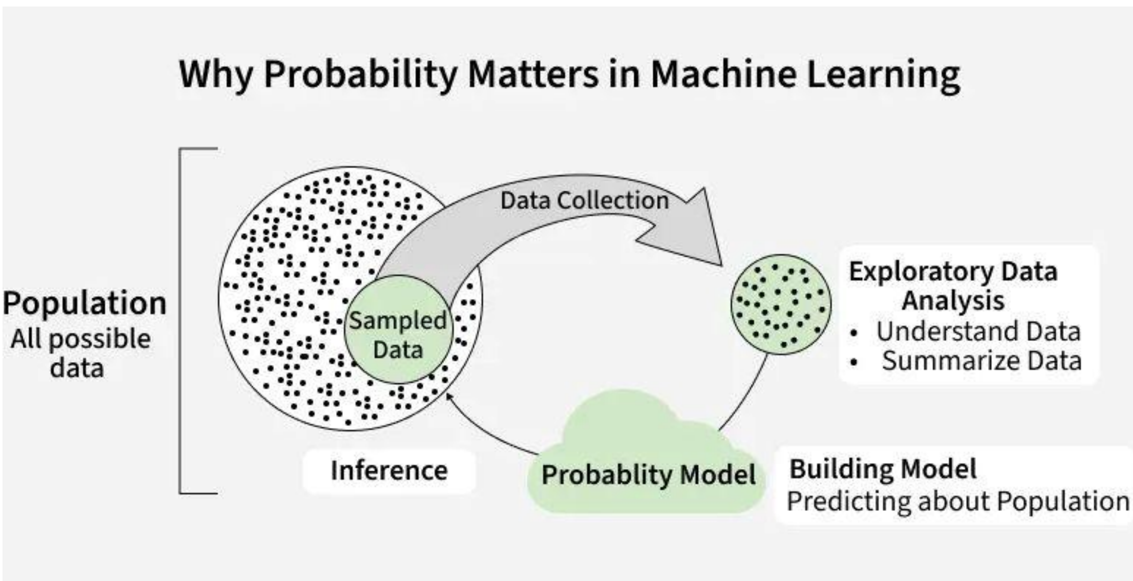
PROBABILITY

Why Probability is Important in Statistics and Machine Learning

Probability is the foundation of statistics. It provides the mathematical rules to measure uncertainty, randomness, and variation in real-world events. All statistical methods such as hypothesis testing, confidence intervals, and sampling techniques are built on probability concepts.

In Machine Learning, probability helps models to make predictions under uncertainty, understand data distributions, and estimate outcomes. Concepts like the Central Limit Theorem and probability distributions allow us to move from theoretical randomness to practical data analysis.

Without probability, statistics cannot exist. It is the language through which data is understood.



📘 1. Probability Theory

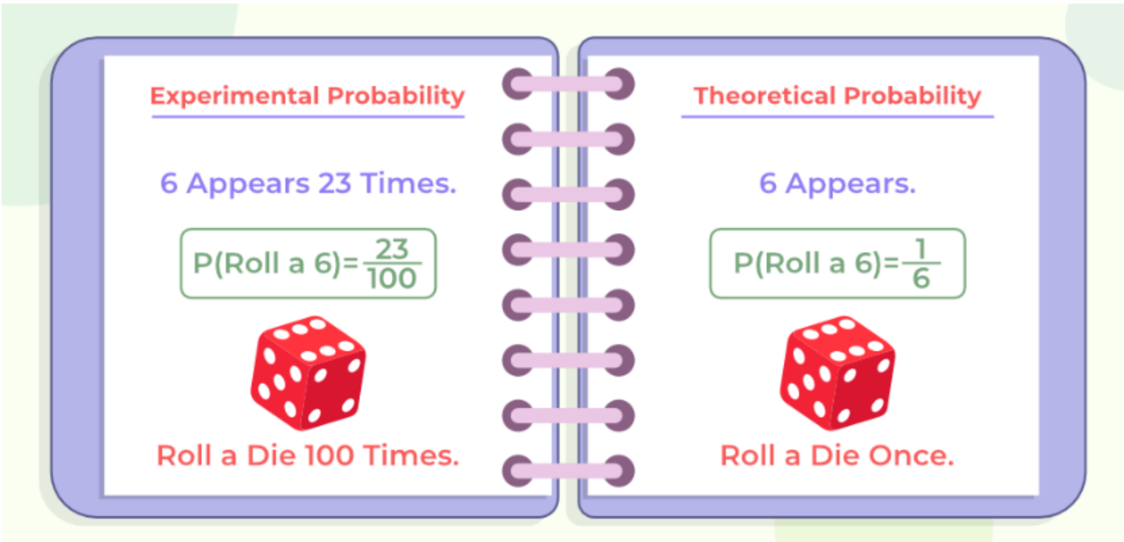
✅ Definition

Probability theory measures how likely events are to occur under uncertainty using ideas like **sample space, events, and random variables**.

Example (fair coin):

$$P(\text{Head}) = \frac{1}{2}, \; P(\text{Tail}) = \frac{1}{2}$$

🧭 Four Approaches to Probability



Approach	Idea	Formula / Example
Theoretical	Assume outcomes are equally likely	$P(A) = \frac{\text{favorable}}{\text{total}}$. Coin → $1/2$
Experimental (Empirical)	Based on trials/observations	$P(A) = \frac{\text{times A occurs}}{\text{trials}}$. 4 heads in 10 → $4/10$
Subjective	Personal belief/experience	Fan predicts 70% win chance
Axiomatic (Kolmogorov)	Formal rules on events	$P(H) + P(T) = 1$for a coin

---

🧩 Basic Terms (Definitions)

- **Random Experiment:** Repeatable process with uncertain outcome (**For example,** tossing a coin, rolling the dice, etc., are [random experiments]).
- **Sample Space (S):** All possible outcomes. **For example**, throwing dice results in six outcomes, which are 1, 2, 3, 4, 5 and 6. Thus, its sample space is (1, 2, 3, 4, 5, 6)

- **Event:** The outcome of any experiment is called an event.Any subset of the sample space.

Types of Events

- **Independent:** The events whose outcomes are not affected by the outcomes of other future and/or past events are called independent events. (**For example**, the output of tossing a coin in repetition is not affected by its previous outcome).
- **Dependent:** The events whose outcomes are affected by the outcome of other events are called dependent events. **For examp**le, picking oranges from a bag that contains 100 oranges without replacement.
- **Mutually Exclusive:** The events that can not occur simultaneously are called mutually exclusive events.
  For example, obtaining a head or a tail in tossing a coin, because both (head and a tail can not be obtained together.
- **Equally Likely:** The events that have an equal chance or probability of happening are known as equally likely events.
  For example, observing any face in rolling the dice has an equal probability of 1/6.

🎲 Random Variables

A variable that can assume the value of all possible outcomes of an experiment is called a random variable in Probability Theory. Random variables in probability theory are of two types, which are discussed below,

- **Discrete:**

  Variables that can take countable values, such as 0, 1, 2, ..., are called discrete random variables.
  Examples: The number of heads when flipping 3 coins, the number of cars arriving at a parking lot in an hour or the number of correct answers on a test.
- **Continuous:** Variables that can take an infinite number of values in a given range are called continuous random variables.
  Examples: The height of a person, the time it takes for a chemical reaction to occur or the temperature of a substance.

📐 Important Probability Formulas

$$\text{Theoretical: } P(A) = \frac{\text{favorable}}{\text{total}}$$
$$\text{Addition: } P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$\text{Complement: } P(A') = 1 - P(A)$$
$$\text{Independent: } P(A \cap B) = P(A)P(B)$$
$$\text{Conditional: } P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$
$$\text{Bayes: } P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

📊 Role in Statistics

- **Descriptive:** Understand distributions.
- **Inferential:** Hypothesis tests, confidence intervals.
- **Regression:** Error distributions.
- **Bayesian:** Update beliefs with data.

🌍 Applications

Weather forecasting, stock markets, risk analysis, product reliability, games & gambling, ML/AI predictions.

📝 Solved Examples (answers)

We can study the concept of probability with the help of the example discussed below.

Question 1: Let's take two random dice and roll them randomly. Now now the probability of getting a total of 10 is calculated.

Solution:

Total Possible events that can occur (sample space) {(1,1), (1,2),..., (1,6),..., (6,6)}. The total spaces are 36.
Now the required events, {(4,6), (5,5), (6,4)} are all which adds up to 10.

So the probability of getting a total of 10 is = 3/36 = 1/12

Question 2: A fair coin is tossed three times. What is the probability of getting exactly two heads?

Solution:

Total possible outcomes when tossing a coin three times = 2^3 = 8.
Possible outcomes: HHH, HHT, HTH, THH, HTT, THT, TTH, TTT.
Outcomes with exactly two heads: HHT, HTH, THH (3 outcomes).

Probability of getting exactly two heads:
P(exactly 2 heads)=Number of favorable outcomes/ Total outcomes.

P(exactly 2 heads)=3/ 8.

Question 3: A standard deck of cards contains 52 cards. What is the probability of drawing an Ace or a King from the deck?

Solution:

Total number of cards = 52.
Number of Aces = 4.
Number of Kings = 4.

Total number of favorable outcomes (Aces or Kings) = 4 + 4 = 8.

Probability of drawing an Ace or a King:

P(Ace or King)=Number of favorable outcomes/Total outcomes
P(Ace or King)=Number of Aces or Kings/Total number of cards.

P(Ace or King) = 8/52 = 2/13.

Question 4: Consider a jar with 7 red marbles, 3 green marbles and 4 blue marbles. What is the probability of randomly selecting a non-blue marble from the jar?

Solution:

Given,
Number of Red Marbles = 7, Number of Green Marbles = 3, Number of Blue Marbles = 4

So, Total number of possible outcomes in this case: 7 + 3 + 4 = 14
Now, Number of non-blue marbles are: 7 + 3 = 10
According to the formula of theoretical Probability we can find, P(Non-Blue) = 10/14 = 5/7

Hence, theoretical probability of selecting a non-blue marble is 5/7.

Question 5: Consider for players Naveena and Isha, playing a table tennis match. The probability of Naveena winning the match is 0.76. What is the probability of Isha winning the match?

Solution:

Let N and M represent the events that Naveena wins the match and Isha wins the match, respectively.

The probability of Naveena's winning P(N) = 0.76(given)
The probability of Isha's winning P(I) = ?
Winning of the match is an mutually exclusive event, since only one of them can win the match.
Therefore,

P(N) + P(I) =1
P(I) = 1 – P(N)
P(I) = 1 – 0.76 = 0.24

Thus, the Probability of Isha winning the match is 0.24.

Question 6: If someone takes out one card from a 52-card deck, what is the probability of the card being a heart? What is the probability of obtaining a 7-number card?

