Probability, Distributions, and Summary Stats

DATA, MSML, BIOI 602 Principles of Data Science

[Slides from Fardina Fathmiul Alam]

Topics we will cover:

1. Probability Theory:
    a. Basic concepts (events, sample space, probability axioms)
    b. Conditional probability
    c. Bayes' theorem
    d. Probability distributions (discrete and continuous)
    e. Common distributions (e.g., normal, binomial, Poisson)
2. Descriptive Statistics:
    a. Measures of central tendency (mean, median, mode)
    b. Measures of dispersion (variance, standard deviation, range)
    c. Percentiles and quartiles
    d. Skewness and kurtosis

Inferential Statistics: Hypothesis testing (Later Topic)

Part 01 Probability Theory

What is Probability?

- People talk loosely about probability all the time:

  “What are the chances the Orioles will win this weekend?”

  “What’s the chance of rain tomorrow?”

- For scientific purposes, we need to be more specific in terms of defining and using probabilities

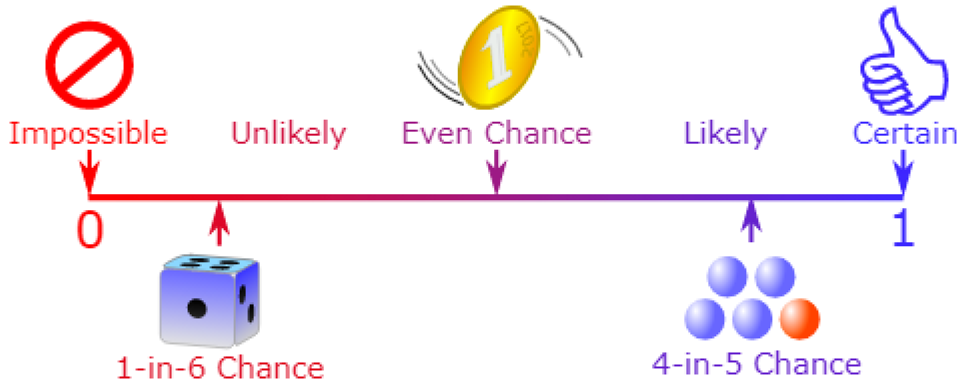
What is Probability?

Probability is simply how likely something is to happen.

Remember, the analysis of events governed by probability is called statistics.

What is Probability? (most commonly used concepts in statistics)

Classic Definition: Probability is a measure between **zero and one** for the **likelihood** that something or some event might occur.



A Classic Example of Probability

- What is the chances of rolling a one (Number 1) on the dice?

Answer is a 1/6

Ques: How do we know that ?

Basic Probability Formula

To find the probability of an event happening we use the formula:

$$P(A) = \frac{\text{Number of times A occurs}}{\text{Total number of possible outcomes}}$$

Here, **P(A)** = The probability of event A

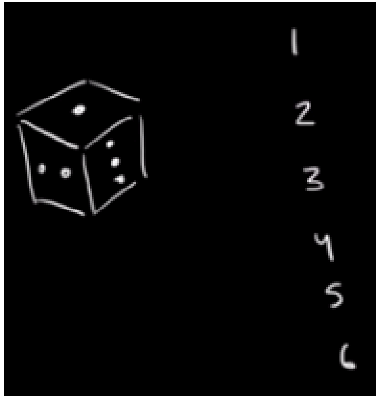
Probabilities range from **0 (impossibility-the event will not occur)) to 1 (certain- the event will definitely occur)** [degrees of likelihood]the and can also be written as a percentage.

Basic Probability Formula

To find the probability of an event happening we use the formula:

$$P(A) = \frac{\text{Number of times A occurs}}{\text{Total number of possible outcomes}}$$

Example: The probability of getting number 1 is: P(1) = 1/6



There are six different outcomes. (**sample space**)

Try Yourself: What is probability of getting an even number when a die is rolled?

Some More Examples

- Stock Market → The chance of stock market's rising above some point, or falling below some point is ‘x’%
- Weather Forecast → The chance for rain is 45% tonight.
    - The likelihood of rainfall in terms of probability is 0.45 that the event, rainfall, ,might occur

So Essentially probability is a measure between equate to 1.

Probability allows us to measure and express uncertainty.

In data science, we often deal with incomplete or noisy data, and probability provides a framework to assess the likelihood of different outcomes. By assigning probabilities to events, we can make informed decisions and evaluate the associated risks.

Random Variable and Probability Distribution: Making Sense of Uncertainty

A random variable is a variable in statistics and probability theory that represents the outcomes of a random event.

- Assigns a number to each possible outcome of an uncertain event.
- Functions as a map of outcomes in a probability space.

E.g: Imagine rolling two dice. The outcome (the combination of numbers we get) is uncertain. So, we use a random variable, X to represent the sum of the numbers on the two dice. X could take on values from 2 to 12, depending on the sum of the two dice.

Random Variable and Probability Distribution: Making Sense of Uncertainty

Probability Distributions: Describe the likelihood of different outcomes occurring.

- Tells us how likely each outcome of the random variable is.
- Act like a map showing the chances of each possible value of our random variable.

Example: Probability Distribution

Probability Distributions: Describe the likelihood of different outcomes occurring.

E.g: Imagine rolling two dice. There are 36 possible outcomes (since each die has 6 sides, giving us 6×6=36 combinations). Each outcome has an equal chance of happening if the dice are fair. So, the **probability distribution for our random variable X** would look like this: →

P(X=2)= 1/36
P(X=3)= 2/36 (rolling a 1 and a 2 or rolling a 2 and a 1)
P(X=4)= 3/36
P(X=5)= 4/36
P(X=6)= 5/36
P(X=7)= 6/36
P(X=8)= 5/36
P(X=9)= 4/36
P(X=10)= 3/36
P(X=11)= 2/36
P(X=12)= 1/36

Random Variable Types:

- Discrete Random Variable: Represents outcomes that can only take on a countable number of distinct values.
  - E.g. Rolling a six-sided die. The possible outcomes (1, 2, 3, 4, 5, 6) are finite and countable.

- Continuous Random Variable: Represents outcomes that can take on any value within a certain range.
  - E.g. Measuring the height of people. Height can take on any value within a range (e.g., from 0 to infinity). It can be 160.5 cm, 178.23 cm, or any value in between.

More Probability Formulas

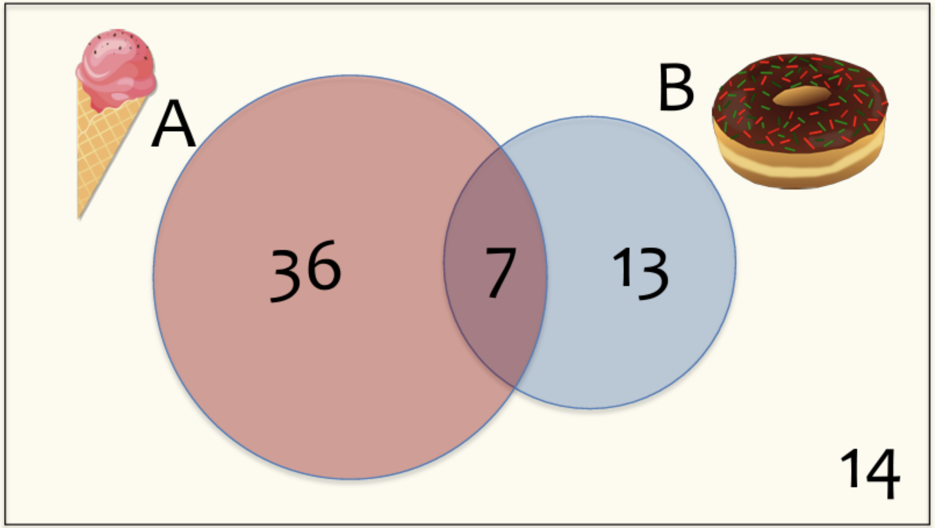
Sometimes you want to

figure out the chance of something happening when we already know else has happened/occured?

“Conditional Probability”

(Understanding Likelihood Given Information)

Conditional Probability (Idea)



What's the probability that someone likes ice cream **given** they like donuts?

Conditional Probability cont.

Example:

What is the probability of passing the class given you didn't sleep the night before?

Conditional Probability cont.

The probability of event A happening given that event B has occurred.

We write it: $P(A \mid B)$ ➜ read as the probability of "A given B."

For the example on the previous slide,
let

- A =Passing the class,
- B =Not sleeping the night before the final

We can expressed as "P(Passing the class | Not sleeping the night before the final)"

Definition: Conditional Probability

Definition. The **conditional probability** of event A given an event B happened (assuming $P(B) \neq 0$) is

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

Formula: Conditional Probability

Conditional Probability Formula

$$\underbrace{P(A \mid B)}_{\text{Probability of } A \text{ given } B} = \frac{\overbrace{P(A \cap B)}^{\text{Probability of } A \text{ and } B}}{\underbrace{P(B)}_{\text{Probability of } B}}$$

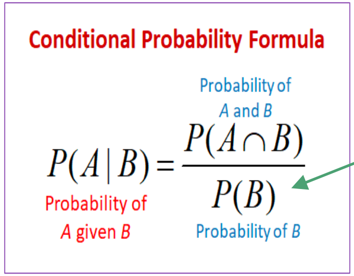
We take the chance of both things happening together, then divide it by the chance of the thing we know

This formula tells us the probability of event A happening, given that event B has happened.

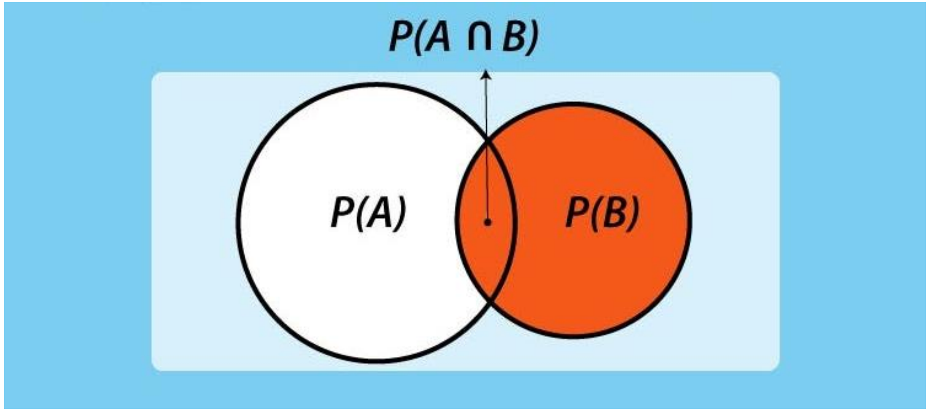
Formula: Conditional Probability

Imagine two overlapping circles representing two events, **A** and **B**:

- **Circle A**: All outcomes where event **A** happens (**P(A)** - the size of Circle A).
- **Circle B**: All outcomes where event **B** happens (**P(B)** - the size of Circle B).
- **Overlap between Circles A and B:** Probability of both events happening together (P(A and B)).



By dividing by P(B), we focus our attention on the subset of cases where event B is true.



Venn diagram for Conditional Probability, P(A|B)

More Example:

Imagine you have a bag of colored marbles - 5 red and 5 blue. If you know that 3 out of the 5 red marbles are also **shiny**, you might wonder:

"What's the chance of picking a shiny marble from the bag if I know it's red?"

A represents the event of picking a shiny marble

B represents the event the event of picking a red marble.

More Example:

Imagine you have a bag of colored marbles - 5 red and 5 blue. If you know that 3 out of the 5 red marbles are also **shiny**, you might wonder:

"What's the chance of picking a shiny marble from the bag if I know it's red?"

Conditional Probability Formula

$$\underbrace{P(A \mid B)}_{\text{Probability of } A \text{ given } B} = \frac{\overbrace{P(A \cap B)}^{\text{Probability of } A \text{ and } B}}{\underbrace{P(B)}_{\text{Probability of } B}}$$

$P(A \cap B)$ = What is the probability of picking a shiny red marble ?

$P(A \cap B) = 3/10$

More Example:

Imagine you have a bag of colored marbles - 5 red and 5 blue. If you know that 3 out of the 5 red marbles are also **shiny**, you might wonder:

"What's the chance of picking a shiny marble from the bag if I know it's red?"

Conditional Probability Formula

$$\underbrace{P(A \mid B)}_{\text{Probability of } A \text{ given } B} = \frac{\overbrace{P(A \cap B)}^{\text{Probability of } A \text{ and } B}}{\underbrace{P(B)}_{\text{Probability of } B}}$$

P (B) = What is the probability of picking a red marble ?

P (B) = 5/10

More Example:

Imagine you have a bag of colored marbles - 5 red and 5 blue. If you know that 3 out of the 5 red marbles are also **shiny**, you might wonder:

"What's the chance of picking a shiny marble from the bag if I know it's red?"

Conditional Probability Formula

Probability of A and B

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

Probability of A given B

Probability of B

$P(A \cap B) = 3/10$

$P(B) = 5/10$

$P(A|B)$?

More Example:

Imagine you have a bag of colored marbles - 5 red and 5 blue. If you know that 3 out of the 5 red marbles are also **shiny**, you might wonder:

"What's the chance of picking a shiny marble from the bag if I know it's red?"

Conditional Probability Formula

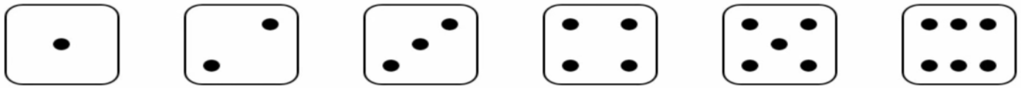
$$\underbrace{P(A \mid B)}_{\text{Probability of } A \text{ given } B} = \frac{\overbrace{P(A \cap B)}^{\text{Probability of } A \text{ and } B}}{\underbrace{P(B)}_{\text{Probability of } B}}$$

P (A ∩ B) = 3/10

P (B) = 5/10

P (A|B) = (3/10) / (5/10) = ⅗ = 0.6

Exercise: Conditional Probability



$$\frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1$$

P(B | A) = What is the Probability of (rolling a dice and it's value is less than 4 | knowing that the value is an odd number)

Exercise: Conditional Probability

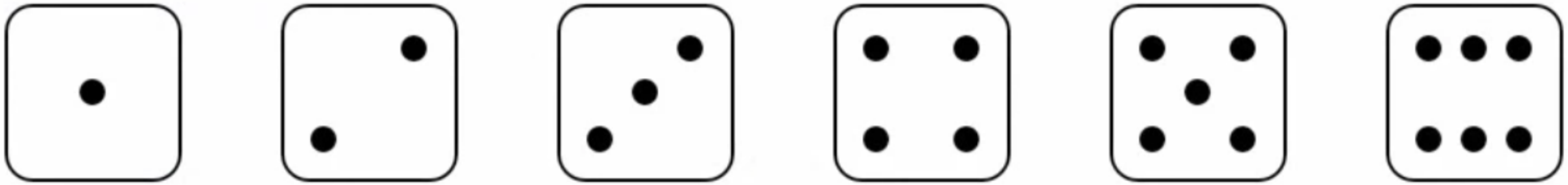
What is the Probability of

rolling a dice and it's value is 1

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

knowing that the value is an odd number

rolling a dice and it's value is 1



Next: Independence

Independence

Two random processes are independent if knowing the outcome of one provides no useful information about the outcome of the other.

- Outcomes of two tosses of a coin are independent.

Independence and Conditional Probability

Conditional independence refers to the situation where the occurrence of one event **doesn't affect** the probability of another event happening

Conditional independence

In mathematical notation, if $P(A|B) = P(A)$ then the events A and B are said to be independent.

- Conceptually: Giving B doesn't tell us anything about A
  - The occurrence (or non-occurrence) of one event doesn't affect the probability of the other event.

Conditional independence

In mathematical notation, if P(A|B) = P(A) then the events A and B are said to be independent.

- Conceptually: Giving B doesn't tell us anything about A
  - The occurrence (or non-occurrence) of one event doesn't affect the probability of the other event.
- Equivalently, one can also check the independency of the events A and B by check whether P(B|A) = P(B).

Examples: Conditional independence

- What if P(Passing the class | I drank Coke before final exam)= P(Passing the class)?
- What is P(Drawing a red card | Not sleeping well) =P(Drawing a red card)?
- Think of P(My relationship working out | I'm a Libra)= P(My relationship working out)?

- Conditional independence can help us to focus on what matters most in our data.
- These concepts are like powerful tools that help us uncover valuable insights and build accurate models, making data science more effective and useful.

Next: Bayes Theorem

One of the most important formulas in statistics (for our purposes) as well as the most important rule in data science!

Bayes Theorem

Bayes' Rule can answer a variety of probability questions, which help us (and machines) understand the complex world we live in.

- Bayes' Theorem is a fundamental theorem in probability theory that describes how to update or revise the probability of a hypothesis based on new evidence or information.
- It is named after Thomas Bayes, an 18th-century British mathematician.

It's particularly useful in situations where we need to make decisions or predictions based on uncertain information.

Bayes Theorem

Theorem. (Bayes Rule) For events A and B, where $P(A), P(B) > 0$,

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

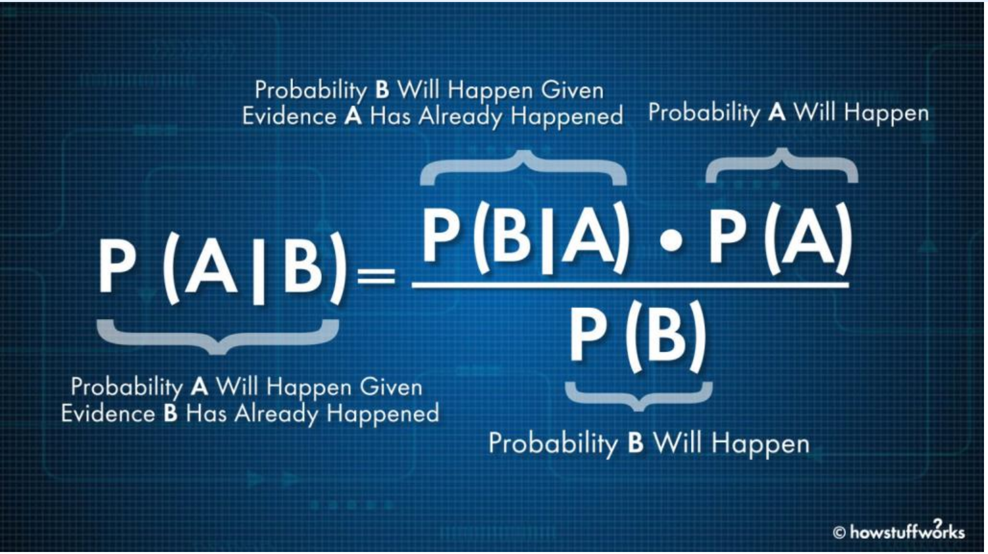
$P(A)$ is called the **prior** (our belief without knowing anything)

$P(A|B)$ is called the **posterior** (our belief after learning B)

Bayes Theorem Formula

Bayes' Rule tells you how to calculate a conditional probability with information (prior beliefs and new evidence) you already have.

Bayes Rules Formula: **Which tells us:** how often A happens given that B happens, written P(A|B),



When we know:

- **P(B|A):** how often B happens given that A happens
- **P(A):** how likely A is on its own
- **P(B):** how likely B is on its own

Bayes Theorem Proof

It provides a way to calculate conditional probabilities when we have **prior beliefs** and **new evidence**.

By definition Cond. Prob: $$P(A|B) = \frac{P(A \cap B)}{P(B)}$$ (Equation 1)

Swapping A, B gives: $$P(B|A) = \frac{P(A \cap B)}{P(A)}$$ Here: $P(A \cap B) = P(B \cap A)$,

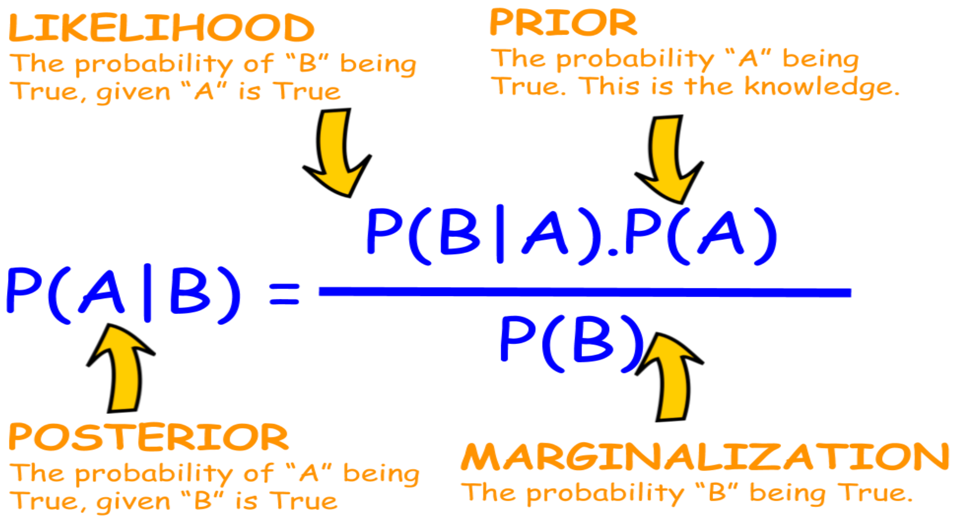
$$\Rightarrow \quad P(A \cap B) = P(B|A) * P(A)$$

Now from Equation 1, substitute value of $P(A \cap B)$:

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

Bayes Theorem Formula

Bayes Theorem Formula with Terminologies



- **Conditional probability** deals with the initial probability of an event given another event.
- While **Bayes' Rule** is a tool for updating this probability based on new evidence.

