Chapter 11

Weibull Distribution

Jaroslav Menčík

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/62375

Abstract

Weibull distribution is very flexible in fitting empirical data, such as strength or time to failure. Several methods for the determination of parameters are described, including direct fitting using solvers available in universal programs. Also finding of parameters of exponential distribution is described. The use of Weibull distribution is illustrated on examples.

Keywords: Probability, reliability, Weibull distribution, exponential distribution, determination of parameters, least squares method, solver

A special position in reliability assessment pertains to Weibull distribution, which offers great flexibility in fitting empirical data. The distribution function (Fig. 1a) is

$$F(t) = 1 - \exp\left\{-\left[\left(t - t_0\right)/a\right]^b\right\}, \tag{1}$$

with parameters a, b, and t_0. The scale parameter a is related to the values of t and ensures that the distribution is independent of the units of t (e.g. minutes or hours). The constant b is shape parameter. Depending on its value, Weibull distribution can approximate various, even very different shapes (Fig. 5 in Chapter 2). It is suitable for the characterization of time to failure as well as strength or load; therefore, it became popular in reliability assessment. The constant t_0 is the threshold value that corresponds to the minimum possible value and characterizes the position of the distribution on the t-axis. (t is the usual symbol for time; for other quantities, other symbols may be used.)

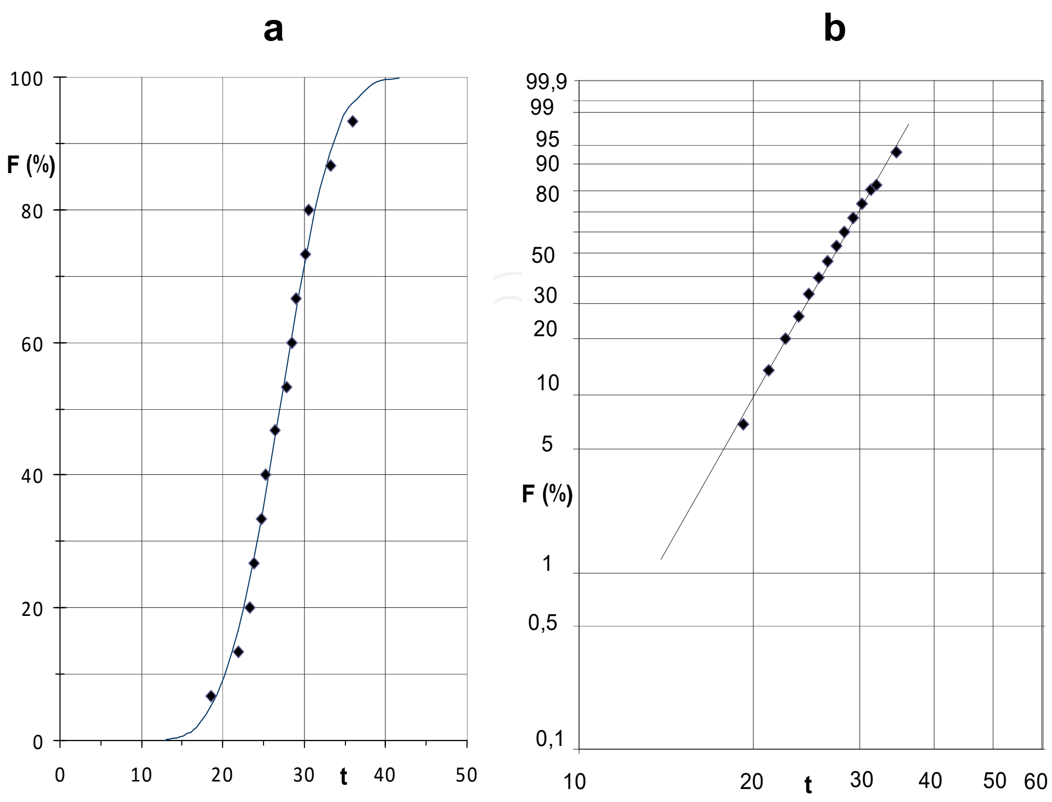

Figure 1. Weibull distribution function *F*(*t*): (a) original coordinate system, (b) transformed coordinates (Weibull probabilistic paper).

1. Determination of parameters in a two-parameter distribution

The strength or time to failure cannot attain negative values, so that the threshold parameter is often assumed zero, $t_0 = 0$. The distribution function (1) will thus have only two parameters:

$$F(t) = 1 - \exp\left[-(t/a)^b \right]. \tag{2}$$

Parameters a and b can be found easily, as the transformed data can be fitted by a straight line. Double logarithmic transformation and rearrangement change Equation (2) to

$$\ln t = \ln a + (1/b) \ln\{\ln[1/(1-F)]\}, \tag{3}$$

which corresponds to the equation of straight line (Fig. 1b)

http://dx.doi.org/10.5772/62375

$$Y = A + BX, \tag{4}$$

where $Y = \ln t$, $X = \ln\{\ln[1/(1-F)]\}$, $A = \ln a$, $B = 1/b$.

