

## 1. Probability

Probability is a measure of the likeliness that an event will occur.

Probability is used to quantify an attitude of mind towards some proposition of whose truth we are not certain. The proposition of interest is usually of the form "Will a specific event occur?" The attitude of mind is of the form "How certain are we that the event will occur?" The certainty we adopt can be described in terms of a numerical measure and this number, between 0 and 1 (where 0 indicates impossibility and 1 indicates certainty), we call probability. Thus the higher the probability of an event, the more certain we are that the event will occur. A simple example would be the toss of a fair coin. Since the 2 outcomes are deemed equiprobable, the probability of "heads" equals the probability of "tails" and each probability is  $1/2$  or equivalently a 50% chance of either "heads" or "tails".

These concepts have been given an axiomatic mathematical formalization in probability theory (see probability axioms), which is used widely in such areas of study as mathematics, statistics, finance, gambling, science (in particular physics), artificial intelligence/machine learning, computer science, and philosophy to, for example, draw inferences about the expected frequency of events. Probability theory is also used to describe the underlying mechanics and regularities of complex systems.

### 1.1 Interpretations

When dealing with experiments that are random and well-defined in a purely theoretical setting (like tossing a fair coin), probabilities can be numerically described by the statistical number of outcomes considered favorable divided by the total number of all outcomes (tossing a fair coin twice will yield head-head with probability  $1/4$ , because the four outcomes head-head, head-tails, tails-head and tails-tails are equally likely to occur). When it comes to practical application however there are two major competing categories of probability interpretations, whose adherents possess different views about the fundamental nature of probability:

Objectivists assign numbers to describe some objective or physical state of affairs. The most popular version of objective probability is frequentist probability, which claims that the probability of a random event denotes the relative frequency of occurrence of an experiment's outcome, when repeating the experiment. This interpretation considers probability to be the relative frequency "in the long run" of

outcomes. A modification of this is propensity probability, which interprets probability as the tendency of some experiment to yield a certain outcome, even if it is performed only once.

Subjectivists assign numbers per subjective probability, i.e., as a degree of belief. The degree of belief has been interpreted as, "the price at which you would buy or sell a bet that pays 1 unit of utility if E, 0 if not E." The most popular version of subjective probability is Bayesian probability, which includes expert knowledge as well as experimental data to produce probabilities. The expert knowledge is represented by some (subjective) prior probability distribution. The data is incorporated in a likelihood function. The product of the prior and the likelihood, normalized, results in a posterior probability distribution that incorporates all the information known to date. Starting from arbitrary, subjective probabilities for a group of agents, some Bayesians claim that all agents will eventually have sufficiently similar assessments of probabilities, given enough evidence (see Cromwell's rule).

## **1.2 Etymology**

The word Probability derives from the Latin *probabilitas*, which can also mean probity, a measure of the authority of a witness in a legal case in Europe, and often correlated with the witness's nobility. In a sense, this differs much from the modern meaning of probability, which, in contrast, is a measure of the weight of empirical evidence, and is arrived at from inductive reasoning and statistical inference.

## **1.3 History**

The scientific study of probability is a modern development. Gambling shows that there has been an interest in quantifying the ideas of probability for millennia, but exact mathematical descriptions arose much later. There are reasons of course, for the slow development of the mathematics of probability. Whereas games of chance provided the impetus for the mathematical study of probability, fundamental issues are still obscured by the superstitions of gamblers.

According to Richard Jeffrey, "Before the middle of the seventeenth century, the term 'probable' (Latin *probabilis*) meant approvable, and was applied in that sense, univocally, to opinion and to action. A probable action or opinion was one such as sensible people would undertake or hold, in the circumstances." However, in legal contexts especially, 'probable' could also apply to propositions for which there was good evidence.

