14th IMEKO TC10 Workshop Technical Diagnostics
New Perspectives in Measurements, Tools and Techniques
for system's reliability, maintainability and safety
Milan, Italy, June 27-28, 2016

Parameter estimation methods for failure rate distributions

Marcantonio Catelani, Lorenzo Ciani, Giulia Guidi, Matteo Venzi
Department of Information Engineering,
University of Florence,
via di S. Marta 3, 50139, Florence (Italy)
lorenzo.ciani@unifi.it

***Abstract* –The main part of the paper deals with the parameter estimation methods that used to determine the best-fitting distribution for a set of data collected during testing or field operations. The methods considered in that case are Least Square Estimation and Maximum Likelihood Estimation. Finally, two different types of field data such as failure times achieved with testing of electronic board for automatic control and with accelerated testing on electronic components were analyzed to find which distribution fits better the data.**

***Keywords* – Failure rate, failure distribution, exponential distribution, Weibull distribution, estimation parameter methods, Weibull analysis, best fit distribution.**

I. INTRODUCTION

The exponential distribution is the most used life distribution in applied reliability analysis since it is quite easy to manage and it represents a realistic lifetime model for a wide variety of items. The failure rate of an item with exponential life distribution is constant (i.e., independent of time), so it may be a realistic life distribution for an item during its useful life period in the bathtub curve [1, 2]. So exponential distribution doesn't consider the degradation of items. The failure rate function is the following:

$$\lambda(t) = \frac{f(t)}{R(t)} = \lambda \quad (1)$$

The Weibull distribution is one of the most widely used life distributions in reliability analysis. It is a very flexible distribution since it is based on different parameters and it can model different behaviors of failure rate functions. The failure rate function is

$$\lambda(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \qquad for\ t > 0 \quad (2)$$

Different values of the shape parameter β can have marked effects on the behavior of the distribution [3].

As shown in the plot, the behavior of the failure rate varying β comprises the three sections of the classic "bathtub curve": the plot of three Weibull distributions with respectively $\beta < 1$, $\beta = 1$ and $\beta > 1$ would reproduce the whole trend of the standard bathtub curve.

Also the modulation of the scale parameter η can influence the trend of curves, a change in this parameter has the same effect on the distribution of a change in the x-axis scale.

Figure 2 shows failure rate functions: a growth of the parameter η produces a corresponding increase of reliability and reduction of number of failures.

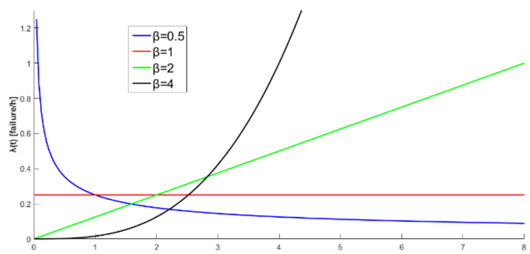

Figure 1. Weibull failure rate function varying β

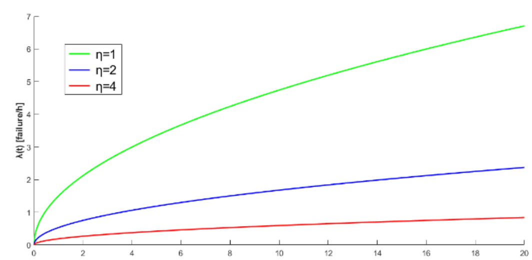

Figure 2. Weibull failure rate function varying η

II. LIFE DATA ANALYSIS AND PARAMETER ESTIMATION

Life data analysis starts with the analysis of a representative sample of units (belonging to the population of interest) in order to make a life prediction for all the products.

Life data analysis is usually developed in four steps:

- Collect product life data;
- Select the best-fitting distribution to model the life of the products;

- Assess the parameters that ensure the distribution to the data;
- Generate plots and results to estimate product life characteristics.

Weibull analysis is typically used to determine the best-fit distribution for a set of failure data because the distribution that best fits these data points provides info about the population from which they are drawn.

In order to fit a statistical model to a life data set, the analyst should estimate the parameters of the life distribution that will make the function fitting the data in the best way [4].

Several methods have been developed to estimate the parameters that will fit a lifetime distribution to a particular data set; the most important are described below.

a. Least Square Estimation (LSE)

The Least Square method requires that a straight line is fitted to a set of data points, such that the sum of the squares of the distance of the points to the fitted line is minimized [4]. This minimization can be performed in either the vertical or horizontal direction. If the regression is on *x*, then the line is fitted so that the horizontal deviations from the points to the line are minimized. If the regression is on *y*, then this means that the distance of the vertical deviations from the points to the line is minimized. This is illustrated in Fig. 3.

