# Lecture 6: Concentration Inequalities

Rafael Oliveira

University of Waterloo
Cheriton School of Computer Science

rafael.oliveira.teaching@gmail.com

September 28, 2020

# Overview

# Final Project

- It is not mandatory to work on an open problem. Doing survey on a topic of your interest is also a very good project!

# Final Project

- It is not mandatory to work on an open problem. Doing survey on a topic of your interest is also a very good project!
- If you have questions about it, you can reach out to the TAs and myself during office hours or piazza to ask your questions!

# Final Project

- It is not mandatory to work on an open problem. Doing survey on a topic of your interest is also a very good project!
- If you have questions about it, you can reach out to the TAs and myself during office hours or piazza to ask your questions!
- Probably many of you may have similar questions about the final project, so if you want to ask us something, piazza would be great so that everyone can participate in the discussion! :)

# Final Project

- It is not mandatory to work on an open problem. Doing survey on a topic of your interest is also a very good project!
- If you have questions about it, you can reach out to the TAs and myself during office hours or piazza to ask your questions!
- Probably many of you may have similar questions about the final project, so if you want to ask us something, piazza would be great so that everyone can participate in the discussion! :)
- There is a post pinned on piazza for you all to look for partners for your final project (undergraduates). So if you have a project in mind and want to check if someone else is interested in working with you on it, please post it there!

# Collaboration on Homework

- Collaboration is highly encouraged in the homework, and I encourage everyone to discuss their questions with their colleagues on piazza!

# Collaboration on Homework

- Collaboration is highly encouraged in the homework, and I encourage everyone to discuss their questions with their colleagues on piazza!
- The only thing that is not allowed is for you to tell your solution verbatim, or to write it down to your colleague. But talking about the *ideas behind it* is highly encouraged! :)

# Collaboration on Homework

- Collaboration is highly encouraged in the homework, and I encourage everyone to discuss their questions with their colleagues on piazza!
- The only thing that is not allowed is for you to tell your solution verbatim, or to write it down to your colleague. But talking about the *ideas behind it* is highly encouraged! :)
- Please refrain from asking us to **read your solutions** and **check if they are correct**. We can help you *work out your ideas*, but not help you check that your proof is correct.

# Collaboration on Homework

- Collaboration is highly encouraged in the homework, and I encourage everyone to discuss their questions with their colleagues on piazza!
- The only thing that is not allowed is for you to tell your solution verbatim, or to write it down to your colleague. But talking about the *ideas behind it* is highly encouraged! :)
- Please refrain from asking us to **read your solutions** and **check if they are correct**. We can help you *work out your ideas*, but not help you check that your proof is correct.
- Writing proofs that are correct (or that correctly showcase your ideas) is part of you mathematical development! (as well as checking that your proof is correct)

# Collaboration on Homework

- Collaboration is highly encouraged in the homework, and I encourage everyone to discuss their questions with their colleagues on piazza!
- The only thing that is not allowed is for you to tell your solution verbatim, or to write it down to your colleague. But talking about the *ideas behind it* is highly encouraged! :)
- Please refrain from asking us to **read your solutions** and **check if they are correct**. We can help you *work out your ideas*, but not help you check that your proof is correct.
- Writing proofs that are correct (or that correctly showcase your ideas) is part of you mathematical development! (as well as checking that your proof is correct)
- Solutions to the homework problems *should be simple*. So, if things are getting very complicated in your solution, there is probably another way (this is a **general hint**)

# Why do we want concentration?

When evaluating performance of randomized algorithms, not enough to know our algorithm runs in expected time $T$. What we want to say is

"our algorithm will run in time $\approx T$ *very often*."

# Why do we want concentration?

When evaluating performance of randomized algorithms, not enough to know our algorithm runs in expected time $T$. What we want to say is

"our algorithm will run in time $\approx T$ *very often*."

That is,

- not only analyse the *expected running times* of the algorithms,
- we would also like to know if the algorithm runs in time close to its expected running time *most of the time*.

# Why do we want concentration?

When evaluating performance of randomized algorithms, not enough to know our algorithm runs in expected time $T$. What we want to say is

"our algorithm will run in time $\approx T$ *very often*."

That is,

- not only analyse the *expected running times* of the algorithms,
- we would also like to know if the algorithm runs in time close to its expected running time *most of the time*.

