Discrete Random Variable Zoo II

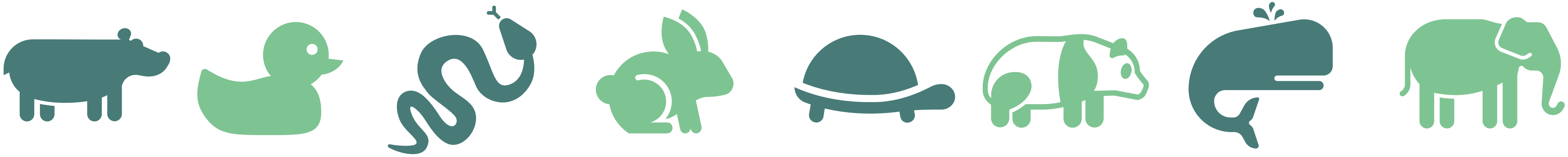
CSE 312 26Su

Lecture 11

Discrete Zoo of Random Variables

There are *common patterns* of random experiments

We're going to **identify some common patterns**, and **compute the support, PMF, CDF, expectation, and variance** for them, so *when we see a random variable that matches that pattern*, we don't have to re-compute everything!



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 $k \in \mathbb{Z}$, $a \le k \le b$

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

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

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

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

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 $k \in \{0, 1, \dots, n\}$

CDF: F_X is ugly.

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

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

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) independently 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.

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 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.

$\Omega_X = \{1,2,3,4, \dots\}$

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

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

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

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

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

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

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

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

Suppose you're flipping coins independently until you see a heads.

$X \sim \mathbf{Geo}(p)$ is number of flips till the first head

> The first three came up tails.

> Y is number of flips *left* until you see the first head

Does Y also follow $\mathbf{Geo}(p)$?

T T T ? ? ? ? …

Fill out the poll everywhere:
pollev.com/jolie312

Geometric Property - *Memoryless*

Geometric random variables are called "**memoryless**"

Suppose you're flipping coins independently until you see a heads.

$\boldsymbol{X \sim \mathbf{Geo}(p)}$ is number of flips till the first head

> The first three came up tails.

> $\boldsymbol{Y}$ is number of flips *left* until you see the first head *after the first 3 tails*

Does $\boldsymbol{Y}$ *also follow* $\mathbf{Geo}(\boldsymbol{p})$? Yes!

The coin "forgot" it already came up tails 3 times.

T T T ? ? ? ? …

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 $p_X(k)$.

A geometric distribution is **memoryless:**

> If $X \sim \mathrm{Geo}(p)$
> $\mathbb{P}(X \geq a + b | X \geq a) = \mathbb{P}(X \geq b)$

Discrete Zoo of Random Variables (today!)

- **Uniform:** Every integer between a and b are equally likely
  $\mathrm{Unif}(a, b)$
- **Bernoulli:** Whether there is success in one trial
  $\mathrm{Ber}(p)$ is 1 with probability 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 till first success
  $\mathrm{Geo}(p)$ - probability p of success on each trial
- **Poisson:** Number of successes in a time interval
  $\mathrm{Poi}(\lambda)$ - average number of successes in the time interval
- **Negative Binomial**: Number of trials till r'th success
  $\mathrm{NegBin}(r, p)$ - probability p of success on each trial, want trials till the r'th success
- **Hypergeometric:** Number of successes when drawing a sample
  $\mathrm{HypGeo}(N, K, n)$ - drawing a sample of n items from a set of N with K successes

Scenario: The Poisson Distribution

We're trying to count the **number of times something happens in some interval of <u>time</u>**.

Examples:

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 / binomial distribution?

What are the flips…the number of cars? Every person who might visit the bakery?
There are too many of these to count exactly or think about dependency between.

Let's try modeling with a Poisson distribution.

