Chapter 18
Tail Bounds

Tails

Defn: The **tail** of random variable X is $\boldsymbol{P}\{X > x\}$.

Examples of why we care about tails:

- Fraction of jobs that queue more than 24 hours
- Fraction of packets that find the router buffer full
- Fraction of hash buckets that have more than 10 items

Unfortunately, determining the tail of even simple r.v.s is often hard – much harder than determining the mean or transform!

Tails Example

Q: Suppose you're distributing n jobs among n servers at random. What's the probability that a particular server gets $\geq k$ jobs?

$$X \sim Binomial(n, p)$$

$$\boldsymbol{P}\{X \geq k\} = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

No closed-form known for this

Tails Example

Q: Jobs arrive to a datacenter according to a Poisson process with rate λ jobs/hour. What's the probability that $\geq k$ jobs arrive during the first hour?

$$X \sim Poisson(\lambda)$$

$$\boldsymbol{P}\{X \geq k\} = \sum_{i=k}^{\infty} e^{-\lambda} \cdot \frac{\lambda^i}{i!}$$

No closed-form known for this

Tails Bounds

Rather than directly compute tails, we will derive upper bounds on the tails, called **tail bounds**!

$$\mathbf{P}\{X \geq k\} = \sum_{i=k}^{\infty} e^{-\lambda} \cdot \frac{\lambda^i}{i!}$$

$$\mathbf{P}\{X \geq k\} = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

We'll soon have tail bounds for both of these!

<u>Definition</u>**:** An upper bound on $\mathbf{P}\{X \geq k\}$ is called a **tail bound**. An upper bound on $\mathbf{P}\{|X - \mu| \geq k\}$ where $\mu = \mathbf{E}[X]$ is called a **concentration bound** or **concentration inequality**.

Running Example

We will develop progressively better (tighter) **tail bounds**.

We will test each bound on the following running example:

Flip a fair coin n times:

Q: What's a tail bound on the probability of getting at least $\frac{3}{4}n$ heads?

Markov's inequality

Theorem: (Markov's inequality) If r.v. X is non-negative, with finite mean $\mu = \boldsymbol{E}[X]$, then $\forall a > 0$,

$$\boldsymbol{P}\{X \geq a\} \leq \frac{\mu}{a}$$

Proof:

$$\begin{aligned}\boldsymbol{E}[X] = \sum_{x=0}^{\infty} x \cdot p_X(x) &\geq \sum_{x=a}^{\infty} x \cdot p_X(x) \\ &\geq \sum_{x=a}^{\infty} a \cdot p_X(x) \\ &= a \sum_{x=a}^{\infty} p_X(x) = a \cdot \boldsymbol{P}\{X \geq a\}\end{aligned}$$

Markov's Inequality on Running Example

Flip a fair coin n times: 

Q: What's a tail bound on the probability of getting at least $\frac{3}{4}n$ heads?

$$X = \text{Number Heads} \sim Binomial\left(n, \frac{1}{2}\right) \qquad \mu = \boldsymbol{E}[X] = \frac{n}{2}$$

$$\boldsymbol{P}\left\{X \geq \frac{3}{4}n\right\} \leq \frac{\mu}{\frac{3}{4}n} = \frac{\frac{n}{2}}{\frac{3}{4}n} = \frac{2}{3}$$

Clearly a terrible bound because doesn't involve n

Chebyshev's inequality

Theorem: (Chebyshev's inequality) Let X be any r.v. with finite mean, μ, and finite variance. Then $\forall a > 0$,

$$\boldsymbol{P}\{|X-\mu| \geq a\} \leq \frac{\boldsymbol{Var}(X)}{a^2}$$

Proof:

Q: Can you see how to apply Markov's inequality here?

$$\boldsymbol{P}\{|X-\mu| \geq a\} = \boldsymbol{P}\{(X-\mu)^2 \geq a^2\}$$

$$\leq \frac{\boldsymbol{E}[(X-\mu)^2]}{a^2}$$

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

Chebyshev's Bound on Running Example

Flip a fair coin n times:

Q: What's a tail bound on the probability of getting at least $\frac{3}{4}n$ heads?

