Tail Bounds

CSE 312 Summer 21
Lecture 19

Announcements

Point values for Question 4 in Problem Set 6 has been updated.

The Distinct Elements question has been updated to direct you to the correct page on the textbook. Look for the CDF on page 334.

What’s a Tail Bound?

When we were finding our margin of error, we didn’t need an exact calculation of the probability.

We needed an inequality: the probability of being outside the margin of error was **at most** 5% (the example discussed mentioned that most of the data lied within the margin of error at least 95% of the time).

A tail bound (or concentration inequality) is a statement that bounds the probability in the “tails” of the distribution (says there’s very little probability far from the center) or (equivalently) says that the probability is concentrated near the expectation.

Our First bound

Two statements are equivalent.
Left form is often easier to use.
Right form is more intuitive.

Markov's Inequality	Markov's Inequality
Let X be a random variable supported (only) on non-negative numbers. For any $t > 0$ $$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$$	Let X be a random variable supported (only) on non-negative numbers. For any $\mathrm{k} > 0$ $$\mathbb{P}(X \geq k\mathbb{E}[X]) \leq \frac{1}{k}$$

To apply this bound you only need to know:

1. it's non-negative
2. Its expectation.

Proof

$$\mathbb{E}[X] = \sum_{x \in \Omega} x \cdot \mathbb{P}(X = x)$$
$$= \sum_{x : x \geq t} x \cdot \mathbb{P}(X = x) + \sum_{x : x < t} x \cdot \mathbb{P}(X = x)$$
$$\geq \sum_{x : x \geq t} x \cdot \mathbb{P}(X = x) + 0$$
$$\geq \sum_{x : x \geq t} t \cdot \mathbb{P}(X = x)$$
$$= t \cdot \sum_{x : x \geq t} \mathbb{P}(X = x)$$
$$= t \cdot \mathbb{P}(X \geq t)$$

$x \geq 0$ whenever $\mathbb{P}(X = x) > 0$

Markov's Inequality

Let X be a random variable supported (only) on non-negative numbers. For any $t > 0$

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

Example with geometric RV

Suppose you roll a fair (6-sided) die until you see a 6. Let X be the number of rolls.

Bound the probability that $X \geq 12$

Markov's Inequality

Let X be a random variable supported (only) on non-negative numbers. For any $t > 0$

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

Example with geometric RV

Suppose you roll a fair (6-sided) die until you see a 6. Let X be the number of rolls.

Bound the probability that $X \geq 12$

$$\mathbb{P}(X \geq 12) \leq \frac{\mathbb{E}[X]}{12} = \frac{6}{12} = \frac{1}{2}.$$

Exact probability?

$$1 - \mathbb{P}(X < 12) \approx 1 - 0.865 = 0.135$$

Markov's Inequality

Let X be a random variable supported (only) on non-negative numbers. For any $t > 0$

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

A Second Example

Suppose the average number of ads you see on a website is 25. Give an upper bound on the probability of seeing a website with 75 or more ads.

Fill out the poll everywhere so Kushal knows how long to explain
Go to pollev.com/cse312su21

Markov's Inequality

Let X be a random variable supported (only) on non-negative numbers. For any $t > 0$

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

A Second Example

Suppose the average number of ads you see on a website is 25. Give an upper bound on the probability of seeing a website with 75 or more ads.

$$\mathbb{P}(X \geq 75) \leq \frac{\mathbb{E}[X]}{75} = \frac{25}{75} = \frac{1}{3}$$

Markov's Inequality

Let X be a random variable supported (only) on non-negative numbers. For any $t > 0$

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

Useless Example

Suppose the average number of ads you see on a website is 25. Give an upper bound on the probability of seeing a website with 20 or more ads.

Fill out the poll everywhere so Kushal knows how long to explain
Go to pollev.com/cse312su21

Markov's Inequality

Let X be a random variable supported (only) on non-negative numbers. For any $t > 0$

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

Useless Example

Suppose the average number of ads you see on a website is 25. Give an upper bound on the probability of seeing a website with 20 or more ads.

$$\mathbb{P}(X \geq 20) \leq \frac{\mathbb{E}[X]}{20} = \frac{25}{20} = 1.25$$

Well, that's…true. Technically.

But without more information we couldn't hope to do much better. What if every page gives exactly 25 ads? Then the probability really is 1.

So…what do we do?

A better inequality!

We’re trying to bound the tails of the distribution.

What parameter of a random variable describes the tails?

The variance!

Chebyshev's Inequality

Two statements are equivalent.
Left form is often easier to use.
Right form is more intuitive.

Chebyshev's Inequality

Let X be a random variable. For any $t > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\text{Var}(X)}{t^2}$$

Chebyshev's Inequality

Let X be a random variable. For any $\mathbf{k} > 0$

$$\mathbb{P}\left(|X - \mathbb{E}[X]| \geq k\sqrt{\text{Var}(X)}\right) \leq \frac{1}{k^2}$$

Proof of Chebyshev

Chebyshev's Inequality

Let X be a random variable. For any $t > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\text{Var}(X)}{t^2}$$

Let $Z = X - \mathbb{E}[X]$

$$\mathbb{P}(|Z| \geq t) = \mathbb{P}\left(Z^2 \geq t^2\right) \leq \frac{\mathbb{E}[Z^2]}{t^2} = \frac{\mathbb{E}[Z^2] - (\mathbb{E}[Z])^2}{t^2} = \frac{\text{Var}(Z)}{t^2} = \frac{\text{Var}(X)}{t^2}$$

Markov's Inequality

$\mathbb{E}[Z] = 0$

Inequalities are equivalent (square each side).

