6.842 Randomness and Computation Feb 04, 2020

Lecture 1

Lecturer: Ronitt Rubinfeld *Scribe: Damian Barabonkov*

1 The Probabalistic Method

Some mathematical objects either exist entirely or not at all; ie) they have binary probabilities of 0 or 1. In such cases, it may be first useful to show that they probably exists with a $Pr > 0$. Since we know the probability is either 0 or 1, and by proving it is greater than 0, then it must be 1. Therefore, the existence has been proven.

1.1 Example: 2-colored Sets

First let us define X to be a set of elements. From this X, we are given an input of m sets such that $S_1 \ldots S_m \subseteq X$. Each set S_i contains l elements from X.

Question: "Can we 2-color X such that each S_i has elements of both colors – is not monochromatic?"

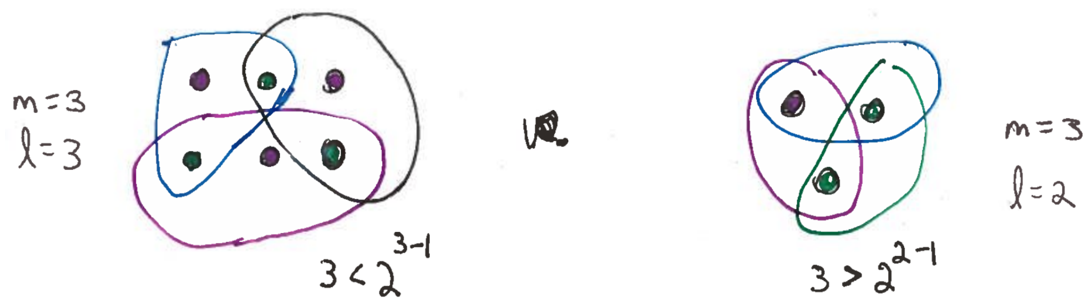

Figure 1: The instance on the left *can* be 2-colored, but the instance on the right *cannot*.

Theorem 1 *If $m < 2^{l-1}$, then there will exist a valid 2-coloring of X.*

Proof Intuition: Show that there are so many ways to 2-color X, so many so, that even by randomly coloring nodes, there will be a slight, albeit extremely unlikely, chance that this coloring produces a valid 2-coloring assignment.

Proof

Randomly color the elements of X red/blue, independently and identically distributed with probability $\frac{1}{2}$. In order to prove that such construction will yield a valid 2-coloring with non-zero probability, the probabilities on a set-by-set basis must be analyzed. For each set i, the probability it is monochromatic is simply the probability that all l elements were either colored all red $\frac{1}{2^l}$ or all blue $\frac{1}{2^l}$. These two events are disjoint and therefore their probabilities are simply summed.

$$Pr[S_i \text{ is monochromatic}] = \frac{1}{2^l} + \frac{1}{2^l} = \frac{1}{2^{l-1}}$$

Now, a union bound may be used over all i sets to get an upper bound on the probability that there exists a monochromatic set.

$$Pr[\exists i \text{ such that } S_i \text{ is monochromatic}] \leq \sum_i Pr[S_i \text{ is monochromatic}] \leq \frac{m}{2^{l-1}} < 1$$

Since there are m sets, their probabilities of being monochromatic ($\frac{1}{2^{l-1}}$) get summed m times. Then, the leap in $\frac{m}{2^{l-1}} < 1$ is achieved based on the theorem's initial assumption that $m < 2^{l-1}$. Taking the complement of $Pr[\exists i$ such that S_i is monochromatic$]$ will yield the $Pr[$all S_i are 2-colored$]$.

$$Pr[\text{all } S_i \text{ are 2-colored}] = 1 - Pr[\exists i \text{ such that } S_i \text{ is monochromatic}] > 0$$

This non-zero probability implies that there exists a 2-coloring of X that gives all m valid non-monochromatic sets S_i.

