The power of two choices

By Sariel Har-Peled, February 17, 2022

The Peace of Olivia. How sweet and peaceful it sounds! There the great powers noticed for the first time that the land of the Poles lends itself admirably to partition.

Consider the problem of throwing $n$ balls into $n$ bins. It is well known that the maximum load is $\Theta(\log n / \log \log n)$ with high probability. Here we show that if one is allowed to pick $d$ bins for each ball, and throw it into the bin that contains less balls, then the maximum load of a bin decreases to $\Theta(\log \log n / \log d)$. A variant of this approach leads to maximum load $\Theta(\log \log n / d)$.

As a concrete example, for $n = 10^9$, this leads to maximum load 13 in the regular case, compared to maximum load of 4, with only two-choices – see Figure 31.1.

31.1. Balls and bins with many rows

Consider throwing $n$ balls into $n$ bins. Every bin can contain a single ball. As such, as we throw the balls, some balls would be rejected because their assigned bin already contains a ball. We collect all the rejected balls, and throw them again into a second row of $n$ bins. We repeat this process till all the balls had found a good home (i.e., empty bin). How many rows one needs before this process is completed?

Lemma 31.1.1. Let $m = \alpha n$ balls be thrown into $n$ bins. Let $Y_{\text{end}}$ the number of bins that are not empty in the end of the process (here, we allow more than one ball into a bin).

(A) For $\alpha \in (0, 1]$, we have $\mu = \mathbb{E}[Y_{\text{end}}] \geq m \exp(-\alpha)$.

(B) If $\alpha \geq 1$, then $\mu = \mathbb{E}[Y_{\text{end}}] \geq n(1 - \exp(-\alpha))$.

(C) We have $\mathbb{P}[|Y_{\text{end}} - \mu| > \sqrt{3cm \log n}] \leq 1/n^c$.

Proof: (A) The probability of the $i$th ball to be the first ball in its bin, is $(1 - \frac{1}{n})^{i-1}$. To see this we use backward analysis – throw in the $i$th ball, and now throw in the earlier $i - 1$ balls. The probability that none of the earlier balls hits the same bin as the $i$th ball is as stated. Now, the expected number of non-empty bins is the number of balls that are first in their bins, which in turn is

$$
\mu = \sum_{i=0}^{m-1} \left(1 - \frac{1}{n}\right)^i \geq m(1 - 1/n)^m \geq m(1 - 1/n)^{(n-1)m/(n-1)} \geq m \exp\left(-\frac{m}{n-1}\right) = m \exp\left(-\frac{\alpha(n-1) + \alpha}{n-1}\right)
$$

$$
= m \exp\left(-\alpha + \frac{\alpha}{n-1}\right) \geq m \exp(-\alpha) \geq \frac{m}{e}.
$$

using $m = \alpha n \leq n$, and $(1 - 1/n)^{n-1} \geq 1/e$, see Lemma 31.5.3.

(B) We repeat the above analysis from the point of view of the bin. The probability of a bin to be empty is $(1 - 1/n)^{\alpha n}$. As such, we have that

$$
\mu = \mathbb{E}[Y_{\text{end}}] = n(1 - (1 - 1/n)^{\alpha n}) \geq n(1 - \exp(-\alpha)),
$$

This work is licensed under the Creative Commons Attribution-Noncommercial 3.0 License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/3.0/ or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.
using $1 - 1/n \leq \exp(-1/n)$.

(C) Let $X_i$ be the index of the bin the $i$th ball picked. Let $Y_i = E[Y_{\text{end}} \mid X_1, \ldots, X_i]$. This is a Doob martingale, with $|Y_i - Y_{i-1}| \leq 1$. As such, Azuma’s inequality implies, for $\lambda = \sqrt{3cm \ln n}$, that

$$
\Pr[|Y_{\text{end}} - E[Y_{\text{end}}]| \geq \lambda] \leq 2 \exp(-\lambda^2/2m) \leq 1/n^c.
$$

\[\blacksquare\]

Remark. The reader might be confused by cases (A) and (B) of Lemma 31.1.1 for $\alpha = 1$, as the two lower bounds are different. Observe that (A) is loose if $\alpha$ is relatively large and close to 1.

