Randomness in Computing

CS 537

LECTURE 19

Last time

- Hashing

Today

- Probabilistic method

The probabilistic method

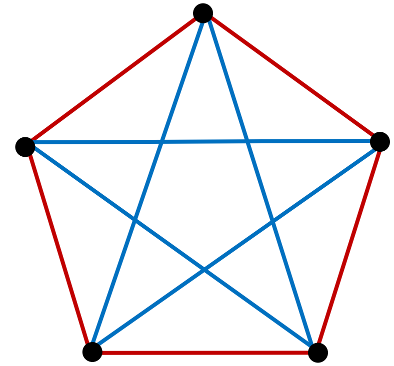
To prove that an object with required properties exists:

1. Define a distribution on objects.
2. Sample an object.
3. Prove that a sampled object has required properties with positive probability.

- Sometimes proofs of existence can be converted into **efficient randomized constructions**.
- Sometimes they can be converted into deterministic constructions (**derandomization**).

Method 1: The counting argument

- K_n = complete graph on n vertices (n-clique)



Theorem

If $\binom{n}{k} \cdot 2^{-\binom{k}{2}+1} < 1$ then it is possible to color the edges of K_n with two colors so that no K_k is monochromatic.

Proof: Define a random experiment:

Color each edge of K_n independently and uniformly blue or red.

- Fix an ordering of the $\binom{n}{k}$ different k-cliques.
- Let M_i be the event that clique i is monochromatic, for $i = 1, \ldots, \binom{n}{k}$

Union Bound

$$\Pr[M_i] = 2 \cdot 2^{-\binom{k}{2}}$$

- $\Pr\left[\bigcup_{i=1}^{\binom{n}{k}} M_i\right] \leq \sum_{i=1}^{\binom{n}{k}} \Pr[M_i] = \binom{n}{k} \cdot 2^{-\binom{k}{2}+1} < 1$
- Probability of a coloring with no monochromatic k-clique is > 0.

Converting an existence proof into an efficient randomized construction

- Can we efficiently sample a **coloring**? *Yes*
- How many samples do we need to generate **a coloring with no monochromatic *k*-clique**?
  - Probability of success is at least $p = 1 - \binom{n}{k} \cdot 2^{-\binom{k}{2}+1}$
  - # of samples $\sim \text{Geom}(p)$, expectation: $1/p$
  - Want: $1/p$ to be polynomial in the problem size
  - If $1 - p = o(1)$, we get a Monte Carlo construction algorithm that errs with probability $o(1)$.
- To get a Las Vegas algorithm (always correct answers), we need a poly-time procedure for checking if **the coloring is monochromatic**.
  - If k is constant, we can check that all $\binom{n}{k}$ cliques are not monochromatic.

Method 2: The expectation argument

- It can't be that everybody is better (or worse) than the average.

Claim

Let X be a R.V. with $\mathbb{E}[X] = \mu$. Then $\Pr[X \geq \mu] > 0$ and $\Pr[X \leq \mu] > 0$.

Proof (by contradiction):

$\leq$

Suppose to the contrary that $\Pr[X \geq \mu] = 0$. Then

$$\mu \; > \; \mathbb{E}[X] = \sum_x x \Pr[X = x]$$

$$< \sum_x \mu \Pr[X = x] = \mu \sum_x \Pr[X = x] = \mu,$$

a contradiction.

Example: Finding a large cut

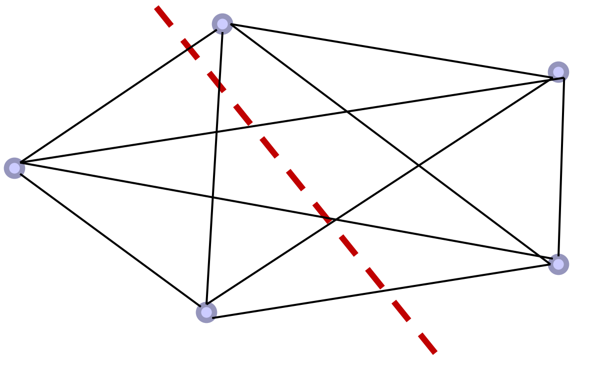
Recall:

- A cut in a graph $G = (V, E)$ is a partition of V into two nonempty sets.
- The size of the cut is the number of edges that cross it.
- Finding a max cut is NP-hard.

