VOSE SOFTWARE
RISK SOFTWARE SPECIALISTS

ModelRisk 3.0
The most advanced RISK MODELING SOFTWARE in the world

A COMPENDIUM OF DISTRIBUTIONS

Created by
Michael Van Hauwermeiren
&
David Vose

WWW.VOSESOFTWARE.COM

If you notice any errors or omissions, please contact david@vosesoftware.com

Referencing this document

Please use the following reference, following the Harvard system of referencing:

> Van Hauwermeiren M and Vose D (2009). A Compendium of Distributions. [ebook]. Vose Software, Ghent, Belgium. Available from www.vosesoftware.com. Accessed dd/mm/yy.

Table of Contents

Introduction

The precision of a risk analysis relies very heavily on the appropriate use of probability distributions to accurately represent the uncertainty and variability of the problem. In our experience, inappropriate use of probability distributions has proved to be a very common failure of risk analysis models. It stems, in part, from an inadequate understanding of the theory behind probability distribution functions and, in part, from failing to appreciate the knock-on effects of using inappropriate distributions. This compendium is intended to alleviate the misunderstanding by providing a practical insight into the various types of probability distributions in common use.

This compendium gives a very complete summary of the distributions used in risk analysis, an explanation of where and why they are used, some representative plots and the most useful descriptive formulae from a risk analysis viewpoint. The distributions are given in alphabetical order. The list comprises all the distributions that we have ever used at Vose Consulting (and have therefore included in ModelRisk), so we are pretty confident that you will find the one you are looking for. Distributions often have several different names depending on the application so if you don't find the distribution you are searching for here do a search and you may find the alternative name is noted.

Most risk analysis and statistical software offer a wide variety of distributions, so the choice can be bewildering. This compendium therefore starts with a list of different applications, and the distributions that you might find most useful. Then we offer a little guide on how to read the probability equations that feature so prominently in this compendium: people's eyes tend to glaze over when they see probability equations but with a few simple rules you will be able to rapidly 'read' the relevant parts of an equation and ignore the rest, which can give you a much more intuitive feel for a distribution's behavior.

DISCRETE AND CONTINUOUS DISTRIBUTIONS

The most basic distinguishing property between probability distributions is whether they are continuous or discrete

Discrete Distributions

A discrete distribution may take one of a set of identifiable values, each of which has a calculable probability of occurrence. Discrete distributions are used to model parameters like the number of bridges a roading scheme may need, the number of key personnel to be employed or the number of customers that will arrive at a service station in a hour. Clearly, variables such as these can only take specific values: one cannot build half a bridge, employ 2.7 people or serve 13.6 customers.

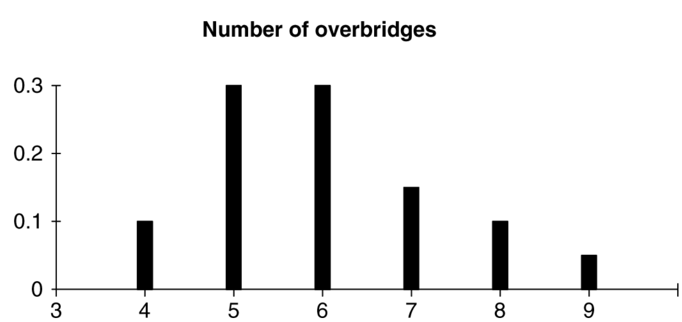


Figure 1 Example of a discrete variable.

The vertical scale of a relative frequency plot of a discrete distribution is the actual probability of occurrence, sometimes called the *probability mass*. The sum of all these values must add up to one. Examples of discrete distributions are: Binomial, Geometric, Hypergeometric, Inverse Hypergeometric Negative Binomial, Poisson and, of course, the Discrete distribution. Figure 1 illustrates a Discrete distribution modeling the number of footbridges that are to be built across a planned stretch of motorway. There is a 30% chance that six bridges will be built, a 10% chance eight bridges will be built, etc. The sum of these probabilities (10% + 30% + 30% + 15% + 10% + 5%) must equal unity.

Continuous Distributions

A continuous distribution is used to represent a variable that can take any value within a defined range (domain). For example, the height of an adult English male picked at random has a continuous distribution because the height of a person is essentially infinitely divisible. We could measure his height to the nearest centimeter, millimeter, tenth of a millimeter, etc. The scale can be repeatedly divided up generating more and more possible values.

Properties like time, mass and distance, that are infinitely divisible, are modeled using continuous distributions. In practice, we also use continuous distributions to model variables that are, in truth, discrete but where the gap between allowable values is insignificant: for example, project cost (which is discrete with steps of one penny, one cent, etc.), exchange rate (which is only quoted to a few significant figures), number of employees in a large organization, etc.

The vertical scale of a relative frequency plot of an input continuous probability distribution is the probability density. It does not represent the actual probability of the corresponding x-axis value since that probability is zero. Instead, it represents the probability per x-axis unit of generating a value within a very small range around the x-axis value.

In a continuous relative frequency distribution, the area under the curve must equal one. This means that the vertical scale must change according to the units used for the horizontal scale. For example, the probability in Figure 2(a) shows a theoretical distribution of the cost of a project using Normal(£4 200 000, £350 000). Since this is a continuous distribution, the cost of the project being precisely £4M is zero. The vertical scale reads a value of $9.7 \times 10–7$ (about one in a million). The x-axis units are £1, so this y-axis reading means that there is a one in a million chance that the project cost will be £4M plus or minus 50p (a range of £1). By comparison, Figure 2(b) shows the same distribution but using million pounds as the scale, i.e. Normal(4.2,0.35). The y-axis value at x = £4M is 0.97, one million times the above value. This does not however mean that there is a 97% chance of being between £3.5M and £4.5M, because the probability density varies very considerably over this range. The logic used in interpreting the $9.7 \times 10–7$ value for Figure 2(a) is an approximation that is valid there because the probability density is essentially constant over that range (£4M ± 50p).

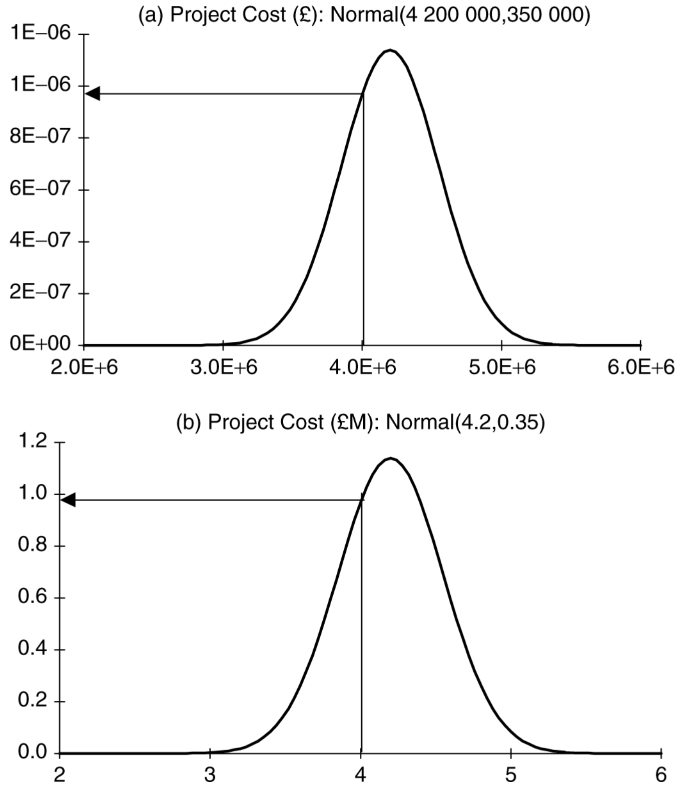


Figure 2 Example of the effect on the y-scale of a probability density function by changing the x-scale units: (a) units of £1, (b) units of £1M.

BOUNDED AND UNBOUNDED DISTRIBUTIONS

A distribution that is confined to lie between two determined values is said to be bounded. Examples of bounded distributions are: Uniform – between minimum and maximum, Triangle – between minimum and maximum, Beta – between 0 and 1, and Binomial – between 0 and n.

A distribution that is unbounded theoretically extends from minus infinity to plus infinity. Examples are: Normal, Logistic, and Extreme Value.

A distribution that is constrained at either end is said to be partially bounded. Examples are: Chi Squared (>0), Exponential (>0), Pareto (>a), Poisson (≥0) and Weibull (>0).

Unbounded and partially bounded distributions may, at times, need to be constrained to remove the tail of the distribution so that nonsensical values are avoided. For example, using a Normal distribution to model sales volume opens up the chance of generating a negative value. If the probability of generating a negative value is significant, and we want to stick to using a Normal distribution, we must constrain the model in some way to eliminate any negative sales volume figure being generated.

Monte Carlo simulation software like ModelRisk usually provides truncated distributions for this purpose as well as filtering facilities. ModelRisk uses the XBounds or PBounds functions to bound a univariate distribution at either specified values or percentiles respectively. One can also build logic into the model that rejects nonsensical values. For example, using the IF function: A2 := IF(A1 < 0, ERR(), 0) only allows values into cell A2 from cell A1 that are ≥0 and produces an error in cell A2 otherwise. However, if there are several distributions being bounded this way, or you are using extreme bounds, you will lose a lot of the iterations in your simulation. If you are faced with the problem of needing to constrain the tail of a distribution, it is also worth questioning whether you are using the appropriate distribution in the first place.

PARAMETRIC AND NON-PARAMETRIC DISTRIBUTIONS

There is a very useful distinction to be made between model-based parametric and empirical non-parametric distributions. By model-based, we mean a distribution whose shape is born of the mathematics describing a theoretical problem. For example: an exponential distribution is a waiting time distribution whose function is the direct result of assuming that the there is a constant instantaneous probability of an event occurring; a lognormal distribution is derived from assuming that ln[x] is Normally distributed, etc.
By “empirical distribution” we mean a distribution whose mathematics is defined by the shape that is required. For example, a Triangle distribution is defined by its minimum, mode and maximum values; a Histogram distribution is defined by its range, the number of classes and the height of each class. The defining parameters for general distributions are features of the graph shape. Empirical distributions include: Cumulative, Discrete, Histogram, Relative, Triangle and Uniform.

Those distributions that fall under the “empirical distribution” or non-parametric class are intuitively easy to understand, extremely flexible and are therefore very useful. Model-based or parametric distributions require a greater knowledge of the underlying assumptions if they are to be used properly. Without that knowledge, the analyst may find it very difficult to justify the use of the chosen distribution type and to gain peer confidence in his or her model. S/he will probably also find it difficult to make alterations should more information become available.

Vose Consulting’s analysts are keen advocates of using non-parametric distributions. We believe that parametric distributions should only be selected if either: (a) the theory underpinning the distribution applies to the particular problem, (b) it is generally accepted that a particular distribution has proven to be very accurate for modeling a specific variable without actually having any theory to support the observation, (c) the distribution approximately fits the expert opinion being modeled (or the estimated moments) and the required level of accuracy is not very high, or (d) one wishes to use a distribution that has a long tail extending beyond the observed minimum or maximum.

Univariate and multivariate distributions

Univariate distributions describe a single parameter or variable and are used to model a parameter or variable that is not probabilistically linked to any other in the model. Multivariate distributions describe several parameters whose values are probabilistically linked in some way. In most cases, we create the probabilistic links via one of several correlation methods. However, there are a few multivariate distributions that have specific, very useful purposes and are therefore worth studying more.

Lists of applications and the most useful distributions

Bounded versus unbounded

The following tables organize distributions according to whether their limits are bounded. ***Italics*** indicate non-parametric distributions.

Univariate Distributions

	Continuous	***Discrete***
Unbounded	Cauchy Error function Error Hyperbolic Secant JohnsonU Laplace Logistic Normal Student T F	
Left bounded	Bradford Burr Chi Chi Squared Dagum Erlang Exponential Extreme value Max Extreme value Min Fatigue Life Fréchet Gamma Generalised Logistic Inverse Gaussian LogGamma LogLogistic LogLaplace Lognormal Lognormal (base E) Lognormal (base B) Pareto (1st kind) Pareto (2nd kind) Pearson 5 Pearson 6 Rayleigh Weibull	Beta-geometric Beta-Negative Binomial Delaporte Geometric Inverse Hypergeometric Logarithmic Negative Binomial Poisson Polya
Left and right bounded	Beta Beta4 *Cumulative ascending* *Cumulative descending* *Histogram* JohnsonB	Bernoulli Beta-binomial Binomial *Discrete* *Discrete uniform* Hypergeometric

	Kumaraswamy Kumaraswamy4 *Modified PERT* Ogive *PERT* Reciprocal *Relative* *Split Triangle* *Triangle* *Uniform*	*Step uniform*

Multivariate Distributions

	Continuous	***Discrete***
Unbounded	Multivariate Normal	
Left bounded		Negative Multinomial 1 Negative Multinomial 2
Left and right bounded	Dirichlet	Multinomial Multivariate Hypergeometric Multivariate Inv Hypergeometric 1 Multivariate Inv Hypergeometric 2

Frequency distributions

The following distributions are often used to model events that are counted, like outbreaks, economic shocks, machine failures, deaths, etc:

Bernoulli	Special case of binomial with one individual who may convert.
Binomial	Used when there is a group of individuals (trials) that may convert (succeed). For example, used in life insurance to answer how many policyholders will claim in a year
Delaporte	Events occur randomly with a randomly varying risk level, the most flexible in modeling frequency patterns
Logarithmic	Peaks at 1, looks exponential
NegBin	Events occur randomly with a randomly varying risk level, more restrictive than the Pólya
Poisson	Events occur randomly with constant risk level
Pólya	Events occur randomly with a randomly varying risk level

Risk impact

A risk is an event that may or may not occur, and the impact distribution describes the 'cost' should it occur. For most applications, a continuous distribution that is right-skewed and left bounded is most applicable. For situations like an FMD outbreak, the number of sheep lost will of course be discrete but such variables are usually modeled with continuous distributions anyway (and you can use ROUND(…,0) of course to convert to integers).

Bradford	Looks exponential with min and max bounds
Burr	Appealing because of its flexibility of shape
Dagum	Lognormal looking with 2 shape parameters for more control, and a scale parameter
Exponential	Ski-slope shape defined by its mean
Ogive	Used to construct a distribution directly from data
LogGamma	Can have very long right tail
LogLogistic	Impact is a function of several variables that are either correlated or one dominates
LogNormal	Impact is a function of several uncorrelated variables
Pareto (2 kinds)	Ski-slope shape with the longest tail so often used to conservatively model the extreme right tail

Time or trials until …

These are either discrete or continuous waiting time distributions. They express how long one must wait before observing a specific event, or a defined number of events.

Beta-Geometric	Failures before 1 beta-binomial success
Beta-Negative Binomial	Failures before *s* beta-binomial successes
Erlang	Time until *m* Poisson counts
Exponential	Time until 1 Poisson count
Fatigue Life	Time until gradual breakdown
Geometric	Failures before 1 binomial success
Gamma	Time until [illegible] Poisson counts but used more generally
Inverse Hypergeometric	Failures before *s* hypergeometric successes
Inverse Gauss	Theoretical waiting time use is esoteric, but the distribution has some flexibility through its parameters
Lognormal	Time until an event that is the product of many variables. Used quite generally
Negative Binomial	Failures before *s* binomial successes
Negative Multinomial	Failures before *s* multinomial successes
Rayleigh	A special case of the Weibull. Also distance to nearest, Poisson distributed, neighbor
Weibull	Time until an event where the instantaneous likelihood of the event occurring changes (usually increases) over time. Used a great deal in reliability engineering.

Variations in a financial market

Analysts used to be happy to assume that random variations in a stock's return, an interest rate, were normally distributed. Normal distributions made the equations easier. Financial analysts now use simulation more so have become a bit more adventurous:

Extreme Value (max, min)	Models the extreme movement but tricky to use
Generalized Error (aka GED, Error)	Very flexible distribution that will morph between a uniform (nearly) a Normal, Laplace, etc
Inverse Gaussian	Used in place of Lognormal when it has right tail too heavy
Laplace	Defined by mean and variance like the Normal, but takes a tent shape. Favored because it gives longer tails.

Lévy	Appealing because it belongs to the 'Stable' family of distributions, gives fatter tails than a Normal
Logistic	Like a Normal but more peaked
LogNormal	Assumes that market is randomly affected by very many multiplicative random elements
Normal	Assumes that market is randomly affected by very many additive random elements
Poisson	Used to model the occurrence of jumps in the market
Student	When rescaled and shifted, it is like the Normal but with more kurtosis when is small

How big something is

How much milk a cow will produce, how big a sale will be, how big a wave, etc. We'll often have data and want to fit them to a distribution, but which one?

Bradford	Like a truncated Pareto. Used in advertising, but take a look
Burr	Appealing because of its flexibility of shape
Dagum	Flexible. Has been used to model aggregate fire losses.
Exponential	Ski-slope shape peaking at zero and defined by its mean.
Extreme Value	Models extremes (min, max) of variables belonging to the Exponential family of distributions. Difficult to use. VoseLargest, VoseSmallest much more flexible and transparent.
Generalized Error (aka GED, Error)	Very flexible distribution that will morph between a uniform (nearly) a Normal, Laplace, etc
Hyperbolic-Secant	Like a Normal but with narrower shoulders, so used to fit to data where a Normal isn't quite right
Inverse Gaussian	Used in place of Lognormal when it has right tail too heavy
Johnson Bounded	Can have any combination of skewness and kurtosis so pretty flexible at fitting to data, but rarely used
LogGamma	If the variable is the product of a number of Exponentially distributed variables it may look LogGamma distributed
LogLaplace	The asymmetric LogLaplace distribution takes a pretty strange shape, but has a history of being fitted to particle size data and similar
LogLogistic	Has a history of being fitted to data for a fair few financial variables
LogNormal	See Central Limit Theorem. Size is a function of the product of a number of random variables. Classic example is oil reserves = Area*Thickness*Porosity*Gas:Oil Ratio*Recovery Rate
Normal	See Central Limit Theorem. Size is a function of the sum of a number of random variables, e.g. a cow's milk yield may be a function of genetics & farm care & mental well being (it's been proven) & nutrition & …
Pareto	Ski-slope shape with the longest tail so often used to conservatively model the extreme right tail and generally fits the main body of data badly, so consider splicing (see VoseSplice function, for example)
Rayleigh	Wave heights, electromagnetic peak power or similar.
Student	If the variance of a Normal distribution is also a random variable (specifically Chi Squared) the variable will take a Student distribution. So think about something that should be roughly Normal with constant mean but where the standard deviation is not constant, e.g. errors in measurement with varying quality of instrument or operator.
Weibull	Much like the Rayleigh, including modeling wind speed

Expert estimates

The following distributions are often used to model subject matter experts' estimates because they are intuitive, easy to control and/or flexible:

Bernoulli	Use to model a risk event occurring or not
Beta4	A min, max and 2 shape parameters. Can be reparameterised (i.e. the PERT distribution). Shape parameters are difficult to use.
Bradford	A min, max and a ski-slope shape between with controllable drop.
Combined	Allows you to combine correctly several SME estimates for the same parameter and weight them
Cumulative (Ascending and Descending)	Good when expert thinks of a series of "probability P of being below x".
Discrete	Specify several possible outcomes with weights for each
Johnson Bounded	VISIFIT software available that will match to expert estimate
Kumaraswamy	Controllable distribution, similar to Beta4.
Modified PERT	PERT distribution with extra control for spread.
PERT	A min, max and mode. Tends to place too little emphasis on tails if distribution is quite skewed.
Relative	Allow you to construct your own shape
Split Triangle	Defined by low, medium and high percentiles. Splices two Triangle distributions together. Intuitive.
Triangle	A min, mode and max. Some software also offer low and high percentiles as inputs. Tends to over-emphasize tails.
Uniform	A min and max. Useful to flag when SME has very little idea.

How to read probability distribution equations

The intention of this section is to help you better understand how to read and use the equations that describe distributions. For each distribution in this compendium we give the following equations:

- Probability mass function (for discrete distributions);
- Probability density function (for continuous distributions);
- Cumulative distribution function (where available);
- Mean;
- Mode;
- Variance;
- Skewness; and
- Kurtosis.

There are many other distribution properties (e.g. moment generating functions, raw moments), but they are of little general use in risk analysis, and would leave you facing yet more daunting pages of equations to wade through.

Location, scale and shape parameters

In this compendium, and in ModelRisk, we have parameterized distributions to reflect the most common usage, and where there are two or more common parameterizations we have used the one that is most useful to model risk. So we use, for example, mean and standard deviation a lot for consistency between distributions, or other parameters that most readily connect to the stochastic process that the distribution is most commonly applied to. Another way to describe parameters is to categorize them as location, scale and shape, which can disconnect the parameters from their usual meaning, but is sometimes helpful in understanding how a distribution will change with variation of the parameter value.

A location parameter controls the position of the distribution on the x-axis. It should therefore appear in the same way in the equations for the mode and mean – two measures of location. So, if a location parameter increases by 3 units, then the mean and mode should increase by three units. For example, the mean μ of a Normal distribution is also the mode, and can be called a location parameter. The same applies for the Laplace, for example. A lot of distributions are extended by including a shift parameter (e.g. VoseShift in ModelRisk), which has the effect of moving the distribution along the x-axis and is a location parameter.

A scale parameter controls the spread of the distribution on the x-axis. Its square should therefore appear in the equation for a distribution's variance. For example, β is the scale parameter for the Gamma, Weibull and Logistic distributions, σ for the Normal and Laplace distributions, b for the Extreme ValueMax, ExtremeValueMin and Rayleigh distributions, etc.

A shape parameter controls the shape (e.g. skewness, kurtosis) of the distribution. It will appear in the probability function in a way that controls the manipulation of x in a non-linear fashion, usually as a coefficient of *x*.

For example, the Pareto distribution has probability density function:

$$f(x) \propto \frac{1}{x^{\theta+1}}$$

ϑ is a shape parameter here as it changes the functional form of the relationship between *f(x)* and *x*. Other examples you can look at are ν for a GED, Student and the ChiSq distribution, and α for a Gamma distribution. A distribution may sometimes have two shape parameters, e.g. α_1 and α_2 for the Beta distribution, v_1 and v_2 for the F distribution.

If there is no shape parameter the distribution always takes the same shape (like the Cauchy, Exponential, Extreme Value, Laplace, Logistic and Normal).

Understanding distribution equations

Probability mass function (pmf) and probability density function (pdf)

The pmf or pdf is the most common equation used to define a distribution, for two reasons. The first is that it gives the shape of the density (or mass) curve, which is the easiest way to recognize and review a distribution. The second is that the pmf (or pdf) is always in a useful form, whereas the cdf frequently doesn't have a closed form (meaning a simple algebraic identity rather than expressed as an integral or summation).

Pmf's must sum to 1, and pdf's must integrate to 1, in order to obey the basic probability rule that the sum of all probabilities = 1. This means that a pmf or pdf equation has two parts: a function of x, the possible value of the parameter; and a multiplying factor that normalizes the distribution to sum to unity. For example, the Error distribution pdf takes the (rather complicated) form:

$$f(x) = \frac{K}{\beta}\exp\left[-\frac{1}{2}\left|\frac{x-\mu}{\beta}\right|^{\nu}\right] \qquad (1)$$

where $K = \dfrac{\nu}{\Gamma\left(\frac{1}{\nu}\right)2^{1+\frac{1}{\nu}}}$ and $\beta = \dfrac{\sigma}{2^{\frac{1}{\nu}}}\sqrt{\dfrac{\Gamma\left(\frac{1}{\nu}\right)}{\Gamma\left(\frac{3}{\nu}\right)}}$

The part that varies with x is simply $\exp\left[-\frac{1}{2}\left|\frac{x-\mu}{\beta}\right|^{\nu}\right]$ so we can write:

$$f(x) \propto \exp\left[-\frac{1}{2}\left|\frac{x-\mu}{\beta}\right|^{\nu}\right] \qquad (2)$$

The rest of Equation (1), i.e. K/β, is a normalizing constant for a given set of parameters and ensures the area under the curve equals unity. Equation (2) is sufficient to define or recognize the distribution and allows us to concentrate on how the distribution behaves with changes to the parameter values. In fact probability mathematicians and Bayesian statisticians frequently work with just the component that is a function of *x*, keeping in the back of their mind that it will be normalized eventually.

For example, the $(x-\mu)/\beta$ part shows us that the distribution is shifted μ along the x-axis (a location parameter), and the division by β means that the distribution is rescaled by this factor (a scale parameter). The parameter ν changes the functional form of the distribution. For example, for ν = 2:

$$f(x) \propto \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\beta}\right)^2\right]$$

Compare that to the Normal distribution density function:

$$f(x) \propto \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

So we can say that when ν = 2, the GED is Normally distributed with mean μ and standard deviation σ. The functional form (the part in *x*) gives us sufficient information to say this, as we know the multiplying constant must adjust to keep the area under the curve equal to unity.

Similarly, for ν = 1 we have:

$$f(x) \propto \exp\left[-\frac{1}{2}\left|\frac{x-\mu}{\beta}\right|\right]$$

Compare that to the Laplace distribution.

So we can say that when $v = 1$, the GED takes a Laplace(μ,σ) distribution.

The same idea applies to discrete distributions. For example, the Logarithmic(ϑ) distribution has pmf:

$$f(x) = \frac{-\theta^x}{x\ln(1-\theta)} \propto \frac{-\theta^x}{x}$$

$\dfrac{1}{\ln(1-\theta)}$ is the normalizing part because, it turns out that log(x) can be expressed as an infinite series so that:

$$\sum_{x=1}^{\infty} \frac{-\theta^x}{x} = \ln(1-\theta)$$

Cumulative distribution function (cdf)

The cdf gives us the probability of being less than or equal to the variable value *x*. For discrete distributions this is simply the sum of the pmf up to *x*, so reviewing its equation is not more informative than the pmf equation. However, for continuous distributions the cdf can take a simpler form than the corresponding pdf. For example, for a Weibull distribution:

$$f(x) = \alpha\beta^{-\alpha} x^{\alpha-1} \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right] \propto x^{\alpha-1} \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right]$$

$$F(x) = 1 - \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right] \qquad (3)$$

The latter is simpler to envisage.

Many cdfs have a component that involves the exponential function (e.g. Weibull, Exponential, Extreme Value, Laplace, Logistic, Rayleigh). Exp(- ∞) = 0 and Exp(0) = 1 which is the range of *F(x)*, so you'll often see functions of the form:

$$F(x) = Exp(-g(x))$$

or

$$F(x) = 1 - Exp(-g(x))$$

where *g(x)* is some function of *x* that goes from zero to infinity or infinity to zero monotonically (meaning always increasing) with increasing *x*. For example, Equation (3) for the Weibull distribution shows us:

- The value β scales *x*
- When *x* = 0, *F(x)* = 1-1 = 0, so the variable has a minimum of 0
- When *x* = ∞, *F(x)* = 1-0 = 1, so the variable has a maximum of ∞
- α makes the distribution shorter, because it 'amplifies' *x*. For example (leaving β=1), if α = 2 and *x* = 3 it calculates 3^2=9, whereas if α = 4 it calculates 3^4=81

Mean μ

The mean of a probability distribution is useful to know for several reasons.

- It gives a sense of the location of the distribution;
- Central Limit Theorem (CLT) uses the mean;
- Knowing the equation of the mean can help us understand the distribution. For example, a Gamma(α,β) distribution can be used to model the time to wait to observe α independent events that occur randomly in time with a mean time to occurrence of β. It makes intuitive sense that, 'on average', you need to wait α*β, which is the mean of the distribution;

- We sometimes want to approximate one distribution with another to make the mathematics easier. Knowing the equations for the mean and variance, even the skewness and kurtosis too, can help us find a distribution with these same moments;
- Because of CLT, the mean propagates through a model much more precisely than the mode or median. So, for example, if you replaced a distribution in a simulation model with its mean the output mean value will usually be close to the output mean when the model includes that distribution. However, the same does not apply as well by replacing a distribution with its median, and often much worse still if one uses the mode;
- We can determine the mean and other moments of the an aggregate distribution if we know the mean and other moments of the frequency and severity distributions; and
- A distribution is often fitted (though not in ModelRisk) to data by matching the data's mean and variance to the mean and variance equations of the distribution – a technique known as Method of Moments.

When the pdf of a distribution is of the form $f(x) = g(x - z)$ where $g(\ )$ is any function and z is a fixed value, the equation for the mean will be a linear function of z.

Mode

The mode is the location of the peak of a distribution, and is the most intuitive parameter to consider – the 'most likely value to occur'.

If the mode has the same equation as the mean it tells us the distribution is symmetric. If the mode is less than the mean (e.g. for the Gamma distribution, mode = $(\alpha-1)\beta$ and mean = $\alpha\beta$) we know the distribution is right-skewed, if the mode is greater than the mean the distribution is left-skewed. The mode is our 'best guess' so it can be informative to see how the mode varies with the distribution's parameters. For example, the Beta(α, β) has a mode of

$$\frac{\alpha-1}{\alpha+\beta-2} \text{ if } \alpha>1, \beta>1$$

A Beta(s+1, n-s+1) distribution is often used to estimate a binomial probability where we have observed s successes in n trials. This gives a mode of s/n: the fraction of our trials that were successes is our 'best guess' at the true (long run) probability, which makes intuitive sense.

Variance V

The variance gives a measure of the spread of a distribution. We give equations for the variance rather than the mean because it avoids having square-root signs all the time, and because probability mathematicians work in terms of variance rather than standard deviation. However, it can be useful to take the square root of the variance equation (i.e. the standard deviation σ) to help make more sense of it. For example, the Logistic(α,β) distribution has variance:

$$V = \frac{\beta^2\pi^2}{3}$$

so

$$\sigma = \sqrt{V} = \frac{\beta\pi}{\sqrt{3}}$$

which shows us that *β* is a scaling parameter: the distribution's spread is proportional to *β*. Another example - the Pareto(ϑ,a) distribution has variance:

$$V = \frac{a^2\theta}{(\theta-1)^2(\theta-2)}$$

so

$$\sigma = a\sqrt{\frac{\theta}{(\theta-1)^2(\theta-2)}}$$

which shows us that *a* is a scaling parameter.

Skewness S

Skewness and kurtosis equations are not that important, so feel free to skip this bit. Skewness is the expected value of $(x-\mu)^3$ divided by $V^{3/2}$, so you'll often see a $\frac{\ldots}{\sqrt{\ldots}}$ or $\frac{\ldots}{(\ldots)^{3/2}}$ component. You can tell whether a distribution is left or right skewed and when by looking at this equation, bearing in mind the possible values of each parameter. For example, the skewness equation for the Negative Binomial is:

$$\frac{2-p}{\sqrt{s(1-p)}}.$$

Since p lies on (0,1) and s is a positive integer, the skewness is always positive. The Beta distribution has skewness:

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

and since α and β are > 0, this means because of the (β-α) term it has negative skewness when α > β, positive skewness when α < β and zero skewness when α = β. An Exponential distribution has a skewness of 2, which you might find a helpful gauge to compare against (i.e. is the distribution more scaled than an Exponential).

Kurtosis K

Kurtosis is the expected value of $(x-\mu)^4$ divided by V^2, so you'll often see a $\frac{\ldots}{(\ldots)^2}$ component. The Normal distribution has a kurtosis of 3 and that's what we usually compare against (the Uniform distribution has a kurtosis of 1.8, the Laplace a kurtosis of 6, which are two fairly extreme points of reference). The Poisson(λ) distribution, for example, has a kurtosis of 3+1/λ which means, when taken together with the behavior of other moments, the bigger the value of λ the closer the distribution is to a Normal. The same story applies for the Student(ν) distribution, for example, which has a kurtosis of 3(ν-2)/(ν-4) so the larger ν the closer the kurtosis is to 3. The kurtosis of a Lognormal(μ,σ) distribution is $z^4 + 2z^3 + 3z^2 - 3$ where $z = 1 + \frac{\mu}{\sigma}$: what does that imply about when the Lognormal will look Normal?

The distributions

Compiled by
Michael van Hauwermeiren and David Vose
Vose Software BVBA

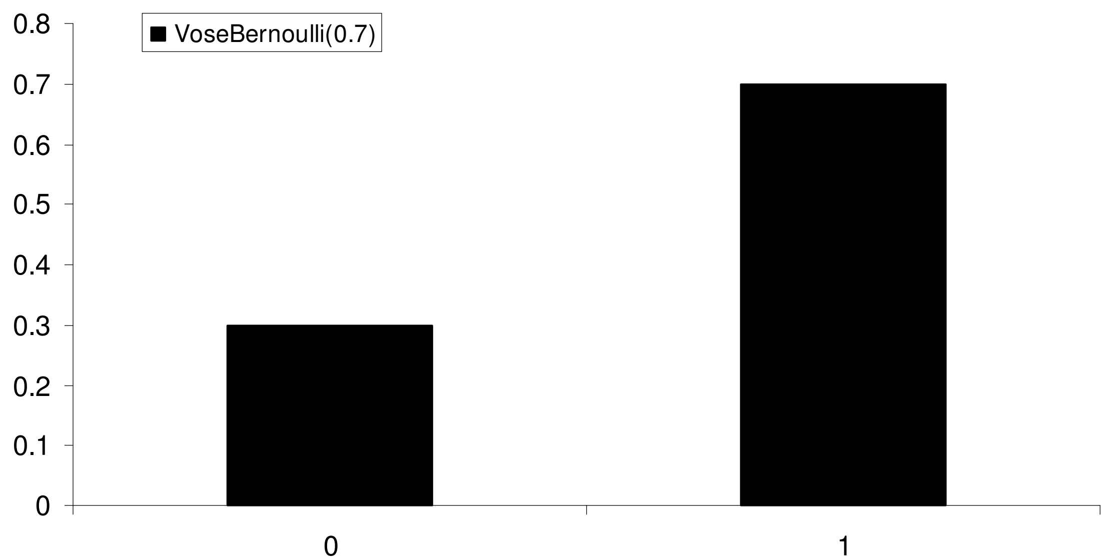
Univariate Distributions

Bernoulli

VoseBernoulli(p)

Graphs

The Bernoulli distribution is a Binomial distribution with n = 1. The Bernoulli distribution returns a 1 with probability p and a zero otherwise.



Uses

The Bernoulli distribution, named after Swiss scientist Jakob Bernoulli, is very useful for modeling a risk event that may or may not occur.

VoseBernoulli(0.2)*VoseLognormal(12,72) models a risk event with a probability of 20% of occurring and an impact, should it occur, equal to Lognormal(12,72).

Equations

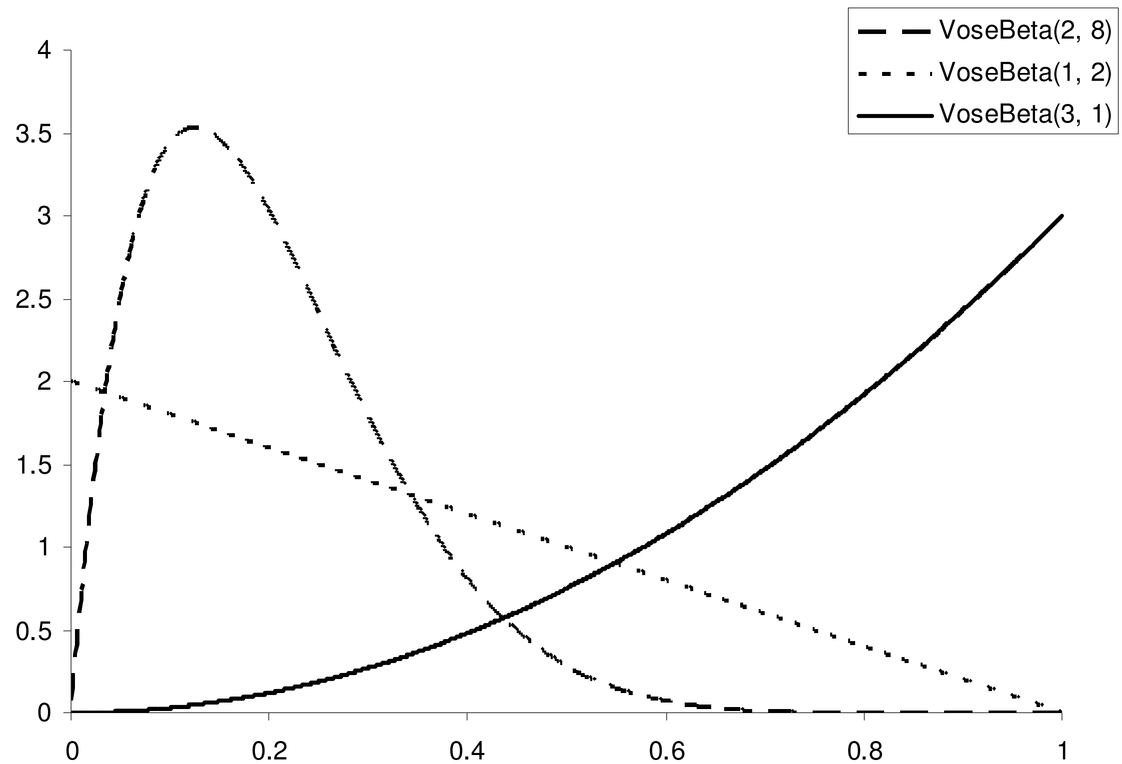
Probability mass function :	$f(x) = p^x (1-p)^{1-x}$
Cumulative distribution function :	$F(0) = 1-p, F(1) = 1$
Parameter restriction :	$0 \le p \le 1$
Domain :	$x = \{0,1\}$
Mean :	p

Mode :	$\lfloor 2p \rfloor$
Variance :	$p(1-p)$
Skewness :	$\dfrac{1-2p}{\sqrt{p(1-p)}}$
Kurtosis :	$\dfrac{1}{p(1-p)}-3$

Beta

VoseBeta(α,β)

Graphs



Uses

The Beta distribution has two main uses:

- As the description of uncertainty or random variation of a probability, fraction or prevalence;
- As a useful distribution one can rescale and shift to create distributions with a wide range of shapes and over any finite range. As such, it is sometimes used to model expert opinion, for example in the form of the PERT distribution.

The Beta distribution is the conjugate prior (meaning it has the same functional form, therefore also often called "convenience prior") to the Binomial likelihood function in Bayesian inference and, as such, is often used to describe the uncertainty about a binomial probability, given a number of trials n have been made with a number of recorded successes s. In this situations, α is set to the value (s + x) and β is set to (n - s + y), where Beta(x, y) is the prior.

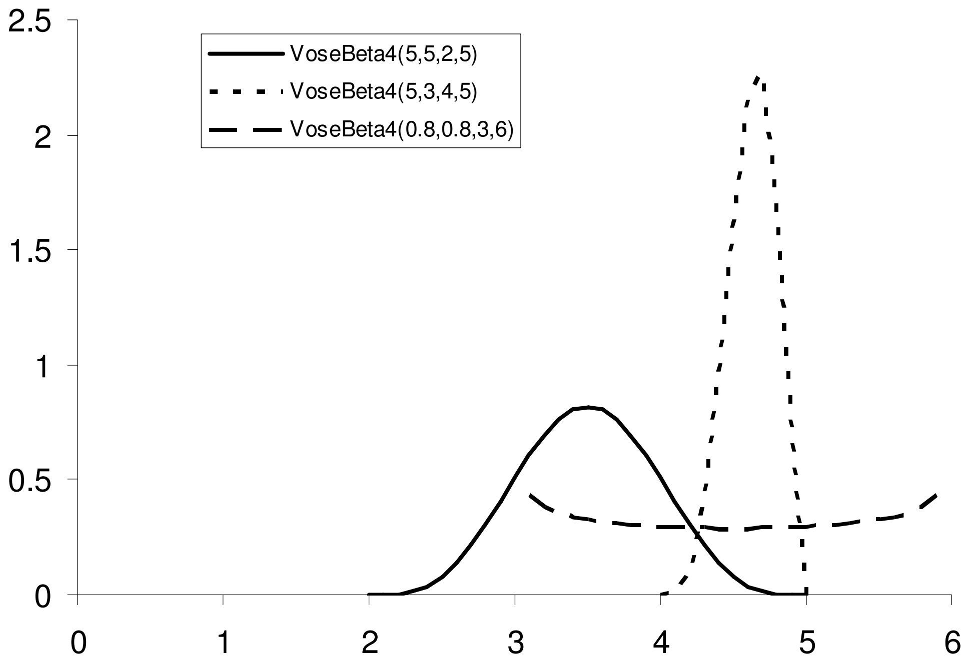
Equations

Probability density function :	$f(x)=\frac{(x)^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$ *where* $B(\alpha,\beta)$ *is a Beta function*
Cumulative distribution function :	*No closed form*
Parameter restriction :	$\alpha>0,\beta>0$
Domain :	$0<x<1$
Mean :	$\frac{\alpha}{\alpha+\beta}$
Mode :	$\frac{\alpha-1}{\alpha+\beta-2}$ *if* $\alpha>1,\beta>1$ $0,1$ *if* $\alpha<1,\beta<1$ 0 *if* $\alpha<1,\beta\geq 1$ *or if* $\alpha=1,\beta>1$ 1 *if* $\alpha\geq 1,\beta<1$ *or if* $\alpha>1,\beta=1$ *does not uniquely exist if* $\alpha=1,\beta=1$
Variance :	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Skewness :	$2\frac{\beta-\alpha}{\alpha+\beta+2}\sqrt{\frac{\alpha+\beta+1}{\alpha\beta}}$
Kurtosis :	$3\frac{(\alpha+\beta+1)\left(2(\alpha+\beta)^2+\alpha\beta(\alpha+\beta-6)\right)}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)}$

Four-parameter Beta

VoseBeta4(α,β,min,max)

Graphs



Uses

See Beta distribution

Equations

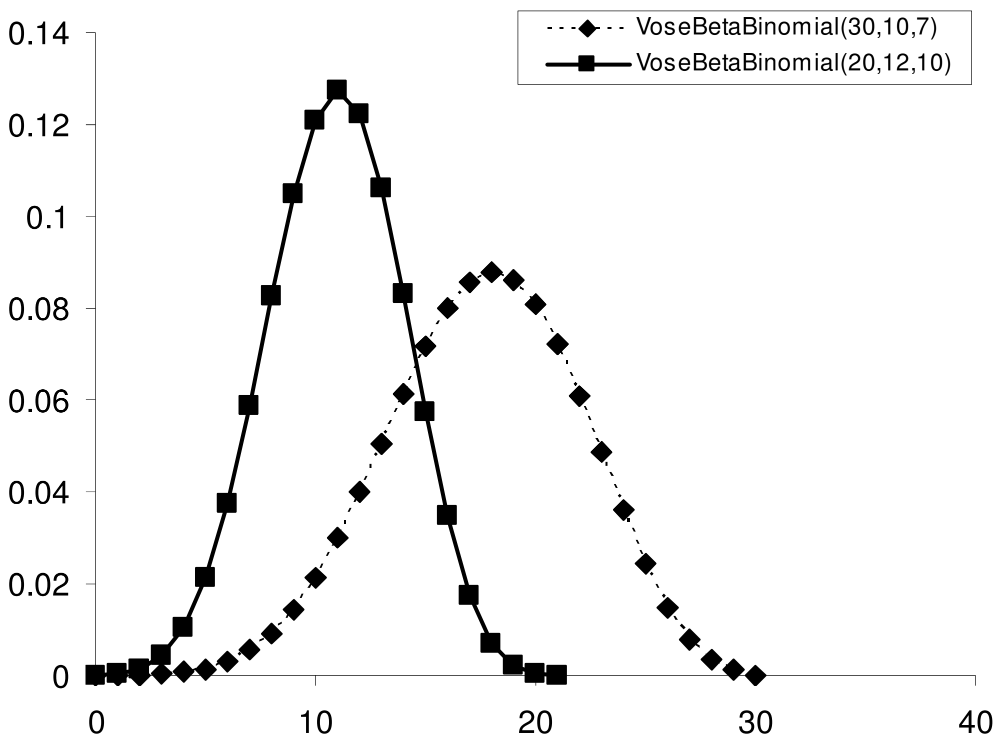
Probability density function :	$f(x)=\dfrac{(x-\min)^{\alpha-1}(\max-x)^{\beta-1}}{B(\alpha,\beta)(\max-\min)^{\alpha+\beta-1}}$ *where* $B(\alpha,\beta)$ *is a Beta function*
Cumulative distribution function :	*No closed form*
Parameter restriction :	$\alpha>0,\beta>0,\min<\max$
Domain :	$\min\le x\le\max$
Mean :	$\min+\dfrac{\alpha}{\alpha+\beta}(\max-\min)$
Mode :	$\min+\dfrac{\alpha-1}{\alpha+\beta-2}(\max-\min)$ *if* $\alpha>1,\beta>1$ $\min\ ,\ \max$ *if* $\alpha<1,\beta<1$ $\min$ *if* $\alpha<1,\beta\ge1$ *or if* $\alpha=1,\beta>1$ $\max$ *if* $\alpha\ge1,\beta<1$ *or if* $\alpha>1,\beta=1$ *does not uniquely exist* *if* $\alpha=1,\beta=1$
Variance :	$\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}(\max-\min)^2$
Skewness :	$2\dfrac{\beta-\alpha}{\alpha+\beta+2}\sqrt{\dfrac{\alpha+\beta+1}{\alpha\beta}}$
Kurtosis :	$3\dfrac{(\alpha+\beta+1)\left(2(\alpha+\beta)^2+\alpha\beta(\alpha+\beta-6)\right)}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)}$

Beta-Binomial

VoseBetaBinomial(n,α,β)

Graphs

A Beta-Binomial distribution returns a discrete value between 0 and n. Examples of a Beta-Binomial(30,10,7) and a Beta-Binomial(20,12,10) distribution are given below:



Uses

The Beta-Binomial distribution is used to model the number of successes in n binomial trials (usually =Binomial(n,p)) but when the probability of success p is also a *random variable*, and can be adequately described by a Beta(α, β) distribution.

