Asymptotics of One-Dimensional Lévy Approximations

Arno Berger[1] · Chuang Xu[2]

Abstract

For arbitrary Borel probability measures on the real line, necessary and sufficient conditions are presented that characterize best purely atomic approximations relative to the classical Lévy probability metric, given any number of atoms, and allowing for additional constraints regarding locations or weights of atoms. The precise asymptotics (as the number of atoms goes to infinity) of the approximation error is identified for the important special cases of best uniform (i.e. all atoms having equal weight) and best (i.e. unconstrained) approximations, respectively. When compared to similar results known for other probability metrics, the results for Lévy approximations are more complete and require fewer assumptions.

Keywords Best (uniform) approximation · Lévy probability metric · Inverse function · Inverse measure · Approximation error · Asymptotic point distribution

Mathematics Subject Classification (2010) 60B10 · 60E15 · 62E15

1 Introduction

Let $\mathcal{P}$ be the set of all Borel probability measures on the real line, and denote the support of $\mu \in \mathcal{P}$ by supp μ. For each positive integer n, let $\mathcal{P}_n^* = \{\mu \in \mathcal{P} : \#\text{supp }\mu \leq n\}$. Recall that $\mathcal{P}$ endowed with the topology of weak convergence is a Polish space that contains $\mathcal{P}_\infty^* := \bigcup_n \mathcal{P}_n^* = \{\mu \in \mathcal{P} : \#\text{supp }\mu < \infty\}$ as a dense subspace [11, Ch. 11]. Many different metrics (and metric-like quantities [15]) on $\mathcal{P}$ or parts thereof have been studied extensively, as they play important roles in probability theory and statistics [30,31]. Given a specific probability metric d and $\mu \in \mathcal{P}\backslash\mathcal{P}_\infty^*$, it is natural to ask whether there exists, for every n, a best d-approximation $\delta_\bullet^{\bullet,n}$ of μ in $\mathcal{P}_n^*$, i.e. $d(\mu, \delta_\bullet^{\bullet,n}) = \inf\{d(\mu, \nu) : \nu \in \mathcal{P}_n^*\}$, perhaps with additional desirable properties

✉ Arno Berger
berger@ualberta.ca

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

2 Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark

 Springer

such as, e.g. all atoms having equal weight; see Sect. 2 for precise terminology and notation. Provided they exist, how can such best d-approximations be characterized and found systematically? How fast do they converge to μ, i.e. at what rate does the approximation error $d(\mu, \delta_\bullet^{\bullet, n})$ tend to 0 as $n \to \infty$? Questions like these, regarding the approximation in $\mathcal{P}$ by elements of $\mathcal{P}_\infty^*$, continue to attract interest in a wide variety of contexts; see, e.g. [5,6,9,10,23,24,28] and the many references therein.

Denoting the distribution function of $\mu \in \mathcal{P}$ by F_μ, that is, $F_\mu(x) = \mu(]-\infty, x])$ for all $x \in \mathbb{R}$, recall the *Kantorovich* (or *Wasserstein*; cf. [6, p. 4] and [15]) *metric*, given by

$$d_{\mathsf{W}}(\mu, \nu) = \int_{\mathbb{R}} |F_\mu(x) - F_\nu(x)| \, \mathrm{d}x = \int_{[0,1]} |F_\mu^{-1}(y) - F_\nu^{-1}(y)| \, \mathrm{d}y, \qquad (1.1)$$

where F_μ^{-1} is an inverse of F_μ; see Sect. 2 for details. Note that strictly speaking d_{W} is not defined on all of $\mathcal{P} \times \mathcal{P}$, but only on $\mathcal{P}_1 \times \mathcal{P}_1$, with $\mathcal{P}_1 = \{\mu \in \mathcal{P} : \int_{\mathbb{R}} |x| \, \mathrm{d}\mu(x) < +\infty\} \supset \mathcal{P}_\infty^*$. The metric space $(\mathcal{P}_1, d_{\mathsf{W}})$ is complete and separable, though its metric topology is finer than the subspace topology inherited from $\mathcal{P}$. Due to its simplicity and functional-analytic flavour, the metric d_{W} figures prominently in many applied areas, e.g. image compression, signal processing, mathematical finance and optimal transport [23,27,32,35,38]. A vast literature exists addressing the basic questions mentioned earlier relative to d_{W}, as well as many generalizations thereof, notably to multi-dimensional settings [6,13,21,22,28].

Another important notion of distance, the *Prokhorov metric* is given by

$$d_{\mathsf{P}}(\mu, \nu) = \inf\left\{y \in \mathbb{R}^+ : \mu(B) \leq \nu(B^y) + y \ \forall \text{ Borel sets } B \subset \mathbb{R}\right\} \quad \forall \mu, \nu \in \mathcal{P}, \tag{1.2}$$

where $B^y = \{x \in \mathbb{R} : \mathrm{dist}(x, B) < y\}$. Note that d_{P} is defined on all of $\mathcal{P} \times \mathcal{P}$, unlike d_{W}, and metrizes precisely the topology of weak convergence [11,15]. Also, $d_{\mathsf{P}}(\mu, \nu) \leq 1$ for all $\mu, \nu \in \mathcal{P}$. A general theory of best d_{P}-approximation in $\mathcal{P}$ by elements of $\mathcal{P}_\infty^*$ has been initiated in [20], where the authors observe that some aspects of the theory are "more difficult [than the corresponding theory for d_{W}] …mainly due to the lack of suitable scaling properties [of d_{P}]".

In a spirit similar to [18,37], this article addresses the approximation problem relative to the classical *Lévy metric*,

$$d_1(\mu, \nu) = \inf\left\{y \in \mathbb{R}^+ : F_{\mu-}(\cdot - y) - y \leq F_\nu \leq F_\mu(\cdot + y) + y\right\} \quad \forall \mu, \nu \in \mathcal{P}, \tag{1.3}$$

where $F_{\mu-}(x) = \lim_{\varepsilon \downarrow 0} F_\mu(x - \varepsilon) = \mu(]-\infty, x[)$. Note that $d_1 \leq 1$, similarly to d_{P}. The values of $d_{\mathsf{W}}, d_{\mathsf{P}}$ and d_1 are not completely unrelated, since $d_1(\mu, \nu) \leq d_{\mathsf{P}}(\mu, \nu) \leq \sqrt{d_{\mathsf{W}}(\mu, \nu)}$ for all $\mu, \nu \in \mathcal{P}_1$; see, e.g. [4,11,15]. When compared to d_{W} and d_{P}, the metric d_1 is particularly attractive: On the one hand, it is a *bona fide* metric [3, p. 100] metrizing the topology of weak convergence on all of $\mathcal{P}$ (similar to d_{P}, but unlike d_{W}). On the other hand, its definition (1.3) is considerably easier to work with than (1.2). Although computing d_1 for concrete problems may still be "not easy" [15, p. 423] (cf. also [34]), especially when compared to (1.1), the main (asymptotic) results of this article suggest that nevertheless d_1 often is more benign than both d_{W} and d_{P}, in that

fewer assumptions (or no assumptions at all, as in Theorem 4.1 and Proposition 4.3) are needed to draw analogous or perhaps even stronger conclusions. With all technical details deferred to later sections, this is illustrated here for two familiar (absolutely continuous) distributions—standard normal and 1-Pareto.

Let μ be the standard normal distribution. By a celebrated asymptotic result for best d_W-approximations [18, Thm. 6.2],

$$\lim_{n\to\infty} n d_W(\mu, \delta_\bullet^{\bullet,n}) = \sqrt{\frac{\pi}{2}} = 1.253, \tag{1.4}$$

whereas by [20, Ex. 5.1],

$$\lim_{n\to\infty} \frac{n}{\sqrt{\log n}} d_P(\mu, \delta_\bullet^{\bullet,n}) = \sqrt{2}. \tag{1.5}$$

Note that (1.4) yields the faster decay of the approximation (or *quantization*) error $d(\mu, \delta_\bullet^{\bullet,n})$, whereas only (1.5) involves a probability metric that actually metrizes the topology of weak convergence. As it turns out, for the Lévy metric these two desirable properties can be achieved simultaneously: Theorem 4.1, one of the main results of this article, yields

$$\lim_{n\to\infty} n d_1(\mu, \delta_\bullet^{\bullet,n}) = -\sqrt{\frac{\pi}{2}} \, \mathrm{Li}_{1/2}\left(-\frac{1}{\sqrt{2\pi}}\right) = 0.3931, \tag{1.6}$$

where $\mathrm{Li}_{1/2}$ denotes the polylogarithm of order $\frac{1}{2}$.

An interesting variant of (1.4)–(1.6) considers best *uniform* approximations of $\mu \in \mathcal{P}$, that is, best approximations of μ by $\nu \in \mathcal{P}_n^*$, subject to the additional requirement that $n\nu(\{x\})$ is a (positive) integer for every $x \in \mathrm{supp}\,\nu$. Best uniform (or, more generally, best constrained) approximations have recently attracted considerable interest, not least in view of potential applications in stochastic processes and differential equations [7,8,16,17,36,37]; they may also be viewed as deterministic analogues of (random) empirical measures [6,9,14]. With $\delta_\bullet^{u_n}$ denoting a best uniform d-approximation of μ, trivially $d(\mu, \delta_\bullet^{\bullet,n}) \leq d(\mu, \delta_\bullet^{u_n})$. For μ being the standard normal distribution, [37, Ex. 5.18] reports that

$$d_W(\mu, \delta_\bullet^{u_n}) = \mathcal{O}\left(\frac{\sqrt{\log n}}{n}\right) \quad \text{as } n \to \infty,$$

and this bound is sharp; cf. also [7,16]. Although the authors do not know of any analogous result regarding best uniform d_P-approximations, (1.5) makes it clear that $d_P(\mu, \delta_\bullet^{u_n})$ is at least $\mathcal{O}(n^{-1}\sqrt{\log n})$ as $n \to \infty$, if not larger. By contrast, Theorem 3.3, another main result of this article, simply yields

$$\lim_{n\to\infty} n d_1(\mu, \delta_\bullet^{u_n}) = \frac{1}{2},$$

which represents a faster and more precise rate than its d_W- and d_P-counterparts.

For a second illustrative example, let μ be the 1-Pareto distribution, i.e. $F_\mu(x) = 1 - x^{-1}$ for all $x \geq 1$. Since $\mu \notin \mathcal{P}_1$, clearly μ is not amenable to d_{W}-approximation, whereas [20, Thm. 5.2] yields

$$\lim_{n\to\infty} \sqrt{n} d_{\mathsf{P}}(\mu, \delta_\bullet^{\bullet,n}) = \frac{1}{\sqrt{2}}.$$

For the Lévy metric, this article again provides faster, more precise rates, namely

$$nd_1(\mu, \delta_\bullet^{u_n}) = \frac{1}{2} - \frac{1}{8} n^{-2} + \mathcal{O}(n^{-3}) \quad \text{as } n \to \infty, \tag{1.7}$$

as well as

$$nd_1(\mu, \delta_\bullet^{\bullet,n}) = \frac{\pi}{8} + \frac{\pi^2(6-\pi)}{3 \times 2^{10}} n^{-2} + \mathcal{O}(n^{-3}) \quad \text{as } n \to \infty. \tag{1.8}$$

Thus, the results of this article make the case that although the Lévy metric d_1, unlike d_{W} and d_{P}, does not extend to higher dimensions in a straightforward way, its usage for one-dimensional probabilities often leads to simpler and stronger results.

This article is organized as follows. Section 2 first introduces all required terminology and notation and then reviews basic facts pertaining to approximations in $\mathcal{P}$ relative to the Lévy metric. Utilizing the latter, Sects. 3 and 4 specifically study best uniform and best (unconstrained) approximations, respectively, and in particular the asymptotics of the approximation error as $n \to \infty$. Also, under a mild assumption the atoms of (asymptotically) best approximations conform to an asymptotic point distribution, as shown by Theorem 4.5.

2 Lévy Probability Metrics

This section reviews basic facts regarding the approximation in $\mathcal{P}$ by measures with finite support, relative to the Lévy probability metric(s). The stated results are straightforward extensions of [4,36], and the reader is referred to these references for further details and elementary proofs. The following, mostly standard notations are used throughout. The sets of all positive integers, nonnegative integers, integers, positive real numbers and real numbers are denoted $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{Z}$, $\mathbb{R}^+$ and $\mathbb{R}$, respectively. Numerical values of real numbers are displayed to four correct significant decimal digits. For every $x \in \mathbb{R}$ and non-empty $A \subset \mathbb{R}$, $\text{dist}(x, A) = \inf_{a \in A} |x - a|$, $\text{diam } A = \sup_{a,b \in A} |a - b|$, and $\mathbf{1}_A$ is the indicator function of A; also, $\lfloor x \rfloor$ denotes the largest integer not larger than x, and the cardinality of A is $\#A$. If the domain of a function f contains A, then $f(A) = \{f(a) : a \in A\}$. Lebesgue measure on the real line is denoted λ.

Since non-decreasing functions play a crucial role in what follows, first a few basic properties of such functions are recorded. Throughout, denote by $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ the extended real line with its usual order and topology, and by $\mathcal{F}$ the family of all functions $f : \mathbb{R} \to \overline{\mathbb{R}}$ that are non-decreasing and right-continuous. Given $f \in \mathcal{F}$, let $f(\pm\infty) = \lim_{x \to \pm\infty} f(x) \in \overline{\mathbb{R}}$, and for every $x \in \overline{\mathbb{R}}$ let $f_-(x) = \lim_{\varepsilon \downarrow 0} f(x - \varepsilon)$.

Note that $f_-(x) \leq f(x) \leq f_-(y)$ whenever $x < y$; in particular, $f_-(x) = f(x)$ if and only if f is continuous at x. With every $f \in \mathcal{F}$ associate its (upper) *inverse function* $f^{-1} : \mathbb{R} \to \overline{\mathbb{R}}$ given by

$$f^{-1}(x) = \sup\{y \in \mathbb{R} : f(y) \leq x\} \quad \forall x \in \mathbb{R};$$

here and throughout the convention $\sup \varnothing = -\infty$ (and $\inf \varnothing = +\infty$) is adhered to. Importantly, $\mathcal{F}$ is closed under inversion and composition.

Proposition 2.1 *Let $f, g \in \mathcal{F}$. Then, $f^{-1} \circ g \in \mathcal{F}$, and $(f^{-1})^{-1} = f$.*

Given $f, g \in \mathcal{F}$ and $\epsilon > 0$, let

$$d_\epsilon(f, g) = \inf\{y \in \mathbb{R}^+ : f_-(\cdot - y/\epsilon) - y \leq g \leq f(\cdot + y/\epsilon) + y\} \in [0, +\infty].$$

Motivated for $\epsilon = 1$ by (1.3), this definition enables a unified treatment of all ϵ-Lévy probability metrics later in this section. It is readily checked that d_ϵ indeed satisfies the axioms of a metric on $\mathcal{F}$, except that $d_\epsilon(f, g)$ may equal $+\infty$. Also, d_ϵ is compatible with inversion.

