Lecture 3

Prof. Nick Harvey *University of British Columbia*

1 The Chernoff Bound

1.1 Formal Statement

Let $X_1, \ldots, X_n$ be independent random variables such that X_i always lies in the interval $[0,1]$. Define $X = \sum_{i=1}^n X_i$ and $p_i = \mathrm{E}[X_i]$. Let $\mu_{\min} \leq \sum_i \mathrm{E}[X_i] \leq \mu_{\max}$.

Theorem 1 *For all $\delta > 0$,*

$$\Pr[X \geq (1+\delta)\mu_{\max}] \overset{(a)}{\leq} \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{\mu_{\max}} \overset{(b)}{\leq} \begin{cases} e^{-\delta^2\mu_{\max}/3} & (\text{if } \delta \leq 1) \\ e^{-(1+\delta)\ln(1+\delta)\mu_{\max}/4} & (\text{if } \delta \geq 1) \\ e^{-\delta\mu_{\max}/3} & (\text{if } \delta \geq 1) \end{cases}$$

$$\Pr[X \leq (1-\delta)\mu_{\min}] \overset{(c)}{\leq} \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu_{\min}} \overset{(d)}{\leq} e^{-\delta^2\mu_{\min}/2}.$$

Inequalities (c) and (d) are only valid for $\delta < 1$, but $\Pr[X \leq (1-\delta)\mu_{\min}] = 0$ if $\delta > 1$.

References: McDiarmid Theorem 2.3, Mitzenmacher-Upfal Theorems 4.4 and 4.5, Motwani-Raghavan Theorem 4.1 and 4.2, Klenke Exercise 5.2.1, Wikipedia.

We will only prove inequality (a) of the Chernoff bound. The other inequalities are discussed in the appendix.

1.2 Proof of inequality (a)

The Chernoff bounds would not be true without the assumption that the X_is are independent. What special properties do independent random variables have? One basic property is that

$$\mathrm{E}[A \cdot B] = \mathrm{E}[A] \cdot \mathrm{E}[B] \tag{1}$$

for any independent random variables A and B.

But the Chernoff bound has nothing to do with *products* of random variables, it is about *sums* of random variables. So one trick we could try is to convert sums into products using the exponential function. Fix some parameter $\theta > 0$ whose value we will choose later. Define

$$\begin{aligned} Y_i &= \exp(\theta X_i) \\ Y &= \exp(\theta X) = \exp(\theta \sum_i X_i) = \prod_i \exp(\theta X_i) = \prod_i Y_i. \end{aligned}$$

It is easy to check that, since the X_is are independent, the Y_is are also independent. Therefore, by (1),

$$\mathrm{E}[Y] = \prod_i \mathrm{E}[Y_i]. \tag{2}$$

So far this all seems quite good. We want to prove that X is small, which is equivalent to proving Y is small. Using (2), we can do this by showing that the $\mathrm{E}[Y_i]$ terms are small. Perhaps we can somehow show $\mathrm{E}[Y_i]$ is small by comparing it to $\mathrm{E}[X_i] = p_i$?

If we were forgetting the rules of probability, we might be tempted to say that $\mathrm{E}\left[e^{\theta X_i}\right]$ equals $e^{\theta\,\mathrm{E}[X_i]}$, but that is false. We might remember one useful probability trick called Jensen's inequality that says $f(\mathrm{E}[A]) \leq \mathrm{E}[f(A)]$ for any random variable A and any convex function f. Applying this with $f(x) = e^{tx}$, we see that

$$e^{\theta p_i} \;=\; e^{\theta\,\mathrm{E}[X_i]} \;\leq\; \mathrm{E}\left[e^{\theta X_i}\right] \;=\; \mathrm{E}[Y_i]. \tag{3}$$

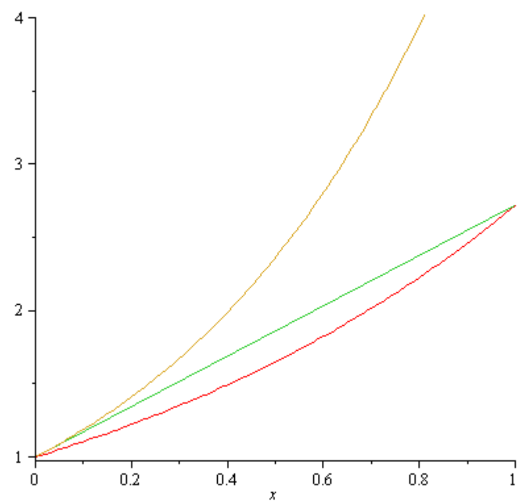
So we get a *lower bound* on $\mathrm{E}[Y_i]$ in terms of p_i, but we actually wanted an *upper bound*.

