J Mater Sci (2008) 43:1914–1919
DOI 10.1007/s10853-008-2457-9

Unbiased estimates of the Weibull parameters by the linear regression method

Murat Tiryakioğlu · David Hudak

Received: 4 October 2007 / Accepted: 8 January 2008 / Published online: 31 January 2008

Abstract For sample sizes from 5 to 100, the bias of the scale parameter was investigated for probability estimators, $P = (i - a)/(n + b)$, which yield unbiased estimates of the shape parameter. A class of unbiased estimators for both the shape and scale parameters was developed for each sample size. In addition, the percentage points of the distribution of unbiased estimate of the shape parameter were determined for all sample sizes. The distribution of the scale parameter was found to be normal by using the Anderson-Darling goodness-of-fit test. How the results can be used to establish confidence intervals on both the shape and scale parameters are demonstrated in the paper.

Introduction

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

$$P = 1 - \exp\left[-\left(\frac{\sigma}{\sigma_0}\right)^m\right] \tag{1}$$

where σ_0 is the scale parameter and m is the shape parameter, alternatively referred to as the Weibull modulus. There are several methods available in the literature to estimate the Weibull parameters: linear regression (least squares), weighted least squares, 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:

$$\ln[-\ln(1 - P)] = m\ln(\sigma) - m\ln(\sigma_0) \tag{2}$$

To estimate m and σ_0 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 [1–4] have been conducted to determine how probability estimators found in the literature perform. All probability estimators were found to give biased Weibull modulus results, i.e., the average of the estimated m values, referred to as $\hat{m}$, is not the same as true m (m_{true}). There is renewed interest in determining a set of probability estimators that yield unbiased estimates of the Weibull modulus. However, to the authors' knowledge, the bias in both the estimated scale ($\hat{\sigma}_0$) and shape parameters has not been investigated. The present study was motivated by the need to address the bias of the both the scale and shape parameters simultaneously.

Background

These probability estimators in the literature can all be written in the form

$$P = \frac{i - a}{n + b} \tag{3}$$

where i is the rank of the data point in the sample in ascending order, n represents the sample size, and a and b are numbers, such that $0 \le a \le 1$ and $0 \le b \le 1.0$. It has

M. Tiryakioğlu (✉)
Department of Engineering, School of Engineering, Mathematics and Science, Robert Morris University, 6001 University Boulevard, Moon Township, PA 15108, USA
e-mail: tiryakioglu@rmu.edu

D. Hudak
Department of Mathematics, School of Engineering, Mathematics and Science, Robert Morris University, 6001 University Boulevard, Moon Township, PA 15108, USA

been shown [1–5] via Monte Carlo simulations that different probability estimators yield different levels of bias, i.e., difference between true value of the Weibull modulus, m_{true}, and the average of the estimated Weibull moduli. A common technique in determining bias is to normalize the estimated Weibull moduli by m_{true}. This estimated normalized moduli, $\frac{\hat{m}}{m_{\text{true}}}$, will be referred to as $\hat{m}^*$. The average of $\hat{m}^*$ (M) is compared with 1. If the normalized average (M) is 1, then the probability estimator is unbiased. The scale parameter can be normalized similarly, $\frac{\hat{\sigma}_0}{\sigma_{0|\text{true}}}$, and will be referred to as $\hat{\sigma}_0^*$. The average of $\hat{\sigma}_0^*$ will be referred to as B.

Recently there has been renewed interest in finding the combination of a and b that yields an unbiased estimate of the Weibull modulus. Wu et al. [5] changed a and b simultaneously to find unbiased probability estimators for sample sizes between 10 and 50 at increments of 5. Tiryakioğlu [6] held $b = 0$ and changed a systematically until 1 was included in the 95% confidence limits of the average of normalized Weibull moduli for sample sizes between 9 and 50. Tiryakioglu and Hudak [7] investigated the bias in Weibull moduli estimated by 34 probability estimators for 27 sample sizes between 5 and 100. The authors developed regression equations for all sample sizes that can be used to estimate the bias of a probability estimator as a function of a and b in Eq. 3. They went on to show that there exist, for each sample size, a series of probability estimators, i.e., combinations of a and b, that yield unbiased estimates of the Weibull modulus, as presented in Fig. 1. Each contour in Fig. 1 represents the series of the unbiased probability estimators for the Weibull modulus. Tiryakioğlu and Hudak did not investigate the bias in the estimated scale parameter.

