# COMPSCI 514: ALGORITHMS FOR DATA SCIENCE

Cameron Musco

University of Massachusetts Amherst. Spring 2020.

Lecture 4

# LOGISTICS

- Week 2 quiz will be released this afternoon and due Monday at 8pm.
- Problem Set 1 is due next Friday, 9/11 at 8pm.

# LAST TIME

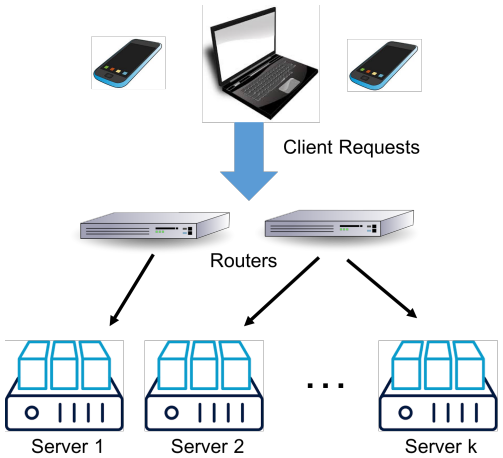
## Last Class:

- 2-Level Hashing Analysis (linearity of expectation and Markov's inequality)
- 2-universal and pairwise independent hash functions

## This Time:

- Random hashing for load balancing. Motivating:
  - Stronger concentration inequalities: Chebyshev's inequality, exponential tail bounds, and their connections to the law of large numbers and central limit theorem.
  - The union bound.

## Randomized Load Balancing:




- $n$ requests randomly assigned to $k$ servers.
- Expected load on server $i$ is $\mathbb{E}[\mathbf{R}_i] = \frac{n}{k}$.
- By Markov's inequality, if we provision each server to handle twice this expected load (so $\frac{2n}{k}$ requests), it will be overloaded with probability $\leq 1/2$.

# Chebyshev's inequality

With a very simple twist Markov's Inequality can be made much more powerful.

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

$$\Pr(|X| \geq t) = \Pr(X^2 \geq t^2).$$

$X^2$ is a nonnegative random variable. So can apply Markov's inequality:

**Chebyshev's inequality:**

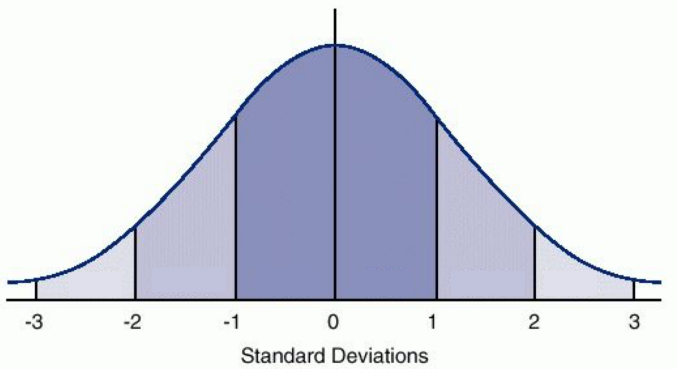
$$\Pr(|X - \mathbb{E}[X]| \geq t) = \Pr(X^2 \geq t^2) \leq \frac{\mathbb{E}[X^2]}{t^2} \frac{\mathrm{Var}[X]}{t^2}.$$

(by plugging in the random variable $X - \mathbb{E}[X]$)

# CHEBYSHEV'S INEQUALITY

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

What is the probability that $\mathbf{X}$ falls $s$ standard deviations from it's mean?




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

Why is this so powerful?

$\mathbf{X}$: any random variable, $t, s$: any fixed numbers.

Consider drawing independent identically distributed (i.i.d.) random variables $\mathsf{X}_1, \ldots, \mathsf{X}_n$ with mean $\mu$ and variance $\sigma^2$.

How well does the sample average $\mathsf{S} = \frac{1}{n}\sum_{i=1}^{n} \mathsf{X}_i$ approximate the true mean $\mu$?

