

Determination of Prediction Intervals for a Future Number of Failures: A Statistical and Monte Carlo Approach

Fortunato S. de Menezes, Mário J. Ferrua Vivanco,

Departamento de Ciências Exatas, Universidade Federal de Lavras, C.P. 3037, 37200-000, Lavras, MG, Brazil

and Luiz C. Sampaio

Centro Brasileiro de Pesquisas Físicas, Rua Dr. Xavier Sigaud, 150, 22290-180, Rio de Janeiro, RJ, Brazil

Received on 7 September, 2005

In this work, we present a new procedure, called sub-sampling, to obtain data concerning time of failure in trials without replacement, (NRT). With this data it is possible to determine the prediction interval (PI) for the future number of failures. We also present an *alternative* way to evaluate the coverage probability of the prediction interval (PI). The results presented show that the method proposed is reliable and can be useful for the statistical analyses of quality control of processes.

Keywords: Sub-sampling; Weibull Distribution; Prediction Intervals; Coverage Probability; Monte Carlo Simulation

I. INTRODUCTION

A. Motivation

The determination of prediction intervals for random quantities has been subject of great interest and intensive research for a long time. Its importance is revealed when one addresses the question: If in a system with N components, X quantities failed in the time interval $[0, T_c]$, how many components will fail in a given future time belonging to the interval $[T_c, T_w]$? or, even more, what is the prediction interval (PI) for the number of components Y that it will fail in the time interval $[T_c, T_w]$? The prediction interval $PI(100(1 - \alpha)\%)$ in the time interval $[T_c, T_w]$ with the confidence level $(1 - \alpha)$ is obtained from the past data of number of components X which have failed in the time interval $[0, T_c]$. We illustrate those questions clearly on Fig. 1.

There are many systems in which the prediction of times of failure of parts is critical, and the implications range from the canonical damage to safety issues. Most of the work which has been developed until now has considered *only* trials with replacement, namely (RT), where the quantities tested are replaced in the sample.

Some authors Meeker and Escobar(1998) [5], Rostum(1999) [10], Nelson(2000) [7] and Nordman and Meeker(2002) [9] have already discussed this subject. In the last paper, Nordman and Meeker were motivated by the problem of large heat exchangers, that transfer energy to cool down the Nuclear power plants. Due to stress and corrosion the heat exchangers crack over time. Such cracks are detected during planned inspections and the heat exchangers are replaced, or alternatively, isolated from the others. New tubes have to be added then after a given time period, when the number of failures is above a critical value.

It is possible then to measure “the time until the failure of a component” (e.g., the crack or stress of heat exchangers). The determination of such “time until the failure of a component” is of great importance to determine prediction intervals (PI).

Usually, the components are replaced in the sample once

they are tested, as long as they have passed in the test. In other words, the components regarded as good according the chosen test are put again the sample. In this way, the good components continues in operation and have a chance to be chosen again in a future time to check its performance and decide whether or not it has failed.

However, there is a class of problems where the replacement of good components is not relevant. Suppose we want to acquire a lot (or sample) of components, with the previous knowledge that the sample has a certain level of failed components. Furthermore, we want to acquire the lot as long as the percentage of failure in the sample is less than a given pre-established value. In this work we have developed a method to determine times of failure and, consequently, prediction intervals (PI) when the experiments are such that the components used for testing are not replaced in the original sample, even though the result of components tested is negative (e.g., the components tested do not fail). Such method is based upon the measure of “time until the failure of a component” using a new procedure called: sub-sampling.

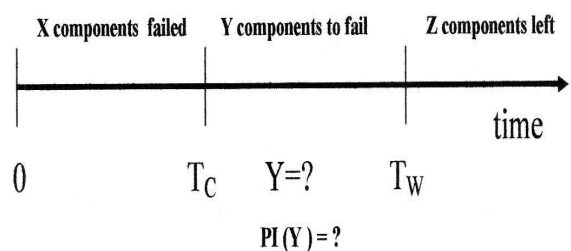


FIG. 1: Illustration of the number of components (Y) to fail in the time interval $[T_c, T_w]$ given that X components failed in the previous time interval $[0, T_c]$ and prediction interval (PI).

We shall consider that Non Replacement Trials (NRT) (or destructive experiments), as the name states, do not replace each component after evaluation (or testing). Nevertheless, we have to bear in mind that the number of components not

replaced in the original sample is a very small fraction of the sample, from where we want to predict something, such that the system as a whole does not change significantly. This is the key point in this procedure and here lies its importance in the prediction intervals (PI).

In the section I.B we present the state of the art concerning the prediction of the future number of failures of components, and the following section I.C shows the purposes to be fulfilled concerning the Non Replacement Trials (NRT) (or destructive experiments). In the section II the Methods are presented, with the sub-sampling strategy (section II.A), the theory of prediction intervals (section II.B), the data generation (section II.C) and coverage probability (section II.D). The results and discussion are presented in the section III, along with the probability distribution which best fits the data generated by the sub-sampling procedure (section III.A), the prediction intervals for the problem posed (section III.B) and the alternative coverage probability (ACP) which shows that the method proposed is reliable (section III.C). Finally, the conclusions and the outlook are presented in section IV.

B. State of the art

Results of research about the prediction of random quantities have enormous potential for application in engineering and are, indeed, of fundamental importance.

A preliminary study was presented by Nelson(1972) [6]. He established methods to build confidence intervals and hypothesis tests for two multinomial proportions. Although the procedures in such work were applied in studies of consumers preferences and election researches, this work was crucial to develop another work published in the year 2000 [7], which has a direct application in engineering.