For the vertical regression, it is assumed that a set of data pairs (x_1, y_1), (x_2, y_2),…., (x_N, y_N) were obtained and plotted, and that the x-values are known exactly. Then, according to the least squares principle, the best fitting straight line to these data is the straight line [5]:

$$y = \hat{b}x + \hat{a} \tag{3}$$

Where:

$$\hat{a} = \frac{1}{N}\sum_{i=1}^{N}\ln[-\ln(1-F(t_i))] - \hat{b}\frac{1}{N}\sum_{i=1}^{N}\ln t_i \tag{4}$$

$$\hat{b} = \frac{N\sum_{i=1}^{N}\ln t_i \ln[-\ln(1-F(t_i))] - \sum_{i=1}^{N}\ln t_i \sum_{i=1}^{N}\ln[-\ln(1-F(t_i))]}{N\sum_{i=1}^{N}(\ln t_i)^2 - \left(\sum_{i=1}^{N}\ln t_i\right)^2} \tag{5}$$

Then the estimated parameters are:

$$\begin{cases} \eta = e^{-\frac{\hat{a}}{\hat{b}}} \\ \beta = \frac{1}{\hat{b}} \end{cases} \tag{6}$$

For the horizontal regression, the same least squares principle is applied. The best fitting straight line to these data is the straight line:

$$y = -\frac{\hat{a}}{\hat{b}} + \frac{1}{\hat{b}}x \tag{7}$$

Where:

$$\hat{a} = \frac{1}{N}\sum_{i=1}^{N}\ln t_i - \hat{b}\frac{1}{N}\sum_{i=1}^{N}\ln[-\ln(1-F(t_i))] \tag{8}$$

$$\hat{b} = \frac{N\sum_{i=1}^{N}\ln t_i \ln[-\ln(1-F(t_i))] - \sum_{i=1}^{N}\ln t_i \sum_{i=1}^{N}\ln[-\ln(1-F(t_i))]}{N\sum_{i=1}^{N}(\ln[-\ln(1-F(t_i))])^2 - \left(\sum_{i=1}^{N}\ln[-\ln(1-F(t_i))]\right)^2} \tag{9}$$

Then the parameters estimated are:

$$\begin{cases} \eta = e^{-\frac{\hat{a}}{\hat{b}}\frac{1}{\beta}} \\ \beta = \frac{1}{\hat{b}} \end{cases} \tag{10}$$

The correlation coefficient is a measure of the quality of data-fitting of the linear regression model and it is usually denoted by ρ. The correlation coefficient of the population is defined as follows:

$$\hat{\rho} = \frac{N\sum_{i=1}^{N}x_i y_i - \sum_{i=1}^{N}x_i\sum_{i=1}^{N}y_i}{\sqrt{\left(N\sum_{i=1}^{N}x_i^2 - \left(\sum_{i=1}^{N}x_i\right)^2\right)\left(N\sum_{i=1}^{N}y_i^2 - \left(\sum_{i=1}^{N}y_i\right)^2\right)}} \tag{11}$$

It assumes values in a range [-1, 1], the closer the value is to ±1, the better is the linear fitting.

The least squares estimation method is quite good for functions that can be linearized: for these distributions, the calculations are relatively easy and straightforward, having closed-form solutions that can readily yield an answer without having to resort to numerical techniques or tables.

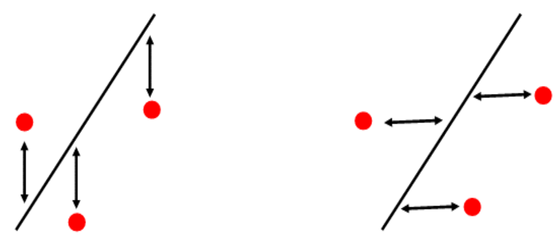

Figure 3. Rank regression on y (on the left) and on x (on the right)

b. Maximum Likelihood Estimation (MLE)

From a statistical point of view, the Maximum Likelihood Estimation method is considered one of the most robust parameter estimation techniques [4]. The basic idea behind MLE is to obtain, for a given distribution, the most likely values of the parameters that will best describe the data.

