Lecture 6: Uniformity Testing

Gabriel Ibagon

April 14, 2017

1 Introduction

We went over property testing algorithms for uniformity, identity, closeness, and independence. The bounds that we proved were all tied to constants, but we didn't actually prove any lower bounds. Today we will go over these lower bounds.

1.1 Approach

We will again be using an adversary method. However, instead of having many possibilities that you would like to extract information about, we only have two possibilities: 1) satisfying the property or 2) far from satisfying the property.

2 Uniformity Testing

We are going to start with the case of uniformity testing.

Let X be a random bit.

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

where u_n is the uniform distribution.

Here we take Poisson(m) samples from p, and count the number of samples in each bin, which we will call A_i.

Note: Notice the symmetry of this distribution. The order of the samples doesn't matter, and doesn't tell you anything about p. We can assume, that the algorithm doesn't see the unordered samples, but rather the sorted samples. In particular, we are going to assume the algorithm sees the samples in terms

of their counts A_i.

We are going to apply an algorithm to these counts: $f(A_i)$. It should be the case that $f(A_i) = X$ with probability of at least $\frac{2}{3}$.

2.1 Information processing inequality

$$\begin{aligned} I(X : A_i) &\geq I(X : f(A_i)) \\ &= H(x) - H(x|f(A_i)) \geq \Omega(1) \end{aligned}$$

Explanation: $I(X : A_i)$ and $I(X : f(A_i))$ are equal most of the time, since $H(x) = 1$ bit, and $H(x|f(A_i))$ is smaller than 1 bit (since X and f agree $\frac{2}{3}$ of the time). Thus, $H(X) - H(X|f(A_i))$ is substantial in size (at least $\Omega(1)$)

2.2 Method

2.2.1 Case Where p is Known

When $X = 1$, we can't take p to be deterministic. If we are given two specific distributions that have separation ϵ, it will take you $\frac{1}{\epsilon^2}$ samples to distinguish which distribution is being sampled from. We are trying to prove $\frac{\sqrt{n}}{\epsilon^2}$. If p is known by the $f(X)$, individual samples would be giving information. We are instead interested in examining collisions of samples.

2.2.2 Testing

Pick p from D. On average, the p is uniform, and w.h.p. has variational distance ϵ from the uniform distance.

Thus, $p_i = \frac{1 \pm \epsilon}{m}$ randomly and independently for each i.

However, the problem is that this p is not normalized, which means that the p_i's can't be chosen independently from each other. This will cause an issue if we want to bound $I(X : A_i)$, where we need all A_i's to be conditionally independent on X.

Fix: Allow p to be a non-normalized distribution. This distribution can then be sampled taking Poisson(mp_i) samples.

$A_i \sim poi(mp_i)$

Taking $poi(mp_i)$ samples from p is the same as $poi(m||p||_1)$ samples from $\frac{p}{||p||_1}$.

2.2.3 Necessities for Using the Non-Normalized Distributions

1. $||p||_1 = \theta(1)$ with 99% probability
   If the $||p||_1$ is too small, taking $\theta(1)$ samples from p is like taking a very small number of samples from the corresponding honest probability distribution.

2. $||p/||p||_1 - u_n||_1 \gg \epsilon$ with probability $\geq 99\%$

2.3 Normalization

Let $s = \sum A_i$ = number of samples we actually took.

Note: if we condition on s, the normalization no longer matters.

$$\begin{aligned} I(X : A_i) &= I(X : s) + I(X : A_i|s) \\ &= \mathrm{E}[H(X) - H(X|s = t) + I(X : A_i|s = t)] \\ &\gg \mathrm{E}[I(X : t)] \end{aligned}$$

where t = samples from normalized distribution.

But, if $s > m$ with $\frac{1}{2}$ probability, it is still $\gg I(X : m$ samples from the normalized distributions)

2.4 Results of Normalization

$p_i = \frac{1 \pm \epsilon}{m}$, where ϵ is at most $\frac{1}{2}$

1. $||p||_1 = \theta(1)$. With high probability, the number of actual samples is some constant multiple of m

2. With high probability,

$$\begin{aligned} &||p||_1 \in (\frac{1 - \epsilon/2}{m}, \frac{1 + \epsilon/2}{m}) \\ &= |p_i/||p||_1 - \frac{1}{n}| \\ &\gg \frac{\epsilon}{n} \end{aligned}$$

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

2.5 Bound

$I(X : A_1, ..., A_n)$

- A_i are conditionally independent on X

- p_i are conditionally independent on X
- $A_i \sim poi(mp_i)$

$I(X : A_1, ..., A_n) \leq \sum I(X : A_i) = nI(X : A_1)$

2.6 Further Abstraction

2.6.1 Setup

X is a uniform random bit

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

2.6.2 Shared Information

$$\begin{aligned} &I(X : A) \\ &= H(X) - H(X|A) \\ &= 1 - \sum_i (\frac{p_i + q_i}{2})(1 - (\frac{p_i}{p_i + q_i})(lg(\frac{p_i + q_i}{p_i}) + (..))) \\ &= \theta(\sum_i (\frac{(p_i - q_i)^2}{p_i + q_i})) \end{aligned}$$

$$A \sim \begin{cases} poi(\frac{m}{n}) & \text{if } X = 0 \\ \frac{1}{2}poi((1+\epsilon)(\frac{m}{n})) + \frac{1}{2}poi((1-\epsilon)(\frac{m}{n})) & \text{if } X = 1 \end{cases}$$

Let $\frac{m}{n} = \lambda$. For each k:

$$\begin{aligned} p_k &= \frac{e^{-\lambda}\lambda^k}{k!} \\ q_k &= \frac{e^{-\lambda}\lambda^k}{k!}[\frac{e^{-\lambda\epsilon}(1+\epsilon)^k + e^{\lambda\epsilon}(1-\epsilon)^k)}{2}] \end{aligned}$$

Substitution:
$I(X : A) \leq \Theta(\sum_k \frac{e^{-k}\lambda^k}{k!}[\frac{e^{-\lambda\epsilon}(1+\epsilon)^2 + e^{\lambda\epsilon}(1+\epsilon)^k}{2} - 1]^2)$

The linear terms cancel each other out, and we can thus bound this by a quadratic term, using Taylor's theorem and the second derivative. Assume $\lambda\epsilon \ll 1$:

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

If $\lambda < 1$, we use $e^{\lambda\epsilon}(1-\epsilon)^k = exp(\epsilon(\lambda - k) + O(k\epsilon^2))$. The $O(k\epsilon^2)$ becomes $\sim \frac{m\epsilon^2}{n} \ll \frac{1}{\sqrt{n}}$. The term in the right hand side can also be approximated by $cosh(\epsilon(\lambda - k)) - 1$. By substituting this into the right hand side of our previous equation:

$\mathbb{E}_{K\sim Poisson(\lambda)}[\epsilon^4(\lambda - K)^4] = O(\epsilon^4\lambda^2)$