TTIC 31150/CMSC 31150 Mathematical Toolkit (Spring 2023)

Avrim Blum and Ali Vakilian

Lecture 12: Tail inequalities 2

Recap

- The probabilistic method, coupon collector problem, DeMillo-Lipton-Schwartz-Zippel lemma, polynomial identity testing, application of DLSZ to finding perfect matchings.

- Basic tail inequalities: Markov's inequality and Chebyshev's inequality.

- Properties of variance: $Var(\sum_i X_i) = \sum_i Var(X_i)$ if pairwise independent.

- Markov vs Chebyshev for coin flips.

- Threshold phenomena in random graphs.

Markov and Chebyshev

Proposition 1.1 (Markov's Inequality) *Let X be non-negative variable. Then,*

$$\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}[X]}{t}. \tag{1}$$

Equivalently,

$$\mathbb{P}[X \geq a \cdot \mathbb{E}[X]] \leq \frac{1}{a}. \tag{2}$$

Proposition 1.2 (Chebyshev's inequality) *Let X be a random variable and let $\mu = \mathbb{E}[X]$. Then,*

$$\mathbb{P}[|X-\mu| \geq t] \leq \frac{\mathsf{Var}[X]}{t^2} = \frac{\mathbb{E}[(X-\mu)^2]}{t^2}. \tag{3}$$

Threshold phenomena in Random Graphs

Consider a graph G on n vertices where each possible edge is placed into the graph independently with probability p. This is called the $G_{n,p}$ random graph model.

It turns out that many graph properties have "threshold phenomena": for some function $f(n)$, for $p \ll f(n)$ the graph will almost surely not have the property and for $p \gg f(n)$ the graph almost surely will have the property (or vice-versa).

One example: the property of containing a 4-clique.

Threshold phenomena in Random Graphs

Theorem 3.1 *Let G be generated randomly according to the model $\mathcal{G}_{n,p}$ graph. Then,*

1. *If $p \ll n^{-2/3}$, then $\mathbb{P}[G$ contains a 4-clique$] \to 0$ as $n \to \infty$.*
2. *If $p \gg n^{-2/3}$, then $\mathbb{P}[G$ contains a 4-clique$] \to 1$ as $n \to \infty$.*

(1) Is the easier case, so let's start with that:

- For each set S of 4 vertices, define indicator R.V. X_S for the event that S is a clique.
- Let $X = \sum_S X_S$ denote the number of 4-cliques in the graph.
- We have $\mathbb{E}[X] = \sum_S \mathbb{E}[X_S] = O\left(n^4 p^6\right) = o(1)$ for $p \ll n^{-2/3}$.
- So, by Markov's inequality, $\mathbb{P}[X \geq 1] \leq \mathbb{E}[X]/1 = o(1)$.

Threshold phenomena in Random Graphs

Theorem 3.1 *Let G be generated randomly according to the model $\mathcal{G}_{n,p}$ graph. Then,*

1. *If $p \ll n^{-2/3}$, then $\mathbb{P}\left[G \text{ contains a 4-clique}\right] \to 0$ as $n \to \infty$.*
2. *If $p \gg n^{-2/3}$, then $\mathbb{P}\left[G \text{ contains a 4-clique}\right] \to 1$ as $n \to \infty$.*

For (2), we have $\mathbb{E}[X] = \Theta\left(n^4 p^6\right) \to \infty$, but this is not sufficient to get $\mathbb{P}[X = 0] = o(1)$.

For this, we will use Chebyshev's inequality with $t = \mathbb{E}[X]$, giving:

$$\mathbb{P}[X = 0] \leq \frac{Var[X]}{\mathbb{E}[X]^2}$$

Second Moment method

So, if we can show that $Var[X] = o(\mathbb{E}[X]^2)$, we will be done.