The extreme flexibility of the shape of the Beta distribution means that it is often a very fair representation of the randomness of p.

The probability of success varies randomly, but in any one scenario that probability applies to all trials. For example, you might consider using the Beta-Binomial distribution to model:

- The number of cars that crash in a race of n cars, where the predominant factor is not the skill of the individual driver, but the weather on the day;

- The number of bottles of wine from a producer that are bad where the predominant factor is not how each bottle is treated, but something to do with the batch as a whole;
- The number of people who get ill at a wedding from n invited, all having a taste of the delicious soufflé, unfortunately made with contaminated eggs, where their risk is dominated not by their individual immune system, or the amount they eat, but by the level of contamination of the shared meal.

Comments

The Beta-Binomial distribution is always more spread than its best fitting Binomial distribution, because the Beta distribution adds extra randomness. Thus, when a Binomial distribution does not match observations, because the observations exhibit too much spread, a Beta-Binomial distribution is often used instead.

Equations

Probability mass function :	$f(x)=\binom{n}{x}\frac{(\alpha+x-1)!(n+\beta-x-1)!(\alpha+\beta-1)!}{(\alpha+\beta+n-1)!(\alpha-1)!(\beta-1)!}$
Cumulative distribution function :	$F(x)=\sum_{i=0}^{\lfloor x\rfloor}\binom{n}{i}\frac{(\alpha+i-1)!(n+\beta-i-1)!(\alpha+\beta-1)!}{(\alpha+\beta+n-1)!(\alpha-1)!(\beta-1)!}$
Parameter restriction :	$\alpha>0;\beta>0;n=\{0,1,2,\ldots\}$
Domain :	$x=\{0,1,2,\ldots,n\}$
Mean :	$n\frac{\alpha}{\alpha+\beta}$
Mode :	$\begin{array}{ll} \left\lfloor n\left(\frac{\alpha-1}{\alpha+\beta-2}+\frac{1}{2}\right)\right\rfloor & \text{if } \alpha>1,\beta>1 \\ 0\,,\,n & \text{if } \alpha<1,\beta<1 \\ 0 & \text{if } \alpha<1,\beta\geq1 \text{ or if } \alpha=1,\beta>1 \\ n & \text{if } \alpha\geq1,\beta<1 \text{ or if } \alpha>1,\beta=1 \\ \text{does not uniquely exist} & \text{if } \alpha=1,\beta=1 \end{array}$
Variance :	$n\frac{\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Skewness :	$(\alpha+\beta+2n)\frac{(\beta-\alpha)}{(\alpha+\beta+2)}\sqrt{\frac{(1+\alpha+\beta)}{n\alpha\beta(n+\alpha+\beta)}}$

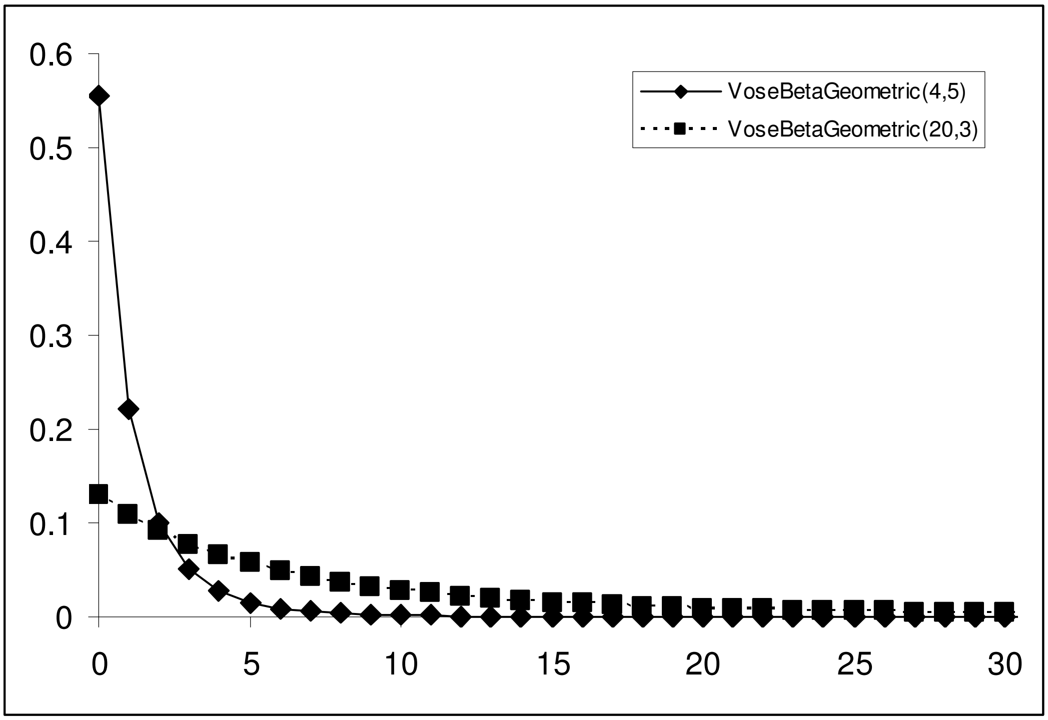
Kurtosis :

$$\frac{(\alpha+\beta)^2(1+\alpha+\beta)}{n\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)(\alpha+\beta+n)}\left((\alpha+\beta)(\alpha+\beta-1+6n)+3\alpha\beta(n-2)+6n^2-\frac{3\alpha\beta n(6-n)}{\alpha+\beta}-\frac{18\alpha\beta n^2}{(\alpha+\beta)^2}\right)$$

Beta-Geometric

VoseBetaGeometric(α,β)

Graphs



Uses

The BetaGeometric(α, β) distribution models the number of failures that will occur in a binomial process before the first success is observed and where the binomial probability p is itself a random variable taking a Beta(α, β) distribution.

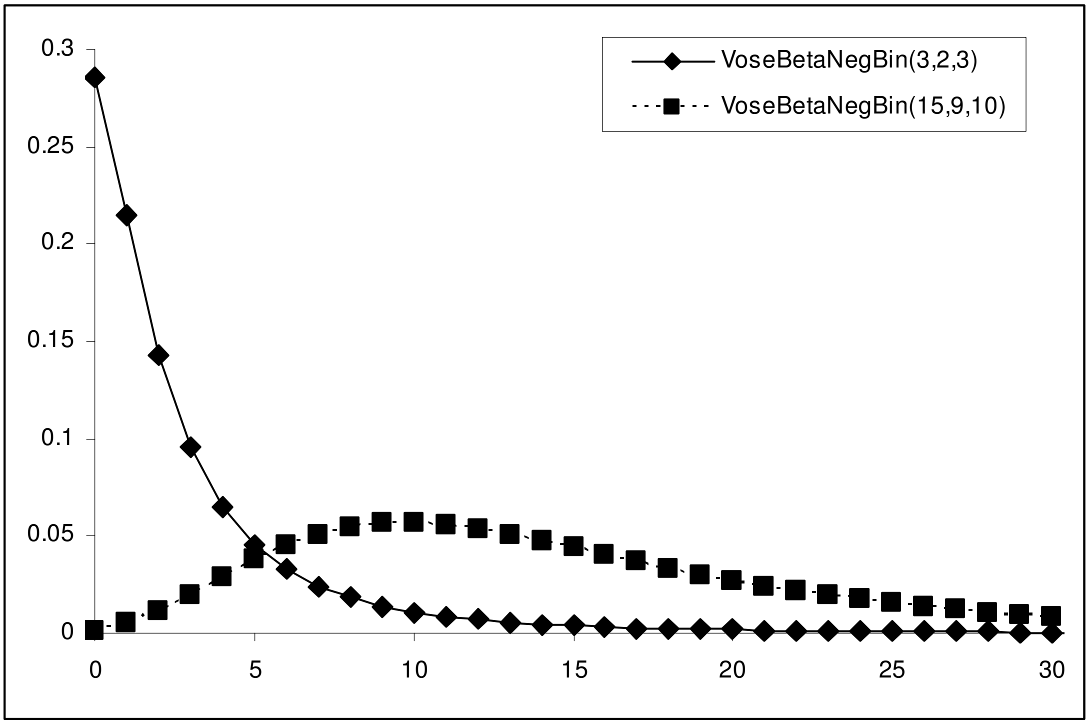
Equations

Probability mass function :	$f(x)=\frac{\beta\,\Gamma(\alpha+\beta)\Gamma(\alpha+x)}{\Gamma(\alpha)\Gamma(\alpha+\beta+x+1)}$
Cumulative distribution function :	$F(x)=\sum_{i=0}^{x}\frac{\beta\,\Gamma(\alpha+\beta)\Gamma(\alpha+i)}{\Gamma(\alpha)\Gamma(\alpha+\beta+i+1)}$
Parameter restriction :	$\alpha>0,\beta>0$
Domain :	$x=\{0,1,2,...\}$
Mean :	$\frac{\alpha}{\beta-1}$ $\quad$ $for\ \beta>1$
Mode :	0
Variance :	$\frac{\alpha\beta(\alpha+\beta-1)}{(\beta-2)(\beta-1)^2}$ $\quad$ $for\ \beta>2$
Skewness :	$\frac{1}{V^{3/2}}\frac{\alpha\beta(\alpha+\beta-1)(2\alpha+\beta-1)(\beta+1)}{(\beta-3)(\beta-2)(\beta-1)^3}$ $\quad$ $for\ \beta>3$
Kurtosis :	*Complicated*

Beta-Negative Bininomial

VoseBetaNegBin(s,α,β)

Graphs



Uses

The Beta-Negative Binomial(s, α, β) distribution models the number of failures that will occur in a binomial process before s successes are observed and where the binomial probability p is itself a random variable taking a Beta(α, β) distribution.

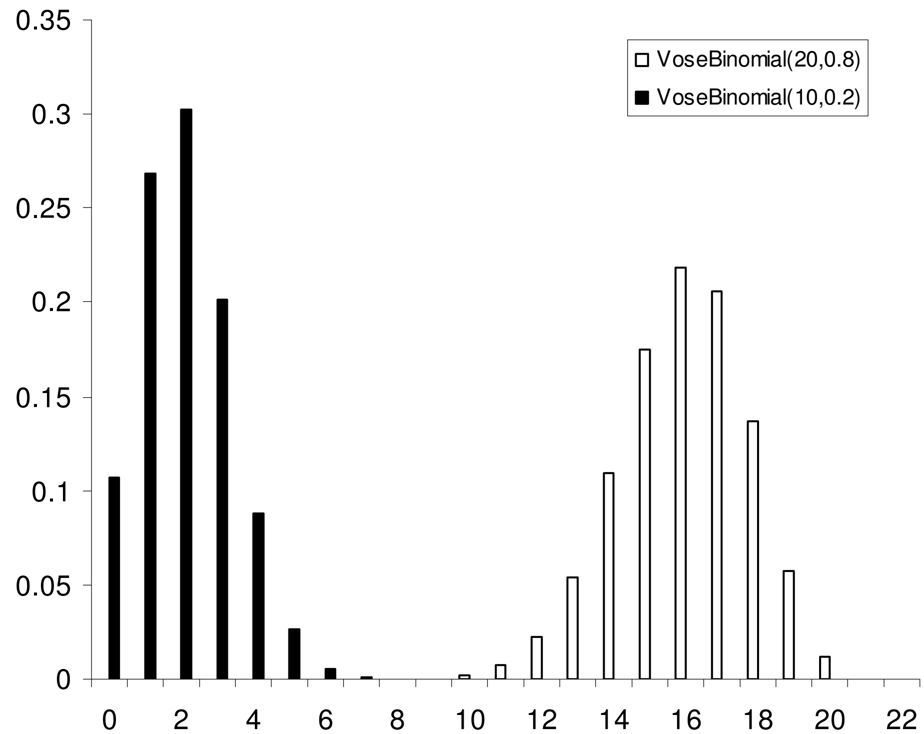
Equations

Probability mass function :	$f(x)=\frac{\Gamma(s+x)\Gamma(\alpha+\beta)\Gamma(\alpha+x)\Gamma(\beta+s)}{\Gamma(s)\Gamma(x+1)\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+s+x)}$
Cumulative distribution function :	$F(x)=\sum_{i=0}^{x}\frac{\Gamma(s+i)\Gamma(\alpha+\beta)\Gamma(\alpha+i)\Gamma(\beta+s)}{\Gamma(s)\Gamma(i+1)\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+s+i)}$
Parameter restriction :	$s>0,\alpha>0,\beta>0$
Domain :	$x=\{0,1,2,\ldots\}$
Mean :	$\frac{s\alpha}{\beta-1}$ $\quad$ $for\ \ \beta>1$
Variance :	$\frac{s\alpha(s\alpha+s\beta-s+\beta^2-2\beta+\alpha\beta-\alpha+1)}{(\beta-2)(\beta-1)^2}\equiv V$ $\quad$ $for\ \ \beta>2$
Skewness :	$\frac{1}{V^{3/2}}\frac{s\alpha(\alpha+\beta-1)(2\alpha+\beta-1)(s+\beta-1)(2s+\beta-1)}{(\beta-3)(\beta-2)(\beta-1)^3}$ $\quad$ $for\ \ \beta>3$
Kurtosis :	*Complicated*

Binomial

VoseBinomial(n,p)

Graphs



Uses

The Binomial distribution models the number of successes from n independent trials where there is a probability p of success in each trial.

The binomial distribution has an enormous number of uses. Beyond simple binomial processes, many other stochastic processes can be usefully reduced to a binomial process to resolve problems. For example:

Binomial process:

- Number of false starts of a car in n attempts;
- Number of faulty items in n from a production line;
- Number of n randomly selected people selected people with some characteristic;

Reduced to binomial:

- Number of machines that last longer than T hours of operation without failure;
- Blood samples that have zero, or >0 antibodies;
- Approximation to a hypergeometric distribution

Comments

The Binomial distribution makes the assumption that the probability p does not change the more trials are performed. That would imply that my aim doesn't get better or worse. It wouldn't be a good estimator, for instance, if the chance of success improved with the number of trials.
Another example: the number of faulty computer chips in a 2000 volume batch where there is a 2% probability that any one chip is faulty = Binomial (2000, 2%).

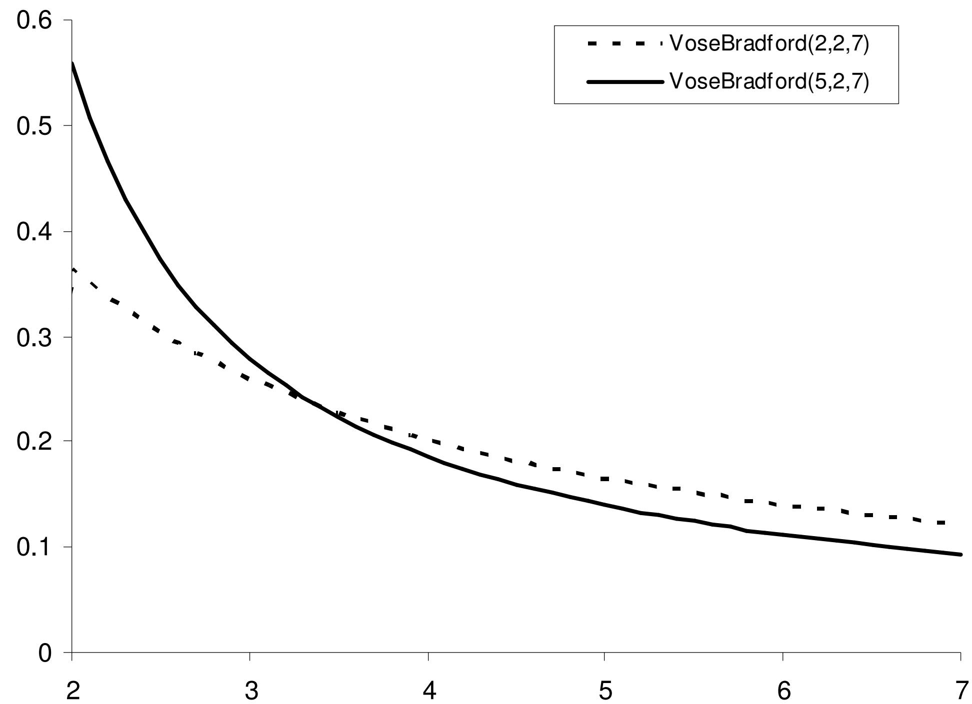
Equations

Probability mass function :	$f(x)=\binom{n}{x}p^{x}(1-p)^{n-x}$
Cumulative distribution function :	$F(x)=\sum_{i=0}^{\lfloor x\rfloor}\binom{n}{i}p^{i}(1-p)^{n-i}$
Parameter restriction :	$0\le p\le 1; n=\{0,1,2,\ldots\}$
Domain :	$x=\{0,1,2,\ldots,n\}$
Mean :	np
Mode :	$p(n+1)-1 \; and \; p(n+1) \quad if \; p(n+1) \; is \; an \; integer$ $p(n+1) \quad otherwise$
Variance :	$np(1-p)$
Skewness :	$\frac{1-2p}{\sqrt{np(1-p)}}$
Kurtosis :	$\frac{1}{np(1-p)}+3\left(1-\frac{2}{n}\right)$

Bradford

VoseBradford(θ,min,max)

Graphs



Comments

The Bradford distribution (also known as the 'Bradford Law of Scattering') is similar to a Pareto distribution that has been truncated on the right. It is right-skewed, peaking at its minimum. The greater the value of theta, the faster its density decreases as one moves away from the minimum. It's genesis is essentially empirical, and very similar to the idea behind the Pareto too. Samuel Clement Bradford originally developed data by studying the distribution of articles in journals in two scientific areas, applied geophysics and lubrication. He studied the rates at which articles relevant to each subject area appeared in journals in those areas. He identified all journals that published more than a certain number of articles in the test areas per year, as well as in other ranges of descending frequency.

He wrote: *If scientific journals are arranged in order of decreasing productivity of articles on a given subject, they may be divided into a nucleus of periodicals more particularly devoted to the subject and several groups or zones containing the same number of articles as the nucleus, when the numbers of periodicals in the nucleus and succeeding zones will be as 1:n:n2... (Bradford, 1948, p. 116)*

Bradford only identified three zones. He found that the value of "n" was roughly 5. So, for example, if a study on a topic finds that six journals contain one-third of the relevant articles found then 6 X 5 = 30 journals will, among them, contain another third of all the relevant articles found, and the last third will be the most scattered of all, being spread out over 6 X 52 = 300 journals.

Bradford's observations are pretty robust. The theory has a lot of implications in researching and investment in periodicals: for example, how many journals an institute should subscribe to, or one should review in a study. It also gives a guide for advertising, by identifying the first third of journals that have the highest impact, helps determine whether journals on a new(ish) topic (or arena like e-journals) have reached a stabilized population, and test the efficiency of Web browsers.

Equations

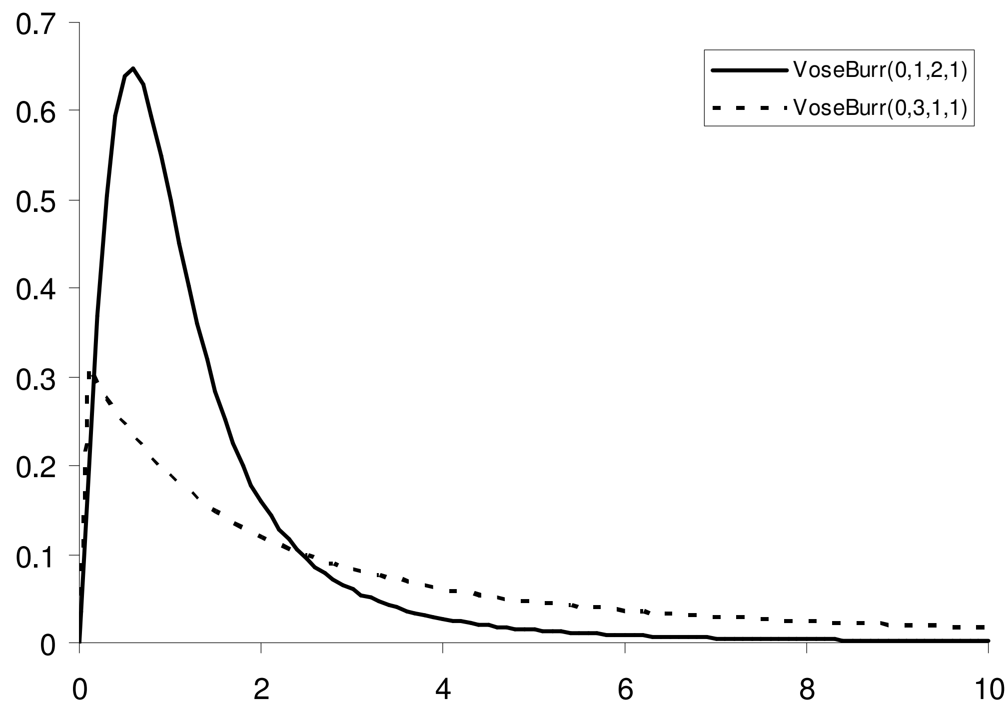
Probability density function :	$f(x)=\dfrac{\theta}{(\theta(x-\min)+\max-\min)\log(\theta+1)}$
Cumulative distribution function :	$F(x)=\dfrac{\log\left(1+\dfrac{\theta(x-\min)}{\max-\min}\right)}{\log(\theta+1)}$
Parameter restriction :	$0<\theta$, $\min<\max$
Domain :	$\min\le x\le\max$
Mean :	$\dfrac{\theta(\max-\min)+k[\min(\theta+1)-\max]}{\theta k}$ *where* $k=\log(\theta+1)$
Mode :	$\min$
Variance :	$\dfrac{(\max-\min)^2[\theta(k-2)+2k]}{2\theta k^2}$
Skewness :	$\dfrac{\sqrt{2}\left(12\theta^2-9k\theta(\theta+2)+2k^2(\theta(\theta+3)+3)\right)}{\sqrt{\theta(\theta(k-2)+2k)}\,(3\theta(k-2)+6k)}$
Kurtosis :	$\dfrac{\theta^3(k-3)(k(3k-16)+24)+12k\theta^2(k-4)(k-3)+6\theta k^2(3k-14)+12k^3}{3\theta(\theta(k-2)+2k)^2}+3$

Burr

VoseBurr(a,b,c,d)

Graphs

The Burr distribution (type III of the list originally presented by Burr) is a right-skewed distribution bounded at a. b is a scale parameter while c and d control its shape. Burr(0,1,c,d) is a unit Burr distribution. Examples of the Burr distribution are given below:



Uses

The Burr distribution has a flexible shape and controllable scale and location which makes it appealing to fit to data. It has, for example, been found to fit tree trunk diameter data for the lumber industry. It is frequently used to model insurance claim sizes, and is sometimes considered as an alternative to a Normal distribution when data show slight positive skewness.

Equations

Probability density function :	$f(x) = \dfrac{cd}{bz^{c+1}\left(1+z^{-c}\right)^{d-1}}$ where $z = \left(\dfrac{x-a}{b}\right)$
Cumulative distribution function :	$F(x) = \dfrac{1}{\left(1+z^{-c}\right)^{d}}$

Parameter restriction :	$b>0, c>0, d>0$
Domain :	$x \geq a$
Mean :	$a+\dfrac{b\Gamma\left(1-\frac{1}{c}\right)\Gamma\left(d+\frac{1}{c}\right)}{\Gamma(d)}$
Mode :	$a+b\left(\dfrac{cd-1}{c+1}\right)^{\frac{1}{c}}$ *if* $c>1$ *and* $d>1$ a *otherwise*
Variance :	$\dfrac{b^2}{\Gamma^2(d)}k$ *where* $k=\Gamma(d)\Gamma\left(1-\frac{2}{c}\right)\Gamma\left(d+\frac{2}{c}\right)-\Gamma^2\left(1-\frac{1}{c}\right)\Gamma^2\left(d+\frac{1}{c}\right)$

Skewness :

$$\frac{\Gamma^2(d)}{k^{3/2}}\left[\frac{2\Gamma^3\left(1-\frac{1}{c}\right)\Gamma^3\left(\frac{1}{c}+d\right)}{\Gamma^2(d)}-\frac{3\Gamma\left(1-\frac{2}{c}\right)\Gamma\left(1-\frac{1}{c}\right)\Gamma\left(\frac{1}{c}+d\right)\Gamma\left(\frac{2}{c}+d\right)}{\Gamma(d)}+\Gamma\left(1-\frac{3}{c}\right)\Gamma\left(\frac{3}{c}+d\right)\right]$$

Kurtosis :

$$\frac{\Gamma^3(d)}{k^2}\left[\frac{6\Gamma\left(1-\frac{2}{c}\right)\Gamma^2\left(1-\frac{1}{c}\right)\Gamma^2\left(\frac{1}{c}+d\right)\Gamma\left(\frac{2}{c}+d\right)}{\Gamma^2(d)}-\frac{3\Gamma^4\left(1-\frac{1}{c}\right)\Gamma^4\left(\frac{1}{c}+d\right)}{\Gamma^3(d)}-\frac{4\Gamma\left(1-\frac{3}{c}\right)\Gamma\left(1-\frac{1}{c}\right)\Gamma\left(\frac{1}{c}+d\right)\Gamma\left(\frac{3}{c}+d\right)}{\Gamma(d)}+\Gamma\left(1-\frac{4}{c}\right)\Gamma\left(\frac{4}{c}+d\right)\right]$$

Cauchy

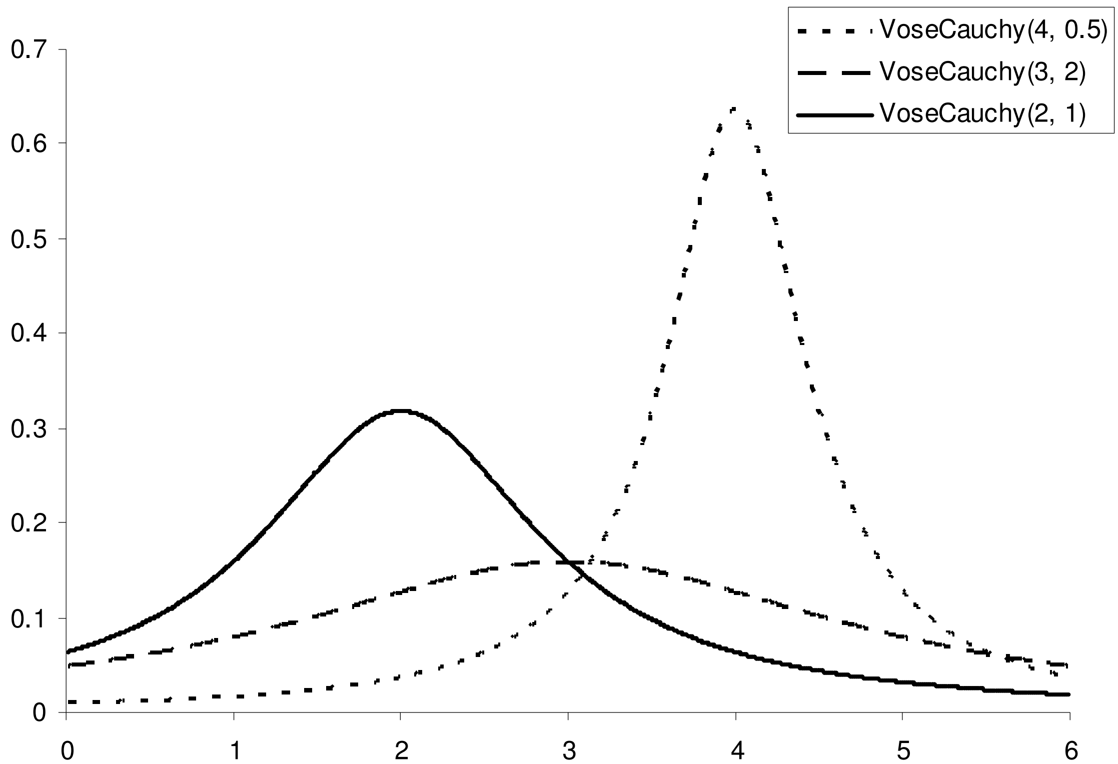
VoseCauchy(a,b)

Graphs

The standard Cauchy distribution is derived from the ratio of two independent Normal distributions, i.e. if X and Y are two independent Normal(0,1) distributions, then

X/Y = Cauchy(0,1)

The Cauchy(a,b) is shifted to have a median at a, and to have b times the spread of a Cauchy(0,1). Examples of the Cauchy distribution are given below:



Uses

The Cauchy distribution is not often used in risk analysis. It is used in mechanical and electrical theory, physical anthropology and measurement and calibration problems. For example, in physics it is usually called a Lorentzian distribution, where it is the distribution of the energy of an unstable state in quantum mechanics. It is also used to model the points of impact of a fixed straight line of particles emitted from a point source.

The most common use of a Cauchy distribution is to show how 'knowledgeable' you are by quoting it whenever someone generalizes about how distributions are used, because it is the exception in many ways: in principle, it has no defined mean (though by symmetry this is usually accepted as being its median = a), and no other defined moments.

Comments

The distribution is symmetric about a and the spread of the distribution increases with increasing b. The Cauchy distribution is peculiar and most noted because none of its moments are well defined (i.e. mean, standard deviation, etc.), their determination being the difference between two integrals that both sum to infinity. Although it looks similar to the Normal distribution, it has much heavier tails. From X/Y = Cauchy(0,1) above you'll appreciate that the reciprocal of a Cauchy distribution is another Cauchy distribution (it is just swapping the two Normal distributions around). The range a - b to a + b contains 50% of the probability area.

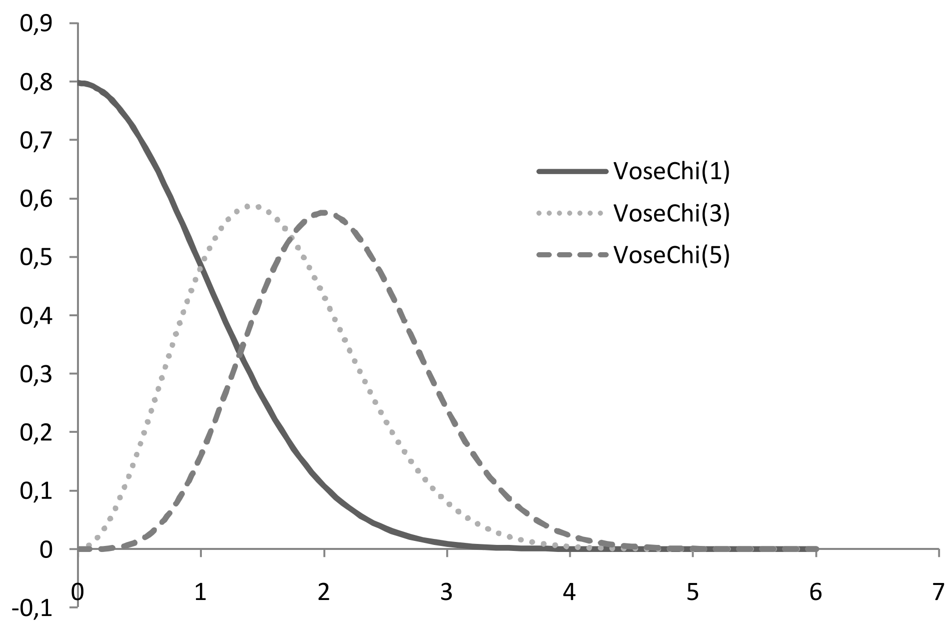
Equations

Probability density function :	$f(x)=\left\{\pi b\left[1+\left(\frac{x-a}{b}\right)^2\right]\right\}^{-1}$
Cumulative distribution function :	$F(x)=\frac{1}{2}+\frac{1}{\pi}\tan^{-1}\left(\frac{x-a}{b}\right)$
Parameter restriction :	$b>0$
Domain :	$-\infty < x < +\infty$
Mean :	*does not exist*
Mode :	a
Variance :	*does not exist*
Skewness :	*does not exist*
Kurtosis :	*does not exist*

Chi

VoseChi(ν)

Graphs



Uses and comments

The standardized chi distribution with v degrees of freedom is the distribution followed by the square root of a Chi-Squared(v) random variable:

$$\mathrm{Chi}(\nu) = \sqrt{\mathrm{ChiSq}(\nu)}$$

Chi(1) is known as a half-normal distribution, i.e.:

$$\mathrm{Chi}(0,1,1) = \sqrt{\mathrm{ChiSq}(1)} = \sqrt{\mathrm{Normal}(0,1)^2} = \left|\mathrm{Normal}(0,1)\right|$$

The Chi distribution usually appears as the length of a k-dimensional vector whose orthogonal components are independent and standard normally distributed. The length of the vector will then have a Chi distribution with k degrees of freedom. For example, the Maxwell distribution of (normalized) molecular speeds is a Chi(3) distribution.

Equations

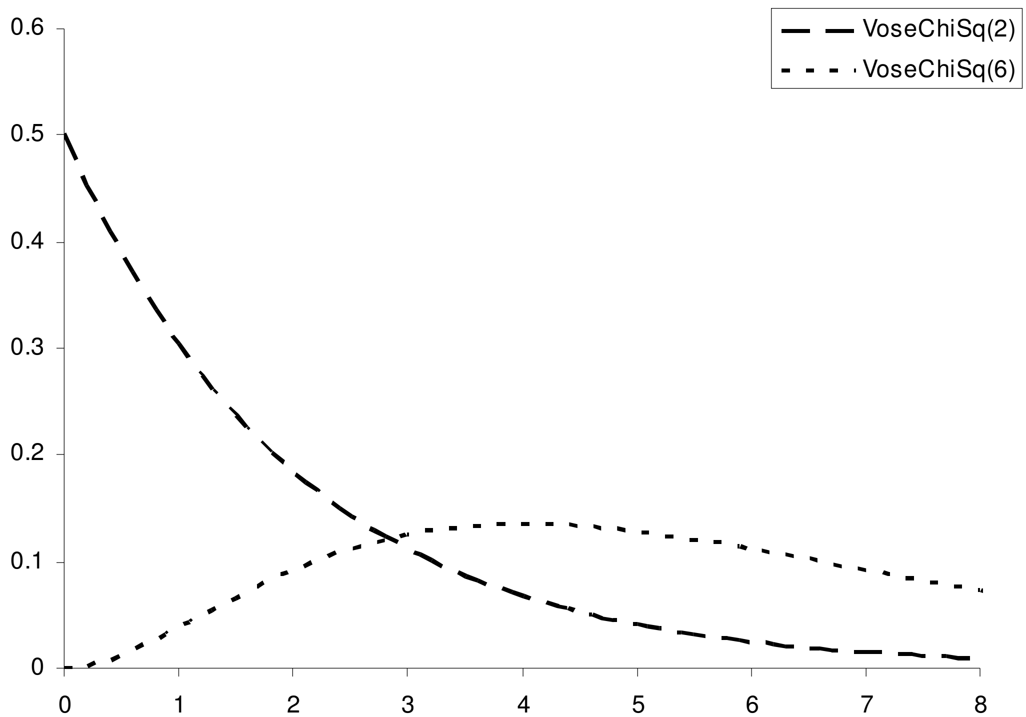
Probability density function :	$f_\nu(x) = \dfrac{x^{\nu-1}\exp\left(-\dfrac{x^2}{2}\right)}{2^{\frac{\nu}{2}-1}\Gamma\left(\dfrac{\nu}{2}\right)}$
Cumulative distribution function :	$F_\nu(x) = \dfrac{\gamma\left(\dfrac{\nu}{2},\dfrac{x^2}{2}\right)}{\Gamma\left(\dfrac{\nu}{2}\right)}$ *where* γ *is the lower incomplete gamma function:* $\gamma(a,x) = \int_0^x t^{a-1}e^{-t}dt$
Parameter restriction :	ν *is an integer*
Domain :	$x \geq 0$
Mean :	$\dfrac{\sqrt{2}\Gamma\left(\dfrac{\nu+1}{2}\right)}{\Gamma\left(\dfrac{\nu}{2}\right)} \equiv \mu$
Mode :	$\sqrt{\nu-1}$ for $\nu \geq 1$
Variance :	$\nu - \mu^2 \equiv V$
Skewness :	*Complicated*
Kurtosis :	*Complicated*

Chi-Square(d)

VoseChiSq(v)

Graphs

The ChiSq distribution is a right-skewed distribution bounded at zero. v is called the 'degrees of freedom' from its use in statistics below.



Uses

The sum of the squares of v unit-Normal distributions (i.e. Normal(0, 1)^2) is a Chisq(v) distribution: so ChiSq(2) = Normal(0,1)^2+Normal(0,1)^2 for example. It is this property that makes it very useful in statistics, particularly classical statistics.

In statistics, we collect a set of observations and from calculating some sample statistics (the mean, variance, etc) attempt to infer something about the stochastic process from which the data came. If the samples are from a Normally distributed population, then the sample variance is a random variable that is a shifted, re-scaled ChiSq distribution.

The Chi Squared distribution is also used to determine the goodness of fit (GOF) of a distribution to a histogram of the available data (a ChiSq test). The method attempts to make a ChiSq distributed statistic by taking the sum of squared errors, normalizing them to be Normal(0,1).

In our view, the ChiSq tests and statistics get over-used (especially the GOF statistic) because the Normality assumption is often tenuous.

Comments

As ν gets large, so it is the sum of a large number of [N(0,1)^2] distributions and, through Central Limit Theorem, approximates a Normal distribution itself.

Sometimes written as $\chi^2(\nu)$. Also related to the Gamma distribution: Chisq(ν) = Gamma(ν/2, 2).

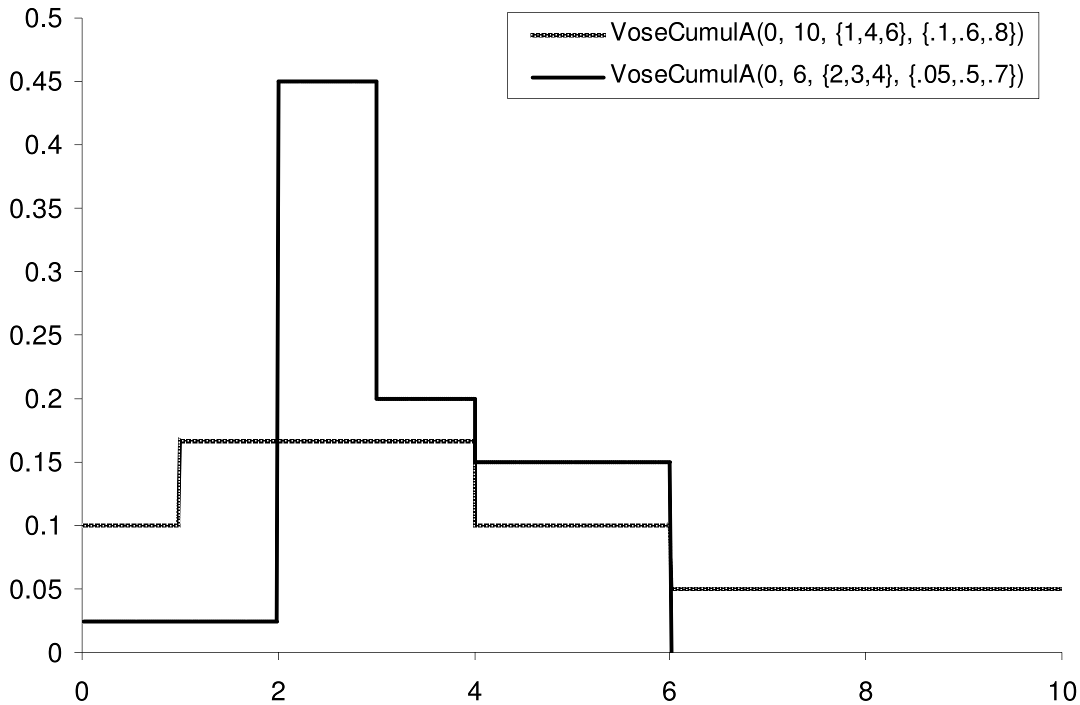
Equations

Probability density function :	$f(x)=\dfrac{x^{\frac{\nu}{2}-1}\exp\left(-\frac{x}{2}\right)}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}$
Cumulative distribution function :	*No closed form*
Parameter restriction :	$\nu>0$, ν *is an integer*
Domain :	$x\geq 0$
Mean :	ν
Mode :	$0 \quad if\ \nu<2$ $\nu\text{-}2 \quad otherwise$
Variance :	2ν
Skewness :	$\sqrt{\dfrac{8}{\nu}}$
Kurtosis :	$3+\dfrac{12}{\nu}$

Cumulative Ascending

VoseCumulA(min,max,{xi},{Pi})

Graphs



Uses

1. Empirical distribution of data

The Cumulative distribution is very useful for converting a set of data values into a first or second order empirical distribution.

