6.889 Sublinear Time Algorithms	April 8, 2019
Lecture 16	
Lecturer: Ronitt Rubinfeld	*Scribe: Thuy-Duong Vuong*

1 Outline

The following topics were covered in class:

- Hypothesis testing
- Cover Method

2 Hypothesis testing

We say a distribution p is **known** to algorithm $\mathcal{A}$ if $\mathcal{A}$ has access to p's probability density function (pdf). We say p is **unknown** to algorithm $\mathcal{A}$ if $\mathcal{A}$ doesn't have access to p's pdf. Unless otherwise stated, $\mathcal{A}$ can take samples from distribution p (in order to "learn" p).

For distributions p, q and parameter ϵ, we say p is ϵ-close to q iff $|p - q|_1 \leq \epsilon$ Input:

- Unknown distribution p.
- Collection $\mathcal{H}$ of known distributions. $\mathcal{H}$ is guaranteed to contain a distribution that is ϵ-close to p.

Output: a distribution $q \in \mathcal{H}$ that is ϵ-close to p

Example 1. *$\mathcal{H}$ is the set of biased coins i.e. $\mathcal{H} = \{Ber(q) | q \in [0,1]\}$[1] and $p = Ber(x)$.*

3 Subtool: Comparing two hypothesis

We break the problem down into smaller pieces. First, let us consider an "easy" case of the problem: when $|\mathcal{H}| = 2$. We can build a solution for the general case from there.

Theorem 1. *There exists algorithm $\mathcal{A}$ that when given input:*

- *Known distributions h_1, h_2.*
- *Unknown distribution p.*
- *parameter $\epsilon' > 0$, confidence level $\delta' \in (0,1)$.*

takes $O(\log(1/\epsilon')/\epsilon'^2)$ samples from p, and output $h \in \{h_1, h_2\}$ such that: if one of h_1, h_2 are ϵ'-close to p, then with probability $\geq 1 - \delta'$ output h is $11\epsilon'$-close to p. Note that, we do not hold any assumption on the output when neither h_1, h_2 are ϵ-close to p.

Actually, we will prove something stronger:

Theorem 2. *There exists algorithm "Choose" when given input:*

- *Unknown distribution p.*
- *Known distributions h_1, h_2, assuming that at least one of them are ϵ-close to p*
- *parameter $\epsilon' > 0$, confidence level $\delta' \in (0,1)$.*

[1] see `https://en.wikipedia.org/wiki/Bernoulli_distribution`

takes $O(\log(1/\delta')/\epsilon'^2)$ *samples from* p,

and outputs tuple out=(outcome, h) where outcome $\in \{win, tie\}$ *and* $h \in \{h_1, h_2\}$ *that with probability* $\geq 1 - \delta'$ *satisfies:*

(1) If h_i *is more than* $12\epsilon'$*-far from* p*, then out* $\neq (outcome, h_i)$

(2) If h_i *is more than* $10\epsilon'$*-far from* p*, then out* $\neq (win, h_i)$ *(but it is probable that out* $= (tie, h_i)$*).*

Proof. Let $A = \{x | h_1(x) > h_2(x)\}$. Let $a_i = h_i(A) = \sum_{x \in A} h_i(a)$ for $i \in \{1, 2\}$.

Claim (1): $|h_1 - h_2|_1 = 2(a_1 - a_2)$.

For a proof by picture, see `https://people.csail.mit.edu/ronitt/COURSE/S19/Handouts/lec16b.pdf`. Here, we formalize the proof in words. for $x \in A$, $|h_1(x) - h_2(x)| = h_1(x) - h_2(x)$ so

$$\sum_{x \in A} |h_1(x) - h_2(x)| = \sum_{x \in A} (h_1(x) - h_2(x)) = h_1(A) - h_2(A) = a_1 - a_2.$$

Similarly, $\sum_{x \notin A} |h_1(x) - h_2(x)| = \sum_{x \in A} (h_2(x) - h_1(x)) = h_2(A^c) - h_1(A^c) = (1 - h_2(A)) - (1 - h_1(A)) = h_1(A) - h_2(A)$, where A^c is the complement of A in the union of the domains of h_1 and h_2.

