etherpad.wikimedia.org/p/312 for (anonymous) questions/comments!

Variance + Discrete Zoo I

CSE 312 24Su

Lecture 10

Fun fact!

Over a large number of trials, the average result will get closer to the expected value. This principle is why casinos make consistent profits. Even though individual results can vary widely, the "law of large numbers" ensures that the casino's profit prevails over time.

Logistics

- HW3 due tonight
- no HW released tonight! (HW4 released next Wednesday)
- Mid-quarter Feedback
- Midterm information on website
  - Similar length to practice midterms but you will have 2x for it (110 minutes)
  - 4 questions
    - *T/F/short answer, counting/probability, conditional probability, random variables*

Outline for Today

- Review linearity of expectation
- Variance (last way to describe random variables!)
- Start learning about the zoo of discrete random variables

One More Linearity of Expectation Example!

Rotating the table

n people are sitting around a circular table. There is a name tag in each place. Nobody is sitting in front of their own name tag.

Rotate the table by a random number k of positions between 1 and n-1 (equally likely)

Let X be the number of people that end up in front of their own name tag. **Find $\mathbb{E}[X]$.**

<u>Decompose:</u>

What X_i can we define that have the needed information?

<u>LOE:</u>

<u>Conquer:</u>

Rotating the table

n people are sitting around a circular table. There is a name tag in each place. Nobody is sitting in front of their own name tag.

Rotate the table by a random number k of positions between 1 and n-1 (equally likely)

X is the number of people that end up in front of their own name tag. Find $\mathbb{E}[X]$.

<u>Decompose:</u> Define X_i as follows:

$$X_i = \begin{cases} 1 & \text{if person i sits infront of their own name tag} \\ 0 & \text{otherwise} \end{cases}$$

Note: $X = \Sigma_{i=1}^{n} X_i$

<u>LOE:</u>

$$\mathbb{E}[X] = \mathbb{E}[\Sigma_{i=1}^{n} X_i] = \Sigma_{i=1}^{n} \mathbb{E}[X_i]$$

<u>Conquer:</u>

These X_i are not independent!
That's ok!!

Rotating the table

n people are sitting around a circular table. There is a name tag in each place. Nobody is sitting in front of their own name tag.

Rotate the table by a random number k of positions between 1 and n-1 (equally likely)

X is the number of people that end up in front of their own name tag. Find $\mathbb{E}[X]$.

<u>Decompose:</u> Define X_i as follows:

$$X_i = \begin{cases} 1 & \text{if person i sits infront of their own name tag} \\ 0 & \text{otherwise} \end{cases} \qquad X = \Sigma_{i=1}^n X_i$$

<u>LOE:</u>

$$\mathbb{E}[X] = \mathbb{E}[\Sigma_{i=1}^n X_i] = \Sigma_{i=1}^n \mathbb{E}[X_i]$$

<u>Conquer:</u>

$$\mathbb{E}[X_i] = P(X_i = 1) = \frac{1}{n-1} \qquad \mathbb{E}[X] = n \cdot \mathbb{E}[X_i] = \frac{n}{n-1}$$

How to know what indicators to define?

- "find the expected number of students (from n in total) who show up in class"
  -> maybe define indicator RVs $\mathbf{X_i}$ that is $\mathbf{1}$ if the i'th student shows up
  -> $\mathbf{X = \sum_{i=1}^{n} X_i = X}$
- "find the expected number of defective items in a batch of 100 items"
  -> maybe define indicator RVs $\boldsymbol{X_i}$ that is $\mathbf{1}$ if the i'th item is defective
  -> $\boldsymbol{X = \sum_{i=1}^{n} X_i}$
- "find the expected number of A that have B (B is some property)"
  -> maybe define indicator RVs $\boldsymbol{X_i}$ for the i'th possible A that is 1 if that A has B

Techniques For Finding Expectation

1. Definition of Expectation

When the support is small enough and we can find the PMF of X

$$\mathbb{E}[X] = \sum_{k \in \Omega_X} k \cdot \mathbb{P}(X = k)$$

2. Linearity of Expectation

When we can break X into a sum

Break into indicator RVs if it's some kind of count

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[\mathrm{Y}]$$

3. LOTUS (Law of the unconscious statistician)

When it's a function of X (e.g, X2)

$$\mathbb{E}[g(X)] = \sum_{k \in \Omega_X} g(k) \cdot \mathbb{P}(X = k)$$

Where are we?

