Probability

Probability is the measure of how likely an event is to happen. It gives a numerical value to uncertainty, helping us predict outcomes in situations where there is randomness.

Probability is a number between 0 and 1 that represents the chance of an event occurring:
•0 means the event will never happen (Impossible event).
•1 means the event will definitely happen (Certain event).
•A probability between 0 and 1 means there is some chance of the event happening.

Formula for Probability:

$$P(E) = \frac{\text{Number of favorable outcomes}}{\text{Total number of possible outcomes}}$$

where:

- $P(E)$ is the probability of an event E.
- Favorable outcomes are the outcomes we are interested in.
- Total outcomes are all possible results.

Example of Probability

Imagine you have a bag with 5 balls:

•3 red balls

•2 blue balls

You pick one ball at random.

What is the probability of picking a red ball?

$$P(\text{Red Ball}) = \frac{\text{Number of Red Balls}}{\text{Total Balls}}$$

$$= \frac{3}{5} = 0.6 \text{ or } 60\%$$

This means there is a **60% chance** of picking a red ball.

What is the probability of picking a blue ball?

$$P(\text{Blue Ball}) = \frac{\text{Number of Blue Balls}}{\text{Total Balls}}$$

$$= \frac{2}{5} = 0.4 \text{ or } 40\%$$

So, there is a **40% chance** of picking a blue ball.

Applications of Probability in AI & Machine Learning

Bayesian Networks (Belief Networks)

- **Use:** Models uncertainty and dependencies between variables.
- **Example:** Medical diagnosis systems use **Bayesian Networks** to predict diseases based on symptoms.

Naïve Bayes Classifier

- **Use:** A simple yet powerful classification algorithm based on Bayes' theorem.
- **Example:** Used in **spam detection**, where it calculates the probability of an email being spam based on words present.

Reinforcement Learning (RL)

- **Use:** Probability helps agents make the best decision under uncertainty.
- **Example:** AI in **robotics or self-driving cars** chooses actions based on probabilities of rewards.

Probabilistic Graphical Models

- **Use:** Represent complex relationships between variables.
- **Example: Hidden Markov Models (HMM)** for speech recognition and **Markov Decision Processes (MDP)** for decision-making in robotics.

Machine Learning Algorithms (Probabilistic Models)

- **Linear Regression:** Uses probability to estimate the best-fit line.
- **Logistic Regression:** Uses **sigmoid probability** to classify data points.

Natural Language Processing (NLP)

- **Use:** Probability is used in language modeling, text generation, and speech recognition.
- **Example: Google Translate** and **Chatbots** use probabilistic models to predict the next word in a sentence.

Basic Terminologies in Probability

Term	Explanation	Example
Experiment(Trial, Observation, Procedure)	An action or process that leads to one or more outcomes.	Tossing a coin or rolling a dice.
Outcome(Result, Observation, Realization)	The result of a single trial of an experiment.	"Heads" or "Tails" when tossing a coin.
Event(Subset of Sample Space, Incident)	A specific result or set of results that we are interested in.	Getting an even number when rolling a dice (2, 4, 6).
Sample Space (S) (Universal Set, Outcome Space)	The set of all possible outcomes of an experiment.	Rolling a dice: {1, 2, 3, 4, 5, 6}.
Probability (P) (Likelihood, Chance, Measure of Uncertainty**)**	The likelihood of an event occurring, expressed as a fraction or decimal between 0 and 1.	P(Heads) = 1/2 when tossing a coin.
Random Variable**(**Stochastic Variable, Aleatory Variable)	A variable that takes numerical values based on outcomes of a random experiment.	Number of heads in 3 coin tosses (0, 1, 2, 3).
Event Space(Event Set, Event Collection)	A subset of the sample space that contains outcomes related to a specific event.	Getting an odd number on a dice: {1, 3, 5}.

Basic Axioms of Probability (Kolmogorov's Axioms)

In 1933, Andrey Kolmogorov introduced three fundamental rules, called the Axioms of Probability, which define how probability works in any situation. These are the foundation of probability theory and are used in statistics, machine learning, and AI.

Non-Negativity Axiom

$$P(A) \geq 0$$

- The probability of any event **A** is always **0 or a positive number** (it can't be negative).
- Example: You cannot say **"The probability of rolling a 6 on a die is -0.2"**—that makes no sense!

Certainty Axiom (Total Probability is 1)

$$P(S) = 1$$

- The probability of **all possible outcomes together is always 1**.
- Example: If you roll a fair die, the probability of getting **any number (1 to 6) must add up to 1**.

Additivity Axiom (For Mutually Exclusive Events)

$$P(A \cup B) = P(A) + P(B), \quad \text{if } A \text{ and } B \text{ are mutually exclusive}$$

- If two events **cannot happen at the same time**, the probability of either happening is the **sum of their individual probabilities**.
- Example: In a deck of cards, the probability of drawing a **King** or a **Queen** is:

$$P(King) + P(Queen) = \frac{4}{52} + \frac{4}{52} = \frac{8}{52}$$

Why Are Kolmogorov's Axioms Important?

1. Defines Probability Clearly– Turns probability into a proper mathematical subject.
2. Keeps Probability Values Correct – Ensures probabilities are always between 0 and 1.
3. Helps in Real-World Applications – Used in AI, ML, finance, and science.
4. Ensures Consistency – Prevents contradictory or illogical probability results.
5. Aids Decision-Making – Helps in predicting risks and making informed choices.

Addition Rule of Probability

The Addition Rule is used when we calculate the probability of either one event or another happening.

When the events are Mutually Exclusive

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

When the events are not mutually exclusive

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

Mutually Exclusive & Not Mutually Exclusive Events

Mutually Exclusive Events

•Example: Rolling Two events are mutually exclusive if they cannot happen at the same time.

•a die and getting a 3 or a 5.

- You can get either 3 or 5, but not both at the same time.
- So, these events are mutually exclusive.

Not Mutually Exclusive Events

•Two events are not mutually exclusive if they can happen together.

•Example: Drawing a card from a deck and getting a red card and a King.

- A card can be both red and a King (e.g., King of Hearts, King of Diamonds).
- So, these events are not mutually exclusive.

Multiplication Rule of Probability

The Multiplication Rule helps us find the probability of two events happening together (at the same time or one after the other). It depends on whether the events are independent or dependent.

