COMPSCI 514: ALGORITHMS FOR DATA SCIENCE

Andrew McGregor

Lecture 4

Last Class:

- 2-Level Hashing Analysis (linearity of expectation and Markov's inequality)
- 2-universal and pairwise independent hash functions
- Chebyshev: $\Pr(|\mathbf{X} - \mathbb{E}[\mathbf{X}]| \geq t) \leq \mathrm{Var}[\mathbf{X}]/t^2$

LAST TIME

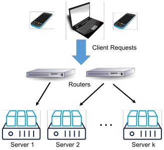
Last Class:

- 2-Level Hashing Analysis (linearity of expectation and Markov's inequality)
- 2-universal and pairwise independent hash functions
- Chebyshev: $\Pr(|\mathbf{X} - \mathbb{E}[\mathbf{X}]| \geq t) \leq \mathrm{Var}[\mathbf{X}]/t^2$

This Time:

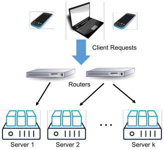
- Random hashing for load balancing. Motivating:
  - Stronger concentration inequalities: Chebyshev's inequality, exponential tail bounds, and their connections to the law of large numbers and central limit theorem.
  - The union bound.

Randomized Load Balancing:



- n requests randomly assigned to k servers.

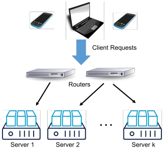
Randomized Load Balancing:



- n requests randomly assigned to k servers.
- Expected load and variance for server i is

$$\mathbb{E}[\mathbf{R}_i] = n/k \quad \text{and} \quad \mathrm{Var}[\mathbf{R}_i] = n(1-1/k)/k \ .$$

Randomized Load Balancing:

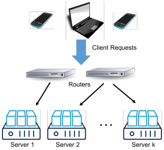


- n requests randomly assigned to k servers.
- Expected load and variance for server i is

$$\mathbb{E}[\mathbf{R}_i] = n/k \quad \text{and} \quad \mathrm{Var}[\mathbf{R}_i] = n(1-1/k)/k \ .$$

- Suppose each server can handle at most $\mathbb{E}[\mathbf{R}_i] = n/k$ requests

Randomized Load Balancing:

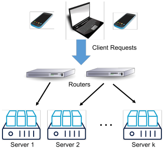


- n requests randomly assigned to k servers.
- Expected load and variance for server i is

$$\mathbb{E}[\mathbf{R}_i] = n/k \quad \text{and} \quad \mathrm{Var}[\mathbf{R}_i] = n(1 - 1/k)/k \ .$$

- Suppose each server can handle at most $\mathbb{E}[\mathbf{R}_i] = n/k$ requests
- By Markov's inequality, $\Pr[\mathbf{R}_i \geq 2\mathbb{E}[\mathbf{R}_i]] \leq 1/2$.

Randomized Load Balancing:



- n requests randomly assigned to k servers.
- Expected load and variance for server i is

$$\mathbb{E}[\mathbf{R}_i] = n/k \quad \text{and} \quad \mathrm{Var}[\mathbf{R}_i] = n(1-1/k)/k\ .$$

- Suppose each server can handle at most $\mathbb{E}[\mathbf{R}_i] = n/k$ requests
- By Markov's inequality, $\Pr[\mathbf{R}_i \geq 2\mathbb{E}[\mathbf{R}_i]] \leq 1/2$.
- By Chebyshev's inequality, $\Pr[\mathbf{R}_i \geq 2\mathbb{E}[\mathbf{R}_i]] \leq \frac{\mathrm{Var}[\mathbf{R}_i]}{\mathbb{E}[\mathbf{R}_i]^2} < \frac{k}{n}$.

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. $\mathbb{E}[\mathbf{R}_i] = \frac{n}{k}$. $\mathrm{Var}[\mathbf{R}_i] = n/k$.

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?

$$\Pr\left(\max_i(\mathbf{R}_i) \geq \frac{2n}{k}\right)$$

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

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?

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

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

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?

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

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

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?

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

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

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. $\mathbb{E}[\mathbf{R}_i] = \frac{n}{k}$. $\mathrm{Var}[\mathbf{R}_i] = n/k$.

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 $\mathbf{R}_1, \ldots, \mathbf{R}_k$ 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. $\mathbb{E}[\mathbf{R}_i] = \frac{n}{k}$. $\mathrm{Var}[\mathbf{R}_i] = n/k$.

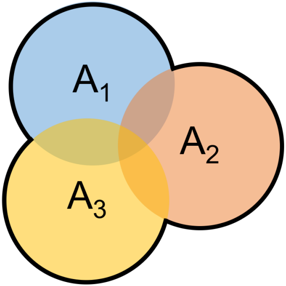
THE UNION BOUND

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

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

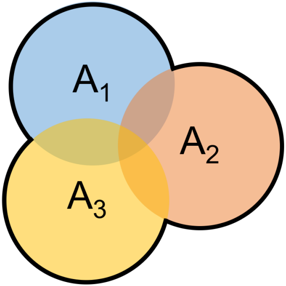
Union Bound: For any random events $A_1, A_2, ..., A_k$,

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



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

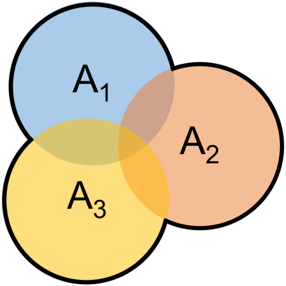
$$\Pr\left(A_1 \cup A_2 \cup \ldots \cup A_k\right) \leq \Pr(A_1) + \Pr(A_2) + \ldots + Pr(A_k).$$



When is the union bound tight?

