CHAPTER FOURTEEN*

Balanced Allocations

In this chapter we examine a simple yet powerful variant of the classic balls-and-bins paradigm, with applications to hashing and dynamic resource allocation.

14.1. The Power of Two Choices

Suppose that we sequentially place n balls into n bins by putting each ball into a bin chosen independently and uniformly at random. We studied this classic balls-and-bins problem in Chapter 5. There we showed that, at the end of the process, the most balls in any bin – the maximum load – is $\Theta(\ln n / \ln \ln n)$ with high probability.

In a variant of the process, each ball comes with d possible destination bins, each chosen independently and uniformly at random, and is placed in the least full bin among the d possible locations at the time of the placement. The original balls-and-bins process corresponds to the case where $d = 1$. Surprisingly, even when $d = 2$, the behavior is completely different: when the process terminates, the maximum load is $\ln \ln n / \ln 2 + O(1)$ with high probability. Thus, an apparently minor change in the random allocation process results in an exponential decrease in the maximum load. We may then ask what happens if each ball has three choices; perhaps the resulting load is then $O(\ln \ln \ln n)$. We shall consider the general case of d choices per ball and show that, when $d \geq 2$, with high probability the maximum load is $\ln \ln n / \ln d + \Theta(1)$. Although having more than two choices does reduce the maximum load, the reduction changes it by only a constant factor, so it remains $\Theta(\ln \ln n)$.

14.1.1. *The Upper Bound*

Theorem 14.1: *Suppose that n balls are sequentially placed into n bins in the following manner. For each ball, $d \geq 2$ bins are chosen independently and uniformly at random (with replacement). Each ball is placed in the least full of the d bins at the time of the placement, with ties broken randomly. After all the balls are placed, the maximum load of any bin is at most* $\ln \ln n / \ln d + O(1)$ *with probability* $1 - o(1/n)$.

The proof is rather technical, so before beginning we informally sketch the main points. In order to bound the maximum load, we need to approximately bound the number of bins with i balls for all values of i. In fact, for any given i, instead of trying to bound the number of bins with load *exactly* i, it will be easier to bound the number of bins with load *at least* i. The argument proceeds via what is, for the most part, a straightforward induction. We wish to find a sequence of values β_i such that the number of bins with load at least i is bounded above by β_i with high probability.

Suppose that we knew that, over the entire course of the process, the number of bins with load at least i was bounded above by β_i. Let us consider how we would determine an appropriate inductive bound for β_{i+1} that holds with high probability. Define the *height* of a ball to be one more than the number of balls already in the bin in which the ball is placed. That is, if we think of balls as being stacked in the bin by order of arrival, the height of a ball is its position in the stack. The number of balls of height at least $i+1$ gives an upper bound for the number of bins with at least $i+1$ balls.

A ball will have height at least $i+1$ only if each of its d choices for a bin has load at least i. If there are indeed at most β_i bins with load at least i at all times, then the probability that each choice yields a bin with load at least i is at most β_i/n. Therefore, the probability that a ball has height at least $i+1$ is at most $(\beta_i/n)^d$. We can use a Chernoff bound to conclude that, with high probability, the number of balls of height at least $i+1$ will be at most $2n(\beta_i/n)^d$. That is, if everything works as sketched, then

$$\frac{\beta_{i+1}}{n} \le 2\left(\frac{\beta_i}{n}\right)^d.$$

We examine this recursion carefully in the analysis and show that β_j becomes $O(\ln n)$ when $j = \ln\ln n/\ln d + O(1)$. At this point, we must be a bit more careful in our analysis because Chernoff bounds will no longer be sufficiently useful, but the result is easy to finish from there.

The proof is technically challenging primarily because one must handle the conditioning appropriately. In bounding β_{i+1}, we assumed that we had a bound on β_i. This assumption must be treated as a conditioning in the formal argument, which requires some care.

We shall use the following notation: the state at time t refers to the state of the system immediately after the tth ball is placed. The variable $h(t)$ denotes the height of the tth ball, and $\nu_i(t)$ and $\mu_i(t)$ refer (respectively) to the number of bins with load at least i and the number of balls with height at least i at time t. We use ν_i and μ_i for $\nu_i(n)$ and $\mu_i(n)$ when the meaning is clear. An obvious but important fact, of which we make frequent use in the proof, is that $\nu_i(t) \le \mu_i(t)$, since every bin with load at least i must contain at least one ball with height at least i.

