WIKIPEDIA

Probability theory

Probability theory is the branch of mathematics concerned with probability. Although there are several different probability interpretations, probability theory treats the concept in a rigorous mathematical manner by expressing it through a set of axioms. Typically these axioms formalise probability in terms of a probability space, which assigns a measure taking values between 0 and 1, termed the probability measure, to a set of outcomes called the sample space. Any specified subset of these outcomes is called an event. Central subjects in probability theory include discrete and continuous random variables, probability distributions, and stochastic processes, which provide mathematical abstractions of non-deterministic or uncertain processes or measured quantities that may either be single occurrences or evolve over time in a random fashion. Although it is not possible to perfectly predict random events, much can be said about their behavior. Two major results in probability theory describing such behaviour are the law of large numbers and the central limit theorem.

As a mathematical foundation for statistics, probability theory is essential to many human activities that involve quantitative analysis of data.[1] Methods of probability theory also apply to descriptions of complex systems given only partial knowledge of their state, as in statistical mechanics or sequential estimation. A great discovery of twentieth-century physics was the probabilistic nature of physical phenomena at atomic scales, described in quantum mechanics.[2]

History of probability

The early form of statistical inference were developed by Arab mathematicians studying cryptography between the 8th and 13th centuries. Al-Khalil (717–786) wrote the *Book of Cryptographic Messages*, which contains the first use of permutations and combinations to list all possible Arabic words with and without vowels. Al-Kindi (801–873) made the earliest known use of statistical inference in his work on cryptanalysis and frequency analysis. An important contribution of Ibn Adlan (1187–1268) was on sample size for use of frequency analysis.[3]

The modern mathematical theory of probability has its roots in attempts to analyze games of chance by Gerolamo Cardano in the sixteenth century, and by Pierre de Fermat and Blaise Pascal in the seventeenth century (for example the "problem of points").[4] Christiaan Huygens published a book on the subject in 1657[5] and in the 19th century, Pierre Laplace completed what is today considered the classic interpretation.[6]

Initially, probability theory mainly considered *discrete* events, and its methods were mainly combinatorial. Eventually, analytical considerations compelled the incorporation of *continuous* variables into the theory.

This culminated in modern probability theory, on foundations laid by Andrey Nikolaevich Kolmogorov. Kolmogorov combined the notion of sample space, introduced by Richard von Mises, and measure theory and presented his axiom system for probability theory in 1933. This became the mostly undisputed axiomatic basis for modern probability theory; but, alternatives exist, such as the adoption of finite rather than countable additivity by Bruno de Finetti.[7]

Treatment

Most introductions to probability theory treat discrete probability distributions and continuous probability distributions separately. The measure theory-based treatment of probability covers the discrete, continuous, a mix of the two, and more.

Motivation

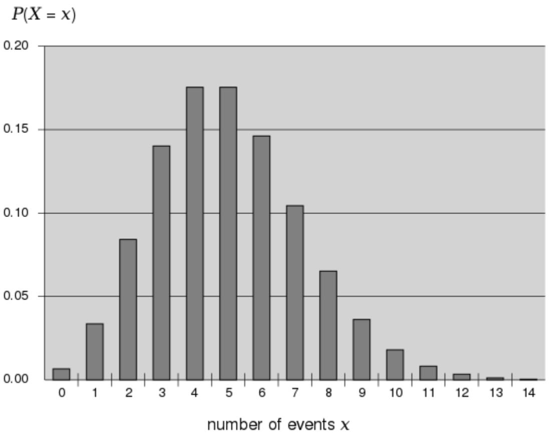
Consider an experiment that can produce a number of outcomes. The set of all outcomes is called the *sample space* of the experiment. The *power set* of the sample space (or equivalently, the event space) is formed by considering all different collections of possible results. For example, rolling an honest die produces one of six possible results. One collection of possible results corresponds to getting an odd number. Thus, the subset {1,3,5} is an element of the power set of the sample space of die rolls. These collections are called *events*. In this case, {1,3,5} is the event that the die falls on some odd number. If the results that actually occur fall in a given event, that event is said to have occurred.

