JOURNAL OF MATERIALS SCIENCE 19 (1984) 4021–4025

Estimation of Weibull parameters by omission of some data in a sample

N. KAMIYA, O. KAMIGAITO
Toyota Central Research and Development Laboratories, 41-1 Aza Yokomichi, Oaza Nagakute, Nagakute-cho, Aichi-gun, Aichi-ken, 480-11 Japan

Methods of estimating Weibull parameters by omission of some data in a sample are considered for various functions for survival probability in linear Weibull plots. Computer simulation is used for the consideration. A use of the survival probability function, $P(i) = 1 - (i - 0.5)/N$, in which $P(i)$ is survival probability, i is the i-th order of failure strength and N is the total number of data in the sample, and the omission of 2 to 4% of the data, which are of the smallest value, the largest one and their neighbours, are recommended to get the best value of the Weibull parameters.

1. Introduction

Weibull statistics are now widely used to describe the distribution of the fracture stress of various groups of ceramics. In the application of the statistics, two independent parameters must be determined to express the distribution of the values of fracture stress as the two-parameter Weibull distribution. The determination is frequently made by linear regression [1–3] due to the simplicity of the method rather than the methods of direct curve fitting [1–4] and maximum likelihood [1–4]. In the linear regression method, each datum is arranged in the order of its value, and a statistical probability such as survival one or failure one is given for an arbitrarily given value of the fracture stress. The value of the probability depends only on the order of each datum, or rank, and the number of the data, or sample size. The actual value of the fracture stress of a given rank, however, is not fixed, but scatters depending on the sampling or the group. The scattering is greater for the smallest and largest values of the data as the density of data is low in the extremes and the difference between the values for neighbouring ranks is greater than in the region near the median. And the proability that the data of the same rank have always the same value or nearly the same one is much lower than that in the region near the mode or median where the density of the data is very high and the difference between the values of the data of neighbouring rank is very small. Therefore, the data which are near the mode or median will give more reliable values for the parameters than the data deviating much from the mode or the median. Thus, to get correct values for the parameters, giving some weight to each datum is thought to be suitable, in which the weight for the data far from the mode or median is light and that for the data near the mode or median is heavy.

In the field of statistics, many works have been published which describe the use of various weighting functions [5–8]. Among these functions, the simplest weighting function is a step function in which the weight is unity or zero. In this work, the step function was studied. To determine the best function or the best weighting, the real values of the two parameters of the sample must be compared with the values estimated from the weighted data. To do this, a computer simulation method, which was studied recently by Trustrum and Jayatilaka [1], was applied.

2. Simulation procedure

The simulation procedure is shown in Fig. 1. The population is assumed to have the Weibull distribution described as follows [1],

$$P = \exp\left\{-\left[\Gamma\left(\frac{m_0+1}{m_0}\right)\frac{\sigma_f}{\mu_0}\right]^{m_0}\right\} \qquad (1)$$

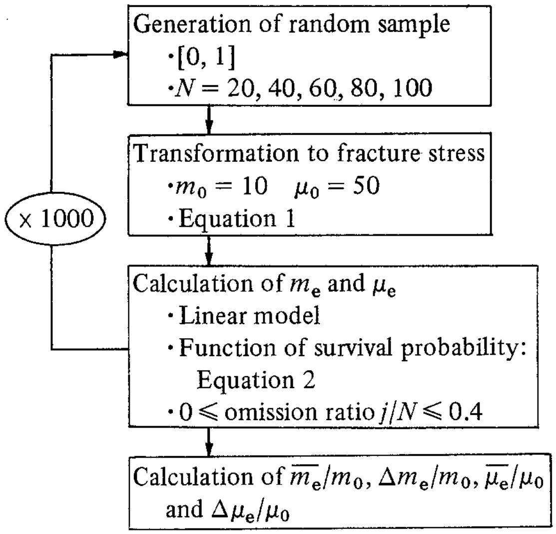

Figure 1 Flow diagram of computer simulation for evaluating the effectiveness of omission of data in a sample and function of survival probability on estimating Weibull parameters.

