Experiment

- There are \( m = n \) balls
- Each ball is thrown uniformly and independently at random into \( n \) bins
- **Objective.** Understand the behavior of \( E[L_{\text{max}}] \)
We shall show the following result

**Theorem (Expected Max-Load)**

Let $m = n$ balls be thrown uniformly and independently at random into $n$ bins. Let $L_{\text{max}}$ be the random variable denoting the maximum load of the bins. Then, we have the following result.

$$\mathbb{E}[L_{\text{max}}] = \Theta \left( \frac{\log n}{\log \log n} \right)$$
Our idea is to prove the following. For some positive constant $c$, we have

$$E[L_{\text{max}}] \leq c \left( \frac{\log n}{\log \log n} \right)$$

Our strategy is to use the following trick to calculate the expectation of a random variable $X$ over natural numbers

$$E[X] = \sum_{i \geq 1} i \cdot P[X = i]$$

$$= \sum_{i \geq 1} \sum_{j \geq i} P[X = j]$$

$$= \sum_{i \geq 1} P[X \geq i]$$
So, we have
\[ E[\mathbb{L}_{\text{max}}] = \sum_{i \geq 1} P[\mathbb{L}_{\text{max}} \geq i] \]

We begin with the following result

**Lemma**

For any \( \ell \in \mathbb{N} \), we have the following bound

\[ P[\mathbb{L}_{j} \geq \ell] \leq \left( \frac{n}{\ell} \right) \frac{1}{n^\ell} \leq \frac{1}{\ell!} \]

**Proof Outline.**

The probability that bin \( j \) receives \( \geq \ell \) balls is (at most) the probability of the following event

- We choose a set of \( \ell \) balls from \( n \) balls in \( \binom{n}{\ell} \) ways
- We compute the probability that these \( \ell \) balls land in bin \( j \)
- The other balls can go anywhere (including falling in bin \( j \))
**Food for Thought.** Why is this probability expression an inequality and not an equality?

- Let $\ell^*$ be the smallest integer such that $(\ell^*)! \geq n^2$

**Exercise.** Prove that $\ell^* \leq c \frac{\log n}{\log \log n}$ for some positive constant $c$

- So, we have $\mathbb{P}[L_j \geq \ell^*] \leq 1/n^2$

- Now, by Union Bound, we have

$$
\mathbb{P}[L_1 \geq \ell^* \text{ or } L_2 \geq \ell^* \text{ or } \cdots \text{ or } L_n \geq \ell^*] \leq n \cdot \frac{1}{n^2} = \frac{1}{n}
$$

- That is, we have

$$
\mathbb{P}[L_{\text{max}} \geq \ell^*] \leq \frac{1}{n}
$$
Now, we are at a position to upper-bound the expected max-load

\[
\mathbb{E} [\mathbb{L}_{\text{max}}] = \sum_{i \geq 1} \mathbb{P} [\mathbb{L}_{\text{max}} \geq i]
\]

\[
= \sum_{i=1}^{\ell^* - 1} \mathbb{P} [\mathbb{L}_{\text{max}} \geq i] + \sum_{i=\ell^*}^{n} \mathbb{P} [\mathbb{L}_{\text{max}} \geq i]
\]

\[
= (\ell^* - 1) \cdot 1 + (n - \ell^*) \cdot \frac{1}{n}
\]

\[
< \ell^*
\]
Let us take a small detour. We shall introduce a very powerful technical tool called the “Poisson Approximation Theorem” and then revisit this problem.
Let us start by calculating the probability that the bin $j$ receives exactly $\ell$ balls

- Suppose we are throwing $m$ balls into $n$ bins
- There are $\binom{m}{\ell}$ ways to choose the set of $\ell$ balls that fall into the bin $j$
- Given this fixed set of balls, the probability that these $\ell$ balls fall into bin $j$, and the remaining $(m - \ell)$ balls do not fall into bin $j$ is given by the following expression

\[
\frac{1}{n^\ell} \left(1 - \frac{1}{n}\right)^{m-\ell}
\]
So, we have the following result

\[ P[L_j = \ell] = \binom{m}{\ell} \frac{1}{n^\ell} \left(1 - \frac{1}{n}\right)^{m-\ell} \]
Rough Calculation below.

