sciendo

uniform distribution theory

DOI: 10.2478/UDT-2023-0003
Unif. Distrib. Theory **18** (2023), no.1, 31–38

APPROXIMATION OF DISCRETE MEASURES BY FINITE POINT SETS

Weiss, Christian

Ruhr West University of Applied Sciences, Mülheim a. d. Ruhr, GERMANY

ABSTRACT. For a probability measure μ on $[0,1]$ without discrete component, the best possible order of approximation by a finite point set in terms of the star-discrepancy is $\frac{1}{2N}$ as has been proven relatively recently. However, if μ contains a discrete component no non-trivial lower bound holds in general because it is straightforward to construct examples without any approximation error in this case. This might explain, why the approximation of discrete measures on $[0,1]$ by finite point sets has so far not been completely covered in the existing literature. In this note, we close the gap by giving a complete description for discrete measures. Most importantly, we prove that for any discrete measures (not supported on one point only) the best possible order of approximation is for infinitely many N bounded from below by $\frac{1}{cN}$ for some constant $6 \geq c > 2$ which depends on the measure. This implies, that for a finitely supported discrete measure on $[0,1]^d$ the known possible order of approximation $\frac{1}{N}$ is indeed the optimal one.

Communicated by Friedrich Pillichshammer

1. Introduction

According to [2], the Lebesgue measure is the hardest Borel measure on $[0,1]$ to approximate by a finite point set. In order to formulate the result in a mathematically precise way, recall first that the star-discrepancy between

2020 Mathematics Subject Classification: 11K36, 11K38, 11J71, 11K31.
Keywords: Star-discrepancy, discrete measures, Borel measures.

two probability measures μ, ν on $[0,1]$ 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 intervals in $[0,1]$ which have one vertex at the origin. Furthermore, the probability measure associated to a finite set $(y_i)_{i=1}^N$ is given by

$$\nu_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{y_i}, \tag{1}$$

where δ_{y_i} denotes the Dirac measure supported on y_i. Borel measures on the interval $[0,1]$ have a particularly comprehensible structure. Lebesgue's decomposition theorem states that 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. It means that, μ_{ac} is zero on sets of Lebesgue measure zero. Measure μ_d is a discrete measure, that is, it is zero on the complement of some countable set, and μ_{cs} is continuous singular. Which menas that μ_{cs} is zero on the complement of some set B of Lesbesgue measure zero but assigns no weight to any countable set of points, see, e.g. [3, Chapter V]. Based on this observation, the following result holds.

Theorem 1.1 ([2], Theorem 1.5)**.** *Fix μ a probability measure on $[0,1]$.*

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

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

(ii) *Suppose μ is a probability measure with no point masses. That is,*

$$\mu = \mu_{ac} + \mu_{cs}.$$

Then

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

for any finite set $(y_i)_{i=1}^N$ and ν_N as in (1).

This result answered the general question from [1], where the authors asked for arbitrary dimensions which Borel measure on $[0,1]^d$ is the hardest to approximate by finite point sets, in the one-dimensional case. However, the lower bound in Theorem 1.1 (ii), i.e., the best theoretically possible speed of approximation by finite point sets, is restricted to measures without discrete components only.

The main purpose of this note is to close this gap and thereby to complete the discussion on approximation of measures by finite point sets in the one-dimensional case.

Already the simplest possible example of approximating the Dirac measure supported on $x_0 \in [0,1]$ by a finite point set yields some insight: indeed, it is possible in the discrete case (in contrast to the other cases) to have an approximation error

$$D_N^*(\mu, \nu_N) = 0.$$

However, this is a rather special situation as our main theorem shows.

Theorem 1.2. *Fix a discrete probability measure μ on $[0,1]$.*

(i) *If μ is supported on one point only, then for every N there exists a finite point set such that*

$$D_N^*(\mu, \nu_N) = 0.$$

(ii) *If μ is supported on more than one point, then*

$$D_N^*(\mu, \nu_N) \geq \frac{1}{6N}$$

holds for infinitely many $N \in \mathbb{N}$ and all finite point sets.

(iii) *If μ is given by*

$$\sum_{i=1}^{\infty} \xi_i \delta_{x_i}$$

with $x_i < x_{i+1}$ and $\xi_i \geq 0$ for all $i \in \mathbb{N}$ and $\sum_{i=1}^{l} \xi_i \notin \mathbb{Q}$ for some $l \in \mathbb{N}$, then for every constant $c > 2$, there are infinitely many

$$N \in \mathbb{N} \quad \text{with} \quad D_N^*(\mu, \nu_N) \geq \frac{1}{cN}$$

for every finite point set.