where P, m_0, σ_f and μ_0 represent, respectively, the cumulative survival probability, Weibull modulus, fracture stress and the mean fracture stress of the sample. The value of P varies from 1 to zero. N random numbers belonging to [0, 1] are generated by a computer, ACOS 850. The value of the generated random number is substituted for the value of P, and it is transformed into σ_f by the use use of the Weibull function given in Equation 1. Through this procedure, N data for fracture stress, σ_f, are generated. These values are arranged in the order of the value. The statistical survival probability of the i-th datum, $P_l(i)$ ($l = 1, 2, 3$), is assumed to be given by one of the following three functions:

$$P_1(i) = 1 - \frac{i}{N+1} \qquad [1\text{–}3, 9]$$

$$P_2(i) = 1 - \frac{i-0.5}{N} \qquad [1, 9] \qquad (2)$$

$$P_3(i) = 1 - \frac{i-0.3}{N+0.4} \qquad [9]$$

$\ln[-\ln P_l(i)]$ is plotted against $\ln \sigma_f$ for each function of $P_l(i)$. Applying the least mean square method to make the function fit to the data, the value of Weibull modulus, m_e, and the mean fracture stress, μ_e, were estimated. Then, $2j$ ($j = 0, 1, 2, \ldots$) data including the lowest, highest and their neighbours are omitted from the sample. Referring to the sample consisting of $(N-2j)$ data, each datum is plotted so as to have the same coordinate as that of the datum before omission, and m_e and μ_e are estimated in the same way as mentioned above. The omission of some data is equivalent to the application of a weight function of step type. The number of omitted data on the lower side is selected to be equal to that on the higher side (j). The simulation was repeated 1000 times for each sample. The mean values of m_e and μ_e, $\overline{m_e}$ and $\overline{\mu_e}$, and their standard deviation, Δm_e and $\Delta\mu_e$, were calculated using the values given from 1000 simulations. The values of N, m_0 and μ_0 were selected respectively as $20 \leqslant N \leqslant 100$, $m_0 = 10$ and $\mu_0 = 50$. There is no loss of generality in choosing $m_0 = 10$ as the distribution of $\overline{m_e}/m_0$ is independent of μ_0 and m_0 [1].

3. Results and discussion

The mean value of the estimated Weibull modulus, $\overline{m_e}$, is plotted as the ratio of $\overline{m_e}$ to the true value, m_0, against the ratio, j/N, in Fig. 2 for the function, $P_2(i) = 1 - (i-0.5)/N$. The standard deviation of m_e, Δm_e, is plotted in the form of $\Delta m_e/m_0$ in Fig. 3 for the function, $P_2(i)$. The estimated value, $\overline{m_e}$, approaches m_0, with the increase of the sample size, N, as shown in Fig. 2. For $N \geqslant 40$, the estimated value can be thought to be practically equal to the true value of m, which agrees well with the conclusion by Trustrum and Jayatilaka [1]. $\overline{m_e}$ varies with the omitting ratio, j/N, and it is a minimum at the value of j/N of about 0.02 to 0.04. The standard deviation, Δm_e, is less for larger sample size and it is a minimum at the value of j/N of about 0.02 to 0.04, as shown in Fig. 3.

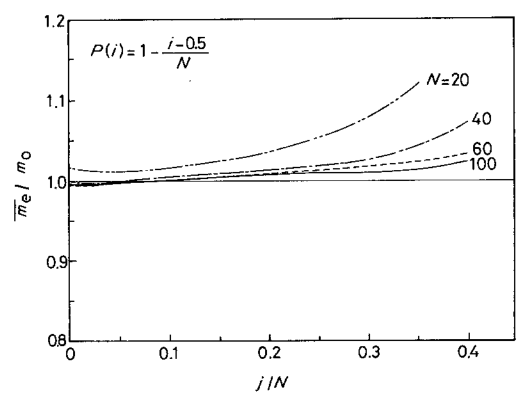

Figure 2 Estimated mean values of Weibull modulus, m_e/m_0, as a function of omission ratio of data, j/N, for various sample sizes. ($P_2(i) = 1 - (i-0.5)/N$ is used as a function for survival probability.)

