Continuous Distributions

Statistical distributions fall into two categories: modelling distributions and sampling distributions. Modelling distributions are used to describe data sets and are divided into two classes: **Continuous distributions** and **Discrete distributions**. Sampling distributions are used to construct confidence intervals and test hypotheses. In this section we shall discuss the Weibull, Normal, Lognormal, and Exponential distributions.

- Normal Distribution
- Lognormal Distribution
- Weibull Distribution
- Exponential Distribution

Normal Distribution

Whenever several random variables are added together, the resulting sum tends to normal regardless of the distribution of the variables being added. Mathematically, if

$$y = x_1 + x_2 + x_3 + \ldots + x_n$$

then the distribution of y becomes normal as n increases. If the random variables being summed are independent, the mean and variance of y are:

$$\mu_y = \mu_{x1} + \mu_{x2} + \mu_{x3} + \ldots + \mu_{xn}$$
$$\sigma_y^2 = \sigma_{x1}^2 + \sigma_{x2}^2 + \sigma_{x3}^2 + \ldots + \sigma_{xn}^2$$

When several random variables are averaged, the resulting average tends to normal regardless of the distribution of the variables being averaged. Mathematically, if

$$y = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

then the distribution of y becomes normal as n increases. If the random variables being averaged have the same mean and variance then the mean of y is equal to the mean of the individual variables being averaged, and the variance of y is:

$$\sigma_y^2 = \frac{\sigma^2}{n}$$

where σ^2 is the variance of the individual variables being averaged.

The tendency of sums and averages to become normally distributed as the number of variables being summed or averaged becomes large is known as the *Central Limit Theorem* or the *Theory of Large Numbers*. For distributions with little skewness, summing or averaging as few as 3 or 4 variables will result in a normal distribution. For highly skewed distributions, more than 30 variables may have to be summed or averaged to obtain a normal distribution.

Interactive Example http://www.engineeredsoftware.com/iqf/normal_ie1.asp

Let us consider one numeric example on this:

Let us consider a simple example with three random variables: X1, X2, and X3. Each variable represents the score obtained by a student in three different exams. Let us assume that X1 follows a uniform distribution between 0 and 50, X2 follows an exponential distribution with a mean of 30, and X3 follows a normal distribution with a mean of 70 and a standard deviation of 10.

We can simulate a scenario where 100 students' scores are generated for each variable and then calculate the sum of the scores for each student. Here is a Python code snippet to illustrate this:

```
import numpy as np
np.random.seed(42)

num_students = 100

X1 = np.random.uniform(0, 50, num_students)
X2 = np.random.exponential(30, num_students)
X3 = np.random.normal(70, 10, num_students)

sum_of_scores = X1 + X2 + X3

print("Sum of scores:", sum_of_scores)
print("Mean:", np.mean(sum_of_scores))
print("Standard deviation:", np.std(sum_of_scores))
```

Running this code will generate the sum of scores for each student. Even though the individual scores (X1, X2, X3) have different distributions, the sum of scores tends to follow a normal distribution due to the central limit theorem.

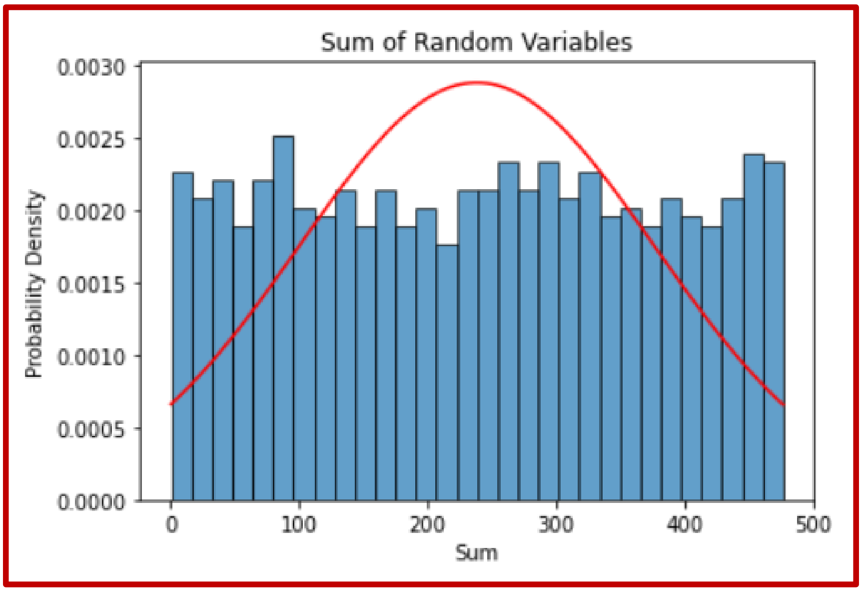
When you run the code, you will obtain different results due to the random nature of the variables. However, on average, you will observe that the mean and standard deviation of the sum of scores approximate a normal distribution.

The normal probability density function is:

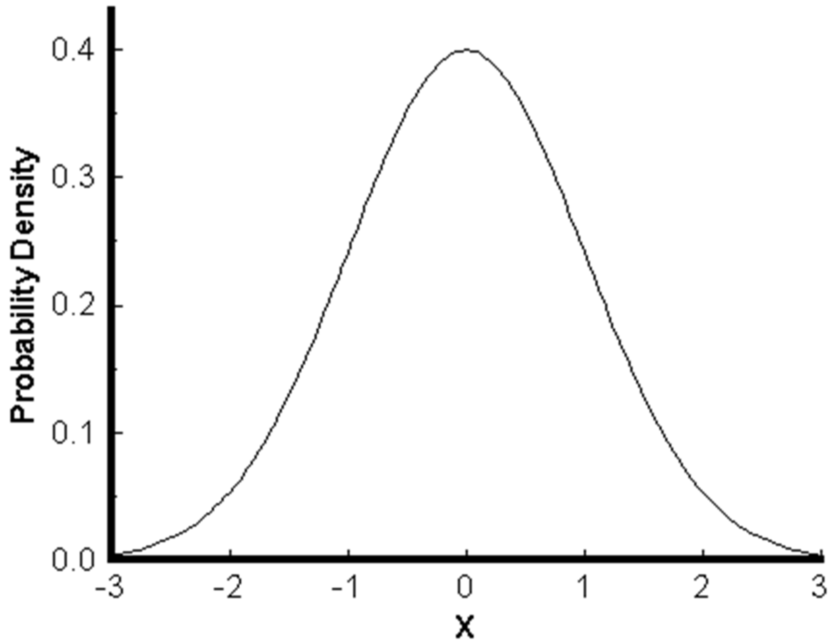
$$f(x) = \frac{1}{\sigma\sqrt{2\Pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < x < \infty$$

where μ is the mean and σ is the standard deviation.

```
# consider the following numerical example:
import numpy as np
import matplotlib.pyplot as plt

# Number of random variables to add
num_variables = 1000

# Generate random variables from a uniform distribution
variables = np.random.uniform(0, 1, (num_variables,))
# print (variables[:10], len(variables))
# Calculate the cumulative sum
cumulative_sum = np.cumsum(variables)
# print (cumulative_sum[:10], len(cumulative_sum), min(cumulative_sum),
max(cumulative_sum))
# Plot the distribution graph
plt.hist(cumulative_sum, bins=30, density=True, edgecolor='black', alpha=0.7)

# Plot a normal distribution curve for comparison
x = np.linspace(np.min(cumulative_sum), np.max(cumulative_sum), 100)
mean = np.mean(cumulative_sum)
std = np.std(cumulative_sum)
y = 1 / (std * np.sqrt(2 * np.pi)) * np.exp(-(x - mean)**2 / (2 * std**2))
plt.plot(x, y, color='red')

# Set plot labels and title
plt.xlabel('Sum')
plt.ylabel('Probability Density')
plt.title('Sum of Random Variables')

# Display the plot
plt.show()
```



The normal probability density function is not skewed and is shown in the figure below.



