Best finite approximations of Benford’s Law

Arno Berger and Chuang Xu
Mathematical and Statistical Sciences
University of Alberta
Edmonton, Alberta, CANADA
March 29, 2018

Abstract

For arbitrary Borel probability measures with compact support on the real line, characterizations are established of the best finitely supported approximations, relative to three familiar probability metrics (Lévy, Kantorovich, and Kolmogorov), given any number of atoms, and allowing for additional constraints regarding weights or positions of atoms. As an application, best (constrained or unconstrained) approximations are identified for Benford’s Law (logarithmic distribution of significands) and other familiar distributions. The results complement and extend known facts in the literature; they also provide new rigorous benchmarks against which to evaluate empirical observations regarding Benford’s Law.

Keywords. Benford’s Law, best uniform approximation, asymptotically best approximation, Lévy distance, Kantorovich distance, Kolmogorov distance.

MSC2010. 60B10, 60E15, 62E15.

1 Introduction

Given real numbers $b > 1$ and $x \neq 0$, denote by $S_b(x)$ the unique number in $[1,b[$ such that $|x| = S_b(x)b^k$ for some (necessarily unique) integer $k$; for convenience, let $S_b(0) = 0$. The number $S_b(x)$ often is referred to as the base-$b$ significand of $x$, a terminology particularly well-established in the case of $b$ being an integer. (Unlike in much of the literature [2, 4, 19, 34], the case of integer $b$ does not carry special significance in this article.) A Borel probability measure $\mu$ on $\mathbb{R}$ is Benford base $b$, or $b$-Benford for short, if

$$\mu(\{x \in \mathbb{R} : S_b(x) \leq s\}) = \frac{\log s}{\log b} \quad \forall s \in [1,b[;$$

(1.1)

here and throughout, log denotes the natural logarithm. Benford probabilities (or random variables) exhibit many interesting properties and have been studied extensively [11, 14, 20, 23, 29]. They provide one major pathway into the study of Benford’s Law, an intriguing, multi-faceted
phenomenon that attracts interest from a wide range of disciplines; see, e.g., [4] for an introduction, and [25] for a panorama of recent developments. Specifically, denoting by $\beta_b$ the Borel probability measure with

$$\beta_b([1,s]) = \frac{\log s}{\log b} \quad \forall s \in [1,b[,$$

note that $\mu$ is $b$-Benford if and only if $\mu \circ S_b^{-1} = \beta_b$.

Historically, the case of decimal (i.e., base-10) significands has been the most prominent, with early empirical studies on the distribution of decimal significands (or significant digits) going back to Newcomb [27] and Benford [2]. If $\mu$ is 10-Benford, note that in particular

$$\mu(\{x \in \mathbb{R} : \text{leading decimal digit of } x = D\}) = \frac{\log(1 + D^{-1})}{\log 10} \quad \forall D = 1, \ldots, 9. \quad (1.2)$$

For theoretical as well as practical reasons, mathematical objects such as random variables or sequences, but also concrete, finite numerical data sets that conform, at least approximately, to (1.1) or (1.2) have attracted much interest [11, 23, 34, 35]. Time and again, Benford’s Law has emerged as a perplexingly prevalent phenomenon. One popular approach to understand this prevalence seeks to establish (mild) conditions on a probability measure that make (1.1) or (1.2) hold with good accuracy, perhaps even exactly [7, 13, 14, 15, 29]. It is the goal of the present article to provide precise quantitative information for this approach.

Concretely, notice that while a finitely supported probability measure, such as, e.g., the empirical measure associated with a finite data set [5], may conform to the first-digit law (1.2), it cannot possibly satisfy (1.1) exactly. For such measures, therefore, it is natural to quantify, as accurately as possible, the failure of equality in (1.1), that is, the discrepancy between $\mu \circ S_b^{-1}$ and $\beta_b$. Utilizing three different familiar metrics $d_*$ on probabilities (Lévy, Kantorovich, and Kolmogorov metrics; see Section 2 for details), the article does this in a systematic way: For every $n \in \mathbb{N}$, the value of $\min_\nu d_*(\beta_b, \nu)$ is identified, where $\nu$ is assumed to be supported on no more than $n$ atoms (and may be subject to further restrictions such as, e.g., having only atoms of equal weight, as in the case of empirical measures); the minimizers of $d_*(\beta_b, \nu)$ are also characterized explicitly.

The scope of the results presented herein, however, extends far beyond Benford probabilities. In fact, a general theory of best (constrained or unconstrained) $d_*$-approximations is developed. As far as the authors can tell, no such theories exist for the Lévy and Kolmogorov metrics — unlike in the case of the Kantorovich metric where it (mostly) suffices to rephrase pertinent known facts [17, 36]. Once the general results are established, the desired quantitative insights for Benford probabilities are but straightforward corollaries. (Even in the context of Kantorovich distance, the study of $\beta_b$ yields a rare new, explicit example of an optimal quantizer [17].)

In particular, it turns out that, under all the various constraints considered here, the limit $Q_* = \lim_{n \to \infty} n \min_\nu d_*(\beta_b, \nu)$ always exists, is finite and positive, and can be computed more or less explicitly. This greatly extends earlier results, notably of [5], and also suggests that $n^{-1}Q_*$ may be an appropriate quantity against which to evaluate the many heuristic claims of closeness to Benford’s Law for empirical data sets found in the literature [3, 25, 26].

The main results in this article, then, are existence proofs and characterizations for the minimizers of $d_*(\mu, \nu)$ for arbitrary (compactly supported) probability measures $\mu$, as provided
by Theorems 3.5, 3.6, 4.1, 5.1, and 5.4 (where additional constraints are imposed on the sizes or locations of the atoms of \( \nu \)), as well as by Theorems 3.12 and 5.6 (where such constraints are absent). As suggested by the title, this work aims primarily at a precise analysis of conformance to Benford’s Law (or the lack thereof). Correspondingly, much attention is paid to the special case of \( \mu = \beta b \), leading to explicit descriptions of best (constrained or unconstrained) approximations of the latter (Corollaries 3.14, 4.4, and 5.9) and the exact asymptotics of \( d_*(\beta b, \nu) \). As indicated earlier, however, the main results are much more general. To emphasize this fact, two other simple but illustrative examples of \( \mu \) are repeatedly considered as well (though in less detail than \( \beta b \)), namely the familiar \( \text{Beta}(2, 1) \) distribution and the (perhaps less familiar) inverse Cantor distribution. It turns out that while the former is absolutely continuous (w.r.t. Lebesgue measure) and its best approximations behave like those of \( \beta b \) in most respects (Examples 3.9, 3.15, 4.9, and 5.10), the latter is discrete and the behaviour of its best approximations is more delicate (Examples 3.10, 3.16, 4.10, and 5.11). Even with only a few details mentioned, these examples will help the reader appreciate the versatility of the main results.

The organization of this article is as follows: Section 2 reviews relevant basic properties of one-dimensional probabilities and the three main probability metrics used throughout. Each of the Sections 3 to 5 then is devoted specifically to one single metric. In each section, the problem of best (constrained or unconstrained) approximation by finitely supported probability measures is first addressed in complete generality, and then the results are specialized to \( \beta b \) as well as other concrete examples. Section 6 summarizes and discusses the quantitative results obtained, and also mentions a few natural questions for subsequent studies.

2 Probability metrics

Throughout, let \( I \subset \mathbb{R} \) be a compact interval with Lebesgue measure \( \lambda(I) > 0 \), and \( \mathcal{P} \) the set of all Borel probability measures on \( I \). Associate with every \( \mu \in \mathcal{P} \) its distribution function \( F_\mu : \mathbb{R} \to \mathbb{R} \), given by

\[
F_\mu(x) = \mu(\{ y \in I : y \leq x \}) \quad \forall \ x \in \mathbb{R},
\]

as well as its (upper) quantile function \( F_\mu^{-1} : [0, 1[ \to \mathbb{R} \), given by

\[
F_\mu^{-1}(x) = \begin{cases} 
\min I & \text{if } 0 \leq x < \mu(\{\min I\}), \\
\sup \{ y \in I : F_\mu(y) \leq x \} & \text{if } \mu(\{\min I\}) \leq x < 1.
\end{cases}
\] (2.1)

Note that \( F_\mu \) and \( F_\mu^{-1} \) both are non-decreasing, right-continuous, and bounded. The support of \( \mu \), denoted \( \text{supp} \mu \), is the smallest closed subset of \( I \) with \( \mu \)-measure 1. Endowed with the weak topology, the space \( \mathcal{P} \) is compact and metrizable.

Three important different metrics on \( \mathcal{P} \) are discussed in detail in this article; for a panorama of other metrics the reader is referred, e.g., to [16, 32] and the references therein. Given probabilities \( \mu, \nu \in \mathcal{P} \), their Lévy distance is

\[
d_*(\mu, \nu) = \omega \inf \{ y \geq 0 : F_\mu(\cdot - y) - y \leq F_\nu \leq F_\mu(\cdot + y) + y \},
\] (2.2)
with $\omega = \max\{1, \lambda(I)\}/\lambda(I)$; their $L^{r}$-Kantorovich (or transport) distance, with $r \geq 1$, is

$$
d_{r}(\mu, \nu) = \lambda(I)^{-1} \left( \int_{0}^{1} |F_{\mu}^{-1}(y) - F_{\nu}^{-1}(y)|^{r} \, dy \right)^{1/r} = \lambda(I)^{-1} \| F_{\mu}^{-1} - F_{\nu}^{-1} \|_{r}; \tag{2.3}
$$

and their Kolmogorov (or uniform) distance is

$$
d_{K}(\mu, \nu) = \sup_{x \in \mathbb{R}} |F_{\mu}(x) - F_{\nu}(x)| = \| F_{\mu} - F_{\nu} \|_{\infty}.
$$

Henceforth, the symbol $d_{s}$ summarily refers to any of $d_{L}, d_{r},$ and $d_{K}$. The (usual) normalizing factors in (2.2) and (2.3) guarantee that all three metrics are comparable numerically in that $\sup_{\mu, \nu \in \mathcal{P}} d_{s}(\mu, \nu) = 1$ in either case. Note that

$$
d_{1}(\mu, \nu) = \lambda(I)^{-1} \int_{1} \left| F_{\mu}(x) - F_{\nu}(x) \right| \, dx \quad \forall \, \mu, \nu \in \mathcal{P},
$$

by virtue of Fubini’s Theorem. The metrics $d_{L}$ and $d_{r}$ are equivalent: They both metrize the weak topology on $\mathcal{P}$, and hence are separable and complete. By contrast, the complete metric $d_{K}$ induces a finer topology and is non-separable. However, when restricted to $\mathcal{P}_{\text{cts}} := \{ \mu \in \mathcal{P} : \mu(\{x\}) = 0 \, \forall \, x \in I \}$, a dense $G_{\delta}$-set in $\mathcal{P}$, the metric $d_{K}$ does metrize the weak topology on $\mathcal{P}_{\text{cts}}$ and is separable. The values of $d_{L}, d_{r},$ and $d_{K}$ are not completely unrelated, as is easily checked,

$$
d_{1} \leq \frac{1 + \lambda(I)}{\omega \lambda(I)} d_{L}, \quad d_{r} \leq d_{s} \text{ (if } r \leq s), \quad d_{1} \leq d_{K}, \quad d_{L} \leq \omega d_{K}, \tag{2.4}
$$

and all bounds in (2.4) are best possible. Beyond (2.4), however, no relative bounds exist between $d_{L}, d_{r},$ and $d_{K}$ in general: If $\ast \neq 1$, $\ast \neq \circ$, and $(\ast, \circ) \notin \{(L, K), (r, s)\}$ with $r \leq s$ then

$$
\sup_{\mu, \nu \in \mathcal{P}, \mu \neq \nu} \frac{d_{s}(\mu, \nu)}{d_{1}(\mu, \nu)} = +\infty.
$$

Each metric $d_{s}$, therefore, captures a different aspect of $\mathcal{P}$ and deserves to be studied independently. To illustrate this further, let $I = [0, 1], \mu = \delta_{0} \in \mathcal{P}$, and $\mu_{k} = (1 - k^{-1}) \delta_{0} + k^{-1} \delta_{1-k}$ for $k \in \mathbb{N}$; here and throughout, $\delta_{a}$ denotes the Dirac (probability) measure concentrated at $a \in \mathbb{R}$. Then $\lim_{k \to \infty} d_{s}(\mu, \mu_{k}) = 0$, but the rate of convergence differs between metrics:

$$
d_{L}(\mu, \mu_{k}) = k^{-2}, \quad d_{r}(\mu, \mu_{k}) = k^{-2-1/r}, \quad d_{K}(\mu, \mu_{k}) = k^{-1} \quad \forall \, k \in \mathbb{N}.
$$

