15-859(M): Randomized Algorithms **Lecturer:** Anupam Gupta
Topic: Balls and bins: the power of two choices **Date:** Oct 13.
Scribe: Páll Melsted

10.1 2-choice method

The objective is to throw n balls into n bins, where balls are inserted on after the other. For every ball we pick two bins uniformly at random, inspect the bins and throw the ball into the bin with lesser load (breaking ties arbitrarily).

Theorem 10.1.1 *In the 2-choice method, the maximum load in any bin is w.h.p.*

$$\frac{\ln(\ln(n))}{\ln(2)} + O(1)$$

Before the going into the technical details behind the proof we will give definitions and outline the general idea of the proof.

The height of a ball is i if it was the i-th ball to be placed in it's bin. Observe that

$$(\text{ total \# of balls of height } i) \geq (\text{ total \# of bins with } \geq i \text{ balls })$$

Let N_i be the number of bins with more than i balls and assume $\frac{N_i}{n} \leq \alpha_i$. Then

$$\mathbf{Pr}[\text{ ball has height } \geq i+1] \leq \alpha_i^2$$

since to place a ball at height $i+1$ we would have to select two bins, both with balls of height more than i. So we expect at most a fraction $\leq \alpha_{i+1} \approx \alpha_i^2$ of the bins to have height $i+1$ (see fig 10.1). We also notice that at most half of the bins can have more than 2 balls so $\alpha_2 \leq \frac{1}{2}$. This gives the recurrence equation

$$\alpha_i \approx \frac{1}{2^{2^{i-1}}}$$

and so we see that $\alpha_{\log(\log(n))} \lessapprox \frac{1}{n}$ so there are no balls with height $\log(\log(n)) = \ln(\ln(n))/\ln(2)$.

10.2 Formal arguments

The previous argumant relied on balls behaving "as expected", we will now give a formal proof to the theorem.

Lemma 10.2.1 *Suppose we have marked at most an α fraction of the bins. We say that a ball is marked if both of the bins it inspects are marked. Let X be the number of marked balls. Then* $\mathbf{E}[X] \leq n\alpha^2$.

Claim 10.2.2 *Suppose $\alpha^2 \leq \frac{9\log(n)}{n}$, then $X \leq 2n\alpha^2$ with probability at least $1 - \frac{1}{n^3}$.*

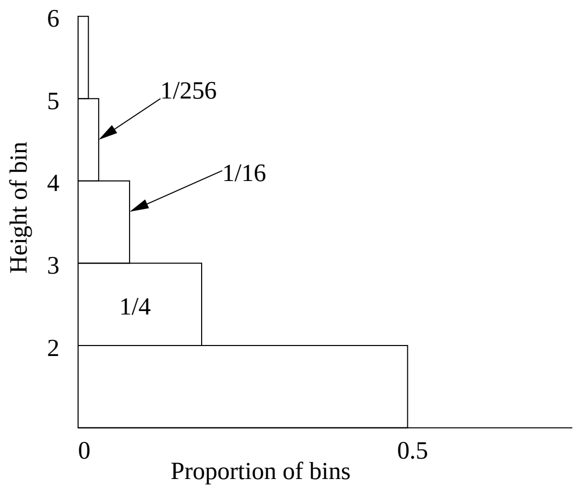

Figure 10.1.1: Expected distribution of load on bins

Proof: We use the Chernoff bound with $\epsilon = \lambda/\mu = \frac{n\alpha^2}{\mu}$

$$\begin{aligned}
\mathbf{Pr}\big[X \geq 2n\alpha^2\big] &\leq \mathbf{Pr}\big[X - \mu \geq n\alpha^2\big] \\
&\leq \exp\left(-\frac{\epsilon^2\mu}{2+\epsilon}\right) = \exp\left(-\frac{(\lambda/\mu)^2\mu}{2+(\lambda/\mu}\right) \\
&= \exp\left(-\frac{\lambda^2}{2\mu+\lambda}\right) \qquad \text{but } \lambda \geq \mu \\
&\leq \exp\left(\frac{\lambda}{3}\right) \qquad \text{but } \lambda = n\alpha^2 \geq 9\log(n) \\
&\leq \exp\left(-3\log(n)\right) \leq \frac{1}{n^3}
\end{aligned}$$

■

Now define $\alpha_3 = \frac{1}{3}$ and $\alpha_i = 2\alpha_{i-1}^2$ for $i \geq 4$. Then we have

$$\alpha_i = 2^{2^{i-3}-1} \cdot \left(\frac{1}{3}\right)^{2^{i-3}}$$

and for $i^* = \log_2(\log_{3/2}(n+3)$ we have $\alpha_{i^*} = \frac{1}{2n} \leq 9\log(n)/n$.

