Discrete RV Zoo

CSE 312 Winter 25
Lecture 12

Where are we?

A random variable is a numerical summary of the outcome of an experiment.

$\mathbb{E}[X]$ is the weighted average of possibilities of X.

For <u>all</u> rv's $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

Indicator rv's are a great trick to simplify expectation computations.

$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$ measures how "spread out" a rv is.

For <u>independent</u> rv's:

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

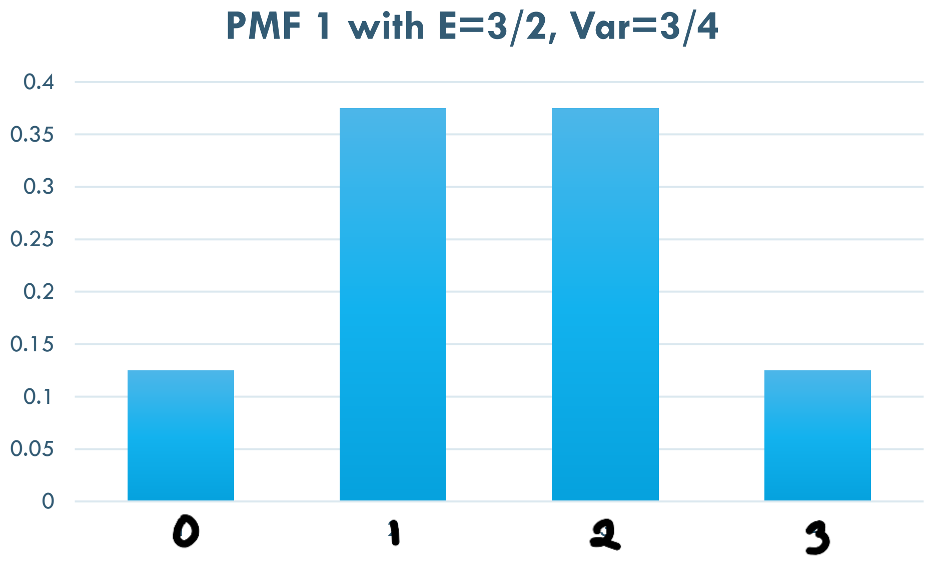
$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

Expectation and Variance aren't everything

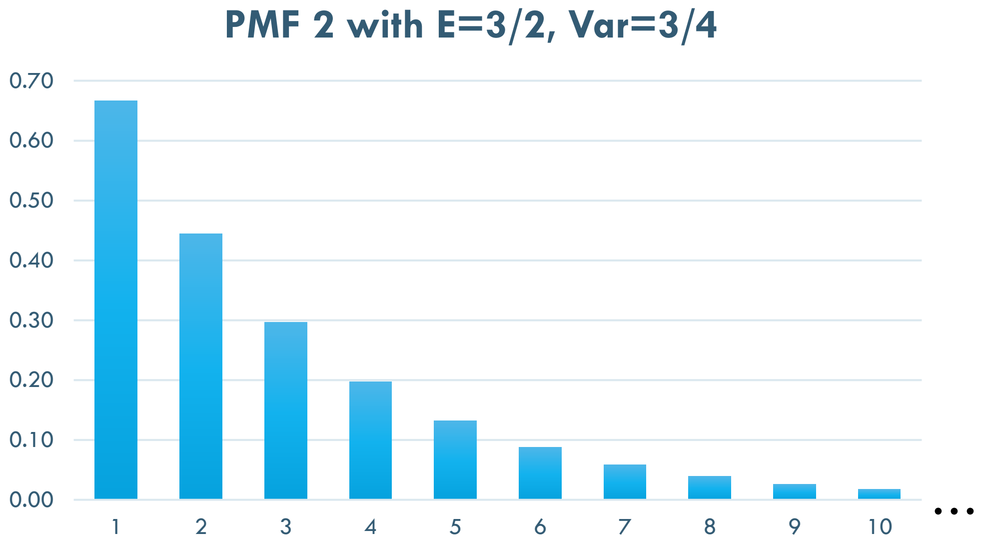
Alright, so expectation and variance is everything right?

