CS1800 Day 15

Admin:
- HW5 released today
- grading estimates updated on canvas

Content:

Parametric Distributions
- Binomial
- Poisson

## In Class Activity

What are the chances that there is somebody in the room who has covid and is contagious right now?

- Get creative about your sources of evidence

- Make assumptions & estimates as necessary to get some value
    - Assumption tip: strike a balance between
        - assumptions which are strong enough to compute a value
        - assumptions which are trustworthy enough to give a meaningful result
    - Estimation tip: some quick googling can get you reasonable / justifiable values

- Evaluate your result, is your probability trustworthy or not?  How much do you think it might be off by?

Georgia Tech Study: in a room of 275 people if there are 800k cases -> 30% covid in room

date: current
assumption: uniformity of covid across the US

stopwatch: 2 mins 30 sec 9 coughes

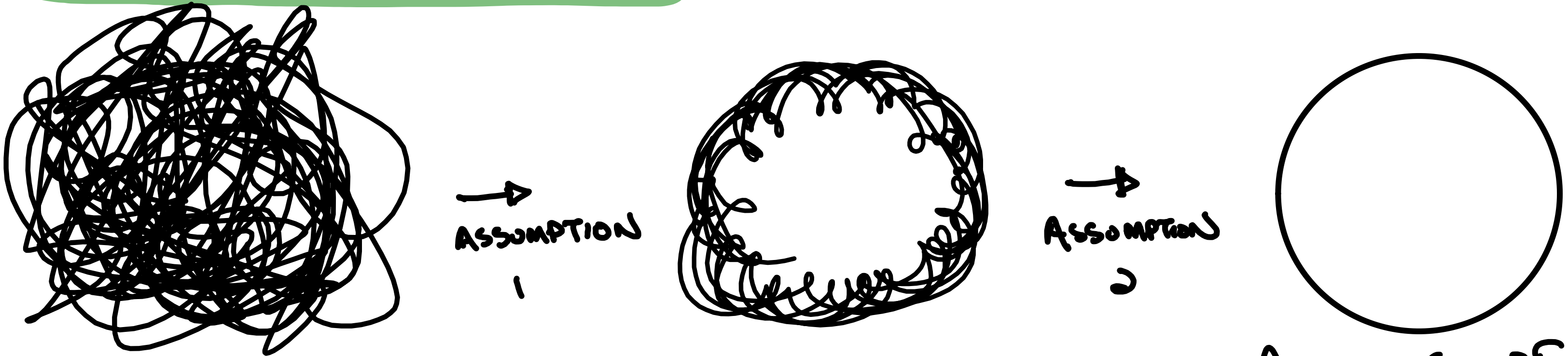
p = prob 1 student has covid in this room
n=250

(1- p)^250 = prob all of us don't have  covid

400 cases in boston / 800k boston = p

assume:
- everyone gets covid with same prob
- every covid event is independent of all others

# Building a math model of the real world




REALITY:

A MODEL OF REALITY

COMPLEX (TOUGH TO UNDERSTAND AND PREDICT)

RELEVANT (CONCLUSIONS DRAWN CAN BE TRUSTED)

SIMPLE (EASY TO UNDERSTAND AND PREDICT)

LESS RELEVANT (CONCLUSIONS DRAWN MAY NOT BE RELEVANT)


# Make assumptions to yield a model which is as simple / relevant as possible

Complex (Tough to understand and predict)

Relevant (Conclusions drawn can be trusted)

Simple (Easy to understand and predict)

Less relevant (Conclusions drawn may not be relevant)

“Essentially, all models are wrong, but some models are useful.” – George Box

## Independence

Intuition: Two experiments are independent if the outcome of one doesn't impact the other

Algebraically: If X and Y are independent then P(X, Y) = P(X) * P(Y)

Example:

Compute the probability of:

- first getting a heads on a fair coin flip A
- then getting a 5 on a fair six-sided die B
- winning a lotto (1 out of a million wins) C

$$P(ABC) = P(A) \cdot P(B) \cdot P(C)$$

$$= 1/2 \cdot 1/6 \cdot 1/1{,}000{,}000 =$$

## In Class Assignment

You flip a coin 10 times.
Each flip is independent of all others (e.g. heads on 2nd flip doesn't change prob heads on others)
Coin is "bent":

- P(heads on any flip) = .6
- P(tails on any flip) = .4

Compute the probabilities of the following events:

- 10 heads (in that order)
- 7 heads, 3 tails (in that order)
- 1 heads, 9 tails (in that order)
- 1 heads, 9 tails (any order)
- 3 heads, 7 tails (any order)
- N heads (any order). Write an expression which is valid for any N

hints:

- rely on your counting expertise
- do the problems in order (each offers insight to the next)

- 10 heads (in that order)

$$.6^{10} = P(HHHH\ldots H) = P(H) * P(H) \ldots * P(H)$$

- 7 heads, 3 tails (in that order)

10 TOTAL

$$.6^{7}\ .4^{3} = P(HHHHHH\ TTTT) = P(H)^{6} \cdot P(T)^{4}$$

- 1 heads, 9 tails (in that order)

$$.6 \cdot .4^{9}$$

H T T T T T T ....

