

COMPSCI 514: Algorithms for Data Science

Cameron Musco

University of Massachusetts Amherst. Fall 2026.

Lecture 4

In-Class Quiz 2

T H T H H H H T H

T H T H H

Question 1: Suppose we flip $n \geq 4$ independent fair coins – each shows H with probability $1/2$ and T with probability $1/2$. A **streak** is a set of 4 coins in a row that all show the same value. Different streaks may overlap. Let S be the number of streaks.

$\frac{1}{8}$
 $(n-3)/8$
:

What is $\mathbb{E}[S]$ as a function of n ? **Hint:** Use linearity of expectation.

$S_i = 1$ if streak starting at pos. i .

$$\frac{\binom{n}{4}}{2 \cdot 16}$$

$$\mathbb{E}S = \sum_{i=1}^{n-3} \mathbb{E}S_i = \sum_{i=1}^{n-3} \frac{1}{8} = \frac{(n-3)}{8}$$

Question 2: Suppose you have a randomized algorithm that sorts a list of n numbers, with expected running time $T(n)$ on any input. Give an upper bound on the probability that the algorithm runs in time $> 4T(n)$.

$$\mathbb{E}[X] = T(n)$$

$$P_r(X \geq 4T(n)) \leq \frac{\mathbb{E}[X]}{4T(n)} = \frac{T(n)}{4T(n)} = \frac{1}{4}$$

- Problem Set 1 posted and due next Friday 10/2, at 11:59pm.

$$\begin{aligned} \mathbb{E} S_i &= \Pr(S_i = 1) = \Pr(\text{4 flips starting at pos. } i \text{ are streak}) \\ &= \frac{|\{HHHH, TTTT\}|}{16} \\ &= \frac{1}{8} \end{aligned}$$

Last Time

Last Class:

- 2-level hashing and its analysis via linearity of expectation.
Gives optimal $O(1)$ query time and $O(m)$ expected space usage.

- Practical random hash functions: 2-universal and pairwise independent hashing.

$$h(x) = a \cdot x + b \bmod P$$

- Start on hashing for load balancing and Chebyshev's inequality.

random seed

Last Time

Last Class:

- 2-level hashing and its analysis via linearity of expectation. Gives optimal $O(1)$ query time and $O(m)$ expected space usage.
- Practical random hash functions: 2-universal and pairwise independent hashing.
- Start on hashing for load balancing and Chebyshev's inequality.

This Time:

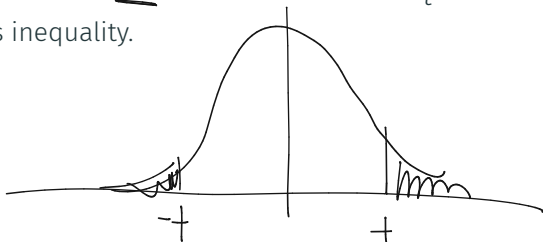
- Finish up Chebyshev's inequality and connection to the **law of large numbers**.
- Exponential tail bounds, and their connections to the law of **central limit theorem**.
- The union bound to bound the probability that one of multiple possible correlated events happens.

Chebyshev's inequality

For any random variable X and any value $t > 0$:

$$\Pr(|X| \geq t) = \Pr(X^2 \geq t^2) \leq \frac{\mathbb{E}[X^2]}{t^2}$$

by Markov's inequality.



Chebyshev's inequality

For any random variable X and any value $t > 0$:

$$\Pr(|X| \geq t) = \Pr(X^2 \geq t^2) \leq \frac{\mathbb{E}[X^2]}{t^2}$$

by Markov's inequality.

Plugging in the random variable $X - \mathbb{E}[X]$ gives: *Variance*

$$\Pr(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{t^2} = \frac{\text{Var}[X]}{t^2}$$

Chebyshev's inequality

For any random variable X and any value $t > 0$:

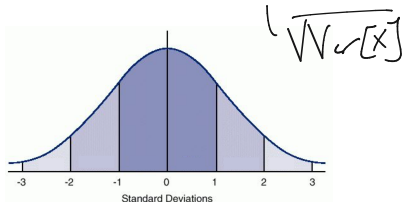
$$\Pr(|X| \geq t) = \Pr(X^2 \geq t^2) \leq \frac{\mathbb{E}[X^2]}{t^2}$$

by Markov's inequality.

Plugging in the random variable $X - \mathbb{E}[X]$ gives:

$$\Pr(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{t^2} = \frac{\text{Var}[X]}{t^2}$$

What is the probability that X falls s standard deviations from its mean?



