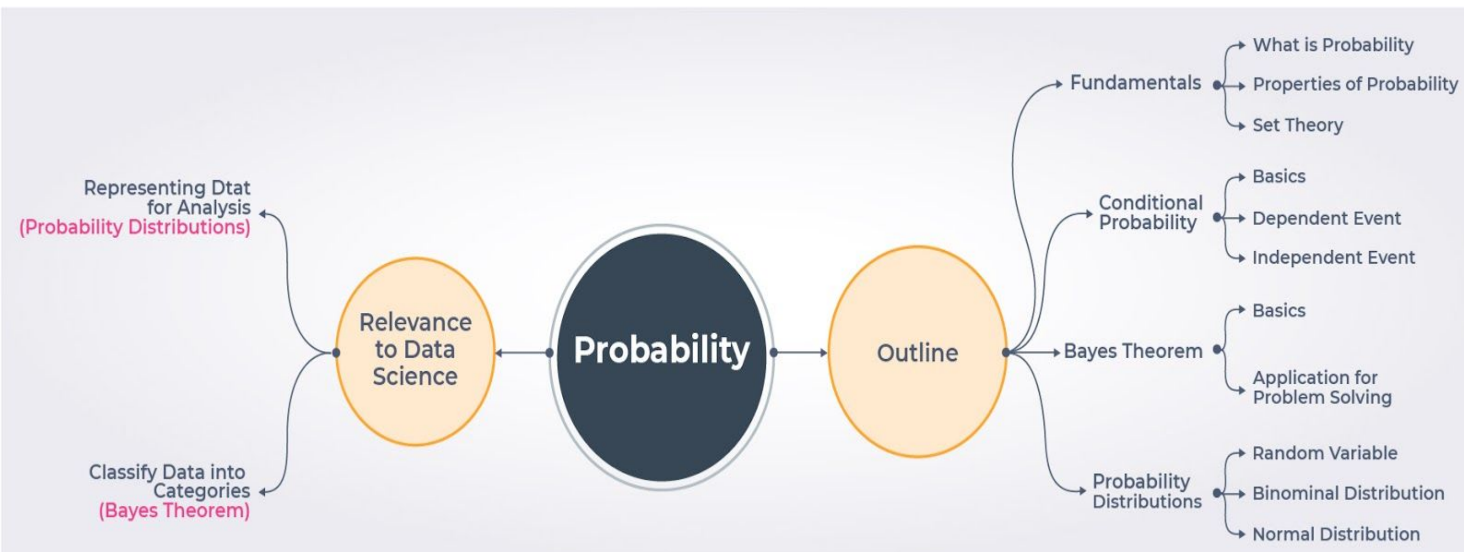
Big picture of probability

One of the key responsibilities of a data scientist is to build systems that understand and interact with the real world. Real-world is filled with uncertainty - you cannot predict how certain events will turn out. When house owners in the USA defaulted on mortgages, the cascading effects led to the subprime housing crisis further leading to recession worldwide. Therefore, it is very important for a data scientist to model uncertainty.

Now, the next question is - how do we model uncertainty? While we cannot predict events accurately, we can form an idea of how likely the events will or will not occur. We often can't prove that something is true but we can ask how likely other events are or what is the most likely explanation. This kind of modeling is possible with probability.

The following image will give you a gist of the concept of probability you will learn along with the applications to data science. The application to data science will become clear as go through the course. But this would give you a sneak peek into the applications of probability to data science.



Before you take a deep dive and understand the concept in detail, have a look at the primer video on probability.

In particular, probability models uncertainty w.r.t time. Uncertainty in time corresponds to those events where the uncertainty is resolved with the passage of time. Think of the following questions:-

- What will the weather be like in the morning?

- Will I like the movie 'Inception'?

Source: Amazon

- What will be the price of the Microsoft stock tomorrow?

Source: Business Insider

The uncertainty in these situations is resolved if we wait a certain amount of time. Once the morning passes, the uncertainty in weather is resolved. Once you watch the movie, you will know whether you liked the movie or not. And the next day will bring you the news regarding the price of your favorite stock.

Let's take the first example in the listed questions - What will the weather be like in the morning? To answer this question, the first thing we need to do is to ask ourselves what are the likely events that can occur. In this case, it is simple - rainy, sunny or cloudy. Only one of these events will occur. The events together form what is called a sample space.

Sample space is defined as the combination (set) of all possible outcomes.

For the first question, the sample space can be defined as

$$S=\{Sunny,Rainy,Cloudy\}$$

$$S=\{Sunny,Rainy,Cloudy\}$$

Similarly, for the second question, where we ask whether you will like the movie, the sample space can be defined as

$$S=\{Yes,No\}$$

$$S=\{Yes,No\}$$

For the third question, the sample space is not so straightforward. The stock could be at any value. Something disastrous could happen and the stock could fall to zero, or a great positive development could send the price skyrocketing. So the way to define the sample space is

$$S=\{x \mid x > 0\}$$

$$S=\{x \mid x>0\}$$

Hence in order to define the sample space, we first need to formulate the questions that need answering. Asking the right questions is a crucial skill for a data scientist. Sample space can contain different types of data i.e. discrete, continuous, or qualitative (categorical). We have learned how to summarize these different types of data in descriptive statistics. Here, we will learn how to answer uncertain questions regarding our data with the help of probability. It is time to define probability formally and explore it further.

Previously we discussed the big picture of probability, let us now define probability. First, look at this video that revises some of the important concepts of probability.

What is Probability?

Uncertainty is all around us. We often ask questions like `What is the chance of ?`. You can simply go to google and type in `What is the chance of` and see what people have been looking up on the internet. Following is a snapshot of some popular search statements.

When we ask a question `What is the chance that it will rain today?`, inherently what we are trying to do is calculate the probability of the `event` raining. Probability is a very intuitive subject but at the same time, it can be treacherous. We will learn various aspects of probability in an intuitive way by looking at some real data. But before starting with the data we will learn about different terminology and concepts through small examples such as the flip of a coin or roll of a die as they are easier to understand.

The first term that we will be learning about is `event`. An `event` is defined as a subset of the `sample space`. Suppose we ask a randomly selected person "how many credit cards do you own?". In this scenario our sample space would technically include all the integer values greater or equal to 0 i.e.

$$S = \{0,1,2,3,\ldots.\}$$

$$S=\{0,1,2,3,....\}$$

Now, let

- A be the event that a randomly selected person hold no credit card i.e.

  $A= \{0\}$

- B be the event that a randomly selected person hold upto 3 credit cards i.e.

  $B=\{0,1,2,3\}$

- C be the event that a randomly selected person hold even number of credit cards i.e.

  $C = \{0,2,4,6,\ldots\}$

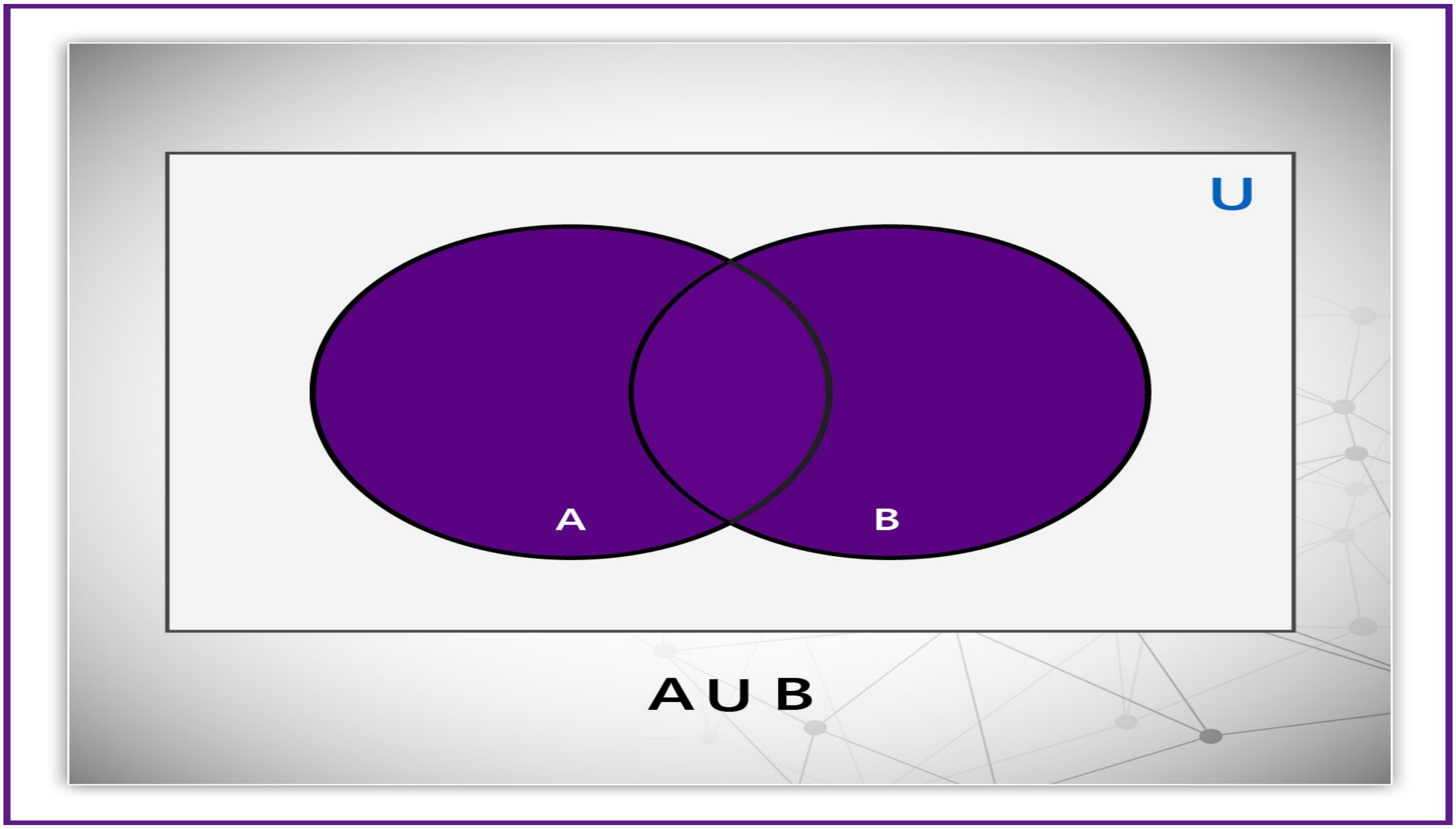
- D be the event that a randomly selected person hold odd number of credit cards i.e.

  $D = \{1,3,5,\ldots\}$

As you can see, these events are nothing but sets. Let us review basic set operations before moving ahead as they will be very helpful going ahead to calculate all the probabilities.

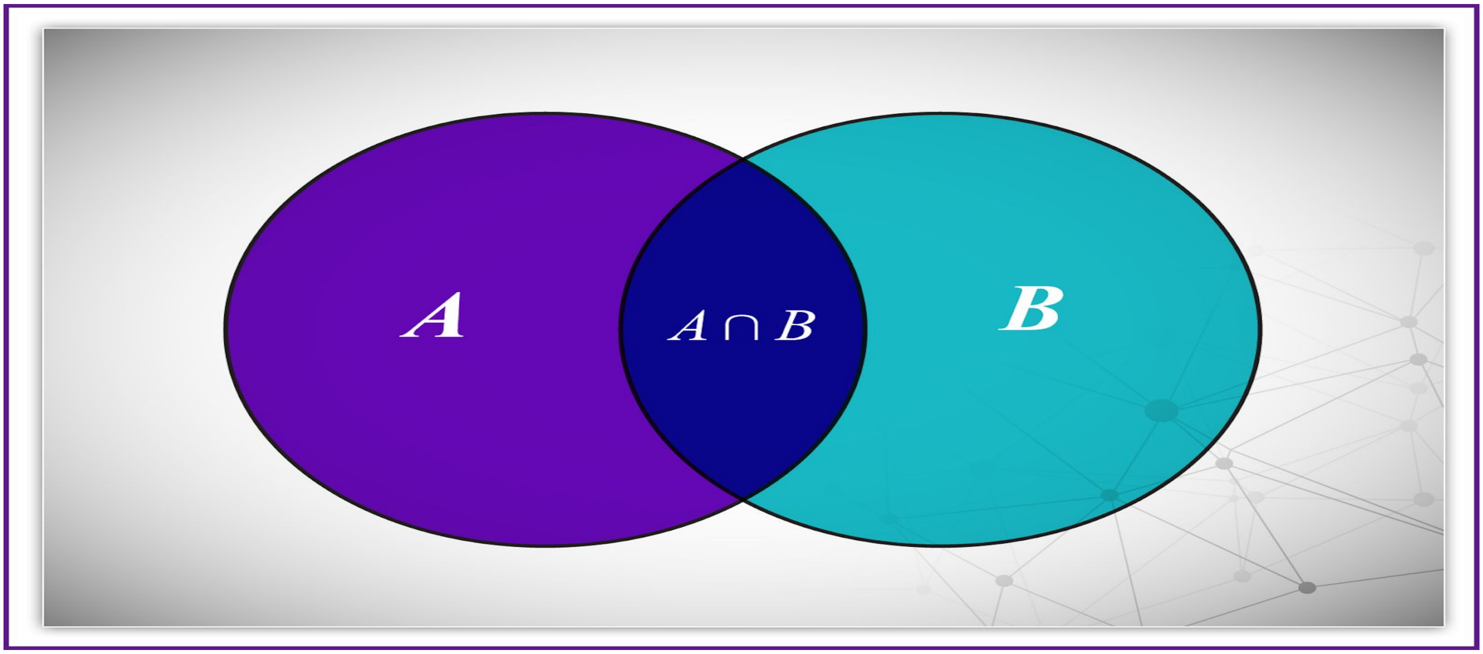
- **Union (∪)** - It is the combination of all the elements of the sets. Ex:

