Computer-aided reliability analysis of mechanical systems; some comments on estimation errors

Olivier Basile¹, Pierre Dehombreux², Enrico Filippi³

¹Faculté Polytechnique de Mons, Service de Génie Mécanique
Rue du Joncquois, 53 à B-7000 Mons (Belgique)
email : olivier.basile@fpms.ac.be, pierre.dehombreux@fpms.ac.be, enrico.filippi@fpms.ac.be

Abstract — Reliability is defined as the probability that a system will perform properly for a specific period of time under a given set of operating conditions. Our research is focused on reliability of mechanical equipment and accurate prediction of Mean Up Time (MUT) of the system. As this problem is quite difficult to solve analytically, two MATLAB libraries were developed: Reliabilitix and Simulatrix, based on numerical analysis of a block diagram description.

Generally, reliability distributions are estimated from data failures. That is why this article presents and compares two methods frequently used: the regression method derived from graphical estimation and the maximum likelihood estimator method. Next, we will discuss on sample dimension and errors committed with small set of failure times. Finally, this article points out that statistical models do not fit perfectly all the reliability distributions of systems, that is another source of errors.

Keywords — Mechanical Reliability, Maintenance, Reliability Estimation Errors, Maximum Likelihood Estimator

I. INTRODUCTION

Reliability is defined as the probability that a system will perform properly for a specific period of time under a given set of operating conditions. Our research is focused on reliability of mechanical equipment and accurate prediction of Mean Up Time (MUT) of the system. Its objective consists in the decrease of spare parts stock and improvement of maintenance plans.

Generally, reliability distributions are estimated from data failures collected on site. That is why literature is crammed with numerous probabilistic developments ([3], [4], [6], [8], [9], [10], [11] and [12]) sometimes completed by incidence analysis of distribution uncertainty on reliability parameters. By sample, Ebeling ([5]) have developed some methods to appreciate errors on Mean Time To Failure (MTTF) in the case of an exponential law. A more theoretical approach of errors induced by distribution estimation was developed in [16] and applied to fatigue test.

Determination of reliability characteristics of an equipment is a problem that become rapidly very difficult to solve analytically. That is the reason why two MATLAB libraries were developed: Reliabilitix and Simulatrix, based on numerical analysis of a block diagram description of the system. This article will present the principles used in these computer tools.

As already mentioned, reliability distributions are estimated from data failures. That is why this article presents and compares two methods frequently used which are a regression method derived from an usual graphical estimation on logarithmical ruled paper and the maximum likelihood estimator method (MLE). As an example, will be considered a bearing which reliability law is a Weibull one.

But, in industry, equipment must have a rather long mean up time otherwise it does not pay. This fact implies that during its life cycle the system will not fail frequently and than samples of failure times will not be very large. So, we wonder about: 'How many data are necessary to ensure convergence to real distribution life?' and 'What is the incidence of this uncertainty on reliability parameters as MUT or R(t)?'. In this article, an example is treated that illustrates errors occurring in reliability distribution and the incidence of this uncertainty on reliability parameters.

Finally, this article points out that standard statistical models are not able to fit perfectly the reliability distribution of a complex system, inducing new errors.

II. RELIABILITIX

Reliabilitix is used to compute reliability of a system that can functionally be described by a block diagram representing serial and/or parallel associations of components for which the standard (exponential, Weibull, log-normal, ... ) reliability distributions are known (fig. 1).

To analyse such a block diagram, Reliabilitix uses a method derived from graph theory known as the "latin composition" method or "concatenation" that allows to generate numerically the reliability function of the system ([7]). The first step of this method consists in associating a reliability network to the block diagram. After what, we have to find the minimal path-sets of the network ([7]). The first step of this method consists in associating a reliability network to the block diagram. After what, we have to find the minimal path-sets of the network ([7]). The first step of this method consists in associating a reliability network to the block diagram. After what, we have to find the minimal path-sets of the network ([7]). The first step of this method consists in associating a reliability network to the block diagram. After what, we have to find the minimal path-sets of the network ([7]).

Determination of reliability characteristics of an equipment is a problem that become rapidly very difficult to solve analytically. That is the reason why two MATLAB libraries were developed: Reliabilitix and Simulatrix, based on numerical analysis of a block diagram description of the system. This article will present the principles used in these computer tools.

As already mentioned, reliability distributions are estimated from data failures. That is why this article presents and compares two methods frequently used which are a regression method derived from an usual graphical estimation on logarithmical ruled paper and the maximum likelihood estimator method (MLE). As an example, will be considered a bearing which reliability law is a Weibull one.

