Statistical Techniques of Sample Size Estimating in Fatigue Tests

B. D. Soh Fotsing [#1], G. F. Anago [*2], M. Fogue [#3]

[#] Department of Mechanical Engineering, University of Dschang
P.O. BOX 134 Bandjoun, Cameroon
[1] sohfotsing@aol.fr
[3] medard_fog@yahoo.com
[*] Ecole Polytechnique d'Abomey Calavi
BP 2009 Cotonou, Benin
[2] ganago@yahoo.fr

***Abstract*—The present paper gives a survey of bibliographical recent models generating statistical predictions of minimum sample size required for fatigue characterisation. Metallurgical defects, roughness and other geometrical defaults are most often considered as the sources of deviations between fatigue lives for the same stress level. The sample size at one given stress level must therefore be carefully chosen in order to find the good compromise between the accuracy of the fatigue strength estimation and the cost of the experiments. The estimations are dependent on the desired confidence level and on the given survival probability. The application of the models to several series of tests issued from the relevant literature or realised within our laboratories is made. The proposed method is based on Gope method which establishes a relationship between the size of the sample, the probability of survival, confidence level and the parameters of Weibull distribution and log-normal. The comparison between the information taken from results of the predicted sample size and the information given by real tested specimens is then developed. It allows to study and to discuss the validity of these models with respect to the former objectives of these theories.**

I. INTRODUCTION

Fatigue tests are generally conducted to evaluate the fatigue strength of materials under some particular loading conditions. This step is mandatory for ensuring the integrity of structures submitted to in-service loading as the assessment of fatigue damage induced by multiaxial stress states cycles or by variable amplitude stress histories requires such fatigue properties. Research programs have therefore concerned during decades both fatigue damage models and experimental investigation about material fatigue strength. The classical S-N curve plots for instance the material lifetimes obtained against constant amplitude stress states.

However fatigue tests applied to several identical specimens or to mechanical structures often result in some large lifetime's discrepancy. Despite the fact that strong efforts are brought to the machine fatigue regulation in order to maintain the same testing conditions for all the specimens, large deviations between fatigue lives are currently observed.

The sample size at one given stress level must therefore be carefully chosen in order to find the good compromise between the accuracy of the fatigue strength estimation and the cost of the experiments. In addition the confidence level of the assessed lifetimes is as a matter of fact directly related to this sample size.

Estimating the parameters of distribution laws is very important for the adjustment of statistical models to experimental data [1]. In a given population, these parameters are estimated from a sample of finite size.

Reference [2] presents a method for estimating the statistical parameters of the fatigue life from a model of SN curves of the form:

$$N(p) = \delta + p^{\alpha} / \left(\sigma^{\alpha} (Y - \rho)^{\alpha/\gamma}\right), \qquad 0 \le p \le 1 \tag{1}$$

Where Y is the stress level, α, σ, γ, δ and ρ parameters related to the material and p the probability of failure.

Several methods for estimating parameters of Weibull model are proposed in order to construct the SN curves [3], [4].

Reference [5], [6] made a statistical analysis of the shape of SN curves by the method of maximum likelihood to estimate the parameters of the normal distribution of data and formulation of SN curves in the form:

$$y = \frac{1}{2}(K + K_1)x + \frac{1}{2}(K - K_1)\left\{|x| + \frac{1}{\alpha}\ln\left[1 + e^{-2\alpha|x|}\right]\right\} \tag{2}$$

with $y = \ln(N/N_E) \quad x = \ln(S/S_E) \qquad 10 \le \alpha \le 100$

Determining the sample size is so important in statistics when we know that the precision of estimates depends heavily on its size. Samples with large numbers of people can make good estimates of various parameters. In fatigue, a large number of tests at a given stress would give a very accurate estimate of expected lifetime of a material and the parameters of the law used. But, it would be very expensive to conduct such tests because they are very long and are essentially destructive.

Several authors have proposed methods for determining the optimal size of statistical samples used to obtain estimates with variations more or less important [7], [8].

