Estimating p for X~Bernoulli(p)

Lecture 23, CS70 Fall 2026

Estimating p for X~Bernoulli(p)

Markov's Inequality

Chebyshev's Inequality

The Chebyshev Bound for Number of Heads

The Chebyshev Bound for Estimating the Fairness of a Coin

Weak Law of Large Numbers

Bounds Beyond CS70

CC BY NC SA

Motivating Problem of the Day

Suppose we wanted to know in advance how many people would vote for James Talarico in the 2026 Texas Senate election.

One approach: Ask random people.

- As long as we ask enough of the right people, we'll get an accurate result.
- But how many people do we need to ask?

CC BY NC SA

Motivating Problem of the Day: Coin Version

We can think of this polling question in terms of a coin flip.

- Every person we ask is effectively one flip of the Texas coin. Let's say "heads" is a James Talarico vote. If 55.1% of the population will vote for James Talarico, then $p = 0.551$.
- Note: The coin flipping is not modeling the randomness inherent in an individual's vote! Instead, it's a sample drawn from the entire population, 55.1% of whom will definitely vote for James Talarico.

Our goal is to flip the coin a bunch of times to estimate the probability of heads p.

CC BY NC SA

Experiment

Suppose we flip the coin 10,000 times and we get heads 5500 times and tails 4500 times. Let p be the true chance of heads. Let $\hat{p}$ be our estimate.

Based on our experiment, what should we use as $\hat{p}$?

CC BY NC SA

Experiment

Suppose we flip the coin 10,000 times and we get heads 5500 times and tails 4500 times. Let p be the true chance of heads. Let $\hat{p}$ be our estimate.

Based on our experiment, what should we use as $\hat{p}$? $\hat{p} = 0.55$

CC BY NC SA

Experiment

Suppose we flip the coin 10,000 times and we get heads 5500 times and tails 4500 times. Let p be the true chance of heads. Let $\hat{p}$ be our estimate.

Based on our experiment, what should we use as $\hat{p}$? $\hat{p} = 0.55$

How confident should we feel about this estimate?

- Example: How sure are we that we're off by less than 0.01? That is, what can we say about $P(|p - 0.55| < 0.01)$?
  - In other words, what is the chance that the true p lies inside $(0.54, 0.56)$?

CC BY NC SA

Experiment

Suppose we flip the coin 10,000 times and we get heads 5500 times and tails 4500 times. Let p be the true chance of heads. Let $\hat{p}$ be our estimate.

Based on our experiment, what should we use as $\hat{p}$? $\hat{p} = 0.55$

How confident should we feel about this estimate?

- Example: How sure are we that we're off by less than 0.01? That is, what can we say about $P(|p - 0.55| < 0.01)$?
  - In other words, what is the chance that the true p lies inside $(0.54, 0.56)$?
  - Today: The "Chebyshev bound" tells us $P(|p - 0.55| < 0.01) \geq 75\%$
  - In other words, we have at LEAST a 75% chance of being "right", where "right" is defined as being within 0.01.

CC BY NC SA

Other Framing

We could also ask the reverse question:

- Suppose we want to create an estimate $\hat{p}$ for the true heads probability p.
- How many times do we need to flip the coin so that we have a 75% chance of being within 0.01 of the true value of p?
  - Want n such that $P(|\hat{p} - p| < 0.01) \geq 75\%$

Or more generally, we want n such that $P(|\hat{p} - p| < \epsilon) \geq C$

- Here, C is our desired confidence level.

CC BY NC SA

Today's Answer (Spoiler)

To achieve our desired goal: $P(|\hat{p} - p| < \epsilon) \geq C$

We can show that $n \geq \frac{1}{4\epsilon^2(1-C)}$ coin flips is sufficient.

- We'll use something called Chebyshev's inequality to derive this later.

Suppose we define being "right" as being within 0.05 of the true probability. How many times do we need to flip the coin to have a 99% chance of being right?

- joshh.ug/A

CC BY NC SA

Today's Answer (Spoiler)

To achieve our desired goal: $P(|\hat{p} - p| < \epsilon) \geq C$

We can show that $n \geq \frac{1}{4\epsilon^2(1-C)}$ coin flips is sufficient.

- We'll use something called Chebyshev's inequality to derive this later.

Example: If $C = 99\%$ and $\epsilon = 0.05$, then target n is $\frac{1}{4(0.05)^2(0.01)} = 10000$

- Here C is our desired confidence level.
- That is, if we flip the coin 10,000 times, then there is only a 1% chance that our estimate $\hat{p}$ is off by 0.05 or more.
    - Example: If we get $\hat{p} = 0.55$, the bound says the chance is less than 1% that the true p lies outside the range $(0.50, 0.60)$.

CC BY NC SA

Today's Answer (Alternate Form)

To achieve our desired goal: $P(|\hat{p} - p| < \epsilon) \geq 1 - \delta$

We can show that $n \geq \frac{1}{4\epsilon^2\delta}$ coin flips is sufficient.

- We'll use something called Chebyshev's inequality to derive this later.

Example: If $\delta = 1\%$ and $\epsilon = 0.05$, then target n is $\frac{1}{4(0.05)^2(0.01)} = 10000$

- Here δ is our acceptable failure level.
- That is, if we flip the coin 10,000 times, then there is only a 1% chance that our estimate $\hat{p}$ is off by 0.05 or more.
    - Example: If we get $\hat{p} = 0.55$, the bound says the chance is less than 1% that the true p lies outside the range $(0.50, 0.60)$.

cc BY NC SA

Today's Answer (Alternate Form #2)

To achieve our desired goal: $P(|\hat{p} - p| \geq \epsilon) \leq \delta$

We can show that $n \geq \frac{1}{4\epsilon^2\delta}$ coin flips is sufficient.

- We'll use something called Chebyshev's inequality to derive this later.

Example: If $\delta = 1\%$ and $\epsilon = 0.05$, then target n is $\frac{1}{4(0.05)^2(0.01)} = 10000$

- Here δ is our acceptable failure level.
- That is, if we flip the coin 10,000 times, then there is only a 1% chance that our estimate $\hat{p}$ is off by 0.05 or more.
    - Example: If we get $\hat{p} = 0.55$, the bound says the chance is less than 1% that the true p lies outside the range $(0.50, 0.60)$.

cc BY NC SA

Today's Answer (Spoiler)

To achieve our desired goal: $P(|\hat{p} - p| \geq \epsilon) \leq \delta$

We can show that $n \geq \frac{1}{4\epsilon^2\delta}$ coin flips is sufficient.

- We'll use something called Chebyshev's inequality to derive this later.

Example: If $\delta = 1\%$ and $\epsilon = 0.05$, then target n is $\frac{1}{4(0.05)^2(0.01)} = 10000$

- That is, if we flip the coin 10,000 times, then there is only a 1% chance that our estimate $\hat{p}$ is off by 0.05 or more.
- I observe that it is also true that $\frac{1}{4(0.01)^2(0.25)} = 10000$. As a result, what else can I say about $\hat{p}$?

CC BY NC SA

Today's Answer (Spoiler)

To achieve our desired goal: $P(|\hat{p} - p| \geq \epsilon) \leq \delta$

We can show that $n \geq \frac{1}{4\epsilon^2\delta}$ coin flips is sufficient.

- We'll use something called Chebyshev's inequality to derive this later.

Example: If $\delta = 1\%$ and $\epsilon = 0.05$, then target n is $\frac{1}{4(0.05)^2(0.01)} = 10000$

- I observe that it is also true that $\frac{1}{4(0.01)^2(0.25)} = 10000$.
  - There is a less than 25% chance we're off by 0.01 or more.
  - Example: If $\hat{p} = 0.55$, there is a $\geq 75\%$ chance the true p is in the range $(0.54,\ 0.56)$.

CC BY NC SA

Understanding this Bound

To achieve our desired goal: $P(|\hat{p} - p| \geq \epsilon) \leq \delta$

We can show (later!) that $n \geq \frac{1}{4\epsilon^2\delta}$ coin flips is sufficient.

Let's get some more intuition from this using a simulator.

