0368.416701 Sublinear Time Algorithms	January 11, 2010
Lecture 13	
Lecturer: Ronitt Rubinfeld	*Scribe: Ran Roth and Boaz Frankel*

1 Lecture Topic

In this lecture we present tests for uniformity assuming monotonicity of the probability distribution. These tests can be carried out in very low time and query complexity due to this assumption (that was not made in the last lecture). In the first part we consider a total order over the domain $[n] = \{1, ..., n\}$. In the second part we consider an example of a partial order – the binary hyper-cube.

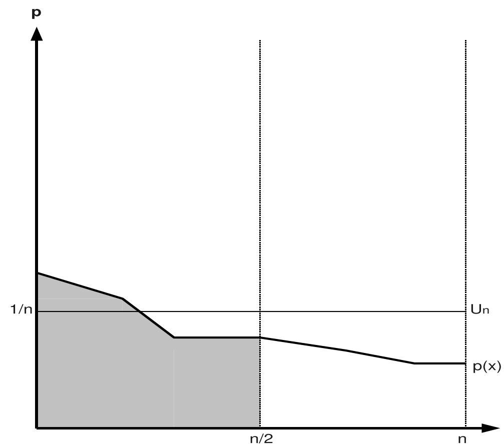
2 Uniformity Testing for monotone distributions over a totally ordered domain

Definition 1. We use the notation u_n to represent the uniform distribution on $D = [n] = \{1, ..., n\}$

Definition 2. Let p be a distribution over $D = [n]$. We say that p is **monotone** if $p_1 = p(1) \leq p_2 = p(2) \ldots \leq p_n = p(n)$

2.1 Intuition

Because the function is monotone, it is unbalanced: The first half of the domain must have higher probability than the second half (the gray area is always greater than the white area). So, if the distribution p is monotone, $\sum_{i=1}^{\frac{n}{2}} p(i)$ is bigger than $\sum_{i=\frac{n}{2}+1}^{n} p(i)$. If we also know that it is ϵ-far from uniform, we can show that this difference reaches a constant factor. This can be checked using sampling.



Lemma 3. If $\sum_{i=1}^{\frac{n}{2}} p(i) \leq (1+\varepsilon) \cdot \sum_{i=\frac{n}{2}+1} p(i)$ then $| p - u_n |_1 \leq \varepsilon$

Proof.

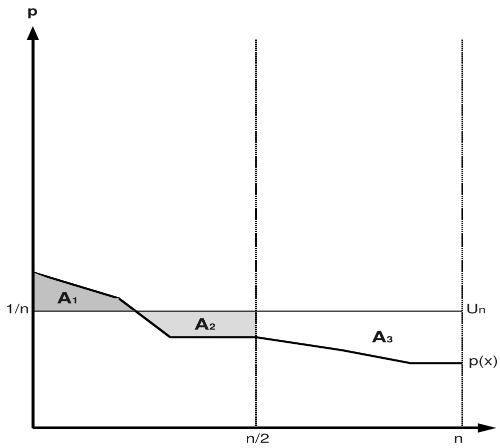
Definition 4. Given $i \in D$, we define the error at point i:

$$\delta_i = \mid p(i) - \frac{1}{n} \mid$$

Let j be the largest member in D such that $p(j) \geq \frac{1}{n}$.
We separate into two cases:

1. $j < \frac{n}{2}$
2. $j \geq \frac{n}{2}$

We give here the full proof for the case (1). Case (2) is left as an exercise and it is nearly symmetric. Assuming $j < \frac{n}{2}$, here is a schematic plot of the distribution function:



Definition 5. We define the three areas shown in the figure above as follows:
The area above uniformity:

$$A_1 \equiv \sum_{i<j} \delta_i$$

The area below uniformity in the first half of the domain:

$$A_2 \equiv \sum_{j \leq i \leq \frac{n}{2}} \delta_i$$

The area below uniformity in the second half of the domain:

$$A_3 \equiv \sum_{i > \frac{n}{2}} \delta_i$$

A few facts can be directly derived:

Fact 6. $\mid p - u \mid_1 = A_1 + A_2 + A_3$

Fact 7. $A_1 = A_2 + A_3$ (because $\sum_i p(i) = 1$, so A_1 and $A_2 + A_3$must be balanced)

