CSE 632: Analysis of Algorithms II: Randomized Algorithms **Spring 2024**

Lecture 15: Balls into Bins

Lecturer: Zongchen Chen

1 Balls into Bins Problem

We have n balls and n bins. Each ball is assigned to a bin independently and uniformly at random. Define the load of bin i to be

$$L(i) = \#\text{ of balls assigned to bin } i.$$

Define the max load to be

$$\textsf{max-load} = \max_i L(i).$$

We would like to know how large $\textsf{max-load}$ is with high probability. This is given by the following lemma.

Lemma 1. *With probability* $1 - o(1)$, *it holds*

$$\textsf{max-load} = (1 + o(1)) \frac{\log n}{\log \log n}.$$

We shall prove a weaker version: with probability $1 - \frac{1}{n}$, it holds

$$\textsf{max-load} \le \frac{3 \log n}{\log \log n}.$$

Proof. For all bin i, for any s, we have

$$\begin{aligned}
\Pr\left(L(i) \ge s\right) &\le \binom{n}{s} \cdot \left(\frac{1}{n}\right)^s && \text{(Union bound)} \\
&\le \left(\frac{en}{s}\right)^s \cdot \left(\frac{1}{n}\right)^s && \left(\tbinom{n}{s} \le \left(\tfrac{en}{s}\right)^s\right) \\
&= \left(\frac{e}{s}\right)^s \\
&= \exp\left(-s(\log s - 1)\right).
\end{aligned}$$

Now take

$$s = \frac{3 \log n}{\log \log n}.$$

We have that

$$\log s - 1 = \log \log n - \log \log \log n + \log 3 - 1 = (1 - o(1)) \log \log n.$$

Therefore,

$$s(\log s - 1) \ge 2 \log n.$$

It follows that

$$\Pr\left(L(i) \geq \frac{3\log n}{\log\log n}\right) \leq \frac{1}{n^2},$$

and the union bound yields

$$\Pr\left(\mathsf{max\text{-}load} \geq \frac{3\log n}{\log\log n}\right) \leq \frac{1}{n},$$

as claimed. □

2 Power of Two Choices

Instead of picking one random bin, if pick two bins uniformly at random and place the ball into the bin with fewer balls, then the max-load turns out to be much smaller. This paradigm is known as the *power of two random choices.*

Algorithm 1 Partially random allocation scheme

for ball $i = 1$ to n **do**
    Choose two bins a and b u.a.r.
    if $L(a) < L(b)$ **then**
        Assign ball i to bin a
    else ▷ $L(a) \geq L(b)$
        Assign ball i to bin b
    end if
end for

In other words, we assign each ball to the less loaded of two randomly chosen bins.

Theorem 2. *With probability* $1 - O(\frac{\log^2 n}{n})$*, it holds*

$$\mathsf{max\text{-}load} = O(\log\log n).$$

Remark 3. More generally, with $d \geq 2$ choices, with high probability it holds

$$\mathsf{max\text{-}load} = O\left(\frac{\log\log n}{\log d}\right).$$

Application in hashing. Consider a hash table in which all keys mapped to the same location are stored in a linked list. The efficiency of accessing a key depends on the length of its list. If we use a single hash function (which selects locations uniformly at random), then with high probability the longest list has size $O(\frac{\log n}{\log\log n})$. If we use two hash functions and put each new key in the shorter of the two lists, then with high probability the longest list has size $O(\log\log n)$.

High-level proof idea of Theorem 2. Let B_i be the number of bins with load at least i at the end. We would like to prove by induction that $B_i \leq \beta_i$ w.h.p. for all $2 \leq i \leq i^*$, where $\{\beta_i\}$ is a sequence of (deterministic) upper bounds to be specified. In particular, we want $i^* = O(\log\log n)$ and $\beta_{i^*} < 1$. Thus, the claim $B_{i^*} \leq \beta_{i^*}$ is equivalent to saying $B_{i^*} = 0$, namely, there is no bin with load at least $i^* = O(\log\log n)$. This implies that the max-load $= O(\log\log n)$ w.h.p.

For the base case, notice that there are at most $\frac{n}{2}$ bins with load at least 2 always. Hence, we have $B_2 \leq \beta_2$ with probability 1 where $\beta_2 = \frac{n}{2}$.

Now suppose we could prove $B_i \leq \beta_i$ w.h.p. for some i, and we aim to prove $B_{i+1} \leq \beta_{i+1}$ w.h.p. We say a ball has height h if it is the h'th ball assigned to its bin. Notice that for every ball k, we have non-rigorously

$$\Pr(\text{ball } k \text{ has height} \geq i+1) \leq \left(\frac{\beta_i}{n}\right)^2,$$

since we need to choose both bins with load at least i which happens with probability at most $(\frac{\beta_i}{n})^2$ assuming $B_i \leq \beta_i$ indeed holds. Therefore, we deduce that

$$\begin{aligned}
B_{i+1} &= \#\text{ of balls with height} = i+1 \\
&\leq \#\text{ of balls with height} \geq i+1 \\
&\leq \mathrm{Bin}\left(n, \left(\frac{\beta_i}{n}\right)^2\right) \\
&\approx n \cdot \left(\frac{\beta_i}{n}\right)^2 = \frac{\beta_i^2}{n},
\end{aligned}$$

where $\mathrm{Bin}(n,p)$ represents a binomial random variable with parameters $n \in \mathbb{N}^+$ and $p \in [0,1]$. The "$\approx$" is expected to hold due to concentration inequalities such as the Chernoff bounds. Therefore, hopefully we could prove $B_{i+1} \leq \beta_{i+1}$ w.h.p. where

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

Thus, we could show that $B_i \leq \beta_i$ w.h.p. for all $i \geq 2$ where $\{\beta_i\}$ is a sequence of upper bounds satisfying the recurrence

$$\frac{\beta_2}{n} = \frac{1}{2} \quad \text{and} \quad \frac{\beta_{i+1}}{n} = \left(\frac{\beta_i}{n}\right)^2.$$

Solving the recurrence gives

$$\beta_i = \frac{n}{2^{2^{i-2}}}.$$

Now we take $i^* = c \log\log n$ so that $\beta_{i^*} < 1$. It follows that $B_{i^*} \leq \beta_{i^*} < 1$ w.h.p., namely, there is no bin with load at least $i^* = c \log\log n$ w.h.p.