

FAMILIES OF WELL APPROXIMABLE MEASURES

SAMANTHA FAIRCHILD¹ — MAX GOERING¹ — CHRISTIAN WEISS²

¹University of Washington, Department of Mathematics, Seattle, Washington, U.S.A

²Ruhr West University of Applied Sciences, Department of Natural Sciences,
Mülheim an der Ruhr, GERMANY

ABSTRACT. We provide an algorithm to approximate a finitely supported discrete measure μ by a measure ν_N corresponding to a set of N points so that the total variation between μ and ν_N has an upper bound. As a consequence if μ is a (finite or infinitely supported) discrete probability measure on $[0, 1]^d$ with a sufficient decay rate on the weights of each point, then μ can be approximated by ν_N with total variation, and hence star-discrepancy, bounded above by $(\log N)N^{-1}$. Our result improves, in the discrete case, recent work by Aistleitner, Bilyk, and Nikolov who show that for any normalized Borel measure μ , there exist finite sets whose star-discrepancy with respect to μ is at most $(\log N)^{d-\frac{1}{2}}N^{-1}$. Moreover, we close a gap in the literature for discrepancy in the case $d = 1$ showing both that Lebesgue is indeed the hardest measure to approximate by finite sets and also that all measures without discrete components have the same order of discrepancy as the Lebesgue measure.

Communicated by Friedrich Pillichshammer.

1. Introduction

In [1], the authors ask whether the Lebesgue measure is the hardest measure to approximate by finite sets. They guess that the answer is *yes* and justify the conjecture because the Lebesgue measure is spread throughout the entire cube $[0, 1]^d$ and treats all points the same. In this paper, the main notion of approximation used is the total variation.

© 2021 BOKU-University of Natural Resources and Life Sciences and Mathematical Institute, Slovak Academy of Sciences.

2010 Mathematics Subject Classification: 11K36, 11K38, 11J71, 11K06.

Keywords: star-discrepancy, low-discrepancy sequences, Borel Measures, total variation.

Licensed under the Creative Commons Attribution-NC-ND 4.0 International Public License.

1.1. Total variation

For two probability measures μ, ν the total variation is defined by

$$\rho(\mu; \nu) := \sup_{A \in \mathcal{B}} |\mu(A) - \nu(A)|, \quad (1.1)$$

where $\mathcal{B}$ is the class of all Borel sets. The total variation generates the strong topology on the set of all probability measures on $[0, 1]^d$. Moreover, it is widely used in statistics and has direct connections to other important (statistical) notions such as the Hellinger distance, the Kullback-Leibler divergence, the star-discrepancy and the Wasserstein metric, see [8, 12, 14, 17]. Exact calculations of the total variation of two arbitrary probability measures can be difficult if not impossible but for many applications approximations suffice. In this paper, we are interested in giving bounds for the total variation between discrete measures and probability measures associated to a finite set $(x_i)_{i=1}^N$. By a probability measure associated to a finite set we mean a measure given by

$$\nu_N = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}, \quad (1.2)$$

where δ_{x_i} denotes the Dirac measure, i.e.,

$$\delta_{x_i}(A) := \begin{cases} 1 & \text{if } x_i \in A, \\ 0 & \text{else.} \end{cases}$$

We show that for large classes of discrete probability measures there exist finite sets $(x_i)_{i=1}^N$ so that the total variation is at most $\frac{\log(N)}{N}$. More precisely, one can combine Theorem 2.3 and Example 2.5 to form the following weaker, but easier to understand theorem.

THEOREM 1.1. *For every $d \geq 1$ and every $0 < r < 1$, there exists a constant c_r independent of the dimension such that the following holds. For every $N \geq 2$ and every probability measure μ on $[0, 1]^d$ of the form $\mu = \sum_{i=1}^{\infty} \alpha_i \delta_{y_i}$ with $\alpha_i \leq r^{i-1} \alpha_1$ for all $i \in \mathbb{N}$, there exists a finite set $x_1, \dots, x_N \in [0, 1]^d$ such that ν_N as in (1.2) satisfies*

$$\rho(\mu; \nu_N) \leq c_r \frac{\log(N)}{N}.$$

The constant c_r is defined explicitly in Example 2.5 and is monotonically increasing in r .

The proof of Theorem 2.3 which produces Theorem 1.1 is composed of two main parts. First, we constructively approximate discrete measures with a finite number of points, and second we use the fact that most points in a discrete

probability measure have negligible total mass allowing the finite approximation to suffice.

REMARK 1.2. The condition $\alpha_{i+1} \leq r^i \alpha_1$ for $r < 1$ should be interpreted as enforcing that the discrete measure is not too close to a uniform distribution over points. Moreover, the precise formula for c_r shows that, at least using this method, measures which are closer to being uniformly distributed are harder to approximate. This matches the heuristic argument mentioned above that the Lebesgue measure should be the hardest measure to approximate, since a measure with equal weights centered at a large number of points almost uniformly covering $[0, 1]^d$, e.g., a lattice is *close to* Lebesgue. Example 2.7 produces a family of measures with faster decay on the weights (hence the measures are further from being uniform) that have an even better approximation by finite sets.

1.2. Comparison to other metrics

If d_A denotes some other (pseudo-)metric on the space of probability measures, often inequalities of the form

$$d_A(\mu; \nu) \leq h(\rho(\mu; \nu)) \tag{1.3}$$

can be proven, where h is an explicitly known function, see [7]. For example, this is true for the star-discrepancy D_N^* and the p -Wasserstein metric W_p . Thus, Theorem 1.1 does not only yield bounds for the total variation distance between the two measures but for any other (pseudo-)metric which satisfies an inequality of the form (1.3). We will focus only on the total variation and star-discrepancy for the remainder of the paper.

