Introduction to Property Testing
Lecture 3

Lecture by Dr. Reut Levi
Typeset by Steven Karas

2020-05-06
Last edited 15:35:30 2020-05-13

Disclaimer These notes are based on the lectures for the course Introduction to Property Testing, taught by Dr. Reut Levi at IDC Herzliyah in the spring semester of 2019/2020. Sections may be based on the lecture slides prepared by Dr. Reut Levi.

1 Agenda

- Recap

2 Proximity oblivious testers

POTs are a common and more generic form of property testers. They are characterized as not receiving a ϵ as an input. Instead the rejection probability increases with the distance from the property.

A proximity oblivious tester (POT) with threshold probability t and detection probability $g : (0, 1] \to (0, 1]$ where g is monotonically non-decreasing for a property Π satisfies the following:

- $\forall f \in \Pi \quad \Pr[T(f) = 1] \geq t$
- $\forall f \notin \Pi \quad \Pr[T(f) = 1] \leq t - g(\delta_\Pi(f))$

2.1 Sorted POT

As an algorithm:

1. select uniformly $i, j \in [n]$ s.t. $i < j$
2. accept iff $x_i \leq x_j$

Correctness Consider $x' = 0^{n-wt(x)}1^{wt(x)}$. Note that $wt(x) = wt(x')$, however x is not necessarily sorted, whereas x' is.

$$D_0 = \{i \in [n - wt(x) + 1, n] : x_i = 0\}$$
$$D_1 = \{i \in [1, n - wt(x)] : x_i = 1\}$$

The rejection probability of this algorithm is at least $\frac{|D_0|\cdot|D_1|}{\binom{n}{2}}$.

Note that $|D_0| = |D_1|$ because any elements that are out of order in the first section must have a corresponding element out of order in the second half. Note that $|D_0| = |D_1| = \frac{1}{2}\delta(x, x') \cdot n \geq \frac{1}{2}n \cdot \delta_{\text{SORTED}}(x)$. Therefore, the rejection probability is at least:

$$\frac{\frac{1}{4}n^2\delta^2_{\text{SORTED}}(x)}{n(n-1)/2} > \frac{\delta^2_{\text{SORTED}(x)}}{2}$$

2.2 Deriving standard testers from POTs

If we have a one sided error POT of query complexity q then we have a one sided error standard property tester of query complexity $q'(\varepsilon) = O\left(\frac{q}{g(\varepsilon)}\right)$

If we have a two sided error POT of query complexity q then we can construct a two sided error tester of query complexity $q'(\varepsilon) = O\left(\frac{q}{g^2(\varepsilon)}\right)$

Proof by construction Invoke a one sided error POT $O\left(\frac{1}{g(\varepsilon)}\right)$ times, accept iff all invocations accept.

Invoke a two sided error POT $O\left(\frac{1}{g^2(\varepsilon)}\right)$ times, accept if at least $t - \frac{g(\varepsilon)}{2}$ fraction of the invocations accept.

If the input has the property, then the tester accepts with probability 1. If it's ε-far from having the property, then the probability a single invocation accepts is no greater than $1-g(\varepsilon)$. Therefore, the probability that all invocations accept is no greater than $(1-g(\varepsilon))^{O(1/g(\varepsilon))} \leq \left(\frac{1}{e}\right)^{O(1)}$. From this, we can apply Chernoff's bound.

3 Group Homomorphism

For simplicity, we will consider the group of integers with addition.

1. Closure under addition: $a+b \in \mathbb{Z}$
2. Associativity: $(a+b)+c = a+(b+c)$
3. Identity element: $0+a = a+0 = a$
4. Inverse element: $a+(-a) = (-a)+a = 0$

A group homomorphism is defined as a function $f : G \to H$ if $\forall x, y \in G$ it holds that $f(x+y) = f(x)+f(y)$.

We will show a POT for testing group homomorphism: select u.a.r. $x, y \in G$, query $f(x)$, $f(y)$, $f(x+y)$ and accept iff $f(x+y) = f(x)+f(y)$.

If f is a homomorphism then our algorithm accepts f with probability 1.

Prop 1 Suppose that $f : C \to H$ has distance δ from homomorphism. Then our algorithm rejects f with probability at least $3\delta - 6\delta^2$

Note that $3\delta - 6\delta^2$ increases with δ only for $\delta \in [0, 1/4]$. for $\delta \geq 1/2$, it is useless.

Proof Let h be the homomorphism closes to f.

$$
\begin{aligned}
&\Pr_{x,y\in G}[f(x)+f(y) \neq f(x+y)] \\
\geq &\Pr_{x,y\in G}[(f(x) \neq h(x)) \wedge (f(y) = h(y)) \wedge (f(x+y) = h(x+y))] \\
+ &\Pr_{x,y\in G}[(f(x) = h(x)) \wedge (f(y) \neq h(y)) \wedge (f(x+y) = h(x+y))] \\
+ &\Pr_{x,y\in G}[(f(x) = h(x)) \wedge (f(y) = h(y)) \wedge (f(x+y) \neq h(x+y))]
\end{aligned}
$$

The three events described above are disjoint, and the inequality follows from the following:

$$
\begin{aligned}
&\Pr_{x,y\in G}[(f(x) \neq h(x)) \wedge (f(y) = h(y)) \wedge (f(x+y) = h(x+y))] \\
= &\Pr_{x\in G}[(f(x) \neq h(x))] - \Pr_{x,y\in G}[(f(x) \neq h(x)) \wedge [(f(y) \neq h(y)) \vee (f(x+y) \neq h(x+y))]] \\
\geq &\Pr_{x\in G}[(f(x) \neq h(x))] - \Pr_{x,y\in G}[(f(x) \neq h(x)) \wedge (f(y) \neq h(y))] \\
- &\Pr_{x,y\in G}[(f(x) \neq h(x)) \wedge (f(x+y) \neq h(x+y))] \\
= &\delta - \delta^2 - \delta^2
\end{aligned}
$$

Note that x, y are selected u.a.r. which means we can use the union bound in this manner.

