**6.842 Randomness and Computation** February 5th, 2014

# Lecture 1

*Lecturer: Ronitt Rubinfeld* *Scribe: Chiyuan Zhang*

## 1 The Probabilistic Method

The *probabilistic method*, pioneered by Paul Erdős, is a way to prove the existence of some mathematical object, by showing the probability that it exists is greater than zero. Since the object should either exist or not deterministically, this proves that it exists. Or we could view this as a way of doing "fancy" counting by employing probability as the language.

### 1.1 Example: Proper 2-coloring

Consider a collection of sets $S_1, \ldots, S_m \subset S$, where $|S_i| = \ell, i = 1, \ldots, m$. Can we 2-color the objects in $S$ such that none of the sets $S_1, \ldots, S_m$ are monochromatic?

Figure 1 shows two examples: the first example is a proper 2-coloring so that none of the subsets are monochromatic. The second example is an impossible case. When you color any object arbitrarily, the other two objects need to be assigned with the other color, but they cannot be both assigned with the (same) other color since they form a subset.

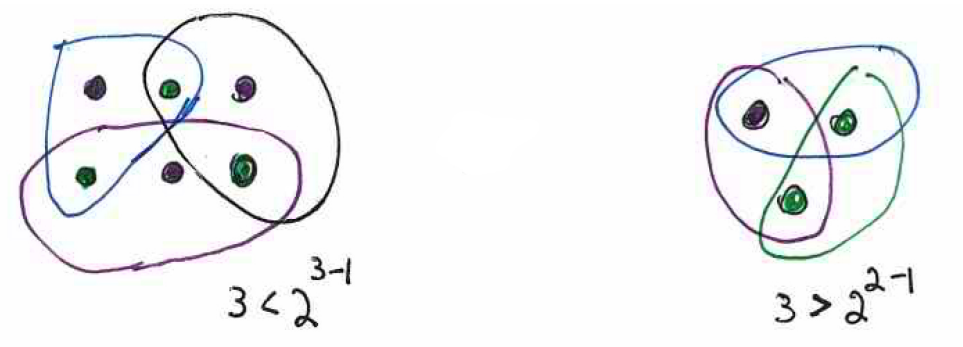


**Figure 1**: Example of proper 2-coloring problem. Left: a solution. Right: no solution possible.

We can prove that when the number of sets is not too large, there is always a proper 2-coloring.

**Theorem 1** *If $m < 2^{\ell-1}$, there exists a proper 2-coloring.*

**Proof** Let us independently and uniformly randomly color each object in $S$ with red or blue. For all $i = 1, \ldots, m$:

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

because it only happens when $S_i$ is colored all-red or all-blue. By union bound, we have

$$\Pr[\exists i \text{ s.t. } S_i \text{ is monochromatic}] \leq \sum_i \Pr[S_i \text{ is monochromatic}] = \frac{m}{2^{\ell-1}} < 1$$

where the last step is by our assumption. So

$$\Pr[\text{proper 2-coloring}] = \Pr[\nexists \text{ a monochromatic } S_i] > 0$$

which implies there must be a setting which gives a proper 2-coloring. ■

Some remarks regarding this example

- We used the language of probability, but our conclusion is deterministic.
- We showed that after ruling out the coloring settings with monochromatic sets, we still have other coloring setting, but did not show how many of them left.
- The theorem does *not* imply that no proper 2-coloring exists when $m \geq 2^{\ell-1}$.

## 1.2 Example: Sum-free Sets

**Definition 2** *A subset $A$ of positive integers is called* sum-free, *if $\nexists\ a_1, a_2, a_3 \in A$ s.t. $a_1 + a_2 = a_3$.*

**Theorem 3 (Erdős '65)** *For any subset $B = \{b_1, \ldots, b_n\}$ of positive integers, $\exists\ A \subset B$ such that $A$ is sum-free and $|A| > n/3$.*

For some simple cases, it is easy to construct such subsets. For example, let $B = \{1, \ldots, n\}$, then $A = \{n/2 + 1, \ldots, n\}$ is clearly sum-free as the sum of any two elements in $A$ will be larger than $n$; and its size is large enough.

The theorem is also quite tight. For example, it could be shown that the statement could not be true for $|A| > 12n/29$ [1].

**Proof** (Theorem 3)

Without loss of generality, let $b_n = \max_i\{b_i\}$. Take a prime $p > 2b_n$ such that $p \equiv 2 \mod 3$, that is, $p = 3k + 2$ for some integer $k$.