We say that the bad event $\mathcal{E}_i$ happens if $(N_i \leq \alpha_i)$, where $i \leq i^*$. Notice that since at most a third of the bins can have 3 balls or more we have

$$\mathbf{Pr}[\mathcal{E}_3] = \mathbf{Pr}[N_3 \leq 1/3] = 1$$

Claim 10.2.3 *If* $\alpha_i^2 \geq \frac{9\log(n)}{n}$ *then* $\mathbf{Pr}[\neg\mathcal{E}_{i+1}] \leq in^3$.

Proof: The base case is trivial since $\mathbf{Pr}[\neg\mathcal{E}_3] = 0$. Notice that $\mathcal{E}_{i+1} \subseteq \mathcal{E}_i$ so we get

$$\mathbf{Pr}[\neg\mathcal{E}_{i+1}] = \mathbf{Pr}[\neg\mathcal{E}_{i+1} \mid \mathcal{E}_i]\mathbf{Pr}[\mathcal{E}_i] + \mathbf{Pr}[\neg\mathcal{E}_i] \leq \mathbf{Pr}[\neg\mathcal{E}_{i+1} \mid \mathcal{E}_i] + \mathbf{Pr}[\neg\mathcal{E}_i]$$

We have that

$$\mathbf{Pr}[\neg\mathcal{E}_{i+1} \mid \mathcal{E}_i] = \mathbf{Pr}\big[\text{ mark at most } \alpha_i \text{ bins, and mark more than } 2\alpha_i^2 \text{ balls}\big]$$

By claim 10.2.2 we see that $\mathbf{Pr}[\neg\mathcal{E}_{i+1} \mid \mathcal{E}_i] \leq 1/n^3$. And by induction we get that

$$\mathbf{Pr}[\neg\mathcal{E}_{i+1}] \leq \frac{1}{n^3} + \frac{i}{n^3} = \frac{i+1}{n^3}$$

■

The union bound and this claim then gives

$$\mathbf{Pr}\left[\bigcup_i \neg\mathcal{E}_{i+1}\right] \leq \frac{1}{n}$$

So w.h.p. $N_i \leq \alpha_i$ as long as $\alpha_i^2 \geq \frac{9\log(n)}{n}$, the point at which $\alpha_i^2 \simeq \frac{9\log(n)}{n}$ is $i^* \approx \frac{\ln(\ln(n))}{\ln(2)} + O(1)$. At this point, w.h.p. the number of bins with $\geq \frac{\ln(\ln(n))}{\ln(2)} + O(1)$ balls is at most $18\log(n)$ by claim 10.2.2. Now we condition on $\mathcal{E}_{i^*}$ happening, this means that the fraction of bins with more then i^* balls is at most $18\log(n)/n$. Then we use the union bound for k balls to get

$$\begin{aligned}\mathbf{Pr}[k \text{ balls have height } > i^*] &\leq \binom{n}{k}\left(\frac{18\log(n)}{n}\right)^{2k} \\ &\leq \left(\frac{en}{k}\right)^k\left(\frac{18\log(n)}{n}\right)^{2k} \\ &\leq \left(\frac{18^2 e\log^2(n)}{nk}\right)^k \\ &\approx \left(\frac{O(\log^(n))}{nk}\right)^k\end{aligned}$$

So for $k = 3$ we get $\mathbf{Pr}[\text{ 3 balls or more have height } > i^*] \leq O(1)\frac{1}{2}$. This shows that w.h.p. the number of balls in any bins is at most $i^* + 2$ or

$$\text{load } \leq \frac{\ln(\ln(n))}{\ln(2)} + O(1) \qquad\qquad \text{w.p. } \Omega(1 - \frac{1}{n})$$

10.3 Random graphs

Another way to show that the maximum load is $O(\log\log(n))$ is to use a random graph process. We will build the graph G iteratively, the vertices of G correspond to the bins and each time we probe to bins we connect them with an edge in G.

Theorem 10.3.1 *If we place $\frac{n}{512}$ balls into n bins, then max load $= O(\log\log(n))$, w.h.p.*

So if we merge every 512 consecutive bins the maximum load will go up by "only" a factor of 512, so we get

Corollary 10.3.2 *For n balls into n bins we get max load $= O(\log\log(n))$, w.h.p.*

We will need two result to prove this

Claim 10.3.3 *Size of G's largest component is $= O(\log(n))$ w.h.p.*

Claim 10.3.4 *For all subsets S of the vertex set, with $|S| \geq K$ the induced graph $G[S]$ has average degree ≤ 5, w.p. $1 - \frac{1}{2n}$.*

Now once we have the graph suppose we remove all vertices of degree ≤ 10 in the graph and repeat.

