CS7800: Advanced Algorithms

Class 22: Randomized Algorithms III

- Balls and Bins: Chernoff Bounds
- Universal Hashing

Jonathan Ullman

November 25, 2025

Balls and Bins: Maximum Load

m balls n bins

- Let L_i be the number of balls in bin i
- Expected maximum load is $\mathbb{E}(\max_i L_i) = \sum_{k=1}^{\infty} \underbrace{\mathbb{P}(\max_i L_i \geq k)}_{\text{want to bound this probability}}$

So far:

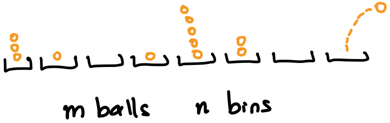
① Trivial bound: $\mathbb{E}(\max L_i) \leq m$

② Markov's inequality: $\mathbb{E}(\max L_i) \leq \infty$

③ Chebyshev's inequality: $\mathbb{E}(\max L_i) \leq O\left(\frac{m}{n} + \sqrt{m}\right)$

Today: Tighter analysis with Chernoff Bounds

Balls and Bins: Maximum Load



- Let L_i be the number of balls in bin i
- Expected maximum load is $\mathbb{E}(\max_i L_i) = \sum_{k=1}^{\infty} \mathbb{P}(\max_i L_i \geq k)$
- Let $L_{i,j} = \begin{cases} 1 & \text{if ball } j \text{ is in bin } i \\ 0 & \text{otherwise} \end{cases}$ $\Rightarrow L_i = L_{i,1} + \dots + L_{i,m}$

Aside: Central Limit Theorem

Chernoff Bounds

$$Z_i = \begin{cases} 1 \text{ with prob } p \\ 0 \text{ with prob } 1-p \end{cases}$$

$$Z = Z_1 + \ldots + Z_m$$

$Z_1, \ldots, Z_m$ independent

$$\mu = \mathbb{E}(Z) = pm$$

Thm: $\mathbb{P}(Z \geq (1+\varepsilon)\mu) \leq \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}}\right)^{\mu}$

$\varepsilon < 1$ $\quad \longrightarrow \quad e^{-\frac{\mu\varepsilon^2}{4}}$

$\varepsilon > 1$ $\quad \longrightarrow \quad e^{-\mu\varepsilon}$

Chernoff Bounds

Thm: $\mathbb{P}\left(Z \geq (1+\varepsilon)\mu\right) \leq \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}}\right)^{\mu}$

$Z_i = \begin{cases} 1 \text{ with prob } p \\ 0 \text{ with prob } 1-p \end{cases}$ $\qquad Z = Z_1 + \dots + Z_m$

$Z_1, \dots, Z_m$ independent $\qquad \mu = \mathbb{E}(Z) = pm$

Proof:

$$
\begin{aligned}
\mathbb{P}(Z \geq t\mu) &= \mathbb{P}\left(e^{sZ} \geq e^{st\mu}\right) && \leftarrow \text{What is this dark magic?} \\
&\leq e^{-st\mu} \cdot \mathbb{E}\left(e^{sZ}\right) && \leftarrow \text{Markov} \\
&= e^{-st\mu} \cdot \mathbb{E}\left(\prod_i e^{sZ_i}\right) \\
&= e^{-st\mu} \cdot \prod_i \underbrace{\mathbb{E}\left(e^{sZ_i}\right)}_{\substack{Z_i \in \{0,1\} \text{ so this} \\ \text{is something "simple"}}} && \leftarrow \text{Independence}
\end{aligned}
$$

Chernoff Bounds

Thm: $\mathbb{P}\left(Z \geq (1+\varepsilon)\mu\right) \leq \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}}\right)^{\mu}$

$Z_i = \begin{cases} 1 \text{ with prob } p \\ 0 \text{ with prob } 1-p \end{cases}$

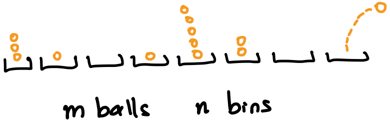
$Z_1, \ldots, Z_m$ independent

$Z = Z_1 + \ldots + Z_m$

$\mu = \mathbb{E}(Z) = pm$

Proof Cont'd:

Balls and Bins: Maximum Load



- Let L_i be the number of balls in bin i
- Expected maximum load is $\mathbb{E}(\max_i L_i) = \sum_{k=1}^{\infty} \mathbb{P}(\max_i L_i \geq k)$
- Let $L_{i,j} = \begin{cases} 1 & \text{if ball } j \text{ is in bin } i \\ 0 & \text{otherwise} \end{cases}$ $\Rightarrow$ $\underbrace{L_i = L_{i,1} + \ldots + L_{i,m}}_{\text{Apply Chernoff Bound}}$

Application: Hash Tables

"universe"

