Class 6

The power of two choices

Warm-Up

- A hypothetical professor wants to assign grading to TAs, but she has no memory.
  - She can't remember which problem has already been assigned to which TA?
- How can she assign n problems to n TAs, so that no TA ends up grading more than $O\left(\frac{\log n}{\log \log n}\right)$ problems?
- Think about how you might do better…
  - It's okay to ask each TA what their current workload is, but you can't remember that info for very long.

Announcements

- HW2 due Friday!

- We realized you hadn't been able to see your quiz grades!
  - **Sorry**, that was not intentional!!
  - Starting now you'll be able to see them soon after the quizzes are due.

Recap I

- Balls and bins!
- Powerful tool: Poissonization (Poissonification?)
- $X \sim Poi(\lambda)$:
  - $\Pr[X = k] = \frac{e^{-\lambda}\lambda^k}{k!}$
  - $\mathrm{E}[X] = \mathrm{Var}[X] = \lambda$
  - $\Pr[|X - \lambda| \geq c] \leq 2\exp\left(\frac{-c^2}{2(c+\lambda)}\right)$

Recap II

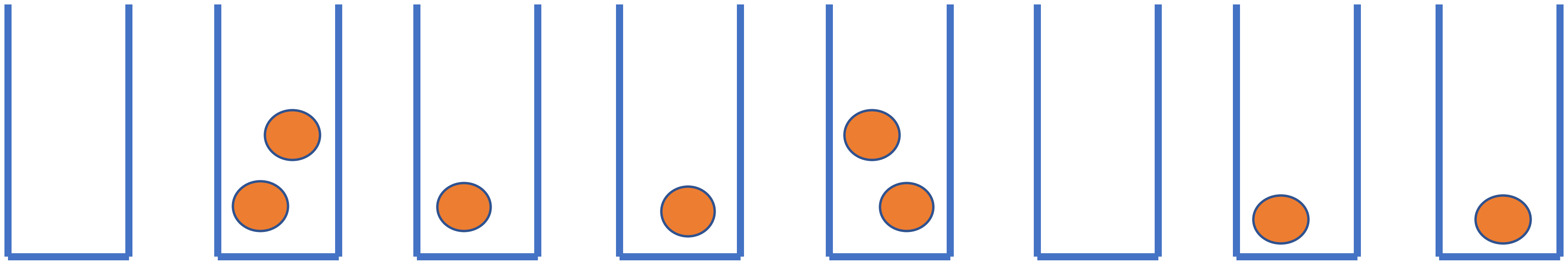
- If you drop $k \sim Poi(n)$ balls into m bins, then:
  - Let X_i = #(Balls in bin i)
  - $X_i \sim Poi\left(\frac{n}{m}\right)$
  - The X_i are all independent

- "Poissonization":
  - #(Balls in bin i when you drop n balls into m bins) $\approx X_i$
  - Work with the X_i instead.

Recap III: Maximum Load

- n balls into n bins.
- Max load is $\Theta\left(\frac{\log n}{\log \log n}\right)$

Solution to warm-up!!

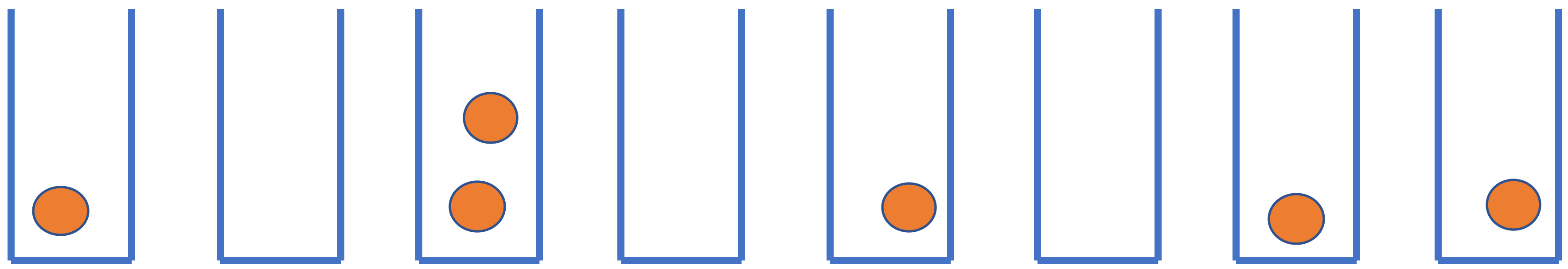


Questions?