The method of linearization was very popular in the past, and it is still often used for the determination of parameters from the operation data via a special diagram, called Weibull paper (Fig. 1b). For its construction, the individual measured values t_j and the corresponding values F_j of the empirical distribution function are needed. The t_j values are obtained by rankordering the n data from operation (e.g. times to failure) from the minimal value ($j = 1$) to maximal ($j = n$). The corresponding values of distribution function are calculated as

$$F_j = j / (n + 1); \tag{5}$$

j is the rank number and n is the total number of measured values. The explanation of formula (5), common for order statistics, is simple. If we have, say, 100 values and order them from the minimal to maximal, then the probability F that t will be smaller or equal to the lowest of 100 values, t_1, is 1:100. The probability of $t \le t_2$ is 2/100, etc.; generally, $F_j = j/n$. In Equation (5), 1 was added to the denominator because of mathematical correctness; the probability F that t will be smaller or equal t_n must be smaller than 1, simply because if more measurements would be done, values higher than t_n could appear. Also other formulas exist for the calculation of empirical F_j values [e.g. $F_j = (j - ½)/n$], but none can be recommended unequivocally, especially when considering the fact that bigger errors in the determination of distribution parameters can arise due to the small amount of data than due to the formula used for F_j.

The regression constants A, B can be obtained by fitting the empirical data by a straight line (using Weibull paper or a program for curve fitting, such as "Insert Trendline" in Excel). Then, the constants in the distribution function (2) are obtained from A and B by inverse transformation:

$$b = 1/B, \ a = \exp(A). \tag{6}$$

Plotting the empirical data into the coordinate system $X = \ln\{\ln[1/(1-F)]\}$, $Y = \ln t$, enables a good visual check. In the ideal case, if Equation (2) is valid, the data lie on a straight line.

2. Determination of parameters in a three-parameter distribution

A two-parameter distribution is not always suitable. Sometimes, the transformed data do not lie on a straight line, or it is obvious that the distribution should have a threshold value t_0 higher than zero. In such case, the use of a two-parameter distribution as a base for dimensioning could lead to uneconomical design, and a three-parameter function (1) would be better.

The parameters in this distribution can be found by the procedure for a two-parameter function if t in Equation (2) is replaced by the expression $t - t_0$; the constant t_0 must be chosen in advance. For various t_0 values, the shape of empirical distribution varies. The best t_0 value is such for which the transformed data best resemble a straight line. However, a more straightforward procedure exists.

Direct determination of parameters

The constants a, b, and t_0 can also be obtained in a simpler way without any transformation. The solution of Equation (1) for t gives the formula for quantiles:

$$t = t_0 + a\left\{\ln\left[1/\left(1 - F\right)\right]^{1/b}\right\}. \tag{7}$$

This equation and the least-squares method are used in search for such values of a, b, and t_0, which minimize the sum of squared differences between the measured and the calculated values of t,

$$(t_{\mathrm{j,meas}} - t_{\mathrm{j,calc}})^2 = \text{ min !} \tag{8}$$

If a suitable solver is available for such minimization (one is present also in Excel), it is then sufficient to prepare one series of measured data, $t_{\mathrm{j,meas}}$, and another series of the $t_{\mathrm{j,calc}}$ values, calculated via Equation (7) for the same values of Fj using the parameters a, b, and t_0. Solver's command to minimize the expression (8) by changing a, b, and t_0 will do the job. An example is shown at the end of this chapter.

Remark: Formula (7) is also suitable for the determination of a "minimum guaranteed value"(e.g. strength or time to failure) for acceptably low probability F.

In addition to flexibility, Weibull distribution has one more advantage. The shape parameter b in Equation (1) or (2) is related to the character of failures. This is well visible at the bathtub curve (Fig. 1 in Chapter 4). The values $b < 1$ are typical of decreasing failure rate λ and may thus indicate the period of early failures. On the contrary, $b > 1$ corresponds to increasing failure rate λ and is typical of the period of aging or wear out. The value $b = 1$ corresponds to the constant failure rate λ = const, with failures from many various reasons (see Chapter 4). The exponent b thus can inform generally about the possible kind of failures and about the period in the life of an object even if the amount of data is not large. However, caution is necessary. If the data from a long period are fitted by Weibull distribution, failures from various reasons and stages can be mixed, and the relation of b to the kind of failures is not unambiguous.

