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

1 Lecture 1, Scribe: Chin Ho Lee

In this first lecture we begin with some background on pseudorandomness and then we move on to the study of bounded independence, presenting in particular constructions and lower bounds.

1.1 Background

Let us first give some background on randomness. There are 3 different theories:

(1) Classical probability theory. For example, if we toss a coin 12 times then the probability of each outcome is the same, i.e., $\Pr[010101010101] = \Pr[011011100011]$. However, intuitively we feel that the first outcome is less random than the second.

(2) Kolmogorov complexity. Here the randomness is measured by the length of the shortest program outputting a string. In the previous example, the program for the second outcome could be "print `011011100011`", whereas the program for the first outcome can be "print `01` six times", which is shorter than the first program.

(3) Pseudorandomness. This is similar to resource-bounded Kolmogorov complexity. Here random means the distribution "looks random" to "efficient observers."

Let us now make the above intuition precise.

Definition 1.[Pseudorandom generator (PRG)] A function $f\colon \{0,1\}^s \to \{0,1\}^n$ is a *pseudorandom generator (PRG)* against a class of tests $T \subseteq \{t\colon \{0,1\}^n \to \{0,1\}\}$ with error ϵ, if it satisfies the following 3 conditions:

(1) the output of the generator must be longer than its input, i.e., $n > s$;

(2) it should *fool* T, that is, for every test $t \in T$, we have $\Pr[t(U_n) = 1] = \Pr[t(f(U_s)) = 1] \pm \epsilon$;

(3) the generator must be efficient.

To get a sense of the definition, note that a PRG is easy to obtain if we drop any one of the above 3 conditions. Dropping condition (1), then we can define our PRG as $f(x) := x$. Dropping condition (2), then we can define our

PRG as $f(x) := 0$. Dropping condition (3), then the PRG is not as obvious to obtain as the previous two cases. We have the following claim.

Claim 2. For every class of tests T, there exists an inefficient PRG with error ϵ and seed length $s = \lg_2 \lg_2(|T|) + 2\lg_2(1/\epsilon) + O(1)$.

Before proving the claim, consider the example where T is the class of circuits of size n^{100} over n-bit input, it is known that $|T| = 2^{n^{O(1)}}$. Hence, applying our claim above we see that there is an inefficient PRG that fools T with error ϵ and seed length $s = O(\lg_2(n/\epsilon))$.

We now prove the claim using the probabilistic method.

Proof. Consider picking f at random. Then by the Chernoff bound, we have for every test $t \in T$,

$$\Pr_f[|\Pr_{U_s}[t(f(U_s)) = 1] - \Pr_{U_n}[t(U_n) = 1]| \geq \epsilon] \leq 2^{-\Omega(\epsilon^2 2^s)} < 1/|T|,$$

if $s = \lg_2 \lg_2(|T|) + 2\lg_2(1/\epsilon) + O(1)$. Therefore, by a union bound over $t \in T$, there exists a fixed f such that for every $t \in T$, the probabilities are within ϵ. □

1.2 k-wise independent distribution

A major goal in research in pseudorandomness is to construct PRGs for (1) richer and richer class T, (2) smaller and smaller seed length s, and making the PRG explicit. For starters, let us consider a simple class of tests.

Definition 3.[d-local tests] The *d-local tests* are tests that depend only on d bits.

We will show that for this class of tests we can actually achieve error $\epsilon = 0$. To warm up, consider what happens when $d = 1$, then we can have a PRG with seed length $s = 1$ by defining $f(0) := 0^n$ and $f(1) := 1^n$.

For $d = 2$, we have the following construction. Define

$$f(x)_y := \langle x, y \rangle = \sum_i x_i y_i \bmod 2.$$

Here the length of x and y is $|x| = |y| = \lg_2 n$, and we exclude $y = 0^{\lg_2 n}$. Note that the output has $n - 1$ bits, but we can append one uniform bit to the output of f. So the seed length would be $\lg_2 n + 1$.

Now we prove the correctness of this PRG.

Claim 4. The f defined above is a PRG against 2-local tests with error $\epsilon = 0$.

