Mathematical Toolkit **Autumn 2023**

Lecture 14: November 14, 2023

Lecturer: Madhur Tulsiani

1 Tail Inequalities

We will develop some inequalities which let us bound the probability of a random variable taking a value very far from its expectation.

1.1 Markov's Inequality

This is the most basic inequality we will use. This is useful if the only thing we know about a random variable is its expectation. It will also be useful to derive other inequalities later.

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

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

Proof: We start by considering the event $E \equiv \{Z \geq t\}$. We can then write,

$$\mathbb{E}[Z] = \mathbb{P}[E] \cdot \mathbb{E}[Z \mid E] + \mathbb{P}[E^c] \cdot \mathbb{E}[Z \mid E^c] .$$

Using non-negativity of Z, we get

$$\mathbb{E}[Z] \geq \mathbb{P}[E] \cdot \mathbb{E}[Z \mid E] \geq \mathbb{P}[E] \cdot t = \mathbb{P}[Z \geq t] \cdot t ,$$

which completes the proof. ∎

1.2 Chebyshev's Inequality

The variance of a random variable X is defined as

$$\mathsf{Var}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2$$

Also, for two random variables X and Y, we define the covariance as

$$\mathsf{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y] .$$

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

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

Proof: Consider the non-negative random variable $(Z-\mu)^2$. Applying Markov's inequality we have

$$\mathbb{P}[|Z-\mu| \geq t] \;=\; \mathbb{P}\left[(Z-\mu)^2 \geq t^2\right] \;\leq\; \frac{\mathbb{E}\left[(Z-\mu)^2\right]}{t^2}.$$

■

1.3 Coin tosses revisited

An unbiased coin is tossed n times. Probability that head shows up in each toss is $\frac{1}{2}$. Let Z be a random variable for the number of heads that have showed up after n tosses. We also have random variables X for i^{th} coin toss, where $X_i = 1$ if head shows up in i^{th} toss and 0 otherwise.

So we have

$$Z = \sum_{i=1}^{n} X_i \quad \text{and} \quad \mathbb{E}[Z] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{n}{2}.$$

Let us now compare the kind of bounds we get using Markov's and Chebyshev's inequalities.

Application of Markov's inequality . Using Markov's inequality we have,

$$\mathbb{P}\left[Z \geq \frac{3n}{4}\right] \leq \frac{\mathbb{E}[Z]}{(3n/4)} \;\Rightarrow\; \mathbb{P}\left[Z \geq \frac{3n}{4}\right] \leq \frac{2}{3} \;\Rightarrow\; \mathbb{P}\left[Z - \frac{n}{2} \geq \frac{n}{4}\right] \leq \frac{2}{3}.$$

Application of Chebyshev's inequality . We will show that Chebyshev's inequality gives a stronger bound on probability. Since Z is a Binomial random variable, we have that

$$\mathsf{Var}[Z] \;=\; n \cdot \frac{1}{2} \cdot \left(1 - \frac{1}{2}\right) \;=\; \frac{n}{4}.$$

Applying Chebyshev's inequality we have,

$$\mathbb{P}\left[\left|Z - \frac{n}{2}\right| \geq t\right] \;\leq\; \frac{n}{4t^2}.$$

Setting $t = n/4$ and $t = \sqrt{n}$, gives the following bounds

$$\mathbb{P}\left[\left|Z - \frac{n}{2}\right| \geq \frac{n}{4}\right] \;\leq\; \frac{4}{n} \quad \text{and} \quad \mathbb{P}\left[\left|Z - \frac{n}{2}\right| \geq \sqrt{n}\right] \;\leq\; \frac{1}{4}$$

Thus, Chebyshev's inequality gives a much stronger bound on a deviation of $n/4$ from the mean, and can also bound the probability of deviations as small as $\sqrt{n}$. In particular, it gives a non-trivial bound whenever the deviation is larger than $\sqrt{\mathrm{Var}[Z]}$, a quantity which is referred to as the *standard deviation* of the random variable Z.

2 Threshold Phenomena in Random Graphs

We consider a model of Random Graphs by Erdős and Rényi [**?**]. To generate a random graph with n vertices, for every pair of vertices $\{i, j\}$, we put an edge independently with probability p. This model is denoted by $\mathcal{G}_{n,p}$.

