J Mater Sci (2006) 41:5011–5013
DOI 10.1007/s10853-006-0123-7

LETTER

An unbiased probability estimator to determine Weibull modulus by the linear regression method

Murat Tiryakioğlu

Received: 17 October 2005 / Accepted: 4 January 2006 / Published online: 27 May 2006

Weibull statistics is widely used to model the variability in the fracture properties of ceramics as well as metals. The probability, P, that a metallic part will fracture at a given stress or strain, x, or below can be predicted as [1]

$$P = 1 - \exp\left[-\left(\frac{x}{x_0}\right)^m\right] \tag{1}$$

where x_0 is the scale parameter and m is the Weibull modulus, alternatively referred to as the shape parameter.

There are several methods available in the literature to calculate the Weibull modulus: linear regression (least square), weighted least square, maximum likelihood method and method of moments. The most popular method is linear regression mainly because of its simplicity; taking the logarithm of Eq. (1) twice yields a linear equation

$$\ln\left[\ln\left(\frac{1}{1-P}\right)\right] = m\ln(x) - m\ln(x_0) \tag{2}$$

with a slope of m and an intercept of $-m\ln(x_0)$. To estimate m by using Eq. (2), probabilities have to be assigned to all experimental data. Since true probabilities are unknown, P has to be estimated. Several studies [2–4] have been conducted to determine which probability estimator performs better. All probability estimators were found to give biased results, i.e., the average of the estimated m values is not the same as true m (m_{true}). For sample sizes (n) above 20, the probability estimator with the least bias is:

$$P = \frac{i - 0.5}{n} \tag{3}$$

where i is the rank of each data point.

To characterize the bias of Weibull moduli estimated by using Eq. (3), Monte-Carlo simulations were used to generate n data from a Weibull distribution with parameters $x_0 = 1$ and $m_{\text{true}} = 10$. For one observation, n random numbers between 0 and 1 were generated to obtain a set of x values, all assigned a rank, which was then used to calculate probabilities using Eq. (3). Equation (2) was used to estimate the Weibull modulus. The sample size was changed systematically between 5 and 50. For each sample size, the experiment was repeated 20,000 times. Estimated Weibull moduli were normalized by dividing them by m_{true} and their average (M) was calculated for each n and probability estimator. Results are presented in Fig. 1 which shows that Eq. (3) consistently overestimates the Weibull modulus. The magnitude of bias, i.e., the difference between M values and the $M = 1$ dashed-line, decreases on n and is almost 0 when $n = 50$. Note that results of Langlois [2] agree very well with those obtained in this study.

In several studies [2, 4, 5], coefficient of variation was used as a measure of the precision of estimated Weibull moduli. The smaller the coefficient of variation, the higher the precision of estimates. However, this approach may be misleading; increases in both the average (bias) and standard deviation (σ_{m}) may be hidden by a decreased coefficient of variation. Several other researchers [3, 6] recommended the use of correction factors to eliminate bias. However, Peterlik [7] showed that each data set gives statistically correct Weibull modulus estimates. The bias arises from adding the results of repeated simulations. Therefore, when there is only one set of data, one should

M. Tiryakioğlu (✉)
Department of Engineering, Robert Morris University, Moon Township, PA 15108, USA
e-mail: tiryakioglu@rmu.edu

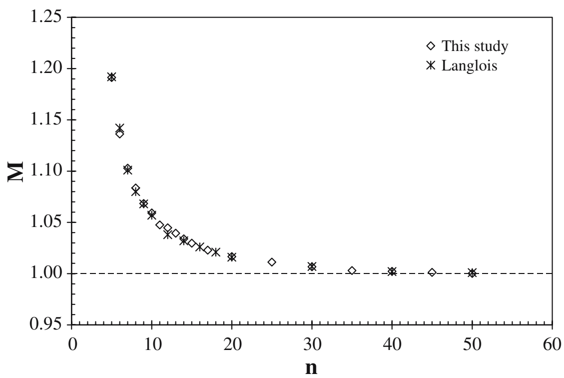

Fig. 1 The effect of n on M as found in this study and by Langlois [2]

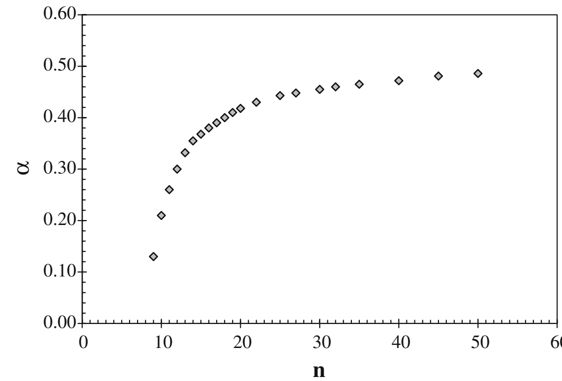

Fig. 2 The change in α with sample size

refrain from using correction factors. Recently, Song et al. [8] demonstrated that better estimates of the Weibull modulus can be made when the probability estimator is written in the form:

$$P = \frac{i - \alpha}{n + \beta} \tag{4}$$

where α and β are empirical values that change with n. Song et al. used the fraction of the distribution of m/m_{true} that lies between 0.9 and 1.1 (f) as the criterion to judge, instead of the coefficient of variation. They ran 10,000 simulations for each sample size and provided α and β values for only 5 sample sizes ranging from 10 to 50. Their results show an increase in f with n, but the averages of resultant distributions were not reported.

