Notes for Lecture 4

In the last lecture we defined the Fourier coefficients of Boolean functions and studied some of their properties. Recall that

- Any Boolean function $f : \{0,1\}^n \to \{-1,1\}$ may be expressed *uniquely* as

$$f(x) = \sum_{S \subseteq \{1,\dots,n\}} \widehat{f}(S) u_S(x)$$

  where $u_S(x) = (-1)^{\sum_{i \in S} x_i}$ and $\widehat{f}(S) = (f, u_S) = \mathbb{E}_{x \sim \{0,1\}^n} f(x) u_S(x)$.
- $\widehat{f}(\emptyset) = \mathbb{E}_{x \sim \{0,1\}^n} f(x)$
- $\sum_S |\widehat{f}(S)| \le \sqrt{2^n}$
- $\sum_S \widehat{f}^2(S) = 1$
- If $z_1, \dots, z_n$ are ε-biased and $f : \{0,1\}^n \to \mathbb{R}$ then

$$\left| \mathop{\mathbb{E}}_{x \sim \{0,1\}^n} f(x) - \mathbb{E} f(z_1, \dots, z_n) \right| \le 2\varepsilon \sum_{S \ne \emptyset} \left| \widehat{f}(S) \right|$$

- If $z_1, \dots, z_n$ are ε-biased and $f : \{0,1\}^n \to \{-1,1\}$ then

$$\left| \mathop{\mathbf{Pr}}_{x \sim \{0,1\}^n} [f(x) = 1] - \mathbf{Pr}[f(z_1, \dots, z_n) = 1] \right| \le \varepsilon \sum_{S \ne \emptyset} \left| \widehat{f}(S) \right|$$

- If a Boolean function f depends on only k of its input bits then $\sum_{S \ne \emptyset} |\widehat{f}(S)| \le \sqrt{2^k}$.

Today we will see some classes of functions for which ε-biased distributions are ε-pseudorandom. Fix $(a_1, \dots a_k) \in \{0,1\}^k$. Let $f : \{0,1\}^k \to \{-1,1\}$ be defined by

$$f(x_1, \dots, x_k) = \begin{cases} -1 & \text{if } \forall i \quad x_i = a_i \\ 1 & \text{otherwise} \end{cases}$$

Let us estimate the Fourier coefficients of f. For any $S \ne \emptyset$ we have

$$\widehat{f}(S) = \mathbb{E} f(x) u_S(x) = 2\,\mathbf{Pr}[f(x) = u_S(x)] - 1$$

Since $\frac{1}{2} - \frac{1}{2^k} \le \mathbf{Pr}[f(x) = u_S(x)] \le \frac{1}{2} + \frac{1}{2^k}$ we have $\left|\widehat{f}(S)\right| \le \frac{2}{2^k}$. It follows that $\sum_{S \ne \emptyset} \left|\widehat{f}(S)\right| \le \frac{2(2^k - 1)}{2^k} \le 2$. Thus for every fixed pattern $a = (a_1, \dots, a_k)$ and every ε-biased random variable $z = (z_1, \dots, z_k)$, the function f as defined above satisfies $|\mathbf{Pr}[f(x) = -1] - \mathbf{Pr}[f(z) = -1]| \le 2\varepsilon$, *i. e.*, $\left|\frac{1}{2^k} - \mathbf{Pr}[z = a]\right| \le 2\varepsilon$.

Now recall that if $f : \{0,1\}^n \to \{-1,1\}$ is a function that depends on only k of its input bits, $i_1, \ldots, i_k$ then $\widehat{f}(S) = 0$ for every S which is not a subset of $\{i_1, \ldots, i_k\}$. It follows that for any k-bit pattern $a_{i_1}, \ldots, a_{i_k}$ the function $f : \{0,1\}^n \to \{-1,1\}$ defined by

$$f(x_1, \ldots, x_k) = \begin{cases} -1 & \text{if } \forall j \quad x_{i_j} = a_{i_j} \\ 1 & \text{otherwise} \end{cases}$$

has the property that $\sum_{S \neq \emptyset} \left| \widehat{f}(S) \right| \leq 2$.