Bayes' Rule is particularly valuable in fields like statistics, machine learning, and Bayesian inference, where we continuously update our beliefs as new data or information becomes available.

Example: Fire and Smoke

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

Home Bonfire

We can then discover the **probability of dangerous Fire when there is Smoke**:

$$P(\text{Fire}|\text{Smoke}) = \frac{P(\text{Fire})\ P(\text{Smoke}|\text{Fire})}{P(\text{Smoke})}$$

$$= \frac{1\% \times 90\%}{10\%}$$

$$= 9\%$$

So it is still worth checking out any smoke to be sure.

Example: Picnic Day

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

What is the chance of rain during the day?

Scenario: You want to find the chance of rain during the day given that the morning is cloudy. You **have** the following probabilities:

1. Probability of Rain (**P(Rain)**) = 10% (because only 3 out of 30 days are rainy).

1. Probability of Cloud, given that Rain happens (**P(Cloud|Rain)**) = 50% (because 50% of rainy days start off cloudy).
2. Probability of Cloudy morning (**P(Cloud)**) = 40% (because 40% of all days start off cloudy).

Now, use Bayes' Theorem:

$$P(Rain|Cloud) = \frac{P(Rain)\;P(Cloud|Rain)}{P(Cloud)}$$

Example: Picnic Day

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

What is the chance of rain during the day?

Scenario: You want to find the chance of rain during the day given that the morning is cloudy. You **have** the following probabilities:

1. Probability of Rain (**P(Rain)**) = 10% (because only 3 out of 30 days are rainy).

1. Probability of Cloudy morning given Rain (**P(Cloud|Rain)**) = 50% (because 50% of rainy days start off cloudy).
2. Probability of Cloudy morning (**P(Cloud)**) = 40% (because 40% of all days start off cloudy).

Now, use Bayes' Theorem:

$$P(Rain|Cloud) = \frac{0.1 \times 0.5}{0.4} = .125$$

Or a 12.5% chance of rain. Not too bad, let's have a picnic!

Try: Exercise:

For example, imagine a rare disease called kittenpox, that turns people into adorable kittens. The symptoms are growing cat ears and meowing. Someone walks into a doctor’s office showing these symptoms and meowing. The doctor wants to know the chance they have kittenpox.

What do each of these terms represent?

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

Try: Bayes Rule Practice

my prior on a student studying is 0.9.

On average, 20% of my students get A’s on the exam.

Of the students who study, 80% gets A’s on the exam.

What is the probability that a student get an A’s if they study?

Try: Bayes Rule and Conditional Prob. Practice

Let's work on a simple NLP problem with Bayes Theorem. By using NLP, I can detect spam e-mails in my inbox. Assume that the word 'offer' occurs in 80% of the spam messages in my account. Also, let's assume 'offer' occurs in 10% of my desired e-mails. If 30% of the received e-mails are considered as a scam, and I will receive a new message which contains 'offer', what is the probability that it is spam?

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

P('offer'|spam) = 0.8, P('offer'|desired) = 0.1,
P(scam) = 0.3, P(desired) = 1- 0.3 = 0.7
P(scam|'offer') = ? <u>P(offer|scam)P(scam)</u>

P(offer) = P(offer|scam)P(scam) + P(offer|desired)P(desired)
P(offer) = P(offer ∩ scam) + P(offer ∩ desired)
P(offer) = P(offer ∩ (scam U desired)) = P(offer in all emails)

P(scam|'offer') =0.8*0.3/(0.8*0.3+0.1*0.7)

Next: Law of Total Probability

Example: Equal Probability of Selecting Each Bag

Example 01: Suppose I have **two bags of marbles**. The first bag contains **6** white marbles and **4** black marbles. The second bag contains **3** white marbles and **7** black marbles. Now suppose I put the two bags in a box. If I close my eyes, grab a bag from the box, and then grab a marble from the bag, what is the probability that it is black?

Total Probability of Drawing a Black Marble:

$$\frac{11 \text{ BLACK}}{20 \text{ total}} = \frac{11}{20} = 0.55$$

$$\frac{1}{2} \cdot \frac{4}{10} + \frac{1}{2} \frac{7}{10} = \frac{4}{20} + \frac{7}{20} = \frac{11}{20}$$

How About Unequal Probability of Selecting Each Bag?

Example: Same scenario as before, but now suppose that the first bag is much larger than the second bag, so that when I reach into the box I'm twice as likely to grab the first bag as the second. What is the probability of grabbing a black marble?

Probability of selecting Bag 1

$P(B_1) = 2/3$

Probability of selecting Bag 2

$P(B_2) = 1/3$

$$\frac{1}{2} \cdot \frac{4}{10} + \frac{1}{2} \frac{7}{10} = \frac{4}{20} + \frac{7}{20} = \frac{11}{20}$$

How About Unequal Probability of Selecting Each Bag?

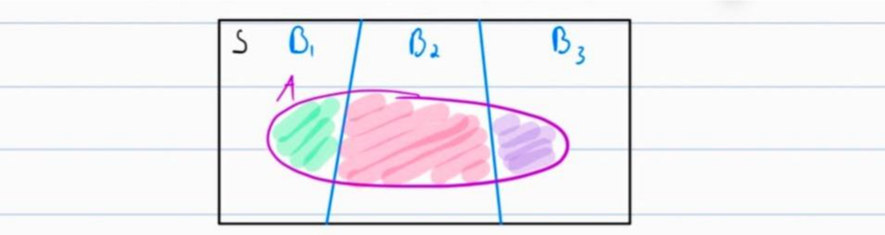
Example: Same scenario as before, but now suppose that the first bag is much larger than the second bag, so that when I reach into the box I'm twice as likely to grab the first bag as the second. What is the probability of grabbing a black marble?

$P(B_1) = 2/3 \qquad P(B_2) = 1/3$

$$\frac{2}{3} \cdot \frac{4}{10} + \frac{1}{3} \cdot \frac{7}{10} = \frac{8}{30} + \frac{7}{30} = \frac{15}{30} = \frac{1}{2}$$

We just applied “Law of Total Probability”

- The Law of Total Probability is a fundamental concept in probability theory that allows us to **compute the probability of an event** by considering all possible ways in which it can occur.
  - If we partition the sample space S into **a collection of mutually exclusive events B1, B2, B3** (i.e., events that cannot occur simultaneously) that together cover the entire sample space, then the **probability of any event A** can be expressed as the sum of the probabilities of that event given each partition, weighted by the probability of each partition occurring.



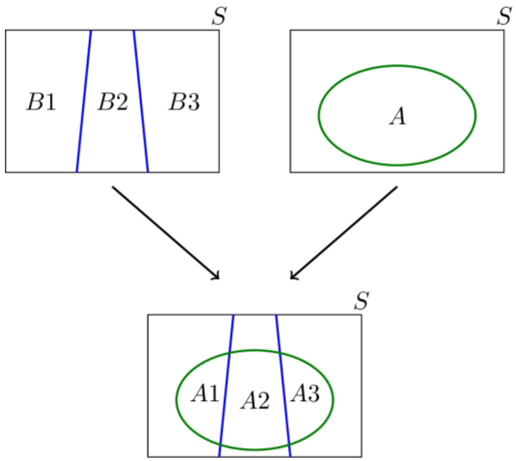
$$P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3)$$

We just applied "Law of Total Probability"

The Law of Total Probability helps us find the probability of an event by considering all possible ways it can happen.

The Law of Total Probability states that if you have a **sample space S** partitioned into n mutually exclusive and exhaustive events B1,B2,…,Bn (events that cannot occur simultaneously and cover the entire sample space), then for any event A, the probability of A can be expressed as:

$$P(A) = \sum_{i=1}^{n} P(B_i) \cdot P(A|B_i)$$



The probability of an event A (P(A): the sum of the probabilities of A given each event, weighted by the probability of each event.

Law of Total Probability: back to the example

Sample Space S: The scenario of choosing a bag and drawing a marble.

1. **Mutually Exclusive Events B1 B1 and B2**:
   - B1: Selecting Bag 1.
   - B2: Selecting Bag 2.

These events are mutually exclusive because you can only choose one bag at a time.

1. **Event A**: Drawing a black marble.

According to the law of total probability, the probability of drawing a black marble (event A) is:

P(A)=P(B1)·P(A | B1)+P(B2)·P(A |B2)P(A)

Try Yourself (Solution is given later in add. Reading slide)

Scenario: Imagine you are organizing a charity event, and there are three possible venues (A, B, and C) where you can hold the event. The probability of each venue being available on a given day is as follows:

Venue A: 40% chance of being available.

Venue B: 30% chance of being available.

Venue C: 30% chance of being available.

You also know that if **Venue A is available, there's a 70% chance of raising a large amount of money for the charity**, while if **Venue B or C is available, there's a 50% chance of raising a large amount of money**.

What is the overall probability L of raising a large amount of money at the charity event?

Expected Value

Expected Value

Expected value uses probability to tell us what outcomes to expect in the long run.

A tool for reasoning under uncertainty.

We use it when we are choosing between two choices, each of which can have variable outcomes.

Expected Value

Expected Value is a measure used to quantify the average outcome or payoff of a random variable.

$$E[X] = \sum x_i p(x_i)$$

x_i = The values that X takes

$p(x_i)$ = The probability that X takes the value x_i

Expected Value: Example

Someone offers for you to go on a game show. On this gameshow, there is:

- A **5%** chance you will be given a million dollars
- A **95%** chance they will hit you with sticks, causing **$10,000** worth of medical bills (the hits are not particularly bad, your insurance just doesn't cover check-ups)

Your feelings about being hit with sticks aside, should you go on the game show?

Expected Value: Example

- There's a 5% chance of winning $1,000,000, which is 0.05× $1,000,000=$50,000
- There's a 95% chance of incurring $10,000 in medical bills, which is 0.95×−$10,000=−$9,500.
- Minus sign: negative financial outcome

Expected value of the game show:

EV=(Probability of Winning×Prize for Winning)+(Probability of Medical Bills×Cost of Medical Bills)

=(1,000,000 * .05) + (-10,000 * .95) = 40500

Net positive!

Conclusion: on average, you can expect to gain $40,500 by participating in the game show.

Expected Value: Example

Someone offers for you to go on a game show. On this gameshow, there is:

- A **5%** chance you will be given a million dollars
- A **95%** chance they will hit you with sticks, causing **$10,000** worth of medical bills (the hits are not particularly bad, your insurance just doesn't cover check-ups)

Your feelings about being hit with sticks aside, should you go on the game show?

Expected value of the game show:

(1,000,000 * .05) + (-10,000 * .95) = 40500

Net positive!

We don't always just care about money though! You can work in other things to your function. If we decide not getting hit with sticks is worth 5,000 dollars to us, we can add that into the equation.

(1,000,000 * .05) + (-(10,000+5000) * .95) = 35750

Doctor's bills

Not getting hit with sticks term

Expected Value: Example

Someone offers for you to go on a game show. On this gameshow, there is:

- A **5%** chance you will be given a million dollars
- A **95%** chance they will hit you with sticks, causing **$10,000** worth of medical bills (the hits are not particularly bad, your insurance just doesn't cover check-ups)

Your feelings about being hit with sticks aside, should you go on the game show?

Expected value of the game show:

(1,000,000 * .05) + (-10,000 * .95) = 40500

Net positive!

We don't always just care about money though! You can work in other things to your function. If we decide not getting hit with sticks is worth 5,000 dollars to us, we can add that into the equation.

(1,000,000 * .05) + (-(10,000+5000) * .95) = 35750

Still totally worth it!

Example #2

You are playing a game where you spin a wheel. The wheel has the following sets of rewards:

20% chance you win nothing

30% chance you win a ten dollar gift card

40% chance you win a twenty dollar gift card

10% chance you win a ten dollar gift card

EV=(.2 * 0) + (.3 * 10) + (.4 * 20) + (.1 * 10) = 0+3+8+1 =12

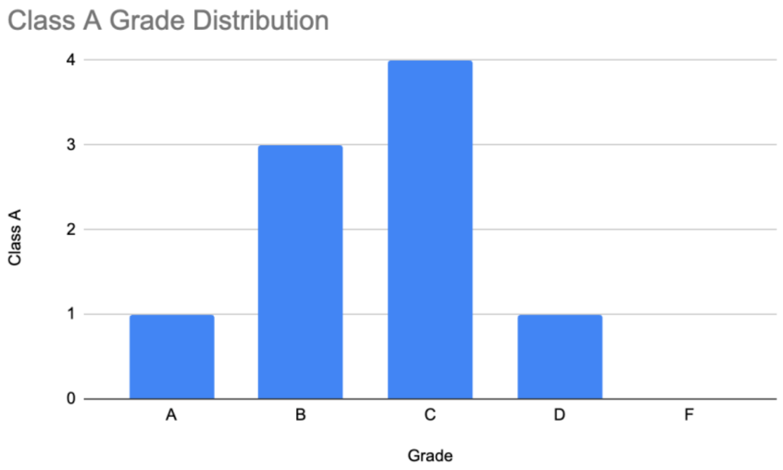
On average, you can expect to win $12 per spin over a large number of spins.

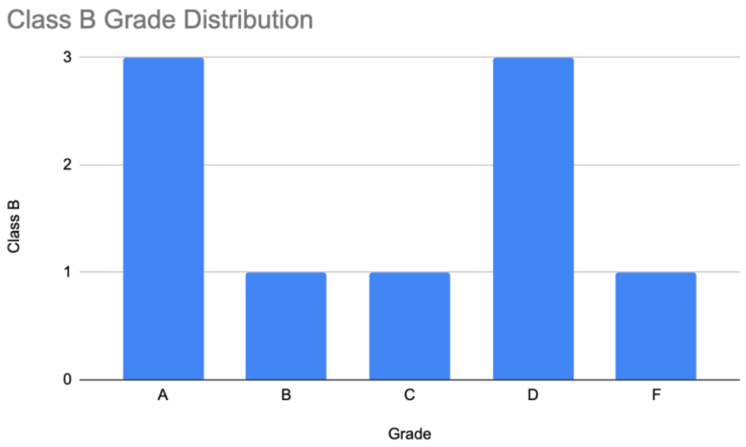
Distributions

What is a distribution?

Class A	Class B
C	A
B	A
C	A
A	B
C	C
B	D
D	D
C	D
B	F

Grade Distributions





These distributions show how the grades are distributed among the students in each class, giving an idea of the proportion of students who received each grade.

- It helps understand the performance of the students in terms of their grades.

RECAP: Probability Distribution

Probability Distributions: Describe the likelihood of different outcomes occurring.

- Tells us how likely each outcome of the random variable is.
- It's like a map that shows us the chances of getting each possible value of our random variable.

E.g: Imagine rolling two dice. There are 36 possible outcomes (since each die has 6 sides, giving us 6×6=36 combinations). Each outcome has an equal chance of happening if the dice are fair. So, the **probability distribution for our random variable X** would look like this: →

P(X=2)= 1/36
P(X=3)= 2/36 (rolling a 1 and a 2 or rolling a 2 and a 1)
P(X=4)= 3/36
P(X=5)= 4/36
P(X=6)= 5/36
P(X=7)= 6/36
P(X=8)= 5/36
P(X=9)= 4/36
P(X=10)= 3/36
P(X=11)= 2/36
P(X=12)= 1/36

Why does this matter?

- Understanding how your data is distributed can tell you a lot about the process generating the data
- The nature of your distribution affects which statistical tools you use

Distribution Types

Discrete Distributions: describe random variables that can only take on a countable number of distinct values. These values are typically integers.

Examples: Bernoulli distribution, Binomial distribution, Poisson distribution.

Continuous Distributions: describe the probabilities associated with continuous random variables. Unlike discrete distributions, which assign probabilities to specific outcomes, continuous distributions assign probabilities to intervals or ranges of outcomes. This is because continuous random variables can take on any value within a certain range, typically real numbers.

Examples: Normal (Gaussian) distribution, Exponential distribution etc.

Next: We will discuss different distribution types

1. Uniform Distribution
2. Normal (Gaussian) Distribution
3. Not Normal:
   a. Poisson Distribution
   b. Zero-Inflated Poisson Distribution
   c. Binomial Distribution

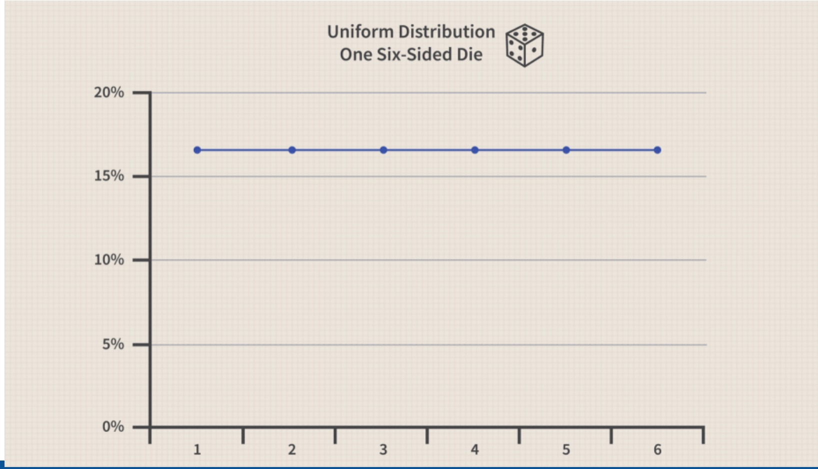
Many more …..

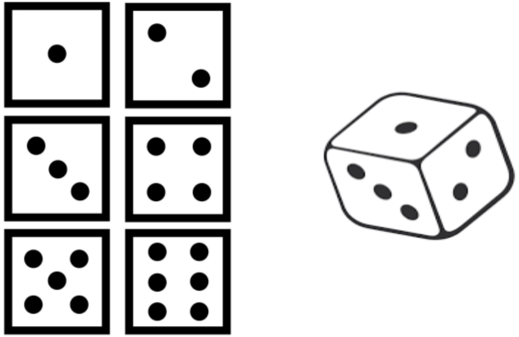
The Uniform Distribution

The Uniform Distribution

Def: describes a set of outcomes where each outcome has an equal probability of occurring within a specified interval.

- All outcomes are equally likely.
- It's like rolling a fair six-sided die, where each number has the same chance of appearing.



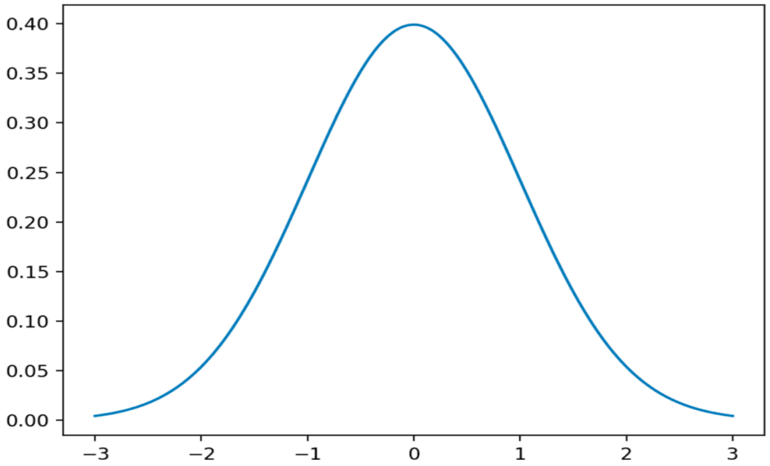


The Normal Distribution (Gaussian)

The Normal Distribution (Gaussian)

- Extremely common
- People tend to assume data is **normally distributed** unless there is a reason to think otherwise.
- Many natural phenomena, such as heights, weights, and measurement errors, follow approximately normal distributions.