$$X = \text{Number Heads} \sim Binomial\left(n, \frac{1}{2}\right) \qquad \mu = E[X] = \frac{n}{2} \qquad \boldsymbol{Var}(X) = \frac{n}{4}$$

$$\boldsymbol{P}\left\{X \geq \frac{3}{4}n\right\} = \boldsymbol{P}\left\{X - \frac{n}{2} \geq \frac{n}{4}\right\} = \frac{1}{2}\boldsymbol{P}\left\{\left|X - \frac{n}{2}\right| \geq \frac{n}{4}\right\} \leq \frac{1}{2} \cdot \frac{\boldsymbol{Var}(X)}{\left(\frac{n}{4}\right)^2} = \frac{1}{2} \cdot \frac{\frac{n}{4}}{\left(\frac{n}{4}\right)^2} = \frac{2}{n}$$

Why?

At least decreases with n

Chernoff Bound

In deriving the Chebyshev bound, we **squared** the r.v. and then applied Markov.

In deriving the Chernoff bound, we **exponentiate** the r.v. and then apply Markov.

$\forall t > 0$:

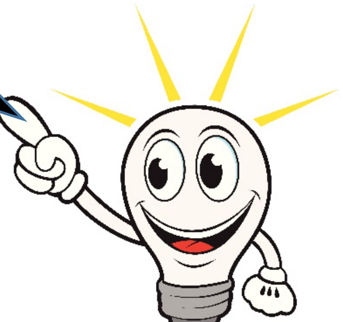
$$\mathbf{P}\{X \geq a\} = \mathbf{P}\{tX \geq ta\}$$

$$= \mathbf{P}\{e^{tX} \geq e^{ta}\}$$

$$\leq \frac{\mathbf{E}[e^{tX}]}{e^{ta}}$$

Why are we allowed to apply Markov to this?

But because this bound holds $\forall t$, it also holds for the minimizing t.



Chernoff Bound

In deriving the Chebyshev bound, we **squared** the r.v. and then applied Markov.

In deriving the Chernoff bound, we **exponentiate** the r.v. and then apply Markov.

$\forall t > 0$:

$$\boldsymbol{P}\{X \geq a\} = \boldsymbol{P}\{tX \geq ta\}$$

$$= \boldsymbol{P}\{e^{tX} \geq e^{ta}\}$$

$$\leq \frac{\boldsymbol{E}[e^{tX}]}{e^{ta}}$$

Theorem 18.3: (Chernoff bound) Let X be any r.v. and a be a constant. Then

$$\boldsymbol{P}\{X \geq a\} \leq \min_{t>0} \left\{ \frac{\boldsymbol{E}[e^{tX}]}{e^{ta}} \right\}$$

Chernoff Bound

In deriving the Chebyshev bound, we **squared** the r.v. and then applied Markov.

In deriving the Chernoff bound, we **exponentiate** the r.v. and then apply Markov.

$\forall t > 0$:

$$\boldsymbol{P}\{X \geq a\} = \boldsymbol{P}\{tX \geq ta\}$$

$$= \boldsymbol{P}\{e^{tX} \geq e^{ta}\}$$

$$\leq \frac{\boldsymbol{E}[e^{tX}]}{e^{ta}}$$

Q: Why do we expect the Chernoff bound to be stronger than the others?

Theorem: (Chernoff bound) Let X be any r.v. and a be a constant. Then

$$\boldsymbol{P}\{X \geq a\} \leq \min_{t>0} \left\{ \frac{\boldsymbol{E}[e^{tX}]}{e^{ta}} \right\}$$

A: Looks a lot like an onion!

Chernoff Bound on c.d.f.

Q: What do we do if we want to upper bound $\boldsymbol{P}\{X \leq a\}$?