Chebyshev's inequality

For any random variable X and any value $t > 0$:

$$\Pr(|X| \geq t) = \Pr(X^2 \geq t^2) \leq \frac{\mathbb{E}[X^2]}{t^2}$$

by Markov's inequality.

Plugging in the random variable $X - \mathbb{E}[X]$ gives:

$$\Pr(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{t^2} = \frac{\text{Var}[X]}{t^2}$$

What is the probability that X falls s standard deviations from its mean?

$$\Pr(|X - \mathbb{E}[X]| \geq s \cdot \sqrt{\text{Var}[X]}) \leq \frac{\text{Var}[X]}{s^2 \cdot \text{Var}[X]} = \frac{1}{s^2}.$$

Law of Large Numbers

Consider drawing independent identically distributed (i.i.d.) random variables $X_1, \dots, X_n$ with mean μ and variance σ^2 .

Law of Large Numbers

Consider drawing independent identically distributed (i.i.d.) random variables $X_1, \dots, X_n$ with mean μ and variance σ^2 .

How well does the sample average $S = \frac{1}{n} \sum_{i=1}^n X_i$ approximate the true mean μ ?

$$\mathbb{E}S = \mu$$

$$|S - \mu| = |S - \mathbb{E}S|$$

Law of Large Numbers

Consider drawing independent identically distributed (i.i.d.) random variables $X_1, \dots, X_n$ with mean μ and variance σ^2 .

How well does the sample average $S = \frac{1}{n} \sum_{i=1}^n X_i$ approximate the true mean μ ?

$$\begin{aligned} \text{Var}[S] &= \text{Var} \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \frac{1}{n^2} \text{Var} (\sum X_i) \stackrel{\text{lin of variance}}{=} \frac{1}{n^2} \cdot \sum_{i=1}^n \text{Var}(X_i) \\ &= \frac{1}{n^2} \cdot n \cdot \sigma^2 \\ &= \frac{\sigma^2}{n} \end{aligned}$$

Law of Large Numbers

Consider drawing independent identically distributed (i.i.d.) random variables $X_1, \dots, X_n$ with mean μ and variance σ^2 .

How well does the sample average $S = \frac{1}{n} \sum_{i=1}^n X_i$ approximate the true mean μ ?

$$\text{Var}[S] = \text{Var} \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i]$$

Law of Large Numbers

Consider drawing independent identically distributed (i.i.d.) random variables $X_1, \dots, X_n$ with mean μ and variance σ^2 .

How well does the sample average $S = \frac{1}{n} \sum_{i=1}^n X_i$ approximate the true mean μ ?

$$\text{Var}[S] = \text{Var} \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] = \frac{1}{n^2} \cdot n \cdot \sigma^2$$

Law of Large Numbers

Consider drawing independent identically distributed (i.i.d.) random variables $X_1, \dots, X_n$ with mean μ and variance σ^2 .

How well does the sample average $S = \frac{1}{n} \sum_{i=1}^n X_i$ approximate the true mean μ ?

$$\text{Var}[S] = \text{Var} \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}.$$

Law of Large Numbers

Consider drawing independent identically distributed (i.i.d.) random variables $X_1, \dots, X_n$ with mean μ and variance σ^2 .

How well does the sample average $S = \frac{1}{n} \sum_{i=1}^n X_i$ approximate the true mean μ ?

$$\text{Var}[S] = \text{Var} \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}.$$

By Chebyshev's Inequality: for any fixed value $\epsilon > 0$,

$$\Pr(|S - \mathbb{E}[S]| \geq \epsilon) \leq \frac{\text{Var}[S]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}.$$

Law of Large Numbers

Consider drawing independent identically distributed (i.i.d.) random variables $X_1, \dots, X_n$ with mean μ and variance σ^2 .

How well does the sample average $S = \frac{1}{n} \sum_{i=1}^n X_i$ approximate the true mean μ ?

$$\text{Var}[S] = \text{Var} \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}.$$

By Chebyshev's Inequality: for any fixed value $\epsilon > 0$,

$$\Pr(|S - \mu| \geq \epsilon) \leq \frac{\text{Var}[S]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}.$$

Law of Large Numbers

Consider drawing independent identically distributed (i.i.d.) random variables $X_1, \dots, X_n$ with mean μ and variance σ^2 .

How well does the sample average $S = \frac{1}{n} \sum_{i=1}^n X_i$ approximate the true mean μ ?

$$\text{Var}[S] = \text{Var} \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}.$$

By Chebyshev's Inequality: for any fixed value $\epsilon > 0$,

$$\Pr(|S - \mu| \geq \epsilon) \leq \frac{\text{Var}[S]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}.$$

$n \rightarrow \infty$
 $\Pr(|S - \mu| \geq \epsilon) \rightarrow 0$

Law of Large Numbers: with enough samples n , the sample average will always concentrate to the mean.

