NNSE 784
Advanced Analytics Methods

Instructor: F Doyle (CESTM L210)

MW 4:30 – 5:50, NFN 203

Slide Set #4
Basic Probability Review

Outline for lecture

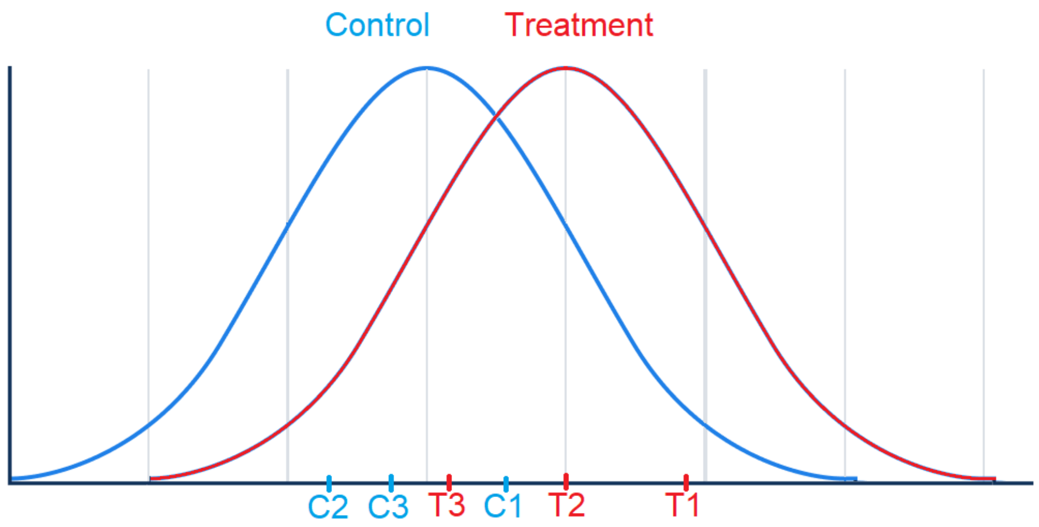
- How does probability fit in with statistics and data science?
- What is the study of probability?
- A tangent note on complexity
- Basics of probability calculations
- Bayes’ Theorem
- The goal is to convey an understanding of the depth of complexity of probability and hopefully remove some common misconceptions so that you have a better intuitive feel for the discipline. We don’t have time to explore the underlying math in a fully rigorous manner.

How does probability fit into statistics?

- We've discussed descriptive statistics and looked at frequency distributions.
- Probability can be describe as a study of "likelihood". How likely is something to happen?
- I alluded to the notion that the more prevalent certain values are in the histogram, the more "likely" you are to see them (e.g., a 5'6" tall vs a 3' or 7' tall person)
- At this point in your studies/career you have likely been exposed to the term "p-value", do you really know what it represents?
- The specifics of a given p-value depend on the statistical test used, but generally speaking it is the **probability** of having observed specified sample values if a hypothesis that we are trying to disprove is actually true. This is why smaller p-values are deemed "better" as supporting indicators of experimental conclusions.

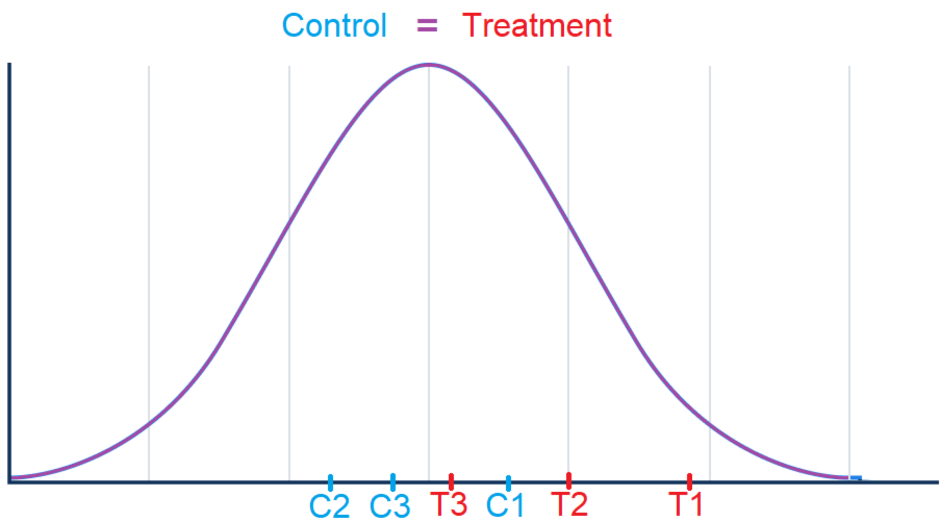
Hypothetical Example

You have performed three replicates of some experimental treatment and collected three values each for treatments and untreated controls.
The goal of the treatment was to have an increase on the measured value.
Understand that the values collected for each category are simply samples of a larger population of all potential values they could have.
The p-value is the probability (from 0 to 1) that the hypothesis you are attempting to invalidate is the correct one.
After applying a test statistic, you calculate a p-value of .04, this means that there is a 4% probability that you could have gathered the sample values that you did if the null hypothesis were true.
Depending on what you have stated as a **significance threshold**, this may be enough for you to reject the null hypothesis and conclude your treatment had an effect.



What you think (and more dangerously…"hope") you're seeing based on the data.

