

# CS7800: Advanced Algorithms

## Class 21: Randomized Algorithms II

- Balls and Bins: maximum vs expected load
- Universal Hashing

Jonathan Ullman

November 21, 2025

# Probability Case Study: Balls and Bins

Throw  $m$  balls into  
 $n$  bins independently

$$\Omega = \{1, \dots, n\}^m$$

$$\omega = (6, 8, 11, 2, 37, \dots)$$

↑      ↑  
Ball 1   Ball 2  
Bin 6   Bin 8



$$\mathbb{P}(\omega_1, \omega_2, \dots, \omega_m) = \frac{1}{n^m}$$

Questions:

- ① How long until bin 1 gets a ball?
- ② How long until no bin is empty?
- ③ What is the maximum number of balls in any bin?

# Application: Hash Tables

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

"universe"

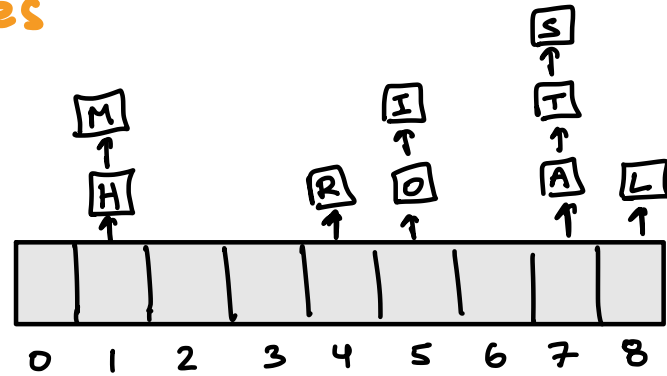


↳ A "dictionary" also lets us associate a value  
with each key  $x$

- 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



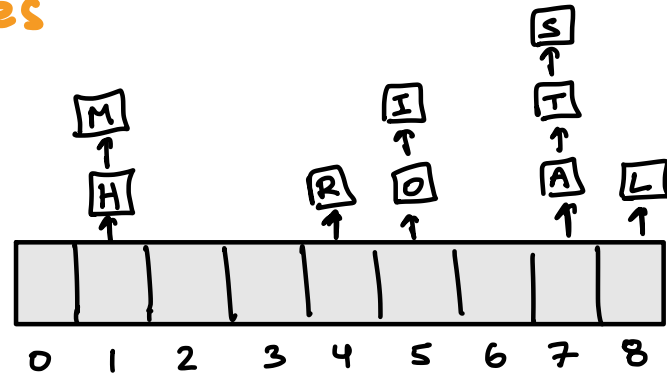
Looks a lot like balls in bins!

- Load factor =  $\frac{m}{n}$
- Let  $l(x) = \# \text{ of elements in the same bin as } x$   
 $\# \{y \in S : h(y) = h(x)\}$   
"collisions"  
Worst-case lookup time =  $\max_{x \in U} l(x)$
- Time to lookup  $x \in U$  is  $O(l(x))$

# Application: Hash Tables

How should we choose  
the hash function

$h: \mathcal{U} \rightarrow \{0, 1, \dots, n-1\}$  to have  
small maximum load?



Looks a lot like balls in bins!

Deterministic hash function  $h$ ?

Suppose  $|\mathcal{U}| \geq nm$  then for every  $h: \mathcal{U} \rightarrow \{0, 1, \dots, n-1\}$

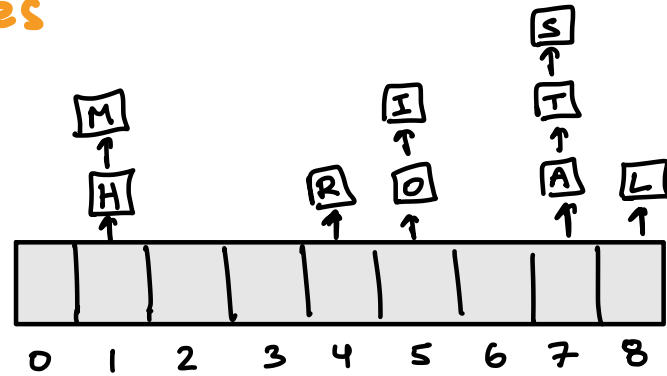
$\exists S \subseteq \mathcal{U}$  of size  $m$  such that

$h(x) = h(y)$  for every  $x, y \in S$

# Application: Hash Tables

How should we choose  
the hash function

$h: \mathcal{U} \rightarrow \{0, 1, \dots, n-1\}$  to have  
small maximum load?



Looks a lot like balls in bins!