Instead of taking the supremum over all Borel sets only boxes anchored at zero are considered for the star-discrepancy. More precisely, the star-discrepancy between two probability measures μ, ν is defined by

$$D_N^*(\mu; \nu) := \sup_{A \in \mathcal{A}} |\mu(A) - \nu(A)|,$$

where $\mathcal{A}$ is the set of all half-open axis-parallel boxes contained in $[0, 1]^d$ which have one vertex at the origin. Hence, the trivial inequality

$$D_N^*(\mu; \nu) \leq \rho(\mu; \nu)$$

holds. Therefore, Theorem 1.1 is also true after replacing the total variation by the star-discrepancy. Moreover, in Remark 2.4 we note how the constant c_r in Theorem 1.1 can be improved for the star-discrepancy in comparison to the total variation.

1.3. Star-discrepancy

The order of convergence in Theorem 1.1 should be compared to the best known result for general measures.

THEOREM 1.3 (Aistleitner, Bilyk, Nikolov, [1]). *For every $d \geq 1$, there exists a constant c_d (depending only on d) such that the following holds. For every $N \geq 2$ and every normalized Borel measure μ on $[0, 1]^d$ there exist points $x_1, \dots, x_N \in [0, 1]^d$ such that ν_N as in (1.2) satisfies*

$$D_N^*(\mu; \nu_N) \leq c_d \frac{(\log N)^{d-\frac{1}{2}}}{N}.$$

Moreover, if $(x_n)_{n=1}^\infty$ is a sequence, then the discrepancy bound becomes

$$\leq c_d N^{-1} (\log N)^{d+\frac{1}{2}} \quad \text{for all } N \geq 2.$$

Theorem 1.3 improved previous results from [2] and [4] where (for finite sets) the exponent of $\log(N)$ was $(3d+1)/2$ and $2d$. For the d -dimensional Lebesgue measure, λ_d , the exponent of $\log(N)$ has an upper bound of $d-1$, and in fact the optimal order of approximation of the Lebesgue measure by a finite atomic measure is conjectured to be $N^{-1}(\log N)^{d-1}$. In other words, it is conjectured that there exists a constant c_d dependent only on the dimension such that for every finite atomic measure ν_N centered at N points, $x_1, \dots, x_N$, the inequality

$$D_N^*(\lambda_d; \nu_N) \geq c_d \frac{(\log N)^{d-1}}{N}$$

holds. In fact, this is standard knowledge for dimensions one, see [14]. In dimension two, this is known by the work of Schmidt, [16].

As a contrast, with upper bound independent of dimension, the measures in Theorem 1.1 can have strictly better rates of convergence in their discrepancy than the conjectured rate for Lebesgue measure when $d \geq 3$, see Examples 2.5 and 2.7.

Sets that achieve the conjectured optimal order of approximation are called low-discrepancy point sets. There is a duality between the discrepancy of point sets and sequences going back to a fundamental result of Roth, [15]: if for every finite point set $x_1, \dots, x_N$ in dimension d the inequality

$$D_N^*(\lambda_d; \nu_N) \geq c_d \frac{(\log N)^{d-1}}{N}$$

holds, then for every sequence in dimension $d-1$ we have

$$D_N^*(\lambda_d; \nu_N) \geq c'_{d-1} \frac{(\log N)^{d-1}}{N} \quad \text{for infinitely many } N,$$

compare also [14, Lemma 3.7].

The opposite statement is true as well: a lower bound of the form

$$D_N^*(\lambda_d; \nu_N) \geq c'_{d-1} \frac{(\log N)^{d-1}}{N}$$

for the discrepancy of every sequence in dimension $d - 1$ and infinitely many N , implies a corresponding inequality for finite point sets in dimension d . Therefore, sequences with the conjectured optimal order of convergence $N^{-1}(\log N)^d$ are called low-discrepancy sequences (for the Lebesgue measure). There are essentially three classical families of low-discrepancy sequences for the one-dimensional Lebesgue measure: Kronecker sequences, digital sequences, and Halton sequences. In dimension one, further classes of examples have recently been found, e.g. [5, 18]. A discussion of the multi-dimensional picture can be found in [14].

For arbitrary normalized Borel measures μ , Theorem 1.3 yields the best known order of approximation. If the measure μ has a non-vanishing continuous component then the method of Roth [15], over the orthogonal functions can be applied in order to find lower bounds for the star-discrepancy. This has been conducted in [6]. The lower bounds which are obtained in this way are the same as those for the Lebesgue measure. This can be regarded as another indication that the Lebesgue measure is particularly hard to approximate by finite sets, and that the results of Theorem 1.3 should be able to be improved to further aid the conjecture that Lebesgue measure is indeed the most difficult measure to approximate by finite sets.

1.4. Dimension 1

Though more is not known for arbitrary Borel measures in higher-dimensions, the bounds in Theorem 1.3 are not optimal when $d = 1$. We show this in Section 3, where we close a gap in the literature when $d = 1$: we demonstrate the Lebesgue measure is the worst measure to approximate by finite sets or sequences, and measures without discrete components have the same optimal order of convergence as the Lebesgue measure.

The core reason for such robust results only with dimension $d = 1$ is that the ordering of $[0, 1]$ is equivalent to the ordering (by inclusion) of axis-parallel boxes in $[0, 1]$ which contain the origin. Hence this case can be treated by generalizing the arguments given in [9, 10].¹

¹Hlawka and Mück concentrated on deriving inequalities of Koksma-Hlawka type in their papers and hence did not work out the details regarding approximation of measures. Moreover, they made a Lipschitz continuity assumption.