VS



What you might actually be seeing and what you are trying to disprove ("null hypothesis")

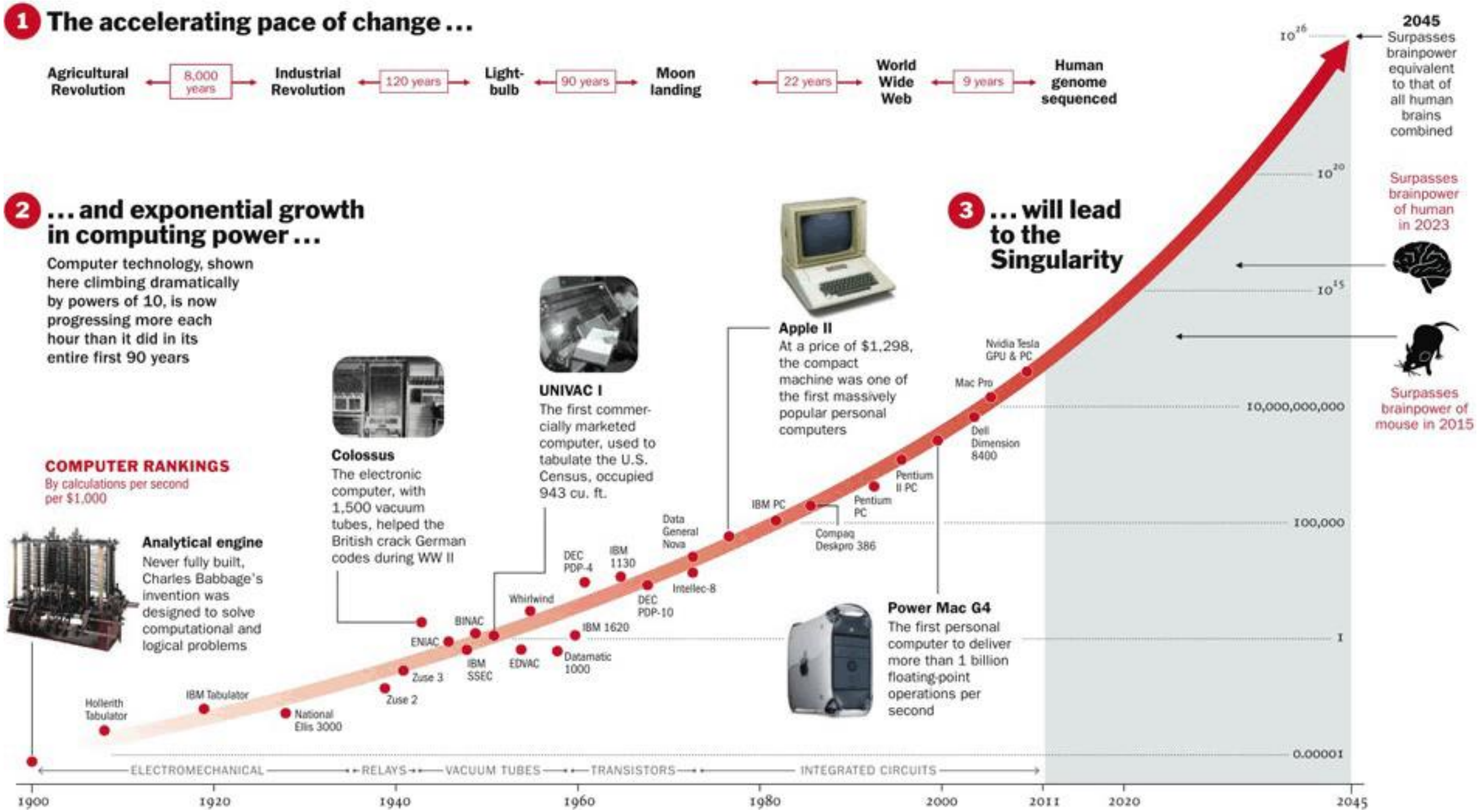
Probability is Core to Inferential Statistics

- We saw in the previous slide that the p-value for a statistical test was actually a probability
- Before we look at how different statistical tests calculate probabilities, we need to get a feel for how general probabilities are calculated
- Probabilities for any given outcomes are generally defined based on an understanding of how many other outcomes are also possible
- Categorizing possible outcomes often involves math pertaining to combinations and permutations, both of which often lead to magnitudes of scope that are not immediately intuitive

First… a word about complexity

- It has been said that the human mind is formed to think, at least intuitively, in a linear manner (based on how we evolved)
- Exponential (or factorial) growth is something we can intellectually grasp but it is not something we have a “gut feeling” for
- Probability is often calculated based on combinations and permutations
- Combinations & permutations are inherently factorial/exponential in scope
- The goal here is not to make you experts in this type of math, but to help you avoid deceiving yourself with regard to intuitive feelings about complexity and probability

Exponential Changes in Technology



https://www.singularity.com/ <- I am not promoting this idea, but it is interesting if you haven't come across it before

