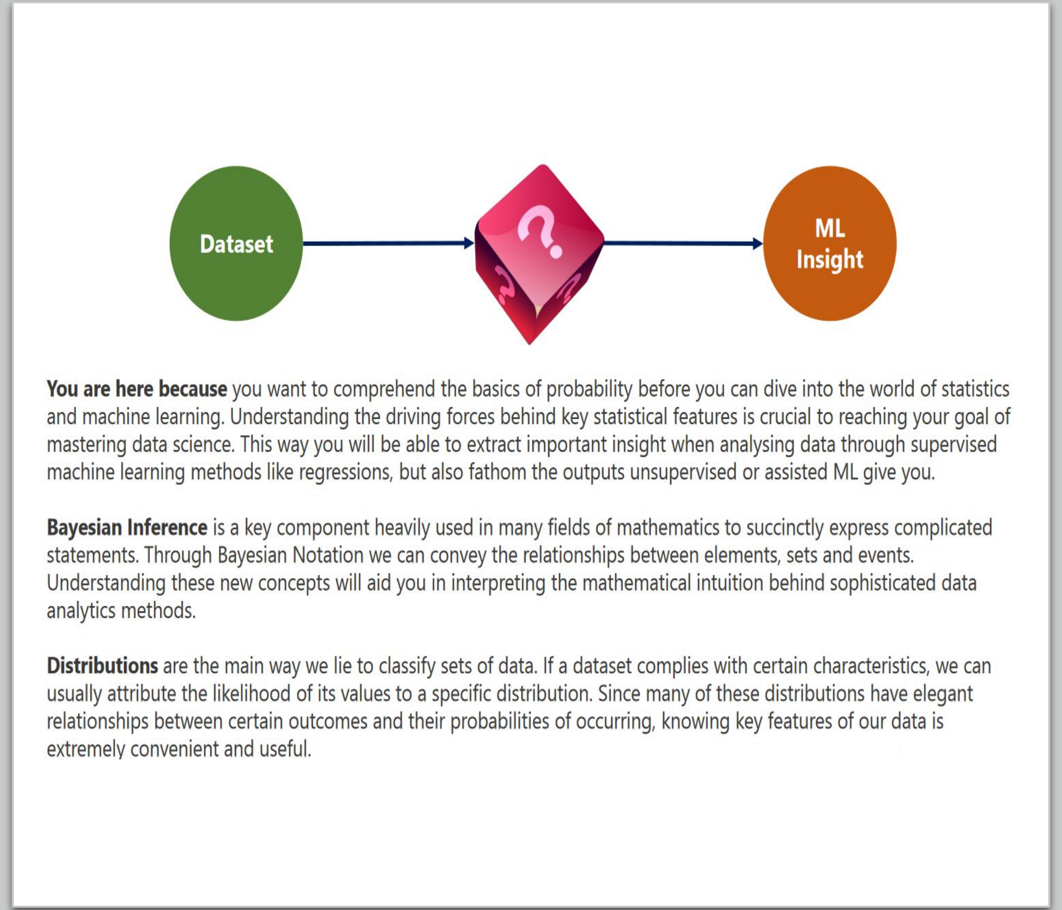
STATISTICS

- *Statistics is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of data.*

- *The practice or science of collecting and analysing numerical data in large quantities, especially for the purpose of inferring proportions in a whole from those in a representative sample.*



You are here because you want to comprehend the basics of probability before you can dive into the world of statistics and machine learning. Understanding the driving forces behind key statistical features is crucial to reaching your goal of mastering data science. This way you will be able to extract important insight when analysing data through supervised machine learning methods like regressions, but also fathom the outputs unsupervised or assisted ML give you.

Bayesian Inference is a key component heavily used in many fields of mathematics to succinctly express complicated statements. Through Bayesian Notation we can convey the relationships between elements, sets and events. Understanding these new concepts will aid you in interpreting the mathematical intuition behind sophisticated data analytics methods.

Distributions are the main way we lie to classify sets of data. If a dataset complies with certain characteristics, we can usually attribute the likelihood of its values to a specific distribution. Since many of these distributions have elegant relationships between certain outcomes and their probabilities of occurring, knowing key features of our data is extremely convenient and useful.

Probability

- Life is filled with uncertain events and often we must consider the possible outcomes before deciding.

- We ask ourselves questions like, What is the chance of success? and what is the probability that we fail to determine whether the risk is worth taking?

- By using probability and statistical data they can predict how likely each outcome is and make the right call for their firm.

- The probability is the chance of something happening. A more academic definition for this would be the likelihood of an event occurring.

- An *event is a specific outcome or a combination of several outcomes.*

- The probability *values having a probability of 1 expresses absolute certainty of the event occurring and a probability of 0 expresses absolute certainty of the event not occurring.*

What is probability?

Probability is **the likelihood of an event occurring**. This event can be pretty much anything – getting heads, rolling a 4 or even bench pressing 225lbs. We measure probability with numeric values between 0 and 1, because we like to *compare* the relative likelihood of events. Observe the general probability formula.

$$P(X)=\frac{\textit{Preferred outcomes}}{\textit{Sample Space}}$$

Probability Formula:

- The Probability of event X occurring equals the *number* of preferred outcomes over the *number* of outcomes in the sample space.
- Preferred outcomes are the outcomes we want to occur or the outcomes we are interested in. We also call refer to such outcomes as "Favorable".
- Sample space refers to all possible outcomes that can occur. Its "size" indicates the amount of elements in it.

If two events are independent:

The probability of them occurring simultaneously equals the product of them occurring on their own.

P(A♥) = P(A) . P(♥)

The probability of an event A occurring, denoted by P(A) is equal to the number of preferred outcomes over the total number of possible outcomes by preferred winning outcomes that we want to see happen.

- A→ Event
- P(A) → probability

$$P(A) = \frac{\text{preferred outcomes (favourable)}}{\text{all (sample space)}}$$

Example:

- Consider ***event A is flipping a coin and getting heads.***
- In this case *heads are our only preferred outcome*
- Assuming the coin doesn't just somehow stay in the air indefinitely.
- There are only two possible outcomes Heads or tails.
- This means that our probability would be a half. So, we write the following,

$$P(A)=1/2=0.5$$

- Imagine we have a standard six-sided die, and we want to roll to get four.
- We have a single preferred outcome i.e. to get 4 but this time we have a greater number of total possible outcomes are 6.
- Therefore, the probability of this event would look as follows.

$$P(A) = 1/6 = 0.167$$

Q. What if we wanted to roll a die for a number divisible by three?

- Note that the probability of two independent events occurring at the same time is equal to the product of all the probabilities of the individual events.

$$P(A \text{ and } B) = P(A).P(B)$$

- For instance, the likelihood of getting the **Ace of Spades** equals the probability of getting an ace times the probability of getting a spade.

$$P(\text{Ace} \spadesuit)=P(\text{Ace}) .P(\spadesuit)$$

Problems:

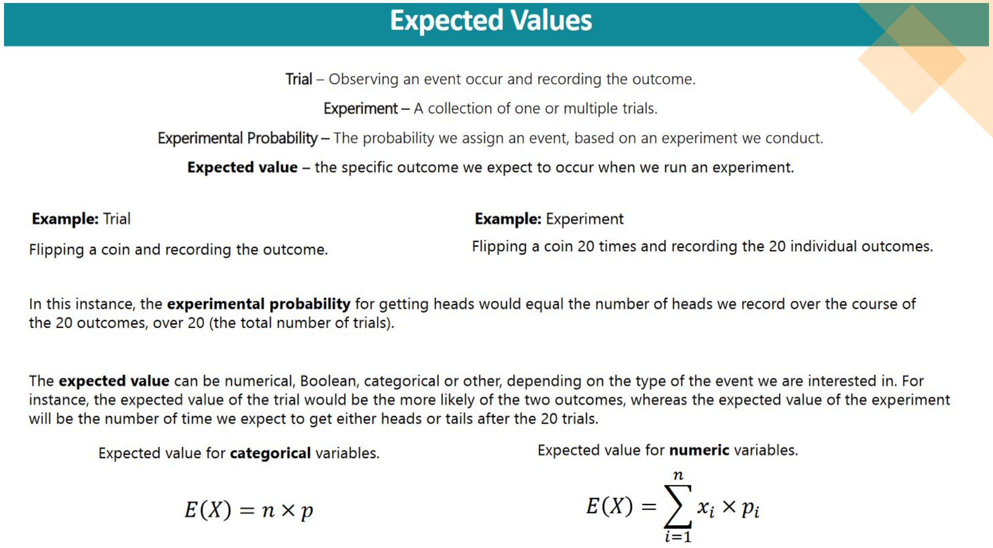
1. **Alice has 2 kids and one of them is a girl. What is the probability that the other child is also a girl? You can assume that there are an equal number of males and females in the world.**

2) A fair six-sided die is rolled twice. What is the probability of getting 2 on the first roll and not getting 4 on the second roll?

3) Cross-fertilizing a red and a white flower produces red flowers 25% of the time. Now we cross-fertilize five pairs of red and white flowers and produce five offspring. What is the probability that there are no red flower plants in the five offspring?

