Mathematical Toolkit **Autumn 2025**

Lecture 14: November 13, 2025

Lecturer: Madhur Tulsiani

1 Chernoff/Hoeffding Bounds

Let's recall the bounds we proved in the previous lecture for sums of independent Bernoulli random variables.

Theorem 1.1 *Let $X_1, ..., X_n$, be n* independent *Bernoulli random variables, where X_i takes value 1 with probability p_i. Let $Z = \sum_{i=1}^{n} X_i$ and let $\mu = \mathbb{E}[Z] = \sum_{i=1}^{n} p_i$. Then, we have for any $\delta > 0$,*

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

Moreover, when $\delta \in (0,1)$ both the above expressions can be bounded by $e^{-\delta^2\mu/3}$.

We'll discuss below a few examples of cases where the exponentially small probability bounds are indeed required. We will need the following union bound

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

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

1.1 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.

1.1.1 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 $\frac{3 \ln n}{\ln \ln n}$.

1.1.2 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).

1.2 Random Max 3-SAT

A few lectures ago, we considered the problem of Max 3-SAT, where given a collection of m clauses in n variables, the goal is choose values for the variables, satisfying as many clauses as possible. An instance φ of the problem is given by

$$\varphi \;\equiv\; (C_1, \ldots, C_m)\,,$$

where each C_i is a clause of the form $C_i = (l_{i_1} \vee l_{i_2} \vee l_{i_3})$ and each l_{i_j} is in turn x_{i_j} or its negation $\overline{x}_{i_j}$. We assume that each clause contains three *distinct* variables. We prove that *for any instance* a random assignment to the variables satisfies $7m/8$ of the clauses in expectation, and thus there always esists an assignment satisfying at least $7m/8$ of the clauses. In fact, we also saw how such an assignment can be found efficiently and deterministically.

Now we revisit the problem, chosing the *instance φ at random*. We will show that with high probability of the choice of the instance, even the best assignment to the variables can satisfy at most $(7/8 + \varepsilon)m$ clauses, and thus the trivial algorithm from earlier finds essentially the best possible assignment.

To define what we mean by a random instance, consider fixing in advance the "structure" of the instance i.e., we fix sets $S_1, \ldots, S_m$ of size 3 each, such that the clause C_i will involve variables in the set S_i. However, when picking the clause C_i, we will decide independently at random for each of the three variables, wether the literal l_{i_j} equals the variable x_{i_j} or its negation $\overline{x}_{i_j}$. This choice is made independently at random for each of the clauses, and each of the 3 variables in the clause. We can now prove the following.

Proposition 1.3 *Let φ be a random instance as above, with $m \geq 10/\varepsilon^2$ clauses. Then*

$$\mathop{\mathbb{P}}_{\varphi}\left[\text{Optimal assignment satisfies more than } \left(\frac{7}{8}+\varepsilon\right)\cdot m \text{ clauses}\right] \;\leq\; e^{-n}\,.$$

Proof: Fix an assignment $\alpha \in \{0,1\}^n$ for the variables, and consider the random variable Z_α corresponding to the number of clauses satisfied by the assignment α (which is a function of the the choice of the instance φ). We can write

$$Z_\alpha \;=\; Y_{1,\alpha} + \cdots + Y_{m,\alpha}\,,$$

where $Y_{i,\alpha} = 1$ if the i-th clause C_i is satisfied by the assignment α and 0 othersise. Since the choices of literals in different clauses are independent, the variables $Y_{1,\alpha}, \ldots, Y_{m,\alpha}$ are independent Bernoulli random variables. Moreover, $\mathbb{E}[Y_{i,\alpha}] = 7/8$ since for each clause there is exactly one literal pattern which falsifies C_i. Thus, by Chernoff-Hoeffding bounds

$$\mathbb{P}_{\varphi}\left[Z_\alpha \geq \left(\frac{7}{8}+\varepsilon\right)\cdot m\right] \;=\; \mathbb{P}_{\varphi}\left[Z_\alpha \geq \left(1+\frac{8\varepsilon}{7}\right)\cdot\frac{7}{8}\cdot m\right] \;\leq\; \exp\left(-\left(\frac{8\varepsilon}{7}\right)^2\cdot\frac{7m}{8}\cdot\frac{1}{3}\right)$$
$$= \;\exp\left(-\left(\frac{8\varepsilon^2}{21}\right)\cdot m\right) \;\leq\; e^{-2n}\,.$$

Taking a union bound over all asignments, we have that

$$\mathbb{P}_{\varphi}\left[\exists\alpha\in\{0,1\}^n.\;\; Z_\alpha \geq \left(\frac{7}{8}+\varepsilon\right)\cdot m\right] \leq 2^n\cdot e^{-2n} \;\leq\; e^{-n}\,,$$

which completes the proof. ■

1.3 Interlude: Dealing with ± 1 random variables

A common variant of the above calculations also arises for the case of random variables which take values in the set $\{-1,1\}$ instead of the set $\{0,1\}$. Let $Y_1,\ldots,Y_n$ be independent random variables, which take values in the set $\{-1,1\}$ with probability $1/2$ each (such random variables are called Rademacher random variables), and let $Z = \sum_{i=1}^n Y_i$. We can easily apply the results for Bernoulli random variables to this case by defining $X_i = (1+Y_i)/2$. Note that the variables $X_1,\ldots,X_n$ are now independent Bernoulli random variables (with parameter $1/2$). Considering $Z' = \sum_{i=1}^n X_i$, we can write

$$Z' \;=\; \sum_{i=1}^n X_i \;=\; \sum_{i=1}^n \frac{1+Y_i}{2} \;=\; \frac{n}{2}+\frac{Z}{2}\,.$$

