10/16: Random Graphs, II

We have two goals for this lecture:

1. To examine threshold phenomena in random graphs. These are properties of graphs that have a critical probability p0. They are unlikely to hold for p < p0, and are likely to hold for p > p0.
2. To learn how to use Chebychev's inequality (and variance bounds) to prove that it is likely that something appears.

In contrast, last lecture we used Markov's inequality to show that it is unlikely that something appears in a graph. That technique is insufficient to handle the reverse: showing that it is likely that something appears.

The two phenomena we will examine are:

1. Whether a graph has a vertex with no attached edges, and
2. Whether a graph has a 4-clique.

We will derive critical probabilities for each.

In the last lecture, we proved that certain structures were unlikely to appear in graphs by defining a random variables $X_1, \ldots, X_m$ that are each one if a certain structure appears, and zero otherwise. We then set X to be the sum of the X_is, and proved that the expectation of X was small. We observed by Markov's inequality that

$$\Pr[X \geq 1] \leq E[X]$$

But, when we want to show that X is unlikely to be zero, this technique does not suffice. Even if we show that E[X] is big, we have not established that X is unlikely to be zero. For example, it could be the case that X is 20000 with probability 1/100 and zero with probability 99/100. So, the expectation of X is 200, but it is usually just 0.

Variance

Recall the definition of the variance of a random variable X:

$$\mathrm{Var}[X] = E[(X - E[X])^2] \qquad (1)$$

We will use to other forms of the definition of the variance. The first is

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

this can be derived from definition 1by exploiting linearity of expectation:

$$= E[X^2] - 2E[X\,E[X]] + E[X]^2$$

$$= E[X^2] - 2E[X]^2 + E[X]^2$$

$$= E[X^2] - 2E[X]^2 + E[X]^2$$

$$= E[X^2] - E[X]^2$$

The other expression we will use applies when X is the sum of many variables.

If $X = \sum_i X_i$, then

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

where $\mathrm{Cov}(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j]$

Note that if two variables are independent and then their covariance is zero

Here is a derivation of this expression for the variance.

$$E[X^2] = E\left[\left(\sum_i X_i\right)^2\right] = E\left[\sum_{i,j} X_i X_j\right]$$

$$= \sum_{i,j} E[X_i X_j] = \sum_i E[X_i^2] + \sum_{i \neq j} E[X_i X_j]$$

and

$$E[X]^2 = \sum_i E[X_i]^2 + \sum_{i \neq j} E[X_i]E[X_j]$$

So, $E[X^2] - E[X]^2$

$$= \sum_i \left(E[X_i^2] - E[X_i]^2\right) + \sum_{i \neq j}\left(E[X_i X_j] - E[X_i]E[X_j]\right)$$

$$= \sum_i \mathrm{Var}[X_i] + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j), \qquad (3)$$

Chebyshev's Inequality

We will use Chebyshev's inequality.

Thm

$$\Pr[|X - E[X]| > \lambda] \leq \frac{Var(X)}{\lambda^2}$$

Proof

$$Var[X] = E[(X - E[X])^2] \geq \lambda^2 \cdot \Pr[|X - E[X]| \geq \lambda]$$

Now, divide both sides by λ^2.

We will use this with $\lambda = E[X]$, which gives

$$\Pr[X \leq 0] \leq \frac{Var[X]}{E[X]^2}$$

So, to prove that a variable is unlikely to be zero, it suffices to prove that its variance is much less than the square of its expectation.

Now let's examine the probability that a graph chosen from $G(n,p)$ has an isolated vertex. (having no edges) We will show that for

is

$p = \frac{(1+\varepsilon)\ln n}{n}$ it is unlikely there is an isolated vertex, and for $p = \frac{(1-\varepsilon)\ln n}{n}$

there probably is an isolated vertex.

