DataSci 520

lesson 3

combinatorics and probability distributions

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

today's agenda

- discrete vs continuous random variables
- understand the 3 axioms of probability
- properties of basic probability distributions
- conditional, mutually exclusive, and independent events

Introduction to Counting & Probability

- Why counting?
  - Counting is fundamental to probability theory.
  - Probability is the extent or likelihood of an event or set of events.
    - Depends heavily on the ability to COUNT up potential outcomes.

Counting

- Counting is one of the biggest areas of mathematics, called "Combinatorics".
- Example:
  - Subway has 4 different breads, 5 different meats, and 4 different toppings. How many sandwich combinations are there?
  - How many different 4-beer tasters can I have in a bar with 10 beers on tap?
- Solve the above with the "Multiplication Principle".
  - Subway:

$$\underset{\text{\# of Breads}}{\underline{\quad 4 \quad}} * \underset{\text{\# of Meats}}{\underline{\quad 5 \quad}} * \underset{\text{\# of Toppings}}{\underline{\quad 4 \quad}} = 80 \text{ Sandwiches}$$

  - Beer Tasters:

$$\underset{\text{\# Beer Choices}}{\underline{\quad 10 \quad}} * \underset{\text{\# Beer Choices}}{\underline{\quad 9 \quad}} * \underset{\text{\# Beer Choices}}{\underline{\quad 8 \quad}} * \underset{\text{\# Beer Choices}}{\underline{\quad 7 \quad}} = 5{,}040 \text{ Beer Tasters}$$

Multiplication Principle

- If there are A ways of doing task-a, B ways of doing task-b, then there are A*B ways of completing both tasks.
- Example:
  - If I have 5 books, how many ways can I ORDER them on the bookshelf

$$\underset{\text{\# Book Choices}}{\underline{\quad 5 \quad}} * \underset{\text{\# Book Choices}}{\underline{\quad 4 \quad}} * \underset{\text{\# Book Choices}}{\underline{\quad 3 \quad}} * \underset{\text{\# Book Choices}}{\underline{\quad 2 \quad}} * \underset{\text{\# Book Choices}}{\underline{\quad 1 \quad}} = 120 \text{ Book Orderings}$$

Factorials

- Factorials
  - Count # of ways to order N things = N!
- Factorials get VERY VERY large quickly.
  - 21! Is larger than the biggest long-int in 64 bit.
    - 21! = 5.1E19
    - Biggest long int (64bit) = 9.2E18
  - Fun fact, every 52 card shuffle is highly likely to be the only time that shuffle has ever occurred.

Counting Subgroups

- Revisit: 10 beers on tap, need a sample of 4 different beers.
- Let's assume order matters.
  - E.g. Amber-Stout-Porter-Red is different than Red-Porter-Stout-Amber
- Use "Permutations" or "Pick":

$$10 * 9 * 8 * 7 = \frac{10!}{6!} = \frac{10!}{(10-4)!} = 10P4 = P(10,4)$$

Counting Subgroups

- Now we assume order doesn’t matter.
- Use ‘Combinations’ or ‘choose’.
- Remember that 10*8*7*6 = orderings of 4 beers.
- Now we are counting some groups multiple times, so we divide out the multiple counts.

$$\frac{(10 * 9 * 8 * 7)}{(\# \ of \ ways \ to \ order \ 4 \ beers)} = \frac{P(10,4)}{4!} = \frac{10!}{(10-4)!\,4!} = C(10,4) = \binom{10}{4}$$

More About Combinations

- Combinations appear on Pascal's Triangle!
- C(N,x) appears on the Nth row, x-th number over (starting counting at 0).

1

1 1

1 2 1

1 3 3 1

1 4 6 4 1

1 5 10 10 5 1

1 6 15 20 15 6 1

1 7 21 35 35 21 7 1

1 8 28 56 70 56 28 8 1

1 9 36 84 126 126 84 36 9 1

1 10 45 120 210 252 210 120 45 10 1

$\binom{10}{4}$

Counting Examples

- There are 10 Light beers on tap, and 10 Dark beers on tap.
- How many ways can Rick get a 4-beer sampler that contains exactly 1 light beer? (Order does not matter)

$$= \frac{(\#\ of\ ways\ for\ 1\ Light) * (\#\ of\ ways\ for\ 3\ D)}{(\#\ of\ ways\ to\ arrange\ 1L\ \&\ 3D)} = \frac{C(10,1) * C(10,3)}{C(4,1)} = \frac{10 * 120}{4}$$

$$= 300$$

Counting Examples

> 6:5 Blackjack is dealt with a 6 shoe deck (52*6=312 cards)

> How many ways can someone get dealt two rank 10 cards?

$$\binom{6decks \ * 4ranks \ * 4suits}{2} = \binom{96}{2} = 4{,}560$$

DICE!

> How many ways can two dice be rolled to get a sum of 10?

	2	3	4	5	6	7
	3	4	5	6	7	8
	4	5	6	7	8	9
	5	6	7	8	9	10
	6	7	8	9	10	11
	7	8	9	10	11	12

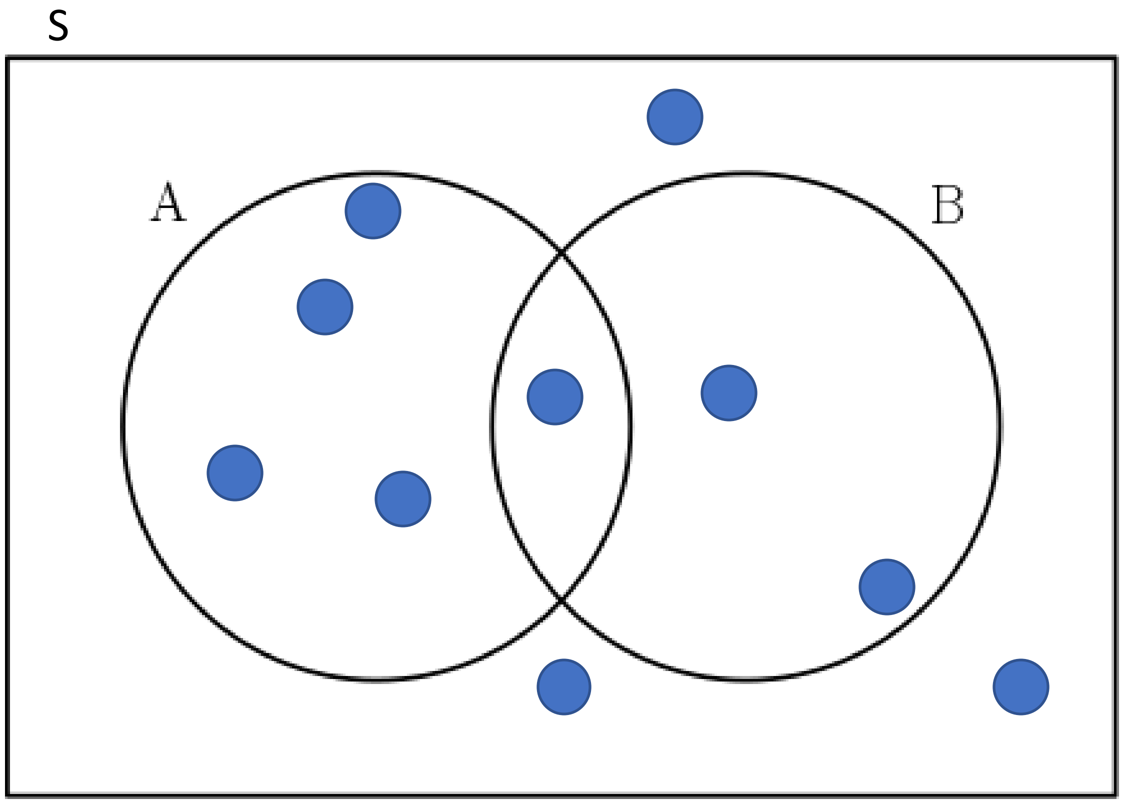
Probability

- The Probability of an event, A, is the number of ways A can occur, divided by the number of total possible outcomes in a “Sample Space”.
  - The sample space is the set of all possible outcomes.