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

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

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

**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 VARIANCE

We can write the number of requests assigned to server $i$, $\mathbf{R}_i$ as:

$$\mathbf{R}_i = \sum_{j=1}^{n} \mathbf{R}_{i,j} \quad \text{Var}[\mathbf{R}_i] = \sum_{j=1}^{n} \text{Var}[\mathbf{R}_{i,j}] \qquad \text{(linearity of variance)}$$

where $\mathbf{R}_{i,j}$ is 1 if request $j$ is assigned to server $i$ and 0 otherwise.

$$\begin{aligned}
\text{Var}[\mathbf{R}_{i,j}] &= \mathbb{E}\left[\left(\mathbf{R}_{i,j} - \mathbb{E}[\mathbf{R}_{i,j}]\right)^2\right] \\
&= \Pr(\mathbf{R}_{i,j} = 1) \cdot \left(1 - \mathbb{E}[\mathbf{R}_{i,j}]\right)^2 + \Pr(\mathbf{R}_{i,j} = 0) \cdot \left(0 - \mathbb{E}[\mathbf{R}_{i,j}]\right)^2 \\
&= \frac{1}{k} \cdot \left(1 - \frac{1}{k}\right)^2 + \left(1 - \frac{1}{k}\right) \cdot \left(0 - \frac{1}{k}\right)^2 \\
&= \frac{1}{k} - \frac{1}{k^2} \leq \frac{1}{k} \implies \text{Var}[\mathbf{R}_i] \leq \frac{n}{k}.
\end{aligned}$$

$n$: total number of requests, $k$: number of servers randomly assigned requests, $\mathbf{R}_i$: number of requests assigned to server $i$.

# BOUNDING THE LOAD VIA CHEBYSHEVS

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

**Applying Chebyshev's:**

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

- Overload probability is extremely small when $k \ll n$!
- 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, $\mathbf{R}_i$: 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?

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

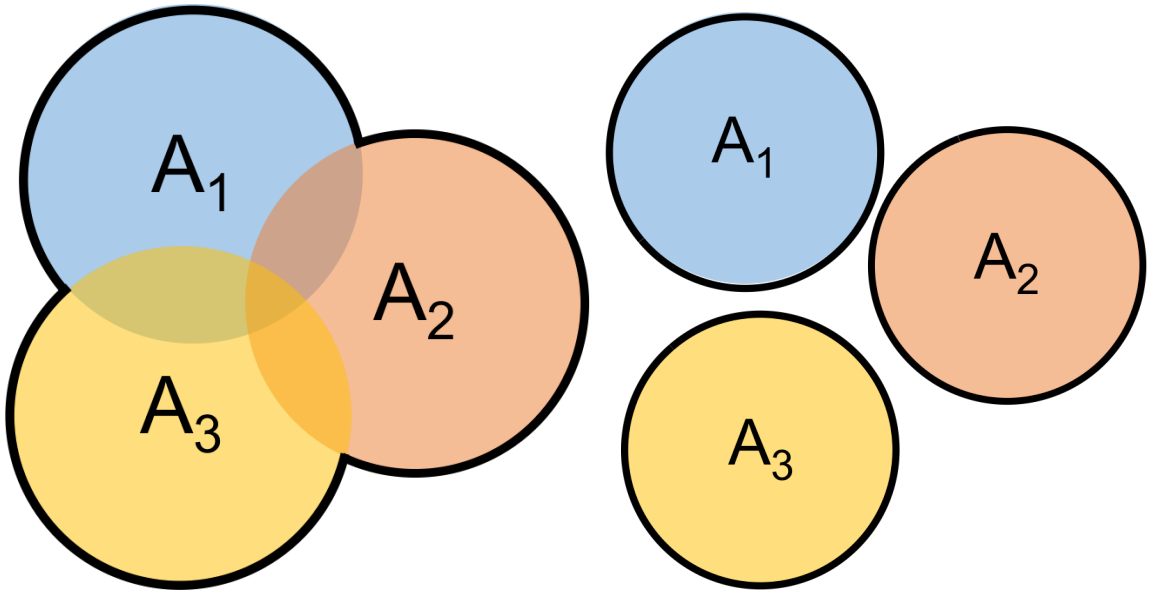
We want to show that $\Pr\left(\bigcup_{i=1}^{k}\left[\mathbf{R}_i \geq \frac{2n}{k}\right]\right)$ is small.

