Today: Power of 2 choices

– useful for hashing schemes

Balls into bins – simple scheme

Setting: n balls & n bins

– Random assignment:

For each ball, independently assign to a uniformly at random bin.

Let $L(i) :=$ load of bin i = # of balls assigned to bin i

Max load $= \max_i L(i)$

How large is the max load?

$O(\log n)$ with high probability (whp)

– Can show it's $(1+o(1))\dfrac{\log n}{\log\log n}$ whp

$$\Pr\left(L(i) > 2\log n\right) \le \binom{n}{2\log n}\left(\frac{1}{n}\right)^{2\log n}$$

$$\le \frac{\cancel{n^{2\log n}}}{(2\log n)!}\left(\cancel{\frac{1}{n}}\right)^{2\log n} \qquad \left(\text{Since } \binom{n}{k} \le \frac{n^k}{k!}\right)$$

$$\le \left(\frac{e}{2\log n}\right)^{2\log n} \qquad \left(\text{Since } k! \ge \left(\frac{k}{e}\right)^k\right)$$

$$\le \left(\frac{1}{2}\right)^{2\log n}$$ for n sufficiently large so that $\log n > e$.

$$\le \frac{1}{n^2}$$

Hence by a union bound,

with prob. $\ge 1 - \frac{1}{n}$, max load $\le 2\log n$.

Can we improve this constant?

Yes, $\Pr\left(L(i) > \log n + \log\log n\right) \le \left(\frac{1}{n}\right)\left(\frac{1}{\log n}\right)$

Hence, max load $\le \log n + \log\log n$ w. prob. $\ge 1 - \frac{1}{\log n}$

this is $(1+o(1))\log n$

Better approach:

Assign balls <u>sequentially</u> into bins

For ball $i = 1 \to n$:

- Choose 2 random bins j & k.
- Check $L(j)$ & $L(k)$ for their current loads.
- If $L(j) < L(k)$ then:
    - assign ball i to bin j
- If $L(k) \leq L(j)$ then:
    - assign ball i to bin k

(In other words, assign ball i to the least loaded of 2 randomly chosen bins)

Theorem: [Azar, Broder, Karlin, Upfal '94]

Max load is $O(\log\log n)$ with high probability.

More generally, with $d \geq 2$ choices, it's $O\left(\frac{\log\log n}{\log d}\right)$

Proof high-level idea:

Let B_i = # of bins with load $\geq i$ at the end of the assignment

Suppose we could prove $B_i \leq \beta_i$ whp

Then,

$$\Pr\left(\begin{array}{c}\text{ball } i \text{ is assigned} \\ \text{to a bin with load} \geq i\end{array}\right) \leq \left(\frac{\beta_i}{n}\right)^2$$

since need $L(j), L(k) \geq i$ for it to occur.

Thus, $B_{i+1} \leq \text{Bin}\left(n, \left(\frac{\beta_i}{n}\right)^2\right)$

the mean of is $\frac{\beta_i^2}{n}$

& it should be close by Chernoff bound.

Thus, $\beta_{i+1} \approx n\left(\frac{\beta_i}{n}\right)^2 = \frac{\beta_i^2}{n}$

Note, $\beta_2 = \frac{n}{2}$ holds since $\leq \frac{n}{2}$ bins have ≥ 2 balls.

Then, the recurrence solves to: $\beta_{i+2} = n/2^{2^i}$

hence for $i^* \approx \log\log n$ we get $\beta_{i+2} < 1$,

So no ~~bins~~ bins have load $> i^* \approx \log\log n$.

Now let's formalize the proof.

Proof:

Base case: $\beta_6 = \frac{n}{2e}$

Note, $\leq \frac{n}{6}$ bins have load ≥ 6.

Since $\frac{n}{6} < \frac{n}{2e}$ we know $B_6 \leq \beta_6$.

For $i > 6$, let

$$\beta_{i+1} = \frac{e\beta_i^2}{n}$$

Let the event $\mathcal{G}_i = \{B_i \leq \beta_i\}$

$\&\ \mathcal{B}_i = \overline{\mathcal{G}_i} = \{B_i > \beta_i\}$

Note,

$$\Pr(\mathcal{B}_{i+1} \mid \mathcal{G}_i) = \Pr(B_{i+1} > \beta_{i+1} \mid \mathcal{G}_i)$$

$$\leq \frac{\Pr\left(\mathrm{Bin}\left(n, \left(\frac{\beta_i}{n}\right)^2\right) > \beta_{i+1}\right)}{\Pr(\mathcal{G}_i)}$$

By a Chernoff bound, $\Pr(X \geq e\mu) \leq e^{-\mu}$,

thus $$\Pr(\mathcal{B}_{i+1} \mid \mathcal{G}_i) \leq \frac{e^{-\beta_i^2/n}}{\Pr(\mathcal{G}_i)} \leq \frac{1/n^2}{\Pr(\mathcal{G}_i)}$$

assuming $\frac{\beta_i^2}{n} \geq 2 \ln n$.