$$P(A) = \frac{N(A)}{N(S)}$$

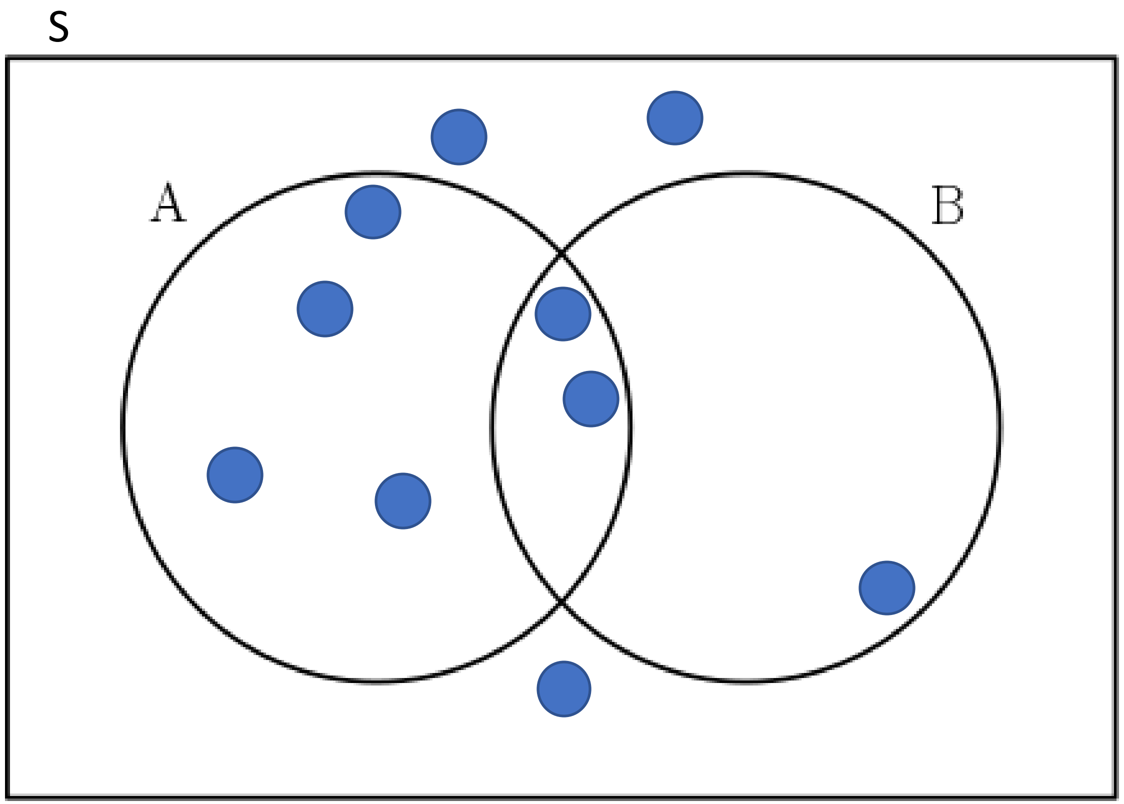
In this example, if ‘‘ is an event, then:

- P(A) = 6/10=3/5
- P(B) = 3/10
- P(S) = 10/10 = 1.0



Probability

- If '●' is an event, then
  - Intersection:
  - $P(A \cap B) = \frac{2}{10} = \frac{1}{5}$
  - Union:
  - $P(A \cup B) = \frac{7}{10}$
  - Negation:
  - $P\left((A \cup B)'\right) = \frac{3}{10}$
  - $P(A' \cap B') = \frac{3}{10}$



Axioms of Probability

1. Probability is bounded between 0 and 1.

$$0 \leq P(A) \leq 1$$

Note: "percent" literally means per one hundred.

2. Probability of the Sample Space is equal to 1.

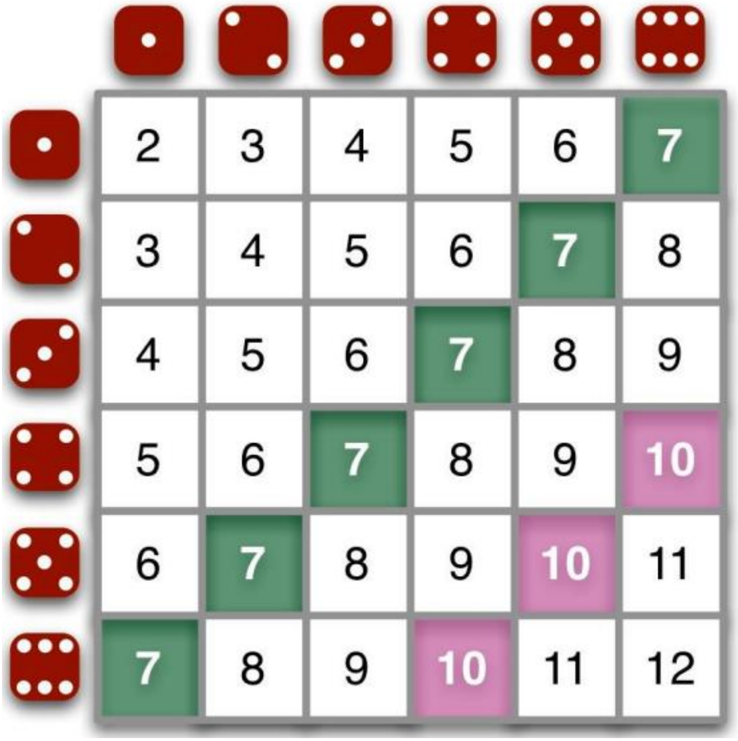
$$P(S) = 1$$

3. The probability of finite *mutually exclusive* events is the sum of their probabilities.

$$P(A \cup B) = P(A) + P(B) \text{ if A and B are Mutually Exclusive}$$

Quick Review – Probabilities

- What is the probability of rolling a 10?
- What is the probability of rolling an even #?
- What is the probability of doubles?

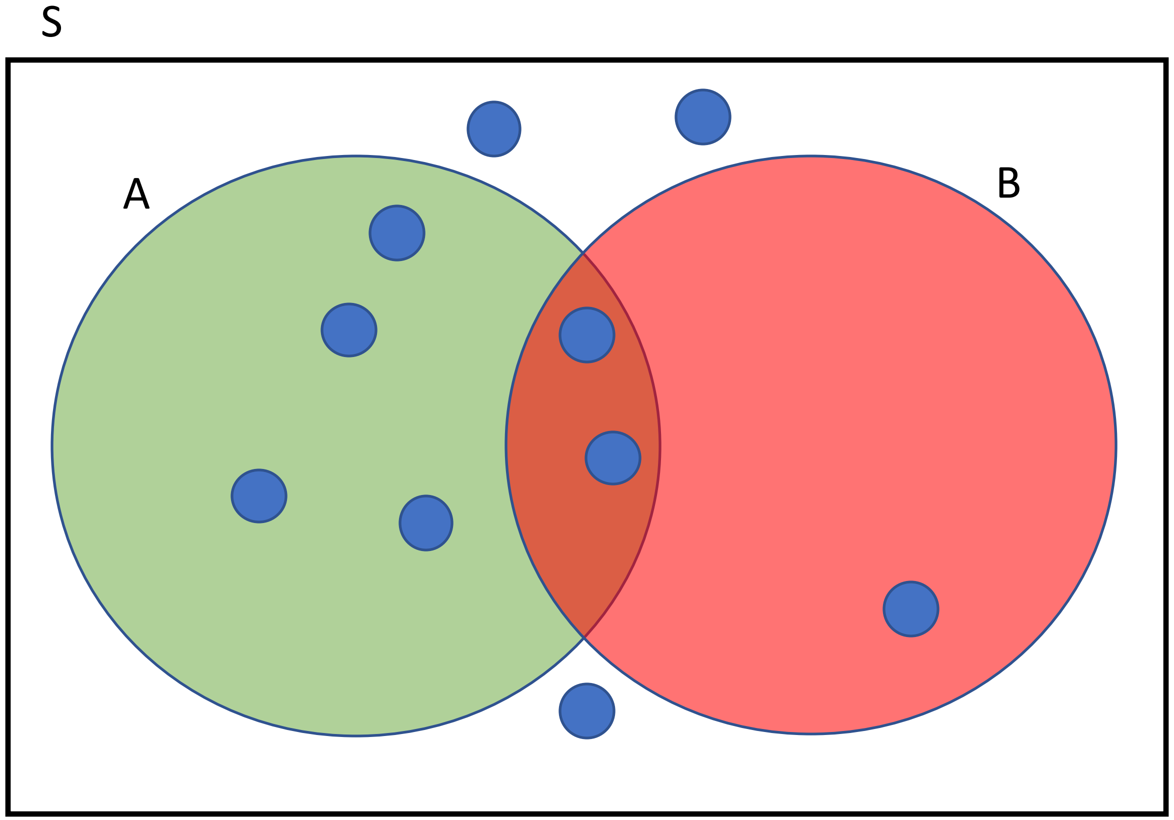


“At Least” or “At Most” Probabilities

- When dealing with problems that are “At Least” or “At Most”, we can think of what else can happen and do 1 minus that.
- Ex:
  - What is the probability that the sum is at least 3?
    - 1-P(sum = 2) = 1-(1/36) = 35/36.
  - If it is always 40% chance of rain each day, what is the probability that it rains on at most 6 days this week?
    - 1-P(rains all week) =1- (0.4)^7

Why is this false?

