For a random variable $X$, the **tails** of $X$ are the parts of the PMF that are “far” from its mean.
PMF for $X \sim \text{Bin}(100,0.5)$
heavy-tailed distribution

Possible?
heavy-tailed distribution
Often, we want to bound the probability that a random variable $X$ is “extreme.” Perhaps:

\[
P(X > \alpha) < \frac{1}{\alpha^3}
\]

\[
P(X > E[X] + t) < e^{-t}
\]

\[
P(\left|X - E[X]\right| > t) < \frac{1}{\sqrt{t}}
\]
We know that randomized quicksort runs in $O(n \log n)$ expected time. But what’s the probability that it takes more than $10 n \log(n)$ steps? More than $n^{1.5}$ steps?

If we know the expected advertising cost is $1500/day, what’s the probability we go over budget? By a factor of 4?

I only expect 10,000 homeowners to default on their mortgages. What’s the probability that 1,000,000 homeowners default?
“Lake Wobegon, Minnesota, where all the women are strong, all the men are good looking, and 

*all the children are above average*…”
In general, an *arbitrary* random variable could have very bad behavior. But knowledge is power; if we know *something*, can we bound the badness?

**Suppose we know that** $X$ **is always non-negative.**

**Theorem:** If $X$ is a non-negative random variable, then for every $\alpha > 0$, we have

$$P(X \geq \alpha) \leq \frac{E[X]}{\alpha}$$

**Corr:**

$$P(X \geq \alpha E[X]) \leq 1/\alpha$$
Markov’s inequality

**Theorem:** If $X$ is a non-negative random variable, then for every $\alpha > 0$, we have

$$P(X \geq \alpha) \leq \frac{E[X]}{\alpha}$$

Example: if $X = \text{daily advertising expenses}$ and $E[X] = 1500$

Then, by Markov’s inequality,

$$P(X \geq 6000) \leq \frac{1500}{6000} = 0.25$$
**Markov’s inequality**

**Theorem:** If $X$ is a non-negative random variable, then for every $\alpha > 0$, we have

$$P(X \geq \alpha) \leq \frac{E[X]}{\alpha}$$

Example: if $X = \text{time to quicksort } n \text{ items}$, expectation $E[X] \approx 1.4 n \log n$. What’s probability that it takes > 4 times as long as expected?

By Markov’s inequality:

$$P(X \geq 4 \cdot E[X]) \leq E[X]/(4 \cdot E[X]) = 1/4$$
Markov’s inequality

**Theorem:** If $X$ is a non-negative random variable, then for every $\alpha > 0$, we have

$$P(X \geq \alpha) \leq \frac{E[X]}{\alpha}$$

**Proof:**

$$E[X] = \sum_x xP(x)$$

$$= \sum_{x<\alpha} xP(x) + \sum_{x\geq\alpha} xP(x)$$

$$\geq 0 + \sum_{x\geq\alpha} \alpha P(x) \quad (x \geq 0; \alpha \leq x)$$

$$= \alpha P(X \geq \alpha)$$
Markov’s inequality

**Theorem:** If $X$ is a non-negative random variable, then for every $\alpha > 0$, we have

$$E[X] = \sum x \cdot P(x) \geq \sum_{x \geq \alpha} x \cdot P(x) \geq \alpha \cdot P(X \geq \alpha) \quad (x \geq 0; \alpha \leq x)$$

**Proof:**

$$E[X] = \sum x \cdot P(x) \geq \sum_{x \geq \alpha} x \cdot P(x) \geq \alpha \cdot P(X \geq \alpha)$$
Chebyshev’s inequality

If we know more about a random variable, we can often use that to get better tail bounds.

Suppose we also know the variance.

**Theorem:** If Y is an arbitrary random variable with \( \mathbb{E}[Y] = \mu \), then, for any \( \alpha > 0 \),

\[
P(|Y - \mu| \geq \alpha) \leq \frac{\text{Var}[Y]}{\alpha^2}
\]
Theorem: If $Y$ is an arbitrary random variable with $\mu = E[Y]$, then, for any $\alpha > 0$,

$$P(|Y - \mu| \geq \alpha) \leq \frac{\text{Var}[Y]}{\alpha^2}$$

Proof: Let $X = (Y - \mu)^2$.

$X$ is non-negative, so we can apply Markov’s inequality:

$$P(|Y - \mu| \geq \alpha) = P(X \geq \alpha^2) \leq \frac{E[X]}{\alpha^2} = \frac{\text{Var}[Y]}{\alpha^2}$$
Theorem: If $Y$ is an arbitrary random variable with $\mu = E[Y]$, then, for any $\alpha > 0$,

$$P(|Y - \mu| \geq \alpha) \leq \frac{\text{Var}[Y]}{\alpha^2}$$

Proof: Let $X = |Y - \mu|$. Since $X$ is non-negative, we can apply Markov's inequality:

$$P(X \geq \alpha) \leq \frac{\text{E}[X]}{\alpha} = \frac{\text{Var}[Y]}{\alpha^2}$$

Chebyshev’s inequality
Chebyshev’s inequality

\[ P\left( |Y - \mu| \geq \alpha \right) \leq \frac{\text{Var}[Y]}{\alpha^2} \]

E.g., suppose:

\( Y = \text{money spent on advertising in a day} \)

\( E[Y] = 1500 \)

\( \text{Var}[Y] = 500^2 \) (i.e. \( \text{SD}[Y] = 500 \))

\[ P(Y \geq 6000) = P\left( |Y - \mu| \geq 4500 \right) \leq \frac{500^2}{4500^2} = \frac{1}{81} \approx 0.012 \]
Chebyshev’s inequality

\[ P\left( |Y - \mu| \geq \alpha \right) \leq \frac{\text{Var}[Y]}{\alpha^2} \]