Quiz, Mini-lectures, Warm-Up?

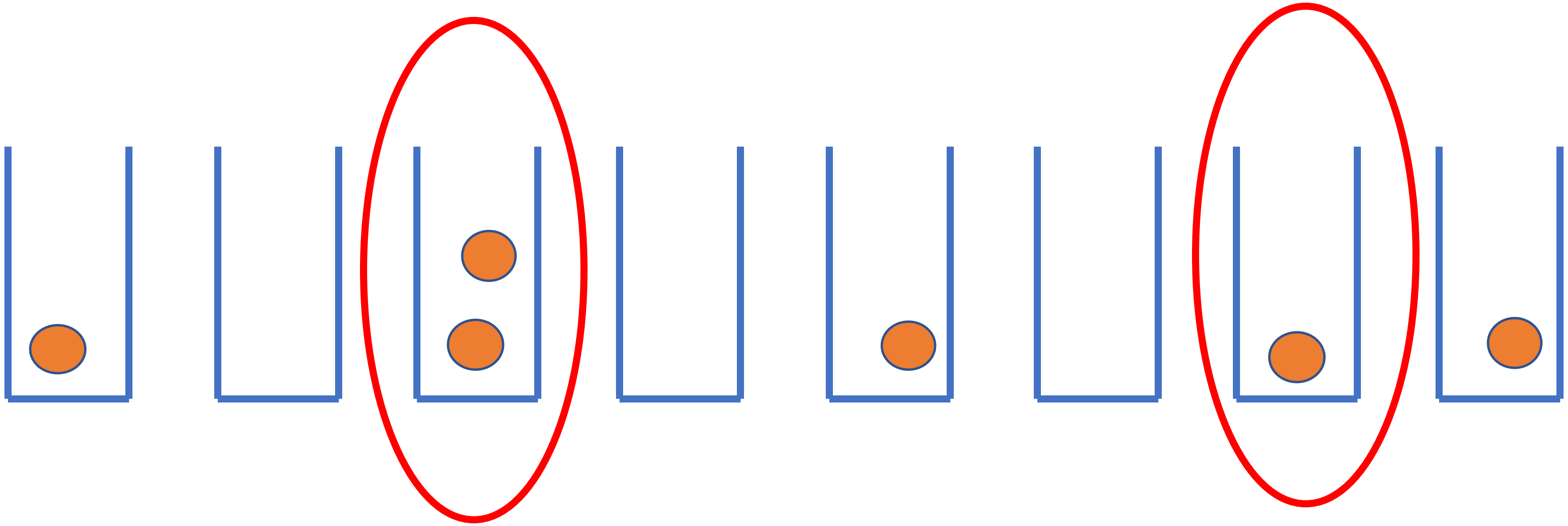
Today: The power of two choices

- Drop n balls into n bins.
- For each ball, pick two bins at random.
- The ball goes in the less-full bin. (Break ties arbitrarily).



Today: The power of two choices

- Drop n balls into n bins.
- For each ball, pick two bins at random.
- The ball goes in the less-full bin. (Break ties arbitrarily).



Today: The power of two choices

- Drop n balls into n bins.
- For each ball, pick two bins at random.
- The ball goes in the less-full bin. (Break ties arbitrarily).

The power of two choices

- n balls into n bins, completely randomly:
  - Max load is $\Theta\left(\frac{\log n}{\log \log n}\right)$

- n balls into n bins, according to the "pick two" scheme:
  - Max load is $\Theta(\log \log n)$

Exponentially smaller!

This is useful, for example, when trying to efficiently assign jobs to processors and wanting to balance the loads.

(Or assigning grading when you have no memory)

For the rest of today…

- We’ll analyze this!