A decision tree on n inputs is a rooted binary tree in which each non-leaf node is labelled with some input variable $x_i, 1 \leq i \leq n$. The two subtrees of the node correspond to the further computation when the variable x_i at the node takes value 0 or 1. Each leaf is labelled with a return value of the function to be computed. Computation on input $x = (x_1, \ldots, x_n)$ proceeds by starting at the root, and examining the variable x_i with which the root is labelled. Depending on the value of x_i computation proceeds recursively in the appropriate subtree. When a leaf is reached, the value of the leaf is returned. Note that each leaf corresponds to a (partial) pattern of variable settings. By *size* of the decision tree we will mean the number of leaves. We will show that the size of a decision tree is an upper bound on the sum of the non-principal Fourier coefficients of the function it computes.

Suppose $f : \{0,1\}^n \to \{-1,1\}$ is computable by a decision tree with m leaves. (Note that we are not making any assumption about the optimality of the tree.) For each leaf ℓ of the decision tree we define the auxiliary function

$$f_\ell(x) = \begin{cases} 0 & \text{if computation on } x \text{ does not lead to } \ell \\ \text{output of } \ell & \text{otherwise} \end{cases}$$

i. e., f_ℓ is non-zero only on inputs that lead the conputation to ℓ. (Note that f_ℓ is not a boolean function.) Since every input x leads the computation to exactly one leaf, we have

$$f(x_1, \ldots, x_n) = \sum_\ell f_\ell(x_1, \ldots, x_n).$$

Taking Fourier expansions of the auxiliary functions we have

$$f(x) = \sum_\ell f_\ell(x) = \sum_\ell \sum_S \widehat{f_\ell}(S) u_S(x) = \sum_S \left(\sum_\ell \widehat{f_\ell}(S) \right) u_S(x).$$

By the uniqueness of Fourier expansions, $\widehat{f}(S) = \sum_\ell \widehat{f_\ell}(S)$ for each S, and we have

$$\sum_{S \neq \emptyset} \left| \widehat{f}(S) \right| = \sum_{S \neq \emptyset} \left| \sum_\ell \widehat{f_\ell}(S) \right| \leq \sum_{S \neq \emptyset} \sum_\ell \left| \widehat{f_\ell}(S) \right| = \sum_\ell \left(\sum_{S \neq \emptyset} \left| \widehat{f_\ell}(S) \right| \right).$$

Thus it suffices to bound the sum of the non-principal Fourier coefficients for each $\widehat{f_\ell}$.

For each leaf ℓ define the function $g_\ell : \{0,1\}^n \to \{-1,1\}$ as follows:

$$g_\ell(x) = 1 - 2(\text{output of } \ell) f_\ell(x) = \begin{cases} -1 & \text{if } x \text{ leads to } \ell \\ 1 & \text{otherwise} \end{cases}$$

Then g_ℓ is a Boolean function corresponding to a fixed (partial) pattern of settings of the input variables. For such g_ℓ we have already seen that $\sum_{S\neq\emptyset}|\widehat{g_\ell}(S)| \leq 2$. Now for $S \neq \emptyset$,

$$\widehat{g_\ell}(S) = (g_\ell, u_S) = (1 \pm 2f_\ell, u_S) = (1, u_S) \pm 2(f_\ell, u_S) = 0 \pm 2(f_\ell, u_S) = \pm 2\widehat{f_\ell}(S)$$

so that $|\widehat{g_\ell}(S)| = 2\left|\widehat{f_\ell}(S)\right|$ and hence $\sum_{S\neq\emptyset}\left|\widehat{f_\ell}(S)\right| \leq 1$. We have shown

Theorem 1 *If a function $f : \{0,1\}^n \to \{-1,1\}$ is computable by a decision tree with m leaves then $\sum_{S\neq\emptyset}\left|\widehat{f}(S)\right| \leq m$.*

In fact the same conclusion holds if instead of decision trees we have generalized decision trees in which the non-leaf nodes are labelled with parities of (subsets of) input variables.

