Failure rate (Updated and Adapted from Notes by Dr. A.K. Nema)

Part 1:

Failure rate is the frequency with which an engineered system or component fails, expressed for example in failures per hour. It is often denoted by the Greek letter λ (lambda) and is important in reliability theory. In practice, the closely related Mean Time Between Failures (MTBF) is more commonly expressed and used for high quality components or systems. Failure rate is usually time dependent, and an intuitive corollary is that the rate changes over time versus the expected life cycle of a system. For example, as an automobile grows older, the failure rate in its fifth year of service may be many times greater than its failure rate during its first year of service—one simply does not expect to replace an exhaust pipe, overhaul the brakes, or have major transmission problems in a new vehicle.

Mean Time Between Failures (MTBF) is closely related to Failure rate. In the special case when the likelihood of failure remains constant with respect to time (for example, in some product like a brick or protected steel beam), and ignoring the time to recover from failure, failure rate is simply the inverse of the Mean Time Between Failures (MTBF). MTBF is an important specification parameter in all aspects of high importance engineering design— such as naval architecture, aerospace engineering, automotive design, etc. —in short, any task where failure in a key part or of the whole of a system needs be minimized and severely curtailed, particularly where lives might be lost if such factors are not taken into account. These factors account for many safety and maintenance practices in engineering and industry practices and government regulations, such as how often certain inspections and overhauls are required on an aircraft. A similar ratio used in the transport industries, especially in railways and trucking is 'Mean Distance Between Failure', a variation, which attempts to correlate actual, loaded distances to similar reliability needs and practices. Failure rates and their projective manifestations are important factors in insurance, business, and regulation practices as well as fundamental to design of safe systems throughout a national or international economy.

Failure rate in the discrete sense

In words appearing in an experiment, the failure rate can be defined as, "The total number of failures within an item population, divided by the total time expended by that population, during a particular measurement interval under stated conditions." Here failure rate λ (t) can be thought of as the probability that a failure occurs in a specified interval, given no failure before time t. It can be defined with the aid of the reliability function or survival function R (t), the probability of no failure before time t, as:

$$\lambda = \frac{R(t_1) - R(t_2)}{(t_2 - t_1) \cdot R(t_1)} = \frac{R(t) - R(t + \triangle t)}{\triangle t \cdot R(t)}$$

Where, t1 (or t) and t2 are respectively the beginning and ending of a specified interval of time spanning Δt. Note that this is a conditional probability, hence the R (t) in the denominator.

Failure rate in the continuous sense

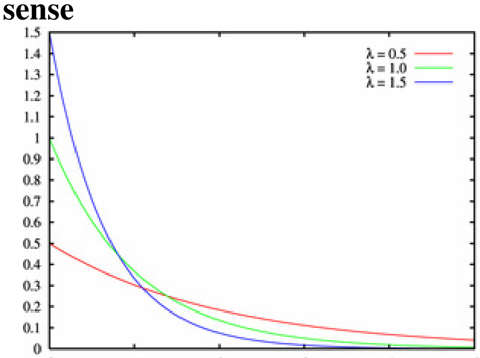


Figure. Exponential failure density functions [f(t)]

By calculating the failure rate for smaller and smaller intervals of time, the interval becomes infinitely small. This results in the hazard function, which is the instantaneous failure rate at any point in time:

$$h(t) = \lim_{\triangle t \to 0} \frac{R(t) - R(t + \triangle t)}{\triangle t \cdot R(t)}.$$

Continuous failure rate depends on a failure distribution, which is a cumulative distribution function that describes the probability of failure prior to time t,

$$P(\mathbf{T} \le t) = F(t) = 1 - R(t), \quad t \ge 0.$$

Where, T is the failure time. The failure distribution function is the integral of the failure density function, f(x),

$$F(t) = \int_0^t f(x)\,dx.$$

The hazard function can be defined now as:

$$h(t) = \frac{f(t)}{R(t)}.$$

Many probability distributions can be used to model the failure distribution. A common model is the exponential failure distribution,

$$F(t) = \int_0^t \lambda e^{-\lambda x}\,dx = 1 - e^{-\lambda t},$$

Which, is based on the exponential density function.