Threshold phenomena in Random Graphs

Theorem 3.1 *Let G be generated randomly according to the model $\mathcal{G}_{n,p}$ graph. Then,*

1. *If $p \ll n^{-2/3}$, then $\mathbb{P}[G \text{ contains a 4-clique}] \to 0$ as $n \to \infty$.*
2. *If $p \gg n^{-2/3}$, then $\mathbb{P}[G \text{ contains a 4-clique}] \to 1$ as $n \to \infty$.*

We can write variance as: $Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sum_{S,S'} \mathbb{E}[X_S X_{S'}] - \mathbb{E}[X]^2$.

Let's now consider a few cases for S, S':

- If S, S' share at most 1 vertex in common, then X_S and $X_{S'}$ are independent, so $\mathbb{E}[X_S X_{S'}] = \mathbb{E}[X_S]\mathbb{E}[X_{S'}]$ and the sum over all of these is at most $\mathbb{E}[X]^2$. We can therefore cover these using the $-\mathbb{E}[X]^2$ term.

So, if we can show that $Var[X] = o(\mathbb{E}[X]^2)$, we will be done.

Threshold phenomena in Random Graphs

Theorem 3.1 *Let G be generated randomly according to the model $\mathcal{G}_{n,p}$ graph. Then,*

1. *If $p \ll n^{-2/3}$, then $\mathbb{P}[G$ contains a 4-clique$] \to 0$ as $n \to \infty$.*
2. *If $p \gg n^{-2/3}$, then $\mathbb{P}[G$ contains a 4-clique$] \to 1$ as $n \to \infty$.*

We can write variance as: $Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sum_{S,S'} \mathbb{E}[X_S X_{S'}] - \mathbb{E}[X]^2$.

Let's now consider a few cases for S, S':

- If S, S' share 2 vertices in common, there are at most $O(n^6)$ such cases and each one has $\mathbb{E}[X_s X_{S'}] = p^{11}$. So, overall, we get $O(n^6 p^{11}) = o(n^8 p^{12}) = o(\mathbb{E}[X]^2)$.

So, if we can show that $Var[X] = o(\mathbb{E}[X]^2)$, we will be done.

Threshold phenomena in Random Graphs

Theorem 3.1 *Let G be generated randomly according to the model $\mathcal{G}_{n,p}$ graph. Then,*

1. *If $p \ll n^{-2/3}$, then $\mathbb{P}[G \text{ contains a 4-clique}] \to 0$ as $n \to \infty$.*
2. *If $p \gg n^{-2/3}$, then $\mathbb{P}[G \text{ contains a 4-clique}] \to 1$ as $n \to \infty$.*

We can write variance as: $Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sum_{S,S'} \mathbb{E}\left[X_S X_{S'}\right] - \mathbb{E}[X]^2$.

Let's now consider a few cases for S, S':

- If S, S' share 3 vertices in common, there are at most $O(n^5)$ such cases and each one has $\mathbb{E}\left[X_s X_{S'}\right] = p^9$. So, overall, we get $O\left(n^5 p^9\right) = o\left(n^8 p^{12}\right) = o\left(\mathbb{E}[X]^2\right)$.

So, if we can show that $Var[X] = o(\mathbb{E}[X]^2)$, we will be done.

Threshold phenomena in Random Graphs

Theorem 3.1 *Let G be generated randomly according to the model $\mathcal{G}_{n,p}$ graph. Then,*

1. *If $p \ll n^{-2/3}$, then $\mathbb{P}[G \text{ contains a 4-clique}] \to 0$ as $n \to \infty$.*
2. *If $p \gg n^{-2/3}$, then $\mathbb{P}[G \text{ contains a 4-clique}] \to 1$ as $n \to \infty$.*

We can write variance as: $Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sum_{S,S'} \mathbb{E}[X_S X_{S'}] - \mathbb{E}[X]^2$.