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.

Shifting a random variable

For any random variable $\mathbf{X}$, and any constants $\mathbf{a}, \mathbf{b}$:

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

For any random variable $\mathbf{X}$, and any constants $\mathbf{a}, \mathbf{b}$:

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

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]

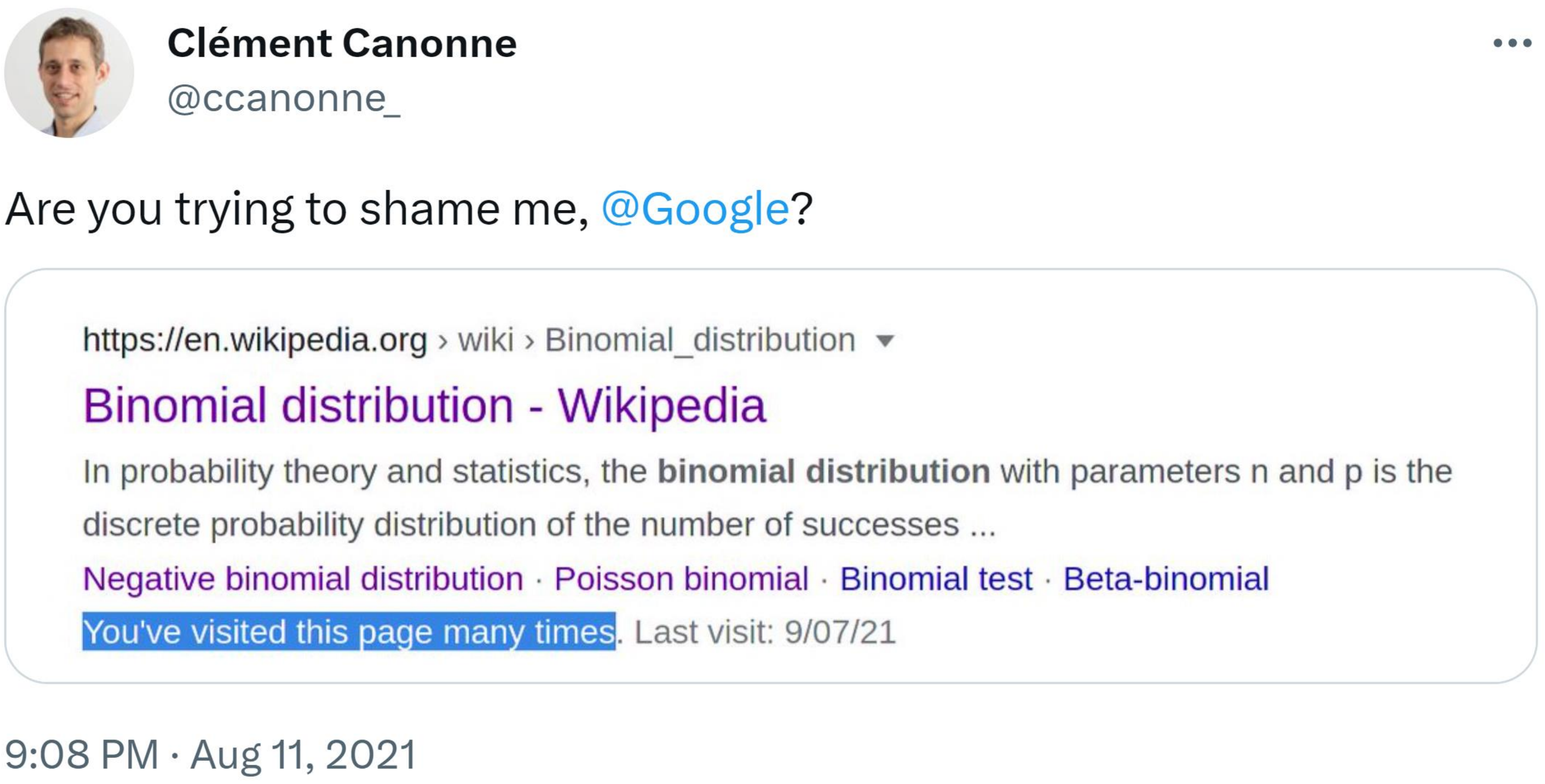
…

Instead of calculating the pmf, cdf, support, expectation, variance,… every time, why not calculate it **once** and look it up 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.

