Lecture 3: The second moment method CSE 525, Fall 2016
Instructor: James R. Lee

1 Threshold phenomena in random graphs

Consider a positive integer n and value $p \in [0,1]$. Perhaps the simplest model of random (undirected) graphs is $\mathcal{G}_{n,p}$. To sample a graph from $\mathcal{G}_{n,p}$, we add every edge $\{i,j\}$ (for $i \neq j$ and $i,j \in \{1,\ldots,n\}$) independently with probability p. For example, if X denotes the number of edges in a $\mathcal{G}_{n,p}$ graph, then $\mathbb{E}\, X = p\binom{n}{2}$.

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 a random graph sampled according to $\mathcal{G}_{n,p}$, and let 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 $\mathbb{P}[C_4]$ will be close to 0. Thus $p = n^{-2/3}$ is a "threshold" for the appearance of a 4-clique.

Remark 1.1. Here we use the asymptotic notation $f(n) \gg g(n)$ to denote that $\lim_{n\to\infty} f(n)/g(n) = \infty$. Similarly, $f(n) \ll g(n)$ means that $\lim_{n\to\infty} f(n)/g(n) = 0$.

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

Lemma 1.2. *If $p \ll n^{-2/3}$ then $\mathbb{P}[C_4] \to 0$ as $n \to \infty$.*

Proof. Let X denote the number of 4-cliques in $G \sim \mathcal{G}_{n,p}$. 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 $X_S = 1$ if there is a 4-clique on S, and $X_S = 0$ otherwise.

We have $\mathbb{P}[X_S = 1] = p^6$ since all 6 edges must be present, thus by linearity of expectation $\mathbb{E}\, X = p^6\binom{n}{4}$. So if $p \ll n^{-2/3}$, then $\mathbb{E}\, X \to 0$ as $n \to \infty$. But now Markov's inequality implies that

$$\mathbb{P}[C_4] = \mathbb{P}[X \geqslant 1] \leqslant \mathbb{E}\, X \to 0\,. \qquad \square$$

On the other hand, proving that $p \gg n^{-2/3} \implies \mathbb{P}[C_4] \to 1$ 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 $\mathbb{P}[C_4] \to 1$. For instance, it could be the case that with probability $1 - \frac{1}{n^2}$, we have no 4-cliques, but with probability $\frac{1}{n^2}$ we see all possible $\binom{n}{4}$ 4-cliques. In that case, $\mathbb{E}\, X \asymp n^2$, but still the probability of seeing a 4-clique would be $\frac{1}{n^2}$.

We need to exploit the fact that the appearances of distinct 4-cliques are mostly independent events. Certainly if S and S' are two disjoint sets of vertices, then the corresponding random variables X_S and $X_{S'}$ are independent.

1.1 Chebyshev's inequality and second moments

First, we recall the notion of the *variance* of a real-valued random variable X: $\mathrm{Var}(X) = \mathbb{E}\,(X - \mathbb{E}[X])^2$. If we know something about the variance, we can improve upon Markov's inequality.

Lemma 1.3 (Chebyshev inequality)**.** *If X is a real-valued random variable with $\mu = \mathbb{E}\, X$, then for every $\alpha > 0$,*

$$\mathbb{P}\left[|X - \mu| \geqslant \alpha\right] \leqslant \frac{\mathrm{Var}(X)}{\alpha^2}\,.$$

Proof. Simply apply Markov's inequality to the nonnegative random variable $(X - \mu)^2$. $\square$

One consequence of the Chebyshev inequality is the following simple fact.

Corollary 1.4. *If X is a real-valued random variable, then*

$$\mathbb{P}(X = 0) \leqslant \frac{\mathrm{Var}(X)}{(\mathbb{E}\, X)^2}\,.$$

Proof. Using the Chebyshev inequality, we have:

$$\mathbb{P}(X = 0) \leqslant \mathbb{P}\left(|X - \mathbb{E}\, X| \geqslant \mathbb{E}\, X\right) \leqslant \frac{\mathrm{Var}(X)}{(\mathbb{E}\, X)^2}\,. \qquad \square$$

Now we are in position to analyze the other side of the threshold. It will help to have the following definition: For two random variables X and Y, we define their *covariance* by $\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$. Note that if X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$.

Lemma 1.5. *If $X = X_1 + \cdots + X_n$, then*

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i=1}^{n} \sum_{j \neq i} \mathrm{Cov}(X_i, X_j)\,.$$

Proof. One observe first that for any random variable X, we have $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}\, X)^2$, hence

$$\mathrm{Var}(X) = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right)^2 - \left(\mathbb{E} \sum_{i=1}^{n} X_i\right)^2\,.$$

Expanding the squares and collecting terms yields the desired result. $\square$

Lemma 1.6. *If $p \gg n^{-2/3}$, then $\mathbb{P}[C_4] \to 1$ as $n \to \infty$*

Proof. As before, let $X = \sum_S X_S$ denote the number of 4-cliques in $G \sim \mathcal{G}_{n,p}$. Our goal is to show that $\mathbb{P}[X = 0] \to 0$ as $n \to \infty$. Using Corollary 1.4, it suffices to show that $\mathrm{Var}(X) \ll (\mathbb{E}\, X)^2$. The main task will be evaluating $\mathrm{Var}(X)$.

