# Ep 10 - Non Gaussian Distributions

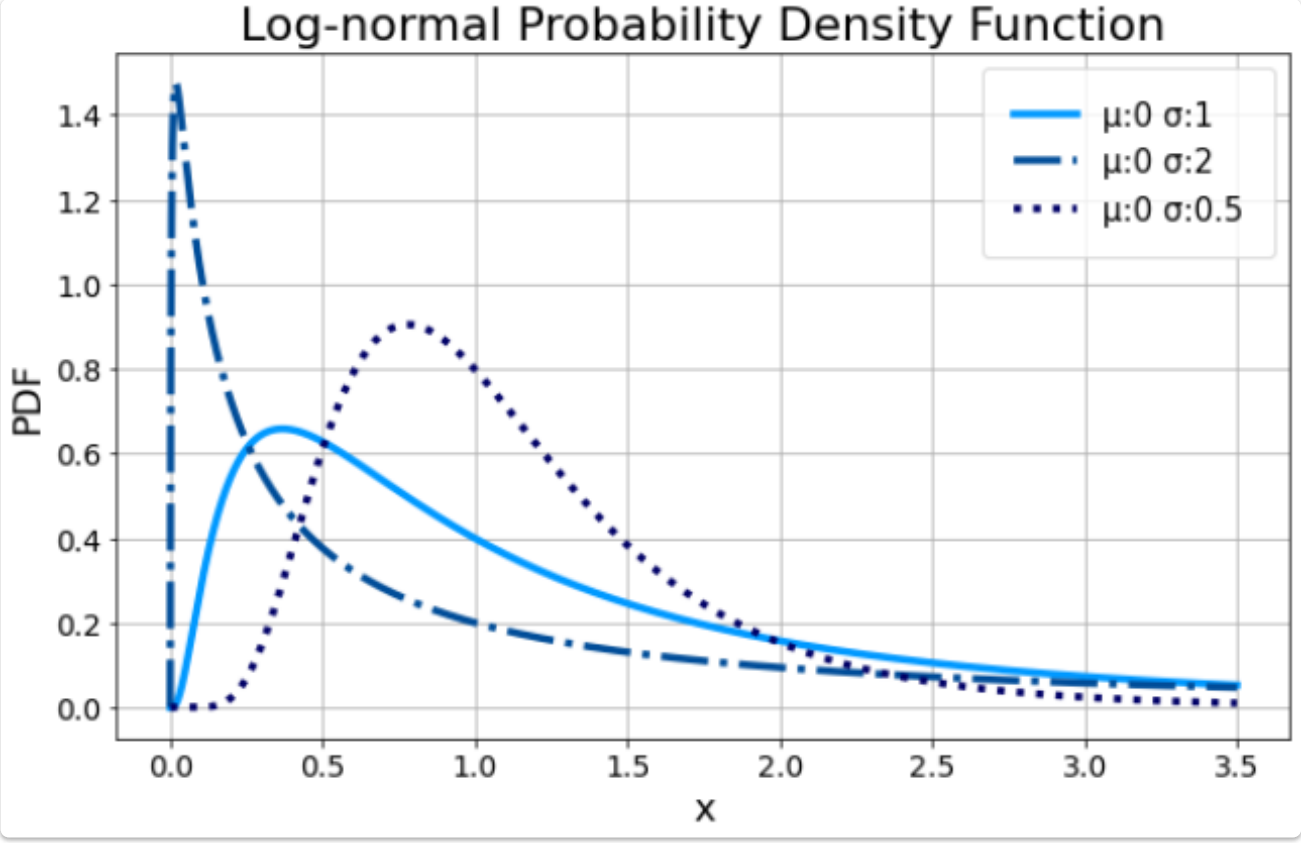
Dec 5,2023

## Student's t Distribution

## Log Normal Distribution

The log - normal distribution is a heavy tailed continuous probability distribution of a random variable whose logarithm is normally distributed.

$$\text{Log-Normal Distribution PDF: } f(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right)$$




### Applications in Data Science and Machine Learning

1. Used to model skewed data, such as income distributions or stock prices, where the logarithm of the data follows a normal distribution.

**Example**

The length of comments posted in Internet discussion forums follows a log-normal distribution.

Users' dwell time on online articles (jokes, news etc.) follows a log-normal distribution.

