STATISTICS ON DISTRIBUTION

Statistics is the fascinating science of collecting, analyzing, interpreting, presenting, and organizing data. One of the foundational concepts in this field is 'distribution.

But what exactly does distribution in statistics mean, and why is it so important?

At the heart of statistics, distribution is simply a way that values of a particular variable or set of data are spread out.

Whether you're studying heights in a population, rainfalls in a year, or the spread of a viral pandemic, distributions can help you understand the data better.

Let's dive deeper into each type and explore their peculiarities. Reading a distribution goes beyond just recognizing its shape. It involves understanding key characteristics such as the mean, median, mode, range, skewness, and kurtosis.

These factors can tell us a lot about our data, from where most values lie to how they deviate from the norm.

Types Of Distributions

- May it be Weather forecasting, banking or finance, Statistics plays a crucial role. Though it appears to be complex. It is the most interesting subject I could say.
- Before getting into the types of distribution. It is essential to know the definition of a distribution. It represents the happening of possibility of the variable and Its frequency.
- Example. Obtaining the number 2 in a dice has a probability of 1/6. Obtaining a number 20 in a dice has a probability of 0. In a nutshell. Probability ranges from zero to 1.

Continuous Distribution

1.Normal Distribution

2.Uniform Distribution

3.Exponential Distribution

4.Chi-Square Distribution

5.Beta Distribution

6.Gamma Distribution

7.Log-normal Distribution

Discrete Distribution

1.Bernoulli Distribution

2.Discrete Uniform Distribution

3.Binomial Distribution

4.Poisson Distribution

5.Geometric Distribution

Continuous Distribution: Represents data that can take on any value within a specific range, and probabilities are represented using a probability density function (PDF).

Discrete Distribution: Represents data that can only take specific, isolated values, and probabilities are represented as individual probabilities of those values using a probability mass function (PMF).

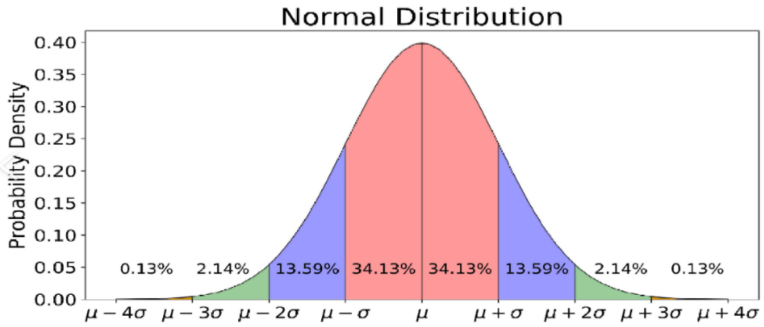
NORMAL DISTRIBUTION:

It is also called the standard normal distribution or the bell curve.

In most real-world scenarios, we follow the normal distribution.

Properties of a normal distribution

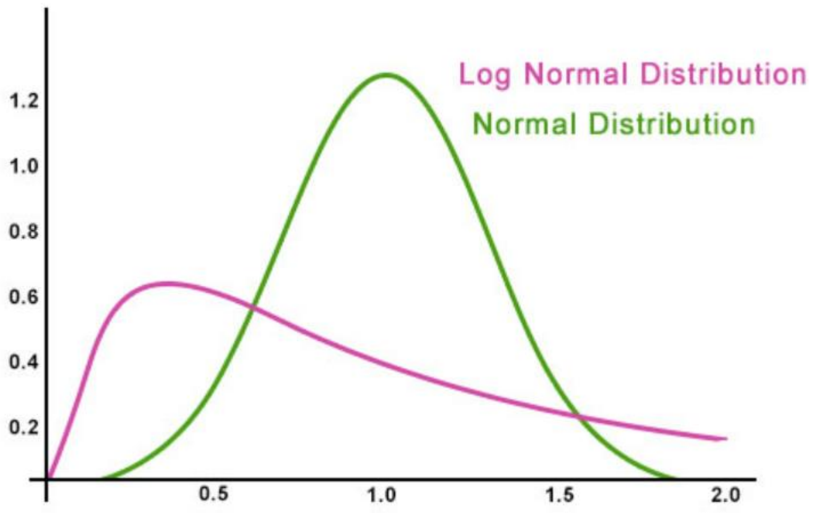
- The curve remains symmetric at the center.
- The area under the curve is 1.
- The mean, median, and mode are always equal.
- Exactly half value is on the left of the center and the other on the right.



