Load Balancing

Feb 10 2022

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

<u>Thm</u>: The max-loaded bin has $\underline{O\left(\frac{\log N}{\log \log N}\right)}$ balls

with probability at least $1 - \frac{1}{N}$

<u>proof sketch</u>:

1. Prob. any bin receives more than $O\left(\frac{\log N}{\log \log N}\right)$ balls.

   ↳ Want to be at most $\frac{1}{N^2}$

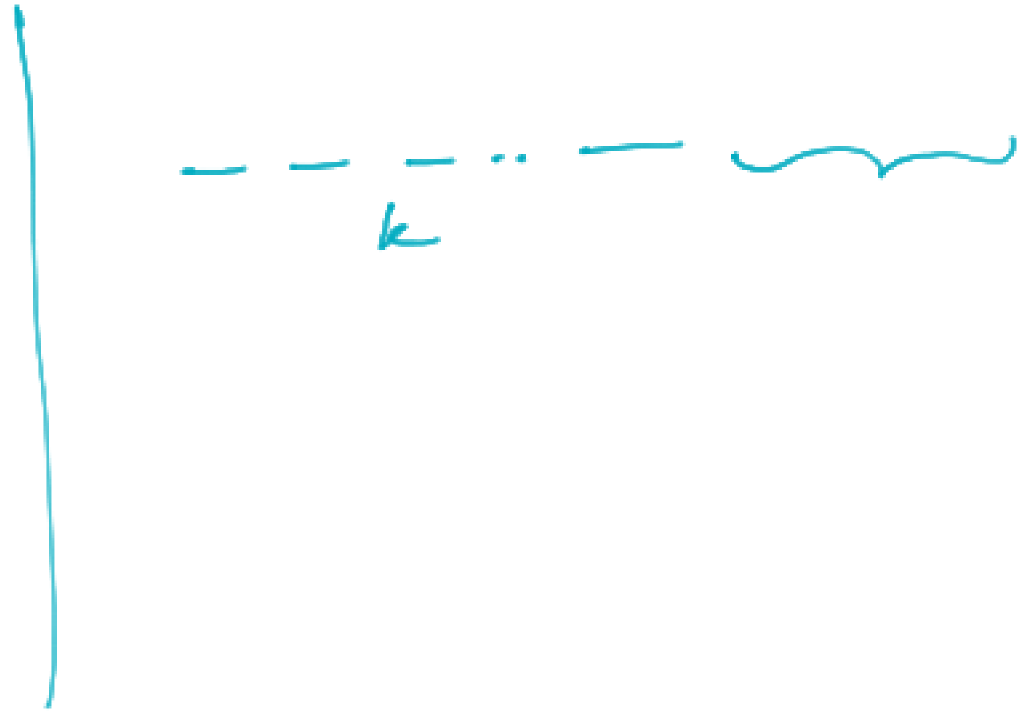
2. Prob. of there being at least one bin with more than

   ↳ want to be at most $\frac{1}{N}$

<u>Union bound</u>

$P(A \cup B) \leq P(A) + P(B)$

Assume fully random hash functions



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

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

$$= \frac{N!}{(N-k)!\,k!} \cdot \frac{1}{N^k}$$

$$\leq \frac{N^k}{k!} \cdot \frac{1}{N^k} \leq \frac{1}{k!}$$

want this to be at most $\frac{1}{N^2}$

Using Sterlings approx: $k! \simeq \sqrt{2\pi k}\left(\frac{k}{e}\right)^k$

Choose $k = O\left(\frac{\log N}{\log\log N}\right)$ gives the desired result

Markov's inequality: X is a non-neg R.V with mean μ

$$P(X \geq \alpha) \leq \frac{\mu}{\alpha}$$

uses only expectation

Chebyshev's inequality: X be a R.V. with μ and variance σ^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 $1/2$

proof: Consider bin i

L_i = load on bin i

$$L_i = \sum_{j=1}^{N} X_{ij}$$

$X_{ij} \in \{0, 1\}$
$= 1$ if j^{th} ball falls into bin i

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

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

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

$\because$ of pairwise indep

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

using chebyshev, $P\left(|L_i - 1| > \sqrt{2N}\,\sigma_i\right) \leq \frac{1}{2N}$

and then use union bound.

p-wise indep. hash family?

Higher-moment Chebyshev

X be a R.V. with mean μ

$$P(|X-\mu| \geq \varepsilon) \leq \frac{E[(X-E(X))^p]}{\varepsilon^p}$$

Hoeffding bound:

Let X_i's be indep. R.V.s taking values in $[0,1]$

$X = X_1 + X_2 + \dots + X_n$ & $\mu = E[X]$ Then

$$P(X > \mu + \lambda) \leq e^{-\frac{\lambda^2}{2\mu+\lambda}}$$

Exponential decay! $\therefore$ much stronger

Thm: With high prob. the max load is $\Omega\left(\frac{\log n}{\log \log n}\right)$

Power-of-2-choice:

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

$\Rightarrow$ Maximum num. of balls drops to $O(\log \log n)$

Proof sketch :

For a ball b, let

height (b) = num. of balls in its bin after placing b

want to show no ball has height more than $O(\log \log n)$

Prob. of a ball getting height 3 is at most ?

Q. Fraction of bins that can have ≥ 2 balls ?

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

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

$\Rightarrow$ Expected num. of bins with 3 balls at most $= \frac{N}{4}$

$$P(\text{ball getting height } 4) \leq \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16} = \frac{1}{2^{2^{4-2}}}$$

$\vdots$

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

Choosing $h = O(\log \log N) + 2$ gives prob. $\frac{1}{N}$

Pick d- bins & place in the bin with smallest # of balls.

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

$$\frac{\log \log N}{\log d} \pm O(1)$$