$\forall t < 0$:

$$\boldsymbol{P}\{X \leq a\} = \boldsymbol{P}\{tX \geq ta\}$$

$$= \boldsymbol{P}\{e^{tX} \geq e^{ta}\}$$

$$\leq \frac{\boldsymbol{E}[e^{tX}]}{e^{ta}}$$

Theorem: (Chernoff bound on c.d.f.) Let X be any r.v. and a be a constant. Then

$$\boldsymbol{P}\{X \leq a\} \leq \min_{t<0} \left\{ \frac{\boldsymbol{E}[e^{tX}]}{e^{ta}} \right\}$$

Chernoff Bound for Poisson Tail

Goal: Bound tail of $X \sim Poisson(\lambda)$

Step 1: Derive $\boldsymbol{E}[e^{tX}]$ where $t > 0$

$$\boldsymbol{E}[e^{tX}] = \sum_{i=0}^{\infty} e^{ti} \cdot \frac{e^{-\lambda} \cdot \lambda^i}{i!}$$

$$= e^{-\lambda} \sum_{i=0}^{\infty} \frac{(\lambda e^t)^i}{i!}$$

$$= e^{-\lambda} \cdot e^{\lambda e^t}$$

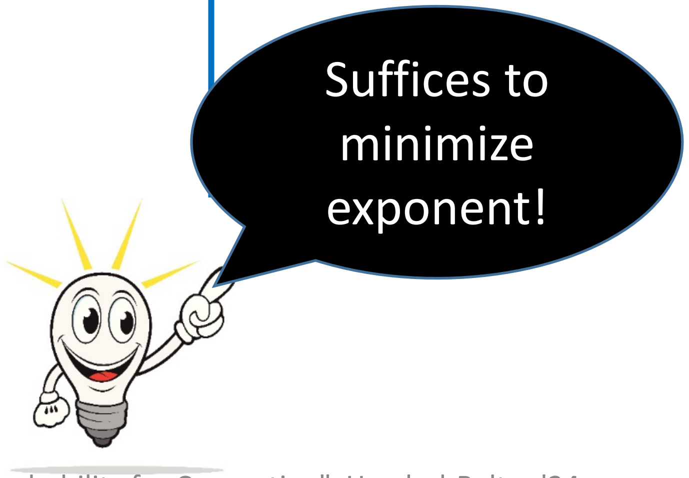
$$= e^{\lambda(e^t - 1)}$$

Step 2: Let $a > \lambda$. Bound $\boldsymbol{P}\{X \geq a\}$

$$\boldsymbol{P}\{X \geq a\} \leq \min_{t>0} \left\{ \frac{\boldsymbol{E}[e^{tX}]}{e^{ta}} \right\}$$

$$= \min_{t>0} \left\{ \frac{e^{\lambda(e^t-1)}}{e^{ta}} \right\}$$

$$= \min_{t>0} \left\{ e^{\lambda(e^t-1)-ta} \right\}$$



Chernoff Bound for Poisson Tail

Goal: Bound tail of $X \sim Poisson(\lambda)$

Step 2: Let $a > \lambda$. Bound $\boldsymbol{P}\{X \geq a\}$

$$\boldsymbol{P}\{X \geq a\} \leq \min_{t>0} \left\{ \frac{\boldsymbol{E}[e^{tX}]}{e^{ta}} \right\}$$

$$= \min_{t>0} \left\{ \frac{e^{\lambda(e^t - 1)}}{e^{ta}} \right\}$$

$$= \min_{t>0} \left\{ e^{\lambda(e^t-1)-ta} \right\}$$

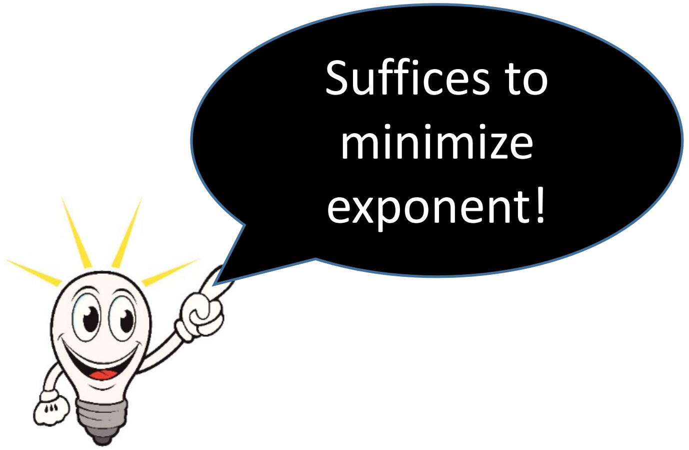
- Exponent is minimized at $t = ln\left(\frac{a}{\lambda}\right)$

