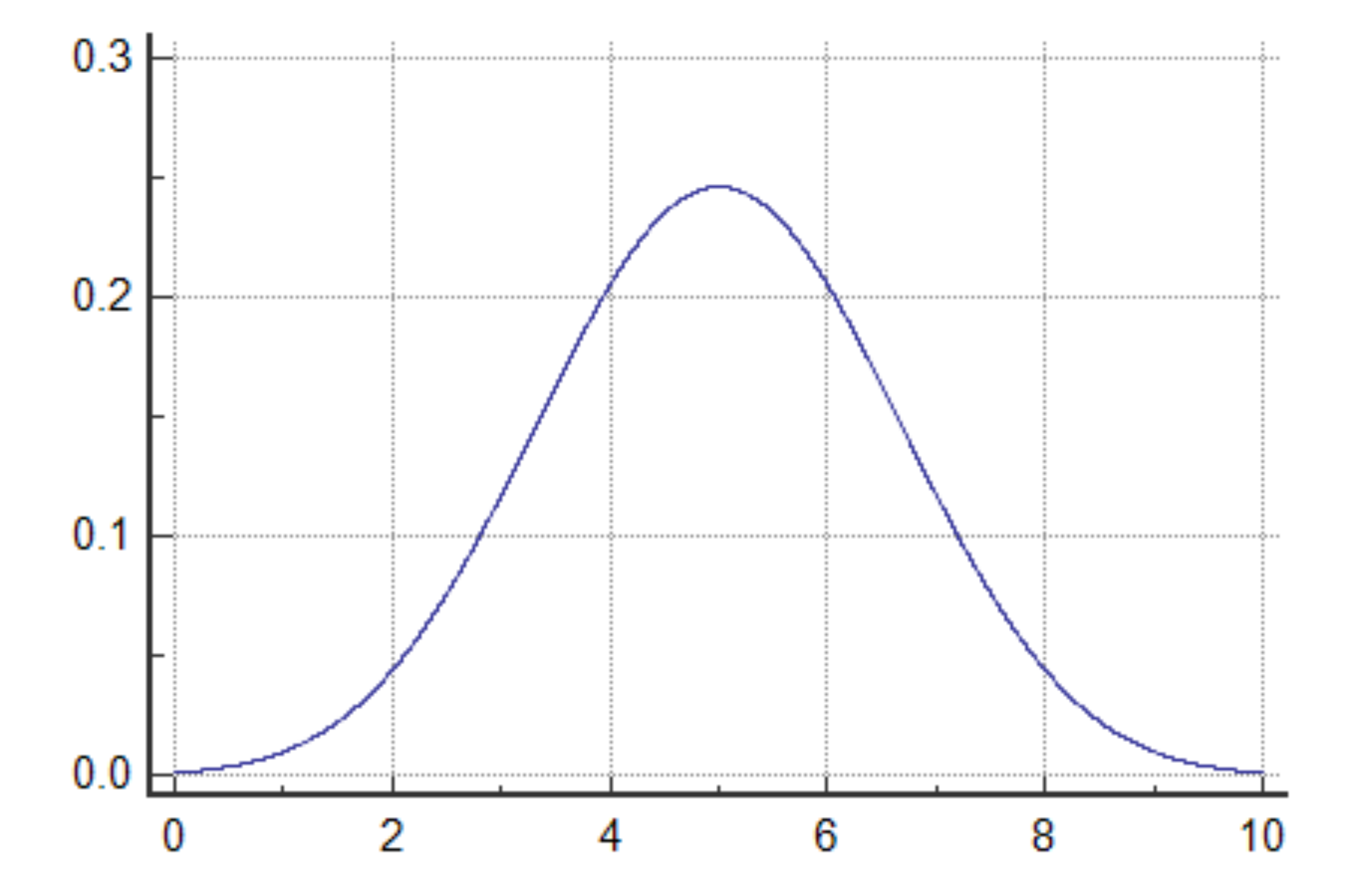
COMP 480/580 Probabilistic Algorithms and Data Structures

