

Class 6: Agenda, Questions, and Links

1 Warm-Up

Go to <http://PollEv.com/cs265> and answer the following question.

Group Work

Important: as you make progress on the question(s), one person in each room should record your progress on <http://PollEv.com/cs265>.

Suppose (completely hypothetically) that you are a professor who has no memory, and you want to allocate grading to TAs. You ask each TA to grade a problem, one problem at a time. e.g.:

- Jay, can you grade 1(a)?
- Mingda, can you grade 1(b)?
- Lunjia, can you grade 1(c)?
- Jay, can you grade 2(a)?
- Bryan, can you grade 2(b)?
-

The problem is, at any point in this process, you can't remember which TAs you've already asked, or how many problems you've already asked them to grade!

Suppose that there are n problems and n TAs.

Devise a randomized scheme to assign grading, that doesn't require any memory (other than which problem you are on) so that no TA ends up with more than $O\left(\frac{\log n}{\log \log n}\right)$ problems to grade.^a

Go to <http://PollEv.com/cs265> to describe your algorithm, or up-vote others' algorithms.

Once you've done that, think about ways that you could do better, while still not using (much) memory. Here, "better" means a more even work distribution for the TAs. (It's okay to ask a TA how many problems they are already grading, but you can't remember that information for long).

^aAssume that all problems take the same amount of time to grade, so this is some proxy for fairness.

Group Work: Solutions

Give each problem to a random TA each time. This is like dropping n balls (problems) into n bins (TAs).

2 Announcements

- HW2 due Friday!
- Please fill out the feedback form here if you haven't already: <https://forms.gle/kd7CREPav5UbsKZb9>
- Add/Drop deadline is Friday. (We hope you all stay, but just a reminder).

3 Questions?

Any questions from the minilectures or warmup? (Moment generating functions; Chernoff bounds)

- Go into small groups and ask each other your questions.
- Ask any questions that the group can't resolve in the chat, or DM one of the course staff and we'll come to your room.

4 The Power of two choices

[Slides with setup; the summary is below]

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 $O(\log \log n)$!

4.1 Group work: Intuition

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

Important: as you make progress on the question(s), one person in each room should record your progress on <http://PollEv.com/cs265>.

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

$$\Pr \{ \text{Ball } t \text{ is the 3rd (or greater) ball to land in its bucket} \} \leq \frac{\beta_2^2}{n^2},$$

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 $\beta(3, t) \leq \beta_3$ 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 $B(i, t) \leq \beta_i$ 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?

Group Work: Solutions

1. By definition, $\beta_2 = n/2$. There can't be more than 2 buckets with at least $n/2$ balls in them each (for any point t in the process), since we are only dropping n balls total.
2. The probability that ball t is (at least) the third in the bucket that it lands in is at least the probability that both buckets chosen by ball t had at least two things in them. The probability that a random bucket has at least 2 things in it is

$B(2, t-1)/n \leq \beta_2/n$, using the fact that $B(2, t-1) \leq \beta_2$ by the previous part. So the probability that both (independently chosen) buckets have this is bounded by $(\beta_2/n)^2$.

3.

$$\begin{aligned} \mathbb{E}[B(3, t)] &\leq \mathbb{E} \left[\sum_{j=1}^t \mathbf{1}\{\text{ball } j \text{ is } \geq 3\text{'rd in its bucket}\} \right] \\ &= \sum_{j=1}^t \Pr \{\text{ball } j \text{ is } \geq 3\text{'rd in its bucket}\} \\ &\leq n \cdot \left(\frac{\beta_2}{n} \right)^2 = \frac{\beta_2^2}{n} = \beta_3. \end{aligned}$$

4. Even though the question said not to be pedantic about conditioning, we will mention it a little bit here (though still not super formally). Let's condition on the event that $B(3, t) \leq \beta_3$ for all t (which is the same as the event that $B(3, n) \leq \beta_3$). Then

$$\begin{aligned} \mathbb{E}[B(4, t) | B(3, n) \leq \beta_3] &= \mathbb{E} \left[\sum_{j \leq t} \mathbf{1}\{\text{ball } j \text{ is } \geq 4\text{th in its bucket}\} | B(3, n) \leq \beta_3 \right] \\ &= \sum_{j \leq t} \Pr[\text{ball } j \text{ is } \geq 4\text{th in its bucket} | B(3, n) \leq \beta_3] \\ &\leq \sum_{j \leq t} \Pr[\text{ball } j \text{ is } \geq 4\text{th in its bucket} | B(3, j-1) \leq \beta_3] \\ &\leq t \left(\frac{\beta_3}{n} \right)^2 \\ &\leq n \left(\frac{\beta_3}{n} \right)^2 \\ &= \beta_4. \end{aligned}$$

Above, we used the fact that conditioning on $B(3, j-1) \leq \beta_3$ (rather than on $B(3, n) \leq \beta_3$, which is stronger) is only going to make it more likely that the t 'th ball lands in a heavy bucket. (This isn't super formal, but intuitively it seems right).

5. We claim that

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

We can see this by induction. For $i = 2$, this reads $\beta_2 = n/2$, which is true by

definition. Assuming it's true for β_{i-1} , we have

$$\beta_i = \frac{\beta_{i-1}^2}{n} = \frac{n^2}{n \cdot 2^{2^{i-3.2}}} = \frac{n}{2^{2^{i-2}}}.$$

(You could also see this by computing $\beta_3, \beta_4, \beta_5$ and so on and noticing the pattern). Thus, if $B(i, n) \leq \beta_i$ for all i , then at some point, when β_i becomes less than 1, since $B(i, n)$ is an integer we'd know that $B(i, n) = 0$ —aka, there are no buckets with more than i bins. For what i does this happen? We solve for i in the equation

$$2^{2^{i-2}} > n$$

and see that this happens as soon as

$$i > \log \log n + 2.$$

[Go over solutions]

4.2 Extending the intuition...

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 $\beta_i \leftarrow \frac{2\beta_{i-1}^2}{n}$. Define $\beta_4 = n/4$ as before. 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 $i = 2$ follows from the definition of β_2 . Inductively assume that $B(i-1, n) \leq \beta_{i-1}$ with probability at least $1 - (i-1)/n^2$. In the case that $B(i-1, n) \leq \beta_{i-1}$, we have

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

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

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

As long as $\beta_i \geq 6 \log n$ (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 $B(i, j) \leq \beta_i$ 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, $B(i, n) \leq \beta_i$ for all $i \leq i^* = \Theta(\log \log n)$. Just as before, this implies that the max load is $\Theta(\log \log n)$.

Group Work

Important: as you make progress on the question(s), one person in each room should record your progress on <http://PollEv.com/cs265>.

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

1. What are the major problems?

Once you have identified the problems, go to pollev.com/cs265 to list them and/or to up-vote others.

2. 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.

Group Work: Solutions

Here are three major problems:

- It wasn’t legit to apply the Chernoff bound because the events aren’t independent.
- The end of the argument doesn’t check out...we actually showed that whp, there are at most $6 \log n$ bins with $\geq i^*$ balls, not zero.
- We completely ignored any subtlety about conditioning on $B(i-1, t) \leq \beta_{i-1}$ when we were doing the induction.

You can check out the original paper if you want to see how to deal with all of these formally and all at once. It’s a bit tedious and beyond the scope of this class—instead, we’ll just wave our hands about strategies to fix these problems in class if there’s time.

[Go over what the problems are, and if time, waving our hands a bit about how to fix them.]

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>