Prague Summer School: Discrepancy Theory Summer 2018

Lecture 1

Aleksandar Nikolov

1 Uniform Sets of Points

We start with one of the classical geometric discrepancy questions: "What is the most uniform set of n points in the unit square $[0,1)^2$?" This is, of course, not a well-posed question and we need to make it more presice before we can give a definite answer.

Let us start from the analogous continuous question: "What is the uniform *probability measure* on the unit square $[0,1)^2$?" In a certain sense, the unique answer to this question is the Lebesgue measure λ^2 restricted to the square, i.e. the measure that assigns every measurable set $S \subseteq [0,1)^2$ its area. For instance, one natural property to ask of a "uniform" measure is that it is translation invariant. Since a translation may take us out of the square, let us look instead at translation with wrap-around, i.e. a translation by vector $u \in \mathbb{R}^2$ maps a point $x \in [0,1)^2$ to $((x_1 + u_1) \bmod 1, (x_2 + u_2) \bmod 1)$, i.e. to the point whose coordinates are the fractional parts of the coordinates of $x + u$. Then the Lebesgue measure is the unique probability measure which is invariant under this notion of translation.

Having decided on the meaning of a uniform probability measure on the square, we are going to define a set of points as uniform if the uniform measure on the points approximates well the Lebesgue measure. In particular, let $P \subset [0,1)^2$ be a set of n points, and let $S \subseteq [0,1)^2$ be some measurable set. The probability that a uniformly random point in P lands in S is $\frac{|P \cap S|}{n}$. The discrepancy of P with respect to S is the difference between this probability and the area of S, denoted

$$D(P,S) = \frac{|P \cap S|}{|P|} - \lambda^2(S).$$

The idea here is that S is a "distinghishing set": if we sample many points from P and they land in S with probability very different from $\lambda^2(S)$, then we know for sure that the points were not sampled uniformly from the square. We are usually interested in a the maximum discrepancy over a whole family of distinguishing sets. To this end, for a collection $\mathcal{S}$ of measurable subsets of $[0,1)^2$, we define the discrepancy of an n-point set P as

$$D(P,\mathcal{S}) = \sup_{S \in \mathcal{S}} |D(P,S)|.$$

If the discrepancy $D(P,\mathcal{S})$ is small, then we cannot easily distinguish the uniform measure on P from the uniform measure on $[0,1)^2$ by sampling and computing how many points land in some set $\mathcal{S}$. Once we fix a family $\mathcal{S}$, the most uniform set of n points in the square (with respect to $\mathcal{S}$) is the one that minimizes the discrepancy, i.e. the one that achieves the discrepancy of $\mathcal{S}$, defined for every n as

$$D(n,\mathcal{S}) = \inf_{P \subset [0,1)^2, |P| = n} D(P,\mathcal{S}).$$

The best studied family of distinguishers is the set of axis aligned rectangles $\mathcal{R}_2$, i.e. all sets of the form $[x_1, y_1] \times [x_2, y_2]$ for $x, y \in [0,1)^2$. We will focus on this family for purposes of illustration.

Exercise 1. *Show that the grid $P = \{(i/\sqrt{n}, j/\sqrt{n})\}$ where $i = 0, \ldots, \sqrt{n}-1$ and $j = 0, \ldots, \sqrt{n}-1$ (assume $\sqrt{n}$ is an integer) has discrepancy $D(P, \mathcal{R}_2) = \Theta(n^{-1/2})$.*

Exercise 2. *Show that $D(n, P) = \Omega(n^{-1})$.*

There are various constructions that give the bound $D(n, \mathcal{R}_2) = O\left(\frac{\log n}{n}\right)$, which is tight up to constants [4]. We give one construction below.

Note that there was nothing special about two dimensions in the discussion above. We can instead look for uniformly distributed sets of points in the unit d-dimensional cube $[0,1)^d$. The unique uniform measure is the (d-dimensional) Lebesgue measure λ^d, i.e. the measure that associates a set $S \subseteq [0,1)^d$ with its d-dimensional volume. The definitions are analogous to the ones above: for an n-point set $P \subset [0,1)^d$, a distinguishing set $S \subseteq [0,1)^d$, and a family $\mathcal{S}$ of distinguishing sets, we define

$$\begin{aligned} D(P, S) &= \frac{|P \cap S|}{|P|} - \lambda^d(S); \\ D(P, \mathcal{S}) &= \sup_{S \in \mathcal{S}} |D(P, S)|; \\ D(n, \mathcal{S}) &= \inf_{P \subset [0,1)^d, |P| = n} D(P, \mathcal{S}). \end{aligned}$$

