

COS 330: Great Ideas in Theoretical Computer Science

Fall 2026

Precept 4

Concentration is Key

RECAP SESSION

Today is a **recap session**: a short lecture collecting in one place the concentration inequalities you have seen in class, with their proofs, to help you prepare for next week's lectures. There are no problems to work on today; everything below is meant to be read (or re-read) at your own pace after precept.

Concentration Inequalities

We spent a lot of time in class discussing a very specific type of inequalities for the probability that a random variable deviates from its expectation. We are looking for inequalities that look like:

$$\Pr[|X - \mathbb{E}[X]| > t] \leq \text{something small.}$$

These inequalities are called *concentration inequalities* because they describe how tightly a random variable is concentrated around its expectation. We care about them because they help us quantify uncertainty, by establishing confidence intervals on estimates derived from random samples. For example, we can prove that algorithms will perform reliably except with some small probability.

The most basic concentration inequality is *Markov's inequality*, which applies to any non-negative random variable.

Theorem 1 (Markov's Inequality). Let X be a non-negative random variable. Then for any $t > 0$,

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}.$$

Proof. Let $\mathbf{1}[X \geq t]$ be the indicator random variable that is 1 if $X \geq t$, and 0 otherwise. Define $\mathbf{1}[X < t]$ similarly. Then we can write X as:

$$X = X \cdot \mathbf{1}[X < t] + X \cdot \mathbf{1}[X \geq t]$$

Now take expectations of both sides:

$$\mathbb{E}[X] = \mathbb{E}[X \cdot \mathbf{1}[X < t]] + \mathbb{E}[X \cdot \mathbf{1}[X \geq t]] \geq \mathbb{E}[X \cdot \mathbf{1}[X \geq t]] \geq t \cdot \mathbb{E}[\mathbf{1}[X \geq t]] = t \cdot \Pr[X \geq t],$$

where the first inequality follows because $X \cdot \mathbf{1}[X < t] \geq 0$ (recall that X is non-negative),

and the second inequality follows because $X \cdot \mathbf{1}[X \geq t] \geq t \cdot \mathbf{1}[X \geq t]$ (because if $X \geq t$, then X is at least t). Rearranging gives the desired result. $\square$

This is the best possible inequality we can use if all we know about X is its expectation. However, it is often too weak to be useful. In addition to the expectation, it is often the case that we also know the variance of X . The *Chebyshev inequality* uses this additional information to give a tighter bound.

Theorem 2 (Chebyshev's Inequality). Let X be a random variable with expectation μ and variance σ^2 . Then for any $t > 0$,

$$\Pr[|X - \mu| \geq t] \leq \frac{\sigma^2}{t^2}.$$

Proof. Apply Markov's inequality to the non-negative random variable $(X - \mu)^2$:

$$\Pr[|X - \mu| \geq t] = \Pr[(X - \mu)^2 \geq t^2] \leq \frac{\mathbb{E}[(X - \mu)^2]}{t^2} = \frac{\sigma^2}{t^2}.$$

$\square$

It might not be immediately clear why Chebyshev's inequality is stronger than Markov's, so let's look at an example. Suppose X is a random variable that describes the number of heads in n fair coin flips. Then $\mathbb{E}[X] = n/2$ and $\text{Var}(X) = n/4$ (see lecture notes). Suppose we want to bound the probability that X deviates from its expectation by more than $n/4$. Markov's inequality gives:

$$\Pr[X \geq n/2 + n/4] \leq \frac{\mathbb{E}[X]}{n/2 + n/4} = \frac{n/2}{n/2 + n/4} = \frac{2}{3}.$$

Chebyshev's inequality gives:

$$\Pr[|X - n/2| \geq n/4] \leq \frac{\text{Var}(X)}{(n/4)^2} = \frac{n/4}{n^2/16} = \frac{4}{n}.$$

For large n , this is much smaller than $2/3$. Thus, Chebyshev's inequality gives a much tighter bound on the probability of deviation from the mean. But is this the full story? Can we get a tighter bound still? The answer is yes, if we know more about the random variable.

If you stare at the proof of Chebyshev's inequality, you will see that it essentially reduces the problem to applying Markov's inequality to the random variable $(X - \mu)^2$. What if we applied the same idea, but to a different function of X ? One natural choice is a higher even power $(X - \mu)^{2k}$, which gives what is known as the *$2k$ -th moment inequality*¹. For the coin flip example, this can be pushed to a bound that is exponentially small in $\sqrt{n}$, which is much better than Chebyshev, but computing the moments takes a while, so we will not do it here.

Instead, let's go one step further: we can go beyond polynomial functions of $X - \mu$ and use any function we want. A particularly useful choice is the exponential function. This gives us the *Chernoff bound*.

¹Note that we use an even power, so that $(X - \mu)^{2k}$ is non-negative regardless of the value of $X - \mu$. This is important because Markov's inequality only applies to non-negative random variables, and we are assuming that X can take any real value.

