



Statistical Distributions

Arena contains a set of built-in functions for generating random numbers from the commonly used probability distributions. These distributions appear on pull-down menus in many Arena modules where they're likely to be used. They also match the distributions in the Arena Input Analyzer. This appendix describes all of the Arena distributions.

Each of the distributions in Arena has one or more parameter values associated with it. You must specify these parameter values to define the distribution fully. The number, meaning, and order of the parameter values depend on the distribution. A summary of the distributions (in alphabetical order) and parameter values is given in the table below.

Summary of Arena's Probability Distributions

Distribution		Parameter Values
Beta	BETA	Beta, Alpha
Continuous	CONT	CumP ₁ , Val ₁ , . . . CumP _n , Val _n
Discrete	DISC	CumP ₁ , Val ₁ , . . . CumP _n , Val _n
Erlang	ERLA	ExpoMean, k
Exponential	EXPO	Mean
Gamma	GAMM	Beta, Alpha
Johnson	JOHN	Gamma, Delta, Lambda, Xi
Lognormal	LOGN	LogMean, LogStd
Normal	NORM	Mean, StdDev
Poisson	POIS	Mean
Triangular	TRIA	Min, Mode, Max
Uniform	UNIF	Min, Max
Weibull	WEIB	Beta, Alpha

To enter a distribution in an Arena field, you type the name of the distribution (or its four-letter abbreviation) followed by its parameters enclosed in parentheses. You may use spaces around punctuation to help read the distribution. A few examples appear below.

UNIF(3.5, 6)

Uniform distribution with a minimum value of 3.5, a maximum value of 6

NORMAL(83, 12.8)

Normal distribution with a mean of 83, a standard deviation of 12.8

DISCRETE(0.3,50, 0.75,80, 1.0,100)

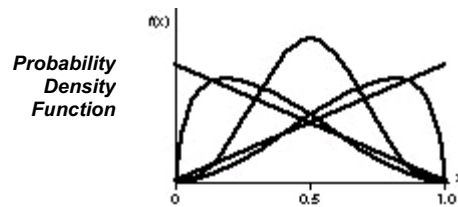
Discrete probability distribution that will return a value of 50 with probability 0.3, a value of 80 with cumulative probability 0.75, and a value of 100 with cumulative probability of 1.0. (See “Discrete probability” for a description of these parameters.)

TRIA(10, 15, 22)

Triangular distribution with a minimum value of 10, mode (most likely value) of 15, and maximum value of 22

In the following pages, we provide a summary of each of the distributions supported by Arena, listed in alphabetical order for easy reference. Each includes the density or mass function, parameters, range, mean, variance, and typical applications for the distribution.

If you have existing data and want to select the appropriate distribution for use in your model, use Arena's Input Analyzer. Click on **Tools > Input Analyzer** to launch the program, or launch it from the Windows Start menu.

Beta(β, α)**BETA(Beta, Alpha)**

$$f(x) = \begin{cases} \frac{x^{\beta-1}(1-x)^{\alpha-1}}{B(\beta, \alpha)} & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

where β is the complete beta function given by

$$B(\beta, \alpha) = \int_0^1 t^{\beta-1} (1-t)^{\alpha-1} dt$$

Parameters Shape parameters Beta (β) and Alpha (α) specified as positive real numbers.

Range $[0, 1]$ (Can also be transformed to $[a, b]$ as described below)

