03683170: Expanders, Pseudorandomness and Derandomization 3/04/16

Lecture 3 – Small bias with respect to linear tests

Amnon Ta-Shma and Dean Doron

1 The Fourier expansion

1.1 Over general domains

Let G be a finite group with operation $+$ and identity 0. Then, the group algebra $\mathbb{C}[G]$ is the set of all functions $f \in \mathbb{C}[G]$. Obviously, it is a vector space on G over the field $\mathbb{C}$ of dimension $|G|$ and a natural basis for $\mathbb{C}[G]$ is

$$\mathbf{1}_g(x) \;=\; \begin{cases} 1, & x = g \\ 0, & \text{otherwise} \end{cases}$$

for every $g \in G$. It is also an inner-product space under the inner product

$$\langle f_1, f_2 \rangle = \mathop{\mathbb{E}}_{x \in G}[f_1(x)\overline{f_2(x)}] = \frac{1}{|G|}\sum_{x \in G} f_1(x)\overline{f_2(x)},$$

and it is easy to see that the basis $\{\mathbf{1}_g\}_{g \in G}$ is an orthogonal basis under this inner product. Often, one writes g instead of $\mathbf{1}_g$ and in this notation an element $f \in \mathbb{C}[G]$ is represented as $\sum_{g \in G} a_g g$ (which is often the notation used in quantum computation).

We now introduce another basis, that contains only functions that are homomorphisms from G to $\mathbb{C}^\times$.

Definition 1. *A* character *of the finite group G is a homomorphism $\chi : G \to \mathbb{C}^\times$, i.e., $\chi(x+y) = \chi(x)\chi(y)$ for every $x, y \in G$, where the addition is the group operation in G, and the multiplication is the group operation in $\mathbb{C}^\times$.*

We have the following easy facts:

Claim 2. *Let G be a finite group. Then:*

1. *$\chi_{\text{trivial}}(x) = 1$ for every $x \in G$ is a character. It is called the trivial character.*
2. *If χ_1 and χ_2 are characters of G then so is $\chi_1 \cdot \chi_2$ (where $\chi_1 \cdot \chi_2(x) = \chi_1(x)\chi_2(x)$).*
3. *For every character χ of G and $x \in G$, $|\chi(x)| = 1$ (the absolute norm is of course in $\mathbb{C}$). In particular, if we define $\overline{\chi_1}(x) = \overline{\chi_1(x)}$, then $\overline{\chi_1}$ is also a character and $\chi \cdot \overline{\chi} = \chi_{\text{trivial}}$.*
4. *This implies that $\widehat{G} = \{\chi \in \mathbb{C}[G] \mid \chi \text{ is a character}\}$ is an Abelian group, with identity as in item (1), multiplication as in item (2) and inverse as in item (3).*
5. *Let χ be a non-trivial character. Then, $\mathbb{E}[\chi] = 0$. This means that every non-trivial character is orthogonal to the trivial character.*

We can then show:

Claim 3. *Let G be a finite group. The set of all characters of G is orthonormal.*

Proof. First, note that $\langle \chi, \chi \rangle = \mathbb{E}[|\chi|^2] = 1$. Next, take χ_1 and χ_2 be two distinct characters of G. Then, $\langle \chi_1, \chi_2 \rangle = \mathbb{E}[\chi_1 \overline{\chi_2}]$. However, $\chi_1 \overline{\chi_2}$ is itself a character, and $\chi_1 \overline{\chi_2} = \chi_1 \chi_2^{-1} \not\equiv 1$ since they are distinct. Thus, $\mathbb{E}[\chi_1 \overline{\chi_2}] = 0$. □

As a consequence, G has at most $\dim(\mathbb{C}[G]) = |G|$ characters.

We will soon see that when G is Abelian, $\widehat{G}$ has a full set of characters. The resulting orthonormal basis for $G[\mathbb{C}]$ is called the *Fourier basis*, and the linear transformation between the natural basis and the Fourier basis is called the *Fourier transform*. Thus, every $f \in G[\mathbb{C}]$ can be (uniquely) written as $f = \sum_{g \in G} \hat{f}(g) \cdot \chi_g$, and the coefficients $\hat{f}(g)$ are called the Fourier coefficients.

Let us see some examples.