Meeker and Escobar(1998) [5], developed a method to determine the prediction limits (upper and lower) for the future number of fails (Y) in the time interval $[T_c, T_w]$. Such procedure is based upon the *conditional binomial distribution* of Y given that X components have failed in the time interval $[0, T_c]$.

Rostum(1999) [10], developed statistical models to predict the state of the pipelines in a network of water distribution. Starting from the historical of the past failures in a network of water supply it was possible to predict the future number of failures in each network. These predictions were then used to make decisions about maintenance of the water network. In other words, with his work it may be possible to answer about the following question: Shall we repair points of failure in the network (or sub-network) of water supply, or shall we change an *entire pipeline* in a given sub-network ?.

Nelson(2000) [7] has established three procedures for the prediction intervals of the number of tubes to fail in components of heat exchangers, namely (I) the procedure of ratio of probabilities, (II) the procedure of ratio of probabilities simplified and (III) the likelihood ratio procedure.

Nordman and Meeker(2002) [9], evaluated the coverage probability for each one of the procedures proposed by Nelson(2000) [7], concluding that the more appropriate is the

likelihood ratio procedure.

Two aspects must be emphasized in relation to the studies made by Nelson(1972) [6], Nelson(2000) [7] and Nordman and Meeker(2002) [9]:

1. The distribution considered for the variable time until the failure of a component is the Weibull distribution.
2. The tests that evaluate *whether or not* a tube is deteriorated **do** replace the tube when it is failed.

By contrast, in Agricultural Sciences, particularly in the area of quality control of seeds, we find in the literature some authors Ellis(1984) [1], who suggested the time until a seed has failed follows a Normal distribution. Furthermore, the usual experiments which evaluate whether or not a seed is failed are such that the sub-samples tested are not replaced in the system, Marcos Filho(1987) [4].

Methods to determine the time of failure of a component (and its respective probability distribution) and the prediction interval (PI) of the future number of failures, for experiments in which the components are **not** replaced in the original sample – as the ones done on seeds, were not found in the literature. This research provides a contribution to fill this lack.

C. Purposes

Considering that exist a problematic for **Non Replacement Trials (NRT)** (e.g., destructive trials), this work has the following objectives:

1. Establish a sampling procedure which allows to obtain data for the variable: "time until the failure of a component".
2. Determine the probability distribution for the variable: "time until the failure of a component".
3. Determine the prediction interval for the future number of failures.
4. Determine the Coverage Probability (CP) of the prediction interval (PI).

II. METHODS

Companies that manufacture **fuses** (a protection device for safeguarding electric circuits) and the shop dealers, quite often face with the problem for the choice of lots which they intend to acquire. The procedure presented in this section will be related with the problem of the quality control of fuses, where the tests applied to know whether or not a fuse is deteriorated, are such that the fuses tested are not replaced in the sample which they came from, even though some of the fuses are good (or have passed in the quality test).

As we are dealing with **Non Replacement Trials (NRT)**, we will evaluate group of fuses rather than one fuse itself. In this work, such group of fuses will be called SUB-SAMPLES.

Therefore, the goal of this research will be evaluate the viability of acquiring a large lot of fuses, starting from the prediction of the *number of sub-samples of fuses failed in a given time interval* (a random quantity). In other words, we start from the test of a small sub-sample of fuses which were taken from a larger sample. With the result which comes up from the test, we can predict what it will be the number of sub-samples of fuses that it will fail in a **future** time interval, thereafter the previous time interval where the test was performed.

The prediction will be made regarding the information obtained on the number of sub-samples of fuses which were deteriorated in a time window interval $([0, T_c])$ between the starting of collection of sub-samples tested (unit of time $T = 0$) until a certain given censured time ($T = T_c$), which is pre-established. The study requires the following steps:

- (i) the determination of time until a sub-sample of fuses is regarded as deteriorated,
- (ii) the determination of a probabilistic distribution of time until the deterioration,
- (iii) the determination of the prediction interval (PI) for the future number of sub-samples to deteriorate and,
- (iv) the evaluation of the coverage probability of such prediction interval.

We shall notice that *although* the procedure to be explained will be applied for the case of **fuses, the method proposed can be applied to any problem in which the sub-samples of components tested are not replaced in the larger sample from where they came from.**

A. Sub-sampling method

As the tests to determine whether or not a fuse is deteriorated are Non Replacement Trials (NRT), it is not possible to apply directly the methods proposed by the authors mentioned in the section IB, especially when we wish to know the data related to "time until the deterioration of a fuse (or sub-sample of fuses)" (and its respective probability distribution). To overcome this problem the following method, called **SUB-SAMPLING**, is proposed (see Fig. 2)

Let us suppose that from a population of N fuses which are packed for sale (illustrated in Fig. 2), we pick up M samples, where each one of them contains n fuses each. From each one of these M samples, we pick *randomly and systematically*, sub-samples of size n_1 . In other words, we now have M sub-samples. In each one of these M sub-samples we apply the quality test to check the performance of the fuse. Such test is applied in a sub-sample of size n_1 (taken from the sample of size n) to determine the percentage of deteriorated fuses (p_{det}). If the **percentage of the deteriorated fuses** (p_{det}) in a given sub-sample of size (n_1) is **greater or equal** than a given value pC (e.g., $p_{det} \geq pC$), which is pre-established, then the whole sub-sample of size n is regarded as **deteriorated** and we register the time in which the test was performed. From

