Stochastic Methods – Definitions, Random Variables, Distributions

Sigma-algebra

In mathematics, a **σ-algebra** (pronounced *sigma-algebra*) or **σ-field** over a set X is a family Σ of subsets of X that is closed under countable set operations; σ-algebras are mainly used in order to define measures on X.

Formally, Σ is a σ-algebra if and only if it has the following properties:

1. The empty set is in Σ.
2. If E is in Σ then so is the complement $X \setminus E$ of E.
3. If $E_1, E_2, E_3, \ldots$ is a (countable) sequence in Σ then their (countable) union is also in Σ.

From 1 and 2 it follows that X is in Σ; from 2 and 3 it follows that the σ-algebra is also closed under countable intersections (via De Morgan's laws).

Elements of the σ-algebra are called measurable sets. An ordered pair (X, Σ), where X is a set and Σ is a σ-algebra over X, is called a **measurable space**. Measures are defined as certain types of functions from a σ-algebra to $[0,\infty]$.

Random variables

Some consider the expression *random variable* a misnomer, as a random variable is not a variable but rather a function that maps events to numbers. Let A be a σ-algebra and Ω the space of events relevant to the experiment being performed. In the die-rolling example, the space of events is just the possible outcomes of a roll, i.e. $\Omega = \{ 1, 2, 3, 4, 5, 6 \}$, and A would be the power set of Ω. In this case, an appropriate random variable might be the identity function $X(\omega) = \omega$, such that if the outcome is a '1', then the random variable is also equal to 1. An equally simple but less trivial example is one in which we might toss a coin: a suitable space of possible events is $\Omega = \{ \mathrm{H}, \mathrm{T} \}$ (for heads and tails), and A equal again to the power set of Ω. One among the many possible random variables defined on this space is

$$X(\omega) = \begin{cases} 0, & \omega = \mathtt{H}, \\ 1, & \omega = \mathtt{T}. \end{cases}$$

Mathematically, a random variable is defined as a measurable function from a probability space to some measurable space. This measurable space is the space of possible values of the variable, and it is usually taken to be the real numbers with the Borel σ-algebra. This is assumed in the following, except where specified.

Let (Ω, *A*, *P*) be a probability space. Formally, a function X: Ω → **R** is a (real-valued) *random variable* if for every subset $A_r = \{ \omega : X(\omega) \leq r \}$. The importance of this technical definition is that it allows us to construct the distribution function of the random variable.

Distribution functions

If a random variable $X : \Omega \to \mathbb{R}$ defined on the probability space (Ω, *A, P*) is given, we can ask questions like "How likely is it that the value of X is bigger than 2?". This is the same as the probability of the event $\{s \in \Omega : X(s) > 2\}$ which is often written as $P(X > 2)$ for short.

Recording all these probabilities of output ranges of a real-valued random variable X yields the probability distribution of X. The probability distribution "forgets" about the particular probability space used to define X and only records the probabilities of various values of X. Such a probability distribution can always be captured by its cumulative distribution function

$$F_X(x) = \mathrm{P}(X \leq x)$$

and sometimes also using a probability density function. In measure-theoretic terms, we use the random variable X to "push-forward" the measure P on Ω to a measure dF on **R**. The underlying probability space Ω is a technical device used to guarantee the existence of random variables, and sometimes to construct them. In practice, one often disposes of the space Ω altogether and just puts a measure on **R** that assigns measure 1 to the whole real line, i.e., one works with probability distributions instead of random variables.

Functions of random variables

If we have a random variable X on Ω and a measurable function f: **R** → **R**, then $Y = f(X)$ will also be a random variable on Ω, since the composition of measurable functions is also measurable. The same procedure that allowed one to go from a probability space (Ω, *A, P*) to (**R**, dF_X) can be used to obtain the distribution of Y. The cumulative distribution function of Y is

$$F_Y(y) = \mathrm{P}(f(X) \leq y).$$

Example

Let X be a real-valued, continuous random variable and let $Y = X^2$. Then,

$$F_Y(y) = \mathrm{P}(X^2 \leq y).$$

If $y < 0$, then $\mathrm{P}(X^2 \leq y) = 0$, so

$$F_Y(y) = 0 \qquad \text{if} \quad y < 0.$$

If $y \geq 0$, then

$$\mathrm{P}(X^2 \leq y) = \mathrm{P}(|X| \leq \sqrt{y}) = \mathrm{P}(-\sqrt{y} \leq X \leq \sqrt{y}),$$

so

$$F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) \qquad \text{if} \quad y \geq 0.$$

Moments

The probability distribution of random variable is often characterised by a small number of parameters, which also have a practical interpretation. For example, it is often enough to know what its "average value" is. This is captured by the mathematical concept of expected value of a random variable, denoted E[X]. Note that in general, E[$f(X)$] is **not** the same as f(E[X]). Once the "average value" is known, one could then ask how far from this average value the values of X typically are, a question that is answered by the variance and standard deviation of a random variable.

