Circling the uniform distribution

Arno Berger*, Ardalan Rahmatidehkordi

Mathematical and Statistical Sciences, University of Alberta, Edmonton, Canada

Abstract

Distributions modulo one of slowly varying sequences are of wide interest in analysis and number theory. For every real sequence (x_n) for which $\lim_{n\to\infty} n(x_{n+1}-x_n)$ exists, this article describes the precise asymptotics of the associated empirical distributions modulo one, utilizing the Kantorovich metric for probability measures on the circle. It is shown in a new, quantitative way that all limit points of these empirical distributions have the same distance from the uniform distribution and form a continuum homeomorphic to a circle. Sharp rates of convergence to that circle are identified for several classes of examples. The results strengthen and complement known facts in the literature. Appropriately adjusted, the main conclusions turn out to be valid also for the sequence $(b \log p_n)$ where b is a real constant and p_n denotes the n-th prime number.

Keywords: Slowly varying sequence, (uniform) distribution modulo one, empirical distribution, Kantorovich metric, discretization error.
2010 MSC: 11K06, 11K31, 26A12, 28A33, 60B10.

1. Introduction

Let (x_n) be an increasing sequence in $\mathbb{R}$ with $x_n \to \infty$ and $x_{n+1} - x_n \to 0$ as $n \to \infty$. The distributions modulo one (mod 1, for short) of such sequences are of wide interest in analysis and number theory, with an extensive literature devoted to them; see, e.g., [1, 7, 17, 19, 26] and the many references therein. It is well known that if $\limsup_{n\to\infty} n(x_{n+1} - x_n) < \infty$ then (x_n) does not have a (unique) distribution mod 1. However, among all distributions mod 1, that is, in the space $\mathcal{P}$ of all probability measures on $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ endowed with the weak topology, the empirical distributions associated with (x_n) are guaranteed to have limit points. It is natural, therefore, to study, as accurately as possible, the structure of these limit points, together with possible rates of convergence, all of which depend, presumably, on properties of (x_n) yet to be identified. In this context, the present article carries out a comprehensive study for the case where

$$\lim_{n\to\infty} n(x_{n+1} - x_n) = b \quad \text{for some } b \in \mathbb{R}\,. \tag{1.1}$$

To informally preview some of the main results, given any (x_n) satisfying (1.1) and any integer $N \geq 1$, denote by ω_N the empirical distribution on the N points $x_1 + \mathbb{Z}, x_2 + \mathbb{Z}, \ldots, x_N + \mathbb{Z}$, that is, $\omega_N(B) = \frac{1}{N}\#\{1 \leq n \leq N : x_n + \mathbb{Z} \in B\}$ for every $B \subset \mathbb{T}$. A suitable metric on $\mathcal{P}$ is used to

*Corresponding author.
Email addresses: berger@ualberta.ca (Arno Berger), rahmatid@ualberta.ca (Ardalan Rahmatidehkordi)

describe the behaviour of ω_N as $N \to \infty$ in an accurate, quantitative manner. Specifically, this article employs the Kantorovich (or 1-Wasserstein) metric $d_{\mathbb{T}}$ on $\mathcal{P}$. Definitions and pertinent properties of $d_{\mathbb{T}}$ are reviewed in Section 2 and Appendix A, for the reader's convenience. For the purpose of this Introduction, it suffices to note that $d_{\mathbb{T}}$ is a bona fide metric, inducing the weak topology and thus making $\mathcal{P}$ a compact metric space. The quantity $d_{\mathbb{T}}$ has many endearing features: For instance, it has a natural interpretation as a minimal "transport cost" from one measure to the other, via the Kantorovich–Rubinstein theorem [10, Sec. 11.8], and as a minimal L^1-distance between appropriately shifted distribution functions; see formula (2.2) below. Its usage in (uniform) distribution theory is also motivated by the fact that $d_{\mathbb{T}}$ equals the largest difference between integrals (w.r.t. one probability measure or the other) incurred over suitable integrands, more precisely,

$$d_{\mathbb{T}}(\mu,\nu) = \sup\left(\int_{\mathbb{T}} h\,\mathrm{d}\mu - \int_{\mathbb{T}} h\,\mathrm{d}\nu\right) \qquad \forall \mu,\nu \in \mathcal{P}\,,$$

where the supremum is over all 1-Lipschitz functions $h : \mathbb{T} \to \mathbb{R}$. The results of this article further highlight the great suitability of $d_{\mathbb{T}}$ for the study of distributions on $\mathbb{T}$.

Now, to preview some of the main results, first consider $(x_n) = (\frac{1}{3}\log n)$ which satisfies (1.1) with $b = \frac{1}{3}$. The sequence (ω_N) of empirical distributions associated with (x_n) is divergent. However, Theorem 4.1, one of the main results, asserts that an appropriately rotated version of ω_N, namely the empirical distribution $\widetilde{\omega}_N$ on the N points $x_1 - x_N + \mathbb{Z}, x_2 - x_N + \mathbb{Z}, \ldots, 0 + \mathbb{Z}$, does converge, to an exponential distribution η "wrapped up mod 1", with distribution function $F_\eta(s) = (e^{3s} - 1)/(e^3 - 1)$. By Corollary 4.2, each limit point of (ω_N) has the same distance from the uniform distribution $\lambda_{\mathbb{T}}$ on $\mathbb{T}$, namely $d_{\mathbb{T}}(\eta, \lambda_{\mathbb{T}}) = \frac{1}{3}\log\cosh\frac{3}{4} \approx 8.609 \cdot 10^{-2}$; moreover, the set of all such limit points is bi-Lipschitz equivalent to the unit circle. Informally put, the sequence (ω_N) asymptotically "circles" the uniform distribution. In addition, Theorem 4.1 yields a $d_{\mathbb{T}}$-rate of convergence for $\widetilde{\omega}_N \to \eta$ of at least $(\log N/N)$, in that $\limsup_{N\to\infty} N d_{\mathbb{T}}(\widetilde{\omega}_N, \eta)/\log N$ is finite. In fact, Remark 5.2 shows that, even more accurately,

$$\lim_{N\to\infty} \frac{N}{\sqrt{\log N}}\, d_{\mathbb{T}}(\widetilde{\omega}_N, \eta) = \frac{1}{3\sqrt{2\pi}} \approx 0.1330\,; \tag{1.2}$$

see Figure 1. Two aspects of (1.2) are particularly noteworthy: On the one hand, as detailed in Sections 3 and 4, it seems plausible that the $d_{\mathbb{T}}$-rate of convergence displayed by (1.2) is best possible in general. For context, notice that $\limsup_{N\to\infty} N d_{\mathbb{T}}(\widetilde{\omega}_N, \lambda_{\mathbb{T}})/\sqrt{\log N} > 0$ always, that is, for *every* sequence (x_n) in $\mathbb{R}$; see, e.g., [14]. In a distributional sense, therefore, the sequence $(\frac{1}{3}\log n)$ exhibits a remarkable (and presumably quite rare) degree of regularity. On the other hand, (1.2) yields the precise rate of convergence for $\widetilde{\omega}_N \to \eta$, up to an explicit positive constant. Other concepts often employed in (uniform) distribution theory, notably versions of *discrepancy*, do not usually allow for such precision; see Remark 5.7 below and, much more comprehensively, [19, 26]. Thus, even though this article is focussed on sequences that are *not* uniformly distributed, its main results suggest that usage of $d_{\mathbb{T}}$ may nonetheless be fruitful in uniform distribution theory as well.

One further illustrative and classical example is $(x_n) = (\frac{1}{3}\log p_n)$ where p_n denotes the n-th prime number. Here (1.1) fails. However, the techniques employed in Section 4 are robust enough to be readily adjustable to this situation. Again, the sequence (ω_N) asymptotically "circles" $\lambda_{\mathbb{T}}$, and $(\widetilde{\omega}_N)$ again converges to η, but does so at the much slower $d_{\mathbb{T}}$-rate $(1/\log N)$: By

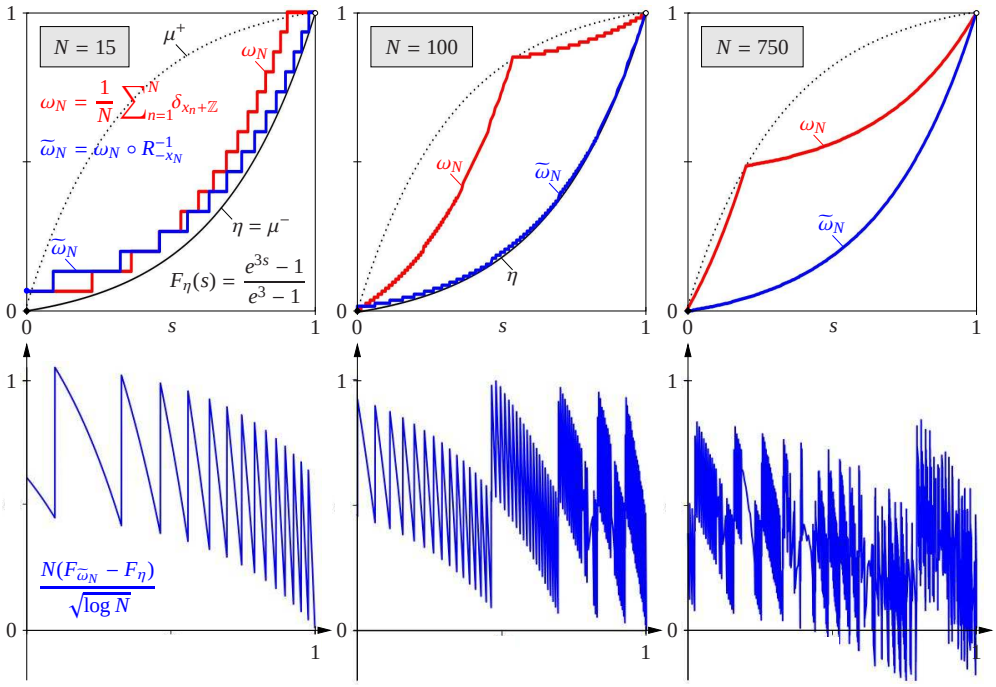


Figure 1: While the empirical distributions ω_N associated with $(x_n) = (\frac{1}{3}\log n)$ do not converge at all (red), the rotated versions $\widetilde{\omega}_N$ do converge (blue) to η, with $d_{\mathbb{T}}$-rate $(\sqrt{\log N}/N)$, by (1.2).

Theorem 5.5,

$$\lim_{N\to\infty} \log N \, d_{\mathbb{T}}(\widetilde{\omega}_N, \eta) = \tfrac{1}{3}\left(e^{\frac{3}{4}} \tanh \tfrac{3}{4} - \log\cosh \tfrac{3}{4} - 1\right) \approx 8.124 \cdot 10^{-2}\,; \tag{1.3}$$

see Figure 2. Note that (1.3), just like (1.2), yields the precise rate for $\widetilde{\omega}_N \to \eta$, up to an explicit positive constant.

To put the approach developed in this article in perspective, the reader may want to recall that the behaviour of ω_N as $N \to \infty$ has traditionally been studied in terms of distribution functions of (x_n); see, e.g., [19, 26, 28] for comprehensive accounts. Specifically, one considers for instance non-decreasing functions with $0 \le F^- \le F^+ \le 1$, given by

$$F^-(s) = \liminf_{N\to\infty} F_{\omega_N}(s)\,, \quad F^+(s) = \limsup_{N\to\infty} F_{\omega_N}(s) \qquad \forall s \in [0,1]\,,$$

and referred to as the *lower* and *upper distribution function* of (x_n) mod 1, respectively. Though usage of $F^\pm$ is quite elementary and may lead to "remarkable results" [19, p. 57], it also has a few unavoidable shortcomings. For instance, $F^\pm$ may not be right-continuous, let alone be the distribution function of a probability measure $\mu^\pm$, say; even if it is, $\mu^\pm$ may fail to be a limit point of (ω_N). Moreover, if F is any distribution function of (x_n), that is, if $F(s) = \lim_{j\to\infty} F_{\omega_{N_j}}(s)$ for almost all $s \in [0,1]$ and some increasing sequence (N_j) in $\mathbb{N}$, then clearly $F^- \le F \le F^+$, but not every (non-decreasing and right-continuous) function with the latter property is a distribution function of (x_n). This suggests that distribution functions of (x_n), notwithstanding their intrinsic

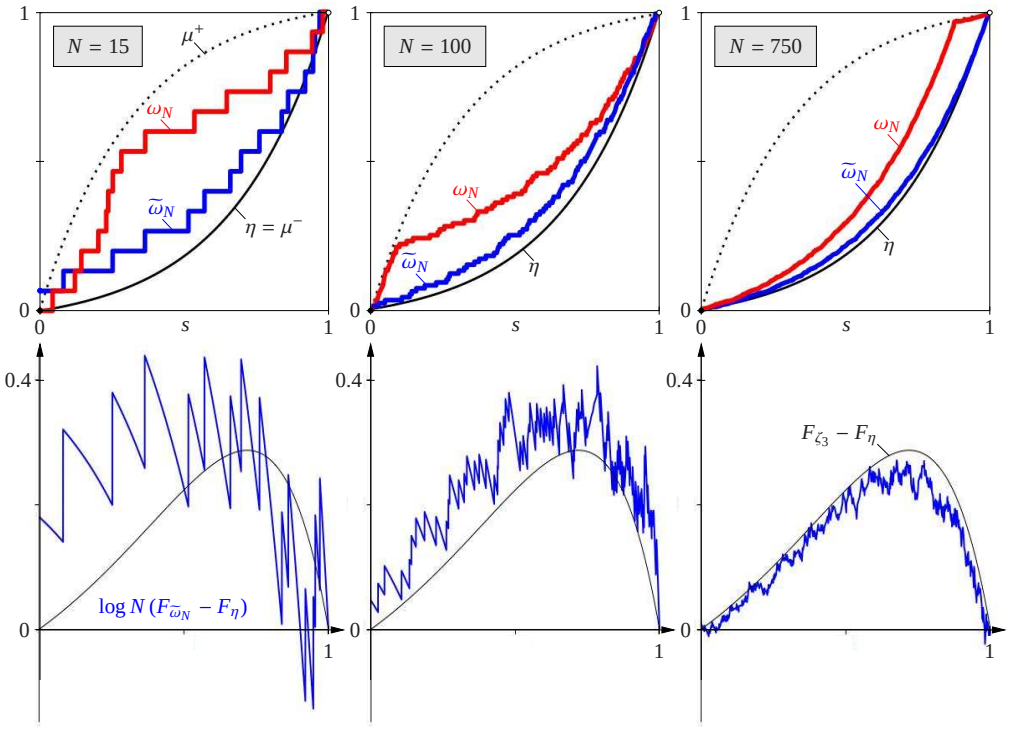


Figure 2: For $(x_n) = (\frac{1}{3}\log p_n)$ the rotated empirical distributions $\widetilde{\omega}_N$ also converge to η, but do so with $d_{\mathbb{T}}$-rate $(1/\log N)$, by (1.3), and in a seemingly more directed or "biased" manner than in Figure 1; see Section 2 for the definition of ζ_3.

analytical appeal, may not be particularly helpful for analyzing the behaviour of ω_N as $N \to \infty$. Finally, one may encounter outright artifacts such as, e.g., for $(x_n) = (\log\log n)$ where $b = 0$ in (1.1) and the constant function $F(s) \equiv c$ is a distribution function for *every* $0 \le c \le 1$; when interpreted as probability measures on $\mathbb{T}$, all these functions are but different representations of one and the same object, namely the Dirac distribution (or unit point mass) concentrated on $0 + \mathbb{Z}$. In light of these shortcomings, the present article avoids usage of distribution functions of (x_n) altogether. As indicated earlier, it instead uses $d_{\mathbb{T}}$ to study, in a direct, quantitative way, the empirical distributions ω_N associated with any (x_n) satisfying (1.1). Such an approach is motivated also by the classical treatise [17].

The organization of this article is as follows. Section 2 reviews definitions of the Kantorovich metric $d_{\mathbb{T}}$ together with its basic properties, as needed in subsequent sections. Two important one-parameter families of probability measures on $\mathbb{T}$ are introduced as well. Section 3 considers the empirical distributions ω_N and $\widetilde{\omega}_N$ associated with arbitrary real sequences mod 1; in particular, it shows that, apart from trivial exceptions, the $d_{\mathbb{T}}$-rate of convergence for (ω_N) to any of its limits points cannot be faster than (N^{-1}). Section 4 precisely describes all limit points of (ω_N) associated with any sequence satisfying (1.1). It also provides a simple bound on the $d_{\mathbb{T}}$-rate of convergence for $(\widetilde{\omega}_N)$. Despite its simplicity, this bound turns out to be essentially sharp, i.e., sharp as $N \to \infty$ up to a positive constant, in many concrete situations, as demonstrated in Section 5. Two basic formulae involving $d_{\mathbb{T}}$ as well as one calculus lemma are proved in Appendix A and Appendix B, respectively. Though elementary and probably known, proofs of these use-

ful results have been included for completeness, but are deferred to appendices for the reader's convenience.

2. Probability measures on the circle

Throughout, the sets of positive integers, integers, rational, positive real, real, and complex numbers are denoted by $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}^+$, $\mathbb{R}$, and $\mathbb{C}$, respectively. For every $t \in \mathbb{R}$, let $\lfloor t \rfloor$ and $\lceil t \rceil$ be the largest integer $\leq t$ and the smallest integer $\geq t$, respectively; also, denote by $\langle t \rangle = t - \lfloor t \rfloor$ the fractional part of t. Intervals in $\mathbb{R}$ are symbolized by square brackets; thus, for example, $[a, b[= \{t \in \mathbb{R} : a \leq t < b\}$ for every $a, b \in \mathbb{R}$. Generally, denote the cardinality and indicator function of any set A by $\#A$ and $\mathbf{1}_A$, respectively, and given $f : A \to \mathbb{R}$ and $t \in \mathbb{R}$, write $\{a \in A : f(a) \leq t\}$ simply as $\{f \leq t\}$, and similarly for $\{f < t\}$, $\{f \neq t\}$ etc.

Recall that $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ is a compact metrizable space when endowed with the usual (quotient) topology. For all $a, b \in \mathbb{R}$, let $d_{\mathbb{T}}(a, b) = \min_{k \in \mathbb{Z}} |a - b + k|$; when restricted to $\mathbb{T} \times \mathbb{T}$, this semi-metric yields a metric that induces the topology of $\mathbb{T}$. Geometrically, one may interpret $\mathbb{T}$ as the unit circle in $\mathbb{C}$: Since $e^{2\pi i x}$ is well-defined for every $x \in \mathbb{T}$, the number $d_{\mathbb{T}}(x, y)$ is the (normalized) arclength of a shortest arc that joins $e^{2\pi i x}$ and $e^{2\pi i y}$ along the unit circle. For every $x \in \mathbb{T}$, let $Q(x) = -x$ and $R_t(x) = x + t$ for all $t \in \mathbb{R}$. Geometrically, Q and R_t are the reflection of the unit circle about the real axis, and its counter-clockwise rotation by $2\pi t$, respectively. Clearly, Q and R_t are isometries of $\mathbb{T}$, with $Q^{-1} = Q$, $R_t^{-1} = R_{-t}$, and $Q \circ R_t = R_{-t} \circ Q$ for every t. To make R_x well-defined for every $x \in \mathbb{T}$ also, simply interpret $R_{t+\mathbb{Z}}$ as R_t.

Denote by $C(\mathbb{T})$ the Banach space of all continuous functions $h : \mathbb{T} \to \mathbb{R}$ with the usual $\|\cdot\|_\infty$-norm, and by $C^\infty(\mathbb{T})$ the (dense) subspace of all smooth functions. For every $a \in \mathbb{R}^+$, let

$$\mathrm{Lip}_a(\mathbb{T}) = \{h \in C(\mathbb{T}) : |h(x) - h(y)| \leq a d_{\mathbb{T}}(x, y)\ \forall x, y \in \mathbb{T}\}\,,$$

and notice that $\mathrm{Lip}(\mathbb{T}) := \bigcup_{a \in \mathbb{R}^+} \mathrm{Lip}_a(\mathbb{T})$, the family of all real-valued Lipschitz functions on $\mathbb{T}$, is a subspace as well, with $\mathrm{Lip}(\mathbb{T}) \supset C^\infty(\mathbb{T})$.