Law of Large Numbers

Consider drawing independent identically distributed (i.i.d.) random variables $X_1, \dots, X_n$ with mean μ and variance σ^2 .

How well does the sample average $S = \frac{1}{n} \sum_{i=1}^n X_i$ approximate the true mean μ ?

$$\text{Var}[S] = \text{Var} \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}.$$

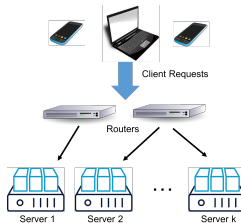
By Chebyshev's Inequality: for any fixed value $\epsilon > 0$,

$$\Pr(|S - \mu| \geq \epsilon) \leq \frac{\text{Var}[S]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}.$$

Law of Large Numbers: with enough samples n , the sample average will always concentrate to the mean.

- Cannot show from vanilla Markov's inequality.

Load Balancing Application



$R_i = \# \text{ request routed to server } i$

Simple Model: n requests randomly assigned to k servers. How many requests must each server handle?

- $\mathbb{E}[R_i] = n/k$.
- If we provision each server to handle $2n/k$ requests, Markov's inequality just gives that

$$\Pr[\text{overload}] = \Pr[R_i > 2n/k] \leq \frac{\mathbb{E}[R_i]}{2n/k} = \frac{1}{2}.$$

- Want to prove something stronger.

Load Balancing Variance

We can write the number of requests assigned to server i , R_i as:

$$R_i = \sum_{j=1}^n R_{i,j}$$

$j = 1, \dots, k$ (servers)
 $j = 1, \dots, n$ (requests)

where $R_{i,j}$ is 1 if request j is assigned to server i and 0 otherwise.

n : total number of requests, k : number of servers randomly assigned requests,
 R_i : number of requests assigned to server i .

Load Balancing Variance

We can write the number of requests assigned to server i , R_i as:

$$\text{Var}[R_i] = \sum_{j=1}^n \text{Var}[R_{i,j}] \quad (\text{linearity of variance})$$

where $R_{i,j}$ is 1 if request j is assigned to server i and 0 otherwise.

$$\begin{aligned} \text{Var}[R_{i,j}] &= \mathbb{E}[R_{i,j}^2] - \mathbb{E}[R_{i,j}]^2 \\ &= \frac{1}{k} - \frac{1}{k^2} = \frac{1}{k} \left(1 - \frac{1}{k}\right) \\ &\quad p \cdot (1-p) \end{aligned}$$

n : total number of requests, k : number of servers randomly assigned requests,
 R_i : number of requests assigned to server i .

Load Balancing Variance

We can write the number of requests assigned to server i , R_i as:

$$\text{Var}[R_i] = \sum_{j=1}^n \text{Var}[R_{i,j}] \quad (\text{linearity of variance})$$

where $R_{i,j}$ is 1 if request j is assigned to server i and 0 otherwise.

$$\text{Var}[R_{i,j}] = \mathbb{E}[R_{i,j}^2] - \mathbb{E}[R_{i,j}]^2$$

n : total number of requests, k : number of servers randomly assigned requests,
 R_i : number of requests assigned to server i .

Load Balancing Variance

We can write the number of requests assigned to server i , R_i as:

$$\text{Var}[R_i] = \sum_{j=1}^n \text{Var}[R_{i,j}] \quad (\text{linearity of variance})$$

where $R_{i,j}$ is 1 if request j is assigned to server i and 0 otherwise.

$$\begin{aligned} \text{Var}[R_{i,j}] &= \mathbb{E}[R_{i,j}^2] - \mathbb{E}[R_{i,j}]^2 \\ &= \mathbb{E}[R_{i,j}] - \mathbb{E}[R_{i,j}]^2 \end{aligned}$$

n : total number of requests, k : number of servers randomly assigned requests,
 R_i : number of requests assigned to server i .