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}) be assigned a value of one. To qualify as a probability distribution, the assignment of values must satisfy the requirement that if you look at a collection of mutually exclusive events (events that contain no common results, e.g., the events {1,6}, {3}, and {2,4} are all mutually exclusive), the probability that any of these events occurs is given by the sum of the probabilities of the events.[8]

The probability that any one of the events {1,6}, {3}, or {2,4} will occur is 5/6. This is the same as saying that the probability of event {1,2,3,4,6} is 5/6. This event encompasses the possibility of any number except five being rolled. The mutually exclusive event {5} has a probability of 1/6, and the event {1,2,3,4,5,6} has a probability of 1, that is, absolute certainty.

When doing calculations using the outcomes of an experiment, it is necessary that all those elementary events have a number assigned to them. This is done using a random variable. A random variable is a function that assigns to each elementary event in the sample space a real number. This function is usually denoted by a capital letter.[9] In the case of a die, the assignment of a number to a certain elementary events can be done using the identity function. This does not always work. For example, when flipping a coin the two possible outcomes are "heads" and "tails". In this example, the random variable *X* could assign to the outcome "heads" the number "0" ($X(heads) = 0$) and to the outcome "tails" the number "1" ($X(tails) = 1$).

Discrete probability distributions



The Poisson distribution, a discrete probability distribution.

Discrete probability theory deals with events that occur in countable sample spaces.

Examples: Throwing dice, experiments with decks of cards, random walk, and tossing coins

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

For example, if the event is "occurrence of an even number when a die is rolled", the probability is given by $\frac{3}{6} = \frac{1}{2}$, since 3 faces out of the 6 have even numbers and each face has the same probability of appearing.

Modern definition: The modern definition starts with a finite or countable set called the sample space, which relates to the set of all *possible outcomes* in classical sense, denoted by Ω. It is then assumed that for each element $x \in \Omega$, an intrinsic "probability" value $f(x)$ is attached, which satisfies the following properties:

1. $f(x) \in [0,1]$ for all $x \in \Omega$;
2. $\sum_{x \in \Omega} f(x) = 1$.

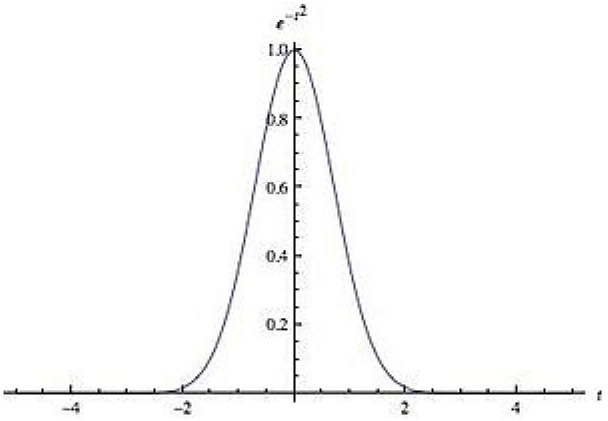
That is, the probability function *f*(*x*) lies between zero and one for every value of *x* in the sample space *Ω*, and the sum of *f*(*x*) over all values *x* in the sample space *Ω* is equal to 1. An *event* is defined as any subset E of the sample space Ω. The *probability* of the event E is defined as

$$P(E)=\sum_{x\in E} f(x)\,.$$

So, the probability of the entire sample space is 1, and the probability of the null event is 0.

The function $f(x)$ mapping a point in the sample space to the "probability" value is called a *probability mass function* abbreviated as *pmf*. The modern definition does not try to answer how probability mass functions are obtained; instead, it builds a theory that assumes their existence.

Continuous probability distributions



The normal distribution, a continuous probability distribution.

Continuous probability theory deals with events that occur in a continuous sample space.

Classical definition: The classical definition breaks down when confronted with the continuous case. See Bertrand's paradox.