Fig. 1. Block diagram example
Reliabilitix interface is presented in figure 2. In this menu, we have to introduce the block diagram file and its corresponding library file where are recorded components characteristics (name, failure distribution, how long the system operated, control, failure function in stock). Once reliability is calculated, results are presented in figure 3, the synthesized function is represented in the top window and can be easily compared to reliability components (bottom windows). This presentation allows to find in a look the critical components of the system. Furthermore, when it calculates the reliability function, Reliabilitix is able to take into consideration the substitution of defective components, the control of the system during a preventive maintenance operation and the ageing of the system, components and spare parts stock.

In practice, it is very rare that we dispose of the reliability distribution of the components because of their sensitivity to working environment. To solve this problem, each entreprise has to collect its own failure times and thus estimate reliability distribution of its own equipment. That is the reason why we have developed Simulatrix which is presented hereafter.

III. Simulatrix

Simulatrix (fig. 4) is a MATLAB library that estimates the system reliability simply on the basis of historical records of components failures and thus is not subject to analytical complexity. For this, we have first to define the mechanical system into a functional block diagram and a system library that includes for each component:

- the identification name,
- a file where are recorded failure times or its failure distribution,
- the maintenance type i.e. none, corrective or preventive (in last case we have to specify a systematic remplacement period),
- the stock quantity,
- the degradation distribution in stock,
- the repair time distribution,
- the number of repairs authorized,
- a systematic replacement period.

On the basis of the block diagram and library system, Simulatrix deduces the failure times for the system. A statistical analysis of these times will give us an estimated reliability model as exponential, Weibull, multi-Weibull, minimum value, maximum value, normal or log-normal. But, if we have not components failure times, Simulatrix can generate these using a Monte Carlo failures simulator.

Figure 5 presents typical results in Simulatrix. It is very easy to compare adequation between the real distribution, the estimated model and the synthesized distribution calculated by Reliabilitix.

IV. Failure time analysis

Once failure times of the system are generated, we have to look for a statistical model corresponding to these times. To this purpose two methods were implemented. The first, a regression method, is directly derived from graphical method using probability plotting; the second is an analytical method called the maximum likelihood estimator method (MLE).

A. Regression method

Probability plotting is a very useful technique that yields estimates of the distribution parameters and provides both a graphical picture and a quantitative estimate of how well the distribution fits the data. Basically, this method consists in transforming the cumulative density function (CDF) in a linear equation \( y = Ax + B \) that can be plotted easily. We illustrate this method with the Weibull function.
Concerning Weibull distribution, only the line that “best” fits the data, but also a measure numerical approach using least-squares fitting provides not evaluating by mean ranks or median ranks methods (see IV-B).

The regression method can be resolved manually but a computer program provides an estimate of the parameters. Let’s consider a random sample \( (T_1, \ldots, T_n) \) extracted from a continuous distribution \( F_T(t) \), and the ordered sample \( T_{(1)} \leq T_{(2)} \leq \ldots \leq T_{(n)} \) the sample sorted. This sequencing allows to define a distribution function \( F_n \) (or empirical cumulative density function) as:

\[
F_n(t) = \begin{cases} 
0 & \text{if } t < T_{(1)} \\
\frac{k/n}{t} & \text{if } T_{(k)} \leq t < T_{(k+1)}, \ k = 1, \ldots, n - 1 \\
1 & \text{if } t > T_{(n)}
\end{cases}
\]  

If we define : \( x = \ln(t) \), \( y = \ln \frac{1}{1 - F(t)} \), \( B = \beta \ln \eta \) and \( A = \beta \), equation 3 becomes:

\[
y = Ax + B
\]  

That means that if we have a Weibull model the data should fall roughly along a straight line. The CDF is calculating by mean ranks or median ranks methods (see IV-B).

The regression method can be resolved manually but a numerical approach using least-squares fitting provides not only the line that “best” fits the data, but also a measure of the goodness of fit. Concerning Weibull distribution, relations that link the slope \( a \) and the axis \( y \) intercept \( b \) to parameters \( \beta \) and \( \eta \) are \( \beta = A \) and \( \eta = \exp(B/\beta) \).