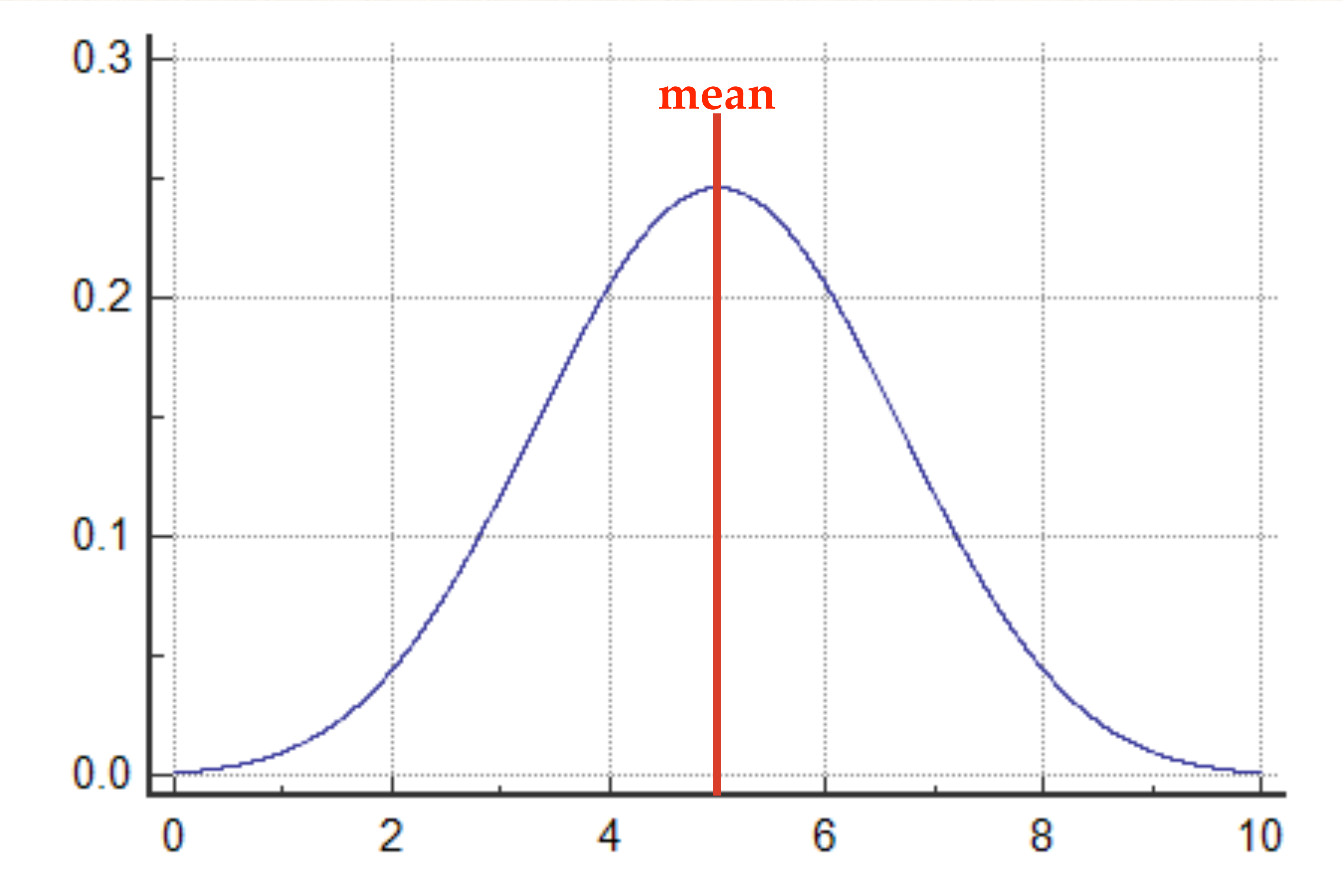
Tail Bounds

Luay Nakhleh
Computer Science
Rice University

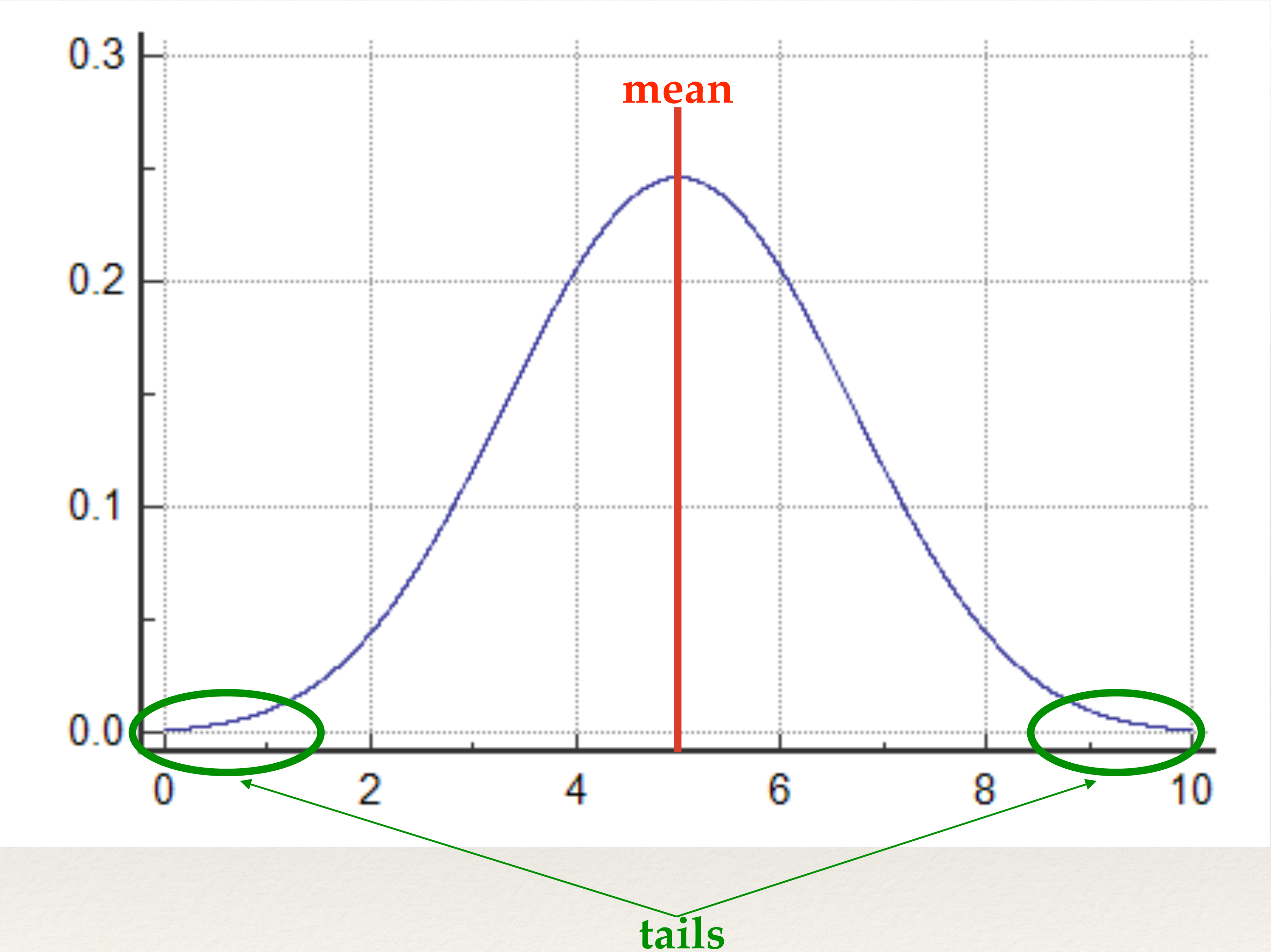
What Is This About?

- How large can a random variable get?
- In other words, how far can a value that the random variable takes be from its mean?