Scenario: The Poisson Distribution

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

> Events occur at a fixed **rate** per time interval, but at random times

> We know the expected number of events in the time interval

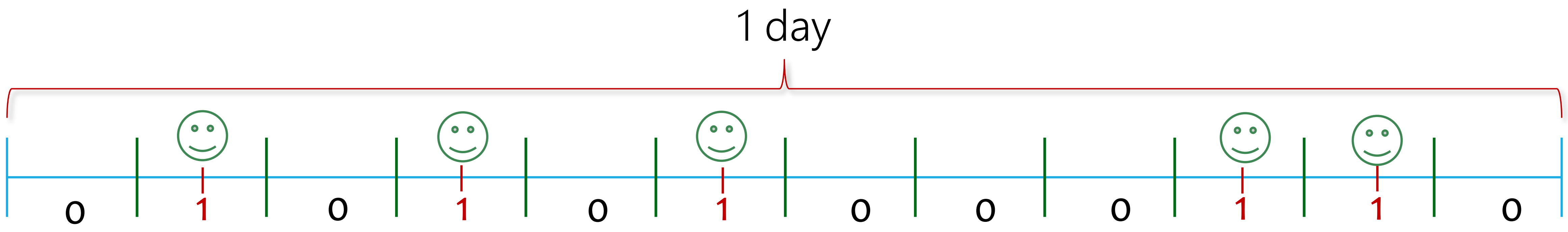
> Event occurrences on disjoint time intervals are independent of the others

Example: model **# of people who buy a mango chili lime donut in a day**

Scenario: The Poisson Distribution

X - # of people who buy a mango chili lime donut in a day, **$E[X] = \lambda > 0$**

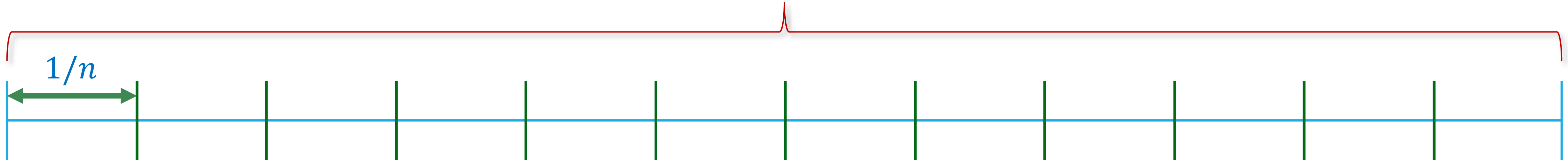
Assume events on disjoint time intervals occur independently



Scenario: The Poisson Distribution

X - # of people who buy a mango chili lime donut in a day, **E[X] = λ > 0**
Assume events on disjoint time intervals occur independently

Divide 1 day into n intervals of length $1/n$

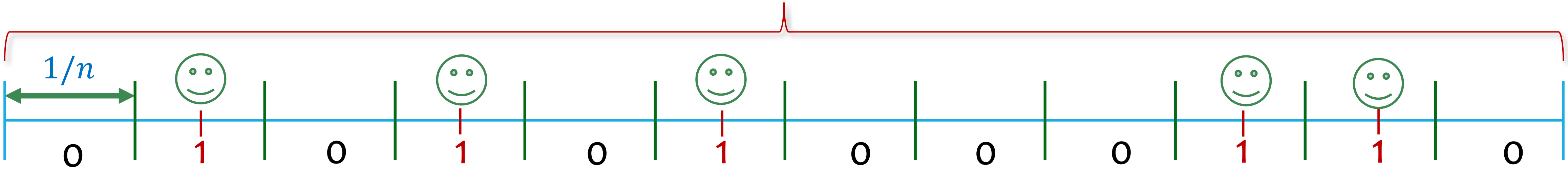


n independent intervals, assume either 0 or 1 per interval
Let p = probability person buys donut in a single interval

Scenario: The Poisson Distribution

X - # of people who buy a mango chili lime donut in a day, **E[X] = λ > 0**
Assume events on disjoint time intervals occur independently

