Probabilistic Methods - lectures 4+5

Theorem: every graph with m edges has cut of size $\geqslant \frac{m}{2}$.

Given two graphs, G_1 and G_2 on the same vertex set V, G_i has m_i edges, can we split V as $A \cup B$ s.t. $\exists$ roughly $m_i/2$ edges of G_i from A to B?

Variance X random variable, $\mu = \mathbb{E}[X]$.

$$\begin{aligned} Var(X) &= \mathbb{E}[(X-\mu)^2] \\ &= \mathbb{E}[X^2 - 2X\mu + \mu^2] \\ &= \mathbb{E}[X^2] - 2\mu\,\mathbb{E}[X] + \mu^2 \quad \text{(linearity of expectation)} \\ &= \mathbb{E}[X^2] - \mu^2 \end{aligned}$$

Def^n: $\sigma = \sqrt{Var(X)}$ is the standard deviation

Markov's inequality: let X be a non-negative R.V., $\mu = \mathbb{E}[X]$

$$\mathbb{P}[X \geqslant \alpha\mu] \leqslant 1/\alpha$$

Proof $\mu = \sum_{i=1}^{n} x_i p_i \geqslant \sum_{x_i \geqslant \alpha\mu} x_i p_i \geqslant \alpha\mu \sum_{x_i \geqslant \alpha\mu} p_i$

$$\Rightarrow \frac{1}{\alpha} \leqslant \sum_{x_i \geqslant \alpha\mu} p_i = \mathbb{P}[X \geqslant \alpha\mu]. \quad \blacksquare$$

Chebychev's inequality let X be a RV, $\mu = \mathbb{E}[X]$, $\sigma = \sqrt{Var(X)}$

$$\mathbb{P}[|X-\mu| \geqslant \lambda\sigma] \leqslant \frac{1}{\lambda^2}$$

Proof: $\mathbb{P}[|X-\mu| \geqslant \lambda\sigma] = \mathbb{P}[(X-\mu)^2 \geqslant \lambda^2\sigma^2]$

let $Z = (X-\mu)^2$ (non-negative RV)

remember $\sigma^2 = \mathbb{E}[(X-\mu)^2] = \mathbb{E}[Z]$

so apply Markov to Z. $\blacksquare$

Applications let $X = \sum_{i=1}^{k} X_i$

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

so if X_i and X_j are independent, $\mathrm{Cov}(X_i, X_j) = 0$

$$\mu = \mathbb{E}[X] = \sum_{i=1}^{k} \mathbb{E}[X_i]$$

$$\begin{aligned} Var(X) &= \mathbb{E}\Big[\Big(\sum_{i=1}^{k} X_i\Big)^2\Big] - \Big(\sum_{i=1}^{k} \mathbb{E}[X_i]\Big)^2 \\ &= \underbrace{\sum_{i=1}^{k}\left(\mathbb{E}[X_i^2] - \mathbb{E}^2[X_i]\right)}_{\sum^k Var(X_i)} + \sum_{i \neq j} \mathrm{cov}[X_i, X_j] \end{aligned}$$

$\therefore$ If $\{X_i\}_{i=1}^{k}$ are independent, $Var(X) = \sum_{i=1}^{k} Var(X_i)$

Distinct sums

Question: (Erdős) What is max k s.t. $\exists x_1, \dots, x_k \in [n]$ with all subset (partial) sums distinct? (Call this $f(n)$)

subset sum: $\sum_{i \in S} x_i \qquad S \subseteq \{1, \dots, k\}$

Total of 2^k subsets of $\{1, \dots, k\} \Rightarrow 2^k$ different partial sums.

Example: take $\{1, 2, 2^2, \dots 2^t\}$ s.t. $2^t \leq n$
ie $t = \lfloor \log_2 n \rfloor$

$\therefore f(n) \geq \lfloor \log_2 n \rfloor + 1$

Q: Is it true that $f(n) \leq \lfloor \log_2 n \rfloor + \underset{\text{constant.}}{C}$?

Warm up:

2^k different partial sums $\subseteq \{1, \dots, kn\}$
(because we add at most k numbers)

need $2^k \leq kn$

take $k = \log_2 n + \log_2 \log_2 n + C$

let $X = \sum_{i=1}^{k} \varepsilon_i x_i$ where $\varepsilon_i = \begin{cases} 1 & \text{probability } 1/2 \\ 0 & \text{" "} \end{cases}$

So X is a random partial sum.

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{k} \varepsilon_i x_i\right] = \sum_{i=1}^{k} x_i \mathbb{E}[\varepsilon_i] = \frac{x_1 + x_2 + \dots + x_k}{2} =: \mu$$

$$\mathrm{Var}[X] = \mathrm{Var}\left[\sum \varepsilon_i x_i\right] = \sum \mathrm{Var}(x_i \varepsilon_i) = \sum x_i^2 \mathrm{Var}(\varepsilon_i) = \frac{\sum x_i^2}{4} \leq \frac{kn^2}{4}$$

so $\sigma \leq \frac{n\sqrt{k}}{2}$

Use Chebychev to conclude that

$$\mathbb{P}[|X - \mu| \geq 2\sigma] \leq \frac{1}{4}$$

so $\mathbb{P}[|X - \mu| \leq 2\sigma] \geq \frac{3}{4} > \frac{1}{2}$

So at least $\frac{1}{2} \times 2^k = 2^{k-1}$ partial sums belong to $[\mu - 2\sigma, \mu + 2\sigma] = [\mu - n\sqrt{k}, \mu + n\sqrt{k}]$

$\therefore 2^{k-1} \leq 2\sqrt{k}\, n$

we can take $k = \log_2 n + \frac{1}{2} \log_2 \log_2 n + C$

Def$^{\underline{n}}$: $G(n,p)$ is a random graph on n vertices s.t $\mathbb{P}[(i,j) \text{ is edge}] = p$

Q: for what probability p does $G(n,p)$ contain K_4 almost surely?

with probability $\to 1$ as $n \to \infty$

Hardy / Ramanujan

Q: given a number x, ~~how many~~ let $\nu(x) = \#$ of prime divisors of x.