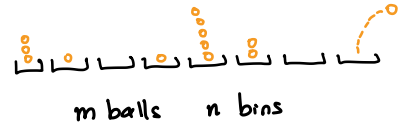
Randomized hash function:

- Model  $h$  as a uniformly random function  $\mathcal{U} \rightarrow \{0, 1, \dots, n-1\}$
- Fix the set  $S$  and study  $\mathbb{E}(\max_x l(x))$

$h(x_1), h(x_2), \dots, h(x_m)$   
uniformly random

$\nwarrow$  expectation over random choice of  $h$

# Balls and Bins: Maximum Load



- Let  $L_i$  be the number of balls in bin  $i$

- Expected maximum load is  $\mathbb{E}(\max_i L_i) = \sum_{k=1}^{\infty} \underbrace{\mathbb{P}(\max_i L_i \geq k)}_{\text{want to bound this probability}}$

# Balls and Bins: Maximum Load



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

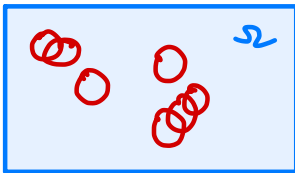
## Step 1: "Union Bound"

Specifically  $\mathbb{P}(\max_i L_i \geq k) \leq n \cdot \mathbb{P}(L_1 \geq k)$

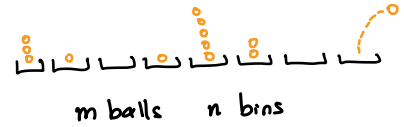
In general:

If  $E_1, \dots, E_n$  are events then  $\mathbb{P}(E_1 \cup E_2 \cup \dots \cup E_n) \leq \mathbb{P}(E_1) + \mathbb{P}(E_2) + \dots + \mathbb{P}(E_n)$

If  $\mathbb{P}(E_i) \leq \frac{1}{10n}$  then  $\mathbb{P}(E_1 \cup \dots \cup E_n) \leq \frac{1}{10}$



# Balls and Bins: Maximum Load



- Let  $L_i$  be the number of balls in bin  $i$

- Expected maximum load is  $\mathbb{E}(\max_i L_i) = \sum_{k=1}^{\infty} \mathbb{P}(\max_i L_i \geq k)$   
 $\leq \sum_{k=1}^{\infty} n \cdot \mathbb{P}(L_1 \geq k)$   
want to bound this

$$\begin{aligned} & \mathbb{P}(\max_i L_i \geq k) \\ &= \mathbb{P}(L_1 \geq k \cup L_2 \geq k \cup \dots \cup L_n \geq k) \\ &\leq \mathbb{P}(L_1 \geq k) + \mathbb{P}(L_2 \geq k) + \dots + \mathbb{P}(L_n \geq k) \\ &= n \cdot \mathbb{P}(L_1 \geq k) \end{aligned}$$

Union Bound

# Markov's Inequality

Thm: For any non-negative random variable  $X$  and every  $k$ ,  $P(X \geq k) \leq \frac{E(X)}{k}$

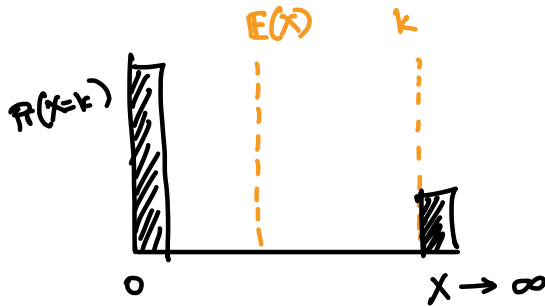
Proof:

$$E(X) = E(X \cdot \mathbb{1}_{\{X < k\}}) + E(X \cdot \mathbb{1}_{\{X \geq k\}})$$

$$\leq E(X \cdot \mathbb{1}_{\{X \geq k\}})$$

$$\leq E(k \cdot \mathbb{1}_{\{X \geq k\}})$$

$$= k \cdot P(X \geq k)$$



# Balls and Bins: Maximum Load



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

$$\begin{aligned}
 \mathbb{P}(\max_i L_i \geq k) &\leq n \cdot \mathbb{P}(L_1 \geq k) \\
 &\leq \frac{n \cdot \mathbb{E}(L_1)}{k} \quad \leftarrow \text{Markov} \\
 &= \frac{n \cdot \left(\frac{m}{n}\right)}{k} = \frac{m}{k} \quad \leftarrow \mathbb{E}(L_1) = \frac{m}{n}
 \end{aligned}$$

$$\sum_{k=1}^{\infty} \frac{1}{k^3} = m \cdot \sum_{k=1}^{\infty} \frac{1}{k} = \infty$$

$$\mathbb{E}(\max_i L_i) \leq \infty$$

# Chebyshev's Inequality

Thm: For any random variable  $X$  with  $\mu = \mathbb{E}(X)$   
and every  $t$ ,  $\mathbb{P}(|X - \mu| \geq t) \leq \frac{\mathbb{E}((X - \mu)^2)}{t^2}$