For $G = \mathbb{Z}_2$, it is easy to check that $\chi_1 \equiv 1$ and $\chi_2(x) = (-1)^x$ are characters (and we know that there are no more than 2). Next consider $G = \mathbb{Z}_m$ with addition modulo m. If χ is a character, and $x \in G$, then $\chi(x)^m = \chi(mx) = \chi(0) = 1$, hence, $\chi(x)$ is an m-th root of unity. Denote $\omega = e^{\frac{2\pi i}{m}}$. For $0 \le j < m$, define $\chi_j : \mathbb{Z}_m \to \mathbb{C}$ by $\chi_j(x) = \omega^{jx}$. It is easy to see that these are distinct characters of $\mathbb{Z}_m$ and since we have m of them, they are all the characters and $|\widehat{G}| = |G|$.

Let $f : \mathbb{Z}_m \to \mathbb{C}$. By now, we know that its Fourier expansion is given by $f(n) = \sum_{k=0}^{m-1} \hat{f}(k) \omega^{kn}$. If we treat f and $\hat{f}$ as vectors in $\mathbb{C}^m$, we get

$$f = \begin{pmatrix} \omega^{0\cdot 0} & \omega^{0\cdot 1} & \dots & \omega^{0\cdot(m-1)} \\ \omega^{1\cdot 0} & \omega^{1\cdot 1} & \dots & \omega^{1\cdot(m-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \omega^{(m-1)\cdot 0} & \omega^{(m-1)\cdot 1} & \dots & \omega^{(m-1)\cdot(m-1)} \end{pmatrix} \cdot \hat{f},$$

and the above matrix is called the Fourier matrix.

We now consider group products. Say $(A, \cdot)$, $(B, \cdot)$ are two groups. A and B are not necessarily Abelian, and we denote their operation by $\cdot$ rather than $+$ to (somewhat) distinguish them from the Abelian case. Let $G = A \times B$ (i.e., the group operation in G is $(a_1, b_1) \cdot (a_2, b_2) = (a_1 \cdot a_2, b_2 \cdot b_2)$). For $f \in \mathbb{C}[A]$ and $g \in \mathbb{C}[B]$, define $f \otimes g \in \mathbb{C}[A \times B]$ by $(f \otimes g)(a, b) = f(a)g(b)$. Then:

Claim 4. *If $f \in \widehat{A}$ and $g \in \widehat{B}$ then $f \otimes g \in \widehat{A \times B}$. Also, all pairs $f_i \otimes g_j$ for $f_i \in \widehat{A}$ and $g_j \in \widehat{B}$ are distinct.*

Back to the Abelian case, we see that if A and B are finite Abelian groups than Claim 4 gives us $|A| \cdot |B| = |A \times B|$ characters, and so we have a full set of characters. As every Abelian group can be decomposed as a product of cyclic groups, we have:

Exercise 5. *Let G be a finite Abelian group. Then $G \simeq \widehat{G}$.*

So what are the characters of $G = \mathbb{Z}_2^n = \mathbb{Z}_2 \times \ldots \times \mathbb{Z}_2$? By the above discussion, for every $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}_2^n$ we have the character

$$\chi_\alpha(x) = (\chi_{\alpha_1} \otimes \ldots \otimes \chi_{\alpha_n})(x_1, \ldots, x_n) = \prod_i \chi_{\alpha_i}(x_i) = \prod_i (-1)^{\alpha_i x_i} = (-1)^{\sum_i x_i \alpha_i}.$$

Equivalently we could say that for every $S \subseteq [n]$ there is the character $\chi_S(x) = (-1)^{\sum_{i \in S} x_i}$. The trivial character is the character of the empty set. Parity is the character of the full set (more precisely, $(-1)^{\text{parity}}$) and every function $f : \{0,1\}^n \to \mathbb{C}$ can be written as

$$f(a) \;=\; \sum_{S \subseteq [n]} \hat{f}(S)\chi_S(a) \;=\; \sum_{S \subseteq [n]} \hat{f}(S) \cdot (-1)^{\langle \mathbf{1}_S, a \rangle}.$$

We also see that the linear transformation converting the natural basis to the Fourier basis or vice versa, is the Hadamard matrix.

Also, because of the orthonormality of the characters:

Theorem 6. *For any $f, f_1, f_2 : G \to \mathbb{C}$,*

- $\hat{f}(S) = \langle f, \chi_S \rangle$.
- *(Parseval's Theorem)* $\langle f, f \rangle = \sum_{g \in G} \hat{f}(g)^2$.
- *(Plancherel's Theorem)* $\langle f_1, f_2 \rangle = \sum_{g \in G} \hat{f}_1(g)\hat{f}_2(g)$.