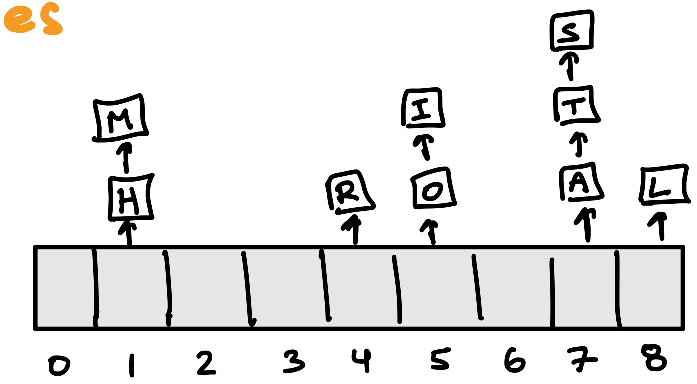
Goal: Store a set of m elements $S \subseteq \mathcal{U}$, such that we can efficiently check if $x \in S$

↳ A "dictionary" also lets us associate a value with each key x

- A hash table $T[1:n]$ stores the elements in n bins
- A hash function $h: \mathcal{U} \rightarrow \{0,1,\dots,n-1\}$ maps elements to bins $x \mapsto T[h(x)]$

Application: Hash Tables

Linear chaining: a common way to deal with collisions



Looks a lot like balls in bins!

- Load factor = $\frac{m}{n}$

- Let $\ell(x)$ = # of elements in the same bin as x

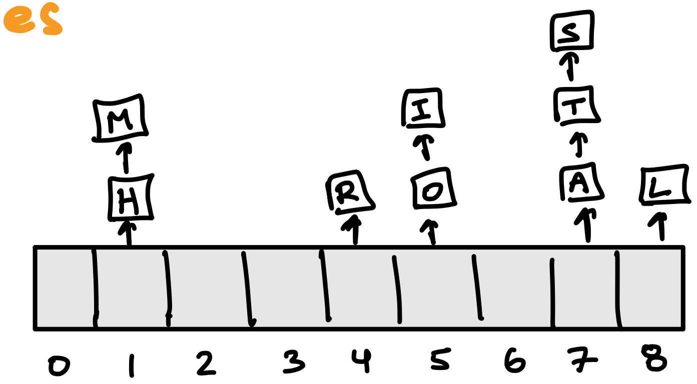
  $\#\{y \in S : h(y) = h(x)\}$ "collisions"

- Time to lookup $x \in \mathcal{U}$ is $O(\ell(x))$

Worst-case lookup time = $\max_{x \in \mathcal{U}} \ell(x)$

Application: Hash Tables

How should we choose the hash function $h: \mathcal{U} \to \{0,1,\ldots,n-1\}$ to have small maximum load?



Looks a lot like balls in bins!

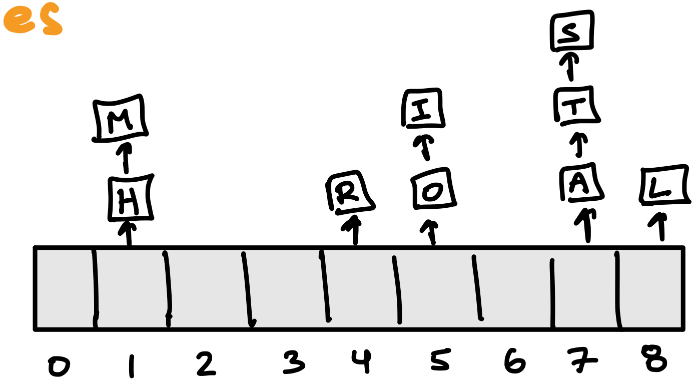
Randomized hash function:

- Model h as a uniformly random function $\mathcal{U} \to \{0,1,\ldots,n-1\}$
- Fix the set S and study $\underbrace{\mathbb{E}}\left(\max_x \ell(x)\right)$

expectation over random choice of h

Application: Hash Tables

How should we choose the hash function $h: \mathcal{U} \to \{0,1,\dots,n-1\}$ to have small maximum load?



Looks a lot like balls in bins!

Uniformly random hash functions: (load factor $\frac{m}{n} = 1$)

- Expected max load is $\Theta\left(\frac{\log n}{\log\log n}\right)$ [Will be true 99.999% of the time]

- For any $x \in \mathcal{U}$, expected lookup time is $\mathbb{E}(\ell(x))$

Universal Hash Families

A <u>hash family</u> is a set of hash functions

$$\mathcal{H} \subseteq \{ h : \mathcal{U} \to \{0, 1, \ldots, n-1\} \}$$

<u>Definition</u>: $\mathcal{H}$ is 2-universal if for every distinct $x, y \in \mathcal{U}$

$$\mathbb{P}_h\left(h(x) = h(y) \right) \leq \frac{1}{n}$$

Behaves like a uniformly random hash fn if we only look at pairs of points

Constructing Universal Hashing

$\mathcal{H}$ is 2-universal if for every distinct $x, y \in \mathcal{U}$

$$\mathbb{P}_{h}(h(x) = h(y)) \leq \frac{1}{n}$$

Construction:

- Fix a prime $p \geq |\mathcal{U}|$, bins n
- Let $h_{a,b}(x) \equiv (ax + b \bmod p) \bmod n$

$$\mathcal{H}_{p,n} = \{ h_{a,b} : a \in \mathbb{Z}_p^{\neq 0}, b \in \mathbb{Z}_p \}$$

Theorem: $\mathcal{H}_{p,n}$ is a 2-universal hash family

Constructing Universal Hashing

$\mathcal{H}$ is 2-universal if for every distinct $x, y \in \mathcal{U}$

$$\mathbb{P}_h\big(h(x) = h(y)\big) \le \frac{1}{n}$$

$$h_{a,b}(x) = (ax + b \mod p) \mod n$$

a,b random
$a \neq 0$

p fixed

<u>Lemma 1:</u> For every prime p and $a \neq 0$ there is a unique $a^{-1} \in \{1, 2, \dots, p-1\}$ such that $a^{-1} \cdot a = 1 \mod p$

<u>Proof:</u> ① $az = c \mod p$ has <u>at most</u> one solution

if $az = az' \mod p \implies a(z - z') = 0 \mod p$

for $z, z' \in \{1, 2, \dots, p-1\} \implies z - z'$ divisible by p

$\implies z - z' = 0$

② $az = 0 \mod p$ has <u>no</u> solutions

Constructing Universal Hashing

$\mathcal{H}$ is 2-universal if for every distinct $x, y \in \mathcal{U}$

$$\underset{h}{\mathbb{P}}(h(x) = h(y)) \leq \frac{1}{n}$$

$$h_{a,b}(x) = (ax + b \mod p) \mod n$$

a,b random
$a \neq 0$

p fixed

Lemma 2: If $x \neq y$ and $r \neq s$ then the system

$$ax + b = r \mod p$$
$$ay + b = s \mod p$$

has a unique solution

Proof:

Constructing Universal Hashing

$\mathcal{H}$ is 2-universal if for every distinct $x, y \in \mathcal{U}$

$$\mathbb{P}_h\left(h(x) = h(y)\right) \le \frac{1}{n}$$

$$h_{a,b}(x) = \left(\underbrace{ax+b}_{\substack{a,b \text{ random} \\ a \neq 0}} \bmod \underbrace{p}_{p \text{ fixed}} \right) \bmod n$$

Thm: $\mathcal{H}_{p,n}$ is 2-universal

Proof: By Lemma 2,

$$\mathbb{P}_{a,b}\left(ax+b = r \bmod p \wedge ay+b = s \bmod p\right) = \frac{1}{p(p-1)}$$

$$\mathbb{P}_{a,b}\left(h_{a,b}(x) = h_{a,b}(y)\right) = \frac{N}{p(p-1)}$$

where N is number of $r \neq s$ in $\mathbb{Z}_p$ so that $r = s \bmod n$

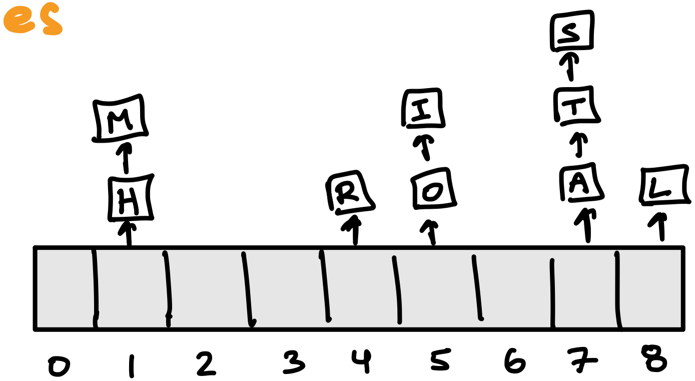
$$N \le \underbrace{p}_{\text{choices of } r} \cdot \underbrace{\frac{p-1}{n}}_{\text{choices of } s \text{ for a given } r}$$

$$\le \frac{p(p-1)}{m\,p(p-1)}$$

$$= \frac{1}{m}$$

Application: Hash Tables

How should we choose the hash function $h: \mathcal{U} \to \{0,1,\dots,n-1\}$ to have small maximum load?



Looks a lot like balls in bins!

Universal

~~Uniformly random~~ hash functions: $\left(\text{load factor } \frac{m}{n} = 1\right)$

- Expected max load is $\cancel{\Theta\left(\frac{\log n}{\log\log n}\right)}$ $\Theta(\sqrt{n})$

- For any $x \in \mathcal{U}$, expected lookup time is

$$\mathbb{E}(\ell(x)) = \sum_{y \in S} \mathbb{P}_h(h(x)=h(y)) \le n \cdot \frac{1}{n} = 1$$

Still true for universal hash families!