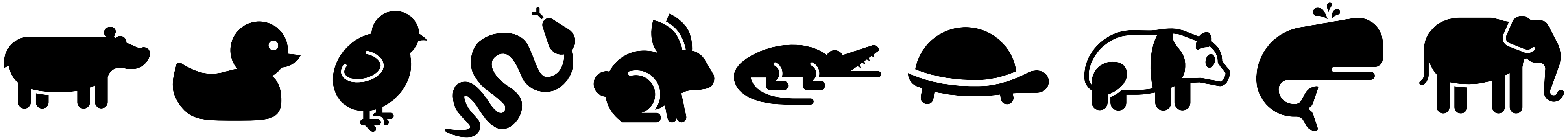
Our goals are:

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!



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

$$p_X(k) = \frac{1}{b-a+1}$$
$$\mathbb{E}[X] = \frac{a+b}{2}$$
$$\text{Var}(X) = \frac{(b-a)(b-a+2)}{12}$$

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

$$p_X(0) = 1-p; \quad p_X(1) = p$$
$$\mathbb{E}[X] = p$$
$$\text{Var}(X) = p(1-p)$$

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

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$
$$\mathbb{E}[X] = np$$
$$\text{Var}(X) = np(1-p)$$

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

$$p_X(k) = (1-p)^{k-1} p$$
$$\mathbb{E}[X] = \frac{1}{p}$$
$$\text{Var}(X) = \frac{1-p}{p^2}$$

$X \sim \text{NegBin}(r,p)$

$$p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$$
$$\mathbb{E}[X] = \frac{r}{p}$$
$$\text{Var}(X) = \frac{r(1-p)}{p^2}$$

$X \sim \text{HypGeo}(N,K,n)$

$$p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$
$$\mathbb{E}[X] = n\frac{K}{N}$$
$$\text{Var}(X) = \frac{K(N-K)(N-n)}{N^2(N-1)}$$

$X \sim \text{Poi}(\lambda)$

$$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$
$$\mathbb{E}[X] = \lambda$$
$$\text{Var}(X) = \lambda$$

Some More Familiar Variables

Situation: Bernoulli

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

You define an indicator random variable, and want to know whether it's 1 or not.

More generally: you have one trial, and some probability p of "success."

Bernoulli Distribution

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

Parameter p is probability of success.

X is the indicator random variable that the trial was a success.

$p_X(0) = 1 - p, p_X(1) = p$

$$F_X(k) = \begin{cases} 0 & \text{if } k < 0 \\ 1 - p & \text{if } 0 \le k < 1 \\ 1 & \text{if } k \ge 1 \end{cases}$$

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

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

Some other uses:
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?

Situation: Binomial

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

More generally: How many success did you see in n independent trials, where each trial has probability p of success?

Binomial Distribution

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

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

X is the number of successes across the n trials.

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

F_X is ugly.

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

$\text{Var}(X) = np(1-p)$

Some other uses:
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?

Situation: Geometric

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

More generally: how many independent trials are needed until the first success?

Geometric Distribution

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

p is the probability of success for one trial.

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

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

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

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

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

Some other uses:
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 two 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 2 times.

Formally…

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

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

…

Which is $p_X(k)$.

Formally…

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

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

$\frac{(1-p)^{k+2-1}p}{(1-p)^2}$

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

Which is $p_X(k)$.

So, the (conditional) PMF for Y matches that of X. The coin "forgot" it did the first two flips.

Scenario: Uniform

You Roll a Fair Die (or draw a random integer from 1,…,n).

More generally: you want an integer in some range, with each equally likely.