Proof. We need to show that for every $y \neq z$, the random variable $(f(x)_y, f(x)_z)$ over the choice of x is identical to U_2, the uniform 2-bit string. Since $y \neq z$, suppose without loss of generality that there exists an i such that $y_i = 1$ and $z_i = 0$. Now $f(x)_z$ is uniform, and conditioned on z, $f(x)_y$ is also uniform, thanks to the index y_i. □

The case for $d = 3$ becomes much more complicated and involves the use of *finite fields*. One can think of a finite field as a finite domain that behaves like $\mathbb{Q}$ in the sense that it allows you to perform arithmetic operations, including division, on the elements. We will use the following fact about finite fields.

Lemma 5. There exist finite fields of size p^k, for every prime p and integer k. Moreover, they can be constructed and operated with in time $\text{poly}(k, p)$.

Remark 6. Ideally one would like the dependence on p to be $\lg_2 p$. However, such construction remains an open question and there have been many attempts to constructing finite fields in time $\text{poly}(k, \lg_2 p)$. Here we only work with finite fields with $p = 2$, and there are a lot of explicit constructions for that.

One simple example of finite fields are integers modulo p.

Theorem 7. Let $D = \{0, 1\}^{\lg_2 n}$. For every k, there exists an explicit construction over D^n such that

(1) elements in D^n can be sampled with $s = k \lg_2 n$ bits, and

(2) every k symbols are uniform in D^k.

For $d = 3$, we can use the above theorem with $k = 3$, and the PRG can output the first bit of every symbol.

Remark 8. There exist other constructions that are similar to the inner product construction for the case $d = 2$, with y carefully chosen, but the way to choose y involves the use of finite fields as well.

Note that we can also apply the theorem for larger d to fool d-local tests with seed length $s = d \lg_2 n$.

We now prove the theorem.

Proof. Pick a finite field $\mathbb{F}$ of size $2^{\lg_2 n}$. Let $a_0, \ldots, a_{n-1} \in \mathbb{F}$ be uniform random elements in $\mathbb{F}$ which we think of as a polynomial $a(x)$ of degree $k-1$. We define the generator f to be

$$f(a_0, \ldots, a_{n-1})_x = a(x) = \sum_{i=0}^{n-1} a_i x^i.$$

(One should think of the outputs of f as lines and curves in the real plane.)

The analysis of the PRG follows from the following useful fact: For every k points $(x_0, y_0), (x_1, y_1), \ldots, (x_{k-1}, y_{k-1})$, there exists exactly one degree $k-1$ polynomial going through them. □

Let us now introduce a terminology for PRGs that fool d-local tests.

Definition 9. We call distributions that look uniform (with error 0) to k-local tests *k-wise independent* (also known as k-wise uniform). The latter terminology is more precise, but the former is more widespread.

We will soon see an example of a distribution where every k elements are independent but not necessarily uniform.

1.3 Lower bounds

We have just seen a construction of k-wise independent distributions with seed length $s = d \lg_2 n$. It is natural to ask, what is the minimum seed length of generating k-wise independent distributions?

Claim 10. For every $k \geq 2$, every PRG for k-local tests over $\{0,1\}^n$ has seed length $s \geq \Omega(k \lg_2(n/k))$.

Proof. We use the linear-algebraic method. See the book by Babai–Frankl [1] for more applications of this method.

To begin, we will switch from $\{0,1\}$ to $\{-1,1\}$, and write the PRG as a $2^s \times n$ matrix M, where the rows are all the possible outputs of the PRG. Since the PRG fools k-local tests and $k \geq 2$, one can verify that every 2 columns of M are orthogonal, i.e., $\langle M_i, M_j \rangle = 0$ for $i \neq j$. As shown below, this implies that the vectors are independent. And by linear algebra this gives a lower bound on s.

However so far we have not used k. Here's how to use it. Consider all the column vectors v obtained by taking the entry-wise products of any of the $k/2$ vectors in M. Because of k-wise independence, these v's are again orthogonal, and this also implies that they are linearly independent.

Claim 11. If $v_1, v_2, \ldots, v_t$ are orthogonal, then they are linearly independent.

Proof. Suppose they are not and we can write $v_i = \sum_{j \in S, i \notin S} v_j$ for some S. Taking inner product with v_i on both sides, we have that the L.H.S. is nonzero, whereas the R.H.S. is zero because the vectors are orthogonal, a contradiction. □

Therefore, the rank of M must be at least the number of v's, and so

$$2^s \geq \text{number of v's} \geq \binom{n}{k/2} \geq (2n/k)^{k/2}.$$

Rearranging gives $s \geq (k/2) \lg_2(2n/k)$. □

1.4 Who is fooled by k-wise independence?

In the coming lectures we will see that k-wise independence fools AC^0, the class of constant-depth circuits with unbounded fan-in. Today, let us see what else is fooled by k-independence in addition to k-local tests.