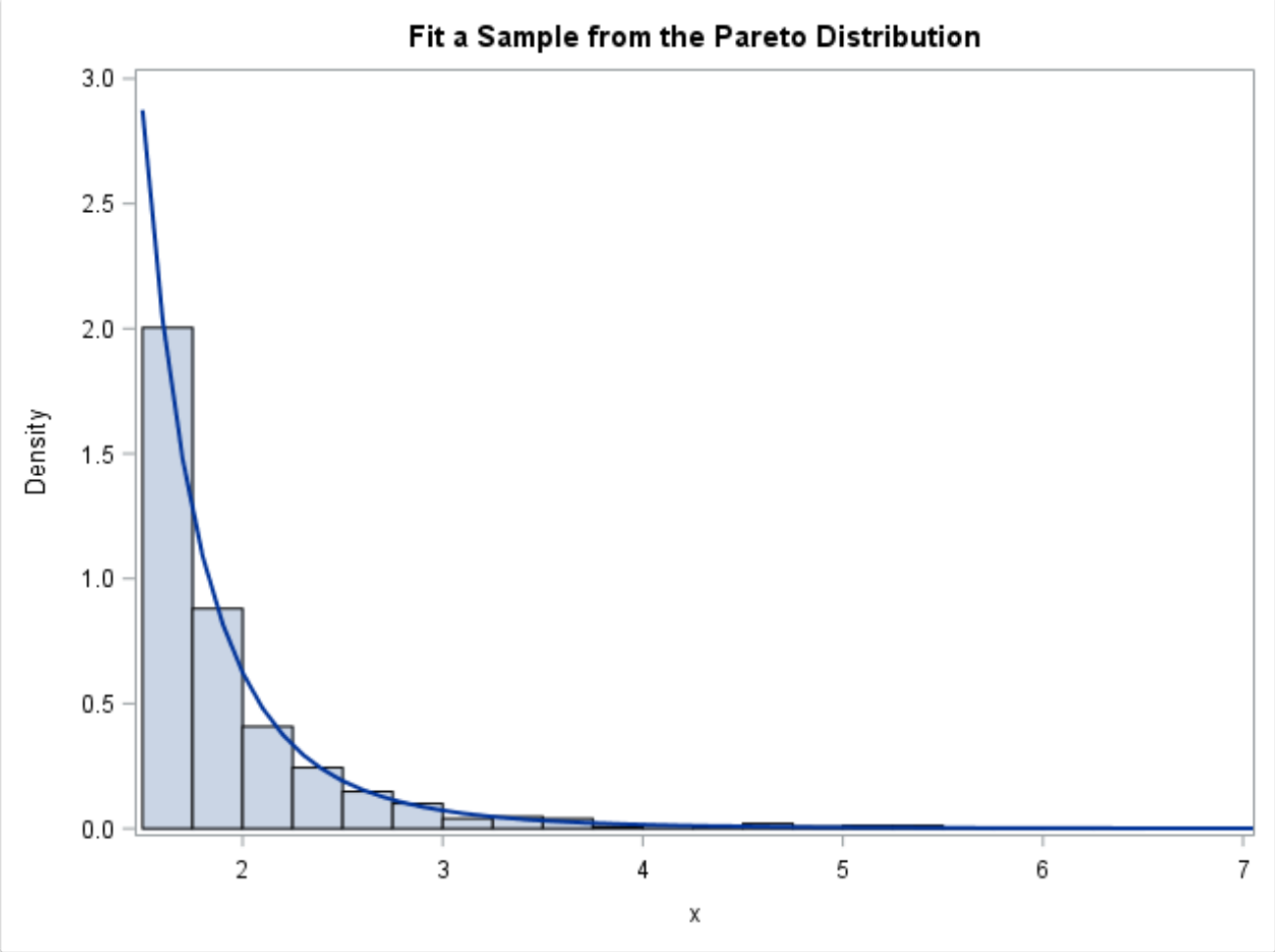
In economics, there is evidence that the income of $97\%$–$99\%$ of the population is distributed log-normally.

