CSC2414: Discrepancy Theory in Computer Science Fall 2015

Lecture 2 — September 20, 2015

Aleksandar Nikolov *Scribe: Lalla Mouatadid*

1 Estimating Integrals

Estimating a complicated integral numerically is a problem which is ubiquitous in applied mathematics. Often we have to deal with functions for which the integral has no simple closed form. Even worse, often we do not even have a formula for the function we are integrating, and instead can only query its value at specific points, for example by performing an experiment. We might, however, have some knowledge about the function, for example that it is bounded or somewhat smooth.

To formalize this integral estimation problem, let us say we want to design a sequence $u = (u_1, \ldots, u_N)$, where $u_i \in [0, 1)$, such that for all $1 \leq n \leq N$ and for all "nice" functions $f : [0, 1) \to \mathbb{R}$ we can bound the error

$$\mathrm{err}(f, u, n) := \left| \int_0^1 f(x)dx - \frac{1}{n} \sum_{i=1}^{n} f(u_i) \right| \leq \epsilon(n). \tag{1}$$

Here $\epsilon(n)$ is an error bound, and we would like it go to 0 with n. Then, we can compute an initial estimate of $\int_0^1 f(x)dx)$ by sampling it at the first n points of u, and refine the estimate by taking more points from u if need be. In order to minimize the number of times we have to query f (remember that each query can be a whole new experiment, and hence be costly) we want to make sure that $\epsilon(n)$ goes to 0 as fast as possible. Hence, a central question when estimating integrals is how fast $\epsilon(n)$ can be made to converge to 0 for a class of "nice" functions.

One obvious approach, known as the Monte Carlo Method (MCM), is to take each u_i to be i.i.d and uniform in $[0, 1)$. Then for any f and n, the error in expectation is:

$$\mathbb{E}\,\mathrm{err}(f, u, n) \leq \frac{1}{\sqrt{n}} \left(\int_0^1 f(x)^2 dx \right)^{1/2}$$

So, if we take the "nice" functions to be those for which the "energy" $\|f\|_2 := \left(\int_0^1 f(x)^2 dx \right)^{1/2}$ is bounded, then $\epsilon(n) = O(n^{-1/2})$. A natural question is whether we can do better, and the answer turns out to be positive, if we are somewhat more strict about what functions are considered "nice". The idea is to choose the sequence u more carefully, rather than randomly, and in fact to use continuous discrepancy to do so.

1.1 The Quasi Monte Carlo Method

Let $\Delta(u)$ denote the discrepancy of a sequence:

$$\Delta(u) = \max_{n=1}^{N} \sup_{0 \leq t \leq 1} |tn - |\{i \leq n : u_i < t\}||$$

If we take n random u_i's, then tn is the expected number of u_i's we expect to see in $[0,t)$, and $|\{i : u_i < t\}$ is the actual number we see. The discrepancy of the sequence is the maximum difference between these two quantities over all n and t.

Consider the indicator function of $[0,t)$:

$$f(x) = \begin{cases} 1, & \text{if } x < t \\ 0, & \text{o.w} \end{cases}$$

Then, for any such f, simply using the definition of $\Delta(u)$, we get the error bound

$$\text{err}(f,u,n) \leq \frac{\Delta(u)}{n},$$

since $|\{i : u_i < t\}| = \sum_{i=1}^{n} f(u_i)$ and $tn = n\int_0^1 f(x)dx$. It turns out that almost the same error bound in fact holds for a much larger family of functions.

Theorem 1 (Koksma-Hlawka Inequality)**.** *For all f, u, and n such that $1 \leq n \leq N$, we have*

$$\text{err}(f,u,n) \leq \frac{\Delta(u)}{n} V(f), \tag{2}$$

where $V(f)$ is the total variation of f and for differentiable functions is equal to $\int_0^1 |f'(x)|\,dx$.

The total variation $V(f)$ is a measure of the smoothness of f. The Koksma-Hlawka suggests that a good strategy for estimating integrals of smooth functions with low error is to take u to be a sequence with low discrepancy. This is known as the quasi Monte Carlo method. Observe that if it is possible to find a sequence u for which $\Delta(u)$ is much smaller than $N^{1/2}$, than we can improve on the convergence of the MCM for functions with bounded total variation.