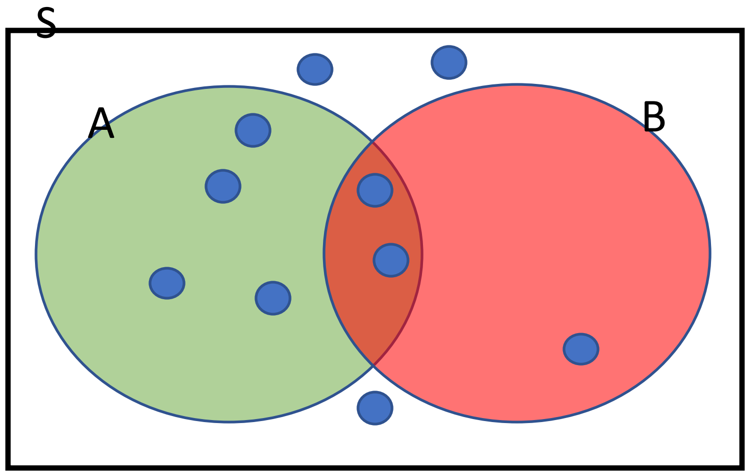
$$P(A \cup B) = P(A) + P(B)$$



Mutually Exclusive Events

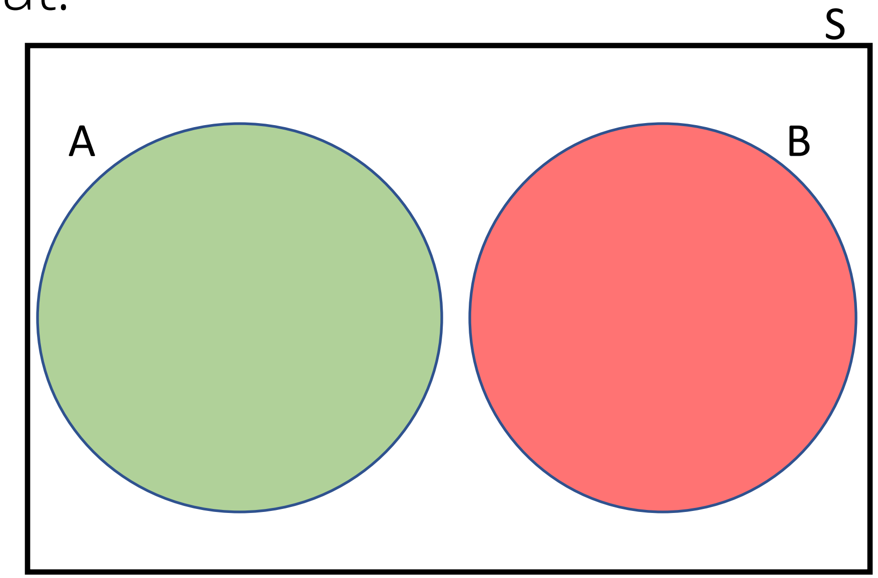
- In all cases, the probability of the union of A and B takes the form:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



- If A and B are *mutually exclusive* then means that:

$$P(A \cap B) = 0$$
$$P(A \cup B) = P(A) + P(B)$$



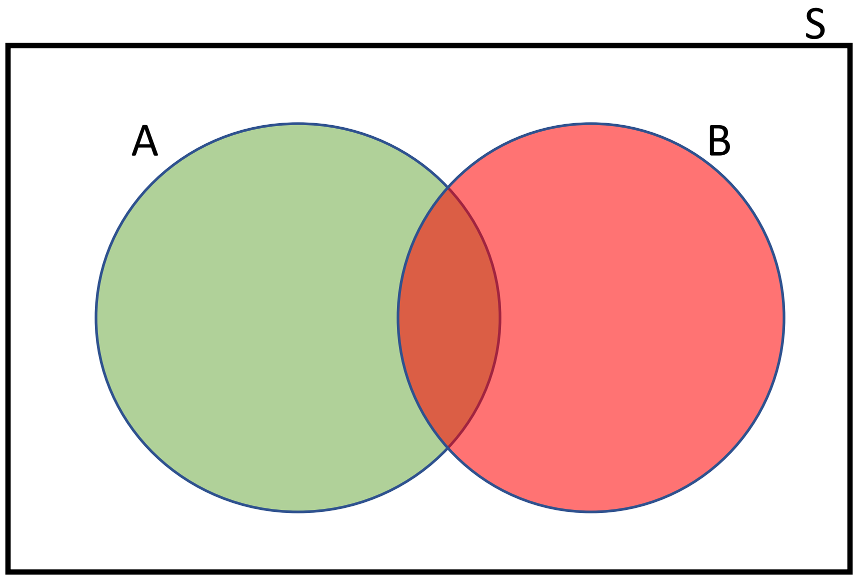
Conditional Probability

> The Probability of A *given* B is written:

$$P(A|B)$$

> And is equal to:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ compare to: } P(E) = \frac{P(E)}{P(S)}$$



Independence VS Mutually Exclusive

> These concepts are not similar AT ALL, and in fact, are nearly opposite ideas.

> If A is M.E. of B, then $P(A|B) = 0$

The event, B, has a HUGE effect on the occurrence of event A.

> If A is independent of B, then $P(A|B) = P(A)$

Example: The probability the sidewalk is wet given it is raining is very high, but the probability that it is raining given the sidewalk is wet could be quite low, if I run my sprinklers often.

ODDS

- Odds are probabilities expressed as
  - (Count of events in favor):(Count of events not in favor), with the reduced form of probability in fractional form.

$$P(A) = \frac{n}{m} = \frac{n}{n + (m - n)}$$

This implies the odds are n:(m-n)

Example:

- If P(A)=50%, then the odds are 1:1 or “one to one”.
- If P(A)=1/3, then the odds are 1:2 or “one to two”.
- If P(A)=5/6, then the odds are 5:1. ‘five to one’.
- If the odds are 3:20, then P(A)=3/23
- A straight up sports bet in Vegas has odds 1:1 (50%) (but pays 0.95Xbet.)

Independent Events

> Event A is *independent* of event B if and ONLY if:

$$P(A|B) = P(A)$$

> A being independent of B does NOT imply B is independent of A

$$P(A|B) = P(A) \not\Rightarrow P(B|A) = P(B)$$

$$P(A|B) = P(A) = \frac{P(A \cap B)}{P(B)}$$

$$P(A)P(B) = P(A \cap B)$$

<u>Used in a sentence:</u> The event that my boss takes a vacation has an impact on when I take vacation. But when I take vacation has no impact on when my boss takes a vacation.

discrete vs continuous random variables

	discrete	continuous
sample space	counted	measured
math	summation	integration
$P(A)$	probability mass function $P(A) = \sum_{a \in A} P(A = a)$	probability density function $P(A) = \int_{a \in A} P(A)$
examples	dice roll, number of arrivals, etc.	price, height, distance
visualizations	bar plot, n-way tables, …	histograms, box plots, …

Consider the probabilities of all potential sums of the two die:

P(2)=1/36

P(3)=2/36

P(4)=3/36

P(5)=4/36

P(6)=5/36

P(7)=6/36

P(8)=5/36

P(9)=4/36

P(10)=3/36

P(11)=2/36

P(12)=1/36

If we consider all the event possibilities together, this is called a "**<u>Probability Distribution</u>**".