$$A \cup B=\{0\} \cup \{0,1,2,3\}=\{0,1,2,3\}$$



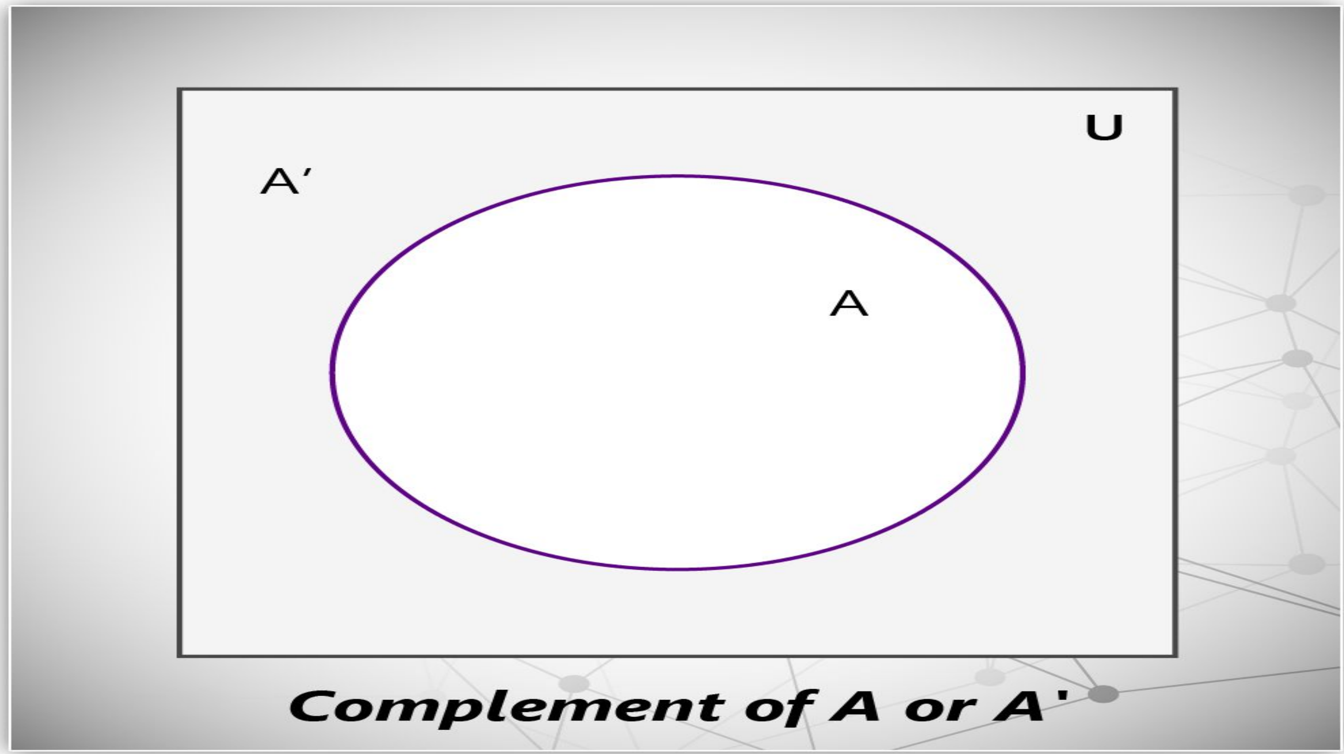
- **Intersection (∩)** - It is the combination of all the common elements of the sets. Ex:

$$B \cap C = \{0,1,2,3\} \cap \{0,2,4,6,...\} = \{0,2\}$$

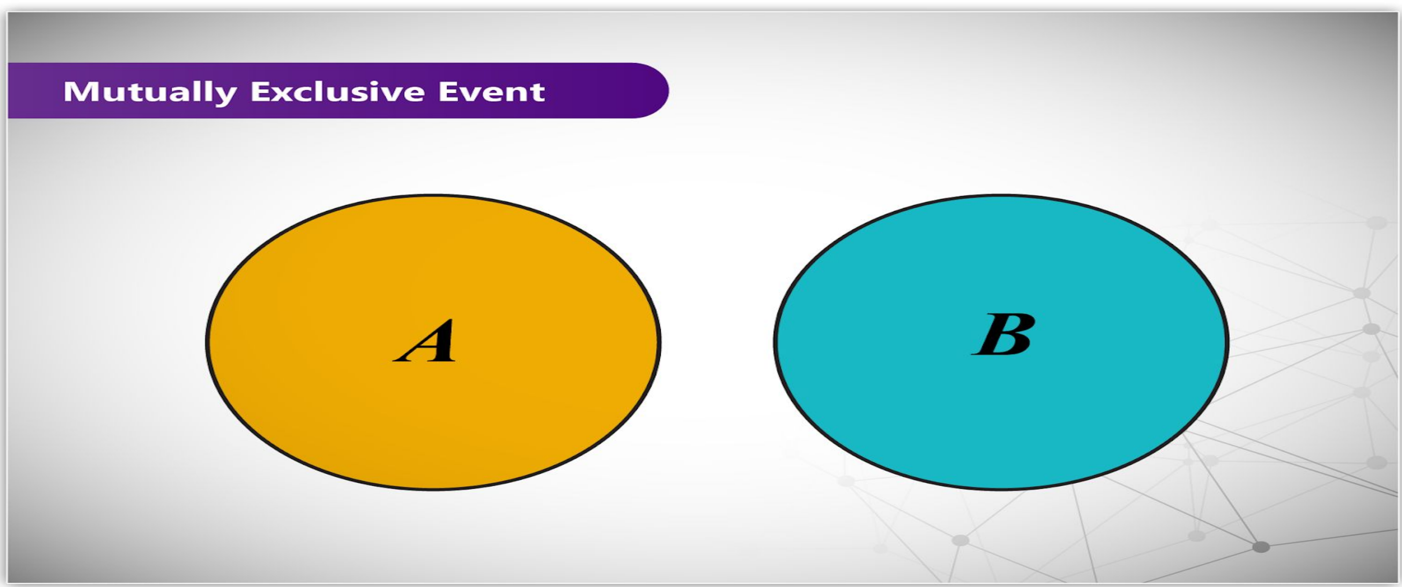


- Complement - It is defined as the elements 'not' in the set. Ex: Complement of `c` is denoted by $\bar{C}$ and can be written as

$$\bar{C} = \{1,3,5,...\}$$



- **Mutually Exclusive** - Two sets are mutually exclusive if their intersection is a null set. Ex. $A \cap D = \{0\} \cap \{1,3,5,...\} = \{\phi\}$, hence we can say that sets A and D are mutually exclusive.



Now let us discuss various approaches to calculate probability. Following are some of the approaches used to calculate the probability of an event

- Theoretical / Classical Approach
- Frequentist Approach
- Bayesian Approach

We will deal with the Bayesian approach in a later part of this concept. Let us now try to understand the classical and frequentist approach with a simple example of flip of a coin.

Theoretical Approach

In an experiment of the flip of a coin, the sample space can be defined as

$$S=\{H,T\}$$

$$S=\{H,T\}$$

. Now, let `A` be the event that we get a Head i.e.

$$A=\{H\}$$

$$A=\{H\}$$

. In this case, the theoretical approach defines the probability of event `A` as

$$P(A) = \frac{no.\ of\ favourable\ outcomes}{sample\ space} = \frac{N(A)}{N(S)} = \frac{1}{2}$$

Frequentist Approach

The frequentist approach involves conducting an experiment repeatedly and observing the number of times the event occurs. In our example, we will have to perform the following steps to calculate probability using a frequentist approach

- Flip the coin n number of times
- Count the number of time a head appears and let it be denoted by N(A)
- Calculate the probability of getting head when you flip a coin can then be calculated as

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

In a frequentist approach as you increase the number of trials of the experiment, the value of probability will approach the theoretical value.

Properties of probability

In the last topic, we defined probability and discussed different approaches to calculate the same. In this topic, we will discuss different properties and rules of probability. Some of the properties of probability are as follows

- Probability of any event will always lie between 0 and 1
- Probability of sample space is 1
- Probability of the complimentary event is given as

$$P(\bar{A}) = 1 - P(A)$$

- If there are two events A and B, the probability of either event A or event B occurring can be given as $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

A university offers four different graduate degrees: business, education, accounting, and science. Enrollment figures show 25 of their graduate students are in each specialty. Although 50 of the students are female, only 15 are female business majors. If a student is randomly selected from the University's registration database:

- What is the probability that the student is not a business major?
- What is the probability the student is a business or education major?
- What is the probability the student is a female or a business major?

Solution

The probability of a student being a business major we know is 0.25 as there are 25 business-major students out of 100. Now to find the probability of the student is not from business major can simply be calculated as follows

$$P(not\ Bus) = 1 - P(Bus) = 1 - 0.25 = 0.75$$

The probability that the student is a business or education major is a mutually exclusive event. Thus:

$P(Bus \cap Edu) = 0$

$$P(Bus \cup Edu) = P(Bus) + P(Edu) - (Bus \cap Edu) = 0.25 + 0.25 - 0 = 0.5$$

The probability that the student is a female or a business major is not mutually exclusive because the student could be a female business major. Thus:

$$P(Fem \cup Bus) = P(Fem) + P(Bus) - P(Fem \cap Bus) = 0.5 + 0.25 - 0.15 = 0.60$$

Conditional Probability

Before defining the conditional probability formally let us take a look at a simple example to examine the intuition.

Suppose we roll a normal 6 faced die. We know that there are 6 possible outcomes for this experiment viz. sample space

$S=\{1,2,3,4,5,6\}$. Now let us define two events as follows A - event that the outcome is 1, 2 or 3 i.e.

A={1,2,3}

B - event that the outcome is an odd number i.e.

