Probability Distributions

Summary

The **Probability Distributions** procedure performs various operations for any of 46 probability distributions. In particular, you may:

1. Plot the probability mass or density function, cumulative distribution, survivor function, log survivor function, or hazard function.
2. Calculate the cumulative distribution or inverse cumulative distribution.
3. Generate random numbers.

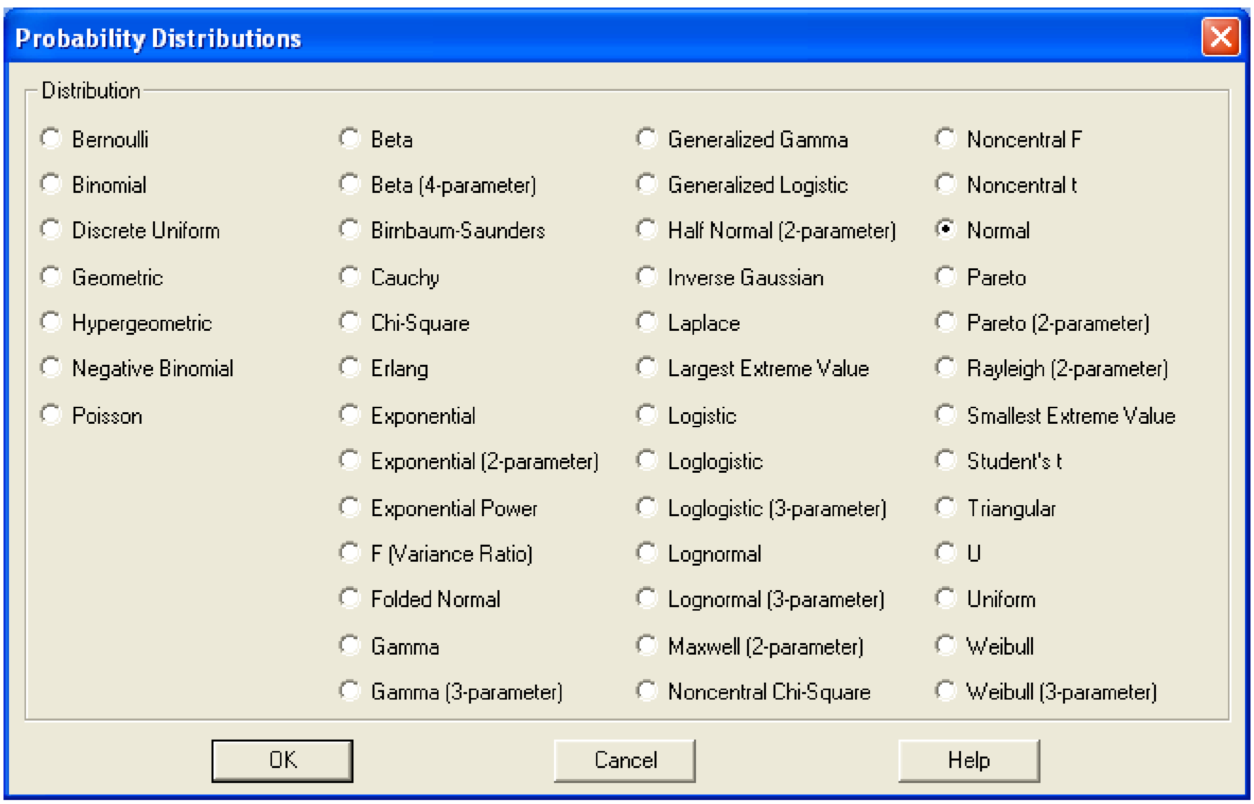
Sample StatFolio: *probdist.sgp*

Sample Data:

None.

Data Input

The data input dialog box is used to select the distribution to be evaluated.



- **Distribution:** select one of the 46 distributions listed.

Analysis Summary

The *Analysis Summary* shows the distribution selected and the values of its parameters.

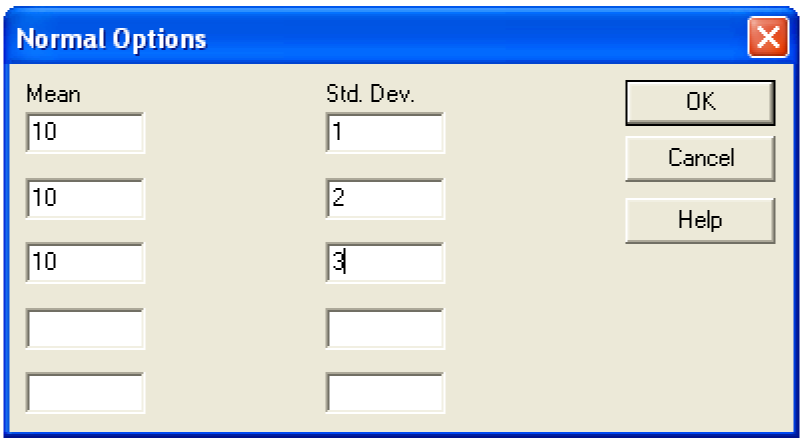
Probability Distributions

Distribution: Normal

Parameters:	*Mean*	*Std. Dev.*
Dist. 1	10	1
Dist. 2	10	2
Dist. 3	10	3
Dist. 4		
Dist. 5		

Analysis Options

Specify the up to 5 sets of parameters for the selected distribution.



The parameters required depend on the distribution selected on the data input dialog box. Definitions of the various distributions are given at the end of this document.

Cumulative Distribution

This pane shows the value of the cumulative distribution and probability density or mass function at up to 5 values of X.

