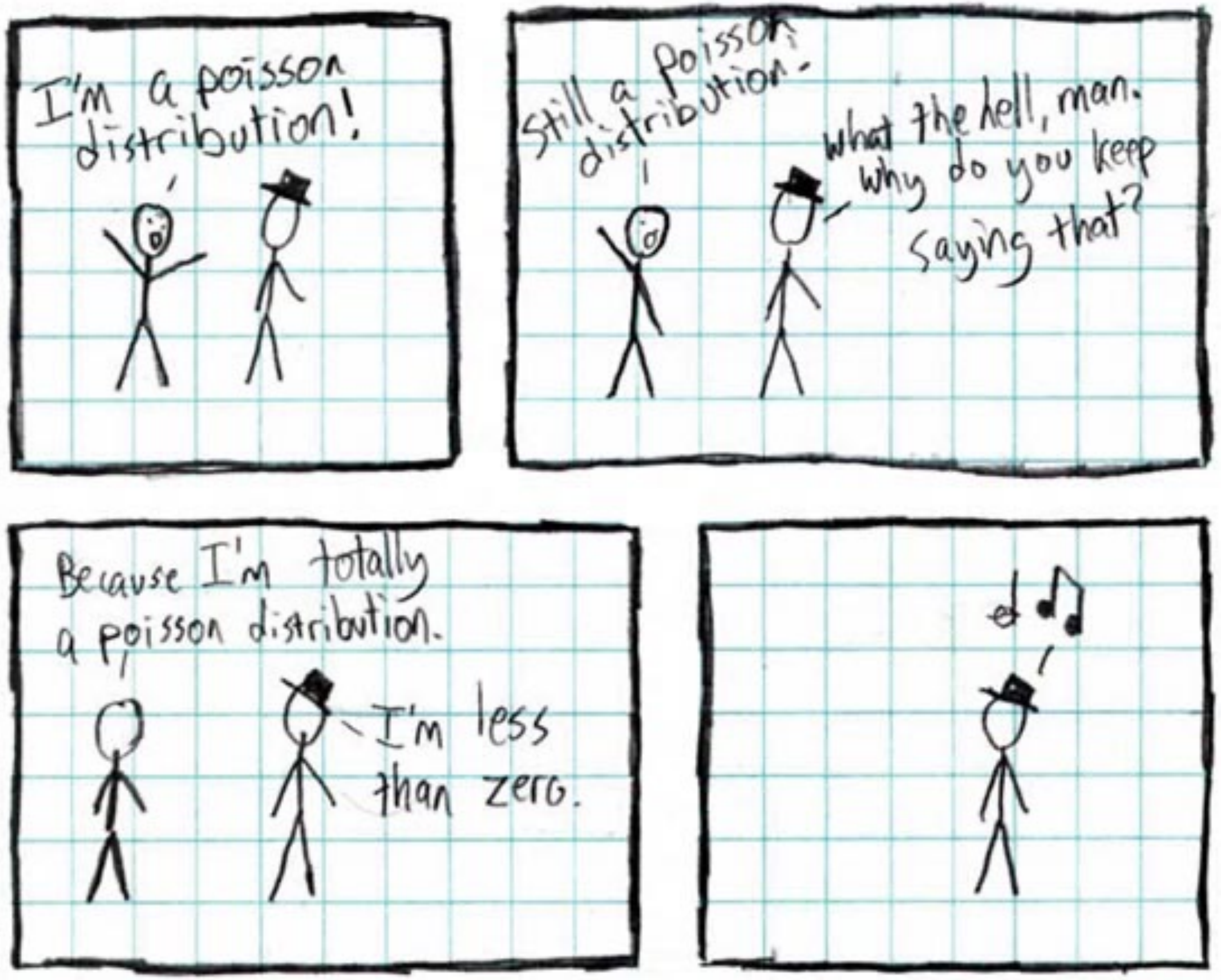
Mid-quarter feedback form in today's CC :)




# Discrete RV Zoo: Part II

CSE 312 Su23
Lecture 11

# Announcements

Midterm on Monday – don’t forget to bring your Husky Card and read all the details on Ed.