B = {1,3,5}

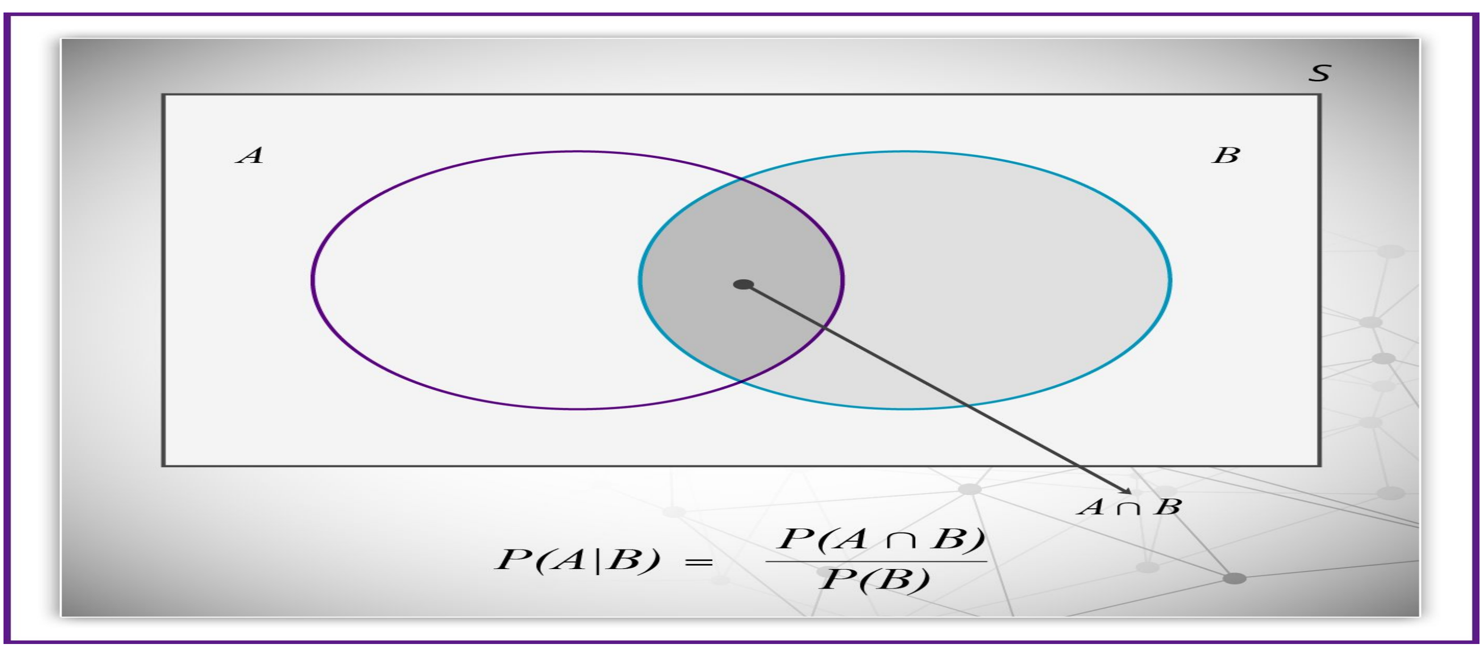
Now let us check the probability of both the events individually,

$$P(A) = \frac{N(A)}{N(S)} = \frac{3}{6} = \frac{1}{2}$$

$$P(B) = \frac{N(B)}{N(S)} = \frac{3}{6} = \frac{1}{2}$$

Now let us change the scenario a little bit. You have rolled the die and one of your friends tells you that an outcome is an odd number (event B), you now wonder what is the probability that the outcome is one of the first three numbers (event A). Will the probability of event A still be ½ as calculated earlier.

No, it will not be the same. In this case, the probability will change as we have some additional information at our disposal. Let us visualize the scenario with the help of a Venn diagram.



We can see that as event B has already occurred we know that the outcome must be one of $\{1,3,5\}$.Out of these, there are only two outcomes that are possible for event A i.e. we can write $A \cap B=\{1,3\}$. Now we can calculate the probability using a familiar method i.e. $\frac{2}{3}$.

What we calculate is nothing but a conditional probability. It is represented as $P(A|B)$ which translates into the probability of event A when we know that event B has already occurred. We can calculate this conditional probability as follows.

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

Based on this equation we can even calculate a probability called joint probability i.e. probability of two events occurring at the same time i.e. $P(A \cap B)$. We can rewrite the above formula as

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

Example:

What is the probability of drawing a king in the first draw and again a king in the second draw as well without replacement?

Ans: 1/221

Explanation:

Event A is drawing a King first, and Event B is drawing a King second.

For the first card the chance of drawing a King is 4 out of 52 (there are 4 Kings in a deck of 52 cards):

P(A) = 4/52

But after removing a King from the deck the probability of the 2nd card drawn is less likely to be a King (only 3 of the 51 cards left are Kings):

P(B) = 3/51

And so: P(A and B) = P(A) x P(B|A) = (4/52) x (3/51) = 12/2652 = 1/221

So the chance of getting 2 Kings is 1 in 221, or about 0.5%

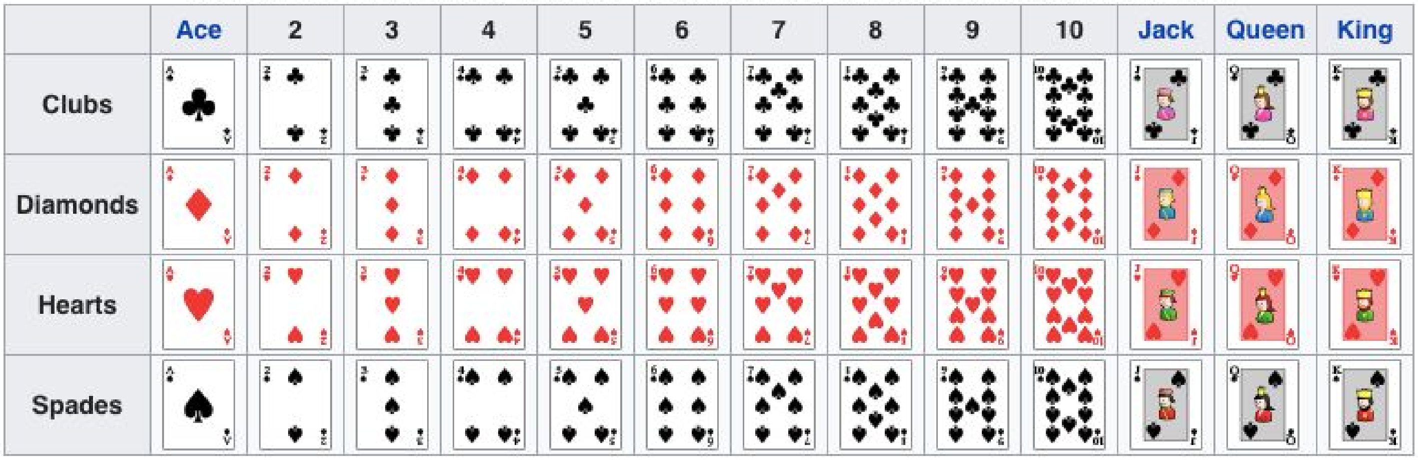
Independent Events

We have defined `event` informally in the previous topics. Following is a more formal definition of the `event`.

An event is a set of outcomes of an experiment to which probability is assigned.

Now that we have defined the event formally, let us discuss different types of events that you would come across while solving problems. We will look at an example of a standard 52 playing cards deck.

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



Independent events

Events are said to be independent if the outcome of one event does not affect the outcome of others.

Ex: Suppose we are to select two cards from the deck of playing cards one after the other. We replace the card we have picked every time before the next draw. In the first draw, we got an ace of hearts, now what is the probability that you draw an ace of hearts in the second draw.

Let us break this problem into steps.

- We have a deck of 52 playing cards hence the `sample space` comprises 52 outcomes. Therefore, $N(S)=52$
- Now let `A` be the event that we draw an ace from the deck. We know that there are 4 aces in a deck, so $N(A) = 4$

- Let us draw the first card and calculate the probability of the event `A`. As discussed in the earlier topics we can calculate the probability as $P(A) = \frac{N(A)}{N(S)} = \frac{4}{52}$.
- Now, we replace the card and draw a second card from the deck, this time if we want to again calculate the probability of event `A` we can simply calculate it as $P(A) = \frac{N(A)}{N(S)} = \frac{4}{52}$.

- We can see that the probability of the second draw does not change depending on the first draw, i.e. we can say that the second draw is `independent` of the first draw.

Dependent events

In contrast to independent events, the events are said to be dependent if the outcome of one event affects the outcome of another. Let us discuss this with a simple example.

Ex: Let us continue with the experiment we did in the previous example. The only difference this time is that we would not replace the card in the deck after each draw. Let us calculate the probability now that you get an ace in the second draw. (Kindly note that the card in the first draw is not an ace and ace is drawn in the second draw.)

Let us break this problem is steps

- We have a deck of 52 playing cards hence the `sample space` comprises 52 outcomes. Therefore, N(S)=52
- Now we draw a card from this deck, let `A`be the event that we draw an ace from the deck. Therefore, N(A)=4

- Hence, the probability of drawing an ace in the first draw is $P(A) = \frac{N(A)}{N(S)} = \frac{4}{52}$
- Now, we are not replacing the card back into the deck, hence the `sample space` for the second draw comprises of only 51 outcomes, $N(S) = 51$
- Let, `B` be the event that we draw an ace in the second draw. We know that there are still 4 aces left in the deck. Hence,

$$P(B) = \frac{N(A)}{N(S)} = \frac{4}{51}$$

As we can see in the example the probability of event B is affected due to event A. Hence events A and B can be termed as dependent events.

Examples:

1.The probability that it is Friday and that a student is absent is 0.03. Since there are 5 school days in a week, the probability that it is Friday is 0.2. What is the probability that a student is absent given that today is Friday?

Ans: 0.15

<u>Explanation:</u>

P(Absent|Friday) = P(Friday and Absent)/P(Friday) = 0.03/0.2 = 0.15

2. Calculate the probability of getting 2 Aces back to back from a deck without replacing the first card?
Ans: 1/221

Explanation:
Event A is drawing a Ace first, and Event B is drawing a Ace second.

For the first card the chance of drawing a Ace is 4 out of 52 (there are 4 Kings in a deck of 52 cards):

P(A) = 4/52

But after removing a King from the deck the probability of the 2nd card drawn is less likely to be a King (only 3 of the 51 cards left are Kings):

P(B|A) = 3/51

And so: P(A and B) = P(A) x P(B|A) = (4/52) x (3/51) = 12/2652 = 1/221

So the chance of getting 2 Kings is 1 in 221, or about 0.5%.

3.Spin a spinner numbered 1 to 7, and toss a coin. What is the probability of getting an odd number on the spinner and a tail on the coin?
Ans: 2/7

Explanation:
When two events, A and B, are independent, the probability of both occurring is:

P(A and B) = P(A) · P(B)

Spinning a number and tossing a coin are two independent events.

P(odd number) = 4/7

P(tail) = 1/2

P(odd number and a tail)= 4/7 x 1/2 = 4/14 = 2/7

4.We have a sample of 38 vehicles, 16 are red, 5 are trucks, and 2 are both. A randomly chosen vehicle is red. What is the probability it is a truck?
Ans: 1/8

Explanation:

P(truck|red) = P(truck and red) / P(red)

P(truck and red) = 2/38

P(red) = 16/38

P(truck|red) = 2/16 = ⅛

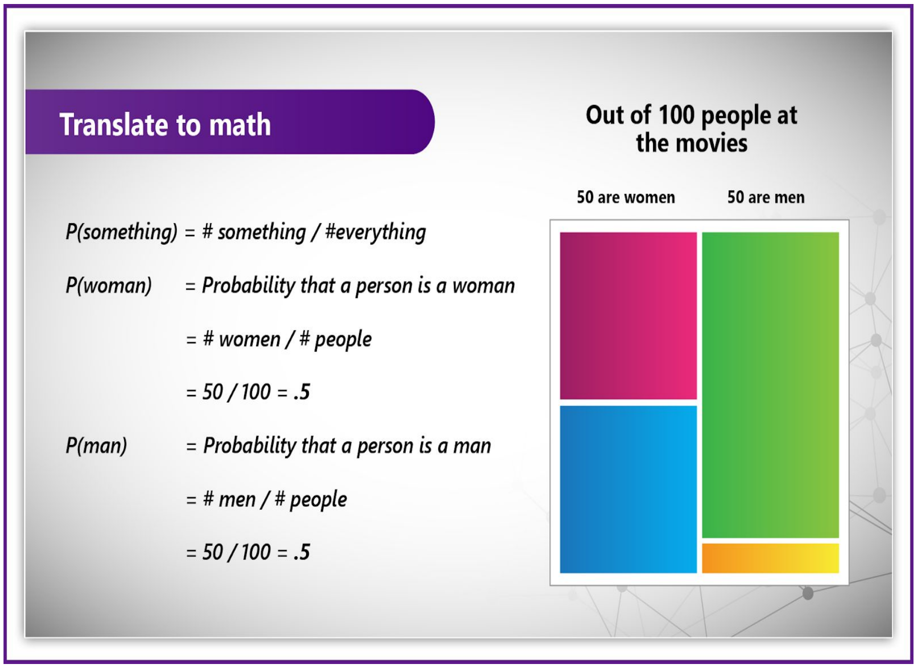
Introduction to Bayes Theorem

Bayes theorem is a different perspective on calculating the probability. In the earlier examples that we saw the probability of an event remains constant over time. In the Bayesian way, the probability of an event changes every time new information with regards to the event arrives. In the previous chapter, we learned how the probability of an event can change based on the information that we already have. Bayes theorem is an extension of conditional probability. Let us look at a video to understand and build intuition for the same.

