Research Article
Methods for Estimating the 2-Parameter Weibull Distribution with Type-I Censored Data

Chris Bambey Guure and Noor Akma Ibrahim

Institute for Mathematical Research, Universiti Putra Malaysia, 43400, Selangor, Malaysia

Abstract: This study is concerned with the two-parameter Weibull distribution which has and is still being used as a model in life testing and reliability engineering. We seek to find out whether Rank Regression Method can be a good alternative to that of the world publicised traditional method known as Maximum Likelihood for estimating two parameters of the Weibull distribution. The methods under consideration are: Maximum Likelihood Estimation, Least Square Estimation on Y and that of Least Square Estimation on X. These estimators are derived for Random Type-I censored samples. These methods were compared using Mean Square Error and Mean Percentage Error through simulation study with small, medium and large sample sizes in estimating the Weibull parameters under Type-I censored data. The observations that are made based on this study are that Maximum Likelihood Estimator stands out when estimating the scale parameter followed by Least Square Estimator on X but for the shape parameter Least Square Estimator on X performed better than Maximum Likelihood Estimator thereby making it a good alternative method to MLE.

Keywords: Ls estimation on x, ls estimation on y, maximum likelihood estimation, random type-i censoring, simulation study, weibull distribution

INTRODUCTION

As a result of the versatility in fitting time-to-failure of a very extensive variety to complex mechanisms, the Weibull distribution has lately assumed the centre stage especially in the field of life-testing and reliability analysis. They have not been many studies that look at the potentials of the two Rank Regression methods against that of the Maximum Likelihood method, although these methods have very good characteristics or properties that could be used as an alternative to Maximum Likelihood Estimator, hence the review of these methods in this study.

Censoring is a feature that is recurrent in lifetime and reliability data analysis, it occurs when exact lifetimes or run-outs can only be collected for a portion of the inspection units. According to Horst (2009), "A data sample is said to be censored when, either by accident or design, the value of the variables under investigation is unobserved for some of the items in the sample". If lifetimes are only known to exceed some given time or assumed to have the potential of exceeding but for certain reasons, may be due to removal or withdrawal then it is referred to as right censoring.

There are two types of right censoring and these are; Type-I and Type-II censoring. Type-I censoring can be classified into two:

- Fixed Type-I censoring
- Random Type-I censoring

The main focus here is on Random Type-I censoring. This is where a study is designed to end after a specified given time T and the censored units or subjects do not all have the same censoring time unlike Fixed Type-I censoring where at the end of the study every unit that did not have an event observed during the course of the study is censored at time T.

The aim of this study is three fold. First, consideration is given to the Maximum Likelihood Estimators (MLE) of the unknown parameters when the data are Random Type-I censored and it has been observed that the MLE cannot be obtained in closed form, therefore Newton-Raphson method has been proposed to solve the non-linear nature of the equations. For prove on how to obtain Maximum Likelihood Estimation of Type-I censoring see, Panani and Saeid (2011), Stefano *et al.* (2007) and Saunders (2007). Secondly, we have considered the estimation of the parameters under Least Square estimators on X (LSX), which is used to minimize the horizontal distance between the data points and the straight line that is fitted to the data and lastly, the Least Square estimators on Y (LSY) which is also used to minimize not the horizontal but the vertical distance between the

---

Corresponding Author: Chris Bambey Guure, Institute for Mathematical Research, Universiti Putra Malaysia, 43400, Selangor Malaysia

data points and the straight line that is fitted to the data. One of the good advantages for using these methods in life testing and reliability analysis is that most of the distributions can be linearized therefore making the calculations very easy and straightforward and have close-form solutions which can readily give an answer without having to employ numerical techniques. A study has been done comparing these two methods in estimating the parameters of the Weibull distribution by Al-Kanani and Shaima (2011) and Zhang *et al.* (2007, 2006), similar work by Berger and Lawrence (1974).

The main objective of this study is to determine whether Rank Regression Method can be a good alternative to that of the world publicised traditional method known as Maximum Likelihood for estimating two parameters of the Weibull distribution. The methods under consideration are: Maximum Likelihood Estimation, Least Square Estimation on Y and that of Least Square Estimation on X.

The rest of the study is arranged as follows. Firstly, the two unknown parameters of the Weibull distribution are derived under Maximum Likelihood Estimation, followed by Least Square Estimation on X and Least Square Estimation on Y. The methods are then compared using Simulation and the results presented and finally Conclusions based on the study are given.