Reminder About Factorials

Factorial

$$n! = n * (n-1) * (n-2) * (n-3) * \cdots * 3 * 2 * 1$$

0! = 1

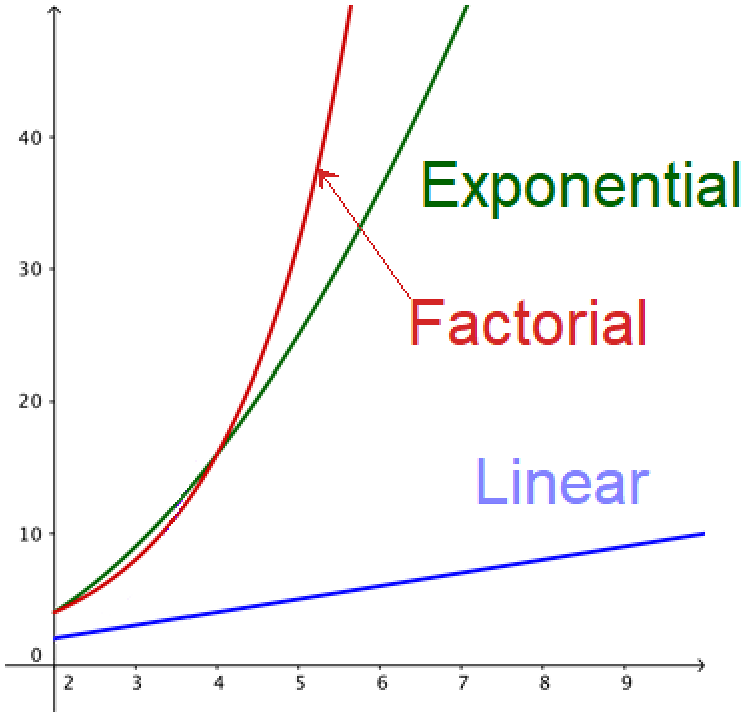
1! = 1

2! = 2 x 1 = 2

3! = 3 x 2 x 1 = 6

4! = 4 x 3 x 2 x 1 = 24

Types of Complexity

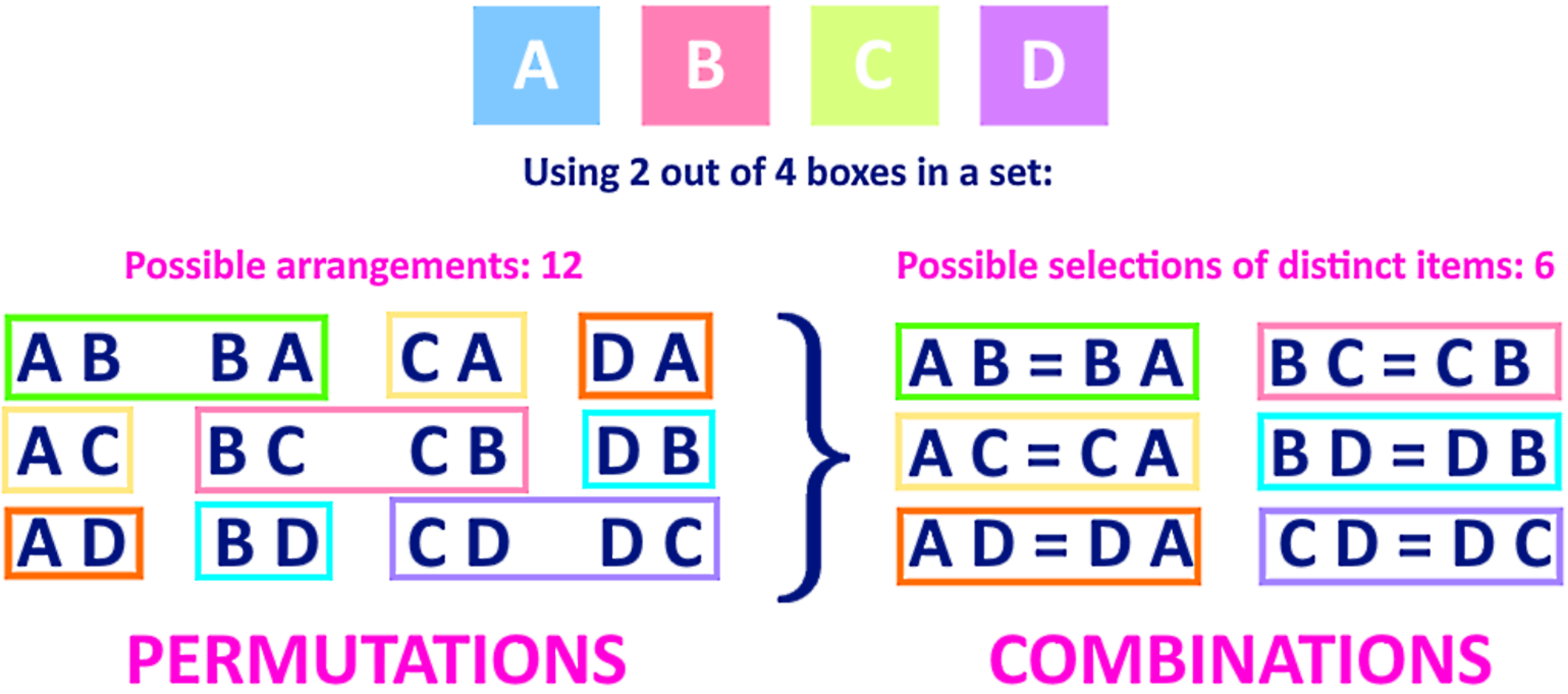


Deceptively Complex

`00000000 00000000 00000000 00000000 01010101 00001111 11100000 00000001`

- This is an example of an 8 byte binary number. Certain computing systems and programming languages, such as Java use this size variable (specifically a "long int" in Java) to store datetime values
- How much information can something like this actually represent? It's only binary…right?
- A 64 bit value can represent any specific millisecond over a period of more than 500 million years.
- 2^{64} = 18,446,744,073,709,551,616 (18 quintillion) permutations

Permutations vs Combinations



Order matters!

Order is not important.

Calculating Permutations and Combinations

(of ***n*** elements taken ***r*** times)

Permutations

Without Replacement

$${}_{n}P_{r} = \frac{n!}{(n - r)!}$$

With Replacement

$${}_{n}P_{r} = nr$$

Combinations

Without Replacement

$${}_{n}C_{r} = \frac{n!}{r! * (n - r)!}$$

With Replacement

$${}_{n}C_{r} = \frac{(r + n - 1)!}{r! * (n - 1)!}$$

Do not memorize these formulas!