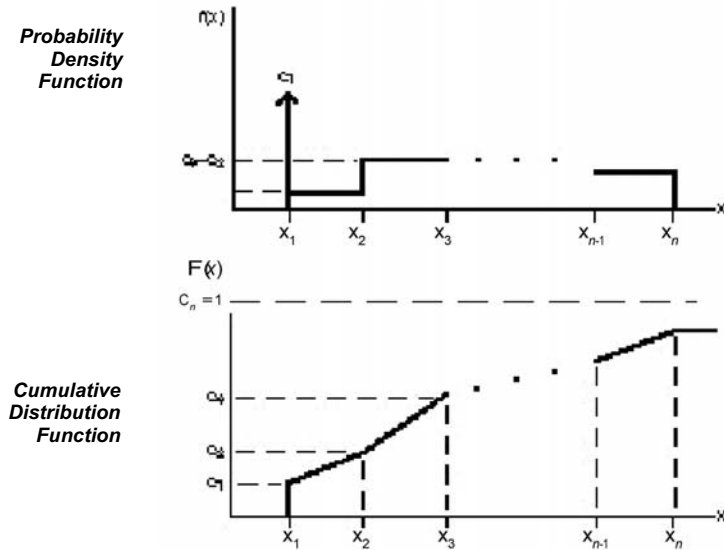
Mean $\frac{\beta}{\beta + \alpha}$

Variance $\frac{\beta \alpha}{(\beta + \alpha)^2 (\beta + \alpha + 1)}$

Applications Because of its ability to take on a wide variety of shapes, this distribution is often used as a rough model in the absence of data. Also, because the range of the beta distribution is from 0 to 1, the sample X can be transformed to the scaled beta sample Y with the range from a to b by using the equation $Y = a + (b - a)X$. The beta is often used to represent random proportions, such as the proportion of defective items in a lot.

Continuous

CONTINUOUS(CumP₁, Val₁, . . . , CumP_n, Val_n)
(c₁, x₁, . . . , c_n, x_n)



$$f(x) = \begin{cases} c_1 & \text{if } x = x_1 \text{ (a mass of probability } c_1 \text{ at } x_1) \\ c_j - c_{j-1} & \text{if } x_{j-1} \leq x < x_j, \text{ for } j = 2, 3, \dots, n \\ 0 & \text{if } x < x_1 \text{ or } x \geq x_n \end{cases}$$

Parameters

The CONTINUOUS function in Arena returns a sample from a user-defined distribution. Pairs of cumulative probabilities c_j ($= \text{CumP}_j$) and associated values x_j ($= \text{Val}_j$) are specified. The sample returned will be a real number between x_1 and x_n , and will be less than or equal to each x_j with corresponding cumulative probability c_j . The x_j 's must increase with j . The c_j 's must all be between 0 and 1, must increase with j , and c_n must be 1.

The cumulative distribution function $F(x)$ is piecewise linear with "corners" defined by $F(x_j) = c_j$ for $j = 1, 2, \dots, n$. Thus, for $j \geq 2$, the returned value will be in the interval



$(x_{j-1}, x_j]$ with probability $c_j - c_{j-1}$; given that it is in this interval, it will be distributed uniformly over it.

You must take care to specify c_1 and x_1 to get the effect you want at the left edge of the distribution. The CONTINUOUS function will return (exactly) the value x_1 with probability c_1 . Thus, if you specify $c_1 > 0$, this actually results in a mixed discrete-continuous distribution returning (exactly) x_1 with probability c_1 , and with probability $1 - c_1$ a continuous random variate on $(x_1, x_n]$ as described above. The graph of $F(x)$ above depicts a situation where $c_1 > 0$. On the other hand, if you specify $c_1 = 0$, you will get a (truly) continuous distribution on $[x_1, x_n]$ as described above, with no “mass” of probability at x_1 ; in this case, the graph of $F(x)$ would be continuous, with no jump at x_1 .

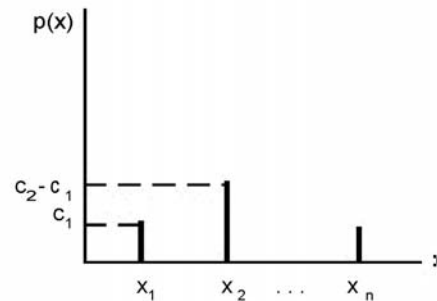
As an example use of the CONTINUOUS function, suppose you have collected a set of data x_1, x_2, \dots, x_n (assumed to be sorted into increasing order) on service times, for example. Rather than using a fitted theoretical distribution from the Input Analyzer, you want to generate service times in the simulation “directly” from the data, consistent with how they’re spread out and bunched up, and between the minimum x_1 and the maximum x_n you observed. Assuming that you don’t want a “mass” of probability sitting directly on x_1 , you’d specify $c_1 = 0$ and then $c_j = (j - 1)/(n - 1)$ for $j = 2, 3, \dots, n$.

Range $[x_1, x_n]$

Applications The continuous empirical distribution is often used to incorporate actual data for continuous random variables directly into the model. This distribution can be used as an alternative to a theoretical distribution that has been fitted to the data, such as in data that have a multimodal profile or where there are significant outliers.