Z is just X shifted. Variance is unchanged.

Example with geometric RV (again)

Suppose you roll a fair (6-sided) die until you see a 6. Let X be the number of rolls.

Bound the probability that $X \geq 12$

Chebyshev's Inequality

Let X be a random variable. For any $t > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\text{Var}(X)}{t^2}$$

Example with geometric RV (again)

Suppose you roll a fair (6-sided) die until you see a 6. Let X be the number of rolls.

Bound the probability that $X \geq 12$

$$\mathbb{P}(X \geq 12) \leq \mathbb{P}(|X - 6| \geq 6) \leq \frac{\frac{5/6}{1/36}}{6^2} = \frac{5}{6}$$

Not any better than Markov ☹

Chebyshev's Inequality

Let X be a random variable. For any $t > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$$

Example with geometric RV (diff bound)

Let X be a geometric rv with parameter p

Bound the probability that $X \geq \frac{2}{p}$

$$\mathbb{P}\left(X \geq \frac{2}{p}\right) \leq \mathbb{P}\left(\left|X - \frac{1}{p}\right| \geq \frac{1}{p}\right) \leq \frac{\frac{1-p}{p^2}}{\frac{1}{p^2}} = 1 - p$$

Markov gives:

$$\mathbb{P}\left(X \geq \frac{2}{p}\right) \leq \frac{\mathbb{E}[X]}{\frac{2}{p}} = \frac{1}{p} \cdot \frac{p}{2} = \frac{1}{2}.$$

For large p, Chebyshev is better.

Chebyshev's Inequality

Let X be a random variable. For any $t > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\text{Var}(X)}{t^2}$$

Better Example

Suppose the average number of ads you see on a website is 25. And the variance of the number of ads is 16. Give an upper bound on the probability of seeing a website with 30 or more ads.

Fill out the poll everywhere so Kushal knows how long to explain
Go to pollev.com/cse312su21

Chebyshev's Inequality

Let X be a random variable. For any $t > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\text{Var}(X)}{t^2}$$

Better Example

Suppose the average number of ads you see on a website is 25. And the variance of the number of ads is 16. Give an upper bound on the probability of seeing a website with 30 or more ads.

$$\mathbb{P}(X \geq 30) = \mathbb{P}(X - 25 \geq 5) \leq \mathbb{P}(|X - 25| \geq 5) \leq \frac{16}{25}$$

Near the mean

Suppose you run a poll of 1000 people where in the true population 60% of the population supports you. What is the probability that the poll is not within 10-percentage-points of the true value?

Chebyshev's Inequality

Let X be a random variable. For any $t > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\text{Var}(X)}{t^2}$$

Near the mean

Suppose you run a poll of 1000 people where in the true population 60% of the population supports you. What is the probability that the poll is not within 10-percentage-points of the true value?

$$\bar{X} = \frac{\sum X_i}{1000}$$

$$\mathbb{E}[\bar{X}] = 1000 \cdot \frac{0.6}{1000} = \frac{3}{5}$$

$$\text{Var}(\bar{X}) = 1000 \cdot \frac{0.6 \cdot 0.4}{1000^2} = \frac{3}{12500}$$

Chebyshev's Inequality

Let X be a random variable. For any $t > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\text{Var}(X)}{t^2}$$

Near the mean

Suppose you run a poll of 1000 people where in the true population 60% of the population supports you. What is the probability that the poll is not within 10-percentage-points of the true value?

$$\bar{X} = \frac{\sum X_i}{1000}$$

$$\mathbb{E}[\bar{X}] = 1000 \cdot \frac{0.6}{1000} = \frac{3}{5}$$

$$\text{Var}(\bar{X}) = 1000 \cdot \frac{0.6 \cdot 0.4}{1000^2} = \frac{3}{12500}$$

$$\mathbb{P}(|\bar{X} - \mathbb{E}[\bar{X}]| \geq 0.1) \leq \frac{3/12500}{0.1^2} = 0.024$$

Chebyshev's Inequality

Let X be a random variable. For any $t > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\text{Var}(X)}{t^2}$$

Chebyshev's – Repeated Experiments

How many coin flips (each head with probability p) are needed until you get n heads?

Let X be the number necessary. What is probability $X \geq \frac{2n}{p}$?

Markov

Chebyshev

Chebyshev's – Repeated Experiments

How many coin flips (each head with probability p) are needed until you get n heads?

Let X be the number necessary. What is probability $X \geq \frac{2n}{p}$?

Markov $\quad \mathbb{P}\left(X \geq \frac{2n}{p}\right) \leq \frac{n/p}{2n/p} = \frac{1}{2}$

Chebyshev $\quad \mathbb{P}\left(X \geq \frac{2n}{p}\right) \leq \mathbb{P}\left(\left|X - \frac{n}{p}\right| \geq \frac{n}{p}\right) \leq \frac{\text{Var}(X)}{n^2/p^2} = \frac{n(1-p)/p^2}{n^2/p^2} = \frac{1-p}{n}$

Takeaway

Chebyshev gets more powerful as the variance shrinks.

Repeated experiments are a great way to cause that to happen.