Fact 8. $\sum_{i=1}^{\frac{n}{2}} p(i) = \frac{1}{2} + A_1 - A_2 = \frac{1}{2} + A_3$

We now estimate an upper bound for these sums:

$$A_3 : 1+\varepsilon \geq \frac{\sum_{i=1}^{\frac{n}{2}} p(i)}{\sum_{i=\frac{n}{2}+1}^{n} p(i)} = \frac{\frac{1}{2}+A_3}{\frac{1}{2}-A_3} \Longrightarrow \frac{1}{2} - A_3 + \frac{\varepsilon}{2} - \varepsilon \cdot A_3 \geq \frac{1}{2} + A_3 \Rightarrow A_3 \leq \frac{\varepsilon}{2\cdot(2+\varepsilon)}$$

$A_2 : A_2 \leq A_3$ because $\delta_j \leq \delta_{j+1} \ldots \leq \delta_n$ (p is monotone)

$A_1 : A_1 \leq 2 \cdot A_3$ because $A_1 = A_2 + A_3$

By summing these upper bounds we get an upper bound for $\mid p-u \mid_1$:
$\mid p-u \mid_1 = A_1 + A_2 + A_3 \leq 4 \cdot A_3 \leq 4 \cdot \frac{\varepsilon}{2\cdot(2+\varepsilon)} \leq \frac{2\varepsilon}{(2+\varepsilon)} = \frac{\varepsilon}{1+\frac{\varepsilon}{2}} \leq \varepsilon$

Using the same technique we can reach the same upper bound (ε) for the second case ($j \geq \frac{n}{2}$).

□

2.2 Algorithm

Using the lemma we have proved, it can be shown that the following algorithm is a uniformity tester for monotone probability distributions over $D = [n]$:

Given a monotone distribution p over $D = [n]$

- take a sample of m elements from p.
- $l \leftarrow \#\, of\ samples \in \left\{1, ..., \frac{n}{2}\right\}$
- $r \leftarrow \#\, of\ samples \in \left\{\frac{n}{2}+1, ..., 1\right\}$
- If $l > \left(1+\frac{\varepsilon}{2}\right) \cdot r$ REJECT
- otherwise ACCEPT

By applying Chernoff bounds one can prove that $m = \bigcirc\left(\frac{1}{\varepsilon^2}\right)$ ensures the algorithm works with probability $\geq \frac{3}{4}$:

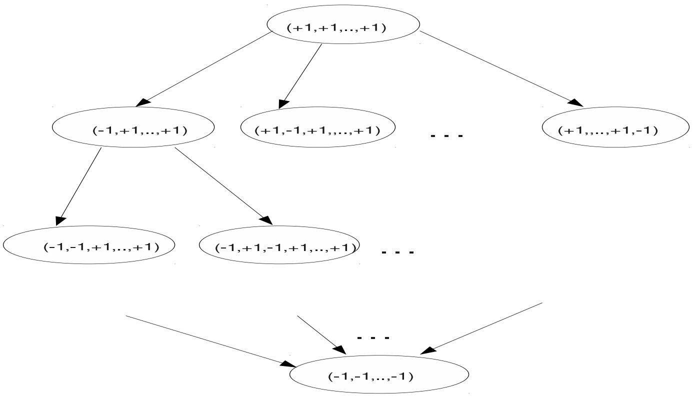
Theorem 9. Using $m \in \bigcirc\left(\frac{1}{\varepsilon^2}\right)$ the algorithm will work with probability $\geq \frac{3}{4}$ (i.e. accept the uniform distribution and reject distributions for which $\mid p - u_n \mid_1 > \varepsilon$).