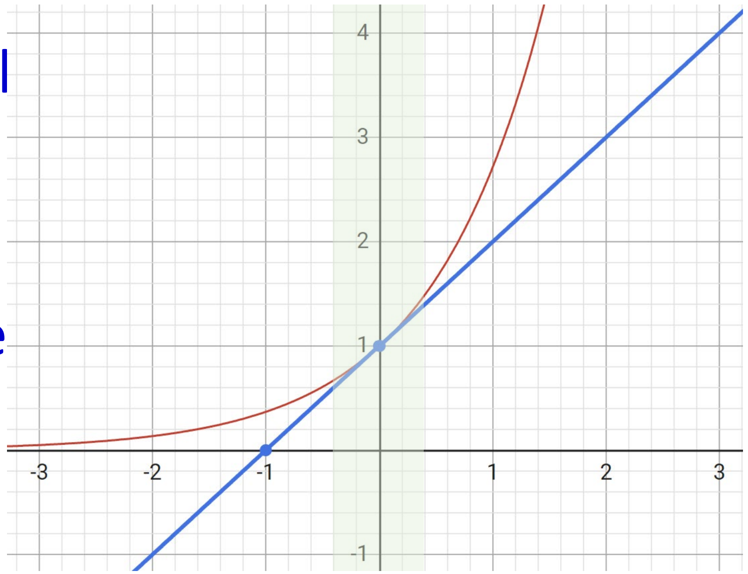
Let's now consider a few cases for S, S':

- And finally, if S, S' share all 4 vertices in common, then the total is just $\mathbb{E}[X] = o(\mathbb{E}[X]^2)$.
- So, overall we have $Var[X] = o(\mathbb{E}[X]^2)$ as desired.

So, if we can show that $Var[X] = o(\mathbb{E}[X]^2)$, we will be done.

Chernoff-Hoeffding bounds

$y = e^x$, $y = 1 + x$



Consider n mutually independent Bernoulli R.V.s $X_1, \dots, X_n$, wl

- Let $X = \sum_i X_i$, and let $\mu = \mathbb{E}[X] = \sum_i p_i$.

Q: how can we use mutual independence to show that it is ve
far from its expectation?

Idea: Define $Y_i = e^{\lambda X_i}$ for some small $\lambda > 0$.

- So, if $X_i = 0$ then $Y_i = 1$, and if $X_i = 1$ then $Y_i \approx 1 + \lambda$. $\mathbb{E}[Y_i] \approx 1 + p_i\lambda \approx e^{p_i\lambda}$.
- Now, consider product Y of the Y_i. $\mathbb{E}[Y] = \prod_i \mathbb{E}[Y_i] \approx e^{\lambda \sum_i p_i} = e^{\lambda \mathbb{E}[X]}$.
- By Markov, $\mathbb{P}\big[Y \geq k\mathbb{E}[Y]\big] \leq \frac{1}{k}$. But since $X = \frac{\ln Y}{\lambda}$, this means $\mathbb{P}\left[X \geq \frac{\ln \mathbb{E}[Y]}{\lambda} + \frac{\ln k}{\lambda}\right] \leq \frac{1}{k}$.
- And for small λ, $\frac{\ln \mathbb{E}[Y]}{\lambda} \approx \mathbb{E}[X]$. So, even for large k, X is just a little bit larger than $\mathbb{E}[X]$.

Chernoff-Hoeffding bounds

Consider n mutually independent Bernoulli R.V.s $X_1, \dots, X_n$, where $\mathbb{P}(X_i = 1) = p_i$.

- Let $X = \sum_i X_i$, and let $\mu = \mathbb{E}[X] = \sum_i p_i$.

But, we are cheating: these "≈" are not exact and require small λ. So, let's now do this carefully.

Idea: Define $Y_i = e^{\lambda X_i}$ for some small $\lambda > 0$.

