CS 761: Randomized Algorithms **University of Waterloo: Winter 2025**

Lecture 3

January 21, 2025

Instructor: Sepehr Assadi

Topics of this Lecture

1 Balls and Bins Experiments

We continue our study of probabilistic analysis of algorithms and concentration inequalities. We shall use the following simple running example in this lecture to review "power" of different concentration inequalities.

Problem 1 (**Maximum Load in Balls-and-Bins**)**.** We are give a collection of n balls and n bins. We throw each ball *independently* to one of the bins chosen *uniformly at random*. We define the **load** of a bin as the number of balls thrown to that bin in this experiment. What is the asymptotically maximum load of a bin in this process with probability at least $2/3$?

More precisely, our goal is to find the smallest possible upper $T(n)$ (as a function of n) such that with probability at least $2/3$, the load of *every* bin in this process is $O(T(n))$.

Let us first see how we can simplify our analysis. The original question involves a quantifier *for-all*—as in, *every* bin should have $O(T(n))$ balls. But, can we turn this into a *for-each* quantifier—for a *fixed* bin, say, bin one, upper bound the load? It is easy to see that if we simply bound the load of a fixed bin with constant probability, it cannot tell us anything meaningful about the maximum load across all bins[1]. However, if we can bound the load of each bin with a much higher probability, then we can extend this to bound the maximum load as well. In particular,

$$\Pr(\text{max load is} > k) = \Pr(\text{load of bin 1 is} > k \text{ OR load of bin 2 is} > k \text{ OR} \cdots \text{ OR load of bin } n \text{ is} > k).$$

But we can now use a very simple—but extremely helpful—inequality, the **union bound**, which is as follows.

Fact 1. *For any two events A and B,* $\Pr(A \text{ OR } B) \leqslant \Pr(A) + \Pr(B)$.

[1] It is an easy calculation to see that with probability $(1 - 1/n)^n \approx 1/e$, the first bin receives zero balls but of course the maximum load is not zero!

So, applying union bound to the above terms gives us that

$$\Pr\left(\text{max load is} > k\right) \leqslant \sum_{i=1}^{n} \Pr\left(\text{load of bin } i \text{ is} > k\right) = n \cdot \Pr\left(\text{a fixed bin, say bin one, has load} > k\right);$$

in the last equality, we used the fact that the distribution of load of all bins is the same.

Consequently, if we define L to be the load of any fixed bin—say, bin one—then, to answer Problem 1, we can instead find a function $T(n)$ such that

$$\Pr\left(L > T(n)\right) \leqslant \frac{1}{3n}. \tag{1}$$

Thus, for the rest of the proof, we will focus on solving this problem. Let us define indicator random variables $B_1, \ldots, B_n$ where for any $i \in [n]$, $B_i = 1$ iff the i-th ball is sent to the fixed bin corresponding to L. As such, we have $L = \sum_{i=1}^{n} B_i$ and hence, using the linearity of expectation,

$$\mathbb{E}\left[L\right] = \mathbb{E}\left[\sum_{i=1}^{n} B_i\right] = \sum_{i=1}^{n} \mathbb{E}\left[B_i\right] = \sum_{i=1}^{n} \Pr\left(\text{ball } i \text{ is sent to bin one}\right) = \sum_{i=1}^{n} \frac{1}{n} = 1.$$

Hence, we know that in expectation, load of every bin is 1. But what can this tell us about Eq (1), namely, the probability that load of a bin is far away from its expectation? For this, we need to take a quick detour.

1.1 Concentration Inequalities

When working with random variables, perhaps the easiest way to "summarize" the variable is to focus on its expected value. However, expected value on its own can often be misleading: for instance, consider a random variable which is 0 with probability 1/2 and is 1 with the remaining value. Expected value of X is 1/2 but of course we do not 'expect' X to ever take the value of 1/2! This, and many other examples, suggest that summarizing a random variable just down to its expectation may lose too much information.

On the other extreme, a random variable can be uniquely identified by its probability distribution:

$$\mathbb{P}_X : k \rightarrow [0, 1] \qquad \text{such that} \qquad \mathbb{P}_X(a) = \Pr(X = a).$$

Yet, the distribution of even very simple random variables can be quite cumbersome to work with. Consider, for instance, the simple example of throwing a fair coin 100 times and defining X to be the number of heads. Here, for every $a \in \{0, \ldots, 100\}$,

$$\Pr(X = a) = \binom{100}{a} \cdot 2^{-100},$$

which, even in this simple form, is rather tedious to work with.