The goal of this article is first to identify, for each metric $d_{s}$ introduced earlier, the best finitely supported $d_{s}$-approximation(s) of any given $\mu \in \mathcal{P}$. The general results are then applied to Benford’s Law, as well as other concrete examples. Specifically, if $\mu = \beta_{b}$ for some $b > 1$ then it is automatically assumed that $I = [1, b]$. The following unified notation and terminology is used throughout: For every $n \in \mathbb{N}$, let $\Xi_{n} = \{ x \in \mathbb{R}^{n} : x_{1} \leq \ldots \leq x_{n} \}$, $\Pi_{n} = \{ p \in \mathbb{R}^{n} : p_{j} \geq 0, \sum_{j=1}^{n} p_{j} = 1 \}$, and for each $x \in \Xi_{n}$ and $p \in \Pi_{n}$ define $\delta_{x}^{p} = \sum_{j=1}^{n} p_{j} \delta_{x_{j}}$. For convenience, $x_{0} := -\infty$ and $x_{n+1} := +\infty$ for every $x \in \Xi_{n}$, as well as $P_{i} = \sum_{j=1}^{n} p_{j}$ for $i = 0, \ldots, n$ and $p \in \Pi_{n}$; note that $P_{0} = 0$ and $P_{n} = 1$. Henceforth, usage of the symbol $\delta_{x}^{p}$ tacitly assumes that $x \in \Xi_{n}$ and $p \in \Pi_{n}$, for some $n \in \mathbb{N}$ either specified explicitly or else clear from the context. Call $\delta_{x}^{p}$ a best $d_{s}$-approximation of $\mu \in \mathcal{P}$, given $x \in \Xi_{n}$ if

$$
d_{s}(\mu, \delta_{x}^{p}) \leq d_{s}(\mu, \delta_{x}^{q}) \quad \forall \, q \in \Pi_{n}.
$$
Similarly, call $\delta_p^n$ a best $d_*$-approximation of $\mu$, given $p \in \Pi_n$ if
\[
d_* (\mu, \delta_p^n) \leq d_* (\mu, \delta_q^n) \quad \forall \, y \in \Xi_n.
\]
Denote by $\delta^*_{\xi}$ and $\delta_p^n$ any best $d_*$-approximation of $\mu$, given $x$ and $p$, respectively. Best $d_*$-approximations, given $p = u_n = (n^{-1}, \ldots, n^{-1})$ are referred to as best uniform $d_*$-approximations, and denoted $\delta^{u_*}$. Finally, $\delta_p^n$ is a best $d_*$-approximation of $\mu \in \mathcal{P}$, denoted $\delta^{u_*}$, if
\[
d_* (\mu, \delta_p^n) \leq d_* (\mu, \delta_q^n) \quad \forall \, y \in \Xi_n, q \in \Pi_n.
\]
Notice that usage of the symbols $\delta^*_{\xi}$, $\delta_p^n$, and $\delta^{u_*}$ always refers to a specific metric $d_*$ and probability measure $\mu \in \mathcal{P}$, both usually clear from the context.

Information theory sometimes refers to $d_* (\mu, \delta^{u_*})$ as the $n$-th quantization error, and to $\lim_{n \to \infty} nd_* (\mu, \delta^{u_*})$, if it exists, as the quantization coefficient of $\mu$; see, e.g., [17]. By analogy, $d_* (\mu, \delta^{u_*})$ and $\lim_{n \to \infty} nd_* (\mu, \delta^{u_*})$, if it exists, may be called the $n$-th uniform quantization error and the uniform quantization coefficient, respectively.

3 Lévy approximations

This section identifies best finitely supported $d_L$-approximations (constrained or unconstrained) of a given $\mu \in \mathcal{P}$. To do this in a transparent way, it is helpful to first consider more generally a few elementary properties of non-decreasing functions. These properties are subsequently specialized to either $F_\mu$ or $F_\mu^{-1}$.

Throughout, let $f : \mathbb{R} \to \overline{\mathbb{R}}$ be non-decreasing, and define $f(\pm \infty) = \lim_{x \to \pm \infty} f(x) \in \overline{\mathbb{R}}$, where $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ denotes the extended real line with the usual order and topology. Associate with $f$ two non-decreasing functions $f_\pm : \mathbb{R} \to \overline{\mathbb{R}}$, defined as $f_\pm (x) = \lim_{\varepsilon \downarrow 0} f(x \pm \varepsilon)$. Clearly, $f_-$ is left-continuous whereas $f_+$ is right-continuous, with $f_- (-\infty) = f(-\infty)$, $f_+(+\infty) = f(+\infty)$, as well as $f_- \leq f \leq f_+$, and $f_-(x) \leq f_+(y)$ whenever $x < y$; in particular, $f_-(x) = f_+(x)$ if and only if $f$ is continuous at $x$. The (upper) inverse function $f^{-1} : \mathbb{R} \to \overline{\mathbb{R}}$ is given by
\[
f^{-1}(t) = \sup \{x \in \mathbb{R} : f(x) \leq t\} \quad \forall \, t \in \mathbb{R};
\]
by convention, $\sup \emptyset := -\infty$ (and $\inf \emptyset := +\infty$). Note that (2.1) is consistent with this notation. For what follows, it is useful to recall a few basic properties of inverse functions; see, e.g., [36] Sec.3] for details.

**Proposition 3.1.** Let $f : \mathbb{R} \to \overline{\mathbb{R}}$ be non-decreasing. Then $f^{-1}$ is non-decreasing and right-continuous. Also, $(f_\pm)^{-1} = f^{-1}$, and $(f^{-1})^{-1} = f_+$.

Given two non-decreasing functions $f, g : \mathbb{R} \to \overline{\mathbb{R}}$, by a slight abuse of notation, and inspired by (2.2), let
\[
d_L (f, g) = \inf \{y \geq 0 : f(\cdot - y) - y \leq g \leq f(\cdot + y) + y\} \in [0, +\infty].
\]
For instance, $d_L (\mu, \nu) = \omega d_L (F_\mu, F_\nu)$ for all $\mu, \nu \in \mathcal{P}$. It is readily checked that $d_L$ is symmetric, satisfies the triangle inequality, and $d_L (f, g) > 0$ unless $f_- = g_-$, or equivalently, $f_+ = g_+$. Crucially, the quantity $d_L$ is invariant under inversion.
Proposition 3.2. Let \( f, g: \mathbb{R} \to \mathbb{R} \) be non-decreasing. Then \( d_\ell(f^{-1}, g^{-1}) = d_\ell(f, g) \).

Thus, for instance, \( d_\ell(\mu, \nu) = \omega d_\ell(F^{-1}_\mu, F^{-1}_\nu) \) for all \( \mu, \nu \in \mathcal{P} \). In general, the value of \( d_\ell(f, g) \) may equal \( +\infty \). However, if the set \( \{ f \neq g \} := \{ x \in \mathbb{R} : f(x) \neq g(x) \} \) is bounded then \( d_\ell(f, g) < +\infty \). Specifically, notice that \( \{ F \neq F^c \} \subset \mathbb{I} \) and \( \{ F^{-1} \neq F^{-1}^c \} \subset [0, 1] \) both are bounded for all \( \mu, \nu \in \mathcal{P} \).

Given a non-decreasing function \( f: \mathbb{R} \to \mathbb{R} \), let \( I \subset \mathbb{R} \) be any interval with the property that
\[
\ell_-(\sup I), -\ell_+(\inf I) < +\infty,
\]
and define an auxiliary function \( \ell_{f,I}: \mathbb{R} \to \mathbb{R} \) as
\[
\ell_{f,I}(x) = \inf \{ y \geq 0 : \ell_-(\sup I - y) - y \leq \ell_+(\inf I + y) + y \}.
\]
Note that for each \( x \in \mathbb{R} \), the set on the right equals \([a, +\infty[\) with the appropriate \( a \geq 0 \), and hence simply \( \ell_{f,I}(x) = a \). Clearly, \( \ell_{f,I} \leq \ell_{f,J} \) whenever \( J \subset I \). Also, for every \( a \in \mathbb{R} \), the function \( \ell_{f,(a]} \) is non-increasing on \([-\infty, f_-(a)]\), vanishes on \([f_-(a), f_+(a)]\), and is non-decreasing on \([f_+(a), +\infty[\). A few elementary properties of \( \ell_{f,I} \) are straightforward to check; they are used below to establish the main results of this section.

Proposition 3.3. Let \( f: \mathbb{R} \to \mathbb{R} \) be non-decreasing, and \( I \subset \mathbb{R} \) an interval satisfying (3.1). Then \( \ell_{f,I} \) is Lipschitz continuous, and
\[
0 \leq \ell_{f,I}(x) \leq |x| + \max \{0, \ell_-(\sup I), -\ell_+(\inf I)\} \quad \forall x \in \mathbb{R}.
\]
Moreover, \( \ell_{f,I} \) attains a minimal value
\[
\ell_{f,I}^* := \min_{x \in \mathbb{R}} \ell_{f,I}(x) = \min \{ y \geq 0 : \ell_-(\sup I - y) - y \leq \ell_+(\inf I + y) + y \} \geq 0
\]
which is positive unless \( \ell_-(\sup I) \leq \ell_+(\inf I) \).

For \( \mu \in \mathcal{P} \), note that (3.1) automatically holds if \( f = F_\mu \), or if \( f = F_\mu^{-1} \) and \( I \subset [0, 1] \). In these cases, therefore, \( \ell_{f,I} \) has the properties stated in Proposition 3.3 and \( \ell_{f,I}^* \leq \frac{1}{2} \).

When formulating the main results, the following quantities are useful: Given \( \mu \in \mathcal{P} \), \( n \in \mathbb{N} \), and \( x \in \Xi_n \), let
\[
L^*(x) = \max \left\{ \ell_{F_\mu, [-\infty, x, 1]}(0), \ell_{F_\mu, [x, 1, x, 2]}, \ldots, \ell_{F_\mu, [x, n-1, x, n]}, \ell_{F_\mu, [x, n, +\infty]}(1) \right\};
\]
similarly, given \( p \in \Pi_n \), let
\[
L^*(p) = \max_{j=1}^n \ell_{F_\mu^{-1}, [p_{j-1}, p_j]}^*.
\]
To illustrate these quantities for a concrete example, consider \( \mu = \beta_b \), where \( \ell_{F_\mu, [x, j, x, j+1]}^* \) is the unique solution of
\[
b^{2\ell} = \frac{x_{j+1} - x_j}{x_{j+1} + \ell} \quad j = 1, \ldots, n - 1,
\]
whereas \( \ell_{F_\mu, [-\infty, x, 1]}(0) \) and \( \ell_{F_\mu, [x, n, +\infty]}(1) \) solve \( b^{\ell} = x_1 - \ell \) and \( b^{\ell} = b/(x_n + \ell) \), respectively. (Recall that \( 1 \leq x, 1 \leq \ldots \leq x_n \leq b \).) Similarly, \( \ell_{F_\mu^{-1}, [p_{j-1}, p_j]}^* \) is the unique solution of
\[
2\ell = b^{P_j - \ell} - b^{P_j - 1 + \ell} \quad j = 1, \ldots, n.
in particular, $j \mapsto \ell_{F_{\mu,n}^{-1}}^{\prime}(j-1/n,j/n)$ is increasing, and hence $L_\bullet(u_n)$ is the unique solution of

$$2L = b^{1-L} - b^{1+L-1/n}. \quad (3.2)$$

By using functions of the form $\ell_{f,i}$, the value of $d_L(\mu,\nu)$ can easily be computed whenever $\nu$ has finite support.