Divide 1 day into n intervals of length $1/n$

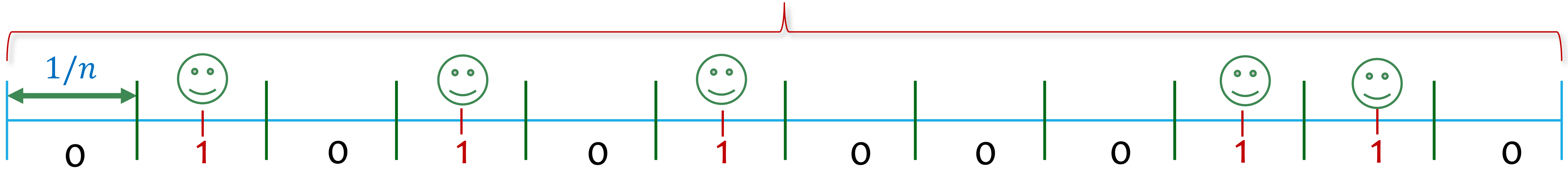


n independent intervals, assume either 0 or 1 per interval
Let p = probability person buys donut in a single interval = λ / n

Scenario: The Poisson Distribution

X - # of people who buy a mango chili lime donut in a day, **E[X] = λ > 0**
Assume events on disjoint time intervals occur independently

Divide 1 day into n intervals of length $1/n$



n independent intervals, assume either 0 or 1 per interval
Let p = probability person buys donut in a single interval = λ / n

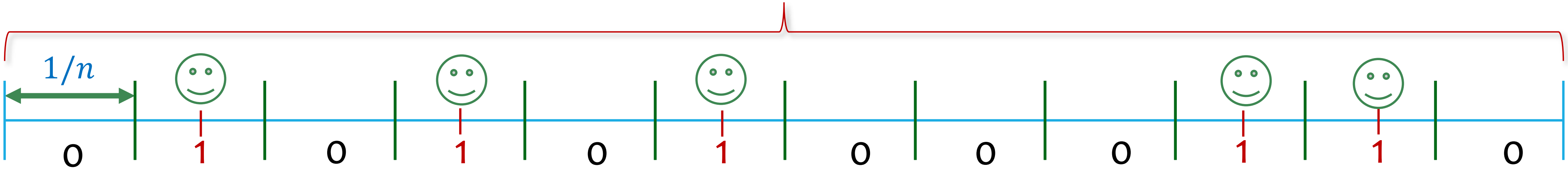
Each interval is $X_i \sim \text{Ber}\left(p = \frac{\lambda}{n}\right)$, $X_i = \begin{cases} 1 & \text{if person in ith interval} \\ 0 & \text{otherwise} \end{cases}$

Scenario: The Poisson Distribution

X - # of people who buy a mango chili lime donut in a day, **E[X] = λ > 0**

Assume events on disjoint time intervals occur independently

Divide 1 day into n intervals of length $1/n$



Each interval is $X_i \sim \text{Ber}\left(p = \frac{\lambda}{n}\right)$, $X_i = \begin{cases} 1 & \text{if person in ith interval} \\ 0 & \text{otherwise} \end{cases}$

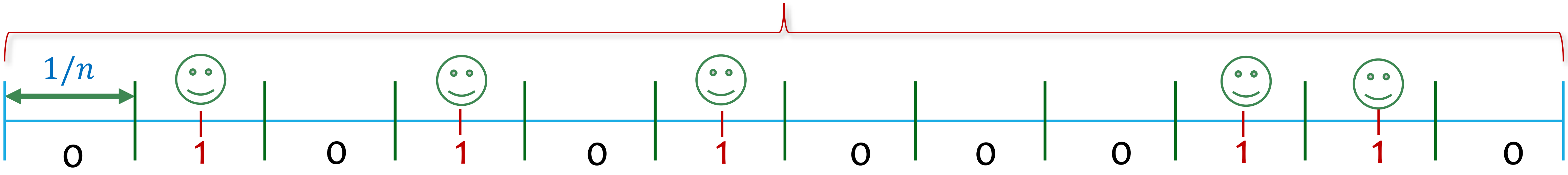
$X = \sum_{i=1}^{n} X_i$

$X \sim \text{Bin}(n, p)$ $\qquad P(X = i) = \binom{n}{i} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i}$ $\qquad$ indeed! $\mathbb{E}[X] = pn = \lambda$

