6.842 Randomness & Computation
Ronitt Rubinfeld
Scribed by Edward Z. Yang <ezyang@mit.edu>

Randomness is a resource: lets us do NEW things, and old things FASTER!
prove existence of combinatorial objects (non-constructively), SIMPLER
exp distributed systems
in proofs its a language for counting, also interactive proofs
learning and testing algorithms
4 (to predict)

Do We Require Randomness? more, less? when?

Learning vs Randomness complexity theory
Hardness vs. Randomness
Average case hardness of probs

Algorithms
Learning
Complexity

a lot of metalevels!
Tools: Fourier representation
Algebraic Techniques
Lovasz Local Lemma

Today's lecture:
- The Probabilistic Method
- The Lovász Local Lemma
The Probabilistic Method

Descartes: “I think, therefore I am.”
Erdős: “I toss coins, therefore I am.” (paraphrased)

Show $E$, by showing it probably exists; if the probability it exists is positive (non-zero), it must exist (existence is a binary proposition) “Fancy counting”

Example 1

```
Input: given $S_1, \ldots, S_m \subseteq S$ (ground set)
Output: can we 2-color objects in $S$ s.t. each $S_i$ not monochromatic (not all the same color)
```

Cobbling is nontrivial

**Def:** Hypergraph is $(V, E)$, where each $e \in E$ is subset of $V$
(an ordinary graph is when all subsets have two elements.)
(So this is hypergraph coloring.)

**Goal:** Show there exists a coloring, when $m < 2^{l-1}$
Proof: Randomly color each element of $S$ red or blue with probability $\frac{1}{2}$ (independently).

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

abuse of notation, we haven't defined the event $S_i$ monochromatic

$$\Pr[\exists i \text{ such that } S_i \text{ monochromatic}] = \Pr[\bigcup_i S_i \text{ monochromatic}]$$

$$\leq \sum_i \Pr[S_i \text{ monochromatic}] \quad (\text{union bound})$$

$$= m \cdot \frac{1}{2^{k-1}} < 1$$

by assumption

$$\therefore \Pr[\text{good coloring}] > 0 \quad \blacksquare$$

Intuitively: There exist many colorings, but even when we rule out monochromatic ones, there are leftover colorings.

Note: We don't know what coloring works, or even how many colorings exist. Algorithm to find this takes exponential time.
Example 2

A is a subset of positive integers (> 0)

Def: A is "sum-free" if \( \neg \exists a_1, a_2, a_3 \in A \) s.t. \( a_1 + a_2 = a_3 \)

Thm [Erdös 65] \( \forall B = \{b_1, \ldots, b_n\}, \exists \text{ sum-free } A \subseteq B \) s.t. \( |A| \geq \frac{n}{3} \) (but this is not true for \( |A| \geq \frac{12}{29} n \))

e.g. \( B = \{1, \ldots, n\} \)
\[ A = \{\frac{n}{2} + 1, \ldots, n\} \] (as all pairs sum to value greater than n)

Proof: wlog. \( b_n \) is max elt of B

pick prime \( p > 2b_n \) s.t. \( p = 2 \pmod{3} \)

i.e. \( p = 3k + 2 \) for some \( k \in \mathbb{Z}_4 \)

\[ \begin{array}{cccccccc}
1 & b_1 & b_2 & \cdots & b_n & \cdots & 2b_n & p \\
\end{array} \]

Let \( C = \{k + 1, \ldots, 2k + 1\} \)

Note: \( C \subseteq \mathbb{Z}^*_p \) (numbers \( \mod p \), relatively prime to \( p \))

\( C \) is sum-free (the sum is outside the range)

\[ \left( \frac{|C|}{p-1} \right) > \frac{1}{3} \quad \left( \frac{|C|}{p-1} = \frac{k+1}{3k+1} \right) \]
Constructing $A$: \[ \mathbb{Z}_p^* \] (a nice set w/ lots of properties)

Pick $X \in \mathbb{Z}_p^* \frac{1}{\ldots} p-1$ \[ \mapsto \text{pick } X \text{ from set uniformly at random} \]

Let $A_X = \{ b_2 \text{ s.t. } (xb_2 \mod p) \in C \}$

---

Claim: $A_X$ is sum-free.

