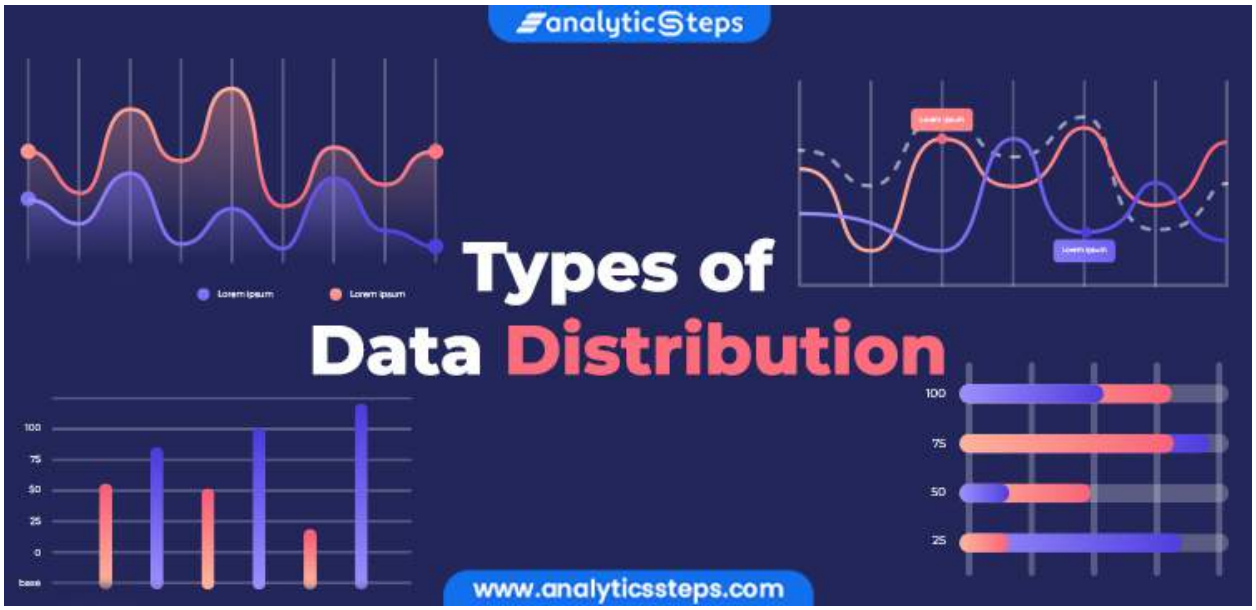
MSP task

Types of data distribution



Bernoulli's Distribution

This is one of the simplest distributions that can be used as an initial point to derive more complex distributions. Bernoulli's distribution has possibly two outcomes (success or failure) and a single trial.

For example, tossing a coin, the success probability of an outcome to head is p, then the probability of having tail as the outcome is (1-p). Bernoulli's distribution is the special case of binomial distribution with a single trial.

Binomial Distribution

The binomial distribution is applied in binary outcomes events where the probability of success is equal to the probability of failure in all the successive trials. Its example includes tossing a biased/unbiased coin for a repeated number of times.

As input, the distribution considers two parameters, and is thus called as bi-parametric distribution. The two parameters are;

- The number of times an event occurs, n, and
- Assigned probability, p, to one of the two classes

For n number of trials, and success probability, p, the probability of successful event (x) within n trials can be determined by the following formula

Normal (Gaussian) Distribution

Being a continuous distribution, the normal distribution is most commonly used in data science. A very common process of our day to day life belongs to this distribution- income distribution, average employees report, average weight of a population, etc.

Where μ = Mean value,

σ = Standard probability distribution of probability,

x = random variable

According to the formula, the distribution is said to be normal if mean (μ) = 0 and standard deviation (σ) = 1

The graph of normal distribution is shown below which is symmetric about the centre (mean).

Poisson Distribution

Being a part of discrete probability distribution, poisson distribution outlines the probability for a given number of events that take place in a fixed time period or space, or particularized intervals such as distance, area, volume.

For example, conducting risk analysis by the insurance/banking industry, anticipating the number of car accidents in a particular time interval and in a specific area.

Poisson distribution considers following assumptions;

- The success probability for a short span is equal to success probability for a long period of time.
- The success probability in a duration equals to zero as the duration becomes smaller.
- A successful event can’t impact the result of another successful event

Exponential Distribution

TLike the poisson distribution, exponential distribution has the time element; it gives the probability of a time duration before an event takes place.

Exponential distribution is used for survival analysis, for example, life of an air conditioner, expected life of a machine,and length of time between metro arrivals.

A variable X is said to possess an exponential distribution when

Multinomial Distribution

The multinomial distribution is used to measure the outcomes of experiments that have two or more variables. It is the special type of binomial distribution when there are two possible outcomes such as true/false or success/failure.

A very popular Mendel experiment where two strains of peas (one green and wrinkled seeds and other is yellow and smooth seeds) are hybridized that produced four different strains of seeds-green and wrinkled, green and round, yellow and round, and yellow and wrinkled. This resulted in multinomial distribution and led to the discovery of the basic principles of gen

The following are properties of multinomial distribution;

- An experiment can have a repeated number of trials, for example, rolling of a dice multiple times.
- Each trial is independent of each other.
- The success probability of each outcome must be the same (constant) for all trials of an experiment.

Beta Distribution

Beta distribution comes under continuous probability distributions having the interval [0,1] with two shape parameters that can be expressed by alpha (α) and beta( ). These two parameters are the exponent of a random variable and control the shape of the distribution.

The distribution shows the family of probabilities and is a suitable model to depict random behaviour of percentages or proportions. It is used for the data models that hold uncertainties of the success probabilities in a random experiment.

The general formulation of beta distribution is also known as the beta distribution of first kind and beta distribution of second kind is another name of beta prime distribution.

Beta distribution has many applications in statistical description of allele frequencies in genetic population, time allocation in project management, sunshine data, proportions of minerals in rocks, etc.

Beta-binomial distribution

A data distribution is said to be beta-binomial if the

- Probability of success, p, is greater than zero.
- And, shape of beat binomial parameter, $\alpha > 0$, as well as $\beta > 0$

Being the simplest form of Bayesian mode, beta-binomial distribution has extensive applications in intelligence testing, epidemiology, and marketing.

The graph of beta-binomial distribution looks as below;

Talking about the key difference amid a beta-distribution and binomial distribution, the success probability, p, is always fixed for a set of trials whereas it is not fixed for beta-binomial distribution and changes trail to trail.

T- Distributions

- Similar to normal distribution, the t-distribution has bell-shaped curve distribution and is symmetric when mean is zero.
- The shape of distribution doesn't alter with degrees of freedom, and has the range $-\infty$ to ∞.
- The variance is always more than one.
- As the sample size, n, increases, t-distribution acts as normal distribution where the considered sample size is greater than 30.

Uniform distribution

Uniform distribution can either be discrete or continuous where each event is equally likely to occur. It has a constant probability constructing a rectangular distribution.

In this type of distribution, an unlimited number of outcomes will be possible and all the events have the same probability, similar to Bernoulli's distribution. For example, while rolling a dice, the outcomes are 1 to 6 that have equal probabilities of ⅙ and represent a uniform distribution.

A variable X is said to have uniform distribution if the probability density function is

The uniform distribution has the following properties;

- The probability density function combines to unity.
- Every input function has an equal weightage.