A pseudorandom distribution is one that appears uniformly random, but contains some correlations between the inputs. Usually we are given a class of computations, or tests $\mathcal{F} = \{f\colon \{0,1\}^n \to [-1,1]\}$, and the objective is to design a pseudorandom distribution $pr$ such that for every $f$ in $\mathcal{F}$,

$$|\mathrm{E}_u[f(u)] - \mathrm{E}_{pr}[f(pr)]| < \varepsilon, \tag{1}$$

where $u$ is the uniform distribution over $\{0,1\}^n$. An example we already saw of such a distribution are small-biased distributions. A distribution $pr$ is $\varepsilon$-biased if it satisfies equation (1) for all character functions $\chi_a(x) = (-1)^{\langle a,x\rangle}$, where $a \in \{0,1\}^n$.

The *support size* of a distribution is the number of possible values in $\{0,1\}^n$ that $pr$ can take. In the theory of pseudorandomness the objective is to construct small-biased distributions with as small support size as possible. Recall that a small-biased distribution can be obtained by uniformly sampling a column vector from a generator matrix of a linear code whose minimum distance is at least $(1-\varepsilon)/2$ and maximum distance is at most $(1+\varepsilon)/2$. A straightforward extension of the Gilbert-Varshamov bound shows that there exist small0biased distributions of support size $O(n/\varepsilon^2)$.

Is this the best possible? Later in the class I hope to show a lower bound of $\Omega(n/\varepsilon^2 \log(1/\varepsilon))$ on the support size. But today I want to show a weaker argument that shows a lower bound of, say $\mathrm{poly}(n)/\varepsilon^{1.9}$, *provided $\varepsilon$ is sufficiently small in terms of $n$*. This will illustrate how small-biased distributions interact nicely with the Fourier expansion of boolean functions and lead us to two applications of small-biased distributions and Fourier analysis to pseudorandomness.

## 1 The support size of distributions with very small bias

Take an arbitrary function $f$. We rewrite the pseudorandomness requirement (1) as:

$$|\mathrm{E}_u[f(u)] - \mathrm{E}_{pr}[f(pr)]| \leq \varepsilon$$

If $pr$ is an $\varepsilon$-biased distribution, then we can take the Fourier transform and rewrite the left-hand side as

$$\begin{aligned}\mathrm{E}_u[f(u)] - \mathrm{E}_{pr}[f(pr)] &= \hat{f}_0 - \mathrm{E}_{pr}\left[\sum\nolimits_{a\in\{0,1\}^n} \hat{f}_a \chi_a(pr)\right] \\ &= \sum\nolimits_{a\neq 0} \hat{f}_a \,\mathrm{E}_{pr}[\chi_a(pr)]\end{aligned}$$

by linearity of expectation and because $\chi_0 = 1$. Taking absolute values and applying the triangle inequality, we obtain

$$|\mathrm{E}_u[f(u)] - \mathrm{E}_{pr}[f(pr)]| \leq \sum\nolimits_{a\neq 0} |\hat{f}_a| \cdot |\mathrm{E}_{pr}[\chi_a(pr)]|$$

and so

$$|\mathrm{E}_u[f(u)] - \mathrm{E}_{pr}[f(pr)]| \leq \varepsilon \cdot \sum\nolimits_{a\neq 0} |\hat{f}_a|. \tag{2}$$

This inequality holds for any boolean function $f$. By the Cauchy-Schwarz inequality, the left-hand side is at most

$$\varepsilon \cdot \sum_{a\neq 0} 1 \cdot |\hat{f}_a| \leq \varepsilon \sqrt{\sum_{a\neq 0} 1} \cdot \sqrt{\sum_{a\neq 0} \hat{f}_a^2} \leq \varepsilon \cdot 2^{n/2}. \tag{3}$$

Now suppose we had an $\varepsilon$-biased distribution with support size, say, $\mathrm{poly}(n)/\varepsilon^{1.9}$ for every $n$ and $\varepsilon$. Plugging in $\varepsilon = 0.1 \cdot 2^{-n/2}$, we would get a distribution of support size less than $2^n/2$ that is 0.1-pseudorandom for the class of *all* functions $\{f\colon \{0,1\}^n \to [-1,1]\}$. In particular, let $f$ be the function that evaluates to 0 when $x$ is in the support of $pr$, and 1 otherwise. Then

$$\mathrm{E}_{pr}[f(pr)] = 0 \quad \text{while} \quad \mathrm{E}_u[f(u)] \geq 1/2$$

which is a contradiction.