Take the limit

X is binomial, $P(X=i)=\binom{n}{i}\left(\frac{\lambda}{n}\right)^i\left(1-\frac{\lambda}{n}\right)^{n-i}$

Divide 1 day into n intervals of length $1/n$, as $n \to \infty$



$$P(X=i)=\binom{n}{i}\left(\frac{\lambda}{n}\right)^i\left(1-\frac{\lambda}{n}\right)^{n-i} = \underbrace{\frac{n!}{(n-i)!\,n^i}}_{\to 1}\frac{\lambda^i}{i!}\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}}\underbrace{\left(1-\frac{\lambda}{n}\right)^{-i}}_{\to 1}$$

$$\to P(X=i)=e^{-\lambda}\cdot\frac{\lambda^i}{i!}$$

Poisson Distribution

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

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

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

Support: $\mathbb{N}$ (all *natural* numbers)

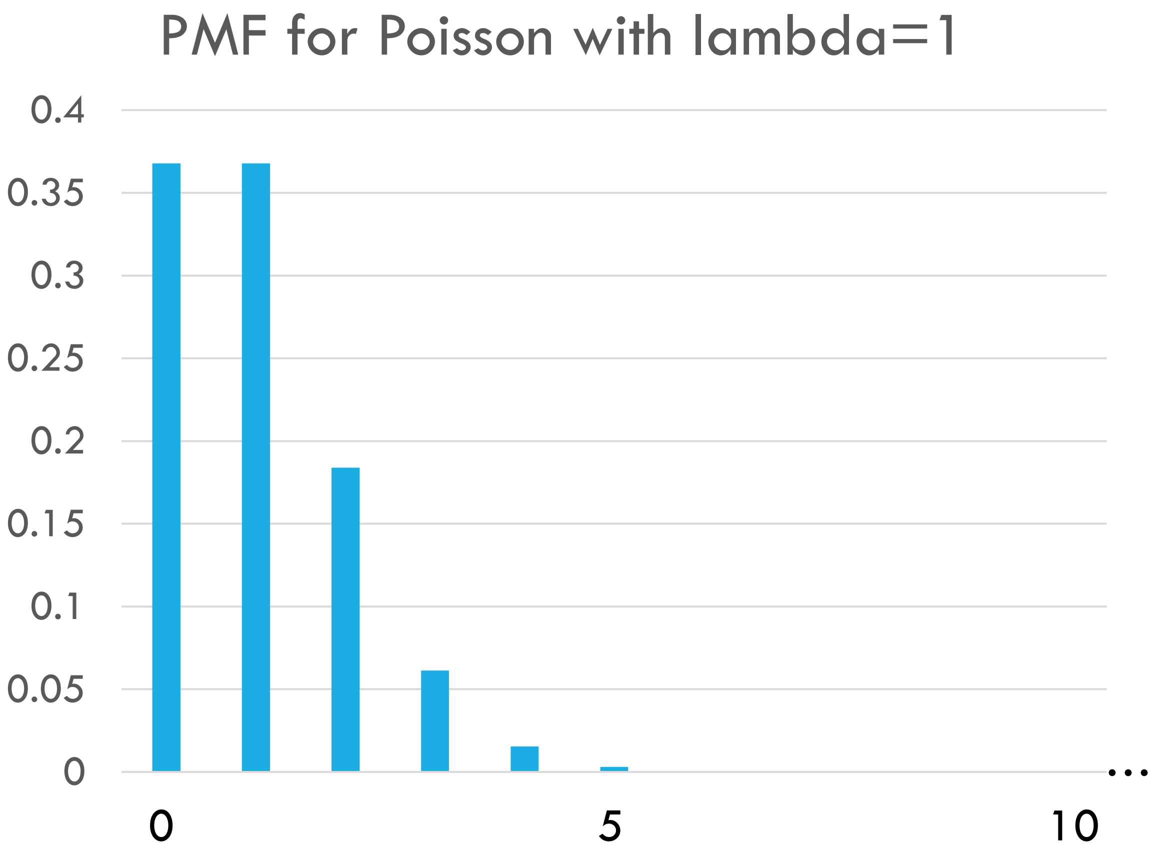
PMF: $p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$ (for $k \in \mathbb{N}$)

