A NEW ALGORITHM FOR ESTIMATING THE PARAMETERS OF THE THREE-PARAMETER WEIBULL DISTRIBUTION

Ogbonna, U.A, Obiora-Ilouno, H.O
Department of Statistics, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria.

Abstract
In this work we propose an algorithm for estimating the parameters of the three-parameter Weibull distribution. The proposed algorithm is an extension of the closed form estimator for the shape parameter proposed by Teimouri and Gupta (2013). We compared the proposed algorithm with Teimouri and Gupta's estimator and found out that the proposed method performs better when $\beta > 1,$ however when $\beta \leq 1$ the Teimouri and Gupta estimator is prefered. We also discovered that the performance of the methods always increases with sample size. The proposed method also provides a prior knowledge of a likely range of the true shape parameter value.
Keywords: Distribution, Estimators, coefficient of Variation, Algorithm, Parameter

1. Introduction

The use of Weibull distribution to describe real phenomena has a long history. This distribution was originally proposed by the Swedish physicist Waloddi Weibull in (1951). Up to the end of the 1950s, lifetime in engineering was nearly always modeled by the exponential distribution which later was gradually substituted by the more flexible Weibull distribution. For more than half a century, the Weibull distribution has attracted the attention of statisticians working on theory and methods as well as various fields of applied statistics. Together with the Normal, Exponential and F distributions, the Weibull distribution is without any doubt the most applied distribution in modern statistics (Rinne, 2008). The Weibull distribution is of great interest to theory-oriented statisticians because of its great number of features. It is also of great interest to practitioners because of its ability to fit to data from various fields, ranging from life data to weather data or observations made in economics and business administration.

Lloyd (1967) as well as so many others have expanded the scope and usefulness of the Weibull distribution to other branches of statistics such as quality control.

Originally, the Weibull distribution has three parameters; the shape parameter (β), the scale parameter (θ) and the location parameter (γ).

The three-parameter Weibull distribution has the following probability density function (Pdf);

$$f(x) = \frac{\beta}{\theta}\left(\frac{x-\gamma}{\theta}\right)^{\beta-1} e^{-\left(\frac{x-\gamma}{\theta}\right)^{\beta}} x > \gamma,\ \beta > 0, \theta > 0 \quad (1)$$

with distribution function given as $F(x) = 1 - e^{-\left(\frac{x-\gamma}{\theta}\right)^{\beta}}$ (2)

which is fitted when all the three parameters are unknown but are assumed to be non-zero, (Cohen and Whitten, 1982a & b).

For $\beta < 2.6$ the Weibull distribution is positively skewed (has a right tail), for $2.6 < \beta < 3.7,$ its coefficient of skewness approaches zero (no tail). Consequently, it may approximate the normal, and for $\beta > 3.7,$ it is negatively skewed (left tail). In this paper, we will be dealing with beta values in the interval 0 <$\beta < 2.6.$

2.0 MATERIAL AND METHOD

2.1Parameter Estimation

In general, one of the best methods for estimating the parameters of a distribution is the maximum likelihood method (MLE). But its application to the three-parameter Weibull distribution has been a problem for the following reasons;

1. the Weibull distribution does not satisfy the regularity condition that is required for the maximum likelihood estimate to be the most efficient method.
2. the MLE solutions are biased and the amount of bias is not known. This bias strongly depends on the shape parameter and the sample size.
3. the MLE solutions are not available in closed form for two of the parameters of a three parameter Weibull distribution.

2.2 Teimouri and Gupta Method

The third reasons above was addressed by Teimouri and Gupta (2013) in an attempt to solve this problem, they used theorem 1.0 below to construct a closed form estimator for the slope parameter.

Theorem 1.0: suppose $x_1, x_2, \dots, x_n$ is a random sample from a weibull distribution. Let ρ denote the sample correlation coefficient between x_i and their ranks. Let C and S denote the sample coefficient of variation and the sample standard deviation respectively. Then,

$$\rho = \left(\frac{\mu_x - \gamma}{\sigma_x}\right)\left(\frac{1}{2} - \frac{1}{2^{1+\frac{1}{\beta}}}\right)\sqrt{\frac{12(n-1)}{n+1}} \tag{3}$$

Where $\mu_x = E(x)$ and $\sigma^2{}_x = Var(x)$ (see Teimouri and Gupta 2013 for proof)

