Prof. Ronitt Rubinfeld

6.842: Randomness and Computation [1]

Day 1: January 31, 2022

Scribe: Priya Malhotra

1 The Probabilistic Method

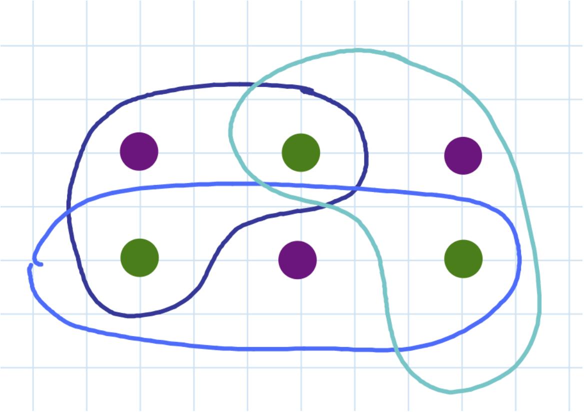
Some mathematical objects either completely exist or do not exist at all. These objects have binary probabilities of 0 or 1. In these cases, by showing that the probability of such an object existing is greater than 0, we can prove it's existence.

1.1 2-Colored sets

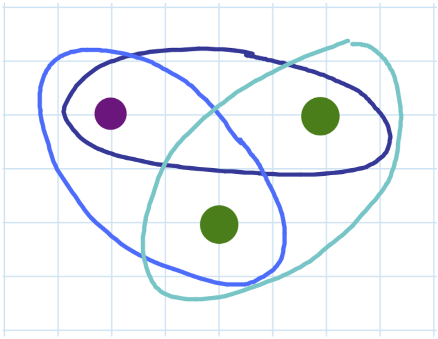
Let X be a set of elements. We have m subsets of X: $S_1, S_2, \cdots, S_m \subseteq X$, where each S_i contains l elements from X.

Question 1.1. *Can we 2-color X (meaning assign each element of X a color) such that each S_i has elements of both colors (is not monochromatic)?*

Example 1.2. *This example of X, where $m = 3$ and $l = 3$, has a 2-coloring:*



Example 1.3. *This example of X, where $m = 3$ and $l = 2$, cannot be 2-colored:*



Theorem 1.4. *If $m < 2^{l-1}$, then there will exist a valid 2-coloring of X*

The proof intuition for that is that there are so many different ways to two color X, that even by randomly coloring nodes, there will be a slight (even if extremely unlikely) chance that a valid 2-coloring assignment is produced. In example 1.1, we have that $3 < 2^{3-1}$ and there is indeed a proper 2-coloring. In example 1.2, $3 < 2^{2-1}$, and there is not a proper 2-coloring.

Proof. Randomly color the elements of X purple and green, independently and identically distributed, each with probability of half. In order to prove that this construction will give us a valid 2-coloring with a non-zero probability, we need to look at the probabilities for each set.

For each set i, the probability that S_i is monochromatic is simply the probability that all l elements were either colored all red or all blue, each of which happens with a probability of $\frac{1}{2^l}$.

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

We can use a union bound [i] over all i sets to get an upper bound on the probability that there exists a monochromatic set:

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

Because there are m sets their probabilities of being monochromatic (each $\frac{1}{2^{l-1}}$) get summed m times. We can conclude that $\frac{m}{2^{l-1}} < 1$ by the theorem's initial assumption that $m < 2^{l-1}$.

We can now take the complement to find the probability that all S_i are 2-colored.

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

Because we have a non-zero probability, we know that there exists a 2-coloring of X that gives all m valid non-monochromatic sets S_i.

□

[i] for any finite or countable set of events, the probability that at least one of the events happens is no greater than the sum of the probabilities of the individual events

This probabilistic method can tell us that there exists some valid 2-coloring, but gives us no insight on what this 2-coloring may actually be. Say we modified the theorem to be such that if $m < 2^{l-2}$, then there exists a proper 2-coloring. Then this would change the final conclusion to mean that the probability that all S_i are 2-colored is greater than half, so a random coloring would give us a stronger probability that we end up with a valid 2-coloring. This would mean that the expected number of random 2-colorings we would need to check is 2.