Supposing *T* is a continuous random variable with PDF $f(t,\beta,\eta,\gamma)$ [3], where t,β,η,γ are unknown parameters which need to be estimated, with *R* independent observations, $x_1, x_2, x_3, \ldots, x_R$, which correspond to failure times (in life data analysis). The likelihood function is given by:

$$L(\beta,\eta,\gamma \mid t_1, t_2, t_3, \ldots, t_R) = L = \prod_{i=1}^{R} f(t_i;\beta,\eta,\gamma) \tag{12}$$

The logarithmic likelihood function is the following:

$$\Lambda = \ln L = \sum_{i=1}^{R}\ln f(t_i;\beta,\eta,\gamma) \tag{13}$$

The maximum likelihood estimators (or parameter values) of β, γ, η are obtained by maximizing Λ or *L*. By maximizing Λ which is much easier to work with than *L*, the maximum likelihood estimators (MLE) are the simultaneous solutions of equations such that:

$$\frac{\partial \Lambda}{\partial \beta} = 0 \quad \frac{\partial \Lambda}{\partial \gamma} = 0 \quad \frac{\partial \Lambda}{\partial \eta} = 0 \tag{14}$$

The log-likelihood function is the following:

$$\Lambda = N \ln \beta - N\beta \ln \eta + (\beta - 1) \sum_{i=1}^{N} \ln(t_i) - \sum_{i=1}^{N} \left(\frac{t_i}{\eta}\right)^{\beta} \tag{15}$$

The parameters are achieved by maximizing the equation above. Deriving and solving to η Eq. 15, the estimated parameter is:

$$\eta = \left(\frac{1}{N} \sum_{i=1}^{N} t_i^{\beta}\right)^{\frac{1}{\beta}} \tag{16}$$

Furthermore, the shape parameter has not a closed form:

$$\frac{\partial \Lambda}{\partial \beta} = N \left(\frac{1}{N} \sum_{i=1}^{N} \ln x_i + \frac{1}{\beta} \frac{\sum_{i=1}^{N} \left(x_i^{\beta} \ln x_i\right)}{\sum_{i=1}^{N} x_i^{\beta}}\right) = 0 \tag{17}$$

The equation can't be solved analytically but needs a numeric technique or a software implementation.

c. Comparison between estimation methods

From a statistical point of view, MLE is usually recommended for large samples because it is versatile, applicable to most models and different types of data, and produces the most precise estimates [6].

The advantages of the MLE method over the LSE method are the following: the distribution parameter estimates are more precise, the estimated variance is smaller and the calculations use more of the information in the data.

The LSE method is also traditionally associated with the use of probability plots to assess goodness-of-fit. However, the LSE method can provide misleading results on a probability plot.

III. CASE STUDY

The data considered come from two different testing procedures: the first is a test on electronic board for automatic control while the second deals with accelerated testing of electronic components. The selected distributions are analyzed to determine how well they fit the data point within the data set; once the analysis is complete, the software shows the ranking results for the selected distributions. To determine the rankings, Λ (log-likelihood function) or ρ (correlation coefficient) are used.

a. Test On Electronic Board For Automatic Control

Table 1 shows the failure times achieved with the testing procedure on a population of ten (10) electronic boards; Table 2, instead, shows the results of the best-fitting procedure using both methods LSE and MLE.

The two methods give different solutions for the second and third distributions; both Weibull and normal distributions fit well the data, but because the estimation methods are different, the rank is different. The other distributions, such as log-normal and exponential don't fit the data and they show a lower value of the coefficient.

Table 1. Failure times of components

Component	Failure Time [h]
1	1200
2	2300
3	2500
4	2800
5	3000
6	3700
7	4000
8	4100
9	4200
10	4800

The distribution that better fits the data is the three-parameter Weibull since the correlation coefficient is almost unitary and the log-likelihood coefficient is higher than the other distributions.

Fig. 4 and Fig. 5 show the data best-fitting as a confirm of the results described above.

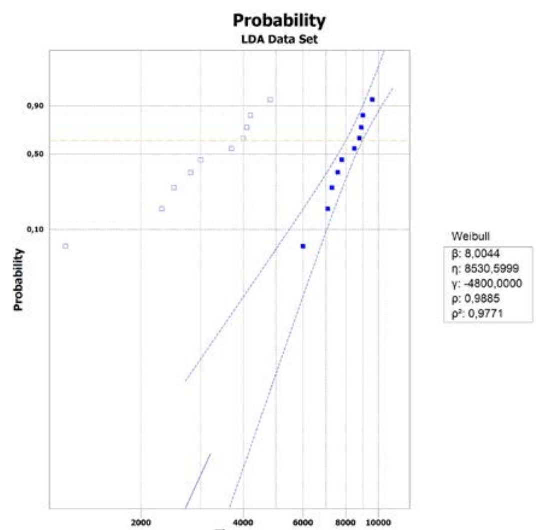

Figure 4. Three-parameter Weibull probability plot

In particular, Fig. 4 shows how the data follows the three-parameter Weibull distribution: all the data are placed almost linearly and the distance between the real curve and the expected Weibull line is minimized. The correlation coefficient ρ=*0,9885* is very high and very close to one.