Log-Normal Distribution:

The lognormal distribution has the following properties:

- The distribution is positively skewed, meaning it has a long right tail.
- The distribution is defined only for positive values since the logarithm of zero or negative values is undefined.
- The mean and variance of the lognormal distribution depend on the parameters of the underlying normal distribution of the algorithm.

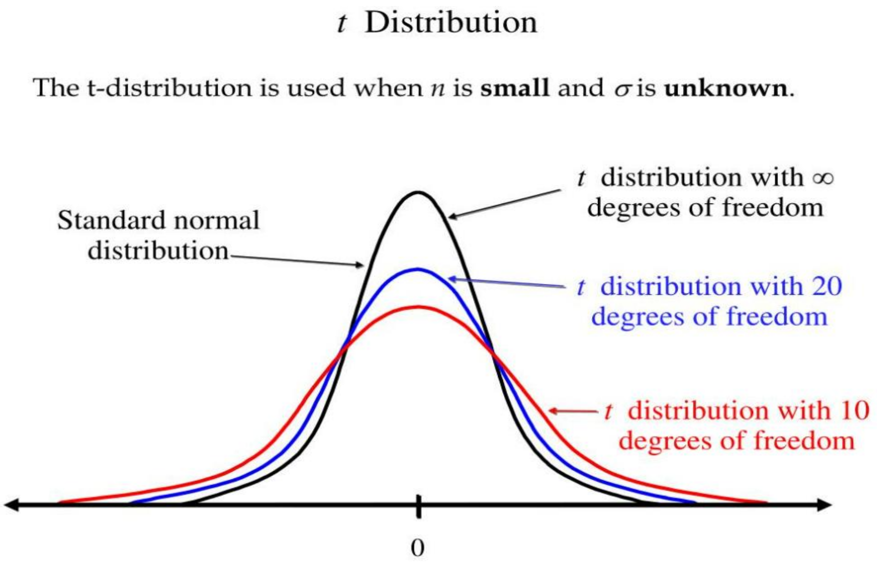


T DISTRIBUTION:

T distribution is also called student's distribution.

Why the name student's distribution?

Statistician Gusset published his work under the pseudonym "Student" due to company policy that prohibited employees from publishing research under their own names.



Degrees Of Freedom:

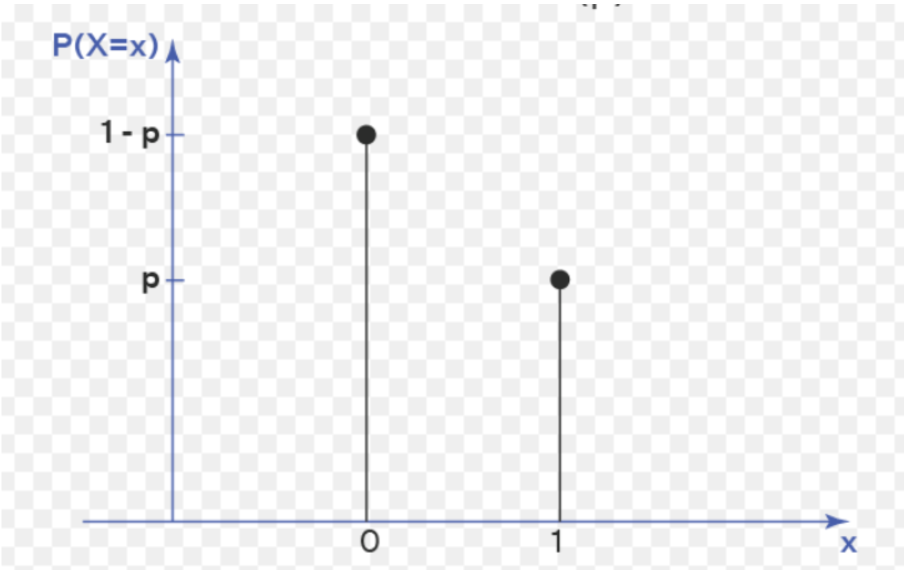
The formal definition of degrees of freedom could be the number of independent pieces of information used to calculate the statistic is called the degrees of freedom. In simpler words, the number of values in the final calculation that are free to vary is your degree of freedom.

