Joint Distributions

CSE 312 Spring 24
Lecture 19

Announcements

CC from Monday’s lecture (CC18) and today’s lecture (CC19) both due on Friday.

HW5 due tonight, HW6 out tonight (back on that normal schedule)

HW6 has a programming question (and a computation/theoretical problem related to it).

The setting is something we haven’t discussed in class, so you might find the textbook sections helpful.

Outline of CLT steps

1. Write event you are interested in, in terms of sum of random variables.

2. Apply continuity correction if RVs are discrete.

For every real number (values produced by $\mathcal{N}$), find the nearest value in the support of original random variable (what would it round to?)

Rephrase event to include all real numbers that round to target values.

3. Standardize RV to have mean 0 and standard deviation 1.

4. Replace RV with $\mathcal{N}(0,1)$.

5. Write event in terms of Φ

6. Look up in table.

Polling

Suppose you know that 60% of CSE students support you in your run for SAC. If you draw a sample of 30 students, what is the probability that you don't get a majority of their votes.

How are you sampling?

Method 1: Get a uniformly random subset of size 30.

Method 2: Independently draw 30 people with replacement.

Which do we use?

Polling

Method 1 is what's accurate to what is actually done…

…but we're going to use the math from Method 2.

Why?
Hypergometric variable formulas are rough, and for increasing population size they're very close to binomial.

And we're going to approximate with the CLT anyway, so…the added inaccuracy isn't a dealbreaker.

If we need other calculations, independence will make any of them easier.

Polling

Let X_i be the indicator for "person i in the sample supports you."

$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{30}$ is the fraction who support you.

We're interested in the event $\mathbb{P}(\bar{X} \leq .5)$.

What is $\mathbb{E}[\bar{X}]$? What is $\text{Var}(\bar{X})$?

$\mathbb{E}[\bar{X}] = \frac{1}{30}\mathbb{E}[\sum X_i] = \frac{.6 \cdot 30}{30} = \frac{3}{5}$.

$\text{Var}(\bar{X}) = \frac{1}{30^2}\text{Var}(\sum X_i) = \frac{1}{30} \cdot .6 \cdot .4 = \frac{1}{125}$.

Using the CLT

$\mathbb{E}[\bar{X}] = \frac{1}{30}\mathbb{E}[\sum X_i] = \frac{.6 \cdot 30}{30} = \frac{3}{5}.$

$\text{Var}(\bar{X}) = \frac{1}{30^2}\text{Var}(\sum X_i) = \frac{1}{30} \cdot .6 \cdot .4 = \frac{1}{125}.$

$\mathbb{P}(\bar{X} \leq .5)$

Using the CLT

$\mathbb{E}[\bar{X}] = \frac{1}{30}\mathbb{E}[\sum X_i] = \frac{.6\cdot 30}{30} = \frac{3}{5}$.
$\text{Var}(\bar{X}) = \frac{1}{30^2}\text{Var}(\sum X_i) = \frac{1}{30}\cdot .6\cdot .4 = \frac{1}{125}$.

$\mathbb{P}(\bar{X} \leq .5)$

$= \mathbb{P}\left(\frac{\bar{X}-.6}{1/\sqrt{125}} \leq \frac{.5-.6}{1/\sqrt{125}}\right)$

$\approx \mathbb{P}\left(Y \leq \frac{.5-.6}{1/\sqrt{125}}\right)$ where $Y \sim \mathcal{N}(0,1)$

$\approx \mathbb{P}(Y \leq -1.12)$

$= \Phi(-1.12) = 1 - \Phi(1.12) \approx 1 - 0.86864 = 0.13136$

Confidence Intervals

A “confidence interval” tells you the probability (how confident you should be) that your random variable fell in a certain range (interval)

Usually “close to its expected value”

$$\mathbb{P}(|X - \mu| > \varepsilon) \leq \delta$$

If your RV has expectation equal to the value you’re searching for (like our polling example) you get a probability of being “close enough” to the target value.

