6.5240 Sublinear Time Algorithms	October 31, 2022
Lecture 14	
Lecturer: Ronitt Rubinfeld	*Scribe: Anton Trygub*

We denote the set $\{1, 2, \ldots, n\}$ by $[n]$.
Our goal for today is testing monotonicity. We consider distributions over domain $[n]$.

1 Introduction

Definition 1 (Monotone decreasing) *A distribution p over set $\{1, 2, \ldots, n\}$ is **monotone decreasing**, if $\forall i \in \{1, 2, \ldots, n-1\}$ holds $p(i) \geq p(i+1)$.*

Definition 2 (ϵ-far from monotone decreasing) *A distribution p over $[n]$ is **ϵ-far from being monotone decreasing**, if for every monotone decreasing distribution q over $[n]$, $||p-q||_1 \geq \epsilon$.*
As a reminder, $||p-q||_1 = \sum_{i=1}^{n} |p(i) - q(i)|$.

We are looking for a monotonicity tester with the following properties:

- If p is monotone decreasing, **pass** with probability $\geq \frac{3}{4}$
- If p is ϵ-far from monotone decreasing, **reject** with probability $\geq \frac{3}{4}$.

Morale time: *If you don't have the strength to stay up at night to work on the problems, you can still be a researcher!*

For testing monotonicity, the following tool will be very useful:

Definition 3 (Birge decomposition) *Decompose the domain $[n]$ into $l = \Theta(\frac{\log \epsilon n}{\epsilon}) \sim \Theta(\frac{\log n}{\epsilon})$ intervals $I_1^\epsilon, I_2^\epsilon, \ldots, I_l^\epsilon$, such that $|I_{k+1}^\epsilon| = \lfloor (1+\epsilon)^k \rfloor$*
*This decomposition is called **Birge decomposition**.*

Notes:

- The last segment might have a smaller length.
- We will drop $^\epsilon$-superscription in the future.
- We will use the terms "intervals," "partitions," "buckets" interchangeably for $I_1, I_2, \ldots, I_l$.
- $\Theta(\frac{1}{\epsilon})$ of these intervals have length 1.

Definition 4 (Flattened distribution) *For any distribution q on $[n]$, and ϵ, define the flattened distribution $\tilde{q}_\epsilon$ as follows:*

$$\forall j \in [l], \forall i \in I_j, \text{ define } \tilde{q}_\epsilon(i) = \frac{q(I_j)}{|I_j|}, \text{ where } q(I_j) = \sum_{i \in I_j} q(i)$$

In other words, we "flatten" the distribution in each of the intervals of Birge decomposition. Note that it immediately follows that $q(I_j) = \tilde{q}_\epsilon(I_j)$.
Let's also denote the maximum and the minimum probabilities of elements from I_j by max_j, min_j correspondingly. Note that $max_j \leq min_{j-1}$.

2 Proof of Birge's theorem

Theorem 5 (Birge's) *If q is a monotone decreasing distribution, then $||\tilde{q}_\epsilon - q||_1 \leq \epsilon$.*

Proof

Consider the error in a single bucket I_j. Clearly, it doesn't exceed $(max_j - min_j) \cdot |I_j|$.
Let's divide buckets into three groups:

- Size 1 intervals: I_j with $|I_j| = 1$
- Short intervals: I_j with $1 < |I_j| < \frac{1}{\epsilon}$
- Long intervals: I_j with $\frac{1}{\epsilon} \leq |I_j|$

Then our error doesn't exceed:

$$\sum_{j=1}^{l}(max_j - min_j) \cdot |I_j| = \sum_{|I_j|=1}(max_j - min_j) \cdot 0 + \sum_{I_j \text{ short}}(max_j - min_j) \cdot |I_j| + \sum_{I_j \text{ long}}(max_j - min_j) \cdot |I_j|$$

Disclaimer: The actual proof is very technical and contains many details. We will only give proof that the error doesn't exceed $O(\epsilon)$, but it should be enough for the intuition and all practical needs.

