Agenda

Professor Hamlin

Day 1

1) Admin — HW5 released
2) Review — Midterm grades released
3) Binomial — regrades on gradescope
4) Poisson — Canvas grade so far

Review

Joint distribution - $Pr[C1=T, C2=T]$

marginalization

Conditional probability: if x then prop y happens

$Pr[Y=y \mid X=x]$

Bayes Rule: if $Pr[X=x \mid Y=y]$ what is $Pr[Y=y \mid X=x]$

Independent Variables: $Pr[X=x, Y=y] = Pr[X=x]\,Pr[Y=y]$

Exercise:

1) State the following in math terms
   a) odds of being cool & having a cat

   $Pr[Cool=1, Cat=1]$

   b) likelyhood of being cool given you have a cat

   $Pr[Cool=1 \mid Cat=1]$

2) if $Pr[Cat=1 \mid Cool=1] = .4$, $Pr[Cat=1 \mid Cool=0] = .2$ & $Pr[Cool=1] = .6$ what is $Pr[Cat=1]$?

$$Pr[Cat=1] = Pr[Cat=1 \mid Cool=1] \cdot Pr[Cool=1] + Pr[Cat=1 \mid Cool=0] \cdot Pr[Cool=0]$$

$$= .4 \cdot .6 + .2 \cdot .4$$

$$= \boxed{.32}$$

In class activity

1) What are the chances that someone in this room has a birthday today? (No spoilers!)

Note: Get creative, make assumptions and estimates as necessary

For assumptions → strong enough to make a guess vs. not realistic

For Estimates → a quick google can give you some numbers

How valid is your guess? How much do you think you are off?