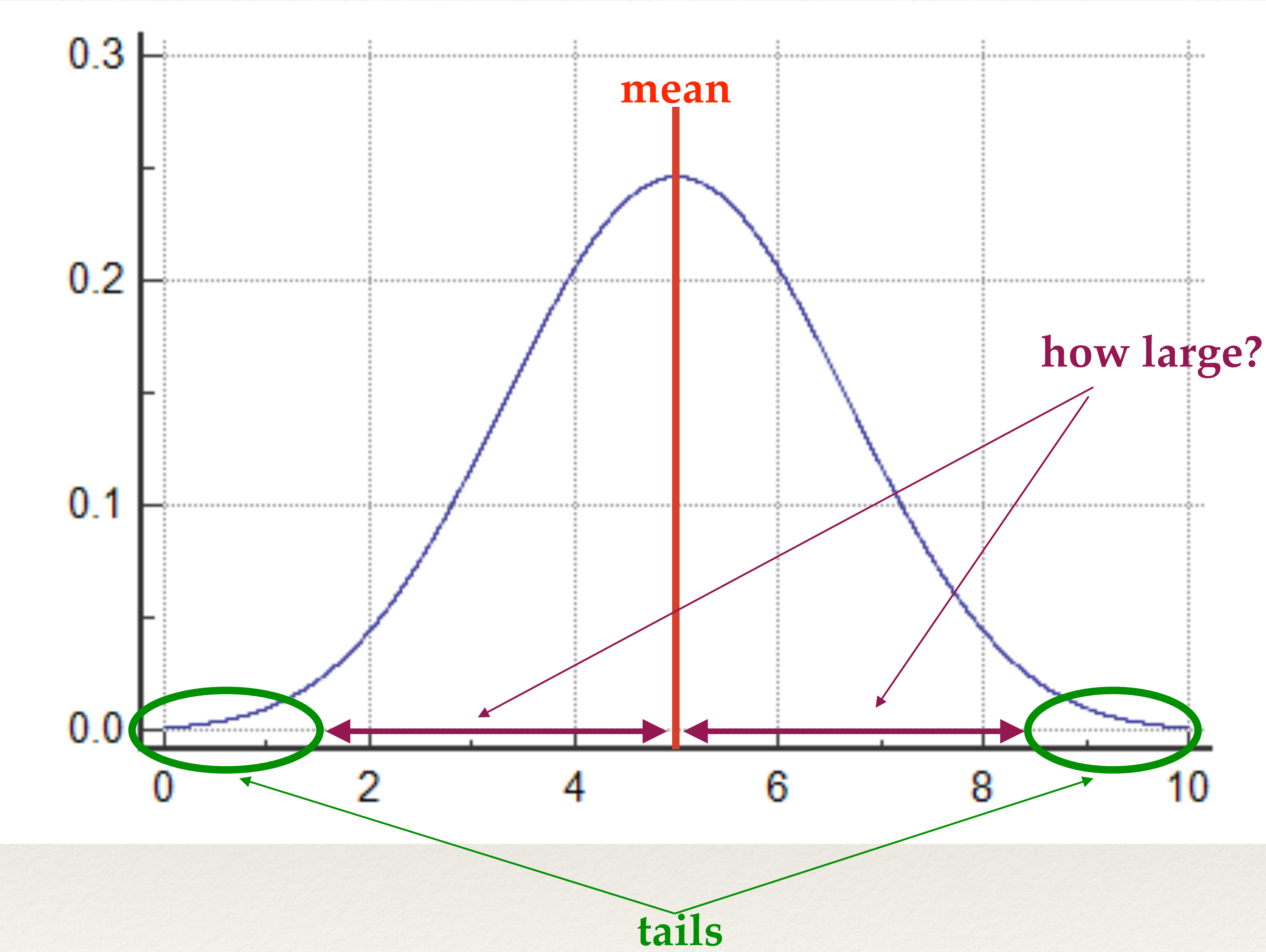

0.3
0.2
0.1
0.0
0
2
4
6
8
10



0.3
0.2
0.1
0.0
mean
0
2
4
6
8
10



mean
0.3
0.2
0.1
0.0
0
2
4
6
8
10
tails


0.3
0.2
0.1
0.0
0
2
4
6
8
10
mean
how large?
tails

Why Do We Care?

- Example:
  - X is the number of steps an algorithm takes.
  - $\mathbb{E}(X)$ is the average-case running-time of the algorithm.
  - Can the algorithm, on average, take $2n$ steps, but on some inputs take, say, $500n^2$ steps?