Langlois [2] investigated how the four methods to estimate Weibull parameters perform between sample sizes of 5 and 50. For the linear regression method, the author used two probability estimators: $a = 0.5$, $b = 0$ and $a = 0.3$, $b = 0.4$. Langlois only commented that all methods yielded a bias of less than 1% for the scale parameter but did not present any results. Khalili and Kromp [1] investigated the bias and the distribution of the estimated scale parameter for the linear regression (with probability estimator $a = 0.5$, $b = 0$), maximum likelihood, and moments' methods for $n = 30$. The authors found that the three methods yield similar bias within 0.2%. They also plotted histograms of the estimated scale parameter and commented that its distribution is negatively skewed, although slightly. Thoman et al. [8] provided percentage points for the distribution of $\hat{m}^* \ln\left(\sigma_0^*\right)$estimated by the maximum likelihood method, but did not address the distribution of the estimated scale parameter. To the authors' knowledge, the distribution of $\hat{\sigma}_0^*$ has not been tested for goodness-of-fit to known distributions.

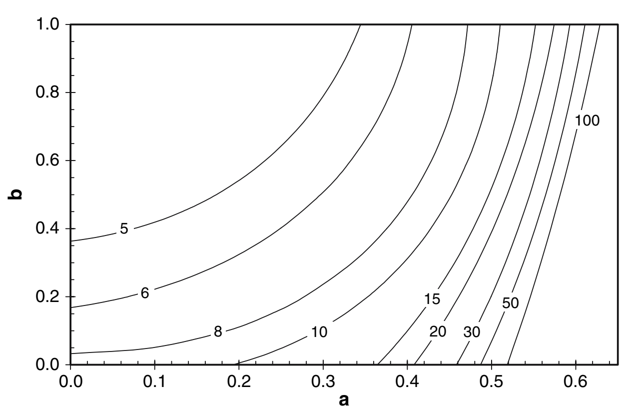

Fig. 1 Contours of unbiased probability estimators for the Weibull modulus [7]

The distribution of the estimated Weibull modulus has received much more attention than that of the estimated scale parameter. Ritter et al. [9] ran Monte Carlo simulations and concluded that the distribution of the estimated Weibull modulus is approximately normal. These researchers ran Monte Carlo simulations only 100 times. It has since been shown [1, 10–12] that the distribution of m is positively skewed. Gong and Wang [10] stated that m follows a lognormal distribution for linear regression (using Eq. 2) and maximum likelihood methods. These authors used the χ^2 goodness-of-fit test for their evaluation. Barbero et al. [11] claimed that the distribution of m estimated by the maximum likelihood method is better expressed by a 3-parameter Weibull distribution. In a later publication [12], the same authors found that 3-parameter log-Weibull distribution provides a better fit to m estimated by the maximum likelihood method than lognormal and 3-parameter Weibull distribution. Recently Tiryakioğlu and Hudak [7] analyzed the distribution of m estimated by the linear regression method using the Anderson-Darling goodness-of-fit test [13–15]:

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n} \left[(2i-1)\ln P_i + (2n-1-2i)\ln(1-P_i)\right] \tag{4}$$

The Anderson-Darling goodness-of-fit test is much more sensitive to tails than the χ^2 test. The authors found that $\hat{m}$ does not follow the normal, lognormal, 3-parameter Weibull or 3-parameter log-Weibull distributions for $5 \le n \le 100$. Because $\hat{m}$ does not follow a known distribution, percentage points for the distribution of $\hat{m}^*$ need to be known to establish confidence intervals for the Weibull modulus.

The literature survey presented above indicates that these issues need to be addressed:

- How does the bias of the estimated scale parameter change among the series of unbiased probability estimators for the Weibull modulus?
- Are there probability estimators that yield unbiased estimates for both Weibull parameters?
- Can equations be developed to estimate percentage points for the distribution of $\hat{m}$?
- Does the estimated scale parameter follow a known distribution?

These issues have been investigated in this study and results are reported.

Research methodology

Monte Carlo simulations were used to generate n data points from a Weibull distribution with $\sigma_{0|\text{true}} = 1$ and $m_{\text{true}} = 3$. Combinations of a and b were chosen using the regression equations provided by Tiryakioğlu and Hudak [7] so that the estimated Weibull modulus was unbiased. In other words, all probability estimators used in this study were along the contours shown in Fig. 1. Thirty sample sizes (n) ranging from 5–100 were investigated. At each iteration of the simulations, n random numbers between 0 and 1 were generated to obtain a set of σ values.