2. Modeling expert opinion

The Cumulative distribution can be used to construct uncertainty distributions when using some classical statistical methods. Examples: p in a Binomial process; ? in a Poisson process.

3. Modelling expert opinion

The Cumulative distribution is used in some texts to model expert opinion. The expert is asked for a minimum, maximum and a few percentiles (e.g. 25%, 50%, 75%). However, we have found it largely unsatisfactory because of the insensitivity of its probability scale. A small change in the shape of the Cumulative distribution that would pass unnoticed produces a radical change in the corresponding relative frequency plot that would not be acceptable.

The cumulative distribution is however very useful to model an expert's opinion of a variable whose range covers several orders of magnitude in some sort of exponential way. For example, the number of bacteria in a kg of meat will increase exponentially with time. The meat may contain 100 units of bacteria or 1 million. In such circumstances, it is fruitless to attempt to use a Relative distribution directly.

Equations

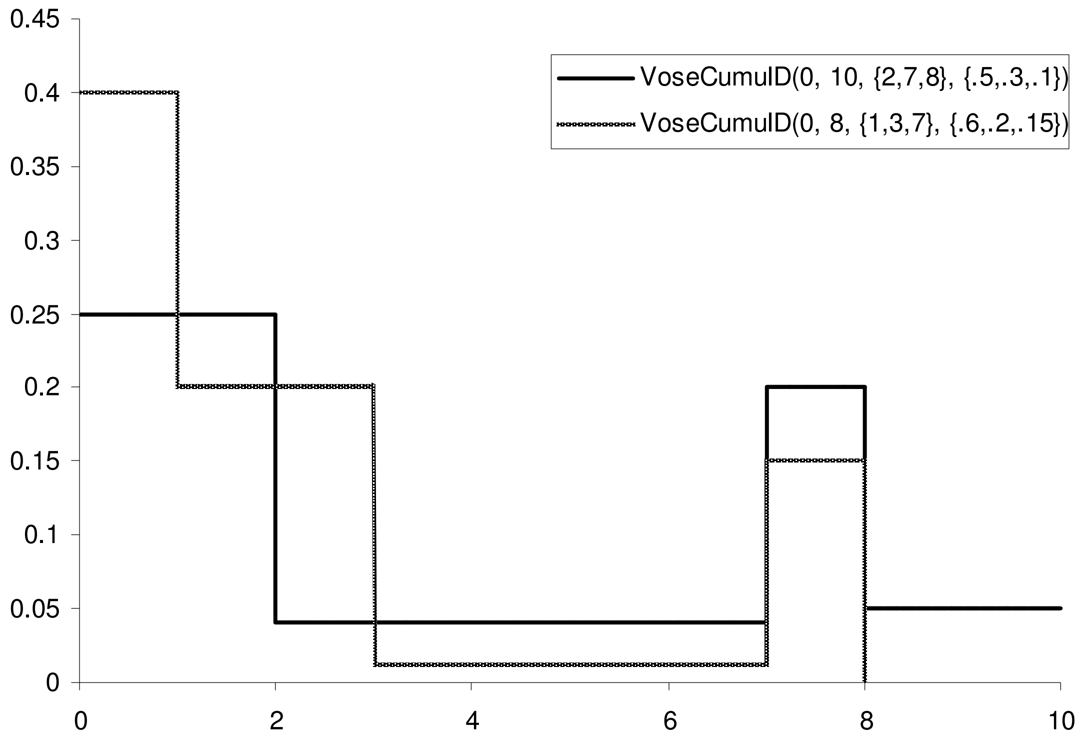
Probability density function :	$f(x)=\dfrac{P_{i+1}-P_i}{x_{i+1}-x_i} \quad for \begin{array}{l} x_i \le x < x_{i+1} \\ i\in\{0,1,\ldots,n\}\end{array}$ $where: \ x_0=\min, x_{n+1}=\max, P_0=0, P_{n+1}=1$
Cumulative distribution function :	$F(x)=\dfrac{x-x_i}{x_{i+1}-x_i}(P_{i+1}-P_i)+P_i \quad for \begin{array}{l} x_i \le x < x_{i+1} \\ i\in\{0,1,\ldots,n\}\end{array}$
Parameter restriction :	$0\le P_i \le 1, P_i \le P_{i+1}, x_i < x_{i+1}, n \ge 0$
Domain :	$\min \le x \le \max$
Mean :	$\sum_{i=0}^{n}\dfrac{f(x_i)}{2}\left(x_{i+1}^2-x_i^2\right)$
Mode :	*No unique mode*
Variance :	*Complicated*
Skewness :	*Complicated*
Kurtosis :	*Complicated*

Cumulative Descending

VoseCumulD(min,max,{xi},{Pi})

Graphs

This is another form of the Cumulative distribution but here the list of cumulative probabilities are the probability of being greater than or equal to their corresponding x-values. Examples of the Cumulative Descending distribution are given below:



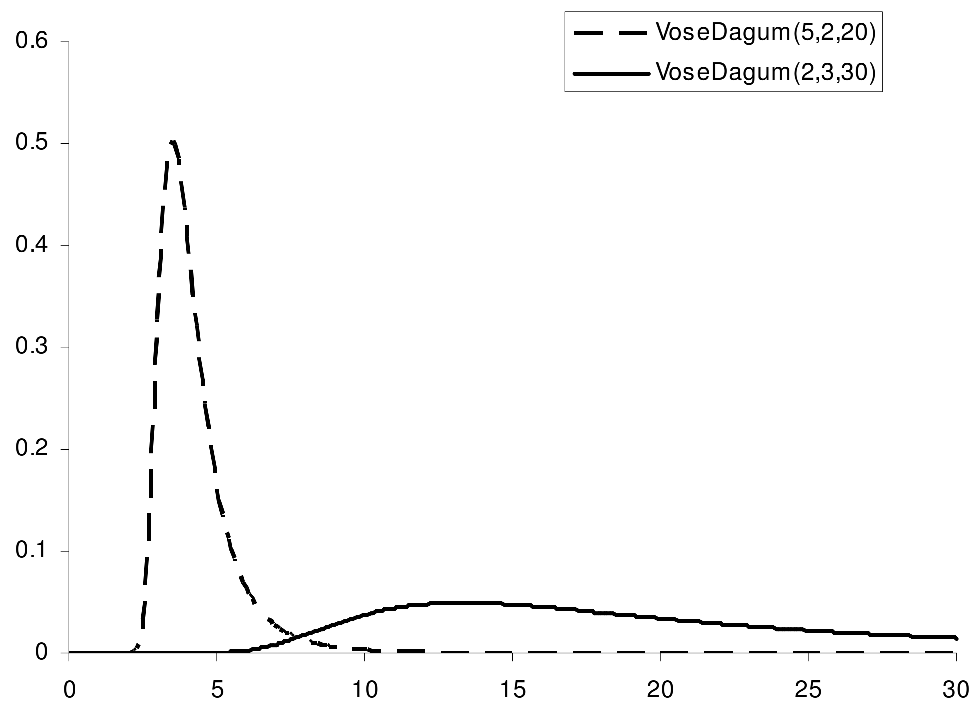
Uses and equations

See Cumulative Ascending distribution (only here $P_{i+1} \leq P_i$, so the Pi values can be converted to those of the CumulA distribution by subtracting them from one: $P_i'=1-P_i$)

Dagum

VoseDagum(a,b,p)

Graphs



The Dagum distribution is often encountered in the actuarial literature or the income distribution literature. The distribution was originally derived by Dagum, when he observed the income elasticity of the c.d.f. of income. a and p are shape parameters and b is a scale parameter.

Uses

The Dagum distribution is sometimes used for fitting to aggregate fire losses.

Equations

Probability density function :	$f(x)=\dfrac{ap\,x^{ap-1}}{b^{ap}\left[1+\left(\dfrac{x}{b}\right)^{a}\right]^{p+1}}$
Cumulative distribution function :	$F(x)=\left[1+\left(\dfrac{x}{b}\right)^{-a}\right]^{-p}$
Parameter restriction :	$a>0, b>0, p>0$
Domain :	$x\geq 0$
Mean :	$\dfrac{b\Gamma\left(p+\frac{1}{a}\right)\Gamma\left(1-\frac{1}{a}\right)}{\Gamma(p)}$
Mode :	$b\left(\dfrac{ap-1}{a+1}\right)^{\frac{1}{a}}$ $\quad if\ ap>1$ 0 $\quad else$
Variance :	$\dfrac{b^2}{\Gamma^2(p)}\left[\Gamma(p)\Gamma\left(p+\frac{2}{a}\right)\Gamma\left(1-\frac{2}{a}\right)-\Gamma^2\left(p+\frac{1}{a}\right)\Gamma^2\left(1-\frac{1}{a}\right)\right]$

Skewness :

$$\frac{b^3}{V^{3/2}\Gamma^3(p)}\left[\Gamma^2(p)\Gamma\left(p+\frac{3}{a}\right)\Gamma\left(1-\frac{3}{a}\right)-3\Gamma(p)\Gamma\left(p+\frac{2}{a}\right)\Gamma\left(1-\frac{2}{a}\right)\Gamma\left(p+\frac{1}{a}\right)\Gamma\left(1-\frac{1}{a}\right)+2\Gamma^3\left(p+\frac{1}{a}\right)\Gamma^3\left(1-\frac{1}{a}\right)\right]$$

Kurtosis :

$$\frac{b^4}{V^2\Gamma^4(p)}\left[\begin{array}{l}\Gamma^3(p)\Gamma\left(p+\frac{4}{a}\right)\Gamma\left(1-\frac{4}{a}\right)-4\Gamma^2(p)\Gamma\left(p+\frac{3}{a}\right)\Gamma\left(1-\frac{3}{a}\right)\Gamma\left(p+\frac{1}{a}\right)\Gamma\left(1-\frac{1}{a}\right)\\ \qquad +6\Gamma(p)\Gamma\left(p+\frac{2}{a}\right)\Gamma\left(1-\frac{2}{a}\right)\Gamma^2\left(p+\frac{1}{a}\right)\Gamma^2\left(1-\frac{1}{a}\right)-3\Gamma^4\left(p+\frac{1}{a}\right)\Gamma^4\left(1-\frac{1}{a}\right)\end{array}\right]$$

Notes

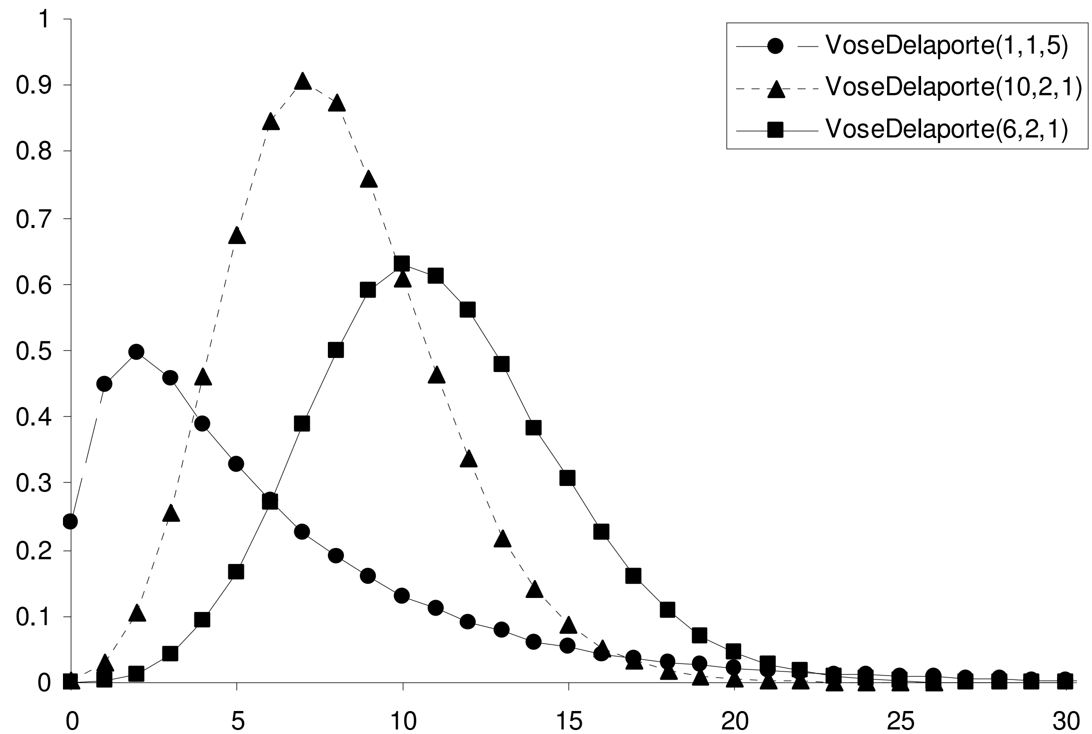
The Dagum distribution is also called the *Inverse Burr distribution* or the *Kappa distribution*.

When a = p, the distribution is also called the *Inverse Paralogistic distribution*.

Delaporte

VoseDelaporte(α,β,λ)

Graphs



Uses

A very common starting point for modeling the numbers of events that occur randomly distributed in time and/or space (e.g. the number of claims that will be received by an insurance company) is the Poisson distribution:

Events = Poisson(λ)

where λ is the expected number of events during the period of interest. The Poisson distribution has a mean and variance equal to λ and one often sees historic data (e.g. frequency of insurance claims) with a variance greater than the mean so that the Poisson model underestimates the level of randomness. A standard method to incorporate greater variance is to assume that λ is itself a random variable (and the resultant frequency distribution is called a mixed Poisson model). A Gamma(α, β) distribution is most commonly used to describe the random variation of λ between periods, so:

Events = Poisson(Gamma(α, β)) (1)

This is the Pólya(α, β) distribution.

Alternatively, one might consider that some part of the Poisson intensity is constant and has an additional component that is random, following a Gamma distribution:

Events = Poisson(λ + Gamma(α, β)) (2)

This is the Delaporte distribution, i.e:

Poisson(λ + Gamma(α, β)) = Delaporte(α, β, γ)

We can split this equation up:

Poisson(λ + Gamma(α, β)) = Poisson(λ) + Poisson(Gamma(α, β))
= Poisson(λ) + Pólya(α, β)

Special cases of the Delaporte distribution:
Delaporte(λ,α,0) = Poisson(λ)
Delaporte(0,α,β) = Pólya(α,β)
Delaporte(0,1,β) = Geometric(1/(1+β))

Equations

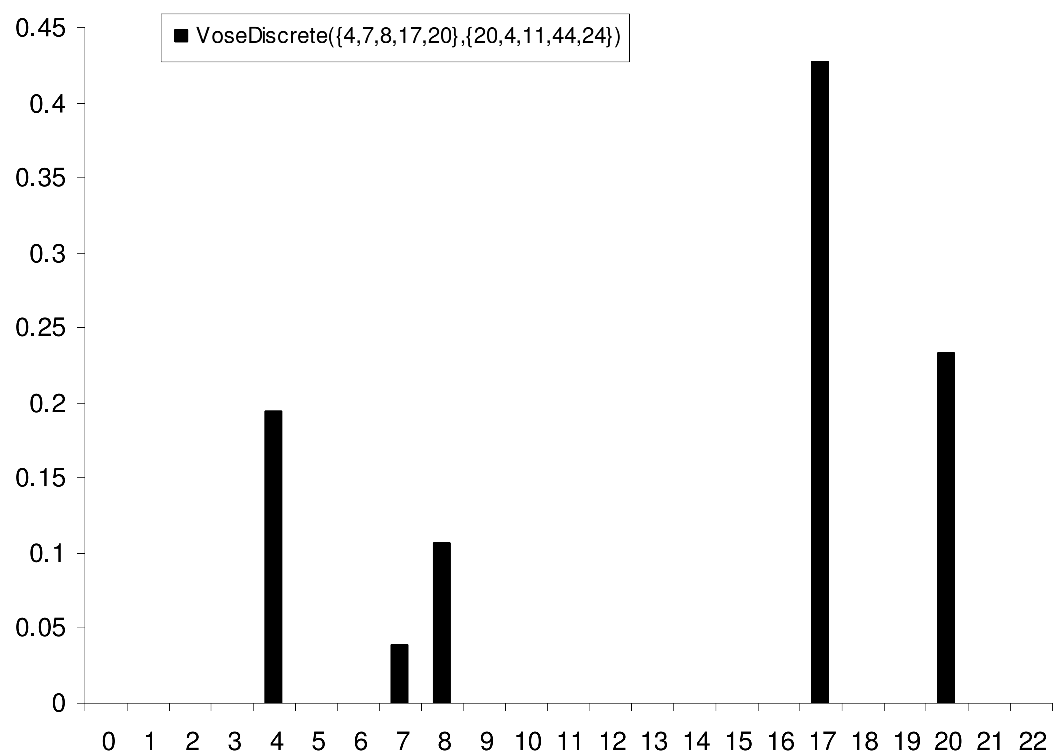
Probability density function :	$f(x)=\sum_{i=0}^{x}\frac{\Gamma(\alpha+i)\beta^{i}\lambda^{x-i}e^{-\lambda}}{\Gamma(\alpha)i!(1+\beta)^{\alpha+i}(x-i)!}$
Cumulative distribution function :	$F(x)=\sum_{j=0}^{x}\sum_{i=0}^{j}\frac{\Gamma(\alpha+i)\beta^{i}\lambda^{j-i}e^{-\lambda}}{\Gamma(\alpha)i!(1+\beta)^{\alpha+i}(j-i)!}$
Parameter restriction :	$\alpha>0,\beta>0,\lambda>0$
Domain :	$x=\{0,1,2,...\}$
Mean :	$\lambda+\alpha\beta$
Mode :	$z,z+1 \quad if\ z\ is\ an\ integer \quad where\ \ z=(\alpha-1)\beta+\lambda$ $\lfloor z\rfloor \quad else$
Variance :	$\lambda+\alpha\beta(\beta+1)$
Skewness :	$\frac{\lambda+\alpha\beta\left(1+3\beta+2\beta^{2}\right)}{\left(\lambda+\alpha\beta(1+\beta)\right)^{3/2}}$
Kurtosis :	$\frac{\lambda+3\lambda^{2}+\alpha\beta\left(1+6\lambda+6\lambda\beta+7\beta+12\beta^{2}+6\beta^{3}+3\alpha\beta+6\alpha\beta^{2}+3\alpha\beta^{3}\right)}{\left(\lambda+\alpha\beta(1+\beta)\right)^{2}}$

Discrete

VoseDiscrete({xi} , {pi})

Graphs

The Discrete distribution is a general type of function used to describe a variable that can take one of several explicit discrete values $\{x_i\}$ and where a probability weight $\{p_i\}$ is assigned to each value. For example, the number of bridges to be built over a motorway extension or the number of times a software module will have to be re-coded after testing. An example of the Discrete distribution is shown below:



Uses

1. Probability branching

A Discrete distribution is also particularly useful to describe probabilistic branching. For example, a firm estimates that it will sell Normal(120,10) tonnes of weed killer next year unless a rival firm comes out with a competing product, in which case it estimates it sales will drop to Normal(85,9) tonnes. It also estimates that there is a 30% chance of the competing product appearing. This could be modeled by:

Sales = VoseDiscrete(A1:A2,B1:B2) where the cells A1:B2 contain the formulae:

A1: =VoseNormal (120, 10)
A2: =VoseNormal (85, 9)
B1: 70%

B2: 30%

2. Combining expert opinion
A Discrete distribution can also be used to combine two or more conflicting expert opinions.

Equations

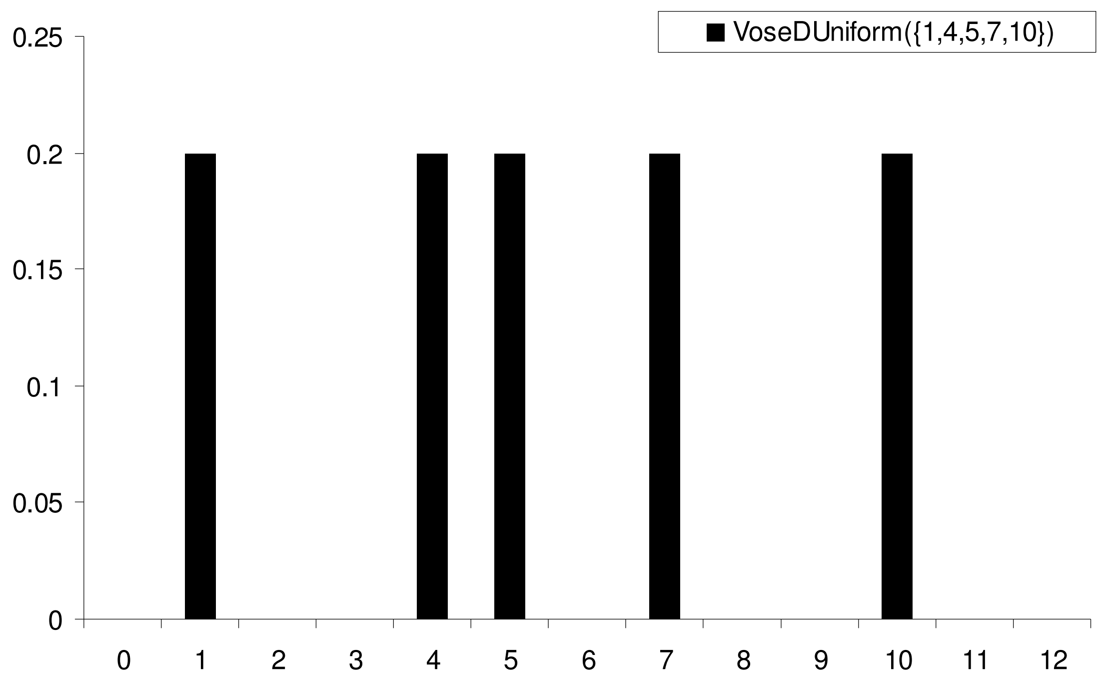
Probability mass function :	$f(x_i) = p_i$
Cumulative distribution function :	$F(x) = \sum_{j=1}^{i} p_j$ if $x_i \le x < x_{i+1}$ assuming the x_i's are in ascending order
Parameter restriction :	$p_i \ge 0, n > 0, \sum_{i=1}^{n} p_i > 0$
Domain :	$x = \{x_1, x_2, \ldots, x_n\}$
Mean :	No unique mean
Mode :	The x_i with the greatest p_i
Variance :	$\sum_{i=1}^{n} (x_i - \bar{x})^2 p_i \equiv V$
Skewness :	$\frac{1}{V^{3/2}} \sum_{i=1}^{n} (x_i - \bar{x})^3 p_i$
Kurtosis :	$\frac{1}{V^2} \sum_{i=1}^{n} (x_i - \bar{x})^4 p_i$

Discrete Uniform

VoseDUniform({xi})

Graph

The Discrete Uniform distribution describes a variable that can take one of several explicit discrete values with equal probabilities of taking any particular value.



Uses

It is not often that we come across a variable that can take one of several values each with equal probability. However, there are a couple of modeling techniques that require that capability:

Bootstrap
Resampling in univariate non-parametric Bootstrap

Fitting empirical distribution to data
Creating an empirical distribution directly from a data set, i.e. where we believe that the list of data values is a good representation of the randomness of the variable.

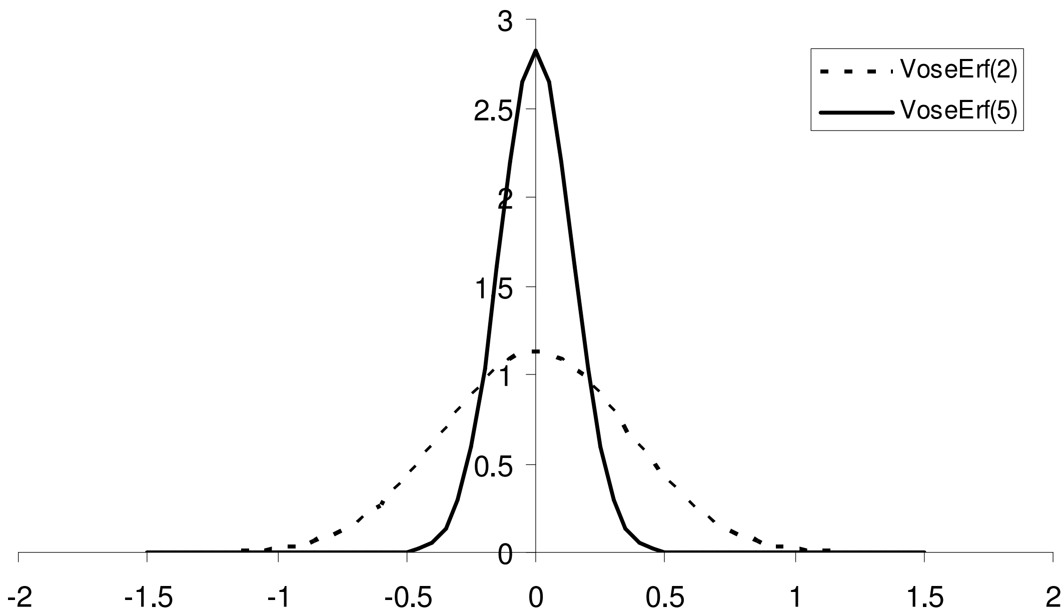
Equations

Probability mass function :	$f(x_i) = \frac{1}{n}$, i = 1 to n
Cumulative distribution function :	$F(x) = \frac{i}{n}$ if $x_i \le x < x_{i+1}$ assuming the x_i's are in ascending order
Parameter restriction :	n > 0
Domain :	$x = \{x_1, x_2, ..., x_n\}$
Mean :	$\frac{1}{n}\sum_{i=1}^{n} x_i$
Mode :	Not uniquely defined
Variance :	$\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 \equiv V$
Skewness :	$\frac{1}{nV^{3/2}}\sum_{i=1}^{n}(x_i - \mu)^3$
Kurtosis :	$\frac{1}{nV^2}\sum_{i=1}^{n}(x_i - \mu)^4$

Error Function

VoseErf(h)

Graphs



Uses

The Error Function distribution is derived from a normal distribution by setting μ = 0 and σ = 1/(h*SQRT(2)). Therefore the uses are the same as those of the Normal distribution.

Equations

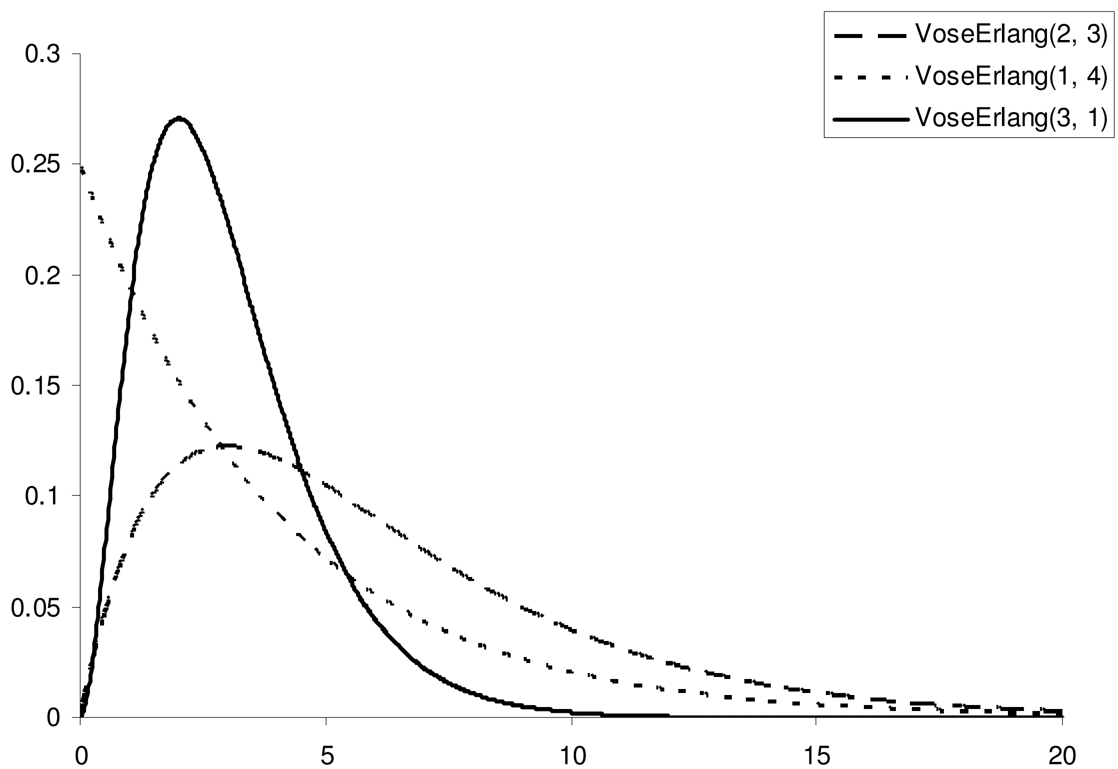
Probability density function :	$f(x)=\frac{h}{\sqrt{\pi}}\exp\left(-(hx)^2\right)$
Cumulative distribution function :	$F(x)=\Phi\left(\sqrt{2}hx\right)$ Φ : Error Function
Parameter restriction :	$h>0$
Domain :	$-\infty<x<+\infty$
Mean :	0
Mode :	0
Variance :	$\frac{1}{2h^2}$
Skewness :	0
Kurtosis :	3

m-Erlang

VoseErlang(m,β)

Graphs

The Erlang distribution (or m-Erlang distribution) is a probability distribution developed by A. K. Erlang. It is a special case of the Gamma distribution. A Gamma(α, β) distribution is equal to an Erlang(m, β) distribution when α is an integer.



Uses

The Erlang distribution is used to predict waiting times in queuing systems, etc. where a Poisson process is in operation, in the same way as a Gamma distribution.

Comments

A.K. Erlang worked a lot in traffic modeling. There are thus two other Erlang distributions, both used in modeling traffic:

Erlang B distribution: this is the easier of the two, and can be used, for example, in a call centre to calculate the number of trunks one need to carry a certain amount of phone traffic with a certain "target service".
Erlang C distribution: this formula is much more difficult and is often used, for example, to calculate how long callers will have to wait before being connected to a human in a call centre or similar situation.

Equations

Probability density function :	$f(x)=\frac{\beta^{-m}x^{m-1}\exp\left(-\frac{x}{\beta}\right)}{(m-1)!}$
Cumulative distribution function :	No closed form
Parameter restriction :	$m>0\,,\beta>0$
Domain :	$x\geq 0$
Mean :	mβ
Mode :	$\beta(m-1)$
Variance :	$m\beta^2$
Skewness :	$\frac{2}{\sqrt{m}}$
Kurtosis :	$3+\frac{6}{m}$

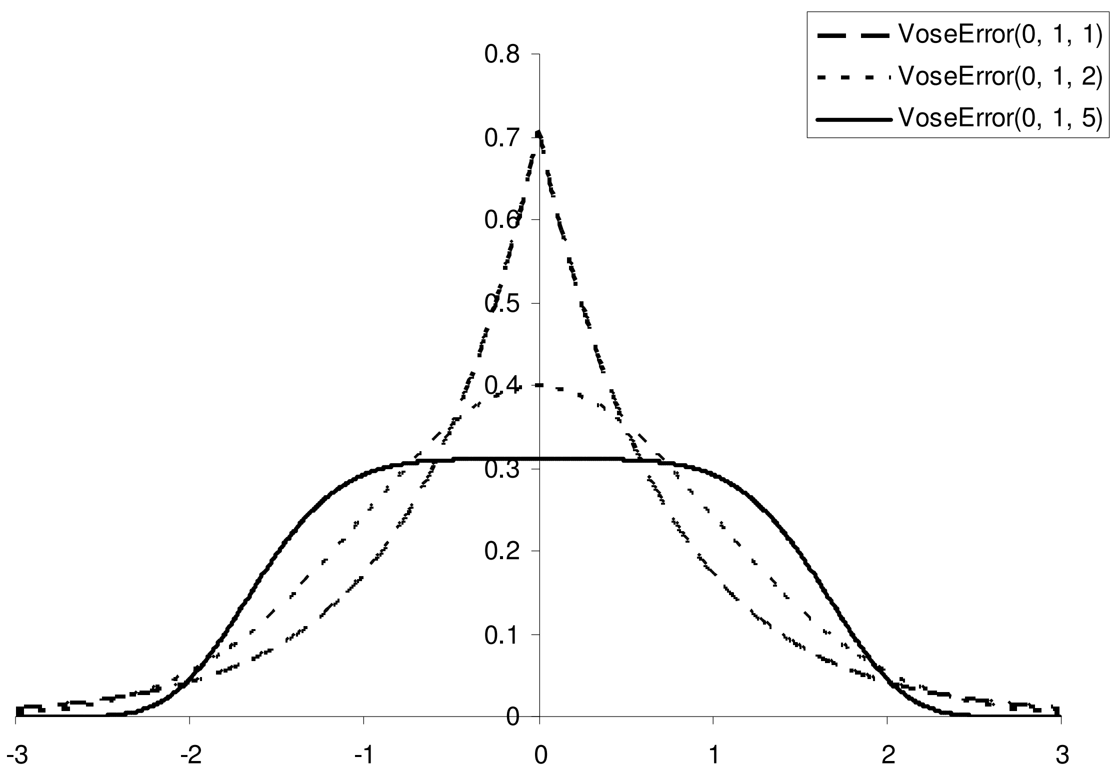
Generalized Error

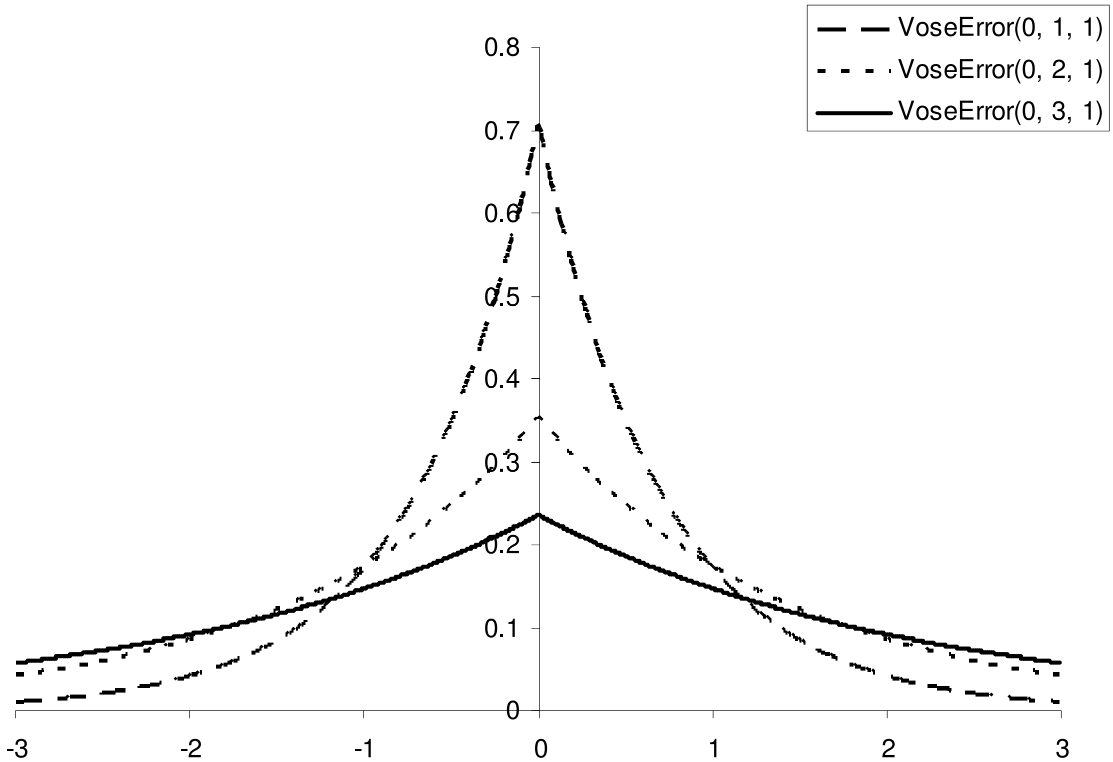
VoseError(μ,σ,ν)

Graphs

The Error distribution goes by the names "*Exponential Power Distribution*" and "*Generalized Error Distribution*".

This three parameter distribution offers a variety of symmetric shapes, as shown in the figures below. The first pane shows the effect on the distribution's shape of varying parameter ν. Note that ν = 2 is a Normal distribution, ν =1 is a Laplace distribution and the distribution approaches a Uniform as ν approaches infinity. The second pane shows the change in the distribution's spread by varying parameter σ, its standard deviation. Parameter μ is simply the location of the distribution's peak, and the distribution's mean.





Uses

The Error distribution finds quite a lot of use as a prior distribution in Bayesian inference because it has greater flexibility than a Normal prior, in that the Error distribution is flatter than a Normal (platykurtic) when $\nu > 2$, and more peaked than a Normal distribution (leptokurtic) when $\nu < 2$. Thus, using the GED allows one to maintain the same mean and variance, but vary the distribution's shape (via the parameter ν) as required.

We have also seen the Error distribution being used to model variations in historic UK property market returns

Equations

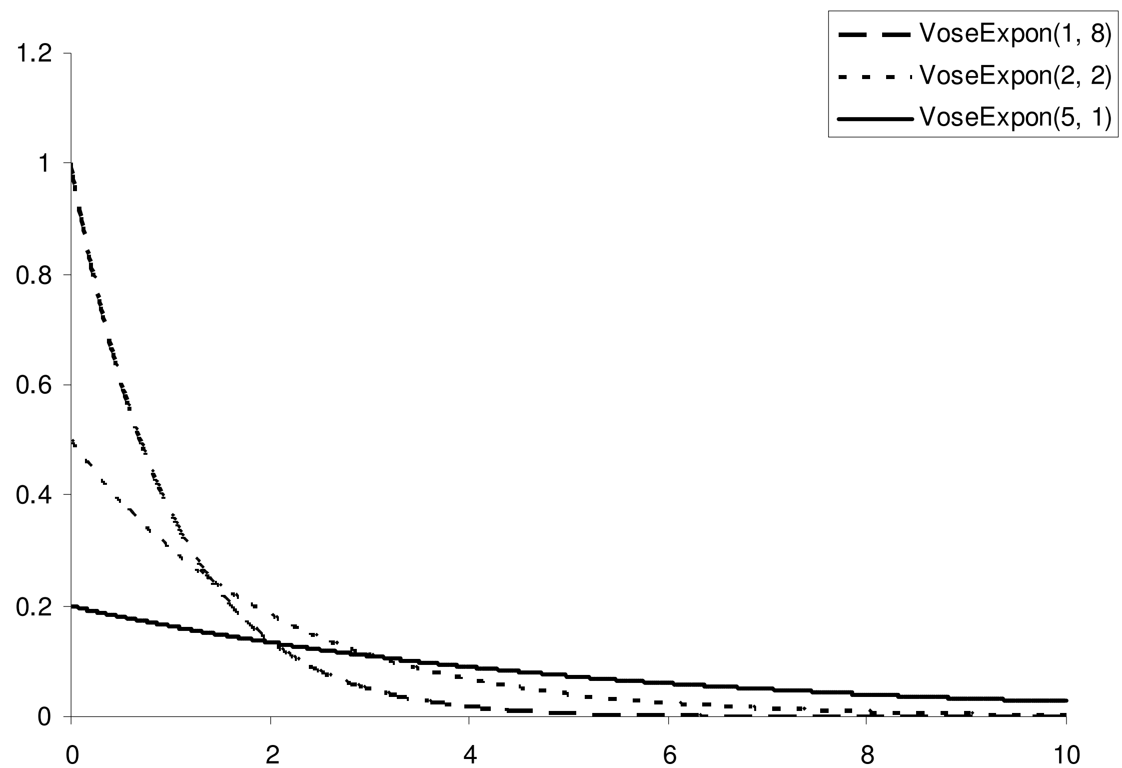
Probability density function :	$f(x) = c_1 \sigma^{-1} \exp\left[-\left\| c_0^{1/2} \sigma^{-1} (x-\mu) \right\|^{\nu} \right]$ where $c_0 = \frac{\Gamma\left(\frac{3}{\nu}\right)}{\Gamma\left(\frac{1}{\nu}\right)}$ and $c_1 = \frac{2c_0^{1/2}}{\nu \Gamma\left(\frac{1}{\nu}\right)}$
Cumulative distribution function :	No closed form
Parameter restriction :	$-\infty < \mu < +\infty, \sigma > 0, \nu > 0$
Domain :	$-\infty < x < +\infty$
Mean :	μ
Mode :	μ
Variance :	σ^2
Skewness :	0
Kurtosis :	$\frac{\Gamma\left(\frac{5}{\nu}\right)\Gamma\left(\frac{1}{\nu}\right)}{\Gamma\left(\frac{3}{\nu}\right)^2}$

Exponential

VoseExpon(β)

Graphs

The Expon(β) is a right-skewed distribution bounded at zero with a mean of β. It only has one shape. Examples of the Exponential distribution are given below:



Uses

The Expon(β) models the time until the occurrence of a first event in a Poisson process. For example:

- The time until the next earthquake;
- The decay of a particle in a mass of radioactive material;
- The length of telephone conversations.

The parameter β is the mean time until the occurrence of the next event.

Example

An electronic circuit could be considered to have a constant instantaneous failure rate, meaning that at small interval of time it has the same probability of failing, given it has survived so far. Destructive tests show that the circuit lasts on average 5 200 hours of operation. The time until failure of any single circuit can be modeled as Expon (5 200) hours. Interestingly, if it conforms to a true Poisson process, this estimate will be independent of how many hours of operation, if any, the circuit has already survived.