Table 2. Ranking of the best fit distribution

Least Square Estimation		Maximum Likelihood Estimation	
Distribution	ρ	Distribution	Λ
Weibull (β, η, γ) • β= 8,0 • η=8530 h • γ= -4800 h	**0,9885**	Weibull (β, η, γ) • β= 9,45 • η=8504 h • γ= -4800 h	**-83,30**
Normal (μ,σ) • μ=3260,00 h • σ=1165 h	**0,9799**	Normal (μ, σ) • μ=3259 h • σ= 1035 h	**-83,62**
Weibull (β, η) • β=2,80 • η=3681 h	**0,9730**	Weibull (β, η) • β=3,68 • η=3621 h	**-83,47**
Log-normal (μ, σ) • μ=8,0 h • σ=0,42 h	**0,9331**	Log-normal (μ, σ) • μ=8 h • σ= 0,39 h	**-84,96**
Exponential (λ, η, γ) • λ=0,00056 h^{-1} • η=1785 h • γ=1188 h	**0,8864**	Exponential (λ, η, γ) • λ=0,000485 h^{-1} • η=2060 h • γ=1200 h	**-86,30**
Exponential (λ, η) • λ=0,00039 h^{-1} • η=2529 h	**0,8864**	Exponential (λ, η) • λ=0,000307 h^{-1} • η=3260 h	**-90,89**

b. Accelerated Test On Electronic Components

Thirty (30) electronic components have been tested with a high stress procedure and the corresponding failure times are shown in the Table 3.

The best-fitting results are listed in Table 4. Since the number of components is higher than the previous case, the data are analyzed only with the MLE technique that is particularly suited for this kind of applications. The rankings show that the best-fitting distributions are the two-parameter exponential and the three-parameter Weibull.

Fig. 5 shows the probability plot of the data-set supposing a two parameters exponential distribution, where most of data are distributed on the exponential line: the software calculates a very high log likelihood coefficient Λ=-340,92 as a confirm that the distribution fits very well the data.

Fig. 6, instead, shows the outcome using the Weibull distribution: it has a lower value of log-likelihood function Λ=-340,93, however it offers a good fitting and a satisfying approximation of the dataset. In fact, almost all data are concentrated on the Weibull line, it's possible to note that in Fig.5 data are located more linearly, anyway both the distribution are a good approximation. The Weibull distribution is a satisfying approximation also in this case, even if the exponential is a more accurate estimate for this dataset, it demonstrates the flexibility of the Weibull distribution to describe different types of data.

Table 3. Failure times of components

Component	Failure time [h]	Component	Failure time [h]
1	2100	16	28100
2	3800	17	29500
3	5400	18	31000
4	6600	19	33800
5	7600	20	34100
6	7800	21	35400
7	12300	22	35800
8	13000	23	43100
9	15200	24	45700
10	15900	25	54500
11	19900	26	56900
12	20100	27	67700
13	20400	28	81800
14	21500	29	94600
15	21800	30	148600

Table 4. Ranking of the best fit distribution

Maximum Likelihood Estimation		Maximum Likelihood Estimation	
Distribution	**Λ**	**Distribution**	**Λ**
Exponential (λ, η, γ) • λ=0,000032 h^{-1} • η=31699 h • γ=2100 h	**-340,92**	Weibull (β, η) • β=1,18 • η=35885 h	**-342,20**
Weibull (β, η, γ) • β=0,98 • η=31486 h • γ=2079 h	**-340,93**	Exponential (λ, η) • λ=0,00003 h^{-1} • η=33800 h	**-342,85**
Log-normal (μ, σ) • μ=10 h • σ= 0,96 h	**-342,18**	Normal (μ, σ) • μ=33799 h • σ=30992 h	**-352,81**

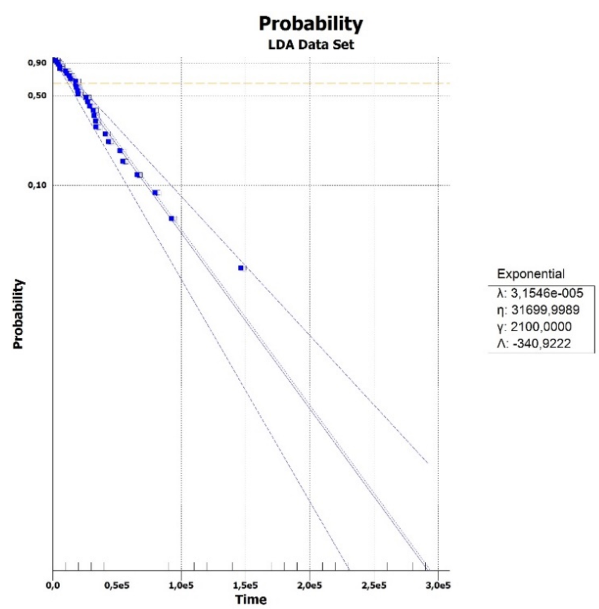

Figure 5. Two-parameter exponential probability plot

III. CONCLUSIONS

The focus of this article is the analysis of the trend of different statistic distributions.