We can expand the other two terms similarly, resulting in the full probability.

Full analysis Our algorithm is a one-sided error POT with detection probability $\min(\frac{\delta}{2}, \frac{1}{6})$.

The vote of y regarding $f(x)$:

$$Q_y(x) \stackrel{\text{def}}{=} f(x+y) - f(y)$$

We define $\phi(x) \stackrel{\text{def}}{=} v$ such that v maximizes $\{y \in G : \phi_y(x) = v\}$ to represent the most frequent vote regarding $f(x)$.

Note that if f is a homomorphism then $f(x) = f(x+y) - f(y)$

If f is a homomorphism then $Q_y(x) = f(x)\ \forall y, x$

Claim 1: ϕ is 2ρ-close to f where ρ is the rejection probability of f by the algorithm above.

We prove this by letting x be a bad point if $f(x)$ disagrees with more than half the votes. Let $B \stackrel{\text{def}}{=} \{x \in G : \Pr_{y \in G}[f(x) \neq Q_y(x)] \geq 1/2\}$.

$\delta(\phi, f) \leq 2\rho$

$$\begin{aligned}
\rho &= \Pr_{x,y\ inG}[f(x) \neq f(x+y) - f(y)] \\
&= \Pr_{x,y \in G}[f(x) \neq \phi_y(x)] \\
&\geq \Pr_{x \in G}[x \in B] \cdot \sum_{x \in B} \Pr_{y \in G}[f(x) \neq \phi_y(x)] \cdot \frac{1}{|B|} \\
&\geq \Pr_{x \in G}[x \in B] \cdot \min_{x \in B} \left\{ \Pr_{y \in G}[f(x) + Q_y(x)] \right\} \\
&\geq \frac{|B|}{|G|} \cdot \frac{1}{2} \\
&\Rightarrow |B| \leq 2\rho |G|
\end{aligned}$$

if $x \in G \setminus B$ then $f(x) = \phi(x)$ because $\Pr_y[f(x) = \phi_y(x)] > 1/2$.

Claim 2: For $x \in G$, $\Pr_{y \in G}[Q_y(x) = \phi(x)] \geq 1 - 2\rho$.

In other words, for all x's, $\phi(x)$ is the majority vote

To prove this, we want to show that $\Pr_{y \in G}[f(x+y) - f(y) = \phi(x)] \geq 1 - 2\rho \ \ \forall x$. Define a random variable $Z_x = Z_x(y) = f(x+y) - f(y)$ and show that it has a high collision probability. Specifically, we will show that:

$$\begin{aligned}
* &= \Pr_{y_1, y_2 \in G}[Z_x(y_1) = Z_x(y_2)] \geq 1 - 2\rho \\
&= \sum_{v \in H} \left(\Pr_{y \in G}[Z_x(y) = v] \right)^2 \\
&\leq^1 \max_{v \in H} \Pr_{y \in G}[Z_x(y) = v] \\
&= \Pr_{y \in G}[Z_x(y) = \phi(x)]
\end{aligned}$$

Overall, we have shown that y_1, y_2 are a good pair if:

1. $f(y_1) + f(-y1 + y_2) = f(y_2)$ which implies that $y_2 = (y_1) + (-y_1 + y_2)$
2. $f(x + y_1) + f(-y_1 + y_2) = f(x + y_2)$ which implies that $x + y_2 = (x + y_1) + (-y_1 + y_2)$

Therefore:

$$\Pr_{y_1, y_2 \in G}[f(y_1) + f(-y_1 + y_2) = f(y_2)] = 1 - \rho$$

note that all the above are distributed uniformly and independently in G.

$$\Pr_{y_1, y_2 \in G}[f(x + y_1) + f(-y_1 + y_2) = f(x + y_2)] = 1 - \rho$$

From this we can conclude that y_1, y_2 is a good pair with probability $1 - 2\rho$.

Let y_1, y_2 be a good pair. From this it follows that:

$$\begin{aligned}
Z_x(y_2) &= f(x + y_2) - f(y_2) \\
&= f(x + y_2) + f(-y_1 + y_2) - [f(y_2) + f(-y_1 + y_2)] \\
&= f(x + y_1) - f(y_1) = Z_x(y_1)
\end{aligned}$$

Thus, $* \geq 1 - 2\rho$ as desired.

References

[1] Oded Goldreich. *Introduction to property testing.* Cambridge University Press, 2017.

[1] this transition holds because $\sum_{v \in H} \Pr_{y \in G}[Z_x(y) = v] = 1$ and $\sum_{x \in A} x^2 \leq \sum_{x \in A} \max_{y \in A}\{y\} \cdot x = \max_{y \in A}\{y\} \cdot \overbrace{\sum_{x \in A} x}^{\leq 1}$