Solution:

Total number of cards in a deck = 52
Total Number of heart cards in a deck = 13

So, the probability of obtaining a heart,

P(heart) = 13/52 = 1/4
Total number of 7-number cards in a deck = 4

So, the probability of obtaining a 7-number card,
P(7-number) = 4/52 = 1/13

Question 7: Find the probability of rolling an even number when you roll a die containing the numbers 1-6. Express the probability as a fraction, decimal, ratio or percent.

Solution:

Out of 1 to 6 number, even numbers are 2, 4 and 6.
So, Number of favorable outcomes = 3.

Total number of outcomes = 6.

Probability of obtaining an even number P(Even)= 1/2 = 0.5 = 1 : 2 = 50%

📘 2.Probability: Joint vs. Marginal vs. Conditional

🎯 Joint, Marginal, and Conditional Probability

Probability is a fundamental concept in statistics that helps us understand the likelihood of different events occurring. Within probability theory, there are three key types of probabilities: joint, marginal, and conditional probabilities.

- Marginal Probability refers to the probability of a single event occurring, without considering any other events.
- Joint Probability is the probability of two or more events happening at the same time. It is the probability of the intersection of these events.
- Conditional Probability deals with the probability of an event occurring given that another event has already occurred.

In this article, we will discuss these probabilities in detail, including examples and differences between them as well.

Table of Content

✅ Probability of an Event

$$P(A) = \frac{\text{favorable outcomes}}{\text{total outcomes}}, 0 \leq P(A) \leq 1$$

- **Sample Space (S):** All possible outcomes. Die → $\{1, 2, 3, 4, 5, 6\}$
- **Event (A):** Subset of S. Even number → $\{2, 4, 6\}$

🔗 Joint Probability — "Together"

Definition: Probability that **two (or more)** events occur at the same time.

$$P(A \cap B)$$

- **Independent:** $P(A \cap B) = P(A)\,P(B)$
- **Dependent:** $P(A \cap B) = P(A)\,P(B \mid A)$

Note: If A and B are dependent, the joint probability is calculated using conditional probability

Examples of Joint Probability

Rolling Two Dice

- Let A be the event that the first die shows a 3.
- Let B be the event that the second die shows a 5.

The joint probability P(A∩B) is the probability that the first die shows a 3 and the second die shows a 5. Since the outcomes are independent,

P(A∩B) = P(A) · P(B).

Given: P(A) = 1/6 and P(B) = 1/6, so

⇒ P(A∩B) = 1/6 × 1/6 = 1/36.

📦 Marginal Probability — "Single, ignoring others"

Definition: Probability of **one** event, regardless of others.
Computed by **summing/integrating** joint probabilities over the other variable.

$$P(A) = \sum_{B} P(A, B)\ \text{(discrete)}, P(A) = \int P(A, B)\, dB\, \text{(continuous)}$$

Examples of Marginal Probability

Consider a table showing the joint probability distribution of two discrete random variables X and Y:

X/Y	Y = 1	Y = 2
X = 1	0.1	0.2
X = 2	0.3	0.4

To find the marginal probability of X = 1:

P(X = 1) = P(X = 1, Y = 1) + P(X = 1, Y = 2) = 0.1 + 0.2 = 0.3

🎯 Conditional Probability — "Given that"

Definition: Probability of A **given** B already happened.

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

Where:

- P(A∩B) is the joint probability of both events A and B occurring.
- P(B) is the probability of event B occurring.

Examples of Conditional Probability

Suppose we have a deck of 52 cards, and we want to find the probability of drawing an Ace given that we have drawn a red card.

- Let A be the event of drawing an Ace.
- Let B be the event of drawing a red card.

There are 2 red Aces in a deck (Ace of hearts and Ace of diamonds) and 26 red cards in total.

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{2}{52}}{\frac{26}{52}} = \frac{2}{26} = \frac{1}{13}$$

🧠 Relationship Between Them

- Joint connects two events.
- Marginal comes **from** joint (by summing).
- Conditional uses **joint + marginal**.

📊 Differences at a Glance

The key differences between joint, marginal and conditional probability are listed in the following table:

- For Independent Events
  - $P(A \cap B) = P(A) \times P(B)$
- For Dependent Events
  - $P(A \cap B) = P(A) \times P(B|A)$

Aspect	**Joint Probability**	**Marginal Probability**	**Conditional Probability**
Definition	The probability of two or more events occurring together.	The probability of a single event irrespective of the occurrence of other events.	The probability of an event given that another event has occurred.
Notation	$P(A \cap B)$ or $P(A, B)$	$P(A)$ or $P(B)$	$P(A\|B)$ or $P(B\|A)$
Formula	$P(A) = \sum B P(A \cap B)$	$P(A\|B) = P(A \cap B) / P(B)$	
Example	Probability of rolling a 2 and flipping heads: $P(2 \cap \text{Heads})$	Probability of rolling a 2: $P(2)$	Probability of rolling a 2 given that the coin flip is heads: $P(2 \mid \text{Heads})$

Aspect	Joint Probability	Marginal Probability	Conditional Probability
Calculation Context	Calculated from a joint probability distribution.	Calculated by summing the joint probabilities over all outcomes of the other variable.	Calculated using the joint probability and the marginal probability of the given condition.
Dependencies	Involves multiple events happening simultaneously.	Does not depend on other events.	Depends on the occurrence of another event.
Use Case	Used to find the likelihood of combined events in probabilistic models.	Used to find the likelihood of a single event in the presence of multiple events.	Used to update the probability of an event based on new information.

✅ **Quick Memory Trick**

- **Joint = AND**
- **Marginal = ALONE**
- **Conditional = GIVEN**

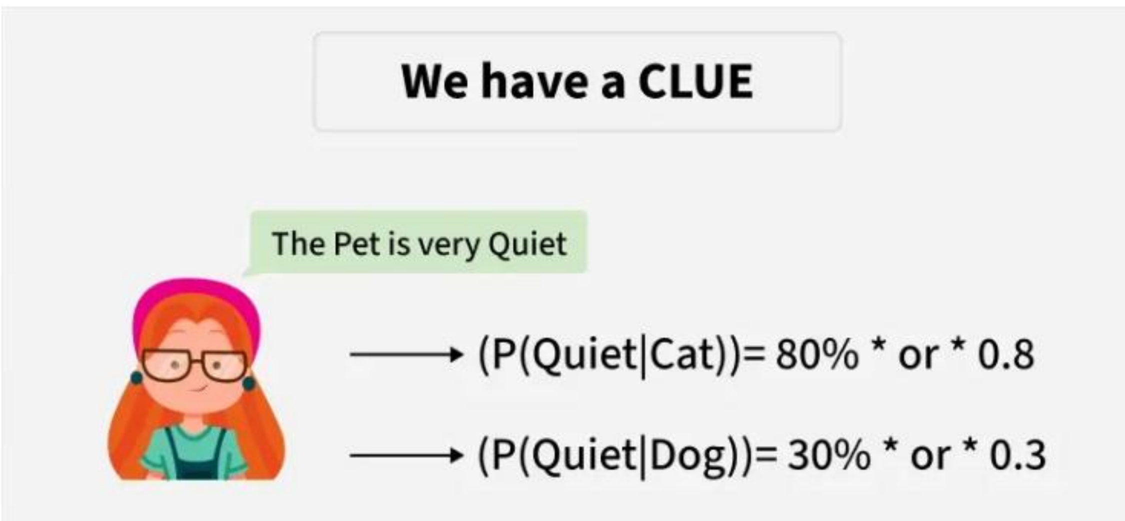
📘 3.Bayes' Theorem

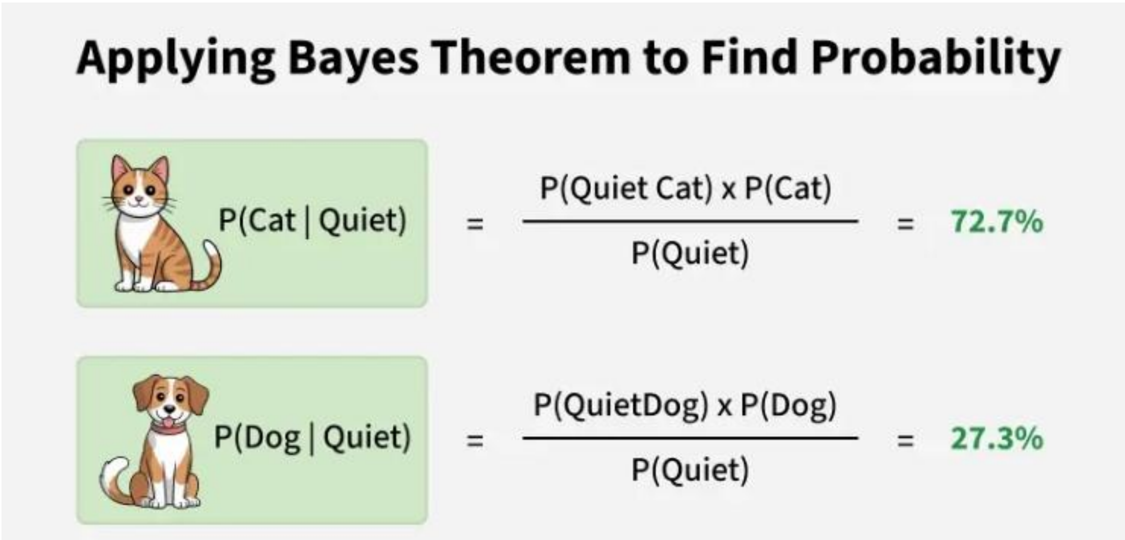
✅ Definition

Bayes' Theorem is a mathematical formula used to determine the conditional probability of an event based on prior knowledge and new evidence.

It adjusts probabilities when new information comes in and helps make better decisions in uncertain situations.


Guessing the Pet
Is it a cat or a dog in the box?
P(Dog) = 0.5
?
P(Cat) = 0.5
Initial Belief: 50/50 chances


We have a CLUE
The Pet is very Quiet
(P(Quiet|Cat))= 80% * or * 0.8
(P(Quiet|Dog))= 30% * or * 0.3



$$P(A \mid B) = \frac{P(B \mid A)\,P(A)}{P(B)}, P(B) \neq 0$$

- $P(A)$: Prior (before evidence)
- $P(B \mid A)$: Likelihood (evidence if A is true)
- $P(A \mid B)$: Posterior (updated belief)
- $P(B)$: Evidence (normalizing term)

  Bayes' Theorem helps us update probabilities based on prior knowledge and new evidence. In this case, knowing that the pet is quiet (new information), we can use Bayes' Theorem to calculate the updated probability of the pet being a cat or a dog, based on how likely each animal is to be quiet.

Bayes Theorem and Conditional Probability

Bayes' theorem (also known as the Bayes Rule or Bayes Law) is used to determine the conditional probability of event A when event B has already occurred.