The density function shown in the figure above is the *standard normal probability density function*. The standard normal probability density function has a mean of 0 and a standard deviation of 1. The normal probability density function cannot be integrated implicitly. Because of this, historically, a transformation to the standard normal distribution is made, and the normal cumulative distribution function or reliability function is read from a table. Of course, with computers readily available, transforming a variable and looking-up values in a table is a waste of time. To obtain the area under the normal probability density function to the left of x (the normal cumulative distribution function) use the expression =Normdist(x,μ,σ,1) in Microsoft Excel. If x is a normal random variable, it can be transformed to standard normal using the expression:

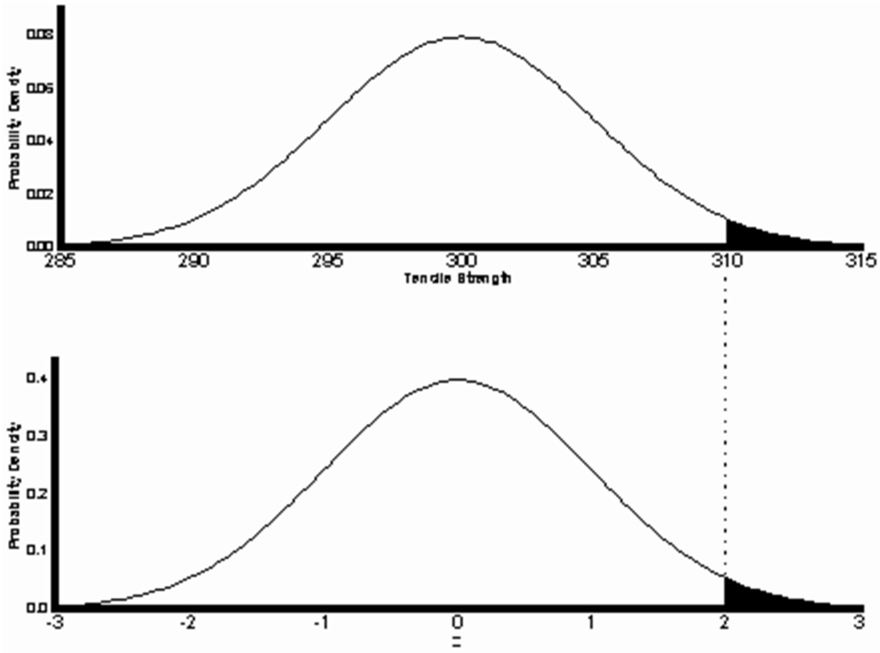
$$z = \frac{x - \mu}{\sigma}$$

Example: The tensile strength of a metal extrusion is normally distributed with a mean 300 and a standard deviation of 5. What percentage of extrusions have a strength greater than 310? What percentage of extrusions have a strength less than 295? What percentage of extrusions have a strength between 295 and 310?

Solution: The shaded area in the graph in the figure below represents the probability of an extrusion being greater than 310. The shaded area in the bottom graph represents the area under the standard normal distribution to the right of z = 2, which is the same as the probability of an extrusion being greater than 310. Transforming to standard normal,

$$z = \frac{x - \mu}{\sigma}$$

$z = (310-300)/5 = 2$



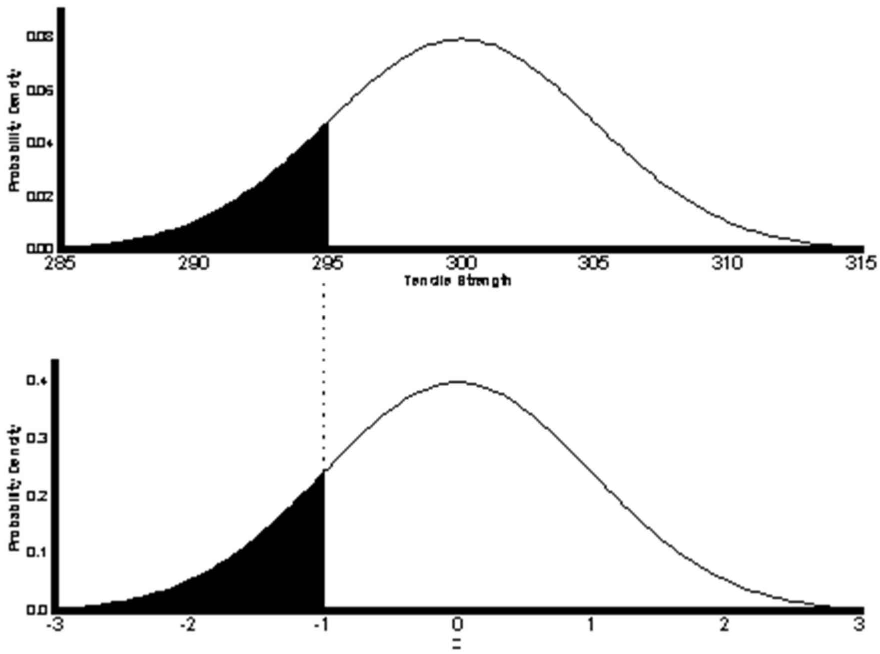
To determine the area under the standard normal probability density function to the right of $z = 2$, lookup $z = 2$ in a standard normal table, which is 0.0228. The percentage of extrusions with strength greater than 310 is 0.0228. This solution can also be found with Microsoft Excel using the expression:

=1-NORMDIST(310,300,5,1)

The shaded area in the top graph in the figure below represents the probability of an extrusion being less than 295. The shaded area in the bottom graph in the figure below represents the area under the standard normal distribution to the left of z = -1, which is the same as the probability of an extrusion being less than 295. Transforming to standard normal,

$$z = \frac{x - \mu}{\sigma}$$

$z = (295-300)/5 = -1$



From a standard normal table, the area to the left of $z = -1$ is equal to 0.1587. Using Microsoft Excel, this value is found from the expression =Normdist(295,300,5,1). The probability of the strength being between 295 and 310 is 1-0.0228-0.1587 = 0.8185.

Detailed calculations are available in the Microsoft Excel file "**Chapt-2 Continuous Distributions.xlsx**".

Example: A type of battery is produced with an average voltage of 60 with a standard deviation of 4 volts. If 9 batteries are selected at random, what is the probability that the total voltage of the 9 batteries is greater than 530? What is the probability that the average voltage of the 9 batteries is less than 62?

