**What is a probability distribution?**

A probability distribution is an idealized [frequency distribution](https://www.scribbr.com/statistics/frequency-distributions/).

A **frequency distribution** describes a specific [sample](https://www.scribbr.com/methodology/population-vs-sample/) or dataset. It’s the number of times each possible value of a variable occurs in the dataset.

The number of times a value occurs in a sample is determined by its **probability** of occurrence. Probability is a number between 0 and 1 that says how likely something is to occur:

* 0 means it’s impossible.
* 1 means it’s certain.

The higher the probability of a value, the higher its frequency in a sample.

More specifically, the probability of a value is its relative frequency in an infinitely large sample.

Infinitely large samples are impossible in real life, so probability distributions are theoretical. They’re idealized versions of frequency distributions that aim to describe the [population](https://www.scribbr.com/methodology/population-vs-sample/) the sample was drawn from.

Probability distributions are used to describe the populations of real-life variables, like coin tosses or the weight of chicken eggs. They’re also used in [hypothesis testing](https://www.scribbr.com/statistics/hypothesis-testing/) to determine [*p* values](https://www.scribbr.com/statistics/p-value/).

**There are two types of probability distributions:**

* [Discrete probability distributions](https://www.scribbr.com/statistics/probability-distributions/#discrete)
* [Continuous probability distributions](https://www.scribbr.com/statistics/probability-distributions/#continuous)

## types of distribution functions:

Based on the types of data we deal with, we have two types of distribution functions.

For discrete data, we have discrete distributions; and for continuous data, we have continuous distributions.

|  |  |
| --- | --- |
| Discrete distributions | Continuous distributions |
| **Uniform distribution** | **Normal distribution** |
| **Binomial distribution** | **Standard Normal distribution** |
| **Bernoulli distribution** | **Student’s T distribution** |
| **Poisson distribution** | **Chi-squared  distribution** |

Before deep-diving into the types of distributions, it is important to revise the fundamental concepts like **Probability Density Function** (PDF), **Probability Mass Function** (PMF), and **Cumulative Density Function** (CDF).

## Discrete probability distributions

A **discrete probability distribution** is a probability distribution of a [categorical or discrete variable](https://www.scribbr.com/methodology/types-of-variables/#quantitative-vs-categorical).

Discrete probability distributions only include the probabilities of values that are possible. In other words, a discrete probability distribution doesn’t include any values with a probability of zero. For example, a probability distribution of dice rolls doesn’t include 2.5 since it’s not a possible outcome of dice rolls.

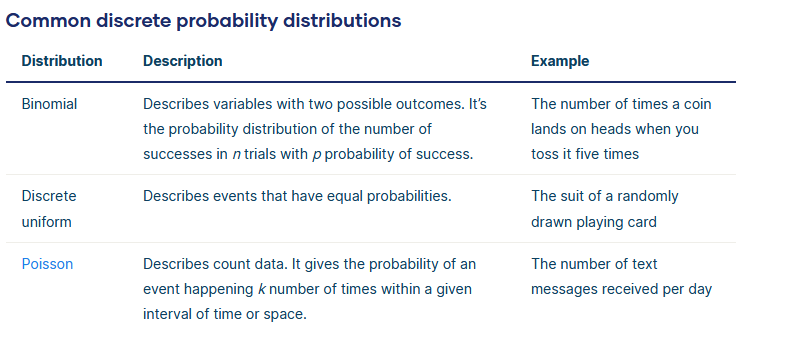
The probability of all possible values in a discrete probability distribution add up to one. It’s certain (i.e., a probability of one) that an observation will have one of the possible values.

### Probability tables

A **probability table** represents the discrete probability distribution of a [categorical variable](https://www.scribbr.com/methodology/types-of-variables/). Probability tables can also represent a discrete variable with only a few possible values or a continuous variable that’s been [grouped into class intervals](https://www.scribbr.com/statistics/frequency-distributions/#grouped).

A probability table is composed of two columns:

* The values or class intervals
* Their probabilities



## Continuous probability distributions

A **continuous probability distribution** is the probability distribution of a [continuous variable](https://www.scribbr.com/methodology/types-of-variables/).

A continuous variable can have any value between its lowest and highest values. Therefore, continuous probability distributions include every number in the variable’s [range](https://www.scribbr.com/statistics/range/).

The probability that a continuous variable will have any specific value is so infinitesimally small that it’s considered to have a probability of zero. However, the probability that a value will fall within a certain interval of values within its range is greater than zero.