Load Balancing Variance

We can write the number of requests assigned to server i , R_i as:

$$\text{Var}[R_i] = \sum_{j=1}^n \text{Var}[R_{i,j}] \quad (\text{linearity of variance})$$

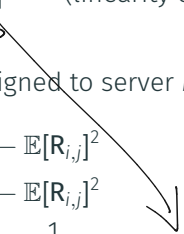
where $R_{i,j}$ is 1 if request j is assigned to server i and 0 otherwise.

$$\begin{aligned}\text{Var}[R_{i,j}] &= \mathbb{E}[R_{i,j}^2] - \mathbb{E}[R_{i,j}]^2 \\ &= \mathbb{E}[R_{i,j}] - \mathbb{E}[R_{i,j}]^2 \\ &= \frac{1}{k} - \frac{1}{k^2} < \frac{1}{k}\end{aligned}$$

n : total number of requests, k : number of servers randomly assigned requests,
 R_i : number of requests assigned to server i .

Load Balancing Variance

We can write the number of requests assigned to server i , R_i as:

$$\text{Var}[R_i] = \sum_{j=1}^n \text{Var}[R_{i,j}] \quad (\text{linearity of variance})$$


where $R_{i,j}$ is 1 if request j is assigned to server i and 0 otherwise.

$$\begin{aligned} \text{Var}[R_{i,j}] &= \mathbb{E}[R_{i,j}^2] - \mathbb{E}[R_{i,j}]^2 \\ &= \mathbb{E}[R_{i,j}] - \mathbb{E}[R_{i,j}]^2 \\ &= \frac{1}{k} - \frac{1}{k^2} < \frac{1}{k} \implies \text{Var}[R_i] \leq \frac{n}{k}. \end{aligned}$$

n : total number of requests, k : number of servers randomly assigned requests,
 R_i : number of requests assigned to server i .

Bounding the Load via Chebyshevs

Letting R_i be the number of requests sent to server i , $\mathbb{E}[R_i] = \frac{n}{k}$ and $\text{Var}[R_i] < \frac{n}{k}$.

n : total number of requests, k : number of servers randomly assigned requests,
 R_i : number of requests assigned to server i .

Bounding the Load via Chebyshev's

Letting R_i be the number of requests sent to server i , $\mathbb{E}[R_i] = \frac{n}{k}$ and $\text{Var}[R_i] < \frac{n}{k}$.

Applying Chebyshev's: A request server can handle n/k

$$\Pr\left(R_i \geq \frac{2n}{k}\right) \leq \Pr\left(|R_i - \mathbb{E}[R_i]| \geq \frac{n}{k}\right)$$

server i is overloaded

$$\begin{aligned} & \Pr\left(R_i \geq \frac{2n}{k}\right) \\ &= \Pr\left(R_i - \mathbb{E}[R_i] \geq \frac{2n}{k} - \mathbb{E}[R_i]\right) = \Pr\left(R_i - \mathbb{E}[R_i] \geq \frac{n}{k}\right) \end{aligned}$$

n : total number of requests, k : number of servers randomly assigned requests,
 R_i : number of requests assigned to server i .

Bounding the Load via Chebyshev's

Letting R_i be the number of requests sent to server i , $\mathbb{E}[R_i] = \frac{n}{k}$ and $\text{Var}[R_i] < \frac{n}{k}$.

Applying Chebyshev's:

$$\Pr\left(R_i \geq \frac{2n}{k}\right) \leq \Pr\left(|R_i - \mathbb{E}[R_i]| \geq \frac{n}{k}\right) \leq \frac{n/k}{n^2/k^2}$$

$$\stackrel{\text{Chebyshev}}{\leq} \frac{\text{Var}(R_i)}{(\frac{n}{k})^2} = \frac{n/k}{n^2/k^2} = \frac{k}{n}$$

n : total number of requests, k : number of servers randomly assigned requests,
 R_i : number of requests assigned to server i .

Bounding the Load via Chebyshev's

Letting R_i be the number of requests sent to server i , $\mathbb{E}[R_i] = \frac{n}{k}$ and $\text{Var}[R_i] < \frac{n}{k}$.

Applying Chebyshev's:

$$\Pr\left(R_i \geq \frac{2n}{k}\right) \leq \Pr\left(|R_i - \mathbb{E}[R_i]| \geq \frac{n}{k}\right) \leq \frac{n/k}{n^2/k^2} = \frac{k}{n}.$$

n : total number of requests, k : number of servers randomly assigned requests,
 R_i : number of requests assigned to server i .

Bounding the Load via Chebyshev's

Letting R_i be the number of requests sent to server i , $\mathbb{E}[R_i] = \frac{n}{k}$ and $\text{Var}[R_i] < \frac{n}{k}$.

Applying Chebyshev's:

$$\Pr\left(R_i \geq \frac{2n}{k}\right) \leq \Pr\left(|R_i - \mathbb{E}[R_i]| \geq \frac{n}{k}\right) \leq \frac{n/k}{n^2/k^2} = \frac{k}{n}.$$

- Overload probability is very small in the typical case that $k \ll n$.

n : total number of requests, k : number of servers randomly assigned requests,
 R_i : number of requests assigned to server i .