Theorem

Let G be an undirected graph with m edges.

Then G has a cut of size $\geq m/2$.



Example: Existence of a large cut

Theorem

Let G be an undirected graph with m edges.

Then G has a cut of size $\geq m/2$.

Proof: Construct sets A and B of vertices by assigning each vertex to A or B uniformly and independently at random.

- For each edge e, let $X_e = \begin{cases} 1 & \text{if edge connects } A \text{ to } B \\ 0 & \text{otherwise} \end{cases}$

  $\mathbb{E}[X_e] = 1/2$

- Let $X = \#$ of edges crossing the cut.

  Linearity of expectation

  $$\mathbb{E}[X] = \mathbb{E}[\sum_{e \in E} X_e] = \sum_{e \in E} \mathbb{E}[X_e] = m \cdot \frac{1}{2} = \frac{m}{2}$$

There exists a cut (A, B) of size at least $m/2$.

Example: Finding a large cut

- It is easy to choose a random cut
- Probability of success: $p = \Pr\left[X \geq \frac{m}{2}\right]$
- An upper bound on X? $X \leq m$

$$\frac{m}{2} = \mathbb{E}[X] = \sum_{i<m/2} i \cdot \Pr[X = i] + \sum_{i \geq m/2} i \cdot \Pr[X = i]$$

$$\leq \frac{m-1}{2} \cdot (1-p) + m \cdot p$$

$$m \leq m - 1 - (m-1) \cdot p + 2m \cdot p$$

$$p \geq \frac{1}{m+1}$$

- Expected # of samples to find a large cut: $\leq m+1$
- Can test if a cut has $\geq \frac{m}{2}$ edges by counting edges crossing the cut (poly time)

Las Vegas

Derandomization: conditional expectations

Finding a large cut

Idea: Place each vertex deterministically, ensuring that

$$\mathbb{E}[X \mid \text{placement so far}] \geq \mathbb{E}[X] = \frac{m}{2}$$

- R.V. Y_i is A or B, indicating which set vertex i is placed in, $\forall i \in [n]$

Base case: $\mathbb{E}[X|Y_1 = A] = \mathbb{E}[X|Y_1 = B] = \mathbb{E}[X]$

By symmetry (it doesn't matter where the first node is)

Inductive step: Let $y_1, \ldots, y_k$ be placements so far (each is A or B) and suppose $\mathbb{E}[X|Y_1 = y_1, \ldots, Y_k = y_k] \geq \mathbb{E}[X]$.

By Law of Total Expectation

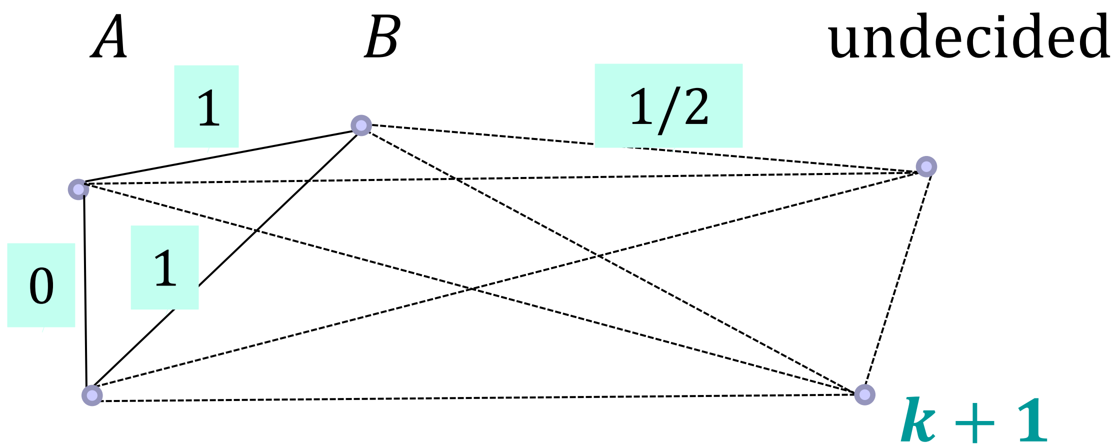
$$\mathbb{E}[X|Y_1 = y_1, \ldots, Y_k = y_k] = \frac{1}{2}\mathbb{E}[X|Y_1 = y_1, \ldots, Y_k = y_k, Y_{k+1} = A] + \frac{1}{2}\mathbb{E}[X|Y_1 = y_1, \ldots, Y_k = y_k, Y_{k+1} = B]$$