Concentration inequalities are a saving grace between these two extremes: morally speaking (but not strictly speaking true), they allow us to extract (perhaps, the most) "important" information about the distribution of our random variables, without getting to compute the very precise distribution itself. More accurately, they allow us to bound the probability of *deviation* of a random variable from its expectation (as a function of its distance from the expectation).

We will study various concentration inequalities in the course of this term, as they arise quite frequently in the analysis of randomized algorithms (and way beyond). This lecture, includes three of the most basic and highly applicable ones. Let us now go back to our original balls and bins questions.

1.2 Concentration Inequalities for Balls and Bins

Equipped with the definition of concentration inequalities, it is easy to see that bounding Eq (1) is precisely the topic of concentration inequalities.

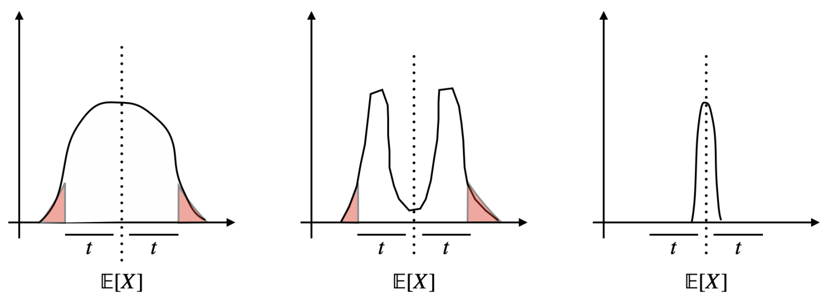

Figure 1: An illustration of three different random variables and their probability distributions. Here, all these variables have the same expected value. Moreover, the first two variables show roughly the same concentration for the particular choice of t; i.e., for the first two variables, the probability that their values are more than t away from the expectation is almost the same (the probability is the part shaded in red) even though the distributions are quite different; however, the third variable is much more concentrated.

Markov Bound

The simplest and most basic variant of concentration results is **Markov's inequality** or **Markov bound**:

Proposition 2 (**Markov Bound**)**.** *For a* non-negative *random variable X and $t > 0$,*

$$\Pr\left(X \geqslant t \cdot \mathbb{E}\left[X\right]\right) \leqslant \frac{1}{t}.$$

Proof. Let $\mu := \mathbb{E}\left[X\right]$. We can use the law of total conditional probabilities to have:

$$\begin{aligned}\mathbb{E}\left[X\right] &= \mathbb{E}\left[X \mid X \geqslant t \cdot \mu\right] \cdot \Pr\left(X \geqslant t \cdot \mu\right) + \mathbb{E}\left[X \mid X < t \cdot \mu\right] \cdot \Pr\left(X < t \cdot \mu\right) \\ &\geqslant t \cdot \mu \cdot \Pr\left(X \geqslant t \cdot \mu\right) + 0.\end{aligned}$$

(the first term since we conditioned on $X \geqslant t \cdot \mu$ and the second term since X is non-negative)

Thus, $\Pr\left(X \geqslant t \cdot \mathbb{E}\left[X\right]\right) = \Pr\left(X \geqslant t \cdot \mu\right) \leqslant 1/t$, otherwise the RHS above will be larger than the LHS. □

Markov bound only bounds the *upper tail* of the distribution[2]: the probability that a random variable takes value t times larger than its expectation is at most $1/t$. This is a basic but extremely useful property. One can alternatively state the Markov bound as follow (by simply picking $t = b/\mathbb{E}\left[X\right]$ in Proposition 2).

Corollary 3. *For a* non-negative *random variable X and $b > 0$,*

$$\Pr\left(X \geqslant b\right) \leqslant \frac{\mathbb{E}\left[X\right]}{b}.$$

Note that a random variable X, which, with probability $\mathbb{E}\left[X\right]/b$ takes the value b and otherwise is 0 will be a tight example for Markov bound; i.e., one cannot expect to improve Markov bound in general.