Denote by $\mathcal{P}$ the set of all (Borel) probability measures on $\mathbb{T}$, endowed with the topology of weak convergence. Recall that $\mathcal{P}$ is compact and metrizable, with $\mu_n \to \mu$ in $\mathcal{P}$ if and only if $\int_{\mathbb{T}} h \,\mathrm{d}\mu_n \to \int_{\mathbb{T}} h \,\mathrm{d}\mu$ for every $h \in C(\mathbb{T})$, or indeed merely for every $h \in \mathrm{Lip}_1(\mathbb{T})$ with $h(0) = 0$; see, e.g., [24]. With each $\mu \in \mathcal{P}$ associate its distribution function $F_\mu : \mathbb{R} \to \mathbb{R}$, given by

$$F_\mu(t) = \mu(\{s + \mathbb{Z} : 0 \leq s \leq \langle t \rangle\}) + \lfloor t \rfloor \qquad \forall t \in \mathbb{R}\,.$$

Thus F_μ is non-decreasing and right-continuous. Also, $t \mapsto F_\mu(t) - t$ is 1-periodic, and $F_\mu(0) \geq 0$ as well as $F_\mu(1-) := \lim_{\varepsilon \downarrow 0} F_\mu(1 - \varepsilon) = 1$. Every function with these properties is the distribution function of one and only one $\mu \in \mathcal{P}$. Examples of elements of $\mathcal{P}$ particularly relevant for this article include the Dirac measure δ_x at $x \in \mathbb{T}$, for which $\delta_x(B) = \mathbf{1}_B(x)$ for every Borel set $B \subset \mathbb{T}$, and the Lebesgue measure $\lambda_{\mathbb{T}}$ characterized by $F_{\lambda_{\mathbb{T}}}(t) \equiv t$. Recall that $\mu \in \mathcal{P}$ is *singular* (w.r.t. $\lambda_{\mathbb{T}}$) if there exists a Borel set B with $\mu(B) = 1$ and $\lambda_{\mathbb{T}}(B) = 0$; by contrast, μ is *absolutely continuous* (w.r.t. $\lambda_{\mathbb{T}}$) if $\mu(B) = 0$ whenever $\lambda_{\mathbb{T}}(B) = 0$. The *support* of $\mu \in \mathcal{P}$, denoted $\mathrm{supp}\,\mu$, is the smallest compact set $C \subset \mathbb{T}$ with $\mu(C) = 1$. For instance, δ_x is singular and $\mathrm{supp}\,\delta_x = \{x\}$, whereas $\lambda_{\mathbb{T}}$ obviously is absolutely continuous and $\mathrm{supp}\,\lambda_{\mathbb{T}} = \mathbb{T}$. Note that the sets $\{\mu \in \mathcal{P} : \mu \text{ is absolutely continuous}\}$ and $\{\mu \in \mathcal{P} : \#\,\mathrm{supp}\,\mu < \infty\}$ both are dense in $\mathcal{P}$.

Given $\mu \in \mathcal{P}$, consider the (upper) inverse function of F_μ, defined by $F_\mu^{-1}(t) = \sup\{F_\mu \leq t\}$ for every $t \in \mathbb{R}$. Notice that $F_\mu^{-1} - F_\mu^{-1}(0-)$ has all the properties of a distribution function mentioned

before, and hence is in fact the distribution function of a unique element of $\mathcal{P}$, denoted μ^{-1} and referred to as the *inverse* of μ. For example, $\mu^{-1} = \delta_0$ precisely if $\mu = \delta_x$ for some $x \in \mathbb{T}$, whereas $\mu^{-1} = \lambda_{\mathbb{T}}$ only if $\mu = \lambda_{\mathbb{T}}$. This shows that taking inverses is not one-to-one; it is not onto either since $\mu^{-1} \neq \delta_x$ for every $\mu \in \mathcal{P}$ and $x \in \mathbb{T} \setminus \{0\}$. It is not hard to see that μ^{-1} is singular if and only if μ is singular [3, Ch. A].

A familiar metric inducing the topology of weak convergence on $\mathcal{P}$, henceforth also denoted $d_{\mathbb{T}}$ for simplicity and often referred to as the *Kantorovich* (or 1-*Wasserstein*) *metric*, is given by

$$d_{\mathbb{T}}(\mu,\nu) = \sup_{h\in \mathrm{Lip}_1(\mathbb{T})} \int_{\mathbb{T}} h \,\mathrm{d}(\mu-\nu) \qquad \forall \mu,\nu \in \mathcal{P}\,. \tag{2.1}$$

(Some authors also refer to $d_{\mathbb{T}}$ as the *Earth Mover's Distance* [9].) While definition (2.1) is easy enough to work with in the context of theoretical considerations such as in, e.g., Sections 4 and 5 below or [10, Sec. 11.8], it may be hard to evaluate directly for concrete examples. An easy-to-use explicit formula for $d_{\mathbb{T}}$ is

$$d_{\mathbb{T}}(\mu,\nu) = \min_{t\in\mathbb{R}} \int_0^1 |F_\mu(s) - F_\nu(s) - t|\,\mathrm{d}s \qquad \forall \mu,\nu \in \mathcal{P}\,. \tag{2.2}$$

As a consequence of (2.2),

$$d_{\mathbb{T}}(\mu,\nu) = \int_0^1 |F_\mu(s) - F_\nu(s) - t|\,\mathrm{d}s \iff \min\{\lambda(\{F_\mu - F_\nu \le t\}), \lambda(\{F_\mu - F_\nu \ge t\})\} \ge \tfrac{1}{2}\,; \tag{2.3}$$

here λ denotes Lebesgue measure on $[0,1]$. Thus the minimum in (2.2) is attained precisely if t is a median of the 1-periodic function $F_\mu - F_\nu$. Although (2.2) apparently has long been part of analysis folklore [5, 9, 14], the authors have been unable to locate a complete proof in the literature; for the reader's convenience, a simple proof of (2.2) and (2.3) is provided in Appendix A.

Using (2.2), it is readily deduced that $d_{\mathbb{T}}(\mu,\lambda_{\mathbb{T}}) \le \frac{1}{4}$, and this inequality is strict unless $\mu = \delta_x$ for some $x \in \mathbb{T}$. As a consequence, $d_{\mathbb{T}}(\mu,\nu) \le \frac{1}{2}$ for all $\mu,\nu \in \mathcal{P}$, and this inequality is strict unless $\mu = \delta_x$, $\nu = \delta_{x+1/2}$ for some $x \in \mathbb{T}$. Note that $d_{\mathbb{T}}(\delta_x,\delta_y) = d_{\mathbb{T}}(x,y)$ for all $x,y \in \mathbb{T}$, so $\{\delta_x : x \in \mathbb{T}\}$ is an isometric copy of $\mathbb{T}$ inside $\mathcal{P}$.

Given any (Borel measurable) map $S : \mathbb{T} \to \mathbb{T}$ and $\mu \in \mathcal{P}$, recall that $\mu \circ S^{-1}$, referred to as the *push-forward* of μ under S, is the unique $\nu \in \mathcal{P}$ with $\nu(B) = \mu(S^{-1}(B))$ for every Borel set B. For example, $\delta_x \circ R_t^{-1} = \delta_{x+t}$ and $\lambda_{\mathbb{T}} \circ R_t^{-1} = \lambda_{\mathbb{T}}$ for all $x \in \mathbb{T}$, $t \in \mathbb{R}$. If $S : \mathbb{T} \to \mathbb{T}$ is continuous, Lipschitz, or an isometry then the same is true for the map $\mu \mapsto \mu \circ S^{-1}$ induced on $\mathcal{P}$; in particular, Q and R_t induce isometries on $\mathcal{P}$. By contrast, $\mu \mapsto \mu^{-1}$ is discontinuous at every μ with $\mu(\{0\}) > 0$; cf. [3, Lem. A5]. However, regarding the inverses of any $\mu,\nu \in \mathcal{P}$ notice two simple facts: $\mu^{-1} = \nu^{-1}$ implies that $\mu = \nu \circ R_t^{-1}$ for some $t \in \mathbb{R}$; and $(\mu^{-1})^{-1} = \mu \circ R_t^{-1}$ with the appropriate $t \in \mathbb{R}$.

Two one-parameter families in $\mathcal{P}$ are particularly relevant for this article: For every $a \in \mathbb{R}\setminus\{0\}$ let η_a and ζ_a be defined via their distribution functions,

$$F_{\eta_a}(s) = \frac{e^{as}-1}{e^a-1}\,, \quad F_{\zeta_a}(s) = \frac{e^{as}(a+1-as)-1}{e^a-1} + a\frac{e^{as}-e^a}{(e^a-1)^2} \qquad \forall s \in [0,1]\,.$$

In probabilistic terms, η_a and ζ_a simply are the distributions of $\langle -Y_1/a\rangle + \mathbb{Z}$ and $\langle -(Y_1+Y_2)/a\rangle + \mathbb{Z}$, respectively, where the real random variables Y_1, Y_2 are independent and standard exponential.

A straightforward calculation using (2.2) yields

$$d_{\mathbb{T}}(\eta_a,\eta_b)=\frac{|a-b|}{ab}\left(\frac{\cosh\left(\frac{1}{4}a\right)^{b/(a-b)}}{\cosh\left(\frac{1}{4}b\right)^{a/(a-b)}}-1\right)\qquad\forall a,b\in\mathbb{R}\setminus\{0\},a\neq b\,,\tag{2.4}$$

and consequently, for every $a\in\mathbb{R}\setminus\{0\}$,

$$\lim_{b\to 0}d_{\mathbb{T}}(\eta_a,\eta_b)=\frac{\log\cosh\left(\frac{1}{4}a\right)}{|a|}=d_{\mathbb{T}}(\eta_a,\lambda_{\mathbb{T}})\,,\quad\lim_{|b|\to\infty}d_{\mathbb{T}}(\eta_a,\eta_b)=\frac{\tanh\left(\frac{1}{4}a\right)}{a}=d_{\mathbb{T}}(\eta_a,\delta_0)\,,\tag{2.5}$$

which suggests defining $\eta_0=\lambda_{\mathbb{T}}$ and $\eta_{-\infty}=\eta_\infty=\delta_0$. With this, $\{\eta_a:-\infty\le a\le\infty\}$ is a homeomorphic (in fact, bi-Lipschitz equivalent) copy of $\mathbb{T}$ inside $\mathcal{P}$. Similarly,

$$d_{\mathbb{T}}(\zeta_a,\eta_a)=\frac{1}{|a|}\left(\frac{e^{\frac{1}{4}a\tanh\left(\frac{1}{4}a\right)}}{\cosh\left(\frac{1}{4}a\right)}-1\right)\qquad\forall a\in\mathbb{R}\setminus\{0\}\,,\tag{2.6}$$

and hence $\lim_{a\to 0}d_{\mathbb{T}}(\zeta_a,\eta_a)=0=\lim_{|a|\to\infty}d_{\mathbb{T}}(\zeta_a,\eta_a)$, which suggests defining $\zeta_0=\lambda_{\mathbb{T}}$ and $\zeta_{-\infty}=\zeta_\infty=\delta_0$ as well, so again $\{\zeta_a:-\infty\le a\le\infty\}$ is a homeomorphic copy of $\mathbb{T}$ inside $\mathcal{P}$.

By (2.4), the family $(\eta_a)_{a\in\mathbb{R}}$ is continuous. In fact, it even is differentiable in the sense that

$$\lim_{\varepsilon\to 0}\int_{\mathbb{T}}h\,\mathrm{d}\left(\frac{\eta_{a+\varepsilon}-\eta_a}{\varepsilon}\right)=\int_{\mathbb{T}}h\,\mathrm{d}\eta_a'\qquad\forall h\in C(\mathbb{T})\,,\tag{2.7}$$

where the signed (Borel) measure η_a' on $\mathbb{T}$, the "tangent vector" to η at η_a, is given by

$$\eta_a'=\begin{cases}(\eta_a-\zeta_a)/a & \text{if } a\in\mathbb{R}\setminus\{0\}\,,\\ \frac{1}{2}(\beta_{2,1}-\lambda_{\mathbb{T}}) & \text{if } a=0\,,\end{cases}$$

with $F_{\beta_{2,1}}(s)=s^2$ for all $s\in[0,1]$. Note that the convergence in (2.7) is uniform on $\mathrm{Lip}_1(\mathbb{T})$. Similarly, the family $(\zeta_a)_{a\in\mathbb{R}}$ is differentiable. Moreover, for every $h\in C^\infty(\mathbb{T})$,

$$\lim_{|a|\to\infty}a\int_{\mathbb{T}}h\,\mathrm{d}(\delta_0-\eta_a)=\lim_{|a|\to\infty}\frac{a}{2}\int_{\mathbb{T}}h\,\mathrm{d}(\delta_0-\zeta_a)=h'(0)=\lim_{\varepsilon\to 0}\int_{\mathbb{T}}h\,\mathrm{d}\left(\frac{\delta_\varepsilon-\delta_0}{\varepsilon}\right),$$

so that, in a sense, the "tangent vectors" to both families (η_a) and (ζ_a) at $\eta_\infty=\zeta_\infty=\delta_0$ are proportional to the "tangent vector" of the family $(\delta_t)_{t\in\mathbb{R}}$ at δ_0.

Reflected and rotated versions of η_a are going to play a prominent role later on. Notice that $\eta_a\circ Q^{-1}=\eta_{-a}$ for every $a\in\mathbb{R}\cup\{-\infty,\infty\}$, but also $\eta_0\circ R_t^{-1}=\lambda_{\mathbb{T}}$ for every $t\in\mathbb{R}$, whereas $\eta_{-\infty}\circ R_t^{-1}=\eta_\infty\circ R_t^{-1}=\delta_t$, and similarly for ζ_a. It is straightforward to show that for every $a\in\mathbb{R}\setminus\{0\}$,

$$\frac{4\log\cosh\left(\frac{1}{4}a\right)}{|a|}\,d_{\mathbb{T}}(s,t)\le d_{\mathbb{T}}(\eta_a\circ R_s^{-1},\eta_a\circ R_t^{-1})\le\left|\tanh\left(\tfrac{1}{4}a\right)\right|d_{\mathbb{T}}(s,t)\qquad\forall s,t\in\mathbb{R}\,;\tag{2.8}$$

when interpreted with caution, (2.8) remains valid for $a=0$ and $|a|=\infty$ also. Thus, for every $a\neq 0$ the set $\{\eta_a\circ R_t^{-1}:t\in\mathbb{R}\}$ is a bi-Lipschitz equivalent copy of $\mathbb{T}$ inside $\mathcal{P}$ which one may interpret as a "circle" with "centre" $\lambda_{\mathbb{T}}$ and "radius" $d_{\mathbb{T}}(\eta_a,\lambda_{\mathbb{T}})\in\,]0,\frac{1}{4}]$; see Figure 3.

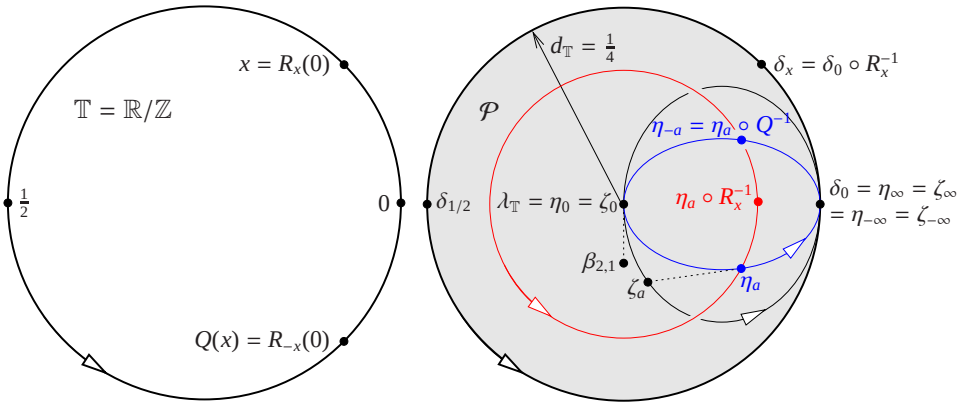


Figure 3: The compact metric spaces $\mathbb{T}$ (left) and $\mathcal{P}$ (right; schematic).

3. Empirical distributions on the circle

Given any sequence (x_n) in $\mathbb{R}$, denote by $\mathbb{A}[x_n]$ the set of all limit points of (x_n) mod 1, that is,

$$\mathbb{A}[x_n] = \{x \in \mathbb{T} : \exists\, N_1 < N_2 < \dots \text{ s.t. } d_{\mathbb{T}}(x_{N_j}, x) \to 0 \text{ as } j \to \infty\}\,.$$

Note that $\mathbb{A}[x_n]$ is non-empty and compact; every subset of $\mathbb{T}$ with these two properties equals $\mathbb{A}[y_n]$ for the appropriate sequence (y_n). Also, for every $N \in \mathbb{N}$ denote by $\omega_N[x_n]$ the empirical distribution of $\{x_1, x_2, \dots, x_N\}$ mod 1, that is,

$$\omega_N[x_n] = \frac{1}{N}\sum\nolimits_{n=1}^{N} \delta_{x_n+\mathbb{Z}} \in \mathcal{P}\,. \tag{3.1}$$

In other words, $\omega_N[x_n]$ is the unique $\mu \in \mathcal{P}$ with $\int_{\mathbb{T}} h\,\mathrm{d}\mu = \frac{1}{N}\sum_{n=1}^{N} h(x_n)$ for every $h \in C(\mathbb{T})$. In later sections, it is going to be convenient also to consider a suitably rotated version of $\omega_N[x_n]$,

$$\widetilde{\omega}_N[x_n] := \omega_N[x_n] \circ R_{-x_N}^{-1} = \frac{1}{N}\sum\nolimits_{n=1}^{N} \delta_{x_n - x_N + \mathbb{Z}}\,.$$

This section collects but a few basic observations regarding $\omega_N[x_n]$ as $N \to \infty$. Given (x_n), denote by $\Omega[x_n]$ the set of all limit points of (ω_N) in $\mathcal{P}$, i.e.,

$$\Omega[x_n] = \{\mu \in \mathcal{P} : \exists\, N_1 < N_2 < \dots \text{ s.t. } \omega_{N_j}[x_n] \to \mu \text{ as } j \to \infty\}\,.$$

For every $\mu \in \Omega[x_n]$, note that $\operatorname{supp}\mu \subset \mathbb{A}[x_n]$, and this inclusion may be strict. Throughout, whenever (x_n) is clear from the context, $\mathbb{A}[x_n]$, $\omega_N[x_n]$, $\widetilde{\omega}_N[x_n]$, and $\Omega[x_n]$ may be written simply as $\mathbb{A}$, ω_N, $\widetilde{\omega}_N$, and Ω, respectively. The basic properties of the set $\Omega[x_n]$ are as follows; see also [19, Sec. 2.4] and [26, Sec. 2.2].