https://www.cdc.gov/nchs/nvss/vsrr/provisional-tables.htm

Being more accurate w/ October birthday

309,000 born in Oct

3,621,000 born over 2023

.275% prob of being born

$1 - \underbrace{90 \cdot \tfrac{1}{365}}$ odds no one has birthday

$\left(1 - \frac{1}{365}\right)$ odds don't have birthday

$\left(1 - \frac{1}{365}\right)^{90}$ odds 90 people don't have birth

$1 - (1 - 1/365)^{90}$ odds some one has Birthday today

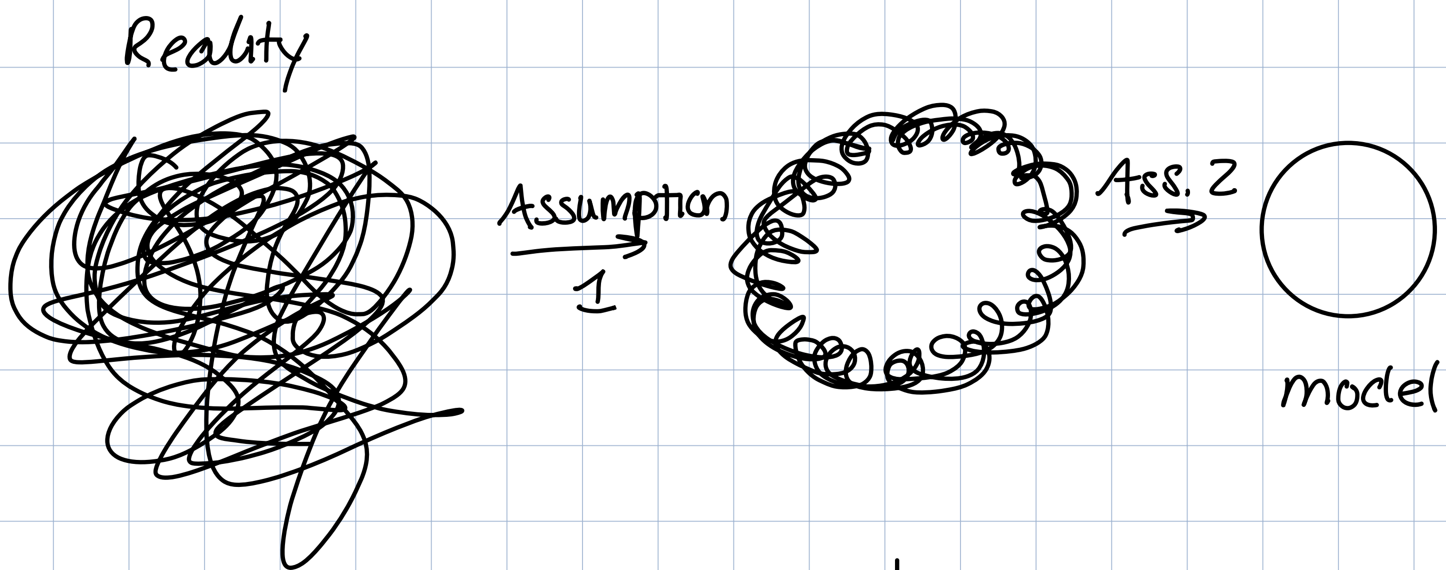
.2739%

$(1 - .0275)$ prob not born Oct 25th

$(1 - .0275)^{90}$ prob

$1 - (1 - .0275)^{90}$

Building a math model of the real world



Complex (tough to understand or predict) | simple (easy to understand and predict)

Relevent (conclusions drawn can be trusted) | Less Relevent (conclusion may not be trusted)

= what we want

> "Essentially, all models are wrong, but some models are useful"
> — George Box

Recall: Independence

Two random variables don't effect each other e.g.

$$P[X=x, Y=y] = Pr[X=x] \cdot Pr[Y=y]$$

Exercise: lucky day

1. What are the odds of
   a. getting heads on fair coin then
   b. 5 on 6-sided die
   c. winning 1 out of 1 million lotto

$$Pr[A=C, B=5, C=1] = \frac{1}{2} \cdot \frac{1}{6} \cdot \frac{1}{1,000,000}$$

2. X = Prob of getting sum of 12 on 2 6-sided dice
   Y = Prob of gettin 1 on that first die

$$Pr[X=1, Y=1] = 0$$

careful! not independent

Working towards our ~~first~~ econ general distribution

(Our first is the uniform distribution)

Certain types of dist. always look a like and if we recognize them we can just use the equations w/o figuring them out again

So let's start with an exercise.

Exercise 1 Flip a coin 10 times, each flip is independent. $Pr[C=H] = .6$

$Pr[C=T] = .4$

What are the following probabilities?

1. 10 Heads, in that order

$$\underline{.6}\cdot\underline{.6}\cdot\underline{.6}\ \underline{\quad}\ \ldots\ldots = (.6)^{10}$$

2. 7 Heads, 3 tails, in that order

H H H H H H H T T T

$$(.6)^7(.4)^3$$

3. 1 Head, 9 tails, in that order

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

4. 1 Head, 9 tails, any order

H _ _ _ _ _ _ _ _ _ $\quad (.6)\cdot(.4)^9$

_ H _ _ _ _ _ _ _ _ $\quad + .4\cdot .6\cdot(.4)^8 = (.6)(.4)^9$

_ _ H _ _ _ _ _ _ _ $\quad + (.6)\cdot(.4)^9$

_ _ _ H _ _ _ _ _ _

_ _ _ _ H _ _ _ _ _

_ _ _ _ _ H _ _ _ _

_ _ _ _ _ _ H _ _ _

_ _ _ _ _ _ _ H _ _

_ _ _ _ _ _ _ _ H _

_ _ _ _ _ _ _ _ _ H

$$10\cdot(.6)\cdot(.4)^9$$

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

5. 3 heads, 7 tails (any order)

$$\underset{\text{\# of cases}}{\binom{10}{3}}\underbrace{(.6)^3\cdot(.4)^7}_{\text{prob. per case}}$$

6. N heads, any order

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

This is an example of a Bernoulli Distribution

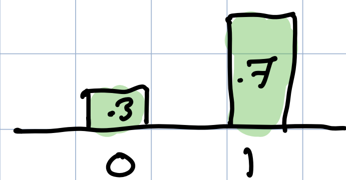
Bernoulli Distribution ~ describes the outcome of a single experiment with 2 outcomes
(1 = success , 0 = failure)

Examples: coinflips, covid tests, raining.

Parameters: p (probability of success)

Sample Space: 0, 1

Distribution $Pr[X=1] = p$ $Pr[X=0] = 1-p$



$p = .7$

Properties $E[X] = p$
$Var(X) = p \cdot (1-p)$

Binomial Distribution - describes outcome of a whole bunch of Bernoulli. The distribution total successes in N trials (remember two outcomes)

Examples N coin flips, N covid tests, etc.