Computing Expected Values

- Expected values represent what we expect the outcome to be if we run an experiment many times to fully grasp the concept.
- So first what is an experiment? Imagine we don't know the probability of getting heads when flipping a coin. We are going to try to estimate it ourselves. So, we toss a coin several times after doing one flip and recording the outcome. We complete a trial by completing multiple trials. We are conducting an *experiment*.
- For example, if we toss a coin 20 times and record the 20 outcomes that entire process is a single experiment with 20 trials.
- The probabilities we get after conducting experiments are called *experimental probabilities.*
- Generally, when we are uncertain what the true probabilities are or how to compute them.
- We like conducting experiments the experimental probabilities we get are not always equal to the theoretical ones but are a good approximation.
- The formula we use to calculate experimental probabilities is similar to the formula applied for the theoretical ones, it is simply the number of successful trials divided by the total number of trials now that we know what an experiment is.
- The expected value of an event A denoted as E(A) is the outcome we expect to occur when we run an experiment.



Expected Values

Trial – Observing an event occur and recording the outcome.

Experiment – A collection of one or multiple trials.

Experimental Probability – The probability we assign an event, based on an experiment we conduct.

Expected value – the specific outcome we expect to occur when we run an experiment.

Example: Trial

Flipping a coin and recording the outcome.

Example: Experiment

Flipping a coin 20 times and recording the 20 individual outcomes.

In this instance, the **experimental probability** for getting heads would equal the number of heads we record over the course of the 20 outcomes, over 20 (the total number of trials).

The **expected value** can be numerical, Boolean, categorical or other, depending on the type of the event we are interested in. For instance, the expected value of the trial would be the more likely of the two outcomes, whereas the expected value of the experiment will be the number of time we expect to get either heads or tails after the 20 trials.

Expected value for **categorical** variables.

$$E(X) = n \times p$$

Expected value for **numeric** variables.

$$E(X) = \sum_{i=1}^{n} x_i \times p_i$$

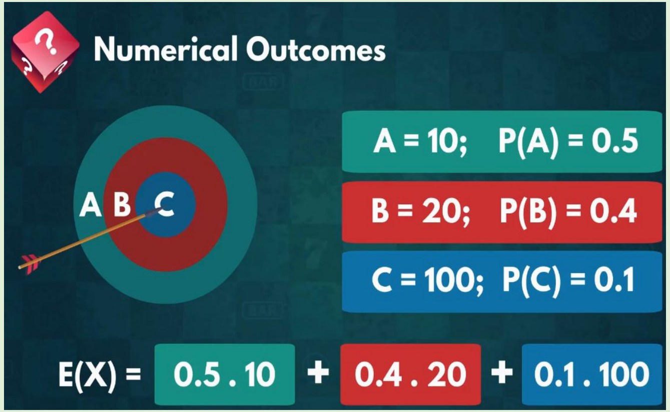
- To calculate Expected value we take the value for every element in the sample space and multiply it by its probability. Then we add all of those up to get the expected value.
- For example, if our random variable were the number obtained by rolling 6-sided die, the expected value would be

$$E(x) = 1*(1/6) + 2*(1/6) + 3*(1/6) + 4*(1/6) + 5*(1/6) + 6*(1/6)$$

$$= 3.5$$

- You are trying to hit a target with a bow and arrow the target has three layers. The outermost one is worth 10 points, the second one is worth 20 points and the innermost is worth 100. You have practiced enough to always be able to hit the target but not so much that you hit the centre every time.
- Suppose the probability of hitting each layer is as follows, 0.5 for the outmost, 0.4 for the second and 0.1 for the centre.
- The expected value for this example would be,

$$E(X) = 0.5*10 + 0.4*20 + 0.1*100 = 23$$



- But we can never get 23 points with a single shot. So why is it important to know what the expected value of an event is? *<u>We can use expected values to make predictions about the future</u>*, based on past data.
- We frequently make predictions using intervals instead of specific values due to the uncertainty the future brings.

Example 1:

A local club plans to invest \$10000 to host a baseball game. They expect to sell tickets worth \$15000 . But if it rains on the day of game, they won't sell any tickets and the club will lose all the money invested. If the weather forecast for the day of game is 20% possibility of rain, is this a good investment?

Example 2:

A company makes electronic gadgets. One out of every 50 gadgets is faulty, but the company doesn't know which ones are faulty until a buyer complains. Suppose the company makes a \$3 profit on the sale of any working gadget, but suffers a loss of \$80 for every faulty gadget because they have to repair the unit. Check whether the company can expect a profit in the long term. Write the probability distribution.

5) Ahmed is playing a lottery game where he must pick 2 numbers from 0 to 9 followed by an English alphabet (from 26-letters). He may choose the same number both times.
If his ticket matches the 2 numbers and 1 letter drawn in order, he wins the grand prize and receives \$10405. If just his letter matches but one or both of the numbers do not match, he wins \$100. Under any other circumstance, he wins nothing. The game costs him \$5 to play. Suppose he has chosen 04R to play. What is the expected net profit from playing this ticket?

Frequency

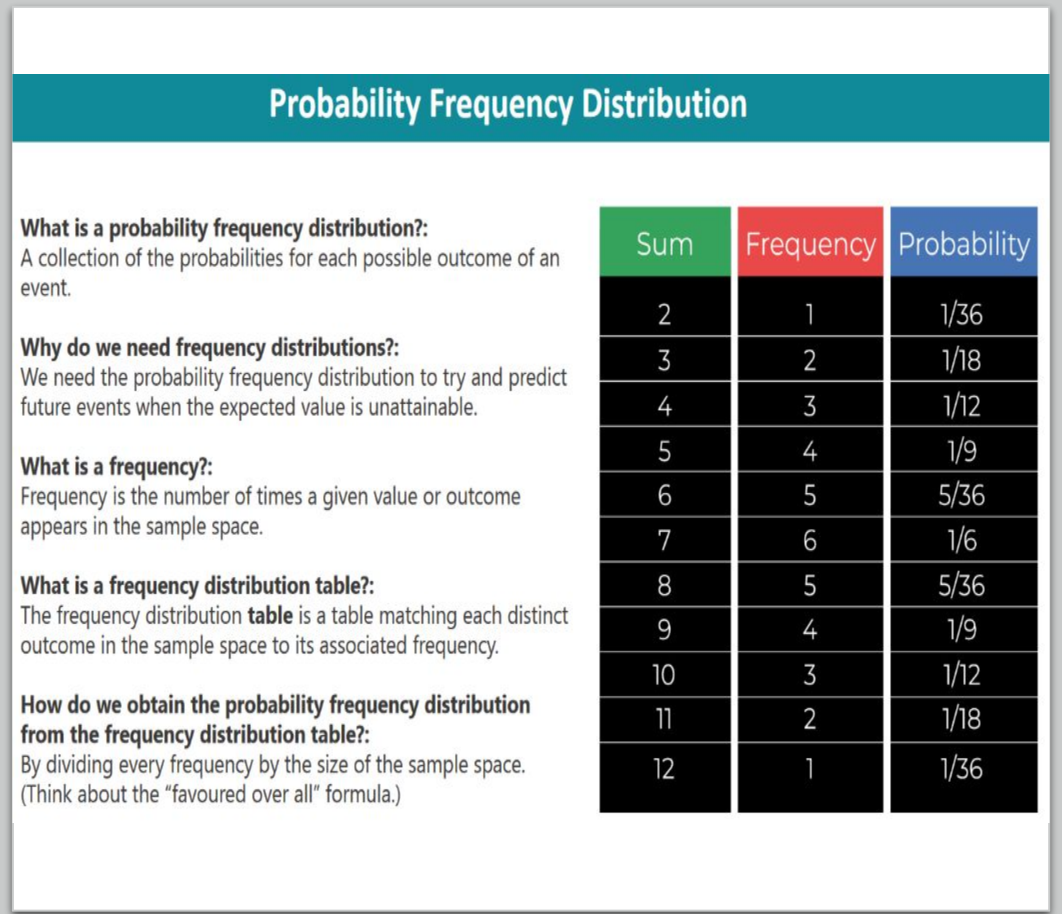
- Sometimes the result of the expected value is confusing or doesn't tell us much.
- Consider a example throwing two standard six sided dice and adding up the numbers on top.
- We have six options for what the result of the first one could be regardless of the number we roll. We still have six different possibilities for what we can roll on the second dice. That gives us a total of 6 * 6 = 36 different outcomes for the two roles.
- We can write out the results in a six by six table where rewrite the sum of the two dice.
- You can clearly see that we have repeating entries along the secondary diagonal a

- Notice how 7 occurs 6 times in the table. This means we have six favourable outcomes to get addition as 7. There are 36 possible outcomes, so the chance of getting a seven is,

$$P(7) = 6/36 = 1/6$$

Probability Frequency Distribution

- ***<u>A probability frequency distribution is a collection of the probabilities for each possible outcome.</u>***
- If we write out all the outcomes in ascending order and the frequency of each one we construct a frequency distribution table. By examining this table we can easily see how the frequency changes with the results.
- We need to transform the frequency of each outcome into a probability. Knowing the size of the sample space we can determine the true probabilities for each outcome. We simply divide the frequency for each possible outcome by the size of the sample space.
- A collection of all the probabilities for the various outcomes is called a <u>probability frequency distribution.</u>



Probability Frequency Distribution

What is a probability frequency distribution?:
A collection of the probabilities for each possible outcome of an event.

Why do we need frequency distributions?:
We need the probability frequency distribution to try and predict future events when the expected value is unattainable.