Equations

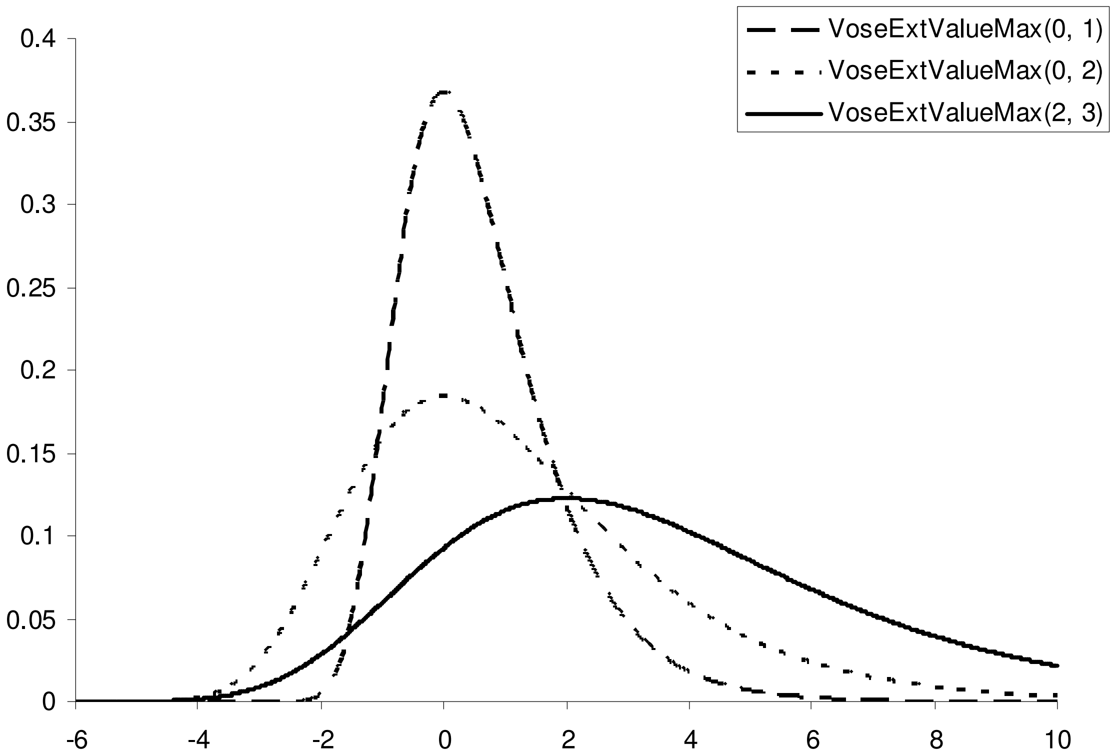
Probability density function :	$f(x)=\dfrac{\exp\left(-\dfrac{x}{\beta}\right)}{\beta}$
Cumulative distribution function :	$F(x)=1-\exp\left(-\dfrac{x}{\beta}\right)$
Parameter restriction :	$\beta>0$
Domain :	$x\geq 0$
Mean :	β
Mode :	0
Variance :	β^2
Skewness :	2
Kurtosis :	9

Extreme Value Max

VoseExtValueMax(a,b)

Graph

The Extreme Value Maximum models the maximum of a set of random variables that have an underlying distribution belonging to the Exponential family, e.g. Exponential, Gamma, Weibull, Normal, Lognormal, Logistic and itself.



Uses

Engineers are often interested in extreme values of a parameter (like minimum strength, maximum impinging force) because they are the values that determine whether a system will potentially fail. For example: wind strengths impinging on a building - it must be designed to sustain the largest wind with minimum damage within the bounds of the finances available to build it; maximum wave height for designing offshore platforms, breakwaters and dikes; pollution emissions for a factory to ensure that, at its maximum, it will fall below the legal limit; determining the strength of a chain, since it is equal to the strength of its weakest link; modeling the extremes of meteorological events since these cause the greatest impact.

Equations

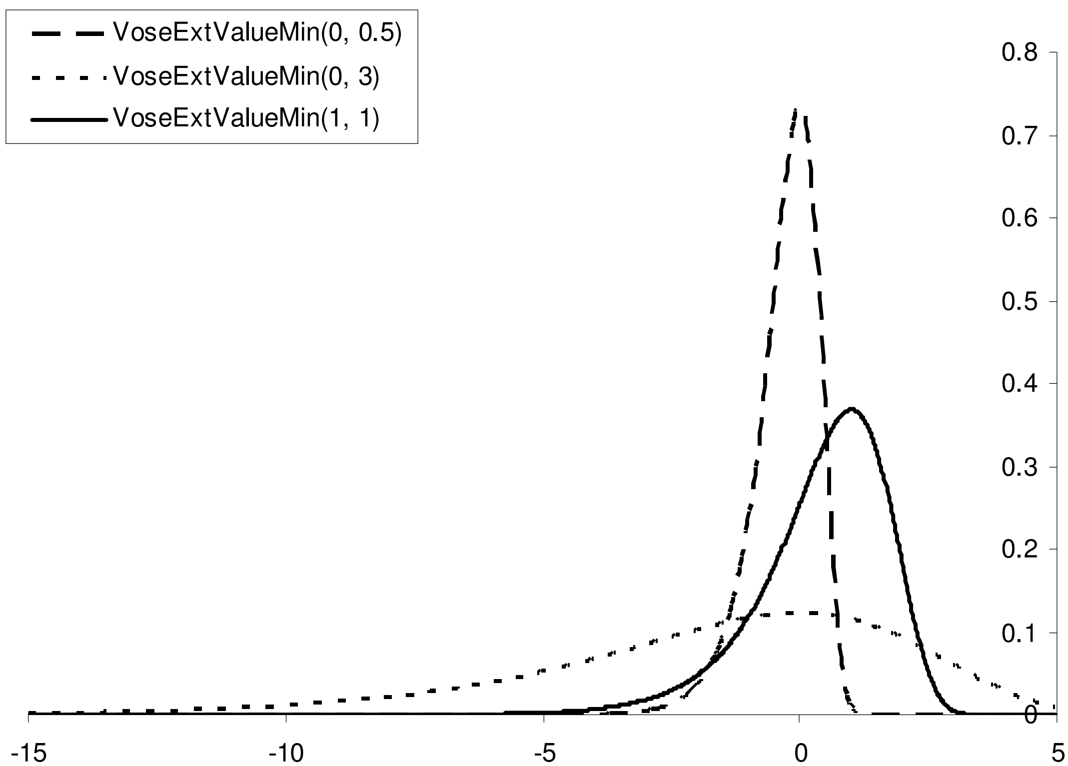
Probability density function :	$f(x)=\left(\frac{1}{b}\right)\exp\left(-\frac{x-a}{b}\right)\exp\left[-\exp\left(-\frac{x-a}{b}\right)\right]$
Cumulative distribution function :	$F(x)=\exp\left[-\exp\left(-\frac{x-a}{b}\right)\right]$
Parameter restriction :	$b>0$
Domain :	$-\infty<x<+\infty$
Mean :	$a-b\Gamma'(1)$ where $\Gamma'(1)\cong -0.577216$
Mode :	a
Variance :	$\frac{b^2\pi^2}{6}$
Skewness :	1.139547
Kurtosis :	5.4

Extreme Value Min

VoseExtValueMin(a,b)

Graph

The Extreme Value Minimum models the minimum of a set of random variables that have an underlying distribution belonging to the Exponential family, e.g. Exponential, Gamma, Weibull, Normal, Lognormal, Logistic and itself.



Uses

Engineers are often interested in extreme values of a parameter (like minimum strength, maximum impinging force) because they are the values that determine whether a system will potentially fail. For example: wind strengths impinging on a building - it must be designed to sustain the largest wind with minimum damage within the bounds of the finances available to build it; maximum wave height for designing offshore platforms, breakwaters and dikes; pollution emissions for a factory to ensure that, at its maximum, it will fall below the legal limit; determining the strength of a chain, since it is equal to the strength of its weakest link; modeling the extremes of meteorological events since these cause the greatest impact.

Equations

Probability density function :	$f(x) = \left(-\frac{1}{b}\right)\exp\left(\frac{x+a}{b}\right)\exp\left[-\exp\left(\frac{x+a}{b}\right)\right]$
Cumulative distribution function :	$F(x) = \exp\left[-\exp\left(\frac{x+a}{b}\right)\right]$
Parameter restriction :	$b > 0$
Domain :	$-\infty < x < +\infty$
Mean :	$a - b\Gamma'(1)$ where $\Gamma'(1) \cong -0.577216$
Mode :	a
Variance :	$\frac{b^2\pi^2}{6}$
Skewness :	-1.139547
Kurtosis :	5.4

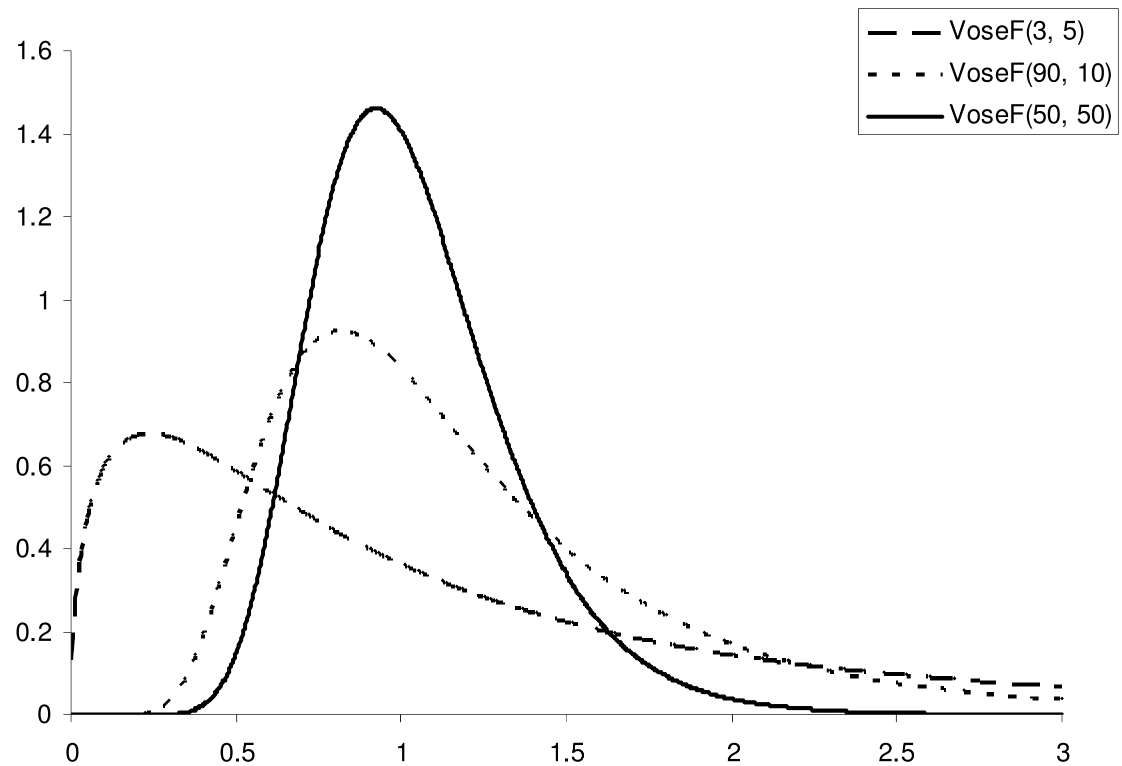
F

VoseF(v1,v2)

Graphs

The F distribution (sometimes known as the Fisher–Snedecor distribution, and taking Fisher's initial) is commonly used in a variety of statistical tests. It is derived from the ratio of two normalized chi-squared distributions with v1 and v2 degrees of freedom as follows:

F(v1, v2)= (ChiSq(v1)/v1)/(ChiSq (v2)/v2)



Uses

The most common use of the F distribution you'll see in statistics text books is to compare the variance between two (assumed Normally distributed) populations. From a risk analysis perspective, it is very infrequent that we would wish to model the ratio of two estimated variances (which is essentially the F-test in this circumstance) so the F distribution is not particularly useful to us.

Equations

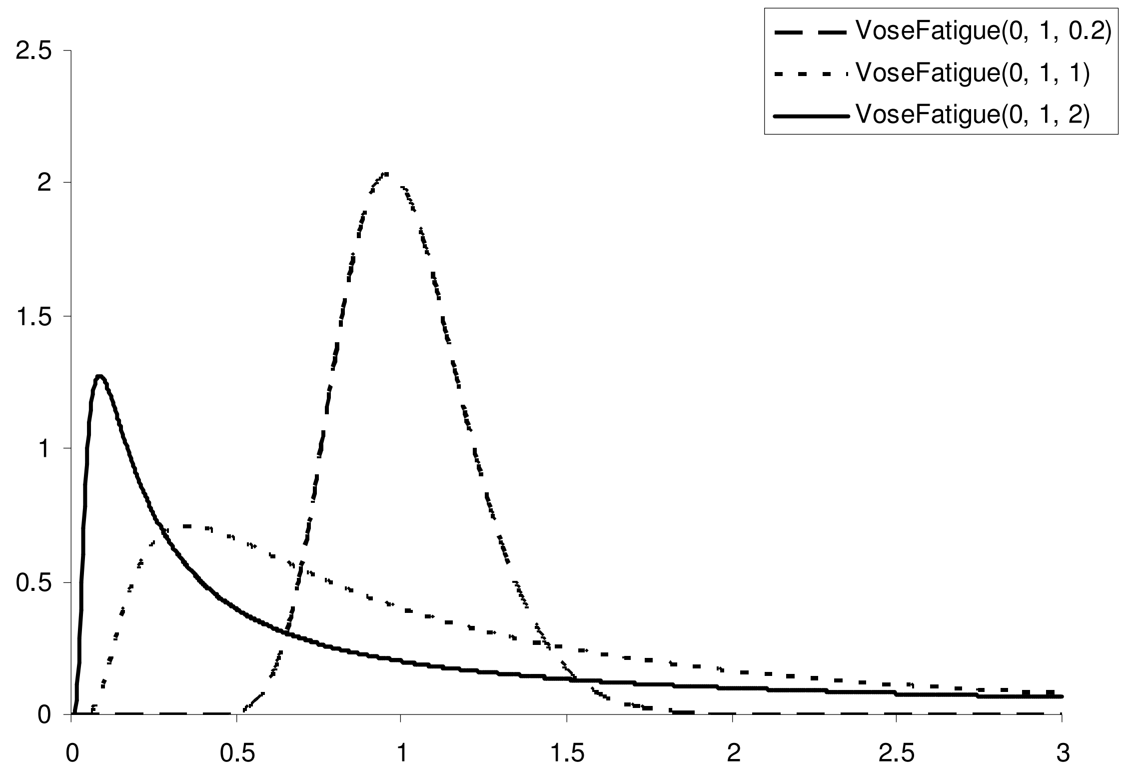
Probability density function :	$f(x)=\dfrac{\Gamma\left(\dfrac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\dfrac{\nu_1}{2}\right)\Gamma\left(\dfrac{\nu_2}{2}\right)}\left(\dfrac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}x^{\frac{\nu_1}{2}-1}\left(1+\left(\dfrac{\nu_1}{\nu_2}\right)x\right)^{-\left(\frac{\nu_1+\nu_2}{2}\right)}$
Cumulative distribution function :	No closed form
Parameter restriction :	$\nu_1>0,\nu_2>0$
Domain :	$x>0$
Mean :	$\dfrac{\nu_2}{\nu_2-1}$ for $\nu_2>2$
Mode :	$\dfrac{\nu_2(\nu_1-2)}{\nu_1(\nu_2+2)}$ if $\nu_1>2$ 0 if $\nu_1=2$
Variance :	$\dfrac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}$ if $\nu_2>4$
Skewness :	$\dfrac{(2\nu_1+\nu_2-2)\sqrt{8(\nu_2-4)}}{(\nu_2-6)\sqrt{\nu_1(\nu_1+\nu_2-2)}}$ if $\nu_2>6$
Kurtosis :	$\dfrac{12\left(20\nu_2-8\nu_2^2+\nu_2^3+44\nu_1-32\nu_1\nu_2+5\nu_2^2\nu_1-22\nu_1^2+5\nu_2\nu_1^2-16\right)}{(\nu_1(\nu_2-6)(\nu_2-8)(\nu_1+\nu_2-2))}+3$ if $\nu_2>8$

Fatigue Life

VoseFatigue(α,β,γ)

Graphs

The Fatigue Lifetime distribution is a right-skewed distribution bounded at a minimum of α. β is a scale parameter while γ controls its shape. Examples of the Fatigue Lifetime distribution are given below:



Uses

The Fatigue Lifetime distribution was originally derived in Birnbaum and Saunders (1969) as the failure of a structure due to the growth of cracks. The conceptual model had a single dominant crack appear and grow as the structure experiences repeated shock patterns up to the point that the crack is sufficiently long to cause failure. Assuming that the incremental growth of a crack with each shock follows the same distribution, that each incremental growth is an independent sample from that distribution, and that there are a large number of these small increases in length before failure, the total crack length will follow a Normal distribution from Central Limit Theorem. Birnbaum and Saunders determined the distribution of the number of these cycles necessary to cause failure. If the shocks occur more or less regularly in time, we can replace the probability that the structure will fail by a certain number of shocks with the probability it fails within a certain amount of time.

Thus, the Fatigue Lifetime distribution is used a great deal to model the lifetime of a device suffering from fatigue. Other distributions in common use to model the lifetime of a devise are the Lognormal, Exponential and Weibull. The Fatigue Life distribution is based on a big assumptions, so be careful in using this distribution despite its popularity. If the growth is likely to be proportional to the crack size, the Lognormal distribution is more appropriate.

Equations

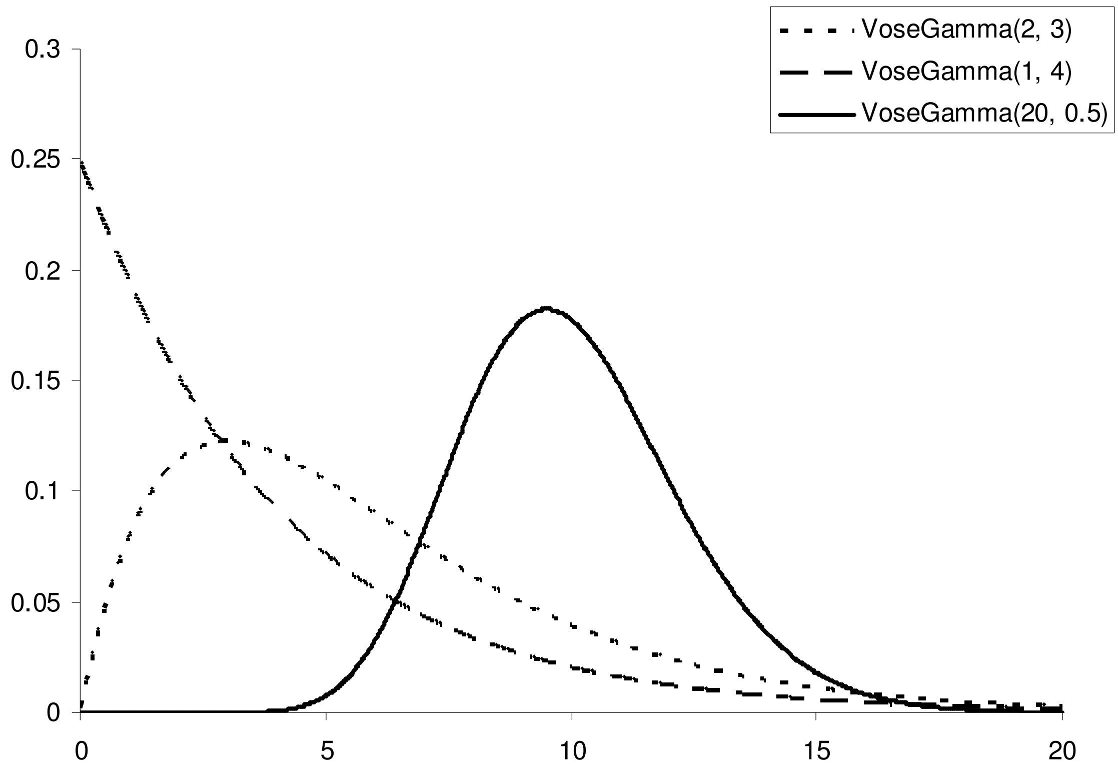
Probability density function :	$f(x)=\frac{z+\frac{1}{z}}{2\gamma z^2}\varphi\left(\frac{z+\frac{1}{z}}{\gamma}\right)$ where $z=\sqrt{\frac{x-\alpha}{\beta}}$ and ϕ = unit normal density
Cumulative distribution function :	$F(x)=\Phi\left(\frac{z-\frac{1}{z}}{\gamma}\right)$ where Φ = unit normal cdf
Parameter restriction :	$\beta>0,\gamma>0$
Domain :	$x>\alpha$
Mean :	$\alpha+\beta\left(1+\frac{\gamma^2}{2}\right)$
Mode :	Complicated
Variance :	$\beta^2\gamma^2(1+\frac{5\gamma^2}{4})$
Skewness :	$\frac{4\gamma^2\left(11\gamma^2+6\right)}{\left(4+5\gamma^2\right)\sqrt{\gamma^2\left(4+5\gamma^2\right)}}$
Kurtosis :	$\frac{3\left(211\gamma^2+120\gamma^2+16\right)}{\left(4+5\gamma^2\right)^2}$

Gamma

VoseGamma(α,β)

Graphs

The Gamma distribution is right-skewed and bounded at zero. It is a parametric distribution based on Poisson mathematics. Examples of the Gamma distribution are given below:



Uses

The Gamma distribution is extremely important in risk analysis modeling, with a number of different uses:

1. Poisson waiting time

The Gamma(α, β) distribution models the time required for α events to occur, given that the events occur randomly in a Poisson process with a mean time between events of β. For example, if we know that major flooding occurs in a town on average every six years, Gamma(4,6) models how many years it will take before the next four floods have occurred.

2. Random variation of a Poisson intensity λ

The Gamma distribution is used for its convenience as a description of random variability of λ in a Poisson process. It is convenient because of the identity:

Poisson(Gamma(α, β)) = NegBin(α, 1/(β+1))

The Gamma distribution can take a variety of shapes, from an Exponential to a Normal, so random variations in ? for a Poisson can often be well approximated by some Gamma, in which case the NegBin distribution becomes a neat combination of the two.

3. Conjugate prior distribution in Bayesian inference

In Bayesian inference, the Gamma distribution is the conjugate to the Poisson likelihood function, which makes it a useful distribution to describe the uncertainty about the Poisson mean λ.

4. Prior distribution for Normal Bayesian inference

If X is Gamma(α, β) distributed, then Y=X^(-1/2) is an Inverted Gamma distribution (InvGamma(α, β)) which is sometimes used as a Bayesian prior for σ for a Normal distribution.

Equations

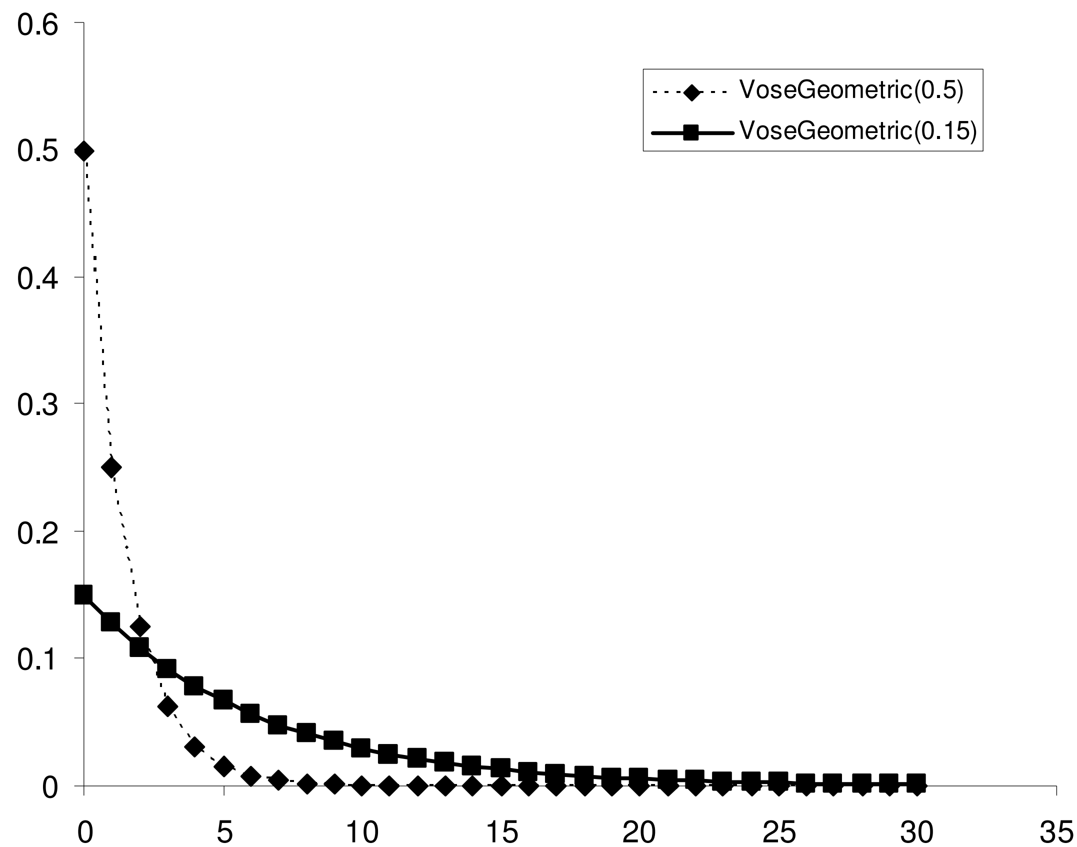
Probability density function :	$f(x)=\dfrac{\beta^{-\alpha}x^{\alpha-1}\exp\left(-\dfrac{x}{\beta}\right)}{\Gamma(\alpha)}$
Cumulative distribution function :	No closed form
Parameter restriction :	$\alpha>0,\beta>0$
Domain :	$x\geq 0$
Mean :	αβ
Mode :	$\beta(\alpha-1)$ if $\alpha\geq 1$ 0 if $\alpha<1$
Variance :	$\alpha\beta^2$
Skewness :	$\dfrac{2}{\sqrt{\alpha}}$
Kurtosis :	$3+\dfrac{6}{\alpha}$

Geometric

VoseGeometric(p)

Graphs

Geometric(p) models the number of failures that will occur before the first success in a set of binomial trials, given that p is the probability of a trial succeeding.



Example uses

Dry oil wells

The Geometric distribution is sometimes quoted as useful to estimate the number of dry wells an oil company will drill in a particular section before getting a producing well. That would, however, be assuming that a) the company doesn't learn from its mistakes; and b) it has the money and obstinacy to keep drilling new wells despite the cost.

More sensible example

You need to purchase some item, conduct a test or operation on that item, and if you fail, go and buy another. For example, you need to find a cow with disease X, but a definitive test involves an expensive procedure, so you randomly select a cow, test it, etc. The number of cows you'll need to buy is 1+Geomet(p) where p is the prevalence of disease X amongst the cows.

Note that this would not work if you bought cows in batches. For example, if you bought batches of 5 cows. Then the number you'd have to buy to get an infected cow is: =(1+Geomet(P))*5, where P = 1-(1-p)5, i.e. the probability that a batch of 5 cows contains at least one infected.

Comments

The Geometric distribution is a special case of the Negative Binomial for s = 1, i.e. Geometric(p) = NegBin(1, p), which means that the sum of s independent Geometric(p) distributions = NegBin(s, p). The Geometric distribution is the discrete analogue of the Exponential distribution, and gets its name because its probability mass function is a geometric progression. The Geometric distribution is occasionally called a *Furry distribution*.

Equations

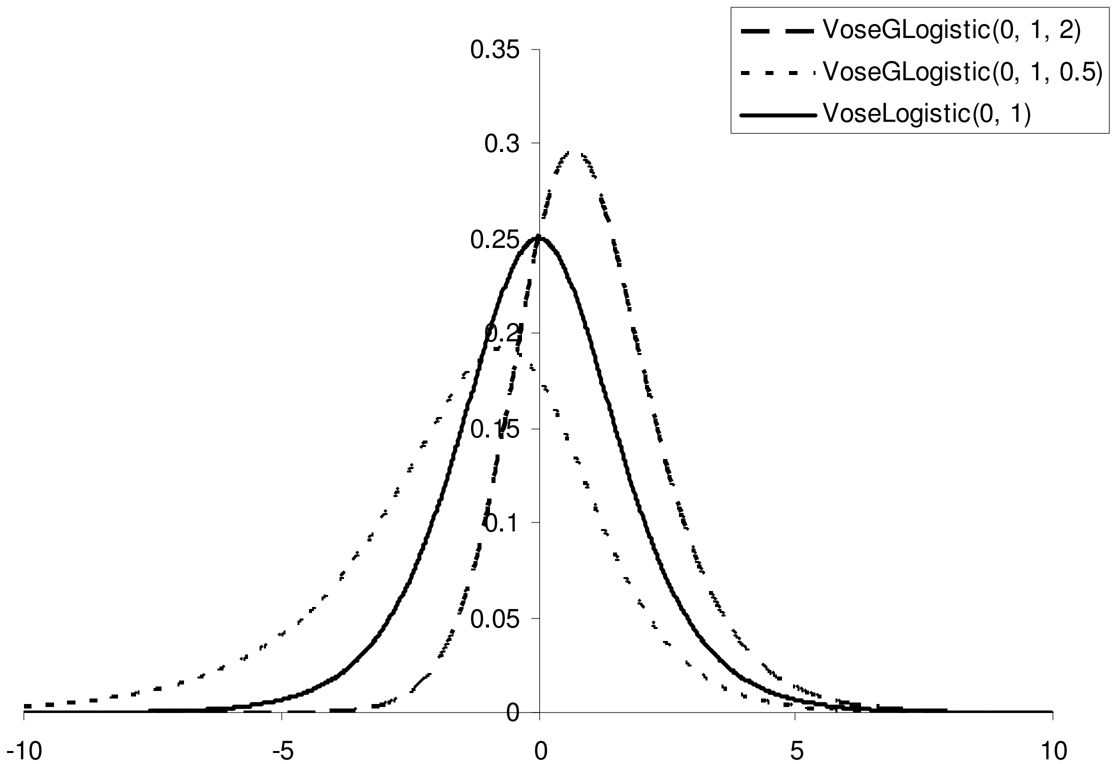
Probability mass function :	$f(x) = p(1-p)^x$
Cumulative distribution function :	$F(x) = 1-(1-p)^{\lfloor x \rfloor + 1}$
Parameter restriction :	$0 < p \leq 1$
Domain :	$x = \{0,1,2,\ldots\}$
Mean :	$\frac{1-p}{p}$
Mode :	0
Variance :	$\frac{1-p}{p^2}$
Skewness :	$\frac{2-p}{\sqrt{1-p}}$
Kurtosis :	$9+\frac{p^2}{1-p}$

Generalized Logistic

VoseGLogistic(α,β,γ)

Graphs

The Generalized logistic distribution can be either left or right skewed (when parameter γ is less than 1 or greater than 1 respectively) or symmetric (γ = 1). When γ = 1 the distribution is Logistic.



Uses

The Generalized logistic distribution has been used to fit values of extremes: for example, extremes of share return fluctuations (Gettinby et al, 2000) and sea levels (van Gelder et al, 2000). It has been used extensively for maximum rainfall modeling and in the UK and elsewhere is used in hydrological risk analysis as the standard model for flood frequency estimation (Institute of Hydrology, 1999).

Equations

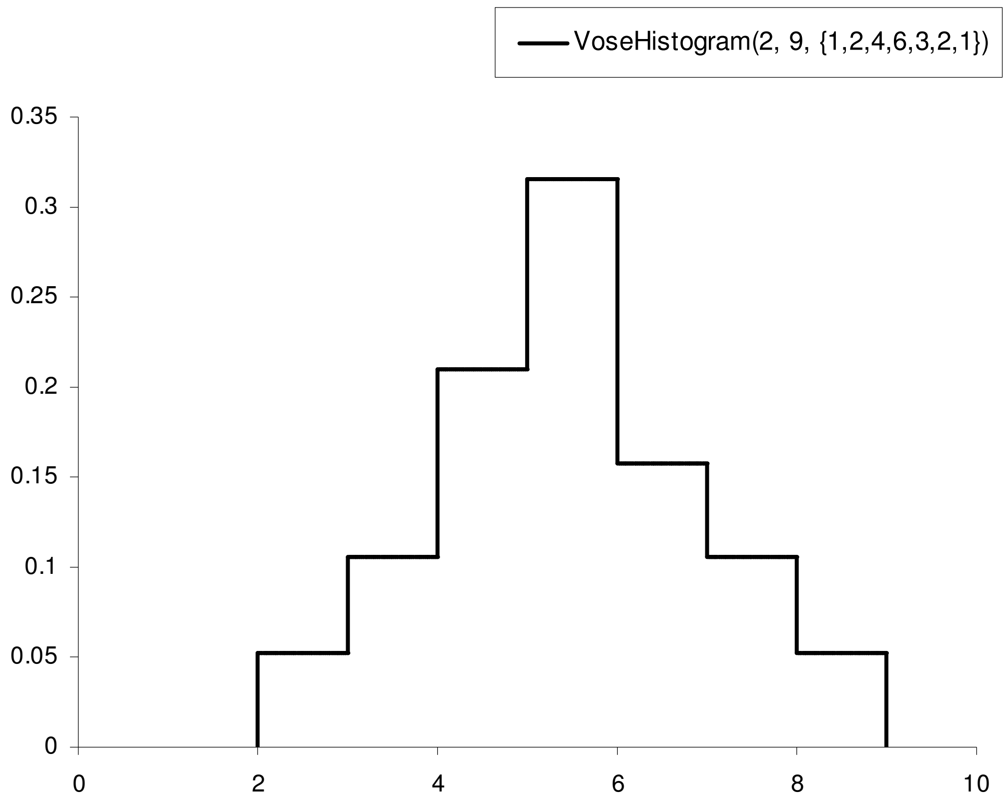
Probability density function :	$f(x)=\dfrac{\gamma}{\beta}\left(\exp\left(\dfrac{x-\alpha}{\beta}\right)\right)\left(1+\exp\left(\dfrac{x-\alpha}{\beta}\right)\right)^{-\gamma-1}$
Cumulative distribution function :	$F(x)=\dfrac{1}{\left(1+\exp\left(\dfrac{x-\alpha}{\beta}\right)\right)^{\gamma}}$
Parameter restriction :	$\beta>0,\gamma>0$
Domain :	$-\infty<x<+\infty$
Mean :	$\alpha+\beta(EM+\Psi(\gamma))$ where $EM\cong 0.57721$ and Ψ the Digamma function
Mode :	$\alpha+\beta\ln[\gamma]$
Variance :	$\beta^2\left(\dfrac{\pi^2}{6}+\Psi'(\gamma)\right)\equiv V$
Skewness :	$\dfrac{\Psi''(1)-\Psi''(\gamma)}{V^{3/2}}$
Kurtosis :	$\dfrac{\Psi'''(1)-\Psi'''(\gamma)}{V^2}$

Histogram

VoseHistogram(min,max,{pi})

Graphs

The Histogram distribution takes three parameters: a minimum; a maximum; and a list of frequencies (or relative frequencies) for a number of equally spaced bands between the minimum and maximum.



Uses

The distribution is useful in a non–parametric technique for replicating the distribution shape of a large set of data. The technique is simply to collate the data into a number of equal bands between a minimum and maximum you determine, calculate the number of data values that fall into each band, and then use this information to define the distribution. It has the disadvantage of 'squaring off' into the histogram shape, but with a lot of data and small bands the technique is a transparent and practical way of fitting a distribution to data.

Equations

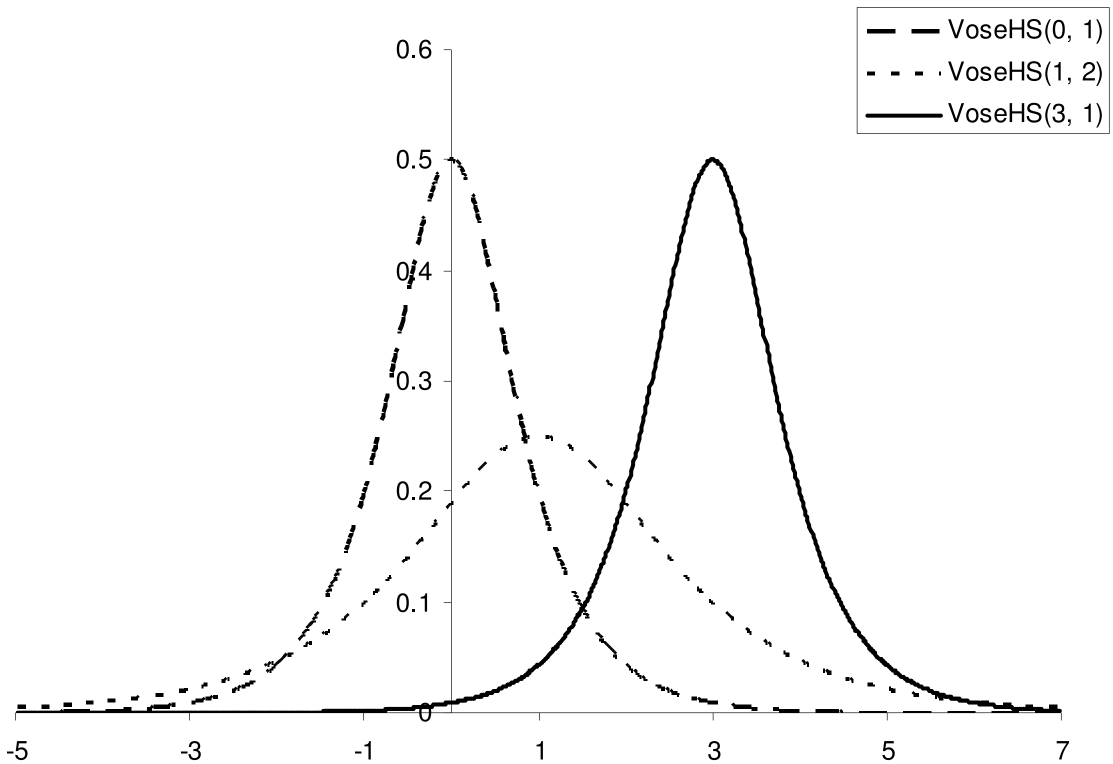
Probability density function :	Assuming the p_i are normalized so that $\frac{n\sum p_i}{\max - \min} = 1$ $f(x) = p_i$ if $x_i \le x < x_{i+1}$ where $x_i = i\frac{\max - \min}{n} + \min$
Cumulative distribution function :	$F(x) = F(x_i) + p_i \frac{x - x_i}{x_{i+1} - x_i}$ if $x_i \le x < x_{i+1}$
Parameter restriction :	$p_i \ge 0, \sum p_i > 0, \min \le \max$
Domain :	$\min \le x \le \max$
Mean :	$\sum_{i=0}^{n} \frac{1}{2}(x_{i+1} + x_1)p_i$
Mode :	No unique mode
Variance :	$\sum_{i=0}^{n-1} \left[\frac{1}{2}(x_{i+1} + x_i) - \mu\right]^2 p_i \equiv V$
Skewness :	$\frac{1}{V^{3/2}} \sum_{i=0}^{n-1} \left[\frac{1}{2}(x_{i+1} + x_i) - \mu\right]^3 p_i$
Kurtosis :	$\frac{1}{V^2} \sum_{i=0}^{n-1} \left[\frac{1}{2}(x_{i+1} + x_i) - \mu\right]^4 p_i$

Hyperbolic Secant

VoseHS(μ,σ)

Graphs

The Hyperbolic-Secant distribution is a symmetric distribution similar to the Normal distribution and defined by its mean and standard deviation, but with a kurtosis of 5, so it is more peaked than the Normal. Examples of the Hyperbolic-Secant distribution are given below:



Uses

The Hyperbolic-Secant distribution can be used to fit data that seem to be approximately Normal in distribution but showing narrower shoulders, just as the Generalized Error and Student distributions are an option for data with wider shoulders than a Normal.

Comments

The Hyperbolic-Secant distribution gets its (rather awful) name from the sech function in its probability density function.

Equations

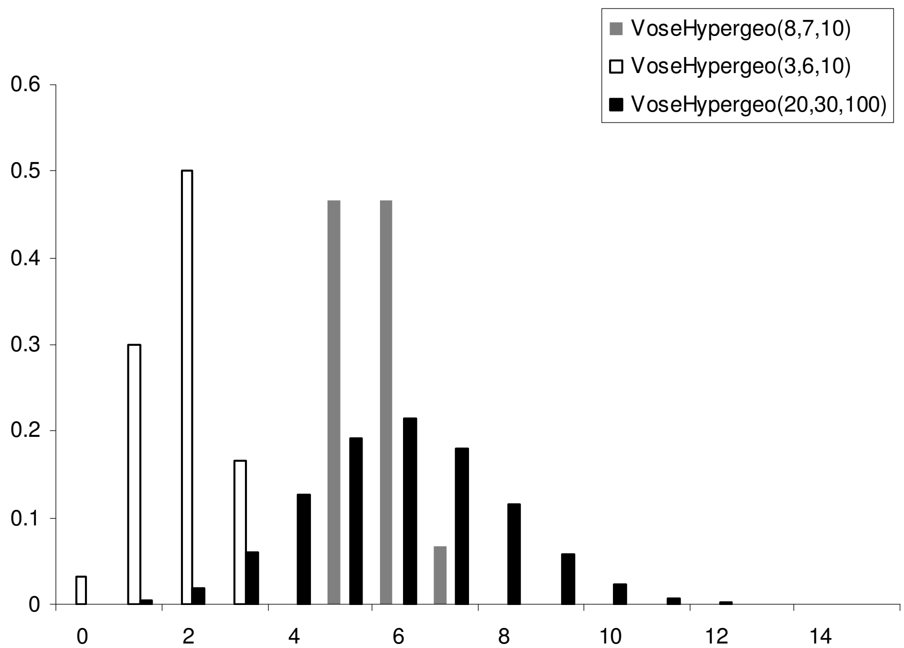
Probability density function :	$f(x) = \frac{\text{sech}(\text{y})}{2\sigma}$ where $y = \frac{\pi}{2\sigma}(x-\mu)$
Cumulative distribution function :	$F(x) = \frac{2}{\pi}\tan^{-1}[\exp(y)]$
Parameter restriction :	$\sigma > 0$
Domain :	$-\infty < x < +\infty$
Mean :	μ
Mode :	μ
Variance :	σ^2
Skewness :	0
Kurtosis :	5

Hypergeometric

VoseHypergeo(n,D,M)

Graphs

The Hypergeo(n, D, M) distribution models the number of items of a particular type that there are in a sample of size n where that sample is drawn from a population of size M of which D are also of that particular type. Examples of the Hypergeometric distribution are shown below:



Examples

Tiles

A company has a stock of 2000 tiles which is known to contain 70 tiles that were not fired properly and will probably crack when exposed to the weather. The tiles are all mixed together and the inferior ones unfortunately cannot be visually identified. A customer orders 800 tiles. The number of faulty tiles he will receive can be estimated by Hypergeo(800, 70, 2000).

Capture-release-recapture experiment to estimate population size

An example of using the Hypergeometric distribution and Bayes' Theorem to estimate the number of tigers on an island is shown in the section uncertainty about a population size. Several animals are captured and tagged, then released back to the wild. Sometime later, another set of animals is captured. The proportion that have tags provide a means, via Bayes' Theorem, to estimate the total population assuming complete diffusion of the tagged sample into the population.