Discrete Distributions

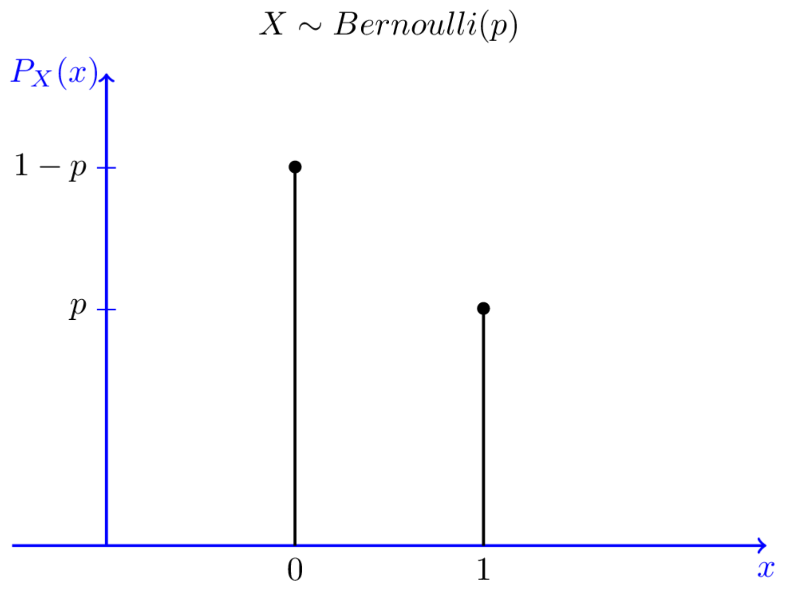
> Properties of Discrete Distributions:
  - Sum of all events MUST be equal to 1.
  - Probability of event, E, is equal to the value of the distribution at that point.
  - No negative values or values greater than 1.

Discrete Distribution: Bernoulli Distribution (1 coin flip)

> Bernoulli (1 event, example: 1 coin flip)

$$P(x) = \begin{cases} p \; if \; x = 1 \\ (1-p) if \; x = 0 \end{cases}$$

$$P(x) = p^x(1-p)^{(1-x)} \; for \; x \in \{0,1\}$$



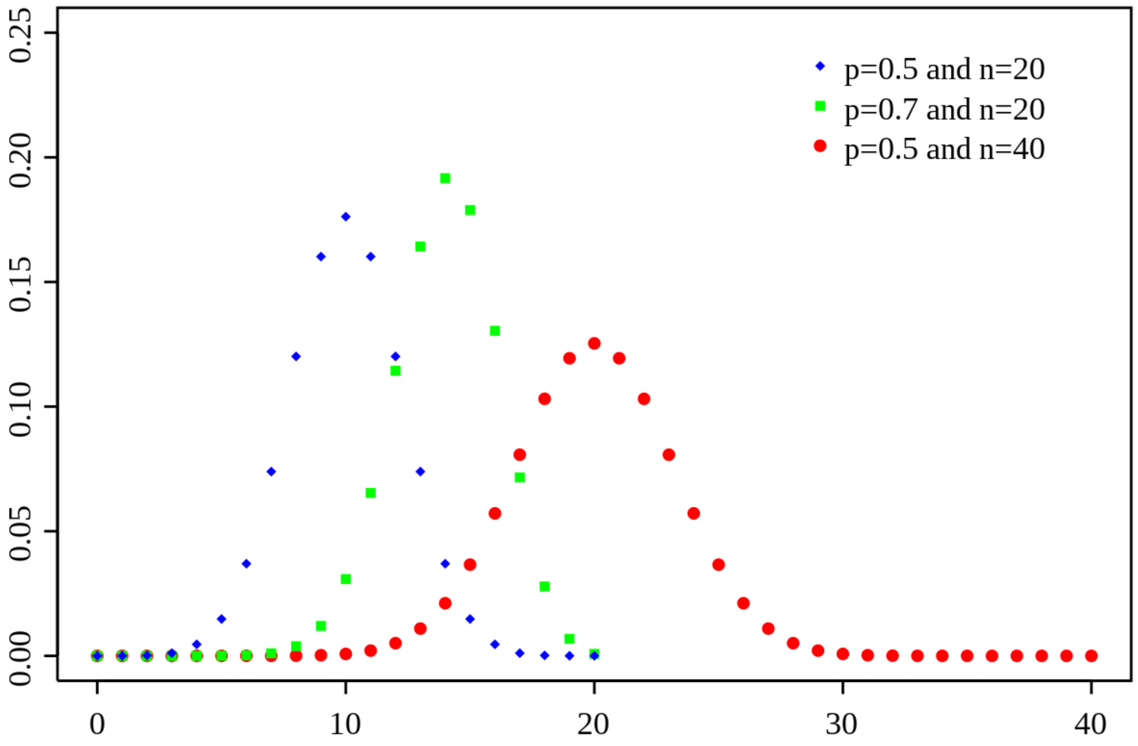
> Mean = p

> Variance = p(1-p)

Discrete Distribution: Binomial Distribution (Multiple Coin Flips)

> Multiple independent events = Product of Bernoilli probabilities.

$$P(x|N,p) = \binom{N}{x} p^x (1-p)^{(N-x)}$$



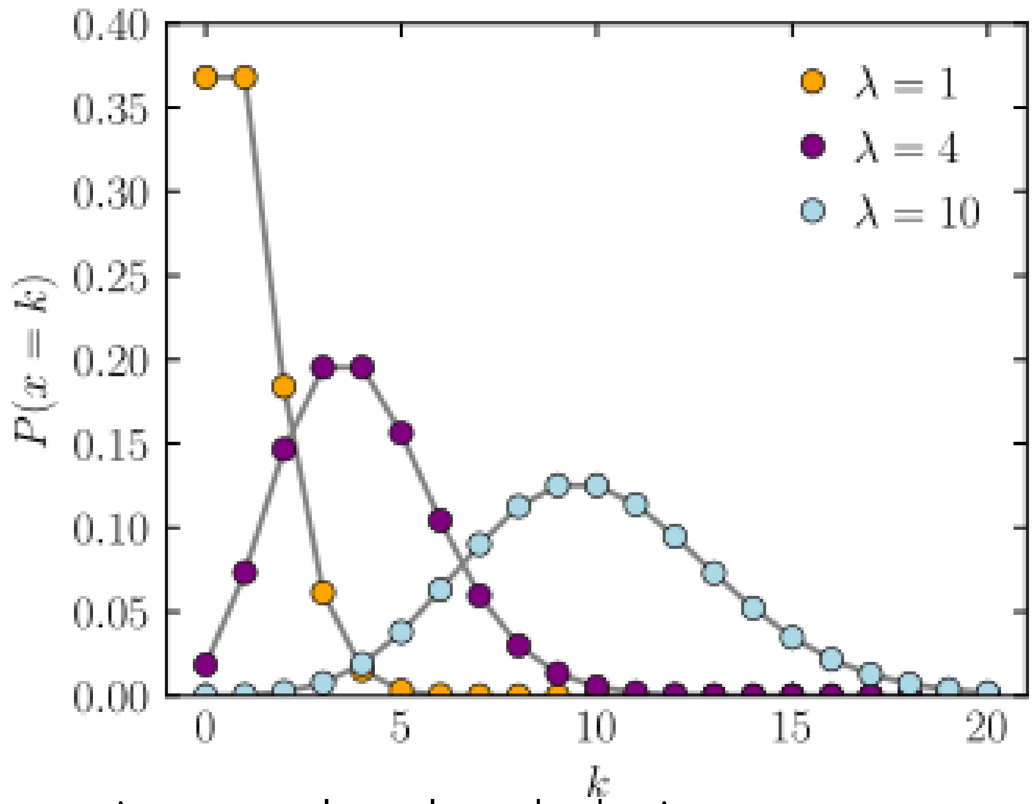
> Mean = np
> Variance = np(1-p)

> Note: for larger n, we approximate this by a normal distribution.