Generally, it is no. Of samples 1

Bernoulli distribution:

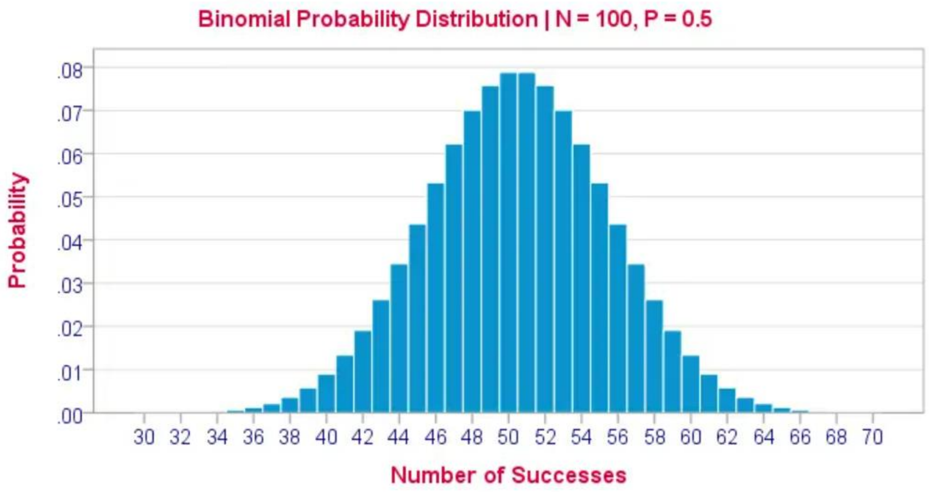
This is a discrete probability. If the probability does not change from one trial to the next trial, then it is known as Bernoulli distribution.

To have more clarity, when we throw a die, we get the probability of getting a 1 as 1/6. When we throw the die 1000 times still the probability is 1/6. This forms a Bernoulli distribution.



Binomial Distribution:

It is often used to model scenarios where there are two possible outcomes (success or failure) with a fixed probability of success. The mean of a binomial distribution is given by μ = np, variance = npq and the standard deviation is given by σ = √(npq) where n is the no. Of trials and p is the success probability in each trial.



Prob = nCx (p^x)(q^n-x)

NCx= n!/(x!)(n-x)!

What makes binomial distribution different from Poisson distribution?

The Poisson distribution models the number of events occurring in a fixed interval of time or space when the events occur independently at a constant average rate whereas the binomial distribution models the number of successes in a fixed number of independent Bernoulli trials.

Conditions for a distribution to be Binomial distribution:

1. Observations should be independent.
2. The probability of success should be constant at all n trials.
3. Mu> variance: which means the difference between the success (p)and failure(q) should be less or the success probability(p) should be more than q whereas in Poisson p is in decimals

   .