Thus

$$|h_1 - h_2|_1 = \sum_{x \in A} |h_1(x) - h_2(x)| + \sum_{x \notin A} |h_1(x) - h_2(x)| = 2(a_1 - a_2)$$

Algorithm "Choose":

1. If $a_1 - a_2 \leq 5\epsilon'$, return (tie, h)
2. Draw $m = \log(1/\delta')/\epsilon'^2$ samples $s_1, \cdots, s_m$ from p
3. Let $\alpha \leftarrow \frac{1}{m} |\{i | s_i \in A\}|$.
4. If $\alpha > a_1 - \frac{3}{2}\epsilon'$ returns (win, h_1)
   else if $\alpha < a_2 + \frac{3}{2}\epsilon'$ returns (win, h_2)
   else return (tie, h_1)

There exists $h^* \in \{h_1, h_2\}$ that is ϵ'-far from p. If algorithm ends at Step 1, then h_2, h_1 are $10\epsilon'$-close to one another thus also $10\epsilon'$-close to h^*; hence, they are $11\epsilon'$-close to p. So algorithm can output "tie" along with either h_1 or h_2. On the other hand, if either h_1 or h_2 is $> 12\epsilon'$-far from p. WLOG, may assume $h^* = h_1$ and h_2 is $12\epsilon'$-far from p, then by triangle inequality, h_2 is $11\epsilon'$-far from h_1, so $a_1 - a_2 = |h_1 - h_2|_1 / 2 > 5\epsilon'$, and algorithm will reach Step 2.

Assume algorithm reaches Step 2. Note that $E(\alpha) = p(A)$, and by Chernoff's bound, with probability $\geq 1 - \delta$, $|\alpha - E(\alpha)| < \epsilon'/2$. Assume this inequality holds. If h_1 is ϵ'-close to p then by triangle inequality, $|p(A) - h_1(A)| \leq \sum_{x \in A} |p(x) - h_1(x)| \leq |p - h_1|_1 \leq \epsilon'$; thus,

$$\alpha > E(\alpha) - \epsilon'/2 = p(A) - \epsilon'/2 \geq (a_1 - \epsilon') - \epsilon'/2 = a_1 - \frac{3}{2}\epsilon'.$$

Similarly, if h_2 is ϵ'-close to p then $|p(A) - a_2| \leq \epsilon'$ so $\alpha < a_2 + \frac{3}{2}\epsilon'$. Note that since we reach step 2, $a_1 - a_2 > 5\epsilon'$ so $a_1 - \frac{3}{2}\epsilon' > a_2 + \frac{3}{2}\epsilon'$, thus the algorithm wouldn't output "tie" (assuming the inequality $|\alpha - E(\alpha)| < \epsilon'/2$ holds!)

□

4 Cover method

Using the subtool in Section 3, we can get an algorithm for the case when $\mathcal{H}$ is finite. But as we see in Example 1, $\mathcal{H}$ might be infinite. How do we deal with that? We revisit the idea of ϵ-net discussed in previous lectures. More concretely, given a set of distributions $\mathcal{D}$, we want to take a smaller set of distributions $\mathcal{C}$ that approximate $\mathcal{D}$ within some ϵ distance. Formally,

Definition 3. *Let $\mathcal{D}$ be a set of distributions. Set of distributions $\mathcal{C}$ is a ϵ-cover of $\mathcal{D}$ if $\forall q \in \mathcal{D}$, there exists $p \in \mathcal{C}$ such that $|p - q|_1 \leq \epsilon$.*

This way, we can save time by running algorithms on $\mathcal{C}$ instead of $\mathcal{D}$.