Discrete Distribution: Poisson Distribution (Counting)

> Count number of events in a time span.

$$P(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$$



> Mean = λ

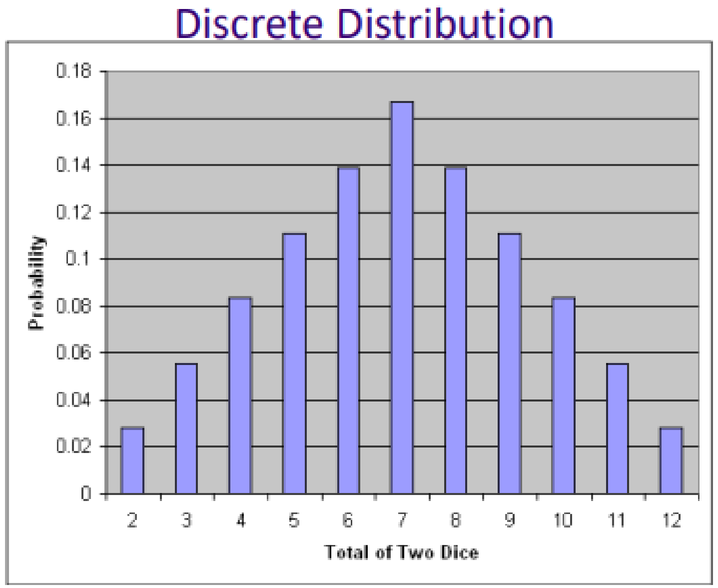
> Variance = λ

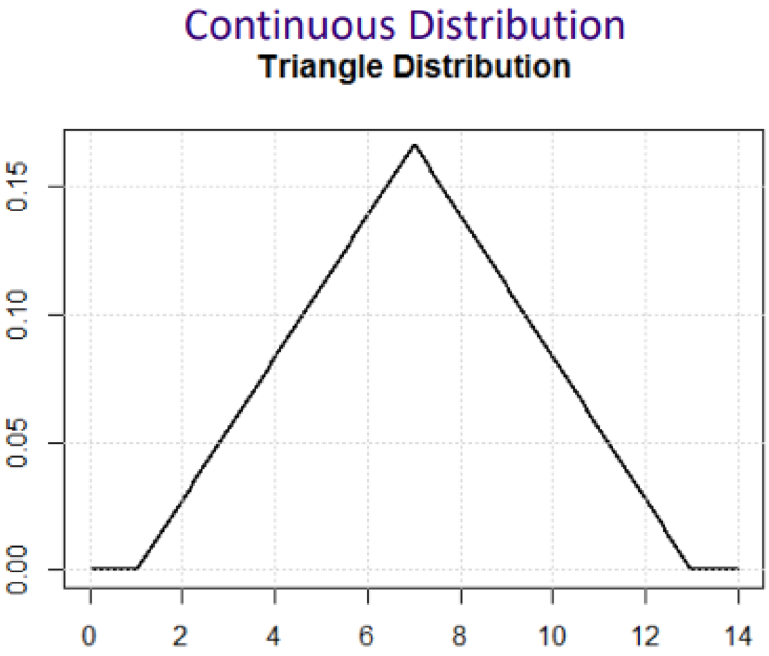
> Interpretation: The rate of occurrence of an event is equal to lambda in a finite period of time

> Can also approximate by a normal at large lambda.

Continuous Distributions

> Properties of Continuous Distributions:
  - AREA under the curve MUST be equal to 1.
  - Probability of event, E, is equal to the AREA under the curve between two points.
  - No negative values or values greater than 1.
  - Probability of a single, exact value is 0.

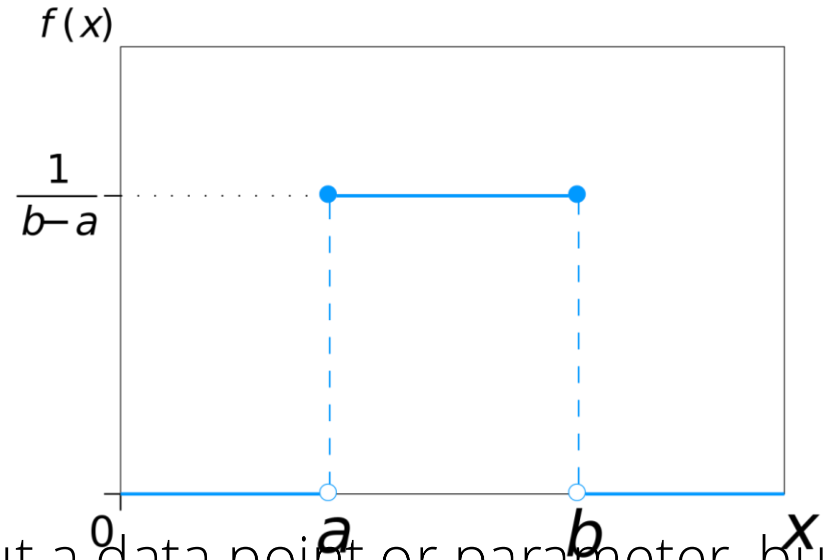




Continuous Distribution: Uniform Distribution (Equal Probability)

- Most commonly *used* in machine learning. (not most commonly occurring in data)
- Uninteresting distribution, all events are equal.
- Flat & Bounded.

$$f(x) = \begin{cases} \dfrac{1}{(b-a)} & if\ a \le x \le b \\ 0 & x < a\ or\ x > b \end{cases}$$



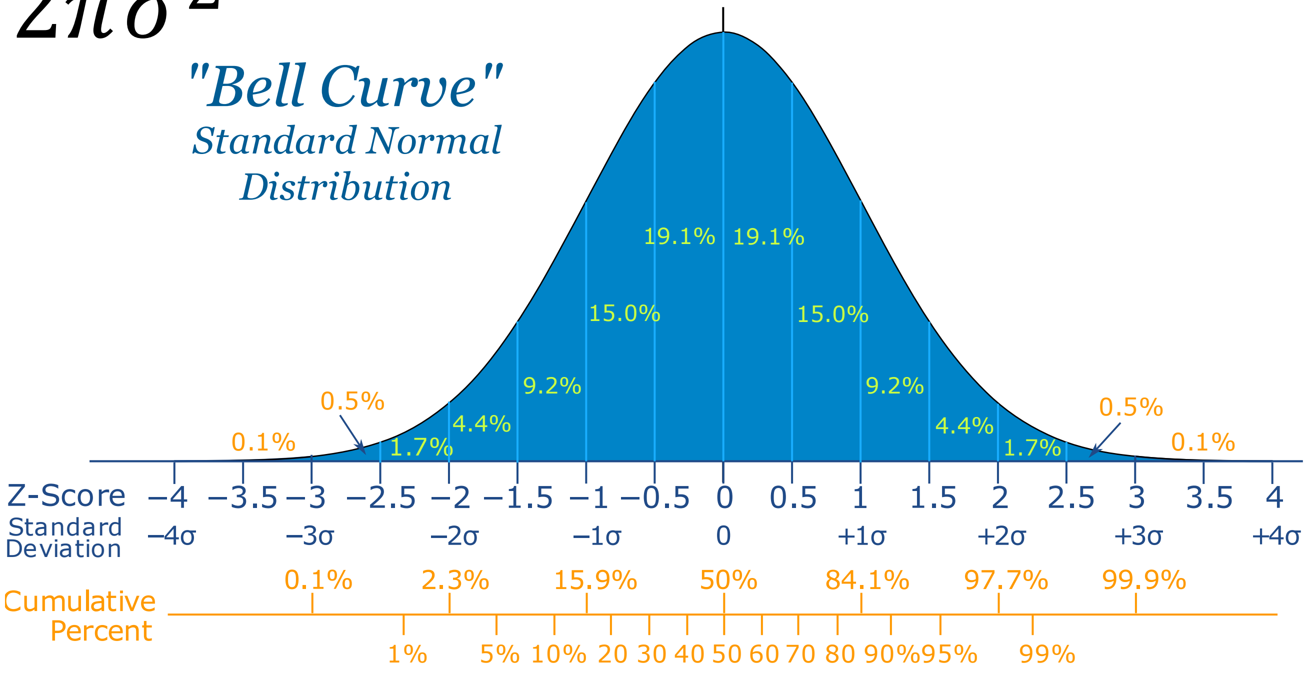
- Mean = $\frac{(a+b)}{2}$
- Variance = $(\frac{1}{12})(b-a)^2$
- Very useful when we *need* to put in information about a data point or parameter, but we don't know much about it.

