Class 6

The power of two choices
Recap I

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

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

• “Poissonization”:
  • \( \#(\text{Balls in bin } i \text{ when you drop } n \text{ balls into } m \text{ bins}) \approx X_i \)
  • Work with the \( X_i \) instead.
Recap III: Maximum Load

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

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

2. Show that

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

3. Show that, for all \( t \),

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

4. \textbf{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. \textbf{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?

Definitions:

\[
\beta_2 = \frac{n}{2} \quad \beta_i = \frac{\beta_{i-1}^2}{n}
\]

\( B(i, t) \) = number of bins with \( \geq i \) balls after step \( t \)
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 $B(2, t) \leq B(2, n) \leq \beta_2$

Definitions:

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

$B(i, t) = \text{number of bins with } \geq i \text{ balls after step } t$
Solutions: Question 2

• 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 $\left( \frac{B(2, t-1)}{n} \right)^2 \leq \left( \frac{\beta_2}{n} \right)^2$

$B(2, t - 1)$ buckets with $\geq 2$ balls in them.
Solutions: Question 3

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

\[
E[B(3, t)] \leq E[B(3, n)]
\]

\[
\leq E\left[ \sum_{t=1}^{n} \mathbb{1}\left\{ \text{ball } t \text{ is } \geq 3\text{rd in its bucket} \right\} \right]
\]

\[
= \sum_{t=1}^{n} P\left\{ \text{ball } t \text{ is } \geq 3\text{rd in its bucket} \right\}
\]

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

def. of \( \beta_3 \)
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 $t$,

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

\[
\begin{align*}
\mathbb{E}[B(4, t)] &\leq \mathbb{E}[B(4, n)] \\
&\leq \mathbb{E} \left[ \sum_{t=1}^{n} \mathbb{I} \left\{ \text{ball } t \text{ is } \geq 4^{th} \text{ in its bucket} \right\} \right] \\
&= \sum_{t=1}^{n} \mathbb{P} \left\{ \text{ball } t \text{ is } \geq 4^{th} \text{ in its bucket} \right\} \\
&\leq \sum_{t=1}^{n} \left( \frac{\beta_3}{n} \right)^2 = \frac{\beta_3^2}{n} = \beta_4
\end{align*}
\]

*Note: As per the instructions in the question, we are ignoring anything about conditioning on the event that $B(3, t) \leq \beta_3$*
Solutions: Question 5

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

- 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.)
Solutions: Question 5

• $\beta_i = \frac{n}{2^{2i-2}}$
5. Suppose that this logic continued, and you could show that $E[B(i, t)] \leq \beta_i$ for all $t$. What would the max load be?

Solutions: Question 5

• $\beta_i = \frac{n}{2^{2i-2}}$

• If we believe that $B(i, n) \leq \beta_i$ for all $i$, then at some point (for large enough $i$), we have $\beta_i < 1$.
  • That would imply that $B(i, n) = 0$.
  • That would mean that the number of bins with load $\geq i$ is 0, aka the max load is $i - 1$. 
Solutions: Question 5

• $\beta_i = \frac{n}{2^{2i-2}}$

• If we believe that $B(i, n) \leq \beta_i$ for all $i$, then at some point (for large enough $i$), we have $\beta_i < 1$.
  • That would imply that $B(i, n) = 0$.
  • That would mean that the number of bins with load $\geq i$ is 0, aka the max load is $i - 1$.

• Set $\frac{n}{2^{2i-2}} < 1$ and solve for $i$: get $i > \log \log n + 2$. 

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

• \( \beta_i = \frac{n}{2^{2^i - 2}} \)

• If we believe that \( B(i, n) \leq \beta_i \) for all \( i \), then at some point (for large enough \( i \)), we have \( \beta_i < 1 \).
  • That would imply that \( B(i, n) = 0 \).
  • That would mean that the number of bins with load \( \geq i \) is 0, aka the max load is \( i - 1 \).

• 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 “in expectation” stuff holds exactly.
Here’s the outline for an argument

WARNING: This is incorrect in a few ways.

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

   this "2" is new.
2. Argue by induction on \( i \) that, with probability \( \geq 1 - \frac{i}{2n} \), \( B(i,n) \leq B_i \):
2. Argue by induction on \( i \) that, with probability \( \ge 1 - \frac{i}{2n} \), \( B(i,n) \le \beta_i \):

- Base case for \( i = 2 \) is by definition.
Define \( A \) to be the event that Chernoff-Hoeffding bound holds for all the \( n \) bins after \( n \) items are tossed.

2. Argue by induction on \( i \) that, with probability \( \geq 1 - \frac{i}{n^2} \), \( B(i, n) \leq \beta_i \cdot n \):

   - Base case for \( i = 2 \) is by definition.