A *random variable* is a way to summarize what outcome you saw.

Ω is set of possible outcomes

Outcome $\boldsymbol{\omega \in \Omega}$ → X → **Quantitative value**

Support (set of values RV can take),

Probability Mass Function & Cumulative Distribution Function **describe** probabilities

The expectation of a RV is its average value. (summarizes a RV)

> Expectation is **linear:** $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.
  - $X + Y$ is a random variable – it's a function that outputs a number given an outcome (or, here, a combination of outcomes).

Variance

Variance

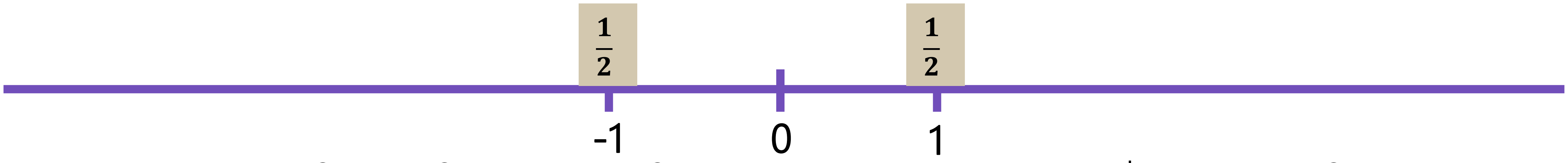
Another one number summary of a random variable.

But wait, we already have expectation, what's this for?

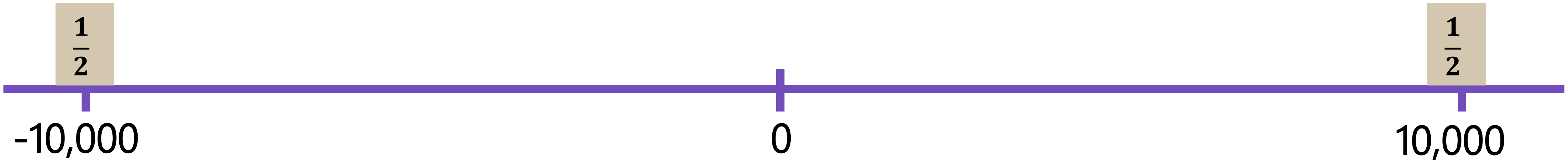
Consider these two games

Would you be willing to play these games?

Game 1: I will flip a fair coin; if it's heads, I pay you \$1. If it's tails, you pay me \$1. Let X_1 be your profit if you play game 1



Game 2: I will flip a fair coin; if it's heads, I pay you \$10,000. If it's tails, you pay me \$10,000. Let X_2 be your profit if you play game 2.



Both games are "fair" ($\mathbb{E}[X_1] = \mathbb{E}[X_2] = 0$)

What’s the difference

Expectation tells you what the average will be…

But it doesn’t tell you how “extreme” your results could be.

Nor how likely those extreme results are.

Game 2 has many (well, only) very extreme results.

In expectation they “cancel out” but if you can only play once…

…it would be nice to measure that.

Designing a Measure – Try 1

Let's measure how **far all the outcomes are away from the center**, weighted on how likely they are

$$\sum_{\omega}(\mathbb{P}(\omega) \cdot (X(\omega) - \mathbb{E}[X]))$$

What happens with Game 1?

$$\frac{1}{2} \cdot (1 - 0) + \frac{1}{2} \cdot (-1 - 0)$$

$$\frac{1}{2} + (-\frac{1}{2}) = 0$$

What happens with Game 2?

$$\frac{1}{2} \cdot (100000 - 0) + \frac{1}{2} \cdot (-100000 - 0)$$

$$5000 + (-5000) = 0$$

Designing a Measure – Try 2

How do we prevent cancelling? Squaring the distances makes everything positive.

$\sum_\omega (\mathbb{P}(\omega) \cdot (X(\omega) - \mathbb{E}[X])^{\mathbf{2}})$

This is the variance!

