Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Lehrstuhl für
INFORMATIK I
Algorithmen & Komplexität

Institut für Informatik

Seminar

Mathematical Foundations of Data Science

Summer Term 2021

Primer on Tail Bounds - April 28, 2021

Chair of Computer Science I - Algorithms and Complexity

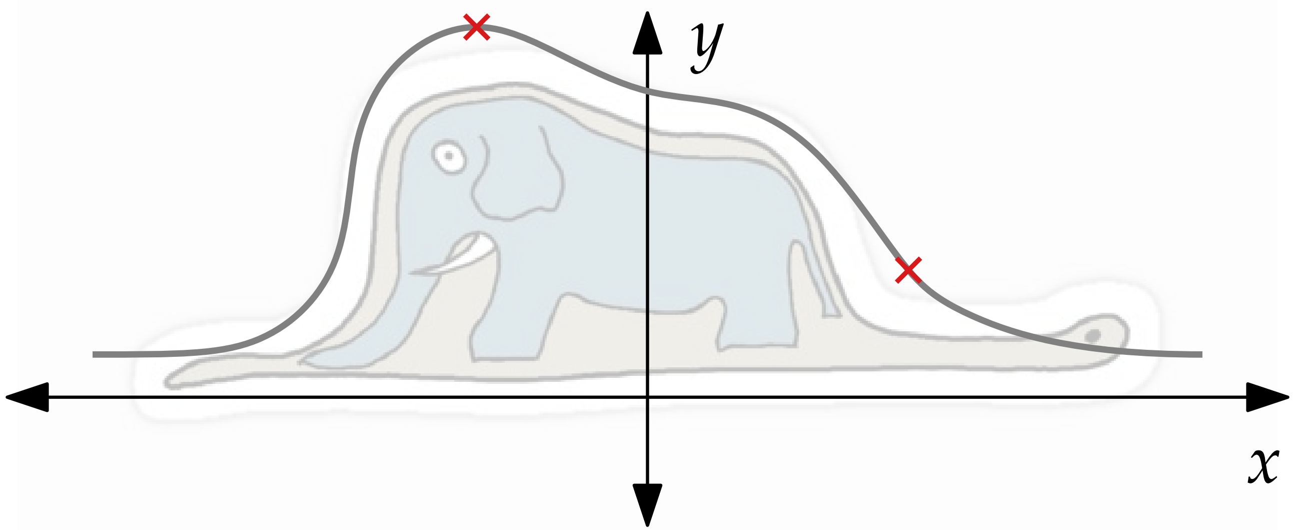
Kamyar Khodamoradi Joachim Spoerhase Alexander Wolff

Agenda

1. Preliminary Definitions
2. Markov's and Chebyshev's Inequalities
   - The Law of Large Numbers
3. Higher Moments and Chernoff's Inequality
4. Master Tail Bounds
   - Applications