Cumulative Distribution
Distribution: Normal

Lower Tail Area (<)

Variable	*Dist. 1*	*Dist. 2*	*Dist. 3*	*Dist. 4*	*Dist. 5*
10.5	0.691464	0.598708	0.566186		
11.5	0.933193	0.773374	0.691464		
12.5	0.99379	0.894351	0.797673		

Probability Density

Variable	*Dist. 1*	*Dist. 2*	*Dist. 3*	*Dist. 4*	*Dist. 5*
10.5	0.352065	0.193334	0.131147		
11.5	0.129518	0.150569	0.117355		
12.5	0.0175283	0.0913245	0.0939706		

Upper Tail Area (>)

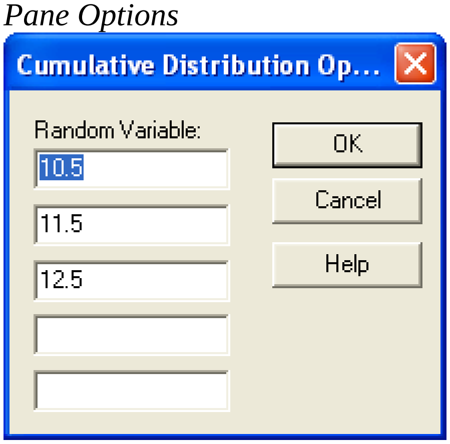
Variable	*Dist. 1*	*Dist. 2*	*Dist. 3*	*Dist. 4*	*Dist. 5*
10.5	0.308536	0.401292	0.433814		
11.5	0.066807	0.226626	0.308536		
12.5	0.00620966	0.105649	0.202327		

Included in the table are:

- **Lower Tail Area**: the probability that a random variable from the specified distribution is less than the value shown in the leftmost column.
- **Probability Density** (continuous distributions only): the height of the probability density function f(X) at the value shown in the leftmost column.
- **Probability Mass** (discrete distributions only): the probability that X equals the value shown in the leftmost column.
- **Upper Tail Area**: the probability that a random variable from the specified distribution is greater than the value shown in the leftmost column.

For example, F(X) = 0.691464 at X=10.5 for the first distribution in the table above.

Pane Options



- **Random Variable:** specify up to 5 values at which the cumulative distribution will be calculated.

Inverse CDF

The *Inverse CDF* (Cumulative Distribution Function) calculates the value of the random variable X at or below which lies a specified probability.

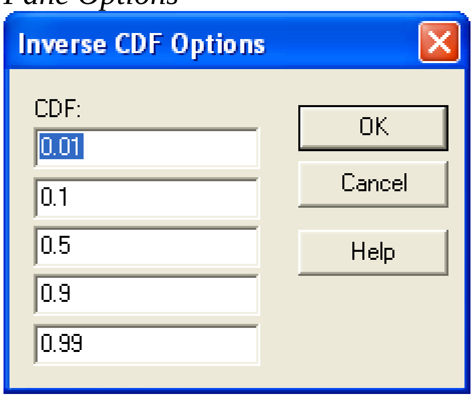
Inverse CDF
Distribution: Normal

CDF	*Dist. 1*	*Dist. 2*	*Dist. 3*	*Dist. 4*	*Dist. 5*
0.01	7.67365	5.3473	3.02094		
0.1	8.71845	7.43689	6.15534		
0.5	10	10	10		
0.9	11.2816	12.5631	13.8447		
0.99	12.3264	14.6527	16.9791		

In the case of a continuous distribution, the value of X is calculated such that the cumulative distribution function F(X) equals the probability shown in the leftmost column. In the case of a discrete distribution, the value displayed is the smallest value of X such that F(X) is greater than or equal to the probability shown in the leftmost column.

For example, F(X) = 0.01 at X=7.67365 for the first distribution in the table above.

Pane Options



- **CDF:** specify up to 5 values of the cumulative distribution function at which values of X will be determined. The values must be greater than 0.0 and less than 1.0.

Random Numbers

Select this pane to generate random numbers from the selected distribution.
The steps for generating random numbers are:

1. Specify the type of probability distribution on the data input dialog box.
2. Use the *Analysis Options* dialog box to specify the parameters of the distribution.
3. Select **Random Numbers** from the list of tabular options and then select *Pane Options.*
4. On the *Pane Options* dialog box, specify how many random numbers should be generated.
5. Select Save Results to place the random numbers in the datasheet.

Each time you select *Save Results,* a different set of random numbers will be generated.

The method used for generating random numbers depends on the distribution selected. Many of the distributions use the *inverse transformation method,* in which a set of *n* random numbers U_i are generated from a uniform distribution over the interval (0,1) and then converted to the desired distribution by letting

$$X_i = F^{-1}(U_i) \tag{1}$$

The uniform random numbers are generated using three linear congruential generators in a manner designed to yield the same random sequence on any computer (given the same seed). STATGRAPHICS sets the seed based on the time it is loaded, so that each session will generate a different sequence of random numbers.

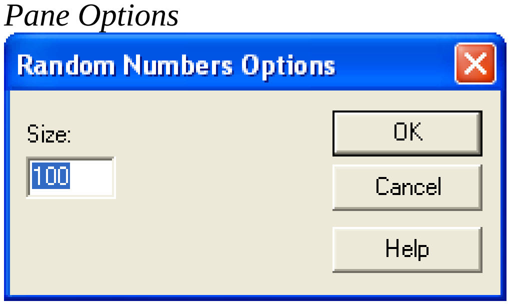
The methods for generating random numbers are summarized below:

Distribution	**Method**
Bernoulli	If U≤p, X=1. Else X=0.
Binomial	Sum of *n* Bernoulli random variables
Discrete uniform	(int)uniform(a,b+1)
Geometric	$(\text{int})\left(\frac{\ln(U_i)}{\ln(1-p)}\right)$
Hypergeometric	Generation of m successes from finite population of size *n* without replacement
Negative binomial	*k* + (sum of *k* geometric random variables)
Poisson	Uses relationship between Poisson and exponential random variables (see Law and Kelton).
Beta	$\frac{Y_1}{Y_1+Y_2}$ where Y_1~gamma(α_1,1) and Y_2~gamma(α_2,1)
Beta (4-parameter)	Translated beta random variable.
Birnbaum-Saunders	$\frac{\left(Z\beta+\sqrt{4Z^2\beta^2}\right)^2\theta}{4}$ where Z ~ normal(0,1)

