CSCI-B609: A Theorist's Toolkit, Fall 2016 Sept 27

Lecture 11: Load Balancing: Balls and Bins

Lecturer: Yuan Zhou *Scribe: Chao Tao*

1 Balls and Bins

Moving on from last lecture, this lecture will talk about load balancing. Suppose now we have n balls and n bins. For each ball, put it into an independently and uniformly randomly chosen bin. We are interested in the number of balls within the max-loaded bin. This problem has its application in load balancing. Suppose we want to assign n tasks to n servers. Prior to making the assignment, we do not know anything about the server. The goal is to ensure that every server gets even load of tasks. And an easy randomized algorithm is to assign each task uniformly at random to any server.

Theorem 1. *The max-loaded bin has* $O\left(\frac{\log n}{\log\log n}\right)$ *balls w.p.* $\geq 1 - \frac{1}{n}$.

Proof. Firstly, we can get the following fact that

$$\begin{aligned}\Pr[\text{Bin } \#i \text{ has} \geq k \text{ balls}] &\leq \binom{n}{k} \cdot \left(\frac{1}{n}\right)^k \\ &\leq \frac{n^k}{k!} \cdot \frac{1}{n^k} = \frac{1}{k!} \leq \frac{1}{k^{k/2}}.\end{aligned}$$

When $k = k^* = \frac{8\log n}{\log\log n}$, we have $k^{k/2} \geq \left(\sqrt{\log n}\right)^{4\log n/\log\log n} = 2^{2\log n} = n^2$. By a union bound, we have

$$\begin{aligned}\Pr[\forall i, \text{Bin } \#i \text{ has} < k^* \text{ balls}] &= 1 - \Pr[\exists i, \text{Bin } \#i \text{ has} \geq k^* \text{ balls}] \\ &\geq 1 - \sum_{i=1}^{n} \Pr[\text{Bin } \#i \text{ has} \geq k^* \text{ balls}] \\ &= 1 - n \cdot \frac{1}{n^2} = 1 - \frac{1}{n}\end{aligned}$$

□

In fact, we also have the following lowerbound.

Theorem 2. *The max-loaded bin has* $O\left(\frac{\log n}{\log \log n}\right)$ *balls w.p.* $\geq 1 - \frac{e^2}{n^{1/3}}$.

Proof. First, note that

$$\begin{aligned}\Pr[\text{Bin } \#i \text{ has} \geq k \text{ balls}] &\geq \binom{n}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{n-k} \\ &\geq \left(\frac{n}{k}\right)^k\left(\frac{1}{n}\right)^k\frac{1}{e} \geq \frac{1}{ek^k}.\end{aligned}$$

When $k = k^{**} = \frac{\log n}{3\log\log n}$, $ek^k \leq e(\log n)^{\log n/(3\log\log n)} = en^{1/3}$.

Let $X_i = \mathbf{1}[\text{Bin } \#i \text{ has} \geq k^{**} \text{ balls}]$. We have $\mathrm{E}[X_i] \geq \frac{1}{en^{1/3}}$. Let $X = \sum_{i=1}^n X_i$. We have $\mathrm{E}X \geq \frac{n^{2/3}}{e}$. Note this doesn't mean $\Pr[X \geq 1] \to 1$ as $n \to +\infty$. We need to explore more information about X. Recall Chebyshev's inequality:

$$\Pr[X = 0] \leq \Pr[|X - \mathrm{E}X| \geq \mathrm{E}X] \leq \frac{\mathrm{Var}[X]}{(\mathrm{E}X)^2}.$$

Here $\mathrm{Var}[X] = \sum_i \mathrm{Var}[X_i] + \sum_{i\neq j}\mathrm{Cov}[X_i, X_j]$. Note that $\mathrm{Cov}[X_i, X_j] = \mathrm{E}[(X_i - \mathrm{E}X_i)(X_j - \mathrm{E}X_j)]$ is the covariance between X_i and X_j. Since X_i and X_j are negatively correlated: some bin having more balls makes it less likely for another bin to do so, we have $\mathrm{Cov}[X_i, X_j] \leq 0$ for $i \neq j$.

