Randomized Algorithms, Summer 2016 — Lecture 5 (5 pages)

Load Balancing and Chernoff Bounds

Instructor: Thomas Kesselheim

This week, we consider a very simple load-balancing problem. Suppose you have n machines and m jobs. You want to assign the jobs to machines such that all machines have approximately the same load. Of course, there is a solution with load at most $\lceil \frac{m}{n} \rceil$ on every machine, but that requires central coordination. Without central coordination, the easiest thing you can do is let each job drawn one machine uniformly and independently.

Such random assignments are prevalent in different settings. In the general setting, one assumes that *balls* (in our case jobs) are thrown into *bins* (in our case machines) at random. We will study the balls-into-bins problems for the case that $m = n$. We are interested in the number of balls within a single bin.

Formally, let L_i be the load of bin i. By symmetry reasons, $\mathbf{E}[L_i] = 1$ for any fixed i. However, the expected maximum load $\mathbf{E}[\max_i L_i]$ is higher. Just consider the case that $n = 2$. Then, only with probability $\frac{1}{2}$ the maximum load is 1 (the balls fall into different bins), with probability $\frac{1}{2}$ it is 2 (the balls fall into the same bin).

In the first lecture, we used a union bound to upper-bound the distribution of the maximum of some random variables. To apply the union bound, we first need to understand the distribution of L_i.

Lemma 5.1 (Markov's Inequality). *Let X be a non-negative random variable. Then for every $a > 0$ we have $\mathbf{Pr}[X \geq a] \leq \frac{\mathbf{E}[X]}{a}$.*

So, Markov's inequality gives us $\mathbf{Pr}[L_i \geq a] \leq \frac{1}{a}$. This means, we could use a union bound to get $\mathbf{Pr}[\max_i L_i \geq a] \leq \sum_i \mathbf{Pr}[L_i \geq a] \leq n \cdot \frac{1}{a}$. Unfortunately, this bound is totally meaningless here. Only for $a > n$, we get a non-trivial bound on the probability. But we already know that $\max_i L_i \leq n$ because there are at most n balls overall. Therefore, we need a stronger bound than the one given by Markovs inequality.

1 Chernoff Bounds

In general, Markov's inequality is tight. However, it is particularly loose for sums of independent random variables. In these settings, we get a lot better guarantees with Chernoff bounds. There are many different variants, also for random variables that are correlated in the right sense. We provide a cheat sheet, giving you an overview many different versions. In the following, we prove the best known basic variant.

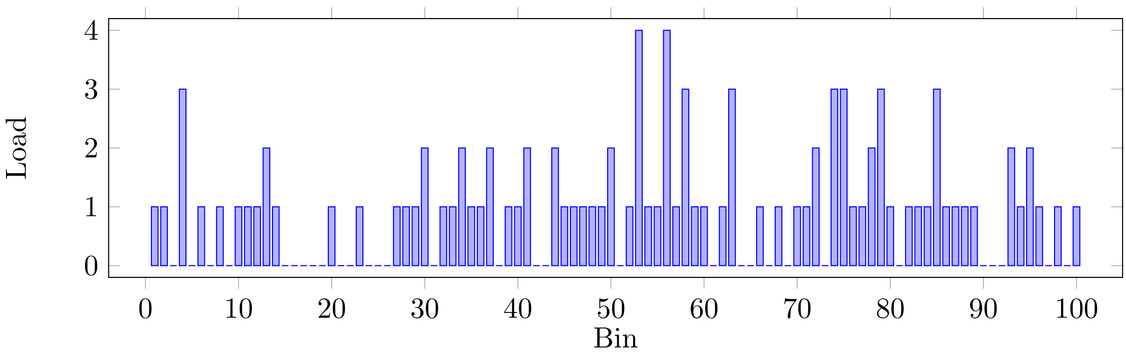

Figure 1: The loads of $n = 100$ bins.

Lemma 5.2 (Chernoff Bound). *Let $X_1, \ldots, X_n$ be independent 0/1 random variables and let X be their sum, i.e., $X = X_1 + \ldots + X_n$. For every $\mu \geq \mathbf{E}[X]$ and every $\delta > 0$, we have*

$$\mathbf{Pr}[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} \quad .$$

At this point, we can do a quick sanity check. This bound is a lot stronger. For a fixed bin, let $X_i = 1$ if ball i falls into this bin. Then we can set $\mu = 1$, $\delta = 2\log_2 n$. If n is large enough, then $\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \leq 2^{-\delta} = \frac{1}{n^2}$. So the probability that a bin gets more than $O(\log n)$ balls is at most by $\frac{1}{n^2}$ and we can apply a union bound. Later on, we will do this calculation more exactly but first, we need to show that the bound is actually correct.