3 Uniformity Testing for monotone distributions over the binary hyper-cube (partial order)

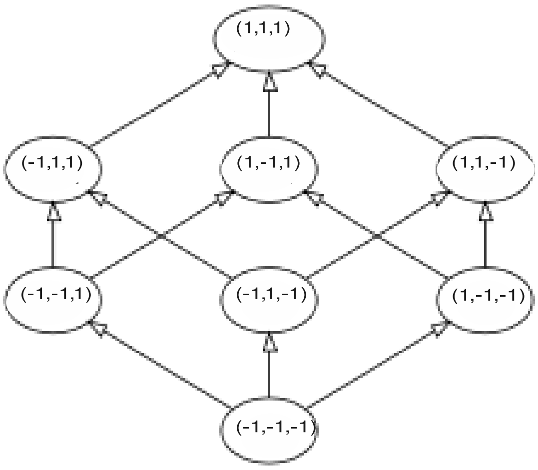
Definition 10. Let $D = \{\pm 1\}^B$ be the binary hyper-cube domain. We define a **partial order**:

$$x \leq y \Longleftrightarrow \forall i\, x_i \leq y_i$$

One can think of the members of the binary hyper-cube as all the subsets of $[B] = \{1, ..., B\}$. $+1$ in the i-th coordinate means i belongs to the subset, whereas -1 means it doesn't belong. From this point of view the partial order is equivalent to the $\subseteq$ ("is a subset of") relation. Another way to picture the hyper-cube partial order is a DAG (Directed Acyclic Graph) where in each edge exactly one coordinate changes from $+1$ to -1. $x \leq y$ if there is a descending path from y to x in the tree.



Example for $B = 4$



Definition 11. We say that a distribution p over domain $\{\pm 1\}^B$ is monotone if $x \leq y \Rightarrow p(x) \leq p(y)$

Definition 12. $bias(x) = \sum x_i$

For example in the domain $\{\pm 1\}^2$:
$bias((1,1)) = 2,\ bias((1,-1)) = 0,\ bias((-1,1)) = 0, bias((-1,-1)) = -2$

Definition 13. For the rest of the lecture we use n instead of B, where $n = \log_2 B \iff B = 2^n$

What is our expectation for the average *bias* when p is the uniform distribution?

$E_{p=u}[bias(x)] = \sum\limits_x \frac{1}{2^n} \cdot \sum\limits_i x_i = \frac{1}{2^n} \sum\limits_x \sum\limits_i x_i = 0$ (because the total number of $+1$ equals the total number of -1)

Corollary 14. We can estimate $\mid p-u\mid_1$ using an estimation of $E\left(bias\left(x\right)\right)$

Definition 15. We define $\delta\left\{\pm 1\right\}\to\mathbb{R}$: $\delta\left(x\right)=2^n p\left(x\right)-1$

Properties of δ:

1. δ is monotone
2. $\sum\limits_x \delta\left(x\right)=\sum\limits_x\left(2^n p\left(x\right)-1\right)=2^n\cdot\sum\limits_x\left(p\left(x\right)\right)-2^n=0$
3. $\sum\limits_x\mid\delta\left(x\right)\mid=\sum\limits_x\left(\mid 2^n p\left(x\right)-1\mid\right)=2^n\cdot\sum\limits_x\left(\mid p\left(x\right)-\frac{1}{2^n}\mid\right)=\underbrace{2^n\mid p-u\mid_1}_{we\ will\ call\ this\ \varepsilon}$
4. $E_p\left[bias\left(x\right)\right]=\sum\limits_x(p\left(x\right)\underbrace{\sum_i x_i}_{bias(x)})=\sum\limits_x\sum\limits_i\left(\underbrace{\frac{\delta\left(x\right)+1}{2^n}}_{p(x)}\cdot x_i\right)=\underbrace{\sum_x\sum_i\left(\frac{x_i}{2^n}\right)}_{0}\cdot\sum\limits_x\sum\limits_i\left(\frac{\delta(x)\cdot x_i}{2^n}\right)=\frac{1}{2^n}\sum\limits_x\sum\limits_i\delta\left(x\right)x_i$

As we did with the complete order, we differentiate the x-s for which $p\left(x\right)$ is above the uniform probability ($\equiv P$) from the ones for which $p\left(x\right)$ is below ($\equiv N$). The sum of probabilities of the two cases should be equal.