For each n, an iterative procedure was employed to calculate the combination of a and b that yielded unbiased results as follows. Using the A and the standard deviation of σ_0^*, $s_{\sigma_0^*}$, for each n, confidence intervals for true mean of distribution (μ_B) were calculated as

$$B - z\frac{s_{\sigma_0^*}}{\sqrt{n_r}} \leq \mu_B \leq B + z\frac{s_{\sigma_0^*}}{\sqrt{n_r}} \tag{5}$$

where z is 1.95996 for 95% confidence. The values of a and b were varied along the unbiased contours for the Weibull modulus until $\mu_B = 1$ was within the confidence intervals. For each sample size and probability estimator, the experiment was repeated 20,000 times (=n_r).

Results and discussion

The effect of unbiased probability estimators for m on B

The value of B was found to increase along the contours for the unbiased parameters of the Weibull modulus with increasing values of b (and a) in Eq. 3 for every sample size, as presented in Fig. 2. Note that b (and a) has an effect on the bias, the strength of which decreases with increasing sample size. To the authors' knowledge, this is the first time that the effect of probability estimators on the bias of the estimated scale parameter is reported. Figure 2 also shows the unbiased scale parameter (as indicated by the dashed line) is obtained at $b = 0$ for sample sizes of 30 and 50. For $n = 10$, however, b is approximately 0.2.

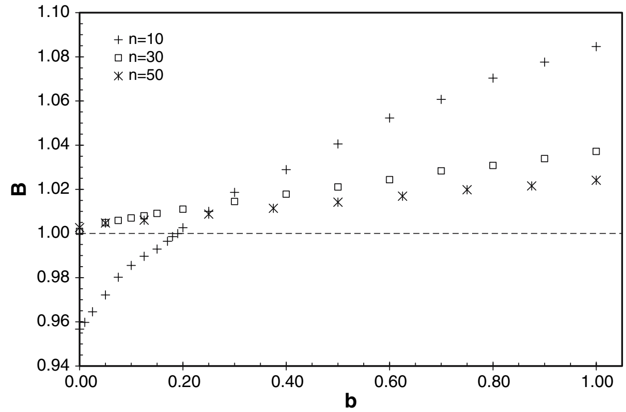

Fig. 2 The effect of b on B for three sample sizes

Table 1 Unbiased probability estimators for both Weibull parameters

n	a	b
5	0.173	0.500
6	0.243	0.390
7	0.280	0.310
8	0.309	0.251
9	0.322	0.210
10	0.348	0.190
11	0.367	0.160
12	0.371	0.130
13	0.382	0.110
14	0.388	0.100
15	0.394	0.080
17	0.407	0.050
20	0.417	0.030
22	0.430	0.000
25	0.443	0.000
27	0.448	0.000
30	0.455	0.000
32	0.460	0.000
35	0.465	0.000
40	0.472	0.000
45	0.481	0.000
50	0.486	0.000
55	0.499	0.000
60	0.503	0.000
65	0.509	0.000
70	0.518	0.000
75	0.522	0.000
80	0.516	0.000
90	0.518	0.000
100	0.519	0.000

Table 2 Percentage points of the distribution of $\hat{m}^*$ obtained by using the unbiased probability estimators in Table 1

