CS1800 Day 13

Admin:
- HW5 released today

Content:

Parametric Distributions
- Binomial
- Poisson

Practice: give me the wrong answer

$1+1 = ?$

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?

google said 0.8%, what is the probability that someone from any sample size has covid

problem: number isn't current
problem: just for one person ... there's a whole bunch in the class

cases in april of 2022 in US / population of the US = 1 case per 256 people

assume: covid spread uniformly across the US

2021 900 cases from Northeastern per semester / 20k students = 4.5% of students have covid once during the whole semester

assume: contagious for 3 weeks with covid and NU semester is 12 weeks: 4.5% -> 4.5% / 4

$$P\left(\text{AT LEAST 1 COVID IN CLASS}\right) = 1 - P\left(\text{NO COVID IN CLASS}\right)$$

$$= 1 - \left(1 - \frac{.045}{4}\right)^{250}$$

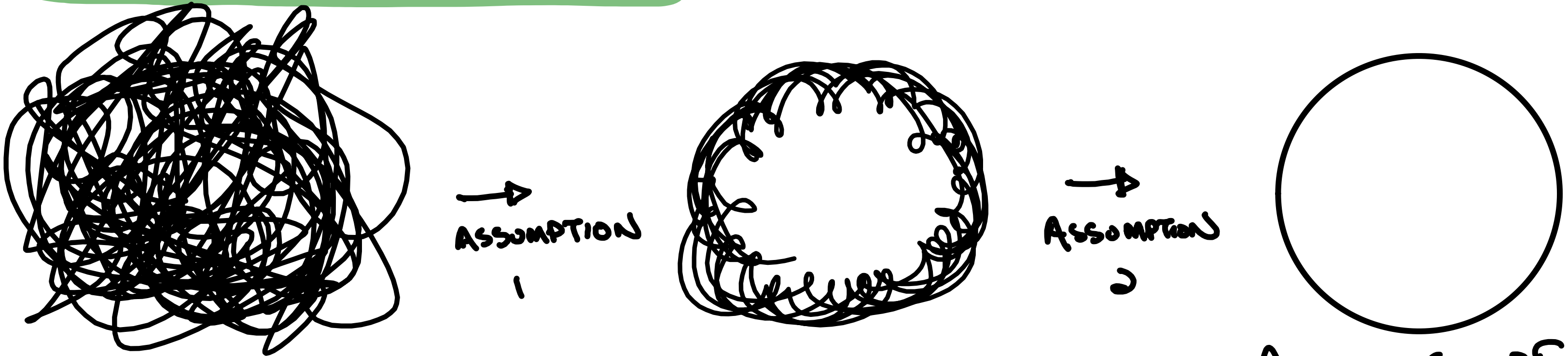
$$= .94$$

ASSUME

EACH COVID IS INDEP

→ EACH COVID CASE OCCURS 4.5% / 4

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)

EVENT COIN IS HEADS

$P(C=1) = 1/2$

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)

$P(D=5) = 1/6$

$P(L=1) = \frac{1}{1\,000\,000}$

$P(C=1\ D=5\ L=1) = 1/2 \cdot 1/6 \cdot \frac{1}{1\,000\,000}$

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)

$$.6^{10} = P(X_1 = \text{HEADS}\ X_2 = \text{HEADS} \ldots) = P(X_1 = \text{HEADS})\, P(X_2 = \text{HEADS}) \ldots$$

$P(\text{HEADS}) = .6$

$P(\text{TAILS}) = .4$

- 7 heads, 3 tails (in that order)

$$.6^7 \cdot .4^3 = P(\text{HHHHHHHTTT}) = P(H) \cdot P(H) \cdot P(H) \cdot \ldots P(T) \cdot P(T) \cdot P(T)$$

- 1 heads, 9 tails (in that order)

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

- 1 heads, 9 tails (any order)

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

WAYS OF PUTTING 1 HEAD IN 10 TOTAL COIN FLIPS

- 3 heads, 7 tails (any order)

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

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

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

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:

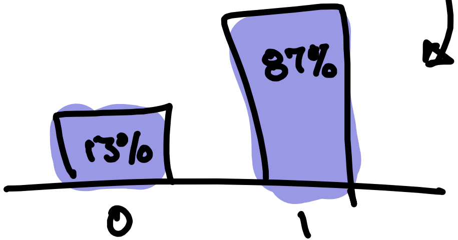
- Expected Value = p
- Variance = p (1 - p)

DISTRIBUTION

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

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

$p = .87$ 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)

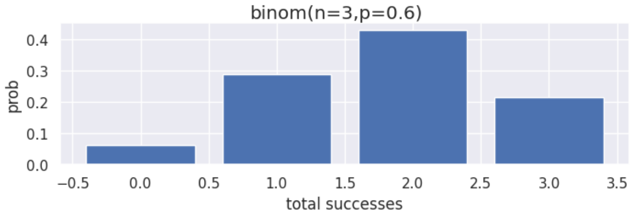
Properties:

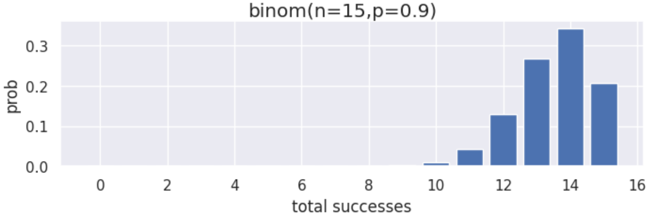
- Expected Value = N * p
- Variance = N p (1 - p)

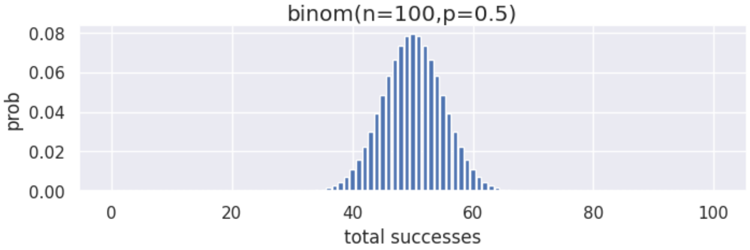
Distribution:

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

PROBABILITY OF GETTING K SUCCESSES AMONG N TRIALS







In Class Activity: Binomial Distribution

N = # songs played

"Success" = kids song $\quad P = \frac{150}{1000} = .15$

Suppose spotify chooses your next song by selecting from among a *fixed set of 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!)

*by fixed, we mean that playing a children's song does not change the probability of another children's song (there may be 1 more children's song in the previous 1000 if it weren't "fixed")

- 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).

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).

- each trial is independent of all others

one song being a children's song does not change the chance of any other song being a children's song

- each trial has same probability of "success"

every song has the same chance of being a children's song (15 percent)

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

$P = .15$

$N = 5 \quad K = 1$

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

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

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

- 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?

$$P(X \leq 1) = P(X=0) + P(X=1)$$

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)

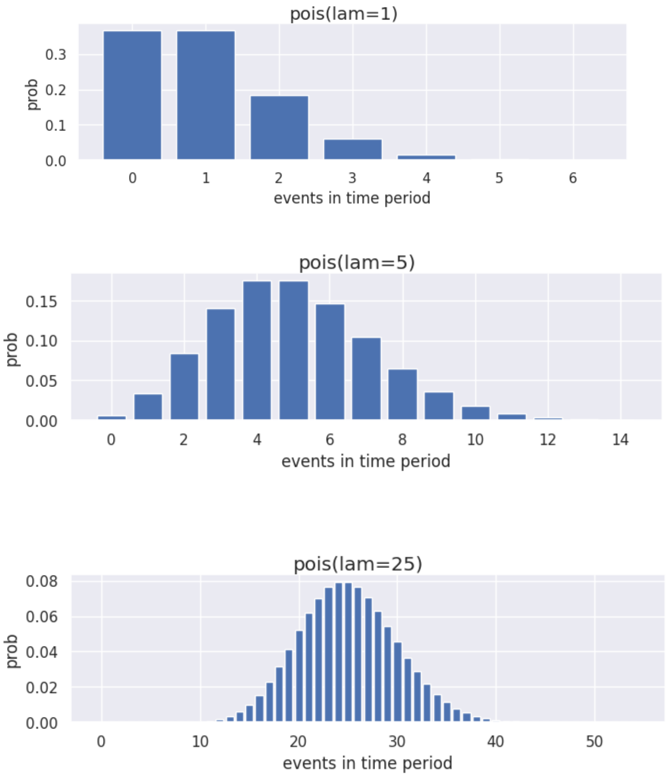
Properties:

- 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

- rate is constant: every mile is as likely as the next to have a flat tire

- one event occuring does not make others more/less likely: if a flat tire occurs on the last mile, another flat tire is not more or less likely on the next mile

- 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 flats / 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…)

0.004676871 flats / mile = 0.4676871 flats / 100 mile

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)

poisson assumptions:

- rate is constant
- one event does not make another more or less likely

- arrival of a subway car in a metro station
    - every hour of the day subway cars arrive as frequently as any other (late at night: run fewer trains)
    - arrival of a subway car doesn't make it more / less likely than another subway car is just behind (no more than 1 car at a station)

- coffees served at starbucks each hour from 6AM to 5PM
    - each hour is as busy (number of coffees) as any other
    - ordering of one coffee doesn't make it more / less likely that another is ordered

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.

$$\lambda = \frac{5 \text{ DRINKS}}{\text{HOUR}}$$

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

$$= 1 - \frac{5^0 e^{-5}}{0!} - \frac{5^1 e^{-5}}{1!} - \frac{5^2 e^{-5}}{2!} - \frac{5^3 e^{-5}}{3!}$$

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