## 2 Branching programs

Inequality (2) tells us that small-bias distribution work well whenever one can show that the quantity $\sum_{a\neq 0}|\hat{f}_a|$ is not too large. Unfortunately, this does not happen too often. Certainly it happens for the character functions $\chi_a$. A simple model that generalizes these functions is that of width two branching programs.

Imagine the following kind of computation. You read the input bits in some fixed in advance order, repetitions allowed, say $x_3$, $x_5$, $x_2$, $x_1$, $x_3$ again, $x_2$ again, $x_4$. In between reading consecutive input bits, you can store one bit of information. The output of the computation is the last bit in memory. This computation, called a branching program of width two, allows you to compute all linear functions, but also functions like arbitrary ANDs of inputs and some more unusual ones ones, like $((x_1 \oplus x_2) \vee x_3) \wedge \overline{x_2}$.

For the following claim it will be easier to think of the branching program as taking values 1 and $-1$.

**Claim 1.** *Let $F\colon \{0,1\}^n \to \{-1,1\}$ be a branching program of width two that reads $\ell$ inputs, possibly with repetitions. Then*

$$\sum_a |\hat{f}_a| \leq \ell + 1.$$

*Proof.* By induction on $\ell$. When $\ell = 0$, the only width two branching programs are the functions 0 and 1, for which the claim clearly holds. Now suppose it holds for width two branching programs of length $\ell$. If $F$ has length $\ell + 1$, we can write $F(x) = h(g(x), x_i)$, where $g(x)$ describes the first $\ell - 1$ steps of the computation and $h$ is the last step. Let $G = (-1)^g$. The Fourier representation of $h\colon \{0,1\}^2 \to \{-1,1\}$ looks like this:

$$\begin{aligned} h(g, x_i) &= \hat{h}_{00} + \hat{h}_{10}G + \hat{h}_{01}\chi(x_i) + \hat{h}_{11}G\chi(x_i) \\ &= (\hat{h}_{00} + \hat{h}_{01}\chi(x_i)) + (\hat{h}_{10} + \hat{h}_{11}\chi(x_i))G. \end{aligned}$$

and therefore

$$h(g(x), x_i) = (\hat{h}_{00} + \hat{h}_{01}\chi(x_i)) + (\hat{h}_{10} + \hat{h}_{11}\chi(x_i)) \sum \hat{G}_a\chi_a(x).$$

By the uniqueness of the Fourier representation, we must have

$$\sum |\hat{F}_a| \leq (|\hat{h}_{00}| + |\hat{h}_{01}|) + (|\hat{h}_{10}| + |\hat{h}_{11}|) \sum |\hat{G}_a|.$$

All of the coefficients $\hat{h}_{bc}$ must be multiples of $1/2$, and so by Parseval's identity we must have $|\hat{h}_{00}| + |\hat{h}_{01}| \leq 1$ and $|\hat{h}_{10}| + |\hat{h}_{11}| \leq 1$, from where the claim follows. □

Therefore any $\varepsilon$-biased generator has bias at most $\varepsilon(\ell + 1)$ against branching programs of width two that read $\ell$ inputs.

**Higher width?** What about branching programs of width three? It turns out that this is not the case anymore – there exist $2^{-\Omega(n)}$-biased distributions that are not (say) $1/3$ pseudorandom against branching programs of width three. This is not too difficult to prove, but there are some calculations, so let's instead give a counterexample for branching programs of width four, which is easier.