REMARK 1.4. Recall that in dimension $d = 1$, Lebesgue's decomposition theorem states any Borel measure μ can be written as

$$\mu = \mu_{ac} + \mu_d + \mu_{cs},$$

where μ_{ac} is absolutely continuous with respect to the Lebesgue measure, that is μ_{ac} is zero on sets of Lebesgue measure zero, μ_d is a discrete measure, that is, it is zero on the complement of some countable set, and μ_{cs} is continuous singular, that is, μ_{cs} is zero on the complement of some set B of Lebesgue measure zero but assigns no weight to any countable set of points. For more details we refer the reader, e.g., to [11, Chapter V],

THEOREM 1.5. *Fix μ a normalized Borel measure on $[0, 1]$.*

- 1) *For all $N \in \mathbb{N}$, there exists a finite set $(x_i)_{i=1}^N$ such that ν_N as in (1.2) satisfies*

$$D_N^*(\mu; \nu_N) \leq \frac{1}{N}.$$

- 2) *Moreover, there exists a sequence $(x_k)_{k \in \mathbb{N}} \in [0, 1]$ such that for all $N \in \mathbb{N}$ and ν_N as in (1.2) we have*

$$D_N^*(\mu; \nu_N) \leq c \frac{\log(N)}{N}$$

with a constant c independent of N .

- 3) *Suppose μ is a Borel measure with no point masses. That is, $\mu = \mu_{ac} + \mu_{cs}$. Then*

$$D_N^*(\mu; \nu_N) \geq \frac{1}{2N} \tag{1.4}$$

for any finite set $(x_i)_{i=1}^N$ and ν_N as in (1.2). Moreover, there exists a constant c so that for infinitely many $N \in \mathbb{N}$,

$$D_N^*(\mu; \nu_N) \geq c \frac{\log N}{N} \tag{1.5}$$

holds for any sequence $(x_k)_{k \in \mathbb{N}}$ with ν_N defined as in (1.2).

REMARK 1.6. Example (3.1) demonstrates that when μ does have point masses, 3) does not hold. When μ does have point masses, so that part 3) of Theorem 1.5 does hold, the theorem states that the optimal rate of convergence is $O(N^{-1})$ and $O(N^{-1} \log(N))$ for sets and sequences respectively. When $\mu = \lambda_1$, the Lebesgue measure, it was already known this is the best rate for convergence for point sets [14, Theorem 2.6] and sequences [16].

REMARK 1.7. From the the duality between the discrepancy (for the Lebesgue measure) of finite point sets in dimension $d + 1$ and sequences in dimension d , it is tempting to assume that part 2) of Theorem 1.5 implies that for every Borel measure μ on $[0, 1]^2$ there exists a finite set $(x_i)_{i=1}^N \in [0, 1]^2$ such that for ν_N as in (1.2) we have

$$D_N^*(\mu; \nu_N) \leq c \frac{\log(N)}{N}$$

with a constant c independent of N . However, the proof of the duality relies on the fact that the Lebesgue measure satisfies

$\lambda_d([0, a_1] \times \cdots \times [0, a_{d-1}] \times [0, a_d]) = \lambda_{d-1}([0, a_1] \times \cdots \times [0, a_{d-1}]) \cdot \lambda_1([0, a_d])$, compare [13, Example 2.2]. Thus, the duality does not transfer (directly) to general Borel measures.

Lemma 1.8 is not only a crucial tool in the proof of Theorem 1.5, but it is also of independent interest. It is a method to construct a point sets allowing one to compare the discrepancy between the Lebesgue measure and a Borel measure on $[0, 1]$. More precisely,

LEMMA 1.8. *Let μ be a probability measure on $[0, 1]$. If $(v_k)_{k=1}^M \subseteq [0, 1]$ is a finite or countably infinite set of points, then there exists $(x_k)_{k=1}^M$ so that for all $2 \leq N \leq M$ and ν_N as in (1.2),*

$$D_N^*(\mu; \nu_N) \leq D_N^*\left(\lambda_1; \frac{1}{N} \sum_{i=1}^N \delta_{v_i}\right). \quad (1.6)$$

Moreover, if μ has no point masses, that is $\mu_a([0, 1]) = 0$ (see Remark 1.4), then in Equation (1.6) we in fact have equality.

2. Total variation

In this section we consider discrete probability measures on $[0, 1]^d$. That is, measures of the form

$$\mu = \sum_{i=1}^{\infty} \alpha_i \delta_{y_i}, \quad (2.1)$$

where $\sum_i \alpha_i = 1$ and $(y_i) \subset [0, 1]^d$. Given sufficient decay of the (α_i) , we provide an algorithm for picking sets $(x_i^N)_{i=1}^N$ that produce finite Dirac measures ν_N as in (1.2) whose total variation distance from μ has quantifiable decay rate.

The interested reader should note that the proofs work identically if you replace the stronger notion of total variation measure, with the weaker notion of star-discrepancy in every instance throughout this proof.