(1) Suppose we have n independent variables $x_1, \ldots, X_n \in [0, 1]$ and we want to understand the behavior of their sum $\sum_i X_i$. Then we can apply tools such as the Chernoff bound, tail bounds, Central Limit Theorem, and the Berry–Esseen theorem. The first two give bounds on large deviation from the mean. The latter two are somewhat more precise facts that show that the sum will approach a normal distribution (i.e., the probability of being larger than t for any t is about the same). One can show that similar results hold when the X_i's are k-wise independent. The upshot is that the Chernoff bound gives error $2^{-\text{samples}}$, while under k-wise independence we can only get an error $(\text{samples})^{-k/2}$.

(2) We will see next time that k-wise independence fools DNF and AC^0.

(3) k-wise independence is also used as hashing in load-balancing.

1.4.1 k-wise independence fools AND

We now show that k-wise independent distributions fool the AND function.

Claim 12. Every k-wise uniform distribution fools the AND functions on bits with error $\epsilon = 2^{-\Omega(k)}$.

Proof. If the AND function is on at most k bits, then by definition the error is $\epsilon = 0$. Otherwise the AND is over more than k bits. Without loss of generality we can assume the AND is on the first $t > k$ bits. Observe that for any distribution D, we have

$$\Pr_D[\text{AND on t bits is 1}] \leq \Pr_D[\text{AND on k bits is 1}].$$

The right-hand-side is the same under uniform and k-wise uniformity, and is 2^{-k}. Hence,

$$|\Pr_{\text{uniform}}[\text{AND} = 1] - \Pr_{\text{k-wise ind.}}[\text{AND} = 1]| \leq 2^{-k}. \qquad \square$$

Instead of working over bits, let us now consider what happens over a general domain D. Given n functions $f_1, \ldots, f_n \colon D \to \{0,1\}$. Suppose $x_1, \ldots, x_n$ are k-wise uniform over D^n. What can you say about the AND of the outputs of the f_i's, $f_1(x_1), f_2(x_2), \ldots, f_n(x_n)$?

This is similar to the previous example, except now that the variables are independent but not necessarily uniform. Nevertheless, we can show that a similar bound of $2^{-\Omega(k)}$ still holds.

Theorem 13.[[2]] Let $X_1, X_2, \ldots, X_n$ be random variables over $\{0,1\}$, which are k-wise independent, but not necessarily uniform. Then

$$\Pr[\prod_{i=1}^{n} X_i = 1] = \prod_{i=1}^{n} \Pr[X_i = 1] \pm 2^{-\Omega(k)}.$$

This fundamental theorem appeared in the conference version of [2], but was removed in the journal version. One of a few cases where the journal version contains *less* results than the conference version.

Proof. Since each X_i is in $\{0,1\}$, by De Morgan's law, we can write

$$\Pr[\prod_{i=1}^{n} X_i = 1] = \mathbb{E}[\prod_{i=1}^{n} X_i] = \mathbb{E}[\text{AND}_{i=1}^{n} X_i] = \mathbb{E}[1 - \text{OR}_{i=1}^{n}(1 - X_i)].$$

If we define the event E_i to be $(1 - X_i)$, then $\mathrm{OR}_{i=1}^{n}(1 - X_i)$ is the same as $\Pr[\bigcup_{I=1}^{n} E_i]$. Now we apply the inclusion-exclusion principle, which says

$$\Pr[\bigcup_{i=1}^{n} E_i] = \sum_{i=1}^{n} \Pr[E_i] - \sum_{i \neq j} \Pr[E_i \cap E_j] + \cdots + (-1)^{J+1} \sum_{S \subseteq [n], |S|=J} \Pr[\bigcap_{i \in S} E_i] + \cdots .$$

we will finish the proof in the next lecture. □

References

[1] László Babai and Peter Frankl. *Linear algebra methods in combinatorics.* 1992.

[2] Guy Even, Oded Goldreich, Michael Luby, Noam Nisan, and Boban Velickovic. Approximations of general independent distributions. In *ACM Symp. on the Theory of Computing (STOC)*, pages 10–16, 1992.