The general statement of Bayes’ theorem is “The conditional probability of an event A, given the occurrence of another event B, is equal to the product of the probability of B, given A, and the probability of A divided by the probability of event B.” i.e.

***For example**, if we want to find the probability that a white marble drawn at random came from the first bag, given that a white marble has already been drawn, and there are three bags each containing some white and black marbles, then we can use Bayes' Theorem.*

Also Check

[Bayes' Theorem for Conditional Probability]

Bayes Theorem Formula

For any two events A and B, **Bayes's** formula for the Bayes theorem is given by:

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

Formula for the Bayes theorem

Where,

- **P(A)** and **P(B)** are the probabilities of events A and B; also, P(B) is never equal to zero.
- **P(A|B)** is the probability of event A when event B happens,
- **P(B|A)** is the probability of event B when A happens.

Bayes Theorem Statement

Bayes' Theorem for n sets of events is defined as,

Let $E_1, E_2, \ldots, E_n$ be a set of events associated with the sample space S, in which all the events $E_1, E_2, \ldots, E_n$ have a non-zero probability of occurrence. All the events $E_1, E_2, \ldots,$ E form a partition of S. Let A be an event in space S for which we have to find the probability, then according to Bayes theorem,

$$P(E_i \mid A) = \frac{P(E_i) \cdot P(A \mid E_i)}{\sum_{k=1}^{n} P(E_k) \cdot P(A \mid E_k)}$$

for k = 1, 2, 3, …., n

Bayes Theorem Derivation

The proof of Bayes' Theorem is given as, according to the conditional probability formula,

$P(E_i \mid A) = \frac{P(E_i \cap A)}{P(A)}$**.....(i)**

Then, by using the multiplication rule of probability, we get

$P(E_i \cap A) = P(E_i) \cdot P(A \mid E_i)$**......(ii)**

Now, by the total probability theorem,

$P(A) = \sum_{k=1}^{n} P(E_k) \cdot P(A \mid E_k)$**.....(iii)**

Substituting the value of $P(E_i \cap A)$ and P(A) from eq (ii) and eq(iii) in eq(i) we get,

$$P(E_i \mid A) = \frac{P(E_i) \cdot P(A \mid E_i)}{\sum_{k=1}^{n} P(E_k) \cdot P(A \mid E_k)}$$

Bayes' theorem is also known as the formula for the probability of "causes". As we know, the E_i's are a partition of the sample space S, and at any given time, only one of the events E_i occurs. Thus, we conclude that the Bayes theorem formula gives the probability of a particular E_i, given that event A has occurred.

🪜 Terms Related to Bayes' Theorem

After learning about Bayes theorem in detail, let us understand some important terms related to the concepts we covered in the formula and derivation.

Hypotheses

- Hypotheses refer to possible events or outcomes in the sample space; they are denoted as $E_1, E_2, \ldots, E_n$.
- Each hypothesis represents a distinct scenario that could explain an observed event.

Priori Probability

- Priori Probability $P(E_i)$ is the initial probability of an event occurring before any new data is taken into account.

- It reflects existing knowledge or assumptions about the event.
- **Example:** The probability of a person having a disease before taking a test.

Posterior Probability

- Posterior probability ($P(E_i|A)$) is the updated probability of an event after considering new information.
- It is derived using the Bayes Theorem.
- **Example:** The probability of having a disease given a positive test result.

Conditional Probability

- The probability of an event A based on the occurrence of another event B is termed conditional Probability.
- It is denoted as **$P(A|B)$** and represents the probability of A when event B has already happened.

Joint Probability

- When the probability of two or more events occurring together and at the same time is measured, it is marked as Joint Probability.
- For two events A and B, it is denoted by joint probability is denoted as $P(A \cap B)$.

Random Variables

- Real-valued variables whose possible values are determined by random experiments are called random variables.
- The probability of finding such variables is the experimental probability.

🏷️ Bayes Theorem Applications

Bayesian inference is very important and has found application in various activities, including medicine, science, philosophy, engineering, sports, law, etc., and Bayesian inference is directly derived from Bayes theorem.

Some of the Key Applications are:

- **AI & Machine Learning** → Used in Naïve Bayes classifiers to predict outcomes.

- **Medical Testing** → Finding the real probability of having a disease after a positive test.
- **Spam Filters** → Checking if an email is spam based on keywords.
- **Weather Prediction** → Updating the chance of rain based on new data.

Difference Between Conditional Probability and Bayes Theorem

The difference between Conditional Probability and Bayes's theorem can be understood with the help of the table given below.

Bayes Theorem	Conditional Probability
Bayes's Theorem is derived using the definition of conditional probability. It is used to find the reverse probability.	Conditional Probability is the probability of event A when event B has already occurred.
Formula: P(A|B) = [P(B|A)P(A)] / P(B)	**Formula:** P(A|B) = P(A∩B) / P(B)
Purpose: To update the probability of an event based on new evidence.	**Purpose**: To find the probability of one event based on the occurrence of another.
Focus: Uses prior knowledge and evidence to compute a revised probability.	**Focus**: Direct relationship between two events.

Theorem of Total Probability

Let E1, E2,…., En be mutually exclusive and exhaustive events of a sample space S, and let E be any event that occurs with some E_i. Then, prove that :

P(E) = n∑i=1P(E/Ei) . P(Ei)

Proof:

Let S be the sample space.
Since the events E_1, E_2,...,E_n are mutually exclusive and exhaustive, we have:

$S = E_1 \cup E_2 \cup E_3 \cup \ldots \cup En$ and $E_i \cap E_j = \emptyset$ for i - j.
Now, consider the event E: $E = E \cap S$
Substituting S with the union of Ei's:
$\Rightarrow E = E \cap (E_1 \cup E_2 \cup E_3 \cup \ldots \cup E_n)$
Using distributive law:
$\Rightarrow E = (E \cap E_1) \cup (E \cap E_2) \cup \ldots \cup (E \cap E_n)$
Since the events Ei are mutually exclusive, the intersections E∩Ei are also ***mutually exclusive****. Therefore:*

$P(E) = P\{(E \cap E_1) \cup (E \cap E_2) \cup \ldots \cup (E \cap E_n)\}$
$\Rightarrow P(E) = P(E \cap E_1) + P(E \cap E_2) + \ldots + P(E \cap E_n)$
{Therefore, $(E \cap E_1)$, $(E \cap E_2)$, . . . ,$(E \cap E_n)$} are pairwise disjoint}

$\Rightarrow P(E) = P(E/E_1) . P(E_1) + P(E/E_2) . P(E_2) + \ldots + P(E/E_n) . P(E_n)$ [by multiplication theorem]
$\Rightarrow P(E) = {}^{n}\sum_{i=1} P(E/E_i) . P(E_i)$

---

olved Examples of Bayes' Theorem

Example 1: A person has undertaken a job. The probability of completing the job on time if it rains is 0.44, and the probability of completing the job on time if it does not rain is 0.95. If the probability that it will rain is 0.45, then determine the probability that the job will be completed on time.

Let:

- *R: event that it rains*
- *R^c: event that it does not rain*
- *C: event that the job is completed on time*

We are given:

$$P(R) = 0.45, P(Rc) = 1 - 0.45 = 0.55$$

$$P(C \mid R) = 0.44, P(C \mid Rc) = 0.95$$

By the law of total probability:

$$P(C) = P(R)P(C \mid R) + P(Rc)P(C \mid Rc)$$

Substitute values:

$$P(C) = (0.45)(0.44) + (0.55)(0.95)$$

$$P(C) = 0.198 + 0.5225 = 0.7205$$

Example 2: There are three urns containing 3 white and 2 black balls, 2 white and 3 black balls, and 1 black and 4 white balls, respectively. There is an equal probability of each urn being chosen. One ball is equal probability chosen at random. What is the probability that a white ball will be drawn?

Solution:

Let E_1, E_2, and E_3 be the events of choosing the first, second, and third urn respectively. Then,
$P(E_1) = P(E_2) = P(E_3) = 1/3$

Let E be the event that a white ball is drawn. Then,
$P(E/E_1) = 3/5$, $P(E/E_2) = 2/5$, $P(E/E_3) = 4/5$

By theorem of total probability, we have
$P(E) = P(E/E_1) . P(E_1) + P(E/E_2) . P(E_2) + P(E/E_3) . P(E_3)$
$\Rightarrow P(E) = (3/5 \times 1/3) + (2/5 \times 1/3) + (4/5 \times 1/3)$
$\Rightarrow P(E) = 9/15 = 3/5$

Example 3: A card from a pack of 52 cards is lost. From the remaining cards of the pack, two cards are drawn and are found to be both hearts. Find the probability of the lost card being a heart.

Solution:

Let E_1, E_2, E_3, and E_4 be the events of losing a card of hearts, clubs, spades, and diamonds respectively.
Then $P(E_1) = P(E_2) = P(E_3) = P(E_4) = 13/52 = 1/4$.

Let E be the event of drawing 2 hearts from the remaining 51 cards. Then,

$P(E|E_1)$ = probability of drawing 2 hearts, given that a card of hearts is missing
$\Rightarrow P(E|E_1) = {}^{12}C_2 / {}^{51}C_2 = (12 \times 11)/2! \times 2!/(51 \times 50) = 22/425$

$P(E|E_2)$ = probability of drawing 2 clubs ,given that a card of clubs is missing
$\Rightarrow P(E|E_2) = {}^{13}C_2 / {}^{51}C_2 = (13 \times 12)/2! \times 2!/(51 \times 50) = 26/425$

$P(E|E_3)$ = probability of drawing 2 spades ,given that a card of hearts is missing
$\Rightarrow P(E|E_3) = {}^{13}C_2 / {}^{51}C_2 = 26/425$

$P(E|E_4)$ = probability of drawing 2 diamonds ,given that a card of diamonds is missing
⇒ $P(E|E_4) = {}^{13}C_2 / {}^{51}C_2 = 26/425$

Therefore,

$P(E_1|E)$ = probability of the lost card is being a heart, given the 2 hearts are drawn from the remaining 51 cards

⇒ $P(E_1|E) = P(E_1) . P(E|E_1)/P(E_1) . P(E|E_1) + P(E_2) . P(E|E_2) + P(E_3) . P(E|E_3) + P(E_4) . P(E|E_4)$
⇒ $P(E_1|E) = (1/4 \times 22/425) / \{(1/4 \times 22/425) + (1/4 \times 26/425) + (1/4 \times 26/425) + (1/4 \times 26/425)\}$
⇒ $P(E_1|E) = 22/100 = 0.22$

Hence, The required probability is 0.22.

Example 4: Suppose 15 men out of 300 men and 25 women out of 1000 are good orators. An orator is chosen at random. Find the probability that a male person is selected.