Confidence Intervals

Using the CLT, we estimated the probability of “missing low”

There’s a few drawbacks though

1. Using the CLT we get an estimate, not a guarantee---what if the CLT estimate is underestimating the probability of failure?

2. We needed to know the true value to do that computation---if we knew the true value, we wouldn’t run the poll!

Some algebra tricks can handle problem 2, but 1 really asks for a new tool; we’ll see concentration inequalities next week.

Multiple Random Variables

This lecture and next lecture

Somewhat out-of-place content.

When we introduced multiple random variables, we've always had them be independent.

Because it's hard to deal with non-independent random variables.

Today is a crash-course in the toolkit for when you have multiple random variables and they aren't independent.

Going to focus on discrete RVs, we'll talk about continuous at the end.

Joint PMF, support

For two (discrete) random variables X, Y their joint pmf

$p_{X,Y}(x, y) = \mathbb{P}(X = x \cap Y = y)$

When X, Y are independent then $p_{X,Y}(x, y) = p_X(x)p_Y(y)$.

Examples

Roll a blue die and a red die. Each die is 4-sided. Let X be the blue die's result and Y be the red die's result.
Each die (individually) is fair. But not all results are equally likely when looking at them both together.

$p_{X,Y}(1,2) = 3/16.$

$p_{X,Y}$	X=1	X=2	X=3	X=4
Y=1	1/16	1/16	1/16	1/16
Y=2	3/16	0	0	1/16
Y=3	0	2/16	0	2/16
Y=4	0	1/16	3/16	0

Marginals

What if I just want to talk about X?

Well, use the law of total probability:

$\mathbb{P}(X = k) = \sum_{\text{partition } \{E_i\}} \mathbb{P}(X = k | E_i)\mathbb{P}(E_i)$

and use E_i to be possible outcomes for Y For the dice example

$\mathbb{P}(X = k) = \sum_{\ell=1}^{4} \mathbb{P}(X = k \,| Y = \ell)\mathbb{P}(Y = \ell)$

$= \sum_{\ell=1}^{4} \mathbb{P}(X = k \cap Y = \ell)$

$p_X(k) = \sum_{\ell=1}^{4} p_{X,Y}(k, \ell)$

$p_X(k)$ is called the "marginal" distribution for X (we "marginalized out" Y) it's the same pmf we've always used; the name emphasizes we have gotten rid of one of the variables.

Marginals

$$p_X(k) = \sum_{\ell=1}^{4} p_{X,Y}(k,\ell)$$

So

$$p_X(2) = \frac{1}{16} + 0 + \frac{2}{16} + \frac{1}{16} = \frac{4}{16}$$

$p_{X,Y}$	X=1	X=2	X=3	X=4
Y=1	1/16	1/16	1/16	1/16
Y=2	3/16	0	0	1/16
Y=3	0	2/16	0	2/16
Y=4	0	1/16	3/16	0

Different dice

Roll two fair dice independently.
Let U be the minimum of the two rolls and V be the maximum

Are U and V independent?

Write the joint distribution in the table

What's $p_U(z)$? (the marginal for U)

$p_{U,V}$	$U=1$	$U=2$	$U=3$	$U=4$
$V=1$		0		
$V=2$		$\frac{1}{16}$		
$V=3$		$\frac{2}{16}$		
$V=4$		$\frac{2}{16}$		

4-sided

U: min
V: max

fill in U=2 col.
give $p_U(2)$

4, 3 U=3 V=4

3, 4 U=3 V=4

2, 2 U=2 V=2

Different dice

Roll two fair dice independently. Let U be the minimum of the two rolls and V be the maximum

$$p_U(z) = \begin{cases} \frac{7}{16} & \text{if } z = 1 \\ \frac{5}{16} & \text{if } z = 2 \\ \frac{3}{16} & \text{if } z = 3 \\ \frac{1}{16} & \text{if } z = 4 \\ 0 & \text{otherwise} \end{cases}$$

$p_{U,V}$	U=1	U=2	U=3	U=4
V=1	1/16	0	0	0
V=2	2/16	1/16	0	0
V=3	2/16	2/16	1/16	0
V=4	2/16	2/16	2/16	1/16

Joint Expectation

$X^2 + 3Y$

Expectations of joint functions

For a function $g(X, Y)$, the expectation can be written in terms of the joint pmf.

$$\mathbb{E}[g(X,Y)] = \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} g(x,y) \cdot p_{X,Y}(x,y)$$

This definition hopefully isn't surprising at this point (it's the value of g times the probability g takes on that value), but it's good to see.