The proposed methods are mostly based on the laws of distributions commonly used to analyze fatigue test data such as normal distribution, the Weibull and log-normal [9].

The first section of this work is a literature review which highlights previous and current efforts made by other researchers in this field. The second section presents the proposal for estimating sample size and the last section is an application and discussion to a fatigue tests campaign.

II. LITERATURE REVIEW

A. Estimation of distribution laws parameters

In a distribution of the expected lifetime N, suppose that for a parameterθ, we have experimentally obtained the unbiased estimate $\hat{\theta}$. To estimate the possible error, if one chooses a number α such an event of probability 1-α can be regarded as certain, we can determine the value ε for which we have

$$P\left(\left|\hat{\theta}-\theta\right|<\varepsilon\right)=1-\alpha \cdot$$

This equality means that with probability 1-α, the random interval $\left[\hat{\theta}-\varepsilon,\hat{\theta}+\varepsilon\right]$will contain the unknown value of θ. The confidence interval at confidence level 1-α, is an estimate of this interval [7].

One method for determining the value ε is to consider another random variable $\prod\left(N_1,\ldots N_n,\theta\right)$, whose distribution function G is given explicitly and does not depend on the parameter θ as

$$\theta\in\left[\hat{\theta}_{\inf},\hat{\theta}_{\sup}\right] \iff Q_T(\alpha')\leq\prod\left(N_1,\ldots N_n,\theta\right)\leq Q_T(1-\alpha+\alpha')$$
$$\iff P\left(Q_T(\alpha')\leq\prod\left(N_1,\ldots N_n,\theta\right)\leq Q_T(1-\alpha+\alpha')\right)=1-\alpha \quad (3)$$

Q_T (α) is the quintile of α by the law G.

In fatigue, the parameter θ is related to the average lifetime for a probability of 50% or the expected lifetime to a probability of γ%.

For large samples (n> 30), N follows the normal law N (μ,S). The estimator of the average μ is $\overline{N}=\frac{1}{n}\sum_{i=1}^{n}N_i$, it follows the normal law $N(\mu,\sigma/\sqrt{n})$.

1) Normal Distribution: Statistics $\sqrt{\frac{n}{\sigma^2}}\left(\overline{N}-\mu\right)$ follows the normal distribution N(0, 1).

According to (1),

$$P\left(-u_{\alpha/2}<\sqrt{\frac{n}{\sigma^2}}\left(\overline{N}-\mu\right)<u_{\alpha/2}\right)=1-\alpha \quad (4)$$

The confidence interval of μ is

$$I_\alpha=\left[\overline{N}-u_{\alpha/2}\frac{\sigma}{\sqrt{n}},\overline{N}+u_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \quad (5)$$

For the samples of small size (n< 30), the normal law is not good any more to approximate the distribution of the sampling of lifetime.

The theory of small samples that provides statistics $T(n-1)=\sqrt{\frac{n-1}{S^2}}\left(\overline{N}-\mu\right)$ follows the Student distribution with (n-1) degree of freedom. S² is the biased estimator of the variance.

$$S^2=\frac{1}{n}\sum_{i=1}^{n}\left(N_i-\mu\right)^2 \quad (6)$$

From (3),

$$P\left(-t_{1-\alpha/2}(n-1)<\sqrt{\frac{n-1}{S^2}}\left(\overline{N}-\mu\right)<t_{1-\alpha/2}(n-1)\right)=1-\alpha$$

And consequently, the tolerance interval of μ is

$$I_\alpha=\left[\overline{N}-t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{n-1}},\overline{N}+t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{n-1}}\right] \quad (7)$$

