Uniform Distribution Theory **6** (2011), no.2, 85–94

uniform distribution theory

A NOTE ON IMPLICITLY DEFINED SETS IN UNIFORM DISTRIBUTION THEORY

Stefan Steinerberger

ABSTRACT. Motivated by some matching results in terms of extreme discrepancy, we describe a packing problem in $[0,1]^d$ with an interpretation in logistics, determine its asymptotic behavior, show its minimizers to be asymptotically uniformly distributed and raise several questions.

Communicated by Michael Drmota

1. Introduction

This short note aims to be a natural outgrowth of some results presented at the 2. International Conference on Uniform Distribution Theory in 2010; these results can be found elsewhere [8]. Except for one proof, which is given fully in [8], this note is self-contained and deals with a packing problem.

As a starting point we consider a result of Steele [5], which gives a lower bound on the shortest traveling salesman path in terms of discrepancy.

Theorem 1 (Steele, 1980). *Let $(y_n)_{n=1}^{\infty}$ be a uniformly distributed sequence in $[0,1]^d$ and let $TSP(y_1,\dots,y_N)$ be the length of the shortest traveling salesman path through $\{y_1,\dots,y_N\}$. Then, for N sufficiently large,*

$$TSP(y_1,\dots,y_N) \geq \frac{2}{3}\frac{1}{6^{d-1}} D_N^{-(d-1)/d},$$

where D_N denotes the extreme discrepancy.

Steele's lower bound was improved by the author [6] by means of a short geometric argument to

$$\mathrm{TSP}(y_1,\dots,y_N) \geq \frac{\sqrt{3}}{4} D_N^{-(d-1)/d},$$

2010 Mathematics Subject Classification: 11K06, 11K38.
Keywords: Optimal matching, Voronoi cells.

where the constant $\sqrt{3}/4$ holds for dimensions $d \geq 2$. This constant is certainly not best possible, can it be improved?

A converse result of Snyder and Steele [4], improving their earlier result [3], implies that the sets with the longest traveling salesman paths are asymptotically uniformly distributed as the number of points goes to infinity. Is it possible to give nontrivial estimates on the discrepancy of these sets? As is clear from the origin of the problem, there is little hope in actually computing these sets for an even moderate number of points. One could, however, try to find other functionals, whose minimizers might have interesting distribution properties and where actual numerical computations are not completely out of reach.

2. A problem in combinatorial geometry...

We nominate the following functional: for a finite set of points $Y \subset [0,1]^d$, consider the average expected distance between a randomly chosen point in $[0,1]^d$ and the closest point from Y. Using $\|\cdot\|$ to denote the Euclidean norm, this can also be written as

$$\int_{[0,1]^d} \inf_{y \in Y} \|x - y\| dx.$$

Problem. *For fixed $N, d \in \mathbb{N}$, how can one find the set $Y = \{y_1, \ldots, y_N\} \subset [0,1]^d$ minimizing this functional? What can be said about the numerical value*

$$\inf \left\{ \int_{[0,1]^d} \inf_{y \in Y} \|x - y\| dx : Y \subset [0,1]^d, \#Y = N \right\}?$$

Our problem can also be motivated by the following real-life example: How should a restaurant chain distribute their stores over a city to minimize the average length a hungry customer has to walk to reach the nearest store (assuming people get hungry at random and are evenly distributed over the city)?

Besides its applicability, the problem is also interesting from a theoretical point of view, since we have to balance out two opposing factors: the average distance to the center of a convex body is minimized for a sphere and is quite large for sausage-like bodies and thus in some sense a measure for the degree in which the body differs from a sphere. On the other hand, it is not possible to decompose the unit cube $[0,1]^d$ into finitely many spheres, i.e. we already lose some optimality there. Furthermore we are quite restricted in the choice of decomposition itself, which is necessarily the Voronoi decomposition induced by finite set of points. The question thus is: How sphere-like can an average cell of

Voronoi decompositions possibly be? In the one-dimensional case, this problem is not hard to solve and we immediately get by direct computation the following result.