$$h(t) = \frac{f(t)}{R(t)} = \frac{\lambda e^{-\lambda x}}{e^{-\lambda x}} = \lambda$$

For an exponential failure distribution the hazard rate is a constant with respect to time (that is, the distribution is "memoryless"). For other distributions, such as a Weibull distribution or a log-normal distribution, the hazard function is not constant with respect to time. For some such as the deterministic distribution it is monotonic increasing (analogous to "wearing out"), for others such as the Pareto distribution it is monotonic decreasing (analogous to "burning in"), while for many it is not monotonic.

Part 2: Failure rate data

Failure rate data can be obtained in several ways. The most common means are:

1) Historical data about the device or system under consideration.

Many organizations maintain internal databases of failure information on the devices or systems that they produce, which can be used to calculate failure rates for those devices or systems. For new devices or systems, the historical data for similar devices or systems can serve as a useful estimate.

2) Government and commercial failure rate data.

Handbooks of failure rate data for various components are available from government and commercial sources. Several failure rate data sources are available commercially that focus on commercial components, including some non-electronic components.

3) Testing.

The most accurate source of data is to test samples of the actual devices or systems in order to generate failure data. This is often expensive, so that the previous data sources are often used instead.

Units

Failure rates can be expressed using any measure of time, but hours is the most common unit in practice. Other units, such as miles, revolutions, etc., can also be used in place of "time" units. Failure rates are often expressed in engineering notation as failures per million, especially for individual components, since their failure rates are often very low. The Failures In Time (FIT) rate of a device is the number of failures that can be expected in one billion (10^9) hours of operation. This term is used particularly by the semiconductor industry.

Additivity

Under certain engineering assumptions (e.g. besides the above assumptions for a constant failure rate, the assumption that the considered system has no relevant redundancies), the failure rate for a complex system is simply the sum of the individual failure rates of its components, as long as the units are consistent, e.g. failures per million hours. This permits testing of individual components or subsystems, whose failure rates are then added to obtain the total system failure rate.

Annualized failure rate (AFR) is the relation between the mean time between failure (MTBF) and the assumed hours that a device is run per year, expressed in percent.

Example1

Suppose it is desired to estimate the failure rate of a certain component. A test can be performed to estimate its failure rate. Ten identical components are each tested until they either fail or reach 1000 hours, at which time the test is terminated for that component. (The level of statistical confidence is not considered in this example.) The results are as follows:

Component	Hours	Failure
Component 1	1000	No failure
Component 2	1000	No failure
Component 3	467	Failed
Component 4	1000	No failure
Component 5	630	Failed
Component 6	590	Failed
Component 7	1000	No failure
Component 8	285	Failed
Component 9	648	Failed

Component 10	882	Failed
Totals	7502	6

Estimated failure rate is = $\frac{6 \text{ failures}}{7502 \text{ hours}} = 0.0007998 \frac{\text{failures}}{\text{hour}} = 799.8 \times 10^{-6} \frac{\text{failures}}{\text{hour}}$,
or 799.8 failures for every million hours of operation.

Example2

A disk drive's MTBF number may be 1,200,000 hours and the disk drive may be running 24 hours a day, seven days a week. One year has 8,760 hours.

$\frac{1,200,000 \, hours}{8760 \, hours/year} = 136.9863 \, years$; then take the reciprocal of 136.9863 years

$\frac{1 \, failure}{136.9863 \, years} \times 100\% = 0.73\%$ => You can expect about 0.73 percent of the population of these disk drives to fail in the average year.

Example 3.

A disk drive's MTBF number may be 700,000 hours and the disk drive may be running 2400 hours a year.