Continuous Distribution: Normal Distribution (Gaussian/Bell Curve)

> Most common & naturally occurring in data/nature.

> Defined by a mean and variance only.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

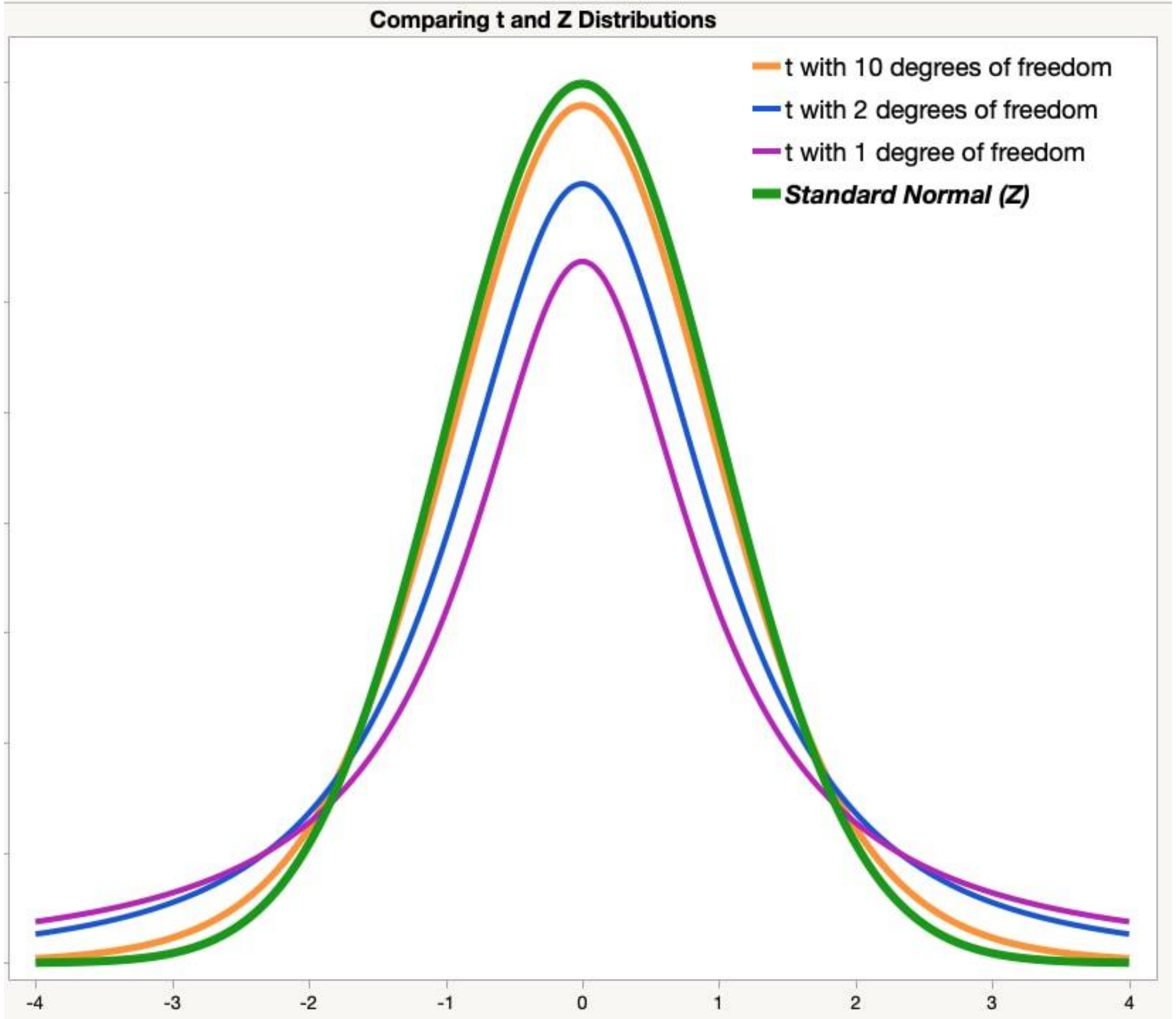


> Has very nice properties.

> Testing data for normality is VERY important.

Continuous Distribution: Student's T Distribution

- Important for investigating / testing smaller data sample sizes.
- Used specifically for:
  - Testing of a mean-value when st.dev. is unknown.
  - Testing difference between two distribution means.
- Looks very similar to the Normal distribution, but has longer tails.



Distribution Transformations

- The purpose of transforming a variable is to make it easier to distinguish between values (graphically, statistically, ML models…)
  - Most commonly, we are looking to transform it to a normal distribution.
- Useful transformations are generally MONOTONIC
  - Functions used are never changing directions.
- Common Transformations to consider:
  - Log based
    - Log(x), log(x+1), log(x-min(x)+1)
  - N-th root based (of positive #s)
    - X^(1/n)
  - Any combination you can think of- but keep it monotonic.
- How to determine? We will cover normality tests in a later class.

notebook time

we return to the lecture later

the Monte Hall game

- you are presented with three doors: two have goats behind, one has a car behind
- the host knows which door has a car behind it, but you don't
- you select a door, then host would open one of the two other doors to reveal a goat
- you have a choice between sticking with the door you selected, or switching to the other door the host didn't open
- does it matter if you switch or not?

Monty Hall Problem

> Start with 3 doors. One prize behind unknown door. Pick a door.

> Host reveals a separate door with no prize.

> Then contestant can switch. Should they?

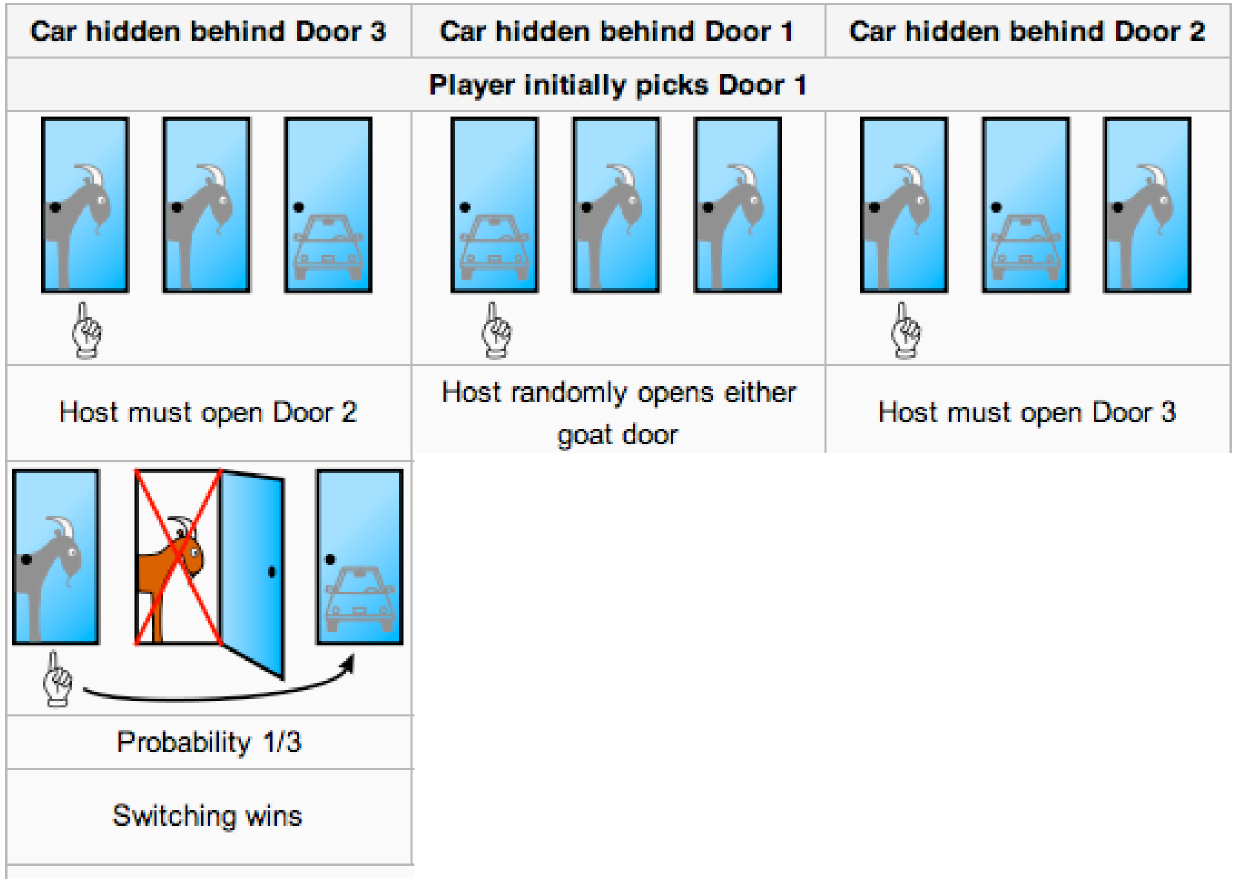
Let’s consider this situation first

Monty Hall Problem

> Start with 3 doors. One prize behind unknown door. Pick a door.