Cauchy	$\theta + \beta\left(\frac{Z_1}{Z_2}\right)$ where Z_1~normal(0,1) and Z_2~normal(0,1)
Chi-square	gamma(ν/2,0.5)
Erlang	$-\frac{\ln\left(\prod_{i=1}^{\alpha} U_i\right)}{\alpha\beta}$
Exponential	Inverse transform method.
Exponential (2-parameter)	Translated exponential random variable.
Exponential power	Numerical inverse transform method.
F	$\frac{Y_1 / v_1}{Y_2 / v_2}$ where Y_1~chisquare(v_1) and Y_2~chisquare(v_2)
Folded normal	|X| where X ~ normal(μ,σ)
Gamma	If α=1, exponential(λ). Else acceptance-rejection method (see Law and Kelton).
Gamma (3-parameter)	Translated gamma random variable.
Generalized gamma	Numerical inverse transform method.
Generalized logistic	Numerical inverse transform method.
Half-normal	Generate X~normal(μ,σ). If X≥μ, return X. Else return 2μ-X.
Inverse Gaussian	Micael/Schucany/Hass method (see Gentle)
Laplace	Inverse transform method.
Largest extreme value	Inverse transform method.
Logistic	Inverse transform method.
Loglogistic	Inverse transform method.
Loglogistic (3-parameter)	Translated loglogistic random variable.
Lognormal	exp[normal(μ,σ)]
Lognormal (3-parameter)	Translated lognormal random variable.
Maxwell	$a + X_1^2 + X_2^2 + X_3^2$ where X_1, X_2, and X_3 ~ normal(0,b)
Noncentral chi-square	If ν is integer, $\sum_{i=1}^{\nu}\left(Z_i + \lambda / \nu\right)^2$ where Z_i~normal(0,1). Else numerical inverse transform method.
Noncentral F	$\frac{Y_1 / v_1}{Y_2 / v_2}$

	where Y_1~noncentral chisquare(v_1,λ) and Y_2~chisquare(v_2)
Noncentral t	$\frac{(Z_1+\lambda)}{\sqrt{Y_1/\nu}}$ where Z_1~normal(0,1) and Y_1~chisquare(ν)
Normal	Polar method (see Law and Kelton).
Pareto	Inverse transform method.
Pareto (2-parameter)	Translated Pareto random variable.
Rayleigh	$a+b\sqrt{-\log(U)}$
Smallest extreme value	Inverse transform method.
Student's t	$\frac{Z_1}{\sqrt{Y_1/\nu}}$ where Z_1~normal(0,1) and Y_1~chisquare(ν)
Triangular	Inverse transform method.
U	Inverse transform method.
Uniform	a+(b-a)U
Weibull	Inverse transform method.
Weibull (3-parameter)	Translated Weibull random variable.

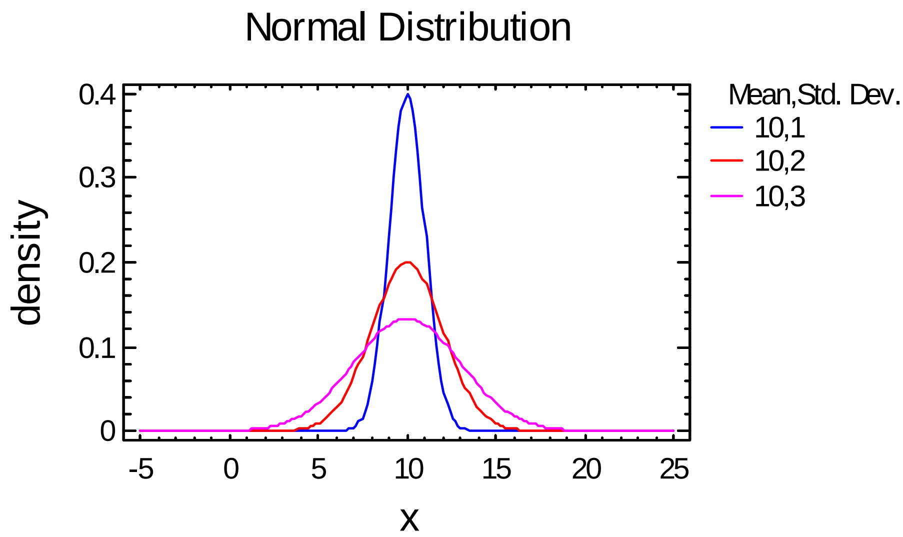
Pane Options



- **Size:** select the number *n* of random numbers to be generated. After selecting the size, click on *Save Results* to save the random numbers to the datasheet.

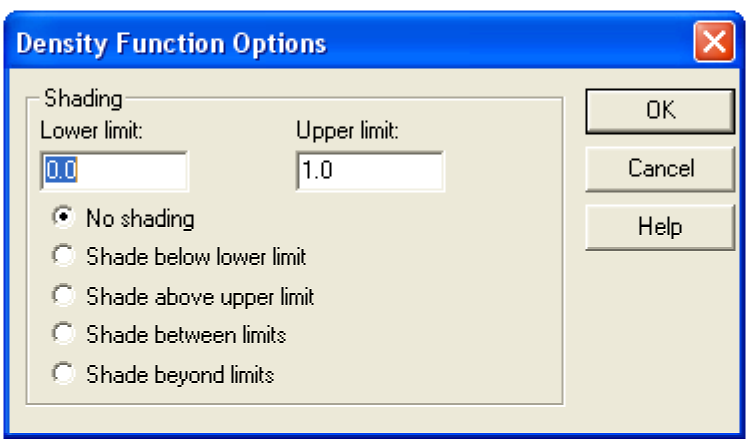
Density/Mass Function

This pane plots the probability density function f(X) for continuous distributions or the probability mass function p(x) for discrete distributions.



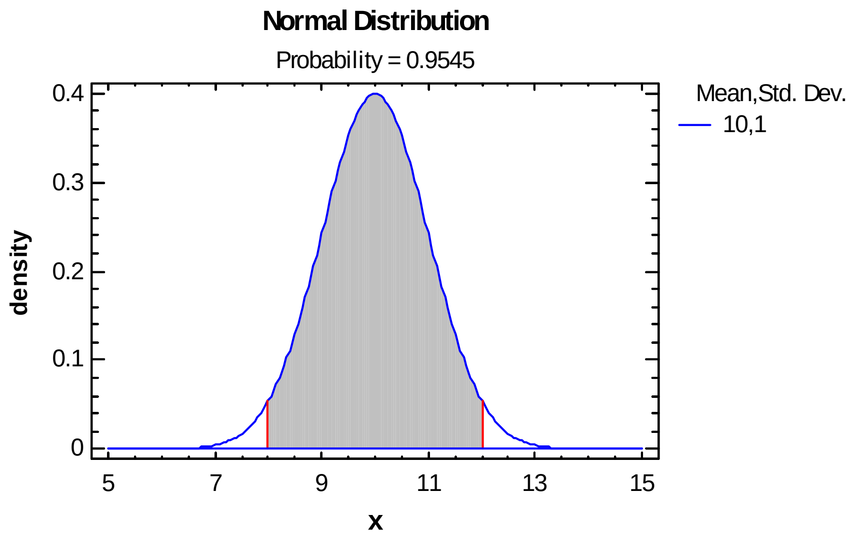
For a continuous distribution such as the normal distribution, the area under the density function over an interval of values for X equals the probability that X falls within that interval.