Running time *small* with *high probability* *better than* small expected running time.

# Why do we want concentration?

When evaluating performance of randomized algorithms, not enough to know our algorithm runs in expected time $T$. What we want to say is

"our algorithm will run in time $\approx T$ *very often*."

That is,

- not only analyse the *expected running times* of the algorithms,
- we would also like to know if the algorithm runs in time close to its expected running time *most of the time*.

Running time *small* with *high probability* *better than* small expected running time.

Often times in algorithm analysis, running time is *concentrated* around expectation. This *concentration of measure* proves that our algorithms will *typically* run in time close to expectation.

# Today's inequalities

**Theorem (Markov's Inequality)**

*Let $X$ be a non-negative discrete random variable. Then*

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}, \quad \forall t > 0.$$

# Today's inequalities

**Theorem (Markov's Inequality)**

*Let $X$ be a non-negative discrete random variable. Then*

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}, \quad \forall t > 0.$$

**Theorem (Chebyshev's Inequality)**

*Let $X$ be a [redacted] discrete random variable. Then*

$$\Pr[|X - \mathbb{E}[X]| \geq t] \leq \frac{\mathrm{Var}[X]}{t^2}, \quad \forall t > 0.$$

# Today's inequalities II

**Theorem (Chernoff-Hoeffding's Inequality)**

*Let* $X_1, \ldots, X_n$ *be independent indicator variables such that* $\Pr[X_i = 1] = p_i$, *where* $0 < p_i < 1$. *Let* $X = \sum_{i=1}^{n} X_i$ *and* $\delta > 0$. *Then*

$$\Pr[X \geq (1+\delta) \cdot \mathbb{E}[X]] \leq \left[ \frac{e^{\delta}}{(1+\delta)^{1+\delta}} \right]^{\mathbb{E}[X]},$$

*and*

$$\Pr[X \leq (1-\delta) \cdot \mathbb{E}[X]] \leq \exp\left( -\mathbb{E}[X] \cdot \delta^2 / 2 \right).$$

# Markov's Inequality

**Theorem (Markov's Inequality)**

*Let $X$ be a non-negative discrete random variable. Then*

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}, \quad \forall t > 0.$$

Proof: $\mathbb{E}[X] = \sum_{n \geq 0} \Pr[X=n] \cdot n =$

$= \sum_{n=0}^{t-1} \underbrace{\Pr[X=n] \cdot n}_{\geq 0} + \sum_{n \geq t} \underbrace{n}_{\geq t} \cdot \Pr[X=n]$

$\geq 0 + t \cdot \sum_{n \geq t} \Pr[X=n] = t \cdot \Pr[X \geq t]$. ☒

# Markov's Inequality

**Theorem (Markov's Inequality)**

*Let $X$ be a non-negative discrete random variable. Then*

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}, \quad \forall t > 0.$$

- **Quicksort:** Expected running time of Quicksort is $2n \ln n$. Markov's inequality tells us that the runtime is at least $2cn \ln n$ with probability $\leq 1/c$, for any $c \geq 1$

# Markov's Inequality

**Theorem (Markov's Inequality)**

*Let $X$ be a non-negative discrete random variable. Then*

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}, \quad \forall t > 0.$$

- **Quicksort:** Expected running time of Quicksort is $2n \ln n$. Markov's inequality tells us that the runtime is at least $2cn \ln n$ with probability $\leq 1/c$, for any $c \geq 1$
- **Coin Flipping:** If we flip $n$ fair coins, the expected number of heads is $n/2$. Markov's inequality tells us that $\Pr[\ \#\text{ heads } \geq 3n/4] \leq 2/3$

# Markov's Inequality

**Theorem (Markov's Inequality)**

*Let $X$ be a non-negative discrete random variable. Then*

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}, \quad \forall t > 0.$$

- **Quicksort:** Expected running time of Quicksort is $2n \ln n$. Markov's inequality tells us that the runtime is at least $2cn \ln n$ with probability $\leq 1/c$, for any $c \geq 1$
- **Coin Flipping:** If we flip $n$ fair coins, the expected number of heads is $n/2$. Markov's inequality tells us that $\Pr[\#\text{ heads } \geq 3n/4] \leq 2/3$

Useful when we have no information beyond expected value (or when random variable difficult to analyze). Otherwise other inequalities much sharper!

