CS 388R: Randomized Algorithms, Fall 2023 September 11th, 2023

Lecture 6: Power of Two Choices

Prof. Eric Price *Scribe: Alexia Atsidakou, Alekhya Kuchimanchi*

NOTE: THESE NOTES HAVE NOT BEEN EDITED OR CHECKED FOR CORRECTNESS

1 Overview

In the last lecture we discussed treaps, coupon collector, and balls and bins. We found the maximum load and the average load over balls for the balls and bins problem. We found that if we randomly throw n balls into n bins the maximum load is with high probability $O(\frac{\log n}{\log \log n})$.

In this lecture we discuss the Power of Two Choices. In the Balls and Bins problem we were only given one choice of a random bin to throw a ball into. If we are given two random bins we expect that the maximum load should be lowered, so in this lecture we aim to answer by how much the maximum load is lowered. Given that each successive ball goes in the less loaded bin out of two random bins we will show that the expected maximum load is $O(\log \log n)$ with probability at least $1 - n^{-c}$.

2 Problem Statement

We have n balls that are thrown into n bins in the following manner:

- For each successive ball we pick two bins uniformly at random.
- The ball is placed into the bin that contains the smaller number of balls (the lighter bin).
- If both bins have the same number of balls, then we break the tie arbitrarily (put the ball in either of the bins).

Maximum Load: We want to find the maximum number of balls that is placed in a bin. We define X_j = the load of bin j at the end of the process. Then, the goal is to find $E[max_j X_j]$. We further define some notation:

- $v_i(t)$:= the number of bins at height $\geq i$ after inserting t balls, $i \epsilon [n]$.
- h_t := height of the t^{th} ball inserted, $t \epsilon [n]$. Notice that $h_1 = 1$ always.

We introduce some quantities β_i, representing an upper bound for the fraction of bins with a height of at least i. The quantities β_i satisfy

- $\beta_4 = 1/4$

- $\beta_{i+1} = 2\beta_i^2$, $\forall i \geq 4$

Lemma 1. *We have that $v_i(t) \leq \beta_i n$ w.h.p $\forall t$, and for $i \geq 4$ and such that $\beta_i^2 n \geq 3c \log n$.*

We first assume that Lemma 1 holds and discuss how we can obtain a bound on the maximum load. At a given point where $v_i(t) \leq \beta_i n$, i.e. there are at most $\beta_i n$ bins at height i, the probability that a ball is placed at height at least $i+1$ is bounded as follows

$$\mathbb{P}\left(\text{ball is placed at height } \geq i+1\right) = \left(\frac{\beta_i n}{n}\right)^2 = \beta_i^2,$$

because we need to sample twice from bins with height at least i. Then, for the expected number of balls at height at least $i+1$ we have

$$\mathbb{E}\left(\#\text{ balls at height} \geq i+1\right) = \beta_i^2 n (\text{considering both bins}).$$

This implies that

$$\mathbb{E}\left(\#\text{ bins at height} \geq i+1\right) \leq \beta_i^2 n.$$

Then, with high probability, the number of bins at height at least $i+1$ is bounded by $2\beta_i^2 n$.

How small is β_i: As the layers increase, we can see that β_i decays faster. In general, by the conditions for β_i we have that $\beta_i = 2^{-(2^{i-4}+1)} \approx 2^{-2^i}$. Then, for $i = O(\log \log n)$ we have $\beta_i < \frac{1}{n}$.

This indicates that the maximum load of a bin is $O(\log \log n)$, which is significantly better than the bound of $O(\frac{\log n}{\log \log n})$ that was obtained for the Balls and Bins problem. Next, we complete the gaps in the proof sketch.

3 Formal Proof

3.1 Proof of Lemma 1

We prove the claim via induction on the height i.

Basis. The base case is when $i = 4$. In that case we trivially have $v_4(n) \leq 1/4$. This is because we have a total of n balls, thus we can have at most $n/4$ bins containing ≥ 4 balls.

Inductive Step. We define the event $Q_i = \{v_i(n) \leq \beta_i n\}$. Suppose that Q_i holds with high probability. We need to show that Q_{i+1} holds with high probability. The challenge here is that the inductive hypothesis only holds with high probability, i.e. $\mathbb{P}\left(Q_i\right) \geq 1 - n^{-c}$ for some c.

The probability of placing a ball at height at least $i+1$ at any time t given the state at $t-1$ is

$$\mathbb{P}\left(h_t \geq i+1 \mid \text{State at time } t-1\right) = \left(\frac{v_i(t-1)}{n}\right)^2,$$

since in order to place a ball at height at least $i+1$ we need to sample (both times) from bins with height at least i, which are $v_i(t-1)$ in number.