Claim 3 gives the desired upper bound; it shows that the inequality in (3) can almost be reversed. The proof is easy once we have the following convexity fact.

Claim 2

$$\exp(tx) \;\leq\; 1 + (e^{\theta} - 1)x \;\leq\; \exp((e^{\theta} - 1)x) \qquad \forall x \in [0, 1],$$

You might be convinced by the following "proof by picture". An actual proof in the appendix.



Claim 3 *Let $\theta \in \mathbb{R}$ be arbitrary. Then $\mathrm{E}[Y_i] \leq \exp((e^{\theta} - 1)p_i)$.*

Proof: Recall that $\mathrm{E}[Y_i] = \mathrm{E}[\exp(\theta X_i)]$. The main idea is as follows. Although we cannot "pull the expectation inside the exponential", we can use Claim 2 to approximate the exponential by a linear function, then "pull the expectation inside" via linearity of expectation, then finally switch back to an exponential function.

The formal argument is

$$\mathrm{E}\left[e^{\theta X_i}\right] \;\leq\; \mathrm{E}\left[1 + (e^{\theta} - 1)X_i\right] \;=\; 1 + (e^{\theta} - 1)\,\mathrm{E}[X_i] \;\leq\; \exp((e^{\theta} - 1)\,\mathrm{E}[X_i]),$$

where both inequalities follow from Claim 2. The claim follows since $p_i = \mathrm{E}[X_i]$. $\square$

Now we are ready to prove the inequality (a) of the Chernoff bound.

$$\begin{aligned}
\Pr[\, X \geq (1+\delta)\mu_{\max} \,] &= \Pr\left[\, \exp(\theta X) \geq \exp\left(\theta(1+\delta)\mu_{\max}\right) \,\right] \quad \text{(by monotonicity)} \\
&\leq \frac{\mathrm{E}[\, \exp(\theta X) \,]}{\exp(\theta(1+\delta)\mu_{\max})} \quad \text{(by Markov's inequality)} \\
&= \frac{\prod_i \mathrm{E}[\, Y_i \,]}{\exp(\theta(1-\delta)\mu_{\min})} \quad \text{(by (2))} \\
&\leq \frac{\prod_i \exp((e^\theta - 1)p_i)}{\exp(\theta(1+\delta)\mu_{\max})} \quad \text{(by Claim 3).}
\end{aligned}$$

Gathering everything inside one exponential we get

$$\Pr[\, X \geq (1+\delta)\mu_{\max} \,] \;\leq\; \exp\Big((e^\theta - 1) \sum_i p_i - \theta(1+\delta)\mu_{\max} \Big).$$

Finally, substituting $\theta = \ln(1+\delta)$ and using $\sum_i p_i \leq \mu_{\max}$ proves inequality (a).

2 Example: Balls and Bins

Let's illustrate the power of the Chernoff bound by using it to analyze a fundamental balls-and-bins problem. But first of all we introduce the other (almost trivial) tool which is often used in conjunction with the Chernoff bound. The reason why this lemma is so useful is that it does *not* require that the events are independent.

Lemma 4 (The Union Bound) *Let $\mathcal{E}_1, \ldots, \mathcal{E}_k$ be any collection of events. Then*

$$\Pr[\, \mathcal{E}_1 \cup \cdots \cup \mathcal{E}_k \,] \;\leq\; \sum_{i=1}^{k} \Pr[\, \mathcal{E}_k \,].$$

References: Mitzenmacher-Upfal Lemma 1.2, Durrett Theorem 1.1.1, Wikipedia.

Many interesting problems can be modeled using simple problems involving balls and bins. Today we are interested in the bin that has the most balls. Suppose there are n bins. We repeat the following experiment n times: throw a ball into a uniformly chosen bin. (The experiments are mutually independent.) Let B_i be the number of balls in bin i. What is $\max_i B_i$?

Theorem 5 *Assume $n \geq 3$ and let $\alpha = 10 \ln n / \ln \ln n$. Then $\max_i B_i \leq \alpha$ with probability at least $1 - 1/n$.*

This theorem is optimal up to constant factors. It is known that $\max_i B_i \geq \ln n / \ln \ln n$ with probability at least $1 - 1/n$. (See, e.g., Lemma 5.12 in Mitzenmacher-Upfal.)