Pick y_{k+1} to maximize conditional expectation

Then $\mathbb{E}[X|Y_1 = y_1, \ldots, Y_{k+1} = y_{k+1}] \geq \mathbb{E}[X|Y_1 = y_1, \ldots, Y_k = y_k] \geq \mathbb{E}[X]$

Finding a large cut: derandomization

When the dust settles



- Place vertex $k+1$ in the set (A or B) with fewer neighbors, breaking ties arbitrarily

Example 2: Maximum satisfiability (MAX-SAT)

Logical formulas

- **Boolean variables:** variables that can take on values T/F (or 1/0)
- **Boolean operations:** $\vee$, $\wedge$, and $\neg$
- **Boolean formula:** expression with Boolean variables and ops

SAT (deciding if a given formula has a satisfying assignment) is NP-complete

- **Literal:** A Boolean variable or its negation. x_i or $\overline{x_i}$
- **Clause:** OR of literals. $C_1 = \overline{x_1} \vee x_2 \vee x_3$
- **Conjunctive normal form (CNF):** AND of clauses. $C_1 \wedge C_2 \wedge C_3 \wedge C_4$

Ex: $(\overline{x_1} \vee x_2 \vee x_3) \wedge (x_1 \vee \overline{x_2} \vee x_3) \wedge (x_2 \vee x_3) \wedge (\overline{x_1} \vee \overline{x_2} \vee \overline{x_3})$

$x_1 = 1$, $x_2 = 1$, $x_3 = 0$ satisfies the formula.

MAX-SAT: Given a CNF formula, find an assignment satisfying as many clauses as possible.

- Assume no clause contains x and $\bar{x}$ (o.w., it is always satisfied).

Example 2: MAX-SAT

Theorem

Given m clauses, let k_i = # literals in clause i, for $i \in [m]$.

Let $k = \min_{i \in [m]} k_i$. There is an assignment that satisfies at least

$m(1 - 2^{-k})$ clauses.

Proof: Assign values 0 or 1 uniformly and independently to each variable.

- X_i = indicator R.V. for clause i being satisfied.
- X = # of satisfied clauses = $\sum_{i \in [m]} X_i$
- $\Pr[X_i = 1] = 1 - 2^{-k_i}$

$$\mathbb{E}[X] = \sum_{i \in [m]} \mathbb{E}[X_i] = \sum_{i \in [m]} (1 - 2^{-k_i}) \geq m(1 - 2^{-k})$$

- There exists an assignment satisfying at least that many clauses.

Example 3: Large sum-free subset

- Given a set $\boldsymbol{A}$ of positive integers, a sum-free subset $\boldsymbol{S} \subseteq \boldsymbol{A}$ contains no three elements $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k} \in \boldsymbol{S}$ satisfying $\boldsymbol{i} + \boldsymbol{j} = \boldsymbol{k}$.
- **Goal:** find as large sum-free subset $\boldsymbol{S}$ as possible.