Recall

- A random variable is a function from the sample space of an experiment/process to the set of real numbers.
- A coin is tossed twice. Let $X(t)$ be the random variable that equals the number of heads that appear when t is the outcome. Then $X(t)$ takes on the following values:
  - $X(HH)=2$
  - $X(HT)=X(TH)=1$
  - $X(TT)=0$

Expected Value

- The expected value (also called the expectation or mean) of a (discrete) random variable X on the sample space S is

$$\mathbb{E}(X) = \sum_{s \in S} P(s) \cdot X(s)$$

(the same as $\mathbb{E}(X) = \sum_{x} x \cdot P(X = x)$)

Linearity of Expectations

- If X_i, $i=1,2,\ldots,n$, are random variables on S, and if a and b are real numbers, then

$$\mathbb{E}(X_1 + X_2 + \cdots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n)$$

$$\mathbb{E}(aX_i + b) = a\mathbb{E}(X_i) + b$$

Variance

- Let X be a random variable on a sample space S. The variance of X, denoted by $V(X)$, is

$$V(X) = \mathbb{E}((X - E(X))^2)$$

(and equals $V(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$)

Bienayme's Formula

- If X_i, i=1,2,…,n, are pairwise independent random variables on S, then

$$V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i)$$

Markov's Inequality

- Let X be a random variable that takes only nonnegative values. Then, for every real number $a>0$ we have

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

Markov's Inequality

- Let X be a random variable that takes only nonnegative values. Then, for every real number $a>0$ we have

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

How large a value can X take?

Markov's Inequality: Proof

?

Markov's Inequality: An Example

- Assume the expected time it takes Algorithm A to traverse a graph with n nodes is $2n$. What is the probability that the algorithm takes more than *10* times that?

- For distributions encountered in practice, Markov’s inequality gives a very loose bound.
  - Why?

Chebyshev's Inequality

- Let X be a random variable. For every real number r>0,

$$P(|X - \mathbb{E}(X)| \geq a) \leq \frac{V(X)}{a^2}$$

Chebyshev's Inequality

- Let *X* be a random variable. For every real number *r>0,*

$$P(|X - \mathbb{E}(X)| \geq a) \leq \frac{V(X)}{a^2}$$

How likely is it that RV *X* takes a value
that's at least distance *a* from its expected value?

Chebyshev's Inequality: Proof

?

Markov vs Chebyshev

$$P(X \geq k\mu) \leq \frac{1}{k}$$

vs

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

Chebyshev's Inequality: An Example

- Assume we have a distribution whose mean is 80 and standard deviation is 10. What is a lower bound on the percentage of values that fall between 60 and 100 (exclusively) in this distribution?

$$P(|X - \mathbb{E}(X)| \geq a) \leq \frac{V(X)}{a^2}$$

Chebyshev's Inequality: An Example

- Assume we have a distribution whose mean is 80 and standard deviation is 10. What is a lower bound on the percentage of values that fall between 60 and 100 (exclusively) in this distribution?

$$P(|X - \mathbb{E}(X)| \geq a) \leq \frac{V(X)}{a^2}$$

$$\mathbb{E}(X) = 80 \qquad V = 100 \qquad r = 20$$

Chebyshev's Inequality: An Example

- Assume we have a distribution whose mean is 80 and standard deviation is 10. What is a lower bound on the percentage of values that fall between 60 and 100 (exclusively) in this distribution?

$$P(|X - \mathbb{E}(X)| \geq a) \leq \frac{V(X)}{a^2}$$

$$\mathbb{E}(X) = 80 \qquad V = 100 \qquad r = 20$$

$$p(|X(s) - 80| \geq 20) \leq \frac{1}{4}$$

Chebyshev's Inequality: An Example

- Assume we have a distribution whose mean is 80 and standard deviation is 10. What is a lower bound on the percentage of values that fall between 60 and 100 (exclusively) in this distribution?

$$P(|X - \mathbb{E}(X)| \geq a) \leq \frac{V(X)}{a^2}$$

$$\mathbb{E}(X) = 80 \qquad V = 100 \qquad r = 20$$

$$p(|X(s) - 80| \geq 20) \leq \frac{1}{4}$$

$$\Rightarrow \text{ lower bound is } 75\%$$

Illustration: Estimating π Using the Monte Carlo Method

- Here's a simple algorithm for estimating π:
  - Throw darts at a square whose area is 1, inside which there's a circle whose radius is 1/2.
  - The probability that it lands inside the circle equals the ratio of the circle area to the square area ($\pi/4$). Therefore, calculate the proportion of times that the dart landed inside the circle and multiply it by 4.

