The Weibull Distribution and the Problem of Guaranteed Minimum Lifetimes

Prof.Dr.-Ing.habil. Harald Rosemann IMKT (Leibniz University Hannover)
Josephine Kelley IMKT (Leibniz University Hannover)
Prof.Dr.-Ing. Gerhard Poll IMKT (Leibniz University Hannover)

Abstract

For service life tests, a shifted Weibull distribution, also known as the translated or three-parameter Weibull distribution, is commonly used. The shifted Weibull distribution promises completely fault-free operation until time $t = L_0$, in other words, in the early stage the process is deterministic. Only after this phase does the distribution allow random behavior, i.e. from the time $t = L_0$ on, the process is stochastic. This model, which is based on two consecutive time periods of quite different nature, is at odds with the idea of a continuously progressing fatigue, wear or decay process as long as there are no influences from outside. To replace this arguably inconsistent model, variants of the Weibull distribution of purely stochastic nature are proposed and investigated that start with a reduced probability of failure before transitioning to normal Weibull behavior.

1 Introduction

Materials wear and fatigue, and, as a result, failures occur. Individual failures as a consequence of fatigue or wear occur at unpredictable, statistically distributed times. It is often assumed that the service lifetimes are distributed according to the Weibull distribution, as this is the distribution that yields the highest target values in parameter estimation using optimization methods such as the maximum likelihood procedure. The original Weibull distribution is defined by two parameters.

Attempts have been made to develop a modified variant of the Weibull distribution by introducing a third parameter in order to describe failure behavior that is initially infrequent. This variant is in constant use, which is clear from some of the first entries from an internet search for the term 'Weibull distribution'. The additional third parameter, also known as threshold, accounts for a minimum initial operating time, during which an (alleged) absolute and total absence of failure is guaranteed. In the following, we consider whether this assumption is justified or should be replaced by a more stringent approach.

2 The problem

2.1 The Weibull distribution with two parameters

For many service life tests, the original Weibull distribution with two parameters can suitably represent the observed values. In general, F(t) denotes the cumulative distribution function of a time-dependent random variable and W(t) specifically denotes the Weibull cumulative distribution function:

$$F(t) = W(t) = \begin{cases} 1 - e^{-\left(t/T\right)^{\beta}}, & t \geq 0,\ \beta > 0,\ T > 0 \\ 0, & t < 0 \end{cases} \tag{1}$$

An important characteristic is that, in the exponential function, the time t itself is raised to the power β. The parameter T is called the characteristic time; regardless of the value of β, one always has $W(T) = 1 - 1/e \approx 0.632$.

At $t = 0$, the cumulative distribution function $W(t)$ is equal to zero and begins to increase monotonically as a function of t, approaching the value 1 for large t. From the values of the cumulative distribution function, one attains the probability that a failure occurs at or before time t. With $W(t) = 0$ for $t < 0$, the distribution shows that the effect cannot occur before the cause, i.e. a failure can only be expected after the start of the damage-inducing loading; this fundamentally excludes the possibility of failure before the damage-inducing loading, and, indeed, the probability of a negative service lifetime is zero.

Instead of the characteristic value T, one commonly uses the L_{10}-lifetime and algebraically manipulates Eqn. (1) into:

$$F(t) = W(t) = \begin{cases} 1 - e^{\ln(0.9)\left[\frac{t}{L_{10}}\right]^{\beta}}, & t \geq 0,\ \beta > 0,\ L_{10} > 0, \\ 0, & t < 0 \end{cases} \tag{2}$$

Once again, there is a value independent from β that the cumulative distribution function depends on: by definition, $W(L_{10}) = 0.1$ and so L_{10} gives the time up to which 10% of failures are to be expected.

2.2 The shifted Weibull distribution (translated or 3-parameter Weibull distribution)

For certain applications, one discovers that the initial number of failures is lower than predicted by the standard Weibull distribution. This deviation is attributed to processes such as wear, deterioration, or fatigue, which usually require a certain amount of time for damage to develop into failure. For this reason, Snare [1] and later on Bergling [2], used a third parameter L_0, also known as threshold, in the evaluation of roller bearing lifetimes to shift the cumulative distribution function to the right, according to

$$F(t) = W(t) = \begin{cases} 1 - e^{\ln(0.9)\left[\frac{t-L_0}{L_{10}-L_0}\right]^{\beta}}, & \begin{cases} t \geq L_0, \\ \beta > 0, \\ L_{10} > L_0 \geq 0 \end{cases} \\ 0, & t < L_0 \end{cases} \tag{3}$$

to obtain a 'better' fit to the data points for early failures. When plotted, this correction can be visually judged to be adequate. Also, if the superiority of a parameter set is to be judged using the target value that arises from the optimization of an estimation process such as the maximum likelihood method, then the three-parameter Weibull distribution should indeed be preferred to the two-parameter Weibull distribution. On the one hand, this is the argumentation in favor of the three-parameter Weibull distribution.