Tail Bounds

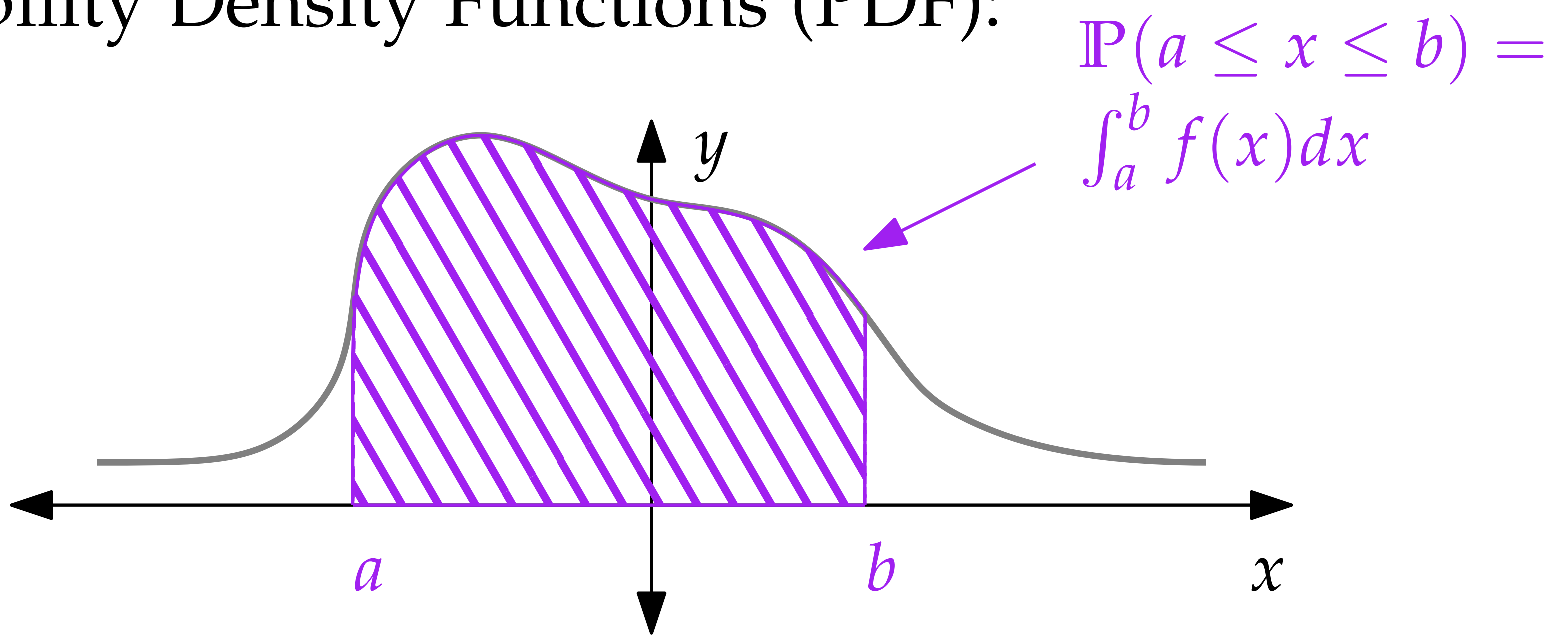
- Continuous Random Variables
- x: Real Random Variable in $(-\infty, +\infty)$
- Probability Density Functions (PDF):



- E.g., $f(x) = \dfrac{e^{-\frac{1}{2}(x-1)^2} + e^{-(x+1)^2}}{(\sqrt{2}+1)\sqrt{\pi}}$ → Relative Likelihood

Tail Bounds

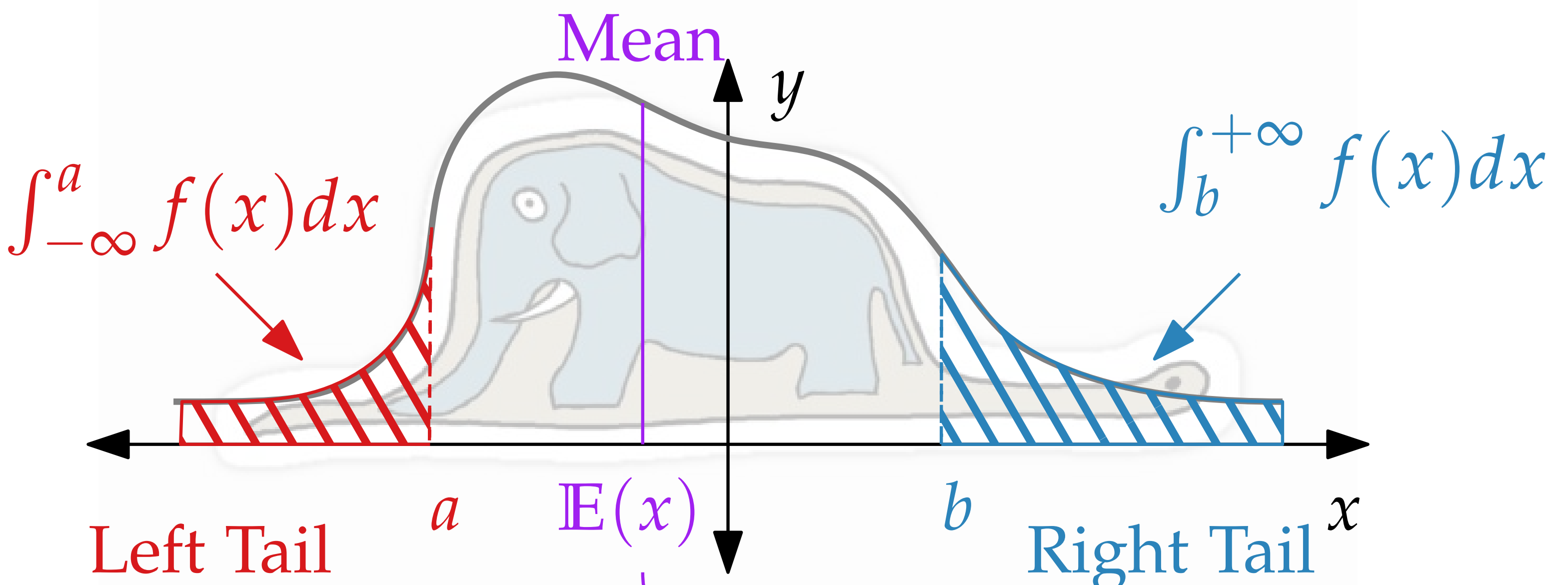
- Continuous Random Variables
- x: Real Random Variable in $(-\infty, +\infty)$
- Probability Density Functions (PDF):