■

1.2 Example: Large Sum-Free Sets

The big picture of this example is to prove that in any set of n numbers, there exists a sub-set of size at least $\frac{n}{3}$ in which no two numbers can be taken and sum to a number that also is in the set.

We introduce some definitions required for the theorem.

Definition 2 $\mathbb{Z}_p \equiv \{0 \dots p-1\}$ *A set of all integer numbers less than p*

Definition 3 $\mathbb{Z}_p^* \equiv \{1 \dots p-1\}$ *A set of all integer numbers less than p that are also co-prime with p. Since p is a prime number itself, this set is virtually equivalent to* $\mathbb{Z}_p$ *without the 0. (As a notational remark, the star denotes the set of numbers that are co-prime with p.)*

Fact 4 *If p is prime, then multiplicative inverses in modular arithmetic modulo p exist* $\forall x \in \mathbb{Z}_p^*$. *In other words:* $\forall x, \exists x^{-1}$ *such that* $x \cdot x^{-1} \equiv 1 \pmod{p}$

Definition 5 *A is a set of some positive integers. A is "sum-free" if* $\nexists a_1, a_2, a_3 \in A$ *such that* $a_1 + a_2 = a_3$. *In plain English, a set is "sum-free" if no two elements in the set sum to another element also in the set.*

Theorem 6 (Erdos '65) $\forall B = \{b_1 \dots b_n\} \exists$ *sum-free* $A \subseteq B$ *such that* $|A| > \frac{n}{3}$

Simple Example: $B = \{1 \dots n\}$ then a possibility is $A = \{\lceil \frac{n}{2} \rceil \dots n\}$ This works because any two elements taken in the set A will sum to a value greater than n.

Theorem Proof Intuition:

1. First we prove that there is a continuous region $C \subseteq \mathbb{Z}_p^*$ whose elements pose a sum-free set.
2. Then we show that there is a way to construct A from B in such a way that each value in A can be randomly and uniquely mapped to this region C. And using this property, we consequently can prove the sum-free nature of A as well.
3. Lastly, we prove that, in expectation, the size of A will be at least $\frac{1}{3}$ the size of B. If the expectation is at least $|B|/3$, then there must be some choice of mapping that achieves $|B|/3$, and we can use that such one to define A.

Proof *For theorem intuition point 1*

Without loss of generality, let b_n be the maximal element in B.

Pick a prime p such that $p > 2b_n$ and $p \equiv 2 \pmod{3}$. In other words, $p = 3k+2$ for some k.

Let a set $C = \{k+1 \dots 2k+1\}$ represent the "middle third" elements.

1. $C \subseteq \mathbb{Z}_p^* \subset \mathbb{Z}_p$
2. C is sum-free, even in $\mathbb{Z}_p$

3. $\frac{|C|}{p-1} = \frac{k+1}{3k+1} > \frac{1}{3}$

The formulation in (1) falls through by definition. That is the range of C is from $k+1$ to $2k+1$ which are well within $\mathbb{Z}_p^*$ as it was defined.

To prove (2), summing the two smallest elements in C will still bring the result out of the range of C. Additionally summing the two largest elements in hopes of a wrap around will get to just before the beginning of C.

More formally:

$$\begin{aligned}(k+1)+(k+1) &= 2k+2 > 2k+1\\ (2k+1)+(2k+1) &= 4k+2 \pmod{p}\\ &= 4k+2 \pmod{3k+2}\\ &\equiv k \pmod{3k+2}\end{aligned}$$

The result of this derivation can be equivalently written as:

$$\begin{aligned}&\forall x, y \in C\\ &x+y \geq 2k+1 \pmod{3k+1}\\ &\qquad OR\\ &x+y \leq k \pmod{3k+1}\end{aligned}$$

Equation (3) relates the size of C with the size of possibilities of numbers, that is $p-1$, to show that $|C|$ is at least a third of the entire set of numbers.