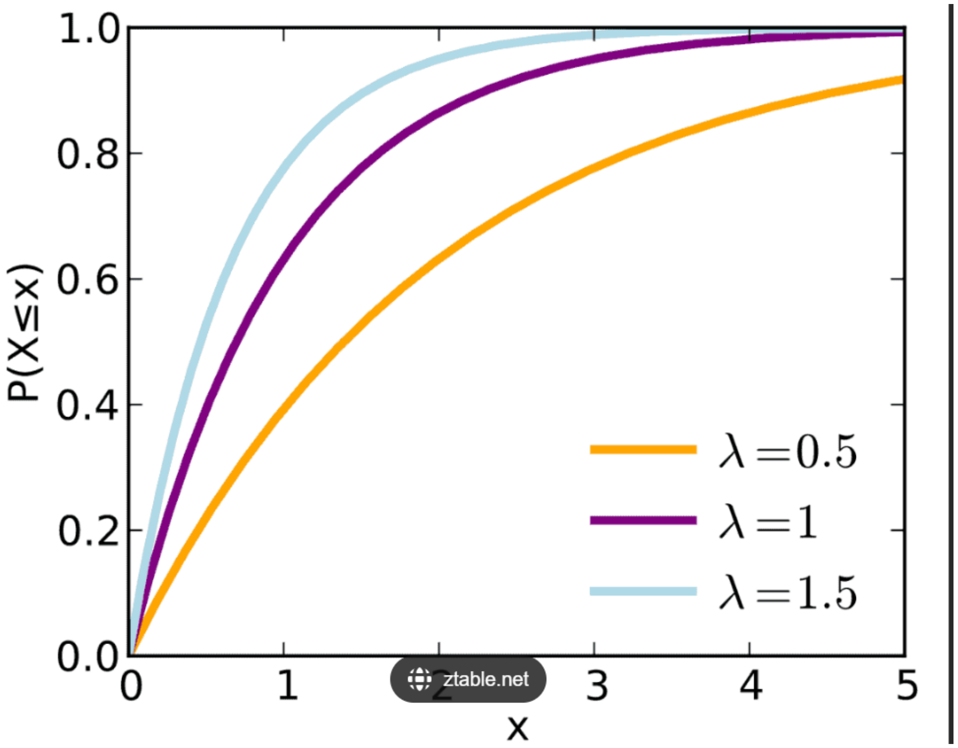
Poisson Distribution:

If X is a random variable and ranges from 0 to infinity, and lambda be the Poisson distribution parameter, then.

P(X=x) = e^-lambda(lambda)^x/x factorial

Exponential distribution:

The exponential distribution is often concerned with the amount of time until some specific event occurs. For example, the amount of time (beginning now) until an earthquake occurs has an exponential distribution. Other examples include the length, in minutes, of long-distance business telephone calls, and the amount of time, in months, a car battery lasts.



The exponential distribution is a probability distribution function that is commonly used to measure the expected time for an event to happen.

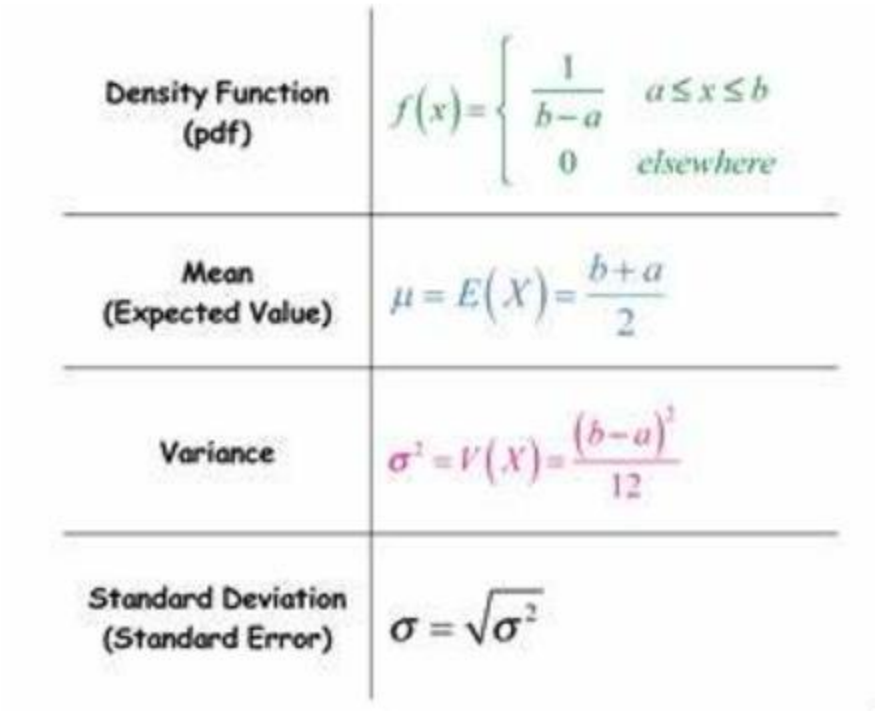
The lambda in exponential distribution represents the rate parameter, and it defines the mean number of events in an interval.