- E.g., $f(x) = \frac{e^{-\frac{1}{2}(x-1)^2} + e^{-(x+1)^2}}{(\sqrt{2}+1)\sqrt{\pi}}$

Tail Bounds

- Continuous Random Variables
- x: Real Random Variable in $(-\infty, +\infty)$
- Probability Density Functions (PDF):



- E.g., $f(x) = \frac{e^{-\frac{1}{2}(x-1)^2} + e^{-(x+1)^2}}{(\sqrt{2}+1)\sqrt{\pi}}$

$$\mathbb{E}(x) = \int_{-\infty}^{+\infty} x f(x)dx$$

Agenda

1. Preliminary Definitions
2. Markov's and Chebyshev's Inequalities
   - The Law of Large Numbers
3. Higher Moments and Chernoff's Inequality
4. Master Tail Bounds
   - Applications

Markov's Inequality

Thm. Let x be a random variable ≥ 0. Then, for $a > 0$:

$$\mathbb{P}(x \geq a) \leq \frac{\mathbb{E}(x)}{a}.$$

Proof. Assume $f(x)$ is the PDF of x.

Ignore

$$\mathbb{E}(x) = \int_0^\infty x f(x) dx = \boxed{\int_0^a x f(x) dx} + \int_a^\infty x f(x) dx$$

$$\geq \int_a^\infty x f(x) dx \geq a \int_a^\infty f(x) dx$$

Markov's Inequality

Thm. Let x be a random variable ≥ 0. Then, for $a > 0$:

$$\mathbb{P}(x \geq a) \leq \frac{\mathbb{E}(x)}{a}.$$

Proof. Assume $f(x)$ is the PDF of x.

$$\begin{aligned}\mathbb{E}(x) &= \int_0^\infty x f(x) dx = \int_0^a x f(x) dx + \int_a^\infty x f(x) dx \\ &\geq \int_a^\infty x f(x) dx \geq a \int_a^\infty f(x) dx \\ &= a\mathbb{P}(x \geq a).\end{aligned}$$

Almost the same proof for *discrete* probabilities

Markov's Inequality

Thm. Let x be a random variable ≥ 0. Then, for $a > 0$:

$$\mathbb{P}(x \geq a) \leq \frac{\mathbb{E}(x)}{a}.$$

Proof. Assume $f(x)$ is the PDF of x.

$$\begin{aligned}\mathbb{E}(x) &= \int_0^\infty x f(x) dx = \int_0^a x f(x) dx + \int_a^\infty x f(x) dx \\ &\geq \int_a^\infty x f(x) dx \geq a \int_a^\infty f(x) dx \\ &= a\mathbb{P}(x \geq a).\end{aligned}$$

Simple, but very useful!

Chebyshev's Inequality

Thm. Let x be a random variable. Then, for $a > 0$:

$$\mathbb{P}(|x - \mathbb{E}(x)| \geq a) \leq \frac{Var(x)}{a^2}.$$

Proof. Note: $\mathbb{P}(|x - \mathbb{E}(x)| \geq a) = \mathbb{P}(|x - \mathbb{E}(x)|^2 \geq a^2)$.

Non-negative R.V.