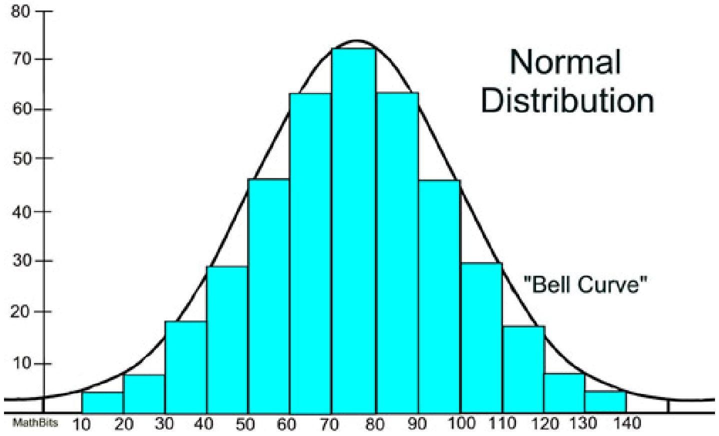
Data is symmetrically distributed with no skew



The Normal Distribution (Gaussian) - A continuous PD

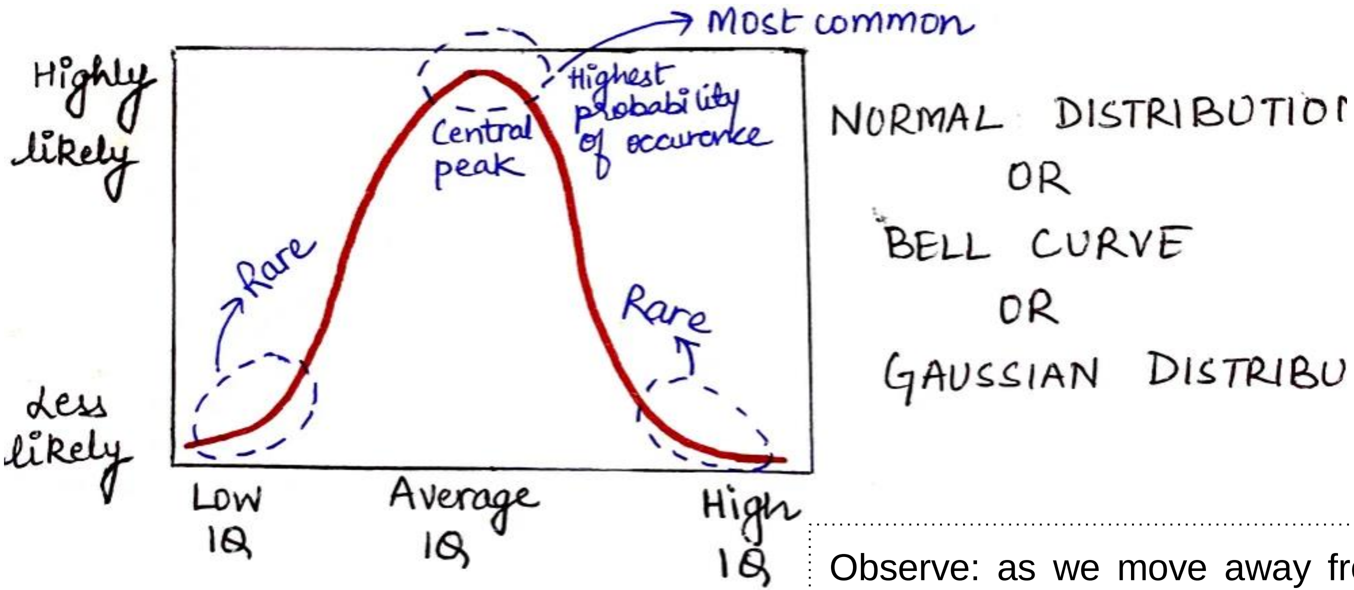
- **Bell-Shaped Curve**: When plotted, the data follows a bell shape, with most values clustering around a central region (with the highest point (the mean) at the center) and tapering off as they go further away from the center.
- **Parameters N(μ, σ²) : The mean** (average) and **standard deviation** (a measure of how spread out the data is) fully describe the distribution.

It is symmetric around its mean and is defined by two parameters: the mean (μ) and the standard deviation (σ).



The Normal Distribution (Gaussian)

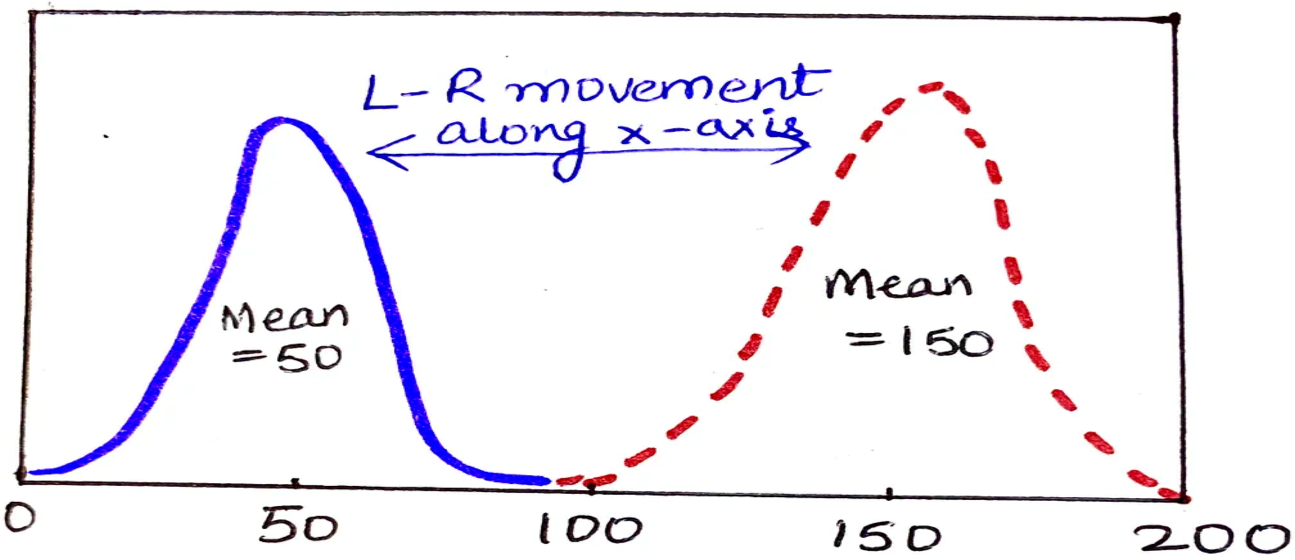
Example: the **IQ levels of a population** follow a normal distribution. It is understandable that people falling in very low IQ levels or very high IQ levels are rare occurrences and the majority of the population lies in the range of Average IQ scores.



Observe: as we move away from the central peak in both the directions, we see that the probability of occurrence of values at the tails of the curve becomes less and less likely.

Parameters of the Normal Distribution

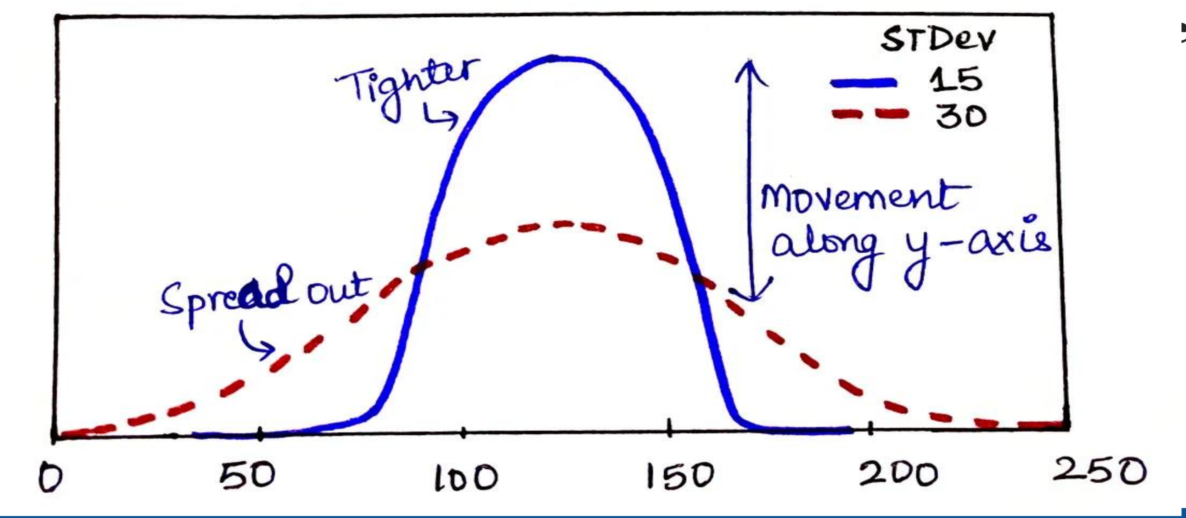
1. **MEAN:** Central tendency of a normal distribution; determines the location of the Peak of the curve. Change in mean results in the horizontal shifting of the curve along the x-axis.



Parameters of the Normal Distribution

2.**Standard deviation (SD):** measure of the variability of a normal distribution which determines the width of the curve.

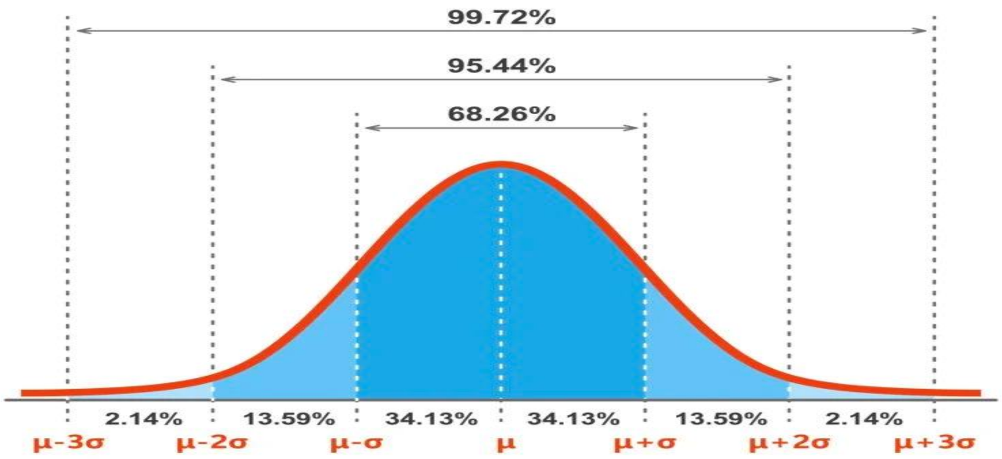
- **Tighter the curve (lesser the width) ->taller the height**
- **Mor** **eight**



Few Key Characteristics of Normal Distribution

The Empirical Rule: The 68, 95, 99 rule: tells you where most of your values lie in a normal distribution

- Approximately 68% of the data falls within one standard deviation of the mean
- 95% falls within two standard deviations, and
- 99.7% falls within three standard deviations.



Because the normal distribution is a probability distribution, the area under the distribution curve is equal to one.

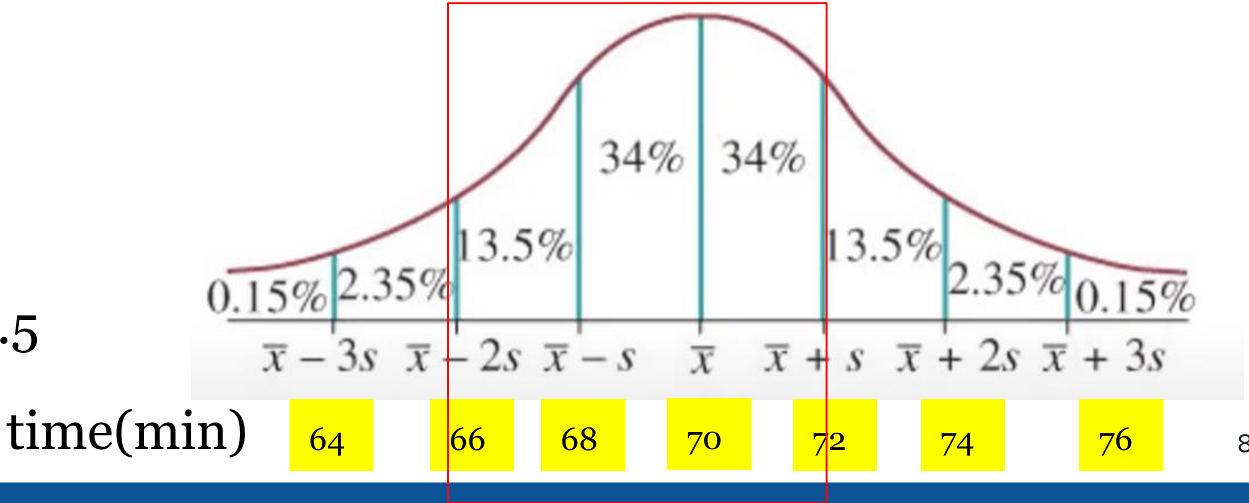
Problem Solving: Finding percentages

Example 1: The time taken to travel between two regional cities is approximately normally distributed with a mean of 70 minutes and a standard deviations of two minutes

Q: What is the percentage of travel times that are between 66 minutes and 72 minutes?

- Mean 70
- Sx 2
- % ?

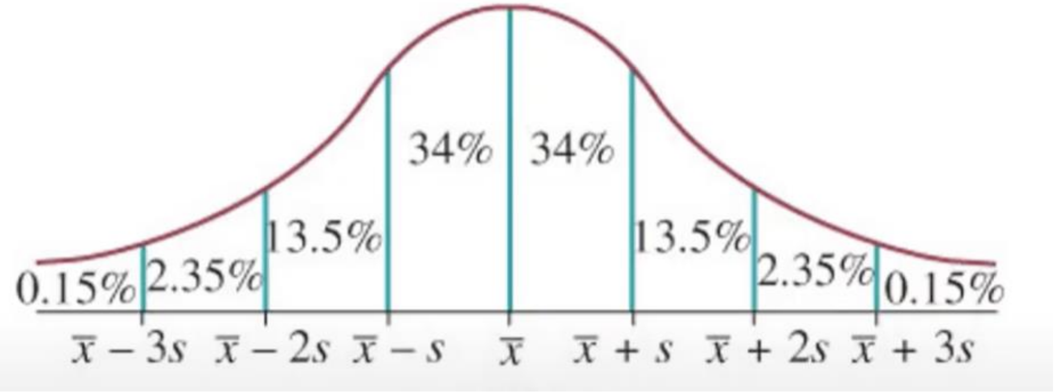
%= 13.5+34+34= 81.5



Try Yourself

The volume of a cup of soup serve via machine is normally distributed with a mean of 240 mL and a standard deviation of 5 mL. A fast store used this machine to serve 160 cups of soup

Ques: What number of these cups of soup are expected to contain less than 230 ml?

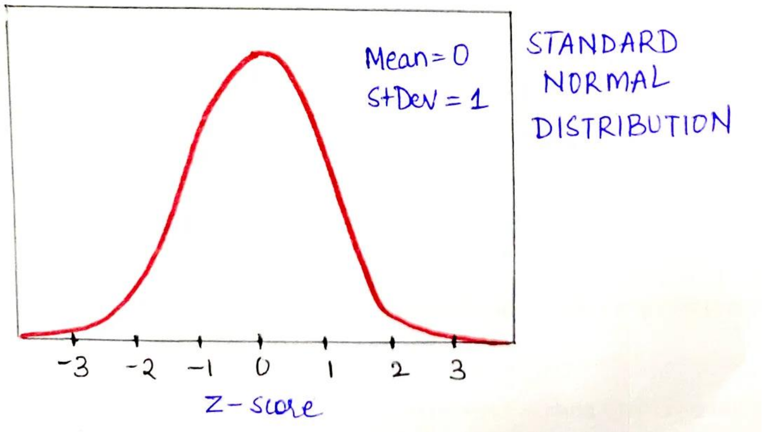


Standard Normal Distribution: Special Case of Normal Distribution

Standard Normal Distribution: Special Case of Normal Distribution

The standard normal distribution, also called the z-distribution, is a special normal distribution where the mean μ is 0 and the standard deviation, σ is 1.

- The standard normal distribution allows for comparison and analysis of data from different distributions by standardizing them into z-scores, making it a fundamental tool in statistical analysis.



Standard Normal Distribution: Special Case of Normal Distribution

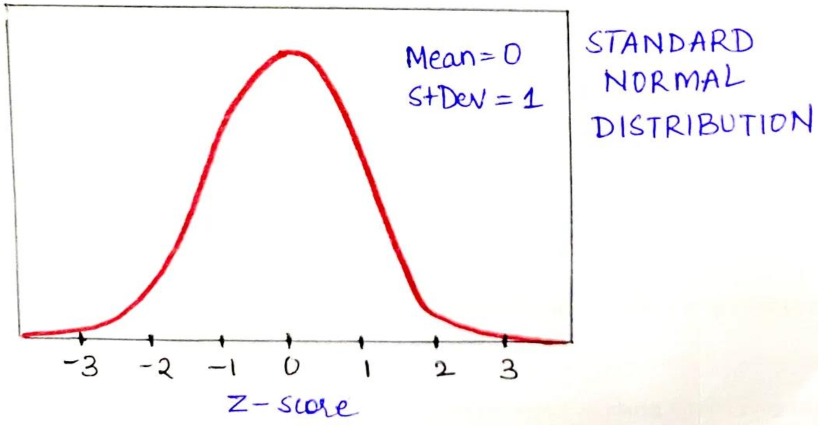
Any normal distribution can be standardized by converting its values into z scores.

A z-score is a numerical measure of **how many standard deviations a data point is away from the mean of the distribution.**

- <u>Calculate Z-Score:</u> subtracting the mean and dividing by the standard deviation, Z.

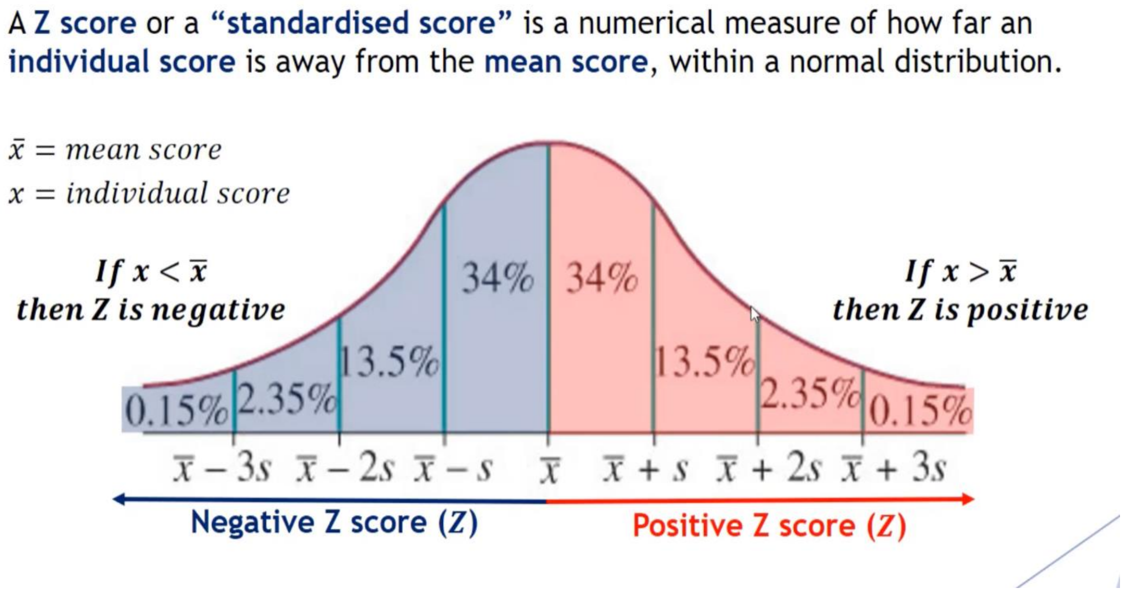
$$z = \frac{x - \mu}{\sigma}$$

- x = individual value
- μ = mean
- σ = standard deviation



Standardization: Interpreting Z-Scores

Any normal distribution can be standardized by converting its values into z scores.



Interpreting Z-Scores:

- **A (+) z score:** your x value is greater than the mean.
- **A (-) z score means** that your x value is less than the mean.
- **A z score of zero** means that your x value is equal to the mean.

Standardization: Interpreting Z-Scores

Calculating Z-Scores: given $\mu = 100$, $\sigma = 15$

If a student scored 115 on the test:

- Z-Score for the score of 115 is,

  Z=115 − 100 / 15 = 1

Interpreting Z-Scores: A z-score of 1 indicates the score is 1 standard deviation above the mean. This indicates that the student performed better than the average student, relative to the spread of scores in the distribution.

Benefit: When you standardize a normal distribution, the mean becomes 0 and the standard deviation becomes 1. This allows you to easily calculate the probability of certain values occurring in your distribution, or to compare data sets with different means and standard deviations.

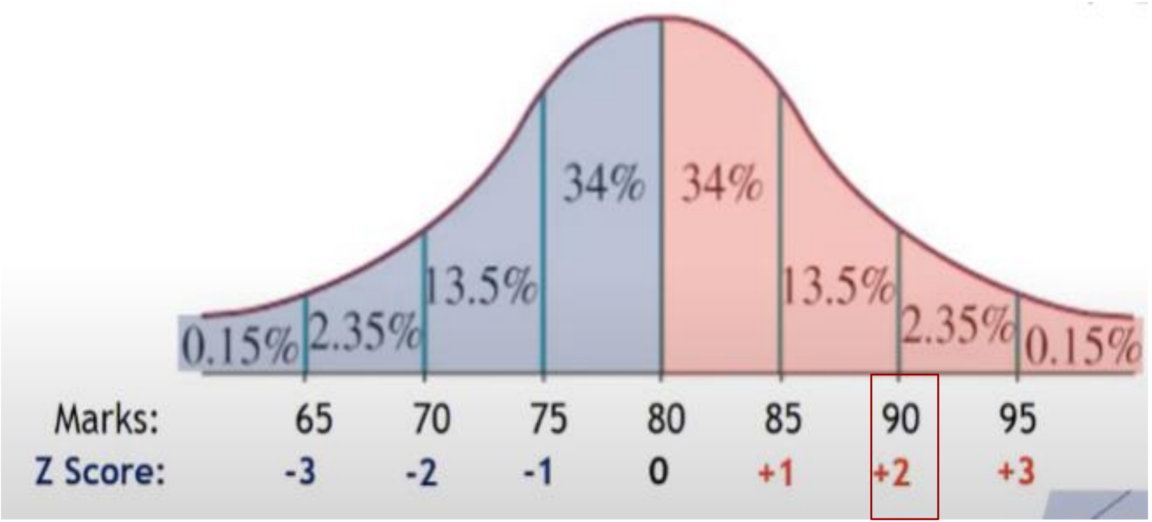
Z-Score Example:

Example: A math class sits a test with a mean score of 80 marks and a standard deviation of 5 marks. The distribution is approximately normally distributed. James achieves score of 90 marks. what is his Z-Score?

