Special Topics in Complexity Theory, Fall 2017. Instructor: Emanuele Viola

1 Lectures 4-5, Scribe: Matthew Dippel

These lectures cover some basics of small-bias distributions, and then a more recent pseudorandom generator for read-once CNF [GMR+12].

2 Small bias distributions

Definition 1.[Small bias distributions] A distribution D over $\{0,1\}^n$ has bias ϵ if no parity function can distinguish it from uniformly random strings with probability greater than ϵ. More formally, we have:

$$\forall S \subseteq [n], S \neq \emptyset, \left| \mathbb{P}_{x \in D} \left[\bigoplus_{i \in S} x_i = 1 \right] - 1/2 \right| \leq \epsilon.$$

In this definition, the $1/2$ is simply the probability of a parity test being 1 or 0 over the uniform distribution. We also note that whether we change the definition to have the probability of the parity test being 0 or 1 doesn't matter. If a test has probability $1/2 + \epsilon$ of being equal to 1, then it has probability $1 - (1/2 + \epsilon) = 1/2 - \epsilon$ of being 0, so the bias is independent of this choice.

This can be viewed as a distribution which fools tests T that are restricted to computing parity functions on a subset of bits.

Before we answer the important question of how to construct and efficiently sample from such a distribution, we will provide one interesting application of small bias sets to expander graphs.

Theorem 2.[Expander construction from a small bias set] Let D be a distribution over $\{0,1\}^n$ with bias ϵ. Define $G = (V, E)$ as the following graph:

$$V = \{0,1\}^n, E = \{(x,y) | x \oplus y \in \text{support}(D)\}.$$

Then, when we take the eigenvalues of the random walk matrix of G in descending order $\lambda_1, \lambda_2, ...\lambda_{2^n}$, we have that:

$$\max\{|\lambda_2|, |\lambda_{2^n}|\} \leq \epsilon.$$

Thus, small-bias sets yields expander graphs. Small-bias sets also turn out to be equivalent to constructing good linear codes. Although all these questions have been studied much before the definition of small-bias sets [NN90], the computational perspective has been quite useful, even in answering old questions. For example Ta-Shma used this perspective to construct better codes [Ta-17].

3 Constructions of small bias distributions

Just like our construction of bounded-wise independent distributions from the previous lecture, we will construct small-bias distributions using polynomials over finite fields.

Theorem 1.[Small bias construction] Let $\mathcal{F}$ be a finite field of size 2^ℓ, with elements represented as bit strings of length ℓ. We define the generator $G : \mathcal{F}^2 \to \{0,1\}^n$ as the following:

$$G(a,b)_i = \langle a^i, b \rangle = \sum_{j \leq \ell} (a^i)_j b_j \mod 2.$$

In this notation, a subscript of j indicates taking the jth bit of the representation. Then the output of $G(a,b)$ over uniform a and b has bias $n/2^\ell$.

Proof. Consider some parity test induced by a subset $S \subset [n]$. Then when applied to the output of G, it simplifies as:

$$\sum_{i \in S} G(a,b)_i = \sum_{i \in S} \langle a^i, b \rangle = \left\langle \sum_{i \in S} a^i, b \right\rangle.$$

Note that $\sum_{i \in S} a^i$ is the evaluation of the polynomial $P_S(x) := \sum_{i \in S} x^i$ at the point a. We note that if $P_S(a) \neq 0$, then the value of $\langle P_S(a), b \rangle$ is equally likely to be 0 or 1 over the probability of a uniformly random b. This follows from the fact that the inner product of any non-zero bit string with a uniformly random bit string is equally likely to be 0 or 1. Hence in this case, our generator has no bias.

In the case where $P_S(a) = 0$, then the inner product will always be 0, independent of the value of b. In these situations, the bias is $1/2$, but this is conditioned on the event that $P_S(a) = 0$.

We claim that this event has probability $\leq n/2^\ell$. Indeed, for non empty S, $P_S(a)$ is a polynomial of degree $\leq n$. Hence it has at most n roots. But we are selecting a from a field of size 2^ℓ. Hence the probability of picking one root is $\leq n/2^\ell$.

Hence overall the bias is at most $n/2^\ell$. □

To make use of the generator, we need to pick a specific ℓ. Note that the seed length will be $|a| + |b| = 2\ell$. If we want to achieve bias ϵ, then we must have $\ell = \log\left(\frac{n}{\epsilon}\right)$. Al the logarithms in this lecture are in base 2. This gives us a seed length of $2\log\left(\frac{n}{\epsilon}\right)$.