Independent Events

- Two events are independent if one event does not affect the other.
- The probability of both happening is found by multiplying their probabilities

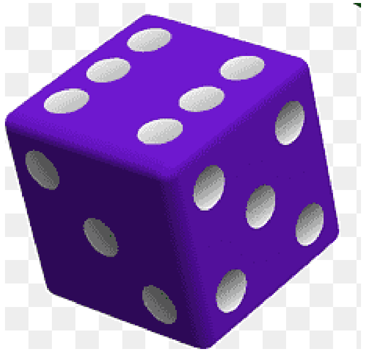
Formula:

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

Example:

- Tossing a coin and rolling a die:
  - Probability of getting **heads** = $1/2$
  - Probability of rolling a **6** = $1/6$
  - Since tossing a coin **does not affect** rolling a die:

$$P(H \cap 6) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$$



Dependent Events

Two events are dependent if one event affects the probability of the other.
The probability of both happening is found by multiplying the probability of the first event by the probability of the second event happening after the first.

Formula:

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

(Where $P(B|A)$ is the probability of **B happening after A has already happened**.)

Example:

- Drawing **two Aces in a row** from a deck of 52 cards:
  - Probability of first Ace = $4/52$
  - Probability of second Ace (since one Ace is removed) = $3/51$
  - Since drawing the first Ace **changes** the probability of the second, we use the formula:

$$P(A_1 \cap A_2) = \frac{4}{52} \times \frac{3}{51} = \frac{12}{2652} = \frac{1}{221}$$

Law of Total Probability

The Law of Total Probability is a fundamental rule that helps calculate the overall probability of an event by considering all possible ways that event can occur. It breaks down a complex probability problem into smaller, manageable parts.

Formula

If $B_1, B_2, B_3, \ldots, B_n$ are **mutually exclusive and exhaustive events** (events that cover the entire sample space without overlapping), the law states:

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

Where:

- $P(A)$ → Total probability of event A occurring.
- $P(A|B_i)$ → Conditional probability of A given that B_i has occurred.
- $P(B_i)$ → Probability of the event B_i.
- The summation ($\sum$) runs over all the **possible outcomes** (B_i).

Example: Diagnosing a Disease

Suppose there are three hospitals (A, B, C) in a city with different accuracy rates in diagnosing a disease:
•Hospital A: 90% accurate, 30% of patients go here.
•Hospital B: 80% accurate, 50% of patients go here.
•Hospital C: 70% accurate, 20% of patients go here.

Step 1: Define Events

- A = Correct diagnosis
- B_1 = Hospital A
- B_2 = Hospital B
- B_3 = Hospital C

Step 2: Apply the Formula

$$P(A) = P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2) + P(A|B_3) \cdot P(B_3)$$

Substituting values:

$$P(A) = (0.9 \times 0.3) + (0.8 \times 0.5) + (0.7 \times 0.2)$$

$$P(A) = 0.27 + 0.40 + 0.14 = 0.81$$

The overall probability of a correct diagnosis from any hospital is 81%.

Bayes' Theorem

Bayes' Theorem is a mathematical formula that helps us **update the probability of an event** based on new information or evidence. It combines **prior knowledge** with **new evidence** to give an updated probability.

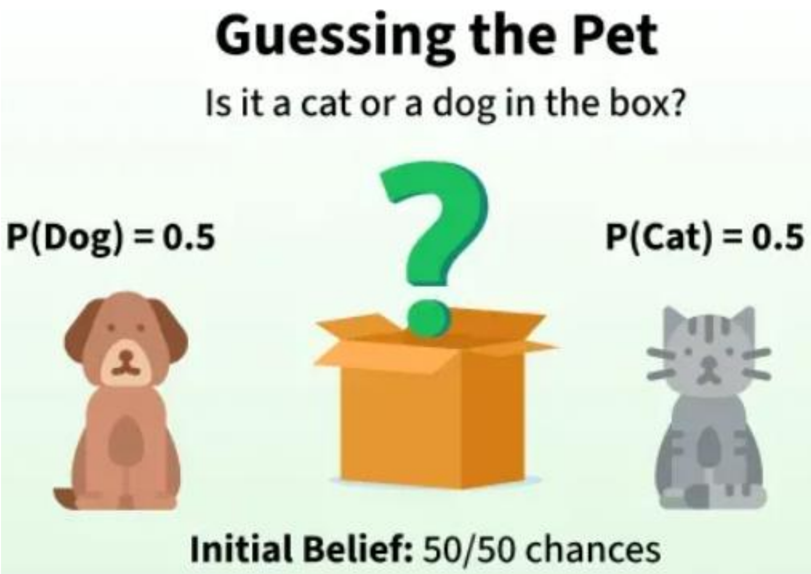
Bayes' Theorem helps us **update our beliefs** when we receive **new information**.

Formula:

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

This means:

- $P(A|B)$ **(Posterior Probability)** → The chance of **A being true**, given that **B has happened**.
- $P(B|A)$ **(Likelihood)** → The chance of **B happening**, assuming **A is true**.
- $P(A)$ **(Prior Probability)** → The original probability of **A happening**, before knowing about B.
- $P(B)$ **(Marginal Probability)** → The total probability of **B happening**, considering all possible cases.



Where is Bayes' Theorem Used?

1. Spam Filters (Text Classification) → Decides if an email is spam based on words like "lottery" or "win".
2. Medical Tests (Disease Diagnosis) → Finds out how likely someone has a disease after a positive test.
3. AI & Machine Learning (Predictive Modeling) → Used in chatbots, recommendation systems, and fraud detection.

Understanding Each Term

1. Prior Probability (P(A)):
The initial probability of an event happening before taking any new evidence into account.
- Example: The probability of having a disease without any test result.

2. Likelihood (P(B | A)):
The probability of observing evidence (B) if the event (A) is true.
- Example: The chance of getting a positive test result if a person actually has the disease.

3. Marginal Probability (P(B)):
The total probability of observing the evidence (regardless of whether A is true or not).
- Example: The overall chance of getting a positive test result from all people tested.

4. Posterior Probability (P(A | B)):
The revised probability of an event after considering the new evidence.
- Example: The probability of having a disease after getting a positive test result.

Example: Is Your Friend Lying?

Imagine your friend is known to lie 30% of the time (Prior Probability of Lying = 0.3). Your friend tells you, “It’s raining outside.” You also know that rain happens 20% of the time in your city (Prior Probability of Rain = 0.2). If your friend says “It’s raining,” what is the chance they are telling the truth?
Using Bayes’ Theorem:

Identify the Given Values

- $P(T|R) = 0.7$ → Chance friend tells the **truth if it’s raining** (**Likelihood**).
- $P(R) = 0.2$ → Chance of **rain happening** (**Prior Probability**).
- $P(T|R^c) = 0.3$ → Chance friend tells the **truth if no rain** (**Likelihood for the Complement Event**).
- $P(R^c) = 0.8$ → Chance of **no rain happening** (**Complement Probability**).

