6.896 Sublinear Time Algorithms February 15, 2007

Lecture 4

Lecturer: Ronitt Rubinfeld *Scribe: Huy Nguyen*

1 Testing uniformity of a monotone distribution over a totally ordered set

Definition 1 *Let $D = 1, 2, .., n = [n]$. D is <u>monotone</u> if $p_1 \geq p_2 \geq \cdots \geq p_n$.*

We want to test whether a monotone distribution over a totally ordered set is uniform or far from uniform. Since all totally ordered sets are equivalent, for simplicity, we assume that the domain is [n]. The key idea of the algorithm is to take samples and check if the number of sample in the left half is close to the number of sample in the right half.

Lemma 2 *If $p(\{1 \cdots k\}) \leq (1+\epsilon)p(\{k+1 \cdots 2k\})$ then*

$$|p - U_{[n]}| \leq \epsilon$$

Proof $\delta_i \leftarrow |p_i - \frac{1}{n}|$
$j \leftarrow$ largest i s.t. $p_i \geq \frac{1}{2k} = \frac{1}{n}$

We only consider the case when $j \leq k$. The case when $j > k$ is similar.

We define A_1, A_2, A_3 as follow:

$$A_1 = \sum_{i \leq j} \delta_i$$

$$A_2 = \sum_{j < i \leq k} \delta_i$$

$$A_3 = \sum_{n \geq i > k} \delta_i$$

$A_1 = A_2 + A_3$ because the sum of distribution p and the uniform distribution are both 1.

$$p(\{1 \cdots k\}) = \frac{1}{2} + A_1 - A_2 \tag{1}$$

$$p(\{k+1 \cdots n\}) = \frac{1}{2} - A_3 \tag{2}$$

Substitute (1) and (2) into the assumption of the lemma, we get:

$$1 + \epsilon \geq \frac{\frac{1}{2} + A_1 - A_2}{\frac{1}{2} - A_3} = \frac{\frac{1}{2} + A_3}{\frac{1}{2} - A_3}$$

Solve for A_3, we get:

$$A_3 \leq \frac{\epsilon}{4 + 2\epsilon}$$

A_2 summed over fewer elements than A_3 ($< k$ vs $= k$). Since $\delta_{j+1} \leq \delta_{j+2} \leq \cdots$, each element in A_2 is smaller then any element in A_3. Therefore,

$$A_2 \leq A_3 \leq \frac{\epsilon}{4 + 2\epsilon}$$

$$A_1 = A_2 + A_3 \leq \frac{\epsilon}{2 + \epsilon}$$

Hence,

$$|p - U_{[n]}| = A_1 + A_2 + A_3 \leq \frac{\epsilon}{1+\frac{\epsilon}{2}} \leq \epsilon$$

■

Using the lemma, we have the following algorithm for testing uniformity:

Test-Uniformity(p)

```
Pick m samples from p, m = Θ(ε^{-2})
l ← number of samples in 1···k
r ← number of samples in k+1···n
if l ≥ (1 + ε/8) r
   then return fail
   else return pass
```

Now we analyze the algorithm. We need to show that it return the correct result both when P is uniform and when P is far from uniform.

Case 1: P is uniform. $E[l] = E[r] = m/2$.

Case 2: P is ϵ far from uniform.

$$\begin{aligned} p(\{1\cdots k\}) &\geq p(\{k+1\cdots n\})(1+\epsilon) \\ E[l] &\geq m(\frac{1}{2} + \frac{\epsilon}{4+2\epsilon}) \end{aligned}$$

The rest of the proof is just computing the variance and applying Chernoff bound to get the desired results.

2 Testing uniformity of a monotone distribution on Boolean cube

In the previous section, we had an algorithm that can, in constant time, test the uniformity of a monotone distribution on a totally ordered set. It is desirable to generalize the problem to partially ordered set. In this section, we will consider the special case of that general problem. The partially ordered set we will look at is a Boolean cube.

Definition 3 *A Boolean cube is the set* $\{\pm 1\}^n$ *for some* n.

Definition 4 $x, y \in \{\pm 1\}^n$. x *is <u>less than or equal to</u>* y *iff* $\forall i\ x_i \leq y_i$.

Definition 5 *A distribution* P *is <u>monotone</u> iff* $\forall x \leq y, p_x \leq p_y$

Notation.

$$\begin{aligned} bias(x) &= \sum_i x_i \\ \delta(x) &= 2^n p(x) - 1 \\ \epsilon &= |p - U|_1 \end{aligned}$$

Note $\frac{\delta(x)}{2^n} = p(x) - \frac{1}{2^n}$

Observation 6 δ *is monotone.*

Observation 7 $\sum_x \delta(x) = \sum_x (2^n p(x) - 1) = 2^n(\sum_x p(x)) - 2^n = 0$.

Observation 8 $\sum_x |\delta(x)| = 2^n |p - U|_1 = 2^n \epsilon$

Now we compute the expectation of the bias:

$$
\begin{aligned}
E_p[bias] &= \sum_x p(x) \sum_{i=1}^{n} x_i \\
&= \sum_x \sum_{i=1}^{n} (\frac{\delta(x)+1}{2^n} x_i
\end{aligned}
$$

Since $\sum_{i=1}^{n} \sum_x x_i = 0$, we have:

$$
\begin{aligned}
E_p[bias] &= \sum_x \sum_{i=1}^{n} \frac{\delta(x)}{2^n} x_i \\
&= \underbrace{\frac{1}{2^n} \sum_x \sum_{i=1}^{n} \delta(x) x_i}_{*}
\end{aligned}
$$

Lemma 9 $E_p[bias] \geq \epsilon$.

Proof Let us define a few quantities:

$$
\begin{aligned}
P &= \{x | \delta(x) \geq 0\} \\
Q &= \{x | \delta(x) < 0\}
\end{aligned}
$$

Proof will be continued next time.