**Lemma 3.4.** Let $\mu \in \mathcal{P}$ and $n \in \mathbb{N}$. For every $x \in \Xi_n$ and $p \in \Pi_n$,

$$d_L(\mu, \delta^p) = \omega \max_{j=0}^n \ell_{F_{\mu,\nu},[x_j,x_{j+1}]}(P_j) = \omega \max_{j=1}^n \ell_{F_{\mu,\nu},[P_{j-1},P_j]}(x_j). \quad (3.3)$$

**Proof.** Label $x \in \Xi_n$ uniquely as

$$x_{j_0+1} = \ldots = x_{j_1} < x_{j_1+1} = \ldots = x_{j_2} < x_{j_2+1} = \ldots < \ldots = x_{j_{m-1}} < x_{j_{m-1}+1} = \ldots = x_{j_m},$$

with integers $i \leq j_i \leq n$ for $1 \leq i \leq m$, and $j_0 = 0$, $j_m = n$, and define $y \in \Xi_m$ and $q \in \Pi_m$ as $y_i = x_{j_i}$, and $q_i = P_{j_i} - P_{j_i-1}$, respectively, for $i = 1, \ldots, m$. For convenience, let $I_j = [x_j, x_{j+1}]$ for $j = 0, \ldots, n$, and $J_i = [y_i, y_{i+1}] = I_{j_i}$ for $i = 0, \ldots, m$. With this, $\delta^p_y = \delta^p_x$, and

$$\omega^{-1}d_L(\mu, \delta^p) = d_L(F_{\mu}, F_{\delta^p}) = \inf \{ t \geq 0 : F_{\mu}^{-1}(y_{i+1} - t) < q_i \leq F_{\mu}(y_i + t) + t, \forall i = 0, \ldots, m \}
= \max_{i=0}^n \ell_{F_{\mu},I_i}(Q_i)
\leq \max_{i=0}^n \ell_{F_{\mu},I_j}(P_j).$$

To prove the reverse inequality, pick any $j = 0, \ldots, n$. If $x_j < x_{j+1}$ then $I_j = J_i$ and $P_{j+1} = Q_i$, with the appropriate $i$, and hence $\ell_{F_{\mu},I_j}(P_j) = \ell_{F_{\mu},J_i}(Q_i)$. If $x_j = x_{j+1}$ then $I_j = \{y_i\}$ for some $i$. In this case either $P_j < F_{\mu}^{-1}(y_i)$ and $Q_i - 1 \leq P_j$, and hence $\ell_{F_{\mu},I_j}(P_j) = \ell_{F_{\mu},I_j}(Q_i) \leq \ell_{F_{\mu},J_i}(Q_{i-1})$; or $F_{\mu}^{-1}(y_i) \leq P_j \leq F_{\mu}(y_i)$, and hence $\ell_{F_{\mu},I_j}(P_j) = \ell_{F_{\mu},I_j}(P_j) = 0$; or $P_j > F_{\mu}(y_i)$ and $Q_i \geq P_j$, and hence $\ell_{F_{\mu},I_j}(P_j) = \ell_{F_{\mu},I_j}(Q_i) \leq \ell_{F_{\mu},I_j}(Q_i)$. In all three cases, therefore, $\omega^{-1}d_L(\mu, \delta^p) \geq \max_{j=0}^n \ell_{F_{\mu},I_j}(P_j)$, which establishes the first equality in (3.3). The second equality, a consequence of Proposition 3.2, is proved analogously. \[\square\]

Utilizing Lemma 3.4 it is straightforward to characterize the best finitely supported $d_L$-approximations of $\mu \in \mathcal{P}$ with prescribed locations.

**Theorem 3.5.** Let $\mu \in \mathcal{P}$ and $n \in \mathbb{N}$. For every $x \in \Xi_n$, there exists a best $d_L$-approximation of $\mu$, given $x$. Moreover, $d_L(\mu, \delta^p_x) = d_L(\mu, \delta^p_x^\bullet)$ if and only if, for every $j = 0, \ldots, n$,

$$x_j < x_{j+1} \implies \ell_{F_{\mu},[x_j,x_{j+1}]}(P_j) \leq L^\bullet(x),$$
(3.4)

and in this case $d_L(\mu, \delta^p_x) = \omega L^\bullet(x)$. 

7
Proof. Fix $\mu \in \mathcal{P}$, $n \in \mathbb{N}$, and $x \in \Xi_n$. As in the proof of Lemma 3.4, write $I_j = [x_j, x_{j+1}]$ for convenience. By (3.3), for every $p \in \Pi_n$,

$$d_L(\mu, \delta^p) = \omega \max_{j=0}^n \ell_{F_{\mu}, I_j}(P_j) \geq \omega \max\{\ell_{F_{\mu}, I_0}(0), \ell_{F_{\mu}, I_1}, \ldots, \ell_{F_{\mu}, I_{n-1}}, \ell_{F_{\mu}, I_n}(1)\} = \omega L^\bullet(x).$$

As seen in the proof of Lemma 3.4, validity of (3.4) implies $\ell_{F_{\mu}, [x_j, x_{j+1}]}(P_j) \leq L^\bullet(x)$ for all $j = 0, \ldots, n$. Thus $\delta^p$ is a best $d_L$-approximation of $\mu$, given $x$, whenever (3.4) holds, i.e., the latter is sufficient for optimality. On the other hand, consider $q \in \Pi_n$ with

$$Q_{j} = \frac{1}{2} \left( F_{\mu, \omega}(x_{j+1} - L^\bullet(x)) + F_{\mu}(x_{j} + L^\bullet(x)) \right) \quad \forall j = 1, \ldots, n - 1.$$ 

Note that $q$ is well-defined, since $j \mapsto Q_{j}$ is non-decreasing, and $0 \leq Q_{j} \leq 1$ for all $j = 1, \ldots, n - 1$. Moreover, by the definition of $L^\bullet(x)$,

$$\ell_{F_{\mu}, I_j}(Q_{j}) \leq L^\bullet(x) \quad \forall j = 0, \ldots, n,$$

and hence $d_L(\delta^p, \mu) = \omega L^\bullet(x)$. This shows that best $d_L$-approximations of $\mu$, given $x$, do exist, and (3.4) also is necessary for optimality. \qed

Best finitely supported $d_L$-approximations of any $\mu \in \mathcal{P}$ with prescribed weights can be characterized in a similar manner. By virtue of (3.3), the proof of the following is completely analogous to the proof of Theorem 3.5 above.

**Proposition 3.6.** Let $\mu \in \mathcal{P}$ and $n \in \mathbb{N}$. For every $p \in \Pi_n$, there exists a best $d_L$-approximation of $\mu$, given $p$. Moreover, $d_L(\mu, \delta^p) = d_L(\mu, \delta^q)$ if and only if, for every $j = 1, \ldots, n$,

$$P_{j-1} < P_j \implies \ell_{F_{\mu}, [P_{j-1}, P_j]}(x_j) \leq L^\bullet(p),$$

and in this case $d_L(\mu, \delta^p) = \omega L^\bullet(p)$.

**Remark 3.7.** (i) With $f, I$ as in Proposition 3.3 for every $a \in \mathbb{R}$ the set $\{\ell_{f, I} \leq a\}$ is a (possibly empty or one-point) interval. Thus, conditions (3.4) and (3.5) are very similar in spirit to the requirements of [36] Thm.5.1 and 5.5], restated in Proposition 4.1 below, though the latter may be quite a bit easier to work with in concrete calculations.

(ii) Note that if $n = 1$ then (3.4) holds automatically, whereas (3.5) shows that $d_L(\mu, \delta_a)$ is minimal precisely if the function $\ell_{F_{\mu, \omega}[0,1]}^{-1}$ attains its minimal value at $a$.

As a corollary, Proposition 3.6 identifies all best uniform $d_L$-approximations of $\beta_b$ with $b > 1$. Recall that $I = [1, b]$, and hence $\omega = \frac{\max\{b, 2\} - 1}{b - 1} =: \omega_b$ in this case.

**Corollary 3.8.** Let $b > 1$ and $n \in \mathbb{N}$. Then $\delta_{\omega_b n}$ is a best uniform $d_L$-approximation of $\beta_b$ if and only if

$$b^{j/n-L} - L \leq x_j \leq b^{(j-1)/n+L} + L \quad \forall j = 1, \ldots, n,$$

where $L$ is the unique solution of (5.2); in particular, $\#\text{supp}^\bullet \delta_{\omega_b n} = n$. Moreover, $d_L(\beta_b, \delta_{\omega_b n}) = \omega_b L$, and

$$\lim_{n \to \infty} n d_L(\beta_b, \delta_{\omega_b n}) = \frac{\max\{b, 2\} - 1}{2b - 2} \cdot \frac{b \log b}{1 + b \log b}.$$
Example 3.9. Consider the Beta(2, 1) distribution on $I = [0, 1]$, i.e., let $F_\mu(x) = x^2$ for all $x \in I$. Given $n \in \mathbb{N}$, it is straightforward to check that, analogously to (3.2), $L_\bullet(u_n)$ is the unique solution of

$$L \sqrt{\frac{2}{n} - 4L^2} = \frac{1}{2n} - L,$$

and $\delta^u_n$ with $x \in \Xi_n$ is a best uniform $d_L$-approximation of $\mu$ if and only if

$$\sqrt{\frac{j}{n} - L - L \leq x_j \leq \sqrt{\frac{j-1}{n} + L + L} \quad \forall j = 1, \ldots, n.}$$

Moreover, $d_L(\mu, \delta^u_n) = L$, and (3.6) yields that $\lim_{n \to \infty} nd_L(\mu, \delta^u_n) = \frac{1}{2}$. Unlike in the case of $\beta_0$, it is possible to have $\#\text{supp} \delta^u_n < n$ whenever $n \geq 10$.

Example 3.10. Let again $I = [0, 1]$ and consider $\mu \in \mathcal{P}$ with $\mu(\{i2^{-m}\}) = 3^{-m}$ for every $m \in \mathbb{N}$ and every odd $1 \leq i < 2^m$. Thus $\mu$ is a discrete measure with $\text{supp} \mu = I$. In fact, $\mu$ simply is the inverse Cantor distribution, in the sense that $F_\mu^{-1}(x) = F_\nu(x)$ for all $x \in I$, where $\nu$ is the log 2/log 3-dimensional Hausdorff measure on the classical Cantor middle-thirds set. Given $n \in \mathbb{N}$, Proposition 3.6 guarantees the existence of a best uniform $d_L$-approximation of $\mu$, though the explicit value of $L_\bullet(u_n)$ is somewhat cumbersome to determine. Still, utilizing the self-similarity of $F_\mu^{-1}$, one finds that

$$\frac{1}{216} \leq \liminf_{n \to \infty} nd_L(\mu, \delta^u_n) \leq \frac{1}{3}, \quad \limsup_{n \to \infty} nd_L(\mu, \delta^u_n) = \frac{1}{2}. \quad (3.7)$$

Thus $(n^{-1})$ is the precise rate of decay of $(d_L(\mu, \delta^u_n))$, just as in the case of $\beta_0$ and Beta(2, 1), but unlike for the latter, $\lim_{n \to \infty} nd_L(\mu, \delta^u_n)$ does not exist.

By combining Theorem 3.5 and Proposition 3.6 it is possible to characterize the best $d_L$-approximations of $\mu \in \mathcal{P}$ as well, that is, to identify the minimizers of $\nu \mapsto d_L(\mu, \nu)$ subject only to the requirement that $\#\text{supp} \nu \leq n$. To this end, associate with every non-decreasing function $f : \mathbb{R} \to \mathbb{R}$ and every number $a \geq 0$ a map $T_{f, a} : \mathbb{R} \to \mathbb{R}$, according to

$$T_{f, a}(x) = f_+((x + a) + 2a) + a \quad \forall x \in \mathbb{R}.$$ 

For every $n \in \mathbb{N}$, denote by $T_{f, a}^{[n]}$ the $n$-fold composition of $T_{f, a}$ with itself. The following properties of $T_{f, a}$ are readily verified.

Proposition 3.11. Let $f : \mathbb{R} \to \mathbb{R}$ be non-decreasing, $a \geq 0$, and $n \in \mathbb{N}$. Then $T_{f, a}^{[n]}$ is non-decreasing and right-continuous. Also, $a \mapsto T_{f, a}^{[n]}(x)$ is increasing and right-continuous for every $x \in \mathbb{R}$, and if $x \leq a + f(+\infty)$ then the sequence $(T_{f, a}^{[k]}(x))$ is non-decreasing.

To utilize Proposition 3.11 for the $d_L$-approximation problem, let $f = F_\mu$ with $\mu \in \mathcal{P}$. Then $(T_{F_\mu, a}^{[k]}(0))$ is non-decreasing; in fact, $\lim_{k \to \infty} T_{F_\mu, a}^{[k]}(0) = a + 1$. On the other hand, given $n \in \mathbb{N}$, clearly $T_{F_\mu, a}^{[n]}(0) \geq 1$ for all $a \geq 1$, and hence

$$L_\bullet^{*, n} := \min \left\{ a \geq 0 : T_{F_\mu, a}^{[n]}(0) \geq 1 \right\} < +\infty.$$
Proof. To see that best with value for some $\nu$ of $\nu$ from which it is easily deduced that $\nu$ whenever $\mu \in \mathcal{P}$ and $n \in \mathbb{N}$. There exists a best $d_L$-approximation of $\mu$, and $d_L(\mu, \delta^{*n}) = \omega L^{*n}$. Moreover, for every $x \in \Xi_n$ and $p \in \Pi_n$, the following are equivalent:

(i) $d_L(\mu, \delta^p) = d_L(\mu, \delta^{*n})$;

(ii) all implications in (3.4) are valid with $L^*(x)$ replaced by $L^{*n}$;

(iii) all implications in (3.5) are valid with $L_n(p)$ replaced by $L^{*n}$.

Proof. To see that best $d_L$-approximations of $\mu$ do exist, simply note that the set $\{\nu \in \mathcal{P} : \#\text{supp} \nu \leq n\}$ is compact, and the function $\nu \mapsto d_L(\mu, \nu)$ is continuous, hence attains a minimal value for some $\nu = \delta^p$ with $x \in \Xi_n$ and $p \in \Pi_n$. Clearly, any such $\delta^p$ also is a best approximation of $\mu$, given $p$. By Proposition 3.6, therefore, $d_L(\mu, \delta^p) = \omega L_n(p)$, as well as

$$F_{\mu}^{-1}(P_{j-1} - L_n(p)) - L_n(p) \leq x \leq P_{j-1} + L_n(p).$$

whenever $P_{j-1} < P_j$, and indeed for every $j = 1, \ldots, n$. It follows that $P_j \leq T_{F_n, L_n(p)}(P_{j-1})$ for all $j$, and hence $1 = T_{F_n, L_n(0)}(0)$, that is, $L^{*n} = L_n(p)$. This shows that $d_L(\mu, \delta^p) \geq \omega L^{*n}$.