By applying Bayes’ Theorem, we can calculate the actual probability of rain, given that your friend says so.

Calculate P(T) (Total Probability of Telling the Truth)

Using the **Law of Total Probability**:

$$P(T) = P(T|R) \cdot P(R) + P(T|\neg R) \cdot P(\neg R)$$

Substituting the values:

$$P(T) = (0.7 \times 0.2) + (0.3 \times 0.8)$$

$$P(T) = 0.14 + 0.24 = 0.38$$

Calculate P(T) (Total Probability of Telling the Truth)

$$P(R|T) = \frac{P(T|R) \cdot P(R)}{P(T)}$$

Substituting the values:

$$P(R|T) = \frac{0.7 \times 0.2}{0.38}$$

$$P(R|T) = \frac{0.14}{0.38} \approx 0.368$$

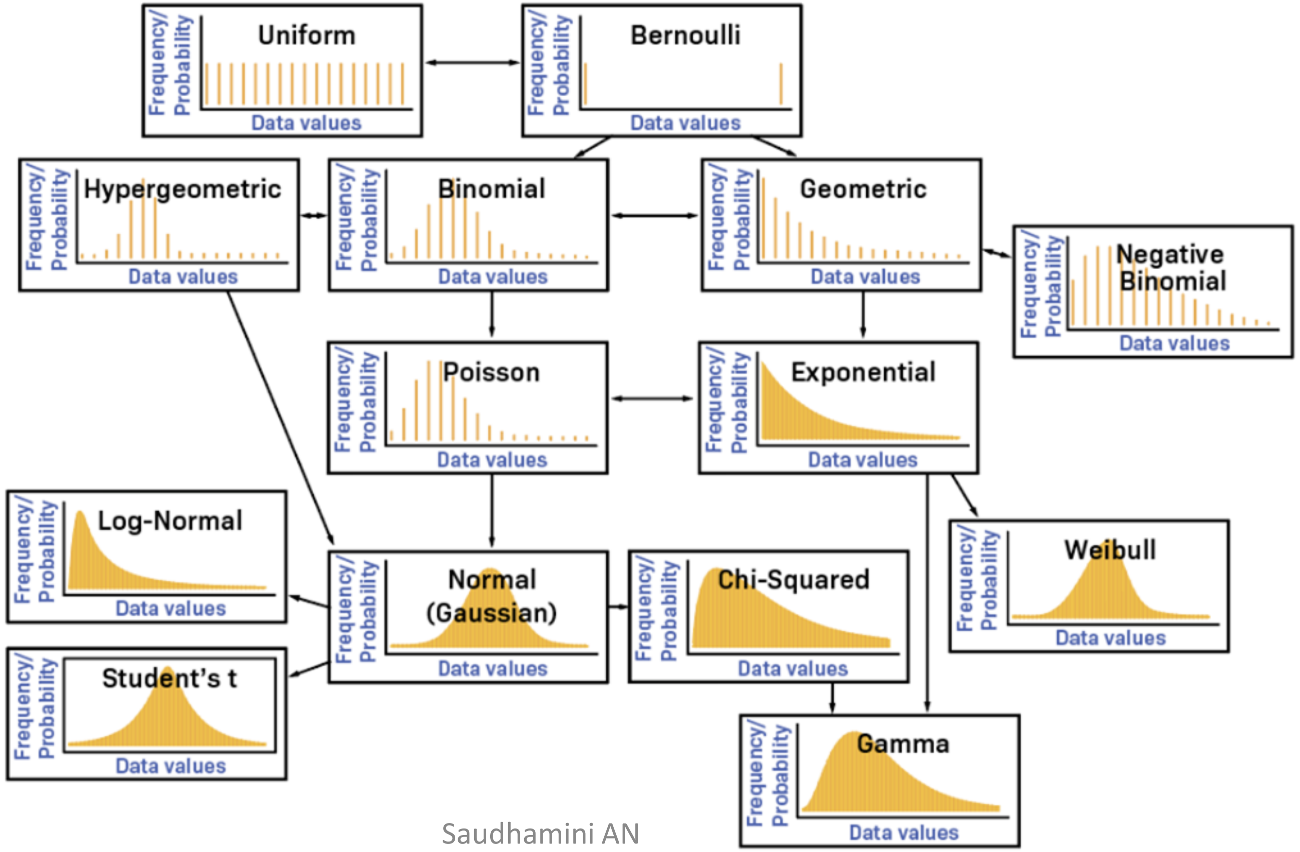
The actual probability that it is raining, given that your friend says so, is approximately 36.8%.

Distributions in Statistics

In statistics, a distribution tells us how the values of a random variable are spread or arranged. It shows us the pattern of how data points are distributed across different possible values.

Types of Distributions

1. Discrete Distributions
2. Continuous Distribution



Properties of Probability Distributions

1. All probabilities are positive or zero
   - You can't have a negative chance.
   - Example: You can have a 0% or 50% chance, but not –10%.
2. Total probability is 1
   - The chances of all possible outcomes add up to 1 (or 100%).
   - This means something must happen.
3. Only possible values have a chance
   - If something can't happen, its probability is 0.
4. Add up probabilities if events can’t happen together
   - If two events don’t happen at the same time, their combined chance is just adding them.
   - Example: Chance of getting 2 or 4 on a die = chance of 2 + chance of 4.
5. Expected value (average outcome)
   - It's the average result you'd expect if you repeated the experiment many times.
   - Like the average score you’d expect when rolling a die.
6. Spread or variation (variance)
   - Shows how far the results are from the average.
   - A small spread = results are close to the average.
7. Standard deviation
   - Just the square root of the spread.
   - Also tells how far results are from the average, but in the same units.
8. PMF (for counting outcomes)
   - Used when outcomes are separate values (like 1, 2, 3 on a die).
   - Shows the exact chance for each value.
9. PDF (for measuring outcomes)
   - Used when outcomes are continuous (like height or weight).
   - You look at areas under a curve to find probabilities.
10. CDF (total chance up to a point)
    - Shows the total chance of getting a value less than or equal to a number.
    - It always increases and ends at 1.