Flips, n **10,000**

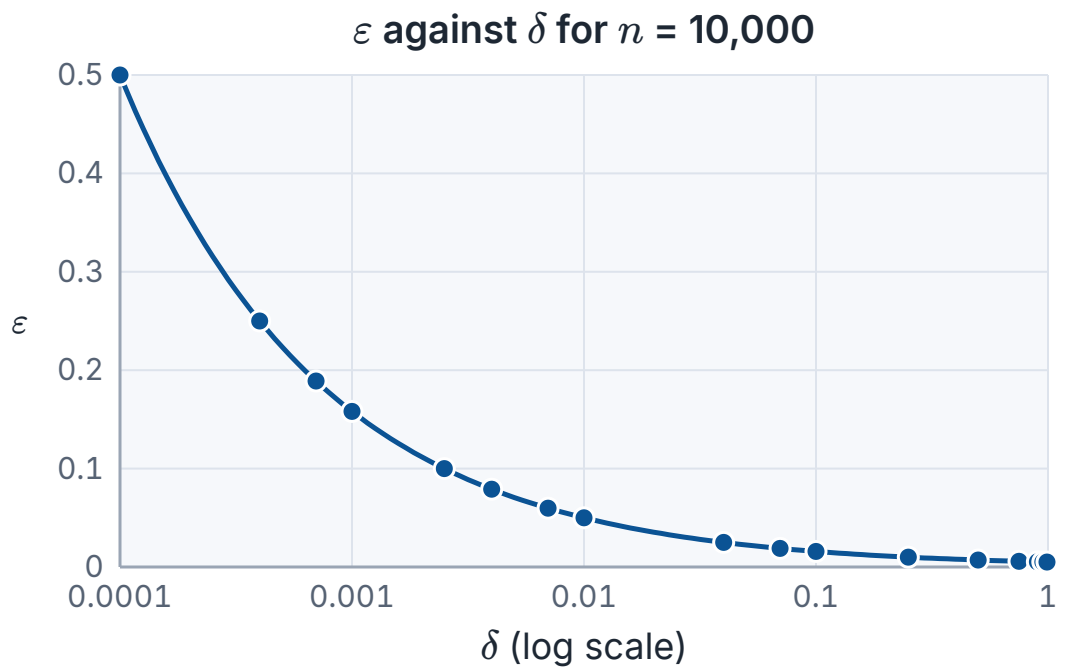
Heads probability, p **0.50**

Simulate

Estimate $\hat{p}$

0.5040

5,040 heads in 10,000 flips



With chance at least 1 – δ, the true p is within ε of $\hat{p}$:

δ	ε	$1 - \delta$	$\hat{p} \pm \varepsilon$
0.0001	0.5	99.99%	[0.004, 1.000]
0.0004	0.25	99.96%	[0.254, 0.754]
0.0007	0.189	99.93%	[0.315, 0.693]
0.001	0.158	99.90%	[0.346, 0.662]
0.0025	0.1	99.75%	[0.404, 0.604]
0.004	0.0791	99.60%	[0.425, 0.583]
0.007	0.0598	99.30%	[0.444, 0.564]
0.01	0.05	99.00%	[0.454, 0.554]
0.04	0.025	96.00%	[0.479, 0.529]
0.07	0.0189	93.00%	[0.485, 0.523]
0.1	0.0158	90.00%	[0.488, 0.520]
0.25	0.01	75.00%	[0.494, 0.514]
0.5	0.00707	50.00%	[0.497, 0.511]
0.75	0.00577	25.00%	[0.498, 0.510]
0.9	0.00527	10.00%	[0.499, 0.509]
0.95	0.00513	5.00%	[0.499, 0.509]
0.99	0.00503	1.00%	[0.499, 0.509]

CC BY NC SA

In Terms of Election Polling

To achieve our desired goal: $P(|\hat{p} - p| \geq \epsilon) \leq \delta$

We can show (later!) that $n \geq \frac{1}{4\epsilon^2\delta}$ coin flips is sufficient.

(Surprising) Observation: It doesn't matter how big the population of the country is, the number of people we need to poll to get a desired degree of accuracy is constant.

- Ask 10,000 people, and our bound says we have a 99% chance of being within 0.05 of p. True if we're surveying Canada (40M) or China (1.4B).

Later, we'll come back to why the actual practice of polling is more difficult than this.

CC BY NC SA

Deriving this Expression and Concentration Inequalities

I've given this formula without any proof or rationale!

$$n \geq \frac{1}{4\epsilon^2\delta}$$

We'll derive this result using a powerful idea called a "concentration inequality".

- A concentration inequality is a mathematical bound on how likely a random variable is to stray from some quantity.
- Example: For a non-negative random variable X with expectation 45, what is the probability that $X > 90$?
  - Surprisingly, we can bound this probability without even knowing anything else about the RV (e.g. variance, distribution, etc).

CC BY NC SA

Markov's Inequality

Lecture 23, CS70 Fall 2026

Estimating p for X~Bernoulli(p)

Markov's Inequality

Chebyshev's Inequality

The Chebyshev Bound for Number of Heads

The Chebyshev Bound for Estimating the Fairness of a Coin

Weak Law of Large Numbers

Bounds Beyond CS70

CC BY NC SA

Thought Experiment 1

Imagine that 100 students take a midterm with average score $\mu = 45$.

- Can all 100 students score greater than $\mu = 45$ on the midterm?

CC BY NC SA

Thought Experiment 1

Imagine that 100 students take a midterm with average score $\mu = 45$.

- Can all 100 students score greater than $\mu = 45$ on the midterm?
- No! If they did, then averaging their scores would yield some number larger than $\frac{100 \times 45}{100}$. Average can't be larger than itself.

Or as a concrete example: Everyone gets just a tiny bit better than the average, say 45.0001, then average would be $\frac{45.0001 + 45.0001 + \cdots + 45.0001}{100} = 45.0001$, but that's more than 45.

CC BY NC SA

Thought Experiment 2

Imagine that 100 students take a midterm with average score $\mu = 45$.

- Can 50 students earn a score greater than or equal to $2\mu = 90$ on the midterm?
- joshh.ug/B



CC BY NC SA

Thought Experiment 2

Imagine that 100 students take a midterm with average score $\mu = 45$.

- Can 50 students earn a score greater than or equal to $2\mu = 90$ on the midterm?
- Sure! Assign all of the "small" outcomes to zero, i.e. 50 students get a 0, and all of the "large" outcomes to 90, i.e. the other 50 students get $2\mu = 90$. In that case, the average is:

$$\frac{50 \times 0 + 50 \times 90}{100} = \frac{4500}{100} = 45$$

Note, the 50 students can't all earn a score greater than 90! But they can all earn a score equal to 90.

cc BY NC SA

Thought Experiment 3

Imagine that we have n people whose average bank account balance is $\mu = \$5{,}000$. What is the maximum possible fraction who have $\$15{,}000$ in their account?

- joshh.ug/C

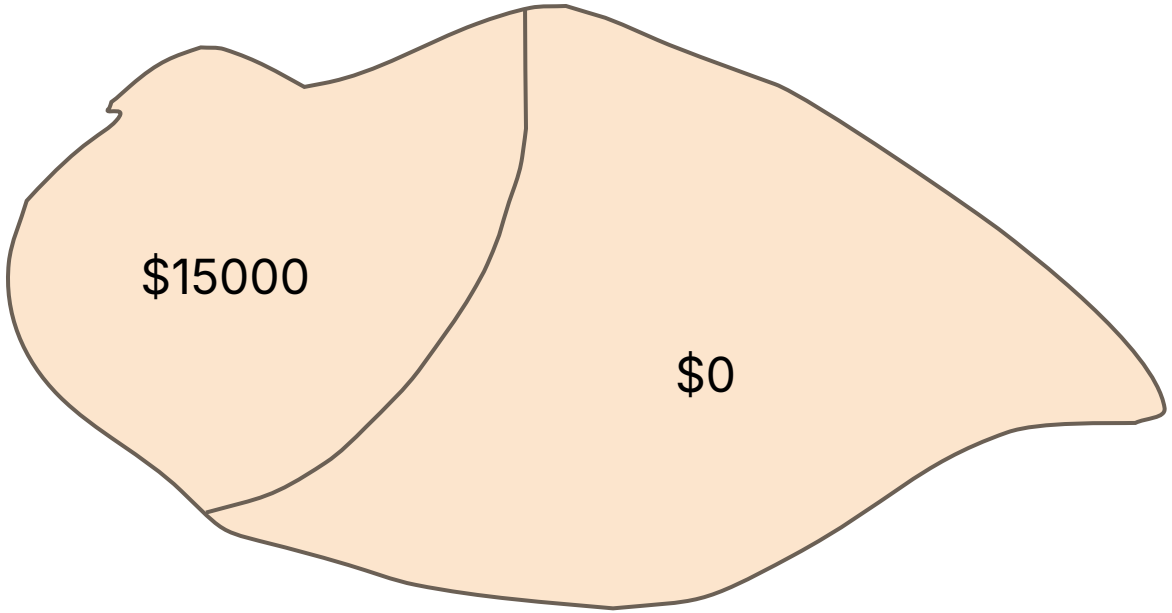


CC BY NC SA

Thought Experiment 3

Imagine that we have n people whose average bank account balance is $\mu = \$5{,}000$. What is the maximum possible fraction who have $\$15{,}000$ in their account?