Discrete Uniform Distribution

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

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

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

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

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

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

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

Some Less Familiar Distributions

The Poisson Distribution

A new kind of random variable.

We use a Poisson distribution when:

We're trying to count the number of times something happens in some interval of time.

We know the average number that happen (i.e. the expectation)

Each occurrence is independent of the others.

There are a VERY large number of "potential sources" for those events, few of which happen.

The Poisson Distribution

Classic applications:

- How many traffic accidents occur in Seattle in a day
- How many major earthquakes occur in a year (not including aftershocks)
- How many customers visit a bakery in an hour.

Why not just use counting coin flips?

What are the flips…the number of cars? Every person who might visit the bakery? There are way too many of these to count exactly or think about dependency between. But a Poisson might accurately model what's happening.

It’s a model

By modeling choice, we mean that we’re choosing math that we think represents the real world as best as possible

Is every traffic accident really independent?

Not *really*, one causes congestion, which causes angrier drivers. Or both might be caused by bad weather/more cars on the road.

But we assume they are (because the dependence is so weak that the model is useful).

Poisson Distribution

$X \sim \text{Poi}(\lambda)$

Let λ be the average number of incidents in a time interval.

X is the number of incidents seen in a particular interval.

Support $\mathbb{N}$

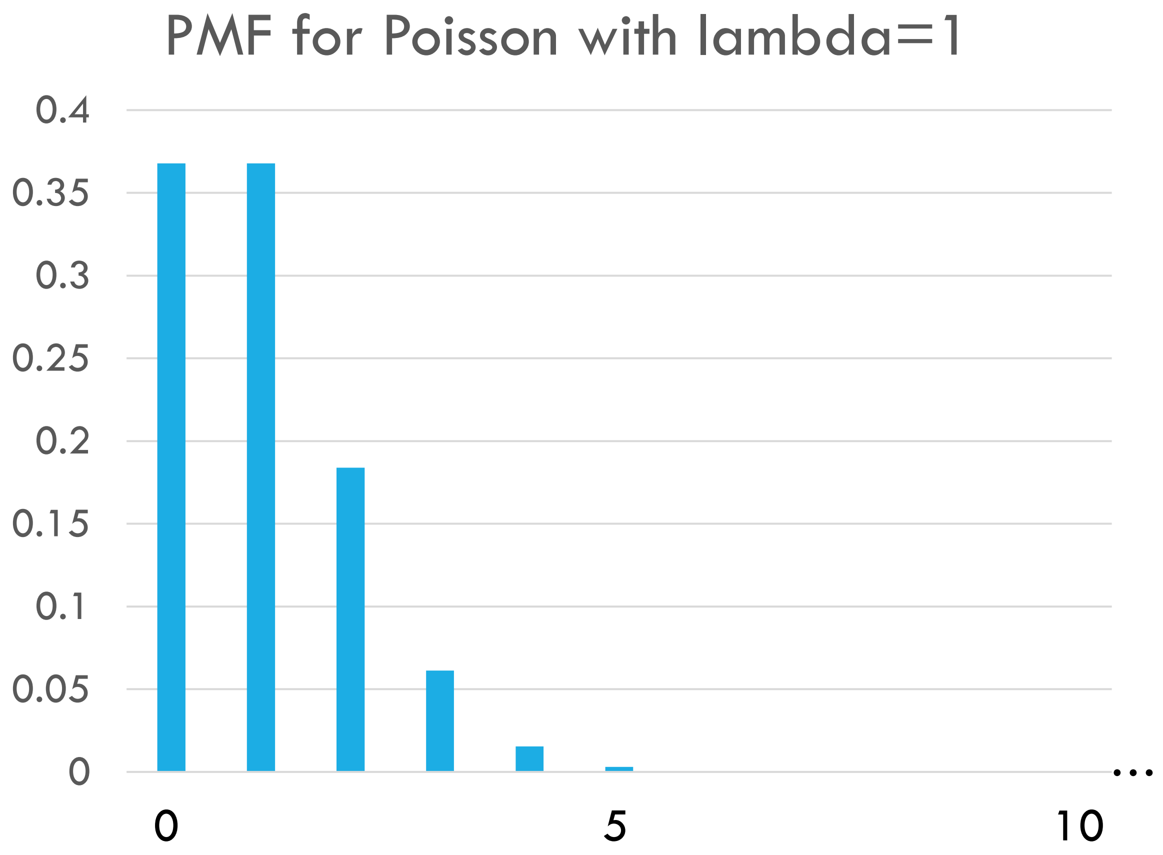
$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$ (for $k \in \mathbb{N}$)