How do we do this? Note that $\mathbf{R}_1, \ldots, \mathbf{R}_k$ 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 $i$. $\mathbb{E}[\mathbf{R}_i] = \frac{n}{k}$. $\mathrm{Var}[\mathbf{R}_i] = \frac{n}{k}$.

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

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




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

On the first problem set, you will prove the union bound, as a consequence of Markov's inquality.

# APPLYING THE UNION BOUND

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?

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

$$\leq \sum_{i=1}^{k} \Pr\left(\left[\mathbf{R}_i \geq \frac{2n}{k}\right]\right) \qquad \text{(Union Bound)}$$

$$\leq \sum_{i=1}^{k} \frac{k}{n} = \frac{k^2}{n} \qquad \text{(Bound from Chebyshev's)}$$

As long as $k \leq O(\sqrt{n})$, 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, $\mathbf{R}_i$: number of requests assigned to server $i$. $\mathbb{E}[\mathbf{R}_i] = \frac{n}{k}$. $\mathrm{Var}[\mathbf{R}_i] = \frac{n}{k}$.

# ANOTHER VIEW ON THIS PROBLEM

The number of servers must be small compared to the number of requests ($k = O(\sqrt{n})$) for the maximum load to be bounded in comparison to the expected load with good probability.

- There are many requests routed to a relatively small number of servers so the load seen on each server is close to what is expected via law of large numbers.
- **A Useful Exercise:** Given $n$ requests, and assuming all servers have fixed capacity $C$, how many servers should you provision so that with probability $\geq 99/100$ no server is assigned more than $C$ requests?

$n$: total number of requests, $k$: number of servers randomly assigned requests.

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

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.

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

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

$\mathbf{H}$ 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.

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

- Markov's: $\Pr(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$. First Moment.
- Chebyshev's: $\Pr(|X - \mathbb{E}[X]| \geq t) = \Pr(|X - \mathbb{E}[X]|^2 \geq t^2) \leq \frac{\mathrm{Var}[X]}{t^2}$. Second Moment.
- What if we just apply Markov's inequality to even higher moments?

# A FOURTH MOMENT BOUND

Consider any random variable $\mathsf{X}$:

$$\Pr(|\mathsf{X} - \mathbb{E}[\mathsf{X}]| \geq t) = \Pr\left((\mathsf{X} - \mathbb{E}[\mathsf{X}])^4 \geq t^4\right) \leq \frac{\mathbb{E}\left[(\mathsf{X} - \mathbb{E}[\mathsf{X}])^4\right]}{t^4}.$$

**Application to Coin Flips:** Recall: $n = 100$ independent fair coins, $\mathsf{H}$ is the number of heads.

- Bound the fourth moment:

$$\mathbb{E}\left[(\mathsf{H} - \mathbb{E}[\mathsf{H}])^4\right] = \mathbb{E}\left[\left(\sum_{i=1}^{100} \mathsf{H}_i - 50\right)^4\right] = \sum_{i,j,k,\ell} c_{ijk\ell}\mathbb{E}[\mathsf{H}_i\mathsf{H}_j\mathsf{H}_k\mathsf{H}_\ell] = 1862.5$$

  where $\mathsf{H}_i = 1$ if coin flip $i$ is heads and 0 otherwise. Then apply some messy calculations...

- Apply Fourth Moment Bound: $\Pr(|\mathsf{H} - \mathbb{E}[\mathsf{H}]| \geq t) \leq \frac{1862.5}{t^4}$.