- The maximum possible fraction is 1/3rd:
  - 2/3rds of people have zero dollars.
  - 1/3rd have $15,000.



CC BY NC SA

Thought Experiment 4

Consider a nonnegative random variable X with $\mathbb{E}[X] = 5000$. Let $p = P(X \geq 15000)$.

What is the largest possible value for p?

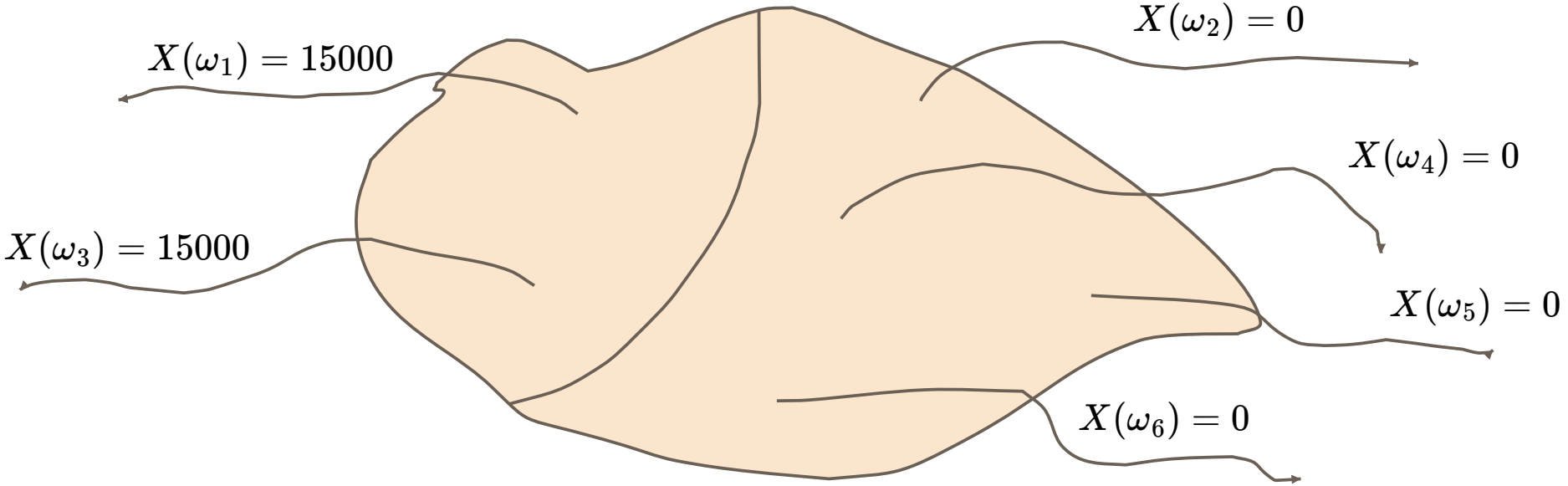
CC BY NC SA

Thought Experiment 4

Consider a nonnegative random variable X with $\mathbb{E}[X] = 5000$. Let $p = P(X \geq 15000)$.

What is the largest possible value for p?

- We know that $p \leq 1/3$.
- In the most extreme case, $X(\omega) = 15000$ for 1/3rd of the sample space and $X(\omega) = 0$ for the other 2/3rds of the sample space.



Thought Experiment 4

Consider a nonnegative random variable X with $\mathbb{E}[X] = \mu$. Let $p = P(X \geq c)$.

What is the largest possible value for p?

- We know that $p \leq \mu/c$.

This is known as Markov's Inequality, usually written concisely as:

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

CC BY NC SA

Markov's Inequality

Markov's Inequality: For a **nonnegative** random variable X (i.e. $X(\omega) \geq 0$ for all $\omega \in \Omega$) with finite mean, we have that for any positive constant c:

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

Proof from the notes, using an indicator random variable I. For all $\omega \in \Omega$:

$$X(\omega) \geq c\, I\{X(\omega) \geq c\} = \begin{cases} c \cdot 0 = 0, & \text{if } X(\omega) < c \\ c \cdot 1 = c, & \text{if } X(\omega) \geq c \end{cases}$$

← Assigning all "small" outcomes to zero.

← Assigning all "large" outcomes to c.

$$\mathbb{E}[X] \geq c\,\mathbb{E}[I\{X \geq c\}]$$

← Taking expectations of both sides.

$$= c\,P(X \geq c)$$

← The expectation of an indicator is the probability of its event.

CC BY NC SA

Example Using Markov's Inequality: Coin Toss

Markov's Inequality: For a **nonnegative** random variable X (i.e. $X(\omega) \geq 0$ for all $\omega \in \Omega$) with finite mean, we have that for any positive constant c:

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

You toss a fair coin n times. Let X be the number of heads. What does Markov's Inequality tell us about the probability that $X \geq \frac{3}{4}n$?

- joshh.ug/D

Hint: What is $\mathbb{E}[X]$, i.e. expected number of heads for n coin flips?

CC BY NC SA

Example Using Markov's Inequality: Coin Toss

Markov's Inequality: For a **nonnegative** random variable X (i.e. $X(\omega) \geq 0$ for all $\omega \in \Omega$) with finite mean, we have that for any positive constant c:

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

You toss a fair coin n times. Let X be the number of heads. What does Markov's Inequality tell us about the probability that $X \geq \frac{3}{4}n$?

CC BY NC SA

Example Using Markov's Inequality: Coin Toss

Markov's Inequality: For a **nonnegative** random variable X (i.e. $X(\omega) \geq 0$ for all $\omega \in \Omega$) with finite mean, we have that for any positive constant c:

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

You toss a fair coin n times. Let X be the number of heads. What does Markov's Inequality tell us about the probability that $X \geq \frac{3}{4}n$?

$$\mathbb{E}[X] = 0.5n$$

$$P\left(X \geq \frac{3}{4}n\right) \leq \frac{0.5n}{(3/4)n}$$

$$P\left(X \geq \frac{3}{4}n\right) \leq \frac{2}{3}$$

CC BY NC SA

Example Using Markov's Inequality: Coin Toss

Markov's Inequality: For a **nonnegative** random variable X (i.e. $X(\omega) \geq 0$ for all $\omega \in \Omega$) with finite mean, we have that for any positive constant c:

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

You toss a fair coin n times. Let X be the number of heads. What does Markov's Inequality tell us about the probability that $X \geq \frac{3}{4}n$?

$$P\left(X \geq \frac{3}{4}n\right) \leq \frac{2}{3}$$

For this example, is this bound "good" or perhaps "tight"?

- joshh.ug/E

Example Using Markov's Inequality: Coin Toss

Markov's Inequality: For a **nonnegative** random variable X (i.e. $X(\omega) \geq 0$ for all $\omega \in \Omega$) with finite mean, we have that for any positive constant c:

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

You toss a fair coin n times. Let X be the number of heads. What does Markov's Inequality tell us about the probability that $X \geq \frac{3}{4}n$?

$$P\left(X \geq \frac{3}{4}n\right) \leq \frac{2}{3}$$

For this example, is this bound "good" or perhaps "tight"?

- No, it says if you flip a coin 1,000 times, chance of getting at least 750 heads is less than 2/3rds. Actual chance is **far** lower (around 6.74×10^{-57} % chance).

Concentration Inequalities and Bound Tightness

Concentration inequalities are not tight bounds!

- Just like the union bound, they are bounds on a probability.

In the case of a binomial random variable, Markov's Inequality is basically useless.

- Markov's Inequality tells us that the probability of getting at least 3/4 heads on a fair coin is at most 66.67%.
- But this bound is absurdly loose. For 1,000 flips, the actual chance is around 6.74×10^{-57} %.

Note: Markov's Inequality can be a tight bound.

- Example: Scaled Bernoulli RV where $X = 10\mu$ with probability $1/10$ and 0 otherwise.

CC BY NC SA

Markov's Inequality

Markov's Inequality: For a **nonnegative** random variable X (i.e. $X(\omega) \geq 0$ for all $\omega \in \Omega$) with finite mean, we have that for any positive constant c:

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

What happens if we drop non-negativity?

- Naturally, Markov's Inequality as stated above is no longer true. Trivial example: one student gets -1,000,000 points on a midterm, then everyone else will do way better than twice the mean.