Difference between Poisson and Exponential distribution:

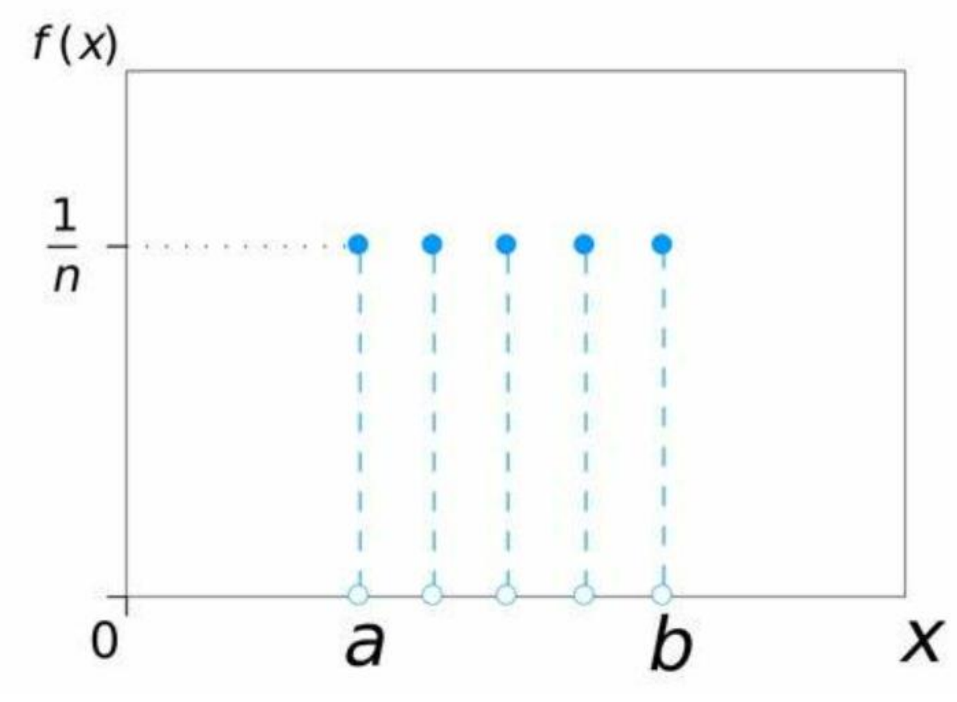
Poisson distribution deals with the number of occurrences of events in a fixed period of time, whereas the exponential distribution is a continuous probability distribution that often concerns the amount of time until some specific event happens.

Uniform Distribution:

A continuous probability distribution is a **Uniform Distribution** and is related to the events which are equally likely to occur. It is defined by two parameters, x and y, where x = minimum value and y = maximum value. It is generally denoted by u (x, y).

