5. (a) Let X be a random variable. Use Chebyshev's inequality to show that

$$\Pr[X = 0] \le \frac{\mathrm{Var}[X]}{E[X]^2}.$$

"second moment method"

Deduce that if X_n is a sequence of random variables and $\mathrm{Var}[X_n] = o(E[X_n]^2)$, then $X_n = 0$ with high probability.

(b) Show that if $p \ge Cn^{-2/3}$ for some large constant C, then the random graph $G(n,p)$ contains a clique of size 4 with high probability.

Let $X = \#$ 4-cliques

$$\Rightarrow \Pr[X=0] \overset{(a)}{\le} \frac{\mathrm{Var}[X]}{\mathbb{E}[X]^2} \xrightarrow{\text{if}} 0$$

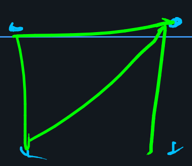
$$X = \sum_C X_C$$ C ranges over all size 4 subsets

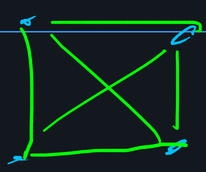
$$\mathbb{E}[X] = \sum_C E[X_C] = \binom{n}{4} p^6$$

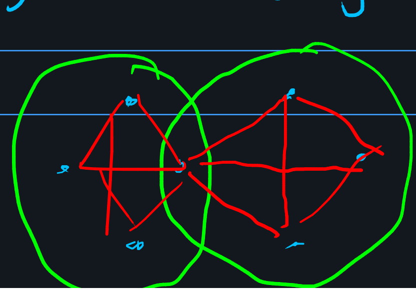
$$X_C X_{C'} = \begin{cases} 1 & \text{if } C \,\&\, C' \text{ clique} \\ 0 & \text{else} \end{cases}$$

$$\mathrm{Var}[X] = \sum_C \mathrm{Var}[X_C] + 2\sum_{C \neq C'} \mathrm{Cov}(X_C, X_{C'})$$

$$\mathrm{Cov}(X_C, X_{C'}) = E[X_C X_{C'}] - E[X_C]E[X_{C'}]$$
$$= \Pr[C \,\&\, C' \text{ cliques}] - \Pr[C \text{ clique}]\Pr[C' \text{ clique}]$$



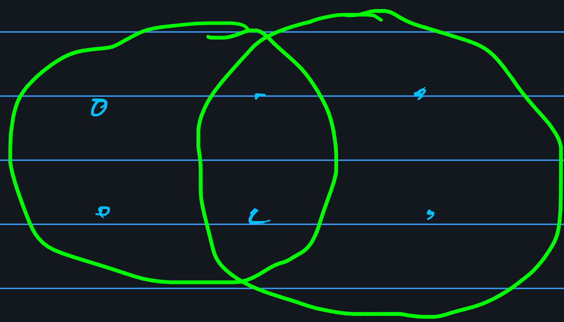




$\Rightarrow$ if $|C \cap C'| = 0, 1$

$\Rightarrow \mathrm{Cov}(X_C, X_{C'}) = 0$

$|C \cap C'| = 2$



$6 + 6 - 1 = 11$ edges

$\Rightarrow \Pr[C \,\&\, C' \text{ 4 clique}] = p^{11}$

$\Rightarrow \mathrm{Cov}(X_C, X_{C'}) = p^{11} - p^{12} = \Theta(p^{11})$

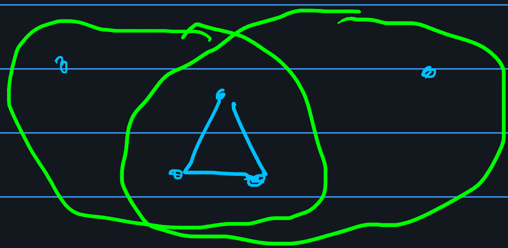
ways to choose $C \,\&\, C'$ s.t. $|C \cap C'| = 2$

$\Rightarrow$ 6 vertices, share 2

$$\binom{n}{6}\binom{6}{2} = \Theta(n^6)$$

$$\Rightarrow \sum_{|C \cap C'| = 2} \mathrm{Cov}(X_C, X_{C'}) = \Theta(n^6 p^{11})$$

if $|C \cap C'| = 3$



$6+6-3$ edges

$=9$

$$P[C \,\&\, C' \text{ clique}] = p^9$$

$$Cov(X_C, X_{C'}) = p^9 - p^{12} = \Theta(p^9)$$

ways to do this: $4+4-3 = 5$ vertices

overlap on 3

$$\Rightarrow \binom{n}{5}\binom{5}{3} = \Theta(n^5)$$

$$\Rightarrow \sum_{|C \cap C'|=3} Cov(X_C, X_{C'}) = \Theta(n^5 p^9)$$

$$\mathrm{Var}[X_c] = E[X_c^2] - E[X_c]^2$$
$$= E[X_c] - E[X_c]^2$$
$$= p^6 - p^{12} = \Theta(p^6)$$

