Determining the Similarity between Observed and Expected Ageing Behavior

P.A.C. Ypma
HAN University of Applied Science
Ruitenberglaan 26
Arnhem, 6826CC Netherlands

R. Ross
TenneT TSO & HAN University of Applied Science
Utrechtseweg 310
Arnhem, 6812AR Netherlands

Abstract- For quality control there is a need to quantify the similarity of observed and reference breakdown behavior. The present paper describes the background of a proposed similarity index, the significance of this similarity index (i.e. confidence limits of the similarity index that expresses how firm conclusions actually may be), the relation with child mortality statistics and how this similarity index may be used in practice.

I. INTRODUCTION

When insulated systems are aged in tests or in practice a distribution of times-to-failure may be found. Such a distribution will be characteristic for the failure mechanism(s) that caused end of life. If materials or components are known then their expected ageing behavior with a characteristic expected failure distribution may be known as well. If a batch of materials or components is produced and put to the test, then for this batch a failure distribution will be observed. The question is to what extent the observed distribution and expected distribution will be the same. The present paper discusses a method to quantify the similarity between the observed and expected failure behavior, namely the similarity index $S(t_1; t_2)$, which is defined as [1]:

$$S(t_1, t_2) = \frac{\int_0^t f(t) \cdot g(t) dt}{\int_0^\infty [f(t)]^2 + [g(t)]^2 - [f(t) \cdot g(t)] dt}$$

(1)

Here $f(t)$ and $g(t)$ are two failure density distributions that are compared. The interval $(t_1, t_2)$ is the evaluation range, which means the similarity can be explored over any arbitrary interval and does not need to cover the full range. The value of the similarity index ranges from zero to unity. If the distributions have nothing in common the similarity index is zero and if they are the same the similarity index is unity. Interestingly, the distributions do not have to belong to the same family. For instance the similarity between a Weibull distribution and a Lognormal distribution can be determined as well as between two exponential distributions.

The approach of comparing the similarity of distributions is quite new. Methods exist to compare data sets by their frequencies of occurrence, but these do not obviously lead to the equation above. Goodness-of-fit tests also provide a means to compare data with a reference distribution, but these generally presume a certain nature of the failure distribution, e.g. a Normal distribution. In weakest link ageing usually the data are distributed according to a lowest extreme value distribution like Weibull or a combination of such functions. The similarity index above can be applied to simple distributions, but also to complicated mixed distributions.

The similarity index enables to evaluate to what extent two distributions from two experiments are the same. A very useful application is to evaluate early failures of a batch of materials or components. E.g. there can be an issue whether the quality of the batch complies with the prescribed quality. The similarity index provides a means to quantify the confidence that the observed failure distribution agrees with the prescribed i.e. expected failure distribution. Fig. 1 shows an example. The green graph is the expected distribution with its 10% and 90% confidence limits for a batch of 87 samples. Six early failures are shown in red together with the best fit and its confidence limits. Although the early failures all fall within the green confidence intervals, the extrapolated best fit is very worrying. In this case the similarity index based on the interval $(0, 36)$ (just after the first six failures) is $S(0, 36) = 0.3$ and based on the interval $(0, \infty)$ the index is $S(0, \infty) = 0.09$. It means that at $t = 36$ days the confidence that the observed behavior is similar to the reference is 30%, which means the quality is doubtful. If the failure behavior continues according to the red graph, then in the end the similarity is only 9% and not compliant at all.
The similarity index provides a means to evaluate concerns and start timely actions. Further elaboration is found in Section V.

II. BACKGROUND

A well-known index in statistics is called the Jaccard index or the Jaccard similarity coefficient [2]:

\[ S_{\text{Jaccard}}(A, B) = \frac{|A \cap B|}{|A \cup B|} \]

in which \( A \) and \( B \) are sets and \(|A|\) is the cardinality of \( A \), c.q. the number of elements in \( A \). The Tanimoto index [3] or the extended Jaccard index compares vectors \( A \) and \( B \) and is given by:

\[ S_{\text{Tanimoto}}(A, B) = \frac{A \cdot B^\top}{A \cdot A^\top + B \cdot B^\top - A \cdot B^\top} \]