2) Weibull distribution: For the Weibull distribution, the probability of the event x is

$$P(x)=1-e^{-((x-x_0/a)^b}$$
$$x=x_0+a\left(-\ln(1-p)\right)^{1/b} \quad (8)$$

For each point x_i, the probability is approximated by

$$p(x_i)=p_i=\frac{i-0,5}{n} \quad (9)$$

The error between x and x_i is

$$dev.=x_0+a\left(-\ln(1-p)\right)^{1/b}-x_i \quad (10)$$

And hence, the estimator of the variance is the sum of squares of the errors at each point x_i is

$$S=\sum\left(dev\right)^2 \quad (11)$$

The criterion for determining a, b and x_0 consist in minimizing S.

$$\left.\frac{dS}{dx_0}=0 \quad \frac{dS}{da}=0 \quad \frac{dS}{db}=0\right\} \Rightarrow a,b \ et \ x_0$$

The estimators of the parameters for the law of Weibull with two parameters can be obtained by the method of maximum likelihood.

$$E(X)=a\Gamma(1+1/b) \quad (12)$$

$$Var(X)=a^2\left[\Gamma\left(1+2/b\right)-\Gamma^2\left(1+1/b\right)\right] \quad (13)$$

$$a^b=\frac{1}{n}\sum_{i=1}^{n}X_i^{\,b} \quad (14)$$

B. Estimation of a sample size

Several methods have been proposed to determine the unknown size of samples in statistics.

1) Standard method: From (5) and (7), we define the maximum absolute difference between the average and its estimated value as

$$E_{\max} = E(\mu) - \mu = u_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad (15)$$

$$E_{\max} = E(\mu) - \mu = t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{n-1}} \quad (16)$$

The relative error is defined by:

$$R_N = \frac{E(\mu)-\mu}{\mu} = u_{\alpha/2}\frac{\sigma}{\mu\sqrt{n}} \quad (17)$$

$$R_T = \frac{E(\mu)-\mu}{\mu} = t_{1-\alpha/2}(n-1)\frac{S}{\mu\sqrt{n-1}} \quad (18)$$

From these equations, n can be determined directly or by iteration on R_T until R_T equals R_0 an acceptable error level for a chosen stress level.

2) Gope Method: Reference [7] presents a method for determining the sample size and estimation of fatigue life for a probability of survival and a given level of confidence with an acceptable error.

He establishes a relationship between the size of the sample, the probability of survival, confidence level and the parameters of Weibull distribution and log-normal. This exercise has two main difficulties that to envisage fatigue life and that to envisage a limited number of samples.

a) Factors of error in determining the sample size by the Weibull distribution

The distribution function of the number of cycles to failure is:

$$F_{Nf}(N_f) = 1 - \exp\left\{\frac{N_f}{\theta}\right\}^{\beta} \quad (19)$$

A change of variable X = ln(N_f) can write.

$$F_x\left[\ln(N_f)\right] = 1 - \exp\left\{\frac{x-\xi}{\delta}\right\} \text{ with } \delta=1/\beta,\ \xi=\ln\theta \quad (20)$$

The expected lifetime μ_x to α% probability and γ% confidence level is

$$\mu_x = \overline{x} + (d+\lambda)\delta$$

where $\lambda = \ln\left[-\ln(1-\alpha)\right]$ and $\overline{x} = \overline{\xi} + \lambda\overline{\delta}$ (21)

The absolute difference between the probable lifetime and its hope is expressed as:

$$E(\mu_x) - \mu_x = \frac{t.\overline{\delta}}{\sqrt{n-1}}.\sqrt{\frac{\chi_n^2.c.n.V_0}{\chi^2[c(n-1)]}} \quad (22)$$

where c is a constant and V_0 statistics on all data.

The error factor is defined by

$$K_{w,\alpha,\gamma} = R_w/\phi_w = \frac{t.}{\sqrt{n-1}}.\sqrt{\frac{\chi_n^2.c.n.V_0}{\chi^2[c(n-1)]}} \quad (23)$$

with $R_w = \frac{E(\mu_x)-\mu_x}{\mu_x}.100$ and $\phi_w = \frac{\delta}{\left[\overline{x}+(d+\lambda).\overline{\delta}\right]}$

b) Factors of error in determining the sample size by the log-normal distribution

The lifetime N_f follows the normal distribution and y = logN_f follows the log-normal distribution.