To establish the reverse inequality, let

$$m = \min \left\{i \geq 1 : T_{F_n, L^{*n}}(0) \geq 1 \right\}.$$ 

Clearly, $1 \leq m \leq n$, and $L^{*m} = L^{*n}$. Define $q \in \Pi_m$ via

$$Q_{i,i} = T_{F_n, L^{*n}}(0) \quad \forall i = 1, \ldots, m - 1.$$ 

Note that $i \mapsto Q_{i,i}$ is non-decreasing, and $0 \leq Q_{i,i} \leq 1$, so $q$ is well-defined. Also, consider $y \in \Xi_m$ with

$$y_{i,i} = \frac{1}{2} \left( F_{\mu}^{-1}(Q_{i,i} - L^{*m}) + F_{\mu}^{-1}(Q_{i,i-1} + L^{*m}) \right) \quad \forall i = 1, \ldots, m.$$ 

By the definitions of $L^{*m}$, $q$, and $y$,

$$\ell_{F_{\mu}^{-1}[Q_{i,i-1}, Q_{i,i}]}(y_{i,i}) \leq L^{*m} \quad \forall i = 1, \ldots, m.$$
and hence
\[
d_{L}(\mu, \delta_{\beta}^{n}) \leq d_{L}(\mu, \delta_{\mu}^{n}) = \omega \max_{1 \leq i \leq n} \ell_{F_{\mu}^{-1},[Q,\nu_{i},Q_{i}]}(y, i) \leq \omega L_{\beta}^{*,m} = \omega L_{\mu}^{*,n}.
\]
This shows that indeed \(d_{L}(\mu, \delta_{\beta}^{n}) = \omega L_{\beta}^{*,n}\) and also proves (i)\(\Rightarrow\)(iii). The implication (i)\(\Rightarrow\)(ii) follows by a similar argument. That, conversely, either of (ii) and (iii) implies (i) is evident from (3.3), together with the fact that, as seen in the proof of Lemma 3.3 above, validity of (3.4) and (3.5) implies \(\max_{j=0}^{n} \ell_{F_{\mu}^{-1}[x_{j},x_{j+1}]}(P_{\beta}) \leq L_{\beta}(x)\) and \(\max_{j=1}^{n} \ell_{F_{\mu}^{-1}[P_{\beta},P_{j}]}(x, j) \leq L_{\mu}(p)\), respectively.

**Remark 3.13.** (i) The above proof of Theorem 3.12 shows that in fact
\[
L_{\mu}^{*,n} = \min_{n \in \mathbb{N}} L_{\mu}(x) = \min_{P \in \Pi_{n}} L_{\mu}(p).
\]
(ii) Theorem 3.12 is similar to classical one-dimensional quantization results as presented, e.g., in [17, Sec.5.2]. What makes the theorem (and its analogue, Theorem 5.6 in Section 5) particularly appealing is that its conditions (ii) and (iii) not only are necessary for optimality, but also sufficient. By contrast, it is well known that sufficient conditions for best \(d_{L}\)-approximations may be hard to come by in general; see, e.g., [17, Sec.4.1], and also Proposition 4.1(iii) below, regarding the case of \(* = 1\).

When specialized to \(\mu = \beta_{b}\), Theorem 3.12 yields the best finitely supported \(d_{L}\)-approximations of Benford’s Law.

**Corollary 3.14.** Let \(b > 1\) and \(n \in \mathbb{N}\). Then the best \(d_{L}\)-approximation of \(\beta_{b}\) is \(\delta_{\beta}^{n}\), with
\[
x_{j} = \frac{(2^{j-1})L + 2L^{2jL-1}}{b^{2jL-1}} - L = b^{j-L} - L,
\]
\[
P_{j} = \frac{1}{\log b} \log \left(\frac{b^{(2j-1)L} + 2L^{2jL-1}}{b^{2jL-1}}\right) + L = \frac{\log(x_{j} + L)}{\log b} + L,
\]
for all \(j = 1, \ldots, n\), where \(L\) is the unique solution of (3.8); in particular, \(\# \text{supp } \delta_{\beta}^{n} = n\). Moreover, \(d_{L}(\beta_{b}, \delta_{\beta}^{n}) = \omega_{b}L\), and
\[
\lim_{n \to \infty} nd_{L}(\beta_{b}, \delta_{\beta}^{n}) = \frac{\max\{b, 2\}}{2b - 2} : \frac{\log(1 + b \log b) - \log(1 + \log b)}{\log b}.
\]

To compare this to Corollary 3.8 note that \(P_{j} \neq j/n\) whenever \(n \geq 2\), and then the \(n\)-th quantization error \(d_{L}(\beta_{b}, \delta_{\beta}^{n})\) is smaller than the \(n\)-th uniform quantization error \(d_{L}(\beta_{b}, \delta_{\mu}^{n})\). The \(d_{L}\)-quantization coefficient of \(\beta_{b}\) also is smaller than its uniform counterpart, since
\[
\frac{\log(1 + b \log b) - \log(1 + \log b)}{\log b} < \frac{b \log b}{1 + b \log b} \quad \forall \ b > 1.
\]

**Example 3.15.** For \(\mu = \text{Beta}(2, 1)\), Theorem 3.12 yields a unique best \(d_{L}\)-approximation. Although the equation determining \(L_{\mu}^{*,n}\) is less transparent than (3.8), it can be shown that \(\lim_{n \to \infty} nd_{L}(\mu, \delta_{\mu}^{n}) = \frac{1}{\tau}(2 - \log 3) < \frac{1}{\tau}\).
10
3
2
1
0
1
10

Figure 1: The best $d_L$-approximation (solid red line) of $\beta_{10}$ is unique, whereas best uniform $d_L$-approximations (broken red lines) are not; see Corollaries 3.14 and 3.8 respectively.

Example 3.16. For the inverse Cantor distribution, a best $d_L$-approximation exists by Theorem 3.12 and utilizing the self-similarity of $F_{\mu}^{-1}$, it is possible to derive estimates such as

$$\frac{1}{216} \leq n^{\log 3/\log 2} d_L(\mu, \delta_n) \leq 3 \quad \forall n \in \mathbb{N},$$

(4.1)

which shows that $(d_L(\mu, \delta_n^n))$ decays like $(n^{-\log 3/\log 2})$, and hence faster than in the case of $\beta_b$ and $\text{Beta}(2, 1)$.

4 Kantorovich approximations

This section studies best finitely supported $d_r$-approximations of Benford’s Law. Mostly, the results are special cases of more general facts taken from the authors’ comprehensive study on $d_r$-approximations [36].

4.1 $d_1$-approximations

With $d_L$ replaced by $d_1$, the main results of the previous section have the following analogues, stated here for the reader’s convenience; see [36 Sec.5] for details.

Proposition 4.1. Let $\mu \in \mathcal{P}$ and $n \in \mathbb{N}$.

(i) For every $x \in \Xi_n$, there exists a best $d_1$-approximation of $\mu$, given $x$. Moreover, $d_1(\mu, \delta_x^n) = d_1(\mu, \delta_x^*)$ if and only if, for every $j = 0, \ldots, n$,

$$x_j < x_{j+1} \implies F_{\mu} (\frac{1}{2}(x_j + x_{j+1})) \leq P_j \leq F_{\mu} (\frac{1}{2}(x_j + x_{j+1})).$$

(4.1)
(ii) For every \( p \in \Pi_n \), there exists a best \( d_1 \)-approximation of \( \mu \), given \( p \). Moreover, \( d_1 (\mu, \delta^p_x) = d_1 (\mu, \delta^x) \) if and only if, for every \( j = 1, \ldots, n \),

\[
P_{j-1} < P_j \implies F_{\mu}^{-1} \left( \frac{1}{2} (P_{j-1} + P_j) \right) \leq x_j \leq F_{\mu}^{-1} \left( \frac{1}{2} (P_{j-1} + P_j) \right).
\]

(4.2)

(iii) There exists a best \( d_1 \)-approximation of \( \mu \), and if \( d_1 (\mu, \delta^p_x) = d_1 (\mu, \delta^x) \) then (4.1) and (4.2) are valid for every \( j = 1, \ldots, n \).

**Remark 4.2.** Though the phrasing of Proposition 4.1 emphasizes its analogy to Theorem 3.5 (and also to Theorem 5.1 below), there nevertheless is a subtle difference: While in (3.4) and (5.1) it can equivalently be stipulated that, respectively, \( \ell F_{\mu}(x, j, x, j+1) \leq \mathbf{L}(x) \) and \( F_{\mu}^{-1} (x, j+1) - K(x) \leq P_j \leq F_{\mu}(x, j) + K(x) \) for all \( j = 0, \ldots, n \), simple examples show that the “only if” part of Proposition 4.1(i) may fail, should (4.1) be replaced by

\[
F_{\mu}^{-1} \left( \frac{1}{2} (x, j + x, j+1) \right) \leq P_j \leq F_{\mu} \left( \frac{1}{2} (x, j + x, j+1) \right) \quad \forall j = 0, \ldots, n.
\]

Similar observations pertain to Proposition 4.1(ii) vis-à-vis Proposition 3.6 and Theorem 5.4.

Proposition 4.1 immediately yields the existence of unique best uniform \( d_1 \)-approximations of \( \beta_b \); see also [5, Cor. 2.10].

**Corollary 4.3.** Let \( b > 1 \) and \( n \in \mathbb{N} \). Then the best uniform \( d_1 \)-approximation of \( \beta_b \) is \( \delta^n_x \), with \( x, j = b^{(2j-1)/(2n)} \) for all \( j = 1, \ldots, n \), and \( \text{supp} \delta^n_x = n \). Moreover, \( d_1 (\beta_b, \delta^n_x) = \frac{1}{\log b} \log b/4n \), and \( \lim_{n \to \infty} nd_1 (\beta_b, \delta^n_x) = \frac{1}{4} \).

Proof. By Proposition 4.1(ii), \( x, j = b^{(2j-1)/(2n)} \) for all \( j = 1, \ldots, n \), and

\[
nd_1 (\beta_b, \delta^n_x) = \frac{n}{b-1} \sum_{j=1}^{n} \int_{(j-1)/n}^{j/n} \left| b^y - b^{(2j-1)/(2n)} \right| \, dy
\]

\[
= \frac{n}{b-1} \left( b^{1/(4n)} - b^{-1/(4n)} \right)^2 \sum_{j=1}^{n} b^{(2j-1)/(2n)} = \frac{n}{\log b} \tanh \left( \frac{\log b}{4n} \right) \xrightarrow{n \to \infty} \frac{1}{4}.
\]

Best (unconstrained) \( d_1 \)-approximations of \( \beta_b \) exist and are unique, too, by virtue of Proposition 4.1 and a direct calculation.

**Corollary 4.4.** Let \( b > 1 \) and \( n \in \mathbb{N} \). Then the best \( d_1 \)-approximation of \( \beta_b \) is \( \delta^p_x \), with

\[
x, j = \left( 1 + \frac{1 - j}{n} \left( b^{1/2} - 1 \right) \right) \left( 1 + \frac{j}{n} \left( b^{1/2} - 1 \right) \right),
\]

\[
P, j = \frac{2}{\log b} \log \left( 1 + \frac{j}{n} \left( b^{1/2} - 1 \right) \right),
\]

for all \( j = 1, \ldots, n \); in particular, \( \text{supp} \delta^n_x = n \). Moreover, \( d_1 (\beta_b, \delta^n_x) = \frac{1}{n \log b} \tanh \left( \frac{\log b}{4} \right) \).
Proof. Let $\delta^p_\beta$ be a best $d_1$-approximation. Then, by Proposition 4.1(iii),

$$b^{P,j} = \frac{x_j + x_{j+1}}{2} \quad \forall j = 1, \ldots, n - 1,$$

but also $x_j = b^{(P,j+1)}$ for all $j = 1, \ldots, n$, and hence $2b^{P,j}/2 = b^{P,j-1} + b^{P,j+1}/2$. Since $P_0 = 0$, $P_n = 1$, it follows that $b^{P,j}/2 = 1 + j(b^{1}/2 - 1)n^{-1}$ for all $j = 0, \ldots, n$. This yields the asserted unique $\delta^p_\beta$, and

$$d_1(\beta_b, \delta^p_n) = \frac{1}{b-1} \sum_{j=1}^{n} \int_{P_j-1}^{P_j} |b^y - x_j| \, dy = \frac{b - x_n - (x_1 - 1)}{(b-1) \log b}$$

$$= \frac{1}{n \log b} \tanh \left( \frac{\log b}{4} \right),$$

via a straightforward calculation.

