6.895 Randomness and Computation	February 15, 2006
Lecture 3	
Lecturer: Ronitt Rubinfeld	*Scribe: David Shin*

1 Linearity testing (continued)

We begin with some definitions and notation from last time. Throughout these notes, G will always represent a finite group, and F will represent a (not necessarily finite) group.

Definition 1 *A function $f : G \to F$ is* linear *if $f(x) + f(y) = f(x+y)$ $\forall x, y \in G$ (also referred to as a* homomorphism*).*

Definition 2 *The* probability of group law failure *of a function $f : G \to F$ is $\Pr_{x,y}[f(x) + f(y) \neq f(x+y)]$. We use δ_f to denote this probability.*

Definition 3 *A function $f : G \to F$ is ϵ-*linear *if there exists a linear function $g : G \to F$ s.t. $\Pr_x[f(x) = g(x)] \geq 1 - \epsilon$. Let ϵ_f denote the smallest value of ϵ for which the given inequality holds.*

Recall the ϵ-linear testing problem: we wish to determine whether a function $f : G \to F$ is ϵ-linear. The algorithm should adhere to the following specifications:

1. If f is linear, it should output *pass* with probability 1.
2. If f is not ϵ-linear, it should output *fail* with high probability.

The proposed test from last time was the following:

Algorithm A
do r times:
  pick $x, y \in_R G$
  if $f(x) + f(y) \neq f(x+y)$ output *fail*
output *pass*

It is straightforward to show by the sampling theorem that we referred to last time as the "gap bound" that if we set $r = \frac{c}{\delta} \log \frac{1}{\beta}$, algorithm A outputs *pass* with probability 1 if f is linear, and outputs *fail* with probability at least $1 - \beta$ if $\delta_f \geq \delta$.[1] To complete the analysis of this algorithm, then, we must simply establish some relationship between δ_f and ϵ_f. Such a relationship, however, is not as straightforward as one might expect. To illustrate, Coppersmith's example from last time shows that ϵ_f can be as big as $1/3$ when $\delta_f = 2/9$, while the following theorem, which we will not prove, implies that ϵ_f must be smaller than $1/9$ if $\delta_f < 2/9$:

Theorem 4 (Coppersmith) *If $\delta_f < 2/9$, then f is $\delta/2$-linear.*

One might capture this unusual phenomenon by stating that the function

$$\alpha(\delta) = \sup_{\delta_f \leq \delta} \epsilon_f$$

is discontinuous, where the supremum is taken over all groups G and all functions f with domain G.

Instead of proving Coppersmith's stronger result, we prove the following:

Theorem 5 *If $\delta_f < 1/16$, then f is $2\delta_f$-linear.*

[1] Using Chernoff bounds, one can actually distinguish the case when f is ϵ'-linear from the case when f is not even ϵ''-linear for constants $\epsilon' < \epsilon''$ in a similar number of samples.

Remark Professor Rubinfeld points out that the first proof of this kind applied only to infinite groups. As more proofs came along (using a variety of techniques), they increased in generality with respect to the domains that they applied to - next came finite cyclic groups, then finite abelian groups, and finally finite non-abelian groups. Later results decreased in generality with respect to the domains, but applied to more general ranges of δ.

Proof We introduce some further definitions and notation. For any function $q : G \to F$, let $\text{plurality}_s[q(s)]$ denote the element of F that occurs most often in the multiset obtained by applying q to each element of G. Ties are broken arbitrarily. Then, define the function $g : G \to F$ by $g(x) = \text{plurality}_y[f(x+y) - f(y)]$. Finally, call an element $x \in G$ *ρ-good* if $\Pr_y[g(x) = f(x+y) - f(y)] > 1 - \rho$.

We now give some claims.

Claim 6 *If $\rho < 1/2$, then $\Pr_x[x$ is ρ-good and $g(x) = f(x)] > 1 - \delta_f/\rho$.*

Claim 7 *If $\rho < 1/4$ and x, y are ρ-good,*

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

Claim 8 *If $\delta_f < 1/16$, then x is $4\delta_f$-good $\forall x$.*

Before we prove these claims, let us see how we can use them to prove the theorem. If $\delta_f < 1/16$, Claim 8 implies that x is $4\delta_f$-good for all $x \in G$. If we set ρ to any number between $4\delta_f < 1/4$ and $1/2$, then, Claim 6 implies that $\Pr_x[g(x) = f(x)] > 1 - \delta_f/\rho$. In particular, setting ρ arbitrarily close to $1/2$ yields $\Pr_x[g(x) = f(x)] > 1 - 2\delta_f$. Finally, since $4\delta_f < 1/4$, Claim 7 implies that g is in fact linear. These last two facts imply that f is $2\delta_f$-linear as desired.