n	0.005	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99	0.995
5	0.292	0.325	0.381	0.434	0.504	1.630	2.014	2.458	3.169	3.653
6	0.320	0.358	0.418	0.475	0.549	1.555	1.852	2.189	2.694	3.052
7	0.356	0.391	0.445	0.501	0.579	1.502	1.747	2.031	2.462	2.753
8	0.374	0.409	0.470	0.530	0.602	1.480	1.705	1.950	2.293	2.579
9	0.392	0.427	0.488	0.545	0.618	1.444	1.650	1.836	2.126	2.371
10	0.404	0.447	0.504	0.563	0.638	1.419	1.605	1.791	2.051	2.298
11	0.429	0.465	0.529	0.582	0.652	1.399	1.574	1.746	2.003	2.194
12	0.436	0.484	0.539	0.593	0.662	1.380	1.543	1.690	1.892	2.042
13	0.450	0.488	0.550	0.610	0.677	1.362	1.510	1.644	1.837	1.982
14	0.459	0.503	0.560	0.618	0.687	1.350	1.493	1.631	1.792	1.930
15	0.462	0.503	0.568	0.625	0.695	1.334	1.465	1.603	1.769	1.900
17	0.491	0.528	0.594	0.643	0.708	1.324	1.449	1.574	1.735	1.859
20	0.523	0.557	0.611	0.664	0.728	1.299	1.412	1.514	1.644	1.760
22	0.536	0.575	0.635	0.683	0.746	1.282	1.384	1.480	1.610	1.714
25	0.565	0.597	0.648	0.697	0.754	1.266	1.363	1.447	1.568	1.643
27	0.559	0.601	0.652	0.702	0.762	1.254	1.341	1.430	1.538	1.613
30	0.576	0.611	0.667	0.716	0.773	1.241	1.328	1.405	1.493	1.565
32	0.592	0.633	0.684	0.729	0.781	1.234	1.311	1.384	1.480	1.537
35	0.612	0.645	0.690	0.738	0.790	1.226	1.300	1.369	1.453	1.523
40	0.632	0.662	0.708	0.749	0.798	1.207	1.276	1.342	1.424	1.486
45	0.633	0.670	0.723	0.764	0.813	1.196	1.260	1.320	1.388	1.440
50	0.654	0.684	0.734	0.773	0.820	1.190	1.250	1.303	1.374	1.420
55	0.670	0.700	0.748	0.785	0.830	1.183	1.239	1.291	1.355	1.406
60	0.676	0.709	0.749	0.790	0.833	1.174	1.229	1.277	1.332	1.380
65	0.691	0.720	0.762	0.798	0.841	1.168	1.218	1.266	1.321	1.361
70	0.703	0.731	0.769	0.804	0.845	1.159	1.210	1.259	1.312	1.348
75	0.704	0.735	0.776	0.810	0.850	1.158	1.207	1.252	1.304	1.347
80	0.713	0.740	0.780	0.817	0.855	1.150	1.197	1.238	1.289	1.327
90	0.723	0.755	0.790	0.824	0.862	1.141	1.184	1.222	1.273	1.304
100	0.742	0.769	0.805	0.834	0.870	1.132	1.173	1.211	1.250	1.287

For all sample sizes, the combinations of a and b that yield unbiased estimates for both the scale and shape parameters were determined. These unbiased probability estimators are listed in Table 1. It is recommended that these probability estimators be used when the linear regression method is employed to estimate the parameters of the Weibull distribution.

Percentage points for the distribution of $\hat{m}^*$

Since the distribution of $\hat{m}^*$ does not follow any distribution tested in the literature and the standard deviation and bias are correlated [7], percentage points (X) of the unbiased probability estimators listed in Table 1 were developed. The results are presented in Table 2. These points can be used to establish confidence limits on the estimated Weibull modulus. For instance, if $\hat{m} = 26.5$ for $n = 30$, then 95% confidence interval for m_{true} can be found as follows:

$$X_{0.025,30} \leq \frac{\hat{m}}{m_{\text{true}}} \leq X_{0.975,30} \tag{6}$$

From Table 2, $X_{0.025,30}$ and $X_{0.975,30}$ are 0.667 and 1.405, respectively. Therefore,

$$0.667 \leq \frac{26.5}{m_{\text{true}}} \leq 1.405 \tag{6.a}$$

$$\frac{26.5}{1.405} \leq m_{\text{true}} \leq \frac{26.5}{0.667} \tag{6.b}$$

Hence m_{true} lies between 18.86 and 39.73 with 95% confidence.

An effort was made to develop an empirical equation to interpolate the percentage points to all sample sizes between 5 and 100. Best results were obtained using

$$X = \frac{\beta_0 + \beta_1 n}{n^{\xi}} \tag{7}$$

where β_0, β_1, and ζ are constants, the values of which are presented in Table 3 for the percentage points determined

Table 3 Constants for Eq. 7 for various percentage points

	β_0	β_1	ζ
0.005	−0.9373	0.3844	0.8519
0.01	−0.9525	0.4246	0.8659
0.025	−1.0544	0.5040	0.8938
0.05	−1.1094	0.5764	0.9152
0.1	−1.0995	0.6612	0.9366
0.9	1.6983	1.4367	1.0542
0.95	3.3098	1.5143	1.0599
0.975	5.1911	1.5501	1.0606
0.99	8.1320	1.5458	1.0563
0.995	10.4428	1.5056	1.0481

in this study. The percentage points listed in Table 2 and the predictions of Eq. 7 are presented in Fig. 3, which shows an excellent agreement. In Fig. 3, the solid line represents the predicted values of the percentile points and the markings represent the actual values from Table 2.