For $\mathrm{Var}[X_i]$, we have $\mathrm{Var}[X_i] = \mathrm{E}(X_i - \mathrm{E}X_i)^2 \leq 1$ since $X_i \in \{0, 1\}$. Therefore $\sum_i \mathrm{Var}[X_i] \leq n$. To summarize, we can get

$$\Pr[X = 0] \leq \frac{\mathrm{Var}[X]}{(\mathrm{E}X)^2} \leq \frac{n}{n^{4/3}/e^2} = \frac{e^2}{n^{1/3}}. \tag{1}$$

By (1), we will have $\Pr[X \geq 1] \geq 1 - \frac{e^2}{n^{1/3}}$ which says the same meaning as this theorem does.

□

Remark 2.1 (Threshold phenomenon)**.** *Here,* $\tau = \Theta\left(\frac{\log n}{\log\log n}\right)$ *is a threshold. The probability that max-loaded bin having balls much below & much above* τ *are both* $o(1)$*. We can observe such threshold phenomenon in many random variables. For example, in PSET 1, we have seen*

a) $\Pr[\exists\ 4\text{-}clique \text{ in } G(n,p)] = o(1)$ *when* $p = o(n^{2/3})$,

b) and $\Pr[\nexists\ 4\text{-}clique \text{ in } G(n,p)] = o(1)$ *when* $p = \omega(n^{2/3})$.

2 The Power of Two Choices

Now let us consider a slightly different strategy: for each ball, independently randomly choose 2 bins, and add the ball to the less-loaded bin. Then, we will have the following theorem.

Theorem 3. *The max-loaded bin has $O(\log\log n)$ balls w.p. $\geq 1 - O\left(\frac{\log^2 n}{n}\right)$.*

2.1 Intuition

Let N_i be the number of bins loaded with $\geq i$ balls. A bin has height i if there are $i-1$ balls in the bin before the ball was added. And let B_i be the number of balls with height $\geq i$. Then, we can find the following two facts:

a) $N_i \leq B_i$,

b) and $N_3 \leq \frac{n}{3}$.

Note that a bin loaded with $\geq i$ balls at least has one ball with height $\geq i$. Then it is not hard to get the fact that $N_i \leq B_i$. We have known that $N_3 \leq \frac{n}{3}$. What about N_4? Intuitively, a ball with height ≥ 4 needs to choose both bins loaded with ≥ 3 balls - with chance $\left(\frac{N_3}{n}\right)^2 \leq \frac{1}{9}$. Therefore, "on expectation", $N_4 \leq B_4 \leq \frac{n}{9}$. Similarly, a ball with height ≥ 5 needs to choose two bins loaded with ≥ 4 balls - with chance $\left(\frac{N_4}{n}\right)^2 \leq \frac{1}{81}$. Therefore, we expect $N_5 \leq B_5 \leq \frac{n}{81}$. In general, we expect $N_i \leq B_i \leq n \cdot 3^{-2^{i-2}}$. When $i = \Theta(\log\log n)$, this number becomes < 1 .

2.2 Full proof

Let $\mathcal{E}_i$ be the event that $N_i \leq \beta_i n$ where $\beta_3 = 1/3$ and $\beta_{i+1} = e(\beta_i)^2$ for $i \geq 3$. We already have $\Pr[\mathcal{E}_3] = 1$. Then, we have the following claim.

Claim 3.1. *Consider the following process: as we put balls in bins, we mark bins. A ball is called "marked" if both associated bins are marked. If throughout the process we mark $\leq \alpha n$ bins, then $\Pr[\#\text{ of marked balls} > e\alpha^2 n] \leq \frac{1}{n^2}$ when $\alpha^2 n \geq 2\ln n$.*

Consider $\Pr[\neg\mathcal{E}_{i+1} \wedge \mathcal{E}_i]$ when $\beta_i^2 \geq 2\ln n/n$. Whenever a bin is loaded with $\geq i$ balls, we mark it. Then a ball is marked if and only if the ball has height $\geq i+1$. Therefore,

$$\begin{aligned}\Pr[\neg\mathcal{E}_{i+1} \wedge \mathcal{E}_i] &\leq \Pr[B_{i+1} > \beta_{i+1} n \wedge N_i \leq \beta_i n] \\ &= \Pr[\#\text{ of marked balls} \geq e\beta_i^2 n \wedge \#\text{ of marked bins} \leq \beta_i n] \\ &\leq \frac{1}{n^2} \text{ (by Claim 3.1).}\end{aligned}$$