On To Probability

Elementary Properties of Probability

- Given some process (e.g. an experiment), with *n* mutually exclusive outcomes (called "events"), $E_1, E_2 \ldots E_n$, the probability of any event E_i must have a non-negative probability

$$P(E_i) \geq 0$$

- The sum of the probabilities of all mutually exclusive outcomes is equal to 1

$$P(E_1) + P(E_2) + \ldots + P(E_n) = 1$$

- For any two mutually exclusive events, E_i and E_j, the probability of the occurrence of either E_i or E_j is equal to the sum of their individual probabilities.

$$P(E_i + E_j) = P(E_i) + P(E_j)$$

Read as "probability of E_i or E_j"

A priori (analyzing concepts)

“The probability of E”

Classical Probability

$$P(E) = \frac{m}{N}$$

If an event can occur in N mutually exclusive and equally likely ways, and if m of these possess a trait, E, the probability of the occurrence of E is equal to m/N.

Measured ***estimate!***

Relative Frequency Probability

$$P(E) = \frac{m}{n}$$

If some process is repeated some large number of times n, and if some event E occurs m times, the relative frequency of occurrence of E, m/n will be approximately equal to the probability of E.

Probability

- Probabilities are denoted as continuous values from 0 to 1
- A probability of 0 for some outcome means that it is impossible
- A probability of 1 for some specific outcome means that it is certain
- The sum of probabilities for all possible outcomes of a given action must equal 1

For the next examples…

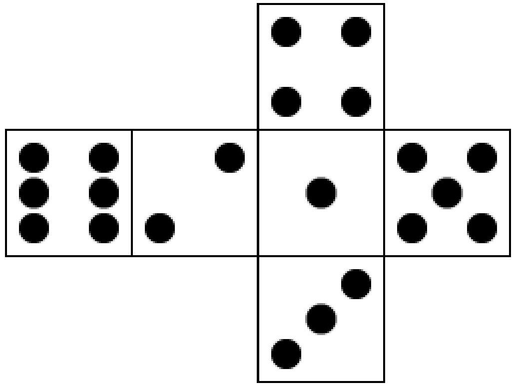
Coin

2 **equally** possible outcomes
(Head [H] or Tail[T])

P(x=H) = P(x=T) = $\frac{1}{2} = \ .5 = 50\%$

Die

6 **equally** possible outcomes
(1,2,3,4,5 or 6)



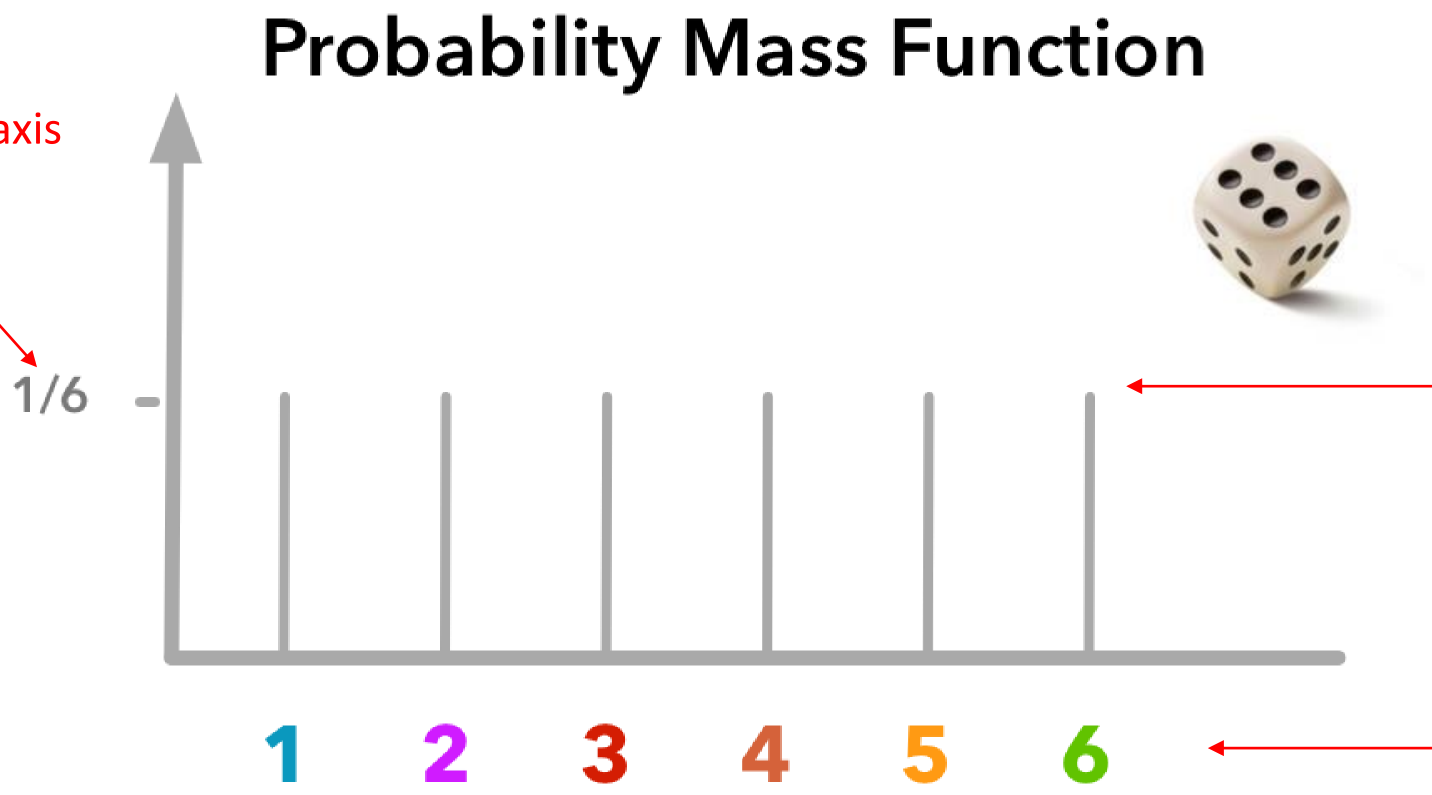
P(y=1) = P(y=2) = P(y=3) = P(y=4) = P(y=5) = P(y=6) = $\frac{1}{6}\ {\sim}.1666 = 16.6\%$

