Lecture 07: Expected Max-Load & Poisson Approximation Theorem

Recall: Upper Bound

Let us quickly recall the strategy that was employed to upper-bound the expected max-load in the previous lecture.

- If there exists a value ℓ^* such that $\mathbb{P}\left[\mathbb{L}_{\max} \geqslant \ell^*\right] \leqslant 1/n$, then we showed that $\mathbb{E}\left[\mathbb{L}_{\max}\right] \leqslant \ell^*$.
- So, our task reduces to finding a meaningful value of ℓ^*
- Next, using the union bound, we reduced the task to finding ℓ^* such that $\mathbb{P}\left[\mathbb{L}_j\right] \leqslant 1/n^2$
- Finally, we demonstrated that there exists $\ell^* = \Theta(\log n / \log\log n)$ that suffices for our purpose

Overview: Today's Plan

Today's plan is as follows

- If there exists ℓ^{**} such that $\mathbb{P}\left[\mathbb{L}_{\mathsf{max}} < \ell^{**}\right] \leqslant 1/n$ then we shall show that $\mathbb{E}\left[\mathbb{L}_{\mathsf{max}}\right] \geqslant \ell^{**}/2$
- Now, we need to find a *meaningful value of* ℓ^{**}
- *This objective shall be achieved via a powerful technical tool, namely, the Poisson Approximation Theorem. We shall not prove this theorem. However, we shall learn how to use this result for our objective. In the homework, we shall see several complex applications of this result*

Lower Bound

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

Poisson Distribution I

Let us start by calculating the probability that the 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

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

- So, we have the following result

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

Rough Calculation below.

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

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

Poisson Distribution.

- The radom 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 $\mu = m/n$

Reality.

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

Poisson Approximation.

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

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

Poisson Approximation Theorem II

We state the following theorem without proof.

Theorem (Poisson Approximation)

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

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

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

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 $c(m) = \mathrm{e}\sqrt{m}$

Revisiting "Lower Bounding Max-Load" I

- Suppose we show that

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

  for as large a value of ℓ^{**} as possible

- Then, we can do the following calculation

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

Revisiting "Lower Bounding Max-Load" II

- To show that $\mathbb{P}\left[\mathbb{L}_{\max} < \ell^{**}\right] \leqslant 1/n$, 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 distribution $(\mathbb{X}^{(1)}, \ldots, \mathbb{X}^{(n)})$ with mean $\mu = m/n = 1$
- By Poisson approximation theorem, the expectation of this function in the real world is

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

- 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, in turn, is 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

Coupon Collector Problem

- **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?