Proof of Lemma 5.2. To prove Lemma 5.2, we apply Markov's inequality on the random variable e^{tX} for some $t \in \mathbb{R}$ to be defined later. Observe that no matter how we choose t, e^{tX} will always be a non-negative random variable. So, we have by Markov's inequality

$$\mathbf{Pr}[X \geq (1+\delta)\mu] = \mathbf{Pr}\left[e^{tX} \geq e^{t(1+\delta)\mu}\right] \leq \frac{\mathbf{E}\left[e^{tX}\right]}{e^{t(1+\delta)\mu}} \quad . \tag{1}$$

Next, we bound $\mathbf{E}\left[e^{tX}\right]$. We have

$$e^{tX} = e^{t\sum_{i=1}^{n} X_i} = \prod_{i=1}^{n} e^{tX_i} \quad ,$$

and therefore by independence of $(X_i)_{i\in[n]}$ also

$$\mathbf{E}\left[e^{tX}\right] = \mathbf{E}\left[\prod_{i=1}^{n} e^{tX_i}\right] = \prod_{i=1}^{n} \mathbf{E}\left[e^{tX_i}\right] \quad .$$

Every X_i is a 0/1 random variable. So let us define p_i by setting $\mathbf{Pr}[X_i = 1] = p_i$. Of course $\mathbf{Pr}[X_i = 0] = 1 - p_i$ by this definition. As we have

$$e^{tX_i} = \begin{cases} e^t & \text{if } X_i = 1 \\ 1 & \text{otherwise} \end{cases}$$

we can write the expectation of e^{tX_i} as

$$\mathbf{E}\left[e^{tX_i}\right] = p_i \cdot e^t + (1-p_i) \cdot 1 = p_i(e^t - 1) + 1 \quad .$$

So far, all steps work for any $t \in \mathbb{R}$. In the following, we set $t = \ln(1+\delta)$. In the previous equation, this gives us

$$\mathbf{E}\left[e^{tX_i}\right] = p_i(e^t - 1) + 1 = p_i(e^{\ln(1+\delta)} - 1) + 1 = p_i\delta + 1 \leq e^{p_i\delta} \quad ,$$

where the last step is true because $e^x \geq x + 1$ for all $x \in \mathbb{R}$.

Overall, this gives us for the expectation of e^{tX}

$$\mathbf{E}\left[e^{tX}\right] \leq \prod_{i=1}^{n} e^{p_i\delta} = e^{\sum_{i=1}^{n} p_i\delta} \quad .$$

Next, we use that $\mathbf{E}[X_i] = \mathbf{Pr}[X_i = 1] = p_i$. So

$$\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} \mathbf{E}[X_i] = \mathbf{E}\left[\sum_{i=1}^{n} X_i\right] = \mathbf{E}[X] \leq \mu \quad ,$$

and therefore

$$\mathbf{E}\left[e^{tX}\right] \leq e^{\mu\delta} \quad .$$

Plugging in this bound and the definition of $t = \ln(1+\delta)$ in Equation (1), we get

$$\mathbf{Pr}[X \geq (1+\delta)\mu] \leq \frac{\mathbf{E}\left[e^{tX}\right]}{e^{t(1+\delta)\mu}} \leq \frac{e^{\mu\delta}}{(1+\delta)^{(1+\delta)\mu}} = \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} \quad . \qquad \square$$

2 Upper Bound

Theorem 5.3. *The maximum bin load is* $O\left(\frac{\log n}{\log\log n}\right)$ *with high probability.*

Proof. We use Lemma 5.2 to bound the load L_j of a single bin j. The load of bin j is given as

$$L_j = X_1 + \ldots + X_n =: X \quad ,$$

where X_i is 1 if ball i falls into bin j and 0 otherwise. The X_i random variables are independent and $\mathbf{Pr}\left[X_i = 1\right] = \frac{1}{n}$.