Corollary 1.0: suppose $x_1, x_2, \dots, x_n$ is a random sample from a weibull distribution with known location parameter. Let ρ denote the sample correlation coefficient between x_i and their ranks. Let C and S denote respectively the sample coefficient of variation and sample standard deviation. Then the estimator of the shape parameter is;

$$\hat{\beta} = \frac{-\ln 2}{\ln\left[1 - \frac{\rho}{\sqrt{3}}\left(\frac{1}{C} - \frac{\gamma}{S}\right)^{-1}\sqrt{\frac{n+1}{n-1}}\right]} \tag{4}$$

Teimouri and Gupta (2013) proposed $\hat{\gamma} = x_{(1)} - \frac{1}{n}$ as an estimator for γ

Therefore,

$$\hat{\beta} = \frac{-\ln 2}{\ln\left[1 - \frac{\rho}{\sqrt{3}}\left(\frac{1}{C} - \frac{x_{(1)} - \frac{1}{n}}{S}\right)^{-1}\sqrt{\frac{n+1}{n-1}}\right]} \tag{5}$$

They also stated that the estimator for the scale parameter can be obtained by maximizing the likelihood function of the three-parameter Weibull distribution with respect to the scale parameter and making the scale parameter subject of the formula gives;

$$\hat{\theta} = \left(\frac{\sum(x_i - \hat{\gamma})^{\hat{\beta}}}{n}\right)^{1/\hat{\beta}} \tag{6}$$

2.3 PROPOSED ALGORITHM *BASED ON THE SAMPLE COEFFICIENT OF VARIATION*

Step 1

Find an initial estimate for the shape parameter based on the sample coefficient of variation

Recall that $CV = \frac{\sigma_x}{\mu_x}$

Substituting in equation 4 and solves gives;

$$\rho = \left(\frac{1}{CV} - \frac{\gamma}{\sigma_x}\right)\left(\frac{1}{2} - \frac{1}{2^{1+\frac{1}{\beta}}}\right)\sqrt{\frac{12(n-1)}{n+1}} \quad (7)$$

Now making CV the subject of the formula we have;

$$CV = \frac{\sigma_x\left(\frac{1}{2} - \frac{1}{2^{1+\frac{1}{\beta}}}\right)\sqrt{\frac{12(n-1)}{n+1}}}{\sigma_x\rho + \gamma\left(\frac{1}{2} - \frac{1}{2^{1+\frac{1}{\beta}}}\right)\sqrt{\frac{12(n-1)}{n+1}}} \quad (8)$$

Using equation 8 we performed multiple simulation experiments and discovered that there is a strong relationship between coefficient of variation and the shape parameter of the distribution. Precisely, we observed that for $0.2 < \beta < 1.4$ the coefficient of variation falls within the interval $0.52 < CV < 3.0$ ($95\%\ confidence$)and for $1.5 < \beta < 2.8$ the CV falls within the interval $0.2 < CV < 0.51$ ($95\%\ confidence$).

Therefore, to find an initial estimate of β $(\widehat{\beta_{in}})$, we first calculate the coefficient of variation (CV), then the mean of the interval of β that corresponds to the sample CV will be an initial estimate of β.

The mean of a uniform interval is given by $\frac{a+b}{2}$.

So, for $0.2 < CV < 0.51$, $\widehat{\beta_{in}} = \frac{1.5+2.8}{2} = 2.15$ (9)

for $0.52 < CV < 3.0$, $\widehat{\beta_{in}} = \frac{0.2+1.4}{2} = 0.9$ (10)

Step 2

Estimate the Scale Parameter

From the simulation experiments, we also discovered that the sample mean of data that follows the weibull distribution is very close to the scale parameter value 99% of the time especially when $1.5 < \beta < 2.8$. Therefore, we propose that the sample mean be used as an estimator of the scale parameter;

$$\hat{\theta} = \bar{x} \quad (11)$$

Step 3

Estimate the location parameter

Recall that the mean equation of the three-parameter Weibull distribution is given by;

$$\bar{X} = \gamma + \theta\Gamma\left(\frac{1}{\beta} + 1\right) \tag{12}$$

Equating this to the sample mean in equation (11) gives;

$$\bar{x} = \gamma + \theta\Gamma\left(\frac{1}{\beta} + 1\right) \tag{13}$$

Making γ the subject of the formula and replacing θ with the estimate obtained in step 2 we have;

$$\gamma = \bar{x} - \left[\Gamma\left(\frac{1}{\beta_{in}} + 1\right)\right]\bar{x} \tag{14}$$

where β_{in} is an initial estimate of the shape parameter obtained from step 1.

Step 4
Remove the effect of the location parameter by subtracting the estimate gotten from step 1 from each element of the sample.

