Lecture 8 — Sep 23, 2015

Prof. Eric Price | *Scribe: An T. Nguyen*

1 Overview

In the last class we studied the Balls in Bins problem and proved the max load bound of $\theta(\frac{\log n}{\log\log n})$. Today, we look at the power of two choices, where for each ball, we look at two random bins and assign to the lesser loaded bin.

2 Intuition

Let v_i be the number of bins with at least i balls, it is easy to see that $v_4 \leq \frac{n}{4}$. Therefore the probability that a ball will push a bin from (at least) 4 to 5 is a most:

$$\left(\frac{v_4}{n}\right)^2 \leq \frac{1}{16}$$

since the two choices need to be two of the v_4 bins with at at least 4 balls. Then $v_5 \approx \frac{n}{16}$. By the same argument,

$$v_6 \leq n\left(\frac{v_5}{n}\right)^2 = \frac{n}{256}$$

and

$$v_k \leq \frac{n}{2^{2^{k-1}}}$$

which suggests that the max load is $\theta(\log\log n)$

3 Analysis

Notations

- $v_i(t)$ = number of bins of height $\geq$ i after t balls.
- h_t = height of the t-th ball.

Claim 1. *$\forall t \leq n, v_i(t) \leq \beta_i n$ with high probability $(1 - \frac{1}{n^c})$, where*

- $\beta_4 = \frac{1}{4}$
- $\beta_{i+1} = 2\beta_i^2$

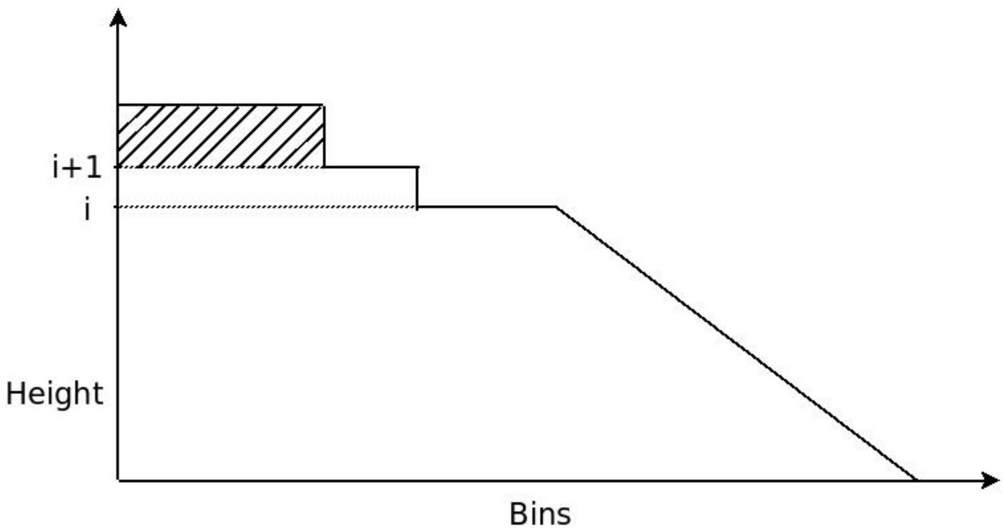

Figure 1: A histogram of the height of the bins. The shaded area is the number of tall balls: $\sum Y_t$

Proof. By induction.

Base case $i = 4$ is trivial.

Inductive Step. Let $Y_t = 1$ if

- $h_t \geq i + 1$
- $v_i(t-1) \leq \beta_i n$

(the first means that the t-th ball is on a 'big' bin, the second means not failing on previous steps). We have:

$$Pr(h_t \geq i+1) = \left(\frac{v_i(t-1)}{n}\right)^2 \tag{1}$$

$$\implies Pr(Y_t) \leq \left(\frac{v_i(t-1}{n}\right)^2 \leq \beta_i^2 \tag{2}$$

$$\implies E\left(\sum_{t=1}^{n} Y_t\right) \leq n\beta_i^2 \tag{3}$$

Let Q_i be the event that $v_i(t) \leq \beta_i n, \forall t$ then:

$$\overline{Q_{i+1}}|Q_i \iff v_{i+1}(t) > \beta_{i+1} n | Q_i \tag{4}$$

and that the number of tall balls is at least the number of tall bins (see Figure 1):

$$\sum Y_t \geq v_{i+1}(t) \tag{5}$$

therefore:

$$Pr(\overline{Q_{i+1}}|Q_i) \leq Pr(\sum_{t=1}^{n} Y_t \geq \beta_{i+1} n | Q_i) \tag{6}$$

We now need to bound $Pr(\sum Y_t \geq 2n\beta_i^2)$. We will use Bernstein type inequality. Consider n coins, each with probability p and is $(\sqrt{p}, 1)$ sub-gamma. Their sum S is $(\sqrt{np}, 1)$ sub-gamma and:

$$Pr(S \geq np + t) \leq \max[e^{t^2/(2np)}, e^{-\Omega(t)}] \tag{7}$$
$$\implies Pr(S \geq 2np) \leq e^{-\Omega(np)} \tag{8}$$

which is a good bound, except we can not (directly) use that on Y_is since they are not independent. However, observe that

$$Pr(Y_t|Y_1..Y_{t-1}) \leq \beta_i^2 \tag{9}$$

regardless of $Y_1..Y_{t-1}$. So the Y_is are *stochastically dominated* by the Z_is such that:

- $Pr(Z_i) = \beta_i^2$
- $Z_i$are independent

Y being stochastically dominated by Z.

We need to sample (Y, Z) such that $Y_i \leq Z_i$ always. Going through the ball in each stage, we want: $Pr(Y_t) = p \leq \beta_i^2 = Pr(Z_t)$

(Y_t, Z_t) is sampled from

- $(0, 0)$ with prob $1 - \beta_i^2$
- $(1, 1)$ with prob p
- $(0, 1)$ with prob $\beta_i^2 - p$

Use the Bernstein bound on Zs (instead of Ys), we have (for $\beta_{i+1} n > \log(n)$):

$$Pr(\overline{Q_{i+1}}|Q_i) \leq Pr(\sum_{t=1}^{n} Y_t \geq \beta_{i+1} n | Q_i) \tag{10}$$
$$\leq e^{-\Omega(\beta_{i+1} n)} \tag{11}$$

Let E_i be the event that $\sum Y_t \geq \beta_{i+1} n$. We have:

$$Pr(E_i) < \frac{1}{n^c} \tag{12}$$

$$Pr(\overline{Q_{i+1}} | Q_i) \leq Pr(E_i | Q_i) \tag{13}$$

and need to bound $Pr(\overline{Q_{i+1}})$. This problem will be added to Homework 2. Now, assume we have:

$$Pr(\overline{Q_{i+1}}) \leq \frac{1}{n^c} \tag{14}$$

then

$$Pr(\text{any } \overline{Q_{i+1}}) \leq \frac{1}{n^c} \tag{15}$$

to be finished next class

□

References

[1] M. Mitzenmacher, A. W. Richa, and R. Sitaraman. The power of two random choices: A survey of techniques and results. In *in Handbook of Randomized Computing*, pages 255–312. Kluwer, 2000.