Individual score, x = 90
Mean = 80
Std dev = 5
Z = ?

Use equation:

Z=(90− 80) / 5 = 2



Find the probability associated with a specific z-score

The z-score itself is not a probability; it is a measure of how many standard deviations a data point is away from the mean of a distribution.

- Once we've standardized the values of interest, we can use **statistical tables (like the z-table) or statistical software** to find the probability associated with those values.

Other Distributions: Not everything is normal

Poisson Distribution

A Poisson distribution is a discrete probability distribution:

- it gives the probability of a discrete (i.e., countable) outcome.
- For Poisson distributions, the discrete outcome is the number of times an event occurs, represented by k.

$$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

We can say, X follows a Poisson distribution with parameter λ

Where,

- P(X=k) = the probability of observing exactly k (0,1,2,3,.....) events.
- λ= average number of events per interval (the rate parameter)
- E = Euler's constant ≈ 2.71828

Poisson Distribution

More Precisely, It gives us the probability of a given specific number of events happening in a fixed interval of time.

We can use to predict or explain the number of events (k) occurring within a given interval of time or space.

- “Events” → disease cases, customer purchases, meteor strikes etc.
- “The interval” → any specific amount of time or space, such as 10 days or 5 square inches.

We to use Poisson distribution: if

- **Individual events happen at random and independently:** the probability of one event doesn’t affect the probability of another event.
- **You know the mean number of events occurring within a given interval of time or space: this number is called λ (lambda)**, and it is assumed to be constant.

Note: When events follow a Poisson distribution, **λ is the only thing you need to know to calculate** the probability of an event occurring a certain number of times.

Poisson Distribution

More Precisely, It gives us the probability of a given specific number of events happening in a fixed interval of time.

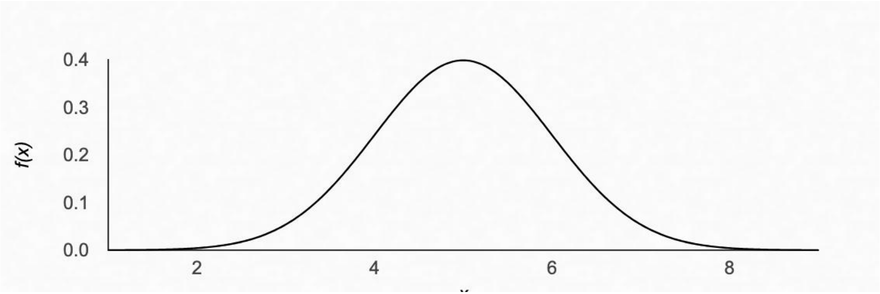
Examples:

Scenario: how many phone calls you receive during your shift (duration of shift 8 hours)?

- Average Rate (Constant Rate): Average number of phone calls received during a shift, $\boldsymbol{\lambda} = 10$.
- **Assumptions:** Phone calls arrive randomly and independently of each other, and the average rate of phone calls per shift remains relatively constant.

Then, we can use the Poisson distribution to model the number of phone calls received during a shift.

Say that I know on average, I get five people (λ) an hour during office hours.



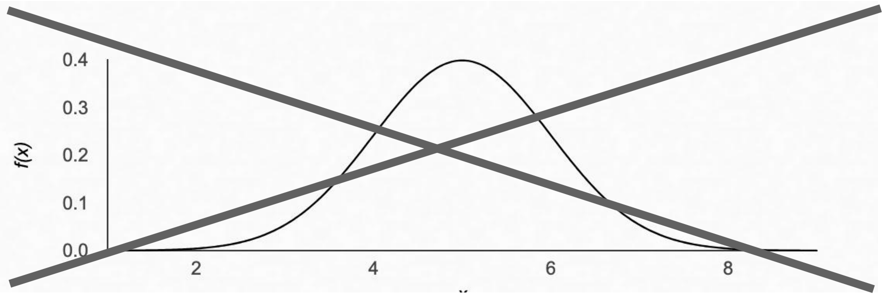
Does this distribution look correct?

Poisson Distribution

Say that I know on average, I get five people (λ) an hour during office hours.

Does this distribution look correct?

- People walking into my office is not continuous
- It's generated by a fundamentally different process

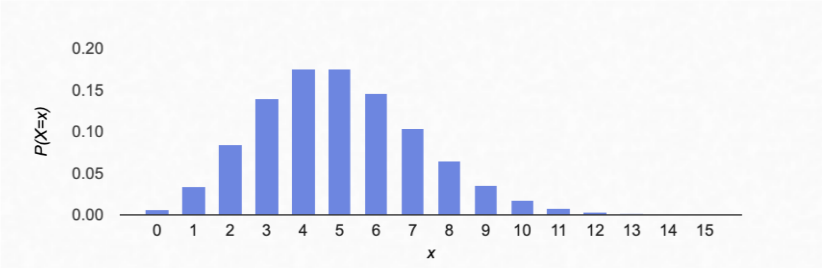


Visualization of Poisson Distribution

Why does it look like this?

- People walking into my office is **not continuous or negative (actually discrete (whole number) as we can't have a fraction of a person arrive)**

- Arrivals are independent events, meaning one arrival does not affect the likelihood of another.

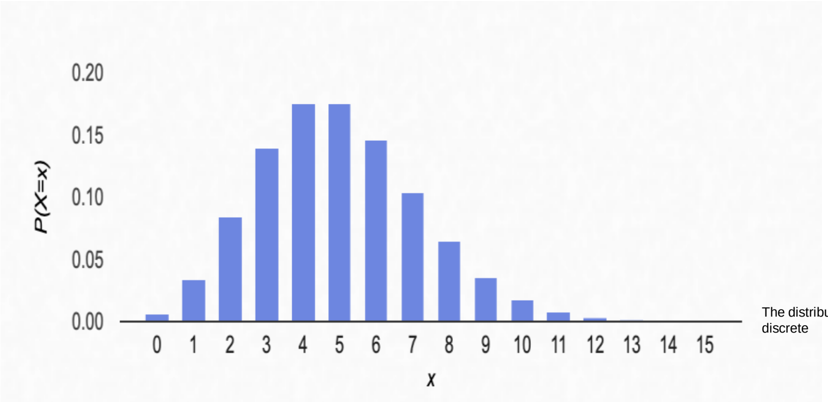


Visualization of Poisson Distribution

Why does it look like this?

- **Right-skewed:** Long tail on the right side and most observations on the left side.
- **Mean λ of 5:** Left side is compressed (dense area), making 5 arrivals per hour the most likely.
- **Plot shows variability:** Most hours have around 5 arrivals; fewer hours have very high numbers.

y-axis →the probability of observing each number of arrivals



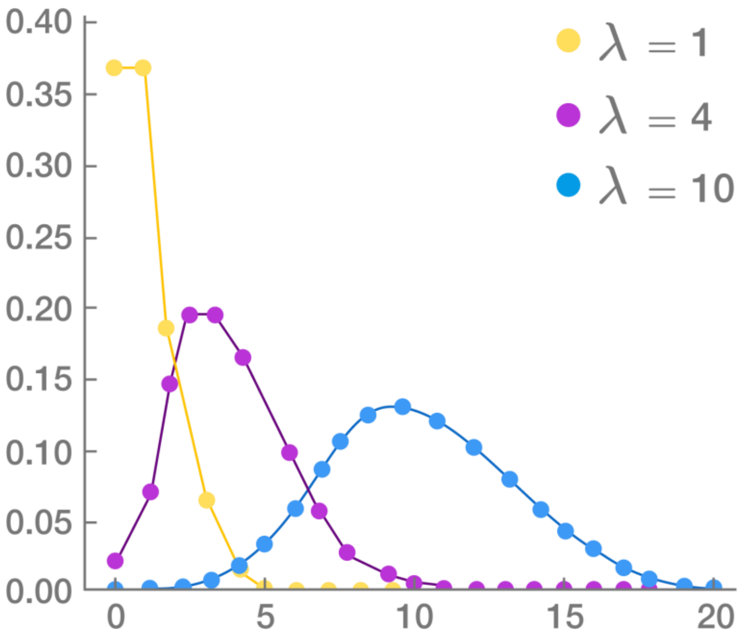
The distribution is discrete

x-axis → the number of people arriving per hour

Effects of Increasing λ in the Poisson Distribution

As λ increases, the distribution becomes more and more similar to a normal distribution, concentrating around the mean and reducing the likelihood of extreme values.

Higher λ means a higher average number of events per interval.



When λ is 10 or greater, a normal distribution is a good approximation of the Poisson distribution.

Example: Poisson Distribution

Example: Births in a hospital occur randomly at an **average rate of 1.8** births per hour. What is the probability of observing **4 births** in a given hour at the hospital?

Sol: Let X =No. of births in a given hour

(i) Events occur randomly

(ii) Mean rate $\lambda = 1.8$

The probability of observing exactly 4 births in a given hour:

$$P(X=4) = \mathbf{e}^{-1.8}\left(\frac{1.8^4}{4!}\right) = 0.0723$$

Zero-Inflated Poisson Distribution

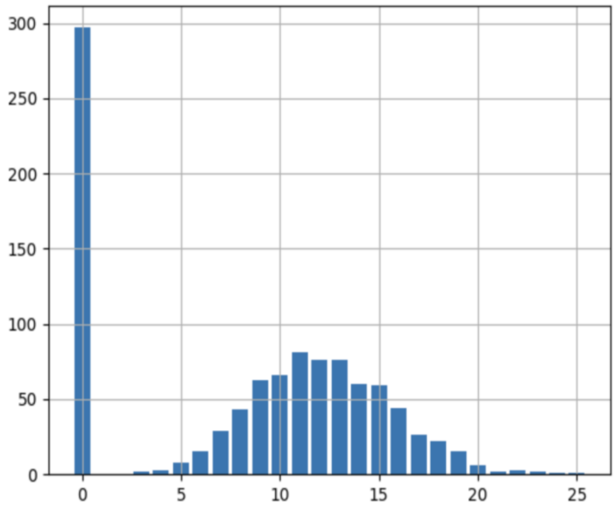
Oftentimes, poisson distributions can have a spike at zero.

- accounts for situations where there's a higher chance of observing zero occurrences.

Remember, Any distribution can be zero inflated

➡ You will need to deal with these differently

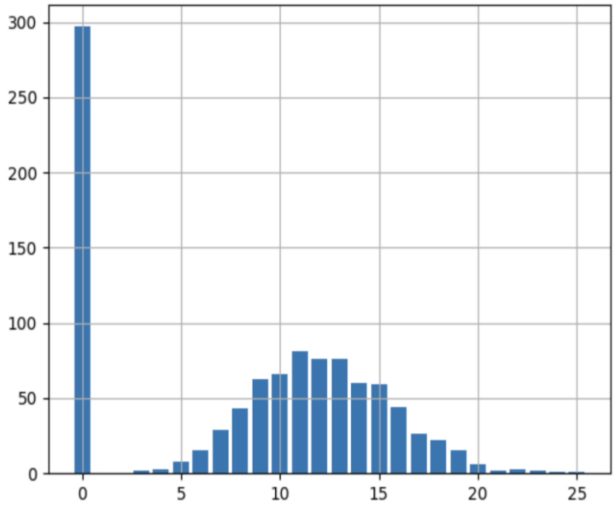
The Zero-Inflated Poisson Distribution addresses this by modeling the probability of excess zeros separately from the rest of the count data.



Example: Zero-Inflated Poisson Distribution

Suppose you have data on the number of patients arriving per hour in an emergency room over a period of time. It is observed that sometimes there are hours with zero patients, even though the average number of patients per hour is around 5.

- Standard Poisson model doesn't fit well due to excess zeros observed.
- Zero-Inflated Poisson distribution considered to model both typical patient arrival process (Poisson distribution) and excess zeros (point mass at zero).



Next: The Bernoulli trial and Binomial Distribution

Next: The Bernoulli trial:

Bernoulli trial

A *Bernoulli* trial is an experiment with the two outcomes "success" and "failure".

We often let p denote the probability of a "success". Two possible outcomes are mutually exclusive ('success': 0, 'failure':1) and often let p denote the probability of a "success".

Examples:

- Flip a coin and call "heads" a "success". If the coin is fair, $p = 1/2$.
- A patient recovers from a disease: recovery(success) or no recovery (failure).

Remember: Each trial is independent (the outcome of one trial does not affect the outcome of another trial.)

Bernoulli Distribution

The Bernoulli distribution is a probability distribution that describes the probability of each possible outcome (success or failure) of a **single Bernoulli trial**.

Models a single trial.

It is a discrete distribution having two possible outcomes (n=0 and n=1)

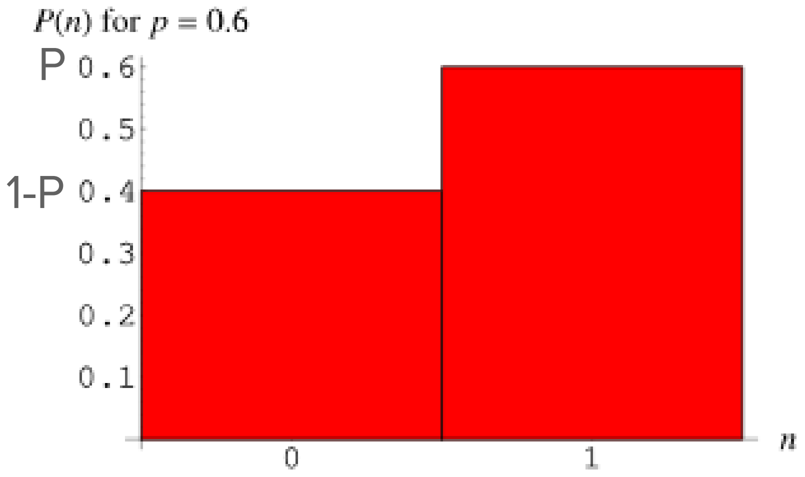
- n=1 ("success") occurs with probability p
- n=0 ("failure") occurs with probability q=1-p,

probability density function

$$P(n) = \begin{cases} 1-p & \text{for } n=0 \\ p & \text{for } n=1, \end{cases}$$

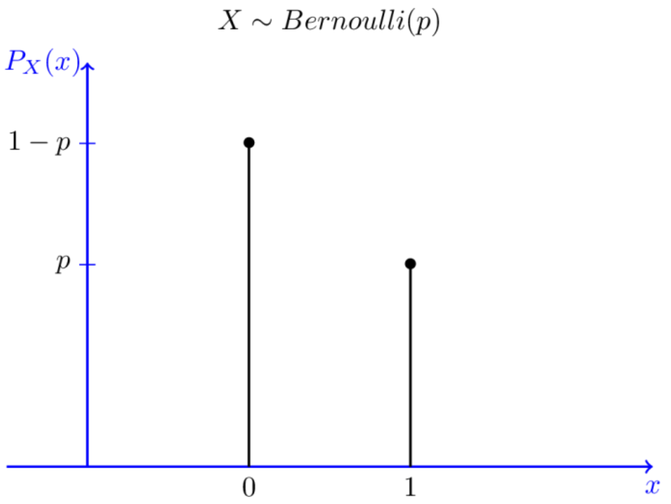
which can also be written

$$P(n) = p^n (1-p)^{1-n}.$$



Ex: Bernoulli Distribution

If a drug has a 10% chance of working, and we observe whether it works for one individual, this scenario follows a Bernoulli distribution.



- It's used when there's only **one trial or event with a fixed probability of success**.
- Essentially, it's just a way to describe the probability of something happening or not happening.

Next: Binomial Distribution

The Binomial Distribution: describes the number of successes and failures in fixed number **n** independent Bernoulli trials for some given value of n.

$$P(X) = {}_{n}C_{x}\, p^{x} (1 - p)^{n-x}$$

Bernoulli: Models a **single** trial.

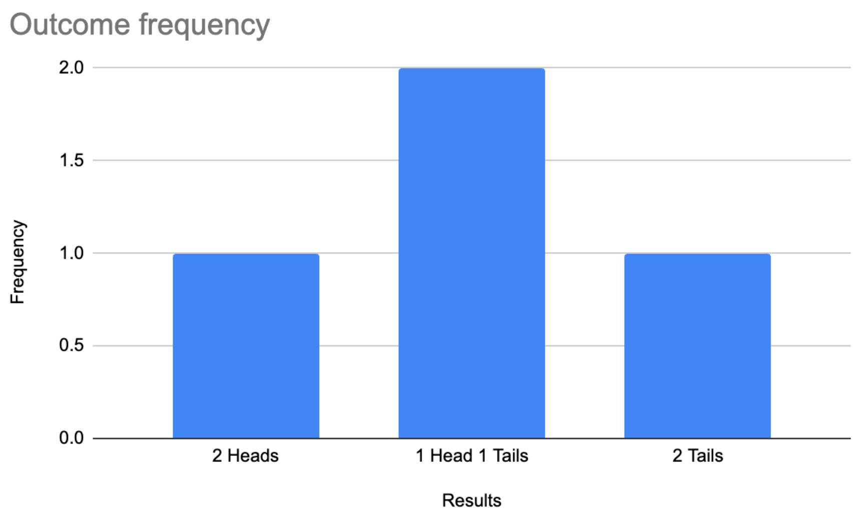
- It is a special case of the Binomial distribution where n=1.

Binomial: Models the number of successes in **n trials**.

- It is the sum of **n** independent Bernoulli random variables.

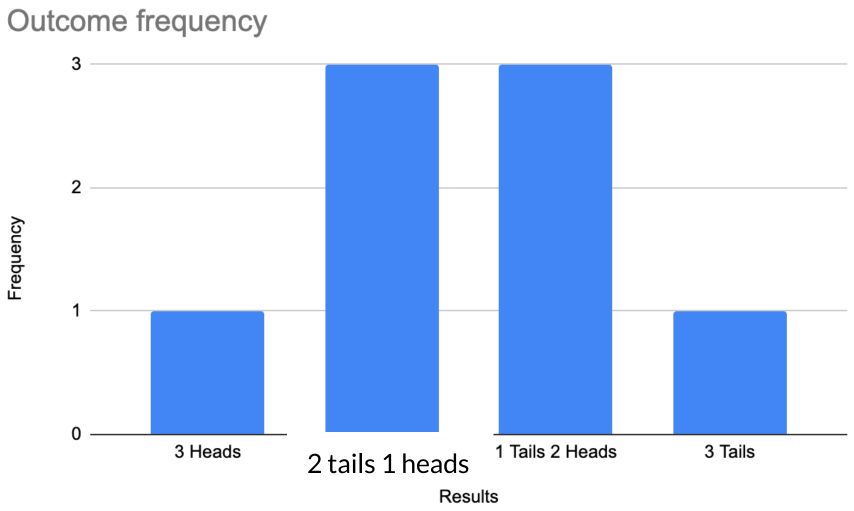
Binomial Distribution

If we flip a coin once, we have a Bernoulli Distribution. What if we flip it twice?



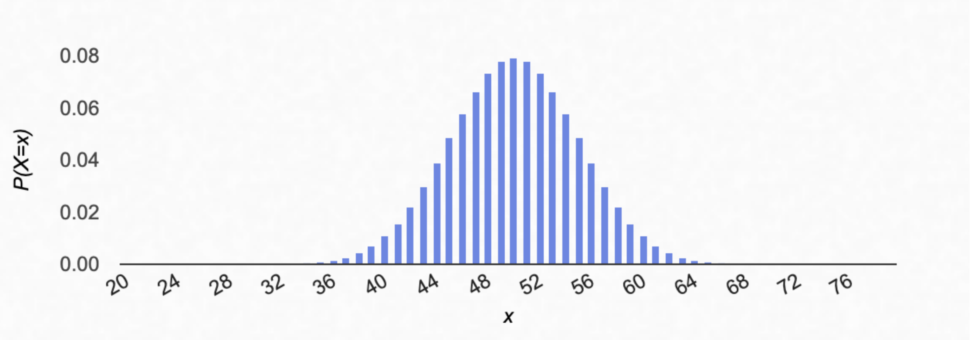
Binomial Distribution

If we flip a coin once, we have a Bernoulli Distribution. What if we flip three times?



Binomial Distribution

A binomial distribution shows us the probability of getting a specific set out of outcomes from a set of Bernoulli Trials.



Binomial vs Bernoulli distribution

Consider :Attempting a quiz that contains 10 True/False questions.

Bernoulli trial ➔ Trying a single T/F question

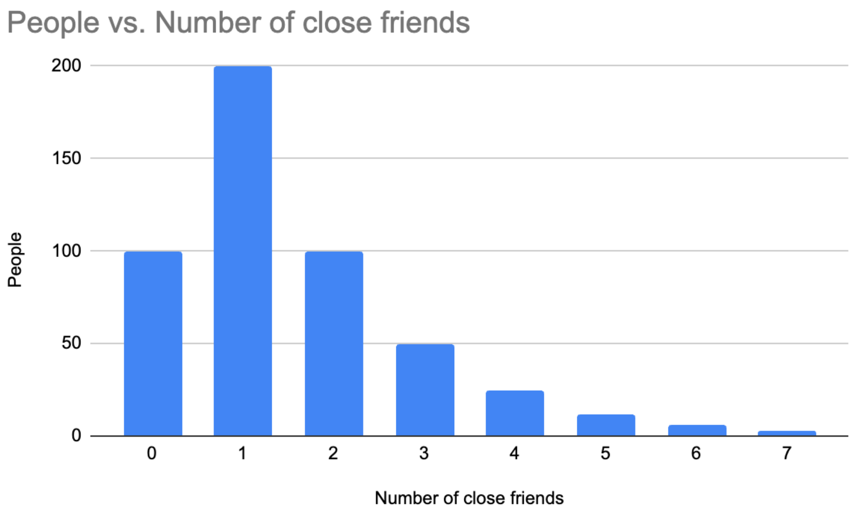
Binomial trial ➔ Attempting the entire quiz of 10 T/F questions

The main characteristics of Binomial Distribution are:

- Independence: Each trial is independent of the others; the outcome of one trial does not affect another.
- Two Outcomes: Each trial can lead to just two possible results (e.g., winning or losing), with probabilities p and (1 – p).

