2 Feb 2025 Load Balance

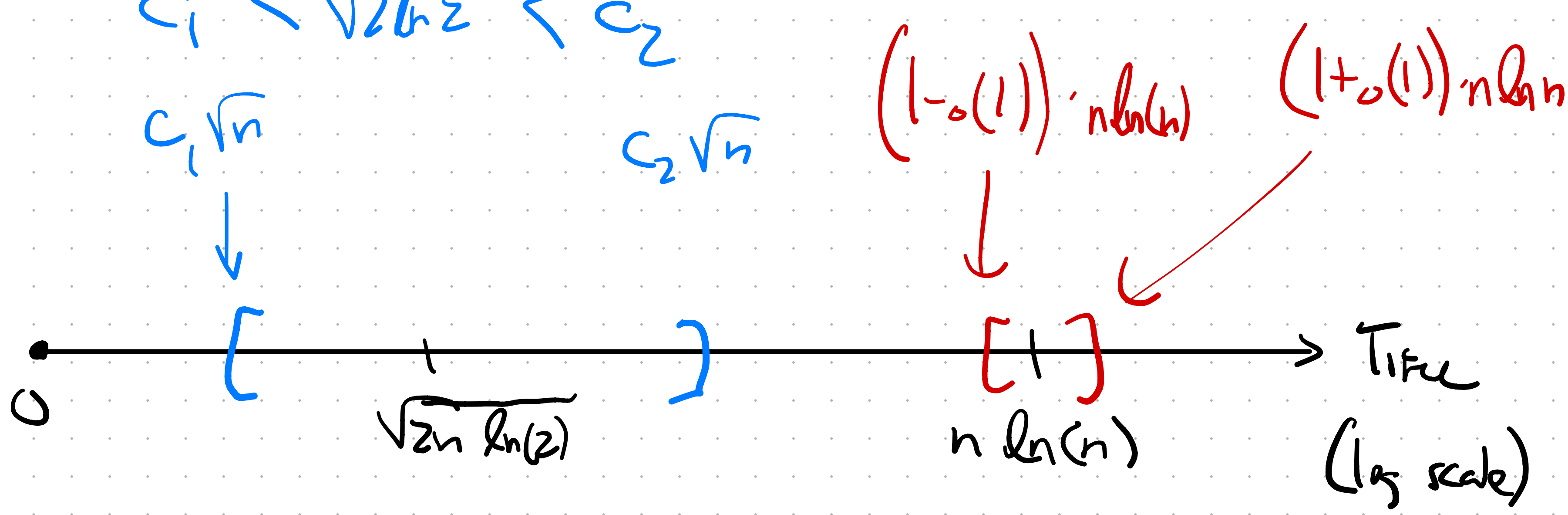
Sequentially throwing balls into n bins.

Random times

T_{bday} = first time when 2 balls occupy same bin.

T_{coupon} = first time when no bin is empty.

$c_1 < \sqrt{2\ln 2} < c_2$



Throw m balls into n bins.

Load vector

$$L = \begin{pmatrix} L_1 \\ L_2 \\ \vdots \\ L_n \end{pmatrix} \qquad L_i = \#\text{ balls in bin } i$$

How large must m be s.t.

with probability $> \frac{1}{2}$,

$$\frac{\max_i \{L_i\}}{\min_i \{L_i\}} < 1 + \varepsilon$$

... for specified $\varepsilon > 0$, e.g. $\varepsilon = 0.1$.

Definitely $m > n \ln(n)$, otherwise denominator likely to be zero.

Correct answer: $m = \Theta\left(\frac{n \log n}{\varepsilon^2}\right)$.

i.e. this happens when

$$\frac{c_1 n \log n}{\varepsilon^2} < m < \frac{c_2 n \log n}{\varepsilon^2}$$

for some constants $0 < c_1 < c_2 < \infty$.

