Probability Distributions on a Circle

by

Ardalan Rahmatidehkordi

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Applied Mathematics

Department of Mathematical and Statistical Sciences
University of Alberta

Abstract

Distributions of sequences modulo one (mod 1) have been studied over the past century with applications in algebra, number theory, statistics, and computer science. For a given sequence, the weak convergence of the associated empirical distributions has been the usual approach to these studies. In this thesis, we give a formula for calculating the Kantrovich distance between mod 1 probability measures. We then use this distance to study the convergence behavior of the (mod 1) empirical distributions associated with real sequences $(x_n)_{n=1}^{\infty}$ for which $\lim_{n\to\infty} n(x_n - x_{n-1})$ exists. We find that for such sequences, every probability distribution in the limit set of the empirical distributions is a rotated version of a certain exponential distribution. We also describe the speed of convergence to this limit set of distributions.

Keywords: distribution modulo one, slow-growing sequence, Kantorovich-Wasserstein distance, empirical measure, weak convergence

Acknowledgements

First and foremost, I am extremely grateful to Dr Arno Berger for his kind and patient help throughout the past years. I consider myself stupendously lucky to have been his student. I'd like to thank Dr Feng Dai for always taking the time to answer my questions. I'd also like to thank University of Alberta, and its Department of Mathematical and Statistical Sciences as well as the chair of the department.

Last but not least, I thank my dad whose sudden death did not let him keep his promise of celebrating my graduation with me. I thank my mom and sister who remained strong nevertheless.

Table of Contents

List of Figures

List of Symbols and Abbreviations

$\mathbb{A}^{(x_n)_{n=1}^{\infty}}$	The set of the accumulation points of the sequence $(x_n)_{n=1}^{\infty}$.
$\overline{A}$	The closure of the subspace A of a topological space X.
$\forall$	'for every'.
$\wedge$	'and'.
aka	'also known as'.
$\rightarrow$	'approaches'.
$\ll$	'is absolutely continuous with respect to' (used for measures).
$\mathcal{B}_X$	Borel σ-algebra on the topological space X.
$\mathbb{C}$	The set of all complex numbers.
$C(X;Y)$	$:= \left\{f \in Y^X : f \text{ is continuous}\right\}$ where X and Y are topological spaces.
$C(\mathbb{T})$	$:= C(\mathbb{T};\mathbb{R})$.
CDF	'cumulative distribution function'.
※	The symbol to indicate a contradiction.
#	Cardinality (of a set), counting measure.
$\searrow$	'is decreasing', 'decreases to'.

$\sim$	'is distributed as', 'is asymptotically equivalent to'.
$\mathbb{E}$	Expectation, the mean value (of a random variable).
$\exists$	'there exists'.
$f(x_0^-)$	$:= \lim_{\epsilon \searrow 0} f(x_0 - \epsilon) \qquad \forall x_0 \in \mathrm{Dom}(f)$.
$f(x_0^+)$	$:= \lim_{\epsilon \searrow 0} f(x_0 + \epsilon) \qquad \forall x_0 \in \mathrm{Dom}(f)$.
i.i.d.	'independent and identically distributed'.
$\overset{P}{\Longrightarrow}$	'implies through statement P', 'therefore, through statement P'.
$\nearrow$	'is increasing', 'increases to'.
inf	Infimum (of a set).
$\cap$	Intersection (of sets).
$\mathbb{1}_X$	The indicator function on X where X is a set.
$L^1(\mathbb{T}, \mathcal{B}_{\mathbb{T}}, \mu)$	The set of all μ-integrable functions $\qquad \forall \mu \in \mathcal{P}$.
$\mathrm{Lip}_a(X;Y)$	$:= \Big\{ h \in C(X;Y) : d_Y\big(h(x_1), h(x_2)\big) \leq a\, d_X(x_1, x_2) \quad \forall x_1, x_2 \in X \Big\}$.
$\mathrm{Lip}_1(\mathbb{T})$	$:= \mathrm{Lip}_1(\mathbb{T};\mathbb{R})$.
$\mathrm{Lip}_{1,0}(\mathbb{T})$	$:= \big\{ h \in \mathrm{Lip}_1(\mathbb{T};\mathbb{R}) : h(0) = 0 \big\}$.
λ	The Lebesgue measure.
$\mathbb{N}$	The set of natural numbers, i.e., $\{1, 2, 3, \cdots\}$.
log	Natural logarithm.
$\Omega^{(x_n)_{n=1}^{\infty}}$	The set of all the accumulation points of the set of the empirical probability distributions of the sequence $(x_n)_{n=1}^{\infty}$.
$\vee$	'or', i.e., either one or the other (or both).

$\mathcal{P}$	The set of all probability measures on the measurable space $(\mathbb{T}, \mathcal{B}_{\mathbb{T}})$.
$\mathcal{P}_0$	$:= \{\mu \in \mathcal{P} : \#\,\mathrm{supp}(\mu) < \infty\}$.
$\mathcal{P}_{AC}$	$:= \{\mu \in \mathcal{P} : \mu \ll \lambda_{\mathbb{T}}\}$.
$\mathcal{P}_C$	$:= \{\mu \in \mathcal{P} : \mu$ continuous $\}$.
$\square$	The QED symbol for a subordinate proof, i.e., the proof of a claim within a larger proof.
$\blacksquare$	The QED symbol for a main proof, i.e., the proof of a stand-alone theorem or lemma. A main proof is not contained within a larger proof.
$\mathbb{R}$	The set of all real numbers.
$\overline{\mathbb{R}}$	The set of extended real numbers.
$\mathbb{R}^+$	$:= \{x \in \mathbb{R} : x \geq 0\}$.
X^*	The dual space of X where X is a normed linear space.
:	'such that'.
s.t.	'such that'.
sup	Supremum (of a set).
$\mathbb{T}$	$:= \mathbb{R}/\mathbb{Z}$.
$\cup$	Union (of sets).
$\uplus$	Union (of sets) where we know the intersection is empty.
w.r.t.	'with respect to'.
Y^X	The set of all Y-valued functions on the domain X, i.e., all $f\colon X \to Y$.
$\mathbb{Z}$	The set of all integers.

Motivation

For a real number x, its fractional part $\langle x \rangle$ is the number in $[0,1)$ such that $x - \langle x \rangle$ is an integer. For example, $\langle 3.14 \rangle$ is 0.14. Consideration of the fractional part of a real number is, in algebraic terms, considering its equivalence class in the quotient space $\mathbb{T} := \mathbb{R}/\mathbb{Z}$. As a natural consequence of taking numbers modulo 1 (mod 1 for short), one is compelled to think of the fractional parts of reals not as the usual interval $[0,1)$, but as the interval bent to form a circle so that the two endpoints overlap. This topology is an obvious choice in light of a sequence like $(0.9, 0.99, 0.999, \cdots)$, and the consideration of the fractional part of its limit.

The question is 'how are the fractional parts of a sequence $(x_n)_{n=1}^{\infty}$ distributed on the circle?' Take, for example, the sequence $x = (\sqrt{2}\, n)_{n=1}^{\infty}$. Plotting the first N elements of this sequence on the circle for larger and larger N seems to suggest that the points are uniformly distributed mod 1 (see Figure 0.1). This is indeed known to be the case [19]. In other words, the sequence $(\omega_N)_{N=1}^{\infty}$ of empirical distributions associated with x given by

$$\omega_N = \frac{1}{N} \sum_{n=1}^{N} \delta_{x_n + \mathbb{Z}} \quad \forall N \in \mathbb{N} \quad ,$$

converges for every interval $[a,b] \subseteq [0,1)$ to the length of that interval, i.e., $\lim_{N\to\infty} \omega_N\big([a,b]\big) = b - a$. This is equivalent to the notion of weak convergence of measures (see (1.4)).

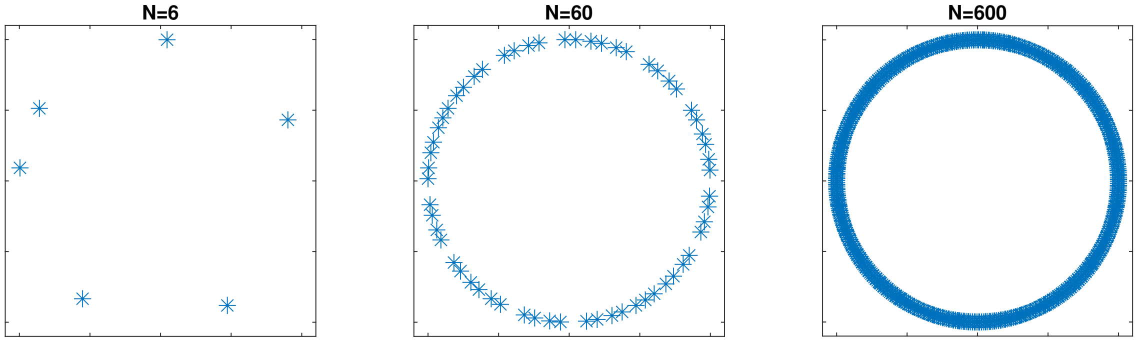

Figure 0.1: Plotting the first N points of the sequence $(\sqrt{2}\,n)_{n=1}^{\infty}$ mod 1. The pattern seems to suggest that in the limit, the points are uniformly distributed. This is known to be the case.

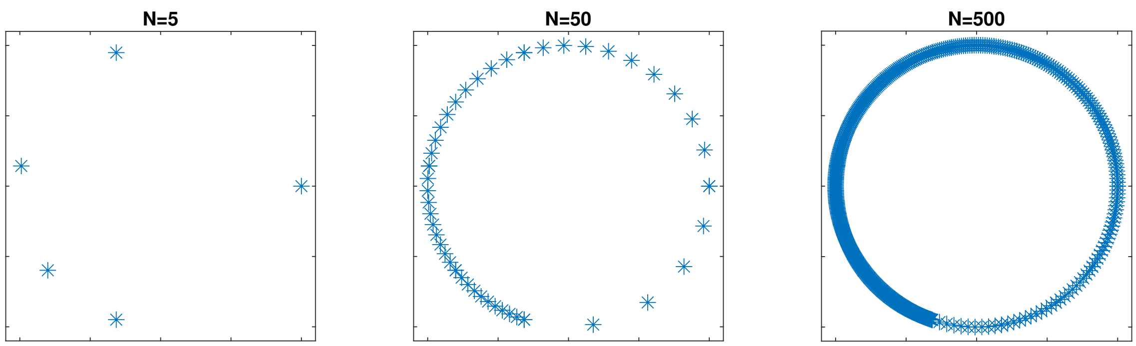

Figure 0.2: Plotting the first N points of the sequence $(\log_{10} n)_{n=1}^{\infty}$ mod 1 for $N = 5$, 50 and 500. The visible points are fewer than N because some points overlap. The pattern seems to suggest that in the limit, the points are most densely distributed to one side of $\log_{10} N$, and least densely to the other side.

In contrast, the sequence $(\log_{10} n)_{n=1}^{\infty}$ is known to not have a unique distribution mod 1 [17]. However, for carefuly chosen Ns, plotting the first N points of the sequence seems to reveal a pattern. For example, if we plot the first 5, 50, and 500 points of $(\log_{10} n)_{n=1}^{\infty}$ mod 1 and compare them side by side as in Figure 0.2, it appears the points tend to be more densely distributed exactly behind the last drawn point namely $\langle \log_{10} 500\rangle = \langle \log_{10} 50\rangle = \langle \log_{10} 5\rangle$ on the circle. While it turns out that indeed for every ω_N the highest density precedes the final point $\log_{10} N$, the observed pattern is an artifact of our chosen Ns. If we choose to plot the first 10, and 100 points, the high density would precede $\langle \log_{10} 1\rangle$ on the circle (see Figure 0.3). The most dense portion of ω_N keeps rotating to $\log_{10} N$ on the circle as N increases, and hence there is no unique distribution mod 1 for $(\log_{10} n)_{n=1}^{\infty}$. We ask the question, however, if something can be said about the pattern observed in the carefully chosen subsequence of empirical distributions plotted Figure 0.2, namely $(\omega_{5\times 10^j})_{j=1}^{\infty}$, or other similar subsequences of $(\omega_N)_{N=1}^{\infty}$ like in Figure 0.3.

To describe the common pattern observed in Figures 0.2 and 0.3, take the points for one particular N, say $N = 10$ in Figure 0.3. Note that the distance between two consecutive points on the circle decreases. That is to say, the distance between the first two points $\log_{10} 1$ and $\log_{10} 2$ is larger than the distance between the next two points $\log_{10} 2$ and $\log_{10} 3$, and so on. These points, circled in red, are also present for $N = 100$, with 9 additional points 'squeezed' between them. The pattern continues for $N = 1000$ (not pictured), and so the windows to squeeze the next 9 points in get smaller, yet more numerous.

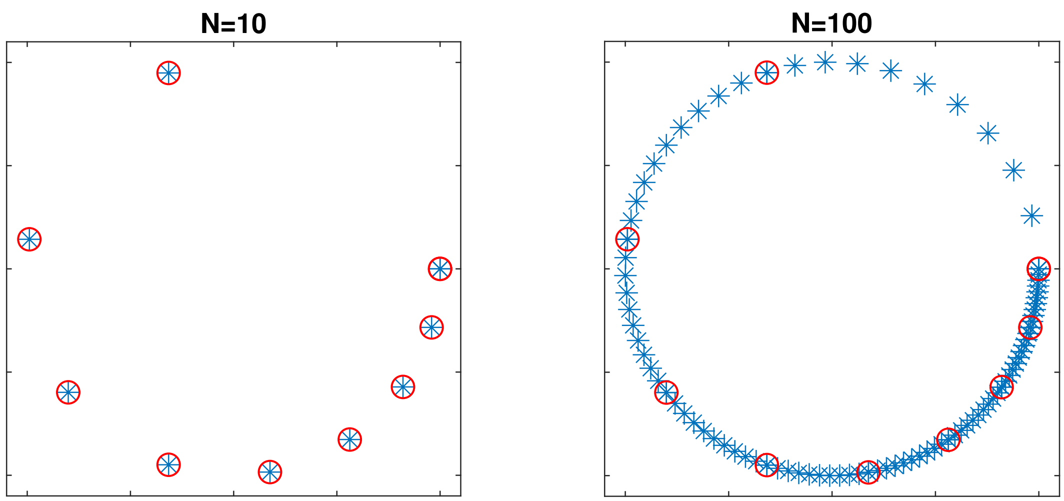

Figure 0.3: Plotting the first N points of the sequence $(\log_{10} n)_{n=1}^{\infty}$ mod 1 for $N = 10$ (circled in red), and 100. For $N = 100$ all the red-circled points are still present, and there are 9 new points inserted in between any two of them.

Once we have set up the apparatus to investigate the above question, we can also get more insight to the curious Benford's law. Reading numbers from left to right, Benford's law informally says that in many different situations, observed numbers most probably begin with the digit 1 (more than 30%), followed by 2 being less likely (about 18%), and so on all the way to 9 being the least likely first digit (less than 5%). This pattern is observed in many seemingly irrelevant recordings of numbers such as lengths of rivers, population sizes of countries, even the numbers appearing on the first page of a newspaper [5]. More precisely, the law states that for any $d \in \{1, 2, \cdots, 9\}$, the probability that the first significant digit of a number is d equals $\log_{10} \frac{d+1}{d}$. Many sequences follow this rule. The sequence of Fibonacci numbers, factorials, powers of 2, or even powers of e all follow Benford's law in the sense that the proportion of the elements whose first digit is d among the first N elements of the sequence approaches $\log_{10} \frac{d+1}{d}$ as N grows [26, 23, 22]. Take for example the sequence $(2^n)_{n=1}^{\infty}$, and for some $N \in \mathbb{N}$, count the elements that have the first digit 1. If we denote this count by $\#\left\{n \in \{1, 2, \cdots, N\} : D_1(2^n) = 1\right\}$, then

$$\frac{\#\left\{n \in \{1, 2, \cdots, N\} : D_1(2^n) = 1\right\}}{N} \xrightarrow{N \to \infty} \log_{10} 2 \quad ,$$

as demonstrated in Figure 0.4. The analogous ratio for the first digit being 2 approaches $\log_{10} \frac{3}{2}$, and so on for the other possible first digits.

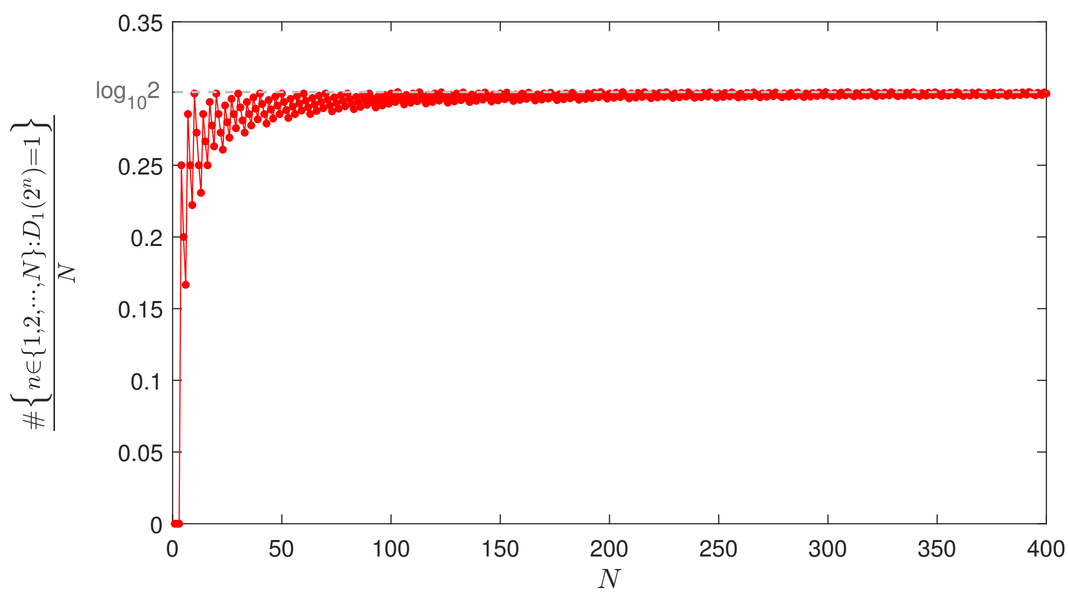

Figure 0.4: The ratio of the powers of 2 that have first digit 1 among the first N powers. The ratio approaches $\log_{10} 2$ as N grows larger.

A natural (no pun intended) question to ask is if the sequence of $\mathbb{N}$ is also Benford in the above sense. In other words, would the proportion of the first N natural numbers whose first digit is d approach $\log_{10}\frac{d+1}{d}$ as N gets larger? The answer is no. The proportion does not converge for any $d \in \{1, 2, \cdots, 9\}$. Figure 0.5 depicts the mentioned ratio for first digit being 1. Instead of approaching the value $\log_{10} 2$, the ratio seems to oscillate up and down in every order of magnitude. The oscillation pattern is present for other possible first digits as well. Focusing on Figure 0.5, the first upward region is for N between 10 and 19; and the downward region after that is for N from 20 to 99. Clearly, that is followed by another upward pattern from 100 to 199, and so on. The up-down oscillation takes a periodic form if the horizontal scale is logarithmic (see Figure 0.6).

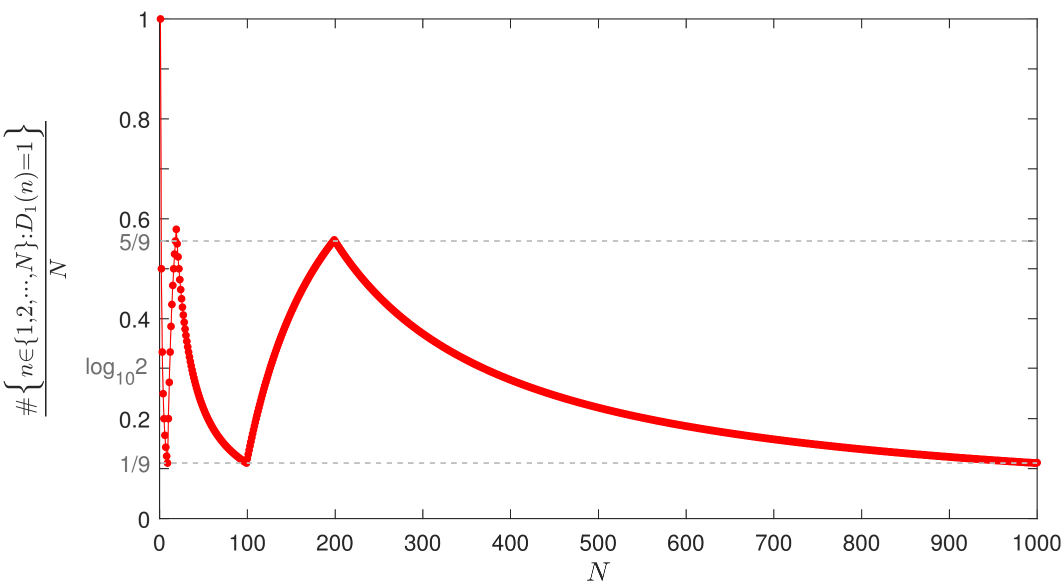

Figure 0.5: The proportion of the first N naturals with first digit 1 keeps oscillating up and down over each order of magnitude.

In this thesis, we work with the Kantorovich/1-Wasserstein metric $d_{\mathbb{T}}$ on the space $\mathcal{P}$ of all probability measures on the circle, which induces the weak topology mentioned above [10]. We introduce an explicit formula for calculating $d_{\mathbb{T}}$ between any two $\mu, \nu \in \mathcal{P}$, and through that formula we explore the basic topology of $\mathcal{P}$ with this metric. We calculate, for example, the distance between an exponential and the uniform distributions mod 1. We will see that for the sequence $(\log_{10} n)_{n=1}^{\infty}$, the subsequence of $(\omega_N)_{N=1}^{\infty}$ that is pictured in Figure 0.2 converges (in the $d_{\mathbb{T}}$ sense) to an exponential distribution mod 1 rotated so that its most dense point is at $\log_{10} 5$. More generally, we will see that if all ω_Ns are rotated so that their most dense points coincide (e.g., at $0+\mathbb{Z}$), they all converge to a unique exponential distribution. We will also see that the pattern observed in Figures 0.5 and 0.6 is the measure of an arc on the circle under a sequence of rotating empirical distributions associated with a log-like sequence, which turn out to be approximations of an exponential distribution rotating around the circle.

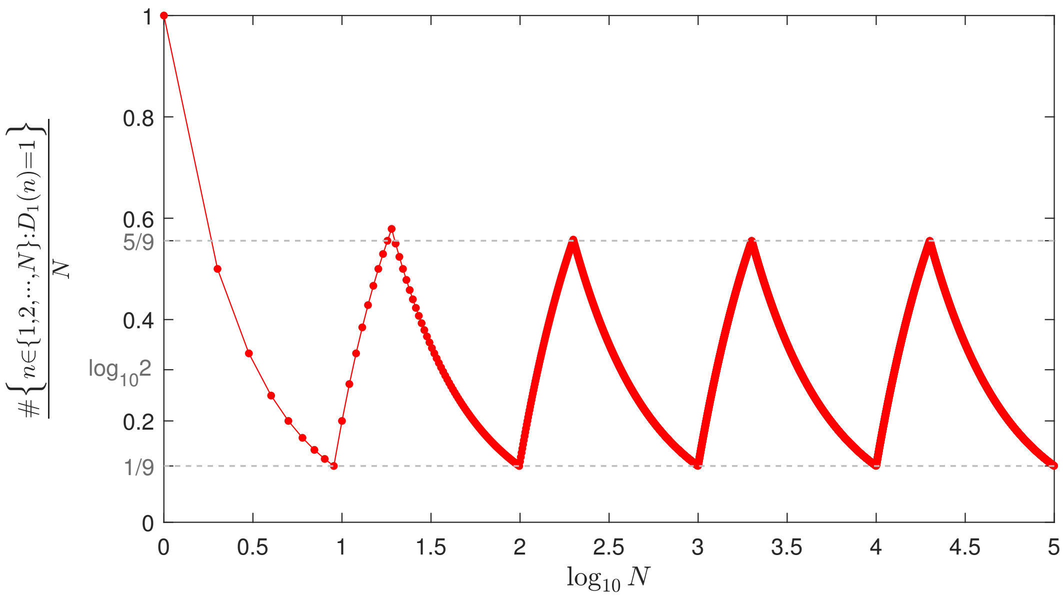

Figure 0.6: The proportion of the first N naturals with first digit 1 plotted against a logarithmic horizontal scale. The ratio seems to follow a periodic pattern.

For the reader's convenience, the preliminary work required for the main results of each chapter are presented separately in the first section of the chapter with the exception of Chapter 1 in which the first section is dedicated to notations. Additionally, all proofs are presented in a different color so that the reader has the option of skipping them at their discretion.

Chapter 1

Preliminaries

In this chapter, we formally introduce the main objects of study in this thesis. Before doing so, we first clarify the notations and conventions in Section 1.1. Additionally, a list of the symbols and their meanings is provided at the beginning of this thesis for reference. Section 1.2 includes preparatory remarks and lemmas required for the rest of the chapter.

1.1 Notations and conventions

Throughout this document, $\mathbb{C}$, $\mathbb{R}$, $\mathbb{R}^{+}$, $\overline{\mathbb{R}}$, $\mathbb{Z}$, and $\mathbb{N}$ denote the set of complex numbers, real numbers, non-negative reals, extended reals, integers, and natural numbers, respectively. The natural logarithm is always denoted log.

For every function $f\colon \mathbb{R} \to \mathbb{R}$, we call f *increasing* if for every $x, y \in \mathbb{R}$ such that $x < y$, we have $f(x) \leq f(y)$. Similarly, f is *decreasing* if for every $x, y \in \mathbb{R}$ such that $x < y$, we have $f(x) \geq f(y)$. A function that is either increasing or decreasing is *monotone*. The adverb "*strictly*" modifies the above adjectives to mean that our monotone function preserves strict inequalities. Furthermore, for every $x_0 \in \operatorname{Dom}(f)$, we denote $\lim_{\epsilon \searrow 0} f(x_0 - \epsilon)$ by $f(x_0^-)$, and likewise $\lim_{\epsilon \searrow 0} f(x_0 + \epsilon)$ by $f(x_0^+)$, provided these limits exist.

For every $t \in \mathbb{R}$, we denote the largest integer that is less than or equal to t by $\lfloor t \rfloor := \max\{k \in \mathbb{Z} : k \leq t\}$, and denote the fractional part of t by

$\langle t \rangle := t - \lfloor t \rfloor$. The number $\langle t \rangle$ is also referred to as *the residue of t modulo 1* [17], and is interpreted geometrically as the distance between t and the first integer to its left on the real line. Furthermore, $t + \mathbb{Z}$ is a shorthand notation for the set $\{\cdots, t-2, t-1, t, t+1, t+2, \cdots\}$.

For any two metric spaces (X, d_X) and (Y, d_Y), denote by $C(X;Y)$ the set of all continuous functions from domain X to codomain Y, i.e., $C(X;Y) := \big\{f \in Y^X : f \text{ is continuous}\big\}$; and similarly denote by $\mathrm{Lip}_a(X;Y)$ the set of all a-Lipschitz continuous functions, i.e.,

$$\mathrm{Lip}_a(X;Y) := \Big\{h \in C(X;Y) : d_Y\big(h(x_1), h(x_2)\big) \leq a\, d_X(x_1, x_2) \ \ \forall x_1, x_2 \in X\Big\}.$$

If Y is omitted in these notations for collections of functions, it is understood that $Y = \mathbb{R}$, e.g., $C(X) = C(X;\mathbb{R})$. Furthermore, for every $f \in Y^X$ and every $B \subseteq Y$, define the notation $\{f \in B\} := \big\{x \in X : f(x) \in B\big\}$. Lastly, denote by $\mathbb{1}_X$, Id_X , $\overline{X}$, and $\#X$, the indicator function, the identity function, the closure, and the cardinality of X, respectively.

For two measures μ and ν defined on the same measurable space, denote by $\mu \ll \nu$ absolute continuity of μ with respect to ν.

1.2 Preparatory work

The preparatory results and lemmas used to prove the main statements in this chapter are gathered in this section and are as follows.

Remark 1.2.1 (composition of continuous functions is continuous)**.** Let (X, τ_X), (Y, τ_Y) and (Z, τ_Z) be topological spaces. Let $f\colon X \to Y$ and $g\colon Y \to Z$ be continuous functions. Then $g \circ f\colon X \to Z$ is also continuous.

Remark 1.2.2 (composition of Lipschitz functions is Lipschitz)**.** Let (X, d_X), (Y, d_Y) and (Z, d_Z) be metric spaces, and let $f\colon X \to Y$ and $g\colon Y \to Z$ be Lipschitz continuous functions. Then $g \circ f\colon X \to Z$ is also Lipschitz continuous. Furthermore, if L' and L'' denote Lipschitz constants of f and g, then a Lipschitz constant of $g \circ f$ is $L'L''$.

Remark 1.2.3 (Range of a continuous function on a compact space is compact)**.** Let (X, τ_X) and (Y, τ_Y) be topological spaces, where (X, τ_X) is compact. Then for every continuous function $f\colon X \to Y$, $\text{Range}\,(f) := f\,(X)$ is also compact.

Corollary 1.2.4 (Continuous functions map compact sets to compact sets)**.** *Let (X, d_X) and (Y, d_Y) be metric spaces, and let $f\colon X \to Y$ be a continuous function. Then $f(K)$ is compact for every compact set $K \subseteq X$.*

Remark 1.2.5 (isometries are one-to-one)**.** Let (X, d_X) and (Y, d_Y) be metric spaces, and let $f\colon X \to Y$ be an isometry. Then f is injective.

Lemma 1.2.6 (auto-isometries on compact domains are onto)**.** *Let (X, d_X) be a compact metric space, and let $f\colon X \to X$ be an isometry. Then f is surjective.*

Proof. We want to show $f(X) = X$.
Clearly, $\text{Range}(f) \subseteq \text{CoDom}(f)$ and therefore $f(X) \subseteq X$. Thus it suffices to show $X \subseteq f(X)$. Assume, by contradiction, $\exists\, x \in X \setminus f(X)$.

Claim 1.2.6.1. The point x is not a limit point of $f(X)$, i.e.,

$$x \notin \overline{f(X)} \quad .$$

Proof. X is compact; thus through Remark 1.2.3, $f(X)$ is compact. If x was a limit point, then there would exist a sequence in $f(X)$ that was convergent (in X) to x. This sequence and all of its subsequences, therefore, would be divergent in $f(X)$. This contradicts the sequential compactness of $f(X)$. $\square$

By Claim 1.2.6.1, we know $d_X\left(x, f(X)\right) > 0$. Let $d := d_X\left(x, f(X)\right)$. Therefore we know

$$\forall y \in f(X), \quad d_X(x, y) \geq d \quad . \tag{1.1}$$

Consider the recursively defined sequence $(x_n)_{n=1}^{\infty} \subseteq X$ in which every element is the image of the previous element under f, i.e., consider

$$x_1 := x \qquad \wedge \qquad \forall n \geq 2, \quad x_n := f(x_{n-1}) \quad .$$

By (1.1), we know that $\forall n \geq 2, \quad d(x_1, x_n) \geq d$. This implies through the following steps that no two elements are ever closer to each other than d.

Let arbitrary $m, n \in \mathbb{N} : m \neq n$ be given. WLOG assume $m < n$. By (1.1) and the fact that composition of isometries is still an isometry we know

$$d_X(x_m, x_n) = d_X\left(f^{m-1}(x_1), f^{m-1}(x_{n-m+1})\right) \overset{\text{isometry}}{=\!=} d_X\left(x_1, x_{n-m+1}\right) \overset{(1.1)}{\geq} d \quad .$$

So no subsequence of $(x_n)_{n=2}^{\infty} \subseteq f(X)$ can be convergent. This contradicts the sequential compactness of $f(X)$. ■

Corollary 1.2.7. *We see immediately from Remark 1.2.5 and Lemma 1.2.6 that an isometry on a compact domain is invertible, and its inverse is an isometry as well.*

Definition 1.2.8. We define the function $d_{\mathbb{T}} \colon \mathbb{R} \times \mathbb{R} \to [0, \frac{1}{2}]$ as

$$\forall\, a, b \in \mathbb{R}, \quad d_{\mathbb{T}}(a, b) := \min_{k \in \mathbb{Z}} |a - b + k| \ .$$

Note that in the above definition, the function $d_{\mathbb{T}} \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$ is invariant under shifts by integers. More precisely,

$$\forall\, a, b \in \mathbb{R},\ \forall k_1, k_2 \in \mathbb{Z}, \quad d_{\mathbb{T}}(a + k_1\,,\, b + k_2) \ = \ d_{\mathbb{T}}(a, b) \ . \qquad (1.2)$$

1.3 The circle

Denote by $\mathbb{T}$ the quotient space $\mathbb{R}/\mathbb{Z}$. Note that every $x \in \mathbb{T}$ has the form $x = t + \mathbb{Z}$ for some $t \in \mathbb{R}$. This real t is not unique, and any two $t \in \mathbb{R}$ yielding the same points in $\mathbb{T}$ differ by an integer, i.e., if $t_1 + \mathbb{Z} = t_2 + \mathbb{Z}$, then $t_1 - t_2 \in \mathbb{Z}$. (The converse of this conditional statement is obviously true as well [4].) It is known that equipped with the quotient topology (i.e., the topology induced by the map $t \mapsto t + \mathbb{Z}$), $\mathbb{T}$ is a compact metrizable space [21]. Note that $d_{\mathbb{T}}$ in Definition 1.2.8 is well-defined on $\mathbb{T} \times \mathbb{T}$ in the sense that for every $x, y \in \mathbb{T}$, the value $d_{\mathbb{T}}(s, t)$ is independent of the chosen representatives $s, t \in \mathbb{R}$ of $x = s + \mathbb{Z}$ and $y = t + \mathbb{Z}$, by (1.2). When interpreted as a function on $\mathbb{T} \times \mathbb{T}$, $d_{\mathbb{T}}$ becomes a metric that induces the quotient topology. We formally define this metric in Definition 1.3.1, and we use the same notation to avoid introducing new notations.

Definition 1.3.1. We define the metric $d_{\mathbb{T}} \colon \mathbb{T} \times \mathbb{T} \to [0, \frac{1}{2}]$ as

$$\forall\, t_1 + \mathbb{Z}\,,\, t_2 + \mathbb{Z} \in \mathbb{T}, \quad d_{\mathbb{T}}\left(t_1 + \mathbb{Z}\,,\, t_2 + \mathbb{Z}\right) := \min_{k \in \mathbb{Z}} |t_1 - t_2 + k| \quad .$$

The set $\mathbb{T}$ corresponds to the unit circle centered at the origin in the complex plane $\mathbb{C}$ in a bijective fashion. One such bijection $\iota_{\mathbb{C}} \colon \mathbb{T} \to \left\{ z \in \mathbb{C} : |z| = 1 \right\}$ is defined in Definition 1.3.2. Furthermore, the bijection $\iota_{\mathbb{C}}$ is in fact a homeomorphism (see Remark 1.3.3) and therefore $(\mathbb{T}, d_{\mathbb{T}})$ is (topologically) isomorphic to $\Big(\left\{ z \in \mathbb{C} : |z| = 1 \right\} , |\cdot| \Big)$, and as such, we are justified in referring to $\mathbb{T}$ as a circle. It is convenient to geometrically interpret $(\mathbb{T}, d_{\mathbb{T}})$ as the unit circle centered at the origin in $\mathbb{C}$ where the metric $d_{\mathbb{T}}(x, y)$ is the (normalized) arclength of a shortest arc connecting the two points $\iota_{\mathbb{C}}(x)$, $\iota_{\mathbb{C}}(y)$ on the circle. To define the bijection $\iota_{\mathbb{C}}$, note that just as was the case with $d_{\mathbb{T}}$ in Definition 1.2.8, the complex-valued function of a real variable $t \mapsto e^{t2\pi i}$ is invariant under shifts by integers, and thus $\iota_{\mathbb{C}}$ in Definition 1.3.2 is well-defined.

Definition 1.3.2. We define $\iota_{\mathbb{C}} \colon \mathbb{T} \to \left\{ z \in \mathbb{C} : |z| = 1 \right\}$ as

$$\forall\, t + \mathbb{Z} \in \mathbb{T}, \qquad \iota_{\mathbb{C}}\left(t + \mathbb{Z}\right) \;:=\; e^{t2\pi i} \quad .$$

Remark 1.3.3. [6] The bijection $\iota_{\mathbb{C}}$ defined in Definition 1.3.2 satisfies

$$\forall x, y \in \mathbb{T}, \quad 4\, d_{\mathbb{T}}(x, y) \leq \big|\iota_{\mathbb{C}}(x) - \iota_{\mathbb{C}}(y)\big| \leq 2\pi\, d_{\mathbb{T}}(x, y) \quad ,$$

and is therefore bi-Lipschitz continuous. Thus $(\mathbb{T}, d_{\mathbb{T}})$ is Lipschitz isomorphic [11] (aka Lipschitz equivalent) to $\Big(\left\{ z \in \mathbb{C} : |z| = 1 \right\} , |\cdot| \Big)$. This implies that $(\mathbb{T}, d_{\mathbb{T}})$ is a compact metric space.

Denote by $\mathcal{P}$ the set of all probability measures on the measurable space $(\mathbb{T}, \mathcal{B}_{\mathbb{T}})$, where $\mathcal{B}_{\mathbb{T}}$ denotes the Borel σ-algebra on $\mathbb{T}$. Every $\mu \in \mathcal{P}$ is uniquely determined by its (cumulative) distribution function F_μ as defined in Definition 1.3.4. Note that despite our use of the usual notation F for a cumulative distribution function, Definition 1.3.4 is not exactly the same as the usual definition, yet is closely related and contains the same information. This definition is comprehended more easily in light of the bijection $\iota_{\mathbb{R}}$ defined in Definition 1.3.6.

Definition 1.3.4. For every $\mu \in \mathcal{P}$, we define the associated distribution function $F_\mu \colon \mathbb{R} \to \mathbb{R}$ as

$$\forall\, t \in \mathbb{R}, \quad F_\mu(t) := \mu\left(\left\{s + \mathbb{Z} : s \in [0, \langle t \rangle]\right\}\right) + \lfloor t \rfloor \quad .$$

Every probability distribution function F_μ, as defined in Definition 1.3.4, is increasing and right-continuous, and $F_\mu(0) \geq 0$. Additionally, the function $t \mapsto F_\mu(t) - t$ is 1-periodic, and $F_\mu(1^-) = 1$. Conversely, every function with these properties is the distribution function of a (unique) $\mu \in \mathcal{P}$.

In talking about measures we often use the concept of a *pushforward measure* or *image measure*. Every measurable function from a measure space to a measurable space induces a measure on the codomain that is in a sense the domain measure 'pushed forward' onto the codomain by the measurable function.

Definition 1.3.5 (Pushforward measure)**.** Let $(X, \mathcal{M}, \mu)$ be a measure space and $(Y, \mathcal{N})$ a measurable space. If $f \colon X \to Y$ is an $\mathcal{M}$-$\mathcal{N}$-measurable function, then $\nu := \mu \circ f^{-1}$ is a measure on $(Y, \mathcal{N})$ — the *pushforward* of μ under f.

We now turn to introduce the bijection $\iota_{\mathbb{R}}$ which provides scaffolding not only for a more intuitive interpretation of probability measures in $\mathcal{P}$ and their distribution functions, but also for the calculations that follow in Chapter 4.

As a set, $\mathbb{T}$ also corresponds to the unit interval $[0, 1)$ in a bijective fashion. One such bijection is $\iota_{\mathbb{R}} \colon \mathbb{T} \to [0,1)$ defined in Definition 1.3.6. If equipped with the right topology, the interval $[0, 1)$ can become topologically isomorphic to $\mathbb{T}$ through $\iota_{\mathbb{R}}$. The metric $d_{\mathbb{T}}$ as defined in Definition 1.2.8 gives the right topology when restricted to $[0, 1)$. With the Euclidean metric $|\cdot|$, however, $[0, 1)$ is not topologically isomorphic to the circle $\mathbb{T}$ (see Remark 1.3.7). To define the bijection $\iota_{\mathbb{R}}$, we again note that just like the map $t \mapsto e^{t2\pi i}$, the map $\langle \cdot \rangle \colon \mathbb{R} \to [0,1)$ as defined in Section 1.1 outputs the same value for all equivalent ts in every $t+\mathbb{Z} \in \mathbb{T}$; and thus $\iota_{\mathbb{R}}$ in Definition 1.3.6 is well-defined.

Definition 1.3.6. We define $\iota_{\mathbb{R}} \colon \mathbb{T} \to [0,1)$ as

$$\forall\, t+\mathbb{Z} \in \mathbb{T}, \qquad \iota_{\mathbb{R}}\,(t+\mathbb{Z}) \,:=\, \langle\, t\,\rangle \quad .$$

Remark 1.3.7. The bijection $\iota_{\mathbb{R}} \colon \mathbb{T} \to [0,1)$ defined in Definition 1.3.6 is bi-measurable. While $\iota_{\mathbb{R}}$ is not continuous, $\iota_{\mathbb{R}}^{-1} \colon [0,1) \to \mathbb{T}$ is 1-Lipschitz.

While $\big([0,1)\,, |\cdot|\big)$ is not topologically isomorphic to $(\mathbb{T}, d_{\mathbb{T}})$, it is of interest because of our familiarity with the calculus of real-valued functions on $[0, 1)$ (see Theorem 2.2.8). Using the measurable map $\iota_{\mathbb{R}}$, we can think of any $\mu \in \mathcal{P}$ as a probability distribution on the measurable space $\big([0,1)\,, \mathcal{B}_{[0,1)}\big)$. The F_μ defined in Definition 1.3.4 is the result of extending the cumulative distribution function of such a pushforward measure by shifting copies of its graph one unit to the right and unit up repeatedly to form a 'diagonal' in the first quadrant of $\mathbb{R}^2$, and similarly one unit to the left and one unit down for the third quadrant's 'diagonal'. To see this, note that for any given $\mu \in \mathcal{P}$,

$$F_\mu(s) = \mu \circ \iota_{\mathbb{R}}^{-1}\left(\,[0,s]\,\right) \qquad \forall\, s \in [0,1) \quad . \tag{1.3}$$

A few examples of probability measures in $\mathcal{P}$ and their distribution functions are given in Section 1.4. In that section, we will see that despite not being our approach later on, each of the examples of interest could be defined as push-forward measures (under $\iota_{\mathbb{R}}^{-1}$) of a probability measure that we know on $[0, 1)$. Similarly, the bijection $\iota_{\mathbb{R}}$ allows us to identify any function on $\mathbb{T}$

as a function on $[0,1)$. Thus from now on, for any function h on $\mathbb{T}$ and any $s \in [0,1)$, the notation $h(s)$ is understood to mean $h \circ \iota_{\mathbb{R}}^{-1}(s)$. Not surprisingly we define the derivative of any h to be that of $h \circ \iota_{\mathbb{R}}^{-1}$.

Definition 1.3.8 (Derivative of a function on $\mathbb{T}$)**.** Let $h \colon \mathbb{T} \to \mathbb{R}$ be a real-valued function, $x_0 \in \mathbb{T}$, and $s_0 := \iota_{\mathbb{R}}(x_0)$. We say h is *differentiable at* x_0 if $h \circ \iota_{\mathbb{R}}^{-1} \circ \langle\cdot\rangle \colon \mathbb{R} \to \mathbb{R}$ is differentiable at s_0 in the usual sense; and we define the derivative of h at x_0 as

$$h'(x_0) := \left(h \circ \iota_{\mathbb{R}}^{-1} \circ \langle\cdot\rangle \right)'(s_0) \quad ,$$

where $\left(h \circ \iota_{\mathbb{R}}^{-1} \circ \langle\cdot\rangle \right)'(s_0)$ is understood to be $\lim_{t \to s_0} \frac{h \circ \iota_{\mathbb{R}}^{-1} \circ \langle\cdot\rangle(t) - h \circ \iota_{\mathbb{R}}^{-1} \circ \langle\cdot\rangle(s_0)}{t - s_0}$.

If h is differentiable at every $x_0 \in \mathbb{T}$, we call it *differentiable (everywhere)*.

1.4 Some examples of measures in $\mathcal{P}$

There are a few elements of $\mathcal{P}$ that are of particular interest to us and they are as follows.

Perhaps the simplest family of probability distributions is the family of *point mass* (or *Dirac*) distributions denoted by δ_{x_0} for every point $x_0 \in \mathbb{T}$, in which all probability mass is assigned to a single point $x_0 \in \mathbb{T}$, or $s_0 \in [0,1)$ if one sees the distribution on $[0,1)$:

$$\forall x_0 \in \mathbb{T}, \qquad \forall B \in \mathcal{B}_{\mathbb{T}}, \quad \delta_{x_0}(B) := \mathbb{1}_B(x_0) \quad .$$

The distribution function of δ_{x_0} satisfies

$$\forall s \in [0,1)\,, \qquad F_{\delta_{x_0}}(s) = \begin{cases} 0 & s < s_0 \quad , \\ 1 & s \geq s_0 \quad , \end{cases}$$

where $s_0 = \iota_{\mathbb{R}}(x_0)$. We could have used this formula of $F_{\delta_{x_0}}$ to *define* the measure δ_{x_0} — an approach we take for the remaining probability measures in this section. Note that the formula above is the same as the distribution function of δ_{s_0} on $[0,1)$. Since every distribution function uniquely determines a probability measure, we see through (1.3) that $\delta_{x_0} \circ \iota_{\mathbb{R}}^{-1} = \delta_{s_0}$. Conversely, we conclude that $\delta_{x_0} = \delta_{s_0} \circ \left(\iota_{\mathbb{R}}^{-1}\right)^{-1}$.

Another simple element of $\mathcal{P}$ is the *uniform* (or *Lebesgue*) distribution denoted by $\lambda_{\mathbb{T}}$ in which every open arc on the circle $\mathbb{T}$ is assigned its (normalized) arclength as its probability mass. We define the distribution function $F_{\lambda_{\mathbb{T}}}$ as follows, and thereby we will have fully defined $\lambda_{\mathbb{T}}$:

$$\forall s \in [0,1) \qquad F_{\lambda_{\mathbb{T}}}(s) := s \quad .$$

Again, note that the above formula is the same as the distribution function of λ, the familiar Lebesgue measure on $[0,1)$. By uniqueness of distribution functions, we see through (1.3) that $\lambda_{\mathbb{T}} \circ \iota_{\mathbb{R}}^{-1} = \lambda$ and therefore $\lambda_{\mathbb{T}} = \lambda \circ \left(\iota_{\mathbb{R}}^{-1}\right)^{-1}$.