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{\frac{-(y-\mu)^2}{2\sigma^2}\right\}$$

The expected lifetime is

$$\mu_y = \overline{\log(N_f)} + u.\psi.s$$

with $\psi = \sqrt{\frac{n-1}{2}}\frac{\Gamma[(n-1)/2]}{\Gamma(n/2)}$ (24)

Following the t distribution,

$$E(\mu_y) - \mu_y = t\sqrt{\frac{\chi_n^2 Var(\mu_y)}{n-1}} \quad (25)$$

The variance of hope is

$$Var(\mu_y) = Var(\overline{\log(N_f)}) + u^2.\psi^2.Var(s) = \frac{\sigma^2}{n} + u^2.\psi^2.Var(s)$$

The error factor is written as

$$K_{N,\alpha,\gamma} = \frac{R}{\phi_N} = t\sqrt{\frac{1}{n} + u^2.(\psi^2-1)}$$

with $\phi_N = \frac{s}{\overline{\log(N_f)} + u.\psi.s}$ (26)

c) Procedures for determining the sample size

- Choice of stress level S and the acceptable level of error R_0
- Estimation of parameters $\overline{\beta},\overline{\theta}$ (tests on three specimens) and choice of α, γ and N
- Calculation of parameters $\overline{\xi},\overline{\delta},\lambda,\overline{x}$,d (Weibull) or ψ (log-normal)
- Calculating ϕ_w (Weibull) or ϕ_N (log-normal)
- Determining the error factor $K_{W,\alpha,\gamma}$ (Weibull) or $K_{N,\alpha,\gamma}$ (log-normal) (tabulated values)
- Calculating R_W (Weibull) or R_N (log-normal)
- If $R_W > R_0$ or $R_N > R_0$, incrementing the value of N and repeat step 4
- N is the sample size desired.

3) Other methods: Reference [8] provides a method for estimating the size of sample Y1, ... Yn from information provided by a statistic T(Y) whose family Pn corresponding to the random variable T and unimodal recursive on {xk}k>1. They determine unbiased estimators g(T) of any function h(n) and the maximum likelihood estimate of n.

Any function h(n) can be estimated by unbiased estimators g(T) such that:

$$E(g(T)) = \sum_{i=1}^{n_0} g(x_i) P_1(x_i) = h(1)$$

$$g(x_{n_0+n}) = \left[\frac{h(n+1) - q_{n+1} h(n)}{1 - q_{n+1}} - \sum_{i=1}^{n_0+n-1} g(x_i) P_n(x_{i-1}) \right] \frac{1}{P_n(x_{n_0+n-1})} \tag{27}$$

If x_{k0} are the values of T, then $n^* = Min\left(n : x_{k_0} \in M(P_n)\right)$, where n* is the maximum likelihood of the sample size n and $M(P_n) = \left(x_{k_0}, x_{k_0+1}, \ldots x_{k_0+k(n)}\right) \quad k(n) \geq 1$.

Assuming known the number of records satisfying a condition to the rank r (eg fractures), 1<r <n, such that

$$N_r^{(1)} = \sum_r^n I_i^{(1)}, \quad N_r^{(2)} = \sum_r^n I_i^{(2)} \tag{28}$$

with $I_i^{(1)} \approx Be\left(\frac{1}{i}\right)$ *and* $I_i^{(2)} \approx Be\left(\frac{r}{i}\right)$

The unbiased estimators of the sample size are:

$$T_{r,\lambda}^{(1)} = r2^{N_r^{(1)}} - 1 + \lambda(1-r)^{N_r^{(1)}}, \quad T_r^{(2)} = r\left(\frac{r+1}{r}\right)^{N_r^{(2)}} - 1, \quad \lambda \in IR \tag{29}$$

Reference [10] proposes a method for estimating of the minimum number G_{min} of test specimens for uniaxial tension–compression fatigue tests on specimens fabricated from acrylic bone cement. For a tolerable error of 5%, we estimated G_{min} to be either 7 (if the fatigue life results are treated using the two-parameter Weibull distribution function) or 11 (if the fatigue life results are treated using the three-parameter Weibull distribution function).