Definition 16. $P\equiv\left\{x\in\left\{\pm 1\right\}^n\mid\delta\left(x\right)\geq 0\right\}$

Definition 17. $N\equiv\left\{\pm 1\right\}^n\backslash P$

Fact 18. $\sum\limits_{x\in P}\mid\delta\left(x\right)\mid=\sum\limits_{x\in P}\delta\left(x\right)=\sum\limits_{x\in N}\mid\delta\left(x\right)\mid=\frac{\varepsilon}{2}\cdot 2^n$

We now define the sum S which we will later use to bound the *bias* with ε:

Definition 19. $\triangle\left(x,y\right)\equiv\mid\delta\left(x\right)-\underbrace{\delta\left(y\right)}_{<0}\mid=\mid\delta\left(x\right)\mid+\mid\delta\left(y\right)\mid$

Definition 20. $S=\frac{\sum\limits_{\substack{x\in P\\ y\in N}}\triangle\left(x,y\right)}{2^{2n-1}}$

We show that S is equal to ϵ.

Fact 21. $S=\frac{\sum\limits_{\substack{x\in P\\ y\in N}}\triangle\left(x,y\right)}{2^{2n-1}}=2^{1-2n}\cdot\left(\underbrace{\sum_{x\in P}\mid\delta\left(x\right)\mid}_{\frac{\varepsilon}{2}}+\underbrace{\sum_{y\in N}\mid\delta\left(y\right)\mid}_{\frac{\varepsilon}{2}}\right)=2^{1-2n}\cdot\left(\frac{\varepsilon}{2}\cdot 2^n\cdot\mid N\mid+\frac{\varepsilon}{2}\cdot 2^n\cdot\mid P\mid\right)=$

$2^{1-2n}\cdot\left(\underbrace{\mid N\mid+\mid P\mid}_{2^n}\right)\cdot\frac{\varepsilon}{2}\cdot 2^n=\varepsilon$

We would like to relate S to the expectation of the bias. We do this by considering the notion of canonical paths. First we define a path in the binary hyper-cube. We do so by looking at the graph induced by the $\geq$ relation as an undirected graph. Formally:

Definition 22. Let $x, y \in \{0,1\}^n$ be two vertices of the binary hyper-cube. A *path between x and y is an ordered list $(z_0 = x, ..., z_k = y)$ where $\forall j, z_j \in \{0,1\}^n$ and z_j, z_{j+1} differ by exactly one bit.*

Given our distribution p, we define the distance of a path:

Definition 23. Let z_{xy} be a path between x and y.

$$dist(z_{xy}) := \sum_{i=0}^{k-1} |\delta(z_i) - \delta(z_{i+1})|$$

Claim 24. Let $x, y \in \{0,1\}^n$. For any path z_{xy} between x and y, $dist(z_{xy}) \geq \Delta(x, y)$

Proof. Straightforward from the triangle inequality:

$$\Delta(x,y) = |\delta(x) - \delta(y)| = |\delta(z_0) - \delta(z_k)| \leq \sum_{i=0}^{k-1} |\delta(z_i) - \delta(z_{i+1})| = dist(z_{xy})$$

□

Our strategy will be to choose canonical paths that have a property of uniformity over the use of the edges - there is a bound on the maximal paths traversing a single edge.

Definition 25. Let $x = (x_1, ..., x_n), y = (y_1, ..., y_n)$ be two vertices in the binary hyper-cube. We look at the following list of vertices:

$$\begin{aligned} Z_0 &= (x_1, x_2, x_3, ..., x_n) \\ Z_1 &= (y_1, x_2, x_3, ..., x_n) \\ Z_2 &= (y_1, y_2, x_3, ..., x_n) \\ &... \\ Z_n &= (y_1, y_2, y_3, ..., y_n) \end{aligned}$$

In this list of vertices **at most** one bit changes in the transition between two successive elements. We remove the redundant edges, so that we get exactly **exactly** one bit changing in each transition. We define the resulting path as **the canonical path Z_{xy} between x and y.**

For example, suppose $x = (+1, -1, +1, -1, -1)$ and $y = (+1, +1, -1, +1, -1)$ then the canonical path between x and y is:

$$\begin{aligned} x = Z_0 &= (+1, -1, +1, -1, -1) \\ Z_1 &= (+1, +1, +1, -1, -1) \\ Z_2 &= (+1, +1, -1, -1, -1) \\ y = Z_3 &= (+1, +1, -1, +1, -1) \end{aligned}$$

We now show a bound on the number of canonical paths traversing an edge in the hyper-cube graph:

Claim 26. Let $z = (z, z^{\oplus i})$ be an edge (where the i-th bit is the one that changes). Then z is traversed by 2^n canonical paths.