Proposition 2.2 *Let $f, g \in \mathcal{F}$ and $\epsilon > 0$. Then, $d_\epsilon(f^{-1}, g^{-1}) = \epsilon d_{1/\epsilon}(f, g)$.*

Given $f, g \in \mathcal{F}$, note that $\epsilon \mapsto d_\epsilon(f, g)$ is non-decreasing and continuous on $\mathbb{R}^+$. Consequently, the limits of $d_\epsilon(f, g)$ exist as $\epsilon \to 0$ or $\epsilon \to +\infty$. For instance, if $f, g \in \mathcal{F}$ are bounded, then simply

$$\lim_{\epsilon \to 0} d_\epsilon(f, g) = \lim \sup_{|x| \to +\infty} |f(x) - g(x)|$$
$$= \max\{|f(-\infty) - g(-\infty)|, |f(+\infty) - g(+\infty)|\},$$

but also

$$\lim_{\epsilon \to +\infty} d_\epsilon(f, g) = \sup_{x \in \mathbb{R}} |f(x) - g(x)| = \|f - g\|_\infty;$$

here, as usual, $\|h\|_\infty = \mathrm{ess} \sup |h| = \inf\{y \in \mathbb{R}^+ : \lambda(\{|h| \geq y\}) = 0\}$ for every measurable function $h : \mathbb{R} \to \overline{\mathbb{R}}$.

Given $f \in \mathcal{F}$, let $I \subset \overline{\mathbb{R}}$ be any interval with the property that

$$f_-(\sup I - x), -f(\inf I + x) < +\infty \quad \text{for some } x \in \mathbb{R}, \tag{2.1}$$

and consider the auxiliary function $\ell_{f,I} : \mathbb{R} \to \mathbb{R}$, introduced in [4], with

$$\ell_{f,I}(x) = \inf\{y \in \mathbb{R}^+ : f_-(\sup I - y) - y \leq x \leq f(\inf I + y) + y\} \quad \forall x \in \mathbb{R};$$

also, let $\ell_{f,I}^* = \inf\{y \in \mathbb{R}^+ : f_-(\sup I - y) - y \leq f(\inf I + y) + y\}$. For any sequence $(I_k)_{k \in \mathbb{N}}$ of intervals in $\overline{\mathbb{R}}$, write $\lim_{k \to \infty} I_k = I$ if $\lim_{k \to \infty} \inf I_k = \inf I$ and $\lim_{k \to \infty} \sup I_k = \sup I$.

Proposition 2.3 *Let $f \in \mathcal{F}$, and let $I \subset \overline{\mathbb{R}}$ be an interval satisfying* (2.1).

(i) *The function $\ell_{f,I}$ is Lipschitz continuous and nonnegative;*
(ii) $\ell_{f,I}(x) \geq \ell^*_{f,I} \geq 0$ *for all $x \in \mathbb{R}$;*
(iii) $\ell^*_{f,I} \leq \frac{1}{2}\lambda(I)$, *and* $\ell^*_{f,I} = 0$ *if and only if $f_-(\sup I) \leq f(\inf I)$;*
(iv) *If $(I_k)_{k \in \mathbb{N}}$ is a sequence of intervals in $\overline{\mathbb{R}}$ with $\lim_{k \to \infty} I_k = I$, then I_k satisfies* (2.1) *for all sufficiently large k, and*

$$\lim_{k \to \infty} \ell^*_{f,I_k} = \ell^*_{f,I} \quad \text{as well as} \quad \lim_{k \to \infty} \ell_{f,I_k}(x) = \ell_{f,I}(x) \quad \forall x \in \mathbb{R}.$$

Remark 1 (i) If $f = F_\mu$ (respectively, $f = F_\mu^{-1}$) for some $\mu \in \mathcal{P}$, then every (respectively, every bounded) interval $I \subset \overline{\mathbb{R}}$ satisfies (2.1). Given $f \in \mathcal{F}$, note that $f = F_\mu$ for some (necessarily unique) $\mu \in \mathcal{P}$ if and only if $f(-\infty) = 0$ and $f(+\infty) = 1$; similarly, $f = F_\mu^{-1}$ for some $\mu \in \mathcal{P}$ if and only if $f_-(0) = -\infty$, $f(1) = +\infty$, and $f(]0, 1[) \subset \mathbb{R}$.

(ii) The function $\ell_{f,I}$ may not attain a minimum value, or when it does, that minimum value may be larger than $\ell^*_{f,I}$. However, mild additional assumptions guarantee that $\ell_{f,I}(x) = \ell^*_{f,I}$ for some $x \in \mathbb{R}$; see [4, Prop. 3.3].

For every $\epsilon > 0$, consider the *ϵ-Lévy metric* on $\mathcal{P}$ given by

$$d_\epsilon(\mu, \nu) = d_\epsilon(F_\mu, F_\nu) \quad \forall \mu, \nu \in \mathcal{P}.$$

The metric d_ϵ is complete, separable and induces the topology of weak convergence (for an authoritative account on the family $(d_\epsilon)_{\epsilon > 0}$, the reader may want to consult [31, Sec. 4.2]; see also [34]). Note that $\epsilon \mapsto d_\epsilon(\mu, \nu)$ is non-decreasing with $\lim_{\epsilon \to 0} d_\epsilon(\mu, \nu) = 0$, whereas

$$\lim_{\epsilon \to +\infty} d_\epsilon(\mu, \nu) = \|F_\mu - F_\nu\|_\infty \quad \forall \mu, \nu \in \mathcal{P},$$

often referred to as the *uniform* or *Kolmogorov* metric, yields a complete yet non-separable metric on $\mathcal{P}$ and induces a finer topology [4, Sec. 5]. For any $\mu \in \mathcal{P}$ and $\epsilon > 0$, and with the dilation $T_\epsilon : x \mapsto \epsilon x$, notice the simple but useful identity

$$d_\epsilon(\mu, \nu) = d_1(\mu \circ T_\epsilon^{-1}, \nu \circ T_\epsilon^{-1}) \quad \forall \mu, \nu \in \mathcal{P}. \tag{2.2}$$

To study finitely supported (and hence purely atomic) d_ϵ-approximations of any $\mu \in \mathcal{P}$, this article employs the following notations: 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 $p \in \Pi_n$ let $\delta_x^p = \sum_{j=1}^n p_{,j} \delta_{x_{,j}}$. For convenience, $x_{,0} := -\infty$ and $x_{,n+1} := +\infty$ for every $x \in \Xi_n$, as well as $P_{,i} := \sum_{j=1}^i p_{,j}$ for $i = 0, \ldots, n$ and every $p \in \Pi_n$; note that $P_{,0} = 0$ and $P_{,n} = 1$. Henceforth, usage of the symbol δ_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. Utilizing (2.2) and [4, Lem. 3.4], the value of $d_\epsilon(\mu, \delta_x^p)$ allows for simple explicit expressions.

Proposition 2.4 *Let $\mu \in \mathcal{P}$, $\epsilon > 0$, and $n \in \mathbb{N}$. For every $x \in \Xi_n$ and $p \in \Pi_n$,*

$$d_\epsilon(\mu, \delta_x^p) = \epsilon \max_{j=0}^n \ell_{F_\mu/\epsilon, [x,j,x,j+1]}(P,j/\epsilon) = \max_{j=1}^n \ell_{\epsilon F_\mu^{-1}, [P,j-1,P,j]}(\epsilon x,j). \tag{2.3}$$

For every $\mu \in \mathcal{P}, \epsilon > 0$ and $n \in \mathbb{N}$, (2.3) suggests considering the following quantities: given $x \in \Xi_n$, let

$$\ell_x^\bullet = \epsilon \max \left\{ \ell_{F_\mu/\epsilon, [-\infty, x,1]}(0), \ \ell_{F_\mu/\epsilon, [x,1,x,2]}^*, \ldots, \ell_{F_\mu/\epsilon, [x,j-1,x,j]}^*, \ldots \right.$$
$$\left. \ldots, \ell_{F_\mu/\epsilon, [x,n-1,x,n]}^*, \ \ell_{F_\mu/\epsilon, [x,n,+\infty]}(1/\epsilon) \right\},$$

and given $p \in \Pi_n$, let

$$\ell_\bullet^p = \max_{j=1}^n \ell_{\epsilon F_\mu^{-1}, [P,j-1,P,j]}^*.$$

Notice that while $\ell_x^\bullet$ and $\ell_\bullet^p$ do depend on μ, ϵ, and, implicitly, also n, in order to keep notations simple, this dependence is not displayed explicitly. By Proposition 2.3(iv), $p \mapsto \ell_\bullet^p$ is continuous on Π_n, and hence

$$\ell_\bullet^{\bullet, n} = \min_{p \in \Pi_n} \ell_\bullet^p$$

is well defined (for a constructive alternative definition of $\ell_\bullet^{\bullet, n}$, see [4, Sec. 3]). The quantities $\ell_x^\bullet$, $\ell_\bullet^p$ and $\ell_\bullet^{\bullet, n}$ control the minimization of $(x, p) \mapsto d_\epsilon(\mu, \delta_x^p)$, with or without constraints, in a sense made precise by Proposition 2.5. To formulate the result, call δ_x^p a *best d_ϵ-approximation of $\mu \in \mathcal{P}$, given $x \in \Xi_n$* if

$$d_\epsilon(\mu, \delta_x^p) \leq d_\epsilon(\mu, \delta_x^q) \quad \forall q \in \Pi_n.$$

Similarly, call δ_x^p a *best d_ϵ-approximation of μ, given $p \in \Pi_n$* if

$$d_\epsilon(\mu, \delta_x^p) \leq d_\epsilon(\mu, \delta_y^p) \quad \forall y \in \Xi_n.$$

Denote by $\delta_x^\bullet$ and $\delta_\bullet^p$ any best d_ϵ-approximation of μ, given x and p, respectively. Best d_ϵ-approximations, given $p = u_n := (n^{-1}, \ldots, n^{-1})$, are referred to as best *uniform d_ϵ-approximations*, and denoted $\delta_\bullet^{u_n}$. Finally, δ_x^p is a *best d_ϵ-approximation of $\mu \in \mathcal{P}$*, denoted $\delta_\bullet^{\bullet, n}$, if

$$d_\epsilon(\mu, \delta_x^p) \leq d_\epsilon(\mu, \delta_y^q) \quad \forall y \in \Xi_n, q \in \Pi_n.$$

Notice that usage of the symbols $\delta_x^\bullet$, $\delta_\bullet^p$ and $\delta_\bullet^{\bullet, n}$ always refers to specific $\mu \in \mathcal{P}$, $\epsilon > 0$ and $n \in \mathbb{N}$, all of which are usually clear from the context.

Proposition 2.5 *Let $\mu \in \mathcal{P}, \epsilon > 0$ and $n \in \mathbb{N}$.*

(i) *For every $x \in \Xi_n$, there exists a best d_ϵ-approximation of μ, given x, and $d_\epsilon(\mu, \delta_x^\bullet) = \ell_x^\bullet$. Moreover, $d_\epsilon(\mu, \delta_x^p) = \ell_x^\bullet$ with $p \in \Pi_n$ if and only if*

$$\epsilon \ell_{F_\mu/\epsilon, [x,j, x,j+1]}(P,j/\epsilon) \le \ell_x^\bullet \quad \forall j = 0, \ldots, n. \tag{2.4}$$

(ii) *For every $p \in \Pi_n$, there exists a best d_ϵ-approximation of μ, given p and $d_\epsilon(\mu, \delta_\bullet^p) = \ell_\bullet^p$. Moreover, $d_\epsilon(\mu, \delta_x^p) = \ell_\bullet^p$ with $x \in \Xi_n$ if and only if*

$$\ell_{\epsilon F_\mu^{-1}, [P,j-1, P,j]}(\epsilon x,j) \le \ell_\bullet^p \quad \forall j = 1, \ldots, n. \tag{2.5}$$

(iii) *There exists a best d_ϵ-approximation of μ, and $d_\epsilon(\mu, \delta_\bullet^{\bullet,n}) = \ell_\bullet^{\bullet,n}$. Moreover, $d_\epsilon(\mu, \delta_x^p) = \ell_\bullet^{\bullet,n}$ with $x \in \Xi_n$, $p \in \Pi_n$ if and only if (2.4) and (2.5) hold with $\ell_\bullet^{\bullet,n}$ instead of $\ell_x^\bullet$ and $\ell_\bullet^p$, respectively.*

The following two examples illustrate Proposition 2.5. Notice that in either example the sequences $(d_\epsilon(\mu, \delta_\bullet^{u_n}))$ and $(d_\epsilon(\mu, \delta_\bullet^{\bullet,n}))$ both converge to 0 at the same rate (n^{-1}). As demonstrated in Sects. 3 and 4 for best uniform and best d_ϵ-approximations, respectively, this rate is not specific to these examples, but rather indicative of much more general mechanisms.

Example 1 Consider the exponential distribution $\exp(a)$ with parameter $a > 0$, i.e. let $F_\mu(x) = 1 - e^{-ax}$ for all $x \ge 0$. From Proposition 2.5, it is easily deduced that $\delta_x^{u_n}$ with $x \in \Xi_n$ is a best uniform d_ϵ-approximation of μ if and only if, for every $j = 1, \ldots, n$,

$$x,j \in \left[-\frac{1}{a} \log\left(1 - \frac{j}{n} + \ell_\bullet^{u_n}\right) - \frac{\ell_\bullet^{u_n}}{\epsilon}, -\frac{1}{a} \log\left(1 - \frac{j-1}{n} - \ell_\bullet^{u_n}\right) + \frac{\ell_\bullet^{u_n}}{\epsilon} \right],$$

with $\ell_\bullet^{u_n} = d_\epsilon(\mu, \delta_\bullet^{u_n})$ being the unique solution of $n\ell(e^{2a\ell/\epsilon} + 1) = 1$. A straightforward analysis of the latter equation yields the asymptotic equality

$$n d_\epsilon(\mu, \delta_\bullet^{u_n}) = \frac{1}{2} - \frac{a}{4\epsilon} n^{-1} + \mathcal{O}(n^{-2}) \quad \text{as } n \to \infty. \tag{2.6}$$
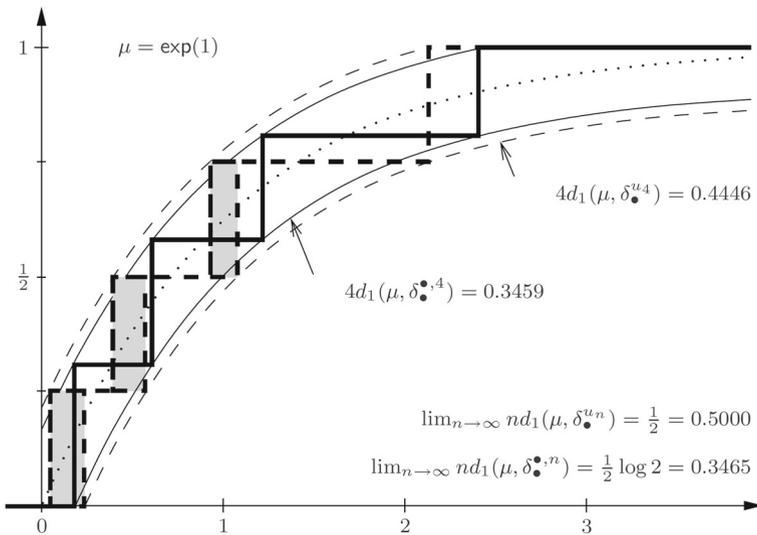
A best d_ϵ-approximation of μ also exists and in fact is unique, with

$$x,j = -\frac{1}{a} \log\left(\frac{e^{2a\ell_\bullet^{\bullet,n}(n-j)/\epsilon} - 1}{e^{2a\ell_\bullet^{\bullet,n}n/\epsilon} - 1} + \ell_\bullet^{\bullet,n} \right) - \frac{\ell_\bullet^{\bullet,n}}{\epsilon}, \quad P,j = \frac{1 - e^{-2a\ell_\bullet^{\bullet,n}j/\epsilon}}{1 - e^{-2a\ell_\bullet^{\bullet,n}n/\epsilon}},$$

for all $j = 1, \ldots, n$, where $\ell_\bullet^{\bullet,n} = d_\epsilon(\mu, \delta_\bullet^{\bullet,n})$ solves $\ell e^{2na\ell/\epsilon} = \ell + \tanh(a\ell/\epsilon)$. Similarly to before, an analysis of this equation yields

$$n d_\epsilon(\mu, \delta_\bullet^{\bullet,n}) = c_1 - \frac{a^2 c_1^2}{6\epsilon(a+\epsilon)} n^{-2} + \mathcal{O}(n^{-4}) \quad \text{as } n \to \infty, \tag{2.7}$$

with $c_1 = \frac{1}{2}\epsilon \log(1 + a/\epsilon)/a < \frac{1}{2}$. Notice that $d_\epsilon(\mu, \delta_\bullet^{\bullet,n}) < d_\epsilon(\mu, \delta_\bullet^{u_n})$ for every $n \ge 2$; see also Fig. 1.