Before beginning, we make note of two simple lemmas. First, we utilize a specific Chernoff bound for binomial random variables, easily derived from Eqn. (4.2) by letting $\delta = 1$.

Lemma 14.2:

$$\Pr(B(n,p) \ge 2np) \le e^{-np/3}. \tag{14.1}$$

The following lemma will help us cope with dependent random variables in the main proof.

Lemma 14.3: *Let $X_1, X_2, \ldots, X_n$ be a sequence of random variables in an arbitrary domain, and let $Y_1, Y_2, \ldots, Y_n$ be a sequence of binary random variables with the property that $Y_i = Y_i(X_1, \ldots, X_i)$. If*

$$\Pr(Y_i = 1 \mid X_1, \ldots, X_{i-1}) \le p,$$

then

$$\Pr\left(\sum_{i=1}^{n} Y_i > k\right) \le \Pr(B(n, p) > k).$$

Proof: If we consider the Y_i one at a time, then each Y_i is less likely to take on the value 1 than an independent Bernoulli trial with success probability p, regardless of the values of the X_i. The result then follows by a simple induction. ■

We now begin the main proof.

Proof of Theorem 14.1: Following the earlier sketch, we shall construct values β_i such that, with high probability, $\nu_i(n) \le \beta_i$ for all i. Let $\beta_4 = n/4$, and let $\beta_{i+1} = 2\beta_i^d/n^{d-1}$ for $4 \le i < i^*$, where i^* is to be determined. We let $\mathcal{E}_i$ be the event that $\nu_i(n) \le \beta_i$. Note that $\mathcal{E}_4$ holds with probability 1; there cannot be more than $n/4$ bins with at least 4 balls when there are only n balls. We now show that, with high probability, if $\mathcal{E}_i$ holds then $\mathcal{E}_{i+1}$ holds for $4 \le i < i^*$.

Fix a value of i in the given range. Let Y_t be a binary random variable such that

$$Y_t = 1 \text{ if and only if } h(t) \ge i + 1 \text{ and } \nu_i(t-1) \le \beta_i.$$

That is, Y_t is 1 if the height of the tth ball is at least $i + 1$ and if, at time $t - 1$, there are at most β_i bins with load at least i. The requirement that Y_t be 1 only if there are at most β_i bins with load at least i may seem a bit odd; however, it makes handling the conditioning much easier.

Specifically, let ω_j represent the bins selected by the jth ball. Then

$$\Pr(Y_t = 1 \mid \omega_1, \ldots, \omega_{t-1}) \le \frac{\beta_i^d}{n^d}.$$

That is, given the choices made by the first $t - 1$ balls, the probability that Y_t is 1 is bounded by $(\beta_i/n)^d$. This is because, in order for Y_t to be 1, there must be at most β_i bins with load at least i; and when this condition holds, the d choices of bins for the tth ball all have load at least i with probability $(\beta_i/n)^d$. If we did not force Y_t to be 0 if there are more than β_i bins with load at least i, then we would not be able to bound this conditional probability in this way.

Let $p_i = \beta_i^d/n^d$. Then, from Lemma 14.3, we can conclude that

$$\Pr\left(\sum_{t=1}^{n} Y_t > k\right) \le \Pr(B(n, p_i) > k).$$

This holds independently of any of the events $\mathcal{E}_i$, owing to our careful definition of Y_t. (Had we not included the condition that $Y_t = 1$ only if $\nu_i(t-1) \le \beta_i$, the inequality would not necessarily hold.)

Conditioned on $\mathcal{E}_i$, we have $\sum_{t=1}^{n} Y_t = \mu_{i+1}$. Since $\nu_{i+1} \le \mu_{i+1}$, we have

$$\begin{aligned}
\Pr(\nu_{i+1} > k \mid \mathcal{E}_i) &\le \Pr(\mu_{i+1} > k \mid \mathcal{E}_i) \\
&= \Pr\left(\sum_{t=1}^{n} Y_t > k \mid \mathcal{E}_i\right) \\
&\le \frac{\Pr\left(\sum_{t=1}^{n} Y_t > k\right)}{\Pr(\mathcal{E}_i)} \\
&\le \frac{\Pr(B(n, p_i) > k)}{\Pr(\mathcal{E}_i)}.
\end{aligned}$$