LEMMA 2.1 (Finite to infinite lemma). *Let μ be an infinitely supported discrete probability measure as in (2.1). For each $K \in \mathbb{N}$ define*

$$\mu_K = \left(\sum_{i=1}^K \alpha_i \right)^{-1} \sum_{i=1}^K \alpha_i \delta_{y_i}.$$

If there exists a continuous decreasing function $h : [2, \infty) \rightarrow [0, 1/2]$ and a constant $c > 0$ so that for all $N \in \mathbb{N}$ there exists $K \geq N$ and a set $(x_i^N)_{i=1}^N$ so that ν_N as in (1.2) satisfies

$$\rho(\mu_K; \nu_N) \leq ch(N),$$

then

$$\rho(\mu; \nu_N) \leq (c + 3)h(N). \quad (2.2)$$

Proof. Fix $N \in \mathbb{N}$ and choose $K_0 \in \mathbb{N}$ such that $\sum_{i=K_0+1}^{\infty} \alpha_i \leq h(N)$. By assumption, there exists $K \geq \max\{K_0, N\}$ and a set $x_1, \dots, x_N$ with

$$\rho(\mu_K; \nu_N) \leq ch(N).$$

It follows from the triangle inequality that

$$\rho(\mu; \nu_N) \leq \rho(\mu; \mu_K) + \rho(\mu_K; \nu_N).$$

To bound the first term, $\sum_{i=K+1}^{\infty} \alpha_i \leq h(N)$ ensures for any $A \in \mathcal{B}([0, 1]^d)$

$$\begin{aligned} |\mu_K(A) - \mu(A)| &= \left| \left(\frac{1}{\sum_{i=1}^K \alpha_i} - 1 \right) \sum_{i=1}^K \alpha_i \delta_{y_i}(A) - \sum_{i=K+1}^{\infty} \alpha_i \delta_{y_i}(A) \right| \\ &\leq \frac{h(N)}{1 - h(N)} + h(N) \\ &= h(N) \left(\frac{1}{1 - h(N)} + 1 \right). \end{aligned}$$

As $0 \leq h(N) \leq \frac{1}{2}$, this verifies (2.2). $\square$

Now we describe an algorithm that, given a measure μ supported on k points, chooses N points defining ν_N as in (1.2) whose total variation distance from μ is at most $\frac{c_k}{N}$ where $c_k \leq 2k$. After Proposition 2.2, we will combine with Lemma 2.1 to get information the total variation distance of infinitely supported discrete measures.

PROPOSITION 2.2. *Let μ be a probability on $[0, 1]^d$ which is supported on no more than k points. Let $N \in \mathbb{N}$. There exists a constant c_k independent of the dimension and a finite set $x_1^N, \dots, x_N^N \in [0, 1]^d$ so that ν_N as in (1.2) satisfies $\rho(\mu; \nu_N) \leq \frac{c_k}{N}$. Notably, $c_k \leq 2k$.*

Proof. If $N \leq k - 1$, then choose

$$x_i^N = \begin{cases} (0, 0, \dots, 0), & i \leq \lceil \frac{N-1}{2} \rceil, \\ (1, 1, \dots, 1), & \lceil \frac{N-1}{2} \rceil < i \leq N \end{cases}$$

that is, all points of the finite sequence lie on the origin or on its opposite corner. For these N we have the trivial bound,

$$\rho(\mu; \nu_N) \leq 2 \leq 2 \cdot \frac{k-1}{N} \leq \frac{c_k}{N}.$$

If $N \geq k$, write $\mu = \sum_{i=1}^k \alpha_i \delta_{y_i}$ with $\sum_{i=1}^k \alpha_i = 1$. For each α_i choose $p_i \in \mathbb{N}$ such that

$$0 \leq \alpha_i - \frac{p_i}{N} < \frac{1}{N}. \quad (2.3)$$

Summing the preceding inequality over i ensures there exists $r \in \mathbb{N}$ such that $r < k$ and

$$\sum_{i=1}^k \alpha_i - \sum_{i=1}^k \frac{p_i}{N} = \frac{r}{N} \iff N - \sum_{i=1}^k p_i = r.$$

Next, we define a finite set $(x_i^N)_{i=1}^N$ by first placing p_j points on each y_j , then placing the remaining points at the origin and the opposite corner. More formally, adopting the notation that $\sum_{i=1}^0 \alpha_i = 0$, for $h = 1, \dots, N - r$ choose

$$x_h^N = y_j \quad \text{whenever} \quad \sum_{i=1}^{j-1} p_i < h \leq \sum_{i=1}^j p_i.$$

To place the remaining r points, define

$$x_h^N = \begin{cases} (0, 0, \dots, 0), & N - r < h \leq N - \lceil \frac{r}{2} \rceil, \\ (1, 1, \dots, 1), & h > N - \lfloor \frac{r}{2} \rfloor. \end{cases}$$

Next, choose some $A \in \mathcal{B}$. Let I be the index set defined by $i \in I \iff y_i \in A$. Then, (2.3) yields

$$|\mu(A) - \nu_N(A)| \leq |\nu_N(\{0\} \cup \{1\})| + \sum_{i \in I} \alpha_i - \frac{p_i}{N} \leq \frac{2r}{N}.$$

Since $r < k$, $c_k = 2k$ suffices. □

We are now ready to prove the main theorem of this section. Examples 2.5 and 2.7 will illuminate how the explicit upper-bound works.

THEOREM 2.3. *Let $(r_k)_{k \in \mathbb{N}}$ be a sequence so that $r_1 = 1$, $\sum_{k=1}^{\infty} r_k < \infty$, and $r_i = 0$ implies $r_j = 0$ for all $j \geq i$. Moreover, suppose there is a constant c_0 so that for all $M \geq 2$, $\sum_{k=M}^{\infty} r_k \leq \frac{c_0}{M}$.*

Fix $d \geq 1$ and a discrete probability measure μ as in (2.1) on $[0, 1]^d$, where $(\alpha_i)_{i=1}^{\infty}$ satisfy

$$\alpha_k \leq r_k \alpha_1. \quad (2.4)$$

For all $N \geq 2$ there exists a finite set $(x_i^N)_{i=1}^N$ so that ν_N as in (1.2) satisfies

$$\rho(\mu; \nu_N) < (6c_0 + 3) \frac{g^{-1}\left(\frac{1}{N\alpha_1}\right)}{N}, \quad (2.5)$$

where $g : [1, \infty) \rightarrow [0, \infty)$ is a continuous, strictly decreasing function such that

$$\begin{cases} \sum_{i=M}^{\infty} r_i \leq g(M), & M \in \mathbb{N}_{\geq 2}, \\ g(s) \leq \frac{c_0}{s}, & s \in [1, \infty). \end{cases} \quad (2.6)$$

We refer to the function g as the gauge function corresponding to the sequence $(r_k)_{k \in \mathbb{N}}$.