Equations

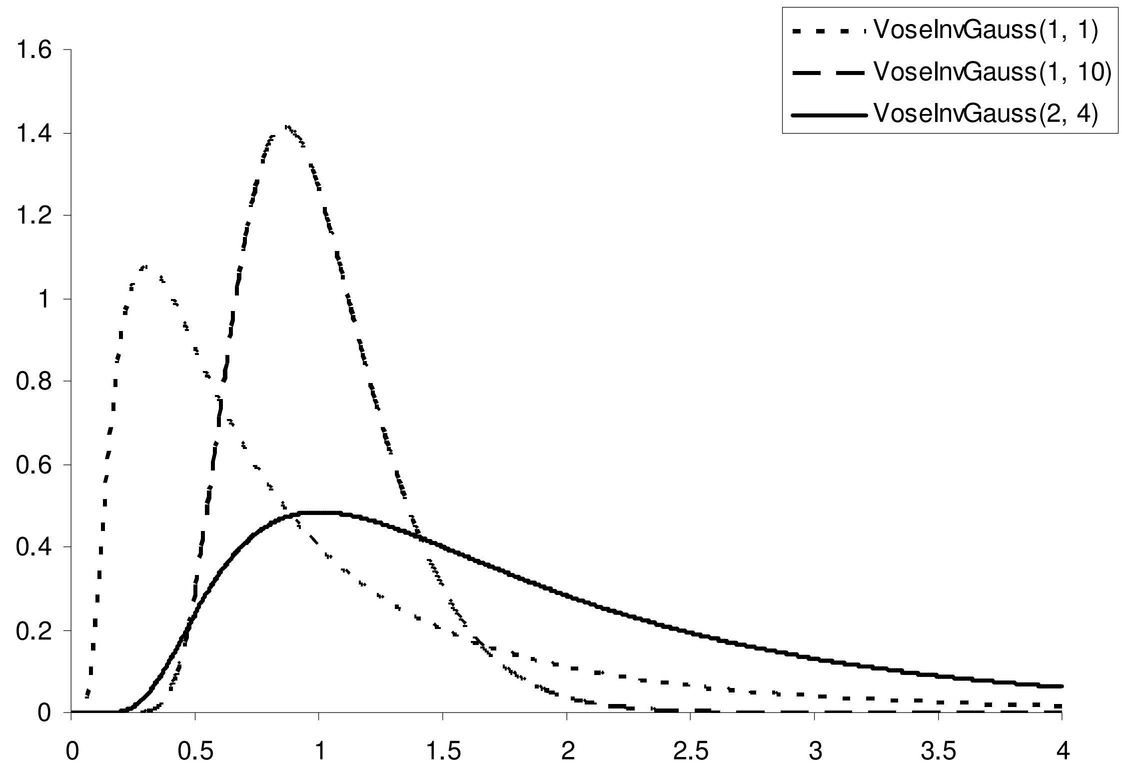
Probability mass function :	$f(x)=\dfrac{\binom{D}{x}\binom{M-D}{n-x}}{\binom{M}{n}}$
Cumulative distribution function :	$F(x)=\sum_{i=0}^{\lfloor x \rfloor}\dfrac{\binom{D}{i}\binom{M-D}{n-i}}{\binom{M}{n}}$
Parameter restriction :	$0<n\le M, 0<D\le M, M>0.$ n, M and D are integers
Domain :	$\max(0,n+D-M)\le x\le \min(n,D)$
Mean :	$\dfrac{nD}{M}$
Mode :	x_m, x_{m-1} if x_m is an integer $\lfloor x_m \rfloor$ otherwise where $x_m=\dfrac{(n+1)(D+1)}{M+2}$
Variance :	$\dfrac{nD}{M^2}\left[\dfrac{(M-D)(M-n)}{(M-1)}\right]$ for $M>1$
Skewness :	$\dfrac{(M-2D)(M-2n)}{M-2}\sqrt{\dfrac{M-1}{nD(M-D)(M-n)}}$ for $M>2$
Kurtosis :	$\left[\dfrac{M^2(M-1)}{n(M-2)(M-3)(M-n)}\right]\left[\dfrac{M(M+1)-6M(M-n)}{D(M-D)}+\dfrac{3n(M-n)(M+6)}{M^2}-6\right]$ for $M>3$

Inverse Gaussian

VoseInvGauss(μ,λ)

Graphs

Right-skewed distribution bounded at zero. Sometimes given the notation IG(μ, λ). Examples of the Inverse Gaussian distribution are given below:



Uses

The Inverse Gaussian is a distribution seldom used in risk analysis. Its primary uses are:

- As a population distribution where a Lognormal distribution has too heavy a right tail
- To model stock returns and interest rate processes (e.g. Madan (1998))

Most uses are rather obscure: it has been used, for example, in physics to model the time until a particle, moving with Brownian motion with a drift, will exceed a certain distance from its original position for the first time.

Equations

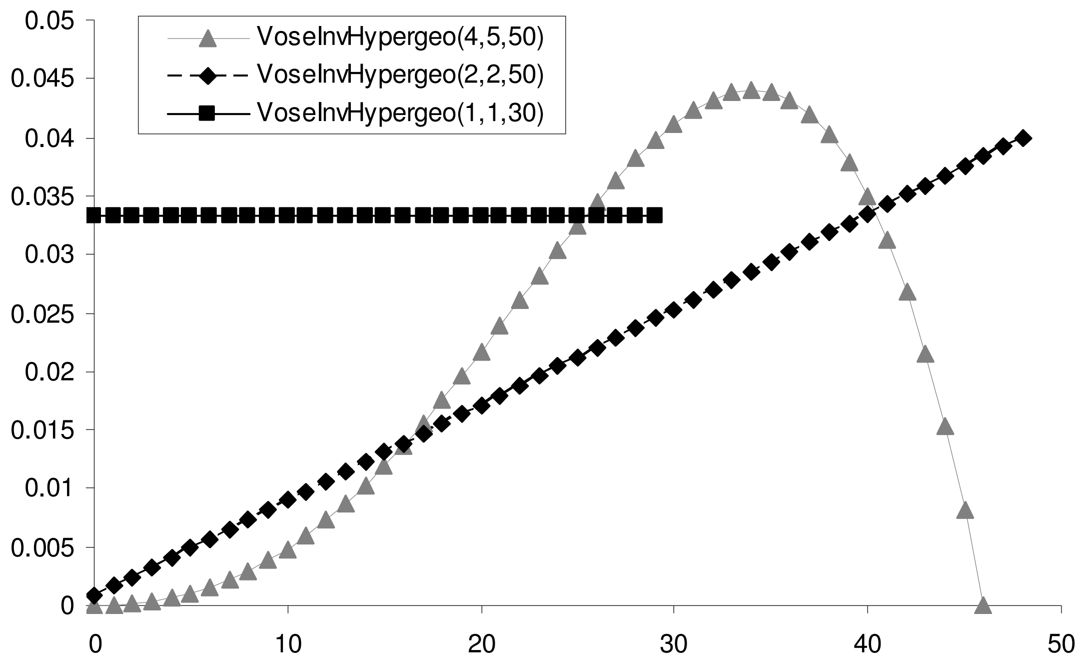
Probability density function :	$f(x)=\sqrt{\frac{\lambda}{2\pi x^3}}\exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right)$
Cumulative distribution function :	No closed form
Parameter restriction :	$\mu>0,\lambda>0$
Domain :	$x>0$
Mean :	μ
Mode :	$\mu\left(\sqrt{1+\frac{9\mu^2}{4\lambda^2}}-\frac{3\mu}{2\lambda}\right)$
Variance :	$\frac{\mu^3}{\lambda}$
Skewness :	$3\sqrt{\frac{\mu}{\lambda}}$
Kurtosis :	$3+15\frac{\mu}{\lambda}$

Inverse Hypergeometric

VoseInvHypergeo(s,D,M)

Graphs

The Inverse Hypergeometric distribution InvHypergeo(s, D, M) models the number of failures one would have before achieving the sth success in a hypergeometric sampling where there are D individuals of interest (their selection is a 'success') in a population of size M. Four examples of the Inverse Hypergeometric distribution are shown below:



Uses

It should be used in any situation where there is hypergeometric sampling and one is asking the question: "How many failures will I observe before I get s successes?", or alternatively "How many samples do I need to s successes?".

Example

The number of cards one needs to turn over to see three hearts will be Inverse Hypergeometric distributed. If the total number of cards is 54, of which 13 are hearts, the number of cards one needs to turn over is 3+InvHypergeo(3,13,54)

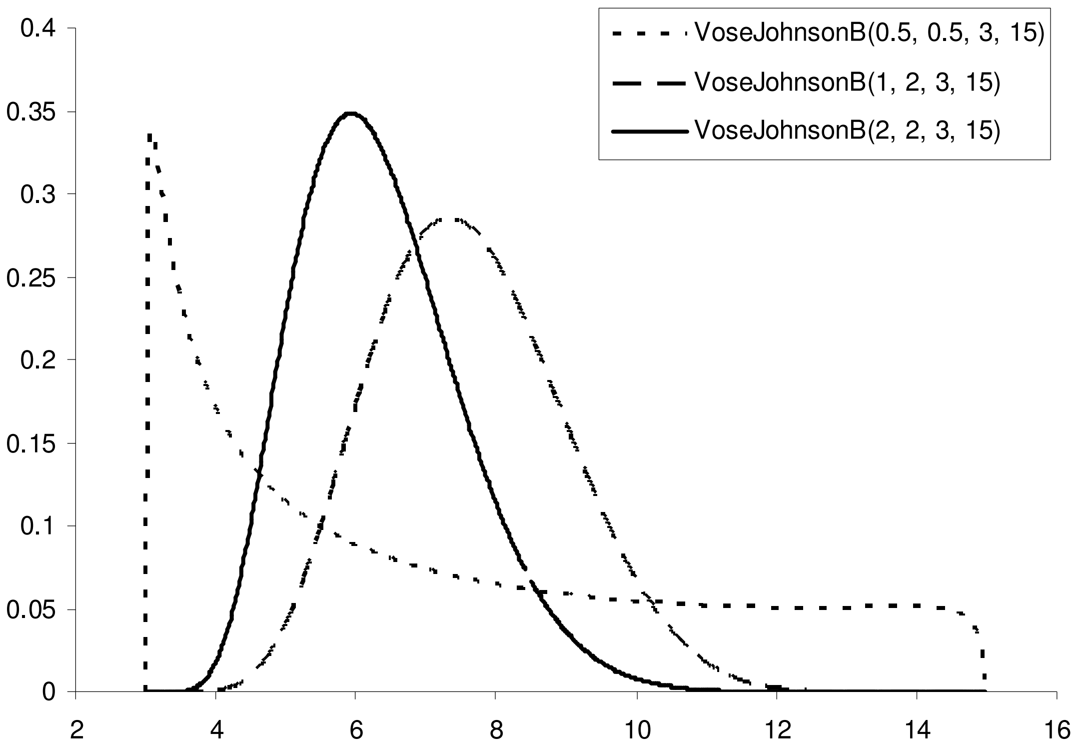
Equations

Probability mass function :	$f(x)=\dfrac{\binom{D}{s-1}\binom{M-D}{x-s}(D-s+1)}{\binom{M}{x-1}(M-x+1)}$
Cumulative distribution function :	$F(x)=\sum_{i=s}^{x}\dfrac{\binom{D}{s-1}\binom{M-D}{i-s}(D-s+1)}{\binom{M}{i-1}(M-i+1)}$
Parameter restriction :	$s \le D,\ D \le M$
Domain :	$s \le x \le M-D+s$
Mean :	$\dfrac{s(M-D)}{D+1}$
Mode :	x_m, x_{m-1} if x_m is an integer $\lfloor x_m \rfloor$ otherwise where $x_m=\dfrac{(s-1)(M-D-1)}{(D-1)}$
Variance :	$\dfrac{s(M-D)(M+1)(D-s+1)}{(D+1)^2(D+2)}\equiv V$
Skewness :	$\dfrac{V(D-2M-1)(2s-D-1)}{(D+1)(D+3)V^{3/2}}$
Kurtosis :	$\dfrac{(D+1)(D-6s)+3(M-D)(M+1)(s+2)+6s^2-3(M-D)(M+1)s(6+s)/(D+1)+18(M-D)(M+1)s^2/(D+1)^2}{(D+3)(D+4)V}$

Johnson Bounded

VoseJohnsonB(α1,α2,min,max)

Graphs



The Johnson bounded distribution has a range defined by the min and max parameters. Combined with its flexibility in shape, this makes it a viable alternative to the PERT, Triangular and Uniform distributions for modeling expert opinion. A public domain software product called VISIFIT allows the user to define the bounds and pretty much any two statistics for the distribution (mode, mean, standard deviation) and will return the corresponding distribution parameters.

Setting min to 0 and max to 1 gives a random variable that is sometimes used to model ratios, probabilities, etc instead of a Beta distribution.

The distribution name comes from Johnson (1949) who proposed a system for categorizing distributions, in much the same spirit that Pearson did. Johnson's idea was to translate distributions to be a function of a unit Normal distribution, one of the few distributions for which there were good tools available at the time to handle.

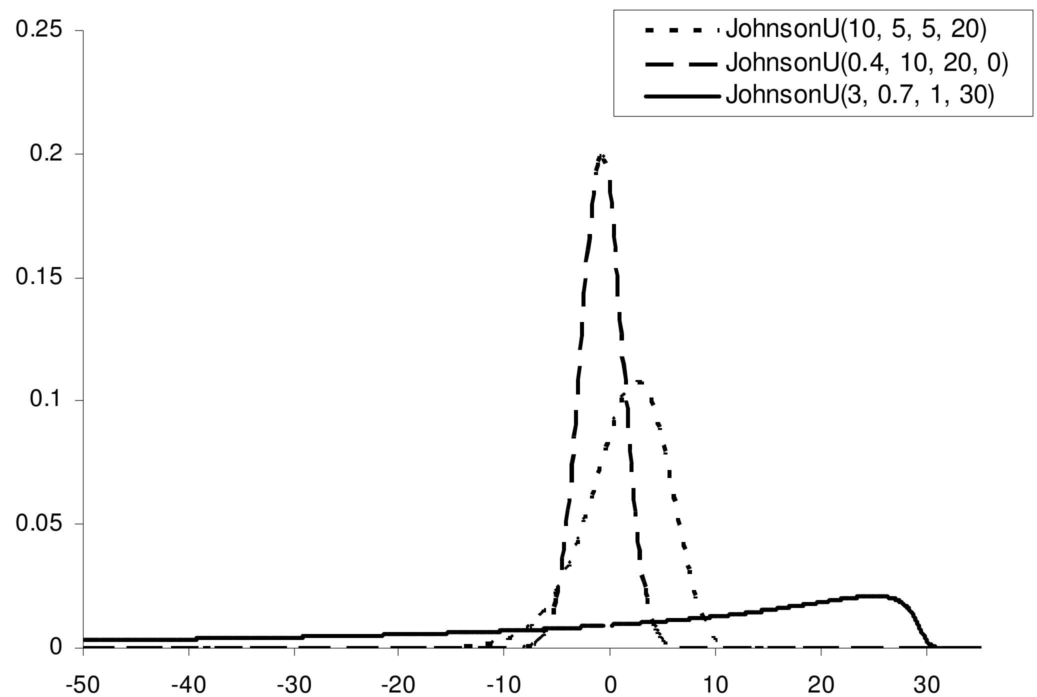
Equations

Probability density function :	$f(x)=\frac{\alpha_2(\max-\min)}{(x-\min)(\max-x)\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\alpha_1+\alpha_2\ln\left[\frac{x-\min}{\max-x}\right]\right)^2\right]$
Cumulative distribution function :	$F(x)=\Phi\left[\alpha_1+\alpha_2\ln\left[\frac{x-\min}{\max-x}\right]\right]$ where $F(x)=\Phi[\bullet]$ is the distribution function for a Normal(0,1).
Parameter restriction :	$\alpha_2>0$, $\max>\min$
Domain :	$\min<x<\max$
Mean :	Complicated
Mode :	Complicated
Variance :	Complicated
Skewness :	Complicated
Kurtosis :	Complicated

Johnson Unbounded

VoseJohnsonU(α1,α2,β,γ)

Graphs



Uses

The main use of the Johnson unbounded distribution is that it can be made to have any combination of skewness and kurtosis. Thus, it provides a flexible distribution to fit to data by matching these moments. That said, it is an infrequently used distribution in risk analysis.

The distribution name comes from Johnson (1949) who proposed a system for categorizing distributions, in much the same spirit that Pearson did. Johnson's idea was to translate distributions to be a function of a unit Normal distribution, one of the few distributions for which there were good tools available at the time to handle.

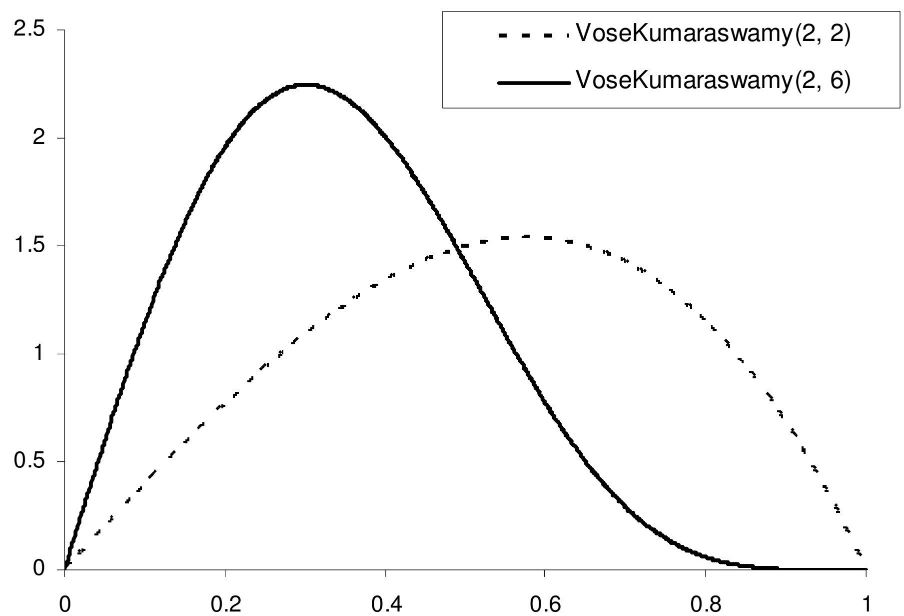
Equations

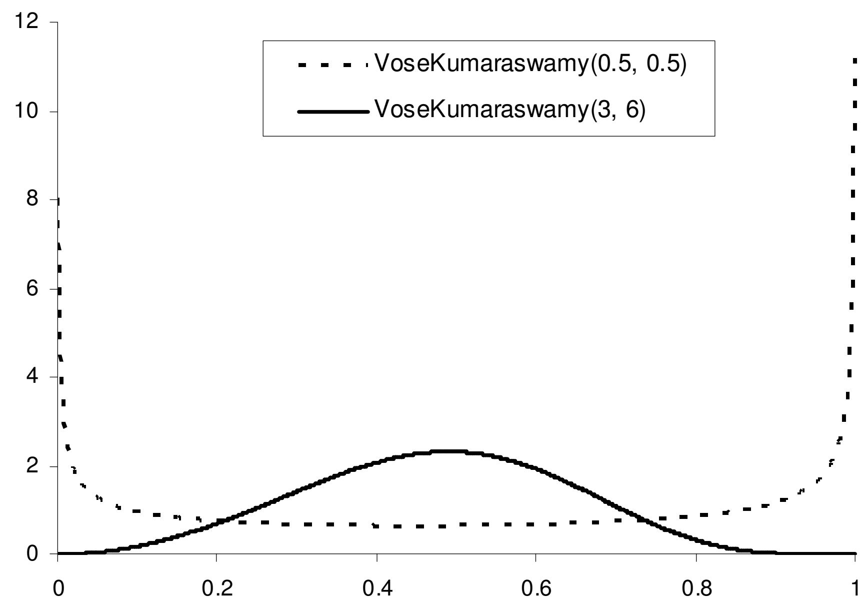
Probability density function :	$f(x)=\frac{\alpha_2}{\sqrt{2\pi\left((x-\gamma)^2+\beta^2\right)}}\exp\left[-\frac{1}{2}\left(\alpha_1+\alpha_2\ln\left[\frac{x-\gamma}{\beta}+\sqrt{\left(\frac{x-\gamma}{\beta}\right)^2+1}\right]\right)^2\right]$
Cumulative distribution function :	$F(x)=\Phi\left[\alpha_1+\alpha_2\ln\left[\frac{x-\gamma}{\beta}+\sqrt{\left(\frac{x-\gamma}{\beta}\right)^2+1}\right]\right]$
Parameter restriction :	$\beta>0,\alpha_2>0$
Domain :	$-\infty<x<+\infty$
Mean :	$\gamma-\beta\exp\left[\frac{1}{2\alpha_2^2}\right]\sinh\left(\frac{\alpha_1}{\alpha_2}\right)$
Mode :	Complicated
Variance :	Complicated
Skewness :	Complicated
Kurtosis :	Complicated

Kumaraswamy

VoseKumaraswamy(α,β)

Graphs





Uses

The Kumaraswamy distribution is not widely used but, for example, it has been applied to model the storage volume of a reservoir (Fletcher and Ponnambalam, 1996) and system design. It has a simple form for its density and cumulative distributions, and is very flexible like the Beta distribution (which does not have simple forms for these functions). It will probably have a lot more applications as it becomes better known.

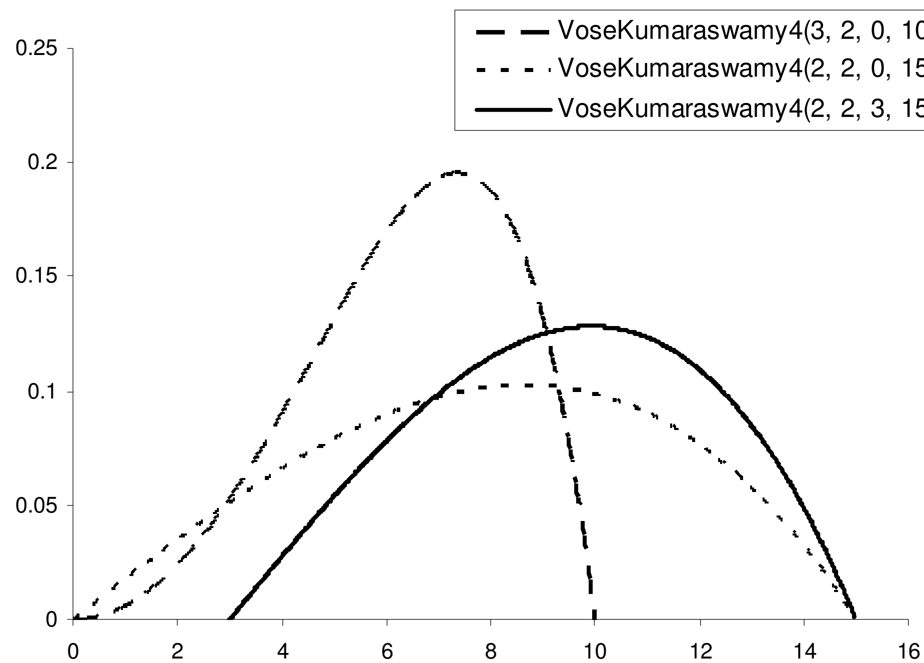
Equations

Probability density function :	$f(x) = \alpha\, \beta\, x^{\alpha-1}(1-x^{\alpha})^{\beta-1}$
Cumulative distribution function :	$F(x) = 1-\left(1-x^{\alpha}\right)^{\beta}$
Parameter restriction :	$\alpha > 0, \beta > 0$
Domain :	$0 \le x \le 1$
Mean :	$\beta \dfrac{\Gamma\left(1+\frac{1}{\alpha}\right)\Gamma(\beta)}{\Gamma\left(1+\beta+\frac{1}{\alpha}\right)} \equiv \mu$
Mode :	$\left(\dfrac{\alpha-1}{\alpha\beta-1}\right)^{\frac{1}{\alpha}}$ if $\alpha > 1 and\ \beta > 1$ if $\alpha < 1$ if $\beta < 1$
Variance :	$\beta \dfrac{\Gamma\left(1+\frac{2}{\alpha}\right)\Gamma(\beta)}{\Gamma\left(1+\beta+\frac{2}{\alpha}\right)} - \mu^2 \equiv V$
Skewness :	$\dfrac{1}{V^{3/2}}\left(\beta \dfrac{\Gamma\left(1+\frac{3}{\alpha}\right)\Gamma(\beta)}{\Gamma\left(1+\beta+\frac{3}{\alpha}\right)} - 3\beta \dfrac{\Gamma\left(1+\frac{2}{\alpha}\right)\Gamma(\beta)}{\Gamma\left(1+\beta+\frac{2}{\alpha}\right)}\mu + 2\mu^3\right)$
Kurtosis :	$\dfrac{1}{V^2}\left(\beta \dfrac{\Gamma\left(1+\frac{4}{\alpha}\right)\Gamma(\beta)}{\Gamma\left(1+\beta+\frac{4}{\alpha}\right)} - 4\beta \dfrac{\Gamma\left(1+\frac{3}{\alpha}\right)\Gamma(\beta)}{\Gamma\left(1+\beta+\frac{3}{\alpha}\right)}\mu + 6\beta \dfrac{\Gamma\left(1+\frac{2}{\alpha}\right)\Gamma(\beta)}{\Gamma\left(1+\beta+\frac{2}{\alpha}\right)}\mu^2 - 3\mu^4\right)$

Four-parameter Kumaraswamy

VoseKumaraswamy4(α,β,min,max)

Graphs



Uses

See Kumaraswamy distribution.

Equations

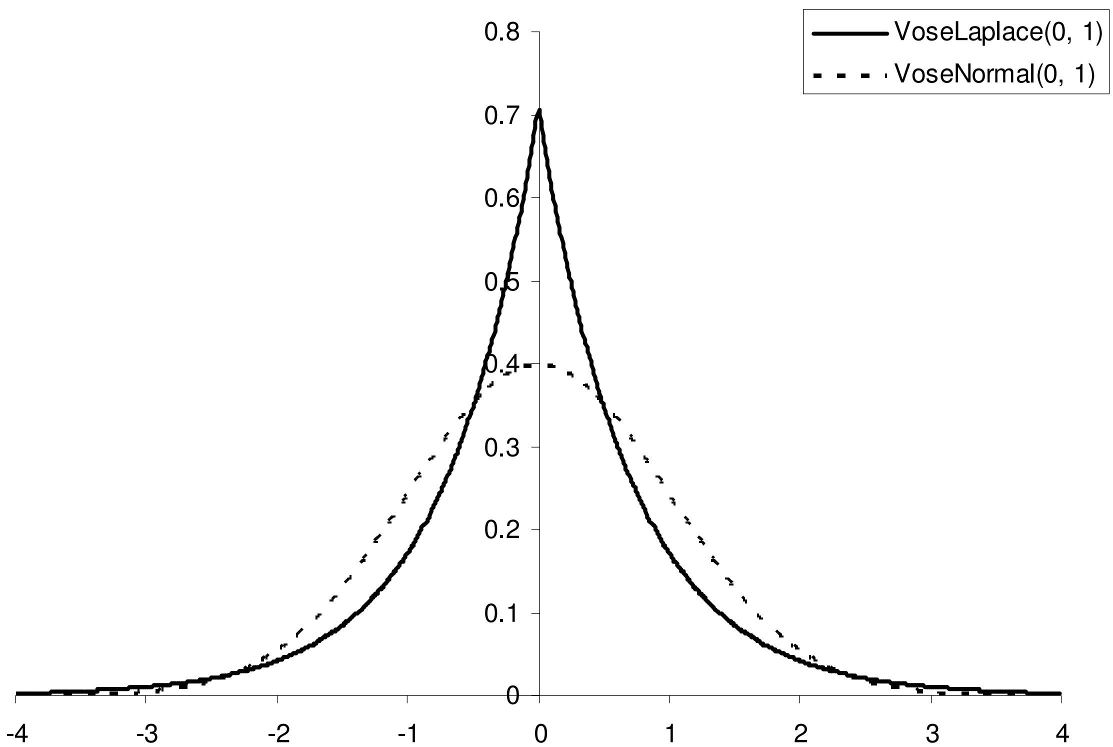
Probability density function :	$f(x)=\frac{\alpha\beta\, z^{\alpha-1}(1-z^{\alpha})^{\beta-1}}{(\max-\min)}$ where $z=\frac{x-\min}{\max-\min}$
Cumulative distribution function :	$F(x)=1-(1-z^{\alpha})^{\beta}$
Parameter restriction :	$\alpha>0,\beta>0$, $\min<\max$
Domain :	$\min\le x\le\max$
Mean :	$\left(\beta\frac{\Gamma\left(1+\frac{1}{\alpha}\right)\Gamma(\beta)}{\Gamma\left(1+\beta+\frac{1}{\alpha}\right)}(\max-\min)\right)+\min\equiv\mu$
Mode :	$\left(\frac{\alpha-1}{\alpha\beta-1}\right)^{\frac{1}{\alpha}}$ if $\alpha>1 and\ \beta>1$ if $\alpha<1$ if $\beta<1$
Variance :	$\left(\beta\frac{\Gamma\left(1+\frac{2}{\alpha}\right)\Gamma(\beta)}{\Gamma\left(1+\beta+\frac{2}{\alpha}\right)}-(\mu_K)^2\right)(\max-\min)^2\equiv V$ where $\mu_K=\mu_{Kumaraswamy(\alpha,\beta)}$
Skewness :	$\frac{1}{V_K^{3/2}}\left(\beta\frac{\Gamma\left(1+\frac{3}{\alpha}\right)\Gamma(\beta)}{\Gamma\left(1+\beta+\frac{3}{\alpha}\right)}-3\beta\frac{\Gamma\left(1+\frac{2}{\alpha}\right)\Gamma(\beta)}{\Gamma\left(1+\beta+\frac{2}{\alpha}\right)}\mu_K+2\mu_K^{3}\right)$
Kurtosis :	$\frac{1}{V_K^{2}}\left(\beta\frac{\Gamma\left(1+\frac{4}{\alpha}\right)\Gamma(\beta)}{\Gamma\left(1+\beta+\frac{4}{\alpha}\right)}-4\beta\frac{\Gamma\left(1+\frac{3}{\alpha}\right)\Gamma(\beta)}{\Gamma\left(1+\beta+\frac{3}{\alpha}\right)}\mu_K+6\beta\frac{\Gamma\left(1+\frac{2}{\alpha}\right)\Gamma(\beta)}{\Gamma\left(1+\beta+\frac{2}{\alpha}\right)}\mu_K^{2}-3\mu_K^{4}\right)$

Laplace

VoseLaplace(μ,σ)

Graphs

If X and Y are two identical independent Exponential(1/σ) distributions, and if X is shifted μ to the right of Y, then (X-Y) is a Laplace(μ, σ) distribution. The Laplace distribution has a strange shape with a sharp peak and tails that are longer than tails of a Normal distribution. The figure below plots a Laplace(0,1) against a Normal(0,1) distribution:



Uses

The Laplace has found a variety of very specific uses, but they nearly all relate to the fact that it has long tails.

Comments

When μ = 0, and σ = 1 we have the standard form of the Laplace distribution, which is also occasionally called 'Poisson's first law of error'. The Laplace distribution is also known as the Double-Exponential distribution (though the Gumbel Extreme Value distribution also takes this name), the Two-Tailed Exponential and the Bilateral Exponential distribution.

Equations

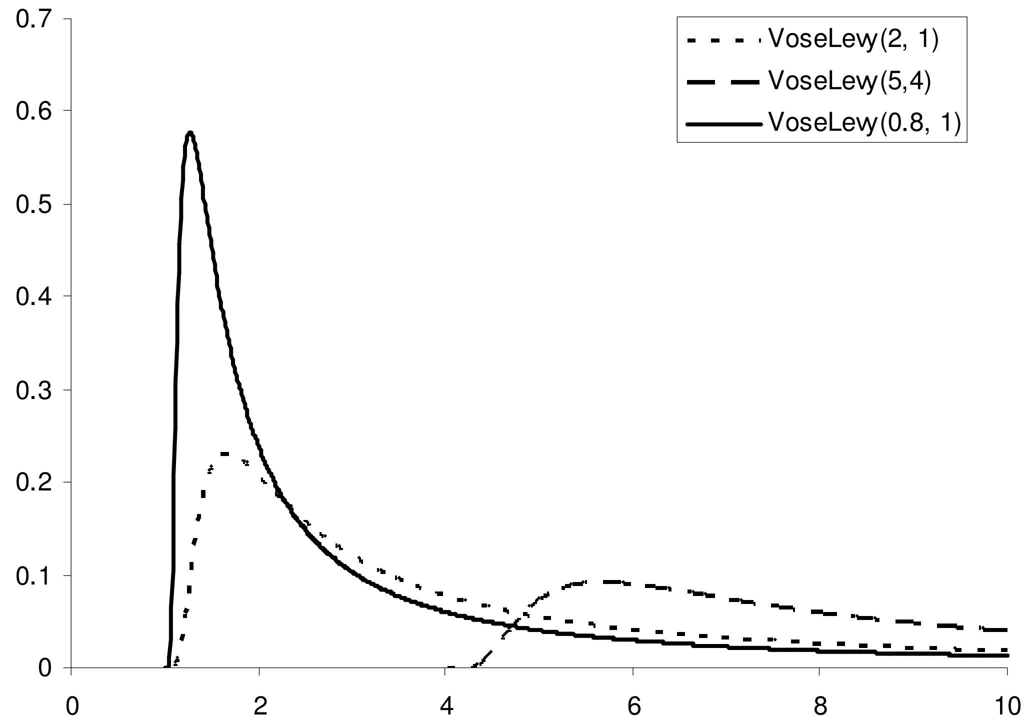
Probability density function :	$f(x)=\frac{1}{\sqrt{2}\,\sigma}\exp\left[-\frac{\sqrt{2}\lvert x-\mu\rvert}{\sigma}\right]$
Cumulative distribution function :	$F(x)=\frac{1}{2}\exp\left[-\frac{\sqrt{2}\lvert x-\mu\rvert}{\sigma}\right]$ if $x<\mu$ $F(x)=1-\frac{1}{2}\exp\left[-\frac{\sqrt{2}\lvert x-\mu\rvert}{\sigma}\right]$ if $x\geq\mu$
Parameter restriction :	$-\infty<\mu<+\infty,\sigma>0$
Domain :	$-\infty<x<+\infty$
Mean :	μ
Mode :	μ
Variance :	σ^2
Skewness :	0
Kurtosis :	6

Levy

VoseLevy(c,a)

Graphs

The Lévy distribution, named after Paul Pierre Lévy, is one of the few distributions that are stable and that have probability density functions that are analytically expressible. The others are the normal distribution and the Cauchy distribution



Uses

The Levy distribution is sometimes used in financial engineering to model price changes.
This distribution takes into account the leptokurtosis ('fat' tails) one sometimes empirically observes in price changes on financial markets.

Equations

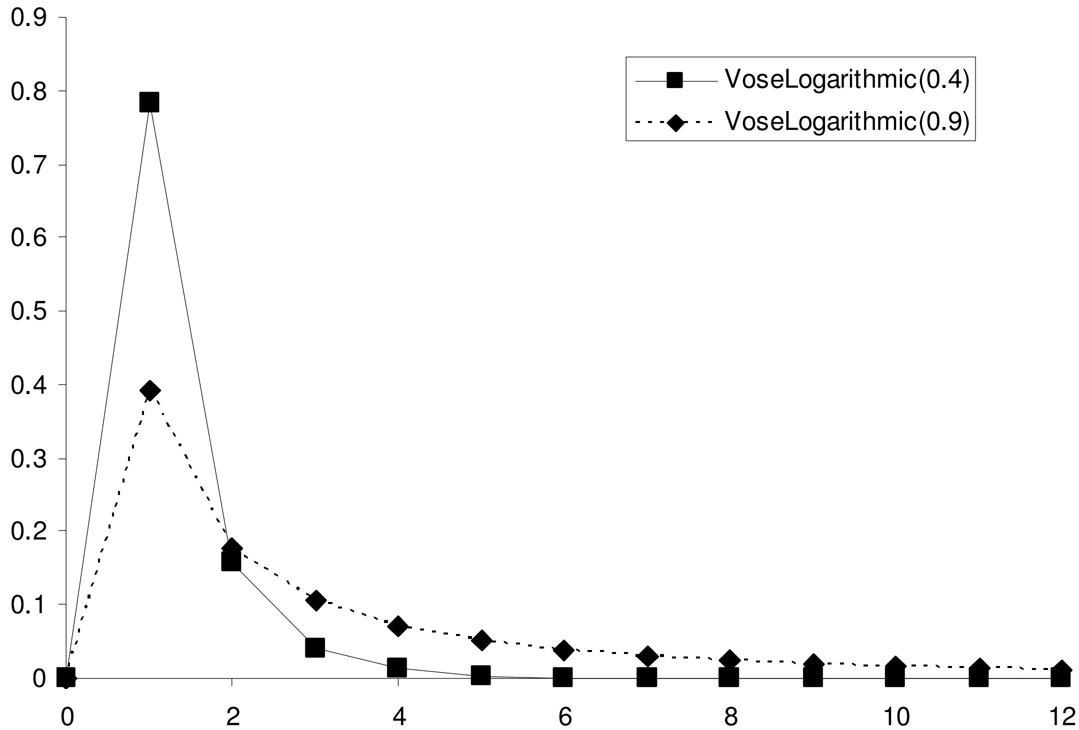
Probability density function :	$f(x) = \sqrt{\frac{c}{2\pi}} \frac{e^{-c/2(x-a)}}{(x-a)^{3/2}}$
Cumulative distribution function :	No closed form
Parameter restriction :	$c > 0$
Domain :	$x \geq a$
Mean :	Infinite
Mode :	$a + \frac{c}{3}$
Variance :	Infinite
Skewness :	Undefined
Kurtosis :	Undefined

Logarithmic (Series)

VoseLogarithmic(θ)

Graphs

The logarithmic distribution (sometimes known as the Logarithmic Series distribution) is a discrete, positive distribution, peaking at x = 1, with one parameter and a long right tail. The figures below show two examples of the Logarithmic distribution.



Uses

The logarithmic distribution is not very commonly used in risk analysis. However, it has been used to describe, for example: the number of items purchased by a consumer in a particular period; the number of bird and plant species in an area; and the number of parasites per host. There is some theory that relates the latter two to an observation by Newcomb (1881) that the frequency of use of different digits in natural numbers followed a Logarithmic distribution.

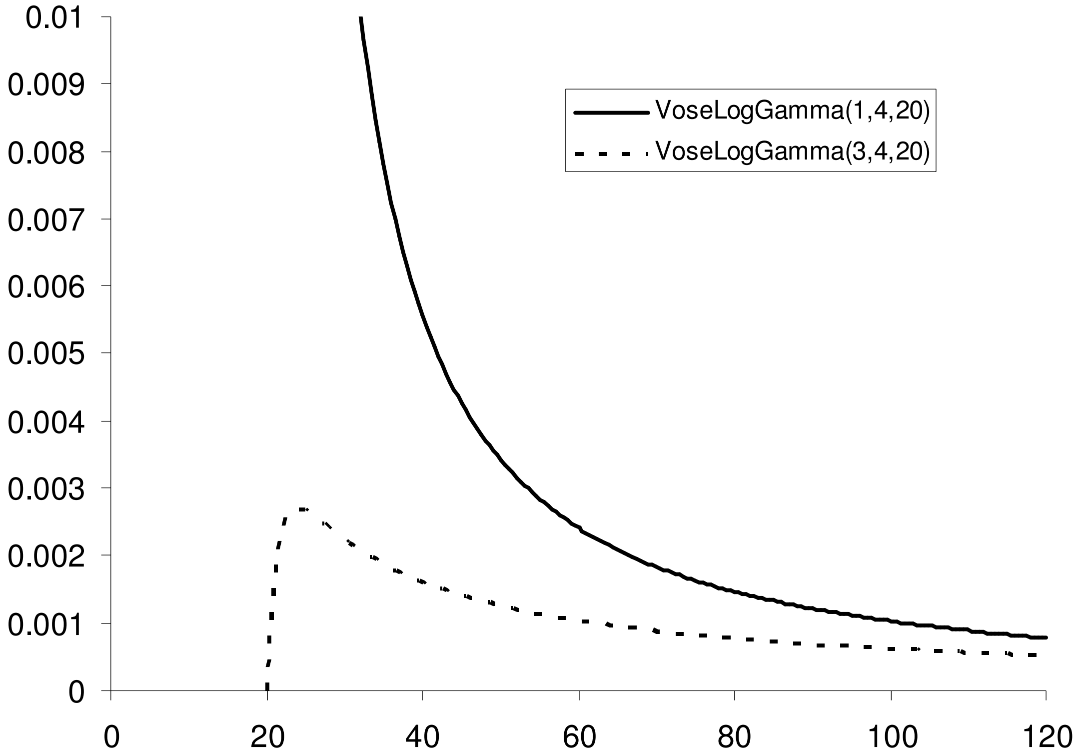
Equations

Probability mass function :	$f(x) = \frac{-\theta^x}{x\ln(1-\theta)}$
Cumulative distribution function :	$F(x) = \sum_{i=1}^{\lfloor x \rfloor} \frac{-\theta^i}{i\ln(1-\theta)}$
Parameter restriction :	$0 < \theta < 1$
Domain :	$x = \{1,2,3,...\}$
Mean :	$\frac{\theta}{(\theta-1)\ln(1-\theta)}$
Mode :	1
Variance :	$\mu\left((1-\theta)^{-1} - \mu\right) \equiv V$ where μ is the mean
Skewness :	$\frac{-\theta}{(1-\theta)^3 V^{3/2}\ln(1-\theta)}\left(1+\theta+\frac{3\theta}{\ln(1-\theta)}+\frac{2\theta^2}{\ln^2(1-\theta)}\right)$
Kurtosis :	$\frac{-\theta}{(1-\theta)^4 V^2\ln(1-\theta)}\left(1+4\theta+\theta^2+\frac{4\theta(1+\theta)}{\ln(1-\theta)}+\frac{6\theta^2}{\ln^2(1-\theta)}+\frac{3\theta^3}{\ln^3(1-\theta)}\right)$

LogGamma

VoseLogGamma(α,β,γ)

Graphs



Uses

A Variable X is LogGamma distributed if its natural log is Gamma distributed. In ModelRisk we include an extra shift parameter γ because a standard LogGamma distribution has a minimum value of 1 when the Gamma variable = 0. Thus:

LogGamma(α, β, γ) = EXP[Gamma(α, β)] +(γ-1)

The LogGamma distribution is sometimes used to model the distribution of claim size in insurance. Set γ = 1 to get the standard LogGamma(α, β) distribution.