For example, let us take $f : \{0,1\}^3 \to \{0,1\}$ to be the majority function. Verify that $\hat{f}(\emptyset) = \frac{1}{2}$, $\hat{f}(\{1\}) = \hat{f}(\{2\}) = \hat{f}(\{3\}) = -\frac{1}{4}$, $\hat{f}(\{1,2\}) = \hat{f}(\{1,3\}) = \hat{f}(\{2,3\}) = 0$ and $\hat{f}(\{1,2,3\}) = \frac{1}{4}$. Also, you can check that Parseval's theorem holds, as $\langle f, f \rangle = \frac{1}{2}$.

We next give an intuitive explanation on the *meaning* of these numbers.

2 ε-biased sets

A set $T \subseteq \Lambda$ ε-fools a function $f : \Lambda \to \{0,1\}$ if $|\Pr_{x \in \Lambda}[f(x) = 1] - \Pr_{x \in T}[f(x) = 1]| \leq \varepsilon$. A set T ε-fools a class of functions $\mathcal{C}$ if it ε-fools every $f \in \mathcal{C}$. A set $T \subseteq \{0,1\}^k$ is called ε-biased if it ε-fools all functions $f : \{0,1\}^k \to \{0,1\}$ that are linear over $\mathbb{F}_2$. Formally:

Definition 7. *Let $T \subseteq \{0,1\}^k$. For a nonzero $w \in \{0,1\}^k$ we denote*

$$\text{bias}_w(T) \;=\; \left| \frac{1}{2} - \Pr_{s \in T}[\langle w, s \rangle = 1] \right|.$$

The bias of T is $\text{bias}(T) = \max_{w \neq 0} \text{bias}_w(T)$. *We say that T is ε-biased if* $\text{bias}(T) \leq \varepsilon$.

An ε-biased set tries to fool a class of functions using samples from a small set (in other words, we try to achieve pseudo-randomness with respect to a (very) limited class of tests). It is then natural to ask how *small* can ε-biased sets be. We shall soon answer this. But first, we interpret ε-bias using Fourier representation.

2.1 ε bias and the Fourier transform

Let X be a distribution over $\{0,1\}^n$ and $w \in \{0,1\}^n$. Then,

$$\begin{aligned}\text{bias}_w(X) &= \frac{1}{2}\left|\Pr_{s\sim X}[\langle w,s\rangle = 0] - \Pr_{s\sim X}[\langle w,s\rangle = 1]\right| = \left|\sum_{s\in\{0,1\}^n}(-1)^{\langle s,w\rangle}\cdot\Pr[X=s]\right| \\ &= \left|\sum_{s\in\{0,1\}^n} X(s)\chi_w(s)\right| = 2^n\cdot|\langle X,\chi_w\rangle| = 2^n\cdot|\hat{X}(w)|.\end{aligned}$$

Thus, we can redefine ε bias in the Fourier language.

Definition 8. *(equivalent to Def 7) Let $T \subseteq \{0,1\}^k$ and X the flat distribution over S. We say that T is ε-biased if $\hat{X}(S) \le \varepsilon 2^{-n}$ for all $S \neq \emptyset$.*

We prove:

Theorem 9 ([3, 2])**.** *Let X be distribution over $\{0,1\}^n$. Then $\| X - U_n \|_2 \le \text{bias}(X)$ and $\| X - U_n \|_1 \le 2^{n/2}\cdot\text{bias}(X)$.*

Proof. Express $X = \sum_w \hat{X}(w)\chi_w$. Now, $X - U_n = \sum_{w\neq\emptyset}\hat{X}(w)\chi_w$ (Why?). Let us first bound the ℓ_2 norm of $X - U_n$. We have:

$$\| X - U_n \|_2^2 = 2^n\langle X-U_n, X-U_n\rangle = 2^n\sum_{w\neq\emptyset}\hat{X}(w)^2 \le 2^n 2^n(\varepsilon 2^{-n})^2 = \varepsilon^2.$$

The bound on the ℓ_1 norm follows from Cauchy-Schwartz. □

In particular we see that if X has zero bias than X must be the uniform distribution.