Bounding the Load via Chebyshev's

Letting R_i be the number of requests sent to server i , $\mathbb{E}[R_i] = \frac{n}{k}$ and $\text{Var}[R_i] < \frac{n}{k}$.

Applying Chebyshev's:

$$\Pr\left(R_i \geq \frac{2n}{k}\right) \leq \Pr\left(|R_i - \mathbb{E}[R_i]| \geq \frac{n}{k}\right) \leq \frac{n/k}{n^2/k^2} = \frac{k}{n}.$$

- Overload probability is very small in the typical case that $k \ll n$.
- Might seem counterintuitive – bound gets worse as k grows.
- When k is large, the number of requests each server sees in expectation is very small so the law of large numbers doesn't 'kick in'.

n : total number of requests, k : number of servers randomly assigned requests,
 R_i : number of requests assigned to server i .

Maximum Server Load

What is the probability that the **maximum server load** exceeds $2 \cdot \mathbb{E}[\mathbf{R}_i] = \frac{2n}{k}$. I.e., that some server is overloaded if we give each $\frac{2n}{k}$ capacity?

n : total number of requests, k : number of servers randomly assigned requests,
 $\mathbf{R}_i$: number of requests assigned to server i . $\mathbb{E}[\mathbf{R}_i] = \frac{n}{k}$. $\text{Var}[\mathbf{R}_i] = \frac{n}{k}$.

Maximum Server Load

What is the probability that the **maximum server load** exceeds $2 \cdot \mathbb{E}[\mathbf{R}_i] = \frac{2n}{k}$. I.e., that some server is overloaded if we give each $\frac{2n}{k}$ capacity?

$$\Pr \left(\max_{i=1}^k (\mathbf{R}_i) \geq \frac{2n}{k} \right)$$

n : total number of requests, k : number of servers randomly assigned requests,
 $\mathbf{R}_i$: number of requests assigned to server i . $\mathbb{E}[\mathbf{R}_i] = \frac{n}{k}$. $\text{Var}[\mathbf{R}_i] = \frac{n}{k}$.

Maximum Server Load

What is the probability that the **maximum server load** exceeds $2 \cdot \mathbb{E}[\mathbf{R}_i] = \frac{2n}{k}$. I.e., that some server is overloaded if we give each $\frac{2n}{k}$ capacity?

$$\Pr \left(\max_i (\mathbf{R}_i) \geq \frac{2n}{k} \right) = \Pr \left(\left[\mathbf{R}_1 \geq \frac{2n}{k} \right] \cup \left[\mathbf{R}_2 \geq \frac{2n}{k} \right] \cup \dots \cup \left[\mathbf{R}_k \geq \frac{2n}{k} \right] \right)$$

n : total number of requests, k : number of servers randomly assigned requests,
 $\mathbf{R}_i$: number of requests assigned to server i . $\mathbb{E}[\mathbf{R}_i] = \frac{n}{k}$. $\text{Var}[\mathbf{R}_i] = \frac{n}{k}$.

Maximum Server Load

What is the probability that the **maximum server load** exceeds $2 \cdot \mathbb{E}[\mathbf{R}_i] = \frac{2n}{k}$. I.e., that some server is overloaded if we give each $\frac{2n}{k}$ capacity?

$$\Pr \left(\max_i (\mathbf{R}_i) \geq \frac{2n}{k} \right) = \Pr \left(\left[\mathbf{R}_1 \geq \frac{2n}{k} \right] \text{ or } \left[\mathbf{R}_2 \geq \frac{2n}{k} \right] \text{ or } \dots \text{ or } \left[\mathbf{R}_k \geq \frac{2n}{k} \right] \right)$$

n : total number of requests, k : number of servers randomly assigned requests,
 $\mathbf{R}_i$: number of requests assigned to server i . $\mathbb{E}[\mathbf{R}_i] = \frac{n}{k}$. $\text{Var}[\mathbf{R}_i] = \frac{n}{k}$.

Maximum Server Load

What is the probability that the **maximum server load** exceeds $2 \cdot \mathbb{E}[\mathbf{R}_i] = \frac{2n}{k}$. I.e., that some server is overloaded if we give each $\frac{2n}{k}$ capacity?

$$\Pr \left(\max_i (\mathbf{R}_i) \geq \frac{2n}{k} \right) = \Pr \left(\bigcup_{i=1}^k \left[\mathbf{R}_i \geq \frac{2n}{k} \right] \right)$$

↑
not independent

n : total number of requests, k : number of servers randomly assigned requests,
 $\mathbf{R}_i$: number of requests assigned to server i . $\mathbb{E}[\mathbf{R}_i] = \frac{n}{k}$. $\text{Var}[\mathbf{R}_i] = \frac{n}{k}$.

