Classical definition: Initially the probability of an event to occur was defined as number of cases favorable for the event, over the number of total outcomes possible in an equiprobable sample space: see Classical definition of probability.

In probability and statistics, a random variable or stochastic variable is a variable whose value is not known. A random variable is a function, which maps events or outcomes (e.g., the possible results of rolling two dice: {1, 1}, {1, 2}, etc.) to real numbers (e.g., their sum). A random variable's possible values might represent the possible outcomes of a yet-to-be-performed experiment, or the potential values of a quantity whose already-existing value is uncertain (e.g., as a result of incomplete information or imprecise measurements). Intuitively, a random variable can be thought of as a quantity whose value is not fixed, but which can take on different values; a probability distribution is used to describe the probabilities of different values occurring. Realizations of a random variable are called random variates.
Random variables are usually real-valued, but one can consider arbitrary types such as boolean values, complex numbers, vectors, matrices, sequences, trees, sets, shapes, manifolds, functions, and processes. The term random element. There are two types of random variables: discrete and continuous.[1][dubious – discuss] A discrete random variable maps outcomes to values of a countable set (e.g., the integers), with each value in the range having probability greater than or equal to zero. A continuous random variable maps outcomes to values of an uncountable set (e.g., the real numbers).

In probability theory, a probability density function (pdf), or density of a continuous random variable is a function that describes the relative likelihood for this random variable to occur at a given point. The probability for the random variable to fall within a particular region is given by the integral of this variable’s density over the region. The probability density function is nonnegative everywhere, and its integral over the entire space is equal to one.
The terms “probability distribution function”[1] and “probability function”[2] have also sometimes been used to denote the probability density function. However, special care should be taken around this usage since it is not standard among probabilists and statisticians. In other sources, “probability distribution function” may be used when the probability distribution is defined as a function over general sets of values, or it may refer to the cumulative distribution function, or it may be a probability mass function rather than the density. Further confusion of terminology exists because density function has also been used for what is here called the “probability mass function”.[3]

In probability theory and statistics, a probability mass function (pmf) is a function that gives the probability that a discrete random variable is exactly equal to some value. The probability mass function is often the primary means of defining a discrete probability distribution, and such functions exist for either scalar or multivariate random variables, given that the distribution is discrete.

In probability theory, the expected value (or expectation, or mathematical expectation, or mean, or the first moment) of a random variable is the weighted average of all possible values that this random variable can take on. The weights used in computing this average correspond to the probabilities in case of a discrete random variable, or densities in case of a continuous random variable. From a rigorous theoretical standpoint, the expected value is the integral of the random variable with respect to its probability measure. [1][2]
The expected value may be intuitively understood by the law of large numbers: The expected value, when it exists, is almost surely the limit of the sample mean as sample size grows to infinity. More informally, it can be interpreted as the long-run average of the results of many independent repetitions of an experiment (e.g. a die roll). The value may not be expected in the general sense—the "expected value" itself may be unlikely or even impossible (such as having 2.5 children), just like the sample mean.

The expected values of the powers of X are called the moments of X; the moments about the mean of X are expected values of powers of $X - E[X]$. The moments of some random variables can be used to specify their distributions, via their moment generating functions.
To empirically estimate the expected value of a random variable, one repeatedly measures observations of the variable and computes the arithmetic mean of the results. If the expected value exists, this procedure estimates the true expected value in an unbiased manner and has the property of minimizing the sum of the squares of the residuals (the sum of the squared differences between the observations and the estimate). The law of large numbers demonstrates (under fairly mild conditions) that, as the size of the sample gets larger, the variance of this estimate gets smaller.

In probability theory and statistics, the variance is used as a measure of how far a set of numbers are spread out from each other. It is one of several descriptors of a probability distribution, describing how far the numbers lie from the mean (expected value). In particular, the variance is one of the moments of a distribution. In that context, it forms part of a systematic approach to distinguishing between probability

distributions. While other such approaches have been developed, those based on moments are advantageous in terms of mathematical and computational simplicity.
The variance is a parameter describing in part either the actual probability distribution of an observed population of numbers, or the theoretical probability distribution of a not-fully-observed population of numbers. In the latter case a sample of data from such a distribution can be used to construct an estimate of its variance: in the simplest cases this estimate can be the sample variance — that is, the variance of the sample data.

In probability theory, the central limit theorem (CLT) states conditions under which the mean of a sufficiently large number of independent random variables, each with finite mean and variance, will be approximately normally distributed (Rice 1995). The central limit theorem also requires the random variables to be identically distributed, unless certain conditions are met. Since real-world quantities are often the balanced sum of many unobserved random events, this theorem provides a partial explanation for the prevalence of the normal probability distribution. The CLT also justifies the approximation of large-sample statistics to the normal distribution in controlled experiments.

n probability theory and statistics, the Poisson distribution (pronounced [pwasɔ̃]) (or Poisson law of small numbers[1]) is a discrete probability distribution that expresses the probability of a number of events occurring in a fixed period of time if these events occur with a known average rate and independently of the time since the last event. (The Poisson distribution can also be used for the number of events in other specified intervals such as distance, area or volume.)
The distribution was first introduced by Siméon Denis Poisson (1781–1840) and published, together with his probability theory, in 1838 in his work Recherches sur la probabilité des jugements en matière criminelle et en matière civile (“Research on the Probability of Judgments in Criminal and Civil Matters”). The work focused on certain random variables N that count, among other things, the number of discrete occurrences (sometimes called “arrivals”) that take place during a time-interval of given length.
If the expected number of occurrences in this interval is λ, then the probability that there are exactly k occurrences (k being a non-negative integer, k = 0, 1, 2, ...) is equal to

where
e is the base of the natural logarithm (e = 2.71828...)
k is the number of occurrences of an event — the probability of which is given by the function
k! is the factorial of k
λ is a positive real number, equal to the expected number of occurrences that occur during the given interval. For instance, if the events occur on average 4 times per minute, and one is interested in the probability of an event occurring k times in a 10 minute interval, one would use a Poisson distribution as the model with λ = 10×4 = 40.