2.2 ε bias and good binary error correcting codes

Definition 10. *An $[n,k]$ error correcting code C is ε-balanced if the Hamming weight of every non-zero codeword in C is between $\left(\frac{1}{2}-\varepsilon\right)n$ and $\left(\frac{1}{2}+\varepsilon\right)n$.*

Claim 11. *$M_{n\times k}$ is a generator matrix of an $[n,k]_2$ error correcting code that is ε-balanced, iff $\{r_i \mid r_i$ is the i'th row of $M\} \subseteq \{0,1\}^k$ is ε-biased.*

Proof. Let M be a generator matrix of an $[n,k]$ ε-balanced code C. For every $x \in \{0,1\}^k$, Mx contains at least $\left(\frac{1}{2}-\varepsilon\right)n$ nonzero entries and at most $\left(\frac{1}{2}+\varepsilon\right)n$. Hence, if we choose a row M_r of M uniformly at random, $\Pr_{r\in[n]}[\langle x, M_r\rangle = 1] \in [\frac{1}{2}\pm\varepsilon]$. It is then clear that the rows of M constitutes an ε-biased set in $\{0,1\}^k$ of size n. The other direction holds as well. We leave this to the reader. □

We are now ready to prove non-explicit existence.

Claim 12. *For every k, there exists an ε-biased set $T \subseteq \{0,1\}^k$ of size $n = O(\frac{k}{\varepsilon^2})$.*

Proof. Choose the entries of A, a binary matrix of dimension $n \times k$, uniformly at random. Fix a nonzero $x \in \{0,1\}^k$ and let W_x be the Hamming weight of Ax. That is, $W_x = \sum_{i=1}^{n} \langle A_i, x \rangle$ where the inner-product is modulo 2.

For a fixed non-zero x, $\mathbb{E}[W_x] = \frac{n}{2}$ (why?). By Chernoff, the probability that Ax is bad is at most

$$\Pr\left[\left|\frac{1}{n}W_x - \frac{1}{2}\right| \geq \varepsilon\right] \;\leq\; 2e^{-2n\varepsilon^2}.$$

By the union bound, the probability that A is a generator matrix for an unbalanced code is at most $2^k \cdot 2e^{-2n\varepsilon^2} \leq 2^{k+1-2n\varepsilon^2} < 1$, for $n \geq \frac{k}{\varepsilon^2}$. □

Non-explicitly the lower bound is $n = \Omega\left(\frac{k}{\epsilon^2 \log(\frac{1}{\varepsilon})}\right)$, and the same lower bound holds for $[n, k, \frac{1}{2} - \epsilon]_2$ codes (that are not necessarily ε-balanced, i.e., they may have high weight codewords).

2.3 An explicit construction

We now show a construction that achieves $n = O(\frac{k^2}{\varepsilon^2})$, due to Alon et al. [1]. The construction is Reed-Solomon concatenated with Hadamard. Specifically, we have the following ingredients:

- The outer code: An $R = [q = \frac{k_1}{\varepsilon}, k_1, 1 - \varepsilon]_q$ Reed-Solomon code, for q that is a power of 2.
- The inner code: An $H = [q, \log(q), \frac{1}{2}]_2$ Hadamard code.

Our code is the concatenation of the two codes, namely,

$$H(R(x)_1), \ldots, H(R(x)_q).$$

Then, the concatenated code $R \circ H$ is a $[n = q^2, k = k_1 \log q, \frac{1}{2}(1 - \varepsilon)]$ linear error correcting code. Now, $q = \frac{k_1}{\varepsilon} = \frac{k}{\varepsilon \log q}$ and so $n \leq (\frac{k}{\varepsilon})^2$. This shows that in the code there are no nonzero codewords of weight smaller than $\frac{1}{2}(1 - \varepsilon)$. In fact, the concatenated code also does have any codewords of length more than $\frac{1}{2}$ (why?) and so we get an ε-balanced code as needed.

2.4 Almost k-wise independence

Definition 13. *Let X be a distribution over $\{0,1\}^n$.*

- *We say X is* (k, ε)-biased*, if it is at most ε-biased with respect to all non-empty, linear tests of size at most k.*
- *We say X is* (k, ε)-wise independent *if for all $S \subseteq [n]$ of size k, $|X_{|S} - U_k| \leq \varepsilon$.*

Theorem 14 ([3])**.** *There exists an explicit distribution that is (k, ε)-biased over $\{0,1\}^n$ and has support size at most $\left(\frac{k \log n}{\varepsilon}\right)^2$.*

Proof. For the construction we combine two ingredients that we already have: k-wise independence and ε-bias. Let

- A of size $n \times h$ be the generator matrix of a k-wise sample space. We saw (in Lecture 2) how to construct A with $2^h = n^k$ (and in fact, we can even get $2^h = n^{k/2}$).
- Sample $b \in B$, where $B \subseteq \{0,1\}^h$ is an ε-biased sample space. We saw how to construct B with support size $\left(\frac{h}{\varepsilon}\right)^2$ (and we mention that, in fact, we can even get support size $O\left(\frac{h}{\varepsilon^2}\right)$).