$$\Pr[X=0] \le \frac{\mathrm{Var}[X]}{E[X]^2} = \frac{\sum \mathrm{Var}[X_c] + 2\sum \mathrm{Cov}(X_c, X_{c'})}{E[X]^2}$$

$$= \frac{\Theta(n^4p^6) + \Theta(n^6p^{11}) + \Theta(n^5p^9)}{\Theta(n^8p^{12})}$$

$$= \Theta\left(\frac{1}{n^4p^6}\right) + O\left(\frac{1}{n^2p}\right) + O\left(\frac{1}{n^3p^3}\right)$$

$$p = \omega(n^{-2/3})$$

$f = \omega(g) \Leftrightarrow \frac{f}{g} \to \infty$

$$\Rightarrow n^4p^6 = \omega(n^4(n^{-2/3})^6) = \omega(1) \to \infty$$

$$n^3p = \omega(n^2 n^{-2/3}) = \omega(n^{4/3}) \to \infty$$
$$n^3p^3 = \omega(n^3(p^{-2/3})^3) = \omega(n) \to \infty$$

$$\Rightarrow \Pr[X=0] \leq O\left(\frac{1}{n^4 p^6}\right) + O\left(\frac{1}{n^3 p}\right) + O\left(\frac{1}{n^3 p^3}\right)$$

$$\to 0$$

4. Suppose that X is uniformly distributed on the set of n even numbers $\{2, 4, 6, 8, \ldots, 2n\}$.

   (a) Calculate $M_X(t)$.

   (b) Calculate $\mathbb{E}[X^3]$.

   (c) Let $X_1, \ldots, X_k$ be independent copies of X and let $S_k = X_1 + \ldots + X_k$. Calculate $M_{S_k}(t)$.

$$M_X(t) = \mathbb{E}\left[e^{tX}\right]$$

$$= \sum_{k=1}^{n} \Pr[X=2k] e^{2kt}$$

$$= \frac{1}{n} \sum_{k=1}^{n} \left(e^{2t}\right)^k$$

↑ geometric sum

$$= \frac{e^{2t}}{n} \sum_{k=0}^{n-1} \left(e^{2t}\right)^k$$

$$= \frac{e^{2t}}{n} \frac{e^{2tn} - 1}{e^{2t} - 1}$$

$$\mathbb{E}[X^3] = \frac{d^3}{dt^3} M_X(t) \Big|_{t=0}$$

$$= \frac{d^3}{dt^3} \left[\frac{1}{n} \sum_{k=1}^{n} e^{2kt} \right] \Bigg|_{t=0}$$

$$= \left[\frac{1}{n} \sum_{k=1}^{n} 8k^3 e^{2kt} \right]_{t=0}$$

$$= \frac{8}{n} \underbrace{\sum_{k=1}^{n} k^3}$$

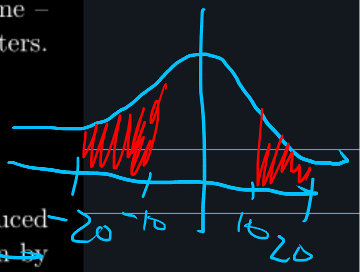
$$= \frac{8}{n} \left(\frac{n(n+1)}{2} \right)^2$$

2. A factory produces bolts of a certain width. The bolts don't come out perfectly each time – their widths deviate from the specified width by a random amount X, measured in micrometers. Suppose X has density function $f(x) = \frac{C}{1+(x/10)^2}$, where $-20 \le x \le 20$.

   (a) Find the value of C such that f is indeed a probability density function.

   (b) Calculate the mean and variance of X.

   (c) Suppose the factory produces 10,000 bolts a day. Assuming that each bolt is produced independently, ~~estimate the number of bolts whose widths deviate from specification by~~ more than 10 micrometers.



estimate $\Pr\left[\geq 2000 \text{ bolts differ from spec by} > 10\,\mu m\right]$