Discrete Distributions

A Discrete Distribution is a statistical method used to describe the likelihood (probability) of outcomes of a discrete random variable.

A discrete probability distribution is a list of possible outcomes of a discrete random variable, along with their associated probabilities. Each probability is between 0 and 1, and the sum of all probabilities is 1.

What is a Discrete Random Variable?

- A random variable is a variable whose value is the result of a random experiment.
- A discrete random variable can only take a finite or countable number of values.

Examples:

- Number of students in a room
- Number of heads in 5 coin tosses
- Number of emails received in an hour
- Number of cars in a parking lot

Key Features of Discrete Distributions

- Values are countable (like 0, 1, 2, 3, …)
- Probabilities are assigned to each possible value.
- Total probability adds up to 1.
- Often visualized with a bar graph (not a smooth curve).
- Used for situations where outcomes happen in whole numbers (not fractions or decimals).

Probability Mass Function (PMF)

For discrete distributions, we use a **PMF** instead of a PDF (which is for continuous variables).

A Probability Mass Function (PMF) gives the probability that a discrete random variable is exactly equal to some value.

Example:
Let X = number of heads in 2-coin tosses
Then:

- $P(X = 0) = 0.25$
- $P(X = 1) = 0.5$
- $P(X = 2) = 0.25$

Sum = 1

Bernoulli Distribution

- Models a single experiment or trial that has only two possible outcomes — typically labeled as "success" and "failure".
- When you're interested in just one yes/no event — like flipping a coin once, or checking if a light bulb works.

Parameters:

- p: Probability of success
- 1−p: Probability of failure

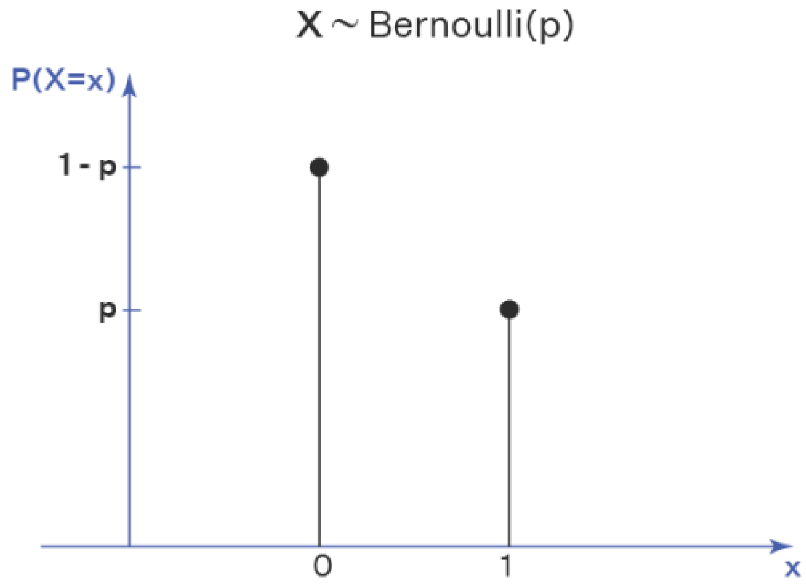
Probability Formula:

$$P(X=x)=\begin{cases} p & \text{if } x=1 \text{ (success)} \\ 1-p & \text{if } x=0 \text{ (failure)} \end{cases}$$

Example: Toss a coin once.

Let "Head" be 1 (success), and "Tail" be 0 (failure). If p=0.5, then:

- P(X=1)=0.5
- P(X=0)=0.5



Binomial Distribution

- Models the number of successes in a fixed number of independent Bernoulli trials.
- Useful when the same experiment is repeated multiple times, like tossing a coin 10 times or conducting 20 quality checks.

Parameters:

- n: Number of trials
- p: Probability of success in each trial

Probability Formula:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

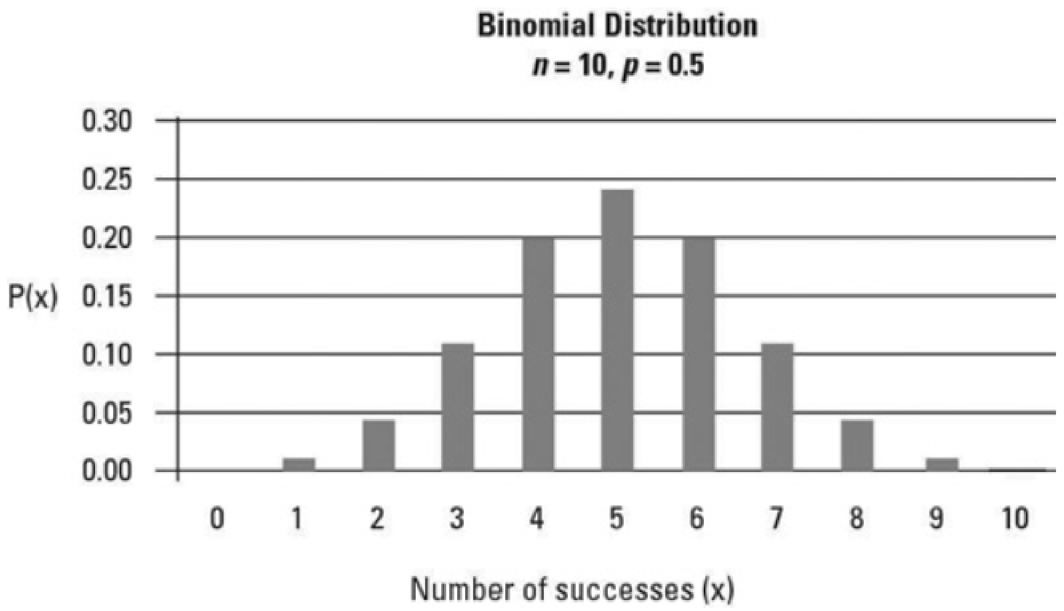
Example:

Toss a fair coin 10 times.

What is the probability of getting exactly 4 heads?

Let n=10, p=0.5, k=4.

$$P(X = 4) = \binom{10}{4} (0.5)^4 (0.5)^6$$



Poisson Distribution

- Models the number of times an event occurs in a fixed interval of time, distance, area, or volume — especially when these events are rare and random.
- Counting events like phone calls per hour, accidents per day, or customer arrivals.