Parameters: p (probability of success)
N (number of trials)

Sample space $[0 - N]$

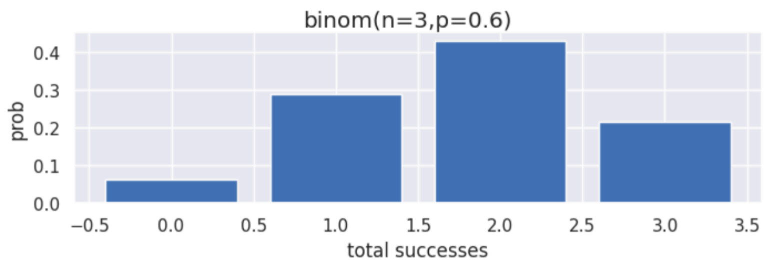
Assumptions: 1. each trial is independent of the others

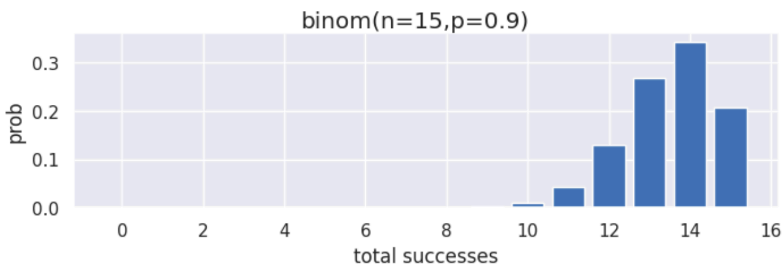
2. each trial has same prob of success

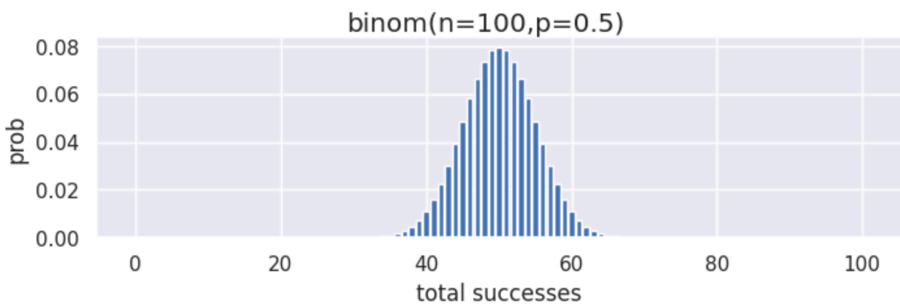
Distribution

$$Pr[X=k] = \binom{N}{k} p^k (1-p)^{N-k}$$

prob of getting k successes in N trials







Properties:

$$E[x] = N \cdot p$$

$$Var(x) = N \cdot p \cdot (1-p)$$

Exercise

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!)

$p = 150/1000 = .15$

1) if play 5 songs, chances of exactly 1 children's song?

$N=5 \quad Pr[X=1] = \binom{5}{1}(.15)(1-.15)^4$

2) Play 10 songs, chances exactly 4 are children's songs?

$N=10 \quad Pr[X=4] = \binom{10}{4}(.15)^4(1-.15)^6$

3) Play 15 songs. Chances no more than 1 is children's song? Binomial

$N=15 \quad Pr[X \leq 1] = Pr[X=0] + Pr[X=1]$

$\binom{15}{0}(.15)^0(1-.15)^{15} + \binom{15}{1}(.15)^1(1-.15)^{14}$

But what about our assumptions? Do they make sense:

1. each trial is independent of the others
2. each trial has same prob of success

Name a scenario that would violate each of these assumptions. In our example above

1. Chance of children's song is the same for each trial

spotify doesn't repeat songs, so playing children's song makes future ones less likely

2. each trial is independent

chosing song makes further songs more like original song.