# TIGHTER BOUNDS

| Chebyshev's: | $4^{th}$ Moment: | In Reality: |
|---|---|---|
| $\Pr(\mathbf{H} \geq 60) \leq .25$ | $\Pr(\mathbf{H} \geq 60) \leq .186$ | $\Pr(\mathbf{H} \geq 60) = 0.0284$ |
| $\Pr(\mathbf{H} \geq 70) \leq .0625$ | $\Pr(\mathbf{H} \geq 70) \leq .0116$ | $\Pr(\mathbf{H} \geq 70) = .000039$ |
| $\Pr(\mathbf{H} \geq 80) \leq .04$ | $\Pr(\mathbf{H} \geq 80) \leq .0023$ | $\Pr(\mathbf{H} \geq 80) < 10^{-9}$ |

Can we just keep applying Markov's inequality to higher and higher moments and getting tighter bounds?

- Yes! To a point.
- In fact – don't need to just apply Markov's to $|\mathbf{X} - \mathbb{E}[\mathbf{X}]|^k$ for some $k$. Can apply to any monotonic function $f(|\mathbf{X} - \mathbb{E}[\mathbf{X}]|)$.
- Why monotonic? $\Pr(|\mathbf{X} - \mathbb{E}[\mathbf{X}]| > t) = \Pr(f(|\mathbf{X} - \mathbb{E}[\mathbf{X}]|) > f(t))$.

$\mathbf{H}$: total number heads in 100 random coin flips. $\mathbb{E}[\mathbf{H}] = 50$.

**Moment Generating Function:** Consider for any $t > 0$:

$$M_t(X) = e^{t\cdot(X-\mathbb{E}[X])} = \sum_{k=0}^{\infty} \frac{t^k (X - \mathbb{E}[X])^k}{k!}$$

- $M_t(X)$ is monotonic for any $t > 0$.
- Weighted sum of all moments, with $t$ controlling how slowly the weights fall off (larger $t$ = slower falloff).
- Choosing $t$ appropriately lets one prove a number of very powerful exponential concentration bounds (exponential tail bounds).
- Chernoff bound, Bernstein inequalities, Hoeffding's inequality, Azuma's inequality, Berry-Esseen theorem, etc.
- We will not cover the proofs in the this class.

**Bernstein Inequality:** Consider independent random variables $X_1, \ldots, X_n$ all falling in $[-M, M]$[-1,1]. Let $\mu = \mathbb{E}[\sum_{i=1}^n X_i]$ and $\sigma^2 = \text{Var}[\sum_{i=1}^n X_i] = \sum_{i=1}^n \text{Var}[X_i]$. For any $t \geq 0 s \geq 0$:

$$\Pr\left(\left|\sum_{i=1}^n X_i - \mu\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2\sigma^2 + \frac{4}{3}Mt}\right).$$

$$\Pr\left(\left|\sum_{i=1}^n X_i - \mu\right| \geq s\sigma\right) \leq 2\exp\left(-\frac{s^2}{4}\right).$$

Assume that $M = 1$ and plug in $t = s \cdot \sigma$ for $s \leq \sigma$.

**Compare to Chebyshev's:** $\Pr\left(\left|\sum_{i=1}^n X_i - \mu\right| \geq s\sigma\right) \leq \frac{1}{s^2}$.

- An exponentially stronger dependence on $s$!

# COMPARISION TO CHEBYSHEV'S

Consider again bounding the number of heads $\mathbf{H}$ in $n = 100$ independent coin flips.

| Chebyshev's: | Bernstein: | In Reality: |
|---|---|---|
| $\Pr(\mathbf{H} \geq 60) \leq .25$ | $\Pr(\mathbf{H} \geq 60) \leq .15$ | $\Pr(\mathbf{H} \geq 60) = 0.0284$ |
| $\Pr(\mathbf{H} \geq 70) \leq .0625$ | $\Pr(\mathbf{H} \geq 70) \leq .00086$ | $\Pr(\mathbf{H} \geq 70) = .000039$ |
| $\Pr(\mathbf{H} \geq 80) \leq .04$ | $\Pr(\mathbf{H} \geq 80) \leq 3^{-7}$ | $\Pr(\mathbf{H} \geq 80) < 10^{-9}$ |

Getting much closer to the true probability.

$\mathbf{H}$: total number heads in 100 random coin flips. $\mathbb{E}[\mathbf{H}] = 50$.