Theorem 3.1. *Let (x_n) be a sequence in $\mathbb{R}$. Then $\Omega[x_n]$ is non-empty, compact, and connected. Given any non-empty, compact, and connected set $\mathcal{A} \subset \mathcal{P}$ there exists an increasing sequence (y_n) in $\mathbb{R}$ such that $\Omega[y_n] = \mathcal{A}$.*

Proof. Given (x_n), for every $N \in \mathbb{N}$ let $\Omega_N = \overline{\{\omega_n : n \geq N\}}$, the closure of $\{\omega_n : n \geq N\}$ in $\mathcal{P}$. Clearly, $N \mapsto \Omega_N$ is non-increasing, and each Ω_N is non-empty and compact, hence closed. By the compactness of $\mathcal{P}$, the intersection $\bigcap_{N\in\mathbb{N}} \Omega_N = \Omega$ is non-empty and closed, hence compact. To show that Ω is connected, pick any continuous function $f : \Omega \to \{0, 1\}$. There exists a continuous function $g : \mathcal{P} \to [0, 1]$ with $g|_\Omega = f$. Suppose that f was not constant, that is, $f(\Omega) = \{0, 1\}$. Then, since $\mathcal{P}$ is connected, $g(\mathcal{P}) = [0, 1]$. Moreover, deduce from

$$(N + 1) \int_{\mathbb{T}} h \, \mathrm{d}\omega_{N+1} - N \int_{\mathbb{T}} h \, \mathrm{d}\omega_N = h(x_{N+1}) \qquad \forall N \in \mathbb{N}, h \in C(\mathbb{T}),$$

that $d_{\mathbb{T}}(\omega_{N+1}, \omega_N) \leq 1/N$ for every $N \in \mathbb{N}$. Since g is continuous, the set $\{g(\omega_N) : N \in \mathbb{N}\}$ is dense in $[0, 1]$. Consequently, there exists an increasing sequence (N_j) in $\mathbb{N}$ with $\frac{1}{3} \leq g(\omega_{N_j}) \leq \frac{2}{3}$ for all $j \in \mathbb{N}$, but also $\lim_{j\to\infty} \omega_{N_j} = \mu$ for some $\mu \in \Omega$. But then, by continuity, $f(\mu) = g(\mu) = \lim_{j\to\infty} g(\omega_{N_j}) \in [\frac{1}{3}, \frac{2}{3}]$, which contradicts the fact that $f(\Omega) = \{0, 1\}$. Thus, f must be constant. In other words, Ω is connected.

To prepare for the proof of the remaining assertion, denote by $\mathcal{P}_0$ the set of all $\mu \in \mathcal{P}$ for which $\#\operatorname{supp}\mu < \infty$. Recall that $\mathcal{P}_0$ is dense in $\mathcal{P}$. Moreover, given any $C \subset \mathcal{P}$ and $\varepsilon > 0$, let $C_\varepsilon = \{\mu \in \mathcal{P} : d_{\mathbb{T}}(\mu, \nu) < \varepsilon \text{ for some } \nu \in C\}$. Note that C_ε is open, $C_\varepsilon \supset \overline{C}$, and $(C_\varepsilon)_{\widetilde{\varepsilon}} \subset C_{\varepsilon+\widetilde{\varepsilon}}$ for all $\varepsilon, \widetilde{\varepsilon} > 0$. Also, $\bigcap_{\varepsilon>0} C_\varepsilon = \overline{C}$, and if C is connected then C_ε is path-connected for all $\varepsilon > 0$; for closed C the converse is also true.

Assume that $\mathcal{A} \subset \mathcal{P}$ is non-empty, compact, and connected. Given any $\varepsilon > 0$ and $\mu_0, \mu_1 \in \mathcal{A}_\varepsilon \cap \mathcal{P}_0$, since $\mathcal{A}_\varepsilon$ is path-connected, since every continuous path can be approximated by a polygonal path, and since $\mathcal{P}_0$ is dense in $\mathcal{P}$, there exists a continuous path $\gamma_\varepsilon : [0, 1] \to \mathcal{P}$ with the following properties:

(i) γ_ε connects μ_0 and μ_1, i.e., $\gamma_\varepsilon(0) = \mu_0$, $\gamma_\varepsilon(1) = \mu_1$;
(ii) γ_ε lies within $\mathcal{A}_\varepsilon$, i.e., $\gamma_\varepsilon(t) \in \mathcal{A}_\varepsilon$ for every $t \in [0, 1]$;
(iii) γ_ε is polygonal, with each of its (straight) segments being directed towards a Dirac measure, i.e., with the appropriate $M_\varepsilon \in \mathbb{N}$, $0 = t_{\varepsilon,0} < t_{\varepsilon,1} < \ldots < t_{\varepsilon,M_\varepsilon-1} < t_{\varepsilon,M_\varepsilon} = 1$, and $x_{\varepsilon,1}, \ldots, x_{\varepsilon,M_\varepsilon} \in \mathbb{T}$,

$$\gamma_\varepsilon(t) = (1 - (t - t_{\varepsilon,m-1}))\gamma_\varepsilon(t_{\varepsilon,m-1}) + (t - t_{\varepsilon,m-1})\delta_{x_m} \qquad \forall m \in \{1, \ldots, M_\varepsilon\}, t \in [t_{\varepsilon,m-1}, t_{\varepsilon,m}],$$

so in particular $\gamma_\varepsilon(t) \in \mathcal{P}_0$ for every t;
(iv) γ_ε approximately covers $\mathcal{A}$, i.e., $\gamma_\varepsilon([0, 1])_\varepsilon \supset \mathcal{A}$.

With these preparations, fix any decreasing sequence (ε_j) in $\mathbb{R}^+$ with $\lim_{j\to\infty} \varepsilon_j = 0$, and for every j pick $\mu_i \in \mathcal{P} \cap \mathcal{A}_{\varepsilon_j}$. By the above, there exists a continuous path $\gamma_{\varepsilon_j} : [0, 1] \to \mathcal{P}_0 \cap \mathcal{A}_{\varepsilon_j}$ with $\gamma_{\varepsilon_j}(0) = \mu_j$ and $\gamma_{\varepsilon_j}(1) = \mu_{j+1}$. Since $\gamma_{\varepsilon_j}([0, 1]) \subset \mathcal{A}_{\varepsilon_j}$, there exists $\widetilde{\varepsilon}_j > 0$ such that $\gamma_{\varepsilon_j}([0, 1])_{\widetilde{\varepsilon}_j} \subset \mathcal{A}_{\varepsilon_j}$. Pick $N_1 \in \mathbb{N}$ and $y_1, \ldots, y_{N_1} \in \operatorname{supp}\mu_1$ so that $d_{\mathbb{T}}(\omega_{N_1}, \mu_1) < \widetilde{\varepsilon}_1$. By property (iii), it is possible to pick $\widetilde{N}_1 > N_1$ and $y_{N_1+1}, \ldots, y_{\widetilde{N}_1} \in \mathbb{T}$ with

$$\{\omega_N[y_n] : N_1 \leq N \leq \widetilde{N}_1\} \subset \mathcal{A}_{\varepsilon_1}, \quad \{\omega_N[y_n] : N_1 \leq N \leq \widetilde{N}_1\}_{\varepsilon_1} \supset \mathcal{A}, \quad d_{\mathbb{T}}(\omega_{\widetilde{N}_1}[y_n], \mu_2) < \widetilde{\varepsilon}_2 . \tag{3.2}$$

Letting $N_{i+1} = \widetilde{N}_i$, continue this process inductively, i.e., pick $\widetilde{N}_{i+1} > N_{i+1}$ and $y_{N_{i+1}+1}, \ldots, y_{\widetilde{N}_{i+1}}$ in $\mathbb{T}$ so that (3.2) holds with $N_{i+1}, \widetilde{N}_{i+1}, \varepsilon_{i+1}$, and $\widetilde{\varepsilon}_{i+2}$ instead of $N_1, \widetilde{N}_1, \varepsilon_1$, and $\widetilde{\varepsilon}_2$, respectively. For the sequence (y_n) in $\mathcal{T}$ thus obtained, $\omega_N[y_n] \in \mathcal{A}_{\varepsilon_j}$ for all $N \geq N_j$, and hence

$$\Omega[y_n] = \bigcap\nolimits_{j\in\mathbb{N}} \overline{\{\omega_N[y_n] : N \geq N_j\}} \subset \bigcap\nolimits_{j\in\mathbb{N}} \overline{\mathcal{A}_{\varepsilon_j}} \subset \bigcap\nolimits_{j\in\mathbb{N}} \mathcal{A}_{2\varepsilon_j} = \mathcal{A},$$

but also $\mathcal{A} \subset \{\omega_N[y_n] : N \geq N_j\}_{\varepsilon_j}$ for every j, and so

$$\mathcal{A} \subset \bigcap\nolimits_{j\in\mathbb{N}} \{\omega_N[y_n] : N \geq N_j\}_{\varepsilon_j} \subset \bigcap\nolimits_{j\in\mathbb{N}} \overline{\{\omega_N[y_n] : N \geq N_j\}}_{\varepsilon_j} = \overline{\Omega[y_n]} = \Omega[y_n]\,.$$

This shows that $\Omega[y_n] = \mathcal{A}$. To make (y_n) increasing, simply replace y_n by $y_n + k_n$ with the appropriate integer k_n, as needed. Since the compact, connected set $\mathcal{A} \neq \varnothing$ has been arbitrary, the proof is complete. (Notice how the connectedness of $\mathcal{A}$ is essential for the argument.) □

This article focusses on sequences (x_n) for which $\Omega[x_n]$ is very small. An extreme situation obviously occurs when $\Omega[x_n] = \{\mu\}$ for some $\mu \in \mathcal{P}$, or equivalently, when $\omega_N \to \mu$ as $N \to \infty$. In this situation, (x_n) is said to *have distribution* μ mod 1, with the most classical special case being $\Omega[x_n] = \{\lambda_{\mathbb{T}}\}$, that is, (x_n) being *uniformly distributed modulo one*, or *u.d.* mod 1 for short; see, e.g., [17, 19]. Typically, though, $\Omega[x_n]$ is not a singleton, i.e., the sequence (ω_N) in $\mathcal{P}$ does not converge; in this case, $\Omega[x_n]$ is uncountable [26, Thm. 6]. Not least for the sequences (x_n) studied in detail in later sections, however, it turns out that, although (ω_N) does not converge, $(\omega_N \circ R_{t_N}^{-1})$ with an appropriate sequence (t_n) may well converge; recall that $\omega_N \circ R_{t_N}^{-1}$ denotes the push-forward of the empirical distribution ω_N under the rotation R_{t_N}. Even in this slightly more general setting, by the nature of (3.1), it seems plausible that $d_{\mathbb{T}}(\omega_N \circ R_{t_N}^{-1}, \mu)$ cannot tend to 0 faster than N^{-1} as $N \to \infty$, except perhaps for (very) singular μ. The following theorem, the main result of this section, makes this precise.

Theorem 3.2. *Let (x_n), (t_n) be sequences in $\mathbb{R}$, and $\mu \in \mathcal{P}$. Then, with ω_N given by* (3.1)*:*

(i) $\limsup_{N\to\infty} N d_{\mathbb{T}}(\omega_N \circ R_{t_N}^{-1}, \mu) > 0$ *unless* $\lim_{N\to\infty} \sum_{n=1}^{N} d_{\mathbb{T}}(x_n + t_N, x) = 0$ *and* $\mu = \delta_x$ *for some* $x \in \mathbb{T}$*;*

(ii) $\liminf_{N\to\infty} N d_{\mathbb{T}}(\omega_N \circ R_{t_N}^{-1}, \mu) > 0$ *provided that μ is not singular.*

The proof of Theorem 3.2 makes use of two elementary observations from calculus regarding monotone functions $f : [0,1] \to \mathbb{R}$. Recall that any such f is Riemann integrable, and has left- and right-sided limits at every $s \in]0,1[$, denoted $f(s-)$ and $f(s+)$, respectively. The first observation is a special case of [31, Cor. 4.4(ii)].

Proposition 3.3. *Let $f : [0,1] \to \mathbb{R}$ be monotone. Then $\int_0^1 |f(s) - t|\,\mathrm{d}s$ is minimal if and only if*

$$\min\left\{f\left(\tfrac{1}{2}-\right), f\left(\tfrac{1}{2}+\right)\right\} \leq t \leq \max\left\{f\left(\tfrac{1}{2}-\right), f\left(\tfrac{1}{2}+\right)\right\};$$

in particular, $\int_0^1 |f(s) - t|\,\mathrm{d}s \geq \int_0^1 \left|f(s) - f\left(\frac{1}{2}\right)\right| \mathrm{d}s$ for every $t \in \mathbb{R}$.

For a second observation, given any $c \in \mathbb{R}$ and $N \in \mathbb{N}$, let

$$s_0 = 0\,, \quad s_n = \frac{\langle c\rangle + n - 1}{N} \ \ (n = 1, \ldots, N)\,, \quad s_{N+1} = 1\,. \tag{3.3}$$

With this, for every (monotone) $f : [0,1] \to \mathbb{R}$ consider the piecewise constant function $\mathcal{F}_{c,N}^{f}$ given by

$$\mathcal{F}_{c,N}^{f}(s) = \sum\nolimits_{n=0}^{N} f\left(\tfrac{1}{2}(s_n + s_{n+1})\right) \mathbf{1}_{[s_n, s_{n+1}[}(s) \quad \forall s \in [0,1[\,; \tag{3.4}$$

for convenience, let $\mathcal{F}_{c,N}^{f}(1) = \mathcal{F}_{c,N}^{f}(1-)$. Note that $\int_0^1 \mathcal{F}_{c,N}^{f}(s)\,\mathrm{d}s$ is a Riemann sum for $\int_0^1 f(s)\,\mathrm{d}s$. The following lemma is a statement about the rate of convergence for such Riemann sums; while the elementary (albeit somewhat lengthy) proof of the lemma is included for completeness, it is deferred to Appendix B for the reader's convenience.

Lemma 3.4. *Given any monotone function $f : [0,1] \to \mathbb{R}$ and any sequence (c_n) in $\mathbb{R}$, the following are equivalent:*

(i) $\lim_{N\to\infty} N \int_0^1 |f(s) - \mathcal{F}^f_{c_N,N}(s)|\,\mathrm{d}s = 0$;
(ii) *there exists $a \in [0,1[$ such that f is constant on $]0,a[$ and $]a,1[$, and either $a = 0$ or else* $\lim_{N\to\infty} d_{\mathbb{T}}(Na, c_N) = 0$.

Proof of Theorem 3.2. For convenience, let $f = F_\mu^{-1}$ and $\nu = \mu^{-1}$. Recall that f is right-continuous and non-decreasing, with $0 \le f(0) \le f(1-) \le 1$, and $F_\nu = f - f(0-)$.

To prove (i), observe that with the appropriate $u_N \in \mathbb{R}$,

$$d_{\mathbb{T}}(\omega_N \circ R_{t_N}^{-1}, \mu) = \int_0^1 |F_\mu(s) - F_{\omega_N}(s - t_N) - u_N|\,\mathrm{d}s = \int_0^1 |f(s) - F^{-1}_{\omega_N}(s - u_N) - t_N|\,\mathrm{d}s\,,$$

where the second equal sign is due to Fubini's theorem and the fact that the integrand is 1-periodic. Note that $F^{-1}_{\omega_N}(\cdot - u_N) - t_N$ is constant on the interval $](k + Nu_N)/N, (k + 1 + Nu_N)/N[$ for every $k \in \mathbb{Z}$. Thus, with $c = Nu_N$ and $s_0, s_1, \ldots, s_{N+1}$ as in (3.3),

$$\begin{aligned} d_{\mathbb{T}}(\omega_N \circ R_{t_N}^{-1}, \mu) &= \sum\nolimits_{n=0}^{N} \int_{s_n}^{s_{n+1}} \left| f(s) - F^{-1}_{\omega_N}\left(\tfrac{1}{2}(s_n + s_{n+1}) - u_N\right) - t_N \right| \mathrm{d}s \\ &\ge \sum\nolimits_{n=0}^{N} \int_{s_n}^{s_{n+1}} \left| f(s) - f\left(\tfrac{1}{2}(s_n + s_{n+1})\right) \right| \mathrm{d}s = \int_0^1 \left| f(s) - \mathcal{F}^f_{Nu_N,N}(s) \right| \mathrm{d}s\,; \end{aligned}$$

here the inequality is due to Proposition 3.3 applied, in appropriately rescaled form, to each interval $[s_n, s_{n+1}]$ individually. By Lemma 3.4, $\limsup_{N\to\infty} N d_{\mathbb{T}}(\omega_N \circ R_{t_N}^{-1}, \mu) > 0$ unless $f = F_\mu^{-1}$ is constant on $]0,a[$ and $]a,1[$ for some $a \in [0,1[$, and so is F_ν. To further analyze the latter possibility, assume first that $a = 0$. In this case, $F_\nu(s) = 1$ for all $s \in [0,1[$, that is, $\nu = \delta_0$, and consequently $\mu = \delta_x$ for some $x \in \mathbb{T}$. From

$$N d_{\mathbb{T}}(\omega_N \circ R_{t_N}^{-1}, \delta_x) = \sum\nolimits_{n=1}^{N} d_{\mathbb{T}}(x_n + t_N, x) \qquad \forall N \in \mathbb{N}\,,$$

it is clear that $\limsup_{N\to\infty} N d_{\mathbb{T}}(\omega_N \circ R_{t_N}^{-1}, \mu) > 0$ unless $\lim_{N\to\infty} \sum_{n=1}^{N} d_{\mathbb{T}}(x_n + t_N, x) = 0$. Next assume that $a \in]0,1[$, where $F_\nu(s) = b$ for some $b \in [0,1[$ and all $s \in]0,a[$. Since $\nu \ne \delta_{a+\mathbb{Z}}$, necessarily $b > 0$, and $\nu = \mu^{-1} = b\delta_0 + (1-b)\delta_a$. With the appropriate $t \in \mathbb{R}$, therefore, $\mu = ((1-a)\delta_0 + a\delta_b) \circ R_t^{-1}$. Letting $\overline{\omega}_N = \omega_N \circ R^{-1}_{t_N - t}$ for convenience, notice that $NF_{\overline{\omega}_N}(s)$ is an integer for every s. Thus, with the appropriate $\overline{u}_N \in \mathbb{R}$ for every N,

$$\begin{aligned} N d_{\mathbb{T}}(\omega_N \circ R_{t_N}^{-1}, \mu) = N d_{\mathbb{T}}(\overline{\omega}_N, \mu \circ R_{-t}^{-1}) &= N \int_0^1 \left| F_{\mu\circ R_{-t}^{-1}}(s) - F_{\overline{\omega}_N}(s) - \overline{u}_N \right| \mathrm{d}s \\ &= N \int_0^b |1 - a - F_{\overline{\omega}_N}(s) - \overline{u}_N|\,\mathrm{d}s + N \int_b^1 |1 - F_{\overline{\omega}_N}(s) - \overline{u}_N|\,\mathrm{d}s \\ &\ge b \min\nolimits_{k\in\mathbb{Z}} |Na + N\overline{u}_N + k| + (1-b) \min\nolimits_{k\in\mathbb{Z}} |N\overline{u}_N + k| \\ &= b d_{\mathbb{T}}(Na, -N\overline{u}_N) + (1-b) d_{\mathbb{T}}(-N\overline{u}_N, 0) \\ &\ge \min\{b, 1-b\} d_{\mathbb{T}}(Na, 0)\,. \end{aligned}$$

Note that $\limsup_{N\to\infty} d_{\mathbb{T}}(Na, 0) \ge \frac{1}{3}$; see, e.g., [2, eq. (3.5)]. Thus

$$\limsup\nolimits_{N\to\infty} N d_{\mathbb{T}}(\omega_N \circ R_{t_N}^{-1}, \mu) \ge \tfrac{1}{3} \min\{b, 1-b\} > 0$$

in this case. In summary, $\limsup_{N\to\infty} Nd_{\mathbb{T}}(\omega_N \circ R_{t_N}^{-1}, \mu) > 0$ unless $\lim_{N\to\infty} \sum_{n=1}^{N} d_{\mathbb{T}}(x_n + t_N, x) = 0$ and $\mu = \delta_x$ for some $x \in \mathbb{T}$.