The distribution of $\hat{\sigma}_0^*$

The histogram of $\hat{\sigma}_0^*$ for $n = 30$ obtained with the unbiased probability estimator in Table 1 is presented in Fig. 4. Note that the distribution is not skewed and there is strong indication that it may be normal. Hence, the distribution of $\hat{\sigma}_0^*$ was tested for normality using the Anderson-Darling goodness-of-fit test. The results are presented in Table 4, which shows that the hypothesis that the distribution is normal could not be rejected because p-values for all sample sizes are larger than 0.05, the value most commonly used in hypothesis testing. This is the first time that the distribution of the scale parameter was shown to follow a known distribution.

That $\hat{\sigma}_0^*$ follows the normal distribution does not agree with the observations of Khalili and Kromp who stated that

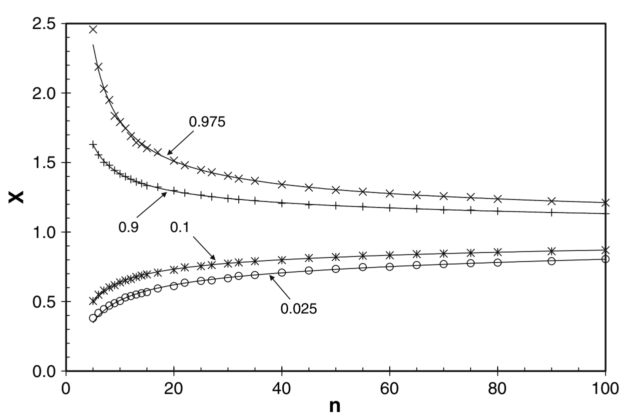

Fig. 3 The percentage points determined from the experiments and those predicted by Eq. 7 for all sample sizes

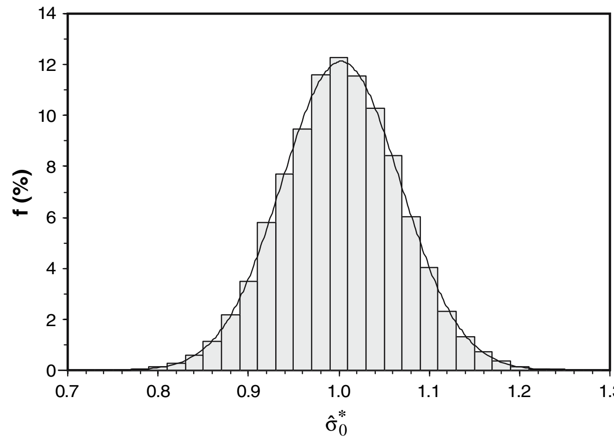

Fig. 4 Histogram of estimated scale parameter using the unbiased probability estimator in Table 1 for $n = 30$

Table 4 The results of the Anderson-Darling goodness-of-fit test and $s_{\sigma_0^*}$ for all sample sizes

n	A^2	p-value	$s_{\sigma_0^*}$
5	0.358	0.453	0.1579
6	0.712	0.063	0.1457
7	0.348	0.479	0.1333
8	0.644	0.093	0.1252
9	0.632	0.100	0.1182
10	0.209	0.865	0.1133
11	0.752	0.051	0.1074
12	0.579	0.132	0.1028
13	0.497	0.212	0.0984
14	0.683	0.075	0.0952
15	0.470	0.247	0.0917
17	0.670	0.080	0.0871
20	0.415	0.334	0.0808
22	0.699	0.068	0.0761
25	0.407	0.350	0.0725
27	0.635	0.098	0.0697
30	0.266	0.690	0.0662
32	0.657	0.086	0.0635
35	0.698	0.069	0.0610
40	0.417	0.331	0.0566
45	0.585	0.128	0.0543
50	0.339	0.503	0.0516
55	0.368	0.431	0.0488
60	0.182	0.913	0.0463
65	0.486	0.226	0.0449
70	0.418	0.328	0.0435
75	0.267	0.689	0.0420
80	0.286	0.625	0.0403
90	0.318	0.537	0.0382
100	0.212	0.856	0.0361

