# TTIC 31150/CMSC 31150 Mathematical Toolkit (Fall 2024)

Avrim Blum

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,*

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

# Chernoff-Hoeffding bounds

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

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

Q: how can we use mutual independence to show that it is very unlikely that $X$ will be *too* far from its expectation?

Intuition: Let's say all $p_i = 0.5$. $X =$ #heads in a sequence of fair coin flips. Want to show it's unlikely we'll have many more heads than tails.

- Consider this random walk: starting at 1, on heads multiply by $1 + \epsilon$ and on tails multiply by $1 - \epsilon$. Final position $Y = \prod_i Y_i$, where $\Pr[Y_i = 1 + \epsilon] = \Pr[Y_i = 1 - \epsilon] = 0.5$.
- By mutual independence, $\mathbb{E}[Y] = \prod_i \mathbb{E}[Y_i] = \prod_i 1 = 1$.
- By Markov, $\Pr[Y > 1{,}000{,}000] < .000001$. But since $Y$ grows multiplicatively, it doesn't take *too* many more heads than tails for $Y$ to get large. (Caveat: $(1 + \epsilon)(1 - \epsilon) = 1 - \epsilon^2$)

# 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 $\mu = \mathbb{E}[X] = \sum_i p_i$.

Now, let's do this for real…

# 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 $\mu = \mathbb{E}[X] = \sum_i p_i$.

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:

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

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

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

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

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

# Chernoff-Hoeffding bounds

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

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

# Chernoff-Hoeffding bounds

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

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

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

# Comparing vs Chebyshev on fair coin tosses

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

- 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 $\mathbb{P}[|X - \mu| \geq k\sigma] = e^{-O(k^2)}$ .

# Random Vectors

Suppose we pick $m$ random vectors $v_1, \ldots, v_m \in \{-1,1\}^n$. 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, $|\langle v_i, v_j \rangle| = O(\sqrt{n \log m})$ 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 $k$th 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:

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

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.
- $\mathbb{P}[Z_i \geq t] \leq \frac{e^{t-1}}{t^t} \leq \left(\frac{e}{t}\right)^t.$

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

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