Characterizing the Relationship between Weibull Location Parameter and the Minimal Observation in a Small Size of Sample Based on Stochastic Simulation

Bo Qin, Ningxiang Wu, Liyang Xie

Department of Mechanical Engineering, Northeastern University, Shenyang, China
Email: lyxie@me.neu.edu.cn

Abstract

Three-parameter Weibull distribution is one of the preferable distribution models to describe product life. However, it is difficult to estimate its location parameter in the situation of a small size of sample. This paper presents a stochastic simulation method to estimate the Weibull location parameters according to a small size of sample of product life observations and a large amount of statistically simulated life date. Big data technique is applied to find the relationship between the minimal observation in a product life sample of size n ($n \geq 3$) and the Weibull location parameter. An example is presented to demonstrate the applicability and the value of the big data based stochastic simulation method. Comparing with other methods, the stochastic simulation method can be applied to very small size of sample such as the sample size of three, and it is easy to apply.

1. Introduction

To estimate product probabilistic life by means of inferential statistics approach, a large size sample of life data are usually required [1] [2] [3]. In the situation of small size of sample, both distribution type identification and distribution parameter estimation are difficult. Weibull distribution is widely applied to describe product life since its flexibility [4], as well as the fact that the shape parameter of a Weibull-distributed product life is mainly determined by failure mechanism.

Lots of experiences have demonstrated that the shape parameter values of the Weibull life distributions of the same kind of products with the same failure mode do not differ from each other considerably.

Due to the difficulty to estimate the location parameter of the Weibull distribution in small size of sample situations, the mostly applied is two-parameter Weibull distribution [5] [6] [7] [8] [9]. To estimate product life distribution in the situation of small size sample, H.K.T. Ng *et al.* studied three-parameter Weibull distribution parameter estimation methods based on progressively Type-II right censored sample [10]. Abbasi *et al.* discussed an approach that takes the advantage of ANN, proposed a simple neural network that simultaneously estimates the three parameters, exploiting the concept of moment method to estimate Weibull parameters using mean, standard deviation, median, skewness and kurtosis [11].

The situations that life distribution estimation has to be carried out according to small size sample of life observations will be more and more encountered in engineering, pertinent theories, methods and techniques have attracted ever increasing investigations. This paper focuses on new principle and new method to estimate the location parameter of three-parameter Weibull distributed product life.

2. On Weibull Location Parameter Estimation

The probability density function of a three-parameter Weibull distribution W(β, η, γ) is

$$f(t) = \frac{\beta(t-\gamma)^{\beta-1}}{\eta^{\beta}} \exp\left[-\left(\frac{t-\gamma}{\eta}\right)^{\beta}\right], \quad t \geq \gamma \tag{1}$$

where, β, η and γ stand for the shape parameter, scale parameter and location parameter, respectively.

2.1. On Weibull Shape Parameter

For a Weibull distributed product life, the shape parameter is believed to be dependent mainly on the underlying failure mechanism, and it is roughly a constant for a certain kind of products operating in similar environments. For instance, experience indicates that the range of the shape parameter of the Weibull distributed ball bearing life, failed in contact fatigue, is between 1.5 and 2.5. That makes a possibility to estimate the location parameter according to a small size of sample, since when the shape parameter is less than 2.5, the left tail of the Weibull distribution curve is not so flat, as shown in Figure 1. In the following ball bearing life distribution estimation, the shape parameter is taken as 2.0.