  Thus:

- $\boldsymbol{P}\{X \geq a\} \leq e^{\lambda(e^t-1)-ta}$, at $t = ln\left(\frac{a}{\lambda}\right)$

$$= e^{\lambda\left(\frac{a}{\lambda}-1\right)-a ln\left(\frac{a}{\lambda}\right)}$$

$$= \boxed{e^{a-\lambda} \cdot \left(\frac{\lambda}{a}\right)^a}$$



Chernoff Bound for Binomial

Theorem 18.4: (Pretty Chernoff Bound for Binomial)

Let $X \sim Binomial(n, p)$ where $\mu = \boldsymbol{E}[X] = np$. Then, for any $\delta > 0$,

$$\boldsymbol{P}\{X - np \geq \delta\} \leq e^{-2\delta^2/n}$$

$$\boldsymbol{P}\{X - np \leq -\delta\} \leq e^{-2\delta^2/n}$$

We will prove this soon, but let's try applying it first!

Bound is strongest when $\delta = \Theta(n)$
Try to use it in this regime.

Chernoff Bound on Running Example

Flip a fair coin n times:

Q: What's a tail bound on the probability of getting at least $\frac{3}{4}n$ heads?

$$X = \text{Number Heads} \sim Binomial\left(n, \frac{1}{2}\right) \qquad \mu = E[X] = \frac{n}{2}$$

$$\boldsymbol{P}\left\{X \geq \frac{3}{4}n\right\} = \boldsymbol{P}\left\{X - \frac{n}{2} \geq \frac{n}{4}\right\} \leq e^{-2\left(\frac{n}{4}\right)^2 \cdot \frac{1}{n}} = e^{-\frac{n}{8}}$$

Decreases exponentially fast in n

Note $\delta = \frac{n}{4} = \Theta(n)$

Comparing the bounds

Flip a fair coin n times: 

Q: What's a tail bound on the probability of getting at least $\frac{3}{4}n$ heads?

Q: What is the exact answer?

$$\boldsymbol{P}\left\{X \geq \frac{3}{4}n\right\} = \sum_{i=\frac{3}{4}n}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^{i} \left(\frac{1}{2}\right)^{n-i} = \boxed{2^{-n} \sum_{i=\frac{3}{4}n}^{n} \binom{n}{i}}$$

Central Limit Theorem

Flip a fair coin n times: 

Q: What's a tail bound on the probability of getting at least $\frac{3}{4}n$ heads?