2.3 The conflict

On the other hand, however, shifting the original Weibull distribution to get the curve of Eqn. (3) introduces a new phase into the model. It is valid for $t < L_0$ and is of purely deterministic nature; the second phase, valid for $t \geq L_0$, is stochastic. These two domains of fundamentally different nature share the predefined, non-random border at $t = L_0$.

In the first part, the model ensures that there are no failures before $t = L_0$. An event in this region representing a failure can not occur and is labeled as 'impossible' by definition of Eqn. (3). Strictly spoken, such a fundamental statement cannot be deduced or validated purely from observation, regardless of the number of data points. Even though an estimator $\widehat{L}_0$ for a sample exists and can be computed according to Park [3], this does not on its own prove the existence of a failure-free period of time L_0.

From a numerical point of view, one hardly notices a difference between 'exactly zero' and very, very small, say one billionth or even less. Qualitatively, on the other hand, the 'impossible event' is fundamentally different from one with a low probability. The first is based on abstract definition, the other is a matter of the real world; in the first case, one can be completely unconcerned, in the other one, precautionary measures may become necessary.

Additionally, this model necessitates an exogenous 'timer setting' that triggers the transition to the second phase after which the ongoing fatigue or wear processes are allowed to develop into a failure.

This is an unsatisfactory situation as there is a conflict. On the one hand, one has the best distribution (among the ones tested), while on the other hand, the statement and core assumptions of the distribution do not apply to the continuously progressing process that generates the observed values. A pragmatic way to resolve this issue would be to consider the Weibull distribution with $L_0 > 0$ an approximation. Nevertheless, one must be prepared to fend off any outside claims that one has guaranteed safety from premature failures. There is a dilemma with only one possible resolution: to find a distribution that yields even higher target values in parameter estimation, that can also be interpreted without any problems.

3 New approach

3.1 Hyperbola instead of the straight lines

The question therefore becomes whether it is possible to find an intermediate solution that preserves the Weibull character and allows for delayed failure behavior without permitting any misinterpretation. It is useful to simplify the equations by using the $(L_{10} - L_0)$-normalized variables $t' = t/(L_{10} - L_0)$ and $L'_0 = L_0/(L_{10} - L_0)$. We then can write what is different in each distribution as auxiliary functions of t' as $g_2(t') = t'$ and $g_3(t') = t' - L'_0$, respectively; the index is counting the parameters. The functions $g(t')$ are both the basis which is taken to the power β in the cumulative distribution function of Weibull.

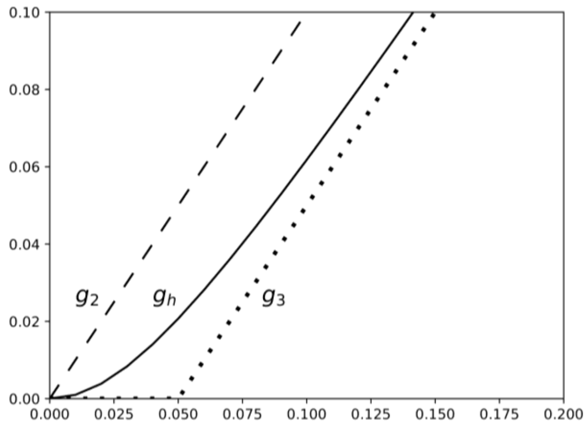

Fig. 1: $g(t')$ over t', $L'_0 = 0.05$

These two functions that depend on t' and L'_0 are shown in Fig. 1 as two parallel lines with g_2 on the left as a dashed line, and shifted by $L'_0 = 0.05$ to the right as g_3, which is represented by a dotted line. In the area between the two lines, we may draw another curve. This curve should increase monotonically from the value 0 at $t' = 0$ and approach the line g_3 for large t'. By taking the same name L'_0 for a similar parameter, an obvious choice would be the branch of a hyperbola, i.e.