What happens with Game 1?

$$\frac{1}{2} \cdot (1 - 0)^{\mathbf{2}} + \frac{1}{2} \cdot (-1 - 0)^{\mathbf{2}}$$

$$\frac{1}{2} + \frac{1}{2} = 1$$

What happens with Game 2?

$$\frac{1}{2} \cdot (100000 - 0)^{\mathbf{2}} + \frac{1}{2} \cdot (-100000 - 0)^{\mathbf{2}}$$

$$5{,}000{,}000{,}000 + 5{,}000{,}000{,}000 = 10^{10}$$

Why Squaring

Why not absolute value? Or Fourth power?

- Squaring is nicer algebraically.
- Our goal with variance was to talk about the spread of results. Squaring makes extreme results even more extreme.
- Fourth power over-emphasizes the extreme results (for our purposes).

Variance

Use LOTUS to compute $\boldsymbol{E}[\boldsymbol{X}^2]$ $(g(X) = X^2)$

Variance

The variance of a random variable X is

$$\mathrm{Var}(X) = \sum_{\omega} \mathbb{P}(\omega) \cdot (X(\omega) - \mathbb{E}[X])^2 = \mathbb{E}[(X - \mathbb{E}[X])^2]$$
$$= \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

The first form forms are the definition. The last one is an algebra trick.

Intuition: Variance is a quantity that measures on average *how "far" the random variable is from its expectation*

> higher values means values are very spread out

> smaller values means values are closer and tend to be closer to expectation

Proof of Calculation Trick

$\mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2 - 2X\mathbb{E}[X] + (\mathbb{E}[X])^2]$ expanding the square

$= \mathbb{E}[X^2] - \mathbb{E}\big[2X\mathbb{E}[X]\big] + \mathbb{E}[(\mathbb{E}[X])^2]$ linearity of expectation.

$= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[(\mathbb{E}[X])^2]$ linearity of expectation.

$= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + (\mathbb{E}[X])^2$ expectation of a constant is the constant

$= \mathbb{E}[X^2] - 2(\mathbb{E}[X])^2 + (\mathbb{E}[X])^2$

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

So $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

Variance of a die

Let X be the result of rolling a fair die.

$\mathrm{Var(X)} = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[(X - 3.5)^2]$

$= \frac{1}{6}(1-3.5)^2 + \frac{1}{6}(2-3.5)^2 + \frac{1}{6}(3-3.5)^2 + \frac{1}{6}(4-3.5)^2 + \frac{1}{6}(5-3.5)^2 + \frac{1}{6}(6-3.5)^2$

$= \frac{35}{12} \approx 2.92.$

Or $\mathbb{E}[X^2] - (E[X])^2 = \sum_{k=1}^{6} \frac{1}{6} \cdot k^2 - 3.5^2 = \frac{91}{6} - 3.5^2 \approx 2.92$

Variance of n Coin Flips

Flip a coin n times, where it comes up heads with probability p each time (independently). Let X be the total number of heads.

We saw last time $\mathbb{E}[X] = np$.

$$X_i = \begin{cases} 1 & \text{if flip } i \text{ is heads} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}[X] = \mathbb{E}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} p = np.$$

Variance of n Coin Flips

Flip a coin n times, where it comes up heads with probability p each time (independently). Let X be the total number of heads.

What about $\mathrm{Var}(X)$

$\mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_{k \in \Omega_X} \mathbb{P}(X = k)(k - np)^2$

$= \sum_{k=0}^{n} \binom{n}{k} \cdot p^k (1-p)^{n-k} \cdot (k - np)^2$

Algebra time?

Variance Adds **<u>If Independent</u>**

If X and Y are <u>independent</u> then
$$\mathbf{Var}(X+Y) = \mathbf{Var}(X) + \mathbf{Var}(Y)$$

Are the X_i independent? Yes!

In this problem X_i is independent of X_j for $i \neq j$ where

$$X_i = \begin{cases} 1 & \text{if flip } i \text{ was heads} \\ 0 & \text{otherwise} \end{cases}$$

Variance Adds **<u>If Independent</u>**

$\mathrm{Var}(X) = \mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$

What's the $\mathrm{Var}(X_i)$?

$$
\begin{aligned}
Var(X_i) &= \mathbb{E}\left[X_i^2\right] - \mathbb{E}[X_i]^2 \\
&= \mathbb{E}[X_i] - p^2 \\
&= p - p^2 \\
&= p(1 - p)
\end{aligned}
$$

Variance Adds **If Independent**

$\mathrm{Var}(X) = \mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$

What's the $\mathrm{Var}(X_i)$?