2.1 Large intervals

So, let's look at the bound for large intervals. Let's suppose that I_{k+1} is the first long interval. Then:

$$\sum_{j=k+1}^{l}(max_j - min_j) \cdot |I_j| \leq \sum_{j=k+1}^{l}(min_{j-1} - min_j) \cdot |I_j| \leq min_k \cdot |I_{k+1}| + \sum_{j=k+1}^{l-1} min_j \cdot (|I_{j+1}| - |I_j|)$$

Now, let's note that if $|I_j| \geq \frac{1}{\epsilon}$, then

$$|I_{j+1}| = \lfloor(1+\epsilon)^j\rfloor \leq (1+\epsilon)^j \leq (|I_j|+1)(1+\epsilon) = |I_j| + |I_j|\epsilon + (1+\epsilon) < |I_j| + 3|I_j|\epsilon$$

So

$$\sum_{j=k+1}^{l-1} min_j \cdot (|I_{j+1}| - |I_j|) \leq \sum_{j=k+1}^{l-1} min_j \cdot 3\epsilon|I_j|$$

Note that this sum doesn't exceed 3ϵ times the area under all long segments, so this doesn't exceed 3ϵ.

Now, let's bound $min_k \cdot |I_{k+1}|$. Note that $min_k(|I_1| + |I_2| + \ldots + |I_k|) \leq 1$. We will show that $|I_{k+1}| \leq 4\epsilon(|I_1| + |I_2| + \ldots + |I_k|)$. Indeed:

$$|I_1| + |I_2| + \ldots + |I_k| \geq \frac{1}{2}((1+\epsilon)^0 + (1+\epsilon)^1 + \ldots + (1+\epsilon)^{k-1}) = \frac{1}{2}\frac{(1+\epsilon)^k - 1}{\epsilon} \geq \frac{1}{4}\frac{(1+\epsilon)^k}{\epsilon} \geq \frac{1}{4\epsilon}|I_{k+1}|$$

Then $min_k \cdot |I_{k+1}| \leq 4\epsilon \cdot min_k(|I_1| + |I_2| + \ldots + |I_k|) \leq 4\epsilon$.
So, the error in the long intervals doesn't exceed 7ϵ.

2.2 Short intervals

Now, let's deal with short intervals − those, for which $1 < |I_j| < \frac{1}{\epsilon}$

Again, for segment I_j, we will bound error on it by $(max_j - min_j) \cdot |I_j| \leq (min_{j-1} - min_j) \cdot |I_j|$.

First, note that all numbers from 2 to $\lfloor \frac{1}{\epsilon} \rfloor$ appear among the lengths of the intervals. By contradiction, suppose that some integer $k \leq \frac{1}{\epsilon}$ doesn't. Then there is some j such that $|I_j| \leq k-1, |I_{j+1}| \geq k+1$. But then

$$1 + \epsilon = \frac{(1+\epsilon)^j}{(1+\epsilon)^{j-1}} > \frac{k+1}{k} = 1 + \frac{1}{k}$$

So $k > \frac{1}{\epsilon}$, contradiction.

Now, let a_k be the smallest number with $|I_{a_k}| = k$, and c be largest integer smaller than $\frac{1}{\epsilon}$. Then, rewrite:

$$\sum_{I_j short} (min_{j-1} - min_j) \cdot |I_j| \sum_{k=2}^{c} \sum_{j=a_k}^{j=a_{k+1}-1} k(min_{j-1} - min_j) \leq 2min_{a_2-1} + min_{a_3-1} + \ldots + min_{a_c-1}$$

Let k_i be the number of partitions with length $2 \leq i \leq c$. It means that there is some j with $|I_j| = i-1$ and $|I_{j+k_i+1}| = i+1$. Then

$$(1+\epsilon)^{k_i+1} = \frac{(1+\epsilon)^{j+k_i}}{(1+\epsilon)^{j-1}} \geq \frac{i+1}{i} = 1 + \frac{1}{i}$$

Now, write

$$\frac{i+1}{i} \leq (1+\epsilon)^{k_i+1} \leq \frac{1}{(1-\epsilon)^{k_i+1}} \leq \frac{1}{1-\epsilon(k_i+1)}$$

After taking inverse, this is equivalent to:

$$1 - \frac{1}{i+1} \geq 1 - \epsilon(k_i+1) \iff (k_i+1)(i+1) \geq \frac{1}{\epsilon}$$

From the last inequality it clearly follows that $k_i i \geq \frac{1}{4\epsilon}$ (for $i \geq 2$). The same thing can be said about k_1 (we will omit this bound).