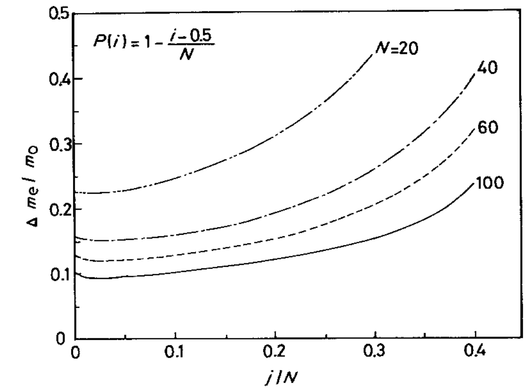

Figure 3 Estimated standard deviation of Weibull modulus, $\Delta m_e/m_0$, as a function of omission ratio of data, j/N, for various sample sizes. ($P_2(i) = 1 - (i-0.5)/N$ is used as a function for survival probability.)

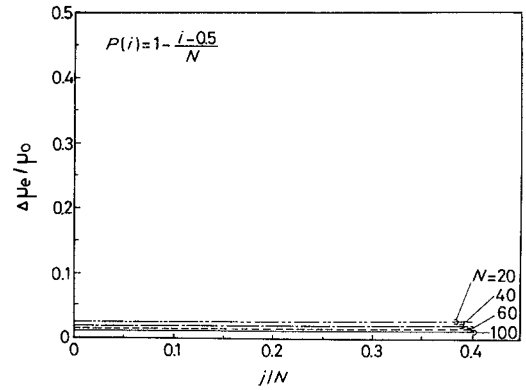

Figure 5 Estimated standard deviation of mean fracture stress, $\Delta\mu_e/\mu_0$, as a function of omission ratio of data, j/N for various sample sizes. ($P_2(i) = 1 - (i-0.5)/N$ is used as a function for survival probability.)

Therefore, the best value can be given when this proportion of data in the regions of the smallest value and of the largest value is omitted from the sample.

The mean value of the estimated mean fracture stress, $\overline{\mu_e}$ and its standard deviation, $\Delta\mu_e$, are shown in Figs. 4 and 5 as a function of j/N for the function $P_2(i) = 1 - (i-0.5)/N$. Also, the mean value of the estimated mean fracture is the same as the value of μ_0 independently of the value of j/N and the sample size. The standard deviation of the mean fracture stress, $\Delta\mu_e$, is less for larger sample size, and independent of the value of j/N.

The ratio of $\overline{m_e}$ to m_0 for the other functions, $P_l(i)$ is given in Fig. 6 together with that for $P_2(i) = 1 - (i-0.5)/N$ for $N = 40$. The standard deviation of m_e for the various functions, $P_l(i)$, is given in Fig. 7. As shown in Fig. 6, the estimated value of $\overline{m_e}$ increases monotonically with j/N, and $\overline{m_e}$ is equal to the true value at the value of j/N of 0.17 for $P_3(i) = 1 - (i-0.3)/(N+0.4)$ and at that of 0.3 for $P_1(i) = 1 - i/(N+1)$. Therefore, omission of a large number of data is thought to be recommended for the probability functions $P_3(i)$, $P_1(i)$. The tangent of the curve for the three functions at the value of $\overline{m_e}/m_0 = 1$ is the smallest for the function $P_2(i) = 1 - (i-0.5)/N$. So the function, $P_2(i) = 1 - (i-0.5)/N$, will be the most desirable for the estimation of m_e of the three functions, as the error due to the variation in the number of omitted values is the smallest. The standard deviation, Δm_e, is almost the same for the three functions, and nearly constant for j/N less than 0.1.