The sixteenth century polymath Gerolamo Cardano demonstrated the efficacy of defining odds as the ratio of favourable to unfavourable outcomes (which implies that the probability of an event is given by the ratio of favourable outcomes to the total number of possible outcomes). Aside from the elementary work by Cardano, the doctrine of probabilities dates to the correspondence of Pierre de Fermat and Blaise Pascal (1654). Christiaan Huygens (1657) gave the earliest known scientific treatment of the subject. Jakob Bernoulli's *Ars Conjectandi* (posthumous, 1713) and Abraham de Moivre's *Doctrine of Chances* (1718) treated the subject as a branch of mathematics. See Ian Hacking's *The Emergence of Probability* and James Franklin's *The Science of Conjecture* for histories of the early development of the very concept of mathematical probability.

The theory of errors may be traced back to Roger Cotes's *Opera Miscellanea* (posthumous, 1722), but a memoir prepared by Thomas Simpson in 1755 (printed 1756) first applied the theory to the discussion of errors of observation. The reprint (1757) of this memoir lays down the axioms that positive and negative errors are equally probable, and that certain assignable limits define the range of all errors. Simpson also discusses continuous errors and describes a probability curve.

The first two laws of error that were proposed both originated with Pierre-Simon Laplace. The first law was published in 1774 and stated that the frequency of an error could be expressed as an exponential function of the numerical magnitude of the error, disregarding sign. The second law of error was proposed in 1778 by Laplace and stated that the frequency of the error is an exponential function of the square of the error. The second law of error is called the normal distribution or the Gauss law. "It is difficult historically to attribute that law to Gauss, who in spite of his well-known precocity had probably not made this discovery before he was two years old."

Daniel Bernoulli (1778) introduced the principle of the maximum product of the probabilities of a system of concurrent errors.

Adrien-Marie Legendre (1805) developed the method of least squares, and introduced it in his *Nouvelles méthodes pour la détermination des orbites des comètes* (New Methods for Determining the Orbits of Comets). [citation needed] In ignorance of Legendre's contribution, an Irish-American writer, Robert Adrain, editor of "The Analyst" (1808), first deduced the law of facility of error,

$$\phi(x) = ce^{-h^2 x^2},$$

Where  $h$  is a constant depending on precision of observation, and  $c$  is a scale factor ensuring that the area under the curve equals 1. He gave two proofs, the second being essentially the same as John Herschel's (1850). Gauss gave the first proof that seems to have been known in Europe (the third after Adrain's) in 1809. Further proofs were given by Laplace (1810, 1812), Gauss (1823), James Ivory (1825, 1826), Hagen (1837), Friedrich Bessel (1838), W. F. Donkin (1844, 1856), and Morgan Crofton (1870). Other contributors were Ellis (1844), De Morgan (1864), Glaisher (1872), and Giovanni Schiaparelli (1875). Peters's (1856) formula for  $r$ , the probable error of a single observation, is well known.

In the nineteenth century authors on the general theory included Laplace, Sylvestre Lacroix (1816), Littrow (1833), Adolphe Quetelet (1853), Richard Dedekind (1860), Helmert (1872), Hermann Laurent (1873), Liagre, Didion, and Karl Pearson. Augustus De Morgan and George Boole improved the exposition of the theory.

Andrey Markov introduced[citation needed] the notion of Markov chains (1906), which played an important role in stochastic processes theory and its applications. The modern theory of probability based on the measure theory was developed by Andrey Kolmogorov (1931).

On the geometric side (see integral geometry) contributors to *The Educational Times* were influential (Miller, Crofton, McColl, Wolstenholme, Watson, and Artemas Martin).

## 1.4 Theory

Like other theories, the theory of probability is a representation of probabilistic concepts in formal terms—that is, in terms that can be considered separately from their meaning. These formal terms are manipulated by the rules of mathematics and logic, and any results are interpreted or translated back into the problem domain.

There have been at least two successful attempts to formalize probability, namely the Kolmogorov formulation and the Cox formulation. In Kolmogorov's formulation (see probability space), sets are interpreted as events and probability itself as a measure on a class of sets. In Cox's theorem, probability is taken as a

primitive (that is, not further analyzed) and the emphasis is on constructing a consistent assignment of probability values to propositions. In both cases, the laws of probability are the same, except for technical details.

