Probability density functions supported by ADSM

Probability density functions are distributions of values representative of the natural range of possible values for some parameter. Values are drawn stochastically from these distributions as a simulation runs.

ADSM supports 22 general pdf types of probability density functions, described in the following sections, along with the parameter order expected by ADSM. Some distributions are more suitable to some applications than others, but all are provided to ensure maximum flexibility to model users. References are provided for users who wish to obtain more detailed information about these distributions; in particular, Vose (2000) provides a very helpful discussion regarding suitable applications of the different types of probability density functions. A discussion of distribution fitting is beyond the scope of this guide, but readers are referred to other sources (Law, 2006; Vose, 2000).

Most of the distributions described here are continuous. Recall, though, that ADSMoperates in discrete time steps of one day. Consequently, for parameters that have x axis units of days, values obtained from these distributions will be rounded to the nearest whole day.

The Beta distribution

A Beta distribution (figure A-1) is a continuous distribution defined by four parameters: α1, α2, a minimum value, and a maximum value. Parameters α1 and α2 must be greater than 0, and the minimum value must be less than the maximum value. The probability density *f(x)* is calculated as:

|  |  |
| --- | --- |
| where *B* denotes the Beta function. |  |
| Figure 0‑1. A sample Beta distribution. |

The BetaPERT distribution

The BetaPERT distribution (figure 0‑2) is a continuous distribution defined by its minimum, its most likely value (mode), and its maximum. In this way, the BetaPERT distribution is similar to the Triangular distribution.

The BetaPERT distribution is related to the Beta distribution; the α1 and α2 parameters used to define a Beta distribution are obtained from the minimum, mode, and maximum values of a BetaPERT. There are several slight variations in how α1 and α2 are calculated from the minimum, mode, and maximum. ADSM uses the same approach as the risk analysis package *@RISK* (Palisade Corporation, 2008). In this approach, α1 and α2 are calculated with the following formulas:

|  |  |
| --- | --- |
| where μ is the distribution mean, defined as: |  |
| Figure 0‑2. A sample BetaPERT distribution. |

The Bernoulli distribution

The Bernoulli distribution is a discrete distribution used to model whether an event will occur. A single parameter *p* represents the probability of the event. For example, the occurrence of getting a head on one toss of a coin could be modeled with the distribution Bernoulli (0.5), or the occurrence of getting a 6 on a single roll of die could be modeled with Bernoulli (0.1667). The outcome of a Bernoulli trial is always 0 or 1. The Bernoulli distribution is a special case of the Binomial distribution, in which the number of trials is 1.

|  |  |
| --- | --- |
| The probability density function *f(x)* for a Bernoulli distribution is calculated as: |  |
| Figure 0‑3. A sample Bernoulli distribution. |

The Binomial distribution

The Binomial distribution is a discrete distribution used most often to model the number of *x* successes from *n* independent trials in which there is a probability *p* of success in each trial (figure 0‑4). It is used when there are exactly two mutually exclusive outcomes of a trial. The probability density function *f(x)* is calculated as:

|  |  |
| --- | --- |
|  |  |
| Figure 0‑4. A sample Binomial distribution. |

The Discreet Uniform distribution

The Discreet Uniform distribution, sometimes called the “equally likely outcomes” distribution, has a set of *n* elements in which each element *i* has the same probability of occurring (figure 0‑5). A simple example of this distribution is the outcome of throwing one die. The probability density function *f(x)* is calculated as:

|  |  |
| --- | --- |
|  |  |
| Figure 0‑5. A sample Discreet Uniform distribution. |

As noted above, ADSM sometimes rounds values from continuous distributions to the nearest integer value, when they are applied to discrete values like the duration in days of a disease state. For parameters like these, the Discrete Uniform distribution might be a better choice than the Uniform distribution in preventing excessive errors due to rounding.

The Exponential distribution

An Exponential distribution (figure 0‑6) is a continuous, highly skewed distribution defined by its mean μ, which must be greater than 0. The probability density function *f(x)* is calculated as:

|  |  |
| --- | --- |
|  |  |
| Figure 0‑6. A sample Exponential distribution. |

The Fixed Value “distribution”

Fixed values are not distributions but are used by ADSM in much the same way. A Fixed values “distribution” is defined by a fixed value *y*. The probability density function *f(x)* for a Fixed values “distribution” always returns this fixed value: *f(x)* = *y*

