APPENDIX D

Arena's Probability Distributions

Arena contains a set of built-in functions for generating random variates from the commonly used probability distributions. These distributions appear on drop-down menus in many Arena modules where they're likely to be used. They also match the distributions in the Arena Input Analyzer (except for the Johnson distribution). 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 Table D-1.

Table D-1. Summary of Arena's Probability Distributions

Distribution			Parameter Values
Beta	BETA	BE	Beta, Alpha
Continuous	CONT	CP	$CumP_1, Val_1, \ldots CumP_n, Val_n$
Discrete	DISC	DP	$CumP_1, Val_1, \ldots CumP_n, Val_n$
Erlang	ERLA	ER	ExpoMean, k
Exponential	EXPO	EX	Mean
Gamma	GAMM	GA	Beta, Alpha
Johnson	JOHN	JO	Gamma, Delta, Lambda, Xi
Lognormal	LOGN	RL	LogMean, LogStd
Normal	NORM	RN	Mean, StdDev
Poisson	POIS	PO	Mean
Triangular	TRIA	TR	Min, Mode, Max
Uniform	UNIF	UN	Min, Max
Weibull	WEIB	WE	Beta, Alpha

The distributions can be specified by using one of two formats: you can select a single format, or you can mix formats within the same model. The format is determined by the name used to specify the distribution. The primary format is selected by using either the variable's full name or a four-letter abbreviation of the name consisting of the first four letters. For example, UNIFORM or UNIF specifies the uniform distribution in the primary format. The secondary format is selected by specifying the distribution with a two-letter abbreviation. For example, UN specifies the uniform distribution in the secondary format. The names are not case-sensitive.

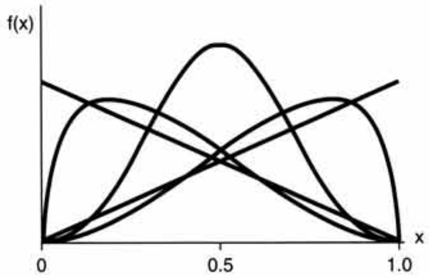
In the primary format, you explicitly enter the parameters of the distribution as arguments of the distribution. For example, UNIFORM(10, 25) specifies a uniform distribution with a minimum value of 10 and a maximum value of 25. In the alternative format, you indirectly define the parameters of the distribution by referencing a parameter set within the Parameters module from the Elements panel. For example, UN(DelayTime) specifies a uniform distribution with the minimum and maximum values defined in the parameter set named DelayTime. The main advantage of the indirect method of defining the parameters provided by the alternative format is that the parameters of the distribution can be modified from within the Parameters module.

The random-number stream, which is used by Arena in generating the sample, can also be specified in both formats. In the primary format, you enter the stream number as the last argument following the parameter value list. For example, UNIFORM(10,25,2) specifies a sample from a uniform distribution using random-number stream 2. In the secondary format, you enter the random-number stream as a second argument following the identifier for the parameter set. For example, UN(DelayTime,PTimeStream) specifies a sample from a uniform distribution using random-number stream PTimeStream (which must be an expression defined to be the stream you want, and is not an entry in the Parameters module).

In the following pages, we provide a summary of each of the distributions supported by Arena, listed in alphabetical order for easy reference. The summary includes the primary and secondary formats for specifying the distribution and a brief description of the distribution. This description includes the density or mass function, parameters, range, mean, variance, and typical applications for the distribution. If you feel in need of a brief refresher on probability and statistics, see Appendix C.

Beta(β, α) **BETA(Beta, Alpha) or BE(ParamSet)**

Probability Density Function



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

where B 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 general range $[a,b]$ as described below.)

Mean

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

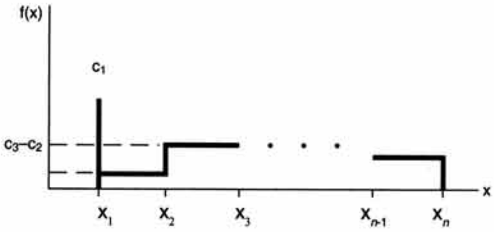
Variance

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

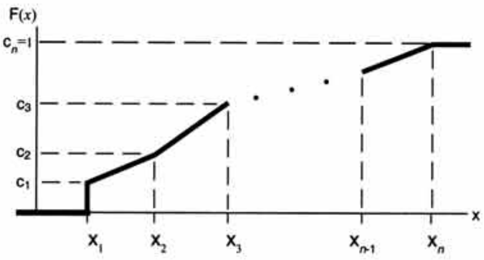
Applications Because of its ability to take on a wide variety of shapes, this distribution is often used as a rough model in the absence of data. Because the range of the beta distribution is from 0 to 1, the sample X can be transformed to the scaled beta sample Y with the range from a to b by using 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. It can also be used as a general and very flexible distribution to represent many input quantities that can be assumed to have a range bounded on both ends.

