Lecture 6: Proving Lower Bound of Uniformity Testing

Daniel Kane
Scribe: Jongha Ryu

Apr 14, 2017

1 Lower Bounds

Again we will use adversary methods with the help of information theory to prove lower bounds.

1.1 Uniformity testing

Generate a random bit $X \sim \text{Bern}(1/2)$, and take

$$p = \begin{cases} u_n & \text{if } X = 0, \\ \text{a distribution far from } u_n & \text{if } X = 1. \end{cases}$$

Then we take Poisson(m) samples from p, and count A_i, the number of samples from i-th bin. (By symmetry, the order of samples does not matter in i.i.d. sampling. In other words, the counts A_i's are sufficient statistic.) Now suppose we have an algorithm f which takes $A^n = A_1, \ldots, A_n$ as an input and $f(A^n) = X$ w.p. $2/3$. (In other words, f is a uniformity tester with probability of success $2/3$.) Then

$$I(X; A^n) \geq I(X; f(A^n)) = H(X) - H(X \,|\, f(A^n)) > \Omega(1),$$

because X and $f(A^n)$ agree w.p. $2/3$.[1]

Now we need to specify a distribution Alice will use when $X = 1$. We first argue that we cannot use a deterministic distribution far from u_n. If that is the case, we only need $O(1/\epsilon^2)$ samples to distinguish between u_n and p. Why? Basically what our algorithm

[1] Let X, Y be binary random variables and assume $\mathsf{P}\{X = Y\} = p$. Let $Z = \mathbb{1}\{X = Y\}$. Then

$$H(X|Y) = H(X, Z|Y) = H(X|Y, Z) + H(Z|Y) = H(Z|Y) \leq H(Z) = h(p).$$

does is to check for every bin whether there are more than or less than $1/n$ fraction of samples. If the distribution p is known to our algorithm, then the algorithm already knows which case we are looking for, and thus every individual sample will give us an information.

Now we construct a class of distributions so that the average of distribution is the uniform distribution u_n. Also we want to make sure that a random p is ϵ-far from u_n in total variation. Suppose we pick p from $\mathcal{D}$ over distributions, where $p_i = \frac{1\pm\epsilon}{n}$ randomly and indpendently from each other. The problem is that p is a pseudo-distribution, i.e., not normalized. Later when we want to bound $I(X; A^n)$, we need A_i's are conditionally independent on X, so we want p_i's to be chosen independently. Thus, this problem is unavoidable.

How can we fix it? Simply we are going to allow p to be a non-normalized distribution, but we do not know how to sample from a non-normalized p! Fortunately, this can be fixed by Poissonization. We can still do Poisson sampling from each bin according to p even if $\|p\| \neq 1$. Then we have $A_i \sim \text{Poisson}(mp_i)$ independently each other. Note that this sampling is equivalent to taking $\text{Poisson}(m\|p\|_1)$ samples from a normalized distribution $\tilde{p} = p/\|p\|_1$, because

$$\text{Poisson}(mp_i) = \text{Poisson}\left(m\|p\|_1 \frac{p_i}{\|p\|_1}\right).$$

To make it work properly, we need

- $\|p\|_1 = \Omega(1)$ w.p. 99%. (If $\|p\|_1$ is too small, this might make $m\|p\|_1$ very small, then the lower bound will not prove anything.)
- $\|p/\|p\|_1 - u_n\|_1 \gg \epsilon$ w.p. 99%.

Assuming these hold, let $S = \sum_i A_i$ be the number of samples we actually took. Note that conditioned on S, normalization no longer matters. Now consider

$$\begin{aligned}
I(X; A^n) &= I(X; A^n, S) && (1.1)\\
&= I(X; S) + I(X; A^n | S) && (1.2)\\
&\gg I(X; A^n | S) && (1.3)\\
&\geq \sum_{s=0}^{\infty} \mathsf{P}\{S = s\} I(X; A^n | S = s) && (1.4)\\
&= \sum_{s=0}^{\infty} \mathsf{P}\{S = s\} I(X; s \text{ samples from } \tilde{p}) && (1.5)\\
&\gg I(X; m \text{ samples from } \tilde{p}). && (1.6)
\end{aligned}$$

Note that (1.6) follows if we assume $S > m$ w.p. $1/2$, because adding samples would only increase mutual information. Therefore, the upshot of this calculation is that $I(X; A^n)$ can be lower bounded by a constant even when we do a sampling from a pseudo-distribution.

