Lecture 05: Balls and Bins Experiments

Description of the Experiment

- We have m balls and n bins. There is no a priori relation between the numbers m and n. For example, based on the context, it is possible that $m < n$, or $m > n$, or $m \ll n$, or $m \gg n$.
- The balls are sequentially thrown into the bins. Again, note that we have not mentioned the "strategy" of throwing the balls. Based on the context, the balls can be thrown according to different strategies. For example, balls can be thrown into bins so that they preferentially land in bins with lower number, or the i-th ball avoids bins that already have a lot of balls in them, etc.
- The "load of bin j" refers to the number of balls in the bin j
- The "max-load" of the bins refers to the maximum load of the bins

Mathematically Formulating the Problem

- Our sample space is $[n]^{\otimes m}$
- Our random variables are $(\mathbb{X}_1, \mathbb{X}_2, \ldots, \mathbb{X}_m)$, where $\mathbb{X}_i$ represents the bin into which the i-th ball falls.
- Now, the load of bin $j \in [n]$ is the number of balls that fall into it. The random variable is represented as follows

$$\mathbb{L}_j := \sum_{i=1}^{m} \mathbf{1}_{\{\mathbb{X}_i = j\}}$$

- The max-load of the bins is expressed as the following random variable

$$\mathbb{L}_{\max} = \max\{\mathbb{L}_1, \mathbb{L}_2, \ldots, \mathbb{L}_n\}$$

Throwing the Balls "Independently"

- For any ball i and any outcomes $x_1, \ldots, x_{i-1}$ of the previous ball throws, we have

$$\mathbb{X}_i \equiv \left(\mathbb{X}_i | \mathbb{X}_1 = x_1, \ldots, \mathbb{X}_{i-1} = x_{i-1}\right)$$

- Basically, we are stating that the strategy to throw the i-th balls does not depend on where the previous $(i-1)$ balls were thrown

Throwing the Balls "Uniformly at random"

- The uniform distribution over a sample space Ω is represented by $\mathbb{U}_\Omega$. This distribution has the property that $\mathbb{P}\left[\mathbb{U}_\Omega = \omega\right] = 1/|\Omega|$, for any element $\omega \in \Omega$
- The random variable $\mathbb{X}_i$ is identical to the distribution $\mathbb{U}_{[n]}$ implies that the balls are thrown uniformly at random

Uniformly and independently at Random

- The default strategy of throwing the balls into the bins will be uniformly and independently at random
- This strategy chooses a uniformly random bin for every ball irrespective of where the previous balls were thrown

Expected Load I

Let us prove an interesting result about the load of any bin

Theorem

For any $j \in [n]$, the expected load of the j-th bin is m/n.

Expected Load II

Here is the proof outline.

$$
\begin{aligned}
\mathbb{E}\left[\mathbb{L}_j\right] &= \mathbb{E}\left[\sum_{i=1}^{m} \mathbf{1}_{\{\mathbb{X}_i = j\}}\right], \text{ By definition of the r.v.}\\
&= \sum_{i=1}^{m} \mathbb{E}\left[\mathbf{1}_{\{\mathbb{X}_i = j\}}\right], \text{ By linearity of expectation}\\
&= \sum_{i=1}^{m} \mathbb{P}\left[\mathbb{X}_i = j\right], \text{ By properties of indicator variables}\\
&= \sum_{i=1}^{m} \frac{1}{n}, \text{ Because } \mathbb{X}_i \text{ is uniform over } [n]\\
&= \frac{m}{n}
\end{aligned}
$$

A Note.

- Observe that the proof does not rely on the fact that the random variables $\mathbb{X}_i$s are independent!
- So, even if the balls are thrown in a "correlated fashion," as long as the distribution $\mathbb{X}_i$ is uniform, for all $i \in [m]$, this proof will hold
- For example, consider the following new way of throwing the balls. "Choose a bin uniformly at random and throw all the balls into that bins."
  Note that in this strategy of throwing balls, we still have $\mathbb{P}[\mathbb{X}_i = j] = 1/n$, for all $i \in [m]$ and $j \in [n]$. So, the expected load of any bin j is m/n as well!

"Birthday Bound" in New Terminology

- The birthday bound problem that we considered in the previous lecture can be formulated as follows. We are interested in computing the following probability

$$P_{m,n} := \mathbb{P}\left[\mathbb{L}_{\max} = 1\right]$$

  as a function of m and n. Because "$\mathbb{L}_{\max} = 1$" is equivalent to the event that "all the balls have fallen into different bins"
- In the previous lecture we concluded that $P_{m,n} = 0.99$ when $m = \alpha\sqrt{n}$ (where α is an appropriate constant) and $P_{m,n} = 0.01$ when $m = \beta\sqrt{n}$ (where β is an appropriate constant). So, is a short range of $(\beta - \alpha)\sqrt{n}$ the probability $P_{m,n}$ transitions from 0.99 to 0.01

Looking Ahead

- The birthday bound only studies the probability that $\mathbb{L}_{\text{max}} = 1$. How does the distribution $\mathbb{L}_{\text{max}}$ behave?
- we will study its expected value and the fact that most of the mass of the probability id close to expectation in the future