# Markov's Inequality

Some practice problems.

- Is Markov's inequality tight? Can you give an example?

# Markov's Inequality

Some practice problems.

- Is Markov's inequality tight? Can you give an example?
- Does it hold for general random variables (not just non-negative)?

# Markov's Inequality

Some practice problems.

- Is Markov's inequality tight? Can you give an example?
- Does it hold for general random variables (not just non-negative)?
- Can it be modified to upper bound $\Pr[X \leq t]$?

# Moments and Variance

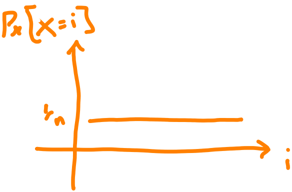
To give better bounds, we need more information about the random variable (beyond expected value).

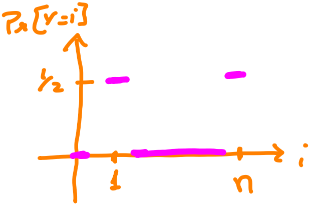
How to distinguish between:

# Moments and Variance

To give better bounds, we need more information about the random variable (beyond expected value).
How to distinguish between:

- $X$ such that $\Pr[X = i] = \begin{cases} 1/n, & \text{if } 1 \leq i \leq n \\ 0, & \text{otherwise} \end{cases}$
- $Y$ such that $\Pr[Y = 1] = 1/2$ and $\Pr[Y = n] = 1/2$ } always far from mean



# Moments and Variance

To give better bounds, we need more information about the random variable (beyond expected value).
How to distinguish between:

- $X$ such that $\Pr[X = i] = \begin{cases} 1/n, & \text{if } 1 \leq i \leq n \\ 0, & \text{otherwise} \end{cases}$
- $Y$ such that $\Pr[Y = 1] = 1/2$ and $\Pr[Y = n] = 1/2$
- same expectation, but very different random variables...

$\mathbb{E}[X] = \mathbb{E}[Y] = \frac{n+1}{2}$

# Moments and Variance

To give better bounds, we need more information about the random variable (beyond expected value).
How to distinguish between:

- $X$ such that $\Pr[X = i] = \begin{cases} 1/n, & \text{if } 1 \leq i \leq n \\ 0, & \text{otherwise} \end{cases}$
- $Y$ such that $\Pr[Y = 1] = 1/2$ and $\Pr[Y = n] = 1/2$
- same expectation, but very different random variables...
- Look at how far variable usually is from its expectation. How to do that?

How far $X$ is from $\mathbb{E}[X]$: $|X - \mathbb{E}[X]|$

Can try to compute

$$\mathbb{E}\left[|X - \mathbb{E}[X]|\right]$$

if close to expectation ← small
if always far from expectation ← large

# Moments and Variance

To give better bounds, we need more information about the random variable (beyond expected value).
How to distinguish between:

- $X$ such that $\Pr[X = i] = \begin{cases} 1/n, & \text{if } 1 \leq i \leq n \\ 0, & \text{otherwise} \end{cases}$
- $Y$ such that $\Pr[Y = 1] = 1/2$ and $\Pr[Y = n] = 1/2$
- same expectation, but very different random variables...
- Look at how far variable usually is from its expectation. How to do that?
- How to bound $\Pr[|X - \mathbb{E}[X]| \geq t]$?

# Moments and Variance

To give better bounds, we need more information about the random variable (beyond expected value).

How to distinguish between:

- $X$ such that $\Pr[X = i] = \begin{cases} 1/n, & \text{if } 1 \le i \le n \\ 0, & \text{otherwise} \end{cases}$
- $Y$ such that $\Pr[Y = 1] = 1/2$ and $\Pr[Y = n] = 1/2$
- same expectation, but very different random variables...
- Look at how far variable usually is from its expectation. How to do that?
- How to bound $\Pr[|X - \mathbb{E}[X]| \ge t]$?

**Theorem (Chebyshev's Inequality)**

*Let $X$ be a [redacted] discrete random variable. Then*

$$\Pr[|X - \mathbb{E}[X]| \ge t] \le \frac{\mathrm{Var}[X]}{t^2}, \quad \forall t > 0.$$

# Chebyshev's inequality

Let $X$ be a random variable.

# Chebyshev's inequality

Let $X$ be a random variable.