$\frac{700,000 \, hours}{2400 \, hours/year} = 291.6667 \, years$ then take the reciprocal of 291.6667 years

$\frac{1 \, failure}{291.6667 \, years} \times 100\% = 0.34\%$ => You can expect about 0.34 percent of the population of these disk drives to fail in the average year.

Now assuming you let the same disk run 24 hours a day, 7 days a week:

$\frac{700,000 \, hours}{8760 hours/year} = 79.9087 \, years$

$\frac{1 \, failure}{79.9087 \, years} \times 100\% = 1.25\%$, i.e., ~1.25% of the population of these disk drives may fail in the average year.

Part 3

Failure Distribution Types

Discrete distributions		**Continuous distributions**	
Binomial	Covered	Normal	Covered
Poisson distribution	Covered	Exponential	Covered
Multinomial distribution	Beyond the scope	lognormal	Covered
		Weibull distribution	Covered
		Extreme value distribution	Beyond the scope

1) Weibull Analysis

This family of distribution has two parameters:

$$R(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} \qquad (1)$$

The constant α is called the *scale parameter*, because it scales the t variable, and the constant β is called the *shape parameter*, because it determines the shape of the rate function. (Occasionally the variable t in the above definition is replaced with t–γ, where γ is a third parameter, used to define a suitable zero point.) If β in greater than 1 the rate increases with t, whereas if β is less than 1 the rate decreases with t. If β = 1 the rate is constant, in which case the Weibull distribution equals the exponential distribution.

Cumulative probability function [F(t)]:

$$F(t) = 1 - e^{-(t/\alpha)^{\beta}} \qquad (2)$$

Failure density distribution [f(t)]

$$f(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} e^{-(t/\alpha)^{\beta}} \qquad (3)$$

Application of Weibull Analysis for failure of components:

Suppose we have a population consisting of n widgets (for large n), all of which began operating continuously at the same time t = 0. (Note: “t” represents "calendar time", which is the hours of operation of each individual widget, not the sum total of the operational hours of the entire population; the latter would be given by (n×t). We can use a single t as the time variable because we have assumed a coherent population consisting of widgets that began operating at the same instant and accumulated hours continuously.) If each widget has a Weibull cumulative failure distribution given by equation (2) for some fixed parameters α and β,, then the expected number N(t) of failures by the time t is: [from Equation 2: F(t)=N(t)/n]

$$N(t) = \left(1 - e^{-(t/\alpha)^{\beta}}\right) n \qquad (4)$$

Dividing both sides by n, and re-arranging terms, this can be written in the form

$$1 - \frac{N(t)}{n} = e^{-(t/\alpha)^{\beta}}$$

Taking the natural log of both sides and negating both sides, we have

$$\ln\left(\frac{1}{1-\frac{N(t)}{n}}\right) = \left(\frac{t}{\alpha}\right)^{\beta}$$

Taking the natural log again, we arrive at

$$\ln\left(\ln\left(\frac{1}{1-\frac{N(t)}{n}}\right)\right) = \beta\ln(t) - \beta\ln(\alpha) \qquad (5)$$

Example 4.

Given an initial population of n = 100 widgets (at time t = 0), and accumulating hours continuously thereafter, suppose the first failure occurs at time $t = t_1$=> Approximately, we could say the expected number of failures at the time of the first failure is about 1, => $F(t_1) = N(t_1)/n = 1/100$. However, this isn't quite optimum, because statistically the first failure is most likely to occur slightly before the expected number of failures reaches 1. To understand why, consider a population consisting of just a single widget, in which case the expected number of failures at any given time t would be simply F(t), which only approaches 1 in the limit as t goes to infinity, and yet the median time of failure is at the value of $t = t_{median}$ such that $F(t_{median}) = 0.5$. In other words, the probability is 0.5 that the failure will occur prior to t_{median}, and 0.5 that it will occur later. Hence in a population of size n = 1 the expected number of failures at the median time of the first failure is just 0.5.