1.2 Discrepancy of Sequences vs. Discrepancy of Pointsets

How does $\Delta(u)$ relate to the discrepancy of sets? Let us recall the discrepancy of sets. Consider a set P of points, $P \subset [0,1)^2, |P| = N$, and let $A \subseteq [0,1)^2$ be Lebesgue measurable. We define the discrepancy of P w.r.t. A as:

$$D(P,A) = n \cdot \text{area}(A) - |P \cap A|$$

As in the first lecture, we will focus on the discrepancy with respect to *corners*, also known as *star discrepancy*. The corner below the point (x,y) is defined as $C_{xy} = \{z : 0 \leq z_1 \leq x, 0 \leq z_2 \leq y\}$, and the collection of all corners in the unit square is $\mathcal{C}_2 = \{C_{xy} : x, y \in [0,1)^2\}$. Then the discrepancy of P with respect to corners is

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

Given a sequence $(u_1, \ldots, u_N)$, we can make a set of size N as follows:

$$P = \left\{ \left(\frac{i-1}{N}, u_i \right) \right\}.$$

It is straightforward to veriy that $\Delta(u) = D(P, \mathcal{C}_2) \pm O(1)$ using the definitions. Conversely, given a set $P = \{p_1, \ldots, p_N\}$, where p_i's are listed in increasing order of the x-coordinates, we construct u by taking u_i to be the y-coordinate of p_i. Once again $D(P, \mathcal{C}_2) = \Delta(u) \pm O(1)$.

The above shows that in order to construct a low discrepancy sequence, it is enough to construct a low discrepancy set one dimension higher.

2 Constructing low discrepancy sets

2.1 Rectangles vs. Corners

We first make two 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

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

Analogously to corners, we can define the discrepancy of a point set $P \subset [0,1)^2$, $|P| = n$, with respect to the set of axis-aligned rectangles $\mathcal{R}_2 = \{[x_1, x_2) \times [y_1, y_2)\}$ as

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

It turns out that the discrepancy with respect to rectangles and the discrepancy with respect to corners are equivalent up to constants:

$$D(P, \mathcal{C}_2) \leq D(P, \mathcal{R}_2) \leq 4D(P, \mathcal{C}_2). \tag{3}$$

The first inequality is trivial as each corner is a rectangle. For the second one we can use the two observations above and the fact that we can express any axis-aligned rectangle in terms of four corners as follows in Figure 1.

From now on we will use discrepancy with respect to corners or rectangles interchangably. The inequalities (3) imply that this does not affect the asymptotics of the discrepancy function.

2.2 The van der Corput Construction

Recall from last class that if we have a grid with $\sqrt{n}$ rows and $\sqrt{n}$ columns, we can select a thin rectangle around a columns such that the area isvery close to 0 but we have $\sqrt{n}$ points, resulting in discrepancy on the order of $\sqrt{n}$. Rotating the grid such that the slope is an irrational value results in better discrepancy. For slope $\sqrt{2}$ for instance, one can show that the discrepancy $D(P, \mathcal{C}_2) = O(\log n)$ for $|P| = n$. It turns out that this bound is optimal up to constants. There

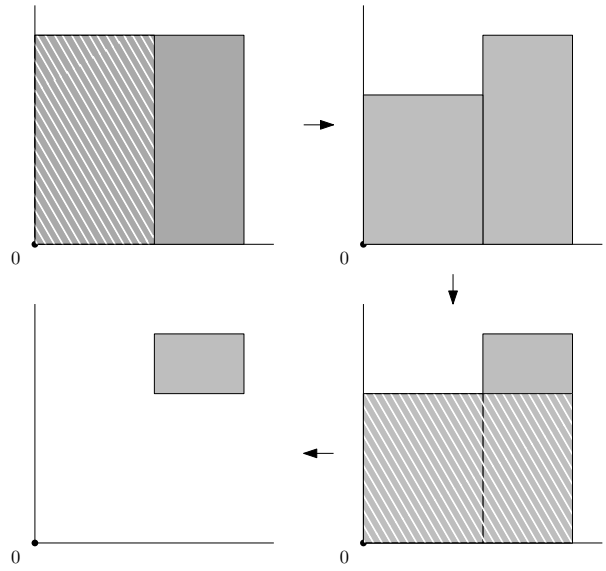


Figure 1: Expressing a rectangle using four corners.

are other constructions that give the same result, and next we will consider one that is particularly easy to analyze.