Note that $k_i i$ is the total length of intervals of length i. So,

$$1 \geq \sum_{l=1}^{c} k_i i \cdot min_{a_{l+1}-1} \geq \frac{1}{4\epsilon} \sum_{l=1}^{c} min_{a_{l+1}-1} \Rightarrow \sum_{l=1}^{c} min_{a_{l+1}-1} \leq 4\epsilon$$

It immediately follows that

$$\sum_{I_j short} (min_{j-1} - min_j) \cdot |I_j| \leq 2min_{a_2-1} + min_{a_3-1} + \ldots + min_{a_c-1} \leq 8\epsilon$$

■

Corollary 6 *If q is ϵ-close to monotone decreasing, then $||\tilde{q}_\epsilon - q||_1 < O(\epsilon)$*

Proof Consider some monotone decreasing p with $||q - p||_1 \leq \epsilon$. Then it's not hard to see that $||\tilde{q}_\epsilon - \tilde{p}_\epsilon||_1 \leq \epsilon$ too. This follows from the following fact: for any array a, b of length n,

$$\sum_{i=1}^{n} |a_i - b_i| \geq n \left| \frac{a_1 + a_2 + \ldots + a_n}{n} - \frac{b_1 + b_2 + \ldots + b_n}{n} \right|$$

This is just $|c_1| + |c_2| + \ldots + |c_n| \geq ||c_1 + c_2 + \ldots + c_n|$ for $c_i = a_i - b_i$. Then, apply this inequality to every bucket.

Now, we have the following inequalities: $||q - p||_1 \leq \epsilon$, $||p - \tilde{p}_\epsilon||_1 \leq \epsilon$, $||\tilde{p}_\epsilon - \tilde{q}_\epsilon||_1 \leq \epsilon$. By the triangle inequality, we get $||\tilde{q}_\epsilon - q||_1 \leq 3\epsilon$. ■

3 Monotonicity Tester

3.1 Algorithm

Let's devise the following testing algorithm. For some ϵ_1 that we will choose later, do:

1. Take a set S of $m = \tilde{O}(\sqrt{n} \cdot \text{poly}(\log n, \frac{1}{\epsilon}))$ samples of q.
2. For each Birge partition I_j, let S_j be the set of samples that fall in I_j ($S_j = S \cap I_j$). Do a uniformity test on each such interval. If greater than ϵ_1-fraction of samples are in a failing interval, output `Reject`.
3. Define $\hat{w}_j = \frac{|S_j|}{m}$ as the estimate of $q(I_j)$.
4. Define q^* as follows: for all $i \in I_j$, $q^*(i) = \frac{\hat{w}_j}{|I_j|}$ as the estimate of q^*.
5. Use linear programming to verify that w is ϵ_1-close to monotone (note that this is an LP on $O(\frac{\log n}{\epsilon})$ variables, so it's feasible). If it is, output Pass, otherwise `Reject`.

3.2 Analysis

3.2.1 Monotone decreasing distributions

Consider any monotone decreasing distribution q. We know that $||q - \tilde{q}_\epsilon|| \leq \epsilon$, and (with Chernoff bounds) that it will pass the uniformity test, and that $||\tilde{q}_\epsilon - q^*||$ won't exceed ϵ. Then, q^* will be at most 2ϵ-far from monotone decreasing.

3.3 ϵ-far from monotone decreasing distributions

If a distribution is likely to pass, then it's almost uniform on all its Birge's partitions, and its q^* is close to monotone decreasing, so we would get that the distribution itself is also close to monotone decreasing.

4 Learning monotone decreasing distributions

It turns out that we can learn such distributions up to an ϵ error in the L_1 error in $O(\frac{1}{\epsilon^3} \log n)$ samples. The intuition is that, by Birge's theorem, it's enough to estimate $\tilde{q}_\epsilon$, which is constant on $O(\frac{\log n}{\epsilon})$ segments.

Happy Halloween!