Types of data distribution and types of density function

Types of data distribution:

There are two main types of data distribution in statistics: **continuous** and **discrete.**

Discrete Distribution Types:

- **Binomial distribution**
  The binomial distribution describes the probability of a certain number of successes (or failures) in a given number of trials or events. This type of distribution is used when there are only two possible outcomes for each trial, such as success or failure, heads or tails, yes or no etc., with equal probabilities for each product. The binomial distribution can be used to calculate the likelihood of achieving a specific result from multiple independent trials that have only two possible effects.
- **Poisson distribution**
  The Poisson distribution describes the probability that an event will occur within a fixed time period when its rate is known but its exact timing cannot be predicted accurately enough to measure it directly. This type of distribution is useful for modelling random occurrences such as customer arrivals at stores, phone calls received by call centers etc., where the average rate of occurrence is known but the exact timing cannot be measured.
- **Hypergeometric distribution**
  The hypergeometric distribution describes the probability of a certain number of successes (or failures) in a given number of draws from an urn or population when the draws are made without replacement. This type of data distribution is used in situations where an urn contains different items, such as colored balls, and you want to evaluate the probability that a certain number of those items will be drawn out with each draw being made without replacing the item that was just taken out.
- **Geometric distribution**
  The geometric distribution describes the probability of a success occurring on any given trial in a series of independent trials when the probability of success

for each trial is known. This type of data distribution can be used to model the number of failures that occur before a success in situations such as manufacturing processes, where there are multiple attempts at creating a product and each attempt has a given probability of success.

Continuous Distribution Types:

- **Normal Distribution**

  Normal Distributions are one of the most commonly used data distributions. This distribution measures data points in a bell-shaped curve, with an equal number of data points to the left and right of the mean value. Normal Distributions can be used to predict future outcomes based on past trends.

  This type of distribution is often used when analysing test scores or financial data**. It is the most common type of data distribution:**

  - **Flat or Uniform distribution** is a type of data distribution where the data is evenly distributed throughout the entire range. This type of distribution is often used when analysing things like wait times or manufacturing processes.
  - **A bimodal distribution** is when data has two peaks or more. This type of distribution often occurs when two different types of data are being measured. For example, you might have a bimodal distribution if you measured the number of people who liked and disliked a particular product.
  - **Non-Symmetric data** has a shape with a long tail. This type of data often occurs when there are outliers in the data set. Outliers are values far from the rest of the data set. Or when the data cannot be below zero, such as waiting times, costs etc.
  - **Skewed data** is similar to Non-Symmetric. This type of data often occurs when there are outliers in the data set. Outliers are values far from the rest of the data set.

- **Lognormal distribution**

  Lognormal distributions measure data points in a curve shaped like a sigmoid function – a curved line beginning at zero and then increasing sharply to a peak and slowly decreasing. This data distribution is often **used in financial data**, allowing us to extrapolate potential future stock prices based on past data.

- **F distribution**

  F Distributions measure data points spread out over a broader range than normal distributions. This data distribution is often used when measuring data with higher variability, such as performance data or customer satisfaction surveys.

- **Chi-Square distribution**

  Chi-Square Distributions measure the difference between observed data and expected results. This data distribution can be used to identify significant differences between two data sets and help us understand which factors may be influencing our results.

- **Exponential distribution**

  Exponential distributions measure data points with an exponential curve – a curve beginning at zero and gradually increasing in value. This data distribution is often used when data points are expected to increase over time, such as population data or customer data in a given market.

- **T Student distribution**

  T Student Distributions measure data points spread out more than normal distributions. This data distribution can be used for data sets with higher variability and outliers, such as performance data.

- **Weibull Distribution**

  Weibull Distributions measure data in an exponential curve – a curve beginning at zero and gradually increasing in value. This data distribution is often used for reliability tests and can help us predict how long it will take for a system to fail.

- **Non-normal distributions**

  Non-normal distributions include data distributions such as the Poisson Distribution, Gamma Distribution, Beta Distribution, Logistic Distribution and

Cauchy Distribution. Non-normal data distributions are often used when data does not fit into the normal data distribution categories, such as highly non-linear or data with outliers.

Types of density function:

Density functions are functions that describe how the proportion of data or likelihood of the proportion of observations change over the range of the distribution.

Types of density functions:

- **Probability density functions:**
  - Calculates the probability of observing a given value.
  - A probability density function, or PDF, can be used to calculate the likelihood of a given observation in a distribution. It can also be used to summarize the likelihood of observations across the distribution's sample space. Plots of the PDF show the familiar shape of a distribution, such as the bell-curve for the Gaussian distribution.
- **Cumulative density functions**:
  - Calculates the probability of an observation equal or less than a value.
  - A cumulative density function, or CDF, is a different way of thinking about the likelihood of observed values. Rather than calculating the likelihood of a given observation as with the PDF, the CDF calculates the cumulative likelihood for the observation and all prior observations in the sample space. It allows you to quickly understand and comment on how much of the distribution lies before and after a given value. A CDF is often plotted as a curve from 0 to 1 for the distribution.