Write

$$\mathrm{Var}(X) = \sum_S \mathrm{Var}(X_S) + \sum_S \sum_{T \neq S} \mathrm{Cov}(X_S, X_T)\,.$$

First, observe that since X_S is a $\{0, 1\}$ random variable, we have $\mathrm{Var}(X_S) \leqslant \mathbb{E}[X_S^2] = \mathbb{E}[X_S]$, yielding

$$\mathrm{Var}(X) \leqslant \mathbb{E}[X] + \sum_S \sum_{T \neq S} \mathrm{Cov}(X_S, X_T)\,. \tag{1.1}$$

Now we evaluate the second sum. The value $\mathrm{Cov}(X_S, X_T)$ depends on $|S \cap T|$. In particular, if $|S \cap T| \leqslant 1$, then $\mathrm{Cov}(X_S, X_T) = 0$ since S and T share no possible edges, hence the events X_S and X_T are independent.

Next, note that $\mathrm{Cov}(X_S, X_T) \leqslant \mathbb{E}[X_S X_T]$, and the latter event is that we have a 4-clique on *both* S and T. Thus if $|S \cap T| = 2$, then $\mathrm{Cov}(X_S, X_T) \leqslant p^{11}$ (since for $X_S = X_T = 1$ to happen, we need 11 edges to be present). Similarly, if $|S \cap T| = 3$, then $\mathrm{Cov}(X_S, X_T) \leqslant p^9$. So now we are simply left to count the possibilities:

$$\sum_S \sum_{T \neq S} \mathrm{Cov}(X_S, X_T) \leqslant O(n^6)p^{11} + O(n^5)p^9\,.$$

(Make sure you understand why this line is true!) From (1.1), we conclude that

$$\mathrm{Var}(X) \leqslant O(n^4)p^6 + O(n^6)p^{11} + O(n^5)p^9\,.$$

Also, recall that $\mathbb{E}\, X = \binom{n}{4}p^6 = \Theta(n^4 p^6)$, so $(\mathbb{E}\, X)^2 = \Theta(n^8 p^{12})$. In particular, if $p \ll n^{-2/3}$, then $\mathrm{Var}(X) \ll (\mathbb{E}\, X)^2$. Combined with Corollary 1.4, this yields the desired result. $\square$

2 Unbiased estimators

Suppose we want to estimate the area of a unit disk in the plane. One way to accomplish this is via a *Monte Carlo* algorithm: We sample a uniformly random point in $x \in [0,1]^2$ in the unit square, and then check whether $x_1^2 + x_2^2 \leqslant 1$. Let X be the indicator of the event that x is in the unit circle. Then

$$\mathbb{E}[X] = \frac{\text{area(disk)}}{\text{area}([0,1]^2)} = \pi\,.$$

So the random variable X is an unbiased estimator for π.

There are many situations in which one wants to estimate the probability of some event, but doing so directly is difficult. In those cases, having an unbiased estimator is quite helpful. Crucially, the usefulness of such an estimator depends on its variance. First, let's recall the following simple fact (you should be able to prove it easily using Lemma 1.5).

Fact 2.1. *If $X_1, X_2, \ldots, X_n$ are independent, then*

$$\mathrm{Var}(X_1 + X_2 + \cdots + X_n) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n)\,.$$

We recall that a family of random variables is said to be "i.i.d." if they are "identical and independently distributed," i.e. they are independent samples from the same distribution.

Theorem 2.2. *For every $\varepsilon > 0$, the following holds. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$. Let $X = \frac{X_1+X_2+\cdots+X_n}{n}$ be their* empirical mean. *If $n \geqslant \frac{4}{\varepsilon^2}\frac{\sigma^2}{\mu^2}$, then*

$$\mathbb{P}\left[|X - \mu| \geqslant \varepsilon\mu\right] \leqslant \frac{1}{4}\,.$$

Proof. By linearity, $\mathbb{E}[X] = \mu$ and by independence and Fact 2.1, $\mathrm{Var}(X) = \sigma^2/n$. So Chebyshev's inequality implies that

$$\mathbb{P}\left[|X - \mu| \geqslant \varepsilon\mu\right] \leqslant \frac{\sigma^2}{n\mu^2\varepsilon^2} \leqslant \frac{1}{4}\,,$$

where the final bound uses our assumption on n. □

Observe that if each X_i is a $\{0,1\}$ random variable, then $\mathrm{Var}(X_i) = \mu - \mu^2$, so $\sigma^2 \leqslant \mu$. In this case, the required number of samples simplifies to $n \geqslant \frac{4}{\varepsilon^2} \cdot \frac{1}{\mu}$. This represents the intuitive fact that if we are trying to estimate probability of a very rare event (so that μ is very small), then we will need many samples to get a decent estimate.

The median trick. Although we stopped at $\frac{1}{4}$, we can improve the probability of a poor estimate with a small additional expense. Suppose that we do N trials of the above experiment (requiring $N \cdot n$ samples overall). Let X' be the median value of these N trials. Then as long as $N > 2\log_{4/3}\frac{1}{\delta}$, we will have $\mathbb{P}[|X' - \mu| \geqslant \varepsilon\mu] \leqslant \delta$.

To see this, let $Y_1, \ldots, Y_N$ be indicators that are equal to 1 if the ith trial gives an empirical mean in the range $[\mu(1-\varepsilon), \mu(1+\varepsilon)]$. We know from the preceding lemma that $\mathbb{P}[Y_i = 1] \geqslant 3/4$, and the variables $\{Y_i\}$ are independent. The only way that $X' \notin [\mu(1-\varepsilon), \mu(1+\varepsilon)]$ is if at least half the trials end negatively, i.e. $Y_1 + \cdots + Y_N \leqslant N/2$.

Claim 2.3. If we perform $2k+1$ coin flips and $\mathbb{P}[\text{heads}] \geqslant 3/4$, then $\mathbb{P}[\leqslant k \text{ flips are heads}] \leqslant (3/4)^k$.

One can prove this directly by counting and estimating some binomial coefficients. In Lecture 5, we will see how to prove this and related results in a more general way, and that will start our foray into the *concentration of measure* phenomenon.