Assume $n$ is even and consider the following distribution $pr$: Choose $x_1, \ldots, x_n$ uniformly at random but conditioned on $p(x) = x_1x_2 + x_3x_4 + \cdots + x_{n-1}x_n = 0$. Clearly the function $P(x) = (-1)^{p(x)}$ (and therefore $p(x)$ itself) is not fooled by $pr$, because $\mathrm{E}[P(pr)] = 1$, while

$$\mathrm{E}[P(u)] = \mathrm{E}[(-1)^{u_1u_2}]^{n/2}] = (-1/2)^{n/2}.$$

On the other hand, we have that $pr$ is $2^{-\Omega(n)}$-biased distribution. For every linear test $a \neq 0$, we have the following two equations:

$$0 = \mathrm{E}[\chi_a(u)] = \mathrm{E}[\chi_a(u) \mid p(u) = 0] \Pr[p(u) = 0] + \mathrm{E}[\chi_a(u) \mid p(u) = 1] \Pr[p(u) = 1]$$
$$\mathrm{E}[P(u)\chi_a(u)] = \mathrm{E}[\chi_a(u) \mid p(u) = 0] \Pr[p(u) = 0] - \mathrm{E}[chi_a(u) \mid p(u) = 1] \Pr[p(u) = 1]$$

from where

$$\mathrm{E}_{pr}[\chi_a(pr)] = \mathrm{E}_u[\chi_a(u) \mid p(u) = 0] = \frac{\mathrm{E}[\chi_a(u)P(u)}{\Pr[p(u) = 0]} = \frac{\mathrm{E}[\chi_a(u)P(u)]}{1/2 + (-1/2)^{n/2}}.$$

But $|\mathrm{E}[\chi_a(u)P(u)]| = |\hat{P}_a| = 2^{-n/2}$ by a calculation we did in the last lecture.

It is an open problem to design pseudorandom distributions with support size $O(\log n)$ for branching programs of width three and higher. The answer is known in some special cases.

## 3 Low-degree polynomials

The last example shows the small-bias distributions also fail in general to be pseudorandom against degree two polynomials over $\mathbb{F}_2$. How can one obtain pseudorandom distributions for low-degree polynomials?

Let's start with degree two polynomials. It turns out that degree two polynomials (in $n$ inputs) have a very simple structure. Up to an affine change of variables (and possible negation of the output), each degree two polynomial is of the form (the operations are modulo 2):

$$p(x) = x_1x_2 + x_3x_4 + \cdots + x_{k-1}x_k \ (+ x_{k+1}).$$

The number $k$ is called the *rank* of $p$.

A simple but important fact is that $\varepsilon$-biased distributions remain $\varepsilon$-biased under affine transformations. So a polynomial of rank $k$ looks to a small-biased distribution like a *$k$-junta*, that is a function that depends on only $k$ out of its $n$ inputs. But a $k$-junta can be viewed as a function from $\{0,1\}^k$ to $[-1,1]$ and by equation (3) we get the following lemma:

**Lemma 2.** *Let $f\colon \{0,1\}^n \to [-1,1]$ be a function that only depends on $k$ of its inputs. Then*

$$|\mathrm{E}_u[f(u)] - \mathrm{E}_{pr}[f(pr)]| \le \varepsilon \cdot 2^{k/2}.$$

What if $k$ is large? By the calculation from the previous section, the bias of $f$ under the uniform distribution is at most $2^{-k/2}$. So we know that in general a small-bias distribution will not work.

One nice thing about the rank is that it is invariant under addition of linear functions: The rank of $p(x)$ equals the rank of $p(x) + \langle a, x\rangle$. So if the rank of $p$ is $k$, we can say that

$$|\hat{P}_a| = \mathrm{E}[(-1)^{p(x)+\langle a,x\rangle}] \le 2^{-k/2} \quad \text{for all } a \in \mathbb{F}^n.$$

where $P(x) = (-1)^{p(x)}$. In other words, if the rank of $p$ is large, then all its Fourier coefficients are small, or $p$ is very far from an affine function. Now recall our analysis of the degree-1 test from last lecture from which

$$\mathrm{U}_2[p] = \mathrm{Ee}_{x,u,u'}[D_{u,u'}p(x)] = \sum_{a\in\{0,1\}^n} \hat{P}_a^4 \le \max_a \hat{P}_a^2 = 2^{-k}.$$