Figure 2: The best (solid blue line) and best uniform (broken blue line) $d_1$-approximations of $\beta_{10}$ both are unique; see Corollaries 4.4 and 4.3, respectively. Coincidentally, best uniform $d_1$-approximations of $\beta_{10}$ are best $d_K$-approximations as well; see Corollary 5.9.

Remark 4.5. (i) Due to the highly non-linear nature of the optimality conditions (4.1) and (4.2), best $d_1$-approximations are rarely given by explicit formulae such as those in Corollary 4.4. Aside from Benford’s Law, the authors know of only two other families of continuous distributions that allow for similarly explicit formulae, namely uniform and (one- or two-sided) exponential distributions.

(ii) A popular family of metrics on $\mathcal{P}$ closely related to $d_1$ are the so-called Fortet–Mourier $r$-distances ($1 \leq r < +\infty$), given by

$$d_{FM_r}(\mu, \nu) = \int \max\{1, |y|\}^{r-1} |F_{\mu}(y) - F_{\nu}(y)| \, dy.$$

Like the Lévy and Kantorovich metrics, the Fortet–Mourier $r$-distance also metrizes the weak topology on $\mathcal{P}$. The reader is referred to [28, 31] for details on the mathematical background of
$d_{\text{FM}}$, and its use for stochastic optimization. Note that if $\mathbb{I} \subset [1, +\infty]$ then

$$d_{\text{FM}}(\mu, \nu) = \frac{\lambda(T(\mathbb{I}))}{r} d_1(\mu \circ T^{-1}, \nu \circ T^{-1})$$

with the homeomorphism $T : x \mapsto x'$ of $[1, +\infty]$. For instance, $\beta_b \circ T^{-1} = \beta_{r_b}$, and hence best (or best uniform) $d_{\text{FM}}$-approximations of $\beta_b$ can easily be identified using Corollary 4.4 (or 4.3).

4.2 $d_r$-approximations ($1 < r < +\infty$)

Similarly to the case of $r = 1$, [36] Thm.5.15 guarantees that, given any $n \in \mathbb{N}$, there exists a (unique) best uniform $d_r$-approximation $\delta^{u_n}$ of $\beta_b$. Except for $r = 2$, however, no explicit formula seems to be available for $\delta^{u_n}$. It is desirable, therefore, to at least identify asymptotically best uniform $d_r$-approximations, that is, a sequence $(x_n)$ with $x_n \in \Xi_n$ for all $n \in \mathbb{N}$ such that

$$\lim_{n \to \infty} \frac{d_r(\beta_b, \delta^{u_n})}{d_r(\beta_b, \delta_{\beta})} = 1.$$ 

Usage of [36] Thm.5.15 accomplishes this and also yields the uniform $d_r$-quantization coefficient of $\beta_b$. (Notice that, as $r \downarrow 1$, the latter is consistent with Corollary 4.3.)

**Proposition 4.6.** Let $b, r > 1$. Then $(\delta^{u_n}_x)$, with $x_{n,j} = b^{(2j-1)/(2n)}$ for all $n \in \mathbb{N}$ and $j = 1, \ldots, n$, is a sequence of asymptotically best uniform $d_r$-approximations of $\beta_b$. Moreover,

$$\lim_{n \to \infty} n d_r(\beta_b, \delta^{u_n}) = \frac{(\log b)^{1-1/r}}{2(b-1)} \left( \frac{b^r - 1}{r(r+1)} \right)^{1/r}.$$ 

The remainder of this section studies best $d_r$-approximations of $\beta_b$. In general, the question of uniqueness of best $d_r$-approximations is a difficult one, for which only partial answers exist; see, e.g., [17] Sec.5]. Specifically, $\beta_b$ does not seem to satisfy any known condition (such as, e.g., log-concavity) that would guarantee uniqueness. However, uniqueness can be established via a direct calculation.

**Theorem 4.7.** Let $b, r > 1$ and $n \in \mathbb{N}$. There exists a unique best $d_r$-approximation $\delta^{u_n}_x$ of $\beta_b$, and $\# \text{supp} \delta^{u_n}_x = n$.

**Proof.** Existence follows as in Theorem 3.12; alternatively, see [17] Sec.4.1 or [36] Prop.5.22. To avoid trivialities, henceforth assume $n \geq 2$. If $d_r(\beta_b, \delta^x) = d_r(\beta_b, \delta^{u_n}_x)$, then by [36] Thm.5.23,

$$b^{P_j} = \frac{x_j + x_{j+1}}{2} \quad \forall j = 1, \ldots, n - 1,$$

but also

$$\int_{P_{j-1}}^{P_j} (x_j - b^y)^{r-1} \, dy = \int_{\log_b x_j}^{P_j} (b^y - x_j)^{r-1} \, dy \quad \forall j = 1, \ldots, n.$$ 

Eliminating $P$ and substituting $z = b^y/x_j$ in (1.3) yields $n$ equations for $x_1, \ldots, x_n$, namely

$$\int_1^{x_1} (z-1)^{r-1} \frac{dz}{z^r} = 2^{1-r} \int_0^{x_1} \frac{dx_2}{x_1},$$

$$g_r \left( \frac{x_j}{x_{j-1}} \right) = g_0 \left( \frac{x_{j+1}}{x_j} \right), \quad \forall j = 2, \ldots, n - 1,$$

$$g_r \left( \frac{x_n}{x_{n-1}} \right) = \left( 2 - \frac{b}{x_n} \right).$$
where the smooth, increasing function \( g_a \), with \( a \in \mathbb{R} \), is given by

\[
g_a(x) = \int_1^x \frac{(z-1)^{r-1}}{z^a(z+1)} \, dz, \quad x \geq 1.
\]

Assume that \( \bar{x} \in \Xi_n \) also solves (4.4). If \( \bar{x},x_1 > x_1 \) then \( \bar{x}/x_1 > x_{j+1}/x_j \) and hence \( \bar{x}_{j+1} > x_{j+1} \) for all \( j = 0, \ldots, n-1 \), but by the last equation in (4.4) also \( 2b/\bar{x},n > 2b/x,n \), an obvious contradiction. Similarly, \( \bar{x}_1 < x_1 \) leads to a contradiction. Thus, \( \bar{x}_1 = x_1 \), and consequently \( \bar{x} = x \). (If \( n = 1 \) then (4.4) reduces to

\[
\int_1^{x_1} \frac{(z-1)^{r-1}}{z^r} \, dz = 2^{1-r} y_0 \left( \frac{2b-x_1}{x_1} \right),
\]

which also has a unique solution since, as \( x_1 \) increases from 1 to \( b \), the left side increases from 0 whereas the right side decreases to 0.) In summary, therefore, \( x \in \Xi_n \) and \( p \in \Pi_n \) are uniquely determined by \( d_r(\beta_b, \delta^p) = d_r(\beta_b, \delta^{p,n}) \).

As in the case of best uniform \( d_r \)-approximations of \( \beta_b \), no explicit formula is available for \( \delta^{p,n}_r \), not even when \( r = 2 \). Still, it is possible to identify asymptotically best \( d_r \)-approximations, that is, a sequence \( (\delta^{p,n}_r) \) with \( x_n \in \Xi_n \) and \( p_n \in \Pi_n \) for all \( n \in \mathbb{N} \) such that

\[
\lim_{n \to \infty} \frac{d_r(\beta_b, \delta^{p_n}_r)}{d_r(\beta_b, \delta^{p,n}_r)} = 1.
\]

In addition, the \( d_r \)-quantization coefficient of \( \beta_b \) can be computed explicitly; for details see [56], Prop.5.26] and the references given there. Notice that, as \( r \downarrow 1 \), the result is consistent with Corollary 4.4.

**Proposition 4.8.** Let \( b,r > 1 \). Then \( (\delta^{p_n}_r) \), with

\[
x_{n,j} = \left( 1 + \frac{j}{n+1} \left( b^{r/(r+1)} - 1 \right) \right)^{1 + 1/r}, \quad P_{n,j} = \frac{1}{\log b} \log \frac{x_{n,j} + x_{n,j+1}}{2},
\]

for all \( n \in \mathbb{N} \) and \( j = 1, \ldots, n-1 \), and \( x_{n,n} = \left( 1 + \left( b^{r/(r+1)} - 1 \right) \frac{n}{n+1} \right)^{1 + 1/r} \), is a sequence of asymptotically best \( d_r \)-approximations of \( \beta_b \). Moreover,

\[
\lim_{n \to \infty} n d_r(\beta_b, \delta^{p,n}_r) = \frac{r+1}{2(b-1)(\log b)^{1/r}} \left( \frac{b^{r/(r+1)} - 1}{r} \right)^{1 + 1/r}.
\]

**Example 4.9.** For \( \mu = \text{Beta}(2,1) \), given any \( n \in \mathbb{N} \), a unique best uniform \( d_r \)-approximation exists for each \( r \geq 1 \). The best uniform \( d_1 \)-approximations \( \delta^{u}_r \), where \( x_j = \sqrt{\frac{2j-1}{2n}} \) for \( j = 1, \ldots, n \), also constitute a sequence of asymptotically best uniform \( d_r \)-approximations for \( 1 < r < 2 \), with

\[
\lim_{n \to \infty} n d_r(\mu, \delta^{u}_r) = \left( \frac{2^{1-2r}}{(r+1)(2-r)} \right)^{1/r}, \quad (4.5)
\]

in analogy to Proposition 4.6. For \( r \geq 2 \), however, this analogy breaks down, as

\[
\lim_{n \to \infty} \frac{n}{\sqrt{\log n}} d_2(\mu, \delta^{u}_r) = \frac{1}{4\sqrt{3}},
\]

16
and \( \lim_{n \to \infty} n^{1/2+1/r} d_r(\mu, \delta_{\mu}^n) \) is finite and positive whenever \( r > 2 \).

Since \( \mu \) is log-concave, or by an argument similar to the one proving Theorem 4.7, there exists a unique best \( d_r \)-approximation of \( \mu \). While the authors do not know of an explicit formula for \( \delta_{\mu}^n \), simple asymptotically best \( d_r \)-approximations in the spirit of Proposition 4.8 exist, and

\[
\lim_{n \to \infty} n d_r(\mu, \delta_{\mu}^n) = 2^{1/r-1} \frac{r+1}{(r+2)^{1+1/r}} \quad \forall r \geq 1; \quad (4.6)
\]

see [36, Ex.5.28]. Note that (4.6) is smaller than (4.5) for every \( 1 \leq n \).

**Example 4.10.** For the inverse Cantor distribution, for every \( r \geq 1 \) let \( \alpha_r = r^{-1} + (1 - r^{-1}) \log 2 / \log 3 \), and note that \( \log 2 / \log 3 < \alpha_r \leq 1 \). With this, \( 3^{\alpha_r} d_r(\mu, \delta_{\mu}^{3^n}) = d_r(\mu, \delta_{\mu}^n) \) for all \( n \in \mathbb{N} \), and it is readily deduced that

\[
2^{2/r-4}3^{-3/r} \leq n^{\alpha_r} d_r(\mu, \delta_{\mu}^n) \leq 2^{1/r} \quad \forall n \in \mathbb{N}.
\]

Thus \( (n^{\alpha_r} d_r(\mu, \delta_{\mu}^n)) \) is bounded below and above by positive constants. (The authors suspect that this sequence is divergent for every \( r \geq 1 \).)

Best \( d_r \)-approximations also exist, and in a similar spirit it can be shown that \( (n^{\alpha_r} d_r(\mu, \delta_{\mu}^n)) \) is bounded below and above by positive constants (and again, presumably, divergent), where \( \bar{\alpha}_r = \alpha_r \log 3 / \log 2 \). Note that \( 1 < \bar{\alpha}_r \leq \log 3 / \log 2 \), and hence \( (d_r(\mu, \delta_{\mu}^n)) \) decays faster than \( (n^{-1}) \) for every \( r \geq 1 \).

## 5 Kolmogorov approximations

This section discusses best finitely supported \( d_k \)-approximations. Though ultimately the results are true analogues of their counterparts in Sections 3 and 4, the underlying arguments are subtly different, which may be seen as a reflection of the fact that \( d_k \) metrizes a topology finer than the weak topology of \( \mathcal{P} \). (Recall, however, that \( d_k \) does metrize the weak topology on \( \mathcal{P}_\text{fin} \).)

Given \( \mu \in \mathcal{P} \) and \( n \in \mathbb{N} \), for every \( x \in \Xi_n \), let

\[
K^*(x) = \max \left\{ F_{\mu^-}(x_{1}), \frac{1}{2} \max_{j=1}^{n-1} \left( F_{\mu^-}(x_{j+1}) - F_{\mu^-}(x_{j}) \right), 1 - F_{\mu^-}(x_{n}) \right\}. 
\]

Note that \( K^*(x) = d_k \left( \mu, \delta_{\mu}^{\Pi(x)} \right) \) with \( \Pi(x) = \frac{1}{2} (F_{\mu^-}(x_{j}) + F_{\mu^-}(x_{j+1})) \) for all \( j = 1, \ldots, n-1 \). Existence and characterization of best \( d_k \)-approximations with prescribed locations are analogous to Theorem 5.5.