Conditional Expectation

Waaaaaay back when, we said conditioning on an event creates a new probability space, with all the laws holding.

So we can define things like "conditional expectations" which is the expectation of a random variable in that new probability space.

$$\mathbb{E}[X|E] = \sum_{x \in \Omega} x \cdot \mathbb{P}(X = x|E)$$

$$\mathbb{E}[X|Y = y] = \sum_{x \in \Omega_X} x \cdot \mathbb{P}(X = x|Y = y)$$

Conditional Expectations

All your favorite theorems are still true.

For example, linearity of expectation still holds

$$\mathbb{E}[(aX + bY + c)\,|E] = a\mathbb{E}[X|E] + b\mathbb{E}[Y|E] + c$$

Law of Total Expectation

$\mathbb{P}(B) = \sum \mathbb{P}(B|A_i)\mathbb{P}(A_i)$

Let $A_1, A_2, \dots, A_k$ be a partition of the sample space, then

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X|A_i]\mathbb{P}(A_i)$$

Let X, Y be discrete random variables, then

$$\mathbb{E}[X] = \sum_{y \in \Omega_Y} \mathbb{E}[X|Y = y]\mathbb{P}(Y = y)$$

Similar in form to law of total probability, and the proof goes that way as well.

LTE

You will flip 2 (independent, fair coins). Call the number of heads Y. Then (independently of the coin flips) draw a ~~geometric~~ random variable X from the distribution $\text{Exp}(Y+1)$.

exponential

What is $\mathbb{E}[X]$?

$\mathbb{E}[X \mid Y=0] = 1$

$\mathbb{E}[X \mid Y=1] = \frac{1}{2}$

$\mathbb{E}[X \mid Y=2] = \frac{1}{3}$

avg # of events

$(X \mid Y=0) \sim \text{Exp}(1)$

$(X \mid Y=1) \sim \text{Exp}(2)$ $\lambda \to \frac{1}{\lambda}$

LTE

You will flip 2 (independent, fair coins). Call the number of heads Y. Then (independently of the coin flips) draw a ~~geometric~~ exponential random variable X from the distribution $\text{Exp}(Y+1)$.

What is $\mathbb{E}[X]$?

$$\mathbb{E}[X]$$
$$= \mathbb{E}[X|Y=0]\mathbb{P}(Y=0) + \mathbb{E}[X|Y=1]\mathbb{P}(Y=1) + \mathbb{E}[X|Y=2]\mathbb{P}(Y=2)$$
$$= \mathbb{E}[X|Y=0]\cdot\frac{1}{4} + \mathbb{E}[X|Y=1]\cdot\frac{1}{2} + \mathbb{E}[X|Y=2]\cdot\frac{1}{4}$$
$$= \frac{1}{0+1}\cdot\frac{1}{4} + \frac{1}{1+1}\cdot\frac{1}{2} + \frac{1}{2+1}\cdot\frac{1}{4} = \frac{7}{12}.$$

Conditional PMFs

When we have a multi-step process, we sometimes want a pmf that will give us conditional probabilities

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

$$\mathbb{P}(X=x|Y=y) = \frac{\mathbb{P}(X=x \cap Y=y)}{\mathbb{P}(Y=y)}$$

Let Y be the number of heads in two (fair, independent) flips.