- **Examples:** A = {2, 3, 4, 5, 6, 8, 10}

  A = {1, 2, 3, 4, 5, 6, 8, 9, 10, 18}

Theorem

Every set $\boldsymbol{A}$ of $\boldsymbol{n}$ positive integers contains a sum-free subset of size greater than $\boldsymbol{n}/3$.

CS 537 Finding a large sum-free subset

A randomized algorithm

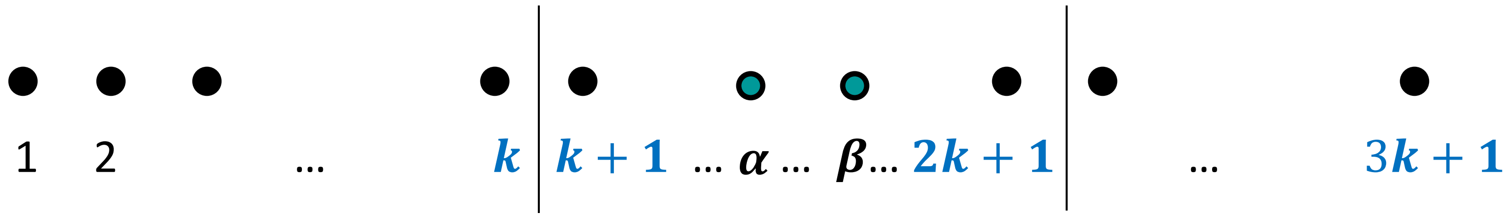
1. Let $p >$ max element of A be a prime, where $p = 3k + 2$.
   //The other choice, $3k + 1$, would also work.
2. Select a number q uniformly at random from $[p - 1]$.
3. Map each element $t \in A$ to tq **mod** p.
4. $S \leftarrow$ all elements of A that got mapped to $\{k + 1, \ldots, 2k + 1\}$.
5. Return S.

Need to prove:

- S is sum-free
- The expected number of elements from A that are mapped to $\{k + 1, \ldots, 2k + 1\}$ is $> n/3$.

Showing that S is sum-free

- Let i and j be any two elements in S.
- Say i is mapped to α; j is mapped to β; $\alpha, \beta \in [k+1, 2k+1]$



- Then $\alpha = iq \textbf{ mod } p$ and $\beta = jq \textbf{ mod } p$
- We need to show that $i + j$, if present in A, is not mapped to $[k+1, 2k+1]$.
- $i + j$ is mapped to $(\alpha + \beta) \textbf{ mod } p$

Argue that

- $(\alpha + \beta)$ must be greater than $2k + 1$.
- If $(\alpha + \beta) > p$, then $(\alpha + \beta) \textbf{ mod } p$ is at most k.

The expected size of S

1. Let $p >$ max element of A be a prime, where $p = 3k + 2$.
2. Select a number q uniformly at random from $[p - 1]$.
3. Map each element $t \in A$ to tq **mod** p.
4. $S \leftarrow$ all elements of A that got mapped to $\{k + 1, \dots, 2k + 1\}$.

Main idea: Every element $t \in A$ gets mapped to tq **mod** p, which is a uniformly random element of $\{1, \dots, 3k + 1\}$.

$$\Pr[t \text{ is selected to be in } S] = \frac{|\{k + 1, \dots, 2k + 1\}|}{|\{1, \dots, 3k + 1\}|} > 1/3$$

Example 3: Large sum-free subset

- Given a set $\boldsymbol{A}$ of positive integers, a sum-free subset $\boldsymbol{S} \subseteq \boldsymbol{A}$ contains no three elements $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k} \in \boldsymbol{S}$ satisfying $\boldsymbol{i} + \boldsymbol{j} = \boldsymbol{k}$.
- **Goal:** find as large as $\boldsymbol{S}$ as possible.

- **Examples:** A = {2, 3, 4, 5, 6, 8, 10}

  A = {1, 2, 3, 4, 5, 6, 8, 9, 10, 18}

Theorem

Every set $\boldsymbol{A}$ of $\boldsymbol{n}$ positive integers contains a sum-free subset of size greater than $\boldsymbol{n}/3$.