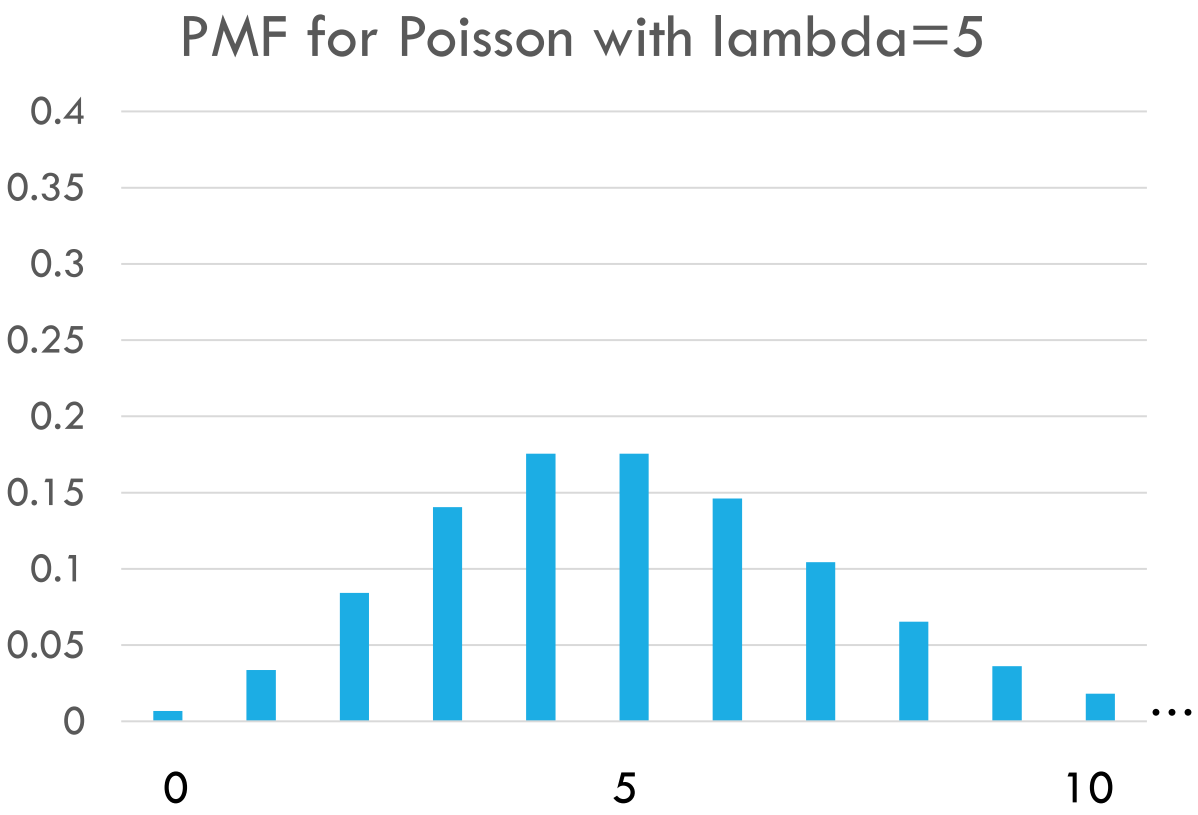
CDF: $F_X(k) = e^{-\lambda} \sum_{i=0}^{\lfloor k \rfloor} \frac{\lambda^i}{i!}$

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

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

Poisson Distribution (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.