Proposition. *Let $Y = \{y_1, \dots, y_N\} \subset [0,1]$ ordered in increasing size. Then*

$$\int_0^1 \inf_{y \in Y} \|x - y\| dx = \frac{1}{2} y_1^2 + \frac{1}{4} \sum_{i=1}^{N-1} (y_{i+1} - y_i)^2 + \frac{1}{2}(1 - y_N)^2 \geq \frac{1}{4N}.$$

Furthermore, the lower bound is tight and the unique minimizer is given by

$$Y = \left\{ \frac{1}{2N}, \frac{3}{2N}, \dots, \frac{2N-1}{2N} \right\}.$$

This sharp result is not surprising at all (it states that the real line can be covered with rescaled translates of the unit interval). Note that the minimizer also minimizes the star discrepancy D_N^*. The situation is rather different in higher dimensions, where we have the following result.

Theorem 2. *Let $d \geq 2$. Then, for all $Y = \{y_1, \dots, y_N\} \subset [0,1]^d$,*

$$\int_{[0,1]^d} \inf_{y \in Y} \|x - y\| dx \geq \frac{d}{d+1} \frac{1}{\lambda(\mathbb{B})^{1/d}} \frac{1}{N^{1/d}}.$$

Furthermore, there exists $Y = \{y_1, \dots, y_N\} \subset [0,1]^d$ with

$$\int_{[0,1]^d} \inf_{y \in Y} \|x - y\| dx \leq \frac{\sqrt{d}}{\sqrt{12}} \frac{1}{N^{1/d}} + o\left(\frac{1}{N^{1/d}} \right).$$

A proof of this is given in full detail in [6]. The idea for the lower bound is based on exploiting the view of the problem in terms of Voronoi decompositions and considerations involving convexity, the upper bound is merely an explicit computation for a cube-like arrangement.

Asymptotically speaking, this result is quite nice, since the constants of lower and upper bound grow like

$$\frac{d}{d+1} \frac{1}{\lambda(\mathbb{B})^{1/d}} \sim 0.2419\sqrt{d} \quad \text{and} \quad \frac{1}{\sqrt{12}}\sqrt{d} \sim 0.2886\sqrt{d}$$

but it is likely that other constructions would yield even better results. It should be remarked that these bounds can be improved through accurate calculations in low dimensions. If $d = 2$, then any computer algebra system will show that

$$\int_0^1 \int_0^1 \sqrt{\left(x - \frac{1}{2}\right)^2 + \left(y - \frac{1}{2}\right)^2} dx dy = \frac{1}{6}\left(\sqrt{2} + \operatorname{arcsinh} 1\right)$$

and thus we get that in the two-dimensional case for square packings Y_N, we get

$$0.3761\dots = \frac{2}{3\sqrt{\pi}} \le \lim_{N\to\infty} \|\{X\} \to Y_N\| \sqrt{N} = \frac{1}{6}\left(\sqrt{2} + \operatorname{arcsinh} 1\right) = 0.3826\dots$$

It is interesting that even computer algebra systems get problems with these classes of integrals very quickly and beyond dimensions $d \ge 5$, no precise results are known (these problems have been studied by Bailey, Borwein and Crandall in two extensive papers [1] and [2], some asymptotic results can be found in [7]).

3. ...and its asymptotic distribution behavior

In this section we will show that the discrepancy of the minimizers of the above functional necessarily tends to 0. The proof is interesting insofar as it neither relies on counting nor on any exponential sums and does not tell us a whole lot about the minimizers.
We also have no estimates on the discrepancy, which could be interesting, especially considering the one-dimensional case where the unique solution minimizes the star discrepancy. We start with the asymptotic behavior and consider a bounded 'reasonably nice' (i.e. for example products of intervals) set $A \subset \mathbb{R}^d$ and study the above problem in this general setting with the functional

$$\int_A \inf_{y\in Y} \|x - y\| dx.$$

We use Y_N to denote a $N-$element minimizer for the functional and a_N to denote the minimal value

$$a_N = \frac{1}{\lambda(A)} \int_A \inf_{y\in Y_N} \|x - y\| dx.$$

It is perhaps not too surprising that the problem is essentially localized and the asymptotic behavior of the functional depends only on its size but in no way on its form (because for these types of problems any boundary effects asymptotically vanish provided the boundary is sufficiently nice).

Theorem 3. *There exists a universal constant $c_d > 0$, depending only on the dimension, such that for any hyperrectangle A*

$$\lim_{N\to\infty} a_N N^{1/d} = c_d \lambda(A)^{1/d},$$

where $\lambda(A)$ denotes the volume of A. Furthermore, $a_N N^{1/d} \ge c_d \lambda(A)^{1/d}$ for every N.