We can thus analyze deviations from the mean (which is 0) for the variable Z as

$$\mathbb{P}\left[|Z|\geq t\right] \;=\; \mathbb{P}\left[\left|Z'-\frac{n}{2}\right|\geq\frac{t}{2}\right] \;=\; \mathbb{P}\left[\left|Z'-\frac{n}{2}\right|\geq\frac{t}{n}\cdot\frac{n}{2}\right] \;\leq\; 2\cdot\exp\left(-t^2/6n\right)\,.$$

1.4 Finding low-discrepancy assignments

We consider the problem of assigning $X_1,\ldots X_n \in \{-1,1\}$ such that the signs in each of the m sets $S_1,\ldots,S_m \subseteq [n]$ are "balanced". We defined the imbalance, also known as the discrepancy, in the set S_i as

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

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,

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

using the bound we proved earlier. By a union bound over all $i \in [m]$, we get that

$$\mathbb{P}\left[\exists i \in [m].\ |Z_{S_i}| \;\geq\; t\right] \;\leq\; 2m \cdot \exp\left(-t^2/(6n)\right) .$$

Thus, when $t = \sqrt{12n \cdot \ln m}$, the probability of the above event is at most $2/m$. Check that just using Chebyshev's inequality does not allow for such a strong bound on the probability of the above event.

2 Hoeffding bounds

Previously, we used bounds for Bernoulli variables to also bound the deviations for a sum of independent Rademacher random variables. Hoeffding's inequality actially also gives a bound for *weighted* sums of independent Rademacher variables.

Lemma 2.1 *Let $X_1, \ldots, X_n$ be independent Rademacher random variables, and let $a \in \mathbb{R}^n$. Then,*

$$\mathbb{P}\left[\left|\sum a_i X_i\right| \geq t\right] \;\leq\; 2 \cdot \exp(-t^2/2\,\|a\|^2)\,.$$

Proof: As before, we use Markov's inequality to say that for any $\lambda > 0$,

$$\mathbb{P}\left[\sum a_i X_i \geq t\right] \;=\; \mathbb{P}\left[\exp(\lambda \sum_i a_i X_i) \geq \exp(\lambda t)\right] \;\leq\; \frac{\mathbb{E}\left[\exp(\lambda \sum_i a_i X_i)\right]}{\exp(\lambda t)}\,.$$

Using independence, we have that $\mathbb{E}\left[\exp(\lambda \cdot \sum_i a_i X_i)\right] = \prod_i \mathbb{E}\left[e^{\lambda \cdot a_i X_i}\right]$. We now calculate the individual expectations as

$$\mathbb{E}\left[e^{\lambda \cdot a_i X_i}\right] \;=\; \frac{1}{2} \cdot \left(e^{\lambda \cdot a_i} + e^{-\lambda \cdot a_i}\right) \;\leq\; e^{\lambda^2 \cdot a_i^2/2}\,,$$

where we used the ineuality $(e^x + e^{-x})/2 \leq e^{x^2/2}$ which can be proved (for example) using Taylor expansions. Plugging the bounds back in, we get

$$\mathbb{P}\left[\sum a_i X_i \geq t\right] \;\leq\; \frac{\prod_i e^{\lambda^2 \cdot a_i^2/2}}{e^{\lambda \cdot t}} \;=\; \exp(\lambda^2 \cdot \|a\|^2/2 - \lambda \cdot t)\,.$$

Optimizing over λ gives the choice $\lambda = \|a\|^2/t$, which gives the bound

$$\mathbb{P}\left[\sum a_i X_i \geq t\right] \;\leq\; \exp(-\lambda^2 \cdot \|a\|^2/2)\,.$$

By switching from a to $-a$, we also get the same bound for $\mathbb{P}\left[\sum a_i X_i \leq -t\right]$. ■

The same proof can also be used to analyze sums of independent random variables $X_1, \ldots, X_n$ where X_i takes values in the interval $[a_i, b_i]$. For any such variable X_i, one can prove the inequality

$$\mathbb{E}\left[\exp(\lambda(X_i - \mathbb{E}[X_i]))\right] \;\leq\; \exp(\lambda^2 \cdot (b_i - a_i)^2/8)\,.$$

Exercise 2.2 *Use the above inequality to prove that for independent random variables* $X_1, \ldots, X_n$ *with* $X_i \in [a_i, b_i]$*, we have that*

$$\mathbb{P}\left[\left|\sum_i X_i - \mathbb{E}\left[\sum_i X_i\right]\right| \geq t\right] \;\leq\; 2 \cdot \exp\left(-\frac{2t^2}{\sum_i (b_i - a_i)^2}\right)\,.$$

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. 3

[ABKU99] ———, *Balanced allocations*, SIAM journal on computing **29** (1999), no. 1, 180–200. 3

[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. 3