standard rate along a subsequence $n = n_k$ growing to infinity. So, the case $p = 1$ may be included in Corollary 5.10.

Proof of Theorem 5.9. As a preliminary step, it was already proved in Theorem 5.6 that if

$$C \,=\, \liminf_{n\to\infty} \big[\sqrt{n}\, \mathbb{E}\big(W_p(\mu_n,\mu)\big)\big]^p \,<\, \infty,$$

then the inverse distribution function F^{-1} associated with μ is absolutely continuous. In particular, F^{-1} is differentiable a.e. (by the Lebesgue differentiation theorem).

Set, for $n \geq 1$, $t \in (0,1)$,

$$\xi_n(t) \,=\, F^{-1}\big(t + \kappa\, \varepsilon_n(t)\big) - F^{-1}\big(t - \kappa\, \varepsilon_n(t)\big),$$

where $\varepsilon_n(t) = \sqrt{\frac{t(1-t)}{n+1}}$ and $\kappa > 0$ is a parameter. Recall the lower-bound

$$\big[\,\mathbb{E}\big(W_p(\mu_n,\nu_n)\big)\big]^p \,\geq\, c^p \int_{\{t(1-t) \geq \frac{4}{\sqrt{n+1}}\}} \xi_n(t)^p \, dt$$

obtained in Theorem 5.7 with $\kappa = \frac{1}{6}$ and a positive numerical constant c, and where ν_n is an independent copy of μ_n. Using the triangle inequality $\mathbb{E}(W_p(\mu_n,\nu_n)) \leq 2\,\mathbb{E}(W_p(\mu_n,\mu))$ and applying our hypothesis, we may conclude that

$$c^p \int_{\{t(1-t) \geq \frac{4}{\sqrt{n+1}}\}} \big(\sqrt{n}\,\xi_n(t)\big)^p \, dt \,<\, 2^p\, C'$$

holds true for infinitely many n with any prescribed value $C' > C$. But for almost all $t \in (0,1)$,

$$\begin{aligned}\lim_{n\to\infty} \sqrt{n}\,\xi_n(t) \;&=\; \lim_{n\to\infty} \bigg[\frac{F^{-1}(t + \kappa\, \varepsilon_n(t)) - F^{-1}(t - \kappa\, \varepsilon_n(t))}{2\kappa\, \varepsilon_n(t)}\; 2\kappa\, \varepsilon_n(t)\sqrt{n}\,\bigg] \\ &=\; (F^{-1})'(t) \cdot 2\kappa \sqrt{t(1-t)}.\end{aligned}$$

Applying Fatou's lemma, we then get

$$\int_0^1 \Big[(F^{-1})'(t)\sqrt{t(1-t)}\,\Big]^p \, dt \,\leq\, \frac{6^p\, C'}{c^p}.$$

Here, the left integral is exactly $J_p(\mu)$, thus proving the second assertion of the theorem.

The first assertion is proved by a similar argument – we then need to apply the first lower integral bound of Theorem 5.7 with $\kappa = \frac{1}{2}$. Theorem 5.9 is therefore established. □

5.6 Standard rate for W_∞ distance

In the last section of this part, we turn to the limit case $p = \infty$. Again, let $(X_k)_{k\geq 1}$ be a sample drawn from a distribution μ on the real line with the distribution function F, and let μ_n be the empirical measure μ_n constructed for the first n observations. As we know from Theorem 5.3, if the inverse function F^{-1} is absolutely continuous, then

$$\mathbb{E}\big(W_p(\mu_n,\mu)\big) \,\leq\, \frac{5p}{\sqrt{n}}\, J_p^{1/p}(\mu)$$

where

$$J_p(\mu) \ = \ \int_0^1 \Big[(F^{-1})'(t)\sqrt{t(1-t)}\,\Big]^p\,dt.$$

However, in the limit as $p \to \infty$, this general upper-bound does not yield a reasonable inequality, since the involved p-dependent constants grow to infinity. On the other hand, there is a limit

$$J_\infty(\mu) = \lim_{p\to\infty} J_p^{1/p}(\mu) \ = \ \operatorname*{ess\,sup}_{0<t<1}\,(F^{-1})'(t). \tag{5.7}$$

As an equivalent definition, one may also write

$$J_\infty(\mu) \ = \ \|F^{-1}\|_{\mathrm{Lip}}.$$

Therefore, if this functional is finite, it is natural to expect that a similar bound still holds for $\mathbb{E}(W_\infty(\mu_n,\mu))$ in terms of J_∞.

Theorem 5.11. *For some positive numerical constants c_0 and c_1,*

$$\mathbb{E}\big(W_\infty(\mu_n,\mu)\big) \ \le \ \frac{c_1}{\sqrt{n}}\,J_\infty(\mu)$$

and

$$\liminf_{n\to\infty}\ \sqrt{n}\ \mathbb{E}\big(W_\infty(\mu_n,\mu)\big) \ \ge \ c_0\,J_\infty(\mu).$$

Necessary and sufficient conditions ensuring the finiteness of $J_\infty(\mu)$ are given in Proposition A.25. As a result, we obtain the following characterization complementing the characterization in case of finite values of p (Corollary 5.10).

Corollary 5.12 (Standard rate for $\mathbb{E}(W_\infty(\mu_n,\mu))$)**.** *If μ is non-degenerate, the following properties are equivalent:*

a) $\mathbb{E}(W_\infty(\mu_n,\mu)) = O(\frac{1}{\sqrt{n}})$ *as $n \to \infty$;*

b) This property holds true along a subsequence $n = n_k \to \infty$ as $k \to \infty$;

c) $J_\infty(\mu) < \infty$*;*

d) μ is supported on a finite interval Δ, and the absolutely continuous component of μ has a density f which is separated from zero on Δ.

If one of these equivalent properties holds,

$$J_\infty(\mu) \ = \ \frac{1}{\operatorname{ess\,inf}_{x\in\Delta} f(x)}\,.$$

Proof of Theorem 5.11. For the upper-bound, the argument is based on Theorem 4.9 (i.e., in essence, on the Dvoretzky-Kiefer-Wolfowitz theorem). Indeed, assuming that $J_\infty(\mu)$ is finite, that is, F^{-1} is Lipschitz, write $X_k^* = F^{-1}(U_k^*)$, $Y_k^* = F^{-1}(V_k^*)$, where U_k^* is the k-th order statistic for a sample of size n drawn from the uniform distribution

μ' on $(0,1)$, and V_k^* is an independent copy of U_k^*. If ν_n is an independent copy of μ_n, we get, by Lemma 4.2,

$$\begin{aligned} W_\infty(\mu_n,\nu_n) &= \max_{1\le k\le n} |X_k^* - Y_k^*| \\ &= \max_{1\le k\le n} \big|F^{-1}(U_k^*) - F^{-1}(V_k^*)\big| \ \le \ \|F^{-1}\|_{\rm Lip} \max_{1\le k\le n} |U_k^* - V_k^*|. \end{aligned}$$

Also, using the convexity of the functional $\nu \mapsto W_\infty(\mu,\nu)$,

$$W_\infty(\mu_n,\mu) \ \le \ \mathbb{E}_V\big(W_\infty(\mu_n,\nu_n)\big)$$

where $\mathbb{E}_V$ denotes expectation with respect to the random vector $(V_1^*,\dots,V_n^*)$. The two bounds give

$$\begin{aligned} \mathbb{E}\big(W_\infty(\mu_n,\mu)\big) &\le J_\infty(\mu)\,\mathbb{E}\Big(\max_{1\le k\le n} |U_k^* - V_k^*|\Big) \\ &= J_\infty(\mu)\,\mathbb{E}\big(W_\infty(\mu_n',\nu_n')\big) \\ &\le 2J_\infty(\mu)\,\mathbb{E}\big(W_\infty(\mu_n',\mu')\big), \end{aligned}$$

where μ_n' is the empirical measure for the sample $U_1,\dots,U_n$ with its independent copy ν_n'. By Theorem 4.9, the last expectation does not exceed $C/\sqrt{n}$.

To get the lower-bound, one may just apply Theorem 5.9 for every $p \ge 1$ to obtain

$$\liminf_{n\to\infty} \sqrt{n}\,\mathbb{E}\big(W_\infty(\mu_n,\mu)\big) \ \ge \ \liminf_{n\to\infty} \sqrt{n}\,\mathbb{E}\big(W_p(\mu_n,\mu)\big) \ \ge \ c\,J_p^{1/p}(\mu),$$

where $c > 0$ is an absolute constant. It remains to let $p \to \infty$. The proof is complete. $\square$

6 Sampling from log-concave distributions

This section turns to the question of obtaining two-sided bounds on the mean transport distances $\mathbb{E}(W_p^p(\mu_n, \mu))$ for samples drawn from log-concave distributions. To this end, the chosen route of study is the one based on the application of Theorem 4.3. It is therefore to focus, as a preliminary step, on the study of the variances of order statistics associated to a sample drawn from a log-concave distribution. We next achieve two-sided bounds on $\mathbb{E}(W_p^p(\mu_n, \mu))$ in terms of the I-function associated to μ and obtain (by different means than the ones developed in the previous section in the general case) necessary and sufficient condition for the standard rate. Upper-bounds involving the variance of μ are studied next. Several examples of interest, with specific rates, are discussed in the last paragraph.

The notation of this section are the same as in the preceding sections. The measure μ is a Borel probability on $\mathbb{R}$ with distribution function F. If $(X_k)_{k\geq 1}$ is a sequence of independent random variables with common law μ, for each $n \geq 1$, μ_n denotes the empirical measure $\frac{1}{n}\sum_{k=1}^n \delta_{X_k}$. If $(X_1, \dots, X_n)$ is a sample of random variables, $(X_1^* \leq \cdots \leq X_n^*)$ denotes the associated order statistics.

We refer besides to Appendix B and the references therein for some basic facts on log-convave measures.

6.1 Bounds on variances of order statistics

If $(X_1, \dots, X_n)$ is drawn independently from a common absolutely continuous law μ with distribution function F and density f, the k-th order statistic X_k^*, $1 \leq k \leq n$, has density

$$f_k(x) \,=\, n\, C_{n-1}^{k-1}\, F(x)^{k-1}\big(1 - F(x)\big)^{n-k} f(x), \qquad x \in \mathbb{R}.$$

This explicit formula implies that, if the function f (and therefore F and $1 - F$) is log-concave, all the functions f_k will be log-concave as well. As another explanation, one may note that the joint distribution μ^n of the sample is log-concave on $\mathbb{R}^n$, so the joint distribution of $(X_1^*, \dots, X_n^*)$ is log-concave as the normalized restriction of μ^n to the cone $x_1 \leq \cdots \leq x_n$ in $\mathbb{R}^n$.

Moreover, once X_k^* has a log-concave distribution, it shares all inequalities mentioned in the general case (cf. sub-Section B.1 of Appendix B). In particular, we have the following relations implied by Proposition B.2,

$$\frac{1}{12\,\mathrm{Var}(X_k^*)} \,\leq\, \operatorname*{ess\,sup}_{x\in\mathbb{R}} f_k(x)^2 \,\leq\, \frac{1}{\mathrm{Var}(X_k^*)}\,.$$

To proceed, let thus μ be a probability measure on the real line with a log-concave density f supported on some interval $(a, b) \subset \mathbb{R}$, finite or not. Recall the associated I-function from Definition A.20

$$I(t) \,=\, I_F(t) \,=\, f\big(F^{-1}(t)\big), \qquad 0 < t < 1, \tag{6.1}$$

where $F^{-1} : (0,1) \to (a,b)$ is the inverse of the distribution function $F(x) = \mu((-\infty, x])$ restricted to $x \in (a,b)$. As is well known (cf. e.g. [Bob2]), μ is log-concave if and only if I is positive and concave on $(0,1)$. (Moreover, in general any such function I generates a certain log-concave probability measure on the real line, which is unique up to a shift parameter.) In terms of the I-function for this measure μ, the density f_k of the k-th order statistic X_k^* has maximum (essential supremum)

$$\begin{aligned} \operatorname*{ess\,sup}_{x\in\mathbb{R}} f_k(x) &= n\,C_{n-1}^{k-1} \sup_{a<x<b} F(x)^{k-1}\big(1-F(x)\big)^{n-k} f(x) \\ &= n\,C_{n-1}^{k-1} \sup_{0<t<1} t^{k-1}(1-t)^{n-k}\, I(t). \end{aligned}$$

As a result, we obtain a preliminary description of the variance in terms of the associated I-function.

Lemma 6.1. *If μ is log-concave with associated I-function I, for any $k = 1, \dots, n$,*

$$\frac{1}{\sqrt{12}\ \sup_{0<t<1} I_{k,n}(t)} \le \sqrt{\mathrm{Var}(X_k^*)} \le \frac{1}{\sup_{0<t<1} I_{k,n}(t)}$$

where

$$I_{k,n}(t) = nC_{n-1}^{k-1}\, t^{k-1}(1-t)^{n-k}\, I(t), \qquad 0<t<1.$$

For example, when μ is a uniform distribution on the unit interval $(0,1)$, then $I(t) = 1$, and $I_{k,n}(t) = nC_{n-1}^{k-1}\, t^{k-1}(1-t)^{n-k}$ is the density of the beta distribution with parameters $(k, n-k+1)$. In this case, $\sup_{0<t<1} I_{k,n}(t)$ is attained at the mode $t = \frac{k-1}{n-1}$ $(n \ge 2)$, but the maximum itself represents a rather complicated expression in variables (k,n). In an opposition direction, recall that for the sample from the uniform distribution we have the simple description

$$\mathrm{Var}(X_k^*) = \frac{k(n-k+1)}{(n+1)^2\,(n+2)}\,.$$

As a consequence of the preceding lemma, we get:

Corollary 6.2. *Given $k = 1, \dots, n$, the maximum*

$$M_{k,n} = \max_{0\le t\le 1} n\,C_{n-1}^{k-1}\, t^{k-1}(1-t)^{n-k}$$

satisfies

$$\frac{1}{12\,M_{k,n}^2} \le \frac{k(n-k+1)}{(n+1)^2\,(n+2)} \le \frac{1}{M_{k,n}^2}\,.$$

To make the bound of Lemma 6.1 effective, we need to properly bound from above and below the maximum of the function $I_{k,n}$. To this aim, a first step is achieved in the next lemma.

Lemma 6.3. *If I is a non-negative convex function on $(0,1)$, for any $k = 1, \dots, n$,*

$$I_{k,n}(t_{k,n}) \le \sup_{0<t<1} I_{k,n}(t) \le e^2 \, I_{k,n}(t_{k,n})$$

where $t_{k,n} = \frac{k}{n+1}$.

Proof. Only the right-hand side inequality requires a proof. To this end, we study the constant C in an inequality of the form $I_{k,n}(t) \le C \, I_{k,n}(t_{k,n})$ for every $0 < t < 1$.

First, assume that I is non-decreasing. Then I represents an envelope of a family of non-negative non-decreasing affine functions $l(t) = a + bt$ on $(0,1)$ with necessarily $a \ge 0$ and $b \ge 0$. The task thus reduces to showing that

$$t^{k-1}(1-t)^{n-k} \, l(t) \le C \, t_{k,n}^{k-1}(1-t_{k,n})^{n-k} \, l(t_{k,n}),$$

for any l with the decribed properties. Actually, since $a, b \ge 0$, the above would follow from the two inequalities

$$t^{k-1}(1-t)^{n-k} \le C \, t_{k,n}^{k-1} \, (1-t_{k,n})^{n-k} \tag{6.2}$$

and

$$t^{k}(1-t)^{n-k} \le C \, t_{k,n}^{k} \, (1-t_{k,n})^{n-k}. \tag{6.3}$$

In (6.2), $t = \frac{k-1}{n-1}$ is a point of maximum for the left-hand side ($n > 1$), so this inequality becomes

$$\left(\frac{k-1}{n-1}\right)^{k-1} \left(\frac{n-k}{n-1}\right)^{n-k} \le C \left(\frac{k}{n+1}\right)^{k-1} \left(\frac{n-k+1}{n+1}\right)^{n-k}.$$

The latter readily follows from $(\frac{n+1}{n-1})^{n-1} = (1 + \frac{2}{n-1})^{n-1} \le C$, so $C = e^2$ works. Also, if $k = n = 1$, the inequality is immeadiate with $C = 1$. In the second inequality (6.3), the left-hand side is maximized for $t = \frac{k}{n}$, and the inequality becomes

$$\left(\frac{k}{n}\right)^{k} \left(\frac{n-k}{n}\right)^{n-k} \le C \left(\frac{k}{n+1}\right)^{k} \left(\frac{n-k+1}{n+1}\right)^{n-k}.$$

This follows from $\left(\frac{n+1}{n}\right)^n = \left(1 + \frac{1}{n}\right)^n \le C$ with $C = e$. Therefore, the lemma is proved when I is non-decreasing.

Now, assume that I is non-increasing. Then I represents an envelope of a family of non-negative non-increasing affine functions $l(t) = a + b(1-t)$ on $(0,1)$ with necessarily $a \ge 0$ and $b \ge 0$. As above, the task reduces to showing that

$$t^{k-1}(1-t)^{n-k} \, l(t) \le C \, t_{k,n}^{k-1}(1-t_{k,n})^{n-k} \, l(t_{k,n}),$$

which would follow from the two inequalities

$$\begin{aligned} t^{k-1}(1-t)^{n-k} &\le C \, t_{k,n}^{k-1} \, (1-t_{k,n})^{n-k}, \\ t^{k-1}(1-t)^{n-k+1} &\le C \, t_{k,n}^{k-1} \, (1-t_{k,n})^{n-k+1}. \end{aligned}$$

Here, the first inequality was already considered in the previous step and was derived with constant $C = e^2$. In the second one, the left-hand side is maximized on $[0,1]$ for $t = \frac{k-1}{n}$, and the inequality becomes

$$\left(\frac{k-1}{n}\right)^{k-1}\left(\frac{n-k+1}{n}\right)^{n-k+1} \leq C\left(\frac{k}{n+1}\right)^{k-1}\left(\frac{n-k+1}{n+1}\right)^{n-k+1}.$$

Again, this follows from $\left(\frac{n+1}{n}\right)^n = \left(1+\frac{1}{n}\right)^n \leq C$ with $c = e$. Therefore, the lemma is also proved when I is non-increasing.

Now, consider the remaining case: For some $t_0 \in (0,1)$, I is non-decreasing on $(0,t_0]$ and is non-increasing on $[t_0,1)$.

Case 1: $t_{k,n} \leq t_0$. Introduce the function

$$\tilde{I}(t) \ = \ \begin{cases} I(t), & \text{if} \quad 0 < t \leq t_0, \\ I(t_0), & \text{if} \quad t_0 \leq t < 1. \end{cases}$$

Clearly, $\tilde{I}$ is non-decreasing, so by the previous step, for any $t \in (0,1)$,

$$t^{k-1}(1-t)^{n-k}\tilde{I}(t) \leq e^2\, t_{k,n}^{k-1}(1-t_{k,n})^{n-k}\tilde{I}(t_{k,n}).$$

But $\tilde{I} \geq I$, while $\tilde{I}(t_{k,n}) = I(t_{k,n})$, so we arrive at the desired inequality.

Case 2: $t_{k,n} \geq t_0$. Introduce the function

$$\tilde{I}(t) \ = \ \begin{cases} I(t_0), & \text{if} \quad 0 < t \leq t_0, \\ I(t), & \text{if} \quad t_0 \leq t < 1. \end{cases}$$

Then, $\tilde{I}$ is non-increasing, and a similar argument leads to the conclusion of the lemma. □

Lemma 6.1 and Lemma 6.3 may be combined to get a two-sided bound

$$\frac{1}{12\, e^4\, I_{k,n}^2(t_{k,n})} \ \leq \ \mathrm{Var}(X_k^*) \ \leq \ \frac{1}{I_{k,n}^2(t_{k,n})} \tag{6.4}$$

where, as before, $t_{k,n} = \frac{k}{n+1}$ and

$$I_{k,n}(t) \ = \ nC_{n-1}^{k-1}\, t^{k-1}(1-t)^{n-k}\, I(t).$$

In fact, one can further simplify the expression $I_{k,n}(t_{k,n})$ by using the following elementary argument involving Corollary 6.2. Indeed, let

$$I(t) \ = \ t(1-t)\, J(t), \qquad 0 < t < 1.$$

Then, since $t^k(1-t)^{n-k+1}$ is maximized on $[0,1]$ for $t = t_{k,n}$, we have

$$\begin{aligned}
\sup_{0<t<1} I_{k,n}(t) &= n\,C_{n-1}^{k-1} \sup_{0<t<1} t^k(1-t)^{n-k+1}\,J(t) \\
&\geq n\,C_{n-1}^{k-1}\, t_{k,n}^k(1-t_{k,n})^{n-k+1}\,J(t_{k,n}) \\
&= n\,C_{n-1}^{k-1} \sup_{0<t<1} t^k(1-t)^{n-k+1}\,J(t_{k,n}) \\
&= v\,J(t_{k,n})\,\frac{nC_{n-1}^{k-1}}{(n+2)\,C_{n+1}^k}
\end{aligned}$$

where

$$v \;=\; (n+2)\,C_{n+1}^k \sup_{0<t<1} t^k(1-t)^{n-k+1}.$$

By Corollary 6.2 applied to the couple $(k+1, n+2)$,

$$v^2 \;=\; M_{k+1,n+2}^2 \;\geq\; \frac{1}{12}\;\frac{(n+3)^2\,(n+4)}{(k+1)\,(n-k+2)}\;.$$

In addition,

$$\frac{nC_{n-1}^{k-1}}{(n+2)\,C_{n+1}^k} \;=\; \frac{k(n-k+1)}{(n+1)\,(n+2)}\,.$$

Hence,

$$\begin{aligned}
\sup_{0<t<1} I_{k,n}(t)^2 &\geq \frac{1}{12}\;\frac{(n+3)^2\,(n+4)}{(k+1)\,(n-k+2)}\cdot\left(\frac{k(n-k+1)}{(n+1)\,(n+2)}\right)^2\cdot J(t_{k,n})^2 \\
&= \frac{1}{12}\;\frac{(n+3)^2\,(n+4)}{(n+1)^2\,(n+2)^2}\cdot\frac{k^2(n-k+1)^2}{(k+1)\,(n-k+2)}\cdot\frac{1}{t_{k,n}^2(1-t_{k,n})^2}\cdot I(t_{k,n})^2 \\
&= \frac{1}{12}\;\frac{(n+1)^2\,(n+3)^2\,(n+4)}{(n+2)^2}\cdot\frac{1}{(k+1)\,(n-k+2)}\cdot I(t_{k,n})^2.
\end{aligned}$$

To simplify, use that

$$(k+1)\,(n-k+2) \;\leq\; 4k\,(n-k+1) \;=\; 4(n+1)^2\,t_{k,n}(1-t_{k,n}),$$

so that

$$\sup_{0<t<1} I_{k,n}(t)^2 \;\geq\; \frac{n+4}{48}\cdot\frac{I(t_{k,n})^2}{t_{k,n}(1-t_{k,n})}\,.$$

Hence, by Lemma 6.1,

$$\mathrm{Var}(X_k^*) \;\leq\; \frac{48}{n+4}\;\frac{t_{k,n}(1-t_{k,n})}{I(t_{k,n})^2}\,.$$

Using Lemma 6.3, the argument may be reversed by choosing different absolute constants. Indeed, with the same value of v,

$$\begin{aligned} e^{-2} \sup_{0<t<1} I_{k,n}(t) \ &\leq \ I_{k,n}(t_{k,n}) \\ &= \ nC_{n-1}^{k-1}\, t_{k,n}^k (1-t_{k,n})^{n-k+1}\, J(t_{k,n}) \\ &= \ v\, J(t_{k,n})\, \frac{nC_{n-1}^{k-1}}{(n+2)\, C_{n+1}^k}\,. \end{aligned}$$

Again, by Corollary 6.2,

$$v^2 \ = \ M_{k+1,n+2}^2 \ \leq \ \frac{(n+3)^2\,(n+4)}{(k+1)\,(n-k+2)}\,.$$

Hence, arguing as before,

$$e^{-2} \sup_{0<t<1} I_{k,n}(t)^2 \ \leq \ \frac{(n+1)^2\,(n+3)^2\,(n+4)}{(n+2)^2} \cdot \frac{1}{(k+1)\,(n-k+2)} \cdot I(t_{k,n})^2.$$

To simplify, use that $(k+1)\,(n-k+2) \geq (n+1)^2\, t_{k,n}(1-t_{k,n})$ and $n+3 \leq 2(n+2)$, so that

$$\sup_{0<t<1} I_{k,n}(t)^2 \ \leq \ 4e^2\,(n+4) \cdot \frac{I(t_{k,n})^2}{t_{k,n}(1-t_{k,n})}\,.$$

Hence, by Lemma 6.1,

$$\mathrm{Var}(X_k^*) \ \geq \ \frac{1}{48e^2\,(n+4)}\ \frac{t_{k,n}(1-t_{k,n})}{I(t_{k,n})^2}\,.$$

The next statement summarizes the conclusions of this investigation.

Theorem 6.4 (Two-sided bound on $\mathrm{Var}(X_k^*)$)**.** *Let μ be log-concave with associated I-function I. For some absolute constants $c_0 > 0$ and $c_1 > 0$, for any $k = 1, \dots, n$,*

$$\frac{c_0}{n+1}\ \frac{t_{k,n}(1-t_{k,n})}{I(t_{k,n})^2} \ \leq \ \mathrm{Var}(X_k^*) \ \leq \ \frac{c_1}{n+1}\ \frac{t_{k,n}(1-t_{k,n})}{I(t_{k,n})^2}$$

where $t_{k,n} = \frac{k}{n+1}$.

Here, the factor $\frac{1}{n+1}$ may be viewed as the step $t_{k,n} - t_{k-1,n}$ of the partition of the interval $[0,1]$.

6.2 Two-sided bounds on $\mathbb{E}(W_p^p(\mu_n, \mu))$

On the basis of Theorem 6.4, we are now ready to reach two-sided bounds on $\mathbb{E}\,(W_p^p(\mu_n, \mu))$ for log-concave probability distributions μ on the real line.

As announced, a basic starting point is the general two-sided bound from Theorem 4.3,

$$\frac{2^{-p}}{n} \sum_{k=1}^{n} \mathbb{E}\big(\big|X_k^* - \mathbb{E}(X_k^*)\big|^p\big) \,\le\, \mathbb{E}\big(W_p^p(\mu_n, \mu)\big) \,\le\, \frac{2^p}{n} \sum_{k=1}^{n} \mathbb{E}\big(\big|X_k^* - \mathbb{E}(X_k^*)\big|^p\big).$$

Thus, the issue is how to effectively bound, both from above and below, the variances of order statistics, $\mathrm{Var}(X_k^*)$, or more generally, the quantities $\mathbb{E}((|X_k^* - \mathbb{E}(X_k^*)|^p)$. By virtue of Theorem 6.4, the problem may be solved for any (individual) log-concave measure μ.

First, let us emphasize that, due to the Khinchine-type inequality (B.4) (in dimension one) for log-concave measures, it is sufficient to study the case $p = 2$, only. Namely, for any $p \ge 1$ and $k = 1, \dots, n$, we have

$$c_0 \sqrt{\mathrm{Var}(X_k^*)} \,\le\, \big[\mathbb{E}\big(\big|X_k^* - \mathbb{E}(X_k^*)\big|^p\big)\big]^{1/p} \,\le\, c_1 p \sqrt{\mathrm{Var}(X_k^*)}.$$

Hence, Theorem 6.4 yields the following lemma.

Lemma 6.5. *Let μ be log-concave distribution with the associated I-function I. For any $p \ge 1$ and $k = 1, \dots, n$,*

$$\left(\frac{c_0}{n}\right)^{p/2} \frac{(t_{k,n}(1-t_{k,n}))^{p/2}}{I(t_{k,n})^p} \,\le\, \mathbb{E}\big(\big|X_k^* - \mathbb{E}(X_k^*)\big|^p\big) \,\le\, p^p \left(\frac{c_1}{n}\right)^{p/2} \frac{(t_{k,n}(1-t_{k,n}))^{p/2}}{I(t_{k,n})^p}$$

where $t_{k,n} = \frac{k}{n+1}$, and c_0 and c_1 are positive absolute constants.