2.2. On Weibull Location Parameter Estimation Approach

A specific probability density function can be identified, and its parameters can be accurately estimated if a large size of sample is available. For a Weibull distributed
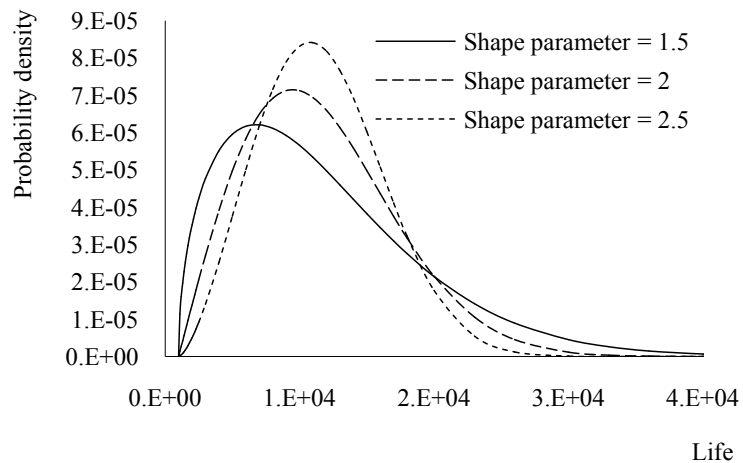
Figure 1. Weibull distributions with different shape parameters.

life random variable, if $t_{(1)}, t_{(2)}, \cdots, t_{(n)}$ are its n observations arranged in ascending order, the location parameter will be close to $t_{(1)}$ if n is large, especially in condition of the shape parameter is small (e.g., $\beta \leq 2.5$).

By means of the median rank estimator of a cumulative distribution function $F(t_i)$, a sample size n_p can be figured out with which the minimal observation $t_{(1)}$ can be taken as the estimator of the $p \times 100$th percentile for a specific Weibull distribution.

The general form of the median rank estimator is:

$$\hat{F}(t_i) = \frac{i}{i + (n+1-i) F_{2(n+1-i), 2i, 0.5}}, \quad i = 1, 2, \cdots, n \tag{2}$$

where, i is the ordinal number of the individual life observations in ascending order; n is the size of the sample; $F_{2(n+1-i), 2i, 0.5}$ is a value with the upper tail area equaling to 0.5 from the F-distribution with the degrees of freedom $2(n + 1 - i)$ and $2i$.

This estimator is usually approximated as

$$\hat{F}(t_i) \approx \frac{i - 0.3}{n + 0.4} \quad (i = 1, 2, \cdots, n) \tag{3}$$

The percentile associated with $t_{(1)}$ can be estimated according to a sample containing n observations as

$$p_{1/n} = \frac{1 - 0.3}{n + 0.4} \tag{4}$$

The difference between the location parameter γ and the $p_{1/n} \times 100$th percentile $t_{(1,n)}$ can be derived from the Weibull distribution function as

$$1 - e^{-\left(\frac{t_{(1,n)} - \gamma}{\eta}\right)^{\beta}} = p_{1/n}$$

i.e.,

$$t_{(1,n)} - \gamma = \eta \left[-\ln\left(1 - p_{1/n}\right) \right]^{1/\beta} \tag{5}$$

This equation, together with Equation (4), can be used in the situation of large size of sample. In the situation that only a small size of sample is available, a new method is required, as the median rank estimator is only meaningful in large size sample conditions.

Evidently, the difference $t_{(1,n)} - \gamma$ is a function of the Weibull distribution parameter η and β. For a specific engineering problem, the Weibull shape parameter can be approximately assigned according to similar products. In the present paper, the shape parameter of the Weibull distribution of the ball bearing life is assigned as 2.0. The scale parameter needs to be estimated through life data, and the estimator is location parameter dependent. Despite the coupling mechanism, suitable technique can be applied to get an approximate relationship between $t_{(1,n)} - \gamma$ and the associated sample size n. Through modifying the minimal observation in a sample of size n (denoted by $t_{(1,n)}$), an estimator for location parameter γ can be constructed. Consequently, a reasonable three-parameter Weibull distribution can be obtained, that will be conservative, yet better than a conventionally estimated two-parameter Weibull distribution.

For engineering applications, a sufficient large size of sample, such as the sample size $n = 518$, can be assigned by that the minimal observation $t_{(1,518)}$ can be taken as the Weibull location parameter γ. The sample size 518 makes $t_{(1,518)}$ corresponds to a very low cumulative probability 0.00135, that is the lower bound of the 6σ for the situation of Gauss distribution (the normal distribution).

