

# 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

## 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

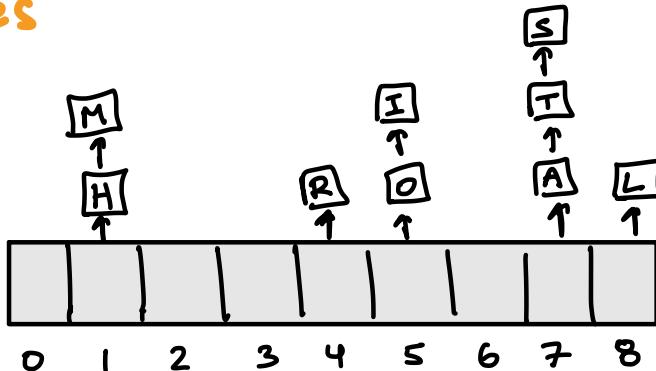
Goal: Store a set of  $m$  elements  $S \subseteq \mathcal{U}$ ,  
such that we can efficiently check if  $x \in S$

↳ 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



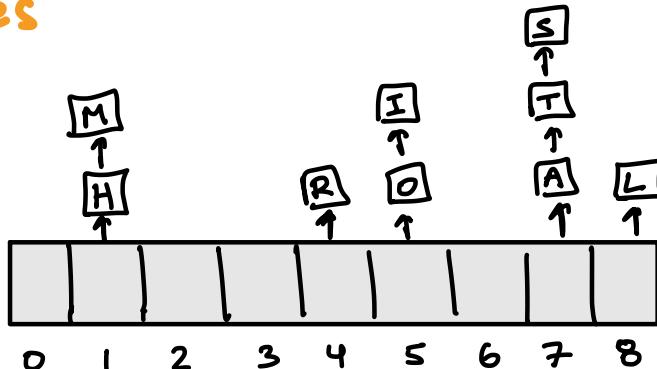
Looks a lot like balls in bins!

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

# Application: Hash Tables

How should we choose  
the hash function

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



Looks a lot like balls in bins!

Deterministic hash function  $h$ ?

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

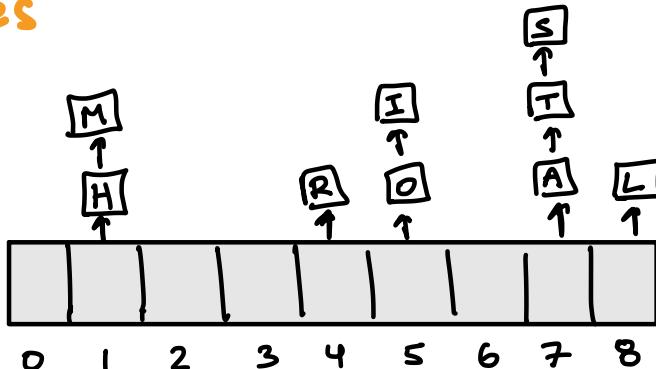
$\exists S \subseteq U$  of size  $m$  such that

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

# Application: Hash Tables

How should we choose  
the hash function

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



Looks a lot like balls in bins!

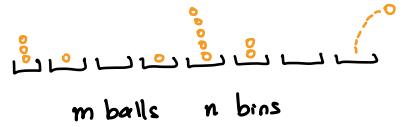
Randomized hash function:

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

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

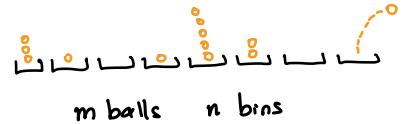
$\mathbb{E}(\max_x l(x))$   
expectation over random choice of  $h$

# Balls and Bins: Maximum Load



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

# Balls and Bins: Maximum Load



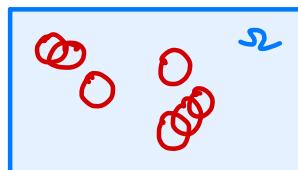
- Let  $L_i$  be the number of balls in bin  $i$
- Expected maximum load is  $\mathbb{E}(\max_i L_i) = \sum_{k=1}^{\infty} \underbrace{\mathbb{P}(\max_i L_i \geq k)}_{\text{want to bound this probability}}$