There exists a generalized version of Markov's Inequality that works even if X can be negative. Won't cover in lecture.

CC BY NC SA

Chebyshev's Inequality

Lecture 23, CS70 Fall 2026

Estimating p for X~Bernoulli(p)

Markov's Inequality

Chebyshev's Inequality

The Chebyshev Bound for Number of Heads

The Chebyshev Bound for Estimating the Fairness of a Coin

Weak Law of Large Numbers

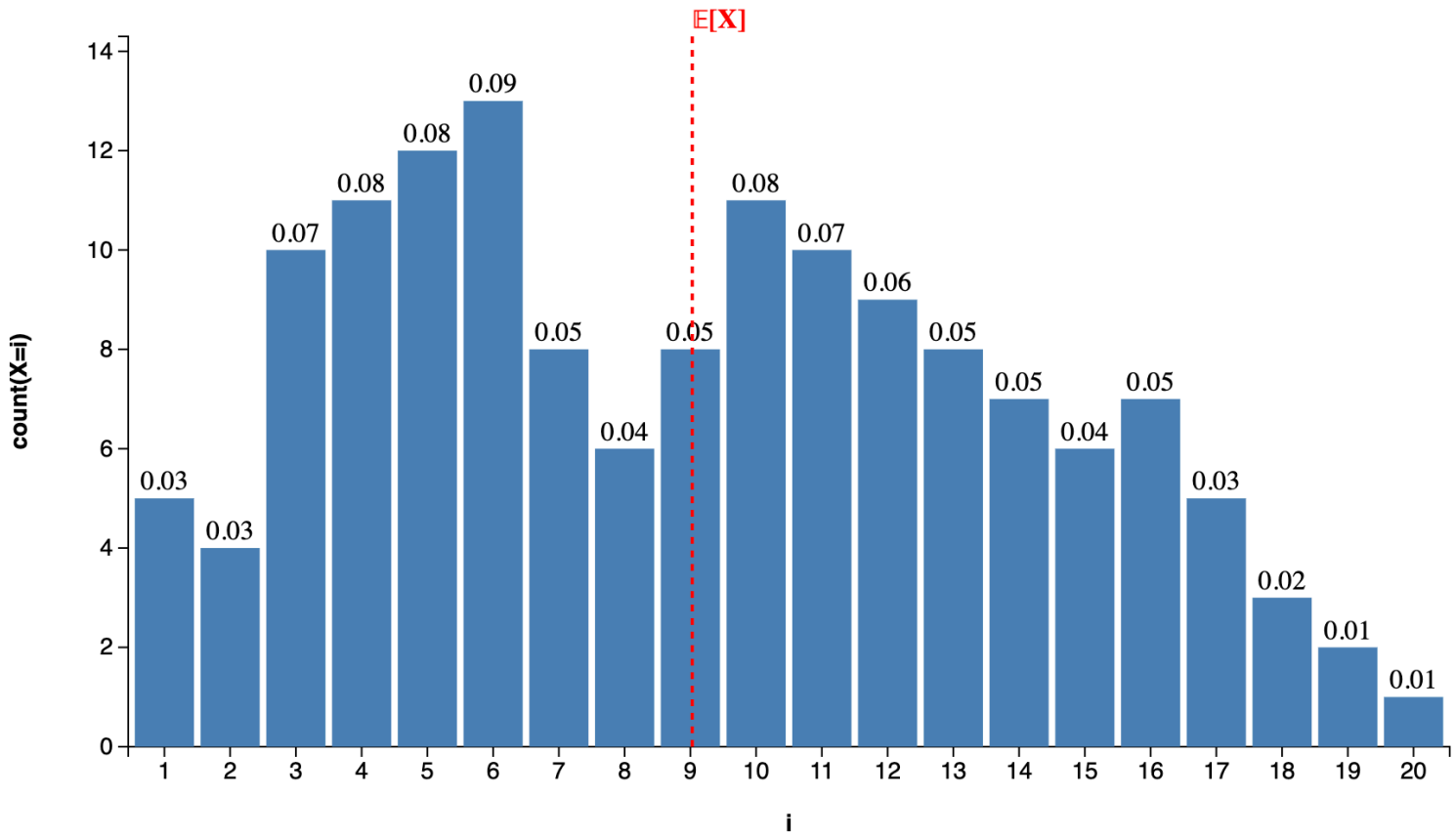
Bounds Beyond CS70

CC BY NC SA

Expectations and Distributions (Visually)

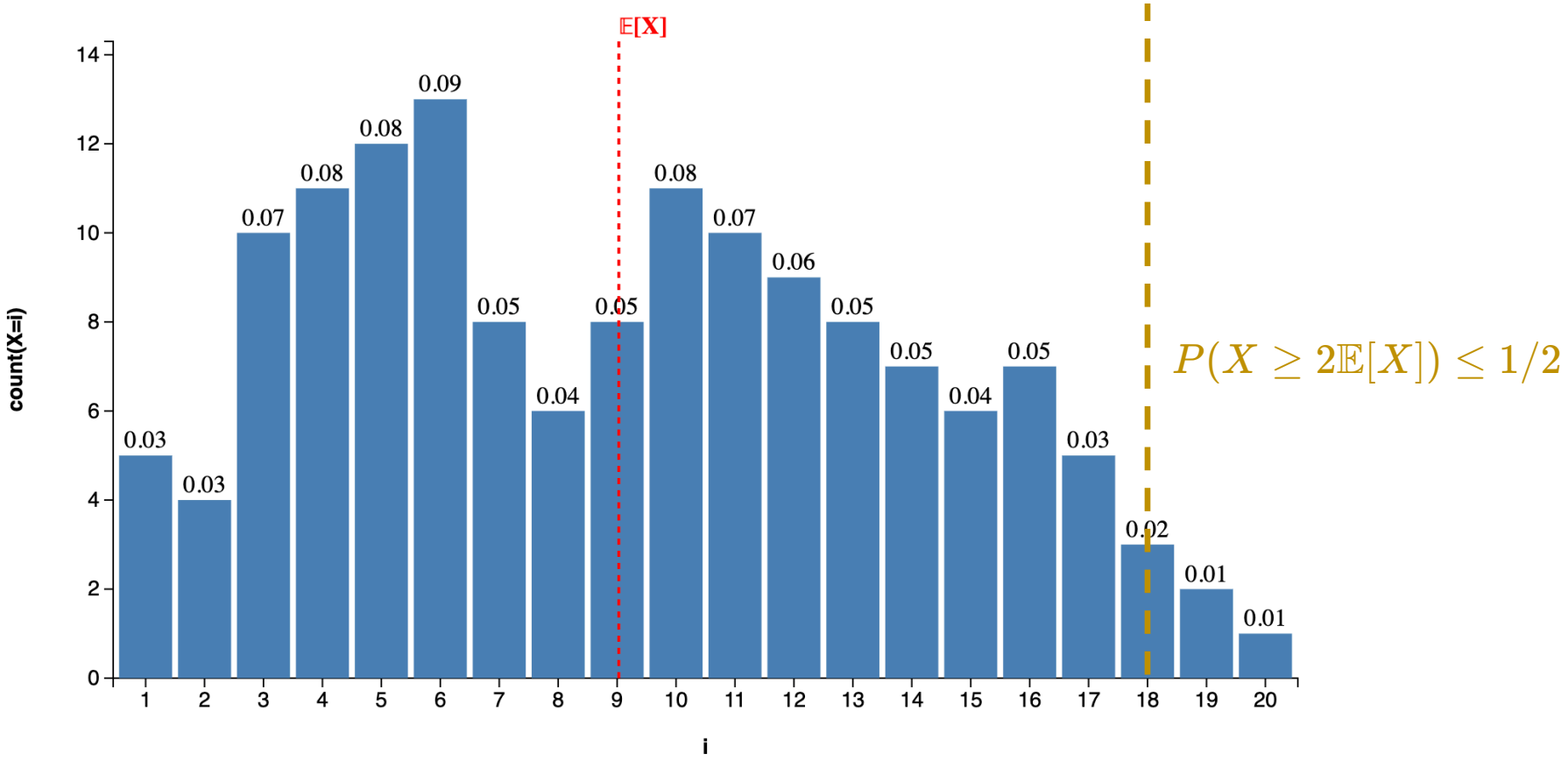
Consider the distribution below.

- We can think of the expectation as the "center of mass" of the distribution.
- Each value has mass equal to the height of the respective bar.