To prove (ii), recall that $\omega_N \circ R_{t_N}^{-1} \in \mathcal{P}$ is discrete, in fact $\#\operatorname{supp}(\omega_N \circ R_{t_N}^{-1}) \le N$. Thus $d_{\mathbb{T}}(\omega_N \circ R_{t_N}^{-1}, \mu) \ge d_{\mathbb{T}}(\nu_N, \mu)$ where ν_N denotes a (not necessarily unique) best approximation of μ in $\{\nu \in \mathcal{P} : \#\operatorname{supp}\nu \le N\}$. By [16, Thm. 1.4],

$$\lim_{N\to\infty} Nd_{\mathbb{T}}(\nu_N, \mu) = \frac{1}{4}\left(\int_{\mathbb{T}} \sqrt{\frac{\mathrm{d}\mu_{\mathsf{A}}}{\mathrm{d}\lambda_{\mathbb{T}}}}\, \mathrm{d}\lambda_{\mathbb{T}}\right)^2 ;$$

here μ_{A} denotes the absolutely continuous part of μ. Consequently,

$$\liminf_{N\to\infty} Nd_{\mathbb{T}}(\omega_N \circ R_{t_N}^{-1}, \mu) \ge \lim_{N\to\infty} Nd_{\mathbb{T}}(\nu_N, \mu) > 0\,,$$

provided that $\mu_{\mathsf{A}} \neq 0$, i.e., provided that μ is not singular. □

Corollary 3.5. *Let (x_n) be a sequence in $\mathbb{R}$, and $\mu \in \mathcal{P}$. Then $\limsup_{N\to\infty} Nd_{\mathbb{T}}(\widetilde{\omega}_N, \mu) > 0$ unless $x_n \in \mathbb{Z}$ for every n, and $\mu = \delta_0$.*

To put Theorem 3.2 in perspective, notice that $\liminf_{N\to\infty} Nd_{\mathbb{T}}(\omega_N \circ R_{t_N}^{-1}, \mu) = 0$ is possible even when μ has no atoms. For instance, if μ is the classical Cantor middle third distribution ("wrapped up" on $\mathbb{T}$) then $\liminf_{N\to\infty} N^r d_{\mathbb{T}}(\omega_N \circ R_{t_N}^{-1}, \mu) = 0$ for every $r < \log 3/\log 2 \approx 1.585$ and the appropriate sequences (x_n), (t_n). Also, the sequence $(Nd_{\mathbb{T}}(\omega_N \circ R_{t_N}^{-1}, \mu))$ may be barely unbounded, in that, for instance, $\limsup_{N\to\infty} Nd_{\mathbb{T}}(\omega_N \circ R_{t_N}^{-1}, \mu)/\sqrt{\log N}$ may be finite even when μ is absolutely continuous; see Example 4.3(i) below. However, it follows from [14, Thm. 3] that

$$\limsup_{N\to\infty} Nd_{\mathbb{T}}(\omega_N \circ R_{t_N}^{-1}, \mu) = \infty\,, \tag{3.5}$$

for every choice of (x_n), provided that $\sup_{n\in\mathbb{N}} nd_{\mathbb{T}}(t_n, 0) < \infty$ and μ is absolutely continuous with bounded density. Much more generally, the authors suspect that (3.5) *always* holds in the context of Theorem 3.2(ii), that is, for *all* sequences (x_n), (t_n) in $\mathbb{R}$ and *every* $\mu \in \mathcal{P}$ that is not singular. If correct, this would constitute a remarkable extension of the classical van Aardenne–Ehrenfest theorem [19, Cor. 2.2.1], with the latter following immediately by taking $\mu = \lambda_{\mathbb{T}}$ in (3.5); see also [25, Sec. 3.1]

4. Empirical distributions of log-like sequences

This section provides an analysis of the empirical distributions $\omega_N[x_n]$ associated with any real sequence (x_n) that satisfies (1.1). Notice that $x_n \equiv f(n)$ with an appropriate smooth (C^∞) function $f : \mathbb{R}^+ \to \mathbb{R}$ for which $\lim_{t\to\infty} tf'(t) = \lim_{n\to\infty} n(x_{n+1} - x_n)$. For convenience, then, let

$$\mathcal{F} = \{f : \mathbb{R}^+ \to \mathbb{R} \text{ is smooth and } \lim_{t\to\infty} tf'(t) \text{ exists (in } \mathbb{R})\}\,,$$

as well as $\mathcal{F}_0 = \{f \in \mathcal{F} : \lim_{t\to\infty} tf'(t) = 0\}$. With every $f \in \mathcal{F}$, associate the smooth function Δ^f given by $\Delta^f(t) = tf'(t) - b^f$, where $b^f = \lim_{t\to\infty} tf'(t)$. In a way, therefore, Δ^f measures the deviation of $f(t)$ from $b^f \log t$, and trivially $\lim_{t\to\infty} \Delta^f(t) = 0$. Given any $f \in \mathcal{F}$, write $\mathbb{A}[f(n)]$, $\omega_N[f(n)]$, $\widetilde{\omega}_N[f(n)]$, and $\Omega[f(n)]$ simply as $\mathbb{A}^f$, ω_N^f, $\widetilde{\omega}_N^f$, and Ω^f, respectively, i.e.,

$$\begin{aligned}
&\mathbb{A}^f = \{x \in \mathbb{T} : \exists\, N_1 < N_2 < \dots \text{ s.t. } d_{\mathbb{T}}(f(N_j), x) \to 0 \text{ as } j \to \infty\}\,;\\
&\omega_N^f = \frac{1}{N}\sum_{n=1}^{N} \delta_{f(n)+\mathbb{Z}}\,, \quad \widetilde{\omega}_N^f = \omega_N^f \circ R_{-f(N)}^{-1} = \frac{1}{N}\sum_{n=1}^{N} \delta_{f(n)-f(N)+\mathbb{Z}} \qquad \forall N \in \mathbb{N}\,;\\
&\Omega^f = \{\mu \in \mathcal{P} : \exists\, N_1 < N_2 < \dots \text{ s.t. } \omega_{N_j}^f \to \mu \text{ as } j \to \infty\}\,.
\end{aligned}$$

Note that $\mathbb{A}^f$ is a closed arc for every $f \in \mathcal{F}$. In fact, $\mathbb{A}^f = \mathbb{T}$ unless $f \in \mathcal{F}_0$; in the latter case, $\mathbb{A}^f$ may be smaller.

Informally put, this section aims to describe, for any given $f \in \mathcal{F}$ and as accurately as possible, the behaviour of ω_N^f as $N \to \infty$, as well as the structure of Ω^f. The quantity

$$D_N^f := d_{\mathbb{T}}(\widetilde{\omega}_N^f, \eta_{1/b^f}) = d_{\mathbb{T}}(\omega_N^f, \eta_{1/b^f} \circ R_{f(N)}^{-1}) \qquad \forall N \in \mathbb{N}\,,$$

is going to be of particular interest in this regard, with the convention $1/0 := \infty$ adhered to throughout. (Recall that $\eta_{1/b^f} \circ R_{f(N)}^{-1}$ denotes the push-forward of the probability measure η_a, with $a = 1/b^f$ and $b^f \in \mathbb{R}$, under the rotation $R_{f(N)}$. The reader may recognize D_N^f as a discrepancy-like quantity; see Remark 5.7 below.) Complementing and refining [17, 22, 30], the main result of this section provides a simple, explicit upper bound on D_N^f; in particular, it implies that $\lim_{N\to\infty} D_N^f = 0$ for every $f \in \mathcal{F}$.

Theorem 4.1. *Let $f : \mathbb{R}^+ \to \mathbb{R}$ be smooth, and assume that $f \in \mathcal{F}$. Then there exists $N_0 \in \mathbb{N}$ such that*

$$D_N^f \le \frac{1 + 4|b^f| \log N}{4N} + \frac{2}{N} \int_1^N |\Delta^f(t)|\, \mathrm{d}t \qquad \forall N \ge N_0\,. \tag{4.1}$$

Proof. Fix any $f \in \mathcal{F}$, and write Δ^f, b^f simply as Δ, b, respectively. Assume for the time being that $b \neq 0$. Since

$$d_{\mathbb{T}}(\widetilde{\omega}_N^f, \eta_{1/b}) = d_{\mathbb{T}}(\omega_N^f, \eta_{1/b} \circ R_{f(N)}^{-1}) = d_{\mathbb{T}}(\omega_N^{-f}, \eta_{-1/b} \circ R_{-f(N)}^{-1}) = d_{\mathbb{T}}(\widetilde{\omega}_N^{-f}, \eta_{-1/b}) \qquad \forall N \in \mathbb{N}\,,$$

it can be assumed that $b \in \mathbb{R}^+$. (Otherwise replace f by $-f$.) Pick $N_1 \in \mathbb{N}$ so large that $|\Delta(t)| \le \frac{1}{4}b$ for all $t \ge N_1$. Noting that $f'(t) > 0$ whenever $t \ge N_1$, denote by $g = g(t)$ the inverse of f, defined for all $t \ge f(N_1)$; for convenience, let $g(t) = N_1$ for all $t \le f(N_1)$. Recall that

$$f(t) = f(1) + b \log t + \int_1^t \frac{\Delta(s)}{s}\, \mathrm{d}s = f(N_1) - b \log N_1 + b \log t + \int_{N_1}^t \frac{\Delta(s)}{s}\, \mathrm{d}s \qquad \forall t \in \mathbb{R}^+\,.$$

Now, fix any $h \in \mathrm{Lip}_1(\mathbb{T})$, and observe that

$$\begin{aligned}
N \int_{\mathbb{T}} h\, \mathrm{d}(\eta_{1/b} \circ R_{f(N)}^{-1}) &= \frac{N}{b} \int_0^1 h(s + f(N)) \frac{e^{s/b}}{e^{1/b} - 1}\, \mathrm{d}s = \frac{N}{b} \int_0^1 h(s + f(N)) e^{s/b} \sum\nolimits_{j=1}^{\infty} e^{-j/b}\, \mathrm{d}s \\
&= \frac{N}{b} \sum\nolimits_{j=1}^{\infty} \int_{-j}^{-j+1} h(s + j + f(N)) e^{s/b}\, \mathrm{d}s = \frac{N}{b} \int_{-\infty}^{0} h(s + f(N)) e^{s/b}\, \mathrm{d}s \\
&= \frac{N}{b} e^{-f(N)/b} \int_{-\infty}^{f(N)} h(s) e^{s/b}\, \mathrm{d}s = N e^{-f(N)/b} \left(h(s) e^{s/b} \Big|_{s=-\infty}^{f(N)} - \int_{-\infty}^{f(N)} h'(s) e^{s/b}\, \mathrm{d}s \right) \\
&= N h \circ f(N) - N e^{-f(N)/b} \int_{-\infty}^{f(N)} h'(s) e^{s/b}\, \mathrm{d}s\,.
\end{aligned}$$

Moreover, by Euler summation for every $N \ge N_1$,

$$\begin{aligned}
N \int_{\mathbb{T}} h\, \mathrm{d}\omega_N^f &= \sum\nolimits_{n=1}^{N} h \circ f(n) \\
&= \sum\nolimits_{n=1}^{N_1 - 1} h \circ f(n) + \tfrac{1}{2}(h \circ f(N_1) + h \circ f(N)) + \int_{N_1}^{N} h \circ f(t)\, \mathrm{d}t + \int_{N_1}^{N} h' \circ f(t) f'(t) \left(\langle t \rangle - \tfrac{1}{2} \right) \mathrm{d}t \\
&= N h \circ f(N) - \int_{f(N_1)}^{f(N)} h'(s) g(s)\, \mathrm{d}s + \int_{f(N_1)}^{f(N)} h'(s) \left(\langle g(s) \rangle - \tfrac{1}{2} \right) \mathrm{d}s + c_N\,;
\end{aligned}$$

here the number c_N depends on f and N_1, but is uniformly bounded, in that $|c_N| \le N_1$ for all $N \ge N_1$. Thus, for every $N \ge N_1$,

$$\begin{aligned} N\left(\int_{\mathbb{T}} h\,\mathrm{d}(\omega_N^f - \eta_{1/b}\circ R_{f(N)}^{-1})\right) &= \\ &= Ne^{-f(N)/b}\int_{-\infty}^{f(N)} h'(s)e^{s/b}\,\mathrm{d}s - \int_{f(N_1)}^{f(N)} h'(s)g(s)\,\mathrm{d}s + \int_{f(N_1)}^{f(N)} h'(s)\left(\langle g(s)\rangle - \tfrac{1}{2}\right)\mathrm{d}s + c_N \\ &= b_N\int_{-\infty}^{f(N_1)} h'(s)e^{s/b}\,\mathrm{d}s + \int_{f(N_1)}^{f(N)} h'(s)g_N(s)\,\mathrm{d}s + \int_{f(N_1)}^{f(N)} h'(s)\left(\langle g(s)\rangle - \tfrac{1}{2}\right)\mathrm{d}s + c_N\,, \end{aligned} \tag{4.2}$$

where, for convenience,

$$b_N = Ne^{-f(N)/b} = e^{-f(1)/b}e^{-1/b\int_1^N \Delta(s)/s\,\mathrm{d}s} > 0 \qquad \forall N \ge N_1\,,$$

and the smooth function $g_N : [f(N_1), f(N)] \to \mathbb{R}$ is given by

$$g_N(s) = Ne^{-f(N)/b}e^{s/b} - g(s) \qquad \forall N \ge N_1\,. \tag{4.3}$$

An upper bound is now going to be derived for each of the three (potentially) unbounded contributions in (4.2). To this end, first notice that if $\int_1^\infty |\Delta(t)|/t\,\mathrm{d}t < \infty$ then the sequence (b_N) is bounded, and hence $\lim_{N\to\infty} b_N/\log N = 0$. If, however, $\int_1^\infty |\Delta(t)|/t\,\mathrm{d}t = \infty$ then $\int_1^\infty |\Delta(t)|\,\mathrm{d}t = \infty$ as well, in which case it is readily seen that $\lim_{N\to\infty} b_N/\int_1^N |\Delta(t)|\,\mathrm{d}t = 0$. Thus, with the appropriate $N_2 \in \mathbb{N}$ independent of h,

$$\left|b_N\int_{-\infty}^{f(N_1)} h'(s)e^{s/b}\,\mathrm{d}s\right| \le \frac{b}{3}\log N + \frac{1}{3}\int_1^N |\Delta(t)|\,\mathrm{d}t \qquad \forall N \ge N_2\,.$$

Next, to derive an upper bound for the second term in (4.2), observe that $g_N(f(N)) = 0$, as well as

$$g_N'(s) = Ne^{-f(N)/b}\frac{e^{s/b}}{b} - \frac{g(s)}{b+\Delta\circ g(s)} = \frac{g_N(s)}{b} + \frac{g(s)\Delta\circ g(s)}{b^2+b\Delta\circ g(s)} \qquad \forall s \in [f(N_1), f(N)]\,,$$

so that consequently

$$g_N(s) = -\int_s^{f(N)} e^{(s-r)/b}\frac{g(r)\Delta\circ g(r)}{b^2+b\Delta\circ g(r)}\,\mathrm{d}r \qquad \forall s \in [f(N_1), f(N)]\,.$$

Reversing the order of integration yields

$$\begin{aligned} \left|\int_{f(N_1)}^{f(N)} h'(s)g_N(s)\,\mathrm{d}s\right| &= \left|-\int_{f(N_1)}^{f(N)} h'(s)\int_s^{f(N)} e^{(s-r)/b}\frac{g(r)\Delta\circ g(r)}{b^2+b\Delta\circ g(r)}\,\mathrm{d}r\,\mathrm{d}s\right| \\ &= \left|\int_{f(N_1)}^{f(N)} \frac{g(r)\Delta\circ g(r)}{b^2+b\Delta\circ g(r)}\int_{f(N_1)}^{r} h'(s)e^{(s-r)/b}\,\mathrm{d}s\,\mathrm{d}r\right| \\ &\le \int_{f(N_1)}^{f(N)} \frac{g(r)|\Delta|\circ g(r)}{b-|\Delta|\circ g(r)}\left|\int_{f(N_1)}^{r} h'(s)\frac{e^{(s-r)/b}}{b}\,\mathrm{d}s\right|\mathrm{d}r \\ &\le \int_{f(N_1)}^{f(N)} \frac{g(r)|\Delta|\circ g(r)}{b-|\Delta|\circ g(r)}\,\mathrm{d}r = \int_{N_1}^{N} |\Delta(t)|\frac{b+\Delta(t)}{b-|\Delta(t)|}\,\mathrm{d}t \\ &\le \frac{b+b/4}{b-b/4}\int_{N_1}^{N}|\Delta(t)|\,\mathrm{d}t = \frac{5}{3}\int_{N_1}^{N}|\Delta(t)|\,\mathrm{d}t\,. \end{aligned}$$

Finally, to derive an upper bound for the remaining terms in (4.2), simply notice that, with the appropriate $N_3 \in \mathbb{N}$ independent of h,

$$\left| \int_{f(N_1)}^{f(N)} h'(s) \left(\langle g(s) \rangle - \tfrac{1}{2} \right) \mathrm{d}s + c_N \right| \leq \frac{f(N) - f(N_1)}{2} + |c_N| \leq \frac{2b}{3} \log N \qquad \forall N \geq N_3 \,.$$

Combining all these estimates back into (4.2) yields

$$\begin{aligned} \left| N \int_{\mathbb{T}} h \, \mathrm{d}(\widetilde{\omega}_N^f - \eta_{1/b}) \right| &\leq \frac{b}{3} \log N + \frac{1}{3} \int_1^N |\Delta(t)| \, \mathrm{d}t + \frac{5}{3} \int_1^N |\Delta(t)| \, \mathrm{d}t + \frac{2b}{3} \log N \\ &= b \log N + 2 \int_1^N |\Delta(t)| \, \mathrm{d}t \qquad \forall N \geq \max\{N_1, N_2, N_3\} \,. \end{aligned}$$

Since $h \in \mathrm{Lip}_1(\mathbb{T})$ has been arbitrary, it follows that with $N_0 = \max\{N_1, N_2, N_3\}$,

$$N D_N^f = N d_{\mathbb{T}}(\widetilde{\omega}_N^f, \eta_{1/b}) \leq b \log N + 2 \int_1^N |\Delta(t)| \, \mathrm{d}t \qquad \forall N \geq N_0 \,,$$

which proves (4.1) for $b > 0$.

It remains to consider the case $b = 0$ where simply $\eta_{1/b} \circ R_{f(N)}^{-1} = \delta_{f(N)}$. As before, fix $h \in \mathrm{Lip}_1(\mathbb{T})$, and deduce from Euler summation that, for every $N \in \mathbb{N}$,

$$\begin{aligned} N \int_{\mathbb{T}} h \, \mathrm{d}\omega_N^f &= \tfrac{1}{2}(h \circ f(1) + h \circ f(N)) + \int_1^N h \circ f(t) \, \mathrm{d}t + \int_1^N h' \circ f(t) f'(t) \left(\langle t \rangle - \tfrac{1}{2} \right) \mathrm{d}t \\ &= \tfrac{1}{2}(h \circ f(N) - h \circ f(1)) + N h \circ f(N) - \int_1^N t h' \circ f(t) f'(t) \, \mathrm{d}t + \int_1^N h' \circ f(t) f'(t) \left(\langle t \rangle - \tfrac{1}{2} \right) \mathrm{d}t \,, \end{aligned}$$

whereas clearly $\int_{\mathbb{T}} h \, \mathrm{d}\delta_{f(N)} = h \circ f(N)$. Thus for every $N \in \mathbb{N}$,

$$\begin{aligned} \left| N \int_{\mathbb{T}} h \, \mathrm{d}(\widetilde{\omega}_N^f - \delta_0) \right| &\leq \frac{1}{2} d_{\mathbb{T}}(f(N), f(1)) + \int_1^N \left(t + \frac{1}{2} \right) |f'(t)| \, \mathrm{d}t \\ &\leq \frac{1}{4} + \int_1^N \left(1 + \frac{1}{2t} \right) |\Delta(t)| \, \mathrm{d}t \leq \frac{1}{4} + \frac{3}{2} \int_1^N |\Delta(t)| \, \mathrm{d}t \,, \end{aligned}$$

which proves (4.1) for $b = 0$. □

Theorem 4.1 immediately yields Ω^f for every $f \in \mathcal{F}$. In essence, this has been established already in [17, Thm. 9], albeit via a very different approach.