TA review session after class today

# Roadmap

1 trial 0/1

Last class

- Bernoulli
- Binomial
- Geometric
- Uniform

n idp Bernoulli # of succes

n idp Bernoulli # of trials until 1st success

equally likely

Today (some unfamiliar distributions)

- Poisson
- Negative Binomial
- Hypergeometric

# 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

Each occurrence is independent of the others.

# 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.

There are a large number of potential sources of these events.

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).

Bulletin of the
Seismological Society of America

Vol. 64 October 1974 No. 5

# IS THE SEQUENCE OF EARTHQUAKES IN SOUTHERN CALIFORNIA, WITH AFTERSHOCKS REMOVED, POISSONIAN?

BY J. K. GARDNER and L. KNOPOFF

ABSTRACT

**Yes.**

# Poisson Distribution

lambda $\mathbb{E}[X]$

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

hour

Let $\lambda$ be the average number of incidents in a time interval.

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

Range/Support: $\mathbb{N}$ → $\{0, \ldots, \infty\}$

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

Euler's numbers

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

$\mathbb{E}[X] = \lambda$ ⟷

$\text{Var}(X) = \lambda$ ✓

# Where did that PMF come from? 🤯

X: # of clicks on an ad in a week, ~~$X \sim Poi(3)$~~ $\lambda = 3 = \mathbb{E}[X]$

What is $\mathbb{P}(X = 5)$? $\rightarrow$

One way to approach this using our existing toolkit is to try using the binomial distribution! We could split a week into 7 sub-intervals (for each day).

$$X = \sum_{i=1}^{7} X_i \qquad X_i = \begin{cases} 1 & \text{if click day } i \\ 0 & \text{o.w.} \end{cases}$$

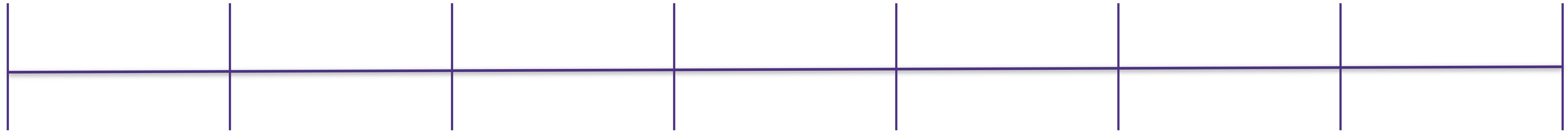

$$\mathbb{E}[X] = n \cdot p = 3 \qquad 7 \cdot p = 3 \Rightarrow p \approx \frac{3}{7}$$

# Where did that PMF come from? 🤯

X: # of clicks on an ad in a week, on average $\lambda = 3$ clicks per week.

We know $\mathbb{E}[X] = 3 = np$, so p should be 3/n.

Specifically: we know have 7 independent intervals. We *assume* 0 or 1 clicks per interval. So, $X = \sum_{i=1}^{7} X_i$ where $X_i$ is 1 if ad is clicked in the interval. Now we can approximate $\mathbb{P}(X = 5)$ with the binomial PMF…



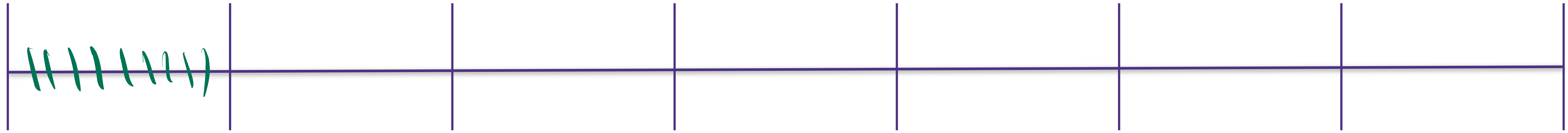
# Where did that PMF come from? 🤯

What's a weakness of this model?

We're assuming only 0 or 1 clicks on an ad per day. We could have more though!

What if we extend that assumption to having 0/1 people click an ad per hour…

$$X = \sum_{i=1}^{24 \cdot 7} X_i$$



# Where did that PMF come from? 🤯

What's a weakness of this model?

We're assuming only 0 or 1 people can click an ad per day. We could have more though!