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

   the same logic from before implies that

\[
\mathbb{E}\left[ \sum_{t=1}^{n} 1\{ \text{ball } t \text{ is the } \text{ith\textsuperscript{th}} \text{ in its bucket} \} \right] \leq \frac{\beta_{i-1}}{n}
\]
2. Argue by induction on \( i \) that, with probability \( \geq \left( 1 - \frac{i}{n^2} \right) \), \( B(i, n) \leq \beta_i \):

- **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

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

- **A Chernoff bound** says that

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

we have \( B(i, n) \leq \text{this, as before.} \)
2. Argue by induction on \( i \) that, with probability \( \geq 1 - \frac{i}{n^2} \), \( B(i, n) \leq \beta_i \):

- A Chernoff bound says that

\[
P\left[ \sum_{t=1}^{n} I\{ \text{ball } t \text{ is the } i^{th} \text{ in its bucket} \} > \frac{2\beta_i^2}{n} \right] \leq \exp\left(-\frac{\beta_i}{3}\right)
\]

we have \( B(i, n) \leq \text{this, as before.} \)

number of bins w/ \( \geq i \) balls after all \( n \) are tossed.
2. Argue by induction on $i$ that, with probability $\geq 1 - \frac{i}{n^2}$, $B(i,n) \leq \beta_i$:

- A Chernoff bound says that

$$
\Pr \left[ \sum_{t=1}^{n} \mathbb{1}\{ \text{ball } t \text{ is the } i^{\text{th}} \text{ in its bucket} \} > \frac{2 \beta_i^2}{n} \right] \leq \exp \left( - \frac{\beta_i}{3} \right)
$$

we have $B(i,n) \leq \beta_i$, but also twice the expectation.

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

This establishes the inductive hypothesis for $i$, as long as $\beta_i \geq 6 \log n$.
3. Choose $i^*$ so that $\beta_{i^*} = 6 \cdot \log n$.

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

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

\[
P \left\{ \sum_{t=1}^{n} 1 \left\{ \text{Ball } t \text{ is } \geq i^{th} \text{ ball to land in its bin} \right\} \right\} \\
= P \left\{ \sum_{t=1}^{n-1} 1 \left\{ \text{Ball } t \text{ is } \geq i^{th} \text{ ball to land in its bin} \right\} \right\} + \left\{ \text{Ball } n \text{ is the } i^{th} \text{ ball to land in its bin} \right\} \\
\text{Inductively assume that this sum behaves like Binomial } (n-1, \left( \frac{p_{i-1}}{n} \right)^2) \\
\text{Conditioned on balls } 1, 2, \ldots, n, \text{ we STILL have } E \left[ \sum_{t=1}^{n-1} 1 \left\{ \text{Ball } t \text{ is the } i^{th} \text{ ball to land in its bin} \right\} \right] \leq \left( \frac{p_{i-1}}{n} \right)^2, \text{ over the randomness in ball } n. \\
\text{both of these together behave like Binomial } (n, \left( \frac{p_{i-1}}{n} \right)^2)
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 \( \beta_{i^*} = 6 \log n \).

Then \( \Pr \left\{ B(i^* + 1,n) \geq 1 \right\} \leq \Pr \left\{ B(i^*, n) \geq 2 \right\} \)

\[ \leq \sum_{s < t} \Pr \left\{ \text{both balls } s \text{ and } t \geq \text{some } i^* \text{ thing in their bin} \right\} \] (union bound over all the balls)

\[ \leq \binom{n}{2} \Pr \left\{ s \geq i^* \text{ thing in its bin} \right\} \cdot \Pr \left\{ t \geq i^* \text{ thing in its bin} \right\} \]

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

This is what we actually claimed to have shown.
Fix for the “The argument didn’t finish!” problem.

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

This is what we actually claimed to have shown.
Fix for the “The argument didn’t finish!” problem.

Say that $B(i, n) \leq \beta_i \quad \forall \ i \leq i^*$, where $i^* = \Theta(\log \log n)$ is such that $\beta_{i^*} = 6 \log n$. Then $P\{ B(i^*+1, n) \geq 1 \} \leq P\{ B(i^*, n) \geq 2 \}$

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

This is what we actually claimed to have shown.
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 $\beta_{i^*} = 6 \log n$.

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

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

$\leq \sum_{s < t} \mathbb{P}\left\{ \text{both balls } s \text{ and } t \text{ are } \geq \text{ the } i^* \text{'th thing in their bin} \right\}$ (union bound over all the balls)

This is what we actually claimed to have shown.
Fix for the “The argument didn’t finish!” problem.

Say that $B(i, n) \leq i$ for $i \leq i^*$, where $i^* = \Theta(\log \log n)$ is such that $B_{i^*} = 6 \log n$.

Then $P\left\{ B(i^* + 1, n) \geq 1 \right\} \leq P\left\{ B(i^*, n) \geq 2 \right\}$ because there have to be at least 2 bins w/ $i^* + 1$ things in them for some bin to end up w/ $i^*$ things.

$$\leq \sum_{s < t} P\left\{ \text{both balls s and t are } \geq \text{ the } i^* \text{ th thing in their bin} \right\}$$ (union bound over all the balls)

$$\leq \binom{n}{2} P\left\{ \text{ s is } \geq i^* \text{ th in its bin} \right\} P\left\{ \text{ t is } \geq i^* \text{ th in its bin} \right\}$$

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

This is what we actually claimed to have shown.
Fix for the “The argument didn’t finish!” problem, ctd.

\[ \Pr \left\{ B(i+1, n) \geq 1 \right\} \leq \binom{n}{2} \Pr \left\{ S \text{ is } \leq i^{*+m} \right\} \cdot \Pr \left\{ t \text{ is } \leq i^{*+m} \mid S \text{ is } \leq i^{*+m} \right\} \]

\[ \leq n^2 \cdot \left( \frac{\beta_i^*}{n} \right)^2 \cdot \left( \frac{\beta_i^*}{n} \right)^2 \]

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