CLT applies because adding i.i.d. r.v.s

$$X \sim Binomial\left(n, \frac{1}{2}\right)$$

$$\mu = \boldsymbol{E}[X] = \frac{n}{2}$$

$$\boldsymbol{Var}(X) = \frac{n}{4}$$

$$\sigma_X = \sqrt{\frac{n}{4}}$$

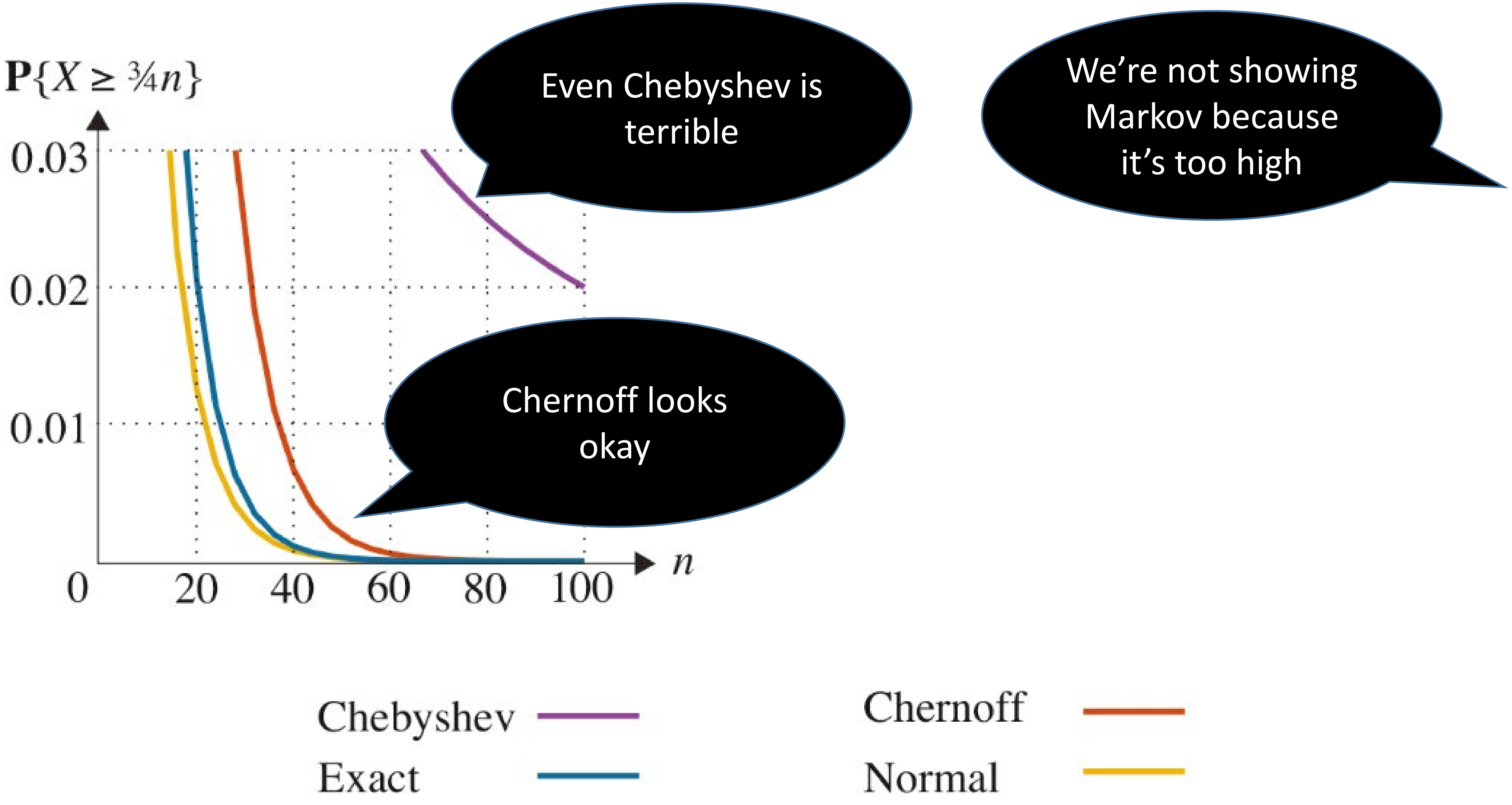
$$\boldsymbol{P}\left\{X \geq \frac{3}{4}n\right\} = \boldsymbol{P}\left\{X - \frac{n}{2} \geq \frac{n}{4}\right\}$$

$$= \boldsymbol{P}\left\{\frac{X - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \geq \frac{\frac{n}{4}}{\sqrt{\frac{n}{4}}}\right\}$$

$$\approx \boldsymbol{P}\left\{Normal(0,1) \geq \sqrt{\frac{n}{4}}\right\} = 1 - \boldsymbol{\Phi}\left(\sqrt{\frac{n}{4}}\right)$$