Proof of Theorem 2.3. Let g be as in the theorem statement. Note, the gauge function g can be chosen to be strictly decreasing on its support since $K \mapsto \sum_{i=K}^{\infty} r_i$ is strictly decreasing for all $K \leq K_0$ and K_0 is the (possibly infinite) number of non-zero r_i .

For $N \in \mathbb{N}$, choose K_N to be the smallest integer so that $K_N \geq g^{-1}\left(\frac{1}{\alpha_1 N}\right)$. In particular, $K_N < g^{-1}\left(\frac{1}{\alpha_1 N}\right) + 1$.

Now we consider the normalized Borel measure

$$\mu_{K_N} = \frac{1}{\sum_{i=1}^{K_N} \alpha_i} \sum_{i=1}^{K_N} \alpha_i \delta_{y_i}.$$

By Proposition 2.2, there exists a set $(x_i^N)_{i=1}^N$ such that

$$\rho(\mu_{K_N}, \nu_N) \leq \frac{2K_N}{N} < 2 \frac{g^{-1}\left(\frac{1}{N\alpha_1}\right) + 1}{N} \leq 3 \frac{g^{-1}\left(\frac{1}{N\alpha_1}\right)}{N} \quad (2.7)$$

because $g^{-1}(x) \geq 1$. To apply Lemma 2.1, we need to find an appropriate function $h : [2, \infty) \rightarrow [0, 1/2]$. Consider,

$$h(N) := (2c_0 N)^{-1} g^{-1}\left((N\alpha_1)^{-1}\right). \quad (2.8)$$

To confirm $h(N) \leq 1/2$, note (2.6) gives $g(c_0 N) \leq N^{-1} \leq (N\alpha_1)^{-1}$. Since g is monotonically decreasing this implies $c_0 N \geq g^{-1}\left(\frac{1}{N\alpha_1}\right)$, verifying $h(N) \leq 1/2$.

Equation (2.7) and (2.8) imply the constant c in Lemma 2.1 is $6c_0$. Thus, Lemma 2.1 verifies (2.5). $\square$

REMARK 2.4. The interested reader may note that nowhere in Lemma 2.1, Proposition 2.2, or Theorem 2.3 do we use the fact that μ is a measure on $[0, 1]^d$. All three of the results could go through identically if (X, μ) were some probability space and μ is a probability measure. In Proposition 2.2, the points $(0, \dots, 0)$ and $(1, \dots, 1)$ can be replaced with any points in the space X .

However we chose to write the proof this way because it allows for a better bound for the constant term when working with discrepancy. Indeed (2.3) yields for any $A \in \mathcal{A}$ an axis parallel box,

$$-\left\lceil \frac{r}{2} \right\rceil \leq \mu(A) - \nu(A) = -\nu(\{0\}) + \sum_{i \in I} \alpha_i - \frac{p_i}{N} \leq \left\lfloor \frac{r}{2} \right\rfloor.$$

The upper and lower bound can potentially be achieved when $I \in \{\emptyset, \{1, \dots, k\}\}$. Finally, since $r < k$, $c_k = k/2$ suffices for a star-discrepancy specific version of Proposition 2.2.

EXAMPLE 2.5. Set $r_k = r^{k-1}$ for some $0 < r < 1$, then for every $N \geq 2$ there exists $(x_1^N)_{i=1}^N$ such that the corresponding ν_N as in (1.2) satisfies

$$\rho(\mu; \nu_N) \leq c_r \frac{\log(N)}{N}. \quad (2.9)$$

Indeed, note

$$\sum_{k=K}^{\infty} r_k = \frac{r^K}{1-r}. \quad (2.10)$$

So we choose $g(s) := \frac{r^s}{1-r}$ for $s \geq 1$ as a strictly decreasing gauge function for the tail of the sequence. We need $g(s) \leq \frac{c_0}{s}$. We write $c_0 = \frac{r^{c_r}}{1-r}$ and compute c_0 . This means, we require

$$\frac{r^s}{1-r} \leq \frac{r^{c_r}}{s(1-r)} \iff sr^s \leq r^{c_r}.$$

Note $G(s) := sr^s$ is maximized when $s = -1/\ln(r)$, so choosing c_r so that $r^{c_r} = G(-1/\ln(r))$ yields $c_0 = \frac{-1}{e \ln(r)(1-r)}$ satisfies $g(s) \leq \frac{c_0}{s}$. Defining $s_N = g^{-1}(\frac{1}{\alpha_1 N})$ implies

$$\frac{r^{s_N}}{1-r} = \frac{1}{N\alpha_1}$$

which is equivalent to

$$s_N = \frac{\log(\alpha_1 N) - \log(1-r)}{-\log(r)} \leq \frac{\log(N) - \log(1-r)}{-\log(r)} \leq \tilde{c}_r \log(N),$$

where

$$\tilde{c}_r = \frac{\log(2) - \log(1-r)}{-\log(r) \log(2)}.$$

Applying Theorem 2.3 verifies (2.9) where $c_r = (6c_0 + 3)\tilde{c}_r$. Since c_0 (which despite the notation depends on r) and $\tilde{c}_r$ are both monotonically increasing for $r \in (0, 1)$, the same can be said about c_r .