Solution: The expected total of the voltage of nine batteries is 540. The expected standard deviation of the voltage of the total of nine batteries is:

$$\sigma = \frac{(9)(4)}{\sqrt{9}} = 12$$

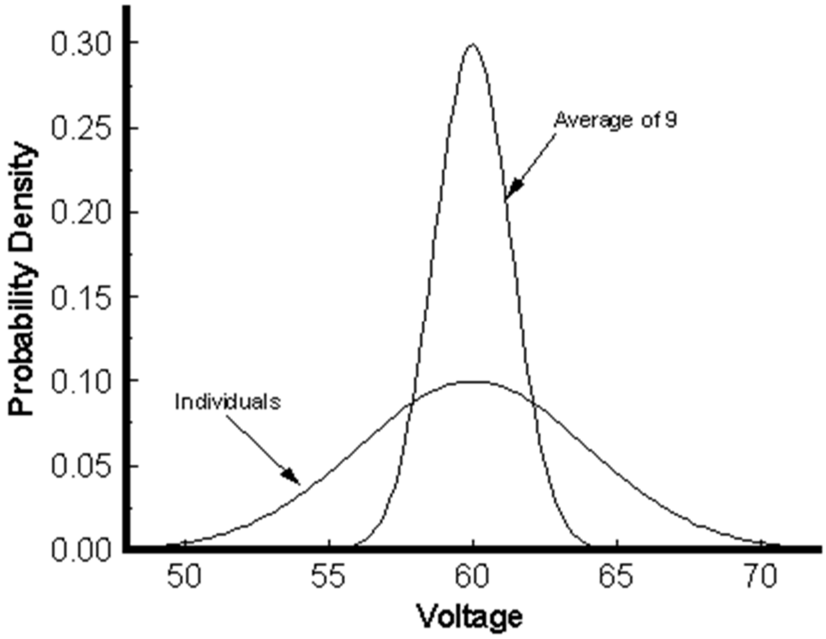
Transforming to standard normal,

$$z = \frac{530 - 540}{12} = -0.833 \qquad z = \frac{x - \mu}{\sigma}$$

From a standard normal table, the area to the left of $z = -0.833$ is 0.2024. The area to the right of $z = -0.833$ is 1-0.2024=0.7976. This problem can also be solved using the expression =1-Normdist(530,540,12,1) in Microsoft Excel.

The probability density function of the voltage of the individual batteries and of the average of nine batteries is shown in the figure below. The distribution of the averages has less variance because the standard deviation of the averages is equal to the standard deviation of the individuals divided by the square root of the sample size.

$$\sigma = \frac{4}{\sqrt{9}} = 1.333$$



The probability of the average voltage of 9 batteries being less than 62 is equal to the probability of a standard normal variable being less than $z = 1.5$; this is shown in the expression below.

$$z = \frac{62 - 60}{1.333} = 1.5 \qquad z = \frac{x - \mu}{\sigma}$$

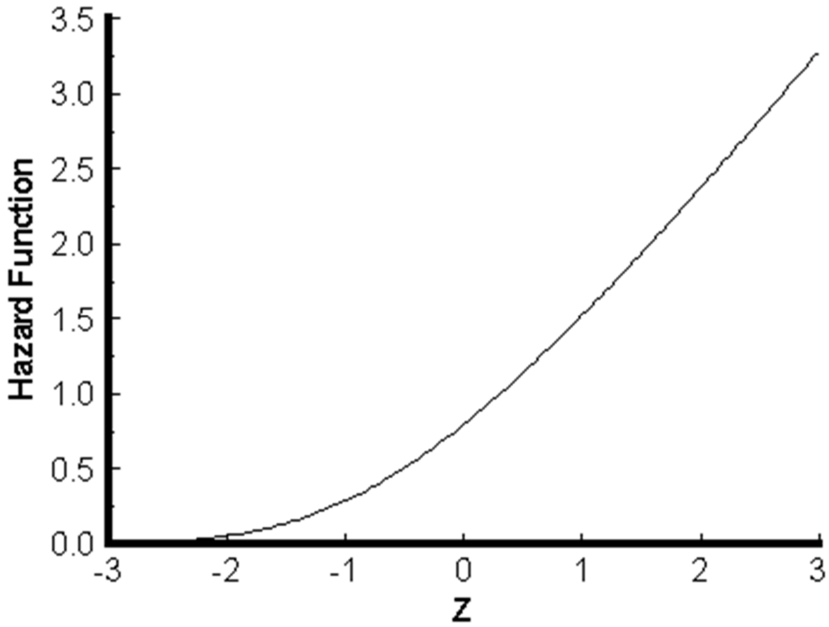
From a standard normal table, the area under the standard normal curve to the left of z = 1.6 is 0.9452, which is the probability of the average voltage of nine batteries being less than 62. In Microsoft Excel, this value is found from the expression =Normdist(62,60,1.25,1).

Detailed calculations are available in the Microsoft Excel file "**Chapt-2 Continuous Distributions.xlsx**".

The normal hazard function is:

$$h(x) = \frac{\phi\left(\frac{x-\mu}{\sigma}\right)}{\sigma\left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]}$$

where $\phi(x)$ is the standard normal probability density function and $\Phi(x)$ is the standard normal cumulative distribution function. The hazard function for the normal distribution is monotonically increasing; this is shown in the figure below.



Lognormal Distribution

If a data set is known to follow a lognormal distribution, transforming the data by taking a logarithm yields a data set that is normally distributed. This is shown in the table below.

Transformation of lognormal data.	
Lognormal	Normal
12	ln(12)
16	ln(16)
28	ln(28)
48	ln(48)
87	ln(87)
143	ln(143)

The most common transformation is made by taking the natural logarithm, but any base logarithm, such as base 10 or base 2, also yields a normal distribution. The remaining discussion will use the natural logarithm denoted as "ln".

When random variables are summed, as the sample size increases, the distribution of the sum becomes a normal distribution regardless of the distribution of the individuals. Since lognormal random variables are transformed to normal random variables by taking the logarithm, when random variables are multiplied, as the sample size increases, the distribution of the product becomes a lognormal distribution regardless of the distribution of the individuals. This is because the logarithm of the product of several variables is equal to the sum of the logarithms of the individuals. This is shown below.

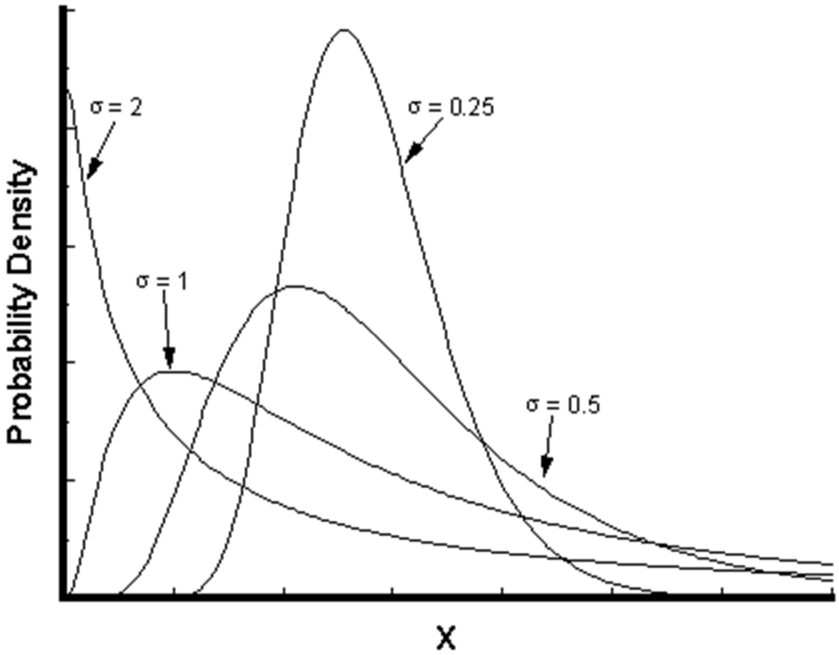
$y = x_1 x_2 x_3$
$\ln y = \ln x_1 + \ln x_2 + \ln x_3$

The lognormal probability density function is:

$$f(x) = \frac{1}{x\sigma\sqrt{2\Pi}} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2}, x > 0$$

where μ is the location parameter or log mean, and σ is the scale parameter or log standard deviation. The location parameter is the mean of the data set after transformation by taking the logarithm, and the scale parameter is the standard deviation of the data set after transformation.