- So, if $X_i = 0$ then $Y_i = 1$, and if $X_i = 1$ then $Y_i \approx 1 + \lambda$. $\mathbb{E}[Y_i] \approx 1 + p_i\lambda \approx e^{p_i\lambda}$.
- Now, consider product Y of the Y_i. $\mathbb{E}[Y] = \prod_i \mathbb{E}[Y_i] \approx e^{\lambda \sum_i p_i} = e^{\lambda \mathbb{E}[X]}$.
- By Markov, $\mathbb{P}\left[Y \geq k\mathbb{E}[Y]\right] \leq \frac{1}{k}$. But since $X = \frac{\ln Y}{\lambda}$, this means $\mathbb{P}\left[X \geq \frac{\ln \mathbb{E}[Y]}{\lambda} + \frac{\ln k}{\lambda}\right] \leq \frac{1}{k}$.
- And for small λ, $\frac{\ln \mathbb{E}[Y]}{\lambda} \approx \mathbb{E}[X]$. So, even for large k, X is just a little bit larger than $\mathbb{E}[X]$.

Chernoff-Hoeffding bounds

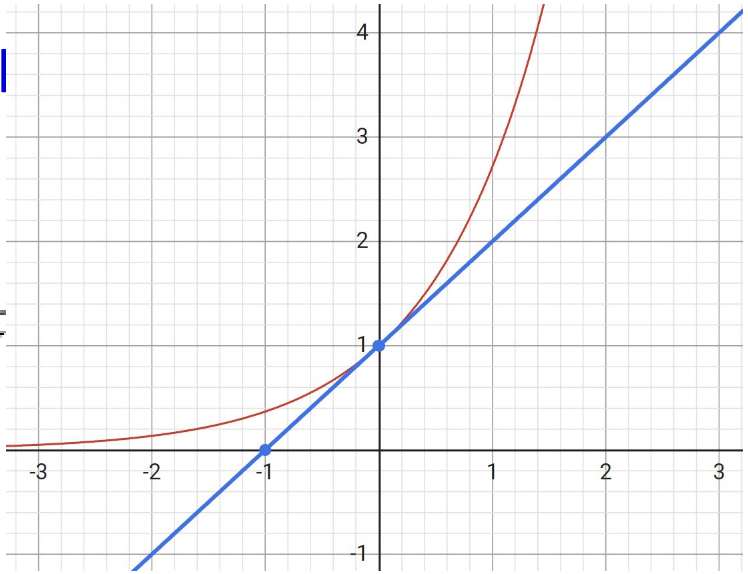
Consider n mutually independent Bernoulli R.V.s $X_1, \ldots, X_n$, wl

- Let $X = \sum_i X_i$, and let $\mu = \mathbb{E}[X] = \sum_i p_i$.

Using the fact that the function e^x is strictly increasing, we get that

$$\mathbb{P}[X \geq (1+\delta)\mu] = \mathbb{P}\left[e^{\lambda X} \geq e^{\lambda(1+\delta)\mu}\right] \overset{\text{(Markov)}}{\leq}$$

$y = e^x, y = 1 + x$



Let's analyze the numerator:

$$\mathbb{E}\left[e^{\lambda X}\right] = \mathbb{E}\left[e^{\lambda(X_1+\ldots+X_n)}\right] = \mathbb{E}\left[\prod_{i=1}^{n} e^{\lambda X_i}\right] \overset{\text{(independence)}}{=} \prod_{i=1}^{n} \mathbb{E}\left[e^{\lambda X_i}\right]$$

$$= \prod_{i=1}^{n} \left[p_i e^{\lambda} + (1-p_i)\right]$$

$$= \prod_{i=1}^{n} \left[1 + p_i(e^{\lambda} - 1)\right]. \leq \prod_{i=1}^{n} e^{p_i(e^{\lambda}-1)}$$

Now use $1 + x \leq e^x$ to get

$$\mathbb{E}\left[e^{\lambda X}\right] \leq e^{\sum_i p_i(e^{\lambda}-1)} = e^{(e^{\lambda}-1)\mu}$$

Chernoff-Hoeffding bounds

So, $\mathbb{P}[X \geq (1+\delta)\mu] \leq e^{(e^{\lambda}-1)\mu - \lambda(1+\delta)\,\mu}$. Set λ to minimize ($e^{\lambda} = 1+\delta,\ \lambda = \ln(1+\delta)$)

