CS 466/666: Algorithm Design and Analysis **University of Waterloo: Fall 2025**

Lecture 4

September 16, 2025

Instructor: Sepehr Assadi

Topics of this Lecture

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

1 Balls and Bins Experiments

Consider the following problem.

Problem 1 (**Maximum Load in Balls-and-Bins**)**.** We are given 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 bound $T(n)$ (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,

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

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

[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!

Fact 1. *For any two events A and B,* $\Pr(A \vee B) \leqslant \Pr(A) + \Pr(B)$.

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

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

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

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

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,

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

Thus, the question we are interested in is precisely a question concerning concentration inequalities. We will thus apply a couple of different concentration inequalities to the random variable L to find out the best approach to bound $T(n)$. Before moving on however, an important remark is in order.

Remark. You may wonder why we consider simple tools like linearity of expectation or union bound as extremely important. The main reason is that these tools allow us to reason about a complex random variable—a one consisting of different components or and other random variables—by "breaking" them down to simpler random variables and analyzing these simpler parts in isolation instead. This task is much simpler than analyzing the probability distribution of the original complex random variable directly (on this front, linearity of expectation is particularly appealing because it is an *equality* and thus this process will actually find the tight answer to our problem).

1.1 Concentration Inequalities for Balls and Bins

Markov bound. Let us start by applying Markov bound from the previous lectures. Recall:

Proposition 2 (**Markov bound**). *For any non-negative random variable X and $b > 0$,*

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

For our purpose, 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 $k = 3n$. But of course this is entirely useless because we only have n balls to begin with and so $L \leqslant n$ happens with probability 1!

Chebyshev's inequality. We can next try Chebyshev's inequality. Recall:

Proposition 3 (**Chebyshev's Inequality**). *For any random variable X and $b > 0$,*

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

To do so, we need to bound $\mathrm{Var}\left[L\right]$. We have,

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

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. Thus,

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

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*. Formally,

Proposition 4 (**Chernoff Bound**). *Suppose $X_1, \ldots, X_n$ are* independent *random variables in $[0, 1]$ and $X = \sum_i X_i$. Then, for any $t \geqslant 1$,*

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

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

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

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[2]). The proof of Proposition 4 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 Appendix A.

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

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

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

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

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

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

1.2 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.

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

So far, in these calculations, we already used one inequality where we applied the union bound. We are going to apply yet another inequality[3] (which in general is a highly useful inequality to know).

Fact 5. *For any integers $a \geqslant b$ (here, $e \approx 2.73...$ is the natural number),*

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

[3] 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.

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,

$$\begin{aligned}
\Pr(L \geqslant k) &\leqslant \binom{n}{k} \cdot \frac{1}{n^k} && \text{(as calculated above)} \\
&\leqslant \left(\frac{e \cdot n}{k}\right)^k \cdot \frac{1}{n^k} && \text{(by Fact 5)} \\
&= \left(\frac{e}{k}\right)^k \\
&= \frac{e^k}{e^{k \cdot \ln k}} && (\ln k \text{ is the log of } k \text{ in base } e) \\
&= \exp\left(-k \cdot (\ln k - 1)\right) \\
&= \exp\left(-k \cdot \ln(k/e)\right).
\end{aligned}$$

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

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

Plugging in all these—honestly, rather tedious[4]—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:

$$\Pr(L \geqslant k) \geqslant \Pr(L = k)$$

[4] 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...

$$
\begin{aligned}
&= \binom{n}{k} \cdot \frac{1}{n^k} \cdot \left(1 - \frac{1}{n}\right)^{n-k} \\
&\geqslant \binom{n}{k} \cdot \frac{1}{n^k} \cdot \frac{1}{4} && \text{(as } (1-1/n)^m \geqslant (1-1/n)^n \geqslant 1/4 \text{ for all } m \leqslant n \text{ and } n \geqslant 2) \\
&\geqslant \left(\frac{n}{k}\right)^k \cdot \frac{1}{n^k} \cdot \frac{1}{4} && \text{(by Fact 5)} \\
&= \frac{1}{4} \cdot \frac{1}{k^k}.
\end{aligned}
$$

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 6 (**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,

$$
\begin{aligned}
\Pr(\text{majority answer is wrong}) &\leqslant \Pr(|X - \mathbb{E}[X]| \geqslant k/6) && \text{(as } 2k/3 - k/2 = k/6) \\
&\leqslant 2 \cdot \exp\left(-\frac{2 \cdot k^2}{36 \cdot k}\right) \\
&\quad \text{(by Proposition 6 for } b = k/6 \text{ and } n = k \text{ and since } X \text{ is a sum of independent random variables in } [0,1]) \\
&= 2 \cdot \exp(-\ln(2/\delta)) = 2 \cdot (\delta/2) = \delta.
\end{aligned}
$$

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 $2/3$?[5] 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[6].

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,

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

[5] E.g., for the streaming distinct element problem where A's answer is a $(1 \pm \varepsilon)$ approximation of the number of distinct elements DE, i.e., is in the range $[(1-\varepsilon) \cdot \mathsf{DE}, (1+\varepsilon) \cdot \mathsf{DE}]$.

[6] You are strongly encouraged to think about why returning the *average* fails (in certain cases miserably).

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.

A 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[7] 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 7 (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 $\varepsilon > 0$,*

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

Notice that in this result, ε is not limited to be less than 1 and thus we can recover the bounds of Proposition 4 by considering $\varepsilon < 1$ and $\varepsilon \geqslant 1$ in the RHS separately.

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

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

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

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

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

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

[7]Recall that a Bernoulli random variable Z with mean p gets value 1 w.p. p and 0 w.p. $1 - p$.

$$
\begin{aligned}
&= \prod_{i=1}^{n} (1 + p_i \cdot (e^{\alpha} - 1)) \\
&\leqslant \prod_{i=1}^{n} \exp\left(p_i \cdot (e^{\alpha} - 1)\right) && (1 + x \leqslant e^x \text{ for all } x) \\
&= \exp\left((e^{\alpha} - 1) \cdot \sum_{i=1}^{n} p_i\right) \\
&= \exp\left((e^{\alpha} - 1) \cdot \mathbb{E}[X]\right).
\end{aligned}
$$

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

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

The bound above is already a ***very strong form*** of Chernoff bound (for Bernoulli random variables)[8] – 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),

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

And thus by Eq (3),

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

This concludes the proof. □

[8] 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.