J Mater Sci (2006) 41:6479–6480
DOI 10.1007/s10853-006-0667-6

LETTER

Comment on the probability indices

S. Nadarajah · S. Kotz

Received: 6 June 2006 / Accepted: 10 July 2006 / Published online: 29 August 2006

The choice of the probability index for fitting the two-parameter Weibull equation has been a subject of research in the *Materials Science* literature. The recent papers by Gong [1] and Wu and Jiang [2] have discussed about the choice of a probability index for fitting the two-parameter Weibull equation given by

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

where m and σ_0 are the Weibull modulus and scale parameter, respectively. In the case of fracture strength data, the most commonly used method for fitting (1) is by linear regression (LR). This involves ranking the strength data in an ascending order and then assigning a probability of failure to each datum. Suppose there are n specimens tested and let P_i denote the probability of failure for the ith ranked datum. Gong [1] proposed the following estimator for P_i:

$$P_i = \frac{i - 0.999}{n + 1000}. \tag{2}$$

In a later paper, Wu and Jiang [2] argued that the estimators given by

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

and

$$P_i = \frac{i}{n + 1} \tag{4}$$

have several advantages over that given by (2).

We would like to point out that the choice of the form for P_i has been well established in the statistics literature. The most commonly known and accepted form for P_i is:

$$P_i = \frac{i - 0.375}{n + 0.25}. \tag{5}$$

There are both theoretical and empirical reasons for the choice of (5), see Blom [3] and Chambers et al. [4]. The choice given by (5) is what is used in most statistical packages, e.g. Genstat, Minitab, R, SPSS, SAS and S-Plus.

We perform a simulation study involving 400 data sets each of size 100 to show that the index given by (5) is superior to (2)–(4). We followed the following scheme:

- Simulate 100 samples each of size 100 from (1) for given σ_0 and m. Suppose $\{\sigma_1, \sigma_2, ..., \sigma_{100}\}$ denotes the enumeration of a sample.
- Compute the following for each simulated sample:

$$M_1 = \sum_{i=1}^{100}\left|1 - \exp\left[-\left(\frac{\sigma_i}{\sigma_0}\right)^m\right] - \frac{i - 0.999}{n + 1000}\right|,$$

$$M_2 = \sum_{i=1}^{100}\left|1 - \exp\left[-\left(\frac{\sigma_i}{\sigma_0}\right)^m\right] - \frac{i - 0.5}{n}\right|,$$

$$M_3 = \sum_{i=1}^{100}\left|1 - \exp\left[-\left(\frac{\sigma_i}{\sigma_0}\right)^m\right] - \frac{i}{n + 1}\right|,$$

and

S. Nadarajah (✉) · S. Kotz
School of Mathematics, University of Manchester,
Manchester M13 9PL, UK
e-mail: saralees.nadarajah@manchester.ac.uk

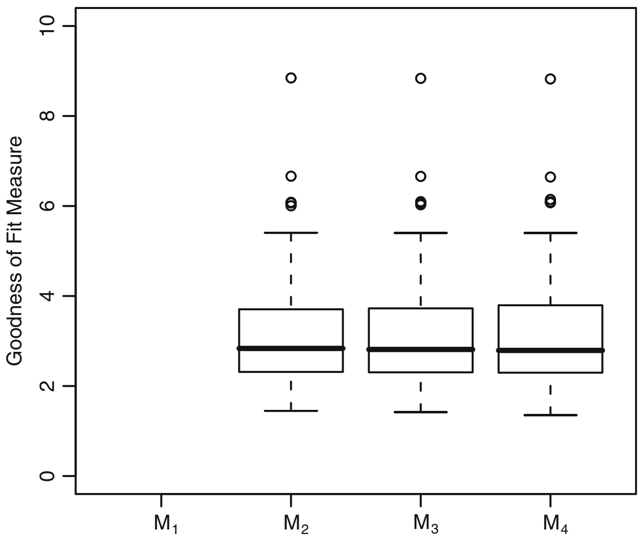

Fig. 1 Box plots of the values of M_i, $i = 1, 2, 3, 4$ for 100 samples (each of size 100) from (1) with $\sigma_0 = 1$ and $m = 2$