1.2 Dominating Set

Definition 1.5. Given a graph $G = (V, E)$, $U \subseteq V$ is a "dominating set" if for every node $v \in V \setminus U$, v has at least one neighbor in U.

Remark 1.6. Finding the minimum size of a dominating set is NP-hard – one of the first known such problems.

Theorem 1.7. *G has minimum degree $\triangle$, then G has a dominating set of size* $\leq \dfrac{4n \cdot \ln(4n)}{\triangle + 1}$

Proof. Construct $\hat{U}$: put each $v \in V$ into $\hat{U}$ independently with probability $\boldsymbol{p} = \dfrac{ln(4n)}{\triangle + 1}$

Is $\hat{U}$ a dominating set?

For $w \in V$, $\mathbf{Pr}[w$ has no neighbor in $\hat{U}$ and is not in $\hat{U}] \leq (1-p)^{\triangle+1}$ (using the independence in constructing $\hat{U}$)

We now consider $\mathbf{Pr}[\exists w \in V$ such that w has no neighbor in $\hat{U}$ and w not in $\hat{U}]$ [ii]

$$\begin{aligned}
\mathbf{Pr}[\exists w \in V \text{ such that } w \text{ has no neighbor in } \hat{U} \text{ and } w \text{ not in } \hat{U}] &\leq n \cdot (1-p)^{\triangle+1} \\
&\leq n \left(1 - \frac{ln(4n)}{\triangle + 1}\right)^{\left(\frac{\triangle+1}{\ln(4n)}\right) \cdot \ln(4n)} \\
&\approx n \cdot e^{-\ln(4n)} \\
&= n \cdot \frac{1}{4n} \\
&= \frac{1}{4}
\end{aligned}$$

Note, the above holds because $\lim_{x\to\infty}(1 - \frac{1}{x})^x \to \frac{1}{e}$. So, $\mathbf{Pr}[\hat{U}$ is not a dominating set$] \leq \frac{1}{4}$

How big is $\hat{U}$?

$$E[\hat{U}] = n \cdot p$$

$$\mathbf{Pr}[|\hat{U}| > 4np] \leq \frac{1}{4}$$

[ii] otherwise $\hat{U}$ is a Dominating Set

by Markov's inequality.

So $\mathbf{Pr}\left[\hat{U} \text{ is a Dominating Set of size } \leq \dfrac{4n \cdot \ln(4n)}{\triangle + 1}\right] \geq 1 - \frac{1}{4} - \frac{1}{4} \geq \frac{1}{2} > 0$, which means it exists!

□

1.3 Sum Free Subsets

A is a subset of positive integers (> 0)

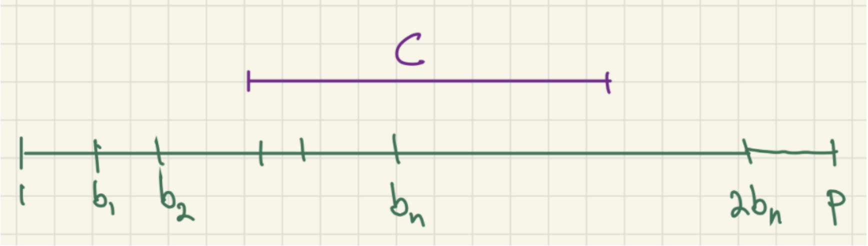
Definition 1.8. A is "sum free" is $\nexists a_1, a_2, a_3 \in A$ such that $a_1 + a_2 = a_3$

Theorem 1.9. *(Erdös '65)* $\forall B = \{b_1, \cdots, b_n\} \exists$ *sum-free* $A \subseteq B$ *such that* $|A| > \frac{n}{3}$

Notation: $[n] = \{1, 2, \cdots, n\}$
$B = [n]$, $A = \{x \mid x \equiv 1 \mod 3$ and $A' = \{\lceil \frac{n}{2} \rceil, \cdots, n\}$

Proof. Without loss of generality, b_n is the maximum element in B. Pick prime $p > 2b_n$ such that $p \equiv 2 \mod (3)$ (i.e. $p = 3k + 2$ for some k).