Let X be the sum of the results of Y fair, independent die-rolls

$$p_{X|Y}(12|2) = \frac{\left(\frac{1}{6}\cdot\frac{1}{6}\right)\cdot\frac{1}{4}}{\frac{1}{4}} = \frac{1}{36}$$

$$p_{X|Y}(12|1) = \frac{(0)\cdot\frac{1}{2}}{\frac{1}{2}} = 0$$

Analogues for continuous

Everything we saw today has a continuous version.

There are “no surprises”– replace pmf with pdf and sums with integrals.

	Discrete	Continuous
Joint PMF/PDF	$p_{X,Y}(x,y) = P(X=x, Y=y)$	$f_{X,Y}(x,y) \neq P(X=x, Y=y)$
Joint CDF	$F_{X,Y}(x,y) = \sum_{t \le x} \sum_{s \le y} p_{X,Y}(t,s)$	$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)\,ds\,dt$
Normalization	$\sum_x \sum_y p_{X,Y}(x,y) = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$
Marginal PMF/PDF	$p_X(x) = \sum_y p_{X,Y}(x,y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$
Expectation	$E[g(X,Y)] = \sum_x \sum_y g(x,y) p_{X,Y}(x,y)$	$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy$
Conditional PMF/PDF	$p_{X\mid Y}(x \mid y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$	$f_{X\mid Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$
Conditional Expectation	$E[X \mid Y=y] = \sum_x x p_{X\mid Y}(x \mid y)$	$E[X \mid Y=y] = \int_{-\infty}^{\infty} x f_{X\mid Y}(x \mid y)\,dx$
Independence	$\forall x,y, p_{X,Y}(x,y) = p_X(x) p_Y(y)$	$\forall x,y, f_{X,Y}(x,y) = f_X(x) f_Y(y)$

Covariance

We sometimes want to measure how "intertwined" X and Y are – how much knowing about one of them will affect the other.

If X turns out "big" how likely is it that Y will be "big" how much do they "vary together"

Covariance

$$\mathbf{Cov(X, Y)} = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Covariance

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$$

That's consistent with our previous knowledge for independent variables. (for X, Y independent, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$).

You and your friend are playing a game, you flip a coin: if heads you pay your friend a dollar, if tails they pay you a dollar. Let X be your profit and Y be your friend's profit.

What is $\mathrm{Var}(X+Y)$?

Before you calculate, make a prediction. What should it be?

Covariance

$X = 1, Y = -1 \quad XY = -1$
$X = -1, Y = 1 \quad XY = -1$

You and your friend are playing a game, you flip a coin: if heads you pay your friend a dollar, if tails they pay you a dollar. Let X be your profit and Y be your friend's profit.

What is $\mathrm{Var}(X + Y)$?

$\mathrm{Var}(X) = \mathrm{Var}(Y) = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 = 1 - 0^2 = 1$

$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

$\mathbb{E}[XY] = \frac{1}{2} \cdot (-1 \cdot 1) + \frac{1}{2}(1 \cdot -1) = -1$

$\mathrm{Cov}(X, Y) = -1 - 0 \cdot 0 = -1.$

$\mathrm{Var}(X + Y) = 1 + 1 + 2 \cdot -1 = 0$

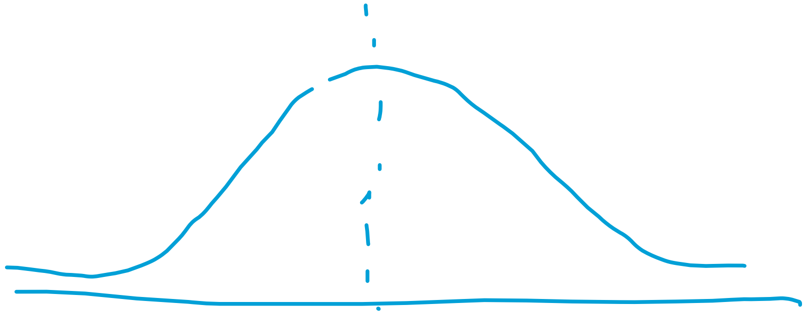
Tail Bounds

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%.

