Discrete RV Zoo

CSE 312 Spring 21
Lecture 14

What Does Independence Give Us?

If X and Y are independent random variables, then

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

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

$$\mathbb{E}\left[\frac{X}{Y}\right] = \mathbb{E}[X] \cdot \mathbb{E}\left[\frac{1}{Y}\right]$$

Shifting the variance

We know that

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

What happens with variance?

i.e., what is $\mathrm{Var}(aX + c)$?

Pause and guess: what is $\mathrm{Var}(X + c)$? What is $\mathrm{Var}(aX)$?

Facts About Variance

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

Proof:

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

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

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

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

$= \mathrm{Var}(X)$

Facts About Variance

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

$$= \mathbb{E}\left[(aX)^2\right] - (\mathbb{E}[aX])^2$$

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

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

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

Shifting a random variable

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

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

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

$$\mathbf{Var}(\boldsymbol{a}\boldsymbol{X} + \boldsymbol{b}) = \boldsymbol{a}^{2}\mathbf{Var}(\mathbf{X})$$

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]

…

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. I check Wikipedia when I forget these.

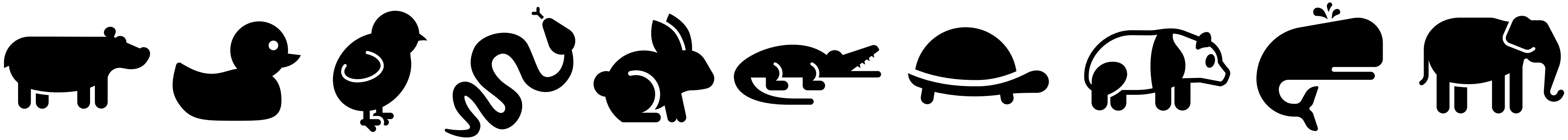
Our goals are:

0. Introduce one new distribution we haven’t seen at all.

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!



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

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

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

$$f_X(0) = 1-p;\ f_X(1) = p$$
$$\mathbb{E}[X] = p$$
$$\mathrm{Var}(X) = p(1-p)$$

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

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

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

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

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

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

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

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

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

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

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 $\mathbb{N}$

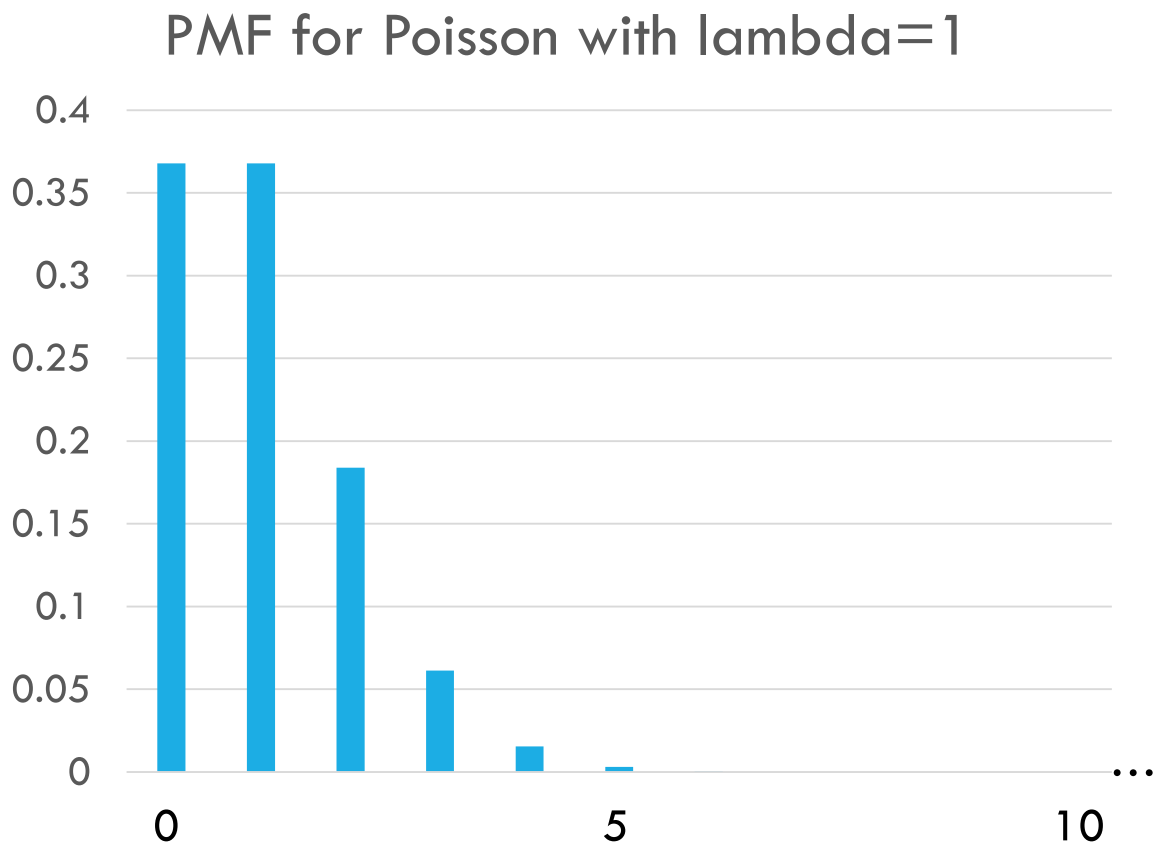
$f_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$ (for $k \in \mathbb{N}$)