MATERIALS AND METHODS

Maximum likelihood estimation: The Weibull Probability Density Function is given as:

$$f(t_i, \alpha, \beta) = \frac{\beta}{\alpha}\left(\frac{t_i}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{t_i}{\alpha}\right)\right]^{\beta} \qquad (1)$$

The Cumulative Distribution Function is:

$$F(t_i, \alpha, \beta) = 1 - \exp\left[-\left(\frac{t_i}{\alpha}\right)\right]^{\beta} \qquad (2)$$

The survival function is also given as:

$$S(t_i, \alpha, \beta) = 1 - \left\{1 - \exp\left[-\left(\frac{t_i}{\alpha}\right)\right]^{\beta}\right\} \qquad (3)$$

The scale and shape parameters are α and β, respectively.

Suppose that $t_i, < \ldots\ldots < t_r$ is known to have failed during the study and the remaining $t_n - t_r = t_q$ censored but the censored units do not all have the same censoring time, then the likelihood function of the 2-parameter Weibull distribution as stated by Saunders (2007) is:

$$L(t_i, \alpha, \beta) = \prod_{i=1}^{r} f(t_i) \prod_{i=r+1}^{n} \bar{F}(t) = \prod_{i=1}^{r} h(t_i) \prod_{i=1}^{r} \bar{F}(t_i) * \bar{F}(t_q)^{n-r} \qquad (4)$$

$$L(t_i, \alpha, \beta) = \prod_{i=1}^{r} \left[\frac{\beta}{\alpha}\left(\frac{t_i}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{t_i}{\alpha}\right)\right]^{\beta}\right]\left[1 - \left(1 - \exp\left[-\left(\frac{t_i}{\alpha}\right)\right]^{\beta}\right)\right]^{n-r} \qquad (5)$$

The log-likelihood is:

$$\ln L = r \ln(\beta) - r\beta \ln(\alpha) + (\beta - 1)\sum_{i=1}^{r} \ln(t_i) - \sum_{i=1}^{r}\left(\frac{t_i}{\alpha}\right)^{\beta} - \left[n - r\left(\frac{t_q}{\alpha}\right)^{\beta}\right] \qquad (6)$$

Differentiating (6) with respect to α and β and equating to zero, we have:

$$\alpha = \left[\frac{1}{r}\left(\sum_{i=1}^{r}(t_i)^{\beta} + (n-r)(t_q)^{\beta}\right)\right]^{\frac{1}{\beta}} \qquad (7)$$

and

$$\frac{r}{\beta} + \sum_{i=1}^{r}\left(\frac{t_i}{\alpha}\right)\ln - \sum_{i=1}^{r}\left(\frac{t_i}{\alpha}\right)^{\beta}\ln\left(\frac{t_i}{\alpha}\right) - (n-r)\left(\frac{t_q}{\alpha}\right)^{\beta}\ln\left(\frac{t_q}{\alpha}\right) = 0 \qquad (8)$$

Substituting (7) into (8) we have:

$$0 = \frac{r}{\beta} - \sum_{i=1}^{r}\left[\frac{(t_i)^{\beta}}{\left[\frac{1}{r}\left(\sum_{i=1}^{r}(t_i)^{\beta} + (n-r)(t_q)^{\beta}\right)\right]}\right]\ln\left[\frac{(t_i)}{\left[\frac{1}{r}\left(\sum_{i=1}^{r}(t_i)^{\beta} + (n-r)(t_q)^{\beta}\right)\right]^{\frac{1}{\beta}}}\right] - \left[\frac{(n-r)(t_q)^{\beta}}{\left[\frac{1}{r}\left(\sum_{i=1}^{r}(t_i)^{\beta} + (n-r)(t_q)^{\beta}\right)\right]}\right]\ln$$

$$\left[\frac{(t_q)}{\left[\frac{1}{r}\left(\sum_{i=1}^{r}(t_i)^{\beta} + (n-r)(t_q)^{\beta}\right)\right]^{\frac{1}{\beta}}}\right] + \sum_{i=1}^{r}\ln\left[\frac{(t_i)}{\left[\frac{1}{r}\left(\sum_{i=1}^{r}(t_i)^{\beta} + (n-r)(t_q)^{\beta}\right)\right]^{\frac{1}{\beta}}}\right] \tag{9}$$

In order to solve (9), Newton-Raphson iterative procedure was employed as follows:
Let f(β) be the same as (8) or (9) and taking the first differential of f (β), we have:

$$f'(\beta) = -\frac{r}{\beta^2} - \sum_{i=1}^{r}\left[\frac{(t_i)^{\beta}}{\left[\frac{1}{r}\left(\sum_{i=1}^{r}(t_i)^{\beta} + (n-r)(t_q)^{\beta}\right)\right]}\right]\ln^2\left[\frac{(t_i)}{\left[\frac{1}{r}\left(\sum_{i=1}^{r}(t_i)^{\beta} + (n-r)(t_q)^{\beta}\right)\right]^{\frac{1}{\beta}}}\right] -$$

$$\left[\frac{(n-r)(t_q)^{\beta}}{\left[\frac{1}{r}\left(\sum_{i=1}^{r}(t_i)^{\beta} + (n-r)(t_q)^{\beta}\right)\right]}\right]\ln^2\left[\frac{(t_q)}{\left[\frac{1}{r}\left(\sum_{i=1}^{r}(t_i)^{\beta} + (n-r)(t_q)^{\beta}\right)\right]^{\frac{1}{\beta}}}\right] \tag{10}$$