Let G be a random $\mathcal{G}_{n,p}$ graph and let H be any fixed graph (on some constant number of vertices independent of n). We will be interested in understanding the probability that G contains a copy of H. We start by computing this when H is K_4, the clique on 4 vertices.

Definition 2.1 *We define k-*clique *to be a fully connected graph with k vertices.*

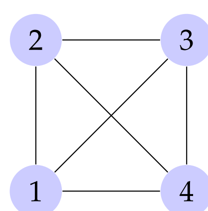

Figure 1: 4-Clique

As a convention, we will count a permutation of a copy of K_4 as the *same* copy. We define the random variable

$$Z \;=\; \text{number of copies of } K_4 \text{ in } G \;=\; \sum_C X_C\,,$$

where C ranges over all subsets of V of size 4 and the random variable X_C is defined as

$$X_C = \begin{cases} 1 & \text{if all pair of vertices in the set } C \text{ have an edge in between them} \\ 0 & \text{otherwise} \end{cases}.$$

We have $\mathbb{E}[X_C] = p^6$, since the probability of connecting all 4 vertices (using 6 edges) in the 4-tuple is p^6. So we have the expectation of Z :

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

We observe that

$$\mathbb{E}[Z] \to 0 \text{ when } p \ll n^{-2/3} \qquad \text{and} \qquad \mathbb{E}[Z] \to \infty \text{ when } p \gg n^{-2/3}.$$

Here, by $p \ll n^{-2/3}$, we mean that $\lim_{n\to\infty}(p/n^{-2/3}) = 0$ and $p \gg n^{-2/3}$ is defined similarly. We will prove that there is in fact a threshold phenomenon in the probability that G contains a copy of K_4. When $p \ll n^{-2/3}$, the probability that a random graph G generated according to model $\mathcal{G}_{n,p}$ contains a copy of K_4, goes to 0 as $n \to \infty$. On the other hand, when $p \gg n^{-2/3}$, this probability tends to 1.

Theorem 2.2 *Let G be generated randomly according to the model $\mathcal{G}_{n,p}$ graph. We have that:*

- *If $p \ll n^{-2/3}$, then $\mathbb{P}[G \text{ contains a copy of } K_4] \to 0$ as $n \to \infty$.*
- *If $p \gg n^{-2/3}$, then $\mathbb{P}[G \text{ contains a copy of } K_4] \to 1$ as $n \to \infty$.*

Proof: As above, we define the random variable Z,

$$Z = \text{ number of copies of } K_4 \text{ in } G = \sum_C X_C .$$

The case when $p \ll n^{-2/3}$ can be easily handled by Markov's inequality. We get that,

$$\mathbb{P}[Z > 0] = \mathbb{P}[Z \geq 1] \leq \frac{\mathbb{E}[Z]}{1}.$$

Since $\mathbb{E}[Z] \to 0$ as $n \to \infty$ when $p \ll n^{-2/3}$, we get that $\mathbb{P}[G \text{ contains a copy of } K_4] \to 0$.

When $p \gg n^{-2/3}$, we want to show that $\mathbb{P}[Z > 0] \to 1$, i.e., $\mathbb{P}[Z = 0] \to 0$. We use Chebyshev's inequality to prove this. We first compute the variance of Z.

$$\mathsf{Var}[Z] = \mathsf{Var}\left[\sum_C X_C\right] = \sum_C \mathsf{Var}[X_C] + \sum_{C \neq D} \mathsf{Cov}[X_C, X_D]$$

Since $\mathbb{E}[X_C] = p^6$, we have $\mathsf{Var}[X_C] = p^6 - p^{12}$. Also, for two distinct sets C and D, we consider four different cases depending on the number of vertices they share.

- **Case 1:** $|C \cap D| = 0$. Since no vertex is shared, X_C and X_D are independent and hence $\mathsf{Cov}[X_C, X_D] = 0$.

- **Case 2:** $|C \cap D| = 1$. Since the variables X_C and X_D depend on *pairs* of vertices in the sets C and D, and the two sets do not share any pair, we still have $\mathsf{Cov}[X_C, X_D] = 0$.

