6.5240 Sublinear Time Algorithms October 30, 2024

Lecture 16

Lecturer: Ronitt A. Rubinfeld *Scribe: Ying Feng*

1 Preliminaries for Testing Monotonicity of Distributions

In this lecture, we will discuss testing for monotone (decreasing) distributions over a totally ordered domain.

Definition 1.1 (Monotone Decreasing)**.** *A distribution $\mathcal{Q}$ over a totally ordered domain $[n]$ is "monotone decreasing" if for all $i \in [n]$, $\mathcal{Q}(i) \geq \mathcal{Q}(i+1)$.*

We aim to find a monotonicity tester, such that

- if $\mathcal{Q}$ is monotone decreasing, then the tester outputs Pass with probability at least $1-\delta$.
- if $\mathcal{Q}$ is ε-far from **any** monotone decreasing distribution $\mathcal{P}$, then the tester outputs Fail with probability at least $1-\delta$.

where ε-far is measured in the L_1 distance, i.e. two distributions $\mathcal{P}$ and $\mathcal{Q}$ are ε-far if and only if

$$\|\mathcal{P}-\mathcal{Q}\|_1 := \sum_{i\in[n]} |\mathcal{P}(i)-\mathcal{Q}(i)| > \varepsilon.$$

1.1 Birge's Decomposition

We will use the following decomposition procedure to construct our tester:

Given a parameter ε, partition the domain $[n]$ into $l = \theta(\frac{\log n}{\varepsilon})$ consecutive intervals:

$$I_1^\varepsilon, I_2^\varepsilon, \cdots, I_l^\varepsilon$$

such that $|I_{j+1}^\varepsilon| = \lfloor (1+\varepsilon)^j \rfloor$ for each $j < l$.

In the following, we will drop the superscript ε since it is fixed in the algorithm.

The unrounded size of the intervals increase by a factor of $(1+\varepsilon)$. Therefore, $\begin{cases} |I_1| = \cdots |I_{\theta(1/\varepsilon)}| = 1 \\ |I_{\theta(1/\varepsilon)+1}| = \cdots = 2 \\ \vdots \end{cases}$.

Given the decomposition, we define "flattening" a distribution as follows:

Definition 1.2. *Given a distribution $\mathcal{Q}$, the "flattened distribution" $\tilde{\mathcal{Q}}$ is defined as follows: for all intervals $j \in [l]$ and for all $i \in I_j$, let*

$$\tilde{\mathcal{Q}}(i) = \frac{\sum_{i\in I_j} \mathcal{Q}(i)}{|I_j|},$$

i.e., the total weight of the interval containing i divided by the number of domain elements in this interval.

In the following, we will use $\mathcal{Q}(I_j)$ to denote $\sum_{i \in I_j} \mathcal{Q}(i)$. We note that $\mathcal{Q}(I_j) = \tilde{\mathcal{Q}}(I_j)$ for all intervals I_j.

The following theorem states that $\tilde{\mathcal{Q}}$ is a good approximation of $\mathcal{Q}$ if $\mathcal{Q}$ is (close to) monotone decreasing.

Theorem 1.3 (Birge's Theorem)**.** *If $\mathcal{Q}$ is monotone decreasing, then $\|\mathcal{Q} - \tilde{\mathcal{Q}}\|_1 \leq O(\varepsilon)$.*

This gives the following immediate corollary:

Corollary 1.4. *If $\mathcal{Q}$ is ε-close to monotone decreasing, i.e., there exists a monotone decreasing distribution $\mathcal{Q}'$ such that $\|\mathcal{Q} - \mathcal{Q}'\|_1 \leq \varepsilon$, then $\|\mathcal{Q} - \tilde{\mathcal{Q}}\|_1 \leq O(\varepsilon)$.*

1.2 Proof of Birge's Theorem

Intuitively, $\tilde{\mathcal{Q}}$ is a bad approximation of $\mathcal{Q}$ on intervals that contains a large drop-off. The proof of Birge's Theorem shows that there cannot be too many such intervals.

Proof (of Birge's Theorem). For an arbitrary interval I_j, define $x_j := \arg\max_{i \in I_j} \mathcal{Q}(i)$ and $y_j := \arg\min_{i \in I_j} \mathcal{Q}(i)$. Observe that $\|\mathcal{Q}_{I_j} - \tilde{\mathcal{Q}}_{I_j}\|_1 \leq (\mathcal{Q}(x_j) - \mathcal{Q}(y_j)) \cdot |I_j|$, where $\mathcal{Q}_{I_j}$ denote the distribution $\mathcal{Q}$ restricted to the interval I_j and $\tilde{\mathcal{Q}}_{I_j}$ denote $\tilde{\mathcal{Q}}$ restricted to I_j.