When plotting the p.d.f. for a single, continuous distribution, *Pane Options* can be used to specify areas that will be shaded on the plot:



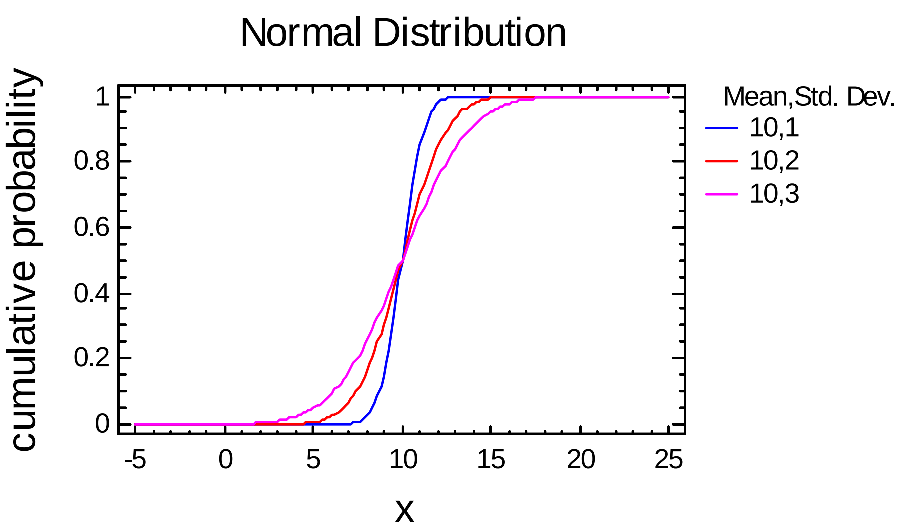
- **Shading**: specifies one or more regions to be shaded.

The specified areas will be indicated on the plot and the probabilities associated with the sum of all shaded areas will be displayed:



CDF

This pane plots the cumulative distribution function F(X).



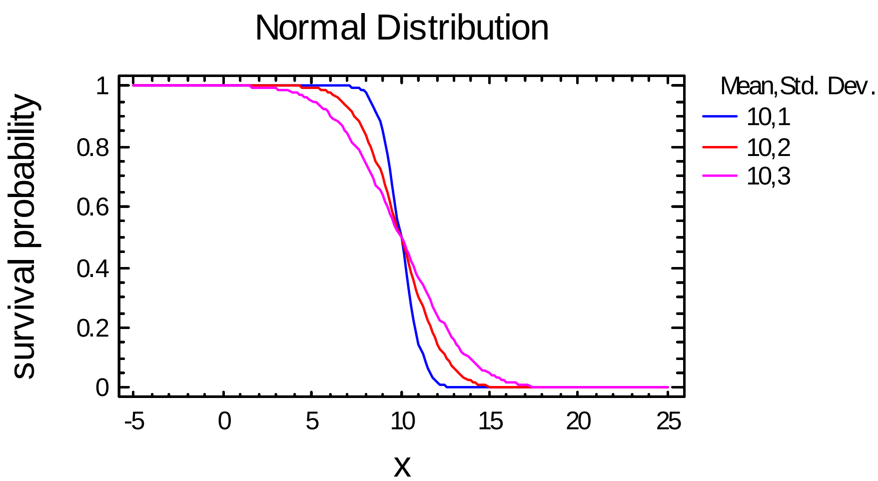
F(X) equals the probability that the random variable will be less than or equal to X.

Survivor Function

This pane plots the survivor function S(X), defined by

$$S(X) = 1 - F(X) \tag{2}$$

where F(X) is the cumulative distribution function.

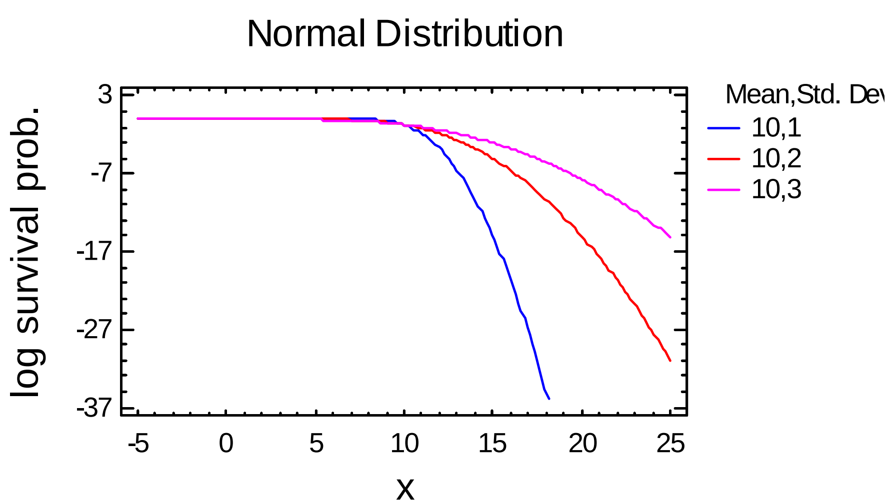


S(X) equals the probability that the random variable will be greater than X. The name of the function is derived from situations where X represents an individual's or product's lifetime. In that case, S(X) is the probability that an individual survives at least X time units.

Log Survivor Function

This pane plots the log of the survivor function S(X), which is defined by

$$S(X) = 1 - F(X) \tag{3}$$



Hazard Function

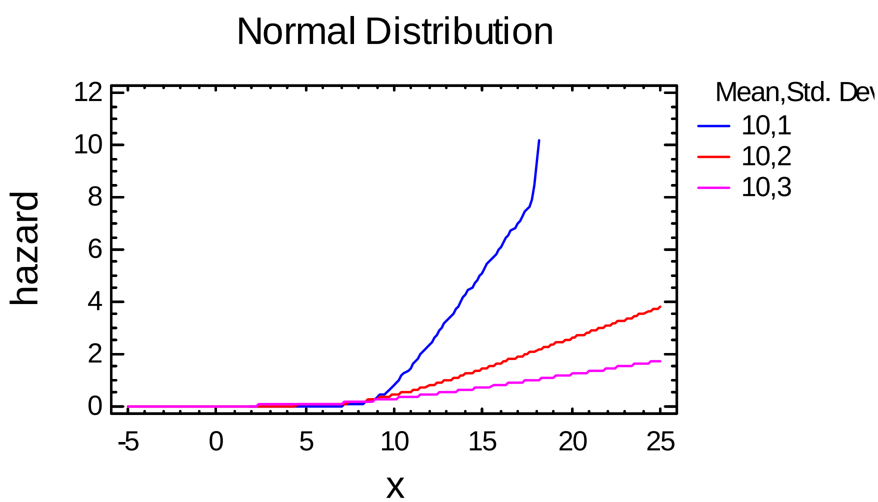
The hazard function represents the conditional distribution of a random variable given than it is at least X. For continuous distributions, it is defined by

$$H(X) = f(x) / S(X) \tag{4}$$

where f(x) is the probability density function and S(X) is the survivor function. For discrete distributions, it is defined by

$$H(X) = p(x+1) / S(X) \tag{5}$$

where p(x) is the probability mass function.



