CSE 521: Design and Analysis of Algorithms I **Fall 2021**

Lecture 2: Second Moment Method

Lecturer: Shayan Oveis Gharan *10/04/2022* *Scribe:*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

Consider a positive integer n and $p \in [0,1]$. Perhaps the simplest model of random (undirected) graphs is $G_{n,p}$. To sample a graph from $G_{n,p}$, we add every edge $\{u,v\}$ (for $u \neq v$ and $u,v \in \{1,\ldots,n\}$) independently with probability p.

For example, if X denotes the number of edges in a $G_{n,p}$ random graph, then we have

$$\mathbb{E}[X] = \binom{n}{2} \cdot p.$$

A 4-clique in a graph is a set of four nodes such that all $\binom{4}{2} = 6$ possible edges between the nodes are present. Let G be a random graph sampled according to $G_{n,p}$, and let $\mathcal{C}_4$ denote the event that G contains a 4-clique. It will turn out that if $p \gg n^{-2/3}$, then G contains a 4-clique with probability close to 1, while if $p \ll n^{-2/3}$, then $\mathbb{P}[\mathcal{C}_4]$ will be close to 0. Thus $p = n^{-2/3}$ is a "threshold" for the appearance of a 4-clique.

Remark 2.1. *Here we use the asymptotic notation $f(n) \gg g(n)$ to denote that $\lim_{n\to\infty} f(n)/g(n) \to \infty$. Similarly, we write $f(n) \ll g(n)$ to denote that $\lim_{n\to\infty} f(n)/g(n) \to 0$.*

We can use a simple first moment calculation for one side of our desired threshold behavior.

Lemma 2.2. *If $p \ll n^{-2/3}$ then $\mathbb{P}[\mathcal{C}_4] \to 0$ as $n \to \infty$.*

Proof. Let X denote the number of 4-cliques in $G \sim G_{n,p}$. We can write $X = \sum_S X_S$ where the set S runs over all $\binom{n}{4}$ subsets of four vertices in G, and X_S be the indicator random variable that there is a 4-clique on S. We have $\mathbb{P}[X_S = 1] = p^6$ since all 6 edges must be present and are independent, thus by linearity of expectation $\mathbb{E}[X] = p^6 \cdot \binom{n}{4}$. So if $p \ll n^{-2/3}$, then $\mathbb{E}[X] \to 0$ as $n \to \infty$. But now Markov's inequality implies that

$$\mathbb{P}[\mathcal{C}_4] = \mathbb{P}[X \geq 1] \leq \mathbb{E}[X] \to 0.$$

□

On the other hand, proving that $p \gg n^{-2/3} \Rightarrow \mathbb{P}[\mathcal{C}_4] \to 1$ is more delicate. Even though a first moment calculation implies that, in this case, $\mathbb{E}[X] \to \infty$, this is not enough to conclude that $\mathbb{P}[\mathcal{C}_4] \to 1$. For instance, it could be the case that with probability $1 - \frac{1}{n^2}$, we have no 4-cliques, but we see all $\binom{n}{4}$ many 4-cliques otherwise. In that case, $\mathbb{E}[X] = \Theta(n^2)$, but still the probability of seeing a 4-clique would be $\frac{1}{n^2}$ In other words, if the only thing we know about the random variable X is its expectation we cannot say it is non-zero with high probability. We need to know higher order moments of X.

2.1 Chebyshev's Inequality

Definition 2.3 (Variance). *The variance of a random variable X is defined as*

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}X)^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$

Theorem 2.4 (Chebyshev's Inequality)**.** *For any random variable X,*

$$\mathbb{P}\left[|X - \mathbb{E}X| > \epsilon\right] < \frac{\mathrm{Var}(X)}{\epsilon^2}$$

In the probabilistic method, the following statement is very handy.

Corollary 2.5. *For any random variable X,*

$$\mathbb{P}\left[X = 0\right] \leq \frac{\mathrm{Var}(X)}{(\mathbb{E}X)^2}$$

Proof. Let $\epsilon = \mathbb{E}X$ in the Chebyshev's inequality. Then,

$$\mathbb{P}\left[X = 0\right] \leq \mathbb{P}\left[|X - \mathbb{E}X| \geq \mathbb{E}X\right] \leq \frac{\mathrm{Var}(X)}{(\mathbb{E}X)^2}.$$

□

Lemma 2.6. *If X is a non-negative random variable, then*

$$\mathbb{P}\left[X > 0\right] \geq \frac{(\mathbb{E}\left[X\right])^2}{\mathbb{E}\left[X^2\right]}.$$