Fig. 1 Illustrating d_1-approximations of the standard exponential distribution (dotted curve) with $n = 4$ atoms: while the best approximation $\delta_\bullet^{\bullet,4}$ (solid line) is unique, best uniform approximations $\delta_\bullet^{u_4}$ (broken lines) are not; see Example 1

Example 2 Fix $b > 1$, and let $F_\mu(x) = \dfrac{\log x}{\log b}$ for all $1 \le x \le b$. Usually referred to as *Benford's law* base b, this distribution has many interesting properties; see, e.g. [2,4] and the references therein. As in the previous example, best uniform d_ϵ-approximations of μ are non-unique, yet $\ell_\bullet^{u_n}$ is the unique solution of $b^{1-\ell} - b^{1+\ell-1/n} = 2\ell/\epsilon$, which in turn yields

$$n d_\epsilon(\mu, \delta_\bullet^{u_n}) = c_2 - \frac{c_2^2}{b\epsilon} n^{-1} + \mathcal{O}(n^{-2}) \quad \text{as } n \to \infty, \tag{2.8}$$

with $c_2 = \frac{1}{2}\epsilon b \log b/(1 + \epsilon b \log b) < \frac{1}{2}$. Also similarly to Example 1, best d_ϵ-approximations of μ are unique, $\ell_\bullet^{\bullet,n}$ solves $b^{2n\ell}(\ell + \epsilon \sinh(\ell \log b)) = \ell + \epsilon b \sinh(\ell \log b)$, and a straightforward analysis yields

$$n d_\epsilon(\mu, \delta_\bullet^{\bullet,n}) = c_3 + \frac{(b-1)c_2^2 c_3^2 b^{2c_3-2}}{3\epsilon} n^{-2} + \mathcal{O}(n^{-4}) \quad \text{as } n \to \infty, \tag{2.9}$$

with $c_3 = \dfrac{\log(1 + \epsilon b \log b) - \log(1 + \epsilon \log b)}{2 \log b} < c_2$; notice again $d_\epsilon(\mu, \delta_\bullet^{\bullet,n}) < d_\epsilon(\mu, \delta_\bullet^{u_n})$ for every $n \ge 2$.

Remark 2 The reader may find it instructive to compare the quantitative results in Examples 1 and 2 to their d_W- and d_P-counterparts. On the one hand, for $\mu = \exp(a)$

[37, Ex. 5.8] shows that

$$nd_W(\mu, \delta_\bullet^{u_n}) = \frac{1}{4a} \log n + \mathcal{O}(1) \quad \text{as } n \to \infty,$$

whereas by [18, Ex. 5.7],

$$nd_W(\mu, \delta_\bullet^{\bullet,n}) = \frac{n}{a} \log \frac{n+1}{n} = \frac{1}{a} + \mathcal{O}(n^{-1}) \quad \text{as } n \to \infty.$$

While the authors are not aware of any result regarding best uniform d_P-approximations, it follows from [20, Prop. 2.1] that

$$nd_P(\mu, \delta_\bullet^{\bullet,n}) = \frac{1}{2a} \log n - \frac{1}{2a} \log \log n + \mathcal{O}(1) \quad \text{as } n \to \infty.$$

On the other hand, for μ being Benford's law base $b > 1$,

$$nd_W(\mu, \delta_\bullet^{u_n}) = \frac{n}{\log b} \tanh\left(\frac{\log b}{4n}\right) = \frac{1}{4} + \mathcal{O}(n^{-2}) \quad \text{as } n \to \infty,$$

$$nd_W(\mu, \delta_\bullet^{\bullet,n}) = \frac{1}{\log b} \tanh\left(\frac{\log b}{4}\right) \quad \forall n \in \mathbb{N},$$

as recorded in [4], and

$$nd_P(\mu, \delta_\bullet^{\bullet,n}) = \frac{b-1}{2} + \mathcal{O}(n^{-1}) \quad \text{as } n \to \infty,$$

again by [20, Prop. 2.1]. When compared to Lévy approximations, therefore, the asymptotics of best (or best uniform) d_W- and d_P-approximations of μ are more sensitive to the tail behaviour of μ. Though perhaps not obvious from (1.1)–(1.3), this is not too surprising either: informally, finiteness of the integral in (1.1) alone necessitates moderate tails for μ, ν, and taking the infimum over all Borel sets in (1.2) makes d_P pick up small deviations near infinity. By contrast, the infimum in (1.3) is, in the parlance of (1.2), taken over *unbounded intervals* only, which in turn causes d_ϵ to be much less sensitive to heavy or irregular tails, an informal observation that is corroborated rigorously by the results in the remaining sections of this article.

3 Best Uniform Lévy Approximations

This section provides a detailed asymptotic analysis of $d_\epsilon(\mu, \delta_\bullet^{u_n})$ for any $\mu \in \mathcal{P}$. Notice the uniform bound $nd_\epsilon(\mu, \delta_\bullet^{u_n}) \leq \frac{1}{2}$, due to Proposition 2.3. Thus, $d_\epsilon(\mu, \delta_\bullet^{u_n}) \to 0$ as $n \to \infty$ at an (upper) rate not slower than (n^{-1}). Except for trivial cases, this rate is sharp.

Lemma 3.1 *Let $\mu \in \mathcal{P}$ and $\epsilon > 0$. Then, $\limsup_{n \to \infty} nd_\epsilon(\mu, \delta_\bullet^{u_n}) > 0$ unless $\mu = \delta_a$ for some $a \in \mathbb{R}$.*

Proof Throughout the proofs of this section, write $g = F_\mu^{-1}$ for convenience, and let $\omega_n = nd_\epsilon(\mu, \delta_\bullet^{u_n})$ for all $n \in \mathbb{N}$, as well as $\omega^- = \liminf_{n \to \infty} \omega_n$ and $\omega^+ = \limsup_{n \to \infty} \omega_n$. Since $\delta_a \circ T_\epsilon^{-1} = \delta_{\epsilon a}$, by (2.2), it suffices to consider the case of $\epsilon = 1$. Fix any $0 < x < y < 1$. Assume that $\omega^+ = 0$, i.e. $\lim_{n \to \infty} \omega_n = 0$. Note that, for every $n \in \mathbb{N}$,

$$g_- \left(\frac{j - 2\omega_n}{n} \right) - \frac{2\omega_n}{n} \leq g \left(\frac{j - 1 + 2\omega_n}{n} \right) + \frac{2\omega_n}{n} \quad \forall j = 1, \ldots, n, \qquad (3.1)$$

by the definition of ω_n. Also, observe that for all sufficiently large n,

$$\frac{j + 2\omega_{n+1}}{n+1} < \frac{j - 2\omega_n}{n} \quad \text{and} \quad \frac{j + 2\omega_n}{n} < \frac{j + 1 - 2\omega_{n+1}}{n+1},$$

for every $j = \lfloor nx \rfloor, \ldots, \lfloor ny \rfloor + 1$, which, together with (3.1), yields

$$\begin{aligned}
g(y) - g(x) &\leq g \left(\frac{\lfloor ny \rfloor + 1 + 2\omega_n}{n} \right) - g_- \left(\frac{\lfloor nx \rfloor - 2\omega_n}{n} \right) \\
&= \sum_{j=\lfloor nx \rfloor}^{\lfloor ny \rfloor + 1} \left(g \left(\frac{j + 2\omega_n}{n} \right) - g_- \left(\frac{j - 2\omega_n}{n} \right) \right) \\
&\quad + \sum_{j=\lfloor nx \rfloor + 1}^{\lfloor ny \rfloor + 1} \left(g_- \left(\frac{j - 2\omega_n}{n} \right) - g \left(\frac{j - 1 + 2\omega_n}{n} \right) \right) \\
&\leq \sum_{j=\lfloor nx \rfloor}^{\lfloor ny \rfloor + 1} \left(g_- \left(\frac{j + 1 - 2\omega_{n+1}}{n+1} \right) - g \left(\frac{j + 2\omega_{n+1}}{n+1} \right) \right) \\
&\quad + \frac{4\omega_n}{n} (\lfloor ny \rfloor - \lfloor nx \rfloor + 1) \\
&\leq \frac{4\omega_{n+1}}{n+1} (\lfloor ny \rfloor - \lfloor nx \rfloor + 2) + \frac{4\omega_n}{n} (\lfloor ny \rfloor - \lfloor nx \rfloor + 1) \\
&\leq 12\omega_{n+1} + 8\omega_n.
\end{aligned}$$

Since $\lim_{n \to \infty} \omega_n = 0$ by assumption, and $0 < x < y < 1$ have been arbitrary, $g(0) = g_-(1) = a$ for some $a \in \mathbb{R}$, that is, $\mu = \delta_a$. $\qquad \square$

For the subsequent finer analysis, the following terminology is useful: for every $f \in \mathcal{F}$, let G_f be the *growth set* of f, i.e. let

$$G_f = \{ x \in \mathbb{R} : f(x - \varepsilon) < f(x + \varepsilon) \; \forall \varepsilon > 0 \}.$$

Note that G_f is closed in $\mathbb{R}$, and $G_f \neq \varnothing$, unless f is constant. For example, $G_{F_\mu} = \operatorname{supp} \mu$ and $\{0, 1\} \subset G_{F_\mu^{-1}} \subset [0, 1]$ for every $\mu \in \mathcal{P}$. Also, $f(x) \in \mathbb{R}$ whenever $f^{-1}(-\infty) < x < f^{-1}(+\infty)$. With $\lambda_f \left( ]-\infty, f^{-1}(-\infty)] \right) := \lambda_f \left([f^{-1}(+\infty), +\infty[\right) := 0$ and

$$\lambda_f(]x, y]) := f(y) - f(x) \quad \forall f^{-1}(-\infty) < x \le y < f^{-1}(+\infty),$$

therefore, λ_f is a σ-finite positive Borel measure concentrated on G_f. For example, $\lambda_{\mathrm{id}_{\mathbb{R}}} = \lambda$, and $\lambda_{F_\mu} = \mu$ for every $\mu \in \mathcal{P}$. Also, $\mu^{-1} := \lambda_{F_\mu^{-1}}$ is a positive Borel measure supported on $G_{F_\mu^{-1}} \subset [0, 1]$, referred to as the *inverse measure* of μ; see, e.g. [6,37]. For convenience, write G_{F_μ} and $G_{F_\mu^{-1}}$ simply as G_μ and $G_{\mu^{-1}}$, respectively. Note that $\mu^{-1}(\mathbb{R}) = \mu^{-1}(]0, 1[) = \mathrm{diam}\, G_\mu$, and hence $\mu^{-1} = 0$ precisely if $\mu = \delta_a$ for some $a \in \mathbb{R}$. When rephrased utilizing this terminology, Lemma 3.1 has the following corollary.

Proposition 3.2 *For every $\mu \in \mathcal{P}$ and $\epsilon > 0$, the following are equivalent:*

(i) $\lim_{n\to\infty} n d_\epsilon(\mu, \delta_{\bullet}^{u_n}) = 0$;
(ii) $d_\epsilon(\mu, \delta_{\bullet}^{u_n}) = 0$ *for every* $n \in \mathbb{N}$;
(iii) $\mu = \delta_a$ *for some* $a \in \mathbb{R}$;
(iv) $\mu^{-1} = 0$.

The first main result in this section asserts that $\bigl(n d_\epsilon(\mu, \delta_{\bullet}^{u_n})\bigr)$ does converge, to an easily determined limit, if μ^{-1} is absolutely continuous. The result is reminiscent of a theorem regarding best uniform d_{W}-approximations [37, Thm. 5.15] (see also [7,16]), but unlike in that theorem, no integrability assumption on $d\mu^{-1}/d\lambda$ is needed, and the limit in question always is finite. When formulating the result, it is helpful to use the function $\Omega : \mathbb{R} \to \mathbb{R}$ with

$$\Omega(x) = \frac{x}{2 + 2|x|} \quad \forall x \in \mathbb{R}.$$

Plainly, Ω is an increasing C^1-function, with $|\Omega(x)| \le \frac{1}{2}|x|$ for all $x \in \mathbb{R}$, and $\Omega(\pm\infty) = \pm\frac{1}{2}$. While the appearance of Ω in the following theorem is a simple consequence of the bound (3.3), the reader may find it curious to note that 2Ω plays a prominent role in the theory of random walks [12].