Performing summation over all $k = 1, \dots, n$ leads to the two-sided bound

$$\left(\frac{c_0}{n}\right)^{p/2} \Sigma_n \,\le\, \mathbb{E}\big(W_p^p(\mu_n, \mu)\big) \,\le\, p^p \left(\frac{c_1}{n}\right)^{p/2} \Sigma_n$$

where

$$\Sigma_n \,=\, \frac{1}{n} \sum_{k=1}^{n} \frac{(t_{k,n}(1-t_{k,n}))^{p/2}}{I(t_{k,n})^p}\,.$$

The natural step at this stage is to replace this Riemann sum with the corresponding integral (at the expense of constants depending on p, only). This can be done using the concavity of the function I. For points $t_{k,n} \le t \le t_{k+1,n}$, $1 \le k \le n-1$, $n \ge 2$, the concavity implies that

$$I(t) \,\ge\, \min\big\{I(t_{k,n}), I(t_{k+1,n})\big\}.$$

In addition, in the same range,

$$t(1-t) \ \le \ 3 \min\big\{t_{k,n}(1-t_{k,n}),\, t_{k+1,n}(1-t_{k+1,n})\big\}.$$

Let us check the last bound. If $t_{k+1,n} \le \frac{1}{2}$, since the function $t(1-t)$ is increasing in $0 \le t \le \frac{1}{2}$, we obviously have

$$t(1-t) \ \le \ t_{k+1,n}(1-t_{k+1,n}) \ \le \ 2t_{k,n}(1-t_{k,n}).$$

The case $t_{k,n} \ge \frac{1}{2}$ is similar (or symmetric). Now, assume that $t_{k,n} \le \frac{1}{2} \le t_{k+1,n}$, i.e. $\frac{n-1}{2} \le k \le \frac{n+1}{2}$. Then,

$$t_{k,n}(1-t_{k,n}) = \frac{k}{n+1}\bigg(1 - \frac{k}{n+1}\bigg) \ \ge \ \frac{n-1}{n+1}\cdot\frac{1}{4} \ \ge \ \frac{1}{3}\cdot\frac{1}{4} \ \ge \ \frac{1}{3}\, t(1-t).$$

Similarly, $t_{k+1,n}(1-t_{k+1,n}) = \frac{k+1}{n+1}\,(1 - \frac{k+1}{n+1}) \ge \frac{1}{3}\, t(1-t)$. Using these bounds, we obtain that the integral

$$\int_{t_{k,n}}^{t_{k+1,n}} \frac{(t(1-t))^{p/2}}{I(t)^p}\, dt$$

does not exceed

$$3^{p/2}\, (t_{k+1,n} - t_{k,n})\ \frac{(\min\{t_{k,n}(1-t_{k,n}),\, t_{k+1,n}(1-t_{k+1,n})\})^{p/2}}{\max\{I(t_{k,n}), I(t_{k+1,n})\}^p}\,,$$

which in turn is bounded from above by

$$\frac{3^{p/2}}{n+1}\,\bigg[\frac{(t_{k,n}(1-t_{k,n}))^{p/2}}{I(t_{k,n})^p} + \frac{(t_{k+1,n}(1-t_{k+1,n}))^{p/2}}{I(t_{k+1,n})^p}\bigg].$$

Hence, after summation over $k = 1,\dots,n-1$, we arrive at

$$\int_{1/(n+1)}^{n/(n+1)} \frac{(t(1-t))^{p/2}}{I(t)^p}\, dt \ \le \ 2\cdot 3^{p/2}\, \Sigma_n.$$

Now, to derive a similar reverse bound, by the concavity of I (and since $I \ge 0$), whenever $0 < a < b < 1$, we have

$$\frac{I(t)}{t} \ \le \ \frac{I(a)}{a} \qquad \text{for} \ \ a \le t < 1 \qquad \text{and} \qquad \frac{I(t)}{1-t} \ \le \ \frac{I(b)}{1-b} \qquad \text{for} \ \ 0 < t \le b.$$

Hence, multiplying these two inequalities, we get in the interval $a \le t \le b$

$$\frac{t(1-t)}{I(t)^2} \ \ge \ \frac{a(1-b)}{I(a)I(b)}.$$

Also, using $\frac{I(t)}{t} \leq \frac{I(a)}{a}$ at $t = b$, we have $\frac{I(b)}{b} \leq \frac{I(a)}{a}$, so $\frac{a(1-b)}{I(a)I(b)} \geq \frac{a^2(1-b)}{bI(a)^2}$ and thus

$$\frac{t(1-t)}{I(t)^2} \geq \frac{a^2(1-b)}{bI(a)^2}, \qquad a \leq t \leq b.$$

In case $a = t_{k,n}$, $b = t_{k+1,n}$, $1 \leq k \leq n-1$,

$$\frac{a^2(1-b)}{b} = a(1-a)\,\frac{a(1-b)}{b(1-a)} = a(1-a)\,\frac{k(n-k)}{(k+1)(n-k+1)} \geq \frac{1}{4}\,a(1-a),$$

hence,

$$\frac{t(1-t)}{I(t)^2} \geq \frac{1}{4}\,\frac{t_{k,n}(1-t_{k,n})}{I(t_{k,n})^2}, \qquad t_{k,n} \leq t \leq t_{k+1,n}.$$

Similarly (so as to involve the value $k = n$ on the right-hand side), one may apply the previous bound $\frac{I(t)}{1-t} \leq \frac{I(b)}{1-b}$ at $t = a$, that is, $\frac{I(a)}{1-a} \leq \frac{I(b)}{1-b}$. It gives, as before,

$$\frac{t(1-t)}{I(t)^2} \geq \frac{a(1-b)}{I(a)I(b)} \geq \frac{a(1-b)^2}{(1-a)I(b)^2}.$$

Again, in case $a = t_{k,n}$, $b = t_{k+1,n}$, $1 \leq k \leq n-1$,

$$\frac{a(1-b)^2}{1-a} = b(1-b)\,\frac{a(1-b)}{b(1-a)} \geq \frac{1}{4}\,b(1-b),$$

hence,

$$\frac{t(1-t)}{I(t)^2} \geq \frac{1}{4}\,\frac{t_{k+1,n}(1-t_{k+1,n})}{I(t_{k+1,n})^2}, \qquad t_{k,n} \leq t \leq t_{k+1,n},$$

and with the previous bound for the left end point

$$\frac{t(1-t)}{I(t)^2} \geq \frac{1}{4}\,\max\left\{\frac{t_{k,n}(1-t_{k,n})}{I(t_{k,n})^2}, \frac{t_{k+1,n}(1-t_{k+1,n})}{I(t_{k+1,n})^2}\right\}.$$

Now, raising this inequality to the power $p/2$, we get, for all $t_{k,n} \leq t \leq t_{k+1,n}$,

$$\frac{(t(1-t))^{p/2}}{I(t)^p} \geq 2^{-p-1}\left[\frac{(t_{k,n}(1-t_{k,n}))^{p/2}}{I(t_{k,n})^p} + \frac{(t_{k+1,n}(1-t_{k+1,n}))^{p/2}}{I(t_{k+1,n})^p}\right].$$

After integration over $[t_{k,n}, t_{k+1,n}]$ and summation over $k = 1, \dots, n-1$, we arrive at

$$\int_{1/(n+1)}^{n/(n+1)} \frac{(t(1-t))^{p/2}}{I(t)^p}\,dt \geq 2^{-p-1}\,\frac{n}{n+1}\,\Sigma_n.$$

Together with Lemma 6.5 and Theorem 4.3, we may summarize the conclusions.

Theorem 6.6 (Two-sided bounds on $\mathbb{E}(W_p^p(\mu_n,\mu))$). *Let μ be log-concave with the associated I-function I. For any $p \geq 1$ and any $n \geq 2$,*

$$\Big(\frac{c_0}{n}\Big)^{p/2} \int_{1/(n+1)}^{n/(n+1)} \frac{(t(1-t))^{p/2}}{I(t)^p}\,dt \,\leq\, \mathbb{E}\big(W_p^p(\mu_n,\mu)\big) \,\leq\, \Big(\frac{c_1 p^2}{n}\Big)^{p/2} \int_{1/(n+1)}^{n/(n+1)} \frac{(t(1-t))^{p/2}}{I(t)^p}\,dt$$

where c_0 and c_1 are positive absolute constants.

In the basic particular cases $p=1$ and $p=2$, these two-sided bounds become

$$\frac{c_0}{\sqrt{n}} \int_{1/(n+1)}^{n/(n+1)} \frac{\sqrt{t(1-t)}}{I(t)}\,dt \,\leq\, \mathbb{E}\big(W_1(\mu_n,\mu)\big) \,\leq\, \frac{c_1}{\sqrt{n}} \int_{1/(n+1)}^{n/(n+1)} \frac{\sqrt{t(1-t)}}{I(t)}\,dt \tag{6.5}$$

and respectively

$$\frac{c_0}{n} \int_{1/(n+1)}^{n/(n+1)} \frac{t(1-t)}{I(t)^2}\,dt \,\leq\, \mathbb{E}\big(W_2^2(\mu_n,\mu)\big) \,\leq\, \frac{c_1}{n} \int_{1/(n+1)}^{n/(n+1)} \frac{t(1-t)}{I(t)^2}\,dt \tag{6.6}$$

for absolute constants $c_0 > 0$ and $c_1 > 0$.

Note that the case $n = 1$ should be excluded from such inequalities, since the above integrals are then vanishing. According to Lemma 6.5, the bounds of Theorem 6.6 for $n = 1$ need to be changed to

$$\bigg(\frac{c_0}{I(\frac{1}{2})}\bigg)^p \,\leq\, \mathbb{E}\big(W_p^p(\mu_1,\mu)\big) \,\leq\, \bigg(\frac{c_1 p}{I(\frac{1}{2})}\bigg)^p$$

with some positive absolute constants c_0 and c_1. Here $\mu_1 = \delta_{X_1}$ is just a delta measure at the random point X_1 distributed according to μ.

6.3 Khinchine-type inequality

To start with in this paragraph, let us rewrite the two-sided bounds in Theorem 6.6 explicitly in terms of the distribution function and the density of the sample. Changing the variable $t = F(x)$ in the involved integrals, we indeed obtain the following statement.

Theorem 6.7. *Let μ be log-concave distribution with distribution function F and density f. For any $p \geq 1$, and any $n \geq 2$, with some absolute constants $c_0 > 0$ and $c_1 > 0$,*

$$\Big(\frac{c_0}{n}\Big)^{p/2} J_{p,n}(\mu) \,\leq\, \mathbb{E}\big(W_p^p(\mu_n,\mu)\big) \,\leq\, \Big(\frac{c_1 p^2}{n}\Big)^{p/2} J_{p,n}(\mu),$$

where

$$J_{p,n}(\mu) \,=\, \int_{F^{-1}(\frac{1}{n+1})}^{F^{-1}(\frac{n}{n+1})} \frac{(F(x)(1-F(x)))^{p/2}}{f(x)^{p-1}}\,dx. \tag{6.7}$$

In the case $p = 1$,

$$J_{1,n}(\mu) \; = \; \int_{F^{-1}(\frac{1}{n+1})}^{F^{-1}(\frac{n}{n+1})} \sqrt{F(x)(1-F(x))}\, dx,$$

and Theorem 6.7 is therefore telling us that

$$\frac{c_0}{\sqrt{n}}\, J_{1,n}(\mu) \; \leq \; \mathbb{E}\big(W_1(\mu_n,\mu)\big) \leq \frac{c_1}{\sqrt{n}}\, J_{1,n}(\mu).$$

This should be compared to Theorem 3.5. In a slightly modified form (due to Lemma 3.8), the two-sided bounds of Theorem 3.5 may be written as

$$c_0\big(\tilde{A}_n + J_{1,n}(\mu)\big) \; \leq \; \mathbb{E}\big(W_1(\mu_n,\mu)\big) \; \leq \; c_1\big(\tilde{A}_n + J_{1,n}(\mu)\big)$$

where

$$\tilde{A}_n \; = \; \int_0^{F^{-1}(\frac{1}{n+1})} F(x)\big(1-F(x)\big)\, dx \; + \; \int_{F^{-1}(\frac{n}{n+1})}^1 F(x)\big(1-F(x)\big)\, dx.$$

Hence, by Theorem 6.7, when μ is log-concave, the term $\tilde{A}_n$ is dominated by $J_{1,n}(\mu)$ and therefore can be removed from the bounds of Theorem 3.5.

One immediate consequence of Theorem 6.7 is necessary and sufficient conditions for the standard rate, which have already been discussed with completely different tools in the general case (cf. Corollary 5.10).

Corollary 6.8 (Characterization of standard rate for log-concave distribution)**.** *Assume that μ is log-concave distribution with distribution function F and density f. Given $p \geq 1$,*

$$\big[\mathbb{E}\big(W_p^p(\mu_n,\mu)\big)\big]^{1/p} \; \leq \; \frac{C}{\sqrt{n}}$$

with a constant $C > 0$ independent of n, if and only if

$$J_p(\mu) \; = \; \int_0^1 \frac{(F(x)(1-F(x)))^{p/2}}{f(x)^{p-1}}\; dx \; < \; \infty.$$

Although Theorem 6.7 provides two-sided bounds for $\mathbb{E}(W_p^p(\mu_n,\mu)$, one may wonder whether or not it is possible to improve upper-bounds for $\mathbb{E}(W_p(\mu_n,\mu))$. This is in fact not possible, at least in the class of log-concave distributions μ. More precisely, in this case, $\mathbb{E}(W_p^p(\mu_n,\mu))^{1/p}$ and $\mathbb{E}(W_p(\mu_n,\mu))$ must have similar rates.

To this end, we make use of the multidimensional Khinchine-type inequality from (B.4). Given fixed $p \geq 1$, this result may be applied, in particular, to the random vector $X = (X_1^*,\dots,X_n^*,Y_1^*,\dots,Y_n^*)$ in the Euclidean space $\mathbb{R}^{2n}$ equipped with the semi-norm

$$\big\|(x,y)\big\| \; = \; \bigg(\frac{1}{n}\sum_{k=1}^n |x_k - y_k|^p\bigg)^{1/p}, \qquad x = (x_1,\dots,x_n) \in \mathbb{R}^n, \; y = (y_1,\dots,y_n) \in \mathbb{R}^n.$$

As before, $(Y_1^*, \dots, Y_n^*)$ is an independent copy of the vector $(X_1^*, \dots, X_n^*)$ of order statistics, constructed for the sample drawn from a given log-concave probability distribution μ on the real line. If μ_n and ν_n denote the empirical measures for the two vectors, we have (cf. sub-Section 4.2) that

$$\|X\| \, = \, W_p(\mu_n, \nu_n).$$

Hence, by the multidimensional Khinchine-type inequality (B.4) applied to this seminorm

$$\big[\,\mathbb{E}\big(W_p^p(\mu_n, \nu_n)\big)\big]^{1/p} \, \leq \, Cp \, \mathbb{E}\big(W_p(\mu_n, \nu_n)\big).$$

In view of Theorem 4.3, this relation refines Theorem 6.7 in the following form.

Theorem 6.9. *Assume that μ is log-concave distribution with distribution function F and density f. For any $p \geq 1$, and any $n \geq 2$,*

$$\frac{c_0}{p\sqrt{n}}\, J_{p,n}^{1/p}(\mu) \, \leq \, \mathbb{E}\big(W_p(\mu_n, \mu)\big) \, \leq \, \big[\,\mathbb{E}\big(W_p^p(\mu_n, \mu)\big)\big]^{1/p} \, \leq \, \frac{c_1 p}{\sqrt{n}}\, J_{p,n}^{1/p}(\mu)$$

where c_0 and c_1 are positive absolute constants.

In fact, the difference between $\mathbb{E}(W_p(\mu_n, \mu))$ and $\big[\,\mathbb{E}(W_p^p(\mu_n, \mu))\big]^{1/p}$ cannot be large at least for $p \leq 2$ in view of concentration inequalities for random variables $W_p(\mu_n, \mu)$. This interesting property also holds for many other non-logconcave distributions μ, and admits further extensions related to the concentration phenomena for non-product measures on $\mathbb{R}^n$ (not considered here). For example, it is possible to control deviations of $W_p(\mu_n, \mu)$ from its mean if μ satisfies certain integro-differential inequalities of additive type. One family of such inequalities will be discussed in sub-Section 7.1.

6.4 Bounds in terms of the variance

To start with in this sub-section, observe that Theorem 5.5 actually covers the family of all log-concave probability distributions on the line. Indeed, as recalled in Appendix B, sub-Section B.1, for such distributions, h (the Cheeger constant) is equal to $2f(m)$ and $\frac{1}{h^2} \leq 3\,\mathrm{Var}(X_1)$, yielding therefore the following corollary.

Corollary 6.10. *Let μ be log-concave on the line with the standard deviation σ. Then, for any $1 \leq p < 2$ and any $n \geq 1$,*

$$\mathbb{E}\big(W_p^p(\mu_n, \mu)\big) \, \leq \, \frac{C}{2-p}\, \Big(\frac{\sigma}{\sqrt{n}}\Big)^p$$

where $C > 0$ is an absolute constant.

In order to involve the important value $p = 2$ in such a statement, an extra condition on μ is required. As another application of Theorem 5.3 – or more precisely, of Theorem 5.1 (which provides a better constant) – we mention here one such sufficient condition.

Corollary 6.11. *Let μ be log-concave supported on a finite interval $[a,b]$. Then, for all $n \geq 1$,*

$$\mathbb{E}\big(W_2^2(\mu_n,\mu)\big) \ \leq \ \frac{C(b-a)^2}{n+1}$$

where $C > 0$ is an absolute constant (one may take $C = \frac{4}{\log 2}$).

Proof. It is enough to provide a suitable lower-bound on the I-function associated to the measure μ. It is known (cf. [Bob4], Proposition 2.1) that

$$(b-a)\, I(t) \ \geq \ t \log \frac{1}{t} + (1-t) \log \frac{1}{1-t}\,, \qquad 0 < t < 1.$$

Hence,

$$\begin{aligned} J_2(\mu) \ &= \ \int_0^1 \frac{t(1-t)}{I(t)^2}\,dt \ \leq \ 2(b-a)^2 \int_0^{1/2} \frac{t(1-t)}{(t\log\frac{1}{t})^2}\,dt \\ &\leq \ 2(b-a)^2 \int_0^{1/2} \frac{1}{t\log^2\frac{1}{t}}\,dt \ = \ \frac{2(b-a)^2}{\log 2}\,. \end{aligned}$$

The conclusion follows from Theorem 5.1. □

As we will now see, for the range $p > 2$, the compactness of the support is however not sufficient to get a standard rate, i.e. for the relation $\mathbb{E}(W_p(\mu_n,\mu)) = O(\frac{1}{\sqrt{n}})$ to hold. Applying Theorem 6.6, we can extend Corollary 6.10 to the range $p \geq 2$, although with weaker rates. Furthermore, the universal estimate of Corollary 6.10 is no longer true for the critical value $p = 2$, although the rate $\frac{1}{\sqrt{n}}$ does hold for W_2 for the class of log-concave probability distributions with a compact support (Corollary 6.11).

Corollary 6.12. *Let μ be log-concave with standard deviation σ. Then, for any $n \geq 2$,*

$$\mathbb{E}\big(W_2^2(\mu_n,\mu)\big) \ \leq \ \frac{C\sigma^2 \log n}{n}\,,$$

while for $p > 2$,

$$\mathbb{E}\big(W_p^p(\mu_n,\mu)\big) \ \leq \ \frac{C_p\, \sigma^p}{n}$$

where $C_p > 0$ depends on p, only.

Proof. One may use the argument described after Theorem 5.5. For any $p \geq 1$, in terms of the Cheeger constant $h = h(\mu)$, it yields a comparison bound

$$\begin{aligned} \mathbb{E}\big(W_p^p(\mu_n, \mu)\big) \ &\leq \ \frac{1}{n} \sum_{k=1}^n \mathbb{E}\big(|X_k^* - Y_k^*|^p\big) \\ &\leq \ \frac{1}{nh^p} \sum_{k=1}^n \mathbb{E}\big(|U_k^* - V_k^*|^p\big) \\ &\leq \ \frac{2^p}{h^p}\, \mathbb{E}\big(W_p^p(\nu_n, \nu)\big) \end{aligned}$$

where $(U_1^*, \dots, U_n^*)$ and $(V_1^*, \dots, V_n^*)$ are order statistics for two independent samples $(U_1, \dots, U_n)$ and $(V_1, \dots, V_n)$ taken from the two-sided exponential distribution ν on the real line with density $\frac{1}{2}\, e^{-|x|}$, and where ν_n is the empirical measure on $(U_1, \dots, U_n)$. Hence, keeping h to be fixed, we are reduced to the case of the measure ν.

Now, since $I_\nu(t) = \min\{t, 1-t\}$, Theorem 6.6 with $p \geq 2$ gives, for any $n \geq 2$,

$$\begin{aligned} \mathbb{E}\big(W_p^p(\nu_n, \nu)\big) \ &\leq \ \Big(\frac{c_1 p^2}{n}\Big)^{p/2} \int_{1/(n+1)}^{n/(n+1)} \frac{(t(1-t))^{p/2}}{I_\nu(t)^p}\, dt \\ &\leq \ 2\,\Big(\frac{c_1 p^2}{n}\Big)^{p/2} \int_{1/(n+1)}^{1/2} t^{-p/2}\, dt \\ &\leq \ \frac{(c_1' p)^p}{p-2}\, \frac{1}{n}\,, \end{aligned}$$

where $c_1, c_1' > 0$ are absolute constants and where we assume that $p > 2$ on the last step. When $p = 2$ the last integral grows like $\log n$. To finish the proof, it remains to use an upper-bound $\frac{1}{h} \leq \sigma\sqrt{3}$ (which was already mentioned before Corollary 6.10). $\square$

Both estimates of Corollary 6.12 are sharp with respect n. Indeed, by the lower-bound of Theorem 6.6, for the two-sided exponential distribution ν we have with some absolute constant $c_0 > 0$

$$\begin{aligned} \mathbb{E}\big(W_p^p(\nu_n, \nu)\big) \ &\geq \ \Big(\frac{c_0}{n}\Big)^{p/2} \int_{1/(n+1)}^{n/(n+1)} \frac{(t(1-t))^{p/2}}{I_\nu(t)^p}\, dt \\ &\geq \ \Big(\frac{c_0}{2n}\Big)^{p/2} \int_{1/(n+1)}^{1/2} t^{-p/2}\, dt. \end{aligned}$$

If $p > 2$, the last expression decays like $\frac{1}{n}$, while for $p = 2$ it decays like $\frac{\log n}{n}$.

Moreover, by Theorem 6.9,

$$\mathbb{E}\big(W_2(\nu_n, \nu)\big) \ \geq \ c\,\sqrt{\frac{\log n}{n}}$$

and, for $p > 2$,

$$\mathbb{E}\big(W_p(\nu_n,\nu)\big) \ \geq \ \frac{c_p}{n^{1/p}}$$

where $c > 0$ is absolute and $c_p > 0$ depends on p, only. Hence, the rates in Corollary 6.12 cannot be improved for the potentially smaller quantity $\mathbb{E}(W_p(\mu_n,\mu))^p$.

Remark 6.13. Without appealing to the lower-bound of Theorem 6.6, the sharpness of the estimates in Corollary 6.12 can be verified directly on the example of the one-sided exponential distribution ν with density e^{-x} $(x > 0)$. Indeed, in this case (cf. e.g. [Ga]) the k-th order statistic U_k^* is equidistributed with the random variable

$$T_k \ = \ \sum_{j=1}^{k} \frac{U_j}{n-j+1}$$

where $(U_1,\dots,U_n)$ is the sample taken from ν. Hence,

$$\begin{aligned}\sum_{k=1}^{n} \mathrm{Var}(U_k^*) \ &= \ \sum_{k=1}^{n}\sum_{j=1}^{k} \frac{1}{(n-j+1)^2} \\ &= \ \sum_{j=1}^{n} \frac{1}{n-j+1} \ = \ 1 + \frac{1}{2} + \dots + \frac{1}{n} \ \sim \ \log n,\end{aligned}$$

so that, by Theorem 4.3, $\mathbb{E}(W_2^2(\nu_n,\nu))$ has the rate $\frac{\log n}{n}$. For $p > 2$, simply use

$$\mathbb{E}\big(\big|U_k^* - \mathbb{E}(U_k^*)\big|^p\big) \ \geq \ \mathrm{Var}(U_k^*)^{p/2}.$$

Since

$$\mathrm{Var}(U_n^*) \ = \ \sum_{j=1}^{n} \frac{1}{(n-j+1)^2} \geq 1,$$

again by Theorem 4.3,

$$\begin{aligned}\mathbb{E}\big(W_p^p(\nu_n,\nu)\big) \ &\geq \ \frac{1}{2^p\,n} \sum_{k=1}^{n} \mathbb{E}\big(\big|U_k^* - \mathbb{E}(U_k^*)\big|^p\big) \\ &\geq \ \frac{1}{2^p\,n}\,\mathbb{E}\big(\big|U_n^* - \mathbb{E}(U_n^*)\big|^p\big) \\ &\geq \ \frac{1}{2^p\,n}\,.\end{aligned}$$

6.5 Some other log-concave examples

For a log-concave probability distribution μ on the real line, Theorem 6.6 may provide a variety of possible rates for $\mathbb{E}(W_p(\mu_n,\mu))$, especially for the range $p > 2$. More precisely, the rate may vary from $\frac{1}{\sqrt{n}}$ to $\frac{1}{n^{1/p}}$ when $p > 2$, while Corollary 6.10 guarantees the standard rate $\frac{1}{\sqrt{n}}$, when $p < 2$.

In the previous section, we considered the example of the exponential distributions (both one-sided and two-sided). Here are some further specific examples.

Corollary 6.14 (Two-sided bounds for Gaussian measure)**.** *Let $n \geq 3$. If μ is the standard Gaussian measure,*

$$\frac{c \log\log n}{n} \leq \big[\mathbb{E}\big(W_2(\mu_n,\mu)\big)\big]^2 \leq \mathbb{E}\big(W_2^2(\mu_n,\mu)\big) \leq \frac{C \log\log n}{n}$$

where $c, C > 0$ are numerical. For $p > 2$,

$$\frac{c_p}{n\,(\log n)^{p/2}} \leq \big[\mathbb{E}\big(W_p(\mu_n,\mu)\big)\big]^p \leq \mathbb{E}\big(W_p^p(\mu_n,\mu)\big) \leq \frac{C_p}{n\,(\log n)^{p/2}}$$

where the involved constants $c_p, C_p > 0$ depend on p, only.

Indeed, in the Gaussian case the associated I-function is symmetric about the point $t = \frac{1}{2}$ and satisfies, for $0 < t \leq \frac{1}{2}$,

$$c_0 t\sqrt{\log(1/t)} \leq I(t) \leq c_1 t\sqrt{\log(1/t)}$$

with some positive constants c_0 and c_1. Hence, by Theorem 6.6, within p-dependent factors, the value $\mathbb{E}(W_p^p(\mu_n,\mu))$ is described by

$$\frac{1}{n^{p/2}}\int_{1/(n+1)}^{1/2} \frac{(t(1-t))^{p/2}}{I(t)^p}\,dt \sim \frac{1}{n^{p/2}}\int_{1/(n+1)}^{1/2} \frac{dt}{(t\,\log(1/t))^{p/2}}\,.$$

When $p = 2$ the last integral grows like $\log\log n$. If $p > 2$, the latter integral (up to the factor $\frac{1}{\frac{p}{2}-1}$) may be integrated by parts to obtain

$$\begin{aligned} -\int_{1/(n+1)}^{1/2} \frac{1}{(\log(1/t))^{p/2}}\,d(1/t)^{\frac{p}{2}-1} \;=\;& \frac{(n+1)^{\frac{p}{2}-1}}{(\log(n+1))^{p/2}} - \frac{2^{\frac{p}{2}-1}}{(\log 2)^{p/2}} \\ &+ \frac{p}{2}\int_{1/(n+1)}^{1/2} \frac{dt}{(t\,\log(1/t))^{p/2}\,\log(1/t)}\,. \end{aligned}$$

Therefore, this integral grows like the first term, i.e. $n^{\frac{p}{2}-1}(\log n)^{-p/2}$.

As a further example, consider the family of the beta distributions μ with parameters $\alpha \geq 1$ and $\beta = 1$. Any such measure is supported on the unit interval $[0,1]$, where it has the density and distribution function

$$f_\alpha(x) = \alpha\, x^{\alpha-1}, \qquad F_\alpha(x) = x^\alpha, \qquad 0 \leq x \leq 1.$$

Hence, the associated I-function is given by

$$I_\alpha(t) \,=\, f_\alpha\big(F_\alpha^{-1}(t)\big) \,=\, \alpha\, t^{(\alpha-1)/\alpha}, \qquad 0<t<1.$$

For example, the value $\alpha = 1$ corresponds to the uniform distribution on $[0,1]$.

If $\alpha \le 2$, then $I_\alpha(t) \ge \alpha\sqrt{t(1-t)}$, and according to Corollary 5.4, we obtain a standard rate

$$\big[\mathbb{E}\big(W_p^p(\mu_n,\mu)\big)\big]^{1/p} \,\le\, \frac{Cp}{\sqrt{n}}$$

where $p \ge 1$ is arbitrary, and $C > 0$ is an absolute constant.

If $\alpha > 2$, the same conclusion remains to hold for the range $1 \le p \le 2$ (recall Corollary 6.11 concerning general compactly supported log-concave measures).

Let now $\alpha > 2$ and $p > 2$. Again, by Theorem 6.6, within (α,p)-dependent factors the value $\mathbb{E}(W_p^p(\mu_n,\mu))$ is given by

$$\frac{1}{n^{p/2}} \int_{1/(n+1)}^{n/(n+1)} \frac{(t(1-t))^{p/2}}{I(t)^p}\,dt \,\sim\, \frac{1}{n^{p/2}} \int_{1/(n+1)}^{1} \frac{dt}{t^{p(\frac{1}{2}-\frac{1}{\alpha})}}\,.$$

Here the last integral is bounded whenever $p < \frac{2\alpha}{\alpha-2}$. It grows like $\log n$ for $p = \frac{2\alpha}{\alpha-2}$, and grows like n^κ with $\kappa = p(\frac{1}{2}-\frac{1}{\alpha}) - 1$ when $p > \frac{2\alpha}{\alpha-2}$. Thus, in the latter case,

$$\frac{1}{n^{p/2}} \int_{1/(n+1)}^{1} \frac{dt}{t^{p(\frac{1}{2}-\frac{1}{\alpha})}} \,\sim\, \frac{1}{n^{\frac{p}{\alpha}+1}}\,.$$

As a conclusion, we get:

Corollary 6.15 (Two-sided bounds for beta distributions)**.** *Let μ be the beta distribution with parameters $\alpha \ge 1$ and $\beta = 1$. If $1 \le \alpha \le 2$, $p \ge 1$, or if $\alpha > 2$, $1 \le p < \frac{2\alpha}{\alpha-2}$, then, for $n \ge 1$,*

$$\frac{c_0}{\sqrt{n}} \,\le\, \big[\mathbb{E}\big(W_p^p(\mu_n,\mu)\big)\big]^{1/p} \,\le\, \frac{c_1}{\sqrt{n}}\,,$$

while for $\alpha > 2$ and $p > \frac{2\alpha}{\alpha-2}$,

$$\frac{c_0}{n^{\frac{1}{\alpha}+\frac{1}{p}}} \,\le\, \big[\mathbb{E}\big(W_p^p(\mu_n,\mu)\big)\big]^{1/p} \,\le\, \frac{c_1}{n^{\frac{1}{\alpha}+\frac{1}{p}}}\,.$$

Here the involved constants c_0 and c_1 are positive and may depend on α and p, only. In the critical case $p = \frac{2\alpha}{\alpha-2}$ with $\alpha > 2$, up to p-dependent constants, we also have

$$\frac{c_0\,(\log(n+1))^{1/p}}{\sqrt{n}} \,\le\, \big[\mathbb{E}\big(W_p^p(\mu_n,\mu)\big)\big]^{1/p} \,\le\, \frac{c_1\,(\log(n+1))^{1/p}}{\sqrt{n}}\,.$$

A similar conclusion may also be made about the beta distributions with parameters $\alpha = \beta \ge 1$ (in which case the distributions are symmetric about the point $x = 1/2$).

Thus, when p is large, the rate can be rather weak, even for compactly supported measures.

7 Miscellaneous bounds and results

This section collects a number of results both supplementing the previous investigations and of independent interest. The first paragraph essentially shows that for classes of distributions μ satisfying a Poincaré-type inequality, the behaviours of $\mathbb{E}(W_p(\mu_n, \mu))$ and $\big[\mathbb{E}(W_p^p(\mu_n, \mu))\big]^{1/p}$ are of the same order. Next, we describe upper-bounds in terms of modulus of continuity of the inverse distribution function F^{-1}. Two-sided bounds for classes of compactly supported distributions are examined in the further paragraph, with in particular a detailed study of connectness (of the support of μ) and absolute continuity (of F^{-1}). While moments cannot achieve the standard rate in general, nevertheless the use of alternative tools such as the Zolotarev ideal metrics allow for the Kantorovich distance W_2. The last sub-section completes the study of convergence in W_∞.