Remark 1.4.1. It is noteworthy that if $\mu \in \mathcal{P}$ is absolutely continuous with respect to $\lambda_{\mathbb{T}}$, then $\mu \circ \iota_{\mathbb{R}}^{-1}$ is absolutely continuous with respect to λ, with density

$$\frac{d\left(\mu \circ \iota_{\mathbb{R}}^{-1}\right)}{d\lambda} = \frac{d\mu}{d\lambda_{\mathbb{T}}} \circ \iota_{\mathbb{R}}^{-1} \quad .$$

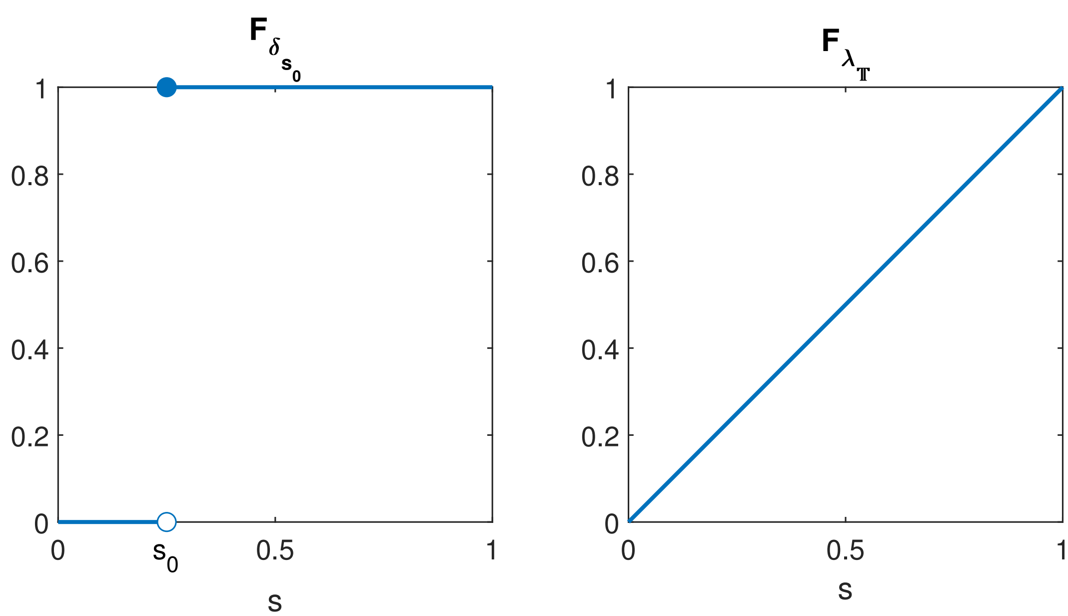

Figure 1.1: The distribution functions of δ_{s_0} and $\lambda_{\mathbb{T}}$ plotted on $[0,1)$.

In this thesis, an important family in $\mathcal{P}$ consists of *exponential* distributions, denoted η_a for every non-zero real a. We define η_a via its distribution function as

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

One may wonder why η_a is called 'exponential'. This is because it can be shown (see Proposition 1.4.2) that η_a is the distribution of $\langle \frac{-1}{a} Y \rangle + \mathbb{Z}$ where $Y \sim \text{Expo}(1)$. Before demonstrating this, we must recall that if X is a random variable such that $X \sim \text{Expo}(\theta)$ where $\theta > 0$ denotes the rate parameter, then $\frac{1}{a} X \sim \text{Expo}(a\,\theta)$ for every $a > 0$ [9]. We also recall that if X has an exponential distribution with rate $\theta > 0$, i.e., if X has the density function

$$\forall t \in \mathbb{R}, \qquad f_X(t) = \begin{cases} \theta\, e^{-\theta t} & t \geq 0 \quad , \\ 0 & t < 0 \quad , \end{cases}$$

then $-X$ has the density function

$$\forall t \in \mathbb{R}, \qquad f_{-X}(t) = \begin{cases} 0 & t > 0 \quad , \\ \theta\, e^{\theta t} & t \leq 0 \quad . \end{cases}$$

Proposition 1.4.2. Let $a \in \mathbb{R}\backslash\{0\}$ be given, and let $Y\colon\ (\Omega, \mathcal{M}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ be a standard exponential random variable, i.e., let $Y \sim \text{Expo}(1)$. Then the function $s \mapsto \frac{e^{as}-1}{e^a-1}$ is the distribution function of $\langle -\frac{1}{a} Y \rangle$ on $[0,1)$.

Proof. Let arbitrary $Y \sim \text{Expo}(1)$ and an arbitrary $a \in \mathbb{R} \setminus \{0\}$ be given. We want to show that $F_{\langle \frac{-1}{a} Y \rangle}(s) = \frac{e^{as}-1}{e^a-1} \quad \forall s \in [0,1)\,.$
Let arbitrary $s \in [0,1)$ be given.

Case 1 $\quad a > 0 \quad .$

We know that $\quad f_{-\frac{1}{a}Y}(t) = \begin{cases} 0 & t > 0 \quad , \\ a\, e^{at} & t \leq 0 \quad . \end{cases}$

By definition of the distribution function,

$$F_{\langle \frac{-1}{a} Y \rangle}(s) \;=\; \mathbb{P}\left(\left\{ \langle -\frac{1}{a} Y \rangle \leq s \right\}\right)$$

$$= \mathbb{P}\left(\biguplus_{n=0}^{\infty}\left\{-1-n \le \frac{-1}{a}Y \le -(1-s)-n\right\}\right)$$

$$= \sum_{n=0}^{\infty}\int_{-1-n}^{s-1-n} a\,e^{at}\,dt \;=\; \lim_{N\to\infty}\left(\sum_{n=0}^{N}\int_{-1}^{s-1} a\,e^{a(t-n)}\,dt\right)$$

$$= \lim_{N\to\infty}\left(\sum_{n=0}^{N} e^{-a\,n}\int_{-1}^{s-1} a\,e^{a\,t}\,dt\right)$$

$$= \int_{-1}^{s-1} a\,e^{a\,t}\,dt \;\lim_{N\to\infty}\sum_{n=0}^{N} e^{-a\,n} \;=\; \left[e^{at}\right]_{-1}^{s-1}\,\frac{1}{1-e^{-a}}$$

$$= \frac{e^{a(s-1)}-e^{-a}}{1-e^{-a}} \;=\; \frac{\cancel{e^{-a}}(e^{as}-1)}{\cancel{e^{-a}}(e^{a}-1)} \;=\; \frac{e^{as}-1}{e^{a}-1}\;.$$

Case 2 $a<0$

The proof of this case is analogous to that of Case 1.

Thus we have shown that $\forall a \in \mathbb{R}\setminus\{0\}, \quad F_{\langle\frac{-1}{a}Y\rangle}(s) = \frac{e^{as}-1}{e^{a}-1} \quad \forall s \in [0,1)\,.$ ■

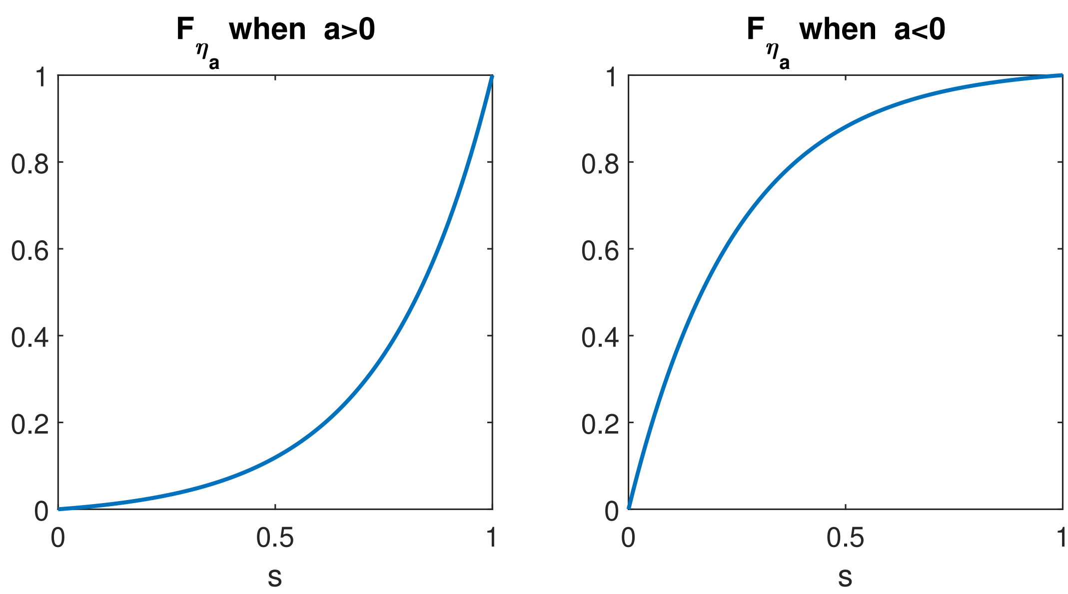

Figure 1.2: The distribution function of η_a plotted on $[0,1)$ when a is positive (left) and when a is negative (right). The closer a is to 0, the more F_{η_a} resembles $F_{\lambda_{\mathbb{T}}}$.

The final family we introduce in $\mathcal{P}$ consists of distributions ζ_a where a is a non-zero real. We define ζ_a via its distribution function as

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

These distributions are examples of Gamma distributions because ζ_a is the distribution of $\langle \frac{-1}{a}Y_1 + \frac{-1}{a}Y_2 \rangle + \mathbb{Z}$ where $Y_1, Y_2 \overset{\text{i.i.d.}}{\sim} \text{Expo}(1)$.

1.5 $\mathcal{P}$ as a compact metric space

In this section we topologize $\mathcal{P}$ by defining a convergence criterion for a sequence in $\mathcal{P}$. We then introduce a metric that induces that topology.

Note that with the notion of distance $d_{\mathbb{T}} \colon \mathbb{T} \times \mathbb{T} \to \mathbb{R}^+$ defined on $\mathbb{T}$ (see Definition 1.3.1), the notion of continuity is well-defined for any real-valued function on $\mathbb{T}$. In addition, the set $\text{Lip}_1(\mathbb{T};\mathbb{R})$ of 1-Lipschitz continuous functions is also well-defined.

Consider the fact that $\mathbb{T}$ is a compact metric space. Thus the space of all finite Borel measures with the total variation norm is isometrically isomorphic to the (normed) space of bounded linear functionals on $C(\mathbb{T})$, i.e., the dual space of $C(\mathbb{T})$ denoted by $C(\mathbb{T})^*$ [15]. Under this isomorphism, every probability measure $\mu \in \mathcal{P}$ is identified with the functional φ_μ given by $\varphi_\mu(h) := \int_{\mathbb{T}} h\, d\mu$ for every $h \in C(\mathbb{T})$. The topology that we define on $\mathcal{P}$ is through the notion of weak convergence in the following sense: A sequence $(\mu_n)_{n=1}^\infty$ converges to μ in $\mathcal{P}$ iff $\left(\varphi_{\mu_n}\right)_{n=1}^\infty$ converges pointwise to φ_μ; in other words, iff for every $h \in C(\mathbb{T})$, the sequence $\left(\int_{\mathbb{T}} h\, d\mu_n\right)_{n=1}^\infty$ converges to $\int_{\mathbb{T}} h\, d\mu$ in $\mathbb{R}$. More succinctly,

$$\mu_n \xrightarrow{n\to\infty} \mu \quad \Longleftrightarrow \quad \forall h \in C(\mathbb{T})\,, \quad \int_{\mathbb{T}} h\, d\mu_n \xrightarrow{n\to\infty} \int_{\mathbb{T}} h\, d\mu \quad . \tag{1.4}$$

In the terminology of linear analysis, this topology corresponds to the weak-$*$ topology on $\mathcal{P}$, understood as a subset of $C(\mathbb{T})^*$. With this topology, $\mathcal{P}$ is compact and metrizable [21]. A metric that induces the weak topology is the following metric known as the Kantorovich or the 1-Wasserstein metric [10].

Definition 1.5.1 (Kantorovich/1-Wasserstein metric)**.** The *Kantorovich metric* $d_{\mathbb{T}} \colon \mathcal{P} \times \mathcal{P} \to \mathbb{R}^{+}$ is defined as

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

In Chapter 2 we will prove that the above supremum is in fact a maximum (see Corollary 2.1.5.) Even then, however, the computation of the Kantorovich distance remains difficult. A less abstract and easier-to-use formula for $d_{\mathbb{T}}$ is introduced in Theorem 2.2.8.

Apart from convenience, the main reason why we use the same symbol $d_{\mathbb{T}}$ for both the metric on $\mathbb{T}$ and the metric on $\mathcal{P}$ is that the Kantorovich metric in Definition 1.5.1 is an extension of $d_{\mathbb{T}}$ as defined in Definition 1.3.1 in that $d_{\mathbb{T}}\left(\delta_x , \delta_y\right) = d_{\mathbb{T}}(x, y)$ for every $x, y \in \mathbb{T}$ (see Theorem 1.5.3). Thus through the map $x \mapsto \delta_x$, the circle $(\mathbb{T}, d_{\mathbb{T}})$ is isometrically isomorphic to $\{\delta_x \in \mathcal{P} : x \in \mathbb{T}\} \subseteq \mathcal{P}$. That is to say, once $\mathcal{P}$ is equipped with the Kantorovich metric $d_{\mathbb{T}}$, the map $x \mapsto \delta_x$ becomes an isometric embedding of $\mathbb{T}$ into $\mathcal{P}$.

Lemma 1.5.2. *Let (X, d) be a metric space, and let an arbitrary $y \in X$ be given. Define the function $h_y \colon X \to \mathbb{R}$ as $h_y(\cdot) := d(\,\cdot\,, y)$. Then $h_y \in \mathrm{Lip}_1(X)$.*

Proof. We want to show $\left|h_y(b) - h_y(a)\right| \leq d(a, b)$ for every $a, b \in X$.
Let arbitrary $a, b \in X$ be given. By the triangle inequality,

$$d(b, y) \leq d(b, a) + d(a, y) \quad . \tag{1.5}$$

It is now readily seen that

$$\left|h_y(b) - h_y(a)\right| \overset{\text{def'n of } h_y}{=\!=} \left|d(b, y) - d(a, y)\right| \overset{(1.5)}{\leq} d(b, a) \overset{d \text{ symmetric}}{=\!=} d(a, b) \quad .$$

■

Theorem 1.5.3. *For every* $x, y \in \mathbb{T}$, $\quad d_{\mathbb{T}}\left(\delta_x, \delta_y\right) = d_{\mathbb{T}}(x, y)$.

Proof. Let arbitrary $x, y \in \mathbb{T}$ be given. On the one hand, by definition,

$$
\begin{aligned}
d_{\mathbb{T}}\left(\delta_x, \delta_y\right) &= \sup_{h \in \mathrm{Lip}_1(\mathbb{T})} \left(\int_{\mathbb{T}} h \, d\delta_x - \int_{\mathbb{T}} h \, d\delta_y \right) \\
&= \sup_{h \in \mathrm{Lip}_1(\mathbb{T})} \big(h(x) - h(y) \big) \le \sup_{h \in \mathrm{Lip}_1(\mathbb{T})} \big| h(x) - h(y) \big| \overset{h \,\in\, \mathrm{Lip}_1(\mathbb{T})}{\le} d_{\mathbb{T}}(x, y) \ .
\end{aligned}
$$

On the other hand, considering the function $h_y \colon \mathbb{T} \to \mathbb{R}$ given by $h_y(\cdot) := d_{\mathbb{T}}(\,\cdot\,, y)$, We know

$$
\begin{aligned}
d_{\mathbb{T}}(x, y) = h_y(x) - h_y(y) \quad &= \quad \int_{\mathbb{T}} h_y \, d\delta_x - \int_{\mathbb{T}} h_y \, d\delta_y \\
&\overset{\text{Lemma 1.5.2}}{\le} \quad \sup_{h \in \mathrm{Lip}_1(\mathbb{T})} \left(\int_{\mathbb{T}} h \, d\delta_x - \int_{\mathbb{T}} h \, d\delta_y \right) = d_{\mathbb{T}}\left(\delta_x, \delta_y\right) \ .
\end{aligned}
$$

Thus $\quad d_{\mathbb{T}}\left(\delta_x, \delta_y\right) = d_{\mathbb{T}}(x, y)$. ■

Note that Theorem 1.5.3 says in particular that the mapping $x \mapsto \delta_x$ is an isometric embedding of $\mathbb{T}$ into $\mathcal{P}$. Thus, the set $\{\delta_x \in \mathcal{P} : x \in \mathbb{T}\}$ may be viewed as a circle. In Chapter 4, we will examine in detail the distance between the probability measures introduced earlier, which will complete this informal view on the space $(\mathcal{P}, d_{\mathbb{T}})$; see Figure 4.4.

1.6 Rotating $\mathbb{T}$ and $\mathcal{P}$

In this section, we introduce the rotation and reflection transformations that will become important later in Chapters 4 and 5. We also consider the push-forward probability measures under rotation. We then conclude that the Kantorovich distance between probability measures is invariant under rotation.

Considering $(\mathbb{T}, d_{\mathbb{T}})$ as a circle in the complex plane, for every $t + \mathbb{Z} \in \mathbb{T}$, we can rotate every point of the circle counterclockwise by $2\pi t$. Another relevant transformation would be to reflect the points about the real axis.

Definition 1.6.1. Let $t \in \mathbb{R}$ be a real number. We define the *rotation (isometry)* $R_t \colon \mathbb{T} \to \mathbb{T}$ as

$$\forall x \in \mathbb{T}, \quad R_t(x) := x + t \quad ,$$

and the *reflection (isometry)* $Q \colon \mathbb{T} \to \mathbb{T}$ as

$$\forall x \in \mathbb{T}, \quad Q(x) := -x \quad .$$

We additionally interpret $R_{t+\mathbb{Z}}$ to mean R_t; and thereby R_y is well-defined for every $y \in \mathbb{T}$.

Remark 1.6.2. Note that for every $s, t \in \mathbb{R}$,

$$R_t^{-1} = R_{-t} \ , \quad R_t \circ R_s = R_{t+s} \ , \quad Q \circ R_t = R_{-t} \circ Q \ , \quad \text{and} \ \ Q^{-1} = Q \quad .$$

Since $R_t \colon \mathbb{T} \to \mathbb{T}$ is (bi)measurable, we can consider the pushforward of any $\mu \in \mathcal{P}$ under R_t (or R_t^{-1}). The same is true for Q (or Q^{-1}). When thinking about $\mu \circ R_t^{-1}$, we can either think that the input set (i.e., the set to be measured) is rotated by $-t$ and then measured by μ, or we can think of the probability measure μ itself being rotated around the circle by t while the input sets remain where they are. See Figure 1.3 for an example. With this view, the mapping $\mu \mapsto \mu \circ R_t^{-1}$ can be thought of as a rotation defined on $\mathcal{P}$. This mapping is an isometry too (see Theorem 1.6.5). In other words, the metric $d_{\mathbb{T}} \colon \mathcal{P} \times \mathcal{P} \to \mathbb{R}^+$ is invariant under rotations. In symbols,

$$\forall \mu, \nu \in \mathcal{P}, \quad \forall t \in \mathbb{R}, \quad d_{\mathbb{T}}(\mu, \nu) = d_{\mathbb{T}}\left(\mu \circ R_t^{-1}, \nu \circ R_t^{-1}\right) \quad . \tag{1.6}$$

Similarly, $\mu \mapsto \mu \circ Q^{-1}$ reflects measures in $\mathcal{P}$ and $d_{\mathbb{T}}$ is invariant under it. It is easy to see that the rotated (or reflected) version of the uniform distribution is the uniform distribution itself, and that the rotated (or reflected) version of a Dirac measure is again a Dirac measure with the massive point rotated (or reflected).

Remark 1.6.3. Note that $\lambda_{\mathbb{T}} = \lambda_{\mathbb{T}} \circ Q^{-1}$, and that for every $t \in \mathbb{R}$ and every $x_0 \in \mathbb{T}$,

$$\lambda_{\mathbb{T}} = \lambda_{\mathbb{T}} \circ R_t^{-1}\,, \quad \delta_{x_0} \circ R_t^{-1} = \delta_{x_0+t}\,, \quad \delta_{x_0} \circ Q^{-1} = \delta_{-x_0}\,, \quad \text{and} \quad \eta_a \circ Q^{-1} = \eta_{-a}\,,$$

for every $a \in \mathbb{R} \setminus \{0\}$.

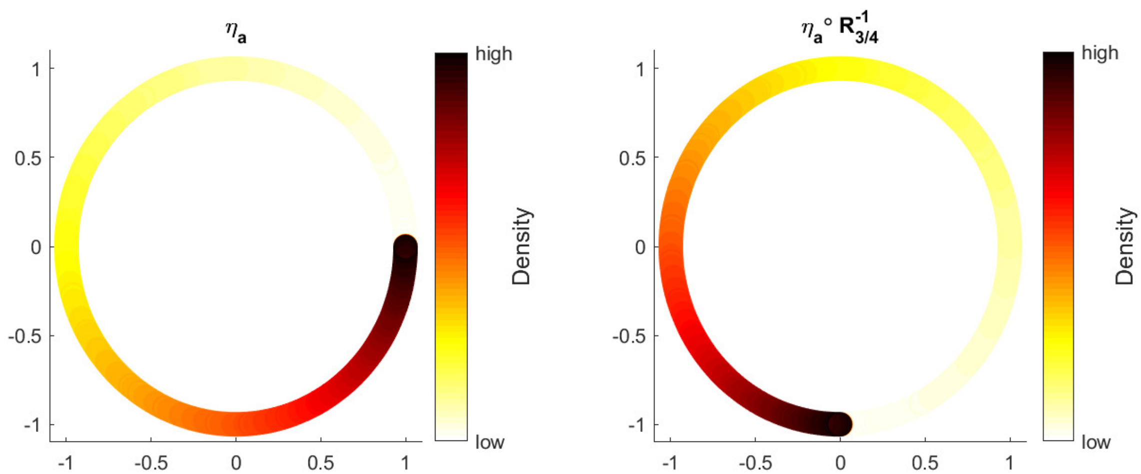

Figure 1.3: Schematic depiction of the probability distributions η_a (left) and $\eta_a \circ R_{3/4}^{-1}$ (right) where $a > 0$. The dark red color represents the highest probability density, and the light yellow color the lowest.

The rotated versions of η_a will be of particular importance in Chapter 5. Considering the example depicted in Figure 1.3, it is easy to see that with the probability density of η_a rotated, the distribution function of $\eta_a \circ R_t^{-1}$ on $[0, 1)$ is essentially the distribution function of η_a on $[0, 1)$ cut at $\langle t \rangle$ and rearranged so that the section on $[0, \langle t \rangle)$ and the section on $[\langle t \rangle, 1)$ are switched to produce a continuous function (see Figure 1.4).

Another way of thinking about this is to first recall that F_{η_a}'s graph on $[0, 1)$ is the graph of F_{η_a} on $\mathbb{R}$ in the sense of Definition 1.3.4 viewed through the $[0, 1) \times [0, 1)$ frame. Similarly, we can think of the graph of $F_{\eta_a \circ R_t^{-1}}$ on $[0, 1)$ to be the graph of F_{η_a} viewed through the 1×1 frame shifted by $1 - \langle t \rangle$ to the right and by $F_{\eta_a}(1 - \langle t \rangle)$ upward (see Figure 1.5). Thinking of rotation

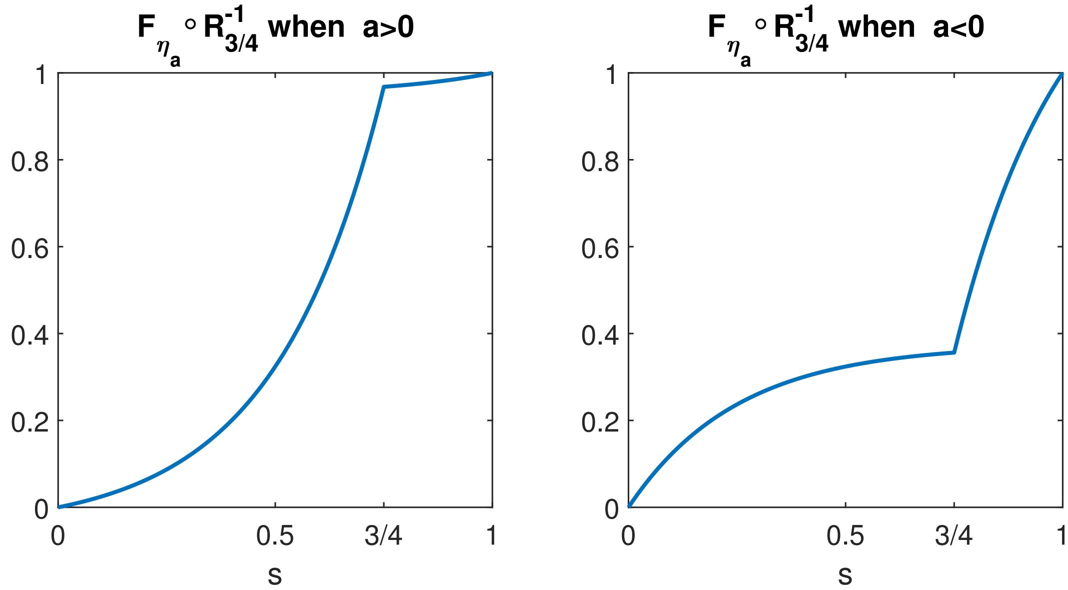

Figure 1.4: The distribution function of $\eta_a \circ R_{3/4}^{-1}$ plotted on $[0, 1)$ when a is positive (left), and when a is negative (right).

as a shift of frame for viewing the distribution function of a measures will be useful in Theorem 4.2.1.

It is thus not hard to see that once η_a is rotated by R_t , we have

$$\forall s \in [0, \langle t \rangle), \quad F_{\eta_a \circ R_t^{-1}}(s) = F_{\eta_a}(1 - \langle t \rangle + s) - F_{\eta_a}(1 - \langle t \rangle) \quad ,$$
$$\forall s \in [\langle t \rangle, 1), \quad F_{\eta_a \circ R_t^{-1}}(s) = F_{\eta_a}(1) - F_{\eta_a}(1 - \langle t \rangle) + F_{\eta_a}(s - \langle t \rangle) \quad .$$

which simplifies to the following

$$\forall s \in [0, 1), \qquad F_{\eta_a \circ R_t^{-1}}(s) = \begin{cases} \frac{e^{a(1-\langle t \rangle)}(e^{as}-1)}{e^a - 1} & \text{if } s \in [0, \langle t \rangle) \quad , \\ 1 - e^{-a\langle t \rangle} \frac{e^a - e^{as}}{e^a - 1} & \text{if } s \in [\langle t \rangle, 1) \quad . \end{cases} \tag{1.7}$$

We will later learn in Chapter 4 that for any fixed exponential distribution η_a, two rotated versions of it, say, $\eta_a \circ R_t^{-1}$ and $\eta_a \circ R_t^{-1}$, are close in $\mathcal{P}$ if and only if s and t are close in $\mathbb{T}$; see Theorem 4.2.8 for a precise statement.

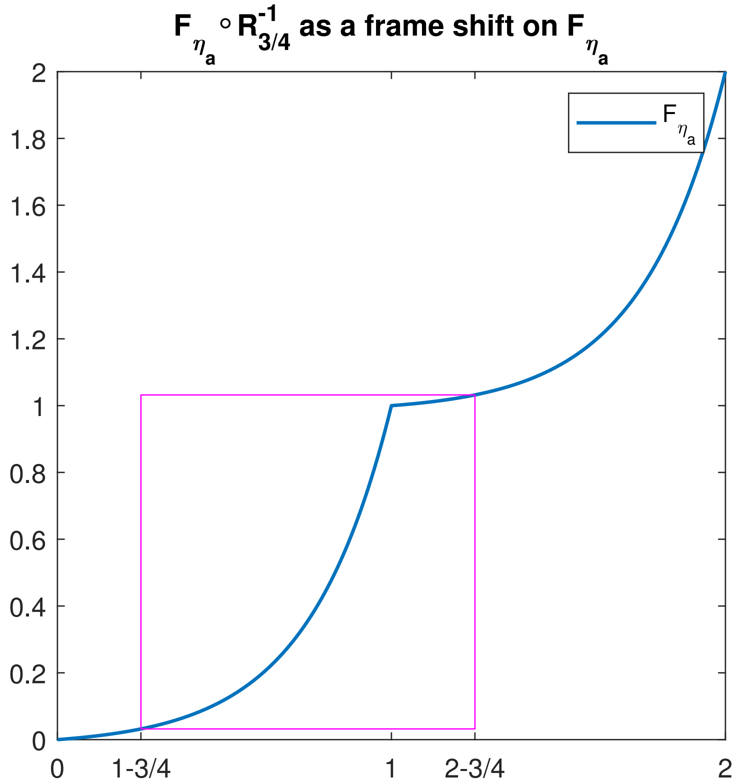

Figure 1.5: The graph of $F_{\eta_a} \circ R_{3/4}^{-1}$ plotted on $[0, 1)$ is the graph of F_{η_a} viewed through the 1×1 frame shifted by $1 - t$ to the right and by $F_{\eta_a}(1 - t)$ upward. In this example a is positive.

The remainder of this section proves a few statements in general about the mapping that takes a $\mu \in \mathcal{P}$ and maps it to its pushforward. By the end of this section the truth of $d_{\mathbb{T}}$'s invariance will have been shown not just for rotation and reflection, but also for any auto-isometry on $\mathbb{T}$.

Theorem 1.6.4. *Let* $S \colon \mathbb{T} \to \mathbb{T}$ *be a continuous function. Then the map* $\mu \mapsto \mu \circ S^{-1}$ *is also continuous.*

Proof. Let an arbitrary convergent sequence $(\mu_n)_{n=1}^{\infty}$ in $\mathcal{P}$ be given, and let $\mu \in \mathcal{P}$ denote $\lim_{n \to \infty} \mu_n$. By definition of weak convergence,

$$\forall \tilde{h} \in C(\mathbb{T}), \quad \lim_{n \to \infty} \int_{\mathbb{T}} \tilde{h} \, d\mu_n = \int_{\mathbb{T}} \tilde{h} \, d\mu \quad . \tag{1.8}$$

We want to show

$$\forall h \in C(\mathbb{T}), \quad \lim_{n\to\infty} \int_{\mathbb{T}} h\, d\left(\mu_n \circ S^{-1}\right) = \int_{\mathbb{T}} h\, d\left(\mu \circ S^{-1}\right) \quad .$$

Let arbitrary $h \in C(\mathbb{T})$ be given. Then,

$$\lim_{n\to\infty} \int_{\mathbb{T}} h\, d\left(\mu_n \circ S^{-1}\right) = \lim_{n\to\infty} \int_{\mathbb{T}} h \circ S \; d\mu_n \overset{(1.8)}{=} \int_{\mathbb{T}} h \circ S \; d\mu = \int_{\mathbb{T}} h\, d\left(\mu \circ S^{-1}\right) \;,$$

where the second equality is due to the fact that $h \circ S \in C(\mathbb{T})$ by Remark 1.2.1.

■

Theorem 1.6.5. *Let $S\colon \mathbb{T} \to \mathbb{T}$ be a an isometry. Then the map $\mu \mapsto \mu \circ S^{-1}$ is also an isometry.*

Proof. By definition of isometry and by Corollary 1.2.7, $S, S^{-1} \in \mathrm{Lip}_1(\mathbb{T}; \mathbb{T})$. We want to show that $d_{\mathbb{T}}(\mu, \nu) = d_{\mathbb{T}}\left(\mu \circ S^{-1}, \nu \circ S^{-1}\right)$ for every $\mu, \nu \in \mathcal{P}$. Let arbitrary $\mu, \nu \in \mathcal{P}$ be given. On the one hand, we know by Remark 1.2.2 that $h \circ S^{-1} \in \mathrm{Lip}_1(\mathbb{T})$ for every $h \in \mathrm{Lip}_1(\mathbb{T})$. Thus considering the supremum in the definition of $d_{\mathbb{T}}\left(\mu \circ S^{-1}, \nu \circ S^{-1}\right)$, we have for every $h \in \mathrm{Lip}_1(\mathbb{T})$,

$$\begin{aligned} d_{\mathbb{T}}\left(\mu \circ S^{-1}, \nu \circ S^{-1}\right) &\geq \int_{\mathbb{T}} h \circ S^{-1} \; d\left(\mu \circ S^{-1}\right) - \int_{\mathbb{T}} h \circ S^{-1} \; d\left(\nu \circ S^{-1}\right) \\ &= \int_{\mathbb{T}} h\, d\mu - \int_{\mathbb{T}} h\, d\nu \;. \end{aligned}$$

Thus by taking the supremum over $h \in \mathrm{Lip}_1(\mathbb{T})$,

$$d_{\mathbb{T}}\left(\mu \circ S^{-1}, \nu \circ S^{-1}\right) \geq d_{\mathbb{T}}(\mu, \nu) \quad .$$

On the other hand, Remark 1.2.2 also tells us that $h \circ S \in \mathrm{Lip}_1(\mathbb{T})$. Applying the same argument to the supremum in the definiton of $d_{\mathbb{T}}(\mu, \nu)$, we get

$$d_{\mathbb{T}}(\mu, \nu) \geq d_{\mathbb{T}}\left(\mu \circ S^{-1}, \nu \circ S^{-1}\right) \quad .$$

Therefore $d_{\mathbb{T}}(\mu, \nu) = d_{\mathbb{T}}\left(\mu \circ S^{-1}, \nu \circ S^{-1}\right)$ for every $\mu, \nu \in \mathcal{P}$. ■

Thus, in particular, we have proved the rotation invariance of $d_{\mathbb{T}}$ mentioned in (1.6). In light of the fact that the graph of $F_{\mu \circ R_t^{-1}}\Big|_{[0,1)}$ is the graph of F_μ viewed through a shifted 1×1 frame, rotation invariance of $d_{\mathbb{T}}$ becomes even more intuitive once we prove in Chapter 2 that the $d_{\mathbb{T}}$ distance between two measures is just the minimal $L^1[0,1)$ distance between vertically shifted versions of their distribution functions. This is because Definition 1.3.4 implies that the difference of two distribution functions is 1-periodic, and thus the integral of this difference remains constant regardless of how much our integration frame is moved.

Chapter 2

A Formula for $d_{\mathbb{T}}(\mu, \nu)$

The goal of this chapter is to give a formula for the Kantorovich distance $d_{\mathbb{T}}$. By the end of this chapter, we will have proved in Theorem 2.2.8 that the Kantorovich distance between two distributions μ and ν is the minimal $L^1\left([0,1)\,,\,\mathcal{B}_{[0,1)}\,,\,\lambda\right)$ distance between vertically shifted distribution functions F_μ and F_ν , i.e.,

$$d_{\mathbb{T}}(\mu,\nu) \;=\; \min_{t\in\mathbb{R}} \left\|F_\mu - F_\nu - t\right\|_{L^1\left([0,1)\,,\,\mathcal{B}_{[0,1)}\,,\,\lambda\right)} \quad .$$

This formula is more convenient to use in light of Theorem 3.2.13 which provides a characterization of a $t_{min} \in \mathbb{R}$ that minimizes such $L^1[0,1)$ distance. Thus, Chapters 2 and 3 together provide a practical way to compute the Kantorovich distance between any two $\mu, \nu \in \mathcal{P}$.

Preparatory lemmas and results are needed for the proof of Theorem 2.2.8. Section 2.1 includes the preparatory work, and Section 2.2 includes Theorem 2.2.8 itself.

2.1 Preparatory work

In this preparatory section, we first recall that the supremum of a real-valued function on a compact domain is attained.

Remark 2.1.1. Let (X, d_X) be a nonempty compact metric space, and let $f\colon X \to \mathbb{R}$ be a continuous function. Then f attains its infimum and supremum; i.e.,

$$\exists\, y, z \in X \,:\, f(y) = \inf_{x \in X} f(x) \quad \wedge \quad f(z) = \sup_{x \in X} f(x) \quad .$$

Next, we introduce a subset of $\mathrm{Lip}_1(\mathbb{T})$ as follows:

$$\mathrm{Lip}_{1,0}(\mathbb{T}) := \Big\{ h \in C(\mathbb{T}; \mathbb{R}) : h(0) = 0 \;\wedge\; \big|h(x) - h(y)\big| \leq d_{\mathbb{T}}(x, y) \quad \forall x, y \in \mathbb{T} \Big\} \;,$$

where $h(0)$ is understood to mean the value of h at $0 + \mathbb{Z}$. The graphs of $\mathrm{Lip}_{1,0}(\mathbb{T})$ functions pass through the origin if we interpret $\mathbb{T}$ as $[0, 1)$.

Remark 2.1.2. Since $h \in \mathrm{Lip}_1(\mathbb{T})$ if and only if $h - h(0) \in \mathrm{Lip}_{1,0}(\mathbb{T})$, Definition 1.5.1 remains unchanged if $\mathrm{Lip}_1(\mathbb{T})$ is replaced by $\mathrm{Lip}_{1,0}(\mathbb{T})$.

With Remark 2.1.1 in mind, we will now prove that $\mathrm{Lip}_{1,0}(\mathbb{T})$ is compact. For this purpose, we first need to remind ourselves of the Arzela-Ascoli theorem.

Remark 2.1.3 (Arzela-Ascoli)**.** [13] Let (X, d_X) be a compact metric space. A set $F \subseteq C(X)$ is relatively compact in $\big(C(X)\,,\, \|.\|_\infty\big)$, i.e., $\overline{F}$ is compact in $\big(C(X)\,,\, \|.\|_\infty\big)$, if and only if F has the following two properties:

(i) F is (pointwise) bounded, i.e., the set $\big\{f(x) : f \in F\big\}$ is bounded for every $x \in X$.

(ii) F is equicontinuous, i.e., for every $\epsilon > 0$ there exists $\delta > 0$ such that $d_X(x, y) < \delta$ implies $\big|f(x) - f(y)\big| < \epsilon$ for every $f \in F$.

Lemma 2.1.4. *The metric space* $\Big(\mathrm{Lip}_{1,0}(\mathbb{T})\,,\, \|.\|_\infty\Big)$ *is compact.*

Proof. We will first use Arzela-Ascoli to prove that $\mathrm{Lip}_{1,0}(\mathbb{T})$ is a relatively compact subset of $\big(C(\mathbb{T})\,,\, \|.\|_\infty\big)$. We will then prove that $\mathrm{Lip}_{1,0}(\mathbb{T})$ is closed. To show the boundedness of $\Big\{h(x) \in \mathbb{R} : h \in \mathrm{Lip}_{1,0}(\mathbb{T})\Big\}$ for every $x \in \mathbb{T}$, let an arbitrary $x \in \mathbb{T}$ be given. Clearly, for every $h \in \mathrm{Lip}_{1,0}(\mathbb{T})$,

$$\big|h(x)\big| = \big|h(0) - h(x)\big| \overset{\text{Lipschitz}}{\leq} d_{\mathbb{T}}(0, x) \leq \frac{1}{2} \quad .$$

Thus $\text{Lip}_{1,0}(\mathbb{T})$ is pointwise bounded. To show the equicontinuiuty of $\text{Lip}_{1,0}(\mathbb{T})$ we need to show

$$\forall \epsilon > 0,\ \exists\, \delta > 0\colon \forall x, y \in \mathbb{T},\quad d_{\mathbb{T}}(x,y) < \delta \implies \big|h(x) - h(y)\big| < \epsilon \ \ \forall h \in \text{Lip}_{1,0}(\mathbb{T})\,.$$

Let arbitrary $\epsilon > 0$ be given, and let $\delta := \frac{\epsilon}{2}$. Clearly for every $x, y \in \mathbb{T}$ such that $d_{\mathbb{T}}(x,y) < \delta$,

$$\big|h_n(x) - h_n(y)\big| \overset{\text{Lipschitz}}{\leq} d_{\mathbb{T}}(x,y) < \delta = \frac{\epsilon}{2} < \epsilon \quad \forall h \in \text{Lip}_{1,0}(\mathbb{T}) \quad .$$

Thus $\text{Lip}_{1,0}(\mathbb{T})$ is also equicontinuous. Therefore by Remark 2.1.3, $\text{Lip}_{1,0}(\mathbb{T})$ is relatively compact. To show that $\text{Lip}_{1,0}(\mathbb{T})$ is closed, let an arbitrary Cauchy sequence $(h_n)_{n=1}^{\infty}$ in $\text{Lip}_{1,0}(\mathbb{T})$ be given. We know

$$\text{Lip}_{1,0}(\mathbb{T}) \subseteq C\,(\mathbb{T}) \xRightarrow{\left(C(\mathbb{T}),\|\cdot\|_{\infty}\right) \text{ is complete}} \exists\, h \in C(\mathbb{T}) : \|h_n - h\|_{\infty} \xrightarrow{n\to\infty} 0 \quad .$$

We want to show that $h \in \text{Lip}_{1,0}(\mathbb{T})$. Since sup-norm convergence implies pointwise convergence, we know $h(0) = 0$. To show $h \in \text{Lip}_1(\mathbb{T})$, let arbitrary $\epsilon > 0$ and $x, y \in \mathbb{T}$ be given. We have for sufficiently large $n \in \mathbb{N}$,

$$\begin{aligned} \big|h(x) - h(y)\big| \;&\leq\; \big|h(x) - h_n(x)\big| + \big|h_n(x) - h_n(y)\big| + \big|h_n(y) - h(y)\big| \\ &<\; \quad\epsilon \quad + \quad d_{\mathbb{T}}(x,y) \quad + \quad \epsilon \quad , \end{aligned}$$

where the first inequality is the triangle inequality, and the second is by the fact that $h_n \in \text{Lip}_1$. Since $\epsilon > 0$ was arbitrary, $\big|h(x) - h(y)\big| \leq d_{\mathbb{T}}(x,y)$ and h is 1-Lipschitz and thereby in $\text{Lip}_{1,0}(\mathbb{T})$. Thus $\overline{\text{Lip}_{1,0}(\mathbb{T})} = \text{Lip}_{1,0}(\mathbb{T})$. Therefore we have proved that $\text{Lip}_{1,0}(\mathbb{T})$ is compact. ■

Next, note that for every $\mu \in \mathcal{P}$, the map $h \mapsto \int_{\mathbb{T}} h\, d\mu$ defines a bounded linear functional on $C\,(\mathbb{T})$ because

$$\left|\int_{\mathbb{T}} h\, d\mu\right| \leq \int_{\mathbb{T}} |h|\; d\mu \leq \|h\|_{\infty} \qquad \forall h \in C(\mathbb{T})\,.$$

We therefore have the following corollary.

Corollary 2.1.5. *The supremum in Definition 1.5.1 is attained with* $\mathrm{Lip}_1(\mathbb{T})$ *replaced by* $\mathrm{Lip}_{1,0}(\mathbb{T})$ *. In other words,*

$$\forall \mu, \nu \in \mathcal{P}, \quad \exists\, h_0 \in \mathrm{Lip}_{1,0}(\mathbb{T}) : \ d_{\mathbb{T}}(\mu, \nu) = \int_{\mathbb{T}} h_0 \, d\mu \ - \int_{\mathbb{T}} h_0 \, d\nu \quad .$$

Proof. Immediate by continuity of $h \mapsto \int_{\mathbb{T}} h \, d\mu$, Remark 2.1.1, and Lemma 2.1.4. ∎

We close the section with a few lemmas about Lipschitz functions that are going to be used in Section 2.2.

Remark 2.1.6. Every Lipschitz continuous function $f : \mathbb{R} \to \mathbb{R}$ is absolutely continuous, and hence differentiable almost everywhere.

Lemma 2.1.7. *Let* $h \in \mathrm{Lip}_1(\mathbb{T})$ *, and let* h' *be as defined in Definition 1.3.8. Then,*

$$\left\| h' \right\|_{\infty} \leq 1 \quad .$$

Proof. Recall from Remark 1.3.7 that $\iota_{\mathbb{R}}^{-1} \in \mathrm{Lip}_1\left([0,1); \mathbb{T}\right)$. By Remark 1.2.2, $\left(h \circ \iota_{\mathbb{R}}^{-1}\right)$ is 1-Lipschitz as well, i.e.,

$$\forall s_1, s_2 \in [0,1), \quad \left| \left(h \circ \iota_{\mathbb{R}}^{-1}\right)(s_1) - \left(h \circ \iota_{\mathbb{R}}^{-1}\right)(s_2) \right| \leq |s_1 - s_2| \quad . \tag{2.1}$$

Through Remark 2.1.6, for almost every $x_0 \in \mathbb{T}$, we have by Definition 1.3.8,

$$\left| h'(x_0) \right| \ = \ \lim_{t \to s_0} \frac{\left| h \circ \iota_{\mathbb{R}}^{-1} \circ \langle \cdot \rangle (t) \ - \ h \circ \iota_{\mathbb{R}}^{-1} \circ \langle \cdot \rangle (s_0) \right|}{|t - s_0|} \quad .$$

Note that $\forall x_0 \in \mathbb{T} \setminus \{0 + \mathbb{Z}\}$, the function $\langle \cdot \rangle$ acts as the identity for any point in close neighborhoods of s_0. In that case,

$$\left| h'(x_0) \right| = \lim_{s \to s_0} \frac{\left| \left(h \circ \iota_{\mathbb{R}}^{-1}\right)(s) - \left(h \circ \iota_{\mathbb{R}}^{-1}\right)(s_0) \right|}{|s - s_0|} \overset{(2.1)}{\leq} \lim_{s \to s_0} 1 = 1 \quad .$$

And in the case where $x_0 = 0 + \mathbb{Z}$, i.e., where $s_0 = 0$, we know that in the right limit $t \to 0^+$ subcase, the function $\langle \cdot \rangle$ acts as the identity again, and the argument is the same as above. In the subcase where $t \to 0^-$,

$$\left|h'(x_0)\right| = \lim_{s\to 1} \frac{\left|\left(h\circ \iota_{\mathbb{R}}^{-1}\right)(s) - \left(h\circ \iota_{\mathbb{R}}^{-1}\right)(s_0)\right|}{\left|s-s_0\right|} \overset{(2.1)}{\leq} \lim_{s\to 1} 1 = 1 \quad .$$

Thus $\|h'\|_\infty \leq 1$. ■

Remark 2.1.8 (reverse triangle inequality)**.** Let (X, d_X) be a metric space. Then,

$$\forall a,b,c \in X, \quad \left|d_X(a,b) - d_X(a,c)\right| \leq d_X(c,b) \quad .$$

2.2 The Kantorovich Formula

As mentioned in the chapter opening, in this section, we prove that the $d_{\mathbb{T}}$ distance between two measures μ and ν is the minimal $L^1\left([0,1)\,,\, \mathcal{B}_{[0,1)}\,,\, \lambda\right)$ distance between vertically shifted distribution functions F_μ and F_ν. Some of the tools we use in this task include the bijection $\iota_{\mathbb{R}}$ introduced in Definition 1.3.6, the Radon-Nikodym theorem, and integration by parts. Below are applicable versions of the latter two statements.