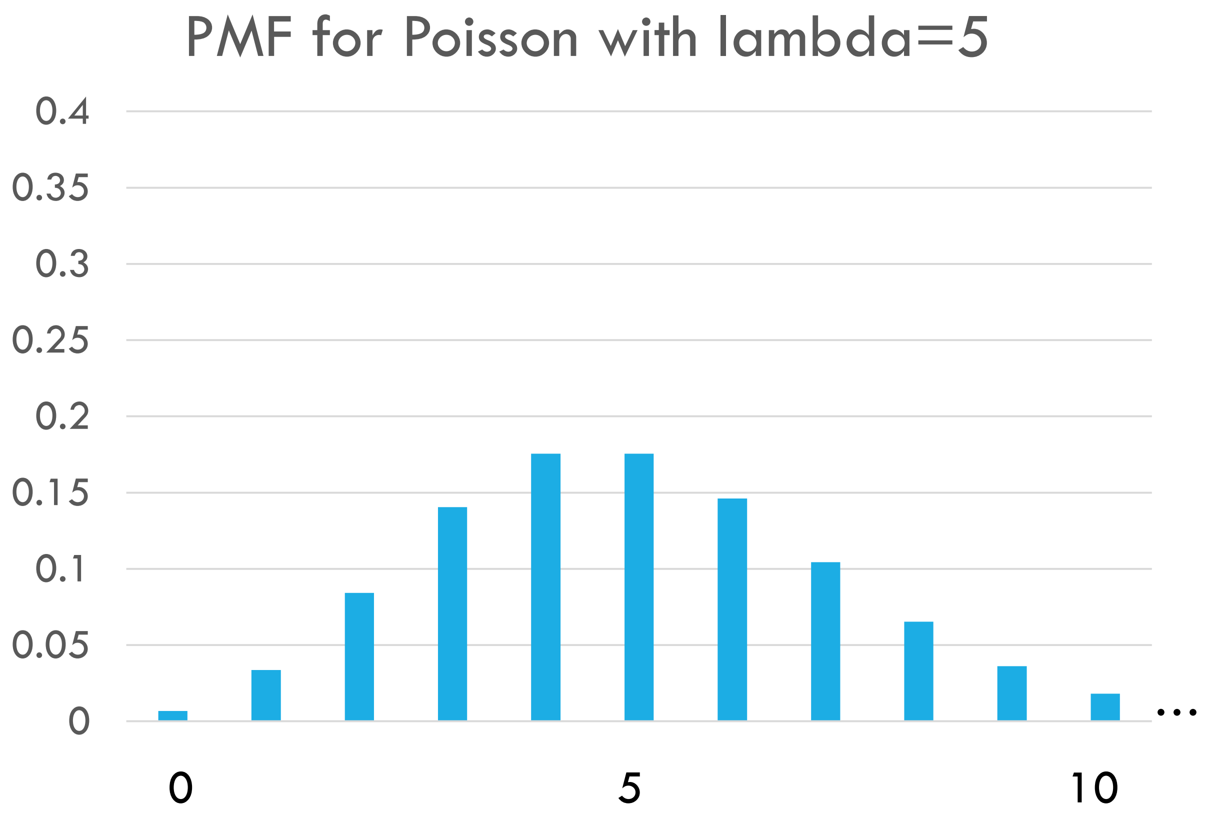
$F_X(k) = e^{-\lambda} \sum_{i=0}^{\lfloor k \rfloor} \frac{\lambda^i}{i!}$

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

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

Some Sample PMFs





Let's take a closer look at that pmf

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

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.

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

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

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

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

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.

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

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

$F_X(k) = 1-(1-p)^k$ 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: Expectation

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

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

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

$\mathbb{E}\left[X^2\right] = \sum_{k=1}^{\infty} k^2(1-p)^{k-1}p = p\sum_{k=1}^{\infty} k^2(1-p)^{k-1}$

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.

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

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

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

Which is $f_X(k)$.