However, such a large size of sample is hard to obtain in engineering application. Obviously, the less the sample size, the greater the difference between the location parameter and the minimal observation. For small size sample situations, the minimal observation has to be revised to be the estimated location parameter. To obtain a reasonable function to modify the minimal observation, Monte Carlo simulation technique is applied as below.

The basic idea is that, for a given Weibull distribution W(β, η, γ), the possible difference between the location parameter γ and the minimal observation $t_{(1,n)}$ in an arbitrary sample of size n can be demonstrated by repeatedly Monte Carlo sampling, and a possible interval can be statistically figured out. Obviously, the more the sampling times, the higher the confidence level. Based on a great number of sampling, an upper envelope curve of the simulated date points can be used to conservatively describe the relationship between the possible maximal difference $\Delta_{\max}$ ($\Delta_{\max} = \max \left\{ t_{(1,n),i} - \gamma \right\}$, $i = 1, 2, \cdots, m$) and sample size n. Where, $t_{(1,n),i}$ stands for the minimal observation in the ith sampling, m stands for the number of the trials. Based on such a relationship, an function can be constructed to modify the minimal observation $t_{(1,n)}$ into the estimated location parameter $\hat{\gamma}$.

With a given Weibull distribution W(β, η, γ), it is easy to get observations $t_1, t_2, \cdots, t_n$ through Monte Carlo sampling. For a sample of size n, the difference between the minimal observation $t_{(1,n)}$ and the location parameter γ, *i.e.* ($t_{(1,n)} - \gamma$) is the value by which the minimal observation should be deducted to be taken as

the estimated location parameter, *i.e.*, $t_{(1,n)} - \left(t_{(1,n)} - \gamma\right) = \gamma$. Demonstrated and discussed below are the simulation results for a Weibull distributed life random variable with $\beta = 2.0$, $\eta = 1000$, and $\gamma = 2000$, *i.e.* W(2.0, 1000, 2000).

Shown in **Figure 2** are a part of the simulation results of the minimal observations in different size of samples. For example, the 10 points with the abscissa of 20 in **Figure 2(a)**, *i.e.*, the column consist of the 10 points corresponding to sample size $n = 20$, are the ten respective minimal observations generated by ten times of sampling of sample size of twenty, respectively.