Remark 2.2.1 (Radon-Nikodym Theorem)**.** [24] Let $(X, \mathcal{M})$ be a measurable space, and let ν and μ be probability measures. Then, $\mu \ll \nu$ if and only if there exists a measurable function $f_0 \colon X \to \overline{\mathbb{R}}$ such that it satisfies the following two conditions:

(i) $\int_X f_0 \, d\nu$ exists (in $\mathbb{R}$).

(ii) $\forall f \in L^1(X, \mathcal{M}, \mu)\,, \quad \int_X f \, d\mu = \int_X f \, f_0 \, d\nu \quad .$

The function f_0 is the *Radon-Nikodym derivative* of μ w.r.t. ν, denoted $\frac{d\mu}{d\nu}$.

Remark 2.2.2 (Integration by parts)**.** [20] Let $[a,b] \subset \mathbb{R}$ be an interval, and let $u, v \colon [a,b] \to \mathbb{R}$ be real-valued functions. If u is absolutely continuous and v is integrable w.r.t. the Lebesgue measure λ on $\mathcal{B}_{[a,b]}$, then

$$\int_{[a,b]} u\, v \, d\lambda \;=\; u(b)\, V(b) \;-\; \int_{[a,b]} u'\, V \, d\lambda \quad ,$$

where $u' \overset{\lambda\text{-a.e.}}{=\!=\!=} \frac{du}{dx}$ and $V(x) = \int_{[a,x]} v \, d\lambda$ for every $x \in [a, b]$.

Recall from Remark 2.1.6 that Lipschitz continuity implies absolute continuity. The fact that every $\mathrm{Lip}_{1,0}(\mathbb{T})$ function is absolutely continuous implies Lemma 2.2.3 through the integration by parts formula. This lemma will be a piece of the proof of the Kantorovich formula.

Lemma 2.2.3. *Let $\mu \in \mathcal{P}$ be absolutely continuous w.r.t. $\lambda_{\mathbb{T}}$. Then,*

$$\forall h \in \mathrm{Lip}_{1,0}(\mathbb{T}) \,, \qquad \int_{\mathbb{T}} h \, d\mu \; = - \int_{[0,1)} h' \, F_\mu \, d\lambda \quad .$$

Proof. Let an arbitrary $\mu \in \mathcal{P}$ with $\mu \ll \lambda_{\mathbb{T}}$ be given. By the Radon-Nikodym theorem we know

$$\exists \, f_\mu \in L^1 (\mathbb{T} \,, \mathcal{B}_{\mathbb{T}} \,, \lambda_{\mathbb{T}}) : \quad \forall f \in L^1 (\mathbb{T} \,, \mathcal{B}_{\mathbb{T}} \,, \mu) \,, \; \int_{\mathbb{T}} f \, d\mu \; = \int_{\mathbb{T}} f \, f_\mu \, d\lambda_{\mathbb{T}} \; . \tag{2.2}$$

Let an arbitrary $h \in \mathrm{Lip}_{1,0}(\mathbb{T})$ be given. Since $\mathbb{T}$ is compact, we know $\mathrm{Lip}_1(\mathbb{T}) \subseteq L^1 (\mathbb{T} \,, \mathcal{B}_{\mathbb{T}} \,, \lambda_{\mathbb{T}})$, and so (2.2) applies to h:

$$\begin{aligned} \int_{\mathbb{T}} h \, d\mu \overset{(2.2)}{=} \int_{\mathbb{T}} h \, f_\mu \, d\lambda_{\mathbb{T}} &= \int_{\iota_{\mathbb{R}}(\mathbb{T})} \left(h \circ \iota_{\mathbb{R}}^{-1}\right) \left(f_\mu \circ \iota_{\mathbb{R}}^{-1}\right) \, d\left(\lambda_{\mathbb{T}} \circ \iota_{\mathbb{R}}^{-1}\right) \\ &= \int_{[0,1)} \left(h \circ \iota_{\mathbb{R}}^{-1}\right) \left(f_\mu \circ \iota_{\mathbb{R}}^{-1}\right) \, d\lambda \quad , \end{aligned} \tag{2.3}$$

where in the second and third equalities $\iota_{\mathbb{R}} \colon \mathbb{T} \to [0,1)$ is as in Definition 1.3.6. We would like to apply integration by parts to (2.3). In order to check the conditions, we note that by Remark 1.2.2, $h \circ \iota_{\mathbb{R}}^{-1} \colon [0,1) \to \mathbb{R}$ is 1-Lipschitz, and hence absolutely continuous through Remark 2.1.6. We also note that

$$\int_{[0,1)} f_\mu \circ \iota_{\mathbb{R}}^{-1} \, d\lambda = \int_{\mathbb{T}} f_\mu \, d\lambda_{\mathbb{T}} \overset{(2.2)}{=} \int_{\mathbb{T}} d\mu \; = 1 \; < \infty \quad .$$

Thus $f_\mu \circ \iota_{\mathbb{R}}^{-1}$ is integrable w.r.t. λ, and we can integrate (2.3) by parts. Recalling from Remark 1.4.1 that $f_\mu \circ \iota_{\mathbb{R}}^{-1}$ is the density of the push-forward measure $\mu \circ \iota_{\mathbb{R}}^{-1}$, and from Definition 1.3.8 that $\left(h \circ \iota_{\mathbb{R}}^{-1}\right)' =: h'$, integration of

(2.3) by parts yields

$$\int_{\mathbb{T}} h\, d\mu = \Big[\cancelto{0}{\left(h \circ \iota_{\mathbb{R}}^{-1}\right)(s)\, F_\mu(s)}\Big]_0^{1^-} - \int_{[0,1)} h'\, F_\mu \, d\lambda \quad ,$$

where the first term vanishes since $h \in \mathrm{Lip}_{1,0}(\mathbb{T})$, hence $h(0) = h(1^-) = 0$. ■

Remark 2.2.4. Through analogous steps to the above proof, one can show that for any $x \in \mathbb{T}$ and absolutely continuous $\mu \in \mathcal{P}$,

$$\forall h \in \mathrm{Lip}_{1,0}(\mathbb{T}) , \qquad \int_{\mathbb{T}} h\, d(\mu \circ R_x^{-1}) = h(x) - \int_{[0,1)} h' \circ R_x \, F_\mu \, d\lambda \quad .$$

As another puzzle piece to be used in the proof of the Kantorovich formula we introduce the following:

Let M_0 denote the set of all essentially bounded measurable functions on $[0,1]$ that have sup-norm ≤ 1, and integrate to zero, i.e.,

$$M_0 := \left\{ g \in L^\infty[0,1] : \|g\|_\infty \leq 1 \quad \wedge \quad \int_{[0,1]} g\, d\lambda = 0 \right\} \quad . \tag{2.4}$$

The functions in this set are exactly the derivatives of the functions in $\mathrm{Lip}_{1,0}(\mathbb{T})$, and in that way M_0 is in one-to-one correspondence with $\mathrm{Lip}_{1,0}(\mathbb{T})$.

Lemma 2.2.5. *With M_0 as in (2.4), the function $\psi \colon \mathrm{Lip}_{1,0}(\mathbb{T}) \to M_0$ given by $\psi(h) := h'$ is a bijection.*

Proof. To show that ψ is surjective, note that for every $h \in \mathrm{Lip}_{1,0}(\mathbb{T})$, h is almost everywhere differentiable and by Lemma 2.1.7, $\|h'\|_\infty \leq 1$. Also, $\int_{[0,1]} h'\, d\lambda = 0$, since by Definition 1.3.8,

$$\int_{[0,1]} h'\, d\lambda = \Big[\left(h \circ \iota_{\mathbb{R}}^{-1}\right)(s)\Big]_0^{1^-} = 0 - 0 = 0 \quad .$$

Thus $\mathrm{Range}(\psi) \subseteq M_0$. To show the reverse inclusion, let an arbitrary $g \in M_0$ be given. Note that g is integrable on $[0,1]$. Consider the anti-derivative of

g denoted $h(s) := \int_{[0,s]} g\, d\lambda$. Clearly $h(0) = 0$. We want to show that h is 1-Lipschitz as well.

Let arbitrary $s_1, s_2 \in [0,1]$ be given. WLOG assume $s_1 < s_2$.

Case 1 $|s_1 - s_2| \leq \frac{1}{2}$.

Note that in this case, $d_{\mathbb{T}}(s_1, s_2) = s_2 - s_1$. By definition of h,

$$\begin{aligned} \left|h(s_2) - h(s_1)\right| &= \left|\int_0^{s_2} g(t)\, dt - \int_0^{s_1} g(t)\, dt\right| \\ &= \left|\int_{s_1}^{s_2} g(t)\, dt\right| \leq \int_{s_1}^{s_2} \left|g(t)\right|\, dt \\ &\leq \|g\|_\infty \int_{s_1}^{s_2} 1\, dt \leq 1\,(s_2 - s_1) = d_{\mathbb{T}}(s_1, s_2) \quad . \end{aligned}$$

Case 2 $|s_1 - s_2| > \frac{1}{2}$.

Note that in this case, $d_{\mathbb{T}}(s_1, s_2) = 1 - s_2 + s_1$. Since $\int_{[0,1]} g\, d\lambda = 0$, we know

$$\int_0^{s_2} g(t)\, dt = -\int_{s_2}^1 g(t)\, dt \quad . \tag{2.5}$$

By definition of h,

$$\begin{aligned} \left|h(s_2) - h(s_1)\right| &= \left|\int_0^{s_2} g(t)\, dt - \int_0^{s_1} g(t)\, dt\right| \\ &\overset{(2.5)}{=} \left|-\int_{s_2}^1 g(t)\, dt - \int_0^{s_1} g(t)\, dt\right| \\ &\leq \int_{s_2}^1 \left|g(t)\right|\, dt + \int_0^{s_1} \left|g(t)\right|\, dt \\ &\leq \|g\|_\infty \int_{s_2}^1 1\, dt + \|g\|_\infty \int_0^{s_1} 1\, dt \\ &\leq 1\,(1 - s_2) + 1\,(s_1) = d_{\mathbb{T}}(s_1, s_2) \quad . \end{aligned}$$

Thus in both cases $h \in \mathrm{Lip}_{1,0}(\mathbb{T})$, and ψ is surjective. To show its injectivity, let $h_1, h_2 \in \mathrm{Lip}_{1,0}(\mathbb{T})$ be such that $\psi(h_1) = \psi(h_2)$. Thus $h_1' = h_2'$ almost everywhere. For every $s \in [0,1)$,

$$0 = \int_{[0,s]} \left(h_1' - h_2'\right) \, d\lambda \;=\; \int_{[0,s]} h_1' \, d\lambda - \int_{[0,s]} h_2' \, d\lambda \;=\; h_1(s) - h_2(s) \quad ,$$

where the second and third equalities are by Remark 2.1.6, the fundamental theorem of calculus, and the fact that $h_1(0) = h_2(0)$. Therefore we have shown that $h_1 = h_2$, and thus ψ is injective too. ■

As the final puzzle piece, we will prove Theorem 2.2.7; but to do so, we need to remind ourselves of the Hahn-Banach theorem.

Remark 2.2.6 (Hahn-Banach)**.** Let $\left(X, \|\cdot\|_X\right)$ be a normed linear space, and $A \subseteq X$ a (not necessarily closed) linear subspace. Then,

$$\forall \varphi \in A^*, \quad \exists\, \widetilde{\varphi} \in X^* : \widetilde{\varphi}\Big|_A = \varphi \quad \wedge \quad \|\widetilde{\varphi}\|_{X^*} = \|\varphi\|_{A^*} \quad ,$$

where A^* and X^* denote dual spaces.

Theorem 2.2.7. *Let $\left(X\,,\, \|\cdot\|\right)$ be a normed vector space over $\mathbb{R}$, and let $x, y \in X$. Then,*

$$\sup\left\{\varphi(x) \in \mathbb{R} \,:\, \varphi \in B_0\right\} \;=\; \min_{\alpha \in \mathbb{R}} \|x - \alpha y\| \quad , \tag{2.6}$$

where $B_0 := \left\{\varphi \in X^* \,:\, \|\varphi\| \leq 1 \quad \wedge \quad \varphi(y) = 0\right\}$ is the set of bounded linear functionals in the closed unit ball centered at $\mathbf{0} \in X^*$, that contain y in their kernels.

Proof. Let arbitrary $x, y \in X$ be given.

Case 1 $\quad x \in \operatorname{span}\{y\}$.

By definition of span, there exists $\alpha_0 \in \mathbb{R}$ such that $x = \alpha_0\, y$. By linearity of φ, clearly $\varphi(x) = 0$ for all $\varphi \in B_0$, and therefore $\sup\left\{\varphi(x) \in \mathbb{R} \,:\, \varphi \in B_0\right\} = 0$. On the other hand, by non-negativity of the norm, $\min_{\alpha \in \mathbb{R}} \|\alpha_0\, y - \alpha y\| = 0$. Thus (2.6) holds in this case.

Case 2 $\quad x \notin \operatorname{span}\{y\}$.

In this case, the distance $\inf_{\alpha \in \mathbb{R}} \|x - \alpha y\|$ of x to $\operatorname{span}\{y\}$ is positive. We will first show that this infimum is attained. Since the attainment is obvious when $y = 0$, assume $y \neq 0$. Note that $\alpha \mapsto \|x - \alpha y\|$ is Lipschitz continuous because for every $\alpha_1, \alpha_2 \in \mathbb{R}$,

$$\big|\, \|x - \alpha_1 y\| - \|x - \alpha_2 y\| \,\big| \;\le\; \|-\alpha_1 y + \alpha_2 y\| \;=\; |\alpha_2 - \alpha_1|\, \|y\| \quad ,$$

where the inequality is by the reverse triangle inequality (see Remark 2.1.8). Note that the reverse triangle inequality also gives us a lower bound

$$\|\alpha y\| - \|x\| \;\le\; \|x - \alpha y\| \qquad \forall \alpha \in \mathbb{R} \quad ;$$

which implies that $\lim_{\alpha \to \pm\infty} \|x - \alpha y\| = +\infty$. Thus considering $\alpha \mapsto \|x - \alpha y\|$ on $[-N, N]$ for a large enough N, we see through Remark 2.1.1 that this map attains a global minimum. Assume the infimum is attained at α_0 and let b denote this minimal value. That is, $b := \|x - \alpha_0 y\| = \min_{\alpha \in \mathbb{R}} \|x - \alpha y\| > 0$. We want to show that $\sup_{\varphi \in B_0} \varphi(x) = b$. Note that $-\varphi \in B_0$ for every $\varphi \in B_0$. Thus $\sup_{\varphi \in B} \varphi(x) = \sup_{\varphi \in B} \big|\varphi(x)\big|$. By the linearity of φ, the definition of B, and the definition of operator norm,

$$\big|\varphi(x)\big| = \big|\varphi(x - \alpha_0 y)\big| \le \|\varphi\|\, \|x - \alpha_0 y\| \le \|x - \alpha_0 y\| = b \qquad \forall \varphi \in B \quad .$$

Therefore

$$\sup_{\varphi \in B} \big|\varphi(x)\big| \le b \quad . \tag{2.7}$$

To show the reverse inequality, consider the vector $z := x - \alpha_0 y$ which is clearly linearly independent of y, and thus the following linear functional ψ on $\operatorname{span}\{y, z\}$ is well-defined:

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

To show that ψ is bounded, note that if $t = 0$, then $\|sy + tz\| \ge |t|$ for all $s \in \mathbb{R}$. And in case where $s \in \mathbb{R}$ and $t \in \mathbb{R} \setminus \{0\}$, we again find

$$\|sy + tz\| = |t| \left\| \frac{s}{t} y + (x - \alpha_0 y) \right\| = |t| \left\| x - \big(\alpha_0 - \frac{s}{t}\big) y \right\| \ge |t|\; b \quad ,$$

where the inequality is true by definition of b. Therefore, for every $s, t \in \mathbb{R}$,

$$\left|\psi(sy+tz)\right| = |t| \le \frac{\|sy+tz\|}{b} \quad ,$$

which implies that $\|\psi\| \le \frac{1}{b}$. By the Hahn-Banach theorem, there exists $\widetilde{\psi} \in X^*$ such that $\|\widetilde{\psi}\| \le \frac{1}{b}$, and $\widetilde{\psi}(sy+tz) = t$ for every $sy+tz \in \operatorname{span}\{y,z\}$. Therefore $b\widetilde{\psi} \in B_0$ and $b\widetilde{\psi}(x) = b\,\psi(-\alpha_0 y + z) = b$. Thus by definition of supremum,

$$b \le \sup_{\varphi \in B} \left|\varphi(x)\right| \quad . \tag{2.8}$$

Through (2.7) and (2.8) we conclude that $\sup_{\varphi \in B} \varphi(x) = b$, and thus (2.6) holds in this case also. ■

Denote by $\mathcal{P}_0$ the space of probability measures on $\mathcal{B}_{\mathbb{T}}$ that have a finite support (i.e., $\mu(C) = 1$ for some finite set $C \subseteq \mathbb{T}$). Since $d_{\mathbb{T}}$ induces the weak topology on $\mathcal{P}$, we know that $\mathcal{P}_0$ is dense in $\mathcal{P}$ [21]. The space $\mathcal{P}_{AC}$ of probability measures that are absolutely continuous w.r.t. $\lambda_{\mathbb{T}}$ can approximate any $\mu \in \mathcal{P}_0$ arbitrarily well, and thus $\mathcal{P}_{AC}$ is dense in $\mathcal{P}$ too. We will use this fact in the proof of Theorem 2.2.8 below. Needless to say $\mathcal{P}_C := \{\mu \in \mathcal{P} : \mu \text{ continuous }\} \supseteq \mathcal{P}_{AC}$ is consequently dense as well.

Theorem 2.2.8. *For every* $\mu, \nu \in \mathcal{P}$*,*

$$d_{\mathbb{T}}(\mu,\nu) \;=\; \min_{t \in \mathbb{R}} \int_0^1 \left|F_\mu(s) - F_\nu(s) - t\right| ds \quad .$$

Proof. We will first show the above holds true for any $\mu, \nu \in \mathcal{P}_{AC}$. We then use the density of $\mathcal{P}_{AC}$ in $\mathcal{P}$ to show the result is true for any $\mu, \nu \in \mathcal{P}$. Let arbitrary $\mu, \nu \in \mathcal{P}_{AC}$ be given. By Corollary 2.1.5,

$$\exists\, h_0 \in \mathrm{Lip}_{1,0}(\mathbb{T}) : \;\; d_{\mathbb{T}}(\mu,\nu) = \int_{\mathbb{T}} h_0\, d\mu \;-\; \int_{\mathbb{T}} h_0\, d\nu \quad .$$

Since $\mu, \nu \ll \lambda_{\mathbb{T}}$, through Lemma 2.2.3 we know that

$$\begin{aligned} d_{\mathbb{T}}(\mu,\nu) &= \int_0^1 h_0'(s) \left(F_\nu(s) - F_\mu(s)\right) ds \\ &\le \sup_{g\in M_0} \int_0^1 g(s) \left(F_\nu(s) - F_\mu(s)\right) ds \quad , \end{aligned} \tag{2.9}$$

where M_0 is as defined in (2.4), and the inequality is because Lemma 2.2.5 implies $h_0' \in M_0$. On the other hand, Lemma 2.2.5 also implies that for every $h' \in M_0$, $\int_{\mathbb{T}} h\, d\mu - \int_{\mathbb{T}} h\, d\nu \le d_{\mathbb{T}}(\mu,\nu)$, and therefore through Lemma 2.2.3, we know for every $h' \in M_0$,

$$\int_0^1 h'(s) \left(F_\nu(s) - F_\mu(s)\right) ds \le d_{\mathbb{T}}(\mu,\nu) \quad .$$

Taking the supremum over all $h' \in M_0$ we have

$$\sup_{g\in M_0} \int_0^1 g(s) \left(F_\nu(s) - F_\mu(s)\right) ds \;\le\; d_{\mathbb{T}}(\mu,\nu) \quad . \tag{2.10}$$

Therefore through (2.9) and (2.10) we have

$$d_{\mathbb{T}}(\mu,\nu) \;=\; \sup_{g\in M_0} \int_0^1 g(s) \left(F_\nu(s) - F_\mu(s)\right) ds \quad . \tag{2.11}$$

Through Corollary 2.1.5 and Lemma 2.2.5 we know that the supremum in (2.11) is attained. Recall that the dual space of the Banach space $\left(L^1[0,1], \|\cdot\|_{L^1}\right)$ is simply $\left(L^\infty[0,1], \|\cdot\|_\infty\right)$ up to an isometric isomorphism. Thus we can identify M_0 in (2.4) with the set

$$N_0 := \left\{\varphi \in L^1[0,1]^* : \|\varphi\| \le 1 \quad \wedge \quad \varphi(\mathbb{1}_{[0,1]}) = 0\right\} \quad .$$

Consequently, (2.11) can be written as

$$d_{\mathbb{T}}(\mu,\nu) = \sup_{\varphi\in N_0} \varphi\left(F_\mu - F_\nu\right) \quad . \tag{2.12}$$

Note that $\mathbb{1}_{[0,1]} \in L^1[0,1]$, and letting $x = F_\mu - F_\nu$ in Theorem 2.2.7, we conclude that

$$d_{\mathbb{T}}(\mu,\nu) = \min_{t\in\mathbb{R}} \left\|F_\mu - F_\nu - t\right\|_{L^1} \quad .$$

Claim 2.2.8.1. (2.11) holds for all $\mu,\nu \in \mathcal{P}$.

Proof. Let arbitrary $\mu,\nu \in \mathcal{P}$ be given. By the density of $\mathcal{P}_{AC}$ in $\mathcal{P}$ there exist sequences $(\mu_n)_{n=1}^{\infty}$ and $(\nu_n)_{n=1}^{\infty}$ in $\mathcal{P}_{AC}$ such that $\lim_{n\to\infty} d_{\mathbb{T}}(\mu_n,\mu) = \lim_{n\to\infty} d_{\mathbb{T}}(\nu_n,\nu) = 0$. Note that the definition of weak convergence implies

$$\lim_{n\to\infty} \int_0^1 F_{\mu_n}\, ds = \int_0^1 F_\mu\, ds \quad \wedge \quad \lim_{n\to\infty} \int_0^1 F_{\nu_n}\, ds = \int_0^1 F_\nu\, ds \quad . \tag{2.13}$$

By what has already been proved, there exists a real sequence $(t_n)_{n=1}^{\infty}$ such that

$$d_{\mathbb{T}}(\mu_n,\nu_n) = \int_0^1 \left|F_{\mu_n}(s) - F_{\nu_n}(s) - t_n\right| ds \qquad \forall n \in \mathbb{N} \quad .$$

Since t_n minimizes $\left\|F_{\mu_n} - F_{\nu_n} - \cdot\right\|_{L^1}$ and $F_{\mu_n} - F_{\nu_n} \le 1$, clearly the sequence $(t_n)_{n=1}^{\infty}$ is bounded and thus has a convergent subsequence $\left(t_{n_j}\right)_{j=1}^{\infty}$. Therefore

$$\begin{aligned} \lim_{j\to\infty} d_{\mathbb{T}}(\mu_{n_j},\nu_{n_j}) &= \lim_{j\to\infty} \int_0^1 \left|F_{\mu_{n_j}}(s) - F_{\nu_{n_j}}(s) - t_{n_j}\right| ds \\ \Longrightarrow \qquad d_{\mathbb{T}}(\mu,\nu) &= \int_0^1 \left|F_\mu - F_\nu - t_0\right| \quad , \end{aligned} \tag{2.14}$$

where the second equality is by dominated convergence and (2.13). The symbol t_0 denotes the limit of $\left(t_{n_j}\right)_{j=1}^{\infty}$. We now show that the right hand side in (2.14) is $\min_{t\in\mathbb{R}} \int_0^1 \left|F_\mu - F_\nu - t\right|$. Assume, by contradiction, that

$$\int_0^1 \left|F_\mu - F_\nu - t_{min}\right| < \int_0^1 \left|F_\mu - F_\nu - t_0\right| \quad ,$$

for some $t_{min} \in \mathbb{R}$. By (2.14), we can choose $\epsilon > 0$ so small that

$$\epsilon < \frac{d_{\mathbb{T}}(\mu,\nu) - \int_0^1 \left|F_\mu - F_\nu - t_{min}\right|}{3} \quad . \tag{2.15}$$

Clearly by the continuity of $d_{\mathbb{T}}$ there exists $N_1 \in \mathbb{N}$ such that for every $j > N_1$,

$$d_{\mathbb{T}}(\mu_{n_j}, \nu_{n_j}) \;>\; d_{\mathbb{T}}(\mu,\nu) - \epsilon \quad .$$

By (2.14) there exists $N_2 \in \mathbb{N}$ such that for every $j > N_2$,

$$\int_0^1 \left| F_{\mu_{n_j}} - F_{\nu_{n_j}} - t_{min} \right| < \int_0^1 \left| F_{\mu} - F_{\nu} - t_{min} \right| + \epsilon \quad .$$

Let $N := \max\{N_1, N_2\}$. We have shown for every $j \geq N$,

$$\begin{aligned} d_{\mathbb{T}}(\mu_{n_j}, \nu_{n_j}) - \int_0^1 \left| F_{\mu_{n_j}} - F_{\nu_{n_j}} - t_{min} \right| \;&>\; d_{\mathbb{T}}(\mu,\nu) - \int_0^1 \left| F_{\mu} - F_{\nu} - t_{min} \right| - 2\epsilon \\ &\overset{(2.15)}{>} 3\epsilon - 2\epsilon = \epsilon \quad . \end{aligned}$$

Taking the limit as $n \to \infty$, this implies $0 > \epsilon$ which is a clear contradiction. Thus we have shown $\int_0^1 \left| F_\mu - F_\nu - t_0 \right| = \min_{t \in \mathbb{R}} \int_0^1 \left| F_\mu(s) - F_\nu(s) - t \right| ds$. □

Thereby $d_{\mathbb{T}}(\mu, \nu) = \min_{t \in \mathbb{R}} \int_0^1 \left| F_\mu(s) - F_\nu(s) - t \right| ds$ for any $\mu, \nu \in \mathcal{P}$. ■

Chapter 3

Minimizing the L^1 Distance

We learned in Chapter 2 that $d_{\mathbb{T}}(\mu, \nu) = \min_{t \in \mathbb{R}} \int_0^1 \left| F_\mu(s) - F_\nu(s) - t \right| ds$ for every $\mu, \nu \in \mathcal{P}$. The goal of this chapter is to show that the median value of $F_\mu - F_\nu$ is a minimizer t for the above integral (see Theorem 3.2.13). We will do so by more generally proving that for any L^1 function $g \colon [0,1) \to \mathbb{R}$ the L^1 distance between g and a constant function t is minimized when t is the median value of g. As with previous chapters, the preparatory lemmas required for the proof of the main result are presented in the first section.

3.1 Preparatory work

In Section 3.2 we will begin the journey toward proving that the median value of g minimizes $\int_0^1 \left| g(s) - t \right| \, ds$ by first proving it for monotone g. Also prominent in an essential theorem (Theorem 3.2.10) will be the cumulative distribution function of g which is a monotone function as well. For these reasons we prove a few useful facts about monotone functions in this preparatory section.

Remark 3.1.1. Let $f \colon [a,b] \to \mathbb{R}$ be monotone. Then for every $c \in (a,b)$, $f(c^-)$ and $f(c^+)$ both exist. Specifically, if f is increasing,

$$f(c^+) = \inf \left\{ f(x) : c < x < b \right\} \quad \wedge \quad f(c^-) = \sup \left\{ f(x) : a < x < c \right\}.$$

Furthermore, for an increasing f we have

$$f(c^-) \le f(c) \le f(c^+) \quad .$$

We will prove that monotone functions have only a countable number of jump discontinuities. We first set the stage to that end.

Definition 3.1.2 (jump of a monotone function)**.** Let $f\colon [0,1] \to \mathbb{R}$ be monotone. We define the *jump* at a point c to be

$$\mathrm{jmp}(c) := f(c^+) - f(c^-) \qquad \forall c \in (0,1) \quad .$$

We additionally define the jump at the end points as $\mathrm{jmp}(0) := f(0^+) - f(0)$, and $\mathrm{jmp}(1) := f(1) - f(1^-)$. Increasing functions have non-negative jumps.

Remark 3.1.3. Let $f\colon [0,1] \to \mathbb{R}$ be increasing, and let $\{x_j\}_{j=0}^{n}$ be a partition on $[0,1]$, i.e., $n \in \mathbb{N}$ and $0 = x_0 < x_1 < x_2 < ... < x_n = 1$. Then,

$$\sum_{j=0}^{n} \mathrm{jmp}(x_j) \le f(1) - f(0) \quad .$$

Lemma 3.1.4 (Monotonicity implies countable discontinuity)**.** *Let $f\colon [0,1] \to \mathbb{R}$ be monotone, and let S be the set of (necessarily jump) discontinuities of f, i.e., let $S := \{x \in [0,1] : \mathrm{jmp}(x) \ne 0\}$. Then S is countable.*

Proof. WLOG assume f is increasing. For every $m \in \mathbb{N}$, define the set $S_m := \{x \in [0,1] : \mathrm{jmp}(x) > \frac{1}{m}\}$. Clearly, $S_m \subseteq S$ for all $m \in \mathbb{N}$, and $(S_m)_{m=1}^{\infty} \nearrow S$, i.e., $S_{m_1} \subseteq S_{m_2}$ and $\bigcup_{m=1}^{\infty} S_m = S$ for every $m_1, m_2 \in \mathbb{N}$ with $m_1 < m_2$.

Claim 3.1.4.1. Every S_m defined above is a finite set.

Proof. Assume, by contradiction, that there exists some $m_0 \in \mathbb{N}$ for which S_{m_0} is infinite. Let n_0 be a natural number such that $(n_0 - 1)\left(\frac{1}{m_0}\right) > f(b) - f(a)$. Since $\#S_{m_0} > n_0$, we can choose $n_0 - 1$ points in S_{m_0}, which we name

$x_1,\, x_2,\, \cdots,\, x_{n_0-1}$ in increasing order. Let $x_0 := 0$ and $x_{n_0} := 1$. Note that $\{x_j\}_{j=0}^{n_0}$ is a partition on $[0,1]$. By definition of S_{m_0},

$$\sum_{j=1}^{n_0-1} \operatorname{jmp}(x_j) > \sum_{j-1}^{n_0-1} \frac{1}{m_0} = (n_0-1)\left(\frac{1}{m_0}\right) > f(b) - f(a) \quad , ※$$

which contradicts Remark 3.1.3. □

By Claim 3.1.4.1, every S_m is a countable set. Since countable union of countable sets is countable, we have proved that S is countable. ■

Lemma 3.1.5. *Let $f\colon [0,1] \to \mathbb{R}$ be a monotone function. Then for every $t_0 \in \operatorname{Range}(f)$, the set $\{f = t_0\}$ is a (possibly degenerate) interval.*

Proof. WLOG assume f is increasing. Let an arbitrary $t_0 \in \operatorname{Range}(f)$ be given. By definition of range, $\{f = t_0\} \neq \emptyset$. We want to show that for every $a, b \in \{f = t_0\}$ such that $a \leq b$, we have $[a,b] \subseteq \{f = t_0\}$.
Let arbitrary $a, b \in \{f = t_0\}$ where $a \leq b$ be given. For every $c \in [a,b]$,

$$t_0 = f(a) \leq f(c) \leq f(b) = t_0 \quad ,$$

since increasing functions preserve order. Thus we have shown that $f(c) = t_0$ or $c \in \{f = t_0\}$. ■

Recall that a monotone function is not necessarily invertible in the usual sense because it may not be one-to-one or onto due to constant regions or jumps, respectively. There is, however, a way to define ‘the inverse’ of a monotone function, and we will make use of this definition because we will need to invert distribution functions (which are increasing) in Section 3.2. In this preparatory section we state the general definition and results about the inverse of an increasing function $f\colon \overline{\mathbb{R}} \to \overline{\mathbb{R}}$, and later in Section 3.2 we use it for cumulative distribution functions $\widehat{g}\colon (0,1) \to \mathbb{R}$ of integrable functions g on $[0,1)$.

Definition 3.1.6 (Inverse of an increasing function)**.** Let $f\colon \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ be an increasing function. We define the *inverse* of f to be the function $f^{-1}\colon \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ given by:

$$\forall a \in \overline{\mathbb{R}}, \quad f^{-1}(a) := \sup\{f \le a\} \quad ,$$

where $\{f \le a\}$ is understood to mean $\{x \in \overline{\mathbb{R}} : f(x) \le a\}$. We follow the usual convention that $\sup \emptyset = -\infty$.

Remark 3.1.7. The f^{-1} in Definition 3.1.6 is increasing and right-continuous.

We close off this section by proving a result which we will use in the proof of Theorem 3.2.13.

Lemma 3.1.8. *Let $(\Omega, \mathcal{M}, \mu)$ be a σ-finite measure space, and let $\varphi\colon [0,+\infty) \to \mathbb{R}$ be a continuously differentiable, increasing function with $\varphi(0) = 0$. Then for every $\mathcal{M}$-$\mathcal{B}_{[0,+\infty)}$-measurable function $f\colon \Omega \to [0,+\infty)$,*

$$\int_\Omega \varphi\left(f(\omega)\right) d\mu(\omega) = \int_0^\infty \varphi'(t)\, \mu\left(\{f \ge t\}\right) dt \quad ,$$

Proof. By the Fundamental Theorem of Calculus,

$$\begin{aligned} \int_\Omega \varphi\left(f(\omega)\right) d\mu(\omega) &= \int_\Omega \left(\int_0^{f(\omega)} \varphi'(t)\, dt\right) d\mu(\omega) \\ &= \int_\Omega \left(\int_0^\infty \varphi'(t)\ \mathbb{1}_{[0,f(\omega)]}(t)\, dt\right) d\mu(\omega) \quad . \end{aligned} \tag{3.1}$$

Note that $(\Omega, \mathcal{M}, \mu)$ and $\left([0,+\infty), \mathcal{B}_{[0,+\infty)}, \lambda_{[0,+\infty)}\right)$ are σ-finite measure spaces. Also note that since φ' is continuous, it is $\mathcal{B}_\mathbb{R}$-$\mathcal{B}_{[0,+\infty)}$-measurable. Thus the map $(t,\omega) \mapsto \varphi'(t)\mathbb{1}_{[0,f(\omega)]}(t)$ is $\left(\mathcal{M} \otimes \mathcal{B}_{[0,+\infty)}\right)$ - $\mathcal{B}_\mathbb{R}$-measurable. We can therefore apply Tonelli's theorem to (3.1):

$$\begin{aligned} \int_\Omega \varphi\left(f(\omega)\right) d\mu(\omega) &= \int_0^\infty \left(\int_\Omega \varphi'(t)\ \mathbb{1}_{[0,f(\omega)]}(t)\, d\mu(\omega)\right) dt \\ &= \int_0^\infty \varphi'(t)\left(\int_\Omega \mathbb{1}_{[0,f(\omega)]}(t)\, d\mu(\omega)\right) dt \\ &= \int_0^\infty \varphi'(t)\ \mu\left(\left\{\omega \in \Omega : f(\omega) \ge t\right\}\right) dt \quad . \end{aligned}$$

■

Corollary 3.1.9. *Let $(\Omega, \mathcal{M}, \mathbb{P})$ be a probability space. Let $X\colon \Omega \to [0,+\infty)$ be an $\mathcal{M}$-$\mathcal{B}_{[0,+\infty)}$-measurable function ("non-negative random variable"). Then,*

$$\mathbb{E}[X] = \int_\Omega X(\omega)\, d\,\mathbb{P}(\omega) \; = \int_0^{+\infty} \mathbb{P}\,(X \geq t)\; dt \quad .$$

3.2 t_{min}

The main goal of this section is to prove Theorem 3.2.13. Throughout this section, we denote by t_{min} any $t \in \mathbb{R}$ that minimizes $\int_0^1 \left|g(s) - t\right|\, ds$ for a given real-valued function g on $[0,1)$. Note that if existent, t_{min} is not necessarily unique. We begin by proving the result for monotone g.

Theorem 3.2.1. *Let $g\colon [0,1] \to \mathbb{R}$ be a monotone function. The value of $\int_0^1 \left|g(s) - t\right|\, ds$ is minimized if and only if t is between the left limit and right limit of g at $\frac{1}{2}$, i.e.,*

$$\min_{t\in\mathbb{R}} \int_0^1 \left|g(s) - t\right| ds = \int_0^1 \left|g(s) - t_0\right|\, ds \quad \Longleftrightarrow \quad t_0 \in \left[\, g^m_{\frac{1}{2}},\, g^M_{\frac{1}{2}} \right] \quad ,$$

where $g^m_{\frac{1}{2}} := \min\left\{g(\tfrac{1}{2}^-), g(\tfrac{1}{2}^+)\right\}$, *and* $g^M_{\frac{1}{2}} := \max\left\{g(\tfrac{1}{2}^-), g(\tfrac{1}{2}^+)\right\}$.

Proof. We prove the theorem for the case where g is increasing. The decreasing case is completely analogous. Let $t_0 := g(\frac{1}{2})$. Note that by Remark 3.1.1,

$$g\Big(\,\frac{1}{2}^-\Big) \;\leq\; t_0 \;\leq\; g\Big(\,\frac{1}{2}^+\Big) \quad . \tag{3.2}$$

Let arbitrary $t \in \mathbb{R}$ be given. We distinguish three cases: $t < g\big(\,\frac{1}{2}^-\big)$, $t > g\big(\,\frac{1}{2}^+\big)$, and $t \in \Big[g(\frac{1}{2}^-)\,,\, g(\frac{1}{2}^+)\Big]$. We will show in Claims 3.2.1.1 and 3.2.1.2 that if $t \notin \Big[g(\frac{1}{2}^-)\,,\, g(\frac{1}{2}^+)\Big]$, then $\int_0^1 \left|g(s) - t\right|\, ds$ is strictly larger than $\int_0^1 \left|g(s) - t_0\right|\, dt$. We then show in Claim 3.2.1.3 that in the case where $t \in \Big[g(\frac{1}{2}^-)\,,\, g(\frac{1}{2}^+)\Big]$, the value $\int_0^1 \left|g(s) - t_0\right|\, ds$ is equal to $\int_0^1 \left|g(s) - t\right|\, ds$. The desired result will thereby follow.

Case 1 $t < g\left(\frac{1}{2}^{-}\right)$

Claim 3.2.1.1. The L^1 distance between g and t is strictly greater than that of g and t_0, i.e.,

$$\int_0^1 \left|g(s) - t_0\right| \, ds < \int_0^1 \left|g(s) - t\right| \, ds \quad .$$

Proof. We want to show $\int_0^1 \left|g(s) - t\right| \, ds - \int_0^1 \left|g(s) - t_0\right| \, ds > 0$. (See Figure 3.2 .) In order to do this, we need to know the sign of the integrands by properly partitioning the domain of the integrals. To that end, we first note that by assumption,

$$t < g\left(\frac{1}{2}^{-}\right) \quad \overset{(3.2)}{\Longrightarrow} \quad t < t_0 \quad . \tag{3.3}$$

Second, note that by Lemma 3.1.5, $g^{-1}\left(\{t_0\}\right)$ is an interval. Let $a < b \in [0, 1]$ denote the endpoints of this interval. Additionally, let $c := 0$ if $\{g < t\} = \emptyset$. Otherwise, denote $c := \sup\{g < t\}$. Lastly denote $d := \inf\{g > t\}$. And note that since g is increasing,

$$0 \leq c \leq d \leq a \leq \frac{1}{2} \leq b \leq 1 \quad , \tag{3.4}$$

as seen in Figure 3.1. The essence of the proof in this case comes down to the following fact: $d < \frac{1}{2}$ because the middle two inequalities in (3.4) cannot simultaneously be equalities. That is, if $a = \frac{1}{2}$ then $d < a$. We will soon demonstrate this fact by distinguishing the subcases $a = \frac{1}{2}$ and $a < \frac{1}{2}$. Before that, however, we break down the integrals into a partition:

$$\int_0^1 \left|g(s) - t\right| \, ds - \int_0^1 \left|g(s) - t_0\right| \, ds =$$

$$\int_{\{g\le t\}} \left|g(s)-t\right|\,ds+\int_{\{g>t\}} \left|g(s)-t\right|\,ds$$
$$-\int_{\{g<t_0\}} \left|g(s)-t_0\right|\,ds-\int_{\{g\ge t_0\}} \left|g(s)-t_0\right|\,ds$$

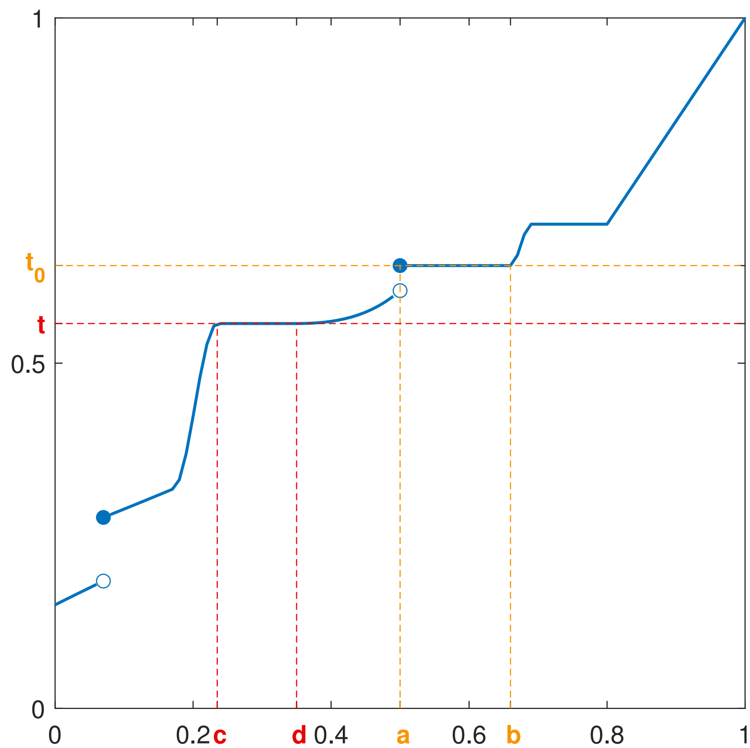

Figure 3.1: An example of a generic increasing g in the case where $t<g(1/2^-)$ is in $\mathrm{Range}(g)$. In this example, $a=\frac{1}{2}$.

$$=\int_0^d \big(-g(s)+t\big)\,ds+\int_d^1 \big(g(s)-t\big)\,ds$$
$$-\int_0^a \big(-g(s)+t_0\big)\,ds-\int_a^1 \big(g(s)-t_0\big)\,ds$$

$$=\int_0^d \big(-g(s)+t\big)\,ds+\int_d^a \big(g(s)-t\big)\,ds+\int_a^1 \big(g(s)-t\big)\,ds$$
$$-\int_0^d \big(-g(s)+t_0\big)\,ds-\int_d^a \big(-g(s)+t_0\big)\,ds-\int_a^1 \big(g(s)-t_0\big)\,ds$$

$$= \int_0^d (t - t_0)\ ds + \int_d^a \big(g(s) - t\big)\ ds + \int_a^1 (t_0 - t)\ ds - \int_d^a \big(-g(s) + t_0\big)\ ds\ . \tag{3.5}$$

Assume $a = \frac{1}{2}$. In this subcase, $d < a$ because otherwise we would have $\inf\{g > t\} = d = a = \frac{1}{2}$, which since g is increasing, implies that $g(s) \leq t$ for all $s < \frac{1}{2}$, which in turn implies the contradiction $g(\frac{1}{2}^-) \leq t$. Also note that the fact that g is increasing, together with the definition of $\inf\{g > t\}$, implies that

$$t < g(s) < t_0 \qquad \forall s \in (d, a)\ \ . \tag{3.6}$$

By (3.5) we know

$$\begin{aligned}
&\int_0^1 \big|g(s) - t\big|\ dt\ - \int_0^1 \big|g(s) - t_0\big|\ dt \\
&\quad = d\,(t - t_0) + (1 - a)(t_0 - t) + \int_d^a \big(g(s) - t\big)\ ds - \int_d^a \big(-g(s) + t_0\big)\ ds \\
&\quad > d\,(t - t_0) + (1 - a)(t_0 - t) - \int_d^a \big(-g(s) + t_0\big)\ ds \\
&\quad > d\,(t - t_0) + (1 - a)(t_0 - t) - \int_d^a (-t + t_0)\ ds\ =\ 0\ ,
\end{aligned}$$

where the strict inequalities are by (3.6) and the fact that $d < a$. Now assume that $a < \frac{1}{2}$. Again, by (3.5) we know

$$\begin{aligned}
&\int_0^1 \big|g(s) - t\big|\ dt\ - \int_0^1 \big|g(s) - t_0\big|\ dt \\
&\quad = d\,(t - t_0) + (1 - a)(t_0 - t) + \int_d^a \big(g(s) - t\big)\ ds - \int_d^a \big(-g(s) + t_0\big)\ ds \\
&\quad \geq d\,(t - t_0) + (1 - a)(t_0 - t) - \int_d^a \big(g(s) - t\big)\ ds - \int_d^a \big(-g(s) + t_0\big)\ ds \\
&\quad = d\,(t - t_0) + (1 - a)(t_0 - t) - (a - d)(t_0 - t) = (1 - 2a)(t_0 - t) > 0\ \ ,
\end{aligned}$$

where the inequalities are by (3.6) and the assumption that $a < \frac{1}{2}$.