Set $\mu = 1$, now Lemma 5.2 gives us for all $\delta > 0$

$$\mathbf{Pr}\left[X \geq 1 + \delta\right] \leq \left(\frac{\mathrm{e}^\delta}{(1+\delta)^{1+\delta}}\right)^\mu = \frac{1}{\mathrm{e}}\left(\frac{\mathrm{e}}{1+\delta}\right)^{1+\delta} \quad .$$

With $1 + \delta = \max\left\{\mathrm{e}\sqrt{\ln n}, 2\alpha\frac{\ln n}{\ln\ln n}\right\}$, we get

$$\mathbf{Pr}\left[X \geq 1 + \delta\right] \leq \frac{1}{\mathrm{e}}\left(\frac{\mathrm{e}}{\mathrm{e}\sqrt{\ln n}}\right)^{2\alpha\frac{\ln n}{\ln\ln n}} \leq \frac{1}{\mathrm{e}}\left(\frac{1}{\ln n}\right)^{\alpha\frac{\ln n}{\ln\ln n}} = \frac{1}{\mathrm{e}n^\alpha} \quad .$$

The bound on $\max_j L_j$ now follows by a union bound. ☐

3 Lower Bound

Theorem 5.4. *The maximum bin load is* $\Omega\left(\frac{\log n}{\log\log n}\right)$ *with constant probability.*

Proof. The probability that bin j gets exactly k balls is given as

$$\mathbf{Pr}\left[L_j = k\right] = \binom{n}{k}\left(\frac{1}{n}\right)^k\left(1 - \frac{1}{n}\right)^{n-k} \geq \left(\frac{n}{k}\right)^k\frac{1}{n^k}\frac{1}{\mathrm{e}} = \frac{1}{\mathrm{e}k^k}$$

For $k = \frac{\ln n}{3\ln\ln n}$ we have $k^k \leq (\ln n)^{\frac{\ln n}{3\ln\ln n}} = n^{\frac{1}{3}}$.

Let $Y_j = 1$ if bin j gets exactly k balls and 0 otherwise. We have

$$\mathbf{Pr}\left[Y_j = 1\right] \geq \mathbf{Pr}\left[L_j = k\right] \geq \frac{1}{\mathrm{e}n^{\frac{1}{3}}} \quad .$$

This tells us that $\mathbf{E}\left[\sum_{j=1}^n Y_j\right] \geq \frac{1}{\mathrm{e}}n^{2/3}$. So, in expectation, $\Omega(n^{2/3})$ bins have a load of exactly k. However, this does not tell us yet anything about the maximal load $\max_j L_j$. In principle, it can happen that the expectation is high but only because with small probability the random variable takes a very high value.

Lemma 5.5 (Chebyshev's inequality)**.** *Let X be a real-valued random variable with expectation μ and variance σ^2. Then for all $a > 0$, we have* $\mathbf{Pr}\left[|X - \mu| \geq a\right] \leq \frac{\sigma^2}{a^2}$.

We naturally have $\max_j L_j < k$ if and only if $\sum_{j=1}^n Y_j = 0$, which is equivalent to $\left|\sum_{j=1}^n Y_j - \mu\right| \geq \mu$ for every $\mu > 0$. So setting $\mu = \mathbf{E}\left[\sum_{j=1}^n Y_j\right]$, we now have

$$\mathbf{Pr}\left[\max_j L_j < k\right] = \mathbf{Pr}\left[\left|\sum_{j=1}^n Y_j - \mu\right| \geq \mu\right] \leq \frac{\sigma^2}{\mu^2} \quad ,$$

where σ^2 is the variance of $\sum_{j=1}^n Y_j$.

So, to get a bound, we need to compute the variance of $\sum_{j=1}^{n} Y_j$.

$$\sigma^2 = Var\left(\sum_{j=1}^{n} Y_j\right) = \sum_{j=1}^{n} Var(Y_j) + \sum_{j=1}^{n}\sum_{\substack{i=1\\ i\neq j}}^{n} Cov(Y_j, Y_i) \enspace .$$

The variance of a single 0/1 random variable Y_j is given as $Var(Y_j) = \mathbf{Pr}\left[Y_j = 1\right]\mathbf{Pr}\left[Y_j = 0\right] \leq \mathbf{Pr}\left[Y_j = 1\right] = \mathbf{E}\left[Y_j\right]$. The covariance for two random variables Y_j and Y_i, $j \neq i$, is defined as $Cov(Y_j, Y_i) = \mathbf{E}\left[Y_j Y_i\right] - \mathbf{E}\left[Y_j\right]\mathbf{E}\left[Y_i\right]$. The intuition is as follows. If one variable has a high value is it then more or less likely that the other one has a high value, too. Independent random variables have covariance 0. But in our case Y_j and Y_i are not independent. As one correctly expects, it is less likely that Y_j has a high value if Y_i already has, which means that the covariance is negative (or possibly zero).

