Lecture 05: Balls and Bins: Max Load

Overview

- In today's lecture we shall study the behavior of the maximum load when $m = n$ balls are thrown into n bins

Bounding the expected value of $\mathbb{L}_{max}$

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 $\mathbb{L}_{max}$ be the random variable denoting the maximum load of the bins. Then, we have the following result.

$$\mathbb{E}\left[\mathbb{L}_{max}\right] = \Theta\left(\frac{\log n}{\log \log n}\right)$$

Upper Bound I

- Our idea is to prove the following. For some positive constant c, we have
$$\mathbb{E}\left[\mathbb{L}_{\max}\right] \leqslant c\left(\frac{\log n}{\log\log n}\right)$$
- Our strategy is to use the following trick to calculate the expectation of a random variable $\mathbb{X}$ over natural numbers

$$\begin{aligned}\mathbb{E}\left[\mathbb{X}\right] &= \sum_{i\geqslant 1} i\cdot\mathbb{P}\left[\mathbb{X}=i\right]\\ &= \sum_{i\geqslant 1}\sum_{j\geqslant i}\mathbb{P}\left[\mathbb{X}=j\right]\\ &= \sum_{i\geqslant 1}\mathbb{P}\left[\mathbb{X}\geqslant i\right]\end{aligned}$$

Upper Bound II

- So, we have

$$\mathbb{E}\left[\mathbb{L}_{\mathsf{max}}\right] = \sum_{i \geqslant 1} \mathbb{P}\left[\mathbb{L}_{\mathsf{max}} \geqslant i\right]$$

Upper Bound III

Lemma

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

$$\mathbb{P}\left[\mathbb{L}_j \geqslant \ell\right] \leqslant \binom{n}{\ell}\frac{1}{n^\ell} \leqslant \frac{1}{\ell!}$$

Proof.

- The probability that bin j receives $\geqslant \ell$ balls is (at most) the probability of the following event
  - We choose a set of ℓ balls from n balls in $\binom{n}{\ell}$ ways
  - We compute the probability that these ℓ balls land in bin j
  - The other balls can go anywhere (including falling in bin j) □

- Think: Why is this an inequality and not an equality?

Upper Bound IV

- Let ℓ^* be an integer such that $(\ell^*)! \geqslant n^2$
- Exercise: Prove that $\ell^* \leqslant c\frac{\log n}{\log\log n}$ for some positive constant c
- So, we have $\mathbb{P}\left[\mathbb{L}_j \geqslant \ell^*\right] \leqslant \frac{1}{n^2}$
- Now, by union bound, we have

$$\mathbb{P}\left[\mathbb{L}_1 \geqslant \ell^* \text{ or } \mathbb{L}_2 \geqslant \ell^* \text{ or } \cdots \text{ or } \mathbb{L}_n \geqslant \ell^*\right] \leqslant n \cdot \frac{1}{n^2} = \frac{1}{n}$$

- That is, we have

$$\mathbb{P}\left[\mathbb{L}_{\max} \geqslant \ell^*\right] \leqslant \frac{1}{n}$$

Upper Bound V

- Now, we are at a position to upper bound the expected max-load

$$
\begin{aligned}
\mathbb{E}\left[\mathbb{L}_{\max}\right] &= \sum_{i \geqslant 1} \mathbb{P}\left[\mathbb{L}_{\max} \geqslant i\right] \\
&= \sum_{i=1}^{\ell^*-1} \mathbb{P}\left[\mathbb{L}_{\max} \geqslant i\right] + \sum_{i=\ell^*}^{n} \mathbb{P}\left[\mathbb{L}_{\max} \geqslant i\right] \\
&\leqslant (\ell^* - 1) \cdot 1 + (n - \ell^*) \cdot \frac{1}{n} \\
&< \ell^*
\end{aligned}
$$

Lower Bound

- Let us take a small detour. We shall introduce a very strong technical tool called "Poisson Approximation Theorem" and then revisit this problem

Poisson Distribution I

Let us start by calculating the property that bin j receives exactly ℓ balls

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

$$\frac{1}{n^\ell}\left(1-\frac{1}{n}\right)^{m-\ell}$$

Poisson Distribution II

- So, we have the following

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

Rough Calculation below.

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

$$\begin{aligned}
\mathbb{P}\left[\mathbb{L}_j = \ell\right] &= \binom{m}{\ell} \frac{1}{n^\ell} \left(1 - \frac{1}{n}\right)^{m-\ell} \\
&\approx \frac{m^\ell}{\ell!} \cdot \frac{1}{n^\ell} \cdot \left(1 - \frac{1}{n}\right)^m \left(1 - \frac{1}{n}\right)^{-\ell} \\
&= \frac{m^\ell}{\ell!} \cdot \frac{1}{(n-1)^\ell} \cdot \left(1 - \frac{1}{n}\right)^m \\
&\approx \exp(-\mu) \frac{\mu^\ell}{\ell!}
\end{aligned}$$

Poisson Distribution IV

Poisson Distribution.