Maximum Server Load

What is the probability that the **maximum server load** exceeds $2 \cdot \mathbb{E}[\mathbf{R}_i] = \frac{2n}{k}$. I.e., that some server is overloaded if we give each $\frac{2n}{k}$ capacity?

$$\Pr \left(\max_i (\mathbf{R}_i) \geq \frac{2n}{k} \right) = \Pr \left(\bigcup_{i=1}^k \left[\mathbf{R}_i \geq \frac{2n}{k} \right] \right)$$

We want to show that $\Pr \left(\bigcup_{i=1}^k \left[\mathbf{R}_i \geq \frac{2n}{k} \right] \right)$ is small.

n : total number of requests, k : number of servers randomly assigned requests,
 $\mathbf{R}_i$: number of requests assigned to server i . $\mathbb{E}[\mathbf{R}_i] = \frac{n}{k}$. $\text{Var}[\mathbf{R}_i] = \frac{n}{k}$.

Maximum Server Load

What is the probability that the **maximum server load** exceeds $2 \cdot \mathbb{E}[\mathbf{R}_i] = \frac{2n}{k}$. I.e., that some server is overloaded if we give each $\frac{2n}{k}$ capacity?

$$\Pr \left(\max_i (\mathbf{R}_i) \geq \frac{2n}{k} \right) = \Pr \left(\bigcup_{i=1}^k \left[\mathbf{R}_i \geq \frac{2n}{k} \right] \right)$$

We want to show that $\Pr \left(\bigcup_{i=1}^k \left[\mathbf{R}_i \geq \frac{2n}{k} \right] \right)$ is small.

How do we do this? Note that $\mathbf{R}_1, \dots, \mathbf{R}_k$ are correlated in a somewhat complex way.

n : total number of requests, k : number of servers randomly assigned requests,
 $\mathbf{R}_i$: number of requests assigned to server i . $\mathbb{E}[\mathbf{R}_i] = \frac{n}{k}$. $\text{Var}[\mathbf{R}_i] = \frac{n}{k}$.

The Union Bound

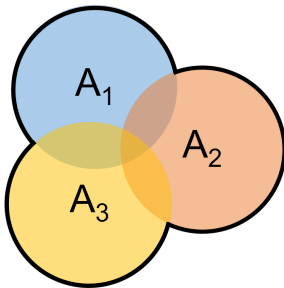
Union Bound: For any random events $A_1, A_2, \dots, A_k$,

$$\Pr(A_1 \cup A_2 \cup \dots \cup A_k) \leq \Pr(A_1) + \Pr(A_2) + \dots + \Pr(A_k).$$

The Union Bound

Union Bound: For any random events $A_1, A_2, \dots, A_k$,

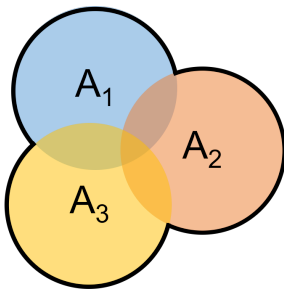
$$\Pr(A_1 \cup A_2 \cup \dots \cup A_k) \leq \Pr(A_1) + \Pr(A_2) + \dots + \Pr(A_k).$$



The Union Bound

Union Bound: For any random events $A_1, A_2, \dots, A_k$,

$$\Pr(A_1 \cup A_2 \cup \dots \cup A_k) \leq \Pr(A_1) + \Pr(A_2) + \dots + \Pr(A_k).$$

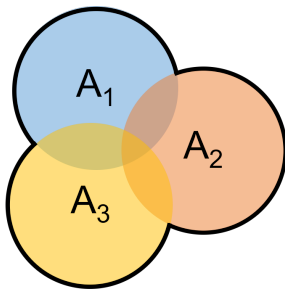


When is the union bound tight?

The Union Bound

Union Bound: For any random events $A_1, A_2, \dots, A_k$,

$$\Pr(A_1 \cup A_2 \cup \dots \cup A_k) \leq \Pr(A_1) + \Pr(A_2) + \dots + \Pr(A_k).$$



When is the union bound tight? When $A_1, \dots, A_k$ are all disjoint.