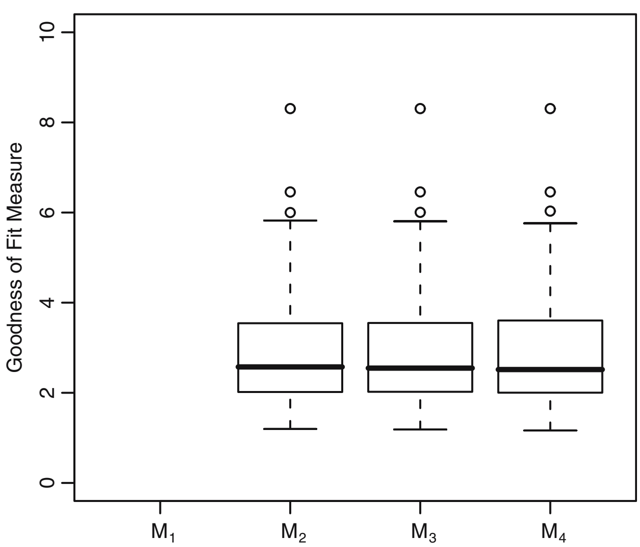

Fig. 2 Box plots of the values of M_i, $i = 1, 2, 3, 4$ for 100 samples (each of size 100) from (1) with $\sigma_0 = 1$ and $m = 3$

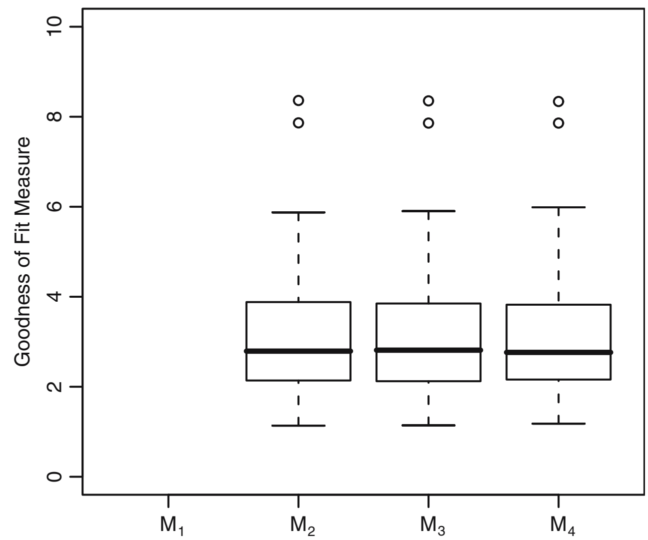

Fig. 3 Box plots of the values of M_i, $i = 1, 2, 3, 4$ for 100 samples (each of size 100) from (1) with $\sigma_0 = 1$ and $m = 5$

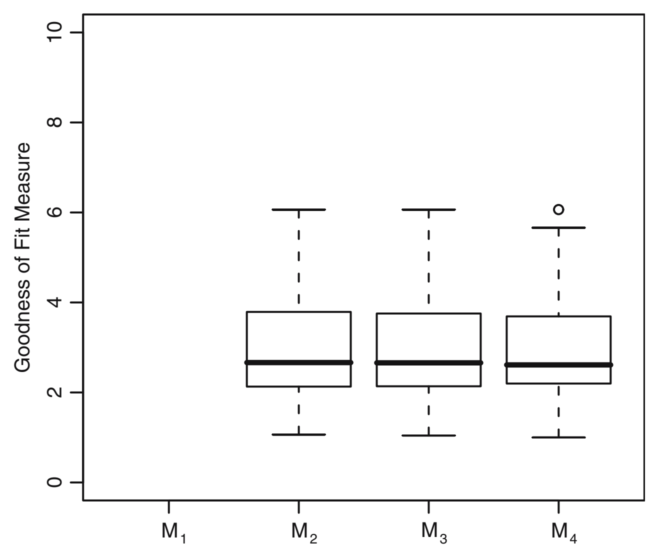

Fig. 4 Box plots of the values of M_i, $i = 1, 2, 3, 4$ for 100 samples (each of size 100) from (1) with $\sigma_0 = 1$ and $m = 10$

$$M_4 = \sum_{i=1}^{100} \left| 1 - \exp\left[-\left(\frac{\sigma_i}{\sigma_0} \right)^m \right] - \frac{i - 0.375}{n + 0.25} \right|.$$

Note that M_1 can be used to measure the goodness of fit of (2), M_2 can be used to measure the goodness of fit of (3), and so on.

The above scheme was executed with $\sigma_0 = 1$ and $m = 2, 3, 5, 10$. The box plots of the values of M_i, $i = 1, 2, 3, 4$ for the four different values of m are shown in Figs. 1–4. In each figure, the smallest values for the goodness of fit measure are given by M_4. The index due to Gong [1] produces the worst performance in each case with the box plot outside the plotting range.

References

1. Gong JH (2000) J Mater Sci Lett 19:827
2. Wu DF, Jiang H (2003) J Mater Sci Lett 22:1745
3. Blom G (1958) Statistical estimates and transformed beta-variables. John Wiley and Sons, New York, Chapter 3
4. Chambers J, Cleveland W, Kleiner B, Tukey P (1983) Graphical methods for data analysis. Chapman and Hall, London, Chapter 3