$$g_h(t') = -L'_0 + \sqrt{t'^2 + L'^2_0}, \quad t' \geq 0,\ L'_0 \geq 0 \tag{4}$$

which is represented by the continuous line in Fig. 1. Near $t' = 0$ the function $g_h(t')$ behaves like $t'^2/2L'_0$, i.e. it begins with a horizontal tangent.[1]

3.2 Comparison of the cumulative distribution functions

The three versions of $g(t')$ lead to three Weibull distribution functions via $W(g(t'))$, where each $g(t')$ replaces the original t'; we apply the notation W_2 to mean $W(g_2(t'))$ for each $g(t')$. Figures 2 and 3 show the curves with linear coordinates on the left and Weibull coordinates on the right, which shows the original Weibull distribution as a straight line. For these calculations, $\beta = 1.35$ was chosen.

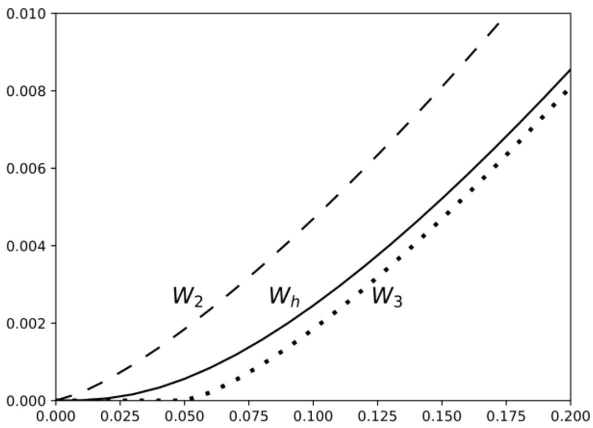

Fig. 2: $W(g(t'))$ over t', $L'_0 = 0.05$, linear coordinates

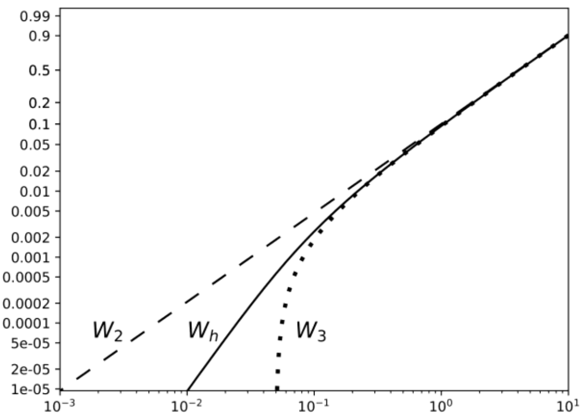

Fig. 3: $W(g(t'))$ over t', $L'_0 = 0.05$, Weibull coordinates

The desired sensible behavior is clearly visible. On the left in Fig. 2, W_h remains close to 0 longer than the original W_2 and in the further course it approaches W_3 more and more. In the Weibull diagram on the right, W_h begins steeper than W_2 but not as abruptly as W_3, which starts at the fixed value $t' = L'_0$.

Thus, early failures are less likely by the hyperbola approach according to Eqn. (4) than for the original Weibull distribution W_2 but not completely impossible before $t' = L'_0$ as it is for W_3. For larger values of t', the curves W_h and W_3 merge as a consequence of Eqn. (4), which can also be seen in the representation with Weibull axes. Fig. 2 with undistorted axes shows only the section with small t'; when these axes are expanded to $t' = 10$ as was done for the Weibull coordinates, one would not be able to distinguish the curves, especially for large t'.

[1] If, on the other hand, one wants to represent particularly frequent early failures rather than delayed ones, one may use a different hyperbola branch that increases quickly at $t' = 0$, just like the square root function: $g_h(t') = \sqrt{t'^2 + 2t'L'_0}$

The stated goal has been achieved since a useful replacement has been found. It is of continuously stochastic nature without a deterministic portion. Using initially small probabilities, it can represent delayed failures. There is no necessity for assumptions of a guaranteed lifetime L_0.

4 Extension of the hyperbola

4.1 Further replacement of the straight lines

Is the potential of the first approach now exhausted or can it be pursued further and expanded? The characteristic course of the hyperbola branch should be preserved; how can it be varied? By generalizing the square root and the second power, we arrive at

$$g_c(t') = -L'_0 + \left[(t')^c + (L'_0)^c \right]^{1/c}, \quad t' \geq 0, \quad L'_0 \geq 0, \quad c \geq 1 \tag{5}$$

with the new parameter c, the name of which is also used as an index for $g_c(t')$, denoting the modified approach.[2] The curve of $g_c(t')$ increases monotonically with t', as was the case with the first hyperbola in Eqn. (4); by replacing t' with $g_c(t')$ in the Weibull formula, the definition of a distribution is still fulfilled.