We bound the tail of the binomial distribution by using the Chernoff bound of Lemma 14.2. Letting $k = \beta_{i+1} = 2np_i$ in the previous equations yields

$$\Pr(\nu_{i+1} > \beta_{i+1} \mid \mathcal{E}_i) \le \frac{\Pr(B(n, p_i) > 2np_i)}{\Pr(\mathcal{E}_i)} \le \frac{1}{\mathrm{e}^{p_i n/3} \Pr(\mathcal{E}_i)},$$

which gives

$$\Pr(\neg\mathcal{E}_{i+1} \mid \mathcal{E}_i) \le \frac{1}{n^2 \Pr(\mathcal{E}_i)} \tag{14.2}$$

whenever $p_i n \ge 6 \ln n$.

We now remove the conditioning by using the fact that

$$\begin{aligned}
\Pr(\neg\mathcal{E}_{i+1}) &= \Pr(\neg\mathcal{E}_{i+1} \mid \mathcal{E}_i) \Pr(\mathcal{E}_i) + \Pr(\neg\mathcal{E}_{i+1} \mid \neg\mathcal{E}_i) \Pr(\neg\mathcal{E}_i) \\
&\le \Pr(\neg\mathcal{E}_{i+1} \mid \mathcal{E}_i) \Pr(\mathcal{E}_i) + \Pr(\neg\mathcal{E}_i);
\end{aligned} \tag{14.3}$$

then, by Eqns. (14.2) and (14.3),

$$\Pr(\neg\mathcal{E}_{i+1}) \le \Pr(\neg\mathcal{E}_i) + \frac{1}{n^2} \tag{14.4}$$

as long as $p_i n \ge 6 \ln n$.

Hence, whenever $p_i n \ge 6 \ln n$ and $\mathcal{E}_i$ holds with high probability, then so does $\mathcal{E}_{i+1}$. To conclude we need two more steps. First, we need to show that $p_i n < 6 \ln n$ when i is approximately $\ln \ln n / \ln d$, since this is our desired bound on the maximum load. Second, we must carefully handle the case where $p_i n < 6 \ln n$ separately, since the Chernoff bound is no longer strong enough to give appropriate bounds once p_i is this small.

Let i^* be the smallest value of i such that $p_i = \beta_i^d / n^d < 6 \ln n / n$. We show that i^* is $\ln \ln n / \ln d + O(1)$. To do this, we prove inductively the bound

$$\beta_{i+4} = \frac{n}{2^{2d^i - \sum_{j=0}^{i-1} d^j}}.$$

This holds true when $i = 0$, and the induction argument follows:

$$\begin{aligned}\beta_{(i+1)+4} &= \frac{2\beta_{i+4}^d}{n^{d-1}}\\ &= \frac{2\left(\dfrac{n}{2^{2d^i-\sum_{j=0}^{i-1}d^j}}\right)^d}{n^{d-1}}\\ &= \frac{n}{2^{2d^{i+1}-\sum_{j=0}^{i}d^j}}.\end{aligned}$$

The first line is the definition of β_i; the second follows from the induction hypothesis. It follows that $\beta_{i+4} \le n/2^{d^i}$ and hence that i^* is $\ln\ln n/\ln d + O(1)$. By inductively applying Eqn. (14.4), we find that

$$\Pr(\neg\mathcal{E}_{i^*}) \le \frac{i^*}{n^2}.$$

We now handle the case where $p_i n < 6\ln n$. We have

$$\begin{aligned}\Pr(\nu_{i^*+1} > 18\ln n \mid \mathcal{E}_{i^*}) &\le \Pr(\mu_{i^*+1} > 18\ln n \mid \mathcal{E}_{i^*})\\ &\le \frac{\Pr(B(n, 6\ln n/n) \ge 18\ln n)}{\Pr(\mathcal{E}_{i^*})}\\ &\le \frac{1}{n^2\Pr(\mathcal{E}_{i^*})},\end{aligned}$$

where the last inequality again follows from the Chernoff bound. Removing the conditioning as before then yields

$$\Pr(\nu_{i^*+1} > 18\ln n) \le \Pr(\neg\mathcal{E}_{i^*}) + \frac{1}{n^2} \le \frac{i^*+1}{n^2}. \tag{14.5}$$