Markov's Inequality Visually

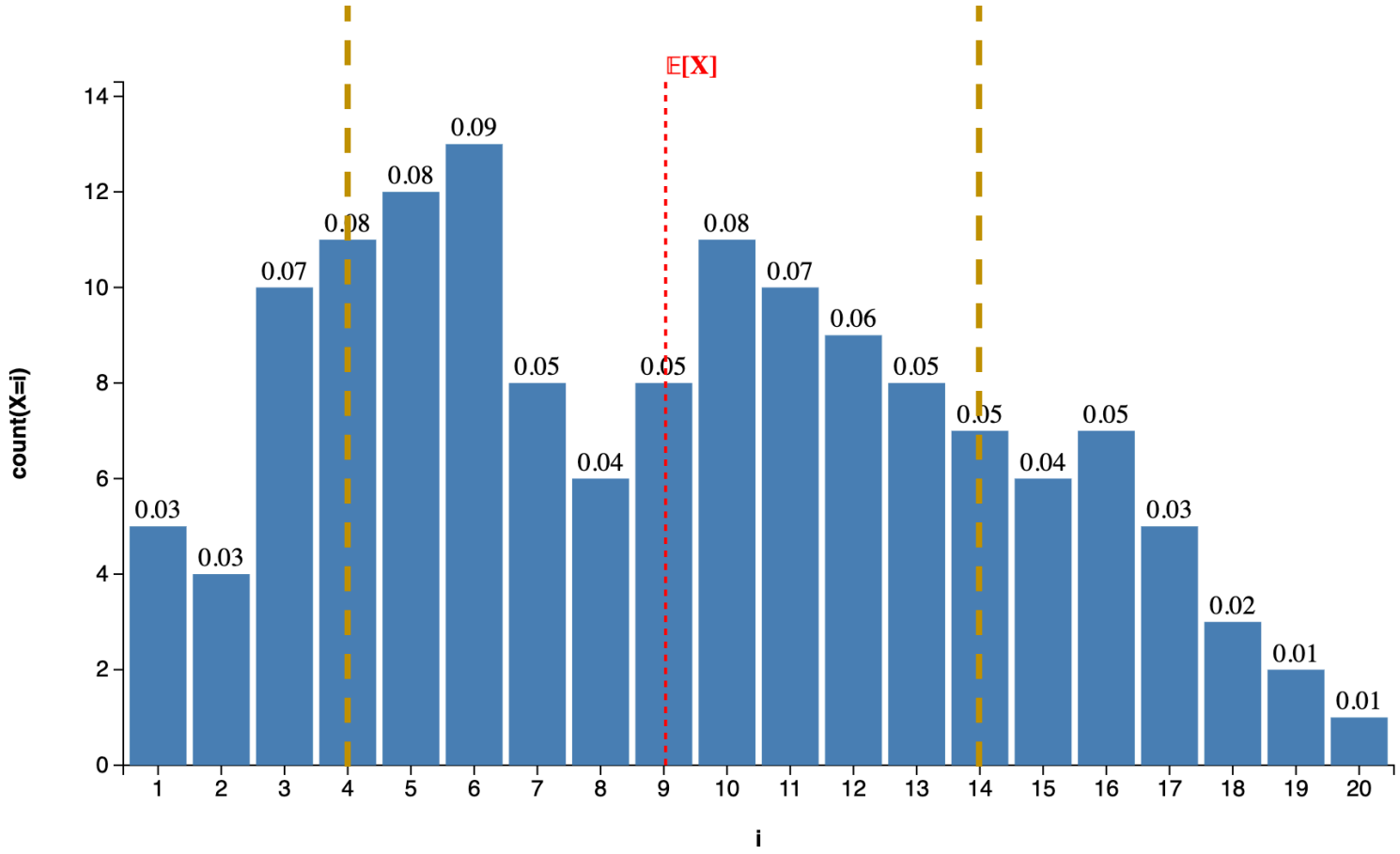
Markov's inequality gives us a bound on how much mass there can be above some constant c. Example with $c = 2\mathbb{E}[X]$ shown below.



Chebyshev's Inequality Visually

Observation: It might be more useful to have an inequality that bounds the amount of mass that is far from the mean (in either direction).

$$P(|X - \mu| \geq c) \leq \ ??$$



Chebyshev's Inequality

We'll use two summary statistics: $\mathbb{E}[X]$ and $\mathbf{var}(X)$ to provide a more useful bound on how far we stray from the mean.

Define **Chebyshev's Inequality** as follows. If X is a random variable with $\mathbb{E}[X] = \mu$, then for any constant $c > 0$:

$$P(|X - \mu| \geq c\,) \leq \frac{\mathbf{var}(X)}{c^2}$$

Before we set about proving this inequality, let's build some intuition for what it means.

cc BY NC SA

Chebyshev's Inequality: Intuition 1, the Alternate Form

Define **Chebyshev's Inequality** as follows. If X is a random variable with $\mathbb{E}[X] = \mu$, then for any constant $c > 0$:

$$P(|X - \mu| \geq c\,) \leq \frac{\text{var}(X)}{c^2}$$

Suppose we pick the constant $c = k\sigma$, what is the upper bound?

- joshh.ug/F

$$P(|X - \mu| \geq k\sigma) \leq \ ??$$



CC BY NC SA

Chebyshev's Inequality

Define **Chebyshev's Inequality** as follows. If X is a random variable with $\mathbb{E}[X] = \mu$, then for any constant $c > 0$:

$$P(|X - \mu| \geq c\,) \leq \frac{\mathbf{var}(X)}{c^2}$$

Suppose we pick the constant $c = k\sigma$, what is the upper bound?

CC BY NC SA

Chebyshev's Inequality

Define **Chebyshev's Inequality** as follows. If X is a random variable with $\mathbb{E}[X] = \mu$, then for any constant $c > 0$:

$$P(|X - \mu| \geq c\,) \leq \frac{\mathrm{var}(X)}{c^2}$$

Suppose we pick the constant $c = k\sigma$, what is the upper bound?

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

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

cc BY NC SA

Chebyshev's Inequality

Define **Chebyshev's Inequality** as follows. If X is a random variable with $\mathbb{E}[X] = \mu$, then for any constant $c > 0$:

$$P(|X - \mu| \geq c\,) \leq \frac{\mathrm{var}(X)}{c^2}$$

Or equivalently:

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

Suppose we have a random variable, what bound can we place on the probability that we are more than three standard deviations from the mean? More than one standard deviation from the mean?

- joshh.ug/G

Chebyshev's Inequality

Define **Chebyshev's Inequality** as follows. If X is a random variable with $\mathbb{E}[X] = \mu$, then for any constant $c > 0$:

$$P(|X - \mu| \geq c\,) \leq \frac{\mathrm{var}(X)}{c^2}$$

Or equivalently:

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

Suppose we have a random variable, what bound can we place on the probability that we are more than three standard deviations from the mean? More than one standard deviation from the mean?

- The chances are $\leq$ 11.11% and 100%, respectively.
- Note, for one standard deviation, the result is trivial, i.e. Chebyshev's tells us literally nothing. In general Chebyshev's is a loose bound.

CC BY NC SA

Chebyshev's Inequality

Define **Chebyshev's Inequality** as follows. If X is a random variable with $\mathbb{E}[X] = \mu$, then for any constant $c > 0$:

$$P(|X - \mu| \geq c\,) \leq \frac{\mathrm{var}(X)}{c^2}$$

Or equivalently:

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

Chebyshev's inequality is true of **ANY POSSIBLE RANDOM VARIABLE** (with finite variance).

- That is, the probability of being more than 3 standard deviations from the mean is less than 11.11% for any distribution (with finite variance)!
- Example implication: It is literally logically impossible for 12% of a class to score at least 3 standard deviations from the mean, no matter what!

cc BY NC SA

Chebyshev's is Still a Loose Bound: Normal Distribution Example

There's a distribution we'll talk about next week called the normal distribution.

- Probability that you are less than 3 standard deviations from the mean is 99.7%.
- Or equivalently: Probability that you are more than 3 standard deviations from the mean is 0.3%.

Chebyshev's is considerably looser. It says the probability of being more than 3 standard deviations from the mean is at most 11.1%.

Key difference: Chebyshev is true of all possible distributions. The 99.7% rule is only for the Normal.

CC BY NC SA

Chebyshev Proof

Define **Chebyshev's Inequality** as follows. If X is a random variable with $\mathbb{E}[X] = \mu$, then for any constant $c > 0$:

$$P(|X - \mu| \geq c\,) \leq \frac{\mathbf{var}(X)}{c^2}$$

Dead End Proof: Observe that $|X - \mu|$ is a random variable.