- Its **Variance** is defined as $\text{Var}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2]$
- and its **standard deviation** is $\sigma(X) := \sqrt{\text{Var}[X]}$

# Chebyshev's inequality

Let $X$ be a random variable.

- Its **Variance** is defined as $\mathrm{Var}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2]$
- and its **standard deviation** is $\sigma(X) := \sqrt{\mathrm{Var}[X]}$

**Theorem (Chebyshev's Inequality)**

*Let $X$ be a discrete random variable. Then*

$$\Pr[|X - \mathbb{E}[X]| \geq t] \leq \frac{\mathrm{Var}[X]}{t^2}, \quad \forall t > 0.$$

Proof: only thing we know is Markov. Let's use it!

$Y := (X - \mathbb{E}[X])^2$ $\quad Y$: discrete if $X$ discrete; $\geq 0$ (can use Markov!)

By Markov: $\Pr[Y \geq t^2] \leq \frac{\mathbb{E}[Y]}{t^2} = \frac{\mathrm{Var}[X]}{t^2}$

and $\Pr[|X - \mathbb{E}[X]| \geq t] = \Pr[Y \geq t^2]$ ∎

# Covariance

How do we measure the correlation between two random variables?

# Covariance

How do we measure the correlation between two random variables?

**Definition (Covariance)**

The *covariance* of two random variables $X, Y$ is defined as

$$\mathrm{Cov}[X, Y] := \mathbb{E}[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])].$$

We say that $X, Y$ are *positively correlated* if $\mathrm{Cov}[X, Y] > 0$ and *negatively correlated* if $\mathrm{Cov}[X, Y] < 0$.

# Covariance

How do we measure the correlation between two random variables?

## Definition (Covariance)

The *covariance* of two random variables $X, Y$ is defined as

$$\mathrm{Cov}[X, Y] := \mathbb{E}[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])].$$

We say that $X, Y$ are *positively correlated* if $\mathrm{Cov}[X, Y] > 0$ and *negatively correlated* if $\mathrm{Cov}[X, Y] < 0$.

## Proposition

- $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{Cov}[X, Y]$
- *If $X, Y$ are independent, then* $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$

Practice problem: prove this proposition!

# Chebyshev & Covariance example

**Coin Flipping:** If $X$ be # heads in $n$ independent unbiased coin flips, let us bound again $\Pr[X \geq 3n/4]$.

$$X_i = \begin{cases} 1 & \text{if } i^{th} \text{ coin flipped heads} \\ 0 & \text{otherwise} \end{cases}$$

$X = \sum_{i=1}^{n} X_i$ , $X_i, X_j$ independent.

By proposition $\text{Var}[X] = \sum_{i=1}^{n} \text{Var}[X_i] = \sum_{i=1}^{n} \frac{1}{4} = \frac{n}{4}$

Chebyshev: $\Pr[X \geq 3n/4] \leq \Pr[|X - \underbrace{n/2}_{E[X]}| \geq n/4]$

$\leq \frac{n/4}{(n/4)^2} = \frac{4}{n}$ $\square$

Much better than Markov!

# Higher Moments

To obtain even more information of a random variable, useful to see more of its moments:

# Higher Moments

To obtain even more information of a random variable, useful to see more of its moments:

- the *$k^{th}$ moment* of random variable $X$ is $\mathbb{E}[X^k]$.

# Higher Moments

To obtain even more information of a random variable, useful to see more of its moments:

- the *$k^{th}$ moment* of random variable $X$ is $\mathbb{E}[X^k]$.
- the *$k^{th}$ central moment* of random variable $X$ is

$$\mu_X^{(k)} := \mathbb{E}[(X - \mathbb{E}[X])^k],$$

if it exists.

$\mathrm{Var}[X] = \mu_X^{(2)}$

Practice problem: when will the $k^{th}$ moment not exist? (Appendix C of MR'07).

(will post on this on my webpage)

# Higher Moments

To obtain even more information of a random variable, useful to see more of its moments:

- the $k^{th}$ *moment* of random variable $X$ is $\mathbb{E}[X^k]$.
- the $k^{th}$ *central moment* of random variable $X$ is

$$\mu_X^{(k)} := \mathbb{E}[(X - \mathbb{E}[X])^k],$$

  if it exists.

**Remark**

Chebyshev's inequality is most useful when we only have information about the *second moment* of our random variable $X$.