- **Case 3:** $|C \cap D| = 2$. Since C and D share a pair of vertices, there are 11 pairs which must all have edges between them in G, for both X_C and X_D to be 1. Thus, we have $\mathbb{E}[X_C X_D] = p^{11}$ and

$$\mathsf{Cov}[X_C, X_D] \;=\; \mathbb{E}[X_C X_D] - \mathbb{E}[X_C] \cdot \mathbb{E}[X_D] \;=\; p^{11} - p^{12}.$$

- **Case 4:** $|C \cap D| = 3$. in this case C and D share 3 pairs and thus there are 9 distinct pairs of vertices which must all have edges between them for both X_C and X_D to be 1. Thus,

$$\mathsf{Cov}[X_C, X_D] \;=\; \mathbb{E}[X_C X_D] - \mathbb{E}[X_C] \cdot \mathbb{E}[X_D] \;=\; p^{9} - p^{12}.$$

Also, there are $\binom{n}{6} \cdot \binom{6}{4}$ pairs C and D such that $|C \cap D| = 2$, and $\binom{n}{5} \cdot \binom{5}{4}$ pairs such that $|C \cap D| = 3$. Combining the above cases we have,

$$\begin{aligned}
\mathsf{Var}[Z] &= \sum_{C} \mathsf{Var}[X_C] + \sum_{C \neq D} \mathsf{Cov}[X_C, X_D] \\
&= \binom{n}{4} \cdot p^6(1-p^6) + \binom{n}{6} \cdot \binom{6}{4} \cdot (p^{11} - p^{12}) + \binom{n}{5} \cdot \binom{5}{4} \cdot (p^9 - p^{12}) \\
&= O(n^4 p^6) + O(n^6 p^{11}) + O(n^5 p^9).
\end{aligned}$$

Applying Chebyshev's inequality gives

$$\begin{aligned}
\mathbb{P}[Z = 0] \;\leq\; \mathbb{P}[|Z - \mathbb{E}[Z]| \geq \mathbb{E}[Z]] &\leq \frac{\mathsf{Var}[Z]}{(\mathbb{E}[Z])^2} \\
&= \frac{1}{\binom{n}{4}^2 \cdot p^{12}} \cdot \left(O(n^4 p^6) + O(n^6 p^{11}) + O(n^5 p^9) \right) \\
&= O\left(\frac{1}{n^4 p^6}\right) + O\left(\frac{1}{n^2 p}\right) + O\left(\frac{1}{n^3 p^3}\right).
\end{aligned}$$

For $p \gg n^{-2/3}$, all the terms on the right tend to 0 as $n \to \infty$. Hence, $\mathbb{P}[Z = 0] \to 0$ as $n \to \infty$. ■

The above analysis can be extended to any graph H of a fixed size. Let Z_H be the number of copies of H in a random graph G generated according to $G_{n,p}$. Using the previous analysis, we have $\mathbb{E}[Z_H] = \Theta\left(n^{|V(H)|} \cdot p^{|E(H)|}\right)$. Hence, $\mathbb{E}[Z] \to 0$ when $p \ll n^{-|V(H)|/|E(H)|}$ and $\mathbb{E}[Z] \to \infty$ when $p \gg n^{-|V(H)|/|E(H)|}$. Thus, it might be tempting to conclude that $p = n^{-|V(H)|/|E(H)|}$ is the threshold probability for finding a copy of H. However, con-

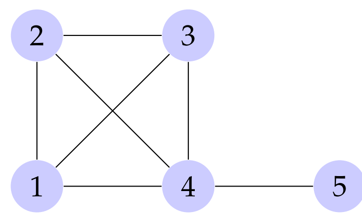

Figure 2: Subgraph H containing K_4

sider the graph in Figure 2. For this graph, we have $|V(H)|/|E(H)| = 5/7$. But for p such that $p \gg n^{-5/7}$ and $p \ll n^{-2/3}$, a random G is extremely unlikely to contain a copy of K_4 and thus also extremely unlikely to contain a copy of H. For an arbitrary graph H, it was shown by Bollobás **[?]** that the threshold probability is $n^{-\lambda}$, where

$$\lambda \;=\; \min_{H' \subseteq H} \frac{|V(H')|}{|E(H')|}\,.$$

3 Chernoff/Hoeffding Bounds

We now derive sharper concentration bounds for sums of independent random variables. We start by considering n *independent* Boolean random variables $X_1, ..., X_n$, where X_i takes value 1 with probability p_i and 0 otherwise. Let $Z = \sum_{i=1}^n X_i$. We set $\mu = \mathbb{E}[Z] = \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n \mu_i$. We will try to derive a bound on the probability $\mathbb{P}[Z \geq t]$ for $t = (1+\delta)\mu$. Using the fact that the function e^x is strictly increasing, we get that for $\lambda > 0$