Parameters:

- λ: Average number of occurrences in a fixed interva

Probability Formula:

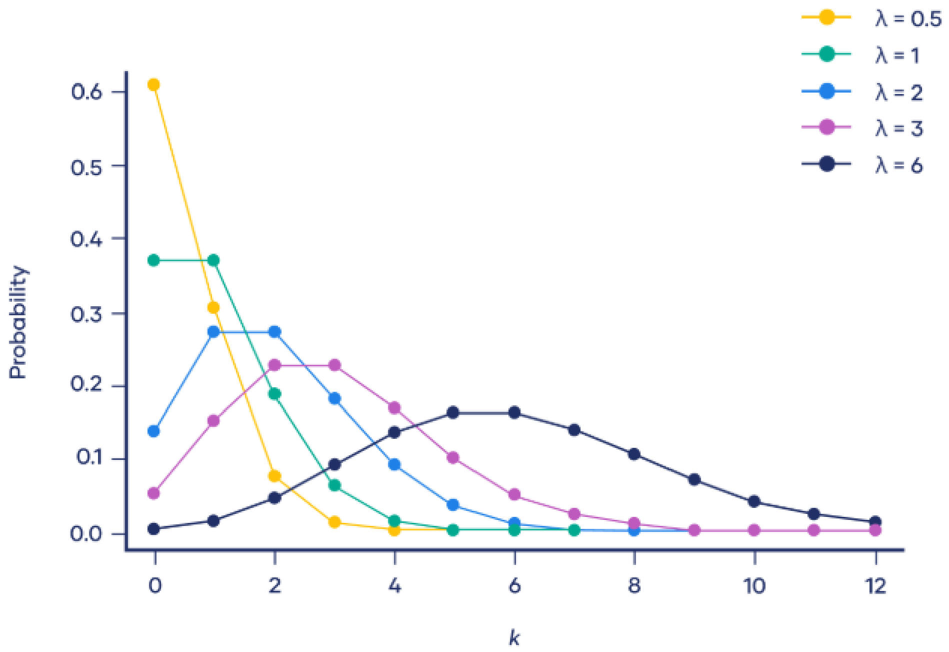
$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots$$

Example:

On average, 3 complaints are received per hour at a help desk. What's the probability of receiving exactly 2 complaints in an hour?

Let λ=3, k=2

$$P(X = 2) = \frac{3^2 e^{-3}}{2!}$$



Geometric Distribution

- Models the number of trials until the first success occurs.
- When you repeat an experiment until you succeed for the first time — like flipping a coin until you get a head.

Parameters:

- p: Probability of success in each trial

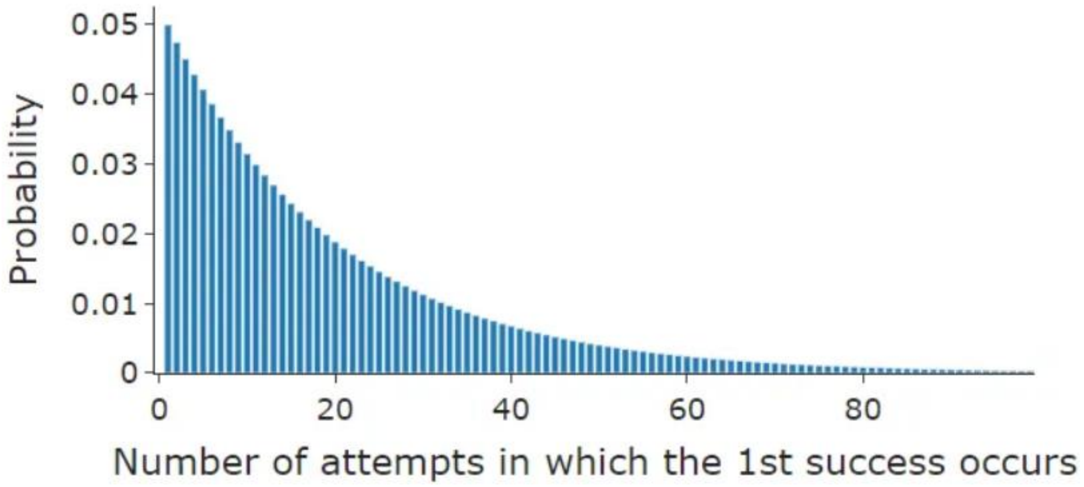
Probability Formula:

$$P(X=k) = (1-p)^{k-1}p, \quad k = 1, 2, 3, \ldots$$

Example:

Toss a fair coin until the first head appears.
What's the probability it takes exactly 3 tosses?
Let p=0.5, k=3

$$P(X=3) = (1-0.5)^2 \cdot 0.5 = 0.125$$



Properties of Discrete Distributions

Distribution	Values of X	PMF (Probability Mass Function)	CDF (Cumulative Distribution Function)	Mean (μ)	Variance (σ^2)
Bernoulli(p)	$x \in \{0,1\}$	$P(X=x)=p^x(1-p)^{1-x}$	$F(x)=\begin{cases}0 & x<0\\ 1-p & 0\le x<1\\ 1 & x\ge 1\end{cases}$	$\mu=p$	$\sigma^2=p(1-p)$
Binomial(n, p)	$x \in \{0,1,\ldots,n\}$	$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}$	$F(x)=\sum_{k=0}^{\lfloor x\rfloor}\binom{n}{k}p^k(1-p)^{n-k}$	$\mu=np$	$\sigma^2=np(1-p)$
Poisson(λ)	$x \in \{0,1,2,\ldots\}$	$P(X=x)=\frac{\lambda^x e^{-\lambda}}{x!}$	$F(x)=\sum_{k=0}^{\lfloor x\rfloor}\frac{\lambda^k e^{-\lambda}}{k!}$	$\mu=\lambda$	$\sigma^2=\lambda$
Geometric(p)	$x \in \{1,2,3,\ldots\}$	$P(X=x)=(1-p)^{x-1}p$	$F(x)=1-(1-p)^x$	$\mu=\frac{1}{p}$	$\sigma^2=\frac{1-p}{p^2}$
Uniform(n)	$x \in \{1,2,\ldots,n\}$	$P(X=x)=\frac{1}{n}$	$F(x)=\frac{\lfloor x\rfloor}{n}$, for $1\le x\le n$	$\mu=\frac{n+1}{2}$	$\sigma^2=\frac{n^2-1}{12}$

Continuous Distributions

A continuous distribution describes the probability of outcomes over a continuous range of values.
Unlike discrete distributions (which deal with specific counts like 0, 1, 2...), continuous distributions work with measurements — like height, weight, time, or temperature — which can take any value within a range, even decimals or fractions.