- Since $|X - \mu|$ is non-negative, we could use Markov's inequality to try to bound $P(|X - \mu| \geq c\,)$, but expectation of an absolute value is a pain.

cc BY NC SA

Chebyshev Proof

Define **Chebyshev's Inequality** as follows. If X is a random variable with $\mathbb{E}[X] = \mu$, then for any constant $c > 0$:

$$P(|X - \mu| \geq c\,) \leq \frac{\mathbf{var}(X)}{c^2}$$

Proof:

CC BY NC SA

Chebyshev Proof

Define **Chebyshev's Inequality** as follows. If X is a random variable with $\mathbb{E}[X] = \mu$, then for any constant $c > 0$:

$$P(|X - \mu| \geq c\,) \leq \frac{\mathbf{var}(X)}{c^2}$$

Proof:

- Observe that $|X - \mu|$ is a random variable.
- Observe that for any outcome ω for which $|X(\omega) - \mu| \geq c$, we also have that $(X(\omega) - \mu\,)^2 \geq c^2$ (and vice versa). This means that:

$$P(|X - \mu| \geq c\,) = P((X - \mu)^2 \geq c^2\,)$$

- Let $Y = (X - \mu)^2$.
- Then by Markov's inequality: $P(Y \geq c^2) \leq \frac{\mathbb{E}[Y]}{c^2}$

cc BY NC SA

Chebyshev Proof

Define **Chebyshev's Inequality** as follows. If X is a random variable with $\mathbb{E}[X] = \mu$, then for any constant $c > 0$:

$$P(|X - \mu| \geq c\,) \leq \frac{\mathrm{var}(X)}{c^2}$$

Proof:

- Observe that $|X - \mu|$ is a random variable.
- Observe that for any outcome ω for which $|X(\omega) - \mu| \geq c$, we also have that $(X(\omega) - \mu\,)^2 \geq c^2$ (and vice versa). This means that:

$$P(|X - \mu| \geq c\,) = P((X - \mu)^2 \geq c^2\,)$$

- Let $Y = (X - \mu)^2$.
- Then by Markov's inequality: $P(Y \geq c^2) \leq \frac{\mathbb{E}[Y]}{c^2}$

What is $\mathbb{E}[Y]$?
joshh.ug/H

CC BY NC SA

Chebyshev Proof

Define **Chebyshev's Inequality** as follows. If X is a random variable with $\mathbb{E}[X] = \mu$, then for any constant $c > 0$:

$$P(|X - \mu| \geq c\,) \leq \frac{\mathbf{var}(X)}{c^2}$$

Proof:

- Observe that $|X - \mu|$ is a random variable.
- Observe that for any outcome ω for which $|X(\omega) - \mu| \geq c$, we also have that $(X(\omega) - \mu\,)^2 \geq c^2$ (and vice versa). This means that:

$$P(|X - \mu| \geq c\,) = P((X - \mu)^2 \geq c^2\,)$$

- Let $Y = (X - \mu)^2$. $\mathbb{E}[Y] = \mathbf{var}(X)$
- Then by Markov's inequality: $P(Y \geq c^2) \leq \frac{\mathbb{E}[Y]}{c^2} = \frac{\mathbf{var}(X)}{c^2}$

CC BY NC SA

Chebyshev's Inequality and Markov's Inequality

Even though Markov's Inequality is weak for understanding X, we were able to enhance its power by considering a new random variable $Y = (X - \mu)^2$ and applying Markov's Inequality to this new variable.

- Since $\mathbb{E}[Y] = \mathbb{E}[(X - \mu)^2] = \mathbf{var}(X)$, our resulting bound was in terms of the variance.

	Equation	Interpretation	Tight?	Condition	Requires Computing
Markov's Inequality	$P(X \geq c) \leq \frac{\mu}{c}$	Gives probability that a random variable exceeds some value.	Very loose	$X \geq 0$	μ
Chebyshev's Inequality	$P(\lvert X - \mu \rvert \geq c) \leq \frac{\mathbf{var}(X)}{c^2}$	Gives probability that a random variable deviates from the mean by at least c.	Loose	$\mathbf{var}(X)$ finite	$\mathbb{E}[(X - \mu)^2] = \mathbf{var}(X)$

CC BY NC SA

Chebyshev's Inequality and Markov's Inequality

Even though Markov's Inequality is weak for understanding X, we were able to enhance its power by considering a new random variable $Y = (X - \mu)^2$ and applying Markov's Inequality to this new variable.

- Yes, you can enhance the bound further by making up new random variables such as $Z = (X - \mu)^{10}$, but then you need to be able to compute $\mathbb{E}[(X - \mu)^{10}]$.
- We call $\mathbb{E}[(X - \mu)^i]$ the ith "central moment" of X (beyond the scope of 70).
- If you use infinite moments $i \in \{1, 2, \ldots\}$, you get the Chernoff bound.

	Equation	Interpretation	Tight?	Condition	Requires Computing
Markov's Inequality	$P(X \geq c) \leq \frac{\mu}{c}$	Gives probability that a random variable exceeds some value.	Very loose	$X \geq 0$	μ
Chebyshev's Inequality	$P(\lvert X - \mu \rvert \geq c) \leq \frac{\mathrm{var}(X)}{c^2}$	Gives probability that a random variable deviates from the mean by at least c.	Loose	$\mathrm{var}(X)$ finite	$\mathbb{E}[(X - \mu)^2] = \mathrm{var}(X)$

CC BY NC SA

The Chebyshev Bound for Estimating Probability of a Coin Flip Event

Lecture 23, CS70 Fall 2026

Estimating p for X~Bernoulli(p)

Markov's Inequality

Chebyshev's Inequality

The Chebyshev Bound for Number of Heads

The Chebyshev Bound for Estimating the Fairness of a Coin

Weak Law of Large Numbers

Bounds Beyond CS70

CC BY NC SA

Application: Probability of Coin Toss Outcomes

Earlier, we tried to use Markov's Inequality to understand the behavior of $X \sim \text{Binomial}(n, 0.5)$.

Specifically, we asked: What is the chance that at least 3/4 of our flips come up heads?

Using Markov's Inequality, we found that $P\left(X \geq \frac{3}{4}n\right) \leq \frac{2}{3}$.

This bound was not useful! As we flip more coins, the actual probability drops.

- If we use Chebyshev's Inequality instead, we'll see a dependence on n.

CC BY NC SA

Application: Probability of Coin Toss Outcomes

For $X \sim \text{Binomial}(n,\ 0.5)$, let's bound $P\left(X \geq \frac{3}{4}n\right)$, using Chebyshev's Inequality:
If X is a random variable with $\mathbb{E}[X] = \mu$, then for any constant $c > 0$:

$$P(|X - \mu| \geq c) \leq \frac{\text{var}(X)}{c^2}$$

CC BY NC SA

Application: Probability of Coin Toss Outcomes

For $X \sim \text{Binomial}(n,\ 0.5)$, let's bound $P\left(X \geq \frac{3}{4}n\right)$, using Chebyshev's Inequality:
If X is a random variable with $\mathbb{E}[X] = \mu$, then for any constant $c > 0$:

$$P(|X - \mu| \geq c) \leq \frac{\text{var}(X)}{c^2}$$

First, we write expectation and variance: $\mathbb{E}[X] = \mu = \frac{n}{2}$ $\text{var}(X) = n \cdot p \cdot (1-p) = \frac{n}{4}$

Then, get $P\left(X \geq \frac{3}{4}n\right)$ into the appropriate form to apply Chebyshev:

$$P\left(X \geq \frac{3}{4}n\right) = P\left(X - \mu \geq \frac{3}{4}n - \mu\right) = P\left(X - \mu \geq \frac{n}{4}\right)$$

CC BY NC SA

Application: Probability of Coin Toss Outcomes

For $X \sim \text{Binomial}(n,\ 0.5)$, let's bound $P\left(X \geq \frac{3}{4}n\right)$, using Chebyshev's Inequality:
If X is a random variable with $\mathbb{E}[X] = \mu$, then for any constant $c > 0$:

$$P(|X - \mu| \geq c) \leq \frac{\text{var}(X)}{c^2}$$

First, we write expectation and variance: $\mathbb{E}[X] = \mu = \frac{n}{2}$ $\text{var}(X) = n \cdot p \cdot (1-p) = \frac{n}{4}$

Then, get $P\left(X \geq \frac{3}{4}n\right)$ into the appropriate form to apply Chebyshev:

$$P\left(X \geq \frac{3}{4}n\right) = P\left(X - \mu \geq \frac{3}{4}n - \mu\right) = P\left(X - \mu \geq \frac{n}{4}\right) \leq P\left(|X - \mu| \geq \frac{n}{4}\right)$$

$$\leq \frac{\text{var}(X)}{c^2} = \frac{\text{var}(X)}{\left(\frac{n}{4}\right)^2} = \frac{n/4}{\left(\frac{n}{4}\right)^2} = \frac{4}{n}$$

$P(R \geq a) \leq P(|R| \geq a)$

CC BY NC SA

Interpretation in Terms of Voting

For $X \sim \text{Binomial}(n, 0.5)$, let's bound $P\left(X \geq \frac{3}{4}n\right)$, using Chebyshev's Inequality:
If X is a random variable with $\mathbb{E}[X] = \mu$, then for any constant $c > 0$:

$$P(|X - \mu| \geq c) \leq \frac{\text{var}(X)}{c^2}$$

Our Chebyshev bound is: $P\left(X \geq \frac{3}{4}n\right) \leq \frac{4}{n}$

Example: In an election where the chance of voting for candidate A is $p = 0.5$.