While part (i) of Theorem 1.2 is trivial and also the proof of (ii) is very short, a bit more involved arguments are needed to show (iii). It is possible to use our approach also in higher dimensions and we again obtain a lower bound of the form

$$D_N^*(\mu, \nu_N) \geq \frac{1}{6N}$$

for infinitely many N. Together with [2, Proposition 2.2], this leads to the following interesting corollary.

Corollary 1.3. *Let μ be a probability measure on $[0,1]^d$ which is supported on a finite number of points $k \in \mathbb{N}$. Then there exists constant $c_\mu \leq 6$, which depends on the measure, such that*

$$D_N^*(\mu, \nu_N) \geq \frac{1}{c_\mu N}$$

for infinitely N and arbitrary point sets $(y_i)_{i=1}^N$. Moreover, for any $N \in \mathbb{N}$, there exists a constant C_k, which only depends on the number of points, and a finite set $(y_i)_{i=1}^N$ such that

$$D_N^*(\mu, \nu_N) \leq \frac{C_k}{N}.$$

In other words, in any dimension any finitely supported discrete measure can be approximated by a finite set of order $\frac{1}{N}$ and this is the best possible order of approximation. If the decay rate of the weights is fast enough, it is furthermore known due to [2, Theorem 1.1]. that certain measures can be approximated by order of convergence at most $\log(N)/N$. Nonetheless, we expect that due to the combinatorial richness of inclusion of half-open intervals in higher dimensions, there exist infinitely supported discrete measures in dimensions $d \geq 3$ with a bigger minimal possible order of approximation. We therefore ask the question under which conditions on the probability measure the lower bounds from Theorem 1.2 are also optimal in higher dimensions. This question is of particular interest because lower bounds for the star-discrepancy, e.g., for the Lebesgue measure, are typically very hard obtained, compare [4, 6].

The main reasons why Theorem 1.2 holds can be easiest understood by considering the second simplest example, namely discrete measures supported on two points x_1, x_2 only. It turns out that **Kronecker sequences**, which are for $\alpha \in \mathbb{R}$ defined by $\{i\alpha\}_{i=1}^\infty$, where $\{\cdot\}$ denotes the fractional part of a real number (see, e.g. [4]), appear here prominently.

Example 1. As an illuminating example let us start with the case that μ consists of two point masses only, supported on $0 \leq x_1 < x_2 \leq 1$. For fixed $N \in \mathbb{N}$ some of the weight, i.e., some of the y_i, needs to be placed at x_1 and the rest at x_2 because all other choices would automatically increase the star-discrepancy. This implies that

$$\left|\mu\big([0,z)\big) - \nu_N\big([0,z)\big)\right| = 0 \quad \text{for } z \leq x_1 \quad \text{and} \quad z \geq x_2.$$

Thus, we only need to consider the case $x_1 < z < x_2$ in the following.

At first, we assume $\xi_1, \xi_2 \in \mathbb{Q}$ and let $\xi_i = p_i/q_i$ with $\gcd(p_i, q_i) = 1$ for $i = 1, 2$. If $q_1 | N$, then also $q_2 | N$. Choosing $k_1 = p_1 \cdot N/q_1$ times $y_i = x_1$ and $k_2 = p_2 \cdot N/q_2$ times $y_i = x_2$ yields a star-discrepancy $D_N^*(\mu, \nu_N) = 0$ as predicted by Theorem 1.2 (i). If on the other hand $q_1 \nmid N$, write $q_1 = r_1 s_1$ with $r_1 | N$ and $\gcd(s_1, N) = 1$ and set $t_1 := N/r_1$. Then for each $1 \le k_1 \le N$ holds

$$\left|\frac{p_1}{q_1} - \frac{k_1}{N}\right| = \left|\frac{p_1 t_1 - k_1 s_1}{N s_1}\right| \ge \frac{1}{N s_1}.$$

This means that no matter how many k_1 of the y_i are placed at x_1 we have

$$\big|\mu\big([0,z)\big) - \nu_N\big([0,z)\big)\big| \ge 1/N \quad \text{for} \quad x_1 < z < x_2.$$

This is essentially the second claim of Theorem 1.2 and also explains entirely, what happens in [2, Example 3.1], where $\xi_1 = \xi_2 = \frac{1}{2}$.