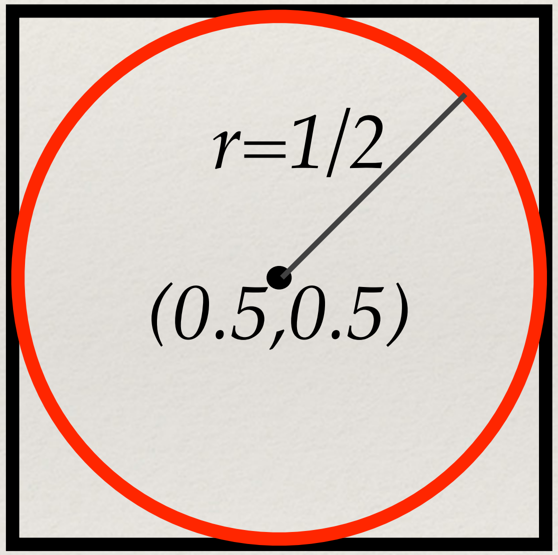


Illustration: Estimating π Using the Monte Carlo Method

Algorithm 1: MonteCarlo_πEstimation.

Input: $n \in \mathbb{N}$.
Output: Estimate $\hat{\pi}$ of π.

for $i = 1$ **to** n **do**
  $a \leftarrow random(0,1)$; `// random number in` $[0,1]$
  $b \leftarrow random(0,1)$; `// random number in` $[0,1]$
  $X_i \leftarrow 0$;
  if $\sqrt{(a-0.5)^2 + (b-0.5)^2} \leq 0.5$ **then**
    $X_i \leftarrow 1$; `// the dart landed inside/on the circle`

$\hat{\pi} \leftarrow 4 \cdot (\sum_{i=1}^{n} X_i)/n$;
return $\hat{\pi}$;

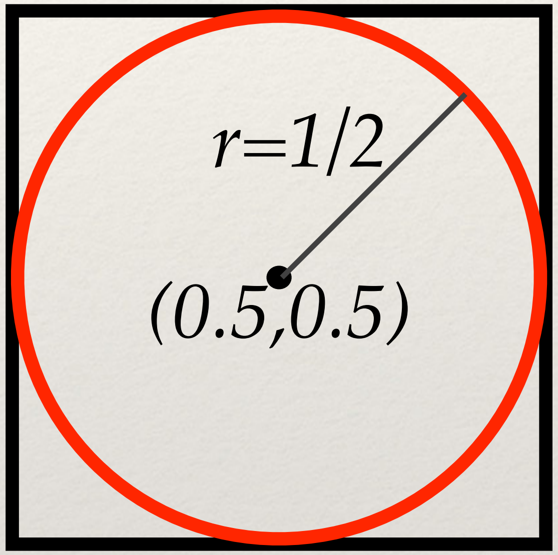


Illustration: Estimating π Using the Monte Carlo Method

- Let X_i be the random variable that denotes whether the i-th dart landed inside the circle (1 if it did, and 0 otherwise).
- Then, $\hat{\pi}(n) = 4\frac{\sum_{i=1}^{n} X_i}{n}$

Illustration: Estimating π Using the Monte Carlo Method

- Let X_i be the random variable that denotes whether the i-th dart landed inside the circle (1 if it did, and 0 otherwise).
- Then, $\hat{\pi}(n) = 4\frac{\sum_{i=1}^{n} X_i}{n}$

$$\mathbb{E}(X_i) = \frac{\pi}{4}1 + (1 - \frac{\pi}{4})0 = \frac{\pi}{4}$$

$$V(X_i) = \frac{\pi}{4}(1 - \frac{\pi}{4})$$

Illustration: Estimating π Using the Monte Carlo Method

- Let X_i be the random variable that denotes whether the i-th dart landed inside the circle (1 if it did, and 0 otherwise).
- Then, $\hat{\pi}(n) = 4\frac{\sum_{i=1}^{n} X_i}{n}$

$$\mathbb{E}(X_i) = \frac{\pi}{4}1 + (1 - \frac{\pi}{4})0 = \frac{\pi}{4}$$

$$V(X_i) = \frac{\pi}{4}(1 - \frac{\pi}{4})$$

$$\mathbb{E}(\hat{\pi}) = \mathbb{E}\left(\frac{4}{n}\sum_{i=1}^{n} X_i\right) = \frac{4}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \pi$$