## Step 1: "Union Bound"

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

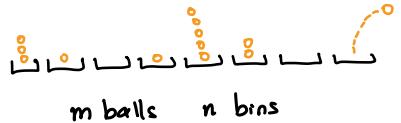
In general:

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  $\mathbb{P}(E_i) \leq \frac{1}{10^n}$  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  $\mathbb{E}(\max_i L_i) = \sum_{k=1}^{\infty} P(\max_i L_i \geq k) \leq \sum_{k=1}^{\infty} n \cdot \underbrace{P(L_1 \geq k)}_{\text{want to bound th.s}}$

$$P(\max_i L_i > k)$$

$$= P((L_1 \geq k) \cup (L_2 \geq k) \dots \cup (L_n \geq k))$$

$$\leq P(L_1 \geq k) + P(L_2 \geq k) + \dots + P(L_n \geq k)$$

$$= n \cdot P(L_1 \geq k)$$

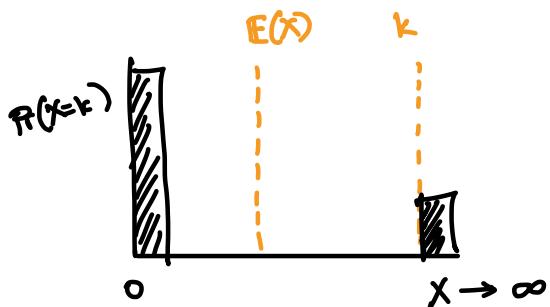
Union Bound

# Markov's Inequality

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

Proof:

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

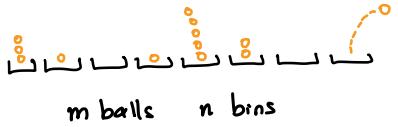


$$\leq \mathbb{E}(X \cdot \mathbb{1}_{\{X \geq k\}})$$

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

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

# Balls and Bins: Maximum Load



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

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

$$\sum_{k=1}^{\infty} \frac{m}{k} = m \cdot \sum_{k=1}^{\infty} \frac{1}{k} = \infty \quad \mathbb{E}(\max_i L_i) \leq \infty$$

# Chebychev's Inequality

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

Proof:

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

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

↑  
Markov

# Balls and Bins: Maximum Load



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

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

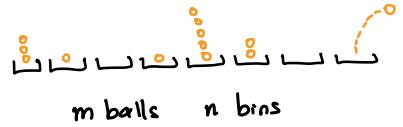
$$\mathbb{E}(L_1) = \frac{m}{n}$$

$$= n \cdot \mathbb{P}(L_1 - \mu > k - \mu)$$

$$\leq n \cdot \mathbb{P}(|L_1 - \mu| > k - \mu)$$

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

# Balls and Bins: Maximum Load

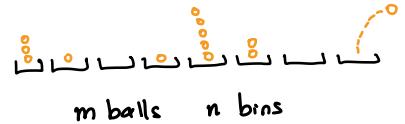


- Let  $L_i$  be the number of balls in bin  $i$
- Expected maximum load is  $\mathbb{E}(\max_i L_i) = \sum_{k=1}^{\infty} \underbrace{\mathbb{P}(\max_i L_i \geq k)}$

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

$$\mathbb{E}((L_i - \frac{m}{n})^2) = \underbrace{\mathbb{E}(L_i^2)}_{\text{Want to bound this probability}} - \frac{m^2}{n^2}$$

# Balls and Bins: Maximum Load



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

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

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

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

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

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

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

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

$$\begin{aligned} \sum_{k=1}^{\infty} \mathbb{P}(\max_i L_i \geq k) &\leq \sum_{k=1}^{\infty} m \cdot \mathbb{E}\left[1_{\{L_i \geq (k-\frac{m}{n})^2\}}\right] \\ &= \sum_{k=1}^{\sqrt{m}} 1 + \sum_{k=\lceil \sqrt{m} \rceil + 1}^{\infty} \frac{m}{(k-\frac{m}{n})^2} = \sqrt{m} + m \cdot \sum_{k=\lceil \sqrt{m} \rceil + 1}^{\infty} \frac{1}{(k-1)^2} \stackrel{?}{\leq} \sqrt{m} + m/\sqrt{m} \\ &= 2\sqrt{m} \end{aligned}$$