The lognormal distribution takes on several shapes depending on the value of the shape parameter. The lognormal distribution is skewed right, and the skewness increases as the value of σ increases. This is shown in the figure below.



http://www.engineeredsoftware.com/iqf/lognormal_ie1.asp

The lognormal cumulative distribution and reliability functions are:

$$F(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$$
$$R(x) = 1 - \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$$

where $\Phi(x)$ is the standard normal cumulative distribution function.

Example: The data below is the time to fail for 4 light bulbs and is known to have a lognormal distribution. What is the reliability at 100 hours?

115 hours 155 hours
183 hours 217 hours

Solution: The table below shows this data and the transformation to normal.

Transformation to normal	
Time to Fail	ln(Time to Fail)
115	4.7449
155	5.0434
183	5.2095
217	5.3799

The parameters of the lognormal distribution are found by computing the mean and standard deviation of the transformed data in the second column of the table above. The mean of the transformed data is

$$\mu = \frac{4.7449 + 5.0434 + 5.2095 + 5.3799}{4} = 5.0944$$

The sum of the second column of the table above is 20.3777. If each value in the second column of the table above is squared, the sum of the squared values is 104.0326. Using these values, the sample standard deviation of the values in the second column of the table above is

$$s = \sqrt{\frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x\right)^2}{n(n-1)}} = \sqrt{\frac{4(104.0326) - 20.3777^2}{4(4-1)}} = 0.2708$$

An easier method of computing the standard deviation is to use the Microsoft Excel function =stdev().

The reliability at 100 hours is

$$R(100) = 1 - \Phi\left(\frac{\ln(100) - 5.0944}{0.2708}\right) = 1 - \Phi(-1.807)$$

From a standard normal table, in Microsoft Excel, the standard normal cumulative distribution function at z = -1.807 is 0.0354, thus the reliability at 100 hours is

R(100) = 1-0.0354 = 0.9646

The reliability at 100 hours can also be computed using the Microsoft Excel function:

=1-NORMDIST(LN(100),5.0944,0.2708,1)

Detailed calculations are available in the Microsoft Excel file "**Chapt-2 Continuous Distributions.xlsx**".

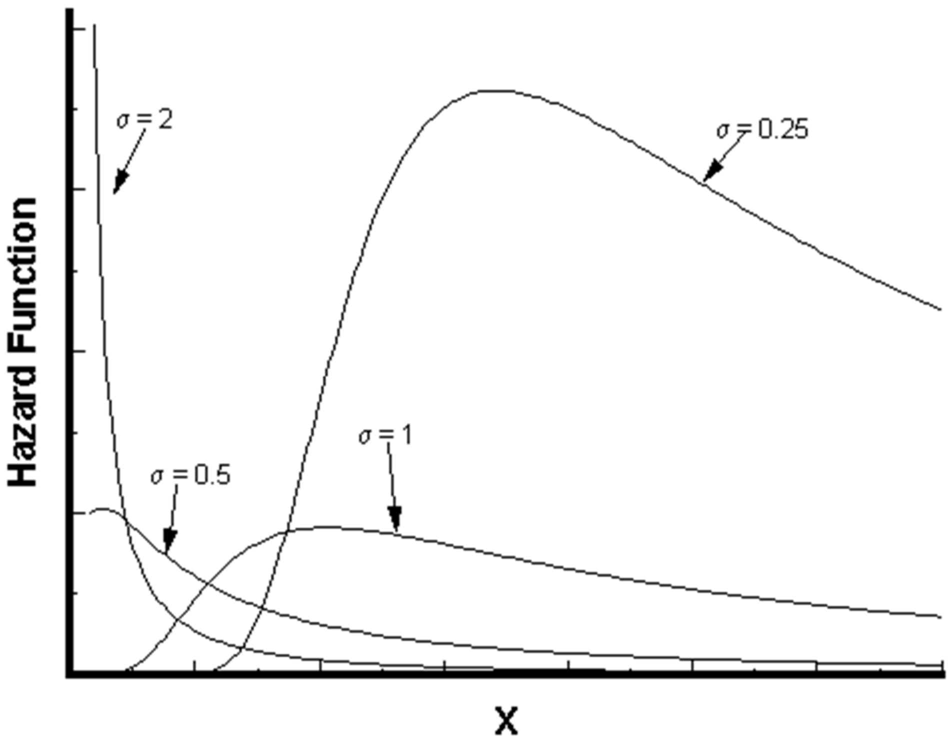
The location parameter, or log mean, is often mistaken for the mean of the lognormal distribution. The mean of the lognormal distribution can be computed from its parameters:

$$\textbf{mean} = e^{\left(\mu+\sigma^2/2\right)}$$

The variance of the lognormal distribution is:

$$\textbf{variance} = \left(e^{\left(2\mu+\sigma^2\right)}\right)\left(e^{\sigma^2}-1\right)$$

The lognormal hazard function has a unique behaviour; it increases initially, then decreases and eventually approaches zero. This means that items with a lognormal distribution have a higher chance of failing as they age for some period of time, but after survival to a specific age, the probability of failure decreases as time increases. The lognormal hazard function is shown in the figure below.



In the figure above, when σ = 2, the hazard function increases so quickly that it cannot be seen on the graph.

Interactive Example http://www.engineeredsoftware.com/iqf/lognormal_ie2.asp

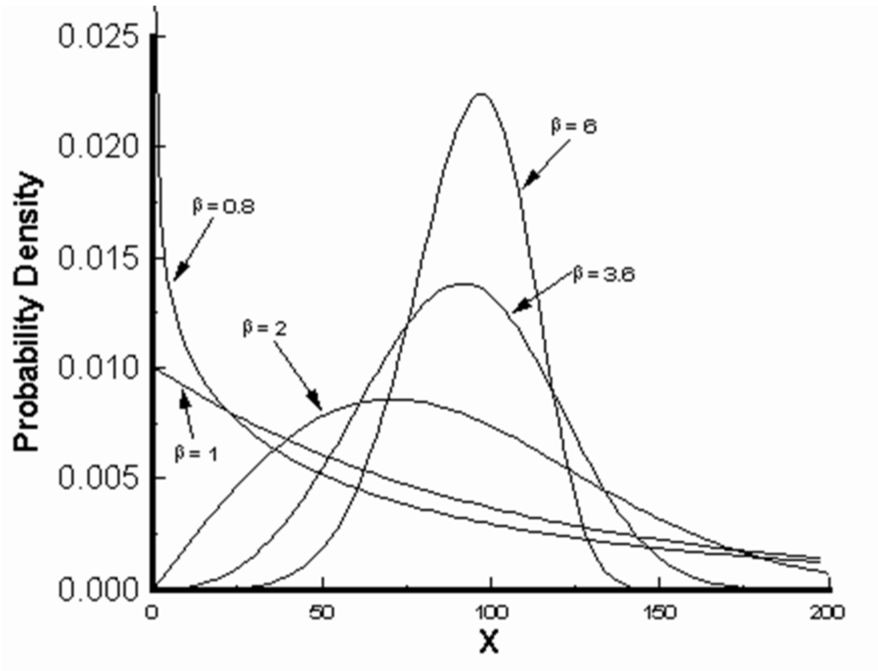
Weibull Distribution

The Weibull distribution is one of the most commonly used distributions in reliability. It is commonly used to model time to fail, time to repair and material strength. The Weibull probability density function is:

$$f(x) = \left(\frac{\beta}{\theta}\left(\frac{x-\delta}{\theta}\right)^{(\beta-1)}\right)\left(e^{-\left[\left(\frac{x-\delta}{\theta}\right)\right]^{\beta}}\right), x \geq \delta$$

