1 Intro

Today in class we will discuss random graph models and doing statistics on such graphs.

2 Statistics Review

Weak Law of Large Number Consider $x_i|_1^n$ i.i.d. random variables with $E[x_i] = \mu$ then for all $\epsilon > 0$ $P(\ |\sum x_i - n\mu\ | > \epsilon\mu\ ) \to 0$ as $n \to \infty$ Also known as asymptotically almost surely or a.a.s.

Strong Law $P(\lim | \text{mean}(x) - \mu| > 0\ )$

Sterling Approximation For any natural number n, $\sqrt{2\pi n}(n/e)^n \leq n! \leq e\sqrt{n}(n/e)^n$ Therefore $\binom{n}{k} \leq (en/k)^k$

Exponential: $\lim_{n\to\infty} (1 - x/n)^n = x^n$

2.1 Linearity of Expectation

A important useful fact to know is that if we have a set of n random variables, X_i then if we define a new variable $Y = \sum_{i=1}^n X_i$ then linearity of expectation says that $E[Y] = E[\sum X_i] = \sum E[X_i]$. In practice, we will often have the $E[X_1] = E[X_2] = \cdots = E[X_n]$ in which case $E[Y] = nE[X_1]$. Also note this is still true even if the random variables X_i are not independent of each other.

A common pattern applied often is to find the expectation of a random variable Y, let X_i be the indicator of some event i.e. $X_i = \begin{cases} 1 & \text{If event i happens} \\ 0 & \text{else} \end{cases}$.

In such a case where $Y = \sum X_i$ we would have $E[Y] = \sum E[X_i] = \sum 1 * Pr[X_i]$.

2.2 Common Distributions

This is just a table of some common distributions and the basic facts about them. This is hopefully mostly reference.

Distribution	definition	E[X]]	Var[X]
Bernoulli	$\Pr[X = 1] = p$, $\Pr[X=0]=1-p$	p	p(1-p)
Binomial	$\Pr[X=k] = \binom{n}{k}p^k(1-p)^{n-k}$	np	np(1-p)
Poisson	$Pr[X = x]\frac{\{^{-\lambda}}{e}\lambda^x\}\{x!\}$	λ	λ
Normal	hi	μ	σ^2

$Pr[X = x] = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma}\left[\frac{-1}{2}\frac{x-\mu}{\sigma}^2\right]$ This is the simplest model where we have a random variable X_i and $Pr[X_i = 1] = p$ and $Pr[X_i = 0] = 1 - p$.

3 Erdos-Renyi Random Graphs

The first graph model we introduce is also the most intuitive for producing a random undirected graph. The Erdos-Renyi random graph model takes two parameters, the number of nodes and probability that there is an edge between two vertices This probability is uniform across every pair of nodes. In the following we will let G(n,p) denote an Erdos-Renyi(ER for the rest of these notes) graph with n nodes and link probability p.

To begin saying interesting things about these graphs we will define the Bernoulli random variables $\ell_{ij} = \begin{cases} 1 & \text{i is connected to j} \\ 0 & \text{Otherwise} \end{cases}$ for all $i\neq$ j, i,j $\in$ [n]. Notice that because the link probability is uniform the set of all ℓ_{ij} variables are independent and specifically $Pr[\ell_{ij} = 1] = p$.

3.1 Number of Edges in ER graphs

This leads to the following expression the expected number of edges.

$$\mathbb{E}[\#Edges] = \mathbb{E}[\sum_{\substack{i,j\in[n] \\ i\neq j}} l_{ij}] \tag{1}$$

$$= \sum_{\substack{i,j\in[n] \\ i\neq j}} \mathbb{E}[l_{ij}] \tag{2}$$

$$= \binom{n}{2}p \tag{3}$$

$$= \frac{n(n-1)}{2}p \tag{4}$$

Using the weak law of large numbers, we have that for any $\alpha > 0$ we have

$$Pr[\left|\sum_{i\neq j} l_{ij} - p\frac{n(n-1)}{2}\right| > \alpha p\frac{n(n-1)}{2}] \to 0$$

as $n \to \infty$.

The implication of this is that for large n the odds we have a random graph that doesn't have close to $\binom{n}{2}p$ goes to 0.

3.2   Degrees of ER Graphs

Now we will figure out what the expected degree of any node v in an ER graph is. Let D be a random variable representing the degree of v.

$$\mathbb{E}[D] = \mathbb{E}[\sum_{j \neq v} l_{vj}] \tag{5}$$

$$= \sum_{j \neq v} \mathbb{E}[l_{vj}] \tag{6}$$

$$= (n - 1) * p \tag{7}$$

And the probability D=d is the well known probability distribution of a binomial random variable. That is $Pr[D = d] = \binom{n-1}{d}p^d(1 - p)^{n-1-p}$ where we use n-1 instead of n because there are n-1 nodes that could match with our target node since self loops aren't permitted.

If we fixed the expected degree with $\lambda = (n-1)p$ then we can approximate this has $Pr[D = d] = \frac{e^\lambda \lambda^d}{d!}$

3.3   Individual clustering coefficient

Clustering coefficient of an ER graph is $Cl_i(p) = p$ **I don't know how get this with what was in class**

4   Transition functions

A thing we will commonly look at in this class is how a property of chosen graph family changes as the graph becomes bigger and the link probablity changes with it. Usually the probablity will of links decreases with n as we are frequently interested in properties of sparse graphs.

Instead of doing the challenging working of fully analzying the probablity of a given property for a fixed n, it will typically be easier to asymptomatic analysis on what happens as $n \to \infty$. We will often see that the probablity of an event goes to 0 if the link probablity p(n) shrinks strictly faster than some speicial threshold function and to 1 if p(n) shrinks slower. We call such behavior a phase transistion.

4.1   Example

The function $f(n) = \frac{log(n)}{n}$ clearly goes to zero as $n \to \infty$. It can be shown that if $p(n)/f(n) \to 0$ with n then odds the graph is connected goes to zero

but if $p(n)/f(n) \to \infty$ then the graph will be connected with incredibly high probablity.

1. Define A = number of edges > 0 Threshold of emergence of the first edge E[#Edges] = C(n,2) * p $\approx n^2/2$ p(n) So is p*$n^2/2 \to 0$ then E[number of edges] $\to 0$ so P(number of edges) $\to 0$

   If instead p(n)$n^2/2$ then E[number of edge] $\to \infty$ **Remark** Does not imply that P(number of edges > 0) $\to 1$

2. $1/n^2$ is the threshold function for the emergence of the first link.

3. E[number of triples] $= n^3$ p^2 using a similar analysis we can show that t(n) = n$^{-3/2}$ is a threshold function

4. Threshold for observing a cycle with k nodes is t(n) = 1/n

5. Also at 1/n the largest component has O(log(n)) nodes

6. At p(n)= log(n)/n threshold the network becomes connected