We fix some height i. We define $Y_t = \mathbb{1}\{h_t \geq i+1 \cap v_i(t-1) \leq \beta_i n\}$. Then we have that

$$\mathbb{P}(Y_t = 1) = \mathbb{P}(h_t \geq i+1 \cap v_i(t-1) \leq \beta_i n) \leq \mathbb{P}(h_t \geq i+1 \mid v_i(t-1) \leq \beta_i n) = \left(\frac{\beta_i n}{n}\right)^2 = \beta_i^2.$$

regardless of the state at time $t-1$. Moreover, we further have $\mathbb{E}\left(\sum_{t\in[n]} Y_t\right) \leq n\beta_i^2$.
The random variables Y_t are not independent. However, we can have $Y_t = 1$ only when event $v_i(t-1) \leq \beta_i n$ holds. Moreover, conditioned on the latter, the probability that Y_t is 1 is always bounded by β_i^2. Therefore, although the variables Y_t are not independent, we can use stochastic dominance in order to apply a multiplicative Chernoff bound, i.e. there exist some variables Z_t such that $\mathbb{P}(Z_t = 1 \mid Z_1, ..., Z_{t-1}) = \beta_i^2, \forall t$ and $\sum_t Y_t \leq \sum_t Z_t$. Then we have that

$$\mathbb{P}\left(\sum_{t\in[n]} Y_t \geq (1+\epsilon)\beta_i^2 n\right) \leq \mathbb{P}\left(\sum_{t\in[n]} Z_t \geq (1+\epsilon)\beta_i^2 n\right) \leq e^{-\frac{\epsilon^2}{2+\epsilon^2}\beta_i^2 n}.$$

Using $\epsilon = 1$ we obtain $\mathbb{P}\left(\sum_{t\in[n]} Y_t \geq 2\beta_i^2 n\right) \leq e^{-\frac{1}{3}\beta_i^2 n}$.

When Q_i holds, i.e. when $v_i(n) \leq \beta_i n$, we have $v_i(t) \leq \beta_i n$, $\forall t \in [n]$. In that case, for any t, we have $Y_t = 1$ if $h_t \geq i+1$. Therefore we can bound

$$\begin{aligned}\sum_{t\in[n]} Y_t &= \#\text{balls at height at least } i+1 \\ &\geq \#\text{bins with height at least } i+1 \\ &= v_{i+1}(n).\end{aligned}$$

Then, we have that

$$\begin{aligned}\mathbb{P}\left(\bar{Q}_{i+1}\right) &= \mathbb{P}\left(\bar{Q}_{i+1} \cap Q_i\right) + \mathbb{P}\left(\bar{Q}_{i+1} \cap \bar{Q}_i\right) \\ &\leq \mathbb{P}\left(v_{i+1}(n) > 2\beta_i^2 n \cap Q_i\right) + \mathbb{P}\left(\bar{Q}_i\right) \\ &\leq \mathbb{P}\left(\sum_{t\in[n]} Y_t > 2\beta_i^2 n \cap Q_i\right) + \mathbb{P}\left(\bar{Q}_i\right) \\ &\leq \mathbb{P}\left(\sum_{t\in[n]} Y_t > 2\beta_i^2 n\right) + \mathbb{P}\left(\bar{Q}_i\right).\end{aligned}$$

Both terms in the above bound are small: by the inductive hypothesis $\mathbb{P}\left(\bar{Q}_i\right) \leq n^{-c}$. Moreover, we showed that $\mathbb{P}\left(\sum_{t\in[n]} Y_t > 2\beta_i^2 n\right) \leq e^{-\frac{1}{3}\beta_i^2 n}$, which is small for a large enough β_i. For instance, if $\beta_i^2 n \geq 3c\log n$, we have $e^{-\frac{1}{3}\beta_i^2 n} \leq n^{-c}$. Therefore, we showed that the Q_{i+1} holds with high probability.

By induction, we conclude that for all $i \geq 4$ and s.t. $\beta_i^2 n \geq 3c\log n$ we have that $v_i(n) \leq \beta_i n$ with high probability.

Note: We would like to find for which heights i a condition such as $\beta_i^2 n \geq 3c \log n$ is satisfied. Using the above observation that $\beta_i \approx 2^{-2^i}$ we obtain that the condition is satisfied for $i = O(\log \log n)$.

Let h_* be the height satisfying the above condition. Next, we will show that the probability of having a larger height that h_* is small.

3.2 Bound For a Larger height

Notice that for $h_* = O(\log \log n)$ is a height such that $v_{h_*}(n) \leq O(\log n)$ with high probability. We define the event $Y_t = \{\text{ball is placed at height at least } h_* + 1 \cap v_{h_*}(n) \leq O(\log n)\}$. Then, we have that

$$\mathbb{P}\left(\max \text{ height} \geq h_* + c\right) \leq \mathbb{P}\left(\sum_{t\in[n]} Y_t \geq c\right) + \mathbb{P}\left(v_{h_*}(n) > O(\log n)\right).$$

We know that by definition of h_* we have $\mathbb{P}\left(v_{h_*}(n) > O(\log n)\right) \leq n^{-c}$. The other term can be bounded as follows: Having $\sum_{t\in[n]} Y_t \geq c$ means that c times out of n a ball was placed on a bin with height $\geq h_* + 1$. Thus,

$$\begin{aligned}
\mathbb{P}\left(\sum_{t\in[n]} Y_t \geq c\right) &\leq \binom{n}{c}\left(\frac{O(\log n)}{n}\right)^{2c} \\
&\leq \left(\frac{en}{c}\right)^c \left(\frac{O(\log n)}{n}\right)^{2c} \\
&\leq \left(\frac{e\, O(\log^2 n)}{nc}\right)^c \\
&\leq n^{-c/2}.
\end{aligned}$$

Therefore, we conclude that for all constants $c \geq 1$, there exists c' such that the maximum load is bounded by $c' \log \log n$ under two choices, with probability $1 - n^{-c}$.