The concentration of rare elements in minerals

## Pareto Distribution

The Pareto distribution is a probability distribution that models extreme outcomes where a small number of variables have a disproportionately large impact.

$$\text{Pareto Distribution PDF: } f(x|x_{\mathrm{m}},k) = \frac{kx_{\mathrm{m}}^{k}}{x^{k+1}} \quad \text{for } x \geq x_{\mathrm{m}}$$



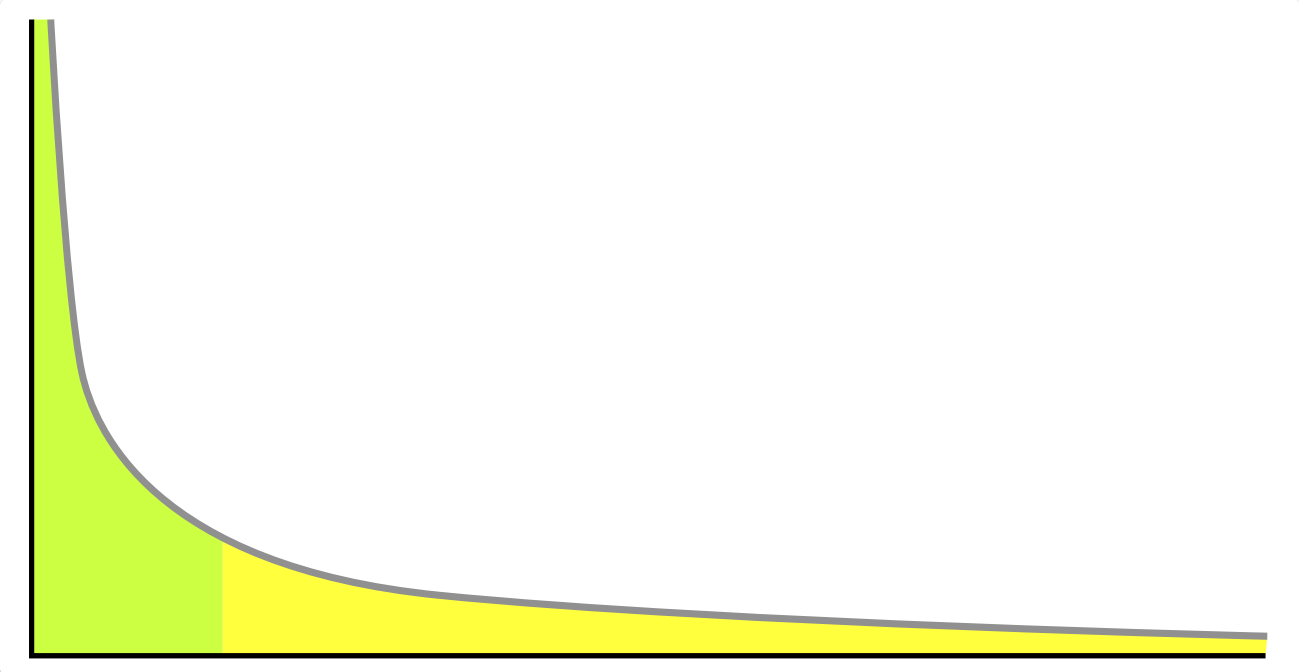
The Pareto distribution is a type of probability distribution that is commonly used to model the distribution of wealth, income, and other quantities that exhibit a similar **power-law behaviour**.

> **Note**
>
> Vilfredo Pareto originally used this distribution to describe the among individuals since it seemed to show rather well the way that a larger portion of the wealth of any society is owned by a smaller percentage of the people in that society. He also used it to describe distribution of income.
> The **Pareto Principle** or the "$80-20$ rule" which says that $20\%$ of the population controls $80\%$ of the wealth.

## Power Law

In mathematics, a power law is a functional relationship between two variables, where one variable is proportional to a power of the other.