$p(1-p)$.

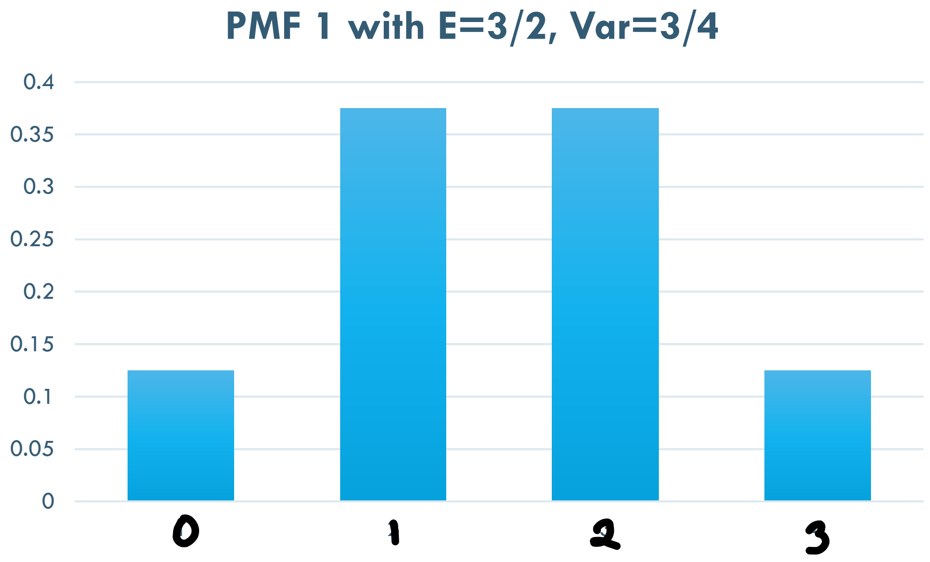
$\mathrm{Var}(X) = \sum_{i=1}^{n} p(1-p) = np(1-p)$.

Expectation and Variance aren't everything

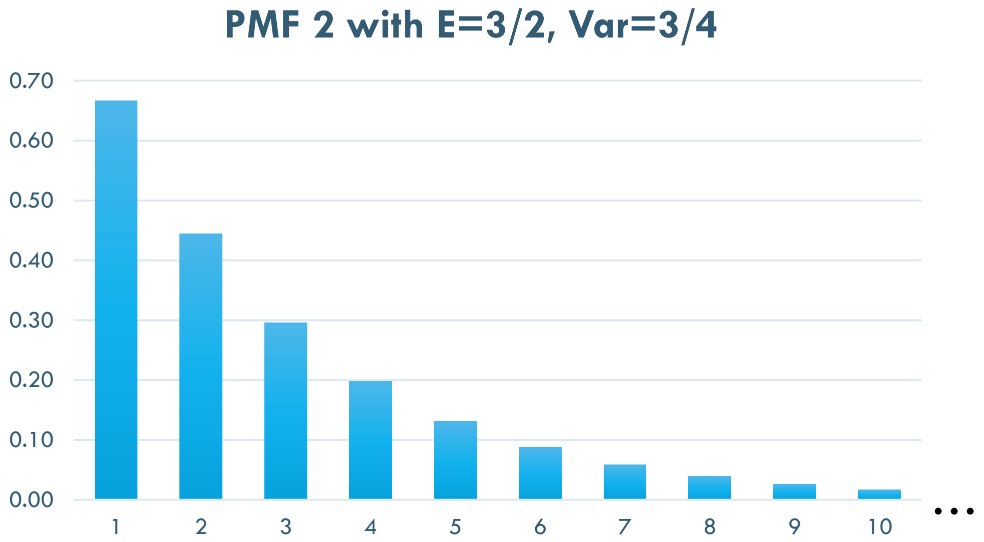
Alright, so expectation and variance is everything right?

No!

Flip a fair coin 3 times indep. Count heads.



Flip a biased coin (prob heads=2/3) until heads. Count flips.



A PMF or CDF *does* fully describe a random variable.

Useful Facts

Make a prediction

How should $\mathrm{Var}(X + c)$ relate to $Var(X)$ if c is a constant?

How should $\mathrm{Var}(\mathrm{aX})$ relate to $\mathrm{Var}(X)$ is a is a constant?