Fig. 2 we can also see that a sub-sample regarded **deteriorated** (e.g., when p_{det} in a given sub-sample of size n_1 has the value $p_{det} \geq pC$) implies to stop of picking more sub-samples of size n_1 from the respective sample. In other words, **the failure of a given sub-sample of size n_1 implies the failure of the respective sample**. A sub-sample regarded **not deteriorated** (e.g., when p_{det} in a given sub-sample of size n_1 is **smaller** than pC ; $p_{det} < pC$) implies in the selection of a new sub-sample of size n_1 , from the sample left. From Fig. 2 it is clear that the sample left now contains $n - n_1$ components, since the first n_1 components which were evaluated in the time unit (say $t = 1$) are set aside once $p_{det} < pC$. Therefore, we apply the "fuses quality test" on each new sub-sample of size n_1 and we check p_{det} in each new sub-sample of size n_1 . In one hand, for those samples in which the sub-samples presents p_{det} which DOES overcome the value pC we record the time in which the test was applied and regarded the whole sample as deteriorated for future tests (for instance, the 2nd sample (column) for the time unit $t = 2$ on Fig. 2). On the other hand, in those samples, in which the p_{det} DOES NOT overcome pC , we continue to apply the sub-sampling and thereafter the "fuses quality test". The procedure continues until the percentage of deterioration in the sub-sample overcomes the value pC (where the time recorded for the correspondent sample will be the time in which the test was applied and provided $p_{det} \geq pC$; for instance, the 1st sample (column) at time unit $t = 3$, see Fig. 2) or until we finish with the fuses in a given sample (in such case, the time to be registered will be a censor data).

In this way we can obtain a set of M data, correspondent to the set of time units until the deterioration of each one of the M samples (see Fig. 2), from which we will adjust a probability distribution.

B. Prediction Intervals

Once the time until the deterioration of the sub-samples has been obtained as well as its respective probability distribution, the following step will be to determine a prediction interval (PI) for the "number of deteriorated sub-samples, in each pack, in a future time interval". The Fig. 3 illustrates the procedure:

Let us suppose that one package of fuses contains a total of M samples. It is possible to show that the variables X , Y and Z follows a Trinomial Distribution, such that $X + Y + Z = M$, with the following probabilities,

- p = probability of occurrence of a failed sub-sample on the time interval $[0, T_c]$.
- q = probability of occurrence of a failed sub-sample on the time interval $[T_c, T_w]$.
- r = probability of occurrence of a failed sub-sample after T_w .

According to Nelson(2000) [7], the prediction interval (PI) for the random variable Y (number of sub-samples to deteriorate in the time interval $[T_c, T_w]$), might be obtained through

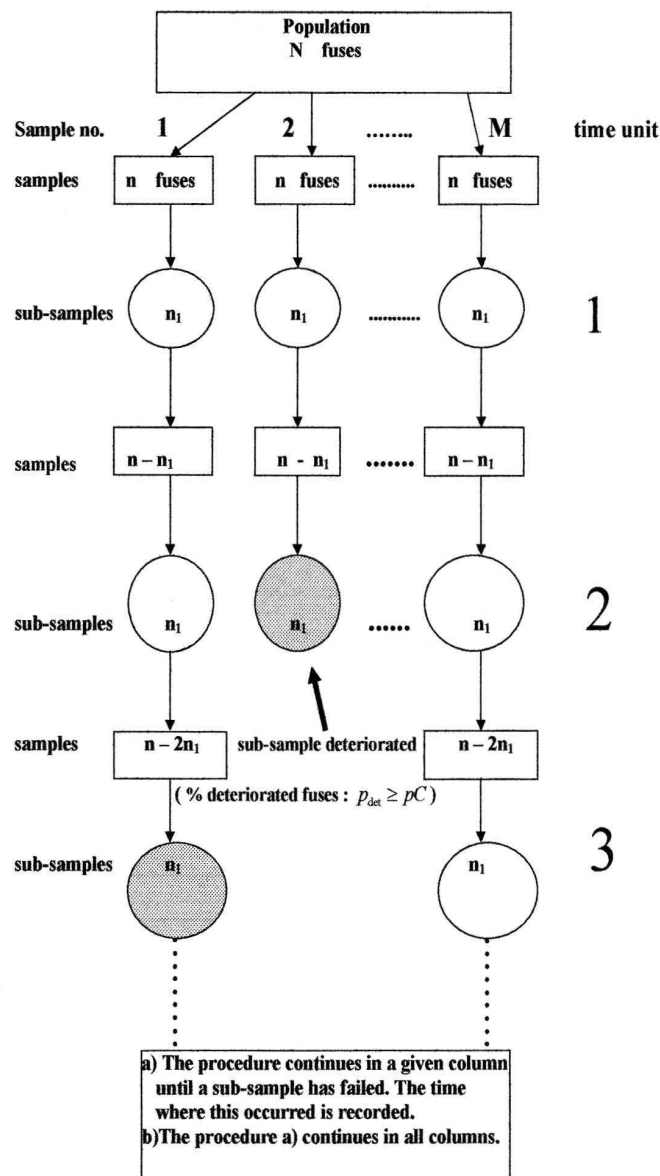


FIG. 2: Representation of the sub-sampling procedure.

the likelihood ratio procedure in the form described in the following:

For the **unrestricted** model the likelihood function (multinomial) will be,

$$L(p, q; x, y) = \frac{M!}{x!y!(M-x-y)!} p^x q^y (1-p-q)^{M-x-y} \quad (1)$$

where the maximum will be given by,

$$L^*(p, q; x, y) = \frac{M!}{x!y!(M-x-y)!} \left(\frac{x}{M}\right)^x \times \left(\frac{y}{M}\right)^y \left(1 - \frac{x+y}{M}\right)^{M-x-y} \quad (2)$$

The values of p , q and r , in the **restricted** model, will depend of the data distribution. For instance, the Weibull probability density has the following form,

$$f(t) = \left(\frac{\beta}{\eta}\right) \left(\frac{t}{\eta}\right)^{\beta-1} \exp(-(t/\eta)^\beta) \quad (3)$$

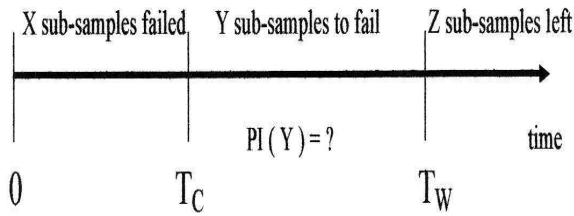


FIG. 3: Illustration of the prediction interval ($PI(Y)$) for the time interval $[T_c, T_w]$.

with accumulated distribution function

$$F(T) \equiv \int_0^T f(t) dt = 1 - \exp(-(T/\eta)^\beta) \quad (4)$$

In other words, if we suppose a Weibull distribution (see Nelson (2000) [7]) with parameters η (scale parameter) and β (shape parameter), for the random variable: time until the deterioration of a sub-sample of fuses, then we have

$$p \equiv \int_0^{T_c} f(t) dt = 1 - \exp(-(T_c/\eta)^\beta) \quad (5)$$

$$q \equiv \int_{T_c}^{T_w} f(t) dt = \exp(-(T_c/\eta)^\beta) - \exp(-(T_w/\eta)^\beta) \quad (6)$$

$$r \equiv 1 - p - q = \exp(-(T_w/\eta)^\beta) \quad (7)$$

Therefore, for a β known (in whatever way), the **restricted** likelihood function to the Weibull distribution will be,

$$K(\eta; x, y) = C \left(1 - e^{-(T_c/\eta)^\beta} \right)^x \times \left(e^{-(T_c/\eta)^\beta} - e^{-(T_w/\eta)^\beta} \right)^y \times \left(e^{-(T_w/\eta)^\beta} \right)^{M-x-y}, \quad (8)$$

with C given by, $C = \frac{M!}{x!y!(M-x-y)!}$

Since that the value of Y is unknown, in order to maximize $K(\bullet)$ with respect to the scale parameter η , it is needed to run all the values of Y between 0 and $M - X$. Each value of Y in the interval $[0, M - X]$ will provide a value of η and, therefore, a maximum value of $K(\bullet)$, that we will call $K^*(\bullet)$.

It is known that the statistics of likelihood ratio given by,

$$Q(x, y) = -2 \mid \log(K^*(x, y)) - \log(L^*(x, y)) \mid \quad (9)$$

distributes itself as a $\chi^2_{(1)}$. Therefore, for any value of the random variable Y , all the values of $Q(X, Y = y)$ generated, will have identical distributions ($\chi^2_{(1gl)}$). So, it will be possible

to generate a prediction interval (PI) for Y considering the following probability,

$$P(Q(x, y) \leq \chi^2_{(1; \alpha)}) = 1 - \alpha \quad (10)$$

In other words, every value of Y which satisfies $Q(x, y) \leq \chi^2_{(1-\alpha)}$ is contained in the prediction interval with a level of confidence $100(1 - \alpha)\%$. Consequently, the lowest value of $Y(Y_{inf})$ and the highest value of $Y(Y_{sup})$ which satisfy Eq. 10 will be considered the lower and upper limits, respectively, of the prediction interval ($PI = [Y_{inf}, Y_{sup}]$).

C. Data generation

The data of the variable time until the deterioration of the sub-samples were generated through Monte Carlo simulation, following the procedure explained in the section II A. For this, it was considered a pack with $N = 120000$ fuses which was divided in $M = 120$ samples of size $n = 1000$. It was considered the values of 1% and 6% for the initial percentages of deterioration in each one of the M samples, since it is supposed that in the beginning of the storage each pack of **fuses** contains a large fraction of **good** fuses but it may contain a certain fraction of failed fuses. The values of 1% and 6% correspond to the fraction of failed fuses in each one of the $M = 120$ initial samples of size $n = 1000$. Afterwards, it was picked out sub-samples of size $n_1 = 10$, in each one of the M samples (see Fig. 2). In the 1st sub-sampling, in each one of the M sub-samples of size n_1 it was made a counting of the number of failed fuses. In the case where the percentage of failed fuses were $p_{det} \geq pC$, where $pC = 15\%$ (for 1% and 6% of initial deterioration of fuses (p_{det}^{ini}) in the sample of size M) or $pC = 50\%$ (for $p_{det}^{ini} = 6\%$), the sub-sample was regarded as failed and it was recorded the value 1 for the respective sample (in other words, the respective sample has failed in the time unit $t = 1$). On the other hand, if the percentage of failed fuses is $p_{det} < pC$ we would pass to the 2nd sub-sampling. Regarding that the percentage of deteriorated fuses increases with time (since we set aside sub-samples of size n_1 which contains **good** fuses and eventually few **failed** fuses) we have to recalculate the percentage of **good** fuses, before the 2nd sub-sampling. This new percentage ($100p^*\%$) of **good** fuses, was obtained subtracting the number of good fuses in 1st sub-sample discarded (of size n_1), from the number of the good fuses existent in the initial sample (of size n) and dividing the result by the number of total fuses in the remained sample (which in the case will be $n - n_1 = 990$ fuses, see Fig. 2). Following this procedure, the percentage of **deteriorated** fuses in the sample left (of size $n - n_1$) will be ($p_{det} = 100(1 - p^*)\%$). In other words, in the 2nd sub-sampling, the source sample will now have $n - n_1 = 990$ fuses, with $100p^*\%$ of **good** fuses and ($p_{det} = 100(1 - p^*)\%$) of **deteriorated** fuses. The 2nd sub-sample of size n_1 is taken (see Fig. 2) from the source sample (of size $n - n_1$), and the comparison between p_{det} (evaluated from this sub-sample size of n_1) and pC is made.