Back to the problem. Let $\alpha_1 = 1$ and $n_1 = \alpha_1 n$. For $i > 1$, inductively, assume that numbers of balls being thrown in the $i$th round is

$$
n_i = \alpha_i n + O(\sqrt{\alpha_i^{-1} n \log n}).
$$

By Lemma 31.1.1, with high probability, the number of balls stored in the $i$th row is

$$
s_i = n_i \exp(-\alpha_i) \pm O(\sqrt{n_i \log n}).
$$

As such, as long as the first term is significantly large than the second term, we have that $s_i = na_i \exp(-\alpha_i)(1 \pm o(1))$. For the time being, let us ignore the $o(1)$ term. We have that

$$
n_{i+1} = n_i - s_i = n(\alpha_i - \alpha_i \exp(-\alpha_i)) \leq n(\alpha_i - \alpha_i (1 - \alpha_i)) = na_i^2,
$$

since $\exp(-\alpha_i) \geq 1 - \alpha_i$.

Observation 31.1.3. Consider the sequence $\alpha_1 = 1$, $c = \alpha_2 = 1 - 1/e$, and $\alpha_{i+1} = \alpha_i^2$, for $i > 2$. We have that $\alpha_{i+1} = c^{2^{i-2}}$. In particular, for $\Delta = 3 + \log \log_{1/e} n$, we have that $\alpha_\Delta < 1/n$.

The above observation almost implies that we need $\Delta$ rows. The problem is that the above calculations (i.e., the high probability guarantee in Lemma 31.1.1) breaks down when $n_i = O(\log n)$ – that is, when $\alpha_i = O((\log n)/n)$. However, if one throws in $O(\log n)$ balls into $n$ bins, the probability of a single collision is at most $O((\log n)^2/n)$. In particular, this implies that after roughly additional $c$ rows, the probability of any ball left is $\leq 1/n^c$.

The above argumentation, done more carefully, implies the following – we omit the details because (essentially) the same analysis for a more involved case is done next (the lower bound stated follows also from the same argumentation).

Theorem 31.1.4. Consider the process of throwing $n$ balls into $n$ bins in several rounds. Here, a ball that can not be placed in a round, because their chosen bin is already occupied, are promoted to the next round. The next round throws all the rejected balls from the previous round into a new row of $n$ empty bins. This process, with high probability, ends after $M = \log \log n + \Theta(1)$ rounds (i.e., after $M$ rounds, all balls are placed in bins).

31.1.1. With only $d$ rows

Lemma 31.1.5. For $\alpha \in (0, 1/4]$, let $\gamma_1 = \alpha$, and $\gamma_i = 2\gamma_{i-1}^2$. We have that $\gamma_{d+1} \leq \alpha^{(2^d+1)/2}$.
Proof: The proof, minimal as it may be, is by induction:

\[ \gamma_{i+1} = 2\gamma_i^2 \leq 2\left(\alpha \left(\frac{(2\gamma_i+1)^2}{2}\right)\right)^2 = 2\alpha \left(\frac{(2\gamma_i+2)^2}{2}\right) \leq \alpha \left(\frac{(2\gamma_i+1)^2}{2}\right), \]

since \(2\sqrt{\alpha} \leq 1\).

\[\Box\]

Lemma 31.1.6. Let \( m = \alpha n \) balls be thrown into \( n \) bins, with \( d \) rows, where \( \alpha > 0 \). Here every bin can contain only a single ball, and if inserting the ball into \( i \)th row failed, then we throw it in the next row, and so on, till it finds an empty bin, or it is rejected because it failed on the \( d \)th row. Let \( Y(d, n, m) \) be the number of balls that did not get stored in this matrix of bins. We have

(A) For a constant \( \alpha < 1/4 \), we have \( Y(d, n, \alpha n) \leq n\alpha (2^d+1)/2 \), with high probability.