Theorem 3.3 *Let $\mu \in \mathcal{P}$ and $\epsilon > 0$. If μ^{-1} is absolutely continuous (w.r.t. λ), then*

$$\lim_{n\to\infty} n d_\epsilon(\mu, \delta_{\bullet}^{u_n}) = \left\| \Omega\left(\epsilon \frac{d\mu^{-1}}{d\lambda}\right) \right\|_\infty. \tag{3.2}$$

Proof Since $\dfrac{d(\mu \circ T_\epsilon^{-1})^{-1}}{d\lambda} = \epsilon \dfrac{d\mu^{-1}}{d\lambda}$, it is enough to prove (3.2) for $\epsilon = 1$. Using the same symbols as in the proof of Lemma 3.1, for every $n \in \mathbb{N}$ let

$$J_{n,j} = \left[\frac{j - 1 + \omega_n}{n}, \frac{j - \omega_n}{n}\right] \quad \forall j = 1, \ldots, n.$$

Note that $\omega_n < \frac{1}{2}$ for every n since g is continuous. Moreover,

$$g\left(\frac{j - \omega_n}{n}\right) - g\left(\frac{j - 1 + \omega_n}{n}\right) \le \frac{2\omega_n}{n} \quad \forall j = 1, \ldots, n,$$

and consequently, by the absolute continuity of g,

$$\frac{2\omega_n}{1 - 2\omega_n} \geq \frac{g\big((j - \omega_n)/n\big) - g\big((j - 1 + \omega_n)/n\big)}{(1 - 2\omega_n)/n}$$

$$= \frac{1}{\lambda(J_{n,j})} \int_{J_{n,j}} g' \, d\lambda \quad \forall j = 1, \ldots, n. \tag{3.3}$$

Equality holds on the left in (3.3) for at least one j, and for that j,

$$\frac{2\omega_n}{1 - 2\omega_n} = \frac{1}{\lambda(J_{n,j})} \int_{J_{n,j}} g' \, d\lambda \leq \|g'\|_\infty,$$

from which it is clear that

$$\omega^+ \leq \frac{\|g'\|_\infty}{2 + 2\|g'\|_\infty} = \|\Omega(g')\|_\infty \leq \frac{1}{2}. \tag{3.4}$$

Since (3.2) trivially holds when $\omega^- = \frac{1}{2}$, henceforth assume $\omega^- < \frac{1}{2}$, and pick $n_1 < n_2 < \ldots$ so that $\lim_{k\to\infty} \omega_{n_k} = \omega^-$. Given any $0 < x < 1$, let $j_k(x) = \lfloor n_k x \rfloor + 1 \in \{1, \ldots, n_k\}$, and note that $J_{n_k, j_k(x)} \subset [x - 1/n_k, x + 1/n_k]$, but also

$$\frac{\lambda(J_{n_k, j_k(x)})}{\lambda([x - 1/n_k, x + 1/n_k])} = \frac{1}{2} - \omega_{n_k} \xrightarrow{k\to\infty} \frac{1}{2} - \omega^- > 0.$$

Thus, the sequence $(J_{n_k, j_k(x)})_{k\in\mathbb{N}}$ shrinks to x nicely, and by [33, Thm. 7.10],

$$g'(x) = \lim_{k\to\infty} \frac{1}{\lambda(J_{n_k, j_k(x)})} \int_{J_{n_k, j_k(x)}} g' \, d\lambda \leq \lim_{k\to\infty} \frac{2\omega_{n_k}}{1 - 2\omega_{n_k}} = \frac{2\omega^-}{1 - 2\omega^-}$$

for λ-almost every $x \in [0, 1]$. It follows that $\omega^- \geq \Omega(g')$ holds λ-almost everywhere, and hence $\omega^- \geq \|\Omega(g')\|_\infty$. Together with (3.4), this completes the proof. $\qquad\square$

If μ^{-1} is singular, then the asymptotic behaviour of $\big(d_\epsilon(\mu, \delta_\bullet^{u_n})\big)$ is controlled by a different mechanism. To prepare for the general result, observe that $\liminf_{n\to\infty} \operatorname{dist}(nx, \mathbb{Z}) = 0$ for every $x \in \mathbb{R}$, whereas

$$\limsup_{n\to\infty} \operatorname{dist}(nx, \mathbb{Z}) = \begin{cases} \frac{1}{2}(q - 1)/q & \text{if } x = p/q \text{ with coprime} \\ & \quad p \in \mathbb{Z}, q \in \mathbb{N}, q \text{ odd,} \\ \frac{1}{2} & \text{otherwise;} \end{cases} \tag{3.5}$$

in particular, $\limsup_{n\to\infty} \operatorname{dist}(nx, \mathbb{Z}) \geq \frac{1}{3}$ unless $x \in \mathbb{Z}$. Defining $\iota : \mathbb{R} \to \mathbb{N}_0 \cup \{+\infty\}$ as

$$\iota(x) = 2 \inf\{n \in \mathbb{N}_0 : (2n + 1)x \in \mathbb{Z}\},$$

notice that the right-hand side in (3.5) is nothing other than $\Omega \circ \iota(x)$. With this, consider the very simple example of $\mu_a = a\delta_{-1} + (1-a)\delta_1$ for some $0 < a < 1$, for which $G_{\mu_a^{-1}} = \{0, a, 1\}$. It is readily confirmed that $nd_\epsilon(\mu_a, \delta_\bullet^{u_n}) = \text{dist}(na, \mathbb{Z})$ for all sufficiently large n, and hence $\liminf_{n\to\infty} nd_\epsilon(\mu_a, \delta_\bullet^{u_n}) = 0$, as well as

$$\limsup_{n\to\infty} nd_\epsilon(\mu_a, \delta_\bullet^{u_n}) = \Omega \circ \iota(a) = \max \Omega \circ \iota(G_{\mu_a^{-1}}).$$

This equality is but one manifestation of a general principle.

Lemma 3.4 *Let $\mu \in \mathcal{P}$ and $\epsilon > 0$. If μ^{-1} is singular (w.r.t. λ), then*

$$\limsup_{n\to\infty} nd_\epsilon(\mu, \delta_\bullet^{u_n}) = \sup \Omega \circ \iota(G_{\mu^{-1}}). \tag{3.6}$$

Proof Using the same symbols as in previous proofs, write $G_{\mu^{-1}}$ simply as G, and let $2m = \sup \iota(G)$ with the appropriate $m \in \mathbb{N}_0 \cup \{+\infty\}$; also, let G^* be the set of atoms of μ^{-1}, i.e. $G^* = \{0 < x < 1 : g_-(x) < g(x)\}$. Assume first that $m \in \mathbb{N}_0$. Since $m = 0$ implies $G = \{0, 1\}$, or equivalently $\mu^{-1} = 0$, and (3.6) is correct in this case by Proposition 3.2, henceforth assume $m \geq 1$. Then, μ^{-1} is concentrated on finitely many atoms, thus

$$G^* = G = \left\{0, \frac{k_1}{2m_1 + 1}, \ldots, \frac{k_l}{2m_l + 1}, 1\right\},$$

with the appropriate positive integers $l, k_1, \ldots, k_l, m_1, \ldots, m_l$, where the numbers $k_i, 2m_i + 1$ are coprime for all i, and $\max_{i=1}^l m_i = m$. As seen in the example above, for all sufficiently large n,

$$\omega_n = \max_{i=1}^l \text{dist}\left(\frac{nk_i}{2m_i + 1}, \mathbb{Z}\right),$$

and hence

$$\omega^+ = \max_{i=1}^l \Omega \circ \iota\left(\frac{k_i}{2m_i + 1}\right) = \max_{i=1}^l \Omega(2m_i) = \Omega(2m),$$

so again (3.6) is correct. It remains to consider the case of $m = +\infty$. Here, it is convenient to consider two subcases, depending on whether $\iota(G^*)$ is unbounded or not. In the former case, fix $a \in \mathbb{R}^+$, and pick $x \in G^*$ with $\iota(x) \geq a$. Moreover, pick $b > 3$, and recall that $y_n := (\text{dist}(nx, \mathbb{Z}) - 1/b)/n > 0$ for infinitely many $n \in \mathbb{N}$. Since x is an atom of μ^{-1}, for every $c \in \mathbb{R}^+$ clearly

$$g_-\left(x + \frac{1}{bn}\right) - g_-\left(x - \frac{1}{bn}\right) \geq \frac{c}{bn} \quad \text{for all sufficiently large } n. \tag{3.7}$$

Choosing $c = b/\epsilon$ in (3.7), note that for infinitely many n,

$$g_- \left(\frac{\lfloor nx \rfloor + 1 - ny_n}{n} \right) - g \left(\frac{\lfloor nx \rfloor + ny_n}{n} \right)$$

$$= g_- \left(x + \frac{\max\{1 - 2\text{dist}(nx, \mathbb{Z}), 0\}}{n} + \frac{1}{bn} \right)$$

$$- g \left(x + \frac{\min\{1 - 2\text{dist}(nx, \mathbb{Z}), 0\}}{n} - \frac{1}{bn} \right)$$

$$\geq g_- \left(x + \frac{1}{bn} \right) - g \left(x - \frac{1}{bn} \right) \geq \frac{1}{n\epsilon} \geq \frac{2y_n}{\epsilon},$$

and consequently $\omega_n \geq ny_n$. It follows that

$$\omega^+ \geq \limsup_{n \to \infty} \left(\text{dist}(nx, \mathbb{Z}) - \frac{1}{b} \right) = \Omega \circ \iota(x) - \frac{1}{b} \geq \Omega(a) - \frac{1}{b}.$$

Since $a, b > 3$ have been arbitrary, $\omega^+ = \frac{1}{2} = \Omega(2m)$. Finally, assume that $\iota(G^*)$ is bounded, and hence G^* is finite, possibly empty. Since $m = +\infty$, clearly $G \backslash G^* \neq \varnothing$, and every $x \in G \backslash G^*$ is a continuity point of g, as well as an accumulation point of G. By [33, Thm. 7.15],

$$\lim_{\varepsilon \downarrow 0} \frac{g_-(x + \varepsilon) - g(x - \varepsilon)}{2\varepsilon} = +\infty \quad \text{for } \mu^{-1}\text{-almost every } 0 < x < 1.$$

From this, it is clear that, given any $b, c > 3$, there exists $x \in G \backslash \mathbb{Q}$ for which (3.7) holds. With $\iota(x) = a = +\infty$, the same argument as before shows that $\omega^+ = \frac{1}{2}$, i.e. (3.6) is correct in this case also. □

Combining Theorem 3.3 and Lemma 3.4 yields a sharp (upper) rate for $(d_\epsilon(\mu, \delta_\bullet^{u_n}))$, for arbitrary $\mu \in \mathcal{P}$. To formulate the result, recall that every σ-finite Borel measure ρ on the real line can be written uniquely as $\rho = \rho_A + \rho_S$, where ρ_A and ρ_S are absolutely continuous and singular (w.r.t. λ), respectively.

Theorem 3.5 *Let $\mu \in \mathcal{P}$ and $\epsilon > 0$. Then,*

$$\limsup_{n \to \infty} nd_\epsilon(\mu, \delta_\bullet^{u_n}) = \max \left\{ \left\| \Omega \left(\epsilon \frac{d(\mu^{-1})_A}{d\lambda} \right) \right\|_\infty, \sup \Omega \circ \iota(G_{(\mu^{-1})_S}) \right\}.$$

$$(3.8)$$

Proof Since there is nothing to prove otherwise, assume that $(\mu^{-1})_A \neq 0$ and $(\mu^{-1})_S \neq 0$. In analogy to the proof of Lemma 3.4, let $g = F_\mu^{-1} = g_A + g_S$, with $g_A, g_S \in \mathcal{F}$ such that $\lambda_{g_A} = (\mu^{-1})_A$ and $\lambda_{g_S} = (\mu^{-1})_S$, as well as $2m = \sup \iota(G_S) \in \mathbb{N}_0 \cup \{+\infty\}$, where $G_S = G_{(\mu^{-1})_S}$ for convenience. Since $g_-(y) - g(x) \geq g_{S-}(y) - g_S(x)$ for all $0 < x < y < 1$, Lemma 3.4 yields $\omega^+ \geq \sup \Omega \circ \iota(G_S) = \Omega(2m)$. Thus, (3.8) clearly is correct when $m = +\infty$, and only the case of $m \in \mathbb{N}$ remains to be considered (note that $m = 0$ is impossible,

as it would imply $(\mu^{-1})_S = 0$). In this case, G_S is finite, say, $G_S = \{0, x_1, \ldots, x_l, 1\}$ with $l \in \mathbb{N}$ and $0 < x_1 < \ldots < x_l < 1$. With $J_{n,j}$ as in the proof of Theorem 3.3, and letting $K_n = \{\lfloor nx_i \rfloor : i = 1, \ldots, l\} \subset \{1, \ldots, n-1\}$ for $n \geq 1/x_1$, observe that

$$g_A\left(\frac{j - \omega_n}{n}\right) - g_A\left(\frac{j - 1 + \omega_n}{n}\right) = g_-\left(\frac{j - \omega_n}{n}\right) - g\left(\frac{j - 1 + \omega_n}{n}\right)$$
$$\leq \frac{2\omega_n}{n\epsilon} \quad \forall j \notin K_n,$$

and consequently

$$\frac{1}{\lambda(J_{n,j})} \int_{J_{n,j}} \epsilon g_A' \, d\lambda \leq \frac{2\omega_n}{1 - 2\omega_n} \quad \forall j \notin K_n.$$

If $\omega^- < \frac{1}{2}$, then the same argument as in the proof of Theorem 3.3 shows that

$$\epsilon g_A'(x) \leq \frac{2\omega^-}{1 - 2\omega^-} \quad \text{for } \lambda\text{-almost every } x,$$

since clearly $j_n(x) \notin K_n$ whenever $x \notin G_S$ and n is sufficiently large. Thus, $\omega^- \geq \|\Omega(\epsilon g_A')\|_\infty$; trivially, the latter also holds when $\omega^- = \frac{1}{2}$. In summary, $\omega^+ \geq \max\{\|\Omega(\epsilon g_A')\|_\infty, \Omega(2m)\} =: \omega$; note that ω simply equals the right-hand side in (3.8).

The reverse inequality is non-trivial only when $\omega < \frac{1}{2}$. In this case, assume $m \in \mathbb{N}$ as before, and pick any z with $\omega < z < \frac{1}{2}$. Then, for all sufficiently large n,

$$g_A\left(\frac{j - z}{n}\right) - g_A\left(\frac{j - 1 + z}{n}\right) \leq \frac{2z}{\epsilon n} \quad \forall j = 1, \ldots, n,$$

but also, since G_S is finite,

$$g_{S-}\left(\frac{j - z}{n}\right) - g_S\left(\frac{j - 1 + z}{n}\right) = 0 \quad \forall j = 1, \ldots, n.$$

Thus, $\omega_n \leq z$ for all sufficiently large n, and since $z > \omega$ was arbitrary, $\omega^+ \leq \omega$. □

Corollary 3.6 *Let $\mu \in \mathcal{P}$ and $\epsilon > 0$.*

(i) *If $\limsup_{n\to\infty} nd_\epsilon(\mu, \delta_\bullet^{u_n}) < \frac{1}{3}$, then $\left(nd_\epsilon(\mu, \delta_\bullet^{u_n})\right)$ converges, and μ^{-1} is absolutely continuous (w.r.t. λ).*

(ii) *If $\limsup_{n\to\infty} nd_\epsilon(\mu, \delta_\bullet^{u_n}) \geq \frac{1}{3}$, then either $\left(nd_\epsilon(\mu, \delta_\bullet^{u_n})\right)$ converges, or*

$$\limsup_{n\to\infty} nd_\epsilon(\mu, \delta_\bullet^{u_n}) \in \left\{\frac{1}{3}, \frac{2}{5}, \frac{3}{7}, \ldots, \frac{1}{2}\right\} = \left\{\Omega(2m) : m \in \mathbb{N} \cup \{+\infty\}\right\}.$$

Remark 3 (i) The proof given above shows that, for every $\mu \in \mathcal{P}$ and $\epsilon > 0$,

$$\liminf_{n \to \infty} n d_\epsilon(\mu, \delta_\bullet^{u_n}) \geq \left\| \Omega \left(\epsilon \frac{d(\mu^{-1})_A}{d\lambda} \right) \right\|_\infty.$$

(ii) Let $\mu \in \mathcal{P}$ be non-atomic. Then, the right-hand side in (3.8) tends to $\frac{1}{2}$ as $\epsilon \to +\infty$. This is consistent with the fact that $n \min_{x \in \Xi_n} \| F_\mu - F_{\delta_{x_n}^{u_n}} \|_\infty = \frac{1}{2}$ for all $n \in \mathbb{N}$ whenever μ is non-atomic [4, Cor. 5.5].