□

Case 2 $t > g\big(\frac{1}{2}^+\big)$

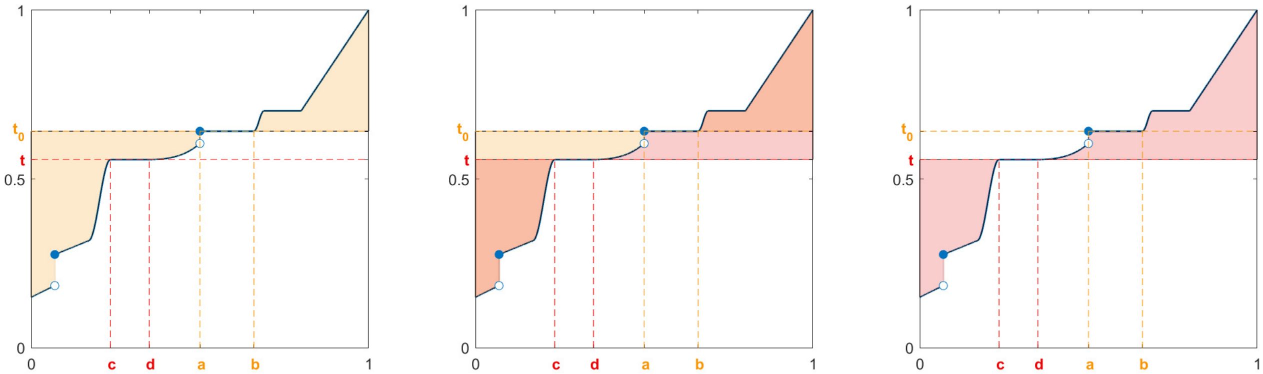

Figure 3.2: The area between g and t_0 shaded in light orange (left) and the area between the same g and $t < g(1/2^-)$ shaded in light red (right), and the superposition of the two pictures (middle). In Claim 3.2.1.1 we proved that the light orange area is strictly bigger than the light red area.

Claim 3.2.1.2. The L^1 distance between g and t is strictly greater than that of g and t_0, i.e.,

$$\int_0^1 \left|g(s) - t_0\right| \, ds < \int_0^1 \left|g(s) - t\right| \, ds \quad .$$

Proof. Analogous to the proof of Claim 3.2.1.1. □

$\boxed{\text{Case 3}}$ $t \in \left[g(\tfrac{1}{2}^-)\,,\, g(\tfrac{1}{2}^+)\right]$

Claim 3.2.1.3. The L^1 distance between g and t is equal to that of g and t_0, i.e.,

$$\int_0^1 \left|g(s) - t_0\right| \, ds = \int_0^1 \left|g(s) - t\right| \, ds \quad .$$

Proof. Breaking down our integral,

$$\int_0^1 \left|g(s) - t_0\right| \, ds - \int_0^1 \left|g(s) - t\right| \, ds =$$

$$\int_0^{\frac{1}{2}} \big(-g(s)+t_0\big)\ ds + \int_{\frac{1}{2}}^1 \big(g(s)-t_0\big)\ ds \\ - \int_0^{\frac{1}{2}} \big(-g(s)+t\big)\ ds - \int_{\frac{1}{2}}^1 \big(g(s)-t\big)\ ds$$

$$= \quad \int_0^{\frac{1}{2}} (t_0-t)\ ds + \int_{\frac{1}{2}}^1 (t-t_0)\ ds \ = \ \frac{1}{2}(t_0-t) - \frac{1}{2}(t_0-t) \ = 0 \ \ ,$$

where the sign of the integrand $g(s)-t$ in the first equality is by Remark 3.1.1.

□

Thus by Claim 3.2.1.2, Claim 3.2.1.1 and Claim 3.2.1.3, $t \mapsto \int_0^1 \big|g(s)-t\big|\ ds$ is minimized iff $t \in \Big[g(\tfrac{1}{2}^-)\ ,\ g(\tfrac{1}{2}^+)\Big]$. ■

Remark 3.2.2. As an immediate consequence of Theorem 3.2.1 we see that $\min_{t\in\mathbb{R}} \int_0^1 |s-t|\, ds \ = \ \frac{1}{4}$.

Informally speaking, Theorem 3.2.1 is saying that for a monotone g, the value of the integral $\int_0^1 \big|g(s)-t\big|\ ds$ is minimized when half the time g is above t, and half the time below. More precisely, it tells us that for every monotone function g,

$$\|g-t\|_1 \text{ is minimized} \iff \begin{cases} \lambda\big(\{g \le t\}\big) \ge \frac{1}{2} \ , \\ \qquad\quad \wedge \\ \lambda\big(\{g \ge t\}\big) \ge \frac{1}{2} \ , \end{cases}$$

which is to say that t_{min} is a median value of g. This fact is reminiscent of a well-known fact in statistics [9] : If X is a random variable and t is a constant estimator for X, then the t that minimizes the *mean absolute error* $\mathbb{E}\big[|X-t|\big]$ is the median value of X ; see Theorem 3.2.3. While not identical, this fact is closely related to Theorem 3.2.1. The exact relationship between these two theorems will become clear in Theorem 3.2.13; but for now, one such relation to know is that Theorem 3.2.1 is used to prove the statistics fact stated in Theorem 3.2.3.

Theorem 3.2.3. *Let $(\Omega, \mathcal{M}, \mathbb{P})$ be a probability space, and let $X\colon \Omega \to \mathbb{R}$ be an $\mathcal{M}$-$\mathcal{B}_{\mathbb{R}}$-measurable, absolutely continuous function (i.e., X has a density). The value of $\int_\Omega \big|X(\omega) - t\big|\, d\mathbb{P}(\omega)$ with $t \in \mathbb{R}$ is minimized iff $t \in \Big[F_X^{-1}(\tfrac{1}{2}^-)\,,\, F_X^{-1}(\tfrac{1}{2}^+)\Big]$, where $F_X(x) := \mathbb{P}\big(\{X \le x\}\big)$ for every $x \in \mathbb{R}$.*

Proof. By definition,

$$\begin{aligned}
\int_\Omega \big|X(\omega) - t\big|\, d\mathbb{P}(\omega) &= \int_{\mathbb{R}} |x - t|\; dP_X(x) \\
&= \int_{\mathbb{R}} |x - t|\, f_X(x)\, dx \\
&= \int_{F_X(\mathbb{R})} \big|F_X^{-1}(z) - t\big|\; dz \;\; = \;\; \int_0^1 \big|F_X^{-1}(z) - t\big|\; dz \;\; ,
\end{aligned}$$

where $P_X := \mathbb{P} \circ X^{-1}$, f_X is the density of X, and the third equality is by the substitution $z = F_X(x)$ which implies $dz = f_X(x)\, dx$ and $x = F_X^{-1}(z)$. Note that F_X^{-1} is understood in the sense of Definition 3.1.6. By Remark 3.1.7, F_X^{-1} is increasing. Thus by Theorem 3.2.1, the above intergral is minimized iff $t \in \Big[F_X^{-1}(\tfrac{1}{2}^-)\,,\, F_X^{-1}(\tfrac{1}{2}^+)\Big]$. ■

Remark 3.2.4. Note that in the above corollary if F_X is strictly increasing, then F_X^{-1} will be continuous, and therefore $F_X^{-1}(\tfrac{1}{2}^-) = F_X^{-1}(\tfrac{1}{2}^+)$ and thus the integral is minimized iff $t = F_X^{-1}(\tfrac{1}{2})$, i.e., iff t is the median value.

If we let the probability space $(\Omega, \mathcal{M}, \mathbb{P})$ in Theorem 3.2.3 be $\big([0,1), \mathcal{B}_{[0,1)}, \lambda\big)$ and rename the random variable X to g, it tells us, just as Theorem 3.2.1 did, that $\|g - t\|_1$ is minimized if t is the median value of g. Unlike Theorem 3.2.1, however, which required g to be a monotone function, Theorem 3.2.3 is making this statement for absolutely continuous g. Both of these results can be thought of as special cases of a general theorem that tells us g only needs to be integrable for this result to hold true (see Theorem 3.2.13).

We will show in Theorem 3.2.10 that the integral of g can be re-written as the integral of an increasing function, namely its cumulative distribution function in the usual sense (CDF). The usual notation for the CDF associated with a random variable g is F_g [9]; however, since we have used the capital F notation for distribution functions in the sense of Definition 1.3.4, we use the notation $\widehat{g}$ for the CDF of g.

Definition 3.2.5 (CDF)**.** Consider the probability space $\big([0,1)\,,\,\mathcal{B}_{[0,1)},\lambda\big)$. Let $g\colon\ [0,1)\to\mathbb{R}$ be a $\mathcal{B}_{[0,1)}$-$\mathcal{B}_{\mathbb{R}}$-measurable function (aka a random variable). Consider the pushforward probability measure $\lambda\circ g^{-1}$ on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$. We define $\widehat{g}$ to be the cumulative distribution function (CDF) of that probability distribution, i.e., we define $\widehat{g}\colon\mathbb{R}\to[0,1]$ as

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

where $\{g\le t\}$ is understood to be $\big\{s\in[0,1):g(s)\le t\big\}$.

Remark 3.2.6. Note that $\widehat{g+a}=\widehat{g}(\,\cdot-a)$ for every $a\in\mathbb{R}$.

Lemma 3.2.7. *Let $g\colon\ [0,1)\to\mathbb{R}$ be $\mathcal{B}_{[0,1)}$-$\mathcal{B}_{\mathbb{R}}$-measurable, and let $\widehat{g}\colon\mathbb{R}\to[0,1]$ be the CDF of g. Then,*

$$\forall t_0\in\mathbb{R},\qquad \widehat{g}\big(t_0^-\big)=\lambda\left(\{g<t_0\}\right)\quad.$$

Proof. Let an arbitrary $t_0\in\mathbb{R}$ be given. Then

$$\begin{aligned}\widehat{g}\big(t_0^-\big)\;&=\;\lim_{t\nearrow t_0}\widehat{g}(t)\;=\;\lim_{t\nearrow t_0}\lambda\left(\{g\le t\}\right)\;=\;\lim_{n\to\infty}\lambda\left(\left\{g\le t_0-\frac{1}{n}\right\}\right)\\&=\;\lambda\left(\bigcup_{n=1}^{\infty}\left\{g\le t_0-\frac{1}{n}\right\}\right)\;=\;\lambda\left(\{g<t_0\}\right)\quad.\end{aligned}$$

■

Consider any CDF $\widehat{g}\colon\mathbb{R}\to[0,1]$ as defined in Definition 3.2.5. Being an increasing function, we can define its inverse as in Definition 3.1.6. However, in this setting, we don't have to be concerned with the extended reals. Since $\lim_{t\to-\infty}\widehat{g}(t)=0$ and $\lim_{t\to+\infty}\widehat{g}(t)=1$, we know that for every s in $(0,1)$,

$\{\widehat{g} \le s\} \neq \emptyset$. We thus define $\widehat{g}^{-1} \colon (0,1) \to \mathbb{R}$ as

$$\forall s \in (0,1)\,, \qquad \widehat{g}^{-1}(s) := \sup\{\widehat{g} \le s\} \quad , \tag{3.7}$$

where $\{\widehat{g} \le s\}$ is understood to mean $\big\{t \in \mathbb{R} : \widehat{g}(t) \le s\big\}$.

Remark 3.2.8. By Remark 3.1.7 it is immediate that $\widehat{g}^{-1}$ as defined in (3.7) is increasing and right-continuous. Therefore, one can interpret this to mean that the set $\big\{F_\mu : \mu \in \mathcal{P}\big\} \subset [0,1]^{[0,1]}$ of all distribution functions on $[0,1)$ is closed under inversion, provided we treat the end-points with care.

Remark 3.2.9. Note that $\widehat{g+a}^{-1} = \widehat{g}^{-1} + a$ for every $a \in \mathbb{R}$.

We now turn to prove the most essential part of characterizing t_{min} which asserts the following: If our measurable function (random variable) g is integrable (has an expected value) then its integral (expectation) can be written as the integral of its CDF.

Theorem 3.2.10. *Let $g \in L^1\big([0,1)\,, \mathcal{B}_{[0,1)}\,, \lambda\big)$, and let $\widehat{g} \colon \mathbb{R} \to [0,1]$ be the CDF of g. Then*

$$\int_{[0,1)} g\, d\lambda \;=\; \int_0^1 \widehat{g}^{-1}(s)\, ds \quad .$$

Proof. Recall that any real-valued measurable function g has a positive part g^+ and a negative part g^- defined as

$$g^+ := \max\{g, 0\} \qquad \wedge \qquad g^- := \max\{-g, 0\} \quad ,$$

both of which are non-negative measurable functions; and $g = g^+ - g^-$.
By definition,

$$\begin{aligned}
\int_{[0,1)} g\, d\lambda \;&= \int_{[0,1)} g^+\, d\lambda \;-\int_{[0,1)} g^-\, d\lambda \\
&= \int_0^{+\infty} \lambda\left(g^+ \ge t\right)\, dt \;-\int_0^{+\infty} \lambda\left(g^- \ge t\right)\, dt \\
&= \int_0^{+\infty} \Big(1 - \lambda\left(g^+ < t\right)\Big)\, dt \;-\int_0^{+\infty} \lambda\left(g^- \ge t\right)\, dt\,,
\end{aligned} \tag{3.8}$$

where the second equality is by Corollary 3.1.9. Note that

$$\begin{aligned}
&\forall t \in (0,+\infty)\,, \quad \left\{g^+ < t\right\} = \{g < t\} \\
\implies &\forall t \in (0,+\infty)\,, \quad \lambda\Big(\left\{g^+ < t\right\}\Big) = \lambda\Big(\{g < t\}\Big) \overset{\text{Lemma 3.2.7}}{=\!=} \widehat{g}\big(t^-\big) \quad . \qquad (3.9)
\end{aligned}$$

Also note that

$$\begin{aligned}
&\forall t \in (0,+\infty)\,, \quad \left\{g^- \geq t\right\} = \left\{-g^- \leq -t\right\} = \{g \leq -t\} \\
\implies &\forall t \in (0,+\infty)\,, \quad \lambda\Big(\left\{g^- \geq t\right\}\Big) = \lambda\Big(\{g \leq -t\}\Big) = \widehat{g}\,(-t) \quad . \qquad (3.10)
\end{aligned}$$

Thus using (3.9) and (3.10), we conclude from (3.8) that

$$\int_{[0,1)} g\, d\lambda = \underbrace{\int_0^{\infty} \big(1 - \widehat{g}(t^-)\big)\, dt}_{I_1} - \underbrace{\int_{-\infty}^{0} \widehat{g}(t)\, dt}_{I_2} \quad . \qquad (3.11)$$

Note that at the points of continuity of $\widehat{g}$, we have $\widehat{g}(t^-) = \widehat{g}(t)$. And since $\widehat{g}$ is monotone, by Lemma 3.1.4 we know the points of discontinuity are countable. Thus

$$\widehat{g}(.^-) \overset{\lambda\text{-a.e.}}{=\!=} \widehat{g}(.) \quad . \qquad (3.12)$$

Also, (3.12) clearly implies

$$1 - \widehat{g}(.^-) \overset{\lambda\text{-a.e.}}{=\!=} 1 - \widehat{g}(.) \quad .$$

Thus I_1 equals the area of the region above $\widehat{g}$ and below the constant 1 on $[0,+\infty)$. Let us call this area E_{pos}, i.e., $E_{\text{pos}} := \big\{(t,s) \in [0,+\infty) \times [0,1] \,:\, \widehat{g}(t) \leq s\big\}$. Thus,

$$\begin{aligned}
I_1 &= \iint_{[0,+\infty)\times[0,1]} \mathbb{1}_{E_{\text{pos}}}(t,s)\, d\,(\lambda \times \lambda)\,(t,s) \qquad (3.13) \\
&= \int_0^1 \left(\int_0^{+\infty} \mathbb{1}_{\left(E_{\text{pos}}\right)^s}(t)\; d\lambda(t) \right) d\lambda(s) = \int_0^1 \lambda\left(\left(E_{\text{pos}}\right)^s \right) ds \quad ,
\end{aligned}$$

where in the first equality, the second λ is understood to mean $\lambda\Big|_{[0,1]}$, and the second equality is by Fubini's theorem. By definition, we know

$$\forall s \in [0,1], \quad \left(E_{\text{pos}}\right)^s := \left\{t \in [0,+\infty) : (t,s) \in E_{\text{pos}}\right\} = \left\{t \in [0,+\infty) : \widehat{g}(t^-) \leq s\right\}$$

$$\overset{(3.12)}{\Longrightarrow} \quad \left(E_{\text{pos}}\right)^s \overset{\lambda\text{-a.e.}}{=\!=} \left\{t \in [0,+\infty) : \widehat{g}(t) \leq s\right\} = \{\widehat{g} \leq s\} \cap [0,+\infty)$$

$$\Longrightarrow \quad I_1 = \int_0^1 \lambda\Big(\; \{\widehat{g} \leq s\} \cap [0,+\infty) \; \Big)\, ds \quad . \tag{3.14}$$

Similarly, I_2 equals the area of the region below $\widehat{g}$ and above the constant 0 on $(-\infty, 0]$. Let us call this area E_{neg}, i.e., $E_{\text{neg}} := \left\{(t,s) \in (-\infty,0] \times [0,1] : s \leq \widehat{g}(t) \;\right\}$. Thus,

$$I_2 \; = \iint_{(-\infty,0]\times[0,1]} \mathbb{1}_{E_{\text{neg}}}(t,s)\, d\,(\lambda \times \lambda)\,(t,s) \quad ,$$

where the second λ is understood to mean $\lambda\Big|_{[0,1]}$. Thus by Fubini's theorem,

$$\begin{aligned} I_2 \; &= \int_0^1 \left(\int_{-\infty}^0 \mathbb{1}_{\left(E_{\text{neg}}\right)^s}(t)\; d\lambda(t) \right) d\lambda(s) = \int_0^1 \lambda\left(\left(E_{\text{neg}}\right)^s \right) ds \\ &= \int_0^1 \lambda\left(\left\{ t \in (-\infty,0] : s \leq \widehat{g} \right\} \right) ds \\ &= \int_0^1 \lambda\Big(\; \{\widehat{g} \geq s\} \cap (-\infty, 0] \; \Big)\, ds \quad . \end{aligned} \tag{3.15}$$

Thus using (3.14) and (3.15), we conclude from (3.11),

$$\begin{aligned} \int_{[0,1)} g\, d\lambda \; &= \int_0^1 \lambda\Big(\; \{\widehat{g} \leq s\} \cap [0,+\infty) \; \Big)\, ds \, - \int_0^1 \lambda\Big(\; \{\widehat{g} \geq s\} \cap (-\infty,0] \; \Big)\, ds \\ &= \int_0^1 \max\left\{\widehat{g}^{-1}(s)\,,\, 0\right\} ds \; - \int_0^1 -\min\left\{\widehat{g}^{-1}(s)\,,\, 0\right\} ds \\ &= \int_0^1 \widehat{g}^{-1}(s)\, ds \quad , \end{aligned}$$

where the penultimate equality is because $\lambda\left(\{\widehat{g} \leq s\}\right) = \lambda\left(\,(-\infty\,,\,\widehat{g}^{-1}(s)]\,\right)$ for every $s \in (0,1)$. ■

Theorem 3.2.11. *Let* $g \in L^1\left([0,1)\,,\,\mathcal{B}_{[0,1)}\,,\,\lambda\right)$ *, and let* $\widehat{g}\colon \mathbb{R} \to [0,1]$ *be the CDF of* g *. Then,*

$$\int_{[0,1)} |g|\; d\lambda \;=\; \int_0^1 \left|\widehat{g}^{-1}(s)\right| ds \quad .$$

Proof. Recall that $|g| = g^+ + g^-$. With that, the rest of the proof is analogous to the proof of Theorem 3.2.10. ■

Corollary 3.2.12. *For every* $g \in L^1\left([0,1)\,,\,\mathcal{B}_{[0,1)}\,,\,\lambda\right)$ *,*

$$\int_{[0,1)} |g - t|\; d\lambda \;=\; \int_0^1 \left|\widehat{g}^{-1}(s) - t\right| ds \quad .$$

Proof. Immediate from Theorem 3.2.11 through Remark 3.2.9. ■

Theorem 3.2.13. *For every* $\mu, \nu \in \mathcal{P}$*,*

$$\int_0^1 \left|F_\mu(s) - F_\nu(s) - t\right| ds \;\; \textit{is minimized} \quad \Longleftrightarrow \quad \begin{cases} \lambda\left(\left\{F_\mu - F_\nu \le t\right\}\right) \ge \frac{1}{2}\ , \\ \qquad\qquad \wedge \\ \lambda\left(\left\{F_\mu - F_\nu \ge t\right\}\right) \ge \frac{1}{2}\ , \end{cases}$$

where λ *is understood to mean* $\lambda\Big|_{[0,1]}$*.*

Proof. We know that every monotone function is (Riemann) integrable [25]. Thus for every probability measure $\mu \in \mathcal{P}$, the distribution function F_μ is integrable when considered as a function on $[0,1)$, and so is $F_\mu - F_\nu$, i.e., $F_\mu - F_\nu \in L^1\left([0,1)\,,\,\mathcal{B}_{[0,1)}\right)$. Therefore by Corollary 3.2.12, we know

$$\int_0^1 \left|F_\mu(s) - F_\nu(s) - t\right| ds = \int_0^1 \left|(\widehat{F_\mu - F_\nu})^{-1} - t\right| ds \quad .$$

By Remark 3.1.7, $(\widehat{F_\mu - F_\nu})^{-1}$ is increasing. Thus by Theorem 3.2.1, the above integral is minimized iff $t \in \left[(\widehat{F_\mu - F_\nu})^{-1}(\frac{1}{2}^-)\,,\,(\widehat{F_\mu - F_\nu})^{-1}(\frac{1}{2}^+)\right]$, which is the case iff $\lambda\Big(\,\big\{F_\mu - F_\nu \le t_{min}\big\}\,\Big) \ge \frac{1}{2}$ and $\lambda\Big(\,\big\{F_\mu - F_\nu \ge t_{min}\big\}\,\Big) \ge \frac{1}{2}$.

■

A number of times in Chapter 4 we find the distance between two (shifted) continuous distribution functions F_μ and $F_\nu + t_{min}$ that intersect at precisely two points $\big(s_1\,,F_\mu(s_1)\big)$ and $\big(s_2\,,F_\mu(s_2)\big)$. One implication of Theorem 3.2.13 is that $s_2 - s_1 = \frac{1}{2}$ through the following steps:

$$
\begin{aligned}
& \lambda\Big(\,\big\{F_\mu \le F_\nu + t_{min}\big\}\,\Big) \ge \frac{1}{2} \quad \wedge \quad \lambda\Big(\,\big\{F_\mu \ge F_\nu + t_{min}\big\}\,\Big) \ge \frac{1}{2} \\
\Longrightarrow \quad & \lambda\left(\,[s_1\,,s_2]\,\right) \ge \frac{1}{2} \quad \wedge \quad \lambda\left(\,[0\,,s_1] \uplus [s_2\,,1]\,\right) \ge \frac{1}{2} \\
\Longrightarrow \quad & s_2 - s_1 \ge \frac{1}{2} \quad \wedge \quad (s_1 - 0) + (1 - s_2) \ge \frac{1}{2} \\
\Longrightarrow \quad & s_2 - s_1 \ge \frac{1}{2} \quad \wedge \quad 1 - (s_2 - s_1) \ge \frac{1}{2} \\
\Longrightarrow \quad & s_2 - s_1 \ge \frac{1}{2} \quad \wedge \quad s_2 - s_1 \le \frac{1}{2} \quad \Longrightarrow \quad s_2 - s_1 = \frac{1}{2}\ .
\end{aligned}
$$

Chapter 4

Elementary Properties of $(\mathcal{P}, d_{\mathbb{T}})$

Chapters 2 and 3 together provide us with a tool to calculate the Kantorovich distance between probability distributions. In this chapter, we build a picture of the points in $(\mathcal{P}, d_{\mathbb{T}})$ by calculating some of these distances. These calculations are simple yet long. The reader may choose to skip the proofs of these distance results.

4.1 Preparatory work

The following lemma will be used in proving that the Kantorovich distance of no probability distribution to the uniform distribution is ever more than $\frac{1}{4}$ (see Theorem 4.2.1).

Lemma 4.1.1. *Let $(a_i)_{i=1}^{\infty}$ be a sequence of non-negative numbers. Then,*

$$\sum_{i=1}^{\infty} a_i^2 \leq \left(\sum_{i=1}^{\infty} a_i \right)^2 \quad . \tag{4.1}$$

Moreover, equality in (4.1) with both sides being finite is achieved if and only if $a_i \neq 0$ for at most one $i \in \mathbb{N}$; i.e.,

$$\sum_{i=1}^{\infty} a_i^2 = \left(\sum_{i=1}^{\infty} a_i \right)^2 \quad \iff \quad \#\{i \in \mathbb{N} : a_i \neq 0\} \leq 1 \quad . \tag{4.2}$$

Proof. Note that (4.1) holds trivially if $\sum_{i=1}^{\infty} a_i = +\infty$. Thus assume that the series is (absolutely) convergent in $\mathbb{R}^+$. Let arbitrary $n \in \mathbb{N}$ be given, and consider the non-negative real numbers $a_1, \cdots, a_n$. This proof essentially comes down to the following inequality

$$\sum_{i=1}^{n} {a_i}^2 \le \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_i a_j \right) , \tag{4.3}$$

which holds true because the RHS sum includes the terms in the LHS sum when $i = j$, as well as some additional non-negative terms $a_i a_j$ when $i \neq j$. Taking the constant factors out of the RHS sums twice, we have

$$\sum_{i=1}^{n} {a_i}^2 \le \sum_{i=1}^{n} \left(a_i \left(\sum_{j=1}^{n} a_j \right) \right) = \left(\sum_{j=1}^{n} a_j \right) \left(\sum_{i=1}^{n} a_i \right) = \left(\sum_{i=1}^{n} a_i \right)^2 ,$$

Taking the limit as $n \to \infty$, we conclude the truth of (4.1). To prove (4.2), we note that the converse implication is immediately true. That is, (4.3) turns into an equality if $a_i a_j = 0$ for all distinct $i, j \in \{1, \cdots, n\}$. To prove the forward implication, assume (4.3) is an equality, and assume by contradiction that there exists $i_0 \neq j_0 \in \{1, \cdots, n\}$ such that $a_{i_0} a_{j_0} > 0$. Then

$$\sum_{i=1}^{n} {a_i}^2 = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_i a_j \right) = \sum_{i=j} a_i a_j + \sum_{i \neq j} a_i a_j \ge \sum_{i=1}^{n} a_i^2 + a_{i_0} a_{j_0} > \sum_{i=1}^{n} a_i^2 , \text{※}$$

a clear contradiction. Taking the limit as $n \to \infty$, we have proven (4.2). ■

Remark 4.1.2. The hyperbolic tangent function $\tanh \colon \mathbb{R} \to (-1, 1)$ given by $x \mapsto \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ is a strictly increasing odd function.

We close this section by talking about *compactifications* of $\mathbb{R}$. A compactification of $\mathbb{R}$ is a compact (topological) space in which $\mathbb{R}$ is a dense subspace [1]. The usual extension $\overline{\mathbb{R}}$ of the real numbers is *the two-point compactification* of $\mathbb{R}$, because we essentially union two additional points, namely $+\infty$ and $-\infty$, with $\mathbb{R}$ and form a compact space in which $\mathbb{R}$ is dense. The easiest way to appreciate the topology of $\overline{\mathbb{R}}$ is through visualizing $\mathbb{R}$ as the open interval $(-1,1)$. The function $\tanh\colon \mathbb{R} \to (-1,1)$ is a topological isomorphism between the two spaces. Intuitively, it feels like two points are 'missing' from $(-1,1)$. While the endpoints -1 and 1 are not part of $(-1,1)$, we know what their open neighborhoods would be if they were to be included to make the interval compact. Through the bijection $\tanh$, the open neighborhoods of -1 and 1 precisely determine the open neighborhoods of $-\infty$ and $+\infty$ respectively. For example, a sequence $(x_n)_{n=1}^{\infty}$ in $\overline{\mathbb{R}}$ converges to $+\infty$ if and only if $(\tanh x_n)_{n=1}^{\infty}$ in $[-1,1]$ converges to 1. In this view, we can think of $\mathbb{R}$ as an interval with two endpoints points missing.

Since $\mathbb{R}$ is locally compact, it is known to also have a *one-point compactification* [14]. As the name suggests, here we only include one additional point, namely ∞, to form a compact space in which $\mathbb{R}$ is dense. Denote this space $\mathbb{R}_\infty$. The open neighborhoods of ∞ are defined to be those sets U whose complement in $\mathbb{R}_\infty$ forms a compact subspace of $\mathbb{R}$. The space $\mathbb{R}_\infty$ is known to be topologically isomorphic to a circle [18]. Again, the easiest way to appreciate this is to visualize $\mathbb{R}$ as the interval $(-1,1)$ bent so that the two endpoints coincide. Therefore in this view, we can think of $\mathbb{R}$ as a circle with one point missing. For example, the mapping $h\colon \mathbb{T} \setminus \{0+\mathbb{Z}\} \to (-1,1)$ given by

$$h\big(r + \frac{1}{2} + \mathbb{Z}\big) := 2r \qquad \forall r \in \big(-\frac{1}{2}, \frac{1}{2}\big) \quad ,$$

is a topological isomorphism.

4.2 Some topological facts about $(\mathcal{P}, d_{\mathbb{T}})$

The main goal of this section is to provide the mental picture of $(\mathcal{P}, d_{\mathbb{T}})$ presented in Figure 4.4. In doing so, we also present some explicit formulae for the distances between Dirac, uniform, and exponential distributions and their rotated versions. We begin by proving that the distance between $\lambda_{\mathbb{T}}$ and any $\mu \in \mathcal{P}$ is at most $\frac{1}{4}$, and thereby we are justified in thinking about $\mathcal{P}$ at a ball centered at $\lambda_{\mathbb{T}}$.

Theorem 4.2.1. *For every* $\mu \in \mathcal{P}$ *,*

$$d_{\mathbb{T}}(\mu, \lambda_{\mathbb{T}}) \leq \frac{1}{4} \quad .$$

Proof. We first prove the result for $\mu \in \mathcal{P}_C$, i.e., when F_μ is a continuous function on $[0,1)$. The assertion will then follow from the fact that $\mathcal{P}_C$ is dense in $\mathcal{P}$.

Let an arbitrary $\mu \in \mathcal{P}_C$ be given. We know by Theorem 2.2.8 that $d_{\mathbb{T}}(\mu, \lambda_{\mathbb{T}}) = \int_{[0,1)} \left| F_\mu(s) - (s + t_{min}) \right| ds$. By Theorem 3.2.13, the continuous functions F_μ and $F_{\lambda_{\mathbb{T}}} + t_{min}$ must have at least one intersection on $[0,1)$.

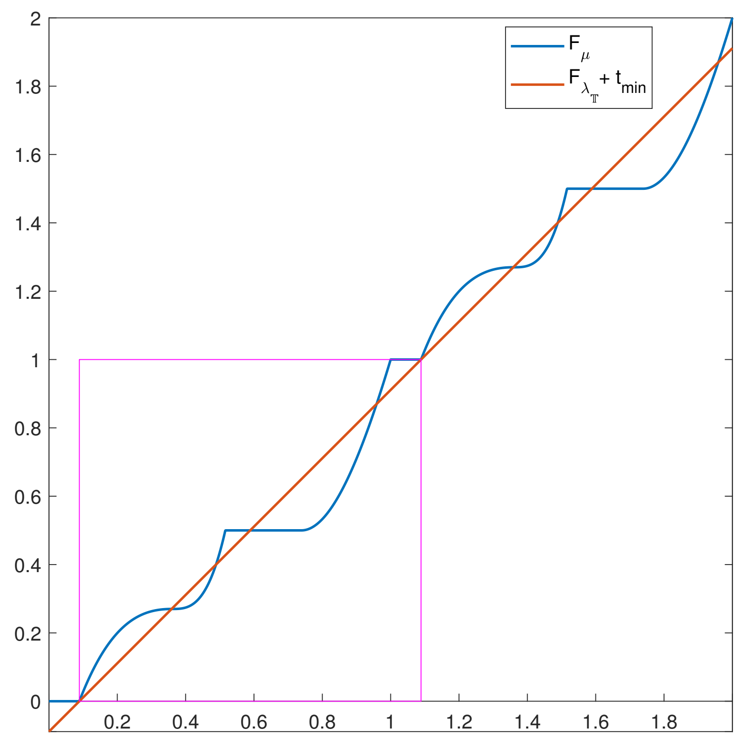

Figure 4.1: The graph of the functions F_μ and $F_{\lambda_{\mathbb{T}}} + t_{min}$ where $\mu \in \mathcal{P}$ is continuous. When these two functions do not intersect at the origin, one can rotate both measures, i.e., shift the 1×1 frame of the distribution functions, so that the origin becomes their intersection.

WLOG assume that both F_μ and $F_{\lambda_{\mathbb{T}}} + t_{min}$ intersect at the origin; otherwise, using the fact that $d_{\mathbb{T}}$ is invariant under rotation, we can rotate both

μ and $\lambda_{\mathbb{T}}$ appropriately. As explained in Section 1.6, this can be viewed as a frame shift when viewing graphs of F_μ and $F_{\lambda_{\mathbb{T}}}$ (see Figures 1.5 and 4.1). Note that the shift of frame in order to shift the intersection to the origin also means we are assuming WLOG that (the rotated) F_μ and $F_{\lambda_{\mathbb{T}}} + t_{min}$ intersect at $s = 1$ as well (see Figure 4.1). The functions F_μ and $F_{\lambda_{\mathbb{T}}} + t_{min}$ (now WLOG assumed to be $F_{\lambda_{\mathbb{T}}}$) may coincide on intervals of positive measure. These intervals' contributions to the L^1 distance is clearly 0. Therefore

$$\int_{[0,1)} \left|F_\mu(s) - s\right| \, ds =$$

$$\cancelto{0}{\int_{\{F_\mu = F_{\lambda_{\mathbb{T}}}\}\big|_{[0,1)}} \left|F_\mu(s) - s\right| \, ds} + \int_{\{F_\mu \neq F_{\lambda_{\mathbb{T}}}\}\big|_{[0,1)}} \left|F_\mu(s) - s\right| \, ds$$

$$= \int_{\{F_\mu > F_{\lambda_{\mathbb{T}}}\}\big|_{[0,1)}} \big(F_\mu(s) - s\big) \, ds + \int_{\{F_\mu < F_{\lambda_{\mathbb{T}}}\}\big|_{[0,1)}} \big(s - F_\mu(s)\big) \, ds \quad .$$

F_μ and $F_{\lambda_{\mathbb{T}}}$ can also intersect at isolated points. Thus $\{F_\mu > F_{\lambda_{\mathbb{T}}}\}\Big|_{[0,1)}$ is a union of intervals, and there are countably many such intervals, i.e.,

$$\{F_\mu > F_{\lambda_{\mathbb{T}}}\}\Big|_{[0,1)} = \biguplus_{j=1}^{\infty} A_j \qquad \text{where } \forall j \in \mathbb{N},\ A_j \subseteq [0,1) \text{ is an open interval}.$$

See Figure 4.2. Similarly,

$$\{F_\mu < F_{\lambda_{\mathbb{T}}}\}\Big|_{[0,1)} = \biguplus_{k=1}^{\infty} B_k \qquad \text{where } \forall k \in \mathbb{N},\ B_k \subseteq [0,1) \text{ is an open interval}.$$

Therefore

$$\int_{[0,1)} \left|F_\mu(s) - s\right| \, ds = \sum_{j=1}^{\infty} \int_{A_j} \big(F_\mu(s) - s\big) \, ds + \sum_{k=1}^{\infty} \int_{B_k} \big(s - F_\mu(s)\big) \, ds \quad . \tag{4.4}$$

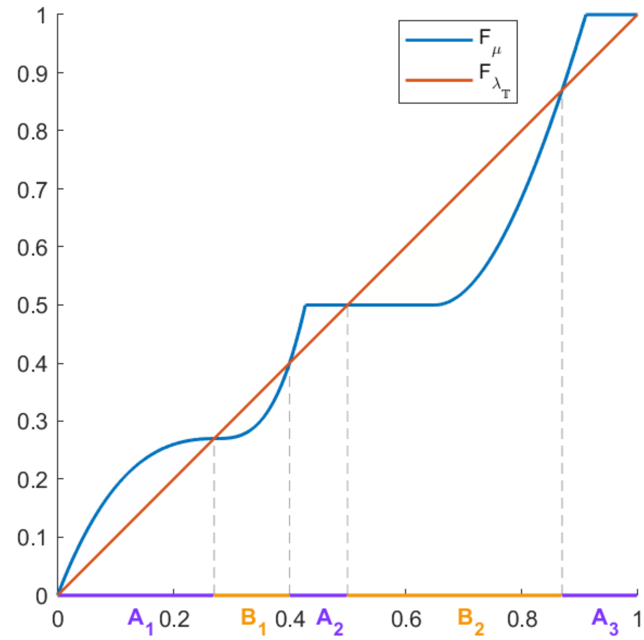

Figure 4.2: The graphs of F_μ and $F_{\lambda_\mathbb{T}}$ partition the interval $[0,1)$ into countably many open intervals $\{A_j\}_j$ and $\{B_k\}_k$.

Claim 4.2.1.1. For the distance between F_μ and $F_{\lambda_\mathbb{T}}$ on every A_j, we have the following strict upper bound (see Figure 4.3):

$$\forall A_j\,, \quad \int_{A_j} \big(F_\mu(s) - s\big)\; ds \;<\; \frac{1}{2}\,\lambda(A_j)^2 \quad .$$

Proof. Consider an arbitrary A_j. Let $a_{j1} < a_{j2}$ denote the two endpoints of the interval A_j. By monotonicity of F_μ, we know $F_\mu(a_{j1}) < F_\mu(a_{j2})$. Let $\epsilon := \frac{F_\mu(a_{j2}) - F_\mu(a_{j1})}{2}$. By continuity of F_μ,

$$\exists\, \delta > 0 \;:\; \forall s \in \big[a_{j1}, a_{j1} + \delta\big]\,, \quad F_\mu(s) - s < F_\mu(a_{j2}) - s \quad .$$

Therefore by monotonicity of integrals,

$$\int_{a_{j1}}^{a_{j1}+\delta} \big(\,F_\mu(s) - s\,\big)\, ds \;<\; \int_{a_{j1}}^{a_{j1}+\delta} \big(\,F_\mu(a_{j2}) - s\,\big)\, ds \quad . \tag{4.5}$$

Thus by breaking down the integral over A_j into partitions,

$$\begin{aligned}\int_{A_j}\big(F_\mu(s)-s\big)\,ds &= \int_{a_{j1}}^{a_{j1}+\delta}\big(F_\mu(s)-s\big)\,ds+\int_{a_{j1}+\delta}^{a_{j2}}\big(F_\mu(s)-s\big)\,ds\\ &\overset{(4.5)}{<} \int_{a_{j1}}^{a_{j1}+\delta}\big(F_\mu(a_{j2})-s\big)\,ds+\int_{a_{j1}+\delta}^{a_{j2}}\big(F_\mu(a_{j2})-s\big)\,ds\\ &= \lambda\left(A_j\right)F_\mu(a_{j2})-\frac{1}{2}\left(a_{j2}^2-a_{j1}^2\right)\\ &= \lambda\left(A_j\right)\left(a_{j2}-\frac{1}{2}\left(a_{j2}+a_{j1}\right)\right)\\ &= \lambda\left(A_j\right)\left(\frac{1}{2}\,a_{j2}-\frac{1}{2}\,a_{j1}\right)\;=\;\frac{1}{2}\,\lambda\left(A_j\right)^2\quad.\end{aligned}$$

□

Claim 4.2.1.2. For the distance between F_μ and $F_{\lambda_{\mathbb{T}}}$ on every B_k, we have the following strict upper bound (see Figure 4.3).

$$\forall B_k\,,\quad \int_{B_k}\big(s-F_\mu(s)\big)\,ds\;<\;\frac{1}{2}\,\lambda(B_k)^2\quad.$$

Proof. Analogous to the proof of Claim 4.2.1.1 . □

Equation (4.4) implies through Claims 4.2.1.1 and 4.2.1.2 that

$$\int_{[0,1)}\big|F_\mu(s)-s\big|\,ds\;<\;\frac{1}{2}\sum_{j=1}^{\infty}\lambda(A_j)^2+\frac{1}{2}\sum_{k=1}^{\infty}\lambda(B_k)^2\quad, \tag{4.6}$$

as visualized in Figure 4.3. We also know from Theorem 3.2.13 that

$$\sum_{j=1}^{\infty}\lambda(A_j)\le\frac{1}{2}\quad\wedge\quad\sum_{k=1}^{\infty}\lambda(B_k)\le\frac{1}{2}\quad. \tag{4.7}$$

Thus (4.6) and (4.7) imply through Lemma 4.1.1 that

$$\int_{[0,1)}\big|F_\mu(s)-s\big|\,ds\;<\;\frac{1}{2}\left(\frac{1}{2}\right)^2+\frac{1}{2}\left(\frac{1}{2}\right)^2\;=\;\frac{1}{4}\quad. \tag{4.8}$$

So we have shown that $d_{\mathbb{T}}(\mu, \lambda_{\mathbb{T}}) < \frac{1}{4}$ for every $\mu \in \mathcal{P}_C$. The density of $\mathcal{P}_C$ in $(\mathcal{P}, d_{\mathbb{T}})$ implies that the upper bound $\frac{1}{4}$ holds for the distance of any $\mu \in \mathcal{P}$ to $\lambda_{\mathbb{T}}$ through the following steps: Let arbitrary $\mu \in \mathcal{P}$ be given. By the density of $\mathcal{P}_C$, there exists $(\mu_n)_{n=1}^{\infty}$ in $\mathcal{P}_C$ such that $d_{\mathbb{T}}(\mu_n, \mu) \xrightarrow{n\to\infty} 0$. By the triangle inequality we have for every $n \in \mathbb{N}$,

$$d_{\mathbb{T}}(\mu, \lambda_{\mathbb{T}}) \leq d_{\mathbb{T}}(\mu, \mu_n) + d_{\mathbb{T}}(\mu_n, \lambda_{\mathbb{T}}) < d_{\mathbb{T}}(\mu, \mu_n) + \frac{1}{4} .$$

Therefore,

$$d_{\mathbb{T}}(\mu, \lambda_{\mathbb{T}}) \leq \cancelto{0}{\lim_{n\to\infty} d_{\mathbb{T}}(\mu, \mu_n)} + \lim_{n\to\infty} \frac{1}{4} = \frac{1}{4} ,$$

as desired. ■

Theorem 4.2.2. *For every* $\mu \in \mathcal{P}$,

$$d_{\mathbb{T}}(\mu, \lambda_{\mathbb{T}}) = \frac{1}{4} \quad \Longleftrightarrow \quad \mu = \delta_x \text{ for some } x \in \mathbb{T} .$$

Proof. We will prove both the forward and the converse implications.

$\boxed{\Longrightarrow}$

Assume $d_{\mathbb{T}}(\mu, \lambda_{\mathbb{T}}) = \frac{1}{4}$. We want to show that $\mu = \delta_x$ for some $x \in \mathbb{T}$. The proof of this implication is essentially the same as the proof of Theorem 4.2.1 with a more careful treatment of the inequalities. Without repeating every detail of the proof, we outline what the more careful treatment is.

By the density of $\mathcal{P}_C$ in $\mathcal{P}$ we know there exists $(\mu_n)_{n=1}^{\infty}$ in $\mathcal{P}_C$ with $\mu_n \xrightarrow{n\to\infty} \mu$. Let arbitrary $n \in \mathbb{N}$ be given. Consider the continuous distribution μ_n. As shown in (4.4), we know

$$\int_{[0,1)} \left|F_{\mu_n}(s) - s\right| \, ds = \sum_{j=1}^{\infty} \int_{A_j} \left(F_{\mu_n}(s) - s\right) \, ds + \sum_{k=1}^{\infty} \int_{B_k} \left(s - F_{\mu_n}(s)\right) \, ds .$$

Note that the inequality in (4.6) is strict. So there exists $\epsilon_n > 0$ that makes the inequality non-strict in the following way:

$$\int_{[0,1)} \left|F_{\mu_n}(s) - s\right| \, ds \;\leq\; \frac{1}{2}\sum_{j=1}^{\infty} \lambda(A_j)^2 + \frac{1}{2}\sum_{k=1}^{\infty} \lambda(B_k)^2 - \epsilon_n \quad , \tag{4.9}$$

where ϵ_n depends on the function F_{μ_n}. In particular, a closer look at the proof of Claim 4.2.1.1 reveals that ϵ_n gets smaller the more F_{μ_n} resembles the step function depicted in Figure 4.3.

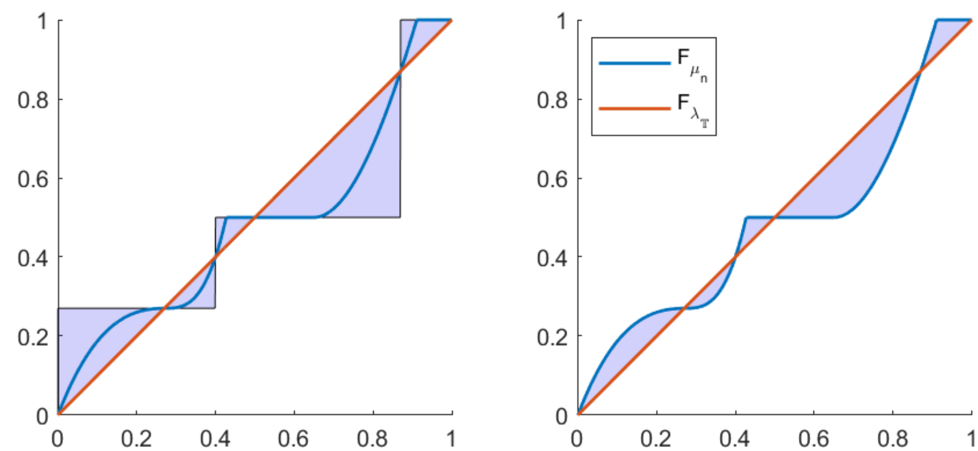

Figure 4.3: The L^1 distance of $F_{\lambda_{\mathbb{T}}}$ to F_{μ_n} (right), and to the step function that is constantly F_{μ_n}'s right-endpoint value on every A_j and left-endpoint value on every B_k (left). By (4.6), the former area is ϵ_n less than the latter for some $\epsilon_n > 0$.