III. Proposal for Estimating Fatigue Test Sample Size

The proposed method is based on the same forms of equations above, but used as estimators of the mean and standard deviation, statistics obtained by the method of maximum likelihood for the Weibull distribution.

$$\begin{gathered} E(X) = a\Gamma(1 + 1/b) \\ Var(X) = a^2\left[\Gamma(1 + 2/b) - \Gamma^2(1 + 1/b)\right] \\ a^b = \frac{1}{n}\sum_{i=1}^{n} X_i^b \end{gathered} \tag{30}$$

The proposed method is to calculate the sample size based on new estimates using the method Gope for the Weibull distribution defined in (23).

The algorithm to determine the sample size N is given in Fig. 1.

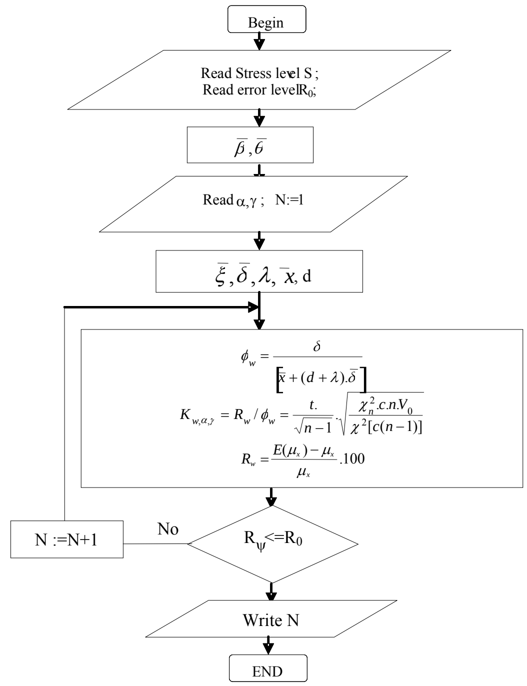

Fig.1. Computational flow chart for the sample size N

IV. Application to a Fatigue Test Campaign

Reference [7] uses data from rotating bending tests of carbon steel containing 0.2 to 25 trials with a stress level of 240 MPa which hope of the population μ and the standard deviation σ are estimated by 4.4896 and 0.48769 respectively. He made a comparison between the sample sizes estimated using four methods to a failure probability of 50% and a confidence level of 90%. Results are shown in Table 1.

TABLE I.
COMPARISON OF RESULTS OBTAINED WITH DIFFERENT MODELS

Method	**Formula used**	**Sample size**	**α and γ**	**μ**	**% diff** $\frac{\mu - \bar{x}}{\bar{x}} \times 100$
t-D^A	$\mu < \bar{x} + \frac{S.t}{\sqrt{n}}$	25	α=50% γ=90%	4,4896	4,61
z-D^A	$\mu < \bar{x} + \frac{\sigma.z_\gamma}{\sqrt{n}}$	25	α=50% γ=90%	4,4634	5,16
Logistic	$\mu < \bar{x} + \frac{m_{0.5}.S}{\sqrt{n}}$	25	α=50% γ=90%	4,5311	3,73
Gope (Log normal)	$\frac{\mu - \bar{x}}{\bar{x}} = R_N$	9	α=50% γ=50%	4,7380	
Gope (Log normal)	$\frac{\mu - \bar{x}}{\bar{x}} = R_N$	7	α=50% γ=90%	4,7066	

D^A: Distribution α: Probability γ: level of confidence

In the method of Gope, the estimated parameters θ and β of the Weibull distribution are estimated experimentally on a sample of three tests and we obtain β = 1.137.

Using this value of β to calculate the parameters of the Weibull distribution obtained by the likelihood method yields the following estimates for θ, μ and σ:

θ=44148.57 μ=42157.9 and s=37159.69. The unbiased estimator of the standard deviation is s=44414.95.