The following example illustrates the results of this section. In particular, it demonstrates that all situations allowed by Theorem 3.5 and Corollary 3.6 do occur. It also shows that (3.2) may fail when μ^{-1} is not absolutely continuous; similarly, (3.6) may fail when μ^{-1} is not singular.

Example 3 For every $0 \leq a < 1 < b$, consider $\mu_{a,b} = a\delta_{-1} + (1-a)U_{1,b}$, where $U_{1,b}$ denotes the uniform distribution (normalized Lebesgue measure) on $[1, b]$. Note that $\mu_{a,1} := \lim_{b \downarrow 1} \mu_{a,b} = \mu_a$, with μ_a considered prior to Lemma 3.4. Since $(\mu_{a,b}^{-1})_A = (b-1)U_{a,1}$ and $(\mu_{a,b}^{-1})_S = 2\delta_a$ provided that $a > 0$, Theorem 3.5 yields

$$\limsup_{n \to \infty} n d_\epsilon(\mu_{a,b}, \delta_\bullet^{u_n}) = \max \left\{ \Omega \left(\epsilon \frac{b-1}{1-a} \right), \Omega \circ \iota(a) \right\},$$

whereas by direct inspection,

$$\liminf_{n \to \infty} n d_\epsilon(\mu_{a,b}, \delta_\bullet^{u_n}) = \Omega \left(\epsilon \frac{b-1}{1-a} \right) < \frac{1}{2}.$$

On the one hand, if $a = a_m = m/(2m+1)$ for some $m \in \mathbb{N}_0$, then $\Omega \circ \iota(a_m) = a_m$, and since $b \mapsto \Omega(\epsilon(b-1)/(1-a_m))$ is increasing continuously from 0 to $\frac{1}{2}$, there exists a unique b_m with $\Omega(\epsilon(b_m - 1)/(1 - a_m)) = a_m$. Thus, $(n d_\epsilon(\mu_{a_m,b}, \delta_\bullet^{u_n}))$ converges precisely if $b \geq b_m$, whereas for $b < b_m$ the lim inf can have any value between 0 and a_m. On the other hand, if $a = \frac{1}{2}$ then $\Omega \circ \iota(a) = \frac{1}{2}$, and again $\liminf_{n \to \infty} n d_\epsilon(\mu_{1/2,b}, \delta_\bullet^{u_n}) = \Omega(2\epsilon(b-1))$ can have any value between 0 and $\frac{1}{2}$. Except for the case of $\lim_{n \to \infty} n d_\epsilon(\mu, \delta_\bullet^{u_n}) = \frac{1}{2}$, which occurs, e.g. for the exponential distributions in Example 1, every possible situation allowed by Theorem 3.5 can be observed in this example by choosing a, b appropriately; see also Fig. 2. Notice that $\mu_{a,b}^{-1}$ is absolutely continuous precisely if $a = 0$ and is singular only if $b = 1$. While Theorem 3.3 and Lemma 3.4 thus cannot in general be reversed, clearly (3.2) and (3.6) may fail if μ^{-1} is not absolutely continuous and singular, respectively.

Example 4 Let μ be a normal distribution with variance $\sigma^2 > 0$. With ϕ denoting the standard normal distribution function, it is readily deduced from Proposition 2.5 that $\ell_\bullet^{u_n} = d_\epsilon(\mu, \delta_\bullet^{u_n})$ is the unique solution of $\phi^{-1}(1-\ell) - \phi^{-1}(1+\ell-1/n) = 2\ell/\sqrt{\epsilon^2 \sigma^2}$. Utilizing the familiar fact [1, Sec. 26.2]

$$1 - \phi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \left(x^{-1} - x^{-3} + \mathcal{O}(x^{-5}) \right) \quad \text{as } x \to +\infty,$$

Fig. 2 Solid black lines indicate, for any $\mu \in \mathcal{P}$, the possible values of the limit inferior and the limit superior of $\left(nd_\epsilon(\mu, \delta_\bullet^{u_n})\right)$; see Theorems 3.3 and 3.5. Example 3 demonstrates that all possible values may indeed occur
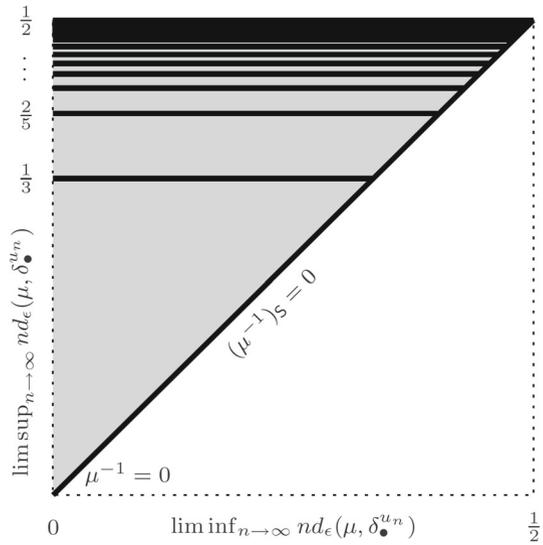
a straightforward analysis of this equation yields

$$nd_\epsilon(\mu, \delta_\bullet^{u_n}) = \frac{1}{2} - \frac{1}{2\epsilon\sqrt{2\sigma^2}} \cdot \frac{\sqrt{\log n}}{n} + \mathcal{O}(n^{-1}) \quad \text{as } n \to \infty, \qquad (3.9)$$

which sharpens (3.2). Notice that convergence occurs here at a slightly slower rate than has so far been observed in this article for examples of absolutely continuous μ^{-1}; cf. [16] and [37, Ex. 5.18].

As Examples 1, 2 and 4 suggest, the results of this section, notably Theorem 3.3, can be refined by imposing further assumptions on μ. For instance, assume that $g = F_\mu^{-1}$ is C^2 on $]0, 1[$ and that $g, g' \neq 0$ both are convex, with $\limsup_{x \uparrow 1}(1 - x)g''(x)/g'(x) < +\infty$ (all mentioned examples meet these requirements; for normal distributions, only the right half of $]0, 1[$ has to be considered due to symmetry). Letting $c = \|\Omega(\epsilon g')\|_\infty > 0$ for convenience, it is straightforward to show that, as a refinement of (3.2),

$$nd_\epsilon(\mu, \delta_\bullet^{u_n}) = c - \frac{2c^2}{\epsilon}e_n + o(e_n) \quad \text{as } n \to \infty, \qquad (3.10)$$

where $\lim_{n \to \infty} e_n = 0$, and more specifically,

$$e_n = \begin{cases} \dfrac{1}{g'_-(1)} + n\dfrac{g(1 - (1-c)/n) - g(1 - c/n)}{(1 - 2c)g'_-(1)^2} & \text{if } c < \frac{1}{2}, \\[2ex] \dfrac{1}{g'\left(1 - 1/(2n)\right)} & \text{if } c = \frac{1}{2}; \end{cases}$$

in particular, if $g''_-(1) < +\infty$ then simply

$$e_n = \frac{g''_-(1)}{2g'_-(1)^2} n^{-1}.$$

As the reader may want to check, for exponential, Benford and normal distributions, the asymptotic equalities (2.6), (2.8) and (3.9), respectively, all are (slightly sharper than, but certainly) consistent with (3.10).

Example 5 Let μ be the *Cantor distribution*, i.e. the $\log 2/\log 3$-dimensional Hausdorff measure on the classical Cantor middle thirds set C. Thus, $G_\mu = C$, and since $\operatorname{diam} C = 1$, the measure μ^{-1}, referred to as the *inverse Cantor distribution* [6, Ex. A.11], is a probability measure as well. Both μ, μ^{-1} are singular: while μ is non-atomic, μ^{-1} is purely atomic; in fact, $\mu^{-1}(\{i2^{-m}\}) = 3^{-m}$ for every $m \in \mathbb{N}$ and every odd $1 \leq i < 2^m$. Obviously, Lemma 3.4 applies to both distributions, showing that

$$\limsup_{n\to\infty} nd_\epsilon(\mu, \delta_\bullet^{u_n}) = \limsup_{n\to\infty} nd_\epsilon(\mu^{-1}, \delta_\bullet^{u_n}) = \frac{1}{2}.$$

By contrast, it is straightforward to check that $\liminf_{n\to\infty} nd_\epsilon(\mu, \delta_\bullet^{u_n}) = 0$. For the inverse Cantor distribution, an elementary analysis [4,36] yields $\frac{1}{216} \leq \liminf_{n\to\infty} nd_1(\mu^{-1}, \delta_\bullet^{u_n}) \leq \frac{1}{3}$, but the authors do not know the precise value of $\liminf_{n\to\infty} nd_\epsilon(\mu^{-1}, \delta_\bullet^{u_n})$ for any $\epsilon > 0$.

4 Best (Unconstrained) Lévy Approximations

This section studies the asymptotics of $d_\epsilon(\mu, \delta_\bullet^{\bullet,n})$ as $n \to \infty$. The following theorem is the section's main result and a counterpart to Theorems 3.3 and 3.5. It asserts that the sequence $(nd_\epsilon(\mu, \delta_\bullet^{\bullet,n}))$ always converges, to a limit smaller than $\frac{1}{2}$ that is easily expressed in terms of Ω and $(\mu^{-1})_A$, the absolutely continuous part of μ^{-1}.

Theorem 4.1 *Let $\mu \in \mathcal{P}$ and $\epsilon > 0$. Then,*

$$\lim_{n\to\infty} nd_\epsilon(\mu, \delta_\bullet^{\bullet,n}) = \int \Omega\left(\epsilon \frac{d(\mu^{-1})_A}{d\lambda}\right) d\lambda. \tag{4.1}$$

As pointed out already in Introduction, Theorem 4.1 may be regarded as an analogue for d_ϵ of (the one-dimensional version of) a classical d_W-quantization theorem [18, Thm. 6.2], but unlike that theorem, it does not impose a moment assumption on $(\mu^{-1})_A$. For d_W-approximations, such an assumption is known to be essential, as the d_W-approximation error for $\mu \in \mathcal{P}_1$ can decay arbitrarily slowly without it; see [18, Ex. 6.4] and [37, Thm. 5.33]. In the light of Theorem 4.1, this may be viewed as an artefact specific to d_W that does not exist for d_ϵ. Also, recall that unlike d_W, the metric d_ϵ metrizes precisely the topology of weak convergence on all of $\mathcal{P}$, and so does d_P. As far as the authors have been able to ascertain, however, all known results pertaining

to the asymptotics of the d_{P}-approximation error also impose additional assumptions [20, Sec. 4], and despite the similarities between d_{P} and d_ϵ, [20, Sec. 5] suggests that the d_{P}-approximation error can decay arbitrarily slowly as well.

When proving Theorem 4.1, the following observation, a direct consequence of Proposition 2.5 together with the argument establishing [4, Thm. 3.9], is helpful; its routine verification is left to the interested reader.

Proposition 4.2 *Let $\mu \in \mathcal{P}$, $\epsilon > 0$, and $n \in \mathbb{N}$. With $\ell = d_\epsilon(\mu, \delta_{\bullet}^{\cdot,n})$, there exists $p \in \Pi_n$ such that for every $j = 1, \ldots, n$,*

$$\mathrm{dist}(P_{,j}, G_{\mu^{-1}}) \le \ell \quad \text{and}$$

$$\mu^{-1}(]P_{,j-1} + \ell, P_{,j} - \ell[) \le \frac{2\ell}{\epsilon} \le \mu^{-1}([P_{,j-1} + \ell, P_{,j} - \ell]).$$

Proof of Theorem 4.1 Throughout this proof, for convenience let $g = F_\mu^{-1}$ and $G = G_{\mu^{-1}}$ as before, but also $\ell_n = d_\epsilon(\mu, \delta_{\bullet}^{\cdot,n})$ and $\omega_n = n\ell_n$ for all $n \in \mathbb{N}$, as well as $\omega^- = \liminf_{n\to\infty} \omega_n$ and $\omega^+ = \limsup_{n\to\infty} \omega_n$. Again it suffices to consider the case of $\epsilon = 1$. Note that $\ell_n = 0$ for some (and hence all sufficiently large) $n \in \mathbb{N}$ if and only if G is finite, in which case (4.1) clearly is correct. Thus, assume G to be infinite from now on, and consequently $\ell_n > 0$ for all $n \in \mathbb{N}$.

Given $n \in \mathbb{N}$, choose $p_n \in \Pi_n$ as in Proposition 4.2 and notice that $\ell_n > 0$ implies $\min_{j=1}^n (P_{n,j} - P_{n,j-1}) \ge 2\ell_n > 0$; in particular, $P_{n,j-1} < P_{n,j}$ for all $j = 1, \ldots, n$. Consequently, for every $x \in [0, 1[$ there exists a unique $j_n(x) \in \{1, \ldots, n\}$ with $P_{n,j_n(x)-1} \le x < P_{n,j_n(x)}$. For convenience, let $J_{n,j} = [P_{n,j-1} + \ell_n, P_{n,j} - \ell_n]$ for all $j = 1, \ldots, n$, and hence $\lambda_{n,j} := \lambda(J_{n,j}) = P_{n,j} - P_{n,j-1} - 2\ell_n$. Next, recall that the set $U := [0,1]\backslash G$ is open, possibly empty. If $U \ne \varnothing$, let $I_1, I_2, \ldots$ be its (at most countably many) connected components, that is, the disjoint open intervals with endpoints in G and $U = \bigcup_k I_k$. Thus, for every $x \in U$ there exists a unique $k(x) \in \{1, 2, \ldots\}$ with $x \in I_{k(x)}$. Finally, consider the subset $G^\dagger$ of G defined as

$$G^\dagger = \{x \in G : I \cap G = \{x\} \text{ for some interval } I \text{ with } \lambda(I) > 0\}. \quad (4.2)$$

Notice that $\{0, 1\} \subset G^\dagger$, and $G^\dagger$ is (at most) countable. Utilizing Proposition 4.2, it is readily checked that

$$\lim_{n\to\infty} [P_{n,j_n(x)-1}, P_{n,j_n(x)}] = \{x\} \quad \forall x \in G\backslash G^\dagger, \quad (4.3)$$

but also

$$\lim_{n\to\infty} [P_{n,j_n(x)-1}, P_{n,j_n(x)}] = I_{k(x)} \quad \forall x \in U. \quad (4.4)$$

With these preparations, the proof is now carried out in three separate steps for the reader's convenience.