The Union Bound

Union Bound: For any random events $A_1, A_2, \dots, A_k$,

$$\Pr(A_1 \cup A_2 \cup \dots \cup A_k) \leq \Pr(A_1) + \Pr(A_2) + \dots + \Pr(A_k).$$

A and B
disjoint

$$A \Rightarrow \overline{B}$$

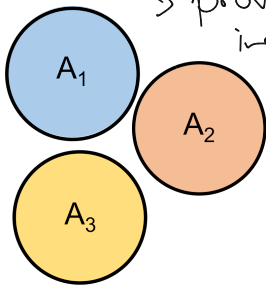
Disjoint:

$$\Pr(A \cap B) = 0$$

Independent:

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

When is the union bound tight? When $A_1, \dots, A_k$ are all disjoint.



prove via Markov's inequality on

$$\mathbb{I}_{A_1} + \mathbb{I}_{A_2} + \dots + \mathbb{I}_{A_k}$$

Applying the Union Bound

What is the probability that the **maximum server load** exceeds $2 \cdot \mathbb{E}[R_i] = \frac{2n}{k}$. I.e., that some server is overloaded if we give each $\frac{2n}{k}$ capacity?

$$\Pr\left(\max_i(R_i) \geq \frac{2n}{k}\right) = \Pr\left(\bigcup_{i=1}^k \left[R_i \geq \frac{2n}{k}\right]\right)$$

n : total number of requests, k : number of servers randomly assigned requests,
 R_i : number of requests assigned to server i . $\mathbb{E}[R_i] = \frac{n}{k}$. $\text{Var}[R_i] = \frac{n}{k}$.

Applying the Union Bound

What is the probability that the **maximum server load** exceeds $2 \cdot \mathbb{E}[R_i] = \frac{2n}{k}$. I.e., that some server is overloaded if we give each $\frac{2n}{k}$ capacity?

$$\begin{aligned}\Pr\left(\max_i(R_i) \geq \frac{2n}{k}\right) &= \Pr\left(\bigcup_{i=1}^k \left[R_i \geq \frac{2n}{k}\right]\right) \\ &\leq \sum_{i=1}^k \Pr\left(\left[R_i \geq \frac{2n}{k}\right]\right) \quad (\text{Union Bound})\end{aligned}$$

n : total number of requests, k : number of servers randomly assigned requests,
 R_i : number of requests assigned to server i . $\mathbb{E}[R_i] = \frac{n}{k}$. $\text{Var}[R_i] = \frac{n}{k}$.

Applying the Union Bound

What is the probability that the **maximum server load** exceeds $2 \cdot \mathbb{E}[R_i] = \frac{2n}{k}$. I.e., that some server is overloaded if we give each $\frac{2n}{k}$ capacity?

$$\begin{aligned}\Pr\left(\max_i(R_i) \geq \frac{2n}{k}\right) &= \Pr\left(\bigcup_{i=1}^k \left[R_i \geq \frac{2n}{k}\right]\right) \\ &\leq \sum_{i=1}^k \Pr\left(\left[R_i \geq \frac{2n}{k}\right]\right) && \text{(Union Bound)} \\ &\leq \sum_{i=1}^k \frac{k}{n} && \text{(Bound from Chebyshev's)}\end{aligned}$$

n : total number of requests, k : number of servers randomly assigned requests,
 R_i : number of requests assigned to server i . $\mathbb{E}[R_i] = \frac{n}{k}$. $\text{Var}[R_i] = \frac{n}{k}$.

Applying the Union Bound

What is the probability that the **maximum server load** exceeds $2 \cdot \mathbb{E}[R_i] = \frac{2n}{k}$. I.e., that some server is overloaded if we give each $\frac{2n}{k}$ capacity?

$$\begin{aligned}\Pr\left(\max_i(R_i) \geq \frac{2n}{k}\right) &= \Pr\left(\bigcup_{i=1}^k \left[R_i \geq \frac{2n}{k}\right]\right) \\ &\leq \sum_{i=1}^k \Pr\left(\left[R_i \geq \frac{2n}{k}\right]\right) \quad (\text{Union Bound}) \\ &\leq \sum_{i=1}^k \frac{k}{n} = \frac{k^2}{n} \quad (\text{Bound from Chebyshev's})\end{aligned}$$

n : total number of requests, k : number of servers randomly assigned requests,
 R_i : number of requests assigned to server i . $\mathbb{E}[R_i] = \frac{n}{k}$. $\text{Var}[R_i] = \frac{n}{k}$.

Applying the Union Bound

What is the probability that the **maximum server load** exceeds

$2 \cdot \mathbb{E}[R_i] = \frac{2n}{k}$. I.e., that some server is overloaded if we give each $\frac{2n}{k}$ capacity?