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

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

We will sketch the proof of the theorem. For the full proof, see Chapter 2.1 in Matoušek's book [1]. First we prove:

Claim 3. *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 $0 \le k \le 2^q - 1$. Then for any $x \in [0, 1)$:*

$$|D(P, [0, x) \times I)| \le 1$$

Proof. To get some intuition for this claim first, consider, for instance, $I = [1/2, 1)$. Then $r(i) \in I$ exactly when i is odd, and it follows that any rectangle of the type $\left[\frac{2i}{n}, \frac{4i}{n}\right) \times I$, for i a non-negative

integer, contains precisely one point of P and has area 1. We can divide $[0, x) \times I$ into rectangles of this type, and one final rectangle $\left[\frac{2i}{n}, x\right) \times I$. All rectangles but the last one have discrepancy 0, and the last one contains at most a single point and has area at most 1, so it has discrepancy at most 1 in absolute value. Using the observations from the beginning of this section, this implies the claim for this particular I.

In general, if $I = \left[\frac{k}{2^q}, \frac{k+1}{2^q}\right)$, then $r(i) \in I$ if and only if $r(i) \equiv 2^q r(k) \pmod{2^q}$. It follows that any rectangle of the type $\left[\frac{i2^q}{n}, \frac{(i+1)2^q}{n}\right) \times I$, for i a non-negative integer, contains exactly one point of P, and has area 1, so it has discrepancy 0. Any rectangle $[0, x) \times I$ can be divided into rectangles of the above type and one final rectangle of discrepancy at most 1. The claim then follows from the observations from the beginning of the section. □

Proof of Theorem 2. To prove the theorem, we use Claim 3 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 3, we have

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

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 3, we have

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

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

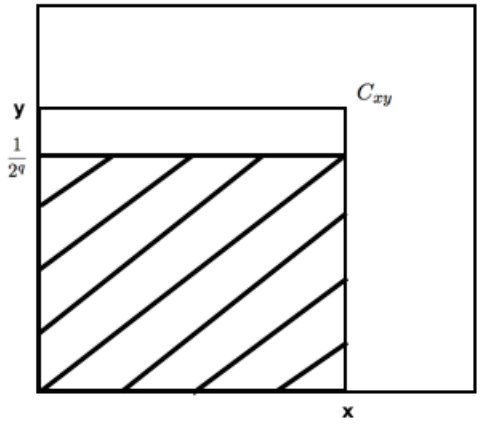


Figure 2: After 1 iteration

After $O(\log n)$ iterations, the remaining rectangle must have area less than $\frac{1}{n}$ and will contain no points of P, so it will have discrepancy ≤ 1. This implies the upper bound of $O(\log n)$ on the discrepancy of the Van der Corput set. □

3 Roth's Lower Bound

Recall that $D(n, \mathcal{C}_2) = \inf_{\substack{P \subset [0,1)^2 \\ |P|=n}} D(P, \mathcal{C}_2)$. In the last section we saw that $D(n, \mathcal{C}_2) = O(\log n)$. Schmidt showed that this bound is tight up to constants [3]. In this section we will sketch Klaus Roth's beautiful proof of the weaker bound $D(n, \mathcal{C}_2) = \Omega(\sqrt{\log n})$.

Let us start with a very simple lower bound:

$$\begin{aligned} D(n, \mathcal{C}_2) &\geq \frac{1}{4} D(n, \mathcal{R}_2) \\ &\geq \frac{1}{4}(1 - \frac{1}{n+1}) \end{aligned}$$

The first inequality is just (3). To see the second inequality, divide the $[0,1)^2$ square into $n+1$ smaller squares of equal area. By the pigeonhole principle at least one of them is empty: denote that square by Q. Then,

$$\begin{aligned} \text{area}(Q) &= \frac{1}{n+1} \\ D(P, Q) &= \text{area}(Q) \cdot n - |P \cap Q| \\ &= \frac{n}{n+1} = 1 - \frac{1}{n+1}. \end{aligned}$$

Roth managed to "lift" this simple pigeonhole argument to $\Omega(\sqrt{\log n})$.