Corollary 4.2. *Let (x_n) be a sequence in $\mathbb{R}$ that satisfies* (1.1)*. Then:*

(i) *$\mathbb{A} \neq \varnothing$ is a closed arc, and $\mathbb{A} = \mathbb{T}$ whenever $b \neq 0$;*
(ii) *$\Omega = \{\eta_{1/b} \circ R_x^{-1} : x \in \mathbb{A}\}$;*
(iii) *for any increasing sequence (N_j) in $\mathbb{N}$ and $t \in \mathbb{R}$, $\omega_{N_j} \to \eta_{1/b} \circ R_t^{-1}$ as $j \to \infty$ if and only if $\lim_{j\to\infty} d_{\mathbb{T}}(x_{N_j}, t) = 0$.*

The following examples illustrate Theorem 4.1 and Corollary 4.2.

Example 4.3. (i) Let $x_n = \log n$ for all $n \in \mathbb{N}$. Taking $f(t) = \log t$, clearly $f \in \mathcal{F}$, with $b^f = 1$ and $\Delta^f = 0$. By Theorem 4.1,

$$\limsup_{N\to\infty} \frac{N d_{\mathbb{T}}(\widetilde{\omega}_N, \eta_1)}{\log N} \leq 1\,,$$

whereas $\liminf_{N\to\infty} N d_{\mathbb{T}}(\widetilde{\omega}_N, \eta_1) > 0$ by Theorem 3.2(ii). A careful analysis shows that in fact

$$\lim_{N\to\infty} \frac{N d_{\mathbb{T}}(\widetilde{\omega}_N, \eta_1)}{\sqrt{\log N}} = \frac{1}{\sqrt{6\pi}}\,;$$

see [1, Thm. 3.1] and Remark 5.2 below.

(ii) Let $x_n = \log\log n$ for all $n \geq 3$. With $f(t) = \log\log t$ for all $t \geq e$, clearly $f \in \mathcal{F}_0$ and $\Delta^f(t) = 1/\log t$. Since

$$\lim_{N\to\infty} \frac{\log N}{N} \int_e^N |\Delta^f(t)|\,\mathrm{d}t = 1\,,$$

it now follows from Theorem 4.1 that

$$\limsup_{N\to\infty} \log N\, d_{\mathbb{T}}(\widetilde{\omega}_N, \delta_0) \leq 2\,.$$

Unlike in (i), this bound on the rate of convergence of $d_{\mathbb{T}}(\widetilde{\omega}_N, \delta_0) \to 0$ is sharp as $N \to \infty$ up to a factor 2: Indeed, it is straightforward to show that

$$\lim_{N\to\infty} \log N\, d_{\mathbb{T}}(\widetilde{\omega}_N, \delta_0) = 1\,.$$

Note that $\mathbb{A} = \mathbb{T}$ and hence $\Omega = \{\delta_x : x \in \mathbb{T}\}$ even though $b = 0$.

(iii) Let $x_n = e^{-n}$ for all $n \in \mathbb{N}$. Here $f(t) = e^{-t}$, so $f \in \mathcal{F}_0$ and $\Delta^f(t) = -te^{-t}$. Theorem 4.1 now yields

$$\limsup_{N\to\infty} N d_{\mathbb{T}}(\widetilde{\omega}_N, \delta_0) \leq \frac{e+16}{4e} \approx 1.722\,.$$

By contrast, Theorem 3.2(i) implies that $\limsup_{N\to\infty} N d_{\mathbb{T}}(\widetilde{\omega}_N, \delta_0) > 0$, whereas Theorem 3.2(ii) does not apply. Still, just like in (ii), the bound provided by (4.1) is essentially sharp, since a short calculation shows that

$$\lim_{N\to\infty} N d_{\mathbb{T}}(\widetilde{\omega}_N, \delta_0) = \frac{1}{e-1} \approx 0.5820\,.$$

Unlike in (ii), however, $\mathbb{A} = \{0\}$ and hence $\Omega = \{\delta_0\}$.

5. Examples and extensions

This concluding section discusses Theorem 4.1 by means of a few concrete examples and possible extensions. While the discussion is quite succinct, the interested reader will have no difficulty filling in the omitted elementary details as desired.

For a first illustrative example, fix any $b, c \in \mathbb{R} \setminus \{0\}$ and let $f : \mathbb{R}^+ \to \mathbb{R}$ be smooth with

$$f(t) = b\log t + c\log\log t \qquad \forall t \geq e\,. \tag{5.1}$$

Clearly, $f \in \mathcal{F} \setminus \mathcal{F}_0$, $b^f = b$, and $\Delta^f(t) = c/\log t$ for $t \geq e$. Deduce from Theorem 4.1 that

$$\limsup_{N\to\infty} \log N\, D_N^f \leq \lim_{N\to\infty} \frac{2|c|\log N}{N} \int_e^N \frac{\mathrm{d}t}{\log t} = 2|c|\,.$$

As the following theorem shows, in this example $\lim_{N\to\infty} \log N\, D_N^f$ is finite and positive, and so the seemingly rough bound (4.1) is essentially sharp, that is, sharp as $N \to \infty$ up to a positive constant.

Theorem 5.1. *Let $x_n = b\log n + c\log\log n$ for some $b, c \in \mathbb{R} \setminus \{0\}$ and all $n \geq 2$. Then $\Omega = \{\eta_{1/b} \circ R_x^{-1} : x \in \mathbb{T}\}$. Moreover,*

$$\lim_{N\to\infty} \log N \int_{\mathbb{T}} h\,\mathrm{d}(\widetilde{\omega}_N - \eta_{1/b}) = \frac{c}{b} \int_{\mathbb{T}} h\,\mathrm{d}(\zeta_{1/b} - \eta_{1/b}) \qquad \forall h \in \mathrm{Lip}(\mathbb{T})\,, \tag{5.2}$$

and the convergence in (5.2) *is uniform on* $\mathrm{Lip}_1(\mathbb{T})$*; in particular,*

$$\lim_{N\to\infty} \log N\, d_{\mathbb{T}}(\widetilde{\omega}_N, \eta_{1/b}) = \frac{|c|}{|b|}\, d_{\mathbb{T}}(\zeta_{1/b}, \eta_{1/b}) = |c| \left(\frac{e^{\frac{1}{4b}} \tanh\frac{1}{4b}}{\cosh\frac{1}{4b}} - 1 \right). \tag{5.3}$$

Proof. Since $\lim_{t\to\infty} tf'(t) = b$, the assertion regarding Ω is immediate from Corollary 4.2(ii). Henceforth, assume w.l.o.g. that $b > 0$, and with $N_1 \geq 3$, $f(N_1) > 1$, and g as in the proof of Theorem 4.1, note that

$$g(s) = e^{s/b} \left(\frac{s}{b} - \frac{c}{b}\log\frac{s}{b} + \widetilde{g}(s)\frac{\log s}{s} \right)^{-c/b} \qquad \forall s \geq f(N_1)\,, \tag{5.4}$$

where $\widetilde{g}$ denotes a *bounded* smooth function; by choosing a larger N_1 if necessary, it can be assumed that

$$\frac{s}{b} - \frac{c}{b}\log\frac{s}{b} + \widetilde{g}(s)\frac{\log s}{s} \geq 1 \qquad \forall s \geq f(N_1)\,.$$

Letting $\ell_N = \log N$ for convenience, deduce from (4.2) that, given any $h \in \mathrm{Lip}_1(\mathbb{T})$, and with the appropriate $N_2 \geq N_1$ independent of h,

$$N \int_{\mathbb{T}} h\,\mathrm{d}(\omega_N^f - \eta_{1/b} \circ R_{f(N)}^{-1}) = a_N \ell_N^{1+|c|/b} + \int_0^{f(N)-f(N_1)} h'(f(N)-s) g_N(f(N)-s)\,\mathrm{d}s \qquad \forall N \geq N_2\,;$$

here the number $a_N \in \mathbb{R}$ depends on h, but $|a_N| \leq 1$ for all $N \geq N_2$. Now, recalling the definition of g_N from (4.3), notice that

$$\frac{g_N(f(N)-s)}{N} = e^{-s/b} - \frac{g(f(N)-s)}{N} \qquad \forall N \geq N_1,\, s \in [0,\, f(N) - f(N_1)]\,.$$

Moreover, from (5.1) and (5.4), it is readily seen that, with the appropriate $N_3 \geq N_2$,

$$\frac{g(f(N)-s)}{N} = e^{-s/b} \left(1 - \frac{s}{b\ell_N} + \frac{\widetilde{g}_N(s)}{\ell_N^{4/3}} \right)^{-c/b} \qquad \forall N \geq N_3,\, s \in [0,\, \sqrt{\ell_N}]\,,$$

where the function $\widetilde{g}_N$ is smooth with $|\widetilde{g}_N(s)| \leq 1$, but also

$$\left| \frac{g(f(N)-s)}{N} \right| \leq e^{-s/b} \ell_N^{1+|c|/b} \qquad \forall N \geq N_3,\, s \in [\,\sqrt{\ell_N},\, f(N) - f(N_1)]\,.$$

Combining these estimates yields, with the appropriate $N_4 \geq N_3$ and $|\widetilde{a}_N| \leq 3b$,

$$\begin{aligned}\ell_N \int_{\mathbb{T}} h \,\mathrm{d}(\omega_N^f - \eta_{1/b} \circ R_{f(N)}^{-1}) \\ = \widetilde{a}_N \ell_N^{2+|c|/b} e^{-\sqrt{\ell_n}/b} + \ell_N \int_0^{\sqrt{\ell_N}} e^{-s/b} \left(1 - \left(1 - \frac{s}{b\ell_N} + \frac{\widetilde{g}_N(s)}{\ell_N^{4/3}}\right)^{-c/b}\right) h'(f(N) - s)\,\mathrm{d}s\end{aligned}$$

for all $N \geq N_4$, from which it is straightforward to deduce that

$$\lim_{N\to\infty} \left(\ell_N \int_{\mathbb{T}} h \,\mathrm{d}(\omega_N^f - \eta_{1/b} \circ R_{f(N)}^{-1}) + \frac{c}{b^2} \int_0^\infty s e^{-s/b} h'(f(N) - s)\,\mathrm{d}s\right) = 0\,,$$

uniformly in $h \in \mathrm{Lip}_1(\mathbb{T})$. Furthermore, observe that

$$\begin{aligned}\int_0^\infty s e^{-s/b} h'(f(N) - s)\,\mathrm{d}s &= \int_0^1 e^{-s/b} h'(f(N) - s) \sum\nolimits_{n=0}^\infty (s+n) e^{-n/b} \mathrm{d}s \\ &= \frac{e^{1/b}}{(e^{1/b} - 1)^2} \int_0^1 e^{-s/b} (s(e^{1/b} - 1) + 1) h'(f(N) - s)\,\mathrm{d}s \\ &= -b \int_{\mathbb{T}} h \,\mathrm{d}((\zeta_{1/b} - \eta_{1/b}) \circ R_{f(N)}^{-1})\,.\end{aligned}$$

For f given by (5.1), therefore, uniformly in $h \in \mathrm{Lip}_1(\mathbb{T})$,

$$\lim_{N\to\infty} \log N \int_{\mathbb{T}} h \,\mathrm{d}(\widetilde{\omega}_N^f - \eta_{1/b}) = \frac{c}{b} \int_{\mathbb{T}} h \,\mathrm{d}(\zeta_{1/b} - \eta_{1/b})\,.$$

This proves (5.2), and since the latter holds uniformly on $\mathrm{Lip}_1(\mathbb{T})$, also (5.3), by (2.1) and (2.6). □

Remark 5.2. As the reader may suspect, the conclusion of Theorem 5.1 is somewhat different in the (excluded) case $c = 0$ for which (4.1) simply yields $\limsup_{N\to\infty} N D_N^f / \log N \leq |b^f|$: It is shown in [1] that if $f \in \mathcal{F}$ satisfies $\int_1^\infty |\Delta^f(t)|\,\mathrm{d}t < \infty$ then

$$\lim_{N\to\infty} \frac{N D_N^f}{\sqrt{\log N}} = \sqrt{\frac{|b^f|}{6\pi}\left(1 + 2V(e^{1/|b^f|})\right)}, \tag{5.5}$$

where the function $V :]1, \infty] \to [0, 1] \cap \mathbb{Q}$ is given by

$$V(t) = \begin{cases} 0 & \text{if } \{t^n : n \in \mathbb{N}\} \cap \mathbb{Q} = \varnothing\,, \\ \dfrac{1}{pq - 1} & \text{if } \{t^n : n \in \mathbb{N}\} \cap \mathbb{Q} = \{(p/q)^j : j \in \mathbb{N}\} \text{ with } p, q \in \mathbb{N} \text{ and } \gcd(p, q) = 1\,. \end{cases}$$

In this case, therefore, the right-hand side in (4.1) is too pessimistic as $N \to \infty$ by a factor $\sqrt{\log N}$. Moreover, utilizing (4.2) with $g_N(s) \equiv 0$, it is readily seen that for f given by (5.1) with $c = 0$,

$$\lim_{N\to\infty} \frac{N}{\sqrt{\log N}} \int_{\mathbb{T}} h \,\mathrm{d}(\widetilde{\omega}_N^f - \eta_{1/b}) = 0 \qquad \forall h \in \mathrm{Lip}(\mathbb{T})\,; \tag{5.6}$$

unlike in (5.2), however, the convergence here is *not* uniform on $\mathrm{Lip}_1(\mathbb{T})$. Thus, whereas the rate of convergence in (5.3) reflects what one may regard as the *deviation error*, corresponding to the right-most term in (4.1), the rate of convergence in (5.5) reflects the *discretization error*, caused by approximating the (absolutely) continuous probability measure $\eta_{1/b}$ with a discrete one, namely $\widetilde{\omega}_N^f$. Notice furthermore that whereas the deviation error is "biased", in the sense that asymptotically it is proportional to $\zeta_{1/b} - \eta_{1/b} \neq 0$, by (5.2), the discretization error is "unbiased", by (5.6). Clearly, a certain discretization error $\widetilde{\omega}_N^f - \eta_{1/b}$ is unavoidable whenever $f \in \mathcal{F} \setminus \mathcal{F}_0$. However, the authors conjecture that the (rate of convergence for the) discretization error appearing in (5.5) is the smallest it can possibly be; more precisely, they conjecture that

$$\limsup_{N\to\infty} \frac{N}{\sqrt{\log N}}\, d_{\mathbb{T}}(\widetilde{\omega}_N^f, \mu) > 0 \qquad \forall\, f \in \mathcal{F} \setminus \mathcal{F}_0, \mu \in \mathcal{P}\,. \tag{5.7}$$

If indeed correct, (5.7) would significantly strengthen (3.5) in the special case where $x_n = -t_n = f(n)$ for all n, with any $f \in \mathcal{F} \setminus \mathcal{F}_0$.

For a second illustrative example, fix again $b, c \in \mathbb{R} \setminus \{0\}$, as well as $0 < r < 1$, and consider $f : \mathbb{R}^+ \to \mathbb{R}$ given by

$$f(t) = b \log t + ct^{-r} \qquad \forall t \in \mathbb{R}^+\,, \tag{5.8}$$

so $f \in \mathcal{F}$, $b^f = b$, and $\Delta^f(t) = -crt^{-r}$. Here Theorem 4.1 yields

$$\limsup_{N\to\infty} N^r D_N^f \leq \lim_{N\to\infty} \frac{2|c|rN^r}{N} \int_1^N t^{-r}\mathrm{d}t = \frac{2|c|r}{1-r}\,.$$

As in the case of (5.1), it turns out that (4.1) is essentially sharp in this example also. The proof of the following result is completely analogous to, yet somewhat simpler than the proof of Theorem 5.1, and hence the details are left to the interested reader.

Proposition 5.3. *Let $x_n = b \log n + cn^{-r}$ for some $b, c \in \mathbb{R} \setminus \{0\}$, $0 < r < 1$, and all $n \in \mathbb{N}$. Then $\Omega = \{\eta_{1/b} \circ R_x^{-1} : x \in \mathbb{T}\}$. Moreover,*

$$\lim_{N\to\infty} N^r \int_{\mathbb{T}} h\,\mathrm{d}(\widetilde{\omega}_N - \eta_{1/b}) = -\frac{c}{b} \int_{\mathbb{T}} h\,\mathrm{d}(\eta_{(1-r)/b} - \eta_{1/b}) \qquad \forall h \in \mathrm{Lip}(\mathbb{T})\,, \tag{5.9}$$

and the convergence in (5.9) *is uniform on* $\mathrm{Lip}_1(\mathbb{T})$*; in particular,*

$$\lim_{N\to\infty} N^r d_{\mathbb{T}}(\widetilde{\omega}_N, \eta_{1/b}) = \frac{|c|}{|b|}\, d_{\mathbb{T}}(\eta_{(1-r)/b}, \eta_{1/b})\,.$$

Remark 5.4. (i) In essence, (5.2) and (5.9) both assert that

$$\lim_{N\to\infty} \frac{1}{|\Delta^f(N)|} \int_{\mathbb{T}} h\,\mathrm{d}(\widetilde{\omega}_N^f - \eta_{1/b^f}) = a \int_{\mathbb{T}} h\,\mathrm{d}(\mu - \eta_{1/b^f}) \qquad \forall h \in \mathrm{Lip}(\mathbb{T})\,, \tag{5.10}$$

for the appropriate constant $a \in \mathbb{R} \setminus \{0\}$ and $\mu \in \mathcal{P}$ with $\mu \neq \eta_{1/b^f}$. As a consequence,

$$\lim_{N\to\infty} \frac{D_N^f}{|\Delta^f(N)|} = |a| d_{\mathbb{T}}(\mu, \eta_{1/b^f}) \in \mathbb{R}^+\,.$$

Thus, $D_N^f \to 0$ at precisely the same rate as $\Delta^f(N) \to 0$. Moreover, it is not hard to see that, as a complement to (5.10),

$$\lim_{N\to\infty} \frac{1}{|\Delta^f(N)|}(F_{\widetilde{\omega}_N} - F_{\eta_{1/bf}}) = a(F_{\zeta_{1/bf}} - F_{\eta_{1/bf}}) \tag{5.11}$$

uniformly on $\mathbb{R}$; see Figure 4. It seems plausible that (5.10) and (5.11) remain valid for any $f \in \mathcal{F} \setminus \mathcal{F}_0$, as long as $|\Delta^f|$ is regularly varying at ∞ with index larger than -1.

(ii) For f as in (5.8), but with $r \geq 1$, it can be shown that (5.5) holds; as noted in Remark 5.2, this is immediate from [1, Thm. 3.1] in case $r > 1$.