We will shortly assert that our assumption requires ϵ_n to vanish as n grows; but before we do that, we introduce another quantity that must vanish: Consider the series in (4.9). Lemma 4.1.1 gives us an upper bound for each of these series through (4.1). Moreover, (4.2) tells us that the inequality in (4.1) is also strict unless there is at most one term in the series. So there exists $\epsilon'_n \geq 0$ such that

$$\int_{[0,1)} \left|F_{\mu_n}(s) - s\right| \, ds \;\leq\; \frac{1}{2}\left(\sum_{j=1}^{\infty} \lambda(A_j)\right)^2 + \frac{1}{2}\left(\sum_{k=1}^{\infty} \lambda(B_k)\right)^2 - \epsilon_n - \epsilon'_n \quad , \tag{4.10}$$

where ϵ'_n depends on how many times F_{μ_n} intersects $F_{\lambda_{\mathbb{T}}}$. Lastly, we apply the upper bounds in (4.7) to (4.10). Note that the inequalities in (4.7) may also be strict because F_{μ_n} and $F_{\lambda_{\mathbb{T}}}$ may coincide on a set of positive measure. In this proof WLOG we treat only the inequality for $\sum_{j=1}^{\infty} \lambda(A_j)$ with this additional care. So there exists $\epsilon''_n \geq 0$ such that

$$\int_{[0,1)} \left|F_{\mu_n}(s) - s\right| \, ds \;\leq\; \frac{1}{2}\left(\frac{1}{2} - \epsilon''_n\right)^2 + \frac{1}{2}\left(\frac{1}{2}\right)^2 - \epsilon_n - \epsilon'_n \;\;,$$

which is a more accurate version of (4.8). Therefore, through the triangle inequality we know that for every $n \in \mathbb{N}$,

$$\frac{1}{4} \;=\; d_{\mathbb{T}}(\mu, \lambda_{\mathbb{T}}) \;\leq\; d_{\mathbb{T}}(\mu, \mu_n) + \frac{1}{2}\left(\frac{1}{2} - \epsilon''_n\right)^2 + \frac{1}{2}\left(\frac{1}{2}\right)^2 - \epsilon_n - \epsilon'_n \quad ,$$

which after letting $n \to \infty$ implies

$$\frac{1}{4} \;\leq\; \frac{1}{2}\left(\frac{1}{2} - \lim_{n\to\infty} \epsilon''_n\right)^2 + \frac{1}{2}\left(\frac{1}{2}\right)^2 - \lim_{n\to\infty} \epsilon_n - \lim_{n\to\infty} \epsilon'_n \;\; . \tag{4.11}$$

It is now clear that ϵ_n, ϵ'_n, and ϵ''_n all converge to 0; otherwise, (4.11) gives the contradictory statement $\frac{1}{4} \leq \frac{1}{4} - \epsilon$ for some $\epsilon > 0$. The vanishing of each of these limits reveals several properties that μ must have:

(i) $\lim_{n\to\infty} \epsilon''_n = 0$ tells us that F_{μ_n} must not equal $F_{\lambda_{\mathbb{T}}}$ on a set of non-vanishing positive measure. Therefore the (L^1) limit F_μ does not coincide with $F_{\lambda_{\mathbb{T}}}$ on a set of positive measure. In other words, by Theorem 3.2.13,

$$\lambda\Big(\big\{F_\mu > F_{\lambda_{\mathbb{T}}}\big\}\Big|_{[0,1)}\Big) = \lambda\Big(\big\{F_\mu < F_{\lambda_{\mathbb{T}}}\big\}\Big|_{[0,1)}\Big) = \frac{1}{2} \;\; . \tag{4.12}$$

(ii) $\lim_{n\to\infty} \epsilon'_n = 0$ tells us that as $n \to \infty$, $\big\{F_{\mu_n} > F_{\lambda_{\mathbb{T}}}\big\}\Big|_{[0,1)}$ consists of at most one interval A_j, and similarly $\big\{F_\mu < F_{\lambda_{\mathbb{T}}}\big\}\Big|_{[0,1)}$ at most one interval B_k. On the other hand, by Theorem 3.2.13, there must be at least one interval A_j and one B_k; so F_μ has precisely one interval on which it is above $F_{\lambda_{\mathbb{T}}}$ and one interval on which it is below. That is

$$\left\{F_\mu > F_{\lambda_\mathbb{T}}\right\}\Big|_{[0,1)} = A \qquad \wedge \qquad \left\{F_\mu < F_{\lambda_\mathbb{T}}\right\}\Big|_{[0,1)} = B \quad , \tag{4.13}$$

where $A, B \subseteq [0,1)$ are intervals. We can WLOG assume the same for every F_{μ_n} as well, i.e., assume $\left\{F_{\mu_n} > F_{\lambda_\mathbb{T}}\right\}\Big|_{[0,1)}$ and $\left\{F_{\mu_n} < F_{\lambda_\mathbb{T}}\right\}\Big|_{[0,1)}$ are respectively A and B for every $n \in \mathbb{N}$. Note that since every F_{μ_n} is continuous, our WLOG assumptions imply that all F_{μ_n}s intersect $F_{\lambda_\mathbb{T}}$ on the same points in $[0,1)$, namely the endpoints of A and B.

(iii) $\lim_{n\to\infty} \epsilon_n = 0$ tells us through (4.13) together with our WLOG assumptions and a careful look at the proof of Claim 4.2.1.1 that F_{μ_n} converges to $F_{\mu_n}(a_2)$ on A in the L^1 sense, where a_2 is the right endpoint of A. Analogously for Claim 4.2.1.2, it tells us that F_{μ_n} converges on B to $F_{\mu_n}(b_1)$ where b_1 is the left endpoint of B. Since these endpoints are the points on which the F_{μ_n} intersect $F_{\lambda_\mathbb{T}}$, we know that $F_{\mu_n}(a_2) = a_2$ and $F_{\mu_n}(b_1) = b_1$ for every $n \in \mathbb{N}$. So F_μ is the step function that takes the value b_1 on B, and a_2 on A, i.e.,

$$\forall s \in [0,1), \quad F_\mu(s) = \begin{cases} F_\mu(b_1) & s \in B \quad , \\ F_\mu(a_2) & s \in A \quad . \end{cases} \tag{4.14}$$

(4.12) and (4.13) imply that A and B are two disjoint subintervals of $[0,1)$ the length of each of which is $\frac{1}{2}$. Recalling that neither include their endpoints, clearly one of them must be $(0, \frac{1}{2})$ and the other $(\frac{1}{2}, 1)$. If B is $(0, \frac{1}{2})$, then by (4.14), the value of F_μ on B is 0. Similarly in that case the value of F_μ on A is 1. The value at $\frac{1}{2}$ must be 1 to preserve right-continuity. So in this case,

$$\forall s \in [0,1), \quad F_\mu(s) = \begin{cases} 0 & s \in [0, \frac{1}{2}) \quad , \\ 1 & s \in [\frac{1}{2}, 1) \quad ; \end{cases}$$

which is the distribution function of $\delta_{\frac{1}{2}}$. On the other hand if A is $(0, \frac{1}{2})$ then F_μ constantly has the value $\frac{1}{2}$, which is the shifted distribution function of δ_0. In either case, the distribution μ ended up a Dirac measure after our WLOG rotation. Once rotated back, it will still be a Dirac distribution.

$\Longleftarrow$

Let arbitrary $x \in \mathbb{T}$ be given and assume $\mu = \delta_x$. We want to show that $d_{\mathbb{T}}(\mu, \lambda_{\mathbb{T}}) = \frac{1}{4}$. By Theorem 2.2.8, $d_{\mathbb{T}}(\lambda_{\mathbb{T}}, \mu) = \min_{t\in\mathbb{R}} \int_0^1 \left|s - F_{\delta_x}(s) - t\right| \, ds$. We will show that $\min_{t\in\mathbb{R}} \int_0^1 \left|s - F_{\delta_x}(s) - t\right| \, ds = \min_{t\in\mathbb{R}} \int_0^1 |s - t| \, ds$, and by Remark 3.2.2 our proof will be complete.

We extend the function $s - F_{\delta_x}(s)\Big|_{[0,1)}$ in a 1-periodic fashion on both sides of $[0, 1)$. To avoid introducing a new notation let the same notation denote the 1-periodic extended version. For convenience, let s_0 denote $\iota_{\mathbb{R}}(x)$.
Partitioning $[0, 1]$ into $[0, 1] \cap \{F_{\delta_x} = 0\}$ and $[0, 1] \cap \{F_{\delta_x} = 1\}$, we have

$$
\begin{aligned}
\forall t \in \mathbb{R}, \quad \int_0^1 \left|s - F_{\delta_x}(s) - t\right| \, ds &= \int_0^{s_0} |s - t| \, ds + \int_{s_0}^1 |s - 1 - t| \, ds \\
&= \int_0^{s_0} |s - t| \, ds + \int_{s_0-1}^0 |s + 1 - 1 - t| \, ds \\
&= \int_0^{s_0} |s - t| \, ds + \int_{s_0-1}^0 |s - t| \, ds \\
&= \int_{s_0-1}^{s_0} |s - t| \, ds = \int_0^1 |s - t + x - 1| \, ds \\
&= \int_0^1 \left|s - (t - x + 1)\right| \, ds \quad .
\end{aligned}
$$

Let $\tilde{t} := t - x + 1$. So we have shown

$$
\min\left\{ \int_0^1 \left|s - F_{\delta_x}(s) - t\right| \, ds \, : \, t \in \mathbb{R} \right\} = \min\left\{ \int_0^1 \left|s - \tilde{t}\,\right| ds \, : \, \tilde{t} \in \mathbb{R} \right\} \quad .
$$

And we know by Remark 3.2.2 that $\min\left\{ \int_0^1 \left|s - \tilde{t}\,\right| ds \, : \, \tilde{t} \in \mathbb{R} \right\} = \frac{1}{4}$. Thus we have shown that $d_{\mathbb{T}}(\mu, \lambda_{\mathbb{T}}) = \frac{1}{4}$. ■

Remark 4.2.3. As a corolllary of Theorem 4.2.1, we see that for every $\mu, \nu \in \mathcal{P}$,

(i) $d_{\mathbb{T}}(\mu, \nu) \leq \frac{1}{2}$,

(ii) $d_{\mathbb{T}}(\mu, \nu) = \frac{1}{2} \iff \mu = \delta_x, \, \nu = \delta_y$ for some $x, y \in \mathbb{T}$ with $d_{\mathbb{T}}(x, y) = \frac{1}{2}$.

As explained at the end of Chapter 1, the set $\{\delta_x \in \mathcal{P} : x \in \mathbb{T}\}$ of all Dirac probability measures can be thought of as an isometric copy of $\mathbb{T}$ in $\mathcal{P}$ (see Theorem 1.5.3).

Theorem 4.2.4. *For every $a \in \mathbb{R} \setminus \{0\}$,*

$$d_{\mathbb{T}}(\eta_a, \delta_0) = \frac{1}{a} \tanh\left(\frac{a}{4}\right) \quad .$$

Proof. Let arbitrary $a \in \mathbb{R} \setminus \{0\}$ be given. By Theorem 2.2.8, $d_{\mathbb{T}}(\eta_a, \delta_0) = \min_{t\in\mathbb{R}} \int_0^1 \left|F_{\eta_a}(s) - F_{\delta_0}(s) - t\right| \, ds$. Thus,

$$d_{\mathbb{T}}(\eta_a, \delta_0) = \min_{t\in\mathbb{R}} \int_0^1 \left| \frac{e^{as}-1}{e^a-1} - 1 - t \right| \, ds \quad .$$

Let $\tilde{t} := 1 + t$. Then,

$$d_{\mathbb{T}}(\eta_a, \delta_0) = \min_{\tilde{t}\in\mathbb{R}} \int_0^1 \left| \frac{e^{as}-1}{e^a-1} - \tilde{t} \right| \, ds \quad .$$

By Theorem 3.2.1 and continuity of F_{η_a}, $\tilde{t} = F_{\eta_a}(\frac{1}{2})$ minimizes the above integral, i.e.,

$$d_{\mathbb{T}}(\eta_a, \delta_0) = \int_0^1 \left| \frac{e^{as}-1}{e^a-1} - \frac{e^{\frac{a}{2}}-1}{e^a-1} \right| \, ds \quad .$$

$\boxed{\text{Case 1}}$ $a > 0$.

Breaking our integral down into a partition,

$$\begin{aligned}
d_{\mathbb{T}}(\eta_a, \delta_0) &= \int_0^{\frac{1}{2}} \left(-\frac{e^{as}-1}{e^a-1} + \frac{e^{\frac{a}{2}}-1}{e^a-1} \right) ds + \int_{\frac{1}{2}}^1 \left(\frac{e^{as}-1}{e^a-1} - \frac{e^{\frac{a}{2}}-1}{e^a-1} \right) ds \\
&= \int_0^{\frac{1}{2}} -\frac{e^{sa}-1}{e^a-1} \, ds + \cancel{\int_0^{\frac{1}{2}} \frac{e^{\frac{a}{2}}-1}{e^a-1} \, ds} + \int_{\frac{1}{2}}^1 \frac{e^{as}-1}{e^a-1} \, ds - \cancel{\int_{\frac{1}{2}}^1 \frac{e^{\frac{a}{2}}-1}{e^a-1} \, ds} \\
&= \left[\frac{-1}{a\,(e^a-1)} \, e^{as} \right]_0^{\frac{1}{2}} + \cancel{\frac{1}{2}\left(\frac{1}{e^a-1}\right)} + \left[\frac{1}{a\,(e^a-1)} \, e^{as} \right]_{\frac{1}{2}}^1 - \cancel{\frac{1}{2}\left(\frac{1}{e^a-1}\right)} \\
&= \frac{-1}{a\,(e^a-1)} \left(e^{\frac{a}{2}} - 1 \right) + \frac{1}{a\,(e^a-1)} \left(e^a - e^{\frac{a}{2}} \right)
\end{aligned}$$

$$= \frac{1}{a}\left(\frac{-2e^{\frac{a}{2}}+e^a+2-1}{e^a-1}\right) = \frac{1}{a}\left(\frac{-2\left(e^{\frac{a}{2}}-1\right)}{e^a-1}+1\right)$$

$$= \frac{1}{a}\left(\frac{-2\cancel{\left(e^{\frac{a}{2}}-1\right)}}{\cancel{\left(e^{\frac{a}{2}}-1\right)}\left(e^{\frac{a}{2}}+1\right)}+1\right) = \frac{1}{a}\left(\frac{-2\,e^{\frac{-a}{4}}}{e^{\frac{a}{4}}+e^{\frac{-a}{4}}}+1\right)$$

$$= \frac{1}{a}\left(\frac{-2\,e^{\frac{-a}{4}}+e^{\frac{a}{4}}+e^{\frac{-a}{4}}}{e^{\frac{a}{4}}+e^{\frac{-a}{4}}}\right) = \frac{1}{a}\left(\frac{-e^{\frac{-a}{4}}+e^{\frac{a}{4}}}{e^{\frac{a}{4}}+e^{\frac{-a}{4}}}\right) = \frac{1}{a}\tanh\left(\frac{a}{4}\right) \quad .$$

Case 2 $a < 0$.

By Theorem 1.6.5 and Remark 1.6.2 we know

$$d_{\mathbb{T}}(\eta_a, \delta_0) \;=\; d_{\mathbb{T}}(\eta_a \circ Q^{-1}, \delta_0 \circ Q^{-1}) \;=\; d_{\mathbb{T}}(\eta_{-a}, \delta_0) \quad . \tag{4.15}$$

On the other hand, since $a < 0$, we have $-a > 0$. Therefore through Case 1 we know that

$$d_{\mathbb{T}}(\eta_{-a}, \delta_0) = \frac{1}{-a}\tanh\left(\frac{-a}{4}\right) = \frac{1}{a}\tanh\left(\frac{a}{4}\right) \quad , \tag{4.16}$$

where the second equality is because tanh is an odd function. By (4.15) and (4.16), therefore,

$$d_{\mathbb{T}}(\eta_a, \delta_0) = \frac{1}{a}\tanh\left(\frac{a}{4}\right) \quad .$$

■

Theorem 4.2.5. *For every* $a \in \mathbb{R} \setminus \{0\}$,

$$d_{\mathbb{T}}(\eta_a, \lambda_{\mathbb{T}}) = \frac{1}{|a|}\log\Big(\cosh\left(\frac{a}{4}\right)\Big) \quad .$$

Proof. By Theorem 2.2.8, $d_{\mathbb{T}}(\eta_a, \lambda_{\mathbb{T}}) \;=\; \min_{t\in\mathbb{R}} \int_0^1 \left|F_{\eta_a} - F_{\lambda_{\mathbb{T}}} - t\right| \, ds$. By its convexity or concavity, F_{η_a} intersects the line $F_{\lambda_{\mathbb{T}}} + t_{min}$ at two points, partitioning $[0,1)$ into the interval between the points and the complement of the interval in $[0,1)$ as explained at the end of Chapter 3. Let $\big(s_1\,, F_{\lambda_{\mathbb{T}}}(s_1)\big)$ and

$\big(s_2\,, F_{\lambda_{\mathbb{T}}}(s_2)\big)$ denote the intersection points, where we assume $s_1 < s_2$. By definition,

$$F_{\eta_a}(s_1) = F_{\lambda_{\mathbb{T}}}(s_1) + t_{min} \implies \frac{e^{as_1}-1}{e^a-1} = s_1 + t_{min} \quad ,$$
$$F_{\eta_a}(s_2) = F_{\lambda_{\mathbb{T}}}(s_2) + t_{min} \implies \frac{e^{as_2}-1}{e^a-1} = s_2 + t_{min} \quad .$$

In addition, we know that $s_2 - s_1 = \frac{1}{2}$ (see the end of Chapter 3). Thus we have three equations for the three unknowns s_1, s_2, and t_{min}. Subtracting the first equation from the second yields

$$\begin{aligned} \frac{e^{as_2}-e^{as_1}}{e^a-1} &= s_2 - s_1 \\ \implies \qquad \frac{e^{a\left(s_1+\frac{1}{2}\right)}-e^{as_1}}{e^a-1} &= \frac{1}{2} \\ \implies \qquad \frac{e^{as_1}\left(e^{\frac{a}{2}}-1\right)}{\left(e^{\frac{a}{2}}-1\right)\left(e^{\frac{a}{2}}+1\right)} &= \frac{1}{2} \quad . \end{aligned}$$

Therefore we have

$$\begin{aligned} s_1 = \frac{1}{a}\log\left(\frac{e^{\frac{a}{4}}+e^{\frac{-a}{4}}}{2}\,e^{\frac{a}{4}}\right) &= \frac{1}{a}\log\left(\cosh\left(\frac{a}{4}\right)\right) + \frac{1}{a}\left(\frac{a}{4}\right) \\ &= \frac{1}{a}\log\left(\cosh\left(\frac{a}{4}\right)\right) + \frac{1}{4} \quad . \qquad (4.17) \end{aligned}$$

With these preparations, we are ready to calculate

$$d_{\mathbb{T}}(\eta_a, \lambda_{\mathbb{T}}) = \int_0^1 \left|\frac{e^{as}-1}{e^a-1} - s - t_{min}\right| ds \quad .$$

Case 1 $a < 0$

Breaking our integral down into a partition,

$$d_{\mathbb{T}}(\eta_a, \lambda_{\mathbb{T}}) = \underbrace{\int_0^{s_1} \left(\frac{1 - e^{as}}{e^a - 1} + s + t_{min} \right) ds}_{I_1} + \underbrace{\int_{s_1}^{s_2} \left(\frac{e^{as} - 1}{e^a - 1} - s - t_{min} \right) ds}_{I_2} + \underbrace{\int_{s_2}^{1} \left(\frac{1 - e^{as}}{e^a - 1} + s + t_{min} \right) ds}_{I_3} \quad .$$

Calculating I_1,

$$\begin{aligned} I_1 &= \frac{s_1}{e^a - 1} - \left[\frac{1}{a\,(e^a - 1)} e^{as} \right]_0^{s_1} + \left[\frac{1}{2} s^2 \right]_0^{s_1} + s_1\, t_{min} \\ &= \frac{s_1}{e^a - 1} - \frac{e^{as_1} - 1}{a\,(e^a - 1)} + \frac{s_1^2}{2} + s_1\, t_{min} \quad . \end{aligned} \tag{4.18}$$

Calculating I_3,

$$\begin{aligned} I_3 &= \frac{1 - s_2}{e^a - 1} - \left[\frac{1}{a\,(e^a - 1)} e^{as} \right]_{s_2}^{1} + \left[\frac{1}{2} s^2 \right]_{s_2}^{1} + (1 - s_2)\; t_{min} \\ &= \frac{1 - s_2}{e^a - 1} - \frac{e^a - e^{as_2}}{a\,(e^a - 1)} + \frac{1 - s_2^2}{2} + (1 - s_2)\; t_{min} \quad . \end{aligned} \tag{4.19}$$

And by (4.18) and (4.19) we know the sum

$$I_1 + I_3 = \frac{1 - (s_2 - s_1)}{e^a - 1} - \frac{e^{as_1} - e^{as_2} + e^a - 1}{a\,(e^a - 1)} + \frac{s_1^2 - s_2^2 + 1}{2} + \big(1 - (s_2 - s_1)\big)\; t_{min} \quad .$$

And since $s_2 - s_1 = \frac{1}{2}$, we have

$$I_1 + I_3 = \frac{\frac{1}{2}}{e^a - 1} - \frac{e^{as_1} - e^{as_2}}{a\,(e^a - 1)} - \frac{1}{a} + \frac{s_1^2 - s_2^2}{2} + \frac{1}{2} + \frac{1}{2} t_{min} \quad . \tag{4.20}$$

Finally, calculating I_2,

$$\begin{aligned}
I_2 &= \frac{-(s_2-s_1)}{e^a-1} + \left[\frac{1}{a\,(e^a-1)}e^{as}\right]_{s_1}^{s_2} - \left[\frac{1}{2}s^2\right]_{s_1}^{s_2} - (s_2-s_1)\,t_{min} \\
&= \frac{-(s_2-s_1)}{e^a-1} + \frac{e^{as_2}-e^{as_1}}{a\,(e^a-1)} - \frac{s_2^2-s_1^2}{2} - (s_2-s_1)\,t_{min} \\
&\overset{s_2-s_1=\frac{1}{2}}{=\!=} \frac{-\frac{1}{2}}{e^a-1} + \frac{e^{as_2}-e^{as_1}}{a\,(e^a-1)} - \frac{s_2^2-s_1^2}{2} - \frac{1}{2}\,t_{min} \quad . \qquad (4.21)
\end{aligned}$$

And since $d_{\mathbb{T}}(\eta_a, \lambda_{\mathbb{T}}) = I_1 + I_1 + I_3$, from (4.20) and (4.21) we get

$$\begin{aligned}
d_{\mathbb{T}}(\eta_a, \lambda_{\mathbb{T}}) &= 2\,\frac{e^{as_2}-e^{as_1}}{a\,(e^a-1)} - \frac{1}{a} - \left(s_2^2 - s_1^2\right) + \frac{1}{2} \\
&= 2\,\frac{e^{a\left(s_1+\frac{1}{2}\right)}-e^{as_1}}{a\,(e^a-1)} - \frac{1}{a} - \frac{1}{2}\left(2s_1 + \frac{1}{2}\right) + \frac{1}{2} \\
&= 2\,\frac{e^{as_1}\left(e^{\frac{a}{2}}-1\right)}{a\left(e^{\frac{a}{2}}-1\right)\left(e^{\frac{a}{2}}+1\right)} - \frac{1}{a} - s_1 - \frac{1}{4} + \frac{1}{2} \\
&\overset{(4.17)}{=} 2\,\frac{e^{\log\left(\cosh\left(\frac{a}{4}\right)\right)+\frac{a}{4}}}{a\left(e^{\frac{a}{2}}+1\right)} - \frac{1}{a} - \frac{1}{a}\log\left(\cosh\left(\frac{a}{4}\right)\right) - \frac{1}{4} + \frac{1}{4} \\
&= 2\,\frac{\cosh\left(\frac{a}{4}\right)e^{\frac{a}{4}}}{a\left(e^{\frac{a}{2}}+1\right)} - \frac{1}{a} - \frac{1}{a}\log\left(\cosh\left(\frac{a}{4}\right)\right) \quad .
\end{aligned}$$

Replacing the cosh in the numerator of the first term by its definition,

$$\begin{aligned}
d_{\mathbb{T}}(\eta_a, \lambda_{\mathbb{T}}) &= \frac{1}{a}\left(\frac{2\,\frac{e^{\frac{a}{4}}+e^{\frac{-a}{4}}}{2}\,e^{\frac{a}{4}}}{e^{\frac{a}{2}}+1} - 1\right) - \frac{1}{a}\log\left(\cosh\left(\frac{a}{4}\right)\right) \\
&= \frac{1}{a}\left(\frac{e^{\frac{2a}{4}}+1}{e^{\frac{a}{2}}+1} - 1\right) - \frac{1}{a}\log\left(\cosh\left(\frac{a}{4}\right)\right) \\
&= \frac{1}{-a}\log\left(\cosh\left(\frac{a}{4}\right)\right) \quad ,
\end{aligned}$$

as desired.

Case 2 $a > 0$

$$d_{\mathbb{T}}\left(\eta_a\,,\,\lambda_{\mathbb{T}}\right) = d_{\mathbb{T}}\left(\eta_a \circ Q^{-1}\,,\,\lambda_{\mathbb{T}} \circ Q^{-1}\right) = d_{\mathbb{T}}\Big(\eta_{-a}\,,\,\lambda_{\mathbb{T}}\Big) = \frac{1}{a}\log\left(\cosh\left(\frac{a}{4}\right)\right) \quad ,$$

where the three equalities are by Theorem 1.6.5, Remark 1.6.3, and Case 1, respectively. Thus the result holds in this case as well. ■

Theorem 4.2.6. *For every* $a, b \in \mathbb{R} \setminus \{0\}$ *such that* $a \neq b$,

$$d_{\mathbb{T}}(\eta_a, \eta_b) = \frac{|a-b|}{ab}\left(\frac{\left(\cosh\left(\frac{a}{4}\right)\right)^{\frac{b}{a-b}}}{\left(\cosh\left(\frac{b}{4}\right)\right)^{\frac{a}{a-b}}} - 1\right) \quad .$$

Proof. By Theorem 2.2.8, $d_{\mathbb{T}}(\eta_a, \eta_b) = \min_{t\in\mathbb{R}} \int_0^1 \left|F_{\eta_a} - F_{\eta_b} - t\right| ds$. By its convexity or concavity, F_{η_a} intersects the curve $F_{\eta_b} + t_{min}$ at two points, partitioning $[0, 1)$ into the interval between the points and the complement of the interval in $[0, 1)$ as explained at the end of Chapter 3. Let $\big(s_1\,, F_{\eta_a}(s_1)\big)$ and $\big(s_2\,, F_{\eta_a}(s_2)\big)$ denote the intersection points, where $s_1 < s_2$. By definition,

$$F_{\eta_a}(s_1) = F_{\eta_b}(s_1) + t_{min} \implies \frac{e^{as_1} - 1}{e^a - 1} = \frac{e^{bs_1} - 1}{e^b - 1} + t_{min} \quad ,$$
$$F_{\eta_a}(s_2) = F_{\eta_b}(s_2) + t_{min} \implies \frac{e^{as_2} - 1}{e^a - 1} = \frac{e^{bs_2} - 1}{e^b - 1} + t_{min} \quad .$$

In addition, we know that $s_2 - s_1 = \frac{1}{2}$ (see the end of Chapter 3). Thus we have three equations for the three unknowns s_1, s_2, t_{min}. Subtracting the first equation from the second yields

$$\frac{e^{as_2} \cancel{-1} - e^{as_1} \cancel{+1}}{e^a - 1} = \frac{e^{bs_2} \cancel{-1} - e^{bs_1} \cancel{+1}}{e^b - 1}$$
$$\implies \quad \frac{e^{a\left(s_1+\frac{1}{2}\right)} - e^{as_1}}{e^a - 1} = \frac{e^{b\left(s_1+\frac{1}{2}\right)} - e^{bs_1}}{e^b - 1}$$

$$\Longrightarrow \quad \frac{e^{as_1}\left(e^{\frac{a}{2}}-1\right)}{\left(e^{\frac{a}{2}}-1\right)\left(e^{\frac{a}{2}}+1\right)} = \frac{e^{bs_1}\left(e^{\frac{b}{2}}-1\right)}{\left(e^{\frac{b}{2}}-1\right)\left(e^{\frac{b}{2}}+1\right)}$$

$$\Longrightarrow \quad e^{(a-b)s_1} = \frac{e^{\frac{a}{2}}+1}{e^{\frac{b}{2}}+1} \quad .$$

$$\Longrightarrow \quad s_1 = \frac{1}{a-b}\log\left(\frac{e^{\frac{a}{2}}+1}{e^{\frac{b}{2}}+1}\right) \quad . \tag{4.22}$$

With these preparations, we are ready to calculate

$$d_{\mathbb{T}}(\eta_a,\eta_b) = \int_0^1 \left|\frac{e^{as}-1}{e^a-1} - \frac{e^{bs}-1}{e^b-1} - t_{min}\right| ds \quad .$$

Case 1 $a < b$

$$d_{\mathbb{T}}(\eta_a,\eta_b) = \underbrace{\int_0^{s_1}\left(\frac{e^{bs}-1}{e^b-1} + t_{min} - \frac{e^{as}-1}{e^a-1}\right) ds}_{I_1} + \underbrace{\int_{s_1}^{s_2}\left(\frac{e^{as}-1}{e^a-1} - \frac{e^{bs}-1}{e^b-1} - t_{min}\right) ds}_{I_2} + \underbrace{\int_{s_2}^{1}\left(\frac{e^{bs}-1}{e^b-1} + t_{min} - \frac{e^{as}-1}{e^a-1}\right) ds}_{I_3} \; .$$

Calculating I_1,

$$\begin{aligned} I_1 &= \left[\frac{1}{b\left(e^b-1\right)}e^{bs}\right]_0^{s_1} - \frac{s_1}{e^b-1} + s_1\, t_{min} - \left[\frac{1}{a\left(e^a-1\right)}e^{as}\right]_0^{s_1} + \frac{s_1}{e^a-1} \\ &= \frac{e^{bs_1}-1}{b\left(e^b-1\right)} - \frac{s_1}{e^b-1} + s_1\, t_{min} - \frac{e^{as_1}-1}{a\left(e^a-1\right)} + \frac{s_1}{e^a-1} \quad . \end{aligned} \tag{4.23}$$

Calculating I_3,

$$I_3 = \left[\frac{1}{b\left(e^b-1\right)}e^{bs}\right]_{s_2}^{1} - \frac{1-s_2}{e^b-1} + (1-s_2)\, t_{min} - \left[\frac{1}{a\left(e^a-1\right)}e^{as}\right]_{s_2}^{1} + \frac{1-s_2}{e^a-1}$$

$$= \frac{e^b - e^{bs_2}}{b\left(e^b-1\right)} - \frac{1-s_2}{e^b-1} + (1-s_2)\, t_{min} - \frac{e^a - e^{as_2}}{a\left(e^a-1\right)} + \frac{1-s_2}{e^a-1}$$

$$= \frac{e^b - e^{bs_2}}{b\left(e^b-1\right)} - \frac{\frac{1}{2}-s_1}{e^b-1} + \left(\frac{1}{2}-s_1\right) t_{min} - \frac{e^a - e^{as_2}}{a\left(e^a-1\right)} + \frac{\frac{1}{2}-s_1}{e^a-1} \quad . \qquad (4.24)$$

Thus by (4.23) and (4.24) we know

$$I_1 + I_3 = \frac{e^{bs_1} - 1 + e^b - e^{bs_2}}{b\left(e^b-1\right)} - \frac{\cancel{s_1} + \frac{1}{2} - \cancel{s_1}}{e^b-1} + \left(\cancel{s_1} + \frac{1}{2} - \cancel{s_1}\right) t_{min}$$
$$- \frac{e^{as_1} - 1 + e^a - e^{as_2}}{a\left(e^a-1\right)} + \frac{\cancel{s_1} + \frac{1}{2} - \cancel{s_1}}{e^a-1} \quad .$$

So we have shown that

$$I_1 + I_3 = \frac{e^{bs_1} - e^{bs_2} + e^b - 1}{b\left(e^b-1\right)} + \frac{-\frac{1}{2}}{e^b-1} + \left(\frac{1}{2}\right) t_{min}$$
$$- \frac{e^{as_1} - e^{as_2} + e^a - 1}{a\left(e^a-1\right)} + \frac{+\frac{1}{2}}{e^a-1} \quad . \qquad (4.25)$$

Finally, calculating I_2,

$$I_2 = \left[\frac{1}{a\left(e^a-1\right)} e^{as}\right]_{s_1}^{s_2} - \frac{s_2-s_1}{e^a-1} - \left[\frac{1}{b\left(e^b-1\right)} e^{bs}\right]_{s_1}^{s_2} + \frac{(s_2-s_1)}{e^b-1} - (s_2-s_1)\, t_{min}$$
$$= \frac{e^{as_2} - e^{as_1}}{a\left(e^a-1\right)} + \frac{\frac{-1}{2}}{e^a-1} + \frac{e^{bs_1} - e^{bs_2}}{b\left(e^b-1\right)} + \frac{\frac{1}{2}}{e^b-1} - \frac{1}{2} t_{min} \quad . \qquad (4.26)$$

Together (4.25) and (4.26) imply that

$$I_1 + I_2 + I_3 = \frac{2e^{bs_1} - 2e^{bs_2} + e^b - 1}{b\left(e^b-1\right)} + \frac{2e^{as_2} - 2e^{as_1} + 1 - e^a}{a\left(e^a-1\right)}$$
$$= -2\,\frac{e^{b\left(s_1+\frac{1}{2}\right)} - e^{bs_1}}{b\left(e^b-1\right)} + \frac{1}{b} + 2\,\frac{e^{a\left(s_1+\frac{1}{2}\right)} - e^{as_1}}{a\left(e^a-1\right)} - \frac{1}{a}$$
$$= -2\,\frac{e^{bs_1}\cancel{\left(e^{\frac{b}{2}}-1\right)}}{b\cancel{\left(e^{\frac{b}{2}}-1\right)}\left(e^{\frac{b}{2}}+1\right)} + 2\,\frac{e^{as_1}\cancel{\left(e^{\frac{a}{2}}-1\right)}}{a\cancel{\left(e^{\frac{a}{2}}-1\right)}\left(e^{\frac{a}{2}}+1\right)} + \frac{1}{b} - \frac{1}{a}$$

$$= 2\left(\frac{e^{as_1}}{a\left(e^{\frac{a}{2}}+1\right)} - \frac{e^{bs_1}}{b\left(e^{\frac{b}{2}}+1\right)}\right) + \frac{a-b}{ab}$$

$$\overset{(4.22)}{=} 2\left(\frac{e^{\frac{a}{a-b}\log\left(\frac{e^{\frac{a}{2}}+1}{e^{\frac{b}{2}}+1}\right)}}{a\left(e^{\frac{a}{2}}+1\right)} - \frac{e^{\frac{b}{a-b}\log\left(\frac{e^{\frac{a}{2}}+1}{e^{\frac{b}{2}}+1}\right)}}{b\left(e^{\frac{b}{2}}+1\right)}\right) + \frac{a-b}{ab}$$

$$= 2\left(\frac{\left(e^{\frac{a}{2}}+1\right)^{\frac{a}{a-b}-1}}{a\left(e^{\frac{b}{2}}+1\right)^{\frac{a}{a-b}}} - \frac{\left(e^{\frac{a}{2}}+1\right)^{\frac{b}{a-b}}}{b\left(e^{\frac{b}{2}}+1\right)^{\frac{b}{a-b}+1}}\right) + \frac{a-b}{ab}$$

$$= \frac{2\left(e^{\frac{a}{2}}+1\right)^{\frac{b}{a-b}}(b-a)}{a\,b\left(e^{\frac{b}{2}}+1\right)^{\frac{a}{a-b}}} + \frac{a-b}{ab}$$

$$= \frac{2\left(\left(\frac{e^{\frac{a}{4}}+e^{\frac{-a}{4}}}{2}\right)\left(\frac{2}{e^{\frac{-a}{4}}}\right)\right)^{\frac{b}{a-b}}(b-a)}{a\,b\left(\left(\frac{e^{\frac{b}{4}}+e^{\frac{-b}{4}}}{2}\right)\left(\frac{2}{e^{\frac{-b}{4}}}\right)\right)^{\frac{a}{a-b}}} + \frac{a-b}{ab} \quad .$$

Thus we have shown that

$$d_{\mathbb{T}}(\eta_a,\eta_b) = \frac{(b-a)\left(\cosh\left(\frac{a}{4}\right)\right)^{\frac{b}{a-b}}\cancel{2^{\frac{a}{a-b}}}\cancel{\frac{1}{\left(e^{\frac{-b}{4}}\right)^{\frac{a}{a-b}}}}}{a\,b\left(\cosh\left(\frac{b}{4}\right)\right)^{\frac{a}{a-b}}\cancel{2^{\frac{a}{a-b}}}\cancel{\frac{1}{\left(e^{\frac{-b}{4}}\right)^{\frac{a}{a-b}}}}} + \frac{a-b}{ab}$$

$$= \frac{b-a}{ab}\left(\frac{\left(\cosh\left(\frac{a}{4}\right)\right)^{\frac{b}{a-b}}}{\left(\cosh\left(\frac{b}{4}\right)\right)^{\frac{a}{a-b}}} - 1\right) \quad .$$

Case 2 $b < a$

By Theorem 1.6.5 and Remark 1.6.2, we know

$$d_{\mathbb{T}}(\eta_a,\eta_b) = d_{\mathbb{T}}(\eta_a \circ Q^{-1}, \eta_b \circ Q^{-1}) = d_{\mathbb{T}}(\eta_{-a},\eta_{-b}) \quad . \tag{4.27}$$

On the other hand, since $b < a$, we know that $-a < -b$. Therefore by Case 1, we know

$$d_{\mathbb{T}}(\eta_{-a}, \eta_{-b}) = \frac{(-b)-(-a)}{(-a)(-b))} \left(\frac{\left(\cosh\left(\frac{-a}{4}\right)\right)^{\frac{-b}{-a+b}}}{\left(\cosh\left(\frac{-b}{4}\right)\right)^{\frac{-a}{-a+b}}} - 1 \right)$$

$$= \frac{a-b}{ab} \left(\frac{\left(\cosh\left(\frac{a}{4}\right)\right)^{\frac{b}{a-b}}}{\left(\cosh\left(\frac{b}{4}\right)\right)^{\frac{a}{a-b}}} - 1 \right) \quad , \tag{4.28}$$

where the second equality is because cosh is an even function. Therefore (4.27) and (4.28) yield

$$d_{\mathbb{T}}(\eta_a, \eta_b) = \frac{a-b}{ab} \left(\frac{\left(\cosh\left(\frac{a}{4}\right)\right)^{\frac{b}{a-b}}}{\left(\cosh\left(\frac{b}{4}\right)\right)^{\frac{a}{a-b}}} - 1 \right) \quad ,$$

as desired. ■

Theorem 4.2.7. *For every $a \in \mathbb{R} \setminus \{0\}$ and $r \in [0,1)$,*

$$d_{\mathbb{T}}\left(\eta_a\,,\, \eta_a \circ R_r^{-1}\right) = \frac{2}{|a|} \log \frac{\cosh\left(\frac{a}{4}\right)}{\cosh\left(a\,\frac{1-2r}{4}\right)} \quad .$$

Proof. By Theorem 3.2.13, there is at least one intersection point between $F_{\eta_a \circ R_r^{-1}}$ and $F_{\eta_a} + t_{min}$. In fact, there are exactly two intersections, and one of these intersection points occurs on $[0, r)$ and the other on $[r, 1)$. This is easy to see in light of the fact that both $F_{\eta_a \circ R_r^{-1}}$ and F_{η_a} start at 0 and end at 1, but they go through this value change at different average speeds on the two intervals because of the swapped concavities (see Figure 1.5). More concretely, the rate of change of F_{η_a} is $\frac{a}{e^a-1}e^{as}$ on both intervals, but the rate of change of $F_{\eta_a \circ R_r^{-1}}$ is $\frac{a}{e^a-1}e^{a(1+s-r)}$ on $[0, r)$, and $\frac{a}{e^a-1}e^{a(s-r)}$ on $[r, 1)$.

Let $\left(s_1, F_{\eta_a \circ R_r^{-1}}(s_1)\right)$ and $\left(s_2, F_{\eta_a \circ R_r^{-1}}(s_2)\right)$ denote the intersection points of $F_{\eta_a \circ R_r^{-1}}$ and $F_{\eta_a} + t_{min}$. WLOG assume $s_1 \in [0, r)$ and $s_2 \in [r, 1)$. By definition,

$$F_{\eta_a}(s_1) + t_{min} = F_{\eta_a \circ R_r^{-1}}(s_1) \implies \frac{e^{as_1}-1}{e^a-1} + t_{min} = \frac{e^{a(1-r)}(e^{as_1}-1)}{e^a-1}\,; \tag{4.29}$$

$$F_{\eta_a}(s_2) + t_{min} = F_{\eta_a \circ R_r^{-1}}(s_2) \implies \frac{e^{as_2}-1}{e^a-1} + t_{min} = 1 - e^{-ar}\,\frac{e^a - e^{as_2}}{e^a-1}. \tag{4.30}$$

Equating the expressions for t_{min} from (4.29) and (4.30),

$$\begin{aligned} \frac{e^{a(1-r)}(e^{as_1}-1) - e^{as_1} + 1}{e^a-1} &= 1 + \frac{e^{-ar}e^{as_2} - e^{-ar}e^a - e^{as_2} + 1}{e^a-1} \\ e^{a(1-r+s_1)} - e^{a(1-r)} - e^{as_1} + 1 &= e^a \cancel{-1} + e^{a(s_2-r)} - e^{a(1-r)} - e^{as_2} \cancel{+1}\ . \end{aligned}$$

Using the fact that $s_2 - s_1 = \frac{1}{2}$,

$$e^{as_1}\left(e^{a(1-r)} - 1\right) + 1 - e^{a(1-r)} = e^{as_1}\left(e^{a(\frac{1}{2}-r)} - e^{\frac{a}{2}}\right) + e^a - e^{a(1-r)}\ .$$

Gathering all terms with the e^{as_1} factor,

$$e^{as_1}\left(e^{a(1-r)} - 1 - e^{a(\frac{1}{2}-r)} + e^{\frac{a}{2}}\right) = e^a - \cancel{e^{a(1-r)}} - 1 + \cancel{e^{a(1-r)}}\ .$$

Thus,

$$e^{as_1} = \frac{e^a-1}{e^{-ar}\left(e^a - e^{\frac{a}{2}}\right) + e^{\frac{a}{2}} - 1} = \frac{\cancel{\left(e^{\frac{a}{2}}-1\right)}\left(e^{\frac{a}{2}}+1\right)}{e^{-ar}e^{\frac{a}{2}}\cancel{\left(e^{\frac{a}{2}}-1\right)} + \cancel{e^{\frac{a}{2}}-1}}\ .$$

Taking the log of both sides,

$$\begin{aligned} as_1 &= \log\frac{e^{\frac{a}{2}}+1}{e^{a(\frac{a}{2}-r)}+1} = \log\left(\frac{e^{\frac{a}{4}} + e^{\frac{-a}{4}}}{e^{\frac{a}{4}(1-2r)} + e^{\frac{-a}{4}(1-2r)}} \times \frac{e^{\frac{-a}{4}(1-2r)}}{e^{\frac{-a}{4}}}\right) \\ &= \log\frac{\left(e^{\frac{a}{4}} + e^{\frac{-a}{4}}\right)\frac{1}{2}}{\left(e^{\frac{a}{4}(1-2r)} + e^{\frac{-a}{4}(1-2r)}\right)\frac{1}{2}} + \log e^{\frac{-a}{4}(1-2r)+\frac{a}{4}}\ . \end{aligned}$$

Therefore

$$s_1 = \frac{1}{a}\,\log\left(\frac{\cosh\left(\frac{a}{4}\right)}{\cosh\left(a\frac{1-2r}{4}\right)}\right) + \frac{r}{2} \quad . \tag{4.31}$$

By Theorem 2.2.8 ,

$$d_{\mathbb{T}}\left(\eta_a\,,\,\eta_a\circ R_r^{-1}\right) = \int_0^1 \left|F_{\eta_a\circ R_r^{-1}}(s) - F_{\eta_a}(s) - t_{min}\right|\,ds \quad .$$

Case 1 $a > 0$

Breaking our integral down into a partition,

$$\begin{aligned} &d_{\mathbb{T}}\left(\eta_a\,,\,\eta_a\circ R_r^{-1}\right) = \\ &\int_0^{s_1}\left(-F_{\eta_a\circ R_r^{-1}}(s) + F_{\eta_a}(s) + t_{min}\right)ds + \int_{s_1}^{s_2}\left(F_{\eta_a\circ R_r^{-1}}(s) - F_{\eta_a}(s) - t_{min}\right)ds \\ &\qquad + \int_{s_2}^{1}\left(-F_{\eta_a\circ R_r^{-1}}(s) + F_{\eta_a}(s) + t_{min}\right)ds \quad , \end{aligned}$$

where t_{min} terms can be canceled using $s_2 - s_1 = 1 - (s_2 - s_1) = \frac{1}{2}$.