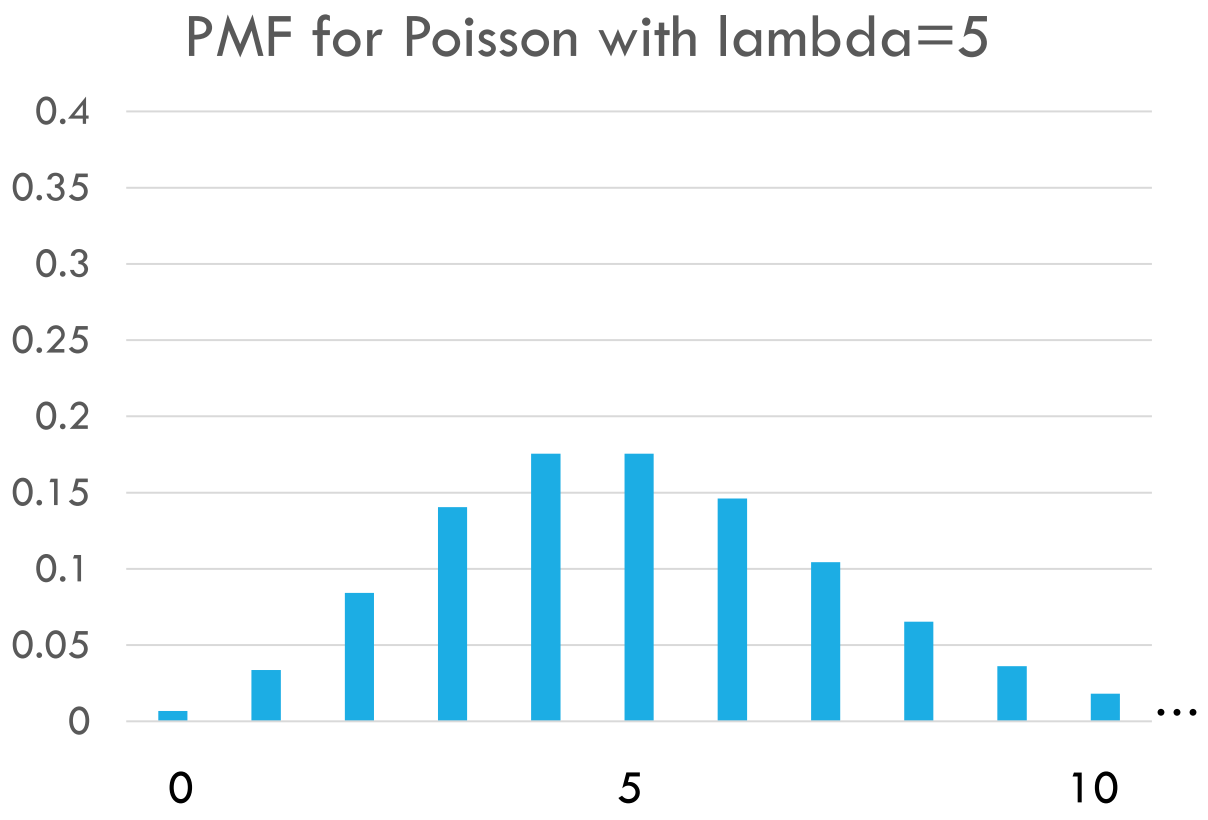
$F_X(k) = e^{-\lambda} \sum_{i=0}^{\lfloor k \rfloor} \frac{\lambda^i}{i!}$

$\mathbb{E}[X] = \lambda$

$\text{Var}(X) = \lambda$

Some Sample PMFs





Let’s take a closer look at that pmf

$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$ (for $k \in \mathbb{N}$)

If this is a real PMF, it should sum to 1.
Let’s check that to understand the PMF a little better.