Mathematically, this is known as the (generalised) problem of moments: for a given class of random variables X, find a collection $\{f_i\}$ of functions such that the expectation values E[$f_i(X)$] fully characterize the distribution of the random variable X.

Equivalence of random variables

There are several different senses in which random variables can be considered to be equivalent. Two random variables can be equal, equal almost surely, equal in mean, or equal in distribution.

In increasing order of strength, the precise definition of these notions of equivalence is given below.

Equality in distribution

Two random variables X and Y are *equal in distribution* if they have the same distribution functions:

$$\mathrm{P}(X \leq x) = \mathrm{P}(Y \leq x) \quad \text{for all} \quad x.$$

To be equal in distribution, random variables need not be defined on the same probability space. The notion of equivalence in distribution is associated to the following notion of distance between probability distributions,

$$d(X, Y) = \sup_x |P(X \le x) - P(Y \le x)|,$$

which is the basis of the Kolmogorov-Smirnov test.

Kolmogorov-Smirnov test

In statistics, the **Kolmogorov-Smirnov** test (often referred to as the **K-S test**) is used to determine whether two underlying probability distributions differ from each other or whether an underlying probability distribution differs from a hypothesized distribution, in either case based on finite samples.

In the one-sample case the KS test compares the empirical distribution function with the cumulative distribution function specified by the null hypothesis. The main applications are for testing goodness of fit with the normal and uniform distributions. For normality testing, minor improvements made by Lilliefors lead to the Lilliefors test. In general the Shapiro-Wilk test or Anderson-Darling test are more powerful alternatives to the Lilliefors test for testing normality.

The two sample KS-test is one of the most useful and general nonparametric methods for comparing two samples, as it is sensitive to differences in both location and shape of the empirical cumulative distribution functions of the two samples.

Mathematical statistics

The empirical distribution function F_n for n observations y_i is defined as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \begin{cases} 1 & \text{if } y_i \le x, \\ 0 & \text{otherwise.} \end{cases}$$

The two one-sided Kolmogorov-Smirnov test statistics are given by

$$\begin{aligned} D_n^+ &= \max(F_n(x) - F(x)) \\ D_n^- &= \max(F(x) - F_n(x)) \end{aligned}$$

where $F(x)$ is the hypothesized distribution or another empirical distribution. The probability distributions of these two statistics, given that the null hypothesis of equality of distributions is true, does not depend on what the hypothesized distribution is, as long as it is continuous. Knuth gives a detailed description of how to analyze the significance of this pair of statistics. Many people use max(D_n^+, D_n^-) instead, but the distribution of this statistic is more difficult to deal with.

Equality in mean

Two random variables X and Y are *equal in p-th mean* if the pth moment of $|X - Y|$ is zero, that is,

$$\mathrm{E}(|X - Y|^p) = 0.$$

Equality in pth mean implies equality in qth mean for all $q<p$. As in the previous case, there is a related distance between the random variables, namely

$$d_p(X, Y) = \mathrm{E}(|X - Y|^p).$$

Almost sure equality

Two random variables X and Y are *equal almost surely* if, and only if, the probability that they are different is zero:

$$\mathrm{P}(X \neq Y) = 0.$$

For all practical purposes in probability theory, this notion of equivalence is as strong as actual equality. It is associated to the following distance:

$$d_\infty(X, Y) = \sup_\omega |X(\omega) - Y(\omega)|,$$

where 'sup' in this case represents the essential supremum in the sense of measure theory.

Equality

Finally, two random variables X and Y are *equal* if they are equal as functions on their probability space, that is,

$$X(\omega) = Y(\omega) \qquad \text{for all} \quad \omega$$

Discrete probability distribution

A probability distribution is called **discrete**, if it is fully characterized by a probability mass function. Thus, the distribution of a random variable X is discrete, and X is then called a **discrete random variable**, if

$$\sum_u \Pr(X = u) = 1 \qquad (1)$$

as u runs through the set of all possible values of X.