B. Cumulative density function estimations : mean and median ranks

Let’s consider a random sample \( (T_1, \ldots, T_n) \) extracted from a continuous distribution \( F_T(t) \) and \( T_{(1)} \leq T_{(2)} \leq \ldots \leq T_{(n)} \) the sample sorted. This sequencing allows to define a distribution function \( F_n \) (or empirical cumulative density function) as:

\[
R(t) = 1 - F(t) = \exp \left( - \left( \frac{t}{\eta} \right)^\beta \right)
\]  

\[
\ln \frac{1}{1 - F(t)} = \left( \frac{t}{\eta} \right)^\beta
\]  

\[
\ln \ln \frac{1}{1 - F(t)} = \beta \ln t - \beta \ln \eta
\]

If we define : \( x = \ln t, \ y = \ln \ln \frac{1}{1 - F(t)} \), \( B = \beta \ln \eta \) and \( A = \beta \), equation 3 becomes:

\[
y = Ax + B
\]

C. Maximum Likelihood Estimator Method

Suppose that \( T \) is a random variable with probability distribution \( f(t|\theta_1, \ldots, \theta_k) \), where \( \theta_1, \ldots, \theta_k \) are unknown parameters. Let \( t_1, \ldots, t_n \) be the observed values in a random sample of size \( n \). Considering the case we have complete data, then the likelihood function is:

\[
L(\theta_1, \ldots, \theta_k) = \prod_{i=1}^{n} f(t_i|\theta_1, \ldots, \theta_k)
\]

The objective is to find the values of \( \theta_1, \ldots, \theta_k \) that maximizes the likelihood function for given values \( t_1, \ldots, t_n \) because of the multiplicative form of the likelihood function, we will consider the maximum of the logarithm of this function. That maximum is obtained by taking the first partial derivatives respect to \( \theta_1, \ldots, \theta_k \) and settings these partials equal to zero.

\[
\frac{\partial \ln L(\theta_1, \ldots, \theta_k)}{\partial \theta_i} = 0 \quad i = 1, 2, \ldots, k
\]

C.1 Weibull maximum likelihood estimator

To reduce the number of parameters we will consider \( \gamma = 0 \) that occurs in most of the cases.

\[
f(t) = \frac{\beta}{\eta} \left( \frac{t}{\eta} \right)^{\beta - 1} \exp \left( - \left( \frac{t}{\eta} \right)^\beta \right)
\]
The likelihood function:

\[ L(\beta, \eta) = \prod_{i=1}^{n} \frac{1}{\eta} \left( \frac{t_i}{\eta} \right)^{\beta-1} \exp \left( -\left( \frac{t_i}{\eta} \right)^{\beta} \right) \]  

(9)

\[ \ln L(\beta, \eta) = \sum_{i=1}^{n} \left( \ln \beta - \eta \ln t_i - (\beta - 1) \ln t_i - \left( \frac{t_i}{\eta} \right)^{\beta} \right) \]  

(10)

Derivatives:

\[ \frac{\partial \ln L(\beta, \eta)}{\partial \beta} = \frac{n \beta}{\eta} - \eta^{\beta+1} \sum_{i=1}^{n} t_i^{\beta} = 0 \]  

(11)

\[ \frac{\partial \ln L(\beta, \eta)}{\partial \eta} = -\frac{n \beta}{\eta} + \beta \eta^{\beta-1} \sum_{i=1}^{n} t_i^\beta = 0 \]  

(12)

To find an estimation of \( \beta \) we have to solve equation (11). For this we use the Newton-Raphson method.

\[ \hat{\beta}_{j+1} = \hat{\beta}_j - \frac{g(\hat{\beta}_j)}{g'(\hat{\beta}_j)} \]  

(13)

\[ g(\beta) = \frac{\partial \ln L(\beta, \eta)}{\partial \beta} \]  

(14)

\[ g'(\beta) = \frac{d g(\beta)}{d \beta} = \frac{-n \beta}{\eta^2} - \sum_{i=1}^{n} \left( \frac{t_i}{\eta} \right)^{\beta} \ln^2 \left( \frac{t_i}{\eta} \right) \]  

(15)

This method is a recursive one so we need an initial estimation of \( \beta \) that may be obtained from a regression analysis (see IV-A).

From equation (12) we deduce:

\[ \eta = \sqrt{\frac{1}{n} \sum_{i=1}^{n} t_i^\beta} \]  

(16)

V. COMPARISON GRAPHICAL METHOD / MLE

Basile ([15]) highlights errors committed in the estimation of a distribution when we have a small set of data (i.e. less than 50 failure times) using the regression method estimation. Now, we will compare MLE and regression methods and see which one is the most accurate. Let’s consider a component (e.g. a bearing) which reliability law is a Weibull with \( \beta = 1.3; \eta = 40000; \gamma = 0 \) and \( \text{MUT}_{\text{theo}} = 39943 \) hours.

We intend to redetermining these parameters distribution. For that purpose, we will generate thanks Simulatrix 100 sets of 5, 10, 20, 50, 100 and 1000 failures. We get the results presented in table I. We note:

1. the MLE method converges more rapidly than the regression method to an accurate estimation of the parameters (50 failures against 1000),
2. estimations are not very precise with a poor collection of data (5,10,20 failures).