Step I: Assume μ^{-1} is absolutely continuous.
Proposition 4.2 with $\epsilon = 1$ yields $\mu^{-1}(J_{n,j}) = 2\ell_n$ or, equivalently,

$$g(P_{n,j} - \ell_n) - \ell_n = g(P_{n,j-1} + \ell_n) + \ell_n \quad \forall j = 1, \ldots, n, \quad (4.5)$$

and since μ^{-1} is absolutely continuous also $\lambda_{n,j} > 0$. Fix any $0 < a < 1$, and recalling that g is differentiable λ-almost everywhere, with $g' \geq 0$ integrable over every compact subinterval of $]0, 1[$, pick a nonnegative continuous function $f_a :$ $]0, 1[\to \mathbb{R}$ with $\int_{[0,1]} |g' - f_a| \, d\lambda < a$ (notice that g', f_a may not be integrable over $[0, 1]$). If $U \neq \varnothing$, then also pick $k_a \in \mathbb{N}$ large enough to ensure $\lambda(\bigcup_{k > k_a} I_k) < a$, and for every $k = 1, \ldots, k_a$ pick a continuous function $e_k :]0, 1[\to [0, 1]$ with $e_k(x) = 1$ for all $x \in]0, 1[\setminus I_k$ such that $\int_{I_k} f_a e_k < a\lambda(I_k)$. Let $f = f_a \prod_{k=1}^{k_a} e_k$; in case $U = \varnothing$, simply let $f = f_a$. Clearly, f is nonnegative and continuous on $]0, 1[$, with

$$\int_{[0,1]} |g' - f| \, d\lambda = \int_G |g' - f_a| \, d\lambda + \int_U |f| \, d\lambda \leq \int_{[0,1]} |g' - f_a| \, d\lambda < a, \quad (4.6)$$

since g' vanishes on U. Next, deduce from (4.5) that

$$\frac{1}{\lambda_{n,j}} \int_{J_{n,j}} f \, d\lambda = \frac{2\ell_n}{\lambda_{n,j}} + \frac{1}{\lambda_{n,j}} \int_{J_{n,j}} (f - g') \, d\lambda$$

$$= \frac{2\ell_n - \int_{J_{n,j}} (g' - f) \, d\lambda}{P_{n,j} - P_{n,j-1} - 2\ell_n} \quad \forall j = 1, \ldots, n,$$

and consequently, for every $j = 1, \ldots, n$,

$$2\,\Omega\left(\frac{1}{\lambda_{n,j}} \int_{J_{n,j}} f \, d\lambda\right) (P_{n,j} - P_{n,j-1}) = 2\ell_n - \frac{\int_{J_{n,j}} (g' - f) \, d\lambda}{1 + \dfrac{1}{\lambda_{n,j}} \int_{J_{n,j}} f \, d\lambda}. \quad (4.7)$$

Summing (4.7) over $j = 1, \ldots, n$ yields

$$\omega_n - \int_{[0,1]} h_n \, d\lambda = \frac{1}{2} \sum_{j=1}^n \frac{\int_{J_{n,j}} (g' - f) \, d\lambda}{1 + \dfrac{1}{\lambda_{n,j}} \int_{J_{n,j}} f \, d\lambda}, \quad (4.8)$$

with the piecewise constant nonnegative function $h_n : [0, 1[\to \mathbb{R}$ given by

$$h_n(x) = \Omega\left(\frac{1}{\lambda_{n,j_n(x)}} \int_{J_{n,j_n(x)}} f \, d\lambda\right) \quad \forall x \in [0, 1[.$$

Recall that $f \geq 0$, and so the right-hand side in (4.8) is bounded, for every $n \in \mathbb{N}$, by $\frac{1}{2} \sum_{j=1}^n \int_{J_{n,j}} |g' - f| \, d\lambda \leq \frac{1}{2} \int_{[0,1]} |g' - f| \, d\lambda < \frac{1}{2}a$. Deduce from (4.3) and the continuity of f that

$$\lim_{n \to \infty} h_n(x) = \Omega(f(x)) \quad \forall x \in G\setminus G^\dagger. \quad (4.9)$$

Similarly, (4.4) and the choice of the functions e_k for $k = 1, \ldots, k_a$ imply that

$$\lim_{n \to \infty} h_n(x) = \Omega \left(\frac{1}{\lambda(I_{k(x)})} \int_{I_{k(x)}} f \, d\lambda \right) \leq \Omega(a) < \frac{a}{2} \quad \forall x \in \bigcup_{k=1}^{k_a} I_k. \quad (4.10)$$

The elementary estimate, valid for all $n \in \mathbb{N}$,

$$\left| \omega_n - \int_{[0,1]} \Omega(g') \, d\lambda \right| \leq \left| \omega_n - \int_{[0,1]} h_n \, d\lambda \right| + \int_{\bigcup_{k=1}^{k_a} I_k} h_n \, d\lambda + \int_{\bigcup_{k > k_a} I_k} h_n \, d\lambda$$

$$+ \int_{[0,1]} |\Omega(f) - \Omega(g')| \, d\lambda + \int_G |h_n - \Omega(f)| \, d\lambda,$$

together with (4.8), Fatou's lemma and (4.10), the choice of k_a and the fact that $0 \leq h_n \leq \frac{1}{2}$, estimate (4.6) as well as the dominated convergence theorem yield

$$\limsup_{n \to \infty} \left| \omega_n - \int_{[0,1]} \Omega(g') \, d\lambda \right| \leq \frac{a}{2} + \frac{a}{2} + \frac{a}{2} + \frac{a}{2} + 0 = 2a.$$

Since $0 < a < 1$ has been arbitrary, $\lim_{n \to \infty} \omega_n = \int_{[0,1]} \Omega(g') \, d\lambda$, i.e. (4.1) holds.

Step II: Assume μ^{-1} is singular.
Given any $0 < a < 1$, let $U_a = \{x \in [0, 1] : \text{dist}(x, G) \geq a\}$. Note that $U_a \subset [a, 1 - a]$ is a compact, possibly empty subset of U, so $U_a \cap G = \varnothing$. Assume for the time being that all atoms of μ^{-1} in $[a, 1 - a]$ are small in that

$$\mu^{-1}(\{x\}) \leq a^2 \quad \forall x \in [a, 1 - a]. \quad (4.11)$$

Recall that $2\ell_n \leq \mu^{-1}(J_{n,j})$ for all $j = 1, \ldots, n$, by Proposition 4.2, and correspondingly

$$\omega_n \leq \sum_{j=1}^{n} \Omega \left(\frac{\mu^{-1}(J_{n,j})}{\lambda_{n,j}} \right) (P_{n,j} - P_{n,j-1}) = \int_{[0,1]} \tilde{h}_n \, d\lambda,$$

with the piecewise constant function $\tilde{h}_n : [0, 1[\to \mathbb{R}$ given by

$$\tilde{h}_n(x) = \Omega \left(\frac{\mu^{-1}(J_{n,j_n(x)})}{\lambda_{n,j_n(x)}} \right) \quad \forall x \in [0, 1[.$$

First, observe that if $x \in U_a$, then (4.4) and (4.11) imply that

$$\limsup_{n \to \infty} \mu^{-1}(J_{n,j_n(x)}) \leq \mu^{-1}(I_{k(x)}) \leq 2a^2,$$

whereas clearly $\lim_{n \to \infty} \lambda_{n,j_n(x)} = \lambda(I_{k(x)}) \geq 2a$. Thus,

$$\limsup_{n \to \infty} \tilde{h}_n(x) \leq \Omega(a) < \frac{a}{2} \quad \forall x \in U_a. \quad (4.12)$$

Next, notice that if $x \in G \backslash G^\dagger$, then $([P_{n, j_n(x)-1}, P_{n, j_n(x)}])$ shrinks to x nicely, and hence

$$\lim_{n \to \infty} \frac{\mu^{-1}([P_{n, j_n(x)-1}, P_{n, j_n(x)}])}{P_{n, j_n(x)} - P_{n, j_n(x)-1}} = 0 \quad \text{for } \lambda\text{-almost every } x \in G,$$

by [33, Thm. 7.13]. Thus, $\lim_{n \to \infty} \ell_n / (P_{n, j_n(x)} - P_{n, j_n(x)-1}) = 0$ for λ-almost every $x \in G$, which in turn shows that $(J_{n, j_n(x)})$ shrinks to x nicely as well. Applying [33, Thm. 7.13] once more yields $\lim_{n \to \infty} \mu^{-1}(J_{n, j_n(x)})/\lambda_{n, j_n(x)} = 0$ for λ-almost every $x \in G$, and thus

$$\lim_{n \to \infty} \widetilde{h}_n(x) = 0 \quad \text{for } \lambda\text{-almost every } x \in G. \tag{4.13}$$

Recalling that $G^\dagger$ is countable, deduce from (4.12) and (4.13) that

$$
\begin{aligned}
\omega^+ &\leq \limsup_{n \to \infty} \int_{[0,1]} \widetilde{h}_n \, d\lambda \\
&= \limsup_{n \to \infty} \left(\int_{U_a} \widetilde{h}_n \, d\lambda + \int_G \widetilde{h}_n \, d\lambda + \int_{[0,1] \backslash (U_a \cup G)} \widetilde{h}_n \, d\lambda \right) \\
&\leq \frac{a}{2} + \frac{1}{2} \big(1 - \lambda(U_a \cup G)\big).
\end{aligned} \tag{4.14}
$$

In summary, (4.14) holds provided that μ satisfies (4.11).

To conclude the argument in the case of μ^{-1} being singular, given $0 < b < 1$, pick $0 < a < b$ so small that $\lambda(U_a \cup G) > 1 - b$. Noting that the set $G_a := \{x \in [a, 1-a] : \mu^{-1}(\{x\}) > a^2\}$ is finite, consider $\widetilde{g} \in \mathcal{F}$ given by

$$\widetilde{g} = g - \sum_{x \in G_a} \mu^{-1}(\{x\}) \mathbf{1}_{[x, +\infty[},$$

as well as the unique $\widetilde{\mu} \in \mathcal{P}$ with $F_{\widetilde{\mu}}^{-1} = \widetilde{g}$. Crucially, (4.11) holds with $\widetilde{\mu}$ instead of μ. Moreover, notice that $\widetilde{G} := G_{\widetilde{\mu}^{-1}} \supset G \backslash G_a$, and clearly $\widetilde{U}_a \supset U_a$, where $\widetilde{U}_a = \{x \in [0,1] : \text{dist}(x, \widetilde{G}) \geq a\}$. Thus, $\widetilde{U}_a \cup \widetilde{G} \supset (U_a \cup G) \backslash G_a$, and (4.14) applied to $\widetilde{\mu}$, with $\widetilde{\ell}_n := d_1(\widetilde{\mu}, \delta_{\bullet}^{\bullet, n})$, yields

$$\limsup_{n \to \infty} n \widetilde{\ell}_n \leq \frac{a}{2} + \frac{1}{2} \big(1 - \lambda(\widetilde{U}_a \cup \widetilde{G})\big) < b.$$

Finally, let $m_a = \#G_a$ and observe that $\ell_{n + m_a} \leq \widetilde{\ell}_n$ for all $n \in \mathbb{N}$, so

$$\omega_n \leq n \widetilde{\ell}_{n - m_a} = (n - m_a) \widetilde{\ell}_{n - m_a} + m_a \widetilde{\ell}_{n - m_a}$$

for all $n > m_a$. Since $\lim_{n \to \infty} \widetilde{\ell}_n = 0$, clearly $\omega^+ \leq \limsup_{n \to \infty} n \widetilde{\ell}_n < b$, and since $0 < b < 1$ has been arbitrary, $\omega^+ = 0$. Thus, (4.1) holds, with vanishing right-hand side, whenever μ^{-1} is singular.

Step III: Let $\mu \in \mathcal{P}$ be arbitrary.

As in the proof of Theorem 3.5, write $g = g_A + g_S$ with $g_A, g_S \in \mathcal{F}$ such that $\lambda_{g_A} = (\mu^{-1})_A$ and $\lambda_{g_S} = (\mu^{-1})_S$. Let $\mu^{\langle A \rangle}$ and $\mu^{\langle S \rangle}$ be the (uniquely determined) probability measures with $(\mu^{\langle A \rangle})^{-1} = (\mu^{-1})_A$ and $(\mu^{\langle S \rangle})^{-1} = (\mu^{-1})_S$, respectively (notice that in general $\mu^{\langle A \rangle} \neq \mu_A$ and $\mu^{\langle S \rangle} \neq \mu_S$). Also, for every $n \in \mathbb{N}$ let $\ell_n^{\langle A \rangle} = d_1(\mu^{\langle A \rangle}, \delta_\bullet^{\bullet, n})$ and $\ell_n^{\langle S \rangle} = d_1(\mu^{\langle S \rangle}, \delta_\bullet^{\bullet, n})$. Given any $m, n \in \mathbb{N}$, pick $p_m^{\langle A \rangle} \in \Pi_m$ and $p_n^{\langle S \rangle} \in \Pi_n$ as in Proposition 4.2. By considering the joint partition of $[0, 1]$ generated by $\left\{ P_{m,i}^{\langle A \rangle} : i = 0, \ldots, m \right\}$ and $\left\{ P_{n,j}^{\langle S \rangle} : j = 0, \ldots, n \right\}$, it is readily seen that $\ell_{m+n} \leq \ell_m^{\langle A \rangle} + \ell_n^{\langle S \rangle}$. For every $0 < a < 1$ and $n \in \mathbb{N}$, therefore

$$\omega_n \leq n\ell_{\lfloor (1-a)n \rfloor + \lfloor an \rfloor} \leq \frac{1 + \lfloor (1-a)n \rfloor}{1-a} \ell_{\lfloor (1-a)n \rfloor}^{\langle A \rangle} + \frac{1 + \lfloor an \rfloor}{a} \ell_{\lfloor an \rfloor}^{\langle S \rangle},$$

and applying Steps I and II to $\mu^{\langle A \rangle}$ and $\mu^{\langle S \rangle}$, respectively, yields

$$\omega^+ \leq \frac{1}{1-a} \int \Omega \left(\frac{d(\mu^{\langle A \rangle})_A^{-1}}{d\lambda} \right) d\lambda = \frac{1}{1-a} \int \Omega(g_A') \, d\lambda.$$

(recall that $\lim_{n \to \infty} \ell_n^{\langle A \rangle} = \lim_{n \to \infty} n\ell_n^{\langle S \rangle} = 0$). Since $0 < a < 1$ has been arbitrary, $\omega^+ \leq \int \Omega(g_A') \, d\lambda$. To obtain a lower bound for ω^-, recall from Proposition 4.2 that

$$g_-(P_{n,j} - \ell_n) - \ell_n \leq g(P_{n,j-1} + \ell_n) + \ell_n \quad \forall j = 1, \ldots, n,$$

and since $g_- = g_A + g_{S-}$,

$$
\begin{aligned}
g_A(P_{n,j} - \ell_n) - \ell_n \leq & \, g_A(P_{n,j-1} + \ell_n) + \ell_n \\
& - \max\left\{ 0, \left(g_{S-}(P_{n,j} - \ell_n) - g_S(P_{n,j-1} + \ell_n) \right) \right\} \\
\leq & \, g_A(P_{n,j-1} + \ell_n) + \ell_n \quad \forall j = 1, \ldots, n,
\end{aligned}
$$

from which it is clear that $\ell_n^{\langle A \rangle} \leq \ell_n$ for every $n \in \mathbb{N}$. Applying Step I to $\mu^{\langle A \rangle}$ yields $\omega_- \geq \liminf_{n \to \infty} n\ell_n^{\langle A \rangle} = \int \Omega(g_A') \, d\lambda$. Hence, $\lim_{n \to \infty} \omega_n = \int \Omega(g_A') \, d\lambda$, and the proof is complete. $\qquad \square$

Along the lines of the above proof, and by considering the absolutely continuous part of μ rather than of μ^{-1}, the following dual version of Theorem 4.1 can be established; the routine details are left to the interested reader.