# Higher Moments

To obtain even more information of a random variable, useful to see more of its moments:

- the $k^{th}$ *moment* of random variable $X$ is $\mathbb{E}[X^k]$.
- the $k^{th}$ *central moment* of random variable $X$ is

$$\mu_X^{(k)} := \mathbb{E}[(X - \mathbb{E}[X])^k],$$

  if it exists.

**Remark**

Chebyshev's inequality is most useful when we only have information about the *second moment* of our random variable $X$.

Practice problem: Can you generalize Chebyshev's inequality to $k^{th}$ order moments?

# Sums of Independent Random Variables

Often times in analysis of algorithms we deal with random variables which are sums of independent random variables (see Distinct Elements analysis from last lecture, hashing, etc). balls & bins lecture!

# Sums of Independent Random Variables

Often times in analysis of algorithms we deal with random variables which are sums of independent random variables (see Distinct Elements analysis from last lecture, hashing, etc).

Can we use this information to get better tail inequalities?

# Sums of Independent Random Variables

Often times in analysis of algorithms we deal with random variables which are sums of independent random variables (see Distinct Elements analysis from last lecture, hashing, etc).

Can we use this information to get better tail inequalities?

**Law of large numbers:** average of independent, identically distributed variables is approximately the expectation of the random variables. That is, if each $X_i$ is an independent copy of random variable $X$

$$\frac{1}{n} \cdot \sum_{i=1}^{n} X_i \approx \mathbb{E}[X]$$

# Sums of Independent Random Variables

Often times in analysis of algorithms we deal with random variables which are sums of independent random variables (see Distinct Elements analysis from last lecture, hashing, etc).

Can we use this information to get better tail inequalities?

**Law of large numbers:** average of independent, identically distributed variables is approximately the expectation of the random variables. That is, if each $X_i$ is an independent copy of random variable $X$

$$\frac{1}{n} \cdot \sum_{i=1}^{n} X_i \approx \mathbb{E}[X]$$

**Central Limit Theorem:** if we let $Z_n = \sum_{i=1}^{n} X_i$, where $X_i$ independent copy of $X$, the random variable

$$Y_n = \frac{Z_n - n \cdot \mathbb{E}[X]}{\sqrt{n \cdot \sigma(X)^2}} \to \mathcal{N}(0, 1)$$

Gaussian distribution

# Chernoff Bounds

*Chernoff bounds* give us quantitative estimates of the probability that $X$ is far from $\mathbb{E}[X]$ for large enough values of $n$.[1]

[1] Also works for sums of random variables which are not identically distributed!

# Chernoff Bounds

*Chernoff bounds* give us quantitative estimates of the probability that $X$ is far from $\mathbb{E}[X]$ for large enough values of $n$.[1]

Simple Setting: we have $n$ coin flips, each is head with probability $p$. So

$$X_i = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{otherwise} \end{cases} \qquad \text{and } X = \sum_{i=1}^{n} X_i.$$

[1] Also works for sums of random variables which are not identically distributed!

# Chernoff Bounds

*Chernoff bounds* give us quantitative estimates of the probability that $X$ is far from $\mathbb{E}[X]$ for large enough values of $n$.[1]

Simple Setting: we have $n$ coin flips, each is head with probability $p$. So

$$X_i = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{otherwise} \end{cases} \qquad \text{and } X = \sum_{i=1}^{n} X_i.$$

- Expected # heads: $n \cdot p$

[1] Also works for sums of random variables which are not identically distributed!

# Chernoff Bounds

*Chernoff bounds* give us quantitative estimates of the probability that $X$ is far from $\mathbb{E}[X]$ for large enough values of $n$.[1]

Simple Setting: we have $n$ coin flips, each is head with probability $p$. So

$$X_i = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{otherwise} \end{cases} \qquad \text{and } X = \sum_{i=1}^{n} X_i.$$

- Expected # heads: $n \cdot p$
- To bound upper tail, need to compute:

$$\Pr[X \geq k] \leq \sum_{i \geq k} \binom{n}{i} p^i (1-p)^{n-i}$$

[1] Also works for sums of random variables which are not identically distributed!