$$\mathrm{y} = k * x^{a}$$

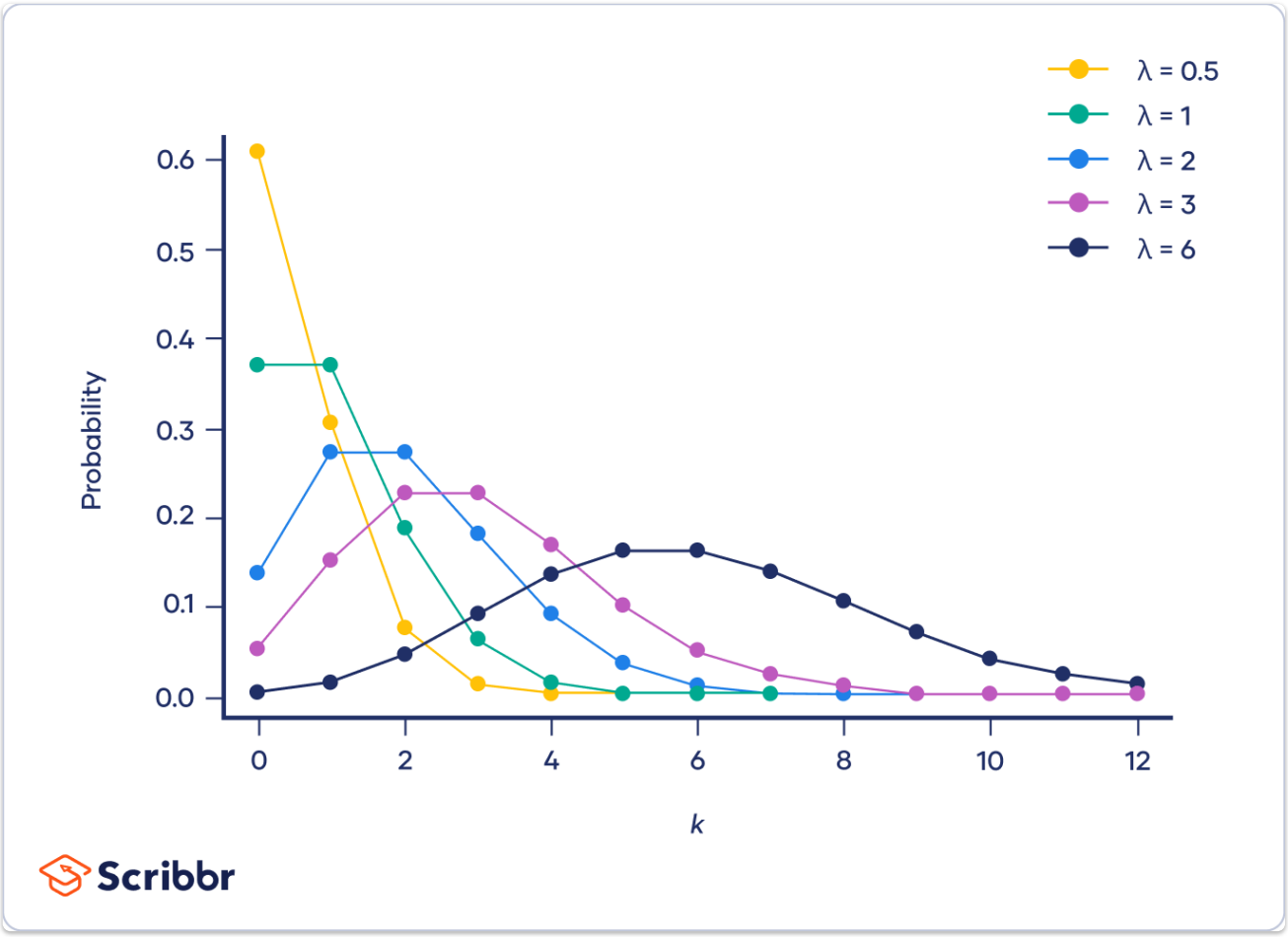


An Explanation of Zipf's Law

> **Example**
>
> The sizes of human settlements (few cities, many hamlets/villages)
>
> The standardized price returns on individual stocks

Amount of time a user on Steam will spend playing different games. (Some games get played a lot, but most get played almost never.)

