CS265, Fall 2023

Class 6: Agenda and Questions

1 Announcements

- HW2 due tomorrow!
- HW3 out now!
- Add/drop deadline tomorrow.

2 Recap and Questions

The mini-lectures covered balls-and-bins and poissonization. Any questions from the minilectures/quiz?

3 The Power of Two Choices

[A bit of lecture to set things up; the summary is below. Note that the lecture notes (at the end) also discuss this material a bit if you miss class and/or want a recap after class.]

Consider the following way to throw n balls into n bins.

- When placing the t'th ball, choose *two* bins at random. (With replacement).
- Put the t'th ball in the less full of those two bins. (Breaking ties arbitrarily).

We saw in the minilecture that without this extra two-choice step, the max load would be $\Theta\left(\frac{\log n}{\log\log n}\right)$. A surprising result, which we'll explore today, is that the max load of the above scheme is $O(\log\log n)$!

3.1 Intuition for the argument

Here is some notation:

- Define $\beta_2 = n/2$. Define β_i for $i > 2$ recursively by

$$\beta_i = \frac{\beta_{i-1}^2}{n}.$$

- Define $B(i,t)$ to be the number of bins with at least i balls after step t.

Group Work

1. Explain why $B(2,t) \leq \beta_2$ for all t.
2. Show that

$$\Pr\{\text{Ball } t \text{ is the} \geq \text{3rd ball to land in its bin}\} \leq \left(\frac{B(2,t-1)}{n}\right)^2 \leq \frac{\beta_2^2}{n^2},$$

   for all t.

3. Show that, for all t,

$$\mathbb{E}[B(3,t)] \leq \beta_3.$$

   ***Hint**: Can you bound $B(3,t)$ in terms of the indicator random variables from the previous part?*

4. **Suppose** that $B(3,t) \leq \beta_3$ for all t. That is, suppose that the thing that you showed in expectation before actually held. Show that, for all t,

$$\mathbb{E}[B(4,t)] \leq \beta_4.$$

   (Note: don't worry about whether or not the **suppose** above means that you should formally condition on anything, or about how you should do that conditioning. The point of this exercise is just a back-of-the-envelope computation.)

5. **Suppose** that this logic continued, and you could show that $\mathbb{E}[B(i,t)] \leq \beta_i$ for all t. What would the max load be?

   ***Hint**: Come up with a closed form for β_i. At what point does β_i become way less than 1?*

3.2 Making the intuition (slightly) more rigorous

Of course, the above argument doesn't work since we can't just assert that the thing that holds in expectation actually holds. Here's a *suggested* way to fix it:

1. Replace β_i with:

   - $\beta_4 = n/4$
   - $\beta_i \leftarrow \frac{2\beta_{i-1}^2}{n}$ for $i > 4$.

   This extra factor of 2 in the recurrence relation will give us a bit of slack.

2. We prove that, with high probability, $B(i,n) \leq \beta_i$ for all $i \leq O(\log\log n)$ using induction on i. The base case for $i = 4$ follows from the definition of β_4, similarly to what

we had before. Inductively assume that $B(i-1,n) \le \beta_{i-1}$ with probability at least $1-(i-1)/n^2$. In the case that $B(i-1,n) \le \beta_{i-1}$, we have

$$\begin{aligned}\Pr\{B(i,n) > \beta_i\} &\le \Pr\left\{\sum_{t=1}^{n} \mathbf{1}\{\text{ ball } t \text{ is the } i\text{'th (or greater) ball to land in its bin}\} > \beta_i\right\} \\ &\le \Pr\left\{\sum_{t=1}^{n} \mathbf{1}\{\text{ ball } t \text{ is the } i\text{'th (or greater) ball to land in its bin}\} > 2\mu\right\},\end{aligned}$$

where $\mu = \beta_{i-1}^2/n$. Our earlier analysis shows that the expectation of the random variable

$$\mathbf{1}\{\text{ ball } t \text{ is the } i\text{'th (or greater) ball to land in its bin}\}$$

is at most $\beta_{i-1}^2/n = \mu$. Thus, we can apply a Chernoff bound, which says that the probability that this sum is more than twice its expectation is at most

$$\Pr[B(i,n) > \beta_i] \le \exp(-2\mu/3) = \exp(-\beta_i/3).$$

As long as $\beta_i \ge 6\log n$ (say), then this probability is at most $1/n^2$, and by a union bound with the event that $B(i-1,n) > \beta_{i-1}$ (which we assume happens with probability at most $(i-1)/n^2$), the probability that both occur and in particular that $B(i,j) \le \beta_i$ is at least $1 - i/n^2$. This establishes the inductive hypothesis for the next round.

3. The inductive argument above works up until $\beta_i = 6\log n$. If we solve for i^* so that $\beta_{i^*} = 6\log n$, we find that $i^* = \Theta(\log\log n)$. (The computation is a bit different than before because now the "1" is "$6\log n$", and the extra factor of 2 in the recurrence relation, but it's basically the same.)

4. We conclude that, with high probability, $B(i,n) \le \beta_i$ for all $i \le i^* = \Theta(\log\log n)$. Just as before, this implies that the max load is $\Theta(\log\log n)$.

Group Work

There are (at least) two or three major problems with the proof above.

1. What are the major problems?
2. If you have time, how might you fix the problems you came up with? Don't worry about trying to write a formal proof that fixes them (the formal proof is a bit tedious...), but rather try to think about, intuitively, why these problems are fixable.

It turns out that these problems *are* fixable. To see the formal proof you can check out this nice survey about different approaches to analyzing the power of two choices: https://www.eecs.harvard.edu/~michaelm/postscripts/tpds2001.pdf