Probability Distribution Types

- **Probability Mass Function** - for discrete variables (e.g., die roll results)

- **Probability Density Function** – for continuous variables (e.g., height of a randomly chosen person)

Probability on y axis

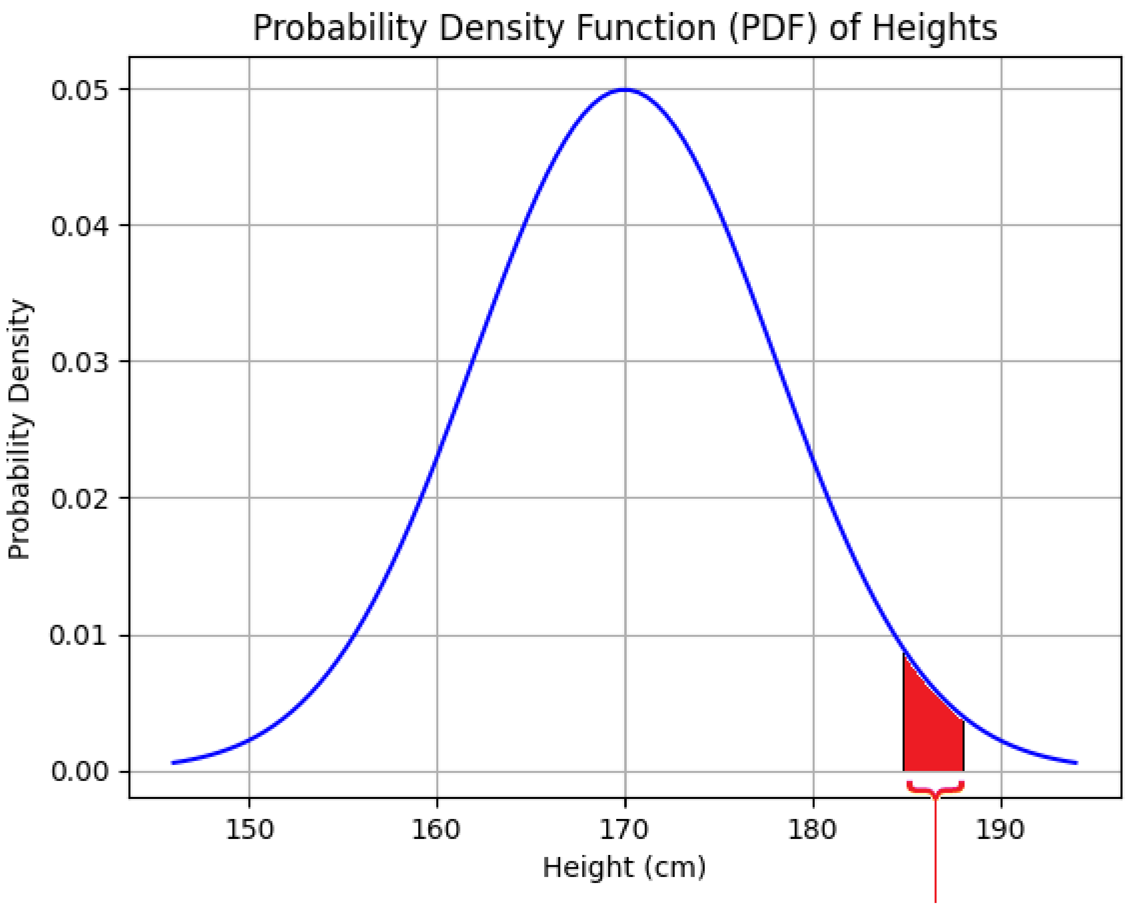


Example of a <u>uniform</u> distribution

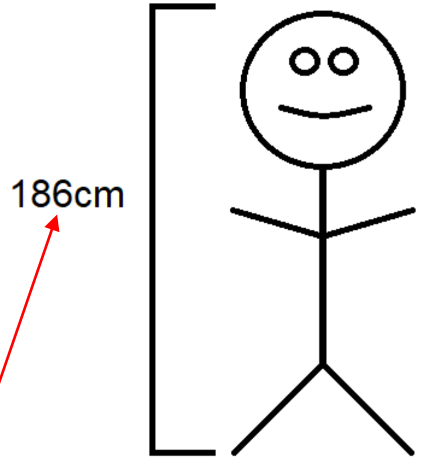
Possible outcomes on x axis

Remember each outcome had same probability: $P(y=1) = P(y=2) = P(y=3) = P(y=4) = P(y=5) = P(y=6) = \frac{1}{6}$

Probability Density Function



Y-axis is often unlabeled with regard to units. The y units equate to the "**probability per unit of the random variable**" on x-axis.