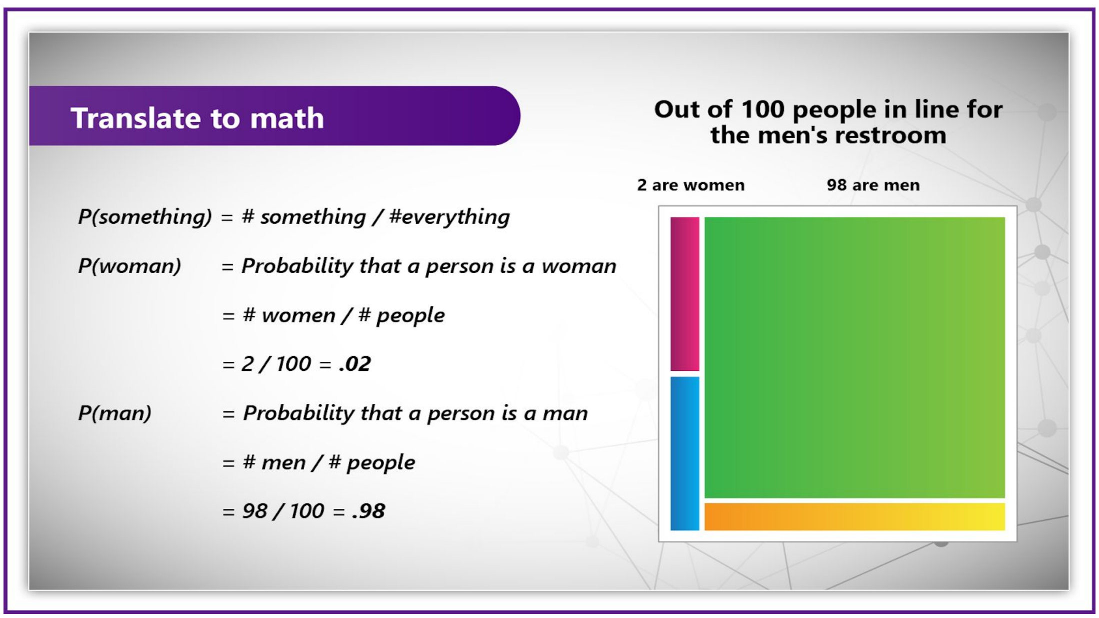
Let's look at the same example in more detail. Suppose you along with your friends are planning to watch a movie. You go to a cinema and see that one of the tickets lying on the ground, you pick up the ticket to search for whom it belongs to. You were only able to get a glimpse and know that the person has long hair. You run towards the person but are now in a dilemma whether to address the person as 'Excuse me ma'am!' or 'Excuse me sir!'. Given what you know in general about hairstyles you would assume that the person is a woman. (this would be a gross oversimplification and generalization).

Now consider that this situation occurs when you are standing in a queue for men's restroom, given this additional piece of information you could safely assume that a person is a man. We make use of this common sense and background knowledge without thinking. Bayes's theorem is the mathematical form to put this in perspective.

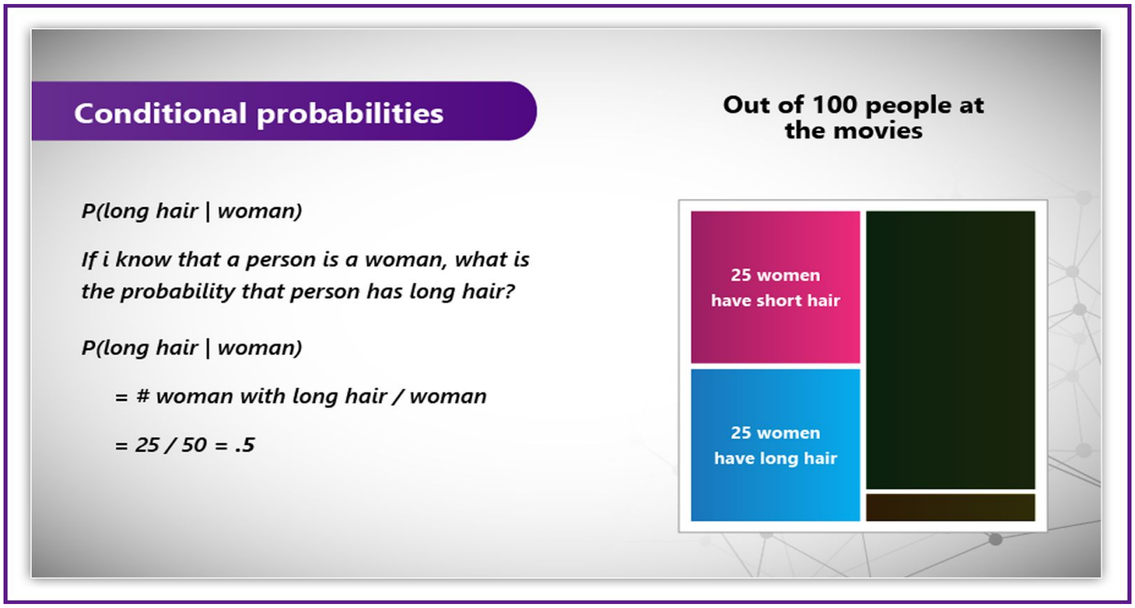
Let us put numbers to the above dilemma and see how the math works. For simplicity let us assume that there are 100 people at the cinema and half of them are men and other half women. Out of the women, half have long hair and half have short hair. Out of men, 48 have short hair and 2 have long hair. Knowing that 25 women have long hair and only 2 men have long hair it is safe to bet that the ticket owner is a woman. The probability that a moviegoer is a male or a female is 0.5 each as shown in the following figure.

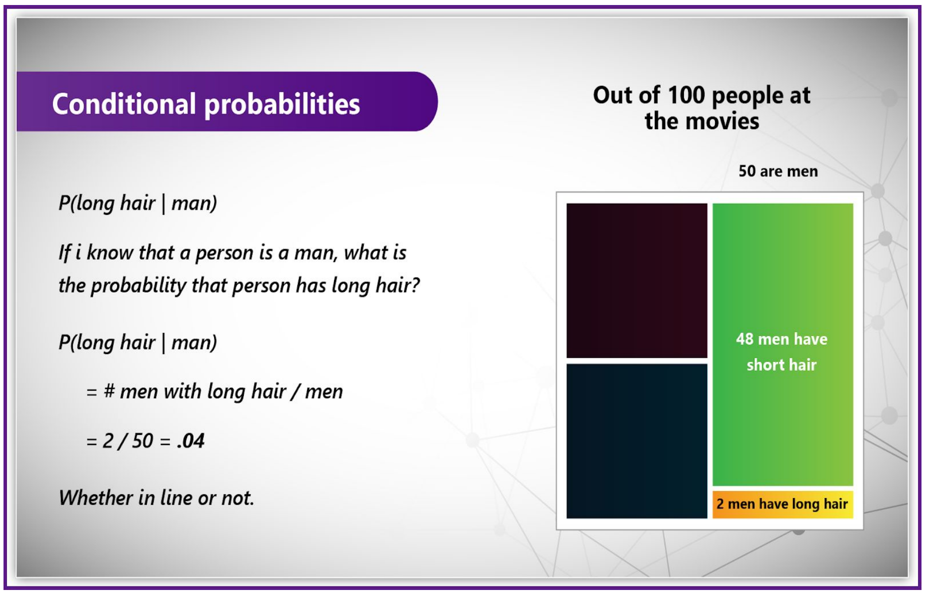


Now at the men's restroom there are 100 people out of which 98 are men and 2 are women who are keeping their partners company. The situation over here is 0.02 for women and 0.98 for men.



Let us move to some conditional probabilities now. Conditional probability answers the question "If I know that the person is a woman, what is the probability that she has long hair?". These probabilities are calculated as follows.



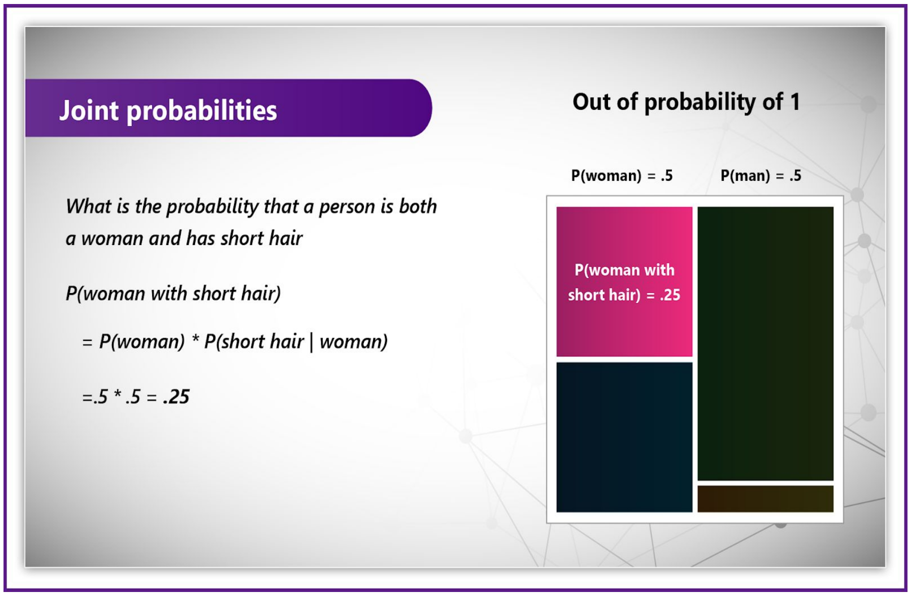


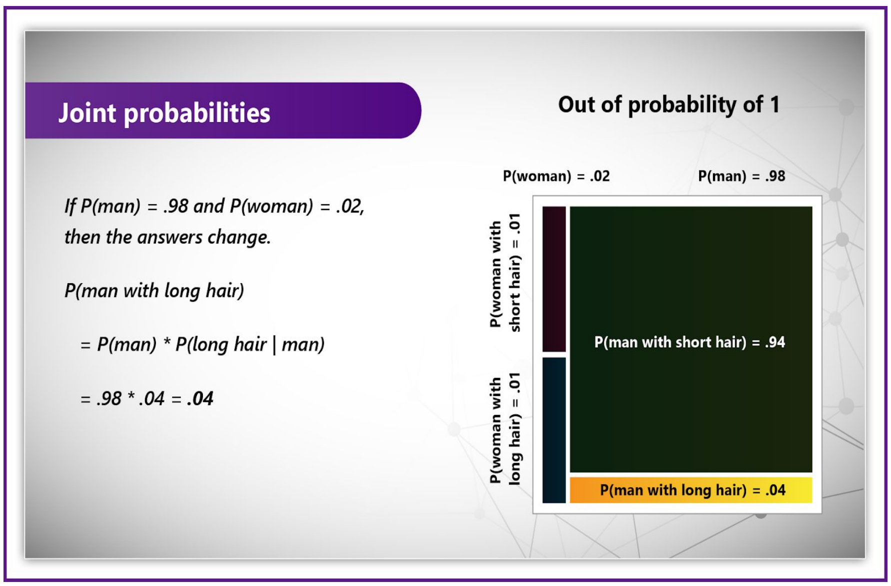
Now moving on to joint probabilities. Joint probabilities help us answer the question "What is the probability that someone is a woman with short hair?".

First, we focus on the probability that someone is a woman, P(woman). Then, we bring in the probability that someone has short hair, given that she is a woman, P(short hair | woman). Combining these by multiplication gives the joint probability,

P(woman with short hair) = P(woman) * P(short hair | woman).

We have computed all these joint probabilities as follows.





The last stop in our journey is marginal probabilities. These are useful for answering the question “What is the probability that someone has long hair?”. To find this, we have to add up the probabilities for all the different ways this could happen - the probability of being a man with long hair plus the probability of being a woman with long hair. Adding up those two joint probabilities gives us P(long hair) of .27 for moviegoers generally, but .05 in the men’s restroom line.

Now that we are armed with the knowledge of different types of probability let us try to answer the dilemma that we faced. "If we know that a person has long hair, what is the probability that the person is a woman (or a man)?" This is an example of conditional probability which can be termed in notations as P(man/long hair), now earlier we have found out the probability of a person having long hair when we know he is a man P(long hair/man), but conditional probabilities are not reversible and hence we need another way to calculate this probability. Thomas Bayes's observations come to our rescue over here.

P(man with long hair) = P(long hair) * P(man/long hair)

P(long hair and man) = P(man) * P(long hair/man)

Because joint probabilities are reversible the above two values are equal and can, therefore, be equated to get the following equation.

$$P(\text{man with long hair}) = \boxed{P(\text{long hair}) * P(\text{man} \mid \text{long hair})}$$

$$P(\text{long hair and man}) = \boxed{P(\text{man}) * P(\text{long hair} \mid \text{man})}$$

$$\text{Because } P(\text{man and long hair}) = P(\text{long hair and man})$$

$$\boxed{P(\text{long hair}) * P(\text{man} \mid \text{long hair})} = \boxed{P(\text{man}) * P(\text{long hair} \mid \text{man})}$$

$$\boxed{P(\text{long hair})} * P(\text{man} \mid \text{long hair}) = P(\text{man}) * P(\text{long hair} \mid \text{man})$$

$$P(\text{man} \mid \text{long hair}) = P(\text{man}) * P(\text{long hair} \mid \text{man}) / \boxed{P(\text{long hair})}$$

The formula that we just derived is nothing but Bayes's theorem.