Discrete **DISCRETE(CumP₁, Val₁, . . . , CumP_n, Val_n)**
(c₁, x₁, . . . , c_n, x_n)

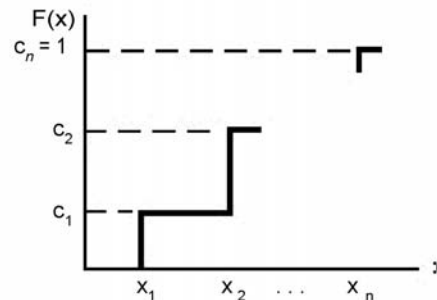
**Probability
Mass
Function**



$$p(x_j) = c_j - c_{j-1}$$

where $c_0 = 0$

**Cumulative
Distribution
Function**

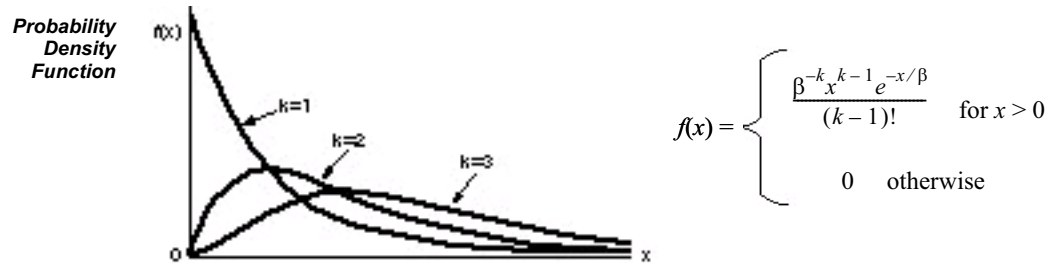


Parameters The DISCRETE function in Arena returns a sample from a user-defined discrete probability distribution. The distribution is defined by the set of n possible discrete values (denoted by x_1, x_2, \dots, x_n) that can be returned by the function and the cumulative probabilities (denoted by c_1, c_2, \dots, c_n) associated with these discrete values. The cumulative probability (c_j) for x_j is defined as the probability of obtaining a value that is less than or equal to x_j . Hence, c_j is equal to the sum of $p(x_i)$ for k going from 1 to j . By definition, $c_n = 1$.

Range $\{x_1, x_2, \dots, x_n\}$

Applications The discrete empirical distribution is often used to incorporate discrete empirical data directly into the model. This distribution is frequently used for discrete assignments such as the job type, the visitation sequence, or the batch size for an arriving entity.

Erlang(β, k) ERLANG(ExpMean, k) or ERLA(ExpMean, k)



Parameters If X_1, X_2, \dots, X_k are independent, identically distributed exponential random variables, then the sum of these k samples has an Erlang- k distribution. The mean (β) of each of the component exponential distributions and the number of exponential random variables (k) are the parameters of the distribution. The exponential mean is specified as a positive real number, and k is specified as a positive integer.

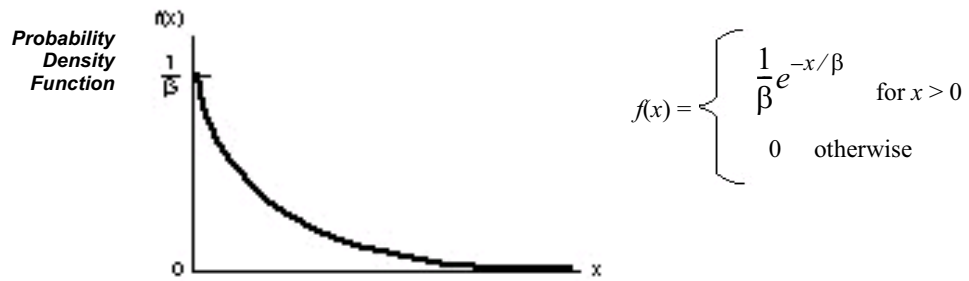
Range $[0, +\infty)$

Mean $k\beta$

Variance $k\beta^2$

Applications The Erlang distribution is used in situations in which an activity occurs in successive phases and each phase has an exponential distribution. For large k , the Erlang approaches the normal distribution. The Erlang distribution is often used to represent the time required to complete a task. The Erlang distribution is a special case of the gamma distribution in which the shape parameter, α , is an integer (k).