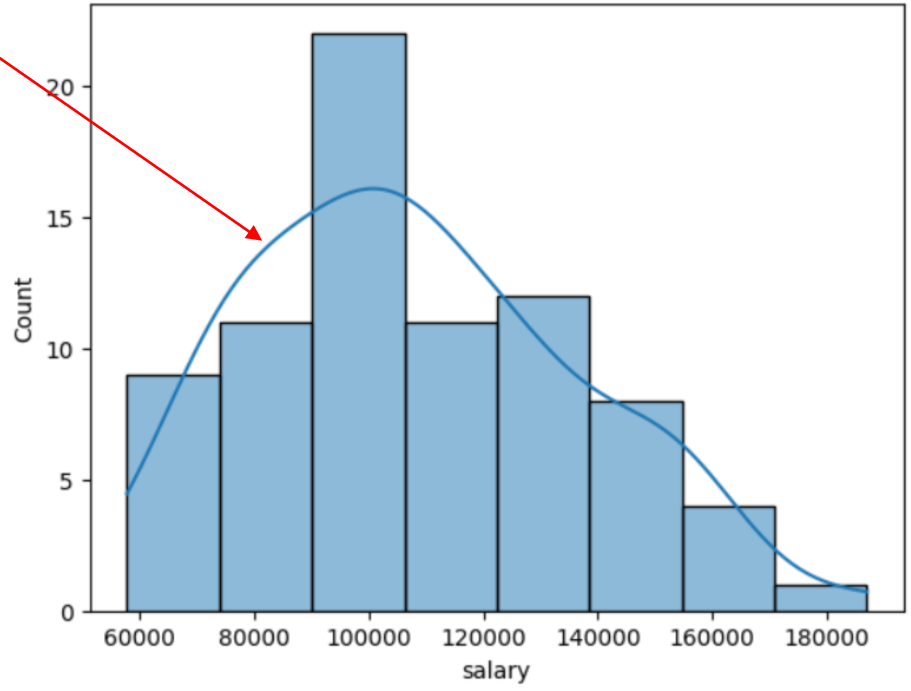
The probability for some interval of x is the area under the curve at that interval

Probability Can Be Directly Derived From A Histogram

The “Kernel Density Estimate” (or KDE) is an estimation of the probability density function for a given histogram.

```
import pandas as pd
import seaborn as sns
df = pd.read_csv("./Salaries.csv")
sns.histplot(df['salary'], kde=True)
```

```
<AxesSubplot:xlabel='salary', ylabel='Count'>
```



Marginal & Joint Probability (Independent Events)

Throwing a die

Joint Probabilities

Tossing a coin

	1	2	3	4	5	6	
Heads	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$
Tails	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$
	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

Marginal probabilities (probability of a simple event irrespective of any other variable outcomes)

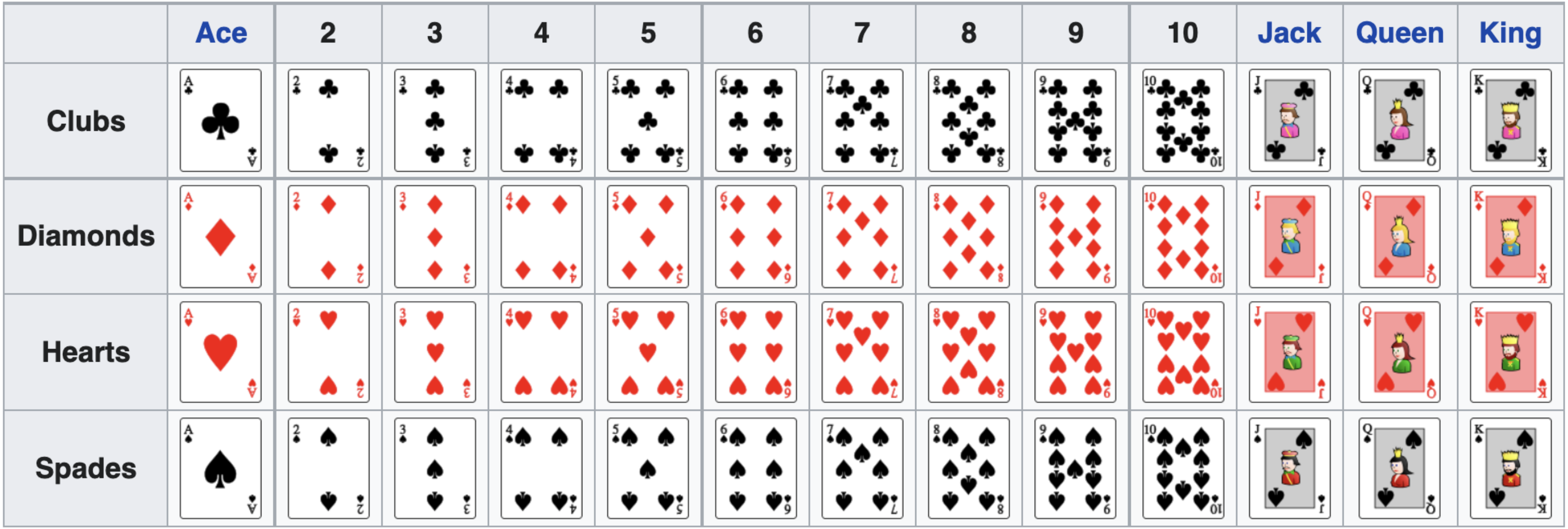
Joint probability is the probability of two independent events occurring together such as tossing a coin and getting heads while rolling a die and getting a 4. Written as P(H ∩ 4) where "∩" is read as "intersection" or "and".

Independent VS Dependent Events

- Independent events do not affect one another and do not increase or decrease the probability of another event happening
- Dependent events influence the probability of other events – or their probability of occurring is affected by other events.