# Chernoff Bounds

*Chernoff bounds* give us quantitative estimates of the probability that $X$ is far from $\mathbb{E}[X]$ for large enough values of $n$.[1]

Simple Setting: we have $n$ coin flips, each is head with probability $p$. So

$$X_i = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{otherwise} \end{cases} \qquad \text{and } X = \sum_{i=1}^{n} X_i.$$

- Expected # heads: $n \cdot p$
- To bound upper tail, need to compute:

$$\Pr[X \geq k] \leq \sum_{i \geq k} \binom{n}{i} p^i (1-p)^{n-i}$$

- Not easy to work with, hard to generalize (homework 1 question 6)

[1] Also works for sums of random variables which are not identically distributed!

# Chernoff Bounds

Generic Chernoff Bounds: apply Markov in the following way:

$$\Pr[X \geq a] = \Pr[e^{tX} \geq e^{ta}] \leq \mathbb{E}[e^{tX}]/e^{ta}, \quad \text{for any } t > 0.$$

# Chernoff Bounds

Generic Chernoff Bounds: apply Markov in the following way:

$$\Pr[X \geq a] = \Pr[e^{tX} \geq e^{ta}] \leq \mathbb{E}[e^{tX}]/e^{ta}, \quad \text{for any } t > 0.$$

What do we gain by doing this?

# Chernoff Bounds

Generic Chernoff Bounds: apply Markov in the following way:

$$\Pr[X \geq a] = \Pr[e^{tX} \geq e^{ta}] \leq \mathbb{E}[e^{tX}]/e^{ta}, \quad \text{for any } t > 0.$$

What do we gain by doing this?

- If $X = X_1 + X_2$, where $X_1, X_2$ are independent, note that

$$\mathbb{E}[e^{tX}] = \mathbb{E}[e^{tX_1} e^{tX_2}] = \mathbb{E}[e^{tX_1}] \cdot \mathbb{E}[e^{tX_2}]$$

easy to bound expectation by splitting it

(yes, but we know of linearity of expectation...)

# Chernoff Bounds

Generic Chernoff Bounds: apply Markov in the following way:

$$\Pr[X \geq a] = \Pr[e^{tX} \geq e^{ta}] \leq \mathbb{E}[e^{tX}]/e^{ta}, \quad \text{for any } t > 0.$$

What do we gain by doing this?

- If $X = X_1 + X_2$, where $X_1, X_2$ are independent, note that

$$\mathbb{E}[e^{tX}] = \mathbb{E}[e^{tX_1} e^{tX_2}] = \mathbb{E}[e^{tX_1}] \cdot \mathbb{E}[e^{tX_2}]$$

- The *moment generating function*

$$M_X(t) := \mathbb{E}[e^{tX}] = \mathbb{E}\left[\sum_{i \geq 0} \frac{t^i}{i!} \cdot X^i\right] = \sum_{i \geq 0} \frac{t^i}{i!} \cdot \mathbb{E}\left[X^i\right]$$

contains information about all moments!

so expect to be more powerful than Chebyshev (2nd moment)

# Chernoff Bounds for Bounded Variables

## Example (Heterogeneous Coin Flips)

Let $X_i = \begin{cases} 1, & \text{with probability } p_i \\ 0, & \text{otherwise} \end{cases}$, $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbb{E}[X]$

1. for $\delta > 0$, $\Pr[X \geq (1+\delta)\mu] \leq \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{\mu}$

$$\mu = \mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} p_i$$

Proof: $\Pr[X \geq (1+\delta)\mu] = \Pr[e^{tX} \geq e^{t(1+\delta)\mu}] \leq \mathbb{E}[e^{tX}] / e^{t(1+\delta)\mu}$

$$= \frac{1}{e^{t(1+\delta)\mu}} \cdot \prod_{i=1}^{n} \mathbb{E}[e^{tX_i}] = \frac{1}{e^{t(1+\delta)\mu}} \cdot \prod_{i=1}^{n} \left(e^{t} \cdot p_i + (1-p_i)\right) \leq$$

$$\leq \frac{1}{e^{t(1+\delta)\mu}} \cdot \prod_{i=1}^{n} e^{p_i(e^t - 1)} = \frac{1}{e^{t(1+\delta)\mu}} \cdot e^{(e^t-1)\sum p_i} = \left(\frac{e^{e^t - 1}}{e^{t(1+\delta)}}\right)^{\mu}$$