Solution:

Given,

- *Total Men = 300*
- *Total Women = 1000*
- *Good Orators among Men = 15*
- *Good Orators among Women = 25*

Total number of good orators = 15 (from men) + 25 (from women) = 40
Probability of selecting a male orator:
P(Male Orator) = Numbers of male orators / total no of orators = 15/40 = 3/8

Example 5: A man is known to speak the lies 1 out of 4 times. He throws a die and reports that it is a six. Find the probability that it **is** actually a six.

Solution:

In a throw of a die, let
E_1 = event of getting a six,
E_2 = event of not getting a six and
E = event that the man reports that it is a six.

Then, $P(E_1) = 1/6$, and $P(E_2) = (1 - 1/6) = 5/6$

$P(E|E_1)$ = probability that the man reports that six occurs when six has actually occurred
⇒ $P(E|E_1)$ = probability that the man speaks the truth
⇒ $P(E|E_1) = 3/4$

$P(E|E_2)$ = probability that the man reports that six occurs when six has not actually occurred

⇒ $P(E|E_2)$ = probability that the man does not speak the truth
⇒ $P(E|E_2) = (1 - 3/4) = 1/4$

Probability of getting a six ,given that the man reports it to be six

$P(E_1|E) = P(E|E_1) \times P(E_1)/P(E|E_1) \times P(E_1) + P(E|E_2) \times P(E_2)$ [by Bayes theorem]
⇒ $P(E_1|E) = (3/4 \times 1/6)/\{(3/4 \times 1/6) + (1/4 \times 5/6)\}$
⇒ $P(E_1|E) = (1/8 \times 3) = 3/8$

Hence the probability required is 3/8.

Practice Problem Based on Bayes' Theorem

Question 1: A medical test for a disease is 95% accurate in detecting the disease (True Positive Rate). The probability of a person having the disease is 0.01 (1%). If a person tests positive for the disease, what is the probability that they actually have the disease? (Assume that the false positive rate is 5%).

Question 2: A bag contains 4 red balls and 6 blue balls. Two balls are drawn at random, and one of them is red. What is the probability that the second ball drawn is also red, given that the first ball was red?

Question 3: In a factory, 80% of the products are produced by Machine A and 20% by Machine B. Machine A produces 2% defective items, while Machine B produces 5% defective items. If a product is found to be defective, what is the probability that it was produced by Machine A?

Question 4: A survey shows that 70% of people like ice cream, and 40% of people like both ice cream and chocolate. What is the probability that a person likes chocolate, given that they like ice cream?

Answer:-

1. ***16.1%.***
2. ***33.33%.***

3. *61.5%.*
4. *57.1%.*

📘 4.Discrete Probability Distribution

✅ Definition

A discrete probability distribution describes the likelihood of each possible outcome for a discrete random variable. A discrete random variable is a variable that can take on a countable number of distinct values, typically whole numbers. In simpler terms, it's a way to assign a probability to each value that the variable can have, showing how likely each outcome is.

The common examples of discrete probability distributions include Bernoulli, Binomial, Poisson, and Geometric distributions.

Conditions

1. *The probability of a discrete random variable lies between 0 and 1:* ***0 ≤ P (X = x) ≤ 1***
2. *Sum of Probabilities is always equal to 1:* ***∑ P (X =x) = 1***

🎲 Example (Two Coins)

Sample space: $\{HH, HT, TH, TT\}$
Let X =number of tails.

x	**Outcomes**	$P(X = x)$
0	HH	$1/4$
1	HT, TH	$2/4 = 1/2$
2	TT	$1/4$

🧮 Discrete Probability Distribution Formulas

The different formulas for the discrete probability distribution, like the probability mass function, the cumulative distribution function, and the mean and variance, are given below.

PMF of Discrete Probability Distribution

PMF of a discrete random variable X is the value completely equal to x. The PMF i.e., probability mass function of a discrete probability distribution is given by:

f(x) = P (X = x)

Example:

A discrete random variable X as the number of heads obtained when tossing two fair coins.

Possible Outcomes:

- *HH → 2 heads*
- *HT, TH → 1 head*
- *TT → 0 heads*

Then PMFs are given as:

- $P(X = 0) = \frac{1}{4}$→ *Only TT*
- $P(X = 1) = \frac{2}{4}$→ *HT or TH*
- $P(X = 2) = \frac{1}{4}$→ *Only HH*

Total Probability : $\frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1$

Properties of PMF:

- The sum of all probabilities must always equal 1. i.e. $\sum P(X = x) = 1$
- P(X = x) ≥ 0

CDF of Discrete Probability Distribution

CDF of a discrete random variable X is less than or equal to value x. The CDF i.e., cumulative distribution function of discrete probability distribution is given by:

f(x) = P (X ≤ x)

Example:

A random variable X as the outcome of rolling a fair 4-sided die. To find $P(1 < X \leq 3)$*:*

X	PMF	CDF
1	0.2	0.2
2	0.3	0.2+0.3 = 0.5
3	0.2	0.5+0.2 = 0.7
4	0.3	0.7+0.3 = 1

$$P(1 < X \leq 3) = F(3) - F(1) = 0.7 - 0.2 = 0.5$$

Discrete Probability Distribution Mean

Mean of discrete probability distribution is the average of all the values that a discrete variable can obtain. It is also called as the expected value of the discrete probability distribution. The mean of discrete probability distribution is given by:

E[X] = ∑x P(X =x)

Discrete Probability Distribution Variance

Variance of discrete probability distribution is defined as the product of squared difference of distribution and mean with PMF. The variance of the discrete probability distribution is given by:

Var[X] = ∑(x - μ)2 P(X = x)

Example:

A random variable X as the outcome of rolling a fair 4-sided die. To find $P(1 < X \leq 3)$*:*

X	PMF	CDF
1	0.2	0.2
2	0.3	0.2+0.3 = 0.5
3	0.2	0.5+0.2 = 0.7

X	PMF	CDF
4	0.3	0.7+0.3 = 1

$$P(1 < X \leq 3) = F(3) - F(1) = 0.7 - 0.2 = 0.5$$

🧭 How to Find Discrete Probability Function

Steps to find the discrete probability function are given below:

- **Step 1:** First determine the sample space of the given event.
- **Step 2:** Define random variable X as the event for which the probability has to be found.
- **Step 3:** Consider the possible values of x and find the probabilities for each value.
- **Step 4:** Write all the values of x and their respective probabilities in tabular form to get the discrete probability distribution.

🧩 Important Discrete Distributions (PMF)

Types of Discrete Probability Distribution

The different types of discrete probability distribution are listed below.

- **Bernoulli Distribution**
- **Binomial Distribution**
- **Poisson Distribution**
- **Geometric Distribution**

Bernoulli Distribution

A discrete probability distribution with the probability of success p if the value of random variable is 1 and the probability of failure 1-p if the value of random variable is zero is called the Bernoulli distribution. The probability mass function of the Bernoulli distribution is given by:

$$P(X = x) = \begin{cases} \mathbf{p, x = 1} \\ \mathbf{1 - p, x = 0} \end{cases}$$

Binomial Distribution

A discrete probability distribution that includes the number of trials n, probability of success and probability of failure is called as Binomial distribution. The probability mass function of the Binomial distribution is given by:

P (X = x) = nCx px (1-p) n-x

Poisson Distribution

A discrete probability distribution that gives the number of events occurred at a specific time period with the help of its mean is called as the Poisson distribution. The probability mass function of Poisson distribution is given by:

P (X = x) = [λx × e-λ] / x!

Geometric Distribution

A discrete probability distribution that includes the successive failure probability until the success probability is encountered is called as Geometric distribution. The probability mass function of the geometric distribution is given by:

P (X = x) = (1 - p) xp

💻 Applications in CS / ML

- Randomized algorithms (e.g., random pivots in QuickSort)
- Network packet loss (Bernoulli/Binomial)
- Naïve Bayes (word counts → Multinomial/Bernoulli)
- Reinforcement learning (time-to-reward patterns)
- Security modeling (attack attempts → Binomial/Poisson)

✏️ Solved Examples on Discrete Probability Distribution

Solved Examples on Discrete Probability Distribution

Example 1: Construct the discrete probability table when a coin is tossed two times and X be random variable representing the number of one head.

Solution:

Sample space of two coin tossed = 4 i.e., {HH, HT, TH, TT}

X: Number of one head

The below table represents the discrete probability.

X	{HT}	{TH}
P (X = x)	1 / 4	1/4

Example 2: Find the value of p from the given discrete probability table.

X	**3**	**4**	**5**	**6**
P (X = x)	0.1	p	0.2	0.4

Solution:

To find the value of p we will use the discrete probability condition.

$\sum P(X = x) = 1$

$0.1 + p + 0.2 + 0.4 = 1$

$0.7 + p = 1$

$p = 1 - 0.7$

$p = 0.3$

Example 3: Find the mean of discrete probability distribution using below table.

X	**2**	**3**	**4**	**5**
P (X = x)	0.16	0.45	0.32	0.07

Solution:

To find the mean of discrete probability distribution we use formula:

$E[X] = \sum [\, x\, P(X = x)]$

$E[X] = 2 \times 0.16 + 3 \times 0.45 + 4 \times 0.32 + 5 \times 0.07$

$E[X] = 0.32 + 1.35 + 1.28 + 0.35$

$E[X] = 3.3$

Example 4: If there are 15 pens in which 3 pens are defective and the probability of pen is defective 0.5 then, find the discrete probability of pen to be defective.

Solution:

To find the required probability we use Binomial Distribution

P (X = x) = nCx px (1-p) n-x

$P(X = 3) = {}^{15}C_3\, p^3\, (1-p)^{15-3}$

$P(X = 3) = {}^{15}C_3\, (0.5)^3\, (1 - 0.5)^{12}$

$P(X = 3) = 455 \times (0.5)^3 \times (0.5)^{12}$

$P(X = 3) = 455 \times (0.5)^{15}$

$P(X = 3) = 0.014$

Practice Questions on Discrete Probability Distribution

Q1. Construct the discrete probability table when a dice is rolled, and X be random variable representing the numbers greater than equal to 3.

Q2. Find the value of a from the given discrete probability table.

X	**0**	**1**	**2**	**4**
P (X = x)	0.6	a	0.1	0.3

Q3. Find the expected value of discrete probability distribution using below table.

X	**4**	**5**	**6**	**7**
P (X = x)	0.21	0.35	0.42	0.02

Q4. Determine the probability if the number of trials is 100, number of successes is 94 and the probability of failure is 0.4.

🧠 **Memory Tip**

PMF → exact value, CDF → up to value, Mean → average, Variance → spread

📘 5.Continuous Probability Distributions

In machine learning, we often face uncertainty in our data. Continuous probability distributions help us understand this uncertainty by showing how likely different values are to occur. Whether predicting prices or classifying images, these distributions let us make smarter, more reliable predictions by accounting for the randomness in the real world.