Practice

What does the following discrete distribution tell us? (This is called a Power Law distribution)



Practice

You are given 1000 datapoints representing how many dollars students have in their bank account. What would you expect the distribution to look like?

Probability distributions in Python

The `SciPy` library contains a `stats` module which has various functions for probability distributions. Functions for the probability distributions discussed in this section are described in the table below. Additional functions and further information about parameters can be found in the SciPy stats documentation.

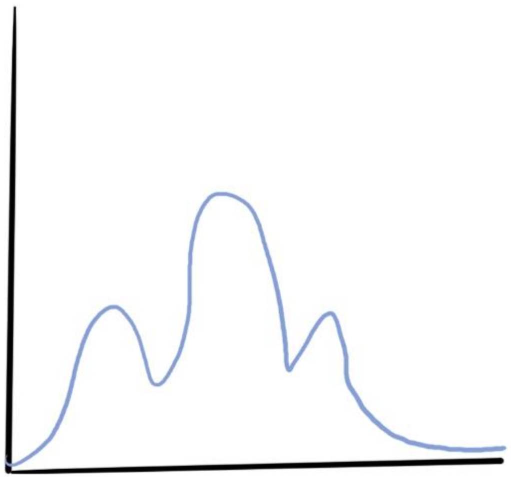
Table 3.4.1: SciPy functions for probability distributions.

Distribution	Functions	Parameters	Description
Bernoulli	`bernoulli.pmf(k, p)` `bernoulli.cdf(k, p)`	`p`$= \pi$ sets the probability of a "success".	`bernoulli.pmf()` returns the probability $P(X = \mathtt{k})$, and the `bernoulli.cdf()` returns the probability $P(X \leq \mathtt{k})$.
Binomial	`binom.pmf(k, n, p)` `binom.cdf(k, n, p)`	`n`$= n$ sets the number of observations. `p`$= \pi$ sets the probability of a "success".	`binomial.pmf()` returns the probability $P(X = \mathtt{k})$, and the `binomial.cdf()` returns the probability $P(X \leq \mathtt{k})$.
Normal	`norm.pdf(x, loc, scale)` `norm.cdf(x, loc, scale)`	`loc`$= \mu$ sets the mean and `scale`$= \sigma$ sets the standard deviation.	`norm.pdf()` returns the density curve's value at `x`, and `norm.cdf()` returns the probability $P(X \leq \mathtt{x})$.

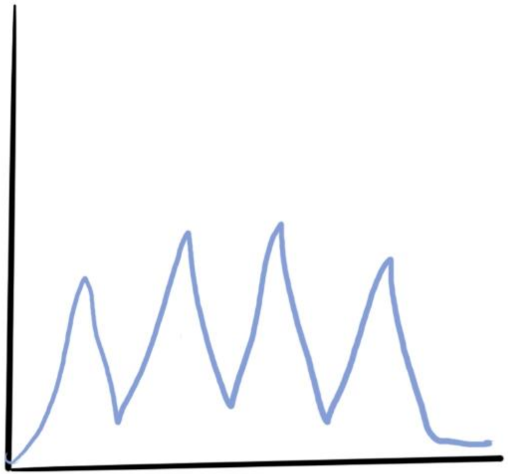
The Central Limit Theorem (CLT):

Before Started

Comparing two non-normal distributions is hard.

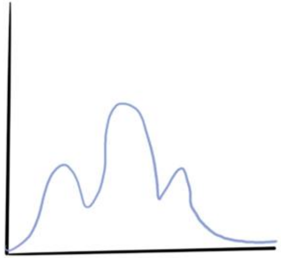


Some annoying distribution



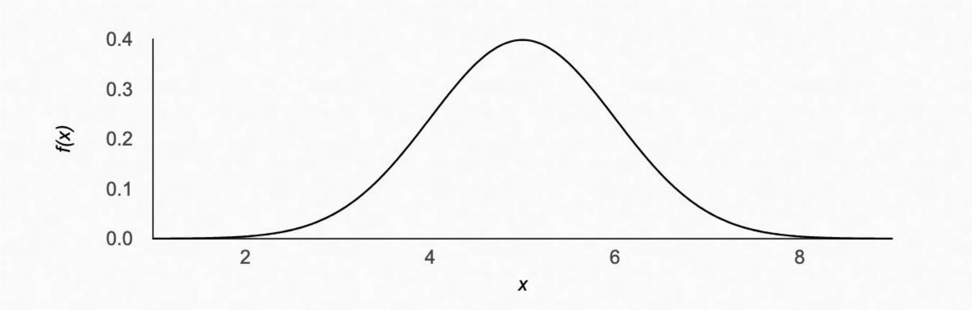
Some other, even more annoying distribution

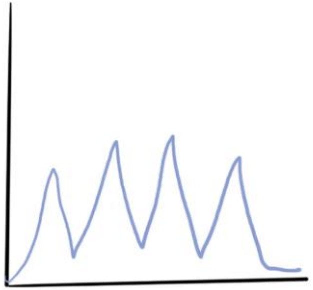
Before Started



Some annoying distribution

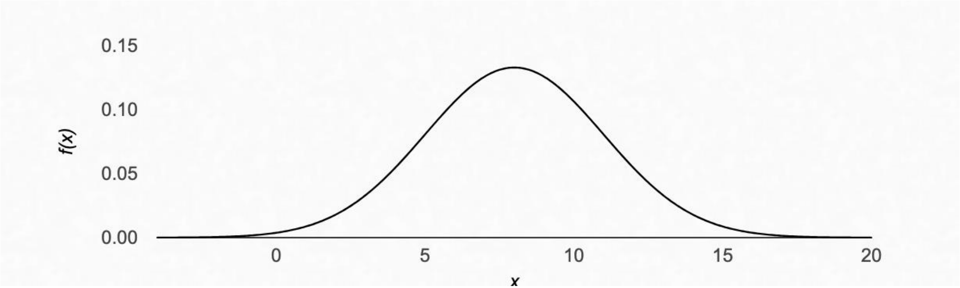
Take a bunch of samples



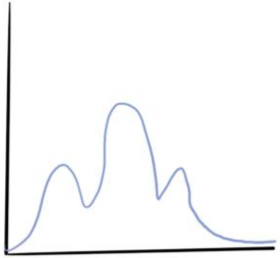


Some other, even more annoying distribution

Take a bunch of samples

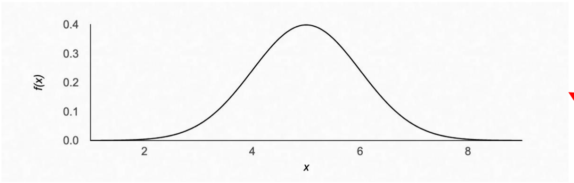


Before Started



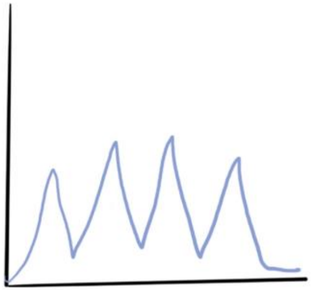
Some annoying distribution

Take a bunch of samples



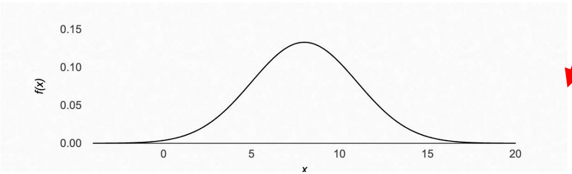
THE OBJECT LEVEL DISTRIBUTION

WE TOOK A BUNCH OF SAMPLES AND HERE ARE THEIR MEANS



Some other, even more annoying distribution

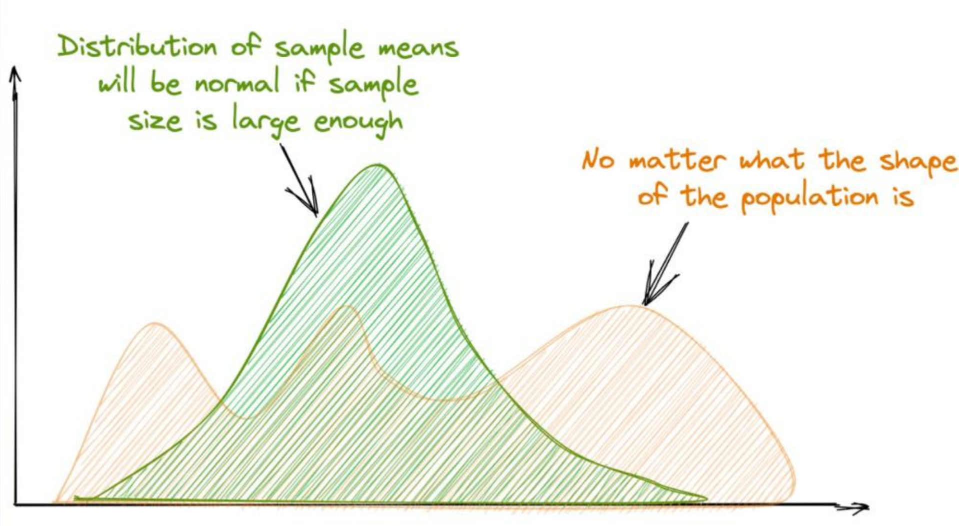
Take a bunch of samples



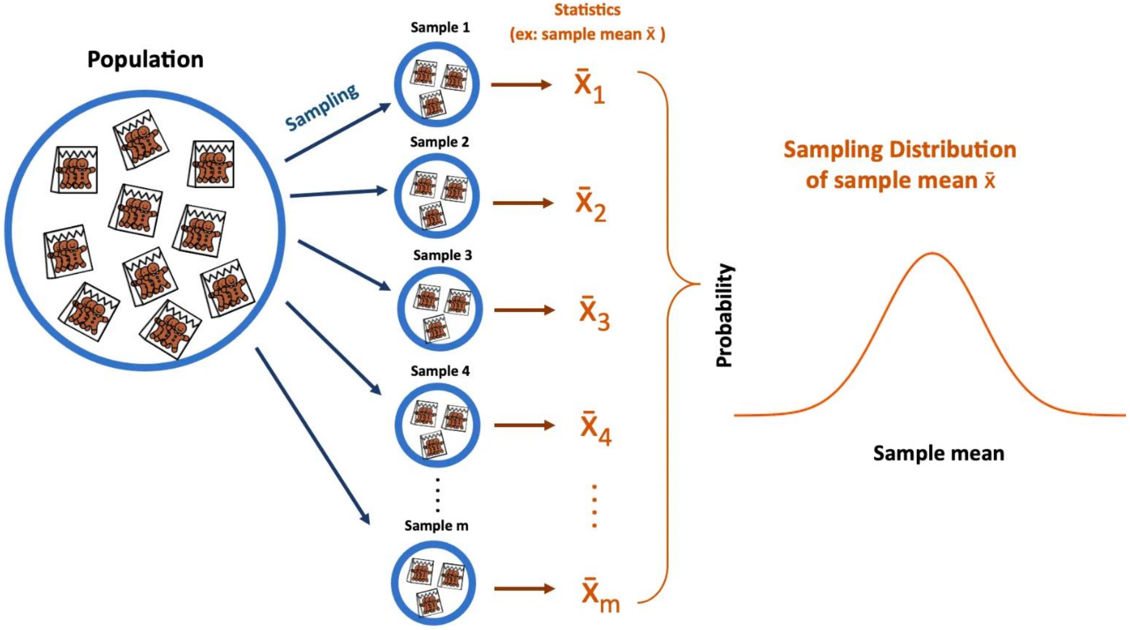
The Central Limit Theorem

- One of the fundamental theorems in statistics.
- States as

> "Given a **sufficiently large sample** size, the sampling distribution of the **mean of a variable** will approximate a normal distribution regardless of that variable's distribution in a population."



The Central Limit Theorem



If we sample a distribution a bunch of times, the set of **sample means** is normally distributed.

Let’s see an example

Sampling

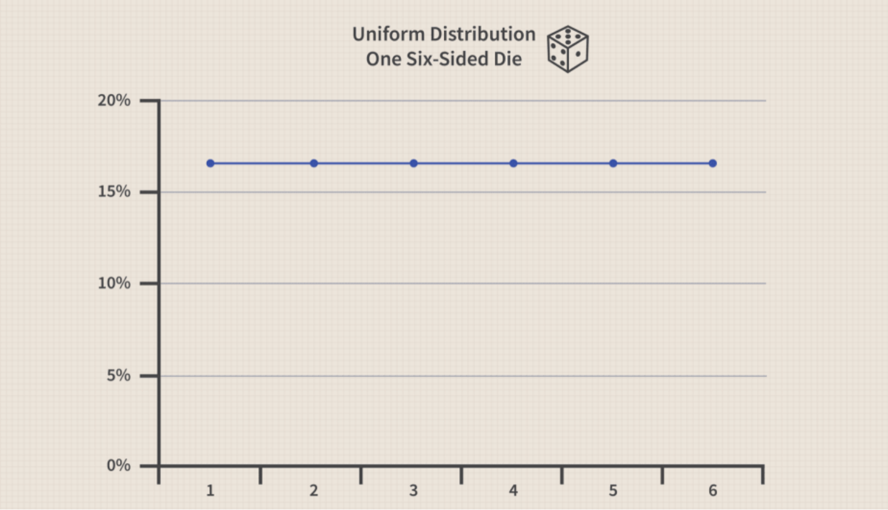
Sampling from a Uniform Distribution

Say I take 5 samples from a uniform distribution

Most of the time, the mean will be 3.5

Sometimes, it won’t be!

What is the rarest mean?



Sampling

To be clear:

If we roll a dice **five times** we might get:

1 2 3 4 5 6 -> Average of 3.5

1 1 6 6 3 4 -> Average of 3.5

1 1 1 1 1 1 -> Average of 1

6 6 6 6 6 6 -> Average of 6

We know we have the highest chances of getting 3.5. Is there a way we can discuss the likelihood of getting the other means?

Yes. A distribution.

Sampling

If you roll a dice five times, here is the frequency with which you will get each of the possible means.

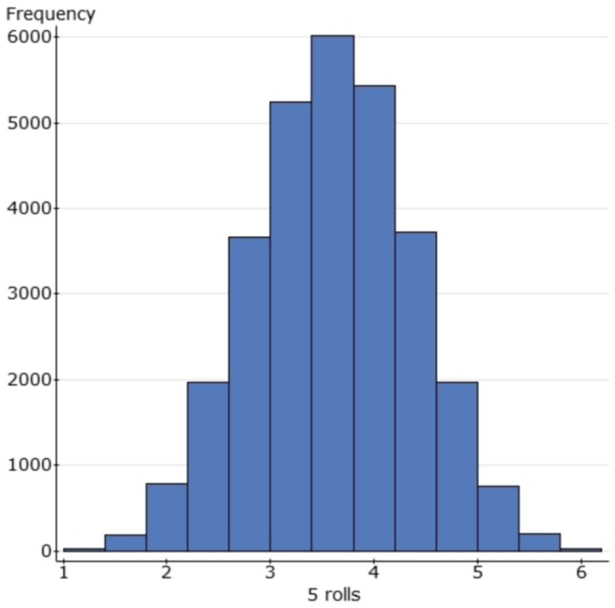
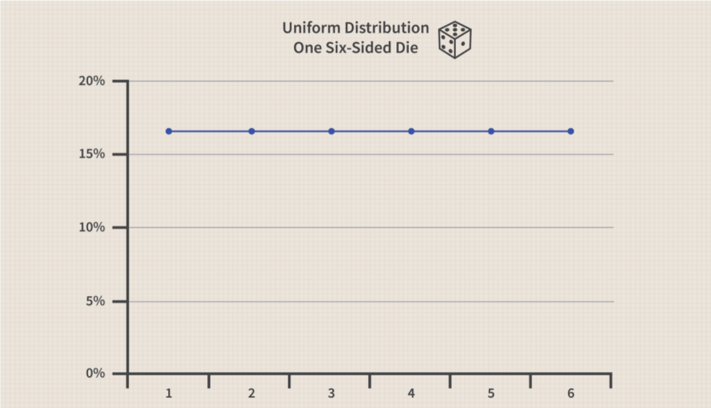


Fig: Histogram showing the frequency of each possible mean when rolling a die five times

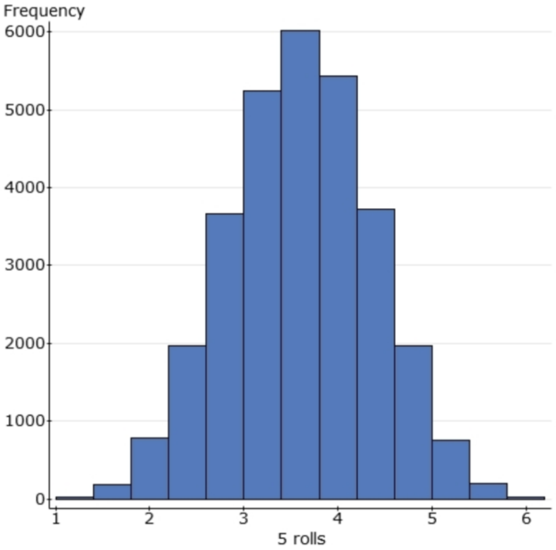
Sampling

According to the Central Limit Theorem, if you roll a die many times and calculate the mean each time, the sample means will follow a normal distribution.

We now have two separate things:



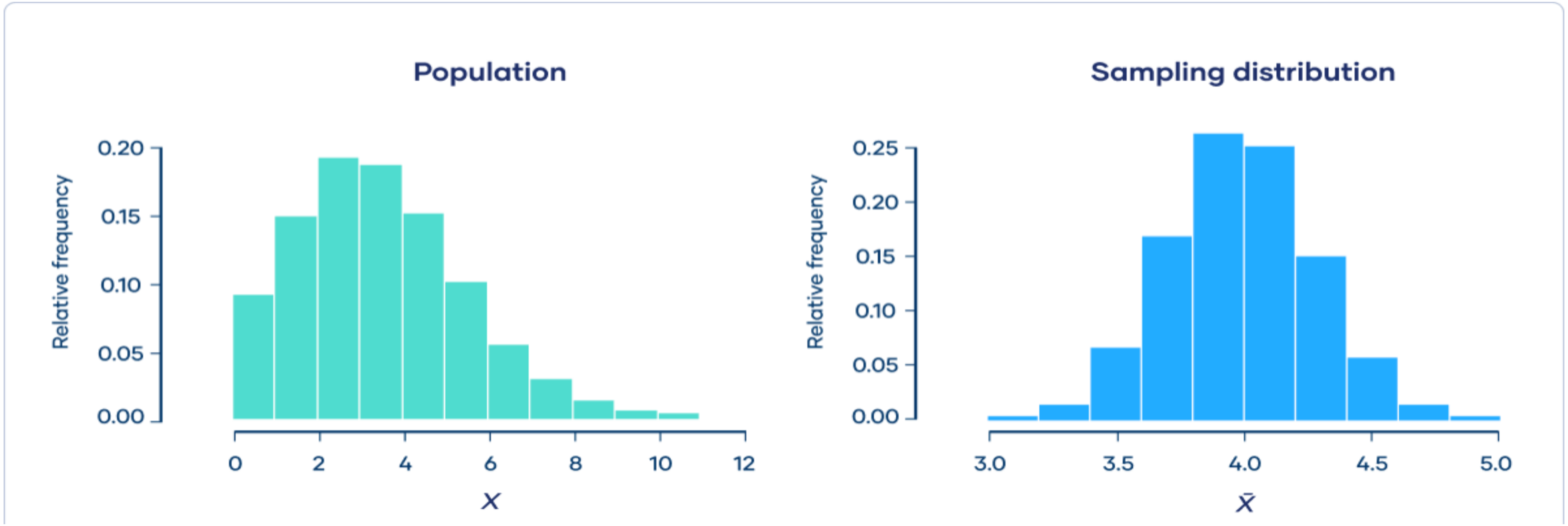
The original distribution



The potential distribution of means when we sample it

Inference: We started with the uniform data distribution but **the means of samples** drawn from it resulted in a normal distribution.

