Lecture 2.1

- Self-correcting for linear fetus
- Testing linearity
Linear Functions:

\[ f: G \to H \]

\( G, H \) finite groups with operations \( \circ, \ast \)

closure, associative, identity, inverse

\( f \) is "linear" (homomorphism) if

\[ \forall x, y \in G \quad f(x \circ y) = f(x) \circ f(y) \]

Examples of finite groups:

\( G = \mathbb{Z}_m \) with operation " + mod m"

\( G = \mathbb{Z}_m^k \) with coordinatewise " + mod m"

\( (x_1, \ldots, x_k) \) s.t. \( x_i \in \mathbb{Z}_m \)

Examples of homomorphisms:

\[ f(x) = x \]

\[ f(x) = 0 \]

\[ f(x) = a \times \text{mod } q \quad \text{for } G = \mathbb{Z}_q \]

Today:

\text{every group is commutative!}
def \( f \) is "linear" (homomorphism) if \( \forall x, y \in G \) \( f(x) + f(y) = f(x + y) \)

\[ f + g \text{ agree on } \geq 1 - \epsilon \text{ fraction of inputs,} \]

\[ \Pr_{x \in G} [f(x) = g(x)] \geq 1 - \epsilon \]

else, \( f \) is "\( \epsilon \)-far" from linear

\[ \text{"}\epsilon\text{-close to linear"} \]
A useful observation:

\[ \forall a, y \in G \quad \Pr_x \left[ y = a + x \right] = \frac{1}{|G|} \]

since only \( x = a - y \) satisfies equation.

\[ \Rightarrow \text{if pick} \quad x \in_R G \]

\[ \text{then} \quad a + x \quad \text{is uniform dist in} \quad G \quad (a + x \in_R G) \]

distributed uniformly.

Example:

If \( G = \mathbb{Z}_2^n \) with operation \((a_1, \ldots, a_n) + (b_1, \ldots, b_n) = (a_1 \oplus b_1, a_2 \oplus b_2, \ldots, a_n \oplus b_n)\)

\[ (0110) + (b_1, b_2, b_3, b_4) = (0 \oplus b_1, 0 \oplus b_2, 1 \oplus b_3, 0 \oplus b_4) \]

is distributed uniformly if \( b_i \)'s are

why?

\[ \text{each coord uniform} \quad \Rightarrow \quad b_i \text{ indep} \quad \Rightarrow \quad a_i \oplus b_i \text{ 's indep} \]
Self-correcting: also known as “random self-reducibility”

Given \( f \) s.t. \( \exists \) linear \( g \) s.t. \( \Pr_{x} [f(x) = g(x)] \geq \frac{2}{3} \)

Can compute \( g(x) \) \( \forall x \): (using \( f \))

\[
\text{for } i = 1 \ldots c \log \frac{1}{\beta} \\
\text{Pick } y \in_R G \\
\text{answer}_i \leftarrow f(y) + f(x-y) \\
\text{Output most common value for answer}_i
\]

**Claim:** \( \Pr [\text{output} = g(x)] \geq 1 - \beta \)

**Pf.**

\[
\Pr [f(y) \neq g(y)] \leq \frac{1}{8} \\
\Pr [f(x-y) \neq g(x-y)] \leq \frac{1}{8} \\
\therefore \Pr [f(y) + f(x-y) \neq g(y) + g(x-y)] \leq \frac{1}{4} \\
\text{so each answer}_i = g(x) \text{ with prob } \geq \frac{3}{4} \\
\Rightarrow \text{most common answer value } = g(x) \text{ with prob } \geq 1 - \beta \text{ (Chernoff)}
\]
Linearity Testing

Goal: Given $f$

- if $f$ linear, pass
- if $f$ \( \varepsilon \)-far from linear, fail with prob \( \geq \frac{2}{3} \)

need to change value of $f$ on \( \geq \varepsilon \) fraction of domain
equivalently, $g$ linear $Pr_{x \in \Omega} [f(x) \neq g(x)] > \varepsilon$