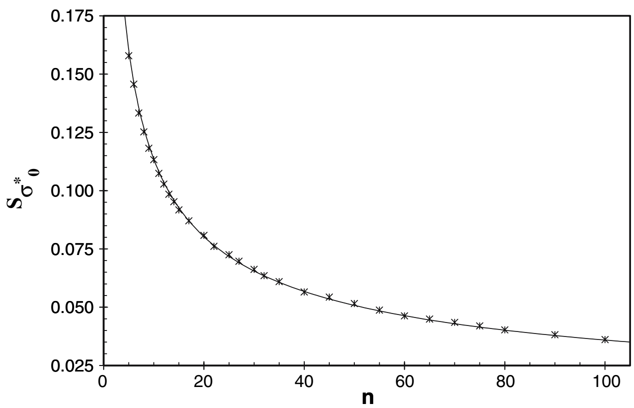

Fig. 5 The change in $s_{\sigma_0^*}$ with sample size and the fit by Eq. 8

the distribution is negatively skewed. Since normal distribution is symmetrical around its mean, it is not a skewed distribution. The reason for this anomaly is unknown. Since our research only considered distributions for unbiased estimators of the scale parameter in combination with unbiased shape parameters, it can be speculated that the bias may have an effect on the skewness of the scale parameter. More research is needed in this area.

The standard deviation of $\hat{\sigma}_0^*$, $s_{s_0^*}$, was determined for all sample sizes and is presented in Table 4. The standard deviation was found to change with $n^{-1/2}$:

$$s_{\sigma_0^*} = \frac{0.359}{\sqrt{n}} \tag{8}$$

The standard deviations and the fit of Eq. 8 to data are presented in Fig. 5, which shows excellent agreement. Equation 8 can be used to interpolate to sample sizes not investigated in this study.

The result that the distribution of $s_{\sigma_0^*}$ is normal and Eq. 8 can be combined to establish confidence limits on the scale parameter. For instance, for $n = 29$ and $\hat{\sigma}_0 = 52.1$ estimated using the probability estimator in Table 1, the confidence limits for 95% confidence are

$$1.000 - 1.95996\frac{0.359}{\sqrt{n}} \leq \frac{\hat{\sigma}_0}{\sigma_{0|\text{true}}} \leq 1.000 + 1.95996\frac{0.359}{\sqrt{n}} \tag{9}$$

Hence

$$\frac{52.1}{1.000 + 1.95996\frac{0.359}{\sqrt{29}}} \leq \sigma_{0|\text{true}} \leq \frac{52.1}{1.000 - 1.95996\frac{0.359}{\sqrt{29}}} \tag{9.a}$$

The 95% confidence limits for the scale parameter are 46.08 and 59.93.

Conclusions

- The values of a and b in Eq. 3 affect the bias in the estimated scale parameter.
- A set of probability estimators that yield unbiased estimates for both the scale and shape parameters were developed in this study for thirty sample sizes ranging from 5–100.
- Using the unbiased probability estimators, percentage points for the Weibull modulus were developed. In addition, empirical equation for percentage points was introduced to interpolate to sample sizes not investigated in this study. These percentage points can be used to develop confidence limits for the Weibull modulus.
- The distribution of the estimated scale parameter was found to be normal by using the Anderson-Darling goodness-of-fit test.
- The use of the empirical equation for the standard deviation of $\hat{\sigma}_0^*$ to develop confidence limits for the scale parameter was demonstrated in this study.

References

1. Khalili A, Kromp K (1991) J Mater Sci 26:6741
2. Langlois R (1991) J Mater Sci Lett 10:1049
3. Bergman B (1984) J Mater Sci Lett 3:689
4. Trustrum K, de Jayatilaka AS (1979) J Mater Sci 14:1080
5. Wu D, Zhoua J, Li Y (2006) J Eur Cer Soc 26:1099
6. Tiryakioğlu M (2006) J Mater Sci 41:5011
7. Tiryakioğlu M, Hudak D (2007) J Mater Sci 42:10173
8. Thoman DR, Bain LJ, Antle CE (1969) Technometrics 11:445
9. Ritter J, Bandyopadhyay N, Jakus K (1981) Amer Cer Soc Bull 60:788
10. Gong J, Wang J (2002) Key Eng Mater 224–226:779
11. Barbero E, Fernandez-Saez J, Navarro C (2000) Composites: Par2 31:375
12. Barbero E, Fernandez-Saez J, Navarro C (2001) J Mater Sci Lett 20:847
13. Anderson TW, Darling DA (1954) J Amer Stat Assoc 49:765
14. Stephens MA (1974) J Amer Stat Assoc 69:730
15. Stephens MA (1986) In: D'Agostino RB, Stephens MA (eds) Goodness of fit techniques. Marcel Dekker, p 97