What if we extend that assumption to having 0/1 people click an ad per hour…

Same limitation!

What we really want is: $n \to \infty$, to give us infinitesimally small intervals!

# Where did that PMF come from? 🤯

$\lambda = np = \mathbb{E}[X]$ $n \to \infty$

What we really want is: $n \to \infty$, to give us infinitesimally small intervals!

$\mathbb{E}[X] = \boxed{\lambda = np}$ $p \downarrow$

We also need to fix $np$ to a constant, allowing us to set $\boxed{p = \frac{\lambda}{n}}$

avg (constant)

$n \uparrow$
$p \downarrow$

And this is where the Poisson PMF comes from:

$\mathbb{P}(X = k) =$

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

$$p_X(k) = \lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \text{ (binomial PMF)}$$

$$= e^{-\lambda} \cdot \frac{\lambda^k}{k!} \text{ (skipping \textit{a lot} of math)}$$

# Where did this expression come from?

For the cars we said: “it’s like every car in Seattle independently might cause an accident.”

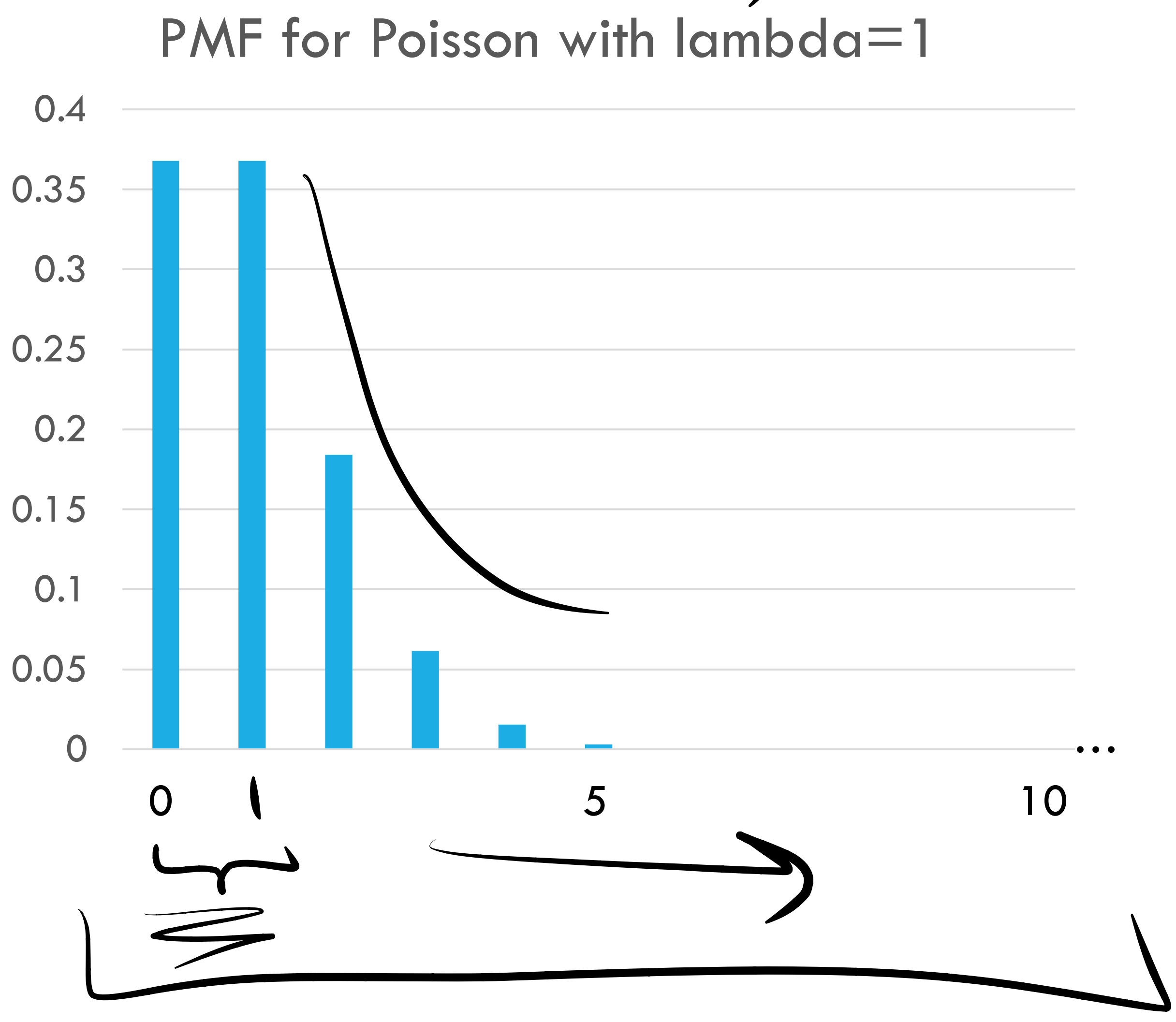
If we knew the exact number of cars, and they all had identical probabilities of causing an accident…