Applying Bayes theorem

Let's connect the dots on what we have understood in the previous chapter and put forward Bayes Theorem formally.

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

While this can be applied for general probability events, it has a particularly nice interpretation in the case where A represents a hypothesis H and B represents some observed evidence E. In this case, the formula can be written as

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)}$$

Here, what we are looking for is the probability of the hypothesis given the evidence. We obtain this through the probability of the hypothesis before getting the evidence P(H). For this reason, P(H) is called the prior probability, while P(H|E) is called the posterior probability. The factor that relates the two, $\frac{P(E|H)}{P(E)}$, is called the likelihood ratio. Using these terms, Bayes' theorem can be rephrased as "the posterior probability equals the prior probability times the likelihood ratio."

A television set produced by ABC Corp was found to be defective (D). There are two factories (A, B) where such televisions are manufactured. A Quality Control Manager (QCM) is responsible for investigating the source of found defects. This is what the QCM knows about the company's television production and the possible source of defects:

Factory	**% of population**	**probability of defective**
A	40% i.e. P(A) = 0.4	1% i.e. P(D/A) = 0.01
B	60% i.e. P(B) = 0.6	1.5% i.e. P(D/B)=0.015

The QCM would like to answer the following question: If a randomly selected television is defective, what is the probability that the television was manufactured in factory B?

Over we know that the television is defective i.e. event D has already occurred and we now need to find the probability that the television is from factory B i.e. event B. We can rewrite the above statement as P(B|D).

From Bayes theorem, we know that the probability can be calculated as

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

(In this case, think about what is the evidence and what is the hypothesis.) Let us calculate the numerator and denominator separately

The numerator is nothing but a joint probability. It can be calculated as follows

$$P(B \cap D) = P(D|B) * P(B) = 0.015 * 0.6 = 0.009$$

Now moving to the denominator, we do not have the probability of a randomly selected television set to be defective i.e. P(D). We have probabilities of television being defective from each of the factories A and B. We can combine theses probabilities as follows

$$P(D) = P((D \cap A)\hat{P}(D \cap B)) = P(D|A) * P(A) + P(D|B) * P(B)$$

$$P(D) = 0.01 * 0.4 + 0.015 * 0.6 = 0.013$$

Now as we have values of both numerator and denominator we can calculate the required probability

$$P(B|D) = \frac{0.009}{0.013} = 0.6923$$

Therefore there is approximately 70% chance that the defective television has been manufactured at factory B.

The clarity in the Bayes theorem improves with a lot of problem-solving. Solve more of the exercise problems to understand Bayes theorem properly.

Random variable

Variable is a value that can change over the course of the experiment. When there is a probability associated with the values that can be taken by a variable, it is called a random variable.

Let's understand more about random variables with the help of the following video - Random Variables explained. Credits to Dr Nic's Math and Stats YouTube Channel.

Ex: When we roll a die the outcome of the experiment is a variable but as we know there is an equal probability of ⅙ for each of the outcomes we can say that the outcome of a roll of a die is a random variable

There are two distinct types of random variables

A random variable X is a discrete random variable if:

- there are a finite number of possible outcomes of X, or
- there are a countably infinite number of possible outcomes of X

A random variable X is a continuous random variable if:

- there are an uncountable infinite number of possible outcomes of X

Generally discrete variable is a variable whose value is obtained by counting and a continuous variable as a variable whose value is obtained by measuring.

Example: Discrete variables

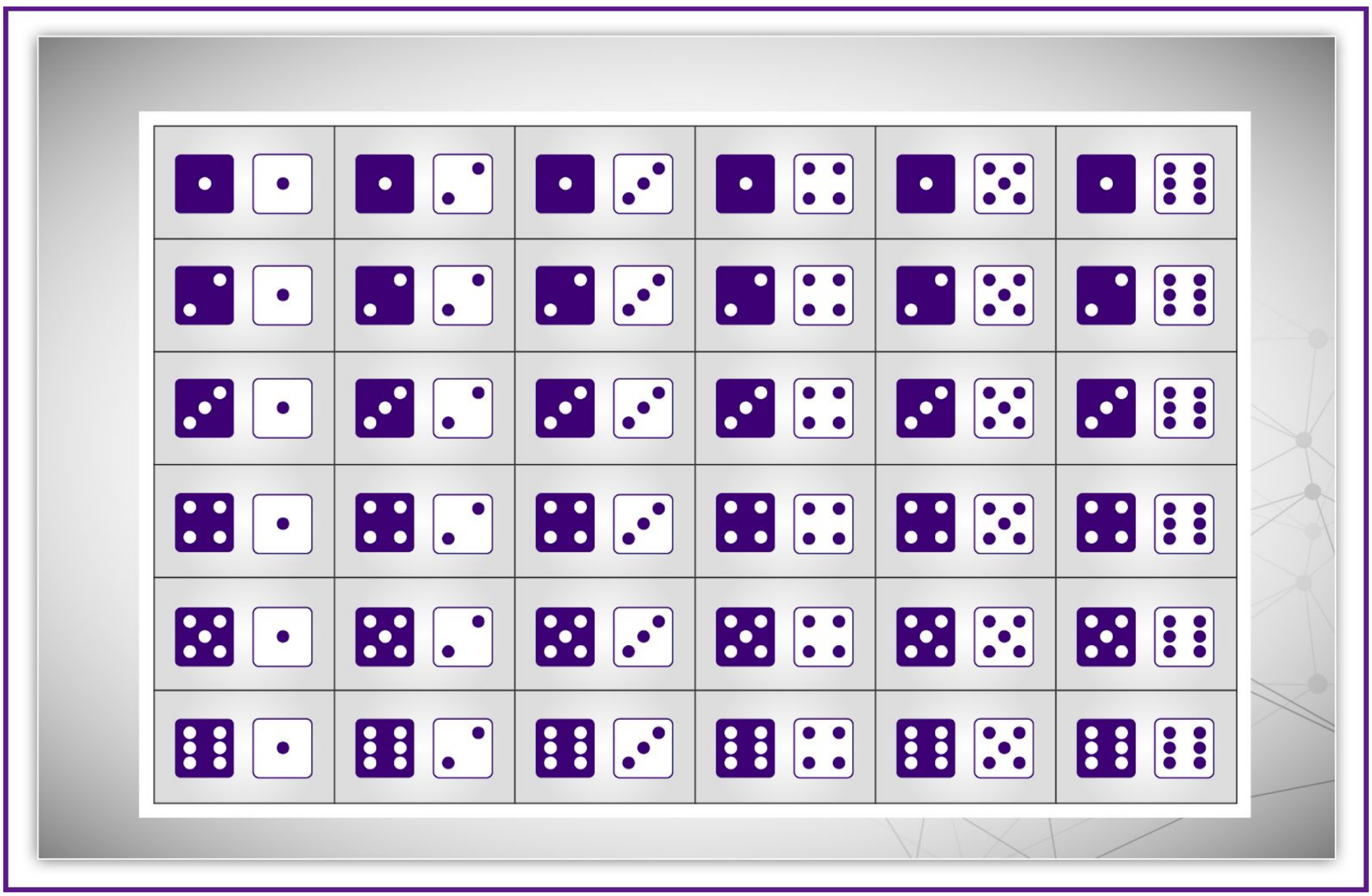
- Number of passengers traveling from an airport
- Number of customers at the store each day

Continuous variables

- Weight of people in a city
- Time take to travel between two cities
- Fuel consumption of an aeroplane

Probability Distributions

Consider an experiment of we rolling two dies. Let X be the random variable representing the sum of two six-sided dice throws. There are 6 possibilities in the first throw we can get any no. from 1 to 6 and similar 6 in the second throw. Therefore we can get a total of 36 combinations. These 36 combinations are enlisted in the table below



The random variable X represents the sum of these combinations. If we look at the outcomes closely there are only 11 values that X can take, these are

$$S = \{2,3,4,5,6,7,8,9,10,11,12\}$$

$$S=2,3,4,5,6,7,8,9,10,11,12$$

Now, let us calculate the probabilities for each of these possible outcomes of random variable X

P(X=2) = 1/36 => (1,1)

$P(X=2)=1/36=>(1,1)$

P(X=3) = 2/36 => (1,2),(2,1)

$P(X=3)=2/36=>(1,2),(2,1)$

P(X=4) = 3/36 => (2,2),(3,1),(1,3)

$P(X=4)=3/36=>(2,2),(3,1),(1,3)$

P(X=5) = 4/36 => (1,4),(4,1),(2,3),(3,2)

$P(X=5)=4/36=>(1,4),(4,1),(2,3),(3,2)$

P(x=6) = 5/36 => (3,3),(1,5),(5,1),(2,4),(4,2)

$P(x=6)=5/36=>(3,3),(1,5),(5,1),(2,4),(4,2)$

P(X=7) = 6/36 => (1,6),(6,1),(2,5),(5,2),(3,4),(4,3)

$P(X=7)=6/36=>(1,6),(6,1),(2,5),(5,2),(3,4),(4,3)$

P(X=8) = 5/36 => (2,6),(6,2),(3,5),(5,3),(4,4)

$P(X=8)=5/36=>(2,6),(6,2),(3,5),(5,3),(4,4)$

P(X=9) = 4/36 => (3,6),(6,3),(5,4),(4,5)

$P(X=9)=4/36=>(3,6),(6,3),(5,4),(4,5)$

P(X=10) = 3/36 => (4,6),(6,4),(5,5)

$P(X=10)=3/36=>(4,6),(6,4),(5,5)$

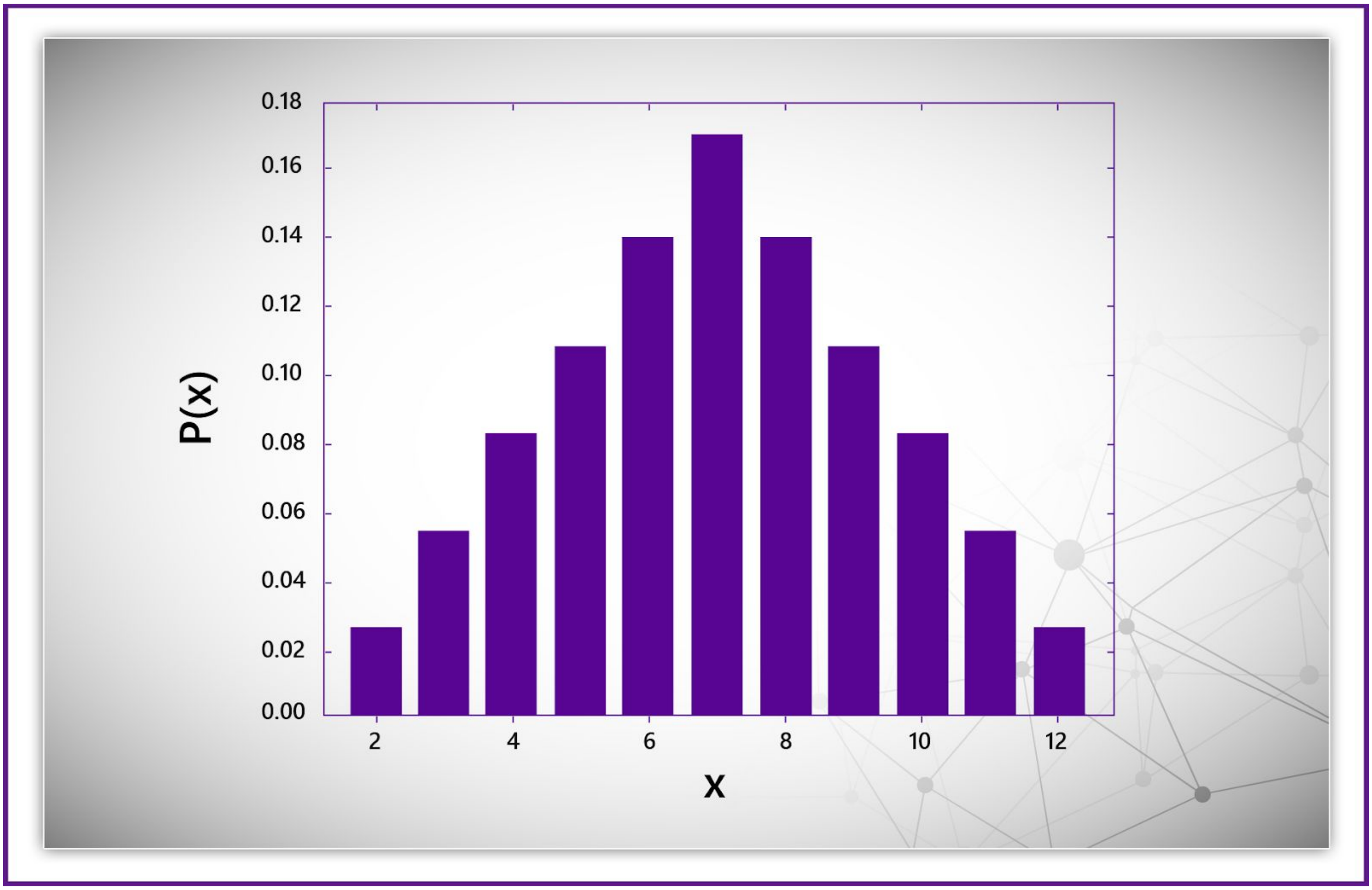
P(X=11) = 2/36 => (5,6),(6,5)

$P(X=11)=2/36=>(5,6),(6,5)$

P(X=12) = 1/36 => (6,6)

$P(X=12)=1/36 \Rightarrow (6,6)$

If we plot the above-derived probabilities on a graph we get a probability distribution



Now that we know what a probability distribution is, let us take the discussion forward.