+++++++++++++++

In general, given a population of n widgets, each with the same failure density f(t), the probability for each individual widget being failed at time t_m is $F(t_m) = N(t_m)/n$. Denoting this value by ϕ, the probability that exactly j widgets are failed and n – j are not failed at time t_m is

$$P[j;n] = \binom{n}{j}\phi^j(1-\phi)^{n-j} \qquad \textbf{(6)}$$

It follows that the probability of *j or more* being failed at the time t_m is

$$P[\geq j;n] = \sum_{i=j}^{n}\binom{n}{i}\phi^i(1-\phi)^{n-i} \qquad \textbf{(7)}$$

This represents the probability that the jth (of n) failure has occurred by the time t_m, and of course the complement is the probability that the jth failure has not yet occurred by the time t_m. Therefore, given that the jth failure occurs at t_m, the "median" value of $F(t_m) = \phi$ is given by putting $P[\geq j;n] = 0.5$ in the above equation and solving for ϕ. This value is called the median rank, and can be computed numerically.

Approximate Formula

An alternative approach is to use the remarkably good approximate formula:

$$F(t_m) = \frac{j-0.3}{n+0.4} \qquad \textbf{(8)}$$

This is the value (rather than j/n) that should be assigned to $N(t_j)/n$ for the jth failure.

Example 5. Determination of model parameters for the Weibull failure distribution

To illustrate the use of approximate formula for determining ranking, consider information given in Example 4. Suppose the first five widget failures occurred at the times t_1 = 1216 hours, t_2 = 5029 hours, t_3 = 13125 hours, t_4 = 15987 hours, and t_5 = 29301 hours, respectively. This gives us five data points. Here use $F(t_m)$ from Equation (8) in place of $[N(t_i)/n]$ in Equation (5):

$\ln(t_j)$	$\ln\left(\ln\left(\frac{1}{1-\frac{j-0.3}{n+0.4}}\right)\right)$
7.103	−4.962
8.523	−4.070
9.482	−3.602
9.680	−3.282
10.285	−3.038

Note: ln(t)=natural logarithm of t on base e(exponential)

By simple linear regression we can perform a least-squares fit of this sequence of k = 5 data points to a line. In terms of variables:

$x_j = \ln(t_j)$ and $v_j = \ln\left(\ln\left(\frac{1}{1-\frac{j-0.3}{n+0.4}}\right)\right)$

=> The estimated Weibull parameters are given by:

$$\beta = \frac{k\sum_{j=1}^{k} x_j v_j - \left(\sum_{j=1}^{k} x_j\right)\left(\sum_{j=1}^{k} v_j\right)}{k\sum_{j=1}^{k} x_j^2 - \left(\sum_{j=1}^{k} x_j\right)^2}$$

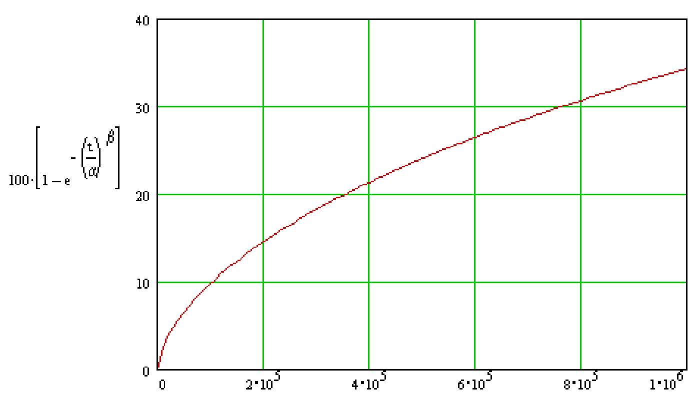
and

$$\alpha = \exp\left(\frac{\left(\sum_{j=1}^{k} v_j\right)\left(\sum_{j=1}^{k} x_j^2\right) - \left(\sum_{j=1}^{k} x_j\right)\left(\sum_{j=1}^{k} x_j v_j\right)}{-\beta\left(k\sum_{j=1}^{k} x_j^2 - \left(\sum_{j=1}^{k} x_j\right)^2\right)}\right)$$

For our example with k = 5 data points, we get □ = 0.609 and □ = 4.14×10^6 hours.