The procedure above was adopted in each sub-sampling until a given sub-sample n_1 has failed. The unit of time in which the sub-sample has failed was regarded as the data for the variable "time until the deterioration of the sample" (the entire column of Fig. 2). At the end of each simulation we end up with a set of 120 data, one for each one of the $M = 120$ samples (or columns of Fig. 2). With this set of 120 data, we have adjusted the probability distribution.

D. Coverage probability

Considering that, given $X = x$, the conditional distribution of the random variable Y is binomial with the parameters $(M - X, \pi = \frac{q}{q+r})$, M defined in the section II A and with q and r defined in Eqs. (5)-(7); we have the value of π given by,

$$\pi = 1 - \exp[(T_c/\eta)^\beta - (T_w/\eta)^\beta] \quad (11)$$

In the work of Nordman and Meeker(2002) [9] it was presented a suggestion given by Wayne Nelson in which the coverage probability (CP) for the prediction interval (PI) with a nominal level of confidence of $100(1 - \alpha)\%$ may be determined by the average of the probabilities $P(Y_{inf} \leq Y \leq Y_{sup}|X)$ found for each value of X . In other words,

$$CP(PI(100(1 - \alpha)\%)) = E_x[P(Y_{inf} \leq Y \leq Y_{sup})] \quad (12)$$

The expression showed on Eq. 12 is true and determine the coverage probability. However, its value is computationally difficulty to obtain for two reasons:

- The procedure of maximization of Eq. 8 generates a maximum likelihood estimation for the parameter η for each value of Y . Therefore, as π depends on η (see Eq. 11), we would have a value of π for each possible value of Y .
- The discrete values of Y that are found inside the prediction interval $PI(Y) = [Y_{inf}, Y_{sup}]$, determined according to Eq. 10, are not necessarily contiguous. In other words, the prediction interval $PI(Y) = [Y_{inf}, Y_{sup}]$ may be composed by sub-intervals not contiguous.

An alternative analysis to evaluate the coverage probability is presented in this work. Such analysis is based upon the counting of values of Y (obtained through simulation, according the procedure presented in Section II A and II C) which are inside the **theoretical** prediction interval $PI(Y) = [Y_{inf}, Y_{sup}]$ (which was obtained according the procedure showed in the Section II B).

The percentage of values of Y which are inside the prediction interval $PI(Y) = [Y_{inf}, Y_{sup}]$ will be regarded as the **alternative coverage probability (ACP)**.

The results of such analysis are presented in Section III C.

TABLE I: Values of β (shape parameter of Weibull distribution) obtained from simulations with the number of samples $M = 20000$. p_i denotes **initial** percentage of deterioration in one sample; pC denotes the **maximum** allowed deterioration in one sub-sample of size n_1 ; and β , the shape parameter of Weibull distribution. In all calculations $n_1 = 10$.

Combination	p_i	pC	β
1	1%	15%	2.7785
2	6%	15%	1.7935
3	6%	50%	3.8542

III. RESULTS AND DISCUSSION

A. Probability distribution

According the procedures detailed in the Sections II A and II C it was performed 6000 simulations for each pair of combination (p_i, pC) (where, p_i , is the initial percentage of deterioration and pC , the maximum acceptable percentage of deterioration in each sub-sample of size n_1) (See Table I). Those simulations generate, each one, $M = 120$ data correspondent to the variable: "time until the failure of the sub-sample" (or sample, as it was emphasize in the Section II A). With those data it was established a probabilistic distribution which offers the best fitting.

From Eq. 4, and using the survival function ($S(T)$) estimated according to Kaplan and Meier (1958) [11],

$$S(T) \equiv 1 - F(T) = \exp(-(T/\eta)^\beta) \quad (13)$$

we can easily see that,

$$\log(-\log(S(T))) = \beta \log\left(\frac{T}{\eta}\right) \quad (14)$$

Fig. 4 presents the illustration of 6 graphs which represents 6 data sets selected randomly among all the simulations performed. These graphs clearly show that the data sets can be adjusted by the Weibull distribution, what allows us to use the methodology to estimate the prediction intervals presented in the Section II B.

The linear tendency of the graphs observed in Fig. 4 demonstrates that the Weibull distribution adjusts very well to the variable "time until the failure of the sub-samples of fuses" in non replacement trials.

Considering now that the distribution which suits best to the simulations performed is the Weibull distribution, it was estimated (through simulation results) the shape parameter (β) more suitable for each combination of (p_i, pC) .