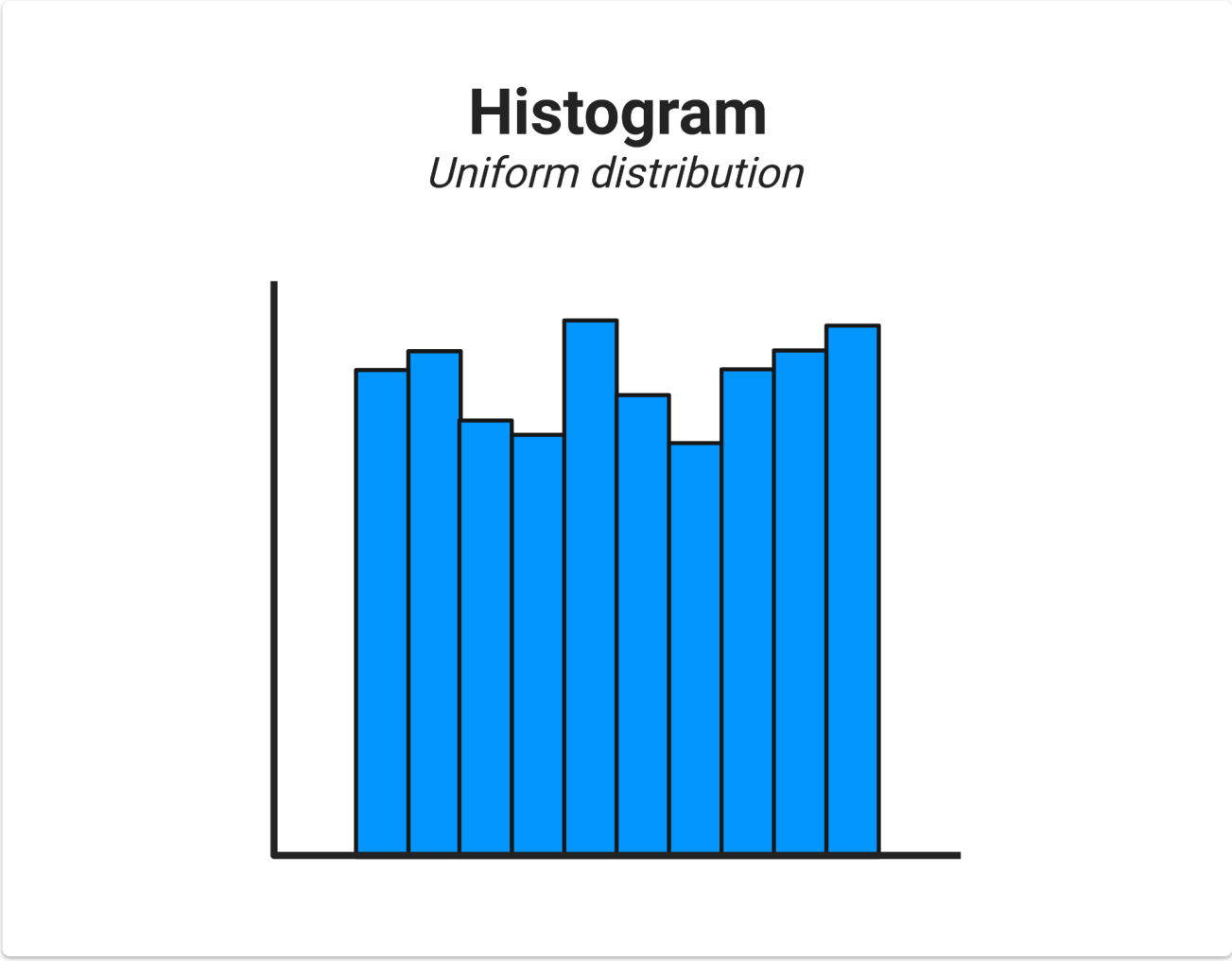
## Poisson Distribution




## Uniform Distribution

The uniform distribution is a probability distribution where all outcomes are equally likely.

$$\text{Uniform Distribution PDF: } f(x|a,b) = u(a,b) = \frac{1}{b-a} \quad \text{for } a \leq x \leq b$$