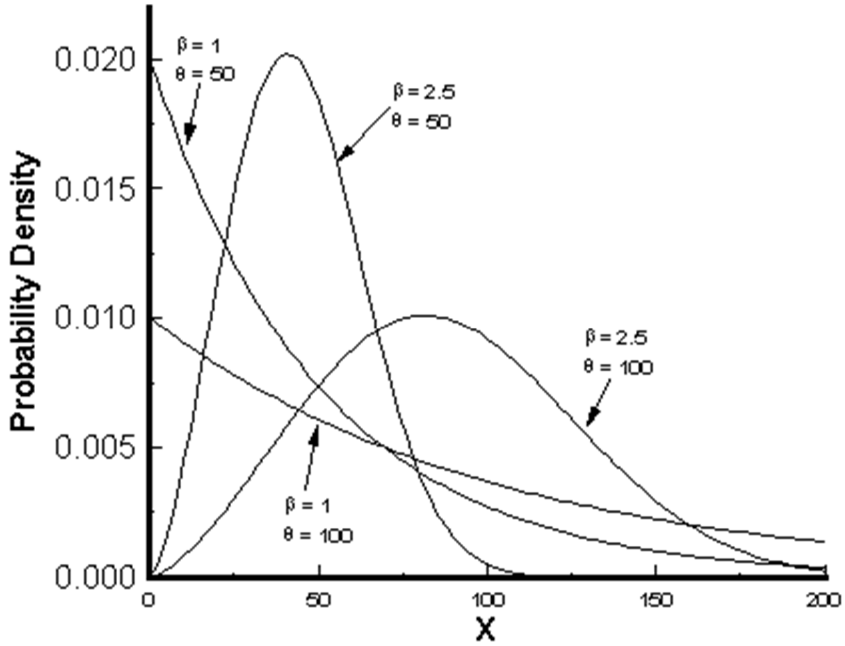
where β is the shape parameter,
θ is the scale parameter, and
δ is the location parameter.

The shape parameter is what gives the Weibull distribution its flexibility. By changing the value of the shape parameter, the Weibull distribution can model a wide variety of data. If $\beta = 1$ the Weibull distribution is identical to the exponential distribution, if $\beta = 2$, the Weibull distribution is identical to the Rayleigh distribution; if β is between 3 and 4 the Weibull distribution approximates the normal distribution. The Weibull distribution approximates the lognormal distribution for several values of β. For most populations more than fifty samples are required to differentiate between the Weibull and lognormal distributions. A sample of the Weibull distribution's flexibility is shown in the figure below with $\theta = 100$ and $\delta = 0$.

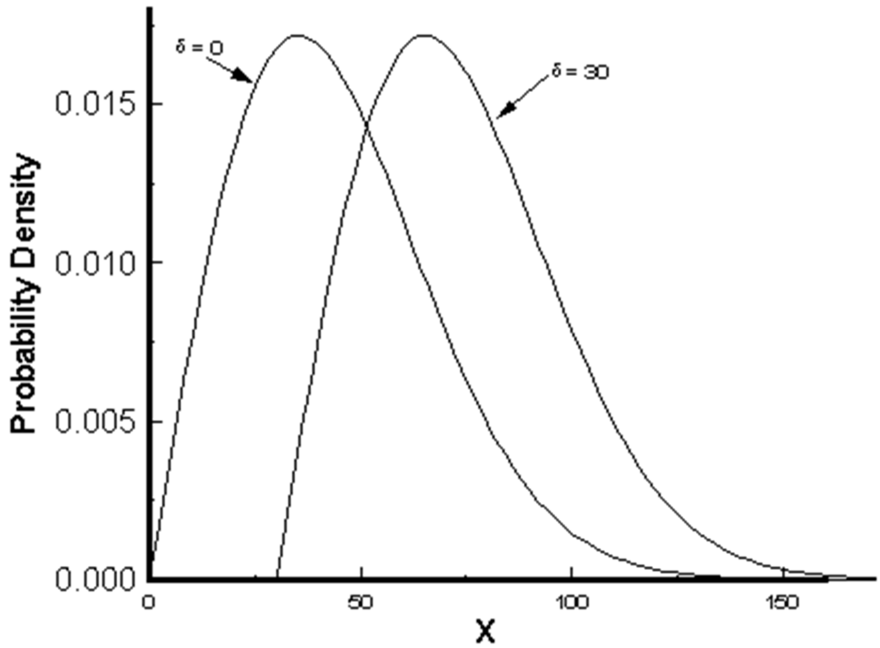


Interactive Example http://www.engineeredsoftware.com/iqf/weibull_ie1.asp

The scale parameter determines the range of the distribution. The scale parameter is also known as the *characteristic life* if the location parameter is equal to zero. If δ does not equal zero, the characteristic life is equal to θ+δ; 63.2% of all values fall below the characteristic life regardless of the value of the shape parameter. The effect of the scale parameter of the probability density function is shown in the figure below.



The location parameter is used to define a failure-free zone. The probability of failure when *x* is less than δ is zero. When δ>0, there is a period when no failures can occur. When δ<0, failures have occurred before time equals 0. At first this seems ridiculous, but a negative location parameter is caused by shipping failed units, failures during transportation, and shelf-life failures. Generally, the location parameter is assumed to be zero. The effect of the location parameter is shown in the figure below.

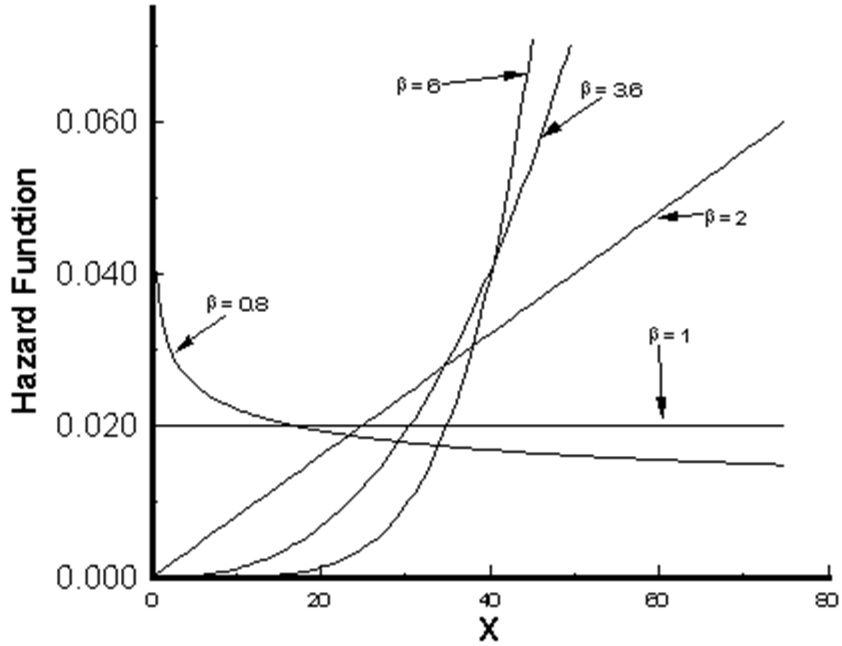


Interactive Example **http://www.engineeredsoftware.com/iqf/weibull_ie2.asp**

The Weibull hazard function is determined by the value of the shape parameter.

$$h(x) = \frac{\beta}{\theta}\left(\frac{x-\delta}{\theta}\right)^{(\beta-1)}$$

When β<1 the hazard function is decreasing; this is known as the infant mortality period. When β=1, the failure rate is constant. When β>1 the failure rate is increasing; this is known as the wear out period. The Weibull hazard function is shown in the figure below.