How is this useful for designing a pseudorandom distribution? Recall last time we showed that

$$\mathrm{Ee}_u[p(u)] = \mathrm{U}_1[p] \le \sqrt{\mathrm{U}_2[p]}$$

We will describe a "pseudorandom version" $\tilde{\mathrm{U}}$ of the uniformity U such that $\tilde{\mathrm{U}}_d[f] \approx \mathrm{U}_d[f]$ for every $f$ and

$$\mathrm{Ee}_{pr}[p(pr)] = \tilde{\mathrm{U}}_1[p] \le \sqrt{\tilde{\mathrm{U}}_2[p]}. \tag{4}$$

Assuming that $\mathrm{U}_2[p]$ is sufficiently small, this will allow us to deduce that so are the quantities $\mathrm{U}_1[p]$, $\tilde{\mathrm{U}}_2[p]$, and $\tilde{\mathrm{U}}_1[p]$, and so $\mathrm{Ee}_u[p(u)]$ and $\mathrm{Ee}_{pr}[p(pr)]$ must both be close to zero, therefore close to one another.

**Low-degree testing with less randomness** Recall that the $d$-uniformity of a function is

$$\mathrm{U}_d[f] = \mathrm{Ee}_{x,a_1,\dots,a_d\sim\{0,1\}^n}[D_{a_1,\dots,a_d}f(x)].$$

Now let

$$\tilde{\mathrm{U}}_d[f] = \mathrm{Ee}_{x\sim\{0,1\}^n,e_1,\dots,e_d\sim E}[D_{e_1,\dots,e_d}f(x)]$$

where $e_1,\dots,e_d$ are independent samples some $\varepsilon$-biased distribution $E$. It turns out that $\tilde{\mathrm{U}}_d[f]$ is a good approximation to $\mathrm{U}_d[f]$ when $\varepsilon$ is small:

**Theorem 3.** *For every $f\colon \{0,1\}^n \to \{0,1\}$, $|\mathrm{U}_d[f] - \tilde{\mathrm{U}}_d[f]| \le d\varepsilon$.*

To prove this theorem, we will use the following lemma, which we will study in more detail later in the course.

**Lemma 4.** *Let $E$ be an $\varepsilon$-biased distribution and $F, G\colon \{0,1\}^n \to [-1,1]$ be any two functions. Then*

$$\left|\mathrm{E}_{x\sim\{0,1\}^n,e\sim E}[F(x)G(x+e)] - \mathrm{E}_{x,y\sim\{0,1\}^n}[F(x)G(y)]\right| \le \varepsilon.$$

*Proof.* We expand $F$ and $G$ by Fourier analysis. Notice that $\mathrm{E}_{x,y\sim\{0,1\}^n}[F(x)G(y)] = \hat{F}_0\hat{G}_0$, while

$$\mathrm{E}_{x,e}[F(x)G(x+e)] = \mathrm{E}\Big[\sum_a \hat{F}_a\chi_a(x)\sum_b \hat{G}_b\chi_b(x+e)\Big] = \hat{F}_0\hat{G}_0 + \sum_{a\neq 0}\hat{F}_a\hat{G}_a\,\mathrm{E}_e[\chi_a(e)].$$

Since $e$ is $\varepsilon$-biased, $|\mathrm{E}_e[\chi_a(e)]| \le \varepsilon$ for every nonzero $a$. By the triangle inequality, the difference between the two absolute values is at most

$$\sum_{a\neq 0}|\hat{F}_a|\cdot|\hat{G}_a|\varepsilon \le \varepsilon\sqrt{\textstyle\sum_{a\neq 0}\hat{F}_a^2}\sqrt{\textstyle\sum_{a\neq 0}\hat{G}_a^2} \le \varepsilon. \qquad \square$$

We can now prove Theorem 3 by induction on $d$. Let $F(x) = (-1)^{f(x)}$. When $d = 1$:

$$\mathrm{U}_1[f] = \mathrm{Ee}_{x,a}[D_af(x)] = \mathrm{E}_{x,y}[F(x)F(y)] \quad \text{and} \quad \tilde{\mathrm{U}}_1[f] = \mathrm{Ee}_{x,e}[D_ef(x)] = \mathrm{E}[F(x)F(x+e)]$$

and $|\mathrm{U}_1[f] - \tilde{\mathrm{U}}_1[f]| \le \varepsilon$ by Lemma 4. You can do the inductive step yourself.