The setting is as in the preceding sections, with a probability μ on $\mathbb{R}$ with distribution function F and the associated empirical measures $\mu_n = \frac{1}{n}\sum_{k=1}^n \delta_{X_k}$, $n \geq 1$, on a sample $(X_k)_{k\geq 1}$ of independent random variables with common law μ.

7.1 Deviations of $W_p(\mu_n, \mu)$ from the mean

This paragraph investigates, in a general setting, the possible deviations of the random variables $W_p(\mu_n, \mu)$ from their expected values. We will pose the hypothesis on the spectral gap of the underlying distribution μ.

Refereing to [B-H2], [L2], [L3], [B-G-L]..., a Borel probability measure P on $\mathbb{R}^n$ is said to satisfy a Poincaré-type inequality with constant $\lambda > 0$, if for any bounded smooth function u on $\mathbb{R}^n$,

$$\lambda \operatorname{Var}_P(u) \leq \int_{\mathbb{R}^n} |\nabla u|^2 \, dP. \tag{7.1}$$

On the real line, a full characterization of such probability measures is well-known. In particular, they have an interval as a support (finite or not), and a finite exponential moment (cf. the preceding references). In particular, for a log-concave probability distribution μ, there is a lower-bound

$$\lambda \geq \frac{1}{12\sigma^2},$$

where σ is the standard deviation of μ (cf. [Bob4]). If μ on the real line satisfies (7.1) with constant $\lambda > 0$, so does $P = \mu^{\otimes n}$ on $\mathbb{R}^n$ with the same constant.

The following theorem is essentially contained in the works of Gozlan and Léonard (see [Go], [G-L1], [G-L2]).

Theorem 7.1 (Concentration of $W_p(\mu_n, \mu)$)**.** *Assume that μ satisfies a Poincaré-type inequality on the real line with constant $\lambda > 0$. Then, for any $p \geq 1$ and $r > 0$, and any $n \geq 1$,*

$$\mathbb{P}\Big\{ \big|W_p(\mu_n, \mu) - \mathbb{E}\big(W_p(\mu_n, \mu)\big)\big| \geq r \Big\} \leq C \exp\big\{ -2n^{1/\max(p,2)}\sqrt{\lambda}\, r\big\},$$

where $C > 0$ is an absolute constant.

The proof of Theorem 7.1 is based on two elementary and known lemmas.

Lemma 7.2. *The map $T_n : \mathbb{R}^n \to Z_p(\mathbb{R})$ assigning to each point $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ the "empirical" measure*

$$T_n(x) \,=\, \frac{1}{n} \sum_{k=1}^n \delta_{x_k}$$

has Lipschitz seminorm $\|T_n\|_{\rm Lip} = n^{-1/\max(p,2)}$ with respect to the Euclidean metric and the metric W_p $(p \geq 1)$.

Proof. Consider the map which assigns to each point $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ the vector $x^* = (x_1^*, \dots, x_n^*) \in \mathbb{R}^n$ whose components are the same as of x, but are arranged in increasing order, $x_1^* \leq \cdots \leq x_n^*$. By Lemma 4.2, this map is Lipschitz with respect to any ℓ^p-norm $\|\cdot\|_p$ $(p \geq 1)$ in the sense that for all $x, y \in \mathbb{R}^n$,

$$\|x^* - y^*\|_p^p \,=\, \sum_{k=1}^n |x_k^* - y_k^*|^p \,\leq\, \|x - y\|_p^p \,=\, \sum_{k=1}^n |x_k - y_k|^p.$$

However, it will be more convenient to relate the latter distance to the Euclidean norm. If $1 \leq p \leq 2$, Hölder's inequality yields $\|x\|_p \leq n^{(2-p)/2p}\, \|x\|_2$, while for $p \geq 2$ we have $\|x\|_p \leq \|x\|_2$. Hence,

$$\|x^* - y^*\|_p \,\leq\, \begin{cases} n^{(2-p)/2p}\, \|x - y\|_2, & \text{if} \quad 1 \leq p \leq 2, \\ \|x - y\|_2, & \text{if} \quad p > 2, \end{cases}$$

where both inequalities are sharp.

Now, recalling Lemma 4.2, for $x, y \in \mathbb{R}^n$,

$$W_p^p\big(T_n(x), T_n(y)\big) \,=\, \frac{1}{n}\, \|x^* - y^*\|_p^p \,\leq\, \frac{1}{n}\, \|x - y\|_p^p.$$

Hence, if $\mathbb{R}^n$ is equipped with the Euclidean distance, the Lipschitz semi-norm of the map T_n equals $n^{-1/p}$ for $p \geq 2$, but is equal to $\frac{n^{(2-p)/2p}}{n^{1/p}} = \frac{1}{n^{1/2}}$, for $1 \leq p \leq 2$. This completes the proof of the lemma. $\square$

Lemma 7.3. *Let P be a Borel probability measure on $\mathbb{R}^n$ with concentration function*

$$\alpha_P(r) \,=\, \sup\Big\{1 - P(A^r) : P(A) \geq \frac{1}{2}\Big\}, \qquad r > 0.$$

Then, the concentration function α_Q for the image $Q = PT_n^{-1}$ of P under the map T_n satisfies, for all $r > 0$,

$$\alpha_Q(r) \,\leq\, \alpha_P\big(n^{1/\max(p,2)}\, r\big).$$

Here, A^r denotes an open r-neighbourhood of a set $A \subset \mathbb{R}^n$ with respect to the Euclidean distance, while the definition of the concentration function α_Q is similar and refers to the metric W_p on $Z_p(\mathbb{R})$. Note that Q is supported on a relatively "small" subset of $Z_p(\mathbb{R})$, namely, the collection of all discrete probability measures on the real line having at most n atoms.

The statement of Lemma 7.3 immediately follows from Lemma 7.2. Indeed, for any set $B \subset Z_p(\mathbb{R})$, there is a general set inclusion

$$\big[T_n^{-1}(B)\big]^{Lr} \subset T_n^{-1}(B^r), \qquad r > 0,$$

where L denotes the Lipschitz seminorm of T_n (explicitly described in Lemma 7.2). Hence, if $Q(B) = P(T_n^{-1}(B)) \geq 1/2$, we have

$$1 - Q(B^r) \,\leq\, 1 - P\left(\left(T_n^{-1}(B)\right)^{Lr}\right) \,\leq\, \alpha_P(Lr).$$

It remains to take the supremum over all B such that $Q(B) \geq 1/2$.

Proof of Theorem 7.1. For any Borel probability measure P on $\mathbb{R}^n$ admitting a Poincaré-type inequality with constant $\lambda > 0$, it is classical that its concentration function satisfies

$$\alpha_P(r) \,\leq\, Ce^{-2\sqrt{\lambda}r}, \qquad r > 0,$$

This general observation is due to Gromov and Milman [G-M] (and Borovkov and Utev [B-U] for dimension $n = 1$). See also later [A-S], [L2], [L3]. The fact that the constant 2 in the exponent is optimal was emphasized in [Bob3].

By the assumption of the theorem, this result may be applied to the product measure $P = \mu^n$. Hence, by Lemma 7.3,

$$\alpha_Q(r) \,\leq\, C\exp\big\{-2\,n^{1/\max(p,2)}\sqrt{\lambda}\,r\big\}, \qquad r > 0.$$

Equivalently (up to an absolute constant $C > 0$), for any function $u : Z_p(\mathbb{R}) \to \mathbb{R}$ with Lipschitz seminorm $\|u\|_{\mathrm{Lip}} \leq 1$ with respect to the metric W_p, for every $r > 0$,

$$Q\bigg\{\bigg|u - \int_{Z_p(\mathbb{R})} u\,dQ\bigg| \geq r\bigg\} \,\leq\, C\exp\big\{-2\,n^{1/\max(p,2)}\sqrt{\lambda}\,r\big\}.$$

In particular, one may apply this inequality to the distance function $u(\nu) = W_p(\nu, \mu)$, and then we arrive at the desired conclusion. Theorem 7.1 is established. $\square$

Corollary 7.4. *Under the assumptions of Theorem 7.1,*

$$\big[\,\mathbb{E}\big(W_p^p(\mu_n, \mu)\big)\big]^{1/p} \,\leq\, \mathbb{E}\left(W_p(\mu_n, \mu)\right) + \frac{Cp}{n^{1/\max(p,2)}\sqrt{\lambda}}$$

where $C > 0$ is an absolute constant.

Indeed, for the random variable $\xi = 2n^{1/\max(p,2)}\sqrt{\lambda}\, W_p(\mu_n,\mu)$, the bound of Theorem 7.1 may be rewritten as

$$\mathbb{P}\big\{\big|\xi - \mathbb{E}(\xi)\big| \geq r\big\} \, \leq \, Ce^{-r}, \qquad r > 0,$$

which implies $\mathbb{E}(|\xi - \mathbb{E}\xi|^p) \leq C\Gamma(p+1)$ and so $\big[\mathbb{E}(\xi^p)\big]^{1/p} \leq \mathbb{E}(\xi) + C^{1/p}\, p$.

For the range $1 \leq p \leq 2$, the inequality of Corollary 7.4 takes the form

$$\big[\,\mathbb{E}\big(W_p^p(\mu_n,\mu)\big)\big]^{1/p} \, \leq \, \mathbb{E}\big(W_p(\mu_n,\mu)\big) + \frac{C}{\sqrt{\lambda n}}$$

with some absolute constant $C > 0$. On the other hand, by Theorem 3.1,

$$\mathbb{E}\big(W_p(\mu_n,\mu)\big) \, \geq \, \mathbb{E}\big(W_1(\mu_n,\mu)\big) \, \geq \, \frac{1}{\sqrt{8n}}\, \mathbb{E}\big(|X-m|\big),$$

where m is a median of X_1. Combining the two inequalities yields the following conclusion. While it appears as a certain extension of Theorem 6.9 in the log-concave case, the constant involved in the result depends on the distribution μ.

Corollary 7.5 (Moment equivalence for $W_p(\mu_n,\mu)$, $1 \leq p \leq 2$)**.** *Under the assumptions of Theorem 7.1, for any $1 \leq p \leq 2$,*

$$\big[\,\mathbb{E}\big(W_p^p(\mu_n,\mu)\big)\big]^{1/p} \, \leq \, C\,\mathbb{E}\big(W_p(\mu_n,\mu)\big)$$

where $C > 0$ is a constant depending on the product $\sqrt{\lambda}\, \mathbb{E}(|X_1 - m|)$.

A similar conclusion may also be made on the basis of Corollary 7.4 for $p > 2$, provided that

$$\lim_{n\to\infty} n\,\mathbb{E}\big(W_p^p(\mu_n,\mu)\big) \, = \, \infty.$$

That is, in this case $\big[\mathbb{E}(W_p^p(\mu_n,\mu))\big]^{1/p}$ and $\mathbb{E}(W_p(\mu_n,\mu))$ have similar rates.

Remark 7.6. In Theorem 7.1, sharper deviation inequalities may be obtained under stronger hypotheses on μ (using Lemma 7.3 or its functional form). For example, starting from a logarithmic Sobolev inequality

$$\rho\bigg[\int_{\mathbb{R}} u^2 \log^2\, d\mu - \int_{\mathbb{R}} u^2\, d\mu\, \log \int_{\mathbb{R}} u^2\, d\mu\bigg] \, \leq \, 2\int_{\mathbb{R}} u'(x)^2\, d\mu(x)$$

with a constant $\rho > 0$, we obtain that, for any $p \geq 1$ and $r > 0$,

$$\mathbb{P}\Big\{\, \big|W_p(\mu_n,\mu) - \mathbb{E}\big(W_p(\mu_n,\mu)\big)\big| \geq r\Big\} \, \leq \, 2\exp\big\{-\rho\, n^{2/\max(p,2)} r^2/2\big\}.$$

In particular, for $1 \leq p \leq 2$, the right-hand side is independent of p, since then

$$\mathbb{P}\Big\{\, \big|W_p(\mu_n,\mu) - \mathbb{E}\big(W_p(\mu_n,\mu)\big)\big| \geq r\Big\} \, \leq \, 2\exp\big\{-\rho\, nr^2/2\big\}.$$

We refer to [L2], [L3] for more on logarithmic Sobolev inequalities and their applications to concentration inqualities.

7.2 Upper-bounds in terms of modulus of continuity

Let μ be a probability measure on $\mathbb{R}$ with distribution function F, and let μ_n, $n \geq 1$, be the associated empirical measures based on a sample from μ. In addition to the functional J_p from (5.1) which is responsible for the standard rate, another general upper-bound on $\mathbb{E}(W_p^p(\mu_n, \mu))$ can be obtained by involving the modulus of continuity of the inverse function,

$$\delta_{F^{-1}}(\varepsilon) \,=\, \sup\big\{\big|F^{-1}(t) - F^{-1}(s)\big| : |t-s| \leq \varepsilon, \ \ t,s \in (0,1)\big\}, \qquad 0 < \varepsilon \leq 1,$$

or equivalently by involving the modulus of increase of F.

We refer to Appendix A, sub-Section A.3, for an account on this modulus of continuity of the inverse function. In particular, recall that according to Proposition A.12 there, for the property $\delta_{F^{-1}}(\varepsilon) < \infty$, μ has to be compactly supported, and moreover, for $\delta_{F^{-1}}(0+) = 0$ we need to assume that the function F^{-1} is continuous. The latter means that the support of μ should be a finite closed interval.

Theorem 7.7 (Upper-bounds in terms of modulus of continuity)**.** *Suppose that the support of μ is a finite closed interval. Then, for all $p \geq 1$, and all $n \geq 1$,*

$$\big[\,\mathbb{E}\big(W_p^p(\mu_n, \mu)\big)\big]^{1/p} \,\leq\, C\,\sqrt{p}\;\delta_{F^{-1}}\Big(\frac{1}{\sqrt{n}}\Big)$$

with some numerical constant $C > 0$.

Proof. Put $\delta(\varepsilon) = \delta_{F^{-1}}(\varepsilon)$. We have $\mathbb{E}(W_p^p(\mu_n, \mu)) \leq \mathbb{E}(W_p^p(\mu_n, \nu_n))$, where ν_n is an independent copy of μ. To bound the last expectation, using Theorem 4.6, write

$$\begin{aligned} \mathbb{E}\big(W_p^p(\mu_n, \nu_n)\big) \ &= \ \int_0^1\!\!\int_0^1 \big|F^{-1}(t) - F^{-1}(s)\big|^p \, dB_n(t,s) \\ &\leq \ \int_0^1\!\!\int_0^1 \delta\big(|t-s|\big)^p \, dB_n(t,s), \end{aligned}$$

where B_n is the mean square beta distribution of order n (cf. Appendix B, sub-Section B.6). If (X,Y) is a random vector, distributed according to B_n, and $\xi = |X - Y|$, we thus have

$$\mathbb{E}\big(W_p^p(\mu_n, \nu_n)\big) \,\leq\, \mathbb{E}\big(\delta(\xi)^p\big).$$

Denote by R the distribution function of ξ and write

$$\begin{aligned} \mathbb{E}\big(\delta^p(\xi)\big) \ &= \ \int_0^1 \delta^p(\varepsilon)\, dR(\varepsilon) \\ &\leq \ \delta^p(1/\sqrt{n})\ R(1/\sqrt{n}) + \int_{1/\sqrt{n}}^1 \delta^p(\varepsilon)\, dR(\varepsilon) \\ &\leq \ \delta^p(1/\sqrt{n}) + \int_1^{\sqrt{n}} \delta^p(x/\sqrt{n})\, dR(x/\sqrt{n}), \end{aligned}$$

where we made the substitution $\varepsilon = x/\sqrt{n}$.

As any other modulus of continuity, δ is subadditive in the sense that whenever $\varepsilon_i \geq 0$, $\sum_{i=1}^N \varepsilon_i \leq 1$, then

$$\delta\Big(\sum_{i=1}^N \varepsilon_i\Big) \,\leq\, \sum_{i=1}^N \delta(\varepsilon_i).$$

In particular, $\delta(N\varepsilon) \leq N\delta(\varepsilon)$, and therefore $\delta(x\varepsilon) \leq ([x]+1)\,\delta(\varepsilon) \leq 2x\delta(\varepsilon)$ for any real $x \geq 1$ ($x\varepsilon \leq 1$). Hence, using this bound in the latter integral, we get

$$\mathbb{E}\big(\delta^p(\xi)\big) \,\leq\, \delta^p(1/\sqrt{n}) + 2^p\,\delta^p(1/\sqrt{n}) \int_1^{\sqrt{n}} x^p\,dR(x/\sqrt{n}).$$

Now, by the subgaussian bound of Proposition B.12, for all $\varepsilon \geq 0$,

$$1 - R(\varepsilon) \,=\, \mathbb{P}\{\xi \geq \varepsilon\} \,\leq\, 2e^{-n\varepsilon^2/16}.$$

Integrating by parts, we then obtain that

$$\begin{aligned}
\int_1^{\sqrt{n}} x^p\,dR(x/\sqrt{n}) \ &= \ \big(1 - R(1/\sqrt{n})\big) + \int_1^{\sqrt{n}} \big(1 - R(x/\sqrt{n})\big)\,dx^p \\
&\leq \ 1 + 2\int_0^\infty e^{-x^2/16}\,dx^p \\
&\leq \ \big(C\sqrt{p}\,\big)^p
\end{aligned}$$

with some constant $C > 0$. The proof is complete. □

Theorem 7.7 includes Theorem 4.9 in the case of the uniform distribution which we considered with different tools. Here is a somewhat general situation (cf. Appendix A, Example A.14).

Corollary 7.8. *If μ is unimodal, symmetric about the point $\frac{1}{2}$, with support $[0,1]$, then for all $p \geq 1$,*

$$\big[\,\mathbb{E}\big(W_p^p(\mu_n,\mu)\big)\big]^{1/p} \,\leq\, C\,\sqrt{p}\ F^{-1}\Big(\frac{1}{\sqrt{n}}\Big)$$

with some numerical constant $C > 0$.

As an example of a different type, let μ be a discrete probability measure on $(0,1)$, for which F^{-1} represents the Cantor stairs. As discussed in Section A.3 from Appendix A, in this case $\delta_{F^{-1}}(\varepsilon) \leq (4\varepsilon)^{\frac{\log 2}{\log 3}}$. Hence,

$$\big[\,\mathbb{E}\big(W_p^p(\mu_n,\mu)\big)\big]^{1/p} \,\leq\, \frac{C\,\sqrt{p}}{n^{\alpha/2}}$$

where $\alpha = \frac{\log 2}{\log 3} = 0.6309\ldots$. When $p < \frac{\log 3}{\log 2}$, it asymptotically improves upon the general bound $(\mathbb{E}\,W_p^p(\mu_n,\mu))^{1/p} \leq n^{-1/2p}$ for probability distributions on $[0,1]$ (to which we return in the next section).

7.3 Two-sided bounds of order $n^{-1/(2p)}$

Consider the situation when μ is supported on a finite interval, say $[0,1]$. In this case,

$$J_1(\mu) \,=\, \int_0^1 \sqrt{F(x)(1-F(x))}\,dx \,\leq\, \frac{1}{2}$$

so that, by Theorem 3.2, for every $n \geq 1$,

$$\mathbb{E}\big(W_1(\mu_n,\mu)\big) \,\leq\, \frac{1}{2\sqrt{n}}$$

which provides the best possible rate. Actually, this bound allows one to control the transport distances $W_p(\mu_n,\mu)$ for $p>1$. Indeed, by the very definition for the setting of an abstract metric space (E,d) with finite diameter $D = \mathrm{diam}(E)$, for all Borel probability measures ν_1 and ν_2,

$$W_p^p(\nu_1,\nu_2) \,\leq\, D^{p-1}\, W_1(\nu_1,\nu_2)$$

for any $p \geq 1$. This simple relation leads to a universal upper-bound for the means of $W_p^p(\mu_n,\mu)$.

Theorem 7.9. *For any probability measure μ supported on $[0,1]$, and any $p \geq 1$ and $n \geq 1$,*

$$\mathbb{E}\big(W_p^p(\mu_n,\mu)\big) \,\leq\, \frac{1}{2\sqrt{n}}\,.$$

In particular, $\mathbb{E}(W_p(\mu_n,\mu)) \leq n^{-1/(2p)}$.

In fact, both estimates cannot be improved asymptotically with respect to n uniformly in the whole class of probability measures on $[0,1]$ as shown by the following example.

Example 7.10. Let $\mu = \frac{1}{2}\,\delta_1 + \frac{1}{2}\,\delta_0$ be the symmetric Bernoulli measure on $\{0,1\}$. In this case

$$\mu_n \,=\, a\delta_1 + (1-a)\delta_0, \qquad a = \frac{S_n}{n}\,,$$

where $S_n = X_1 + \cdots + X_n$ is the number of successes in n independent trials with probability $\frac{1}{2}$ of success in each trial. By Example 2.2, for any $p \geq 1$,

$$W_p^p(\mu_n,\mu) \,=\, W_1(\mu_n,\mu) \,=\, \bigg|\frac{S_n}{n} - \frac{1}{2}\bigg|.$$

Therefore, in terms of the independent random variables $\varepsilon_k = 2X_k - 1$, $k = 1,\dots,n$, taking the values ± 1 with probability $\frac{1}{2}$,

$$\mathbb{E}\big(W_p(\mu_n,\mu)\big) \,=\, 2^{-1/p}\, n^{-1/(2p)}\, \mathbb{E}\big(|Z_n|^{1/p}\big)$$

where $Z_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n \varepsilon_k$. Since, by the central limit theorem, $\mathbb{P}\{|Z_n| \geq 1\} \to \mathbb{P}\{|Z| \geq 1\}$ with $Z \sim N(0,1)$, we have $\mathbb{P}\{|Z_n| \geq 1\} \geq 2c$ for all $n \geq 1$ with some absolute constant $c > 0$. But, for all $p \geq 1$,

$$\mathbb{E}\big(|Z_n|^{1/p}\big) \,\geq\, \mathbb{P}\big\{|Z_n| \geq 1\big\},$$

so that

$$\mathbb{E}\big(W_p(\mu_n,\mu)\big) \,\geq\, 2c\,2^{-1/p}\,n^{-1/(2p)} \,\geq\, c\,n^{-1/(2p)}$$

showing therefore that the upper-bound in Theorem 7.9 may not be improved.

This example also shows that

$$\mathbb{E}\big(W_\infty(\mu_n,\mu)\big) \,\geq\, c > 0, \qquad n \geq 1.$$

Therefore, there is no convergence of the empirical measures μ_n with respect to the W_∞ transport distance.

In order to judge the sharpness of the bound of Theorem 7.9, we clarify here the role of the connectedness of the support of the measure (which means the continuity of the inverse distribution function).

Theorem 7.11. *If the support of a probability distribution μ on $\mathbb{R}$ is not an interval, then for any $p \geq 1$, and any $n \geq 1$,*

$$\mathbb{E}\big(W_p(\mu_n,\mu)\big) \,\geq\, \frac{c}{n^{1/2p}}$$

with some positive constant c depending on μ and p, only.

For example, if μ is the uniform measure on the set $\Delta = (-2,-1) \cup (1,2)$, then $\mathbb{E}(W_p(\mu_n,\mu)) \geq \frac{c}{n^{1/2p}}$ with some positive constant $c = c_p$.

As a result, we obtain a wide class of measures for which the rate is completely determined.

Corollary 7.12 (Two-sided bounds of order $n^{-1/(2p)}$)**.** *If the support of a compactly supported probability distribution μ is not an interval, then for any $p \geq 1$, and any $n \geq 1$,*

$$\frac{c_0}{n^{1/2p}} \,\leq\, \mathbb{E}\big(W_p(\mu_n,\mu)\big) \,\leq\, \big[\mathbb{E}\big(W_p^p(\mu_n,\mu)\big)\big]^{1/p} \,\leq\, \frac{c_1}{n^{1/2p}}$$

with some positive constants c_0 and c_1 depending on μ and p, only.

Proof of Theorem 7.11. The support $\Delta = \mathrm{supp}(\mu)$ is a closed subset of the real line, consisting of all points $x \in \mathbb{R}$ such that $\mu(A) > 0$ for any open set $A \subset \mathbb{R}$ containing x. Let $a = \inf \Delta$, $b = \sup \Delta$, $-\infty \leq a < b \leq \infty$. If Δ is not an interval with endpoints a and b, there is a proper interval $(a_0,b_0) \subset \Delta$ of μ-measure zero, with $a < a_0 < b_0 < b$.

In that case, the intervals $\Delta_0 = [a, a_0]$ and $\Delta_1 = [b_0, b]$ have positive μ-measures, say q and $1 - q$, respectively.

When picking a point x at random according to μ, it will belong to Δ_0 and Δ_1 with probabilities q and $1 - q$, respectively. Therefore, if S_n denotes the number of points in the sample $X_1, \dots, X_n$ which belong to Δ_0, then S_n has a binomial distribution with parameters (n, q). Let S_n' denote the number of points in Δ_0 for an independent sample $Y_1, \dots, Y_n$ drawn from μ. By the definition of order statistics, with probability one

$$S_n \geq k \iff X_k^* \leq a_0 \qquad \text{and} \qquad S_n' < k \iff Y_k^* \geq b_0.$$

Hence, for the corresponding empirical measures μ_n and ν_n we obtain that

$$\begin{aligned} W_p^p(\mu_n, \nu_n) &= \frac{1}{n} \sum_{k=1}^n |X_k^* - Y_k^*|^p \\ &\geq \frac{1}{n} \sum_{\min(S_n, S_n') < k \leq \max(S_n, S_n')} |X_k^* - Y_k^*|^p \\ &\geq \frac{(b_0 - a_0)^p}{n} \, |S_n - S_n'|. \end{aligned}$$

Therefore, by the triangle inequality,

$$W_p(\mu_n, \mu) + W_p(\nu_n, \mu) \geq W_p(\mu_n, \nu_n) \geq \frac{b_0 - a_0}{n^{1/p}} \, |S_n - S_n'|^{1/p},$$

so that

$$\mathbb{E}\big(W_p(\mu_n, \mu)\big) \geq \frac{b_0 - a_0}{2n^{1/p}} \, \mathbb{E}\big(|S_n - S_n'|^{1/p}\big).$$

But in case of the binomial distribution, L^λ-norms $\|S_n - S_n'\|_\lambda = (\mathbb{E}(|S_n - S_n'|^\lambda))^{1/\lambda}$ ($\lambda > 0$) are equivalent to each other within factors depending on λ, only. This follows, for example, from a dimension free concentration inequality for convex Lipschitz functions on $\mathbb{R}^n$ under product measures on $[-1, 1]^n$, cf. [L3]. In particular,

$$\|S_n - S_n'\|_{1/p} \geq c \, \|S_n - S_n'\|_2 = c\sqrt{2nq(1-q)}$$

with some $c > 0$ depending on p and q, only. Therefore,

$$\mathbb{E}\big(W_p(\mu_n, \mu)\big) \geq \frac{b_0 - a_0}{2n^{1/p}} \, \big(c\sqrt{2nq(1-q)}\,\big)^p.$$

Theorem 7.11 is therefore established. □

7.4 General upper-bounds on $\mathbb{E}(W_p^p(\mu, \nu))$

Since, as we have seen, the rate $[\mathbb{E}(W_p^p(\mu_n, \mu))]^{1/p} = O(\frac{1}{\sqrt{n}})$ requires that μ possess rather strong properties, one may ask whether or not one can obtain weaker reasonable rates under weaker standard assumptions such as moments conditions. For example, compactness of the support of μ is a sufficient condition (cf. Corollary 7.12). Our next purpose is to extend this observation to larger families of probability distributions, including those that have appropriate finite moments. To this end, it is necessary to reach working estimates on $W_p(\mu,\nu)$ which would be explicit in terms of the associated distribution functions of μ and ν. For example, one can use the following general bound.

Lemma 7.13. *Let $\varphi,\psi : \mathbb{R} \to \mathbb{R}$ be continuous functions such that φ is non-negative, ψ is non-decreasing, and, for all $x,y \in \mathbb{R}$,*

$$\varphi(x-y) \,\leq\, \big|\psi(x)-\psi(y)\big|.$$

Then, for any distribution functions F and G,

$$\int_0^1 \varphi\big(F^{-1}(t)-G^{-1}(t)\big)\,dt \,\leq\, \int_{-\infty}^{\infty} \big|F(x)-G(x)\big|\,d\psi(x).$$

Proof. One may assume that ψ is strictly increasing and that $\psi(-\infty)=-\infty$, $\psi(\infty)=\infty$ (since otherwise one may apply the result to $\psi_\varepsilon(x)=\psi(x)+\varepsilon x$ with $\varepsilon \downarrow 0$). In that case, introduce the inverse function $\psi^{-1} : \mathbb{R} \to \mathbb{R}$ and the new distribution functions

$$\widetilde{F}(x) \,=\, F\big(\psi^{-1}(x)\big), \qquad \widetilde{G}(x) \,=\, G\big(\psi^{-1}(x)\big).$$

By the assumption on φ,ψ and using Theorem 2.10 with $p=1$,

$$\begin{aligned}\int_0^1 \varphi\big(F^{-1}(t)-G^{-1}(t)\big)\,dt \ &\leq\ \int_0^1 \big|\psi\big(F^{-1}(t)\big)-\psi\big(G^{-1}(t)\big)\big|\,dt \\ &=\ \int_0^1 \big|\widetilde{F}^{-1}(t)-\widetilde{G}^{-1}(t)\big|\,dt \\ &=\ \int_{-\infty}^{\infty} \big|\widetilde{F}(x)-\widetilde{G}(x)\big|\,dx.\end{aligned}$$

Changing the variable $x=\psi(y)$,

$$\int_{-\infty}^{\infty} \big|\widetilde{F}(x)-\widetilde{G}(x)\big|\,dx \ =\ \int_{-\infty}^{\infty} \big|F(y)-G(y)\big|\,d\psi(y)$$

and the lemma is proved. □

Under certain restrictions on φ,ψ,F,G, inequalities such as in Lemma 7.13 were introduced by Ebralidze, and the above proof is based on his simple argument (cf. [Eb]). Later, Borisov and Shadrin generalized Ebralidze-type inequalities to the form

$$\int_0^1 \varphi\big(F^{-1}(t)-G^{-1}(t)\big)dt \,\leq\, C\int_{-\infty}^{\infty} \big|F(x)-G(x)\big|\,\big|d\varphi(x)\big|$$

with $C = |A - B| + B$, assuming that φ is a non-negative, continuous, even function on $\mathbb{R}$, which is non-decreasing for $x > 0$ and satisfying

$$\varphi(x+y) \ \leq \ A\,\big(\varphi(x) + \varphi(y)\big), \qquad \varphi(x-y) \ \leq \ B\,\big|\varphi(x) - \varphi(y)\big|$$

for all $x, y \geq 0$ (see also Theorem 1, [B-S]). This result is included in Lemma 7.13 by applying it with

$$\psi(x) \ = \ C' \operatorname{sign}(x)\varphi(x),$$

where $C' = \max(A, B)$ is even slightly better in case $B > A$.