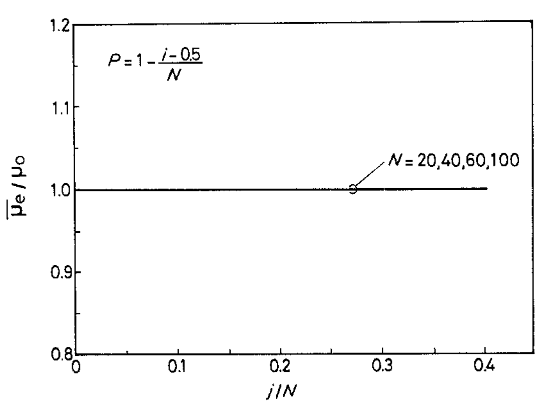

Figure 4 Estimated mean values of mean fracture stress, $\overline{\mu_e}/\mu_0$, as a function of omission ratio of data, j/N, for various sample sizes. ($P_2(i) = 1 - (i-0.5)/N$ is used as a function for survival probability.)

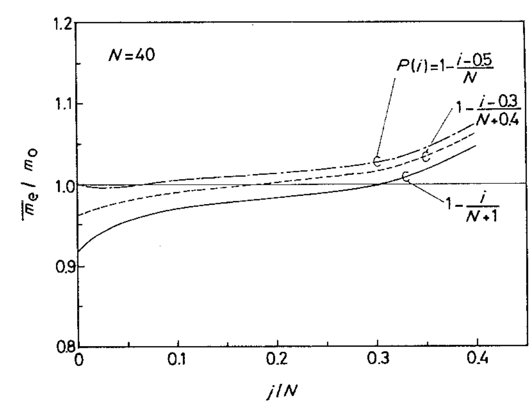

Figure 6 Estimated mean values of Weibull modulus, $\overline{m_e}/m_0$, as a function of omission ratio of data, j/N for various functions for survival probability. Sample size is forty.

TABLE I Estimated mean values of mean fracture stress, $\overline{\mu_e}/\mu_0$, in a relation to omission ratio of data for various functions for survival probability. Sample size is forty

j/N	Function		
	$1-\frac{i-0.5}{N}$	$1-\frac{i-0.3}{N+0.4}$	$1-\frac{i}{N+1}$
0	0.9996	0.9990	0.9982
0.025	0.9994	0.9990	0.9985
0.05	0.9994	0.9991	0.9987
0.075	0.9994	0.9992	0.9988
0.10	0.9994	0.9992	0.9989
0.125	0.9994	0.9993	0.9990
0.15	0.9994	0.9992	0.9990
0.175	0.9994	0.9992	0.9991
0.2	0.9993	0.9992	0.9991
0.225	0.9994	0.9993	0.9991
0.25	0.9994	0.9993	0.9992
0.275	0.9994	0.9993	0.9992
0.30	0.9995	0.9994	0.9993
0.325	0.9995	0.9995	0.9994
0.35	0.9996	0.9995	0.9995
0.375	0.9997	0.9996	0.9996
0.4	0.9998	0.9998	0.9998

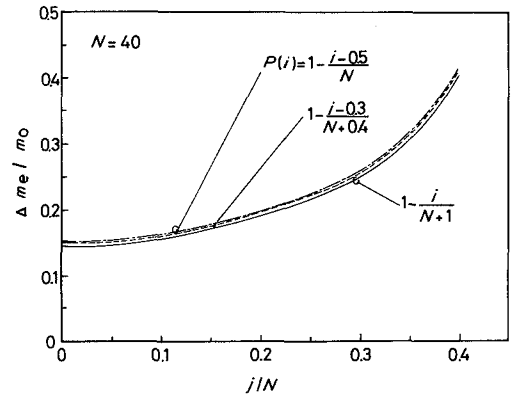

Figure 7 Estimated standard deviation of Weibull modulus, $\Delta m_e/m_0$, as a function of omission ratio of data for various functions for survival probability. Sample size is forty.

