

## Section 1: Introduction

Lecturer: Jiri Matousek

Scribe: Lily Li

## 1.1 Discrepancy for Rectangles and Uniform Distribution

Lets begin with the following problem:

**Example 1.1** *How can we uniformly distribute  $n$  points in a  $d$ -dimensional unit cube  $[0, 1]^d$ . The case where  $d = 1$  is quite straight-forward. Simply place point  $i$  at position  $i/d + 1/(2d)$  — evenly spaced. Next consider the case where  $d = 2$ . Now the precise definition of uniform is not so clear. On possible criterion is **with respect to axis-parallel rectangles**. Let  $P$  be an  $n$ -point set in the unit square  $[0, 1]^2$  and  $R$  is an axis-parallel rectangle where  $R = [a_1, b_1] \times [a_2, b_2] \subseteq [0, 1]^2$ . Then  $P$  is **uniformly distributed** if the number of points of  $P$  that fall in  $R$  is approximately  $n \cdot \text{vol}(R)$  where  $\text{vol}(R)$  is the area of  $R$  (further note that the choice of  $P$  is uniformly and independently at random thus  $n \cdot \text{vol}(R)$  is just the expected number of points in  $R$ , there is no adversary picking the worst point set  $P$  possible).*

Let  $P$  be **justly distributed** if the deviation (defined below)

$$|n \cdot \text{vol}(R) - |P \cap R||$$

is at most  $c$  — some constant — for all axis-parallel rectangles  $R$ . As it turns out justly distribution is impossible for sufficiently large sets. There will always exist some rectangle  $R$  for which the magnitude of irregularity grows to infinite as  $n \rightarrow \infty$ .

From this example we can extract some useful notation. Let the **deviation** of  $P$  from a uniform distribution on a particular rectangle  $R$  be

$$D(P, R) = n \cdot \text{vol}(R) - |P \cap R|.$$

Let  $\mathcal{R}_2$  be the set of all axis-parallel rectangles in the unit square. Then

$$D(P, \mathcal{R}_2) = \sup_{R \in \mathcal{R}_2} |D(P, R)|$$

is the **discrepancy of  $P$**  for axis-parallel rectangles. Further, the function

$$D(n, \mathcal{R}_2) = \inf_{P \subset [0, 1]^2, |P|=n} D(P, \mathcal{R}_2)$$

quantifies the smallest possible discrepancy of an  $n$ -point set.

**Problem 1.2** *Is  $D(n, \mathcal{R}_2)$  bounded above by a constant for all  $n$ ? (or is  $\limsup_{n \rightarrow \infty} D(n, \mathcal{R}_2) = \infty$ )*

Remark that your intuitive notion of what is or is not evenly distributed should be kept in check. The triangle lattice arrangement of points looks pretty even, but has pretty bad discrepancy (about  $\sqrt{n}$ ). It is possible that good discrepancy results from a distribution which puts two points very close together.

### 1.1.1 Uniform Distribution of Infinite Sequences

Previously we thought about a finite number of point in unit cubes of increasing dimension. Now let us consider what would happen if we have infinitely many points. Again consider the one-dimensional interval  $I = [0, 1]$  and a infinite sequence  $u = (u_1, u_2, \dots)$  points in  $I$ . Uniformity here means that as the points of  $u$  are added one by one, they *sweep out* all subintervals of  $I$  as evenly as possible.

**Definition 1.3**  $u = (u_1, u_2, \dots)$  is **uniformly distributed** in  $[0, 1]$  if for each subinterval  $[a, b] \subset [0, 1]$ ,

$$\lim_{n \rightarrow \infty} \left( \frac{|\{u_1, \dots, u_n\} \cap [a, b]|}{n} \right) = b - a$$

Note: for any Riemann-integrable function  $f : [0, 1] \rightarrow \mathcal{R}$ ,

$$\lim_{n \rightarrow \infty} \left( \frac{\sum_{i=1}^n f(u_i)}{n} \right) = \int_0^1 f(x) dx$$

(do you still remember what a Riemann-integrable function is? Actually it is just normal integrable function, they just need to be continuous everywhere).