Equations

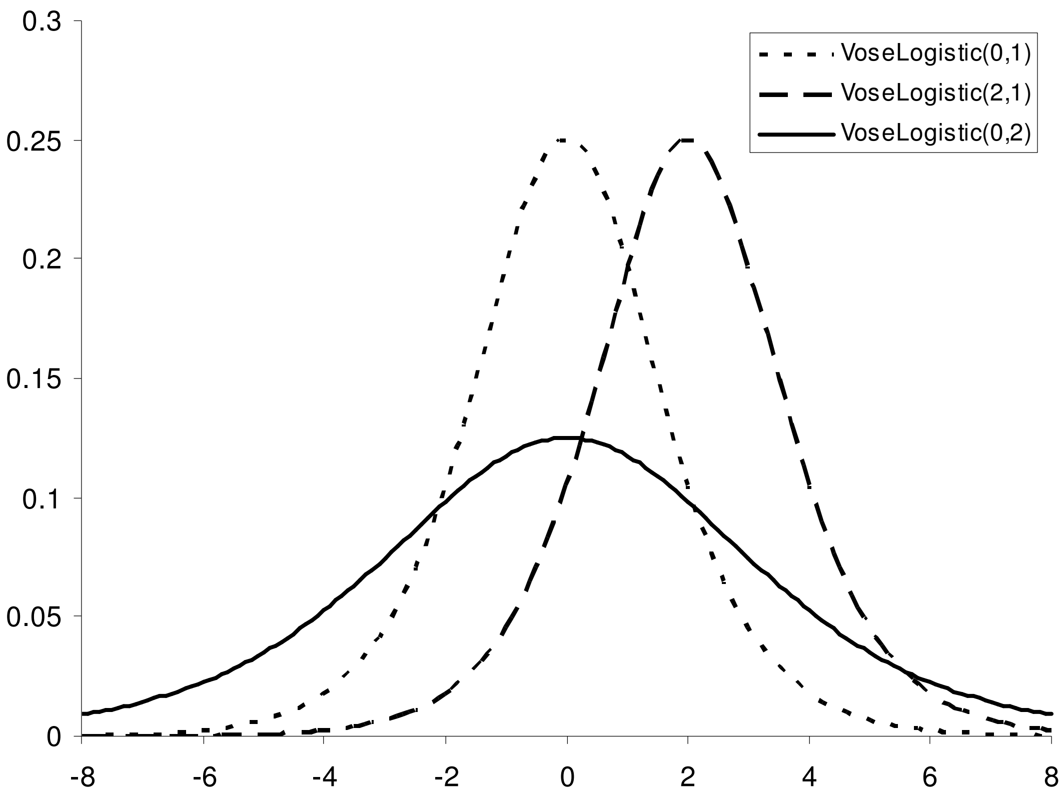
Probability density function :	$f(x)=\frac{(\ln[x-\gamma+1])^{\alpha-1}(x-\gamma+1)^{-\left(\frac{1+\beta}{\beta}\right)}}{\beta^{\alpha}\Gamma(\alpha)}$	
Cumulative distribution function :	No closed form	
Parameter restriction :	$\alpha>0,\beta>0$	
Domain :	$x\geq\gamma$	
Mean :	$(1-\beta)^{-\alpha}+\gamma-1$	if $\beta<1$
Mode :	$\exp\left[\frac{\beta(\alpha-1)}{\beta+1}\right]+\gamma-1$	if $\alpha>1$
	0	else
Variance :	$(1-2\beta)^{-\alpha}-(1-\beta)^{-2\alpha}\equiv V$	if $\beta<1/2$
Skewness :	$\frac{(1-3\beta)^{-\alpha}-3(1-3\beta+2\beta^{2})^{-\alpha}+2(1-\beta)^{-3\alpha}}{V^{3/2}}$	if $\beta<1/3$
Kurtosis :	$\frac{(1-4\beta)^{-\alpha}-4(1-4\beta+3\beta^{2})^{-\alpha}+6(1-2\beta)^{-\alpha}(1-\beta)^{-2\alpha}-3(1-\beta)^{-4\alpha}}{V^{2}}$	if $\beta<1/4$

Logistic

VoseLogistic(α,β)

Graphs

The Logistic distribution looks similar to the Normal distribution but has a kurtosis of 4.2 compared to the Normal kurtosis of 3. Examples of the Logistic distribution are given below:



Uses

The Logistic distribution is popular in demographic and economic modeling because it is similar to the Normal distribution but somewhat more peaked. It does not appear often in risk analysis modeling.

Comments

The cumulative function has also been used as a model for a “growth curve”. Sometimes called the Sech-Squared distribution because its distribution function can be written in a form that includes a sech. Its mathematical derivation is the limiting distribution as n approaches infinity of the standardized mid-range (average of the highest and lowest values) of a random sample of size n from an exponential-type distribution.

Equations

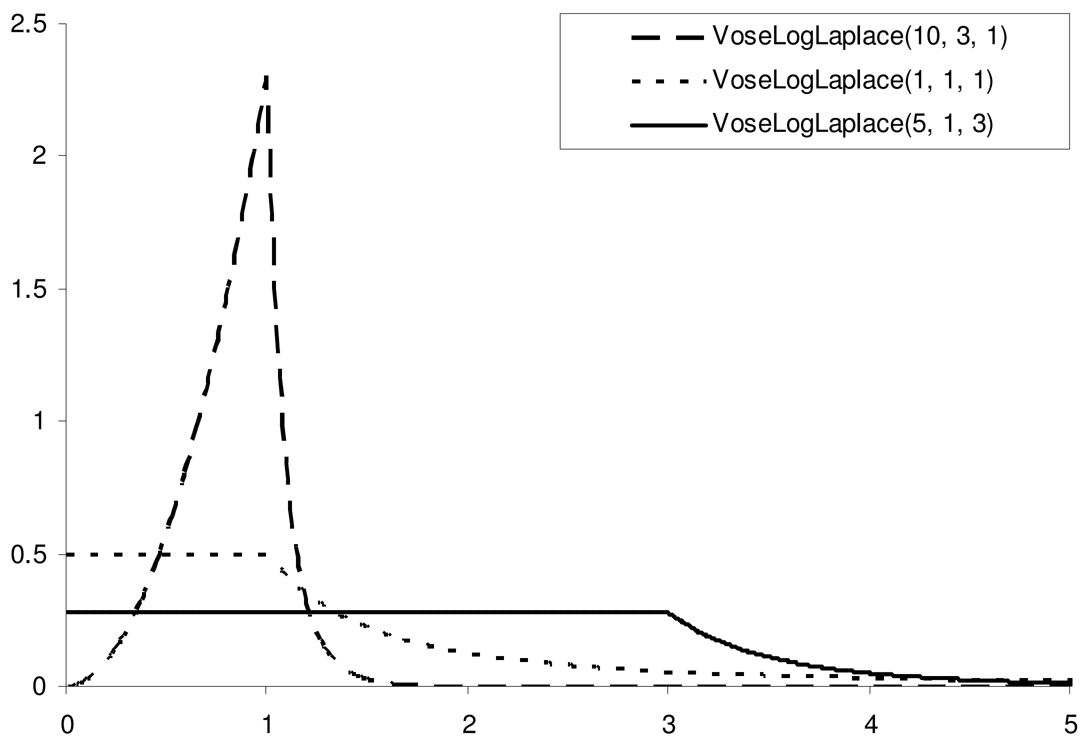
Probability density function :	$f(x)=\dfrac{z}{\beta(1+z)^2}$ where $z=\exp\left(-\dfrac{x-\alpha}{\beta}\right)$
Cumulative distribution function :	$F(x)=\dfrac{1}{1+z}$
Parameter restriction :	$\beta>0$
Domain :	$-\infty<x<+\infty$
Mean :	α
Mode :	α
Variance :	$\dfrac{\beta^2\pi^2}{3}$
Skewness :	0
Kurtosis :	4.2

LogLaplace

VoseLogLaplace(α,β,δ)

Graphs

Examples of the LogLaplace distribution are given below. δ is just a scaling factor, giving the location of the point of inflection of the density function. The LogLaplace distribution takes a variety of shapes, depending on the value of β. For example, when β = 1, the LogLaplace distribution is uniform for x < δ.



Uses

Kozubowski TJ and Podgórski K review many uses of the LogLaplace distribution. The most commonly quoted use (for the symmetric LogLaplace) has been for modeling 'moral fortune', a state of well-being that is the logarithm of income, based on a formula by Daniel Bernoulli.

The asymmetric LogLaplace distribution has been fit to pharmacokinetic and particle size data (particle size studies often show the log size to follow a tent'shaped distribution like the Laplace). It has been used to model growth rates, stock prices, annual gross domestic production, interest and forex rates. Some explanation for the goodness of fit of the LogLaplace has been suggested because of its relationship to Brownian motion stopped at a random exponential time.

Equations

Probability density function :	$f(x)=\frac{\alpha\beta}{\delta(\alpha+\beta)}\left(\frac{x}{\delta}\right)^{\beta-1}$ for $0\le x<\delta$ $f(x)=\frac{\alpha\beta}{\delta(\alpha+\beta)}\left(\frac{\delta}{x}\right)^{\alpha+1}$ for $x\ge\delta$
Cumulative distribution function :	$F(x)=\frac{\alpha}{(\alpha+\beta)}\left(\frac{x}{\delta}\right)^{\beta}$ for $0\le x<\delta$ $F(x)=1-\frac{\beta}{(\alpha+\beta)}\left(\frac{\delta}{x}\right)^{\alpha}$ for $x\ge\delta$
Parameter restriction :	$\alpha>0,\beta>0,\delta>0$
Domain :	$0<x<+\infty$
Mean :	$\delta\frac{\alpha\beta}{(\alpha-1)(\beta+1)}\equiv\mu$ for $\alpha>1$
Mode :	0 for $0<\beta<1$ No unique mode for $\beta=1$ δ for $\beta>1$
Variance :	$\delta^2\left(\frac{\alpha\beta}{(\alpha-2)(\beta+2)}-\left[\frac{\alpha\beta}{(\alpha-1)(\beta+1)}\right]^2\right)\equiv V$ for $\alpha>2$
Skewness :	$\frac{1}{V^{3/2}}\left(\frac{\delta^3\alpha\beta}{(\alpha-3)(\beta+3)}-3(V+\mu^2)\mu+2\mu^3\right)$ for $\alpha>3$
Kurtosis :	$\frac{1}{V^2}\left(\frac{\delta^4\alpha\beta}{(\alpha-4)(\beta+4)}-4\frac{\delta^3\alpha\beta}{(\alpha-3)(\beta+3)}\mu+6(V+\mu^2)\mu^2-3\mu^4\right)$ for $\alpha>4$

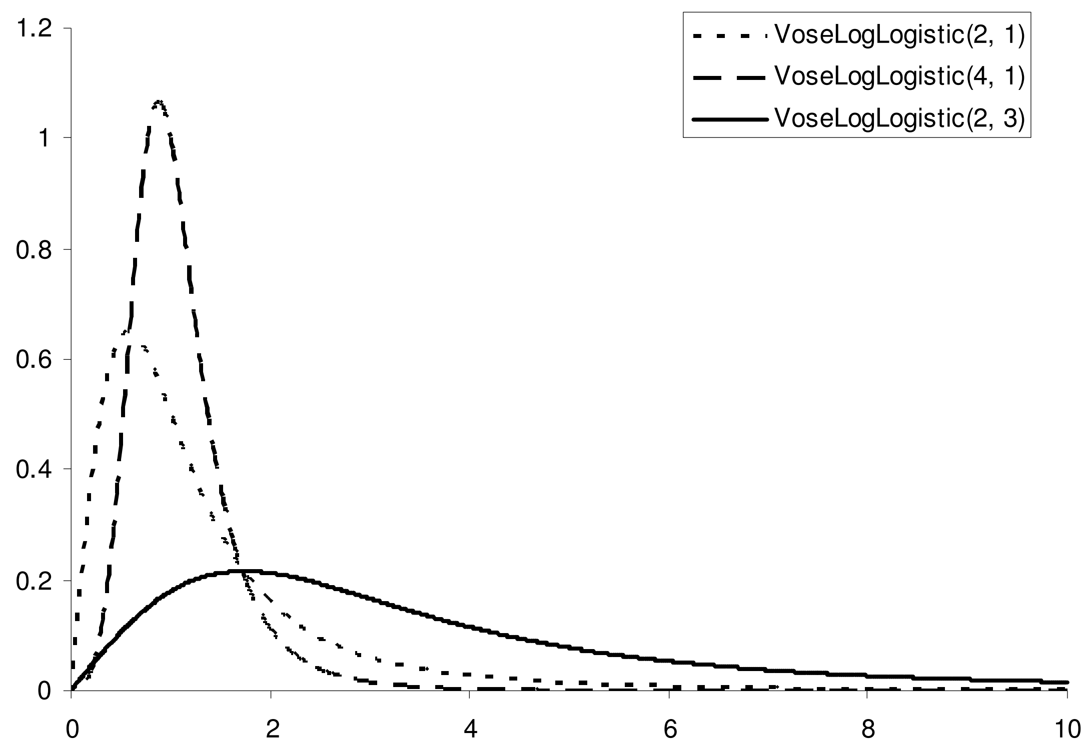
LogLogistic

VoseLogLogistic(α,β)

Graphs

When Log(X) takes a Logistic distribution then X takes a LogLogistic distribution. Their parameters are related as follows:

EXP(Logistic(α, β)) = LogLogistic(1/β, EXP(α))
LogLogistic(α, 1) is the standard LogLogistic distribution.



Uses

The LogLogistic distribution has the same relationship to the Logistic distribution that the Lognormal distribution has to the Normal distribution. If you feel that a variable is driven by some process that is the product of a number of variables, then a natural distribution to use is the Lognormal because of Central Limit Theorem. However, if one or two of these factors could be dominant, or correlated, so that the distribution is less spread than a lognormal, then the LogLogistic may be an appropriate distribution to try

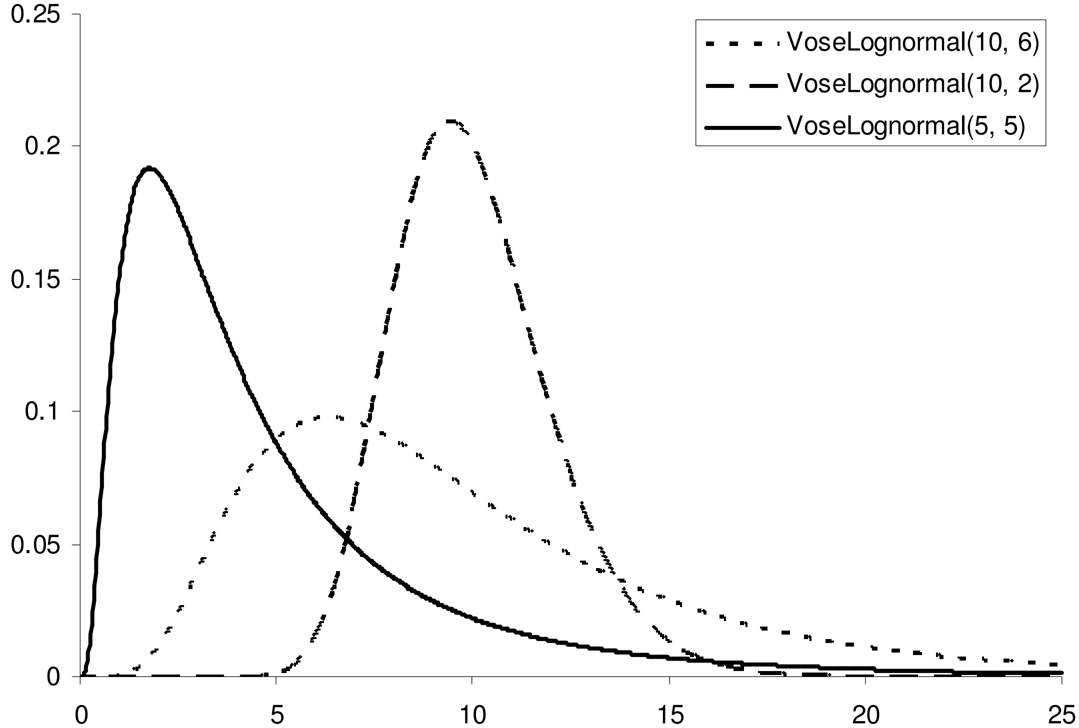
Equations

Probability density function :	$f(x)=\dfrac{\alpha\,x^{\alpha-1}}{\beta^{\alpha}\left[1+\left(\dfrac{x}{\beta}\right)\right]^{2}}$
Cumulative distribution function :	$F(x)=\dfrac{1}{1+\left(\dfrac{\beta}{x}\right)^{\alpha}}$
Parameter restriction :	$\alpha>0,\beta>0$
Domain :	$0\le x<+\infty$
Mean :	$\beta\,\theta\csc(\theta)$ where $\theta=\dfrac{\pi}{\alpha}$
Mode :	$\beta\left[\dfrac{\alpha-1}{\alpha+1}\right]^{\frac{1}{\alpha}}$ for $\alpha>1$ 0 for $\alpha\le 1$
Variance :	$\beta^{2}\,\theta\left[2\csc(2\theta)-\theta\csc^{2}(\theta)\right]$ for $\alpha>2$
Skewness :	$\dfrac{3\csc(3\theta)-6\theta\csc(2\theta)\csc(\theta)+2\theta^{2}\csc^{3}(\theta)}{\sqrt{\theta}\left[2\csc(2\theta)-\theta\csc^{2}(\theta)\right]^{\frac{3}{2}}}$ for $\alpha>3$
Kurtosis :	$\dfrac{6\theta^{2}\csc^{3}(\theta)\sec(\theta)+4\csc(4\theta)-3\theta^{3}\csc^{4}(\theta)-12\theta\csc(\theta)\csc(3\theta)}{\theta\left[2\csc(2\theta)-\theta\csc^{2}(\theta)\right]^{2}}$ for $\alpha>4$

Lognormal

VoseLognormal(μ,σ)

Graphs



Uses

The Lognormal distribution is useful for modeling naturally occurring variables that are the product of a number of other naturally occurring variables. Central Limit Theorem shows that the product of a large number of independent random variables is Lognormally distributed. For example, the volume of gas in a petroleum reserve is often Lognormally distributed because it is the product of the area of the formation, its thickness, formation pressure, porosity and the gas:liquid ratio.

Lognormal distributions often provide a good representation for a physical quantity that extend from zero to + infinity and is positively skewed, perhaps because some Central limit Theorem type of process is determining the variable's size. Lognormal distributions are also very useful for representing quantities that are thought of in orders of magnitude. For example, if a variable can be estimated to within a factor of 2 or to within an order of magnitude, the Lognormal distribution is often a reasonable model.

Lognormal distributions have also been used to model lengths of words and sentences in a document, particle sizes in aggregates, critical doses in pharmacy and incubation periods of infectious diseases, but one reason the Lognormal distribution appears so frequently is because it is easy to fit and test (one simply transforms the data to logs and manipulate as a Normal distribution), and so observing its use in your field does not necessarily mean it is a good model: it may just have been a convenient one. Modern software and statistical techniques have removed much of the need for assumptions of normality, so be cautious about using the Lognormal because it has always been that way.

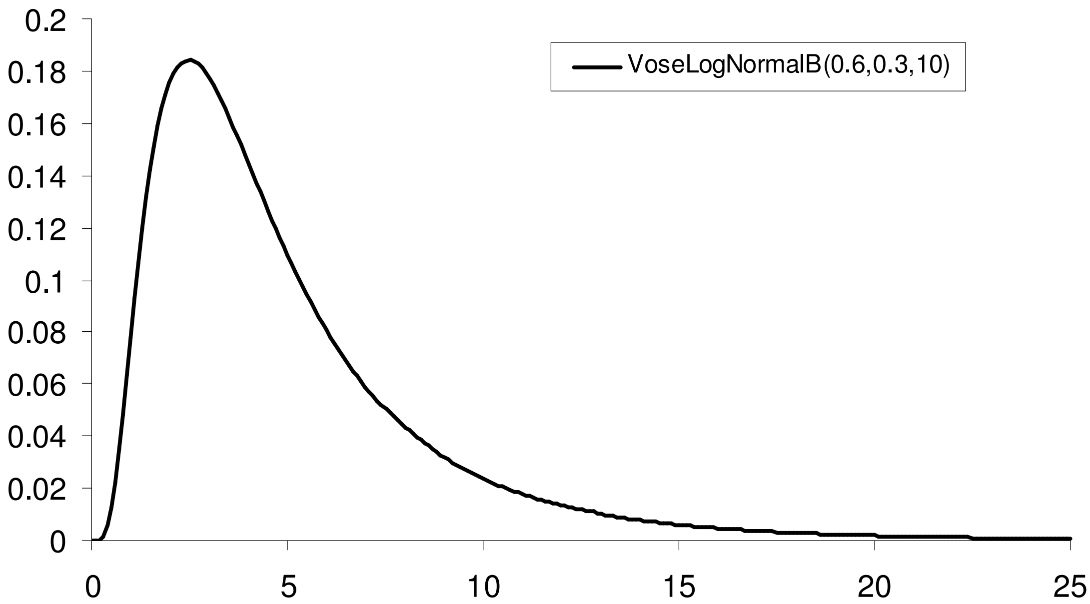
Equations

Probability density function :	$f(x)=\frac{1}{x\sqrt{2\pi\sigma_1^2}}\exp\left[-\frac{\left(\ln[x]-\mu_1^2\right)}{2\sigma_1^2}\right]$ where $\mu_1=\ln\left[\frac{\mu^2}{\sqrt{\sigma^2+\mu^2}}\right]$ and $\sigma_1=\sqrt{\ln\left[\frac{\sigma^2+\mu^2}{\mu^2}\right]}$
Cumulative distribution function :	No closed form
Parameter restriction :	$\sigma>0,\mu>0$
Domain :	$x\geq 0$
Mean :	μ
Mode :	$\exp\left(\mu_1-\sigma_1^2\right)$
Variance :	σ^2
Skewness :	$\left(\frac{\sigma}{\mu}\right)^3+3\left(\frac{\sigma}{\mu}\right)$
Kurtosis :	$z^4+2z^3+3z^2-3$ where $z=1+\frac{\sigma}{\mu}$

LognormalB

VoseLognormalB(μ,σ,B)

Graphs



Uses

See Lognormal distribution.

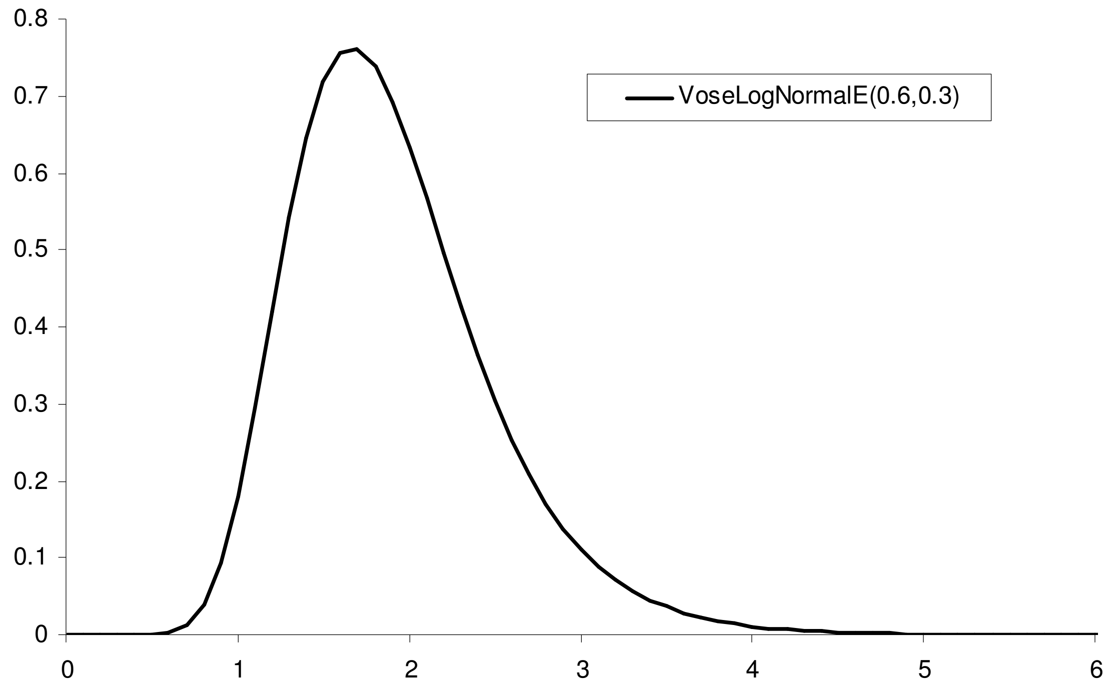
Equations

Probability density function :	$f(x)=\frac{\exp\left[-\frac{(\ln x-m)^2}{2V}\right]}{x\sqrt{V\,2\pi}}$ where $V=\sigma^2\ln B$ and $m=\mu\ln B$
Cumulative distribution function :	No closed form
Parameter restriction :	$\sigma>0,\mu>0$, $B>0$
Domain :	$x\geq 0$
Mean :	$\exp\left(m+\frac{V}{2}\right)$
Mode :	$\exp(m-V)$
Variance :	$\exp(2m+V)(\exp(V)-1)$
Skewness :	$(\exp(V)+2)\sqrt{\exp(V)-1}$
Kurtosis :	$\exp(4V)+2\exp(3V)+3\exp(2V)-3$

LognormalE

VoseLognormalE(μ,σ)

Graphs



Uses

See Lognormal distribution.

Equations

Probability density function :	$f(x)=\dfrac{\exp\left[-\dfrac{(\ln x-\mu)^2}{2\sigma^2}\right]}{x\sigma\sqrt{2\pi}}$
Cumulative distribution function :	No closed form
Parameter restriction :	$\sigma>0,\mu>0$
Domain :	$x\geq 0$
Mean :	$\exp\left(\mu+\dfrac{\sigma^2}{2}\right)$
Mode :	$\exp(\mu-\sigma^2)$
Variance :	$\exp(2\mu+\sigma^2)(\exp(\sigma^2)-1)$
Skewness :	$(\exp(\sigma^2)+2)\sqrt{\exp(\sigma^2)-1}$
Kurtosis :	$\exp(4\sigma^2)+2\exp(3\sigma^2)+3\exp(2\sigma^2)-3$

Modified PERT

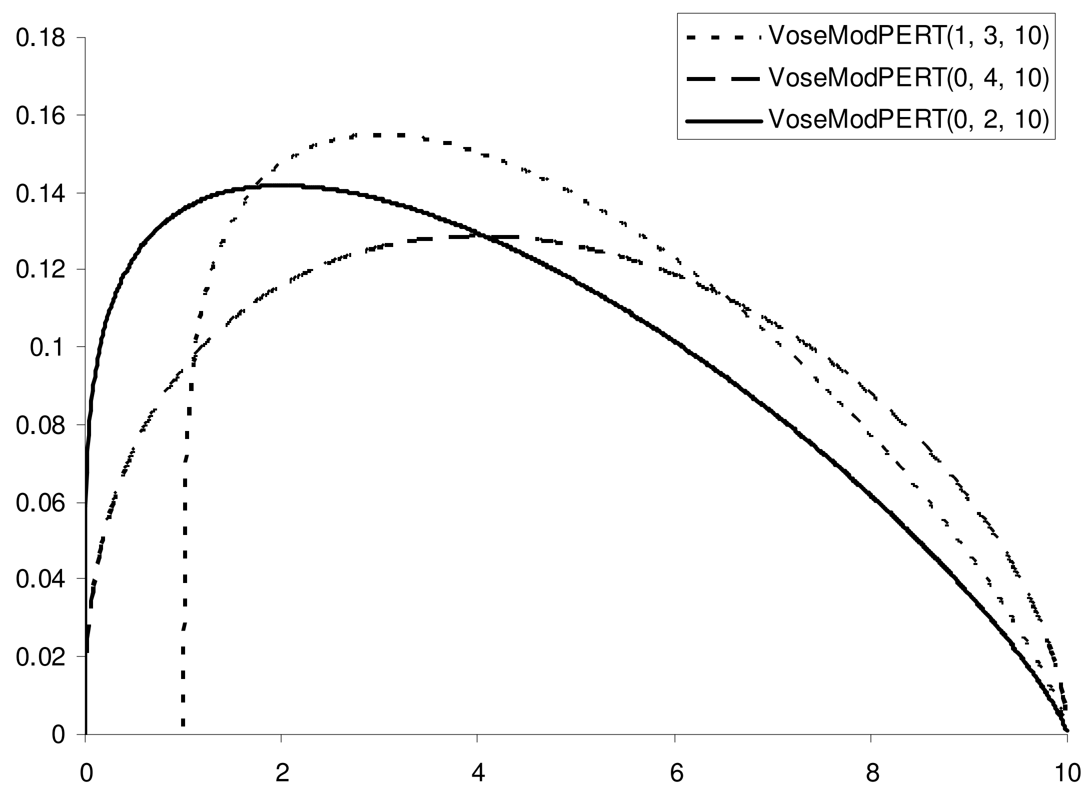
VoseModPERT(min,mode,max,γ)

Graphs

David Vose developed a modification of the PERT distribution with minimum min, most likely mode and maximum max to produce shapes with varying degrees of uncertainty for the min,mode,max values by changing the assumption about the mean:

$$\frac{\min + \gamma\,\text{mode} + \max}{\gamma + 2} \equiv \mu$$

In the standard PERT, γ = 4, which is the PERT network assumption that the best estimate of the duration of a task = (min + 4mode + max) /6. However, if we increase the value of γ, the distribution becomes progressively more peaked and concentrated around Mode (and therefore less uncertain). Conversely, if we decrease γ the distribution becomes flatter and more uncertain. The figure below illustrates the effect of three different values of γ for a modified PERT(5,7,10) distribution.



Uses

This modified PERT distribution can be very useful in modeling expert opinion. The expert is asked to estimate three values (minimum, most likely and maximum). Then a set of modified PERT distributions are plotted and the expert is asked to select the shape that fits his/her opinion most accurately.

Equations

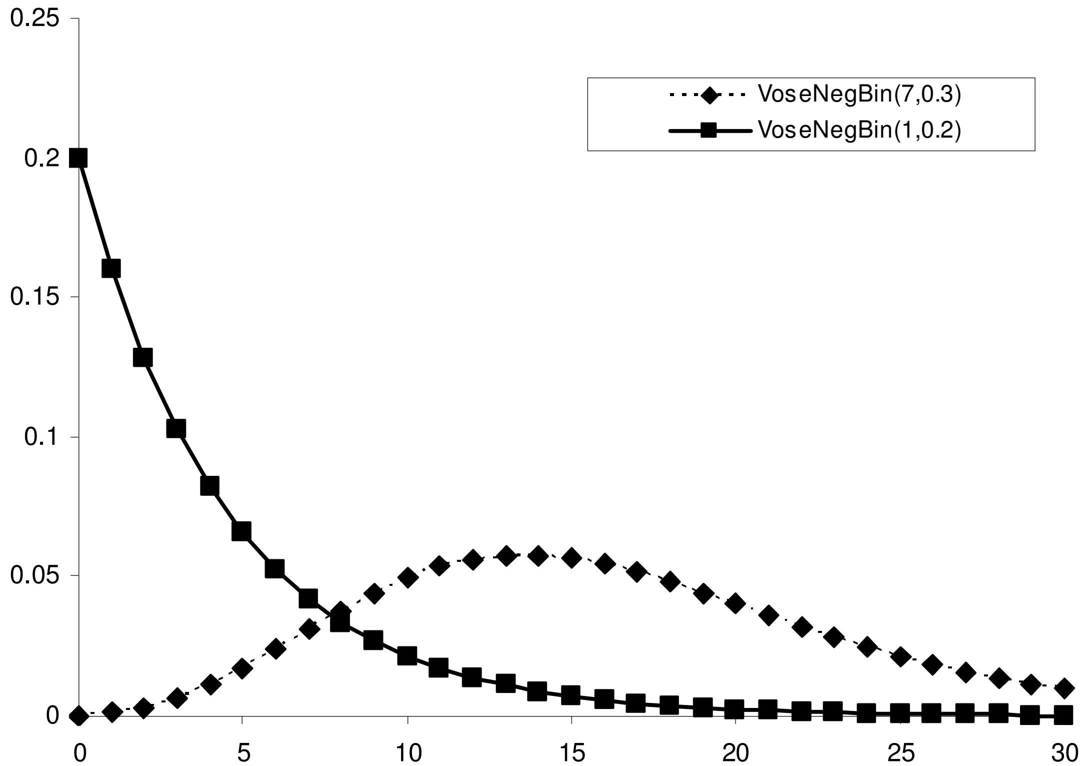
Probability density function :	$f(x)=\frac{(x-\min)^{\alpha_1-1}(\max-x)^{\alpha_2-1}}{B(\alpha_1,\alpha_2)(\max-\min)^{\alpha_1+\alpha_2-1}}$ where $\alpha_1=1+\gamma\left(\frac{\text{mode}-\min}{\max-\min}\right)$, $\alpha_2=1+\gamma\left(\frac{\max-\text{mode}}{\max-\min}\right)$ and $B(\alpha_1,\alpha_2)$ is a Beta function
Cumulative distribution function :	$F(x)=\frac{B_z(\alpha_1,\alpha_2)}{B(\alpha_1,\alpha_2)}\equiv I_z(\alpha_1,\alpha_2)$ where $z=\frac{x-\min}{\max-\min}$ and $B_z(\alpha_1,\alpha_2)$ is an incomplete Beta function
Parameter restriction :	$\min<\text{mode}<\max$, $\gamma>0$
Domain :	$\min\leq x\leq\max$
Mean :	$\frac{\min+\gamma\,\text{mode}+\max}{\gamma+2}\equiv\mu$
Mode :	mode
Variance :	$\frac{(\mu-\min)(\max-\mu)}{\gamma+3}$
Skewness :	$\frac{\min+\max-2\mu}{4}\sqrt{\frac{7}{(\mu-\min)(\max-\mu)}}$
Kurtosis :	$3\frac{(\alpha_1+\alpha_2+1)\left(2(\alpha_1+\alpha_2)^2+\alpha_1\alpha_2(\alpha_1+\alpha_2-6)\right)}{\alpha_1\alpha_2(\alpha_1+\alpha_2+2)(\alpha_1+\alpha_2+3)}$

Negative Binomial

VoseNegBin(s,p)

Graphs

The Negative Binomial distribution estimates the number of failures there will be before s successes are achieved where there is a probability p of success with each trial. Examples of the Negative Binomial distribution are shown below:



Uses

Binomial examples

The NegBin distribution has two applications for a binomial process:

- The number of failures in order to achieve s successes = NegBin(s,p);
- The number of failures there might have been when we have observed s successes = NegBin(s+1,p)

The first use is when we know that we will stop at the sth success. The second is when we only know that there had been a certain number of successes.

For example, a hospital has received a total of 17 people with a rare disease in the last month. The disease has a long incubation period. There have been no new admissions for this disease for a fair number of days. The hospital knows that people infected with this problem have a 65% chance of showing symptoms and thus turning up at the hospital. They are worried about how many people there are infected in the outbreak that have not turned up in hospital and may therefore infect others. The answer is NegBin(17+1,65%). IF we knew (we don't) that the last person to be infected was ill, the answer would be NegBin(17,65%). The total number infected would be 17+NegBin(17+1,65%).

Poisson example

The Negative Binomial distribution is frequently used in accident statistics and other Poisson processes because the Negative Binomial distribution can be derived as a Poisson random variable whose rate parameter lambda is itself random and Gamma distributed, i.e.:

Poisson(Gamma(α, β)) = NegBin(α, 1/(β+1))

The Negative Binomial distribution therefore also has applications in the insurance industry, where for example the rate at which people have accidents is affected by a random variable like the weather, or in marketing. This has a number of implications: it means that the Negative Binomial distribution must have a greater spread than a Poisson distribution with the same mean; and it means that if one attempts to fit frequencies of random events to a Poisson distribution but find the Poisson distribution too narrow, then a Negative Binomial can be tried and if that fits well, this suggests that the Poisson rate is not constant but random, and can be approximated by the corresponding Gamma distribution.

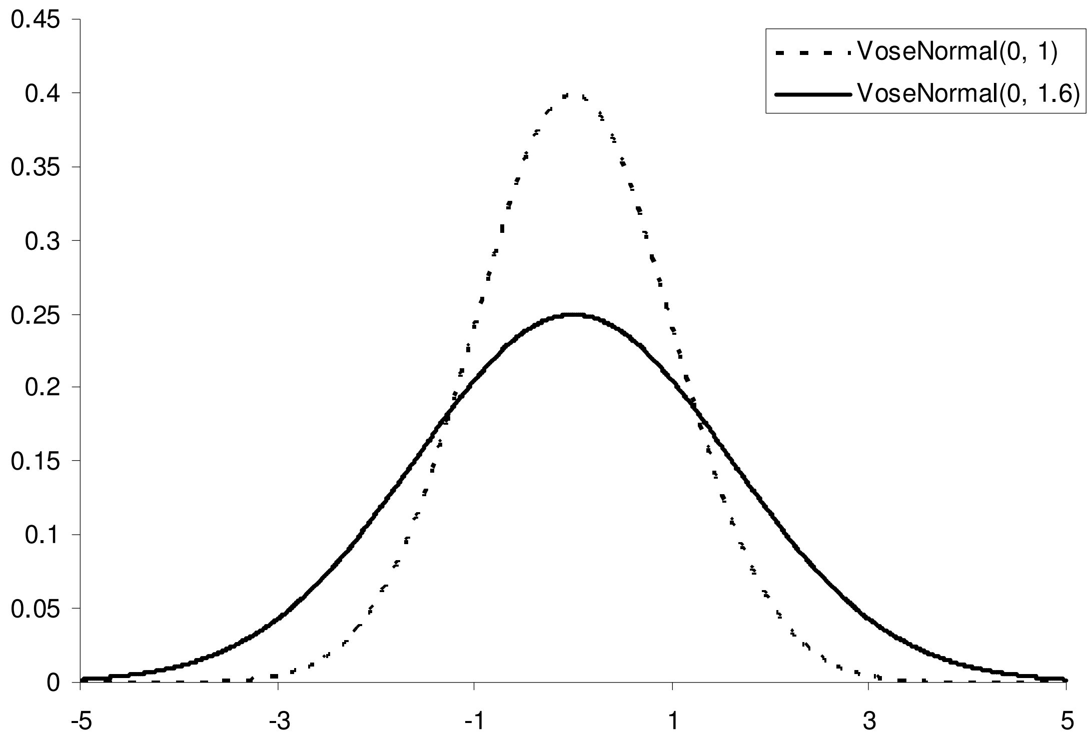
Equations

Probability mass function :	$f(x) = \binom{s+x-1}{x} p^s (1-p)^x$
Cumulative distribution function :	$F(x) = \sum_{i=0}^{\lfloor x \rfloor} \binom{s+i-1}{i} p^s (1-p)^i$
Parameter restriction :	$0 < p \le 1$ $s > 0$, s is an integer
Domain :	$x = \{0,1,2,\ldots\}$
Mean :	$\frac{s(1-p)}{p}$
Mode :	$z, z+1$ if z is an integer $\lvert z \rvert + 1$ otherwise where $z = \frac{s(1-p)-1}{p}$
Variance :	$\frac{s(1-p)}{p^2}$
Skewness :	$\frac{2-p}{\sqrt{s(1-p)}}$
Kurtosis :	$3 + \frac{6}{s} + \frac{p^2}{s(1-p)}$

Normal

VoseNormal(μ,σ)

Graphs



Uses

1. Modelling a naturally occurring variable

The Normal, or Gaussian, distribution occurs in a wide variety of applications due, in part, to Central Limit Theorem which makes it a good approximation to many other distributions. It is frequently observed that variations of a naturally occurring variable are approximately Normally distributed: for example, the height of adult European males, arm span, etc. Population data tend to approximately fit to a normal curve, but the data usually have a little more density in the tails.

2. Distribution of errors

A Normal distribution is frequently used in statistical theory for the distribution of errors (for example, in least squares regression analysis).

3. Approximation of uncertainty distribution

A basic rule of thumb in statistics is that the more data you have, the more the uncertainty distribution of the estimated parameter approaches a Normal. There are various ways of looking at it: from a Bayesian perspective, a Taylor series expansion of the posterior density is helpful; from a frequentist perspective, a Central limit Theorem argument is often appropriate: Binomial example; Poisson example.

4. Convenience distribution

The most common use of a Normal distribution is simply for its convenience. For example to add Normally distributed (uncorrelated and correlated) random variables, one combines the means and variances in simple ways to obtain another normal distribution.

Classical statistics has grown up concentrating on the Normal distribution, including trying to transform data so that they look Normal. The Student-t distribution, and the Chi Squared distribution are based on a Normal assumption. It's the distribution we learn at college. But take care that when you select a Normal distribution it is not simply through lack of imagination: that you have a good reason for its selection, because there are many other distributions that may be far more appropriate.

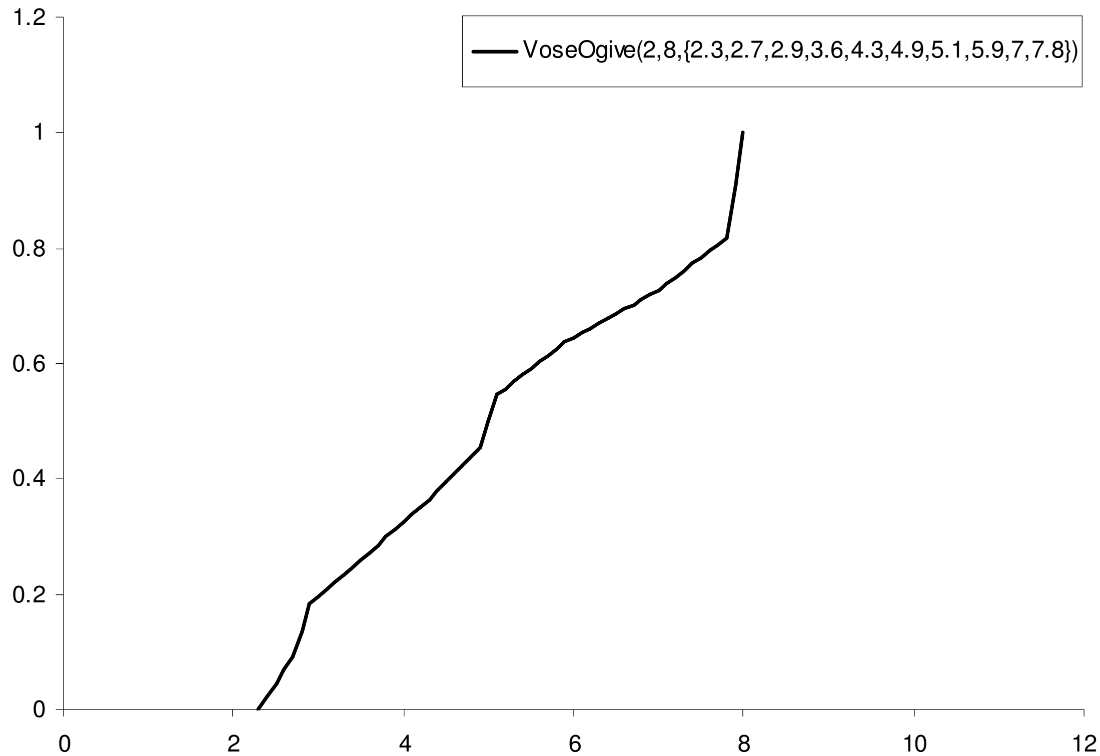
Equations

Probability density function :	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
Cumulative distribution function :	No closed form
Parameter restriction :	$\sigma > 0$
Domain :	$-\infty < x < +\infty$
Mean :	μ
Mode :	μ
Variance :	σ^2
Skewness :	0
Kurtosis :	3

Ogive

VoseOgive(min, max, {data})

Graphs



Cumulative graph of the Ogive distribution

Uses

The Ogive distribution is used to convert a set of data values into an empirical distribution (see section 10.2.1).