If additionally φ is convex, then $B = 1$ (which is best possible) and necessarily $A \geq 1$, so that $C' = C = A$. As emphasized in [B-S], in the important case $\varphi(x) = |x|^p$, $p \geq 1$, we have $A = 2^{p-1}$. Hence, using Theorem 2.10, we reach the following statement.

Proposition 7.14. *Let μ and ν be probability measures on $\mathbb{R}$ with distribution functions F and G respectively. For all $p \geq 1$,*

$$W_p^p(\mu,\nu) \ \leq \ p\, 2^{p-1} \int_{-\infty}^{\infty} |x|^{p-1} \left|F(x) - G(x)\right| dx. \tag{7.2}$$

In particular,

$$W_2^2(\mu,\nu) \ \leq \ 4 \int_{-\infty}^{\infty} |x| \left|F(x) - G(x)\right| dx.$$

Remark 7.15. Another way to obtain similar bounds is to connect the Kantorovich distances with the Zolotarev so-called ideal metrics. In one partial case, it is defined by

$$\zeta_2(\mu,\nu) \ = \ \sup_{\|u'\|_{\mathrm{Lip}} \leq 1} \bigg| \int_{-\infty}^{\infty} u\, d\mu \ - \int_{-\infty}^{\infty} u\, d\nu \bigg|$$

where the supremum is taken over all continuously differentiable functions $u : \mathbb{R} \rightarrow \mathbb{R}$ whose derivative u' has Lipschitz semi-norm $\|u'\|_{\mathrm{Lip}} \leq 1$. The quantity $\zeta_2(\mu,\nu)$ is called the Zolotarev ideal metric of order 2.

It is easy to see that $\zeta_2(\mu,\nu) < \infty$ if μ and ν have finite second moments and equal first moments, that is, $\int_{-\infty}^{\infty} x\, d\mu(x) = \int_{-\infty}^{\infty} x\, d\nu(x)$. However, $\zeta_2(\mu,\nu) = \infty$ if the first moments are different. Restricting ζ_2 to the class of probability measures on the real line with a fixed first moment (and bounded p-th moments, $p > 2$), it represents a metric generating the topology of weak convergence. So, it is still of weak type like W_p.

Recall from Theorem 2.5 that there is a similar representation for W_1, namely

$$W_1(\mu,\nu) \ = \ \sup_{\|u\|_{\mathrm{Lip}} \leq 1} \bigg| \int_{-\infty}^{\infty} u\, d\mu \ - \int_{-\infty}^{\infty} u\, d\nu \bigg|$$

where, however, the supremum is taken over all $u : \mathbb{R} \to \mathbb{R}$ with $\|u\|_{\mathrm{Lip}} \le 1$. As for W_2, there is the following relation proved by Rio [Ri]: For all probability measures μ and ν on the real line with finite second moments,

$$W_2^2(\mu,\nu) \,\le\, 4\,\zeta_2(\mu,\nu)$$

(in fact, Rio considered more general distances W_p and ζ_p). On the other hand, the Zolotarev metric of order 2 admits an equivalent representation

$$\zeta_2(\mu,\nu) \,=\, \int_{-\infty}^{\infty} \bigg| \int_{-\infty}^{x} \big(F(y) - G(y)\big)\, dy \bigg|\, dx,$$

valid when μ and ν have equal first moments. This identity is elementary and may be found in [Z]. It can be used to show that without the first moment restriction

$$\begin{aligned} W_2^2(\mu,\nu) \ &\le\ 8 \int_{-\infty}^{\infty} |x|\, \big|F(x) - G(x)\big|\, dx \\ &\quad + 8\, \big|\mathbb{E}(X) - \mathbb{E}(Y)\big|\, \mathbb{E}\big(|X|\big) + 6\, \big(\mathbb{E}(X) - \mathbb{E}(Y)\big)^2, \end{aligned}$$

where X and Y are random variables with respective laws μ and ν and distributions functions F and G, having finite second moments. The last bound is not as sharp as the (second) bound of Proposition 7.14, although both lead to similar conclusions when applying them to the empirical measures.

7.5 Moment upper-bounds of order $n^{-1/2}$ on $\mathbb{E}(W_p^p(\mu_n,\mu))$

Proposition 7.14 of the preceding paragraph may now be used towards upper-bounds of order $n^{-1/2}$ on $\mathbb{E}(W_p^p(\mu_n,\mu))$. Let thus $X_1,\dots,X_n$ be a sample drawn from a Borel probability μ on $\mathbb{R}$ with distribution function F. We are in position to apply Proposition 7.14 to the couple (μ_n,μ), where μ_n is the empirical measure of the sample. As a result, we arrive at the following extension of Theorem 3.2.

Theorem 7.16. *Let $p \ge 1$. In the preceding notation, for all $n \ge 1$,*

$$\mathbb{E}\big(W_p^p(\mu_n,\mu)\big) \,\le\, \frac{p\,2^{p-1}}{\sqrt{n}} \int_{-\infty}^{\infty} |x|^{p-1} \sqrt{F(x)(1-F(x)}\, dx. \tag{7.3}$$

In particular,

$$\mathbb{E}\big(W_2^2(\mu_n,\mu)\big) \,\le\, \frac{4}{\sqrt{n}} \int_{-\infty}^{\infty} |x|\, \sqrt{F(x)(1-F(x)}\, dx.$$

The finiteness of the integral in (7.3) is equivalent to saying that the random variable $|X|^p$ (where X has law μ) belongs to the Lorentz Banach space $L^{2,1}$, cf. (3.4). It is so in particular if X belongs to L^s for some $s > 2p$. Indeed, if $\mathbb{E}(|X|^s) < \infty$, then $F(x)(1 - F(x)) = O(|x|^{-s})$ as $|x| \to \infty$, so that

$$\int_{-\infty}^{\infty} |x|^{p-1} \sqrt{F(x)(1 - F(x)}\, dx \, < \, \infty$$

as long as $s > 2p$.

Corollary 7.17 (Upper-bound on $\mathbb{E}(W_p^p(\mu_n, \mu))$ under moment conditions)**.** *Let $p \geq 1$. If $\mathbb{E}(|X|^s) < \infty$ for some $s > 2p$, then*

$$\mathbb{E}\big(W_p^p(\mu_n, \mu)\big) \, = \, O\bigg(\frac{1}{\sqrt{n}}\bigg) \qquad as \quad n \to \infty.$$

The expression $\mathbb{E}(|F_n(x) - F(x)|)$ may be bounded from above in a sharper way, since, as in the proof of Theorem 3.5, for each $n \geq 1$,

$$\mathbb{E}\big(\big|F_n(x) - F(x)\big|\big) \, \leq \, \frac{1}{n} \, \min \Big\{ 2nF(x)\big(1 - F(x)\big), \sqrt{nF(x)(1 - F(x))} \, \Big\}.$$

Using this in the general bound (7.2), the inequality (7.3) of Theorem 7.16 may be sharpened to

$$\begin{aligned} \mathbb{E}\big(W_p^p(\mu_n, \mu)\big) \, &\leq \, p\, 2^p \int_{\{4nF(x)(1-F(x)) \leq 1\}} |x|^{p-1} \, F(x)\big(1 - F(x)\big)\, dx \\ &\quad + \frac{p\, 2^{p-1}}{\sqrt{n}} \int_{\{4nF(x)(1-F(x)) > 1\}} |x|^{p-1} \sqrt{F(x)(1 - F(x))}\, dx. \end{aligned} \tag{7.4}$$

This estimate is applicable for arbitrary probability measures μ on $\mathbb{R}$ with finite moment of order p, although it might lead to weaker rates for $\mathbb{E}(W_p^p(\mu_n, \mu))$ in comparison with Corollary 7.17 with its moment condition.

For example, if we know that $c = \mathbb{E}(|X|^{2p}) < \infty$, then, by Chebyshev's inequality,

$$F(x)(1 - F(x)) \, \leq \, \frac{1 + c}{1 + |x|^{2p}}\,, \qquad x \geq 0.$$

Hence, the second integral in (7.4) is bounded by

$$\sqrt{1 + c} \int_{\{|x|^{2p} < 4n(1+c) - 1\}} \frac{|x|^{p-1}}{\sqrt{1 + |x|^{2p}}}\, dx \, = \, O\big(\log n\big).$$

In order to bound the first integral in (7.4), let $u(t) = F^{-1}(t)$ so that $c = \int_0^1 |u(t)|^{2p}\, dt$. Here, the region of integration is contained in the set $\min\{F(x), 1 - F(x)\} \leq \frac{1}{8n}$ which is part of $(-\infty, r_n] \cap [r_n, \infty)$ for some $r_n \to \infty$. Integrating by parts and using the Cauchy-Schwarz inequality, we have, for all n large enough,

$$\begin{aligned} p \int_{\{1-F(x) \leq \frac{1}{8n}\}} x^{p-1} \big(1 - F(x)\big)\, dx \ &\leq \ \int_{\{1-F(x) \leq \frac{1}{8n}\}} x^p \, dF(x) \\ &= \ \int_{1-\frac{1}{8n}}^{1} u(t)^p \, dt \ \leq \ \frac{\sqrt{c}}{\sqrt{8n}}\,. \end{aligned}$$

A similar bound also holds for the integral of $|x|^{p-1} F(x)$ over the part $\{F(x) \leq \frac{1}{8n}\}$. We may therefore conclude to the following statement.

Corollary 7.18. *If* $\mathbb{E}(|X|^{2p}) < \infty$*, then*

$$\mathbb{E}\big(W_p^p(\mu_n, \mu)\big) \ = \ O\bigg(\frac{\log n}{\sqrt{n}}\bigg) \qquad as \quad n \to \infty.$$

Analogues of Corollaries 7.17 and 7.18 for samples of vector-valued random variables have been achieved recently in [D-S-S] and [F-G] by coupling methods.

7.6 W_∞-convergence of empirical distributions

We conclude this section with the issue of the convergence of μ_n to μ with respect to the W_∞ distance, in the mean or with probability one (analogously to the Glivenko-Cantelli theorem). The standard rate in W_∞ distance was investigated in sub-Section 5.6. We complete the study here with the following characterization.

Theorem 7.19 (Characterization of convergence of $W_\infty(\mu_n, \mu)$)**.** *Each of the following three statements*

a) $W_\infty(\mu_n, \mu) \to 0$ *as* $n \to \infty$ *with probability one,*

b) $W_\infty(\mu_n, \mu) \to 0$ *as* $n \to \infty$ *in probability,*

c) $\mathbb{E}(W_\infty(\mu_n, \mu)) \to 0$ *as* $n \to \infty$

is equivalent to the property that the support of μ *is a finite closed interval.*

Proof. First recall the formula

$$W_\infty(\mu_n, \mu) \ = \ \sup_{0<t<1} \big|F_n^{-1}(t) - F^{-1}(t)\big|,$$

where F is the distribution function of the sample, and F_n is the empirical distribution function. In particular, this distance is finite, if and anly if μ is compactly supported. The latter property is thus necessary for each of the three statements.

Hence, one may assume that μ is supported on an interval with finite length ℓ. But then $W_\infty(\mu_n,\mu) \le \ell$ with probability one, and therefore $a) \Rightarrow b) \Rightarrow c)$, by the Lebesgue dominated convergence theorem for the last implication.

Let us now derive from $c)$ the conclusion about the connectedness of the support. Due to the triangle inequality for W_∞, this hypothesis implies that

$$\lim_{n\to\infty} \mathbb{E}\big(W_\infty(\mu_n,\nu_n)\big) \, = \, 0,$$

where ν_n is an independent copy of μ_n. We employ the second lower-bound of Theorem 5.7, namely

$$\big[\mathbb{E}\big(W_p(\mu_n,\nu_n)\big)\big]^p \, \ge \, c^p \int_{\{t(1-t)\ge \frac{4}{\sqrt{n+1}}\}} \bigg[F^{-1}\Big(t+\frac{1}{6}\,\varepsilon_n(t)\Big) - F^{-1}\Big(t-\frac{1}{6}\,\varepsilon_n(t)\Big)\bigg]^p \, dt.$$

Here $\varepsilon_n(t) = \sqrt{\frac{t(1-t)}{n+1}}$, $p \ge 1$, and c is a positive numerical constant. Raising both sides of this bound to the power $1/p$ and letting $p \to \infty$, we get in the limit

$$\mathbb{E}\big(W_\infty(\mu_n,\nu_n)\big) \, \ge \, c \sup_{t(1-t)>\frac{4}{\sqrt{n+1}}} \bigg[F^{-1}\Big(t+\frac{1}{6}\,\varepsilon_n(t)\Big) - F^{-1}\Big(t-\frac{1}{6}\,\varepsilon_n(t)\Big)\bigg].$$

Hence, for any fixed $t_0 \in (0,\frac{1}{2})$ and for all sufficiently large n so that $t_0(1-t_0) > \frac{4}{\sqrt{n+1}}$, we have

$$\mathbb{E}\big(W_\infty(\mu_n,\nu_n)\big) \, \ge \, c \bigg[F^{-1}\Big(t+\frac{1}{6}\,\varepsilon_n(t_0)\Big) - F^{-1}\Big(t-\frac{1}{6}\,\varepsilon_n(t_0)\Big)\bigg],$$

whenever $t_0 \le t \le 1-t_0$. Letting now $n \to \infty$, we get

$$\lim_{n\to\infty} \bigg[F^{-1}\Big(t+\frac{1}{6}\,\varepsilon_n(t_0)\Big) - F^{-1}\Big(t-\frac{1}{6}\,\varepsilon_n(t_0)\Big)\bigg] \, = \, 0.$$

But this means that the inverse function F^{-1} is continuous on $(t_0, 1-t_0)$, and therefore it is continuous on the whole unit interval $(0,1)$. The latter is equivalent to the connectedness of the support (cf. Proposition A.7).

It remains to derive the statement $a)$ from the property $\mathrm{supp}(\mu) = [a,b]$. In that case, the inverse function $F^{-1} : [0,1] \to [a,b]$ is continuous, and hence its modulus of continuity is vanishing at zero, $\delta_{F^{-1}}(0+) = 0$.

Again, let ν_n be an independent copy of μ_n constructed for the first n values in the sample $(Y_n)_{n\ge1}$, and let $(X_k^*)_{1\le k\le n}$ and $(Y_k^*)_{1\le k\le n}$ be the corresponding order statistics. Since, by the convexity of the distance,

$$W_\infty(\mu_n,\mu) \, \le \, \mathbb{E}_Y\big(W_\infty(\mu_n,\nu_n)\big),$$

and since $W_\infty \le b - a$, it will be sufficient to see that $W_\infty(\mu_n, \nu_n) \to 0$ a.s. as $n \to \infty$. We now employ another formula, given in Lemma 4.2, namely

$$W_\infty(\mu_n, \nu_n) \ = \ \max_{1 \le k \le n} |X_k^* - Y_k^*|.$$

Write $X_k^* = F^{-1}(U_k^*)$ and $Y_k^* = F^{-1}(V_k^*)$ by using order statistics for two independent samples $U_1, \dots, U_n$ and $V_1, \dots, V_n$ drawn from the uniform distribution μ' on $(0,1)$, and denote by μ_n' and ν_n' the empirical measures for these samples. Then,

$$\begin{aligned} W_\infty(\mu_n, \nu_n) \ &\le \ \max_{1 \le k \le n} \delta_{F^{-1}}\big(|U_k^* - V_k^*|\big) \\ &= \ \delta_{F^{-1}}\big(\max_{1 \le k \le n} |U_k^* - V_k^*|\big) \ = \ \delta_{F^{-1}}(W_\infty(\mu_n', \nu_n')). \end{aligned}$$

In addition, by the triangle inequality.

$$W_\infty(\mu_n', \nu_n') \ \le \ W_\infty(\mu_n', \mu') + W_\infty(\nu_n', \mu').$$

But, as was already emphasized in sub-Section 4.2, in the particular case of the uniform distribution, with probability one we have

$$W_\infty(\mu_n', \mu') \ = \ \sup_{0<x<1} \big|\mu_n'\big([0,x]\big) - \mu'\big([0,x]\big)\big| \ = \ \sup_{0<x<1} \big|\mu_n'\big([0,x]\big) - x\big|.$$

These random variables do tend to zero a.s. by the Glivenko-Cantelli theorem. Similarly, $W_\infty(\nu_n', \mu') \to 0$ a.s. The theorem is therefore established. $\square$

Appendices

A Inverse distribution functions

Throughout this appendix, μ denotes a probability measure on the Borel sets of $\mathbb{R}$, with associated distribution function $F = \mu((-\infty, x])$, $x \in \mathbb{R}$, uniquely determining μ. The measure μ (or its distribution function F) is said to be degenerate if it is a Dirac mass. We sometimes denote by X a random variable on a probability space $(\Omega, \Sigma, \mathbb{P})$ with distribution (law) μ.

At significant occasions in this work, the study is reduced to the uniform distribution by means of the inverse function F^{-1} (cf. Proposition A.1 below). It is therefore important to freely work with the inverse distribution functions and their analytic properties. This appendix thus collects material on supports and continuity, modulus of continuity and absolute continuity of inverse distribution functions. The notion of I-function which plays an important role in this investigation is addressed in this framework and the study of integrals containing the derivative of F^{-1}.

A number of results presented in this appendix are classical, although often they are not always stated in full generality. Some results seem to be new, or we could not precisely determine suitable references. Among the bibliography pertinent to this appendix, in particular in connection with empirical measures and processes, let us mention [S-W], [C-H]. In particular, the study of quantile processes undertaken in these monographs involves similarly inverse distribution functions and associated I-functions as presented here, usually however under some regularity conditions.

A.1 Inverse distribution functions

Let F be a distribution function on the real line. Recall its inverse function

$$F^{-1}(t) \,=\, \inf\big\{x \in \mathbb{R} : F(x) \geq t\big\}, \qquad 0 < t < 1. \tag{A.1}$$

Since F is non-decreasing and continuous from the right, the infimum in the definition of $F^{-1}(t)$ is always attained at some point, so

$$F^{-1}(t) \,=\, \min\big\{x \in \mathbb{R} : F(x) \geq t\big\}, \qquad 0 < t < 1.$$

It is convenient furthermore to extend this function to $[0,1]$ by monotonicity, setting

$$\begin{aligned} F^{-1}(0) \,&=\, F^{-1}(0-) \,=\, \inf\big\{x \in \mathbb{R} : F(x) > 0\big\}, \\ F^{-1}(1) \,&=\, F^{-1}(1-) \,=\, \sup\big\{x \in \mathbb{R} : F(x) < 1\big\}, \end{aligned}$$

similarly to the standard convention $F(-\infty) = 0$, $F(\infty) = 1$.

The use of the inverse functions is mainly explained by the following well-known observation.

Proposition A.1. *Let F be a distribution function. If U is a random variable uniformly distributed in $(0,1)$, then the random variable $F^{-1}(U)$ has F as its distribution function.*

Another related (although less universal) property is that, when a random variable X has a continuous distribution function F, the random variable $U = F(X)$ is uniformly distributed in the interval $(0,1)$.

The next statement describes how the transform $F \mapsto F^{-1}$ acts between its domain and image.

Proposition A.2. *Any inverse distribution function is non-decreasing and left-continuous. Moreover, for any non-decreasing, left-continuous function $G : (0,1) \to \mathbb{R}$, there exists a unique distribution function F such that $F^{-1} = G$.*

Thus, when thinking of F in terms of F^{-1}, there should be no constraint on the latter function except for the property of being non-decreasing and left-continuous. The next statement lists a number of basic properties and relations between F and F^{-1}. The various claims are elementary and are verified in a straightforward manner.

Lemma A.3 (Properties of inverse distribution functions)**.** *Given a distribution function F, the following hold for all $0 < t < s < 1$ and $x \in \mathbb{R}$.*

1) $F^{-1}(t) \le x$ *if and only if* $F(x) \ge t$.
2) $F^{-1}(t) > x$ *if and only if* $F(x) < t$.
3) $F^{-1}(t) \le x < F^{-1}(s)$ *if and only if* $t \le F(x) < s$.
4) $F(F^{-1}(t)) \ge t$ *with equality if and only if* $t = F(y)$ *for some* $y \in \mathbb{R}$. *In particular,*
5) $F(F^{-1}(t)) = t$ *if* F *is continuous.*
6) $F^{-1}(F(x)) \le x$ *if and only if* $x \ge F^{-1}(0)$. *Moreover,*
7) $F^{-1}(F(x)) < x$ *if and only if* $F(y) = F(x)$ *for some* $y < x$ $\;(x > F^{-1}(0))$.
8) F^{-1} *is strictly increasing on* $(0,1)$ *if and only if* F *is continuous.*

Note that Properties 1)-2) also hold for $t = 1$.

On the basis of this lemma, we address the proofs of Propositions A.1 and A.2.

Proof of Propositions A.1 and A.2. Proposition A.1 is immediate since by Lemma A.3, for all $x \in \mathbb{R}$,

$$\lambda\big\{t \in (0,1) : F^{-1}(t) \le x\big\} \;=\; \lambda\big\{t \in (0,1) : t \le F(x)\big\} \;=\; F(x)$$

where λ denotes Lebesgue measure on $(0,1)$.

Turning to Proposition A.2, assume first that F^{-1} is not left-continuous at some point t. That is, setting $x = F^{-1}(t)$, there exists a $\delta > 0$ such that, for all $\varepsilon > 0$,

$$F^{-1}(t - \varepsilon) \;\le\; F^{-1}(t) - \delta \;=\; x - \delta.$$

By Lemma A.3, this is equivalent to saying that $F(x - \delta) \ge t - \varepsilon$. Letting $\varepsilon \downarrow 0$, we get $F(x - \delta) \ge t$, which in turn is equivalent to $x - \delta \ge F^{-1}(t)$, a contradiction. The first assertion of Proposition A.2 is established.

Next turn to the existence part of the second claim. Set $G(1) = G(1-)$ and define on $\mathbb{R}$ the function

$$F(x) \,=\, \sup\big\{t \in (0,1] : G(t) \le x\big\},$$

using (if necessary) the convention that $\sup\emptyset = 0$. In particular, $F(x) > 0$ if and only if $G(t) \le x$ for some $t \in (0,1)$. By construction, F is non-decreasing and takes values in $[0,1]$. To prove that it is right-continuous, observe the following. Fix $t_0 \in (0,1]$ and $x \in \mathbb{R}$. If $G(t_0) \le x$, then the set $\{G(t) \le x\}$ is non-empty, and $F(x) \ge t_0$. Conversely, if $F(x) \ge t_0$, then $F(x) > 0$, so the set $\{G(t) \le x\}$ is non-empty. Hence, using the left continuity of G,

$$\sup\big\{t \in (0,1] : G(t) \le x\big\} \ge t_0 \implies \exists\, t \ge t_0\,, G(t) \le x \implies G(t_0) \le x.$$

Thus, for all $t \in (0,1]$ and $x \in \mathbb{R}$,

$$G(t) \le x \iff t \le F(x). \tag{A.2}$$

Now, assume that F is not-right continuous at some point x_0 and put $t_0 = F(x_0)$. Then, there exists $\delta > 0$ such that, for all $\varepsilon > 0$,

$$F(x_0 + \varepsilon) \,\ge\, F(x_0) + \delta \,=\, t_0 + \delta.$$

In particular, $t = t_0 + \delta \in (0,1]$. By (A.2), $G(t_0 + \delta) \le x_0 + \varepsilon$. Letting $\varepsilon \downarrow 0$, we are led to $G(t_0 + \delta) \le x_0$, which in turn is equivalent to $t_0 + \delta \le F(x_0)$, a contradiction with $t_0 = F(x_0)$. The existence of the distribution function F is therefore established.

We are left with uniqueness. Let $\widetilde{F}$ be another distribution function such that $\widetilde{F}^{-1} = G$. By (A.2) applied to both F and $\widetilde{F}$, for all $t \in (0,1]$ and $x \in \mathbb{R}$,

$$t \le \widetilde{F}(x) \iff t \le F(x)$$

which clearly amounts to $F(x) = \widetilde{F}(x)$. The proof of Proposition A.2 is therefore complete. □

Being non-decreasing and left-continuous, the inverse function F^{-1} of a given distribution function F generates a non-negative Borel measure μ^{-1} on $(0,1)$, defined for semi-open intervals by

$$\mu^{-1}\big([t,s)\big) \,=\, F^{-1}(s) - F^{-1}(t), \qquad 0 < t < s < 1.$$

It may be called the inverse measure (with respect to the probability measure μ with the distribution function F). The next statement describes μ^{-1} for any Borel sets.

Proposition A.4 (Inverse measure)**.** *Any non-degenerate distribution function F restricted to the interval $\Delta = \{x \in \mathbb{R} : 0 < F(x) < 1\}$ pushes forward the Lebesgue measure λ on Δ onto the inverse measure μ^{-1}. That is, for any Borel set $A \subset (0,1)$,*

$$\lambda\big\{x \in \Delta : F(x) \in A\big\} \,=\, \mu^{-1}(A).$$

It is sufficient to verify this equality for $A = [t, s)$ with arbitrary $0 < t < s < 1$ and indeed, by Lemma A.3,

$$\lambda\big\{x \in \Delta : t \le F(x) < s\big\} \;=\; \lambda\big\{x \in \Delta : F^{-1}(t) \le x < F^{-1}(s)\big\} \;=\; \mu^{-1}\big([t, s)\big).$$

There are other interesting properties of the transform $F \mapsto F^{-1}$ such as the following one which is equivalent to the so-called "Elementary Skorokhod Theorem" (cf. [S-W], pp. 9-10). We provide a proof for completeness.

Lemma A.5 (Elementary Skorokhod Theorem)**.** *Let $(F_n)_{n\in\mathbb{N}}$ be a sequence of distribution functions, and assume that $F_n \to F$ weakly, i.e. $\lim_{n\to\infty} F_n(x) = F(x)$ for any point x of continuity of F. Then*

$$\lim_{n\to\infty} F_n^{-1}(t) \;=\; F^{-1}(t),$$

for any point t of continuity of F^{-1}.

Proof. Let t be a point of continuity of F^{-1} and let $T \subset \mathbb{R}$ be the set of all continuity points of all F_n's (which is dense on the real line). We first show that, given $\varepsilon > 0$, $F_n^{-1}(t) \le F^{-1}(t) + \varepsilon$ for all n large enough. By Lemma A.3, this is equivalent to $F_n(F^{-1}(t) + \varepsilon) \ge t$. Choose $x \in T$ such that

$$F^{-1}(t) + \varepsilon \;>\; x \;>\; F^{-1}(t) + \frac{\varepsilon}{2}\,.$$

Then $F_n(x) \ge F(F^{-1}(t) + \frac{\varepsilon}{2}) - \delta$, for all n large enough with any prescribed $\delta > 0$. Hence, it suffices to prove that $F(F^{-1}(t) + \frac{\varepsilon}{2}) \ge t + \delta$. But if $0 < \delta < 1 - t$, the latter is equivalent to $F^{-1}(t) + \frac{\varepsilon}{2} \ge F^{-1}(t + \delta)$. An appropriate value of δ can be then chosen once F^{-1} is continuous at the point t. By a similar argument, $F_n^{-1}(t) > F^{-1}(t) - \varepsilon$ for all n large enough, concluding therefore the proof of the lemma. □

The preceding Lemma A.5 may be used, for example, to justify the identity in Theorem 2.10,

$$W_p^p(\mu, \nu) = \int_0^1 \big|F^{-1}(t) - G^{-1}(t)\big|^p \, dt$$

where F and G are the distribution functions associated with the probability measures μ and ν from the space $Z_p(\mathbb{R})$. Indeed, if these measures are supported on n distinct points with mass $\frac{1}{n}$, the above identity is reduced to Lemma 4.2 (based on the elementary Lemma 4.1). In the general case, one can approximate μ and ν by such discrete measures μ_n and ν_n in the metric W_p. If F_n, F, and G_n, G are the associated distribution functions, then necessarily $F_n \to F$ and $G_n \to G$ weakly. Hence, by Lemma A.5 and Fatou's lemma,

$$\begin{aligned}\int_0^1 \left|F^{-1}(t) - G^{-1}(t)\right|^p dt &\leq \liminf_{n\to\infty} \int_0^1 \left|F_n^{-1}(t) - G_n^{-1}(t)\right|^p dt \\ &= \liminf_{n\to\infty}\ W_p^p(\mu_n, \nu_n) \\ &= W_p^p(\mu, \nu).\end{aligned}$$

On the other hand, the joint distribution F^{-1} of G^{-1} under the Lebesgue measure on $(0,1)$ has μ and ν as marginals. Hence, by Definition 2.1, there is an opposite inequality

$$W_p^p(\mu, \nu) \leq \int_0^1 \left|F^{-1}(t) - G^{-1}(t)\right|^p dt.$$

A.2 Supports and continuity

In this paragraph, we comment on issues connected with the support of measures. Given a distribution function F on $\mathbb{R}$ and its inverse function F^{-1}, observe first that $[F^{-1}(0), F^{-1}(1)]$ represents the smallest closed interval in $[-\infty, \infty]$ on which the measure μ with the distribution function F is supported. In general, the support $\mathrm{supp}(\mu)$ is defined as the smallest closed subset of the real line $\mathbb{R}$ of full μ-measure. This set can be defined as the collection of all points x of growth of F, i.e. such that $F(x+\varepsilon) > F(x-\varepsilon)$ for every $\varepsilon > 0$.

A similar definition is applied to a general Borel measure generated by a non-decreasing function on a given interval, and in particular to the inverse measure μ^{-1} on $(0,1)$ generated by the inverse distribution function F^{-1}. Its support, i.e. the smallest closed subset of $(0,1)$ of full μ^{-1}-measure, is described in terms of the image set

$$\mathrm{Im}(F) \,=\, \big\{F(x) : -\infty \leq x \leq \infty\big\}.$$

As an equivalent definition, one may involve the inverse function to write the representation

$$\mathrm{Im}(F) \cap (0,1) \,=\, \big\{t \in (0,1) : F^{-1}(t) < F^{-1}(s) \text{ for all } \ s \in (t,1)\big\}. \tag{A.3}$$

Indeed, using Lemma A.3, we have

$$\begin{aligned}F^{-1}(t) < F^{-1}(s) &\iff \exists\, x \in \mathbb{R},\ F^{-1}(t) \leq x < F^{-1}(s) \\ &\iff \exists\, x \in \mathbb{R},\ t \leq F(x) < s.\end{aligned}$$

By the continuity of F from the right, the latter property holds true for all $s \in (t,1)$ if and only if $t = F(x)$ for some $x \in \mathbb{R}$, thus proving the claim.