**Cauchy-Schwarz-Gowers for arbitrary distributions** To obtain equation (4) we apply the following generalization of the Cauchy-Schwarz-Gowers inequality (it is proved in a similar way):

**Lemma 5.** *Let $x, y, y', a_1, \ldots, a_{d-1}$ be independent random variables in $\{0,1\}^n$. Assume $y$ and $y'$ are identically distributed. Then*

$$\mathrm{Ee}_{x,y,a_1,\ldots,a_{d-1}}[D_{a_1,\ldots,a_{d-1}}f(x+y)]^2 \le \mathrm{Ee}_{x,a_1,\ldots,a_{d-1},y,y'}[D_{a_1,\ldots,a_{d-1}}D_{y+y'}f(x+y)].$$

Iterating this inequality $d$ times, we obtain

$$\mathrm{Ee}_{x,y_1,\ldots,y_d}[f(x+y_1+\cdots+y_d)]^{2^d} \le \mathrm{Ee}_{y_1,\ldots,y_d,y'_1,\ldots,y'_d}[D_{y_1+y'_1,\ldots,y_d+y'_d}f(x+y_1+\cdots+y_d)]$$

Now notice that if $p$ is a polynomial of degree $d$, then $D_{y_1+y'_1,\ldots,y_d+y'_d}p(x+y_1+\cdots+y_d)]$ is completely independent of $x$, since taking $d$ derivatives of a degree-$d$ polynomial gives a constant. So we can fix $x = 0$ and conclude that

$$\mathrm{Ee}[p(y_1+\cdots+y_d)]^{2^d} \le \mathrm{Ee}[D_{y_1+y'_1,\ldots,y_d+y'_d}p(\text{anything})].$$

Now if we set $y_i = e_i, y'_i = e'_i$, all independent copies of some $\varepsilon$-biased distribution, then $y_1 + y'_1, \ldots, y_d + y'_d$ are also $\varepsilon$ biased (in fact $\varepsilon^2$-biased) so we get that

$$|\mathrm{Ee}[p(e_1+\cdots+e_d)]| \le \tilde{\mathrm{U}}_d[p]^{1/2^d}$$

which in particular gives equation (4).

Let's now recall what we showed for polynomials of degree two: If $p$ has small rank, then under a suitable change of basis it depends on few inputs, so to an $\varepsilon$-biased distribution $e$ is looks like a junta, and by Lemma 2 $\mathrm{Ee}_e[p(e)] \approx \mathrm{Ee}_{u\sim\{0,1\}^n}[p(u)]$. If $p$ has large rank, then it has no correlation with any linear functions so $\mathrm{U}_2[p]$ is small, and so are in turn $\mathrm{Ee}_{u\sim\{0,1\}^n}[p(u)]$, $\tilde{\mathrm{U}}_2[p]$, and finally $\mathrm{Ee}_{e_1,e_2}[p(e_1+e_2)]$, where $e_1, e_2$ are independent samples from an $\varepsilon$-biased distribution. Working out the parameters carefully we obtain that

**Theorem 6.** *Let $e_1, e_2$ be independent copies of an $\varepsilon$-biased distribution. Then for every degree 2 polynomial $p$ over $\mathbb{F}_2$,*

$$|\mathrm{E}[p(e_1+e_2)] - \mathrm{E}[p(u)]| = O(\varepsilon^{1/2}).$$

The theorem generalizes to higher degree: If $e_1, \ldots, e_d$ are independent copies of an $\varepsilon$-biased distribution, then $e_1 + \cdots + e_d$ is $O(\varepsilon^{1/2^{d-1}})$-pseudorandom for the class of degree-$d$ polynomials over $\mathbb{F}_2$.