Chebyshev's Inequality

Thm. Let x be a random variable. Then, for $a > 0$:

$$\mathbb{P}(|x - \mathbb{E}(x)| \geq a) \leq \frac{Var(x)}{a^2}.$$

Proof. Note: $\mathbb{P}(|x - \mathbb{E}(x)| \geq a) = \mathbb{P}(|x - \mathbb{E}(x)|^2 \geq a^2)$.

Apply Markov's Inequality

$$\leq \frac{\mathbb{E}(|x - \mathbb{E}(x)|^2)}{a^2}$$

Further Information

$$= \frac{Var(x)}{a^2}$$

Better Bound

The Law of Large Numbers

- So far: tail bounds for a singe sample of a R.V.
- What about $x_1 + x_2 + \ldots + x_n$?
  - x_i: Independet sample of a random variable x
  - x: Finite variance.

Thm. Assume $x_1, x_2, \ldots, x_n$ are n independent samples of x.Then

$$\mathbb{P}\left(\mid \frac{x_1 + x_2 + \ldots + x_n}{n} - \mathbb{E}(x) \mid \geq \varepsilon \right) \leq \frac{Var(x)}{n\varepsilon^2}.$$

The Law of Large Numbers

Fact. If x and y are independent random variables, Then

$$Var(x+y) = Var(x) + Var(y).$$

Fact. $Var(cx) = c^2 Var(x).$

The Law of Large Numbers

Target:

$$\mathbb{P}\left(\left|\frac{x_1 + x_2 + \ldots + x_n}{n} - \mathbb{E}(x)\right| \geq \varepsilon\right) \leq \frac{Var(x)}{n\varepsilon^2}.$$

Proof.

$$\mathbb{P}\left(\left|\frac{x_1 + x_2 + \ldots + x_n}{n} - \mathbb{E}(x)\right| \geq \varepsilon\right)$$

$$\mathbb{E}(x_i) = \mathbb{E}(x)$$

$$= \mathbb{P}\left(\left|\frac{x_1 + x_2 + \ldots + x_n}{n} - \mathbb{E}\left(\frac{x_1 + x_2 + \ldots + x_n}{n}\right)\right| \geq \varepsilon\right)$$

The Law of Large Numbers

Target:

$$\mathbb{P}\left(\left|\frac{x_1+x_2+\ldots+x_n}{n}-\mathbb{E}(x)\right|\geq\varepsilon\right)\leq\frac{Var(x)}{n\varepsilon^2}.$$

Proof.

$$\mathbb{P}\left(\left|\frac{x_1+x_2+\ldots+x_n}{n}-\mathbb{E}(x)\right|\geq\varepsilon\right)$$

$$=\mathbb{P}\left(\left|\underbrace{\frac{x_1+x_2+\ldots+x_n}{n}}_{y}-\underbrace{\mathbb{E}\left(\frac{x_1+x_2+\ldots+x_n}{n}\right)}_{\mathbb{E}(y)}\right|\geq\varepsilon\right)$$

$$\leq\frac{Var(y)}{\varepsilon^2}$$

Fact 2.

$$=\frac{Var(\frac{x_1+x_2+\ldots+x_n}{n})}{\varepsilon^2}=\frac{Var(x_1+x_2+\ldots+x_n)}{n^2\varepsilon^2}$$

The Law of Large Numbers

Target:

$$\mathbb{P}\left(\left|\frac{x_1 + x_2 + \ldots + x_n}{n} - \mathbb{E}(x)\right| \geq \varepsilon\right) \leq \frac{Var(x)}{n\varepsilon^2}.$$

Proof.

$$\mathbb{P}\left(\left|\frac{x_1 + x_2 + \ldots + x_n}{n} - \mathbb{E}(x)\right| \geq \varepsilon\right)$$

$$= \mathbb{P}\left(\left|\underbrace{\frac{x_1 + x_2 + \ldots + x_n}{n}}_{y} - \underbrace{\mathbb{E}\left(\frac{x_1 + x_2 + \ldots + x_n}{n}\right)}_{\mathbb{E}(y)}\right| \geq \varepsilon\right)$$

$$\leq \frac{Var(y)}{\varepsilon^2}$$

Fact 1.

$$= \frac{Var(\frac{x_1+x_2+\ldots+x_n}{n})}{\varepsilon^2} = \frac{Var(x_1 + x_2 + \ldots + x_n)}{n^2\varepsilon^2} = \frac{nVar(x)}{n^2\varepsilon^2}.$$