Situation	Type of Data	Continuous?
Weight of a watermelon	2.35 kg	Yes
Time to cook rice	14.5 minutes	Yes
Number of eggs in a basket	10 eggs	No (This is discrete)
Distance traveled by a car	120.8 km	Yes

Key Concept of Continuous Distribution

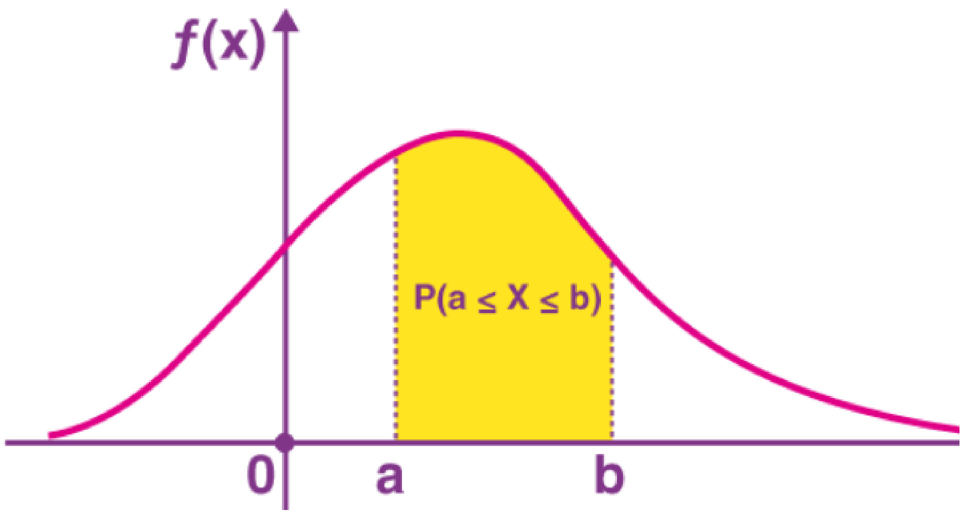
In a continuous distribution, the probability of getting one exact value is zero because there are infinitely many possible values. So, we do not calculate $P(X=x)$.

Instead, we calculate the probability over a range, like: $\boldsymbol{P(a \le X \le b)}$
This gives a meaningful (non-zero) probability.

Probability Density Function (PDF)

A Probability Density Function (PDF) is a mathematical function that shows how the probability is distributed over a continuous range of values.
It helps us understand:
Where the values are more likely to occur in a continuous distribution.

Feature	**Explanation**
Non-negative	PDF values $f(x) \geq 0$ for all x
Total Area = 1	The total probability over all possible values is 1
Area = Probability	The area under the curve between two points gives the probability
No spikes	It's a smooth curve, not bars like in discrete cases



Uniform Distribution

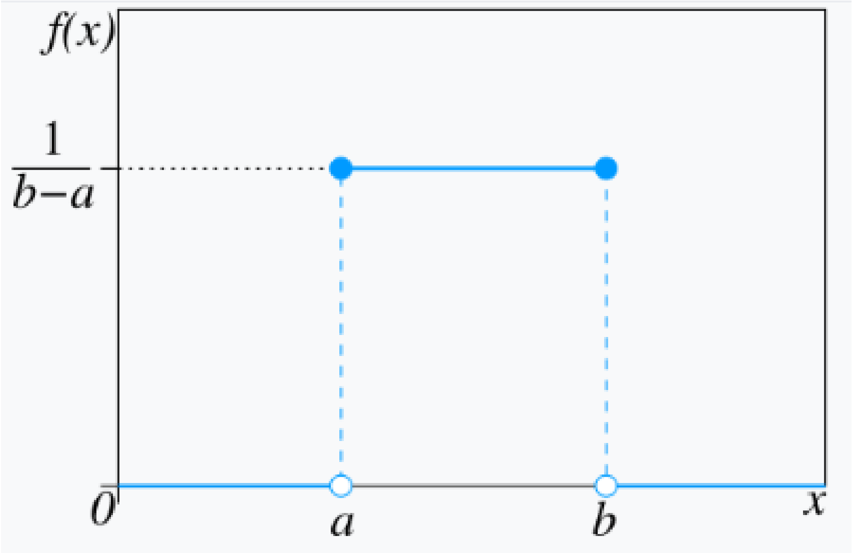
Models a situation where every value in a given range is equally likely to occur.
Think of it as choosing a random value from a range — like picking any number between 1 and 10,

Parameters:

- a: Lower limit
- b: Upper limit

Probability Formula:

$$P(x) = \frac{1}{b-a}$$



Example:
Pick a random number between 2 and 8.
What's the probability the number lies between 3 and 5?
Let a=2, b=8
Range of interest = 5 – 3 = 2

$$P(3 \leq X \leq 5) = \frac{2}{8-2} = \frac{2}{6} = 0.333$$

Normal Distribution

The Normal Distribution is a way to describe data where most values are close to the average, and fewer values are very high or very low. It looks like a smooth hill or bell shape when you draw it. This kind of distribution happens a lot in real life — like people's heights, test marks, or weights.

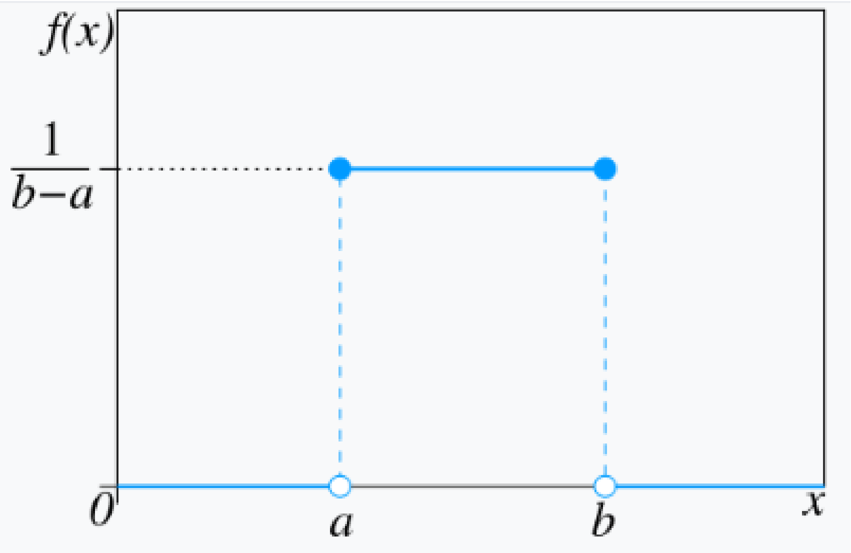
Most people are average, and only a few are much shorter or taller, or score very high or very low. It is controlled by two things: the mean (which shows the center) and the standard deviation (which shows how spread out the values are).