Let A_i be the event that vertex i is isolated,

$$A = \text{OR}_i\, A_i$$

$$X_i = \begin{cases} 1 & \text{if } A_i \\ 0 & \text{o.w.} \end{cases}$$

$$X = \sum_{i=1}^{n} X_i$$

Then $\Pr[A_i] = (1-p)^{n-1}$

Then, $\Pr[A_i] = (1-p)^{n-1}$

As we will frequently encounter expressions of this form, let we recall that

$$(1-p)^p e^{-p} \le (1-p) \le e^{-p}$$

So, $E[X_i] \le e^{-p(n-1)}$, and $E[X] \le n e^{-p(n-1)}$

If we substitute $p = \frac{(1+\varepsilon)\ln n}{n-1}$, this gives

$$E[X] \le n e^{-p(n-1)} = n e^{-(1+\varepsilon)\ln n} = n \cdot n^{-(1+\varepsilon)} = n^{-\varepsilon}$$

So, $\Pr[X \ge 1] \le n^{-\varepsilon}$

On the other hand, if $p = \frac{(1-\varepsilon)\ln n}{n-1}$

$$E[X_i] = (1-p)^{(n-1)} \ge (1-p)^{p(n-1)} e^{-p(n-1)}$$

$$= (1-p)^{(1-\varepsilon)\ln n} e^{-(1-\varepsilon)\ln n} \ge (1-p)^{\ln n} n^{-(1-\varepsilon)}$$

$$\ge \left(1 - \frac{(1-\varepsilon)\ln^2 n}{n-1}\right) \frac{1}{n^{1-\varepsilon}}$$

(by the inequality $(1-\alpha)^k \ge 1-k\alpha$)

(happens if n sufficiently large)

Assuming $\frac{\ln^2 n}{n-1} \le \frac{1}{2}$, we get

$$\ge \frac{1}{2n^{1-\varepsilon}}$$

So, $E[X] = \sum_{i=1}^{n} E[X_i] \ge n \frac{1}{2n^{1-\varepsilon}} = \frac{n^\varepsilon}{2}$.

So, $E[X] = \sum_{i=1}^{n} E[X_i] \geq n \frac{1}{2n^{1+\varepsilon}} = \frac{n^{\varepsilon}}{2}$.

But, this is not enough to conclude $\Pr[X>0]$ is non-negligible!

So, let's compute $Var(X)$.

We will show $Var(X) \leq 2n^{\varepsilon}$.

This will imply

$$\Pr[X=0] \leq \frac{Var[X]}{E[X]^2} \leq \frac{2n^{\varepsilon}}{(n^{\varepsilon}/2)^2}$$

$$= \frac{8}{n^{\varepsilon}} \to 0 \quad \text{as } n \to \infty$$

Now, let's calculate the variance using expression (3)

In our case, X_i only takes the values 0 or 1, so $X_i = X_i^2$, and

$E[X_i^2] = E[X_i] = (1-p)^{(n-1)}$

So, $Var[X_i] = E[X_i^2] - E[X_i]^2 \leq E[X_i^2]$

$$= E[X_i] = (1-p)^{(n-1)} \leq e^{-p(n-1)} = \frac{1}{n^{1+\varepsilon}}$$

On the other hand,

$\mathrm{Cov}(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j]$

now, $E[X_i X_j] = \Pr[A_i \wedge A_j]$, and $A_i \wedge A_j$

happens only if none of the $2(n-2)+1 = 2n-3$

possible edges attached to i and j appear,

So $E[X_i X_j] = (1-p)^{2n-3}$

As $E[X_i] = (1-p)^{n-1}$,

$$\mathrm{Cov}(X_i X_j) = (1-p)^{2n-3} - (1-p)^{2(n-1)}$$

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

$$= \frac{p}{1-p}\left(\frac{n^{\varepsilon}}{n}\right)^2$$

$$= \frac{(1-\varepsilon)\ln n/(n-1)}{1-\frac{(1-\varepsilon)\ln n}{n-1}}\left(\frac{n^{\varepsilon}}{n}\right)^2$$

