tail bounds
Often, we want to bound the probability that a random variable $X$ is far from its expectation.

A random variable $X$ has mean $\mu$

Can we bound

\[ Pr(X \geq 100\mu) \]
\[ Pr(X \geq 1,000\mu) \]
\[ Pr(X \geq 1,000,000\mu) \]

Not without additional information…
We know that randomized quicksort runs in $O(n \log n)$ \textit{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?

We 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…*”
Markov’s inequality

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:

$$P(X \geq \alpha E[X]) \leq 1/\alpha$$
**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:

$$P(X \geq 4 \cdot E[X]) \leq E[X]/(4 E[X]) = 1/4$$
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:**

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

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

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

$$= \alpha P(X \geq \alpha)$$
Theorem: If $X$ is a non-negative random variable, then for every $\alpha > 0$, we have

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

Proof:

\[
\mathbb{E}[X] = \sum x \mathbb{P}(X = x) \\
\geq \sum \alpha \mathbb{P}(X = x) \\
= \alpha \sum \mathbb{P}(X = x) = \alpha \mathbb{P}(X \geq \alpha) \]

(x $\geq$ 0; $\alpha \leq x$)
For a random variable $X$, the *tails* of $X$ are the parts of the PMF that are “far” from its mean.
Binomial distribution, n=100, p=.5
heavy-tailed distribution
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 $E[Y] = \mu$, then, for any $\alpha > 0$,

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

**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:

$$P(|Y - \mu| \geq \alpha) = P(X \geq \alpha^2) \leq \frac{E[X]}{\alpha^2} = \frac{\text{Var}[Y]}{\alpha^2}$$
Chebyshev’s inequality

**Theorem:** If $Y$ is an arbitrary random variable with $\mu = \mathbb{E}[Y]$, 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:

$$P(X \geq \alpha^2) \leq \frac{\text{Var}[Y]}{\alpha^2}$$
Chebyshev’s inequality

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

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

**Corr:** If

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

Then:

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

\[ P(|Y - \mu| \geq t\sigma) \leq \frac{1}{t^2} \]
Chebyshev’s inequality

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

\[ Y = \text{comparisons in quicksort for } n=1024 \]

\[ E[Y] = 1.4 \, n \, \log_2 n \approx 14000 \]

\[ \text{Var}[Y] = \left( \frac{21-2\pi^2}{3} \right) n^2 \approx 441000 \]

(i.e. \( \text{SD}[Y] \approx 664 \))

\[ P(Y \geq 4 \mu) = P(Y - \mu \geq 3 \mu) \leq \frac{\text{Var}(Y)}{(9 \mu^2)} < .000242 \]

1000 times smaller than Markov but still overestimated?: \( \sigma/\mu \approx 0.05 \), so \( 4\mu \approx \mu + 60\sigma \)
Chebyshev’s inequality

$X \text{ Binomial } (n, 1/2)$

$$Pr(X \geq \frac{3}{4}n)$$

Markov: $\frac{2}{3}$

Chebyshev: $\frac{2}{n}$

If $n=1000$, Probability > 750 H’s at most 0.002

Truth:

0.000000000000000000000000000000000000000000000000000000000067
Suppose $X \sim \text{Bin}(n,p)$

$\mu = E[X] = pn$

Chernoff bound:

For any $\delta$ with $0 < \delta < 1$,

$$P(X > (1 + \delta)\mu) \leq e^{-\frac{\delta^2 \mu}{2}}$$

$$P(X < (1 - \delta)\mu) \leq e^{-\frac{\delta^2 \mu}{3}}$$
Suppose $X \sim \text{Bin}(n,p)$

$$\mu = E[X] = np$$

**Chernoff bound:**

For any $\delta$ we have

$$P(X < \mu - \delta \mu) \leq e^{-\frac{\delta^2 \mu}{2}}$$

$$P(X < \mu - \frac{\delta \mu}{3}) \leq e^{-\frac{\delta^2 \mu}{3}}$$
Model: 100,000 computers each independently send a packet with probability $q = 0.01$ each second. The router processes its buffer every second. How many packet buffers so that router drops a packet:

- Never?
  100,000

- With probability at most $10^{-6}$, every hour?
  1210

- With probability at most $10^{-6}$, every year?
  1250

- With probability at most $10^{-6}$, since Big Bang?
  1331
Let $p = \text{probability of buffer overflow in 1 second}$

By the Chernoff bound

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

Overflow probability in $T$ seconds

$$= 1 - (1-p)^T \leq Tp \leq T \exp(-\frac{\delta^2 \mu}{2}),$$

which is $\leq \varepsilon$ provided $\delta \geq \sqrt{\frac{2}{\mu} \ln(T/\varepsilon)}$.

For $\varepsilon = 10^{-6}$ per hour: $\delta \approx .210$, buffers = 1210

For $\varepsilon = 10^{-6}$ per year: $\delta \approx .250$, buffers = 1250

For $\varepsilon = 10^{-6}$ per 15BY: $\delta \approx .331$, buffers = 1331
Tail bounds – bound probabilities of extreme events

Three (of many):

Markov: \( \Pr(X \geq k \mu) \leq 1/k \) (weak, but general; only need \( X \geq 0 \) and \( \mu \))

Chebyshev: \( \Pr(|X - \mu| \geq k \sigma) \leq 1/k^2 \) (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 “\( \approx \)”