In the d-dimensional setting we will focus on the set of d-dimensional axis-aligned boxes $\mathcal{R}_d$, i.e. sets of the form $[x_1, y_1] \times \ldots \times [x_d, y_d]$ for $x, y \in [0,1)^d$. Here, the best known lower bound on $D(n, \mathcal{R}_d)$ is $D(n, \mathcal{R}_d) = \Omega\left(\frac{(\log n)^{(d-1)/2+\eta_d}}{n}\right)$ for some η_d that goes to 0 as d goes to infinity [2] (see also the excellent survey [1]). The best known upper bound is $D(n, \mathcal{R}_d) = \Omega\left(\frac{(\log n)^{d-1}}{n}\right)$. Closing this significant gap is known as the "Great Open Problem" in geometric discrepancy theory.

2 Combinatorial Discrepancy

We now consider *combinatorial discrepancy*, which is an interesting measure in and of itself (and is also related to the usual continuous discrepancy, as we will see). Let $\mathcal{U}$ be a set and let $\mathcal{S} \subseteq 2^{\mathcal{U}}$ a family of subsets of $\mathcal{U}$; together the pair $(\mathcal{U}, \mathcal{S})$ is called a *set system*. Given a coloring $\chi : \mathcal{U} \to \{-1, 1\}$ of the elements of $\mathcal{U}$ with ± 1 we define

$$\operatorname{disc}(\chi, \mathcal{S}) = \max_{S \in \mathcal{S}} |\chi(S)|$$

where $\chi(S) = \sum_{u \in S} \chi(u)$. Intuitively, this measures how balanced χ is. A coloring χ has small discrepancy if it colors about half the the elements of every set $S \in \mathcal{S}$ with -1 and about half with $+1$. The *discrepancy* of $\mathcal{S}$ is then defiend as

$$\operatorname{disc}(\mathcal{S}) = \min_{\chi} \operatorname{disc}(\chi, \mathcal{S})$$

where the minimum is taken over all colorings $\chi : \mathcal{U} \to \{-1, 1\}$.

Notice that if $A \in \{0,1\}^{\mathcal{S} \times U}$ is the incidence matrix of $\mathcal{S}$, i.e. the matrix defined by

$$a_{S,u} = \begin{cases} 1 & u \in S \\ 0 & u \notin S \end{cases},$$

then $\mathrm{disc}(S) = \min_{x\in\{-1,1\}^U} \|Ax\|_\infty$, where $\|\cdot\|_\infty$ is the infinity norm defined by $\|y\|_\infty = \max_i |y_i|$ with the maximum taken over the coordinates of y. We can now define the discrepancy $\mathrm{disc}(A) = \min_{x\in\{-1,1\}^n} \|Ax\|_\infty$ for any $m \times n$ matrix A.

It will be important for us to consider a more robust version of discrepancy, known as *hereditary discrepancy*. The hereditary discrepancy is defined for an $m \times n$ matrix A by $\mathrm{herdisc}(A) = \max_S \mathrm{disc}(A_S)$ where the maximum is over all subsets S of $[n] = \{1, \dots, n\}$, and A_S is the submatrix of A consisting of the columns indexed by S. The hereditary $\mathrm{herdisc}(\mathcal{S})$ of a set system $(\mathcal{S}, U)$ can be defined as $\mathrm{herdisc}(A)$, where A is the incidence matrix of $\mathcal{S}$ defined above, and equals the maximum discrepancy over restricted set systems $\mathcal{S}|_W = \{S \cap W : S \in \mathcal{S}\}$, with W ranging over subsets of the ground set U.

We have the following important lemma, due to Lovasz, Spencer, and Vesztergombi [3].

Lemma 1. *For any $m \times n$ matrix A, and any $w \in [0,1]^n$, there exists a $x \in \{0,1\}^n$ such that $\|Ax - Aw\|_\infty \le \mathrm{herdisc}(A)$.*

Proof. First an easy exercise.