Step 5
Obtain the final estimate of the shape parameter
From equation 6, we have;

$$\hat{\beta} = \frac{-\ln 2}{\ln\left[1 - \frac{\rho}{\sqrt{3}}\left(\frac{1}{C} - \frac{x_{(1)} - \frac{1}{n}}{S}\right)^{-1}\sqrt{\frac{n+1}{n-1}}\right]} \tag{15}$$

Removing the effect of the location parameter by letting $x_{(1)} - \frac{1}{n} = 0$, the equation becomes;

$$\hat{\beta} = \frac{-l\,n\,2}{l\,n\left[1 - \frac{\rho(CV)}{\sqrt{3}}\sqrt{\frac{n+1}{n-1}}\right]} \tag{16}$$

The estimate for the shape parameter is obtained using equation 16.

2.4 Method of Comparison
We compare the proposed algorithm to the method proposed by Teimouri and Gupta (2013) using simulation. We generated samples from the Weibull distribution with different shape parameter values. We chose 0.5, 1, 1.5 and 2.5 to cover the interval $0 < \beta < 2.6$ and we put the scale and location parameters at 100 and 10 respectively. The scale and location parameters need not to be varied because they are just scaling parameters. We also varied the sample size as 10, 20 and 50 to represent small, medium and large samples sizes. 1000 samples were generated for each sample category and the methods were applied on each sample. All simulation experiments and computation were done using the R software.

The root means square error (RMSE) was used as a measure of accuracy. RMSE is given as;

$$RMSE = \sqrt{Var(x) + bias^2}. \qquad (17)$$

Also, to assess the performance of the methods in predicting the three parameters for each category we used the Euclidean norm of the vector containing the RMSE of the shape, scale and location parameters. The Euclidean norm is given as;

$\|X\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$ where x_is are the elements of the vector

3. Results

Table 3.1: Categories of data simulation

True β	True θ	True γ	Sample size n
0.5	100	10	10
0.5	100	10	20
0.5	100	10	50
1	100	10	10
1	100	10	20
1	100	10	50
1.5	100	10	10
1.5	100	10	20
1.5	100	10	50
2.5	100	10	10
2.5	100	10	20
2.5	100	10	50

Table 3.2: Results for n=10 and beta= 0.5

		BIAS	RMSE	EUCLIDEAN NORM
Proposed Method	β	-0.03445494	0.1971911	186.801
	θ	8.926951	185.5696	
	γ	14.82412	22.01566	
Teimouri And Gupta Method	β	0.1327512	0.1626008	72.095
	θ	18.42228	71.90500	
	γ	-2.163484	5.309800	

Table 3.3: Results for n=20 and beta= 0.5

		BIAS	RMSE	EUCLIDEAN NORM
Proposed Method	β	-0.0616117	0.1457542	150.083
	θ	8.926951	149.4446	
	γ	13.98512	18.04774	
Teimouri And Gupta Method	β	-0.00234304	0.1188928	52.015
	θ	-11.15802	52.17167	
	γ	0.4591412	1.247136	

Table 3.4: Results for n=50 and beta= 0.5

		BIAS	RMSE	EUCLIDEAN NORM
Proposed Method	β	0.120495	0.1492751	126.751
	θ	8.926951	125.1546	
	γ	-20.83912	21.10285	
Teimouri And Gupta Method	β	4.37837e-05	0.07617271	31.000
	θ	-5.792638	31.03455	
	γ	0.06195151	0.1941643	

Table 3.5: Results for n=10 and beta= 1

		BIAS	RMSE	EUCLIDEAN NORM
Proposed Method	β	0.2167867	0.4133144	36.276
	θ	8.926951	32.75525	
	γ	-15.68215	15.76821	
Teimouri And Gupta Method	β	-0.1689997	0.2847648	42.522
	θ	-23.16872	38.70501	
	γ	10.29491	17.60728	

Table 3.6: Results for n=20 and beta= 1

		BIAS	RMSE	EUCLIDEAN NORM
Proposed Method	β	0.2125245	0.3094314	29.322
	θ	8.926951	24.7096	
	γ	-15.75905	15.80246	
Teimouri And Gupta Method	β	-0.08500607	0.1989304	26.527
	θ	-11.91453	25.06324	
	γ	4.938966	8.725661	

Table 3.7: Results for n=50 and beta= 1

		BIAS	RMSE	EUCLIDEAN NORM
Proposed Method	β	0.2251297	0.2648941	23.326
	θ	8.926951	17.2042	
	γ	-15.74039	15.75732	
Teimouri And Gupta Method	β	-0.0327187	0.4812885	15.803
	θ	-4.030083	15.42769	
	γ	1.97096	3.39117	