> Host reveals a separate door with no prize.

> Then contestant can switch. Should they?

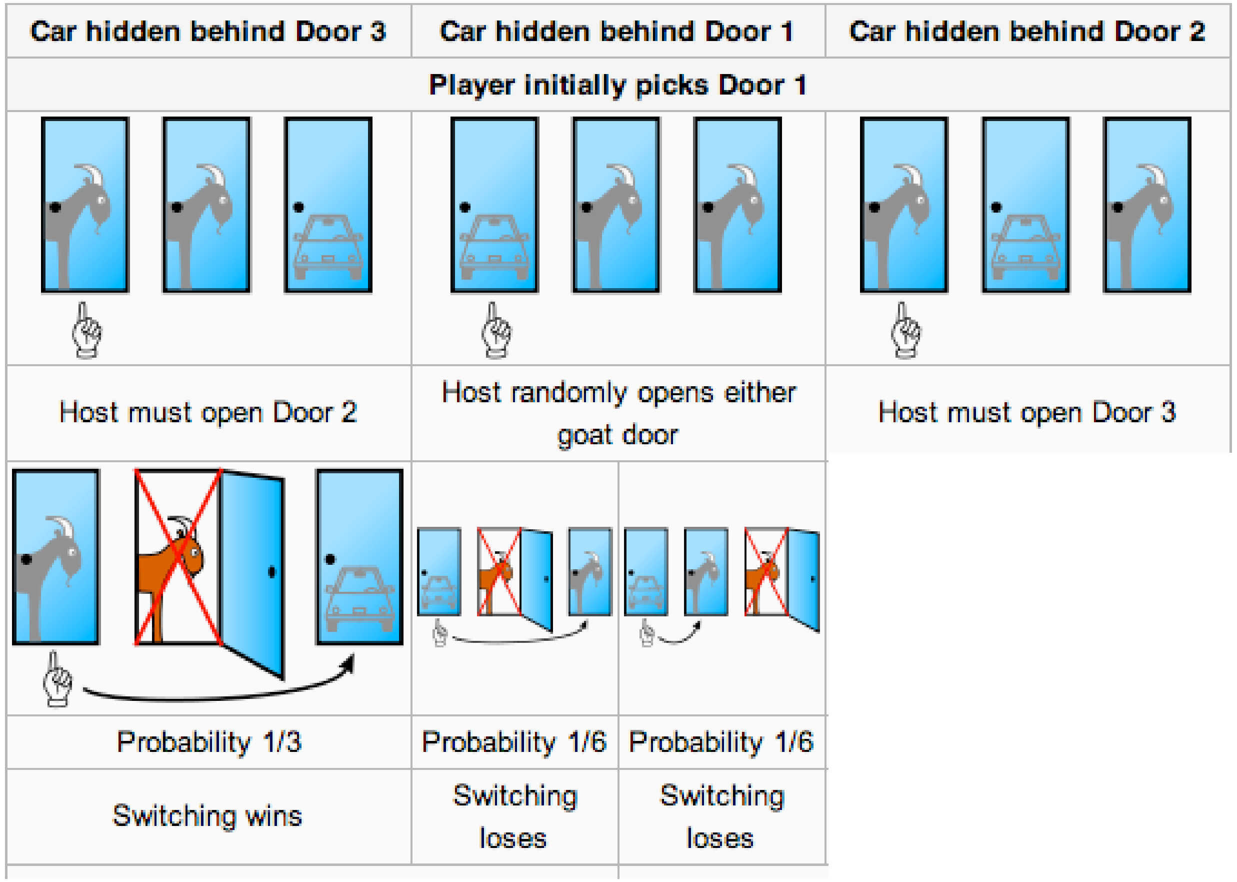


Monty Hall Problem

> Start with 3 doors. One prize behind unknown door. Pick a door.

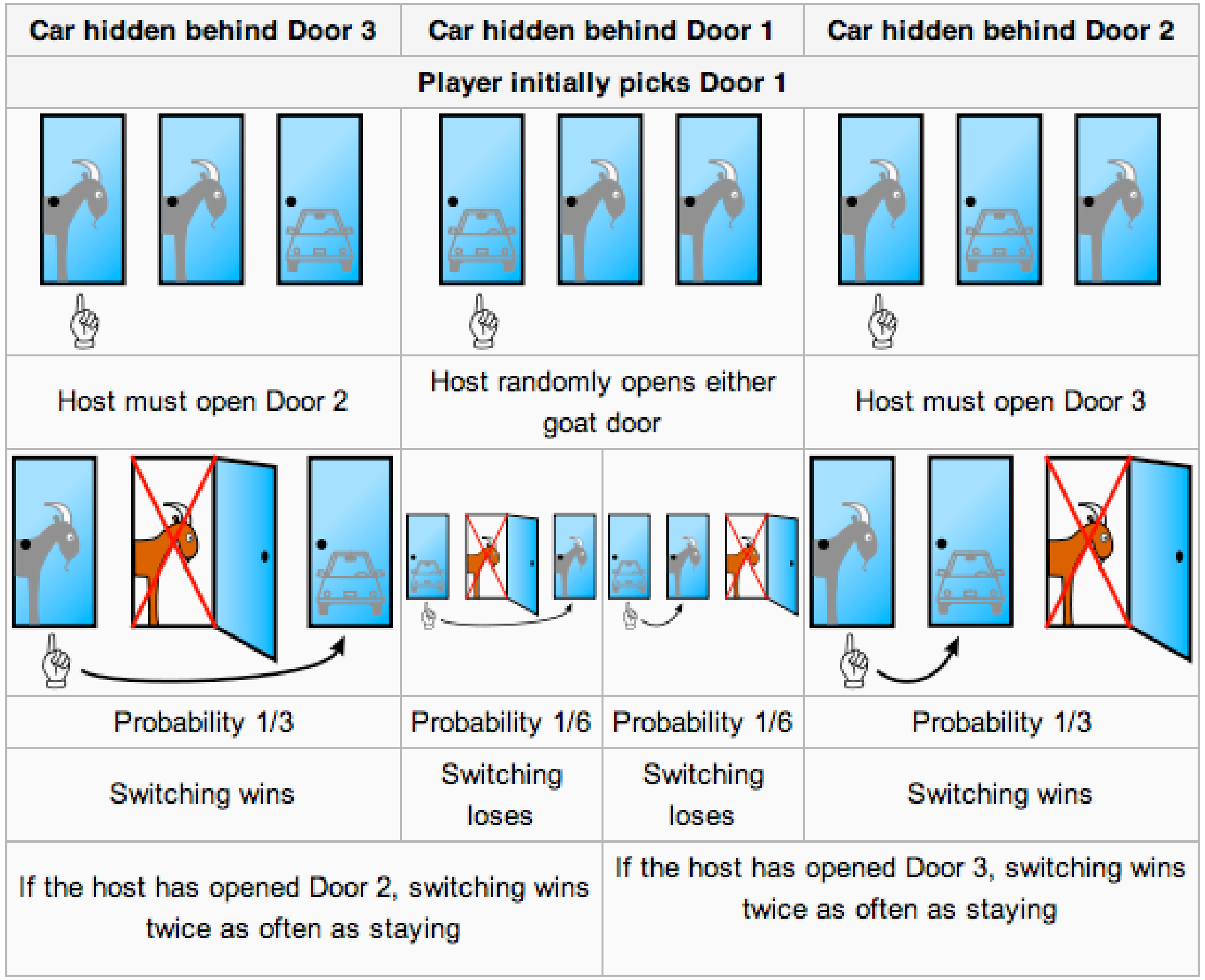
> Host reveals a separate door with no prize.