In life data analysis, the hazard function represents the conditional failure rate, i.e., the probability of failure in the next small increment of time given that an individual has survived until time X.

Save Results

You can save the following results to the datasheet:

- **Random number for Dist. #**: a set of random numbers generated from the specified distribution. The size of the set is determined from the *Pane Options* dialog box for the *Random Numbers* pane.

Definitions

STATGRAPHICS generates results for 46 different probability distributions, 7 for discrete random variables and the other 39 for continuous random variables. Each of the distributions contains 1 or more parameters, which are either specified by the user or estimated from a data sample.

Bernoulli Distribution

Range of X: 0 or 1

Common use: representation of an event with 2 possible outcomes. In the distributions below, the primary outcome will be referred to as a "success".

PMF: $p(x) = p^{x}(1-p)^{1-x}$

Parameters: event probability $0 \le p \le 1$

Mean: p

Variance: $p(1-p)$

Binomial Distribution

Range of X: 0, 1, 2, …, n

Common use: distribution of number of successes in a sample of *n* independent Bernoulli trials. Commonly used for number of defects in a sample of size *n*.

PMF: $p(x) = \binom{n}{x} p^{x}(1-p)^{n-x}$

Parameters: event probability $0 \le p \le 1$, number of trials $n \ge 1$.

Mean: np

Variance: $np(1-p)$

Discrete Uniform Distribution

Range of X: a, a+1, a+2, …, b

Common use: distribution of an integer valued variable with both a lower bound and an upper bound.

PMF: $p(x) = \frac{1}{b-a+1}$

Parameters: lower limit a, upper limit $b \ge a$.

Mean: $\frac{a+b}{2}$

Variance: $\frac{(b-a+1)^{2}-1}{12}$

Geometric Distribution

Range of X: 0, 1, 2, …

Common use: waiting time until the occurrence of the first success in a sequence of independent Bernoulli trials. Number of items inspected before the first defect is found.

PMF: $p(x) = p(1-p)^{x}$

Parameters: event probability $0 \le p \le 1$

Mean: $\frac{1-p}{p}$

Variance: $\frac{1-p}{p^{2}}$

Hypergeometric Distribution

Range of X: max(0,n-m), 1, 2, …, min(m,n)

Common use: number of items of a given type selected from a finite population with two types of items, such as good and bad. Acceptance sampling from lots of fixed size.

PMF: $p(x) = \dfrac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$

Parameters: population size N, number of items 0 ≤ m ≤ N, sample size n

Mean: $\dfrac{mn}{N}$

Variance: $\dfrac{\left(\frac{mn}{N}\right)\left(1-\frac{m}{N}\right)(N-n)}{(N-1)}$

Negative Binomial (Pascal) Distribution

Range of X: 0, 1, 2, …

Common use: waiting time until the occurrence of k successes in a sequence of independent Bernoulli trials. Number of good items inspected before the k^{th} defect is found.

PMF: $p(x) = \binom{x+k-1}{x} p^k (1-p)^x$

Parameters: event probability p, number of successes k

Mean: $k\dfrac{(1-p)}{p}$

Variance: $\dfrac{k(1-p)}{p^2}$

NOTE: the definition of this distribution has changed from earlier versions. Earlier versions included k as part of the definition of the random variable, so that the range of X was *k* or greater instead of 0 or greater. The change has been made to allow the negative binomial distribution to be more easily used as a model for overdispersed count data, i.e, integer data in which the variance exceeds the mean.

Poisson Distribution

Range of X: 0, 1, 2, …

Common use: number of events occurred in an interval of fixed size when events occur independently. Common model for number of defects per unit.

PMF: $p(x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$

Parameters: mean $\lambda > 0$

Mean: λ

Variance: λ

Beta Distribution

Range of X: $0 \le X \le 1$

Common use: distribution of a random proportion.

PDF: $f(x) = \dfrac{x^{\alpha_1-1}(1-x)^{\alpha_2-1}}{B(\alpha_1,\alpha_2)}$

Parameters: shape $\alpha_1 > 0$, shape $\alpha_2 > 0$

Mean: $\dfrac{\alpha_1}{\alpha_1+\alpha_2}$

Variance: $\dfrac{\alpha_1\alpha_2}{(\alpha_1+\alpha_2)^2(\alpha_1+\alpha_2+1)^2}$

Beta Distribution (4-parameter)

Range of X: a ≤ X ≤ b

Common use: model for variable with both lower and upper limits. Often used as a prior distribution for Bayesian analysis.

PDF: $f(x) = \dfrac{(x-a)^{\alpha_1-1}(b-x)^{\alpha_2-1}}{B(\alpha_1,\alpha_2)(b-a)^{\alpha_1+\alpha_2-1}}$

Parameter: shape $\alpha_1 > 0$, shape $\alpha_2 > 0$, lower limit a, upper limit b > a

Mean: $a + \dfrac{b\alpha_1}{\alpha_1+\alpha_2}$

Variance: $\dfrac{\alpha_1\alpha_2(b-a)^2}{(\alpha_1+\alpha_2)^2(\alpha_1+\alpha_2+1)^2}$

Birnbaum-Saunders Distribution

Range of X: X > 0

Common use: model for the number of cycles needed to cause a crack to grow to a size that would cause a fracture to occur.

PDF: $f(x) = \dfrac{\sqrt{\dfrac{x}{\theta}}+\sqrt{\dfrac{\theta}{x}}}{2\beta x}\phi\left(\dfrac{1}{\beta}\left(\sqrt{\dfrac{x}{\theta}}-\sqrt{\dfrac{\theta}{x}}\right)\right)$ where $\phi(z)$ is the standard normal pdf

Parameters: shape $\beta > 0$, scale $\theta > 0$

Mean: $\theta\left(1+\dfrac{\beta^2}{2}\right)$

Variance: $(\theta\beta)^2\left(1+\dfrac{5\beta^2}{4}\right)$

Cauchy Distribution

Range of X: all real X

Common use: model for measurement data with longer and flatter tails than the normal distribution.