Proof: Let us focus on the first bin. Let $X_1, \ldots, X_n$ be indicator random variables where X_j is 1 if the jth ball lands in the first bin. Obviously $\mathrm{E}[\, B_1 \,] = \sum_j \mathrm{E}[\, X_j \,] = 1$. What is the probability that this bin has more than α balls?

We will analyze this using the Chernoff bound. This is possible since B_1 is a sum of independent random variables, each of which take values in $[0, 1]$. Applying inequality (b) of the Chernoff bound (with $\alpha = 1 + \delta$), we obtain

$$\Pr[\, X \geq \alpha \, \mathrm{E}[\, X \,] \,] \;\leq\; \exp\Big(-\mathrm{E}[\, X \,] \cdot \alpha \ln(\alpha)/4 \Big) \qquad \forall \alpha \geq 2.$$

Applying this to B_1 and using $\mathrm{E}[B_1] = 1$, we obtain

$$\Pr[B_1 \geq \alpha] \;\leq\; \exp(-\alpha \ln(\alpha)/4).$$

A rough calculation gives

$$\alpha \cdot \ln \alpha \;=\; 10\frac{\ln n}{\ln \ln n} \cdot \ln\Big(\frac{10 \ln n}{(\underbrace{\ln \ln n}_{\text{negligible}})}\Big) \;\approx\; 10\frac{\ln n}{\ln \ln n} \cdot \ln \ln n \;=\; 10 \ln n.$$

More precisely, one can show that $\alpha \ln \alpha \geq 8 \ln n$ for all $n \geq 1$. Plugging that in,

$$\Pr[B_1 \geq \alpha] \;\leq\; \exp\big(-\alpha \ln(\alpha)/4\big) \;\leq\; \exp\big(-2 \ln n\big) \;\leq\; n^{-2}. \tag{4}$$

So bin 1 is unlikely to have more than α balls. By symmetry, the same is true for each bin *individually*.

Now we will show the same is true for every bin *simultaneously*. The key tool is the union bound. Let $\mathcal{E}_i$ be the event $\{B_i \geq \alpha\}$. Then

$$\begin{aligned}
\Pr[\text{any bin } i \text{ has } B_i \geq \alpha] \;&=\; \Pr[\cup_i \mathcal{E}_i] \;\leq\; \sum_{i=1}^{n} \Pr[\mathcal{E}_i] \qquad \text{(by the union bound)} \\
&=\; \sum_{i=1}^{n} \Pr[B_i \geq \alpha] \;\leq\; \sum_{i=1}^{n} n^{-2} \qquad \text{(by (4))} \\
&=\; n \cdot n^{-2} \;=\; n^{-1}.
\end{aligned}$$

Thus, with probability at least $1 - 1/n$, all bins have less than α balls. $\square$

3 Congestion Minimization

One of the classically important areas in algorithm design and combinatorial optimization is *network flows*. A central problem in that area is the maximum flow problem. We now look at a generalization of this problem.

An instance of the problem consists of a directed graph $G = (V, A)$ and a sequence $(s_1, t_1), \ldots, (s_k, t_k)$ of pairs of vertices. Let $n = |V|$. (It is not crucial that the graph be directed; the problem is equally interesting in undirected graphs. However in network flow problems it is often more convenient to look at directed graphs. Feel free to think about whatever variant you find easier.)

A natural question to ask is: do there exist paths P_i from s_i to t_i for every i such that these paths share no arcs? This is called the **edge-disjoint paths** problem. Quite remarkably, it is NP-hard even in the case $k = 2$, assuming the graph is directed. For undirected graphs, it is polynomial time solvable if k is a fixed constant, but NP-hard if k is a sufficiently large function of n.

We will look at a variant of this problem called the **congestion minimization** problem. The idea is to allow each arc to be used in multiple paths, but not too many. The number of paths using a given arc is the "congestion" of that arc. We say that a solution has congestion C if it is a collection of paths P_i from s_i to t_i, where each arc is contained in at most C of the paths. The problem is to find the minimum value of C such that there is a solution of congestion C. This problem is still NP-hard, since determining if $C = 1$ is the edge-disjoint paths problem.

We will look at the congestion minimization problem from the point of view of approximation algorithms. Let OPT be the minimum congestion of any solution. We would like to give an algorithm which can produce a solution with congestion at most $\alpha \cdot OPT$ for some $\alpha \geq 1$. This factor α is the called the approximation factor of the algorithm.