- The above experiment of rolling two dies was fairly simple and we could easily derive the probability distribution for the same.
- But, it would be very tedious to derive a probability distribution every time we have a different experiment.
- Luckily, there are enough similarities between a certain type of experiments for which we can develop a distribution which would represent their general characteristics.
- We will now take a look at some of these general probability distributions which are widely used in statistical computations

Binomial Distribution

- It would be a very tedious task to derive a probability distribution every time we have a different experiment.
- Luckily, there are enough similarities between a certain type of experiment for which we can develop a distribution which would represent their general characteristics.
- We will not take a look at some of these general probability distributions which are widely used in statistical computations

The binomial distribution is a discrete probability distribution. Let us now derive a binomial distribution using a simple experiment and then we will proceed to the mathematical formula.

- Consider kids playing in a pit full of red and blue coloured balls. We will consider that on an average 25% of the balls in the pit are red. Now if I pick up 5 balls randomly from the pit, what is the probability that 2 of them are red in colour.
- Now one possible outcome of the random pick could be Red - Red - Blue - Blue - Blue (RRBBB). Let us calculate the probability for this outcome.
- We know from the given data that

- We know from the given data that $P(R) = \frac{1}{4}$ and $P(B) = P(\bar{R}) = 1 - P(R) = \frac{3}{4}$. From the axioms of probability that we have studied earlier we can write

$$P(RRBBB) = P(R) \cdot P(R) \cdot P(B) \cdot P(B) \cdot P(B) = \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} = \left(\frac{1}{4}\right)^2 \cdot \left(\frac{3}{4}\right)^3 = 0.026$$

- Now there are 9 other possibilities of which satisfy our requirement of getting 2 red balls.

$$RBRBB, RBBRB, RBBBR, BRRBB, BRBRB, BRBBR, BBRRB, BBRBR, BBBRR$$

The probabilities for all these outcomes can be calculated in the above manner and the value would be same i.e. 0.026

- So now we know that there are 10 different possibilities of a favourable outcome of our experiment. We can thus calculate the probability of getting 2 red balls in a random pick of 5 balls by simply adding up these probabilities or we can write it as

$$P(2\ red\ balls) = 10 * 0.026$$

- If we observe carefully we can calculate the above probability by simply multiplying 0.026 by **number of arrangement possible** i.e. a combination calculation given by ${}^{5}_{2}\mathrm{C}$

- Remember we can calculate the number of arrangements possible as follow

$${}^{n}_{r}\mathrm{C} = \frac{n!}{r! \cdot (n-r)!}$$

- From above observations we can now write down the **probability mass function** for Binomial Distribution

$$P = {}^{n}_{r}\mathrm{C} \cdot p^r \cdot q^{n-r}$$

where
n = total number of items in the group
r = number of items in the group of the desired type
p = probability that any given item in the group is of the desired type
q = probability that any given item in the group is not of the desired type i.e. (1-p)

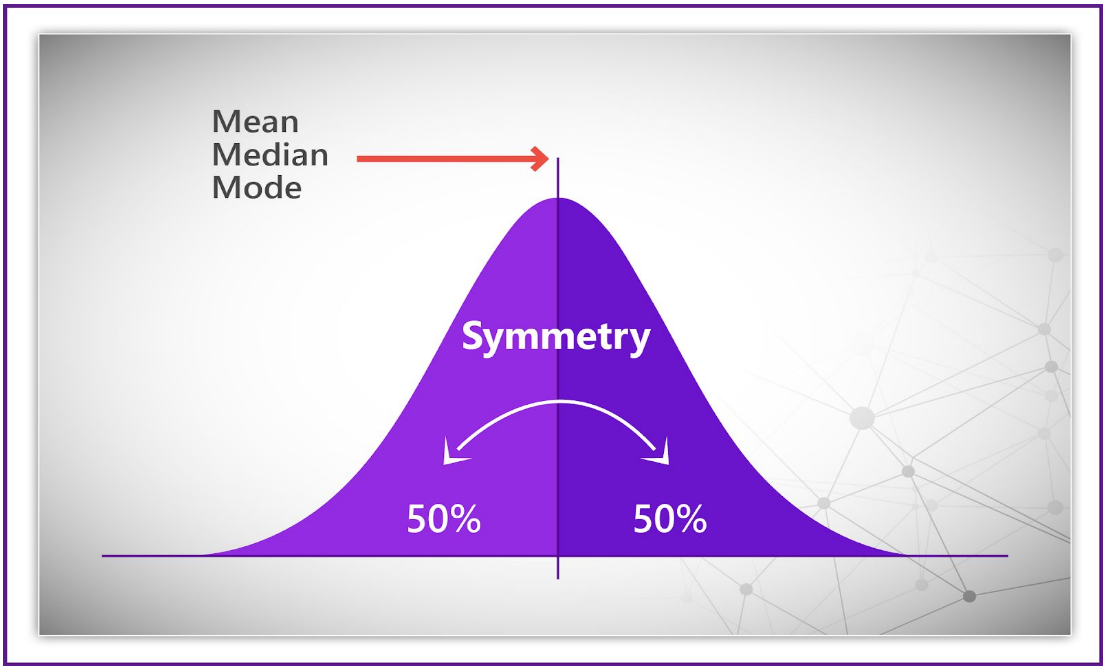
Normal Distribution

A continuous random variable X is said to follow a normal distribution if its **probability density function** can be written as

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

Above equation is not very intuitive, let us look at the properties of **normal** distribution to gain insights

Normal Distribution has: - mean = median = mode - symmetry about the center - 50% of values less than the mean - and 50% greater than the mean - The total area under the curve is 1.

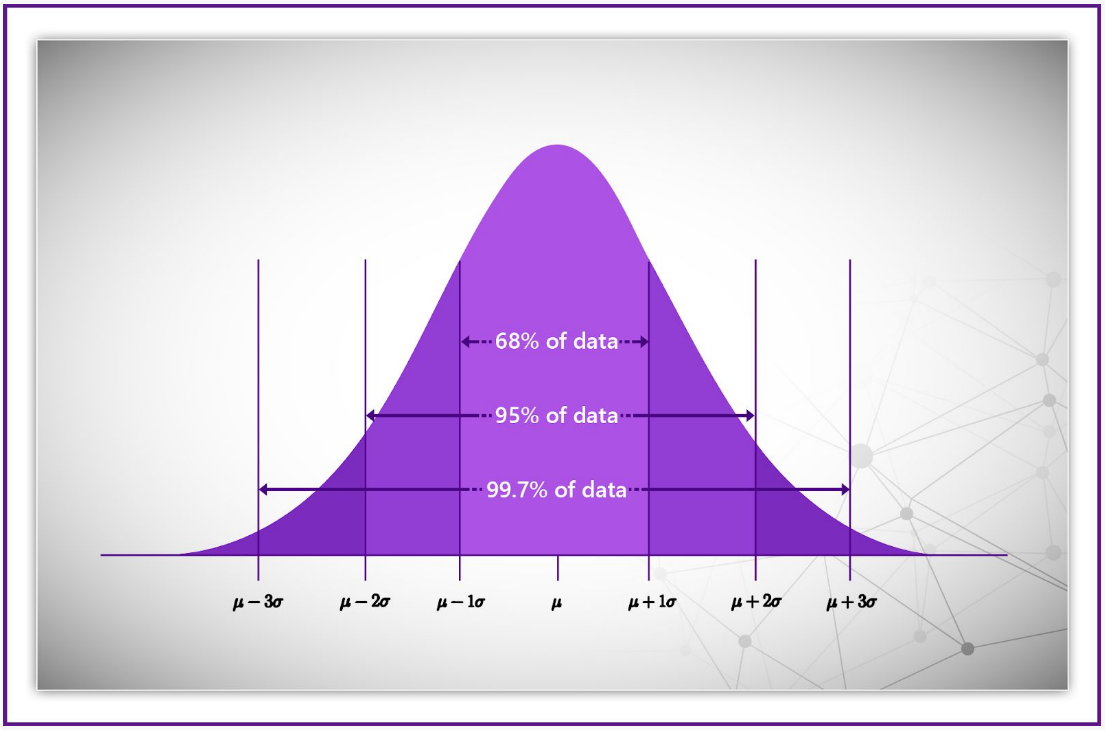


Important Normal Distribution property

We have learned about Standard Deviation and how it's the most commonly used measure of how spread out observations of our data are

When we calculate the standard deviation of a normally distributed data we find that (generally):

```
- 68% of values are within 1 standard deviation of the mean
- 95% of values are within 2 standard deviations of the mean
- 99.7% of values are within 3 standard deviations of the mean
```



Let us take an example. Let us take an example, suppose X is a random variable representing the waiting time in a queue at an airport security checkpoint. The average waiting time is 10 minutes with a standard deviation of 2 minutes. Now we can say that there is a 68% probability that the wait time in the queue is between 1 standard deviation of the mean i.e. between 8 and 12 minutes.

Similarly, we can extend the result saying there is a 95% probability that the wait time will be between 6 and 14 minutes and so on.

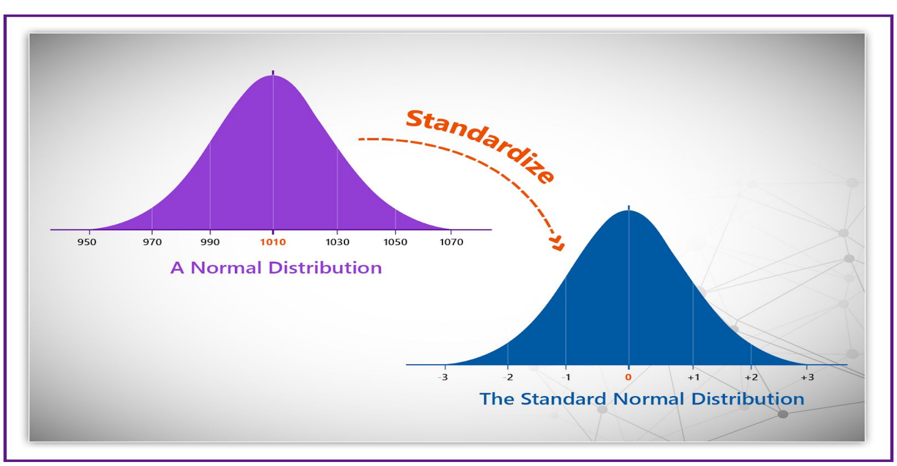
Standard Normal Distribution

A standard normal distribution is a normal distribution with mean = 0 and standard deviation of 1.

We can convert any normal distribution to a standard normal distribution using the following steps

- first, subtract the mean,
- then divide by the Standard Deviation

This process is called standardizing. The normal random variable of a standard normal distribution is called a standard score or a z-score.