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

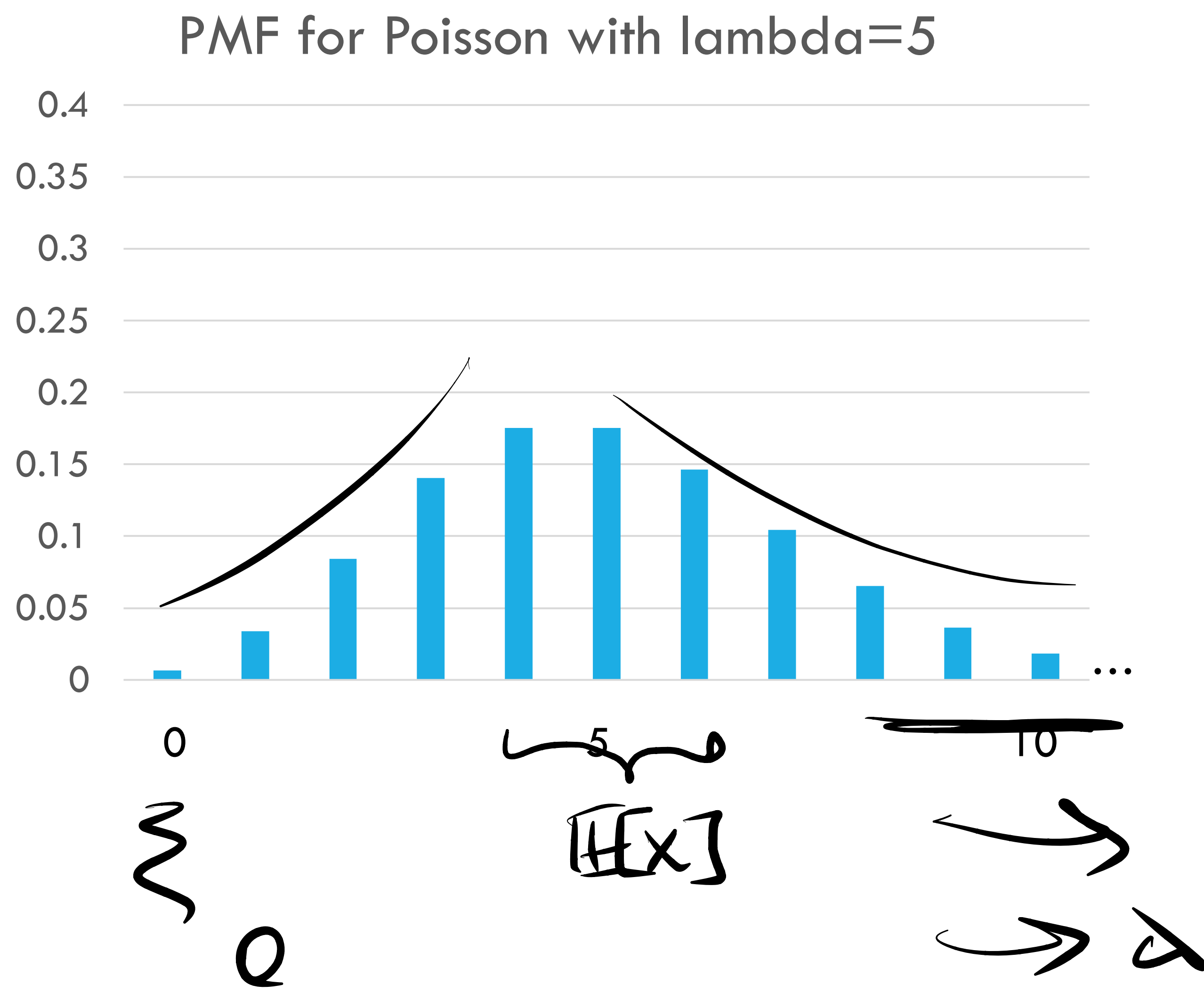
The Poisson is a certain limit as $n \to \infty$ but $np$ (the expected number of accidents) stays constant.