Dependent Events – Multiplication Rule

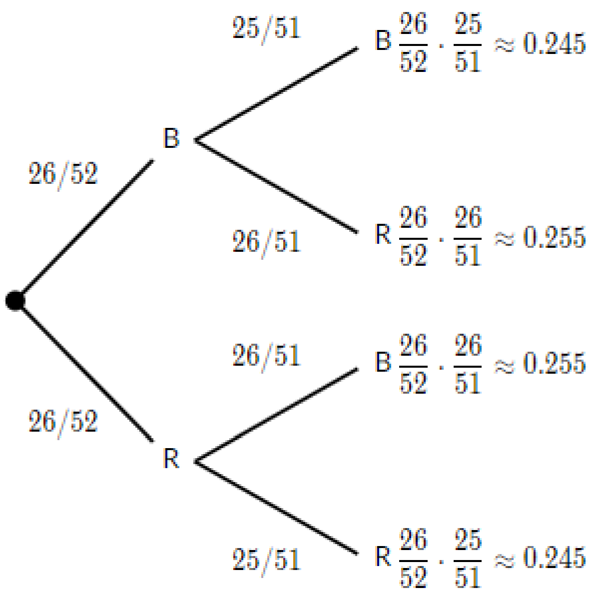
Example set of 52 playing cards; 13 of each suit: clubs, diamonds, hearts, and spades



What is the probability of drawing two black cards?

Probability of Picking Two Black Cards - Multiplication Rule

Half of the 52 cards are black, so the probability that the first card is black is 26/52. But the probability of getting a black card changes on the next draw, since the number of black cards and the total number of cards have both been decreased by 1.

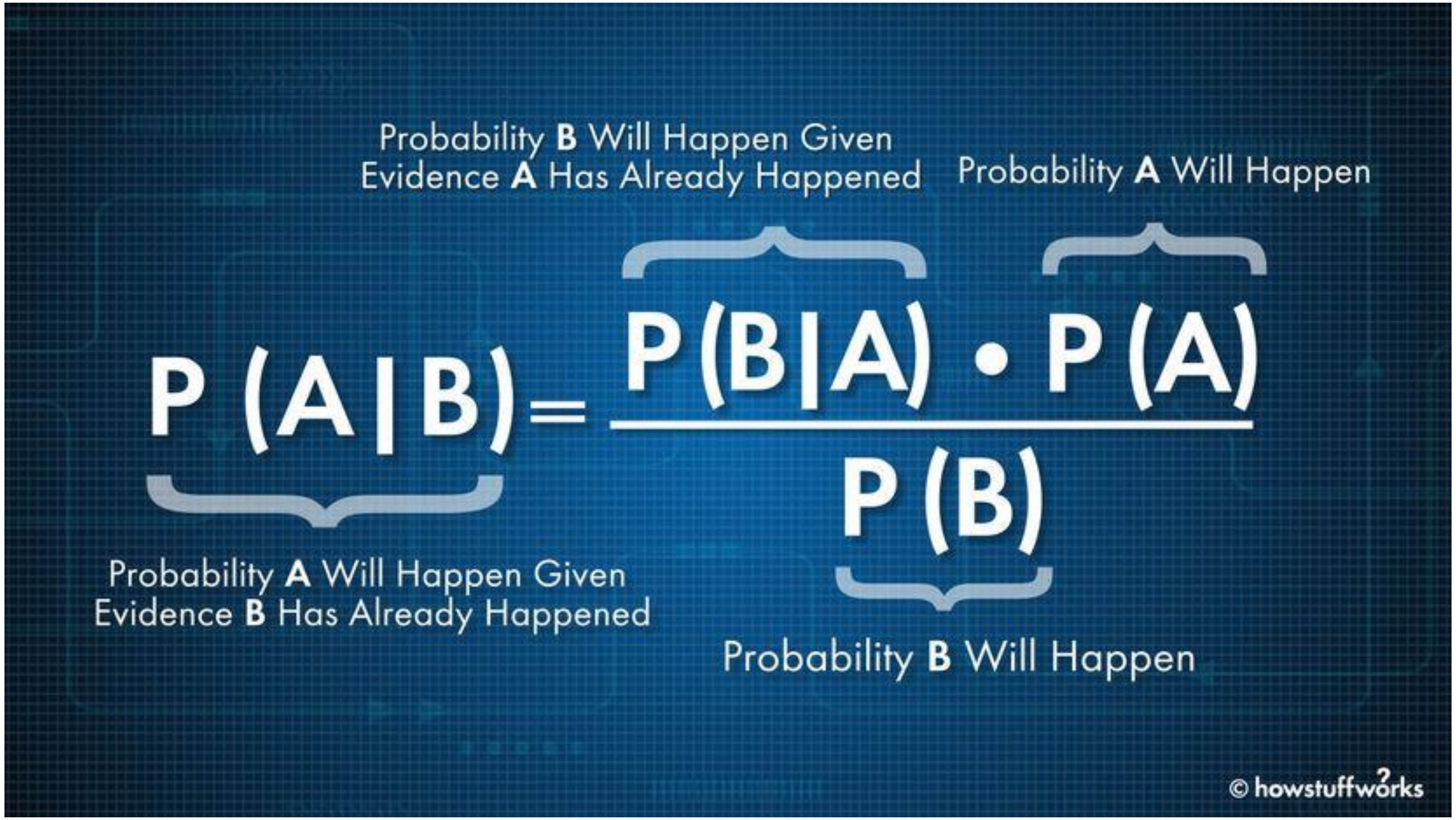


$$P(A \cap B) = P(A) * \boxed{P(B|A)}$$

So the probability that both cards are black is:

$$P(\text{both black}) = \frac{26}{52} \cdot \frac{25}{51} \approx 0.245$$

Conditional Probability & Bayes' Theorem

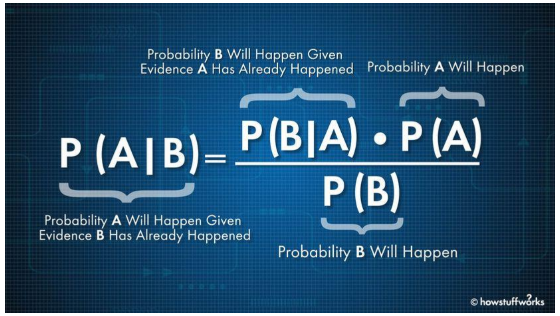


Work to understand it, but do not memorize this formula!