We find out that for increasing values of M the shape parameter β stabilize as showed in Fig. 5. Following the sample size used by Nordman and Meeker(2002) [9] and the result presented in Fig. 5, we decided to use the values of β related to the number of samples $M = 20000$. For each one of the combinations of (p_i, pC) , the stable values of β are presented

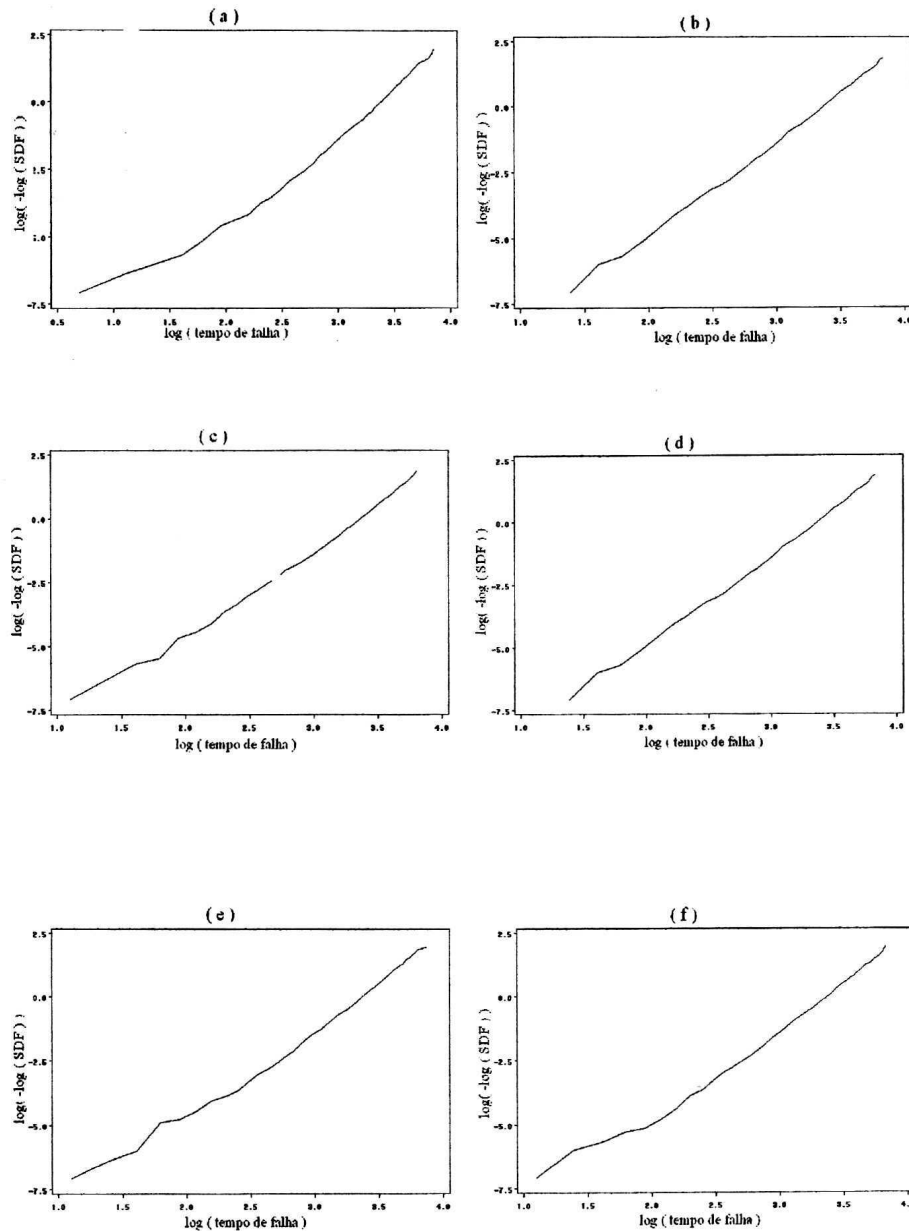


FIG. 4: Plots of the Weibull probabilities for the set of data of time until the failure of samples of fuses, for 6 simulations selected among the 6000 simulations. The **horizontal axis** represents $\log(T_f)$ and the **vertical axis** represents $\log(-\log(S(T_f)))$, where $S(T_f) = 1 - F(T_f)$ is the survival function estimated according the Kaplan-Meier (1958) method [11], for the time until the failure, T_f .

in Table 1. Such values will be used to determine the prediction intervals.

The values of β (the shape parameter of the Weibull distribution) used for each combination of (p_i, pC) are presented in Table I. Those values of β are used to determine the prediction intervals (PI) for the number of samples $M = 120$. The reason for using the values of β obtained for $M = 20000$ and **not** the β values obtained for the number of samples $M = 120$, is that we are looking for a stable value of β . This argument is

inspired on **Physical arguments of Computational Physics** called **finite size scaling** (see for instance, Newman and Bak-erma (1999) [8], Landau and Binder (2000) [2]). When we are looking at some *parameter* that is obtained through Computer simulation, we have to look at the value of the *parameter* for a sample size such that the parameter does not changes for increasing values of M . This is what is called the **thermodynamic limit**, where the value of all parameters are stable. As Fig. 5 shows, this happens for $M \geq 5000$, what justify the

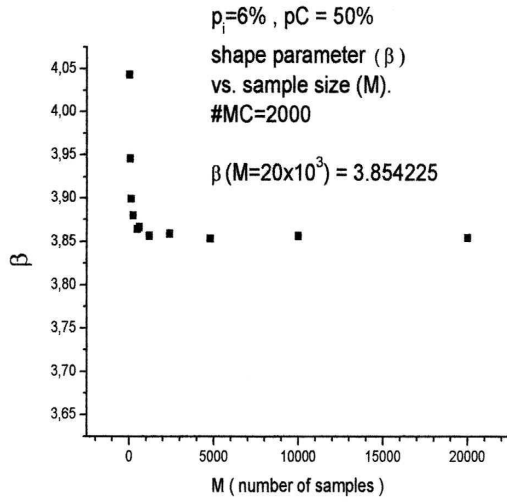


FIG. 5: Plot of the shape parameter (β), as function of the number of samples M .

choice made.

