Mathematical Tools for CS - Lecture 2

February 20, 2024

Today's Plan

- See some tools/examples for the analysis of probabilities and random variables.
- Threshold Phenomena
- ..

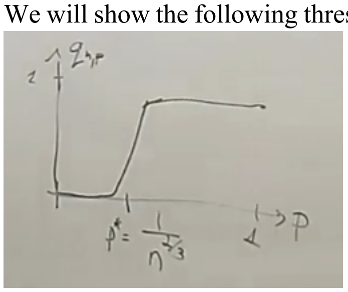
1 Copies of k_4 in $G(n,p)$

Let $G \sim G(n,p)$.

$X_{n,p} = X$ Counts the nuber of copies of k_4 (clique of size 4) in G.

$q_{n,p} = q = Pr(X_{n,p} \geq 1)$.

We will show the following threshold phenomonon:



Lemma. $\mathbb{E}(X_{n.p}) = \binom{n}{4} \cdot p^6$

Proof. For $S \subseteq [n]$ where $|S| = 4$, define $X_S = \begin{cases} 1 & S \text{ induces} k_4 \\ 0 & \text{otherwise} \end{cases}$. So:

$$X = \sum_{S \in \binom{[n]}{4}} X_S$$

by linearity of expectation $\implies$

$$\mathbb{E}(X) = \sum_{S \in \binom{[n]}{4}} \mathbb{E}(X_S) = \sum_{S \in \binom{[n]}{4}} Pr(S \text{ induces} k_4) = \sum_{S \in \binom{[n]}{4}} p^6 = \binom{n}{4} \cdot p^6$$

$\square$

Corollaries $\Rightarrow$ By markov inequality: $q_{n,p} \leq \frac{\mathbb{E}(X)}{1} = \binom{n}{4} \cdot p^6 = \Theta(n^4 \cdot p^6)$

$\Rightarrow$ If $p << n^{-\frac{2}{3}}$ then $q_{n,p} << 1$

Lemma. $\frac{Var(X)}{\mathbb{E}(X)^2} \leq \Theta(n^{-2} \cdot p^{-1} + n^{-3} \cdot p^{-3} + n^{-4} \cdot p^{-6})$

Proof. Same decomposition as previous proof:

$$X = \sum_{S \in \binom{[n]}{4}} X_S$$

$$
\begin{aligned}
Var(X) &= \mathbb{E}(X - \mathbb{E}(X))^2 \\
&= \mathbb{E}(\sum_{S \in \binom{[n]}{4}} X_S - \mathbb{E}(\sum_{S \in \binom{[n]}{4}} X_S))^2 \\
&= \mathbb{E}(\sum_{S \in \binom{[n]}{4}} X_S - \sum_{S \in \binom{[n]}{4}} \mathbb{E}(X_S))^2 \\
&= \mathbb{E}(\sum_{S \in \binom{[n]}{4}} (X_S - \mathbb{E}(X_S)))^2 \\
&= \mathbb{E}\left(\sum_{S \in \binom{[n]}{4}} \sum_{T \in \binom{[n]}{4}} \left(X_s - \mathbb{E}(X_S)\right)\left(X_T - -\mathbb{E}(X_T)\right)\right) \\
\text{linearity of expectation} &= \sum_{S,T \in \binom{[n]}{4}^2} \underbrace{\mathbb{E}\left(\left(X_s - \mathbb{E}(X_S)\right)\left(X_T - \mathbb{E}(X_T)\right)\right)}_{Cov(X_s, X_t)} \\
&= \sum_{i=0}^{4} \sum_{\substack{S,T \in \binom{[n]}{4}^2 \\ |S \cap T| = i}} Cov(X_s, X_T) \\
X_s, X_T \text{are independet if } |S \cap T| \leq 1 &= \sum_{i=2}^{4} \sum_{\substack{S,T \in \binom{[n]}{4}^2 \\ |S \cap T| = i}} Cov(X_s, X_T) \\
\text{see } * &\leq \sum_{i=2}^{4} \sum_{\substack{S,T \in \binom{[n]}{4}^2 \\ |S \cap T| = i}} \underbrace{\mathbb{E}(X_s X_t)} \\
&= \sum_{i=2}^{4} \sum_{\substack{S,T \in \binom{[n]}{4}^2 \\ |S \cap T| = i}} Pr(S \text{ and } T \text{ induces} k_4) \\
&= \underbrace{p^{11} \cdot \binom{n}{4} \cdot 6 \cdot \binom{n-4}{2}}_{i=2} + \underbrace{p^9 \cdot \binom{n}{4} \cdot 4 \cdot (n-4)}_{i=3} + \underbrace{p^6 \cdot \binom{n}{4}}_{i=4} \\
&= \Theta(p^{11} \cdot n^6 + p^9 \cdot n^5 + p^6 \cdot n^4)
\end{aligned}
$$

$$
\Longrightarrow \frac{Var(X)}{\mathbb{E}(X)^2} = \frac{Var(X)}{\Theta(n^8 \cdot p^{12})} \leq \Theta(n^{-2} \cdot p^{-1} + n^{-3} \cdot p^{-3} + n^{-4} \cdot p^{-6})
$$

$*$: because X_S, X_T are positive:

$$\begin{aligned}\mathbb{E}\left(\left(X_s - \mathbb{E}(X_S)\right)\left(X_T - \mathbb{E}(X_T)\right)\right) &= \mathbb{E}\left(X_s X_t - X_S\mathbb{E}(X_T) - \mathbb{E}(X_S)X_T + \mathbb{E}(X_S)\mathbb{E}(X_T)\right) \\ &= \mathbb{E}(X_s X_t) - \mathbb{E}(X_S)\mathbb{E}(X_T) \leq \mathbb{E}(X_s X_t)\end{aligned}$$

□

Corollaries ⇒by chevishev inequality (ex1/2) $q_{n,p} \geq 1 - \frac{Var(X)}{\mathbb{E}(X)^2}$
⇒if $p >> n^{-\frac{2}{3}}$ then $q_{n,p} = 1 - o(1)$

2 Birthday Paradox & Coupon Collector Problem

Consider the following: n bins, and k balls. We throw the balls into the bins uniformally at random.