Remark. Even though Markov bound may sound almost trivial (and it is indeed straightforward), it is the basis for proving all other concentration inequalities that we use in this course; moreover, Markov bound is used one way or another in analysis of almost every randomized algorithm.

[2]Although one can use it to bound the lower tail in special cases as well, but the bounds there are generally very weak.

For our application to the balls and bins experiment, we can thus have for any $k \geqslant 1$,

$$\Pr(L \geqslant k) \leqslant \frac{\mathbb{E}[L]}{k} = \frac{1}{k},$$

and thus to bound this probability with $1/3n$, we need to take $k = 3n$ – notice that this is entirely useless because we only have n balls to begin with and so $L \leqslant n$ happens with probability 1 already!

Chebyshev's Inequality

We consider our second concentration inequality: **Chebyshev's inequality**. Unlike Markov bound that only required a knowledge of the expected value of the random variable to bound its deviation probability, Chebyshev's inequality applies to the settings in which we could additionally bound the *variance* of the random variable as well.[3]

Recall that for a random variable X, **variance** of X is:

$$\mathrm{Var}[X] := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2.$$

Chebyshev inequality allows us to bound deviation of a random variable based on its variance.

Proposition 4 (**Chebyshev's Inequality**). *For any random variable X and $t > 0$,*

$$\Pr(|X - \mathbb{E}[X]| \geqslant t \cdot \mathbb{E}[X]) \leqslant \frac{\mathrm{Var}[X]}{\mathbb{E}[X]^2 \cdot t^2}.$$

Proof. Define a new random variable $Y := (X - \mathbb{E}[X])^2$. Clearly, Y is non-negative. Moreover, $|X - \mathbb{E}[X]| \geqslant t \cdot \mathbb{E}[X]$ if and only if $Y = (X - \mathbb{E}[X])^2 \geqslant t^2 \cdot \mathbb{E}[X]^2$. Hence,

$$\Pr(|X - \mathbb{E}[X]| \geqslant t \cdot \mathbb{E}[X]) = \Pr\left(Y \geqslant t^2 \cdot \mathbb{E}[X]^2\right) \leqslant \frac{\mathbb{E}[Y]}{\mathbb{E}[X]^2 \cdot t^2} \qquad \text{(by Markov bound of Corollary 3)}$$

$$= \frac{\mathrm{Var}[X]}{\mathbb{E}[X]^2 \cdot t^2}. \qquad \text{(as } \mathbb{E}[Y] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathrm{Var}[X] \text{ by definition)}$$

□

A useful variant of Chebyshev's inequality is the following (by simply picking $t = b/\mathbb{E}[X]$ in Proposition 4).

Corollary 5. *For a random variable X and $b > 0$,*

$$\Pr(|X - \mathbb{E}[X]| \geqslant b) \leqslant \frac{\mathrm{Var}[X]}{b^2}.$$

For our balls and bins experiment, to apply Chebyshev's inequality, we need to bound $\mathrm{Var}[L]$. We have,

$$\mathrm{Var}[L] = \mathrm{Var}\left[\sum_{i=1}^{n} B_i\right] = \sum_{i=1}^{n} \mathrm{Var}[B_i] \leqslant \sum_{i=1}^{n} \mathbb{E}[B_i] = \mathbb{E}[L] = 1,$$

where in the second equality, we used the fact that $B_1, \ldots, B_n$ are *independent* and hence variance of their sum is equal to their sum of variances, as proven in the following fact[4].

[3] As we showed earlier, Markov bound *can* be tight for certain random variables; thus, naturally whenever we need a stronger bound we should show that our variable satisfies additional guarantees than what is only required by Markov bound.

[4] Recall that in general, we do not have linearity of variance, only linearity of expectation ...