(B) We \( \mathbb{E}[Y(d, n, dn)] = O(n \log d) \).

(C) For a constant \( c > 1 \), we have \( \mathbb{E}[Y(d, n, cn \log d)] = n/e^{-d/2} \), assuming \( d \) is sufficiently large.

\[\Box\]

Proof: (A) By Lemma 31.1.1, in expectation, at least \( s_1 = n\alpha \exp(-\alpha) \) balls are placed in the first row. As such, in expectation \( n_2 = n\alpha (1 - \exp(-\alpha)) \leq n\alpha^2 \) balls get thrown into the second row. Using Chernoff inequality, we get that \( n_2 \leq 2\alpha^2 n \), with high probability. Setting \( \gamma_1 = \alpha \), and \( \gamma_i = 2\gamma_{i-1}^2 \), we get the claim via Lemma 31.1.5.

(B) As long as we throw \( \Omega(n \log d) \) balls into a row, we expect by Lemma 31.1.1 that at least \( n(1 - 1/d^{O(1)}) \) balls to be stored in this row. As such, let \( D = O(\log d) \), and observe that the first \( d - D \) rows in expectation contains \( n(d - D)(1 - 1/d^{O(1)}) \) balls. This implies that only \( O(Dn) \) are not stored in these first \( d - D \) rows, which implies the claim.

(C) Break the \( d \) rows into two groups. The first group of size \( D = O(\log d) \), and the second group is the remaining group. As long as the number of balls arriving to a row is larger than \( n \), we expected at least \( n(1 - 1/e) \) of them to be stored in this row. As such, after the first \( D \) rows, we expect the number of remaining balls to be \( \leq n \). But them. the same argumentation implies that the number of balls arriving to the \( D + i \) row, in expectation, is at most \( n/e^i \). In particular, we get that the number of balls failed to be placed is at most \( n/e^{D-d} \leq n/e^{-d/2} \).

\[\Box\]

31.2. The power of two choices

Making \( d \) choices. Let us throw \( n \) balls into \( n \) bins. For each ball, we first pick randomly \( d \geq 2 \) bins, and place the ball in the bin that currently contains the smallest number of balls (here, a bin might contain an arbitrary number of balls). If there are several bins with the minimum number of bins, we resolve it arbitrarily.

Here, we will show the surprising result that the maximum number of balls in any bin is bounded by \( O(\log \log n/\log d) \) with high probability in the end of this process. For \( d = 1 \), which is the regular balls into bins setting, we already seen that this quantity is \( \Theta(\log n/\log \log n) \), so this result is quite surprising.

31.2.1. Upper bound

Definition 31.2.1. The load of a bin is the number of balls in it. The height of a ball, is the load of the bin it was inserted into, just after it was inserted.

Some notations:
(A) $\beta_i$: An upper bound on the number of bins that have load at least $i$ by the end of the process.
(B) $h(i)$: The height of the $i$th ball.
(C) $\cup_{\geq i}(t)$: Number of bins with load at least $i$ at time $t$.
(D) $\mathfrak{B}_{\geq i}(t)$: Number of balls with height at least $i$ at time $t$.

Observation 31.2.2. $\cup_{\geq i}(t) \leq \mathfrak{B}_{\geq i}(t)$.

Let $\mathfrak{U}_{\geq i} = \cup_{\geq i}(n)$ be the number of bins, in the end of the process, that have load $\geq i$.

Observation 31.2.3. Since every bin counted in $\mathfrak{U}_{\geq i}$ contains at least $i$ balls, and there are $n$ balls, it follows that $\mathfrak{U}_{\geq i} \leq n/i$.