If $\xi_1,\ \xi_2 \notin \mathbb{Q}$, we again fix $N \in \mathbb{N}$ and only place weight at x_1 and x_2. There exists exactly one

$$p(N) \in \mathbb{N}$$

with

$$\big|\mu\big([0,z)\big) - \nu_N\big([0,z)\big)\big| = |\xi_1 - p(N)/N| \le 1/2N \quad \text{for} \quad x_1 < z < x_2.$$

Multiplying the inequality by N results in

$$|N\xi_1 - p(N)| \le 1/2.$$

Hence, for arbitrary N the best possible error term is

$$\|N\xi_1\|, \quad \text{where} \quad \|x\| := \min(|x|, 1-|x|).$$

This means that the lower bound is governed by the distance of the Kronecker sequence elements $N\xi_1$ from the closest integer. As the Kronecker sequence is a uniformly distributed sequence, see, e.g. [6, Theorem 3.3], we thus have for every $c > 2$ that

$$D_N^*(\mu, \nu_N) \ge \frac{1}{cN},$$

because $\|N\xi_1\| \ge \frac{1}{c}$ holds for infinitely many $N \in \mathbb{N}$. This is claim (iii) of Theorem 1.2. On the other hand, uniform distribution of the Kronecker sequence also implies

$$D_N^*(\mu, \nu_N) \le \frac{1}{cN}$$

for infinitely many $N \in \mathbb{N}$, because $\|N\xi_1\| \le \frac{1}{c}$ holds for infinitely many $N \in \mathbb{N}$, too. Thus, the approximation error

$$\big|\mu\big([0,z)\big) - \nu_N\big([0,z)\big)\big| \quad \text{for} \quad x_1 < z < x_2$$

can get arbitrarily small as well.

In fact, the question of the best possible approximation of the measure μ is in this particular case closely related to the Diophantine properties of ξ_1 and more precise statements than in our example can be made by using methods from the respective field of research, see, e.g. [5].

Example 1 also shows that the bounds in Theorem 1.2 are sharp. In fact corresponding examples which proof the sharpness of our bounds can be constructed in the same manner for arbitrary $n \in \mathbb{N}$ (and also for infinitely supported discrete measures).

Finally, we compare Theorem 1.2 to classical results from Diophantine approximation: If the measure is supported on finitely many points $x_1, \ldots, x_n$ and the weights ξ_i are linearly independent (over $\mathbb{Q}$), irrational, algebraic numbers, then Schmidt's subspace Theorem, see [7], can be applied and yields an error term of at least

$$N^{-(1+1/n+\epsilon)} \quad \text{for all} \quad N \geq N_0.$$

If N is big enough, this implies a lower bound for the star-discrepancy of order $N^{-(1+1/n+\epsilon)}$ which is worse than what we obtain. On the contrary, the simultaneous Dirichlet theorem implies that it is possible to find infinitely N such that each individual point mass is approximated of order $\leq \frac{1}{N^{1+1/n}}$. This is worse than the result $1/N$ from Theorem 1.1 and thus does not impose an obstacle for what follows.

2. Proof of the Main Result

This section is dedicated to the proof of parts (ii) and (iii) in Theorem 1.2. In order to show (ii), let μ be an arbitrary discrete measure and let again $\|\cdot\|$ denote the distance from the nearest integer. Choose an arbitrary $x \in [0,1)$ with $\mu(\{x\}) \notin \{0,1\}$. Then since $N\nu_N(A)$ is an integer for every interval A we get

$$\begin{aligned}\|N\mu(\{x\})\| &= \lim_{y \searrow x} \big\|N\mu\big([0,x)\big) - N\mu\big([0,y)\big)\big\| \\ &= \lim_{y \searrow x} \big\|N\mu\big([0,x)\big) - N\nu_N\big([0,x)\big) - N\mu\big([0,y)\big) + N\nu_N\big([0,y)\big)\big\| \\ &\leq 2ND_N^*(\mu, \nu_N).\end{aligned}$$

The left hand side is $\geq 1/3$ for infinitely many $N \in \mathbb{N}$ no matter if $\mu(\{x\})$ is rational or irrational. This finishes the proof of (ii).