**Theorem 5.1.** Assume that \( \mu \in \mathcal{P} \), and \( n \in \mathbb{N} \). For every \( x \in \Xi_n \), there exists a best \( d_k \)-approximation of \( \mu \), given \( x \). Moreover, \( d_k \left( \mu, \delta_{\mu}^x \right) = d_k \left( \mu, \delta_{\mu}^* \right) \) if and only if, for every \( j = 0, \ldots, n, \)

\[
x_j < x_{j+1} \implies F_{\mu^-}(x_{j+1}) - K^*(x) \leq P_j \leq F_{\mu^-}(x_{j}) + K^*(x),
\]

and in this case \( d_k \left( \mu, \delta_{\mu}^* \right) = K^*(x) \).
Proof. Given $x \in \Xi_n$ and $p \in \Pi_n$, let $y \in \Xi_m$ and $q \in \Pi_m$ as in the proof of Lemma 3.4. Then

$$d_K(\mu, \delta^p) = \max_{i=0}^m \sup_{t \in [y_i, y_{i+1}]} |F_\mu(t) - Q_i|$$

$$
\geq \max \left\{ F_\mu(y_i) - \frac{1}{2} \max_{i=1}^{m-1} (F_\mu(y_i) - F_\mu(y_{i+1})), 1 - F_\mu(y_m) \right\}
$$

$$= \max \left\{ F_\mu(x) - \frac{1}{2} \max_{i=1}^{n-1} (F_\mu(x) - F_\mu(x_{j+1})), 1 - F_\mu(x_n) \right\}
$$

$$= K^*(x).$$

This shows that $\delta^p_x(x)$ is a best $d_K$-approximation, given $x$, and $d_K(\mu, \delta^p) = K^*(x)$. Moreover, $d_K(\mu, \delta^p) = K^*(x)$ if and only if

$$\max \{|F_\mu(y_{i+1}) - Q_i|, |F_\mu(y_i) - Q_i|\} \leq K^*(x) \quad \forall \; i = 1, \ldots, m - 1,$$

that is,

$$F_\mu(y_{i+1}) - K^*(x) \leq Q_i \leq F_\mu(y_i) + K^*(x) \quad \forall \; i = 0, \ldots, m,$$

which in turn is equivalent to the validity (5.1) for every $j$.

To address the approximation problem with prescribed weights, an auxiliary function analogous to $\ell_{f,I}$ in Section 3 is useful. Specifically, given a non-decreasing function $f : \mathbb{R} \to \mathbb{R}$, let $I \subset \mathbb{R}$ be any bounded, non-empty interval, and define $\kappa_{f,I} : \mathbb{R} \to \mathbb{R}$ as

$$\kappa_{f,I}(x) = \max \{|f_-(x) - \inf I|, |f_+(x) - \sup I|\}.$$ 

A few basic properties of $\kappa_{f,I}$ are easily established.

**Proposition 5.2.** Let $f : \mathbb{R} \to \mathbb{R}$ be non-decreasing, and $\emptyset \neq I \subset \mathbb{R}$ a bounded interval. Then, with $s := f^{-1}(\frac{1}{2}(\inf I + \sup I))$, the function $\kappa_{f,I}$ is non-increasing on $]-\infty, s[$, and non-decreasing on $]s, +\infty[$. Moreover, $\kappa_{f,I}$ attains a minimal value whenever $\inf I \leq \frac{1}{2}(f_-(s) + f_+(s)) \leq \sup I$.

It is worth noting that $\kappa_{f,I}$ may in general not attain its infimum, as the example of $f = 15F_\mu$, with $\mu = \frac{1}{10} \lambda [0, 5] + \frac{2}{5} \delta_5$, and $I = [6, 8]$ shows, for which $s = 5$, and $\kappa_{f,I}(5) = 3$, $\kappa_{f,I}(5) = 7$, $\kappa_{f,I}(5) = 9$; correspondingly, $\frac{1}{2}(f_-(5) + f_+(5)) \notin I$.

By using functions of the form $\kappa_{f,I}$, the value of $d_K(\mu, \nu)$ can easily be bounded above whenever $\nu$ has finite support. For convenience, for every $n \in \mathbb{N}$ let $\Xi^+_n = \{x \in \Xi_n : x_1 < \ldots < x_n\}$.

The proof of the following analogue of Lemma 3.4 is straightforward.

**Proposition 5.3.** Let $\mu \in \mathcal{P}$ and $n \in \mathbb{N}$. For every $x \in \Xi_n$ and $p \in \Pi_n$,

$$d_K(\mu, \delta_x^p) \leq \max_{j=1}^n \kappa_{\mu, [p_{j-1}, p_j]}(x_j),$$

and equality holds in (5.2) whenever $x \in \Xi^+_n$.

Consider for instance $\mu = \frac{1}{6} \lambda [0, 2] + \frac{2}{3} \delta_1$, and $x = (1, 1)$. Then, for every $p \in \Pi_2$, clearly $d_K(\mu, \delta^p_x) = \frac{1}{6}$, whereas $\max_{j=1}^2 \kappa_{\mu, [p_{j-1}, p_j]}(x_j) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3} \geq \frac{1}{3}$. Thus the inequality (5.2) may be strict if $x \notin \Xi^+_n$. This, together with the fact that a function $\kappa_{f,I}$ may not attain
Similarly to the proof of Lemma 3.4, label $d$ as $\eta$ and this will now be done. To this end, notice that by the definition of $\xi(p)$, with integers $i \in x$, clearly, to establish (5.4) it is enough to show that $\eta$.

Thus assume $\min_{i=1}^{n} F_{i}^{-1} = 0$ throughout. For convenience, write $F_{i}^{-1}$ as $G$. To prove the existence of a best $d_{K}$-approximation of $\eta$, given $p$, as well as $d_{K}(\mu, \delta_{\xi}^{p}) = K_{\ast}(p)$, clearly it suffices to show that

$$d_{K}(\mu, \delta_{\xi}^{p}) \geq d_{K}(\mu, \delta_{\xi}^{p}) \quad \forall x \in \Xi_n.$$  

Similarly to the proof of Lemma 3.3, label $\xi$ uniquely as

$$\xi_{1} = \ldots = \xi_{j_{1}} < \ldots < \xi_{j_{2}} < \ldots < \xi_{j_{m-1}} < \ldots < \xi_{j_{m}},$$

with integers $i \leq j_{i} \leq m$ for $1 \leq i \leq m$, and $j_{0} = 0$, $j_{m} = n$, and define $\eta \in \Xi_n$ and $q \in \Pi_{m}$ as $\eta_{i} = \xi_{j_{i}}$, and $q_{i} = F_{i} - P_{i-1}$, respectively. With this, $\delta_{\xi}^{p} = \delta_{\xi}^{q}$, and by Proposition 5.3

$$K_{\ast}(p) = d_{K}(\mu, \delta_{\xi}^{p}) = \max_{i=1}^{m} \kappa_{F_{i}, [Q_{i-1}, Q_{i}]} (\eta_{i}).$$

Pick $i$ such that $\kappa_{F_{i}, [Q_{i-1}, Q_{i}]} (\eta_{i}) = K_{\ast}(p)$, that is,

$$\max \{|F_{i} (\eta_{i}) - Q_{i-1}|, |F_{i} (\eta_{i}) - Q_{i}|\} = K_{\ast}(p).$$

Clearly, to establish 5.4 it is enough to show that

$$\max \{|F_{i} (\eta_{i}) - G_{i} (\eta_{i})|, |F_{i} (\eta_{i}) - G_{i} (\eta_{i})|\} \geq K_{\ast}(p)$$

and this will now be done. To this end, notice that by the definition of $\eta$,

$$\frac{1}{2} (P_{j_{i-1}} + P_{j_{i}}) \leq \frac{1}{2} (P_{j_{i-1}} + P_{j_{i}+1}),$$

but also

$$\frac{1}{2} (P_{j_{i-1}} + P_{j_{i}}) \leq \frac{1}{2} (P_{j_{i}} + P_{j_{i}+1}).$$
with the convention that $P_{r-1} = 0$ and $P_{n+1} = 1$.

Assume first that $K_\bullet(p) = |F_{\mu^-} (\eta_i) - Q_{i-1}|$. If $\eta_i \leq x_{j_{i-1}}$ then $G_- (\eta_i) \leq P_{j_{i-1} - 1}$, and hence $F_{\mu^-} (\eta_i) - G_- (\eta_i) \geq F_{\mu^-} (\eta_i) - P_{j_{i-1}}$, but also, by (5.6),

$$F_{\mu^-} (\eta_i) - G_- (\eta_i) \geq F_{\mu^-} (\eta_i) - P_{j_{i-1}} - (2F_{\mu^-} (\eta_i) - P_{j_{i-1} - 1} - P_{j_{i-1}})$$

$$= P_{j_{i-1}} - F_{\mu^-} (\eta_i),$$

and consequently

$$F_{\mu^-} (\eta_i) - G_- (\eta_i) \geq |F_{\mu^-} (\eta_i) - P_{j_{i-1}}| = |F_{\mu^-} (\eta_i) - Q_{i-1}| = K_\bullet(p).$$

If $x_{j_{i-1}} < \eta_i \leq x_{j_{i-1}+1}$ then $G_- (\eta_i) = P_{j_{i-1}}$ and hence

$$|F_{\mu^-} (\eta_i) - G_- (\eta_i)| = K_\bullet(p).$$

Finally, if $\eta_i > x_{j_{i-1}+1}$ then $G_- (\eta_i) \geq P_{j_{i-1}+1}$, and hence $G_- (\eta_i) - F_{\mu^-} (\eta_i) \geq P_{j_{i-1}} - F_{\mu^-} (\eta_i)$, but also, again by (5.6),

$$G_- (\eta_i) - F_{\mu^-} (\eta_i) \geq P_{j_{i-1}+1} - F_{\mu^-} (\eta_i) - (P_{j_{i-1}+1} + P_{j_{i-1}+1} - 2F_{\mu^-} (\eta_i))$$

$$= F_{\mu^-} (\eta_i) - P_{j_{i-1}},$$

and therefore

$$G_- (\eta_i) - F_{\mu^-} (\eta_i) \geq |F_{\mu^-} (\eta_i) - P_{j_{i-1}}| = K_\bullet(p).$$

Thus (5.5) holds whenever $K_\bullet(p) = |F_{\mu^-} (\eta_i) - Q_{i-1}|$.

Next assume that $K_\bullet(p) = |F_{\mu^-} (\eta_i) - Q_{i-1}|$. Utilizing (5.7) instead of (5.6), completely analogous arguments show that $|F_{\mu^-} (\eta_i) - G (\eta_i)| \geq K_\bullet(p)$ in this case as well, which again implies (5.5). The latter therefore holds in either case. As seen earlier, this proves the existence of a best $d_K$-approximation of $\mu$, given $p$, and also that $d_K(\mu, \delta^p_\mu) = K_\bullet(p)$.

Finally, with $y \in \Xi_m^+$ and $p \in \Pi_m$ as in the proof of Lemma 3.3 observe that $d_K(\mu, \delta^y_\mu) = K_\bullet(p)$ if and only if $\max_{i=1}^m \kappa_{F_{\mu^-}[Q_{i-1}, Q_i]}(y,y) = K_\bullet(p)$, by Proposition 5.3. As seen in the proof of Theorem 5.1, this means that

$$F_{\mu^-}(y_{i+1}) - K_\bullet(p) \leq Q_{i} \leq F_{\mu^-}(y_{i}) + K_\bullet(p) \quad \forall \ i = 0, \ldots, m,$$

or equivalently,

$$F_{\mu^-}^{-1}(Q_{i} - K_\bullet(p)) \leq y_{i} \leq F_{\mu^-}^{-1}(Q_{i-1} + K_\bullet(p)) \quad \forall \ i = 1, \ldots, m,$$

which in turn is equivalent to the validity of (5.3) for every $j$.