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

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

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!

From Binomial to Poisson

Binomial approaches Poisson in the limit as $n \to \infty$ (equivalently, $p \to 0$) while holding $np = \lambda$

Poisson approximates Binomial when:

n is very large, p is very small, and $\lambda = np$ is "moderate"

e.g. ($n > 20$ and $p < 0.05$), ($n > 100$ and $p < 0.1$)

Example: From Binomial to Poisson

Consider sending bit string over a network

Send bit string of length $n = 10^4$

Probability of (independent) bit corruption is $p = 10^{-6}$

What is probability that message arrives uncorrupted?

Using $X \sim \mathrm{Poi}(\lambda = np = 10^4 \cdot 10^{-6} = 0.01)$

$$P(X = 0) = e^{-\lambda} \cdot \frac{\lambda^0}{0!} = e^{-0.01} \cdot \frac{0.01^0}{0!} \approx 0.990049834$$

Using $Y \sim \mathrm{Bin}(10^4, 10^{-6})$

$$P(Y = 0) \approx 0.990049829$$

Scenario: Negative Binomial

Run <u>independent</u> trials with probability $\boldsymbol{p}$. How many trials do you need *until $\boldsymbol{r}$ successes*?

Example

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?

Try it

Run independent trials with probability p. How many trials do you need *until* r *successes*?

X is the number of trials till (and including) the r'th success

What is the **support** of X?

What's the **PMF**?

i.e., what is the probability it takes exactly k trials till the r'th success?

Fill out the poll everywhere: pollev.com/jolie312

Try it

Run <u>independent</u> trials with probability $\boldsymbol{p}$. How many trials do you need *until* $\boldsymbol{r}$ *successes*?

X is the number of trials till (and including) the r'th success

What is the **support** of X? $\Omega_X = \{r, r+1, r+2, \dots\}$

What's the **PMF**?

i.e., what is the probability it takes exactly k trials till the r'th success?

Negative Binomial Analysis

Run independent trials with probability p
X is the number of trials till (and including) the r'th success

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

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.

1. We want **exactly $\boldsymbol{r-1}$ of the first $\boldsymbol{k-1}$ to be successes** – this sounds like a binomial! It's the $p_Y(r-1)$ where $Y \sim \text{Bin}(k-1, r-1)$:
$\binom{k-1}{r-1}(1-p)^{k-1-(r-1)}p^{r-1} = \binom{k-1}{r-1}(1-p)^{k-r}p^{r-1}$

2. Multiply by p, probability **$\boldsymbol{k}$'th trial is success**

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

Negative Binomial Analysis

X is the number of trials till we see r successes

To see r successes:

We do trials until we see success 1.

Then do trials until success 2.

…do trials until success r.

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

How can we write X as a sum of random variables?

Fill out the poll everywhere: pollev.com/jolie312

Negative Binomial Analysis

X is the number of trials till we see r successes

To see r successes:

We do trials until we see success 1.

Then do trials until success 2.

...do trials 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 $\boldsymbol{Z_i}$ are independent.**

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

Negative Binomial

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

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.

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

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

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

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

Scenario: Hypergeometric

You have an urn with N balls, of which K are purple. You are going to draw n balls out of the urn **without** replacement. How many purple balls do we get in this sample?

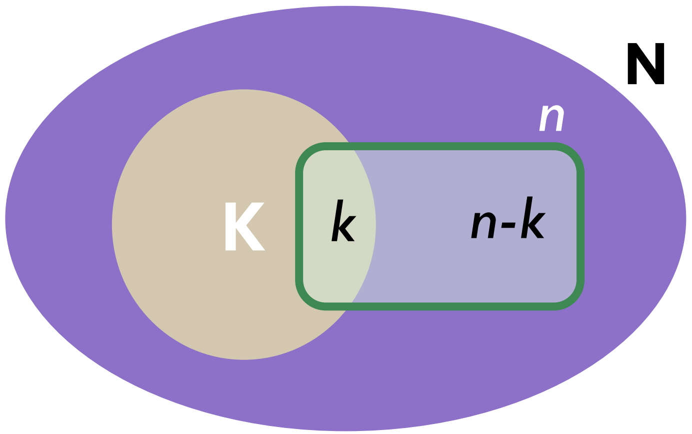
X is the number of purple balls in this sample