Using the fact that the function e^x is strictly increasing, we get that for $\lambda > 0$

$$\mathbb{P}\left[X \geq (1+\delta)\mu\right] \;=\; \mathbb{P}\left[e^{\lambda X} \geq e^{\lambda(1+\delta)\mu}\right] \;\overset{\text{(Markov)}}{\leq}\; \frac{\mathbb{E}\left[e^{\lambda X}\right]}{e^{\lambda(1+\delta)\mu}}.$$

Let's analyze the numerator:

$$\mathbb{E}\left[e^{\lambda X}\right] \;=\; \mathbb{E}\left[e^{\lambda(X_1+\ldots+X_n)}\right] \;=\; \mathbb{E}\left[\prod_{i=1}^{n} e^{\lambda X_i}\right] \;\overset{\text{(independence)}}{=}\; \prod_{i=1}^{n} \mathbb{E}\left[e^{\lambda X_i}\right]$$

$$= \prod_{i=1}^{n}\left[p_i e^{\lambda} + (1-p_i)\right]$$

$$= \prod_{i=1}^{n}\left[1 + p_i(e^{\lambda}-1)\right].$$

Now use $1 + x \leq e^x$ to get

$$\mathbb{E}[e^{\lambda X}] \leq e^{\sum_i p_i(e^{\lambda}-1)} = e^{(e^{\lambda}-1)\mu}$$

Chernoff-Hoeffding bounds

So, $\mathbb{P}[X \geq (1+\delta)\mu] \leq e^{(e^\lambda - 1)\mu - \lambda(1+\delta)\,\mu}$. Set λ to minimize ($e^\lambda = 1+\delta,\ \lambda = \ln(1+\delta)$)

Get: $\mathbb{P}[X \geq (1+\delta)\mu] \leq e^{\mu(\delta - (1+\delta)\ln(1+\delta))} = \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu$.

Similarly, $\mathbb{P}[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^\mu$.

For $\delta \in [0,1]$ can use Taylor series to simplify to:

- $\mathbb{P}[X \geq (1+\delta)\mu] \leq e^{-\delta^2\mu/3}$
- $\mathbb{P}[X \leq (1-\delta)\mu] \leq e^{-\delta^2\mu/2}$

$$e^{\delta - (1+\delta)\ln(1+\delta)}$$
$$\leq e^{\delta - (1+\delta)\left(\delta - \frac{\delta^2}{2} + \frac{\delta^3}{3} - \frac{\delta^4}{4} + \cdots\right)}$$
$$= e^{-\frac{\delta^2}{2} + \frac{\delta^3}{6} - \frac{\delta^4}{12} + \cdots} \leq e^{-\frac{\delta^2}{3}}$$

Comparing vs Chebyshev on fair coin tosses

Consider n independent fair coin flips $X_1, \dots, X_n$, $\mathbb{P}(X_i = 1) = \frac{1}{2}$, $X = \sum_i X_i$, $\mu = \mathbb{E}[X] = \frac{n}{2}$

- Chebyshev: $\mathbb{P}[|X - \mu| \geq \delta\mu] \leq \frac{Var[X]}{\delta^2\mu^2} = \frac{n/4}{\delta^2(n/2)^2} = \frac{1}{\delta^2 n}$.
- Chernoff/Hoeffding: $\mathbb{P}[|X - \mu| \geq \delta\mu] \leq 2e^{-\delta^2 n/6}$.
    - ➢ Using $\delta = k/\sqrt{n}$, get $\mathbb{P}[|X - \mu| \geq k\sigma] = e^{-O(k^2)}$.