Small-bias are so important that a lot of attention has been devote to optimizing the constant "2" above. A lower bound of $\log n + (2 - o(1))\log(1/\epsilon)$ on the seed length was known. Ta-Shma recently [Ta-17] gave a nearly matching construction with seed length $\log n + (2 + o(1))\log(1/\epsilon)$.

We next give a sense of how to obtain different tradeoffs between n and ϵ in the seed length. We specifically focus on getting a nearly optimal dependence on n, because the construction is a simple, interesting "derandomization" of the above one.

3.1 An improved small bias distribution via bootstrapping

We will show another construction of small bias distributions that achieves seed length $(1 + o(1))\log n + O(\log(1/\epsilon))$. It will make use of the previous construction and proof.

The intuition is the following: the only time we used that b was uniform was in asserting that if $P_S(a) \neq 0$, then $\langle P_S(a), b\rangle$ is uniform. But we don't need b to be uniform for that. What do we need from b? We need that it has small-bias!

Our new generator is $G(a, G'(a', b'))$ where G and G' are as before but with different parameters. For G, we pick a of length $\ell = \log n/\epsilon$, whereas G' just needs to be an ϵ-biased generator on ℓ bits, which can be done as we just saw with $O(\log \ell/\epsilon)$ bits. This gives a seed length of $\log n + \log\log n + O(\log 1/\epsilon)$, as promised.

We can of course repeat the argument but the returns diminish.

4 Connecting small bias to k-wise independence

We will show that using our small bias generators, we can create distributions which are almost k-wise independent. That is, they are very close to a k-wise independent distribution in statistical distance, while having a substantially shorter seed length than what is required for k-wise independence. In particular, we will show two results:

- Small bias distributions are themselves close to k-wise independent.
- We can improve the parameters of the above by feeding a small bias distribution to the generator for k-wise independence from the previous lectures. This will improve the seed length of simply using a small bias distribution.

Before we can show these, we'll have to take a quick aside into some fundamental theorems of Fourier analysis of boolean functions.

4.1 Fourier analysis of boolean functions 101

Let $f : \{-1,1\}^n \to \{-1,1\}$. Here the switch between $\{0,1\}$ and $\{-1,1\}$ is common, but you can think of them as being isomorphic. One way to think of f is as being a vector in $\{-1,1\}^{2^n}$. The xth entry of f indicates the value of $f(x)$. If we let $\mathbf{1_S}$ be the indicator function returning 1 iff $x = S$, but once again written as a vector like f is, then any function f can be written over the basis of the $\mathbf{1_S}$ vectors, as:

$$f = \sum_S f(s)\mathbf{1_S}.$$

This is the "standard" basis.

Fourier analysis simply is a different basis in which to write functions, which is sometimes more useful. The basis functions are $\chi_S(x) : \{-1,1\}^n \to \{-1,1\} = \prod_{i \in S} x_i$. Then any boolean function f can be expressed as:

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S)\chi_S(x),$$

where the $\hat{f}(S)$, called the "Fourier coefficients," can be derived as:

$$\hat{f}(S) = \mathbb{E}_{x\ U_n}\left[f(x)\chi_S(x)\right],$$

where the expectation is over uniformly random x.

Claim 1. For any function f with range $\{-1, 1\}$, its Fourier coefficients satisfy:

$$\sum_{S\subseteq[n]} \hat{f}(S)^2 = 1.$$

Proof. We know that $\mathbb{E}[f(x)^2] = 1$, as squaring the function makes it 1. We can re-express this expectation as:

$$\mathbb{E}[f(x)f(x)] = \mathbb{E}\left[\sum_S \hat{f}(s)\chi_S(x) \cdot \sum_T \hat{f}(T)\chi_T(x)\right] = \mathbb{E}\left[\sum_{S,T} \hat{f}(s)\chi_S(x)\hat{f}(T)\chi_T(x)\right].$$

We make use of the following fact: if $S \neq T$, then $\mathbb{E}[\chi_S(x)\chi_T(x)] = \mathbb{E}[\chi_{S\oplus T}(x)] = 0$. If they equal each other, then their difference is the empty set and this function is 1.

Overall, this implies that the above expectation can be simply rewritten as:

$$\sum_{S=T} \hat{f}(S)\hat{f}(T) = \sum_S \hat{f}(S)^2.$$

Since we already decided that the expectation is 1, the claim follows. □

5 Small bias distributions are close to k-wise independent

Before we can prove our claim, we formally introduce what we mean for two distributions to be close. We use the most common definition of statistical difference, which we repeat here:

Definition 1. Let D_1, D_2 be two distributions over the same domain H. Then we denote their statistical distance $\mathrm{SD}(D_1, D_2)$, and sometimes written as $\Delta(D_1, D_2)$, as

$$\Delta(D_1, D_2) = \max_{T\subseteq H} \left|\mathcal{P}[D_1 \in T] - \mathcal{P}[D_2 \in T]\right|.$$

Note that the probabilities are with respect to the individual distributions D_1 and D_2. We may also say that D_1 is ϵ-close to D_2 if $\Delta(D_1, D_2) \leq \epsilon$.