Exponential(β) EXPONENTIAL(Mean) or EXPO(Mean)



Parameters The mean (β) specified as a positive real number.

Range $[0, +\infty)$

Mean β

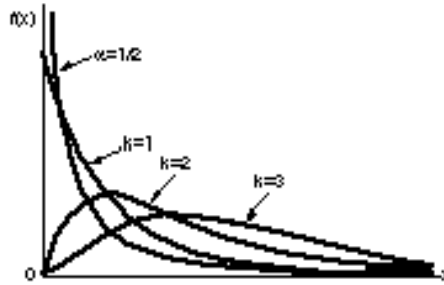
Variance β^2

Applications This distribution is often used to model inter-event times in random arrival and break-down processes, but it is generally inappropriate for modeling process delay times.

In Arena's Create module, the Schedule option automatically samples from an exponential distribution with a mean that changes according to the defined schedule. This is particularly useful in service applications, such as retail business or call centers, where the volume of customers changes throughout the day.

Gamma(β, α) GAMMA(Beta, Alpha) or GAMM(Beta, Alpha)

**Probability
Density
Function**



$$f(x) = \begin{cases} \frac{\beta^{-\alpha} x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

where Γ is the complete gamma function given by

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$$

Parameters Shape parameter (α) and scale parameter (β) specified as positive real values.

Range $[0, +\infty)$

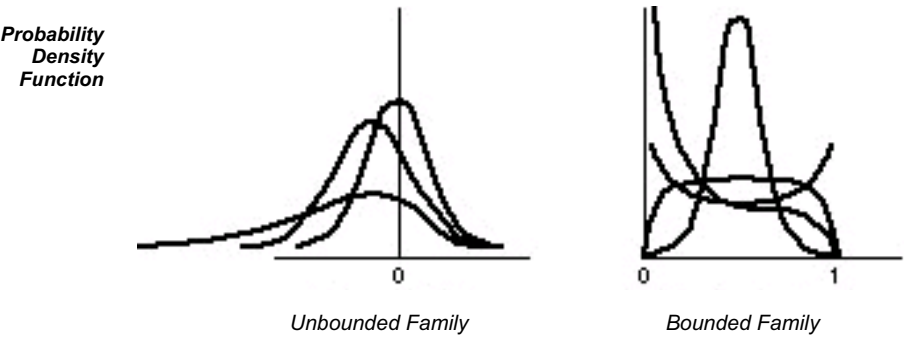
Mean $\alpha\beta$

Variance $\alpha\beta^2$

Applications For integer shape parameters, the gamma is the same as the Erlang distribution. The gamma is often used to represent the time required to complete some task (e.g., a machining time or machine repair time).

Johnson

JOHNSON(Gamma, Delta, Lambda, Xi) or JOHN(Gamma, Delta, Lambda, Xi)



Parameters

Gamma shape parameter (γ), Delta shape parameter ($\delta > 0$), Lambda scale parameter (λ), and Xi location parameter (ξ).

Range

$(-\infty, +\infty)$

Unbounded Family

$[\xi, \xi + \lambda]$

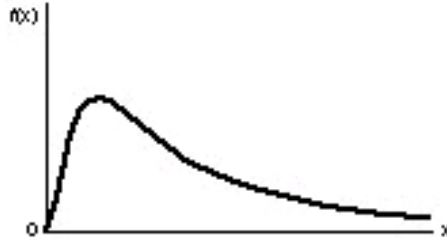
Bounded Family

Applications

The flexibility of the Johnson distribution allows it to fit many data sets. Arena can sample from both the unbounded and bounded form of the distribution. If Delta (δ) is passed as a positive number, the bounded form is used. If Delta is passed as a negative value, the unbounded form is used with $|\delta|$ as the parameter.

Lognormal(μ, σ) LOGNORMAL(LogMean, LogStd) or LOGN(LogMean, LogStd)

Probability
Density
Function



Denote the user-specified input parameters as
LogMean = μ_l and LogStd = σ_l .

Then let $\mu = \ln(\mu_l^2 / \sqrt{\sigma_l^2 + \mu_l^2})$ and

$$\sigma = \sqrt{\ln[(\sigma_l^2 + \mu_l^2) / \mu_l^2]}$$

the probability density function can then be written as

$$f(x) = \begin{cases} \frac{1}{\sigma x \sqrt{2\pi}} e^{-(\ln(x) - \mu)^2 / (2\sigma^2)} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Parameters Mean LogMean ($\mu_l > 0$) and standard deviation LogStd ($\sigma_l > 0$) of the lognormal random variable. Both LogMean and LogStd must be specified as strictly positive real numbers.