Result is approximation not bound

Comparing the approximation and bounds

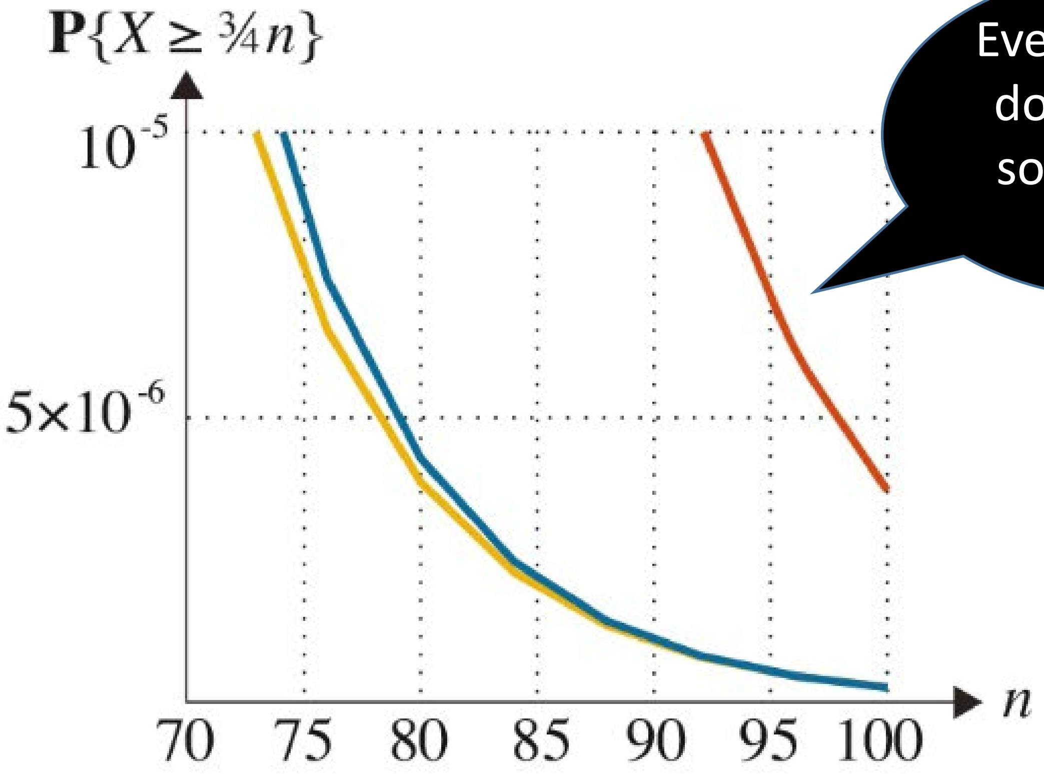


"Introduction to Probability for Computing", Harchol-Balter '24

Comparing the approximation and bounds

An up close view at higher $n > 70$

Chebyshev no longer visible. It's too high



Even Chernoff doesn't look so great any more

Proof of Thm 18.4 – Pretty Chernoff Bound

Theorem 18.4: (Pretty Chernoff Bound for Binomial)

Let $X \sim Binomial(n, p)$ where $\mu = \boldsymbol{E}[X] = np$. Then, for any $\delta > 0$,

$$\boldsymbol{P}\{X - np \geq \delta\} \leq e^{-2\delta^2/n}$$

$$\boldsymbol{P}\{X - np \leq -\delta\} \leq e^{-2\delta^2/n}$$

We will now prove Thm 18.4 (top half). The bottom half is an Exercise in your book.

Our proof requires using Lemma 18.5, which is proven in your book.

Lemma 18.5: For any $t > 0$ and $0 < p < 1$ and $q = 1 - p$, we have that:

$$pe^{tq} + qe^{-tp} \leq e^{t^2/8}$$

Proof of Thm 18.4 – Pretty Chernoff Bound

Theorem 18.4: Let $X \sim Binomial(n,p)$ where $\mu = \boldsymbol{E}[X] = np$. Then, for any $\delta > 0$,

$$\boldsymbol{P}\{X - np \geq \delta\} \leq e^{-2\delta^2/n}$$

Proof: For any $t > 0$,

$$\begin{aligned}
\boldsymbol{P}\{X - np \geq \delta\} &= \boldsymbol{P}\{t(X - np) \geq t\delta\} \\
&= \boldsymbol{P}\left\{e^{t(X-np)} \geq e^{t\delta}\right\} \\
&\leq e^{-t\delta}\boldsymbol{E}\left[e^{t(X-np)}\right] = e^{-t\delta}\boldsymbol{E}\left[e^{t((X_1-p)+(X_2-p)+\cdots+(X_n-p))}\right] \\
&= e^{-t\delta} \cdot \prod_{i=1}^{n} \boldsymbol{E}\left[e^{t(X_i-p)}\right]
\end{aligned}$$