We can now show our result, which is known as **Vazirani's XOR Lemma**:

Theorem 2. If a distribution D over $\{0,1\}^n$ has bias ϵ, then D is $\epsilon 2^{n/2}$ close to the uniform distribution.

Proof. Let T be a test. To fit the above notation, we can think of T as being defined as the set of inputs for which $T(x) = 1$. Then we want to bound:

$$|\mathbb{E}[T(D)] - \mathbb{E}[T(U)]|.$$

Expanding T in Fourier basis we rewrite this as

$$|\mathbb{E}[\sum_S \hat{T_S}\chi_S(D)] - \mathbb{E}[\sum_S \hat{T_S}\chi_S(U)]| = |\sum_S \hat{T_S}\left(\mathbb{E}[\chi_S(D)] - \mathbb{E}[\chi_S(U)]\right)|.$$

We know that $\mathbb{E}_U[\chi_S(x)] = 0$ for all non empty S, and 1 when S is the empty set. We also know that $\mathbb{E}_D[\chi_S(x)] \leq \epsilon$ for all non empty S, and is 1 when S is the empty set. So the above can be bounded as:

$$\leq \sum_{S \neq \emptyset} |\hat{T_S}||\mathbb{E}_D[\chi_S(x)] - \mathbb{E}_U[\chi_S(x)]| \leq \sum_S |\hat{T_S}|\epsilon = \epsilon \sum_S |\hat{T_S}|.$$

Lemma 3. $\sum_S |\hat{T_S}| \leq 2^{n/2}$

Proof. By Cauchy Schwartz:

$$\sum |\hat{T_S}| \leq 2^{n/2}\sqrt{\sum \hat{T_S}^2} \leq 2^{n/2}$$

Where the last simplification follows from Claim 1. □

Using the above lemma completes the upper bound and the proof of the theorem. □

Corollary 4. Any k bits of an ϵ biased distribution are $\epsilon 2^{k/2}$ close to uniform.

Using the corollary above, we see that we can get ϵ close to a k-wise independent distribution (in the sense of the corollary) by taking a small bias distribution with $\epsilon' = \epsilon/2^{k/2}$. This requires seed length $\ell = O(\log(n/\epsilon') = O(\log(2^{k/2}n/\epsilon) = O(\log(n) + k + \log(1/\epsilon))$. Recall that for exact k-wise we required seed length $k \log n$.

5.1 An improved construction

Theorem 5. Let $G : \{0,1\}^{k \log n} \to \{0,1\}^n$ be the generator previously described that samples a k-wise independent distribution (or any linear G). If we replace the input to G with a small bias distribution of $\epsilon' = \epsilon/2^k$, then the output of G is ϵ-close to being k-wise independent.

Proof. Consider any parity test S on k bits on the output of G. It can be shown that G is a linear map, that is, G simply takes its seed and it multiplies it by a matrix over the field GF(2) with two elements. Hence, S corresponds to a test S' on the input of G, on possibly many bits. The test S' is not empty because G is k-wise independent. Since we fool S' with error ϵ', we also fool S with error ϵ, and the theorem follows by Vazirani's XOR lemma. □

Using the seed lengths we saw we get the following.

Corollary 6. There is a generator for almost k-wise independent distributions with seed length $O(\log \log n + \log(1/\epsilon) + k)$.

6 Tribes Functions and the GMRTV Generator

We now move to a more recent result. Consider the Tribes function, which is a read-once CNF on $k \cdot w$ bits, given by the And of k terms, each on w bits. You should think of $n = k \cdot w$ where $w \approx \log n$ and $k \approx n/\log n$.

We'd like a generator for this class with seed length $O(\log n/\epsilon)$. This is still open! (This is just a single function, for which a generator is trivial, but one can make this challenge precise for example by asking to fool the Tribes function for any possible negation of the input variables. These are 2^n tests and a generator with seed length $O(\log n/\epsilon)$ is unknown.)