$t = \ln(1+\delta)$

# Chernoff Bounds for Bounded Variables

## Example (Heterogeneous Coin Flips)

Let $X_i = \begin{cases} 1, & \text{with probability } p_i \\ 0, & \text{otherwise} \end{cases}$, $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbb{E}[X]$

1. for $\delta > 0$, $\Pr[X \geq (1+\delta)\mu] \leq \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{\mu}$
2. for $0 < \delta < 1$, $\Pr[X \geq (1+\delta)\mu] \leq e^{-\delta^2 \mu/3}$

just note that $0 < \delta < 1 \Rightarrow \frac{e^{\delta}}{(1+\delta)^{1+\delta}} < e^{-\delta^2/3}$

Then consider $f(\delta) = \delta - (1+\delta)\ln(1+\delta) + \delta^2/3$ and

show $f(\delta) \leq 0$ in $[0,1]$.

# Chernoff Bounds for Bounded Variables

**Example (Heterogeneous Coin Flips)**

Let $X_i = \begin{cases} 1, & \text{with probability } p_i \\ 0, & \text{otherwise} \end{cases}$, $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbb{E}[X]$

1. for $\delta > 0$, $\Pr[X \geq (1+\delta)\mu] \leq \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{\mu}$
2. for $0 < \delta < 1$, $\Pr[X \geq (1+\delta)\mu] \leq e^{-\delta^2\mu/3}$
3. for $R \geq 6\mu$, $\Pr[X \geq R] \leq 2^{-R}$

$R \geq 6\mu$ then $\delta \geq 5$ in ①

# Chernoff Bounds for Bounded Variables

What about the lower tail?

[2] See [Motwani & Raghavan 2007, Theorem 4.2] or [Mitzenmacher & Upfal, Theorem 4.5]

# Chernoff Bounds for Bounded Variables

What about the lower tail?

Similar proof, by setting $t < 0$.[2]

[2] See [Motwani & Raghavan 2007, Theorem 4.2] or [Mitzenmacher & Upfal, Theorem 4.5]

# Chernoff Bounds for Bounded Variables

What about the lower tail?

Similar proof, by setting $t < 0$.[2]

**Theorem (Heterogeneous Coin Flips - lower tail)**

1. $\Pr[X \leq (1-\delta)\cdot\mu] \leq \left[\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right]^{\mu}$
2. *if* $0 < \delta < 1$ *then* $\Pr[X \leq (1-\delta)\cdot\mu] \leq e^{-\mu\delta^2/2}$

Practice problem: Prove this theorem!

[2] See [Motwani & Raghavan 2007, Theorem 4.2] or [Mitzenmacher & Upfal, Theorem 4.5]

# Hoeffding's generalization

What if the variables $X_i$ took values in $[a_i, b_i]$?

# Hoeffding's generalization

What if the variables $X_i$ took values in $[a_i, b_i]$?

**Theorem (Hoeffding's Inequality)**

*Let $X_i$ be independent random variables, taking values in $[a_i, b_i]$, $X = \sum_{i=1}^{n} X_i$. Then*

$$\Pr[|X - \mathbb{E}[X]| \geq \ell] \leq 2 \cdot \exp\left(-\frac{2\ell^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

# Hoeffding's generalization

What if the variables $X_i$ took values in $[a_i, b_i]$?

**Theorem (Hoeffding's Inequality)**

*Let $X_i$ be independent random variables, taking values in $[a_i, b_i]$, $X = \sum_{i=1}^{n} X_i$. Then*

$$\Pr[|X - \mathbb{E}[X]| \geq \ell] \leq 2 \cdot \exp\left(-\frac{2\ell^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

Proof uses *Hoeffding's lemma:* $\mathbb{E}[e^{t(X_i - \mathbb{E}[X_i])}] \leq \exp\left(\frac{t^2(b_i - a_i)^2}{8}\right)$

Practice problem: prove this theorem.