Theorem 6 *There is an algorithm for the congestion minimization problem with approximation factor* $O(\log n / \log\log n)$*.*

To design such an algorithm we will use linear programming. We write down an integer program (IP) which captures the problem exactly, relax that to a linear program (LP), then design a method for "rounding" solutions of the LP into solutions for the IP.

The Integer Program. Writing an IP formulation of an optimization problem is usually quite simple. That is indeed true for the congestion minimization problem. However, we will use an IP which you might find rather odd: our IP will have exponentially many variables. This will simplify our explanation of the rounding.

Let $\mathcal{P}_i$ be the set of *all* paths in G from s_i to t_i. (Note that $|\mathcal{P}_i|$ may be exponential in n.) For every path $P \in \mathcal{P}_i$, we create a variable x^i_P. This variable will take values only in $\{0,1\}$, and setting it to 1 corresponds to including the path P in our solution.

The integer program is as follows

$$\begin{array}{lllll}
\min & C & & & \\
\text{s.t.} & \displaystyle\sum_{P \in \mathcal{P}_i} x^i_P & & = 1 & \forall i = 1,\ldots,k \\
& \displaystyle\sum_{i} \sum_{P \in \mathcal{P}_i \text{ with } a \in P} x^i_P & & \leq C & \forall a \in A \\
& x^i_P \in \{0,1\} & & & \forall i = 1,\ldots,k \text{ and } P \in \mathcal{P}_i
\end{array}$$

The last constraint says that we must decide for every path whether or not to include it in the solution. The first constraint says that the solution must choose exactly one path between each pair s_i and t_i. The second constraint ensures that the number of paths using each arc is at most C. The optimization objective is to find the minimum value of C such that a solution exists.

Every solution to the IP corresponds to a solution for the congestion minimization problem with congestion C, and vice-versa. Thus the optimum value of the IP is OPT, which we previously defined to be the minimum congestion of any solution to the original problem.

A Appendix

A.1 Proof of Claim 2

Consider the first inequality $e^{\theta x} \leq 1 + (e^\theta - 1)x$ for all $x \in [0,1]$. This follows from Inequality 3 in the Notes on Convexity Inequalities. by setting $c = e^\theta$.

The second inequality $1 + (e^\theta - 1)x \leq \exp((e^\theta - 1)x)$ follows from our favorite inequality $1 + x \leq \exp(x)$, which is Inequality 1 in the Notes on Convexity Inequalities.

A.2 Proof of inequality (b), $\delta \in [0,1]$

Claim 7 *Suppose* $\delta \in [0,1]$. *Then* $(1+x)\ln(1+x) - x \geq x^2/3$.

Proof: Note that the LHS and RHS both vanish at $x = 0$. So the claim holds if the derivative of the LHS is at least the derivative of the RHS on the interval $[0,1]$. By simple calculus,

$$\frac{d}{dx}\big[(1+x)\ln(1+x) - x\big] = \ln(1+x) \qquad \text{and} \qquad \frac{d}{dx}x^2/3 = 2x/3.$$

At $x = 0$, $\ln(1+x)$ equals $2x/3$. At $x = 1$, we have $\ln(1+x) = \ln(2) > 0.69$ and $2x/3 < 0.67$. Since $\ln(1+x)$ is concave, we have $\ln(1+x) \geq 2x/3$ for all $x \in [0,1]$. $\square$

Corollary 8 *For all* $\delta \in [0,1]$,

$$\Pr\left[\textstyle\sum_{i=1}^k X_i \geq (1+\delta)\mu_{\max}\right] \;\leq\; \exp\big(-(\delta^2/3)\mu_{\max}\big).$$

Proof: Claim 7 implies that $\left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right) \leq e^{-\delta^2/3}$. $\square$

A.3 Proof of inequality (b), $\delta \geq 1$

Claim 9 *Define*

$$\begin{aligned} f(x) &= (1+x)\ln(1+x) - x \\ g(x) &= (1+x)\ln(1+x)/4 \\ h(x) &= x/3 \end{aligned}$$

Then $f(x) > g(x) > h(x)$ *for* $x \geq 1$.