Good news for LB:

$$\forall i \quad E[L_i] = \frac{m}{n}.$$

Proof Let $X_{ij} = \begin{cases} 1 & \text{if ball } j \text{ lands in bin } i \\ \emptyset & \text{if not.} \end{cases}$

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

$$\mathbb{E}(L_i) = \sum_{j=1}^{m} \mathbb{E}[X_{ij}] = \sum_{j=1}^{m} \frac{1}{n} = \frac{m}{n}.$$

Set $\delta = \frac{\varepsilon}{3}$. Suppose

$$\boxed{\forall i \quad (1-\delta)\frac{m}{n} \leq L_i \leq (1+\delta)\frac{m}{n}.}$$

$\Downarrow$

$\frac{\max}{\min} \leq 1+\varepsilon.$

Then

$$\frac{\max_i \{L_i\}}{\min_i \{L_i\}} \leq \frac{(1+\delta)\frac{m}{n}}{(1-\delta)\frac{m}{n}}$$

$$= \frac{1+\delta}{1-\delta} = 1 + \frac{2\delta}{1-\delta}$$

$$= 1 + \frac{2\varepsilon/3}{1-\varepsilon/3} \leq 1+\varepsilon$$

as long as $0 < \varepsilon \leq 1$.

The Tail Bound (TB) plus Union Bound (UB) method.

Plan of attack.

Define

$\mathcal{E}_i$ = event that

$L_i \notin \left[(1-\delta)\frac{m}{n}, (1+\delta)\frac{m}{n}\right]$

View $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n$ as "bad events" to be avoided by taking m large enough that

$\mathcal{E}_1 \cup \ldots \cup \mathcal{E}_n$ is improbable.

$$\Pr(\text{blue-box event}) \leq \Pr\left(\frac{\max}{\min} \leq 1+\varepsilon\right)$$

$$\| $$

$$1 - \Pr(\mathcal{E}_1 \cup \dots \cup \mathcal{E}_n) \leftarrow$$

work on making this diff $\geq \frac{1}{2}$.

Equivlt.: $\Pr(\mathcal{E}_1 \cup \dots \cup \mathcal{E}_n) \leq \frac{1}{2}$.

UNION BOUND.

$$\Pr(\mathcal{E}_1 \cup \dots \cup \mathcal{E}_n) \leq \sum_{i=1}^{n} \Pr(\mathcal{E}_i).$$

$$\text{LHS} = \mathbb{E}\left[\begin{cases} 1 & \text{if } \mathcal{E}_1 \cup \dots \cup \mathcal{E}_n \\ 0 & \text{if not} \end{cases}\right]$$

$$\text{RHS} = \mathbb{E}\left[\sum_{i=1}^{n} \begin{cases} 1 & \text{if } \mathcal{E}_i \\ 0 & \text{if not} \end{cases}\right]$$

<u>New goal:</u> Make m large enough that

$$\Pr(\mathcal{E}_i) \le \frac{1}{2n} \quad \forall i.$$

$\mathcal{E}_i$ is $\left\{ L_i \notin \left[(1-\delta)\frac{m}{n}, (1+\delta)\frac{m}{n} \right] \right\}$

$$\| $$

$$\left| L_i - \frac{m}{n} \right| \ge \frac{\delta m}{n}.$$

<u>Chebyshev</u> For rand var Y with expectation $\mathbb{E}Y$ and variance $\mathrm{Var}(Y)$

$$\Pr\left(|Y - \mathbb{E}Y| \geq \lambda\right) \leq \frac{\mathrm{Var}(Y)}{\lambda^2}.$$