Proposition 4.3 *Let $\mu \in \mathcal{P}$ and $\epsilon > 0$. Then,*

$$\lim_{n \to \infty} nd_\epsilon(\mu, \delta_\bullet^{\bullet, n}) = \epsilon \int \Omega \left(\frac{1}{\epsilon} \cdot \frac{d\mu_A}{d\lambda} \right) d\lambda.$$

Notice that Theorem 4.1 and Proposition 4.3 together imply the familiar fact [6, Ch. A] that μ^{-1} is singular if and only if μ is singular, and hence yield a direct analogue of Proposition 3.2 in the context of best approximations.

Proposition 4.4 *For every $\mu \in \mathcal{P}$ and $\epsilon > 0$, the following are equivalent:*

(i) $\lim_{n \to \infty} nd_\epsilon(\mu, \delta_\bullet^{\bullet,n}) = 0$;

(ii) $\mu_A = 0$;

(iii) $(\mu^{-1})_A = 0$.

Just as in the case of best uniform approximations, Theorem 4.1 can be refined through further assumptions on μ. For instance, if $g = F_\mu^{-1}$ is C^4 on $]0, 1[$, if both $g, g' \neq 0$ are convex and if $\Omega(\epsilon g')$ has a C^3-extension to $\mathbb{R}$, then mild boundedness assumptions on g and its derivatives (ensuring all relevant integrals are finite) guarantee that, as a refinement of (4.1),

$$nd_\epsilon(\mu, \delta_\bullet^{\bullet,n}) = c_1 + \frac{c_1^2 c_2}{12} n^{-2} + o(n^{-2}) \quad \text{as } n \to \infty, \tag{4.15}$$

where $c_1 = \int \Omega(\epsilon g') \, d\lambda$ and

$$c_2 = \int \frac{2(1 + \epsilon g')(g'')^2 - (2 + \epsilon g')g' g'''}{(1 + \epsilon g')^2 (g')^2} \, d\lambda.$$

When compared to $(nd_\epsilon(\mu, \delta_\bullet^{u_n}))$, therefore, not only does $(nd_\epsilon(\mu, \delta_\bullet^{\bullet,n}))$ converge to a smaller value (unless g' is constant), but also it converges at the rate (n^{-2}) which often is faster than the rate in (3.10). For example, (4.15) applies to exponential as well as Benford distributions, and the reader may want to check that (2.7) and (2.9) both are consistent with it. If μ is a normal distribution with variance $\sigma^2 > 0$, then $\Omega(\epsilon g')$ does not have even a C^1-extension to $\mathbb{R}$, and correspondingly $c_2 = -\infty$, which suggests that $(nd_\epsilon(\mu, \delta_\bullet^{\bullet,n}))$ converges at a slower rate. This indeed is the case: an elementary albeit lengthy analysis yields

$$nd_\epsilon(\mu, \delta_\bullet^{\bullet,n}) = -\epsilon \sqrt{\frac{\pi \sigma^2}{2}} \, \text{Li}_{1/2}\left(-\frac{1}{\epsilon \sqrt{2\pi \sigma^2}}\right) + \mathcal{O}\left(\frac{\log n}{n^2}\right) \quad \text{as } n \to \infty,$$

where $\text{Li}_{1/2}$ denotes the polylogarithm of order $\frac{1}{2}$; see, e.g. [26, §25.12]. Though slower than (4.15), this rate of convergence again is considerably faster than its counterpart (3.9) for best uniform approximations. It should be noted, however, that such a hierarchy of rates, though observed for many familiar distributions, is by no means universal: as mentioned already in Introduction, for the 1-Pareto distribution both sequences $(nd_\epsilon(\mu, \delta_\bullet^{u_n}))$ and $(nd_\epsilon(\mu, \delta_\bullet^{\bullet,n}))$ converge to their respective limits $\frac{1}{2}$ and $\frac{\pi}{8}$ at the same rate (n^{-2}), as is evident from (1.7) and (1.8). For the $\frac{1}{2}$-Pareto distribution, i.e. $F_\mu(x) = 1 - x^{-1/2}$ for all $x \geq 1$, (3.10) yields

$$nd_1(\mu, \delta_\bullet^{u_n}) = 0.5000 - 0.03125 \, n^{-3} + o(n^{-3}) \quad \text{as } n \to \infty,$$

whereas (4.15), with $c_1 = \int_0^1 (2 + t^3)^{-1} dt = 0.4508$ and $c_2 = 6 \int_0^1 t(2 - t^3)(2 + t^3)^{-2} dt = 0.9102$, reads

$$nd_1(\mu, \delta_\bullet^{\bullet,n}) = 0.4508 + 0.01541\, n^{-2} + o(n^{-2}) \quad \text{as } n \to \infty.$$

Here, $\left(nd_1(\mu, \delta_\bullet^{u_n})\right)$ converges at an even faster rate than $\left(nd_1(\mu, \delta_\bullet^{\bullet,n})\right)$.

Example 6 For the Cantor distribution μ and its inverse μ^{-1} in Example 5, Theorem 4.1 yields $\lim_{n \to \infty} nd_\epsilon(\mu, \delta_\bullet^{\bullet,n}) = \lim_{n \to \infty} nd_\epsilon(\mu^{-1}, \delta_\bullet^{\bullet,n}) = 0$. An elementary analysis shows that $\left(n^{1/c} d_\epsilon(\mu, \delta_\bullet^{\bullet,n})\right)$ and $\left(n^{1/c} d_\epsilon(\mu^{-1}, \delta_\bullet^{\bullet,n})\right)$ both are divergent, yet bounded above and below by positive constants, where $c = \log 2 / \log 3 < 1$ is the Hausdorff dimension of both the set $C = G_\mu$ and the measure μ. It seems plausible that Theorem 4.1 can similarly be refined for a wide class of self-similar (singular) distributions, thus complementing known d_W-quantization results [18,19,25,29].

To establish one further interesting property of best d_ϵ-approximations, recall from Proposition 2.5 that if $d_\epsilon(\mu, \delta_{x_n}^{P_n}) = \ell_\bullet^{\bullet,n}$, then p_n can easily be determined from x_n (or vice versa). Thus, x_n (or p_n) alone already captures $\delta_{x_n}^{P_n}$ to a large extent, and it is natural to ask, for instance, whether $x_{n,1}, \ldots, x_{n,n}$, i.e. the locations of best d_ϵ-approximations of $\mu \in \mathcal{P}$ conform to an *asymptotic point distribution* as $n \to \infty$, referred to as the *point density measure* of μ in [19]. In the context of best d_W-approximations (or -quantizations), and under the appropriate assumptions, this question has a positive answer; see, e.g. the "empirical measure theorem" [18, Thm. 7.5] and variants thereof [19]. As is the case with Theorem 4.1 and Proposition 4.3, the result for best d_ϵ-approximations again is simpler than its d_W-counterpart in that the asymptotic point distribution is readily identified whenever $\mu \in \mathcal{P}$ is non-singular, and no further assumptions on μ are needed. In fact, it even is possible to allow for slightly more general x_n. To concisely state the result, for every $\mu \in \mathcal{P}$ with $\mu_A \neq 0$, define $\mu_\epsilon^* \in \mathcal{P}$ via

$$\frac{d\mu_\epsilon^*}{d\lambda} = \frac{\Omega\left(\frac{1}{\epsilon} \cdot \frac{d\mu_A}{d\lambda}\right)}{\int_{\mathbb{R}} \Omega\left(\frac{1}{\epsilon} \cdot \frac{d\mu_A}{d\lambda}\right) d\lambda} \quad \forall \epsilon > 0.$$

Clearly, μ_ϵ^* is absolutely continuous, and $\mu_\epsilon^* = \mu$ for some (in fact, all) $\epsilon > 0$ if and only if μ is uniform, i.e. $\mu = \lambda(\cdot \cap B)/\lambda(B)$ for some Borel set B with $\lambda(B) \in \mathbb{R}^+$. Also, given any $\mu \in \mathcal{P} \backslash \mathcal{P}_\infty^*$, i.e. $\# \operatorname{supp} \mu = \infty$, call a sequence (x_n), with $x_n \in \Xi_n$ for every $n \in \mathbb{N}$, *asymptotically d_ϵ-minimal for μ* if

$$\lim_{n \to \infty} \frac{d_\epsilon(\mu, \delta_{x_n}^\bullet)}{d_\epsilon(\mu, \delta_\bullet^{\bullet,n})} = 1.$$

Thus, for instance, (x_n) is asymptotically d_ϵ-minimal for $\mu \in \mathcal{P} \backslash \mathcal{P}_\infty^*$ whenever $\delta_{x_n}^{P_n}$, with $x_n \in \Xi_n$, $p_n \in \Pi_n$, is a best d_ϵ-approximation of μ for every $n \in \mathbb{N}$.

Theorem 4.5 *Let $\mu \in \mathcal{P}$ and $\epsilon > 0$. If $\mu_A \neq 0$ and (x_n) is asymptotically d_ϵ-minimal for μ, then*

$$\lim_{n \to \infty} \frac{\#\{1 \leq j \leq n : x_{n,j} \in I\}}{n} = \mu_\epsilon^*(I) \quad \forall I \subset \mathbb{R}, \; I \text{ an interval.} \qquad (4.16)$$

Proof For convenience, write $f = F_\mu = f_A + f_S$ with $f_A, f_S \in \mathcal{F}$ such that $\lambda_{f_A} = \mu_A$ and $\lambda_{f_S} = \mu_S$ (the functions f_A, f_S can be made unique, for instance, by requiring that $f_A(-\infty) = f_S(-\infty) = 0$). Also, let $G = G_\mu$, $\ell_n = d_\epsilon(\mu, \delta_{x_n}^\bullet)$ for all $n \in \mathbb{N}$, and define $G^\dagger$ as in (4.2). Once again, it suffices to consider the case of $\epsilon = 1$. Note that $\mu_A \neq 0$ implies $\ell_n > 0$ for every n, and $\lim_{n \to \infty} n\ell_n = \int_{\mathbb{R}} \Omega(d\mu_A/d\lambda) > 0$, by Proposition 4.3 and the assumed asymptotic d_ϵ-minimality of (x_n).

Fix a non-empty interval $I =]y, z]$ with $y, z \in \mathbb{R}$. Perturbing x_n slightly if necessary, without altering $\#\{1 \leq j \leq n : x_{n,j} \in I\}$ or increasing $d_1(\mu, \delta_{x_n}^\bullet)$ for any n, it may be assumed that $x_{n,j} < x_{n,j+1}$ for all $n \in \mathbb{N}$ and $j = 0, \ldots, n$. Thus, for every $x \in \mathbb{R}$, there exists a unique $j_n(x) \in \{0, \ldots, n\}$ with $x_{n,j_n(x)} \leq x < x_{n,j_n(x)+1}$. By Proposition 2.5,

$$f_A(x_{n,j+1} - \ell_n) + f_{S-}(x_{n,j+1} - \ell_n) - \ell_n \leq f_A(x_{n,j} + \ell_n) + f_S(x_{n,j} + \ell_n) + \ell_n,$$

for every $j = 0, \ldots, n$, and consequently also

$$f_A(x_{n,j+1} - \ell_n) - f_A(x_{n,j} + \ell_n) \leq 2\ell_n \quad \forall j = 0, \ldots, n. \qquad (4.17)$$

Fix any $a > 0$, and recalling that f_A is differentiable λ-almost everywhere with $f_A' \geq 0$ and $0 < \int_{\mathbb{R}} f_A' d\lambda = \mu_A(\mathbb{R}) \leq 1$, pick a continuous function $g : \mathbb{R} \to \mathbb{R}^+$ with $\int_{\mathbb{R}} |f_A' - g| d\lambda < a$.