In the following, we consider intervals as three different types based on their sizes. Given an interval I_j, it is a

$$\begin{cases} \text{size-1 interval if } |I_j| = 1 \\ \text{short interval if } 1 < |I_j| < 1/\varepsilon \\ \text{long interval if } |I_j| \geq 1/\varepsilon \end{cases}$$

We observe that $\max_{i \in I_j : |I_j| > 1} \mathcal{Q}(i) = O(\varepsilon)$ because by definition there are $\theta(1/\varepsilon)$ many size-1 intervals. For monotone decreasing $\mathcal{Q}$, if $\max_{i \in I_j : |I_j| > 1} \mathcal{Q}(i) = \omega(\varepsilon)$ then the weight of each size-1 interval is $\omega(1)$, thus the total weight of size-1 intervals becomes $\theta(1/\varepsilon) \cdot \omega(\varepsilon) \gg 1$, a contradiction.

We now bound the error incurred by each type of intervals separately.

- (Size-1 Intervals.) $\sum_{|I_j|=1} \|\mathcal{Q}_{I_j} - \tilde{\mathcal{Q}}_{I_j}\|_1 = 0$ because for size-1 intervals, the weight of each element is the same as the sum of weights divided by one.
- (Short Intervals.) $\sum_{1 < |I_j| < 1/\varepsilon} \|\mathcal{Q}_{I_j} - \tilde{\mathcal{Q}}_{I_j}\|_1 \leq \sum_{1 < |I_j| < 1/\varepsilon} |I_j| \cdot (\mathcal{Q}(x_j) - \mathcal{Q}(y_j))$ as we argued before. We can split the summation according to the sizes $|I_j| = 2, 3, \cdots, 1/\varepsilon - 1$. Define $j_t := \min\{j : |I_j| = t\}$ for $t = 2, 3, \cdots, 1/\varepsilon - 1$, i.e., j_t is the index of the first interval of size t, we obtain

$$\begin{aligned} &\sum_{1 < |I_j| < 1/\varepsilon} \|\mathcal{Q}_{I_j} - \tilde{\mathcal{Q}}_{I_j}\|_1 \\ &\qquad \leq 2 \cdot (\mathcal{Q}(x_{j_2}) - \mathcal{Q}(x_{j_3})) + 3 \cdot (\mathcal{Q}(x_{j_3}) - \mathcal{Q}(x_{j_4})) + \cdots + (1/\varepsilon - 1) \cdot (\mathcal{Q}(x_{j_{1/\varepsilon - 1}}) - \mathcal{Q}(x_{j_{1/\varepsilon}})) \\ &\qquad = \mathcal{Q}(x_{j_2}) + \sum_{1 < |t| < 1/\varepsilon} \mathcal{Q}(x_{j_t}) - (1/\varepsilon - 1) \cdot \mathcal{Q}(x_{j_{1/\varepsilon}}) \end{aligned}$$

- (Long Intervals.) We still use the notation that $j_t := \min\{j : |I_j| = t\}$, i.e., j_t is the index of the first interval of size t. We get

$$\begin{aligned}\sum_{|I_j|\geq 1/\varepsilon} \|\mathcal{Q}_{I_j} - \tilde{\mathcal{Q}}_{I_j}\|_1 &\leq \sum_{|I_j|\geq 1/\varepsilon} |I_j| \cdot (\mathcal{Q}(x_j) - \mathcal{Q}(y_j)) \\ &\leq |I_{j_{1/\varepsilon}}| \cdot (\mathcal{Q}(x_{j_{1/\varepsilon}}) - \mathcal{Q}(x_{j_{1/\varepsilon}+1})) + |I_{j_{1/\varepsilon}+1}| \cdot (\mathcal{Q}(x_{j_*+1}) - \mathcal{Q}(x_{j_{1/\varepsilon}+2})) + \cdots \\ &= |I_{j_{1/\varepsilon}}| \cdot \mathcal{Q}(x_{j_{1/\varepsilon}}) + \sum_{j>j_{1/\varepsilon}} (|I_j| - |I_{j-1}|)\mathcal{Q}(x_j)\end{aligned}$$

By construction, $|I_j| - |I_{j-1}| \approx (1+\varepsilon)|I_{j-1}| - |I_{j-1}| \approx \varepsilon |I_{j-1}|$. Therefore

$$\sum_{|I_j|\geq 1/\varepsilon} \|\mathcal{Q}_{I_j} - \tilde{\mathcal{Q}}_{I_j}\|_1 \leq |I_{j_{1/\varepsilon}}| \cdot \mathcal{Q}(x_{j_{1/\varepsilon}}) + O(\varepsilon) \cdot \sum_{|I_j|\geq 1/\varepsilon} |I_j|\mathcal{Q}(x_{j+1}).$$