Remark: Weibull distribution was proposed in 1939 by the Swedish engineer Waloddi Weibull, who studied the strength of materials, life endurance of ball bearings, and fatigue life of mechanical components and other quantities. Later, it appeared that this very useful distribution belongs to the family of extreme value distributions [1, 2]. More on Weibull distribution and its applications can be found, for example, in [3 - 5].

http://dx.doi.org/10.5772/62375

3. Exponential distribution

Let us now look at a special and very important case. With the shape parameter $b = 1$, Weibull distribution simplifies to exponential distribution

$$F(t) = 1 - \exp\left[-(t/a)\right], \text{ or } F(t) = 1 - \exp\left[-\left(t - t_0\right)/a\right]. \tag{9}$$

The probability density and distribution function are depicted in Fig. 5. The parameters a and t_0 can be determined similarly as described above. If $t_0 = 0$, the remaining parameter a is usually calculated from the mean time to failure, as it will be shown in Chapter 20. Typical of exponential distribution is that the standard deviation has the same or similar value as the mean.

The determination of parameters and use of Weibull and exponential distribution will be demonstrated in the following examples.

Example 1

The strength (S) of a new alloy was measured on seven specimens, with the following results: 203, 223, 248, 265, 290, 313, and 342 MPa. Solve the following three problems:

A. Determine the parameters of Weibull distribution for this alloy using:

a. Two-parameter distribution and linearized data;

b. Two-parameter distribution, applying Solver on the original data without transformation); and

c. Three-parameter distribution, applying Solver on the nontransformed data.

B. Calculate (for each case) the probability that the strength will be lower than 120 MPa.

C. Calculate (for each distribution) the "minimum guaranteed" strength such that the probability of the actual strength being lower equals: 0.05 – 0.01 – 0.001.

Solution.

Task A. Determination of distribution parameters

a. Linearized two-parameter Weibull distribution. The strength values, ordered from minimum to maximum, are given in Table 1 together with the values of distribution function, calculated as $F_j = j/(n + 1)$, with $n = 7$; see also Fig. 2. The distribution function $F(t) = 1 - \exp[-(t/a)^b]$ was transformed to linear form; see Equation (4) and the following formulas. The transformed values are in the columns X_j and Y_j. Note: The values of distribution function are fixed (deterministic), as they correspond to the number of measured values, whereas the strengths exhibit random variations. Therefore, F is the independent variable and t is the dependent variable.

j	S_j	F_j	X_j	Y_j	$S_{j,c,lin2}$	$S_{j,c,sol2}$	$S_{j,c,sol3}$
1	203	0.1250	-2.0134	5.3132	196.6331	194.501	201.13921
2	223	0.2500	-1.2459	5.4072	228.3515	227.021	226.32602
3	248	0.3750	-0.7550	5.5134	251.2710	250.615	247.17612
4	265	0.5000	-0.3665	5.5797	271.0299	271.015	266.95925
5	290	0.6250	-0.0194	5.6699	289.9970	290.645	287.54013
6	313	0.7500	0.3266	5.7462	310.2209	311.624	311.21631
7	342	0.8750	0.7321	5.8348	335.7233	338.143	343.64209

Subscript c means calculated; lin2 – linearized, two parameters; sol2 – nonlinearized, Solver, two parameters; sol3 – nonlinearized, Solver, three parameters

Table 1. Measured values $S(F_j)$ and those calculated using three methods.

The transformed values were fitted by linear function (4); see columns X_j and Y_j in Table 2. The regression constants were $A = 5.673642$ and $B = 0.194844$. The inverse transformation has given $a = \exp A = 291.0928$ and $b = 1/B = 5.132311$, so that the two-parameter distribution function is $F(t) = 1 - \exp[-(S/291.0928)^{5.13231}]$. The corresponding calculated values S_j are in column $S_{j,c,lin2}$ and depicted by a curve in Fig. 2.

b. Two-parameter distribution, application of Solver on untransformed data. In this case, the strength values $t_{j,calc}$ were calculated for the individual F_j values using Equation (7), with $t_0 = 0$. (See column $t_{j,calc,2p}$.) Now, a quantity for characterization of the quality of the fit was defined: sum of the squared differences of the measured and calculated strengths, $\sum(t_{meas} - t_{calc})^2$. The Solver then changes the constants a and b of the distribution function (7) automatically until the sum of squared differences attains a minimum. In the investigated example, the "optimum" constants were $a = 291.7807$ and $b = 4.964505$, near to the results of the linearized problem. The calculated values are in column $S_{j,c,sol2}$.

c. Three-parameter distribution, application of Solver on untransformed data. The difference from the previous case is the full form of distribution function (7). Also here the sum of the squared differences of the measured and calculated strength, $\sum(t_{meas} - t_{calc})^2$, was minimized. The resultant constants were $a = 154.9796$, $b = 2.4156$, and $t_0 = 133.7975$. The calculated values are in column $S_{j,c,sol2}$. The calculated distribution function is plotted by the thick curve in Fig. 2. Also the distribution function of a two-parameter distribution function is shown (thin curve). The curves for cases (a) and (b) were very close to each other.