The set C is simply a theoretical sum-free construction of proven minimal size. We now need to construct a sum-free set A which contains the actual values b_i using the help of C. This is done by mapping numbers from B to locations in C using a random linear map. ■

Claim 7 *A_x is sum-free*
Constructing A:

- *Pick* $x \in_R \{1 \ldots p-1\} \equiv \mathbb{Z}_p^*$.
- *Use x to define a random linear map* $f_x(a) = x \cdot a \pmod{p}$.
- *Then* $A_x \leftarrow \{b_i \text{ such that } f_x(b_i) \in C\}$. *In other words "the elements of B mapped to C by x"*

Proof *For theorem intuition point 2*

If $\exists b_i, b_j, b_k \in A_x$ such that $b_i + b_j = b_k$ then $xb_i + xb_j = xb_k \pmod{p}$.

All of xb_i, xb_j, xb_k are in C by construction which is sum-free. Therefore so are b_i, b_j, b_k all sum-free as well.

■

Claim 8 *$\exists x$ such that $|A_x| > \frac{n}{3}$*

Proof Intuition: We calculate the probability to map a value into C by utilizing how multiplicative inverses are unique in a prime number space and knowing the size of $|C|$. Then an indicator random variable can represent whether a value was mapped into C, and over all of the n elements, the expectation is that at least $\frac{n}{3}$ values will map into the sum-free C. Furthermore, we can conclude that there must be some combination of elements that achieve the expectation.

Proof *For theorem intuition point 3*

Fact 9 $\forall y \in \mathbb{Z}_p^* \; \exists$ *unique* $x \in \mathbb{Z}_p^*$ *such that* $y \equiv xb \pmod p$
$\Rightarrow \forall y \in \mathbb{Z}_p^*$, $\forall i Pr[y$ *mapped via* f_x *to* $b_i] = \frac{1}{p-1}$ *uses* $x \equiv yb^{-1} \pmod p$

This statement arises from the notion that only one x exists which can map a given y to b_i.

From this follows that $\forall i, |C|$ choices of x map b_i into C

Let us define an indicator random variable $\sigma_i^{(x)}$ which describes whether x mapped b_i into C, ie $(xb_i \in C)$.

More formally: $\sigma_i^{(x)} = \begin{cases} 1 \text{ if } x \text{ maps } b_i \text{ into } C \\ 0 \text{ otherwise} \end{cases}$

The expected value of this indicator value will show us with what frequency b_i gets mapped into C.
$E_x(\sigma_i^{(x)}) = Pr_x[\sigma_i^{(x)} = 1] = \frac{|C|}{p-1} > \frac{1}{3}$.

The numerator in $\frac{|C|}{p-1}$ comes from the number of choices for x to map b_i into C and the denominator are the total number of choices of x possible. So this value is proven above to be greater than $\frac{1}{3}$.

Now the average value of $|A_x|$ will be the sum of expectations for all n elements that land in C.
$|A_x| = E_x[|A_x|] = E_x[\sum_i \sigma_i^{(x)}] = \sum_i E[\sigma_i^{(x)}] > \sum_i \frac{1}{3} = \frac{n}{3}$
And from this it follows that if the average size of $|A_x| > \frac{n}{3}$, there must exist a specific x that is able to map A to C such that $|A_x| > \frac{n}{3}$.
■

Finally to prove the theorem that $\forall B = \{b_1 \ldots b_n\} \; \exists$ *sum-free* $A \subseteq B$ *such that* $|A| > \frac{n}{3}$
Proof

1. We proved that $C \subseteq \mathbb{Z}_p^*$ and C is sum-free.
2. We proved that if elements in A are mapped into C, then those elements of A also form a sum-free constituent.
3. We proved that there will always exist a selection of A for which $\frac{n}{3}$ can be mapped to C.

Therefore, there always exists $A \subseteq B$ of size at least $|A| \geq \frac{n}{3}$ which can be mapped to C. And that if they are mapped to C, those elements are all mutually sum-free as well. This concludes the theorem's proof!
■