P r o o f. Standard arguments show that it suffices to prove the result for $A = [0,1]^d$ (scaling reduces to the case $\lambda(A) = 1$ and scaled hyperrectangles cover the cube and vice versa). This will be done by showing that $a_N N^{1/d}$ is bounded from below (this we already know from the lower bound of Theorem 1) and satisfies

$$\limsup_{M\to\infty} a_M M^{1/d} \le a_N N^{1/d}$$

for every $N \in \mathbb{N}$. For this, consider an arbitrary number $k \in \mathbb{N}$ and use

$$a_{k^d N}(k^d N)^{1/d} \le a_N N^{1/d}$$

which follows immediately from rescaling Y_N by a factor k and placing k^d copies of it into $[0,1]^d$. In general, one will expect some interaction between the cells (i.e. points at the border of one cell, which get actually matched to a point of an adjacent cell) and therefore one will expect even $a_{k^d N}(k^d N)^{1/d} < a_N N^{1/d}$. Using the above inequality in combination with the trivial inequality

$$a_{N+M}(N+M)^{1/d} \le \left(1 + \frac{M}{N}\right)^{1/d} a_N N^{1/d}$$

yields the result. □

Before using this result to study the asymptotic distribution properties, we derive an additional very weak result concerning the *dispersion* of Y_N, which we use later on.

Estimate. There exists a constant c depending only on the dimension such that

$$\sup_{x\in[0,1]} \inf_{y\in Y_N} \|x-y\| \le cN^{-\frac{1}{d(d+1)}}.$$

P r o o f. c will denote a constant, changing from line to line and depending only on the dimension. It is easy to see that there is a constant c with $a_N \le cN^{-1/d}$. Pick a point x such that the distance to the closest point from Y_N becomes maximal and suppose this distance to be $r > 0$. Then we can consider the set

$$B = \left\{ z \in [0,1]^d : |x - z| \le \frac{r}{2} \right\}.$$

There exists a constant $c > 0$ with $\lambda(B) > cr^d$. Furthermore, by definition of x, the distance from any point of B to the closest point of Y_N is at least $r/2$. Thus there is a constant $c > 0$ with

$$a_N = \int_{[0,1]^d} \inf_{y\in Y_N} \|z-y\| dz \ge \int_B \inf_{y\in Y_N} \|z-y\| dz \ge \int_B \frac{r}{2} dz \ge cr^{d+1}.$$

□

This result is extremely weak but sufficient for our purpose (we only need the dispersion to go to 0); it seems reasonable to conjecture that for an optimal set Y_N all induced Voronoi cells have roughly the same measure with none of them being too eccentric in shape, which would suggest that maybe

$$\sup_{x\in[0,1]} \inf_{y\in Y_N} \|x-y\| \leq cN^{-\frac{1}{d}}$$

but this could turn out to be a difficult problem (the statement is trivial for grid-like structures but showing that an extremal structure is necessarily grid-like seems more challenging). These results can now be combined to prove the following result.

Corollary. *We have*

$$\lim_{N\to\infty} D_N(Y_N) = 0.$$

P r o o f. Proof by contradiction. We assume the discrepancy does not tend to 0; then there exists a hyperrectangle, which has infinitely many times either too many or too few points. We look at this subsequence, decompose the problem into two problems by removing a critical area around the hyperrectangle, whose volume tends to 0 and which will be shown to be harmless with respect to removing it. For both of the new problems we do have a lower bound given by the theorem concerning the asymptotic behavior and these facts combined will yield a contradiction.

Setting up. Suppose the statement is false. Then there exists $\varepsilon > 0$, and sequences (N_k) of integers and sequences of hyperrectangles (R_k) such that

$$\left|\frac{\#Y_{N_k} \cap R_k}{N_k} - \lambda(R_k)\right| > \varepsilon.$$

The sequence of rectangles can be interpreted as a sequence of points in $[0,1]^{2d}$ and thus has a convergent subsequence and hence there exists $\delta > 0$ and a sequence (N_k) of integers and a fixed hyperrectangle $R \subset [0,1]^d$ with

$$\left|\frac{\#Y_{N_k} \cap R}{N_k} - \lambda(R)\right| > \delta.$$

We will from now on assume that $\#Y_{N_k} \cap R \geq (\lambda(R) + \delta)N_k$, the other case can be dealt with in precisely the same fashion. Let us now decompose $[0,1]^d = R \cup ([0,1]^d \setminus R)$ and write

$$a_{N_k} = \int_{[0,1]^d} \inf_{y\in Y_{N_k}} \|z-y\| dz = \int_R \inf_{y\in Y_{N_k}} \|z-y\| dz + \int_{[0,1]^d\setminus R} \inf_{y\in Y_{N_k}} \|z-y\| dz.$$

