Chapter one

1-1 Probability distribution

In [probability theory](https://en.wikipedia.org/wiki/Probability_theory) and [statistics](https://en.wikipedia.org/wiki/Statistics), a **probability distribution** is a mathematical function that, stated in simple terms, can be thought of as providing the probabilities of occurrence of different possible outcomes in an [experiment](https://en.wikipedia.org/wiki/Experiment_%28probability_theory%29). For instance, if the [random variable](https://en.wikipedia.org/wiki/Random_variable) *X* is used to denote the outcome of a coin toss ("the experiment"), then the probability distribution of *X* would take the value 0.5 for *X* = heads, and 0.5 for *X* = tails (assuming the coin is fair).

In more technical terms, the probability distribution is a description of a [random](https://en.wikipedia.org/wiki/Randomness) phenomenon in terms of the [probabilities](https://en.wikipedia.org/wiki/Probability) of [events](https://en.wikipedia.org/wiki/Event_%28probability_theory%29). Examples of random phenomena can include the results of an [experiment](https://en.wikipedia.org/wiki/Experiment_%28probability_theory%29) or [survey](https://en.wikipedia.org/wiki/Survey_methodology). A probability distribution is defined in terms of an underlying [sample space](https://en.wikipedia.org/wiki/Sample_space), which is the [set](https://en.wikipedia.org/wiki/Set_%28mathematics%29) of all possible [outcomes](https://en.wikipedia.org/wiki/Outcome_%28probability%29) of the random phenomenon being observed. The sample space may be the set of [real numbers](https://en.wikipedia.org/wiki/Real_numbers) or a higher-dimensional [vector space](https://en.wikipedia.org/wiki/Vector_space), or it may be a list of non-numerical values; for example, the sample space of a coin flip would be {heads, tails} .

Probability distributions are generally divided into two classes. A **discrete probability distribution** (applicable to the scenarios where the set of possible outcomes is [discrete](https://en.wikipedia.org/wiki/Discrete_probability_distribution), such as a coin toss or a roll of dice) can be encoded by a discrete list of the probabilities of the outcomes, known as a [probability mass function](https://en.wikipedia.org/wiki/Probability_mass_function). On the other hand, a **continuous probability distribution** (applicable to the scenarios where the set of possible outcomes can take on values in a continuous range (e.g. real numbers), such as the temperature on a given day) is typically described by [probability density functions](https://en.wikipedia.org/wiki/Probability_density_function) (with the probability of any individual outcome actually being 0). The [normal distribution](https://en.wikipedia.org/wiki/Normal_distribution) is a commonly encountered continuous probability distribution. More complex experiments, such as those involving [stochastic processes](https://en.wikipedia.org/wiki/Stochastic_processes) defined in [continuous time](https://en.wikipedia.org/wiki/Continuous_time), may demand the use of more general [probability measures](https://en.wikipedia.org/wiki/Probability_measure).

A probability distribution whose sample space is the set of real numbers is called [univariate](https://en.wikipedia.org/wiki/Univariate_distribution%22%20%5Co%20%22Univariate%20distribution), while a distribution whose sample space is a [vector space](https://en.wikipedia.org/wiki/Vector_space) is called [multivariate](https://en.wikipedia.org/wiki/Multivariate_distribution). A univariate distribution gives the probabilities of a single [random variable](https://en.wikipedia.org/wiki/Random_variable) taking on various alternative values; a multivariate distribution (a [joint probability distribution](https://en.wikipedia.org/wiki/Joint_probability_distribution)) gives the probabilities of a [random vector](https://en.wikipedia.org/wiki/Random_vector)—a list of two or more random variables—taking on various combinations of values. Important and commonly encountered univariate probability distributions include the [binomial distribution](https://en.wikipedia.org/wiki/Binomial_distribution), the [hypergeometric distribution](https://en.wikipedia.org/wiki/Hypergeometric_distribution%22%20%5Co%20%22Hypergeometric%20distribution), and the [normal distribution](https://en.wikipedia.org/wiki/Normal_distribution). The [multivariate normal distribution](https://en.wikipedia.org/wiki/Multivariate_normal_distribution) is a commonly encountered multivariate distribution.