Task B. Determination of probability S ≤ 120 MPa

The probabilities are as follows:

a. 0.010533, (b) 0.012069, and (c) 0; the minimum possible value is $t_0 = 133.8$ MPa.

Task C. Determination of guaranteed strength

http://dx.doi.org/10.5772/62375

The results are in the following table.

Probability of lower strength	**Guaranteed strength (MPa) with the constants from the method:**		
	(a)	**(b)**	**(c)**
0.05	163.19	160.41	179.12
0.01	118.79	115.51	156.88
0.001	75.78	72.58	142.68

Note the big difference between the two- and three-parameter distributions for very low failure probabilities (cf. also Fig. 2). According to the three-parameter model, the minimum (threshold) strength is 133.8 MPa.

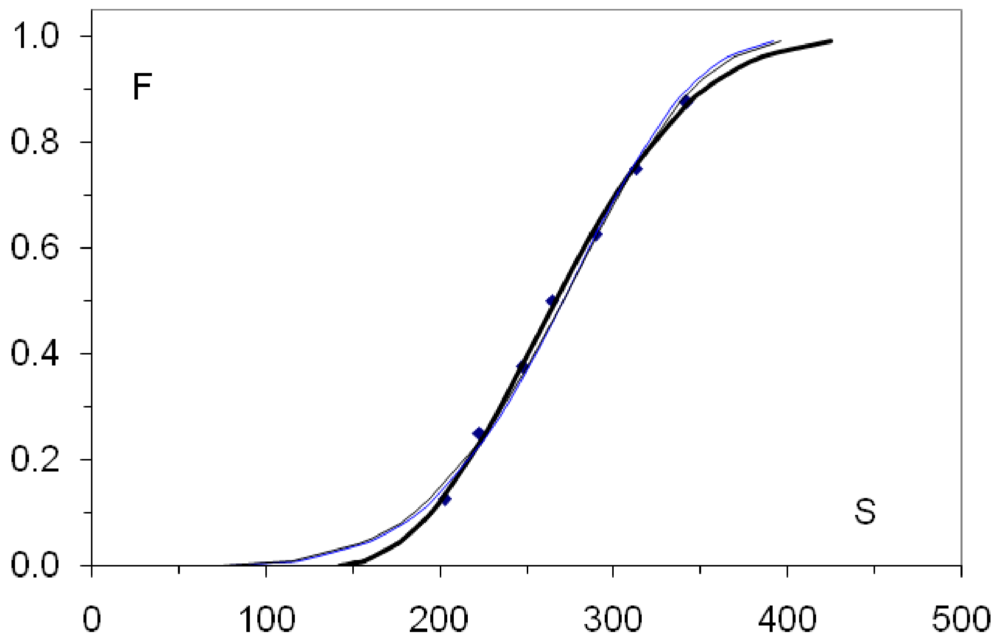

Figure 2. Measured values of strength (S) and approximate distribution functions (F) for various approximations in Example 1. Thick curve – case c, three-parameter function; thin curves – cases a, b, two-parameter curves.

Example 2

Eight components ($n = 8$) were tested until failure. The failures occurred at the following times t_j: 65, 75, 90, 120, 250, 510, 520, and 760 h. Calculate the mean time to failure and failure rate. Calculate also the standard deviation, so that you can assess whether exponential distribution may be used for the time to failure.

Solution.

$MTTF = \sum t_j/n$ = (65+75+90+120+250+510+520+760)/8 = 298.750 h.

The sample standard deviation [Equation (4) in Chapter 2] is σ_{MTTF} = 264.288 h. This is reasonably close to the sample mean, and an exponential distribution may be assumed. For this case, failure rate $\lambda = 1/MTTF = 1/298.75 = 0.003347\ h^{-1}$. The determination of confidence interval for λ will be demonstrated in a similar case in Chapter 20.

Author details

Jaroslav Menčík

Address all correspondence to: jaroslav.mencik@upce.cz

Department of Mechanics, Materials and Machine Parts, Jan Perner Transport Faculty, University of Pardubice, Czech Republic

References

[1] Rao S S. Reliability-Based Design. New York: McGraw-Hill; 1992. 569 p.

[2] Gumbel J E. Statistics of Extremes. New York: Columbia University Press; 1958. 375 p.

[3] Bentley J P. Introduction to Reliability and Quality Engineering. Harlow, England: Addison-Wesley; 1999. 202 p.

[4] Abernethy R B. The New Weibull Handbook. 4th ed. North Palm Beach, Florida: Robert B Abernethy; 2001. ISBN 0-9653062-1-6.

[5] VDA 3. Zuverlässigkeitssicherung bei Automobilherstellern und Lieferanten. Frankfurt am Main: Verband der Automobilindistrie e.V; 1984.