Summarizing the above,

$$\begin{aligned}&\|\mathcal{Q}_{I_j} - \tilde{\mathcal{Q}}_{I_j}\|_1 \\ &\quad = \sum_{|I_j|=1} \|\mathcal{Q}_{I_j} - \tilde{\mathcal{Q}}_{I_j}\|_1 + \sum_{1<|I_j|<1/\varepsilon} \|\mathcal{Q}_{I_j} - \tilde{\mathcal{Q}}_{I_j}\|_1 + \sum_{|I_j|\geq 1/\varepsilon} \|\mathcal{Q}_{I_j} - \tilde{\mathcal{Q}}_{I_j}\|_1 \\ &\quad \leq \mathcal{Q}(x_{j_2}) - (1/\varepsilon - 1) \cdot \mathcal{Q}(x_{j_{1/\varepsilon}}) + O(\varepsilon) \cdot \sum_{|I_j|\geq 1/\varepsilon} |I_j|\mathcal{Q}(x_{j+1}) + \Big(\sum_{1<t<1/\varepsilon} \mathcal{Q}(x_{j_t}) + |I_{j_{1/\varepsilon}}| \cdot \mathcal{Q}(x_{j_{1/\varepsilon}})\Big) \\ &\quad \leq O(\varepsilon) + O(\varepsilon) \cdot \sum_{|I_j|\geq 1/\varepsilon} |I_j|\mathcal{Q}(x_{j+1}) + \Big(\sum_{1<t<1/\varepsilon} \mathcal{Q}(x_{j_t}) + (|I_{j_{1/\varepsilon}}| - 1/\varepsilon + 1) \cdot \mathcal{Q}(x_{j_{1/\varepsilon}})\Big)\end{aligned}$$

It suffices to show that $\sum_{|I_j|\geq 1/\varepsilon} |I_j|\mathcal{Q}(x_{j+1}) = O(1)$ and $\sum_{1<t<1/\varepsilon} \mathcal{Q}(x_{j_t}) + (|I_{j_{1/\varepsilon}}| - 1/\varepsilon + 1) \cdot \mathcal{Q}(x_{j_{1/\varepsilon}}) = O(\varepsilon)$. The former holds because the summation is at most the cumulative mass of the probability distribution $\mathcal{Q}$ which sums to 1. For the latter, we observe that

$$\sum_{1<t\leq 1/\varepsilon} \Big(\mathcal{Q}(x_{j_t}) \cdot \sum_{|I_j|=t} |I_j|\Big) \leq 1$$

because this summation is also at most the cumulative mass of the entire distribution. Moreover, for each $1 < t \leq 1/\varepsilon$, we have $\sum_{|I_j|=t} |I_j| \geq \theta(1/\varepsilon)$ by our definition of intervals. This implies $\sum_{1<t\leq 1/\varepsilon} \mathcal{Q}(x_{j_t}) = O(\varepsilon)$. Finally, since $(|I_{j_{1/\varepsilon}}| - 1/\varepsilon + 1) = 1/\varepsilon - 1/\varepsilon + 1 = 1$, we obtain

$$\begin{aligned}\sum_{1<t<1/\varepsilon} \mathcal{Q}(x_{j_t}) + (|I_{j_{1/\varepsilon}}| - 1/\varepsilon + 1) \cdot \mathcal{Q}(x_{j_{1/\varepsilon}})) &= \sum_{1<t<1/\varepsilon} \mathcal{Q}(x_{j_t}) + \mathcal{Q}(x_{j_{1/\varepsilon}}) \\ &= \sum_{1<t\leq 1/\varepsilon} \mathcal{Q}(x_{j_t}) \\ &= O(\varepsilon)\end{aligned}$$

which concludes our proof that $\|\mathcal{Q}_{I_j} - \tilde{\mathcal{Q}}_{I_j}\|_1 = O(\varepsilon)$.

□

2 Monotonicity Tester

We start by presenting a simplified tester (which unfortunately does not work) and then discuss its issues and fixes.

2.1 Simplified Tester

Testing Algorithm for Monotonicity

Input: (Samples of) a distribution $\mathcal{Q}$ to be tested.