$\sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!}$

$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$

Taylor Series for e^x

$$\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x$$

$= e^{-\lambda} e^{\lambda} = e^0 = 1$

Let's check something…the expectation

$\mathbb{E}[X] = \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \frac{\lambda^k}{k!}$

$= \sum_{k=1}^{\infty} k \cdot e^{-\lambda} \frac{\lambda^k}{k!}$ first term is 0.

$= \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^k}{(k-1)!}$ cancel the k.

$= \lambda \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^{k-1}}{(k-1)!}$ factor out λ.

$= \lambda \sum_{j=0}^{\infty} e^{-\lambda} \frac{\lambda^j}{(j)!}$ Define $j = k - 1$

$= \lambda \cdot 1$ The summation is just the pmf!

Where did this expression come from?

For the cars we said "it's like every car in Seattle independently might cause an accident."

If we knew the exact number of cars, and they all had identical probabilities of causing an accident…

It'd be just like counting the number of heads in n flips of a bunch of coins (the coins are just VERY biased).

The Poisson is a certain limit as $n \to \infty$ but np (the expected number of accidents) stays constant.

Scenario: Negative Binomial

You're playing a carnival game, and there are r little kids nearby who all want a stuffed animal. You can win a single game (and thus win one stuffed animal) with probability p (independently each time) How many times will you need to play the game before every kid gets their toy?

More generally, run independent trials with probability p. How many trials do you need for r successes?

Try it

More generally, run independent trials with probability p. How many trials do you need for r successes?

What's the pmf?

What's the expectation and variance (hint: linearity)

Negative Binomial Analysis

What's the pmf? Well how would we know $X = k$?

Of the first $k-1$ trials, $r-1$ must be successes.
And trial k must be a success.

That first part is a lot like a binomial!

It's the $p_Y(r-1)$ where $Y \sim \text{Bin}(k-1, p)$

First part gives $\binom{k-1}{r-1}(1-p)^{k-1-(r-1)}p^{r-1} = \binom{k-1}{r-1}(1-p)^{k-r}p^{r-1}$

Second part, multiply by p

Total: $p_X(k) = \binom{k-1}{r-1}(1-p)^{k-r}p^r$

Negative Binomial Analysis

What about the expectation?

To see r successes:

We flip until we see success 1.

Then flip until success $\mathbf{2.}$

… Flip until success r.

The total number of flips is…the sum of geometric random variables!

Negative Binomial Analysis

Let $Z_1, Z_2, \ldots, Z_r$ be independent copies of $\text{Geo}(p)$

Z_i are called "independent and identically distributed" or "i.i.d.'

Because they are independent…and have identical pmfs.

$X \sim \text{NegBin}(r, p)$ $X = Z_1 + Z_2 + \cdots + Z_r$.

$\mathbb{E}[X] = \mathbb{E}[Z_1 + Z_2 + \cdots Z_r] = \mathbb{E}[Z_1] + \mathbb{E}[Z_2] + \cdots + \mathbb{E}[Z_r] = r \cdot \frac{1}{p}$

Negative Binomial Analysis

Let $Z_1, Z_2, \dots, Z_r$ be independent copies of $\mathtt{Geo}(p)$

$X \sim \text{NegBin}(r, p)$ $X = Z_1 + Z_2 + \cdots + Z_r$.

$\text{Var}(X) = \text{Var}(Z_1 + Z_2 + \cdots + Z_r)$

Up until now we've just used the observation that $X = Z_1 + \cdots + Z_r$.

$= \text{Var}(Z_1) + \text{Var}(Z_2) + \cdots + \text{Var}(Z_r)$ because the Z_i are independent.

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

Negative Binomial

$X \sim \mathrm{NegBin(r, p)}$

Parameters: r: the number of successes needed, p the probability of success in a single trial

X is the number of trials needed to get the r^{th} success.