Lemma 31.2.4. Let $\beta_1 = n/4$, and let $\beta_{i+1} = 2n(\beta_i/n)^d$, for $i \geq 4$. Let $I$ be the last iteration, such that $\beta_I \geq 16c \ln n$, where $c > 1$ is an arbitrary constant. Then, with probability $\geq 1 - 1/n^c$, we have that

(A) $\mathfrak{U}_{\geq i} \leq \beta_i$, for $i = 4, \ldots, I$.
(B) $\mathfrak{U}_{\geq i+1} \leq c' \log n$, for some constant $c'$.
(C) For $j > 0$, and any constant $\varepsilon > 0$, we have $\mathbb{P}[\mathfrak{U}_{\geq i+1+j} > 0] \leq O(1/n^{(d-1-\varepsilon)j})$.
(D) With probability $\geq 1 - 1/n^c$, the maximum load of a bin is $I + O(c)$.

Proof: (A) Let $\mathcal{B}_i$ be the bad event that $\mathfrak{U}_{\geq i} > \beta_i$, for $i = 1, \ldots, n$. The following analysis is conditioned on none of these bad events happening. Let $\mathcal{G}$ be the good event that is the complement of $\cup_i \mathcal{B}_i$. Let $Y_i$ be an indicator variable that is one $\iff h(i) \geq i+1$ conditioned on $\mathcal{G}$ (for clarity, we omit mentioning this conditioning explicitly). We have that
\[
\tau_i = \mathbb{P}[Y_j = 1] \leq p_i \quad \text{for} \quad p_i = (\beta_i/n)^d,
\]
as all $d$ probes must hit bins of height at least $i$, and there are at most $\beta_i$ such bins. This readily implies that $\mathbb{E}[\mathfrak{U}_{\geq i+1}(n)] \leq p_in$. The variables $Y_1, \ldots, Y_n$ are not independent, but consider a variable $Y_i'$ that is 1 if $Y_i = 1$, or if $Y_i = 0$, then $Y_i'$ is 1 with probability $p - \tau_j$. Clearly, the variables $Y_1', \ldots, Y_n'$ are independent, and $\sum_i Y_i' \geq \sum_i Y_i$. For $i < I$, setting
\[
\beta_{i+1} = 2np_i = 2n(\beta_i/n)^d,
\]
we have, by Chernoff’s inequality, that
\[
\alpha_{i+1} = \mathbb{P}[\mathcal{B}_{i+1} \cap \bigcap_{k=1}^i \mathcal{B}_k] = \mathbb{P}[\mathfrak{U}_{\geq i+1}(n) > \beta_{i+1}] = \mathbb{P}[\mathfrak{U}_{\geq i+1}(n) > 2np_i] \leq \mathbb{P}\left[\sum_i Y_i' > (1 + 1)np_i\right] \\
\leq \exp(-np_i/4) = \exp(-\beta_{i+1}/8) < 1/n^{2c},
\]
and for $\beta_{i+1}$, we have $\beta_{i+1} \leq 16c \log n$. Setting $A = 2e \cdot 16c \log n$, we have
\[
\alpha_{i+1} = \mathbb{P}[\mathfrak{U}_{\geq i+1}(n) > \beta_{i+1}] \leq \mathbb{P}\left[\sum_i Y_i' > \frac{A}{\beta_{i+1}}\beta_{i+1}\right] \leq 2^{-32c \log n} \leq \frac{1}{n^c},
\]
by Lemma 31.5.4.

As for the conditioning used in the above, we have that $\mathbb{P}[\mathcal{G}] = \prod_{i=4}^{I+1} \mathbb{P}[\mathcal{B}_{i+1} \cap \bigcap_{k=1}^i \mathcal{B}_k] = \prod_i (1 - a_i) \geq 1 - 1/n^{c-1}$, since $I \leq n$. 