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

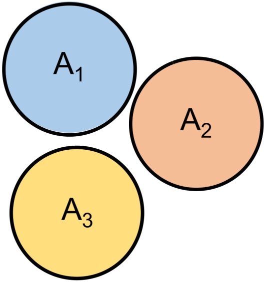
$$\Pr\left(A_1 \cup A_2 \cup \ldots \cup A_k\right) \leq \Pr(A_1) + \Pr(A_2) + \ldots + Pr(A_k).$$



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

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

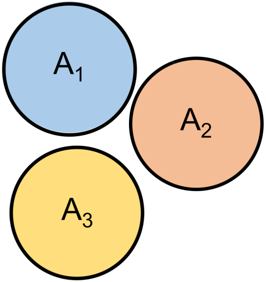
$$\Pr(A_1 \cup A_2 \cup \ldots \cup A_k) \leq \Pr(A_1) + \Pr(A_2) + \ldots + Pr(A_k).$$



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

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

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



When is the union bound tight? When $A_1, ..., A_k$ 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?

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

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

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?

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

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

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

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?

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

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

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?

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

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

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

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

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?

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

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

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

BACK TO CHEBYSHEV'S INEQUALITY

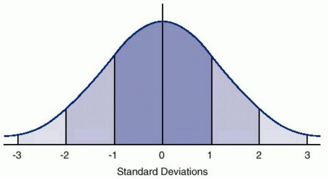
$$\Pr(|\mathbf{X} - \mathbb{E}[\mathbf{X}]| \geq t) \leq \frac{\mathrm{Var}[\mathbf{X}]}{t^2}$$

X: any random variable, t, s: any fixed numbers.

BACK TO CHEBYSHEV'S INEQUALITY

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

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

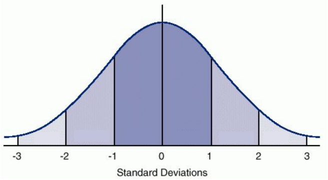


$\mathbf{X}$: any random variable, t, s: any fixed numbers.

BACK TO 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?



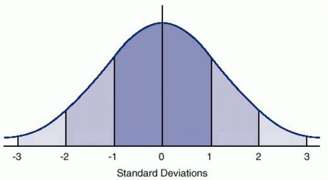
$$\Pr(|\mathbf{X} - \mathbb{E}[\mathbf{X}]| \geq s \cdot \sqrt{\mathrm{Var}[\mathbf{X}]}) \leq \frac{\mathrm{Var}[\mathbf{X}]}{s^2 \cdot \mathrm{Var}[\mathbf{X}]} = \frac{1}{s^2}.$$

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

BACK TO 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?

X: any random variable, t, s: any fixed numbers.

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

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

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

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

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

$$\mathrm{Var}[\mathbf{S}] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i\right]$$

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

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

$$\mathrm{Var}[\mathbf{S}] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}\left[\mathbf{X}_i\right]$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

- Cannot show from vanilla Markov's inequality.

SERVER LOAD AND LAW OF LARGE NUMBERS

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.

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

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.

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.

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.

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

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.

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

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.

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

Markov's:

$\Pr(\mathbf{H} \geq 60) \leq .833$

$\Pr(\mathbf{H} \geq 70) \leq .714$

$\Pr(\mathbf{H} \geq 80) \leq .625$

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.

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

Markov's:	**Chebyshev's:**
$\Pr(\mathbf{H} \geq 60) \leq .833$	$\Pr(\mathbf{H} \geq 60) \leq .25$
$\Pr(\mathbf{H} \geq 70) \leq .714$	$\Pr(\mathbf{H} \geq 70) \leq .0625$
$\Pr(\mathbf{H} \geq 80) \leq .625$	$\Pr(\mathbf{H} \geq 80) \leq .0278$

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.

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

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.

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

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?

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: $\Pr(\mathbf{X} \geq t) \leq \frac{\mathbb{E}[\mathbf{X}]}{t}$. First Moment.

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: $\Pr(\mathbf{X} \geq t) \leq \frac{\mathbb{E}[\mathbf{X}]}{t}$. First Moment.
- Chebyshev's: $\Pr(|\mathbf{X} - \mathbb{E}[\mathbf{X}]| \geq t) = \Pr(|\mathbf{X} - \mathbb{E}[\mathbf{X}]|^2 \geq t^2) \leq \frac{\mathrm{Var}[\mathbf{X}]}{t^2}$. Second Moment.