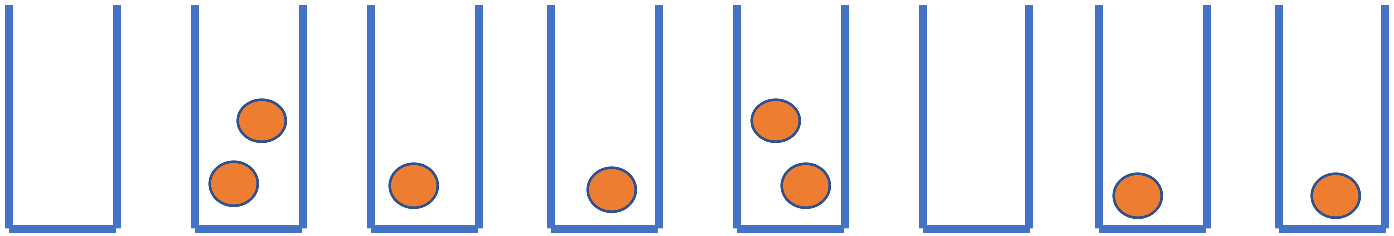
Group Work

- Notation:

$$\beta_2 = \frac{n}{2} \qquad \beta_i = \frac{\beta_{i-1}^2}{n} \text{ for all } i > 2$$

$$\frac{n}{2} \qquad \frac{\left(\frac{n}{2}\right)^2}{n} = \frac{n}{4} \qquad \frac{\left(\frac{n}{4}\right)^2}{n} = \frac{n}{16} \qquad \dots$$

$$B(i,t) = \text{number of bins with} \geq i \text{ balls after step } t$$



This is step $t = 8$

$B(1,8) = 6$

$B(2,8) = 2$

$B(3,8) = 0$

Group Work

Intuition

Definitions:

$$\beta_2 = \frac{n}{2} \qquad \beta_i = \frac{\beta_{i-1}^2}{n}$$

$B(i,t)$ = number of bins with $\geq i$ balls after step t

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,

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

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,

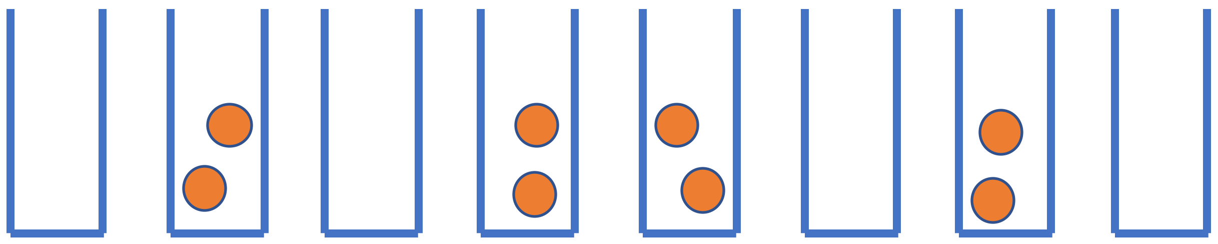
$$\mathbb{E}[B(4,t)] \leq \beta_4.$$

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?

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

Solutions: Question 1

- $\beta_2 = n/2$.
- There can't be more than 2 buckets with $\geq n/2$ balls in them (since there's only n balls total).
- So $B(2,t) \leq B(2,n) \leq \beta_2$



This is step $t = 8$

Definitions:

$$\beta_2 = \frac{n}{2} \qquad \beta_i = \frac{\beta_{i-1}^2}{n}$$

$B(i,t)$ = number of bins with $\geq i$ balls after step t

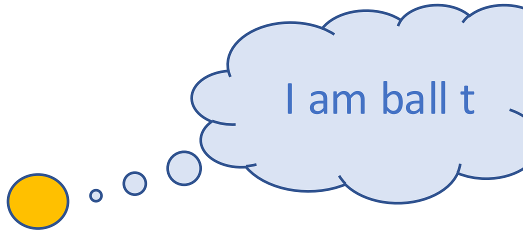
Solutions: Question 2

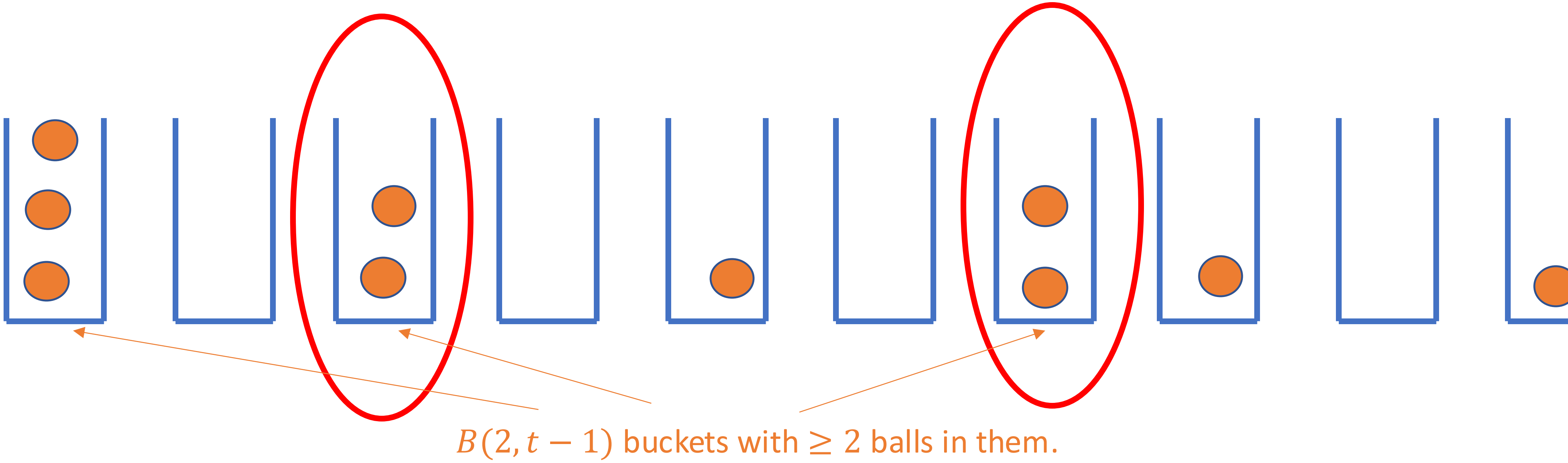
2. Show that

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

for all t.

- Probability that ball t is the third (or greater) ball in its bucket:
  - Need to choose two buckets with at least two things in them.
  - The probability of that is at most $\left(\frac{B(2,t-1)}{n}\right)^2 \leq \left(\frac{\beta_2}{n}\right)^2$





Solutions: Question 3

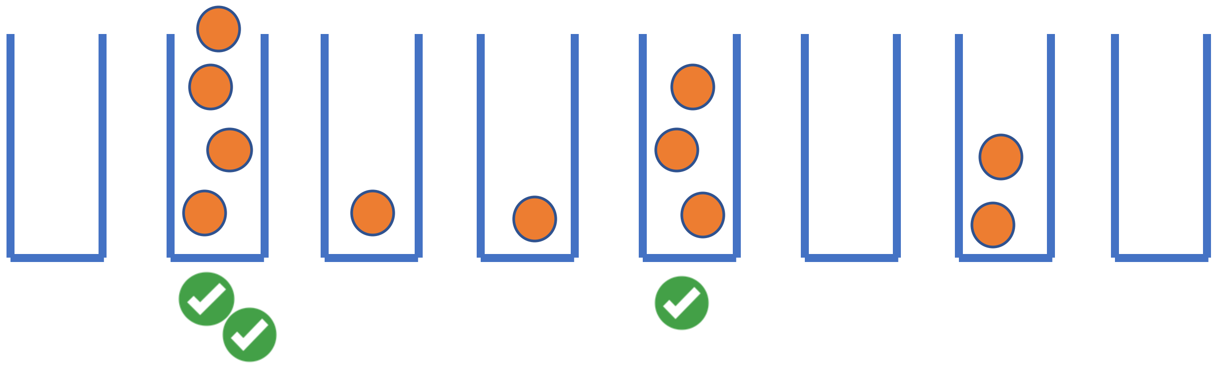
3. Show that, for all t,
$$\mathbb{E}[B(3,t)] \leq \beta_3.$$

$$\mathbb{E}[B(3,t)] \leq \mathbb{E}\left[\sum_{t=1}^{t} 1\{\text{ball t is} \geq 3^{rd} \text{ in bucket}\}\right]$$

$$= \sum_{t=1}^{t} Pr\{\text{ball t is} \geq 3^{rd} \text{ in bucket}\}$$

$$\leq \sum_{t=1}^{n} \left(\frac{\beta_2}{n}\right)^2 = \frac{\beta_2^2}{n} = \beta_3$$

This is because the number of bins with at least 3 balls is at most the number of balls that were at least 3rd in their bin.