4
(C) Observe that $\cup_{i=1}^{2i+1}(n) \leq \cup_{i=1}^{2i}(n)$. As such, for all $j > 0$, we have that $\cup_{i=1}^{j+1}(n) \leq \cup_{i=1}^{j+1}(n) \leq \Delta = 2\varepsilon \cdot 16c \log n$, by (B). As such, we have
\[ E[\Delta_{j+1}(n)] \leq n(\Delta/n)^d = O(\log^d n/n^{d-1}) = O(1/n^{d-1-\varepsilon}) \ll 1, \]
for $\varepsilon > 0$ an arbitrary constant, and $n$ sufficient large. Using Markov’s inequality, we get that
\[ q = \mathbb{P}[\Delta_{j+1}(n) \geq 1] = O(1/n^{d-1-\varepsilon}). \]
The probability that the first $j$ such rounds fail (i.e., that $\Delta_{j+1}(n) > 0$) is at most $q^j$, as claimed.

(D) This follows immediately by picking $\varepsilon = 1/2$, and then using (C) with $j = O(c)$. □

Lemma 31.2.5. For $i = 4, \ldots, I$, we have that $\beta_i \leq n/2^{d-i+1}$.

Proof: The proof is by induction. For $i = 4$, we have $\beta_4 \leq n/4$, as claimed. Otherwise, we have
\[ \beta_{i+1} = 2n(\beta_i/n)^d \leq 2n\left(1/2^{d-i+1}\right)^d = n/2^{d-i+1} \leq n/2^{d-i+1}. \]
□

Theorem 31.2.6. When throwing $n$ balls into $n$ bins, with $d$ choices, with probability $\geq 1 - 1/n^{O(1)}$, we have that the maximum load of a bin is $O(1) + \lg \lg n/\lg d$.

Proof: By Lemma 31.2.4, with the desired probability the $\beta_i$s bound the load in the bins for $i \leq I$. By Lemma 31.2.5, it follows that for $I = O(1) + (\lg \lg n)/\lg d$, we have that $\beta_I \leq o(\log n)$. Thus giving us the desired bound. □

It is not hard to verify that our upper bounds (i.e., $\beta_i$) are not too badly off, and as such the maximum load in the worst case is (up to additive constant) the same. We state the result without proof.

Theorem 31.2.7. When throwing $n$ balls into $n$ bins, with $d$ choices (where the ball is placed with the bin with the least load), with probability $\geq 1 - o(1/n)$, we have that the maximum load of a bin is at least $\lg \lg n/\lg d - O(1)$.

31.2.2. Applications

As a direct application, we can use this approach for open hashing, where we use two hash functions, and place an element in the bucket of the hash table with fewer elements. By the above, this improves the worst case search time from $O(\log n/\log \log n)$ to $O(\log \log n)$. This comes at the cost of doubling the time it takes to do lookup on average.

31.2.3. The power of restricted $d$ choices: Always go left

The always go left rule. Consider throwing a ball into $n$ bins (which might already have some balls in them) as follows – you pick uniformly a number $X_j \in [n/d]$, and you try locations $Y_1, \ldots, Y_d$, where $Y_j = X_j + j(n/d)$, for $j = 1, \ldots, d$. Let $L_j$ be the load of bin $Y_j$, for $j = 1, \ldots, d$, and let $L = \min_j L_j$ be the minimum load of any bin. Let $\tau$ be the minimum index such that $L_j = L$. We throw the ball into $Y_\tau$.

What the above scheme does, is to partition the $n$ bins into $d$ groups, placed from left to right. We pick a bin uniformly from each group, and always throw the ball in the leftmost location that realizes the minimum load.

The following proof is informal for the sake of simplicity.
<table>
<thead>
<tr>
<th># balls in bin</th>
<th>Regular</th>
<th>2-choices</th>
<th>2-choices+go left</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>369,899,815</td>
<td>240,525,897</td>
<td>228,976,604</td>
</tr>
<tr>
<td>1</td>
<td>365,902,266</td>
<td>528,332,061</td>
<td>546,613,797</td>
</tr>
<tr>
<td>2</td>
<td>182,901,437</td>
<td>221,765,420</td>
<td>219,842,639</td>
</tr>
<tr>
<td>3</td>
<td>61,604,865</td>
<td>9,369,389</td>
<td>4,566,915</td>
</tr>
<tr>
<td>4</td>
<td>15,760,559</td>
<td>7,233</td>
<td>45</td>
</tr>
<tr>
<td>5</td>
<td>3,262,678</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>568,919</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>86,265</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>11,685</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1,347</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>143</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Theorem 31.2.8. When throwing \( n \) balls into \( n \) bins, using the always-go-left rule, with \( d \) groups of size \( n/d \), the maximum load of a bin is \( O(1) + (\log \log n)/d \), with high probability.