Suppose we have a normally distributed random variable X, we calculate a new random variable **z** as

$$z = \frac{X - \mu}{\sigma}$$

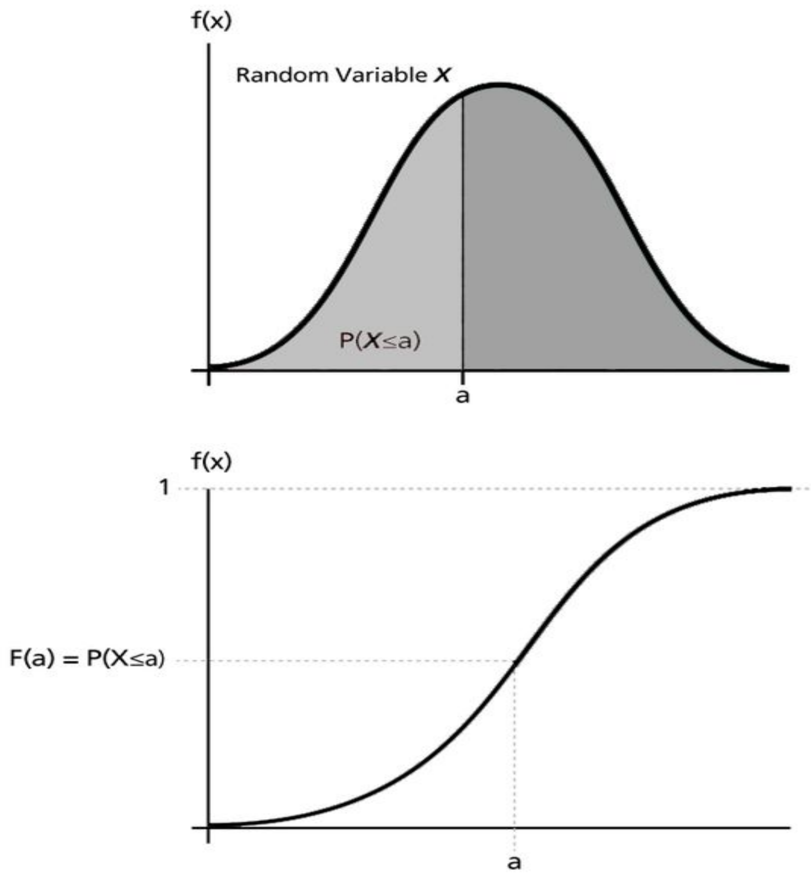
where

μ- mean of random variable X

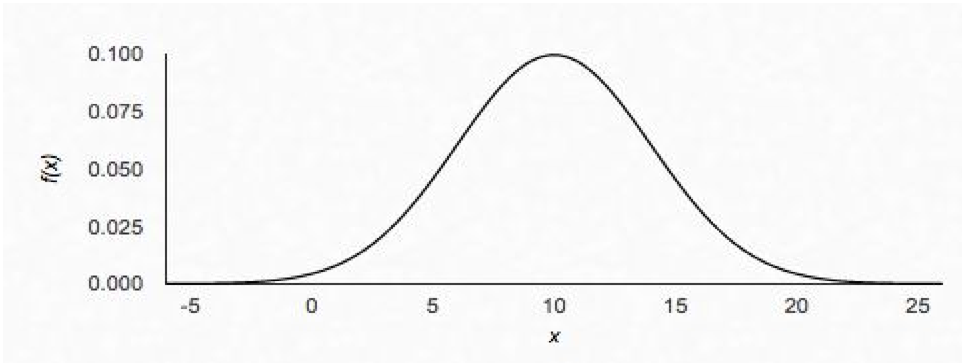
σ- std. dev. of random variable X

Calculating Probability using Standard Normal Distribution

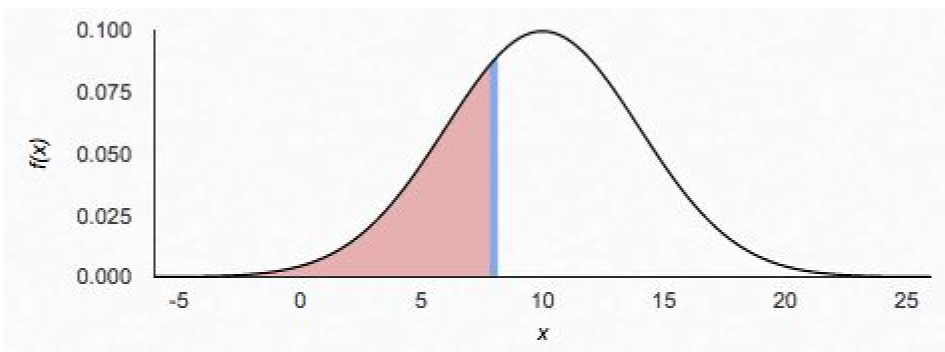
- While deriving a simple probability distribution earlier, we saw the concept of **probability mass function** where on the chart every point represents the probability of that event
- For a continuous variable, we can similarly calculate the probability from its **probability distribution function (PDF)** using the area under the curve
- The **cumulative distribution function (CDF)** for a continuous random variable X is defined as $F(x) = \int_{-\infty}^{x} f(t)dt$



Now let us continue with the airport security waiting time example. Earlier we saw that the waiting time at the airport



Let's say you want to know the probability that the wait time will be less than 8 minutes. We can easily calculate this probability by calculating the area under the normal curve below 8 minutes as represented by the following figure



The first step to calculate this probability is to convert the random variable X in a standard normal variable `z`, this can be achieved by using the formula seen earlier. Let us put things in notations to have a better sense of what is going on.

$$P(X \leqslant 8) = P\left(z \leqslant \frac{X-\mu}{\sigma}\right) = P\left(z \leqslant \frac{8-10}{2}\right) = P(z \leqslant -1)$$

Now the second step is to calculate the probability from the Standard Normal Distribution Table. Search for a value of 1.0 (-1.0) in the first column of the table. The values are the same for positive and negative numbers as the distribution is symmetric. Now for the above example, we need value for 1.00, check the value corresponding to column `0` besides 1.0. You can see that the value is 0.1587. Now using this value we calculate the probability as follows

$$P(X \leqslant 8)=P(z \leqslant -1)=1-0.8413=0.1587$$

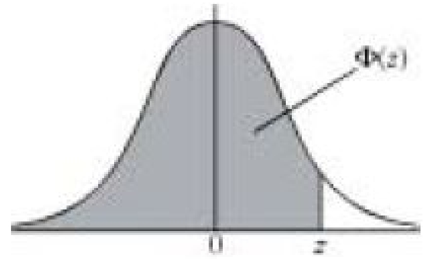
Hence the probability that the wait time in the queue will be less than 8 minutes in 15.87%

THE NORMAL DISTRIBUTION FUNCTION

If Z has a normal distribution with mean 0 and variance 1 then, for each value of z, the table gives the value of $\Phi(z)$, where

$$\Phi(z) = P(Z \leq z).$$

For negative values of z use $\Phi(-z) = 1 - \Phi(z)$.



z	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
															ADD				
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	4	8	12	16	20	24	28	32	36
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	4	8	12	16	20	24	28	32	36
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	4	8	12	15	19	23	27	31	35
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	4	7	11	15	19	22	26	30	34
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	4	7	11	14	18	22	25	29	32
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224	3	7	10	14	17	20	24	27	31
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	3	7	10	13	16	19	23	26	29
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	3	6	9	12	15	18	21	24	27
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	3	5	8	11	14	16	19	22	25
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389	3	5	8	10	13	15	18	20	23
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	2	5	7	9	12	14	16	19	21
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830	2	4	6	8	10	12	14	16	18
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015	2	4	6	7	9	11	13	15	17
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	2	3	5	6	8	10	11	13	14
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	1	3	4	6	7	8	10	11	13
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441	1	2	4	5	6	7	8	10	11
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545	1	2	3	4	5	6	7	8	9
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633	1	2	3	4	4	5	6	7	8
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706	1	1	2	3	4	4	5	6	6
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767	1	1	2	2	3	4	4	5	5
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817	0	1	1	2	2	3	3	4	4
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857	0	1	1	2	2	2	3	3	4
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890	0	1	1	1	2	2	2	3	3
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916	0	1	1	1	1	2	2	2	2
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936	0	0	1	1	1	1	1	2	2
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952	0	0	0	1	1	1	1	1	1
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964	0	0	0	0	1	1	1	1	1
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974	0	0	0	0	0	1	1	1	1
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981	0	0	0	0	0	0	0	1	1
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986	0	0	0	0	0	0	0	0	0

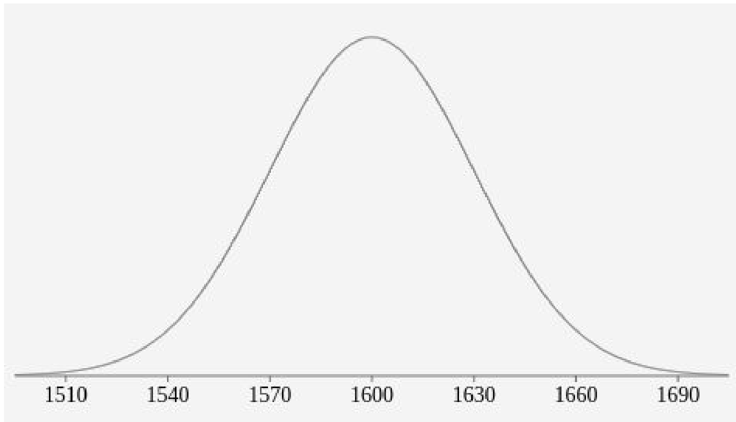
Example of Normal Distribution

Life of bulbs manufactured by a company follows a normal distribution with Mean = 1600 hrs and S.D. = 30 hrs

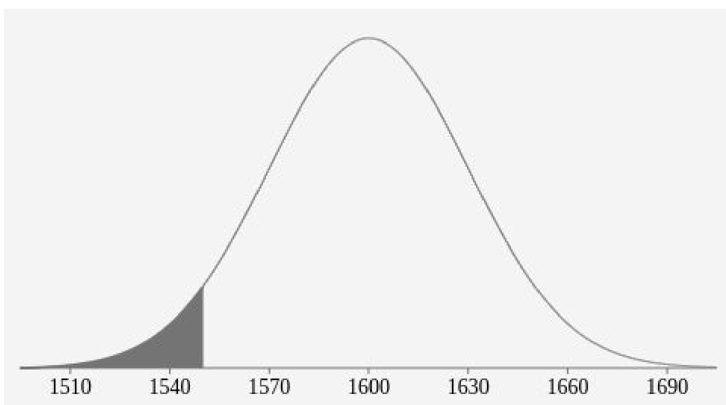
- What is the probability that the life of a bulb selected at random will be less than or equal to 1550 hrs?
- What is the proportion or percentage of bulbs having life less than or equal to 1660 hrs?
- What is the probability that the life of a bulb selected at random will be between 1550 and 1580 hrs?
- What is the probability that the life of a bulb selected at random will be between 1630 and 1680 hrs?
- What is the probability that the life of a bulb selected at random will be between 1550 and 1630 hrs?

Answer

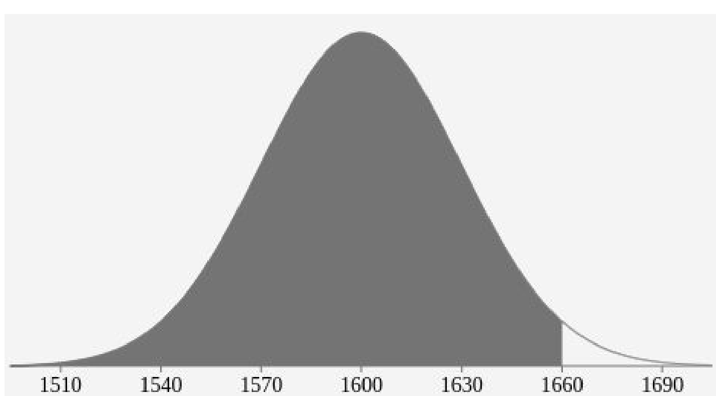
- Let X be the random variable representing the life of the bulb
- Original distribution



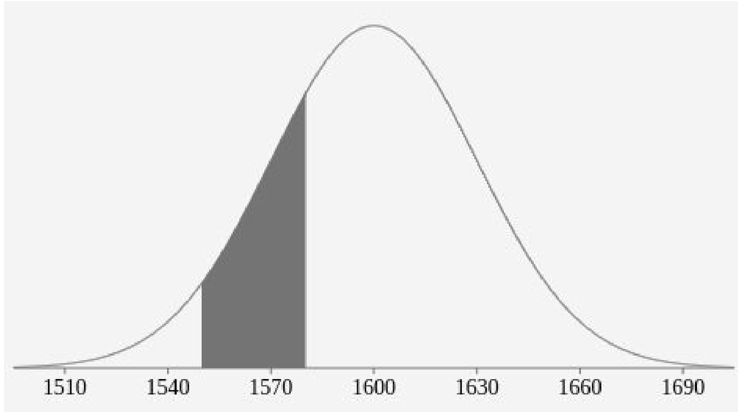
$$P(X \leqslant 1550) = P\left(z \leqslant \frac{1550-1600}{30}\right) = P(z \leqslant -1.67)$$



