CS1800 Day 13

Admin:
- HW5 released today

Content:

Parametric Distributions
- Binomial
- Poisson

$1 + 1 = ?$

In Class Activity

Imagine its the spring of 2022 and you're sitting in a classroom ...

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 as needed

- 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?

3-5%

average students in class 18
CDC: during height of covid 3-4% had covid

esetimate of percentage of population which covid now
20% of people with covid show no symptoms

assume: people with symptoms don't come to class

contagious: 10 days after onset of symptoms

20-30% more likely after spring break (people are travelling), estimate based on most students going home or going on vacation (rough estimate) assume: 100 people in the class

assume: everybody gets covid independent of all others

assume: 2% of students have covid and would come to class (not contagious, or aware)

X = # CONTAGIOUS STUDENTS IN CLASS

$X_i = 1$ IF STUDENT i HAS COVID + CONTAGIOUS

$$P(X \geq 1) = 1 - P(X=0)$$

$$= 1 - P(X_1 = 0 \; X_2 = 0 \; X_3 = 0 \; \dots X_{100} = 0)$$

$$= 1 - P(X_1 = 0) P(X_2 = 0) P(X_3 = 0) \dots$$

$$= 1 - (1 - .02)^{100} \approx 86\%$$

Building a math model of the real world

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)

COIN HEADS   DIE ROLL   LOTTO WIN

$$P(C=1, D=5, L=1)$$

$$= P(C=1)\,P(D=5)\,P(L=1)$$

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

Example:

Compute the probability of:

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

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)

$P(THTH\ THTHTH) = (1/2)^{10}$

ASSUMES FAIR DIE (NOT LIKE YOUR PROBLEM)

- 10 heads (in that order) $(.6)^{10}$

- 7 heads, 3 tails (in that order) $(.6)^7 (1-.6)^3$

HHHHHHHTTT

- 1 heads, 9 tails (in that order) $(.6)^1 (1-.6)^9$

- 1 heads, 9 tails (any order) $(.6)^1 (1-.6)^9 \binom{10}{1}$

10

.6

HTTTTTTTTT

THTTTTTTTT $\to (.4)(.6)(.4)^8$

TTHTTTTTTT $\to .4 \cdot .4 \cdot .6 \cdot .4^7$

TTTHTTTTTT

$\vdots$

$.4^9 \cdot .6$

- 3 heads, 7 tails (any order) $(.6)^3 (1-.6)^7 \binom{10}{3}$

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

$$P(X=N) = (.6)^N (1-.6)^{10-N} \binom{10}{N}$$

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:
- quick intuition on new problems of this form (they're just like the old ones!)
- formulas 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)

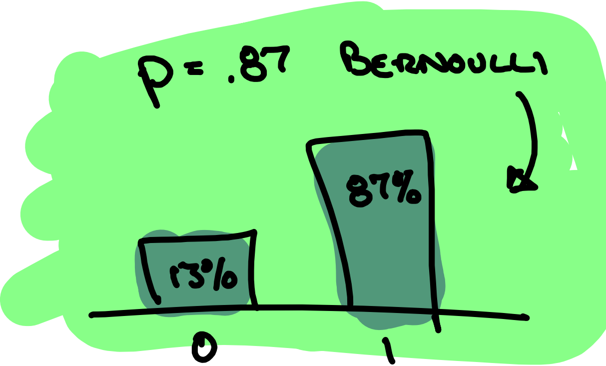
Assumes:
- sample space is {0, 1}

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

DISTRIBUTION

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

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



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)

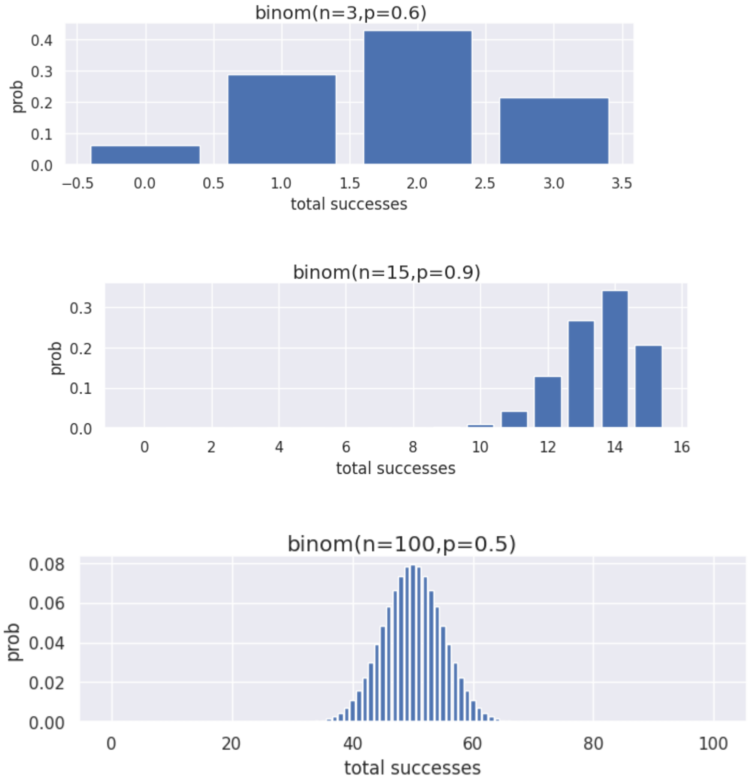
Properties:

- 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

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

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

Defining binomial parameters: N $\underline{p} = .15$

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

$N=5$

$$P(X=1) = \binom{5}{1} (.15)^1 (1-.15)^4 \approx .39$$

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

$N=10$

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

- 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?

$$P(X<2) = P(X=0) + P(X=1)$$

Assumes:
- each trial is independent of all others

selection of any song as a children's song does not impact the chance of another song being /not being a children's song

violation: if I play 1000 kid songs in a row, I'm garaunteed another kids song

- each trial has same probability of "success"

each song selected has the same chance of being a kids song
violation: history of 1000 songs changes its composition away from 150 kids songs

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:

- λ (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)

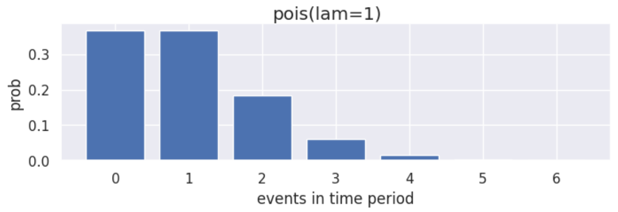
Properties:

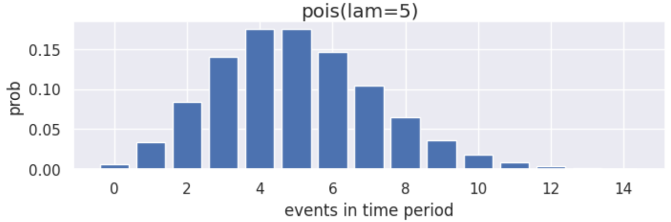
- Expected Value = λ
- Variance = λ

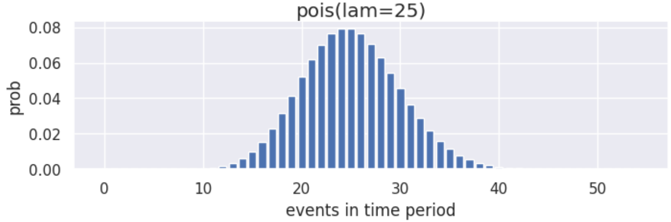
Distribution:

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

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

constant rate: a flat tire is as likely to occur at any mile of the trip

one event doesn't make others more / less likely: having a flat tire on mile 1, doesn't make it more or less likely that a flate tire occurs on mile 2 (or any other)

- Build a poisson model (i.e. find a rate parameter) of flat bike tire events per mile on the bike

    (trust your first intuition about estimating this rate parameter, it is that simple)

lambda = 11 flat tires / 2352 miles = 0.004676871 flat tires per mile

- Compute the chance of not getting another flat in the next mile on the bike (from the poisson)

P(x=0) = $\frac{\lambda^k e^{-\lambda}}{k!} = \frac{.0046^0 e^{-.0046}}{0!}$

-

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

(just modify your rate parameter to be valid for 100 mile stretches…)

lambda = 11 flat tires / 2352 miles = 0.004676871 flat tires per mile

0.4676871 flat tires per 100 mile

$$P(X=0) = \frac{.4676^{0}\, e^{-.4676}}{0!}$$

In Class Activity:
Skill: applying & critiquing assumptions

For each of the situations below, clearly state each Poisson assumption in the context of the problem and give a real-life circumstance which violates just this one assumption (not the other)

- arrival of a subway train in a metro station

constrant rate: subway cars as likely to arrive at any moment as any other
violation: late at night fewer cars arrive (fewer cars available)

one event doesn't make others more / less likely: subway car arrival at one moment doesn't make it more or less likely that a subway car arrives at any other moment
violation: arrival of one train means that another shouldn't come soon (logistics)

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

constrant rate:
one event doesn't make others more / less likely:

Skill: Computing with a Poisson $\lambda = 5$

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.

$$P\left(\text{Run out of cups}\right) = P(X \geq 4) = 1 - P(X \leq 3)$$

$$= 1 - P(X=0) - P(X=1) - P(X=2) - P(X=3)$$

$$P(X=2) = \frac{5^2 e^{-5}}{2!} = \frac{\lambda^k e^{-\lambda}}{k!}$$