There are other methods for quantifying uncertainty, such as the Dempster–Shafer theory or possibility theory, but those are essentially different and not compatible with the laws of probability as usually understood.

## **1.5 Applications**

Probability theory is applied in everyday life in risk assessment and in trade on financial markets. Governments apply probabilistic methods in environmental regulation, where it is called pathway analysis. A good example is the effect of the perceived probability of any widespread Middle East conflict on oil prices—which have ripple effects in the economy as a whole. An assessment by a commodity trader that a war is more likely vs. less likely sends prices up or down, and signals other traders of that opinion. Accordingly, the probabilities are neither assessed independently nor necessarily very rationally. The theory of behavioral finance emerged to describe the effect of such groupthink on pricing, on policy, and on peace and conflict.

The discovery of rigorous methods to assess and combine probability assessments has changed society. It is important for most citizens to understand how probability assessments are made, and how they contribute to decisions.

Another significant application of probability theory in everyday life is reliability. Many consumer products, such as automobiles and consumer electronics, use reliability theory in product design to reduce the probability of failure. Failure probability may influence a manufacturer's decisions on a product's warranty.

The cache language model and other statistical language models that are used in natural language processing are also examples of applications of probability theory.

## **1.6 Mathematical treatment**

Consider an experiment that can produce a number of results. The collection of all results is called the sample space of the experiment. The power set of the sample space is formed by considering all different collections of possible results. For example, rolling a dice can produce six possible results. One collection of possible results gives an odd number on the dice. Thus, the subset  $\{1,3,5\}$  is an element of

the power set of the sample space of dice rolls. These collections are called "events." In this case,  $\{1,3,5\}$  is the event that the dice falls on some odd number. If the results that actually occur fall in a given event, the event is said to have occurred.

A probability is a way of assigning every event a value between zero and one, with the requirement that the event made up of all possible results (in our example, the event  $\{1,2,3,4,5,6\}$ ) is assigned a value of one. To qualify as a probability, the assignment of values must satisfy the requirement that if you look at a collection of mutually exclusive events (events with no common results, e.g., the events  $\{1,6\}$ ,  $\{3\}$ , and  $\{2,4\}$  are all mutually exclusive), the probability that at least one of the events will occur is given by the sum of the probabilities of all the individual events.

The probability of an event A is written as  $P(A)$ ,  $p(A)$  or  $\Pr(A)$ . This mathematical definition of probability can extend to infinite sample spaces, and even uncountable sample spaces, using the concept of a measure.

The opposite or complement of an event A is the event [not A] (that is, the event of A not occurring); its probability is given by  $P(\text{not } A) = 1 - P(A)$ . As an example, the chance of not rolling a six on a six-sided die is  $1 - (\text{chance of rolling a six})$

$$= 1 - \frac{1}{6} = \frac{5}{6} \text{ See Complementary event for a more complete treatment.}$$

If two events A and B occur on a single performance of an experiment, this is called the intersection or joint probability of A and B, denoted as  $P(A \cap B)$

### 1.6.1 Independent events

If two events, A and B are independent then the joint probability is

$$P(A \text{ and } B) = P(A \cap B) = P(A)P(B),$$

for example, if two coins are flipped the chance of both being heads is  $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ .

Mutually exclusive events[edit]

If either event A or event B or both events occur on a single performance of an experiment this is called the union of the events A and B denoted as  $P(A \cup B)$  If two events are mutually exclusive then the probability of either occurring is  $P(A \text{ or } B) = P(A \cup B) = P(A) + P(B)$ .

For example, the chance of rolling a 1 or 2 on a six-sided die is  $P(1 \text{ or } 2) = P(1) + P(2) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$ .

