6.896 Sublinear Time Algorithms	February 22, 2007

Lecture 6

Lecturer: Ronitt Rubinfeld *Scribe: Ning Xie*

In this lecture, we continue studying the problem of testing whether a monotone distribution is uniform. We will restrict to distributions over the Boolean cube. First recall some notation. Let $p : \{-1,1\}^n \to \mathbb{R}^{\geq 0}$ be a monotone distribution over the Boolean cube. Define $\delta(x) = 2^n p(x) - 1$, that is, the weight deviation from uniform distribution. Note that since $p(x)$ is monotone, so is $\delta(x)$. As we proved last time, $\sum_x \delta(x) = 0$ and $\sum_x |\delta(x)| = 2^n\epsilon$, where ϵ is the ℓ_1-distance between p and uniform distribution, $\epsilon = |p - U|_1$. We also define, for each $x \in \{-1,1\}^n$, the bias of x to be $bias(x) = \sum_{i=1}^n x_i$. It is easy to see that for uniform distribution $\mathbb{E}_U[bias(x)] = 0$. Our main result for this lecture is the follow lemma:

Lemma 1 *Let p be a monotone distribution over Boolean cube and $\epsilon = \sum_{x\in\{-1,1\}^n} |p(x) - \frac{1}{2^n}|$, then*

$$\mathbb{E}_p[bias(x)] \geq \epsilon.$$

Proof Let $S = \mathbb{E}_p[bias(x)]$, our goal is to obtain a lower bound for it in term of ϵ. Since $p(x) = \frac{1}{2^n} + \frac{\delta(x)}{2^n}$ and the expectation of bias(x) over the uniform distribution is zero, we can rewrite S as

$$S = \sum_x \sum_{i=1}^n x_i p(x) = \frac{1}{2^n} \sum_x \sum_{i=1}^n x_i \delta(x).$$

Our proof strategy is to first transform this sum over nodes to a sum over the edges in the Boolean cube. Then we get a lower bound of the sum over the edges by considering a so-called "canonical path". To get an idea of how the first step works, consider the 2-dimensional Boolean cube with $A = (1,1), B = (1,-1), C = (-1,1)$ and $D = (-1,-1)$. Then we have

$$\begin{aligned} 2^n S =& \delta_A 1 + \delta_A 1 + \delta_B 1 + \delta_B(-1) + \delta_C(-1) + \delta_C 1 + \delta_D(-1) + \delta_D(-1) \\ =& (\delta_A - \delta_B) + (\delta_A - \delta_C) + (\delta_B - \delta_D) + (\delta_C - \delta_D). \end{aligned}$$

If we assign the weight $\delta(u) - \delta(v)$ to each edge $(u,v) \in E$, then S can be expressed as a sum of the weights of all edges in the Boolean cube[1]. Also note that we only sum over *downward* edges, that is, edges (u,v) with $u > v$.

Next we define two sets of points:

$$P = \{x | \delta(x) \geq 0\}$$

and

$$N = \{x | \delta(x) < 0\}.$$

For each pair of points x and y with $x \in P$ and $y \in N$, define

$$\begin{aligned} \Delta(x,y) &= |\delta(x) - \delta(y)| \\ &= |\delta(x)| + |\delta(y)|, \end{aligned}$$

because $\delta(y) < 0$. We have

$$\begin{aligned} \sum_{x\in P, y\in N} \Delta(x,y) &= \sum_{x\in P, y\in N} \delta(x) + \sum_{x\in P, y\in N} \delta(y) \\ &= |P|\frac{\epsilon}{2}2^n + |N|\frac{\epsilon}{2}2^n = \frac{\epsilon}{2}2^{2n}. \end{aligned}$$

[1]Note that the Boolean cube naturally gives rise to a graph: the vertices are the 2^n elements in $\{-1,1\}^n$ and there is an edge between vertices x and y if and only if they differ at exactly one coordinate.

Let $x, y \in \{-1,1\}^n$. Let $(z_0 = x, z_1, \ldots, z_{k-1}, z_k = y)$ be a path from x to y along the edges of the Boolean cube, where $(z_i, z_{i+1}) \in E$, for $0 \le i \le k-1$. The "z-distance path between x and y" is defined to be

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

where the last step follows from triangle inequality.

Now we can rewrite the sum over distance pairs in P and N as

$$\begin{aligned}\frac{\epsilon}{2}2^{2n} = \sum_{x\in P, y\in N} \Delta(x,y) &\le \sum_{x\in P, y\in N} dist_{z_{xy}}(x,y) \qquad \text{(for arbitrary choice of path } z_{xy}\text{)}\\ &= \sum_{x\in P, y\in N} \sum_{j=0}^{|z_{xy}-1|} |\delta(z_{xy}^j) - \delta(z_{xy}^{j+1})|.\end{aligned}$$

The key point here is to pick a routing path so that there is no edge get crossed too many times.

Now we introduce a very useful concept called "canonical path". For $x \in P$ and $y \in N$, z_{xy} is the canonical path from x to y obtained by scanning from left to right and flip bits to match y whenever the scanned bit of current string is different from that of y. Here is an example:

$$\begin{aligned} x =& (1, -1, 1, 1, -1)\\ & (1, 1, 1, 1, -1)\\ & (1, 1, -1, 1, -1)\\ y =& (1, 1, -1, 1, 1).\end{aligned}$$

Now let us count the number of downward paths between $x \in P$ and $y \in N$ crossing a given edge. A naive bound is 2^{2n}. A more careful analysis is the following. Suppose the edge is $(a1b, a-1b)$, where $a \in \{-1,1\}^k$ and $b \in \{-1,1\}^{n-k-1}$. Then according to the definition of canonical path, the starting vertex must be of the form $(\alpha 1 b)$ and the ending vertex must be of the form $(a-1\beta)$, where $\alpha \in \{-1,1\}^k$ and $\beta \in \{-1,1\}^{n-k-1}$. Then the number of downward paths crossing the edge is simply the total number of choices of strings α and β, which is 2^{n-1}.