Illustration: Estimating π Using the Monte Carlo Method

- Let X_i be the random variable that denotes whether the i-th dart landed inside the circle (1 if it did, and 0 otherwise).
- Then, $\hat{\pi}(n) = 4\frac{\sum_{i=1}^{n} X_i}{n}$

$$\mathbb{E}(X_i) = \frac{\pi}{4}1 + (1 - \frac{\pi}{4})0 = \frac{\pi}{4}$$

$$V(X_i) = \frac{\pi}{4}(1 - \frac{\pi}{4})$$

$$\mathbb{E}(\hat{\pi}) = \mathbb{E}\left(\frac{4}{n}\sum_{i=1}^{n} X_i\right) = \frac{4}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \pi$$

$$V(\hat{\pi}) = V(\frac{4}{n}\sum_{i=1}^{n} X_i) = \frac{16}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{\pi(4-\pi)}{n}$$

Illustration: Estimating π Using the Monte Carlo Method

- The question of interest is: How big should n be for us to get a good estimate?

Illustration: Estimating π Using the Monte Carlo Method

- In a probabilistic setting, the question can be asked as:
  - What should the value of n be so that the estimation error of π is within δ with probability at least ε?
  - (of course, we want δ to be very small and ε to be as close to 1 as possible. For example, δ=0.001 and ε=0.95)

Illustration: Estimating π Using the Monte Carlo Method

- In other words, we are interested in the value of n that yields

$$p(|\hat{\pi}(n) - \pi| < \delta) > \varepsilon$$

$$(\text{equivalently, } p(|\hat{\pi}(n) - \pi| \geq \delta) \leq 1 - \varepsilon)$$

Illustration: Estimating π Using the Monte Carlo Method

- For δ=0.001 and ε=0.95, we seek n such that

$$p(|\hat{\pi}(n) - \pi| \geq 0.001) \leq 0.05$$

Illustration: Estimating π Using the Monte Carlo Method

- For δ=0.001 and ε=0.95, we seek *n* such that

$$p(|\hat{\pi}(n) - \pi| \geq 0.001) \leq 0.05$$

Chebyshev's inequality: $\hat{\pi}(n)$ $\mathbb{E}(\hat{\pi})$ *a* *V*/*a*²

Illustration: Estimating π Using the Monte Carlo Method

- For δ=0.001 and ε=0.95, we seek n such that

$$p(|\hat{\pi}(n) - \pi| \geq 0.001) \leq 0.05$$

Chebyshev's inequality: $\hat{\pi}(n)$ $\mathbb{E}(\hat{\pi})$ *a* *V/a*²

$$V(\hat{\pi}(n)) = \frac{\pi(4-\pi)}{n}$$

Illustration: Estimating π Using the Monte Carlo Method

- For δ=0.001 and ε=0.95, we seek n such that

$$p(|\hat{\pi}(n) - \pi| \geq 0.001) \leq 0.05$$

Chebyshev's inequality: $\hat{\pi}(n)$ $\mathbb{E}(\hat{\pi})$ *a* *V/a²*

$$V(\hat{\pi}(n)) = \frac{\pi(4-\pi)}{n}$$

So, we would like n such that $\frac{\pi(4-\pi)}{n(0.001)^2} \leq 0.05$

Illustration: Estimating π Using the Monte Carlo Method

$$\frac{\pi(4-\pi)}{n(0.001)^2} \leq 0.05$$

Illustration: Estimating π Using the Monte Carlo Method

$$\frac{\pi(4-\pi)}{n(0.001)^2} \leq 0.05$$

$$\pi(4-\pi) \leq 4$$

Illustration: Estimating π Using the Monte Carlo Method

$$\frac{\pi(4-\pi)}{n(0.001)^2} \leq 0.05$$

$$\pi(4-\pi) \leq 4$$

$$\Rightarrow \frac{\pi(4-\pi)}{n(0.001)^2} \leq \frac{4}{n(0.001)^2} \leq 0.05$$