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,

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

Solutions: Question 4

$$\mathbb{E}[B(4,t)] \leq \mathbb{E}\left[\sum_{t=1}^{t} 1\{\text{ball t is} \geq 4th \text{ in bucket}\}\right]$$

$$= \sum_{t=1}^{t} Pr\{\text{ball t is} \geq 4th \text{ in bucket}\}$$

$$\leq \sum_{t=1}^{n} \left(\frac{\beta_3}{n}\right)^2 = \frac{\beta_3^2}{n} = \beta_4$$

This is because the number of bins with at least 4 balls is at most the number of balls that were at least 4th in their bin.

*Note: As per the instructions in the question, we are ignoring anything about conditioning on the event that $B(3,t) \leq \beta_3$

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?

Solutions: Question 5

- $\beta_i = \dfrac{n}{2^{2^{i-2}}}$

- You can see this by doing out a bunch and guessing the pattern:
  - $\beta_2 = \frac{n}{2}$
  - $\beta_3 = \frac{1}{n}\left(\frac{n}{2}\right)^2 = \frac{n}{2^2}$
  - $\beta_4 = \frac{1}{n}\left(\frac{n}{2^2}\right)^2 = \frac{n}{2^{2^2}}$
  - $\beta_5 = \frac{1}{n}\left(\frac{n}{2^{2^2}}\right)^2 = \frac{n}{2^{2^3}}$
  - $\beta_6 = \frac{1}{n}\left(\frac{n}{2^{2^3}}\right)^2 = \frac{n}{2^{2^4}}$

(And formally you can prove it by induction.)

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?

Solutions: Question 5

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

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?

Solutions: Question 5

- $\beta_i = \frac{n}{2^{2^{i-2}}}$
- Say $\mathbb{E}B(i,n) \leq \beta_i$ for all i.
- Suppose that at some point (for large enough i), we have $\beta_i \ll 1$.
  - That would imply that $\mathbb{E}B(i,n) \ll 1$.
  - By Markov's inequality, $\Pr[\, B(i,n) \geq 1] \ll 1$.
  - So with high probability, there are no bins with more than i balls!
- Set $\frac{n}{2^{2^{i-2}}} < 1$ and solve for i: get $i > \log\log n + 2$.
- Conclude that max load is $\Theta(\log\log n)$, ***assuming*** that the induction works… (but we were cheating!)

Here's the outline for a fixed argument

WARNING: This is incorrect in a few ways.

1. Define $\beta_2 = n/2$, $\beta_i = 2(\beta_{i-1})^2/n$

this "2" is new.

number of bins w/ $\geq i$ balls after all n are tossed.

2. Argue by induction on i that, with probability $\geq 1 - i/n^2$, $B(i,n) \leq \beta_i$:

number of bins w/ $\geq i$ balls after all n are tossed.

2. Argue by induction on i that, with probability $\geq 1 - i/n^2$, $B(i,n) \leq \beta_i$:

- Base case for $i=2$ is by definition.

number of bins w/ $\geq i$ balls after all n are tossed.

2. Argue by induction on i that, with probability $\geq 1 - i/n^2$, $B(i,n) \leq \beta_i$:

- Base case for $i=2$ is by definition.

- Assuming that $B(i-1, n) \leq \beta_{i-1}$ (which holds w/ prob $\geq 1 - \frac{i-1}{n^2}$ by induction)

  the same logic from before implies that

$$\mathbb{E}\left[\sum_{t=1}^{n} \mathbb{1}\left\{ \begin{array}{c} \text{ball } t \text{ is the } \geq i^{th} \\ \text{in its bucket} \end{array} \right\} \right] \leq \beta_{i-1}^2 / n$$

number of bins w/ $\geq i$ balls after all n are tossed.

