CS 388R: Randomized Algorithms, Fall 2019 September 17

Lecture 6: Power of Two Choices

Prof. Eric Price *Scribe: Mofei Zhang, Jack Youstra*

NOTE: THESE NOTES HAVE NOT BEEN EDITED OR CHEKED FOR CORECTNES

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

$$V_i(t) = \text{number of bins after } t \text{ balls with more than } i \text{ balls inside}$$

$$h_t = \text{height at which the } t_{th} \text{ ball is placed}$$

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

$$\begin{aligned} h = (h_1, h_2, ..., h_7) &= (1, 1, 1, 2, 1, 2, 3) \\ V_1(t) &= 4 && \text{(count the number of ones)} \\ V_2(t) &= 2 \\ V_3(t) &= 1 \\ V_i(t) &= \text{number total where } h_t = i \end{aligned}$$

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) $\forall i \geq 4,\ 1 \leq t \leq n$

$$V_i(t) \leq \beta_i * n$$

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 ≥ 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, define $Y_t = 1$ if $h_t = i + 1$ and $V_i(t-1) \leq \beta_i n$. $Y_t = 0$ otherwise.

This implies that

$$\mathbb{E}[Y_t] = \Pr[Y_t = 1] \leq \Pr[h_t \geq i + 1 \mid V_i(t-1) \leq \beta_i n] \leq \left(\frac{\beta_i n}{n}\right)^2 = \beta_i^2.$$

Thus $\mathbb{E}[\sum_t Y_t] \leq \beta_i^2 n = \frac{1}{2}\beta_{i+1} n$.

The next step would be to analyze $\Pr\left[\sum Y_t \geq \beta_{i+1} n\right]$, but 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 $\mathbb{E}[Y_t \mid Y_1, \ldots, Y_{t-1}] \leq \beta_i^2$, we could say that there exists set of independent random variables $Z = \{Z_1, ..., Z_t\}$ that **stochastically dominate** y. We say that Z stochastically dominates Y if there exists correspondents distribution over $Z \times Y$ with marginals Z, Y such that $Z_i \geq Y_i$ always. This is true since if $V_i(t-1) > \beta_i n$ then all $Y_t, Y_{t+1}, \ldots, Y_n$ would be equal to 0, otherwise $\Pr[Y_t = 1] \leq \beta_i^2$. Thus we have

$$\begin{aligned}\Pr\left[\sum Y_t \geq \beta_{i+1} n\right] &\leq \Pr\left[\sum Z_t \geq \beta_{i+1} n\right] \\ &\leq e^{-\frac{1}{6}\beta_{i+1} n}\end{aligned}$$

where the last inequality comes from Bernstein's inequality applied to independent variables Z_t. Hence the Y_t's are unlikely to collide, which we cared about because the Y_t's are condition on the previous value. For large enough β_{i+1} we can assume that

$$\Pr\left[\sum Y_t \geq \beta_{i+1} n\right] < \frac{1}{n^c}$$

for fixed constant $c > 0$. Precisely speaking, β_{i+1} needs to be $\Omega(\log n/n)$ with appropriate constant.

Let $Q_i = \mathbb{I}[V_{i+1}(t) \geq \beta_{i+1} n]$, then for $\beta_{i+1} = \Omega(\log n/n)$ we have that

$$\begin{aligned}
\Pr[Q_{i+1}] &\leq \Pr\left[Q_{i+1} \wedge \overline{Q_i}\right] + \Pr[Q_i] \\
&\leq \Pr\left[\sum Y_t \geq \beta_{i+1} n\right] + \Pr[Q_i] \\
&\leq \Pr[Q_i] + \frac{1}{n^c}.
\end{aligned}$$

Letting h^* be the largest i such that $\beta_i = \Omega(\log n/n)$, by union bound with probability at least $1 - n^{1-c}$, the condition $V_i(t) \leq \beta_i n$ holds for $i = 1, \ldots, h^*$.

Till now, we proved that with high probability there exists h^* of $O(\log \log n)$ such that

$$V_{h^*}(t) < O\left(\frac{\log n}{n}\right). \tag{1}$$

Then, we have to determine Pr[any ball gets height larger than $h^* + 1$]. Observe that given (**??**) the probability of ball put higher than h^* would be bounded by $O\big(\big(\frac{\log(n)}{n}\big)^2\big)$. Using this result we have

$$\Pr\left[\sum Y_t \geq c\right] \leq \binom{n}{c}\left(\frac{\log n}{n}\right)^{2c} \leq \left(\frac{ne}{c} \cdot \frac{\log^2 n}{n^2}\right)^c < \frac{1}{n^{c/2}}$$

for large enough n.

Hence the max height is $\log(\log(n)) + O(1)$ with high probability. □

What's $\mathbb{E}$[max height]?

$$\begin{aligned}
\mathbb{E}[\text{max height}] &\leq h^* + c + \Pr[\max h \geq h^* + c] \cdot (\text{max possible } h) \\
&\leq h^* + c + \frac{n}{n^{c/2}} \\
&\leq h^* + O(1)
\end{aligned}$$