The construction: Sample $b \in B$ output $Ab \in \{0,1\}^n$.

Let $S \subseteq [n]$ be a set of size at most k. We want to bound $\text{bias}_S(Ab)$. Let A_i be the i-th row of A. It holds that:

$$\oplus_{i \in S} \ (Ab)_i = \oplus_{i \in S} \ \langle A_i, b \rangle = \left\langle \sum_{i \in S} A_i, b \right\rangle,$$

so $\Pr_{b \in B}[\oplus_{i \in S} Ab_i = 1] \in [\frac{1}{2} \pm \varepsilon]$ because the vectors $\{A_i\}_{i \in S}$ are linearly independent and so $\sum_{i \in S} A_i$ is nonzero and B forms an ε-biased distribution. The support size is $(\frac{h}{\varepsilon})^2 = (\frac{k \log n}{\varepsilon})^2$. □

It therefore follows:

Corollary 15. *There exists an explicit distribution that is (k, ε)-wise independent over $\{0,1\}^n$ and has support size at most $2^k \left(\frac{k \log n}{\varepsilon}\right)^2$.*

Proof. By Theorem 9, an (k, ε')-biased distribution is $(k, \varepsilon' \cdot 2^{k/2})$-wise independent. Setting $\varepsilon' = 2^{-k/2}\varepsilon$, we are finished. □

3 The Fourier transform as a multilinear representation

We now choose to work with the group $\mathbb{Z}_{2,\cdot} = (\{1,-1\}, \cdot)$ instead of $\mathbb{Z}_{2,+} = (\{0,1\}, + \bmod 2)$ as we did so far. The two groups are isomorphic with the isomorphism $\psi : b \mapsto (-1)^b$. The two characters of $\mathbb{Z}_{2,\cdot}$ are $\mathbf{1}(x) = 1$ and $\mathbf{x}(x) = x$. Consequently, the characters of $\mathbb{Z}_{2,\cdot}^n$ are $\prod_{i \in S} x_i$. If we take $f' : \mathbb{Z}_{2,\cdot}^n \to \mathbb{C}$, then its Fourier representation tells us how to open f' as a multi-linear function over $\mathbb{C}$.

We can identify a boolean function $f : \{0,1\}^n \to \{0,1\}$ with a function $f' : \{-1,1\}^n \to \{1,-1\}$ defined by

$$f'(\psi(b_1), \ldots, \psi(b_n)) = \psi(f(b_1, \ldots, b_n)).$$

It turns out that the Fourier representation of f in $\mathbb{Z}_{2,+}^n$ is closely related to the Fourier representation of f' in $\mathbb{Z}_{2,\cdot}^n$:

Exercise 16. *Suppose $f(x) = \sum_S \hat{f}(S)\chi_S(x)$ and $f'(y) = \sum_S \hat{f}(S) \prod_{i \in S} y_i$. Then $\hat{f'}(\emptyset) = 1 - 2\hat{f}(\emptyset)$ and $\hat{f'}(S) = -2\hat{f}(S)$ for all $S \neq \emptyset$.*

Hint: $\psi(b) = 1 - 2b$.

Thus, the Fourier expansion of f tells how f' can be represented as a multilinear function. However, the translation between f and f' is linear ($f = \frac{1}{2}(1 - f')$), as is the translation between the variables ($y_i = 1 - 2x_i$) and so this also tells how f can be represented as a multilinear function over $\mathbb{C}$. Also, $\max_{S : \hat{f}(S) \neq 0} |S|$ is the degree of the multilinear polynomial computing f' over $\mathbb{C}$. Thus, the Parity function $f(x_1, \ldots, x_n) = \sum_i x_i$ is linear over $\mathbb{F}_2$ but has degree n over $\mathbb{C}$.

From this discussion it is clear that the Fourier representation can help determine how close a function on $\{0,1\}^n$ is to being linear, or to a low-degree multilinear function. We will see that it also helps in unexpected places, e.g., in determining the resiliency of a function.