The Law of Large Numbers

Target: **Interesting Applications**

$$\mathbb{P}\left(\left|\frac{x_1+x_2+\ldots+x_n}{n}-\mathbb{E}(x)\right|\geq\varepsilon\right)\leq\frac{Var(x)}{n\varepsilon^2}.$$

Proof.

$$\mathbb{P}\left(\left|\frac{x_1+x_2+\ldots+x_n}{n}-\mathbb{E}(x)\right|\geq\varepsilon\right)$$

$$=\mathbb{P}\left(\left|\underbrace{\frac{x_1+x_2+\ldots+x_n}{n}}_{y}-\underbrace{\mathbb{E}\left(\frac{x_1+x_2+\ldots+x_n}{n}\right)}_{\mathbb{E}(y)}\right|\geq\varepsilon\right)$$

$$\leq\frac{Var(y)}{\varepsilon^2}$$

$$=\frac{Var\left(\frac{x_1+x_2+\ldots+x_n}{n}\right)}{\varepsilon^2}=\frac{Var(x_1+x_2+\ldots+x_n)}{n^2\varepsilon^2}=\frac{nVar(x)}{n^2\varepsilon^2}.$$

Agenda

1. Preliminary Definitions
2. Markov's and Chebyshev's Inequalities
   - The Law of Large Numbers
3. Higher Moments and Chernoff's Inequality
4. Master Tail Bounds
   - Applications

Moments

- Markov: $\mathbb{P}(x \geq a) \leq \frac{\mathbb{E}(x)}{a}$ → 1^{st} Moment
- Chebyshev: $\mathbb{P}(|x - \mathbb{E}(x)| \geq a) \leq \frac{Var(x)}{a^2}$ → $\mathbb{E}((x - \mathbb{E}(x))^2)$

Def. The r^{th} moment of x is $\mathbb{E}(x^r)$ for $r \in \mathbb{N}_{>0}$.

Moments

- Markov: $\mathbb{P}(x \geq a) \leq \frac{\mathbb{E}(x)}{a}$ → 1^{st} Moment
- Chebyshev: $\mathbb{P}(|x - \mathbb{E}(x)| \geq a) \leq \frac{Var(x)}{a^2}$ → 2^{nd} Moment

Def. The r^{th} moment of x is $\mathbb{E}(x^r)$ for $r \in \mathbb{N}_{>0}$.

- Can we use higher moments for better bounds?

For a non-negative even integer r:

$$\mathbb{P}(|x| \geq a) = \mathbb{P}(x^r \geq a^r) \leq \frac{\mathbb{E}(x^r)}{a^r}.$$

Polynomial Drop

- Even better?

Chernoff's Inequality

- Focus: n independet binomial random variables
- $x_1, x_2, \ldots, x_n$:

Bernoulli ← $x_i = \begin{cases} 0 & \text{with prob. } 1-p \\ 1 & \text{with prob. } p \end{cases}$

- Interest: $s \triangleq x_1 + x_2 + \ldots + x_n$
- $\mathbb{E}(s) = \sum_{i=1}^{n} \mathbb{E}(x_i) = np$ → μ

Thm. For any $\delta > 0$:

$$\mathbb{P}(s > (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \right)^{\mu}.$$

Moment Generating Function

Def. For a random variable x, the moment-generating function is:

$$M_x(\lambda) \triangleq \mathbb{E}(e^{\lambda x})$$

- $\frac{d^{(r)}M}{(d\lambda)^r}(0) = \boxed{\mathbb{E}(x^r)}$ r^{th} Moment of x
- Chernoff uses moment-generating functions

Proof of Chernoff's Inequality

Target:

$$\mathbb{P}(s > (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}.$$

Proof. For $\lambda > 0$, $e^{\lambda x}$ in monotone.

$$\mathbb{P}(s > (1+\delta)\mu) = \mathbb{P}(e^{\lambda s} > e^{\lambda(1+\delta)\mu})$$

By Markov's Inequality

$$\leq e^{-\lambda(1+\delta)\mu} \cdot \mathbb{E}(e^{\lambda s})$$

$$\mathbb{E}(e^{\lambda s}) = \mathbb{E}\left(e^{\lambda \sum_{i=1}^{n} x_i}\right) = \mathbb{E}\left(\prod_{i=1}^{n} e^{\lambda x_i}\right) = \prod_{i=1}^{n} \mathbb{E}(e^{\lambda x_i})$$

Why?