Another alternative to obtain approximate values of β is through the determination of the confidence intervals optimal conditioned, discussed by Mahdi (2003) [3].

B. Prediction Intervals

It was determined the prediction intervals for the random variable Y : Future number of failures of samples in the time interval $[T_c, T_w] = [10, 20]$ units of time, for each combination (p_i, p_C) , following the procedure detailed in the Section II B. The nominal level of confidence $1 - \alpha$ used was 0.95 and the number of samples was $M = 120$.

On Fig. 6 it is presented the prediction intervals (PI), obtained as a function of the number of sub-samples failed (X) in the time interval $[0, T_c) = [0, 10)$, for each set of values presented on Table I, namely $(p_i = 1\%, p_C = 15\%)$, $(p_i = 6\%, p_C = 15\%)$ and $(p_i = 6\%, p_C = 50\%)$.

We shall notice that in all 3 cases presented on Fig. 6, we observe that as the number of fails X in the time interval $[0, T_c) = [0, 10)$ **increases**, the **width** of the prediction interval **decreases**. Such behavior is reasonable if we bear in mind that as we have **more** information (the number of components failed, X) in the initial time interval $([0, T_c))$, the determination of the prediction interval (PI), which depends on this previous information, will be more precise, and therefore its width decreases. Following this reasoning we shall believe that for values of X very close to zero provides prediction intervals (PI) quite wide as shown on Fig. 6. Nevertheless, we shall notice that the level $(1 - \alpha)$ for all the intervals is the same. We believe that the points of discontinuity observed on Fig. 6 ((a) at $X = 80$, (b) at $X = 108$ and (c) at $X = 45$) should neither be interpreted as prediction intervals of zero width, nor as a pointed prediction of Y . As in the discontinuity points the

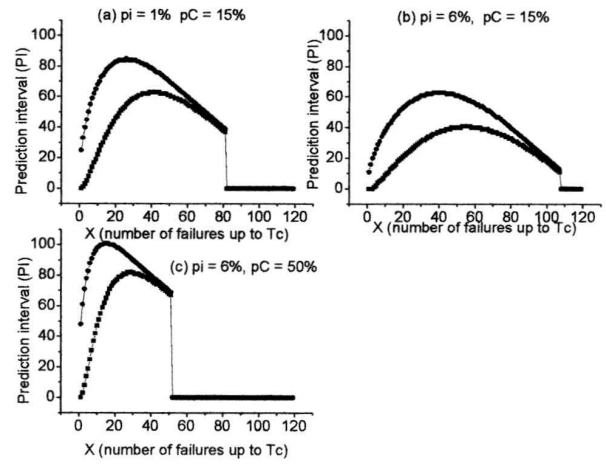


FIG. 6: Prediction interval for Y : future number samples failed in the time interval $[T_c, T_w] = [10, 20]$ units of time, for a nominal level of probability of 0.95.

prediction intervals do not offer any width, there would not be any sense to evaluate the prediction intervals for those points.

C. Alternative Coverage probability

As it was mentioned in Section II D, we will present an **alternative method** to evaluate the coverage probability (CP), hereafter called **ACP**. Here we propose a simpler method, as we judge, to confirm this. This method require 4 stages:

1. The generation, through Monte Carlo simulation, of 6000 data sets, according the procedure explained in Section II A. Each data set contains $M = 120$ data of time of failure for the $M = 120$ samples used (see Fig. 2).
2. The recording, for each simulation in step (1), of X : the number of samples failed in the time interval $[0, T_c) = [0, 10)$. As we have used $M = 120$ samples, obviously $X \leq M$.
3. Afterwards, according the **theoretical** procedure showed on Section II B, the determination of the prediction interval for Y : the future number of failures in the time interval $[T_c, T_w] = [10, 20]$ for each possible value of $X \leq M$, regardless this value has (or has not) appeared in the simulation results (step 2).
4. The recording, for each simulation in step (1), the number of failures Y^* which have occurred in the time window interval $[T_c, T_w] = [10, 20]$. The Y^* is obtained in the following way. Let be $X1$ the number of failures in the time interval $[0, T_c) = [0, 10)$, and let be $Y1$ the number of failures in the time interval $[0, T_w] = [0, 20]$. Y^* is given by $Y^* = Y1 - X1$. Therefore, Y^* is the **net** number of failures in the time interval $[T_c, T_w] = [10, 20]$.

5. The determination of the percentage (or fraction) of values recorded in step (4) (The Y^* values obtained through simulation results), which are inside the respective theoretical prediction interval (PI) obtained in step (3).

We can notice from Fig. 7, that in the 6000 simulations performed for each combination (p_i, pC) , we did not find necessarily all the possible values of X with the value of ACP near the nominal level (95%) or alternatively, 0.95. The reason for that is that Monte Carlo simulations for each pair (p_i, pC) provide an interval of values of X that is finite and dependent upon the values (p_i, pC) . This means that in the region where the ACP is null we **do not** have data set and therefore we cannot calculate **properly** the ACP. On the other hand, for each value of X found in the Monte Carlo simulation, the number of failures occurred in the time interval $[T_c, T_w] = [10, 20]$, is inside the respective theoretical prediction interval (PI) obtained in the step 3 of this procedure and illustrated on Fig. 6. This result is a good indicator of the reliability of the method.