Parameters:

- μ: Mean (center of the distribution)
- σ: Standard deviation (spread or width)

Probability Formula:

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



Exponential Distribution

Models the time between events in a process where events occur continuously and independently at a constant average rate.

Think of it like measuring how long you have to wait:

•for the next bus to arrive

•for the next customer to walk into a shop

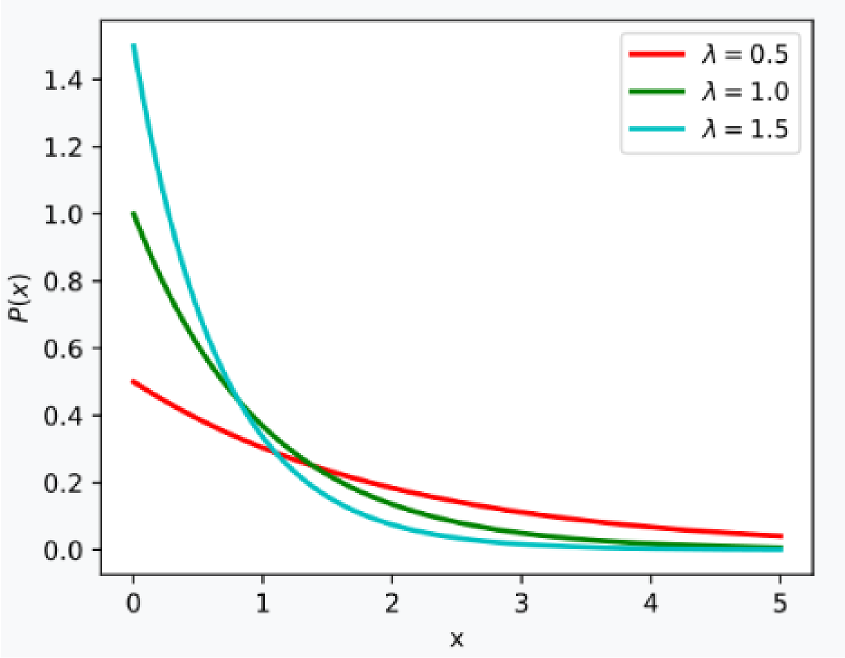
•for the next call in a call center

Parameters:

- λ: Rate (events per unit time)

Probability Formula:

$$f(x;\lambda) = \lambda e^{-\lambda x}, \quad \text{for } x \geq 0$$



Example:

You run a coffee shop. On average, 3 customers arrive per hour.

So,λ=3

What's the probability that the next customer will arrive within 10 minutes?

Convert time: 10 minutes = 1/6 hours

We need: $P(X \leq \frac{1}{6}) = 1 - e^{-\lambda x} = 1 - e^{-3\cdot\frac{1}{6}} = 1 - e^{-0.5}$

P≈1−0.6065=0.3935P

Gamma Distribution

Models the waiting time until *k* events occur, assuming events happen independently and at a constant rate.
Think of it as an extension of the Exponential Distribution:

- Exponential: Time until 1 event
- Gamma: Time until k events

Used when you're waiting for multiple events, like:

- Time until a machine breaks down for the third time
- Time until receiving 5 customer calls

Parameters:

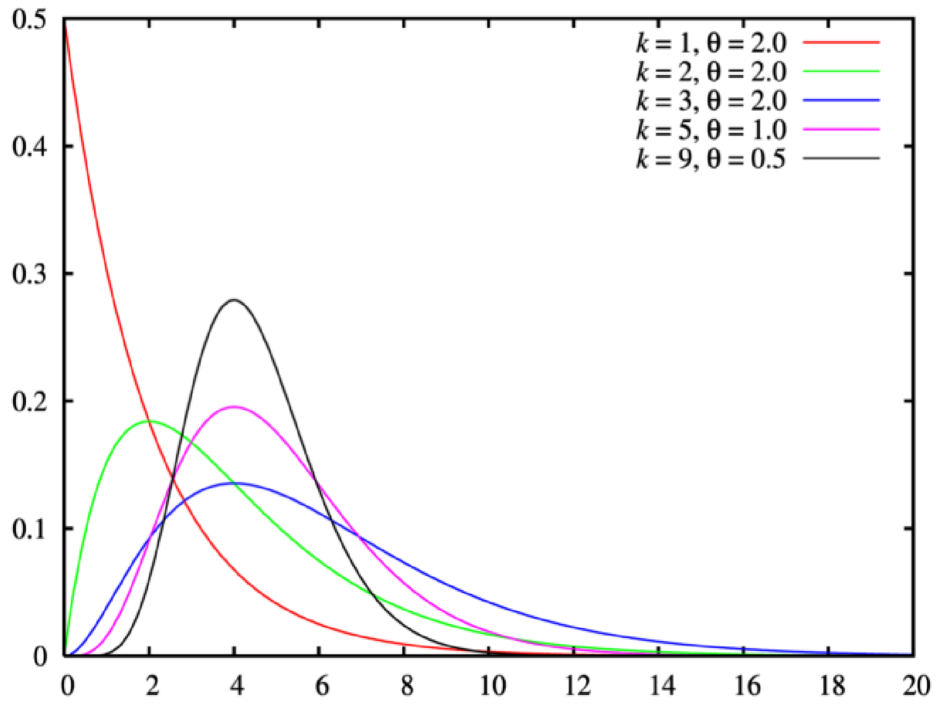
- k (shape): Number of events to wait for (also written as α)
- θ (scale): Average time between events
  (can also use rate λ=1/θ)

Probability Formula:

$$f(x;k,\theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta}, \quad \text{for } x > 0$$

Where $\Gamma(k)$ is the gamma function

(For integers, $\Gamma(k) = (k-1)!$)



Beta Distribution

The Beta distribution is used when we want to model a probability that is between 0 and 1.
It tells us:
How likely is it that the true probability is close to a certain value?
Where is it used?
When you are not sure about a probability
For example:
What's the chance a person clicks a button on a website?
What's the chance it will rain tomorrow?
Parameters:
- α: Number of successes (how many times something happened)
- β: Number of failures (how many times it didn't happen)

Probability Formula:

$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, \quad \text{for } 0 < x < 1$$

Where B(α,β) is the **Beta function**, a normalization constant.