- Chebyshev bound says the chance of getting at least 75% of the vote is less than 0.4% if 1000 people vote.
- But this is a somewhat silly use of the Chebyshev bound, we already know how to compute this chance exactly (and it's much smaller than 0.4%).
- Let's instead use Chebyshev in a more useful way: Estimating p itself.

CC BY NC SA

The Chebyshev Bound for Estimating the Fairness of a Coin

Lecture 23, CS70 Fall 2026

Estimating p for X~Bernoulli(p)

Markov's Inequality

Chebyshev's Inequality

The Chebyshev Bound for Number of Heads

The Chebyshev Bound for Estimating the Fairness of a Coin

Weak Law of Large Numbers

Bounds Beyond CS70

CC BY NC SA

Back to Where We Started: Estimating p for a Coin

We have a coin with unknown heads probability p. We estimate it by tossing a coin n times.

- X is the number of heads. $X = X_1 + X_2 + \cdots + X_n$.
- Our empirical mean is $\hat{p} = \frac{X}{n}$

Goal: How big does n need to be so that $P(|\hat{p} - p| \geq \epsilon) \leq \delta$?

- In other words, how big does n need to be so that we have at most a δ chance of being ϵ or more away from the true p?

CC BY NC SA

Back to Estimating p for a Coin

We have a coin with unknown heads probability p. We estimate it by tossing a coin n times.

- X is the number of heads. $X = X_1 + X_2 + \cdots + X_n$.
- Our empirical mean is $\hat{p} = \frac{X}{n}$

Observation: we can bound $P(|\hat{p} - p| \geq \epsilon)$ using Chebyshev's inequality.

CC BY NC SA

Back to Estimating p for a Coin

We have a coin with unknown heads probability p. We estimate it by tossing a coin n times.

- X is the number of heads. $X = X_1 + X_2 + \cdots + X_n$.
- Our empirical mean is $\hat{p} = \dfrac{X}{n}$

Observation: we can bound $P(|\hat{p} - p| \geq \epsilon)$ using Chebyshev's inequality.

- Since $\mathbb{E}[\hat{p}] = p$, we have that:
- So now we need to know $\mathrm{var}(\hat{p})$.

$$P(|\hat{p} - \mathbb{E}[\hat{p}]| \geq \epsilon) = P(|\hat{p} - p| \geq \epsilon) \leq \frac{\mathrm{var}(\hat{p})}{\epsilon^2}$$

$$\mathrm{var}(\hat{p}) = \mathrm{var}\left(\frac{X}{n}\right) = \frac{\mathrm{var}(X)}{n^2} = \frac{\mathrm{var}(X_1 + \cdots + X_n)}{n^2} = \frac{n \cdot \mathrm{var}(X_1)}{n^2} = \frac{n \cdot p(1-p)}{n^2}$$

$$= \frac{p(1-p)}{n}$$

CC BY NC SA

Back to Estimating p for a Coin

We have a coin with unknown heads probability p. We estimate it by tossing a coin n times.

- X is the number of heads. $X = X_1 + X_2 + \cdots + X_n$.
- Our empirical mean is $\hat{p} = \dfrac{X}{n}$

Observation: we can bound $P(|\hat{p} - p| \geq \epsilon)$ using Chebyshev's inequality.

- Since $\mathbb{E}[\hat{p}] = p$, and $\mathbf{var}(\hat{p}) = p(1-p)/n$ we have that:

$$P(|\hat{p} - \mathbb{E}[\hat{p}]| \geq \epsilon) = P(|\hat{p} - p| \geq \epsilon) \leq \frac{p(1-p)}{n\epsilon^2}$$

At this point, it might seem like we're a bit stuck! The variance of $\hat{p}$ seems to depend on p, which is what we're trying to estimate! Any idea how to proceed?

CC BY NC SA

Back to Estimating p for a Coin

We have a coin with unknown heads probability p. We estimate it by tossing a coin n times.

- X is the number of heads. $X = X_1 + X_2 + \cdots + X_n$.
- Our empirical mean is $\hat{p} = \frac{X}{n}$

Observation: we can bound $P(|\hat{p} - p| \geq \epsilon)$ using Chebyshev's inequality.

- Since $\mathbb{E}[\hat{p}] = p$, and $\mathbf{var}(\hat{p}) = p(1-p)/n$ we have that:

$$P(|\hat{p} - \mathbb{E}[\hat{p}]| \geq \epsilon) = P(|\hat{p} - p| \geq \epsilon) \leq \frac{p(1-p)}{n\epsilon^2}$$

Solution: Be conservative and replace $p(1-p)$ with its maximum possible value. What is it? joshh.ug/J

CC BY NC SA

Back to Estimating p for a Coin

We have a coin with unknown heads probability p. We estimate it by tossing a coin n times.

- X is the number of heads. $X = X_1 + X_2 + \cdots + X_n$.
- Our empirical mean is $\hat{p} = \dfrac{X}{n}$

Observation: we can bound $P(|\hat{p} - p| \geq \epsilon)$ using Chebyshev's inequality.

- Since $\mathbb{E}[\hat{p}] = p$, and $\mathbf{var}(\hat{p}) = p(1-p)/n$ we have that:

$$P(|\hat{p} - \mathbb{E}[\hat{p}]| \geq \epsilon) = P(|\hat{p} - p| \geq \epsilon) \leq \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}$$

Solution: Be conservative and replace $p(1-p)$ with its maximum possible value of 1/4, which we get when $p = 0.5$.

CC BY NC SA

Back to Estimating p for a Coin

We have a coin with unknown heads probability p. We estimate it by tossing a coin n times.

- X is the number of heads. $X = X_1 + X_2 + \cdots + X_n$.
- Our empirical mean is $\hat{p} = \frac{X}{n}$

Using Chebyshev's inequality we have shown: $P(|\hat{p} - p| \geq \epsilon) \leq \frac{1}{4n\epsilon^2}$

We want an n such that $\frac{1}{4n\epsilon^2} \leq \delta$

So, to achieve this goal, we multiply both sides by n and divide by δ yielding:

$$\frac{1}{4\delta\epsilon^2} \leq n$$

CC BY NC SA

Weak Law of Large Numbers

Lecture 23, CS70 Fall 2026

Estimating p for X~Bernoulli(p)

Markov's Inequality

Chebyshev's Inequality

The Chebyshev Bound for Number of Heads

The Chebyshev Bound for Estimating the Fairness of a Coin

Weak Law of Large Numbers

Bounds Beyond CS70

CC BY NC SA

Weak Law of Large Numbers

As a quick detour before wrapping up today, let's talk about the **Weak Law of Large Numbers**.

If $X_1, X_2, X_3, \ldots$ are independent and identically distributed (i.i.d.) random variables with $\mathbb{E}[X_i] = \mu$, and $\mathbf{var}(X_i) = \sigma^2 < \infty$, then for every $\epsilon > 0$, we have that:

$$\lim_{n \to \infty} P\left(\left| \frac{X_1 + X_2 + \cdots + X_n}{n} - \mu \right| \geq \epsilon \right) = 0$$

Informal interpretation: As we run the same experiment over and over, the difference between the empirical average and the expectation converges to zero.