Example 6. Prediction of number of failures following Weibull failure distributions

Using equation (4), we can now predict the expected number of failures into the future, as shown in the figured below. Here Y-axis shows [100*N(t)/n] and x-axis shows time (hours) values.



What happens when we replace a failed unit?

Exponential failure distribution: In this case, we assume replacement of the failed units, so that the size of the overall population remains constant (for constant failure rate).

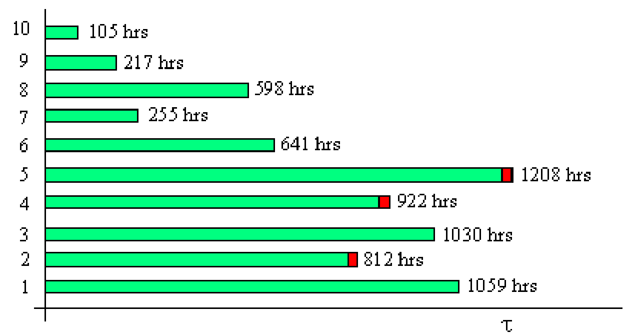
Weibull failure distribution: In this case, the failure rate of each widget depends on age of that particular widget, so, if we replace a unit that has failed at 10000 hours with a new unit, the overall failure rate of the total population changes abruptly, depending on whether □ is less than or greater than 1.**This is the primary reason** why we considered a coherent population of widgets whose ages are all synchronized. The greatly simplifies the analysis.

In more realistic situations the population of widgets will be changing, and the "age" of each widget in the population will be different, as will the rate at which it accumulates operational hours as a function of calendar time. More generally we could consider a population of widgets such that each widget has it's own "proper time" τ_j given by Equation (9):

$\tau_j = \tau_j \, (t \; \tau_j)$ for all $t > \tau_j$ (9)

where t is calendar time, τ_j is the birth date of the jth widget, and τ_j is the operational usage factor.
This proper time is then the time variable for the Weibull density function for the jth widget, and the overall failure rate for the whole population at any given calendar time is composed of all the individual failure rates. In this non-coherent population, each widget has its own distinct failure distribution.

Example 7: At a given calendar time, the experience basis of a particular population might be as illustrated in the following figure.



(Note: Y-axis denotes widget number and X-axis denotes "proper time" (hours) (green: working and red: not working).
So far there have been three failures, widgets 2, 4, and 5. The other seven widgets are continuing to accumulate operational hours at their respective rates. These are sometimes called "right censured" data points, or "suspensions", because we imagine that the testing has been suspended on these seven units prior to failure. We do not know when they will fail, so we can not directly use them as data points to fit the distribution, but we would still like to make some use of the fact that they accumulated their respective hours without failure.

Part 3

Steps in reliability analysis:

Step 1: Rank all the data points according to their accumulated hours in increasing order, as shown below.

Failure Rank	Overall Rank	Hours	
	1	105	
	2	217	
	3	255	
	4	598	
	5	641	
1	6	812	failed
2	7	922	failed
	8	1030	
	9	1059	
3	10	1208	failed

Step 2: Assign adjust ranks to the widgets that have actually failed. Letting k_j denote the overall rank of the jth failure, and letting r(j) denote the adjusted rank of the jth failure (with r(0) defined as 0), the adjusted rank of the jth failure is given by the Equation (10:

$$r(j) = r(j-1) + \frac{N+1-r(j-1)}{N+1-(k_j-1)} \qquad \textbf{(10)}$$

So, for the example above, we have (here N=10; j=1 to 3; k=1 to 10 as N=10)

$$r(1) = r(0) + \frac{10+1-r(0)}{10+1-(6-1)} = 1.667$$ [here, r(0)=0]