Equations

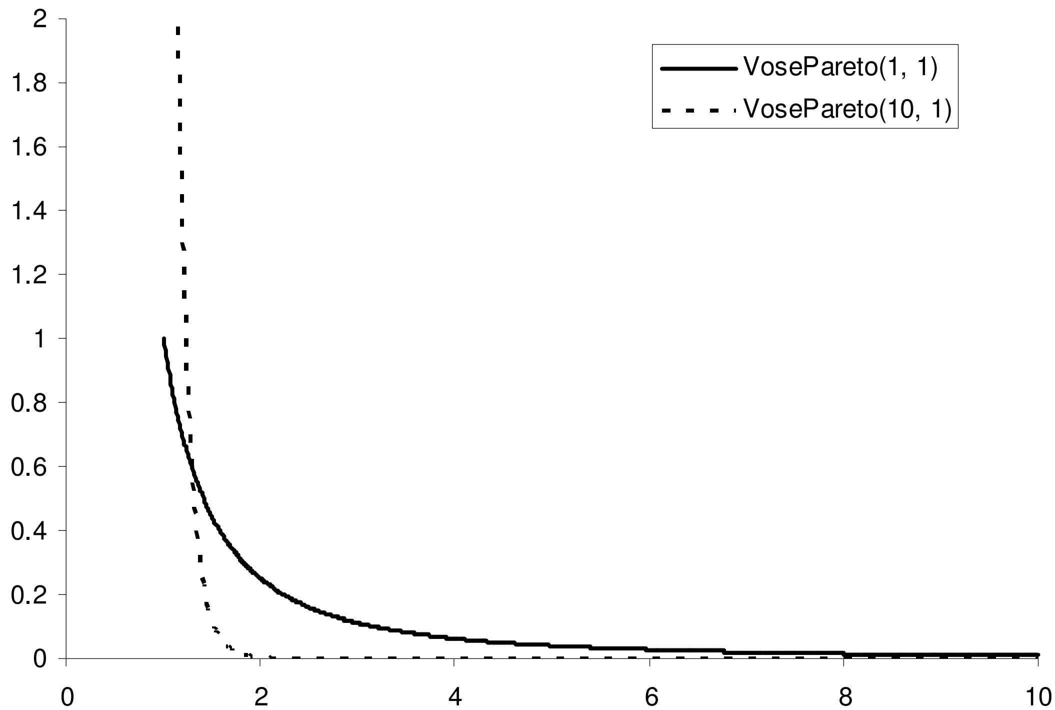
Probability density function :	$f(x)=\frac{1}{n+1}\cdot\frac{1}{x_{i+1}-x_i}$ for $\begin{matrix} x_i \le x < x_{i+1} \\ i \in \{0,1,\ldots,n+1\}\end{matrix}$ where : $x_0 = \min, x_{n+1} = \max, P_0 = 0, P_{n+1} = 1$
Cumulative distribution function :	$F(x_i)=\frac{i}{n+1}$
Parameter restriction :	$x_i < x_{i+1}, n \ge 0$
Domain :	$\min < x < \max$
Mean :	$\frac{1}{n+1}\sum_{i=0}^{n}\frac{x_{i+1}+x_i}{2}$
Mode :	No unique mode
Variance :	Complicated
Skewness :	Complicated
Kurtosis :	Complicated

Pareto

VosePareto(θ,a)

Graphs

The Pareto distribution has an exponential type of shape: right skewed where mode and minimum are equal. It starts at a, and has a rate of decrease determined by θ: The larger θ, the quicker its tail falls away. Examples of the Pareto distribution are given below:



Uses

1. Demographics
The Pareto distribution was originally used to model the number of people with an income of at least x, but it is now used to model any variable that has a minimum, but also most likely, value and for which the probability density decreases geometrically towards zero.

The Pareto distribution has also been used for city population sizes, occurrences of natural resources, stock price fluctuations, size of companies, personal income and error clustering in communication circuits.

An obvious use of the Pareto is for insurance claims. Insurance policies are written so that it is not worth claiming below a certain value (a) and the probability of a claim greater than a is assumed to decrease as a power function of the claim size. It turns out, however, that the Pareto distribution is generally a poor fit.

2. Long-tailed variable

The Pareto distribution has the longest tail of all probability distributions. Thus, while it is not a good fit for the bulk of a variable like a claim size distribution, it is frequently used to model the tails by splicing with another distribution like a Lognormal. That way an insurance company is reasonably guaranteed to have a fairly conservative interpretation of what the (obviously rarely seen, but potentially catastrophic) very high claim values might be. It can also be used to model a longer-tailed discrete variable than any other distribution.

Equations

Probability density function :	$f(x)=\dfrac{\theta a^{\theta}}{x^{\theta+1}}$
Cumulative distribution function :	$F(x)=1-\left(\dfrac{a}{x}\right)^{\theta}$
Parameter restriction :	$\theta>0, a>0$
Domain :	$a \le x$
Mean :	$\dfrac{\theta a}{\theta-1}$ for $\theta>1$
Mode :	a
Variance :	$\dfrac{\theta a^2}{(\theta-1)^2(\theta-2)}$ for $\theta>2$
Skewness :	$2\dfrac{\theta+1}{\theta-3}\sqrt{\dfrac{\theta-2}{\theta}}$ for $\theta>3$
Kurtosis :	$\dfrac{3(\theta-2)(3\theta^2+\theta+2)}{\theta(\theta-3)(\theta-4)}$ for $\theta>4$

Pareto2

VosePareto2(b,q)

Graphs

This distribution is simply a standard Pareto distribution but shifted along the x-axis so that it starts at x = 0. This is most readily apparent by studying the cumulative distribution functions for the two distributions:

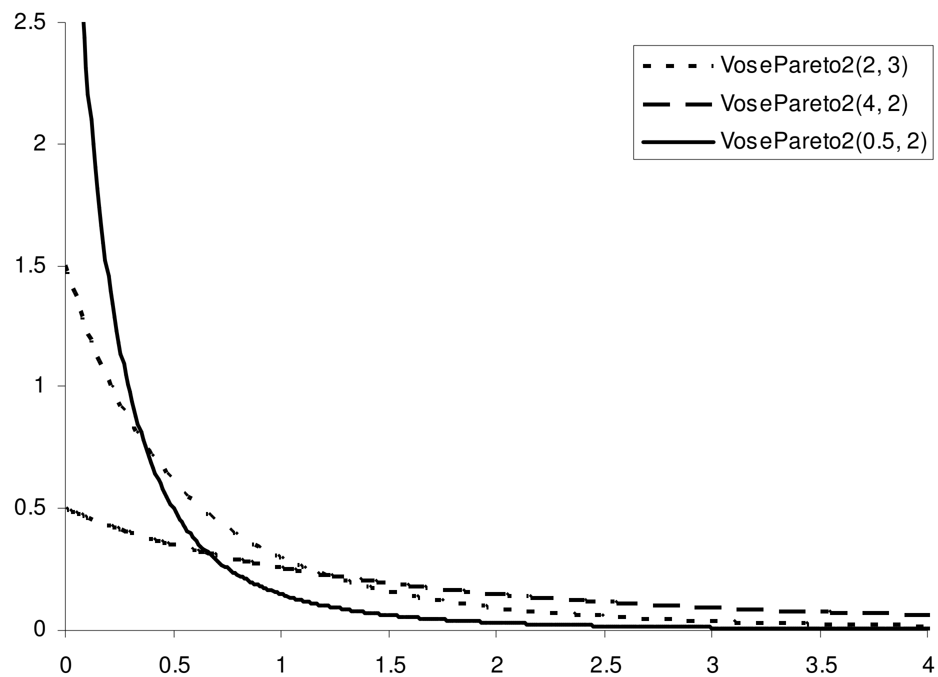
Pareto: $$F(x) = 1 - \left(\frac{a}{x}\right)^{\theta}$$

Pareto2: $$F(x) = 1 - \left(\frac{b}{x+b}\right)^{q}$$

The only difference between the two equations is that x for the Pareto has been replaced by (x+b) for the Pareto2. In other words, using the notation above:

Pareto2(b,q) = Pareto(θ,a) – a
where a = b, and q = θ

Thus both distributions have the same variance and shape when a = b and q = θ, but different means.



Uses

See Pareto distribution.

Equations

Probability density function :	$f(x)=\frac{qb^q}{(x+b)^{q+1}}$
Cumulative distribution function :	$F(x)=1-\frac{b^q}{(x+b)^q}$
Parameter restriction :	$b>0, q>0$
Domain :	$0\le x\le +\infty$
Mean :	$\frac{b}{q-1}$ for $q>1$
Mode :	0
Variance :	$\frac{b^2 q}{(q-1)^2(q-2)}$ for $q>2$
Skewness :	$2\frac{q+1}{q-3}\sqrt{\frac{q-2}{q}}$ for $q>3$
Kurtosis :	$\frac{3(q-2)(3q^2+q+2)}{q(q-3)(q-4)}$ for $q>4$

Pearson Type 5

VosePearson5(α,β)

Graphs

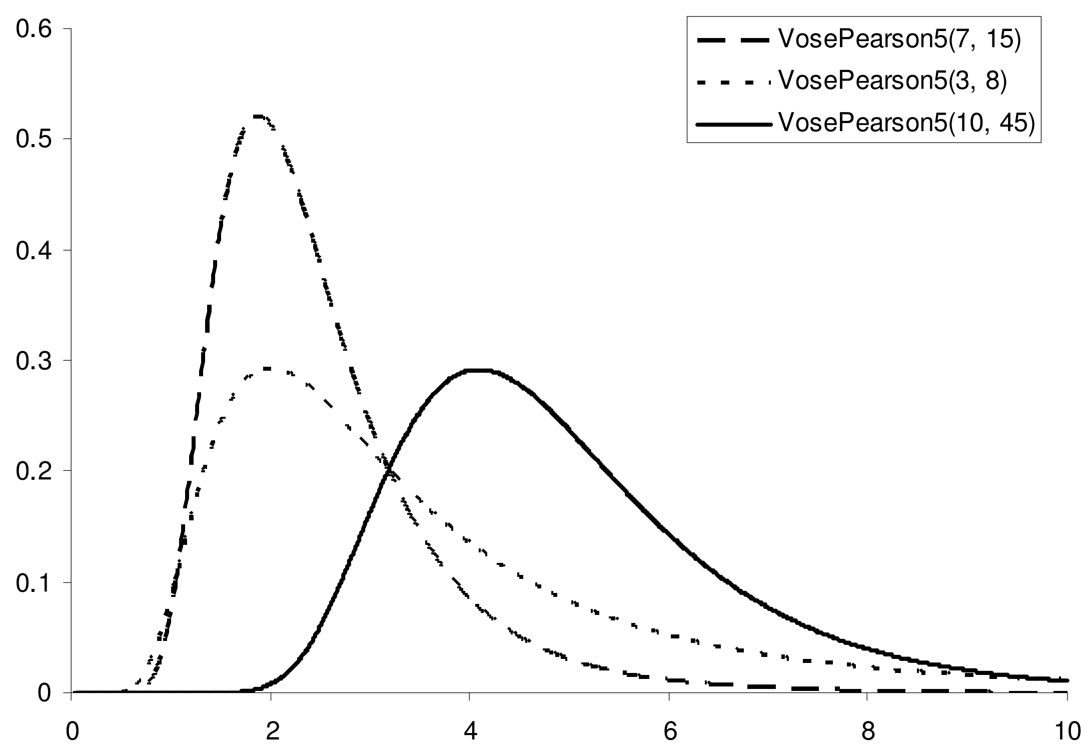
The Pearson family of distributions was designed by Pearson between 1890 and 1895. It represents a system whereby for every member the probability density function f(x) satisfies a differential equation:

$$\frac{1}{p}\frac{dp}{dx} = -\frac{a+x}{c_0 + c_1 x + c_2 x^2} \quad (1)$$

where the shape of the distribution is depends on the values of the parameters a, c0, c1, and c2. The Pearson Type V corresponds to the case where c0 + c1x + c2x2 is a perfect square (c2=4c0c2). Thus, equation (1) can be rewritten as:

$$\frac{d \log f(x)}{dx} = -\frac{a+x}{c_2 (x+c_1)^2}$$

Examples of the Pearson Type 5 distribution are given below:



Uses

This distribution is very rarely properly used in risk analysis.

Equations

Probability density function :	$f(x)=\frac{1}{\beta\Gamma(\alpha)}\frac{e^{-\beta/x}}{(x/\beta)^{\alpha+1}}$
Cumulative distribution function :	No closed form
Parameter restriction :	$\alpha>0,\beta>0$
Domain :	$0\le x<+\infty$
Mean :	$\frac{\beta}{\alpha-1}$ for $\alpha>1$
Mode :	$\frac{\beta}{\alpha+1}$
Variance :	$\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$ for $\alpha>2$
Skewness :	$\frac{4\sqrt{\alpha-2}}{\alpha-3}$ for $\alpha>3$
Kurtosis :	$\frac{3(\alpha+5)(\alpha-2)}{(\alpha-3)(\alpha-4)}$ for $\alpha>4$

Pearson Type 6

VosePearson6(α1,α2,β)

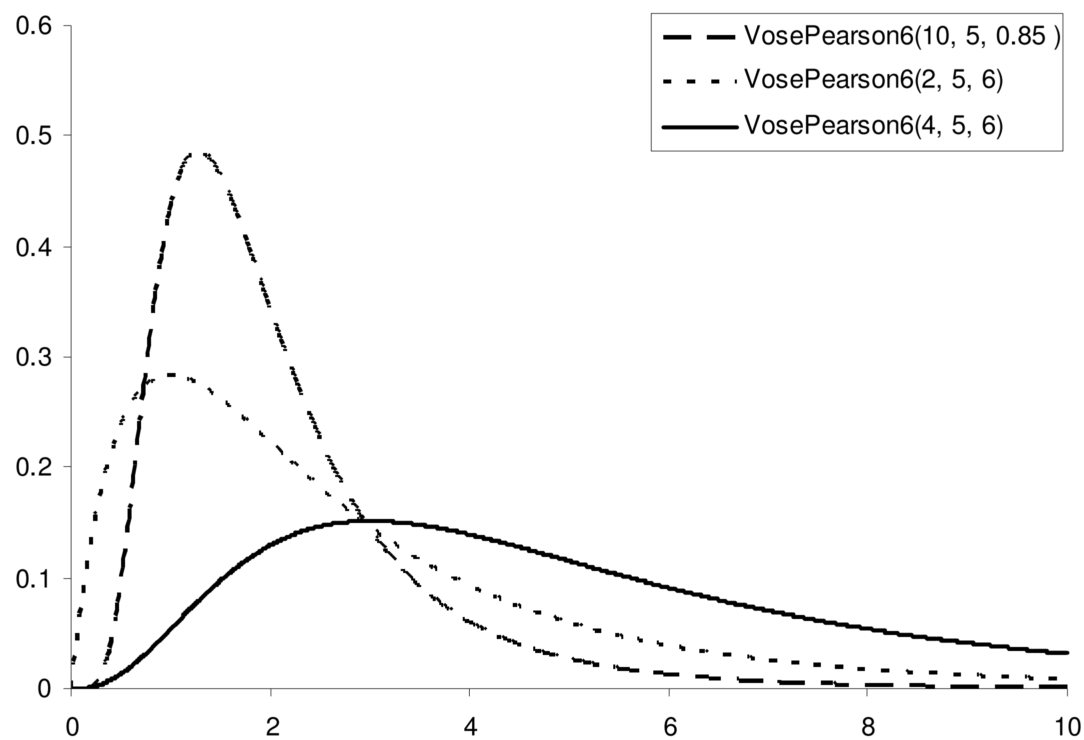
Graphs

The Pearson Type 6 distribution corresponds in the Pearson system to the case when the roots of c0 + c1x + c2x2 = 0 are real and of the same sign. If they are both negative, then:

$$f(x) = K(x - a_1)^{m_1}(x - a_2)^{m_2}$$

Since the expected value is greater than α_2, it is clear that the range of variation of x must be $x>\alpha_2$.

Examples of the Pearson Type 6 distribution are given below:



Uses

At Vose Consulting we don't find much use for this distribution (other than to generate an F distribution or to match the four moments of a data set). The distribution is very unlikely to reflect any of the processes that the analyst may come across, but it's three parameters – plus a fourth if you shift it - (giving it flexibility), sharp peak and long tail make it a possibly candidate to be fitted to a very large (so you know the pattern is real) data set that other distributions won't fit well to.

Like the Pearson Type 5 distribution, the Pearson Type 6 distribution hasn't proven to be very useful in risk analysis.

Equations

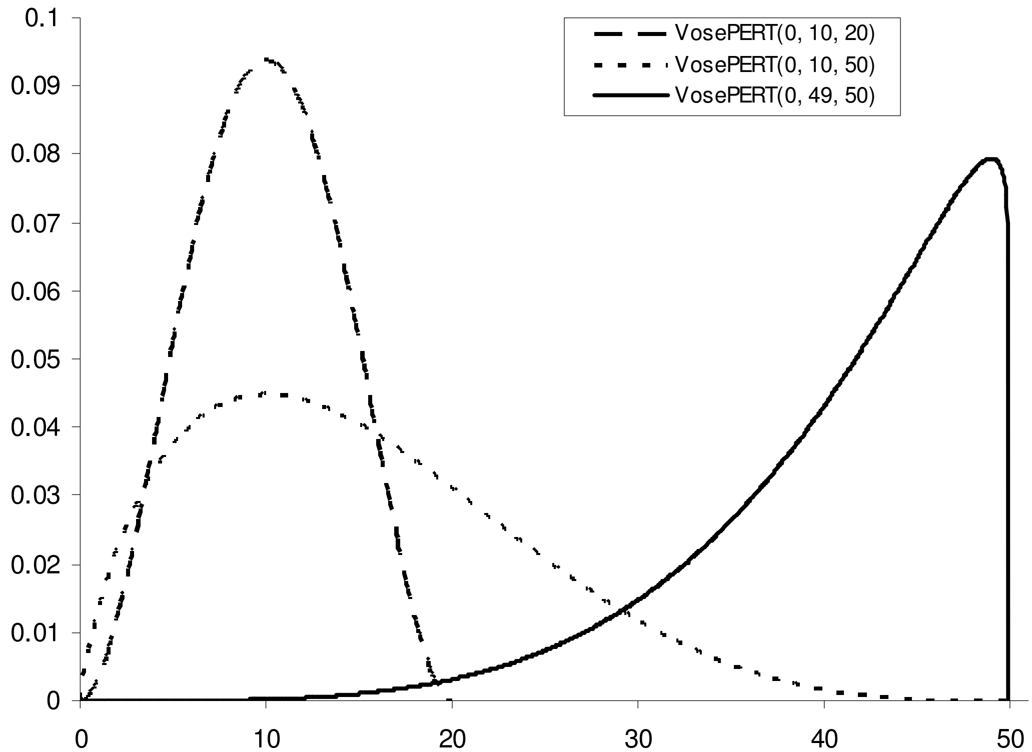
Probability density function :	$f(x)=\frac{1}{\beta B(\alpha_1,\alpha_2)}\frac{(x/\beta)^{\alpha_1-1}}{\left(1+\frac{x}{\beta}\right)^{\alpha_1+\alpha_2}}$ where $B(\alpha_1,\alpha_2)$ is a Beta function
Cumulative distribution function :	No closed form
Parameter restriction :	$\alpha_1>0,\alpha_2>0,\beta>0$
Domain :	$0\le x<+\infty$
Mean :	$\frac{\beta\alpha_1}{\alpha_2-1}$ for $\alpha_2>1$
Mode :	$\frac{\beta(\alpha_1-1)}{\alpha_2+1}$ for $\alpha_1>1$ 0 otherwise
Variance :	$\frac{\beta^2\alpha_1(\alpha_1+\alpha_2-1)}{(\alpha_2-1)^2(\alpha_2-2)}$ for $\alpha_2>2$
Skewness :	$2\sqrt{\frac{\alpha_2-2}{\alpha_1(\alpha_1+\alpha_2-1)}}\left[\frac{2\alpha_1+\alpha_2-1}{\alpha_2-3}\right]$ for $\alpha_2>3$
Kurtosis :	$\frac{3(\alpha_2-2)}{(\alpha_2-3)(\alpha_2-4)}\left[\frac{2(\alpha_2-1)^2}{\alpha_1(\alpha_1+\alpha_2-1)}+(\alpha_2+5)\right]$ for $\alpha_2>4$

PERT

VosePERT(min,mode,max)

Graphs

The PERT (aka BetaPERT) distribution gets its name because it uses the same assumption about the mean (see below) as PERT networks (used in the past for project planning). It is a version of the Beta distribution and requires the same three parameters as the Triangular distribution, namely minimum (a), most likely (b) and maximum (c). The figure below shows three PERT distributions whose shape can be compared to the Triangular distributions:



Uses

The PERT distribution is used exclusively for modeling expert estimates, where one is given the expert's minimum, most likely and maximum guesses. It is a direct alternative to a Triangular distribution.

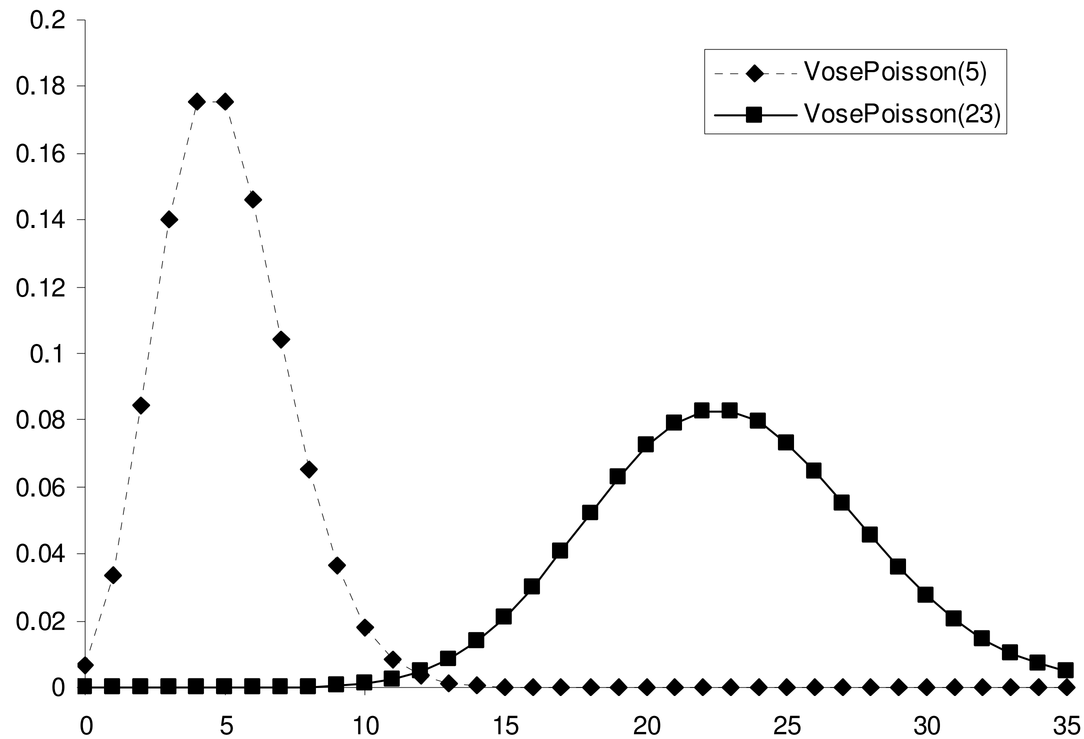
Equations

Probability density function :	$f(x)=\dfrac{(x-\min)^{\alpha_1-1}(\max-x)^{\alpha_2-1}}{B(\alpha_1,\alpha_2)(\max-\min)^{\alpha_1+\alpha_2-1}}$ where $\alpha_1=6\left[\dfrac{\mu-\min}{\max-\min}\right]$, $\alpha_2=6\left[\dfrac{\max-\mu}{\max-\min}\right]$ with $\mu(=\text{mean})=\dfrac{\min+4\,\text{mode}+\max}{6}$ and $B(\alpha_1,\alpha_2)$ is a Beta function
Cumulative distribution function :	$F(x)=\dfrac{B_z(\alpha_1,\alpha_2)}{B(\alpha_1,\alpha_2)}\equiv I_z(\alpha_1,\alpha_2)$ where $z=\dfrac{x-\min}{\max-\min}$ and $B_z(\alpha_1,\alpha_2)$ is an incomplete Beta function
Parameter restriction :	$\min<\text{mode}<\max$
Domain :	$\min\le x\le\max$
Mean :	$\dfrac{\min+4\,\text{mode}+\max}{6}\equiv\mu$
Mode :	mode
Variance :	$\dfrac{(\mu-\min)(\max-\mu)}{7}$
Skewness :	$\dfrac{\min+\max-2\mu}{4}\sqrt{\dfrac{7}{(\mu-\min)(\max-\mu)}}$
Kurtosis :	$3\dfrac{(\alpha_1+\alpha_2+1)\left(2(\alpha_1+\alpha_2)^2+\alpha_1\alpha_2(\alpha_1+\alpha_2-6)\right)}{\alpha_1\alpha_2(\alpha_1+\alpha_2+2)(\alpha_1+\alpha_2+3)}$

Poisson

VosePoisson(λ)

Graphs



Uses

The Poisson(λ) distribution models the number of occurrences of an event in a given time with an expected rate of λ events when the time between successive events follows a Poisson process.

Example

If β is the mean time between events, as used by the Exponential distribution, then λ = 1/β. For example, imagine that records show that a computer crashes on average once every 250 hours of operation (β=250 hours), then the rate of crashing λ is 1/250 crashes per hour. Thus a Poisson (1000/250) = Poisson(4) distribution models the number of crashes that could occur in the next 1000 hours of operation.

Equations

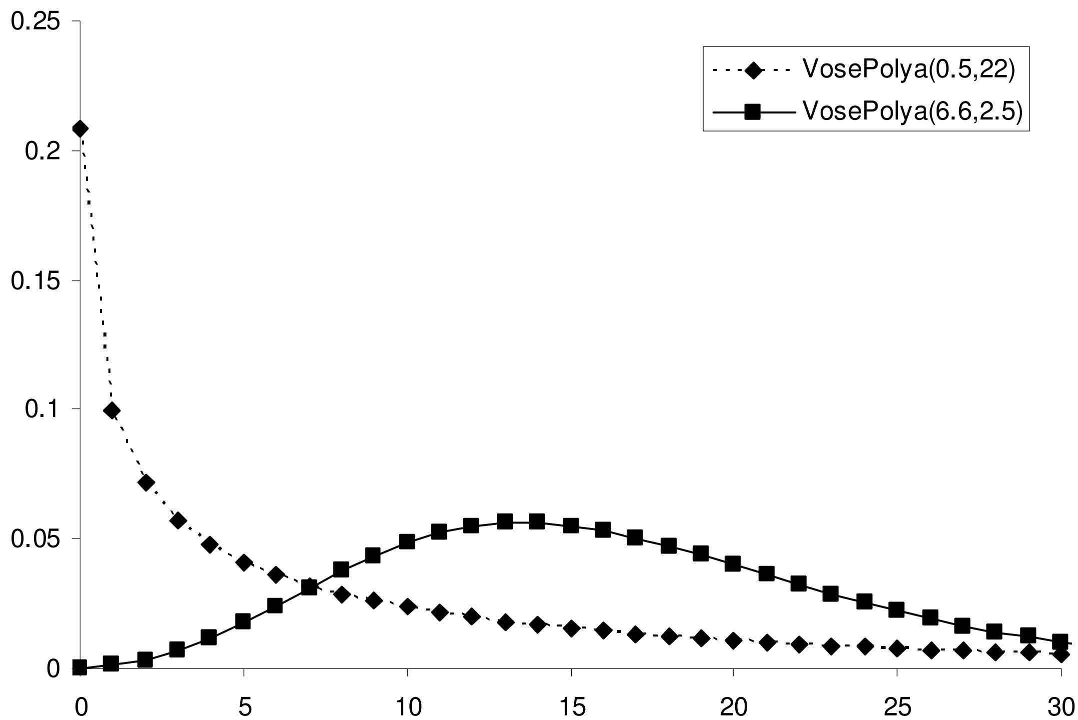
(using λ = expected rate per unit time, t = time over which counts are to be estimated)

Probability mass function :	$f(x)=\dfrac{e^{-\lambda t}(\lambda t)^x}{x!}$
Cumulative distribution function :	$F(x)=e^{-\lambda t}\sum_{i=0}^{\lfloor x \rfloor}\dfrac{(\lambda t)^i}{i!}$
Parameter restriction :	$\lambda t>0$
Domain :	$x=\{0,1,2,\ldots\}$
Mean :	λt
Mode :	$\lambda t, \lambda t-1$ if λt is an integer $\lvert\lambda t\rvert$ otherwise
Variance :	λt
Skewness :	$\dfrac{1}{\sqrt{\lambda t}}$
Kurtosis :	$3+\dfrac{1}{\lambda t}$

Pólya

VosePolya(α,β)

Graphs



Uses

There are several types of distribution in the literature that have been given the Pólya name. We employ the name for a distribution that is very common in the insurance field. A standard initial assumption of the frequency distribution of the number of claims is Poisson:

Claims = Poisson(λ)

where λ is the expected number of claims during the period of interest. The Poisson distribution has a mean and variance equal to λ and one often sees historic claim frequencies with a variance greater than the mean so that the Poisson model underestimates the level of randomness of claim numbers. A standard method to incorporate greater variance is to assume that λ is itself a random variable (and the claim frequency distribution is called a mixed Poisson model). A Gamma(α, β) distribution is most commonly used to describe the random variation of λ between periods, so:

Claims = Poisson(Gamma(α, β)) (1)

This is the Pólya(α, β) distribution.

Comments

If α is an integer, we have:

Claims = Poisson(Gamma(α, β)) = NegBin(α, 1/(1+β)) (2)

so one can say that the Negative Binomial distribution is a special case of the Pólya.

Equations

Probability mass function :	$f(x)=\dfrac{\Gamma(\alpha+x)\beta^x}{\Gamma(x+1)\Gamma(\alpha)(1+\beta)^{\alpha+x}}$
Cumulative distribution function :	$F(x)=\sum_{i=0}^{x}\dfrac{\Gamma(\alpha+i)\beta^i}{\Gamma(i+1)\Gamma(\alpha)(1+\beta)^{\alpha+i}}$
Parameter restriction :	$\alpha>0,\beta>0$
Domain :	$x=\{0,1,2,\ldots\}$
Mean :	$\alpha\beta$
Mode :	0 if $\alpha\le 1$ z, z+1 if z is an integer $\lceil z\rceil$ if z is not an integer where $z=\beta(\alpha-1)-1$
Variance :	$\alpha\beta(1+\beta)$
Skewness :	$\dfrac{1+2\beta}{\sqrt{(1+\beta)\alpha\beta}}$
Kurtosis :	$3+\dfrac{6}{\alpha}+\dfrac{1}{\alpha\beta(1+\beta)}$

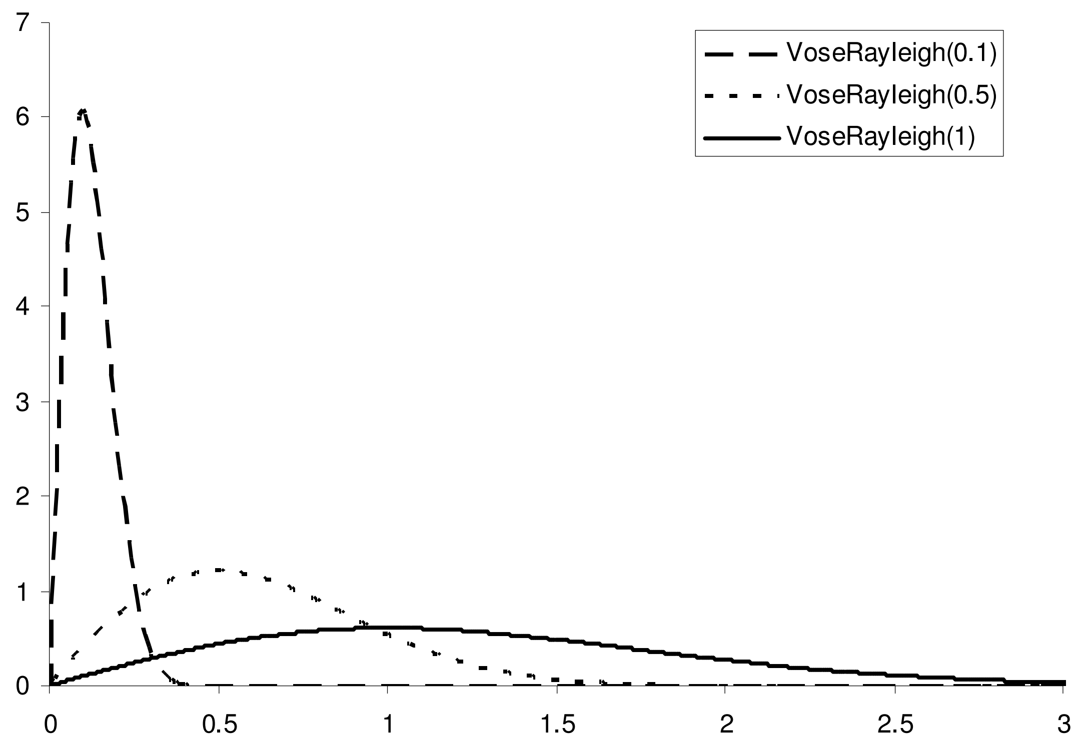
Rayleigh

VoseRayleigh(b)

Graphs

Originally derived by Lord Rayleigh (or by his less glamorous name J.W. Strutt) in the field of acoustics.

The graph below shows various Rayleigh distributions. The distribution in black is a Rayleigh(1), sometimes referred to as the standard Rayleigh distribution.



Uses

The Rayleigh distribution is frequently used to model wave heights in oceanography, and in communication theory to describe hourly median and instantaneous peak power of received radio signals.

The distance from one individual to its nearest neighbor when the spatial pattern is generated by a Poisson distribution follows a Rayleigh distribution.

The Rayleigh distribution is a special case of the Weibull distribution since Rayleigh(b) = Weibull(2, b√2), and as such is a suitable distribution for modeling the lifetime of a device that has a linearly increasing instantaneous failure rate: $z(x) = x/b^2$.

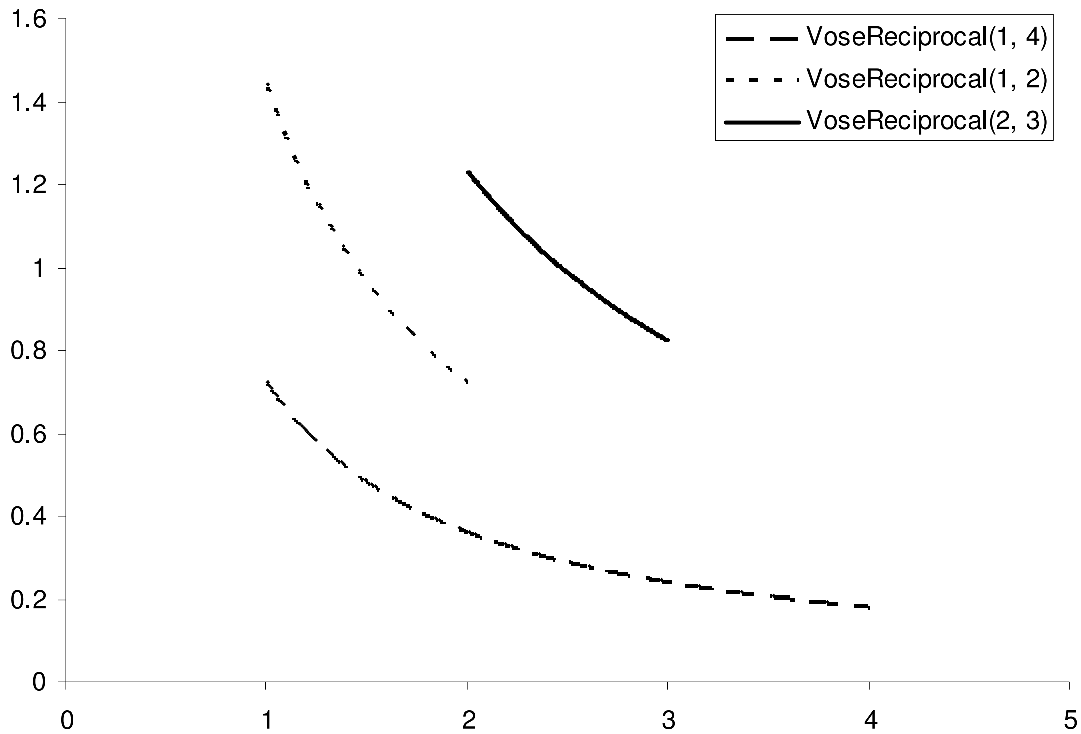
Equations

Probability density function :	$f(x)=\frac{x}{b^2}\exp\left[-\frac{1}{2}\left(\frac{x}{b}\right)^2\right]$
Cumulative distribution function :	$F(x)=1-\exp\left[-\frac{1}{2}\left(\frac{x}{b}\right)^2\right]$
Parameter restriction :	$b>0$
Domain :	$0\le x<+\infty$
Mean :	$b\sqrt{\frac{\pi}{2}}$
Mode :	b
Variance :	$b^2\left(2-\frac{\pi}{2}\right)$
Skewness :	$\frac{2(\pi-3)\sqrt{\pi}}{(4-\pi)^{3/2}}\approx 0.6311$
Kurtosis :	$\frac{32-3\pi^2}{(4-\pi)^2}\approx 3.2451$

Reciprocal

VoseReciprocal(min,max)

Graphs



Uses

The Reciprocal distribution is widely used as an uninformed prior distribution in Bayesian inference for scale parameters.

It is also used to describe '1/f noise'. One way to characterize different noise sources is to consider the spectral density, i.e. the mean squared fluctuation at any particular frequency f and how that varies with frequency. A common approach is to model spectral densities that vary as powers of inverse frequency: the power spectra P(f) is proportional to $f^{-\beta}$ for $\beta \geq 0$. ⍰ = 0 equates to white noise (i.e. no relationship between P(f) and f), ⍰ = 2 is called Brownian noise, and $\beta = 1$ takes the name '1/f noise' which occurs very often in nature.

Equations

Probability density function :	$f(x)=\frac{1}{xq}$ where $q=\log(\max)-\log(\min)$
Cumulative distribution function :	$F(x)=\frac{\log(x)-\log(\min)}{q}$
Parameter restriction :	$0<\min<\max$
Domain :	$\min\le x\le\max$
Mean :	$\frac{\max-\min}{q}$
Mode :	$\min$
Variance :	$\frac{(\max-\min)[\max(q-2)+\min(q+2)]}{2q^2}$
Skewness :	$\frac{\sqrt{2}\left[12q(\max-\min)^2+q^2\left(\max^2(2q-9)+2\min\max q+\min^2(2q+9)\right)\right]}{3q\sqrt{\max-\min}\,(\max(q-2)+\min(q+2))^{3/2}}$

Kurtosis :

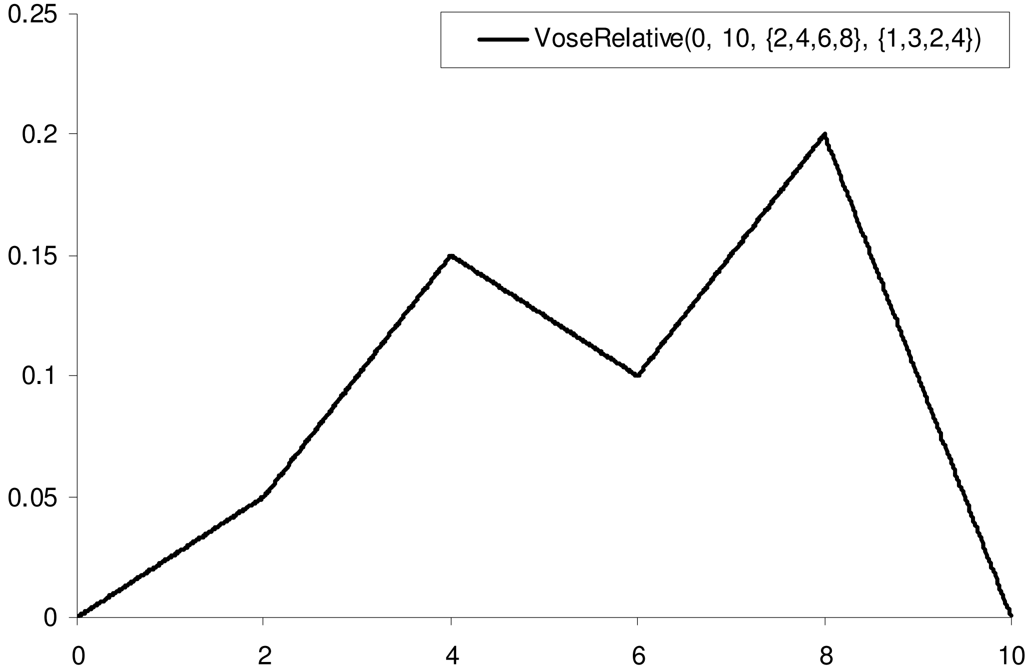
$$\frac{36q(\max-\min)^2(\max+\min)-36(\max-\min)^3-16q^2\left(\max^3-\min^3\right)+3q^3\left(\max^2+\min^2\right)(\max+\min)}{3(\max-\min)(\max(q-2)+\min(q+2))^2}$$

Relative

VoseRelative(min,max,{xi},{pi})

Graphs

The Relative distribution is a non-parametric distribution (i.e. there is no underlying probability model) where {xi} is an array of x-values with probability densities {pi} and where the distribution falls between the minimum and maximum. An example of the Relative distribution is given below:



Uses

1. Modelling expert opinion

The Relative distribution is very useful for producing a fairly detailed distribution that reflects an expert's opinion. The Relative distribution is the most flexible of all of the continuous distribution functions. It enables the analyst and expert to tailor the shape of the distribution to reflect, as closely as possible, the opinion of the expert.

2. Modelling posterior distribution in Bayesian inference

If you use the construction method of obtaining a Bayesian posterior distribution, you will have two arrays: a set of possible value in ascending order; and a set of posterior weights for each of those values.

This exactly matches the input parameters for a Relative distribution which can then be used to generate values from the posterior distribution. Examples: Turbine blades; Fitting a Weibull distribution.