Proposed Test

Do \( ? \) times:

Pick $x, y \in \Omega$
if $f(x) + f(y) \neq f(x+y)$ output "FAIL" and halt

Output "PASS"
Behavior of Test

$\frac{f \text{ linear}}{} \Rightarrow \text{ always passes}$

if $f$ $\varepsilon$-far from linear?

to show (contrapositive):

if $f$ likely to pass then $f$ is $\varepsilon$-linear

(equivalent: if $f$ $\varepsilon$-far from linear, $f$ likely to fail)
Plan

• if \( f \) \( \varepsilon \)-close to linear then find \( g \) you get from self-correcting \( f' \):

\[
g(x) = \text{majority}_y \left[ f(x + y) - f(y) \right]
\]

will be

1. \( f \) linear
2. \( f \) close to \( f' \)

• if \( f \) not close to linear, then no guarantees on \( g(x) \)
  
  but if test fails rarely, then you do get guarantees

  e.g., if most \( x \) satisfy \( f(x) = \text{majority}_y \left[ f(x + y) - f(y) \right] \)

  if \( f \) satisfies \( f(x) = \text{majority}_y \left[ f(x + y) - f(y) \right] \)
Theorem Suppose \( \delta = \Pr_{x,y} [f(x)+f(y) \neq f(xy)] \leq \frac{1}{16} \). Then \( f \) is \( 2\delta \)-close to linear.

# times we need to repeat lin test is \( \frac{1}{\delta} \) so \( \gg 16 \)

\( \frac{1}{\delta} \)

Proof Let \( g \) be the self-correction of \( f \):

\[
\text{def } g(x) = \text{plurality}_y [f(xy) - f(y)]
\]

\( y \)'s vote for \( f(x) \)

\( \leftarrow \) break ties arbitrarily

we will show: no ties

\[ \text{def } x \text{ is } \rho \text{-good if } \Pr_y [g(x) = f(xy) - f(y)] > 1 - \rho \]

\( \text{how many votes did } g(x) \text{ disagree with?} \)

Suppose \( 1 - \rho > \frac{1}{2} \), \( g(x) \) defined via clear majority

\( \frac{1}{2} \)-good \( x \): clear winner
First: $g + f$ usually agree

**Claim 1:** for $p < \frac{1}{2}$

$$ Pr_x [ x \text{ is } p\text{-good } \text{ and } g(x) = f(x)] > 1 - \frac{8}{p} $$

\[ \Rightarrow \text{fraction of } x \text{ for which } f + g \text{ agree is } > 1 - \frac{2\delta}{p} \geq \frac{7}{8} \text{ since } \delta < \frac{1}{6} \]

**Proof of Claim 1**

let $\alpha_x = Pr_y [ f(x) \neq f(x+y) - f(y)]$

If $\alpha_x < \frac{1}{2}$ then $x$ is $p$-good and $g(x) = f(x)$

$$ E_x [\alpha_x] = \frac{1}{|G|} \sum_{x \in G} Pr_y [ f(x) \neq f(x+y) - f(y)] $$

$$ = Pr_{x,y \in G} [ f(x) \neq f(x+y) - f(y)] = \delta $$

So $Pr [ \alpha_x > \rho ] \leq \frac{\delta}{p}$ \( \triangleq \text{Markov's} \)

So $Pr [ \alpha_x > \rho ] \leq \frac{\delta}{p} \leq \frac{\delta}{\frac{1}{6}}$
Second: Show $g$ "is a homomorphism" (at least, where it is defined)

Claim 2: $p < \frac{1}{4}$. If $x, y$ both $p$-good then

1. $x + y$ is $2p$-good
2. $g(x + y) = g(x) + g(y)$

Pt of Claim 2:

Let $h(x + y) = g(x) + g(y)$

\[
\begin{align*}
\Pr_x \left[ g(y) \neq f(y) - f(z) \right] &< p \quad \text{since $y$ is $p$-good} \\
\Pr_x \left[ g(x) \neq f(x + (y - z)) - f(y + z) \right] &< p \quad \text{since $x$ is $p$-good} \\
\end{align*}
\]

So

\[
\Pr_x \left[ h(x + y) = g(x) + g(y) \right] = \Pr_x \left[ g(y) \neq f(y - z) + f(x + (y - z)) - f(y + z) \right] = \Pr_x \left[ f(x + y - z) - f(z) = f(x + y) \right] > 1 - 2p
\]

$\geq \frac{1}{2}$ since $p < \frac{1}{4}$

\[
\Rightarrow \quad g(x + y) = h(x + y) \quad \text{by def of $g$} \quad \text{and since $f(x + y + z) - f(z) = h(x + y)$ for $\geq \frac{1}{2}$ $z$'s}
\]
Third: Show that $g$ is actually defined for all $x$.

**Claim 3** $\delta < \frac{1}{16}$, $\forall x$, $x$ is $4\delta$-good + $g(x)$

is defined via majority element

**Pf of Claim 3**

if $\exists y$ s.t. $y + (x - y)$ both $2\delta$-good

then claim 2 $\Rightarrow$ $x$ is $4\delta$-good

$4g(x) \neq g(y) + g(x - y)$

To show $y$ exists:

$Pr_y [y + (x - y) both 2\delta$-good] $\geq$ $1 - 2 \cdot \frac{\delta}{2\delta} = 0$

$\forall$ both uniform

$\delta = Pr_{xy} [f(x) + f(y) \neq f(x+y)] < \frac{1}{16}$

$def \ g(x) = plurality [f(xy) - f(y)]$

$def \ x is \ p$-good if $Pr_y [g(y) = f(xy) - f(y)] > 1 - \frac{\delta}{p}$

**Claim 1** for $p < \frac{1}{4}$$

$Pr_x [x is p$-good $\& g(x) = f(x)] > 1 - \frac{\delta}{p}$

$\Rightarrow$ fraction of $x$ for which $f \& g$ agree

is $> 1 - 2\delta > \frac{7}{8}$

**Claim 2** $p < \frac{1}{4}$. If $x, y$ both $p$-good then

1. $x + y$ is $2p$-good
2. $g(x+y) = g(x) + g(y)$

since $Pr > 0$

$\exists y \& y + x - y both 2\delta$-good

Claim 1

$\Rightarrow 1 - \frac{\delta}{p}$

$\geq 1 - \frac{1}{4} \cdot \frac{3}{4}$

of $x$ are $p$-good
Claim 1

Claim 2

Claim 3

\[ g(x) = \sum_1^{l(x)} y \text{ is defined via majority} \]

\[ \frac{1}{18} < p \leq \frac{1}{4} \]

\[ g(x) = \text{plurality of } \sum x_i \text{ of } X \]

\[ \sum x_i \text{ are } p \text{-good} \]

\[ g(x) \text{ is } 2p \text{-good} \]

\[ X \text{, } Y \text{ both } p \text{-good then} \]

\[ p \text{ agree on } 2-2p \]

\[ g(x+y) = g(x) + g(y) = 2p \]

\[ x+y \text{ is } 2p \text{-close} \]
Improvements: only need $\delta < 2/9$

$O(9^{1/2})$ tests give constant prob of failure instead of $O(16)$

big deal? can lead to improvements in exponents of hardness of approximation results.

Over $\mathbb{GF}(2)$, can get better $\delta$ in general $2/\delta$ is tight: (Coppersmith's example)

$$f(x) = \begin{cases} 1 & \text{if } x = 1 \mod 3 \\ 0 & \text{if } x = 0 \mod 3 \\ -1 & \text{if } x = 2 \mod 3 \end{cases}$$

f fails when $x = y = 1 \mod 3$ \# prob = $2/9$

else passes

gives:

$$\Pr[f(x) = g(x)] = \frac{1}{3}$$

$2/3$ - far