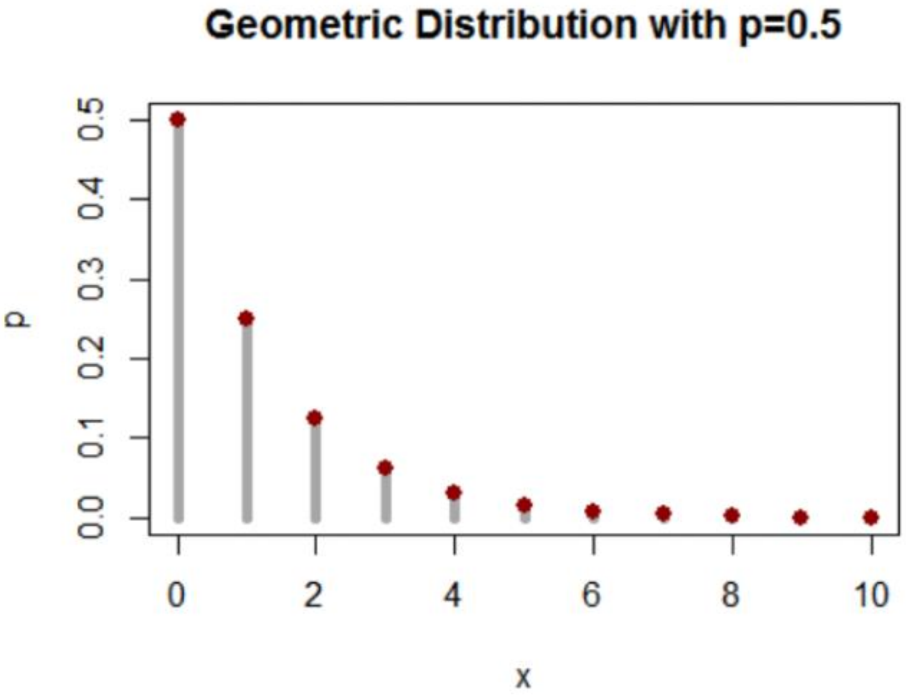


Density Function (pdf)	$f(x)=\begin{cases}\frac{1}{b-a} & a\le x\le b\\ 0 & elsewhere\end{cases}$
Mean (Expected Value)	$\mu = E(X)=\frac{b+a}{2}$
Variance	$\sigma^2 = V(X)=\frac{(b-a)^2}{12}$
Standard Deviation (Standard Error)	$\sigma = \sqrt{\sigma^2}$



Geometric Distribution:

$$f\left(x;p\right) = pq^{x}; x = 0, 1, 2..., \infty$$



- The probability distribution of the number X of Bernoulli trials needed to get one success, supported on the set.
- The probability distribution of the number $Y = X - 1$ of failures before the first success, supported on the set {0,1, 2,…..}.
- Geometric distribution makes it simple to depict the likelihood of how many times a coin must be thrown before it lands on its head. Other examples are Number of Network Failures, Number of Bugs in a Code, Number of Faulty Products Manufactured in an Industry etc.

Hypergeometric distribution:

1.A hypergeometric distribution has a specified number of dependent trials having two possible outcomes, success, or failure. The random variable is the number of successful outcomes in the specified number of trials. The Individual outcomes cannot be repeated within these trials.

2. The probability of x successes in r dependent trials is

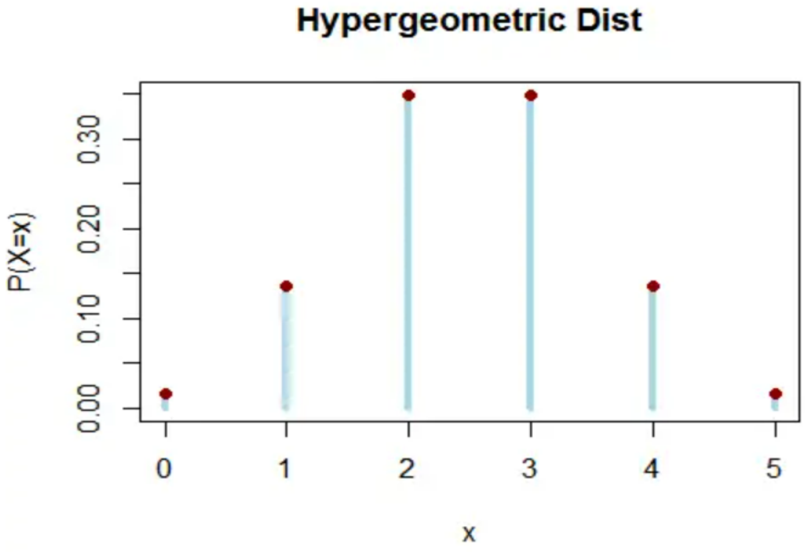
$$P(x) = \frac{{}_{a}C_{x} \times {}_{n-a}C_{r-x}}{{}_{n}C_{r}},$$