$$\leq \frac{\ln(n)}{n-1}\frac{1}{n^{2-2\varepsilon}}$$

So, $\mathrm{Var}[X] \leq \sum_i \mathrm{Var}[X_i] + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$

$$\leq n\frac{1}{n^{1-\varepsilon}} + n(n-1)\frac{\ln(n)}{n-1}\frac{1}{n^{2-2\varepsilon}}$$

$$= n^{\varepsilon} + \frac{\ln(n)\cdot n^{2\varepsilon}}{n} \leq 2n^{\varepsilon}$$ for n sufficiently large.

Now, let's do a more interesting example.

We will show that the property of containing

We will show that the property of containing a 4-clique has a threshold at $n^{-2/3}$.

First, for each set $|S|=4$, let A_S denote the event that the graph contains a clique on the vertices in S.

Let $X_S = \begin{cases} 1 & \text{if } A_S \\ 0 & \text{o.w.} \end{cases}$ $\qquad X = \sum_{|S|=4} X_S$

We first compute $E[X] = \sum_{|S|=4} E[X_S]$

$$= \sum_{|S|=4} \Pr[A_S] = \binom{n}{4} p^6$$

So, if $p = cn^{-2/3}$, $E[X] \le \frac{n^4}{24} c^6 (n^{-2/3})^6$

$$= \frac{c^6}{24} \to 0 \text{ as } c \to 0$$

For example, if $c = \frac{1}{\ln(n)}$ this goes to zero.

Now, let's consider the case in which $c \to \infty$.

For this case, we need to compute $\mathrm{Var}[X]$.

We upper bound $\mathrm{Var}[X]$ by

$$\mathrm{Var}[X] \le E[X] + \sum_{S \ne T} \mathrm{Cov}(X_S X_T)$$

To bound $\mathrm{Cov}(X_S, X_T)$, we recall that

$\mathrm{Cov}(X_S, X_T) = 0$ if X_S and X_T are

independent, which happens if $|S \cap T| = 0$
or $|S \cap T| = 1$

In the other cases, we apply $Cov(X_S, X_T) \le Pr[A_S \wedge A_T]$

So, for $|S \cap T| = 2$, $Cov(X_S, X_T) \le p^{11}$,

as 11 edges have to appear for both S and T
to be cliques.

For $|S \cap T| = 3$, $Cov(X_S, X_T) \le p^9$.

The number of pairs S, T for which $|S \cap T| = 2$

$$\le \binom{n}{2}\binom{n-2}{2}\binom{n-4}{2} \le n^6$$

and for $|S \cap T| = 3$

$$\le \binom{n}{3}\binom{n-3}{1}\binom{n-4}{1} \le n^5$$

So, $\sum_{S \neq T} Cov(X_S, X_T) \le n^6 p^{11} + n^5 p^9$

$$= c^{11} n^{\left(6 - \frac{2 \cdot 11}{3}\right)} + c^9 n^{\left(5 - \frac{9 \cdot 2}{3}\right)}$$

$$= c^{11} n^{-4/3} + c^9 n^{-1},$$

so

$$Var[X] \le \frac{c^6}{24} + c^{11} n^{-4/3} + c^9 n^{-1} \le \frac{c^6}{12},$$

for n sufficiently large.

We also need to know

$$E[X] = \binom{n}{4} p^6 \ge \frac{n^4}{25} p^6 \quad \text{for } n \text{ sufficiently large}$$

$$= \frac{c^6}{25}$$

25

So, for n sufficiently large,

$$\Pr[\omega(G) < 4] = \Pr[X=0] \le \frac{Var[X]}{E[X]^2} \le \frac{c^6/12}{(c^6/25)^2}$$

$$= \frac{25^2}{12 \cdot c^6} \to 0 \text{ as } c \to \infty,$$

For example, if $c = \ln(n)$