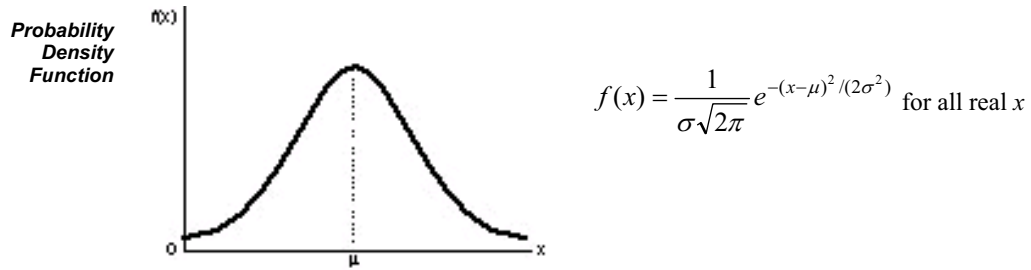
Range $[0, +\infty)$

Mean LogMean = $\mu_l = e^{\mu + \sigma^2/2}$

Variance $(\text{LogStd})^2 = \sigma_l^2 = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$

Applications The lognormal distribution is used in situations in which the quantity is the product of a large number of random quantities. It is also frequently used to represent task times that have a distribution skewed to the right. This distribution is related to the normal distribution as follows. If X has a lognormal (μ_l, σ_l) distribution, then $\ln(X)$ has a normal (μ, σ) distribution. Note that μ and σ are *not* the mean and standard deviation of the lognormal random variable X , but rather the mean and standard deviation of the normal random variable $\ln X$.

Normal(μ, σ) NORMAL(Mean, StdDev) or NORM(Mean, StdDev)



Parameters The mean (μ) specified as a real number and standard deviation (σ) specified as a positive real number.

Range $(-\infty, +\infty)$

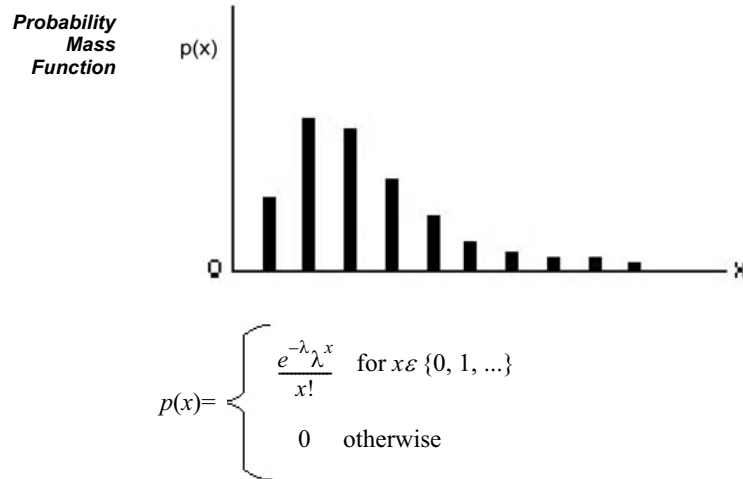
Mean μ

Variance σ^2

Applications The normal distribution is used in situations in which the central limit theorem applies; i.e., quantities that are sums of other quantities. It is also used empirically for many processes that appear to have a symmetric distribution. Because the theoretical range is from $-\infty$ to $+\infty$, the distribution should only be used for positive quantities like processing times.



Poisson(λ) POISSON(Mean) or POIS(Mean)



Parameters The mean (λ) specified as a positive real number.

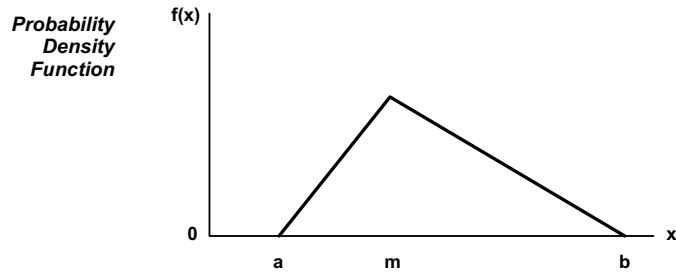
Range $\{0, 1, \dots\}$

Mean λ

Variance λ

Applications The Poisson distribution is a discrete distribution that is often used to model the number of random events occurring in a fixed interval of time. If the time between successive events is exponentially distributed, then the number of events that occur in a fixed-time interval has a Poisson distribution. The Poisson distribution is also used to model random batch sizes.