Proof of Chernoff's Inequality

Target:

$$\mathbb{P}(s > (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}.$$

Proof. For $\lambda > 0$, $e^{\lambda x}$ in monotone.

$$\mathbb{P}(s > (1+\delta)\mu) = \mathbb{P}(e^{\lambda s} > e^{\lambda(1+\delta)\mu})$$

By Markov's Inequality ←

$$\leq e^{-\lambda(1+\delta)\mu} \cdot \mathbb{E}(e^{\lambda s})$$

$$\mathbb{E}(e^{\lambda s}) = \mathbb{E}\left(e^{\lambda \sum_{i=1}^{n} x_i}\right) = \mathbb{E}\left(\prod_{i=1}^{n} e^{\lambda x_i}\right) = \prod_{i=1}^{n} \mathbb{E}(e^{\lambda x_i})$$

Independence

$\mathbb{E}(x \cdot y) = \mathbb{E}(x) \cdot \mathbb{E}(y)$

Proof of Chernoff's Inequality

Target:

$$\mathbb{P}(s > (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \right)^{\mu}.$$

Proof. $\mathbb{E}(e^{\lambda x_i})$ → with prob. p: $e^{\lambda \times 1} = e^{\lambda}$

→ with prob. $1-p$: $e^{\lambda \times 0} = 1$

$$\mathbb{E}(e^{\lambda s}) = \prod_{i=1}^{n} \mathbb{E}(e^{\lambda x_i})$$

$$= \prod_{i=1}^{n} (e^{\lambda} p + 1 - p)$$

$$= \prod_{i=1}^{n} (p(e^{\lambda} - 1) + 1) < \prod_{i=1}^{n} e^{p(e^{\lambda} - 1)}$$

$1 + x < e^{x}$

Proof of Chernoff's Inequality

Target:

$$\mathbb{P}(s > (1+\delta)\mu) \le \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}.$$

Proof.

- $\mathbb{P}(s > (1+\delta)\mu) \le e^{-\lambda(1+\delta)\mu} \cdot \mathbb{E}(e^{\lambda s})$
- $\mathbb{E}(e^{\lambda s}) = \prod_{i=1}^{n} \mathbb{E}(e^{\lambda x_i})$
- $\prod_{i=1}^{n} \mathbb{E}(e^{\lambda x_i}) < \prod_{i=1}^{n} e^{p(e^{\lambda}-1)}$

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

$$\mathbb{P}(s > (1+\delta)\mu) \le e^{-\lambda(1+\delta)\mu} \cdot \prod_{i=1}^{n} e^{p(e^{\lambda}-1)}$$

Agenda

1. Preliminary Definitions
2. Markov's and Chebyshev's Inequalities
   - The Law of Large Numbers
3. Higher Moments and Chernoff's Inequality
4. Master Tail Bounds
   - Applications

A Unification Theorem

- Chernoff's Bound
- Gaussian Annulus
- Power Law
- $\ldots$

Special cases of Master Tail Theorem

Master Tail Theorem

Thm. Let $x_1, x_2, \ldots, x_n$:

- be mutually independent random variables,
- have $\mathbb{E}(x_i) = 0$ and $Var(x_i) \leq \sigma^2$.

Suppose

- $a \in [0, \sqrt{2}n\sigma^2]$,
- $s \leq n\sigma^2/2$ is a positive **even** integer,
- $|\mathbb{E}(x_i^r)| \leq \sigma^2 r!$ for $r = 3, 4, \ldots, s$:

$$\mathbb{P}(|x_1 + x_2 + \ldots + x_n| \geq a) \leq \left(\frac{2sn\sigma^2}{a^2}\right)^{s/2}.$$

Further, if $s \geq a^2/(4n\sigma^2)$, then

$$\mathbb{P}(|x_1 + x_2 + \ldots + x_n| \geq a) \leq 3e^{-a^2/(12n\sigma^2)}.$$

Master Tail Bound $\Rightarrow$ Chernoff

- Slightly weaker version:

Thm. Let $y_1, y_2, \ldots, y_n$ be independent Bernoulli variables with $\mathbb{E}(y_i) = p$. For any $c \in [0,1]$

$$\mathbb{P}(|y - \underbrace{\mathbb{E}(y)}_{np}| \geq c\mu) \leq 3e^{-\mu c^2/12},$$

where $y = y_1 + y_2 + \ldots + y_n$, and $\mu = np$.