E.g., suppose:

\begin{align*}
Y &= \text{comparisons in quicksort for } n=1024 \\
E[Y] &= 1.4 \cdot n \cdot \log_2 n \approx 14000 \\
\text{Var}[Y] &= ((21-2\pi^2)/3) \cdot n^2 \approx 441000 \\
&\quad \text{(i.e. SD}[Y] \approx 664) \\
P(Y \geq 4\mu) &= P(Y-\mu \geq 3\mu) \leq \frac{\text{Var}(Y)}{(9\mu^2)} < 0.000242
\end{align*}

1000 times smaller than Markov but still overestimated?: \( \sigma/\mu \approx 5\% \), so \( 4\mu \approx \mu + 60\sigma \)
Chebyshev’s inequality

**Theorem:** If $Y$ is an arbitrary random variable with $\mu = \mathbb{E}[Y]$, then, for any $\alpha > 0,$

$$P\left( |Y - \mu| \geq \alpha \right) \leq \frac{\text{Var}[Y]}{\alpha^2}$$

**Corr:** If

$$\sigma = SD[Y] = \sqrt{\text{Var}[Y]}$$

Then:

$$P\left( |Y - \mu| \geq t\sigma \right) \leq \frac{\sigma^2}{t^2\sigma^2} = \frac{1}{t^2}$$
Chebyshev’s inequality

$$P(|Y - \mu| \geq t\sigma) \leq \frac{1}{t^2}$$

For comparison, normal & many others would decline exponentially in t, or faster I.e., Chebyshev is much weaker, but much more general$^{23}$
Y \sim \text{Bin}(15000, 0.1)
\mu = E[Y] = 1500, \sigma = \sqrt{\text{Var}(Y)} = 36.7

P(Y \geq 6000) = P(Y \geq 4\mu) \leq \frac{1}{4} \quad (\text{Markov})
P(Y \geq 6000) = P(Y-\mu \geq 122\sigma) \leq 7 \times 10^{-5} \quad (\text{Chebyshev})

Poisson approximation: Y \sim \text{Poi}(1500)
Rough computer calculation:

P(Y \geq 6000) \ll 10^{-1600}

And the exact value is \approx 4 \times 10^{-2031}
Suppose $X \sim \text{Bin}(n,p)$

$\mu = E[X] = pn$

Chernoff bound:

For any $0 < \delta < 1$,

$$P(X > (1 + \delta)\mu) \leq \exp\left(-\frac{\delta^2 \mu}{3}\right)$$

$$P(X < (1 - \delta)\mu) \leq \exp\left(-\frac{\delta^2 \mu}{2}\right)$$
Chernoff bounds

Suppose $X \sim \text{Bin}(n,p)$

$\mu = E[X] = np$

**Chernoff bound:**

For any $0 < \delta < 1$,

$$P(X > \mu + \delta \mu) \leq \exp \left( -\frac{\delta^2 \mu}{3} \right)$$

$$P(X < \mu - \delta \mu) \leq \exp \left( -\frac{\delta^2 \mu}{2} \right)$$
router buffers
**Model:** \( n = 100,000 \) computers each independently send a packet with probability \( p = 0.01 \) each second. The router processes its buffer every second. How many packet buffers so that router drops a packet:

- **Never?**
  
  \( 100,000 \)

- **With probability \( \approx \frac{1}{2} \), every second?**
  
  \( \approx 1000 \)  
  \( \approx 1000 \)  
  \( (P(X > E[X]) \approx \frac{1}{2} \text{ when } X \sim \text{Binomial}(100000, .01)) \)

- **With probability at most \( 10^{-6} \), every hour?**
  
  \( 1257 \)

- **With probability at most \( 10^{-6} \), every year?**
  
  \( 1305 \)

- **With probability at most \( 10^{-6} \), since Big Bang?**
  
  \( 1404 \)

Exercise: How would you formulate the exact answer to this problem in terms of binomial probabilities? Can you get a numerical answer?
$X \sim \text{Bin}(100,000, 0.01), \quad \mu = E[X] = 1000$

Let $p =$ probability of buffer overflow in 1 second
By the Chernoff bound

$$p = P(X > (1 + \delta)\mu) \leq \exp \left( -\frac{\delta^2 \mu}{3} \right)$$

Overflow probability in $n$ seconds

$$= 1-(1-p)^n \leq np \leq n \exp(- \frac{\delta^2 \mu}{3}),$$

which is $\leq \epsilon$ provided $\delta \geq \sqrt{(3/\mu)\ln(n/\epsilon)}$.

For $\epsilon = 10^{-6}$ per hour: $\delta \approx .257$, buffers = 1257
For $\epsilon = 10^{-6}$ per year: $\delta \approx .305$, buffers = 1305
For $\epsilon = 10^{-6}$ per 15BY: $\delta \approx .404$, buffers = 1404
Tail bounds – bound probabilities of extreme events
Important, e.g., for “risk management” applications
Three (of many):

- **Markov**: \( P(X \geq k\mu) \leq 1/k \) (weak, but general; only need \( X \geq 0 \) and \( \mu \))
- **Chebyshev**: \( P(|X-\mu| \geq k\sigma) \leq 1/k^2 \) (often stronger, but also need \( \sigma \))
- **Chernoff**: various forms, depending on underlying distribution; usually \( 1/\text{exponential} \), vs \( 1/\text{polynomial} \) above

Generally, more assumptions/knowledge \( \Rightarrow \) better bounds

“Better” than exact distribution?

Maybe, e.g. if latter is unknown or mathematically messy

“Better” than, e.g., “Poisson approx to Binomial”?

Maybe, e.g. if you need rigorously “\( \leq \)” rather than just “\( \approx \)”