Let $C = \{k + 1, \cdots, 2k + 1\}$, which is the "middle third"



Number theory reminder: $\mathbb{Z}_p = \{0, \cdots, p - 1\}$ and $\mathbb{Z}_p^* = \{1, \cdots, p - 1\}$. Every value in $\mathbb{Z}_p^*$ has exactly one multiplicative inverse.

Example 1.10. *The multiplicative inverses in the values of* $\mathbb{Z}_3^* = \{1, 2\}$ *are themselves:*
Observe, $1 \cdot 1 \equiv 1 \mod 3$ *and* $2 \cdot 2 \equiv 1 \mod 3$

Observation 1.11. *1.* $C \subseteq \mathbb{Z}_p^*$

2. *C is sum-free even in $\mathbb{Z}_p$.*

   Say we take the largest element in C, $2k+1$, and add it together twice, then $2k+1+2k+1 \equiv 4k + 2 \mod 3k + 2 \equiv k$. Even this element does not wrap back around to a value in C (because the smallest element of C is $k + 1$).

   In addition, any pair of elements in C will add to an integer that is larger than the maximum element of C.

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

Notation: $\in_R$ means picking an element randomly.

Observation 1.12. *Sum-freeness extends to linear functions of elements:*
If $x_1 + x_2 = x_3$, then $a \cdot x_1 + a \cdot x_2 = a \cdot x_3$

We will use this notion "backwards."

Construction A: pick $x \in_R \{1, \cdots, p-1\} = \mathbb{Z}_p^*$, use x to define a (random) linear map $f_x(a) \equiv x \cdot a \mod p$

Let $A_x = \{b_i$ such that $f_x(b_i) = x \cdot b_i \mod p \in C\}$

So A_x are elements of B in the preimage of C under f_x. "x maps the above to the middle third."

Claim 1.13. *A_x is sum-free.*

Proof. If not, then there exists a $b_i, b_j, b_k \in A_x$ such that $b_i + b_j = b_k$ so that $xb_i + x_j = xb_k \mod p$, which means C is not sum-free in $\mathbb{Z}_p$, so we get a contradiction. Therefore, A_x is sum-free. □

Claim 1.14. *$\exists x$ such that $|A_x| > \frac{n}{3}$*

Observation 1.15. *$\forall\ y \in \mathbb{Z}_p^*$ and $\forall\ i$, exactly one $x \in \mathbb{Z}_p^*$ satsifies $y \equiv x \cdot b_i (\mod p)$.*
Then $\forall\ y \in \mathbb{Z}_p^*$ and $\forall\ i$, $\mathbf{Pr}_x[y$ mapped to $b_i] = \frac{1}{p-1}$.
$\forall i$, this means that $|c|$ choices of x such that $x \cdot b_i \mod p \in C$.

Define $\sigma_i^{(x)} = \begin{cases} 1 & \text{if } x \cdot b_i \mod p \in C \\ 0 & \text{otherwise} \end{cases}$

$E_x[\sigma_i^{(x)}] = \mathbf{Pr}_x[\sigma_i^{(x)} = 1] = \frac{|c|}{p-1} > \frac{1}{3}$

$\sum_i \sigma_i^{(x)}$ is the number of b's that map to C under x.

$E_x[|A_x|] = E_x[\sum_i \sigma_i^{(x)}] = \sum_i E_x[\sigma_i^{(x)}]$

Therefore, there exists at least one x such that $|A_x| > \frac{n}{3}$

□

References

[1] Ronitt Rubinfeld. Massachusetts Institute of Technology, 6.842: Randomness and Computation, 2022. https://people.csail.mit.edu/ronitt/COURSE/S22/index.html.