Fact 6. *For any two* independent *random variables* X, Y*:*

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y].$$

Proof. We have,

$$\begin{aligned}
\mathrm{Var}[X+Y] &= \mathbb{E}\left[(X+Y)^2\right] - (\mathbb{E}[X+Y])^2 \\
&= \mathbb{E}\left[X^2\right] + \mathbb{E}[Y]^2 + 2\cdot\mathbb{E}[X\cdot Y] - (\mathbb{E}[X])^2 - (\mathbb{E}[Y])^2 - 2\cdot\mathbb{E}[X]\cdot\mathbb{E}[Y] \\
&\qquad\qquad\text{(by expanding the terms)} \\
&= \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 + \mathbb{E}\left[Y^2\right] - (\mathbb{E}[Y])^2 \qquad \text{(as } \mathbb{E}[X\cdot Y] = \mathbb{E}[X]\cdot\mathbb{E}[Y] \text{ when } X \perp Y\text{)} \\
&= \mathrm{Var}[X] + \mathrm{Var}[Y].
\end{aligned}$$

□

Thus,

$$\Pr(L \geqslant k) \leqslant \Pr(|L - \mathbb{E}[L]| \geqslant k-1) \leqslant \frac{\mathrm{Var}[L]}{(k-1)^2} = \frac{1}{(k-1)^2}.$$

This time, we will get something non-trivial: by setting $k = \sqrt{3n}+1$, we obtain that $\Pr(L \geqslant k) \leqslant 1/3n$ which is the bound we wanted. In other words, using Chebyshev's inequality allows us to bound $T(n) = O(\sqrt{n})$. While this is non-trivial, as we shall see this is still very far from the *right* bounds for the problem.

Chernoff Bound

We now switch to one of the strongest concentration inequalities that we will encounter in this course, the **Chernoff bound**. To be able to use Chernoff bound for bounding deviation of a random variable, we need a much stronger knowledge than variance in Chebyshev's inequality and expectation in Markov bound; we now need to know that the random variable X is a *sum of bounded-value independent random variables.*

Proposition 7 (Chernoff Bound). *Suppose $X_1, \ldots, X_n$ are* independent *random variables in* $[0,1]$ *and $X = \sum_i X_i$. Then, for any $t \geqslant 1$,*

$$\Pr(|X - \mathbb{E}[X]| \geqslant t\cdot\mathbb{E}[X]) \leqslant 2\cdot\exp\left(-\frac{t\cdot\mathbb{E}[X]}{3}\right),$$

and, for any $\varepsilon \in (0,1]$,

$$\Pr(|X - \mathbb{E}[X]| \geqslant \varepsilon\cdot\mathbb{E}[X]) \leqslant 2\cdot\exp\left(-\frac{\varepsilon^2\cdot\mathbb{E}[X]}{3}\right).$$

The above bound is often called the *multiplicative* Chernoff bound (as opposed to the *additive* Chernoff bound that we shall see later and is weaker than the above[5]). The proof of Proposition 7 is not as straightforward as those of Markov bound or Chebyshev's inequality, but it is not also particularly hard. For interested reader (and to showcase the main ideas), we present a proof of a simpler form in Section 3.

Let us now apply Chernoff bound to our problem. We have that $L = B_1 + \ldots + B_n$, each $B_i \in [0,1]$, and B_i's are independent of each other. So, we can indeed apply Chernoff bound to L and obtain that

$$\Pr(L \geqslant k) \leqslant \Pr(|L - \mathbb{E}[L]| \geqslant k-1) \leqslant 2\cdot\exp\left(-\frac{(k-1)\cdot\mathbb{E}[X]}{3}\right) = 2\cdot\exp\left(-\frac{(k-1)}{3}\right).$$

[5] There is also a more general version of Chernoff bound which is rather unwieldy to use in practice (at least initially), but is a bit stronger than the above bounds (especially for "small" values of $\mathbb{E}[X]$); however, the above bounds are quite more convenient for us to use and as such we will ignore the stronger bound for now.

By setting $k = 3 \cdot \ln(6n) + 1$, we have,

$$\Pr(L \geqslant k) \leqslant 2 \cdot \exp(-\ln(6n)) = 2 \cdot \frac{1}{6n} = \frac{1}{3n}.$$

Thus, Chernoff bound allows us to say $T(n) = O(\log n)$, exponentially stronger than the bound obtained via Chebyshev's inequality.

1.3 Optimal Bounds for Balls and Bins?

The bound of $O(\log n)$ on maximum load obtained via Chernoff bound is quite close to the optimal solution but it is *not* the right answer still. While we can use a stronger version of Chernoff bound to obtain the optimal bound via the use of a "generic" concentration inequality, we are going to instead use a more direct approach: by calculating the probability of the event explicitly ourself (as somewhat of a last resort).

