Randomness in Computing

LECTURE 20

CS
537

Last time

- Probabilistic method
  - The Counting Argument
  - The Expectation Argument
  - Derandomization using conditional expectations

Today

- Probabilistic method
  - Sample and Modify
  - The Second Moment Method

CS 537

The expectation argument

To prove that an object of required value exists:

1. Define a distribution on objects.
2. Sample an object from the distribution.
   - Compute the expected value of the sampled object.
3. Conclude that there exists an object with value equal to at least (at most) the expectation.

CS 537 Example: 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 $\boldsymbol{n}/3$.

Finding a large sum-free subset

Algorithm

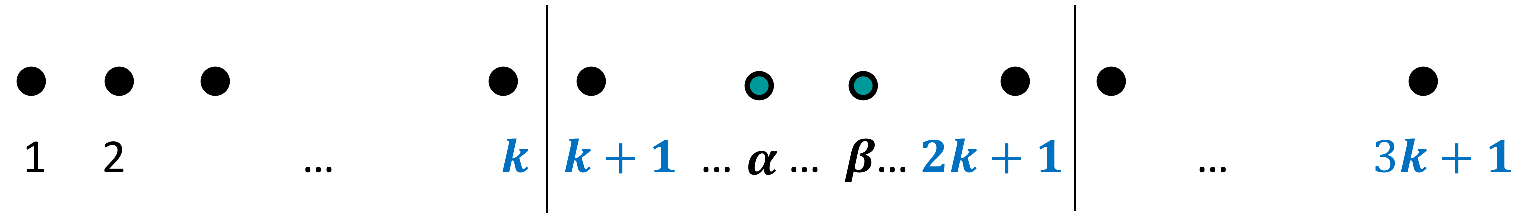
1. Let $p >$ max element of A be a prime, where $p = 3k + 2$.
   //The other choice, $3k + 1$, would also work.
2. Select a number q uniformly at random from $[p - 1]$.
3. Map each element $t \in A$ to tq **mod** p.
4. $S \leftarrow$ all elements of A that got mapped to $\{k + 1, \dots, 2k + 1\}$.
5.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

Need to prove:

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

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

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

Example: 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 $\boldsymbol{n}/3$.

Sample and Modify

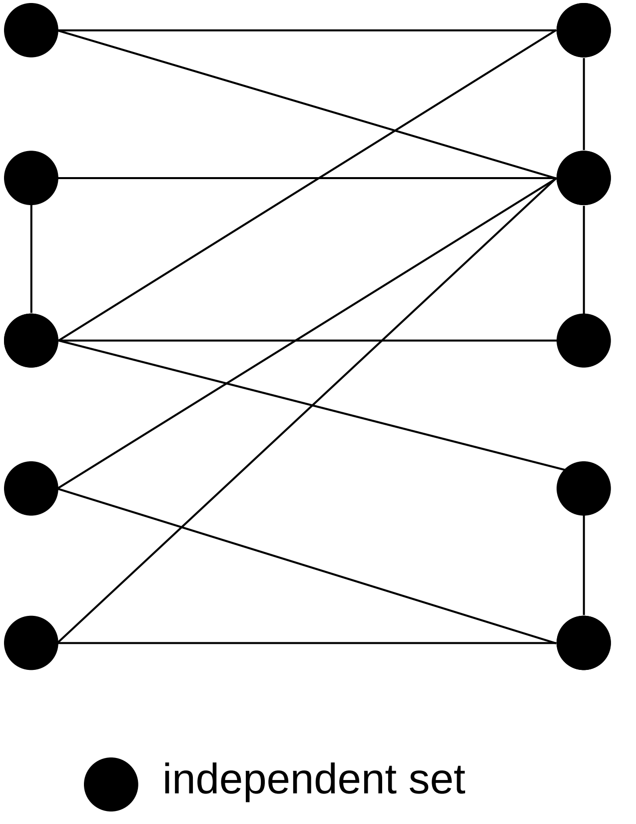
To prove that an object of required value exists:

1. Define a distribution on objects.
2. Sample an object from the distribution.
3. *Modify the sampled object.*
   - *Compute the expected value of the modified object.*
4. Conclude that there exists an object with value equal to at least (at most) the expectation.

Example: Finding an independent set

An **independent set** in an undirected graph G is
a set of nodes that includes at most one endpoint of every edge.

- What is the size of the largest independent set in this graph?



Finding a largest independent set in a given graph is NP-hard.

Example: a large independent set

Theorem

Let G be a connected graph with n nodes and m edges.

Then G has an independent set of size $\geq \frac{n^2}{4m}$.