Breaking the middle integral down further,

$$\begin{aligned} &d_{\mathbb{T}}\left(\eta_a\,,\,\eta_a\circ R_r^{-1}\right) = \\ &\underbrace{\int_0^{s_1}\left(-F_{\eta_a\circ R_r^{-1}}(s) + F_{\eta_a}(s) + t_{min}\right)ds}_{I_1} + \underbrace{\int_{s_1}^{r}\left(F_{\eta_a\circ R_r^{-1}}(s) - F_{\eta_a}(s) - t_{min}\right)ds}_{I_2} \\ &\underbrace{\int_r^{s_2}\left(F_{\eta_a\circ R_r^{-1}}(s) - F_{\eta_a}(s) - t_{min}\right)ds}_{I_3} + \underbrace{\int_{s_2}^{1}\left(-F_{\eta_a\circ R_r^{-1}}(s) + F_{\eta_a}(s) + t_{min}\right)ds}_{I_4} \,. \end{aligned}$$

Calculating I_1 ,

$$\begin{aligned}
I_1 &= \int_0^{s_1} \left(\frac{-e^{a(1-r)}(e^{as}-1)}{e^a-1} + \frac{e^{as}-1}{e^a-1} \right) ds \\
&= \int_0^{s_1} \left(\frac{-e^{a(1-r)}}{e^a-1} e^{as} + \frac{e^{a(1-r)}}{e^a-1} + \frac{1}{e^a-1} e^{as} - \frac{1}{e^a-1} \right) ds \\
&= \left[\frac{-e^{a(1-r)}}{a\left(e^a-1\right)} e^{as} \right]_0^{s_1} + \left[\frac{e^{a(1-r)}}{e^a-1}\, s \right]_0^{s_1} + \left[\frac{1}{a\left(e^a-1\right)} e^{as} \right]_0^{s_1} - \left[\frac{1}{e^a-1}\, s \right]_0^{s_1} \\
&= \frac{-e^{a(1-r)}\left(e^{as_1}-1\right)}{a\left(e^a-1\right)} + \frac{s_1\, e^{a(1-r)}}{e^a-1} + \frac{e^{as_1}-1}{a\left(e^a-1\right)} - \frac{s_1}{e^a-1} \\
&= \left(\frac{1-e^{a(1-r)}}{e^a-1} \right) \left(\frac{e^{as_1-1}}{a} - s_1 \right) \quad . \qquad (4.32)
\end{aligned}$$

Calculating I_2 ,

$$\begin{aligned}
I_2 &= \int_{s_1}^{r} \left(\frac{e^{a(1-r)}(e^{as}-1)}{e^a-1} - \frac{e^{as}-1}{e^a-1} \right) ds \\
&= \int_{s_1}^{r} \left(\frac{e^{a(1-r)}}{e^a-1} e^{as} - \frac{e^{a(1-r)}}{e^a-1} - \frac{1}{e^a-1} e^{as} + \frac{1}{e^a-1} \right) ds \\
&= \left[\frac{e^{a(1-r)}}{a\left(e^a-1\right)} e^{as} \right]_{s_1}^{r} - \left[\frac{e^{a(1-r)}}{e^a-1}\, s \right]_{s_1}^{r} - \left[\frac{1}{a\left(e^a-1\right)} e^{as} \right]_{s_1}^{r} + \left[\frac{1}{e^a-1}\, s \right]_{s_1}^{r} \\
&= \frac{e^{a(1-r)}\left(e^{ar}-e^{as_1}\right)}{a\left(e^a-1\right)} - \frac{\left(r-s_1\right)\, e^{a(1-r)}}{e^a-1} - \frac{e^{ar}-e^{as_1}}{a\left(e^a-1\right)} + \frac{r-s_1}{e^a-1} \\
&= \frac{1}{e^a-1} \left(\frac{e^a - e^{a(1-r+s_1)} - e^{ar} + e^{as_1}}{a} + (r-s_1)\left(1-e^{a(1-r)}\right) \right) \\
&= \frac{1}{e^a-1} \left(\frac{e^{ar}\left(e^{a-ar}-1\right) - e^{as_1}\left(e^{a(1-r)}-1\right)}{a} + (r-s_1)\left(1-e^{a(1-r)}\right) \right) \\
&= \frac{1-e^{a(1-r)}}{e^a-1} \left(\frac{-e^{ar}+e^{as_1}}{a} + r - s_1 \right) \quad . \qquad (4.33)
\end{aligned}$$

Calculating I_3,

$$
\begin{aligned}
I_3 &= \int_r^{s_2} \left(1 - e^{-ar}\,\frac{e^a - e^{as}}{e^a - 1} - \frac{e^{as} - 1}{e^a - 1}\right) ds \\
&= \left[1 - \frac{e^{a(1-r)}}{e^a - 1}\right]_r^{s_2} + \left[\frac{e^{-ar}}{a(e^a - 1)}\,e^{as}\right]_r^{s_2} - \left[\frac{1}{a(e^a - 1)}\,e^{as}\right]_r^{s_2} + \left[\frac{1}{e^a - 1}\right]_r^{s_2} \\
&= (s_2 - r)\left(1 - \frac{e^{a(1-r)}}{e^a - 1}\right) + \frac{(e^{-ar})(e^{as_2} - e^{ar})}{a(e^a - 1)} - \frac{e^{as_2} - e^{ar}}{a\,(e^a - 1)} + \frac{s_2 - r}{e^a - 1} \\
&= (s_2 - r)\left(1 - \frac{e^{a(1-r)} - 1}{e^a - 1}\right) + \frac{(e^{-ar} - 1)(e^{as_2} - e^{ar})}{a(e^a - 1)} \\
&= (s_2 - r)\left(\frac{e^a\left(1 - e^{-ar}\right)}{e^a - 1}\right) + \frac{(e^{-ar} - 1)(e^{as_2} - e^{ar})}{a(e^a - 1)} \\
&= \left(\frac{1 - e^{-ar}}{e^a - 1}\right)\left((s_2 - r)e^a + \frac{e^{ar} - e^{as_2}}{a}\right) \quad . \qquad (4.34)
\end{aligned}
$$

Calculating I_4,

$$
\begin{aligned}
I_4 &= \int_{s_2}^{1} \left(-1 + e^{-ar}\,\frac{e^a - e^{as}}{e^a - 1} + \frac{e^{as} - 1}{e^a - 1}\right) ds \\
&= \left[-1 + \frac{e^{a(1-r)}}{e^a - 1}\right]_{s_2}^{1} - \left[\frac{e^{-ar}}{a(e^a - 1)}\,e^{as}\right]_{s_2}^{1} + \left[\frac{1}{a(e^a - 1)}\,e^{as}\right]_{s_2}^{1} - \left[\frac{1}{e^a - 1}\right]_{s_2}^{1} \\
&= (1 - s_2)\left(-1 + \frac{e^{a(1-r)}}{e^a - 1}\right) - \frac{(e^{-ar})(e^a - e^{as_2})}{a(e^a - 1)} + \frac{e^a - e^{as_2}}{a(e^a - 1)} - \frac{s_2 - 1}{e^a - 1} \\
&= s_2 - 1 + \frac{(s_2 - 1)\left(1 - e^{a(1-r)}\right)}{e^a - 1} + \frac{(e^a - e^{as_2})\left(1 - e^{-ar}\right)}{a(e^a - 1)} \quad . \qquad (4.35)
\end{aligned}
$$

By (4.32) and (4.33) we have

$$
I_1 + I_2 = \frac{1 - e^{a(1-r)}}{e^a - 1}\left(\frac{2e^{as_1} - 1 - e^{ar}}{a} - 2s_1 + r\right) \quad .
$$

By (4.34) and (4.35) we have

$$I_3 + I_4 \;=$$

$$\frac{1-e^{-ar}}{e^a-1}\left((s_2-r)e^a + \frac{e^a - e^{as_2} + e^{ar} - e^{as_2}}{a}\right) + \frac{(s_2-1)\left(1-e^{a(1-r)}\right)}{e^a-1} + s_2 - 1\,.$$

And therefore, since $d_{\mathbb{T}}\left(\eta_a\,,\,\eta_a\circ R_r^{-1}\right) = I_1 + I_2 + I_3 + I_4$,

$$\begin{aligned} d_{\mathbb{T}}\left(\eta_a\,,\,\eta_a\circ R_r^{-1}\right) = &\\ &\frac{1-e^{a(1-r)}}{e^a-1}\left(\frac{2e^{as_1}-1-e^{ar}}{a} - s_1 + r - \frac{1}{2}\right)\\ &+ \frac{1-e^{-ar}}{e^a-1}\left((s_2-r)e^a + \frac{-2e^{as_2}+e^a+e^{ar}}{a}\right) + s_2 - 1 \quad . \end{aligned}$$

Multiplying the numerator of each term into the parentheses,

$$\begin{aligned} d_{\mathbb{T}}\left(\eta_a\,,\,\eta_a\circ R_r^{-1}\right) = &\\ &\frac{1}{e^a-1}\Bigg(\frac{2e^{as_1}-1-e^{ar}-2e^{a(1-r+s_1)}+e^{a(1-r)}+e^a}{a}\\ &\qquad - e^{a(1-r)}\Big(-s_1+r-\frac{1}{2}\Big) - s_1 + r - \frac{1}{2}\Bigg)\\ &+ \frac{1}{e^a-1}\Bigg(\frac{-2e^{as_2}+e^a+e^{ar}+2e^{a(s_2-r)}-e^{a(1-r)}-1}{a}\\ &\qquad + (s_2-r)e^a - (s_2-r)e^{a(1-r)}\Bigg) + s_2 - 1 \quad . \end{aligned}$$

Factoring $\frac{1}{e^a-1}$ out, and replacing s_2 with $s_1+\frac{1}{2}$, we get

$$\begin{aligned} d_{\mathbb{T}}\left(\eta_a\,,\,\eta_a\circ R_r^{-1}\right) = &\\ &\frac{1}{e^a-1}\Bigg(\frac{2e^{as_1}-2e^{a(s_1+\frac{1}{2})}+2e^a+2e^{a(s_1-r+\frac{1}{2})}-2e^{a(s_1-r+1)}-2}{a}\\ &\qquad + \Big(s_1+\frac{1}{2}-r\Big)e^a - s_1 + r - \frac{1}{2}\Bigg) + s_1 - \frac{1}{2} \quad . \end{aligned}$$

By factorizing further,

$$d_{\mathbb{T}}\left(\eta_a\,,\,\eta_a\circ R_r^{-1}\right)=$$
$$\frac{1}{e^a-1}\left(\frac{2}{a}\left(e^{as_1}(1-e^{\frac{a}{2}})+e^{a(s_1-r)}(e^{\frac{a}{2}}-e^a)+e^a-1\right)+\left(s_1+\frac{1}{2}-r\right)(e^a-1)\right)+s_1-\frac{1}{2}\ .$$

Multiplying the $\frac{1}{e^a-1}$ factor back inside the parentheses,

$$d_{\mathbb{T}}\left(\eta_a\,,\,\eta_a\circ R_r^{-1}\right)=$$
$$\frac{2}{a}\left(\frac{e^{as_1}\cancel{(1-e^{\frac{a}{2}})}^{-1}}{\cancel{(e^{\frac{a}{2}}-1)}(e^{\frac{a}{2}}+1)}+\frac{e^{a(s_1-r)}e^{\frac{a}{2}}\cancel{(1-e^{\frac{a}{2}})}^{-1}}{\cancel{(e^{\frac{a}{2}}-1)}(e^{\frac{a}{2}}+1)}+1\right)+s_1+\cancel{\frac{1}{2}}-r+s_1-\cancel{\frac{1}{2}}\ .$$

Factoring out $e^{a(s_1-r)}$ from the two fractions in the parentheses as well as commuting them with the $+1$,

$$d_{\mathbb{T}}\left(\eta_a\,,\,\eta_a\circ R_r^{-1}\right)\;=\;\frac{2}{a}\left(1-e^{a(s_1-r)}\,\frac{e^{ar}+e^{\frac{a}{2}}}{(e^{\frac{a}{2}}+1)}\right)+2s_1-r\ . \tag{4.36}$$

We now prove that the expression inside the parentheses in (4.36) equates to 0, and thereby demonstrate that $d_{\mathbb{T}}\left(\eta_a\,,\,\eta_a\circ R_r^{-1}\right)=2s_1-r$. In other words, we show that $e^{a(s_1-r)}\,\frac{e^{ar}+e^{\frac{a}{2}}}{(e^{\frac{a}{2}}+1)}=1$.

We know by (4.31) that

$$\begin{aligned}e^{a(s_1-r)}\,\frac{e^{ar}+e^{\frac{a}{2}}}{(e^{\frac{a}{2}}+1)}&\overset{(4.31)}{=}\left(\frac{\cosh\left(\frac{a}{4}\right)}{\cosh\left(a\frac{1-2r}{4}\right)}\right)e^{\frac{-ar}{2}}\,\frac{e^{ar}+e^{\frac{a}{2}}}{(e^{\frac{a}{2}}+1)}\\&=\left(\frac{\cosh\left(\frac{a}{4}\right)}{\cosh\left(a\frac{1-2r}{4}\right)}\right)\frac{e^{\frac{ar}{2}}+e^{\frac{a(1-r)}{2}}}{(e^{\frac{a}{2}}+1)}\ .\end{aligned}$$

Multiplying both the numerator and the denominator by $e^{\frac{-a}{4}}$,

$$e^{a(s_1-r)}\,\frac{e^{ar}+e^{\frac{a}{2}}}{(e^{\frac{a}{2}}+1)}\;=\;\left(\frac{\cosh\left(\frac{a}{4}\right)}{\cosh\left(a\frac{1-2r}{4}\right)}\right)\frac{e^{\frac{2ar-a}{4}}+e^{\frac{2a(1-r)-a}{4}}}{(e^{\frac{a}{4}}+e^{\frac{-a}{4}})}\ .$$

Factoring out $\frac{a}{4}$ in the powers of the numerator,

$$\begin{aligned} e^{a(s_1-r)}\,\frac{e^{ar}+e^{\frac{a}{2}}}{(e^{\frac{a}{2}}+1)} &= \left(\frac{\cosh\left(\frac{a}{4}\right)}{\cosh\left(a\frac{1-2r}{4}\right)}\right)\frac{e^{\frac{a}{4}(2r-1)}+e^{\frac{a}{4}(2-2r-1)}}{(e^{\frac{a}{4}}+e^{\frac{-a}{4}})} \\ &= \left(\frac{\cosh\left(\frac{a}{4}\right)}{\cosh\left(a\frac{1-2r}{4}\right)}\right)\left(\frac{\cosh\left(a\frac{1-2r}{4}\right)}{\cosh\left(\frac{a}{4}\right)}\right) = 1 \quad ; \end{aligned}$$

which implies through (4.36) that

$$\begin{aligned} d_{\mathbb{T}}\left(\eta_a\,,\,\eta_a\circ R_r^{-1}\right) &= 2s_1 - r \\ &\overset{(4.31)}{=} 2\left(\frac{1}{a}\log\left(\frac{\cosh\left(\frac{a}{4}\right)}{\cosh\left(a\frac{1-2r}{4}\right)}\right)+\frac{r}{2}\right)-r \\ &= \frac{2}{a}\log\left(\frac{\cosh\left(\frac{a}{4}\right)}{\cosh\left(a\frac{1-2r}{4}\right)}\right) \quad . \end{aligned}$$

∎

Theorem 4.2.8. *For every* $a\in\mathbb{R}\setminus\{0\}$ *and* $s,t\in\mathbb{R}$*,*

$$\frac{4\log\cosh\left(\frac{a}{4}\right)}{|a|}\,d_{\mathbb{T}}(s,t) \le d_{\mathbb{T}}\left(\eta_a\circ R_s^{-1}\,,\,\eta_a\circ R_t^{-1}\right) \le \left|\tanh\left(\frac{a}{4}\right)\right| d_{\mathbb{T}}(s,t) \quad .$$

Proof. Let arbitrary $s,t\in\mathbb{R}$ be given. We know through Theorem 1.6.5 and Remark 1.6.2 that

$$\begin{aligned} d_{\mathbb{T}}\left(\eta_a\circ R_s^{-1},\eta_a\circ R_t^{-1}\right) &= d_{\mathbb{T}}\left(\eta_a\circ R_s^{-1}\circ\left(R_t^{-1}\right)^{-1},\eta_a\circ R_t^{-1}\circ\left(R_t^{-1}\right)^{-1}\right) \\ &= d_{\mathbb{T}}\left(\eta_a\circ R_{s-t}^{-1}\,,\,\eta_a\right) \quad . \end{aligned}$$

Letting $r:=s-t$, it suffices to prove this theorem only for when one measure is η_a rotated by r, and the other is not rotated. WLOG assume $r\in[0,1)$.

Claim 4.2.8.1.

$$d_{\mathbb{T}}\left(\eta_a\,,\,\eta_a\circ R_r^{-1}\right) \le \left|\tanh\left(\frac{a}{4}\right)\right| d_{\mathbb{T}}(0,r) \quad .$$

Proof. For every $r \in [0,1]$, let $d(r) := d_{\mathbb{T}}\left(\eta_a\,,\, \eta_a \circ R_r^{-1}\right)$. By Theorem 4.2.7,

$$d(r) = \frac{2}{|a|}\,\log\Big(\cosh\big(\frac{a}{4}\big)\Big) - \frac{2}{a}\,\log\Big(\cosh\big(a\,\frac{1-2r}{4}\big)\Big) \quad .$$

Clearly, $d(0) = d(1) = 0$. As a function on $[0,1]$, d is differentiable everywhere, and its derivative is

$$d'(r) \;=\; \frac{2}{a}\,(-2)\,\frac{a}{4}\,\sinh\big(\frac{a(1-2r)}{4}\big)\,\frac{-1}{\cosh\big(\frac{a(1-2r)}{4}\big)} \;=\; \tanh\big(\frac{a(1-2r)}{4}\big) \quad .$$

Clearly, d' is continuous, and by Remark 2.1.1 has a maximum. To find this maximum, note that $d'\big(\frac{1}{2}\big) = 0$ and thus through Remark 4.1.2 we know $|d'|$ is symmetric about $\frac{1}{2}$. Remark 4.1.2 also tells us that d' is strictly decreasing (if $a > 0$) or decreasing (if $a < 0$). Therefore

$$\max_{s\in[0,1]}\big|d'(s)\big| = \big|d'(0)\big| = \big|d'(1)\big| = \Big|\tanh\big(\frac{a}{4}\big)\Big| \quad .$$

Consequently, through the mean value theorem we know

$$d(r) - d(0) \leq \Big|\tanh\big(\frac{a}{4}\big)\Big|\,|r-0| \quad .$$

In particular, this implies that

$$\forall r \in [0,\frac{1}{2}]\,,\quad d_{\mathbb{T}}\left(\eta_a\,,\,\eta_a\circ R_r^{-1}\right) \;\leq\; \Big|\tanh\big(\frac{a}{4}\big)\Big|\,d_{\mathbb{T}}(0,r) \quad . \tag{4.37}$$

The mean value theorem also tell us that

$$d(r) - d(1) \leq \Big|\tanh\big(\frac{a}{4}\big)\Big|\,|r-1| \quad .$$

In particular, this implies that

$$\forall r \in [\frac{1}{2},1]\,,\quad d_{\mathbb{T}}\left(\eta_a\,,\,\eta_a\circ R_r^{-1}\right) \;\leq\; \Big|\tanh\big(\frac{a}{4}\big)\Big|\,d_{\mathbb{T}}(0,r) \quad . \tag{4.38}$$

Thus (4.37) and (4.38) prove the claim. □

Claim 4.2.8.2.
$$\frac{4\log\cosh\left(\frac{a}{4}\right)}{|a|}\, d_{\mathbb{T}}\,(0,r)\;\leq\; d_{\mathbb{T}}\left(\eta_a\,,\,\eta_a\circ R_r^{-1}\right)\quad .$$

Proof. Let d be as in the proof of Claim 4.2.8.1. By Remark 4.1.2, d is concave (if $a>0$) or convex (if $a<0$). Consider the secant passing through $\begin{bmatrix} 0 \\ d(0) \end{bmatrix}$ and $\begin{bmatrix} \frac{1}{2} \\ d(\frac{1}{2}) \end{bmatrix}$. By concavity,

$$\forall r\in[0,\frac{1}{2}]\,,\quad 2d(\frac{1}{2})\,r\leq d(r)\quad .$$

In particular, this implies that

$$\forall r\in[0,\frac{1}{2}]\,,\quad \frac{4}{|a|}\,\log\cosh\left(\frac{a}{4}\right)d_{\mathbb{T}}\,(0,r)\leq d_{\mathbb{T}}\left(\eta_a\,,\,\eta_a\circ R_r^{-1}\right)\quad .\tag{4.39}$$

Similarly, consider the secant passing through $\begin{bmatrix} \frac{1}{2} \\ d(\frac{1}{2}) \end{bmatrix}$ and $\begin{bmatrix} 1 \\ d(1) \end{bmatrix}$. By concavity,

$$\forall r\in[\frac{1}{2},1]\,,\quad -2d(\frac{1}{2})\,(r-1)\leq d(r)\quad .$$

In particular, this implies that

$$\forall r\in[\frac{1}{2},1]\,,\quad \frac{4}{|a|}\,\log\cosh\left(\frac{a}{4}\right)d_{\mathbb{T}}\,(0,r)\leq d_{\mathbb{T}}\left(\eta_a\,,\,\eta_a\circ R_r^{-1}\right)\quad .\tag{4.40}$$

Thus (4.39) and (4.40) prove the claim. □

The truth of the theorem has thus been shown through Claims 4.2.8.1 and 4.2.8.2. (The argument above shows that the left and right inequality becomes an equality when $d_{\mathbb{T}}(s,t)=\frac{1}{2}$ and $d_{\mathbb{T}}(s,t)\to 0$, respectively.) ■

Recall the distribution function $F_{\eta_a}(s)=\frac{e^{as}-1}{e^a-1}$ (see Figure 1.2). While not defined for $a=0$, one can see that as a approaches 0, the continuous functions F_{η_a} approach (monotonically from below if $a\to 0^+$, and from above if $a\to 0^-$)

the continuous distribution function $F_{\lambda_{\mathbb{T}}}(s) = s$ pointwise:

$$\forall s \in [0,1], \qquad \lim_{a\to 0} \frac{e^{as}-1}{e^{a}-1} \overset{\text{L'Hospital}}{=\!=} \lim_{a\to 0} \frac{s\,e^{as}}{e^{a}} = \lim_{a\to 0} s\,(e^{a})^{s-1} = s \quad .$$

One therefore suspects the Kantorovich distance between η_a and $\lambda_{\mathbb{T}}$ to go to 0 as well, and indeed it does (see Theorem 4.2.9). In fact, one could use Dini's theorem to show that the convergence is uniform [12]. As a result, we are justified to define $\eta_0 := \lambda_{\mathbb{T}}$.

Theorem 4.2.9. *The exponential distribution η_a converges (in $\mathcal{P}$) to $\lambda_{\mathbb{T}}$ as $a \to 0$, i.e.,*

$$d_{\mathbb{T}}(\eta_a, \lambda_{\mathbb{T}}) \xrightarrow{a\to 0} 0 \quad .$$

Proof. By Theorem 4.2.5,

$$\lim_{a\to 0} d_{\mathbb{T}}(\eta_a, \lambda_{\mathbb{T}}) = \lim_{a\to 0} \frac{1}{a} \log\Big(\cosh\big(\frac{a}{4}\big)\Big) = \lim_{a\to 0} \frac{\frac{1}{4}\sinh\left(\frac{a}{4}\right)}{\cosh\left(\frac{a}{4}\right)} = 0 \quad ,$$

where the second equality is by L'Hospital's rule. ∎

Additionally, we can see pointwise convergence of F_{η_a} to δ_0 as $a \to +\infty$.

$$\text{For } s = 1, \qquad \lim_{a\to+\infty} \frac{e^{as}-1}{e^{a}-1}\bigg|_{s=1} = \lim_{a\to+\infty} 1 = 1 \quad ,$$

$$\forall s \in [0,1), \qquad \lim_{a\to+\infty} \frac{e^{as}-1}{e^{a}-1} = \lim_{a\to+\infty} \frac{e^{as}}{e^{a}} = \lim_{a\to+\infty} (e^{a})^{s-1} \overset{s-1<0}{=\!=} 0 \quad ,$$

which implies a jump at 1; but since the points 0 and 1 correspond to the same point on our circle, we recognize this point mass distribution as δ_0 in $\mathcal{P}$. Through analogous steps we see the pointwise convergence of F_{η_a} to δ_0 as $a \to -\infty$. We therefore define $\eta_{-\infty} := \eta_{+\infty} := \delta_0$ as justified by Theorem 4.2.10. We lastly agree that $\eta_{\frac{1}{0}} = \eta_{\infty} := \delta_0$.

Theorem 4.2.10. *The exponential distribution η_a converges (in $\mathcal{P}$) to δ_0 as $a \to \pm\infty$, i.e.,*

$$d_{\mathbb{T}}(\eta_a, \delta_0) \xrightarrow{|a|\to\infty} 0 \quad .$$

Proof. By Theorem 4.2.4 we see

$$\lim_{|a|\to\infty} d_{\mathbb{T}}(\eta_a, \delta_0) = \lim_{|a|\to\infty} \frac{1}{a} \tanh\left(\frac{a}{4}\right) = \lim_{|a|\to\infty} \frac{1}{|a|} = 0 \quad ,$$

where the second equality is by the fact that $|\tanh|$ is bounded by 1. ■

Note that while originally η_a was only defined for any $a \in \mathbb{R}\setminus\{0\}$, we now have it defined for any $a \in \mathbb{R} \cup \{\infty\}$, where the symbol ∞ represents both $\pm\infty$ as one point since $\eta_{-\infty} = \eta_{+\infty}$. Noting how $F_{\eta_a}(s) := \frac{e^{as}-1}{e^{a}-1}$ is continuous with respect to $a \in \mathbb{R}\setminus\{0\}$, and how we defined η_0 and η_∞ by taking limits, we can see that the map $a \mapsto \eta_a$ is not only a bijection between $\mathbb{R} \cup \{\infty\}$ and $E := \left\{\eta_a \in \mathcal{P} : a \in \mathbb{R} \cup \{\infty\}\right\}$, but also a topological isomorphism if $\mathbb{R} \cup \{\infty\}$ has the one-point compactification topology. On the other hand, as explained at the end of Section 4.1, the one-point compactification of $\mathbb{R}$ is topologically isomorphic to $\mathbb{T}$. So we conclude that the set E of all unrotated mod 1 exponential distributions forms a circle in $\mathcal{P}$ that on one side passes through $\mathcal{P}$'s center, namely $\lambda_{\mathbb{T}}$, and on the other passes through δ_0. This set is depicted in blue in Figure 4.4. Note that E is closed under reflection but not under rotation. For any $t \in \mathbb{R}$, the set $\left\{\eta_a \circ R_t^{-1} : a \in \mathbb{R} \cup \{\infty\}\right\}$ too is a circle, and passes through $\lambda_{\mathbb{T}}$ and $\delta_{t+\mathbb{Z}}$.

Figure 4.4 pieces together the results of this chapter: Recall that Theorem 4.2.1 tells us that the space $(\mathcal{P}, d_{\mathbb{T}})$ can be thought of as a disc or a ball of radius 1/4 centered at $\lambda_{\mathbb{T}}$. Theorem 4.2.2 states that the outermost edge of this disc is the set $\{\delta_x : x \in \mathbb{T}\}$ of all Dirac distributions, which itself forms a Lipschitz-isomorphic copy of $\mathbb{T}$ by Theorem 1.5.3. This circle is depicted in Figure 4.4 in black. Note that through Theorem 4.2.8 we know that for every $a \in \mathbb{R} \setminus \{0\}$, the mapping $s + \mathbb{Z} \mapsto \eta_a \circ R_s^{-1}$ is bi-Lipschitz continuous. As a result, for every $a \in \mathbb{R} \setminus \{0\}$, $\left\{\eta_a \circ R_s^{-1} : s \in \mathbb{R}\right\}$ is a Lipschitz-isomorphic copy of the circle $\mathbb{T}$ inside $\mathcal{P}$. The rotation-invariance of $d_{\mathbb{T}}$ implies that the center of this circle must be $\lambda_{\mathbb{T}}$. One such circle is depicted in red in Figure 4.4. Theorem 4.2.9 implies that the closer a is to 0, the smaller the radius of the circle $\left\{\eta_a \circ R_s^{-1} : s \in \mathbb{R}\right\}$. This radius approaches $\frac{1}{4}$ as $a \to \infty$ by Theorem 4.2.10.

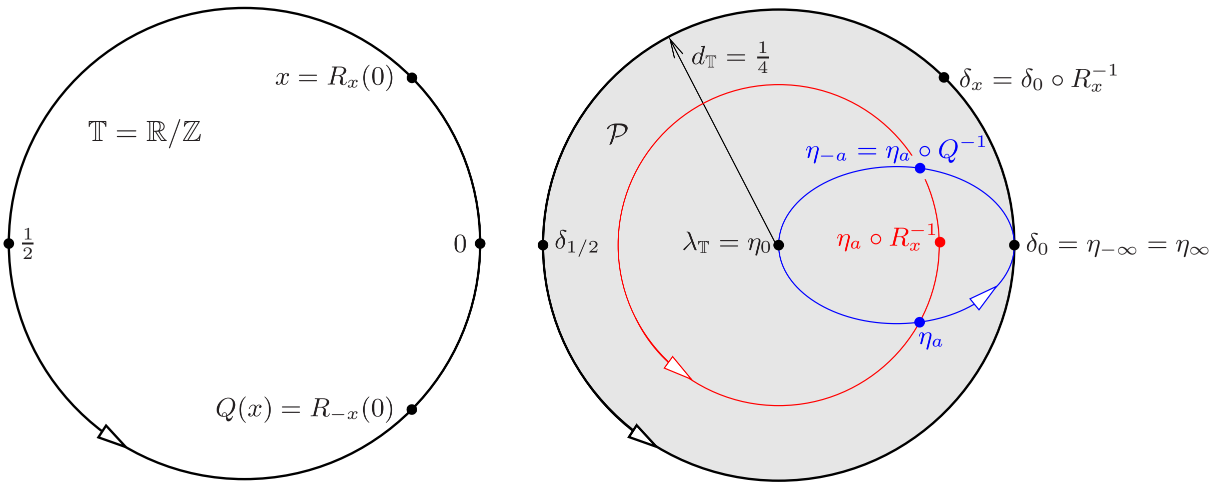

Figure 4.4: Schematic depictions of $\mathbb{T}$ (left), and $\mathcal{P}$ (right). The latter is a ball of radius $\frac{1}{4}$ with $\lambda_{\mathbb{T}}$ at the center. For a fixed nonzero a, the set $\{\eta_a \circ R_s^{-1} : s \in \mathbb{R}\}$ (red) is Lipschitz-isomorphic to $\mathbb{T}$.

We will learn in Chapter 5 that the empirical distributions of $(\log n)_{n=1}^{\infty}$ have the circular limit set $\{\eta_1 \circ R_s^{-1} : s \in \mathbb{R}\}$.

Chapter 5

Empirical Distributions of Slow-varying Sequences

In light of the questions raised in the Motivation, this chapter demonstrates that for any sequence that grows to infinity at the pace of $(\log n)_{n=1}^{\infty}$, the sequence of suitably rotated empirical distributions converges to an exponential distribution in $(\mathcal{P}, d_{\mathbb{T}})$. Theorem 5.2.10 provides an upper bound for the $d_{\mathbb{T}}$ distance between the rotated empirical distributions and the exponential limit. Theorem 5.2.10 also applies to sequences that grow slower than $(\log n)_{n=1}^{\infty}$. We will even be able to get information about the speed of the mentioned convergence using this theorem. These tools are then used to explain the patterns observed in the Motivation.

5.1 Preparatory work

Definition 5.1.1 (Asymptotic equivalence of sequences)**.** Let $(t_n)_{n=1}^{\infty}$ and $(r_n)_{n=1}^{\infty}$ be sequences in $\mathbb{R}^+ \setminus \{0\}$. We say $(t_n)_{n=1}^{\infty}$ and $(r_n)_{n=1}^{\infty}$ are *asymptotically equivalent*, in symbols $t_n \sim r_n$, if

$$\lim_{n\to\infty} \frac{t_n}{r_n} = 1 \quad .$$

Definition 5.1.2 (Asymptotic order of decay of sequences)**.** Let $(t_n)_{n=1}^{\infty}$ and $(r_n)_{n=1}^{\infty}$ be positive real sequences that converge to 0 . We say $(t_n)_{n=1}^{\infty}$ *decays* *no slower than* (or *is of the order*) $(r_n)_{n=1}^{\infty}$, in symbols $t_n = O\,(\,r_n\,)$, if

$$\limsup_{n\to\infty} \frac{t_n}{r_n} < \infty \quad .$$

To numerically approximate the integral of a function w.r.t. Lebesgue measure on an interval, one can sum the values of that function at certain sample points in the interval. It is natural to ask what the difference between the sum and the integral is. The Euler-Maclaurin summation formula answers this question, and is also useful for approximating a sum by an integral [2]. The latter is what we will use the formula for in Lemma 5.2.9.

Remark 5.1.3 (Euler-Maclaurin summation formula)**.** [3] Let $g \in C^1[t_1, t_2]$ where $t_1 < t_2$. Then,

$$\sum_{n\in\mathbb{Z}\cap[t_1,t_2]} g(n) \;=\; \int_{t_1}^{t_2} g(t)\,dt \;+\; \int_{t_1}^{t_2} \Big(\langle t\rangle - \frac{1}{2}\Big) g'(t)\,dt \;+\; \frac{g(t_1)\;+\;g(t_2)}{2} \quad .$$

One other tool used in Theorem 5.2.10 is the variation of parameters method for solving a linear differential equation.

Remark 5.1.4 (First-order variation of parameters)**.** [16] Consider the initial value problem

$$\begin{cases} g' + \alpha g = G(t) \quad , \\ g(t_0) = g_0 \quad , \end{cases}$$

where α is a constant, and $G\colon [t_0, +\infty) \to \mathbb{R}$ is continuous. The unique solution to this problem is

$$g(t) = \left(g_0 + \int_{t_0}^{t} e^{\alpha r} G(r)\,dr \right) e^{-\alpha t} \qquad ; t \ge t_0 \quad .$$

For the main convergence result to follow from Theorem 5.2.10, we will use the fact that if the value of a locally integrable function on $\mathbb{R}^+$ converges to a real number as the input tends to $+\infty$, then the average value of that function over increasingly larger intervals converges to the same real number.

Lemma 5.1.5. *Let* $g \in L^1_{\text{loc}}\left(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \lambda\right)$. *If* $\lim_{t\to\infty} g(t) = a$ *for some* $a \in \mathbb{R}$, *then*

$$\frac{1}{T}\int_0^T g(t)\,dt \xrightarrow{T\to\infty} a \quad .$$

Proof. Assume $g(t) \xrightarrow{t\to\infty} a \in \mathbb{R}$, i.e., assume

$$\forall \epsilon_1 > 0,\ \exists\, M_1 \in \mathbb{R}^+ : \forall t \geq M_1,\ \left|g(t) - a\right| < \epsilon_1 \quad .$$

We want to show that $\forall \epsilon > 0,\ \exists\, M \in \mathbb{R}^+ : \forall\, T \geq M,\ \left|\frac{1}{T}\int_0^T g(t)\,dt - a\right| < \epsilon$. Let an arbitrary $\epsilon > 0$ be given, and let $\epsilon_1 := \frac{\epsilon}{2}$. By assumption,

$$\exists\, M_1 > 0 :\ \forall t \geq M_1,\ \left|g(t) - a\right| < \frac{\epsilon}{2} \quad . \tag{5.1}$$

Note that since g is locally integrable, $\int_0^{M_1} \left|g(t) - a\right|\,dt$ is finite. Thus,

$$\exists\, M_2 > 0 :\ \forall\, T \geq M_2,\ \frac{\int_0^{M_1} \left|g(t) - a\right|\,dt}{T} < \frac{\epsilon}{2} \quad . \tag{5.2}$$

Let $M := \max\{M_1, M_2\}$. For every $T \geq M + 1$,

$$\begin{aligned}
\left|\frac{1}{T}\int_0^T g(t)\,dt - a\right| &= \left|\frac{\int_0^T g(t)\,dt - aT}{T}\right| = \left|\frac{\int_0^T g(t)\,dt - \int_0^T a\,dt}{T}\right| \\
&= \frac{\left|\int_0^T \left(g(t) - a\right)dt\right|}{T} \leq \frac{\int_0^T \left|g(t) - a\right|dt}{T} \\
&= \frac{1}{T}\int_0^{M_1} \left|g(t) - a\right|dt + \frac{1}{T}\int_{M_1}^T \left|g(t) - a\right|dt \\
&\leq \frac{\epsilon}{2} + \frac{1}{T}\int_{M_1}^T \frac{\epsilon}{2}\,dt \\
&= \frac{\epsilon}{2} + \frac{\epsilon}{2} - \frac{\epsilon M_1}{2T} < \epsilon \quad .
\end{aligned}$$

Thus we have shown that $\lim_{T\to\infty} \frac{1}{T}\int_0^T g(t)\,dt = a$. ■

In Example 5.2.13 we use the following useful fact:

Lemma 5.1.6. *The logarithmic integral function is given by* $\mathrm{Li}(x) := \int_e^x \frac{1}{\log t}\,dt$ *for every* $x \in [e,\infty)$ *. This function is asymptotically equivalent to* $\frac{x}{\log x}$ *, i.e.,*

$$\lim_{x\to\infty} \frac{\int_e^x \frac{1}{\log t}\,dt}{\frac{x}{\log x}} = 1 \quad .$$

Proof. By the L'Hospital rule, we have

$$\lim_{x\to\infty} \frac{\int_e^x \frac{1}{\log t}\,dt}{\frac{x}{\log x}} = \lim_{x\to\infty} \frac{\frac{1}{\log x}}{\frac{\log x - \frac{x}{x}}{\log(x)^2}} = \lim_{x\to\infty} \frac{\log x}{\log x - 1} = 1 \quad .$$

■

Remark 5.1.7. Let $(\epsilon_n)_{n=1}^{\infty}$ be a sequence in $\mathbb{R} \setminus \{0\}$ that converges to 0. Then $\big(\log(1+\epsilon_n)\big)_{n=1}^{\infty}$ converges to 0 at the same rate, i.e.,

$$\log(1+\epsilon_n) \,\sim\, \epsilon_n \quad .$$

5.2 Convergent subsequences of empirical distributions

This section describes the limit set of the empirical distributions in $(\mathcal{P}, d_{\mathbb{T}})$ associated with slow-varying real sequences mod 1; it also provides an estimate of the speed of convergence to those limits.

Definition 5.2.1 (Empirical distribution)**.** Let $x = (x_n)_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$. For every $N \in \mathbb{N}$, we define the associated (mod 1) empirical distribution to be

$$\omega_N^x := \frac{1}{N}\sum_{n=1}^{N} \delta_{x_n+\mathbb{Z}} \quad ,$$

where $\delta_{x_n+\mathbb{Z}}$ is as defined in Section 1.4. When the sequence with which the empirical distribution is associated is clear from the context, we simplify the notation to ω_N.

We call a sequence $(x_n)_{n=1}^{\infty}$ in $\mathbb{R}$ *log-like* if the distance between two consecutive elements decreases asymptotically equivalent to a constant multiple of $\left(\frac{1}{n}\right)_{n=1}^{\infty}$, i.e., if

$$\lim_{n\to\infty} n(x_n - x_{n-1}) \in \mathbb{R}\setminus\{0\} \quad . \tag{5.3}$$

Such a sequence will necessarily satisfy $\lim_{n\to\infty} x_n = \pm\infty$. Clearly, if the limit in (5.3) is 0, then the sequence's rate of change is eventually slower than that of $(\log n)_{n=1}^{\infty}$. The main results of this chapter (Theorem 5.2.10 and its implications) apply to these sequences as well as log-like ones. Hence we define *slow-varying* sequences as follows.

Definition 5.2.2 (Slow-varying sequence)**.** A sequence $x = (x_n)_{n=1}^{\infty}$ in $\mathbb{R}$ is *slow-varying* if it satisfies

$$\lim_{n\to\infty} n(x_n - x_{n-1}) \in \mathbb{R} \ . \tag{5.4}$$

The limit in (5.4) is denoted b^x.

As mentioned in the Motivation, the sequence $(\omega_N)_{N=1}^{\infty}$ of empirical distributions for a log-like sequence does not converge in $\mathcal{P}$ in the sense of (1.4). However, Theorem 5.2.10 will tell us that if every ω_N is rotated by $-x_N$, we will have a convergent sequence. We denote this rotated version of every ω_N by $\widetilde{\omega}_N$.

Definition 5.2.3 (Suitably rotated empirical distribution)**.** Let $x = (x_n)_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$. For every empirical distribution ω_N^x, we define its suitably rotated version to be

$$\widetilde{\omega}_N^x := \omega_N^x \circ R_{-x_N}^{-1} \quad .$$

As with Definition 5.2.1, we simplify the notation to $\widetilde{\omega}_N$ when the sequence is clear from the context.

Given a log-like sequence, in Theorem 5.2.16 we will characterize the limit set of that sequence. For log-like sequences these limits end up being accumulation points of the set $\{x_n\}_{n=1}^{\infty}$. For this reason, we denote the limit set by $\mathbb{A}$.

Definition 5.2.4. Let $x = (x_n)_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$. The set of limits of all convergent subsequences of x in $\mathbb{T}$ is

$$\mathbb{A}^x := \left\{ y \in \mathbb{T} \,:\, d_{\mathbb{T}}\left(x_{\phi(n)}, y\right) \xrightarrow{n\to\infty} 0 \text{ for a strictly increasing } \phi\colon \mathbb{N} \to \mathbb{N} \right\} .$$

We simplify the notation to $\mathbb{A}$ when there is no ambiguity about the sequence.

For a sequence of empirical distributions, we define the limits of convergent subsequences in $\mathcal{P}$ similarly.

Definition 5.2.5. Let $x = (x_n)_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$, and let $(\omega_N)_{N=1}^{\infty}$ be the sequence of the associated empirical distributions. Then

$$\Omega^x := \left\{ \mu \in \mathcal{P} \,:\, d_{\mathbb{T}}\left(\omega_{\phi(N)}, \mu\right) \xrightarrow{n\to\infty} 0 \text{ for a strictly increasing } \phi\colon \mathbb{N} \to \mathbb{N} \right\} .$$

We simplify the notation to Ω when there is no ambiguity about the sequence.

A sequence is in essence a function on $\mathbb{N}$. Thus restricting the domain of any function on $\mathbb{R}^+ \setminus \{0\}$ to $\mathbb{N}$ specifies a sequence, and clearly every sequence can be generated that way. The advantage of considering these *functions* as opposed to *sequences* is the ability to use the Euler-Mclaurin summation formula (see Remark 5.1.3). This summation formula is a central part of Theorem 5.2.10. Thus henceforth, instead of a slow-varying sequence x, we consider a function $f\colon \mathbb{R}^+ \to \mathbb{R}$ that is a smooth interpolant for the graph $\left(\left[\begin{smallmatrix} n \\ x_n \end{smallmatrix}\right]\right)_{n=1}^{\infty}$ of x, and satisfies $\lim_{t\to\infty} t\, f'(t) = b^x$. We denote the collection of such functions by $\mathcal{F}$, thus,

$$\mathcal{F} := \left\{ f \in C^{\infty}\left(\mathbb{R}^+ \setminus \{0\}\right) \,:\, \lim_{t\to\infty} t\, f'(t) \text{ exists in } \mathbb{R} \right\} . \tag{5.5}$$

Hence for any given $f \in \mathcal{F}$, we re-define the relevant quantities consistent with their discrete counterparts:

$$b^f := \lim_{t\to\infty} t\, f'(t) \;;$$

$$\mathbb{A}^f := \left\{ x \in \mathbb{T} \,:\, d_{\mathbb{T}}\Big(f\left(\phi(n)\right), x\Big) \xrightarrow{n\to\infty} 0 \text{ for a strictly increasing } \phi\colon \mathbb{N} \to \mathbb{N} \right\} \;;$$

$$\omega_N^f := \frac{1}{N}\sum_{n=1}^{N}\delta_{f(n)+\mathbb{Z}} \quad \wedge \quad \widetilde{\omega}_N^f := \omega_N^f \circ R_{-f(N)}^{-1} \quad \forall N \in \mathbb{N} \ ;$$
$$\Omega^f := \left\{\mu \in \mathcal{P} \ : \ d_{\mathbb{T}}\big(\omega_{\phi(N)}^f, \mu\big) \xrightarrow{n\to\infty} 0 \text{ for a strictly increasing } \phi\colon \mathbb{N}\to\mathbb{N}\right\} \ .$$

As before, we simplify the notation by omitting the superscript f whenever the function in question is clear from the context. Note that since $d_{\mathbb{T}}$ is invariant under reflection (Theorem 1.6.5) we have $\mathbb{A}^{-f} = Q(\mathbb{A}^f)$. Next, we denote the subset of functions that are asymptotically slower than log by $\mathcal{F}_0$, i.e.,

$$\mathcal{F}_0 := \left\{f \in \mathcal{F} \ : \ b^f = 0\right\} \ .$$

Thus the set of all log-like functions is $\mathcal{F} \setminus \mathcal{F}_0$. It is convenient for the proof of Theorem 5.2.10 to introduce two more quantities for any given $f \in \mathcal{F}$. The first is the smooth function $\Delta^f\colon \mathbb{R}^+ \setminus \{0\} \to \mathbb{R}$ given by

$$\Delta^f(t) := tf'(t) - b^f \quad \forall t \in \mathbb{R}^+ \setminus \{0\} \ . \tag{5.6}$$

It is clear from the definition of $\mathcal{F}$ that $\lim_{t\to+\infty}\Delta^f(t) = 0$ for every $f \in \mathcal{F}$. For a given $f \in \mathcal{F}$, one interpretation of $\Delta^f(t)$ is that it tells us how different the slope of f is compared to that of $b^f \log t$. Note how $\Delta^g(t) = 0$ for all $t \in \mathbb{R}^+ \setminus \{0\}$ if $g(t) := b^f \log t$. So in a sense, $\Delta^f(t)$ measures how 'far' $f(t)$ is from $b^f \log t$ at every $t \in \mathbb{R}^+$.