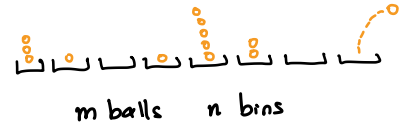
$$\mathbb{E}((X - \mu)^2) = \text{Var}(X)$$

Proof:

$$\mathbb{P}(|X - \mu| \geq t) = \mathbb{P}((X - \mu)^2 \geq t^2) \leq \frac{\mathbb{E}((X - \mu)^2)}{t^2}$$

↑  
Markov

# Balls and Bins: Maximum Load



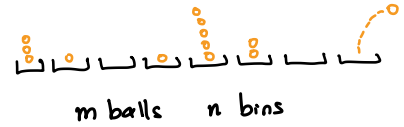
- Let  $L_i$  be the number of balls in bin  $i$
- Expected maximum load is  $\mathbb{E}(\max_i L_i) = \sum_{k=1}^{\infty} \underbrace{\mathbb{P}(\max_i L_i \geq k)}_{\text{want to bound this probability}}$

$$\begin{aligned}\mathbb{P}(\max_i L_i \geq k) &\leq n \cdot \mathbb{P}(L_1 \geq k) \\ &= n \cdot \mathbb{P}(L_1 - \mu \geq k - \mu) \\ &\leq n \cdot \mathbb{P}(|L_1 - \mu| \geq k - \mu)\end{aligned}$$

Chebyshev  $\rightarrow$   $\leq \frac{n \cdot \text{Var}(L_1)}{(k - \mu)^2}$

$$\mathbb{E}(L_1) = \frac{m}{n}$$

# Balls and Bins: Maximum Load



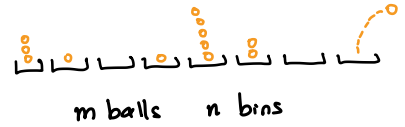
- Let  $L_i$  be the number of balls in bin  $i$

- Expected maximum load is  $\mathbb{E}(\max_i L_i) = \sum_{k=1}^{\infty} \underbrace{\mathbb{P}(\max_i L_i \geq k)}_{\text{want to bound this probability}}$

$$\text{Var}(L_i) = ???$$

$$\mathbb{E}\left(\left(L_i - \frac{m}{n}\right)^2\right) = \underbrace{\mathbb{E}(L_i^2)} - \frac{m^2}{n^2}$$

# 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  $L_{i,j} = \begin{cases} 1 & \text{if ball } j \text{ is in bin } i \\ 0 & \text{otherwise} \end{cases}$

$$\Rightarrow L_i = L_{i,1} + \dots + L_{i,m}$$

$$\text{Var}(L_i) = \mathbb{E}(L_i^2) - \frac{m^2}{n^2}$$

$$= \mathbb{E}((L_{i,1} + \dots + L_{i,m})^2) - \frac{m^2}{n^2}$$

$$= \mathbb{E}\left(\sum_{i,j} L_{i,i} L_{i,j}\right) - \frac{m^2}{n^2}$$

$$= \sum_{i,j} \mathbb{E}(L_{i,i} L_{i,j}) - \frac{m^2}{n^2} = \frac{m}{n} + \frac{m(m-1)}{n^2} - \frac{m^2}{n^2} = \frac{m}{n} - \frac{m}{n^2} \leq \frac{m}{n}$$

How can we reason  
about sums of independent  
random variables?

# Balls and Bins: Maximum Load



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

$$\mathbb{P}(\max_i L_i \geq k) \leq \frac{n \cdot \text{Var}(L_1)}{(k - \mu)^2} \leq \frac{n \cdot \frac{m}{n}}{(k - \mu)^2} = \frac{m}{(k - \mu)^2} = \frac{m}{(k - \frac{m}{n})^2}$$

$\mathbb{E}(L_1) = \frac{m}{n}$

(Assume  $\frac{m}{n} = 1$ )

$$\sum_{k=1}^{\infty} \mathbb{P}(\max_i L_i \geq k) \leq \sum_{k=1}^{\infty} \min\left\{1, \frac{m}{(k-1)^2}\right\}$$

$$= \sum_{k=1}^{\sqrt{m}} 1 + \sum_{k=\sqrt{m}+1}^{\infty} \frac{m}{(k-1)^2} = \sqrt{m} + m \cdot \sum_{k=\sqrt{m}+1}^{\infty} \frac{1}{(k-1)^2} \lesssim \sqrt{m} + m/\sqrt{m} = 2\sqrt{m}$$