$p_X(k) = \binom{k-1}{r-1}(1-p)^{k-r}p^r$

$F_X(k)$ is ugly, don't bother with it.

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

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

Scenario: Hypergeometric

You have an urn with N balls, of which K are purple. You are going to draw balls out of the urn **without** replacement.

If you draw out n balls, what is the probability you see k purple ones?

Hypergeometric: Analysis

You have an urn with N balls, of which K are purple. You are going to draw balls out of the urn **without** replacement.

If you draw out n balls, what is the probability you see k purple ones?

Of the K purple, we draw out k choose which k will be drawn

Of the $N - K$ other balls, we will draw out $n - k$, choose which $N - K - (n - k)$ will be removed.

Sample space all subsets of size n

$$\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Hypergeometric: Analysis

$X = D_1 + D_2 + \cdots + D_n$

Where D_i is the indicator that draw i is purple.

D_1 is 1 with probability K/N.

What about D_2?

$$\mathbb{P}(D_2 = 1) = \frac{K-1}{N-1} \cdot \frac{K}{N} + \frac{K}{N-1} \cdot \frac{K-N}{N} = \frac{K(K-N+K-1)}{N(N-1)} = \frac{K}{N}$$

In general $\mathbb{P}(D_i = 1) = \frac{K}{N}$

It might feel counterintuitive, but it's true!

Hypergeometric: Analysis

$\mathbb{E}[X]$

$= \mathbb{E}[D_1 + \cdots D_n] = \mathbb{E}[D_1] + \cdots + \mathbb{E}[D_n] = n \cdot \frac{K}{N}$

Can we do the same for variance?

No! The D_i are dependent. Even if they have the same probability.

Hypergeometric Random Variable

$X \sim \text{HypGeo}(N, K, n)$

Parameters: A total of N balls in an urn, of which K are successes. Draw n balls without replacement.

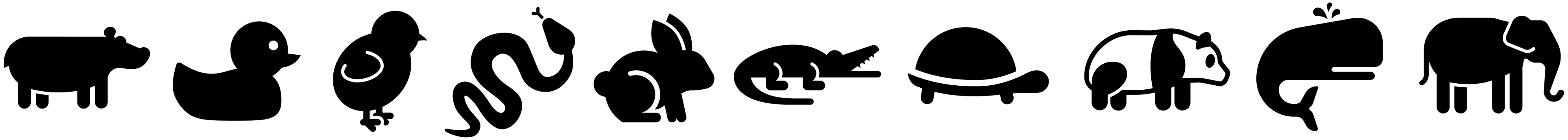
X is the number of success balls drawn.

$$p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

$$\mathbb{E}[X] = \frac{nK}{N}$$

$$\text{Var}(X) = n \cdot \frac{K}{N} \cdot \frac{N-K}{N} \cdot \frac{N-n}{N-1}$$

Zoo!



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

$$p_X(k) = \frac{1}{b - a + 1}$$

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

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

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

$$p_X(0) = 1 - p; \quad p_X(1) = p$$

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

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

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

$$p_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

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

$$\text{Var}(X) = np(1 - p)$$

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

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

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

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

$X \sim \text{NegBin}(r, p)$

$$p_X(k) = \binom{k - 1}{r - 1} p^r (1 - p)^{k-r}$$

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

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

$X \sim \text{HypGeo}(N, K, n)$

$$p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

$$\mathbb{E}[X] = n\frac{K}{N}$$

$$\text{Var}(X) = \frac{K(N - K)(N - n)}{N^2(N - 1)}$$

$X \sim \text{Poi}(\lambda)$

$$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$$\mathbb{E}[X] = \lambda$$

$$\text{Var}(X) = \lambda$$

Zoo Takeaways

You can do relatively complicated counting/probability calculations much more quickly than you could week 1!

You can now explain why your problem is a zoo variable and save explanation on homework (and save yourself calculations in the future).

Don't spend extra effort memorizing…but be careful when looking up Wikipedia articles.