4 Back to resilient functions

Throughout, f is a function from $\{0,1\}^n$ to $\{0,1\}$, although other variants can be considered as well. We begin by noticing that the expectation and variance of f both have simple formulas using the Fourier coefficients:

$$\begin{aligned} \mathbb{E}[f] &= \langle f, \mathbf{1} \rangle = \langle f, \chi_\emptyset \rangle = \hat{f}(\emptyset) \\ \mathrm{Var}[f] &= \mathbb{E}[f^2] - \mathbb{E}[f]^2 = \sum_{S \subseteq n} \hat{f}(S)^2 - \hat{f}(\emptyset)^2 = \sum_{S \neq \emptyset} \hat{f}(S)^2. \end{aligned}$$

We already mentioned that the influence of a single variable x_i is defined as the probability that changing this variable will alter the result of the function f. We abbreviate $I_{\{x_i\}} = I_i$, and note that

$$I_i(f) = \Pr_x[f(x) \neq f(x \oplus e_i)].$$

If we denote $f_i(x) = f(x) - f(x \oplus e_i)$, it holds that $I_i(f) = \Pr_x[f_i(x) \neq 0]$. It is then immediate that $I_i(f) = \mathbb{E}[f_i^2]$.

Claim 17. $f_i = 2 \sum_{S: i \in S} \hat{f}(S) \chi_S$.

Proof. For every $S \subseteq [n]$,

$$\hat{f}_i(S) = \langle f_i, \chi_S \rangle = \langle f, \chi_S \rangle - \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x \oplus e_i) \chi_S(x).$$

Denote $y = x \oplus e_i$, so $x = y \oplus e_i$ and

$$\hat{f}_i(S) = \langle f, \chi_S \rangle - \frac{1}{2^n} \sum_{y \in \{0,1\}^n} f(y) \chi_S(y) \chi_S(e_i).$$

If $i \in S$ then $\chi_S(e_i) = -1$ and $\frac{1}{2^n} \sum_y f(y) \chi_S(y) \chi_S(e_i) = -\langle f, \chi_S \rangle$. Otherwise, $\chi_S(e_i) = 1$ and $\frac{1}{2^n} \sum_y f(y) \chi_S(y) \chi_S(e_i) = \langle f, \chi_S \rangle$. This finished the proof. □

We can then conclude:

Corollary 18. $I_i(f) = 4 \sum_{S: i \in S} \hat{f}(S)^2$.

Proof. By Parseval's theorem and our previous observations,

$$I_i(f) = \mathbb{E}[f_i^2] = \sum_{S \subseteq [n]} \hat{f}_i(S)^2 = \sum_{S: i \in S} (2\hat{f}(S))^2 = 4 \sum_{S: i \in S} \hat{f}(S)^2.$$

□

In words, the influence of a variable x_i is proportional to the sum of the squares of the Fourier coefficients related to sets that contain i.[1] We define the *total influence* of f by $I(f) = \sum_{i=1}^{n} I_i(f)$. We then have:

Corollary 19. $I(f) = 4 \sum_{S \subseteq [n]} \hat{f}(S)^2 |S|$.

Proof. By the previous observation,

$$\sum_{i=1}^{n} I_i(f) = 4 \sum_{i=1}^{n} \sum_{S: i \in S} \hat{f}(S)^2 = 4 \sum_{S \subseteq [n]} \sum_{i \in S} \hat{f}(S)^2 = 4 \sum_{S \subseteq [n]} \hat{f}(S)^2 |S|.$$

□

Now, note that $I(f) = 4 \sum_S \hat{f}(S)^2 |S| \geq \sum_{S \neq \emptyset} \hat{f}(S)^2 = \mathrm{Var}[f]$, so the variance is bounded by the total influence. We can hence conclude:

Theorem 20. *For every $f : \{0,1\}^n \to \{0,1\}$ there exists an $i \in [n]$ for which $I_i(f) \geq \frac{\mathrm{Var}[f]}{n}$.*

To get to KKL's bound of $\frac{\log n}{n} \cdot \mathrm{Var}[f]$, one has to work a bit harder, and use the Bonami-Beckner inequality.

5 Linearity testing

See Section 1.6 of Ryan O'Donnell's book [4].

References

[1] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992.

[2] Oded Goldreich. Three xor-lemmasan exposition. In *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*, pages 248–272. Springer, 2011.

[3] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM journal on computing*, 22(4):838–856, 1993.

[4] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[1] When normalized appropriately so that $\|\hat{f}\| = 1$, this means that the influence of i is proportional to the probability that $i \in S$ when we sample according to $\hat{f}$.