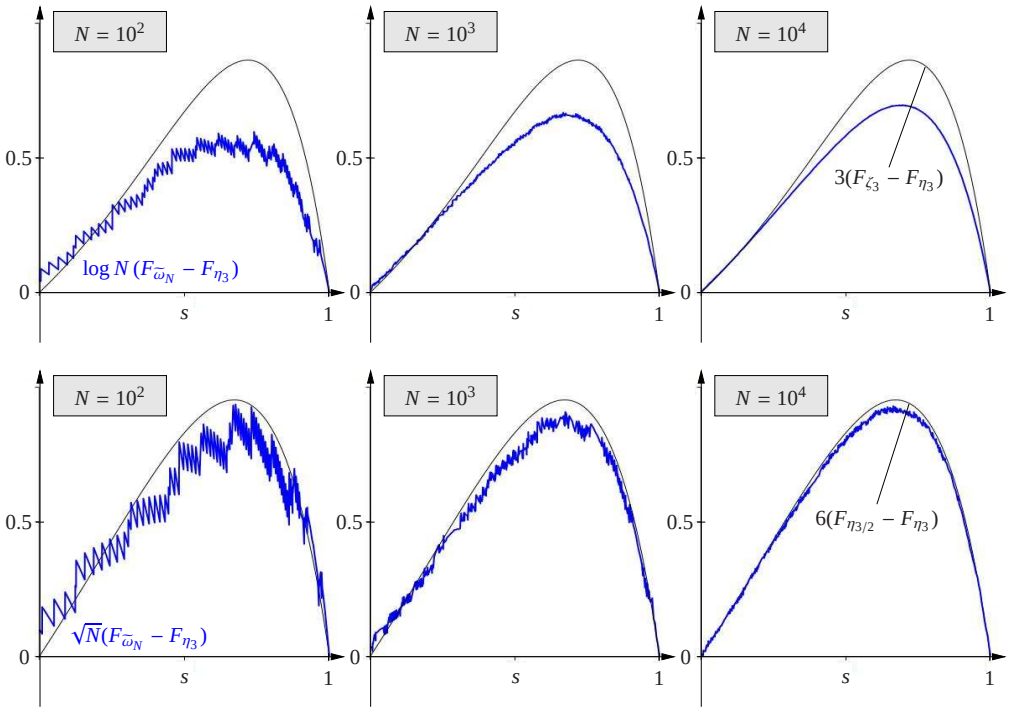


Figure 4: By Remark 5.4(i), the deviation $\log N\,(F_{\widetilde{\omega}_N} - F_{\eta_3})$ associated with the sequence $(x_n) = (\frac{1}{3}\log n + \log\log n)$ converges uniformly (top), and so does the deviation $\sqrt{N}(F_{\widetilde{\omega}_N} - F_{\eta_3})$ associated with $(x_n) = (\frac{1}{3}\log n - 2/\sqrt{n})$; see also Figure 2.

For a third illustrative example, denote by p_n the n-th prime number. The distributions mod 1 of the sequence $(\log p_n)$ have been studied by many authors; see, e.g., [15, 21, 22, 27, 29]. To appreciate this sequence within the context of the present article, notice first that known results about prime gaps [13, 32] imply

$$\liminf_{n\to\infty} n\log n\log\frac{p_{n+1}}{p_n} < \infty\,, \quad \limsup_{n\to\infty} \frac{n}{\sqrt{\log\log n}}\log\frac{p_{n+1}}{p_n} > 0\,,$$

and consequently

$$\liminf_{n\to\infty} n(\log p_{n+1} - \log p_n) = 0\,, \quad \limsup_{n\to\infty} n(\log p_{n+1} - \log p_n) = \infty\,.$$

Hence Corollary 4.2 does not apply directly to $(x_n) = (\log p_n)$. However, a conclusion very similar to that of Theorem 5.1 holds nonetheless.

Theorem 5.5. *Let $x_n = b \log p_n$ for some $b \in \mathbb{R} \setminus \{0\}$ and all $n \in \mathbb{N}$. Then $\Omega = \{\eta_{1/b} \circ R_x^{-1} : x \in \mathbb{T}\}$. Moreover,*

$$\lim_{N\to\infty} \log N \int_{\mathbb{T}} h \, \mathrm{d}(\widetilde{\omega}_N - \eta_{1/b}) = \int_{\mathbb{T}} h \, \mathrm{d}(\zeta_{1/b} - \eta_{1/b}) \qquad \forall h \in \mathrm{Lip}(\mathbb{T}) \,, \tag{5.12}$$

and the convergence in (5.12) *is uniform on* $\mathrm{Lip}_1(\mathbb{T})$*; in particular*

$$\lim_{N\to\infty} \log N \, d_{\mathbb{T}}(\widetilde{\omega}_N, \eta_{1/b}) = d_{\mathbb{T}}(\zeta_{1/b}, \eta_{1/b}) \,.$$

As presented here, the proof of Theorem 5.5 makes use of the following two quantitative refinements of the prime number theorem; see, e.g., [6, 11, 12] and the references therein. On the one hand, with $\pi(x) := \#\{n \in \mathbb{N} : p_n \le x\}$ for every $x \in \mathbb{R}$,

$$\pi(x) = \frac{x}{\log x}\left(1 + \frac{1}{\log x} + \frac{q(x)}{(\log x)^2}\right) \qquad \forall x \ge e \,, \tag{5.13}$$

where the function $q : [e, \infty[\to \mathbb{R}$ is bounded. On the other hand,

$$\frac{p_n}{n} = \log n + \log\log n - 1 + \frac{q_n \log\log n}{\log n} \qquad \forall n \ge 2 \,, \tag{5.14}$$

where the sequence (q_n) is bounded.

Proof of Theorem 5.5. Pick any smooth $f : \mathbb{R}^+ \to \mathbb{R}$ with

$$f(t) = b \log t + b \log(\log t + \log\log t - 1) \qquad \forall t \ge 2e \,; \tag{5.15}$$

to avoid confusing it with $\widetilde{\omega}_N$, denote $\omega_N^f \circ R_{-f(N)}^{-1}$ by μ_N throughout. Note that Theorem 4.1 applies, since obviously $f \in \mathcal{F}$, whereas $\widetilde{f} \notin \mathcal{F}$ for every smooth $\widetilde{f} : \mathbb{R}^+ \to \mathbb{R}$ with $\widetilde{f}(n) = b \log p_n$ for all $n \in \mathbb{N}$. By (5.14),

$$x_n - f(n) = b \log\left(1 + \frac{q_n \log\log n}{\log n(\log n + \log\log n - 1)}\right) \qquad \forall n \ge 6 \,,$$

and consequently, with the appropriate $a \in \mathbb{R}^+$ and $N_1 \in \mathbb{N}$,

$$|x_n - f(n)| \le a|b| \frac{\log\log n}{(\log n)^2} \qquad \forall n \ge N_1 \,.$$

With this, given any $h \in \mathrm{Lip}_1(\mathbb{T})$,

$$\begin{aligned}
\left|N \int_{\mathbb{T}} h \, \mathrm{d}(\widetilde{\omega}_N - \mu_N)\right| &= \left|\sum_{n=1}^N \Big(h(x_n - x_N) - h(f(n) - f(N))\Big)\right| \\
&\le \sum_{n=1}^N d_{\mathbb{T}}(x_n - x_N, f(n) - f(N)) = \sum_{n=1}^N d_{\mathbb{T}}(x_n - f(n), x_N - f(N)) \\
&\le \tfrac{1}{2}(N_1 - 1) + a|b| \sum_{n=N_1}^N \left(\frac{\log\log n}{(\log n)^2} + \frac{\log\log N}{(\log N)^2}\right) \qquad \forall N \ge N_1 \,.
\end{aligned}$$

Next, observe that

$$\lim_{N\to\infty} \frac{(\log N)^2}{N \log\log N} \sum_{n=2}^{N} \frac{\log\log n}{(\log n)^2} = 1\,.$$

This implies that, with the appropriate $N_2 \geq N_1$, independent of h,

$$\left| N \int_{\mathbb{T}} h \,\mathrm{d}(\widetilde{\omega}_N - \mu_N) \right| \leq 3a|b| \frac{N \log\log N}{(\log N)^2} \qquad \forall N \geq N_2\,,$$

and consequently, uniformly on $\mathrm{Lip}_1(\mathbb{T})$,

$$\lim_{N\to\infty} \log N \int_{\mathbb{T}} h \,\mathrm{d}(\widetilde{\omega}_N - \mu_N) = 0 \qquad \forall h \in \mathrm{Lip}(\mathbb{T})\,.$$

Thus, to complete the proof, it suffices to establish (5.12) with μ_N instead of $\widetilde{\omega}_N$. As in the proof of Theorem 5.1, assume w.l.o.g. that $b > 0$, and denote by $g = g(s)$ the inverse of f for all sufficiently large s. Deduce from (5.15) that, with the appropriate $s_1 \in \mathbb{R}^+$,

$$g(s) = e^{s/b} \left(\frac{s}{b} - 1 + \widetilde{g}(s) \frac{\log s}{s} \right)^{-1} \qquad \forall s \geq s_1\,,$$

where $\widetilde{g}$ denotes a *bounded* smooth function. Utilizing this analogue of (5.4), the remainder of the proof is now carried out in complete analogy to the proof of Theorem 5.1. □

Corollary 5.6. *Let $x_n = b \log p_n$ for some $b \in \mathbb{R} \setminus \{0\}$ and all $n \in \mathbb{N}$. Then, for every increasing sequence (N_j) in $\mathbb{N}$ with $\lim_{j\to\infty} \log N_j d_{\mathbb{T}}(x_{N_j}, 0) = 0$,*

$$\lim_{j\to\infty} \log N_j \int_{\mathbb{T}} h \,\mathrm{d}(\omega_{N_j} - \eta_{1/b}) = \int_{\mathbb{T}} h \,\mathrm{d}(\zeta_{1/b} - \eta_{1/b}) \qquad \forall h \in \mathrm{Lip}(\mathbb{T})\,,$$

and this convergence is uniform on $\mathrm{Lip}_1(\mathbb{T})$.

For a simple example illustrating Corollary 5.6, fix $b \in \mathbb{R}^+$ and let $N_j = \pi(e^{j/b})$ for all $j \in \mathbb{N}$. By (5.13),

$$\log N_j = \frac{j}{b} - \log j + \log b + O\left(j^{-1}\right) \quad \text{as } j \to \infty\,,$$

and (5.14) implies that

$$d_{\mathbb{T}}(x_{N_j}, 0) = O\left(\frac{(\log\log N_j)^2}{(\log N_j)^2} \right) \quad \text{as } j \to \infty\,.$$

Thus, by Corollary 5.6,

$$\lim_{j\to\infty} j \int_{\mathbb{T}} h \,\mathrm{d}(\omega_{N_j} - \eta_{1/b}) = b \int_{\mathbb{T}} h \,\mathrm{d}(\zeta_{1/b} - \eta_{1/b}) \qquad \forall h \in \mathrm{Lip}(\mathbb{T})\,.$$

Put informally, $\omega_{\pi(e^{j/b})}$ converges to $\eta_{1/b}$ at precisely the rate (j^{-1}), with the deviation $\omega_{\pi(e^{j/b})} - \eta_{1/b}$ being asymptotically proportional to $\zeta_{1/b} - \eta_{1/b}$.

Remark 5.7. (i) By Theorem 5.5, for $(x_n) = (b \log p_n)$ the (absolutely continuous) probability measure $\eta_{1/b} \circ R_{x_N}^{-1}$ is a reasonable approximation of the (discrete) probability measure ω_N as $N \to \infty$, since $\lim_{N\to\infty} d_{\mathbb{T}}(\omega_N, \eta_{1/b} \circ R_{x_N}^{-1}) = 0$. A slightly better approximation of ω_N is obtained as follows: By (2.7) and (5.12), and with $a_n := 1/b - 1/(b \log n)$ for all $n \geq 2$,

$$\lim_{N\to\infty} \log N \int_{\mathbb{T}} h \, \mathrm{d}(\widetilde{\omega}_N - \eta_{a_N}) = 0 \qquad \forall h \in \mathrm{Lip}(\mathbb{T}) ;$$

since this convergence is uniform on $\mathrm{Lip}_1(\mathbb{T})$,

$$\lim_{N\to\infty} \frac{d_{\mathbb{T}}(\omega_N, \eta_{a_N} \circ R_{x_N}^{-1})}{d_{\mathbb{T}}(\omega_N, \eta_{1/b} \circ R_{x_N}^{-1})} = \lim_{N\to\infty} \frac{d_{\mathbb{T}}(\widetilde{\omega}_N, \eta_{a_N})}{d_{\mathbb{T}}(\widetilde{\omega}_N, \eta_{1/b})} = 0 .$$

Thus, $\eta_{a_N} \circ R_{x_N}^{-1}$ is a better approximation of ω_N than is $\eta_{1/b} \circ R_{x_N}^{-1}$.

(ii) The results in this article use only the metric $d_{\mathbb{T}}$ on $\mathcal{P}$. As noted, e.g., in [8, 14, 25], the existing literature records relatively few precise quantitative facts regarding $d_{\mathbb{T}}$. By contrast, a vast literature exists studying the metric d_∞ on $\mathcal{P}$, given by

$$d_\infty(\mu, \nu) = \sup_{0\leq s<1} |F_\mu(s) - F_\nu(s)| \qquad \forall \mu, \nu \in \mathcal{P} .$$

Clearly, $d_{\mathbb{T}} \leq d_\infty$, and $(\mathcal{P}, d_\infty)$ is complete but not separable (let alone, compact), its topology being much finer than that of $(\mathcal{P}, d_{\mathbb{T}})$. Unlike $d_{\mathbb{T}}$, the metric d_∞ on $\mathcal{P}$ is not induced in a natural way by a metric on $\mathbb{T}$. Correspondingly, the maps Q and R_t do not induce d_∞-isometries. However, if $\mu \in \mathcal{P}$ has no atoms then $d_\infty(\mu_n, \mu) \to 0$ if and only if $d_{\mathbb{T}}(\mu_n, \mu) \to 0$. Thus, for instance, a sequence (x_n) is u.d. mod 1 precisely if $D_N^*(x_n) := d_\infty(\omega_N[x_n], \lambda_{\mathbb{T}}) \to 0$ as $N \to \infty$. The quantity $D_N^*(x_n)$, referred to as *star discrepancy* of (x_n), is a central object of study in uniform distribution theory [19, Ch. 2]. For $(x_n) = (b \log p_n)$ it has (in essence) been shown in [22] that $\limsup_{N\to\infty} \log N \, d_\infty(\widetilde{\omega}_N, \eta_{1/b}) < \infty$. From Theorem 5.5, it is clear that $\liminf_{N\to\infty} \log N \, d_\infty(\widetilde{\omega}_N, \eta_{1/b}) > 0$. It is unknown whether $\lim_{N\to\infty} \log N \, d_\infty(\widetilde{\omega}_N, \eta_{1/b})$ exists. In fact, the bottom part of Figure 2 suggests that it does not: If indeed $(\log N \, (F_{\widetilde{\omega}_N} - F_{\eta_{1/b}}))$ converges uniformly, to $F_{\zeta_{1/b}} - F_{\eta_{1/b}}$, then

$$2 \liminf_{N\to\infty} \log N \, d_\infty(\widetilde{\omega}_N, \eta_{1/b}) = \limsup_{N\to\infty} \log N \, d_\infty(\widetilde{\omega}_N, \eta_{1/b}) = d_\infty(\zeta_{1/b}, \eta_{1/b}) > 0 ,$$

so $d_\infty(\widetilde{\omega}_N, \eta_{1/b})$ behaves less regularly than $d_{\mathbb{T}}(\widetilde{\omega}_N, \eta_{1/b})$ as $N \to \infty$.

Finally, to put the examples of this section in perspective, briefly consider the case of a smooth $f : \mathbb{R}^+ \to \mathbb{R}$ with $\lim_{t\to\infty} t|f'(t)| = \infty$. With $1/\infty := 0$, note that $\eta_0 = \lambda_{\mathbb{T}}$ and $D_N^f = d_{\mathbb{T}}(\omega_N^f, \lambda_{\mathbb{T}})$, so Theorem 4.1 suggests that $\lim_{N\to\infty} D_N^f = 0$, i.e., the sequence $(f(n))$ is u.d. mod 1. As is well known, under mild assumptions on f this conclusion is indeed correct. For instance, a classical result by Fejér [19, Cor. 1.2.1] implies that if $f'(t) \to 0$ monotonically as $t \to \infty$ then $\lim_{N\to\infty} D_N^f = 0$. To relate this result to Theorem 4.1 and the examples discussed earlier, assume that

$$f : \mathbb{R}^+ \to \mathbb{R} \text{ is smooth, } t|f'(t)| \to \infty \text{ as } t \to \infty, \text{ but } \lim_{t\to\infty} f'(t) = 0. \tag{5.16}$$

The following is an analogue of Theorem 4.1 tailor-made for such f; the elementary proof is left to the interested reader.

Proposition 5.8. *Let $f, g : \mathbb{R}^+ \to \mathbb{R}$ be smooth, and assume that f satisfies* (5.16)*, as well as $g(|f(t)|) = t$ for all sufficiently large $t \in \mathbb{R}^+$. Then there exists $N_0 \in \mathbb{N}$ such that*

$$d_{\mathbb{T}}(\omega_N^f, \lambda_{\mathbb{T}}) \le \frac{2}{N} \sum\nolimits_{1 \le j \le |f(N)|} \left(|g(j+2) - g(j)| \sup\nolimits_{j \le s,t \le j+2} \frac{|g'(s) - g'(t)| + 1}{|g'(s)| + 1} \right) \qquad \forall N \ge N_0 \,. \tag{5.17}$$

To illustrate Proposition 5.8, consider $f(t) = \sqrt{t}$ for all $t \in \mathbb{R}^+$, hence $g(s) = s^2$, and a short calculation using (5.17) yields

$$\sqrt{N} d_{\mathbb{T}}(\omega_N^f, \lambda_{\mathbb{T}}) \le \frac{40}{\sqrt{N}} \sum\nolimits_{1 \le j \le \sqrt{N}} \frac{j+1}{2j+1} \le 20 + \frac{20}{\sqrt{N}} \sum\nolimits_{1 \le j \le \sqrt{N}} \frac{1}{2j+1} \qquad \forall N \ge N_0 \,,$$

so in particular $\limsup_{N\to\infty} \sqrt{N} d_{\mathbb{T}}(\omega_N^f, \lambda_{\mathbb{T}}) \le 20$. A careful analysis reveals that

$$\lim\nolimits_{N\to\infty} \sqrt{N} \int_{\mathbb{T}} h \, \mathrm{d}(\widetilde{\omega}_N^f - \lambda_{\mathbb{T}}) = \int_{\mathbb{T}} h \, \mathrm{d}(\beta_{2,1} - \lambda_{\mathbb{T}}) \qquad \forall h \in \mathrm{Lip}(\mathbb{T}) \,,$$

and this convergence is uniform on $\mathrm{Lip}_1(\mathbb{T})$, so in particular

$$\lim\nolimits_{N\to\infty} \sqrt{N} d_{\mathbb{T}}(\omega_N^f, \lambda_{\mathbb{T}}) = d_{\mathbb{T}}(\beta_{2,1}, \lambda_{\mathbb{T}}) = \tfrac{1}{16} \,. \tag{5.18}$$

Thus, despite its simplicity, the bound in (5.17) is essentially sharp in this example.

Notice that Proposition 5.8 does not require f' to be (eventually) monotone, and hence may be regarded as a quantitative complement to, rather than a special case of, Fejér's theorem. With regard to Remark 5.7(ii), also notice that (5.18) compares favourably to the fact that $(\sqrt{N} D_N^*(\sqrt{n}))$ is divergent, as

$$\liminf\nolimits_{N\to\infty} \sqrt{N} D_N^*(\sqrt{n}) = 1 < \tfrac{9}{8} = \limsup\nolimits_{N\to\infty} \sqrt{N} D_N^*(\sqrt{n}) \,.$$

This is yet another illustration of how usage of $d_{\mathbb{T}}$ may lead to simpler, more clear-cut results than does usage of other metrics on $\mathcal{P}$, including d_∞.

Acknowledgements

The first author was partially supported by an N Discovery Grant. Both authors are grateful to the anonymous referees for several insightful comments which helped improve the presentation and put it in a wider perspective.

Appendix A. Computing $d_{\mathbb{T}}$ on $\mathcal{P}$

The purpose of this section is to provide simple complete proofs for two basic facts regarding the metric $d_{\mathbb{T}}$ on $\mathcal{P}$, namely (2.2) and (2.3), both of which are useful for concrete computations.