The mean value of the estimated mean fracture stress and its standard deviation are shown for $N = 40$ in Table I and Table II for three functions of $P_l(i)$. As shown in these tables, the mean value and the standard deviation of the estimated mean fracture stress are almost the same for the three functions and nearly constant, independent of the value of j/N.

As discussed above, the application of the function, $P_2(i) = 1-(i-0.5)/N$, to represent the statistical cumulative survival probability and the omission of some data comprising 2 to 4% of the smallest and largest fracture stresses is recommended for the estimation of the Weibull parameters. Moreover, as shown by Trustrum and Jayatilaka [1], a sample size, N, larger than 40 is recommended. As the result of the application of the recommended function and the omission of some data in the sample whose size is larger than 40, the estimated value of m_e can be made nearly equal to the true value and the standard deviation is made less than 0.15.

TABLE II Estimated standard deviation of mean fracture stress, $\Delta\mu_e/\mu_0$, in a relation with omission ratio of data for various functions for survival probability. Sample size is forty

j/N	Function		
	$1-\frac{i-0.5}{N}$	$1-\frac{i-0.3}{N+0.4}$	$1-\frac{i}{N+1}$
0	0.19218 ($\times 10^{-1}$)	0.19205 ($\times 10^{-1}$)	0.19185 ($\times 10^{-1}$)
0.025	0.19207	0.19203	0.19194
0.05	0.19331	0.19326	0.19318
0.075	0.19497	0.19492	0.19483
0.1	0.19683	0.19677	0.19668
0.125	0.19893	0.19887	0.19878
0.15	0.20093	0.20086	0.20076
0.175	0.20269	0.20262	0.20251
0.2	0.20464	0.20456	0.20445
0.225	0.20663	0.20656	0.20644
0.25	0.20864	0.20856	0.20844
0.275	0.21055	0.21046	0.21033
0.30	0.21230	0.21220	0.21205
0.325	0.21428	0.21418	0.21402
0.35	0.21654	0.21642	0.21625
0.375	0.21886	0.21874	0.21856
0.4	0.22074	0.22062	0.22043

4. Conclusions

For the estimation of Weibull parameters, three kinds of functions to calculate the statistical cumulative survival probability and the omission of some data of smaller and of larger value were studied. The function, $P(i) = 1 - (i - 0.5)/N$, for the statistical probability and the omission of 2 to 4% of the data which are the smallest, the largest and their neighbours, are recommended to get the best values of the Weibull parameters. A sample size larger than 40 is also recommended. Application of the function and the omission to any sample with size larger than 40 will give estimated parameters which should be practically equal to the true value of the parameters.

References

1. K. TRUSTRUM and A. DE S. JAYATILAKA, *J. Mater. Sci.* **14** (1979) 1080.
2. J. W. HEAVENS and P. N. MURGATROYD, *J. Amer. Ceram. Soc.* **53** (1970) 503.
3. W. D. SCOTT and A. GADDIPATI, "Fracture Mechanics of Ceramics 3" (Plenum Press, New York, 1978) p. 125.
4. D. G. S. DAVIES, *Proc. Brit. Ceram. Soc.* **22** (1973) 429.
5. J. W. TUKEY, Proceedings of the Conference on Critical Evaluation of Chemical and Physical Structural Information (National Academy of Science, Washington, 1974) Vol. 3, p. 48.
6. D. F. ANDREW, *ibid.* p. 36.
7. J. F. CLAERBOUT, *Geophysics* **38** (1973) 826.
8. P. J. HUBER, *Ann. Math. Stat.* **35** (1964) 73.
9. E. J. GUMBEL, "Statistics of Extremes" (Columbia Univ. Press, New York, 1958) Chap. 2.

Received 22 November 1983
and accepted 8 February 1984