Interactive Example http://www.engineeredsoftware.com/iqf/weibull_ie3.asp

The Weibull reliability and cumulative distribution functions are:

$$R(x) = e^{-\left[\left(\frac{x-\delta}{\theta}\right)\right]^{\beta}}$$

$$F(x) = 1 - e^{-\left[\left(\frac{x-\delta}{\theta}\right)\right]^{\beta}}$$

Example: The time to fail for a flexible membrane follows the Weibull distribution with β = 2 and θ = 300 months. What is the reliability at 200 months? After how many months is 90% reliability achieved?

Solution: After 200 months, the reliability of the flexible membrane is

$$R(200) = e^{\left[-\left(\frac{200}{300}\right)^{2}\right]} = 0.6412$$

By manipulating the expression for reliability, 90% reliability is achieved after:

$t = \theta(-\ln R)^{(1/\beta)}$

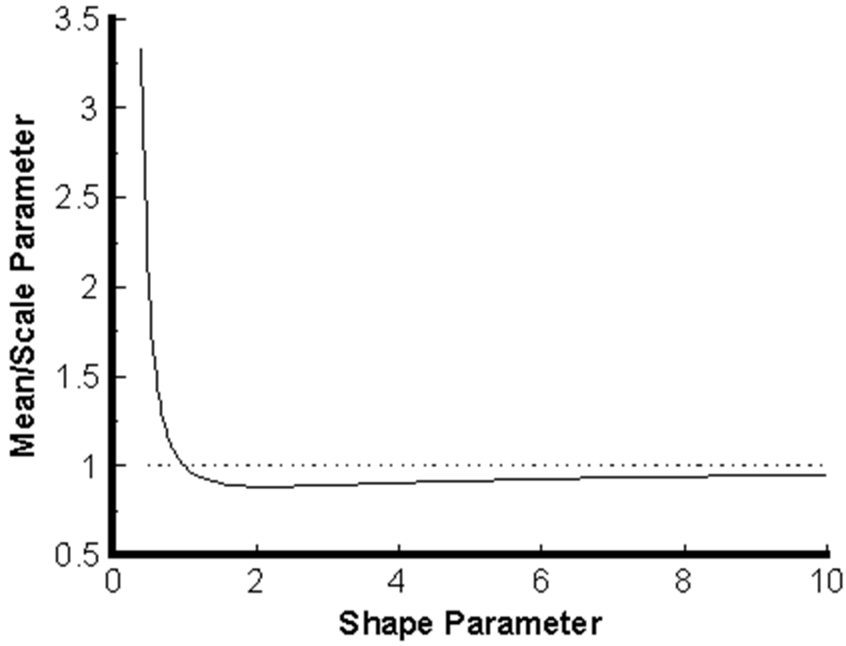
$t = 300\sqrt{-\ln(0.9)} = 97.38$ months

Detailed calculations are available in the Microsoft Excel file "**Chapt-2 Continuous Distributions.xlsx**".

The mean and variance of the Weibull distribution are computed using the gamma distribution which is available in Microsoft Excel. The mean of the Weibull distribution is:

$$\mu = \theta\Gamma\left(1+\frac{1}{\beta}\right)$$

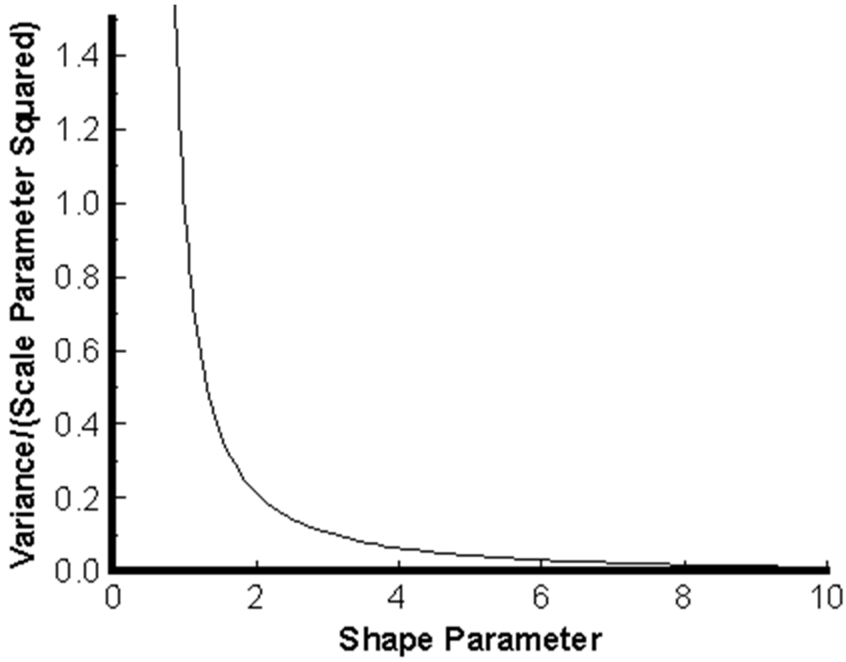
The mean of the Weibull distribution is equal to the characteristic life if the shape parameter is equal to one. The mean as a function of the shape parameter is shown in the figure below.



The variance of the Weibull distribution is

$$\sigma^2 = \theta^2\left[\Gamma\left(1+\frac{2}{\beta}\right) - \Gamma^2\left(1+\frac{1}{\beta}\right)\right]$$

The variance of the Weibull distribution decreases as the value of the shape parameter increases. This is shown in the figure below.



The reliability measure of many components is the mean time to fail. Consumer electronics companies often advertise the mean time to fail of the products. The mean time to fail is a deceptive measure because the variance of the time to fail distribution is not considered. To achieve the same reliability with a larger variance, requires a larger mean.

Consider two components: A and B. Component A has a mean time to fail of 4645 hours, and component B has a mean time to fail of 300 hours. If both components sell for the same price, which component should be used to maximize reliability at 100 hours?

This question cannot be answered without knowing more information about the distribution of the time to fail. Component A has a mean of 4645 hours and a Weibull time to fail distribution with a shape parameter of 0.8. Using the mean and the shape parameter, the scale parameter of component A can be computed to be 4100 hours. The reliability at 100 hours is:

$$R(100) = e^{-\left(\frac{100}{4100}\right)^{0.8}} = 0.95$$

Component B has a mean of 300 hours and a Weibull time to fail distribution with a shape parameter of 3. Using the mean and the shape parameter, the scale parameter of component B can be computed to be 336 hours. The reliability at 100 hours is:

$$R(100) = e^{-\left(\frac{100}{336}\right)^{3}} = 0.974$$

Although the mean of component A is more than 10 times as large as the mean of component B, the reliability of component B is greater than the reliability of component A at 100 hours. Continuing with this example, if the reliability at 1000 hours is to be maximized, component A has a reliability of 0.723 and component B has a reliability of approximately zero.

Interactive Example http://www.engineeredsoftware.com/iqf/weibull_ie4.asp

Exponential Distribution

The exponential distribution is used to model items with a constant failure rate, usually electronics. The exponential distribution is closely related to the Poisson distribution. If a random variable, x, is exponentially distributed, then the reciprocal of x, $y=1/x$ follows a Poisson distribution. Likewise, if x is Poisson distributed, then $y=1/x$ is exponentially distributed. Because of this behaviour, the exponential distribution is usually used to model the mean time between occurrences, such as arrivals or failures, and the Poisson distribution is used to model occurrences per interval, such as arrivals, failures, or defects.