Remark. Think of working directly with the probability distribution of a random variable as the "ultimate solution" – technically speaking, this approach *always* give you the optimal answer as you will be calculating the exact probability. But in most cases, these probabilities are just too cumbersome or downright impossible to calculate, hence the need for slightly more general concentration inequalities.

Also, despite the fact that Markov, Chebyshev, and Chernoff bounds are without a doubt the most used concentration inequalities in the analysis of the algorithms, they are only the tip of the iceberg among the vast array of concentration inequalities known. See the suggested reading materials and in particular the excellent book of

> *Devdatt P. Dubhashi and Alessandro Panconesi, Concentration of Measure for the Analysis of Randomised Algorithms.*

that provides a more detailed overview of concentration inequalities in this context.

We have,

$$\Pr(L \geqslant k) = \sum_{b=k}^{n} \Pr(L = b) = \sum_{b=k}^{n} \binom{n}{b} \cdot \left(\frac{1}{n}\right)^{b} \cdot \left(1 - \frac{1}{n}\right)^{n-b}.$$

Notice that in this approach, we already have a rather unwieldily calculation to work with. So, even in this case, it would help to use some inequalities although they will be more ad-hoc compared to a generic concentration inequality. In particular, we are going to calculate the above probability rather differently by upper bounding it first.

$$\begin{aligned}
\Pr(L \geqslant k) &= \Pr(\text{there exists some } k \text{ balls that are mapped to the bin}) \\
&\qquad (\text{a sufficient and necessary condition for } L \geqslant k \text{ is that some } k \text{ balls are mapped to this bin}) \\
&\leqslant \sum_{\text{balls } a_1, \ldots, a_k} \Pr(a_1, \ldots, a_k \text{ are mapped to the bin}) \\
&\qquad (\text{by union bound where the sum ranges over all possible choices of picking } k \text{ balls out of } n) \\
&= \sum_{\text{balls } a_1, \ldots, a_k} \left(\frac{1}{n}\right)^{k} \qquad (\text{as the probability of a fixed } k \text{ balls mapping to the bin is } n^{-k}) \\
&= \binom{n}{k} \cdot \left(\frac{1}{n}\right)^{k}. \qquad (\text{the number of ways of picking } k \text{ balls out of } n \text{ is } \tbinom{n}{k})
\end{aligned}$$

So far, in these calculations, we already used one inequality where we applied the union bound. We are going to apply yet another inequality[6] (which in general is a highly useful inequality to know).

Fact 8. *For any integers $a \geqslant b$ (here, $e \approx 2.73...$ is the natural number),*

$$\left(\frac{a}{b}\right)^b \leqslant \binom{a}{b} \leqslant \left(\frac{e \cdot a}{b}\right)^b.$$

Proof. We have,

$$\binom{a}{b} = \frac{a!}{b! \cdot (a-b)!} = \underbrace{\frac{a}{b} \cdot \frac{(a-1)}{(b-1)} \cdot \frac{(a-2)}{(b-2)} \cdots \frac{(a-b+1)}{1}}_{b \text{ terms in total}} \geqslant \left(\frac{a}{b}\right)^b,$$

because each of the b terms is at least a/b. This proves the LHS. For the RHS, we have,

$$\binom{a}{b} \leqslant \frac{a^b}{b!},$$

because each of the terms in the nominator of the first equation is at most a. Furthermore, by writing the Taylor expansion of e^b, we have that

$$e^b = \sum_{k=0}^{\infty} \frac{b^k}{k!} \geqslant \frac{b^b}{b!}. \qquad \text{(as this is just one term, } k = b\text{, in the whole series)}$$

Plugging in $b! \geqslant (b/e)^b$ in the above inequality, we get the RHS as well. ☐

Continuing our calculations from before, we now have,

$$\begin{aligned}
\Pr(L \geqslant k) &\leqslant \binom{n}{k} \cdot \frac{1}{n^k} && \text{(as calculated above)} \\
&\leqslant \left(\frac{e \cdot n}{k}\right)^k \cdot \frac{1}{n^k} && \text{(by Fact 8)} \\
&= \left(\frac{e}{k}\right)^k \\
&= \frac{e^k}{e^{k \cdot \ln k}} && (\ln k \text{ is the log of } k \text{ in base } e) \\
&= \exp(-k \cdot (\ln k - 1)) \\
&= \exp(-k \cdot \ln(k/e)).
\end{aligned}$$