Modern definition: If the outcome space of a random variable *X* is the set of real numbers ($\mathbb{R}$) or a subset thereof, then a function called the *cumulative distribution function* (or *cdf*) F exists, defined by $F(x) = P(X \leq x)$. That is, *F*(*x*) returns the probability that *X* will be less than or equal to *x*.

The cdf necessarily satisfies the following properties.

1. F is a monotonically non-decreasing, right-continuous function;
2. $\lim_{x \to -\infty} F(x) = 0$;
3. $\lim_{x \to \infty} F(x) = 1$.

If F is absolutely continuous, i.e., its derivative exists and integrating the derivative gives us the cdf back again, then the random variable *X* is said to have a *probability density function* or *pdf* or simply *density* $f(x) = \dfrac{dF(x)}{dx}$.

For a set $E \subseteq \mathbb{R}$, the probability of the random variable *X* being in E is

$$P(X \in E) = \int_{x \in E} dF(x).$$

In case the probability density function exists, this can be written as

$$P(X \in E) = \int_{x \in E} f(x)\, dx.$$

Whereas the *pdf* exists only for continuous random variables, the *cdf* exists for all random variables (including discrete random variables) that take values in $\mathbb{R}$.

These concepts can be generalized for multidimensional cases on $\mathbb{R}^n$ and other continuous sample spaces.

Measure-theoretic probability theory

The *raison d'être* of the measure-theoretic treatment of probability is that it unifies the discrete and the continuous cases, and makes the difference a question of which measure is used.

Furthermore, it covers distributions that are neither discrete nor continuous nor mixtures of the two.

An example of such distributions could be a mix of discrete and continuous distributions—for example, a random variable that is 0 with probability 1/2, and takes a random value from a normal distribution with probability 1/2. It can still be studied to some extent by considering it to have a pdf of $(\delta[x] + \varphi(x))/2$, where $\delta[x]$ is the Dirac delta function.

Other distributions may not even be a mix, for example, the Cantor distribution has no positive probability for any single point, neither does it have a density. The modern approach to probability theory solves these problems using measure theory to define the probability space:

Given any set Ω (also called *sample space*) and a σ-algebra $\mathcal{F}$ on it, a measure P defined on $\mathcal{F}$ is called a *probability measure* if $P(\Omega) = 1.$

If $\mathcal{F}$ is the Borel σ-algebra on the set of real numbers, then there is a unique probability measure on $\mathcal{F}$ for any cdf, and vice versa. The measure corresponding to a cdf is said to be *induced* by the cdf. This measure coincides with the pmf for discrete variables and pdf for continuous variables, making the measure-theoretic approach free of fallacies.

The *probability* of a set E in the σ-algebra $\mathcal{F}$ is defined as

$$P(E) = \int_{\omega \in E} \mu_F(d\omega)$$

where the integration is with respect to the measure μ_F induced by $F.$

Along with providing better understanding and unification of discrete and continuous probabilities, measure-theoretic treatment also allows us to work on probabilities outside $\mathbb{R}^n$, as in the theory of stochastic processes. For example, to study Brownian motion, probability is defined on a space of functions.

When it's convenient to work with a dominating measure, the Radon-Nikodym theorem is used to define a density as the Radon-Nikodym derivative of the probability distribution of interest with respect to this dominating measure. Discrete densities are usually defined as this derivative with respect to a counting measure over the set of all possible outcomes. Densities for absolutely continuous distributions are usually defined as this derivative with respect to the Lebesgue measure. If a theorem can be proved in this general setting, it holds for both discrete and continuous distributions as well as others; separate proofs are not required for discrete and continuous distributions.

Classical probability distributions

Certain random variables occur very often in probability theory because they well describe many natural or physical processes. Their distributions, therefore, have gained *special importance* in probability theory. Some fundamental *discrete distributions* are the discrete uniform, Bernoulli, binomial, negative binomial, Poisson and geometric distributions. Important *continuous distributions* include the continuous uniform, normal, exponential, gamma and beta distributions.

Convergence of random variables

In probability theory, there are several notions of convergence for random variables. They are listed below in the order of strength, i.e., any subsequent notion of convergence in the list implies convergence according to all of the preceding notions.