Proof: Lemma 31.1.6 (B) tells us that \( n(1 - O(\log d/d)) \) balls get placed as the first ball in their bin, and their height is one.

Lemma 31.1.6 (C) implies that at most \( dn/e^{-d/2} \) balls have height larger than 2.

Lemma 31.1.6 (A) implies that now we can repeat the same analysis as the power of two choices, the critical difference is that every one of the \( d \) groups, behaves like a separate height. Since there are \( O(\log \log n) \) maximum height in the regular analysis, this implies that we get \( O((\log \log n)/d) \) maximum load, with high probability.

**31.3. Avoiding terrible choices**

Interestingly, one can prove that two choices are not really necessary. Indeed, consider the variant where the \( i \)th ball randomly chooses a random location \( r_i \). The ball then is placed in the bin with least load among the bins \( r_i \) and \( r_{i-1} \) (the first ball inspects only a single bin – \( r_1 \)). It is not difficult to show that the above analysis applies in this settings, and the maximum load is \( O(\log \log n) \) – despite making only \( n \) choices for \( n \) balls. Intuitively, what is going on is that the power of two choices lies in the ability to avoid following a horrible, no good, terrible choice, by having an alternative. This alternative choice does not have to be quite of the same quality as the original choice - it can be stolen from the previous ball, etc.
31.4. Bibliographical notes

The multi-row balls into bins (Section 31.1) is from the work by Broder and Karlin [BK90]. The power of two choices (Section 31.2) is from Azar et al. [ABKU99].

The restricted $d$ choices structure, the always go-left rule, described in Section 31.2.3, is from [Vöc03].

31.5. From previous lectures

Theorem 31.5.1 (Azuma’s Inequality - Stronger Form). Let $X_0, X_1, \ldots,$ be a martingale sequence such that for each $k$, $|X_k - X_{k-1}| \leq c_k$, where $c_k$ may depend on $k$. Then, for all $t \geq 0$, and any $\lambda > 0$, we have

$$P\left[|X_t - X_0| \geq \lambda\right] \leq 2 \exp\left(-\frac{\lambda^2}{2 \sum_{k=1}^{t} c_k^2}\right).$$

Lemma 31.5.2. Let $X_1, \ldots, X_n$ be $n$ independent Bernoulli trials, where $P[X_i = 1] = p_i$, and $P[X_i = 0] = 1 - p_i$, for $i = 1, \ldots, n$. Let $X = \sum_{i=1}^{n} X_i$, and $\mu = E[X] = \sum_i p_i$. For $\delta \in (0, 1)$, we have

$$P\left[X > (1 + \delta)\mu\right] < \exp(-\mu \delta^2 / 4),$$

Lemma 31.5.3. For any positive integer $n$, we have:

(i) $(1 + 1/n)^n \leq e$.

(ii) $(1 - 1/n)^{n-1} \geq e^{-1}$.

(iii) $n! \geq (n/e)^n$.

(iv) For any $k \leq n$, we have: $\left(\frac{n}{k}\right)^k \leq \left(\frac{n}{k}\right)^k$.

Lemma 31.5.4. Let $X_1, \ldots, X_n$ be $n$ independent Bernoulli trials, where $P[X_i = 1] = p_i$, and $P[X_i = 0] = 1 - p_i$, for $i = 1, \ldots, n$. Let $X = \sum_{i=1}^{n} X_i$, and $\mu = E[X] = \sum_i p_i$. For $\delta > 2e - 1$, we have $P\left[X > (1 + \delta)\mu\right] < 2^{-\mu(1+\delta)}$.

References