Now let's bound $\Pr(\mathcal{G}_i)$.

<u>Claim:</u> $\Pr(\mathcal{B}_i) \leq i/n^2$ assuming $\frac{\beta_i^2}{n} \geq 2\ln n$.

Using the claim, let i^* be the min i where $\beta_i^2 < 2n\ln n$.

Since $\beta_{i+1} = \frac{e\beta_i^2}{n}$ then $i^* = \frac{\ln\ln n}{\ln 2}$

Thus, for $i^* \approx \ln\ln n$ we have $\beta_{i^*} \leq \sqrt{2n\ln n}$

& we conclude that:

$\leq \sqrt{2n\ln n}$ bins have load $\geq \ln\ln n$

with high probability $\geq 1 - \frac{i^*}{n^2} \geq 1 - \frac{1}{n}$.

$\cancel{\geq 1 - \frac{1/n^2}{\left(\frac{k\ln\ln n}{n^2}\right)} \geq 1 - 1/n}$

Proof of claim:

Base case: we know $\Pr(\mathcal{B}_0) = 0$ ✓

In general, recall $\Pr(\mathcal{B}_{i+1} \mid \mathcal{G}_i) \leq \dfrac{1/n^2}{\Pr(\mathcal{G}_i)}$

thus:

$$\Pr(\mathcal{B}_{i+1}) \leq \Pr(\mathcal{B}_{i+1} \mid \mathcal{G}_i)\Pr(\mathcal{G}_i) + \Pr(\mathcal{B}_{i+1} \mid \mathcal{B}_i)\Pr(\mathcal{B}_i)$$

$$\leq \frac{1/n^2}{\Pr(\mathcal{G}_i)}\Pr(\mathcal{G}_i) + \frac{\Pr(\mathcal{B}_{i+1}, \mathcal{B}_i)}{\cancel{\Pr(\mathcal{B}_i)}}\cancel{\Pr(\mathcal{B}_i)}$$

$$\leq \frac{1}{n^2} + \Pr(\mathcal{B}_i)$$

$$\leq \frac{1}{n^2} + \frac{i}{n^2} \quad \text{by induction}$$

$$\leq \frac{(i+1)}{n^2}. \qquad \blacksquare$$

Once again we ~~know~~ have that:

$$\Pr(B_i) \le \frac{1}{n^2} \le \frac{1}{n} \text{ for all } i \text{ where}$$

$$\frac{\beta_i^2}{n} \ge 2\ln n$$

Let i^* be the min i where $\beta_i^2 < 2n\ln n$.

Solving the recurrence $\beta_{i+1} = \frac{e\beta_i^2}{n}$

we have $i^* = \frac{\ln\ln n}{\ln 2}$ & since $\beta_{i^*}^2 \le 2n\ln n$

so $\beta_{i^*} \le \sqrt{2n\ln n}$.

Therefore, $\le \sqrt{2n\ln n}$ bins have load $\ge \ln\ln n$

with prob. $\ge 1 - 1/n$.

To finish off the proof:

Claim: $\Pr(B_{i^*+2} \geq 1) = O\left(\frac{\log^2 n}{n}\right)$

Proof:

Let $\mathcal{G}_{i^*+1} = \{B_{i^*+1} \leq 6 \ln n\}$

$$\Pr(B_{i^*+1}) \leq \Pr(B_{i^*+1} \geq 6 \ln n \mid \mathcal{G}_{i^*}) \Pr(\mathcal{G}_{i^*}) + \Pr(B_{i^*})$$

$$\leq \Pr\left(\mathrm{Bin}\left(n, \frac{2 \ln n}{n}\right) \geq 6 \ln n\right) + \frac{1}{n}$$

from claim.

$$\leq \frac{1}{n^2} + \frac{1}{n} \quad \text{by Chernoff bound}$$

$$= O(1/n)$$

Now for i^*+2:

$$\Pr(B_{i^*+2} \geq 1) \leq \Pr(B_{i^*+2} \geq 1 \mid \mathcal{G}_{i^*+1}) \Pr(\mathcal{G}_{i^*+1}) + \Pr(\mathcal{B}_{i^*+1})$$

$$\leq \Pr\left(\mathrm{Bin}\left(n, \left(\frac{6\ln n}{n}\right)^2\right) \geq 1\right) + O(1/n)$$

$$\leq n\left(\frac{6\ln n}{n}\right)^2 + O(1/n)$$

$$= O\left(\frac{(\ln n)^2}{n}\right) = o(1). \quad \blacksquare$$