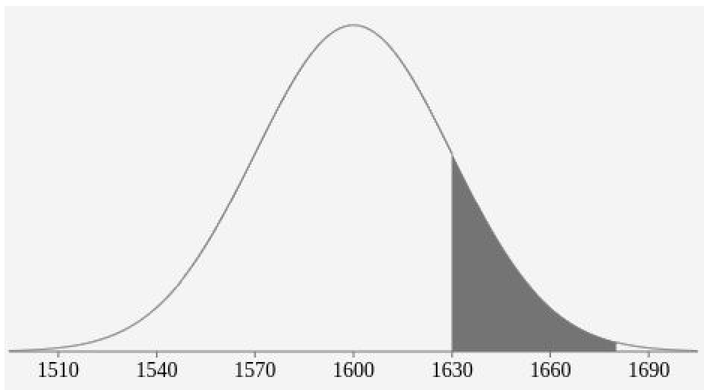
$$P(X \leqslant 1660) = P\left(z \leqslant \frac{1660-1600}{30}\right) = P(z \leqslant 2)$$



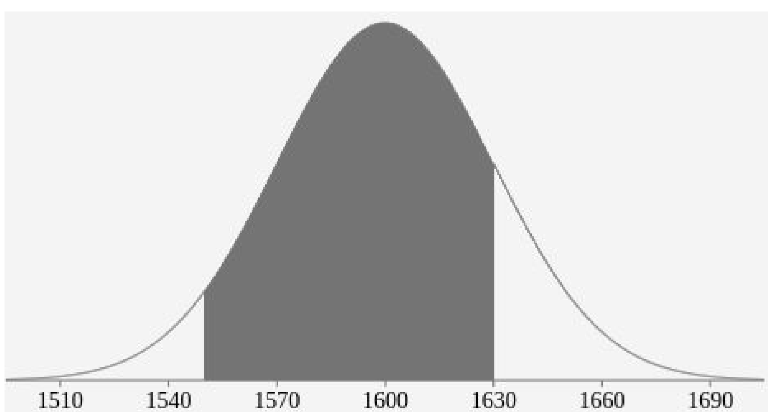
$$P(1550 \leqslant X \leqslant 1580) = P\left(\frac{1550-1600}{30} \leqslant z \leqslant \frac{1580-1600}{30}\right) = P(-1.67 \leqslant z \leqslant -1)$$



$$P(1630 \leqslant X \leqslant 1680) = P\left(\frac{1630-1600}{30} \leqslant z \leqslant \frac{1680-1600}{30}\right) = P(1 \leqslant z \leqslant 2.67)$$



$$P(1550 \leqslant X \leqslant 1630) = P\left(\frac{1550-1600}{30} \leqslant z \leqslant \frac{1630-1600}{30}\right) = P(-1.67 \leqslant z \leqslant 1)$$



Example:

We have the following details, X = 495,μ = 500 and σ = 110, Find the range in which 68% of the data lies.

Ans: 390 to 610

Explanation:

68% of the data lies within one standard deviation from the mean. So, we will add and subtract one standard deviation that is 110 to mean value of 500.

2.The average life of a certain type of motor is 10 years, with a standard deviation of 2 years. If the manufacturer is willing to replace only 3% of the motors because of failures, how long a guarantee should she offer?

Ans: 6.24

Explanation:

X = life of motor

x = guarantee period

We need to find the value (in years) that will give us the bottom 3% of the distribution. These are the motors that we are willing to replace under the guarantee.

$$P(X<x)=0.03$$

The area that we can find from the z-table is

$$0.5-0.03=0.47$$

The corresponding z-score is

$$z=-1.88$$

Since

$$Z=x-\mu\sigma$$

we can write:

$$(x-10)/2 = -1.88$$

Solving this gives x=6.24.

2.Given that X is a random variable which is normally distributed with μ = 30 and σ = 4. Determine the probability of X > 25.

Ans: 0.8944

Explanation:

First we will find the Z-score, Z = (X- μ)/(σ)

Now, we will look for the probability value for this Z-score from the Z-score table which is 0.8944.

3. It was found that the mean length of 100 parts produced by a lathe was 20.05 with a standard deviation of 0.02 mm. Find the probability that a part selected at random would have a length between 20.05 mm and 20.08mm.

Ans: 0.4332

Explanation:

First we will find the Z-score,

$$Z = \frac{X - \mu}{\sigma}$$

Z(X=20.05) = $\frac{25.05 - 20.05}{0.02}$ = 0

P(Z = 0) = 0.5 from the Z-score table

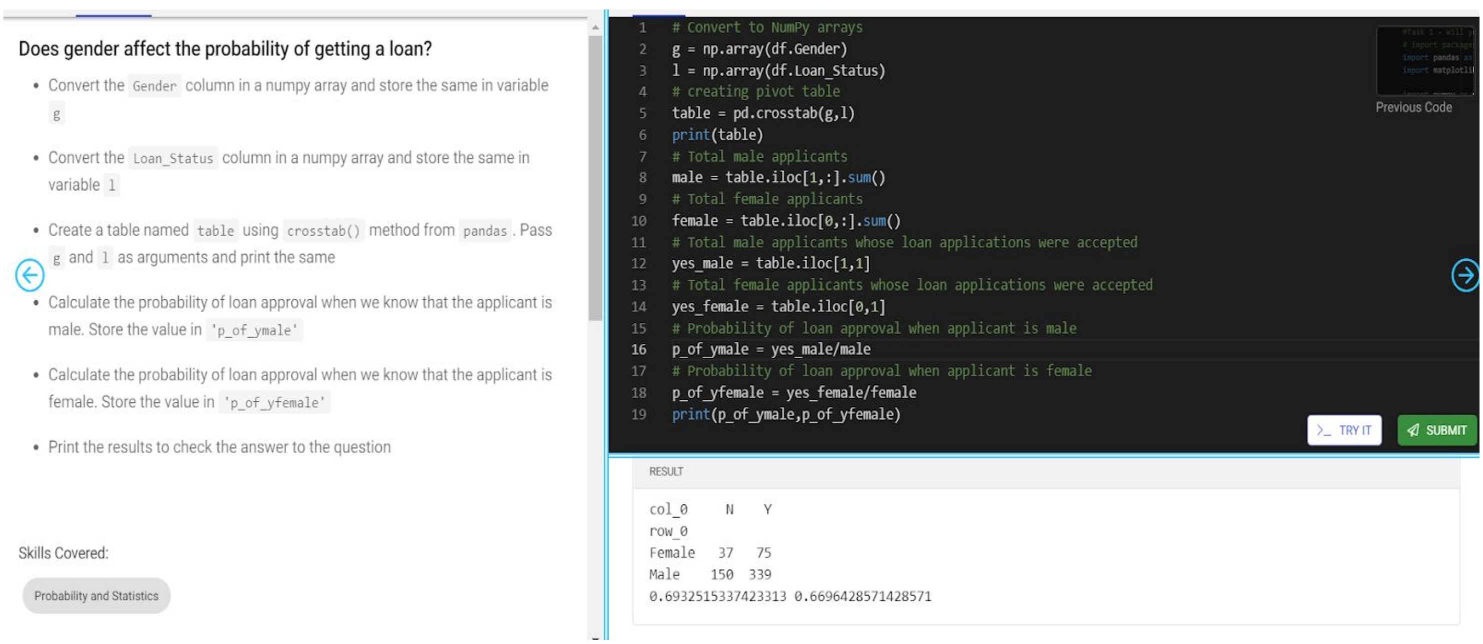
Z(X=20.08) = $\frac{25.08 - 20.05}{0.02}$ = 1.5

P(Z = 1.5) = 0.9332 from the Z-score table

So the probability that a part selected at random would have a length between 20.05 and 20.08 is 0.9332 - 0.5 = 0.4332

P(Z = 1.5) = 0.9332 from the Z-score table

So the probability that a part selected at random would have a length between 20.05 and 20.08 is 0.9332 - 0.5 = 0.4332.


Does gender affect the probability of getting a loan?

- Convert the `Gender` column in a numpy array and store the same in variable `g`
- Convert the `Loan_Status` column in a numpy array and store the same in variable `l`
- Create a table named `table` using `crosstab()` method from `pandas`. Pass `g` and `l` as arguments and print the same
- Calculate the probability of loan approval when we know that the applicant is male. Store the value in `'p_of_ymale'`
- Calculate the probability of loan approval when we know that the applicant is female. Store the value in `'p_of_yfemale'`
- Print the results to check the answer to the question

Skills Covered:

Probability and Statistics

```
# Convert to NumPy arrays
g = np.array(df.Gender)
l = np.array(df.Loan_Status)
# creating pivot table
table = pd.crosstab(g,l)
print(table)
# Total male applicants
male = table.iloc[1,:].sum()
# Total female applicants
female = table.iloc[0,:].sum()
# Total male applicants whose loan applications were accepted
yes_male = table.iloc[1,1]
# Total female applicants whose loan applications were accepted
yes_female = table.iloc[0,1]
# Probability of loan approval when applicant is male
p_of_ymale = yes_male/male
# Probability of loan approval when applicant is female
p_of_yfemale = yes_female/female
print(p_of_ymale,p_of_yfemale)
```

Previous Code

TRY IT SUBMIT

RESULT

```
col_0     N    Y
row_0
Female   37   75
Male    150  339
0.6932515337423313 0.6696428571428571
```

1. A roulette wheel has 26 slots, 12 are blue, 10 are green, and 4 are red. You play six games and always bet on red. What is the probability that you win all the 6 games?

Ans: 0.000013

Explanation:

The probability that it would be Red in any spin is 4/26. Now, you are playing for game 6 times and all the games are independent of each other. Thus, the probability that you win all the games is $(4/26)^6$ = 0.000013

2.Events A and B are independent if and only if

Ans: P(A and B) = P(A) . P(B)

3.A group of 80 students is randomly split into 4 classes of equal size. All partitions are equally likely. Smith and David are two students belonging to that group. What is the probability that Smith and David will end up in the same class?

Ans: 19/79

Explanation:

All possible partitions are obtained with equal probability by a random assignment if these numbers, it doesn't matter with which students we start, so we are free to start by assigning a random number to Joe and then we assign a random number to Smith. After Smith has been assigned a random number there are 79 random numbers available for David and 19 of these will land her in the same group as Joe. Therefore the probability is 19/79.

4.Suppose you were interviewed for a data science role. 60% of the people who sat for the first interview received the call for the second interview. 80% of the people who got a call for the second interview felt good about their first interview. 70% of people who did not receive a second call, also felt good about their first interview. If you felt good after your first interview, what is the probability that you will receive a second interview call?

Ans: 63%

Explanation:

The probability of receiving a call for the second interview is 0.6 and for not receiving a call is 0.4. The probability that you feel good about the first interview given that you got a call for the second interview is 0.6x0.8 = 0.48. The probability that you feel good about the first interview given that you did not get a call for the second interview is 0.4x0.7 = 0.28. So the probability that you will receive a second call given that you felt good after your first interview = 0.48}/{0.48+0.28}= 0.63 i.e. 63%

5.In a data science class , there are 12 boys and 13 girls. Three students are selected at random. The probability that 2 girl and 1 boys are selected, is:

Ans: 0.4069

Explanation:

Since we are selecting 1 of 12 boys, 2 of 13 girls, and 3 of 25 total people, the probability is:= {12C1 * 13C2}/{25C3}=12*78/2300=0.4069

6.From a very large population two people are selected, and their blood type is checked.

- A) person 1 has blood type A
- B) person 2 has blood type A

Are the two events independent or dependent?

Ans:Independent

Explanation:

Person 1 having blood type A had no effect on the probability of person 2 having blood type A.

7.Let A and B be events on the same sample space, with P (A) = 0.8 and P (B) = 0.4. Can these two events be disjoint?

Ans: No

Explanation:

Assuming you've figured out that your events are disjoint (using the definition above), you can figure out the probabilities by adding them together:

P(A or B) = P(A) + P(B)

The addition of both the probabilities here is going beyond 1. So, the two events cannot be disjoint events.

8.You are planning to play cricket today, but the morning is cloudy.50% of all rainy days start off cloudy and cloudy mornings are common about 40% of days.This month is a dry one where there is only a 10% chance to be rainy. What is the chance of rain today?

Ans:0.125

Explanation:

We will use Rain to mean rain during the day, and Cloud to mean cloudy morning.

The chance of Rain given Cloud is written P(Rain|Cloud)

So let's put that in the formula: $P(Rain|Cloud) = \{P(Rain)\ P(Cloud|Rain)\}/\{P(Cloud)\}$

$$P(Rain|Cloud) = \{0.1 * 0.5\}/\{0.4\} = 0.125$$

9.Which of the following options cannot be the probability of any event?

Ans: -0.656

Explanation:Probability of an event lies between 0 and 1 only.

10.Kevin has 2 kids and one of them is a boy. What is the probability that the other child is also a boy?

Ans: 0.333