Power of two choices

Huy L. Nguyễn

1 Power of two choices

Now we consider a radical new idea to improve the load: use two random bins instead of one. When ball i arrives, we consider two random bins and assign i to the least loaded bin. How does the maximum load change?

First let's try to have a hand-wavy argument on the load. Let B_i denote the number of bins with *at least* i balls. Let $B_{i,j}$ denote the number of bins with *at least* i balls after j balls are assigned. Notice that $0 = B_{i,0} \le B_{i,1} \le B_{i,2} \le \cdots \le B_{i,n} = B_i$.

Suppose we stack the balls in each bin. Let $Height(i)$ be the height of ball i in its bin. For example, if a bin has 3 balls $5, 2, 7$, then $Height(5) = 1, Height(2) = 2, Height(7) = 3$.

Suppose we have already assigned $j-1$ balls. In order for ball j to have height at least $i+1$, both of its assigned bins $h(j), g(j)$ must have load at least i. Therefore,

$$\Pr[Height(j) \ge i+1] \le \left(\frac{B_{i,j-1}}{n}\right)^2$$

$$B_{i+1} \le \text{Number of balls with height at least } i+1 \le n \cdot \left(\frac{B_i}{n}\right)^2$$

As the base case, since we have n balls, the number of bins with at least 2 balls is bounded by $n/2$. Thus, $\frac{B_2}{n} \le \frac{1}{2}$. Using our formula above, we have

$$\frac{B_3}{n} \le \frac{1}{2^2}$$
$$\frac{B_4}{n} \le \frac{1}{(2^2)^2}$$
$$\cdots$$

In general, $\frac{B_i}{n} \le \frac{1}{2^{2^{i-2}}}$. Thus, for $i > 2 + \log_2 \log_2 n$, we have $\frac{B_i}{n} < 1$. In other words, we do not expect to have bins with more than $\Theta(\log\log n)$ balls. Note that this is an exponential improvement over our previous load of $\Theta(\log n/(\log\log n))$.

Next, let's try to have a formal argument. We will use a strategy similar to above, and bound the number of bins with at least i balls for $i = 6, 7, 8, \ldots$ using induction. Let $\beta_6 = \frac{n}{2e} > \frac{n}{6}$ and $\beta_{i+1} = e \cdot \frac{\beta_i^2}{n}$. Notice that β_i decays at a similar speed compared with our hand-wavy bounds above i.e. $\beta_{6+i} = \frac{1}{2^{2^i} e}$.

Let E_i be the event that $B_i \le \beta_i$. Notice that the event E_6 ($B_6 \le \beta_6$) always holds. We will show that for $i = O(\log\log n)$, E_i holds with probability at least $1 - \frac{2i}{n^2}$.

Let Y_j be the indicator random variable that is 1 if $Height(j) \ge i+1$ *and* after $j-1$ balls are assigned, $B_{i,j-1} \le \beta_i$. We have

$$\Pr[Y_j = 1|\text{choices for balls } 1, 2, \ldots, j-1] \leq \left(\frac{B_{i,j-1}}{n}\right)^2 \leq \frac{\beta_i^2}{n^2}$$

The sum of Y_j's is bounded by the sum of n Bernoulli random variables X_j's, each of them is 1 with probability $\frac{\beta_i^2}{n^2}$. The expectation of this sum is $\frac{\beta_i^2}{n}$. By the Chernoff bound, the probability that it is more than $\frac{e\beta_i^2}{n}$ is at most $\exp(-\beta_i^2/n) \leq \frac{1}{n^2}$ as long as $\beta_i^2/n \geq 2\ln n$.

Thus,

$$\begin{aligned}\Pr[\neg E_{i+1}|E_i] &= \Pr[\sum_j Y_j \geq \beta_{i+1}|E_i] \\ &\leq \Pr[\sum_j X_j \geq \beta_{i+1}|E_i] \\ &= \frac{\Pr[\sum_j X_j \geq \beta_{i+1}]}{\Pr[E_i]} \\ &\leq \frac{1}{n^2 \Pr[E_i]}\end{aligned}$$

Note that $\Pr[E_i]$ is close to 1 so $\Pr[\neg E_{i+1}|E_i]$ is at most $\frac{2}{n^2}$. Thus,

$$\Pr[E_{i+1}] \geq \left(1 - \frac{2}{n^2}\right)\Pr[E_i] \geq \Pr[E_i] - \frac{2}{n^2}$$

This concludes the induction.

Let i^* be the minimum number of balls such that $\beta_{i^*}^2 \leq 2n\ln n$. Note that $i^* \leq \log\log n + 6$. Let $B(n,p)$ denote a random variable distributed according to the binomial distribution with parameters n and p. By a similar argument as above,

$$\Pr[\text{at least } 6\log n \text{ balls at height } i^*+1|E_i] \leq \frac{\Pr[B(n,(2\ln n)/n) \geq 6\log n]}{\Pr[E_i]} \leq \frac{1}{n^2\Pr[E_i]}$$

Thus, with high probability, there are at most $6\log n$ balls at height i^*+1. We now have

$$\begin{aligned}\Pr[\text{at least 1 balls at height } i^*+2|\text{at most } 6\log n \text{ balls at height } i^*+1] &\leq \frac{\Pr[B(n,(6\log n)^2/n^2) \geq 1]}{\Pr[E_i]} \\ &\leq \frac{36\log^2(n)}{n\Pr[E_i]}\end{aligned}$$

Therefore, with high probability, there is no ball at height i^*+2.

The next question is, what happens if we use 3 random bins?