Lecture 32: Basic probability theory

Nitin Saxena [*]

IIT Kanpur

1 The probabilistic method

1.1 Sum-free subsets

Let us take another example. Given a set of integers S, $S+S$ is defined as the subset of integers which contain all possible sums of pair of elements in S,

$$S+S := \{t : t = s_1 + s_2,\ s_1, s_2 \in S\}\,.$$

A set S is called *sum-free* if S does not contain any element of $S+S$.

Exercise 1. Construct a set of 10 elements which is sum-free. Construct a set of n elements which is sum-free.

Eg. you can take a sequence of rapidly growing integers. Eg. $\{3, 3^2, \ldots, 3^n\}$.

Fermat's last theorem says that the set $\{1, 2^a, 3^a, \ldots, n^a\}$ is sum-free, for every $a \in \mathbb{N}_{>2}$.

Using probabilistic method, we will show that every subset of integers contains a large sum-free subset.

Theorem 1. *For any subset S of n non-zero integers, There exists a subset of S which is sum-free and has size more than $n/3$.*

Proof. Suppose $S = \{s_1, s_2, \cdots, s_n\}$. The idea would be to randomly transform S to $rS = \{rs_1, rs_2, \cdots, rs_n\}$, for a random $r \mod p$ (p is a fixed prime). If some subset of rS is sum-free then the corresponding set in S will also be sum-free (Why?).

First, pick a prime p of the form $3k+2$, such that, p is at least 3 times bigger than the absolute value of any element of S.

Exercise 2. Show that there are infinitely many primes of the form $3k+2$.

Modify Euclid's proof for infinite primes.

We will do the calculations modulo prime $p = 3k+2$.

Notice that the set $T = \{k+1, k+2, \cdots, 2k+1\}$ is a sum-free subset when we do addition modulo p. It is the middle-1/3rd of $[0, \ldots, p-1]$.

For applying the probabilistic method, pick a random r and consider the set $rS \mod p := \{rs_1 \mod p, rs_2 \mod p, \cdots, rs_n \mod p\}$.

Exercise 3. Show that if we pick an r at random from $0, 1, \cdots, p-1$ then $rs_1 \mod p$ is also random with uniform probability.

Since, $s_1 \neq 0 \mod p$ as we assumed p to be large and the elements of S to be nonzero.

Define a random variable Y which is the intersection size of $rS \mod p$ and T.

Using linearity of expectation,

$$E[Y] = \sum_i E[rs_i \mod p \in T]\,.$$

[*] Edited from Rajat Mittal's notes.

Exercise 4. Show that $E[Y] > \frac{|S|}{3}$.

This implies that there exists at least one r for which $rS \mod p \cap T$ is of size at least $|S|/3$. Call that particular r, x_0. Then $T' := x_0 S \mod p \cap T$ is sum-free when addition is considered modulo p ($\because T$ is sum-free). This implies that the pre-image in S which maps to T' is also sum-free.

Exercise 5. Show that $x_0^{-1}T'$ is sum-free with respect to addition over integers.

□

This also gives a fast randomized algorithm to find a sum-free subset of a given set S.

2 Using linearity of expectation

We have already discussed linearity of expectation. It is a simple result to prove, but has profound implications. Again, the importance of linearity lies in the fact that we can even take dependent random variables and still decompose the expectation into components.

$$E[X+Y] = E[X] + E[Y].$$

for any two random variables X and Y.

Notice that we used linearity of expectation for the proof in the previous section. We will take some more examples now.

2.1 Ramsey number revisited

First, let us look at the example of Ramsey number in the light of expectation.

Suppose we color each edge of K_n uniformly at random with blue or red. Define T to be the random variable which counts the number of monochromatic K_k in the coloring. We are interested in the expectation of T.

Define T_i (for i from 1 to $\binom{n}{k}$) to be the random variable which assigns 1 if a particular K_k is monochromatic otherwise 0. Convince yourself that $T = \sum_i T_i$.

Note 1. The random variables T_i are dependent on each other.

Then,

$$E[T] = \sum_i E[T_i] = \sum_i 2^{1-\binom{k}{2}} = \binom{n}{k} 2^{1-\binom{k}{2}}.$$

If $E[T] < 1$ then there exists a coloring which has less than or equal to $E[T]$ number of monochromatic K_k's. Since number of monochromatic K_k's is an integer, there exists a coloring for which number of monochromatic K_k's is zero.

Let us take another example of probabilistic method which utilizes linearity of expectation.

2.2 Discrepancy

Given n arbitrary vectors we want to partition them into two parts such that the "discrepancy" between the two sums is "small".

Theorem 2. *Given n unit vectors $v_i \in \mathbb{R}^n$, $i \in [n]$, there always exists a "bit" string $b \in \{-1,1\}^n$, such that,*

$$\left\| \sum_i b_i v_i \right\| \leq \sqrt{n}\,.$$

References

1. N. Alon and J. H. Spencer. The Probabilistic Method. *Wiley*, 2008.
2. H. Tijms. Understanding Probability. *Cambridge University Press*, 2012.
3. D. Stirzaker. Elementary Probability. *Cambridge University Press*, 2003.
4. U. Schöning. Gems of Theoretical Computer Science. *Springer-Verlag*, 1998.