Theorem 4 ([2])**.** $D(n, \mathcal{C}_2) = \Omega(\sqrt{\log n})$

Proof. Fix an arbitrary P of size n. Let us use the notation C_u for C_{xy} where $u = (x, y) \in [0,1)$, and $D(u) = D(P, C_u)$. To prove the theorem, we prove the following inequality:

$$\begin{aligned} \sup_{u \in [0,1)^2} |D(u)| &\geq \left(\int_{[0,1)^2} D(u)^2 du \right)^{\frac{1}{2}} \\ &= \Omega(\sqrt{\log n}). \end{aligned}$$

We use this opportunity to remark that this bound on the L_2 discrepancy $\left(\int_{[0,1)^2} D(u)^2 du \right)^{\frac{1}{2}}$ is in fact tigh.

Our strategy is to find a function $F : [0,1)^2 \to \mathbb{R}$, which depends on P, such that

$$\int_{[0,1)^2} F(u)^2 du \leq C \log n, \tag{4}$$

$$\int_{[0,1)^2} F(u) D(u) du \geq c \log n, \tag{5}$$

for constants c, C. If we can find such an F, then we are done, because by Cauchy-Schwarz we get

$$\left(\int F(u)^2 du \right)^{\frac{1}{2}} \left(\int D(u)^2 du \right)^{\frac{1}{2}} \geq \int F(u) D(u) du,$$

and, therefore,

$$\left(\int D(u)^2 du\right)^{\frac{1}{2}} \geq \frac{\int F(u)D(u)du}{\left(\int F(u)^2 du\right)^{\frac{1}{2}}} \geq (c/C)\sqrt{\log n}.$$

(Above and in the rest of the proof we omit the domain of integration when it is equal to $[0,1)^2$.)

The remaining questions then is how to find such an F. Let us pick an integer m such that $2n \leq 2^m \leq 4n$. We define $F(u)$ to be

$$F(u) = f_0(u) + f_1(u) + \ldots + f_m(u),$$

where the functions $f_i : [0,1]^2 \to \{-1, 0, 1\}$ satisfy the following key properties:

$$\forall i : \int f_i(u)^2 du \leq 1, \tag{6}$$

$$\forall i \neq j : \int f_i(u) f_j(u) du = 0 \tag{7}$$

$$\forall i : \int f_i(u) D(u) du \geq c_0. \tag{8}$$

for some constant c. Let us first see why this suffices to prove (4) and (5), and, therefore, finish the proof. First, properties (6) and (7) together imply:

$$\int F(u)^2 du = \sum_{i,j} \int f_i(u) f_j(u) du = \sum_i \int f_i(u)^2 du \leq m = O(\log n)$$

Second, property (8) implies

$$\int F(u)D(u)du = \sum_i \int f_i(u)D(u)du \geq c_0 m = \Omega(\log n).$$

This proves (4) and (5) based on (6)–(8).

For the remainder of the proof we define the functions f_i and prove they satisfy (6)–(8). To define f_i, let us divide the unit square into 2^{m-i} rows and 2^i columns, so that $[0,1)^2$ is partitioned into 2^m smaller rectangles. The function f_i is defined as 0 in every rectangle which contains a point $p \in P$. For the remaining (empty) rectangles, devide each one into 4 smaller identical rectangles and set f_i to 1 in the upper right and lower left corners and to -1 in the remaining two corners. Figure 2 is an example for $|P| = 3, m = 3, i = 2$.

(This definition of the f_i may appear to come out of nowhere, but in fact it is based on the Haar wavelets, a well-knwon system of orthogonal functions that is widely used in signal processing.)

It remains to show the f_i's satisfy properties (6)–(8) above. To see (6), observe that $f_i(u)^2 \leq 1$ for all u, so $\int f_i(u)^2 \leq 1$. To show the orthogonality property (7) for f_i, f_j, $i < j$, observe that we can divide the unit square into $2^{m-i} \times 2^j$ rectangles, so that the product $f_i(u)f_j(u)$ is either zero on