First, with the definition of $d_{\mathbb{T}}$ in mind, notice that $h(0) = 0$ can be assumed in (2.1), and also that $\mathrm{Lip}_{1,0}(\mathbb{T}) := \{h \in \mathrm{Lip}_1(\mathbb{T}) : h(0) = 0\}$ is a compact subset of $C(\mathbb{T})$. Thus the supremum in (2.1) is attained. Moreover, $h \in \mathrm{Lip}_1(\mathbb{T})$ if and only if the 1-periodic extension $t \mapsto h(t + \mathbb{Z})$, also denoted h for simplicity, is absolutely continuous with $\|h'\|_\infty \le 1$ and $\int_{[0,1]} h' \, \mathrm{d}\lambda = 0$, where h' is the (λ-a.e. defined) derivative of h. Thus, $\mathrm{Lip}_{1,0}(\mathbb{T})$ can be identified with the set

$$M_0 := \left\{ g \in L^\infty[0,1] : \|g\|_\infty \le 1, \int_{[0,1]} g \, \mathrm{d}\lambda = 0 \right\} ,$$

with a bijection $\phi : M_0 \to \mathrm{Lip}_{1,0}(\mathbb{T})$ being given by $\phi(g) = \int_{[0,t]} g \,\mathrm{d}\lambda$.

Next, assume that $\mu, \nu \in \mathcal{P}$ are absolutely continuous with densities f_μ, f_ν, respectively. Then, with the appropriate $h \in L_{1,0}(\mathbb{T})$,

$$\begin{aligned} d_{\mathbb{T}}(\mu,\nu) &= \int_{\mathbb{T}} h \,\mathrm{d}(\mu-\nu) = \int_{[0,1]} h(f_\mu - f_\nu)\,\mathrm{d}\lambda \\ &= h(s)(F_\mu(s) - F_\nu(s))\big|_{s=0}^{1} - \int_{[0,1]} h'(F_\mu - F_\nu)\,\mathrm{d}\lambda = \int_{[0,1]} h'(F_\nu - F_\mu)\,\mathrm{d}\lambda\,, \end{aligned}$$

and hence $d_{\mathbb{T}}(\mu,\nu) \le \sup_{g\in M_0} \int_{[0,1]} g(F_\mu - F_\nu)\,\mathrm{d}\lambda$, but also $d_{\mathbb{T}}(\mu,\nu) \ge \int_{\mathbb{T}} \phi(g)\,\mathrm{d}(\mu-\nu)$ for every $g \in M_0$, so that really

$$d_{\mathbb{T}}(\mu,\nu) = \sup_{g\in M_0} \int_{[0,1]} g(F_\mu - F_\nu)\,\mathrm{d}\lambda\,; \tag{A.1}$$

as in (2.1), the supremum is attained in (A.1). Since absolutely continuous probabilities are dense in $\mathcal{P}$, plainly (A.1) is valid for *all* $\mu, \nu \in \mathcal{P}$.

Recalling that $L^1[0,1]^*$, the dual of the Banach space $L^1[0,1]$, is, up to an isometric isomorphism, simply $L^\infty[0,1]$, and with a view on (A.1), notice that M_0 can in turn be identified with the set

$$N_0 := \{\varphi \in L^1[0,1]^* : \|\varphi\| \le 1,\ \varphi(\mathbf{1}_{[0,1]}) = 0\}\,.$$

Thus (A.1) can equivalently be expressed as

$$d_{\mathbb{T}}(\mu,\nu) = \sup_{\varphi\in N_0} \varphi(F_\mu - F_\nu) \qquad \forall \mu,\nu \in \mathcal{P}\,. \tag{A.2}$$

Now the desired identity (2.2) follows immediately from (A.2), together with the following simple observation from linear analysis, applied specifically with $X = L^1[0,1]$, $x = F_\mu - F_\nu$, and $y = \mathbf{1}_{[0,1]}$.

Lemma A.1. *Let $(X, \|\cdot\|)$ be a normed space (over $\mathbb{R}$). Then, for every $x, y \in X$,*

$$\sup\{\varphi(x) : \varphi \in X^*,\ \|\varphi\| \le 1,\ \varphi(y) = 0\} = \min_{a\in\mathbb{R}} \|x + ay\|\,. \tag{A.3}$$

Proof. Fix $x, y \in X$. Since (A.3) clearly is correct otherwise, henceforth assume that x, y are linearly independent. Note that $a \mapsto \|x + ay\|$ is Lipschitz and bounded below by $|a|\,\|y\| - \|x\|$, so attains a global minimum, say for $a = a_0$. For convenience, let $b = \|x+a_0y\| = \min_{a\in\mathbb{R}} \|x+ay\| > 0$ and $B = \{\varphi \in X^* : \|\varphi\| \le 1,\ \varphi(y) = 0\}$. On the one hand, $\sup_{\varphi\in B} |\varphi(x)| \le b$ because

$$|\varphi(x)| = |\varphi(x + a_0y)| \le \|x + a_0y\| = b \qquad \forall \varphi \in B\,.$$

On the other hand, let $z = x + a_0y$ and note that y, z are linearly independent. Thus letting

$$\psi(sy + tz) = t \qquad \forall t \in \mathbb{R}\,,$$

yields a well-defined linear functional on $\mathrm{span}\,\{y, z\}$. Since $\|sy + tz\| \ge |t|b$,

$$|\psi(sy + tz)| = |t| \le \frac{\|sy + tz\|}{b} \qquad \forall s, t \in \mathbb{R}\,,$$

which shows that $\|\psi\| \le 1/b$. Use the Hahn–Banach theorem to extend ψ to $\widetilde{\psi} \in X^*$ with $\|\widetilde{\psi}\| \le 1/b$ and $\widetilde{\psi}(sy + tz) = t$ for all $s, t \in \mathbb{R}$. Then $b\widetilde{\psi} \in B$ and $b\widetilde{\psi}(x) = b\psi(-a_0y + z) = b$. Thus $\sup_{\varphi\in B} |\varphi(x)| \ge b$, and hence $\sup_{\varphi\in B} |\varphi(x)| = b$, as claimed. □

With (2.2) established, it remains to deduce (2.3) from it. To this end, a few preparatory considerations are in order. Throughout, fix any $g \in L^1[0,1]$.

First, given g, let

$$\widehat{g}(t) = \lambda(\{g \le t\}) \qquad \forall t \in \mathbb{R}\,,$$

so $\widehat{g} : \mathbb{R} \to \mathbb{R}$ is non-decreasing, right-continuous, with $\lim_{t\to-\infty} \widehat{g}(t) = 0$ and $\lim_{t\to\infty} \widehat{g}(t) = 1$, as well as $\widehat{g}(t-) = \lambda(\{g < t\})$ for every $t \in \mathbb{R}$. In probabilistic terms, $\widehat{g}$ is the distribution function of $g(Z)$, where the random variable Z is uniform on $[0,1]$. Note that choosing a different representative for $g \in L^1[0,1]$ does not alter $\widehat{g}$. For every $s \in]0,1[$ the set $\{\widehat{g} \le s\}$ is non-empty and bounded above, and hence

$$\widehat{g}^{-1}(s) = \sup\{\widehat{g} \le s\} \qquad \forall s \in]0,1[$$

is well-defined. Again, $\widehat{g}^{-1} :]0,1[\to \mathbb{R}$ is non-decreasing and right-continuous, with

$$\widehat{g}^{-1}(s-) = \inf\{\widehat{g} \ge s\} \qquad \forall s \in]0,1[\,.$$

Now, with g_-, g_+ denoting the negative and positive part of g, respectively, observe that

$$\begin{aligned}
\int_{[0,1]} g\,\mathrm{d}\lambda &= \int_{[0,1]} g_+\,\mathrm{d}\lambda - \int_{[0,1]} g_-\,\mathrm{d}\lambda = \int_0^\infty \lambda(\{g_+ \ge t\})\,\mathrm{d}t - \int_0^\infty \lambda(\{g_- \ge t\})\,\mathrm{d}t \\
&= \int_0^\infty (1 - \widehat{g}(t-))\,\mathrm{d}t - \int_{-\infty}^0 \widehat{g}(t)\,\mathrm{d}t \\
&= \int_0^1 \lambda(\{\widehat{g} \le s\} \cap [0,\infty[)\,\mathrm{d}s - \int_0^1 \lambda(\{\widehat{g} \ge s\} \cap]-\infty,0])\,\mathrm{d}s \\
&= \int_0^1 \max\{\widehat{g}^{-1}(s), 0\}\,\mathrm{d}s - \int_0^1 -\min\{\widehat{g}^{-1}(s), 0\}\,\mathrm{d}s = \int_0^1 \widehat{g}^{-1}(s)\,\mathrm{d}s\,;
\end{aligned}$$

here the second and fourth equal signs are due to Fubini's theorem. The same argument, applied to $|g| = g_+ + g_-$ instead of $g = g_+ - g_-$, shows that also

$$\int_{[0,1]} |g|\,\mathrm{d}\lambda = \int_0^1 |\widehat{g}^{-1}(s)|\,\mathrm{d}s \qquad \forall g \in L^1[0,1]\,.$$

Moreover, notice that $\widehat{g + a} = \widehat{g}(\cdot - a)$ and $\widehat{g + a}^{-1} = \widehat{g}^{-1} + a$ for every $a \in \mathbb{R}$, and hence in fact

$$\int_{[0,1]} |g - t|\,\mathrm{d}\lambda = \int_0^1 |\widehat{g}^{-1}(s) - t|\,\mathrm{d}s \qquad \forall g \in L^1[0,1],\, t \in \mathbb{R}\,.$$

(The function $\widehat{g}^{-1}$ may be thought of as a *symmetrization* or monotone *rearrangement* of g; see, e.g., [18].)

For one other preparatory consideration, let R_g be the essential range of g, that is, let

$$R_g = \{t \in \mathbb{R} : \lambda(\{|g - t| < \varepsilon\}) > 0\ \forall \varepsilon \in \mathbb{R}^+\}\,.$$

Note that R_g coincides with the set of growth points of $\widehat{g}$, i.e.,

$$R_g = \{t \in \mathbb{R} : \widehat{g}(t + \varepsilon) - \widehat{g}(t - \varepsilon) > 0\ \forall \varepsilon \in \mathbb{R}^+\};$$

just like $\widehat{g}$, therefore, R_g is independent of the choice of representative for g. Recall that R_g is non-empty and closed. Moreover, the function $t \mapsto \int_{[0,1]} |g-t|\,\mathrm{d}\lambda$, which is 1-Lipschitz and bounded below by $|t| - \int_{[0,1]} |g|\,\mathrm{d}\lambda$, has constant derivative on each (open) component of $\mathbb{R} \setminus R_g$, and hence attains its global minimum (also) on R_g. Thus,

$$\min_{t\in\mathbb{R}} \int_{[0,1]} |g-t|\,\mathrm{d}\lambda = \min_{t\in R_g} \int_{[0,1]} |g-t|\,\mathrm{d}\lambda = \inf_{t\in[0,1]} \int_{[0,1]} |g-g(t)|\,\mathrm{d}\lambda\,, \tag{A.4}$$

since obviously the closure of $\{g(s) : s \in [0,1]\}$ contains R_g.

With the above preparations, (2.3) is readily established: Given $\mu, \nu \in \mathcal{P}$, it is clear from

$$\int_0^1 |F_\mu(s) - F_\nu(s) - t|\,\mathrm{d}s = \int_0^1 \left|\widehat{(F_\mu - F_\nu)}^{-1}(s) - t\right| \mathrm{d}s \qquad \forall t \in \mathbb{R}\,, \tag{A.5}$$

together with Proposition 3.3, that either side in (A.5) is minimal precisely if

$$t \in \left[\widehat{(F_\mu - F_\nu)}^{-1}\left(\tfrac{1}{2}-\right), \widehat{(F_\mu - F_\nu)}^{-1}\left(\tfrac{1}{2}\right)\right],$$

or equivalently if $\widehat{(F_\mu - F_\nu)}(t-) \le \frac{1}{2}$ and $\widehat{(F_\mu - F_\nu)}(t) \ge \frac{1}{2}$, which in turn is the case if and only if

$$1 - \lambda(\{F_\mu - F_\nu \ge t\}) \le \tfrac{1}{2} \quad \text{and} \quad \lambda(\{F_\mu - F_\nu \le t\}) \ge \tfrac{1}{2}\,.$$

This proves (2.3). Also, notice that (A.4) yields

$$d_{\mathbb{T}}(\mu,\nu) = \inf_{t\in[0,1]} \int_0^1 |F_\mu(s) - F_\nu(s) - F_\mu(t) + F_\nu(t)|\,\mathrm{d}s \qquad \forall \mu,\nu \in \mathcal{P}\,.$$

Remark A. 2. This section simplifies and extends the main results of [5]; in particular, (A.4) explains why the (restrictive) continuity assumption in [5, Cor. 3.9] is superfluous.

Appendix B. A calculus lemma

This section provides a self-contained proof of Lemma 3.4. To prepare for it, given any Riemann integrable function $f : [0,1] \to \mathbb{R}$ for which $f(0+)$ and $f(1-)$ exist and any $a \in [0,1]$, consider the quantity

$$I^f(a) := (f(1-) - f(0+))(1-a) - 2\int_0^{\frac{1}{2}a} f(s)\,\mathrm{d}s + 2\int_{\frac{1}{2}a}^{\frac{1}{2}(a+1)} f(s)\,\mathrm{d}s - 2\int_{\frac{1}{2}(a+1)}^1 f(s)\,\mathrm{d}s\,. \tag{B.1}$$

With $f^\dagger : [0,1] \to \mathbb{R}$ given by $f^\dagger(s) = -f(1-s)$ for all $s \in [0,1]$, notice that

$$I^{f^\dagger}(a) + I^f(1-a) = f(1-) - f(0+) \qquad \forall a \in [0,1]\,. \tag{B.2}$$

If, for example, $f(s) = \pi \sin(2\pi s)$ then $f^\dagger = f$, and $I^f(a) = 4\cos(\pi a)$, so I^f may attain positive as well as negative values. As presented here, the proof of Lemma 3.4 hinges on the elementary observation that such indefiniteness of I^f cannot occur in case f is monotone.

Lemma B.1. *Let $f : [0,1] \to \mathbb{R}$ be monotone, and $a \in [0,1]$. If f is non-decreasing then*

$$0 \le I^f(a) \le f(1-) - f(0+)\,; \tag{B.3}$$

if f is non-increasing then both inequalities in (B.3) *are reversed. Moreover, equality in* (B.3) *holds on the left precisely if f is constant on $\left]0, \frac{1}{2}(a+1)\right[$ and $\left]\frac{1}{2}(a+1), 1\right[$, whereas equality holds on the right precisely if f is constant on $\left]0, \frac{1}{2}a\right[$ and $\left]\frac{1}{2}a, 1\right[$.*

Proof. Fix $a \in [0,1]$. Assume that f is non-decreasing (otherwise replace f by $-f$), and also that $f(0+) < f(1-)$, as all assertions clearly are correct in case $f(0+) = f(1-)$. To see that $I^f(a) \ge 0$ for all $a \in]0,1[$, simply observe that

$$\begin{aligned}
I^f(a) &= (f(1-) - f(0+))(1-a) + 2\int_0^{\frac{1}{2}a} \left(f\left(\tfrac{1}{2}a-\right) - f(s)\right) \mathrm{d}s - f\left(\tfrac{1}{2}a-\right) a + \\
&\qquad + 2\int_{\frac{1}{2}a}^{\frac{1}{2}(a+1)} \left(f(s) - f\left(\tfrac{1}{2}a+\right)\right) \mathrm{d}s + f\left(\tfrac{1}{2}a+\right) + 2\int_{\frac{1}{2}(a+1)}^{1} (f(1-) - f(s))\,\mathrm{d}s - f(1-)(1-a) \\
&= \left(f\left(\tfrac{1}{2}a-\right) - f(0+)\right)(1-a) + f\left(\tfrac{1}{2}a+\right) - f\left(\tfrac{1}{2}a-\right) + 2\int_0^{\frac{1}{2}a} \left(f\left(\tfrac{1}{2}a-\right) - f(s)\right) \mathrm{d}s + \\
&\qquad + 2\int_{\frac{1}{2}a}^{\frac{1}{2}(a+1)} \left(f(s) - f\left(\tfrac{1}{2}a+\right)\right) \mathrm{d}s + 2\int_{\frac{1}{2}(a+1)}^{1} (f(1-) - f(s))\,\mathrm{d}s \\
&\ge 0\,.
\end{aligned}$$

Moreover, this inequality is strict unless f is constant on each of the three non-empty intervals $\left]0, \frac{1}{2}a\right[$, $\left]\frac{1}{2}a, \frac{1}{2}(a+1)\right[$, $\left]\frac{1}{2}(a+1), 1\right[$, and does not have a jump at $s = \frac{1}{2}a$. In other words, $I^f(a) \ge 0$ always, and $I^f(a) > 0$ unless f is constant on $\left]0, \frac{1}{2}(a+1)\right[$ and $\left]\frac{1}{2}(a+1), 1\right[$. Conversely, if $f(s) = b$ for all $s \in \left]0, \frac{1}{2}(a+1)\right[$, and $f(s) = c$ for all $s \in \left]\frac{1}{2}(a+1), 1\right[$ then

$$I^f(a) = (c-b)(1-a) - 2b \cdot \tfrac{1}{2}a + 2b \cdot \tfrac{1}{2} - 2c \cdot \tfrac{1}{2}(1-a) = 0\,.$$

For $a = 0$, observe that

$$\begin{aligned}
I^f(0) &= f(1-) - f(0+) + 2\int_0^{\frac{1}{2}} f(s)\,\mathrm{d}s - 2\int_{\frac{1}{2}}^{1} f(s)\,\mathrm{d}s \\
&= 2\int_0^{\frac{1}{2}} (f(s) - f(0+))\,\mathrm{d}s + 2\int_{\frac{1}{2}}^{1} (f(1-) - f(s))\,\mathrm{d}s \ge 0\,,
\end{aligned}$$

and this inequality is strict unless f is constant on $\left]0, \frac{1}{2}\right[$ and $\left]\frac{1}{2}, 1\right[$. Similarly, for $a = 1$,

$$I^f(1) = -2\int_0^{\frac{1}{2}} f(s)\,\mathrm{d}s + 2\int_{\frac{1}{2}}^{1} f(s)\,\mathrm{d}s = 2\int_0^{\frac{1}{2}} \left(f\left(s+\tfrac{1}{2}\right) - f(s)\right) \mathrm{d}s \ge 0\,,$$

and this inequality is strict unless f is constant on $]0,1[$. This proves all assertions regarding the left part of (B.3). To prove the assertions regarding the right part, simply notice that $f^\dagger$ is non-decreasing as well, so by (B.2) and by what has already been shown,

$$I^f(a) = f(1-) - f(0+) - I^{f^\dagger}(1-a) \le f(1-) - f(0+)\,,$$

and this inequality is strict unless $I^{f^\dagger}(1-a)=0$, that is, unless $f^\dagger$ is constant on $\left]0,\frac{1}{2}(2-a)\right[$ and $\left]\frac{1}{2}(2-a),1\right[$, or equivalently, f is constant on $\left]0,\frac{1}{2}a\right[$ and $\left]\frac{1}{2}a,1\right[$. □

Proof of Lemma 3.4. Though in essence an elementary calculus exercise, the argument is somewhat intricate and requires a certain amount of notational and analytic preparations that are going to be carried out first. Again, assume throughout that f is non-decreasing. (Otherwise replace f by $-f$.)