$\sim \text{Bin}(n,p)$

S_n = # bolts in n that differ by > 10

$$= \sum_{i=1}^{n} X_i \qquad X_i = \begin{cases} 1 & \text{if } i\text{-th bolt differs} \\ 0 & \text{else} \end{cases}$$

$$\mathbb{E}[X_i] = p$$

De Moivre : Laplace

$$\frac{S_n - np}{\sqrt{np(1-p)}} \to N(0,1)$$

$p = \Pr[\text{bolt differs f/ spec by} \geq 10]$

if X = bolt width

$$\Rightarrow p = \Pr[|X| \geq 10] = 2\Pr[10 \leq X \leq 20]$$

$$= 2\int_{10}^{20} f(x)\,dx = 2C\int_{10}^{20} \frac{dx}{1+\left(\frac{x}{10}\right)^2}$$

we want

$$\Pr[S_{10k} \geq 2000]$$

$$= \Pr\left[\frac{S_{10k} - 10000p}{(10000p(1-p))^{1/2}} \geq \frac{2000 - 10000p}{(10000p(1-p))^{1/2}}\right]$$

$$\approx \Pr\left[g \geq \frac{2000 - 10000p}{(10000p(1-p))^{1/2}}\right]$$

$$= 1 - \Phi\left(\frac{2000 - 10000p}{(10000p(1-p))^{1/2}}\right)$$

look up in a table

where Φ is cdf of $\mathcal{N}(0,1)$

$$\Phi(x) = \Pr[g \leq x]$$

$$\Pr[\text{prize} \mid X = x] = \frac{\Pr[\text{prize}, X = x]}{\Pr[X = x]}$$

3c) $\Pr[\text{prize} \mid \text{matching ticket}]$

$$= \sum_{X=0}^{1M} \Pr[\text{prize}, X=x]$$

$\sim \text{Pois}(1)$

$$= \sum_{X=0}^{1M} \Pr[\text{prize} \mid X=x] \Pr[X=x]$$

$$= \sum_{x=0}^{1M} \frac{1}{x+1} \frac{1^x e^{-1}}{x!} = \frac{1}{e} \sum_{x=0}^{1M} \frac{1}{(x+1)!}$$

$$\approx \frac{1}{e}(e-1) = 1 - 1/e$$

$$e = \sum_{n=0}^{\infty} \frac{1}{n!}$$

1. A biased coin lands heads with probability 1/10. The coin is flipped 200 times. Use Markov's inequality to give an upper bound on the probability that the coin shows heads at least 120 times. Improve this bound using Chebyshev's inequality. Improve it even further with Chernoff's inequality.

Chernoff: Let $X = \sum_{i=1}^{n} X_i$, $X_i \sim \text{Bern}(p_i)$

independent

$$\mu = E[X] = \sum_{i=1}^{n} p_i$$

Then

a) (upper) $\Pr[X \geq (1+\delta)\mu] \leq e^{-\delta^2\mu/2+\delta}$

$\forall \delta > 0$

b) (lower) $\Pr[X \leq (1-\delta)\mu] \leq e^{-\delta^2/2}$

$\forall\ 0 < \delta < 1$

1. A biased coin lands heads with probability 1/10. The coin is flipped 200 times. Use Markov's inequality to give an upper bound on the probability that the coin shows heads at least 120 times. Improve this bound using Chebyshev's inequality. Improve it even further with Chernoff's inequality.

Let $X = \#$ heads. $X = \sum_{i=1}^{200} X_i$, $X_i \sim \text{Bern}(1/10)$

indep

$$\mu = 200 \, 1/10 = 20$$

$\delta = 5$

Want $\Pr[X \geq 120] = \Pr[X \geq (1+\delta)20]$

$$\leq e^{-\delta^2 \mu / 2+\delta}$$

$$= \exp\left(-\frac{25 \cdot 20}{7}\right)$$

1. The weak law of large numbers states that, if $X_1, X_2, \ldots$ are iid random variables with mean μ and variance σ^2, then for any $\epsilon > 0$ we hve

$$\Pr\left[\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| > \epsilon\right] \to 0.$$

That is, the sample mean $\frac{1}{n}\sum_{i=1}^{n} X_i$ converges to the mean μ *in probability*. Prove the weak law of large numbers.

Let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$

$$\mathbb{E}[\bar{X}_n] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{n} \cdot n\mu = \mu$$

we want $\Pr[|\bar{X}_n - \mu| \geq \varepsilon]$

Chebyshev says $\Pr[|\bar{X}_n - \mathbb{E}[\bar{X}_n]| \geq \varepsilon]$

$$= \Pr[|\bar{X}_n - \mu| \geq \varepsilon] \leq \mathrm{Var}[\bar{X}_n] / \varepsilon^2$$

$$\mathrm{Var}[\bar{X}_n] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n^2}\sum^{n}\mathrm{Var}[X_i] = \frac{n\sigma^2}{n^2}$$

$$= \frac{\sigma^2}{n}$$

$$\Rightarrow \Pr\left[|\bar{X}_n - \mu| \geq \varepsilon\right] \leq \mathrm{Var}[\bar{X}_n]/\varepsilon^2$$

$$= \frac{\sigma^2}{n\varepsilon^2} \to 0$$