We can break this up!
$X = \sum_{i=1}^{n} X_i$ where
$X_i \sim Bernoulli(p)$

Proof of Thm 18.4 – Pretty Chernoff Bound

Theorem 18.4: Let $X \sim Binomial(n, p)$ where $\mu = \boldsymbol{E}[X] = np$. Then, for any $\delta > 0$,

$$\boldsymbol{P}\{X - np \geq \delta\} \leq e^{-2\delta^2/n}$$

Proof, cont: So, for any $t > 0$,

$$\boldsymbol{P}\{X - np \geq \delta\} \leq e^{-t\delta} \prod_{i=1}^{n} \boldsymbol{E}\left[e^{t(X_i - p)}\right]$$

$$= e^{-t\delta} \prod_{i=1}^{n} \left(p \cdot e^{t(1-p)} + (1-p) \cdot e^{-tp}\right)$$

by Lemma 18.5

$$= e^{-t\delta} \prod_{i=1}^{n} \left(e^{t^2/8}\right) = e^{-t\delta + nt^2/8}$$

Q: What do we do next?

A: Find the minimizing t

Proof of Thm 18.4 – Pretty Chernoff Bound

Theorem 18.4: Let $X \sim Binomial(n, p)$ where $\mu = \boldsymbol{E}[X] = np$. Then, for any $\delta > 0$,

$$\boldsymbol{P}\{X - np \geq \delta\} \leq e^{-2\delta^2/n}$$

Proof, cont: So, for any $t > 0$,

$$\boldsymbol{P}\{X - np \geq \delta\} \leq \boxed{e^{-t\delta + nt^2/8}}$$

The exponent is minimized at

$$t = \frac{4\delta}{n}$$

$$\Rightarrow \boldsymbol{P}\{X - np \geq \delta\} \leq e^{-\left(\frac{4\delta}{n}\right)\delta + n\left(\frac{4\delta}{n}\right)^2/8} = \boxed{e^{-2\delta^2/n}}$$

Stronger (?) Chernoff Bound for Binomial

Theorem 18.6 presents an alternative, sometime stronger, bound.
The bound holds for a more general definition of a Binomial.

Theorem 18.6: (Sometimes stronger Chernoff Bound)

Let $X = \sum_{i=1}^{n} X_i$ where $X_i \sim Bernoulli(p_i)$ and $\mu = \boldsymbol{E}[X] = \sum_{i=1}^{n} p_i$. Then, $\forall \epsilon > 0$,

$$\boldsymbol{P}\{X \leq (1-\epsilon)\mu\} \leq \left(\frac{e^{\epsilon}}{(1+\epsilon)^{(1+\epsilon)}}\right)^{\mu}$$

Stronger (?) Chernoff Bound for Binomial

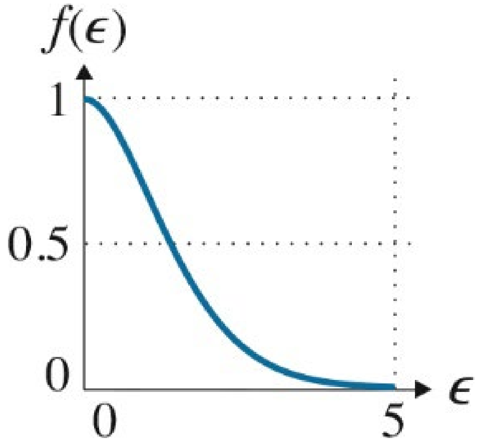
Theorem 18.6: (Sometimes stronger Chernoff Bound)

Let $X = \sum_{i=1}^{n} X_i$ where $X_i \sim Bernoulli(p_i)$ and $\mu = \boldsymbol{E}[X] = \sum_{i=1}^{n} p_i$. Then, $\forall \epsilon > 0$,

$$\boldsymbol{P}\{X \geq (1-\epsilon)\mu\} < \left(\frac{e^{\epsilon}}{(1+\epsilon)^{(1+\epsilon)}}\right)^{\mu}$$

Plot of inner term:

$$f(\epsilon) = \frac{e^{\epsilon}}{(1+\epsilon)^{(1+\epsilon)}}$$



Two observations:

1. $f(\epsilon) < 1$, so bound is exponentially decreasing.