Facts About Variance

$$\mathrm{Var}(X + c) = \mathrm{Var}(X)$$

Proof:

$$\mathrm{Var}(X + c) = \mathbb{E}[(X + c)^2] - \mathbb{E}[X + c]^2$$

$$= \mathbb{E}[X^2] + \mathbb{E}[2Xc] + \mathbb{E}[c^2] - (\mathbb{E}[X] + c)^2$$

$$= \mathbb{E}[X^2] + 2c\mathbb{E}[X] + c^2 - \mathbb{E}[X]^2 - 2c\mathbb{E}[X] - c^2$$

$$= \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

$$= \mathrm{Var}(X)$$

***Intuition**: Adding a constant just shifts the distribution – the spread says the same*

Facts about Variance

$\mathrm{Var}(aX) =$

Facts about Variance

$$\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$$
$$= \mathbb{E}[(aX)^2] - (\mathbb{E}[aX])^2$$
$$= a^2\mathbb{E}[X^2] - (a\mathbb{E}[X])^2$$
$$= a^2\mathbb{E}[X^2] - a^2\mathbb{E}[X]^2$$
$$= a^2(\mathbb{E}[X^2] - \mathbb{E}[X]^2)$$

***Intuition**: Multiplying by a positive constant makes it more spread out*

Summary of *Expectation* and *Variance*

Expectation ("on average")

$$\mathbb{E}[X] = \sum_{k \in \Omega_X} k \cdot \mathbb{P}(X = k)$$

> $\mathbb{E}[g(X)] = \sum_{k \in \Omega_X} g(k) \cdot \mathbb{P}(X = k)$

Linearity of Expectation (LoE)

$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ *(always true)*

Shifting and Scaling:

$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$

If X and Y are independent:

$\mathbb{E}[\boldsymbol{X} \cdot \boldsymbol{Y}] = \mathbb{E}[\boldsymbol{X}]\mathbb{E}[\boldsymbol{Y}]$

Variance ("how spread out")

$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

Linearity of Variance

only if X and Y are independent

Shifting and Scaling:

$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$

If X and Y are independent:

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

Discrete Random Variable Zoo

Discrete Random Variable Zoo

There are common **patterns** of experiments:

> Flip a [fair/unfair] coin [blah] times and count the number of heads.

> Flip a [fair/unfair] coin until the first time that you see a heads

> Draw a uniformly random element from [set]

> Define an indicator random variable for [event]

…

Instead of calculating the support, PMF, CDF, expectation, variance every time, why not calculate it **once** and look it up every time?

for example…

E.g.

- the number of coin tosses till we get the first heads
- the number of games we need to play till we first win
- the number of dice rolls till the first 6