Weak convergence

A sequence of random variables $X_1, X_2, \dots,$ converges *weakly* to the random variable X if their respective cumulative *distribution functions* $F_1, F_2, \dots$ converge to the cumulative distribution function F of X, wherever F is continuous. Weak convergence is also called *convergence in distribution*.

Most common shorthand notation: $X_n \xrightarrow{\mathcal{D}} X$

Convergence in probability

The sequence of random variables $X_1, X_2, \dots$ is said to converge towards the random variable X *in probability* if $\lim_{n\to\infty} P\left(|X_n - X| \geq \varepsilon\right) = 0$ for every ε > 0.

Most common shorthand notation: $X_n \xrightarrow{P} X$

Strong convergence

The sequence of random variables $X_1, X_2, \ldots$ is said to converge towards the random variable X *strongly* if $P(\lim_{n\to\infty} X_n = X) = 1$. Strong convergence is also known as *almost sure convergence*.

Most common shorthand notation: $X_n \xrightarrow{\text{a.s.}} X$

As the names indicate, weak convergence is weaker than strong convergence. In fact, strong convergence implies convergence in probability, and convergence in probability implies weak convergence. The reverse statements are not always true.

Law of large numbers

Common intuition suggests that if a fair coin is tossed many times, then *roughly* half of the time it will turn up *heads*, and the other half it will turn up *tails*. Furthermore, the more often the coin is tossed, the more likely it should be that the ratio of the number of *heads* to the number of *tails* will approach unity. Modern probability theory provides a formal version of this intuitive idea, known as the *law of large numbers*. This law is remarkable because it is not assumed in the foundations of probability theory, but instead emerges from these foundations as a theorem. Since it links theoretically derived probabilities to their actual frequency of occurrence in the real world, the law of large numbers is considered as a pillar in the history of statistical theory and has had widespread influence.[10]

The *law of large numbers* (LLN) states that the sample average

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$$

of a sequence of independent and identically distributed random variables X_k converges towards their common expectation μ, provided that the expectation of $|X_k|$ is finite.

It is in the different forms of convergence of random variables that separates the *weak* and the *strong* law of large numbers

Weak law: $\overline{X}_n \xrightarrow{P} \mu$ for $n \to \infty$

Strong law: $\overline{X}_n \xrightarrow{\text{a. s.}} \mu$ for $n \to \infty.$

It follows from the LLN that if an event of probability *p* is observed repeatedly during independent experiments, the ratio of the observed frequency of that event to the total number of repetitions converges towards *p*.

For example, if $Y_1, Y_2, \ldots$ are independent Bernoulli random variables taking values 1 with probability *p* and 0 with probability 1-*p*, then $\mathrm{E}(Y_i) = p$ for all *i*, so that $\bar{Y}_n$ converges to *p* almost surely.

Central limit theorem

"The central limit theorem (CLT) is one of the great results of mathematics." (Chapter 18 in[11]) It explains the ubiquitous occurrence of the normal distribution in nature.

The theorem states that the average of many independent and identically distributed random variables with finite variance tends towards a normal distribution *irrespective* of the distribution followed by the original random variables. Formally, let $X_1, X_2, \ldots$ be independent random variables with mean μ and variance $\sigma^2 > 0.$ Then the sequence of random variables

$$Z_n = \frac{\sum_{i=1}^n (X_i - \mu)}{\sigma\sqrt{n}}$$

converges in distribution to a standard normal random variable.

For some classes of random variables the classic central limit theorem works rather fast (see Berry–Esseen theorem), for example the distributions with finite first, second, and third moment from the exponential family; on the other hand, for some random variables of the heavy tail and fat tail variety, it works very slowly or may not work at all: in such cases one may use the Generalized Central Limit Theorem (GCLT).