If a random variable is discrete, then the set of all values that it can assume with nonzero probability is finite or countably infinite, because the sum of uncountably many positive real numbers (which is the smallest upper bound of the set of all finite partial sums) always diverges to infinity.

In the cases most often considered, this set of possible values is a topologically discrete set in the sense that all its points are isolated points. But there are discrete random variables for which this countable set is dense on the real line.

The Poisson distribution, the Bernoulli distribution, the binomial distribution, the geometric distribution, and the negative binomial distribution are among the most well-known discrete probability distributions.

Representation in terms of indicator functions

For a discrete random variable X, let u_0, u_1, ... be the values it can assume with non-zero probability. Denote

$$\Omega_i = \{\omega : X(\omega) = u_i\},\ i = 0, 1, 2, \dots$$

These are disjoint sets, and by formula (1)

$$\Pr\left(\bigcup_i \Omega_i\right) = \sum_i \Pr(\Omega_i) = \sum_i \Pr(X = u_i) = 1.$$

It follows that the probability that X assumes any value except for u_0, u_1, ... is zero, and thus one can write X as

$$X = \sum_i \alpha_i 1_{\Omega_i}$$

except on a set of probability zero, where $\alpha_i = \Pr(X = u_i)$ and 1_A is the indicator function of A. This may serve as an alternative definition of discrete random variables.

List of important probability distributions

Several probability distributions are so important in theory or applications that they have been given specific names:

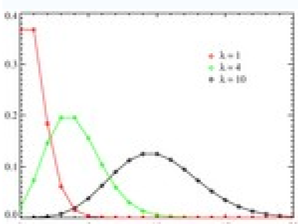
Discrete distributions

With finite support

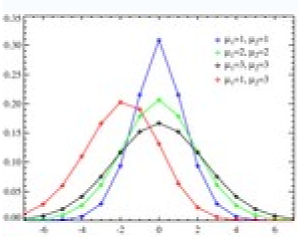
- The Bernoulli distribution, which takes value 1 with probability p and value 0 with probability $q = 1 - p$.
- The Rademacher distribution, which takes value 1 with probability 1/2 and value −1 with probability 1/2.
- The binomial distribution describes the number of successes in a series of independent Yes/No experiments.
- The degenerate distribution at x_0, where X is certain to take the value x_0. This does not look random, but it satisfies the definition of random variable. This is useful because it puts deterministic variables and random variables in the same formalism.
- The discrete uniform distribution, where all elements of a finite set are equally likely. This is supposed to be the distribution of a balanced coin, an unbiased die, a casino roulette or a well-shuffled deck. Also, one can use measurements of quantum states to generate uniform random variables. All these are "physical" or "mechanical" devices, subject to design flaws or perturbations, so the uniform distribution is only an approximation of their behaviour. In digital computers, pseudo-random number generators are used to produced a statistically random discrete uniform distribution.
- The hypergeometric distribution, which describes the number of successes in the first m of a series of n independent Yes/No experiments, if the total number of successes is known.
- Zipf's law or the Zipf distribution. A discrete power-law distribution, the most famous example of which is the description of the frequency of words in the English language.
- The Zipf-Mandelbrot law is a discrete power law distribution which is a generalization of the Zipf distribution.

With infinite support

- The Boltzmann distribution, a discrete distribution important in statistical physics which describes the probabilities of the various discrete energy levels of a system in thermal equilibrium. It has a continuous analogue. Special cases include:
- The Gibbs distribution
- The Maxwell-Boltzmann distribution
- The Bose-Einstein distribution
- The Fermi-Dirac distribution
- The geometric distribution, a discrete distribution which describes the number of attempts needed to get the first success in a series of independent Yes/No experiments.
- The logarithmic (series) distribution
- The negative binomial distribution, a generalization of the geometric distribution to the *n*th success.
- The parabolic fractal distribution
- The Poisson distribution, which describes a very large number of individually unlikely events that happen in a certain time interval.



Poisson distribution

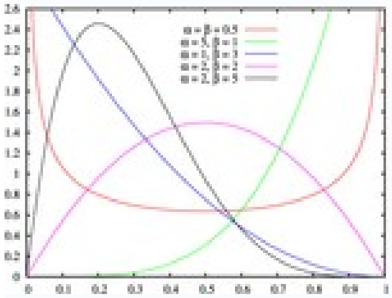


Skellam distribution

- The Skellam distribution, the distribution of the difference between two independent Poisson-distributed random variables.
- The Yule-Simon distribution
- The zeta distribution has uses in applied statistics and statistical mechanics, and perhaps may be of interest to number theorists. It is the Zipf distribution for an infinite number of elements.