Bayes' Theorem Example

- A given test for cancer has:
  - false negative (does not detect cancer when it is present) rate of 20%
  - false positive (detects cancer when not present) of 9.6%
- Population rate for this type of cancer is 1%
- During routine ***screening***, a patient has a positive test result
- What is that patient's probability of having this cancer based on this one test?

Bayes’ Formula for Cancer Test Example



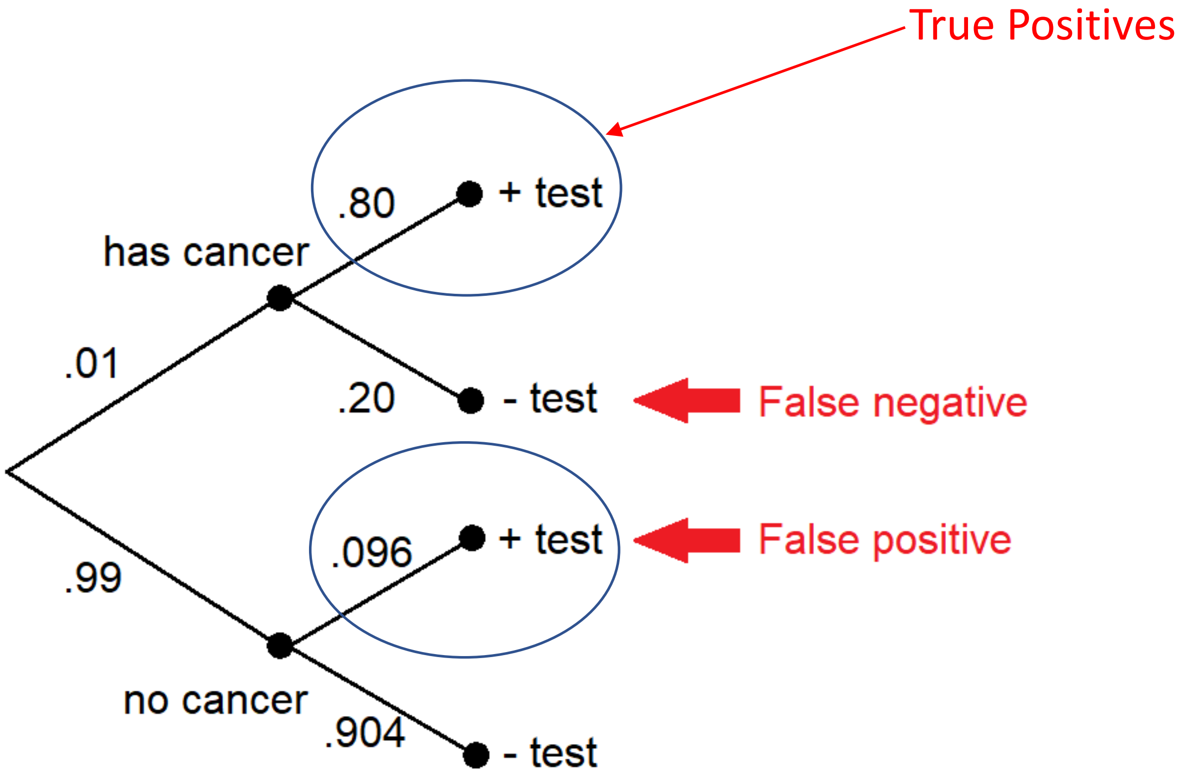
P(have Cancer | Test positive) =

$$\frac{\overbrace{\text{P(Test positive|have Cancer)}}^{\text{True positive (.80)}}\ \overbrace{\text{P(have Cancer)}}^{\text{Population Stat (.01)}}}{\text{P(Test positive|have Cancer) P(have Cancer)} + \underbrace{\text{P(Test positive | no Cancer)}}_{\text{False positive (.096)}}\ \underbrace{\text{P(no Cancer)}}_{\text{Population Stat (.99)}}}$$

$$.078 = \frac{.80 * .01}{(.80 * .01) + (.096 * .99)}$$

Tree Diagram of Bayes’ Theorem Applied

$$.078 = \frac{.80 * .01}{(.80 * .01) + (.096 * .99)}$$



Posterior vs Prior Probabilities

- A key concept of Bayes' Theorem is the notion of revisiting conclusions with updated evidence.
- The "posterior probability" is the updated probability gicen new evidence
- The "prior probability" is the previously calculated/estimated probability. In the example we just looked at, it was the population statistic for the probability of the cancer. Upon applying a second test, we would use the calculated probability from the first (.078)

Bayes' Formula
update with confirmatory test

Instead of population statistic of .01

Previously Calculated Probability (.078)

$$P(\text{have Cancer} \mid \text{Test positive}) = \frac{\underbrace{P(\text{Test positive}|\text{have Cancer})}_{\text{True positive (.80)}}\ \underbrace{P(\text{have Cancer})}_{(.078)}}{P(\text{Test positive}|\text{have Cancer})\ P(\text{have Cancer}) + \underbrace{P(\text{Test positive} \mid \text{no Cancer})}_{\text{False positive (.096)}}\ \underbrace{P(\text{no Cancer})}_{(.922)}}$$

(1 - .078)

Understand… **reality isn't changing** here! Our understanding of it (given new information) is!!!

$$.413 = \frac{.80 * .078}{(.80 * .078) + (.096 * .922)}$$

Still not an overwhelming diagnosis!