all follow the same format as,

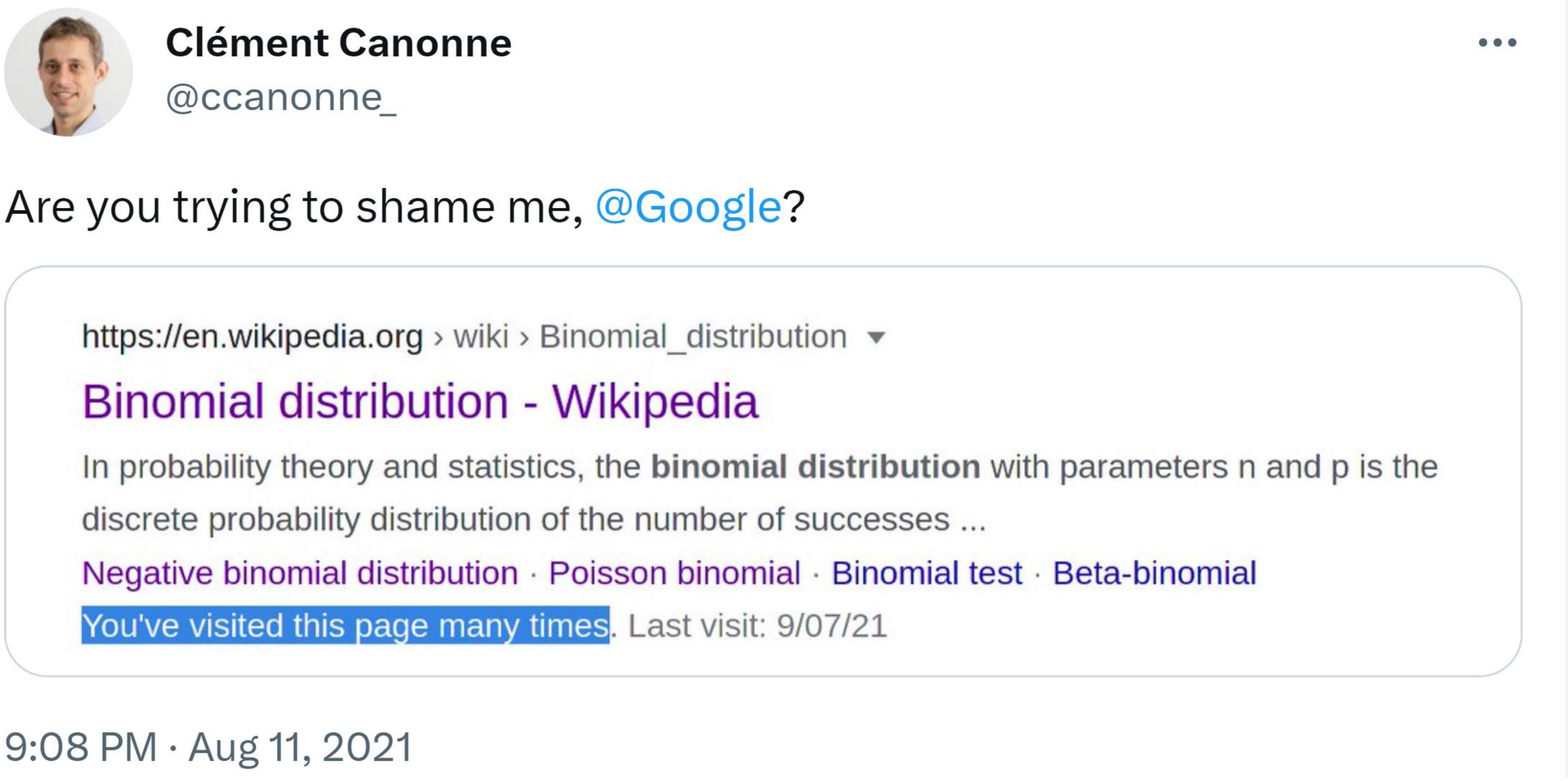
- the number of trials till we get the first success

The discrete zoo defines some common patterns and gives us the PMF, expectation, and variance, so we don't have to compute it every time!

What’s our goal?

Your goal is NOT to memorize these facts (it’ll be convenient to memorize some of them, but don’t waste time making flash cards).

Everything is on Wikipedia anyway. Everyone checks Wikipedia when they forget these.



What’s our goal?

Your goal is NOT to memorize these facts (it’ll be convenient to memorize some of them, but don’t waste time making flash cards).

Everything is on Wikipedia anyway. Everyone checks Wikipedia when they forget these.

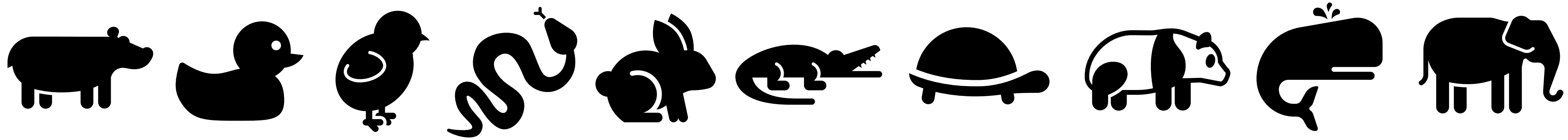
While learning about this zoo, we will also get the chance too:

0. Introduce one new distribution we haven’t seen at all (next time).

1. Practice expectation, variance, etc. for ones we have gotten hints of.

2. Review the first half of the course with some probability calculations.

Zoo!