In detail, the three observations (t_1, t_2, t_3) with the abscissa $n = 3$ are randomly taken from the Weibull distributed population in each trial. The minimum, denoted by $t_{(1,3)}$, is presented as a date point located at the vertical line with the abscissa $n = 3$ in the graph, that will be used with other data points together to evaluate the possible difference between the Weibull location parameter and the minimal observation with respective to different sample sizes.
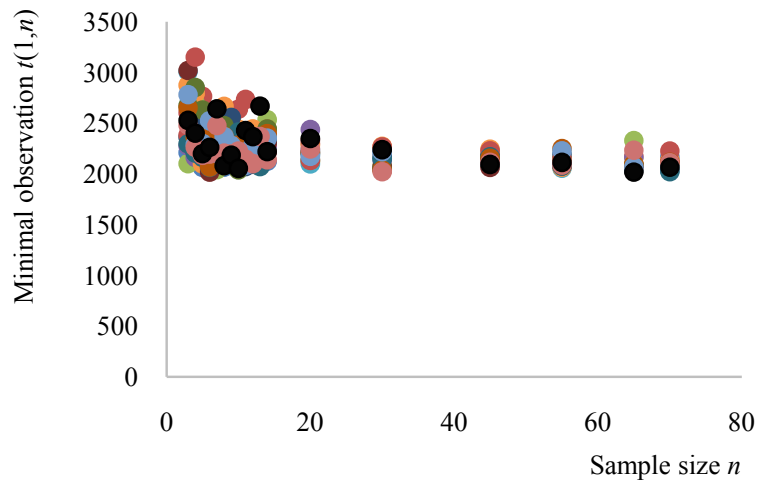
The minimal sample life values represented by the ordinates of the individual points in **Figure 2**, which are the respective minimums in the individual samples of the respective trials, clearly show a tendency that when sample size becomes
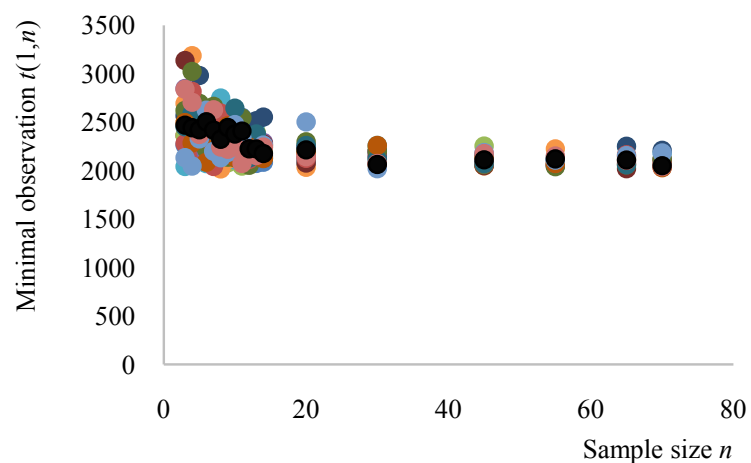
(a)
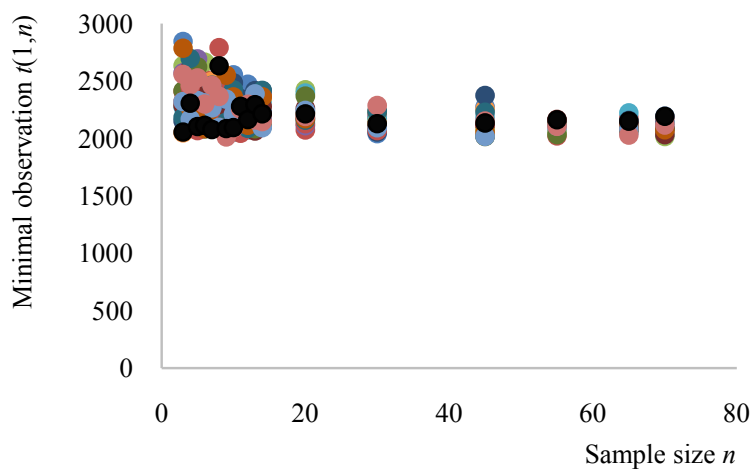
(b)

(c)

Figure 2. Simulated minimal life observations corresponding to different size of samples (*n* = 1, 2, …, 70, respectively) from W(2.0, 1000, 2000). (a) Results of the first group of the simulations; (b) Results of the second group of the simulations; (c) Results of the third group of the simulations.

small, the minimal observation might be much greater than the location parameter; when the sample size is large enough, the minimal observation will not be greater considerably than the location parameter. To obtain a reasonable estimation of the location parameter based on a small size sample of observations, the difference between the location parameter and the minimal observation has to be appropriately figured out. Owing to the stochastic characteristic of the minimal observation, a conservative decrement should be applied to accommodate the deviations yielded by the majority of observations, such as 99% of the observations.

To get a conservative model to modify the minimal observation into the location parameter, the upper boundary of the individual minimal observations from the different size of samples needs to be described by a suitable function. Many methods, including the variety of traditional methods and machine leaning, can be used to fit the envelop line according to the simulated data. By trial and error, the following equation is constructed to describe the upper boundary:

$$t_{(1,n)}^{\max} = \gamma - 2\eta \lg\left(n/70\right)^{1/\beta} \tag{6}$$

where, 70 is the sample size $n_{0.01}$ in such a sample the minimum $t_{(1,70)}$ being corresponded to the first percentile of the random life as $(1-0.3)/(70+0.4) \approx 0.01$.