The Gamma distribution

A Gamma distribution (figure 0‑7) is continuous and defined by two parameters: its shape (α) and its scale (β), in which α and β must be greater than 0. Gamma distributions can take a wide variety of shapes, depending on the values of the shapes and scale parameters. This distribution has a mean of α×β. The probability density function *f(x)* is calculated as:

|  |  |
| --- | --- |
| where Γ is the Gamma function. |  |
| Figure 0‑7. A sample Gamma distribution. |

The Gaussian (Normal) distribution

A Gaussian or Normal distribution (figure 0‑8) is a continuous, bell-shaped curve described by two parameters: its mean μ and its standard deviation σ. The standard deviation must be greater than 0. The Gaussian distribution is inherently symmetric. The probability density function *f(x)* is calculated as:

|  |  |
| --- | --- |
|  |  |
| Figure 0‑8. A sample Gaussian distribution. |

The Histogram distribution

The Histogram distribution (figure 0‑9) is a continuous empirical distribution. A histogram distribution directly makes use of data to define its shape and properties, rather than mathematically defining a formula and its parameters. The probability density function *f(x)* is calculated from the set of minimum and maximum values for each histogram bin as follows:

|  |  |
| --- | --- |
| where  and the pi are normalized so that |  |
| Figure 0‑9. A sample Histogram distribution. |

The Hypergeometric distribution

A Hypergeometric distribution (figure 0‑10) is a discrete distribution commonly used to estimate the number of items of type *X* in sample *n* when the sample is drawn from population *M* that has *D* items of type *X*. For example, the number of infected animals in a shipment of animals selected at random from a herd of a particular size with a known prevalence of disease could be modeled using a hypergeometric distribution. In this example, *n* is the number animals in the shipment, *M* is the herd size, and *D* is the herd size times the prevalence. The probability density function *f(x)* for a Hypergeometric distribution is calculated as:

|  |  |
| --- | --- |
|  |  |
| Figure 0‑10. A sample Hypergeometric distribution. |

The Inverse Gaussian distribution

An Inverse Gaussian distribution (figure 0‑11) is a continuous distribution characterized by two parameters: a mean (µ) and a scaling factor (λ). This distribution has been used in epidemiology to model mean infectious and latent periods. The probability density function *f(x)* is calculated as:

|  |  |
| --- | --- |
|  |  |
| Figure 0‑11. A sample Inverse Gaussian distribution. |

The Logistic distribution

A Logistic function (figure 0‑12) is a continuous distribution defined by two parameters: its location α and scale β. The scale parameter must be greater than 0. The probability density function *f(x)* is calculated as:

|  |  |
| --- | --- |
| where *sech* is the hyperbolic secant function |  |
| Figure 0‑12. A sample Logistic distribution. |

The Loglogistic distribution

The Loglogistic function (figure 0‑13) is a continuous distribution defined by three parameters: its shape α, scale β, and location γ. Scale and shape must be greater than 0. The probability density function *f(x)* is calculated as:

|  |  |
| --- | --- |
|  |  |
| Figure 0‑13. A sample Loglogistic distribution. |

The Lognormal distribution

The Lognormal distribution (figure 0‑14) is a logarithmic transformation of the normal distribution. It is described using the same parameters as the normal distribution: its mean μ and its standard deviation σ, where both μ and σ are greater than 0.

The Lognormal distribution, which is continuous, is extremely asymmetric (skewed to the right) when the mean is close to 0; the further the mean is from 0, the more the Lognormal distribution approaches the symmetry and shape of a Normal distribution. The probability density function *f(x)* for a Lognormal distribution is calculated as:

|  |  |
| --- | --- |
| where  and |  |
| Figure 0‑14. A sample Lognormal distribution. |

A Lognormal distribution may also be defined by zeta(ζ) and σ’, as shown previously. ADSM supports both sets of parameters for Lognormal distributions, and automatically handles the conversion of μ and σ to ζ and σ’.

The Negative Binomial distribution

A Negative Binomial distribution (figure 0‑15) is a continuous distribution often used to estimate the number of failures that will occur before there are *s* successes in which there is a probability *p* of success in each trial. The parameters s and p are specified. The probability density function *f(x)* is calculated as:

|  |  |
| --- | --- |
|  |  |
| Figure 0‑15. A sample Negative Binomial distribution. |