Question (Birthday Paradox): What is the probability of the event $B_{n,k} = B$ where all balls fall into different pits.

Answer: (Threshold Phenomenon)

$k >> \sqrt{2n} \implies Pr(B) = o(1)$

$k << \sqrt{2n} \implies Pr(B) = 1 - o(1)$

Question (Coupon Collector): Assuming we have infinite balls, how many balls should we throw until (with high probability) each bin contains at least 1 balls.

Answer: (Threshold Phenomenon)

$k = nlog(n) + \omega(n) \implies$ with high prob ($1 - o_n(1)$) all bins have a ball

$k = nlog(n) - \omega(n) \implies$ with high prob ($1 - o_n(1)$) all bins have a ball there is an empty bin.

Definition. A random variable X is Geometric with parameter $0 \leq p \leq 1$ if $Pr(X = k) = (1 - P)^{k-1} \cdot p$. (Amount of tries until first success).

Fact. $\mathbb{E}(X) = \frac{1}{p}$,$Var(X) = \frac{1-p}{p^2} \leq \frac{1}{p^2}$

Fact. $\sum_{k=1}^{n} \frac{1}{k} = ln(n) + \Theta(1)$

Coupon Collector

T_n- number of balls that were thrown until all bins had a ball.

Lemma. $\mathbb{E}(T_n) = n \cdot ln(n) + \Theta(1)$

Proof. $T_{n,m} :=$ Number of balls until m bins are full.

$$T_n = T_{n,n} = \sum_{m=1}^{n} \underbrace{(T_{n,m} - T_{n,m-1})}_{:=S_m}$$

S_m is a geometric R.V. with parameter $\frac{n-(m-1)}{n}$.

$$\begin{aligned}
\implies \mathbb{E}(T_n) &= \sum_{m=1}^{n} \mathbb{E}(S_m) \\
&= \sum_{m=1}^{n} \frac{n}{n-(m-1)} \\
&= n \cdot \sum_{m=1}^{n} \frac{1}{n-(m-1)} \\
&= n \cdot \sum_{m=1}^{n} \frac{1}{m} \\
&= n \cdot (ln(n) + \Theta(1))
\end{aligned}$$

□

Now for the variance:

$$\begin{aligned}
Var(T_n) &= Var(\sum_{m=1}^{n} S_m) \\
\text{independence} &= \sum_{m=1}^{n} Var(S_m) \\
&\leq \sum_{m=1}^{n} \frac{n^2}{(n-(m-1))^2} \\
&= n^2 \sum_{m=1}^{n} \frac{1}{(n-(m-1))^2} \\
&= n^2 \sum_{m=1}^{n} \frac{1}{m^2} \\
&= \Theta(n^2)
\end{aligned}$$

Birthday Paradox

Mark $B_{n,k}$ The event that all balls fall into different bins. so:

$$Pr(B_{n,k}) = 1 \cdot (\frac{n-1}{n}) \cdot (\frac{n-2}{n}) \cdot ... \cdot (\frac{n-(k-1)}{n})$$

Fact. for any x, it hold that $1 + x \leq e^x$

additionaly, for sufficiently small x, $1 + x \geq e^{x-x^2}$

so:

$$\begin{aligned} Pr(B_{n,k}) &= \prod_{i=1}^{k-1} \left(1 - \frac{i}{n}\right) \\ &\approx \prod_{i=1}^{k-1} e^{-\frac{i}{n}} \\ &= e^{-\sum_{i=1}^{k-1} \frac{i}{n}} \\ &= e^{-\frac{1}{n} \cdot \sum_{i=1}^{k-1} i} \\ &= e^{-\frac{k \cdot (k-1)}{2n}} \end{aligned}$$

But $\approx$ is not formal. formally we know for the upper bound:

$$Pr(B_{n,k}) \leq e^{-\frac{k \cdot (k-1)}{2n}}$$

Also:

$$k >> \sqrt{2n} \implies Pr(B_{n,k}) = o(1)$$

Now, assuming $k < \sqrt{2n}$:

$$\begin{aligned} Pr(B_{n,k}) &\geq \prod_{i=1}^{k-1} e^{-\frac{i}{n} - (\frac{i}{n})^2} \\ &= e^{-\sum_{i=1}^{k-1} \frac{i}{n}} \cdot e^{-\sum_{i=1}^{k-1} (\frac{i}{n})^2} \\ &\geq e^{-\frac{k \cdot (k-1)}{2n}} \cdot e^{-\sum_{i=1}^{k} (\frac{k}{n})^2} \\ &= e^{-\frac{k \cdot (k-1)}{2n}} \cdot \underbrace{e^{-\frac{k^3}{n^2}}}_{1+o(1)} \end{aligned}$$