Therefore, as long as $\beta_i^2 \geq 2\ln n/n$, we have

$$\begin{aligned}\Pr[\neg\mathcal{E}_{i+1}] &= \Pr[\neg\mathcal{E}_i] + \Pr[\neg\mathcal{E}_{i+1} \wedge \mathcal{E}_i] \\ &\leq \Pr[\neg\mathcal{E}_i] + \frac{1}{n^2} \\ &\leq i/n^2 \text{ (by induction).}\end{aligned}$$

What is the largest i^* such that $\beta_{i^*}^2 \geq \frac{2\ln n}{n}$? From the following claim, we have $i^* = O(\ln\ln n)$.

Claim 3.2. *The largest i^* such that $\beta_{i^*}^2 \geq \frac{2\ln n}{n}$ satisfies $i^* = O(\ln\ln n)$.*

Therefore,

$$\Pr[\neg\mathcal{E}_{i^*+1}] \leq O\left(\frac{\ln\ln n}{n^2}\right),$$

which means w.p. $\geq 1 - O\left(\frac{\ln\ln n}{n^2}\right)$, there are at most $\beta_{i^*+1} < \frac{2\ln n}{n}$ bins loaded with $i^* + 1 = O(\ln\ln n)$ balls. We also have the following claim.

Claim 3.3. $\Pr[B_{i^*+2} \geq 1] \leq O\left(\frac{\ln^2 n}{n}\right)$.

2.2.1 Proof of Claim 3.1

Each ball is marked w.p. $\leq \alpha^2$. Therefore expected number of marked balls $\mu \leq \alpha^2 n$. By Chernoff bound (the worst case happens when $\mu = \alpha^2 n$), we have

$$\Pr[X_1 + \ldots X_n \geq e\mu] \leq \left(\frac{e^{e-1}}{e^e}\right)^\mu = e^{-\mu} \leq \frac{1}{n^2}.$$

2.2.2 Proof of Claim 3.2

Let $\gamma_i = \ln\beta_i$. We have $\gamma_3 = -\ln 3$ and $\gamma_{i+1} = 2\gamma_i + 1$ for $i \geq 3$. Then we will have $(\gamma_{i+1} + 1) = 2(\gamma_i + 1)$. From this equation, we can further get $\gamma_i + 1 = 2^{i-3} \cdot (1 - \ln 3)$ for $i \geq 3$. Then, $\gamma_i = 2^{i-3} \cdot (1 - \ln 3) - 1$.

In order to have $\beta_i^2 \geq \frac{2\ln n}{n}$, we need $2\gamma_i \geq \ln\ln n - \ln\frac{n}{2}$. Then, we have

$$\begin{aligned}&\Rightarrow 2^{i-2}(1 - \ln 3) - 2 \geq \ln\ln n - \ln\frac{n}{2} \\ &\Rightarrow 2^{i-2} \leq \frac{\ln\ln n - \ln n + \ln 2 + 2}{1 - \ln 3} \\ &\Rightarrow i - 2 \leq \ln\left(\frac{\ln n - \ln\ln n - \ln 2 - 2}{\ln 3 - 1}\right).\end{aligned}$$

Therefore $i^* = O(\ln\ln n)$.

2.2.3 Proof of Claim 3.3

Note the fact that

$$\begin{aligned}
\Pr[B_{i^*+2} \geq 1] &\leq \Pr[\neg \mathcal{E}_{i^*+1}] + \Pr[B_{i^*+2} \geq 1 \wedge \mathcal{E}_{i^*+1}] \\
&\leq O\left(\frac{\ln \ln n}{n^2}\right) + n \cdot \beta_{i^*+1}^2 \\
&= O\left(\frac{\ln \ln n}{n^2}\right) + O\left(\frac{\ln^2 n}{n}\right) \\
&= O\left(\frac{\ln^2 n}{n}\right).
\end{aligned}$$

Therefore, we have $\Pr[N_{i^*+2} \geq 1] \leq \Pr[B_{i^*+2} \geq 1] = O\left(\frac{\ln^2 n}{n}\right)$.