By Proposition A.4, since $F : \mathbb{R} \to \mathrm{Im}(F)$, the measure μ^{-1} is supported on $\mathrm{Im}(F)$ (once we realize that the image set is Borel measurable). This set does not need be closed, but its closure is just

$$\mathrm{clos}(\mathrm{Im}(F)) \,=\, \mathrm{Im}(F) \cup \big\{F(x-) : x \in \mathbb{R}\big\}.$$

Proposition A.6 (Support of the inverse measure)**.** *Let μ be a probability measure on $\mathbb{R}$ with distribution function F. A number $t \in (0,1)$ is a point of growth of F^{-1} if and only if $t = F(x)$ or $t = F(x-)$ for some $x \in \mathbb{R}$. Equivalently,*

$$\mathrm{supp}(\mu^{-1}) \,=\, \mathrm{clos}\big(\mathrm{Im}(F)\big) \cap (0,1).$$

Proof. The support of μ^{-1}, $\mathrm{supp}(\mu^{-1})$, represents the collection of all points t of growth of F^{-1} on $(0,1)$, i.e. such that $F^{-1}(t_0) < F^{-1}(t_1)$ whenever $0 < t_0 < t < t_1 < 1$. Using Lemma A.3 and arguing as before, we have

$$\begin{aligned} F^{-1}(t_0) < F^{-1}(t_1) \quad &\Longleftrightarrow \quad \exists\, x \in \mathbb{R},\ F^{-1}(t_0) \le x < F^{-1}(t_1) \\ &\Longleftrightarrow \quad \exists\, x \in \mathbb{R},\ t_0 \le F(x) < t_1 \\ &\Longleftrightarrow \quad [t_0, t_1) \cap \mathrm{Im}(F) \neq \emptyset. \end{aligned}$$

Hence, t is a point of growth of F^{-1}, if and only if $[t_0,t_1) \cap \mathrm{Im}(F) \neq \emptyset$ for all t_0, t_1 such that $t_0 < t < t_1$. But the latter is equivalent to the property $t \in \mathrm{clos}(\mathrm{Im}(F))$. □

Next, we comment more on the structure of the image set $\mathrm{Im}(F)$. If the measure μ with distribution function F is non-atomic (F is continuous), then $\mathrm{Im}(F) = [0,1]$, and F^{-1} is strictly increasing. In the general case, let us return to the representation (A.3) and consider the complement of the image in $(0,1)$,

$$A \,=\, (0,1) \setminus \mathrm{Im}(F) \,=\, \big\{t \in (0,1) : F^{-1}(t) = F^{-1}(s) \text{ for some } s \in (t,1)\big\}.$$

With every point t in A, this set also contains some non-empty interval $[t,s)$. For a rational number $r \in (0,1)$, denote by e_r the union of all such intervals that contain r. Clearly, if e_r is non-empty, it is an interval either of type $[a,b)$ or (a,b). Thus, $A = \bigcup_r e_r$, which shows in particular that $\mathrm{Im}(F)$ is always Borel measurable.

The following proposition collects conditions insuring the continuity of the inverse distribution functions in terms of the support of the measure.

Proposition A.7 (Continuity and support)**.** *Let μ be a probability measure on $\mathbb{R}$ with distribution function F. The following properties are equivalent:*

a) The function F is strictly increasing on the interval $\Delta_0 = \{x \in \mathbb{R} : 0 < F(x) < 1\}$;
b) The inverse function F^{-1} is continuous;
c) The inverse measure μ^{-1} is non-atomic;
d) The support of μ is a closed interval on the real line, finite or not.

Thus, for the continuity of F^{-1}, the support $\Delta = \mathrm{supp}(\mu)$ should be one of the following types

$$1)\ \Delta = (-\infty,\infty), \quad 2)\ \Delta = (-\infty,b], \quad 3)\ \Delta = [a,\infty), \quad 4)\ \Delta = [a,b]$$

with some finite a, b. Then, in the corresponding cases

1) $\Delta_0 = (-\infty, \infty)$, 2) $\Delta = (-\infty, b)$, 3) $\Delta_0 = [a, \infty)$ or (a, ∞), 4) $\Delta_0 = [a, b)$ or (a, b)

(in the two last cases depending on whether μ has an atom at the point a). Anyhow, all points of Δ will be points of growth of F. In particular, according to Lemma A.3, $F^{-1}(F(x)) = x$ for every $x \in \Delta$.

Proof. The equivalence of b) and c) is standard. For $b) \Rightarrow a)$, assume by contradiction that F is not strictly increasing on Δ_0, that is, $t = F(y) = F(x)$ for some $y < x$ with $0 < t < 1$. Let y be the smallest number satisfying this equality with fixed x. Then, by Lemma A.3, $F^{-1}(t) = y$, while $F^{-1}(s) > x$ whenever $s > t$. Hence, F^{-1} is discontinuous at t, thus proving the implication. Conversely, if F^{-1} is discontinuous at $t \in (0,1)$ and $y = F^{-1}(t)$, then $x = F^{-1}(t+) > y$. Hence F is constant on $[y, x)$ proving the implication $a) \Rightarrow b)$. Finally, assuming a), any point in Δ_0 is a point of growth of F^{-1}. Hence, $\mathrm{supp}(\mu) = \mathrm{clos}(\Delta_0)$ which implies d). In turn, d) implies that $\mu(x-\varepsilon, x+\varepsilon) > 0$ for every $x \in \Delta$ and $\varepsilon > 0$, and we arrive at a). The proof is complete. □

To conclude this paragraph, we illustrate the preceding results with some examples.

Example A.8. For the mass point $\mu = \delta_x$, $x \in \mathbb{R}$ (the degenerate case),

$$F^{-1}(t) \ = \ x, \qquad 0 < t < 1.$$

Hence $\mu^{-1} = 0$.

Example A.9. For the Bernoulli measure $\mu = p\delta_x + (1-p)\delta_y$, $x < y$, $0 < p < 1$,

$$F^{-1}(t) \ = \ \begin{cases} x, & \text{if} \quad 0 < t \leq p, \\ y, & \text{if} \quad p < t < 1. \end{cases}$$

Hence $\mu^{-1} = (y-x)\delta_p$ which is a multiple of the mass point.

Example A.10. Let λ denote the uniform measure on $(0,1)$. For a mixture of the Bernoulli and the uniform measure $\mu = \frac{1}{4}\,\delta_0 + \frac{1}{4}\,\delta_1 + \frac{1}{2}\,\lambda$, the inverse distribution function is continuous and is given by

$$F^{-1}(t) \ = \ \begin{cases} 0, & \text{if} \quad 0 < t \leq \frac{1}{4}, \\ 2t - \frac{1}{2}, & \text{if} \quad \frac{1}{4} < t \leq \frac{3}{4}, \\ 1, & \text{if} \quad \frac{1}{4} \leq t < 1. \end{cases}$$

In this case μ^{-1} represents a multiple of the uniform distribution on $[\frac{1}{4}, \frac{3}{4}]$.

Example A.11. Let X be a discrete random variable taking the values

$$\mathbb{P}\Big\{X = \frac{2i-1}{2^k}\Big\} = \frac{1}{3^k}, \qquad i = 1, 2, \dots, 2^{k-1}, \quad k = 1, 2, \dots$$

Since the points $\frac{2i-1}{2^k}$ form a dense subset of $[0,1]$, the distribution μ of X has the interval $[0,1]$ as the support. Hence, the inverse distribution function F^{-1} is continuous. In fact, F^{-1} coincides with the classical Cantor's stairs, so it is not absolutely continuous (in contrast with the previous example). Note that the function F^{-1} is constant on intervals having the total measure 1. This example also shows that the image $\mathrm{Im}(F)$ may have the continuum cardinality, even if F corresponds to a discrete distribution.

A.3 Modulus of continuity

Once the inverse function F^{-1} of a distribution function F is continuous, one can try to quantify this property by considering its modulus of continuity

$$\delta_{F^{-1}}(\varepsilon) = \sup\big\{\big|F^{-1}(t) - F^{-1}(s)\big| : |t-s| \le \varepsilon, \ \ t, s \in (0,1)\big\}, \qquad 0 < \varepsilon \le 1,$$

which is an optimal function δ such that

$$\big|F^{-1}(t) - F^{-1}(s)\big| \le \delta\big(|t-s|\big)$$

for all $t, s \in (0,1)$. However, as is made clear by the next statement, the study of moduli of continuity is restricted to the class of compactly supported measures.

Proposition A.12. *Let μ be a probability measure on $\mathbb{R}$ with distribution function F. The following properties are equivalent:*

a) $\delta_{F^{-1}}(\varepsilon) < \infty$ *for some* $\varepsilon \in (0,1)$*;*
b) $\delta_{F^{-1}}(\varepsilon) < \infty$ *for all* $\varepsilon \in (0,1)$*;*
c) μ *is compactly supported.*

If μ is supported on an interval of length ℓ, then $\delta_{F^{-1}} \le \ell$. Furthermore, F^{-1} is continuous if and only if $\delta_{F^{-1}}(0+) = 0$.

The statement is obvious. Let us only stress that μ is not compactly supported if and only if $F^{-1}(0) = -\infty$ or $F^{-1}(1) = \infty$. In the latter case, we have $\delta_{F^{-1}}(\varepsilon) \ge F^{-1}(1) - F^{-1}(1-\varepsilon) = \infty$ for every $\varepsilon \in (0,1)$.

If μ is compactly supported, and $[a,b]$ is the smallest segment where μ is supported, the behaviour of the modulus of continuity $\delta_{F^{-1}}$ near zero can be connected with the dual notion – an analogous "modulus of increase" of the distribution function F. This function may be defined as

$$\begin{aligned} \varepsilon_F(\delta) &= \inf\big\{F(y) - F(x) : \ y - x > \delta, \ \ x, y \in [a,b]\big\} \\ &= \inf\big\{F(y) - F(x-) : \ y - x \ge \delta, \ \ x, y \in [a,b]\big\} \end{aligned}$$

for every $0 < \delta < b - a$.

Proposition A.13. *Let F be the distribution function associated with a probability measure μ such that* $\mathrm{supp}(\mu) = [a,b]$*. Then for all $0 < \varepsilon < 1$ and $0 < \delta < b - a$,*

$$\delta_{F^{-1}}(\varepsilon) \le \delta \iff \varepsilon_F(\delta) \ge \varepsilon.$$

In particular,

$$\delta_{F^{-1}}(\varepsilon) \, = \, \inf\big\{\delta \in (0, b-a) : \, \varepsilon_F(\delta) \ge \varepsilon\big\}.$$

Proof. The support assumption means that the inverse function F^{-1} is continuous on $(0,1)$ according to Proposition A.7. We may assume that μ is non-degenerate so that $a < b$.

By definition, $\varepsilon_F(\delta) \ge \varepsilon$ means that $s - t \le \varepsilon \Rightarrow F^{-1}(s) - F^{-1}(t) \le \delta$ whenever $0 < t < s < 1$. Since F^{-1} is continuous, this implication may be rewritten as

$$s - t < \varepsilon \implies F^{-1}(s) - F^{-1}(t) \le \delta.$$

Moreover, this description will not change if we require additionally that $F^{-1}(t) < F^{-1}(u) < F^{-1}(s)$ for all $u \in (t,s)$. Indeed, otherwise, the interval (t,s) may be decreased without change of the value $F^{-1}(s) - F^{-1}(t)$. As explained in the proof of Proposition A.1, such a requirement is equivalent to the property that $[t,u) \cap \mathrm{Im}(F) \neq \emptyset$ and $[u,s) \cap \mathrm{Im}(F) \neq \emptyset$, for all $t < u < s$. But then $t = F(x)$ and $s = F(y-)$ for some $x < y$ in $[a,b]$, and hence $F^{-1}(t) = x$. In addition, by the left-continuity of F^{-1},

$$F^{-1}(s) \, = \, F^{-1}\big(F(y-)\big) \, = \, \lim_{z \uparrow y} F^{-1}\big(F(z)\big) \, = \, y$$

since $F^{-1}(F(z)) = z$ for all $z \in [a,b]$. Thus, the inequality $\varepsilon_F(\delta) \ge \varepsilon$ is reduced to the statement

$$F(y) - F(x) < \varepsilon \implies y - x \le \delta$$

(for $a \le x < y \le b$), or equivalently,

$$y - x > \delta \implies F(y) - F(x) \ge \varepsilon.$$

The latter amounts to $\varepsilon_F(\delta) \ge \varepsilon$ and hence Proposition A.13 is established. □

As in the preceding paragraph, we conclude with some examples illustrating these results.

Example A.14. Let a random variable X have a unimodal distribution, symmetric about the point $\frac{1}{2}$, with support $[0,1]$. The latter means that the distribution function F of X is convex on $[0,\frac{1}{2}]$ and concave on $[\frac{1}{2},1]$, with the symmetry property $F(1-x) = 1 - F(x)$ for all $0 \le x \le 1$ (for simplicity, let us exclude the case where F has a jump at the point $\frac{1}{2}$). Then the probabilities

$$\mathbb{P}\{y \le X \le x\} = F(y) - F(x), \qquad y - x \ge \delta, \ \ x, y \in [0,1],$$

are minimized for $y = 0$, $x = \delta$, so $\varepsilon_F(\delta) = F(\delta)$. Hence,

$$\delta_{F^{-1}}(\varepsilon) = F^{-1}(\varepsilon), \quad \text{for every} \;\; 0 < \varepsilon < 1.$$

This function is concave on the interval $[0, \frac{1}{2}]$ and convex on $[\frac{1}{2}, 1]$.

Example A.15. Let us return to the Example A.11 of the previous paragraph of a discrete random variable X with values in $(0, 1)$ for which F^{-1} represents the Cantor stairs. Under the constraints $y - x \geq \delta$, $x, y \in [0, 1]$, consider the probabilities

$$\mathbb{P}\{y \leq X \leq x\} \; = \sum_{y \leq \frac{2i-1}{2^k} \leq x} \frac{1}{3^k}$$

with allowed values $i = 1, 2, \dots, 2^{k-1}$, $k = 1, 2, \dots$ Here, for each fixed k the summation is performed over all positive integers i from the interval $[\frac{1+2^k y}{2}, \frac{1+2^k x}{2}]$. If this interval has length $\ell_k \geq 1$, it contains at least $[\ell_k] = [2^{k-1}(x - y)] \geq [2^{k-1}\delta]$ integers. Hence, if $2^{k-1}\delta \geq 1$,

$$\sum_{i:\; y \leq \frac{2i-1}{2^k} \leq x} \frac{1}{3^k} \; \geq \; \frac{[2^{k-1}\delta]}{3^k} \; \geq \; \frac{1}{2}\,\frac{2^{k-1}\delta}{3^k} \; = \; \frac{\delta}{4}\left(\frac{2}{3}\right)^k.$$

Putting k_0 to be the least integer $\geq 1 + \log_2 \frac{1}{\delta}$, this gives

$$\begin{aligned} \mathbb{P}\{y \leq X \leq x\} \;\; &\geq \;\; \sum_{k=k_0}^{\infty} \frac{\delta}{4}\left(\frac{2}{3}\right)^k \\ &= \;\; \frac{3\delta}{4}\left(\frac{2}{3}\right)^{k_0} \; \geq \; \frac{3\delta}{4}\left(\frac{2}{3}\right)^{2+\log_2(1/\delta)} \; = \; \frac{1}{3}\,\delta^{\frac{\log 3}{\log 2}}. \end{aligned}$$

As a result, $\varepsilon_F(\delta) \geq \frac{1}{3}\,\delta^{\frac{\log 3}{\log 2}}$ and thus $\delta_{F^{-1}}(\varepsilon) \leq (3\varepsilon)^{\frac{\log 2}{\log 3}}$.

A.4 Absolute continuity

In this paragraph, we study explicit characterizations of the (potential) absolute continuity property of the inverse distribution functions. This property will always be understood in the local sense. Namely, a function u defined on an interval (a, b), finite or not, will be called absolutely continuous if for all $a < a' < b' < b$ and $\varepsilon > 0$, there exists $\delta > 0$ such that, for any sequence of non-overlapping intervals $(a_i, b_i) \subset [a', b']$,

$$\sum_i (b_i - a_i) < \delta \;\; \Longrightarrow \;\; \sum_i \big|u(b_i) - u(a_i)\big| < \varepsilon.$$

Equivalently, for some locally integrable function v on (a, b), for all $a < t_0 < t_1 < b$,

$$u(t_1) - u(t_0) \; = \int_{t_0}^{t_1} v(t)\, dt.$$

The function v is uniquely determined up to a set of measure zero, is denoted by u', and is called the Radon-Nikodym derivative of u. As a possible variant, one may put

$$u'(t) \,=\, \limsup_{\varepsilon \to 0} \frac{u(t+\varepsilon) - u(t)}{\varepsilon}$$

or use lim inf instead.

For example, any locally Lipschitz function is absolutely continuous. Such functions have Radon-Nikodym derivatives that are bounded on proper subintervals of (a, b).

If an absolutely continuous function u is non-decreasing, the Radon-Nikodym derivative of u can always be chosen to be non-negative and a.e. finite. Any such u generates a non-negative Borel measure

$$\nu(A) \,=\, \int_A u'(t)\,dt, \qquad A \subset (a,b) \text{ Borel},$$

which is absolutely continuous with respect to the Lebesgue measure on (a, b) in the usual sense of Measure Theory. We do not require that ν be finite, but ν should be finite on compact subintervals of (a, b). Note that any non-negative absolutely continuous measure ν, which is finite on compact subintervals of (a, b), is generated by some non-decreasing absolutely continuous function u.

Here we consider such properties for the inverse function F^{-1} on $(a, b) = (0, 1)$. First let us state one immediate important consequence of the absolute continuity assumption on F^{-1}.

Proposition A.16. *If F is a non-degenerate distribution function such that F^{-1} is absolutely continuous, then the image set* $\mathrm{Im}(F)$ *has a positive Lebesgue measure.*

Proof. The non-degeneracy of F insures that F^{-1} generates a non-zero inverse measure μ^{-1} where μ is the probability measure associated with F. By the second assumption, the inverse measure is absolutely continuous with respect to the Lebesgue measure λ. So, $\lambda(\mathrm{Im}(F)) = 0$ would imply that $\mu^{-1}(\mathrm{Im}(F)) = 0$ which is impossible since μ^{-1} is supported on $\mathrm{Im}(F)$. □

Next, we turn to a full characterization of the absolute continuity of F^{-1} in terms of the distribution function F. In general, the measure μ generated by F admits a unique decomposition $\mu = \mu_0 + \mu_1 + \mu_2$, where μ_0 is a discrete measure, μ_1 is a singular continuous measure which is orthogonal to the Lebesgue measure λ on $\mathbb{R}$, and μ_2 is a measure which is absolutely continuous with respect to λ. They are respectively called the discrete component, the singular continuous component, and the absolutely continuous component of μ. Furthermore, by the Lebesgue differentiation theorem, the limit

$$f(x) = \lim_{y \to x,\ y \neq x} \frac{F(y) - F(x)}{y - x}$$

exists and is finite for almost all x, and represents the density of μ_2 with respect to Lebesgue measure. That is, $f(x) = \frac{d\mu_2(x)}{dx}$ in the sense of Measure Theory (Radon-Nikodym derivative).

Proposition A.17 (Characterization of absolute continuity of F^{-1})**.** *Let μ be a non-degenerate probability measure on $\mathbb{R}$ with distribution function F. The inverse function F^{-1} is absolutely continuous on $(0,1)$ if and only if μ is supported on an interval, finite or not, and the absolutely continuous component of μ has on that interval an a.e. positive density (with respect to the Lebesgue measure).*

Since the absolute continuity is stronger than just continuity, necessarily the support of μ should be a closed interval Δ, as already indicated in Proposition A.7. In that case, an additional requirement concerning the density which is needed for the absolute continuity of F^{-1} is equivalent to the property that the Lebesgue measure on Δ is absolutely continuous with respect to μ.

Proof. We may assume that F^{-1} is continuous on $(0,1)$, so that $\Delta = \mathrm{supp}(\mu)$ is an interval (not shrinking to a point by the non-degeneracy assumption). In particular, $F^{-1}(F(x)) = x$, for all $x \in \Delta$.

By definition, F^{-1} is absolutely continuous on $(0,1)$ if and only if, for all $0 < a < b < 1$ and $\varepsilon > 0$, there exists $\delta > 0$ such that, for any sequence of non-overlapping intervals $(a_i, b_i) \subset [a,b]$,

$$\sum_i (b_i - a_i) < \delta \quad \Longrightarrow \quad \sum_i \big(F^{-1}(b_i) - F^{-1}(a_i)\big) < \varepsilon.$$

Note that when a continuous function u is non-decreasing, in the definition of the absolute continuity one may require without loss of generality that $u(a_i) < u(t) < u(b_i)$ for $a_i < t < b_i$ (otherwise, the intervals (a_i, b_i) may be decreased without change of the value $u(b_i) - u(a_i)$). In the case $u = F^{-1}$, as was already explained in the proof of Proposition A.13, such a requirement implies that $a_i = F(x_i)$ and $b_i = F(y_i-)$ for some $x_i < y_i$ in Δ and, moreover, $F^{-1}(a_i) = x_i$ and $F^{-1}(y_i) = b_i$.

Thus, the definition of the absolute continuity of F^{-1} reduces to the statement that, for any finite interval $[x,y] \subset \Delta$ and any $\varepsilon > 0$, there exists $\delta > 0$ such that, for any sequence of non-overlapping intervals $(x_i, y_i) \subset [x,y]$,

$$\sum_i \big(F(y_i-) - F(x_i)\big) < \delta \quad \Longrightarrow \quad \sum_i (y_i - x_i) < \varepsilon.$$

Equivalently, if $\mu(A) < \delta$ then $\mathrm{mes}(A) < \varepsilon$, for the open set $A = \bigcup_i (x_i, y_i)$. Using regularity of measures, this implication can easily be extended to the class of all Borel subsets A of $[x,y]$. Therefore, the Lebesgue measure on $[x,y]$ is absolutely continuous with respect to the measure μ restricted to $[x,y]$. Extending $[x,y]$ to the whole support interval, we finally conclude that F^{-1} is absolutely continuous on $(0,1)$ if and only if the Lebesgue measure on Δ is absolutely continuous with respect to μ. Proposition A.17 is established. $\square$

In case μ is absolutely continuous, a more precise statement is available. When F has a positive continuous derivative f in a neighbourhood of $F^{-1}(t)$, then F^{-1} is differentiable at t and has derivative $(F^{-1})'(t) = 1/f(F^{-1}(t))$. In a more relaxed form, the following statement is valid.

Proposition A.18. *Let μ be a probability measure on $\mathbb{R}$ supported on an open interval, finite or not, where it has an a.e. positive density f, and let F be the distribution function of μ. Then, the inverse function F^{-1} is strictly increasing, absolutely continuous and, moreover, for all $0 < t_0 < t_1 < 1$,*

$$F^{-1}(t_1) - F^{-1}(t_0) \;=\; \int_{t_0}^{t_1} \frac{1}{f(F^{-1}(t))}\,dt.$$

In particular, a.e. F^{-1} is differentiable and has derivative $(F^{-1})'(t) = 1/f(F^{-1}(t))$.

Proof. Let μ be supported on $(a,b) \subset \mathbb{R}$. Since $f(x) > 0$ a.e. on this interval, F is continuous and strictly increasing on (a,b), and so is the inverse function $F^{-1} : (0,1) \to (a,b)$. Now, if a random variable U is uniformly distributed in $(0,1)$, the random variable $X = F^{-1}(U)$ has the distribution function F and the density f. Hence,

$$\begin{aligned}
\int_{t_0}^{t_1} \frac{1}{f(F^{-1}(t))}\,dt \;&=\; \mathbb{E}\bigg(\frac{1}{f(F^{-1}(U))}\;\mathbb{1}_{\{t_0<U<t_1\}}\bigg) \\
&=\; \mathbb{E}\bigg(\frac{1}{f(F^{-1}(U))}\;\mathbb{1}_{\{F^{-1}(t_0)<F^{-1}(U)<F^{-1}(t_1)\}}\bigg) \\
&=\; \mathbb{E}\bigg(\frac{1}{f(X)}\;\mathbb{1}_{\{F^{-1}(t_0)<X<F^{-1}(t_1)\}}\bigg) \\
&=\; \int_a^b \frac{1}{f(x)}\;\mathbb{1}_{\{F^{-1}(t_0)<x<F^{-1}(t_1)\}}\,f(x)\,dx \\
&=\; F^{-1}(t_1) - F^{-1}(t_0).
\end{aligned}$$

□

In the general case, when μ has a non-zero absolute continuous component, but also may have a non-zero discrete or continuous singular component, the Radon-Nikodym derivative of F^{-1} can be expressed in a similar way. Anyway, once F^{-1} is absolutely continuous, for this derivative we may take the function

$$(F^{-1})'(t) \;=\; \liminf_{s\to t,\; s>t} \frac{F^{-1}(s) - F^{-1}(t)}{s-t} \tag{A.4}$$

by the Lebesgue differentiation theorem.

If $t \notin \mathrm{Im}(F)$, then $F^{-1}(s) = F^{-1}(t)$ for some $s > t$, so $(F^{-1})'(t) = 0$. Otherwise, $t = F(x)$ for some x from the support Δ of μ, and $F^{-1}(t) = x$. Let us see that the above lim inf may be taken along the values $s = F(y-)$ with $y > x$. Indeed, in case s is not of the type $F(y-)$, we would have $F^{-1}(s') = F^{-1}(s)$ for some $t < s' < s$, and then

$$\frac{F^{-1}(s') - F^{-1}(t)}{s' - t} > \frac{F^{-1}(s) - F^{-1}(t)}{s - t}.$$

Hence, one may exclude such points s from the lim inf in (A.4). That is, it suffices to consider the values $s = F(y-)$ with $y \in \Delta$, $y > x$. In that case, by the left-continuity of the inverse function, necessarily $F^{-1}(s) = y$, so that

$$\begin{aligned} (F^{-1})'(t) &= \liminf_{y \to x,\ y > x} \frac{y - x}{F(y-) - F(x)} \\ &= \liminf_{y \to x,\ y > x} \frac{y - x}{F(y) - F(x)} \\ &= \frac{1}{\limsup_{y \to x,\ y > x} \frac{F(y) - F(x)}{y - x}}. \end{aligned}$$

The following proposition summarizes the conclusion at this point.

Proposition A.19. *Let F be a non-degenerate distribution function. If the inverse function F^{-1} is absolutely continuous, then it has the Radon-Nikodym derivative*

$$(F^{-1})'(t) = \frac{1}{f(F^{-1}(t))}, \qquad t \in \mathrm{Im}(F) \cap (0,1),$$

where

$$f(x) = \limsup_{y \to x,\ y > x} \frac{F(y) - F(x)}{y - x}.$$

Here, the function f represents a specific representative of the density of the absolutely continuous component of the measure μ generated by F. According to Proposition A.17, the assumption that F^{-1} is absolutely continuous is equivalent to saying that f is a.e. positive on the supporting interval for μ.

For $t \notin \mathrm{Im}(F)$, one may set $(F^{-1})'(t) = 0$. Recall that the measure μ^{-1} generated by F^{-1} is supported on the set $\mathrm{Im}(F) \cap (0,1)$, so it does not matter how to define the Radon-Nikodym derivative on its complement.

The preceding proposition emphasizes the concept of I-function associated to a distribution function F on the real line, extensively used throughout this investigation.

Definition A.20 (I-function)**.** The I-function of a distribution function F, whose inverse function F^{-1} is absolutely continuous on $(0,1)$, is defined as

$$I_F(t) = \frac{1}{(F^{-1})'(t)}, \qquad 0 < t < 1.$$

This function is well-defined a.e., and then $(F^{-1})'$ denotes the corresponding Radon-Nikodym derivative.

In particular, if an absolutely continuous probability measure on μ on $\mathbb{R}$ with distribution function F is supported on an open interval and has there an a.e. positive density f, then the I-function is well-defined and is given, according to Proposition A.18, by

$$I_F(t) \,=\, f\big(F^{-1}(t)\big) \qquad (a.e.)$$

These I-functions are somehow related to isoperimetric profiles, and indeed do represent isoperimetric properties for log-concave distributions. For example, if μ is the standard normal law on $\mathbb{R}$, the associated I-function is $\varphi(\Phi^{-1})$ where Φ is the usual notation of the distribution function of μ and φ is its density. If $d\mu(x) = \frac{1}{2}\, e^{-|x|}dx$ on $\mathbb{R}$, then $I(t) = \min\{t, 1-t\}$, $0 < t < 1$.

According to Proposition A.19, the formula $I_F(t) \,=\, f\big(F^{-1}(t)\big)$ remains to hold in the general case of the absolutely continuous inverse function F^{-1} – however for a specific representative of the density of the absolutely continuous component of μ.

To see that in general f cannot be chosen in an arbitrary way, let μ be supported on an open interval and such that its absolutely continuous component has on that interval an a.e. positive density f (so that the I-function is well-defined). Assume that μ is continuous and has a non-zero singular continuous component μ_1. Let $\tilde f$ be another representative of the density such that $f(x) \neq \tilde f(x)$ on a set A of Lebesgue measure zero with $\mu_1(A) > 0$. Then

$$f\big(F^{-1}(t)\big) \,\neq\, \tilde f\big(F^{-1}(t)\big)$$

on the set $B = \{t \in (0,1) : F^{-1}(t) \in A\}$. But, by Proposition A.1, the Lebesgue measure of B is equal to $\mu(A) = \mu_1(A)$. Thus, $f\big(F^{-1}(t)\big) \neq \tilde f\big(F^{-1}(t)\big)$ on a set of positive Lebesgue measure.

Nevertheless, a number of important relations and integrals containing the I-functions may be expressed explicitely in terms of f and do not depend on how we choose the density f. Some of them are considered in the next section.

A.5 Integrals containing the derivative of F^{-1}

The study of rates for Kantorovich distances for empirical measures requires representation formulas for integrals containing the Radon-Nikodym derivative of F^{-1}. The following general formula involves a weight which will take concrete forms in specific examples.

Proposition A.21. *Let F be a distribution function such that the inverse function F^{-1} is absolutely continuous and has a Radon-Nikodym derivative $(F^{-1})'$. For any Borel measurable function $w : (0,1) \to [0,\infty)$, and any $p \geq 1$,*

$$\int_0^1 \big|(F^{-1})'(t)\big|^p\, w(t)\, dt \,=\, \int_{\{0<F(x)<1\}} \frac{w(F(x))}{f(x)^{p-1}}\, dx, \tag{A.5}$$

where f is a density of the absolutely continuous component of F.

Proof. The formula (A.5) is a variant of integration by using a change of variable. If F is degenerate (a mass point), both integrals are vanishing, so let F be non-degenerate.