Table 3.8: Results for n=10 and beta= 1.5

		BIAS	RMSE	EUCLIDEAN NORM
Proposed Method	β	-0.03818768	0.4261418	20.602
	θ	8.926951	20.42073	
	γ	1.377303	2.700893	
Teimouri And Gupta Method	β	-0.4584609	0.5364848	44.585
	θ	-31.26084	37.86138	
	γ	19.34675	23.54054	

Table 3.9: Results for n=20 and beta= 1.5

		BIAS	RMSE	EUCLIDEAN NORM
Proposed Method	β	-0.05372831	0.2906152	14.004
	θ	8.926951	13.84383	
	γ	1.40345	2.109452	
Teimouri And Gupta Method	β	-0.2805753	0.3622391	27.613
	θ	-17.20969	23.52643	
	γ	12.10774	14.44512	

Table 3.10 Results for n=50 and beta= 1.5

		BIAS	RMSE	EUCLIDEAN NORM
Proposed Method	β	-0.03608594	0.1684401	9.082
	θ	8.926951	8.927062	
	γ	1.438244	1.759658	
Teimouri And Gupta Method	β	-0.1574609	0.2237386	16.478
	θ	-9.816728	14.2841	
	γ	6.803893	8.214172	

Table 3.11 Results for n=10 and beta= 2.5

		BIAS	RMSE	EUCLIDEAN NORM
Proposed Method	β	-0.04431032	0.7789568	12.157
	θ	8.926951	11.99993	
	γ	1.259622	1.853849	
Teimouri And Gupta Method	β	-1.231671	1.283226	56.797
	θ	-43.98572	41.46263	
	γ	35.7258	38.80891	

Table 3.12 Results for n=20 and beta= 2.5

		BIAS	RMSE	EUCLIDEAN NORM
Proposed Method	β	-0.05643378	0.515282	8.905
	θ	8.926951	8.745896	
	γ	1.264323	1.604951	
Teimouri And Gupta Method	β	-0.9306541	0.9873057	46.119
	θ	-32.20311	35.10124	
	γ	27.42744	29.91266	

Table 3.13 Results for n=50 and beta= 2.5

		BIAS	RMSE	EUCLIDEAN NORM
Proposed Method	β	-0.04752342	0.2900169	5.685
	θ	8.926951	5.518459	
	γ	1.228817	1.371911	
Teimouri And Gupta Method	β	-0.6004832	0.655485	23.272
	θ	-21.00493	23.08954	
	γ	1.970963	2.759584	

4. Discussion and Summary

From the results of the experiments, it is evident that when $\beta \geq 1$, the accuracy of the proposed method is relatively high (especially when $\beta > 1$). However, when $\beta < 1$, the proposed method estimator for the scale parameter ($\hat{\theta} = \bar{x}$) does not perform too well. This was actually expected because, when $\beta < 1$, the Weibull distribution is very positively skewed and the mean of a skewed distribution favors extreme values. This is why the estimates produced at $\hat{\theta} = \bar{x}$ deviates from the true scale. When $\beta \leq 1$, the Mahdi and Gupta method performed well but it loses its accuracy as β increases.

5. Conclusion

From the above findings, the proposed method is the best method for estimating the parameters of the three parameter Weibull distribution when $\beta > 1$. This means that as the distribution gradually becomes symmetric and other methods begin to lose accuracy, the proposed method produces the most accurate estimates and should be preferred. This also means that when modeling the wear out period of a device, the proposed method should be adopted. However, the Teimouri and Gupta method should be adopted for estimating the parameters of the three-parameter Weibull distribution when $\beta \leq 1$. This means that when the Weibull distribution is very positively skewed. Therefore, for modelling the infant mortality period of a device, the Teimouri and Gupta method of parameter estimation should be the choice. Sample size does not affect the choice of method, though the performances of the methods always improve as sample size increases.

References

[1] Cohen, A. C. and Whitten, B. (1982a) Modified Maximum Likelihood For The Three-Parameter Weibull Distribution, Communications in Statistics Theory and Methods 11, 197-214.

[2] Cohen, A. C. and Whitten, B. (1982b) Modified Moment Estimators For The Three-Parameter Weibull Distribution, Communications in Statistics Theory and Methods 11, 2631-2656.

[3] Lloyd, N. S. (1967) Weibull Probability Paper, Industrial Quality Control 23, 452–453.

[4] Rinne, H. (2008) The Weibull Distribution: A Hand Book, CRC press, London.

[5] Teimouri, M. and Gupta, A.K. (2013) On the three-parameter weibull distribution shape parameter estimation, journal of data science 11, 403-414.

[6] Weibull, W. (1951) A Statistical Distribution Function of Wide Applicability, Journal of Applied Mechanics, Transactions of The American Society of Mechanical Engineers, 51, 293-279.