The Pareto distribution

A Pareto distribution (figure 0‑16) is a power-law-type probability distribution. A power-law implies that small occurrences are extremely common, whereas large instances are extremely rare. This distribution is heavily skewed to the right and has a mode and minimum that are equal. The distribution starts at the mode *a* and has a rate of decrease determined by the parameter *θ*. The probability density function *f(x)* is calculated as:

|  |  |
| --- | --- |
|  |  |
| Figure 0‑16. A sample Pareto distribution. |

The Pearson 5 distribution

The Pearson 5 distribution (figure 0‑17) is defined by its shape α and scale β, both of which must be greater than 0. A Pearson 5 distribution has a mean of β/(α-1). The probability density function *f(x)* is calculated as:

|  |  |
| --- | --- |
| where Γ is the Gamma function. |  |
| Figure 0‑17. A sample Pearson 5 distribution. |

The Piecewise (General) distribution

A Piecewise or General distribution (figure 0‑18) is an empirical distribution defined by an array of points, each of which has an *x* and a *y* value. Each *x* value must be larger than the previous *x* value. Each *y* value must be at least 0. Finally, the *y* values of the first and last points must be 0. The probability density function *f(x)* is calculated as.

|  |  |
| --- | --- |
| where *xi* ≤ *x* ≤ *xi*+1 |  |
| Figure 0‑18. A sample Piecewise distribution. |

The Poisson distribution

Unlike any of the other distributions described in this appendix, the Poisson distribution (figure 0‑19) is discrete, rather than continuous. Poisson distributions have one specific role in a ADSM scenario: Poisson distributions are used to determine the number of contacts that will be initiated by each herd that is a source of disease.

A Poisson distribution is defined by its mean, designated λ. The probability mass function *f(x)* is calculated as:

|  |  |
| --- | --- |
|  |  |
| Figure 0‑19. A sample Poisson distribution. |

The Triangular distribution

A Triangular distribution (figure 0‑20) is, as its name implies, a triangle. It is described by three parameters: minimum, peak (mode or “most likely”), and maximum values. The Triangular distribution can be symmetric or asymmetric, depending on the relation of the peak to the minimum and maximum values. The probability density function *f(x)* is calculated as:

 if *Min* ≤ *x* ≤ *Mode*

or

 if *Mode* ≤ *x* ≤ *Max*

|  |  |
| --- | --- |
|  |  |
| Figure 0‑20. Two sample Triangular distributions (symmetric and asymmetric). | |

The Uniform distribution

The Uniform distribution (figure 0‑21) is a rectangular block, indicating that all values within a range occur with equal frequency. It is described by two parameters: the minimum and maximum of the range. The Uniform distribution is inherently symmetric. The probability density function *f(x)* is calculated as:

|  |  |
| --- | --- |
| *f(x)* = 1/(*Max*-*Min*) |  |
| Figure 0‑21. A sample Uniform distribution. |

The Weibull distribution

A Weibull distribution (figure 0‑22) is defined by its shape α and scale β, both of which must be greater than 0. The probability density function *f(x)* is calculated as:

|  |  |
| --- | --- |
|  |  |
| Figure 0‑22. A sample Weibull distribution. |

Useful references probability density functions

The following resources provide additional detailed information about the calculation and application of pdfs:

* Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Booth, M., and Rossi, F. 2004. *GNU Scientific Library Reference Manual*. The GNU Scientific Library (GSL) is a collection of routines for numerical computing. ADSM uses the GSL for many of its probability density functions. GSL source code is distributed under the GNU General Public License. The *Reference Manual* is available online at *<http://www.gnu.org/software/gsl>*. An Adobe PDF file is available for download from *<http://gnuwin32.sourceforge.net/‌packages/gsl.htm>* (follow the documentation download link).
* McLaughlin, M.P. 2000. *A Compendium of Common Probability Density Functions*. A reference document assembled by the creators of the mathematical modeling program *Regress+*. The *Compendium* is available for download as an Adobe PDF file from *<http://www.causascientia.org/math\_stat/Dists/Compendium.html>*.
* Palisade Corporation**.** 2002. *A Concise Summary of @RISK Probability Distribution Functions*. A guide published by the creators of *@RISK*, a risk analysis package for *Microsoft Excel*. The *Summary* is available for download as an Adobe PDF file from *<http://project.zf.jcu.cz/risk/data/distfunc.pdf>*.
* Van Hauwermeiren, M. and Vose, D. 2009. *A Compendium of Distributions*. [ebook]. Vose Software, Ghent, Belgium. The *Compendium* is available for download as an Adobe PDF file from *<http://www.vosesoftware.com/content/ebook.pdf>*.
* Vose, D.. 2000. *Risk Analysis: A Quantitative Guide*, 2nd edition. New York: John Wiley & Sons. Vose’s book includes a very useful chapter on probability density functions and their applications.