As we know, the absolute continuity of F^{-1} is equivalent to the property that the probability measure μ with distribution function F is supported on a non-degenerate interval Δ of the real line (open or closed, finite or not) and, moreover an absolutely continuous component of μ should have an a.e. positive density f on Δ. Note that, up to the endpoints, Δ is the same as $\Delta_0 = \{0 < F(x) < 1\}$. Furthermore, the integral on the right-hand side in (A.5) will not change when changing f on a set of null Lebesgue measure. So, we may take for f the one defined in Proposition A.19, i.e.

$$f(x) \,=\, \limsup_{y \rightarrow x,\ y > x} \frac{F(y) - F(x)}{y - x}\,.$$

Now, by Proposition A.4, the function F pushes forward the Lebesgue measure on Δ_0 to the the inverse measure μ^{-1}. Equivalently,

$$\int_{\Delta_0} R\big(F(x)\big)dx \,=\, \int_0^1 R(t)\, d\mu^{-1}(t)$$

for any R such that at least one of the above integrals is defined in the Lebesgue sense. But μ^{-1} is supported on the image set $\Delta' \equiv \mathrm{Im}(F) \cap (0,1)$, which has a positive Lebesgue measure, cf. Proposition A.16. Moreover, by the assumption, μ^{-1} has the density $(F^{-1})'$, hence

$$\int_{\Delta_0} R\big(F(x)\big)dx \,=\, \int_{\Delta'} R(t)\,(F^{-1})'(t)\,dt.$$

Applying Proposition A.19, we get

$$\int_{\Delta_0} R\big(F(x)\big)dx \,=\, \int_{\Delta'} \frac{R(t)}{f(F^{-1}(t))}\,dt.$$

In particular, since $F^{-1}(F(x)) = x$ on Δ,

$$\begin{aligned} \int_{\Delta_0} \frac{w(F(x))}{f(x)^{p-1}}\,dx \ &= \ \int_{\Delta_0} \frac{w(F(x))}{f(F^{-1}(F(x))^{p-1}}\,dx \\ &= \ \int_{\Delta'} \frac{w(t)}{f(F^{-1}(t))^p}\,dt \\ &= \ \int_{\Delta'} \big|(F^{-1})'(t)\big|^p\, w(t)\,dt. \end{aligned}$$

Finally, recall that $(F^{-1})' = 0$ a.e. outside $\mathrm{Im}(F)$. Therefore, without any change, the last integral may be extended to the whole interval $(0,1)$, and it will not depend on the choice of the Radon-Nikodym derivate for F^{-1}. The proof is complete. □

Proposition A.21 will be used with the weight functions $w(t) = (t(1-t))^{p/2}$.

Corollary A.22. *Let F be a distribution function such that the inverse function F^{-1} is absolutely continuous and has a Radon-Nikodym derivative $(F^{-1})'$. Then, for all $p \geq 1$,*

$$\int_0^1 \big|(F^{-1})'(t)\big|^p \, \big(t(1-t)\big)^{p/2} \, dt \, = \, \int_{0<F(x)<1} \frac{[F(x)(1-F(x))]^{p/2}}{f(x)^{p-1}} \, dx,$$

where f is a density of the absolutely continuous component of F. In particular,

$$\int_0^1 \big((F^{-1})'(t)\big)^2 \, t(1-t) \, dt \, = \, \int_{0<F(x)<1} \frac{F(x)(1-F(x))}{f(x)} \, dx.$$

Another case of interest is the constant weight function $w(t) = 1$. Taking $p = 1$, we arrive at the description of the length of the supporting interval.

Corollary A.23. *Let F be a distribution function such that F^{-1} is absolutely continuous and has a Radon-Nikodym derivative $(F^{-1})'$. Then, the integral*

$$\int_0^1 \big|(F^{-1})'(t)\big| \, dt$$

represents the length of the smallest interval supporting the probability measure with the distribution function F.

A.6 Monotone Lipschitz transforms

In the last section of this appendix, we examine Lipschitz transformations of measures and how they translate on the associated distribution functions.

If ν is a given non-atomic probability measure on the Borel sets of $\mathbb{R}$ with a (continuous) distribution function G, any other probability measure μ on the real line with the distribution function F can be obtained as a monotone transform of ν, i.e. as the distribution of a monotone map T under ν. One then says that T pushes forward ν to μ, or G to F, and writes in symbols $\mu = \nu T^{-1} = T(\nu)$.

A canonical map $T : \mathbb{R} \rightarrow \mathbb{R}$ is given by

$$T(x) \, = \, F^{-1}\big(G(x)\big), \qquad x \in \mathbb{R}.$$

Indeed, since G is continuous, it pushes forward ν to the uniform measure on $(0, 1)$, while the inverse function F^{-1} pushes forward the uniform measure to μ.

In a number of questions, it is desirable to know whether or not T is Lipschitz, and how to estimate its Lipschitz semi-norm

$$\|T\|_{\mathrm{Lip}} \, = \, \sup_{x<y} \frac{F(y) - F(x)}{y - x}.$$

A simple characterization can be given by comparing the associated I-functions

$$I_F(t) = \frac{1}{(F^{-1})'(t)}, \qquad I_G(t) = \frac{1}{(G^{-1})'(t)}, \qquad 0 < t < 1,$$

from Definition A.20.

Proposition A.24. *Suppose that G is continuous and G^{-1} is absolutely continuous. Given a distribution function F, let T be the canonical map pushing forward G to F. The map T has a finite Lipschitz semi-norm with $\|T\|_{\rm Lip} \leq \frac{1}{c}$ for some $c > 0$ if and only if the inverse function F^{-1} is absolutely continuous, and*

$$I_F(t) \geq c\, I_G(t) \qquad a.e.$$

A characterization of the absolute continuity of the inverse functions is given in Proposition A.17, while Proposition A.19 describes the associated I-functions. In many practical situations, ν is absolutely continuous, is supported on some interval, and has there an a.e. positive density g. In that case,

$$I_G(t) = g\big(G^{-1}(t)\big), \qquad 0 < t < 1.$$

For example, the two-sided exponential distribution ν with density $g(x) = \frac{1}{2}\, e^{-|x|}$, $x \in \mathbb{R}$, has the the associated function $I_G(t) = \min\{t, 1-t\}$. Hence, the inequality of the form

$$I_F(t) \geq c\, \min\{t, 1-t\} \qquad \text{a.e.} \tag{A.6}$$

means that the measure μ is obtained from ν as a Lipschitz monotone transform with $\|T\|_{\rm Lip} \leq \frac{1}{c}$.

Proof of Proposition A.24. (Sufficiency) Assume that F^{-1} is absolutely continuous. By the assumption, the Radon-Nikodym derivative of the inverse function admits a pointwise upper-bound

$$(F^{-1})' \leq \frac{1}{c}\, (G^{-1})' \qquad a.e.$$

Hence, integrating this inequality, for all $0 < t < s < 1$,

$$F^{-1}(s) - F^{-1}(t) \leq \frac{1}{c}\, \big(G^{-1}(s) - G^{-1}(t)\big). \tag{A.7}$$

Note that ν must have an interval Δ as its support (Proposition A.17), and therefore $G^{-1}(G(x)) = x$ for all $x \in \Delta$. Changing the variables $t = G(x)$, $s = G(y)$ with $x, y \in \Delta$, $x < y$, this gives $T(y) - T(x) \leq \frac{1}{c}\,(y - x)$, which means that $\|T\|_{\rm Lip} \leq \frac{1}{c}$.

(Necessity) We can start with the above relation (A.7) for the inverse functions, which immediately implies that F^{-1} is absolutely continuous (since G^{-1} is). Moreover, dividing this inequality by $s - t$ and letting $s \downarrow t$, we obtain, by the Lebesgue differentiation theorem, that $(F^{-1})' \leq \frac{1}{c}\,(G^{-1})'$ a.e. $\square$

If ν is the uniform measure on $(0, 1)$, then $I_G(t) \equiv 1$, and Proposition A.24 may be formulated in a different manner.

Proposition A.25 (Lipschitz inverse distribution function)**.** *Let F be the distribution function of a non-degenerate probability measure μ on $\mathbb{R}$. The inverse function F^{-1} has a finite Lipschitz semi-norm $\|F^{-1}\|_{\rm Lip}$ on $(0,1)$ if and only if μ is supported on a finite interval Δ, and the absolutely continuous component of μ has a density f which is separated from zero on Δ. In this case,*

$$\|F^{-1}\|_{\rm Lip} \, = \, \frac{1}{\operatorname{ess\,inf}_{x \in \Delta} f(x)}\,.$$

Proof. In terms of the modulus of continuity, the Lipschitz property is equivalent to the relation

$$\delta_{F^{-1}}(\varepsilon) \, \leq \, C\varepsilon, \qquad 0 < \varepsilon < 1,$$

where the optimal value of C represents the Lipschitz semi-norm $\|F^{-1}\|_{\rm Lip}$. In this case, the support Δ of μ has to be a finite closed interval, finite or not (Proposition A.12). Moreover, since F^{-1} is absolutely continuous, an absolutely continuous component of μ has a density f which is a.e. positive on Δ (Proposition A.17).

Assuming that $\delta = C\varepsilon < 1$, the relation $\delta_{F^{-1}}(\varepsilon) \leq C\varepsilon$ is equivalent to $\varepsilon_F(\delta) \geq \varepsilon$, where ε_F is the modulus of increase of F (by Proposition A.13). That is,

$$\varepsilon_F(\delta) \, = \, \inf\big\{F(y) - F(x): \, y - x > \delta, \ \ x,y \in \Delta\big\} \, \geq \, \frac{\delta}{C}\,.$$

But, by the Lebesgue differentiation theorem, $\frac{1}{\delta}\,(F(x+\delta) - F(x)) \rightarrow f(x)$ as $\delta \downarrow 0$, for almost all $x \in \Delta$. Hence, $\|F^{-1}\|_{\rm Lip} \leq C$ implies $f(x) \geq 1/C$ a.e. on Δ. Conversely, the latter easily implies $\varepsilon_F(\delta) \geq \delta/C$ and hence $\|F^{-1}\|_{\rm Lip} \leq C$. The proof is complete. $\square$

B Beta distributions

This appendix is devoted to special properties of the beta distributions needed in the analysis of formulas such as the one given in Theorem 4.6. As we believe, many of such properties are of independent interest. Some of them can be studied by using only the fact that all beta distributions with parameters $\alpha, \beta \geq 1$ are log-concave. Therefore, it is natural first to collect together a number of relevant general results about such measures. We next investigate Poincaré-type inequalities for beta distributions, both in the standard L^2-norm but also for L^p-norms, $p \geq 1$, including the $p = 1$ case corresponding to Cheeger-type inequalities. Mean square beta distributions are analyzed in the subsequent paragraph. Quite a bit of work is then devoted, in the last part, to refined lower-bounds and lower integral bounds for beta densities. While the topic is classical, most of results developed here seem to be new.

The beta distribution is denoted by $B_{\alpha,\beta}$, $\alpha, \beta > 0$, which is always treated as a probability measure on $(0, 1)$ with density

$$\frac{dB_{\alpha,\beta}(x)}{dx} = \frac{1}{B(\alpha,\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}, \qquad 0 < x < 1, \tag{B.1}$$

where

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, dx = \frac{\Gamma(\alpha)\Gamma(\alpha)}{\Gamma(\alpha+\beta)}$$

is the normalizing factor (the classical beta function in two variables). If a random variable X has the distribution $B_{\alpha,\beta}$, its moments are given by

$$\mathbb{E}(X^p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+p)}\,\frac{\Gamma(\alpha+p)}{\Gamma(\alpha)}, \qquad p \geq 0. \tag{B.2}$$

In particular,

$$\mathbb{E}(X) = \frac{\alpha}{\alpha+\beta}, \qquad \mathbb{E}(X^2) = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$$

and

$$\mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2\,(\alpha+\beta+1)}\,.$$

B.1 Log-concave measures on the real line

In this sub-Section, we first recall some general facts on log-concave measures on $\mathbb{R}^k$, mostly specializing to the one-dimensional case. We refer to the classical papers by Borell [Bor2] and Brascamp and Lieb [B-L] as general references on the subject.

A probability measure μ on the Euclidean space $\mathbb{R}^k$ is called log-concave, if it satisfies a Brunn-Minkowski-type inequality

$$\mu\big((1-t)A + tB\big) \geq \mu(A)^{1-t}\mu(B)^t, \qquad 0 < t < 1, \tag{B.3}$$

in the class of all non-empty Borel subsets A and B of $\mathbb{R}^k$, where

$$(1-t)A + tB \,=\, \big\{(1-t)x + ty : x \in A,\ y \in B\big\}$$

denotes the usual Minkowski average. Equivalently (cf. [Bor2]), μ must be supported on a convex set V in $\mathbb{R}^n$, where it has a log-concave density with respect to the Lebesgue measure of the same dimension as V. Such characterization can easily be obtained, by applying the Prékopa-Leindler theorem, cf. [B-L], [L2-3].

By the dilation-type Lemma 3.1 of Borell [Bor2], log-concave measures have moments of all orders. In a more precise sharp form, letting X be a random vector in $\mathbb{R}^k$ with log-concave distribution μ, and given any norm $\|\cdot\|$ on $\mathbb{R}^k$, for all $t \geq 1$ and $x > 0$,

$$\mathbb{P}\big\{\|X\| \geq tx\big\} \,\leq\, \mathbb{P}\big\{\|X\| \geq x\big\}^{(t+1)/2}.$$

When the norm is Euclidean, this inequality was first obtained by Lovász and Simonovits [L-S], and later Guédon [Gu] extended it to arbitrary norms. This results entails a multidimensional Khinchine-type inequality

$$\mathbb{E}\big(\|X\|^p\big)^{1/p} \,\leq\, Cp\, \mathbb{E}\big(\|X\|\big) \tag{B.4}$$

where $C > 0$ is an absolute constant. Actually, it remains to hold for an arbitrary semi-norm.

On the real line, the definition of log-concave measure reduces to the requirement that either μ is a mass point, or it is supported on some interval $(a,b) \subset \mathbb{R}$, finite or not, where it has a positive density f such that the function $\log f$ is concave on (a,b). If F is the distribution function associated with μ, then necessarily the associated I-function $I(t) = f(F^{-1}(t))$ from Definition A.20 is concave on $(0,1)$. This already shows that all log-concave measures on $\mathbb{R}$ are unimodal. In fact, the property that I is concave characterizes the class of all absolutely continuous log-concave probability measures on the real line. As a consequence of the Brunn-Minkowski-type inequality (applied to half-axes), one immediately obtains that the functions F and $1-F$ are also log-concave on the supporting interval (a,b).

According to (B.4), any real random variable X having a log-concave distribution μ has finite moments of any order, and moreover an exponential moment $\mathbb{E}(e^{\varepsilon|X|})$ is finite for some $\varepsilon > 0$. The global behavior of μ is mostly determined or can be controled by the two parameters – the expectation $\mathbb{E}(X)$ and the variance $\mathrm{Var}(X)$. Many other quantities or characteristics of μ are often related to these parameters. The following statement is one such example.

Proposition B.1. *Let $X > 0$ be a random variable with median m, having a log-concave distribution. Then*

$$\mathbb{E}(X) \,\leq\, \frac{m}{\log 2}\,. \tag{B.5}$$

The inverse bound $m \leq 2\,\mathbb{E}(X)$ obviously holds without the log-concavity assumption. With an existing constant, the inequality $\mathbb{E}X \leq Cm$ is well-known and holds in much greater generality in the form of the bound $\mathbb{E}(\|X\|) \leq Cm(\|X\|)$ (where X is a random vector in $\mathbb{R}^k$ having an arbitrary log-concave distribution and $\|\cdot\|$ is an arbitrary norm) as a consequence of Borell's result. In the current formulation, the constant $\frac{1}{\log 2}$ is sharp and is attained for the standard exponential distribution ν with density e^{-x}, $x > 0$.

Proof. It may be assumed that $a = \operatorname{ess\,inf} X = 0$ since the stated inequality (B.5) being written for $X - a$ is getting stronger. So, let the distribution of X be supported on an interval $(0, b)$, $0 < b \leq \infty$, with the distribution function F. Using homogeneity, assume that $m = \log 2$, so that $F(\log 2) = \frac{1}{2}$.

We only use the property that the function $1 - F(x)$ is log-concave on $(0, b)$. Hence, the function $T(x) = -\log(1 - F(x))$, which pushes forward ν to the distribution of X, is concave. Let l be a linear function whose graph is tangent to the graph of T at the point $\log 2$. That is, $l(\log 2) = T(\log 2) = \log 2$ and, for some $c \in \mathbb{R}$,

$$l(x) \;=\; l(\log 2) + c(x - \log 2) \;=\; \log 2 + c(x - \log 2).$$

Since $l \geq T$ and $T \geq 0$ with $T(0+) = 0$, necessarily $c \leq 1$. Hence, if ξ is distributed according to ν, we obtain that

$$\mathbb{E}(X) \;=\; \mathbb{E}\big(T(\xi)\big) \leq \mathbb{E}\big(l(\xi)\big) \;=\; \log 2 + c(1 - \log 2) \;\leq\; 1.$$

□

Proposition B.2 (Variance of a log-concave distribution)**.** *Let X be a random variable with median m, having a log-concave distribution μ with (log-concave) density f. Then*

$$\frac{1}{12\,\mathrm{Var}(X)} \;\leq\; f(m)^2 \;\leq\; \frac{1}{2\,\mathrm{Var}(X)}\,. \tag{B.6}$$

Furthermore,

$$\frac{1}{12\,\mathrm{Var}(X)} \;\leq\; \sup_{x \in \mathbb{R}} f(x)^2 \;\leq\; \frac{1}{\mathrm{Var}(X)}\,, \tag{B.7}$$

$$\frac{1}{3e^2\,\mathrm{Var}(X)} \;\leq\; f\big(\mathbb{E}(X)\big)^2 \leq \frac{1}{\mathrm{Var}(X)}\,. \tag{B.8}$$

For the proof of (B.6), we refer to [Bob4], Proposition 4.1. The left-hand side of (B.7) is weaker, but is still sharp, since it is attained for the uniform distribution in the unit interval. Here, the constant $\frac{1}{12}$ actually serves for the class of all probability densities on the real line. This fact was already mentioned without giving details in 1960's by Statulevičius [Sta] and later was emphasized by Hensley [He], who also considered upper estimates in the class of log-concave densities that are symmetric about the origin (cf. also [Ba]). As for the right-hand sides of (B.6) and (B.7), they are also sharp and both are attained for the one-sided exponential distribution.

The upper-bound on the maximum of the density in (B.7) is due to Fradelizi [F], who proved it for marginals of convex bodies in isotropic position. A simple proof of this bound in the general log-concave case may be found in [B-C]. It is based on the concavity of the function $I(t) = f(F^{-1}(t))$. Note that this property implies that $\sup_x f(x) \le 2f(m)$.

The upper-bound in (B.8) follows from a similar bound for the maximum of the density. To comment on the lower-bound, first let us mention that the value $h = 2f(m)$ is also known as an optimal constant in the Cheeger-type analytic inequality

$$h \int_{-\infty}^{\infty} \big|u(x) - m(u)\big|\, d\mu(x) \ \le \ \int_{-\infty}^{\infty} \big|u'(x)\big|\, d\mu(x), \tag{B.9}$$

holding in the class of all absolutely continuous functions u on the real line with median $m(u)$ under μ (cf. [Bob4]). Being applied to the indicator functions of half-axes $(-\infty, x]$ (in the approximate sense), it yields the relation

$$h \min\big\{F(x), (1 - F(x)\big\} \ \le \ f(x), \qquad a < x < b,$$

where F is the distribution function of X and (a, b) is the supporting interval of f. In turn, this relation implies the Cheeger-type analytic inequality. Therefore, by the upper-bound in (B.6),

$$f(x) \ge \frac{1}{\sqrt{3\,\mathrm{Var}(X)}}\ \min\big\{F(x), (1 - F(x)\big\}$$

for all $x \in (a, b)$. On the other hand, there is a two-sided bound

$$\frac{1}{e} \ \le \ \mathbb{P}\big\{X \le \mathbb{E}(X)\big\} \ \le \ 1 - \frac{1}{e}$$

(cf. [Bob5], Lemma 3.3). So, taking $x = \mathbb{E}(X)$, we arrive at the lower-bound in (B.8).

Finally, it is worth mentioning that together with the lower-bound in (B.6), the Cheeger-type inequality (B.9) yields the Poincaré-type inequality

$$\mathrm{Var}_\mu(u) \ \le \ 12\,\mathrm{Var}(X) \int_{-\infty}^{\infty} u'(x)^2\, d\mu(x) \tag{B.10}$$

for all absolutely continuous functions u on the real line (see e.g. [Bob4], Corollary 4.3).

B.2 Log-concave measures of high order

A number of results about general log-concave measures can further be sharpened for certain subclasses, and here we discuss one of them.

A random variable $X > 0$ is said to have a log-concave distribution of order $\alpha \geq 1$ if it has a density of the form

$$f(x) \,=\, x^{\alpha-1}\rho(x),$$

for some log-concave function ρ on $(0,\infty)$. For example, the standard Gamma-distribution with $\alpha \geq 1$ degrees of freedom, which has the density

$$f(x) \,=\, \frac{1}{\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-x}, \qquad x > 0,$$

is log-concave of order α. The beta distribution $B_{\alpha,\beta}$ is also log-concave of order α for all $\alpha \geq 1$ and $\beta \geq 1$.

When α is large, such probability measures are more concentrated about a point in comparison with general log-concave measures. This can already be seen from the following proposition.

Proposition B.3. *If X has a log-concave distribution of order $\alpha \geq 1$, then*

$$\mathrm{Var}(X) \,\leq\, \frac{1}{\alpha}\,\big(\mathbb{E}(X)\big)^2.$$

Here, equality is attained for the standard Gamma-distribution with α degrees of freedom. This assertion follows from a result of Borell [Bor1] about reverse Lyapunov inequalities on convex bodies (and from the results of [B-M-P] in case α is integer). For more details and some interesting concentration applications, we refer to [Bob6-7] and [B-Ma]. As shown in [Bob7], Proposition B.3 may be used to get a distributional self-improvement of Gaussian type. (A similar statement with integer values of α – about deviations of X from the maximum of the density – was also studied by Klartag [Kl]).

Proposition B.4 (Concentration inequalities)**.** *If X has a log-concave distribution of order $\alpha \geq 1$, and Y is an independent copy of X, then, for all $0 \leq r \leq 1$,*

$$\mathbb{P}\big\{\big|X - \mathbb{E}(X)\big| \geq r\,\mathbb{E}(X)\big\} \,\leq\, 2\,e^{-\alpha r^2/4}$$

and

$$\mathbb{P}\big\{|X - Y| \geq r\,\mathbb{E}(X)\big\} \,\leq\, 2\,e^{-\alpha r^2/8}.$$

We briefly recall the argument of proof. Without loss of generality, it may be assumed that $\mathbb{E}(X) = 1$ and that the density $f(x) = x^{\alpha-1}\rho(x)$ is compactly supported. For $t \in \mathbb{R}$, consider random variables X_t with densities

$$f_t(x) \,=\, \frac{x^{\alpha-1}e^{tx}\rho(x)}{\int_0^\infty x^{\alpha-1}e^{tx}\rho(x)\,dx}\,, \qquad x > 0.$$

Since all X_t have log-concave distributions of the same order α, by Proposition B.3, $\mathrm{Var}(X_t) \,\leq\, \frac{1}{\alpha}\,\mathbb{E}(X_t)^2$ or, in terms of X,

$$\mathbb{E}(X^2e^{tX})\,\mathbb{E}(e^{tX}) - \big[\mathbb{E}(Xe^{tX})\big]^2 \,\leq\, \frac{1}{\alpha}\,\big[\mathbb{E}(Xe^{tX})\big]^2. \tag{B.11}$$

The function $u(t) = \log \mathbb{E}(e^{tX})$, $t \in \mathbb{R}$, is convex, and the function

$$v(t) \,=\, u'(t) \,=\, \frac{\mathbb{E}(Xe^{tX})}{\mathbb{E}(e^{tX})}$$

is strictly positive and has a positive derivative on the whole real line. One may rewrite (B.3) in terms of v as $v'(t) \le \frac{1}{\alpha}\, v(t)^2$. Equivalently $(-\frac{1}{v(t)})' \le \frac{1}{\alpha}$, so that, after integration and using the assumption $v(0) = \mathbb{E}(X) = 1$,

$$\left|\frac{v(t)-1}{v(t)}\right| \,=\, \left|\frac{1}{v(t)} - \frac{1}{v(0)}\right| \,\le\, \frac{|t|}{\alpha}$$

for all $t \in \mathbb{R}$. Set $u_0(t) = \log \mathbb{E}\, e^{t(X-\mathbb{E}X)} = u(t) - t$ and $v_0(t) = u_0'(t) = v(t) - 1$, so that

$$u_0(t) \,\ge\, 0, \qquad u_0(0) \,=\, v_0(0) \,=\, 0, \qquad v_0(t) \,>\, -1$$

for all $t \in \mathbb{R}$. Hence

$$\frac{|v_0(t)|}{1+|v_0(t)|} \,\le\, \frac{|v_0(t)|}{1+v_0(t)} \,\le\, \frac{|t|}{\alpha}\,.$$

In particular, for $|t| \le \frac{\alpha}{2}$, necessarily $|v_0(t)| \le 1$ and thus $|v_0(t)| \le 2\,\frac{|v_0(t)|}{1+v_0(t)} \le \frac{2\,|t|}{\alpha}$. Integrating from 0 to t, we obtain that

$$\big|u_0(t)\big| \,\le\, \left|\int_0^t \big|v_0(s)\big|\, ds\right| \,\le\, \frac{t^2}{\alpha}\,.$$

Since $\mathbb{E}(X) = 1$, subgaussian bounds on the Laplace transform follow in the form

$$\mathbb{E}\big(e^{t(X-\mathbb{E}X)}\big) \,\le\, e^{t^2/\alpha}, \qquad \mathbb{E}\big(e^{t(X-Y)}\big) \,\le\, e^{2t^2/\alpha}$$

in the interval $|t| \le \frac{\alpha}{2}$. Finally, an application of Chebyshev's inequality easily yields, for $0 \le r \le 1$,

$$\mathbb{P}\big\{X - \mathbb{E}(X) \ge r\,\mathbb{E}(X)\big\} \,\le\, e^{-\alpha r^2/4}\,, \qquad \mathbb{P}\big\{X - Y \ge r\,\mathbb{E}(X)\big\} \,\le\, e^{-\alpha r^2/8}\,.$$

Similar bounds hold for the left deviations, concluding therefore the proof of the proposition.

B.3 Spectral gap

After these preliminaries, we next turn to the beta distributions themselves. Let us start with a spectral gap or Poincaré-type inequality. It will however differ in the right-hand side from the usual Poincaré-type inequalities (B.10) by a specific weight adapted to the beta distributions.

Recall the beta distribution $B_{\alpha,\beta}$, $\alpha, \beta > 0$, from (B.1). We denote by $\mathrm{Var}_{B_{\alpha,\beta}}(u)$ the variance of a function u (with finite second moment) under this measure.

Proposition B.5 (Poincaré inequality for $B_{\alpha,\beta}$). *For any absolutely continuous function u on $(0,1)$ with finite second moment under $B_{\alpha,\beta}$,*

$$\mathrm{Var}_{B_{\alpha,\beta}}(u) \,\leq\, \frac{1}{\alpha+\beta} \int_0^1 x(1-x)\, u'(x)^2 \, dB_{\alpha,\beta}(x), \tag{B.12}$$

where u' denotes the (Radon-Nikodym) derivative of u.

For example, for the uniform distribution (when $\alpha = \beta = 1$), the inequality is translated equivalently as

$$\int_0^1 \int_0^1 (u(x) - u(y))^2 \, dxdy \,\leq\, \int_0^1 x(1-x)\, u'(x)^2 \, dx.$$

In this inequality, as well as in the general one (B.12), the constant $\frac{1}{\alpha+\beta}$ is optimal and is attained for linear functions. Indeed, for $u(x) = x$, i.e. for a random variable X distributed according to $B_{\alpha,\beta}$, we have

$$\mathrm{Var}_{B_{\alpha,\beta}}(u) \,=\, \mathrm{Var}(X) \,=\, \frac{\alpha\beta}{(\alpha+\beta)^2\,(\alpha+\beta+1)}\,.$$

On the other hand,

$$\begin{aligned}
\int_0^1 x(1-x)\, u'(x)^2 \, dB_{\alpha,\beta}(x) \;&=\; \mathbb{E}\big(X(1-X)\big) \\
&=\; \frac{\alpha}{\alpha+\beta} - \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} \\
&=\; \frac{\alpha\beta}{(\alpha+\beta)\,(\alpha+\beta+1)}
\end{aligned}$$

so that indeed equality holds in Proposition B.5 for $u(x) = x$.

Let us also mention that in order to prove Poincaré-type inequalities on the real line in the class of all absolutely continuous functions, such as in Proposition B.5, it suffices to establish them for the class of smooth functions.

Proof. It is based on a standard expansion in orthogonal polynomials (cf. e.g. [B-G-L]). The normalized Jacobi polynomials $J^{\ell}_{\alpha,\beta}$ with $\ell = 0, 1, 2, \dots$ form an orthonormal basis in $L^2((0,1), B_{\alpha,\beta})$. Moreover, they are eigenvectors of the second-order linear differential operator

$$\mathrm{L}u(x) \,=\, x(1-x)\, u''(x) + \big[\alpha - (\alpha+\beta)x\big]\, u'(x)$$

with eigenvalues

$$\lambda^{\ell}_{\alpha,\beta} \,=\, \ell(\ell+\alpha+\beta-1), \qquad \ell = 0, 1, 2, \dots$$

Furthermore, by integration by parts, for any smooth function u on $(0,1)$,

$$\int_0^1 u\,(-\mathrm{L}u)\,dB_{\alpha,\beta} \,=\, \int_0^1 x(1-x)\,u'(x)^2\,dB_{\alpha,\beta}(x).$$

Note that all eigenvalues are non-negative, and the smallest one $\lambda^0_{\alpha,\beta} = 0$ corresponds to the constant eigenfunction $J^0_{\alpha,\beta}(x) = 1$. The next eigenvalue $\lambda^1_{\alpha,\beta} = \alpha + \beta$ corresponds to the linear eigenfunction $J^1_{\alpha,\beta}(x) = \mathrm{const} \cdot (x - \frac{\alpha}{\alpha+\beta})$.

Now, if u is expanded into the Fourier series $u = \sum_{\ell=0}^\infty a_\ell J^\ell_{\alpha,\beta}$, we have

$$\mathrm{Var}_{B_{\alpha,\beta}}(u) \,=\, \sum_{\ell=1}^\infty a_\ell^2$$

and

$$\int_0^1 x(1-x)\,u'(x)^2\,dB_{\alpha,\beta}(x) \,=\, \sum_{\ell=1}^\infty a_\ell^2\,\lambda^\ell_{\alpha,\beta}\,.$$

The conclusion follows by noting that $\lambda^\ell_{\alpha,\beta} \geq \lambda^1_{\alpha,\beta} = \alpha+\beta$ for all $\ell \geq 1$. $\square$

The following is a restatement of Proposition B.5 in the particular case $\alpha = k$ and $\beta = n - k + 1$ in terms of the samples taken from the uniform distribution.