PDF: $f(x) = \dfrac{1}{\pi\beta}\left[\left(\dfrac{x-\theta}{\beta}\right)^2+1\right]^{-1}$

Parameters: mode θ, scale $\beta > 0$

Mean: not defined

Variance: not defined

Chi-Squared Distribution

Range of X: X ≥ 0

Common use: distribution of the sample variance s^2 from a normal population.

PDF: $f(x) = \dfrac{x^{(\nu-2)/2} e^{-x/2}}{2^{\nu/2}\Gamma\left(\dfrac{\nu}{2}\right)}$

Parameters: degrees of freedom $\nu > 0$

Mean: ν

Variance: 2ν

Erlang Distribution

Range of X: X ≥ 0

Common use: length of time before α arrivals in a Poisson process.

PDF: $f(x) = \dfrac{\lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}$

Parameters: integer shape $\alpha \geq 1$, scale $\lambda > 0$

Mean: $\dfrac{\alpha}{\lambda}$

Variance: $\dfrac{\alpha}{\lambda^2}$

Exponential Distribution

Range of X: X > 0

Common use: time between consecutive arrivals in a Poisson process. Lifetime of items with a constant hazard rate.

PDF: $f(x) = \lambda e^{-\lambda x}$

Parameters: mean $\lambda > 0$

Mean: $\dfrac{1}{\lambda}$

Variance: $\dfrac{1}{\lambda^2}$

Exponential Distribution (2-parameter)

Range of X: X > θ

Common use: model for lifetimes with a lower limit.

PDF: $f(x) = \lambda e^{-\lambda(x-\theta)}$

Parameters: threshold θ, scale $\lambda > 0$

Mean: $\theta + \dfrac{1}{\lambda}$

Variance: $\dfrac{1}{\lambda^2}$

Exponential Power Distribution

Range of X: all real X

Common use: symmetric distribution with parameter controlling the kurtosis. Special cases include normal and Laplace distributions.

PDF: $f(x) = \dfrac{1}{\Gamma\left(1+\dfrac{1+\beta}{2}\right)2^{1+(1+\beta)/2}\phi} \exp\left(-\dfrac{1}{2}\left|\dfrac{x-\mu}{\phi}\right|^{2/(1+\beta)}\right)$

Parameters: mean μ, shape $\beta \geq -1$, scale $\phi > 0$

Mean: μ

Variance: $2^{(1+\beta)}\left\{\dfrac{\Gamma\left[\dfrac{3}{2}(1+\beta)\right]}{\Gamma\left[\dfrac{1}{2}(1+\beta)\right]}\right\}\phi^2$

F Distribution

Range of X: $X \geq 0$

Common use: distribution of the ratio of two independent variance estimates from a normal population.

PDF: $f(x) = \dfrac{\Gamma\left(\dfrac{v+w}{2}\right)v^{v/2}w^{w/2}x^{(v-2)/2}}{\Gamma\left(\dfrac{v}{2}\right)\Gamma\left(\dfrac{w}{2}\right)(w+vx)^{(v+w)/2}}$

Parameters: numerator degrees of freedom v > 0, denominator degrees of freedom w > 0

Mean: $\dfrac{w}{w-2}$ if w > 2

Variance: $\dfrac{2w^2(v+w-2)}{v(w-2)^2(w-4)}$ if w > 4

Folded Normal Distribution

Range of X: $X \geq 0$

Common use: absolute values of data that follows a normal distribution.

PDF: $f(x) = \dfrac{1}{\sigma}\sqrt{\dfrac{2}{\pi}}\cosh\left(\dfrac{\mu x}{\sigma^2}\right)e^{-\frac{x^2+\mu^2}{2\sigma^2}}$

Parameters: location $\mu > 0$, scale $\sigma \geq 0$

Mean: $\sigma\sqrt{\dfrac{2}{\pi}}\exp\left(-\dfrac{\mu^2}{2\sigma^2}\right) - \mu\left|1-2\Phi\left(\dfrac{\mu}{\sigma}\right)\right|$ where $\Phi(z)$ is the standard normal cdf

Variance: $\mu^2 + \sigma^2 - \left[\sigma\sqrt{\dfrac{2}{\pi}}\exp\left(-\dfrac{\mu^2}{2\sigma^2}\right) + erf\left(\dfrac{\mu}{\sqrt{2}\sigma}\right)\mu\right]^2$

Gamma Distribution

Range of X: $X \geq 0$

Common use: model for positively skewed measurements. Time to complete a task, such as a repair.

PDF: $f(x) = \dfrac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)}$

Parameters: shape $\alpha > 0$, scale $\lambda > 0$

Mean: $\dfrac{\alpha}{\lambda}$

Variance: $\dfrac{\alpha}{\lambda^2}$

Gamma Distribution (3-parameter)

Range of X: X ≥ θ

Common use: model for positively skewed data with a fixed lower bound.

PDF: $f(x) = \dfrac{\lambda^{\alpha} x^{\alpha-1} e^{-\lambda(x-\theta)}}{\Gamma(\alpha)}$

Parameters: shape α >0, scale λ > 0, threshold θ

Mean: $\theta + \dfrac{\alpha}{\lambda}$

Variance: $\dfrac{\alpha}{\lambda^2}$

Generalized Gamma Distribution

Range of X: X > 0

Common use: general distribution containing the exponential, gamma, Weibull, and lognormal distributions as special cases.

PDF: $f(x) = \dfrac{\lambda}{\sigma x} \dfrac{1}{\Gamma(\lambda^{-2})} \exp\left[\dfrac{w}{\lambda} + \dfrac{\ln(\lambda^{-2})}{\lambda^2} - \exp(\lambda w + \ln(\lambda^{-2}))\right]$

where w = [log(x) – μ] / σ.

Parameters: location μ, scale σ > 0, shape λ > 0

Mean: $\exp\left[\mu + \dfrac{\sigma}{\lambda}\ln(\lambda^{-2})\right] \dfrac{\Gamma\left(\dfrac{\sigma}{\lambda} + \lambda^{-2}\right)}{\Gamma(\lambda^{-2})}$

Variance: $\exp\left[\mu + \dfrac{\sigma}{\lambda}\ln(\lambda^{-2})\right]^2 \left[\dfrac{\Gamma\left(\dfrac{2\sigma}{\lambda} + \lambda^{-2}\right)}{\Gamma(\lambda^{-2})} - \dfrac{\Gamma^2\left(\dfrac{2\sigma}{\lambda} + \lambda^{-2}\right)}{\Gamma^2(\lambda^{-2})}\right]$

Generalized Logistic Distribution

Range of X: all real x

Common use: used for the analysis of extreme values. May be either left-skewed or right-skewed, depending on the shape parameter.