CC BY NC SA

Weak Law of Large Numbers

If $X_1, X_2, X_3, \ldots$ are independent and identically distributed (i.i.d.) random variables with $\mathbb{E}[X_i] = \mu$, and $\mathbf{var}(X_i) = \sigma^2 < \infty$, then for every $\epsilon > 0$, we have that:

$$\lim_{n \to \infty} P\left(\left| \frac{X_1 + X_2 + \cdots + X_n}{n} - \mu \right| \geq \epsilon \right) = 0$$

Proof: Let $\sigma^2 = \mathbf{var}(X_i)$. Let $S_n = X_1 + \cdots + X_n$. Let $Y_n = S_n / n$

Questions:

- What is $\mathbb{E}[Y_n]$?
- What is $\mathbf{var}(Y_n)$?
- Consider doing these as an exercise rather than just watching / reading.

CC BY NC SA

Weak Law of Large Numbers

If $X_1, X_2, X_3, \ldots$ are independent and identically distributed (i.i.d.) random variables with $\mathbb{E}[X_i] = \mu$, and $\mathbf{var}(X_i) = \sigma^2 < \infty$, then for every $\epsilon > 0$, we have that:

$$\lim_{n \to \infty} P\left(\left| \frac{X_1 + X_2 + \cdots + X_n}{n} - \mu \right| \geq \epsilon \right) = 0$$

Proof: Let $\sigma^2 = \mathbf{var}(X_i)$. Let $S_n = X_1 + \cdots + X_n$. Let $Y_n = S_n/n$

Questions:

- What is $\mathbf{var}(Y_n)$?

CC BY NC SA

Weak Law of Large Numbers

If $X_1, X_2, X_3, \ldots$ are independent and identically distributed (i.i.d.) random variables with $\mathbb{E}[X_i] = \mu$, and $\mathbf{var}(X_i) = \sigma^2 < \infty$, then for every $\epsilon > 0$, we have that:

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) = 0$$

Proof: Let $\sigma^2 = \mathbf{var}(X_i)$. Let $S_n = X_1 + \cdots + X_n$. Let $Y_n = S_n/n$

Questions:

- What is $\mathbb{E}[Y_n]$? $\mathbb{E}[Y_n] = \mathbb{E}[(X_1 + \cdots + X_n)/n] = (\mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n])/n = \mu n/n = \mu$
- What is $\mathbf{var}(Y_n)$?

$$\mathbf{var}(Y_n) = \mathbf{var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\mathbf{var}(X_1 + \cdots + X_n)}{n^2} = \frac{n\,\mathbf{var}(X_i)}{n^2} = \frac{\sigma^2}{n}$$

CC BY NC SA

Weak Law of Large Numbers

If $X_1,\ X_2,\ X_3, \ldots$ are independent and identically distributed (i.i.d.) random variables with $\mathbb{E}[X_i] = \mu$, and $\mathbf{var}(X_i) = \sigma^2 < \infty$, then for every $\epsilon > 0$, we have that:

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) = 0$$

Proof: Let $\sigma^2 = \mathbf{var}(X_i)$. Let $S_n = X_1 + \cdots + X_n$. Let $Y_n = S_n/n$

- $\mathbb{E}[Y_n] = \mu$
- $\mathbf{var}(Y_n) = \sigma^2/n$

Chebyshev's inequality tells us:

Weak Law of Large Numbers

If $X_1, X_2, X_3, \ldots$ are independent and identically distributed (i.i.d.) random variables with $\mathbb{E}[X_i] = \mu$, and $\mathbf{var}(X_i) = \sigma^2 < \infty$, then for every $\epsilon > 0$, we have that:

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) = 0$$

Proof: Let $\sigma^2 = \mathbf{var}(X_i)$. Let $S_n = X_1 + \cdots + X_n$. Let $Y_n = S_n/n$

- $\mathbb{E}[Y_n] = \mu$
- $\mathbf{var}(Y_n) = \sigma^2/n$

Chebyshev's inequality tells us:

$$P(|Y_n - \mu| \geq \epsilon) \leq \frac{\mathbf{var}(Y_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \qquad \text{So:} \qquad \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0$$

CC BY NC SA

Bounds Beyond CS70

Lecture 23, CS70 Fall 2026

Estimating p for X~Bernoulli(p)

Markov's Inequality

Chebyshev's Inequality

The Chebyshev Bound for Number of Heads

The Chebyshev Bound for Estimating the Fairness of a Coin

Weak Law of Large Numbers

Bounds Beyond CS70

CC BY NC SA

Bounds Beyond 70

Wrapping up, let's consider a few extra ideas that are a bit outside of the main scope of CS70:

- Confidence Intervals.
- Chernoff Bounds and the Cost of Accuracy vs. Confidence.
- Why Polling is Hard

CC BY NC SA

Confidence Intervals

Suppose we're trying to find the probability p that a coin comes up heads.

- You tossed it 1,000,000 times and you got heads 35% of the time.
- How confident are you that $p = 0.35$ exactly?
- How confident are you that $p \in (0.3,\ 0.4)$?

CC BY NC SA

Confidence Intervals

Suppose we're trying to find the probability p that a coin comes up heads.

- You tossed it 1,000,000 times and you got heads 35% of the time.
- How confident are you that $p = 0.35$ exactly? It's definitely not.
- How confident are you that $p \in (0.3,\ 0.4)$? Use $\delta = 1/(4n\epsilon^2)$, where $\epsilon = 0.05$ and $n = 1,000,000$.

CC BY NC SA

Confidence Intervals

Suppose we're trying to find the probability p that a coin comes up heads.

- You tossed it 1,000,000 times and you got heads 35% of the time.
- How confident are you that $p \in (0.3,\ 0.4)$? Use $\delta = 1/(4n\epsilon^2)$, where $\epsilon = 0.05$ and $n = 1,000,000$.

This is what is called a "confidence interval". Common in experimental science.

- Definition: An interval $[a,\ b]$ is a **95% confidence interval** for an unknown quantity p if:

$$P(p \in [a,\ b]) \geq 0.95$$

This interval $[a, b]$ is calculated based on an experiment.

- Note: In this definition, a and b are random variables, and p is some constant.

CC BY NC SA

Coin Flip Bound Tightness

We can do better than $n \geq \frac{1}{4\epsilon^2\delta}$

Turns out we can get a tighter bound using something called a Chernoff bound.

$$n \geq \frac{1}{2\epsilon^2} \ln\left(\frac{2}{\delta}\right)$$

Observations:

- Confidence is cheap. Example:
  - $\epsilon = 0.01$, $\delta = 10\%$ requires ~15,000 flips. $\epsilon = 0.01$, $\delta = 1\%$ requires 26,000.
- Accuracy is expensive.
  - $\epsilon = 0.01$, $\delta = 10\%$ requires ~15,000 flips. $\epsilon = 0.001$, $\delta = 10\%$ requires 1,500,000 flips.

CC BY NC SA

Polling in the Real World

The Chernoff bound tells us that if we want to know the election results to within 1% with 99% certainty, we just need to poll 26,000 people.

Why is this hard in reality?

- You need to get 26,000 people to respond.
- Then you need to ensure that "person responds" and "person votes for James Talarico" are independent.

Note: It's possible to make an even tighter bound than the Chernoff bound, but to do that, we'll need to consider continuous rather than discrete random variables.

- Discrete random variables have a countable range.
- Continuous random variables have an uncountable range. More next time!

CC BY NC SA

Extra Slide: Markov's Inequality (Alternate Proof)

Markov's Inequality: For a **nonnegative** random variable X (i.e. $X(\omega) \geq 0$ for all $\omega \in \Omega$) with finite mean, we have that for any positive constant c:

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

Proof from the notes (exact same idea that we used) using mega-verbose form of expectation:

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \times P(\omega)$$

Assigning all "small" outcomes to zero. ⟶

$$\geq \sum_{\omega : X(\omega) \geq c} X(\omega) \times P(\omega)$$

Assigning all "large" outcomes to c. ⟶

$$\geq \sum_{\omega : X(\omega) \geq c} c \times P(\omega)$$

$$= c \sum_{\omega : X(\omega) \geq c} P(\omega) = cP(X \geq c)$$

CC BY NC SA

Extra Slide: Markov's Inequality (Alternate Proof)

Markov's Inequality: For a **nonnegative** random variable X (i.e. $X(\omega) \geq 0$ for all $\omega \in \Omega$) with finite mean, we have that for any positive constant c:

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

The same proof, summing over the values a of X, as in the notes:

$$\mathbb{E}[X] = \sum_{a} a \times P(X = a)$$

Assigning all "small" outcomes to zero. ⟶

$$\geq \sum_{a \geq c} a \times P(X = a)$$

Assigning all "large" outcomes to c. ⟶

$$\geq \sum_{a \geq c} c \times P(X = a)$$

$$= c\,P(X \geq c)$$

CC BY NC SA