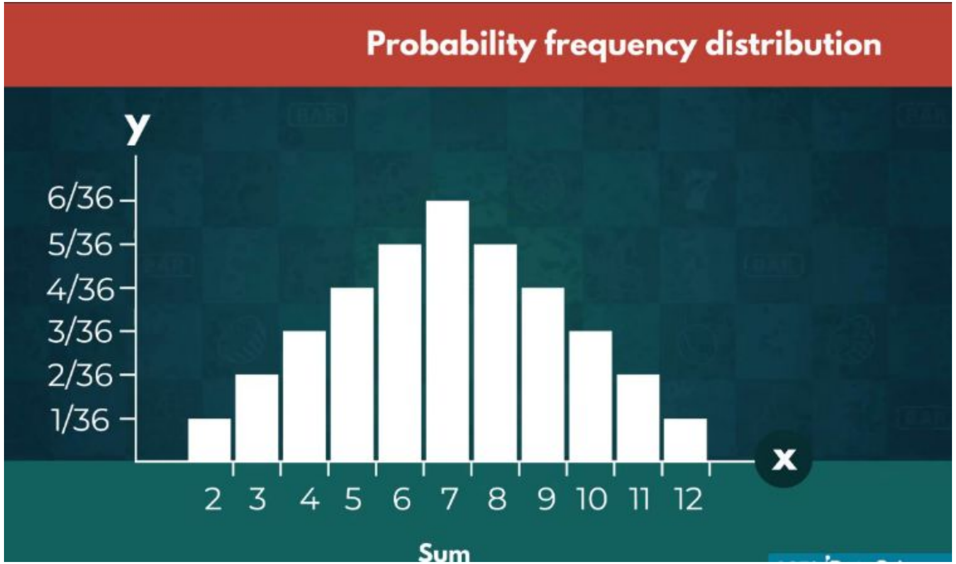
What is a frequency?:
Frequency is the number of times a given value or outcome appears in the sample space.

What is a frequency distribution table?:
The frequency distribution **table** is a table matching each distinct outcome in the sample space to its associated frequency.

How do we obtain the probability frequency distribution from the frequency distribution table?:
By dividing every frequency by the size of the sample space. (Think about the "favoured over all" formula.)

Sum	Frequency	Probability
2	1	1/36
3	2	1/18
4	3	1/12
5	4	1/9
6	5	5/36
7	6	1/6
8	5	5/36
9	4	1/9
10	3	1/12
11	2	1/18
12	1	1/36

- We can express this probability frequency distribution through a table or a graph.
- On the graph we see the probability frequency distribution. The x axis depicts the different possible numbers of sums we can get and the y axis represents the probability of getting each outcome when making predictions.
- We generally want our interval to have the highest probability.
- We can see that the individual outcomes with the high best probability are the ones with the highest bars in the graph usually the highest bars will form around the expected value.
- Thus the values around it would also be the values with the highest probability.



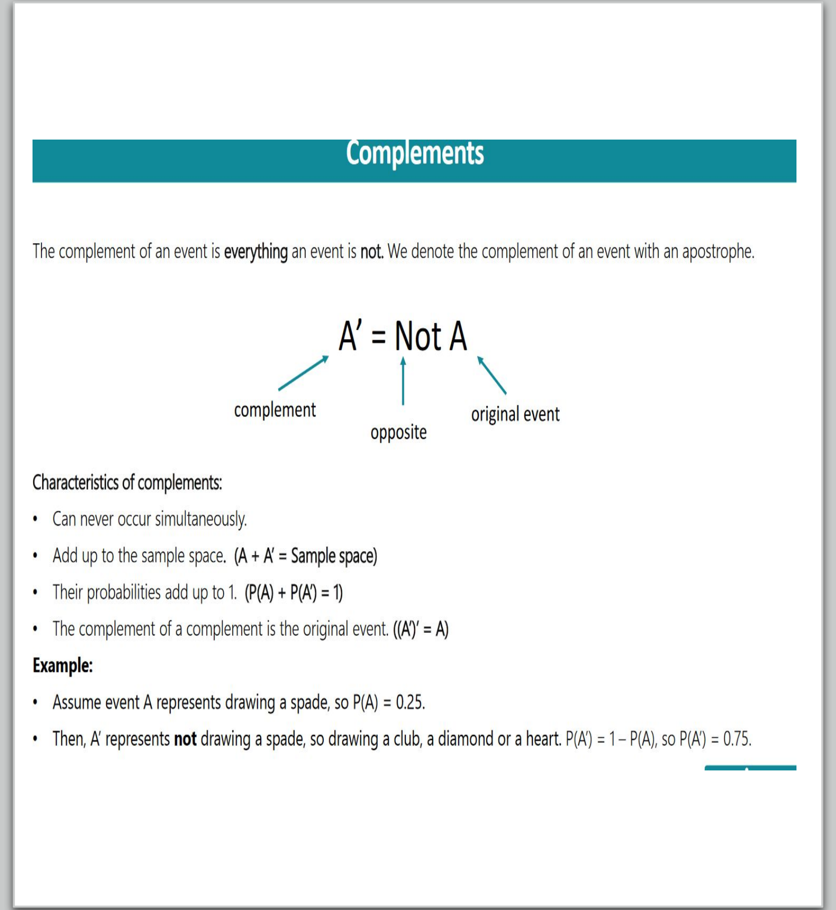
Events and Their Complements

- *A complement of an event is everything the event is not.*
- If we add the probabilities of different events we get their sum of probabilities. Now if we add up all the possible outcomes of an event we should always get one.
- *Example:*

  P(head) + P(tail) =1

- All events have complements and we denote them by adding an apostrophe for example the complement of the event A is denoted as A'.
- It is also worth noting that the complement of a complement is the event itself so a apostrophe would equal A.

  (A')'= A



Complements

The complement of an event is **everything** an event is **not.** We denote the complement of an event with an apostrophe.

Characteristics of complements:

- Can never occur simultaneously.
- Add up to the sample space. (A + A' = Sample space)
- Their probabilities add up to 1. (P(A) + P(A') = 1)
- The complement of a complement is the original event. ((A')' = A)

Example:

- Assume event A represents drawing a spade, so P(A) = 0.25.
- Then, A' represents **not** drawing a spade, so drawing a club, a diamond or a heart. P(A') = 1 – P(A), so P(A') = 0.75.

- Example, if you were rolling a standard six sided die and want to roll an even number, the opposite of that would be not rolling an even number which is the same as wanting to roll an odd number
- ***Complements are often used when the event we want to occur is satisfied by many outcomes***.
- For example you want to know the probability of rolling a one, two, four, five or six. That is the same as the probability of not rolling a three.
- We already said that the sum of the probabilities of all possible outcomes equals one.
- So you can probably guess how we calculate complements, the probability of the inverse equals 1 minus the probability of the event itself
    - $P(A) + P(B) + P(C) = 1$
    - $A' = B + C$
    - $P(A) = 1 - P(A)$

Example:

The sum of probabilities of getting 1, 2, 4, 5 or 6 is equal to the sum of the separate probabilities the likelihood of each outcome is equal to one sixth so the sum of their probabilities adds up to five sixth.

$$\begin{aligned} P(A) &= P(1) + P(2) + P(4) + P(5) + P(6) \\ &= 1/6 + 1/6 + 1/6 + 1/6 + 1/6 \\ &= 5/6 \end{aligned}$$

- Now another way of describing getting one, two, four, five or six is not getting a three.
- **Calculate the probability of not getting a three.** This is the complement of getting a three so we know that the two should add up to 1.

$$P(A') = 1 - 5/6 = 1/6$$

- Therefore the probability of not getting a three equals one minus the probability of getting a three.
- We know that P(3) equals one sixth so the probability of not getting three is equal to one minus one sixth.
- **Therefore the probability of not getting three is 5/6**. This shows that the probability of getting one, two, four, five or six is equal to the probability of not getting a three.

Probability – Combinatorics

- **Combinatorics deals with combinations of objects from a specific finite set.**
- In addition, we will also consider certain restrictions that can be applied to form combinations. These restrictions can be in terms of repetition, order or a different criterion.
- We will explore the three integral parts of combinatorix ***permutations, variations and combinations.*** Then we will use each of these parts to determine the number of favourable outcomes or the number of all elements in a sample space.

1) Permutations

Permutations represents the number of different possible ways we can arrange a set of elements. These elements can be digits, letters, objects or even people.

Permutations

Permutations represent the number of different possible ways we can **arrange** a number of elements.

$$P(n) = n \times (n-1) \times (n-2) \times \cdots \times 1$$

Permutations

Options for who we put first

Options for who we put second

Options for who we put last

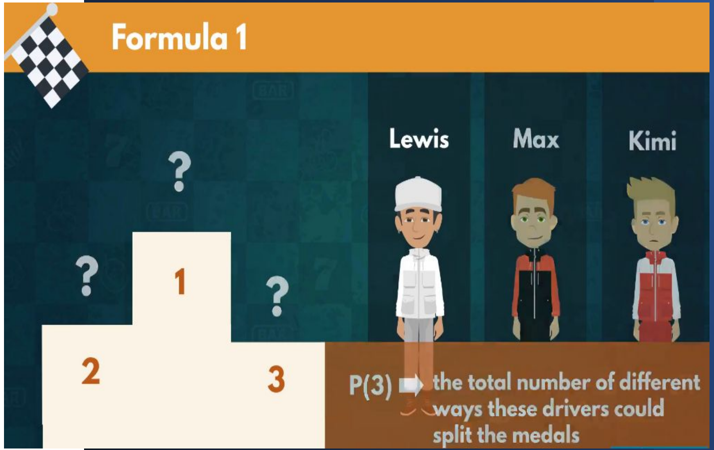
Characteristics of Permutations:

- Arranging **all** elements within the sample space.
- No repetition.
- $P(n) = n \times (n-1) \times (n-2) \times \cdots \times 1 = n!$ (Called "n factorial")

Example:

- If we need to arrange 5 people, we would have P(5) = 120 ways of doing so.