Proof: First let us consider the point $x = 1$. We have

$$\begin{aligned} f(1) &= 2\ln(2) - 1 \;>\; 0.38 \\ g(1) &= \ln(2)/4 \;\approx\; 0.346 \\ h(1) &= 1/3. \end{aligned}$$

The claim follows if we can show that

$$\frac{d}{dx}f(x) \;\geq\; \frac{d}{dx}g(x) \;\geq\; \frac{d}{dx}h(x) \qquad \forall x \geq 1.$$

Simple calculus gives

$$\frac{d}{dx} f(x) = \ln(1+x) \qquad \frac{d}{dx} g(x) = (1+\ln(1+x))/4 \qquad \text{and} \qquad \frac{d}{dx} h(x) = 1/3.$$

Clearly $y \geq (1+y)/4$ for $y \geq 1/3$, so $\ln(1+x) \geq (1+\ln(1+x))/4$ for $x \geq e^{1/3}-1 \approx 0.395$. This shows $\frac{d}{dx} f(x) \geq \frac{d}{dx} g(x)$ for $x \geq 1$. On the other hand, $x \geq 1$ yields $(1+\ln(1+x))/4 \geq (1+\ln(2))/4 \approx 0.423$. This shows $\frac{d}{dx} g(x) \geq \frac{d}{dx} h(x)$ for $x \geq 1$. □

Corollary 10 *For all $\delta \geq 1$,*

$$\Pr\Big[\textstyle\sum_{i=1}^{k} X_i \geq (1+\delta)\mu_{\max}\Big] \;\leq\; \begin{cases} \exp\big(-(1+\delta)\ln(1+\delta)\mu_{\max}/4\big) \\ \exp\big(-\delta\mu_{\max}/3\big). \end{cases}$$

Proof: Claim 9 implies that

$$\left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right) \;\leq\; \exp\big(-(1+\delta)\ln(1+\delta)/4\big) \;\leq\; \exp\big(-\delta/3\big).$$

□

A.4 Proof of inequality (d)

Claim 11 *Suppose $x \in [0,1]$. Then $(1-x)\ln(1-x)+x \geq x^2/2$.*

Proof: Note that the LHS and RHS both vanish at $x=0$. So the claim holds if the derivative of the LHS is at least the derivative of the RHS on the interval $[0,1)$. By simple calculus,

$$\frac{d}{dx}\big[(1-x)\ln(1-x)+x\big] = -\ln(1-x) \qquad \text{and} \qquad \frac{d}{dx} x^2/2 = x.$$

The linear approximation of $-\ln(1-x)$ at $x=0$ is

$$x \cdot \tfrac{d}{dx}\big(-\ln(1-x)\big)\Big|_{x=0} = x \cdot \big(\tfrac{1}{1-x}\big)\Big|_{x=0} = x.$$

Furthermore, $-\ln(1-x)$ is convex on $[0,1)$ because its second derivative is $1/(1-x)^2 \geq 0$. Thus $-\ln(1-x) \geq x$ on $[0,1)$. □

Corollary 12 *For all $\delta \in [0,1]$,*

$$\Pr\Big[\textstyle\sum_{i=1}^{k} X_i \leq (1-\delta)\mu_{\min}\Big] \leq \exp\big(-(\delta^2/2)\mu_{\min}\big)$$

Proof: Claim 11 implies that $\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right) \leq e^{-\delta^2/2}$. □

A.5 Proof of inequality (c)

The argument is very similar to the proof of (a). Assume $\theta < 0$. Then

$$
\begin{aligned}
\Pr\left[\, X \le (1-\delta)\mu_{\min} \,\right] &= \Pr\left[\exp(\theta X) \ge \exp\big(\theta(1-\delta)\mu_{\min}\big)\right] \quad \text{(by monotonicity and } \theta < 0\text{)} \\
&\le \frac{\mathrm{E}\left[\, \exp(\theta X) \,\right]}{\exp(\theta(1-\delta)\mu_{\min})} \quad \text{(by Markov's inequality)} \\
&\le \frac{\prod_i \exp((e^\theta - 1)p_i)}{\exp(\theta(1-\delta)\mu_{\min})} \qquad \text{(by Claim 3)} \\
&= \exp\Big((e^\theta - 1)\sum_i p_i - \theta(1-\delta)\mu_{\min}\Big) \\
&\le \exp\Big((e^\theta - 1)\mu_{\min} - \theta(1-\delta)\mu_{\min}\Big) \qquad \text{(since } e^\theta - 1 < 0 \text{ and } \mu_{\min} \le \textstyle\sum_i p_i\text{)} \\
&= \Big(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\Big)^{\mu_{\min}},
\end{aligned}
$$

by choosing $\theta = \ln(1-\delta) < 0$.