Further note: the trigonometric polynomials of the form

$$f(x) = \sum_{k=0}^n (a_k \sin(2\pi kx) + b_k \cos(2\pi kx))$$

with  $a_0, \dots, a_n$  and  $b_0, \dots, b_n$  real or complex are also sufficient for testing the uniform distribution.

**Theorem 1.4** (Weyl's criterion) A sequence  $u = (u_1, u_2, \dots)$  is uniformly distributed in  $[0, 1]$  if and only if for all integers  $k \neq 0$ ,

$$\lim_{n \rightarrow \infty} \left( \frac{\sum_{j=1}^n e^{2\pi i k u_j}}{n} \right) = \int_0^1 e^{2\pi i k u_j} dx = 0$$

An application of the above criterion is the following:

**Theorem 1.5** For each irrational number  $\alpha$ , the sequence  $u = (u_1, u_2, \dots)$  given by  $u_n = \{\alpha n\}$  is uniformly distributed in  $[0, 1]$ , where  $\{x\}$  is the fractional part of  $x$ .

**Proof:** Lets be honest, we are not really sure what is going on here... well, no actually it is not that bad. If you take Weyl's criterion as a given — which is a pretty reasonable thing to do really — then this proof is just a simple application of the criterion. In fact it seems tailor made for this problem since the notation wipes out the annoying “fractional part” function. ■

### 1.1.2 Dynamic Setting

Not all uniformly distributed sequences are made equal. Discrepancy was originally introduced to measure the amount of non-uniformity in an infinite sequence.

**Definition 1.6** *Discrepancy of an infinite sequence  $u$  in  $[0, 1]$  is defined as*

$$\Delta(u, n) = \sup_{0 \leq a \leq b \leq 1} |n(b - a) - |\{u_1, \dots, u_n\} \cap [a, b]|$$

In-fact, as we will soon see, there is a very tight connection between the two-dimensional and infinite one-dimensional problems.

## 1.2 Combinatorial Discrepancy

Let  $X$  be an  $n$ -element set and  $\mathcal{S}$  a system of subsets of  $X$  called a **set system**. Our goal is to color  $X$  red or blue so that every set of  $\mathcal{S}$  has roughly the same number of red and blue points.

**Definition 1.7** A **coloring** of  $X$  is a mapping  $\chi : X \rightarrow \{-1, +1\}$ .

The **discrepancy** of the set system  $\mathcal{S}$  is the maximum deviation from an even split over all sets of  $\mathcal{S}$ . Formally the **discrepancy** of  $\mathcal{S}$ , denoted  $\text{disc}(\mathcal{S})$  is the minimum overall colorings  $\chi$  of

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

Note that  $\chi(S) = \sum_{x \in S} \chi(x)$  so taking the norm makes sense.

This type of discrepancy is often denoted *combinatorial discrepancy* to contrast with the type defined in the previous section known as *Lebesgue-measure discrepancy*.

Examples of when the geometry sneaks into the problem involves the following: given an  $n$ -point set  $P$ , color each  $p \in P$  red or blue to minimize the maximum the difference between the number of points of each color in a halfplane  $h$  over all halfplanes.