The exponential probability density function is:

$$f(x) = \lambda e^{-\lambda x}, x > 0$$

The exponential probability density is also written as

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, x > 0$$

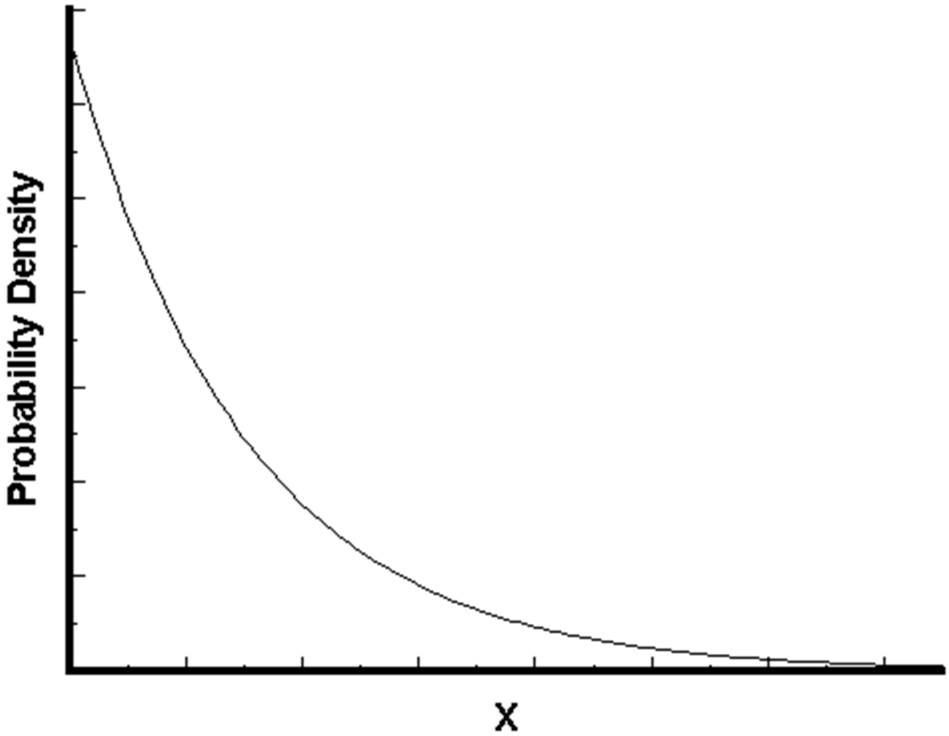
where θ is the mean.

From the equations above, it can be seen that $\lambda=1/\theta$. The variance of the exponential distribution is equal to the mean squared.

$$\sigma^2 = \theta^2$$

$$\sigma^2 = \frac{1}{\lambda^2}$$

The exponential probability density function is shown in the figure below.



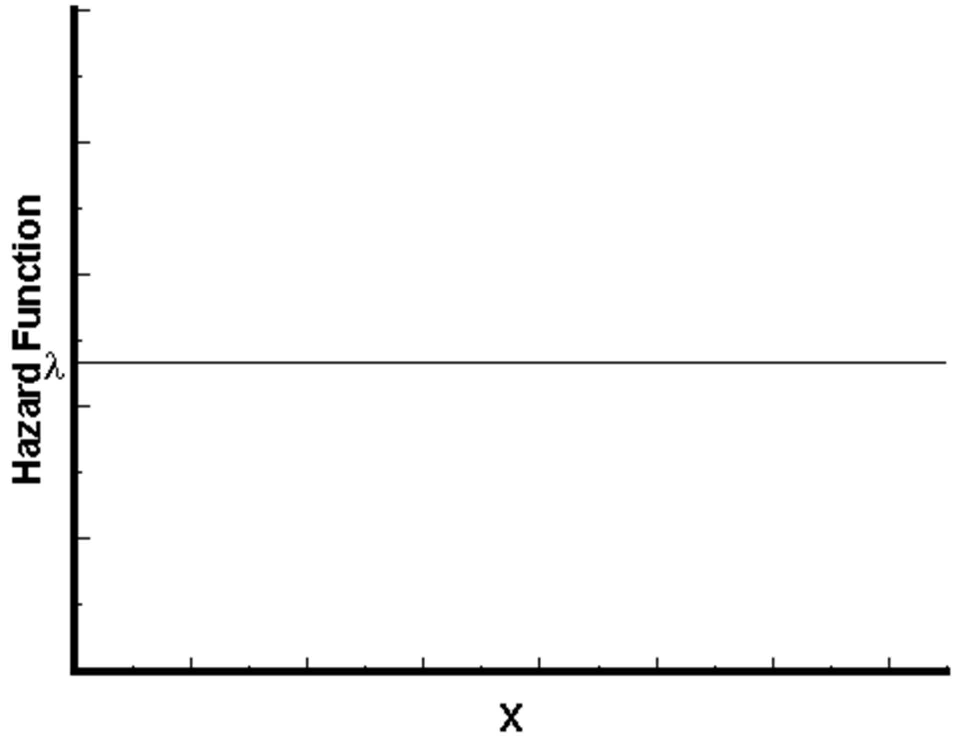
The exponential reliability function is:

$$R(x) = e^{-\frac{x}{\theta}}, x > 0$$
$$R(x) = e^{-\lambda x}, x > 0$$

The exponential hazard function is:

$$h(x) = \frac{1}{\theta}$$
$$h(x) = \lambda$$

The exponential hazard function is shown in figure the figure below.



Example: A resistor has a constant failure rate of 0.04 per hour. What is the resistor's reliability at 100 hours? If 100 resistors are tested, how many would be expected to be in a failed state after 25 hours?

Solution: The reliability at 100 hours is

$$R(100) = e^{-0.04(100)} = 0.0183$$

The probability of failing before 25 hours is given by the cumulative distribution function which is equal to one minus the reliability function.

$$F(25) = 1 - R(25) = 1 - e^{0.04(25)} = 1 - 0.368 = 0.632$$

The expected number of resistors in a failed state is 100(0.632)=63.2.

Detailed calculations are available in the Microsoft Excel file "**Chapt-2 Continuous Distributions.xlsx**".

The exponential distribution is characterized by its hazard function which is constant. Because of this, the exponential distribution exhibits a lack of memory. That is, the probability of a survival for a time interval, given survival to the beginning of the interval, is dependent ONLY on the length of the interval, and not on the time of the start of the interval. For example, consider an item that has a mean time to fail of 150 hours that is exponentially distributed. The probability of surviving through the interval 0 to 20 hours is:

$$R(20) = e^{-\frac{20}{150}} = 0.8751$$

The probability of surviving the interval 100 to 120 is equal to

$$R(120, \text{given survival to } t = 100) = \frac{R(120)}{R(100)} = \frac{e^{-\frac{120}{150}}}{e^{-\frac{100}{150}}} = \frac{0.4493}{0.5134} = 0.8751$$