Exercise 3. *Show that for any $w \in \{0, \frac{1}{2}, 1\}^n$ there exists an $x \in \{0,1\}^n$ such that $\|Ax - Aw\|_\infty \le \frac{1}{2}\mathrm{herdisc}(A)$.*

To finish the proof of the lemma, it suffices to prove it for w such that w_i has a finite binary expansion for each i. We do so by induction. Let us assume that the lemma is proved for all $w \in 2^{-k}\mathbb{Z}^n \cap [0,1)^n$ for some $k \ge 1$. (Note that the case $k = 1$ is the exercise above.) We will show that it then also holds for every $w \in 2^{-k-1}\mathbb{Z}^n \cap [0,1)^n$. Fix such a w, and write it as $w = w' + \frac{1}{2}w''$, where $w' \in \{0, \frac{1}{2}\}^n$ and $w'' \in 2^{-k}\mathbb{Z}^n \cap [0,1)^n$. By the induction hypothesis, there exists an $x'' \in \{0,1\}^n$ such that $\|Ax'' - Aw''\|_\infty \le \mathrm{herdisc}(A)$. We have $w' + \frac{1}{2}x'' \in \{0, \frac{1}{2}, 1\}$, and, by the exercise above, there exists an $x \in \{0,1\}^n$ such that

$$\|Ax - A(w' + \frac{1}{2}x'')\|_\infty \le \frac{1}{2}\mathrm{herdisc}(A). \tag{1}$$

The inequality (1) and the triangle inequality then imply

$$\begin{aligned}
\|Ax - Aw\|_\infty &= \|Ax - A(w' + \frac{1}{2}w'')\|_\infty \\
&\le \|Ax - A(w' + \frac{1}{2}x'')\|_\infty + \frac{1}{2}\|Ax'' - Aw''\|_\infty \\
&\le \mathrm{herdisc}(A).
\end{aligned}$$

This completes the proof of the lemma. □

We are now ready to relate combinatorial and geometric discrepancy.

Theorem 2. *Let $\mathcal{S}$ be a class of Lebesgue measurable sets in $[0,1)^d$ such that $[0,1)^d \in \mathcal{S}$, and let $n \le N$ positive integers. Then,*

$$D(n, \mathcal{S}) \le \frac{2\,\mathrm{disc}(N, \mathcal{S})}{n} + D(N, \mathcal{S}).$$

Above $\mathrm{disc}(N, \mathcal{S}) = \max\{\mathrm{disc}(\mathcal{S}|_P) : P \subset [0,1)^d, |P| \le N\}$.

Proof. Take P to be a set of size N such that $D(P, \mathcal{S}) = D(N, \mathcal{S})$, and let A be the incidence matrix of $\mathcal{S}|_P = \{S \cap P : S \in \mathcal{S}\}$. Let $w \in [0,1)^P$ equal $w_p = n/N$ for each $p \in P$. Then the coordinate corresponding to a set $S \cap P$ of Aw equals $(Aw)_{S\cap P} = \frac{n|P\cap S|}{N}$. On the other hand, by Lemma 1, there exists an $x \in \{0,1\}^P$ such that

$$\|Ax - Aw\|_\infty \leq \text{herdisc}(S|_P) \leq \text{disc}(N, \mathcal{S}), \tag{2}$$

where the last inequality is trivial from the definition of $\text{disc}(N, \mathcal{S})$. Define the point set $Q \subseteq P$ to contain all points in P for which $x_p = 1$, so that for any $S \in \mathcal{S}$ we have $(Ax)_{S\cap P} = |S \cap Q|$. Then by (2), for any $S \in \mathcal{S}$ we have

$$\left| |Q \cap S| - \frac{n|P \cap S|}{N} \right| \leq \text{disc}(N, \mathcal{S}).$$

Together with the discrepancy bound on P, this means that

$$\left| \frac{|Q \cap S|}{n} - \lambda^d(S) \right| \leq \left| \frac{|Q \cap S|}{n} - \frac{|P \cap S|}{N} \right| + \left| \frac{|P \cap S|}{n} - \lambda^d(S) \right| \leq \frac{\text{disc}(N, \mathcal{S})}{n} + D(N, \mathcal{S}). \tag{3}$$

Moreover, because $[0,1)^d \in \mathcal{S}$, we have that $||Q \cap S| - n| \leq \text{disc}(N, \mathcal{S})$. Let us construct Q' by arbitrarily adding or removing at most $\text{disc}(N, \mathcal{S})$ points to Q so that $|Q| = n$. Since $||Q \cap S| - |Q' \cap S|| \leq \text{disc}(N, \mathcal{S})$, together with (3) we get

$$\left| \frac{|Q' \cap S|}{n} - \lambda^d(S) \right| \leq \frac{2\,\text{disc}(N, \mathcal{S})}{n} + D(N, \mathcal{S}).$$

Then,

$$D(n, \mathcal{S}) \leq D(Q', \mathcal{S}) \leq \frac{2\,\text{disc}(N, \mathcal{S})}{n} + D(N, \mathcal{S}),$$

this completes the proof. $\square$

For example, we can take $N = n^2$ and using an $n \times n$ grid we see thatf $D(n^2, \mathcal{R}_2) \leq \frac{1}{n}$. Later we will see that $\text{disc}(n, \mathcal{R}_2) = O((\log n)^{1.5})$. So Theorem 2 gives us taht $D(n, \mathcal{R}_2) = O((\log n)^{1.5}/n)$, which is not quite the best bound we know, but comes quite close.