The exponential distribution is widely used in reliability applications since it describes the constant failure rate section and it is used for component with a long useful life (e.g. electronic components). In all other cases, data generally has a non-constant failure rate trend and the most used distribution to describe it is the Weibull one: this is a very flexible distribution thanks to its parameters β conditioning the shape of the curves and η that extend or compress the curves.

The parameters that define all the distributions should be estimated using different methods and, in this study, the most common procedures were described: Probability Plotting, Least Square Estimation and Maximum Likelihood Estimation.

The LSE is an analytic method and it gives good outcomes in case of small number of samples.

The MLE is the most accurate technique and it is also suitable for large numbers of samples but has a high computational complexity and requires complex calculation; for these reasons it is the default method implemented on many software.

The final part of this study concerns two test cases on boards for automatic control and electronic components under accelerated test: the probability plot of all the distributions and a rank of the best-fit distributions were achieved by using a dedicated software.

The first case has been analyzed with both LSE and MLE methods and both confirmed that the data don't fit a constant failure rate distribution and that the Weibull and normal distribution provide the best-fitting.

The second case, instead, was assessed on an elevate number of samples and for this reason it was analyzed only with MLE procedure: the two-parameter exponential and the three-parameter Weibull resulted to be the best to fit the data. This test, thanks to the large number of samples, provides realistic results and it confirms that the Weibull distribution is very flexible and can describe a lot of life models although for this particular test case the distribution that better fits the samples is the two-parameter exponential one.

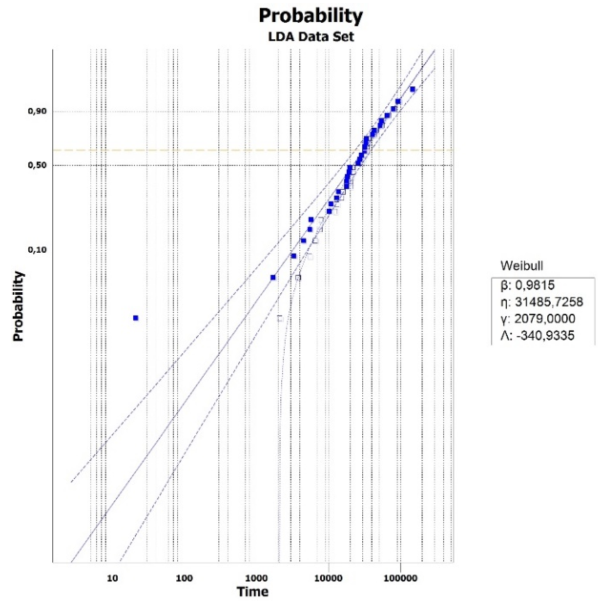

Figure 6. Three-parameter Weibull probability plot

REFERENCE

[1] Rausand M., Hoyland A., *System Reliability Theory*, John Wiley & Sons Inc. Publication, 2004

[2] M. Catelani, L. Ciani, M. Venzi, "Sensitivity analysis with MC simulation for the failure rate evaluation and reliability assessment", *Measurement*, Volume 74, October 2015, Pages 150-158, ISSN 0263-2241, http://dx.doi.org/10.1016/j.measurement.2015.07.003.

[3] N. Balakrishnan, M. Kateri, On the maximum likelihood estimation of parameters of Weibull distribution based on complete and censored data, *Statistics & Probability* Letters, Volume 78, Issue 17, 1 December 2008, Pages 2971-2975

[4] Meeker, W.Q., and Escobar, L.A., Statistical Methods for Reliability Data, John Wiley & Sons, Inc., New York, 1998

[5] Ivana Pobocikova, Zuzana Sedliackova, " Comparison of four methods for estimating the Weibull distribution parameters", Applied Mathematical Sciences, Vol. 8, 2014, no. 83, 4137-4149

[6] W.C. Navidi, Statistics for Engineers and Scientists, McGraw-Hill Higher Education, 2008.