Proof: Let $d = \frac{2m}{n}$ be the average degree in G.

Since G is connected, $d \geq 1$.

Algorithm

1. Delete each node in G (together with adjacent edges) independently w.p. $1 - 1/d$.
2. For each remaining edge: remove it and one (arbitrary) adjacent node.
3. Output remaining nodes.

Analysis: Algorithm returns an independent set.

Claim

The expected size of the returned set is $\geq \frac{n^2}{4m}$.

Example: a large independent set

Claim

The expected size of the returned set is $\geq \frac{n^2}{4m}$.

1. Delete each node w.p. $1 - 1/d$.
2. Remove each edge with one adjacent node.

Proof: Recall: $d = \frac{2m}{n}$ is the average degree in G.

- Let X = the number of nodes that remain after Step 1.

$$\mathbb{E}[X] = n \cdot \frac{1}{d}$$

- Let Y = the number of edges that remain after Step 1.

  An edge remains iff both of its endpoints remain, i.e. w.p. $1/d^2$.

$$\mathbb{E}[Y] = m \cdot \frac{1}{d^2} = \frac{nd}{2} \cdot \frac{1}{d^2} = \frac{n}{2d}$$

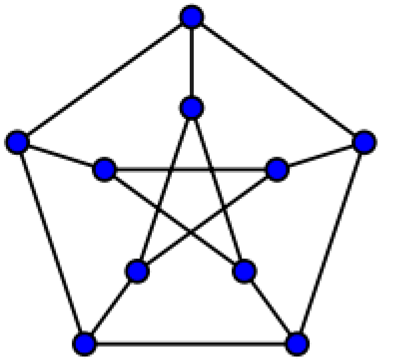
- Step 2 removes at most Y nodes.
- Let Z = the number of nodes in the output: $Z \geq X - Y$

$$\mathbb{E}[Z] \geq \mathbb{E}[X] - \mathbb{E}[Y] = \frac{n}{d} - \frac{n}{2d} = \frac{n}{2d} = \frac{n}{2} \cdot \frac{n}{2m} = \frac{n^2}{4m}$$

Example 2: Graphs with large girth

The **girth** of an undirected graph G is the length of the shortest cycle contained in G.

- What is the girth of this graph?



Ex. 2: Dense graphs with large girth

Theorem

$\forall$ integer $k \geq 3$, for sufficiently large n, there is a graph with n nodes, at least $\frac{n^{1+1/k}}{4}$ edges and girth at least k.

Proof:

Algorithm

1. Sample a graph $G \sim G_{n,p}$ with $p = n^{1/k-1}$.
2. Delete an (arbitrary) edge in G from each cycle of length $\leq k - 1$.
3. Return G.

Analysis: G has n nodes and girth at least k.

- Let $X =$ number of edges in the graph sampled in Step 1.

Question

1. Sample a graph $G \sim G_{n,p}$ with $p = n^{1/k-1}$.
2. Delete an edge from each cycle of length $\leq k-1$.

Let $X =$ number of edges in the graph sampled in Step 1.

What is the expectation of X?

A. np

B. $\binom{n}{2} p$

C. $n^2 p(1-p)$

D. None of the above.

Ex. 2: Dense graphs with large girth

Theorem

$\forall$ integer $k \geq 3$, for sufficiently large n, there is a graph with n nodes, at least $\frac{n^{1+1/k}}{4}$ edges and girth at least k.

Proof:

Algorithm

1. Sample a graph $G \sim G_{n,p}$ with $p = n^{1/k-1}$.
2. Delete an (arbitrary) edge in G from each cycle of length $\leq k-1$.
3. Return G.

Analysis: G has n nodes and girth at least k.

- Let $X =$ number of edges in the graph sampled in Step 1.

$$\mathbb{E}[X] = p \cdot \binom{n}{2} = n^{1/k-1} \cdot \frac{n(n-1)}{2} = \frac{1}{2} n^{1+1/k} \left(1 - \frac{1}{n}\right)$$
$$\geq \frac{1}{3} n^{1+1/k} \quad \text{for } n \geq 3$$

Ex. 2: Dense graphs with large girth

Claim

G has at least $\frac{n^{1+1/k}}{4}$ edges

1. Sample a graph $G \sim G_{n,p}$ with $p = n^{1/k-1}$.
2. Delete an edge from each cycle of length $\leq k-1$.

Proof: Recall: $\mathbb{E}[X] \geq \frac{1}{3} n^{1+1/k}$ for sufficiently large n.