Lemma 5.2.6. *Let $f \in \mathcal{F}$. Then for every $m, t \in \mathbb{R}^+ \setminus \{0\}$,*

$$f(t) = f(m) + b\log t - b\log m + \int_m^t \frac{\Delta(s)}{s}\, ds \ .$$

Proof. Let arbitrary $m, t \in \mathbb{R}^+$ be given. By (5.6),

$$f'(t) = \frac{\Delta(t) + b}{t}$$

Integrating both sides from m to t yields through the Fundamental Theorem of Calculus,

$$f(t) - f(m) = \int_m^t \frac{\Delta(s)}{s}\, ds + b\log t - b\log m \quad ,$$

which is rearranged to show the desired result. ■

Lemma 5.2.7. *Let $f \in \mathcal{F}$. Then*

$$\lim_{N\to\infty} \frac{\int_1^N \frac{|\Delta(s)|}{s}\, ds}{\log N} = 0 \quad .$$

Proof. Let arbitrary ϵ be given. Since $\lim_{t\to\infty}\Delta(t) = 0$, there exists $N_\epsilon \in \mathbb{N}$ such that $|\Delta(t)| < \frac{\epsilon}{2}$ for every $N \geq N_\epsilon$. Clearly,

$$\int_1^N \frac{|\Delta(s)|}{s}\, ds \leq \int_1^{N_\epsilon} \frac{\Delta(s)}{s}\, ds + \int_{N_\epsilon}^N \frac{\epsilon}{2s}\, ds = \int_1^{N_\epsilon} \frac{\Delta(s)}{s}\, ds - \frac{\epsilon}{2}\log N_\epsilon + \frac{\epsilon}{2}\log N \quad .$$

Note that since Δ is continuous, $\int_1^{N_\epsilon} \frac{\Delta(s)}{s}\, ds$ and thereby $\int_1^{N_\epsilon} \frac{\Delta(s)}{s}\, ds - \frac{\epsilon}{2}\log N_\epsilon$ is bounded. After division by $\log N$ we find

$$0 \leq \lim_{N\to\infty} \frac{\int_1^N \frac{|\Delta(s)|}{s}\, ds}{\log N} \leq \frac{\epsilon}{2} < \epsilon \quad .$$

Since ϵ was arbitrary, the limit must equal 0. ■

Lemma 5.2.7 states that $\log N$ eventually grows faster than $\int_1^N \frac{|\Delta(s)|}{s}\, ds$. Needless to say, this means that any constant multiple of $\int_1^N \frac{|\Delta(s)|}{s}\, ds$ or its sum with any constant still grows eventually slower than $\log N$.

The other quantity we define is D_N^f for every $N \in \mathbb{N}$:

$$D_N^f := d_{\mathbb{T}}\left(\widetilde{\omega}_N^f\,,\, \eta_{\frac{1}{bf}}\right) = d_{\mathbb{T}}\left(\omega_N^f\,,\, \eta_{\frac{1}{bf}} \circ R_{f(N)}^{-1}\right) \quad ,$$

which is the distance between $\eta_{\frac{1}{bf}}$ and the suitably rotated empirical distribution of the first N elements of $\big(f(n)\big)_{n=1}^\infty$.

In Theorem 5.2.10 we introduce a sequence of (eventual) upper bounds for each of these distances. Since the upper bounds converge to 0, as a corollary it becomes apparent that the distances D_N^f also approach 0.

Lemma 5.2.8. *Let* $f \in \mathcal{F}$ *be such that* $b > 0$. *Then for every* $h \in \mathrm{Lip}_1(\mathbb{T})$ *and every* $N \in \mathbb{N}$,

$$\int_{\mathbb{T}} h\, d\big(\eta_{\frac{1}{b}} \circ R_{f(N)}^{-1}\big) = h \circ f(N) - \int_{-\infty}^{f(N)} h'(s)\, e^{\frac{s-f(N)}{b}}\, ds \quad .$$

Proof. Let arbitrary $h \in \mathrm{Lip}_1(\mathbb{T})$ and $N \in \mathbb{N}$ be given. Then

$$\begin{aligned}\int_{\mathbb{T}} h\, d\big(\eta_{\frac{1}{b}} \circ R_{f(N)}^{-1}\big) &= \int_{R_{f(N)}^{-1}(\mathbb{T})} h \circ R_{f(N)}\, d\big(\eta_{\frac{1}{b}} \circ R_{f(N)}^{-1} \circ R_{f(N)}\big) = \int_{\mathbb{T}} h \circ R_{f(N)}\, d\eta_{\frac{1}{b}} \\ &= \int_{\mathbb{T}} h \circ R_{f(N)}\, f_{\eta_{\frac{1}{b}}}\, d\lambda_{\mathbb{T}} = \int_{\iota_{\mathbb{R}}(\mathbb{T})} \big(h \circ R_{f(N)} \circ \iota_{\mathbb{R}}^{-1}\big)\, \big(f_{\eta_{\frac{1}{b}}} \circ \iota_{\mathbb{R}}^{-1}\big)\, d\lambda\ ,\end{aligned}$$

where $f_{\eta_{\frac{1}{b}}}$ is the Radon-Nikodym derivative $\frac{d\eta_{1/b}}{d\lambda_{\mathbb{T}}}$, and by Remark 1.4.1 equals $F'_{\eta_{\frac{1}{b}}}$. Thus

$$\begin{aligned}\int_{\mathbb{T}} h\, d\big(\eta_{\frac{1}{b}} \circ R_{f(N)}^{-1}\big) &= \int_0^1 h\left(s + f(N)\right) \frac{e^{s/b}}{b\left(e^{1/b} - 1\right)}\, ds \\ &= \int_0^1 h\left(s + f(N)\right) \frac{e^{s/b}}{b} \sum_{j=1}^{\infty} \left(e^{-1/b}\right)^j ds \\ &= \sum_{j=1}^{\infty} \int_{0-j}^{1-j} h\left(s + j + f(N)\right) \frac{e^{\frac{s+j}{b}}}{b}\, e^{\frac{-j}{b}}\, ds \\ &= \sum_{j=1}^{\infty} \int_{-j}^{1-j} h\left(s + f(N)\right) \frac{e^{\frac{s}{b}}}{b}\, ds = \int_{-\infty}^{0} h\left(s + f(N)\right) \frac{e^{\frac{s}{b}}}{b}\, ds\ ,\end{aligned}$$

where the geometric series formula is true by positivity of b, the third equality is by dominated convergence, and the penultimate equality is by recalling from Section 1.3 that by $h\left(s + j + f(N)\right)$ we mean $h\left(s + j + f(N) + \mathbb{Z}\right)$.

Therefore

$$\begin{aligned}\int_{\mathbb{T}} h\, d\big(\eta_{\frac{1}{b}} \circ R^{-1}_{f(N)}\big) &= \int_{-\infty}^{f(N)} h\,(s)\, \frac{e^{\frac{s-f(N)}{b}}}{b}\, ds \\ &= \Big[h(s)\, e^{\frac{s-f(N)}{b}}\Big]_{-\infty}^{f(N)} - \int_{-\infty}^{f(N)} h'(s)\, e^{\frac{s-f(N)}{b}}\, ds \\ &= h\big(f(N)\big) - \int_{-\infty}^{f(N)} h'(s)\, e^{\frac{s-f(N)}{b}}\, ds \quad ,\end{aligned}$$

where the middle equality is because Remark 2.1.6 allows us to integrate by parts, and the last equality is by the fact that h is bounded. ■

Lemma 5.2.9. *Let $f \in \mathcal{F}$. Then for every $h \in \mathrm{Lip}_1(\mathbb{T})$ and every $N_1, N \in \mathbb{N}$ such that $N_1 \leq N$,*

$$\begin{aligned}N \int_{\mathbb{T}} h\, d\omega_N = N\, h \circ f(N) &+ \int_{f(N_1)}^{f(N)} \Big(\langle f^{-1}(s)\rangle - \frac{1}{2}\Big) h'(s)\, ds \\ &- \int_{f(N_1)}^{f(N)} f^{-1}(s) h'(s)\, ds + a_N \quad ,\end{aligned}$$

where $a_N = \sum_{n=1}^{N_1-1} \big(h \circ f(n) - h \circ f(N_1)\big) + \frac{1}{2}\big(h \circ f(N) - h \circ f(N_1)\big)$.

Proof. Let arbitrary $h \in \mathrm{Lip}_1(\mathbb{T})$ and $N_1, N \in \mathbb{N}$ be given where $N_1 \leq N$. Then

$$\begin{aligned}N \int_{\mathbb{T}} h\, d\omega_N &= \sum_{n=1}^{N_1-1} h \circ f(n) + \sum_{n=N_1}^{N} h \circ f(n) \\ &= \sum_{n=1}^{N_1-1} h \circ f(n) + \int_{N_1}^{N} h \circ f(t)\, dt + \int_{N_1}^{N} \Big(\langle t\rangle - \frac{1}{2}\Big)\, (h \circ f)'\,(t)\, dt \\ &\quad + \frac{h \circ f(N_1) + h \circ f(N)}{2} \quad ,\end{aligned}$$

where the second equality is by Remark 5.1.3. Since f' is continuous and $f'(t) \xrightarrow{t\to\infty} 0$, we know that f, and thereby $h \circ f$, are Lipschitz on $[N_1, +\infty)$.

Thus Remark 2.1.6 allows us to integrate the first integral by parts and get

$$N\int_{\mathbb{T}} h\,d\omega_N = \sum_{n=1}^{N_1-1} h\circ f(n) + \big[t\,h\circ f(t)\big]_{N_1}^{N} - \int_{N_1}^{N} t\,h'\circ f(t)\,f'(t)\,dt$$
$$+\int_{N_1}^{N}\Big(\langle t\rangle - \frac{1}{2}\Big)h'\circ f\,f'(t)(t)\,dt + \frac{h\circ f(N_1)+h\circ f(N)}{2}\ ;$$

And by rearranging terms and using the substitution $s = f(t)$, we get

$$N\int_{\mathbb{T}} h\,d\omega_N = \sum_{n=1}^{N_1-1} h\circ f(n) - N_1\,h\circ f(N_1) + \frac{h\circ f(N_1)+h\circ f(N)}{2}$$
$$+\,N\,h\circ f(N) + \int_{f(N_1)}^{f(N)}\Big(\langle f^{-1}(s)\rangle - \frac{1}{2}\Big)h'(s)\,ds - \int_{f(N_1)}^{f(N)} f^{-1}(s)h'(s)\,ds\,;$$

Thus we have shown

$$N\int_{\mathbb{T}} h\,d\omega_N = N\,h\circ f(N) + \int_{f(N_1)}^{f(N)}\Big(\langle f^{-1}(s)\rangle - \frac{1}{2}\Big)h'(s)\,ds$$
$$-\int_{f(N_1)}^{f(N)} f^{-1}(s)h'(s)\,ds + a_N\quad ,$$

where $a_N = \sum_{n=1}^{N_1-1}\big(h\circ f(n) - h\circ f(N_1)\big) + \frac{1}{2}\big(h\circ f(N) - h\circ f(N_1)\big)$. ■

The following theorem is the main result in this chapter.

Theorem 5.2.10. *For every $f\in\mathcal{F}$, there exists $N_0\in\mathbb{N}$ such that*

$$\forall N\geq N_0\,,\qquad D_N^f \leq \frac{1}{4N} + \frac{\left|b^f\right|\log N}{N} + \frac{2}{N}\int_1^N \left|\Delta^f(t)\right|dt\quad .$$

Proof. Fix $f\in\mathcal{F}$. When the superscript in Δ^f, b^f, and ω_N^f is omitted it is understood to refer to that fixed function. For a fixed $N\in\mathbb{N}$, we want to find an upper bound for

$$\left\{\int_{\mathbb{T}} h\,d\omega_N - \int_{\mathbb{T}} h\,d\big(\eta_{\frac{1}{b}}\circ R_{f(N)}^{-1}\big)\ :\ h\in \mathrm{Lip}_1(\mathbb{T})\right\}\quad .$$

Let an arbitrary $h \in \mathrm{Lip}_1(\mathbb{T})$ be given.

Case 1 $b > 0$

Choose $N_1 \in \mathbb{N}$ such that $\left|\Delta(t)\right| \leq \frac{b}{4}$ for every $t \geq N_1$. Since $f'(t) > 0$ for every $t \geq N_1$, we know that f^{-1} is defined on $[f(N_1), +\infty)$. Let $g(t)$ be defined to be $f^{-1}(t)$ for all $t \geq f(N_1)$, and the constant N_1 for all $t \leq f(N_1)$. By Lemmas 5.2.8 and 5.2.9,

$$
\begin{aligned}
-N\int_{\mathbb{T}} h\, d\big(\eta_{\frac{1}{b}} \circ R^{-1}_{f(N)}\big) + N\int_{\mathbb{T}} h\, d\omega_N = \\
N\int_{-\infty}^{f(N_1)} h'(s)\, e^{\frac{s-f(N)}{b}}\, ds + N\int_{f(N_1)}^{f(N)} h'(s)\, e^{\frac{s-f(N_1)}{b}}\, ds \\
-\int_{f(N_1)}^{f(N)} h'(s)g(s)\, ds + \int_{f(N_1)}^{f(N)} \Big(\langle g(s)\rangle - \frac{1}{2}\Big)h'(s)\, ds + a_N\ .
\end{aligned}
$$

Therefore

$$
\begin{aligned}
N\left(\int_{\mathbb{T}} h\, d\omega_N - \int_{\mathbb{T}} h\, d\big(\eta_{\frac{1}{b}} \circ R^{-1}_{f(N)}\big)\right) = \underbrace{b_N\int_{-\infty}^{f(N_1)} h'(s)\, e^{\frac{s}{b}}\, ds}_{I_1} + \underbrace{\int_{f(N_1)}^{f(N)} h'(s) g_N(s)\, ds}_{I_2} \\
+ \underbrace{\int_{f(N_1)}^{f(N)} \Big(\langle g(s)\rangle - \frac{1}{2}\Big)h'(s)\, ds + a_N}_{I_3}\ , \quad (5.7)
\end{aligned}
$$

where $b_N := Ne^{\frac{-f(N)}{b}}$, and g_N is a smooth real function on $[f(N_1), f(N)]$ given by $g_N(s) := b_N\, e^{\frac{s}{b}} - g(s)$ for every $N \geq N_1$. We now find an upper bound for each of the labeled terms in (5.7).

(i) Firstly, for I_1, by Lemma 2.2.5, $\left|\int_{-\infty}^{f(N_1)} h'(s)\, e^{\frac{s}{b}}\, ds\right|$ is bounded by $b\, e^{\frac{f(N_1)}{b}}$. Therefore $|I_1|$ is bounded by some constant multiple of b_N, namely $|b_N|\, b\, e^{\frac{f(N_1)}{b}}$. Considering b_N, note that by Lemma 5.2.6,

$$
\begin{aligned}
b_N = N\, e^{\frac{1}{b}\left(-f(1) - \int_1^N \frac{\Delta(s)}{s}\, ds\right)} e^{\log(1) - \log(N)} = e^{\frac{-f(1)}{b}} e^{\frac{-1}{b}\int_1^N \frac{\Delta(s)}{s}\, ds} \\
\leq e^{\frac{-f(1)}{b}}\, e^{\frac{1}{b}\int_1^N \frac{|\Delta(s)|}{s}\, ds}\ . \qquad (5.8)
\end{aligned}
$$

Clearly, if $\int_1^N \frac{|\Delta(s)|}{s}\, ds < \infty$, then (5.8) implies that b_N is bounded too, and thus $\lim_{N\to\infty} \frac{b_N}{\log N} = 0$. That is, in this subcase, clearly $\log N$ eventually grows faster than any constant multiple of b_N. In the subcase where $\int_1^N \frac{|\Delta(s)|}{s}\, ds = \infty$, we know that both the upper bound in (5.8) and $\int_1^N |\Delta(s)|\, ds$ approach ∞ as N grows. We show that $\lim_{N\to\infty} \frac{b_N}{\int_1^N |\Delta(s)|\, ds} = 0$ in this subcase. By (5.8) and the L'Hospital rule we have,

$$
\begin{aligned}
0 \le \lim_{N\to\infty} \frac{b_N}{\int_1^N |\Delta(s)|\, ds} &\overset{(5.8)}{\le} \lim_{N\to\infty} \frac{e^{\frac{-f(1)}{b}}\, e^{\frac{1}{b}\int_1^N \frac{|\Delta(s)|}{s}\, ds}}{\int_1^N |\Delta(s)|\, ds} \\
&= \lim_{N\to\infty} \frac{e^{\frac{-f(1)}{b}}\, e^{\frac{1}{b}\int_1^N \frac{|\Delta(s)|}{s}\, ds}\, \frac{\cancel{|\Delta(N)|}}{b\,N}}{\cancel{|\Delta(N)|}} \\
&= \frac{1}{b} e^{\frac{-f(1)}{b}} \lim_{N\to\infty} e^{\frac{1}{b}\int_1^N \frac{|\Delta(s)|}{s}\, ds}\, e^{-\log N} \\
&\le \frac{1}{b} e^{\frac{f(1)}{b}}\, e^{\lim_{N\to\infty}\left(\frac{1}{b}\int_1^N \frac{|\Delta(s)|}{s}\, ds - \log N\right)} = 0 \quad ,
\end{aligned}
$$

where the last equality is because Lemma 5.2.7 tells us the limit in the exponent is $-\infty$. Thus we have shown that in this subcase, $\int_1^N |\Delta(s)|\, ds$ eventually grows faster than any constant multiple of b_N. Therefore it is true in either subcase that there exists $N_2 \in \mathbb{N}$ such that

$$
|I_1| \le \frac{b}{3}\log N + \frac{1}{3}\int_1^N |\Delta(s)|\, ds \qquad \forall N \ge N_2 \quad . \tag{5.9}
$$

(ii) Secondly, for I_2, noting that $g_N(s) = b_N\, e^{\frac{s}{b}} - g(s)$, we have

$$
g_N'(s) = \frac{b_N}{b} e^{\frac{s}{b}} - \frac{1}{f'\big(g(s)\big)} \quad ,
$$

where the last term is because $g = f^{-1}$ for every $s \ge f(N_1)$. Therefore we know by (5.6) that for every $s \ge f(N_1)$,

$$
\begin{aligned}
g_N'(s) &= \frac{b_N}{b} e^{\frac{s}{b}} - \frac{g(s)}{\Delta \circ g(s) + b} \\
&= \frac{\big(\Delta \circ g(s) + b\big)\, b_N e^{\frac{s}{b}} - b\, g(s)}{b\,\big(\Delta \circ g(s) + b\big)} + \frac{g(s)\Delta \circ g(s) - g(s)\Delta \circ g(s)}{b\,\big(\Delta \circ g(s) + b\big)}
\end{aligned}
$$

$$= \frac{b_N\, e^{\frac{s}{b}} - g(s)}{b} + \frac{g(s)\Delta \circ g(s)}{b\left(\Delta \circ g(s) + b\right)} = \frac{g_N(s)}{b} + \frac{g(s)\Delta \circ g(s)}{b\left(\Delta \circ g(s) + b\right)} \quad ,$$

which, together with the fact that $g_N\left(f(N)\right) = 0$ poses a first-order initial value problem. Since the right hand side of the last equation above is continuous and its derivative w.r.t. g_N is also continuous, we have a unique solution which, by Remark 5.1.4, is:

$$g_N(s) = -\int_s^{f(N)} e^{\frac{s-r}{b}} \frac{g(r)\Delta \circ g(r)}{b\left(\Delta \circ g(s) + b\right)}\, dr \qquad \forall s \in [f(N_1), f(N)] \quad .$$

Substituting this formula for g_N in I_2, we get

$$\begin{aligned}
|I_2| &= \left| -\int_{f(N_1)}^{f(N)} \int_s^{f(N)} h'(s) e^{\frac{s-r}{b}} \frac{g(r)\Delta \circ g(r)}{b\left(\Delta \circ g(s) + b\right)}\, dr\, ds \right| \\
&= \left| -\int_{f(N_1)}^{f(N)} \frac{g(r)\Delta \circ g(r)}{b\left(\Delta \circ g(s) + b\right)} \int_{f(N_1)}^{r} h'(s) e^{\frac{s-r}{b}}\, ds\, dr \right| \\
&\le \int_{f(N_1)}^{f(N)} \left| \frac{g(r)\Delta \circ g(r)}{b\left(\Delta \circ g(s) + b\right)} \right| \left| \int_{f(N_1)}^{r} h'(s) e^{\frac{s-r}{b}} \right| ds\, dr \quad ,
\end{aligned}$$

where the second equality is by reversing the order of integration. Therefore, using the fact that $g = f^{-1}$ and that $\left|\Delta(r)\right| < \frac{b}{4}$ for every $r \ge N_1$,

$$\begin{aligned}
|I_2| &\le \int_{f(N_1)}^{f(N)} \frac{g(r)\, |\Delta| \circ g(r)}{b - |\Delta| \circ g(s)} \int_{f(N_1)}^{r} \left| h'(s) \frac{1}{b} e^{\frac{s-r}{b}} \right| ds\, dr \\
&\le \int_{f(N_1)}^{f(N)} \frac{g(r)\, |\Delta| \circ g(r)}{b - |\Delta| \circ g(s)} \int_{f(N_1)}^{r} \frac{1}{b} e^{\frac{s-r}{b}}\, ds\, dr \\
&\le \int_{f(N_1)}^{f(N)} \frac{g(r)\, |\Delta| \circ g(r)}{b - |\Delta| \circ g(s)}\, dr = \int_{N_1}^{N} \left|\Delta(t)\right| \frac{\Delta(t) + b}{b - \left|\Delta(t)\right|}\, dt \quad ,
\end{aligned}$$

where the second inequality is by Lemma 2.1.7 and the fact that $\frac{1}{b} e^{\frac{s-r}{b}} > 0$, and the last equality is by the substitution $t = g(r)$. Therefore since $\left|\Delta(t)\right| <$

$\frac{b}{4}$ for $t \geq N_1$, we have

$$|I_2| \leq \frac{b+\frac{b}{4}}{b-\frac{b}{4}} \int_{N_1}^{N} \big|\Delta(t)\big| \, dt = \frac{5}{3} \int_{N_1}^{N} \big|\Delta(t)\big| \, dt \quad . \tag{5.10}$$

(iii) Finally, to bound I_3, note that by the triangle inequality,

$$\begin{aligned} |a_N| &\leq \sum_{n=1}^{N_1-1} \Big| h\left(f(n)\right) - h\left(f(N_1)\right) \Big| + \frac{1}{2} \Big| h\left(f(N)\right) - h\left(f(N_1)\right) \Big| \\ &\leq \sum_{n=1}^{N_1-1} d_{\mathbb{T}}\left(f(n), f(N_1)\right) + \frac{1}{2} d_{\mathbb{T}}\left(f(N), f(N_1)\right) \\ &\leq \sum_{n=1}^{N_1-1} \frac{1}{2} + \frac{1}{2}\Big(\frac{1}{2}\Big) = \frac{N_1-1}{2} + \frac{\frac{1}{2}}{2} = \frac{N_1 - \frac{1}{2}}{2} \quad , \end{aligned} \tag{5.11}$$

where the second inequality is because $h \in \mathrm{Lip}_1(\mathbb{T})$, and the third inequality is by Definition 1.3.1. Thus $(a_N)_{N=N_1}^{\infty}$ is bounded. Note that

$$\begin{aligned} |I_3| &\leq \int_{f(N_1)}^{f(N)} \left| \langle g(s) \rangle - \frac{1}{2} \right| \big|h'(s)\big| \, ds + |a_N| \leq \frac{f(N) - f(N_1)}{2} + |a_N| \\ &\overset{(5.11)}{\leq} \frac{f(N) - f(N_1) + N_1 - \frac{1}{2}}{2} \\ &\leq \frac{1}{2} \left(b \log N + \int_1^N \frac{\big|\Delta(s)\big|}{s} \, ds + f(1) - f(N_1) + N_1 - \frac{1}{2} \right) \quad , \end{aligned}$$

where the last inequality is by Lemma 5.2.6. Thus $|I_3|$ is bounded by half of the sum of $b \log N$ and $\int_1^N \frac{|\Delta(s)|}{s} \, ds$ shifted by a constant. It therefore follows from Lemma 5.2.7 that there exists $N_3 \in \mathbb{N}$ such that

$$|I_3| \leq \frac{2b}{3} \log N \qquad \forall N \geq N_3 \quad . \tag{5.12}$$

Letting $N_0 := \max\{N_1, N_2, N_3\}$ and summing (5.9), (5.10) and (5.12), we find an upper bound for (5.7).

$$N \left(\int_{\mathbb{T}} h \, d\omega_N - \int_{\mathbb{T}} h \, d\big(\eta_{\frac{1}{b}} \circ R_{f(N)}^{-1}\big) \right) \leq b \log N + 2 \int_1^N \big|\Delta(s)\big| \, ds \qquad \forall N \geq N_0 \ ,$$

which proves the result for Case 1 once we notice that N_1, N_2, N_3, and hence also N_0 are independent of h.

$\boxed{\text{Case 2}}$ $b < 0$

Define $\hat{f} := -f$ and note that $b^{\hat{f}} = -b^f > 0$. By definition of D_N^f and Theorem 1.6.5,

$$D_N^f = d_{\mathbb{T}}\left(\omega_N^f\,,\, \eta_{\frac{1}{b^f}} \circ R_{f(N)}^{-1}\right) = d_{\mathbb{T}}\left(\omega_N^f \circ Q^{-1}\,,\, \eta_{\frac{1}{b^f}} \circ R_{f(N)}^{-1} \circ Q^{-1}\right)$$
$$= d_{\mathbb{T}}\Big(\omega_N^{-f}\,,\, \eta_{\frac{1}{-b^f}} \circ R_{-f(N)}^{-1}\Big) = D_N^{\hat{f}} \quad ,$$

where the penultimate equality is by the definition of Q, Remark 1.6.2, and Remark 1.6.3. Thus the truth of this case has already been shown in Case 1.

$\boxed{\text{Case 3}}$ $b = 0$

Let an arbitrary $N \in \mathbb{N}$ be given. On the one hand, recalling that $\eta_\infty = \delta_0$,

$$\int_{\mathbb{T}} h\, d\big(\eta_{\frac{1}{b}} \circ R_{f(N)}^{-1}\big) = \int_{\mathbb{T}} h\, d\delta_{f(N)} = h \circ f(N) \quad .$$

On the other hand, using Remark 5.1.3 and integration by parts through similar steps to that of the proof of Lemma 5.2.9,

$$\int_{\mathbb{T}} h\, d\omega_N = h \circ f(N) + \frac{1}{N}\Bigg(\frac{h \circ f(N) - h \circ f(1)}{2}$$
$$- \int_1^N \Big(\langle t \rangle - \frac{1}{2} - t\Big)\big(h \circ f\big)'(t)\, dt\Bigg) \quad .$$

Therefore

$$\left|\int_{\mathbb{T}} h\, d\omega_N - \int_{\mathbb{T}} h\, d\big(\eta_{\frac{1}{b}} \circ R_{f(N)}^{-1}\big)\right| \le \frac{1}{N}\Bigg(\frac{\big|h \circ f(N) - h \circ f(1)\big|}{2}$$
$$+ \int_1^N \left|\langle t \rangle - \frac{1}{2} - t\right| \big|h' \circ f(t)\big|\, \big|f'(t)\big|\, dt\Bigg)$$

$$\leq \frac{1}{N}\Bigg(\frac{d_{\mathbb{T}}\left(f(N),f(1)\right)}{2}$$
$$+\int_1^N \Big|-\lfloor t\rfloor-\frac{1}{2}\Big|\,\|h'\|_\infty\,|f'(t)|\;dt\Bigg)$$
$$\leq \frac{1}{N}\Bigg(\frac{d_{\mathbb{T}}\left(f(N),f(1)\right)}{2}+\int_1^N (t+\frac{1}{2})\,|f'(t)|\;dt\Bigg)\,,$$

where the second inequality is by the fact that $h\in \mathrm{Lip}_1(\mathbb{T})$, and the third inequality is by Lemma 2.1.7 as well as the fact that $t>0$. Thus, by definition of $d_{\mathbb{T}}$, as well as the fact that $tf'=\Delta+b$,

$$\left|\int_{\mathbb{T}} h\,d\omega_N-\int_{\mathbb{T}} h\,d\big(\eta_{\frac{1}{b}}\circ R^{-1}_{f(N)}\big)\right| \leq \frac{1}{N}\left(\frac{1}{4}+\int_1^N (1+\frac{1}{2t})\,|\Delta(t)|\;dt\right)$$
$$\leq \frac{1}{N}\left(\frac{1}{4}+\frac{3}{2}\int_1^N |\Delta(t)|\;dt\right)$$
$$\leq \frac{1}{4N}+\frac{2}{N}\int_1^N |\Delta(t)|\;dt\quad .$$

Thus the theorem is true for this case as well as the other cases. ■

Corollary 5.2.11. *For every $f\in\mathcal{F}$,*

$$D_N^f \xrightarrow{N\to\infty} 0\quad .$$

Proof. Immediate from Theorem 5.2.10 through Lemma 5.1.5. ■

In particular, Corollary 5.2.11 proves the promised result that for a slow-growing sequence in $\mathbb{R}$, the associated sequence $(\widetilde{\omega}_N)_{N=1}^\infty$ converges to $\eta_{\frac{1}{b}}$.

The upper bounds in Theorem 5.2.10 do more than just show convergence, they also allow for gauging the speed of convergence, as shown in the following examples.

Example 5.2.12. Consider $x = (\log n)_{n=1}^{\infty}$. This sequence is clearly log-like with $f = \log \in \mathcal{F}$, and $b = 1$, and $\Delta = 0$ constantly. By Theorem 5.2.10 we know that eventually,

$$\begin{aligned} d_{\mathbb{T}}(\widetilde{\omega}_N, \eta_1) &\leq \frac{1}{4N} + \frac{\log N}{N} \\ \implies \qquad \limsup_{N\to\infty} \frac{d_{\mathbb{T}}(\widetilde{\omega}_N, \eta_1)}{\frac{\log N}{N}} &\leq \limsup_{N\to\infty} \left(\frac{1}{4\log N} + 1\right) = 1 < \infty \quad , \end{aligned}$$

which means $d_{\mathbb{T}}(\widetilde{\omega}_N, \eta_1) = O\big(\frac{\log N}{N}\big)$. A careful analysis [6] of $d_{\mathbb{T}}(\widetilde{\omega}_N, \eta_1)$ reveals that the precise speed of convergence is

$$d_{\mathbb{T}}(\widetilde{\omega}_N, \eta_1) \sim \frac{1}{\sqrt{6\pi}} \frac{\sqrt{\log N}}{N} \quad .$$

Thus the bound on $d_{\mathbb{T}}(\widetilde{\omega}_N, \eta_1)$ provided by Theorem 5.2.10 is too pessimistic by a factor $\sqrt{\log N}$ as $N \to \infty$.

Example 5.2.13. Consider $x = (\log\log n)_{n=3}^{\infty}$. Letting $f := \log\log$ on $[e, \infty)$, we see that

$$b = \lim_{t\to\infty} \frac{\frac{1}{t\,\log t}}{\frac{1}{t}} = \lim_{t\to\infty} \frac{1}{\log t} = 0 \quad ,$$

i.e., $f \in \mathcal{F}_0$, and $\Delta(t) = \frac{1}{\log t}$ for every $t > e$. By Theorem 5.2.10 we know that eventually,

$$\begin{aligned} d_{\mathbb{T}}(\widetilde{\omega}_N, \delta_0) &\leq \frac{1}{4N} + \frac{2}{N}\int_e^N \frac{1}{\log t}\,dt \\ \implies \qquad \log N\, d_{\mathbb{T}}(\widetilde{\omega}_N, \delta_0) &\leq \frac{\log N}{4N} + \frac{2\,\log N}{N}\int_e^N \frac{1}{\log t}\,dt \quad , \end{aligned}$$

which through Lemma 5.1.6 implies

$$\limsup_{N\to\infty} \frac{d_{\mathbb{T}}(\widetilde{\omega}_N, \delta_0)}{\frac{1}{\log N}} \leq 2 < \infty \quad ,$$

which means $d_{\mathbb{T}}\left(\widetilde{\omega}_N, \delta_0\right) = O\left(\frac{1}{\log N}\right)$. Again, it can be shown [8] that

$$d_{\mathbb{T}}\left(\widetilde{\omega}_N, \delta_0\right) \sim \frac{1}{\log N} \quad .$$

Thus the bound on $d_{\mathbb{T}}\left(\widetilde{\omega}_N, \delta_0\right)$ provided by Theorem 5.2.10 is sharp as $N \to \infty$, up to a constant factor.

Example 5.2.14. Consider $x = \left(e^{-n}\right)_{n=1}^{\infty}$. Letting $f(t) := e^{-t}$ for every $t \in [1, +\infty)$, we see that

$$b = \lim_{t\to\infty} \frac{-e^{-t}}{1/t} = 0 \quad ,$$

and thus $f \in \mathcal{F}_0$, and $\Delta(t) = -te^{-t}$. Theorem 5.2.10 tells us that eventually,

$$\begin{aligned} & d_{\mathbb{T}}\left(\widetilde{\omega}_N, \delta_0\right) \leq \frac{1}{4N} + \frac{2}{N}\int_1^N \left|-te^{-t}\right| dt \\ \implies \quad & N\, d_{\mathbb{T}}\left(\widetilde{\omega}_N, \delta_0\right) \leq \frac{1}{4} + 2\int_1^N te^{-t}\, dt = \frac{1}{4} + 2\left(\frac{2}{e} - \frac{N+1}{e^N}\right) \\ \implies \quad & \limsup_{N\to\infty} N\, d_{\mathbb{T}}\left(\widetilde{\omega}_N, \delta_0\right) \leq \frac{1}{4} + \frac{4}{e} \approx 1.722 < \infty \quad , \end{aligned}$$

implying that $d_{\mathbb{T}}\left(\widetilde{\omega}_N, \delta_0\right) = O\left(\frac{1}{N}\right)$. This upper bound which Theorem 5.2.10 yielded for $\limsup_{N\to\infty} N\, d_{\mathbb{T}}\left(\widetilde{\omega}_N, \delta_0\right)$ is consistent with the precise value of the limit superior calculated from

$$d_{\mathbb{T}}\left(\widetilde{\omega}_N, \delta_0\right) = \sup_{h\in \mathrm{Lip}_{1,0}(\mathbb{T})} \left| \int_{\mathbb{T}} h\, d\widetilde{\omega}_N - \cancelto{0}{\int_{\mathbb{T}} h\, d\delta_0} \right| = \sup_{h\in \mathrm{Lip}^+_{1,0}(\mathbb{T})} \frac{1}{N}\sum_{n=1}^{N} h\left(e^{-n} - e^{-N}\right) \quad .$$

Let $I(x) := d_{\mathbb{T}}(x, 0)$ for all $x \in \mathbb{T}$. Note that $I \in \mathrm{Lip}^+_{1,0}(\mathbb{T})$. For every $h \in \mathrm{Lip}_{1,0}(\mathbb{T})$, we have $h \leq I$ in the sense that $h(x) \leq I(x)$ for every $x \in \mathbb{T}$. If otherwise, we would contradict Lemma 2.1.7 through the Mean Value theorem. Thus $\sup_{h\in \mathrm{Lip}^+_{1,0}(\mathbb{T})} h(x) = x$ for every $x \in [0, \frac{1}{2}]$. We have therefore shown

$$d_{\mathbb{T}}\left(\widetilde{\omega}_N, \delta_0\right) = \frac{1}{N}\sum_{n=1}^{N}\left(e^{-n} - e^{-N}\right) \quad ,$$

and hence

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

implying $d_{\mathbb{T}}\left(\widetilde{\omega}_N, \delta_0\right) \sim \frac{1}{(e-1)N}$. Note that the bound on $d_{\mathbb{T}}\left(\widetilde{\omega}_N, \delta_0\right)$ provided by Theorem 5.2.10 again is sharp as $N \to \infty$, up to a constant factor.

Example 5.2.15. Consider $x = (\log n + \log\log n)_{n=3}^{\infty}$. Letting $f(t) := \log t + \log\log t$ for every $t \in [3, +\infty)$, we see that

$$b = \lim_{t\to\infty} \frac{\frac{1}{t} + \frac{1}{t\log t}}{\frac{1}{t}} = \lim_{t\to\infty}\left(1 + \frac{1}{\log t}\right) = 1 \quad ,$$

and that $\Delta(t) = \frac{1}{\log t}$. Theorem 5.2.10 tells us that eventually,

$$\begin{aligned} d_{\mathbb{T}}\left(\widetilde{\omega}_N, \eta_1\right) &\leq \frac{1}{4N} + \frac{\log N}{N} + \frac{2}{N}\int_3^N \frac{1}{\log t}\, dt \\ \implies \qquad \log N\, d_{\mathbb{T}}\left(\widetilde{\omega}_N, \eta_1\right) &\leq \frac{\log N}{4N} + \frac{\log(N)^2}{N} + \frac{2\log N}{N}\int_3^N \frac{1}{\log t}\, dt \quad . \end{aligned}$$

Therefore through Lemma 5.1.6 we know

$$\limsup_{N\to\infty} \log N\, d_{\mathbb{T}}\left(\widetilde{\omega}_N, \eta_1\right) \leq 2 \quad , \tag{5.13}$$

implying $d_{\mathbb{T}}\left(\widetilde{\omega}_N, \eta_1\right) = O\left(\frac{1}{\log N}\right)$. As with the previous two examples, this asymptotic analysis gives the correct rate of convergence. In other words, $\lim_{N\to\infty} \log N\, d_{\mathbb{T}}\left(\widetilde{\omega}_N, \eta_1\right)$ is finite and positive. We now show what this limit is.

Since the domain of f is $[3, +\infty)$, we have $f'(t) = \frac{\log t + 1}{t \log t} > 0$, and thus our continuous f is also strictly increasing. Denote the inverse by g. While not explicitly expressed, g has a useful implicit formula. By definition of inverse, for every $s \in [f(3), +\infty)$,

$$\begin{aligned} s &= \log g(s) + \log\log g(s) \\ \implies \qquad e^s &= g(s) \log g(s) \\ \implies \qquad g(s) &= \frac{e^s}{\log g(s)} \quad . \end{aligned}$$

Substituting this implicit formula into itself, we get

$$g(s) = \frac{e^s}{\log \frac{e^s}{\log g(s)}} = \frac{e^s}{s - \log\log g(s)} = \frac{e^s}{s - \log\log \frac{e^s}{\log g(s)}} = \frac{e^s}{s - \log\left(s - \log\log g(s)\right)} \ .$$

Therefore for every $s \in [f(3), +\infty)$,

$$g(s) = \frac{e^s}{s - \log s + \frac{\log s}{s}\widetilde{g}(s)} \quad , \tag{5.14}$$

where $\widetilde{g}(s) := \frac{s}{\log s} \log\left(1 - \frac{\log\log g(s)}{s}\right)$.

Claim 5.2.15.1. The function $\widetilde{g}$ is bounded.

Proof. Note that $f(t) > \log t$ for every $t \in [3, +\infty)$. Since f is increasing and continuous, so is g. Thus g is order-preserving and we have $g \circ f(t) > g \circ \log t$. Since $g \circ f = \exp \circ \log = \mathrm{Id}_{[3,+\infty)}$, we have

$$\exp(y) > g(y) \qquad \forall y \geq f(3) \quad .$$

Thus clearly, for every $s \in [f(3), +\infty)$,

$$\frac{\log\log g(s)}{\log(s)} < 1 \quad \implies \quad \frac{\log\log g(s)}{s} = O\Big(\frac{\log s}{s}\Big) \quad .$$

Therefore by Remark 5.1.7, $\log\left(1 - \frac{\log\log g(s)}{s}\right) = O\left(\frac{\log s}{s}\right)$, and hence $\widetilde{g}$ is bounded. □

Consider the denominator in (5.14). Note that for a large enough $N_1 \in \mathbb{N}$,

$$s - \log s + \frac{\log s}{s}\, \widetilde{g}(s) \geq 1 \qquad \forall s \geq f(N_1) \quad . \tag{5.15}$$

For every $N \geq N_1$, let $l_N := \log N$ for convenience. Let arbitrary $h \in \mathrm{Lip}_1(\mathbb{T})$ be given. By (5.7),

$$N \int_{\mathbb{T}} h\, d\big(\omega_N^f - \eta_1 \circ R_{f(N)}^{-1}\big) = \underbrace{N e^{-(l_N + \log l_N)} \int_{-\infty}^{f(N_1)} h'(s)\, e^{\frac{s}{b}}\, ds}_{J_1} + \underbrace{\int_{f(N_1)}^{f(N)} h'(s) g_N(s)\, ds}_{J_2}$$
$$+ \underbrace{\int_{f(N_1)}^{f(N)} \Big(\langle g(s)\rangle - \frac{1}{2}\Big) h'(s)\, ds + a_N}_{J_3} \quad .$$

By (5.12), we know $\|J_3\| < l_N$ eventually. Additionally, Lemma 2.2.5 implies that $\left|\int_{-\infty}^{f(N_1)} h'(s)\, e^s\, ds\right|$ is bounded by $e^{f(N_1)}$ and thereby $|J_1| \leq \cancel{N} \frac{1}{\cancel{N} l_N}\, N_1 \log N_1$. Thus we know that eventually,

$$\left| N \int_{\mathbb{T}} h\, d\big(\omega_N^f - \eta_1 \circ R_{f(N)}^{-1}\big) - \int_{f(N_1)}^{f(N)} h'(s)\, g_N(s)\, ds \right| < \frac{N_1 \log N_1}{l_N} + l_N \quad ,$$

which after the change of variable $u = f(N) - s$ in the integral, implies that eventually,

$$\left| N \int_{\mathbb{T}} h\, d\big(\omega_N^f - \eta_1 \circ R_{f(N)}^{-1}\big) - \int_0^{f(N) - f(N_1)} h'(f(N) - s) g_N(f(N) - s)\, ds \right|$$
$$< \frac{N_1 \log N_1}{l_N} + l_N \quad .$$

Thus the LHS difference is $O(l_N)$. This implies that the ratio of this difference to l_N^2 vanishes as N grows. Therefore there exists $N_2 \in \mathbb{N}$ such that for every $N \geq N_2$,

$$N \int_{\mathbb{T}} h\, d\big(\omega_N^f - \eta_1 \circ R_{f(N)}^{-1}\big) = c_N\, l_N^2 + \int_0^{f(N) - f(N_1)} h'(f(N) - s) g_N(f(N) - s)\, ds \quad , \tag{5.16}$$

where we know that the number c_N, which clearly depends on N_1 and thereby on h, satisfies $|c_N| \leq 1$ for all $N \geq N_2$.

We will partition the RHS integral in (5.16) into $[0, \sqrt{l_N}]$ and $[\sqrt{l_N}, f(N) - f(N_1)]$. Consider the integrand on the RHS. By definition of g_N, we have for every $N \geq N_2$,

$$\frac{g_N(f(N)-s)}{N} = e^{-s} - \frac{g(f(N)-s)}{N} \qquad \forall s \in [0, f(N) - f(N_1)] \ . \tag{5.17}$$

On the other hand, by (5.14), we know that for every $s \in [0\,,\, f(N) - f(N_1)]$,

$$\frac{g(f(N)-s)}{N} = \frac{1}{N} \frac{e^{f(N)-s}}{f(N) - s - \log(f(N)-s) + \frac{\log(f(N)-s)}{f(N)-s}\widetilde{g}(f(N)-s)} \tag{5.18}$$

$$= \frac{l_N\, e^{-s}}{l_N + \log l_N - s - \log(l_N + \log l_N - s) + \frac{\log(l_N + \log l_N - s)}{l_N + \log l_N - s}\widetilde{g}(f(N)-s)}$$

$$= \frac{l_N\, e^{-s}}{l_N - s - \log(1 + \frac{\log l_N - s}{l_N}) + \frac{\log l_N + \log(1 + \frac{\log l_N - s}{l_N})}{l_N + \log l_N - s}\widetilde{g}(f(N)-s)} \ .$$

Therefore,

$$\frac{g(f(N)-s)}{N} = \frac{e^{-s}}{1 - \frac{s}{l_N} + \frac{1}{l_N^{4/3}}\widetilde{g}_N(s)} \qquad \forall s \in [0, f(N) - f(N_1)] \ , \tag{5.19}$$

where $\widetilde{g}_N(s) := -l_N^{1/3}\log(1 + \frac{\log l_N - s}{l_N}) + l_N^{1/3}\frac{\log l_N + \log(1 + \frac{\log l_N - s}{l_N})}{l_N + \log l_N - s}\widetilde{g}(f(N)-s)$.

Claim 5.2.15.2. For every $N \in \mathbb{N}$, the value $\widetilde{g}_N(s)$ vanishes as $N \to \infty$ for all $s \in [0, \sqrt{l_N}]$.