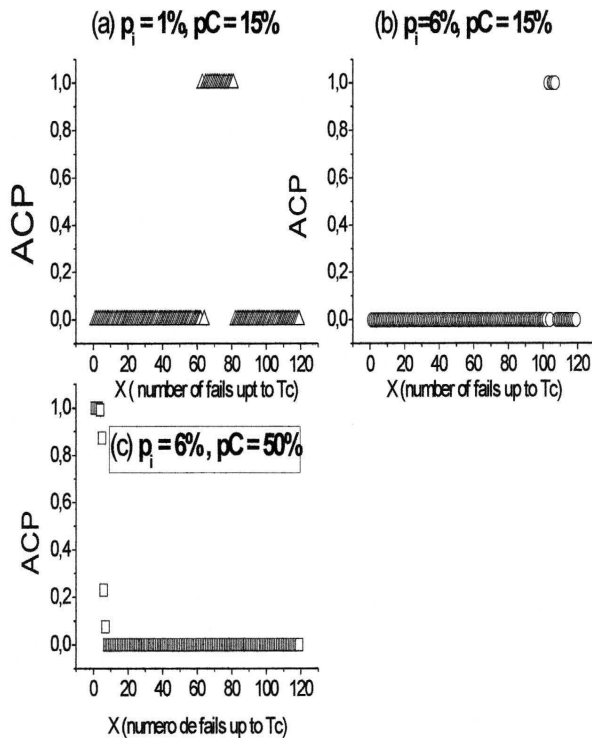


FIG. 7: Alternative Coverage probability (ACP) for the prediction interval $PI(Y)$: number of failures in the time interval $[10, 20]$ units of time for each combination of values (p_i, pC) showed on Table 1.

What Fig. 7 also shows is that for fixed pC (the maximum allowed percentage of deterioration in each sub-sample of size $n_1 = 10$), the number of failures X increases for increasing values of p_i (the initial percentage of deterioration in the sample size $n = 1000$).

IV. CONCLUSIONS

In this work we have proposed a method called SUB-SAMPLING, to obtain data of the time of failure in trials without replacement, (NRT), through Monte Carlo simulation. The data set obtained through this method has been showed to be well adjusted by the Weibull distribution for the data of time of failure. It was possible to obtain the prediction interval for a **future** number of failures with the knowledge **only** of the number of failures in the **past** time.

The results obtained with the alternative procedure to evaluate the coverage probability (ACP), indicate that the method, suggested in this work, to determine the prediction intervals in non replacement trials is reliable.

As a forthcoming work is the determination of the pair of probabilities (p_i, pC) from the histogram of frequencies of X (i.e., the number of failed components up to T_c) by the simulation data for a number of samples M . Then, through a single real experiment in a single sample, we determine the number of failed components until T_c (e.g., the value of X_{EXP}); with this date, and in association with the most probable value of the number of failed components for the same T_c (X_{SIM}^*) from the simulation data for M samples (which is dependent on the simulated pair of probabilities (p_i, pC)), we obtain through the matching of X_{EXP} with X_{SIM}^* the correspondent pair of probabilities (p_i, pC) . This is a key point, as we can determine from a single real experiment the value of X_{EXP} for a given T_c and, with the complementary simulation data through the most probable value of X_{SIM}^* from the histogram of frequency of X for M samples, the degree of deterioration p_i (for a given cut-off probability pC) for the single sample where the real experiment was made.

Acknowledgements

We wish to thank Willian Q. Meeker and Daniel J. Nordman for clarifying us on the subtle points regarding the Coverage Probability and the reliability of the method in their paper Nordman and Meeker (2002) [9].

[1] R.H. Ellis, The meaning of viability, *Hortscience*, **26**,9 (1984).
 [2] D.P. Landau and K. Binder, *A Guide to Monte Carlo Simulation in Statistical Physics* Cambridge: Cambridge University Press (2000).

[3] S. Mahdi, Optimal conditional interval for the shape parameter of a Weibull distribution, *Brazilian Journal of Probability and Statistics*, **17**, 57-74 (2003).
 [4] W.F.J. Silva Marcos Filho and S.M. Cícero, Teste de tetrazólio,

Departamento de Agricultura e Horticultura, ESALQ/USP (1987).

- [5] W. Meeker and L. Escobar, *Statistical Methods for Reliability Data*, John Wiley: New York (1998).
- [6] W. Nelson, Statistical methods for the ratio of two multinomial proportions, *The American Statistician*, **26**, 22-27 (1972).
- [7] W. Nelson, Weibull prediction of a future number of failures, *Quality Reliability Engineering International*, **16**, 23-26 (2000).
- [8] M. Newman and G. Bakema, *Monte Carlo Methods in Statistical Physics*, Oxford: Oxford University Press (1999).
- [9] D.J. Nordman and W.Q. Meeker, Weibull prediction for a future number of failures, *Technometrics*, **44**, no.1, 15-23 (2002).
- [10] J. Rostum Decision Support Tools for Sustainable Water Network Management, *A research project supported by the European Commission under the fifth framework program*, bf <http://www.unife.it/care-w> (1999).
- [11] E.L. Kaplan and P. Meier, Nonparametric Estimation from Incomplete Observations, *Journal of the American Statistical Association*, **53**, 457-481 (1958).