- Let Y = the number of cycles of length $\leq k-1$ in the sampled graph.
- For $i \in [3, k-1]$, there are $\boxed{\binom{n}{i} \cdot \frac{(i-1)!}{2}}$ possible cycles of length i, each occurring w.p. $\boxed{p^i}$.

$$\mathbb{E}[Y] = \sum_{i=3}^{k-1} \binom{n}{i} \cdot \frac{(i-1)!}{2} \cdot p^i \leq \sum_{i=3}^{k-1} (np)^i = \sum_{i=3}^{k-1} (n^{1/k})^i < k \cdot n^{\frac{k-1}{k}}$$

$$\leq \frac{1}{12} n^{1+1/k} \text{ for sufficiently large } n$$

- Let Z = the number of edges remaining in G: $Z \geq X - Y$

$$\mathbb{E}[Z] \geq \mathbb{E}[X] - \mathbb{E}[Y] > \frac{1}{3} n^{1+1/k} - \frac{1}{12} n^{1+1/k} = \frac{n^{1+1/k}}{4}$$

The 2nd moment method

- Consider a R.V. X with $\mathbb{E}[X] > 0$.
- We want to give an upper bound on $\Pr[X = 0]$.
- By Chebyshev, for all $a > 0$,

$$\Pr[|X - \mathbb{E}[X]| \geq a] \leq \frac{\text{Var}\,[X]}{a^2}$$

$$\Pr[X = 0] \leq \Pr\big[|X - \mathbb{E}[X]| \geq \mathbb{E}[X]\big] \leq \frac{\text{Var}[X]}{(\mathbb{E}[X])^2}$$

Theorem

If X is a random variable with $\mathbb{E}[X] > 0$, then

$$\Pr[X = 0] \leq \frac{\text{Var}[X]}{(\mathbb{E}[X])^2}$$

Threshold behavior in random graphs $G \sim G(n,p)$

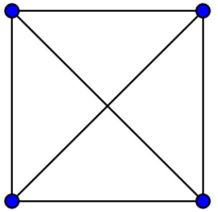
For many properties $\mathcal{P}$, there exists function $f(n)$ s.t.

1. when $p \ll f(n)$, probability that G has $\mathcal{P}$ $\rightarrow 0$ as $n \rightarrow \infty$
2. when $p \gg f(n)$, probability that G has $\mathcal{P}$ $\rightarrow 1$ as $n \rightarrow \infty$

(It holds for all nontrivial monotone properties.)

Review question

What is the expected number of k-cliques in $G \sim G_{n,p}$?



A. $\binom{n}{k} \cdot p^k$

B. $\binom{n}{k} \cdot p^{k(k-1)/2}$

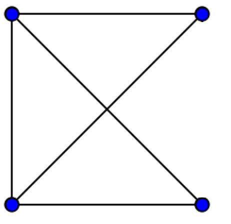
C. $\binom{k}{2}\binom{n}{k} \cdot p^k$

D. $n \cdot p^k(1-p)$

E. None of the above

Review question

What is the expected number of copies of this graph in $G \sim G_{n,p}$?



A. $\binom{n}{4} \cdot p^6$

B. $4\binom{n}{4} \cdot p^6$

C. $4\binom{n}{4} \cdot p^5(1-p)$

D. $6\binom{n}{4} \cdot p^5(1-p)$

E. None of the above

Example: having a 4-clique

Theorem

Let $G \sim G_{n,p}$ and $p^* = \Pr[G \text{ has a } K_4]$.

1. If $p \boxed{= \boldsymbol{o}\left(n^{-2/3}\right)}$ then $p^* \to 0$ as $n \to \infty$
2. If $p \boxed{= \boldsymbol{\omega}\left(n^{-2/3}\right)}$ then $p^* \to 1$ as $n \to \infty$

Proof: Let X = number of 4-cliques in G.

For every subset C of 4 nodes, let X_C be the indicator for C being a K_4.

$$\mathbb{E}[X] = \sum_C \mathbb{E}[X_C] = \binom{n}{4} \cdot p^6$$

1. $p = o(n^{-2/3})$

Markov

$$p^* = \Pr[X \geq 1] \leq \frac{\mathbb{E}[X]}{1} = \mathbb{E}[X]$$

$$\leq \frac{n^4}{4!} \cdot p^6 = \frac{n^4}{4!} \cdot o\left(n^{-(2/3)\cdot 6}\right) = \frac{n^4}{4!} \cdot o(n^{-4}) = o(1)$$

Example: having a 4-clique

Theorem

Let $G \sim G_{n,p}$ and $p^* = \Pr[G \text{ has a } K_4]$.