### 1.6.2 Not mutually exclusive events

If the events are not mutually exclusive then

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B).$$

For example, when drawing a single card at random from a regular deck of cards, the chance of getting a heart or a face card (J,Q,K) (or one that is both) is  $\frac{13}{52} + \frac{12}{52} - \frac{3}{52} = \frac{11}{26}$ , because of the 52 cards of a deck 13 are hearts, 12 are face cards, and 3 are both: here the possibilities included in the "3 that are both" are included in each of the "13 hearts" and the "12 face cards" but should only be counted once.

### 1.6.3 Conditional probability

*Conditional probability* is the probability of some event  $A$ , given the occurrence of some other event  $B$ . Conditional probability is written  $P(A | B)$ , and is read "the probability of  $A$ , given  $B$ ". It is defined by<sup>[23]</sup>

$$P(A | B) = \frac{P(A \cap B)}{P(B)}.$$

If  $P(B) = 0$  then  $P(A | B)$  is formally undefined by this expression. However, it is possible to define a conditional probability for some zero-probability events using a  $\sigma$ -algebra of such events (such as those arising from a continuous random variable),

For example, in a bag of 2 red balls and 2 blue balls (4 balls in total), the probability of taking a red ball is  $1/2$ ; however, when taking a second ball, the probability of it being either a red ball or a blue ball depends on the ball previously taken, such as, if a red ball was taken, the probability of picking a red ball again would be  $1/3$  since only 1 red and 2 blue balls would have been remaining.

### 1.6.4 Inverse probability

In probability theory and applications, **Bayes' rule** relates the odds of event  $A_1$  to event  $A_2$ , before (prior to) and after (posterior to) conditioning on another event  $B$ . The odds on  $A_1$  to event  $A_2$  is simply the ratio of the probabilities of

the two events. When arbitrarily many events  $A$  are of interest, not just two, the rule can be rephrased as **posterior is proportional to prior times likelihood**,  $P(A|B) \propto P(A)P(B|A)$  where the proportionality symbol means that the left hand side is proportional to (i.e., equals a constant times) the right hand side as  $A$  varies, for fixed or given  $B$  (Lee, 2012; Bertsch McGrayne, 2012). In this form it goes back to Laplace (1774) and to Cournot (1843); see Fienberg (2005). See Inverse probability and Bayes' rule.

### 1.6.5 Relation to randomness

In a deterministic universe, based on Newtonian concepts, there would be no probability if all conditions were known (Laplace's demon), (but there are situations in which sensitivity to initial conditions exceeds our ability to measure them, i.e. know them). In the case of a roulette wheel, if the force of the hand and the period of that force are known, the number on which the ball will stop would be a certainty (though as a practical matter, this would likely be true only of a roulette wheel that had not been exactly levelled — as Thomas A. Bass' Newtonian Casino revealed). Of course, this also assumes knowledge of inertia and friction of the wheel, weight, smoothness and roundness of the ball, variations in hand speed during the turning and so forth. A probabilistic description can thus be more useful than Newtonian mechanics for analyzing the pattern of outcomes of repeated rolls of a roulette wheel. Physicists face the same situation in kinetic theory of gases, where the system, while deterministic *in principle*, is so complex (with the number of molecules typically the order of magnitude of Avogadro constant  $6.02 \cdot 10^{23}$ ) that only a statistical description of its properties is feasible.

Probability theory is required to describe quantum phenomena. A revolutionary discovery of early 20th century physics was the random character of all physical processes that occur at sub-atomic scales and are governed by the laws of quantum mechanics. The objective wave function evolves deterministically but, according to the Copenhagen interpretation, it deals with probabilities of observing, the outcome being explained by a wave function collapse when an observation is made. However, the loss of determinism for the sake of instrumentalism did not meet with universal approval. Albert Einstein famously remarked in a letter to Max Born: "I am convinced that God does not play dice". Like Einstein, Erwin Schrödinger, who discovered the wave function, believed quantum mechanics is a statistical approximation of an underlying deterministic reality. In modern interpretations, quantum decoherence accounts for subjectively probabilistic behavior.