The exact definitions of the parameters can differ (is a geometric random variable the number of failures before the first success, or the total number of trials including the success?)

First Half Review

What have we done over the past 5 weeks?

Counting

Combinations, permutations, indistinguishable elements, starts and bars, inclusion-exclusion…

Probability foundations

Events, sample space, axioms of probability, expectation, variance

Conditional probability

Conditioning, independence, Bayes' Rule

Refined our intuition

Especially around Bayes' Rule

What’s next?

Continuous random variables.

- So far our sample spaces have been countable. What happens if we want to choose a random real number?
- How do expectation, variance, conditioning, etc. change in this new context?
- Mostly analogous to discrete cases, but with integrals instead of sums.

Analysis when it’s inconvenient (or impossible) to exactly calculate probabilities.

- Central Limit Theorem (approximating discrete distributions with continuous ones)
- Tail Bounds/Concentration (arguing it’s unlikely that a random variable is far from its expectation)

A first taste of making predictions from data (i.e., a bit of ML)

Practice Problem: Coin Flips

There are two coins, heads up, on a table in front of you. One is a trick coin – both sides are heads. The other is a fair coin.

You are allowed 2 coin flips (total between the two coins) to figure out which coin is which. What is your strategy? What is the probability of success?

Flip each once

With probability 1 when we flip the trick coin it shows heads.

With probability ½ the fair coin shows tails, and we know it's the fair one.

With probability 1/2, both the coins were heads and we have learned nothing. So we have a ½ chance of guessing which is which.

$\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4}$ chance of success

Flip one twice.

Now flip the same coin twice.

We'll see a tails with probability $\frac{1}{2} \cdot \frac{3}{4} = \frac{3}{8}$

If we don't see a tails, just guess the other one? What's our probability of guessing right? Let T be the event "we're flipping the trick coin" N be the event we saw no tails

$$\mathbb{P}(T|N) = \frac{\mathbb{P}(N|T)\mathbb{P}(T)}{\mathbb{P}(N)} = \frac{1 \cdot \frac{1}{2}}{1 \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{1}{2}} = \frac{4}{5}$$

Guess right: $1 \cdot \frac{3}{8} + \frac{4}{5} \cdot \frac{5}{8} = \frac{7}{8}$ Better to flip the same coin twice!

Practice Problem: Donuts

You are buying at most 7 donuts (could be 0, could be 1,…, could be 7).

There are chocolate, strawberry, and vanilla donuts.

How many different orders could you make – give a simple formula!

Donuts: Approach 1

Use the sum rule over the possible numbers of donuts.

For n donuts, by the stars and bars formula there are $\binom{n+3-1}{3-1}$

So we have $\sum_{n=0}^{7}\binom{n+3-1}{3-1}$ correct. But not simple yet…

Use pascal's rule. Rewrite $\binom{2}{2}$ as $\binom{3}{3}$

We'll get $\binom{j}{2} + \binom{j}{3} = \binom{j+1}{3}$, that can combine with $\binom{j+1}{2}$ until you get

$\binom{7+3-1+1}{3} = \binom{10}{3}$

Donuts: Approach 2

Clever way: a fourth type of donut: the don't-buy-one donut.

Then we're buying exactly seven donuts of the four types (chocolate, strawberry, vanilla, don't-buy-one)

By stars and bars $\binom{7+4-1}{4-1} = \binom{10}{3}$.

Practice Problem: Poisson

Seattle averages 3 days with snowfall per year.

Suppose that the number of days with snow follows a Poisson distribution. What is the probability of getting exactly 5 days of snow?

According to the Poisson model, what is the probability of getting 367 days of snow?

Practice: Poisson

Let $X \sim \text{Poi}(3)$.

$f_X(5) = \frac{3^5 e^{-3}}{5!} \approx .1008$

Or about once a decade.

Probability of 367 snowy days, err…

The distribution says

$f_X(367) \approx 1.8\text{x}10^{-610}$.

Definition of a "year" says probability should be $\mathbf{0}$.