To see this, note that since parity is addition mod 2, a path from the root to a leaf corresponds to a system $Ax = b$ of linear equations over $\mathbb{F}_2$. Let r be the rank of A. For a leaf ℓ define

$$g_\ell(x) = \begin{cases} -1 & \text{if } Ax = b \\ 1 & \text{otherwise} \end{cases}$$

Since $f(x) = \sum_\ell \frac{g(x)-1}{2}(\text{output of } \ell)$, once again it is sufficient to show that $\sum_{S\neq\emptyset}|\widehat{g_\ell}(S)| \leq 2$.
If the system $Ax = b$ is inconsistent, then $g_\ell \equiv 1$ and for all non-empty S, $\widehat{g_\ell}(S) = 0$. If it is consistent, then without loss of generality (by deleting redundant rows) we may assume that the rows of A are linearly independent mod 2 (*i. e.*, A is an $r \times n$ matrix). Let M be any full-rank $n \times n$ matrix over $\{0,1\}$ which agrees with A on the first r rows. (Such a matrix M exists because a linearly independent set of vectors may always be extended to a basis.) Let $h : \{0,1\}^n \to \{-1,1\}$ be defined by

$$h(y) = \begin{cases} -1 & \text{if for } 1 \leq i \leq r, \quad y_i = b_i \\ 1 & \text{otherwise} \end{cases}$$

Then we know that $\sum_{S\neq\emptyset}\left|\widehat{h}(S)\right| \leq 2$. Also $g_\ell(x) = h(Mx)$ and we have

$$\begin{aligned}
\widehat{g_\ell}(S) &= (g_\ell, u_S) \\
&= \mathbb{E}\, g_\ell(x)(-1)^{S^t x} \\
&= \mathbb{E}\, h(Mx)(-1)^{S^t(M^{-1}Mx)} \\
&= \mathbb{E}\, h(Mx)(-1)^{(S^t M^{-1})(Mx)} \\
&= (h, \mathcal{U}_{(M^{-1})^t S}) \\
&= \widehat{h}((M^{-1})^t S)
\end{aligned}$$

where the superscript t denotes transpose and by abuse of notation we use S to denote the $\{0,1\}$-vector of inclusion in set S. Under this identification, $(M^{-1})^t S$ is some other subset of $\{1,\ldots,n\}$ and we've shown that the non-principal Fourier coefficients of g_ℓ are just some permutation of the non-principal Fourier coefficients of h. It follows that $\sum_{S\neq\emptyset}|\widehat{g_\ell}(S)| \leq 2$.

Thus we have shown that an (ε/m)-biased distribution is ε-pseudorandom for the class of all Boolean functions f that are computable by decision trees (with nodes labelled by parities) with m leaves.

Next we will consider consider functions representable by k-CNF formulas of small size. We'll say $f : \{0,1\}^n \to \{-1,1\}$ is a k-CNF with m clauses if there is a k-CNF formula φ with m clauses such that

$$f(x_1, \ldots, x_n) = \begin{cases} -1 & \text{if } \varphi \text{ is satisfied by } x_1, \ldots, x_n \\ 1 & \text{otherwise} \end{cases}$$

It turns out that there are 2-CNF functions f with $m = O(n)$ clauses for which $\sum_{S \neq \emptyset} \left| \widehat{f}(S) \right| = 2^{\Omega(n)}$. Therefore we cannot apply the previous arguments to show pseudorandomness of ε-biased distributions for this class. However it can be shown that if f is a k-CNF with m clauses and $z_1, \ldots, z_n$ is $\varepsilon^{O(k2^k)}$-biased then $(z_1, \ldots, z_n)$ is ϵ-pseudorandom for f.

Conjecture 1 *If $z_1, \ldots, z_n$ is poly(ε/m)-biased then it is ε-pseudorandom for the class of k-CNFs with m clauses.*

Conjecture 2 *If $z_1, \ldots, z_n$ is $\left(\frac{1}{2^{\log(s/\varepsilon)^{O(d)}}}\right)$-biased then it is ε-pseudorandom for the class of functions that are computable by circuits of size s and depth d.*