2. Argue by induction on i that, with probability $\geq 1 - i/n^2$, $B(i,n) \leq \beta_i$:

- Base case for $i=2$ is by definition.

- Assuming that $B(i-1,n) \leq \beta_{i-1}$ (which holds w/ prob $\geq 1 - \frac{i-1}{n^2}$ by induction) the same logic from before implies that

$$\mathbb{E}\left[\sum_{t=1}^{n} \mathbb{1}\left\{ \begin{array}{c} \text{ball } t \text{ is the } \geq i^{th} \\ \text{in its bucket} \end{array} \right\} \right] \leq \beta_{i-1}^2 / n$$

- A Chernoff bound says that

$$\mathbb{P}\left[\sum_{t=1}^{n} \mathbb{1}\left\{ \begin{array}{c} \text{ball } t \text{ is the } \geq i^{th} \\ \text{in its bucket} \end{array} \right\} > \frac{2\beta_{i-1}^2}{n} \right] \leq \exp\left(-\beta_i/3\right)$$

this is β_i, but also twice the expectation

we have $B(i,n) \leq$ this, as before.

number of bins w/ $\geq i$ balls after all n are tossed.

2. Argue by induction on i that, with probability $\geq 1 - i/n^2$, $B(i,n) \leq \beta_i$:

- A Chernoff bound says that

this is β_i, but also twice the expectation

$$\mathbb{P}\left[\sum_{t=1}^{n} \mathbb{1}\left\{ \begin{array}{c} \text{ball } t \text{ is the } \geq i^{th} \\ \text{in its bucket} \end{array} \right\} > \frac{2\beta_{i-1}^2}{n} \right] \leq \exp\left(-\beta_i/3\right)$$

we have $B(i,n) \leq$ this, as before.

number of bins w/ $\geq i$ balls after all n are tossed.

2. Argue by induction on i that, with probability $\geq 1 - 1/n^2$, $B(i,n) \leq \beta_i$:

- A Chernoff bound says that

this is β_i, but also twice the expectation

$$\mathbb{P}\left[\sum_{t=1}^{n} \mathbb{1}\left\{ \begin{array}{c} \text{ball } t \text{ is the } \geq i^{th} \\ \text{in its bucket} \end{array} \right\} > \frac{2\beta_{i-1}^2}{n} \right] \leq \exp(-\beta_i/3)$$

we have $B(i,n) \leq$ this, as before.

- Therefore, if $\beta_i \geq 6\log n$, this probability is $\leq 1/n^2$. By a union bound w/ the event that $B(i-1,n) > \beta_{i-1}$, $\mathbb{P}[B(i,n) > \beta_i] \leq \frac{i}{n^2}$

prob. is $\leq \frac{i-1}{n^2}$

This establishes the inductive hypothesis for i, as long as $\beta_i \geq 6\log n$

Now we know that $B(i,n) \leq \beta_i$ with probability at least $1 - i/n^2$, as long as $\beta_i \geq 6 \log n$

3. Choose i^* so that $\beta_{i^*} \geq 6 \cdot \log n$.

So the argument above shows that, whp, $B(i,n) \leq \beta_i \;\forall\, i \leq i^*$.

You can check that $\beta_{i^*} \approx 6 \log n$ when $i^* = \Theta(\log\log n)$.

This part is not the problem!

Now we know that $B(i,n) \leq \beta_i$ with probability at least $1 - i/n^2$, as long as $\beta_i \geq 6 \log n$

3. Choose i^* so that $\beta_{i^*} \geq 6 \cdot \log n$.

   So the argument above shows that, whp, $B(i,n) \leq \beta_i \ \forall\ i \leq i^*$.

   You can check that $\beta_{i^*} \approx 6 \log n$ when $i^* = \Theta(\log \log n)$.

4. We conclude that, whp, $B(i,n) \leq \beta_i$ 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

- What was wrong with this argument/sketch?
- There are at least two or three major problems
  - depending on what you count as “major”

Three problems

1. Can't apply the Chernoff bound – the random variables are not independent!
2. The end of the argument doesn't make any sense! We showed that $B(i,n) \leq \beta_i$ whenever $i \leq i^*$, but $\beta_{i^*} \approx 6 \log n$.
   - So there are still about $6 \log n$ buckets with at least i^* balls in them.
   - (It is true that $i^* = \Theta(\log \log n)$ though).
3. We are not being careful about the conditioning.