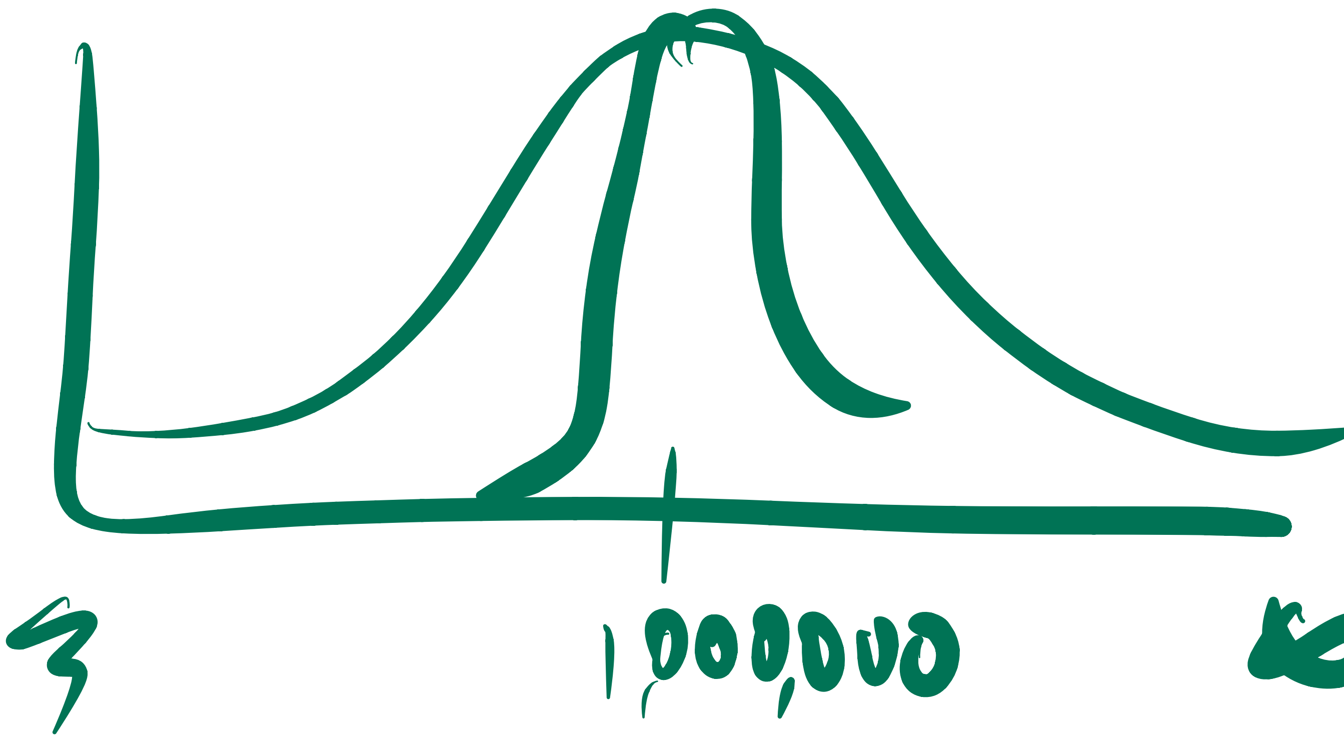
# Some Sample PMFs

X ~ (1)




X ~ Poi(5)