**Definition 1.8** If  $(X, \mathcal{A})$  is a set system ( $X$  might be infinite) and  $Y \subset X$  is a set. Then the **set system induced by  $\mathcal{A}$  on  $Y$**  (sometimes called the **trace** of  $\mathcal{A}$  on  $Y$ ) is the set system:

$$\mathcal{A}|_Y = \{A \cap Y : A \in \mathcal{A}\}$$

Typically  $\mathcal{A} \subset \mathcal{R}^d$  and the combinatorial discrepancies set system  $\mathcal{A}|_P$  where  $P \subseteq \mathcal{R}^d$  is a finite set we are considering. Often, for notational simplicity  $\text{disc}(P, \mathcal{A}) = \text{disc}(\mathcal{A}|_P)$ . Explicitly  $\text{disc}(P, \mathcal{A})$  is the minimum over all colorings  $\chi : P \rightarrow \{-1, 1\}$  of

$$\text{disc}(\chi, P, \mathcal{A}) = \max_{A \in \mathcal{A}} |\chi(P \cap A)|.$$

The **discrepancy function** of  $\mathcal{A}$  is

$$\text{disc}(n, \mathcal{A}) = \max_{|P|=n} \text{disc}(P, \mathcal{A})$$

We will show later that upper bounds for combinatorial discrepancy for some class  $\mathcal{A}$  imply upper bounds for the Lebesgue-measure discrepancy for  $\mathcal{A}$  (the reverse does not hold).

### 1.2.1 $\epsilon$ -approximations

From probability theory, the notion of an  $\epsilon$ -approximation is a generalization of both types of discrepancy we have discussed so far. Let  $X$  be some finite or infinite ground set,  $\mu$  is a measure on  $X$  with  $\mu(X) < \infty$ ,

and  $\mathcal{S}$  is a system of  $\mu$ -measurable subsets of  $X$ . Let  $\epsilon \in [0, 1]$  be a real number. Then  $Y \subseteq X$  is an  $\epsilon$ -approximation for the set system  $(X, \mathcal{S})$  with respect to measure  $\mu$  if for all  $S \in \mathcal{S}$ ,

$$\left| \frac{|Y \cap S|}{|Y|} - \frac{\mu(S)}{\mu(X)} \right| \leq \epsilon$$

Notice that this notion is indeed quite similar to that of discrepancy, particularly combinatorial discrepancy. Here we want the ratio of points in  $Y$  which are also in  $S$  to be proportional to the ratio of  $S$  in the entire population. General we let  $\mu$  be the counting measure i.e.  $\mu(S) = |S|$ .

According to the book, the connection to the Lebesgue-measure discrepancy is the obvious one:

**Proposition 1.9** *If  $\mathcal{A}$  is a class of Lebesgue-measurable sets in  $\mathcal{R}^d$  and  $P \subset [0, 1]^d$  is an  $n$ -point set, then  $D(P, \mathcal{A}) \leq \epsilon n$  if and only if  $P$  is an  $\epsilon$ -approximation for  $(\mathcal{R}^d, \mathcal{A})$  with respect to the measure  $\text{vol}_{\square} = \text{vol}(\mathcal{A} \cap [0, 1]^d)$ .*

Here we will go into a little foray into measure theory: recall

### Definition 1.10

Let  $X$  be a set and  $\sigma$  a  $\sigma$ -algebra — collection of subsets of  $X$  including the empty subset which is closed under complement, countable unions and countable intersections — over  $X$ . A function  $\mu : \sigma \rightarrow \mathbb{R}$  is a **measure** if it has the following properties:

1. Non-negativity: for all  $S \in \Sigma : \mu(S) \geq 0$
2. Null empty set:  $\mu(\emptyset) = 0$  (and for all non-empty sets,  $S, \mu(S) \neq 0$ ?)
3. Countable additivity (aka  $\sigma$ -additivity): for all countable collections  $\{S_i\}_{i=1}^{\infty}$  of pairwise disjoint sets in  $\Sigma$ :  $\mu(\cup_{k=1}^{\infty} S_k) = \sum_{k=1}^{\infty} \mu(S_k)$ .

The pair  $(X, \Sigma)$  is a **measurable space** and the members of  $\Sigma$  are measurable sets. Further the triple  $(X, \Sigma, \mu)$  is called a **measure space** (think: it is a *measurable space* with a measure).