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

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

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

A FOURTH MOMENT BOUND

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

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

A FOURTH MOMENT BOUND

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

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

A FOURTH MOMENT BOUND

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

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

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

- Bound the fourth moment:

A FOURTH MOMENT BOUND

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

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

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

- Bound the fourth moment:

$$\mathbb{E}\left[(\mathbf{H} - \mathbb{E}[\mathbf{H}])^4\right] = \mathbb{E}\left[\left(\sum_{i=1}^{100} \mathbf{H}_i - 50\right)^4\right]$$

  where $\mathbf{H}_i = 1$ if coin flip i is heads and 0 otherwise.

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

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

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

- Bound the fourth moment:

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

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

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

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

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

- Bound the fourth moment:

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

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

A FOURTH MOMENT BOUND

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

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

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

- Bound the fourth moment:

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

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

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

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

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

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?

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.

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 $f(|\mathbf{X} - \mathbb{E}[\mathbf{X}]|)$.

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?

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 $f\left(|\mathbf{X} - \mathbb{E}[\mathbf{X}]|\right)$.
- Why monotonic?
  $\Pr\left(|\mathbf{X} - \mathbb{E}[\mathbf{X}]| > t\right) = \Pr\left(f\left(|\mathbf{X} - \mathbb{E}[\mathbf{X}]|\right) > f(t)\right)$.

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

$$M_t(\mathbf{X}) = e^{t \cdot (\mathbf{X} - \mathbb{E}[\mathbf{X}])}$$

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

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

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

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

- $M_t(\mathbf{X})$ is monotonic for any $t > 0$.

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

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

- $M_t(\mathbf{X})$ is monotonic for any $t > 0$.
- Weighted sum of all moments, with t controlling how slowly the weights fall off (larger t = slower falloff).

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

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

- $M_t(\mathbf{X})$ 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).

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

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

- $M_t(\mathbf{X})$ 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.

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

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

- $M_t(\mathbf{X})$ 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 explore the basic proof approach in homework.

Bernstein Inequality: Consider independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ all falling in $[-M, M]$. Let $\mu = \mathbb{E}[\sum_{i=1}^n \mathbf{X}_i]$ and $\sigma^2 = \mathrm{Var}[\sum_{i=1}^n \mathbf{X}_i] = \sum_{i=1}^n \mathrm{Var}[\mathbf{X}_i]$. For any $t \geq 0$:

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

Bernstein Inequality: Consider independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ all falling in $[-M, M]$. Let $\mu = \mathbb{E}[\sum_{i=1}^n \mathbf{X}_i]$ and $\sigma^2 = \mathrm{Var}[\sum_{i=1}^n \mathbf{X}_i] = \sum_{i=1}^n \mathrm{Var}[\mathbf{X}_i]$. For any $t \geq 0$:

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

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

Bernstein Inequality: Consider independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ all falling in [-1,1]. Let $\mu = \mathbb{E}[\sum_{i=1}^n \mathbf{X}_i]$ and $\sigma^2 = \text{Var}[\sum_{i=1}^n \mathbf{X}_i] = \sum_{i=1}^n \text{Var}[\mathbf{X}_i]$. For any $s \geq 0$:

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

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

Bernstein Inequality: Consider independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ all falling in [-1,1]. Let $\mu = \mathbb{E}[\sum_{i=1}^n \mathbf{X}_i]$ and $\sigma^2 = \text{Var}[\sum_{i=1}^n \mathbf{X}_i] = \sum_{i=1}^n \text{Var}[\mathbf{X}_i]$. For any $s \geq 0$:

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

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

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

Bernstein Inequality: Consider independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ all falling in [-1,1]. Let $\mu = \mathbb{E}[\sum_{i=1}^n \mathbf{X}_i]$ and $\sigma^2 = \text{Var}[\sum_{i=1}^n \mathbf{X}_i] = \sum_{i=1}^n \text{Var}[\mathbf{X}_i]$. For any $s \geq 0$:

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

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

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

- An exponentially stronger dependence on s!

COMPARISION TO CHEBYSHEV'S

Consider again bounding the number of heads **H** in $n = 100$ independent coin flips.

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

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

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. $\mathbb{E}[\mathbf{H}] = 50$.

A useful variation of the Bernstein inequality for binary (indicator) random variables is:

Chernoff Bound (simplified version): Consider independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ taking values in $\{0, 1\}$. Let $\mu = \mathbb{E}[\sum_{i=1}^n \mathbf{X}_i]$. For any $\delta \geq 0$

$$\Pr\left(\left|\sum_{i=1}^{n} \mathbf{X}_i - \mu\right| \geq \delta\mu\right) \leq 2\exp\left(-\frac{\delta^2\mu}{2+\delta}\right).$$

A useful variation of the Bernstein inequality for binary (indicator) random variables is:

Chernoff Bound (simplified version): Consider independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ taking values in $\{0, 1\}$. Let $\mu = \mathbb{E}[\sum_{i=1}^n \mathbf{X}_i]$. For any $\delta \geq 0$

$$\Pr\left(\left|\sum_{i=1}^n \mathbf{X}_i - \mu\right| \geq \delta\mu\right) \leq 2\exp\left(-\frac{\delta^2\mu}{2+\delta}\right).$$

As δ gets larger and larger, the bound falls of exponentially fast.