# Remarks

- In coin flips example from beginning of lecture, by flipping $n$ independent fair coins, expected # heads is $n/2$. Chernoff-Hoeffding implies:

$$\Pr[|\#\text{ heads } - \mu| \geq \delta\mu] \leq 2\exp(\mu\delta^2/3) = 2\exp(n\delta^2/6)$$

# Remarks

- In coin flips example from beginning of lecture, by flipping $n$ independent fair coins, expected # heads is $n/2$. Chernoff-Hoeffding implies:

$$\Pr[|\#\text{ heads } - \mu| \geq \delta\mu] \leq 2\exp(\mu\delta^2/3) = 2\exp(n\delta^2/6)$$

- Setting $\delta = \sqrt{60/n}$, probability above is $\leq 2e^{-10}$. Thus

$$\Pr[|\#\text{ heads } - n/2| \geq \sqrt{15 \cdot n}] \leq 2e^{-10}.$$

# Remarks

- In coin flips example from beginning of lecture, by flipping $n$ independent fair coins, expected # heads is $n/2$. Chernoff-Hoeffding implies:

$$\Pr[|\#\text{ heads } - \mu| \geq \delta\mu] \leq 2\exp(\mu\delta^2/3) = 2\exp(n\delta^2/6)$$

- Setting $\delta = \sqrt{60/n}$, probability above is $\leq 2e^{-10}$. Thus

$$\Pr[|\#\text{ heads } - n/2| \geq \sqrt{15 \cdot n}] \leq 2e^{-10}.$$

- With high probability, # heads is within $O(\sqrt{n})$ of the expected value (this comes up in many places). **Practice problem:** prove that with constant probability that $|\#\text{ heads } - n/2| = \Omega(n)$.

# Remarks

- In coin flips example from beginning of lecture, by flipping $n$ independent fair coins, expected # heads is $n/2$. Chernoff-Hoeffding implies:

$$\Pr[|\#\text{ heads } - \mu| \geq \delta\mu] \leq 2\exp(\mu\delta^2/3) = 2\exp(n\delta^2/6)$$

- Setting $\delta = \sqrt{60/n}$, probability above is $\leq 2e^{-10}$. Thus

$$\Pr[|\#\text{ heads } - n/2| \geq \sqrt{15 \cdot n}] \leq 2e^{-10}.$$

- With high probability, # heads is within $O(\sqrt{n})$ of the expected value (this comes up in many places). **Practice problem:** prove that with constant probability that $|\#\text{ heads } - n/2| = \Omega(n)$.
- Recall from previous slides that Markov gave us that $\Pr[\#\text{ heads } \geq 3n/4] \leq 2/3$, and Chebyshev gave us $\Pr[\#\text{ heads } \geq 3n/4] \leq 4/n$. Chernoff gives us $\Pr[\#\text{ heads } \geq 3n/4] \leq e^{-n/24}$.

# Remarks

- It is often easier to compute moments by computing the moment generating functions

# Remarks

- It is often easier to compute moments by computing the moment generating functions
- Why do we want to compute moments? See Sum-of-Squares and pseudo-distributions references in course webpage. These methods give very powerful tools to solve many challenging problems! (great final project topic!)

# Remarks

- It is often easier to compute moments by computing the moment generating functions
- Why do we want to compute moments? See Sum-of-Squares and pseudo-distributions references in course webpage. These methods give very powerful tools to solve many challenging problems! (great final project topic!)
- Chernoff-Hoeffding bounds also hold for *negatively correlated* variables, because all we need is

$$\mathbb{E}[e^{t(X+Y)}] \leq \mathbb{E}[e^{tX}] \cdot \mathbb{E}[e^{tY}]$$

  This observation is very useful in many applications (also great source of final projects!)

# Remarks

- It is often easier to compute moments by computing the moment generating functions
- Why do we want to compute moments? See Sum-of-Squares and pseudo-distributions references in course webpage. These methods give very powerful tools to solve many challenging problems! (great final project topic!)
- Chernoff-Hoeffding bounds also hold for *negatively correlated* variables, because all we need is

$$\mathbb{E}[e^{t(X+Y)}] \leq \mathbb{E}[e^{tX}] \cdot \mathbb{E}[e^{tY}]$$

  This observation is very useful in many applications (also great source of final projects!)
- For instance: two edges appear in a random spanning tree is a negatively correlated event, thus Chernoff bounds are useful to analyze random spanning trees.

# Acknowledgement


- Lecture based largely on Lap Chi's notes and [Motwani & Raghavan 2007, Chapters 3 and 4].
- See Lap Chi's notes at https://cs.uwaterloo.ca/~lapchi/cs466/notes/L02.pdf

# References I


Motwani, Rajeev and Raghavan, Prabhakar (2007)
Randomized Algorithms

Mitzenmacher, Michael, and Eli Upfal (2017)
Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis.
Cambridge university press, 2017.