**Lemma 1.11** (*Combinatorial discrepancy and  $\epsilon$ -approximations*) *Let  $\mathcal{S}$  be a system of subsets of a  $2n$ -point set  $X$ .*

1. *If  $Y \subset X$  is an  $n$ -point set that is an  $\epsilon$ -approximation for  $(X, \mathcal{S})$  then  $\text{disc}(\mathcal{S}) \leq 2\epsilon n$*
2. *If  $\mathcal{S}$  is such that  $X \in \mathcal{S}$  and  $\text{disc}(\mathcal{S}) \leq \epsilon n$  then there exists an  $n$ -point set  $Y \subset X$  such that  $Y$  is an  $\epsilon$ -approximation for  $(X, \mathcal{S})$ .*

**Proof:** To prove part (1) we take the  $n$ -point set  $Y \subset X$  and provide a coloring  $\chi$  such that  $\text{disc}(\chi, \mathcal{S}) \leq 2\epsilon n$  (notice since  $\text{disc}(\mathcal{S})$  is taken to be the minimum over all colorings, it suffices to find one coloring which is upper bounded by  $2\epsilon n$ ). The coloring is the obvious one:  $\chi(p) = 1$  if  $p \in Y$  and  $\chi(p) = -1$  otherwise. Now we can calculate  $\chi(S)$  for all  $S \in \mathcal{S}$

$$\chi(S) = \sum_{x \in S} \chi(x) = |Y \cap S| - (|S| - |Y \cap S|) = 2|Y \cap S| - |S|.$$

Since  $Y$  is an  $\epsilon$ -approximation for  $(X, \mathcal{S})$ :

$$\left| \frac{|Y \cap S|}{|Y|} - \frac{|S|}{|X|} \right| = \frac{1}{2n} |2|Y \cap S| - |S|| \leq \epsilon.$$

Notice that we basically get what we want if we move the  $2n$  to the other side since  $|\chi(S)| \leq 2\epsilon n$ .

Now for part (2) we need the opposite direction to be true. So given that there exists a coloring  $\chi$  such that  $|\chi(S)| \leq \epsilon n$  for all  $S \in \mathcal{S}$  (including  $X$ ) we need to construct a  $n$ -point set  $Y$  which is an  $\epsilon$ -approximation for  $(X, \mathcal{S})$ . The subset to take should be inspired from the first part namely we will take  $Y' = \arg \max\{\chi^{-1}(1), \chi^{-1}(-1)\}$  (we need to take the larger set because the subset we actually want has size  $n$ ). Next we will use the fact that  $X \in \mathcal{S}$  to get a bound on  $Y'$ :

$$|\chi(X)| = |2|Y' \cap X| - |X|| = 2|Y'| - 2n \leq \epsilon n$$

thus  $|Y'| - n \leq \epsilon/2$ . Now we can show that a subset  $Y \subset Y'$  of size  $n$  is an  $\epsilon$ -approximation of  $(X, \mathcal{S})$  since for any  $S \in \mathcal{S}$

$$\begin{aligned} \left| \frac{|Y \cap S|}{|Y|} - \frac{|S|}{|X|} \right| &= \frac{1}{2n} |2|Y \cap S| - |S|| \\ &= \frac{1}{2n} |2|Y \cap S| - |S| + 2|Y' \cap S| - 2|Y' \cap S|| \\ &\leq \frac{1}{n} ||Y \cap S| - |Y' \cap S|| + \frac{1}{2n} |2|Y' \cap S| - |S|| \quad \text{By triangle inequality} \\ &\leq \frac{n\epsilon}{2n} + \frac{\chi(S)}{2n} \\ &\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \end{aligned}$$

■

**Proposition 1.12** *If  $Y_0$  is an  $\epsilon$ -approximation for  $(X, \mathcal{S})$  with respect to measure  $\mu$  and  $Y_1$  is a  $\delta$ -approximation for set system  $(Y_0, \mathcal{S}|_{Y_0})$ . Then  $Y_1$  is an  $(\epsilon + \delta)$ -approximation for  $(X, \mathcal{S})$  with respect to  $\mu$ .*