Illustration: Estimating π Using the Monte Carlo Method

$$\frac{\pi(4-\pi)}{n(0.001)^2} \leq 0.05$$

$$\pi(4-\pi) \leq 4$$

$$\Rightarrow \frac{\pi(4-\pi)}{n(0.001)^2} \leq \frac{4}{n(0.001)^2} \leq 0.05$$

$$\Rightarrow n \geq 80,000,000$$

A Corollary of Chebyshev's Inequality

- Let $X_1, X_2, \ldots, X_n$ be independent random variables with

$$\mathbb{E}(X_i) = \mu_i \qquad \text{and} \qquad V(X_i) = \sigma_i^2$$

Then, for any $a > 0$:

$$P\left(\left| \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i \right| \geq a \right) \leq \frac{\sum_{i=1}^{n} \sigma_i^2}{a^2}$$

The Weak Law of Large Numbers

- Let $X_1, X_2, \ldots, X_n$ be independently and identically distributed (i.i.d.) random variables, where the (unknown) expected value μ is the same for all variables (that is, $\mathbb{E}(X_i) = \mu$) and their variance is finite. Then, for any ε>0, we have

$$P\left(\left|\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu\right| \geq \varepsilon\right) \xrightarrow{n\to\infty} 0$$

Chernoff Bounds

- The question is: Can we do better (give tighter bounds) than Markov's and Chebyshev's inequalities if we know something about the distribution of the random variables?
- The answer is YES, and there are many forms of Chernoff bounds depending on the assumptions.

Chernoff Bound

- Let $X=X_1+X_2+\ldots+X_n$, where all the X_i's are independent and $X_i \sim Bernoulli(p_i)$.
- Let $\mu = \mathbb{E}(X) = \sum_{i=1}^{n} p_i$.
- Then, for $\delta>0$,

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

Chernoff Bound

- The bound can also be written as

$$P(X \geq (1+\delta)\mu) \leq e^{-\frac{\delta^2 \mu}{2+\delta}} \qquad \text{for } \delta>0$$

$$P(X \leq (1-\delta)\mu) \leq e^{-\frac{\delta^2 \mu}{2}} \qquad \text{for } 1>\delta>0$$

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

Lemma 1 *Given random variable $Y \sim Bernoulli(p)$, we have for all $s \in \mathbb{R}$*

$$\mathbb{E}(e^{sY}) \leq e^{p(e^s - 1)}.$$

Lemma 2 *Let $X_1, \ldots, X_n$ be independent random variables, and $X = \sum_{i=1}^n X_i$. Then, for $s \in \mathbb{R}$*

$$\mathbb{E}(e^{sX}) = \prod_{i=1}^n \mathbb{E}(e^{sX_i}).$$

Lemma 3 *Let $X_1, \ldots, X_n$ be independent random variables (Bernoulli distributed), and $X = \sum_{i=1}^n X_i$ and $\mathbb{E}(X) = \sum_{i=1}^n p_i = \mu$. Then, for $s \in \mathbb{R}$*

$$\mathbb{E}(e^{sX}) \leq e^{(e^s - 1)\mu}.$$

To establish the result, use Markov's inequality on the rhs of $P(X \geq a) = P(e^{sX} \geq e^{sa})$,
the inequality $\ln(1+x) \geq 2x/(2+x)$ for $x>0$ and
set $a=(1+\delta)\mu$ and $s=\ln(1+\delta)$ (why?)

Tossing a Fair Coin

- A fair coin is tossed 200 times. How likely is it to observe at least 150 heads?
  - Markov: ≤ 0.6666
  - Chebyshev: ≤ 0.02
  - Chernoff: ≤ 0.017

Another Chernoff Bound

- Let $X = X_1 + X_2 + \ldots + X_n$, where all the X_i's are independent and $a \leq X_i \leq b$ for all i.
- Let $\mu = \mathbb{E}(X)$.
- Then, for $\delta > 0$,

$$P(X \geq (1+\delta)\mu) \leq e^{-\frac{2\delta^2\mu^2}{n(b-a)^2}}$$

$$P(X \leq (1-\delta)\mu) \leq e^{-\frac{\delta^2\mu^2}{n(b-a)^2}}$$

Questions?