Mathematical Toolkit **Fall 2026**

Lecture 11: November 4, 2026

Lecturer: Avrim Blum (notes based in part on notes from Madhur Tulsiani)

1 Tail Inequalities

Tail inequalities are bounds on the probability mass of the tail of a distribution. They give us a way to say that it is unlikely that a random variable will take on a value that is too far away 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.

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,

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

Proof: Immediate from basic facts about expectation.

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{P}[X \geq t] \cdot \mathbb{E}[X|X \geq t] + \mathbb{P}[X < t] \cdot \mathbb{E}[X|X < t] \\ &\geq \mathbb{P}[X \geq t] \cdot t + 0 \end{aligned}$$

■

1.2 Chebyshev's Inequality

The variance of a random variable X is defined as

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

It is often convenient to simplify this as follows:

$$\begin{aligned}\mathbb{E}\left[(X-\mathbb{E}[X])^2\right] &= \mathbb{E}\left[X^2-2X\,\mathbb{E}[X]+\mathbb{E}[X]^2\right]\\ &= \mathbb{E}\left[X^2\right]-2\,\mathbb{E}[X]\,\mathbb{E}[X]+\mathbb{E}[X]^2\\ &= \mathbb{E}\left[X^2\right]-\mathbb{E}[X]^2\,.\end{aligned}$$

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

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

So, if two random variables are independent, their covariance is 0.

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

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

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

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

■

Chebyshev's inequality is particularly powerful when the overall random variable X can be decomposed into a sum of pairwise independent random variables X_i. (Later we will see even more powerful inequalities that can be used when X can be decomposed into a sum of *mutually* independent random variables X_i).

Proposition 1.3 *Let $X=X_1+\ldots+X_n$ where the X_i are pairwise independent. Then* $\mathsf{Var}[X]=\mathsf{Var}[X_1]+\ldots+\mathsf{Var}[X_n]$.

Proof:
$$\begin{aligned}\mathsf{Var}[X] &= \mathbb{E}[X^2]-\mathbb{E}[X]^2\\ &= \mathbb{E}\left[\sum_i\sum_j X_iX_j\right]-\left(\sum_i\mathbb{E}[X_i]\right)^2\\ &= \sum_i\mathbb{E}[X_i^2]+\sum_i\sum_{j\neq i}\mathbb{E}[X_iX_j]-\sum_i\mathbb{E}[X_i]^2-\sum_i\sum_{j\neq i}\mathbb{E}[X_i]\mathbb{E}[X_j]\\ &= \sum_i\mathsf{Var}[X_i]\quad\text{(using pairwise independence)}\end{aligned}$$

■

2 Coin tosses revisited

An unbiased coin is tossed n times. Let X_i be the indicator for the event that the ith toss is a heads, and let $X = \sum_i X_i$ be the total number of heads. So, by linearity of expectation, $\mathbb{E}[X] = n/2$. What is the chance that X is much larger than its expectation? Let us now compare the kind of bounds we get using Markov's and Chebyshev's inequalities.

2.1 Application of Markov's inequality

Using Markov's inequality we have,

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

2.2 Application of Chebyshev's inequality

Chebyshev's inequality gives a better bound on the chance of this occuring (and then we'll see tail bounds that give even stronger bounds). For this we need to calculate the variance of X, which we can do using Proposition 1.3.

In particular, for each X_i we have:

$$\begin{aligned} \mathsf{Var}\left[X_i\right] &= \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 \\ &= \frac{1}{2} - \frac{1}{4} = \frac{1}{4}. \end{aligned}$$

So, by Proposition 1.3, we have $\mathsf{Var}\left[X\right] = \frac{n}{4}$. Applying Chebyshev's inequality, we get:

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

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

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

Thus, Chebyshev's inequality gives a much stronger bound on the probability of X being larger than its expectation by $n/4$, 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 $\sqrt{\mathsf{Var}\left[X\right]}$, a quantity which is referred to as the *standard deviation* of the random variable X.

3 Threshold Phenomena in Random Graphs

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

For many properties, these graphs have what are called "threshold phenomena": for some function $f(n)$, for $p \ll f(n)$ (i.e., $\lim_{n\to\infty}(p/f(n)) = 0$) they almost surely do not have the property, and for $p \gg f(n)$ (i.e., $\lim_{n\to\infty}(p/f(n)) = \infty$) they almost surely do have the property. We will see one such property here, namely the property of containing a 4-clique.

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 $n \to \infty$.*

Proof: The easier direction here is part (1), which we can prove using Markov's inequality. For each of the $\binom{n}{4}$ sets S of four vertices, define indicator random variable X_S for the event that S is a clique. Let $X = \sum_S X_S$ be the total number of 4-cliques in the graph. Then $\mathbb{E}[X] = \sum_S \mathbb{E}[X_S] = \binom{n}{4}p^6$. If $p \ll n^{-2/3}$ then $\mathbb{E}[X] = o(1)$. By Markov's inequality, $\mathbb{P}[X \geq 1] \leq \mathbb{E}[X]/1 = o(1)$ as desired.

For part (2), if $p \gg n^{-2/3}$, then $\mathbb{E}[X] \gg 1$. But this in itself is not enough to guarantee that $\mathbb{P}[X = 0] = o(1)$. For this, we will use Chebyshev's inequality. In particular, plugging in $t = \mathbb{E}[X]$, Chebyshev's inequality tells us that $\mathbb{P}[X = 0] \leq \frac{\mathsf{Var}[X]}{\mathbb{E}[X]^2}$. So, if we can show that $\mathsf{Var}[X] = o(\mathbb{E}[X]^2)$, then $\mathbb{E}[X] \gg 1$ will indeed be sufficient.

We can break down $\mathsf{Var}[X]$ into its component pieces:

$$\mathsf{Var}[X] \quad = \quad \mathbb{E}[X^2] - \mathbb{E}[X]^2 \quad = \quad \sum_{S,S'} \mathbb{E}[X_S X_{S'}] - \mathbb{E}[X]^2.$$

Let us now consider a few cases for S, S'. First, if S and S' share at most one vertex in common then X_S and $X_{S'}$ are independent, so $\mathbb{E}[X_S X_{S'}] = \mathbb{E}[X_S]\mathbb{E}[X_{S'}]$. The sum of all of these is at most $\mathbb{E}[X]^2$ and so is covered by the $-\mathbb{E}[X]^2$ term in the variance. If S and S' share two vertices in common, then X_S and $X_{S'}$ are not independent, but we can use the fact that there are at most $O(n^6)$ such cases, and each case has $\mathbb{E}[X_S X_{S'}] = p^{11}$, so the overall contribution is $O(n^6 p^{11})$. This is $o(\mathbb{E}[X]^2)$ since $\mathbb{E}[X] = \Theta(n^4 p^6)$ and $p \gg n^{-2/3}$. If S and S' share three vertices in common, we have at most $O(n^5)$ such cases and each case has $\mathbb{E}[X_S X_{S'}] = p^9$, so the overall contribution is $O(n^5 p^9) = o(\mathbb{E}[X]^2)$. Finally, if S and S' share all four vertices, we get $\mathbb{E}[X] = o(\mathbb{E}[X]^2)$. So, we have $\mathsf{Var}[X] = o(\mathbb{E}[X]^2)$ as desired. ■