Given any sequence (c_n) in $\mathbb{R}$, for every $N\in\mathbb{N}$ define the following $N+4$ numbers:

$$s_{N,0}:=0\,,\qquad s_{N,n}:=\frac{\langle c_N\rangle+n-1}{N}\quad(n=1,\dots,N)\,,\qquad s_{N,N+1}:=1\,;$$

$$\epsilon_N:=\langle Nc_{N+1}-(N+1)c_N\rangle\,,\qquad \ell_N:=\lfloor N\langle c_{N+1}\rangle-(N+1)\langle c_N\rangle\rfloor+1\,.$$

Notice that $0=s_{N,0}\le s_{N,1}<s_{N,2}<\ldots<s_{N,N}<s_{N,N+1}=1$ and $0\le\epsilon_N<1$, but also $|\ell_N|\le N$ for every N. Furthermore, given any (non-decreasing) function $f:[0,1]\to\mathbb{R}$ let

$$F_{N,n}=f\left(\tfrac{1}{2}(s_{N,n}+s_{N,n+1})\right)\qquad\forall n=0,1,\dots,N\,,$$

as well as

$$f_N(s)=\sum\nolimits_{n=0}^{N}F_{N,n}\mathbf{1}_{[s_{N,n},s_{N,n+1}[}(s)\qquad\forall s\in[0,1[\,;$$

with $f_N(1):=f_N(1-)$ the function $f_N:[0,1]\to\mathbb{R}$ is piecewise constant (and non-decreasing). Notice that $f(0+)\le F_{N,1}\le\ldots\le F_{N,N}\le f(1-)$. In the notation of (3.4) simply $f_N=\mathcal{F}^f_{c_N,N}$.

Now, the main thrust of the subsequent argument is to find a (positive) lower bound for the deviation

$$J_N:=\int_0^1|f_{2N}(s)-f_{2N+1}(s)|\,\mathrm{d}s\qquad(N\in\mathbb{N})$$

along a suitable increasing sequence (N_j) in $\mathbb{N}$. In order to ensure good readability for what follows, write $s_{2N,n}$, $s_{2N+1,n}$, $F_{2N,n}$, $F_{2N+1,n}$ simply as a_n, b_n, F_n, G_n, respectively, so that, for instance,

$$f_{2N}(s)=\sum\nolimits_{n=0}^{2N}F_n\mathbf{1}_{[a_n,a_{n+1}[}(s)\,,\quad f_{2N+1}(s)=\sum\nolimits_{n=0}^{2N+1}G_n\mathbf{1}_{[b_n,b_{n+1}[}(s)\qquad\forall s\in[0,1[\,.$$

Since f_{2N}, f_{2N+1} are piecewise constant and non-decreasing, in order to determine J_N precisely, it is necessary to first clarify the position of the $2N+1$ intervals $[a_0,a_1[,\dots,[a_{2N},a_{2N+1}[$ relative to the $2N+2$ intervals $[b_0,b_1[,\dots,[b_{2N+1},b_{2N+2}[$. To this end, observe that $a_n\le b_n$ whenever $0\le n\le\max\{0,\ell_{2N}\}$, whereas $a_n>b_n$ whenever $1+\max\{0,\ell_{2N}\}\le n\le 2N+1$, and

$$a_n-b_n=\frac{n-\epsilon_{2N}-\ell_{2N}}{2N(2N+1)}\qquad\forall n=1,\dots,2N\,.$$

Similarly, $a_n\le b_{n+1}$ whenever $0\le n\le 2N+\min\{0,\ell_{2N}\}$, whereas $a_n>b_{n+1}$ whenever $2N+1+\min\{0,\ell_{2N}\}\le n\le 2N+1$, and

$$a_n-b_{n+1}=\frac{n-2N-\epsilon_{2N}-\ell_{2N}}{2N(2N+1)}\qquad\forall n=1,\dots,2N\,.$$

Finally, $a_{n+1}>b_n$ and $a_{n-1}<b_{n+1}$ for all $n=1,\dots,2N$.

Next, recall that $|\ell_{2N}| \le 2N$. It is going to be convenient to distinguish the four (slightly overlapping) cases

$$-2N \le \ell_{2N} \le -N\,, \quad -N \le \ell_{2N} \le 0\,, \quad 0 \le \ell_{2N} \le N\,, \quad N \le \ell_{2N} \le 2N\,. \tag{B.4}$$

Assume for the time being that $N \le \ell_{2N} \le 2N$. (As detailed below, the argument for the other three cases is completely analogous.) In this case, if $0 \le n < \ell_{2N}$ then

$$|f_{2N}(s) - f_{2N+1}(s)| = \begin{cases} F_n - G_{n-1} & s \in [a_n, b_n[\,, \\ G_n - F_n & s \in [b_n, a_{n+1}[\,, \end{cases}$$

whereas if $n = \ell_{2N}$ then

$$|f_{2N}(s) - f_{2N+1}(s)| = \begin{cases} F_n - G_{n-1} & s \in [a_n, b_n[\,, \\ |F_n - G_n| & s \in [b_n, b_{n+1}[\,, \\ G_{n+1} - F_n & s \in [b_{n+1}, a_{n+1}[\,, \end{cases}$$

and if $\ell_{2N} < n \le 2N$ then

$$|f_{2N}(s) - f_{2N+1}(s)| = \begin{cases} F_n - G_n & s \in [a_n, b_{n+1}[\,, \\ G_{n+1} - F_n & s \in [b_{n+1}, a_{n+1}[\,. \end{cases}$$

With this, and temporarily writing ϵ_{2N}, ℓ_{2N} as ϵ, ℓ, respectively, for convenience, it follows that

$$\begin{aligned}
J_N &= \sum\nolimits_{n=0}^{2N} \int_{a_n}^{a_{n+1}} |f_{2N}(s) - f_{2N+1}(s)|\,\mathrm{d}s \\
&= \sum\nolimits_{n=0}^{\ell-1}((F_n - G_{n-1})(b_n - a_n) + (G_n - F_n)(a_{n+1} - b_n)) + \\
&\quad + (F_\ell - G_{\ell-1})(b_\ell - a_\ell) + |F_\ell - G_\ell|(b_{\ell+1} - b_\ell) + (G_{\ell+1} - F_\ell)(a_{\ell+1} - b_{\ell+1}) + \\
&\quad + \sum\nolimits_{n=\ell+1}^{2N}((F_n - G_n)(b_{n+1} - a_n) + (G_{n+1} - F_n)(a_{n+1} - b_{n+1})) \\
&\ge \sum\nolimits_{n=2}^{\ell}((F_n - G_{n-1})(b_n - a_n) + (G_{n-1} - F_{n-1})(a_n - b_{n-1})) + \\
&\quad + \sum\nolimits_{n=\ell+1}^{2N}((F_n - G_n)(b_{n+1} - a_n) + (G_n - F_{n-1})(a_n - b_n)) \\
&= \frac{1}{2N(2N+1)} \sum\nolimits_{n=2}^{\ell}((F_n - G_{n-1})(\epsilon + \ell - n) + (G_{n-1} - F_{n-1})(2N + n - \epsilon - \ell)) + \\
&\quad + \frac{1}{2N(2N+1)} \sum\nolimits_{n=\ell+1}^{2N}((F_n - G_n)(2N - n + \epsilon + \ell) + (G_n - F_{n-1})(n - \epsilon - \ell))\,.
\end{aligned}$$

All terms in the above sums are non-negative, so

$$\begin{aligned}
2N(2N+1)J_N &\geq \sum_{n=2}^{\ell}(F_n - F_{n-1})\min\{\ell - n, 2N + n - \ell\} + \\
&\quad + \sum_{n=\ell+1}^{2N}(F_n - F_{n-1})\min\{2N - n + \ell, n - \ell\} + \\
&\quad + \epsilon\left(\sum_{n=2}^{\ell}(F_n + F_{n-1} - 2G_{n-1}) + \sum_{n=\ell+1}^{2N}(F_n + F_{n-1} - 2G_n)\right) \\
&= \sum_{n=2}^{\ell-N}(F_n - F_{n-1})(2N + n - \ell) + \\
&\quad + \sum_{n=\ell-N+1}^{\ell}(F_n - F_{n-1})(\ell - n) + \sum_{n=\ell+1}^{2N}(F_n - F_{n-1})(n - \ell) + \\
&\quad + \epsilon\left(2\sum_{n=1}^{2N}(F_n - G_n) - F_1 - F_{2N} + 2G_\ell\right) \\
&= (F_{2N} - F_1)(2N + 1 - \ell) - \sum_{n=2}^{\ell-N-1} F_n + \sum_{n=\ell-N}^{\ell-1} F_n - \sum_{n=\ell}^{2N} F_n + \\
&\quad + \epsilon\left(2\sum_{n=1}^{2N} F_n - 2\sum_{n=1}^{2N} G_n - F_1 - F_{2N} + 2G_\ell\right).
\end{aligned}$$

Since $2G_\ell - F_1 - F_{2N} \geq f(0+) - f(1-)$,

$$\begin{aligned}
(2N+1)J_N &\geq (F_{2N} - F_1)\frac{2N + 1 - \ell}{2N} - \frac{1}{2N}\sum_{n=1}^{\ell-N-1} F_n + \frac{1}{2N}\sum_{n=\ell-N}^{\ell-1} F_n - \frac{1}{2N}\sum_{n=\ell}^{2N} F_n + \\
&\quad + \epsilon\left(\frac{1}{N}\sum_{n=1}^{2N} F_n - \frac{1}{N}\sum_{n=1}^{2N} G_n - \frac{f(1-) - f(0+)}{2N}\right),
\end{aligned}$$

provided that $N \leq \ell \leq 2N$. Notice that

$$F_1 = f\left(\tfrac{1}{2}(s_{2N,1} + s_{2N,2})\right) \overset{N\to\infty}{\longrightarrow} f(0+), \quad F_{2N} = f\left(\tfrac{1}{2}(s_{2N,2N} + 1)\right) \overset{N\to\infty}{\longrightarrow} f(1-),$$

but also, since f is Riemann integrable,

$$\frac{1}{2N}\sum_{n=1}^{2N} F_n = \sum_{n=0}^{2N} F_n(a_{n+1} - a_n) + \frac{\langle c_{2N}\rangle}{2N}(F_{2N} - F_0) \overset{N\to\infty}{\longrightarrow} \int_0^1 f(s)\,\mathrm{d}s,$$

and similarly

$$\frac{1}{2N}\sum_{n=1}^{2N} G_n \overset{N\to\infty}{\longrightarrow} \int_0^1 f(s)\,\mathrm{d}s.$$

Suppose now that $N \leq \ell_{2N} \leq 2N$ for every $N \in \mathbb{N}$, and that furthermore $\lim_{N\to\infty} \ell_{2N}/N = c$ for some $c \in [1, 2]$. Then

$$a_{\ell-N} = \frac{\langle c_{2N}\rangle + \ell - N - 1}{2N} \overset{N\to\infty}{\longrightarrow} \tfrac{1}{2}(c-1), \quad a_\ell = \frac{\langle c_{2N}\rangle + \ell - 1}{2N} \overset{N\to\infty}{\longrightarrow} \tfrac{1}{2}c,$$

from which it is readily seen that

$$\frac{1}{2N}\sum_{n=1}^{\ell-N-1} F_n \overset{N\to\infty}{\longrightarrow} \int_0^{\frac{1}{2}(c-1)} f(s)\,\mathrm{d}s,$$

but also

$$\frac{1}{2N}\sum_{n=\ell-N}^{\ell-1} F_n \overset{N\to\infty}{\longrightarrow} \int_{\frac{1}{2}(c-1)}^{\frac{1}{2}c} f(s)\,\mathrm{d}s, \quad \frac{1}{2N}\sum_{n=\ell}^{2N} F_n \overset{N\to\infty}{\longrightarrow} \int_{\frac{1}{2}c}^1 f(s)\,\mathrm{d}s.$$

Since $0 \le \epsilon < 1$, clearly $\epsilon/N \overset{N\to\infty}{\longrightarrow} 0$, and hence

$$\begin{aligned} 4\liminf_{N\to\infty} NJ_N &= \liminf_{N\to\infty} 2(2N+1)J_N \\ &\ge (f(1-)-f(0+))(2-c) - 2\int_0^{\frac{1}{2}(c-1)} f(s)\,\mathrm{d}s + 2\int_{\frac{1}{2}(c-1)}^{\frac{1}{2}c} f(s)\,\mathrm{d}s - 2\int_{\frac{1}{2}c}^{1} f(s)\,\mathrm{d}s \\ &= I^f(c-1)\,, \end{aligned} \tag{B.5}$$

with I^f given by (B.1). Arguments completely analogous to the one leading to (B.5) show that if either of the other three cases in (B.4) holds for all $N \in \mathbb{N}$, and if furthermore $\lim_{N\to\infty} \ell_{2N}/N = c$ then (B.5) remains valid, provided that its right-hand side is replaced by

$$I^{f^\dagger}(-c-1)\,, \quad I^f(c+1)\,, \quad I^{f^\dagger}(1-c)\,, \tag{B.6}$$

respectively. Notice that by Lemma 1, the number $I^f(c-1)$ is positive unless f is piecewise constant on $]0,1[$ with at most one jump, and the same is true for each of the numbers in (B.6).

At long last, the scene is now set for showing that (i) and (ii) in Lemma 3.4 indeed are equivalent. If f is constant on $]0,1[$ then obviously (i) and (ii) both hold, so assume w.l.o.g. that $f(0+) < f(1-)$. First assume (i). At least one of the four cases in (B.4) occurs for infinitely many N, and by choosing an appropriate increasing sequence (N_j) in $\mathbb{N}$, it can be assumed that case occurs for every N_j, and furthermore $\lim_{j\to\infty} \ell_{2N_j}/N_j$ exists. Since

$$\begin{aligned} NJ_N &\le N\int_0^1 |f(s) - f_{2N}(s)|\,\mathrm{d}s + N\int_0^1 |f(s) - f_{2N+1}(s)|\,\mathrm{d}s \\ &\le \tfrac{1}{2}(2N+1)\int_0^1 |f(s) - \mathcal{F}^f_{c_{2N+1},2N+1}(s)|\,\mathrm{d}s + \tfrac{1}{2}\cdot 2N\int_0^1 |f(s) - \mathcal{F}^f_{c_{2N},2N}(s)|\,\mathrm{d}s \qquad \forall N\in\mathbb{N}\,, \end{aligned}$$

clearly $\lim_{j\to\infty} N_j J_{N_j} = 0$. As seen above, there exists $a \in [0,1[$ such that f is constant on $]0,a[$ and $]a,1[$; in fact, $a > 0$ as f is not constant on $]0,1[$. For every sufficiently large N there exists a unique $1 \le n \le N$ with $a \in [s_{N,n}, s_{N,n+1}[$, and therefore

$$\begin{aligned} N\int_0^1 |f(s) - \mathcal{F}^f_{c_N,N}(s)|\,\mathrm{d}s &= N\int_{s_{N,n}}^{s_{N,n+1}} |f(s) - F_{N,n}|\,\mathrm{d}s \\ &= N(f(1-)-f(0+))\min\{a - s_{N,n}, s_{N,n+1} - a\} \\ &= (f(1-)-f(0+))d_{\mathbb{T}}(Na, c_N)\,. \end{aligned} \tag{B.7}$$

Thus (i) implies $\lim_{N\to\infty} d_{\mathbb{T}}(Na, c_N) = 0$ and hence (ii). Conversely, (B.7) immediately shows that (ii)⇒(i). □

Remark B. 2. If, instead of being monotone, the function $f : [0,1] \to \mathbb{R}$ is assumed to be C^1 then $\lim_{N\to\infty} N\int_0^1 |f(s) - \mathcal{F}^f_{c_N,N}(s)|\,\mathrm{d}s = \frac{1}{4}\int_0^1 |f'(s)|\,\mathrm{d}s$ for every sequence (c_n) in $\mathbb{R}$. Thus, the equivalence (i)⇔(ii) of Lemma 3.4 remains valid in this situation also. Notice, however, that whereas (ii)⇒(i) always, the reverse implication may fail if no assumptions at all are imposed on the (Riemann integrable) function f.

References

[1] A. B , A note on the distributions of $(\log n)$ mod 1, *Unif. Distrib. Theory* **17**(2022), 77–100.

[2] A. B C. X , Asymptotics of one-dimensional Lévy approximations, *J. Theoret. Probab.* **33**(2020), 1164–1195.

[3] S. B M. L , One-dimensional empirical measures, order statistics, and Kantorovich transport distances, *Mem. Amer. Math. Soc.* **261**(2019), 1259.

[4] L. B S. S , On the Wasserstein distance between classical sequences and the Lebesgue measure, *Trans. Amer. Math. Soc.* **373**(2020), 8943–8962.

[5] C.A. C U.M. M , The Kantorovich metric for probability measures on the circle, *J. Comput. Appl. Math.* **57**(1995), 345–361.

[6] E. C ` , Sur une formule empirique de M. Pervouchine, *Comptes rendus hebdo. des séances de l'académie des sciences* **CXIX**(1895), 848–849.

[7] J. C , Asymptotische Verteilung reeller Zahlen mod 1, *Monatsh. Math.* **64**(1960), 201–225.

[8] F. D'A , F. L , P. M , Deformations of the Canonical Commutation Relations and Metric Structures? *SIGMA* **10**(2014), 14 pp.

[9] J. D , J. R , Y. G , Transportation Distances on the Circle, *J. Math. Imaging Vision* **41**(2011), 147–167.

[10] R.M. D , *Real analysis and probability*, Cambridge Studies in Advanced Mathematics **74**, Cambridge University Press, 2002.

[11] P. D , The kth prime is greater than $k(\ln k + \ln\ln k - 1)$ for $k \geq 2$, *Math. Computation* **68**(1999), 411–415.

[12] P. D , Estimates of some functions over primes without R.H., `arXiv:1002.0442v1`, 2010.

[13] K. F , B. G , S. K , J. M , T. T , Long gaps between primes, *J. Amer. Math. Soc.* **31**(2018), 65–105.

[14] C. G , Irregularity of distribution in Wasserstein distance, *J. Fourier Anal. Appl.* **26**(2020), 21pp.

[15] H.L. H , On the distribution of $\log(p)$, *Int. J. Pure Appl. Math.* **7**(2003), 499–508.

[16] M. I , Asymptotic quantization for probability measures on Riemannian manifolds, *ESAIM Control Optim. Calc. Var.* **22**(2016), 770–785.

[17] J.H.B. K , Distributions modulo 1 of slowly changing sequences, *Nieuw Arch. Wisk.* **21**(1973), 138–163.

[18] S. K , *Symmetrization and applications*, World Scientific Publishing, 2006.

[19] L. K H. N , *Uniform distribution of sequences*, Wiley-Interscience, 1974.

[20] H. N , Distribution mod 1 of monotone sequences, *Nederl. Akad. Wetensch. Indag. Math.* **46**(1984), 315–327.

[21] Y. O , On sequences involving primes, *Unif. Distrib. Theory* **6**(2011), 221–238.

[22] Y. O O. S Distribution of leading digits of numbers, *Unif. Distrib. Theory* **11**(2016), 23–45.

[23] Y. O O. S , Distribution of leading digits of numbers II, *Unif. Distrib. Theory* **14**(2019), 19–42.

[24] K.R. P , *Probability measures on metric spaces*, Academic Press, 1967.

[25] S. S , Wasserstein distance, Fourier series and applications, *Monatsh. Math.* **194**(2021), 305–338.

[26] O. S , *Distribution of sequences: a theory*, Veda a Academia Nakladatelství, 2019.

[27] O. S O. B ˇ ´, Distribution of the sequence p_n/n mod 1, *Unif. Distrib. Theory* **1**(2006), 45–63.

[28] O. S Š. P ´, *Distribution of sequences: a sampler*, Schriftenreihe der Slowakischen Akademie der Wissenschaften **1**, Peter Lang, 2005.

[29] A. W , On the cyclical distribution of the logarithms of the prime numbers, *Quart. J. Math. Oxford Ser.* **6**(1935), 65–68.

[30] C. X , The distributional asymptotics mod 1 of $(\log_b n)$, *Unif. Distrib. Theory* **14**(2019), 105–122.

[31] C. X A. B , Best finite constrained approximations of one-dimensional probabilities, *J. Approx. Theory* **244**(2019), 1–36.

[32] Y. Z , Bounded gaps between primes, *Ann. Math.* **179**(2014), 1121–1174.