- For example, you have to decide Formula for race in which the three drivers on the podium are Lewis, Max and Kimmie
- A permutation of three, denoted P(3) would express the total number of different ways these drivers could split the medals among one another
- Suppose Lewis won the race. Then we have two possible scenarios Max finished second and Kimmie finished third or Kimmie finished second and Max finished third.
- Now suppose that Max won. Once again we have two possible outcomes but this time it is Lewis and Kimmie who have to split the silver and bronze medals. Either Kimmie got silver and Lewis got bronze or the other way around.
- If Kimmie won the race we would have two more ways the drivers can be arranged on the podium either Max gets silver and Lewis gets bronze or Lewis gets silver and Max gets bronze in total.
- This leaves us with six unique ways.
- We start filling out the positions one by one the order in which we fill them out is completely up to us for convenience. We usually start with the first slot which represents the race winner in our example since anybody out of the end many drivers in the set could have won the race.
- We have n different possible winners after that we have n-1 possible drivers left and any one of those can finish second regardless of which out of the n elements we chose to take the first slot.
- We have n-1possibilities for the second slot similarly we would have n-2 possible outcomes for who finishes third and so on.
- Generally, the further down the ranking we go the more options we exhaust and the more options we exhaust the fewer options we have left.
- This trend will continue until we get to the last element for which we will only have a single option available therefore mathematically the number of permutations is represented as,



Simple Operations with Factorials

- The factorial is the notion n! is used to express the product of the natural numbers from 1 to n.

- Negative numbers don't have a factorial and 0! is equal to 1.
- The first property for any natural number n we know that,
  - n! = (n-1)! * n
  - (n+1)! = (n+1) * n!

Factorials

Factorials express the **product** of all integers from **1** to **n** and we denote them with the "!" symbol.

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 1$$

Key Values:

- 0! = 1.
- If n<0, n! does not exist.

Rules for factorial multiplication. (For n>0 and n>k)

- $(n+k)! = n! \times (n+1) \times \cdots \times (n+k)$
- $(n-k)! = \frac{n!}{(n-k+1)\times\cdots\times(n-k+k)} = \frac{n!}{(n-k+1)\times\cdots\times n}$
- $\frac{n!}{k!} = \frac{k!\times(k+1)\times\cdots\times n}{k!} = (k+1) \times \cdots \times n$

Examples: n = 7, k = 4

- $(7+4)! = 11! = 7! \times 8 \times 9 \times 10 \times 11$
- $(7-4)! = 3! = \frac{7!}{4\times5\times6\times7}$
- $\frac{7!}{4!} = 5 \times 6 \times 7$

Solving Variations with Repetition

- ***Variations express the total number of ways we can pick and arrange some elements of a given set.***

- For example, imagine you went on vacation and forgot the code for the combination lock on your carryon. Luckily for you the lock requires a two-letter code using only the letters A B and C to unlock it.
- As we have three different options A B or C. Suppose we chose A and move on to the next letter since we can repeat values. We once again have these same three options for the second letter A B or C. This indicates that there are three different variations if we decide to start with A.
- Now if we put B in the first position, again we would have three options for what we choose for the second letter in general regardless of which one of the three letters we decide to start with.
- We are going to have three different options for the second letter therefore the total number of variations we can get is 3*3 = 9.
- The formula we use to calculate variations with repetition is the following
  - $\overline{V_p^n} = n^p$
- where n is the total number of elements, we have available and P is the number of positions we need to fill the way we interpret this notion is the number of variations with repetition.
- When picking p many elements out of n elements is equal to n^p. If we apply this to the combination lock example, we will write
  - $\overline{V_2^3} = 3^2 = 9$

Variations

Variations represent the number of different possible ways we can pick and arrange a number of elements.

Variations **with** repetition → $\bar{V}(n,p) = n^p$ (n: Number of different elements available; p: Number of elements we are arranging)

Variations without repetition → $V(n,p) = \frac{n!}{(n-p)!}$ (n: Number of different elements available; p: Number of elements we are arranging)

Intuition behind the formula. (With Repetition)

- We have n-many options for the first element.
- We still have n-many options for the second element because repetition is allowed.
- We have n-many options for each of the p-many elements.
- $n \times n \times n \ldots n = n^p$

Intuition behind the formula. (Without Repetition)

- We have n-many options for the first element.
- We only have (n-1)-many options for the second element because we cannot repeat the value for we chose to start with.
- We have less options left for each additional element.
- $n \times (n-1) \times (n-2) \ldots (n-p+1) = \frac{n!}{(n-p)!}$

** We interpret this as there are nine different variations of two letter pass codes consisting of A, B or C only, what happens if the law could use any of the 26 letters, we would have 26 to the power of 2 which is 676 different variations. **

Solving Variations without Repetition

- Imagine you are a track and field coach and need to choose which 4 members of your team run the relay and in what order. The team consists of 5 people Tom, Eric, David, Kevin and Josh and you must decide who starts, who anchors and who runs in between.
- We have 5 members in the team which means 5 different scenarios for who we want to start. Let's say we know that David is the best guy for the job. That decision leaves us with only 4 options for who gets the second position namely Tom, Eric, Kevin or Josh suppose we chose Josh to run after David.
- That leaves us with only 3 options for who runs third Tom, Erik or Kevin. Now if we pick Kevin then we know one of either Tom or Erik finishes the race. Not surprisingly if we chose Erik to run third instead, we once again have two choices for who runs last Tom or Kevin finally.
- If we choose Tom to run third we would have two options available Erik and Kevin. This means we can have six different variations for who runs the last two positions.
- If we have chosen David and Josh to run first and second respectively using a similar logic if we pick somebody different than Josh to run second we would still have six possible variations for who runs third and fourth.
- In fact regardless of who out of the four members of the team we chose to run second we would always have six options for who fills out the remaining spots on the team that suggests there are four times six or 24 different ways to arrange the three remaining positions.
- If we knew David starts, what if we aren't sure that David is the best runner to start well. If somebody out of the remaining four people started we would still have 24 ways of filling out the remaining spots on the team. What happens is the further down the order we go the fewer options we are left with.
- *Since nobody is permitted to run multiple legs this is what variations without repetition is about.*
- We cannot use the same element or in this case person twice in terms of numbers. We have five times four times three times two. This makes 120 different options of how to arrange our team for the competition.
- It's time to introduce the formula for calculating variations without repetition
- So the number of variations without repetition when arranging p elements out of a total of n is equal to,
  - $\overline{V_p^n} = \frac{n!}{(n-p)!} = \frac{5!}{1!} = 120$

Combinations

- The combinations represent the number of different ways we can pick certain elements of a set.

- Imagine you were trying to pick 3 people to represent your company on a very important technology related conference. There are 10 people working in the office.

- So how many different combinations are there?

- If you calculate this as a variation your answer would be 720 but you would be counting every group of 3 people several times over. This is because picking Alex, Sarah and Dave to go to the conference is the same as picking Alex, Dave and Sarah as variations don't take into account double counting elements.

- ***<u>We can say that all the different permutations of a single combination are different variations.</u>***

- Let us look at the Sarah, Alex and Dave example choosing those three to represent the company is a single combination since the order in which we pick them is not relevant. Choosing Sarah, Alex and Dave is exactly the same as choosing Sarah, Dave and Alex. Dave, Sarah and Alex; Dave, Alex and Sarah; Alex Dave and Sarah or Alex ,Sarah and Dave any of the six permutations we wrote is a different variation but not a different combination.

- That is what we meant when we said that combinations take into account double counting.

- The formula for calculating permutations of n many elements is simply n! since n is 3 in this case.

    - P(n)= P(3)=6

- There would be a total of 6 permutations for choosing Alex, Dave and Sarah since

Combinations

Combinations represent the number of different possible ways we can **pick** a number of elements.

$$C(n,p) = \frac{n!}{(n-p)!\,p!}$$

Combinations

Total number of elements in the sample space

Number of elements we need to select

Characteristics of Combinations:

- Takes into account double-counting. (Selecting Johny, Kate and Marie is the same as selecting Marie, Kate and Johny
- All the different permutations of a single combination are different variations.
- $C = \frac{V}{P} = \frac{n!/(n-p)!}{p!} = \frac{n!}{(n-p)!p!}$
- Combinations are symmetric, so $C_p^n = C_{n-p}^n$, since selecting p elements is the same as omitting n-p elements.

- We are going to have six variations for any combination. This means that we are going to end up with 6 times fewer combinations than variations using the formulas we already know.

- V_3^{10}= 10*9*8=720

- There are ten times nine times eight or 720 variations

- In terms of combinations, We have 720 divided by six or 120 ways of choosing who represents the company.