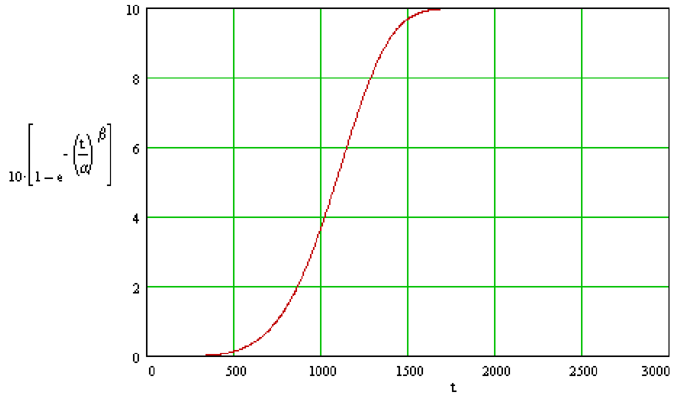
$$r(2) = r(1) + \frac{10+1-r(1)}{10+1-(7-1)} = 3.533$$ [here, r(1)=1.667 from previous calculation]

$$r(3) = r(2) + \frac{10+1-r(2)}{10+1-(10-1)} = 7.267$$

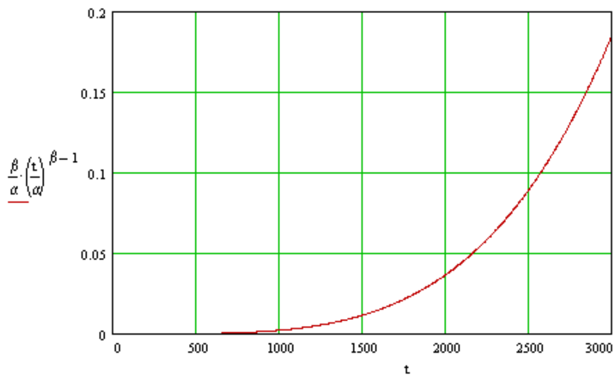
Using these adjusted ranks, we have the three data points (Note: ln(812 hours)= 6.7 for widget #2.Similarly calculate ln(proper time) for other widgets as well.

$\ln(t_j)$	$\ln\left(\ln\left(\frac{1}{1-\frac{j-0.3}{N+0.4}}\right)\right)$
6.700	−1.960
6.826	−0.988
7.096	0.102

Now objective is to determine Weibull failure distribution parameters to predict future number of failed widget. Fitting these three points using linear regression (as discussed above), we get the Weibull parameters α = 1169 and β = 4.995. The expected number of failures (which is just n times the cumulative distribution function(from Equation 2)) is shown below. This shows a clear "wear out" characteristic, consistent with the observed failures (and survivals).



The failure rate is quite low until the unit reaches about 500 hours, at which point the rate begins to increase, as shown in the figure below.



The examples discussed above are classical applications of the Weibull distribution, but the Weibull distribution is also sometimes used more loosely to model the "maturing system" effect of a high level system being introduced into service. In such a context the variation in the failure rate is attributed to gradual increase in familiarity of the operators with the system, improvements in maintenance, incorporation of retro-fit modifications to the design or manufacturing processes to fix unforeseen problems, and so on. Whether the Weibull distribution is strictly suitable to model such effects is questionable, but it seems to have become an accepted practice. In such applications it is common to lump all the accumulated hours of the entire population together, as if every operating unit has the same failure rate at any given time. This makes the analysis fairly easy, since it avoids the need to consider "censored" data, but, again, the validity of this assumption is questionable.

Questions

Q.1 Define the following terms: Failure Rate; Cumulative distribution function; Density function; Mean time between failure.

Q.2 Define the relationship between cumulative distribution function (cdf) $F(t)$ and the density function $f(t)$.