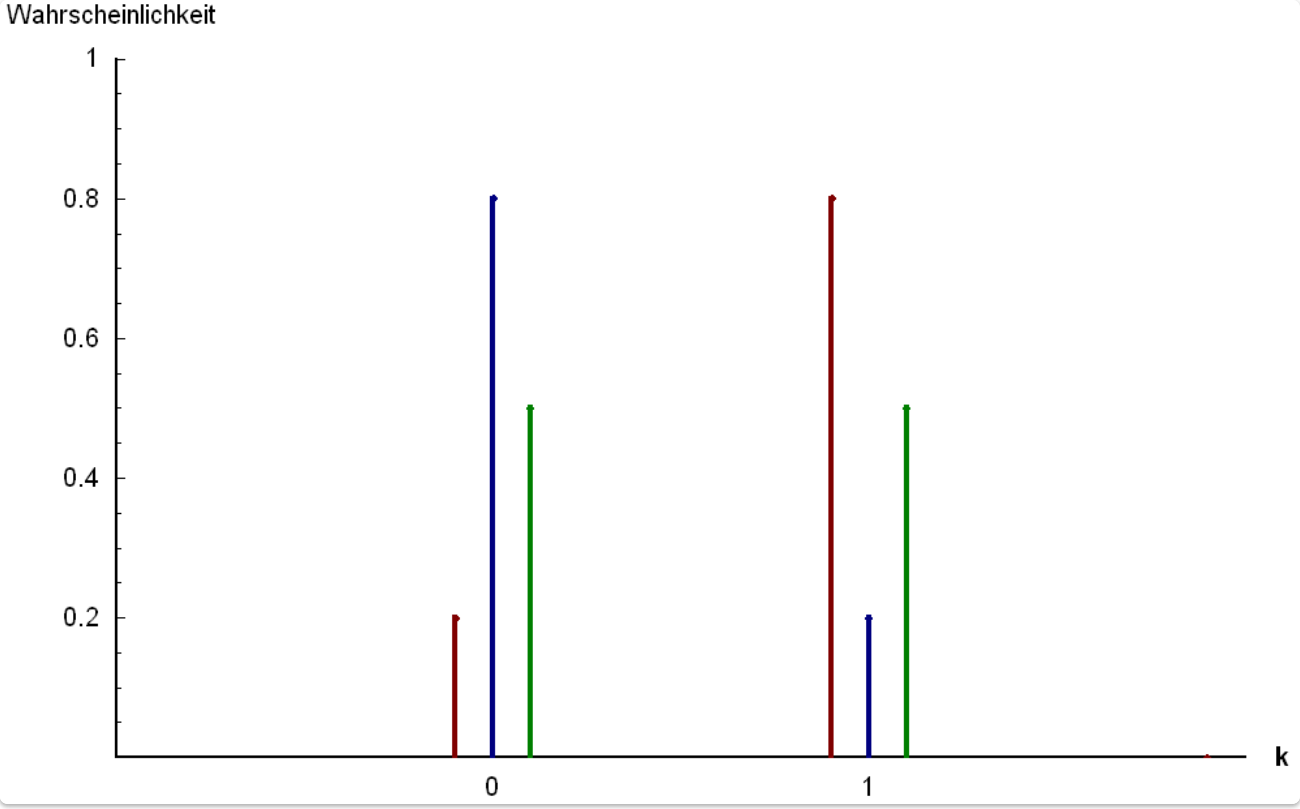

## Applications in Data Science and Machine Learning

1. **Random Initialization**: In many machine learning algorithms, such as neural networks and k-means clustering, the initial values of the parameters can have a significant impact on the final result.
2. **Sampling**
3. **Data Augmentation**: Uniform distribution can be used to generate new data points that are within a specified range of the original data.
4. **Hyper-parameter Tuning**

# Bernoulli Distribution

The Bernoulli distribution is a discrete probability distribution representing a random variable that can take on one of two possible outcomes, usually labelled as success and failure.
The probability mass function (PMF) of a Bernoulli distribution with parameter $p$, where $0 \leq p \leq 1$, is given by:

$$\text{Bernoulli Distribution PMF: } f(x|p) = p^x * (1-p)^{1-x}$$




## Applications in Data Science and Machine Learning

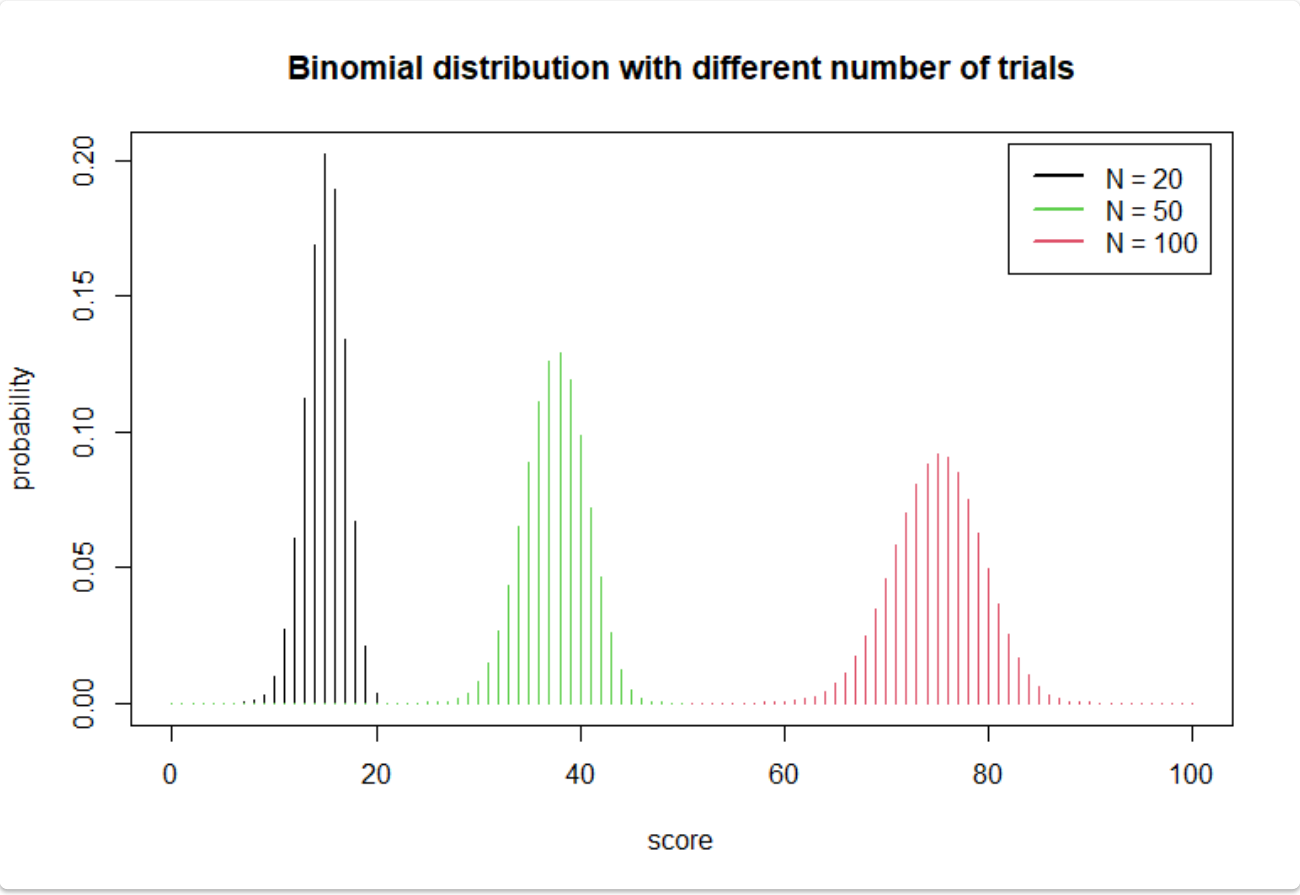
**Binary Classification**: The Bernoulli distribution is commonly used in machine learning for modelling binary outcomes, such as whether a customer will make a purchase or not, whether an email is spam or not, or whether a patient will have a certain disease or not.

# Binomial Distribution

The binomial distribution is a discrete probability distribution that describes the number of successes (usually denoted as $k$) in a fixed number of *independent and identically distributed Bernoulli trials*.
Each trial has *only two possible outcomes*, typically labelled as success (with probability $p$) and failure (with probability $1 - p$).

The probability mass function (PMF) of a binomial distribution with parameters $n$ (number of trials) and $p$ (probability of success on each trial) is given by:

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




## Criteria

1. The process consists of n trials
2. Only 2 exclusive outcomes are possible, a success and a failure.
3. $P$(success) = $p$ and $P$(failure) = $1 - p$ and it is fixed from trial to trial
4. The trials are independent.

## Applications in Data Science and Machine Learning

1. **Binary Classification**
2. **Hypothesis Testing**
3. **Logistic Regression**
4. **A/B Testing**

Binomial Distribution and Test
A Very Beautiful Video on Binomial Distribution