$X \sim \text{Unif}(a,b)$	$X \sim \text{Ber}(p)$	$X \sim \text{Bin}(n,p)$	$X \sim \text{Geo}(p)$
$p_X(k) = \frac{1}{b-a+1}$ $\mathbb{E}[X] = \frac{a+b}{2}$ $\text{Var}(X) = \frac{(b-a)(b-a+2)}{12}$	$p_X(0) = 1-p;$ $p_X(1) = p$ $\mathbb{E}[X] = p$ $\text{Var}(X) = p(1-p)$	$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$ $\mathbb{E}[X] = np$ $\text{Var}(X) = np(1-p)$	$p_X(k) = (1-p)^{k-1}p$ $\mathbb{E}[X] = \frac{1}{p}$ $\text{Var}(X) = \frac{1-p}{p^2}$

$X \sim \text{NegBin}(r,p)$	$X \sim \text{HypGeo}(N,K,n)$	$X \sim \text{Poi}(\lambda)$
$p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$ $\mathbb{E}[X] = \frac{r}{p}$ $\text{Var}(X) = \frac{r(1-p)}{p^2}$	$p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$ $\mathbb{E}[X] = n\frac{K}{N}$ $\text{Var}(X) = \frac{K(N-K)(N-n)}{N^2(N-1)}$	$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$ $\mathbb{E}[X] = \lambda$ $\text{Var}(X) = \lambda$

Scenario: Uniform

You want an integer in some range, with each integer equally likely.

Familiar example:

Outcome of rolling a fair die (or draw a random integer from 1,…,n)

equally likely to be an integer between 1 and 6

Discrete Uniform Distribution

$X \sim \mathrm{Unif}(a, b)$

X is a uniformly random integer between a and b (inclusive)

Parameter a is the minimum value in the support, b is the maximum value in the support.

PMF: $p_X(k) = \frac{1}{b-a+1}$ for $k \in \mathbb{Z}$, $a \leq k \leq b$

CDF: $F_X(k) = \frac{k-a+1}{b-a+1}$ for $k \in \mathbb{Z}$, $a \leq k \leq b$.

Expectation: $\mathbb{E}[X] = \frac{a+b}{2}$

Variance: $\mathrm{Var}(X) = \frac{(b-a)(b-a+2)}{12}$

Situation: Bernoulli

Do I get a success in <u>one</u> trial with <u>probability $\boldsymbol{p}$ of "success"</u>?

Familiar examples:

You flip a biased coin (once) and want to record whether its heads.

An indicator random variable that is either 1 or 0

Bernoulli Distribution

$X \sim \mathrm{Ber}(p)$

X is the <u>indicator random variable that the trial was a success</u>.

Parameter p is probability of success on the trial.

PMF: $p_X(0) = 1 - p,\, p_X(1) = p$

CDF: $F_X(k) = \begin{cases} 0 & \text{if } k < 0 \\ 1 - p & \text{if } 0 \leq k < 1 \\ 1 & \text{if } k \geq 1 \end{cases}$

Expectation: $\mathbb{E}[X] = p$

Variance: $\mathrm{Var}(X) = p(1 - p)$

Bernoulli Distribution Examples

Did a particular bit get written correctly on the device?

Did you guess right on a multiple-choice test?

Did a server in a cluster fail?

Did a disk drive crash?

Did buying a particular share of a stock pay off?

Did a user like or dislike a YouTube video? (Back in ye olden days, at least)

Does a user click an ad?

In each of these, we find the probability of success on a trial, define a RV as an indicator for whether the event occurred, say the RV follows a bernoulli distribution with relevant and we have the PMF, E[X] and Var(X)!

Situation: Binomial

How many success did you see in $\boldsymbol{n}$ *independent* trials, where each trial has probability $\boldsymbol{p}$ of success?

Familiar example:

You flip a coin n times independently, each with a probability p of coming up heads. How many heads are there?

Binomial Distribution

$X \sim \mathrm{Bin}(n, p)$

X is the number of successes across n independent trials.

n is the number of independent trials.
p is the probability of success for one trial.

PMF: $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k \in \{0,1,\ldots,n\}$

CDF: F_X is ugly.

Expectation: $\mathbb{E}[X] = np$

Variance: $\mathrm{Var}(X) = np(1-p)$

Binomial Distribution *Examples*

How many bits were written correctly on the device?

How many questions did you guess right on a multiple-choice test?

How many servers in a cluster failed?

How many keys went to one bucket in a hash table?

Example: Unpopular Donuts

A donut shop serves 50 people a day and serves a *mango chili lime* donut. The probability that a customer chooses this donut is 0.2. All customers' choices are independent of each other.