Another Example: The Central Limit Theorem



A **population** follows a **Poisson distribution** **(left image).** If we take 10,000 **samples** from the population, each with a sample size of 50, the sample means follow a normal distribution, as predicted by the **central limit theorem (right image).**

Formula: The Central Limit Theorem

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Sample mean = Population mean = μ

$$\text{Sample standard deviation} = \frac{\text{(Standard deviation)}}{\sqrt{n}}$$

$$= \frac{\sigma}{\sqrt{n}}$$

Formula: The Central Limit Theorem

- The mean of the sampling distribution is the mean of the population.

$$\mu_{\bar{x}} = \mu$$

- The standard deviation of the sampling distribution is the standard deviation of the population divided by the square root of the sample size.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

We can describe the sampling distribution of the mean using this notation:

$$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

Where:

- $\bar{X}$ is the sampling distribution of the sample means
- ~ means "follows the distribution"
- *N* is the normal distribution
- μ is the mean of the population
- σ is the standard deviation of the population
- *n* is the sample size

Central Limit Theorem Application

The Central Limit Theorem (CLT) is useful in various scenarios, particularly when analyzing an entire population is difficult. Here are some key applications:

- **Data Science:** The CLT helps make accurate assumptions about a population to build robust statistical models.
- **Applied Machine Learning:** The CLT aids in making inferences about model performance.
- **Statistical Hypothesis Testing:** The CLT is used to determine if a given sample belongs to a specific population.

Part 02 Descriptive Statistics/ Summary Statistics:

Descriptive Statistics/ Summary Statistics

Data can be presented in many different formats but, what are the main characteristics that describe the data set?

Descriptive statistics involves summarizing, organizing, and presenting data (the features of a dataset) meaningfully and concisely.

- Focus: Describes and analyzes dataset's main features without generalizing to larger populations.
- Enables insights: Helps researchers understand patterns, trends, and distributions within the dataset

What are summary statistics?

- They communicate something about the dataset without needing to understand the whole thing
    - Summarize known data in a way that can be used for further predictions and analysis.
- Very useful for quickly understanding what’s going on
- Examples: Mean, median, mode

How not to use summary statistics

Do NOT take your dataset, compute the mean or median, and then call it a day. If you ever do this I will pop out of your computer and shake you

How not to use summary statistics

Do NOT take your dataset, compute the mean or median, and then call it a day. If you ever do this I will pop out of your computer and shake you

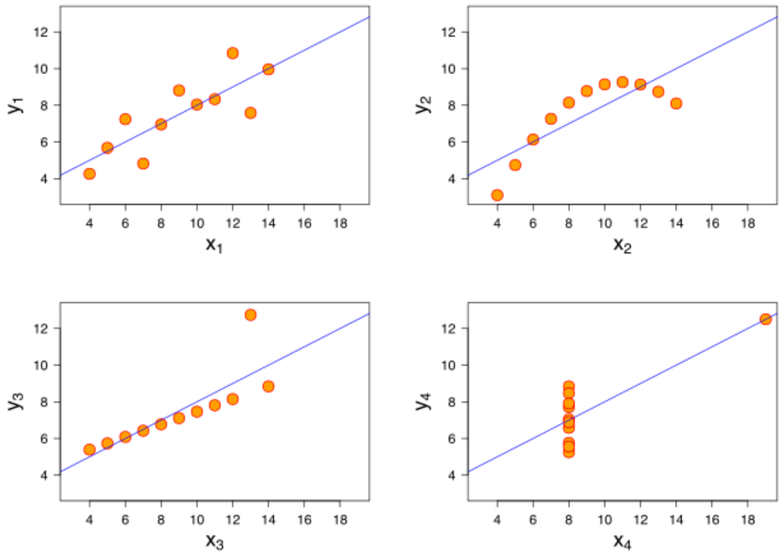
How not to use summary statistics

Do NOT take your dataset, compute the mean or median, and then call it a day. If you ever do this I will pop out of your computer and shake you

If you do not understand your data ahead of computing the summary statistics, they can end up being useless, or actively misleading.

Why not to rely on summary statistics

All four of these have the same mean and variance, but are clearly generated by four different processes.



Summary stats are for *after* you understand your data holistically (though you can use them to help that process)

Types of Descriptive/ Summary Statistics

1. Measures of central tendency
2. Measures of variability (spread)
3. Measures of Skewness
4. Modality

1. Measures of Central Tendency

1. Measures of Central Tendency

Measures of Central Tendency tell you about the center of your distribution. It represents the whole set of data by a single value. These include:

- **The Pythagorean means**
  - **Arithmetic Mean**
  - **Geometric Mean**
  - **Harmonis Mean**
- **Median**
- **Mode**

Measures of Central Tendency

Let y denote a quantitative variable, with observations y_1 , y_2 , y_3 , ... , y_n

a. Describing the center

Median: Middle measurement of ordered sample.

- The value in the dataset that has an equal number of items greater than and less than it.

Mean: Average of all observations in the dataset.

Mode: The most common item in your dataset

Pythagorean Means

Pythagorean means refer to **three separate concepts:** the arithmetic mean, the geometric mean, and the harmonic mean.

- These means are named after the Pythagorean theorem because they are all forms of averaging.
- Provide different ways of summarizing a set of numbers.

Pythagorean Means

- **Arithmetic Mean**: Your typical average
  - Sensitive to extreme values and is affected by outliers as it incorporates every value in the dataset into its calculation.
    - Gives each value equal weight in the calculation, regardless of its distance from the center of the dataset.

$$\mathrm{AM}(x_1, \ldots, x_n) = \frac{x_1 + \cdots + x_n}{n}$$

Pythagorean Means

- **Geometric Mean**: Multiply all the values in a dataset and then take the nth root of the product (n = number of values).
  - Suitable for datasets with proportional values, such as growth rates or investment returns, as it accounts for relative changes rather than absolute values.
    - Ideal for data that changes or grows over time, like COVID cases, investment returns, or population growth

$$\mathrm{GM}(x_1, \ldots, x_n) = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} = \sqrt[n]{x_1 \cdot x_2 \cdots x_n}$$

Geometric Means is Less sensitive to outliers

- Why? It considers the relative magnitude of values, extreme values have less influence on the geometric mean)

Example: Consider two datasets of investment returns (in percentage):

Dataset A: 2%, 4%, 6%, 8% Dataset B (with an outlier): 2%, 4%, 6%, 100%

C

1. **Arithmetic Mean (sensitive to outliers)**:
   - Dataset A: (2% + 4% + 6% + 8%) / 4 = 5%
   - Dataset B: (2% + 4% + 6% + 100%) / 4 = 28%
2. **Geometric Mean (less sensitive to outliers)**:
   - Dataset A: $\sqrt[4]{2 \times 4 \times 6 \times 8} \approx 4.76$
   - Dataset B: $\sqrt[4]{2 \times 4 \times 6 \times 100} \approx 6.94$

Observe: In Dataset B, the geometric mean is less affected by the extreme 100% value compared to the arithmetic mean.

This is because the geometric mean looks at the product of the values (considering their relative sizes) rather than just their sum.

Pythagorean Means: (Harmonic Mean)

- **Harmonic Mean**: is calculated by taking the reciprocal of the arithmetic mean of the reciprocals of all the values in a set.

- Used to calculate the mean of rates or ratios.
- Ideal for averaging rates or inverse values, such as speeds, work rates, or densities.
- Ex: Investment Returns: when calculating the average of multiple ratios or fractions, such as price-to-earnings ratios in finance.

$$\text{Harmonic Mean} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

- **Less sensitive to outliers:** Gives more weight to smaller values, providing robustness against extremely large values.
  - **Undefined for datasets containing zero values.**

Harmonic Means: Examples

Example for Speed: Consider the following example where a car travels over equal distances at different speeds:

- 60 km/h for the first half of the trip
- 40 km/h for the second half of the trip

Try Arithmetic Mean = $(60+40)/2=50$ km/h does not accurately reflect the total time taken for the trip because it **assumes equal weighting by speed rather than distance**.

Try Geometric Mean= $\sqrt{(60*40)}/2 \sim 48.99$ km/h: handle proportional growth rates better than the arithmetic mean, it still **does not correctly address the relationship between speed and time over equal distances**.

Solution: Harmonic Mean $=2/\ (1/60 + 1/40) = 2/\ (2/120 + 3/120) = 2/(5/120) = (2\times120)/5 = 48$ km/h

The harmonic mean accurately reflects the average speed over the entire trip because **it takes into account the time spent traveling at each speed**. It gives <u>more weight to lower speeds</u>, which is important when averaging rates where the time spent at each rate matters.

2. Measures of Variance

2. Measures of Variance

An important characteristic of any set of data is the variation in the data; it reflects how tightly or widely data points are distributed around the mean.

The **standard deviation**. is the most common measure of this spread.

Good one: https://www.shiksha.com/online-courses/articles/variance-and-standard-deviation/

The standard deviation σ and The Variance

The Standard Deviation: provides a numerical measure of the overall amount of variation in a data set, and

- Can be used to determine whether *a particular data value is close to or far from the mean*.
- It's always positive or zero.
  - When the data is clustered close to the average (mean), the **standard deviation is small**.
  - When the data is spread out from the mean, **the standard deviation is larger.**

The Variance: the average of the squared differences from the Mean; indicates how far individuals in a group are spread out.

- **Standard deviation is the square root of the variance.**

Formulas: Measures of Variance

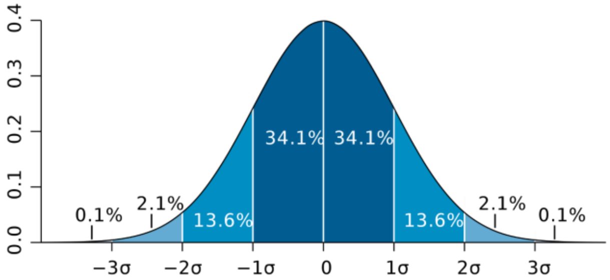
The procedure to calculate the standard deviation depends on whether the **numbers are the entire population or are data from a sample**. The calculations are similar, but not identical.

Sample Variance

$$s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$$

Sample Standard Deviation

$$s = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}$$



NOTE: In general, the sample standard deviation is preferred over the sample variance

- ***Interpretability***: standard deviation measured in same units as the original data
- **Scale Sensitivity:** Variance involves squaring differences, making comparisons across datasets with different units or scales challenging.

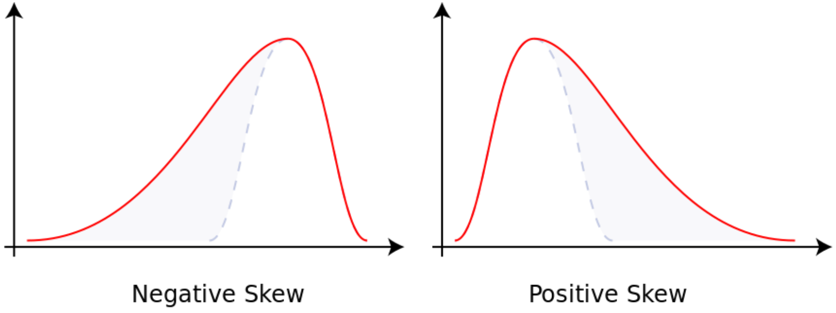
3. Measures of Skewness: Other Descriptors

3. Measures of Skewness: Other Descriptors

Skewness is essential for understanding the shape and structure of the distribution of the data.

Skew (left or right tailed)

- Skew measure asymmetry in the distribution of data.
- whether the distribution is shifted to the left or right.



Positive skew (right-tailed): the tail of the distribution extends more towards the higher values

- Mean is typically greater than the median; more data points are on the left side of the distribution. Ex: Income distribution, where there are a few extremely high incomes.

Negative skew (left-tailed): indicates the opposite, with the tail extending more towards the lower values.

- Mean is typically less than the median; more data points are on the right side of the distribution. Ex: Test scores of students, where there are a few extremely low scores.

4. Modality: Other Descriptors

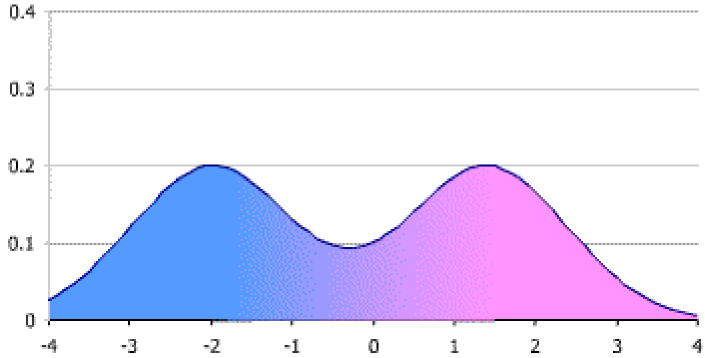
4. Modality: Other Descriptors

Modality is useful because it provides insights into the underlying structure (pattern or cluster) and characteristics of the data

Modality: Refers to the number of peaks or modes in a distribution. A distribution can be:

- Unimodal (one peak),
- Bimodal (two peaks),
- Trimodal (three peaks), or
- Multimodal (more than three peaks)

depending on how the modes are distributed within the data.



The End

Additional Reading Slides

What is probability?

- We are taught that probability is the chance that something will happen: The chances of rolling a one on the dice is a ⅙
- If we built a machine that could predict the exact physics of a dice roll, would the chance of a 1 still be ⅙?

Probability is a measure of *uncertainty*

If I predict I will see a 1 on the dice, I will expect to be correct 1/6th of the time.

If I was smarter, my probability might be different, because I would have less uncertainty.

Some Basic Concepts to revise

Consider rolling a fair six-sided die:

Events: A = "rolling an even number" (comprising outcomes {2, 4, 6}) and B = "rolling a number less than 4" (comprising outcomes {1, 2, 3}).

Sample Space: $\boldsymbol{S}$ = {1, 2, 3, 4, 5, 6}.

Random Variable: Let X be the random variable representing the outcome of the roll. X can take on values from the sample space: $=\{1,2,3,4,5,6\}$ $X=\{1,2,3,4,5,6\}$.

Probability Axioms:

- $P(A)=3/6=0.5$ (since there are 3 outcomes in A out of 6 total outcomes).
- $P(B)=0.5$ (similarly, there are 3 outcomes in B out of 6 total outcomes).
- $P(A\cup B)=P(A)+P(B)=0.5+0.5=1$, as events A and B are mutually exclusive.

Conditional Probability

Conditional Probability Definition

This Example is from Professor Max's Slides

Conditional Probability Formula

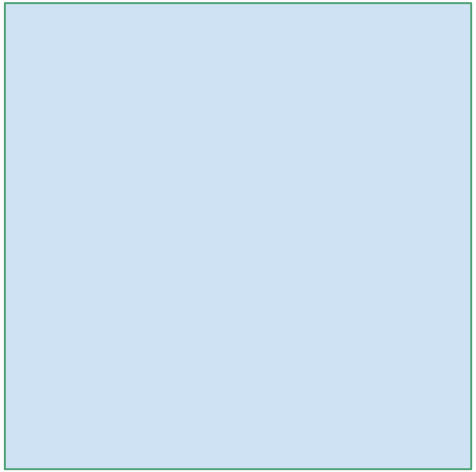
Probability of A and B

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

Probability of A given B

Probability of B

Imagine we have event B, which has a 80% chance of happening.



The World

Conditional Probability Definition

Conditional Probability Formula

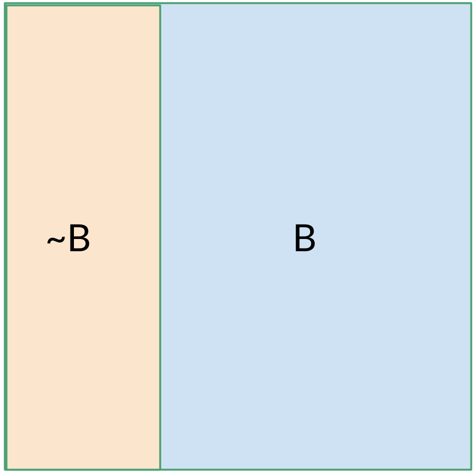
Probability of A and B

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

Probability of A given B

Probability of B

Imagine we have event B, which has a 80% chance of happening.



The World

This Example is from Professor Max's Slides

Conditional Probability Definition

Conditional Probability Formula

Probability of *A* and *B*

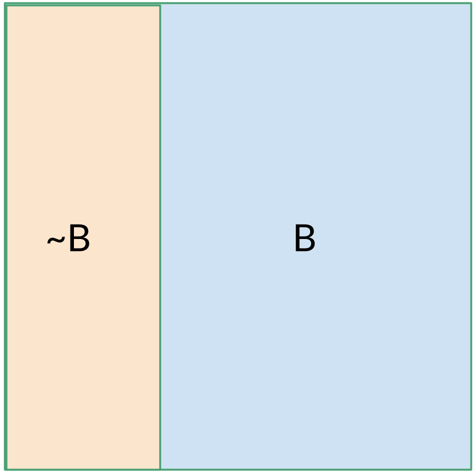
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Probability of *A* given *B*

Probability of *B*

Imagine we have event B, which has a 80% chance of happening.

P(B) is the area of the blue box



The World (Area 1)

This Example is from Professor Max's Slides

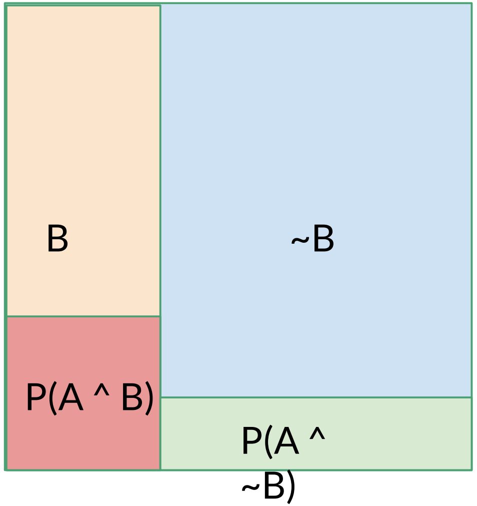
Conditional Probability Definition

Conditional Probability Formula

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

Probability of *A* given *B* — Probability of *A* and *B* — Probability of *B*

Imagine we have event B, which has a 20% chance of happening.



The World (Area 1)

This Example is from Professor Max’s Slides

Conditional Probability Definition

Conditional Probability Formula

Probability of A and B

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

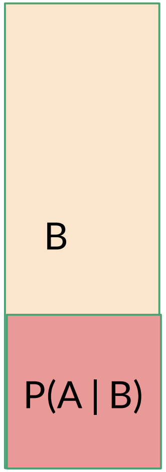
Probability of A given B

Probability of B

Imagine we have event B, which has a 80% chance of happening.

P(A | B) is the **proportion** of of the B square taken up by the red square.

Hence P(A ^ B) / P(B)

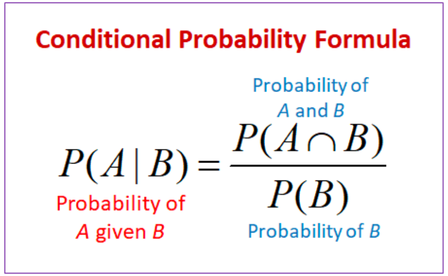


The World We Care About (Area Now 1)

This Example is from Professor Max's Slides

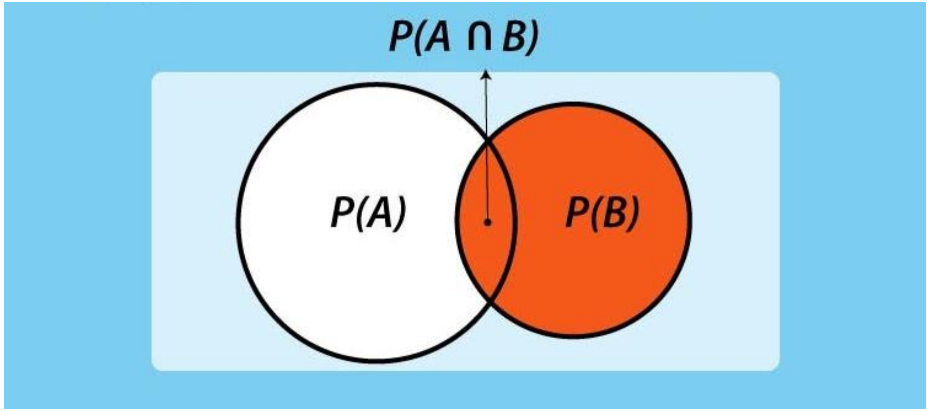
More Explanation to understand Conditional Probability

Conditional Probability Definition



Imagine you have two events, **A** and **B**, represented by two overlapping circles:

- **Circle A** represents all the outcomes where event A happens.
- **Circle B** represents all the outcomes where event B happens.