Theorem: (H/R) Most of the numbers $1 \leq x \leq n$ have the same number of prime divisors

so

$$\frac{\left|\left\{1 \leq x \leq n : \left|\nu(x) - \log\log x\right| \geq \omega(n)\sqrt{\log\log n}\right\}\right|}{n} = o(1) \quad (\text{or} \to 0)$$

where $\omega(n) \to \infty$ arbitrarily slowly.

In other words $\nu(x) \in [\log\log x - \omega(n)\sqrt{\log\log n}, \log\log x + \omega(n)\sqrt{\log\log n}]$

Proof: pick X between 1 and n uniformly at random.

so $\nu(X) = \sum_{p \leq n} X_p$ where $X_p = \begin{cases} 1 & \text{if } p \mid x \\ 0 & \text{if } p \nmid x \end{cases}$ $\forall p$ prime.

let $Z = \sum_{p \leq M} X_p$ where $M = n^{1/10}$

Now $|Z - \nu(X)| \leq 10$

$\mathbb{E}[X_p] = \frac{\lfloor n/p \rfloor}{n} = \frac{1}{p} + O(\frac{1}{n})$

$\Rightarrow \mathbb{E}[Z] = \sum_{p \leq M} \mathbb{E}[X_p] = \sum_{p \leq M} \frac{1}{p} + O(1)$

$$\boxed{\text{Fact: } \forall t, \quad \sum_{p \leq t} \frac{1}{p} = \log\log t + O(1)}$$

(Aside) Theorem # of primes $p \leq x = (1 + o(1)) \frac{x}{\log x}$

so fact follows from: $\sum_{p \leq x} \frac{1}{p} \approx \int_1^t \frac{1}{x \log x} dx = \int_1^{\log t} \frac{1}{w} dw \quad (w = \log x) = \log\log t$.

So $\mathbb{E}[Z] = \log\log n + O(1)$

$$\mathrm{Var}[Z] = \sum_{p \le M} \mathrm{Var}[X_p] + \sum_{\substack{p \ne q \\ p,q \le M}} \mathrm{cov}[X_p, X_q]$$

$$\mathrm{Var}[X_p] = \mathbb{E}[X_p^2] - \mathbb{E}[X_p]^2 = \frac{1}{p} + O(\tfrac{1}{n}) - \left(\frac{1}{p} + O(\tfrac{1}{n})\right)^2$$

$$= \frac{1}{p}\left(1 - \frac{1}{p}\right) + O(\tfrac{1}{n})$$

$$\therefore \sum_{p \le M} \mathrm{Var}[X_p] = \sum_{p \le M}\left(\frac{1}{p} - \frac{1}{p^2}\right) + \cancel{o(n)}\, O(1) = \log\log n + O(1)$$

$$\mathrm{cov}[X_p X_q] = \mathbb{E}[X_p X_q] - \mathbb{E}[X_p]\mathbb{E}[X_q]$$

$$= \frac{\lfloor \frac{n}{pq} \rfloor}{n} - \left(\frac{1}{p} + O(\tfrac{1}{n})\right)\left(\frac{1}{q} + O(\tfrac{1}{n})\right)$$

$$= \frac{1}{pq} + O(\tfrac{1}{n}) - \frac{1}{pq} + O(\tfrac{1}{n})$$

$$= O(\tfrac{1}{n})$$

$$\therefore \sum_{p,q \le M} \mathrm{cov}[X_p X_q] \le M^2 O(\tfrac{1}{n}) \ll O\left(\frac{M^2}{n}\right) << 1$$

So by Chebychev's inequality

$$\mathbb{P}[\,|Z - \mathbb{E}[Z]| > \omega(n)\sigma] \le \frac{1}{\omega(n)^2} \longrightarrow 0$$

ie $\quad \mathbb{P}[\,|\nu(x) - \log\log n| > \omega(n)\sqrt{\log\log n}\,] \longrightarrow 0$ ∎

Theorem:

$$\frac{\#\{1 \le x \le n : \nu(x) \ge \log\log n + \sqrt{\log\log n}\,\lambda\}}{n}$$

$$\longrightarrow \int_{\lambda}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \qquad \text{as } n \to \infty.$$

Local lemma

If you have n events, each with probability $1/2$,
s.t. every event depends only on a few other events
then with probability > 0 no event happens.
(e.g. n dependent coin tosses)

Large deviation inequalities

Example: let RV $Z = f(x_1, \dots, x_n)$ where $x_i = \begin{cases} 1 \\ 0 \end{cases}$ indicators

call f Lipschitz if

$|f(y) - f(y')| \leq 1$ for any two vectors $y, y' \in \{0,1\}^n$ which differ in 1 coordinate.

Then

$$\mathbb{P}[\,|f - \mathbb{E}(f)| \geq t\,] \leq 2e^{-t^2/n}$$