What is the probability that exactly 10 *people choose this flavor?*

What is the probability that at least 3 *people choose this flavor?*

Example: Unpopular Donuts

A donut shop serves 50 people a day and serves a *mango chili lime* donut. The probability that a customer chooses this donut is 0.2. All customers' choices are independent of each other.

What is the probability that <u>exactly</u> 10 *people choose this flavor?*

PMF: $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k \in \{0,1,\ldots,n\}$

What is the probability that <u>at least</u> 3 *people choose this flavor?*

Example: Unpopular Donuts

A donut shop serves 50 people a day and serves a *mango chili lime* donut. The probability that a customer chooses this donut is 0.2. All customers' choices are independent of each other.

What is the probability that exactly 10 *people choose this flavor?*

$X\sim$ number of people who choose this flavor. $X\sim\mathrm{Bin}(50, 0.2)$

$$\mathbb{P}(X = 10) = \binom{50}{10}0.2^{10}(1-0.2)^{50-10}$$

What is the probability that at least 3 *people choose this flavor?*

$$\mathbb{P}(X \geq 3) = 1 - \mathbb{P}(X < 3) = 1 - (\mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2))$$
$$= 1 - (\binom{50}{0}0.2^{0}(0.8)^{50-0} + \binom{50}{1}0.2^{1}(0.8)^{50-1} + \binom{50}{2}0.2^{2}(0.8)^{50-2})$$

Situation: Geometric

How many *independent* trials are needed until the first success?

Familiar Example:

You flip a coin (which comes up heads with probability p) until you get a heads. How many flips did you need?

Geometric Distribution

$X \sim \text{Geo}(p)$

X is the number of trials needed to see the first success.

p is the probability of success for one trial.

PMF: $p_X(k) = (1-p)^{k-1}p$ for $k \in \{1,2,3,\ldots\}$

$F_X(k) = 1-(1-p)^k$ for $k \in \mathbb{N}$

$\mathbb{E}[X] = \frac{1}{\mathrm{p}}$

$\text{Var}(X) = \frac{1-p}{p^2}$

Geometric Distribution *Examples*

How many bits can we write *before one is incorrect*?

How many questions do you have to answer *until you get one right*?

How many times can you run an experiment *until it fails for the first time*?

Geometric: Analysis

Both the expectation and variance are new to us.

The derivations of both are uninformative

- Every derivation I've ever seen has wild algebra tricks.

Geometric: Expectation

$\mathbb{E}[X] = \sum_{k=1}^{\infty} k(1-p)^{k-1}p$

$= p\sum_{k=1}^{\infty} k(1-p)^{k-1} = p \cdot \frac{1}{p^2} = \frac{1}{p}.$

Intuition: Smaller p means longer wait

$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

$= \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$

Intuition: for small p lots of variance (might have to wait a long time, and it's variable)

For large p very little variance (for $p = 1$ there's no variation at all!)

Geometric Property

Geometric random variables are called “memoryless”

Suppose you’re flipping coins (independently) until you see a heads.

The first three came up tails.

How many flips are ***left*** until you see the first heads?

It’s another independent copy of the original!

The coin “forgot” it already came up tails 3 times.

Formally…

Let X be the number of flips needed, Y be the flips after the third.

$\mathbb{P}(Y = k | X \geq 3) =?$

…

Which is $p_X(k)$.

Formally…

Let X be the number of flips needed, Y be the flips after the third.

$\mathbb{P}(Y = k | X \geq 3) = \mathbb{P}(Y = k \cap X \geq 3)/\mathbb{P}(X \geq 3)$

$\frac{(1-p)^{k+3-1}p}{(1-p)^3}$

$= (1-p)^{k-1}p$

Which is $p_X(k)$.

Summary

Today we covered the following random variables from the zoo:

- **Bernoulli:** Whether there is success in one trial
  $\mathrm{Ber}(p)$ **is 1 with probability $\boldsymbol{p}$ and 0 otherwise**
- **Binomial:** Number of successes in n independent trials
  $\mathrm{Bin}(n, p)$ - n independent trials, probability p of success on each trial
- **Geometric:** Number of trials until the first success
  $\mathrm{Geo}(p)$ – probability p of success on a single trial
- **Uniform:** Every integer between a and b are equally likely
  $\mathrm{Uni}f(a, b)$

More on Friday ☺