The curve expressed by Equation (6) is drawn in **Figure 3** together with the individual minimal observations generated by Monte Carlo sampling at different trials. Great number of simulation results show that almost all the minimal observations from the different size of samples are below this curve when the sample size *n* is three or larger. Therefore, the minimal observation should be modified by deducting ($t_{(1,n)}^{\max} - \gamma$) to become the estimated location parameter.
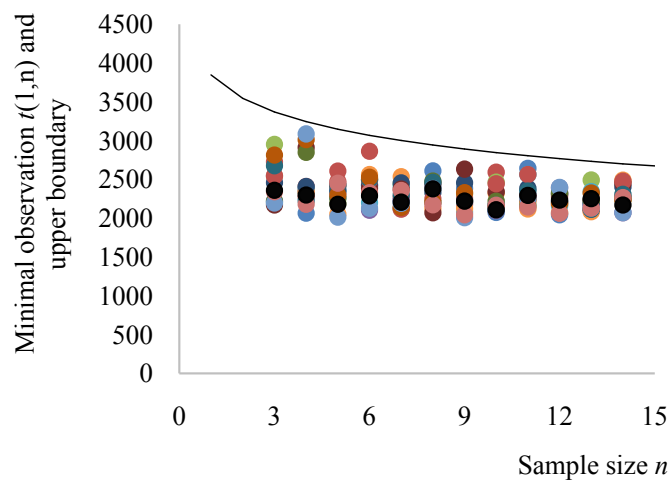
Correspondingly, for a Weibull life probability density function with known
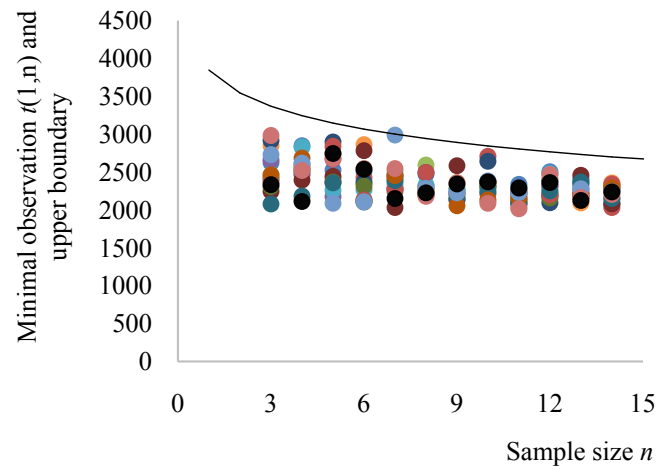
(a)
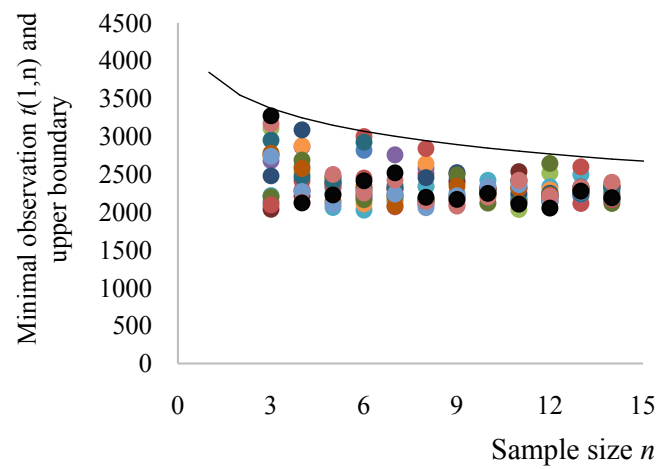
(b)

(c)

Figure 3. Upper envelope curves (boundaries) of the minimal observations from W(2.0, 1000, 2000). (a) The first group of the simulations and upper boundary expressed by Equation (6); (b) The second group of the simulations and upper boundary expressed by Equation (6); (c) The third group of the simulations and upper boundary expressed by Equation (6).

shape parameter and scale parameter, as well as n observations with the minimum $t_{(1,n)}$, the location parameter can be estimated as

$$\gamma = t_{(1,n)} + 2\eta \lg\left(n/70\right)^{1/\beta} \tag{7}$$

Equation (6) indicates that the location parameter revision amount $2\eta \lg\left(n/70\right)^{1/\beta}$ is a function of the scale parameter, and the larger the scale parameter is, the larger the difference between the minimal observation and the location parameter for a given sample size.