- Let \( \mu = m/n \), the expected load of a bin
- Let us now perform a rough calculation

\[
P[ \mathbb{L}_j = \ell] = \binom{m}{\ell} \frac{1}{n^\ell} \left(1 - \frac{1}{n}\right)^{m-\ell}
\]

\[
\approx \frac{m^\ell}{\ell!} \cdot \frac{1}{n^\ell} \left(1 - \frac{1}{n}\right)^m \left(1 - \frac{1}{n}\right)^{-\ell}
\]

\[
= \frac{m^\ell}{\ell!} \cdot \frac{1}{(n-1)^\ell} \left(1 - \frac{1}{n}\right)^m
\]

\[
\approx \exp(-\mu) \frac{\mu^\ell}{\ell!}
\]
Poisson Distribution

The random variable $X$ over $\Omega = \{0, 1, \ldots\}$ is a Poisson distribution with mean $\mu$ if the following condition is satisfied for all $i \in \Omega$

$$P[X = i] = \exp(-\mu) \frac{\mu^i}{i!}$$

So, the load $L_j$ is (roughly) distributed like a Poisson distribution with mean $\mu = m/n$
Reality.

- We throw \( m \) balls into \( n \) bins uniformly and independently at random. Let \((L_1, L_2, \ldots, L_n)\) be the joint distribution of the loads of the bins.

Poisson Approximation.

- Let \((X^{(1)}, X^{(2)}, \ldots, X^{(n)})\) be the distribution corresponding to \( n \) independent Poisson distributions with mean \( \mu \).

Objective.

- We can approximate the behavior of the function \( f \) in the reality using its behavior in the Poisson approximation world. That is, we approximate the random variable \( f(L_1, \ldots, L_n) \) using the random variable \( f(X^{(1)}, \ldots, X^{(n)}) \).
We state the following theorem without proof.

**Theorem (Poisson Approximation)**

If \( f \) is “well-behaved” (for some positive function \( c(m) \))

\[
\mathbb{E} \left[ f(L_1, \ldots, L_n) \right] \leq c(m) \mathbb{E} \left[ f(X^{(1)}, \ldots, X^{(n)}) \right]
\]


For example, if \( f \) is a non-negative and monotonically increasing function in \( m \) (the number of balls) then we have \( c(m) = 2 \)

If \( f \) is non-negative function then \( c(m) = e \sqrt{m} \)
Suppose we show that

$$\mathbb{P} [L_{\text{max}} < \ell^{**}] \leq \frac{1}{n}$$

for as large a value of $\ell^{**}$ as possible.

Then, we can do the following calculation:

$$\mathbb{E} [L_{\text{max}}] = \sum_{i \geq 1} i \cdot \mathbb{P} [L_{\text{max}} = i]$$

$$\geq \sum_{i \geq \ell^{**}} i \cdot \mathbb{P} [L_{\text{max}} = i]$$

$$\geq \sum_{i \geq \ell^{**}} \ell^{**} \cdot \mathbb{P} [L_{\text{max}} = i]$$

$$= \ell^{**} \mathbb{P} [L_{\text{max}} \geq \ell^{**}]$$

$$\geq \ell^{**} \left(1 - \frac{1}{n}\right)$$
To show that \( P[L_{\text{max}} < \ell^{**}] \leq 1/n \), let us define a random variable \( 1_{\{L_{\text{max}} < \ell^{**}\}} \).

We can equivalently write this random variable as a function \( f(L_1, \ldots, L_n) \).

Consider \( n \) independent Poisson distribution \( (X^{(1)}, \ldots, X^{(n)}) \) with mean \( \mu = m/n = 1 \).

By Poisson approximation theorem, the expectation of this function in the real world is

\[
e^{\sqrt{n}E[f(X^{(1)}, \ldots, X^{(n)})]}
\]

So, it shall suffice to show that

\[
\left( P[X < \ell^{**}] \right)^n \leq \frac{1}{e n^{3/2}} = \exp \left( -1 - \frac{3}{2} \log n \right)
\]
Which, in turn, is equivalent to showing that

$$\Pr[X < \ell^{**}] \leq \exp\left(-\frac{1 + \frac{3}{2} \log n}{n}\right)$$

To prove the above statement, it suffices to prove the following statement

$$\Pr[X < \ell^{**}] \leq 1 - \left(\frac{1 + \frac{3}{2} \log n}{n}\right),$$

because $1 - x \leq \exp(-x)$

To find $\ell^{**}$ such that this bound holds, note the following

$$\Pr[X < \ell^{**}] = 1 - \Pr[X \geq \ell^{**}] \leq 1 - \Pr[X = \ell^{**}] = 1 - \frac{\exp(-1)}{(\ell^{**})!}$$

Now, we solve for $\ell^{**} = \frac{n}{1 + \frac{3}{2} \log n}$, which gives $\ell^{**} \geq d \frac{\log n}{\log \log n}$, for some positive constant $d$.
**Problem Statement.** What is the number $m$ of balls that one should throw such that each bin receives at least one ball?

This problem is referred to as the Coupon Collector’s Problem. Basically, how many cereal boxes to buy so that you get all the toys?

Think: How to solve this problem using the Poisson approximation theorem. The answer is $m \approx n \log n$

How many balls should one throw to ensure that there are at least $r$ balls in each bin?