CS 683: Advanced Design and Analysis of Algorithms	May 2, 2008

Lecture Notes 41

Professor: John Hopcroft *Scribes: Brook Li and Sucheta Soundarajan*

1 Minimizing Discrepancy in Regions Enclosed by Circles

How can we put points on the blackboard to minimize discrepancy?

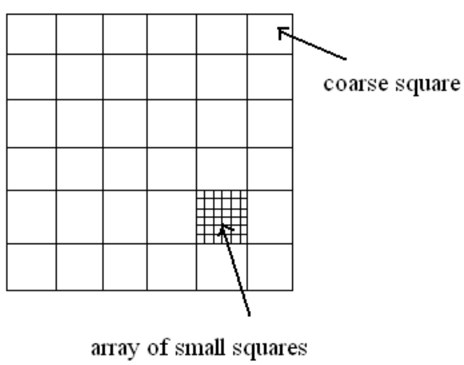
Theorem: There is set of n points P in the plane whose discrepancy is O($n^{1/4}\sqrt{\log n}$).

Proof: The points will be placed in a square area of the plane in the following manner:

1. Divide the square into an array of $\sqrt{n} \times \sqrt{n}$ 'coarse' squares.

2. Divide each such 'coarse' square into an $\sqrt{n} \times \sqrt{n}$ array of 'small' squares.

3. For each coarse square, select one small square at random and place a point in the center of that square.



Now let us consider circles inside the large square. A circular disk C completely includes some coarse squares, completely excludes coarse squares, and intersects the rest. The coarse squares which are completely inside or outside the disk contribute nothing to the discrepancy. If a coarse square is inside the disk, then it contributes the appropriate number of points (1) for its area inside the disk, and if it is outside the disk, it still contributes the appropriate number of points (0) for its area inside the disk. So we really only need to consider those coarse squares which the circle intersects. How many such coarse squares are there, and what is the discrepancy due to these coarse squares?

Let us first bound the number of coarse squares which intersect the boundary of the circle. Imagine starting at some point on the circle and traversing the circumference. Each coarse square on the boundary is entered through one of its four edges. At most $\sqrt{n}$ squares are entered from the top, and same for the bottom, left, or right. Since each square intersecting the boundary must be entered at some point, there are at most $4\sqrt{n}$ coarse squares on the boundary of the circle.

Now let's examine the discrepancy due to these squares on the boundary. Consider one coarse square intersected by the circle. Let Q be the set of small squares corresponding to that coarse square. Let $q \in Q$ be the small square in which the point is placed. Let k = $|Q \cap C|$, so k is the number of centers of small squares in Q which are contained inside C. So what is the contribution of q to $Q \cap C$? If $q \in C$, it contributes 1, otherwise it contributes 0.

Discrepancy is defined as the difference between the actual number of points contained within an area and the estimated number of points contained within that area. So if the point is inside the circle, then the discrepancy is 1 - $\frac{k}{n}$. This occurs with probability $\frac{k}{n}$. If the point is outside the circle, the discrepancy is $-\frac{k}{n}$. This occurs with probability 1 - $\frac{k}{n}$. The expected value of the discrepancy contributed by this square is thus 0.

The total discrepancy is then a random variable with expected value 0. What is the probability that the random variable differs from its expected value by more than Δ.

Define X = $x_1 + x_2 + ... + x_m$, where each x_i is the discrepancy due to coarse square i, and X is the total discrepancy. Then Prob(X $\geq \Delta$) $< e^{\frac{-2\Delta^2}{\sigma^2}}$. (Here, σ is the standard deviation). $\sigma^2 \leq m$, and $m = 4\sqrt{n}$. Then Prob(X $\geq \Delta$) $< e^{\frac{-2\Delta^2}{4\sqrt{n}}}$.

Let $\Delta = cn^{\frac{1}{4}}\sqrt{\log n}$. Then Prob(X $\geq \Delta$) $< e^{-2\frac{c^2\sqrt{n}\log n}{4\sqrt{n}}} = e^{-\frac{c^2 \log n}{2}} = n^{-\frac{c^2}{2}}$.

For a fixed disk, the probability that the discrepancy $< \Delta$ is non-zero. Take the union over all disks. Define two disks to be equivalent if they have the same set of small square centers inside them. There are $(n^2)^3$ such equivalence classes. Take the union over all equivalence classes. So then Prob(Discrepancy $> \Delta$) $< n^6 n^{-\frac{c^2}{2}}$. So if we pick $c^2 = 14$, we get that this probability is $\frac{1}{n}$.

2 Partial Shatter Functions

A partial shatter function of an integer n is the maximum number of different subsets of n points that can be enclosed with a particular bounding figure, such as a rectangle. That is, there is a different function for each figure being considered.

It can be shown that a partial shatter function of a figure grows exponentially from $n = 0$ to the VC dimension of the figure, and then grows only at the rate of a polynomial.

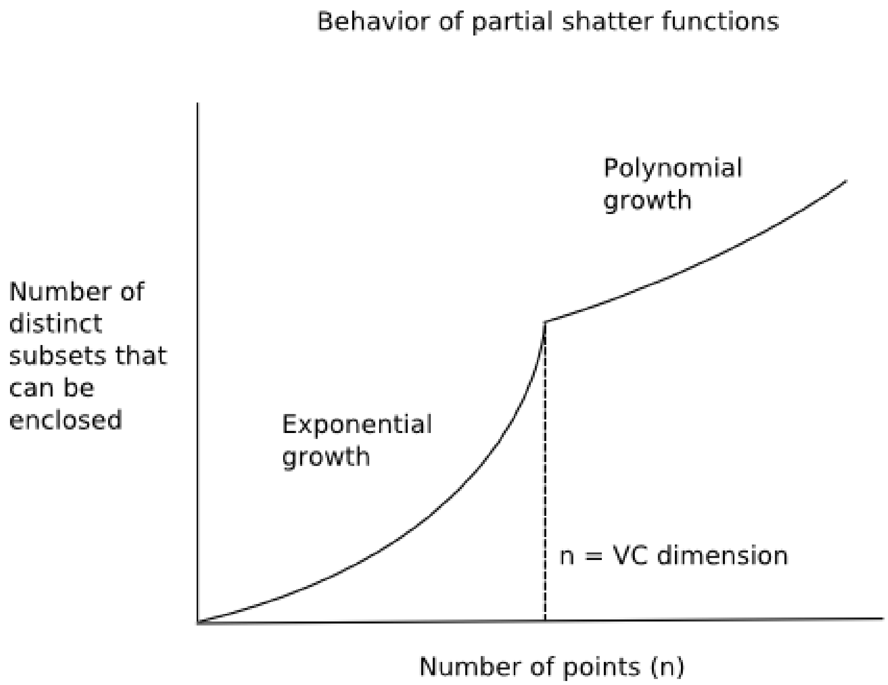


Figure 1: Diagram showing general behavior of partial shatter functions

A proof of the assertion about the growth rates of partial shatter functions can be done via induction over n and d, where d is the VC dimension of the figure, by using the following induction hypothesis:

$\pi(n) = \sum_{i=0}^{d-1} \binom{n-1}{i}$

Here π is the partial shatter function of a certain figure.