1 Power of Two Choices

Take the balls in bins problem, and modify it so that there’s two choices (instead of one choice) for each ball to go into each bin (with each ball opting for the less-full bin). This should reduce the number of collisions.

1.1 Expected number of balls in any bin

Is one, same as with only one choice. A more interesting question arises about the maximum number of balls.

1.2 Expected maximum number of balls in any given bin

Assisted with a claim: Let

\[ V_i(t) = \text{number of bins after } t \text{ balls with } \geq i \text{ balls inside} \]

\[ h_t = \text{height at which the } t^{th} \text{ ball is placed} \]

An example. Suppose \( h_i \) is the new height of the bin that the \( i^{th} \) ball lands in, and suppose \( t = 7 \).

\[ h = (h_1, h_2, \ldots, h_7) = (1, 1, 1, 2, 1, 2, 3) \]

\[ V_1(t) = 4(\text{count the number of ones}) \]

\[ V_2(t) = 2 \]

\[ V_3(t) = 1 \]

\[ V_i(t) = \{\text{number total where } h_t = i\} \]

An aside - whenever “with high probability” is mentioned, it refers to \( P = 1 - n^{-c} \) for some arbitrary constant \( c \).

We have another claim - increasingly higher heights are associated with an increasingly smaller number of bins with that height. Specifically, we have (with high probability \( \forall i \geq 4, \forall t \)


We can see a base case
\[ \beta_4 = \frac{1}{4} \]
This sequence decays faster than exponentially:
\[ \beta_{i+1} = 2 \beta_i^2 \]
And with high probability, there exists an \( h = O(\log(\log(n))) \) such that the following statement holds.
\[ V_h(t) < \frac{O(\log(n))}{n} \]
We prove this by inducting on \( i \).
We choose a base case of \( i = 4 \) because one and two are too small to make the doubling every step irrelevant (relative to the exponential), and three could be inconvenient. The proof for the base case is trivial (\( n \) balls can lead to max \( \frac{n}{4} \) bins of height \( \geq 4 \)).
For now, let’s suppose something a bit more strong than the inductive hypothesis: suppose that the hypothesis is deterministically true for tree with height \( i \).
However, we can’t suppose that it always happens - it may fail at a higher step, and if that happens, it’ll steadily get worse. So, at each height \( i \), \( Y_t = 1 \) if \( (h_t = i + 1) \cap (v_t(t - 1) \leq \beta_i n) \).
This implies \( \mathbb{E}[y_t] \leq \beta_i^2 n \leq \frac{1}{2} \beta_{i+1} n \).
The next step would be to easily analyze \( \Pr[\sum y_t \geq \beta_{i+1} n] \), but we can’t because \( y_t \) aren’t independent. However, there’s a way out of this! If we’re able to show that no matter what happens over the preceding steps \( 1, ..., t - 1 \), that \( \mathbb{E}[y_t|this] \leq \frac{1}{2} \beta_{i+1} n \), we could say that \( \exists \) set of independent random variables \( Z = \{Z_1, ..., Z_t\} \) that stochastically dominate \( y \).
Here, we say that \( Z \) stochastically dominates \( y \) if \( \exists \) correspondents distribution over \( Z \times y \) with marginals \( Z, y \) such that \( Z_i \geq y_i \) always. Then,
\[
\Pr[\sum y_t \geq \beta_{i+1} n] \leq \Pr[\sum Z_k \geq \beta_{i+1} n] \\
\leq e^{-\frac{1}{2} \beta_{i+1} n}
\]
Which means that the ys are unlikely to collide, which we cared about because the ys are condition on the previous value. So now,
\[ V_{i+1}(t) \geq \beta_{i+1} n \text{ vs. } \sum y_t \geq \beta_{i+1} n \]
Let \( Q_i = \text{event } V_{i+1}(t) \geq \beta_{i+1} n \)
\[ \Pr[Q_i] \leq \Pr[Q_{i+1} \cap \overline{Q}_i] + \Pr[Q_i] \]
\[ \Pr[\sum y_t \geq \beta_{i+1}n] \leq \Pr[Q_i] + \frac{1}{n^c} \leq \Pr[Q_{i-1}] + \frac{2}{n^c} \leq \ldots \]

This shows that the chance that any given level is bad is small, which proves our claim: we don’t get too many bins with too many balls.

So, with high probability \( \exists h \) of \( O(\log(\log(n))) \) such that \( V_{h^*}(t) < \frac{O(\log(n))}{n} \).

Then, we have to determine \( \Pr[\text{any ball gets height } h^* + 1] \). Given ball chance \( \leq O\left(\left(\frac{\log(n)}{n}\right)^2\right) \)

\[ \Pr[\text{any ball has both of the places it lands in to be } h] \leq \frac{O(\log^2(n))}{n} \]

\[ \Pr[\sum y_t \geq c] \leq \left(\frac{n}{c}\right)\left(\frac{\log n}{n}\right)^{2c} \leq \left(\frac{n}{c} \cdot \frac{\log^2 n}{n^2}\right) \leq \frac{1}{n^{c^2}} \]

Which means the max height is \( \log(\log(n)) + O(1) \) with high probability.

So, what’s \( E(\max \text{ height}) \)?

\[ \leq h^* + c + \Pr[\max h \geq h^* + c] \cdot \max \text{ possible } h \]

\[ \leq h^* + c + \frac{n}{n^{2c}} \]

\[ \leq h^* + O(1) \]