The underlying principle is that, for a given shape parameter, the larger the scale parameter, the larger the life dispersion, and the larger the difference between the minimal observation and the location parameter. For a particular Weibull location parameter estimation problem, a temporary scale parameter value larger than the true value can be used for Equation (6) and Equation (7) to get a conservative location parameter estimated.

It is well known that, with known shape parameter and location parameter, the Weibull scale parameter can be estimated as

$$\eta = \left(\sum_{i=1}^{n}\left(t_i - \gamma\right)^{\beta}/n\right)^{1/\beta} \tag{8}$$

where, t_i stands for the ith observation in a sample, n stands for sample size.

For a conservative location parameter estimation, temporarily use a location parameter equaling to $t_{(1,n)}/2$ in Equation (8) when estimates the temporary value of the scale parameter for Equation (6) and Equation (7). Consequently, a scale parameter greater than the true value can be obtained. For instance, if the five observations from a Weibull distribution W(2.0, 1000, 2000) are 3136, 3070, 2276, 2350 and 2191, respectively (generated by Monte Carlo sampling), a scale parameter can be estimated by Equation (8) as 1564 based on a temporary location parameter value of 2191/2. With this scale parameter and the minimal observation 2191, a location parameter is eventually estimated by Equation (7) as 398.

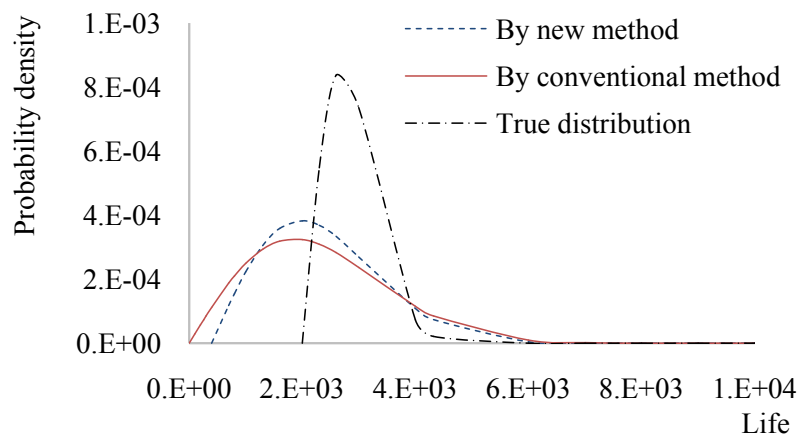Though such an estimated location parameter is much lower than the true value

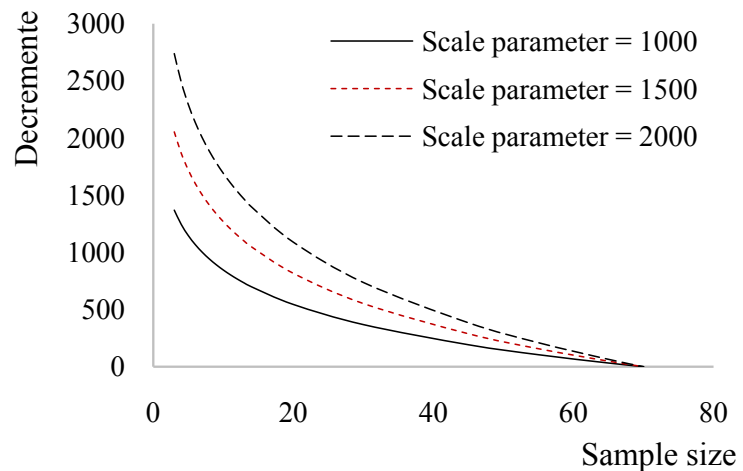**Figure 4.** True distribution and those estimated by different methods.

Figure 5. Revision amount based on different scale parameters.

of 2000, it still can yield a better estimation W(2.0, 2244, 398) for the product life than the conventional two-parameter Weibull distribution W(2.0, 2637, 0), as shown in **Figure 4**.

Furthermore, the effect of the temporarily estimated scale parameter on the revision amount, *i.e.* $t_{(1,n)}^{\max} - \gamma$, can be clearly demonstrated by Equation (6). Shown in **Figure 5** are the situations of scale parameters of 1000, 1500, and 2000, respectively, for the case of the true scale parameter of 1000. It illustrates that larger scale parameter applied in Equation (6) and Equation (7) yields smaller location parameter estimation.