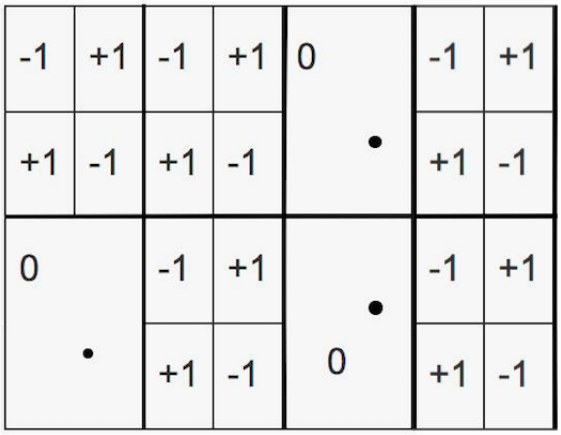


Figure 3: Illustration of f_i

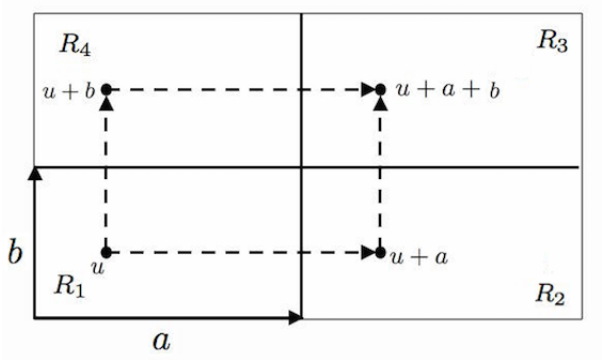
such a rectangle, or is "checkered", i.e. either 1 in the top left and bottom right quadrants of the rectangle, and -1 in the remaining quadrants, or the other way around. In all these cases $f_i(u)f_j(u)$ integrates to 0 on such a rectangle, and therefore integrates to 0 on the entire unit square.

Observe that among the 2^m rectangles used in the definition of f_i, at least n are empty by the pigeonhole principle, since 2^m was chosen to be at least $2n$, and $|P| = n$. Because of this fact, and because f_i is 0 on non-empty rectangles, to prove (8) it suffices to show that for any empty rectangle R

$$\int_R f_i(u)D(u)du = \Omega\left(\frac{1}{n}\right).$$

To prove the above bound, we divide R into 4 rectangles, enumerated R_1, R_2, R_3, and R_4 counterclockwise, starting from the lower left (see Figure 3). Then we can rewrite the integral above as:

$$\int_R f_i(u)D(u)du = \int_{R_1} D(u) - \int_{R_2} D(u) + \int_{R_3} D(u) - \int_{R_4} D(u)du$$



Let a be the vector $(0, 2^{-i-1})$, and let b be the vector $(2^{i-m-1}, 0)$. Notice that a is parallel to the horizontal side of R_1 and has the same length as it, and b is parallel to the vertical side of R_1 and

has the same length as it. Using the vectors a and b, we rewrite the integral as

$$\begin{aligned}\int_R f_i(u)D(u)du &= \int_{R_1} D(u) - \int_{R_2} D(u) + \int_{R_3} D(u) - \int_{R_4} D(u)du \\ &= \int_{R_1} \left(D(u) - D(u+a) + D(u+a+b) - D(u+b)\right) du \\ &= n\left(\text{area}(C_u) - \text{area}(C_{u+a}) + \text{area}(C_{u+a+b}) - \text{area}(C_{u+b})\right) \\ &\quad - \left(|P \cap C_u| + |P \cap C_{u+a}| - |P \cap C_{u+a+b}| + |P \cap C_{u+b}|\right), \end{aligned} \tag{9}$$

where the final equality is by the definition of discrepancy of sets. The second term in (9) is equal to $|P \cap (R_1 + u)|$, which is equal to 0 because $R_1 + u \subset R$ for any $u \in R_1$, and R was assumed to be empty. The first term is just n times the area of $R_1 + u$, which is $n \cdot \text{area}(R_1)$. Therefore,

$$\begin{aligned}\int_R f_i(u)D(u)du = \int_{R_1} n \cdot \text{area}(R_1)du &= n \cdot \text{area}(R_1)^2 \\ &= n(2^{-m-2})^2 \\ &\geq \frac{1}{16n},\end{aligned}$$

thereby completing the proof. □

References

[1] Jiri Matousek. *Geometric discrepancy: An illustrated guide*, volume 18. Springer Science & Business Media, 2009.

[2] Karl F Roth. On irregularities of distribution. *Mathematika*, 1(02):73–79, 1954.

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