- $C_3^{10} = \frac{720}{6}$ =120

- Once again we can say that all the different permutations of a single combination are different variations. There are 6 permutations, 120 combinations and 720 variations. Now let's construct it together what's the number of combinations for choosing p many elements out of a sample space of n elements as you saw in the last example the number of combinations equals the number of variations over the number of permutations mathematically we would write that as,

- C $= \frac{V}{P}$

- If we plug in the formulas associated with variations and permutations, we get the following formula

- $C_p^n = \frac{n!}{p!(n-p)!}$

- Let's apply the formula to our example the number of combinations equals

- $C_3^{10} = \frac{10!}{3!(10-3)!} = \frac{10*9*8}{1*2*3} = \frac{720}{6} = 120$

- Let's go through another example what if we had to choose 4 out of 10 people to go to the conference according to the formula we are going to have

- $C_4^{10} = \frac{10!}{4!(10-4)!} = \frac{10*9*8*7}{1*2*3*4} = \frac{5040}{24} = 210$

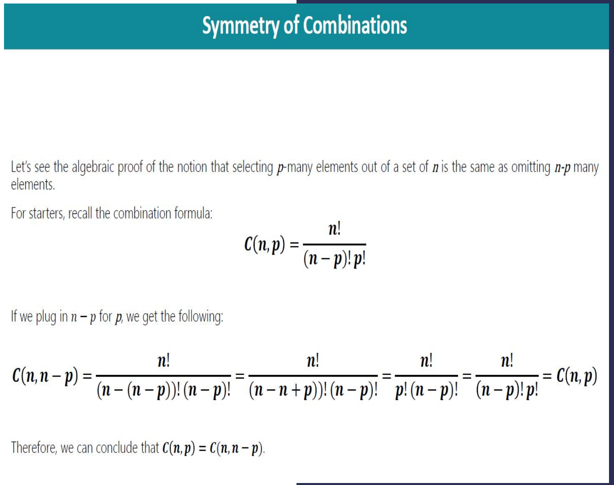
- ten factorial over four factorial times six factorial combinations of doing so after some simplifications this would equal seven times eight times nine times ten over one times two times three times four. After computing the values this equals five thousand forty over twenty four which is to ten.

Example:

Amita randomly picks 4 cards from a deck of 52-cards and places them back into the deck (Any set of 4 cards is equally likely). Then, Babita randomly chooses 8 cards out of the same deck (Any set of 8 cards is equally likely). Assume that the choice of 4 cards by Amita and the choice of 8 cards by Babita are independent. What is the probability that all 4 cards chosen by Amita are in the set of 8 cards chosen by Babita?

Symmetry of Combinations

- Imagine you are going on a picnic and have 6 pieces of fruit that you want to take with you
- However your basket only has room for 4 of them. Using the combinations formula, we are going to have 15 possible choices.
    - $C_4^6 = \frac{6*5*4*3}{1*2*3*4} = 15$
    - 6!/5!1!=6 6!/1!5!=1
- Therefore you go out and buy a bigger basket. But it turns out it can only fit 5 pieces of fruit. According to the formula there are just 6 ways of picking the 5 fruits.
- What if you get an even bigger basket which can fit 6 fruits well in how many ways can you pick 6 fruits out of six fruits only one. Picking all of them.
- So we see that in this case ***picking more elements leads to having fewer combinations.***
- This is because we can construct the question in a different way instead of picking which pieces of fruit to take. We can choose the pieces to leave behind therefore picking 4 fruits out of 6 is the same as choosing 2 fruits that will be left out.
- Mathematically
    - $C_2^6 = \frac{6!}{2!*4!} = 15$
- How about 5 out of 6 fruits. It's the same as choosing which one out of the 6 to leave behind. In the general case, we can pick p many elements in as many ways as we can pick (n – p) many elements.
- This shows us that when it comes to combinations **the number of possible ways in which p many elements can be selected is symmetric with respect to n/2.**
- Consider recall the example where we had to select 3 of our 10 employees to represent the company at a conference as we already showed there are 120 possible selections.
- What if instead of choosing 3, we had to pick 7 people to go to the conference. The number of combinations we would have is thus
    - $C_7^{10} = \frac{10!}{7!*3!} = \frac{8*9*10}{1*2*3} = \frac{720}{6} = 120$
- As we would also have 120 different ways of picking the seven employees. That's because picking 7 out of 10 employees to take to the conference is the same as choosing 3 out of 10 to leave behind.
- To sum up when **p > n/2 > n-p** in such instances we can apply symmetry to avoid calculating factorial of large numbers.
- Generally we use symmetry to simplify the calculations we need to make.

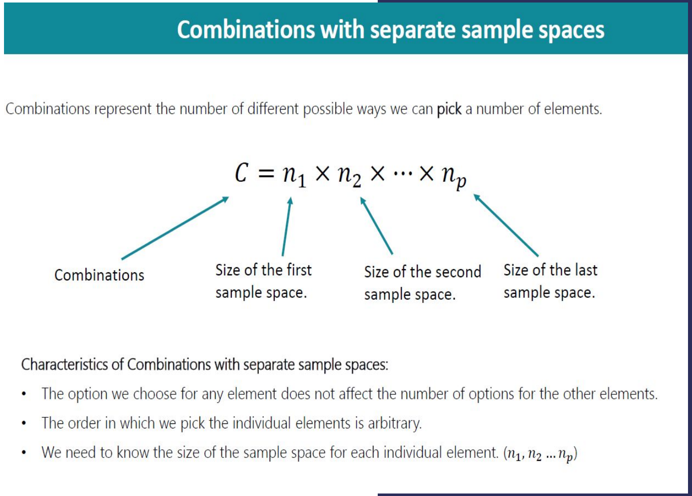


Solving Combinations with Separate Sample Spaces

- Imagine the diner near work just introduced a lunch menu which consists of a sandwich, a drink, and a side.
- Assuming you go there every day. How long will it take for you to be able to try out every possible item on the menu? To solve this you need to know what is included in their lunch deal.
- Each menu consists of a sandwich, a side and a drink. They offer three types of sandwiches a panini, a toast and a veggie wrap. The sides they have available are only fries and onion rings and the drinks they offer are cola or water.
- The way to tackle such problems is by thinking about the different parts of the menu as separate positions. If we start by choosing a sandwich first we have 3 options for each of them. We can pick one of 2 sides fries or rings to complete our menu. We would also have to add a drink which can either be water or coke.
- Therefore for any combination of Sandwich and side we have two ways of completing the menu thus we would have a total of three dishes times two sides times two drinks or 12 different lunch menus at the diner.

  3* 2 * 2 =12

- This is important because it shows us how many different possibilities there are available despite the choices for each part seeming limited.
- Furthermore, this allows us to determine the appropriate amount of time it would take for such a task to be completed when the components are simply too many. We can remove several of the options to tremendously decrease the workload.
- The way of calculating the total number of combinations for these kinds of questions is by simply multiplying the number of options available for each individual event.



Example:
Consider to win lottery you need to satisfy two independent events
1) Correctly guess “Powerball” number (From 1 to 26)
2)Correctly guess 5 regular numbers (From 1 to 69)
Find the probability of winning single ticket.

In how many different ways could 23 children sit on 23 chairs in a Maths Class? If you have 4 lessons a week and there are 52 weeks in a year, how many years does it take to get through all different possibilities? *Note: The age of the universe is about 14 billion years.*

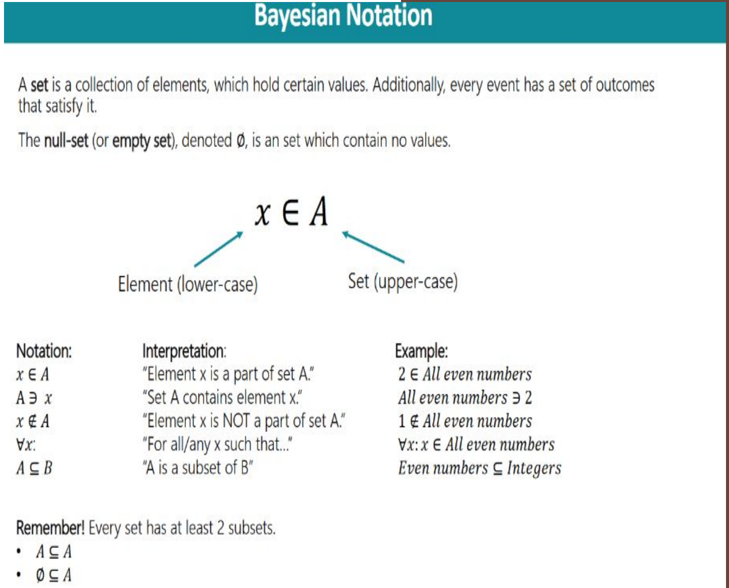
Unfortunately, you can’t remember the code for your four-digit lock. You only know that you didn’t use any digit more than once. How many different ways do you have to try? What do you conclude about the safety of those locks?

In a shop there are five different T-shirts you like, coloured red, blue, green, yellow and black. Unfortunately you only have enough money to buy three of them. How many ways are there to select three T-shirts from the five you like?

Four children, called A, B, C and D, sit randomly on four chairs. What is the probability that A sits on the first chair?

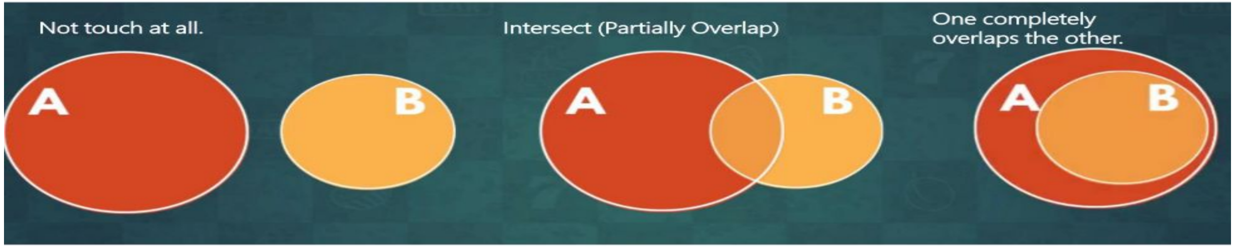
Sets And Events

- Every event has a set of outcomes that satisfy it. These are the favourable outcomes.
- For example, the event could be even, and the set of values would consist of 2, 4, 6 and all other even numbers.
- However, values of a set do not always have to be numerical. For instance, an event can be being a member of the European Union. Values like France or Germany would be a part of this set and values like USA or Japan would not.
- Convention dictates that we use upper case letters to denote these sets and lower-case letters to express individual elements.
- In the numerical example upper case X will express all even numbers and lowercase x.
- Any set can be either empty or have values in it.
- If it does not contain any values, we call it the empty set or null set and denote it with ∅
- The non-empty sets can be finite or infinite depending on the number of elements they have when working with them.
- We often want to express if an element is part of a set the symbol, we use to denote that is the "belongs to (∈)" symbol.
    - x ∈ **X**
- We read it as x is an element of or simply in set X.
- But what if we want to show that an element is not contained in a set then we can use the same notations but simply cross out the symbol with a single diagonal line like,
    - x ∉ **X**
- So the statements now mean x IS NOT IN X and X does not contain x.
- A subset is a set that is fully contained in another set.
- If every element of A is also an element of B then A is a subset of B. We do know that with a subset B as you can see not all elements of B are necessarily part of a going forward.