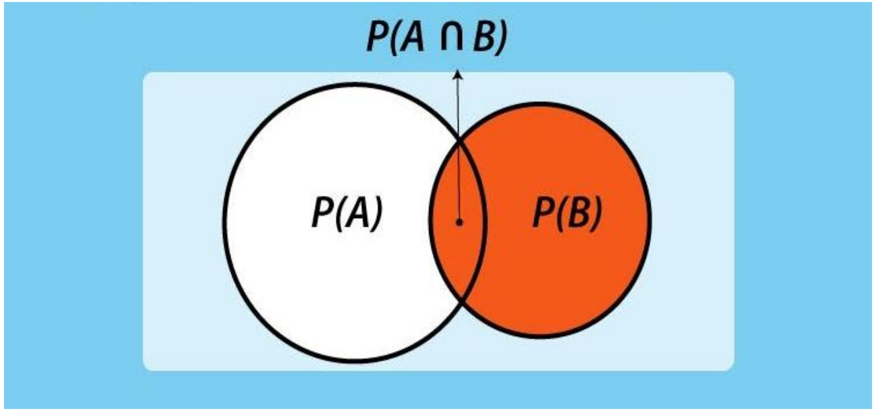


Venn diagram for Conditional Probability, P(A|B)

Conditional Probability Definition

The area inside each circle represents the probability of each event happening:

- **P(A)** is the probability of event A occurring (the size of Circle A).
- **P(B)** is the probability of event B occurring (the size of Circle B).
- **P(A and B)** is the probability of both A and B happening together (the overlap between Circles A and B)



Venn diagram for Conditional Probability, P(A|B)

Conditional Probability Formula

Probability of A and B

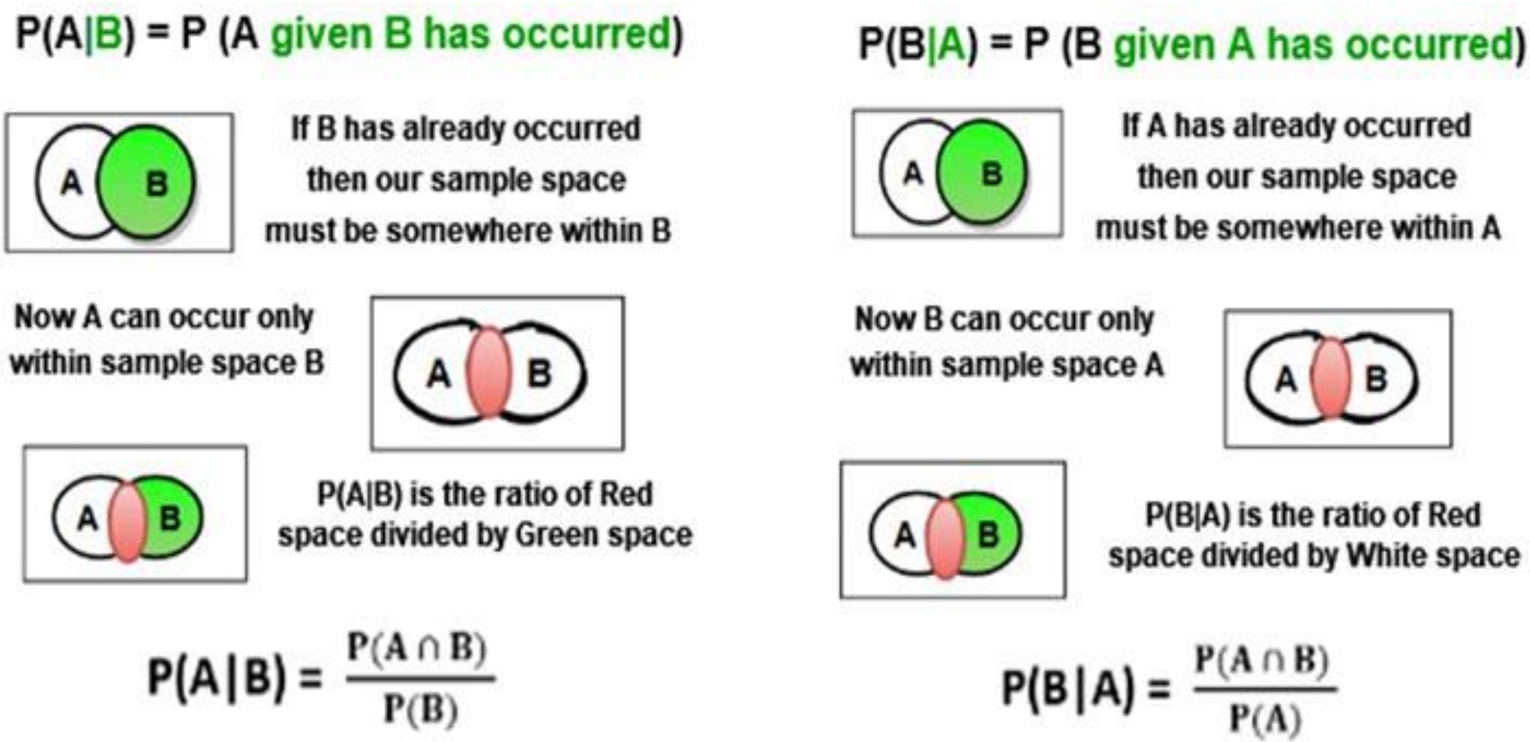
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Probability of A given B

Probability of B

By dividing by P(B), we focus our attention on the subset of cases where event B is true.

Conditional Probability Definition



Conditional Probability Definition

Conditional Probability Formula

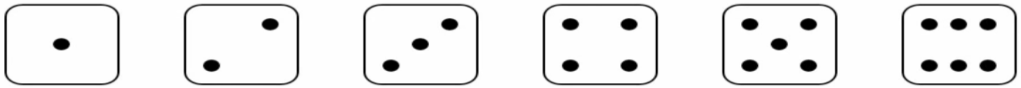
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Probability of A given B

Probability of A and B

Probability of B

Example:



$$\frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1$$

P(B | A) = What is the Probability of (rolling a dice and it's value is less than 4 | knowing that the value is an odd number)

Conditional Probability Definition

Example:

P(B | A) = What is the Probability of (rolling a dice and it's value is less than 4 | knowing that the value is an odd number)

Conditional Probability Formula

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

Probability of A given B

Probability of A and B

Probability of B

P(B): rolling a dice and it's value is less than 4

$$\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$

P(A): rolling a dice and it's value is an odd number

$$\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$

Conditional Probability Definition

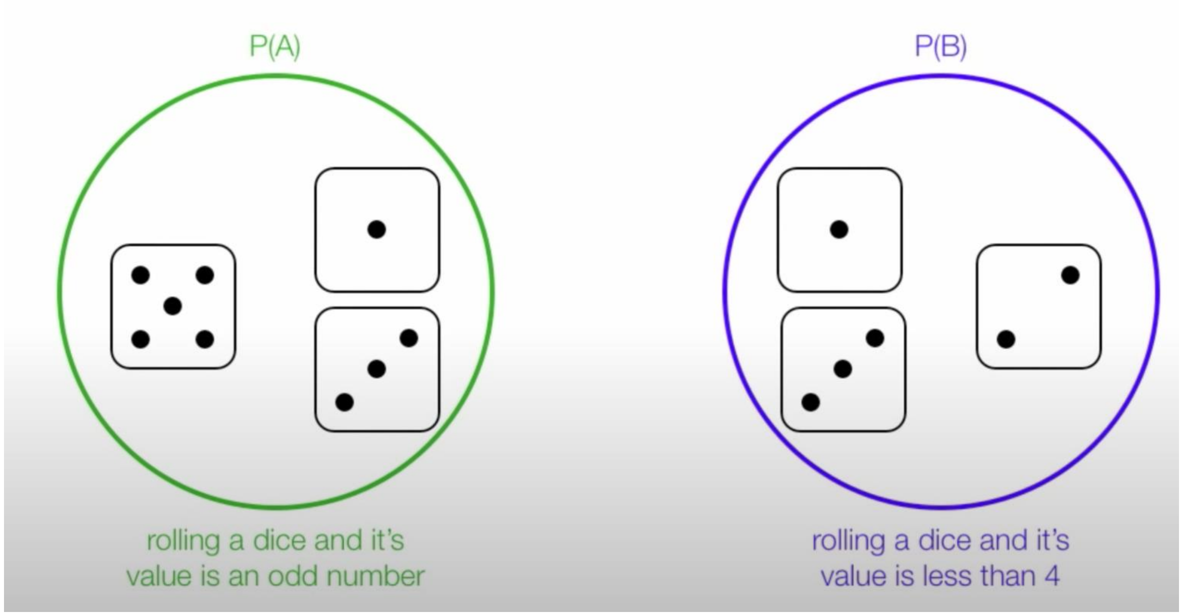
Example:

P(B | A) = What is the Probability of (rolling a dice and it's value is less than 4 | knowing that the value is an odd number)

Conditional Probability Formula

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

Probability of *A* given *B*; Probability of *A* and *B*; Probability of *B*



Conditional Probability Definition

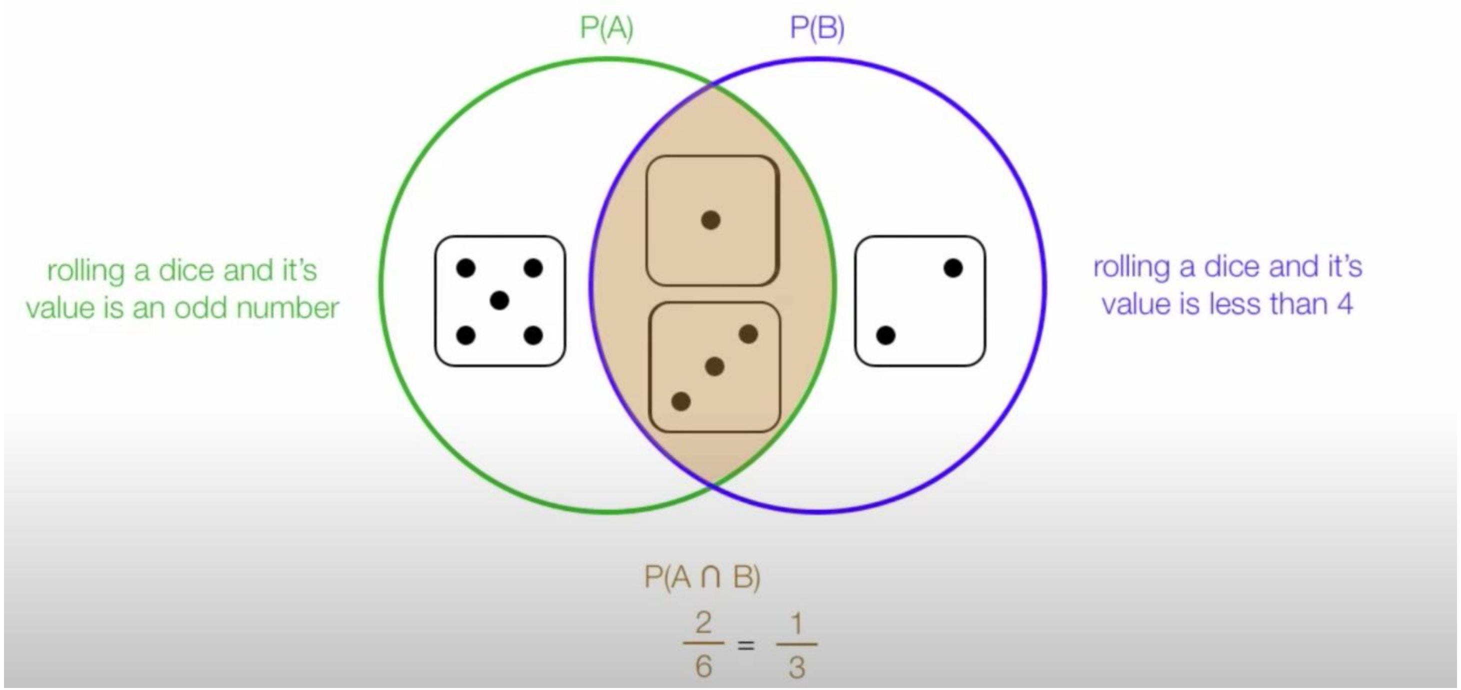
Example:

P(B | A) = What is the Probability of (rolling a dice and it's value is less than 4 | knowing that the value is an odd number)

Conditional Probability Formula

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

Probability of A given B = Probability of A and B / Probability of B



Conditional Probability Definition

Example:

P(B | A) = What is the Probability of (rolling a dice and it's value is less than 4 | knowing that the value is an odd number)

Conditional Probability Formula

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

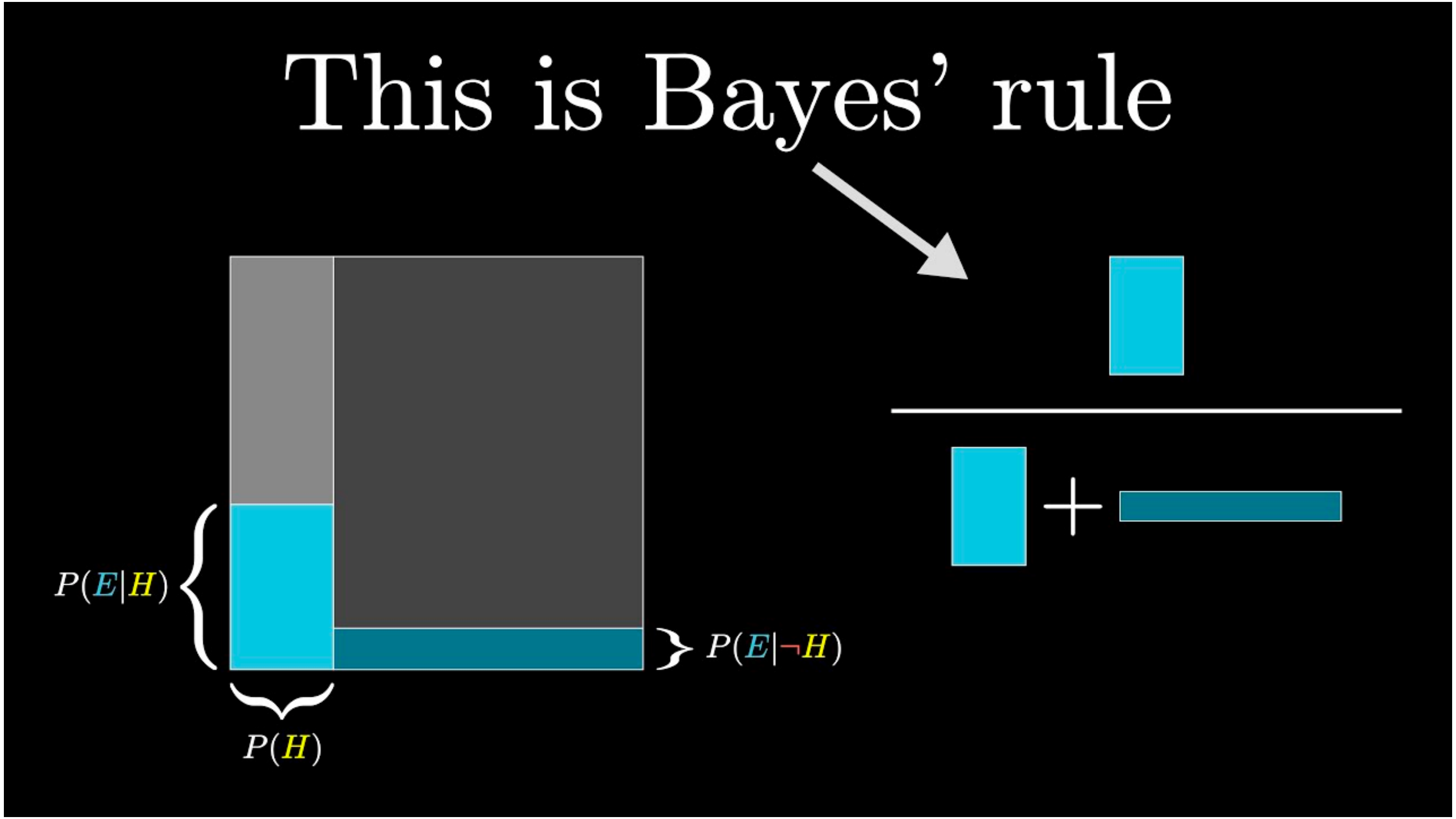
Probability of A given B; Probability of A and B; Probability of B

What is the Probability of

rolling a dice and it's value is less than 4

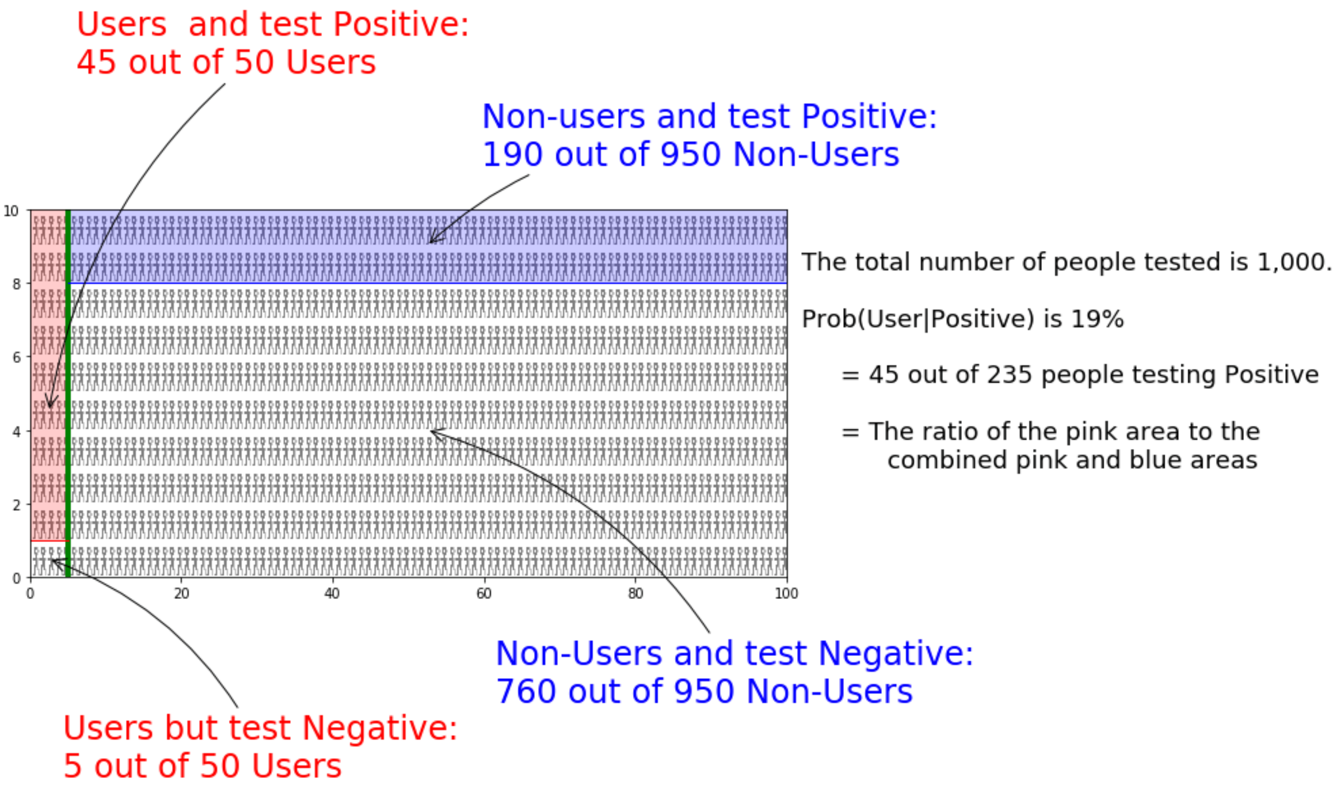
knowing that the value is an odd number

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$


This is Bayes' rule
$P(E|H)$
$P(E|\neg H)$
$P(H)$

Example:

Ref: https://en.wikipedia.org/wiki/Bayes%27_theorem



Bayes Theorem

Example Scenario: Bayes Theorem

Let's consider a medical diagnosis scenario. Imagine a patient undergoes a diagnostic test for a particular disease. The test result may come back positive or negative, indicating whether the patient has the disease or not. However, no test is perfect, and there's always some degree of uncertainty associated with the result.

- We want to know the probability that the patient actually has the disease (A) given the result of the diagnostic test (B).

Event A: **Represents the occurrence of the disease.**
Event B: **Represents the outcome of the diagnostic test.**

Example Scenario: Bayes Theorem

Let's consider a medical diagnosis scenario. Imagine a patient undergoes a diagnostic test for a particular disease. The test result may come back positive or negative, indicating whether the patient has the disease or not. However, no test is perfect, and there's always some degree of uncertainty associated with the result.

Prior Probability P(A) (our initial belief or probability regarding the occurrence of an event before considering any new evidence): the likelihood of a randomly chosen person having the disease based on population data.

Likelihood P(B|A)): Suppose there's a diagnostic test for the disease. Clinical studies have shown that this test correctly identifies 95% of individuals who have the disease (true positives). **This is the likelihood of observing a positive test result given that the person actually has the disease**.

Event A: Represents the occurrence of the disease.
Event B: Represents the outcome of the diagnostic test.

Let's consider a medical diagnosis scenario. Imagine a patient undergoes a diagnostic test for a particular disease. The test result may come back positive or negative, indicating whether the patient has the disease or not. However, no test is perfect, and there's always some degree of uncertainty associated with the result.

Total Probability (P(B)): Now, let's say we administer the test to a large group of people. Some of them may have the disease, and some may not. The total probability of observing a positive test result P(B)) is the **sum of two probabilities**:

- The probability of a **true positive result** (observing a positive test result given that the person has the disease).
- The probability of a **false positive result** (observing a positive test result given that the person does not have the disease).

Event A: Represents the occurrence of the disease.
Event B: Represents the outcome of the diagnostic test.

Let's consider a medical diagnosis scenario. Imagine a patient undergoes a diagnostic test for a particular disease. The test result may come back positive or negative, indicating whether the patient has the disease or not. However, no test is perfect, and there's always some degree of uncertainty associated with the result.

Posterior Probability (P(A|B)): After administering the test, we obtain a positive result for a particular individual.

- Now, we want to calculate the **probability that this person actually has the disease**.
- This is the posterior probability of the person **having the disease given the positive test result**.