Master Tail Bound $\Rightarrow$ Chernoff

- Slightly weaker version:

Thm. Let $y_1, y_2, \ldots, y_n$ be independent Bernoulli variables with $\mathbb{E}(y_i) = p$. For any $c \in [0,1]$

$$\mathbb{P}(|y - \mathbb{E}(y)| \geq c\mu) \leq 3e^{-\mu c^2/12},$$

where $y = y_1 + y_2 + \ldots + y_n$, and $\mu = np$.

- Need variables with $\mathbb{E}(x) = 0$.
- $x_i \stackrel{\Delta}{=} y_i - p$

For $s = 2$, $\mathbb{E}(x_i^2) = Var(x_i) = p(1-p)$

- $|\mathbb{E}(x_i^s)| = |\mathbb{E}((y_i - p)^s)| = |p(1-p)((1-p)^{s-1} - (-p)^{s-1})|$

$(1-p)^s$ p $1-p$ $(0-p)^s$

Master Tail Bound $\Rightarrow$ Chernoff

- Slightly weaker version:

> *Thm.* Let $y_1, y_2, \ldots, y_n$ be independent Bernoulli variables with $\mathbb{E}(y_i) = p$. For any $c \in [0,1]$
>
> $$\mathbb{P}(|y - \mathbb{E}(y)| \geq c\mu) \leq 3e^{-\mu c^2/12},$$
>
> where $y = y_1 + y_2 + \ldots + y_n$, and $\mu = np$.

- Need variables with $\mathbb{E}(x) = 0$.
- $x_i \stackrel{\Delta}{=} y_i - p$

For all $s \geq 2$: $\in [-1,1]$

- $|\mathbb{E}(x_i^s)| = |\mathbb{E}((y_i - p)^s)| = |p(1-p)((1-p)^{s-1} - (-p)^{s-1})|$

$(1-p)^s$ p $1-p$ $(0-p)^s$

Master Tail Bound $\Rightarrow$ Chernoff

- Slightly weaker version:

Thm. Let $y_1, y_2, \ldots, y_n$ be independent Bernoulli variables with $\mathbb{E}(y_i) = p$. For any $c \in [0, 1]$

$$\mathbb{P}(|y - \mathbb{E}(y)| \geq c\mu) \leq 3e^{-\mu c^2/12},$$

where $y = y_1 + y_2 + \ldots + y_n$, and $\mu = np$.

- Need variables with $\mathbb{E}(x) = 0$.

- $x_i \triangleq y_i - p$

- $|\mathbb{E}(x_i^s)| \leq p$

Master Bound $\Rightarrow$ Chernoff

- $Var(x_i) \leq \sigma^2$ ⟶ $\sigma^2 \leftarrow p$
- $a < \sqrt{2}n\sigma^2$ ⟶ $a \leftarrow c\mu$
- $s \geq a^2/(4n\sigma^2)$ ⟶ $s \leftarrow \mu c^2/4$

$$
\begin{aligned}
&\mathbb{P}(|x_1 + x_2 + \ldots + x_n| \geq a) \\
=&\mathbb{P}(|y - \mathbb{E}(y)| \geq c\mu) \\
\leq&3e^{-a^2/(12n\sigma^2)} \\
=&3e^{-\frac{c^2n^2p^2}{12np}} \\
=&3e^{-\mu c^2/12}
\end{aligned}
$$

Using the second inequality

Agenda

1. Preliminary Definitions
2. Markov's and Chebyshev's Inequalities
   - The Law of Large Numbers
3. Higher Moments and Chernoff's Inequality
4. Master Tail Bounds
   - Applications

Summary

- PDFs, and Tail Bounds
- Markov's Inequality (1st Moment)
- Chebyshev's Inequality (2nd Moment)
- The Law of Large Numbers
- Chernoff's Inequality (Moment Generating Functions)
- Master Tail Bounds Theorem
- Alternative Proof of Chernoff