Going back to our specific choice $p_i = (1 \pm \epsilon)/n$, first note that $\|p\|_1 = \Theta(1)$ always. Hence, the number of samples we are going to take is $\gg m$ w.h.p. To check that $\tilde{p}$ is far from u_n, notice that w.h.p.

$$\|p\|_1 \in \left(1 - \frac{\epsilon}{2}, 1 + \frac{\epsilon}{2}\right).$$

This implies that for each i

$$\left|\frac{p_i}{\|p\|_1} - \frac{1}{n}\right| \gg \frac{\epsilon}{n},$$

and therefore

$$\|\tilde{p} - u_n\|_1 \gg \epsilon.$$

Now we finally find an upper bound of $I(X; A^n)$. Recall that A_i's are conditionally independent on X, because p_i's are conditionally independent on X and $A_i \sim \text{Poisson}(mp_i)$. Therefore,

$$I(X; A^n) \leq \sum_{i=1}^{n} I(X; A_i) = nI(X; A_1).$$

Now we abstract our situation further. We have a uniform random bit X, and introduce a new random variable

$$A \sim \begin{cases} p & \text{if } X = 0; \\ q & \text{if } X = 1. \end{cases}$$

Here, we would like to find $I(X; A)$. We know that $\mathsf{P}\{A = k\} = (p_k + q_k)/2$, and

$$X|\{A = k\} = \begin{cases} 0 & \text{w.p. } \frac{p_k}{p_k + q_k}, \\ 1 & \text{w.p. } \frac{q_k}{p_k + q_k}. \end{cases}$$

Hence,

$$\begin{aligned} I(X; A) &= H(X) - H(X|A) \\ &= 1 - \sum_k \frac{p_k + q_k}{2} h\left(\frac{p_k}{p_k + q_k}\right) \\ &= \sum_k \frac{p_k + q_k}{2}\left(1 - h\left(\frac{p_k}{p_k + q_k}\right)\right) \\ &= \Theta\left(\sum_k \frac{(p_k - q_k)^2}{p_k + q_k}\right). \end{aligned}$$

Here, $h\colon [0, 1] \to [0, 1]$ is the binary entropy function, and we use the fact that $1 - h(p) = \Theta\left((1/2 - p)^2\right)$.

Back to our problem, we have

$$p = \text{Poisson}\left(\frac{m}{n}\right), \quad q = \frac{1}{2}\text{Poisson}\left((1 + \epsilon)\frac{m}{n}\right) + \frac{1}{2}\text{Poisson}\left((1 - \epsilon)\frac{m}{n}\right), \tag{1.7}$$

and let $\lambda = m/n$ for simplicity. Accordingly, for each k we would get

$$p_k = \frac{e^{-\lambda}\lambda^k}{k!}, \quad q_k = \frac{e^{-\lambda}\lambda^k}{k!}\left[\frac{e^{-\lambda\epsilon}(1+\epsilon)^k + e^{\lambda\epsilon}(1-\epsilon)^k}{2}\right].$$

Plugging these into the last expression,

$$I(X;A) \le \Theta\left(\sum_k \frac{(p_k - q_k)^2}{p_k}\right) = \Theta\left(\sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!}\left[\frac{e^{-\lambda\epsilon}(1+\epsilon)^k + e^{\lambda\epsilon}(1-\epsilon)^k}{2} - 1\right]^2\right). \tag{1.8}$$

Now assuming $\lambda\epsilon \ll 1$, we can upper bound the squared term by Taylor expansion. Since the constant and the linear term cancel out, we can bound the term by the quadratic term

$$O\left([\lambda^2 + \lambda k + k(k-1)]\epsilon^2\right).$$

If $\lambda < 1$, then the (R.H.S.) of (1.8) becomes

$$\mathsf{E}_{K\sim\text{Poisson}(\lambda)}\left[\epsilon^4\left(\lambda^2 + \lambda K + K(K-1)\right)^2\right] = O(\epsilon^4\lambda^2).$$

Hence, the total mutual information $I(X;A^n) = O(\epsilon^4 m^2/n)$, and finally we conclude that

$$m \gg \frac{\sqrt{n}}{\epsilon^2}.$$

If $\lambda > 1$, we need a different analysis. Consider the term $e^{\lambda\epsilon}(1-\epsilon)^k = \exp(\epsilon(\lambda - k) + O(k\epsilon^2))$. Note that we only need to care $k \in \lambda \pm \sqrt{\lambda}$ for Poisson distribution. Hence, the term $k\epsilon^2 \sim m\epsilon^2/n \ll 1/\sqrt{n}$ will be small around this range, so it is negligible. Finally, the squared term in the expression becomes

$$\cosh\left(\epsilon(\lambda - k)\right) - 1 \simeq \epsilon^2(\lambda - k)^2,$$

and finally the (R.H.S.) of (1.8) becomes

$$\mathsf{E}_{K\sim\text{Poisson}(\lambda)}\left[\epsilon^4(\lambda - K)^4\right] = O(\epsilon^4\lambda^2),$$

which implies the same bound.