Multiple Events

The sets of outcomes that satisfy two events A and B can interact in one of the following 3 ways.



Examples:
A -> Diamonds
B -> Hearts

Diamonds
Queens

Red Cards
Diamond

- Take events A and B for example, we express the set of values that satisfy each of them as Circle's one for A and another one for B
- Any element that is part of either set will be represented by a point in the appropriate circle. Since we can have additional events the more events, we have the more circles we draw.
- Let's only focus on A and B, the two circles can either not touch it all, intersect or one can completely overlap the other
- If the **two circles never touch, then the two events can never happen simultaneously.** Essentially event A occurring guarantees that event B is not occurring and vice versa Example getting a diamond and getting a heart would be such a situation. If we get a heart, we can't get a diamond and if we get a diamond, we can't get a heart since each card has exactly one suit.
- Now if these circles **intersect** it means that the two events can occur at the same time.
- Imagine we draw a card from a standard deck of playing cards. If event A is drawing a diamond and event B is drawing a queen, the area where they intersect will be represented solely by the queen of diamonds the remaining area of a will represent all other diamonds whereas the area of B outside of that will represent all other queens.
- The third case happens if **one circle completely overlaps another**. That means that one event can only ever occur if the other one does as well. For instance, event A could be drawing a red card and event B could be drawing a diamond, then the circle of B is completely contained inside A. So, we can only ever get a diamond if we get a red card notice that, if the card we drew is black it cannot be a diamond.
- Thus, if event A does not occur then neither does event B. However, because we can draw a heart it is possible to get a red card that isn't a diamond.
- Therefore, event B not occurring does not guarantee event A not occurring. In short if an outcome is not part of a set it cannot be part of any of its subsets. However, an outcome not being part of some subset does not exclude from the entirety of the Greater set.
- Of course, we can have 3, 4 or as many events as we want and the relationships between each two will always be represented by one of the three ways

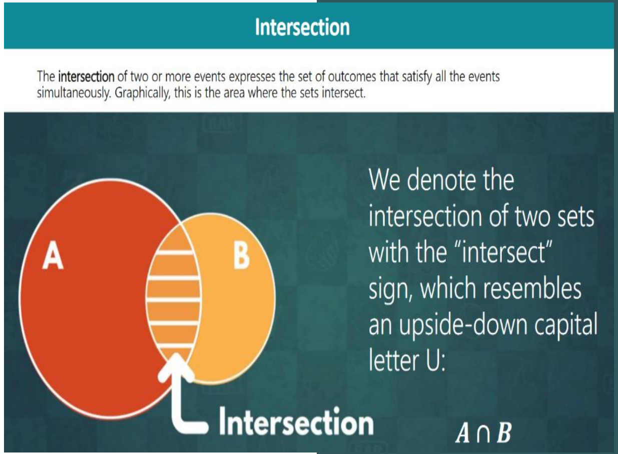
- The intersection of two events occur when we want both A and B to happen at the same time.
- Graphically the intersection is exactly as the name suggests the area where these events intersect.
- It consists of all the outcomes that are favourable for both event A and event B simultaneously as we denote it as A intersect B
- Consider the examples the intersection of all hearts and all diamonds is the empty set as there are no outcomes which satisfy both events simultaneously, we would write this as,

$$A \cap B = \emptyset$$

- Consider example, the intersection of all diamonds and all queens is represented by the queen of diamonds. That card is the only one that satisfies being a queen and being a diamond at the same time.
- In the example with red cards and diamonds the intersection of the two would simply be all diamonds. That is because any diamond is simultaneously red and a diamond. We would write this as

$$A \cap B = B$$

- Everybody we use intersections only when we want to denote instances where both events A and B happens simultaneously.



If we only require one of A or B to occur regardless which one that is the same as asking either A or B to happen, in such cases we need to find the union of A and B

The union of two sets is a combination of all outcomes preferred for either A or B.

We denote the union of two sets as a U symbol.

Let us examine what the unions would be in the three different cases if the sets A and B do not touch it all then their intersection would be the empty set, therefore their union would simply be their sum.

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

Going back to the card example the union of hearts and diamonds would be all red cards, no card can have multiple suits so we need not worry about counting a card twice therefore the number of red cards equal the union of cards which are either diamonds or hearts

If the events intersect the area of the Union, then it is represented by the sum of the two sets minus their intersection

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

That is because if we simply add up the area of the two sets we would be double counting every element that is part of the intersection.

In fact the union formula we just showed you is universally true regardless of the relationship between A and B

So what happens if B is a subset of A. In that case the union would simply be the entire set A

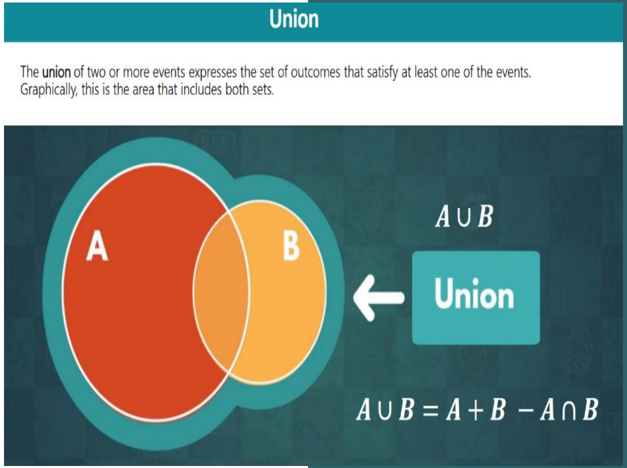
Imagine event A is being from the U.S. an event B is being from California if you talk about all the people who are either from California or the United States you are simply talking about all the people from the USA

Remember that the intersection of these two sets is equal to the entire set B, so the intersection of A and B represents all the people from California

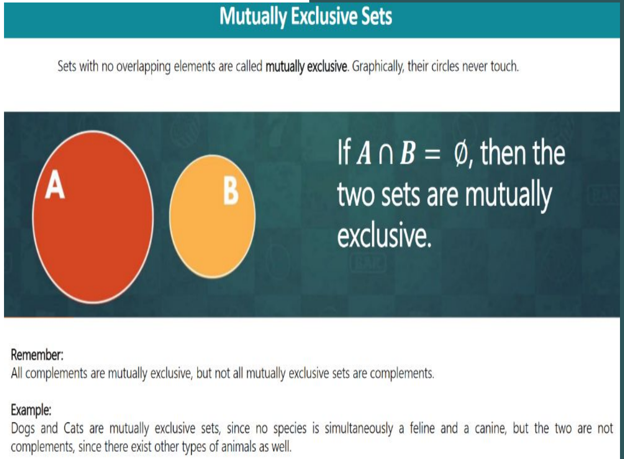
If we plug this into the formula, that would give us the following statement,

The union of all people from California or the United States = all California natives + all Americans - all Californians

we get exactly what we expect the union of people from either California or the United States equals the entire population of the USA now



- Mutually exclusive sets are sets in which you're not allowed to have any overlapping elements graphically their circles never intersect.
- Mutually exclusive sets have the empty set as their intersection.
- Therefore, if the intersection of any number of sets is the empty set then they must be mutually exclusive and vice versa.
- What about their union if some sets are mutually exclusive? Their union is simply the sum of all separate individual sets.
- Sets have complements which consist of all values that are parts of the sample space but not part of the set.
- Consider a set consisting of all the odd numbers, its complement would be the set of all even numbers. It means complements are always mutually exclusive.
- However not all mutually exclusive sets are complements.
- For instance, imagine A is the set of all even numbers and B is the set of all numbers ending in five. We know that any number ending with five is odd.
- So these two sets are definitely mutually exclusive.
- However the complement of all even is all odd and not just the ones ending with 5.
- Therefore a number like 13 would be part of the complement but not the set B.



Dependence and Independence of Sets

- We can have dependent events as their probabilities vary as conditions change.
- For instance, take the probability of drawing the Queen of Spades normally. The answer is 1/52. Since we have exactly one favourable outcome and fifty-two elements in the sample space.
- Now imagine we know that the card we drew was a spade. Our chances of getting the queen of spades suddenly go up since the new sample space contains the 13 cards from the suit only. Therefore, the probability becomes 1/13.
- Now imagine a different scenario instead of a spade. We know our card is a queen.
- So, the sample space only consists of 4 cards. Therefore, the probability of drawing the Queen of Spades becomes 1/4
- With this example you could clearly see how the probability of an event changes depending on the information we have.
- Lets introduce some new notation as usual. Suppose we have two events A and B to express the probability of getting A, if we are given that B has occurred, we use the following notation, **P(A/B), we read this as P of A given B.**
- Going back to our card example, event A is drawing the queen of spades and event B is drawing a spade.
- Therefore P(A/B) would represent the probability of drawing the queen of spades. If we know the card is a spade so

$$P(A/B)=1/13$$

- Similarly if event C represents getting a queen then P(A/C) expresses the likelihood of getting the Queen of Spades assuming we drew a queen.
- Thus

$$P(A/C) = 1/4$$

- We call this probability the **conditional probability** and we use it to distinguish dependent from independent events

Independent and Dependent Events

If the likelihood of event A occurring (P(A)) is affected event B occurring, then we say that A and B are **dependent** events.