T H T T T T T ...

T T H T T T T ...

- 1 heads, 9 tails (any order)

HOW MANY 1 HEAD 10 SPOTS →

$$\binom{10}{1} \cdot .6 \cdot .4^{9}$$

- 3 heads, 7 tails (any order)

HOW MANY WAYS PUT 3 HEADS IN 10 SPOTS →

$$\binom{10}{3}\ .6^{3} \cdot .4^{7}$$

$\boxed{H}\ \boxed{\ }\ \boxed{\ }\ \boxed{H}\ \boxed{\ }\ \boxed{H}\ \boxed{\ }\ \boxed{\ }\ \boxed{\ }\ \boxed{\ }$

- N heads (any order). Write an expression which is valid for any N

$$\binom{10}{N}\ .6^{N}\ .4^{10-N}$$

## Parametric Distributions (e.g. Binomial & Poisson)

Inutition:
A parametric distribution is a "template" distribution which can be used to model the real world

requires:
- a set of assumptions be satisfied

offers:
- an intuition honed on all the other examples of this distribution
- expressions for the expected value & variance of the random variable
- expressions for the probability of every outcome

## Bernoulli Distribution (a big name for a tiny little thing)

Describes the outcome of a single experiment with two possible outcomes.
(Conventionally, we call outcome 1 a "success" and 0 a "failure")

**Examples:**

| coin flip | covid test | raining |
|---|---|---|
| {1=heads, 0=tails} | {1=positive, 0=negative} | {1=raining, 0=not-raining} |

**Parameters:**
- p (probability of the "success" event)

**Assumes:**
- sample space is {0, 1}

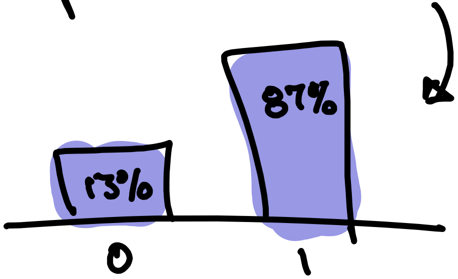
**Properties:**
- Expected Value = p
- Variance = p (1 - p)

DISTRIBUTION

$$P(X=1) = p$$

$$P(X=0) = 1-p$$

p = .87 BERNOULLI



## Binomial Distribution (adding together a bunch of Bernoullis)

Total successes in N trials with two possible outcomes.
(Conventionally, we call outcome 1 a "success" and 0 a "failure")

**Examples:**

| N coin flips | N covid test | rain in N days |
|---|---|---|
| {1=heads, 0=tails} | {1=positive, 0=negative} | {1=raining, 0=not-raining} |

**Parameters:**
- N (number of trials)
- p (probability of the "success" event)

**Assumes:**
- each trial is independent of all others
- each trial has same probability of "success"

**Properties:**
- Expected Value = N * p
- Variance = N p (1 - p)

## Binomial Distribution (whats it look like?)

### Parameters:

- N (number of trials)
- p (probability of the "success" event)

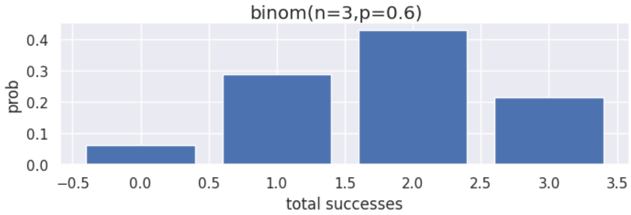
### Properties:

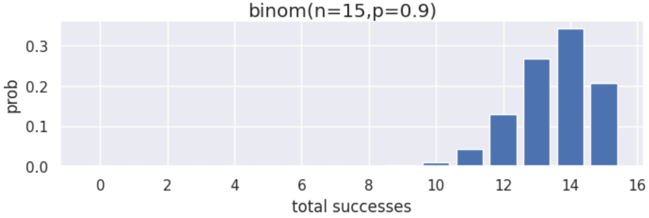
- Expected Value = N * p
- Variance = N p (1 - p)

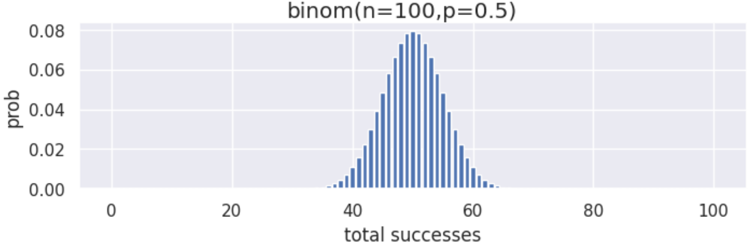
### Distribution:

$$P(X=k) = \binom{N}{k} p^k (1-p)^{N-k}$$

PROBABILITY OF GETTING K SUCCESSES AMONG N TRIALS



## In Class Activity: Binomial Distribution

$p = .15 = \frac{150}{1000}$ "SUCCESS" = KIDS SONG

Suppose spotify chooses your next song by selecting from among the 1000 previous songs you've listened to (each with an equal chance of being chosen). In my spotify history, 150 of my previous songs are childrens songs (e.g. Baby Beluga & PJ Masks are all too well represented!)