Proof. Let an arbitrary $s \in [0, \sqrt{l_N}]$ be given. Note that $\widetilde{g}_N$ is smooth. Additionally, note that

$$0 = \lim_{N\to\infty} \frac{\log l_N - \sqrt{l_N}}{l_N} \leq \lim_{N\to\infty} \frac{\log l_N - s}{l_N} \leq \lim_{N\to\infty} \frac{\log l_N}{l_N} = 0 \ . \tag{5.20}$$

Therefore by Remark 5.1.7,

$$\lim_{N\to\infty} l_N^{1/3}\log(1+\frac{\log l_N - s}{l_N}) = \lim_{N\to\infty} l_N^{1/3}\frac{\log l_N - s}{l_N}$$
$$= \lim_{N\to\infty}\frac{\log l_N}{l_N^{2/3}} - \lim_{N\to\infty}\frac{s}{l_N^{2/3}} = 0 \quad ,$$

where $\frac{s}{l_N^{2/3}} \xrightarrow{N\to\infty} 0$ because $s\in[0, l_N^{1/2}]$. As for the second term in $\widetilde{g}_N$, recall that by Claim 5.2.15.1, $\widetilde{g}\left(f(N)-s\right)$ is bounded for all N, s. For convenience, we denote the bound simply by $\widetilde{g}$, so

$$\lim_{N\to\infty}\frac{\log l_N + \log(1+\frac{\log l_N - s}{l_N})}{l_N^{2/3}\left(1+\frac{\log l_N - s}{l_N}\right)}\widetilde{g} = \lim_{N\to\infty}\frac{\log l_N}{l_N^{2/3}\left(1+\frac{\log l_N - s}{l_N}\right)}\widetilde{g}$$
$$+ \lim_{N\to\infty}\frac{\log(1+\frac{\log l_N - s}{l_N})}{l_N^{2/3}\left(1+\frac{\log l_N - s}{l_N}\right)}\widetilde{g} \; = 0 \quad ,$$

where the reason is Remark 5.1.7 as in the first term. Thus $\widetilde{g}_N$ is a continuous function with $\lim_{N\to\infty}\widetilde{g}_N(s) = 0$ uniformly on $s\in[0,\sqrt{l_N}]$. □

Claim 5.2.15.2 implies that there exists $N_3\in\mathbb{N}$ with $N_3 \geq f(N_1)$ such that $|\widetilde{g}_N| < 1$ for all $N\geq N_3$. Turning our focus to $s\in[\sqrt{l_N}, f(N)-f(N_1)]$, note that

$$f(N_1) \leq f(N) - s \leq f(N) - \sqrt{l_N} \quad ,$$

and so (5.15) applies to (5.18), and therefore

$$\left|\frac{g(f(N)-s)}{N}\right| \leq e^{-s}\, l_N^2 \qquad \forall s\in[\sqrt{l_N}\,,\, f(N)-f(N_1)] \quad . \tag{5.21}$$

Partitioning (5.16) implies through (5.17) that

$$l_N \int_{\mathbb{T}} h\, d\big(\omega_N^f - \eta_1 \circ R_{f(N)}^{-1}\big) = \frac{c_N\, l_N^3}{N} + \underbrace{l_N \int_0^{\sqrt{l_N}} h'(f(N)-s)\left(e^{-s} - \frac{g(f(N)-s)}{N}\right) ds}_{K_1}$$

$$+ \underbrace{l_N \int_{\sqrt{l_N}}^{f(N)-f(N_1)} h'(f(N)-s)\left(e^{-s} - \frac{g(f(N)-s)}{N}\right) ds}_{K_2} \ . \qquad (5.22)$$

Claim 5.2.15.3. The limit of K_2 as N grows is 0, i.e.,

$$\lim_{N\to\infty} K_2 = 0 \ .$$

Proof. By the triangle inequality, Lemma 2.2.5, and (5.21),

$$|K_2| \le l_N \int_{\sqrt{l_N}}^{\infty} \left(e^{-s} + e^{-s} l_N^2\right) = l_N(1+l_N^2)\left[-e^{-s}\right]_{\sqrt{l_N}}^{\infty} = \left(l_N + l_N^3\right) e^{-\sqrt{l_N}} \ ,$$

and thus $\lim_{N\to\infty} K_2 = 0$. □

Claim 5.2.15.4. The term K_1 approaches $\int_0^\infty h'(f(N)-s)e^{-s}(-s)\, ds$ as $N \to \infty$, i.e.,

$$\lim_{N\to\infty} \left| K_1 + \int_0^\infty s\, h'(f(N)-s)e^{-s}\, ds \right| = 0 \ .$$

Proof. We first note that by (5.19) we know

$$K_1 = l_N \int_0^{\sqrt{l_N}} h'(f(N)-s)e^{-s}\left(1 - \frac{1}{1 - \frac{s}{l_N} + \frac{1}{l_N^{4/3}}\widetilde{g}_N(s)}\right) ds$$

$$= \int_0^{\sqrt{l_N}} h'(f(N)-s)e^{-s}\left(\frac{-s + \frac{1}{l_N^{1/3}}\widetilde{g}_N(s)}{1 - \frac{s}{l_N} + \frac{1}{l_N^{4/3}}\widetilde{g}_N(s)}\right) ds \ . \qquad (5.23)$$

Also note that by Lemma 2.1.7,

$$\left| \int_{\sqrt{l_N}}^{\infty} s\, h'(f(N)-s)e^{-s}\, ds \right| \le \int_{\sqrt{l_N}}^{\infty} s\, e^{-s}\, ds = (\sqrt{l_N}+1)e^{-\sqrt{l_N}} \ ,$$

and thus $\lim_{N\to\infty}\int_{\sqrt{l_N}}^{\infty} s\,h'(f(N)-s)e^{-s}\,ds = 0$. Therefore,

$$\lim_{N\to\infty}\left|K_1 + \int_0^{\infty} s\,h'(f(N)-s)e^{-s}\,ds\right| = \lim_{N\to\infty}\left|K_1 + \int_0^{\sqrt{l_N}} s\,h'(f(N)-s)e^{-s}\,ds\right| .$$

On the other hand, by (5.23) we know that

$$\begin{aligned}\left|K_1 + \int_0^{\sqrt{l_N}} s\,h'(f(N)-s)e^{-s}\,ds\right| \\ &= \left|\int_0^{\sqrt{l_N}} h'(f(N)-s)e^{-s}\left(\frac{-s+\frac{\widetilde{g}_N(s)}{l_N^{1/3}}}{1-\frac{s}{l_N}+\frac{\widetilde{g}_N(s)}{l_N^{4/3}}}+s\right)ds\right| \\ &= \left|\int_0^{\sqrt{l_N}} h'(f(N)-s)e^{-s}\frac{\frac{\widetilde{g}_N(s)}{l_N^{1/3}}-\frac{s^2}{l_N}+\frac{s\widetilde{g}_N(s)}{l_N^{4/3}}}{1-\frac{s}{l_N}+\frac{\widetilde{g}_N(s)}{l_N^{4/3}}}\,ds\right| .\end{aligned}$$

Therefore by the triangle inequality,

$$\begin{aligned}\left|K_1 + \int_0^{\sqrt{l_N}} s\,h'(f(N)-s)e^{-s}\,ds\right| \le \\ &\frac{1}{l_N^{1/3}}\underbrace{\left|\int_0^{\sqrt{l_N}} h'(f(N)-s)e^{-s}\frac{\widetilde{g}_N(s)}{1-\frac{s}{l_N}+\frac{\widetilde{g}_N(s)}{l_N^{4/3}}}\,ds\right|}_{\widetilde{K}_1} \\ &+\frac{1}{l_N}\underbrace{\left|\int_0^{\sqrt{l_N}} h'(f(N)-s)e^{-s}\frac{s^2}{1-\frac{s}{l_N}+\frac{\widetilde{g}_N(s)}{l_N^{4/3}}}\,ds\right|}_{\widetilde{K}_2} \\ &+\frac{1}{l_N^{4/3}}\underbrace{\left|\int_0^{\sqrt{l_N}} h'(f(N)-s)e^{-s}\frac{s\,\widetilde{g}_N(s)}{1-\frac{s}{l_N}+\frac{\widetilde{g}_N(s)}{l_N^{4/3}}}\,ds\right|}_{\widetilde{K}_3} .\end{aligned} \tag{5.24}$$

All three terms in the above equation converge to zero because $\widetilde{K}_1$, $\widetilde{K}_2$, and $\widetilde{K}_3$ are all bounded. To see this, note that for all $N \geq N_3$ and $s \in [0, \sqrt{l_N}]$, we have $\frac{-\sqrt{l_N}}{l_N} \leq \frac{-s}{l_N} \leq 0$ and $\frac{-1}{l_N^{4/3}} \leq \frac{\widetilde{g}_N}{l_N^{4/3}} \leq \frac{1}{l_N^{4/3}}$. Therefore

$$-\left(\frac{\sqrt{l_N}}{l_N} + \frac{1}{l_N^{4/3}}\right) \leq \frac{\widetilde{g}_N}{l_N^{4/3}} - \frac{s}{l_N} \leq \frac{1}{l_N^{4/3}} \quad .$$

Thus for a large enough $N_4 \geq N_3$, we have for every $N \geq N_4$,

$$\left|1 - \frac{s}{l_N} + \frac{\widetilde{g}_N(s)}{l_N^{4/3}}\right| = 1 - \frac{s}{l_N} + \frac{\widetilde{g}_N(s)}{l_N^{4/3}} \geq 1 - \frac{\sqrt{l_N}}{l_N} - \frac{1}{l_N^{4/3}} \quad . \tag{5.25}$$

For $\widetilde{K}_1$, through Lemma 2.1.7 we have

$$\left|\widetilde{K}_1\right| \leq \int_0^{\sqrt{l_N}} e^{-s} \frac{\left|\widetilde{g}_N(s)\right|}{\left|1 - \frac{s}{l_N} + \frac{\widetilde{g}_N(s)}{l_N^{4/3}}\right|} < \infty \quad ,$$

because the integrand is integrable on any finite interval in $\mathbb{R}^+$, and for every $N \geq N_4$, the integral remains below some constant through (5.25) and Claim 5.2.15.2. Analogously one can see that $\widetilde{K}_2$ and $\widetilde{K}_3$ are bounded as well, and thereby (5.24) is 0, and thus we have shown that

$$\lim_{N\to\infty} \left|K_1 + \int_0^\infty s\, h'(f(N) - s)e^{-s}\, ds\right| = 0 \quad .$$

□

Thus (5.22) implies through Claims 5.2.15.3 and 5.2.15.4 that

$$\lim_{N\to\infty} l_N \int_{\mathbb{T}} h\, d\big(\omega_N^f - \eta_1 \circ R_{f(N)}^{-1}\big) = \lim_{N\to\infty} \int_0^\infty h'(f(N) - s)e^{-s}(-s)\, ds \quad , \tag{5.26}$$

uniformly in $h \in \mathrm{Lip}_1(\mathbb{T})$. On the other hand, we observe that

$$-\int_0^\infty h'\left(f(N) - s\right) se^{-s}\, ds = -\sum_{n=0}^{\infty} \int_n^{n+1} h'\left(f(N) - s\right) se^{-s}\, ds$$

$$
\begin{aligned}
&= -\sum_{n=0}^{\infty}\int_0^1 h'\big(f(N)-s\big)\,(s+n)e^{-s-n}\,ds \\
&= -\int_0^1 h'\big(f(N)-s\big)\,e^{-s}\Bigg(s\sum_{n=0}^{\infty}e^{-n}+\sum_{n=0}^{\infty}ne^{-n}\Bigg)\,ds \\
&= -\int_0^1 h'\big(f(N)-s\big)\left(\frac{se^{1-s}}{e-1}+\frac{e^{1-s}}{(e-1)^2}\right)ds \quad ,
\end{aligned}
$$

where the second equality is because h' is 1-periodic. This also implies that $h'\big(f(N)-s\big)=h'\big(f(N)+1-s\big)$. Using the substitution $u=1-s$, we have

$$
\begin{aligned}
-\int_0^{\infty} h'\big(f(N)-s\big)\,se^{-s}\,ds &= -\int_0^1 h'\big(f(N)+u\big)\left(\frac{(1-u)e^u}{e-1}+\frac{e^u}{(e-1)^2}\right)du \\
&= -\int_0^1 h'\big(f(N)+u\big)\left(\frac{(1-u)e^u}{e-1}+\frac{e^u-e}{(e-1)^2}\right)du \\
&= -\int_0^1 h'\big(f(N)+u\big)\big(F_{\zeta_1}(u)-F_{\eta_1}(u)\big)\,du \\
&= \int_{\mathbb{T}} h\,d\big(\zeta_1\circ R_{f(N)}^{-1}\big)-\int_{\mathbb{T}} h\,d\big(\eta_1\circ R_{f(N)}^{-1}\big) \quad ,
\end{aligned}
$$

where the ultimate and antepenultimate equalities are by Remark 2.2.4 and Lemma 2.2.5 respectively. Thus by (5.26) we have shown

$$
\lim_{N\to\infty}\left(l_N\int_{\mathbb{T}} h\,d\big(\omega_N^f-\eta_1\circ R_{f(N)}^{-1}\big)-\int_{\mathbb{T}} h\,d\big((\zeta_1-\eta_1)\circ R_{f(N)}^{-1}\big)\right)=0 \quad ,
$$

uniformly in $h\in \mathrm{Lip}_1(\mathbb{T})$. Replacing h with $h\circ R_{f(N)}^{-1}\in \mathrm{Lip}_1(\mathbb{T})$, we get

$$
\lim_{N\to\infty} l_N\int_{\mathbb{T}} h\,d\big(\widetilde{\omega}_N^f-\eta_1\big)=\lim_{N\to\infty}\int_{\mathbb{T}} h\,d(\zeta_1-\eta_1) \quad .
$$

Taking the supremum over $h\in \mathrm{Lip}_1(\mathbb{T})$ yields

$$
\lim_{N\to\infty} l_N\, d_{\mathbb{T}}\left(\widetilde{\omega}_N,\eta_1\right)=d_{\mathbb{T}}(\zeta_1\,,\,\eta_1) \quad .
$$

Thus we have shown that $d_{\mathbb{T}}\left(\widetilde{\omega}_N,\eta_1\right)$ is asymptotically equivalent to a constant multiple of $\frac{1}{\log N}$.

Remarkably, the upper bound that Theorem 5.2.10 gauges for the rate of the decay of D_N^f was sharp in Examples 5.2.13 to 5.2.15,up to a constant factor.

Consider a sequence of real numbers $(x_n)_{n=1}^{\infty}$. Slow-varying or not, this sequence has a convergent subsequence mod 1 simply because $(\mathbb{T}, d_{\mathbb{T}})$ is compact. In other words, $\mathbb{A} \neq \emptyset$. The same is true for the sequence $(\omega_N)_{N=1}^{\infty}$ in the compact space $(\mathcal{P}, d_{\mathbb{T}})$. In other words, $\Omega \neq \emptyset$. Theorem 5.2.17 uses Corollary 5.2.11 to characterize the limit set Ω and its relation to $\mathbb{A}$. It states that precisely for the subsequences $\left(x_{N_j}\right)_{j=1}^{\infty}$ that converge to a t mod 1, the sequence $\left(\omega_{N_j}\right)_{j=1}^{\infty}$ of empirical distributions converge to an exponential distribution mod 1 rotated by t. On the other hand, Theorem 5.2.16 below tells us that a slow-varying sequence $(x_n)_{n=1}^{\infty}$ accumulates everywhere on $\mathbb{T}$.

Theorem 5.2.16. *For every $f \in \mathcal{F} \setminus \mathcal{F}_0$, we have $\mathbb{A}^f = \mathbb{T}$.*

Proof. Let arbitrary $f \in \mathcal{F} \setminus \mathcal{F}_0$ be given.

Case 1 $b > 0$.

Since $\lim_{t\to\infty} \frac{f'(t)}{1/t} = b > 0$, there exists $t_0 \in \mathbb{R}^+$ such that f is strictly increasing on $[t_0, +\infty)$. For this proof, it suffices to show $\mathbb{T} \subseteq \mathbb{A}^f$. To that end, let arbitrary $x + \mathbb{Z} \in \mathbb{T}$ be given. We want to show

$$\exists \left(f(n_j)\right)_{j=1}^{\infty} : \forall \epsilon > 0, \ \exists j_0 \in \mathbb{N} : \forall j \geq j_0 \,, \ d_{\mathbb{T}}\left(f(n_j), x\right) < \epsilon \quad .$$

Let an arbitrary $\epsilon > 0$ be given. WLOG assume $\epsilon < \frac{1}{2}$. Set $j = 1$. To find n_j, let $\epsilon_j = \frac{1}{j}\epsilon$. Since $\left|f(n) - f(n-1)\right| \xrightarrow{n\to\infty} 0$, there exists $n_{0j} \in \mathbb{N}$ such that $n_{0j} \geq t_0$, and

$$\left|f(n) - f(n-1)\right| < \epsilon_j \quad \forall n \geq n_{0j} \,. \tag{5.27}$$

The fact that $f(n) \xrightarrow{n\to\infty} +\infty$ tells us that the following set N is non-empty:

$$N := \left\{N_0 \in \mathbb{N} \,:\, N_0 \geq n_{0j} \quad \wedge \quad \text{dist}\left(f(N_0), x + \mathbb{Z}\right) < \epsilon_j\right\} .$$

Define $n_j := \min N$. Set $t_0 = n_j$, increment j by 1, and repeat the above steps. Since $\epsilon_j \xrightarrow{j\to\infty} 0$, we have shown the desired result.

Case 2 $b < 0$.

Define $\hat{f} := -f$ and note that

$$\mathbb{A}^{\hat{f}} = Q(\mathbb{A}^f) = Q(\mathbb{T}) = \mathbb{T} \quad ,$$

where the penultimate equality is by Case 1. ■

Theorem 5.2.17. *Let $(x_n)_{n=1}^{\infty}$ be a slow-varying sequence of real numbers in the sense of Definition 5.2.2, and let $t \in \mathbb{R}$. Consider any subsequence $\left(x_{N_j}\right)_{j=1}^{\infty}$. We have*

$$\omega_{N_j} \xrightarrow{j\to\infty} \eta_{\frac{1}{b}} \circ R_t^{-1} \iff \lim_{j\to\infty} d_{\mathbb{T}}\left(x_{N_j}, t\right) = 0 \quad .$$

Proof.

$\Longleftarrow$

Assume $x_{N_j} \xrightarrow{j\to\infty} t$. By the triangle inequality,

$$\begin{aligned} d_{\mathbb{T}}\left(\omega_{N_j}, \eta_{\frac{1}{b}} \circ R_t^{-1}\right) &\le d_{\mathbb{T}}\left(\omega_{N_j}, \eta_{\frac{1}{b}} \circ R_{x_{N_j}}^{-1}\right) + d_{\mathbb{T}}\left(\eta_{\frac{1}{b}} \circ R_{x_{N_j}}^{-1}, \eta_{\frac{1}{b}} \circ R_t^{-1}\right) \\ &\le D_{N_j} + C\, d_{\mathbb{T}}\left(t, x_{N_j}\right) \quad , \end{aligned}$$

where the second inequality is by Theorem 4.2.8, and $C = \left|\tanh\left(\frac{1}{4b}\right)\right|$. Taking the limit as $j \to \infty$,

$$\lim_{j\to\infty} d_{\mathbb{T}}\left(\omega_{N_j}, \eta_{\frac{1}{b}} \circ R_t^{-1}\right) \le \cancelto{0}{\lim_{j\to\infty} D_{N_j}} + C \cancelto{0}{\lim_{j\to\infty} d_{\mathbb{T}}\left(t, x_{N_j}\right)},$$

where $\lim_{j\to\infty} d_{\mathbb{T}}\left(t, x_{N_j}\right)$ and $\lim_{j\to\infty} D_{N_j}$ vanish by assumption and Corollary 5.2.11 respectively. Thus $\lim_{j\to\infty} d_{\mathbb{T}}\left(\omega_{N_j}, \eta_{\frac{1}{b}} \circ R_t^{-1}\right) = 0$.

$\Longrightarrow$

Assume $d_{\mathbb{T}}\left(\omega_{N_j}, \eta_{\frac{1}{b}} \circ R_t^{-1}\right) \xrightarrow{j\to\infty} 0$. By Theorem 4.2.8 and the triangle inequality,

$$\begin{aligned} d_{\mathbb{T}}\left(x_{N_j}, t\right) &\le C' \, d_{\mathbb{T}}\left(\eta_{\frac{1}{b}} \circ R^{-1}_{x_{N_j}} \, , \, \eta_{\frac{1}{b}} \circ R^{-1}_t\right) \\ &\le C' \left(d_{\mathbb{T}}\left(\eta_{\frac{1}{b}} \circ R^{-1}_{x_{N_j}}, \omega_{N_j}\right) + d_{\mathbb{T}}\left(\omega_{N_j}, \eta_{\frac{1}{b}} \circ R^{-1}_t\right)\right) \quad , \end{aligned}$$

where $C' = \frac{\left|\frac{1}{b}\right|}{4 \log \cosh\left(\frac{1}{4b}\right)}$. Taking the limit as $j \to \infty$,

$$\lim_{j\to\infty} d_{\mathbb{T}}\left(x_{N_j}, t\right) \le C' \lim_{j\to\infty} D_{N_j} + \lim_{j\to\infty} d_{\mathbb{T}}\left(\omega_{N_j}, \eta_{\frac{1}{b}} \circ R^{-1}_t\right) \quad ,$$

where $\lim_{j\to\infty} d_{\mathbb{T}}\left(\omega_{N_j}, \eta_{\frac{1}{b}} \circ R^{-1}_t\right)$ and $\lim_{j\to\infty} D_{N_j}$ vanish by assumption and Corollary 5.2.11 respectively. Thus $\lim_{j\to\infty} d_{\mathbb{T}}\left(x_{N_j}, t\right) = 0$. ■

Clearly the above theorem says that the set of all empirical distributions of $(x_n)_{n=1}^{\infty}$ and its subsequences accumulates at every mod 1 exponential distribution $\eta_{\frac{1}{b}}$ that is rotated by an accumulation point of $(x_n)_{n=1}^{\infty}$. We formally state this characterization of Ω in Corollary 5.2.18.

Corollary 5.2.18. *Let* $x = (x_n)_{n=1}^{\infty}$ *be a slow-varying sequence in* $\mathbb{R}$*. Then,*

$$\Omega^x = \left\{\eta_{\frac{1}{b^x}} \circ R^{-1}_{x_0} : x_0 \in \mathbb{A}\right\} \quad .$$

Proof. Immediate from Theorem 5.2.17. ■

Thus, together with Theorem 5.2.16, we see that given a slow-varying sequence, for every $x \in \mathbb{A}$, there exists a subsequence $\left(\omega_{N_j}\right)_{j=1}^{\infty}$ of mod 1 empirical distributions that converge to $\eta_{\frac{1}{b}} \circ R^{-1}_x$.

In summary, given a log-like sequence $(x_n)_{n=1}^{\infty}$, Theorem 5.2.10 told us that if every mod 1 empirical distribution ω_N is rotated by $-x_N$, then $\left(\omega_N \circ R^{-1}_{-x_N}\right)_{N=1}^{\infty}$ converges to $\eta_{\frac{1}{b}}$. Theorem 5.2.17 offered an alternative view of this convergence: we do not need to rotate the empirical distributions if they are associated with a convergent (in $\mathbb{T}$) sequence. They converge to a

$\eta_{\frac{1}{b}}$ rotated by the limit in $\mathbb{T}$. This view explains the observations made in Figures 0.2 and 0.3 regarding $x = (\log_{10} n)_{n=1}^{\infty}$. The pattern observed in Figure 0.2 was for the empirical distributions associated with the subsequence $\left(\log_{10}(5 \times 10^j)\right)_{j=1}^{\infty}$ which mod 1 is the constant sequence $(\log_{10} 5)_{j=1}^{\infty}$. The empirical distributions do therefore, as suspected, converge to a mod 1 exponential distribution, namely $\eta_{\frac{1}{\log_{10} e}} \circ R_{\log_{10} 5}^{-1}$. Similarly, the empirical distributions associated with $\left(\log_{10}(1 \times 10^j)\right)_{j=1}^{\infty}$ depicted in Figure 0.3 converges to $\eta_{\frac{1}{\log 10}} \circ R_0^{-1} = \eta_{\frac{1}{\log 10}}$. When it comes to the sequence x, any sequence of empirical distributions that does converge, converges to some rotated version of the same exponential distribution, namely $\eta_{\frac{1}{\log 10}}$.

5.3 Distribution of the first significant digits

In this section we explain how $\mathbb{T}$ and $\mathcal{P}$ are related to the question of significant digits, and how they explain the pattern observed in Figures 0.5 and 0.6.

5.3.1 Setting the Benford stage

Most likely because of the number of fingers on human hands, the radix 10 is central to perceiving and recording numbers. For example, the symbol 472, as children learn in elementary school, means 4 hundreds plus 7 tens plus 2 ones. In other words, every digit is associated with a power of 10. Informally, the non-zero digit associated with the highest power of 10 is the *first significant digit*. For a positive number, the defining property of this digit is that if increased by 1, it will bound the number from above. In the given example, 5 hundreds bound 472 from above. That is to say, 472 is between 4 hundreds and 5 hundreds. The same is clearly not the case for other digits: 472 is not between 7 tens and 8 tens, nor is it between 2 ones and 3 ones.

Definition 5.3.1 (First significant digit)**.** [7] Let $t \in \mathbb{R}$ be non-zero. The first significant (decimal) digit of t, denoted $D_1(t)$, is a unique integer $j \in \{1, 2, \cdots, 9\}$ satisfying

$$10^k j \ \leq \ |t| \ < \ 10^k (j+1) \ \ ,$$

for some (necessarily unique) $k \in \mathbb{Z}$.

The centrality of the radix 10 is also tangible in the scientific notation for recording a number, which writes every $t \in \mathbb{R} \setminus \{0\}$ as

$$t = \operatorname{sgn}(t) \times 10^{\langle \log_{10}|t| \rangle} \times 10^{\lfloor \log_{10}|t| \rfloor} \quad , \tag{5.28}$$

where the left-most factor is the sign of t given by $\frac{t}{|t|}$, and the right-most factor is an integer power of 10. The integer $\lfloor \log_{10}|t| \rfloor$ is called the *order of magnitude* of t. The middle factor, known as the *significand* of t, is a number in $[1, 10)$ because $\langle \log_{10}|t| \rangle$ is in $[0, 1)$. One can think of the significand of t as the value of a function $S\colon \mathbb{R} \setminus \{0\} \to [1, 10)$ at t.

Definition 5.3.2 (The significand)**.** The significand is the function $S\colon \mathbb{R} \setminus \{0\} \to [1, 10)$ given by

$$S(t) := 10^{\langle \log_{10}|t| \rangle} \qquad \forall t \in \mathbb{R} \setminus \{0\} .$$

We additionally define $S(0) := 0$. It is easy to see that S is $\mathcal{B}_{\mathbb{R}}$-$\mathcal{B}_{[1,10)}$-measurable [7].

Given a real number t, the value $S(t) \in [1, 10)$ contains all the information about the significant digits of t. That is to say, in the representation (5.28) of t, changes in the sign or the order of magnitude do not affect the significant digits. For this reason, when it comes to probability measures for events describing significant digits of reals, $\mathcal{B}_{\mathbb{R}}$ is too fine a σ-algebra on $\mathbb{R}$. A more suitable σ-algebra would include all real numbers that share a significand in the same measurable set regardless of their order of magnitude or sign. For example, if the event of interest is having precisely the three significant digits 472 (in that order), the set to be assigned a probability must be

$$\left\{ \cdots , \pm\frac{472}{100} , \pm\frac{472}{10} , \pm 472 , \pm 4720 , \pm 47200 , \cdots \right\} .$$

The σ-algebra generated by S achieves this. We denote this σ-algebra by $\mathcal{S}$.

$$\mathcal{S} := \sigma(S) = \left\{S^{-1}(B) : B \in \mathcal{B}_{[1,10)}\right\} \subset 2^{\mathbb{R}} \quad .$$

Despite what might be inferred from the Motivation, Benford's law does not only describe the distribution of the first significant digits, but all significant digits as well as their combinations. This law is characterized by whether it gives rise to $\log_{10}$ as the the CDF of S.

Definition 5.3.3 (Benford probability measure)**.** A probability measure $\mathbb{P}$ on $(\mathbb{R}, \mathcal{S})$ is *Benford* if and only if

$$\mathbb{P}\left(\{S \leq t\}\right) = \log_{10} t \qquad \forall t \in [1,10) \quad . \tag{5.29}$$

This probability measure is unique.

It is clearly seen that (5.29) implies the first-digit rule described in the Motivation which stated

$$\mathbb{P}\left(\{D_1 = d_1\}\right) = \log_{10}(d_1+1) - \log_{10}(d_1) \qquad \forall d_1 \in \{1, 2, \cdots, 9\} \quad . \tag{5.30}$$

Since the distribution of S on $[1,10)$ assigns a probability to any event concerning significant digits (i.e., any set in $\mathcal{S}$), it is sometimes convenient to skip a level of abstraction and consider only the pushforward measure $\mathrm{P} := \mathbb{P} \circ S^{-1}$ as the main probability distribution. This poses no problems because P and $\mathbb{P}$ fully determine each other. In this view, Definition 5.3.3 can be re-written as the following.

Definition 5.3.4. A probability measure P on $\left([1,10), \mathcal{B}_{[1,10)}\right)$ is *Benford* if and only if

$$\mathrm{P}\left([1,t]\right) = \log_{10} t \qquad \forall t \in [1,10) \quad .$$

This probability measure is unique.

Just as a real sequence $x = (x_n)_{n=1}^{\infty}$ has a sequence of associated mod 1 empirical distributions (see Definition 5.2.1), it also has a sequence of significand empirical distributions defined for every $N \in \mathbb{N}$ to be

$$\alpha_N^x := \frac{1}{N} \sum_{n=1}^{N} \delta_{S(x_n)} \quad .$$

Every α_N^x is a probability distribution on $\big([1,10), \mathcal{B}_{[1,10)}\big)$. Figures 0.4 and 0.5 plot the value of $\alpha_N\left(\{D_1 = 1\}\right)$ for the sequences $(2^n)_{n=1}^{\infty}$ and $(n)_{n=1}^{\infty}$, respectively.

We would like to call a sequence Benford if α_N^x somehow approaches the Benford probability measure P of Definition 5.3.4 as N increases. All we have to define is the sense in which this convergence must occur.

5.3.2 The relevance to $\mathbb{T}$

The range of S is easily perceived to have a circular structure if one takes a real sequence like $(9, 9.9, 9.99, 9.999, \cdots)$, and considers the significand of its limit. This is the result of the fact that Definition 5.3.2 has a fractional part as the power of 10. Thus the function $\log_{10} \circ S$ takes us to the familiar space $[0,1)$ which we identify with $\mathbb{T}$ through the bijection $\iota_{\mathbb{R}}$. The probability distribution of $\log_{10} \circ S$ on $(\mathbb{T}, \mathcal{B}_{\mathbb{T}})$ fully determines the distribution of S on $\big([1,10), \mathcal{B}_{[1,10)}\big)$ and vice versa, and thereby fully determines the probability measure defined on $(\mathbb{R}, \mathcal{S})$ and vice versa.

Example 5.3.5. Let $\mathbb{P}$ be a probability measure on $(\mathbb{R}, \mathcal{S})$. Assume the event of interest is $\{D_1 = 1\}$. We have

$$\begin{aligned}\mathbb{P}\left(\{D_1 = 1\}\right) = \mathbb{P}\Big(\big\{S \in [1,2)\big\}\Big) &= \mathbb{P}\Big(\big\{\log_{10} \circ S \in [0, \log_{10} 2)\big\}\Big) \\ &= \mu_S\Big(\big\{[0, \log_{10} 2) + \mathbb{Z}\big\}\Big) \quad ,\end{aligned}$$

where $\mu_S \in \mathcal{P}$ is the pushforward of $\mathbb{P}$ under $\log_{10} \circ S$. In other words, if μ_S is known, then the probability of $\{D_1 = 1\}$ is just the μ_S-measure of the arc from $0 + \mathbb{Z}$ to $\log_{10} 2 + \mathbb{Z}$ on the circle.

Analogously, in general, for every $d_1 \in \{1, 2, \cdots, 9\}$,

$$\mathbb{P}\left(\{D_1 = d_1\}\right) = \mu_S\left(\left\{\left[\log_{10} d_1\,,\, \log_{10}(d_1+1)\right) + \mathbb{Z}\right\}\right)\ , \tag{5.31}$$

Note that if μ_S simply returns the length of the arc of interest, then the probability turns out exactly what the Benford law implies in (5.30). This is no accident. In fact, it is well-known that $\mathbb{P}$ is Benford if and only if μ_S is uniform [7]. In summary, the study of the distribution S on $[1, 10)$ and whether it is Benford, can be replaced with the study of the distribution of $\langle \log_{10} |\cdot| \rangle$ on $\mathbb{T}$ and whether it is uniform.

Therefore, given a sequence $(x_n)_{n=1}^{\infty}$ in $\mathbb{R}$, the study of the distribution of $y = \left(S(x_n)\right)_{n=1}^{\infty}$ on $[1, 10)$ reduces to the study of the distribution of $x = \left(\log_{10}|x_n|\right)_{n=1}^{\infty}$ mod 1. The sequence $(\alpha_N^y)_{N=1}^{\infty}$ converges to P if and only if $(\omega_N^x)_{N=1}^{\infty}$ converges to $\lambda_{\mathbb{T}}$.

For the sequence of natural numbers, the question becomes whether or not $l = (\log_{10} n)_{n=1}^{\infty}$ is uniformly distributed mod 1. Noting that

$$\log_{10} n = \frac{1}{\log 10} \log n \qquad \forall n \in \mathbb{N}\ ,$$

the sequence is slow-growing with $b^l = \frac{1}{\log 10}$, and therefore $(\omega_N^l)_{N=1}^{\infty}$ does not converge. It is therefore no surprise that the sequence $(n)_{n=1}^{\infty}$ is not Benford. However, we know by Theorem 5.2.16 that for every $M \in \mathbb{N}$, the point $\log_{10} M$ is an entry in a subsequence $(\log_{10} N_j)_{j=1}^{\infty}$ that converges mod 1 to $\log_{10} M$. Thus through Theorem 5.2.17 we know that ω_N^l is an approximation of $\eta_{\frac{1}{\log_{10} e}} \circ R_{\log_{10}(N)}^{-1}$ for every $N \in \mathbb{N}$. Therefore ω_N^l is just more or less the exponential distribution $\eta_{\frac{1}{\log_{10} e}}$ rotated such that its most dense region is "behind" $\log_{10} N$. From the fact that $\log_{10} N \xrightarrow{N\to\infty} +\infty$ and $\log N - \log(N-1) \xrightarrow{N\to\infty} 0$, we see that the approximation of the rotated $\eta_{\log 10}$ that ω_N^l is, keeps rotating around the circle endlessly and in increasingly fine angles as N grows (see Figure 5.1). It is this endless rotation that causes the periodic up-down pattern seen in Figures 0.5 and 0.6. This example captures the essence of the convergence behavior (or lack thereof) of slow-growing sequences: As N grows the empirical distributions approximate an exponential distribution increasingly well, yet there is no convergence because the exponential distribution they approximate keeps rotating.

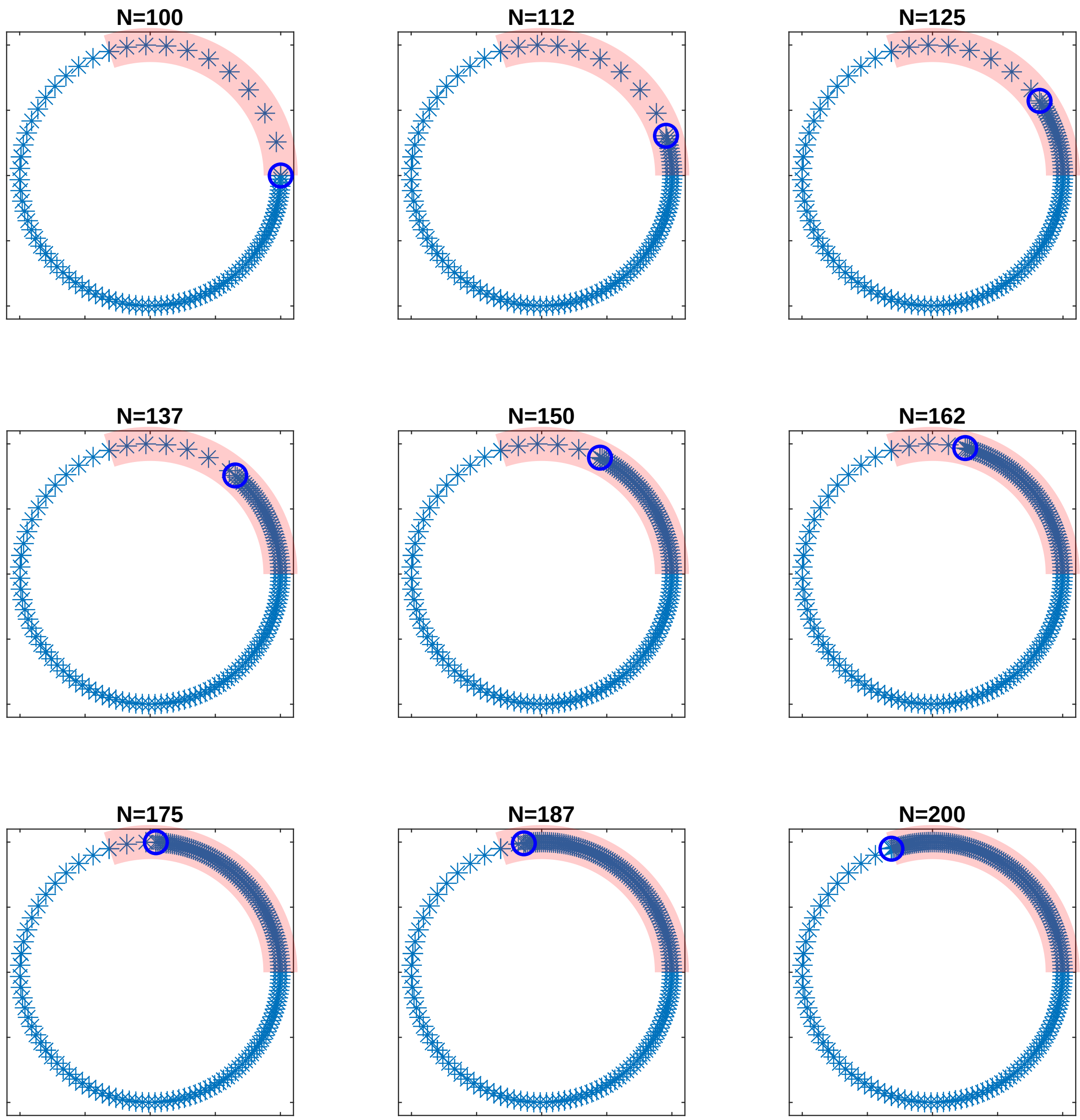

Figure 5.1: A depiction of ω_N^l (in blue) where $l = (\log_{10} n)_{n=1}^{\infty}$. Each ω_N is an approximation of a rotated version of $\eta_{\log 10}$, with the higher density "behind" the point $l_N + \mathbb{Z}$ (circled in blue). As N increases, the approximation rotates with $x_N + \mathbb{Z}$. The event $\{D_1(n) = 1\}$ (in pink) has a small measure under ω_{100}^l because it coincides with the least dense region of ω_N^l. As N increases from 100 to 199 the measure increases, and achieves a local max at $N = 199$ where the event coincides with the most dense region of ω_N^l. The measure of the event then starts to decrease for any N after 200 and before 1000 (not pictured). This increase and decrease is precisely the pattern observed in Figure 0.5.

Conclusion

The Kantorovich metric $d_{\mathbb{T}}$ induces the weak topology on the space $\mathcal{P}$ of all mod 1 probability measures. For any two $\mu, \nu \in \mathcal{P}$, the distance $d_{\mathbb{T}}(\mu, \nu)$ can be calculated using $\int_0^1 \left|F_\mu(s) - F_\nu(s) - t_{min}\right| ds$ where t_{min} is a median value of $F_\mu - F_\nu$. The metric $d_{\mathbb{T}}$ is invariant under rotations and reflections. With this metric, $\mathcal{P}$ can be perceived as a compact ball of radius $\frac{1}{4}$ centered at $\lambda_{\mathbb{T}}$. In other words, no probability measure in $\mathcal{P}$ is more than $\frac{1}{4}$ away from $\lambda_{\mathbb{T}}$. The set of probability measures that are precisely $\frac{1}{4}$ away from $\lambda_{\mathbb{T}}$ is the set $\{\delta_x : x \in \mathbb{T}\}$ of Dirac measures, which itself is topologically isomorphic to $\mathbb{T}$. Given any exponential distribution η_a, the set $\left\{\eta_a \circ R_s^{-1} : s \in \mathbb{R}\right\}$ of all its rotated versions is again topologically isomorphic to $\mathbb{T}$.

Given a real sequence $x = (x_n)_{n=1}^{\infty}$ for which $b := \lim_{n\to\infty} n(x_n - x_{n-1})$ exists in $\mathbb{R}$, the limit set of the associated mod 1 empirical distributions is precisely the set of mod 1 exponential distributions $\eta_{\frac{1}{b}}$ that are rotated by every point of in the limit set of x, with the convention that $\eta_{\frac{1}{0}} = \eta_\infty = \delta_0$. More precisely, for any subsequence $(x_{N_j})_{j=1}^{\infty}$ of x, the subsequence $(\omega_{N_j})_{j=1}^{\infty}$ of $(\omega_N)_{N=1}^{\infty}$ converges to $\eta_{\frac{1}{b}} \circ R_{x_0}^{-1}$ where $x_0 \in \mathbb{T}$ is the limit of $(x_{N_j})_{j=1}^{\infty}$, in symbols,

$$\omega_{N_j} \xrightarrow{j\to\infty} \eta_{\frac{1}{b}} \circ R_{x_0}^{-1} \iff \lim_{j\to\infty} d_{\mathbb{T}}\left(x_{N_j}, x_0\right) = 0 \ .$$

Specifically, for a log-like sequence $l = (l_n)_{n=1}^{\infty}$, this means that the associated empirical distributions accumulate at every rotated version of the exponential distribution $\eta_{\frac{1}{b}}$. In other words, the limit set of $(\omega_N^l)_{N=1}^{\infty}$ is the circle $\left\{\eta_{\frac{1}{b}} \circ R_x^{-1} : x \in \mathbb{T}\right\}$ provided that $b \neq 0$. For example, the pattern seen in Figure 0.2 is a subsequence of the empirical distributions associated with $(\log_{10} n)_{n=1}^{\infty}$ that approximate $\eta_{\frac{1}{\log_{10} e}} \circ R^{-1}_{\log_{10}(5)}$ increasingly well since $\left(\log_{10}(5 \times 10^j)\right)_{j=1}^{\infty}$ is a constant sequence mod 1.

The fact that $(\omega^l_N)_{N=1}^{\infty}$ approximates some rotated version of $\eta_{\frac{1}{b}}$ is a result of Theorem 5.2.10 which provides the following upper bound for the distance D^f_N between $\eta_{\frac{1}{b}}$ and the suitably rotated ω^l_N:

$$D^f_N \leq \frac{1}{4N} + \frac{\left|b^f\right| \log N}{N} + \frac{2}{N} \int_1^N \left|\Delta^f(t)\right| dt \quad ,$$

for every large enough N. This upper bound allows us to describe the speed of convergence of $(\widetilde{\omega}_N)_{N=1}^{\infty}$ as big-O of a sequence. As shown in Examples 5.2.13 to 5.2.15, in some cases the upper bound is sharp, in that the big-O asymptotics that it yields is in fact asymptotically equivalent to D^f_N, up to a constant factor.

Natural questions to investigate in the future include finding the speed of convergence for the empirical distributions associated with sequences that are known to have a distribution mod 1. For example, how fast does the sequence of empirical measures associated with $(\log_{10} F_n)_{n=1}^{\infty}$ of the logarithm of Fibonacci numbers converge to the uniform distribution mod 1? Another interesting question naturally arises in light of Example 5.2.15: The prime number theorem implies that the sequence $(p_n)_{n=1}^{\infty}$ of prime numbers is asymptotically equivalent to $(n \log n)_{n=1}^{\infty}$, which implies that $\log p_n \sim \log n + \log \log n$. Since in Example 5.2.15 we were able to find the limit set of the mod 1 empirical distributions associated with $(\log n + \log \log n)_{n=1}^{\infty}$ as well as the precise rate of convergence, it is natural to ask whether anything can be said for the mod 1 empirical distributions associated with $(\log p_n)_{n=1}^{\infty}$.

Bibliography

[1] *The Concise Oxford Dictionary of Mathematics*. Oxford University Press, 2021.

[2] Tom M Apostol. An Elementary View of Euler's Summation Formula. *The American Mathematical Monthly*, 106(5):409–418, 1999.

[3] Tom M Apostol. *Introduction to Analytic Number Theory*. Undergraduate texts in mathematics. Springer, New York, NY, December 2010.

[4] Sheldon Axler. *Linear Algebra Done Right*. Undergraduate Texts in Mathematics. Springer International Publishing, Basel, Switzerland, 3rd edition, December 2014.

[5] Frank Benford. The Law of Anomalous Numbers. *Proceedings of the American Philosophical Society*, 78(4):551–572, 1938.

[6] Arno Berger. A Note on the Distributions of (log n) mod 1. *Uniform Distribution Theory*, 17(2):77–100, 2022.

[7] Arno Berger and Theodore P Hill. *An Introduction to Benford's Law*. Princeton University Press, May 2015.

[8] Arno Berger and Ardalan Rahmatidehkordi. Circling the uniform distribution. *Journal of Mathematical Analysis and Applications*, 527(2), 2023.

[9] Joseph K Blitzstein and Jessica Hwang. *Introduction to Probability*. Crc Press Boca Raton, FL, 2015.

[10] Sergey Bobkov and Michel Ledoux. *One-dimensional empirical measures, order statistics, and kantorovich transport distances*. Memoirs of the American Mathematical Society. American Mathematical Society, Providence, RI, October 2020.

[11] Stefan Cobzas, Radu Miculescu, and Adriana Nicolae. *Lipschitz Functions*. Lecture notes in mathematics. Springer Nature, Cham, Switzerland, 1st edition, May 2019.

[12] Donald L Cohn. *Measure theory*. Birkhauser Advanced Texts Basler Lehrbucher. Birkhauser Boston, Secaucus, NJ, 2nd edition, July 2013.

[13] John B Conway. *A Course in Abstract Analysis*. Graduate Studies in Mathematics. American Mathematical Society, Providence, Rhode Island, October 2012.

[14] John B Conway. *A Course in Point Set Topology*. Undergraduate texts in mathematics. Springer, 1st edition, 2014.

[15] Gerald B Folland. *Real Analysis: Modern Techniques and Their Applications*. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. John Wiley & Sons, Nashville, TN, 2nd edition, March 1999.

[16] Morris W Hirsch, Stephen Smale, and Robert L Devaney. *Differential Equations, Dynamical Systems, and an Introduction to Chaos*. Elsevier, 3rd edition, 2013.

[17] Lauwerens Kuipers and Harald Niederreiter. *Uniform Distribution of Sequences*. Pure & Applied Mathematics Monograph. John Wiley & Sons, Nashville, TN, June 1974.

[18] James R Munkres. *Topology*. Pearson custom library. Pearson Education, London, England, 2 edition, July 2013.

[19] Ivan Niven. Uniform distribution of sequences of integers. *Transactions of the American Mathematical Society*, 98(1):52–61, 1961.

[20] Endre Pap. *Handbook of Measure Theory*. Elsevier Science, Amsterdam, Netherlands, October 2002.

[21] Kalyanapuram R. Parthasarathy. *Probability Measures on Metric Spaces*. Elsevier, 1967.

[22] Andreas N Philippou, Alwyn F Horadam, and G E Bergum, editors. *Applications of Fibonacci numbers*. Springer, Dordrecht, Netherlands, 1988 edition, March 2013.

[23] Ralph A Raimi. The first digit problem. *The American Mathematical Monthly*, 83(7):521–538, 1976.

[24] Elias M Stein and Rami Shakarchi. *Real Analysis: Measure Theory, Integration, & Hilbert Spaces*. Princeton lectures in analysis. Princeton University Press, Princeton, NJ, March 2005.

[25] Terence Tao. *Analysis I*, volume 37 of *Texts and Readings in Mathematics*. Springer, Singapore, 3rd edition, August 2016.

[26] Lawrence C Washington. Benford's Law for Fibonacci and Lucas Numbers. *The Fibonacci Quarterly*, 19(2):175–177, 1981.