### Probability density functions

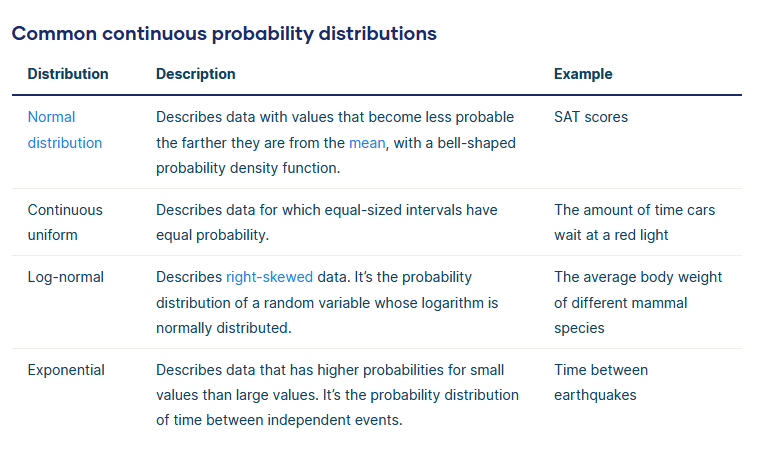
A **probability density function** (PDF) is a mathematical function that describes a continuous probability distribution. It provides the **probability density** of each value of a variable, which can be greater than one.

A probability density function can be represented as an equation or as a graph.

In graph form, a probability density function is a curve. You can determine the probability that a value will fall within a certain interval by calculating the area under the curve within that interval. You can use reference tables or software to calculate the area.

The area under the whole curve is always exactly one because it’s certain (i.e., a probability of one) that an observation will fall somewhere in the variable’s range.

A **cumulative distribution function** is another type of function that describes a continuous probability distribution.



### Exponential Distribution

Let’s consider the call center example one more time. What about the interval of time between the calls? Here, the exponential distribution comes to our rescue. Exponential distribution models the interval of time between the calls.

Other examples are:

1. Length of time between metro arrivals
2. Length of time between arrivals at a gas station
3. The life of an air conditioner

The exponential distribution is widely used for survival analysis. From the expected life of a machine to the expected life of a human, exponential distribution successfully delivers the result.

A random variable X is said to have an **exponential distribution** with PDF:

f(x) = { λe-λx,  x ≥ 0

And parameter **λ>0,** which is also called the rate.

For survival analysis, λ is called the failure rate of a device at any time t, given that it has survived up to t.

Mean and Variance of a random variable X following an exponential distribution:

* Mean -> E(X) = 1/λ
* Variance -> Var(X) = (1/λ)²

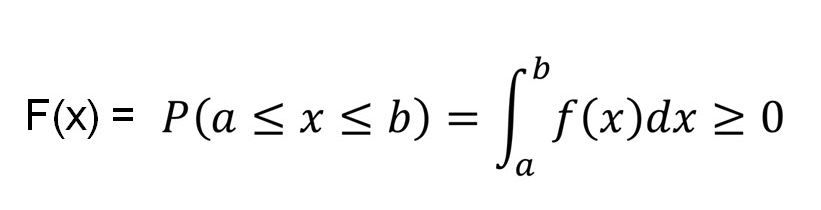
Also, the greater the rate, the faster the curve drops, and the lower the rate, the flatter the curve. This is explained better with the graph shown below.

To ease the computation, there are some formulas given below:

* P{X≤x} = 1 – e-λx corresponds to the area under the density curve to the left of x
* P{X>x} = e-λx corresponds to the area under the density curve to the right of x
* P{x1<X≤ x2} = e-λx1 – e-λx2, corresponds to the area under the density curve between x1 and x2.

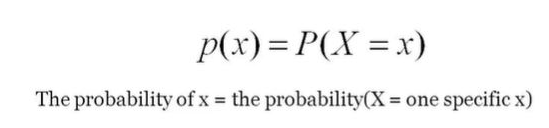
**the probability density function (PDF)**

the probability density function (PDF) of a continuous random variable serves as a function that interprets the relative likelihood of the random variable matching a given sample within the dataset or sample space. The PDF represents the probability per unit length. Essentially, it allows us to gauge the higher likelihood of the random variable being near one sample compared to another by comparing the values of the PDF at these two samples.



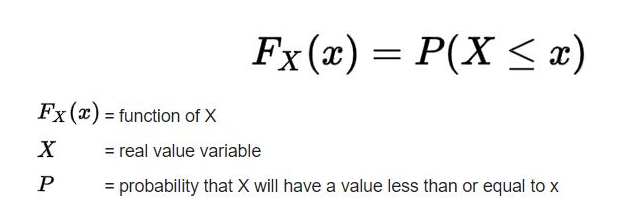
### Probability Mass Function (PMF)

It is a statistical term that describes the probability distribution of a discrete random variable.



### Cumulative Distribution Function (CDF)

It is another method to describe the distribution of a random variable (either continuous or discrete).



**Reverences** :

<https://www.scribbr.com/statistics/probability-distributions/>

<https://www.analyticsvidhya.com/blog/2017/09/6-probability-distributions-data-science/#Distribution_Function_in_Probability>

<https://datasciencedojo.com/blog/types-of-statistical-distributions-in-ml/>