- If I play 5 spotify-chosen songs, what are the chances that exactly 1 is a children's song?

$n=5$
$p=.15$

$$P(K=1) = \binom{N}{K} p^K (1-p)^{N-K} = \binom{5}{1} (.15)^1 (1-.15)^4$$

- If I play 10 spotify-chosen songs, what are the chances that exactly 4 are children's songs?

$n=10$
$p=.15$

$$P(K=4) = \binom{10}{4} .15^4 (1-.15)^6$$

- If I play 15 spotify-chosen songs, what are the chances that no more than 1 are children's songs?
    - hint: whare are the chances that 0 or 1 are children's songs?

$$P(K=0) + P(K=1)$$

State each of the two binomial assumptions so they're easily understood by a non-technical in this context.  For each, give a circumstance which would violate this assumption (feel free to be creative).

assumption: each new trial is independent of all others

having chosen (or not) a kids song makes you no more or lress likely to choose a kids song next

violation: because no song is played back-to-back, we can't play same kids song back-to-back

assumption: the probability of "success" is the same across all trials

every new song has the same chance of being a kids song

violation: spotify gives more kids songs during the times of day when the kids are more likely
to be listening

# Poisson Distribution

Describes how many events occur in a given period of time

**Examples:**
Customers per minute in a shop, cars at a stoplight each hour,
engine failures per hour in a fleet of cars, text messages per hour in group of phones,
moose per square mile in a forest, illness cases per year in a country

**Parameters:**
- $\lambda$ (rate that events occur)

**Assumes:**
- rate is constant
(cars as likely to enter intersection at any moment)

- one event occuring does not make others more/less likely
(one car arriving at intersection doesn't make another more/less likely)

Poisson Distribution: whats it look like?

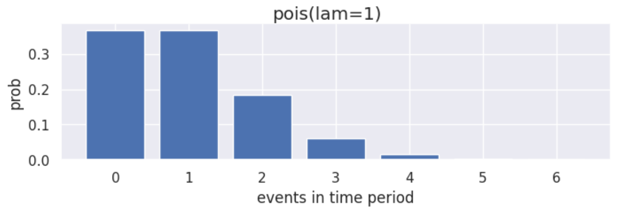
Parameters:
- λ (rate that events occur)

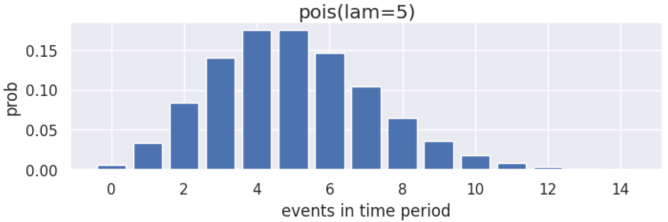
Properties:
- Expected Value = λ
- Variance = λ

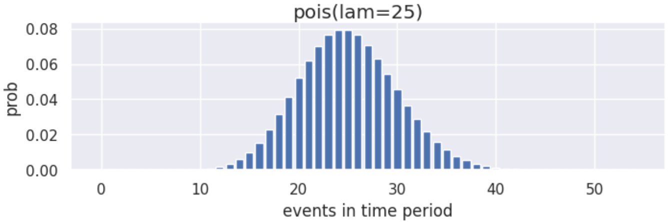
Distribution:

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

PROB OF HAVING K EVENTS OCCUR IN SOME TIME WINDOW



## Example: Flat Bike Tires

Over the past 2352 miles I've riden my bike, I've gotten 11 flat tires.

- State and critique each poisson assumption in this context

assumption: rate is constant.  each mile I"ve riden my bike I"m equally likely to get a flat tire

assumption: one event doesn't impact the chance of another another happening sooner / later.
if I do get a flat tire, I'm as likely to have a second flat in the next mile as if I never got the flat tire.

- Build a poisson model of flat bike tire events per 100 miles on the bike

2352 miles = 100 miles * 23.52

lambda = 11 / 23.52 flat tires per 100 miles

- Compute the chance of not getting ~~another~~ 1 flat in the next 100 miles on the bike (from the poisson)

$K=1$

$$P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{\left(11/23.52\right)^1 e^{-(11/23.52)}}{1!}$$

In Class Activity:

Skill: applying & critiquing assumptions
For each of the situations below, clearly state each Poisson assumption and give a real-life circumstance which violates just this assumption (not the other)

- how many subway cars arrive in a metro station each hour

- coffees served at starbucks each hour from 6AM to 5PM

Skill: Computing with a Poisson
A starbucks serves, on average, 5 drinks in an hour.  This starbucks has only 3 coffee cups left.
Estimate the chances that the starbucks runs out of coffee cups in the next hour with a Poisson Distribution.

$$\lambda = 5 \text{ cups}/\text{HR}$$

$$1 - \big(P(X=0) + P(X=1) + P(X=2)\big)$$