2. Bound in Thm 18.6 is particularly strong when ϵ is high.

Comparison of Chernoff Bounds

Theorem 18.4: (Pretty bound)
Let $X \sim Binomial(n, p)$ where $\mu = \boldsymbol{E}[X] = np$. Then, for any $\delta > 0$,

$$\boldsymbol{P}\{X - np \geq \delta\} \leq e^{-2\delta^2/n}$$

Theorem 18.6: (Sometimes stronger bound)
Let $X = \sum_{i=1}^{n} X_i$ where $X_i \sim Bernoulli(p_i)$ and $\mu = \boldsymbol{E}[X] = \sum_{i=1}^{n} p_i$. Then, $\forall \epsilon > 0$,

$$\boldsymbol{P}\{X \geq (1+\epsilon)\mu\} < \left(\frac{e^{\epsilon}}{(1+\epsilon)^{(1+\epsilon)}}\right)^{\mu}$$

Q: Which gives best bound on probability of getting $\geq \frac{3}{4}n$ heads, when flipping fair coin n times?

$$\boldsymbol{P}\left\{X \geq \frac{3n}{4}\right\} = \boldsymbol{P}\left\{X - \frac{n}{2} \geq \frac{n}{4}\right\} \leq e^{-\frac{n}{8}}$$

This is the better bound!

$$\boldsymbol{P}\left\{X \geq \frac{3n}{4}\right\} = \boldsymbol{P}\left\{X \geq \left(1 + \frac{1}{2}\right) \cdot \frac{n}{2}\right\} \leq \left(\frac{e^{0.5}}{1.5^{1.5}}\right)^{n/2} \approx (1.54)^{-\frac{n}{8}}$$

Comparison of Chernoff Bounds

Theorem 18.4: (Pretty bound)
Let $X \sim Binomial(n, p)$ where $\mu = \boldsymbol{E}[X] = np$. Then, for any $\delta > 0$,

$$\boldsymbol{P}\{X - np \geq \delta\} \leq e^{-2\delta^2/n}$$

Theorem 18.6: (Sometimes stronger bound)
Let $X = \sum_{i=1}^{n} X_i$ where $X_i \sim Bernoulli(p_i)$ and $\mu = \boldsymbol{E}[X] = \sum_{i=1}^{n} p_i$. Then, $\forall \epsilon > 0$,

$$\boldsymbol{P}\{X \geq (1+\epsilon)\mu\} < \left(\frac{e^{\epsilon}}{(1+\epsilon)^{(1+\epsilon)}}\right)^{\mu}$$

Q: Which is the better bound on $\boldsymbol{P}\{X \geq 21\}$ if $p_i = p = \frac{1}{n}$?

$$\boldsymbol{P}\{X \geq 21\} = \boldsymbol{P}\{X - 1 \geq 20\}$$

$$\leq e^{-\frac{2\cdot(20)^2}{n}}$$

$$\leq e^{-\frac{800}{n}} \to \boxed{1}$$

$$\boldsymbol{P}\{X \geq 21\} = \boldsymbol{P}\{(X \geq (1+20)\cdot 1\}$$

$$\leq \frac{e^{20}}{21^{21}}$$

$$\approx \boxed{8.3 \cdot 10^{-20}}$$

Much better bound!

More general bound: Hoeffding's Inequality

Theorem 18.7: (Hoeffding's Inequality)

Let $X_1, X_2, \ldots, X_n$ be independent r.v.s, where $a_i \leq X_i \leq b_i, \ \forall i$.

Let:

$$X = \sum_{i=1}^{n} X_i$$

More general because X_i's don't have to be independent

Then,

$$\boldsymbol{P}\{X - \boldsymbol{E}[X] \geq \delta\} \leq \ exp\left(-\frac{2\,\delta^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

$$\boldsymbol{P}\{X - \boldsymbol{E}[X] \leq -\delta\} \leq \ exp\left(-\frac{2\,\delta^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$