## 1-2Cumulative distribution function

Because a probability distribution *P* on the real line is determined by the probability of a [scalar](https://en.wikipedia.org/wiki/Scalar_%28mathematics%29) random variable *X* being in a half-open interval (−∞, *x*], the probability distribution is completely characterized by its [cumulative distribution function](https://en.wikipedia.org/wiki/Cumulative_distribution_function):

F(X)=P[X≤x] for all x

## 1-3 Discrete probability distribution

A **discrete probability distribution** is a probability distribution characterized by a [probability mass function](https://en.wikipedia.org/wiki/Probability_mass_function). Thus, the distribution of a [random variable](https://en.wikipedia.org/wiki/Random_variable) *X* is discrete, and *X* is called a **discrete random variable**, if

∑P[X≤u] =1

Well-known discrete probability distributions used in statistical modeling include the [Poisson distribution](https://en.wikipedia.org/wiki/Poisson_distribution), the [Bernoulli distribution](https://en.wikipedia.org/wiki/Bernoulli_distribution), the [binomial distribution](https://en.wikipedia.org/wiki/Binomial_distribution), the [geometric distribution](https://en.wikipedia.org/wiki/Geometric_distribution), and the [negative binomial distribution](https://en.wikipedia.org/wiki/Negative_binomial_distribution). Additionally, the [discrete uniform distribution](https://en.wikipedia.org/wiki/Uniform_distribution_%28discrete%29) is commonly used in computer programs that make equal-probability random selections between a number of choices.

## 1-4Continuous probability distribution

A **continuous probability distribution** is a probability distribution that has a cumulative distribution function that is continuous. Most often they are generated by having a [probability density function](https://en.wikipedia.org/wiki/Probability_density_function). Mathematicians call distributions with probability density functions **absolutely continuous**, since their [cumulative distribution function](https://en.wikipedia.org/wiki/Cumulative_distribution_function) is [absolutely continuous](https://en.wikipedia.org/wiki/Absolute_continuity) with respect to the [Lebesgue measure](https://en.wikipedia.org/wiki/Lebesgue_measure%22%20%5Co%20%22Lebesgue%20measure) *λ*. If the distribution of *X* is continuous, then *X* is called a **continuous random variable**. There are many examples of continuous probability distributions: [normal](https://en.wikipedia.org/wiki/Normal_distribution), [uniform](https://en.wikipedia.org/wiki/Uniform_distribution_%28continuous%29), [chi-squared](https://en.wikipedia.org/wiki/Chi-squared_distribution), and [others](https://en.wikipedia.org/wiki/List_of_probability_distributions#Continuous_distributions).

Intuitively, a continuous random variable is the one which can take a [continuous range of values](https://en.wikipedia.org/wiki/Continuous_and_discrete_variables)—as opposed to a discrete distribution, where the set of possible values for the random variable is at most [countable](https://en.wikipedia.org/wiki/Countable_set). While for a discrete distribution an [event](https://en.wikipedia.org/wiki/Event_%28probability_theory%29) with [probability](https://en.wikipedia.org/wiki/Probability) zero is impossible (e.g., rolling 31/2 on a standard dice is impossible, and has probability zero), this is not so in the case of a continuous random variable. For example, if one measures the width of an oak leaf, the result of 3½ cm is possible; however, it has probability zero because uncountably many other potential values exist even between 3 cm and 4 cm. Each of these individual outcomes has probability zero, yet the probability that the outcome will fall into the [interval](https://en.wikipedia.org/wiki/Interval_%28mathematics%29) (3 cm, 4 cm) is nonzero. This apparent [paradox](https://en.wikipedia.org/wiki/Paradox) is resolved by the fact that the probability that *X* attains some value within an [infinite](https://en.wikipedia.org/wiki/Infinity) set, such as an interval, [cannot be found by naively adding](https://en.wikipedia.org/wiki/Integral) the probabilities for individual values. Formally, each value has an [infinitesimally](https://en.wikipedia.org/wiki/Infinitesimal) small probability, which [statistically is equivalent](https://en.wikipedia.org/wiki/Almost_surely) to zero.

Formally, if *X* is a continuous random variable, then it has a [probability density function](https://en.wikipedia.org/wiki/Probability_density_function) *ƒ*(*x*), and therefore its probability of falling into a given interval, say [*a*, *b*] is given by the integral

P[a<x<b] =