PDF: $f(x) = \dfrac{\gamma}{\kappa} \dfrac{\exp(-(x-\mu)/\kappa)}{[1 + \exp(-(x-\mu)/\kappa)]^{1+\gamma}}$

Parameters: location μ, scale κ > 0, shape γ > 0

Mean: $\mu + [0.5226 + \Psi(\gamma)]\kappa$ where Ψ(z) is the digamma function

Variance: $\left[\dfrac{\pi^2}{6} + \Psi'(\gamma)\right]\kappa^2$

Half Normal Distribution

Range of X: X ≥ μ

Common use: normal distribution folded about its mean.

PDF: $f(x) = \frac{1}{\sigma}\sqrt{\frac{2}{\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$

Parameters: scale $\sigma > 0$, threshold μ

Mean: $\mu + \sqrt{\frac{2}{\pi}}\sigma$

Variance: $\sigma^2\left(1-\frac{2}{\pi}\right)$

Inverse Gaussian Distribution

Range of X: X > 0

Common use: first passage time in Brownian motion.

PDF: $f(x) = \frac{1}{x}\frac{\sqrt{\beta}}{\exp(z/2)}\phi\left[\sqrt{\beta}\left(\frac{e^z-1}{e^{z/2}}\right)\right]$ where z = ln(x/θ)

Parameters: mean $\theta > 0$, scale $\beta > 0$

Mean: θ

Variance: $\frac{\theta^2}{\beta}$

Laplace (Double Exponential) Distribution

Range of X: all real X

Common use: symmetric distribution with a very pronounced peak and long tails

PDF: $f(x) = \frac{\lambda}{2}e^{-\lambda|x-\mu|}$

Parameters: mean μ, scale $\lambda > 0$

Mean: μ

Variance: $\frac{2}{\lambda^2}$

Largest Extreme Value Distribution

Range of X: all real X

Common use: distribution of the largest value in a sample from many distributions. Also used for positively skewed measurement data.

PDF: $f(x) = \frac{1}{\beta}\exp\left\{-\left(\frac{x-\alpha}{\beta}\right)-\exp\left(-\frac{x-a}{\beta}\right)\right\}$

Parameters:

Mean: $\alpha+\beta\Gamma^{-1}(1)$

Variance: $\frac{\beta^2\pi^2}{6}$

Logistic Distribution

Range of X: all real X

Common use: used as a model for growth and as an alternative to the normal distribution.

PDF: $f(x) = \frac{1}{\sigma} \frac{\exp(z)}{[1+\exp(z)]^2}$ where $z = \frac{x-\mu}{\sigma}$

Parameters: mean μ, standard deviation $\sigma > 0$

Mean: μ

Variance: σ^2

Loglogistic Distribution

Range of X: X > 0

Common use: used for data where the logarithms follow a logistic distribution.

PDF: $f(x) = \frac{1}{\sigma x} \frac{\exp(z)}{[1+\exp(z)]^2}$ where $z = \frac{\ln(x)-\mu}{\sigma}$

Parameters: median exp(μ), scale $\sigma > 0$

Mean: $\exp(\mu)\Gamma(1+\sigma)\Gamma(1-\sigma)$

Variance: $\exp(2\mu)\left[\Gamma(1+2\sigma)\Gamma(1-2\sigma) - \Gamma^2(1+\sigma)\Gamma^2(1-\sigma)\right]$

Loglogistic Distribution (3-parameter)

Range of X: X > θ

Common use: used for data where the logarithms follow a logistic distribution after subtracting a threshold value.

PDF: $f(x) = \frac{1}{\sigma x} \frac{\exp(z)}{[1+\exp(z)]^2}$ where $z = \frac{\ln(x-\theta)-\mu}{\sigma}$

Parameters: median exp(μ), scale $\sigma > 0$, threshold θ

Mean: $\theta + \exp(\mu)\Gamma(1+\sigma)\Gamma(1-\sigma)$

Variance: $\exp(2\mu)\left[\Gamma(1+2\sigma)\Gamma(1-2\sigma) - \Gamma^2(1+\sigma)\Gamma^2(1-\sigma)\right]$

Lognormal Distribution

Range of X: X > 0

Common use: used for data where the logarithms follow a normal distribution.

PDF: $f(x) = \frac{1}{x\sqrt{2\pi}\sigma} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$

Parameters: location μ, scale σ > 0

Mean: $e^{\mu+\sigma^2/2}$

Variance: $e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)$

Lognormal Distribution (3-parameter)

Range of X: X > θ

Common use: used for data where the logarithms follow a normal distribution after subtracting a threshold value.

PDF: $f(x) = \frac{1}{(x-\theta)\sqrt{2\pi}\sigma} e^{-\frac{(\ln(x-\theta)-\mu)^2}{2\sigma^2}}$

Parameters: location μ, scale σ > 0, threshold θ

Mean: $\theta + e^{\mu+\sigma^2/2}$

Variance: $e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)$

Maxwell Distribution

Range of X: X > θ

Common use: the speed of a molecule in an ideal gas.

PDF: $f(x) = \sqrt{\frac{2}{\pi}} \frac{(x-\theta)^2}{\beta^3} \exp\left[-\frac{1}{2}\left(\frac{x-\theta}{\beta}\right)^2\right]$

Parameters: scale $\beta > 0$, threshold θ

Mean: $\theta + \beta\sqrt{8}/\sqrt{\pi}$

Variance: $\beta^2(3 - 8/\pi)$

Noncentral Chi-squared Distribution

Range of X: X ≥ 0

Common use: used to calculate the power of chi-squared tests.

PDF: $f(x) = \sum_{j=0}^{\infty} \frac{1}{j!}\left(\frac{c}{2}\right)^j e^{-c/2} \frac{x^{\nu/2+j-1} e^{-x/2}}{2^{\nu/2+j}\,\Gamma(\nu/2+j)}$

Parameters: degrees of freedom ν > 0, noncentrality c ≥ 0

Mean: $\nu + c$

Variance: $2(\nu + 2c)$

Noncentral F Distribution

Range of X: X ≥ 0

Common use: used to calculate the power of F tests.