**Corollary 5.5.** Assume $\mu \in P_{cs}$, and $n \in \mathbb{N}$. Then $d_K(\mu, \delta^y_\mu) \geq \frac{1}{2}n^{-1}$ for all $x \in \Xi_n$, with equality holding if and only if

$$F_{\mu^-}^{-1}\left(\frac{2j - 1}{2n}\right) \leq x \leq F_{\mu^-}^{-1}\left(\frac{2j - 1}{2n}\right) \quad \forall j = 1, \ldots, n.$$
By combining Theorems \textbf{5.1} and \textbf{5.4} it is possible to characterize best \(d_{K}\)-approximations of \(\mu \in \mathcal{P}\) as well. For this, associate with every non-decreasing function \(f : \mathbb{R} \to \mathbb{R}\) and every number \(a \geq 0\) a map \(S_{f,a} : \mathbb{R} \to \mathbb{R}\), given by

\[
S_{f,a}(x) = f_{+}(f^{-1}(x + a)) + a \quad \forall \ x \in \mathbb{R}.
\]

This map is a true analogue of \(T_{f,a}\) in Section 3, and in fact, Proposition\textbf{5.11} with \(T_{f,a}\) replaced by \(S_{f,a}\), remains fully valid. Identical reasoning then shows that

\[
K_{t}^{\infty} := \min \left\{ a \geq 0 : S_{f,a}^{[n]}(0) \geq 1 \right\} < + \infty;
\]

again, \((K_{t}^{\infty})\) is non-increasing, \(nK_{t}^{\infty} \leq \frac{1}{2}\) for every \(n\), and \(K_{t}^{\infty} = 0\) if and only if \#supp \(\mu \leq n\). Notice that if \(\mu \in \mathcal{P}_{cts}\) then

\[
S_{f,a}(x) = \begin{cases} 
  a & \text{if } x < -a, \\
  2a + x & \text{if } -a \leq x < 1 - a, \\
  a + 1 & \text{if } x \geq 1 - a,
\end{cases}
\]

from which it is clear that \(K_{t}^{\infty} = \frac{1}{2}n^{-1}\).

\textbf{Theorem 5.6.} Let \(\mu \in \mathcal{P}\) and \(n \in \mathbb{N}\). There exists a best \(d_{K}\)-approximation of \(\mu\), and \(d_{K}(\mu, \delta_{t}^{\infty}) = K_{t}^{\infty}\). Moreover, for every \(x \in \Xi_{n}\) and \(p \in \Pi_{n}\), the following are equivalent:

(i) \(d_{K}(\mu, \delta_{t}^{p}) = d_{K}(\mu, \delta_{t}^{\infty})\);

(ii) all implications in \textbf{5.1} are valid with \(K_{t}(x)\) replaced by \(K_{t}^{\infty}\);

(iii) all implications in \textbf{5.3} are valid with \(K_{t}(p)\) replaced by \(K_{t}^{\infty}\).

\textbf{Proof.} Note that once the existence of a best \(d_{K}\)-approximation of \(\mu\) is established, the proof is virtually identical to that of Theorem \textbf{5.12}. Thus, only the existence is to be proved here. To this end, let \(a = \inf_{x \in \Xi_{n}, p \in \Pi_{n}} d_{K}(\mu, \delta_{t}^{p})\), and pick sequences \((x_{k})\) and \((p_{k})\) in \(\Xi_{n}\) and \(\Pi_{n}\), respectively, with the property that \(\lim_{k \to \infty} d_{K}(\mu, \delta_{x_{k}}^{p_{k}}) = a\). By the compactness of \(\Xi_{n}\), assume w.o.l.g. that \(\lim_{k \to \infty} x_{k} = \eta \in \Xi_{n}\). Since \(a \leq K_{t}(x_{k}) \leq d_{K}(\mu, \delta_{x_{k}}^{p_{k}})\), it suffices to show that \(K_{t}(\eta) \leq a\). To see the latter, assume that \(\eta_{j} < \eta_{j+1}\) for any \(j = 1, \ldots, n - 1\). Then \(x_{k,j} < x_{k,j+1}\) for all sufficiently large \(k\), and hence by Theorem \textbf{5.1} \(F_{\mu} - (x_{k,j+1}) - F_{\mu}(x_{k,j}) \leq 2K_{t}(x_{k})\), which in turn implies

\[
F_{\mu}(\eta_{j+1}) - F_{\mu}(\eta_{j}) \leq \lim \inf_{k \to \infty} (F_{\mu}(x_{k,j+1}) - F_{\mu}(x_{k,j})) \leq 2a.
\]

Since, similarly, \(F_{\mu}(\eta_{1}) \leq a\) and \(1 - F_{\mu}(\eta_{n}) \leq a\), it follows that \(K_{t}(\eta) \leq a\), as claimed. \(\square\)

\textbf{Corollary 5.7.} Assume \(\mu \in \mathcal{P}_{cts}\), and \(n \in \mathbb{N}\). Then \(K_{t}^{\infty} = K_{t}(u_{n}) = \frac{1}{2}n^{-1}\), and \(\delta_{t}^{p}\) with \(x \in \Xi_{n}\), \(p \in \Pi_{n}\) is a best \(d_{K}\)-approximation of \(\mu\) if and only if it is a best uniform \(d_{K}\)-approximation of \(\mu\).

\textbf{Remark 5.8.} (i) By Theorem \textbf{5.6} \(K_{t}^{\infty} = \min_{x \in \Xi_{n}} K_{t}(x) = \min_{p \in \Pi_{n}} K_{t}(p)\).

(ii) If \(\mu\) has even a single atom, then \(K_{t}^{\infty}\) may be smaller than \(K_{t}(u_{n})\), and thus a best uniform \(d_{K}\)-approximation may not be a best \(d_{K}\)-approximation. A simple example illustrating this is \(\mu = \frac{3}{4}\delta_{0} + \frac{1}{4}\lambda_{[0,1]}\), where \(K_{t}^{\infty} = \frac{1}{4}(2n-1)^{-1}\) whereas \(K_{t}(u_{n}) = \frac{1}{4}\max\{n, 2\}^{-1}\), and hence \(K_{t}^{\infty} < K_{t}(u_{n})\) for every \(n \geq 2\).
For Benford’s Law, the best $d_K$-approximations are the same as the best uniform $d_1$-approximations; see also Figure 1.

**Corollary 5.9.** Assume $b > 1$, and $n \in \mathbb{N}$. Then $\delta_n^{u_n}$ with $x_{n,i} = b^{(2j-1)/(2n)}$ for all $j = 1, \ldots, n$ is the unique best (uniform) $d_K$-approximation of $\beta_b$. Moreover, $d_K(\beta_b, \delta_n^{u_n}) = \frac{1}{2}n^{-1}$.

**Example 5.10.** For $\mu = \text{Beta}(2, 1)$, both $F_\mu$ and $F_\mu^{-1}$ are continuous. By Corollaries 5.3 and 5.4, the best (or best uniform) $d_K$-approximation of $\mu$ is $\delta_n^{u_n}$, with $x_{n,j} = \frac{2j-1}{2n}$ for $j = 1, \ldots, n$, and $d_K(\mu, \delta_n^{u_n}) = d_K(\mu, \delta_1^{u_1}) = \frac{1}{2}n^{-1}$. With Examples 3.9, 3.15, and 4.9, therefore, the sequences $(nd_\ast(\mu, \delta_n^{u_n}))$ all converge to a finite, positive limit, and so do $(nd_\ast(\mu, \beta_n))$, provided that $r < 2$ in case $\ast = r$.

**Example 5.11.** Even though the inverse Cantor distribution is discrete with infinitely many atoms, a best uniform $d_K$-approximation exists, by Theorem 5.6, Utilizing (2.4), a tedious but elementary analysis of $F_\mu$ reveals that $\beta_n^{\ast}$ is valid with $d_K$ instead of $d_L$. With Examples 3.10 and 3.11, therefore, $(nd_\ast(\mu, \delta_n^{u_n}))$ is bounded below and above by positive constants for $\ast = L, 1, K$, but tends to $+\infty$ for $\ast = r > 1$.

Very similarly, a best $d_k$-approximation exists, by Theorem 5.6, and the estimates (3.9) hold with $d_K$ instead of $d_L$. Thus, $(n^3 \log 2 d_\ast(\mu, \delta_n^{u_n}))$ is bounded below and above by positive constants for $\ast = L, 1, K$, but tends to $+\infty$ for $\ast = r > 1$.

### 6 Conclusion

As the title of this article suggests, and the introduction explains, the general results have been motivated by a quantitative analysis of Benford’s Law, and the precise statements regarding the latter are but simple corollaries of the former. In particular, Sections 3 to 5 show that the quantization coefficients $Q_\ast = \lim_{n \to \infty} nd_\ast(\beta_b, \delta_n^{u_n})$ and their uniform counterparts $Q_{\ast,u} = \lim_{n \to \infty} nd_\ast(\beta_b, \delta_n^{u_n})$ all are finite and positive for each metric $d_\ast$ considered. Clearly, $Q_\ast \leq Q_{\ast,u}$ for all $b > 1$. Also, note that $(nd_\ast(\beta_b, \delta_n^{u_n}))$ is non-increasing, possibly constant, whereas $(nd_\ast(\beta_b, \delta_n^{u_n}))$ is non-decreasing. Figure 5 summarizes the results obtained earlier.

The dependence of $Q_\ast$ and $Q_{\ast,u}$ on $b$ is illustrated in Figure 4. On the one hand, $Q_L$ and $Q_{L,u}$ tend to $\frac{1}{2}$ as $b \downarrow 1$, but also as $b \to +\infty$, both attaining their respective minimal value for $b = 2$. On the other hand, $Q_r$ and $Q_{r,u}$ both tend to $\frac{1}{2}(r + 1)^{-1/r}$ as $b \downarrow 1$, whereas $\lim_{b \to +\infty}(\log b)^{1/r}Q_r = \frac{1}{2}(r + 1)r^{(r+1)/r}$ and $\lim_{b \to +\infty}(\log b)^{1/r}Q_{r,u} = \frac{1}{2}r^{-1/r}(r + 1)^{-1/r}$. Finally, $Q_K = Q_{K,u} = \frac{1}{2}$ for all $b$.

**Remark 6.1.** In the context of Benford’s Law, $\mathbb{I} = [1, b]$, and since $S_b < b$ always, it may seem more natural to study the approximation problem not on all of $\mathcal{P}$, but rather on the (dense) subset $\tilde{\mathcal{P}} := \{\mu \in \mathcal{P} : \mu([b]) = 0\}$. Clearly, $d_L$ and $d_\ast$ both metrize the weak topology on $\tilde{\mathcal{P}}$ but are not complete. (By contrast, $d_K$ is complete but not separable, and induces a finer topology.) Since $\tilde{\mathcal{P}}$ is a $G_\delta$-set in $\mathcal{P}$, a classical theorem [12, Thm.2.5.4] yields, for instance,

$$
\tilde{d}(\mu, \nu) = \int_0^1 |G_\mu - G_\nu| + \sum_{k=1}^\infty 2^{-k} \left| \frac{f_{1-k}^{-1} (G_\mu - G_\nu)}{f_{1-k}^{-1} G_\mu f_{1-k}^{-1} G_\nu} \right|.
$$
\begin{align*}
Q_\star & = \frac{\max\{b, 2\} - 1}{2b - 2} \cdot \log(1 + b \log b) - \frac{\log(1 + \log b)}{\log b} \quad \text{for } \star = L, 1, 2, K \\
Q_{\star,u} & = \frac{\max\{b, 2\} - 1}{2b - 2} \cdot \frac{b \log b}{1 + b \log b} \\
& = \frac{r + 1}{2(b - 1)(\log b)^{1/r}} \left( \frac{b^{r/(r+1)} - 1}{r} \right)^{1+1/r} \\
& = \frac{(\log b)^{1-1/r}}{2(b - 1)} \left( \frac{b^r - 1}{r(r+1)} \right)^{1/r} \\
& \quad \text{for } \star = L, 1, 2, K
\end{align*}

\begin{figure}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
\star & $Q_\star$ & $Q_{\star,u}$ \\
\hline
L & $\frac{\max\{b, 2\} - 1}{2b - 2} \cdot \log(1 + b \log b) - \frac{\log(1 + \log b)}{\log b}$ & $\frac{\max\{b, 2\} - 1}{2b - 2} \cdot \frac{b \log b}{1 + b \log b}$ \\
\hline
$r \geq 1$ & $\frac{r + 1}{2(b - 1)(\log b)^{1/r}} \left( \frac{b^{r/(r+1)} - 1}{r} \right)^{1+1/r}$ & $\frac{(\log b)^{1-1/r}}{2(b - 1)} \left( \frac{b^r - 1}{r(r+1)} \right)^{1/r}$ \\
\hline
K & $\frac{1}{2}$ & $\frac{1}{2}$ \\
\hline
\end{tabular}
\caption{The quantization ($Q_\star$) and uniform quantization ($Q_{\star,u}$) coefficients of $\beta_b$ for $d_\star$; see also Figure 4.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.png}
\caption{Comparing the quantization coefficients $Q_\star$ (solid curves) and uniform quantization coefficients $Q_{\star,u}$ (broken curves) of $\beta_b$, for $\star = L$ (red), $\star = 1, 2$ (blue), and $\star = K$ (black), respectively; see also Figure 3.}
\end{figure}

With $G_\mu = b - F_\mu^{-1}$, $G_\nu = b - F_\nu^{-1}$, as an equivalent complete, separable metric on $\tilde{P}$. However, $\tilde{d}$ appears to be quite unwieldy, and the authors do not know of an equivalent complete metric on $\tilde{P}$ for which explicit results similar to those in Sections 3 and 4 could be established.