1. Take m samples S of $\mathcal{Q}$.
2. For each interval I_j:
   (a) Let $S_j \leftarrow S \cap I_j$.
   (b) Let $\hat{w}_j \leftarrow \frac{|S_j|}{m}$. *// Guess of the weight of interval I_j, i.e. guess of $\mathcal{Q}(I_j) = \tilde{\mathcal{Q}}(I_j)$.*
3. Define $\mathcal{Q}^*$ as follows: for all $i \in I_j$ and for all intervals I_j, $\mathcal{Q}^* := \frac{\hat{w}_j}{|I_j|}$. *// Guess of $\tilde{\mathcal{Q}}(I_j)$.*
4. If $\mathcal{Q}^*$ is monotone decreasing then output Pass; otherwise output Fail.

The intuition behind the above tester is that when $\mathcal{Q}$ is (close to) monotone decreasing, $\tilde{\mathcal{Q}}$ and $\mathcal{Q}$ are close by Birge's Theorem. Therefore, the hope is that if $\mathcal{Q}^*$ well-approximates $\tilde{\mathcal{Q}}$ then testing on $\mathcal{Q}^*$ works for the original distribution $\mathcal{Q}$. However, there are multiple problems about this idealized argument, and we are going to fix them by modifying line 4 of the tester (marked in blue).

2.2 Problems and Fixes

Problem 1. Sampling Errors. Even when both $\mathcal{Q} = \tilde{\mathcal{Q}}$ is monotone decreasing, it is possible that $\mathcal{Q}^*$ induced by our samples S is not *exactly* monotone decreasing due to sampling errors. Consider $\mathcal{Q} = \tilde{\mathcal{Q}}$ being uniform distributions as an example. Sampling errors can easily cause $\mathcal{Q}^*$ to deviate from being exactly monotone.

To fix this, we will tolerate a small error when testing $\mathcal{Q}^*$. i.e., Instead of line 4 of the simplified tester, we use the following Test A:

A. If $\mathcal{Q}^*$ is $c \cdot \varepsilon$-close to monotone decreasing then output Pass; otherwise output Fail.

for some constant $c < 1$. We note that Test A does not require any new samples from $\mathcal{Q}$, and can be implemented in $\text{poly}\log(n)$ time using linear programming with $O(\log n)$ variables.

Problem 2. Errors within Intervals. Recall that we flatten each intervals such that all weights inside a same interval is regarded the same. However, it is possible for an input distribution $\mathcal{Q}$ to be monotone decreasing across different intervals, but non-monotone inside intervals. To detect this, we need another test to verify that $\mathcal{Q}$ looks good inside each interval. This is described as the following Test B:

B. For each interval I_j, run a uniformity test using samples S_j. If more than $c' \cdot \varepsilon$-fraction of uniformity test fails, output Fail; otherwise output Pass.

for some constant $c' < 1$. The idea is that if $\mathcal{Q}$ is (close to) monotone, then by Birge's Theorem it is close to $\tilde{\mathcal{Q}}$ which is uniform within each intervals. We note that the uniformity test is guaranteed to pass only on *exactly* uniform distributions. However, one can show that for distributions where the maximum and minimum probabilities are close (which is the case within each interval of a close-to-monotone $\mathcal{Q}$), the uniformity test is likely to pass distributions that are *close* to uniform.

Analysis. Summarizing the above, we replace line 4 in the simplified tester with two tests A and B. The modified tester outputs Pass if and only if both tests A and B pass. We need roughly $\sqrt{I_j}$ samples for each interval I_j and the total number of intervals is $O(\frac{\log n}{\varepsilon})$. This results in a total number of samples $m = O(\sqrt{n} \cdot \mathrm{poly}(\log n, 1/\varepsilon))$.

For the correctness, when $\mathcal{Q}$ is monotone thus $\tilde{\mathcal{Q}}$ is monotone, Test A passes with high probability because we can show via a Chernoff bound that each $\hat{w}_j$ is close to $\tilde{\mathcal{Q}}(I_j)$, which implies $\|\mathcal{Q}^* - \tilde{\mathcal{Q}}\|_1 \le c \cdot \varepsilon$ for some properly chosen constant c. Moreover, Test B passes with high probability because of our argument on (tolerating) uniformity test above.

For the negative case, we show the contrapositive that if the tester is likely to pass, then $\mathcal{Q}$ is ε-close to monotone decreasing. Passing Test A implies that $\mathcal{Q}^*$ is close to monotone. Passing Test B means $\mathcal{Q}$ is almost uniform on all intervals thus close to $\tilde{\mathcal{Q}}$. Therefore, since $\mathcal{Q}^*$ is likely to well-approximate $\tilde{\mathcal{Q}}$ by the Chernoff bound, $\mathcal{Q}$ is close to monotone by the triangle inequality.