where n is the population size and # is the number of successes in the population.

3. The expectation for hypergeometric distribution is.

$$E(X) = \frac{ra}{n}.$$

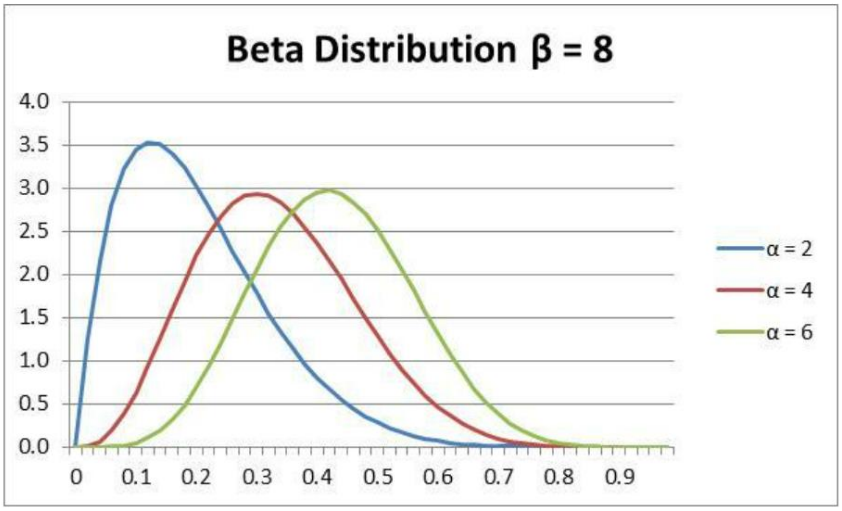
4. To simulate a hypergeometric experiment, ensure that the number of trials is representative of the situation and that each trial is dependent (no replacement or resetting between trials). Record the number of successes and summarize the results by calculating probabilities and expectations.



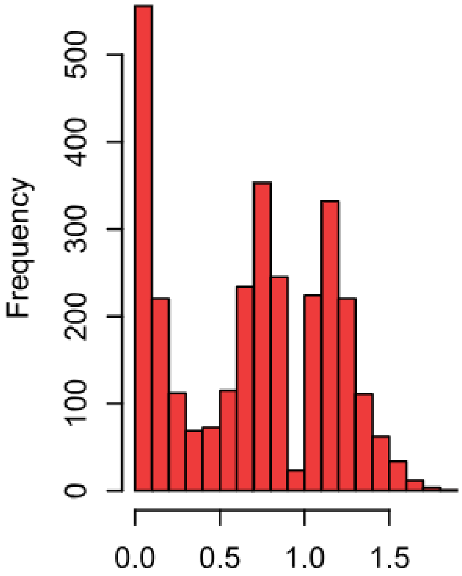
Beta Distribution:

The Beta distribution is **a probability distribution** ***on probabilities***. It is a versatile probability distribution that could be used to model probabilities in different scenarios.

The difference between the binomial and the beta distribution is that **the former models the number of successes (x), while the latter models the probability (p) of success.**



Multimodal distribution:



What do you call a distribution with two modes?

Yes, it’s called a bimodal distribution. A three-mode distribution is called trimodal distribution. A distribution with more than two nodes is called a multi modal distribution. This tells us how the modality is distinct in our dataset.