CS761 Derandomization and Pseudorandomness **2022-23 Sem I**

Lecture 4: 18-08-22

Scribe: Punna Hitesh Kumar *Lecturer: Rohit Gurjar*

1 A lower bound for pairwise independence

We have seen that we can generate n pairwise independent random bits by using $O(\log n)$ independent random bits. We will now argue that it is optimal.

Claim 4.1. *Suppose there is a (deterministic) generator $Gen : \{0,1\}^k \to \{0,1\}^n$ that takes k truly random bits as input and generates n pairwise random bits then we must have $2^k \geq n$ i.e., $k > \log n$.*

Proof. Consider a $2^k \times n$ matrix G defined as follows: for a k-bit string s, the corresponding row of the matrix G is $Gen(s)$. In other words, the output of the algorithm Gen comes uniformly randomly among the rows of G. We know that the ith output bit of Gen is 0 or 1 with probability 1/2. This means

1) For any $1 \leq i \leq n$, the ith column of G should have $2^k/2$ of its entries as 1 and remaining as 0.

Pairwise independence tells us that for any $i \neq j$,

$$\Pr[i\text{th output bit} = 0, j\text{th output bit} = 0] = 1/4.$$

Hence,

2) when we put ith and jth column together, we should see:

- 00 occurs $2^k/4$ times
- 01 occurs $2^k/4$ times
- 10 occurs $2^k/4$ times
- 11 occurs $2^k/4$ times.

Now using this pattern, we will argue that the rank of matrix G is at least n, but we also know that rank of matrix G can be at most the number rows in it, i.e, 2^k. This will prove $2^k \geq n$, as desired.

Approach 1 The key Idea is to Replace 1, 0 with 1, -1. Then for any $i \neq j$, the inner Product of the ith and jth column will be equal to

$$2^k/4(1 \times 1) + 2^k/4(1 \times -1) + 2^k/4(-1 \times -1) + 2^k/4(-1 \times 1) = 0.$$

This Implies that any pair of columns in G are orthogonal, i.e., we have n orthogonal vectors. That means, these n vectors must be linearly independent. Hence, rank$(G) \geq n$.

Approach 2 (suggested by Amit Rajaraman) Consider the matrix product G^TG. Observe that the ith diagonal entry of G^TG will be simply the the number of 1s in the ith column of G, which is 2^{k-1}. And for any $i \neq j$, the (i,j) entry of G^TG will be the number of common 1s in the ith and jth columns, which is 2^{k-2}. Hence,

$$G^TG = 2^{k-2}(I_n + J_n)$$

where I_n is the identity matrix and J_n is the all 1s matrix of size $n \times n$. One can verify that this matrix has rank n. This implies that G must have rank at least n. □

Now, we will see a construction for pairwise independence different from the previous lecture. This construction will naturally generalize to k-wise independence, for any k.

2 An algebraic construction of pairwise independence

Let $\mathbb{F}$ is a field of size q. Choose two elements $a, b \in \mathbb{F}$, uniformly randomly. Then consider the set

$$\{ax + b : x \in F\}$$

We claim that these q random variables are pairwise independent.

Example. Let $\mathbb{F}$ be of size 3. That is, the elements $\{0, 1, 2\}$ with $\pmod 3$ addition and multiplication. We choose $a, b \in \mathbb{F}$ randomly. $b, a + b, 2a + b$ will be the three different values generated. The below table gives the values of these three random variables for various choices of a and b

a	b	b	a+b	2a+b
0	0	0	0	0
0	1	1	1	1
0	2	2	2	2
1	0	0	1	2
1	1	1	2	0
1	2	2	0	1
2	0	0	2	1
2	1	1	0	2
2	2	2	1	0

Now, observe from the table that if, for example, we fix $a + b$ to 1 then $2a + b$ is equally likely to be $0, 1$ or 2. Similarly, if we fix $2a + b$ to 0 then b is equally likely to be $0, 1$ or 2. One can verify the pairwise independence from the table. Note that the three variables are not completely independent. If you fix two of them, say for example, b and $a + b$ then the third variable $2a + b$ gets fixed.