In this study, the primary aim is to provide a probability estimator that is unbiased for all sample sizes investigated. To accomplish this task, only α was changed while β was kept constant at 0, following Eq. (3). For each n, an iterative procedure was employed to calculate the value of α that yielded unbiased results as follows. Using the M and σ_m for each n, confidence intervals for true mean of the m/m_{true} distribution (μ_M) were calculated as

$$M - z\frac{\sigma_m}{\sqrt{n_m}} \leq \mu_M \leq M + z\frac{\sigma_m}{\sqrt{n_m}} \tag{5}$$

where z is 1.95996 for 95% confidence. The value of α was varied until $\mu_M = 1$ was within the confidence intervals. Experiments were repeated for 50,000 times (=n_m) for each n.

Results are presented in Table 1, in which α, average, standard deviation and 95% confidence intervals and the fraction of the distribution for $0.9 \leq m/m_{\text{true}} \leq 1.1$ (f) are

Table 1 Values of α for sample sizes between 9 and 50 along with statistics calculated

n	α	M	σ_m	95% confidence limits		f
				Upper	Lower	
9	0.130	1.0025	0.3574	1.0056	0.9994	0.2365
10	0.210	0.9994	0.3358	1.0023	0.9965	0.2490
11	0.260	1.0005	0.3153	1.0032	0.9977	0.2631
12	0.300	1.0003	0.3007	1.0029	0.9976	0.2709
13	0.332	1.0012	0.2889	1.0037	0.9986	0.2860
14	0.355	1.0003	0.2764	1.0027	0.9979	0.2956
15	0.368	1.0010	0.2645	1.0033	0.9986	0.3097
16	0.380	1.0014	0.2558	1.0037	0.9992	0.3195
17	0.390	0.9990	0.2470	1.0012	0.9969	0.3273
18	0.400	1.0008	0.2399	1.0029	0.9986	0.3357
19	0.410	1.0008	0.2331	1.0028	0.9987	0.3426
20	0.418	1.0006	0.2277	1.0026	0.9986	0.3494
22	0.430	1.0007	0.2174	1.0026	0.9988	0.3675
25	0.443	1.0006	0.2051	1.0024	0.9988	0.3810
27	0.448	0.9997	0.1961	1.0014	0.9980	0.3999
30	0.455	1.0000	0.1858	1.0017	0.9984	0.4183
32	0.460	1.0001	0.1813	1.0016	0.9985	0.4276
35	0.465	1.0002	0.1735	1.0017	0.9987	0.4441
40	0.472	0.9998	0.1623	1.0013	0.9984	0.4724
45	0.481	1.0001	0.1530	1.0014	0.9987	0.4931
50	0.486	0.9999	0.1459	1.0012	0.9986	0.5168

Table 2 Results of simulations using α and β values reported by Song et al

n	α	β	M	σ_m	95% confidence limits		f
					Upper	Lower	
10	0.68	0.82	1.0859	0.3506	1.0890	1.0829	0.2628
20	0.62	1.00	1.0129	0.2251	1.0149	1.0109	0.3619
30	0.66	0.99	1.0139	0.1835	1.0155	1.0123	0.4282
40	0.59	0.92	0.9981	0.1589	0.9995	0.9967	0.4785
50	0.68	1.00	1.0113	0.1444	1.0126	0.9967	0.5191

listed for sample sizes between 9 and 50. The function did not yield unbiased estimates for $n < 9$. The values of α versus sample size in Table 1 are plotted in Fig. 2. Note that α increases sharply at low sample sizes and would be a negative value for $n = 8$. In Table 1, σ_m decreases and f increases with n, as expected. σ_m values in Table 1 almost match with the ones reported by Langlois and those produced in this study for Eq. (3). Hence α seems to affect only the average but not the standard deviation of m/m_{true}.

For comparison, simulations were run using α and β values reported by Song et al. for the five sample sizes they studied. For each sample size, 50,000 groups of data were generated. The results are presented in Table 2. Note that for all sample size with the exception of 50, α and β values of Song et al. yield biased estimates, as evidenced by the value of 1 being not included within the confidence limits. For $n = 10$, their findings are more biased than those for Eq. (3) ($M = 1.0592$ for this study in Fig. 1). The f results in Table 2 are slightly lower than those reported by Song et al., which may be due to more accurate prediction of distribution percentiles in this study because of higher number of replications.

When we compare standard deviations in Tables 1 and 2, the ones in Table 2 are only slightly less than those in Table 1 with the exception of $n = 10$. Since α seems to affect only the average, the slightly lower standard deviations are probably a result of β not being held at zero, although the effect of β seems to be very small. For $n = 10$, it seems possible to obtain α and β values that yield higher f values than those reported by Song et al.

The f values in Table 2 are higher for respective sample sizes in Table 1. This is due to the positive bias and/or lower standard deviation reported in Table 2. The distribution of m/m_{true} is positively skewed [4] and therefore increasing the average (bias) results in a higher fraction of the distribution to be within 0.9 and 1.1. When estimates have no bias ($n = 50$) or slightly negative bias ($n = 40$), f values in Tables 1 and 2 are almost identical, with slight differences as a result of lower standard deviations in Table 2.

In conclusion, a probability estimator that yields unbiased estimates of the Weibull modulus for sample sizes between 9 and 50 is provided. This estimator performs just as well as if not better than those reported in the literature.

References

1. Weibull W (1951) J App Mech 8:293
2. Langlois R (1991) J Mater Sci Lett 10:1049
3. Bergman B (1984) J Mater Sci Lett 3:689
4. Khalili A, Kromp K (1991) J Mater Sci 26:6741
5. Gong J (2000) J Mater Sci Lett 19:827
6. Davies IJ (2001) J Mater Sci Lett 20:997
7. Peterlik H (1995) J Mater Sci 30:1972
8. Song L, Wu D, Li Y (2003) J Mater Sci Lett 22:1651