1. If p $\boxed{= \boldsymbol{o}(n^{-2/3})}$ then $p^* \to 0$ as $n \to \infty$
2. If p $\boxed{= \boldsymbol{\omega}(n^{-2/3})}$ then $p^* \to 1$ as $n \to \infty$

Proof: Expected number of 4-cliques: $\mathbb{E}[X] = \binom{n}{4} \cdot p^6$

2. $p = \omega(n^{-2/3})$

$\mathbb{E}[X] \to \infty$ as $n \to \infty$

Goal: Show $\mathrm{Var}[X] \ll (\mathbb{E}[X])^2$

$$\boxed{\begin{aligned}\boldsymbol{Cov}(Y,Z) &= \mathbb{E}[(Y-\mu_Y)\cdot(Z-\mu_Z)] \\ &= \mathbb{E}[YZ] - \mu_Y\mu_Z \le \mathbb{E}[YZ]\end{aligned}}$$

$$\mathrm{Var}[X] = \mathrm{Var}\left[\sum_C X_C\right] = \sum_C \mathrm{Var}[X_C] + \sum_{C \ne D} \mathrm{Cov}[X_C, X_D]$$

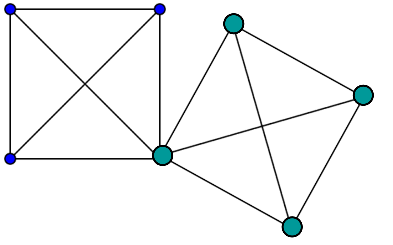
$$\mathrm{Var}[X_C] = \mathbb{E}[X_C^2] - (\mathbb{E}[X_C])^2 = \mathbb{E}[X_C] - (\mathbb{E}[X_C])^2 = p^6 - p^{12} \le p^6$$

$$\sum_C \mathrm{Var}[X_C] \le \binom{n}{4} \cdot p^6 = O(n^4 p^6)$$

Bounding the covariance

$\boxed{\boldsymbol{Cov(X_C, X_D) \leq \mathbb{E}[X_C \cdot X_D]}}$

Case 1: $|C \cap D|$ is 0 or 1

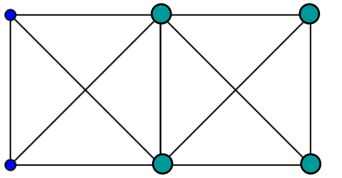


- Corresponding cliques do not share an edge.
- X_C and X_D are independent.
- $\boldsymbol{Cov(X_C, X_D) = 0}$

Bounding the covariance

$\boxed{\boldsymbol{Cov}(\boldsymbol{X_C}, \boldsymbol{X_D}) \leq \mathbb{E}[\boldsymbol{X_C} \cdot \boldsymbol{X_D}]}$

Case 2: $|C \cap D| = 2$

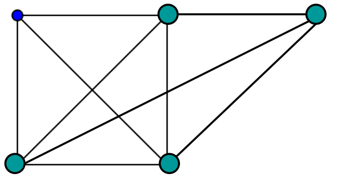


$$\boldsymbol{Cov}(\boldsymbol{X_C}, \boldsymbol{X_D}) \leq \mathbb{E}[\boldsymbol{X_C} \cdot \boldsymbol{X_D}]$$

Bounding the covariance

$$\boxed{\boldsymbol{Cov(X_C, X_D) \le \mathbb{E}[X_C \cdot X_D]}}$$

Case 3: $|C \cap D| = 3$



$$\boldsymbol{Cov(X_C, X_D) \le \mathbb{E}[X_C \cdot X_D]}$$

Putting it all together

Theorem

Let $G \sim G_{n,p}$ and $p^* = \Pr[G \text{ has a } K_4]$.

2. If p $\boxed{= \boldsymbol{\omega}(n^{-2/3})}$ then $p^* \to 1$ as $n \to \infty$

- $\mathrm{Var}[X] \leq \mathrm{Var}[\sum_C X_C] = \sum_C \mathrm{Var}[X_C] + \sum_{C \neq D} \mathrm{Cov}[X_C, X_D]$
  $= O(n^4p^6 + n^6p^{11} + n^5p^9)$
- $\Pr[X = 0] \leq \frac{\mathrm{Var}[X]}{(\mathbb{E}[X])^2} \quad = O\left(\frac{n^4p^6 + n^6p^{11} + n^5p^9}{n^8p^{12}}\right)$

$$= O\left(\frac{1}{n^4p^6} + \frac{1}{n^2p^6} + \frac{1}{n^3p^3}\right)$$

$$= o(1) \text{ for } p = \boldsymbol{\omega}(n^{-2/3})$$