A more sophisticated way to use Theorem 2 is by repeated halving. Suppose that $\text{disc}(2n, \mathcal{S}) \leq (2 - 2\delta)\,\text{disc}(n, \mathcal{S})$ and that $D(n, \mathcal{S}) = o(1)$. Then, by applying Theorem 2 with n and $2n$, and then $2n$ and $4n$, and so on, we get

$$\begin{aligned} D(n, \mathcal{S}) &\leq \frac{2\,\text{disc}(2n, \mathcal{S})}{n} + \frac{2\,\text{disc}(4n, \mathcal{S})}{2n} + \ldots + \frac{2\,\text{disc}(N, \mathcal{S})}{N/2} + D(N, \mathcal{S}) \\ &\leq \left(1 + (1-\delta) + (1-\delta)^2 + \ldots\right) \frac{2\,\text{disc}(2n, \mathcal{S})}{n} + D(N, \mathcal{S}) \\ &\leq \frac{2\,\text{disc}(2n, \mathcal{S})}{\delta n} + D(N, \mathcal{S}). \end{aligned}$$

Then we can take N big enough so that $D(N, \mathcal{S}) \leq \frac{2\,\text{disc}(2n, \mathcal{S})}{\delta n}$, and we get

$$D(n, \mathcal{S}) \leq \frac{4\,\text{disc}(2n, \mathcal{S})}{\delta n}.$$

In other words, under these mild assumptions, $D(n, \mathcal{S}) = O(\text{disc}(n, \mathcal{S})/n)$.

3 A Low Discrepancy Set for Rectangles

We first make two trivial but useful observations. Given two sets $A, B \subseteq [0,1)^2$:

First, if $A \cap B = \emptyset$ then

$$|D(P, A \cup B)| = |D(P,A) + D(P,B)| \leq |D(P,A)| + |D(P,B)|.$$

Secondly, if $B \subseteq A$, then

$$\begin{aligned} |D(P, A \backslash B)| &= |D(P,A) - D(P,B)| \\ &\leq |D(P,A)| + |D(P,B)|. \end{aligned}$$

It is often convenient to argue about the discrepancy of corners rather than rectangles. We define the set of d-dimensional corners as $\mathcal{C}_d$ as the collection of sets of the form $[0, x_d] \times \ldots \times [0, x_d]$ for $x \in [0,1)^d$. It turns out that the discrepancy with respect to rectangles and the discrepancy with respect to corners are equivalent up to constants.

Exercise 4. *Show that for any $d \geq 1$, we have*

$$D(P, \mathcal{C}_d) \leq D(P, \mathcal{R}_d) \leq 2^d D(P, \mathcal{C}_d). \tag{4}$$

From now on we will use discrepancy with respect to corners or rectangles interchangably. The inequalities (4) imply that this does not affect the asymptotics of the discrepancy function.

The bit reversal function: We define the bit reversal function $r(\cdot) : \mathbb{N} \to [0,1)$ to be the function that takes an integer i, converts it into binary then reverses the bits and precedes them by 0.; to put it another way, $r(i)$ flips the bits of the binary representation of i around the radix point. For instance $1 = 1_2 \implies r(1) = 0.1_2 = 0.5, 2 = 10_2 \implies r(2) = 0.01_2 = 0.25$, and so on. Formally, if $a_0, \ldots, a_{k-1} \in \{0,1\}$ is the unique sequence such that $i = \sum_{i=0}^{k-1} a_i 2^i$, then

$$r(i) := \sum_{i=0}^{k-1} a_i 2^{-i-1}.$$

The *van der Corput Set* is the set of points defined as : $P = \{(\frac{i}{n}, r(i)) : i = 0 \ldots n-1\}$. For the rest of this subsection we will fix P to be this set.

