Lecture 04: Balls and Bins: Birthday Paradox & Maximum Load

Elementary Inequalities

- $\left(1-\frac{1}{k}\right)^k \leqslant e^{-1}$, for $k \geqslant 1$
- $\left(1+\frac{1}{k}\right)^k \leqslant e$, for $k \geqslant 0$
- $\left(\frac{n}{k}\right)^k \leqslant \binom{n}{k} \leqslant \left(\frac{en}{k}\right)^k$, for $0 \leqslant k \leqslant n$
- $1-x \leqslant e^{-x}$, for all $x \geqslant 0$
- There exists a constant $c \in (0,1)$ such that $e^{-x-x^2} \leqslant 1-x$, for $x \in [0,c]$

Markov Inequality

Theorem (Markov Inequality)

Let X be a random variable that takes non-negative values. Then $\Pr[X \geqslant t] \leqslant \mathbb{E}[X]/t$.

- Suppose not, then $\Pr[X \geqslant t] > \mathbb{E}[X]/t$
- $\mathbb{E}[X] \geqslant 0 \cdot \Pr[0 \leqslant X < t] + t \cdot \Pr[X \geqslant t] > \mathbb{E}[X]$
- Hence contradiction

- Think: Tightness

Chebyshev's Inequality

Theorem (Chebyshev's Inequality)

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

- Use Markov on $\Pr[(X - \mathbb{E}[x])^2 \geqslant t^2]$
- Think: Tightness

Birthday Paradox

$p_{m,n}$ is the probability of encountering a collision when m balls are thrown in n bins

$$\begin{aligned}
1 - p_{m,n} &= \frac{n}{n} \cdot \frac{(n-1)}{n} \cdots \frac{(n-m+1)}{n} \\
&= \prod_{i=0}^{m-1} \left(1 - \frac{i}{n}\right) \\
&\leqslant \prod_{i=0}^{m-1} \exp\left(-\frac{i}{n}\right) = \exp\left(-\frac{m(m-1)}{2n}\right)
\end{aligned}$$

- Use $m \sim \sqrt{2n \ln(1/p)}$, to achieve $p_{m,n} \geqslant (1-p)$
- Think: Tightness

Maximum Load

Theorem (Maximum Load Bound)

When n balls are thrown into n bins, the maximum load is $\Theta\left(\frac{\log n}{\log\log n}\right)$ with high probability.

Upper Bound

- Let X_i be the indicator variable for bin i getting $\geqslant k$ balls
- $\Pr[X_i = 1] \leqslant \binom{n}{k} \left(\frac{1}{n}\right)^k \leqslant \frac{e^k}{k^k}$
- There exists a suitable constant c such that for $k = k^* := c \log n / \log\log n$, we have $\Pr[X_i = 1] \leqslant 1/n^2$
- Let $X := \sum_{i=1}^{n} X_i$
- $\Pr[X \geqslant 1] \leqslant 1/n$, by union bound

Abstraction: First Moment Method

- $\mathbb{E}[X] = o(1) \implies \Pr[X = 0] = 1 - o(1)$

Lower Bound

- $\Pr[X_i = 1] \geqslant \binom{n}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{n-k} \geqslant \frac{e^k}{k^k}\left(1-\frac{1}{n}\right)^n\left(1-\frac{1}{n}\right)^{-k} \geqslant \frac{e^k}{4k^k}$
- There exists a constant d such that for $k = k^{**} = c\log n/\log\log n$, we have $\Pr[X_i = 1] \geqslant n^{-1/3}$
- $\mathbb{E}[X] \geqslant n^{2/3}$, by linearity of expectation
- $\Pr[X = 0] \leqslant \Pr[|X - \mathbb{E}[X]| \geqslant \mathbb{E}[x]] \leqslant \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2} = \frac{\sum_{i=1}^n \mathrm{Var}(X_i) + 2\sum_{1\leqslant i<j\leqslant n}\mathrm{Cov}(X_i,X_j)}{\mathbb{E}[X]^2} \leqslant \frac{n+0}{n^{4/3}} \leqslant n^{-1/3}$
- We used the fact that $\mathrm{Var}(X_i) \leqslant 1$ for indicator variables
- We used the fact that $\mathrm{Cov}[X_i, X_j] \leqslant 0$ (Prove this)

Abstraction: Second Moment Method

- $\Pr[X = 0] = o(1)$, if $\mathbb{E}[X] \to \infty$ and $\mathbb{E}[X_iX_j] = (1 + o(1))\,\mathbb{E}[X_i]^2$