Proof. Let Z_{xy} be a canonical path traversing the edge z. Then by definition:

$$\begin{aligned} x_0 &= z_0 \\ x_1 &= z_1 \\ &\dots \\ x_{i-1} &= z_{i-1} \\ y_i &= z_i \\ y_{i+1} &= z_{i+1} \\ &\dots \\ y_n &= z_n \end{aligned}$$

The canonical path is uniquely determined by its end-points. So, x has the freedom to choose its $i, ..., n$ coordinates whereas y has the freedom to choose its $1, ..., i-1$ coordinates. (Note that y_i is fixed once we choose x_i, since it's known to be its negation). Choosing freely n coordinates gives 2^n possibilities for x, y and hence 2^n canonical paths going through z. □

Now we are ready to see the connection with the expected bias:

Lemma 27. For a distribution p that is ϵ-far from uniform,

$$Exp_p[bias(x)] \geq \frac{\epsilon}{2}$$

Proof.

$$\epsilon = S = 2^{1-2n} \sum_{x \in P, y \in N} \Delta(x, y)$$

By claim (24):

$$\epsilon \leq 2^{1-2n} \sum_{x \in P, y \in N} \sum_{i=0}^{|Z_{xy}|-1} |\delta(z_{xy}^{(i)}) - \delta(z_{xy}^{(i+1)}|$$

Since we sum on edges, we can bound this sum by multiplying the weight of each edge by the maximal number of times we cross it. Since the graph is undirected, we can assume without loss of generality that all edges (u, v) have $u \geq v$

$$\epsilon \leq 2^{1-2n} \sum_{(u,v) s.t. u_i=1, v=u^{\oplus i}} |\delta(u) - \delta(v)| \times (\text{Max num. of times } (u, v) \text{is traversed})$$

Because we use canonical paths, by claim(26) we have an upper bound on the number of times each edge is traversed by paths (this upper bound is for all paths, not only ones going from P to N, but this is good enough for our purpose). So

$$\epsilon \leq 2^{1-2n} \sum_{(u,v) s.t. u_i=1, v=u^{\oplus i}} |\delta(u) - \delta(u^{\oplus i})| \times 2^n$$

By the monotonicity of p, we can remove the absolute value:

$$\epsilon \leq 2^{1-n} \sum_{(u,v) s.t. u_i=1, v=u^{\oplus i}} \delta(u) - \delta(u^{\oplus i})$$

We separate the sum of edges to the single component that changes in this edge. The inner sum can go over all the vertices x, summing for each one the edge changing its i-th component with the correct sign:

$$\epsilon \leq 2^{1-n} \sum_{i=1}^{n} \sum_{x \in \{0,1\}^n} x_i \delta(x)$$

But this is just the formula of the bias expectation. So:

$$\frac{\epsilon}{2} \leq Exp_{x \in p}[bias(x)]$$

□

Using the above theorem and by applying the Chernoff bounds one can prove the following theorem:

Theorem 28. The following algorithm is a probabilistic tester for uniformity of a distribution over the binary hyper-cube. It accepts with high probability the uniform distribution and fails with high probability distributions that are ϵ-far from uniform in $L1$ metric.

- Pick $s = O(\frac{n}{\epsilon^2} log(\frac{n}{\epsilon}))$ samples from the distribution p. Let $x^{(1)}, ... x^{(s)}$ be the picked samples.
- If for any of the samples $x^{(i)}$ it holds that $|bias(x^{(i)}| \geq \sqrt{2nlog(20s)}$, stop and fail.
- Let $\mu = \frac{1}{s} \sum_{i=1}^{s} bias(x^{(i)})$
- If $\mu \leq \frac{\epsilon}{4}$, pass. Otherwise fail.

The second step is necessary in order to improve the bound on the maximal value used in the Chernoff inequality. However, recent work shows that one may be able to avoid it.