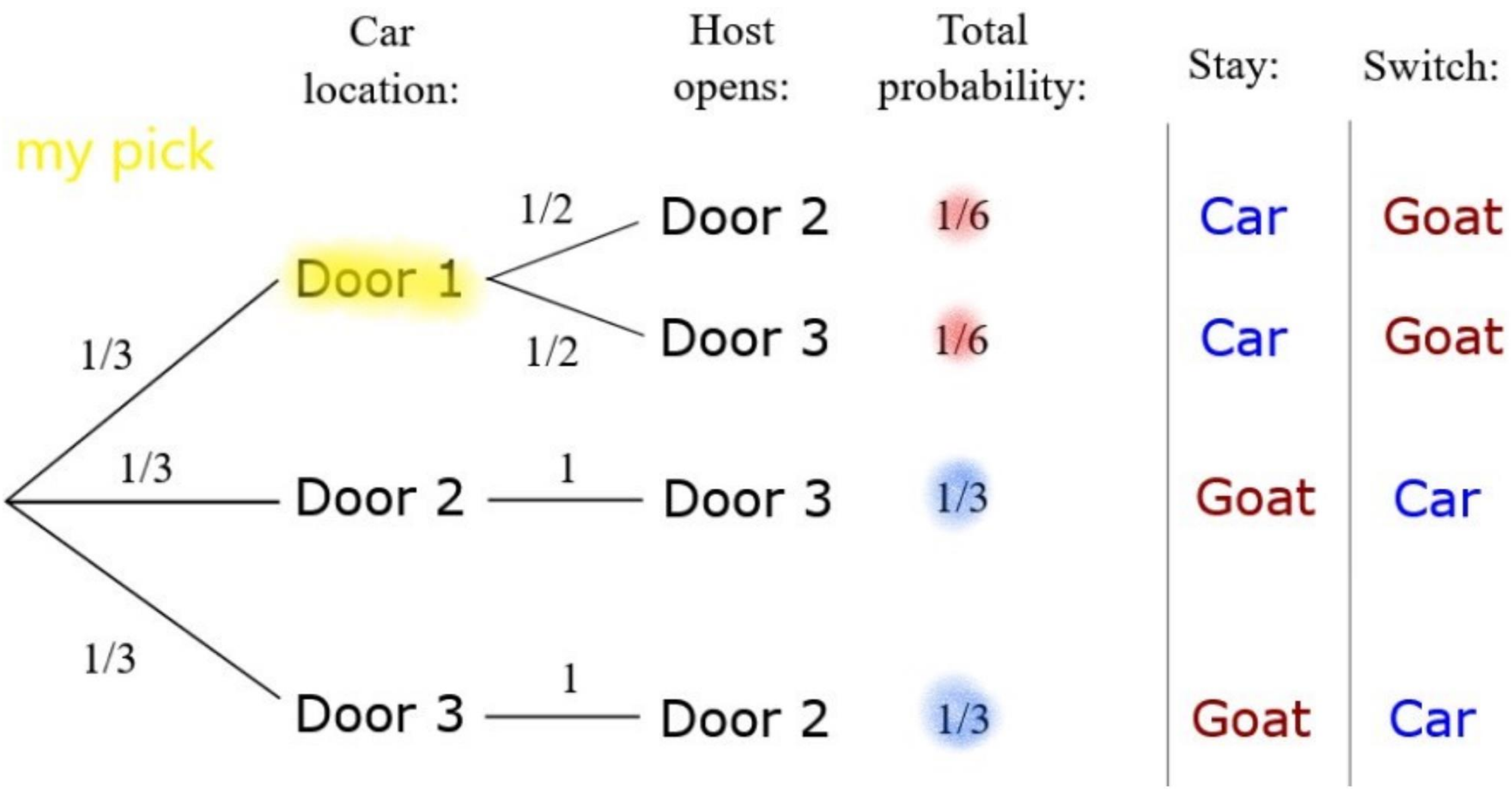
> Then contestant can switch. Should they?



Monty Hall Problem

> Start with 3 doors. One prize behind unknown door. Pick a door.
> Host reveals a separate door with no prize.
> Then contestant can switch. Should they?




Car location:
Host opens:
Total probability:
Stay:
Switch:
my pick
1/3
Door 1
1/2
Door 2
1/6
Car
Goat
1/2
Door 3
1/6
Car
Goat
1/3
Door 2
1
Door 3
1/3
Goat
Car
1/3
Door 3
1
Door 2
1/3
Goat
Car

- let $C = \{1, 2, 3\}$ represent where the car is
- let $H = \{1, 2, 3\}$ represent the door opened by host
- **assume by symmetry that we pick door 1**, and compute conditionals probabilities:
  - note that $P(H = 1|C = 1) = 0$, which leaves us with $P(H = 2|C = 1) = 0.5$ and $P(H = 3|C = 1) = 0.5$
  - note that $P(H = 1|C = 2) = P(H = 1|C = 3) = 0$ because Montey won't open a door containing the car, leaving us with $P(H = 3|C = 2) = 1$ and $P(H = 2|C = 3) = 1$ respectively
- now we compute the joint probabilities $P(H \text{ and } C) = P(H|C)P(C)$
- finally, we add up joint probabilities for goat and car under switch scenario

Dealing with Missing Data (Intro)

- Reasons for missing data
  - Recording failure (mechanical/software failures)
  - Reporting failure (human decisions)
  - Translation failure (data parsing / data transferring errors)
- Many shapes and types of missing data:
  - Block missing – Physical shapes of data missing
    - Examples: Time series- collection was down for some time.
  - Regular missing
    - Rule based missing: Temperature gauge doesn't work below freezing.
  - Random missing
    - Probability of missing data point equal across all data points.
  - Sparsity
    - Very low observations. Hard to observe data (endangered species, rare events)
- NOTE: Outliers *may* be treated as missing data. Depending on reasons.

Dealing with Missing Data

Type	Benefits	Disadvantages	Notes
Drop Missing	Speed	Data Loss	
Mean/Median/Mode Fill	No Data Loss	Variation Reduction	
X ~ F(independents)	More accurate No Data Loss	Slower	Needs most columns to be filled out Harder on truly independent data
KNN	More accurate No Data Loss	Slower Dependent on a Distance Function	
X~F(y,independents)	Very Accurate No Data Loss	Slower Need Y	Can ONLY do this on a TRAINING dataset.

Dealing with Missing Data in the Real World: Using Outside or New Data Sources as Supplements

- Don’t forget to explore outside or new data sources to help with missing data.
- With the advent of free public data and bigger data sources, this is gaining popularity as a tool for imputation.
- This includes using unstructured text as a major source of data.
- Will eventually include asking Large-Language-Models to generate missing data.
- Ex
  - Caesar’s uses public reviews on websites to mine for customer sentiment about hotel rooms.
  - Zillow uses text descriptions of properties to fill in missing data about homes (# BDs, # BTHMs, sq footage, etc)
  - Subject to human stupidity.:

> Yelp Rating for Circus-Circus: 2/5
> Text Description: “My son and I stayed here. The service and room was great, but it turns out my son is deathly afraid of clowns.

Breakout Discussion

- Below are some example situations for missing data.
  - How would / could we address these? (Both statistically AND business wise)

Situations
An intern at your company accidentally deleted 10 rows in a large user-events table.
A set of door-to-door poll workers in a county were all sick over the same time period. (They are interviewing residents about news topic sentiments).
A malicious SQL injection attack dropped data from user-sign in logging records. Only from users who have a email that starts with the letter ‘A’ or ‘a’.
The IoT coffee maker company you work for has a defect in the devices – it disconnects from wifi when boiling water is spilled in a certain spot.
A rental car company cannot collect GPS car locations during severe weather.

the end