Pf: Let $b_i, b_j, b_k \in A_X$ s.t. $b_i + b_j = b_k$

But then $xb_i + xb_j = xb_k \pmod{p}$

by construction these $\in C$

Contradiction with $C$ being sum-free. \(\blacksquare\)

---

Warning: Why don’t we just take the $b_i$ which are in $C$?

Look closely at what the direction is.

Also note: $C$ is sum-free \(\pmod{p}\) (since it’s a third of the space)
Next goal: show $A_x$ is big. (Will show exists one $X$ w/ property)

**Claim** $\exists X \text{ s.t. } |A_x| > \frac{n}{3}$

**Fact** $\forall y \in \mathbb{Z}_p^*$ and $\forall z$, there is exactly one $x \in \mathbb{Z}_p^*$ that satisfies $y = x \cdot z \pmod{p}$

(by existence of inverses, linear equation has unique solution)

Proof of fact in last year's notes.

Idea: show how many choices of $X$ make a given $b_z$ land in center area.

$\forall y$, $\text{Fact } \implies |C|$ choices of $X$ such that $X \cdot b_z \in C$

(i.e. one for each element of $C$)

Define $g_z(x) = \begin{cases} 1 & \text{if } x \cdot b_z \in C \\ 0 & \text{otherwise} \end{cases}$ (Indicator value)

$$\mathbb{E}_x[|A_x|] = \mathbb{E}_x\left[ \sum_{z} g_z(x) \right]$$

$$= \sum_{z} \mathbb{E}_x[g_z(x)]$$ (Linearity of expectation)

Intuitively, this is the average.

So there must be some value that hits the average

$$\Pr_x[ g_z(x) = 1 ] = \frac{|C|}{p-1} > \frac{1}{3}$$ (property of indicator variable)

$$> \frac{n}{3}$$ so since at least one $X$ gives at least expectation; theorem follows.
Lovász Local Lemma

A₁, …, Aₙ, bad events

Naive way: (best we can do in general)

Pr[UAᵢ] ≤ ∑ Pr[Aᵢ] (union bound)

In general, need that Pr[Aᵢ] < 1/n for each i to show Pr[UAᵢ] < 1 i.e. Pr[∩Aᵢ] > 0

(That is, it is possible no bad events happen)

Very Strong condition

If Aᵢ’s are independent and “non-trivial”

Pr[∩Aᵢ] = ∏ Pr[Aᵢ] > 0

In the naive case, we have stringent requirement on Pr[Aᵢ], but no independence condition. In the second case, we have stringent indep. req., but relaxed Pr[Aᵢ]. We want something in the middle, i.e.

[\{n\} = \{1, …, n\}]

Def A “independent” of B₁, …, Bₖ if ∀ J ⊆ [K] s.t. J ≠ \∅

Pr[A ∩ ∩ Bᵢ] = \prod_{j \in J} Pr[Aᵢ] Pr[∩ Bᵢ]

(Note: this is not pair-wise independence.)

Def Given events A₁, …, Aₙ, D = (V, E) with V = [n]

is a “dependency digraph of A₁, …, Aₙ” if each Aᵢ is independent of the set of all Aⱼ that don’t neighbor it in D.
Lovász Local Lemma (symmetric version)

Given $A_1, \ldots, A_n$ s.t. $\Pr[A_i] \leq p \ \forall i$

and dependency digraph $D$ of degree $\leq d$,

If $e^p \cdot (d+1) \leq 1$

then $\Pr[\bigcap_{i=1}^n \overline{A_i}] > 0$

Note the requirement doesn't rely on $n$; only the degree $d$.

Next Time: New version of hypergraph 2-coloring w/ bounding on intersection, rather than bound on number of subsets.