Load Balancing

N jobs, M machines
Assume N = M

Balls and bins problem

N balls
N bins
Randomly put balls into bins
Expected # balls in each bin = 1
Thm: The max-loaded bin has $O\left(\frac{\log N}{\log \log N}\right)$ balls with probability at least $1 - \frac{1}{N}$.

Proof sketch:
1. Prob. any bin receives more than $O\left(\frac{\log N}{\log \log N}\right)$ balls.
2. Prob. of there being at least one bin with more than $\frac{1}{N^2}$ balls.
   $\Rightarrow$ want to be at most $\frac{1}{N}$

Union Bound
$P(A \cup B) \leq P(A) + P(B)$
Assume fully random hash functions

\[ P(\text{bin } i \text{ has at least } k \text{ balls}) \]

\[ \leq \binom{N}{k} \left( \frac{1}{N} \right)^k \]

\[ = \frac{N!}{(N-k)! k!} \cdot \frac{1}{N^k} \]

\[ \leq \frac{N^k}{k!} \cdot \frac{1}{N^k} \leq \frac{1}{k!} \]

\[ \text{want this to be at most } \frac{1}{N^2} \]

Using Stirling's approximation:

\[ k! = \sqrt{2\pi k} \left( \frac{k}{e} \right)^k \]

Choose \( k = O \left( \frac{\log N}{\log \log N} \right) \) give the desired result.
Markov's inequality: \( X \) is a non-negative r.v. with mean \( \mu \)

\[
P(X \geq x) \leq \frac{\mu}{x}
\]

Uses only expectation

Chebyshev's inequality: \( X \) be a r.v. with \( \mu \) and variance \( \sigma^2 \)

\[
P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}
\]

Uses variance

Using Chebyshev in load balancing

E.g. 2-wise indep. hash family \( H \), \( N \) balls, \( N \) bins

Lemma: Max. load over all bins is \( O(\sqrt{N}) \) w.p. at least \( \frac{1}{2} \)
Consider bin $i$

$$L_i = \text{load on bin } i$$

$$L_i = \sum_{j=1}^{N} x_{ij} \quad x_{ij} \in \{0, 1\}$$

where the $i^{th}$ ball falls into bin $i$.

$$E[L_i] = \frac{N \cdot 1}{N} = 1$$

$$\sigma_i^2 = \text{Var} [L_i] = \text{Var} \left( \sum_{j=1}^{N} x_{ij} \right)$$

$$= \sum_{j=1}^{N} \text{Var} (x_{ij})$$

$$= \frac{N}{N} \cdot \frac{1}{N} = \frac{1}{N}$$

$$\sigma_i^2 = 1 - \frac{1}{N}$$

Using Chernoff, $P \left( |L_i - 1| > \sqrt{2N} \sigma_i \right) \leq \frac{1}{2N}$ and then use union bound.
p-wise independent hash family?

Higher-moment Chebyshev

\[ P \left( |X - \mu| > \varepsilon \right) \leq \frac{E \left[ (X - E(X))^p \right]}{\varepsilon^p} \]

Hoeffding bound:

Let \( X_i \)'s be independent r.v.'s taking values in \([0, 1]\)

\[ X = X_1 + X_2 + \ldots + X_n \quad \text{and} \quad \mu = E[X] \]

Then

\[ P \left( X > \mu + \lambda \right) \leq e^{-\frac{\lambda^2}{2\lambda + \lambda'}} \]

Exponential decay! Much stronger
With high prob. the max load is \( \Omega \left( \frac{\log n}{\log \log n} \right) \)

**Power-2 choice:**

1. Pick 2 bins & place the ball in the bin with smaller number of balls.

2. Maximum num. of balls drops to \( O \left( \log \log n \right) \)
Proof sketch:

For a ball $b$, let $\text{height}(b) =$ num. of balls in its bin after placing $b$ want to show no ball has height more than $0$ (log log)

Prob: of a ball getting height $3$ is at most?

Q: Fraction of bins that can have $\geq 2$ balls?

- at most $\frac{1}{2}$

at most $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \quad (2 \text{ choices})$

$\Rightarrow \text{Expected num. of bins with 3 balls at most} = \frac{N}{4}$
\[ P(\text{ball getting height 4}) \leq \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16} = \frac{1}{2^{4-2}} \]

\[ P(\text{ball getting height } h) \leq \frac{1}{2^{h-2}} \]

Choosing \( h = O(\log \log n) + 2 \) gives \( P \approx \frac{1}{n} \)
pick d-bins at place in the bin with smallest # of balls.

Thm: For any \( d \geq 2 \) d-choice gives a max. load of

\[
\frac{\log \log N}{\log d} + O(1)
\]