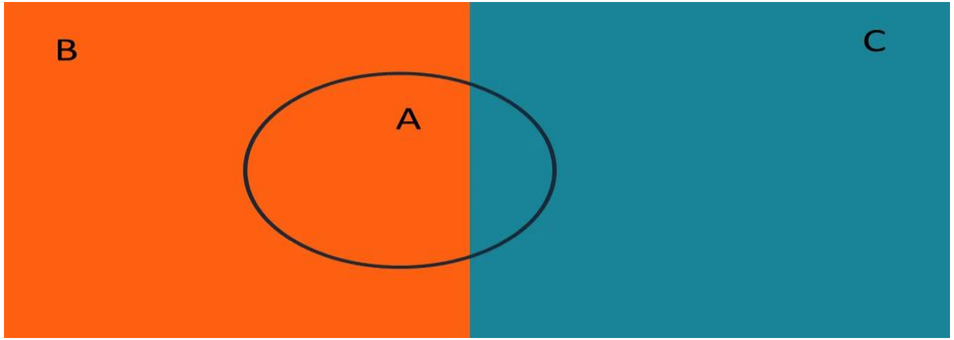
$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Event A: Represents the occurrence of the disease.
Event B: Represents the outcome of the diagnostic test.

- The Total Probability Rule/ Law of Total Probability is a fundamental rule in statistics relating to conditional and marginal probabilities.
- States that "*if the probability of an event is unknown, it can be calculated using the known probabilities of several distinct events*".
- Allows to calculate the probability of an event by considering all possible ways or sources that the event can occur.

Law of Total Probability

Law of Total Probability



- Consider three events: A, B, and C. Events B and C are distinct from each other, while event A intersects with both events.
- We do not know the probability of event A.
- However, we **know** the probability of e**vent A under condition B** and the probability of **event A under condition C**.
- **The total probability rule states that by using the two conditional probabilities, we can find the probability of event A.**

The Law of Total Probability

"An **event B** can be expressed as the sum of the conditional probabilities of B occurring given different possible scenarios or "partitions" Ai of the sample space."

- If A_1, A_2, ..., A_k are mutually exclusive with $\sum_{i=1}^{k} P(A_i) = 1$, i.e. collectively exhaustive, then for any event B,

$$P(B) = P(B|A_1) \cdot P(A_1) + P(B|A_2) \cdot P(A_2) + \cdots + P(B|A_k) \cdot P(Ak)$$

For more read:
https://drive.google.com/file/d/172DG9tr1rARfNfB4qYN06nTjVnl2GFyJ/view?usp=sharing

Try Yourself

Scenario: Imagine you are organizing a charity event, and there are three possible venues (A, B, and C) where you can hold the event. The probability of each venue being available on a given day is as follows:

Venue A: 40% chance of being available.

Venue B: 30% chance of being available.

Venue C: 30% chance of being available.

You also know that if **Venue A is available, there's a 70% chance of raising a large amount of money for the charity**, while if **Venue B or C is available, there's a 50% chance of raising a large amount of money**.

What is the overall probability L of raising a large amount of money at the charity event?

Solution

A: Venue A is available; B: Venue B is available; C: Venue C is available.

L: Raising a large amount of money.

We want to find : **P(L), the probability of raising a large amount of money at the charity event.**

P(L) = the sum of probabilities weighted by the probability of each venue being available.

$$P(L) = P(L|A) \cdot P(A) + P(L|B) \cdot P(B) + P(L|C) \cdot P(C)$$

Solution

- P(A)=0.40 (40% chance of Venue A being available).
- P(B)=0.30 (30% chance of Venue B being available).
- P(C)=0.30 (30% chance of Venue C being available).
- P(L|A)=0.70 (70% chance of raising a large amount if Venue A is available).
- P(L|B)=0.50 (50% chance of raising a large amount if Venue B is available).
- P(L|C)=0.50 (50% chance of raising a large amount if Venue C is available).

$$P(L) = P(L|A) \cdot P(A) + P(L|B) \cdot P(B) + P(L|C) \cdot P(C)$$

Solution

- $P(A)=0.40$ (40% chance of Venue A being available).
- $P(B)=0.30$ (30% chance of Venue B being available).
- $P(C)=0.30$ (30% chance of Venue C being available).
- $P(L|A)=0.70$ (70% chance of raising a large amount if Venue A is available).
- $P(L|B)=0.50$ (50% chance of raising a large amount if Venue B is available).
- $P(L|C)=0.50$ (50% chance of raising a large amount if Venue C is available).

$$P(L) = P(L|A) \cdot P(A) + P(L|B) \cdot P(B) + P(L|C) \cdot P(C)$$

$$P(L)=(0.70 \cdot 0.40)+(0.50 \cdot 0.30)+(0.50 \cdot 0.30)=0.28+0.15+0.15=0.58$$

So, the overall probability of raising a large amount of money at the charity event is 58%.

Expected Value

Expected Value

Example: A lottery ticket costs 5 dollars

- The chances of winning are 1%
- Not winning makes you 0 dollars
- Winning makes you 100 dollars

How much is the lottery ticket worth?

$$E[X] = \sum x_i p(x_i)$$

x_i = The values that X takes

$p(x_i)$ = The probability that X takes the value x_i

This Example is from Professor Max's Slides

Expected Value

A lottery ticket costs 5 dollars

- The chances of winning are 1%
- Not winning makes you 0 dollars
- Winning makes you 100 dollars

How much is the lottery ticket worth?

$$E[X] = \sum x_i p(x_i)$$

x_i = The values that X takes

$p(x_i)$ = The probability that X takes the value x_i

Expected value of the lottery ticket:

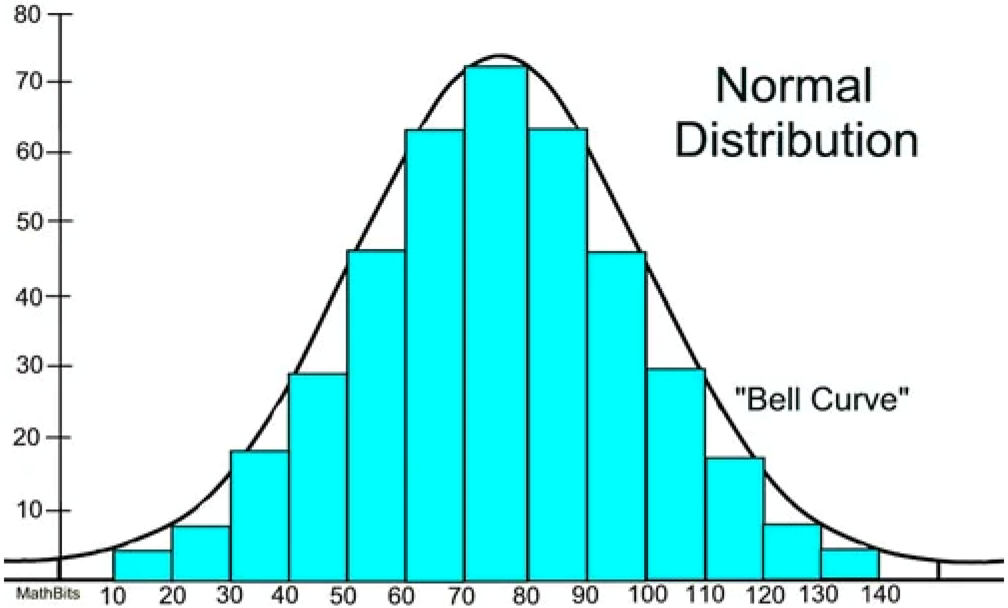
100 * .01 = $1

This Example is from Professor Max's Slides

Normal Distributions (Gaussian)

More Example: The Normal Distribution (Gaussian)

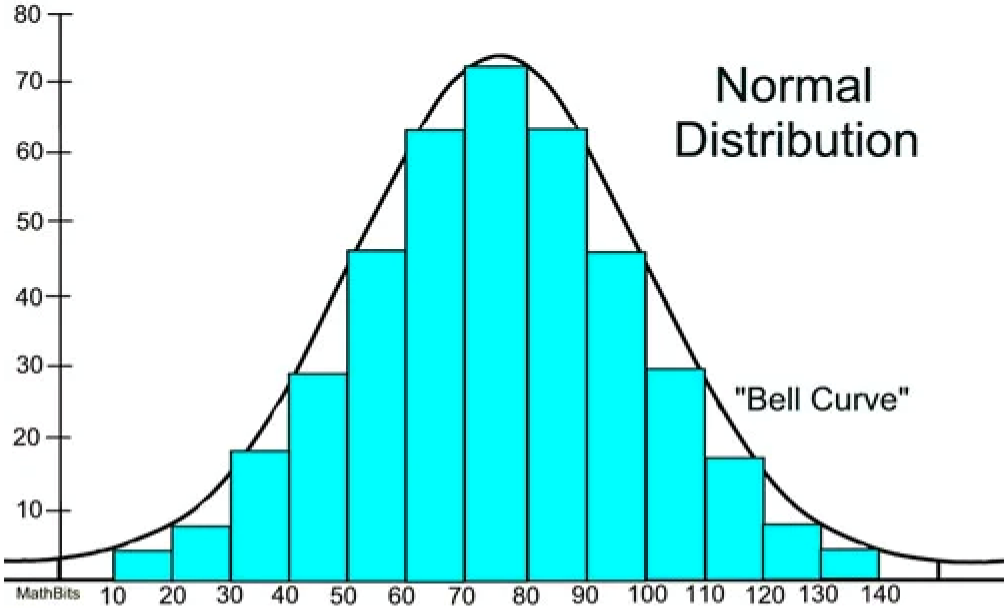
Example: the **scores of a quiz** follow a normal distribution. Many of the students scored between 60 and 80 as illustrated in the graph below. Of course, students with scores that fall outside this range are deviating from the center.



More Example: The Normal Distribution (Gaussian)

- The “bell-shaped” curve around the central region, indicating that most data points exist there.
- The normal distribution is represented as $N(\mu, \sigma^2)$ here,
    - μ represents the mean, and σ^2 represents the variance, one of which is mostly provided.

The expected value of a normal distribution is equal to its mean.

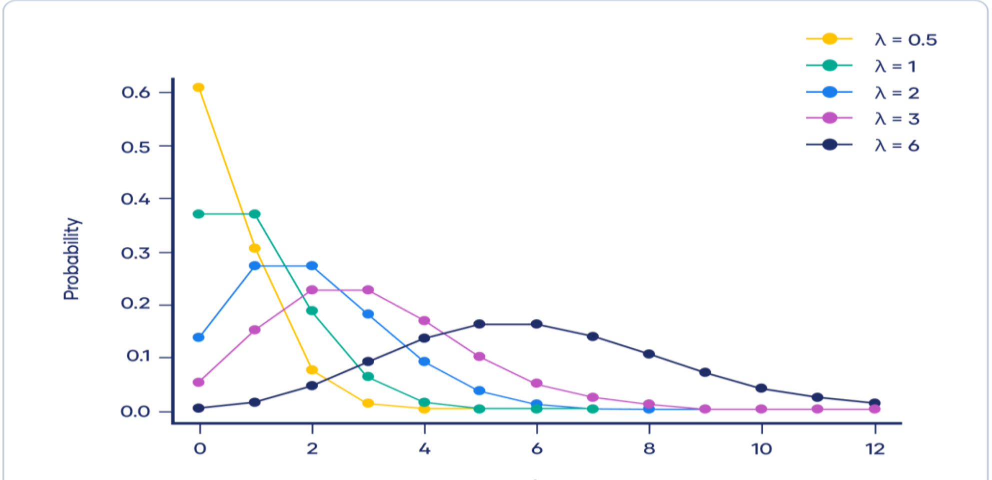


Poisson Distribution

Poisson Distribution

- A Poisson distribution is a **discrete probability distribution**. It gives the probability of an event happening a certain number of times (k) within a given interval of time or space.
- Gives the probability of a discrete (i.e., countable) outcome.
- For Poisson distributions, the discrete outcome is the number of times an event occurs, represented by λ (lambda), which is the mean number of events (only one parameter).

The graph below shows examples of Poisson distributions with different values of λ.



Try with different lamba and x values:

https://homepage.divms.uiowa.edu/~mbognar/applets/pois.html

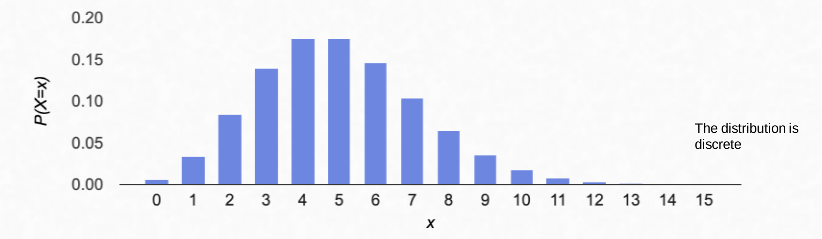
The Poisson distribution assumes a constant rate of events within the specified interval.

Let's say, During the office hour, the average rate (λ) of student arrivals per hour is 5.

- We wanted to calculate the probabilities of different numbers of students arriving in each hour.

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

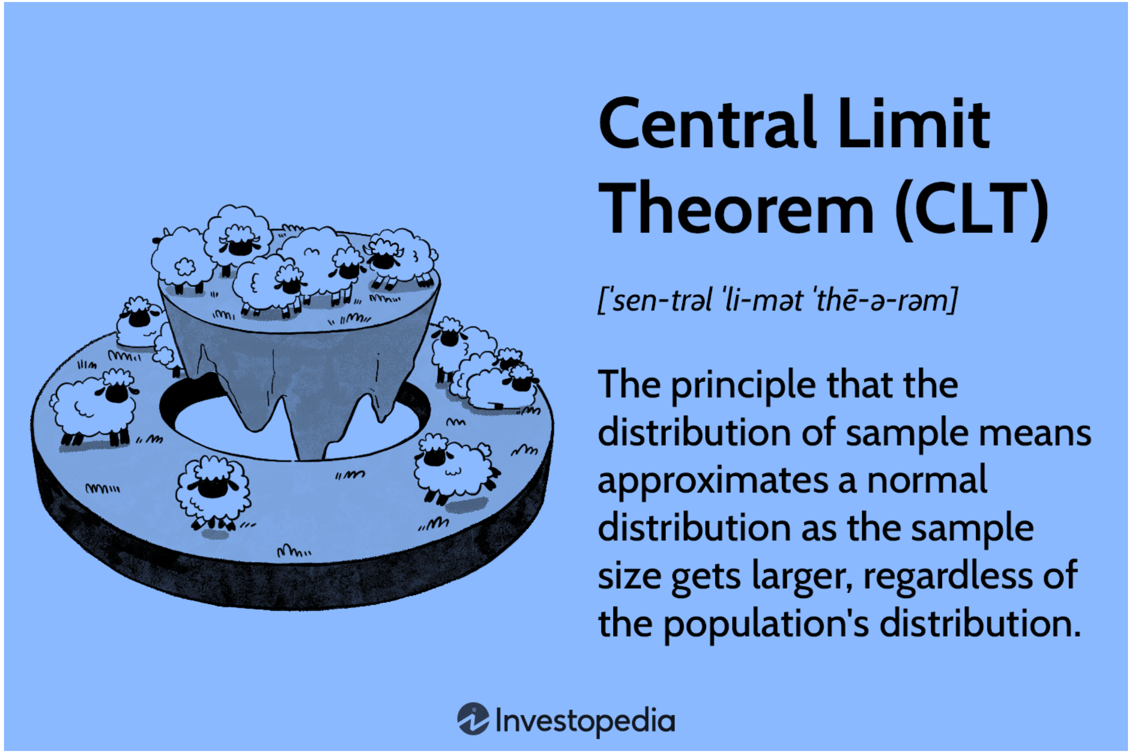
y-axis →the probability of observing each number of arrivals



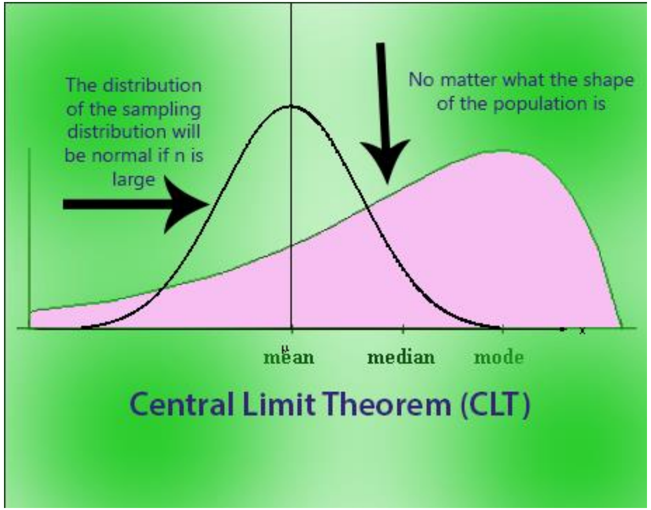
x-axis → the number of people arriving per hour

- **P(X=k)=** the probability of observing k events occurring in the given interval.
- **e** =base of the natural logarithm (approximately 2.71828)
- **λ** =the average rate of events occurring in the interval.
  - λ=5 (average number (mean) of students per hour)
- ***k*=** represents no of events = 0,1,2,3,... (number of students arriving)

CLT Theorem



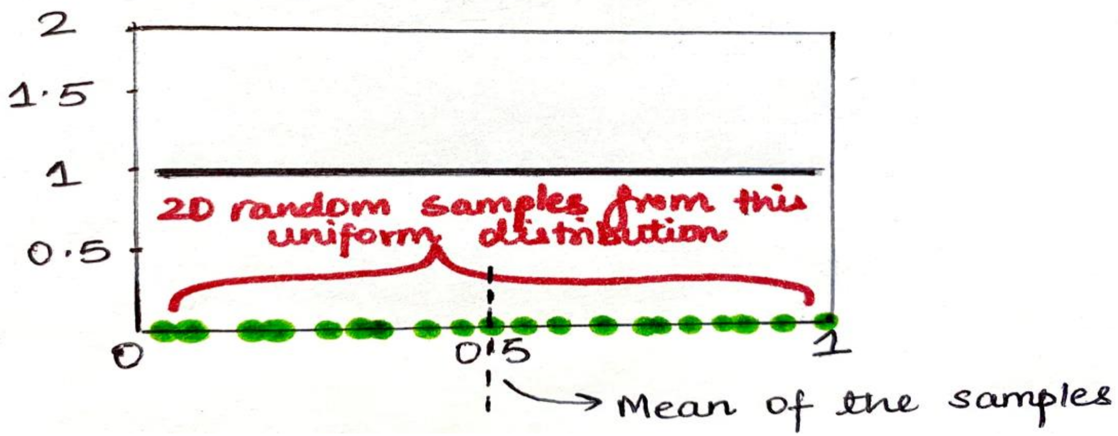
Central Limit Theorem (CLT)
['sen-trəl 'li-mət 'thē-ə-rəm]
The principle that the distribution of sample means approximates a normal distribution as the sample size gets larger, regardless of the population's distribution.
Investopedia



The distribution of the sampling distribution will be normal if n is large
No matter what the shape of the population is
mean
median
mode
Central Limit Theorem (CLT)

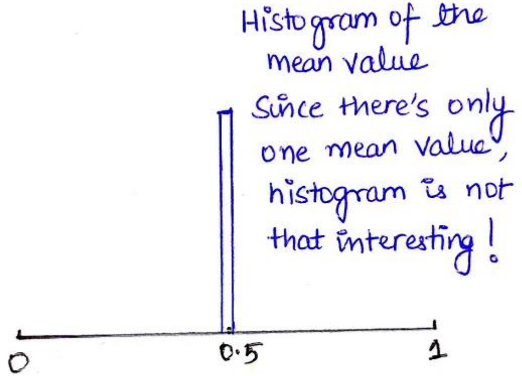
Key Point

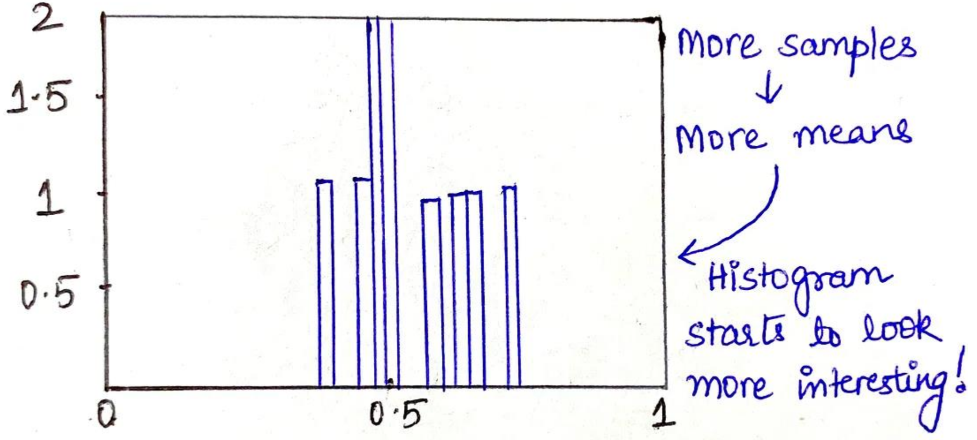
CLT: "When you repeatedly sample from a population (in your case, rolling a die) and calculate means, the distribution of those sample means becomes approximately normal, even if the original population distribution is not normal.**"**



Key Point

CLT: **“**When you repeatedly sample from a population (in your case, rolling a die) **and calculate means, the distribution of those sample means** becomes approximately normal, even if the original population distribution is not normal.**”**





Key Point

CLT: **“**When you repeatedly sample from a population (in your case, rolling a die) and calculate means, the distribution of those sample means **becomes approximately normal, even if the original population distribution is not normal.”**