Continuous	**CONTINUOUS(CumP$_1$, Val$_1$, . . ., CumP$_n$, Val$_n$) or**
$(c_1, x_1, \ldots, c_n, x_n)$	**CONT(CumP$_1$, Val$_1$, . . ., CumP$_n$, Val$_n$) or CP(ParamSet)**

Probability Density Function



Cumulative Distribution Function



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

Parameters The CONTINUOUS function in Arena returns a sample from a user-defined empirical distribution. Pairs of cumulative probabilities c_j (= CumP$_j$) and associated values x_j (= 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 $j = 1, 2, \ldots, n$. Thus, for $j \ge 2$, 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.

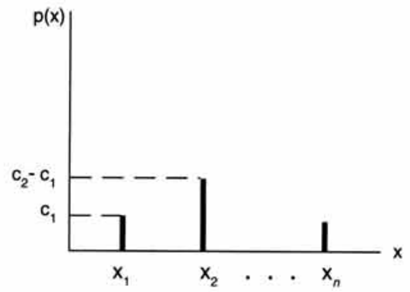
As an example use of the CONTINUOUS function, suppose you have collected a set of data $x_1, x_2, \ldots, x_n$ (assumed to be sorted into increasing order) on service times, for example. Rather than using a fitted theoretical distribution from the Input Analyzer (Section 4.5), you want to generate service 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 $j = 2, 3, \ldots, n$.

Range $[x_1, x_n]$

Applications The continuous empirical distribution is used to incorporate empirical 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(CumP$_1$, Val$_1$, . . ., CumP$_n$, Val$_n$) or**
$(c_1, x_1, \ldots, c_n, x_n)$ **DISC(CumP$_1$, Val$_1$, . . ., CumP$_n$, Val$_n$) or DP(ParamSet)**

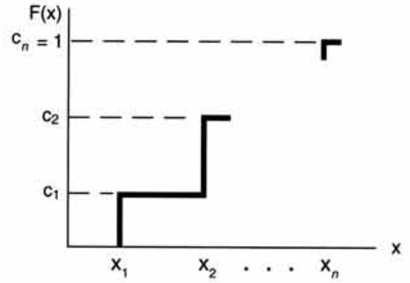
Probability Mass Function



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

where $c_0 = 0$

Cumulative Distribution Function



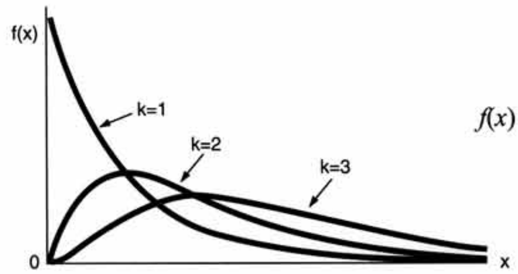
Parameters The DISCRETE function in Arena returns a sample from a user-defined discrete probability distribution. The distribution is defined by the set of n possible discrete values (denoted by $x_1, x_2, \ldots, x_n$) that can be returned by the function and the cumulative probabilities (denoted by $c_1, c_2, \ldots, c_n$) 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 $p(x_k)$ for k going from 1 to j. By definition, $c_n = 1$.

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

Applications The discrete empirical distribution is 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) or ER(ParamSet)**

Probability Density Function



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

Parameters If $X_1, X_2, \ldots, X_k$ are IID 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 $[0, +\infty)$

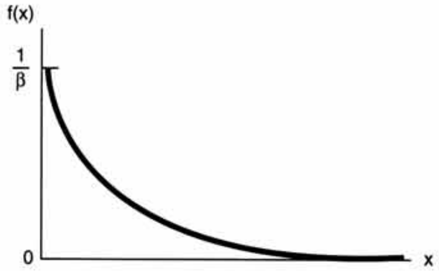
Mean $k\beta$

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) or EX(ParamSet)**

Probability Density Function



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

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

Range $[0, +\infty)$

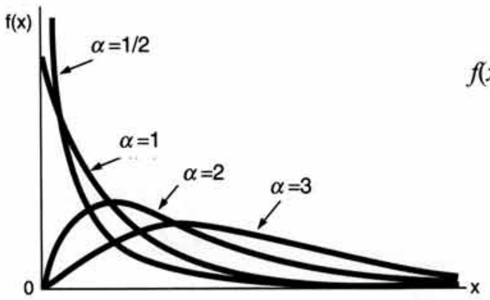
Mean β

Variance β^2

Applications This distribution is often used to model interevent times in random arrival and breakdown processes, but it is generally inappropriate for modeling process delay times.

Gamma(β, α) **GAMMA(Beta, Alpha) or GAMM(Beta, Alpha) or GA(ParamSet)**

Probability Density Function



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

where Γ is the complete gamma function given by

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

Parameters Scale parameter (β) and shape parameter (α) specified as positive real values.