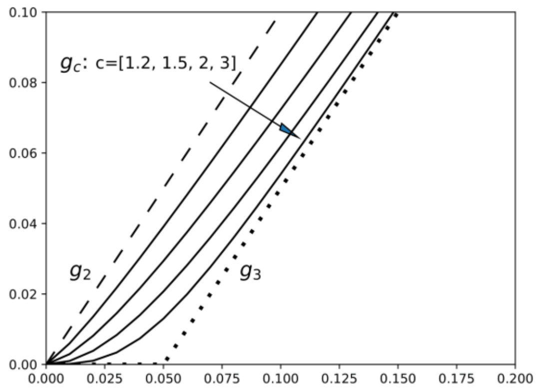

Fig. 4: $g(t')$ over t', $L'_0 = 0.05$

Figure 4 shows a sheath of continuous curves between the original straight lines, which are represented by dashed line and dotted line, respectively. The list shows the corresponding values for c, where the arrow is pointing in the direction of increasing values. For $t' = 0$, the curves increase with t', with almost horizontal tangent lines, like t'^c / cL'^{c-1}_0, and with increasing c they can thus lie along the time axis more closely and for a longer duration.

The new formula does not just fill the area between the first two straight lines, it also has the nice property of including the original Weibull distribution for $c = 1$, while the other shifted one is boundary case for $c \to \infty$.

[2] Values in the range $0 < c < 1$ generate more frequent early failures

4.2 Comparison of the cumulative distribution functions

The appearance of the corresponding cumulative distribution functions, on the left in equally divided coordinates and on the right with Weibull axes, now turns out as one might expect; between the two original curves, there are arbitrarily many intermediate variants. In Fig. 6 with Weibull coordinates, the curves run from the bottom almost straight up towards the line W_2 with varying curvature. Because the series expansion of $g_c(t')$ begins with order t'^c for small times t', the initial slope of the W_c in the Weibull coordinates is $c\beta$.

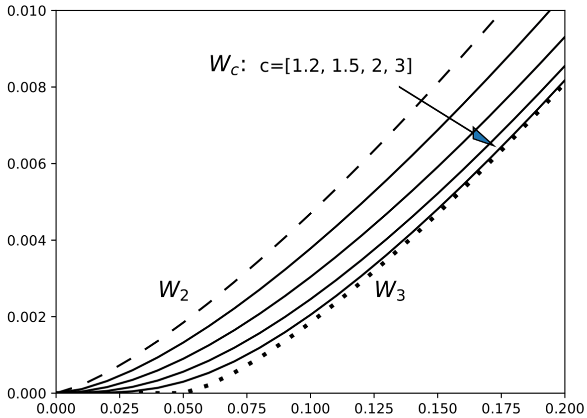

Fig. 5: $W(g(t'))$ over t', $L'_0 = 0.05$, linear coordinates

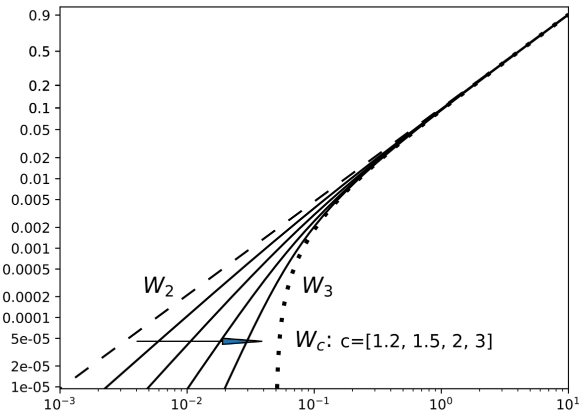

Fig. 6: $W(g(t'))$ over t', $L'_0 = 0.05$, Weibull coordinates

4.3 Special properties

As an example, Fig. 7 repeats the representation of the first hyperbola approach according to Eqn. (4). Additionally, a series of small circles shows the nearly linear initial slope of 2β and continues it to larger values. We see that this line, together with W_2, can be pieced together to conservatively approximate W_h. This is reminiscent of the old rule for the design of ball bearings, according to which the value of β should be increased to 1.5 for service lifetimes below L_{10}.[3]