Chi-Square Distribution(χ^2)

Models the sum of squared values of independent standard normal variables.
If you square several values from a normal distribution and add them, the result follows a Chi-Square distribution.
The Chi-Square distribution helps us check if things match what we expect.
It is mostly used in statistics when we want to answer questions like:

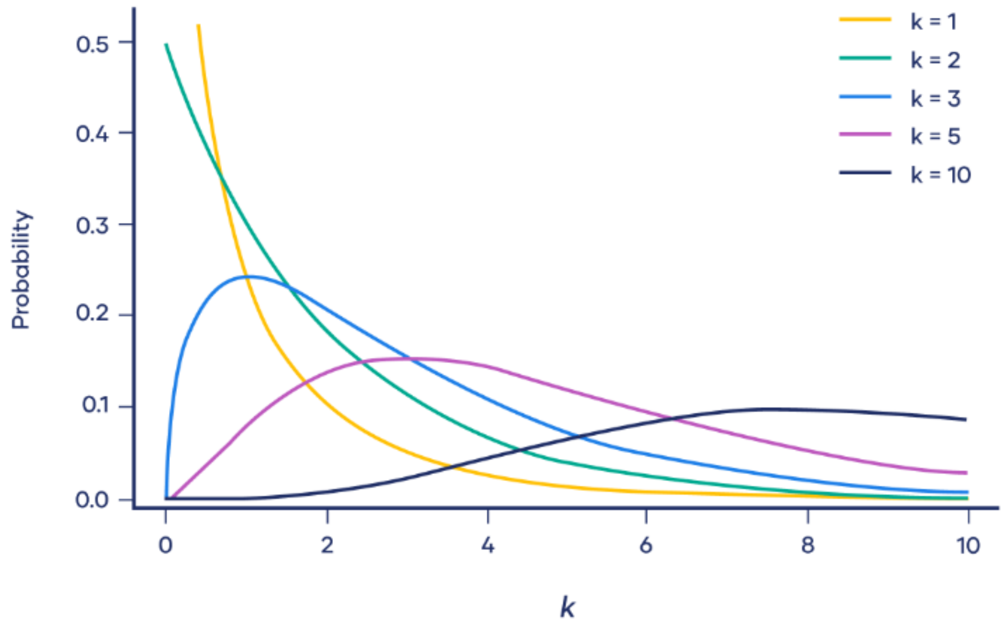
- "Do these categories occur equally often?"
- "Is there a relationship between two things?"
- "Is the variation too much or normal?"

- **Parameters**:

k: Degrees of freedom (df)
(The number of independent squared values)

Probability Formula:

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



Student-t Distribution

Used when estimating the mean of a population based on a small sample size, especially when population standard deviation is unknown.

Think of it like this:

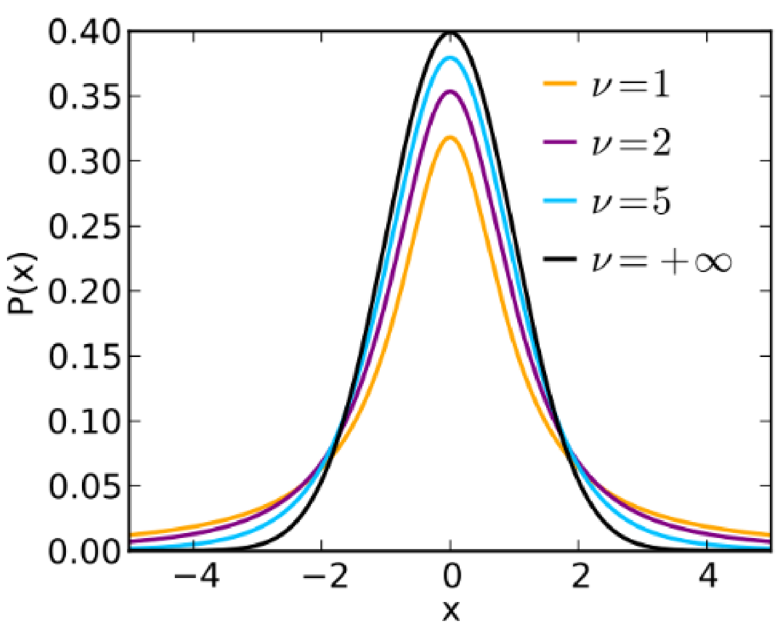
When you're working with small samples, you can't trust the normal distribution — the t-distribution accounts for that extra uncertainty.

When to use it?

•Sample size is small ($n < 30$)

•You are comparing averages

•You don't know the true population standard deviation



•**Parameters**:

ν: Degrees of freedom (df)

Usually, $\nu = n - 1$, where n is the sample size

Probability Formula:

$$f(t;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Properties of Continuous Distributions

Distribution	Values of X	PDF (Probability Density Function)	CDF (Cumulative Distribution Function)	Mean (μ)	Variance (σ^2)
Uniform(a, b)	$x \in [a, b]$	$f(x) = \frac{1}{b-a}$	$F(x) = \frac{x-a}{b-a},\ a \le x \le b$	$\mu = \frac{a+b}{2}$	$\sigma^2 = \frac{(b-a)^2}{12}$
Normal(μ, σ^2)	$x \in (-\infty, \infty)$	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	No simple form; use standard normal table or erf function	μ	σ^2
Exponential(λ)	$x \ge 0$	$f(x) = \lambda e^{-\lambda x}$	$F(x) = 1 - e^{-\lambda x}$	$\mu = \frac{1}{\lambda}$	$\sigma^2 = \frac{1}{\lambda^2}$
Gamma(k, θ)	$x \ge 0$	$f(x) = \frac{x^{k-1} e^{-x/\theta}}{\theta^k \Gamma(k)}$	No simple closed form	$\mu = k\theta$	$\sigma^2 = k\theta^2$
Beta(α, β)	$x \in [0, 1]$	$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$	No simple closed form	$\mu = \frac{\alpha}{\alpha+\beta}$	$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Chi-Square(k)	$x \ge 0$	$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2}$	No simple closed form	$\mu = k$	$\sigma^2 = 2k$
Student t(ν)	$x \in (-\infty, \infty)$	$f(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}$	No simple closed form	$\mu = 0$ (for $\nu > 1$)	$\sigma^2 = \frac{\nu}{\nu-2}$ (for $\nu > 2$)