The next example will use a fact which we recall here for the reader's convenience. See for instance, [3, 5.1.20].

PROPOSITION 2.6. *For any $a > 0$, the following holds*

$$\int_a^\infty \frac{e^{-t}}{t} dt < e^{-a} \log \left(1 + \frac{1}{a} \right).$$

We emphasize that in the next example, the discrete measures have total variation distance that, under the algorithm of Proposition 2.2 converges faster than the Lebesgue discrepancy, even when $d = 2$ (see Remark 1.2). In particular, the star-discrepancy for any measure in this family of converges faster than the Lebesgue.

EXAMPLE 2.7. Let $r_k = \frac{r^{e^k}}{r^e}$ where $0 < r < \frac{1}{2}$. Then for each N , there exists a set $(x_i^N)_{i=1}^N$ and a constant c_r so that the associated measure ν_N as in (1.2) satisfies

$$\rho(\mu; \nu_N) \leq c_r \frac{\log \left(\frac{\log(N)}{|\log(r)|} \right)}{N}. \quad (2.11)$$

Indeed, $g(m) = r^{e^m}$ satisfies the assumptions of Theorem 2.3 since, using Proposition 2.6

$$\begin{aligned} \sum_{k=m+1}^\infty r^{e^k} &\leq \int_m^\infty r^{e^s} ds = |\log(r)| \int_{e^m}^\infty \frac{e^{-|\log(r)|t}}{|\log(r)|t} dt = \int_{|\log(r)|e^m}^\infty \frac{e^{-u}}{u} du \\ &< e^{-|\log(r)|e^m} \log \left(1 + \frac{1}{|\log(r)|e^m} \right) \\ &\leq e^{-|\log(r)|e^m} = r^{e^m}. \end{aligned}$$

The penultimate line comes from the fact that $r < \frac{1}{2} < e^{(e^{-1}-1)}$, and thus

$$1 + \frac{1}{|\log(r)|e^m} < e$$

for all $m \geq 1$. Moreover, $g(s) \leq \frac{c_0}{s}$ with $c_0 = r^e$. Define $s_N = g^{-1} \left(\frac{1}{\alpha_1 N} \right)$.

That is,

$$r^{e^{s_N}} = \frac{1}{\alpha_1 N} \iff \log(r)e^{s_N} = -\log(\alpha_1 N) \iff$$

$$s_N = \log\left(\frac{\log(\alpha_1 N)}{|\log(r)|}\right) \leq \log\left(\frac{\log(N)}{|\log(r)|}\right).$$

Applying Theorem 2.3 verifies (2.11) with $c_r = (6r^e + 3)$.

The last example is intended to demonstrate that this technique of approximating discrete measures in general cannot be the optimal way to do so. In this example, it only yields something equivalent to the trivial bound for total variation distance. One reason to expect this method is not sharp is that it completely ignores the location of the points, which should be very important especially in higher dimensions.

EXAMPLE 2.8. Let $r_k = 1/k^2$. Then the same method only guarantees that there exists a set $(x_i^N)_{i=1}^N$ so that the associated ν_N as in (1.2) satisfies

$$\rho(\mu; \nu_N) \leq 9. \quad (2.12)$$

Indeed,

$$\sum_{k=m+1}^{\infty} \frac{1}{k^2} \leq \int_m^{\infty} \frac{1}{s^2} ds = \frac{1}{m} =: g(m).$$

Then

$$K_N = g^{-1}((\alpha_1 N)^{-1}) = \alpha_1 N \leq N.$$

Applying Theorem 2.3 gives the bound in (2.12).

3. Discrepancy in one dimension

In order to prove Theorem 1.5, we will make use of the following lemma which constructs finite sets and/or sequences for which one can compare the discrepancy with respect to Lebesgue and to an arbitrary Borel measure.

Lemma 1.8. Let μ be a probability measure on $[0, 1]$. If $(v_k)_{k=1}^M \subseteq [0, 1]$ is a finite or countably infinite set of points, then there exists $(x_k)_{k=1}^M$ so that for all $2 \leq N \leq M$ and ν_N as in (1.2),

$$D_N^*(\mu; \nu_N) \leq D_N^*\left(\lambda_1; \frac{1}{N} \sum_{i=1}^N \delta_{v_i}\right). \quad (3.1)$$

Moreover, if μ has no point masses, that is $\mu_d([0, 1]) = 0$ (see Remark 1.4), then in Equation (3.1) we in fact have equality.

The strategy of the proof is to use a cumulative distribution type function for μ to create an injective function from $\mathcal{A} \rightarrow \mathcal{A}$ (recall $\mathcal{A}$ is the set of half-open intervals contained in $[0, 1]$ containing 0) that can be used to pull back a set or sequence (v_k) with given Lebesgue discrepancy to find a set or sequence (x_k) that has the same μ discrepancy. When μ_d is zero, the distribution function creates a bijection not just an injection. This fails in higher-dimensions because axis parallel boxes do not have a well-ordering that also respects the geometry. Namely different boxes can see different points in a different order.

Proof of Lemma 1.8. Note that the function

$$f(a) := \mu([0, a))$$

is non-decreasing and continuous from the left.

Let $(v_k)_{k \geq 1}$ be an arbitrary finite or infinite set. For each $k \geq 1$ define x_k so that

$$\begin{cases} f(a) \leq v_k, & \forall a \leq x_k, \\ f(a) > v_k, & \forall a > x_k. \end{cases} \quad (3.2)$$

This can be done since f is non-decreasing and continuous from the left. Fix $N \in \mathbb{N}_{\geq 2}$ and an interval $[0, b)$.