What is the incidence of this uncertainty on reliability parameters as MUT, estimation time to failure and reliability? First, we have to note that \( \beta \) and \( \eta \) follow normal distributions which means \( \beta \) and \( \eta \). Moreover we know that for a normal distribution 95% of the values are included in interval \( m \pm 2s \) (\( s \) is the standard deviation). It has been verified that reliability law is included in a variation field delimited by the four curves corresponding to extreme values of \( \beta \) and \( \eta \). This frontier allows us to determine errors on MUT, time to failure and reliability (fig. 6). Results are presented in tables II and III.

From these results, we note that the regression method and MLE method generate quite the same errors with an advantage for MLE method concerning determination of MUT.

VI. DEFINITION OF A DISTRIBUTION MODEL

Is it possible to find every time an exact distribution model corresponding to reliability law of a system which components have identical or different failures distributions? To answer this question we will study different series systems associating Weibull distributions.

A. Association of n identical Weibull

Suppose \( n \) components associated in series whose reliability laws are identical Weibull. We propose to determine
the reliability distribution of the system. We can prove easily that this distribution has the same \( \beta \) parameter that the components.

For example, the reliability of a system of 10 components in series which reliability law a Weibull with parameters \( \beta = 1,3 \) and \( \eta = 40000 \) (hours) results to be a Weibull with \( \beta = 1,3 \) and \( \eta = 6805 \) (hours).

B. Association of different Weibull distributions

First, let’s consider three components in series which reliability laws are Weibull ones with identical \( \beta \) parameter and different \( \eta \) parameter. To determine the reliability function of the system, we will consider :

<table>
<thead>
<tr>
<th>Component</th>
<th>( \beta )</th>
<th>( \eta ) (hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,3</td>
<td>4000</td>
</tr>
<tr>
<td>2</td>
<td>1,3</td>
<td>40000</td>
</tr>
<tr>
<td>3</td>
<td>1,3</td>
<td>400000</td>
</tr>
</tbody>
</table>

The synthesized reliability distribution of the system calculated by Reliabilitix is a Weibull with \( \beta = 1,3 \) and \( \eta = 3850 \) (hours). From this example (and others), we conclude that the distribution of such a system (same \( \beta \)) is a Weibull distribution too with same \( \beta \) than the components.

Secondly, let’s consider three components in series which reliability laws are Weibull ones with different \( \beta \) parameter and identical \( \eta \) parameter. In this case, we will consider the following values :

<table>
<thead>
<tr>
<th>Component</th>
<th>( \beta )</th>
<th>( \eta ) (hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0,8</td>
<td>40000</td>
</tr>
<tr>
<td>2</td>
<td>1,3</td>
<td>40000</td>
</tr>
<tr>
<td>3</td>
<td>1,8</td>
<td>40000</td>
</tr>
</tbody>
</table>

The reliability distribution that best fits the synthesized distribution is a Weibull law with \( \beta = 1,093 \) and \( \eta = 15150 \) (hours). The two distributions are plotted in the figure 7, the difference between the two distributions corresponds to an error of 1,55% on MUT. From this simple example, we conclude that it is not possible to find a standard model that perfectly fits the synthesized distribution.

Practically, this will be a general rule for real industrial systems which include components whose reliability is described by different families of laws as Weibull, log-normal, etc. As a consequence, it appears a new source of errors for reliability parameters estimation due to a deficiency of a standard reliability model. A system can not be modelled as accurately as its components and/or failure modes.

VII. Conclusion

The difficulty of determining reliability of a system imposes the use of computer tools. That is the reason why we have developed Reliabilitix and Simulatrix under MATLAB. Simulatrix algorithm (fig. 8) summarizes the methodology used to solve this problem.

![Simulatrix algorithm](image_url)

This flowchart points out that we need three types of data. The first type is relative to the system, the second to the components and the third to the maintenance. In practice, the system has to be described in a block diagram ; the
data relative to its components are the failure times or reliability distributions (Reliabilitix needs only reliability laws) and the maintenance parameters are the maintenance type (i.e. preventive or corrective), the number of reparations, the stock size and replacement policy by sample.

These tools allow to realize parametric studies of maintenance models and to appreciate errors affecting reliability distributions when deduced from small sets of sample and the incidence of this uncertainty on reliability parameters as MUT ; that was illustrated by an example in this paper.

Furthermore, this article points out that in most cases statistical models do not fit perfectly the reliability distribution of a system, that induces new errors on reliability parameters. The aggregation of these two errors makes illusory the search of a perfect reliability model for a system.

References