Triangular(*a, m, b*) TRIANGULAR(Min, Mode, Max) or TRIA(Min, Mode, Max)



$$f(x) = \begin{cases} \frac{2(x-a)}{(m-a)(b-a)} & \text{for } a \leq x \leq m \\ \frac{2(b-x)}{(b-m)(b-a)} & \text{for } m \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Parameters The minimum (a), mode (m), and maximum (b) values for the distribution specified as real numbers with $a < m < b$.

Range $[a, b]$

Mean $(a + m + b)/3$

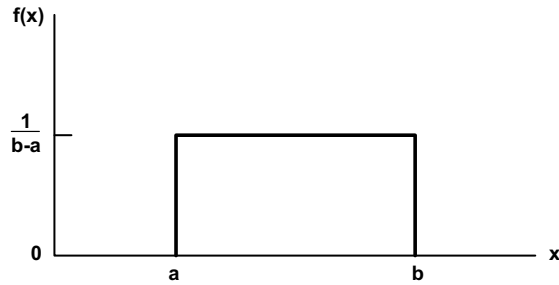
Variance $(a^2 + m^2 + b^2 - ma - ab - mb)/18$

Applications The triangular distribution is commonly used in situations in which the exact form of the distribution is not known, but estimates (or guesses) for the minimum, maximum, and most likely values are available. The triangular distribution is easier to use and explain than other distributions that may be used in this situation (e.g., the beta distribution).



Uniform(a, b) UNIFORM(Min, Max) or UNIF(Min, Max)

**Probability
Density
Function**



$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Parameters The minimum (a) and maximum (b) values for the distribution specified as real numbers with $a < b$.

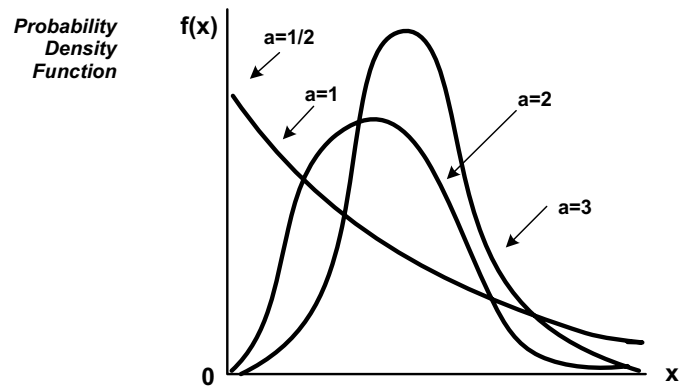
Range $[a, b]$

Mean $(a + b)/2$

Variance $(b - a)^2/12$

Applications The uniform distribution is used when all values over a finite range are considered to be equally likely. It is sometimes used when no information other than the range is available. The uniform distribution has a larger variance than other distributions that are used when information is lacking (e.g., the triangular distribution).

Weibull(β, α) WEIBULL(Beta, Alpha) or WEIB(Beta, Alpha)



$$f(x) = \begin{cases} a\beta^{-a}x^{a-1}e^{-(x/\beta)^a} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Parameters Shape parameter (α) and scale parameter (β) specified as positive real values.

Range $[0, +\infty)$

Mean $\frac{\beta}{a}\Gamma\left(\frac{1}{a}\right)$, where Γ is the complete gamma function (see gamma distribution).

Variance $\frac{\beta^2}{a}\left\{2\Gamma\left(\frac{2}{a}\right) - \frac{1}{a}\left[\Gamma\left(\frac{1}{a}\right)\right]^2\right\}$



Applications The Weibull distribution is widely used in reliability models to represent the lifetime of a device. If a system consists of a large number of parts that fail independently, and if the system fails when any single part fails, then the time between successive failures can be approximated by the Weibull distribution. This distribution is also used to represent non-negative task times that are skewed to the left.