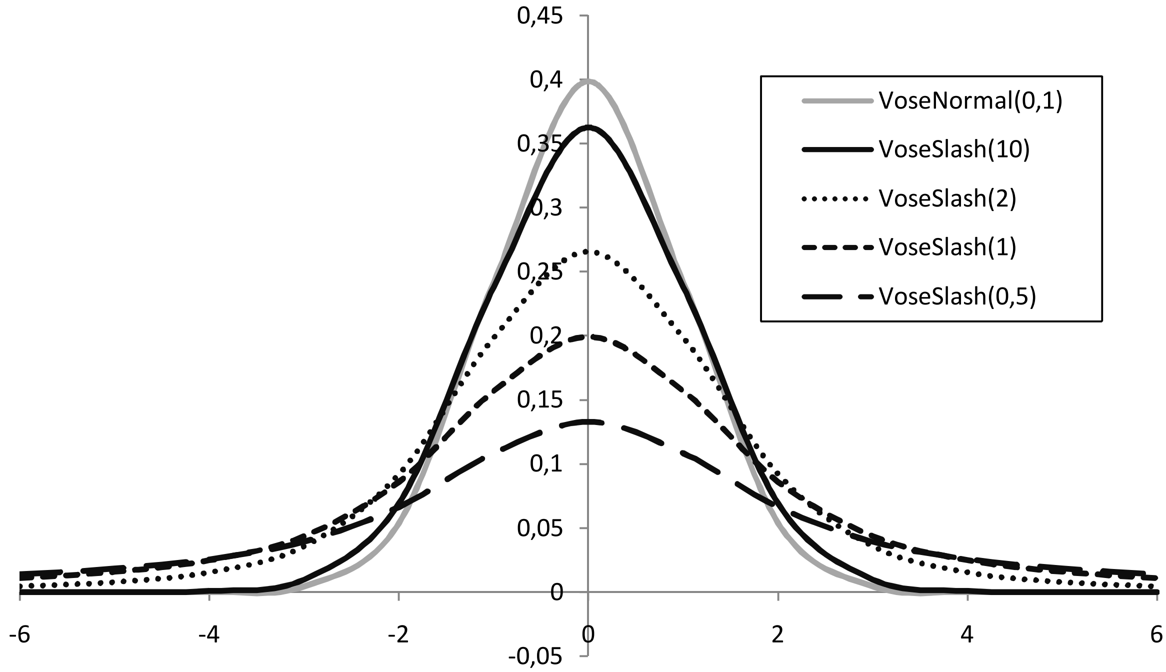
Equations

Probability density function :	$f(x) = \frac{x - x_i}{x_{i+1} - x_i}(p_{i+1} - p_i) + p_i$ if $x_i \le x < x_{i+1}$
Cumulative distribution function :	$F(x) = F(x_i) + \frac{x - x_i}{x_{i+1} - x_i} \cdot \frac{p_i + p_{i+1}}{2(x_{i+1} - x_i)}$ if $x_i \le x < x_{i+1}$
Parameter restriction :	$p_i \ge 0, x_i < x_{i+1}, n > 0, \sum_{i=1}^{n} p_i > 0$
Domain :	$\min \le x \le \max$
Mean :	No closed form
Mode :	No closed form
Variance :	No closed form
Skewness :	No closed form
Kurtosis :	No closed form

Slash

VoseSlash(q)

Graphs



Uses

The (standard) Slash distribution is defined as the ratio of a standard Normal and a Uniform distribution: Slash(q,0,1)=Normal(0,1)/Uniform(0,1)1/q

The Slash distribution may be used for perturbations on a time series in place of a more standard normal distribution assumption, because it has longer tails. The q parameter controls the extent of the tails. As q approaches infinity, the distribution looks more and more like a Normal.

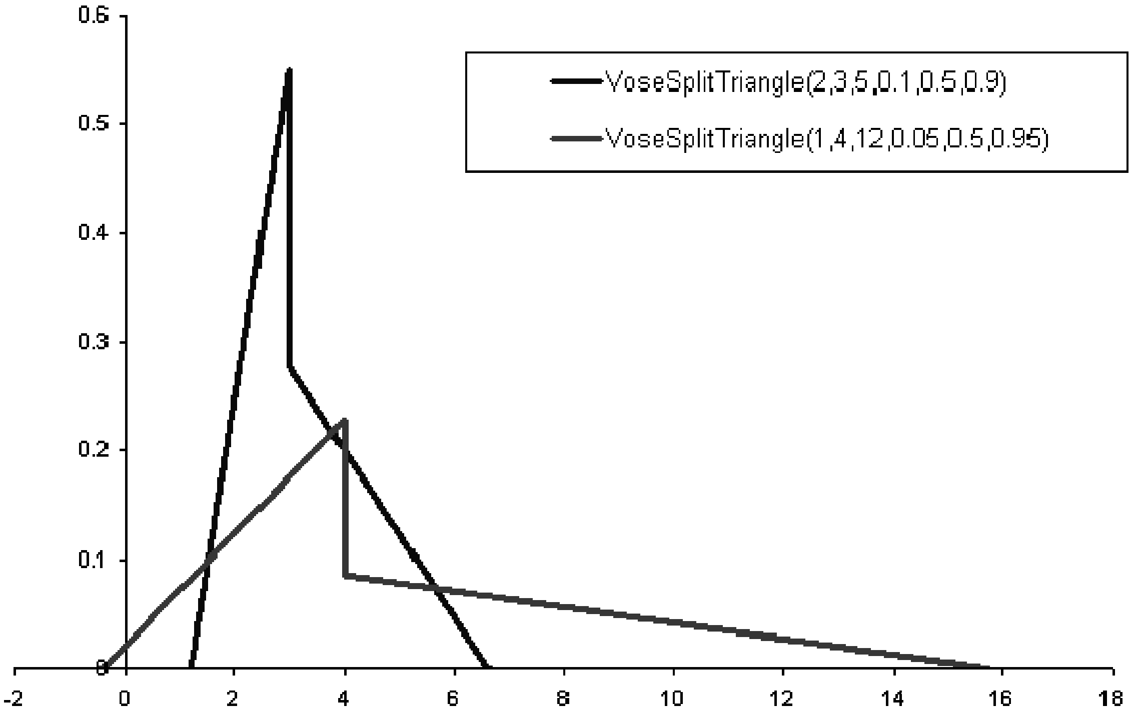
Equations

Probability density function :	$f(x) = q\int_0^1 u^q \phi(ux)du$ Where $\phi(x)$ is the standard Normal pdf
Cumulative distribution function :	$F(x) = q\int_0^1 u^{q-1}\Phi(xu)du$ Where $\Phi(x)$ is the standard Normal cdf
Parameter restriction :	$q > 0$
Domain :	$-\infty < x < +\infty$
Mean :	0 for $q > 1$
Mode :	0
Variance :	$\frac{q}{q-2}$ for $q > 2$
Skewness :	Complicated
Kurtosis :	Complicated

Split Triangle

VoseSplitTriangle(low,medium,high,lowP,mediumP,highP)

Graphs



Uses

The SplitTriangle is used to model expert opinion where the SME is asked for three points on the distribution: for each point the SME provides a value of the variable and their view of the probability of being less than that value. The SplitTriangle then extrapolates from these values to create a distribution composed of 2 triangles as in the figure above.

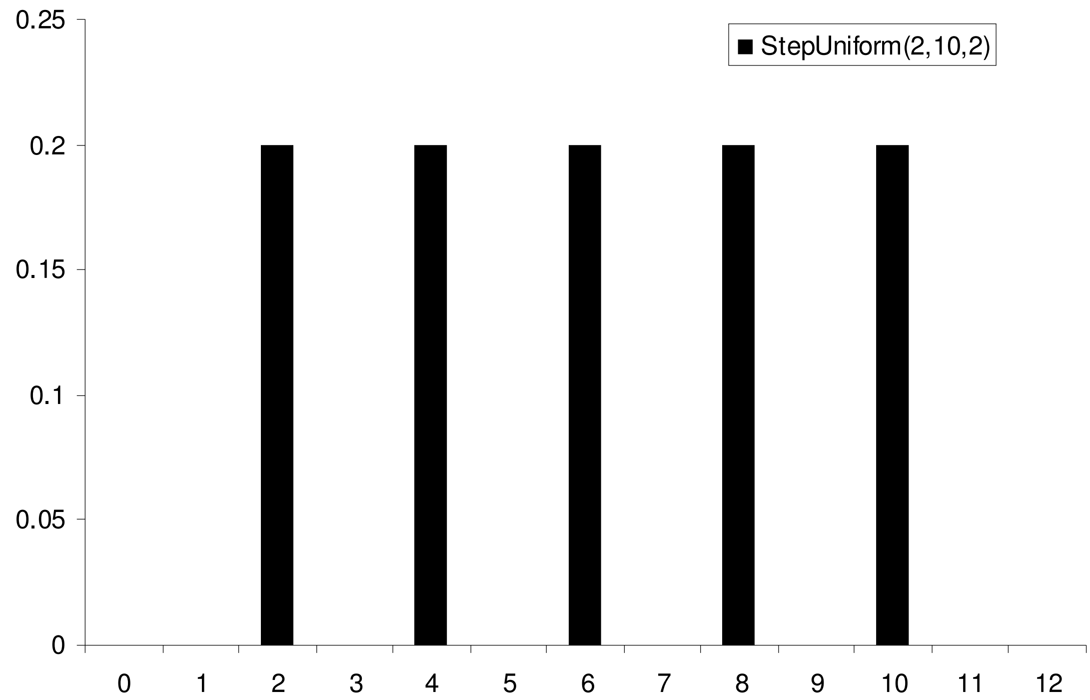
Equations

Probability density function :	$f(x)=\frac{Height1(x-\min)}{(\text{mode}-\min)}$ if $\min \le x \le \text{mode}$ $f(x)=\frac{Height2(\max-x)}{(\max-\text{mode})}$ if $\text{mode} < x \le \max$ Where : $Height1=\frac{2*mediumP}{\text{mode}-\min}$ $Height2=\frac{2*(1-mediumP)}{\max-\text{mode}}$ and: $\min=\frac{(low-\text{mode}*\sqrt{\text{lowP/mediumP}})}{(1-\sqrt{\text{lowP/mediumP}})}$ $\text{mode}=\text{medium}$ $\max=\frac{(\text{mode - high}*\sqrt{\text{1 - mediumP/1 - highP}})}{(1-\sqrt{\text{1 - mediumP/1 - highP}})}$
Cumulative distribution function :	$F(x)=0$ if $x<\min$ $F(x)=\frac{Height1(x-\min)^2}{2(\text{mode}-\min)}$ if $\min \le x \le \text{mode}$ $F(x)=1-\frac{Height2(\max-x)^2}{2(\max-\text{mode})}$ if $\text{mode} < x \le \max$ $F(x)=1$ if $\max < x$
Parameter restriction :	$\min \le \text{mode} \le \max, \min < \max$
Domain :	$\min < x < \max$
Mean :	$\frac{mediumP+(1-mediumP)\max+2*\text{mode}}{3}$
Mode :	mode
Variance :	$\frac{\text{mode}^2-2*\text{mode}*\text{mediumP}*\min+2*\text{mode}*\text{mediumP}*\max-2*\text{mode}*\max-2*\text{mediumP}^2*\min^2+\max^2}{18}$ $+\frac{-2*mediumP^2*\max^2+4*mediumP^2*\max*\min+3*mediumP*\min^2+mediumP*\max^2-4*\min*mediumP*\max}{18}$
Skewness :	Complicated
Kurtosis :	Complicated

Step Uniform

VoseStepUniform(min,max,step)

Graph



Uses

The StepUniform distribution returns values between the min and max at the defined step increments. If the step value is omitted the function assumes a default value of 1.

The StepUniform function is generally used as a tool to sample in a controlled way along some dimension (e.g. time, distance, x) and can be used to perform simple one-dimensional numerical integrations.

StepUniform(A,B) where A,B are integers will generate a random integer variable that can be used as an index variable as an efficiently way to randomly select paired data from a database.

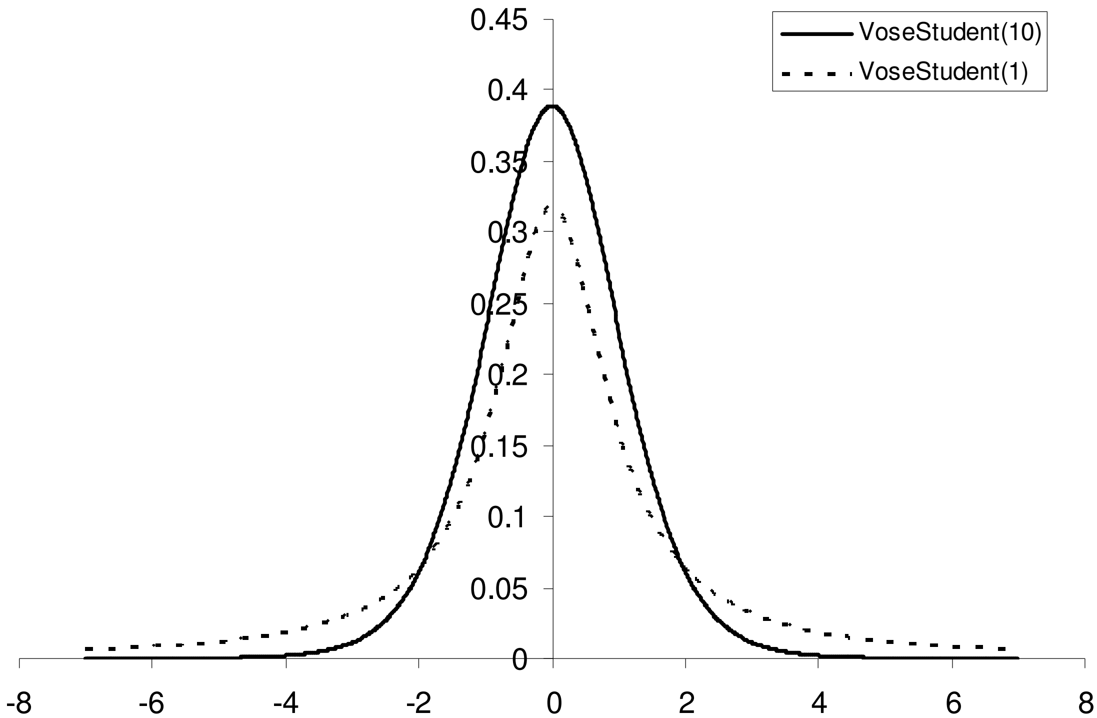
Equations

Probability mass function :	$f(x)=\frac{\text{step}}{\max-\min+\text{step}}$ for $x=\min+i\cdot\text{step}$, $i=0\ to\ \frac{\max-\min}{\text{step}}$ 0 otherwise
Cumulative distribution function :	$F(x)=0$ for $x<\min$ $\left(\left\lfloor\frac{x-\min}{\text{step}}\right\rfloor+1\right)*\frac{\text{step}}{\max-\min+\text{step}}$ for $\min\le x<\max$ for $\max\le x$
Parameter restriction :	$\frac{\max-\min}{\text{step}}$ must be an integer
Domain :	$\min\le x\le\max$
Mean :	$\frac{\min+\max}{2}$
Mode :	Not uniquely defined
Variance :	$\frac{(\max-\min)(\max-\min+2\,\text{step})}{12}$
Skewness :	0
Kurtosis :	$\frac{3}{5}\left[\frac{3(\max-\min)^2+2\,\text{step}(3\max-3\min-2\,\text{step})}{(\max-\min)(\max-\min+2\,\text{step})}\right]$

Student

VoseStudent(ν)

Graphs



Uses

The most common use of the Student distribution is for the estimation of the mean of a(n assumed Normally distributed) population where random samples from that population have been observed, and its standard deviation is unknown. The relationship:

Student(ν) = Normal(0,SQRT(ν/ChiSq(ν)))

is at the centre of the method. This is equivalent to a t-test in classical statistics.

Other sample statistics can be approximated to a Student distribution, and thus a Student distribution can be used to describe one's uncertainty about the parameter's true value: in regression analysis, for example.

Comments

First discovered by the English statistician William Sealy Gossett (1876-1937), whose employer (the brewery company, Guinness) forbade employees from publishing their work, so he wrote a paper under the pseudonym 'Student'. As ν increases, the Student-t distribution tends to a Normal(0, 1) distribution.

Equations

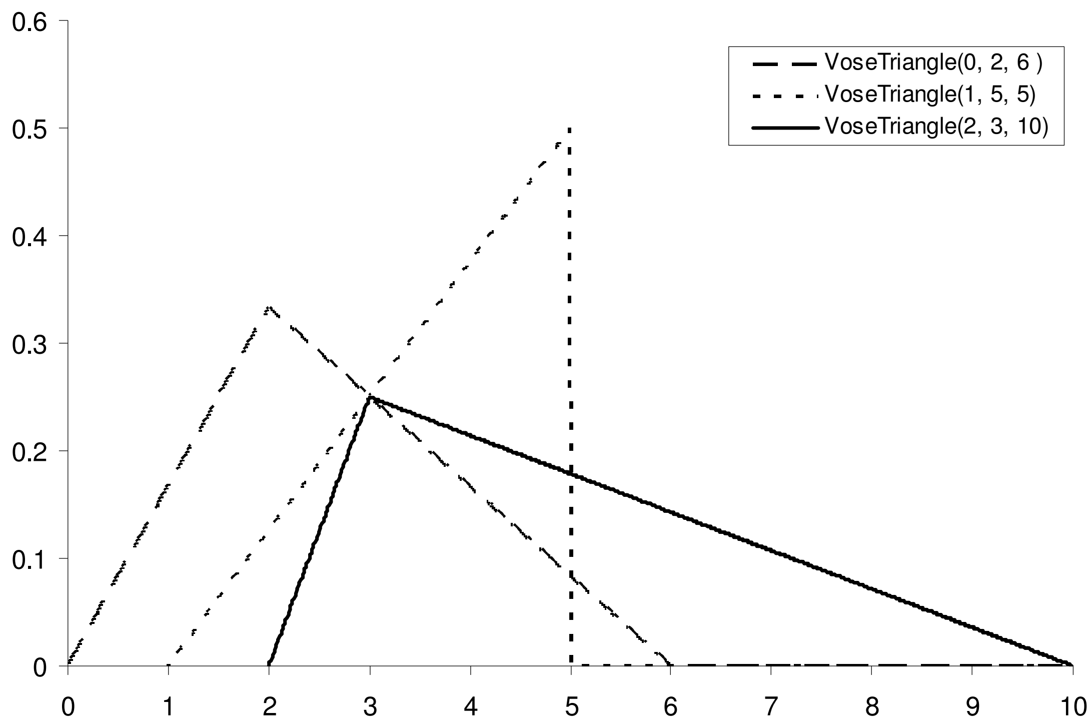
Probability density function :	$$f(x)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)\left[1+\left(\frac{x^2}{\nu}\right)\right]^{\frac{\nu+1}{2}}}$$
Cumulative distribution function :	$$F(x)=\frac{1}{2}+\frac{1}{\pi}\left[\tan^{-1}\left(\frac{x}{\sqrt{\nu}}\right)+\frac{x\sqrt{\nu}}{\nu+x^2}\sum_{j=0}^{\frac{\nu-3}{2}}\frac{a_j}{\left(1+\frac{x^2}{\nu}\right)^j}\right]$$ if ν is odd $$F(x)=\frac{1}{2}+\frac{x}{2\sqrt{\nu+x^2}}\sum_{j=0}^{\frac{\nu-2}{2}}\frac{b_j}{\left(1+\frac{x^2}{\nu}\right)^j}$$ if ν is even where $a_j=\left(\frac{2j}{2j+1}\right)a_{j-1};a_0=1$ and $b_j=\left(\frac{2j-1}{2j}\right)b_{j-1};b_0=1$
Parameter restriction :	ν is a positive integer
Domain :	$-\infty < x < +\infty$
Mean :	0 for $\nu > 1$
Mode :	0
Variance :	$\frac{\nu}{\nu-2}$ for $\nu > 2$
Skewness :	0 for $\nu > 3$
Kurtosis :	$3\left(\frac{\nu-2}{\nu-4}\right)$ for $\nu > 4$

Triangle

VoseTriangle(min,mode,max)

Graphs

The Triangle distribution constructs a triangular shape from the three input parameters. An example of the Triang distribution is given below:



Uses

The Triangle distribution is used as a rough modeling tool where the range (min to max) and the most likely value within the range (mode) can be estimated. It has no theoretical basis but derives its statistical properties from its geometry.

The Triangle distribution offers considerable flexibility in its shape, coupled with the intuitive nature of its defining parameters and speed of use. It has therefore achieved a great deal of popularity among risk analysts. However, min and max are the absolute minimum and maximum estimated values for the variable and it is generally a difficult task to make estimates of these values.

Equations

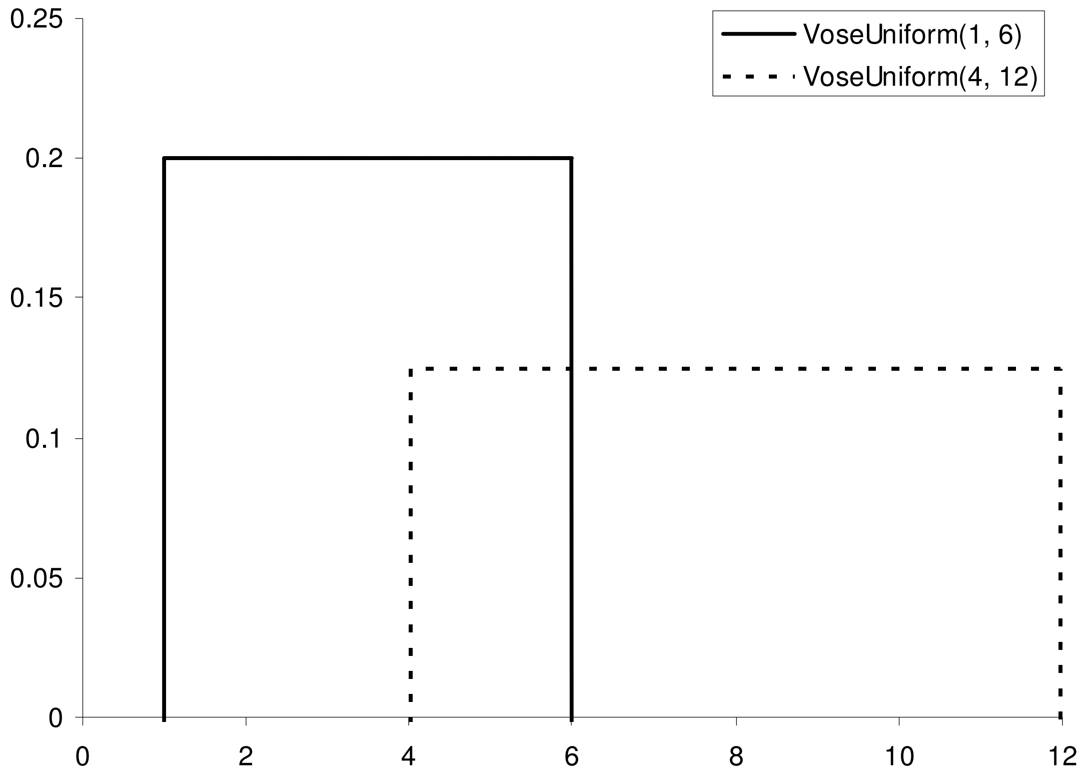
Probability density function :	$f(x)=\frac{2(x-\min)}{(\text{mode}-\min)(\max-\min)}$ if $\min \le x \le \text{mode}$ $f(x)=\frac{2(\max-x)}{(\max-\min)(\max-\text{mode})}$ if $\text{mode} < x \le \max$
Cumulative distribution function :	$F(x)=0$ if $x<\min$ $F(x)=\frac{(x-\min)^2}{(\text{mode}-\min)(\max-\min)}$ if $\min \le x \le \text{mode}$ $F(x)=1-\frac{(\max-x)^2}{(\max-\min)(\max-\text{mode})}$ if $\text{mode} < x \le \max$ $F(x)=1$ if $\max < x$
Parameter restriction :	$\min \le \text{mode} \le \max, \min < \max$
Domain :	$\min < x < \max$
Mean :	$\frac{\min+\text{mode}+\max}{3}$
Mode :	mode
Variance :	$\frac{\min^2+\text{mode}^2+\max^2-\min\text{mode}-\min\max-\text{mode}\max}{18}$
Skewness :	$\frac{2\sqrt{2}}{5}\frac{z(z^2-9)}{(z^2+3)^{3/2}}$ where $z=\frac{2(\text{mode}-\min)}{\max-\min}-1$
Kurtosis :	2.4

Uniform

VoseUniform(min,max)

Graphs

A Uniform distribution assigns equal probability to all values between its minimum and maximum. Examples of the Uniform distribution are given below:



Uses

1. Rough estimate
The Uniform distribution is used as a very approximate model where there are very little or no available data. It is rarely a good reflection of the perceived uncertainty of a parameter since all values within the allowed range have the same constant probability density, but that density abruptly changes to zero at the minimum and maximum. However, it is sometimes useful for bringing attention to the fact that a parameter is very poorly known.

2. Crude sensitivity analysis
Sometimes we want to get a rough feel for whether it is important to assign uncertainty to a parameter. You could give the parameter a Uniform distribution with reasonably wide bounds, run a crude sensitivity analysis, and see whether the parameter registered as having influence on the output uncertainty: if not,

it may as well be left crudely estimated. The Uniform distribution assigns the most (reasonable) uncertainty to the parameter, so if the output is insensitive to the parameter with a Uniform, it will be even more insensitive for another distribution.

3. Rare Uniform variable

There are some special circumstances where a Uniform distribution may be appropriate, for example a Uniform(0, 360) distribution for the angular resting position of a camshaft after spinning; or a Uniform(0, L/2) for the distance from a random leak in a pipeline of segments of length L to its nearest segment end (where you'd break the pipeline to get access inside).

4. Plotting a function

Sometimes you might have a complicated function you wish to plot for different values of an input parameter, or parameters. For a one parameter function (like y=GAMMALN(ABS(SIN(x)/((x-1)^0.2+COS(LN(3*x))))) for example), you can make two arrays: the first with the x-values (say between 1 and 1000), the second the correspondingly calculated y-values. Alternatively, you could write one cell for x: =Uniform(1,1000) and another for y using the generated x-value, name both as outputs, run a simulation, and export the generated values into a spreadsheet. Perhaps not worth the effort for one parameter, but when you have two or three it is. Graphic software like S-PLUS will draw surface contours for {x,y,z} data arrays.

5. Uninformed prior

A Uniform distribution is often used as an uninformed prior in Bayesian inference.

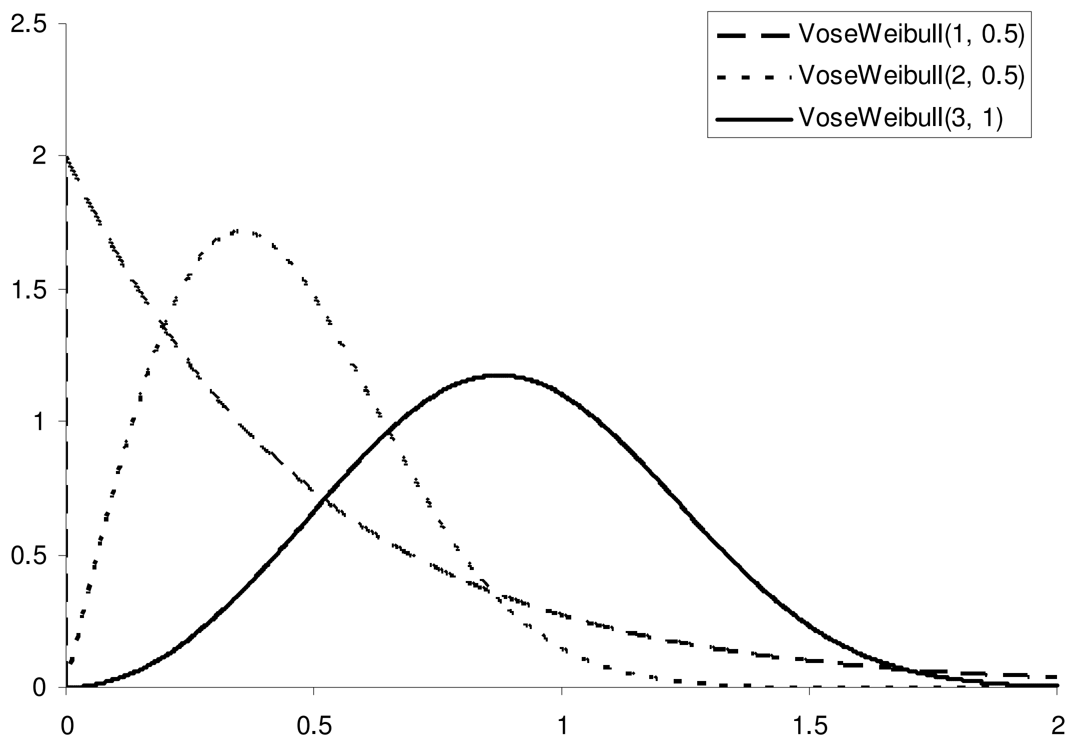
Equations

Probability density function :	$f(x)=\dfrac{1}{\max-\min}$
Cumulative distribution function :	$F(x)=\dfrac{x-\min}{\max-\min}$
Parameter restriction :	$\min<\max$
Domain :	$\min<x<\max$
Mean :	$\dfrac{\min+\max}{2}$
Mode :	No unique mode
Variance :	$\dfrac{(\max-\min)^2}{12}$
Skewness :	0
Kurtosis :	1.8

Weibull

VoseWeibull(α,β)

Graphs



Uses

The Weibull distribution is often used to model the time until occurrence of an event where the probability of occurrence changes with time (the process has "memory"), as opposed to the Exponential distribution where the probability of occurrence remains constant ("memoryless"). It has also been used to model variation in wind speed at a specific site.

Comments

The Weibull distribution becomes an exponential distribution when $\alpha = 1$, i.e. Weibull(1, β) = Expon(β). The Weibull distribution is very close to the Normal distribution when $\alpha = 3.25$. The Weibull distribution is named after the Swedish physicist Dr E. H. Wallodi Weibull (1887-1979) who used it to model the distribution of the breaking strengths of materials.

Equations

Probability density function :	$f(x) = \alpha\beta^{-\alpha}x^{\alpha-1}\exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right]$
Cumulative distribution function :	$F(x) = 1-\exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right]$
Parameter restriction :	$\alpha > 0, \beta > 0$
Domain :	$-\infty < x < +\infty$
Mean :	$\frac{\beta}{\alpha}\Gamma\left(\frac{1}{\alpha}\right)$
Mode :	$\beta\left(\frac{\alpha-1}{\alpha}\right)^{\frac{1}{\alpha}}$ if $\alpha \geq 1$ 0 if $\alpha < 1$
Variance :	$\frac{\beta^2}{\alpha}\left[2\Gamma\left(\frac{2}{\alpha}\right)-\frac{1}{\alpha}\Gamma\left(\frac{1}{\alpha}\right)^2\right]$
Skewness :	$\dfrac{3\Gamma\left(\frac{3}{\alpha}\right)+\frac{6}{\alpha}\Gamma\left(\frac{2}{\alpha}\right)\Gamma\left(\frac{1}{\alpha}\right)+\frac{2}{\alpha^2}\Gamma^3\left(\frac{1}{\alpha}\right)}{\sqrt{\frac{1}{\alpha}}\left[2\Gamma\left(\frac{2}{\alpha}\right)-\frac{1}{\alpha}\Gamma^2\left(\frac{1}{\alpha}\right)\right]^{\frac{3}{2}}}$
Kurtosis :	$\dfrac{4\Gamma\left(\frac{4}{\alpha}\right)-\frac{12}{\alpha}\Gamma\left(\frac{1}{\alpha}\right)\Gamma\left(\frac{3}{\alpha}\right)-\frac{12}{\alpha}\Gamma^2\left(\frac{2}{\alpha}\right)+\frac{24}{\alpha^2}\Gamma^2\left(\frac{1}{\alpha}\right)\Gamma\left(\frac{2}{\alpha}\right)-\frac{6}{\alpha^3}\Gamma^4\left(\frac{1}{\alpha}\right)}{\frac{1}{\alpha}\left[2\Gamma\left(\frac{2}{\alpha}\right)-\frac{1}{\alpha}\Gamma^2\left(\frac{1}{\alpha}\right)\right]}$

Multivariate distributions

Dirichlet

VoseDirichlet({α})

Uses

The Dirichlet distribution is the multivariate generalization of the beta distribution. {VoseDirichlet} is input in Excel as an array function.

It is used in modeling probabilities, prevalence of fractions where there are multiple states to consider. It is the multinomial extension to the beta distribution for a binomial process.

Example 1:

You have the results of a survey conducted in the premises of a retail outlet. The age and sex of 500 randomly selected shoppers were recorded:

<25 years, male: 38 people
25 to < 40 years, male: 72 people
> 40 years, male: 134 people
<25 years, female: 57 people
25 to < 40 years, female: 126 people
> 40 years, female: 73 people

In a manner analogous to the beta distribution, by adding 1 to each number of observations we can estimate the fraction of all shoppers to this store that are in each category as follows:

=VoseDirichlet({38+1,72+1,134+1,57+1,126+1,73+1})

The Dirichlet then returns the uncertainty about the fraction of all shoppers that are in each group.

Example 2:

A review of 1000 companies that were S&P AAA rated last year in your sector shows their rating one year later:

AAA: 908
AA: 83
A: 7
BBB or below: 2

If we assume that the market has similar volatilities to last year, we can estimate the probability that a company rated AAA now will be in each state next year as:

=VoseDirichlet({908+1,83+1,7+1,2+1})

The Dirichlet then returns the uncertainty about these probabilities.

Comments

The Dirichlet distribution is named after Johann Peter Gustav Lejeune Dirichlet. It is the conjugate to the multinomial distribution.

Equations

The probability density function of the Dirichlet distribution of order K is:

$$f(x) = \frac{1}{B(\alpha)} \prod_{i=1}^{K} x_i^{\alpha_i - 1}$$

where x is a K-dimensional vector $x = (x_1, x_2, \ldots, x_K)$, $\alpha = (\alpha_1, \ldots, \alpha_K)$ a parameter vector and B(α) is the multinomial Beta function:

$$B(\alpha) = \frac{\prod_{i=1}^{K} \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^{K} \alpha_i\right)}$$

Parameter restrictions: $\alpha_i > 0$

Domain: $0 \le x_i \le 1$; $\sum_{i=1}^{k} x_i = 1$

Inverse Multivariate Hypergeometric

VoseInvMultiHypereo({s},{d})

The Inverse Multivariate Hypergeomeric distribution answers the question: how many extra (wasted) random multivariate hypergeometric samples will occur before the required numbers of successes {s} are selected from each sub-population {D}.

For example, imagine that our population is split up into four sub-groups {A,B,C,D} of sizes {20,30,50,10} and that we are going to randomly sample from this population until we have {4,5,2,1} of each sub-group respectively. The number of extra samples we will have to make is modeled as:

=VoseInvMultiHypergeo({4,5,2,1},{20,30,50,10})

The total number of trials that need to be performed is:

=SUM({4,5,2,1}) + VoseInvMultiHypergeo({4,5,2,1},{20,30,50,10})

The InvMultiHypergeo2 is a multivariate distribution that responds to the same question, but breaks down the number of extra samples into their sub-groups.

Inverse Multivariate Hypergeometric 2

VoseInvMultiHypergeo2({s},{D})

The 2nd Inverse Multivariate Hypergeomeric distribution answers the question: how many extra (wasted) random multivariate hypergeometric samples would be drawn from each sub-population before the required numbers of successes {s} are selected from each sub-population {D}.

For example, imagine that our population is split up into four sub-groups {A,B,C,D} of sizes {20,30,50,10} and that we are going to randomly sample from this population until we have {4,5,2,1} of each sub-group respectively. The number of extra samples we will have to make for each sub-population A to D is modeled as the array function:

{=VoseInvMultiHypergeo({4,5,2,1},{20,30,50,10})}

So, for example, in the random scenario of the model shown below, there were a total of 21 extra samples taken over those required. Note that at least one category must be zero, since once the last category to be filled has the required number of samples the sampling stops so for that category at least there will be no extra samples.

The InvMultiHypergeo2 responds to the same question as the MultiHypergeo distribution but breaks down the number of extra samples into their sub-groups, whereas the Multihypergeo simply returns the total number of extra samples.

Multivariate Hypergeometric

VoseMultiHypergeo(N,{Dj})

The Multivariate Hypergeometric distribution is an extension of the Hypergeomeric distribution where more than two different states of individuals in a group exist.

Example

In a group of 50 people, of whom 20 were male, a VoseHypergeo(10,20,50) would describe how many from ten randomly chosen people would be male (and by deduction how many would therefore be female). However, let's say we have a group of 10 people as follows:

German	English	French	Canadian
3	2	1	4

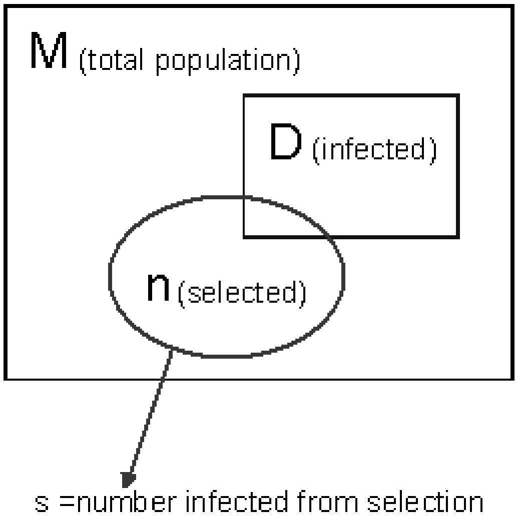
Now let's take a sample of 4 people at random from this group. We could have various numbers of each nationality in our sample:

German	English	French	Canadian
3	1	0	0
3	0	1	0
3	0	0	1
2	2	0	0
2	1	1	0
2	1	0	1
2	0	2	0
2	0	1	1
2	0	0	2
...	...	...	...
Etc.			

and each combination has a certain probability. The Multivariate Hypergeometric distribution is an array distribution, in this case generating simultaneously four numbers, that returns how many individuals in the random sample came from each sub-group (e.g. German, English, French, and Canadian).

Generation

The Multivariate Hypergeometric distribution is created by extending the mathematics of the Hypergeometric distribution. For the Hypergeometric distribution with a sample of size n, the probability of observing s individuals from a sub-group of size M, and therefore (n-s) from the remaining number (M-D):

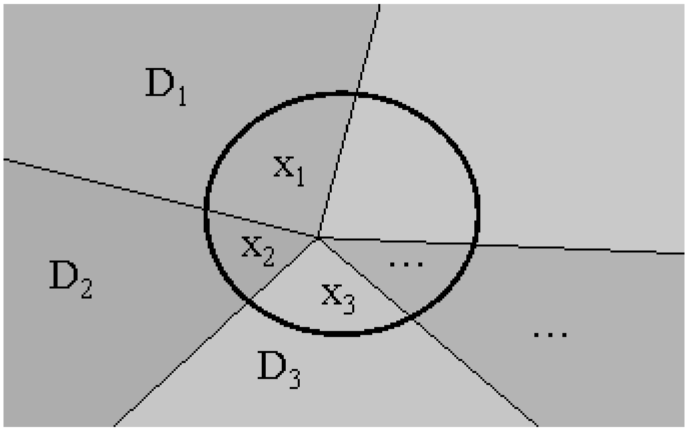


and results in the probability distribution for s:

$$f(x)=\frac{\binom{D}{x}\binom{M-D}{n-x}}{\binom{M}{n}}$$

where M is the group size, and D is the sub-group of interest. The numerator is the number of different sampling combinations (each of which has the same probability because each individual has the same probability of being sampled) where one would have exactly s from the sub-group D (and by implication (n-s) from the sub-group (M-D). The denominator is the total number of different combinations of individuals one could have in selecting n individuals from a group of size M. Thus the equation is just the proportion of different possible scenarios, each of which has the same probability, that would give us s from D.

The Multivariate Hypergeometric probability equation is just an extension of this idea. The figure below shows the graphical representation of the multivariate hypergeometric process: D1, D2, D3 and so on are the number of individuals of different types in a population, and x1, x2, x3, ... are the number of successes (the number of individuals in our random sample (circled) belonging to each category).



and results in the probability distribution for {s}:

$$f(x) = \frac{\binom{D_1}{x_1}\binom{D_2}{x_2}\cdots\binom{D_k}{x_k}}{\binom{M}{n}}$$

where $\sum_{i=1}^{k} D_i = M$, $\sum_{i=1}^{k} x_i = n$

Equations

Probability mass function :	$f(\{x_1, x_2, \ldots, x_j\}) = \frac{\binom{D_1}{x_1}\binom{D_2}{x_2}\cdots\binom{D_j}{x_j}}{\binom{M}{n}}$ where $M = \sum_{i=1}^{j} D_i$
Parameter restriction :	$0 < n \le M$, n, M, D_i are integers
Domain :	$\max(0, n + D_i - M) \le x_i \le \min(n, D_i)$

Multinomial

VoseMultinomial(N,{p})

The Multinomial distribution is an array distribution and is used to describe how many independent trials will fall into each of several categories where the probability of falling into any one category is constant for all trials. As such, it is an extension of the Binomial distribution where there are only two possible outcomes ('successes' and, by implication, 'failures').

Uses

For example, consider the action people might take on entering a shop:

Code	Action	Probability
A1	Enter and leave without purchase or sample merchandise	32%
A2	Enter and leave with a purchase	41%
A3	Enter and leave with sample merchandise	21%
A4	Enter to return a product and leave without purchase	5%
A5	Enter to return a product and leave with a purchase	1%

If 1000 people enter a shop, how many will match each of the above actions?

The answer is {Multinomial(1000,{32%, 41%, 21%, 5%, 1%})} which is an array function that generates five separate values. The sum of those five values must, of course, always add up to the number of trials (1000 in this example).

Equations

Probability mass function:	$f(\{x_1, x_2, \ldots, x_j\}) = \dfrac{n!}{x_1! \ldots x_j!} p_1^{x_1} \ldots p_j^{x_j}$
Parameter restrictions:	$p_i \geq 0, n \in \{1,2,3,\ldots\}, \sum_{i=1}^{n} p_i = 1$
Domain:	$x_i = \{0,1,2,\ldots,n\}$

MultiNormal

VoseMultiNormal({µi }, {cov_matrix})

Uses

A multinormal distribution, also sometimes called a multivariate normal distribution, is a specific probability distribution, which can be thought of as a generalization to higher dimensions of the one-dimensional normal distribution.

Equations

The probability density function of the MultiNormal distribution is the following function of an N-dimensional vector $x = (x_1, ..., x_N)$:

$$f(x) = \frac{1}{(2\pi)^{N/2}|C|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)\right]$$

where $\mu = (\mu_1, ..., \mu_N)$, C is the covariance matrix ($N \times N$) and $|C|$ is the determinant of C

Parameter restrictions: C must be a symmetric, positive semi-definite matrix.

Negative Multinomial

VoseNegMultinomial({s},{p})

The NegMultinomial distribution is a generalization of the NegBin distribution. The NegBin(s,p) distribution estimates the total number of binomial trials that are failures before s successes are achieved where there is a probability p of success with each trial.
For the NegMultinomial distribution, instead of having a single value for s, we now have a set of success values {s} representing different 'states' of successes (si) one can have, with each 'state' i having a probability pi of success.

Now, the NegMultinomial distribution tells us how many failures we will have before we have achieved the total number of successes ($\sum_{i=1}^{k} s_i$) .

Example

Suppose you want to do a telephone survey about a certain product you made by calling people you pick randomly out of the phone book.
You want to make sure that at the end of the week you have called 50 people who never heard of your product, 50 people who don't have internet at home and 200 people who use internet almost daily.

If you know the probabilities of succes pi , the NegMultinomial({50,50,200},{p1,p2,p3}) will tell you how many failures you'll have before you've called all the people you wanted en so you also know the total number of phone calls you'll have to make to reach the people you wanted.
The total number of phone calls = the total number of successes (300) + the total number of failures (NegMultinomial({50,50,200},{p1,p2,p3})).

Negative Multinomial 2

VoseNegMultinomial2({s},{p})

The NegMultinomial2 is the same distribution as the NegMultinomial, but instead of giving you the global number of failures before reaching a certain number of successes, the NegMultinomial2 gives you the number of failures in each ‘group’ or ‘state’.

So, in the example of the telephone survey (see NegMultinomial) where the total number of phone calls was equal to the total number of successes plus the total number of failures, the total number of failures would now be a sum of the number of failures in each group (3 groups in the example).