See also

- Catalog of articles in probability theory
- Expected value and Variance

- Fuzzy logic and Fuzzy measure theory
- Glossary of probability and statistics
- Likelihood function
- List of probability topics
- List of publications in statistics
- List of statistical topics
- Notation in probability
- Predictive modelling
- Probabilistic logic – A combination of probability theory and logic
- Probabilistic proofs of non-probabilistic theorems
- Probability distribution
- Probability axioms
- Probability interpretations
- Probability space
- Statistical independence
- Statistical physics
- Subjective logic
- Probability of the union of pairwise independent events

Notes

1. *Inferring From Data (http://home.ubalt.edu/ntsbarsh/stat-data/Topics.htm)*
2. *"Why is quantum mechanics based on probability theory?" (https://physics.stackexchange.com/q/69718) . StackExchange. July 1, 2014.*

3. Broemeling, Lyle D. (1 November 2011). "An Account of Early Statistical Inference in Arab Cryptology". The American Statistician. **65** (4): 255–257. doi:10.1198/tas.2011.10191 (https://doi.org/10.1198%2Ftas.2011.10191) .

4. LIGHTNER, JAMES E. (1991). "A Brief Look at the History of Probability and Statistics" (https://www.jstor.org/stable/27967334) . The Mathematics Teacher. **84** (8): 623–630. ISSN 0025-5769 (https://www.worldcat.org/issn/0025-5769) .

5. Grinstead, Charles Miller; James Laurie Snell. "Introduction". Introduction to Probability. pp. vii.

6. Hájek, Alan (Fall 2019). "Interpretations of Probability". In Zalta, Edward (ed.). The Stanford Encyclopedia of Philosophy (https://plato.stanford.edu/archives/fall2019/entries/probability-interpret/) .

7. " "The origins and legacy of Kolmogorov's Grundbegriffe", by Glenn Shafer and Vladimir Vovk" (http://www.probabilityandfinance.com/articles/04.pdf) (PDF). Retrieved 2012-02-12.

8. Ross, Sheldon (2010). A First Course in Probability (https://books.google.com/books?id=Bc1FAQAAIAAJ&pg=PA26) (8th ed.). Pearson Prentice Hall. pp. 26–27. ISBN 978-0-13-603313-4. Retrieved 2016-02-28.

9. Bain, Lee J.; Engelhardt, Max (1992). Introduction to Probability and Mathematical Statistics (2nd ed.). Belmont, California: Brooks/Cole. p. 53. ISBN 978-0-534-38020-5.

10. "Leithner & Co Pty Ltd - Value Investing, Risk and Risk Management - Part I" (https://web.archive.org/web/20140126113323/http://www.leithner.com.au/circulars/circular17.htm) . Leithner.com.au. 2000-09-15. Archived from the original (http://www.leithner.com.au/circulars/circular17.htm) on 2014-01-26. Retrieved 2012-02-12.

11. David Williams, "Probability with martingales", Cambridge 1991/2008

References

- Pierre Simon de Laplace (1812). *Analytical Theory of Probability*.

  The first major treatise blending calculus with probability theory, originally in French: *Théorie Analytique des Probabilités*.

- A. Kolmogoroff (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. doi:10.1007/978-3-642-49888-6 (https://doi.org/10.1007%2F978-3-642-49888-6) . ISBN 978-3-642-49888-6.

  An English translation by Nathan Morrison appeared under the title *Foundations of the Theory of Probability* (Chelsea, New York) in 1950, with a second edition in 1956.

- Patrick Billingsley (1979). *Probability and Measure*. New York, Toronto, London: John Wiley and Sons.
- Olav Kallenberg; *Foundations of Modern Probability,* 2nd ed. Springer Series in Statistics. (2002). 650 pp. ISBN 0-387-95313-2
- Henk Tijms (2004). *Understanding Probability*. Cambridge Univ. Press.
  A lively introduction to probability theory for the beginner.
- Olav Kallenberg; *Probabilistic Symmetries and Invariance Principles*. Springer -Verlag, New York (2005). 510 pp. ISBN 0-387-25115-4
- Gut, Allan (2005). *Probability: A Graduate Course*. Springer-Verlag. ISBN 0-387-22833-0.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Probability_theory&oldid=1035224029"

Last edited 6 days ago by Guest2625