**Lemma 1.13** (Transference Lemma) *Let  $\mathcal{A}$  be a class of Lebesgue measurable sets in  $\mathcal{R}^d$  containing a set  $A_0$  with  $[0, 1]^d \subseteq A_0$ . Suppose that  $D(n, \mathcal{A}) = o(n)$  as  $n \rightarrow \infty$ , and that  $\text{disc}(n, \mathcal{A}) \leq f(n)$  for all  $n$ , where  $f(n)$  is a function satisfying  $f(2n) \leq (2 - \delta)f(n)$  for all  $n$  and some fixed  $\delta > 0$ . Then  $D(n, \mathcal{A}) = O(f(n))$ .*

*On the other hand, if  $D(n, \mathcal{A}) = o(n)$  and  $D(n, \mathcal{A}) \geq f(n)$  for all  $n$ , with a class  $\mathcal{A}$  and a function  $f(n)$  as above, then  $\text{disc}(n, \mathcal{A}) \geq cf(n)$  holds for infinitely many  $n$  with a suitable constant  $c = c(\delta) > 0$ .*

**Proof:** We begin by taking note of some suggestive features of the lemma definition: first note that  $A_0$  is not used anywhere in the proof (why it is necessary then?) further, why does the function  $f$  need to be sub-linear? With these thoughts in mind we are going to prove the first part of the lemma. Remember we want to show that  $D(n, \mathcal{A}) = O(\text{disc}(n, \mathcal{A}))$ . Thus it helps to pick an  $n$  and work from there.

What can we do with some random  $n$ ? Let's define a constant  $\epsilon = f(n)/n$ . Since  $\epsilon < 1$  and  $D(n, \mathcal{A}) = o(n)$  we can find some  $k$  such that

$$\frac{D(2^k n, \mathcal{A})}{2^k n} \leq \epsilon.$$

The reason for doing this is so that we can use the second part of lemma 1.11 to halve the number of points in each iteration. We get the hint for this from the  $A_0$  since the lemma requires that the ground set is a member of the set system (with  $[0, 1]^d \subset A_0$  this is no problem). The above inequality really says that there exists some point set  $P_0$  of size  $2^k n$  such that  $P_0$  is an  $\epsilon$  approximation of  $(\mathcal{R}^d, \mathcal{A})$  with measure  $\text{vol}_\square$ . Reduce the size of the point set by halving as described above). Generally at step  $i + 1$  we have the set system  $(P_i, \mathcal{A}|_{P_i})$  and we take a coloring  $\chi_i$  with discrepancy at most  $f(|P_i|)$ . Use lemma 1.11 (ii) to find a subset

$P_{i+1} \subset P_i$  of  $2^{k-1-i}n$  point and which is an  $\epsilon_i$  approximation for  $(P_i, \mathcal{A}|_{P_i})$  where  $\epsilon_i = f(2^k n)/2^k n$ . After  $k$  iterations we get a set  $P_k$  of size  $n$  which is a  $\eta$ -approximation of the original set system  $(\mathcal{R}^d, \mathcal{A})$  where

$$\eta = \epsilon + \sum_{i=0}^{k-1} \epsilon_i = \frac{f(n)}{n} + \sum_{i=1}^k \frac{f(2^i n)}{2^i n} \leq \frac{f(n)}{n} \left( 1 + \sum_{i=1}^{\infty} \left( \frac{2-\delta}{2} \right)^i \right) \in O\left(\frac{f(n)}{n}\right)$$

by proposition 1.12. Thus  $D(n, \mathcal{A}) \leq O(\text{disc}(n, \mathcal{A}))$  by an application of proposition 1.9.

Now we will show the second part of the lemma in the following ■