Claim 10.3.5 *This process ends after $O(\log\log(n))$ w.h.p.*

Proof: Condition on the events in the two previous claims. Now look at any component C of the graph. Its average degree is ≤ 5 so the number of nodes with degree ≥ 10 is at most $|C|/2$ (this is just Markov). So as long as we have at least K nodes in a component we will remove at least half of the nodes in the component. Thus

$$\# \text{ of rounds } = O(\log(|C|/K)) = O(log(|C|)) - \log(K)$$

but since $|C| = O(log(n))$ we get that after $O(\log\log(n))$ steps we're down to K nodes. ■

Claim 10.3.6 *If the removal process gives us a graph with K vertices and if edge e_i is removed in round t then the maximum height of ball i is $\leq 10t + K$.*

Proof: The balls that are removed in the first process have height ≤ 10, because each vertex that was removed had degree ≤ 10. Similarly in the next round each ball collides with at most 9 other balls in the process, since in the graph a vertex corresponding to a bin has at most 10 incoming balls. Thus in the worst case max height of a ball is ≤ 20. This continues untill we have only K balls left, but then every vertex has degree at most $K - 1$ so we have that max height is at most $10t + K$. ■

These two claims prove that with high probability the maximum load is $O(\log\log(n))$. We now only need to prove claims 10.3.3 and 10.3.4.

Proof of claim 10.3.3: We have a graph with n vertices and $m = \frac{n}{512}$ edges where we connect vertices at random.

$$\begin{aligned}\mathbf{Pr}[k+1 \text{ vertices connected }] &\leq \mathbf{Pr}[k \text{ edges in a set of size } k+1] \\ &\leq \binom{m}{k}\left(\frac{\binom{k+1}{2}}{\binom{n}{2}}\right)^k \\ &\approx \binom{m}{k}\left(\frac{8k}{n}\right)^{2k}\end{aligned}$$

Now the probability that any such set exists can bounded above by the union bound

$$
\begin{aligned}
\mathbf{Pr}[\exists \text{ a connected set of size } k+1] &\leq \binom{n}{k+1}\binom{m}{k}\left(\frac{8k}{n}\right)^{2k} \\
&\leq n\left(\frac{ne}{k}\right)^{k}\left(\frac{ne}{512k}\right)^{k}\left(\frac{8k}{n}\right)^{2k} \\
&\leq n\left(\frac{e^2}{8}\right)^{k} \leq \frac{1}{2n} \quad \text{if } k = O(\log(n))
\end{aligned}
$$

which proves the claim. ■

Proof of claim 10.3.4: We want to show that the average degree of the graph $G[S]$ induced by a subset S no more than 5 w.h.p. as long as $|S| \geq K$ for some fixed K. Now if $G[S]$ would have average degree greater than 5 it would have to contain at least $\frac{5|S|}{2}$ edges. We will now bound the probability that S gets more than $\frac{5|S|}{2}$ edges.

$$
\begin{aligned}
\mathbf{Pr}\left[\text{ a set of size } k \text{ gets more than } \frac{5k}{2} \text{ edges }\right] &= \binom{m}{5k/2}\left(\frac{k^2}{n^2}\right)^{5k/2} \\
&= \binom{m}{5k/2}\left(\frac{k}{n}\right)^{5k}
\end{aligned}
$$

now the probability that such a set exists is

$$
\begin{aligned}
\mathbf{Pr}[\exists \text{ such a set }] &\leq \binom{n}{k}\binom{m}{5k/2}\left(\frac{k}{n}\right)^{5k} \\
&\leq \left(\frac{ne}{k}\right)^{k}\left(\frac{ne}{512(5k/2)}\right)^{5k/2}\left(\frac{k}{n}\right)^{5k} \\
&= \left(\frac{k}{n}\right)^{3k/2}\alpha^{k}
\end{aligned}
$$

where $\alpha = \frac{e^{7/2}}{1280^{5/2}} < 1$ then we use the union bound again and sum over k from K to n.

$$
\begin{aligned}
\mathbf{Pr}[\text{ a violating set exists }] &\leq \sum_{k=K}^{n} \left(\frac{k}{n}\right)^{3k/2} \alpha^k \\
&\leq \sum_{k=K}^{\sqrt{n}} \left(\frac{k}{n}\right)^{3k/2} \alpha^k + \sum_{k=\sqrt{n}}^{n} \left(\frac{k}{n}\right)^{3k/2} \alpha^k \\
&\leq \left(\frac{1}{\sqrt{n}}\right)^{3K/2} \sum_{k=K}^{\sqrt{n}} \alpha^k + \sum_{k=\sqrt{n}}^{n} \alpha^k \\
&\leq \frac{1}{n^{3K/4}} \sum_{k=K}^{\sqrt{n}} \alpha^k + \alpha^{\sqrt{n}} \sum_{k=0}^{n-\sqrt{n}} \alpha^k \\
&\leq \left(\frac{1}{n^{3K/4}} + \alpha^{\sqrt{n}}\right) \sum_{k=0}^{\infty} \alpha^k \\
&= \left(\frac{1}{n^{3K/4}} + \alpha^{\sqrt{n}}\right) \frac{1}{1-\alpha}
\end{aligned}
$$

So we see that for K large enough (like 4) the probability of the existance of such a set is $O(\frac{1}{n^3})$. ■