Continuous Probability Distributions

A probability distribution is a mathematical function that describes the likelihood of different outcomes for a random variable. Continuous probability distributions (CPDs) are probability distributions that apply to continuous random variables. It describes events that can take on any value within a specific range, like the height of a person or the amount of time it takes to complete a task.

✅ A **continuous probability distribution** describes a **continuous random variable** that can take **any real value in an interval** (e.g., height, time, price).

In continuous probability distributions, two key functions describe the likelihood of a variable taking on specific values:

🔑 **Two Core Functions**

1) Probability Density Function (PDF)

The probability density function gives the probability density at a specific point or interval for a continuous random variable. It indicates how likely the variable is to fall within a small interval around a particular value.

- The height of the PDF curve at any point represents the probability density at that value.
- Higher density implies a higher probability of the variable taking on values around that point.

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$

- Properties: $f(x) \geq 0$, and $\int_{-\infty}^{\infty} f(x)\, dx = 1$

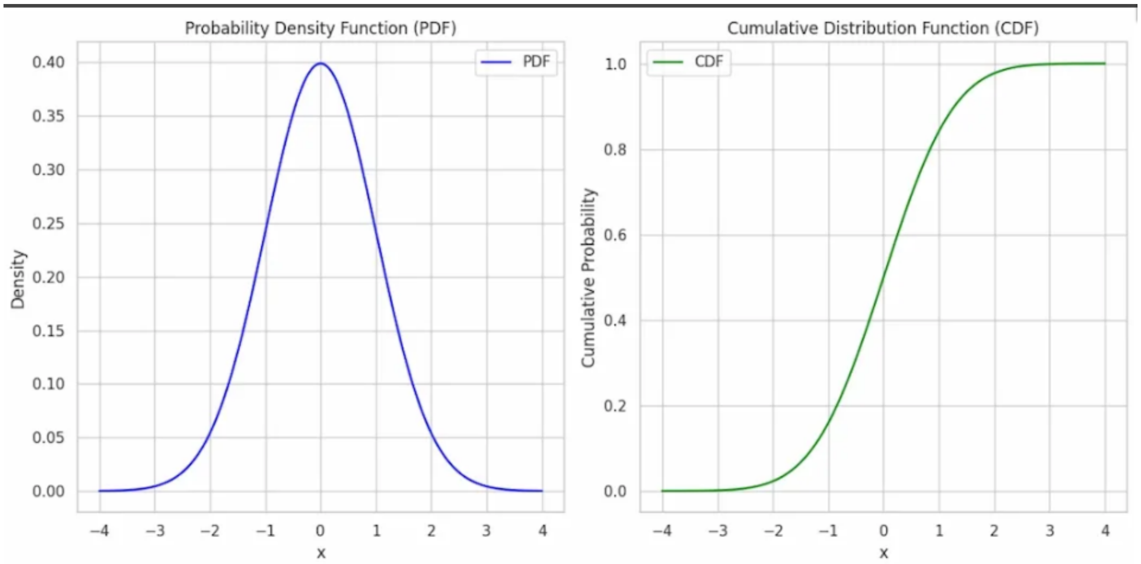
2) Cumulative Distribution Function (CDF)

The Cumulative Distribution Function gives the probability that a random variable is less than or equal to a specific value.It provides a cumulative view of the probability distribution, starting at 0 and increasing to 1 as the value of the random variable increases.

- The CDF starts at 0 for the smallest possible value of the random variable (since there is no probability below this value) and approaches 1 as the value approaches infinity (since the probability of the variable being less than or equal to infinity is 1).
- Probability up to a value:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt$$

- Relation: **CDF is integral of PDF**, **PDF is derivative of CDF**.



🎯 Why CPD Matters in Machine Learning

- Models **uncertainty** with probabilities, not single guesses.
- Assumptions of **normality** in regression, error terms, Bayesian models.
- **MLE** and **Bayesian inference** rely on continuous likelihoods.
- Used in **GMMs, VAEs**, density estimation, anomaly detection.

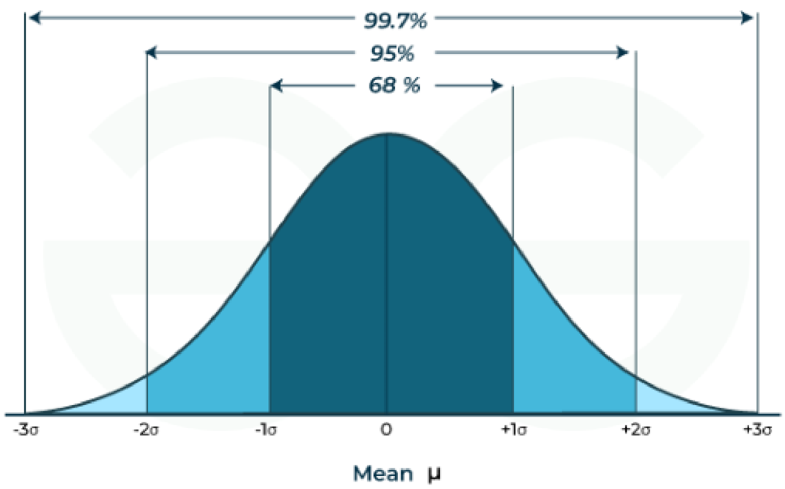
📊 Types of Continuous Probability Distributions

1) **Normal (Gaussian) Distribution/Bell curve:**

The Gaussian Distribution is a bell-shaped, symmetrical basic continuous probability distribution. Two factors define it:

- the standard deviation (σ), which indicates the distribution's spread or dispersion,
- the mean (μ), which establishes the distribution.
  For a random variable x, it is expressed as,

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



- Parameters: mean μ, std σ
- 68–95–99.7 rule around μ

ML uses: regression errors, Naïve Bayes (Gaussian), Kalman filters.

---

2) Uniform Distribution

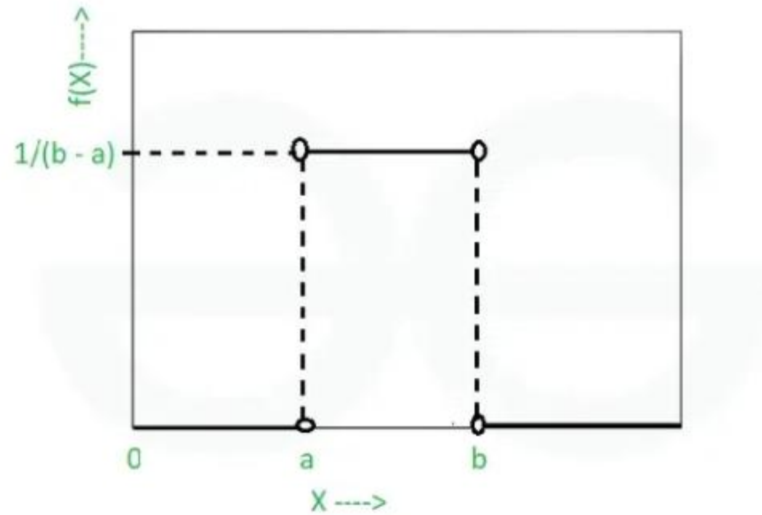
The Uniform Distribution is a continuous probability distribution where all values within a specified range are equally likely to occur.

- Parameters: Lower bound (a) and upper bound (b).
- The mean of a uniform distribution is $\mu = \frac{a+b}{2}$ and the variance is $\sigma^2 = \frac{(b-a)^2}{12}$

It is expressed as:

$$f(x) = \frac{1}{b-a}, a \leq x \leq b$$

$$\mu = \frac{a+b}{2}, \sigma^2 = \frac{(b-a)^2}{12}$$



ML uses: random initialization, random sampling.

---

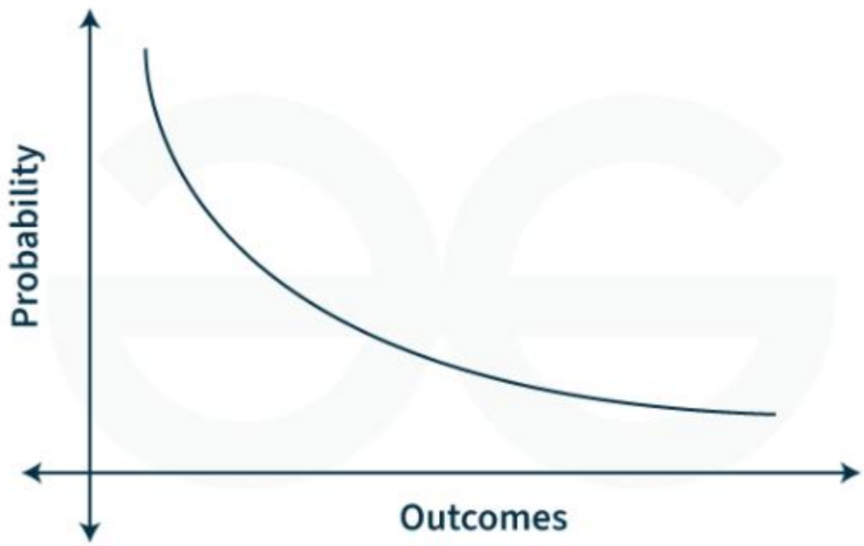
3) Exponential Distribution

The exponential distribution is a continuous probability distribution that represents the duration between occurrences in a Poisson process, which occurs continuously and independently at a constant average rate.

- Parameter: Rate parameter (λ).
- The mean of the exponential distribution is $1/\lambda$ and the variance is $1/\lambda^2$.

For a random variable x, it is expressed as

$$f(x) = \lambda e^{-\lambda x}, x \geq 0$$
$$\mu = \frac{1}{\lambda}, \sigma^2 = \frac{1}{\lambda^2}$$



Exponential Distribution

ML uses: time between events, survival/time-to-event modeling.

---

4) Chi-Squared Distribution

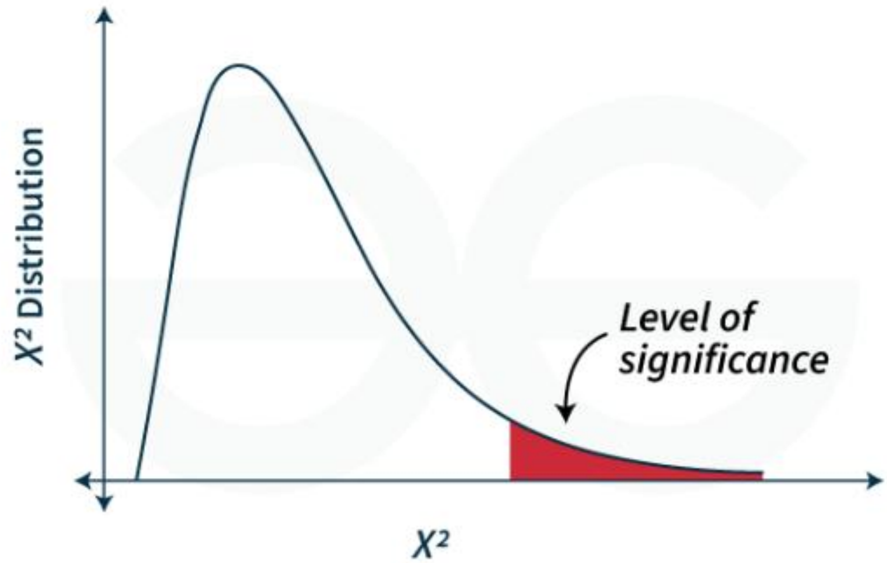
The Chi-Squared Distribution is a continuous probability distribution that arises in statistics, particularly in hypothesis testing and confidence interval estimation.