Alternatively, if it isn't – the two events are **independent**.

We express the probability of event A occurring, given event B has occurred the following way $\boldsymbol{P(A|B)}$.
We call this the conditional probability.

Independent:

- All the probabilities we have examined so far.
- The outcome of A does not depend on the outcome of B.
- $P(A|B) = P(A)$

Example

- A -> Hearts
- B -> Jacks

Dependent

- New concept.
- The outcome of A depends on the outcome of B.
- $P(A|B) \neq P(A)$

Example

- A -> Hearts
- B -> Red

- Example:

1) In a manufacturing unit 3 parts from assembly are selected. You are observing whether they are defective or non-defective.

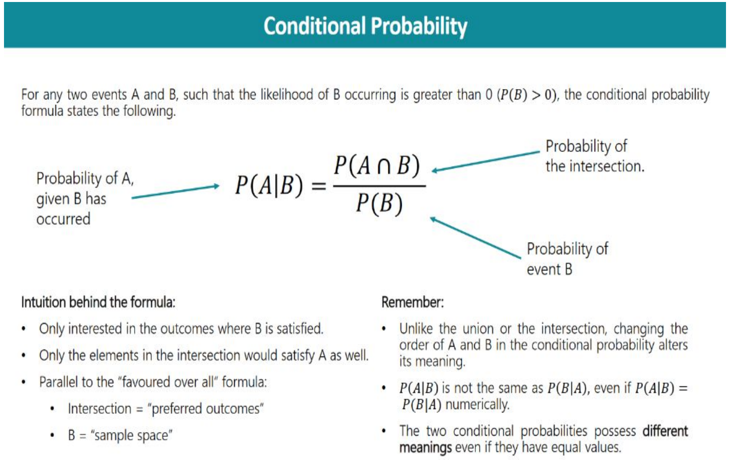
   - Determine

     a) The sample space.

     b) The probability of the event of getting 2 defective parts if there is chance of 1 defective in 10.

The Conditional Probability Formula

- **The conditional probability is the likelihood of an event occurring assuming a different one has already happened.**
- Consider the coin flip example, A represents getting heads and B represents getting heads on the previous flip the probability of getting heads.
- Now after getting heads last time, P(A) is still 0.5 therefore P(A) = P(A/B).
- This is equivalent to saying the <u>two events are independent.</u>
- We also mentioned that if any two events are independent the probability of their intersection is the product of the individual probabilities.
    - $P(A \cap B) = P(A) \times P(B)$
- Now let us examine the Queen of Spades example, where A represented drawing the exact card, B represented drawing the correct suit and C represented getting a queen
- Normally the probability of drawing the queen of spades is equal to 1/52
- However it increases if we know it's a spade since P(A/B) = 1/13 and P(A) is 1/52, we can say the two events are dependent
- Similarly, because the probability of drawing our desired card alters if we know it is a queen we can say A and C are also dependent
- Let's formalize these observations with a formula by definition the conditional probability of an event A given an event B equals,
    - $P(A/B) = \frac{P(A \cap B)}{P(B)}$
- This holds true <u>if the probability of event B is greater than zero.</u>



- So if P(B) is equal to zero then event B would never occur. Thus A given B would not be interpretable.
- Now let's look at the conditional probability formula more closely if we compare it to the favourable-overall formula

$$P(A)= \frac{favourable}{all}$$

We have been using so far, we can see many similarities to satisfy the conditional probability we need both events B and A to occur simultaneously.

- This suggests that the intersection of A and B would consist of all favourable outcomes for this probability, secondly the conditional probability requires that event B occurs.
- So the sample space would simply be all outcomes where event B is satisfied
- Firstly, try to remember that <u>the order in which we write the elements for the conditional probability is crucial.</u>
- P(A/B) is definitely not the same as P (B/A) even if the numeric values are equal
- For example, consider the characteristics of Hamilton College's Class of 2018. The 5% of the students who got a degree in economics graduated with honors
    - P(E/H) = 5%

  at the same time 5 % of all students who graduated with honors completed a concentration in economics

    - P(H/E) = 5%

  these two statements might have the same conditional probability but they hold completely different meanings
- In particular the first one suggests that only four of the 80 economics majors graduated with distinction.
- The second one suggests that four out of the 80 students who graduated with high grades completed a degree in economics

Ex 1

Suppose you draw two cards from a deck and you win if you get a jack followed by an ace (without replacement). What is the probability of winning, given we know that you got a jack in the first turn?

Ex 2:

Suppose you have a jar containing 6 marbles – 3 black and 3 white. What is the probability of getting a black given the first one was black too without replacement

Marginal Probability

- Contingency table consists of rows and columns of two attributes at different levels with frequencies or number in each of the cells.
- It is matrix of frequencies assigned to rows and columns.
- The term marginal is used to indicate that the probabilities are calculated using a contingency table.
- Also called as <u>joint probability table.</u>

A survey of 200 families were conducted. Information regarding family income per year & whether family buys a car is given in following table

Family	Income below Rs 10 lakh	Income above Rs 10lakh	Total
Buyer of car	38	42	80
Non buyer	82	38	120
Total	120	80	200

a) What is the probability that a randomly selected family is buyer of the car?

b) What is the probability that a family is both buyer of the car & belonging to income of Rs 10 lakh & above.

c) A family selected at random is found to be belonging to income of Rs 10 lakh & above. What is the probability that this family is buyer of car?

Example 3

A research group collected the yearly data of road accidents with respect to the conditions of following and not following the traffic rules of an accident prone area. They are interested in calculating the probability of accident given that a person followed the traffic rules. The table of the data is given as follows:

Condition	Follow Traffic Rule	Does not follow Traffic Rule
Accident	50	500
No Accident	2000	5000

The Law of Total Probability

- Many scientific papers rely on conducting experiments or surveys
- They often provide summarized statistics we use to analyse and interpret how certain factors affect one another.
- An example would illustrate this better imagine you conducted a survey where 100 men and women of all ages were asked if they eat meat.
- The results are summarized in the table, we see 15 of the 47 women that participated are vegetarian as our 29 out of the 53 men.

			total
♀	15	32	47
♂	29	24	53
total	44	56	100

- Now if A represents being vegetarian and B represents being a woman then P (A/B) and P(B/A) expressed different events
- The former equals 15/47 represents the likelihood of a woman being vegetarian while the other equals 15/44 expresses the likelihood of a vegetarian being a woman since 15/44 is greater than 15/47.

$$P(A/B)= 15/47 \qquad \neq \qquad P(B/A) = 15/44$$

- It is more likely for a vegetarian to be female than for a woman not to eat meat.
- It shows you that in probability theory things are never straightforward

- Now we will discuss important concept the law of total probability.
- Imagine A is the union of some finitely many events B1, B2 and so on.

A=B1 U B2 U …….. U Bn

- This law dictates that the P(A) is the sum of all the conditional probabilities of a given some B multiplied by the probability of the associated B

P(A) = P(A/B1) * P(B1) + P(A/B2) * P(B2) …………

Let us go back to the survey example. The probability of being vegetarian equals

Law of total probability

The **law of total probability** dictates that for any set A, which is a union of many mutually exclusive sets $B_1, B_2, \dots, B_n$, its probability equals the following sum.

$$P(A) = \underbrace{P(A|B_1)}_{\text{Conditional Probability of A, given } B_1 \text{ has occurred.}} \times \underbrace{P(B_1)}_{\text{Probability of } B_1 \text{ occurring.}} + \underbrace{P(A|B_2)}_{\text{Conditional Probability of A, given } B_2 \text{ has occurred.}} \times \underbrace{P(B_2)}_{\text{Probability of } B_2 \text{ occurring.}} + \cdots + P(A|B_n) \times P(B_n)$$

Intuition behind the formula:

- Since P(A) is the union of mutually exclusive sets, so it equals the **sum of the individual sets**.
- The **intersection** of a union and one of its subsets is the entire subset.
- We can rewrite the conditional probability formula $\boldsymbol{P(A|B) = \frac{P(A \cap B)}{P(B)}}$ to get $\boldsymbol{P(A \cap B) = P(A|B) \times P(B)}$.
- Another way to express the law of total probability is:
  - $P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_n)$

P(vegetarian) =P(vegetarian/male)*P(male) + P(vegetarian/female)*P(female)

= 29/53*53/100 + 15/47 * 47/100

= 0.44

			total
♀	15	32	47
♂	29	24	53
total	44	56	100

The Additive Rule

- The additive law recall when we introduced the concept of unions
- The union of two events A and B is equal to their sum minus their intersection
- The additive law states something very similar the probability of the union of two sets is equal to the sum of the individual probabilities of each event minus the probability of their intersection.

Additive Law

The additive law calculates the probability of the union based on the probability of the individual sets it accounts for.

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

Probability of the union

Probability of the intersection

Intuition behind the formula

- Recall the formula for finding the size of the Union using the size of the Intersection:
  - $A \cup B = A + B - A \cap B$
- The probability of each one is simply its size over the size of the sample space.
- This holds true for any events A and B.