The result we saw earlier about fooling And gives a generator with seed length $O(\log n)$, however the dependence on ϵ is poor. Achieving a good dependence on ϵ has proved to be a challenge. We now describe a recent generator [GMR$^+$12] which gives seed length $O(\log n/\epsilon)(\log \log n)^{O(1)}$. This is incomparable with the previous $O(\log n)$, and in particular the dependence on n is always suboptimal. However, when $\epsilon = 1/n$ the generator [GMR$^+$12]

gives seed length $O(\log n) \log\log n$ which is better than previously available constructions.

The high-level technique for doing this is based on iteratively restricting variables, and goes back about 30 years [AW89]. This technique seems to have been abandoned for a while, possibly due to the spectacular successes of Nisan [Nis91, Nis92]. It was revived in [GMR$^+$12] (see also [GLS12]) with an emphasis on a good dependence on ϵ.

A main tool is this claim, showing that small-bias distributions fool products of functions with small variance. Critically, we work with non-boolean functions (which later will be certain averages of boolean functions).

Claim 1. Let $f_1, f_2, ..., f_k : \{0,1\}^w \to [0,1]$ be a series of boolean functions. Further, let $D = (v_1, v_2, ..., v_k)$ be an ϵ-biased distribution over wk bits, where each v_i is w bits long. Then

$$\mathbb{E}_D[\prod_i f_i(v_i)] - \prod_i \mathbb{E}_U[f_i(U)] \leq \left(\sum_i \mathrm{var}(f_i)\right)^d + (k2^w)^d\epsilon,$$

where $\mathrm{var}(f) := \mathbb{E}[f^2] - \mathbb{E}^2[f]$ is variance of f with respect to the uniform distribution.

This claim has emerged from a series of works, and this statement is from a work in progress with Chin Ho Lee. For intuition, note that constant functions have variance 0, in which case the claim gives good bounds (and indeed any distribution fools constant functions). By contrast, for balanced functions the variance is constant, and the sum of the variances is about k, and the claim gives nothing. Indeed, you can write Inner Product as a product of nearly balanced functions, and it is known that small-bias does not fool it. For this claim to kick in, we need each variance to be at most $1/k$.

In the tribes function, the And fucntions have variance 2^{-w}, and the sum of the variances is about 1 and the claim gives nothing. However, if you perturb the Ands with a little noise, the variance drops polynomially, and the claim is useful.

Claim 2. Let f be the AND function on w bits. Rewrite it as $f(x,y)$, where $|x| = |y| = w/2$. That is, we partition the input into two sets. Define $g(x)$ as:

$$g(x) = \mathbb{E}_y[f(x,y)],$$

where y is uniform. Then $\text{var}(g) = \Theta(2^{-3w/2})$.

Proof.

$$\text{var}(g) = \mathbb{E}[g(x)^2] - (\mathbb{E}[g(x)])^2 = \mathbb{E}_x[\mathbb{E}_y[f(x,y)]^2] - (\mathbb{E}_x[\mathbb{E}_y[f(x,y)]])^2 .$$

We know that $(\mathbb{E}_x[\mathbb{E}_y[f(x,y)]])$ is simply the expected value of f, and since f is the AND function, this is 2^{-w}, so the right term is 2^{-2w}.

We reexpress the left term as $\mathbb{E}_{x,y,y'}[f(x,y)f(x,y')]$. But we note that this product is 1 iff $x = y = y' = \mathbf{1}$. The probability of this happening is $(2^{-w/2})^3 = 2^{-3w/2}$.

Thus the final difference is $2^{-3w/2}(1 - 2^{-w/2}) = \Theta(2^{-3w/2})$. □

We'll actually apply this claim to the Or function, which has the same variance as And by De Morgan's laws.

We now present the main inductive step to fool tribes.

Claim 3. Let f be the tribes function, where the first $t \leq w$ bits of each of the terms are fixed. Let $w' = w - t$ be the free bits per term, and $k' \leq k$ the number of terms that are non-constant (some term may have become 0 after fixing the bits).

Reexpress f as $f(x,y) = \bigwedge_{k'} (\bigvee(x_i, y_i))$, where each term's input bits are split in half, so $|x_i| = |y_i| = w'/2$.

Let D be a small bias distribution with bias ϵ^c (for a big enough c to be set later). Then

$$\left|\mathbb{E}_{(x,y)\in U^2}[f(x,y)] - \mathbb{E}_{(x,y)\in(D,U)}[f(x,y)]\right| \leq \epsilon.$$

That is, if we replace half of the free bits with a small bias distribution, then the resulting expectation of the function only changes by a small amount.

To get the generator from this claim, we repeatedly apply Claim 3, replacing half of the bits of the input with another small bias distribution. We repeat this until we have a small enough remaining amount of free bits that replacing all of them with a small bias distribution causes an insignificant change in the expectation of the output.