The table below shows the sample sizes obtained with the method and the method proposed by Gope.

TABLE II

COMPARISON OF GOPE RESULTS AND THE PROPOSED METHOD

Method	Relation used	Mean	% diff.	Size
Gope (log-normal)	$K_{N,\alpha,\gamma} = \frac{R}{\phi_N} = t\sqrt{\frac{1}{n} + u^2.(\psi^2 - 1)}$	4,7066	5,4	7
Gope (Weibull)	$K_{w,\alpha,\gamma} = R_w / \phi_w = \frac{t.}{\sqrt{n-1}}.\sqrt{\frac{\chi_n^2.c.n.V_0}{\chi^2[c(n-1)]}}$	3,6224	4,76	10
Proposed Method	$K_{w,\alpha,\gamma} = R_w / \phi_w = \frac{t.}{\sqrt{n-1}}.\sqrt{\frac{\chi_n^2.c.n.V_0}{\chi^2[c(n-1)]}}$	4,2158	5,08	9

V. CONCLUSION

Many bibliographic models for predicting the sample size in a fatigue test were investigated. The estimates are based on major statistical models commonly used (Weibull law, Student law, normal law...).

The proposed method is based on the same laws but used as estimators of the mean and standard deviation statistics obtained by the method of maximum likelihood for the Weibull distribution.

Different models were applied to a series of rotating bending tests of carbon steel containing 0.2 to 25 trials with a stress level of 240 MPa which expectancy of the population m and the standard deviation s are estimated respectively by 4.4896 and 0.48769.

The results are quite consistent with those of the literature. However one or the other model depends on the confidence level chosen and the estimator of the mean and standard deviation.

REFERENCES

[1] Xiulin Zheng. *On some basic problems of fatigue research in engineering.* International Journal of Fatigue, 23, 751–766, 2001.

[2] Enrique Castillo, A. Fernández-Canteli and Ali S. Hadi. *On fitting a fatigue model to data,* International Journal of Fatigue, Volume 21, Issue 1, pp.97-105, 1999.

[3] Lee J., B. Harris, D. P. Almond and F. Hammett. *Fibre composite fatigue-life determination,* Composites Part A: Applied Science and Manufacturing, Volume 28, Issue 1, pp.5-15, 1997.

[4] T. P. Philippidis and A. P. Vassilopoulos. *Fatigue of composite laminates under off-axis loading,* International Journal of Fatigue, Volume 21, Issue 3, pp.253-262, 1999.

[5] J. E. Spindel and E. Haibach. *The method of maximum likelihood applied to the statistical analysis of fatigue data,* International Journal of Fatigue, Volume 1, Issue 2, pp. 81-88, 1979.

[6] J. E. Spindel and E. Haibach. *Some considerations in the statistical determination of the shape of S-N curves*, Statistical analysis of fatigue data,ASTM. STP744, R.E Little &J.C. EKWEL, Eds. American Society of Testing and Materials, pp. 89-113, 1981.

[7] P. C. Gope. *Determination of sample size for estimation of fatigue life by using Weibull or log-normal distribution,* International Journal of Fatigue, Volume 21, Issue 8, pp. 745-752,1999.

[8] J. L. Moreno-Rebollo, F. López-Blázquez, I. Barranco-Chamorro and A. Pascual-Acosta. *Estimating the unknown sample size*, Journal of Statistical Planning and Inference, Volume 83, Issue 2, pp. 311-318, 2000.

[9] J. Schijve. *A normal distribution or a Weibull distribution for fatigue lives.* Fatigue, Fract. Eng. Mater. Struct. Vol. 16 N° 8. pp. 851-859, 1993.

[10] Gladius Lewis, *Anupama Sadhasivini. Estimation of the minimum number of test specimens for fatigue testing of acrylic bone cement*. Biomaterials, Volume 25, Issue 18, August 2004, Pages 4425-4432.