COS 330: Great Ideas in Theoretical Computer Science

Fall 2025

Precept 4

My Concentration

Learn

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. □

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 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 $(X-\mu)^2$:

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

□

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? Let's try to do this for some $2k$-th power of $X-\mu$[1].

Theorem 3 ($2k$-th Moment Inequality)**.** Let X be a random variable with expectation μ. Then for any $t > 0$ and any integer $k \geq 1$,

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

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

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

□

In our example of the number of heads in n fair coin flips, we can compute the $2k$-th moment of $X-\mu$ as follows:

$$\mathbb{E}[(X-\mu)^{2k}] = \sum_{i=0}^{n}(i-n/2)^{2k} \cdot \Pr[X=i] = \sum_{i=0}^{n}(i-n/2)^{2k} \cdot \binom{n}{i} \cdot (1/2)^n.$$

This is a bit messy, but it can be shown that $\mathbb{E}[(X-\mu)^{2k}] \leq (k^2 n)^k$ by doing some calculations and approximations[2]. Plugging this into the $2k$-th moment inequality gives:

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

[1] Note that we are using 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.

[2] You can obtain this using only Stirling's approximation, which says that $m! \approx \sqrt{2\pi m}(m/e)^m$.

For large enough n, this is much smaller than the Chebyshev bound of $4/n$, for any constant k we get a probability that decreases with the k-th power of n. In fact, you can try to pick the best k to optimize the bound; if you pick $k = \sqrt{n}/4e$, you get a bound of

$$\Pr[|X - n/2| \geq n/4] \leq \left(\frac{16k^2}{n}\right)^k = \left(\frac{16(\sqrt{n}/4e)^2}{n}\right)^{\sqrt{n}/4e} = \left(\frac{1}{e^2}\right)^{\sqrt{n}/4e} = e^{-\sqrt{n}/2e}.$$

The above is even better, this is an exponentially small probability in $\sqrt{n}$. However, we can do even better. 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*.

Theorem 4 (Chernoff Bound 1)**.** Let $X_1, X_2, \ldots, X_n$ 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 $e^{\lambda X} = 1 + \lambda X + \frac{(\lambda X)^2}{2!} + \frac{(\lambda X)^3}{3!} + \ldots$. So in a way, we are applying the $2k$-th moment inequality with all moments at once.

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 an improvement over the $2k$-th moment bound, giving an exponentially small probability 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 $\Pr[X \geq \mu - t]$.

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

Theorem 5 (Chernoff Bound)**.** Let $X_1, X_2, \ldots, X_n$ 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 6 (Two-Sided Chernoff Bound)**.** Let $X_1, X_2, \ldots, X_n$ 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 $t > 0$,

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

Practice

Problem 1

Consider the following model of a random graph, known as the Erdős-Rényi model $G(n, p)$. A graph G sampled from $G(n, p)$ has n vertices, and each pair of vertices is connected by an edge with probability p, independently of all other pairs.

(a) What is the expected degree of a vertex in a graph sampled from $G(n, p)$?

(b) Show that if $p = \frac{500 \ln n}{n-1}$, then with probability at least $1 - 2n^{-2/3}$, every vertex in a graph sampled from $G(n, p)$ has degree between $450 \ln n$ and $550 \ln n$.

Challenge

Problem 1

Suppose there is some algorithm A that given a decision task, outputs the correct answer with probability $1/2 + \epsilon$, for some $\epsilon > 0$ (so this algorithm is just slightly better than random guessing). We want to boost the accuracy of this algorithm to be at least $1 - \delta$, for some $\delta \in (0, 1)$. So, we run the algorithm n times independently and take the majority answer as the final answer. How large does n need to be to achieve this accuracy?