In our application

$$Y = L_i$$

$$\mathbb{E}Y = \frac{m}{n}$$

$$\mathrm{Var}(Y) = m \cdot \frac{1}{n} \cdot \left(1 - \frac{1}{n}\right)$$

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

$$\mathrm{Var}(Y) = \sum_{j=1}^{m} \mathrm{Var}(X_{ij})$$

For $X = \begin{cases} 1 & \text{with prob } p \\ \emptyset & \text{" " " } 1-p \end{cases}$ $\quad p = \frac{1}{n}$

$$\mathbb{E}[X] = p$$

$$\mathbb{E}[X^2] = p$$

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$$

$$= p - p^2$$

$$= p(1-p).$$

$$\mathrm{Var}(Y) = \sum_{j=1}^{m} \frac{1}{n}\left(1 - \frac{1}{n}\right)$$

$$= \frac{m}{n}\left(1 - \frac{1}{n}\right) < \frac{m}{n}.$$

Cheby:

$$\Pr\left(\left| L_i - \frac{m}{n} \right| \geq \frac{\delta m}{n} \right) < \frac{m/n}{\left(\delta m / n \right)^2}$$

$$= \frac{\cancel{m/n}}{\delta^2 \cdot m/n \cdot \cancel{m/n}}$$

$$= \frac{n}{\delta^2 m}$$

We want this to be $\leq \frac{1}{2n}$.

Solve for m:

$$\frac{n}{\delta^2 m} \leq \frac{1}{2n}$$

$$\frac{18 n^2}{\varepsilon^2} = \frac{2n^2}{\delta^2} \leq m$$

Chebyshev is too pessimistic.

Instead for the Tail Bound (TB)

$$\Pr\left(\left| L_i - \frac{m}{n} \right| \geq \delta \cdot \frac{m}{n} \right) \leq ??$$

use Chernoff Bound.

Chernoff. If $X_1, \dots, X_m$ are independent random variables taking values in $[0,1]$ and

$$X = X_1 + \cdots + X_n$$

then for all $0 \leq \varepsilon \leq 1$

$$\boxed{\begin{aligned} \Pr\left(X \geq (1+\varepsilon)\mathbb{E}X \right) &\leq e^{-\frac{1}{3}\varepsilon^2 \cdot \mathbb{E}X} \\ \Pr\left(X \leq (1-\varepsilon)\mathbb{E}X \right) &\leq e^{-\frac{1}{2}\varepsilon^2 \cdot \mathbb{E}X} \end{aligned}}$$

Goal: $\Pr\left(L_i > (1+\delta)\frac{m}{n}\right) + \Pr\left(L_i < (1-\delta)\frac{m}{n}\right) \leq \frac{1}{2n}$.

Will be using Chernoff with

$$X = L_i$$

local variable ε in Chernoff Bound statement.

$$\mathbb{E}X = \frac{m}{n}$$

$$\varepsilon \cdot \mathbb{E}X = \frac{\delta m}{n} \Longrightarrow \varepsilon = \delta,$$

$$\leq e^{-\frac{1}{3}\delta^2 m/n}$$

$$\leq e^{-\frac{1}{2}\delta^2 m/n}$$

// local variable ε out of scope.

Solve for m:

$$e^{-\frac{1}{3}\delta^2 m/n} + e^{-1/2\,\delta^2 m/n} \leq \frac{1}{2n}.$$

Easier:

$$e^{-\frac{1}{3}\delta^2 m/n} + e^{-\frac{1}{3}\delta^2 m/n} \le \frac{1}{2n}$$

$$e^{-1/3\,\delta^2\, m/n} \le \frac{1}{4n}$$

$$\frac{1}{3}\delta^2 \frac{m}{n} \ge \ln(4n)$$

$$m \ge 3n \ln(4n)/\delta^2$$

$$= 27n \ln(4n)/\varepsilon^2.$$

<u>Conclusion</u>: Throwing $m \ge 27n\ln(4n)/\varepsilon^2$ balls ensures w. prob $\ge 1/2$

$$\frac{\text{max load}}{\text{min load}} \leq 1+\varepsilon.$$

Chernoff, unlike Chebyshev, gives an answer which is tight within constant factor.