3. Conclusions

A new method is presented to estimate the location parameter of a Weibull probability density function to describe product life distribution in the situation of a very small size of life sample. Simulated big data, generated by Monte Carlo sampling, is applied to identify the relationship between the Weibull location parameter and the minimal observation. A logarithmic model is established to modify the minimal observation into the estimated Weibull location parameter.

The example to estimate the three-parameter Weibull life distribution of the ball bearing according to five life data illustrates that the three-parameter Weibull life probability density function estimated with the new method differs from the two-parameter Weibull distribution conventionally estimated. Although it is far from the true distribution, such a result is still significant for product reliability estimation, contrasting to the traditionally used two-parameter Weibull distribution.

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] McLain, A.C. and Ghosh, S.K. (2011) Nonparametric Estimation of the Conditional Mean Residual Life Function with Censored Data. *Lifetime Data Anal*, **17**, 514-532. https://doi.org/10.1007/s10985-011-9197-x

[2] Soliman, A.A., Abd-Ellah, A.H. and Abou-Elheggag, N.A. (2012) Estimation of the Parameters of Life for Gompertz Distribution Using Progressive First-Failure Censored Data. *Computational Statistics and Data Analysis*, **56**, 2471-248. https://doi.org/10.1016/j.csda.2012.01.025

[3] Zhao, M., Jiang, H. and Liu, X. (2013) A Note on Estimation of the Mean Residual Life Function with Left-Truncated and Right-Censored Data. *Statistics and Probability Letters*, **83**, 2332-2336. https://doi.org/10.1016/j.spl.2013.06.020

[4] Elmahdy, E.E. (2015) A New Approach for Weibull Modeling for Reliability Life Data Analysis. *Applied Mathematics and Computation*, **250**, 708-720. https://doi.org/10.1016/j.amc.2014.10.036

[5] Jia, X., Wang, D. and Jiang, P. (2016) Inference on the Reliability of Weibull Distribution with Multiply Type-I Censored Data. *Reliability Engineering and System Safety*, **150**, 171-181. https://doi.org/10.1016/j.ress.2016.01.025

[6] Ducrosadn, F. and Pamphile, P. (2018) Bayesian Estimation of Weibull Mixture in Heavily Censored Data Setting. *Reliability Engineering and System Safety*, **180**, 453-462. https://doi.org/10.1016/j.ress.2018.08.008

[7] Ahmed, A.O.M. and Bayesian, N.A.I. (2010) Estimator for Weibull Distribution with Censored Data Using Extension of Jeffrey Prior Information. *Procedia Social and Behavioral Sciences*, **8**, 663-669. https://doi.org/10.1016/j.sbspro.2010.12.092

[8] Al Sobhi, M.M. and Soliman, A.A. (2016) Estimation for the Exponentiated Weibull Model with Adaptive Type-II Progressive Censored Schemes. *Applied Mathematical Modelling*, **40**, 1180-1192. https://doi.org/10.1016/j.apm.2015.06.022

[9] Mweleli, R., Orawo, L., Tamba, C. and Okenye, J. (2020) Interval Estimation in a Two Parameter Weibull Distribution Based on Type-2 Censored Data. *Open Journal of Statistics*, **10**, 1039-1056. https://doi.org/10.4236/ojs.2020.106059

[10] Ng, H.K.T., Luo, L., Hu, Y. and Duan, F. (2012) Parameter Estimation of Three-Parameter Weibull Distribution Based on Progressively Type-II Censored Samples. *Journal of Statistical Computation & Simulation*, **82**, 1661-1678. https://doi.org/10.1080/00949655.2011.591797

[11] Abbasi, B., Rabelo, L. and Hosseinkouchack, M. (2008) Estimating Parameters of the Three-Parameter Weibull Distribution Using a Neural Network. *European Journal of Industrial Engineering*, **2**, 428-445. https://doi.org/10.1504/EJIE.2008.018438