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

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

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

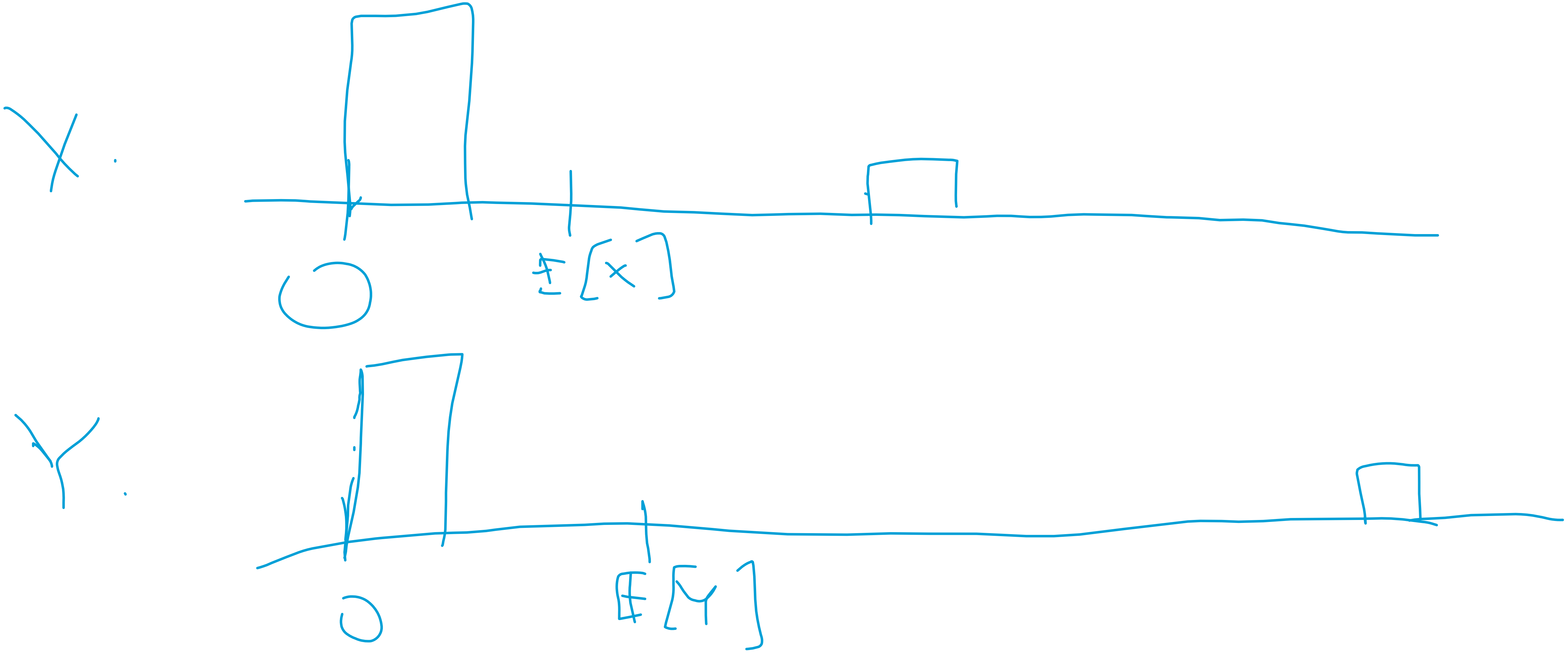
Markov's Inequality

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

$$\mathbb{P}(X \geq k\mathbb{E}[X]) \leq \frac{1}{k}$$

To apply this bound you only need to know:

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


X:
0
E[X]
Y:
0
E[Y]

Proof

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

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

$$\geq \sum_{x:x\geq t} t \cdot \mathbb{P}(X = x)$$
$$= t \cdot \sum_{x:x\geq t} \mathbb{P}(X = x)$$
$$= t \cdot \mathbb{P}(X \geq t)$$

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

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?

$$1 - \mathbb{P}(X < 12) \approx 1 - 0.865 = .135$$

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.

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.

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!