Let $C$ be the "middle third" of $\{1, \ldots, p\}$. More precisely, let $C = \{k + 1, \ldots, 2k + 1\}$. Notice that

1. $C \subset \mathbb{Z}_p^*$
2. $C$ is sum-free under both regular addition and addition in $\mathbb{Z}_p^*$: for regular addition, this is obvious since the sum of two numbers is always greater than $2k + 2$. For addition in $\mathbb{Z}_p^*$, one more thing we need to worry about is that when the sum goes beyond $p$ and wrap around, it will fall in $C$ again. But that is impossible, as $(2k + 1) + (2k + 1) \mod (3k + 2) = k \mod (3k + 2) < k + 1$. So the summation will always be less than $k + 1$, thus cannot fall in $C$.
3. it is also easy to see
$$\frac{|C|}{p-1} = \frac{k+1}{p-1} = \frac{k+1}{3k+1} > \frac{1}{3} \tag{1}$$

In the following, we will construct a random $A$ that is always sum-free, and we will show that the probability that $A$ is large is positive.

For $x \in_R \mathbb{Z}_p^*$, where $\in_R$ means choose at uniform random. Let

$$A_x \leftarrow \{b_i \in B : (xb_i \mod p) \in C\}$$

which contains elements of $B$ in the preimage of $C$ under the map by $x$.

**Claim 4** *$A_x$ is sum-free.*

**Claim 5** *$\exists x$ s.t. $|A_x| > n/3$.*

The theorem is proven by the two claims. ■

**Proof** (Claim 4) If the claim does not hold. Let $b_i, b_j, b_k \in A_x$ such that $b_i + b_j = b_k$. Then

$$\begin{aligned} & xb_i + xb_j = xb_k \\ \Rightarrow\quad & xb_i + xb_j \equiv xb_k \mod p \end{aligned}$$

But $xb_i, xb_j, xb_k \in C$, which is sum-free, a contradiction. ■

**Proof** (Claim 5) We first note that $\forall y \in \mathbb{Z}_p^*$ and $\forall i$, exactly one $x \in \mathbb{Z}_p^*$ that satisfies

$$y \equiv xb_i \mod p$$

by the uniqueness of the inverse of a group. This implies $\forall y \in \mathbb{Z}_p^*$ and $\forall i$,

$$\Pr_x [y \text{ mapped to } b_i] = \frac{1}{p-1} \tag{2}$$

This further implies $\forall i$, there are $|C|$ choices of $x$ such that $xb_i \mod p \in C$. Let

$$\sigma_i^{(x)} \leftarrow \begin{cases} 1 & \text{if } xb_i \mod p \in C \\ 0 & \text{otherwise} \end{cases}$$

then we have

$$\mathbb{E}\left[\sigma_i^{(x)}\right] = \Pr\left[\sigma_i^{(x)} = 1\right] = \frac{|C|}{p-1} > \frac{1}{3}$$

by Eq. (1). Thus

$$\mathbb{E}\left[|A_x|\right] = \mathbb{E}\left[\sum_{i=1}^{n} \sigma_i^{(x)}\right] = \sum_{i=1}^{n} \mathbb{E}\left[\sigma_i^{(x)}\right] > \frac{n}{3}$$

Since the expectation is the average, there must $\exists x$ such that $|A_x| > n/3$. ■

Remarks: in this example, we showed another common paradigm of probabilistic method: when we want to show the existence of a setting under which the variable $\eta$ satisfy some lower/upper bounds, we did that by random construction and showing $\mathbb{E}[\eta]$ satisfies the bounds, which implies the existence.

# 2 The Lovász Local Lemma

A useful tool when we want to control the probability that some "bad events" happen. Let $p = \Pr[A_i]$ be the probability that one bad event $A_i$ happens. The two extreme cases are

1. When we have no assumption / prior knowledge about the (in)dependency of the events: we use union bound:
$$\Pr\left[\bigcup_{i=1}^{n} A_i\right] \leq \sum_{i=1}^{n} \Pr\left[A_i\right]$$
so $p < 1/n$ is needed to control the probability.

2. When the events are independent: we have
$$\Pr\left[\bigcup_{i=1}^{n} A_i\right] = 1 - \Pr\left[\bigcap_{i=1}^{n} \overline{A}_i\right]$$
so we could control the probability by just requiring that $A_i$s are "non-trivial", i.e. does not happen with probability 1.

Lovász Local Lemma deals with the intermediate case, when the events are not completely independent, but still "mostly" independent, we could still control the probability in a much nicer way than the union bound. The details will be covered in the next lecture.

# References

[1] N. Alon and D. J. Kleitman. Sum-free subsets. In *A tribute to Paul Erdős*, pages 13–26. Cambridge Univ. Press, Cambridge, 1990.