Now we have

$$\begin{aligned}\frac{\epsilon}{2}2^{2n} =& \sum_{x\in P, y\in N} \Delta(x,y)\\ =& \sum_{x\in P, y\in N} \sum_{j=0}^{|z_{xy}|-1} |\delta(z_{xy}^j) - \delta(z_{xy}^{j+1})|\\ \le& \sum_{(u,v)\in E} |\delta(u) - \delta(v)| \cdot (\text{max. \# of times } (u,v) \text{ used in any canonical path})\\ =& 2^{n-1} \sum_{(u,v)\in E} |\delta(u) - \delta(v)|.\end{aligned}$$

Note that, since p is monotone and we are summing over downward edges, $\delta(u) \ge \delta(v)$. Therefore we have

$$\begin{aligned}
\epsilon \leq & \frac{1}{2^n} \sum_{(u,v)\in E} |\delta(u) - \delta(v)| \\
= & \frac{1}{2^n} \sum_{i=1}^{n} \sum_{(u,v)\in E_i} |\delta(u) - \delta(v)| \\
= & \frac{1}{2^n} \sum_{i=1}^{n} \sum_{(u,v)\in E_i} (\delta(u) - \delta(v)) \\
= & \frac{1}{2^n} \sum_{i=1}^{n} \sum_{x\in\{-1,1\}^n} x_i \delta(x) \\
= & \mathbb{E}[bias(x)].
\end{aligned}$$

This completes the proof of the lemma. ■

The testing algorithm is the following

Testing Uniformity of Monotone Distributions over Boolean Cube
Pick s samples $X^{(1)}, \ldots, X^{(s)} \in p$, with $s = \Theta(\frac{n}{\epsilon^2} \log \frac{n}{\epsilon})$
Fail if any $X^{(i)}$ has $\lvert bias(X^{(i)})\rvert \geq \sqrt{2n \log(20s)}$
Let $\hat{\mu} = \frac{1}{s} \sum_{i=1}^{s} bias(X^{(i)})$
If $\hat{\mu} \geq \frac{\epsilon}{4}$ fail
Else pass

The tricky part of the algorithm is in the first step. We check all the sample strings to see if anyone has large bias. If so, we fail immediately. This is for a technical reason. Due to the following additive Chernoff bound, in order to reduce the number of samples needed to get desired concentration around mean, we need to reduce the range of the i.i.d. random variables. Note that this is essentially the best we can do: if we further reduce the threshold to a smaller value, we will not be able to let uniform distribution pass the test with high probability.

Theorem 2 (Additive Chernoff bound, a.k.a. Hoeffding bound) *Let $X_1, \ldots, X_m$ be i.i.d. random variables ranging in $[-a, a]$. Let $\hat{\mu} = \frac{1}{m} \sum_{i=1}^{m} X_i$ and $\mu = \mathbb{E}[\hat{\mu}]$. Then for all $\rho > 0$, we have $\Pr[|\hat{\mu} - \mu| \geq \rho] \leq 2 \cdot e^{-\frac{\rho^2}{2a^2} m}$.*

Now we will show that if p is the uniform distribution, then the testing algorithm will pass with high probability. The analysis for distributions that are far from uniform is left for next lecture. The intuition is, if p is the uniform distribution, then for large enough n, the distribution of $bias(x)$ approaches a Gausssian distribuion with mean 0 and variance $\Theta(n)$. Therefore, $bias(x)$ is concentrated in $[-\Theta(\sqrt{n}), \Theta(\sqrt{n})]$ and we are unlikely to see any sample with large bias.

Formally, let s be the number of samples and define $\rho = \sqrt{\frac{2\log(20s)}{n}}$. Let $X^{(i)}$ be any sample and consider the random variable $Y_1, \ldots, Y_n$ to be the n bits of $X^{(i)}$. Applying the additive Chernoff bound with $a = 1$ and $m = n$ gives

$$\Pr[|bias(X^{(i)})| \geq \sqrt{2n\log(20s)}] = \Pr[|bias(X^{(i)})| \geq \rho n] = \Pr[\frac{1}{n}|bias(X^{(i)})| \geq \rho] \leq 2e^{-\rho^2 n/2} = \frac{1}{10s}$$

By union bound, $\Pr[\text{there is an } i \text{ such that } bias(X^{(i)}) \geq \sqrt{2n\log(20s)}] \leq \frac{1}{10}$. Therefore, if p is the uniform distribution, the probability that step 1 outputs fail is at most $1/10$.

If we reach step 3, then there is no sample has large bias. Let $D' = \{x \in D | |bias(X^{(i)})| < \sqrt{2n\log(20s)}\}$ and $p' = p_{|D}$ be the restriction of p to D'. Since p is uniform over D, p' is uniform over D'. Then by symmetry, $\mathbb{E}_{p'}[bias(x)] = 0$. Apply the additive Chernoff bound again, but this time set $a = \sqrt{2n\log(20s)}$, $\rho = \frac{\epsilon}{4}$ and $m = s$, we have

$$\Pr[\bar{\mu} \geq \frac{\epsilon}{4}] \leq 2e^{-\epsilon^2 s/64n\log(20s)} < \frac{1}{10}.$$

Combined with the previous calculation, we see that the total probability of output fail is at most 1/5.