We now proceed with the proofs of the claims.

Proof of Claim 6: Suppose $\rho < 1/2$. Let $\alpha_x = \Pr_y[f(x) \neq f(x+y) - f(y)]$. Note that if $\alpha_x < 1/2$, then $f(x) = f(x+y) - f(y)$ with probability strictly greater than $1/2$; i.e., $f(x) = \text{plurality}_y[f(x+y) - f(y)] = g(x)$. If $\alpha_x < \rho$, then, we have that x is ρ-good. It suffices to show that $\Pr[\alpha_x \geq \rho] < \delta_f/\rho$. But this follows directly from Markov's inequality[2], since

$$\begin{aligned}
\text{Exp}[\alpha_x] &= \frac{1}{|G|}\sum_{x \in G} \Pr_y[f(x) \neq f(x+y) - f(y)] \\
&= Pr_{x,y}[f(x) \neq f(x+y) - f(y)] \\
&= \delta_f.
\end{aligned}$$

■

Proof of Claim 7: Assume that $\rho < 1/4$ and that x, y are ρ-good. This means that

$$\Pr_z[g(y) \neq f(y+z) - f(z)] \leq \rho \tag{1}$$
$$\Pr_z[g(x) \neq f(x+y+z) - f(y+z)] \leq \rho \tag{2}$$

[2]Markov's inequality states that for any positive random variable Y and constant $a > 0$, $\Pr[Y \geq a \cdot \text{Exp}[Y]] \leq 1/a$.

The latter equation simply follows from the fact that for any fixed y, the distribution of $z + y$ over uniformly chosen z is again uniform. Now let $h = g(x) + g(y)$. With some basic manipulation, we find:

$$\begin{aligned}
& \Pr_z[h \neq f(x+y+z) - f(z)] \\
= & \Pr_z[g(x) + g(y) \neq (f(x+y+z) - f(y+z)) + (f(y+z) - f(z))] \\
\leq & \Pr_z[g(x) \neq f(x+y+z) - f(y+z)] + \Pr_z[g(y) \neq f(y+z) - f(z)] \qquad (3) \\
\leq & 2\rho. \qquad (4)
\end{aligned}$$

Here, (3) follows from the union bound, and (4) follows from (1) and (2). Thus, we have that $\Pr_z[h = f(x+y+z) - f(z)] \geq 1 - 2\rho > 1/2$, implying that $g(x+y) = h = g(x) + g(y)$ and that $x + y$ is 2ρ-good. ■

Proof of Claim 8: Let x be fixed. Claim 7 implies that if there exists some y such that y and $x - y$ are both $2\delta_f$-good, then x is $4\delta_f$-good. It suffices to show that such a y must exist. We use the probabilistic method. Applying Claim 6 twice with $\rho = 2\delta_f$ with the union bound, we find

$$\Pr_y[y \text{ and } x - y \text{ are } 2\delta_f\text{-good}] > 1 - \frac{\delta_f}{2\delta_f} - \frac{\delta_f}{2\delta_f} = 0.$$

Here, we are again making use of the fact that for any fixed y, the distribution of $z + y$ over uniformly chosen z is again uniform. ■

This proves the three claims, and hence proves the theorem. ■

Testing properties of convolutions of distributions As an aside, observe that if we let $\mathcal{F}$ denote the distribution induced by selecting $x \in_R G$ and outputting $f(x)$, then the linearity condition ($f(x+y) = f(x) + f(y)$) can be expressed as $\mathcal{F} = \mathcal{F} * \mathcal{F}$; i.e., $\mathcal{F}$ is equal to its own self-convolution. It can be shown that the only distributions that satisfy this relationship are distributions that are uniform over a subgroup of G (including the distribution that puts all weight on zero and the uniform distribution). It can also be shown that tests for ϵ-linearity are intimately related to the problem of testing whether a distribution $\mathcal{F}$ is *approximately* equal to its self convolution (according to some well chosen metric).