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

We now have:

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda Z}\right] \;=\; \mathbb{E}\left[e^{\lambda(X_1+...X_n)}\right] \;=\; \mathbb{E}\left[\prod_{i=1}^n e^{\lambda X_i}\right] &\overset{\text{(independence)}}{=} \prod_{i=1}^n \mathbb{E}\left[e^{\lambda X_i}\right] \\
&= \prod_{i=1}^n \left[\mu_i e^{\lambda} + (1-\mu_i)\right] \\
&= \prod_{i=1}^n \left[1 + \mu_i(e^{\lambda} - 1)\right].
\end{aligned}$$

At this point, we utilize the simple but very useful inequality:

$$\forall x \in R, \;\; 1 + x \leq e^x.$$

Since all the quantities in the previous calculation are non-negative, we can plug the above inequality in the previous calculation and we get:

$$\mathbb{E}\left[e^{\lambda Z}\right] \;\leq\; \prod_{i=1}^{n} \exp\left((e^{\lambda}-1)\mu_i\right) \;=\; \exp\left((e^{\lambda}-1)\mu\right)$$

Thus, we get

$$\mathbb{P}\left[Z \geq (1+\delta)\mu\right] \;\leq\; \exp\left((e^{\lambda}-1)\mu - \lambda(1+\delta)\mu\right).$$

We now want to minimize the right hand-side of the above inequality, with respect to λ. Setting the derivative of the exponent to zero, we get

$$e^{\lambda}\mu - (1+\delta)\mu = 0 \quad \Rightarrow \quad \lambda = \ln(1+\delta)\,.$$

Using this value for λ, we get

$$\mathbb{P}\left[Z \geq (1+\delta)\mu\right] \;\leq\; \frac{\exp\left((e^{\lambda}-1)\mu\right)}{\exp\left(\lambda(1+\delta)\mu\right)} \;=\; \frac{e^{\delta\mu}}{(1+\delta)^{(1+\delta)\mu}} \;=\; \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}.$$

Exercise 3.1 *Prove similarly that*

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

(Note that $\mathbb{P}\left[Z \leq (1-\delta)\mu\right] = \mathbb{P}\left[e^{-\lambda Z} \geq e^{-\lambda(1-\delta)\mu}\right]$.) When $\delta \in (0,1)$, the bounds above expressions can be simplified further. It is easy to check that

$$\left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} \leq e^{-\delta^2\mu/3}, \;\; 0 < \delta < 1\,.$$

So we get:

$$\mathbb{P}\left[Z \geq (1+\delta)\mu\right] \leq e^{-\delta^2\mu/3}, \quad \text{for } 0 < \delta < 1.$$

Similarly:

$$\mathbb{P}\left[Z \leq (1-\delta)\mu\right] \leq e^{-\delta^2\mu/3}, \quad \text{for } 0 < \delta < 1.$$

Combining the two we get

$$\mathbb{P}\left[|Z-\mu| \geq \delta\mu\right] \leq 2 \cdot e^{-\delta^2\mu/3}, \quad \text{for } 0 < \delta < 1.$$

The above is only one of the proofs of the Chernoff-Hoeffding bound. A delighful paper by Mulzer **[?]** gives several other proofs with different applications.

References

[Bol81] Béla Bollobás, *Threshold functions for small subgraphs*, Mathematical Proceedings of the Cambridge Philosophical Society, vol. 90, Cambridge Univ Press, 1981, pp. 197–206.

[ER60] Paul Erdős and A Rényi, *On the evolution of random graphs*, Publ. Math. Inst. Hungar. Acad. Sci **5** (1960), 17–61.

[Mul18] Wolfgang Mulzer, *Five proofs of Chernoff's bound with applications*, Bulletin of EATCS **1** (2018), no. 124.