We can now pick $k = \frac{e \cdot \ln n}{\ln \ln n}$ to have,

$$\begin{aligned}
k \cdot \ln(k/e) &= \left(\frac{e \cdot \ln n}{\ln \ln n}\right) \cdot \ln\left(\frac{\ln n}{\ln \ln n}\right) \\
&= \left(\frac{e \cdot \ln n}{\ln \ln n}\right) \cdot (\ln \ln n - \ln \ln \ln n) \\
&\geqslant \left(\frac{e \cdot \ln n}{\ln \ln n}\right) \cdot \frac{1}{2} \cdot \ln \ln n && \text{(as for sufficiently large } n\text{, } \ln \ln n > 2 \cdot \ln \ln \ln n) \\
&\geqslant n^{1.3} \geqslant 3n. && \text{(for sufficiently large } n)
\end{aligned}$$

[6] The reason we are pointing out the inequalities explicitly is that we need to be extra careful with them if our goal is to compute an optimal bound – any loose inequality may negate all our effort of calculating the probability explicitly.

Plugging in all these—honestly, rather tedious[7]—calculations back into our bounds, we get that

$$\Pr\left(L \geqslant \frac{e \cdot \ln n}{\ln \ln n}\right) \leqslant \frac{1}{3n},$$

for sufficiently large n. This allows us to say $T(n) = O(\log n / \log\log n)$ also holds for our problem in Eq (1).

Now, is this the right answer? Yes:

$$\begin{aligned}
\Pr(L \geqslant k) &\geqslant \Pr(L = k) \\
&= \binom{n}{k} \cdot \frac{1}{n^k} \cdot \left(1 - \frac{1}{n}\right)^{n-k} \\
&\geqslant \binom{n}{k} \cdot \frac{1}{n^k} \cdot \frac{1}{4} && \text{(as } (1-1/n)^m \geqslant (1-1/n)^n \geqslant 1/4 \text{ for all } m \leqslant n \text{ and } n \geqslant 2) \\
&\geqslant \left(\frac{n}{k}\right)^k \cdot \frac{1}{n^k} \cdot \frac{1}{4} && \text{(by Fact 8)} \\
&= \frac{1}{4} \cdot \frac{1}{k^k}.
\end{aligned}$$

Thus, if we set $k = \frac{1}{2} \cdot \frac{\ln n}{\ln \ln n}$, we get that $\Pr(L \geqslant k) \geqslant \frac{1}{4 \cdot \sqrt{n}}$; this means that $T(n)$ is also $\Omega(\log n / \log\log n)$, and hence the correct answer asymptotically is $\Theta(\log n / \log\log n)$.

2 Success Amplification of Randomized Algorithms

Let us conclude by studying another application: *boosting* the probability of success of a randomized algorithm to any desired bound (with repeating it a certain number of times in parallel). To do so, we need the following weaker—but easier to work with—variant of Chernoff bound called the *additive* Chernoff bound.

Proposition 9 (**Additive Chernoff Bound**)**.** *Suppose $X_1, \ldots, X_n$ are* independent *random variables in $[0,1]$ and $X = \sum_i X_i$. Then, for any $b \geqslant 1$,*

$$\Pr(|X - \mathbb{E}[X]| \geqslant b) \leqslant 2 \cdot \exp\left(-\frac{2b^2}{n}\right).$$

We now formalize the problem. Suppose A is a randomized algorithm for some problem P with probability of success at least 2/3. How can we boost its probability of success to $1 - \delta$ for a given $\delta \in (0,1)$?

Majority trick for decision problems: If P is a decision problem and thus A returns a *Yes/No* answer, we can do the following: run A independently for $k := 18 \ln(2/\delta)$ times and return the *majority* answer.