Proof of pairwise independence. First let us show that each variable is uniformly distributed. That is, for any $x \in \mathbb{F}, \alpha \in \mathbb{F}$ we have

$$\Pr_{a,b}[ax + b = \alpha] = 1/q$$

Here the probability is over the choice of a and b from $\mathbb{F}$ and $q = |\mathbb{F}|$. To see the above probability observe that for any fixing of a, there is a unique b that gives $ax + b = \alpha$. So, out of q^2 choices of (a, b), exactly q of them give $ax + b = \alpha$. Hence, the $1/q$ probability.

Now, we will argue pairwise independence. For any $x \neq y \in F$, and $\alpha, \beta \in \mathbb{F}$, we want to show

$$\Pr_{a,b}[ax + b = \alpha \text{ and } ay + b = \beta] = 1/q^2$$

Que. How many tuples (a, b) will satisfy $ax + b = \alpha$ and $ay + b = \beta$?

For any given $x \neq y$ and α, β, there will be a unique solution for (a, b). This is given by

$$a = (\alpha - \beta)(x - y)^{-1} \text{ and } b = (\beta x - \alpha y)(x - y)^{-1}.$$

Thus, the Probability will be $1/q^2$

□

Comparison with construction 1 from the previous lecture. Here, we generated q many random variables $R_1, R_2, \ldots, R_q$ taking values from $\mathbb{F}$. Thus, each of these random variables can be viewed as $\log q$ random bits. If we want to generate these using construction 1 (xor of each possible subset), then we should generate $\log q$ independent instances of q pairwise independent random bits. The ith bits of $R_1, R_2, \ldots, R_q$ will come from the ith instance of q pairwise independent random bits. Recall that for generating q pairwise independent random bits we need $O(\log q)$ random bits. Hence, we would need in total $O(\log^2 q)$ random bits.

On the other hand, in the algebraic construction above, we chose a, b randomly. That is, we needed only $2 \log q$ random bits.

3 Generalization to k-wise independence

Again, let $\mathbb{F}$ be a field of size q. We choose $a_0, a_1, , a_{k-1} \in \mathbb{F}$, uniformly randomly. Then consider the set

$$\{a_0 + a_1 x + a_2 x^2 + \cdots + a_{k-1} x^{k-1} : x \in \mathbb{F}\}$$

We claim that these q random variables are k-wise independent.

Proof. We want to show that for any distinct $x_1, x_2, \ldots, x_k \in \mathbb{F}$, and $\alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{F}$, we have

$$\begin{aligned} \Pr_{a_0, a_1, , a_{k-1}} [\quad & a_0 + a_1 x_1 + a_2 x_1^2 + \cdots + a_{k-1} x_1^{k-1} && = \alpha_1 \\ & \text{and } a_0 + a_1 x_2 + a_2 x_2^2 + \cdots + a_{k-1} x_2^{k-1} && = \alpha_2 \\ & \vdots \\ & \text{and } a_0 + a_1 x_k + a_2 x_k^2 + \cdots + a_{k-1} x_k^{k-1} && = \alpha_k] = 1/q^k \end{aligned}$$

We argue that this set of k linear equations (viewing $a_0, a_1, \ldots, a_{k-1}$ as unknowns) will have a unique solution. That will immediately give the probability. To show the uniqueness of the solution, we will prove that the following matrix, which is known as Vandermonde Matrix, is invertible.

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{k-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{k-1} \\ & & \vdots & & \\ 1 & x_k & x_k^2 & \cdots & x_k^{k-1} \end{bmatrix}$$

It's known that the determinant of the matrix is $\Pi_{1 \le i < j \le k}(x_j - x_i)$.

Homework:- Show that the Vandermonde matrix V is Invertible.

Hint: If the matrix is not invertible then it must have a nonzero null vector, say $0 \neq u \in \mathbb{F}^k$. Now, argue that this is not possible because of the fact that a polynomial of degree $k-1$ can have at most $k-1$ roots. □

4 Applications

One area with many interesting applications of k-wise independence is that of streaming algorithms. Here the the input comes as a stream and our storage capacity is much smaller than the input size. For instance, suppose our working space is bounded by $O(\log n)$, while the input size is n. Moreover, we can make only one pass over the stream. We would still like to compute something useful from the input.

One popular example is to count the number of distinct elements in a stream of elements, say,

$$a_1, a_2, \ldots, a_m \in \{1, 2, 3, \ldots, n\}$$

Clearly there is a trivial solution that takes $O(n)$ space. Can we use randomization and approximately count the number using only $O(\log n)$ space? We will see an approach in the next lecture.