$k = n$

$C = 10000$

$\frac{2n}{k}$

$$\Pr\left(\max_i (R_i) \geq \frac{2n}{k}\right) = \Pr\left(\bigcup_{i=1}^k \left[R_i \geq \frac{2n}{k}\right]\right)$$

union bound

$$\leq \sum_{i=1}^k \Pr\left(\left[R_i \geq \frac{2n}{k}\right]\right)$$

$\Pr(R_i \geq \frac{2n}{k}) \leq \frac{k}{n}$
by Chebyshev's inequality

(Union Bound)

$$\leq \sum_{i=1}^k \frac{k}{n} = \frac{k^2}{n}$$

(Bound from Chebyshev's)

Again, as long as $k \ll n$, with good probability, the maximum server load will be small (compared to the expected load).

n : total number of requests, k : number of servers randomly assigned requests,
 R_i : number of requests assigned to server i . $\mathbb{E}[R_i] = \frac{n}{k}$. $\text{Var}[R_i] = \frac{n}{k}$.

Questions on union bound, Chebyshev's inequality,
random hashing?

Stronger Concentration Bounds

Flipping Coins

We flip $n = 100$ independent coins, each are heads with probability $1/2$ and tails with probability $1/2$. Let $\mathbf{H}$ be the number of heads.

Flipping Coins

We flip $n = 100$ independent coins, each are heads with probability $1/2$ and tails with probability $1/2$. Let H be the number of heads.

$$\mathbb{E}[H] = \frac{n}{2} = 50 \text{ and } \text{Var}[H] = \sum_{i=1}^{100} \text{Var}[H_i] = \sum_{i=1}^{100} \frac{1}{4} = 25$$

Flipping Coins

We flip $n = 100$ independent coins, each are heads with probability $1/2$ and tails with probability $1/2$. Let $\mathbf{H}$ be the number of heads.

$$\mathbb{E}[\mathbf{H}] = \frac{n}{2} = 50 \text{ and } \text{Var}[\mathbf{H}] = \frac{n}{4} = 25$$

Flipping Coins

We flip $n = 100$ independent coins, each are heads with probability $1/2$ and tails with probability $1/2$. Let H be the number of heads.

$$\mathbb{E}[H] = \frac{n}{2} = 50 \text{ and } \text{Var}[H] = \frac{n}{4} = 25$$

Markov's: $\leq \frac{50}{60}$

$$\Pr(H \geq 60) \leq .833$$

$$\Pr(H \geq 70) \leq .714$$

$$\Pr(H \geq 80) \leq .625$$

Flipping Coins

We flip $n = 100$ independent coins, each are heads with probability $1/2$ and tails with probability $1/2$. Let H be the number of heads.

$$\mathbb{E}[H] = \frac{n}{2} = 50 \text{ and } \text{Var}[H] = \frac{n}{4} = 25 \rightarrow s.d. = 5$$

Markov's:

$$\Pr(H \geq 60) \leq .833$$

$$\Pr(H \geq 70) \leq .714$$

$$\Pr(H \geq 80) \leq .625$$

Chebyshev's:

$$\Pr(H \geq 60) \leq .25$$

$$\Pr(H \geq 70) \leq .0625$$

$$\Pr(H \geq 80) \leq .0278$$

Flipping Coins

We flip $n = 100$ independent coins, each are heads with probability $1/2$ and tails with probability $1/2$. Let H be the number of heads.

$$\mathbb{E}[H] = \frac{n}{2} = 50 \text{ and } \text{Var}[H] = \frac{n}{4} = 25 \rightarrow \text{s.d.} = 5$$

Markov's:	Chebyshev's:	In Reality:
$\Pr(H \geq 60) \leq .833$	$\Pr(H \geq 60) \leq .25$	$\Pr(H \geq 60) = 0.0284$
$\Pr(H \geq 70) \leq .714$	$\Pr(H \geq 70) \leq .0625$	$\Pr(H \geq 70) = .000039$
$\Pr(H \geq 80) \leq .625$	$\Pr(H \geq 80) \leq .0278$	$\Pr(H \geq 80) < 10^{-9}$

H has a simple Binomial distribution, so can compute these probabilities exactly.

Tighter Concentration Bounds

To be fair.... Markov and Chebyshev's inequalities apply much more generally than to Binomial random variables like coin flips.

Tighter Concentration Bounds

To be fair.... Markov and Chebyshev's inequalities apply much more generally than to Binomial random variables like coin flips.

Can we obtain tighter concentration bounds that still apply to very general distributions?