These indices have found applications mainly in the field of pattern recognition. A nice overview of possible similarity indices and distances in this field is given by [4]. Distribution functions like failure function, failure rate, hazard function and hazard rate can be embedded in an inner product function space [5]. In a similar fashion we could define an inner product similarity as a generalization of the two indices above by

\[ S_{\text{inner product}}(x_1, x_2) = \langle x_1, x_2 \rangle = \begin{cases} < x_1, x_2 > & < x_1, x_2 > + < x_2, x_2 > - < x_1, x_2 > \end{cases} \]

in which \( x_1 \) and \( x_2 \) are elements of an inner product space with the inner product being \( < x_1, x_2 > \) and with the restriction that \( < x_1, x_2 > \geq 0 \). In [1] it is proposed to use the failure density distribution \( f \) of which the inner product can be defined as:

\[ < f_1, f_2 >(t_1, t_2) = \int_{t_1}^{t_2} f_1(t) \cdot f_2(t) \, dt \]

where \( t_1 \) and \( t_2 \) define the time interval over which the inner product and the similarity index have to be evaluated. Applying the inner product of (5) to (4) yields the similarity index of (1).

III. SINGULARITY ISSUES

A singularity occurs where a function has an infinite value. This happens for instance with delta functions \( \delta(x) \) at \( x = 0 \). At singularities (5) may become infinite and can cause special behavior of the similarity index. A well-known singularity of the probability density function also occurs at \( t = 0 \) for Weibull distributions with the shape parameter \( \beta < 1 \). For Weibull the cross term \( f_1 f_2 \) in (5) also features a singularity at \( t = 0 \) when \( \beta_1 + \beta_2 \leq 2 \). An integration over a singularity may not be problem as the Weibull cumulative function exists also for \( \beta < 1 \). But a discontinuity appears with \( \beta_1 + \beta_2 \leq 1 \) in:

\[ < f_1, f_2 > (0, t) = \frac{\beta_1 \beta_2}{\alpha_1 \alpha_2} \int_0^t t^{\beta_1 - 1} \cdot e^{-\alpha_1 t} \cdot t^{\beta_2 - 1} \cdot e^{-\alpha_2 t} \, dt \]

It can be proven that the corresponding \( S(0, t) \) still will have values in \([0, 1]\) but jumps to 0 for \( \beta_1 \neq \beta_2 \) and \( \beta_1 < \frac{1}{2} \) or \( \beta_2 < \frac{1}{2} \). This discontinuity problem seems to be resolved by using

\[ S(t_1, t_2) > 0 \text{, but the result will be very sensitive for } t_1 \text{ when this value is small compared to the observed failure times. As } \beta_1 \neq \beta_2 \text{, the similarity index may rightly be } 0 \text{ though.}

Identifying reference and observed child mortalities with \( \beta_1 + \beta_2 \leq 1 \) would yield a discontinuous \( S \) if at least one \( \beta < \frac{1}{2} \) as mentioned above (and \( \beta_1 \neq \beta_2 \)). However, if the child mortality actually is rather due to a fast ageing subpopulation with \( \beta > 1 \) combined with a too small \( \alpha \) (as often occurs in practice), then there is no singularity issue and no discontinuity occurs either.

An example is given in Fig. 2. Here is a mix of two subpopulations: 2% child mortality with \((\alpha, \beta) = (1.75, 4)\) and 98% wear-out with \((\alpha, \beta) = (4000, 4)\). Both subpopulations also suffer from a random failure effect with \((\alpha, \beta) = (1000, 1)\). Fig. 2 shows the resulting bath tub curve above and below the distribution for the child mortality fraction \((F_1)\), for the correct wear-out fraction \((F_2)\) and the random failure \((F_3)\). \( F_{\text{total}} \) is the combined distribution for the total group. The similarity index for \( F_{\text{total}} \) and \( F_2 \) and \( F_2 \) is free from discontinuities.

IV. THE STATISTICS OF THE SIMILARITY INDEX

The more failures are observed, the more accurate the observed distribution is defined and the better the similarity between observed and reference distribution can be determined. The question is how accurate the similarity index is with a limited set of observed failure data. This was explored by finding the confidence intervals as a function of percentage failed population. The calculation of \( S \) is carried out by first defining two distributions, e.g. a reference distribution defined by known parameters and an observed distribution defined by estimated parameters. In case of early evaluation of the similarity index the observed distribution is based on parameter estimation with a censored data set, i.e. only a part of the failure data is known.
In order to explore how the significance (or confidence limits) of the similarity depends on the completeness of observed data, a Monte Carlo simulation is carried out to generate datasets based on the reference distribution. In the simulation these sets can be censored either in time or in number of observations. Given the nature of the inner product defined in (5) only time censoring is applied in the simulations based on the choice of the time interval \([0, t_c]\) that has to be evaluated. For each censored data set the parameters of the observed distribution are estimated. Next the similarity index of the reference distribution with each of the observed distributions is determined. In this way a distribution of similarity indices is found. From this distribution the confidence level of the similarity index can be determined. The procedure then is as follows:

1) Define \(n\), the total number of samples of which the failure time has to be observed.
2) Define the reference distribution \(f_{\text{ref}}(t)\).
3) Define the censoring interval, mostly \(0 < t < t_c\). That means that after \(t = t_c\) no failure times were observed.
4) Based on the observed failure times within the censoring interval, estimate the parameters of the observed distribution. In this paper the parameter estimation method proposed by [6] is used.
5) Calculate \(S_o(t)\) with respect to \(f_{\text{ref}}(t)\) and with the estimated parameters.
6) Generate a large number of sets of failure times with size \(n\) by Monte Carlo simulation based on the reference distribution. For this paper 16000 sets are generated for each experiment.
7) Estimate for each set, based on the generated failure times within the censoring interval, the parameters of each 'observed' distribution. Use the same parameter estimation method as in step 4.
8) Calculate for each set \(S_o(t)\) with respect to \(f_{\text{ref}}(t)\) and with the corresponding estimated parameters.
9) Create the statistics of \(S_o\). This can be done by estimating the cumulative density distribution \(F_S(S_o)\).
10) Estimate the confidence level of \(S_o(t)\) as \(F_S(S_o(t))\).

**V. EXAMPLES**

As a first example the similarity index is applied to data found in [7] as shown in Fig. 3. The total number of samples is \(n = 10\). The parameters of \(f_{\text{ref}}\) are shown in Table I. The given procedure is followed until step 6 for \(t_c = 60\) minutes (after 5 failure times) and for \(t_c = 80\) minutes (after 8 failure times) respectively. The values of the similarity index can be found in Table I. The values of \(S\) are very close to 1. This confirms what already can be seen in Fig. 3. Namely that the estimations from the censored data (red and green lines) are very close to the reference (black line).

The second example is the dataset obtained from the example presented in Fig. 1. Now the complete procedure is followed for different censoring intervals. The total number of samples is \(n = 87\) and the parameters of \(f_{\text{ref}}\) are shown in Table II together with the results. The cumulative distributions of \(S\) for the different censoring intervals are shown in Fig. 4. The grey areas indicate the value range outside 10% and 90% limits. Note that \(F_S(S_o)\) (see Table II) is in fact the confidence level of \(S\) of the observed data. So e.g. for \(t_c = 92\) days the calculated \(S(0, 92) = .66\) corresponds to a confidence level of 19%, so 81% of all possible observations with the same \(t_c\) and coming from the reference distribution will have a higher \(S\)-value. Furthermore the values of \(S\) for \(t_c\) of 40, 92 and 175 days are within in \(S_{0.95}\) and \(S_{0.05}\) limits. Only the for \(t_c = 4750\) days the \(S\)-value is above the \(S_{0.05}\) limit. This value could therefore be regarded as a remarkable good result.

![Fig. 3. Example data from [7] (see text)](image1)

![Fig. 4. Plot of \(S(0, t_c)\) and its statistics from data of Fig. 1](image2)
The third example data are given in Fig. 5. Now the trend of the first failure times that are still within the confidence limits of the reference distribution, continues and the later failure times fall outside these confidence limits. The results are given in table III. Here can be seen that the confidence levels of the corresponding $S$-values all are below the 10% limit and that they are even decreasing for larger values of $t_e$. These confidence levels are that small that a graph equivalent to Fig. 4 is omitted. This decreasing level of confidence means that the found similarity index becomes less and less explainable from the reference distribution, i.e. the observed distribution cannot be explained from scatter within the reference distribution. The observed failure behavior thus strongly deviates from the reference and is not compliant. This confidence limit of 10% may be taken as a level of significance of the similarity index.

VI. DISCUSSION

In this paper the concept of the similarity index $S$ based on an inner product function space, is explained and demonstrated. The statistics of $S$ enables the estimation of a confidence level, up to which an observed failure behavior can be regarded to be from the reference distribution or not with respect to $S$.

In general the statistics of $S$ can be estimated by Monte Carlo simulations which is practically seen not a big problem with the computers and the random number generators at the state of the art of 2016. Nevertheless it seems to be possible to find explicit expressions for these statistics at least in approximation and only in special cases. This topic will be discussed in another paper.

For parameter estimation only the least square method is applied in this paper as proposed in [6], without any bias correction. Because the similarity index, as defined in this paper, uses estimated parameters, it will be of interest to investigate the influence of changing the estimations method on the similarity index like maximum likelihood methods [8] and other least square methods [9].

REFERENCES