To show this formally, observe that

$$\begin{aligned}
\mathbf{Pr}\left[L_j = k \mid L_i = k\right] &= \binom{n-k}{k}\left(\frac{1}{n-1}\right)^k\left(1 - \frac{1}{n-1}\right)^{n-k} \\
&\leq \binom{n}{k}\left(\frac{1}{n}\right)^k\left(1 - \frac{1}{n}\right)^{n-k} \\
&= \mathbf{Pr}\left[L_j = k\right] \enspace .
\end{aligned}$$

So, we have $\mathbf{Pr}\left[Y_j = 1 \mid Y_i = 1\right] \leq \mathbf{Pr}\left[Y_j = 1\right]$ and therefore $\mathbf{E}\left[Y_j Y_i\right] = \mathbf{Pr}\left[Y_j = 1, Y_i = 1\right] \leq \mathbf{Pr}\left[Y_j = 1\right]\mathbf{Pr}\left[Y_i = 1\right] = \mathbf{E}\left[Y_j\right]\mathbf{E}\left[Y_i\right]$. This implies $Cov(Y_j, Y_i) = \mathbf{E}\left[Y_j Y_i\right] - \mathbf{E}\left[Y_j\right]\mathbf{E}\left[Y_i\right] \leq 0$ for $j \neq i$

So, therefore $\sigma^2 \leq \sum_{j=1}^{n} \mathbf{E}\left[Y_j\right] = \mu$.

Overall, this implies

$$\mathbf{Pr}\left[\max_j L_j < k\right] \leq \frac{\sigma^2}{\mu^2} \leq \frac{1}{\mu} \leq \frac{\mathrm{e}}{n^{2/3}} \enspace .$$

□

4 An Application: Coin Tosses

We toss n coins. How likely is it that we see more than $n/2 + c\sqrt{n}$ heads?

Let X_i, $1 \leq i \leq n$ be independent 0-1 random variables with $\mathbf{Pr}\left[X_i = 1\right] = 1/2$. Let $X = \sum_i X_i$ and $\mu = n/2 = \mathbf{E}\left[X\right]$. Let $\delta = c\sqrt{n}/(n/2)$. Then

$$\begin{aligned}
\mathbf{Pr}\left[X \geq \frac{n}{2} + c\sqrt{n}\right] &= \mathbf{Pr}\left[X \geq \left(1 + \frac{2c}{\sqrt{n}}\right)\frac{n}{2}\right] \\
&\leq \left(\frac{e^{2c/\sqrt{n}}}{(1+\frac{2c}{\sqrt{n}})^{1+2c/\sqrt{n}}}\right)^{n/2} \\
&= \exp\left(\frac{n}{2}\left(\frac{2c}{\sqrt{n}} - (1 + \frac{2c}{\sqrt{n}})\ln((1 + \frac{2c}{\sqrt{n}}))\right)\right) \\
&\leq \exp\left(\frac{n}{2}\left(\frac{2c}{\sqrt{n}} - (1 + \frac{2c}{\sqrt{n}})(\frac{2c}{\sqrt{n}} - \frac{4c^2}{n})\right)\right) \\
&= \exp\left(\frac{n}{2}\left(-\frac{4c^2}{n} + (1 + \frac{2c}{\sqrt{n}})\frac{4c^2}{2n}\right)\right) \\
&= \exp\left(\frac{n}{2}\left(-\frac{4c^2}{n} + \frac{3c^2}{n}\right)\right) \\
&= \exp\left(-c^2/2\right)
\end{aligned}$$

The second inequality uses $\ln(1+x) \geq x - x^2/2$ for small positive x. The last inequality holds for $2c/\sqrt{n} \leq 1/2$.

For $c = 2$, the probability is bounded by e^{-2}, for $c = 10$, the probability is bounded by e^{-50} and for $c = \sqrt{2 \log n}$ the probability is bounded by $1/n$.