Corollary B.6. *Let $(U_1, \dots, U_n)$ be a sample drawn from the uniform distribution on $(0,1)$. For any $k = 1, \dots, n$ and any absolutely continuous function $u : (0,1) \to \mathbb{R}$,*

$$\mathrm{Var}\big(u(U_k^*)\big) \,\leq\, \frac{1}{n+1}\,\mathbb{E}\big(U_k^*(1-X_k^*)\,u'(U_k^*)^2\big).$$

B.4 Poincaré-type inequalities for L^p norms

While the Poincaré inequalities for the beta distributions of Proposition B.5 have been obtained from a standard L^2 orthogonal decomposition, when $\alpha, \beta \geq 1$, they actually follow, up to an absolute constant, from a stronger result for the L^1-norm in the form of a weighted Cheeger inequality.

Proposition B.7 (Weighted Cheeger inequality for beta distributions)**.** *Given $\alpha, \beta \geq 1$, for any absolutely continuous function u on $(0,1)$ with median m under the beta distribution $B_{\alpha,\beta}$,*

$$\int_0^1 \big|u(x) - m\big|\,dB_{\alpha,\beta}(x) \,\leq\, \frac{C}{\sqrt{\alpha+\beta+1}} \int_0^1 \sqrt{x(1-x)}\,\big|u'(x)\big|\,dB_{\alpha,\beta}(x),$$

where C is an absolute constant. One may take $C = 2.5$.

Proof. Consider an inequality of the form

$$\int_0^1 \big|u(x) - m\big| \, dB_{\alpha,\beta}(x) \, \leq \, C_{\alpha,\beta} \int_0^1 \sqrt{x(1-x)} \, \big|u'(x)\big| \, dB_{\alpha,\beta}(x). \tag{B.13}$$

Here we do not loose generality by assuming that $u \geq 0$ and $m = 0$. Moreover, it is equivalent to the particular case when u is asymptotically the indicator function of an interval $(0, x)$ (like in many other similar Sobolev-type inequalities, see e.g. [B-H2]). In this case (B.13) becomes

$$\min\big\{F_{\alpha,\beta}(x), 1 - F_{\alpha,\beta}(x)\big\} \, \leq \, C_{\alpha,\beta} \sqrt{x(1-x)} \, f_{\alpha,\beta}(x), \qquad 0 < x < 1,$$

where $f_{\alpha,\beta}(x) = \frac{1}{B(\alpha,\beta)} \, x^{\alpha-1}(1-x)^{\beta-1}$ is the density of $B_{\alpha,\beta}$ with respect to the Lebesgue measure and

$$F_{\alpha,\beta}(x) \, = \, B_{\alpha,\beta}([0,x]) \, = \, \frac{1}{B(\alpha,\beta)} \int_0^x y^{\alpha-1}(1-y)^{\beta-1} \, dy$$

is its associated distribution function.

Replacing the role of α and β, that is, using the identity $1 - F_{\alpha,\beta}(1-x) = F_{\beta,\alpha}(x)$, one may further assume that $F_{\alpha,\beta}(x) \leq \frac{1}{2}$, or equivalently, $0 < x \leq m_{\alpha,\beta}$, where $m_{\alpha,\beta}$ denotes the median of a random variable $X_{\alpha,\beta}$ with distribution $B_{\alpha,\beta}$. Thus, under the requirement that $C_{\alpha,\beta} = C_{\beta,\alpha}$, it suffices to show that

$$\int_0^x y^{\alpha-1}(1-y)^{\beta-1} \, dy \, \leq \, C_{\alpha,\beta} \sqrt{x(1-x)} \, x^{\alpha-1}(1-x)^{\beta-1}, \qquad 0 < x \leq m_{\alpha,\beta}.$$

Changing the variable $y = tx$, the above is simplified to

$$\sqrt{\frac{x}{1-x}} \, \int_0^1 t^{\alpha-1}\Big(\frac{1-tx}{1-x}\Big)^{\beta-1} dt \, \leq \, C_{\alpha,\beta}, \qquad 0 < x \leq m_{\alpha,\beta}.$$

But the functions $x \rightarrow \frac{x}{1-x}$ and $x \rightarrow \frac{1-tx}{1-x} = 1 + \frac{x}{1-x}\,(1-t)$ are increasing. Hence, the extremal situation corresponds to the point $x = m_{\alpha,\beta}$, which leads us back to the partial case of the previous inequality, namely

$$\min\big\{F_{\alpha,\beta}(m_{\alpha,\beta}), 1 - F_{\alpha,\beta}(m_{\alpha,\beta})\big\} \, \leq \, C_{\alpha,\beta} \sqrt{m_{\alpha,\beta}\,(1 - m_{\alpha,\beta})} \, f_{\alpha,\beta}(m_{\alpha,\beta}).$$

In addition, $1 - m_{\alpha,\beta} = m_{\beta,\alpha}$ and $f_{\alpha,\beta}(m_{\alpha,\beta}) = f_{\beta,\alpha}(m_{\beta,\alpha})$, so the preceding inequality is symmetric in (α, β). Since the left-hand side is equal to $\frac{1}{2}$, the optimal constant is thus given by

$$C_{\alpha,\beta} \, = \, \frac{1}{2\sqrt{m_{\alpha,\beta}\,(1 - m_{\alpha,\beta})} \, f_{\alpha,\beta}(m_{\alpha,\beta})}\,.$$

Now, since $\alpha \geq 1$ and $\beta \geq 1$, the distribution $B_{\alpha,\beta}$ is log-concave. By Proposition B.1,

$$m_{\alpha,\beta} \, \geq \, (\log 2)\, \mathbb{E}(X_{\alpha,\beta}) \, = \, \log 2 \, \frac{\alpha}{\alpha + \beta}$$

and $1 - m_{\alpha,\beta} = m_{\beta,\alpha} \geq \log 2\, \frac{\beta}{\alpha+\beta}$. Also, by Proposition B.2,

$$\frac{1}{f_{\alpha,\beta}^2(m_{\alpha,\beta})} \ \leq\ 12\,\mathrm{Var}(X_{\alpha,\beta}) \ =\ \frac{12\,\alpha\beta}{(\alpha+\beta)^2\,(\alpha+\beta+1)}\,.$$

The two bounds yield

$$C_{\alpha,\beta} \leq \frac{1}{2\log 2}\,\frac{\sqrt{12}}{\sqrt{\alpha+\beta+1}} \ <\ \frac{2.5}{\sqrt{\alpha+\beta+1}}\,.$$

Proposition B.7 follows. □

Within universal factors, the inequality of Proposition B.7 is actually equivalent to the Poincaré-type inequality of Proposition B.5, which is due to the log-concavity of the beta distributions. We refer to [L4] for discussions of a similar property in the class of general log-concave probability distributions on $\mathbb{R}^n$.

We now extend Proposition B.7 to arbitrary L^p norms.

Proposition B.8 (L^p-Poincaré-type inequality for beta distributions)**.** *Given $\alpha, \beta \geq 1$, for any absolutely continuous function u on $(0,1)$ and any $p \geq 1$,*

$$\int_0^1\!\!\int_0^1 \big|u(x) - u(y)\big|^p\, dB_{\alpha,\beta}(x) dB_{\alpha,\beta}(y) \leq \ \Big(\frac{5p}{\sqrt{\alpha+\beta+1}}\Big)^p \int_0^1 \big(x(1-x)\big)^{p/2} \big|u'(x)\big|^p\, dB_{\alpha,\beta}(x).$$

Proof. First suppose that u has median zero under $B_{\alpha,\beta}$. Then the same is true for $u_+(x) = \max\{u(x), 0\}$ and $u_-(x) = \max\{-u(x), 0\}$, and Proposition B.7 being applied to u_+^p and u_-^p yields

$$\int_0^1 \big|u(x)\big|^p\, dB_{\alpha,\beta}(x) \ \leq\ p\, C_{\alpha,\beta} \int_0^1 \Big(\big(x(1-x)\big)^{1/2} \big|u'(x)\big|\Big) \big|u(x)\big|^{p-1}\, dB_{\alpha,\beta}(x),$$

where

$$C_{\alpha,\beta} \ =\ \frac{2.5}{\sqrt{\alpha+\beta+1}}\,.$$

By Hölder's inequality with exponents p and $q = p/(p-1)$, the last integral does not exceed

$$\Big(\int_0^1 \big(x(1-x)\big)^{p/2} \big|u'(x)\big|^p\, dB_{\alpha,\beta}(x)\Big)^{1/p} \Big(\int_0^1 \big|u(x)\big|^p\, dB_{\alpha,\beta}(x)\Big)^{1/q},$$

and we arrive at

$$\int_0^1 \big|u(x)\big|^p\, dB_{\alpha,\beta}(x) \ \leq\ (p\, C_{\alpha,\beta})^p \int_0^1 \big((x(1-x)\big)^{p/2} \big|u'(x)\big|^p\, dB_{\alpha,\beta}(x).$$

To remove the assumption concerning the median of u, one may also write

$$\int_0^1 \big|u(x) - m\big|^p \, dB_{\alpha,\beta}(x) \ \leq \ (p\, C_{\alpha,\beta})^p \int_0^1 \big((x(1-x)\big)^{p/2} \big|u'(x)\big|^p \, dB_{\alpha,\beta}(x)$$

where m is a median of u under $B_{\alpha,\beta}$. It remains to integrate the inequality

$$\big|u(x) - u(y)\big|^p \ \leq \ 2^{p-1}\big(\big|u(x) - m\big|^p + \big|u(y) - m\big|^p\big)$$

over $B_{\alpha,\beta} \otimes B_{\alpha,\beta}$. The proposition is proved. □

As another variant, we also obtain that

$$\int_0^1 \bigg|u(x) - \int_0^1 u \, dB_{\alpha,\beta}\bigg|^p \, dB_{\alpha,\beta}(x) \ \leq \ \bigg(\frac{5p}{\sqrt{\alpha+\beta+1}}\bigg)^p \int_0^1 \big(x(1-x)\big)^{p/2} \big|u'(x)\big|^p \, dB_{\alpha,\beta}(x).$$

To conclude this paragraph, we restate Proposition B.8 in the particular case $\alpha = k$ and $\beta = n - k + 1$ (in analogy with Corollary B.6).

Corollary B.9. *Let $(U_1, \dots, U_n)$ be a sample drawn from the uniform distribution on $(0,1)$, and let $p \geq 1$. For any $k = 1, \dots, n$, and any absolutely continuous function $u : (0,1) \rightarrow \mathbb{R}$,*

$$\mathbb{E}\big(\big|u(U_k^*) - u(V_k^*)\big|^p\big) \ \leq \ \bigg(\frac{5p}{\sqrt{n+2}}\bigg)^p \ \mathbb{E}\Big(\big(U_k^*(1-U_k^*)\big)^{p/2} \big|u'(U_k^*)\big|^p\Big),$$

where V_k^* is an independent copy of U_k^*.

B.5 Gaussian concentration

If one of the parameters α and β is large, the beta distribution is close to the delta measure at the point $\frac{\alpha}{\alpha+\beta}$, the baricenter of the measure. This is already seen from the obvious bound

$$\mathrm{Var}(X) \ \leq \ \frac{1}{4(\alpha+\beta)}$$

where X is a random variable distributed according to $B_{\alpha,\beta}$. It also follows from Proposition B.5, which yields a Poincaré-type inequality

$$\mathrm{Var}_{B_{\alpha,\beta}}(u) \ \leq \ \frac{1}{4(\alpha+\beta)} \int_0^1 u'(x)^2 \, dB_{\alpha,\beta}(x)$$

with a weakened gradient side. As is well-known (cf. [G-M], [B-U], [Bob2], [L3]), such an analytic inequality provides much more, namely,

$$\mathbb{E}\big(e^{c\sqrt{\alpha+\beta}\,|X - \mathbb{E}(X)|}\big) \ \leq \ 2$$

with some absolute constant $c > 0$.

In fact, this exponential bound may further be sharpened here to a Gaussian bound.

Proposition B.10 (Gaussian concentration of beta distributions)**.** *If X is a random variable distributed according to $B_{\alpha,\beta}$ with $\alpha, \beta \geq 1$, and if Y is an independent copy, then for all $r \geq 0$,*

$$\mathbb{P}\big\{\big|X - \mathbb{E}(X)\big| \geq r\big\} \ \leq \ 2\ e^{-(\alpha+\beta)r^2/8}$$

and

$$\mathbb{P}\big\{|X - Y| \geq r\big\} \ \leq \ 2\, e^{-(\alpha+\beta)r^2/16}.$$

Proof. A main point of the argument is that X has a log-concave distribution of order α, while $1-X$ has a log-concave distribution of order β. Hence, the general Proposition B.4 may be applied to both X and $1 - X$.

Note first that, due to the fact that $|X - \mathbb{E}(X)| < 1$ a.s., it is enough to consider the values $0 \leq r \leq 1$. The double application of Proposition B.4 then yields

$$\mathbb{P}\big\{\big|X - \mathbb{E}(X)\big| \geq r\,\mathbb{E}(X)\big\} \ \leq \ 2\, e^{-\max(\alpha,\beta)\, r^2/4}\,.$$

It remains to use $\max(\alpha, \beta) \geq \frac{\alpha+\beta}{2}$ together with $\mathbb{E}(X) \leq 1$. Similarly, from the second bound of Proposition B.4,

$$\mathbb{P}\big\{|X - Y| \geq r\,\mathbb{E}(X)\big\} \ \leq \ 2\, e^{-\max(\alpha,\beta)\, r^2/8}\,.$$

□

Remark B.11. The subgaussian bound of Proposition B.11 implies that with some absolute constant $c > 0$

$$\mathbb{E}\Big(e^{(\alpha+\beta)\,(X-\mathbb{E}(X))^2/c^2}\Big) \ \leq \ 2.$$

The best value of $c > 0$ in such an inequality represents the so-called ψ_2-norm, i.e. the Orlicz norm generated by the Young function $\psi_2(t) = e^{t^2} - 1$. Thus, for $X \sim B_{\alpha,\beta}$, we have

$$\big\|X - \mathbb{E}(X)\big\|_{\psi_2} \ \leq \ \frac{1}{c\sqrt{\alpha+\beta}}\,.$$

As well as for any other log-concave probability distribution on the line with finite ψ_2-norm, the latter means that the beta distribution shares a logarithmic Sobolev inequality. More precisely, we get that, for any absolutely continuous function $u : (0, 1) \to \mathbb{R}$,

$$\int_0^1 u^2 \log u^2\, dB_{\alpha,\beta} - \int_0^1 u^2\, dB_{\alpha,\beta}\, \log \int_0^1 u^2\, dB_{\alpha,\beta} \ \leq \ \frac{C}{\alpha+\beta} \int_0^1 u'^2\, dB_{\alpha,\beta}$$

with some absolute constant C (cf. [Bob4]). This improves upon the usual Poincaré-type inequality (although does not imply Proposition B.5).

B.6 Mean square beta distributions

This paragraph is concerned with square products of the beta distributions $B_{\alpha,\beta}$ with positive integer parameters $\alpha = k$, $\beta = n-k+1$. More precisely, the representation of Theorem 4.6 leads to investigate the properties of the probability measures on the unit square $(0,1)\times(0,1)$ given by

$$B_n \;=\; \frac{1}{n}\sum_{k=1}^{n} B_{k,n-k+1}\otimes B_{k,n-k+1}.$$

Such a measure B_n will be called the mean square beta distribution of order n.

Every such measure is symmetric around the diagonal $x=y$ on the xy plane and has the uniform distribution on $(0,1)$ as its marginals. It may also be introduced explicitly via the density

$$\frac{dB_n(x,y)}{dx\,dy} \;=\; n\sum_{k=1}^{n}\big(C_{n-1}^{k-1}\big)^2\,(xy)^{k-1}\,\big((1-x)(1-y)\big)^{n-k}, \qquad 0<x,y<1.$$

This expression is however not quite tractable.

If n is large, B_n is nearly concentrated on the diagonal $x=y$. To get an idea about the rate of concentration, we refer to the developments in sub-Section 4.2 of Section 4. Indeed,

$$\begin{aligned}\int_0^1\!\int_0^1 |x-y|^2\,dB_n(x,y) \;&=\; \frac{1}{n}\sum_{k=1}^{n}\int_0^1\!\int_0^1 |x-y|^2\,dB_{k,n-k+1}(x)dB_{k,n-k+1}(y)\\ &=\; \frac{2}{n}\sum_{k=1}^{n}\mathrm{Var}(U_k^*).\end{aligned}$$

Here, $U_1^* \le \dots \le U_n^*$ is the order statistics associated with a sample $(U_1,\dots,U_n)$ drawn from the uniform distribution. But, from Theorem 4.7, the last expression is equal to $\frac{1}{3(n+1)}$. This means that, roughly speaking, B_n is almost supported on the $\frac{1}{\sqrt{n}}$-neighbourhood of the diagonal. A more precise statement is contained in the following assertion, which is an immediate consequence of Proposition B.10.

Proposition B.12. *If (X,Y) is a random vector distributed according to B_n, then for all $r\ge 0$,*

$$\mathbb{P}\big\{|X-Y|\ge r\big\} \;\le\; 2\,e^{(n+1)\,r^2/16}.$$

Indeed, assuming that X_k is a random variable distributed according to $B_{k,n-k+1}$, and that Y_k is an independent copy,

$$\mathbb{P}\{|X-Y|\ge r\} \;=\; \frac{1}{n}\sum_{k=1}^{n}\mathbb{P}\big\{|X_k-Y_k|\ge r\big\}.$$

It then remains to apply the second subgaussian bound of Proposition B.10 (which is independent of k).

A similar application may be developed about Poincaré-type inequalities. Direct consequences of Proposition B.5 and Proposition B.7 are the following Poincaré-type inequalities for B_n restricted to the class of functions of the form $(x,y) \mapsto u(x) - u(y)$.

Proposition B.13 (Poincaré-type inequality for mean square beta distributions)**.** *For any absolutely continuous function u on $(0,1)$,*

$$\int_0^1 \int_0^1 \big|u(x) - u(y)\big|^2 \, dB_n(x,y) \, \leq \, \frac{2}{n+1} \int_0^1 x(1-x)\, u'(x)^2 \, dx.$$

Moreover, for any $p \geq 1$,

$$\int_0^1 \int_0^1 \big|u(x) - u(y)\big|^p \, dB_n(x,y) \, \leq \, \Big(\frac{5p}{\sqrt{n+2}}\Big)^p \int_0^1 \big(x(1-x)\big)^{p/2} \big|u'(x)\big|^p \, dx.$$

To judge sharpness the subgaussian bound of Proposition B.12, one can look at the behaviour of the measures B_n in the topology of the weak convergence. If $k \sim \frac{t}{n}$ with fixed $0 < t < 1$, the order statistics U_k^* associated with a sample $(U_1, \dots, U_n)$ from the uniform distribution are known to be asymptotically normal. More precisely, weakly

$$\sqrt{n}\, \big(U_k^* - \mathbb{E}(U_k^*)\big) \, \to \, N\big(0, t(1-t)\big), \qquad \text{as } n \to \infty,$$

where $N(0, t(1-t))$ is the centered Gaussian measure on the real line with variance $t(1-t)$. Hence, $\sqrt{2n}\,(U_k^* - V_k^*) \to N(0, t(1-t))$ with V_k^* an independent copy of U_k^*. It is not difficult to conclude that, for (X,Y) distributed according to B_n, we have

$$\sqrt{2n}\, (X - Y) \, \to \, \int_0^1 N\big(0, t(1-t)\big)\, dt \qquad \text{as } n \to \infty.$$

Being a mixture of Gaussian measures, this limit distribution has tails that are bounded both from above and from below by the Gaussian tails (up to absolute factors and scaling parameters).

B.7 Lower-bounds on the beta densities

While Proposition B.13 provides upper integral bounds over the mean square beta distributions B_n, we now focus on lower-bounds. This section is mostly technical and deals with pointwise lower-bounds on the densities.

Recall the densities of the beta distributions $B_{k,n-k+1}$, $k = 1, \dots, n$,

$$p_{k,n}(x) \, = \, n\, C_{n-1}^{k-1}\, x^{k-1} (1-x)^{n-k}, \qquad 0 < x < 1,$$

and write the density of the mean square beta distribution B_n as

$$b_n(x,y) \;=\; \frac{1}{n}\sum_{k=1}^{n} p_{k,n}(x)\, p_{k,n}(y), \qquad 0<x,y<1.$$

The analysis is divided is several steps. To this end, set, for every $k=1,\dots,n$, $x_{k,n} = \frac{k}{n+1}$ and $k^* = \min\{k, n-k+1\}$.

Lemma B.14. *For all* $0<x<1$,

$$p_{k,n}(x) \;\geq\; \frac{1}{e\sqrt{3}}\,\frac{n+1}{\sqrt{k^*}}\,\left(\frac{x}{x_{k,n}}\right)^{k-1}\left(\frac{1-x}{1-x_{k,n}}\right)^{n-k}.$$

Proof. Write

$$p_{k,n}(x) \;=\; \frac{p_{k,n}(x)}{p_{k,n}(x_{k,n})}\,p_{k,n}(x_{k,n}) \;=\; p_{k,n}(x_{k,n})\left(\frac{x}{x_{k,n}}\right)^{k-1}\left(\frac{1-x}{1-x_{k,n}}\right)^{n-k}.$$

In order to bound $p_{k,n}(x_{k,n})$ from below, we recall that the k-th order statistic U_k^* for a sample from the uniform distribution has mean $\mathbb{E}(U_k^*) = x_{k,n}$ and variance

$$\mathrm{Var}(U_k^*) \;=\; \frac{k(n-k+1)}{(n+1)^2\,(n+2)} \leq \frac{k^*}{(n+1)^2}\,.$$

Applying the third lower-bound (B.8) of Proposition B.2 with $X = U_k^*$ and its density $f = p_{k,n}$ (which is log-concave), we get

$$p_{k,n}(x_{k,n}) \;\geq\; \frac{1}{\sqrt{3e^2\,\mathrm{Var}(U_k^*)}} \;\geq\; \frac{1}{e\sqrt{3}}\,\frac{n+1}{\sqrt{k^*}}\,.$$

□

The next step is to properly bound the obtained expression in Lemma B.14 for the values x that are sufficiently close to $x_{k,n}$.

Lemma B.15. *If* $x(1-x) \geq \frac{1}{n+1}$ *and* $|x - x_{k,n}| \leq 4\sqrt{\frac{x(1-x)}{n+1}}$ $(0<x<1)$, *then*

$$p_{k,n}(x) \;\geq\; \frac{1}{e^{21}\sqrt{3}}\,\frac{n+1}{\sqrt{k^*}}\,.$$

Proof. Write $x_{k,n} = \frac{k}{n+1} = x + \varepsilon x(1-x)$, so that

$$\frac{x_{k,n}}{x} \;=\; 1+\varepsilon(1-x), \qquad \frac{1-x_{k,n}}{1-x} \;=\; 1-\varepsilon x,$$

and thus

$$\left(\frac{x}{x_{k,n}}\right)^{k-1}\left(\frac{1-x}{1-x_{k,n}}\right)^{n-k} = e^{-A}$$

with

$$A = (k-1)\log\big(1+\varepsilon(1-x)\big) + (n-k)\log\big(1-\varepsilon x\big).$$

Note that

$$\varepsilon^2 = \frac{(x-x_{k,n})^2}{(x(1-x))^2} \le \frac{16}{(n+1)\,x(1-x)} \le 16$$

in view of the two assumptions on x. Hence, $|\varepsilon| \le 4$.

Using the inequality $\log(1+z) \le z$, we have $A \le [(k-1)-(n-1)x]\,\varepsilon$. On the other hand,

$$\begin{aligned}(k-1)-(n-1)x &= (n+1)x_{k,n} - (n-1)x - 1\\ &= (n+1)(x+\varepsilon x(1-x)) - (n-1)x - 1\\ &= (n+1)\,\varepsilon x(1-x) + (2x-1).\end{aligned}$$

Hence, $|(k-1)-(n-1)x| \le 1 + (n+1)\,|\varepsilon|\,x(1-x)$, and thus

$$\begin{aligned}A &\le |\varepsilon| + (n+1)\,\varepsilon^2\,x(1-x)\\ &\le 4 + (n+1)\,\frac{(x-x_{k,n})^2}{x(1-x)} \le 20.\end{aligned}$$

It remains to apply Lemma B.14 to conclude. □

The next lemma provides further lower-bounds on the densities $p_{k,n}$.

Lemma B.16. *Given $0 < t < 1$ such that $t(1-t) \ge \frac{1}{\sqrt{n+1}}$ and $|\,t - \frac{k}{n+1}| \le \sqrt{\frac{t(1-t)}{n+1}}$, we have*

$$p_{k,n}(t+s)\,p_{k,n}(t-s) \ge \frac{1}{3e^{42}}\,\frac{(n+1)^2}{k^*}$$

for all $|s| \le \sqrt{\frac{t(1-t)}{n+1}}$.

Proof. We only need to justify the application of Lemma B.15 to the values $x = t+s$ and $y = t-s$. Note that $n \ge 15$, by the first assumption on t. Since $t \ge \frac{1}{\sqrt{n+1}}$ and $|s| \le \frac{1}{2\sqrt{n+1}}$, necessarily $t - |s| > 0$. Similarly, $t + |s| < 1$, so that $0 < x, y < 1$. Now,

$$|x - x_{k,n}| \le \left|t - \frac{k}{n+1}\right| + |s| \le 2\sqrt{\frac{t(1-t)}{n+1}} \le 4\sqrt{\frac{x(1-x)}{n+1}}\,.$$

More precisely, the last inequality is fulfilled if and only if $t(1-t) \le 4x(1-x)$. To verify it, consider the function $\psi(t) = t(1-t)$. It is 1-Lipschitz in the interval $0 \le t \le 1$, so

$$\psi(x) \,=\, \psi(t+s) \ge \psi(t) - |s| \,\ge\, \psi(t) - \sqrt{\frac{t(1-t)}{n+1}}\,.$$

The required inequality will thus follow from $2\sqrt{\frac{t(1-t)}{n+1}} \le t(1-t)$, hence from $t(1-t) \ge \frac{4}{n+1}$. But $\frac{4}{n+1} \le \frac{1}{\sqrt{n+1}} \le t(1-t)$, and we are done.

To verify the other condition of Lemma B.15, i.e. $\psi(x) \ge \frac{1}{n+1}$ and $\psi(y) \ge \frac{1}{n+1}$, one can return to the previous argument and further note that

$$\psi(x) \,\ge\, \psi(t) - \sqrt{\frac{t(1-t)}{n+1}} \,\ge\, t(1-t) - \frac{1}{2\sqrt{n+1}} \,\ge\, \frac{1}{2\sqrt{n+1}} \,>\, \frac{1}{n+1}\,.$$

Similarly, $\psi(y) > \frac{1}{n+1}$. The proof is complete □

On the basis of the previous lemmas, we are prepared to derive a non-uniform lower-bound on the density of the random vector $(X-Y, X+Y)$, when (X,Y) is distributed according to B_n.

Lemma B.17. *If* $t(1-t) \ge \frac{1}{\sqrt{n+1}}$ $(0 < t < 1)$ *and* $|s| \le \sqrt{\frac{t(1-t)}{n+1}}$, *then*

$$b_n(t+s, t-s) \,\ge\, e^{-45}\sqrt{\frac{n+1}{t(1-t)}}\,.$$

Proof. Let us perform summation over all k in the bound of Lemma B.16. For $0 < t < 1$ fixed, the admissible values of k are given by $a \le k \le b$, where

$$\begin{aligned} a &= (n+1)\,t - \sqrt{(n+1)\,t(1-t)}, \\ b &= (n+1)\,t + \sqrt{(n+1)\,t(1-t)}. \end{aligned}$$

Hence, restricting ourselves to these values and using that $k^* \le k$, we obtain by Lemma B.16 that

$$b_n(t,s) \,\ge\, \frac{n+1}{3e^{42}} \sum_{a \le k \le b} \frac{1}{k}\,.$$

It remains to estimate the last sum by a simpler expression. Assuming that $t \le \frac{1}{2}$, note that

$$D = \sqrt{(n+1)\,t(1-t)} \,\ge\, 2,$$

since it is the same as $t(1-t) \ge \frac{4}{n+1}$, which is true, since $\frac{4}{n+1} \le \frac{1}{\sqrt{n+1}} \le t(1-t)$, as was already used before (recall that necessarily $n \ge 15$). In particular, $b - a \ge 4$. In addition, $a > 1 \Longleftrightarrow (n+1)t > 1-t$, which is also true.

Now, for integer values a_0, b_0 that are closest to a, b respectively, and such that $1 \le a \le a_0 \le b_0 \le b$, we have

$$\sum_{a \le k \le b} \frac{1}{k} = \sum_{k=a_0}^{b_0} \frac{1}{k} \ge \int_{a_0}^{b_0} \frac{dx}{x} = \log \frac{b_0}{a_0} \ge \log \frac{b-1}{a+1}.$$

But $D \ge 2$, so $D - 1 \ge \frac{1}{2} D$, and thus $b - 1 \ge (n+1)\,t + \frac{1}{2} D$. For a similar reason, $a + 1 \le (n+1)\,t$. Hence,

$$\sum_{a \le k \le b} \frac{1}{k} \ge \log \frac{(n+1)\,t + \frac{1}{2} D}{(n+1)\,t} = \log(1+\delta), \qquad \delta = \frac{1}{2} \sqrt{\frac{1-t}{(n+1)\,t}}.$$

Here,

$$\delta = \frac{1-t}{2} \frac{1}{\sqrt{(n+1)\,t(1-t)}} \le \frac{1-t}{2} \frac{1}{(n+1)^{1/4}} < \frac{1}{4}.$$

Hence, $\log(1+\delta) \ge \delta - \frac{1}{2}\delta^2 > \frac{7}{8}\delta$. As a result,

$$\begin{aligned} b_n(t,s) &\ge \frac{n+1}{3e^{42}} \frac{7}{16} \sqrt{\frac{1-t}{(n+1)\,t}} \\ &\ge \frac{(n+1)^{1/2}}{3e^{42}} \frac{7}{16} \frac{1}{2} \frac{1}{\sqrt{t(1-t)}} \\ &> \frac{(n+1)^{1/2}}{e^{45}\sqrt{t(1-t)}}. \end{aligned}$$

The case $t \ge \frac{1}{2}$ is symmetric. Lemma B.17 is established in this way. □

B.8 Lower integral bounds

The pointwise bounds on the densities obtained in the previous section are used here to derive lower integral bounds over the beta distributions $B_{k,n-k+1}$ and the mean square beta distribution B_n. Such bounds will be given in terms of the integrals of the form

$$L_{n,\kappa}(u) = \int_{\{t(1-t) \ge \frac{4}{\sqrt{n+1}}\}} \big[u(t + \kappa\varepsilon_n(t)) - u(t - \kappa\varepsilon_n(t))\big]^p \, dt, \qquad \kappa > 0,$$

where we put $\varepsilon_n(t) = \sqrt{\frac{t(1-t)}{n+1}}$.