At each step, w is cut in half, so the required number of repetitions to reduce w' to constant is $R = \log(w) = \log\log(n)$. Actually, as explained

below, we'll stop when $w = c' \log\log 1/\epsilon$ for a suitable constant c' (this arises from the error bound in the claim above, and we).

After each replacement, we incur an error of ϵ, and then we incur the final error from replacing all bits with a small bias distribution. This final error is negligible by a result which we haven't seen, but which is close in spirit to the proof we saw that bounded independence fools AND.

The total accumulated error is then $\epsilon' = \epsilon \log\log(n)$. If we wish to achieve a specific error ϵ, we can run each small bias generator with $\epsilon/\log\log(n)$.

At each iteration, our small bias distribution requires $O(\log(n/\epsilon))$ bits, so our final seed length is $O(\log(n/\epsilon))\text{poly}\log\log(n)$.

Proof of Claim 3. Define $g_i(x) = \mathbb{E}_y[\bigvee_i(x_i, y_i)]$, and rewrite our target expression as:

$$\mathbb{E}_{x\in U}\left[\prod g_i(x_i)\right] - \mathbb{E}_{x\in D}\left[\prod g_i(x_i)\right].$$

This is in the form of Claim 1. We also note that from Claim 2 that $\text{var}(g_i) = 2^{-3w'/2}$.

We further assume that $k' \leq 2^{w'}\log(1/\epsilon)$. For if this is not true, then the expectation over the first $2^{w'}\log(1/\epsilon)$ terms is $\leq \epsilon$, because of the calculation

$$(1 - 2^{-w'})^{2^{w'}\log(1/\epsilon)} \leq \epsilon.$$

Then we can reason as in the proof that bounded independence fools AND (i.e., we can run the argument just on the first $2^{w'}\log(1/\epsilon)$ terms to show that the products are close, and then use the fact that it is small under uniform, and the fact that adding terms only decreases the probability under any distribution).

Under the assumption, we can bound the sum of the variances of g as:

$$\sum \text{var}(g_i) \leq k' 2^{-3w'/2} \leq 2^{-\Omega(w')}\log(1/\epsilon).$$

If we assume that $w' \geq c\log\log(1/\epsilon)$ then this sum is $\leq 2^{-\Omega(w')}$.

We can then plug this into the bound from Claim 1 to get

$$(2^{-\Omega(w')})^d + (k2^{w'})^d\epsilon^c = 2^{-\Omega(dw')} + 2^{O(dw')}\epsilon^c.$$

Now we set d so that $\Omega(dw') = \log(1/\epsilon) + 1$, and the bound becomes:

$$\epsilon/2 + (1/\epsilon)^{O(1)}\epsilon^c \leq \epsilon.$$

By making c large enough the claim is proved. □

In the original paper, they apply these ideas to read-once CNF formulas. Interestingly, this extension is more complicated and uses additional ideas. Roughly, the progress measure is going to be number of terms in the CNF (as opposed to the width). A CNF is broken up into a small number of Tribes functions, the above argument is applied to each Tribe, and then they are put together using a general fact that they prove, that if f and g are fooled by small-bias then also $f \wedge g$ on disjoint inputs is fooled by small-bias.

References

[AW89] Miklos Ajtai and Avi Wigderson. Deterministic simulation of probabilistic constant-depth circuits. *Advances in Computing Research - Randomness and Computation*, 5:199–223, 1989.

[GLS12] Dmitry Gavinsky, Shachar Lovett, and Srikanth Srinivasan. Pseudorandom generators for read-once acc^0. In *IEEE Conf. on Computational Complexity (CCC)*, pages 287–297, 2012.

[GMR+12] Parikshit Gopalan, Raghu Meka, Omer Reingold, Luca Trevisan, and Salil Vadhan. Better pseudorandom generators from milder pseudorandom restrictions. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, 2012.

[Nis91] Noam Nisan. Pseudorandom bits for constant depth circuits. *Combinatorica. An Journal on Combinatorics and the Theory of Computing*, 11(1):63–70, 1991.

[Nis92] Noam Nisan. Pseudorandom generators for space-bounded computation. *Combinatorica*, 12(4):449–461, 1992.

[NN90] J. Naor and M. Naor. Small-bias probability spaces: efficient constructions and applications. In *22nd ACM Symp. on the Theory of Computing (STOC)*, pages 213–223. ACM, 1990.

[Ta-17] Amnon Ta-Shma. Explicit, almost optimal, epsilon-balanced codes. In *ACM Symp. on the Theory of Computing (STOC)*, pages 238–251, 2017.