- It is characterized by a single parameter, often denoted as k ,which represents the degrees of freedom.
- The mean of the Chi-Squared Distribution is k and the variance is 2k.

For a random variable x, it is expressed as

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}, x \geq 0$$

$$\mu = k, \sigma^2 = 2k$$



Chi squared Distribution

ML uses: hypothesis tests, variance estimation, feature tests.

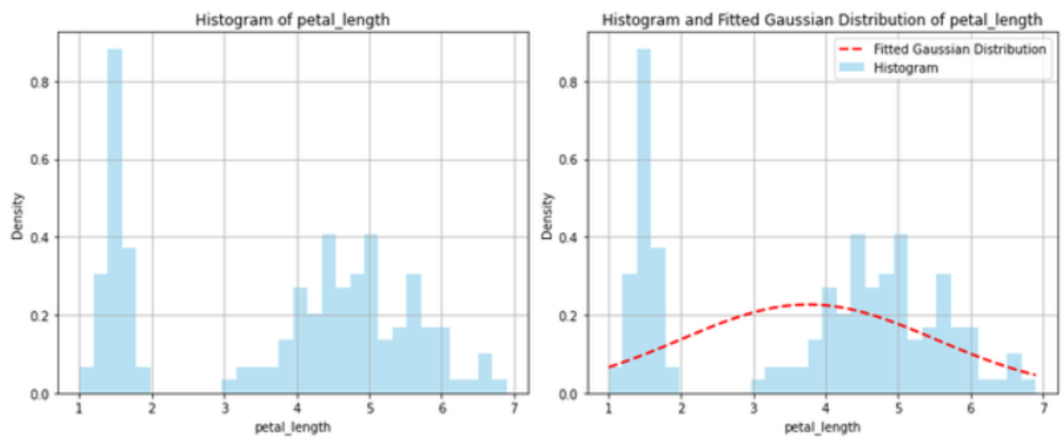
🧪 Determining Distribution from Data (Iris Example)

Example : Lets understand the distribution of a variable with the help of iris dataset

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from scipy.stats import norm


url = "https://raw.githubusercontent.com/uiuc-cse/data-fa14/gh-pages/data/iris.csv"

iris_data = pd.read_csv(url)


selected_feature = 'petal_length'

selected_data = iris_data[selected_feature]


plt.figure(figsize=(12, 5))

plt.subplot(1, 2, 1)
```

```
plt.hist(selected_data, bins=30, density=True, color='skyblue', alpha=0.6)
plt.title('Histogram of {}'.format(selected_feature))
plt.xlabel(selected_feature)
plt.ylabel('Density')
plt.grid(True)

estimated_mean, estimated_std = np.mean(selected_data), np.std(selected_data)

plt.subplot(1, 2, 2)
plt.hist(selected_data, bins=30, density=True, color='skyblue', alpha=0.6)

x = np.linspace(np.min(selected_data), np.max(selected_data), 100)
pdf = norm.pdf(x, estimated_mean, estimated_std)
plt.plot(x, pdf, color='red', linestyle='--', linewidth=2)

plt.title('Histogram and Fitted Gaussian Distribution of {}'.format(
  selected_feature))
plt.xlabel(selected_feature)
plt.ylabel('Density')
plt.legend(['Fitted Gaussian Distribution', 'Histogram'])
plt.grid(True)