The main purpose is to prove the following statement.

Proposition B.18 (Lower integral bounds on beta densities)**.** *For any non-decreasing function u on $(0,1)$ and any $p \geq 1$,*

$$\int_0^1 \int_0^1 \big|u(x) - u(y)\big|^p \, dB_n(x,y) \, \geq \, c\, L_{n,1/2}(u) \tag{B.14}$$

where $c > 0$ is an absolute constant. Moreover,

$$\frac{1}{n} \sum_{k=1}^n \bigg(\int_0^1 \int_0^1 \big|u(x) - u(y)\big| \, dB_{k,n-k+1}(x) \, dB_{k,n-k+1}(y) \bigg)^p \, \geq \, c^p \, L_{n,1/6}(u). \tag{B.15}$$

For the proof, recall that $B_{k,n-k+1}$ and B_n have respective densities $p_{k,n}(x)$ and $b_n(x,y)$, and that $B_{k,n-k+1}$ has mean $t_{k,n} = \frac{k}{n+1}$.

Proof of (B.14)*.* After the change of variables $x = t + s$, $y = t - s$, the integral

$$J \, = \, \int_0^1 \int_0^1 \big|u(x) - u(y)\big|^p \, dB_n(x,y)$$

may be rewritten as

$$J \;\; = \;\; 2 \iint_{\{|s| \leq \min(t,1-t)\}} \big|u(t+s) - u(t-s)\big|^p \; b_n(t+s, t-s) \, dt \, ds.$$

Restricting this integral to the smaller region $R = \{|s| \leq \varepsilon_n(t), \;\, t(1-t) \geq \frac{1}{\sqrt{n+1}}\}$ and applying Lemma B.17, we get that

$$J \, \geq \, 2e^{-45} \iint_R \big|u(t+s) - u(t-s)\big|^p \, \sqrt{\frac{n+1}{t(1-t)}} \, dt \, ds.$$

Moreover, if we further restrict the integration to the values $\frac{1}{2}\,\varepsilon_n(t) \leq |s| \leq \varepsilon_n(t)$, then in view of the monotonicity of the function u, one may use an s-independent bound

$$\big|u(t+s) - u(t-s)\big| \, \geq \, u\Big(t + \frac{1}{2}\,\varepsilon_n(t)\Big) - u\Big(t - \frac{1}{2}\,\varepsilon_n(t)\Big).$$

Hence, after integration over admissible values of s, we arrive at the (B.14) with constant $c = e^{-45}$ (and with better region of integration $t(1-t) \geq \frac{1}{\sqrt{n+1}}$ in comparison with the region appearing in the definition of L_n). □

For the second assertion (B.15) in Proposition B.18, we first relate the integrals

$$J_k \, = \, \int_0^1 \int_0^1 \big|u(x) - u(y)\big| \, dB_{k,n-k+1}(x) \, dB_{k,n-k+1}(y)$$

to the values of u at the points $t_{k,n} \pm \mathrm{const}\, \varepsilon_n(t_{k,n})$.

Lemma B.19. *If* $t_{k,n}(1-t_{k,n}) \geq \frac{2}{\sqrt{n+1}}$, *then*

$$J_k \;\geq\; \frac{1}{60\,e^{42}}\,\Big[u\Big(t_{k,n}+\frac{8}{17}\,\varepsilon_n(t_{k,n})\Big)-u\Big(t_{k,n}-\frac{8}{17}\,\varepsilon_n(t_{k,n})\Big)\Big].$$

Proof. Again change the variables so that to write the integral J_k as

$$J_k \;=\; 2\iint_{\{|s|\leq\min(t,1-t)\}}\big|u(t+s)-u(t-s)\big|\;p_{k,n}(t+s)p_{k,n}(t-s)\,dt\,ds. \qquad \text{(B.16)}$$

Denote by $\Delta_{k,n}$ the collection of all points $t\in(0,1)$ such that

$$t(1-t)\;\geq\;\frac{1}{\sqrt{n+1}} \qquad \text{and} \qquad |t-t_{k,n}|\leq\frac{1}{4}\,\varepsilon_n(t).$$

We restrict the integration in (B.16) to the rectangle $\Delta_{k,n}\times\big[\frac{3}{4}\,\varepsilon_n(t),\varepsilon_n(t)\big]$ and use the monotonicity of the function u (together with the Fubini theorem). The lower pointwise bound of Lemma B.16

$$p_{k,n}(t+s)\,p_{k,n}(t-s)\;\geq\;\frac{1}{3e^{42}}\,\frac{(n+1)^2}{k^*}$$

then yields

$$J_k\;\geq\;\frac{1}{6e^{42}}\,\frac{(n+1)^2}{k^*}\int_{\Delta_{k,n}}\Big[u\Big(t+\frac{3}{4}\,\varepsilon_n(t)\Big)-u\Big(t-\frac{3}{4}\,\varepsilon_n(t)\Big)\Big]\,\varepsilon_n(t)\,dt, \qquad \text{(B.17)}$$

where we recall that $k^*=\min\{k,n-k+1\}$.

To simplify the latter integral, first recall that the condition $t(1-t)\geq\frac{1}{\sqrt{n+1}}$ implies $n\geq 15$ which will be assumed (otherwise, $\Delta_{k,n}$ is empty). The function $\varepsilon_n(t)$ has derivative $\frac{1-2t}{2\sqrt{(n+1)\,t(1-t)}}$, so that

$$\big|\varepsilon_n'(t)\big|\;\leq\;\frac{1}{2(n+1)^{1/4}}\leq\frac{1}{4}\,, \qquad t\in\Delta_{k,n}.$$

Hence, for all $t\in\Delta_{k,n}$,

$$\varepsilon_n(t_{k,n})\;\leq\;\varepsilon_n(t)+\frac{1}{4}\,|t_{k,n}-t|\;\leq\;\frac{17}{16}\,\varepsilon_n(t)\;<\;2\varepsilon_n(t).$$

Moreover, for $\kappa>0$,

$$\begin{aligned} t_{k,n}+\kappa\,\varepsilon_n(t_{k,n}) \;&\leq\; \big(t+|t_{k,n}-t|\big)+\kappa\,\Big(\varepsilon_n(t)+\frac{1}{4}\,|t_{k,n}-t|\Big)\\ &\leq\; t+\frac{4+17\kappa}{16}\,\varepsilon_n(t)\\ &\leq\; t+\frac{3}{4}\,\varepsilon_n(t) \end{aligned}$$

for $\kappa \le \frac{8}{17}$ in the last step. Similarly, for the same range of κ,

$$\begin{aligned} t_{k,n} - \kappa\,\varepsilon_n(t_{k,n}) &\geq \big(t - |t_{k,n} - t|\big) - \kappa\left(\varepsilon_n(t) + \frac{1}{4}\,|t_{k,n} - t|\right) \\ &\geq t - \frac{4 + 17\kappa}{16}\,\varepsilon_n(t) \\ &\geq t - \frac{3}{4}\,\varepsilon_n(t). \end{aligned}$$

Thus, $\varepsilon_n(t) \ge \frac{1}{2}\,\varepsilon_n(t_{k,n})$, and choosing $\kappa = \frac{8}{17}$, we also have

$$t + \frac{3}{4}\,\varepsilon_n(t) \,\geq\, t_{k,n} + \frac{8}{17}\,\varepsilon_n(t_{k,n}), \qquad t - \frac{3}{4}\,\varepsilon_n(t) \,\leq\, t_{k,n} - \frac{8}{17}\,\varepsilon_n(t_{k,n}).$$

Using these bounds in the last estimate (B.17) for J_k, we obtain that

$$J_k \,\geq\, \frac{1}{12\,e^{42}}\,\frac{(n+1)^2}{k^*}\,|\Delta_{k,n}|\left[u\Big(t_{k,n} + \frac{8}{17}\,\varepsilon_n(t_{k,n})\Big) - u\Big(t_{k,n} - \frac{8}{17}\,\varepsilon_n(t_{k,n})\Big)\right]\varepsilon_n(t_{k,n}). \tag{B.18}$$

Next we need to estimate from below the (Lebesgue) measure $|\Delta_{k,n}|$ of $\Delta_{k,n}$. Let us see that it contains the interval $[t_{k,n} - \frac{1}{5}\,\varepsilon_n(t_{k,n}), t_{k,n} + \frac{1}{5}\,\varepsilon_n(t_{k,n})]$. To this aim, we should show that

$$|t - t_{k,n}| \,\leq\, \frac{1}{5}\,\varepsilon_n(t_{k,n}) \,\Rightarrow\, \left(\, t(1-t) \,\geq\, \frac{1}{\sqrt{n+1}} \quad \text{and} \quad |t - t_{k,n}| \,\leq\, \frac{1}{4}\,\varepsilon_n(t)\right),$$

under the condition $t_{k,n}(1 - t_{k,n}) \ge \frac{2}{\sqrt{n+1}}$. Using the inequality $\varepsilon_n(t) \le \frac{1}{2\sqrt{n+1}}$ holding for all $t \in (0,1)$, the assumption $|t - t_{k,n}| \le \frac{1}{5}\,\varepsilon_n(t_{k,n})$ insures that $0 < t < 1$. Indeed,

$$t \,\geq\, t_{k,n} - |t - t_{k,n}| \,\geq\, \frac{2}{\sqrt{n+1}} - \frac{1}{5}\,\varepsilon_n(t_{k,n}) \,\geq\, \frac{2}{\sqrt{n+1}} - \frac{1}{10\sqrt{n+1}} \,>\, 0.$$

Similarly, $t < 1$. Moreover, using the Lipschitz property of the function $t \mapsto t(1-t)$ on $(0,1)$, we get that

$$\begin{aligned} t(1-t) &\geq t_{k,n}(1 - t_{k,n}) - |t - t_{k,n}| \\ &\geq t_{k,n}(1 - t_{k,n}) - \frac{1}{5}\,\varepsilon_n(t_{k,n}) \\ &\geq \frac{2}{\sqrt{n+1}} - \frac{1}{10\sqrt{n+1}} \\ &> \frac{1}{\sqrt{n+1}}\,. \end{aligned}$$

This gives the first required bound $t(1-t) \ge \frac{1}{\sqrt{n+1}}$. But this also implies $|\varepsilon_n'(t)| \le \frac{1}{4}$, so

$$\varepsilon_n(t) \,\geq\, \varepsilon_n(t_{k,n}) - \frac{1}{4}\,|t - t_{k,n}| \,\geq\, \left(5 - \frac{1}{4}\right)|t - t_{k,n}| \,\geq\, 4\,|t - t_{k,n}|,$$

which yields the second required bound $|t - t_{k,n}| \le \frac{1}{4}\,\varepsilon_n(t)$.

As a result, $\Delta_{k,n}$ contains the interval $\{t : |t_{k,n} - t| \le \frac{1}{5}\,\varepsilon_n(t_{k,n})\}$, and thus it has length

$$|\Delta_{k,n}| \ge \frac{2}{5}\,\varepsilon_n(t_{k,n}),$$

as long as $t_{k,n}(1 - t_{k,n}) \ge \frac{2}{\sqrt{n+1}}$. Under this condition, we therefore obtain from (B.18) that

$$J_k \,\ge\, \frac{1}{30\,e^{42}}\,\frac{(n+1)^2}{k^*}\,\varepsilon_n^2(t_{k,n})\,\Big[u\Big(t_{k,n} + \frac{8}{17}\,\varepsilon_n(t_{k,n})\Big) - u\Big(t_{k,n} - \frac{8}{17}\,\varepsilon_n(t_{k,n})\Big)\Big].$$

Finally,

$$\frac{(n+1)^2}{k^*}\,\varepsilon_n^2(t_{k,n}) \,=\, \frac{k(n-k+1)}{k^*(n+1)} \,\ge\, \frac{1}{2}\,.$$

The proof of the lemma is therefore complete. □

On the basis of Lemma B.19, we may now complete the proof of Proposition B.18.

Proof of (B.15). The region of integration $t(1-t) \ge \frac{4}{\sqrt{n+1}}$ in the definition of L_n is non-empty, as long as $n \ge 255$, so we assume this below.

By Lemma B.19, up to the factor c^p with $c = \frac{1}{60}\,e^{-42}$, the left-hand side of (B.15) may be bounded from below by

$$\Sigma_n \,=\, \frac{1}{n}\sum_k d_{k,n}^p, \qquad d_{k,n} \,=\, u\Big(t_{k,n} + \frac{8}{17}\,\varepsilon_n(t_{k,n})\Big) - u\Big(t_{k,n} - \frac{8}{17}\,\varepsilon_n(t_{k,n})\Big),$$

where the sum is running over all $k = 1, \dots, n$, such that $t_{k,n}(1-t_{k,n}) \ge \frac{2}{\sqrt{n+1}}$. For such k, consider the intervals $\delta_{k,n} = \big[t_{k,n} - \frac{1}{2(n+1)}, t_{k,n} + \frac{1}{2(n+1)}\big]$. If $t \in \delta_{k,n}$, we have

$$t(1-t) \,\ge\, t_{k,n}(1-t_{k,n}) - \frac{1}{2(n+1)} \,\ge\, \frac{1}{\sqrt{n+1}}\,.$$

This gives $|\varepsilon_n'(t)| \le \frac{1}{4}$, and in addition $\varepsilon_n(t) = \sqrt{\frac{t(1-t)}{n+1}} \ge \frac{1}{(n+1)^{3/4}}$. Hence,

$$\begin{aligned}
t_{k,n} + \frac{8}{17}\,\varepsilon_n(t_{k,n}) \;&\ge\; \big(t - |t_{k,n} - t|\big) + \frac{8}{17}\Big(\varepsilon_n(t) - \frac{1}{4}\,|t_{k,n} - t|\Big) \\
&\ge\; t + \frac{8}{17}\,\varepsilon_n(t) - \frac{19}{17\,(n+1)} \\
&\ge\; t + \Big(\frac{8}{17} - \frac{19}{17\,(n+1)^{1/4}}\Big)\,\varepsilon_n(t) \\
&\ge\; t + \frac{1}{6}\,\varepsilon_n(t),
\end{aligned}$$

where we used $(n+1)^{1/4} \geq 4$ on the last step. Similarly, $t_{k,n} - \frac{8}{17}\,\varepsilon_n(t_{k,n}) \leq t - \frac{1}{6}\,\varepsilon_n(t)$. Therefore, applying the monotonicity of u, we get

$$d_{k,n}^p \,\geq\, (n+1) \int_{\delta_{k,n}} \Big[u\Big(t + \frac{1}{6}\,\varepsilon_n(t)\Big) - u\Big(t - \frac{1}{6}\,\varepsilon_n(t)\Big)\Big]^p \, dt.$$

It remains to perform summation over k and note that the union of admissible intervals $\delta_{k,n}$ contains the interval $t(1-t) \geq \frac{4}{\sqrt{n+1}}$. Indeed, given any such t, choose k so that $t \in \delta_{k,n}$. Then,

$$t_{k,n}(1 - t_{k,n}) \,\geq\, t(1-t) - |t - t_{k,n}| \,\geq\, \frac{4}{\sqrt{n+1}} - \frac{1}{2(n+1)} \,\geq\, \frac{2}{\sqrt{n+1}}\,.$$

As a result,

$$\Sigma_n \,\geq\, \frac{n+1}{n} \int_{\{t(1-t) \geq \frac{4}{\sqrt{n+1}}\}} \Big[u\Big(t + \frac{1}{6}\,\varepsilon_n(t)\Big) - u\Big(t - \frac{1}{6}\,\varepsilon_n(t)\Big)\Big]^p \, dt.$$

Inequality (B.15) is proved. □

References

[A-S] Aida, S., Stroock, D., Moment estimates derived from Poincare and logarithmic Sobolev inequalities. Math. Res. Letters 1 (1994), 75–86.

[A-G-Z] Anderson, G. W., Guionnet, A., Zeitouni, O., An introduction to random matrices. Cambridge Studies in Advanced Mathematics, 118. Cambridge University Press (2010).

[B-G-L] Bakry, D., Gentil, I., Ledoux, M., Analysis and geometry of Markov diffusion operators. Grundlehren der mathematischen Wissenschaften 348. Springer (2014).

[Ba] Ball, K., Logarithmically concave functions and sections of convex sets in R^n. Studia Math. 88 (1988), 69–84.

[B-M-P] Barlow, R. E., Marshall, A. W., and Proshan, F., Properties of probability distributions with monotone hazard rate. Ann. Math. Stat. 34 (1963), 375–389.

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

[B-G-U] del Barrio, E., Giné, E., Utzet, F., Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. Bernoulli 11 (2005), 131–189.

[B-M1] del Barrio, E., Matrán, C., Rates of convergence for partial mass problems. Probab. Theory Rel. Fields 155 (2013), 521–542.

[B-M2] del Barrio, E., Matrán, C., The empirical cost of optimal incomplete transportation Ann. Probab. 41 (2013), 3140–3156.

[B-B] Barthe, F., Bordenave, C., Combinatorial optimization over two random point sets. Séminaire de Probabilités XLV, Lecture Notes in Math. 2078, 483–535. Springer (2013).

[Bi] Billingsley, P., Convergence of probability measures. John Wiley & Sons Inc. (1968).

[Bob1] Bobkov, S. G., Some extremal properties of the Bernoulli distribution. (Russian) Teor. Veroyatnost. i Primenen. 41 (1996), 877–884.

[Bob2] Bobkov, S. G., Extremal properties of half-spaces for log-concave distributions. Ann. Probab. 24 (1996), 35–48.

[Bob3] Bobkov, S. G., Remarks on the Gromov-Milman inequality. (Russian) Vestn. Syktyvkar. Univ., Ser. 1, Mat. Mekh. Inform. 3 (1999), 15–22.

[Bob4] Bobkov, S. G., Isoperimetric and analytic inequalities for log-concave probability measures. Ann. Probab. 27 (1999), 1903–1921.

[Bob5] Bobkov, S. G., Concentration of distributions of the weighted sums with Bernoullian coefficients. Geometric aspects of functional analysis, Lecture Notes in Math. 1807, 27–36. Springer (2003).

[Bob6] Bobkov, S. G., Spectral gap and concentration for some spherically symmetric probability measures. Geom. Geometric aspects of functional analysis, Lecture Notes in Math.1807, 37–43. Springer (2003).

[Bob7] Bobkov, S. G., Gaussian concentration for a class of spherically invariant measures. J. Math. Sciences (New York), vol. 167 (2010), 326–339. Translated from: Problems in Math. Analysis, vol. 46 (2010), 45–56.

[B-C] Bobkov, S. G., Chistyakov, G. P., On concentration functions of random variables. J. Theor. Probab. Published online: 3 July 2013.

[B-G] Bobkov, S. G., Götze, F., Concentration of empirical distribution functions with applications to non-i.i.d. models. Bernoulli 16 (2010), 1385–1414.

[B-G-T] Bobkov, S. G., Götze, F., Tikhomirov, A., On concentration of empirical measures and convergence to the semi-circle law. J. Theor. Probab. 23, (2010), 792–823.

[B-H1] Bobkov, S. G., Houdré, C., Isoperimetric constants for product probability measures. Ann. Probab. 25 (1997), 184–205.

[B-H2] Bobkov, S. G., Houdré, C., Some connections between isoperimetric and Sobolev-type inequalities. Mem. Amer. Math. Soc. 129, American Mathematical Society (1997).

[B-Ma] Bobkov, S. G., Madiman, M., Concentration of the information in data with log-concave distributions. Ann. Probab. 39 (2011), 1528–1543.

[B-K] Bogachev, V. I., Kolesnikov, A. V., The Monge-Kantorovich problem: achievements, connections, and prospects. (Russian) Uspekhi Mat. Nauk 67 (2012), (407), 3–110; translation in Russian Math. Surveys 67 (2012), 785–890.

[Boi] Boissard, E., Simple bounds for the convergence of empirical and occupation measures in 1-Wasserstein distance. Elect. J. Prob. 16 (2011), 2296–2333.

[B-G-V] Bolley, F., Guillin, A., Villani, C., Quantitative concentration inequalities for empirical measures on non-compact spaces. Probab. Theory Related Fields 137 (2007), 541–593.

[Bor1] Borell, C., Complements of Lyapunov's inequality. Math. Ann. 205 (1973), 323–331 (1973).

[Bor2] Borell, C., Convex measures on locally convex spaces. Ark. Math. 12 (1974), 239–252.

[B-S] Borisov, I. S., Shadrin, A. V., Some remarks on the Sh. S. Ebralidze inequality. (Russian) Teor. Veroyatnost. i Primenen. 41 (1996),177–181; translation in Theory Probab. Appl. 41 (1996), 143–146 (1997).

[B-U] Borovkov, A. A., Utev, S. A., On an inequality and a characterization of the normal distribution connected with it. Probab. Theory Appl. 28 (1983), 209–218.

[B-L] Brascamp, H. J., Lieb, E. H., On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. J. Funct. Anal. 22 (1976), 366–389.

[C-S-S] Cambanis, S., Simons, G., Stout, W., Inequalities for $Ek(X, Y)$ when the marginals are fixed. Z. Wahrsch. Verw. Gebiete 36 (1976), 285–294.

[C-L] Chatterjee, S., Ledoux, M., An observation about submatrices. Electron. Commun. Probab. 14 (2009), 495–500.

[C-H] Csörgő, M., Horváth, L., Weighted Approximations in Probability and Statistics. John Wiley & Sons Inc. (1993).

[C-R] Csörgő, M., Révész, P., Strong approximation of the quantile process. Ann. Statist. 6 (1978), 882–894.

[Da] Dall'Aglio, G., Sugli estremi di momenti detle funzioni di ripartizione doppia. Annali Scuola Normale Superiore di Pisa 10 (1956), 35-74.

[Dal] Dallaporta, S., Eigenvalue variance bounds for Wigner and covariance random matrices. Random Matrices, Theory and Applications 1 (2012), 1250007, 28 pp.

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

[Do] Dobrushin, R. L., Definition of random variables by conditional distributions. (Russian) Teor. Veroyatnost. i Primenen. 15 (1970), 469–497.

[Du1] Dudley, R. M., Real analysis and probability. The Wadsworth & Brooks/Cole Mathematics Series (1989).

[Du2] Dudley, R. M., Uniform central limit theorems. Cambridge Studies in Advanced Mathematics, vol. 63. Cambridge University Press (1999)

[D-K-W] Dvoretzky, A., Kiefer, J., Wolfowitz, J., Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. Ann. Math. Statist. 27 (1956), 642–669.

[Eb] Ebralidze, Sh. S., Inequalities for the probabilities of large deviations in terms of pseudomoments. (Russian) Teor. Verojatnost. i Primenen. 16 (1971), 760–765.

[Et] Etemadi, N., Maximal inequalities for averages of i.i.d. and 2-exchangeable random variables. Statist. Probab. Lett. 44 (1999), 195–200.

[F] Fradelizi, M., Hyperplane sections of convex bodies in isotropic position. Beiträge Algebra Geom. 40 (1999), 163–183.

[F-G] Fournier, N., Guillin, A., On the rate of convergence in Wasserstein distance of the empirical measure (2014).

[Ga] Galambos, J., The asymptotic theory of extreme order statistics. Second edition. Robert E. Krieger Publishing Co. (1987).

[G-Z] Giné, E., Zinn, J., Empirical processes indexed by Lipschitz functions. Ann. Probab. 14 (1986), 1329–1338.

[Go] Gozlan, N., A characterization of dimension free concentration in terms of transportation inequalities. Ann. Probab. 37 (2009), 2480–2498.

[G-L1] Gozlan, N., Léonard, C., A large deviation approach to some transportation cost inequalities. Probab. Theory Related Fields 139 (2007), 235–283.

[G-L2] Gozlan, N., Léonard, C., Transport inequalities. A survey. Markov Process. Related Fields 16 (2010), 635–736.

[G-M] Gromov, M., Milman V. D., A topological application of the isoperimetric inequality. Amer. J. Math. 105 (1983), 843–854.

[Gu] Guédon, O., Kahane-Khinchine type inequalities for negative exponent. Mathematika 46 (1999), 165–173.

[Ha] Haagerup, U., The best constants in the Khintchine inequality. Studia Math. 70 (1981), 231–283.

[H-K] Hájek, J., Rényi, A., Generalization of an inequality of Kolmogorov. Acta Math. Acad. Sci. Hungar 6 (1955), 281–283.

[He] Hensley, D., Slicing convex bodies – bounds for slice area in terms of the body's covariance. Proc. Amer. Math. Soc. 79 (1980), 619–625.

[Ka1] Kantorovitch, L. V., On the translocation of masses. (Russian) Dokl. Akad. Nauk SSSR 37 (1942), 227–229; Journal of Math. Sciences 133 (2006), 1381–1382. Translated from: Zapiski Nauchn. Semin. POMI, vol. 312 (2004), 11–14.

[Ka2] Kantorovitch, L. V., On a problem of Monge. (Russian) Uspekhi Mat. Nauk 3 (1948), 225–226; Journal of Math. Sciences 133 (2006), 1383. Translated from: Zapiski Nauchn. Semin. POMI, vol. 312 (2004), 15–16.

[K-A] Kantorovich, L. V., Akilov, G. P., Functional analysis. Translated from the Russian by Howard L. Silcock. Second edition. Pergamon Press (1982).

[K-R] Kantorovitch, L. V., Rubinstein, G. Sh., On a space of completely additive functions. (Russian) Vestnik Leningrad. Univ. 13 (1958), Ser. Mat. Astron. Phys. 2: 52–59.

[Ki] Kim, T. Y., On tail probabilities of Kolmogorov-Smirnov statistics based on uniform mixing processes. Statist. Probab. Lett. 43 (1999), 217–223.

[Kl] Klartag, B., A central limit theorem for convex sets. Invent. Math. 168 (2007), 91–131.

[L1] Ledoux, M., On Talagrand's deviation inequalities for product measures. ESAIM Probab. Statist. 1 (1995/97), 63–87 (electronic).

[L2] Ledoux, M., Concentration of measure and logarithmic Sobolev inequalities. Séminaire de Probabilités XXXIII, Lecture Notes in Math. 1709, 120–216. Springer (1999).

[L3] Ledoux, M., The concentration of measure phenomenon. Math. Surveys and monographs, vol. 89, American Mathematical Society (2001).

[L4] Ledoux, M., Spectral gap, logarithmic Sobolev constant, and geometric bounds. Surveys in differential geometry. Vol. IX, 219–240, Surv. Differ. Geom., IX, Int. Press (2004).

[L-S] Lovász, L., Simonovits, M., Random walks in a convex body and an improved volume algorithm. Random Structures and Algorithms, 4 (1993), 359–412.

[M-Z] Marcinkiewicz, J., Zygmund, A., Sur les fonctions independantes. Fund. Math., 28 (1937), 60–90. Reprinted in Jozef Marcinkiewicz, Collected papers, edited by Antoni Zygmund, Panstwowe Wydawnictwo Naukowe, Warsaw, 1964, 233–259.

[Mas] Massart, P., The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. Ann. Probab. 18 (1990), 1269–1283.

[M-M] Meckes, E., Meckes, M., Concentration and convergence rates for spectral measures of random matrices. Probab. Theory Related Fields 156 (2013), 145–164.

[P-S] Pastur, L., Shcherbina, M., Eigenvalue distribution of large random matrices. Mathematical Surveys and Monographs, 171. American Mathematical Society (2011).

[P] Prokhorov, Yu. V., Convergence of random processes and limit theorems in probability theory. (Russian) Teor. Veroyatnost. i Primenen. 1 (1956), 177-238.

[Ra] Rachev. S. T., Monge–Kantorovich problem on mass transition and its applications in stochastics. Teor. Veroyatnost. i Primenen. 29:4 (1984), 625–653.

[R-R] Rachev. S. T., Rüschendorf, L., Mass transportation problems. Vol. I. Springer (1998).

[Ri] Rio, E., Upper bounds for minimal distances in the central limit theorem. Ann. Inst. Henri Poincaré Probab. Stat. 45 (2009), 802–817.

[Ru] Rüschendorf, L., The Wasserstein distance and approximation theorems. Z. Wahrsch. Verw. Gebiete 70 (1985), 117–129.

[Se] Sen, P. K., Weak convergence of multidimensional empirical processes for stationary ϕ-mixing processes. Ann. Probab. 2 (1974), 147–154.

[S-W] Shorack, G. R., Wellner, J. A., Empirical processes with applications to statistics. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons (1986).

[Sta] Statulevičius, V. A., Limit theorems for densities and the asymptotic expansions for distributions of sums of independent random variables. (Russian) Teor. Verojatnost. i Primenen. 10 (1965), 645–659.

[Str] Strassen, V., The existence of probability measures with given marginals. Ann. Math. Statist. 36 (1965), 423–439.

[T1] Talagrand, M., The transportation cost from the uniform measure to the empirical measure in dimension ≥ 3. Ann. Probab. 22, 919–959 (1994).

[T2] Talagrand, M., Concentration of measure and isoperimetric inequalities in product spaces. Inst. Hautes Etudes Sci. Publ. Math. 81 (1995), 73–205.

[T-Y] Talagrand, M., Yukich, J. E., The integrability of the square exponential transportation cost. Ann. Appl. Probab. 3 1100–1111 (1993).

[Val] Vallander, S. S., Calculations of the Vasserstein distance between probability distributions on the line. (Russian) Teor. Verojatnost. i Primenen. 18 (1973), 824–827.

[Var] Varadarajan, V. S., On the convergence of sample probability distributions. Sankhya, 19 (1958), 23–26.

[Vas] Vaserstein, L. N., Markov processes over denumerable products of spaces describing large system of automata. Problems of Information Transmission 5 (1969), 47–52. Translated from: Problemy Peredachi Informacii 5 (1969), 64–72 (Russian).

[Ve1] Vershik, A. M., Long history of the Monge-Kantorovich transportation problem. Math. Intell. no. 4 (2013).

[Ve2] Vershik, A. M., Two ways to define consistent metrics on the simplex of measures. (Russian) To appear in: Zapiski Nauch. Semin. POMI, 2013.

[Vi1] Villani, C., Topics in optimal transportation. Graduate Studies in Mathematics, vol. 58. American Mathematical Society (2003).

[Vi2] Villani, C., Optimal transport, Old and new. Grundlehren der mathematischen Wissenschaften 338. Springer (2009).

[Y] Yoshihara, K., Weak convergence of multidimensional empirical processes for strong mixing sequences of stochastic vectors. Z. Wahrsch. Verw. Gebiete 33 (1975/76), 133–137.

[Z] Zolotarev, V. M., Modern theory of summation of random variables. Modern Probability and Statistics. VSP, Utrecht, 1997.