Consider $c := f(b)$. Due to the monotonicity and one-sided continuity of f one of the following holds:

$$f^{-1}(\{c\}) = \begin{cases} \{b\}, \\ (b - \delta_1, b + \delta_2] \end{cases} \quad \text{for some } \delta_1, \delta_2 \geq 0.$$

The first case occurs when f is strictly increasing at b , and the second case occurs if f is constant on a sub-interval containing b . In either case, define $b_0 = \max\{f^{-1}(f(b))\}$.

Claim 1. *The discrepancy of $(x_k)_{k=1}^N$ with respect to μ over the interval $[0, b_0)$ is the same as the Lebesgue discrepancy of $(v_k)_{k=1}^N$ over $[0, c)$. That is*

$$\left| \frac{\#\{x_k < b_0 : k \leq N\}}{N} - \mu([0, b_0)) \right| = \left| \frac{\#\{v_k < c : k \leq N\}}{N} - \lambda_1([0, c)) \right|. \quad (3.3)$$

Indeed, since $f(b_0) = c = \lambda_1([0, c))$ we have $\mu([0, b_0)) = \lambda_1([0, c))$. Moreover, if $x_k < b_0$, then $f(b_0) > v_k$ by (3.2). In particular, $v_k < c$. On the other hand, if $v_k < c$, then

$$f(x_k) \leq v_k < c = f(b_0),$$

so the non-decreasing nature of f combined with the strict inequality forces $x_k < b_0$. This verifies Claim 1.

Claim 2. *The discrepancy of $(x_k)_{k=1}^N$ for μ over the intervals $[0, b)$ and $[0, b_0)$ are equal. That is*

$$\left| \frac{\#\{x_k < b_0 : k \leq N\}}{N} - \mu([0, b_0)) \right| = \left| \frac{\#\{x_k < b : k \leq N\}}{N} - \mu([0, b)) \right|. \quad (3.4)$$

Indeed, if $b = b_0$, this statement is trivial. If $b_0 > b$, then we still have

$$f(b_0) = f(b),$$

so we only need to show that $(x_k)_{k \leq N} \cap [b, b_0) = \emptyset$. Since f is constant on the interval $[b, b_0)$, by (3.2), either $x_k = b_0$ which is outside $[b, b_0)$, or $x_k \leq b - \delta_1$, so $(x_k)_{k \leq N} \cap [b, b_0) = \emptyset$ as desired.

Claim 1 and Claim 2 ensure that for every interval $[0, b)$ associated to μ , the interval $[0, f(b))$ has the same Lebesgue discrepancy. Taking the supremum over all b ensures

$$D_N^*(\mu; \nu_N) = \sup_b \left| \sum_{i=1}^N \frac{\mathbb{1}_{[0, f(b))}(v_i)}{N} - \lambda_1[0, f(b)) \right| \leq D_N^* \left(\lambda_1; \frac{1}{N} \sum_{i=1}^N \delta_{v_i} \right), \quad (3.5)$$

verifying (3.1).

When $\mu = \mu_{ac} + \mu_{sc}$ has no discrete part, then f is in fact continuous. Therefore f maps onto $[0, 1)$. Hence taking the supremum over $[0, f(b))$ is now equivalent to taking the supremum over $[0, c)$ forcing an equality in (3.5). $\square$

We now utilize Lemma 1.8 to prove Theorem 1.5, which we repeat for convenience.

Theorem 1.5. Fix μ a normalized Borel measure on $[0, 1]$.

- 1) For all $N \in \mathbb{N}$, there exists a finite set $(x_i)_{i=1}^N$ such that ν_N as in (1.2) satisfies

$$D_N^*(\mu; \nu_N) \leq \frac{1}{N}.$$

- 2) Moreover, there exists a sequence $(x_k)_{k \in \mathbb{N}} \in [0, 1]$ such that for all $N \in \mathbb{N}$ and ν_N as in (1.2) we have

$$D_N^*(\mu; \nu_N) \leq c \frac{\log(N)}{N}$$

with a constant c independent of N .

- 3) Suppose μ is a Borel measure with no point masses. That is, $\mu = \mu_{ac} + \mu_{cs}$. Then

$$D_N^*(\mu; \nu_N) \geq \frac{1}{2N} \quad (3.6)$$

for any finite set $(x_i)_{i=1}^N$ and ν_N as in (1.2). Moreover, there exists a constant c so that for infinitely many $N \in \mathbb{N}$,

$$D_N^*(\mu; \nu_N) \geq c \frac{\log N}{N} \quad (3.7)$$

holds for any sequence $(x_k)_{k \in \mathbb{N}}$ with ν_N defined as in (1.2).

Proof of Theorem 1.5.

- 1) For $k = 1, \dots, N$, set $v_k = \frac{2k-1}{2N}$. Let x_k be defined as in (3.2). Then by Theorem 2.7 of [14] combined with Lemma 1.8,

$$D_N^*(\mu; \nu_N) \leq D_N^*\left(\lambda_1; \frac{1}{N} \sum_{i=1}^N \delta_{v_i}\right) = \frac{1}{N}.$$

- 2) Now let $(v_k)_{k \in \mathbb{N}}$ be any low-discrepancy sequence with respect to Lebesgue measure and let $(x_k)_{k \in \mathbb{N}}$ be defined as in (3.2). Then by Lemma 1.8,

$$D_N^*(\mu; \nu_N) \leq D_N^*\left(\lambda_1; \frac{1}{N} \sum_{i=1}^N \delta_{v_i}\right) \leq c \frac{\log(N)}{N}.$$

- 3) Finally, suppose $\mu = \mu_{ac} + \mu_{sc}$. Combining Lemma 1.8 with the results of Schmidt and Niederreiter (see Remark 1.6) verifies (3.6) and (3.7).