2 Fourier Representations

In this setting, we focus on functions $f : G \to F$ with $G = \{\pm 1\}^n$ and $F = \{\pm 1\}$. Our group operator will be component-wise multiplication (represented by $\cdot$).

It turns out that there are exactly 2^n linear functions from G to F. Namely, for each subset $S \subseteq [1 \ldots n]$, the function $\chi_S(x) = \prod_{i \in S} x_i$ is linear. One can easily check that these functions are in fact linear.

What happens if we apply a single iteration of algorithm A to test the linearity of a function f in this setting? It is easy to see that

$$f(x)f(y)f(x \cdot y) = \begin{cases} 1 & \text{if test passes} \\ -1 & \text{if test fails} \end{cases}$$

This suggests the introduction of an indicator function:

$$f'(x,y) = \frac{1 - f(x)f(y)f(x \cdot y)}{2} = \begin{cases} 0 & \text{if test passes} \\ 1 & \text{if test fails,} \end{cases}$$

giving us the nice relation

$$\Pr[\text{algorithm } A \text{ fails}] = E_{xy}[f'(x,y)].$$

At this point, we must turn to the tools of Fourier analysis. For now, we lay out some of the basic concepts.

We begin with widening our current framework to include functions $f : G \to F$ with $F = \mathbb{R}$ (G, as before, is $\{\pm 1\}^n$). If we let $\mathbb{G}$ represent the set of all such linear functions, it turns out that $\mathbb{G}$ is a vector space with dimension 2^n. We can characterize this vector space with the obvious basis $\{e_a(x)\}_{a\in G}$, where

$$e_a(x) = \begin{cases} 1 & \text{if } x = a \\ 0 & \text{else} \end{cases}.$$

A less obvious, but more useful basis, is $\{\chi_S \mid S \subseteq [1\ldots n]\}$. We conclude these notes by proving orthnormality of this proposed basis (but not the required spanning property). First, we need some notion of an inner product. For this, we define the inner product of f and g to be

$$\langle f, g\rangle = \frac{1}{2^n} \sum_{x\in\{\pm1\}^n} f(x)g(x).$$

One can easily verify that this satisfies the required properties of an inner product (namely, linearity, symmetry, and positive definiteness).

Claim 9 $\langle \chi_S, \chi_S\rangle = 1$ *for all* $S \subseteq [1\ldots n]$.

Proof Note that $\langle \chi_S, \chi_S\rangle = \frac{1}{2^n}\sum_x \chi_S^2(x)$. Since $\chi_S^2 \equiv 1$, the sum reduces to $\sum_x 1 = 2^n$, which gives the desired result. ■

Claim 10 $\langle \chi_S, \chi_T\rangle = 0$ *for all* $S, T \subseteq [1\ldots n]$ *with* $S \neq T$.

Proof Note that

$$\begin{aligned} \langle \chi_S, \chi_T\rangle &= \frac{1}{2^n}\sum_x \chi_S(x)\chi_T(x) \\ &= \frac{1}{2^n}\sum_x \prod_{i\in S} x_i \prod_{j\in T} x_j \end{aligned}$$

Now, we may write $S\cup T$ as the disjoint union $S\cup T = (S\cap T)\cup\Delta$, where $\Delta \neq \emptyset$. Additionally, if we define the Δ-*conjugate* of x, $\Delta(x)$, to be the element of G taken by switching the $\min(\Delta)$-th coordinate of x, we may note that summing over all x is equivalent to summing over all $\Delta(x)$. This is because G can be partitioned into Δ-conjugate pairs. The sum then becomes

$$\begin{aligned} & \frac{1}{2^n}\sum_x \prod_{i\in S\cap T} x_i^2 \prod_{j\in\Delta} x_j \\ = \; & \frac{1}{2^n}\sum_x \prod_{j\in\Delta} x_j \\ = \; & \frac{1}{2^{n+1}}\left(\sum_x \prod_{j\in\Delta} x_j + \sum_x \prod_{j\in\Delta} \Delta(x)_j\right) \\ = \; & \frac{1}{2^{n+1}}\sum_x (1 + (-1)) \prod_{\substack{j\in\Delta \\ j\neq\min(\Delta)}} 2x_j \\ = \; & 0. \end{aligned}$$

■