Decomposing the problem. Now, we wish to decompose the problem into two seperate problems; for this, we consider the problem area

$$P_{N_k} = \left\{ z \in [0,1]^d : \inf_{w \in \partial R} \|z - w\| \le \sup_{x \in [0,1]} \inf_{y \in Y_{N_k}} \|x - y\| \right\},$$

which has the property that for any point in $R \setminus P_{N_k}$, the closest point from Y_{N_k} lies in R and, at the same time, for any point in $[0,1]^d \setminus R \setminus P_{N_k}$, the closest point from Y_{N_k} lies in $[0,1]^d \setminus R$ and thus we have eliminated any interaction between the sets. As a consequence of the above estimate,

$$\lambda_d(P_{N_k}) \le c N_k^{-\frac{1}{d(d+1)}} \lambda_{d-1}(\partial R)$$

for some constant c. We now wish to show that there aren't too many points within the critical area, i.e.

$$\lim_{N \to \infty} \frac{\# Y_{N_k} \cap P_{N_k}}{N_k} = 0.$$

Suppose this is not the case and there is a further subsequence N_{k_l} with at least ηN_{k_l} points in $P_{N_{k_l}}$ for some $\eta > 0$. Then we build another set, the problem set of the problem set $PP_{N_{k_l}}$, which is given by

$$PP_{N_{k_l}} = \left\{ z \in [0,1]^d : \inf_{w \in P_{N_{k_l}}} \|z - w\| < \sup_{x \in [0,1]} \inf_{y \in Y_{N_{k_l}}} \|x - y\| \right\}.$$

This set has the same purpose as before and leaves us with a remaining set $[0,1]^d \setminus PP_{N_{k_l}}$. Trivially,

$$a_{N_{k_l}} \ge \int_{[0,1]^d \setminus PP_{N_{k_l}}} \inf_{y \in Y_{N_{k_l}} \cap ([0,1]^d \setminus P_{N_{k_l}})} \|z - y\| dz.$$

At the same time, however, we do have a lower bound given by the asymptotic law and thus, because at least ηN_{k_l} points have been removed,

$$\int_{[0,1]^d \setminus PP_{N_{k_l}}} \inf_{y \in Y_{N_{k_l}} \cap ([0,1]^d \setminus P_{N_{k_l}})} \|z - y\| dz \ge c_d \frac{\lambda([0,1]^d \setminus PP_{N_{k_l}})}{((1-\eta) N_{k_l})^{1/d}}.$$

However, for $l \to \infty$, the volume $\lambda([0,1]^d \setminus PP_{N_{k_l}})$ converges to 1 and this contradicts the limiting law.

Concluding the argument. We write, for short, $f(N_k) = \# Y_{N_k} \cap P_{N_k} \cap R$ and $g(N_k) = \# Y_{N_k} \cap P_{N_k} \cap ([0,1]^d \setminus R)$; we have seen that $f(N_k) + g(N_k) = o(N_k)$. The argument can now be concluded by considering three possibilities. We aim to calculate the distance of a random variable to the nearest point of a set;

either the random variables lands in $[0,1]^d \setminus R \setminus P_{N_k}$ or in $R \setminus P_{N_k}$ in which case they are mapped to the closest element of the set within $[0,1]^d \setminus R$ or R, respectively, or it lands in P_{N_k}, in which case we estimate the expected distance to the closest element from below by 0. If we now abuse notation by choosing δ_k so that $(\lambda(R) + \delta_k)N$ is the actual number of points within R (and thus $\delta_k \geq \delta$ by assumption), then

$$\begin{aligned} a_{N_k} \geq{} & c_d \lambda(R \setminus P_{N_k}) \frac{\lambda(R \setminus P_{N_k})^{1/d}}{((\lambda(R) + \delta_k)N_k - f(N_k))^{1/d}} \\ & + c_d \lambda([0,1] \setminus R \setminus P_{N_k}) \frac{\lambda([0,1] \setminus R \setminus P_{N_k})^{1/d}}{((1 - \lambda(R) - \delta_k)N_k - g(N_k))^{1/d}}. \end{aligned}$$