□

Thus for an arbitrary probability measure μ on $[0, 1]$, there exist sequences whose μ discrepancy converges to zero at least as fast as in the Lebesgue case. Moreover, when μ has no discrete part, the rate from the Lebesgue measure is in fact optimal. The following example shows that in the discrete case there exist sequences whose discrepancy converge strictly faster than in the Lebesgue case.

EXAMPLE 3.1. Let δ_y denote the Dirac measure centered at y and consider the Borel measure $\mu := \frac{1}{2}\delta_0 + \frac{1}{2}\delta_{0.5}$. Then the sequence $(x_k)_{k \in \mathbb{N}}$ defined by $x_{2k+1} = 0$ and $x_{2k} = 1/2$ has

$$D_N^*(\mu; \nu_N) = \begin{cases} \frac{1}{2N}, & N \text{ odd}, \\ 0, & N \text{ even}. \end{cases}$$

ACKNOWLEDGEMENT. This research began at the trimester program *Dynamics: Topology and Numbers* at the Hausdorff Research Institute for Mathematics (HIM) in Bonn. The authors would like to thank HIM for bringing them together and for their hospitality. Moreover, the authors would like to thank Stefan Steinerberger for insightful comments on the first draft of this document and also the anonymous referee for valuable comments. The second author was partially supported by PIMS and by FRG DMS-1853993.

FAMILIES OF WELL APPROXIMABLE MEASURES

REFERENCES

- [1] AISTLEITNER, C.—BILYK, D.—NIKOLOV, A.: *Tusnady’s problem, the transference principle, and non-uniform QMC sampling*. In: Monte Carlo and Quasi-Monte Carlo Methods—MCQMC 2016, Springer Proc. Math. Stat. Vol. 241, Springer, Cham, 2018 pp. 169–180.
- [2] AISTLEITNER, C.—DICK, J.: *Low-discrepancy point sets for non-uniform measures*, Acta Arith. **163** (4) (2014), 345–369.
- [3] ABRAMOWITZ, M.—STEGUN, I. A.: *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. National Bureau of Standards Applied Mathematics Series, Vol. 55. U.S. Government Printing Office, Washington, D. C. 1948.
- [4] BECK, J.: *Some upper bounds in the theory of irregularities of distribution*, Acta Arith. **43** (2) (1984), 115–130.
- [5] CARBONE, I.: *Discrepancy of LS-sequences of partitions and points*, Ann. Mat. Pura Appl. **191** (2012), 819–844.
- [6] CHEN, W.: *On irregularities of distribution and approximate evaluation of certain functions*, Quart. J. Math. Oxford Ser. (2) **142** (1985), no. 36, 173–182.
- [7] GIBBS, A.—SU, F. E.: *On choosing and bounding probability metrics*, Int. Stat. Review **70** (2002), 419–435.
- [8] HELLINGER, E.: *Neue Begrundung der Theorie quadratischer Formen von unendlich vielen Veranderlichen*, J. Reine Angew Math. **136** (1909), 210–271.
- [9] HLAWKA, E.—MUCK, R.: *A transformation of equidistributed sequences*. In: Proc. Sympos., Univ. Montreal, Montreal, Que., 1971. Applications of Number Theory to Numerical Analysis (1972), pp. 371–388.
- [10] HLAWKA, E.—MUCK, R.: *Uber eine Transformation von gleichverteilten Folgen*. II. Computing (Arch. Elektron. Rechnen) **9** (1972), 127–138.
- [11] HEWITT, E.—STROMBERG, K.: *Real and Abstract Analysis: A Modern Treatment of the Theory of Functions of a Real Variable*. Graduate Texts in Math. Vol. 25, Springer-Verlag, Berlin, 1975.
- [12] KULLBACK, S.—LEIBLER, R.: *On information and sufficiency*, Ann. Math. Stat. **22** (1951), no. 1, 79–86.
- [13] KUIPERS, L.—NIEDERREITER, H.: *Uniform Distribution of Sequences*. John Wiley & Sons, New York, 1974.
- [14] NIEDERREITER, H.: *Random Number Generation and Quasi-Monte Carlo Methods*. In: CBMS-NSF Regional Conference Series in Appl. Math. Vol. 63. Society for Industrial and Appl. Math. (SIAM), Philadelphia, PA, 1992.
- [15] ROTH, K. F.: *On irregularities of distribution*, Mathematika **1** (1954), 73–79.
- [16] SCHMIDT, W. M.: *Irregularities of distribution VII*, Acta Arith. **21** (1972), 45–50.

- [17] VILLANI, C.: *Optimal Transport, Old and New*. Springer-Verlag, Berlin, 2009.
- [18] WEISS, C.: *Interval exchange transformations and low-discrepancy*, Ann. Mat. Pura Appl. (4) **198** (2019), no. 2, 399–410.

Received September 2, 2020

Accepted April 13, 2021

Samantha Fairchild

Max Goering

University of Washington

Department of Mathematics

P.O. Box 354350

98195 Seattle

WA, U.S.A

E-mail: skayf@uw.edu

mgoering@uw.edu

Christian Weiß

Ruhr West University of Applied Sciences

Department of Natural Sciences

Duisburger Str. 100

45479 Mülheim an der Ruhr

GERMANY

E-mail: christian.weiss@hs-ruhrwest.de