- The random variable $\mathbb{X}$ over $\Omega = \{0, 1, \dots, \}$ is a Poisson distribution with mean μ if the following condition is satisfied for all $i \in \Omega$
$$\mathbb{P}[\mathbb{X} = i] = \exp(-\mu)\frac{\mu^\ell}{\ell!}$$
- So, the load $\mathbb{L}_j$ is (roughly) distributed like a Poisson distribution with mean $\mu = m/n$

Poisson Approximation I

Reality.

- We throw m balls into n bins uniformly and independently at random. Let $(\mathbb{L}_1, \mathbb{L}_2, \dots, \mathbb{L}_n)$ be the joint distribution of the loads of the bins

Poisson Approximation.

- Let $(\mathbb{X}^{(1)}, \mathbb{X}^{(2)}, \dots, \mathbb{X}^{(n)})$ be the distribution corresponding to n independent Poisson distributions with mean μ

Goal.

- 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 $f(\mathbb{L}_1, \dots, \mathbb{L}_n)$ using the random variable $f(\mathbb{X}^{(1)}, \dots, \mathbb{X}^{(n)})$.

Poisson Approximation II

We state the following theorem without proof.

Theorem (Poisson Approximation)

If f is "well-behaved" then (for some function $c(m)$)

$$\mathbb{E}\left[f(\mathbb{L}_1, \ldots, \mathbb{L}_n)\right] \leqslant c(m) \cdot \mathbb{E}\left[f(\mathbb{X}^{(1)}, \ldots, \mathbb{X}^{(n)})\right]$$

Refer to the book "Probability and Computing: Randomized Algorithms and Probabilistic Analysis," by Michael Mitzenmacher and? Eli Upfal for a full proof.

For example, if f is 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 $c(m) = \mathrm{e}\sqrt{m}$.

Revisiting "Lower Bounding Max Load" I

- Suppose we show that

$$\mathbb{P}\left[\mathbb{L}_{\mathsf{max}} < \ell^{**}\right] \leqslant \frac{1}{n}$$

- Then, we can do the following calculation

$$\begin{aligned}
\mathbb{E}\left[\mathbb{L}_{\mathsf{max}}\right] &= \sum_{i \geqslant 0} i\mathbb{P}\left[\mathbb{L}_{\mathsf{max}} = i\right] \\
&\geqslant \sum_{i \geqslant \ell^{**}} i\mathbb{P}\left[\mathbb{L}_{\mathsf{max}} = i\right] \\
&\geqslant \sum_{i \geqslant \ell^{**}} \ell^{**}\mathbb{P}\left[\mathbb{L}_{\mathsf{max}} = i\right] \\
&= \ell^{**}\mathbb{P}\left[\mathbb{L}_{\mathsf{max}} \geqslant \ell^{**}\right] \\
&\geqslant \ell^{**}\left(1 - \frac{1}{n}\right)
\end{aligned}$$

Revisiting "Lower Bounding Max Load" II

- To show that $\mathbb{P}[\mathbb{L}_{\max} < \ell^{**}] \leqslant \frac{1}{n}$, let us define a random variable $\mathbf{1}_{\{\mathbb{L}_{\max} < \ell^{**}\}}$
- We can equivalently write this random variable as a function $f(\mathbb{L}_1, \ldots, \mathbb{L}_n)$
- Consider n independent Poisson distributions $(\mathbb{X}^{(1)}, \ldots, \mathbb{X}^{(n)})$ with mean $\mu = m/n = 1$
- By Poisson Approximation theorem, the expectation of this function in the real world is

$$\leqslant \mathrm{e}\sqrt{n}\mathbb{E}\left[f(\mathbb{X}^{(1)}, \ldots, \mathbb{X}^{(n)})\right]$$

- So, it shall suffice to show that

$$\left(\mathbb{P}\left[\mathbb{X} < \ell^{**}\right]\right)^n \leqslant \frac{1}{\mathrm{e}n^{3/2}} = \exp\left(-1 - \frac{3}{2}\log n\right)$$

Revisiting "Lower Bounding Max Load" III

- Which is, in turn, equivalent to showing that

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

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

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

  because $1 - x \leqslant \exp(-x)$.
- To find ℓ^{**} such that this bound holds, note the following.
  - $\mathbb{P}\left[\mathbb{X} < \ell^{**}\right] = 1 - \mathbb{P}\left[\mathbb{X} \geqslant \ell^{**}\right] \leqslant 1 - \mathbb{P}\left[\mathbb{X} = \ell^{**}\right] = 1 - \frac{\exp(-1)}{(\ell^{**})!}$
  - Now we solve for $(\ell^{**})! = \frac{n}{1+\frac{3}{2}\log n}$, which gives $\ell^{**} \geqslant d\frac{\log n}{\log\log n}$, 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?