Hypergeometric: Analysis (PMF)

You have an urn with N balls, of which K are purple. You are going to draw n balls out of the urn **without** replacement. How many purple balls do we get in this sample?

<u>X is the number of purple balls in this sample</u>

If you draw out $\boldsymbol{n}$ balls, what is the probability you see $\boldsymbol{k}$ purple ones?



Hypergeometric: Analysis (PMF)

You have an urn with N balls, of which K are purple. You are going to draw n balls out of the urn **without** replacement. How many purple balls do we get in this sample?

X is the number of purple balls in this sample

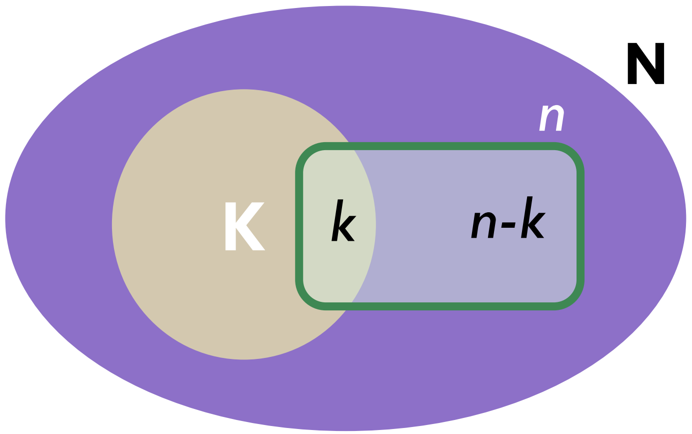
If you draw out $\boldsymbol{n}$ balls, what is the probability you see $\boldsymbol{k}$ purple ones?

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

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



Hypergeometric: Analysis (Expectation)

X is the number of purple balls in the sample

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

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

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

X is the number of success balls drawn in the sample.

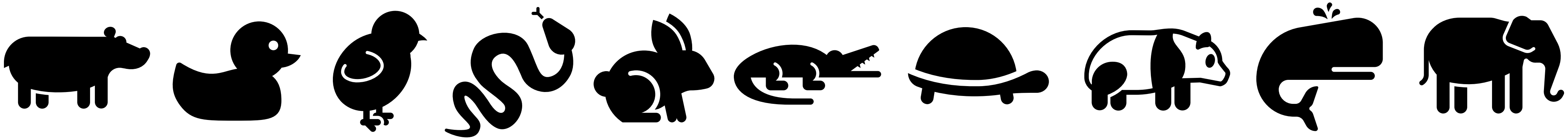
Parameters: A total of N balls in an urn, of which K are successes. Draw n balls without replacement.

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

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

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

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?)

Discrete Zoo of Random Variables

- **Uniform:** Every integer between a and b are equally likely
  $\mathrm{Unif}(a, b)$
- **Bernoulli:** Whether there is success in one trial
  $\mathrm{Ber}(p)$ is 1 with probability 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 till first success
  $\mathrm{Geo}(p)$ - probability p of success on each trial
- **Poisson:** Number of successes in a time interval
  $\mathrm{Poi}(\lambda)$ - average number of successes in the time interval
- **Negative Binomial**: Number of trials till r'th success
  $\mathrm{NegBin}(r, p)$ - probability p of success on each trial, want trials till the r'th success
- **Hypergeometric:** Number of successes when drawing a sample
  $\mathrm{HypGeo}(N, K, n)$ - drawing a sample of n items from a set of N with K successes

Halfway Point!

What have we done over the past 4 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)