Hence we come to the proof of part (iii). At first, we consider finite discrete measures with (possibly) irrational weights ξ_i. In comparison to Example 1, the situation can be treated as follows: at first, we take the best individual approximation of $\{N\xi_1\}$. Then we inductively define $p_l(N)$ by

$$\left| \sum_{j=1}^{l} (\xi_j - p_j(N)/N) \right| < 1/2N.$$

As we assume $x_1 < x_2 < \cdots < x_n$, the discrepancy cannot be diminished. Since $\sum_{i=1}^{l} \xi_i \notin \mathbb{Q}$ and the one-dimensional Kronecker sequence $\left\{N \sum_{i=1}^{l} \xi_i\right\}$ is uniformly distributed in $[0,1]$ (see, e.g. [4, Section 2.3]), also

$$\left\| \left\{ N \sum_{i=1}^{l} \xi_i \right\} \right\|$$

is uniformly distributed in $[0, 1/2]$. Hence, for any $c > 2$, there exist infinitely many N with $D_N(\mu, \nu_N) \geq \frac{1}{cN}$ for every finite point set $y_1, \ldots, y_N$ (this is similar to the case for $n = 2$ points from Example 1).

Finally, we consider the infinite case and make use of what we have already proven for finitely supported measures. Let $N \in \mathbb{N}$ and choose z_1 as the supremum over all

$$x \in [0,1] \quad \text{with} \quad \mu([0,x]) \leq 1/2N.$$

At first let us assume that $\mu(\{z_1\}) = 0$. If we put weight of ν_N at z_1 (or any point smaller than z_1), then $D_N^*(\mu, \nu_N) \geq 1/2N$ because the minimum weight we can choose is $1/N$. If we do not put weight of ν_N at z_1, then the interval $[0, z_1]$ does not contain a point and we also have

$$D_N^*(\mu, \nu_N) \geq 1/2N.$$

Therefore, we obtain the desired lower bound in this case (which is not surprising because it very much resembles the continuous case). So the remaining case to solve is

$$\mu(\{z_1\}) > 0.$$

Again putting weight of ν_N left of z_1 would result in

$$D_N^*(\mu, \nu_N) \geq 1/2N.$$

Let $\tilde{c} > 0$ be arbitrary. As long as $0 < \mu(\{z_1\}) < 1/\tilde{c}N$ and we put weight at z_1 we have

$$D_N^*(\mu, \nu_N) \geq (1/2 - 1/\tilde{c})N.$$

Now choose an arbitrary $N^* > N$ big enough such that $\mu(\{z_1\}) > 1/2N^*$ and consider the measure $\mu([0, z_1]) = \xi$. Since we have to put weight of ν_N on z_1 (because otherwise $D_{N^*}^*(\mu, \nu_{N^*}) > 1/2N^*$ by the assumption on z_1 and we would be done), we cannot approximate it any better than $\|N^*\xi\|$ and we can apply the result for the irrational finite case. Thus, claim (iii) follows.

Acknowledgement. Parts of the research on this paper was conducted during a stay at the Max-Planck-Institute for Mathematics in Bonn. The author thanks the MPI for hospitality. Furthermore, the author would like to thank the referee for their useful comments.

REFERENCES

[1] AISTLEITNER, C.—BILYK, D.—NIKOLOV, A.: *Tusnády´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] FAIRCHILD, S.—GOERING, M.—WEISS C.: *Families of well approximable measures*, Unif. Distrib. Theory, **16** (2021), no. 1, 53–70.

[3] HEWITT, E.—STROMBERG, H.: *Real and Abstract Analysis: A Modern Treatment of the Theory of Functions of a Real Variable*. In: *Graduate Texts in Math. Vol. 25*, Springer-Verlag, Berlin, 1975.

[4] KUIPERS, L.—NIEDERREITER. H.: *Uniform Distribution of Sequences*. In: *Pure and Applied Mathematics*. Wiley-Interscience [John Wiley & Sons], New York, 1974.

[5] LANG, S.: *Introduction to Diophantine Approximations*. (Second ed.) Springer-Verlag, Berlin, 1995.

[6] NIEDERREITER, H.: *Random Number Generation and Quasi-Monte Carlo Methods*. In: CBMS-NSF *Series in Appl. Math. Vol. 63*, SIAM, Philadelphia, PA, 1992.

[7] SCHMIDT, W. M.: *Norm form equations*, Ann. Math. **96** (1972), no. 3, 526–551.

Received July 7, 2022
Accepted February 9, 2023

Christian Weiß
Department of Natural Sciences
Ruhr West University of Applied Sciences
Duisburger Str. 100
45479 Mülheim a. d. Ruhr
GERMANY
E-mail: christian.weiss@hs-ruhrwest.de