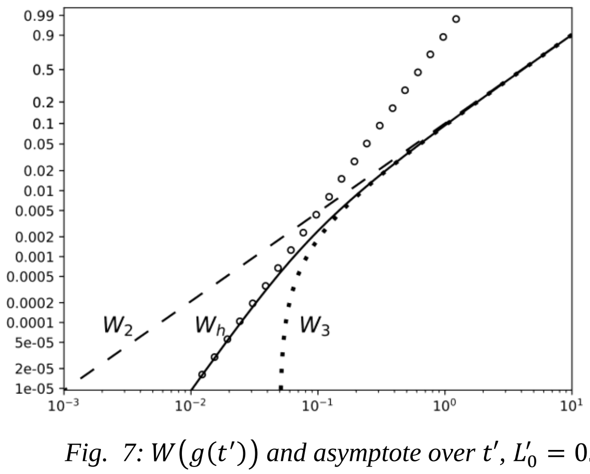

Fig. 7: $W(g(t'))$ and asymptote over t', $L'_0 = 0.05$, Weibull coordinates

[3] This modification is taken into account in the calculation of the reliability factor a_1 according to ISO 281 (2007 and previous versions) [4].

4.4 A short look at parameter estimation

For the original Weibull distribution with two parameters, one calculates the estimators $\hat{\beta}$ and $\hat{L}_{10}$ from measured service lifetimes. Every measurement has an influence on each of those two values. At most, the extreme failure times with low and high values have more influence on the result of the slope $\hat{\beta}$ in the Weibull coordinates and the intermediate values have more weight in the calculation of $\hat{L}_{10}$.

This changes for the four parameters of the extended approach. The new values L_0 and c arise on their own as the influence and efficacy in the initial range; as a result, their estimation $\hat{L}_{10}$ and $\hat{c}$ depend mainly on the times of the first early failure cases. This is related to a reduced dependence of both estimators $\hat{\beta}$ and $\hat{L}_{10}$ on the first early failure cases. A sufficiently large number of early failure cases is therefore necessary in order to estimate the new parameters accurately and reliably. If so far the number of early failures appeared to be sufficient to calculate the estimate $\hat{L}_0$ of the shifted Weibull distribution alone, such a number might now also be good enough to get usable values for $\hat{L}_0$ and $\hat{c}$ for the proposal. Moreover, typical values for certain special applications can be considered, such as the typical values of β equal to 1.11 for roller bearings primarily with point contacts versus β equal to 1.35 for cases with point and line contacts.

5 Conclusion

For continuously progressing wear and fatigue processes, the Weibull distribution with three parameters is not a suitable model for the distribution of service lifetimes as long as there are no external influences; it can only be viewed as a pragmatic approximation. In the approach presented here, the linear dependence on time t is replaced by a hyperbolic dependence. This new variant can represent delayed failure behavior in a fully stochastic model while avoiding difficulties with interpretation of the parameters, in particular with respect to guaranteed service lifetimes.

Acknowledgements

I would like to thank my highly esteemed colleagues M.Sc. Josephine Kelley and Prof.Dr.-Ing. Gerhard Poll at the IMKT, who translated this script from German and gave valuable support and encouragement for the preparation of this publication.

References

[1] Snare, B., Neuere Erkenntnisse über die Zuverlässigkeit von Wälzlagern; Die Kugellager-Zeitschrift, Heft Nr. 162 (1969) S. 3-7.

Remark: Figure 2 refers to an L_{10} of 15 Million rotations; in the text, however, it states: "Die Lager liefen . . . bei . . . einer Belastung, die nach dem Katalog einer L_{10}-Lebensdauer von 10 Millionen Umdrehungen entspricht." (The bearings ran for a load that, according to the catalogue, corresponds to an L_{10} service lifetime of 10 Million revolutions.)

[2] Bergling, G., Betriebszuverlässigkeit von Wälzlagern; Die Kugellager-Zeitschrift, Jahrgang 51, Heft Nr.188 (1976) S. 1-10.

Remark: Figure 3 (agrees with Figure 2 in [1]) shows an L_{10} of 15 Million revolutions; in the legend, a different value of 10 Million revolutions is stated.

[3] Park, C., A Note on the Existence of the Location Parameter Estimate of the Three-Parameter Weibull Model Using the Weibull Plot; Mathematical problems in engineering (2018), S. 1-6.

[4] ISO 281:2007, Rolling bearings - Dynamic load ratings and rating life.
(2007, and previous versions)