Also, it is readily confirmed that, given any $\mu \in \tilde{P}$, there exists a best (or best uniform) $d_\star$-approximation $\delta^{\star,n} \in \tilde{P}$ (or $\delta^{\star,u} \in \tilde{P}$), i.e., these approximation problems always have a solution in $(\tilde{P}, d_\star)$, notwithstanding the fact that the latter space is not complete (if $\star = L, r$) or not separable (if $\star = K$).

For Benford’s Law, as seen above, all best (or best uniform) approximations considered converge at the same rate, namely $(n^{-1})$; the same is true for the Beta$(2, 1)$ distribution whenever $1 \leq r < 2$. These are not coincidences. Rather, for many other probability metrics $n^{-1}$ turns

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{Comparing the quantization coefficients $Q_\star$ (solid curves) and uniform quantization coefficients $Q_{\star,u}$ (broken curves) of $\beta_b$, for $\star = L$ (red), $\star = 1, 2$ (blue), and $\star = K$ (black), respectively; see also Figure 3.}
\end{figure}
out to yield the correct order of magnitude of the \(n\)-th quantization error as well. Specifically, consider a metric \(d\) on \(P\) for which

\[
a_1 \|F_{\mu}^{s_1} - F_{\nu}^{s_1}\|_1 \leq d(\mu, \nu) \leq a_2 \left( \epsilon \|F_{\mu}^{s_2} - F_{\nu}^{s_2}\|_\infty + (1 - \epsilon) \|F_{\mu}^{-1} - F_{\nu}^{-1}\|_\infty \right) \quad \forall \mu, \nu \in P, \tag{6.1}
\]

with positive constants \(a_1, a_2, s_1, s_2, \) and \(\epsilon \in \{0, 1\}\); see, e.g., [8, 31, 32] for examples and properties of such metrics. Note that validity of (6.1) causes \(d\) to metrize a topology at least as fine as the weak topology, and clearly (6.1) holds for any \(d = d_\ast\). The latter fact, together with [17, Thm.6.2] yields a simple observation regarding the prevalence of the rate \((n^{-1})\).

**Proposition 6.2.** Let \(d\) be a metric on \(P\) satisfying (6.1). Then, for every \(\mu \in P\),

\[
\limsup_{n \to \infty} n \inf_{x \in \Xi, \rho \in \Pi_n} d(\mu, \delta_x^\rho) < +\infty,
\]

and if \(\mu\) is non-singular (w.r.t. \(\lambda\)) then also

\[
\lim \inf_{n \to \infty} n \inf_{x \in \Xi, \rho \in \Pi_n} d(\mu, \delta_x^\rho) > 0.
\]

**Remark 6.3.** (i) Apart from \(d_\ast\), examples of familiar probability metrics that satisfy (6.1) include the discrepancy distance \(\sup_{I \subset \mathbb{R}} |\mu(I) - \nu(I)|\) and the \(L^p\)-distance \(\|F_{\mu} - F_{\nu}\|_p\) between distribution functions [31]. For the important Prokhorov distance, validity of the right-hand inequality in (6.1) appears to be unknown [16], but best approximations are suspected to converge at the rate \((n^{-1})\) regardless [18, Sec.4]. Also, \((n^{-1})\) is established in [10] as the universal rate of convergence for best approximations under Orlicz norms, which contains \(d_r\) as a special case.

(ii) In [32, Sec.4.2], for any \(a \geq 0\), the \(a\)-Lévy distance

\[
d_{L_a}(\mu, \nu) = \inf \{ y \geq 0 : F_\mu(\cdot - ay) - y \leq F_\nu \leq F_\mu(\cdot + ay) + y \}
\]

is considered. Every \(d_{L_a}\) satisfies (6.1), and \(d_{L_0} = d_K, d_{L_1} = \omega^{-1} d_L\). Usage of \(a\)-Lévy distances may enable a unified treatment of the results in Sections 3 and 5.

(iii) Under additional assumptions on \(\mu\), the value of \(n \inf_{x \in \Xi_n} d(\mu, \delta_x^\nu)\) can similarly be bounded above and below by positive constants [36, Thm.5.15].

Finally, it is worth pointing out that, though motivated here by Benford’s Law, compactness of the interval \(I\) was assumed largely for convenience, and can easily be dispensed with for many of the general results in this article. For instance, if \(I\) is (closed but) unbounded then (2.2), with \(\omega = 1\), still yields \(d_L\) as a complete, separable metric inducing the weak topology on \(P\), though the latter no longer is compact. Clearly, Theorem 3.7 is valid in this situation, as (3.1) holds for \(f = F_\mu\) and any interval \(I \subset \mathbb{R}\). Even though (3.1) may fail for \(f = F_{\mu}^{-1}\) when \(\sup \mu\) is unbounded, it is readily checked that nevertheless the conclusions of Proposition 3.3 remain intact for \(\ell_{F_{\mu}^{-1},I}\), provided that \(I \subset [0, 1]\) but \(I \neq \{0\}\) and \(I \neq \{1\}\). With \(\ell_{F_{\mu}^{-1},(0)}^* := \ell_{F_{\mu}^{-1},(1)}^* := 0\), then, Proposition 3.3 holds verbatim, and so does Theorem 3.12. Analogously, Theorems 5.1, 5.4, and 5.9 all can be seen to be correct, with the definition of \(K_s(\mu)\) understood to assume that \(p, \mu, n > 0\). By contrast, the classical \(L^1\)-Kantorovich distance \(d_1(\mu, \nu) = \|F_{\mu}^{-1} - F_{\nu}^{-1}\|_1\)
is defined only on the (dense) subset \( \mathcal{P}_1 = \{ \mu \in \mathcal{P} : \int |x| \, d\mu(x) < +\infty \} \) where it metrizes a topology finer than the weak topology. Still, with \( \mathcal{P} \) replaced by \( \mathcal{P}_1 \), Proposition 4.1 also remains intact; see, e.g., [36, Sec.5]. Note that the sequence \( (nd_{\ast}(\mu, \delta_{\ast}^{\mu})) \) is bounded when \( \ast = L, K \) because \( d_L \leq d_K \), whereas \( (nd_1(\mu, \delta_{1}^{\mu})) \) may decay arbitrarily slowly; see [36, Thm.5.32]. For a simple application of these results to a probability measure with unbounded support, let \( \mu \) be the standard exponential distribution, i.e., \( F_\mu(x) = \max\{0, 1 - e^{-x}\} \). Calculations quite similar to the ones shown earlier for Benford’s Law yield

\[
\lim_{n \to \infty} nd_L (\mu, \delta_{\ast}^{\mu,n}) = \frac{\log 2}{2}, \quad \lim_{n \to \infty} nd_L (\mu, \delta_{\ast}^{\mu,n}) = \frac{1}{2},
\]

whereas

\[
\lim_{n \to \infty} nd_1 (\mu, \delta_{1}^{\mu,n}) = 1 \quad \text{but} \quad \lim_{n \to \infty} \frac{n}{\log n} d_1(\mu, \delta_{1}^{\mu,n}) = \frac{1}{4},
\]

and clearly \( nd_k(\mu, \delta_{1}^{\mu,n}) = nd_k(\mu, \delta_{1}^{\mu,n}) = \frac{1}{2} \) for all \( n \). Even though \( \mu \) has finite moments of all orders, there exist probability metrics \( d \) for which \( (nd(\mu, \delta_{1}^{\mu,n})) \) is unbounded; see [18, Ex.5.1(d)].

Acknowledgements

The first author was partially supported by an NSERC Discovery Grant. Both authors gratefully acknowledge helpful suggestions made by F. Dai, B. Han, T.P. Hill, and an anonymous referee.

References

[1] P.C. Allaart, An invariant-sum characterization of Benford’s law, J. Appl. Probab., 34 (1997), 288–291.
[2] F. Benford, The law of anomalous numbers, Proc. Amer. Philos. Soc., 78 (1938), 551–572.
[3] A. Berger and T.P. Hill, Benford’s law strikes back: no simple explanation in sight for mathematical gem, Math. Intelligencer, 33 (2011), 85–91.
[4] A. Berger and T.P. Hill, An Introduction to Benford’s Law, Princeton University Press, Princeton, 2015.
[5] A. Berger, T.P. Hill, and K.E. Morrison, Scale-distortion inequalities for mantissas of finite data sets, J. Theoret. Probab., 21 (2008), 97–117.
[6] A. Berger, T.P. Hill, and E. Rogers, Benford Online Bibliography, [http://www.benfordonline.net](http://www.benfordonline.net) 2009. (Last accessed March 16th, 2018.)
[7] A. Berger and I. Twelves, On the significands of uniform random variables, to appear in: J. Appl. Probab. (2018).
[8] I. Bloch and J. Atif, Defining and computing Hausdorff distances between distributions on the real line and on the circle: link between optimal transport and morphological dilations, Math. Morphol. Theory Appl., 1 (2016), 79–99.
[9] S.G. Bobkov and M. Ledoux, One-dimensional empirical measures, order statistics and Kantorovich transport distances, preprint (2016). http://www-users.math.umn.edu/~bobko001/preprints/2016_BL_Order.statistics_Revised.version.pdf To appear in: Memoirs of the AMS.

[10] S. Dereich and C. Vormoor, The high resolution vector quantization problem with Orlicz norm distortion, J. Theoret. Probab., 24 (2011), 517–544.

[11] P. Diaconis, The distribution of leading digits and uniform distribution mod 1, Ann. Probab., 5 (1977), 72–81.

[12] R. Dudley, Real Analysis and Probability. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, 2004.

[13] L. Dümbgen and C. Leuenberger, Explicit bounds for the approximation error in Benford’s law, Elect. Comm. in Probab., 13 (2008), 99–112.

[14] W. Feller, An Introduction to Probability Theory and Its Applications. Vol. II, John Wiley and Sons, New York, 1966.

[15] N. Gauvrit and J.-P. Delahaye, Scatter and Regularity Imply Benford’s Law... and More, pp. 53–69 in: H. Zenil (ed.), Randomness Through Complexity, World Scientific, Singapore, 2011.

[16] A.L. Gibbs and F.E. Su, On choosing and bounding probability metrics, Int. Stat. Rev., 70 (2002), 419–435.

[17] S. Graf and H. Luschgy, Foundations of Quantization for Probability Distributions, Lecture Notes in Mathematics 1730, Springer, 2000.

[18] S. Graf and H. Luschgy, Quantization for probability measures in the Prokhorov metric, Theory Probab. Appl., 53 (2009), 216–241.

[19] T.P. Hill. A statistical derivation of the significant-digit law. Statistical Science, 10 (1995), 354–363.

[20] T.P. Hill, Base-invariance implies Benford’s law, Proc. Amer. Math. Soc., 123 (1995), 887–895.

[21] P.J. Huber, Robust Statistics, John Wiley and Sons, New York, 1981.

[22] L.V. Kantorovich and G. Rubinstein, On a space of completely additive functions, Vestnik Leningradskogo Universiteta, 13 (1958), 52–59.

[23] D.E. Knuth, The Art of Computer Programming, Addison-Wesley, Reading, 1975.

[24] T. Linder, On asymptotically companding quantization, Probl. Control. Inform., 20 (1991), 475–484.
[25] S.J. Miller, *Benford’s Law: Theory and Applications*, Princeton University Press, Princeton, 2015.

[26] Y. Mori and K. Takashima, *On the distribution of the leading digit of $a^n$: a study via $\chi^2$ statistics*, Period. Math. Hung., 73 (2016), 224–239.

[27] S. Newcomb, *Note on the frequency of use of the different digits in natural numbers*, Amer. J. Math., 4 (1881), 39–40.

[28] G.C. Pflug and A. Pichler, *Approximations for probability distributions and stochastic optimization problems*, Internat. Ser. Oper. Res. Manage. Sci. 1633, Springer, New York, 2011, 343–387.

[29] R.S. Pinkham, *On the distribution of first significant digits*, Ann. Math. Statist., 32 (1961), 1223–1230.

[30] K. Pötzelberger and K. Felsenstein, *An asymptotic result on principal points for univariate distributions*, Optimization, 28 (1994), 397–406.

[31] S.T. Rachev, *Probability Metrics and the Stability of Stochastic Models*, John Wiley and Sons, New York, 1991.

[32] S.T. Rachev, L.B. Klebanov, S.V. Stoyanov, and F.J. Fabozzi, *A structural classification of probability distances*, In: *The Methods of Distances in the Theory of Probability and Statistics*, Springer, New York, 2013.

[33] S.T. Rachev and L. Rüschendorf, *Mass Transportation Problems. Vol. II: Applications*, Springer-Verlag, 1998.

[34] R. A. Raimi, *The first digit problem*, Amer. Math. Monthly, 83 (1976), 521–538.

[35] P. Schatte, *On mantissa distributions in computing and Benford’s law*, J. Inform. Process. Cybernet., 24 (1988), 443–455.

[36] C. Xu and A. Berger, *Best finite constrained approximations of one-dimensional probabilities*, preprint (2017), [arXiv:1704.07871](https://arxiv.org/abs/1704.07871)