To wrap up, we note that

$$\Pr(\nu_{i^*+3} \ge 1) \le \Pr(\mu_{i^*+3} \ge 1) \le \Pr(\mu_{i^*+2} \ge 2)$$

and bound the latter quantity as follows:

$$\Pr(\mu_{i^*+2} \ge 2 \mid \nu_{i^*+1} \le 18\ln n) \le \frac{\Pr(B(n, (18\ln n/n)^d) \ge 2)}{\Pr(\nu_{i^*+1} \le 18\ln n)} \le \frac{\binom{n}{2}(18\ln n/n)^{2d}}{\Pr(\nu_{i^*+1} \le 18\ln n)}.$$

Here the last inequality comes from applying the crude union bound; there are $\binom{n}{2}$ ways of choosing two balls, and for each pair the probability that both balls have height at least i^*+2 is $(18\ln n/n)^{2d}$.

Removing the conditioning as before and then using Eqn. (14.5) yields

$$\begin{aligned}\Pr(\nu_{i^*+3} \ge 1) &\le \Pr(\mu_{i^*+2} \ge 2)\\ &\le \Pr(\mu_{i^*+2} \ge 2 \mid \nu_{i^*+1} \le 18\ln n)\Pr(\nu_{i^*+1} \le 18\ln n)\\ &\quad + \Pr(\nu_{i^*+1} > 18\ln n)\\ &\le \frac{(18\ln n)^{2d}}{n^{2d-2}} + \frac{i^*+1}{n^2},\end{aligned}$$

showing that $\Pr(\nu_{i^*+3} \geq 1)$ is $o(1/n)$ for $d \geq 2$ and hence that the probability the maximum bin load is more than $i^* + 3 = \ln \ln n / \ln d + O(1)$ is $o(1/n)$. ■

Breaking ties randomly is convenient for the proof, but in practice any natural tie-breaking scheme will suffice. For example, in Exercise 14.1 we show that if the bins are numbered from 1 to n then breaking ties in favor of the smaller-numbered bin is sufficient.

As an interesting variation, suppose that we split the n bins into two groups of equal size. Think of half of the bins as being on the left and the other half on the right. Each ball now chooses one bin independently and uniformly at random from each half. Again, each ball is placed in the least loaded of the two bins – but now, if there is a tie, the ball is placed in the bin on the left half. Surprisingly, by splitting the bins and breaking ties in this fashion, we can obtain a slightly better bound on the maximum load: $\ln \ln n / 2 \ln\big((1 + \sqrt{5})/2\big) + O(1)$. One can generalize this approach by splitting the bins into d ordered equal-sized groups; in case of a tie for the least-loaded bin, the bin in the lowest-ranked group obtains the ball. This variation is the subject of Exercise 14.13.

14.2. Two Choices: The Lower Bound

In this section we demonstrate that the result of Theorem 14.1 is essentially tight by proving a corresponding lower bound.

Theorem 14.4: *Suppose that n balls are sequentially placed into n bins in the following manner. For each ball, $d \geq 2$ bins are chosen independently and uniformly at random (with replacement). Each ball is placed in the least full of the d bins at the time of the placement, with ties broken randomly. After all the balls are placed, the maximum load of any bin is at least* $\ln \ln n / \ln d - O(1)$ *with probability* $1 - o(1/n)$.

The proof is similar in spirit to the upper bound, but there are some key differences. As with the upper bound, we wish to find a sequence of values γ_i such that the number of bins with load at least i is bounded below by γ_i with high probability. In deriving the upper bound, we used the number of balls with height at least i as an upper bound on the number of bins with height at least i. We cannot do this in proving a lower bound, however. Instead, we find a lower bound on the number of balls with height exactly i and then use this as a lower bound on the number of bins with height at least i.

In a similar vein, for the proof of the upper bound we used that the number of bins with at least i balls at time n was at least $\nu_i(t)$ for any time $t \leq n$. This is not helpful now that we are proving a lower bound; we need a lower bound on $\nu_i(t)$, not an upper bound, to determine the probability that the tth ball has height $i + 1$. To cope with this, we determine a lower bound γ_i on the number of bins with load at least i that exist at time $n(1 - 1/2^i)$ and then bound the number of balls of height $i + 1$ that arise over the interval $(n(1 - 1/2^i), n(1 - 1/2^{i+1})]$. This guarantees that appropriate lower bounds hold when we need them in the induction, as we shall clarify in the proof.