The analysis is as follows. Let $X_1, \ldots, X_k$ be k indicator random variables where $X_i = 1$ iff the i-th run of A returns the correct answer. Let $X := \sum_{i=1}^k X_i$. Firstly, we have that $\mathbb{E}[X] \geqslant k \cdot 2/3$ as each $X_i = 1$ with probability at least 2/3. Secondly, for the majority answer to be wrong, we need $X \leqslant k/2$. Thus,

$$\begin{aligned}
\Pr(\text{majority answer is wrong}) &\leqslant \Pr(|X - \mathbb{E}[X]| \geqslant k/6) && \text{(as } 2k/3 - k/2 = k/6) \\
&\leqslant 2 \cdot \exp\left(-\frac{2 \cdot k^2}{36 \cdot k}\right)
\end{aligned}$$

(by Proposition 9 for $b = k/6$ and $n = k$ and since X is a sum of independent random variables in $[0,1]$)

$$= 2 \cdot \exp(-\ln(2/\delta)) = 2 \cdot (\delta/2) = \delta.$$

[7] Well, this was kind of the whole point of showing how generic concentration inequalities are much easier to work with than an ad-hoc argument...

Median trick for estimation problems: What if P was an estimation problem and thus A returns a number that is in the correct range $[x : y]$ with probability at least $2/3$? In this case, returning the majority will not work because it is possible that none of the numbers returned by the different copies of A are even the same. The solution however is almost the same with a simple tweak: run A independently for $k := 18 \ln(2/\delta)$ times and return the *median* answer[8].

The analysis is also similar to before. Let $X_1, \ldots, X_k$ be k indicator random variables where $X_i = 1$ iff the i-th run of A returns an answer in the correct range $[x : y]$. Let $X := \sum_{i=1}^{k} X_i$. We again have that $\mathbb{E}[X] \geqslant k \cdot 2/3$. Moreover, for the median to be a wrong answer, *at least* half of X_i's has to be zero (this is a necessary condition for failing but is not sufficient). So, again,

$$\Pr(\text{median answer is wrong}) \leqslant \Pr(|X - \mathbb{E}[X]| \geqslant k/6) \leqslant \delta,$$

where the rest of the calculation is exactly as done for the majority trick.

> **Remark.** It is worth emphasizing that both the majority trick and the median trick completely treated the algorithm (and analysis) of their input algorithm in a black-box way. One can use these tricks to boost the probability of success of *any* algorithm (in many different settings) – this is indeed the reason that for most algorithms, the dependence of resources on δ is almost always $O(\ln(1/\delta))$ factor[a].
>
> [a]We emphasize that this is O-notation and *not* Θ; there are a good number of cases also that one can do considerably better than this bound also using more sophisticated (and often ad-hoc) approaches.

3 Appendix: Proof of (a Special Case of) Chernoff Bound

We are *not* going to give the entire proof of the Chernoff bound as it is somewhat tedious. However, to provide enough intuition, we will prove a simpler variant of Chernoff (which is actually sufficient for our balls in bins argument and many other settings as well).

Let us assume that each X_i is a Bernoulli random variable[9] with mean p_i (instead of arbitrary random variable in $[0, 1]$). For simplicity, we are also going to only prove the *upper tail* of the deviation bound instead of both tails (but the lower tail can be proven symmetrically). In particular, we prove the following result.

Proposition 10 (A weaker form of Chernoff bound). *Suppose $X_1, \ldots, X_n$ are* independent *Bernoulli random variables in $\{0, 1\}$ with mean $p_1, \ldots, p_n$ and $X = \sum_i X_i$. Then, for any $\varepsilon > 0$,*

$$\Pr(X \geqslant (1 + \varepsilon) \cdot \mathbb{E}[X]) \leqslant \exp\left(-\frac{\varepsilon^2}{2 + \varepsilon} \cdot \mathbb{E}[X]\right).$$

Notice that in this result, ε is not limited to be less than 1 and thus we can recover the bounds of Proposition 7 by considering $\varepsilon < 1$ and $\varepsilon \geqslant 1$ in the RHS separately.