Multiplying with $N_k^{1/d}$ and letting $k \to \infty$ implies for the LHS that

$$c_d = \lim_{k \to \infty} a_{N_k} N_k^{1/d},$$

whereas the RHS need not converge (since δ_k) may not). However, for large values of k it simplifies (ignoring lower order terms f, g and using the fact that the measure of the problem set goes to 0) to

$$c_d \left(\lambda(R) \frac{\lambda(R)^{1/d}}{(\lambda(R) + \delta_k)^{1/d}} + (1 - \lambda(R)) \frac{(1 - \lambda(R))^{1/d}}{(1 - \lambda(R) - \delta_k)^{1/d}} \right).$$

The expression in the brackets, however, has a particular structure and we may use the fact that for $0 < x, y < 1$ and $d \geq 1$ always

$$x \left(\frac{x}{y} \right)^{1/d} + (1 - x) \left(\frac{1 - x}{1 - y} \right)^{1/d} \geq 1$$

with equality if and only if $x = y$ to derive a contradiction since equality is ruled out by $\delta_k \geq \delta > 0$. □

Concluding remarks. The classical way towards uniformly distributed sequences is constructing a sequence and using either counting techniques or exponential sums to get some ideas about its discrepancy; the sequence is already given, one has to determine its structure. Another more subtle way is given, for example, by $(t, m, s)-$nets: one constructs an abstract structure and uses the underlying properties to determine the discrepancy. Here, the difficulty lies not in controlling the discrepancy but rather in finding the set. The approach sketched above belongs to neither: we determined asymptotic uniform distribution by using only the fact that these sets minimize a certain functional and have very little idea how such a set actually looks like (while, at the same time, the functional does not immediately imply discrepancy bounds). If one could get results about the convergence speed of $a_N N^{1/d}$, this would effectively imply bounds on the discrepancy along the same lines of thought as above.

It is not inconceivable that minimizers of certain functionals might actually produce sets of wonderful distribution qualities: trivially, minimizing the discrepancy D_N over all $N-$element point sets will do this but computing the discrepancy is hard; however, there might exist an effectively computable simple functional, which naturally encodes the idea of uniform distribution and is simple to approximately minimize, at least using a computer (the above functional, minimizing the average distance to the closest point, can certainly be said to describe some sort of notion of uniform distribution but we make no predictions about it being easy to compute/minimize).

Finally, we believe the problem of finding an $N-$element set in $[0,1]^d$ minimizing the average distance of a random variable to its closest point to be of independent interest, giving rise to many questions regarding

- asymptotic behavior: what is the numerical value of c_d? what about convergence speed?
- structure: what can be said about the structure of the minimizers? how do the convex bodies resulting from the induced Voronoi decomposition behave?
- computability: how effectively can the minimizers be computed or approximated?
- applications: can this set be profitably used (excluding the optimal distribution of restaurant chains)?

REFERENCES

[1] D. Bailey, J. Borwein and R. Crandall, *Box integrals*, J. Comput. Appl. Math. **206** (2007), 196–208.

[2] D. Bailey, J. Borwein and R. Crandall, *Advances in the Theory of Box Integrals*, Math. Comp. **79** (2010), 1839–1866.

[3] T. Snyder and J. Steele, *Equidistribution of point sets for the traveling salesman and related problems*, Proceedings of the Fourth Annual ACM-SIAM Symposium on Discrete Algorithms (1993), 462–466.

[4] T. Snyder and J. Steele, *Equidistribution in all Dimensions of Worst-case Point Sets for the Travelling Salesman Problem*, SIAM J. Discrete Mathematics **8** (1995), 678–683.

[5] J. Steele, *Shortest paths through pseudo-random points in the $d-$cube*, Proc. Amer. Math. Soc. **80** (1980), 130–134.

[6] S. Steinerberger, *A new lower bound for the geometric traveling salesman problem in terms of discrepancy*, Operations Research Letters **38** (2010), 318–319.

[7] S. Steinerberger, *Extremal Uniform Distribution and Random Chord Lengths*, Acta Mathematica Hungarica **130** (2011), 321–339.
[8] S. Steinerberger, *Random restricted matching and lower bounds for combinatorial optimization*, Journal of Combinatorial Optimization, accepted.
[9] M. Talagrand, *The transportation cost from the uniform measure to the empirical measure in dimension* ≥ 3, Ann. Probab. **22** (1994), 919–959.

Received January 31, 2011
Accepted May 9, 2011

Stefan Steinerberger
Mathematisches Institut
Universität Bonn
Endenicher Allee 60
53115 Bonn
Germany
E-mail: stefan.steinerberger@gmail.com