Theorem 3 (Van der Corput)**.** *For P the van der Corput set defined above,*

$$D(P, \mathcal{R}_2) \leq 4 \cdot D(P, \mathcal{C}_2) = O(\log n).$$

We will sketch the proof of the theorem. First we prove:

Claim 4. *Let I be an interval of the form $I = \left[\frac{k}{2^q}, \frac{k+1}{2^q}\right)$ where q is a positive integer and $0 \leq k < 2^q - 1$. Then for any $x \in [0,1)$:*

$$|D(P, [0,x] \times I)| \leq \frac{1}{n}$$

Exercise 5. *Prove Claim 4.*

Hint*: It is enough to show that for any rectangle of the form* $R = \left[\frac{\ell 2^q}{n}, \frac{(\ell+1)2^q}{n}\right) \times \left[\frac{k}{2^q}, \frac{k+1}{2^q}\right)$*, for* $0 \leq k \leq 2^q - 1$ *and* $0 \leq \ell \leq \frac{n}{2^q} - 1$*, we have* $|P \cap R| = 1$ *and* $D(P, R) = 0$*.*

Proof of Theorem 3. To prove the theorem, we use Claim 4 repeatedly. Let $x, y \in [0,1)^2$ be arbitrary. We need to show that $|D(P, C_{xy})| = O(\log n)$. First we choose the smallest integer q_0 such that $\frac{1}{2^{q_0}} \leq y$; by Claim 4, we have

$$\left|D(P, [0,x) \times [0, 2^{-q_0}))\right| \leq \frac{1}{n}.$$

Then we choose the smallest integer $q_1 > q_0$ such that $\frac{1}{2^{q_0}} + \frac{1}{2^{q_1}} \leq y$; again by Claim 4, we have

$$\left|D(P, [0,x) \times [0, (2^{q_1 - q_0} + 1)2^{-q_1}))\right| \leq \frac{1}{n}.$$

We continue in this mannar for $O(\log n)$ iterations. We illustrate the first iteration in Figure 1 below.

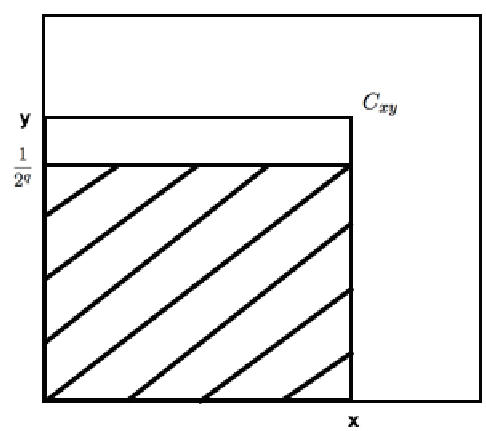

Figure 1: After 1 iteration

After $O(\log n)$ iterations, the remaining rectangle must have area less than $\frac{1}{n}$ and will contain at most one point of P, so it will have discrepancy $\leq \frac{1}{n}$. This implies the upper bound of $O((\log n)/n)$ on the discrepancy of the van der Corput set. □

Exercise 6. *Let us define* $r_2(i) = r(i)$*, as defined above, and, for a prime* p*, let*

$$r_p(i) = \sum_{i=0}^{k-1} a_i p^{-i-1},$$

where $a_0, \ldots, a_{k-1}$ *are the digits of* i *in base* p*, i.e. the unique sequence in* $\{0, \ldots, p-1\}$ *such that* $i = \sum_{i=0}^{k-1} a_i p^i$*. Let* $p_1 = 2, \ldots, p_{d-1}$ *be the first* $d-1$ *primes. Show that the* n*-point set* $P = \{(\frac{i}{n}, r_2(i), \ldots, r_{p_{d-1}}(i)) : i = 0 \ldots n-1\}$ *has discrepancy* $D(P, \mathcal{C}_d) = O\left(\frac{(\log n)^{d-1}}{n}\right)$*.*

References

[1] Dmitriy Bilyk. On Roth's orthogonal function method in discrepancy theory. *Unif. Distrib. Theory*, 6(1):143–184, 2011.

[2] Dmitriy Bilyk, Michael T. Lacey, and Armen Vagharshakyan. On the small ball inequality in all dimensions. *J. Funct. Anal.*, 254(9):2470–2502, 2008.

[3] L. Lovász, J. Spencer, and K. Vesztergombi. Discrepancy of set-systems and matrices. *European J. Combin.*, 7(2):151–160, 1986.

[4] Wolfgang Schmidt. Irregularities of distribution, vii. *Acta Arithmetica*, 1(21):45–50, 1972.