Theorem 4. *There exists an algorithm, that given $p \in \mathcal{D}$, takes $O(\frac{1}{\epsilon^2} \log |\mathcal{C}|)$ samples of p and output $h \in \mathcal{C}^{\mathcal{D}}$ such that $|h - p| \leq 11\epsilon$.*

Proof. Since $p \in \mathcal{D}$, there exists $q \in \mathcal{C}$ such that $|p - q|_1 \leq \epsilon$. We run Choose on every pair q_1, q_2 in $\mathcal{C}$ with parameter $\epsilon' = \epsilon$ and $\delta' = \frac{1}{10\binom{|\mathcal{C}|}{2}}$. Then by union bound, with probability $\geq 1 - \binom{|\mathcal{C}|}{2}\delta' = 9/10$, all output of calls to Choose satisfy their guarantee. Assuming this happens. We can show that there is a q' that wins or ties all matches (q', q_2) where $q_2 \in \mathcal{C} \setminus \{q'\}$. For example, let $q' = q$ then by Definition of Algorithm Choose, any match (q, q_2) either ends in a "tie" at Step 1, or reaches Step 2 and ends in a "win" for q.

But what if there is multiple q' that wins or ties all matches? We can just pick an arbitrary such q' and output it, since any such q' satisfies $|q' - p|_1 \leq 11\epsilon$. Indeed, if $q' = q$ then we are done, as q is ϵ-close to p. Assume $q' \neq q$, and consider the match between q' and q: if q' wins, then q' is $\leq 10\epsilon'$-close to p, else if q' tie, then q' is 10ϵ-close to q, thus 11ϵ-close to p. □

Example 1 revisited. We abuse notation and write q in place of $Ber(q)$ for brevity's sake. We write $\mathcal{H} = \{q | q \in [0,1]\}$. Then $\mathcal{C} = \{0, \frac{1}{k}, \frac{2}{k}, \cdots, \frac{k-1}{k}, 1\}$ where $k = 2/\epsilon$ is a ϵ-cover of $\mathcal{H}$. Indeed, let $r \in \{0, \cdots, k\}$ be such that $\frac{r}{k} \leq x < \frac{r+1}{k}$ then $\left|Ber(x) - Ber(\frac{r}{k})\right|_1 = 2\left|\frac{r}{k} - x\right| \leq 2/k = \epsilon$. Note that $|\mathcal{C}| = \theta(1/\epsilon)$. So by Theorem 3, setting $\epsilon' = \epsilon/11$, can learn $Ber(q)$ ϵ-close to $p = Ber(x)$ by taking $O(\frac{1}{\epsilon^2} \log(\frac{1}{\epsilon}))$ samples from p.

Example 2 (2-Bucket distributions). *A 2-bucket distribution $B_{\alpha,\beta}$ is defined by*

$$Pr_{X \sim B_{\alpha,\beta}}[X = i] = \begin{cases} \frac{\alpha}{n/2} & \text{if } i \in [n/2] \\ \frac{\beta}{n/2} & \text{if } i \in [n] \setminus [n/2] \\ 0 & \text{else} \end{cases}$$

Let $\mathcal{D}$ be the set of all 2-bucket distributions $B_{\alpha,\beta}$ where $\alpha, \beta \in [0,1]$. Similar to in Example 1, we can create an ϵ-cover using ϵ-net for each of α, β i.e. $\mathcal{C} = \{B_{i/k, j/k} | i, j \in \{0, \cdots, k\}\}$ where $k = 1/\epsilon$. The size of this cover is $\theta(\frac{1}{\epsilon^2})$ thus can learn unknown $B_{\alpha,\beta}$ in $O(\frac{1}{\epsilon^2} \log(\frac{1}{\epsilon}))$.

Example 3 (Monotone distributions). *Let $\mathcal{D}$ be the set of monotone (decreasing) distributions over $[n] = \{1, 2, \cdots, n\}$. By lecture 15, the set of Birge distributions*[2] *$\mathcal{C} = \{(w_1, \cdots, w_{\theta(\frac{\log n}{\epsilon})}) | w_i = \frac{j_i}{k}, j_i \in \{0, \cdots, k\}\}$ where $k = 1/\epsilon$ forms an ϵ-cover. The size of this cover is $|\mathcal{C}| = \theta(\frac{1}{\epsilon^{\log n/\epsilon}})$, so we can learn $p \in \mathcal{D}$ in $O(\frac{\log n}{\epsilon^3} \log(\frac{1}{\epsilon}))$.*

[2] see Lecture 15