Proof of a weaker form. Fix $\alpha > 0$. Note that $X \geqslant (1 + \varepsilon)\,\mathbb{E}[X]$ if and only if

$$\exp(\alpha \cdot X) \geqslant \exp(\alpha \cdot (1 + \varepsilon) \cdot \mathbb{E}[X]).$$

Define a random variable $Y = \exp(\alpha \cdot X)$. Using Markov bound on random variable Y, we have,

$$\Pr(X \geqslant (1 + \varepsilon) \cdot \mathbb{E}[X]) = \Pr(Y \geqslant \exp(\alpha \cdot (1 + \varepsilon) \cdot \mathbb{E}[X])) \leqslant \frac{\mathbb{E}[Y]}{\exp(\alpha \cdot (1 + \varepsilon) \cdot \mathbb{E}[X])} \tag{2}$$

[8]You are strongly encouraged to think about why returning the *average* fails (in certain cases miserably).
[9]Recall that a Bernoulli random variable Z with mean p gets value 1 w.p. p and 0 w.p. $1 - p$.

As such, to bound the probability of deviation of X from $\mathbb{E}[X]$, we only need to bound $\mathbb{E}[Y]$ and then we can apply Eq (2). We can now upper bound $\mathbb{E}[Y]$ as follows:

$$\begin{aligned}
\mathbb{E}[Y] = \mathbb{E}[\exp(\alpha \cdot X)] &= \mathbb{E}\left[\exp\left(\alpha \cdot \sum_{i=1}^{n} X_i\right)\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{n} \exp(\alpha \cdot X_i)\right] \\
&= \prod_{i=1}^{n} \mathbb{E}[\exp(\alpha \cdot X_i)] && \text{(for \underline{independent} random variables } A, B\text{: } \mathbb{E}[AB] = \mathbb{E}[A] \cdot \mathbb{E}[B]) \\
&= \prod_{i=1}^{n} (1 - p_i + p_i \cdot e^{\alpha}) && \text{(by the assumption that } X_i \text{ is Bernoulli with mean } p_i) \\
&= \prod_{i=1}^{n} (1 + p_i \cdot (e^{\alpha} - 1)) \\
&\leqslant \prod_{i=1}^{n} \exp(p_i \cdot (e^{\alpha} - 1)) && (1 + x \leqslant e^x \text{ for all } x) \\
&= \exp\left((e^{\alpha} - 1) \cdot \sum_{i=1}^{n} p_i\right) \\
&= \exp((e^{\alpha} - 1) \cdot \mathbb{E}[X]).
\end{aligned}$$

Let us now set $\alpha = \ln(1 + \varepsilon)$ and use Eq (2) to obtain that:

$$\begin{aligned}
\Pr(X \geqslant (1 + \varepsilon) \cdot \mathbb{E}[X]) &\leqslant \exp((e^{\alpha} - 1) \cdot \mathbb{E}[X] - \alpha \cdot (1 + \varepsilon) \cdot \mathbb{E}[X]) \\
&= \exp(\varepsilon \cdot \mathbb{E}[X] - (1 + \varepsilon) \cdot \ln(1 + \varepsilon) \cdot \mathbb{E}[X]) \\
&= \left(\frac{e^{\varepsilon}}{(1 + \varepsilon)^{(1+\varepsilon)}}\right)^{\mathbb{E}[X]}. \qquad (3)
\end{aligned}$$

The bound above is already a ***very strong form*** of Chernoff bound (for Bernoulli random variables)[10] – note also that in this equation, we do *not* need ε to be in $(0, 1)$. We can simplify this bound to get the bounds we want in the statement of the proposition (this will weaker the bound slightly; very rarely this weakening can be problematic and we may need to use the stronger bound above directly).

We are going to use the following inequality (the proof is omitted) to simplify Eq (3): For any $x > 0$,

$$1 + x \geqslant \frac{e^x}{1 + x/2}.$$

By applying this, we have (the proof is again omitted),

$$\left(\frac{e^{\varepsilon}}{(1 + \varepsilon)^{(1+\varepsilon)}}\right) = \exp(\varepsilon - (1 + \varepsilon) \cdot \ln(1 + \varepsilon)) \leqslant -\frac{\varepsilon^2}{2 + \varepsilon}.$$

And thus by Eq (3),

$$\Pr(X \geqslant (1 + \varepsilon) \cdot \mathbb{E}[X]) \leqslant \exp\left(-\frac{\varepsilon^2}{2 + \varepsilon} \cdot \mathbb{E}[X]\right).$$

This concludes the proof. □

[10] And we could have recovered the bound of $T(n) = O(\log n / \log \log n)$ for balls-and-bins expriments by using this stronger Chernoff bound; you are encouraged to try this on your own.