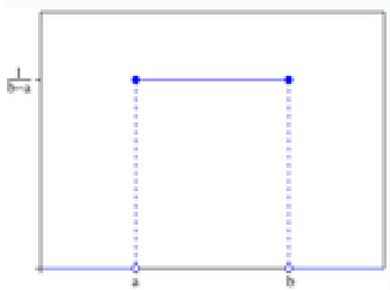
Continuous distributions

Supported on a bounded interval



Beta distribution

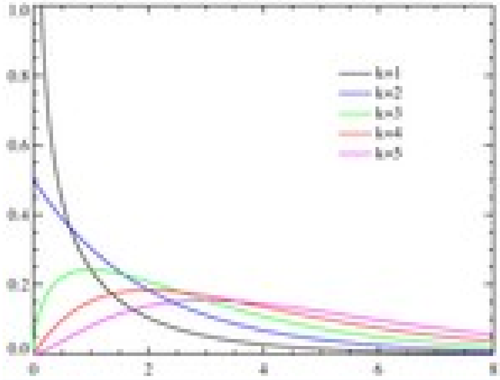
- The Beta distribution on [0,1], of which the uniform distribution is a special case, and which is useful in estimating success probabilities.



continuous uniform distribution

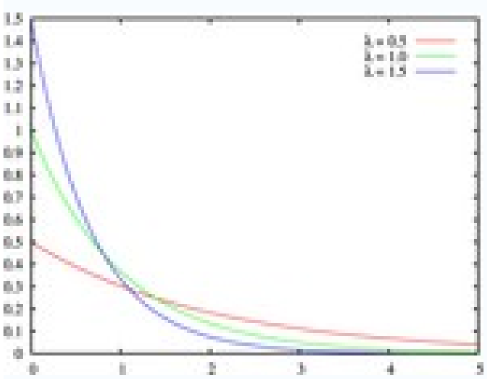
- The continuous uniform distribution on [*a*,*b*], where all points in a finite interval are equally likely.
- The rectangular distribution is a uniform distribution on [-1/2,1/2].
- The Dirac delta function although not strictly a function, is a limiting form of many continuous probability functions. It represents a *discrete* probability distribution concentrated at 0 — a degenerate distribution — but the notation treats it as if it were a continuous distribution.
- The Kumaraswamy distribution is as versatile as the Beta distribution but has simple closed forms for both the cdf and the pdf.
- The logarithmic distribution (continuous)
- The triangular distribution on [*a*, *b*], a special case of which is the distribution of the sum of two uniformly distributed random variables (the *convolution* of two uniform distributions).
- The von Mises distribution
- The Wigner semicircle distribution is important in the theory of random matrices.

Supported on semi-infinite intervals, usually [0,∞)



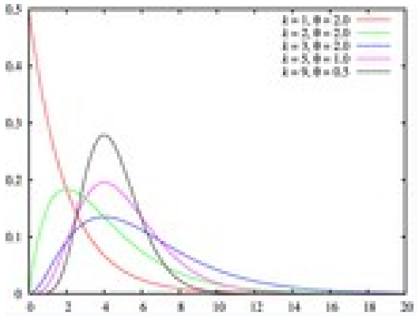
[chi-square distribution]

- The chi distribution
- The noncentral chi distribution
- The chi-square distribution, which is the sum of the squares of *n* independent Gaussian random variables. It is a special case of the Gamma distribution, and it is used in goodness-of-fit tests in statistics.
- The inverse-chi-square distribution
- The noncentral chi-square distribution
- The scale-inverse-chi-square distribution
- The exponential distribution, which describes the time between consecutive rare random events in a process with no memory.



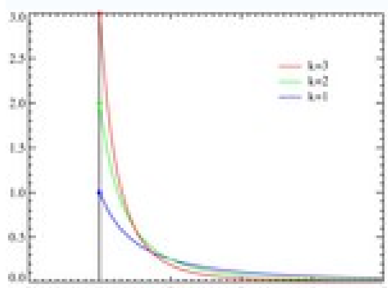
Exponential distribution

- The F-distribution, which is the distribution of the ratio of two (normalized) chi-square distributed random variables, used in the analysis of variance.
- The noncentral F-distribution



[Gamma distribution](#)

- The [Gamma distribution](#), which describes the time until *n* consecutive rare random events occur in a process with no memory.
- The [Erlang distribution](#), which is a special case of the gamma distribution with integral shape parameter, developed to predict waiting times in [queuing systems](#).
- The [inverse-gamma distribution](#)
- [Fisher's z-distribution](#)
- The [half-normal distribution](#)
- The [Lévy distribution](#)
- The [log-logistic distribution](#)
- The [log-normal distribution](#), describing variables which can be modelled as the product of many small independent positive variables.