Proof. We use the Cauchy-Schwartz inequality: For any two random variables X, Y we can write

$$\mathbb{E}\left[X \cdot Y\right] \leq \sqrt{\mathbb{E}\left[X^2\right]} \cdot \sqrt{\mathbb{E}\left[Y^2\right]}.$$

Having this we write,

$$\mathbb{E}\left[X\right] = \mathbb{E}\left[X \mathbf{1}_{X>0}\right] \leq \sqrt{\mathbb{E}\left[X^2\right]}\sqrt{\mathbb{E}\left[\mathbf{1}_{X>0}\right]} = \sqrt{\mathbb{E}\left[X^2\right]}\sqrt{\mathbb{P}\left[X > 0\right]}.$$

□

For random variables X, Y let

$$\mathrm{Cov}(X, Y) = \mathbb{E}\left[XY\right] - \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right].$$

In particular, if X, Y is independent, then $\mathrm{Cov}(X, Y) = \mathbb{E}\left[XY\right]$.

Fact 2.7. *If $X = X_1 + \cdots + X_n$, then*

$$\mathrm{Var}(X) = \sum_i \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j).$$

In particular, if all X_i's are independent then $\mathrm{Var}(X) = \sum_i \mathrm{Var}(X_i)$.

Proof. First, observe

$$\mathrm{Var}(X) = \mathbb{E}(\sum_i X_i)^2 - \left(\mathbb{E}\sum_i X_i\right)^2$$

Expanding the terms and combining the terms corresponding to X_i, X_j gives the desired identity. □

Lemma 2.8. *If $p \gg n^{-2/3}$, then $\mathbb{P}\left[\mathcal{C}_4\right] \to 1$ as $n \to \infty$.*

Proof. Let X_S be the indicator random variable of having a clique on S and $X = \sum_S X_S$ as before. Using Corollary 2.5,

$$\mathbb{P}[\mathcal{C}_4] = \mathbb{P}[X > 0] \geq 1 - \frac{\text{Var}(X)}{(\mathbb{E}X)^2}$$

our goal is to show that $\text{Var}(X) \ll (\mathbb{E}X)^2$.

First, notice that for any S,

$$\text{Var}(X_S) = \mathbb{E}[X_S] - (\mathbb{E}[X_S])^2 \leq \mathbb{E}[X_S] = p^6.$$

So, $\sum_S \text{Var}(X_S) \leq \binom{n}{4} p^6$.

Now, fix two sets $S, T \in \binom{n}{4}$. Obviously if $|S \cap T| \leq 1$, then S, T do not share any "potential" edges. So, by independence of edges $\mathbb{P}[X_S X_T] = \mathbb{P}[X_S]\,\mathbb{P}[X_T] = p^{12}$.

On the other hand, if $|S \cap T| = 2$. Then,

$$\mathbb{P}[X_S X_T] = \mathbb{P}[X_S]\,\mathbb{P}[X_T | X_S] = p^6 \mathbb{P}[X_T | X_S] = p^{11}.$$

The last identity is because since X_S occurs we know that there is an edge in the common pair. So, we only need 5 more edges to get X_T. Similarly, if $|S \cap T| = 3$, then $\mathbb{P}[X_S X_T] = p^9$. In summary,

$$\mathbb{P}[X_S X_T] = \begin{cases} \mathbb{P}[X_S]\,\mathbb{P}[X_T] & \text{if } |S \cap T| \leq 1 \\ p^{11} & \text{if } |S \cap T| = 2 \\ p^9 & \text{if } |S \cap T| = 3. \end{cases}$$

It follows that

$$\begin{aligned} \sum_{S \neq T} \text{Cov}(X_S, X_T) &= \sum_S \left(\sum_{T:|T \cap S| = 2} \text{Cov}(X_S, X_T) + \sum_{T:|T \cap S| = 3} \text{Cov}(X_S, X_T) \right) \\ &= \sum_S \left(6 \binom{n-4}{2} (p^{11} - p^{12}) + 4 \binom{n-4}{1} (p^9 - p^{12}) \right) \\ &\leq \binom{n}{4} (3n^2 p^{11} + 4np^9) \end{aligned}$$

Lastly,

$$\mathbb{P}[X = 0] \leq \frac{\text{Var}(X)}{(\mathbb{E}X)^2} \leq \frac{\binom{n}{4} p^6 + \binom{n}{4} (3n^2 p^{11} + 3np^9))}{(\binom{n}{4} p^6)^2} \leq \frac{1 + 3n^2 p^5 + 4np^3}{\binom{n}{4} p^6}$$

Observe that for $p \gg n^{-2/3}$ the ratio goes to infinity as $n \to \infty$ and $p \to \infty$ (at a slower rate). $\square$