$n = 1000, \mu = 500$

- Markov $\mathbb{P}[X > 600] \leq 5/6 \approx 0.83$
- Chebyshev $\mathbb{P}[X > 600] \leq \mathbb{P}[|X - 500| > 0.2 \times 500] \leq 250/(0.2 \times 500)^2 \approx 0.025$
- Chernoff $\mathbb{P}[X > 600] \leq \mathbb{P}[|X - 500| > 0.2 \times 500] \leq 2e^{-.2^2 \times 1000/6} \approx 0.001$

Random Vectors

Suppose we pick m random vectors $v_1, \dots, v_m \in \{-1,1\}^n$. Clearly, $\langle v_i, v_i \rangle = n$.

What about $\langle v_i, v_j \rangle$ for $i \neq j$? Claim: whp, $|\langle v_i, v_j \rangle| = O\left(\sqrt{n \log m}\right)$ for all $i \neq j$.

So, even though can only have n truly orthogonal vectors, can have a much larger number of nearly-orthogonal vectors.

Random Vectors

Suppose we pick m random vectors $v_1, \dots, v_m \in \{-1,1\}^n$. Clearly, $\langle v_i, v_i \rangle = n$.

What about $\langle v_i, v_j \rangle$ for $i \neq j$? Claim: whp, $|\langle v_i, v_j \rangle| = O\left(\sqrt{n \log m}\right)$ for all $i \neq j$.

Proof: First, fix some i, j s.t. $i \neq j$ (then will do a union bound over all $\binom{m}{2}$ such pairs).

- For $k \in \{1, \dots, n\}$ let X_k be indicator RV for event that kth coordinate of v_i, v_j are equal.
- Let $X = \sum_k X_k$. By Chernoff/Hoeffding, $\mathbb{P}\left(\left|X - \frac{n}{2}\right| \geq \frac{\delta n}{2}\right) \leq 2e^{-\delta^2 n/6}$.
- Notice that $|\langle v_i, v_j \rangle| = 2\left|X - \frac{n}{2}\right|$. So, using $\delta = 6\sqrt{\frac{\ln m}{n}}$ we get:

$$\mathbb{P}\left(|\langle v_i, v_j \rangle| \geq 6\sqrt{n \ln m}\right) = \mathbb{P}\left(2\left|X - \frac{n}{2}\right| \geq 2\frac{\delta n}{2}\right) \leq 2e^{-6 \ln m} = 2m^{-6}.$$

Finally, do a union bound over all $\binom{m}{2}$ pairs. Overall prob of failure $\leq m^{-4}$.

Balls and Bins revisited

We saw earlier that if we toss balls independently at random into n bins, it will take an expected $\Theta(n \log n)$ tosses until there are no empty bins.

Other statistics:

- If toss n balls into n bins, what is the expected fraction of empty bins?
  - Let X_i be indicator R.V. for event that bin i is empty. $\mathbb{E}[X_i] = \left(1 - \frac{1}{n}\right)^n \approx \frac{1}{e}$. So, expected fraction of empty bins is $\approx 1/e$.
- If toss n balls into n bins, how loaded will the most-loaded bin be?

Balls and Bins revisited

Claim: if we toss n balls into n bins, whp no bin will have more than $t = \frac{3 \ln n}{\ln \ln n}$ balls.

Proof:

- Let X_{ij} be indicator RV for event that ball j is in bin i. Let $Z_i = \sum_j X_{ij}$. What is $\mathbb{E}[Z_i]$?
- $\mathbb{E}[Z_i] = 1$ and is a sum of independent Bernoulli R.V.s, so can apply Chernoff/Hoeffding.
- $\mathbb{P}[Z_i \geq t] \leq \frac{e^{t-1}}{t^t} \leq \left(\frac{e}{t}\right)^t$.

$$\mathbb{P}[Z_i \geq \underbrace{(1+\delta)}_{=t}\mu] \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu$$

- For $t = \frac{3 \ln n}{\ln \ln n}$ we have $\left(\frac{e}{t}\right)^t \leq \left(\frac{\ln \ln n}{\ln n}\right)^t = O\left(\left(\frac{1}{\ln n}\right)^{0.9t}\right) = O\left(e^{-2.7 \ln n}\right) = O\left(n^{-2.7}\right)$.
- Now do a union bound over all i.