plt.tight_layout()
plt.show()
```

Output:



What you see

- Histogram: empirical density
- Red curve: fitted normal PDF using μ, σ

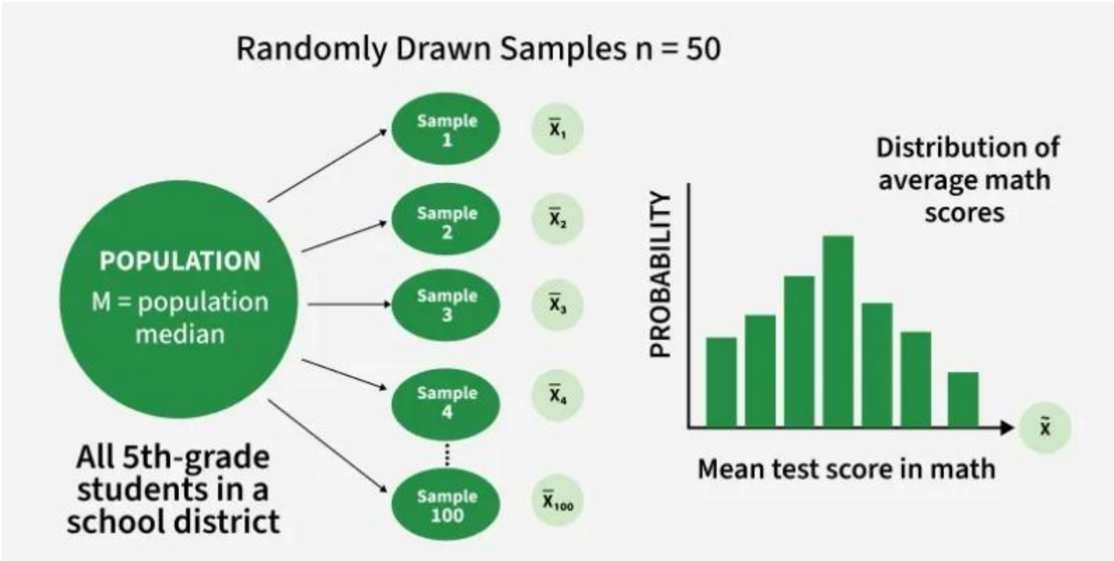
🧠 **Quick Memory Table**

Concept	**Meaning**	**Formula**
PDF	Density at a value	$\int f(x)dx = 1$
CDF	Probability up to x	$F(x) = \int_{-\infty}^{x} f(t)dt$
Normal	Bell curve	μ, σ
Uniform	Equal likelihood	$1/(b-a)$
Exponential	Time between events	$\lambda e^{-\lambda x}$
Chi-Square	Variance tests	$k$dof

PMF is for discrete. PDF is for continuous. Probability = area under the curve.

📘 6.Sampling Distributions

Sampling distribution is essential in various aspects of real life, essential in inferential statistics. A sampling distribution represents the probability distribution of a statistic (such as the mean or standard deviation) that is calculated from multiple samples of a population. It helps us to understand how a statistic varies across different samples and is crucial for making inferences about the population.



Sampling distribution is the probability distribution of a statistic based on random samples of a given population. It is also know as finite distribution.

In this article, we will discuss the Sampling Distribution in detail and its types, along with examples, and go through some practice questions, too.

Important Terminologies in Sampling Distribution

Some important terminologies related to sampling distribution are given below:

- **Statistic:** Summary value from a sample (e.g., mean, median).
- **Parameter:** Summary value from a population.
- **Sample:** A subset of a population.
- **Population:** The entire group being studied.
- **Sampling Distribution:** Distribution of a statistic across many samples.
- **Central Limit Theorem (CLT):** Sample means follow a normal distribution as the sample size increases.
- **Standard Error:** Standard deviation of the sampling distribution.
- **Bias:** Systematic error causing deviation from the true value.
- **Confidence Interval:** Range likely to contain the population parameter.

- **Sampling Method:** How samples are chosen (random, stratified, etc.).
- **Inferential Statistics:** Concluding a population from samples.
- **Hypothesis Testing:** Making decisions about population parameters using sample data.

Factors Influencing Sampling Distribution

The variability of a sampling distribution is measured by standard error or population variance, depending on the context and the type of inference required. Both measure how spread out the data is around the mean.

Main factors influencing the variability of a sampling distribution are:

1. **Number Observed in a Population:** The symbol for this variable is "N." It is the measure of observed activity in a given group of data.
2. **Number Observed in Sample:** The symbol for this variable is "n." It is the measure of observed activity in a random sample of data that is part of the larger grouping.
3. **Method of Choosing Sample:** How you chose the samples can account for variability in some cases.

Types of Distributions

3 main types of sampling distributions are:

- Sampling Distribution of Mean
- Sampling Distribution of Proportion
- T-Distribution

Sampling Distribution of Mean

The sampling distribution of the mean refers to the probability distribution of sample means that you get by repeatedly taking samples (of the same size) from a population and calculating the mean of each sample.

Key concepts of Sampling Distribution of Mean

- **Population Mean (μ):** The average of the entire population.
- **Sample Mean ($\bar{x}$):** The average of a sample taken from the population.

- **Sampling Distribution of the Mean:** If you take multiple samples and plot their means, that plot will form the sampling distribution of the mean.

For any population with mean μ and standard deviation σ:

- Mean, or center of the sampling distribution of x̄, is equal to the population mean, μ.

$\mu_{x^-} = \mu$

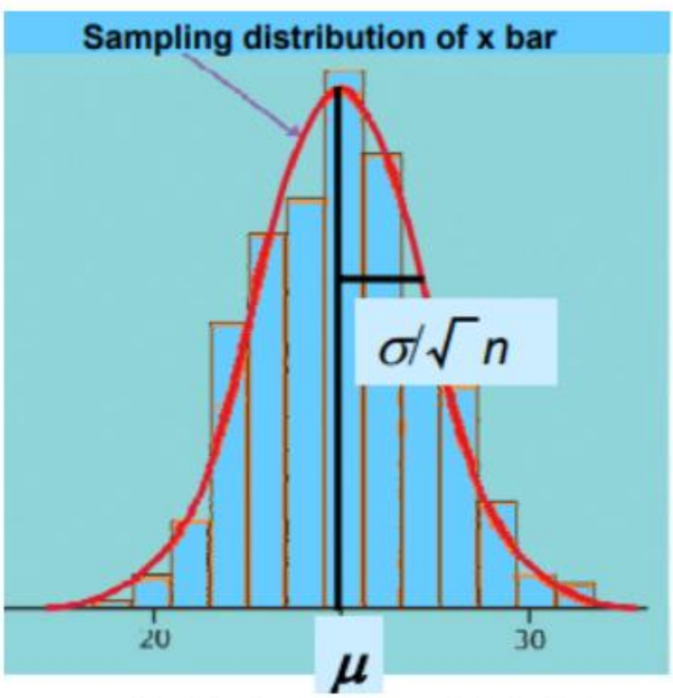
There is no tendency for a sample mean to fall systematically above or below μ, even if the distribution of the raw data is skewed. Thus, the mean of the sampling distribution is an unbiased estimate of the population mean μ.

- Standard deviation of the sampling distribution is σ/√n, where n is the sample size.

$\sigma_x = \sigma/\sqrt{n}$

Where

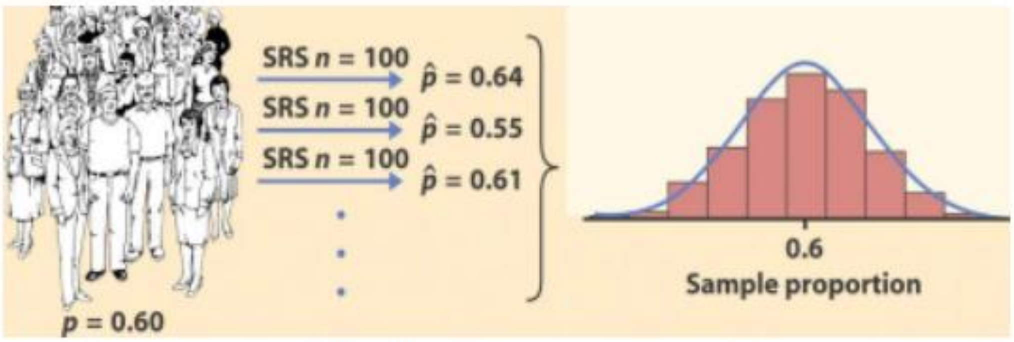
- σ_x - Standard Deviation of Sampling Deviation
- σ - Population Standard Deviation
- n - Sample size



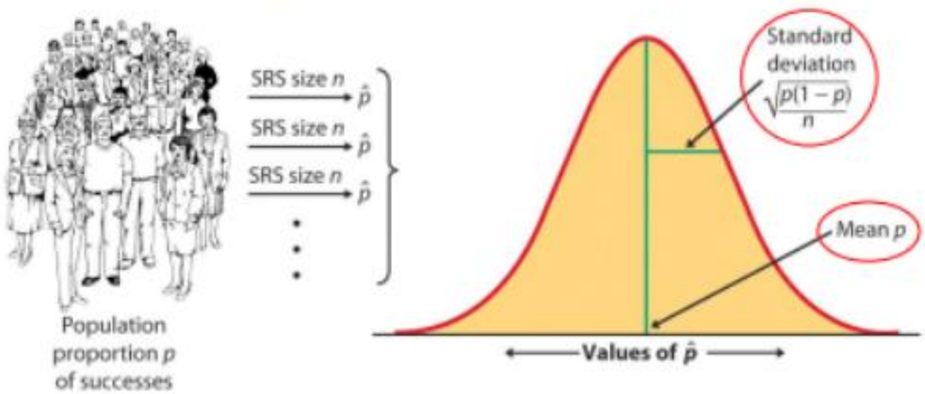
sampling distribution of standard deviation

Sampling Distribution of Proportion

Sampling distribution of a proportion focuses on proportions in a population. Here, you select samples and calculate their corresponding proportions. The means of the sample proportions from each group represent the proportion of the entire population.



Sampling distribution of proportion - 1



Sampling distribution of proportion - 2

The formula for the sampling distribution of a proportion (often denoted as $\hat{p}$) is:

$\hat{p} = x/n$

Where:

- **$\hat{p}$** is the Sample Proportion
- **x** is the Number of "successes" or occurrences of the Event of Interest in the Sample
- **n** is Sample Size

This formula calculates the proportion of occurrences of a certain event (e.g., success, positive outcome) within a sample.

T-Distribution

Sampling distribution involves a small population or a population about which you don't know much. It is used to estimate the mean of the population and other statistics such as confidence intervals, statistical differences, and linear regression. T-distribution uses a t-score to evaluate data that wouldn't be appropriate for a normal distribution.

The formula for the t-score, denoted as t, is:

$$t = [x - \mu] / [s / \sqrt{(n)}]$$

Where:

- **x** is the Sample Mean
- **μ** is Population Mean (or an estimate of it)
- **s** is the Sample Standard Deviation
- **n** is Sample Size

This formula calculates the difference between the sample mean and the population mean, scaled by the standard error of the sample mean. The t-score helps to assess whether the observed difference between the sample and population means is statistically significant.

Solved Examples of Sampling Distribution

Example 1: Mean and standard deviation of the tax value of all vehicles registered in a certain state are μ=\$13,525 and σ=\$4,180. Suppose random samples of size 100 are drawn from the population of vehicles.
Find

- mean $\mu_{\bar{x}}$
- standard deviation $\sigma_{\bar{x}}$ of the sample mean $\bar{x}$

Solution:

Since n = 100, the formulas yield

$\mu_{\bar{x}} = \mu =$ ***\$13,525***

$\sigma_{\bar{x}} = \sigma / \sqrt{n} = \$4180 / \sqrt{100}$

$\boldsymbol{\sigma_{\bar{x}} = \$418}$

Example 2: A prototype automotive tire has a design life of 38,500 miles with a standard deviation of 2,500 miles. Five such tires are manufactured and tested. On the assumption that the actual population mean is 38,500 miles and the actual population standard deviation is 2,500 miles, find the probability that the sample mean will be less than 36,000 miles. Assume that the distribution of lifetimes of such tires is normal.

Solution:

Here, we will assume and use units of thousands of miles.

Then sample mean $\bar{x}$ has

- *Mean: $\mu_{\bar{x}} = \mu =$ **38.5***
- *Standard Deviation: $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 2.5/\sqrt{5} =$ **1.11803***

Since the population is normally distributed, so is $\bar{x}$, hence,

$P\ (X < 36) = P(Z < \{36 - \mu_{\bar{x}}\}/\sigma_{\bar{x}})$

$P\ (X < 36) = P(Z < \{36 - 38.5\}/1.11803)$

$P\ (X < 36) = P(Z < -2.24)$

$P(X < 36) = 0.0125$

*Therefore, if the tires perform as designed then there is only about a **1.25%** chance that the average of a sample of this size would be so low.*

Practice Questions on Sample Distribution

Question 1: Random samples of size 225 are drawn from a population with a mean of 100 and a standard deviation of 20. Find the mean and standard deviation of the sample mean.

Question 2: Random samples of size 64 are drawn from a population with a mean of 32 and a standard deviation of 5. Find the mean and standard deviation of the sample mean.

Question 3: A population has a mean of 75 and a standard deviation of 12.

1. Random samples of size 121 are taken. Find the mean and standard deviation of the sample mean.
2. How would the answers to part
3. Change if the size of the samples were 400 instead of 121?

Question 4: A population has a mean of 5.75 and a standard deviation of 1.02.

1. Random samples of size 81 are taken. Find the mean and standard deviation of the sample mean.
2. How would the answers to part
3. Change if the size of the samples were 25 instead of 81?

Question 5: The Numerical population of grade point averages at a college has a mean of 2.61 and a standard deviation of 0.5. If a random sample of size 100 is taken from the population, what is the probability that the sample mean will be between 2.51 and 2.71?

Question 6: Random samples of size 1,600 are drawn from a population in which the proportion with the characteristic of interest is 0.05. Decide whether or not the sample size is large enough to assume that the sample proportion is normally distributed.

📘 7.Chi-Square Distributions

The Chi-Squared distribution (also chi-square or χ^2-distribution) represents the distribution of the sum of the squares of k independent standard normal random variables. If $Z_1, Z_2, \ldots, Z_k$ are independent standard normal random variables, the Chi-Squared variable is: $X^2 = Z_1^2 + Z_2^2 + \cdots + Z_k^2$

The Chi-Squared distribution is parameterised by the **degrees of freedom (df)**, which corresponds to the number of independent random variables being summed.

- The chi-square distribution is actually a series of distributions that vary in shape according to their degrees of freedom. As the degrees of freedom increase, the distribution becomes more symmetric and approaches a normal distribution.
- The chi-square test is a hypothesis test designed to test for a statistically significant relationship between nominal and ordinal variables organized in a bivariate table. In other words, it tells us whether two variables are independent of one another.

Probability Density function (PDF) of Chi-Squared Distribution

The PDF of a Chi-Squared distribution with k degrees of freedom is:

$$f(x;k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2}, x \geq 0$$

where Γ is the Gamma function.

Properties of Chi-Squared Distribution

Some of the common properties of Chi-Squared Distribution are discussed below:

Non-Negativity

The Chi-Squared distribution is defined only for non-negative values ($x \geq 0$) because it is based on the sum of squared standard normal variables, which are always non-negative.

Degrees of Freedom

The shape of the Chi-Squared distribution depends on the number of degrees of freedom (k).

- For small k, the distribution is positively skewed.
- As $k \to \infty$ the distribution approaches a normal distribution (via the Central Limit Theorem).

Mean

The mean of the Chi-Squared distribution is equal to its degrees of freedom: Mean = k

Variance

The variance of the Chi-Squared distribution is twice its degrees of freedom: Variance = 2k

Standard Deviation (SD)

The standard deviation of the Chi-Squared distribution is the square root of the variance, so: $SD = \sqrt{2 \times \text{df}}$

Skewness

The skewness decreases as the degrees of freedom increase: Skewness = $\sqrt{\frac{8}{k}}$

- For small k, the distribution is heavily skewed to the right.
- As k increases, the skewness approaches 0.

Kurtosis (Excess)

The kurtosis (excess) of the Chi-Squared distribution is: Excess Kurtosis = 12/k

- This shows that the distribution becomes less peaked as k increases.

Additive Property

If $X_1, X_2, \ldots, X_m$ are independent Chi-Squared random variables with degrees of freedom $k_1, k_2, \ldots, k_m$, then their sum is also a Chi-Squared random variable: $X = X_1 + X_2 + \cdots + X_m \sim \chi^2(k_1 + k_2 + \cdots + k_m)$

Moment-Generating Function (MGF)

The MGF of a Chi-Squared random variable with k degrees of freedom is: $M(t) = (1 - 2t)^{-k/2}, t < \frac{1}{2}$

Chi-Square Distribution with 1 Degree of Freedom

The $\chi 2(1)$ distribution is a special case of the $\chi 2$ distribution, which is based on the sum of the squares of independent standard normal variables. Specifically, if Z is a standard normal variable (mean = 0, standard deviation = 1), then:

$X = Z^2$follows a$\chi^2(1)$ distribution.

Since squaring a standard normal variable always results in a non-negative value, the $\chi^2(1)$ distribution is skewed to the right with its values ranging from 0 to positive infinity.

Chi-Squared Distributions in R Language

R is a popular high level programming language used for statistical analysis. It is open-source programming language as it has a huge community and users can contribute to the development as well. It has vast number of packages which allows the data miners to perform statistical analysis and data visualizations in an interactive manner.

R has 4 built in functions for chi-square distribution.

dchisq() function

The dchisq() function calculates the probability density of a chi-squared distribution. It can compute cumulative probabilities with lower.tail = TRUE for the left tail or lower.tail = FALSE for the right tail.

Syntax:

dchisq (x, df, ncp = 0, log= FALSE)

where x= vector of quantiles.
p= vector of probabilities.
n= number of observations

df= degrees of freedam
ncp= non-centrality parameter (non-negative).
log.p– logical; if TRUE, probabilities p are given as log(p).
lower.tail= this is a logical value, if TRUE (default),probabilities are $P[X \leq x]$, *otherwise P[X>x].*

pchisq() function

The pchisq() function gives the cumulative distribution function (CDF) for the chi-squared distribution. The dchisq(x, df) function calculates the probability density of a chi-squared distribution for a value x with df degrees of freedom. It can also be used to determine the area under the chi-squared curve for specified intervals with a given number of degrees of freedom.

Syntax:

pchisq(q, df, ncp = 0, lower.tail = TRUE, log.p = FALSE)

qchisq() function

The qchisq() function returns the quantile function for the chi-squared distribution. When the non-centrality parameter (ncp) is set to 0, it computes for the central chi-squared distribution. This method gives the value of x at the qth percentile, where lower.tail = TRUE corresponds to the cumulative probability up to x.

syntax:

qchisq(p, df, ncp = 0, lower.tail = TRUE, log.p = FALSE)

rchisq() function:

The rchisq(n, df) function generates n random numbers from the chi-square distribution with df degrees of freedom. It is used to produce random deviates from the chi-square distribution.

Syntax:

rchisq (n, df, ncp = 0)

read more about [Chi-Squared Distributions in R.]()

Chi-squared Distributions

Non-Central Chi-Square distribution

The non-central chi-square distribution is a generalization of the chi-square distribution, often used in power analyses. It introduces an additional parameter, λ known as the non-central parameter. This parameter shifts the distribution's peak to the right and increases the variance as λ increases.

The λ parameter influences the mean of the normal distributions that make up the chi-square distribution. For instance, a non-central chi-square distribution with $\lambda=2$ and $k=3$ can be generated by squaring and summing values from three normal distributions, each with a mean of 2 and a variance of 1.

Generalized Chi-squared Distribution

The Generalized Chi-squared Distribution is a more flexible version of the standard Chi-squared distribution. It's based on a mathematical expression called a *quadratic form*, which looks like $z'Az$.

Here's what each part means:

- z is a vector of random variables that follow a **Gaussian distribution** (i.e., they are normally distributed) with a **mean of zero**.
- A is a matrix that helps define how the variables in z are related to each other (i.e., it defines their **covariance** or how they vary together).
- $z'Az$ represents a mathematical operation where you multiply the vector z by the matrix A and then by the vector z again. This produces a single number, which is the value from the distribution.

In simpler terms, the generalized chi-squared distribution describes how a set of correlated normal variables behave when combined in a specific way. This allows for more complex scenarios than the basic chi-squared distribution.

📘 8.Student's t-Distribution

The Student's t-distribution (or simply the t-distribution) is a probability distribution used in statistics when making inferences about a population mean, particularly when the sample size is small ($n \le 30$) or the population standard deviation (σ) is unknown. It resembles the standard normal distribution but has heavier tails, allowing it to better handle variability in small samples. The t-score indicates how many estimated standard errors the sample mean ($\bar{x}$) is away from the population mean (μ).

Formula for the t-Score

The t-score (or t-statistic) quantifies how many estimated standard errors the sample mean ($\bar{x}$) is from the population mean (μ):

$$t = \frac{\bar{x} - \mu}{s\sqrt{n}}$$

where,

- t = t-score,
- x̄ = sample mean
- μ = population mean,
- s = standard deviation of the sample,
- n = sample size

The t-score helps determine how far the sample mean is from the population mean under the assumption of random sampling.

When to Use the t-Distribution

Student's t Distribution is used when :

- The sample size is 30 or less than 30.
- The population standard deviation(σ) is unknown.
- The population distribution must be unimodal and skewed.

Interpretation of t-Distribution with Example

Suppose a researcher wants to estimate the average daily study time of students before exams. A random sample of 20 students reports an average (x̄) of 4 hours with a sample standard deviation (s) of 1.5 hours. We want to construct a 90% confidence interval for the true population mean.

Given:

- x̄ = 4 hours, s = 1.5 hours, n = 20 and a 90% confidence level.
- Degrees of freedom = n – 1 = 19
- Critical t-value (from t-table) ≈ 1.729

$$CI = \bar{x} \pm t \times \frac{s}{\sqrt{n}}$$

Substituting the given values:

$$CI = 4 \pm 1.729 \times \frac{1.5}{\sqrt{20}} = (3.42, 4.58)$$

Interpretation: We are 90% confident that the true average study time for all students lies between 3.42 and 4.58 hours per day. This range indicates where the actual population mean is most likely to fall, given the data and confidence level.

Implementation

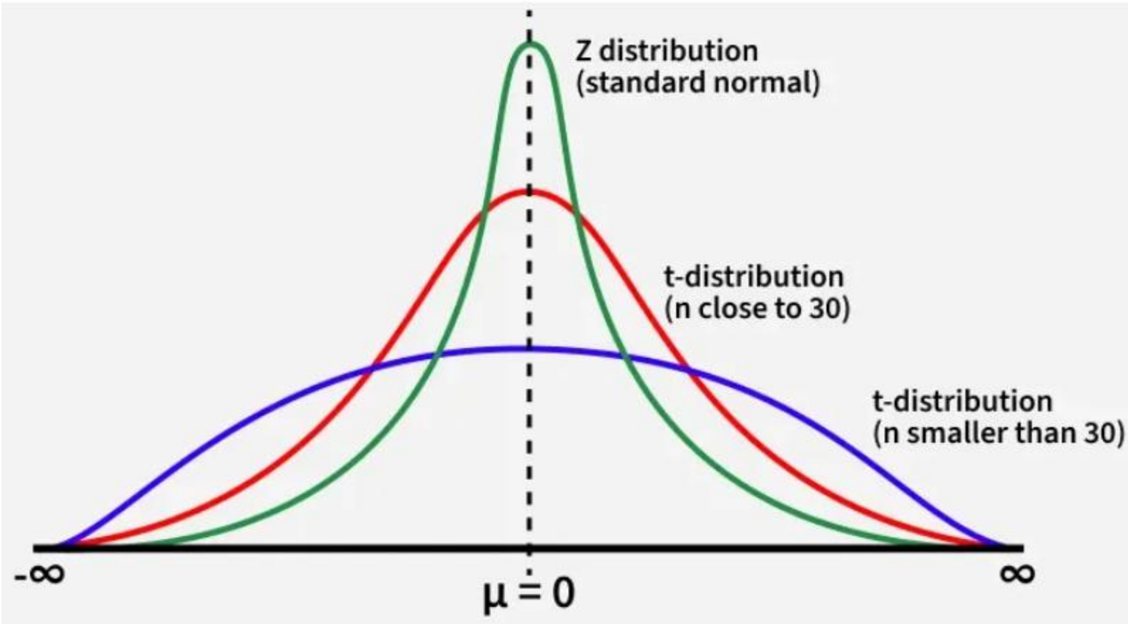
Let's implement the example in Python using scipy.stats:

```
import numpy as np
from scipy import stats

x_bar = 4
s = 1.5
n = 20
confidence = 0.90

df = n - 1

t_critical = stats.t.ppf((1 + confidence) / 2, df)

margin_of_error = t_critical * (s / np.sqrt(n))

lower_bound = x_bar - margin_of_error
upper_bound = x_bar + margin_of_error

print(f"t-critical value: {t_critical:.3f}")
print(f"Confidence Interval (90%): ({lower_bound:.2f}, {upper_bound:.2f})")
```

Output:

t-critical value: 1.729
Confidence Interval (90%): (3.42, 4.58)

Properties of the t-Distribution



- t is symmetric and bell-shaped, like the normal distribution.
- The variable t ranges from $-\infty$ to $+\infty$.
- As degrees of freedom (df) increase, the t-distribution approaches the standard normal distribution (Z).
- The mean, median and mode are all zero (for df > 1).
- Variance = df / (df − 2) for df > 2, indicating that as df increases, variance decreases.
- It has heavier tails than the normal distribution, making it more flexible for small samples.

Key Elements in Student's t-Distribution

1. t-Distribution Table

- Provides critical t-values for various confidence levels and degrees of freedom.
- Used to determine whether the calculated t-score falls in the rejection region of a hypothesis test.
- **Example**: At a 5% significance level ($\alpha = 0.05$), if the calculated $|t|$ exceeds the tabulated t-value, the difference between sample and population means is considered statistically significant.

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

T- Distribution table

2. t-Score

- t-Score measures how far a sample mean deviates from the population mean in terms of standard errors.
- Helps determine statistical significance and construct confidence intervals.
- Larger |t| values indicate a greater difference between sample and population means.

3. p-Value

- p-Value Represents the probability of obtaining a test result at least as extreme as the one observed, assuming the null hypothesis is true.

- A small p-value (typically < 0.05) suggests strong evidence against the null hypothesis, leading to its rejection.
- It can be obtained from statistical software or a t-table using the calculated t-score and degrees of freedom.

Applications

- **Testing a Population Mean**: Determine if a sample mean significantly differs from a known or hypothesized population mean.
- **Comparing Two Means**: Evaluate whether two independent or paired samples have different means.
- **Testing Correlation**: Assess if the correlation coefficient between two variables significantly differs from zero.

Limitations

- **Assumes Normality**: The t-distribution relies on the assumption that the underlying population is normally distributed. Significant deviations can lead to inaccurate results.
- **Less Useful for Large Samples**: As the sample size increases, the t-distribution approaches the normal distribution, making the latter more suitable for large datasets.
- **Sensitive to Outliers**: Extreme values can distort results, especially in small samples where the t-distribution's heavy tails amplify their effect.
- **Requires Random Sampling**: Valid conclusions depend on random, independent observations. Violating these assumptions reduces the reliability of inferences.

Difference Between T-Distribution and Normal Distribution

Aspect	t-Distribution	Normal Distribution
Definition	Defined by degrees of freedom (df) depending on sample size	Defined by mean (μ) and standard deviation (σ)

Aspect	t-Distribution	Normal Distribution
Sample Size	Used for small samples ($n \leq 30$)	Used for large samples ($n > 30$)
Standard Deviation	Unknown (estimated from sample)	Known
Shape	Heavier tails more prone to extreme values	Lighter tails, data closer to mean
Application	Hypothesis testing when σ is unknown	When σ is known or sample size is large
Range of Critical Values	Wider range due to more uncertainty	Narrower range with less uncertainty