Fix for the Chernoff bound problem (sketch)

$$\mathbb{P}\left\{ \sum_{t=1}^{n} \mathbb{1}\left\{ \text{Ball } t \text{ is } \geq i^{th} \text{ ball to land in its bin} \right\} \right\}$$

$$= \mathbb{P}\left\{ \underbrace{\sum_{t=1}^{n-1} \mathbb{1}\left\{ \text{Ball } t \text{ is } \geq i^{th} \text{ ball to land in its bin} \right\}} + \underbrace{\mathbb{1}\left\{ \text{Ball } n \text{ is the } \geq i^{th} \text{ ball to land in its bin} \right\}} \right\}$$

Inductively assume that this sum behaves like $\text{Binomial}\left(n-1, \leq \left(\frac{\beta_{i-1}}{n}\right)^2\right)$

Conditioned on balls $1, 2, \ldots, n$, we STILL have $\mathbb{E}\left[\mathbb{1}\left\{ \text{Ball } n \text{ is the } \geq i^{th} \text{ ball to land in its bin} \right\}\right] \leq \left(\frac{\beta_{i-1}}{n}\right)^2$, over the randomness in ball n

both of these together behave like $\text{Binomial}\left(n, \leq \left(\frac{\beta_{i-1}}{n}\right)^2\right)$

Fix for the “The argument didn’t finish!” problem.

Say that $B(i,n) \leq \beta_i \;\; \forall \; i \leq i^*$, where $i^* = \Theta(\log\log n)$ is such that $\beta_{i^*} = 6\log n$.

This is what we actually claimed to have shown.

Then $\mathbb{P}\{B(i^*+1, n) \geq 1\} \leq \mathbb{P}\{B(i^*, n) \geq 2\}$

because there have to be at least 2 bins w/ i^* + 1 things in them for some bin to end up w/ i^* things.

$$\leq \sum_{s<t} \mathbb{P}\left\{ \begin{array}{l} \text{BOTH balls } s \text{ and } t \\ \text{are} \geq \text{the } i^{*\prime}\text{th} \\ \text{thing in their bin} \end{array} \right\}$$

(union bound over all the balls)

$$\leq \binom{n}{2} \mathbb{P}\left\{ \begin{array}{l} s \text{ is} \geq i^{*\text{th}} \\ \text{in its} \\ \text{bin} \end{array} \right\} \cdot \mathbb{P}\left\{ \begin{array}{l} t \text{ is} \geq i^{*\text{th}} \\ \text{in its bin} \end{array} \;\middle|\; \begin{array}{l} s \text{ is} \geq i^{*\text{th}} \\ \text{in its} \\ \text{bin} \end{array} \right\}$$

this doesn't affect our argument before for t, assuming $B(i^*, n) \leq \beta_{i^*}$.

Fix for the "The argument didn't finish!" problem, ctd.

$$\mathbb{P}\{B(i^*+1,n)\geq 1\} \leq \binom{n}{2}\,\mathbb{P}\left\{\begin{array}{l}s \text{ is} \geq i^{*\,th} \\ \text{in its} \\ \text{bin}\end{array}\right\}\cdot\mathbb{P}\left\{\begin{array}{l}t \text{ is} \geq i^{*\,th} \\ \text{in its bin}\end{array}\;\middle|\;\begin{array}{l}s \text{ is} \geq i^{*\,th} \\ \text{in its} \\ \text{bin}\end{array}\right\}$$

this doesn't affect our argument before for t, assuming $B(i^*,n) \leq \beta_{i^*}$.

$$\leq \binom{n}{2}\left(\frac{\beta_{i^*}}{n}\right)^2\cdot\left(\frac{\beta_{i^*}}{n}\right)^2$$

$$\leq n^2 \cdot \left(\frac{6\log n}{n}\right)^4 = \frac{6^4\log^4 n}{n^2} = o(1).$$

Fix for the "We are being sloppy about the conditioning" problem.

- Be less sloppy about the conditioning.
  - (It's a bit delicate but not that interesting…)

Next time!

- Metric Embeddings!!