Theorem 3 (Chernoff Bound 1). Let $X_1, X_2, \dots, X_n$ be independent Bernoulli random variables with parameters p_i , or equivalently, random variables whose value is 1 with probability p_i and 0 with probability $1 - p_i$. Let $X = \sum_{i=1}^n X_i$ and let $\mu = \mathbb{E}[X]$. Then for any $t > 0$,

$$\Pr[X \geq \mu + t] \leq \exp\left(-\frac{t^2}{2\mu + t}\right).$$

Proof. The idea is the same as before, but we apply Markov's inequality to the non-negative random variable $e^{\lambda X}$, for some $\lambda > 0$ to be chosen later. We have:

$$\Pr[X \geq \mu + t] = \Pr[e^{\lambda X} \geq e^{\lambda(\mu+t)}] \leq \frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda(\mu+t)}}.$$

Now we need to compute $\mathbb{E}[e^{\lambda X}]$. Since $X = \sum_{i=1}^n X_i$ and the X_i are independent, we have:

$$\mathbb{E}[e^{\lambda X}] = \mathbb{E}\left[e^{\lambda \sum_{i=1}^n X_i}\right] = \mathbb{E}\left[\prod_{i=1}^n e^{\lambda X_i}\right] = \prod_{i=1}^n \mathbb{E}[e^{\lambda X_i}],$$

where the last equality follows from the independence of the X_i . Now, since X_i is a Bernoulli random variable with parameter p_i , we have:

$$\mathbb{E}[e^{\lambda X_i}] = p_i e^{\lambda} + (1 - p_i) = 1 + p_i(e^{\lambda} - 1) \leq e^{p_i(e^{\lambda} - 1)},$$

where the last inequality follows from the fact that $1 + x \leq e^x$ for all x . Thus, we have:

$$\mathbb{E}[e^{\lambda X}] \leq \prod_{i=1}^n e^{p_i(e^{\lambda} - 1)} = e^{(e^{\lambda} - 1) \sum_{i=1}^n p_i} = e^{(e^{\lambda} - 1)\mu}.$$

Plugging this back into our bound gives:

$$\Pr[X \geq \mu + t] \leq \frac{e^{(e^{\lambda} - 1)\mu}}{e^{\lambda(\mu+t)}} = e^{(e^{\lambda} - 1)\mu - \lambda(\mu+t)}.$$

Now we need to choose λ to minimize the right-hand side. Some simple calculus shows that the optimal choice is $\lambda = \ln(1 + t/\mu)$. Plugging this in gives:

$$\Pr[X \geq \mu + t] \leq e^{t - (\mu+t) \ln(1+t/\mu)}.$$

Finally, we can use the inequality $\ln(1+x) \geq \frac{x}{1+x/2}$ for $x > 0$ to get:

$$\Pr[X \geq \mu + t] \leq e^{t - (\mu+t) \frac{t/\mu}{1+t/(2\mu)}} = e^{-\frac{t^2}{2\mu+t}}.$$

□

Why did we pick the exponential function? Notice that in the proof, the bulk of the work was in computing $\mathbb{E}[e^{\lambda X}]$. This quantity is also called the *moment generating function* of X , because it generates all the moments of X . You can see this by looking at the Taylor series expansion of $e^{\lambda X} = 1 + \lambda X + \frac{(\lambda X)^2}{2!} + \frac{(\lambda X)^3}{3!} + \dots$. So in a way, we are using all the moments of X at once,

instead of just one of them.

Now let's get back to our coin flip example. In the setting of the above theorem, we have n independent Bernoulli random variables with parameter $p = 1/2$, so $\mu = n/2$. If we want to bound the probability that X deviates from its expectation by more than $n/4$, we can set $t = n/4$ in the above theorem to get:

$$\Pr[X \geq n/2 + n/4] \leq \exp\left(-\frac{(n/4)^2}{2(n/2) + n/4}\right) = \exp\left(-\frac{n^2/16}{5n/4}\right) = \exp(-n/20).$$

This is yet another improvement: compared to the $4/n$ we got from Chebyshev, we now get a probability that is exponentially small in n .

To close out this section, let's ask one last question: does the Chernoff bound only apply in the above setting? Note that in our proof, we only used the fact that the X_i are independent and that we could bound their moment generating functions. So, as long as we have independent random variables whose moment generating functions we can bound, we can get a Chernoff bound like the above. For example, we can get a similar bound for sums of independent random variables that are bounded in $[0, 1]$, which you can prove by an almost identical argument. We can also get a similar bound for the probability that X is smaller than its expectation by more than t , so $\Pr[X \leq \mu - t]$.

So here is the most general form of the Chernoff bound that we will state.

Theorem 4 (Chernoff Bound). Let $X_1, X_2, \dots, X_n$ be independent random variables such that $0 \leq X_i \leq 1$ for all i . Let $X = \sum_{i=1}^n X_i$ and let $\mu = \mathbb{E}[X]$. Then for any $t > 0$,

$$\Pr[X \geq \mu + t] \leq \exp\left(-\frac{t^2}{2\mu + t}\right).$$

and

$$\Pr[X \leq \mu - t] \leq \exp\left(-\frac{t^2}{2\mu}\right).$$

Combining the two above inequalities, we get a two-sided bound:

Theorem 5 (Two-Sided Chernoff Bound). Let $X_1, X_2, \dots, X_n$ be independent random variables such that $0 \leq X_i \leq 1$ for all i . Let $X = \sum_{i=1}^n X_i$ and let $\mu = \mathbb{E}[X]$. Then for any $0 < t \leq \mu$,

$$\Pr[|X - \mu| \geq t] \leq 2 \exp\left(-\frac{t^2}{3\mu}\right).$$