Let $K_n = \{0 \leq j \leq n : x_{n,j} + \ell_n < x_{n,j+1} - \ell_n\}$ which may not be all of the set $\{0, \ldots, n\}$ but does contain $0, n$ in any case. On the one hand, if $j \in K_n \setminus \{0, n\}$ let $J_{n,j} = [x_{n,j} + \ell_n, x_{n,j+1} - \ell_n]$ and $\lambda_{n,j} = \lambda(J_{n,j}) = x_{n,j+1} - x_{n,j} - 2\ell_n > 0$, and deduce from (4.17) that

$$\frac{1}{\lambda_{n,j}} \int_{J_{n,j}} g \, d\lambda \leq \frac{2\ell_n - \int_{J_{n,j}} (f_A' - g) \, d\lambda}{x_{n,j+1} - x_{n,j} - 2\ell_n},$$

and consequently

$$\ell_n \geq \Omega \left(\frac{1}{\lambda_{n,j}} \int_{J_{n,j}} g \, d\lambda \right) (x_{n,j+1} - x_{n,j}) - \frac{1}{2} \int_{J_{n,j}} |f_A' - g| \, d\lambda; \qquad (4.18)$$

with the usual convention $0 \cdot (\pm \infty) = 0$, (4.18) is correct also for $j = 0, n$. On the other hand, if $j \notin K_n$ then clearly $\ell_n \geq \frac{1}{2}(x_{n,j+1} - x_{n,j})$. With (4.18) as well as the definitions of $j_n(x)$ and K_n, therefore,

$$\left(1 + j_n(z) - j_n(y)\right)\ell_n$$

$$\geq \sum_{j \in \{j_n(y),\ldots,j_n(z)\} \cap K_n} \Omega\left(\frac{1}{\lambda_{n,j}} \int_{J_{n,j}} g \, d\lambda\right)(x_{n,j+1} - x_{n,j})$$

$$- \frac{1}{2} \sum_{j \in \{j_n(y),\ldots,j_n(z)\} \cap K_n} \int_{J_{n,j}} |f_A' - g| \, d\lambda$$

$$+ \sum_{j \in \{j_n(y),\ldots,j_n(z)\} \setminus K_n} \frac{1}{2}(x_{n,j+1} - x_{n,j})$$

$$\geq \int_{[x_{n,j_n(y)}, x_{n,j_n(z)+1}]} h_n \, d\lambda - \frac{1}{2} \int_{[x_{n,j_n(y)}, x_{n,j_n(z)+1}]} |f_A' - g| \, d\lambda$$

$$\geq \int_I h_n \, d\lambda - \frac{a}{2}, \tag{4.19}$$

where the piecewise constant function $h_n : \mathbb{R} \to \mathbb{R}^+$ is given by

$$h_n(x) = \begin{cases} \Omega\left(\dfrac{1}{\lambda_{n,j_n(x)}} \displaystyle\int_{J_{n,j_n(x)}} g \, d\lambda\right) & \text{if } j_n(x) \in K_n, \\ \dfrac{1}{2} & \text{if } j_n(x) \notin K_n. \end{cases}$$

If $j_n(x) \notin K_n$ for all sufficiently large n, then $\lim_{n\to\infty} h_n(x) = \frac{1}{2}$, whereas if $x \in G \setminus G^\dagger$ and $j_n(x) \in K_n$ for infinitely many n, then $\liminf_{n\to\infty} h_n(x) \geq \Omega(g(x))$ because similarly to (4.3),

$$\lim_{n\to\infty}[x_{n,j_n(x)}, x_{n,j_n(x)+1}] = \{x\} \quad \forall x \in G \setminus G^\dagger.$$

In summary, therefore,

$$\liminf_{n\to\infty} h_n(x) \geq \Omega\big(g(x)\mathbf{1}_G(x)\big) \quad \text{for } \lambda\text{-almost every } x \in \mathbb{R}. \tag{4.20}$$

Note that $j_n(z) - j_n(y) = \#\{1 \leq j \leq n : x_{n,j} \in I\}$. Consequently, (4.19), Proposition 4.3 with $\int_{\mathbb{R}} \Omega(f_A') \, d\lambda > 0$, and Fatou's lemma applied to (4.20), together yield

$$\liminf_{n\to\infty} \frac{\#\{1 \leq j \leq n : x_{n,j} \in I\}}{n} \geq \frac{\displaystyle\int_I \Omega(g\mathbf{1}_G) \, d\lambda - \frac{a}{2}}{\displaystyle\int_{\mathbb{R}} \Omega(f_A') \, d\lambda}.$$

Recall that $f_A' = 0$ on $\mathbb{R} \setminus G$; hence,

$$\int_I \Omega(g\mathbf{1}_G)\,d\lambda = \int_I \Omega(f'_A)\,d\lambda + \int_{I\cap G}\big(\Omega(g\mathbf{1}_G) - \Omega(f'_A)\big)\,d\lambda$$

$$\geq \int_I \Omega(f'_A)\,d\lambda - \frac{1}{2}\int_I |f'_A - g|\,d\lambda \geq \int_I \Omega(f'_A)\,d\lambda - \frac{a}{2},$$

and consequently

$$\liminf_{n\to\infty}\frac{\#\{1 \leq j \leq n : x_{n,j} \in I\}}{n} \geq \mu_1^*(I) - \frac{a}{\displaystyle\int_{\mathbb{R}}\Omega(f'_A)\,d\lambda}.$$

Since the number $a > 0$ and the interval $I \subset \mathbb{R}$ have been arbitrary, and since $\mu_1^*(\mathbb{R}) = 1$,

$$\lim_{n\to\infty}\frac{\#\{1 \leq j \leq n : x_{n,j} \in I\}}{n} = \mu_1^*(I) \quad \forall I \subset \mathbb{R},\ I \text{ an interval,}$$

i.e. (4.16) holds as claimed. □

Note that Theorem 4.5 in particular asserts that if $\mu \in \mathcal{P}$ is non-singular and $(\delta_{x_n}^{p_n})$, with $x_n \in \Xi_n$ and $p_n \in \Pi_n$ for every $n \in \mathbb{N}$, is *any* sequence of best d_ϵ-approximations of μ, then the sequence $(\delta_{x_n}^{u_n})$, obtained by "forgetting" the optimal weights and instead assigning equal weight $1/n$ to each atom, converges weakly to μ_ϵ^*. It seems rather remarkable that $(\delta_{x_n}^{u_n})$ always converges, and to a limit that is independent of (x_n). By contrast, simple examples show that $(\delta_{x_n}^{u_n})$ may diverge if μ is singular; cf. [19].

Example 7 Let $\mu = \exp(a)$ with $a > 0$. With $\ell_\bullet^{\bullet,n}$ and the (unique) best d_ϵ-approximation $\delta_{x_n}^{p_n}$ of μ found in Example 1, it is readily confirmed that for any $n \in \mathbb{N}$ and $x \in \mathbb{R}^+$ the number $\#\{1 \leq j \leq n : x_{n,j} \leq x\}$ equals the largest integer not larger than

$$n - \frac{\epsilon}{2a\ell_\bullet^{\bullet,n}}\log\left(1 + \left(\frac{e^{-a(x+\ell_\bullet^{\bullet,n}/\epsilon)}}{\ell_\bullet^{\bullet,n}} - 1\right)\tanh\frac{a\ell_\bullet^{\bullet,n}}{\epsilon}\right).$$

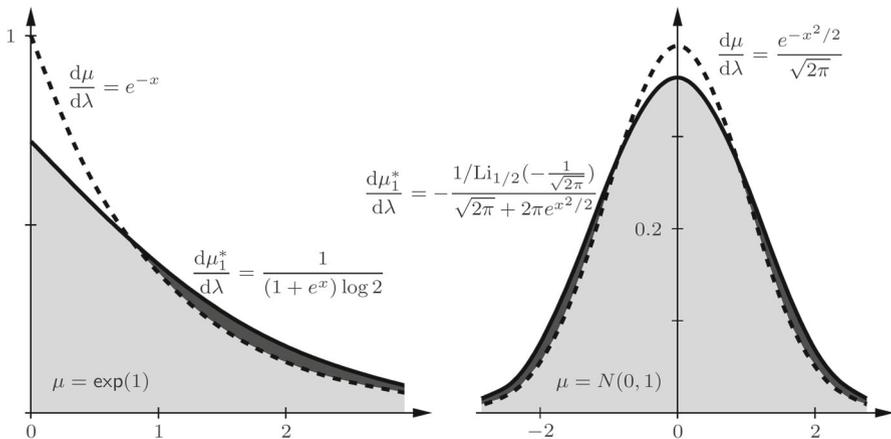From this, a straightforward calculation utilizing (2.7) yields

$$\lim_{n\to\infty}\frac{\#\{1 \leq j \leq n : x_{n,j} \leq x\}}{n} = 1 - \frac{\log(1 + ae^{-ax}/\epsilon)}{\log(1 + a/\epsilon)} \quad \forall x \in \mathbb{R}^+.$$

Thus, the asymptotic point density of (x_n) is

$$\frac{a^2}{\log(1 + a/\epsilon)} \cdot \frac{1}{a + \epsilon e^{ax}} = \frac{\Omega(ae^{-ax}/\epsilon)}{\int_{\mathbb{R}}\Omega(ae^{-ay}/\epsilon)\,dy} = \frac{d\mu_\epsilon^*}{d\lambda}(x) \quad \forall x \in \mathbb{R}^+,$$

in perfect agreement with Theorem 4.5. Note that unlike for best d_W-approximations [18, Thm. 7.5], this asymptotic point distribution is not exponential; see also Fig. 3.

Fig. 3 Comparing the standard exponential (left) and standard normal densities (broken curves) to the asymptotic point densities of their respective best d_1-approximations (solid curves); see Example 7

For another simple example, let μ be a normal distribution with mean 0 and variance $\sigma^2 > 0$. While no explicit formula is available for the (unique) best d_ϵ-approximation $\delta_{x_n}^{p_n}$ of μ in general, Theorem 4.5 yields

$$\frac{d\mu_\epsilon^*}{d\lambda}(x) = -\frac{1}{\text{Li}_{1/2}\left(-\dfrac{1}{\epsilon\sqrt{2\pi\sigma^2}}\right)} \cdot \frac{1}{\sqrt{2\pi\sigma^2} + 2\pi\epsilon\sigma^2 e^{x^2/(2\sigma^2)}} \quad \forall x \in \mathbb{R},$$

as the asymptotic point density of (x_n). Again, this asymptotic point distribution is not normal, unlike its d_W-counterpart.

References

1. Abramowitz, M., Stegun, I.A.: Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, National Bureau of Standards Applied Mathematics Series, vol. 55. Dover Publications, New York (1964)
2. Berger, A., Hill, T.P.: An Introduction to Benford's Law. Princeton University Press, Princeton (2015)
3. Berger, A., Hill, T.P., Morrison, K.E.: Scale-distortion inequalities for mantissas of finite data sets. J. Theor. Probab. **21**, 97–117 (2008)
4. Berger, A., Xu, C.: Best finite approximations of Benford's Law. J. Theor. Probab. (2018). https://doi.org/10.1007/s10959-018-0827-z
5. Bloch, I., Atif, J.: Hausdorff distances between distributions using optimal transport and mathematical morphology. In: Mathematical Morphology and its Applications to Signal and Image Processing, Lecture Notes in Computer Science, vol. 9082, Springer (2015)
6. Bobkov, S.G., Ledoux, M.: One-dimensional empirical measures, order statistics and Kantorovich transport distances. Mem. Am. Math. Soc. (2016). http://www-users.math.umn.edu/~bobko001/preprints/2016_BL_Order.statistics_Revised.version.pdf. Accessed 1 March 2019

7. Chevallier, J.: Uniform decomposition of probability measures: quantization, classification, clustering and rate of convergence. J. Appl. Prob. **55**, 1037–1045 (2018)
8. Chevallier, J., Duarte, A., Löcherbach, E., Ost, G.: Mean field limits for nonlinear spatially extended Hawkes processes with exponential memory kernels. Stoch. Proc. Appl. (2018). https://doi.org/10.1016/j.spa.2018.02.007
9. Dereich, S., Scheutzow, M., Schottstedt, R.: Constructive quantization: approximation by empirical measures. Ann. Inst. Henri Poincaré Probab. Stat. **49**, 1183–1203 (2013)
10. Dereich, S., Vormoor, C.: The high resolution vector quantization problem with Orlicz norm distortion. J. Theor. Probab. **24**, 517–544 (2011)
11. Dudley, R.: Real Analysis and Probability. Cambridge University Press, Cambridge (2002)
12. Dvoretzky, V., Erdős, P.: Some problems on random walk in space. In: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, pp. 353–367 (1950)
13. Fort, J.-C., Pagès, G.: Asymptotics of optimal quantizers for some scalar distributions. J. Comput. Appl. Math. **146**, 253–275 (2002)
14. Fournier, N., Guillin, A.: On the rate of convergence in Wasserstein distance of the empirical measure. Probab. Theory Related Fields **162**, 707–738 (2015)
15. Gibbs, A.L., Su, F.E.: On choosing and bounding probability metrics. Int. Stat. Rev. **70**, 419–435 (2002)
16. Giles, M.B., Hefter, M., Mayer, L., Ritter, K.: Random bit quadrature and approximation of distributions on Hilbert spaces. Found. Comput. Math. **19**, 205–238 (2019)
17. Giles, M.B., Hefter, M., Mayer, L., Ritter, K.: Random bit multilevel algorithms for stochastic differential equations. J. Complex. (2019). https://doi.org/10.1016/j.jco.2019.01.002
18. Graf, S., Luschgy, H.: Foundations of Quantization for Probability Distributions. Lecture Notes in Mathematics, vol. 1730. Springer, Berlin (2000)
19. Graf, S., Luschgy, H.: The point density measure in the quantization of self-similar probabilities. Math. Proc. Camb. Philos. Soc. **138**, 513–531 (2005)
20. Graf, S., Luschgy, H.: Quantization for probability measures in the Prokhorov metric. Theory Probab. Appl. **53**, 216–241 (2009)
21. Graf, S., Luschgy, H., Pagès, G.: Optimal quantizers for Radon random vectors in a Banach space. J. Approx. Theory **144**, 27–53 (2007)
22. Graf, S., Luschgy, H., Pagès, G.: The local quantization behavior of absolutely continuous probabilities. Ann. Probab. **40**, 1795–1828 (2012)
23. Gray, R.M., Neuhoff, D.L.: Quantization. IEEE Trans. Inf. Theory **44**, 2325–2383 (1998)
24. Gruber, P.M.: Optimum quantization and its applications. Adv. Math. **186**, 456–497 (2004)
25. Kreitmeier, W.: Optimal quantization for uniform distributions on Cantor-like sets. Acta Appl. Math. **105**, 339–372 (2009)
26. Olver, F.W., Lozier, D.W., Boisvert, R.F., Clark, C.W.: NIST handbook of mathematical functions, U.S. Department of Commerce, National Institute of Standards and Technology. Cambridge University Press, Cambridge (2010)
27. Pagès, G., Pham, H., Printems, J.: Optimal quantization methods and applications to numerical problems in finance. In: Rachev, S.T. (ed.) Handbook of Computational and Numerical Methods in Finance, pp. 253–297. Birkhäuser, Boston (2004)
28. Pflug, G.C., Pichler, A.: Approximations for Probability Distributions and Stochastic Optimization Problems. International Series in Operations Research and Management Science, vol. 163, pp. 343–387. Springer, Berlin (2011)
29. Pötzelberger, K.: The quantization error of self-similar distributions. Math. Proc. Camb. Philos. Soc. **137**, 725–740 (2004)
30. Rachev, S.T.: Probability Metrics and the Stability of Stochastic Models. Wiley, Hoboken (1991)
31. Rachev, S.T., Klebanov, L.B., Stoyanov, S.V., Fabozzi, F.J.: A structural classification of probability distances. In: Rachev, S.T., Klebanov, L.B., Stoyanov, S.V., Fabozzi, F.J. (eds.) The Methods of Distances in the Theory of Probability and Statistics. Springer, New York (2013)
32. Rachev, S.T., Rüschendorf, L.: Mass Transportation Problems: Applications, vol. II. Springer, Berlin (1998)
33. Rudin, W.: Real and Complex Analysis. McGraw Hill, New York (1973)
34. Thompson, J.W.: A note on the Lévy distance. J. Appl. Prob. **12**, 412–414 (1975)
35. Villani, C.: Optimal Transport: Old And New, Grundlehren der Mathematischen Wissenschaften, vol. 338. Springer, Berlin (2009)

36. Xu, C.: Approximation and invariance properties of one-dimensional probabilities. Ph.D. thesis, University of Alberta (2018)
37. Xu, C., Berger, A.: Best finite constrained approximations of one-dimensional probabilities. J. Approx. Theory (2019, to appear)
38. Zador, P.L.: Asymptotic quantization error of continuous signals and the quantization dimension. IEEE Trans. Inf. Theory **28**, 139–149 (1982)