Poisson Distribution - distribution of how many events occur in a given time or unit (like mile)

Examples: cars at stop light per hour
customers per minute at coffe shop
moose per square mile

Parameter: λ (lambda) (rate (something per something) that events occur)

Sample Space $[0 - \infty)$

Assumptions: 1. rate is constant
- cars are likely to enter intersection in any given moment

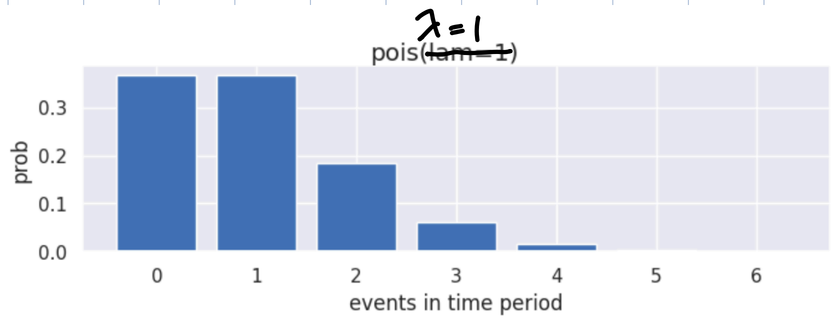
2. one event occurring does not make others more/less likely
one car does not make other cars more/less likely

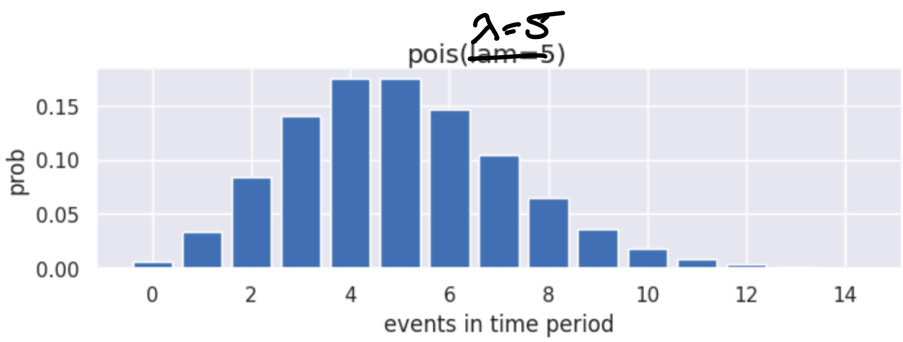
Distribution

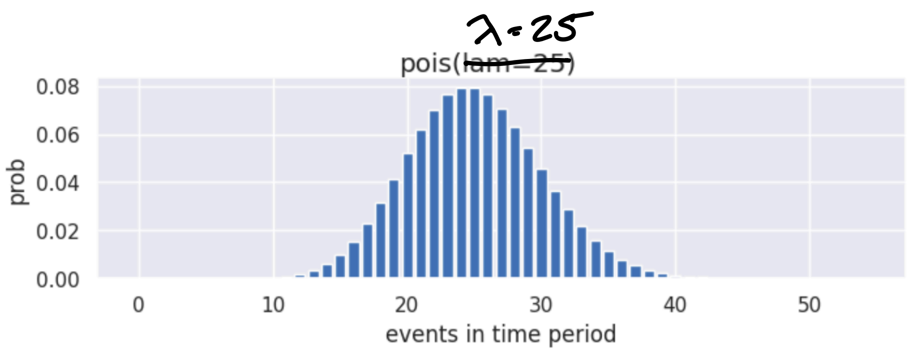
$$Pr[X=k] = \frac{\lambda^k e^{-\lambda}}{k!}$$

prob of k events happening during window

2.71828..... a common constant (Euler's #)







Example: Over the last 235 ski runs I've been on, I've fallen 5 times

Build a poisson model of falls per 10 runs then compute the chance of not falling in my next 10 runs from the model

$\boxed{\text{rate should match window}}$

$$\lambda = ?\ \text{falls}/10\text{runs} = \frac{5 \text{ falls}}{23.5\ \ 10\text{runs}} = .212\ \text{falls}/10\text{runs}$$

$$Pr[X=0] = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{.212^0 \cdot e^{-.212}}{0!}$$

Exercise

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)

1) arrival of a subway car at T station
   constant rate: more cars during rush hour
   independence: controller tries to evenly space things

2) coffees served at starbucks each hour from 6AM to 5PM
   constant: busier during rush hour
   independence: a group order

Exercise

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

$\lambda = 5$ drinks/hour

$$Pr[X \geq 4] = 1 - Pr[X=0] - Pr[X=1] - Pr[X=2] - Pr[X=3]$$

$\boxed{.735}$