3
1,000,000
∞

# Let's take a closer look at that PMF

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

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

normalization constant

$x = \lambda$

TS

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

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

$$\boxed{\begin{array}{c}\textit{Taylor Series for } e^x \\ \sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x = e^\lambda\end{array}}$$

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

# Let's check something…the expectation

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

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

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

$= \lambda \cdot 1$ The summation is just the PMF! $p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$ (for $k \in \mathbb{N}$)

# Applications: Poisson

Instagram posts per hour

Hard drive crashes per year

Website visitors per month

Ad clicks per month

# 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

k trials in order to see rth success

$$\underbrace{1 \quad 2 \quad \cdots \quad k-1}_{r-1} \qquad \underset{\text{rth success}}{k}$$

More generally, run $n$ 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: think about using linearity of expectation)

$$\mathbb{P}(X=k) = \underbrace{\binom{k-1}{r-1} p^{r-1} (1-p)^{k-1-(r-1)}} p$$

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

# 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 $Y \sim \mathrm{Bin}(k-1, r-1)$

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

# Try it

Geo(p) 1st Success

$\longrightarrow$ 1st $\longrightarrow$ 2nd $\longrightarrow$ - - - - $\longrightarrow$ rth

More generally, run independent trials with probability $p$. How many trials do you need for $\underbrace{r \text{ successes}}$?

What's the expectation and variance?

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{r} X_i\right] = \sum_{i=1}^{r} \mathbb{E}[X_i] = \sum_{i=1}^{r} \frac{1}{p} = r \cdot \frac{1}{p}$$

$$X = \sum_{i=1}^{r} X_i$$

$X_i$ # of trials ith success

$\boxed{X_i}$ $\underbrace{\text{i.i.d}}$ $X_i \sim \text{Geo}(p)$

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

$Z_i$ 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 $\mathrm{Geo}(p)$

$\frac{1-p}{p^2}$ Var(Zi)

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

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

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

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

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

iid

$$\mathrm{Var}(X) = \mathrm{Var}\left(\sum_{i=1}^{r} Z_i\right) = \sum_{i=1}^{r} \mathrm{Var}(Z_i) = r\,\frac{1-p}{p^2}$$

# 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 $r^{\text{th}}$ 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 $N$ marbles, of which $K$ are purple. You are going to draw marbles out of the urn **without** replacement.

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

purple · non-purple

$$\mathbb{P}(X = k) = \frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}}$$

# Hypergeometric: Analysis

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

If you draw out $n$ marbles, 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 marbles, 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

$\mathbb{P}(D_1 = 1) = \frac{K}{N}$

$\mathbb{E}[X]$

$N, K$

$n$

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

Where $D_i$ is the indicator that draw $i$ is purple.

# of purple marbles

$1 \quad \ldots \quad n$

+1 – purple

+0 – non purple

$D_1$ is 1 with probability $K/N$.

What about $D_2$?

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

$\mathbb{P}(D_2 = 1) = \mathbb{P}(D_2 = 1 \mid D_1 = 1)\, \mathbb{P}(D_1 = 1)$

$+ \mathbb{P}(D_2 = 1 \mid D_1 = 0)\, \mathbb{P}(D_1 = 0)$

(LTP)

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

It might feel counterintuitive, but it's true!

# Hypergeometric: Analysis

K

$D_1 \cdots D_K$

$D_{K+1}$

$\mathbb{E}[X]$

LOE

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

Can we do the same for variance?

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

$$X = \sum_{i=1}^{n} D_i$$

$$\mathbb{E}[D_i] = \mathbb{P}(D_i = 1) = \frac{K}{N}$$

# Hypergeometric Random Variable

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

Parameters: A total of $N$ objects, of which $K$ are successes.
Draw $n$ objects without replacement.

$X$ is the number of successes drawn.

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

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

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

$$\underbrace{N \qquad K}_{n}$$

# Zoo!

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

# 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?)

# Wrap-up & preview of rest of quarter :D

# 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, zoo

Conditional probability

Conditioning, independence, Bayes’ Rule

Refined our intuition

Especially around Bayes’ Rule

# What's next?

Distribution $\longrightarrow$ Prediction.

$\uparrow$

Data

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?

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)