Sampling Distributions

- t Distribution
- F Distribution
- Chi-Square Distribution

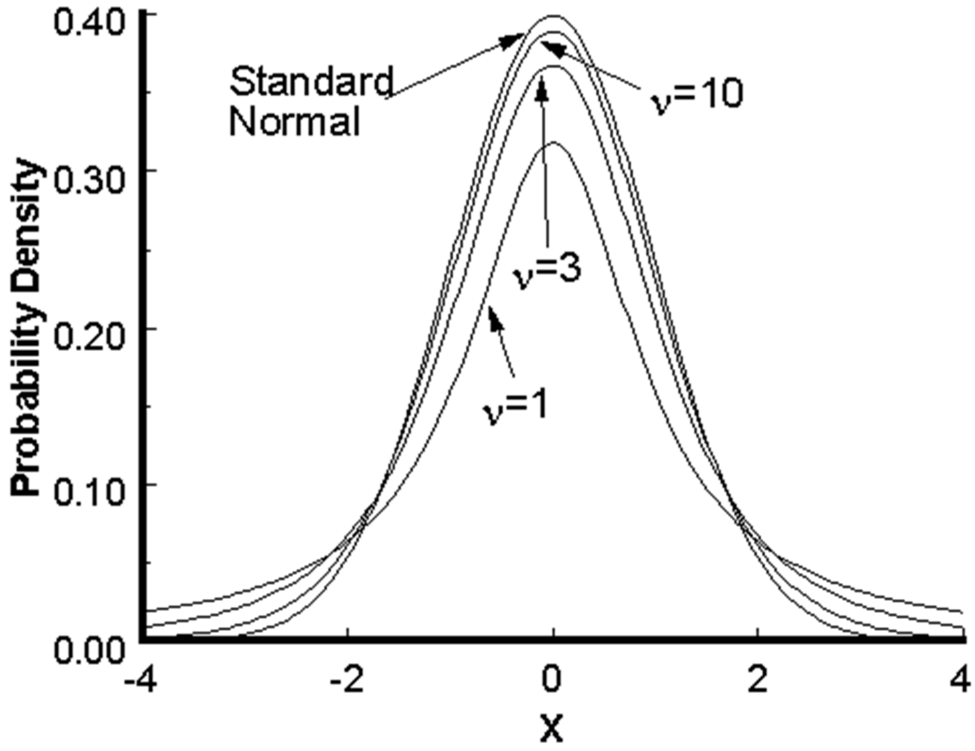
The *t, F* and Chi-square distributions are formed from combinations of random variables. Because of this, they are generally not used to model physical phenomena, like time to fail, but are used to make decisions and construct confidence intervals.

t Distribution

The t-distribution is formed by combining a standard normal random variable and a chi-square random variable. If *z* is a standard normal random variable, and χ^2 is a chi-square random variable with ν degrees of freedom, then a random variable with a t-distribution is obtained by dividing z by the square root of χ^2 divided by ν. The *t* probability density function is:

$$t = \frac{z}{\sqrt{\frac{\chi^2}{\nu}}}$$

The t-distribution is equivalent to the F-distribution with 1 and n degrees of freedom. The t-distribution is commonly used for hypothesis testing and constructing confidence intervals for means. It is used in place of the normal distribution when the standard deviation is unknown. The t-distribution compensates for the error in the estimated standard deviation. If the sample size is large, n>100, the error in the estimated standard deviation is small, and the t-distribution is approximately normal.



The mean and variance of the t-distribution are:

$\mu = 0$, and

$$\sigma^2 = \frac{\nu}{\nu - 2}, \nu \geq 3$$

From a random sample of *n* items, the probability that

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

falls between any two specified values is equal to the area under the *t*-probability density function between the corresponding values on the *x*-axis with *n*-1 degrees of freedom.

Example

The burst strength of 15 randomly selected seals is given below. What is the probability that the average burst strength of the population is greater than 500?

480	489	491	508	501
500	486	499	479	496
499	504	501	496	498

Solution

The mean of these 15 data points is 495.13. The sample standard deviation of these 15 data points is 8.467. The probability that the population mean is greater than 500 is equal to the area under the t probability density function, with 14 degrees of freedom, to the left of

$$t = \frac{495.13 - 500}{8.467/\sqrt{15}} = -2.227$$

The area under the *t* probability density function, with 14 degrees of freedom, to the left of -2.227 is 0.0215. This value may be found from standard statistical tables, but it is easier to use an electronic spreadsheet. The value can be computed directly in Excel using the function:

=TDIST(-2.227,14,1)

Simply stated, making an inference from the sample of 15 data points, there is a 2.15% that the true population mean is greater than 500.

F Distribution

If *X* is a chi square random variable with ν_1, degrees of freedom and *Y* is a chi-square random variable with ν_2 degrees of freedom and if *X* and *Y* are independent, then,

$$F = \frac{X/\nu_1}{Y/\nu_2}$$

is *F* distributed with ν_1 and ν_2 degrees of freedom.

The *F* distribution is used extensively to test for equality of variances from two normal populations.

Example

Find $F_{0.05}$ with $\nu_1 = 9$ and $\nu_2 = 10$ and $F_{0.95}$ with $\nu_1 = 10$ and $\nu_2 = 9$.

Solution

The following Excel functions provide the solutions.

=FINV(0.05,9,10)

=FINV(0.95,10,9)

Chi-Square Distribution

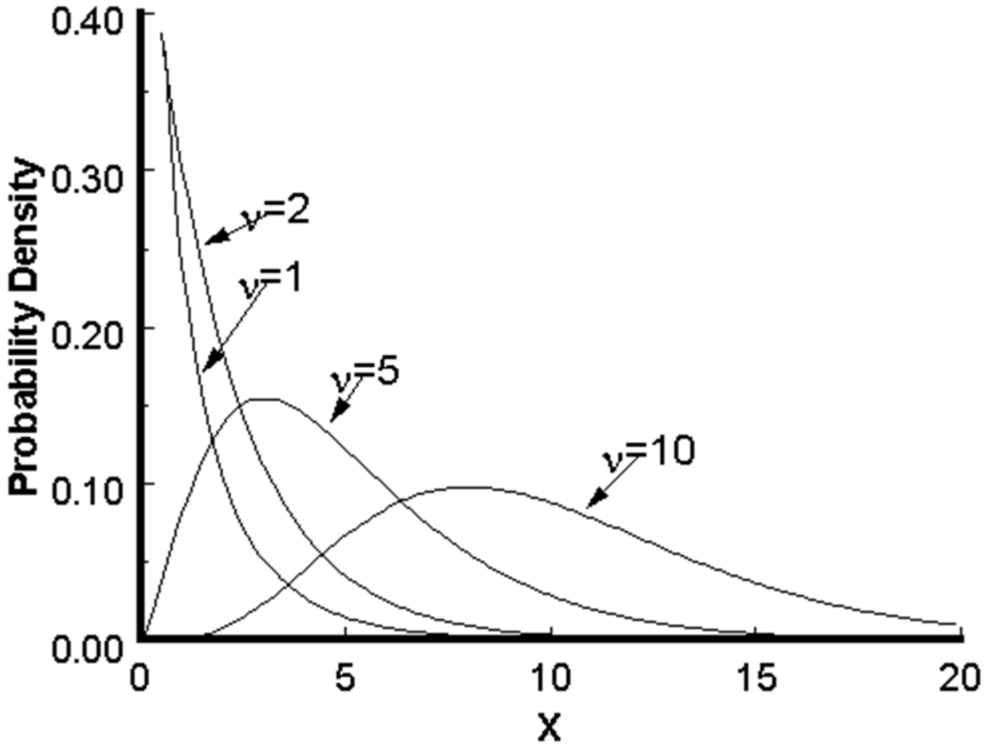
The chi-square distribution is formed by summing the square of standard normal random variables. For example, if *z* is a standard normal random variable, then:

$$y = z_1^2 + z_2^2 + z_3^2 + \ldots + z_n^2$$

is a chi-square random variable with n degrees of freedom.

A chi-square random variable is also created by summing two or more chi-square random variables. A distribution having this property is regenerative. The chi-square distribution is a special case of the gamma distribution with a failure rate of 2, and degrees of freedom equal to 2 divided by the number of degrees of freedom for the corresponding chi-square distribution.

The chi-square density function is shown below.



The critical value of the chi-square distribution is given standard statistical tables, but it is easier to use the Excel function:

=CHIINV(probability, degrees_freedom)