Range $[0, +\infty)$

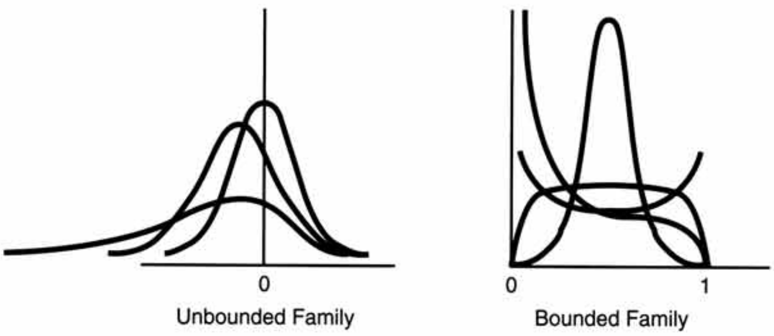
Mean $\alpha\beta$

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 (for example, a machining time or machine repair time).

Johnson **JOHNSON(Gamma, Delta, Lambda, Xi) or JOHN(Gamma, Delta, Lambda, Xi) or JO(ParamSet)**

Probability Density Function



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

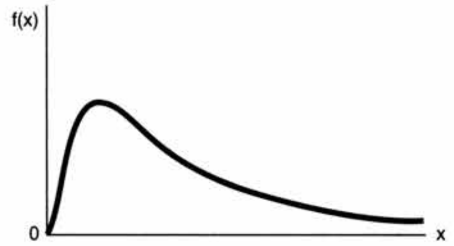
Range

$(-\infty, +\infty)$	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. (At present, the Input Analyzer does not support fitting Johnson distributions to data.)

Lognormal(μ, σ) **LOGNORMAL(LogMean, LogStd) or LOGN(LogMean, LogStd) or RL(ParamSet)**

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 $\sigma = \sqrt{\ln\left[(\sigma_l^2 + \mu_l^2) / \mu_l^2\right]}$. The probability density function can then be written as

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

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 $[0, +\infty)$

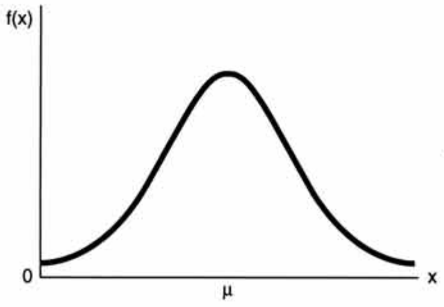
Mean LogMean = $\mu_l = e^{\mu + \sigma^2/2}$

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$X$; the mean LogMean = μ_l and variance (LogStd)^2 = σ_l^2 of X are given by the formulas earlier on this page.

Normal(μ, σ) **NORMAL(Mean, StdDev) or NORM(Mean, StdDev) or RN(ParamSet)**

Probability Density Function



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} \quad \text{for all real } x$$

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

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

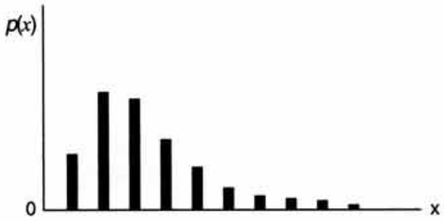
Mean μ

Variance σ^2

Applications The normal distribution is used in situations in which the central limit theorem applies—that is, 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 not be used for positive quantities like processing times.

Poisson(λ) **POISSON(Mean) or POIS(Mean) or PO(ParamSet)**

Probability Mass Function



$$p(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & \text{for } x \varepsilon \{0, 1, \ldots\} \\ 0 & \text{otherwise} \end{cases}$$

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

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

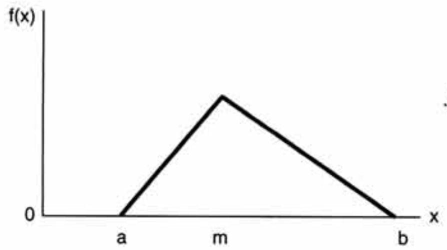
Mean λ

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) or TR(ParamSet)

Probability Density Function



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

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

Range $[a, b]$

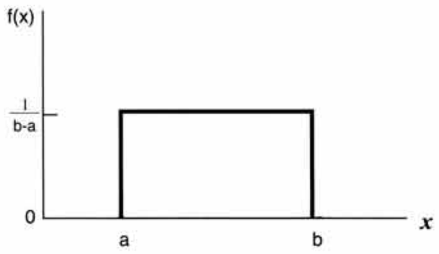
Mean $(a + m + b)/3$

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) or UN(ParamSet)**

Probability Density Function



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

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

Range $[a, b]$

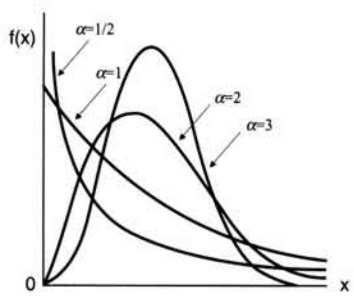
Mean $(a + b)/2$

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) or WE(ParamSet)**

Probability Density Function



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

Parameters Scale parameter (β) and shape parameter (α) specified as positive real numbers.

Range $[0, +\infty)$

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

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

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.