| **Mathematical Toolkit** | **Autumn 2019** |
|---|---|

Lecture 16: November 21, 2019

Lecturer: Madhur Tulsiani

# 1 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 $\lambda > 0$

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

We now have:

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

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

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

We now want to minimize the right hand-side of the above inequality, with respect to $\lambda$. 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 $\lambda$, 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 1.1** *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 [Mul18] gives several other proofs with different applications.

## 1.1 Coin tosses once more

We will now compare the above bound with what we can get from Chebyshev's inequality. Let's assume that $X_1, ..., X_n$ are independent coin tosses, with $\mathbb{P}\left[X_i = 1\right] = \frac{1}{2}$. We want to get a bound on the value of $Z = \sum_{i=1}^{n} X_i$. Using Chebyshev's inequality, we get that

$$\mathbb{P}\left[|Z-\mu| \geq \delta\mu\right] \;\leq\; \frac{\mathsf{Var}\left[Z\right]}{\delta^2\mu^2}.$$

And since in this particular case we have that $\mathsf{Var}[Z] = n/4$ and $\mu = n/2$, we get that

$$\mathbb{P}[|Z-\mu| \geq \delta\mu] \;\leq\; \frac{1}{\delta^2 n}.$$

The above bound is only inversely polynomial in $n$, while the Chernoff-Hoeffding bound gives

$$\mathbb{P}[|Z-\mu| \geq \delta\mu] \;\leq\; 2\cdot\exp\left(-\delta^2 n/6\right),$$

which is exponentially small in $n$. This fact will prove very useful when taking a union bound over a large collection of events, each of which may be bounded using a Chernoff-Hoeffding bound. For example, consider the case where for $m$ sets $S_1,\ldots,S_m \subseteq [n]$, we define

$$Z_{S_i} \;=\; \sum_{j\in S_i} X_j\,.$$

While the variables $Z_{S_1},\ldots,Z_{S_m}$ are *not* necessarily indepdent, each of these is a sum of few $X_j$ variables, which are indepdent. Thus, we can say that for any $S_i$,

$$\mathbb{P}\left[\left|Z_{S_i} - \frac{|S_i|}{2}\right| \;\geq\; t\right] \;\leq\; 2\cdot\exp\left(-2t^2/(3\,|S_i|)\right) \;\leq\; 2\cdot\exp\left(-2t^2/(3n)\right)\,,$$

where we choose $\delta = 2t/\,|S_i|$ so that $\delta\,|S_i|\,/2 = t$. Thus, by a union bound over all $i\in[m]$, we get that

$$\mathbb{P}\left[\exists i\in[m].\; \left|Z_{S_i} - \frac{|S_i|}{2}\right| \;\geq\; t\right] \;\leq\; 2m\cdot\exp\left(-2t^2/(3n)\right)\,.$$

Thus, when $t = \sqrt{3n\cdot\ln m}$, the probability of the above event is at most $2/m$. Check that it just using Chebyshev's inequality does not allow for such a strong bound on the probability of the above event.

Note that the above calculation used the following union bound

**Exercise 1.2** *Let $E_1,\ldots,E_k$ be events on the same outcome space $\Omega$. Then*

$$\mathbb{P}[E_1\cup\cdots\cup E_k] \;\leq\; \sum_{i=1}^{k}\mathbb{P}[E_i]\,.$$

## 2 Balanced Allocations

We consider the following problem of allocating jobs to servers: We are given a set of $n$ servers $1,\ldots,n$ and $m$ jobs arrive one by one. We seek to assign each job to one of the servers so that the loads placed on all servers are as balanced as possible.

In developing simple, effective load balancing algorithms, randomization often proves to be a useful tool. We consider two approaches for this problem:

- **Random Choice**: one possible way to distribute the jobs is to simply place each job on a random server, chosen independently of the previous allocations.
- **Two Random Choices**: For each job, we choose two servers independently and uniformly at random and place the job on the server with less load (breaking ties arbitrarily).

We will show that using two random choices significantly reduces the maximum load on any server. For the entire analysis, we will work with the case when $m = n$. The analysis easily extends to an arbitrary $m$, but it easier to appreciate the bounds when $m = O(n)$ (and in particular when $m = n$).

It is convenient to think of the above in terms of the so called "balls and bins" model. Each job can be thought of a s ball and each server is a bin. We think of assigning job $j$ to a server $i$ as throwing the $j^{th}$ ball in bin $i$. The load of a server is the same as the number of balls in the corresponding bin.

### 2.1 Random choice

Suppose $Z_i$ = number of balls in bin $i$. We can write

$$Z_i = \sum_j X_{ij}, \quad \text{where} \quad X_{ij} = \begin{cases} 1 & \text{if ball } j \text{ is thrown in bin } i \\ 0 & \text{otherwise} \end{cases} .$$

Then, we have that each $Z_i$ is a sum of $m(= n)$ independent random variables with $\mathbb{E}[Z_i] = 1$. Let $t = \frac{3 \ln n}{\ln \ln n}$. By Chernoff/Hoeffding bounds, we have that for each $i$,

$$\mathbb{P}[Z_i \geq t] \;\leq\; \left(\frac{e}{t}\right)^t .$$

To bound the maximum load in across all bins, we use a union bound to say that

$$\mathbb{P}[\exists i \in [n].\, Z_i \geq t] \;\leq\; \sum_{i=1}^{n} \mathbb{P}[Z_i \geq t] \;\leq\; n \cdot \left(\frac{e}{t}\right)^t ,$$

which is at most $\frac{1}{n}$ for the above value of $K$. Hence, with probability at least $1 - \frac{1}{n}$, the maximum number of balls in a bin is at most $\frac{3 \ln n}{\ln \ln n}$.

### 2.2 The power of two random choices

It is a somehwat surprising result (which can still be proved using Chernoff bounds) that two random choices can reduce the maximum load to $O(\ln \ln n)$. The proof technique is due to Azar et al. [ABKU94, ABKU99] and various applications were explored by Mitzenmacher in his thesis [Mit96]. We will not discuss the proof of this result, but you are encouraged to look up the analysis from the notes in 2016 (or from the book by Mitzenmacher and Upfal).

## References


[ABKU94] Yossi Azar, Andrei Z Broder, Anna R Karlin, and Eli Upfal, *Balanced allocations*, Proceedings of the twenty-sixth annual ACM symposium on Theory of computing, ACM, 1994, pp. 593–602. 4

[ABKU99] ______, *Balanced allocations*, SIAM journal on computing **29** (1999), no. 1, 180–200. 4

[Mit96] Michael David Mitzenmacher, *The power of two random choices in randomized load balancing*, Ph.D. thesis, PhD thesis, Graduate Division of the University of California at Berkley, 1996. 4

[Mul18] Wolfgang Mulzer, *Five proofs of Chernoff's bound with applications*, CoRR **abs/1801.03365** (2018). 2