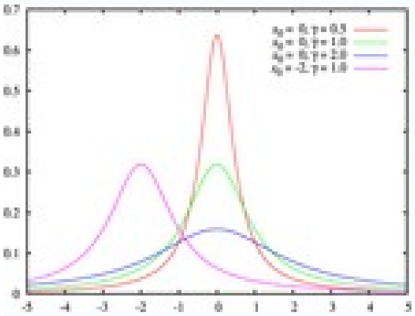


[Pareto distribution](#)

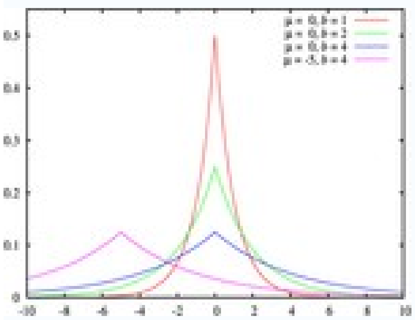
- The [Pareto distribution](#), or "power law" distribution, used in the analysis of financial data and critical behavior.
- The [Rayleigh distribution](#)
- The [Rayleigh mixture distribution](#)
- The [Rice distribution](#)
- The [type-2 Gumbel distribution](#)
- The [Wald distribution](#)

- The [Weibull distribution](), of which the exponential distribution is a special case, is used to model the lifetime of technical devices.

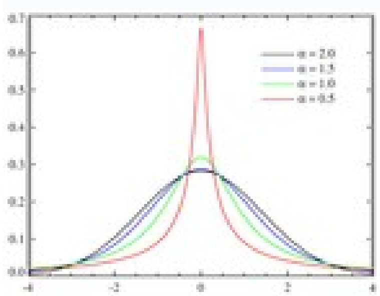
Supported on the whole real line



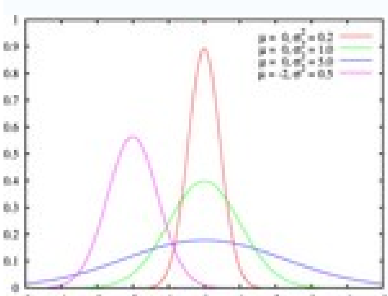
[Cauchy distribution]()



[Laplace distribution]()



[Levy distribution]()



[Normal distribution]()

- The Beta prime distribution
- The Cauchy distribution, an example of a distribution which does not have an expected value or a variance. In physics it is usually called a Lorentzian profile, and is associated with many processes, including resonance energy distribution, impact and natural spectral line broadening and quadratic stark line broadening.
- The Fisher-Tippett, extreme value, or log-Weibull distribution
- The Gumbel distribution, a special case of the Fisher-Tippett distribution
- The generalized extreme value distribution
- The hyperbolic secant distribution
- The Landau distribution
- The Laplace distribution
- The Lévy skew alpha-stable distribution is often used to characterize financial data and critical behavior.
- The map-Airy distribution
- The normal distribution, also called the Gaussian or the bell curve. It is ubiquitous in nature and statistics due to the central limit theorem: every variable that can be modelled as a sum of many small independent variables is approximately normal.
- Student's t-distribution, useful for estimating unknown means of Gaussian populations.
- The noncentral t-distribution
- The type-1 Gumbel distribution
- The Voigt distribution, or Voigt profile, is the convolution of a normal distribution and a Cauchy distribution. It is found in spectroscopy when spectral line profiles are broadened by a mixture of Lorentzian and Doppler broadening mechanisms.

Joint distributions

For any set of independent random variables the probability density function of the joint distribution is the product of the individual ones.

Two or more random variables on the same sample space

- Dirichlet distribution, a generalization of the beta distribution.
- The Ewens's sampling formula is a probability distribution on the set of all partitions of an integer *n*, arising in population genetics.
- multinomial distribution, a generalization of the binomial distribution.
- multivariate normal distribution, a generalization of the normal distribution.

Matrix-valued distributions

- Wishart distribution
- matrix normal distribution
- matrix t-distribution
- Hotelling's T-square distribution

Miscellaneous distributions

- The Cantor distribution