- The union of women and vegetarians equal the sum of probabilities of being a woman and being a vegetarian minus the probability of being a vegetarian woman

P(women U vegetarian) = P(women) + P(vegetarian) – P(being a vegetarian women)

$P(W \cup V) = P(W) + P(V) - P(W \cap V)$

$= 47/100 + 44/100 - 15/100$

$= 0.47 + 0.44 - 0.15$

$= 0.76$

- Thus if we picked a random person from the survey there is a 76% chance they're either female vegetarian or both.
- Now consider a different example.

			total
♀	15	32	47
♂	29	24	53
total	44	56	100

- Now consider a different example.
- Suppose we know 38% of our colleagues can proficiently use Tableau and 45% are experts in SQL
- Additionally 66% of the people in the office are good with at least one of the two. What is the probability of somebody being able to simultaneously implement SQL and Tableau
- To answer this, We can rearrange the additive law to get the intersection of Tableau and SQL users equals the

P(Tableau) = 38%

P(SQL) = 45%

P(Tableau U SQL) = 66%

- According to additive rule,

$P(T \cup S) = P(T) + P(S) – P(T \cap S)$

$P(T \cap S) = P(T) + P(S) – P(T \cup S)$

$= 38\% + 45\% - 66\%$

$= 17\%$

Transforming this into a probability gives us a likelihood of 0.17 for somebody in the office to be able to proficiently implement SQL and Tableau

Example:

1) From a pack of well shuffled cards, a card is picked up at random.

- **a) What is probability that the selected card is a king or a queen?**
- **b) What is probability that the selected card is king or a diamond?**

2) The probability that you will get an A grade in quantitative method is 0.7. The probability that you get an A grade in marketing is 0.5. Assuming these 2 courses are independent, compute the probability that you will get an A grade in both this subjects.

The Multiplication Law

- Let us examine the conditional probability formula once again

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

- We can multiply both sides of the equation by P(B) to get the probability of the intersection of A and B

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

$$\mathbf{P(A\cap B) = P(A/B) * P(B)}$$

- We call this equation the **multiplication rule**

The Multiplication Rule

The multiplication rule calculates the probability of the intersection based on the conditional probability.

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

Probability of the Intersection

Conditional Probability

Probability of event B

Intuition behind the formula

- We can multiply both sides of the conditional probability formula $P(A|B) = \frac{P(A\cap B)}{P(B)}$ by $P(B)$ to get $P(A \cap B) = P(A|B) \times P(B)$.
- If event B occurs in 40% of the time ($P(B) = 0.4$) and event A occurs in 50% of the time B occurs ($P(A|B) = 0.5$), then they would simultaneously occur in 20% of the time ($P(A|B) \times P(B) = 0.5 \times 0.4 = 0.2$).

- Consider a numerical example and see why this makes sense.
- Suppose the probability of event B is 0.5 and the probability of event A given B, P(A/B) is 0.8.
- This suggests that event B occurs 50% of the time and event A also appears in 80% of those 50% when B occurred.
- Therefore, the likelihood of A and B occurring simultaneously is 0.8*0.5 or 0.4

Example:

Suppose we draw two cards from a standard deck of 52 playing cards we draw one. Shuffle the deck without returning the card and then draw a second one. What is the probability of drawing a spade on the second draw and not drawing a spade on the first draw?

If we express these as a single conditional probability event A would be drawing a spade on the second try and event B would be not drawing a spade on the first try.

As stated before the likelihood of drawing a specific suit is 1/4th or 0.25.

We already discussed how to calculate the probability of complements so the probability of B would equal,

$P(B) = 1 - 0.25 = 0.75$

Now be careful when estimating the probability of drawing a spade on the second turn.

There are only 51 cards left. So we must adjust the favourable overall formula to find the new likelihood.

We have assumed we did not draw a spade on the first go so the favourable outcomes would still be the 13 spades left.

However, we are one card short from having a complete deck. So the new sample space would be 51.

Therefore, the probability would be,

$P(A/B) = 13/51 = 0.255.$

So far we have calculated the likelihood of not drawing a spade on the first turn and the probability of drawing a spade on the second go.

Given we do something else first however we still haven't answered the question we are interested in.

What is the probability of drawing a spade on the second draw and not drawing a spade on the first draw to answer it.

We need to apply the multiplication rule,

$$\begin{aligned} \mathbf{P(A \cap B)} &= \mathbf{P(A/B) * P(B)} \\ &= \mathbf{0.255 * 0.75} \\ &= \mathbf{0.191} \end{aligned}$$

So we have a probability of 0.191 of drawing a spade on the second turn assuming we did not draw 1 initially

- Example:

From a pack of cards, 2 cards are drawn in succession one after another. After every draw, the selected card is not replaced. What is probability that in both the draws you will get spades?

Bayes' Law

- Here we are going to introduce one of the most important formulas in the world of probability.
- The Bayes rule people also refer to it as Bayes theorem or Bayes law.
- For starters take two events A and B according to the conditional probability formula.

$$P(A/B) = P(A \cap \mathbf{B})\ /\ \mathbf{P(B)} \text{ ------------------- } \mathbf{Eq1}$$

- The conditional probability of A given B equals the probability of their intersection over the probability of event B
- Using the multiplication rule, we can write,

$$\mathbf{P(A \cap B) = P(B/A) * P(A)} \text{ ------------------------} \mathbf{Eq2}$$

- We can transform the numerator of Eq1 to get the probability of the intersection of A and B equals,

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

From Eq2

$$\mathbf{P(A/B) = \frac{P\left(\frac{B}{A}\right) * P(A)}{P(B)}}$$

- This equation is known as **Bayes theorem**.
- It is crucial because it allows us to find a relationship between the different conditional probabilities of two events

Bayes' Law

Bayes' Law helps us understand the relationship between two events by computing the different conditional probabilities. We also call it Bayes' Rule or Bayes' Theorem.

Conditional probability of B, given A.

Conditional probability of A, given B. $\longrightarrow$ $$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Intuition behind the formula

- According to the multiplication rule $P(A \cap B) = P(A|B) \times P(B)$, so $P(B \cap A) = P(B|A) \times P(A)$.
- Since $P(A \cap B) = P(B \cap A)$, we plug in $P(B|A) \times P(A)$ for $P(A \cap B)$ in the conditional probability formula $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- Bayes' Law is often used in medical or business analysis to determine which of two symptoms affects the other one more.

- One of the most prominent examples of using Bayes Rule is in medical research when trying to find a causal relationship between symptoms.
- Knowing both conditional probabilities between the two helps us make more reasonable arguments about which one causes the other.
- For instance, there is certain correlation between patients with back problems and patients wearing glasses.
- More specifically 67% of people with spinal problems wear glasses (P(VI/BP) =67%) while only 41% of patients with eyesight issues have back pains (P(BP/VI)=41%)
- These conditional probabilities suggest that is much more likely for someone with back problems to wear glasses than the other way around even though we cannot find a direct causal link between the two.
- There exists some arguments to support such claims.
- For instance, most patients with back pain are either elderly or work a desk job where they remained stationary for long periods
- Old age and a lot of time in front of the desktop computer can have a deteriorating effect on an individual's eyesight however many healthy and young individuals wear glasses from a young age.
- In those cases, there is no other underlying factor that would suggest incoming back pains.
- Similarly, we can also apply Bayes Theorem in business. Let's explore this fictional scenario.

- Your boss want you to do research about what companies are looking for in recent college graduates.
- Good academic performance or working experience. You go over the resumes of the last 200 people who match the requirements and got the job they applied for.
- Out of those candidates 45% had the relevant experience. In addition 60% had good grades.

  P(Exp) =45%

  P(Good Grades) = 60%

- Furthermore, we know that out of those 45% who had relevant experience 50% also performed well academically

  P(Good Grades/Exp) =50%

  P(Exp/Good Grades) = ?

- To avoid any harsh decisions, we need to compute the conditional probability of the candidate to have relevant experience provided they had a high GPA
- If we use Bayes Theorem we get that the conditional probability of a candidate performing well academically to have relevant experience is

$$P(\text{Exp/Good Grades}) = \frac{\text{P(Good Grades|Exp)} * \text{P(Exp)}}{P(Good\ Grades)}$$

$$= \frac{0.5 * 0.45}{0.6}$$

$$= 0.375$$

- Since 0.5 is greater than 0.375 then it is more likely for an experienced candidate to excel academically than for a student with high grades to have the required working pedigree.
- Thus candidates who had internships are more likely to also have a high GPA therefore firms are much more likely to get their ideal candidate if they go for somebody who has experience rather than somebody who thrived academically

Example1: A bag I contain 4 white and 6 black balls while another Bag II contains 4 white and 3 black balls. One ball is drawn at random from one of the bags, and it is found to be black. Find the probability that it was drawn from Bag I.

Example 2:

You are planning a picnic today, but the morning is cloudy

- Oh no! 50% of all rainy days start off cloudy!
- But cloudy mornings are common (about 40% of days start cloudy)
- And this is usually a dry month (only 3 of 30 days tend to be rainy, or 10%)

What is the chance of rain during the day?

- Let us say P(Fire) means how often there is fire, and P(Smoke) means how often we see smoke, then:

- P(Fire|Smoke) means how often there is fire when we can see smoke
  P(Smoke|Fire) means how often we can see smoke when there is fire

- **dangerous fires are rare (1%)**
- **but smoke is fairly common (10%) due to barbecues,**
- **and 90% of dangerous fires make smoke**

Discover the **probability of dangerous Fire when there is Smoke**