PDF: $f(x) = \sum_{j=0}^{\infty} \frac{1}{j!}\left(\frac{c}{2}\right)^j e^{-c/2} \frac{(v/w)^{v/2+j}}{B(v/2+j, w/2)}\left(1 + \frac{v}{w}x\right)^{-((v+w)/2+j)}$

Parameters: number degrees of freedom v > 0, denominator degrees of freedom w > 0, noncentrality c > 0

Mean: $\frac{w(v+c)}{v(w-2)}$ if w > 2

Variance: $2\left(\frac{w}{v}\right)^2 \frac{(v+c)^2 + (v+2c)(w-2)}{(w-2)^2(w-4)}$ if w > 4

Noncentral t Distribution

Range of X: all real X

Common use: used to calculate the power of t tests.

PDF: $f(x) = \sum_{j=0}^{\infty} \frac{1}{j!}\left(c\sqrt{2}\right)^j e^{-c^2/2} \frac{\Gamma[(\nu+j+1)/2]}{\Gamma(\nu/2)\Gamma(1/2)} \frac{x^j}{\nu^{(j+1)/2}}\left(1 + \frac{x^2}{\nu}\right)^{-(\nu+j+1)/2}$

Parameters: degrees of freedom ν > 0, noncentrality c ≥ 0

Mean: $(\nu/2)^{1/2} \frac{\Gamma[(\nu-1)/2]}{\Gamma(\nu/2)} c$

Variance: $\frac{\nu}{\nu-2}(1+c^2) - \left\{(\nu/2)^{1/2} \frac{\Gamma[(\nu-1)/2]}{\Gamma(\nu/2)} c\right\}^2$

Normal Distribution

Range of X: all real X

Common use: widely used for measurement data, particularly when variability is due to many sources.

PDF: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Parameters: mean μ, standard deviation $\sigma > 0$

Mean: μ

Variance: σ^2

Pareto Distribution

Range of X:X ≥ 1

Common use: model for many socio-economic quantities with very long upper tails.

PDF: $f(x) = cx^{-c-1}$

Parameters: shape c > 0

Mean: $\frac{c}{c-1}$ if c > 1

Variance: $\frac{c}{(c-2)(c-1)^2}$ if c > 2

Pareto Distribution (2-parameter)

Range of X:X ≥ θ

Common use: distribution of socio-economic quantities with a lower bound.

PDF: $f(x) = c\theta^c x^{-c-1}$

Parameters: shape c >0, threshold θ ≥ 0

Mean: $\frac{\theta c}{c-1}$ if c > 1

Variance: $\frac{\theta^2 c}{(c-2)(c-1)^2}$ if c > 2

Rayleigh Distribution

Range of X:X > θ

Common use: the distance between neighboring items in a pattern generated by a Poisson process.

PDF: $f(x) = \frac{2}{x-\theta}\left(\frac{x-\theta}{\beta}\right)^2 \exp\left[-\left(\frac{x-\theta}{\beta}\right)^2\right]$

Parameters: scale $\beta > 0$, threshold θ

Mean: $\theta + \beta\sqrt{\pi}/2$

Variance: $\beta^2(1-\pi/4)$

Smallest Extreme Value Distribution

Range of X: all real X

Common use: distribution of the smallest value in a sample from many distributions. Also used for negatively skewed measurement data.

PDF: $f(x) = \frac{1}{\beta}\exp\left\{\left(\frac{x-\alpha}{\beta}\right) - \exp\left(\frac{x-a}{\beta}\right)\right\}$

Parameters: mode α, scale $\beta > 0$

Mean: $\alpha - \beta\Gamma^{-1}(1)$

Variance: $\frac{\beta^2 \pi^2}{6}$

Student's t Distribution

Range of X: all real X

Common use: reference distribution for the sample mean when sampling from a normal population with unknown variance.

PDF: $f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)\left[1+\frac{x^2}{v}\right]^{-(v+1)/2}}{\sqrt{\pi v}\,\Gamma\left(\frac{v}{2}\right)}$

Parameters: degrees of freedom $v \geq 1$

Mean: 0

Variance: $\frac{v}{v-2}$ if $v > 2$

Triangular Distribution

Range of X: $a \leq X \leq b$

Common use: often used as a rough model in the absence of data.

PDF:
$$f(x) = \frac{2(x-a)}{(b-a)(c-a)}, \; x \leq c$$
$$f(x) = \frac{2(b-x)}{(b-a)(b-c)}, \; x \geq c$$

Parameters: lower limit a, mode $c \geq a$, upper limit $b \geq c$

Mean: $\frac{a+b+c}{3}$

Variance: $\frac{a^2+b^2+c^2-ab-ac-bc}{18}$

U Distribution

Range of X: $b - a \leq X \leq b + a$

Common use: used in metrology for the distribution of quantities that oscillate around a specific value.

PDF: $f(x) = \frac{1}{\pi a\sqrt{1-\left(\frac{x-b}{a}\right)^2}}$

Parameters: scale $a > 0$, mean b

Mean: b

Variance: $\frac{a^2}{2}$

Uniform Distribution

Range of X: $a \leq X \leq b$

Common use: model for variable with equal probability everywhere over an interval.

PDF: $f(x) = \dfrac{1}{b-a}$

Parameters: lower limit a, upper limit b≥a

Mean: $\dfrac{a+b}{2}$

Variance: $\dfrac{(b-a)^2}{12}$

Weibull Distribution

Range of X: X ≥ 0

Common use: widely used in reliability analysis to model product lifetimes.

PDF: $f(x) = \dfrac{\alpha}{\beta^{\alpha}} x^{\alpha-1} e^{-(x/\beta)^{\alpha}}$

Parameters: shape α > 0, scale β > 0

Mean: $\dfrac{\beta}{\alpha}\Gamma\left(\dfrac{1}{\alpha}\right)$

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

Weibull Distribution (3-parameter)

Range of X: X ≥ θ

Common use: used for product lifetimes with a lower bound.

PDF: $f(x) = \dfrac{\alpha}{\beta^{\alpha}} (x-\theta)^{\alpha-1} \exp\left[-\left(x-\theta\right)/\beta\right]^{\alpha}$

Parameters: shape α > 0, scale β > 0, threshold θ

Mean: $\theta + \dfrac{\beta}{\alpha}\Gamma\left(\dfrac{1}{\alpha}\right)$

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