β is estimated by assuming an initial value and solving (11) below repeatedly till it converges, after which α can be determined:

$$\beta_{i+1} = \beta_i - \left\{\frac{\frac{r}{\beta} - s_{i1} - s_{i2} + s_{i3}}{-\frac{r}{\beta^2} - s_{i4} - s_{i5}}\right\} \tag{11}$$

where,

$$s_{i1} = \sum_{i=1}^{r}\left[\frac{(t_i)^{\beta}}{\left[\frac{1}{r}\left(\sum_{i=1}^{r}(t_i)^{\beta} + (n-r)(t_q)^{\beta}\right)\right]}\right]\ln\left[\frac{(t_i)}{\left[\frac{1}{r}\left(\sum_{i=1}^{r}(t_i)^{\beta} + (n-r)(t_q)^{\beta}\right)\right]^{\frac{1}{\beta}}}\right]$$

$$s_{i2} = \left[\frac{(n-r)(t_q)^{\beta}}{\left[\frac{1}{r}\left(\sum_{i=1}^{r}(t_i)^{\beta} + (n-r)(t_q)^{\beta}\right)\right]}\right]\ln\left[\frac{(t_q)}{\left[\frac{1}{r}\left(\sum_{i=1}^{r}(t_i)^{\beta} + (n-r)(t_q)^{\beta}\right)\right]^{\frac{1}{\beta}}}\right]$$

$$s_{i3} = \sum_{i=1}^{r} \ln \left[\frac{(t_i)}{\left[\frac{1}{r} \left(\sum_{i=1}^{r} (t_i)^\beta + (n-r)(t_q)^\beta \right) \right]^{\frac{1}{\beta}}} \right]$$

$$s_{i4} = \sum_{i=1}^{r} \left[\frac{(t_i)^\beta}{\left[\frac{1}{r} \left(\sum_{i=1}^{r} (t_i)^\beta + (n-r)(t_q)^\beta \right) \right]} \right] \ln^2 \left[\frac{(t_i)}{\left[\frac{1}{r} \left(\sum_{i=1}^{r} (t_i)^\beta + (n-r)(t_q)^\beta \right) \right]^{\frac{1}{\beta}}} \right]$$

$$s_{i5} = \left[\frac{(n-r)(t_q)^\beta}{\left[\frac{1}{r} \left(\sum_{i=1}^{r} (t_i)^\beta + (n-r)(t_q)^\beta \right) \right]} \right] \ln^2 \left[\frac{(t_q)}{\left[\frac{1}{r} \left(\sum_{i=1}^{r} (t_i)^\beta + (n-r)(t_q)^\beta \right) \right]^{\frac{1}{\beta}}} \right]$$

Least square estimation on Y: Taking logarithm twice on the (2) we have:

$$\ln[-\ln(1-F(t_i))] = \beta \ln(t_i) - \beta \ln(\alpha), \qquad (12)$$

Eq. (12) can be represented by:

$$y_i = \ln[-\ln(1-F(t_i))] \text{ and } x_i = \ln(t_i) \qquad (13)$$

where,

$$y_i = \beta x_i - \beta \ln(\alpha), \qquad (14)$$

if $A = \sum_{i=1}^{r} (y_i - Y)^2$

then

$$A = \sum_{i=1}^{r} (y_i - [\beta x - \beta \ln(\alpha)])^2 \qquad (15)$$

Differentiating (15) with respect to α and β and equating the partial derivatives to zero (0), the estimating equations of LSY as obtained by Zhang *et al.* (2007) is:

$$\hat{\beta}_{xy} = \frac{\sum_{i=1}^{r} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{r} (x_i - \overline{x})^2} \qquad (16)$$

and

$$\hat{\alpha}_{xy} = \exp \left[-\left(\frac{\overline{y}}{\hat{\beta}_{xy}} - \overline{x} \right) \right] \qquad (17)$$

where,

$$\overline{x} = \sum_{i=1}^{r} \frac{x_i}{r} \text{ and } \overline{y} = \sum_{i=1}^{r} \frac{y_i}{r}$$

Least square estimation on X:
This can be obtained in a similar way as above but:

$$A = \sum_{i=1}^{r} [x_i - (1/\beta y_i + \ln(\alpha))]^2$$

where,

$$\hat{\beta}_{yx} = \frac{\sum_{i=1}^{r} (y_i - \overline{y})^2}{\sum_{i=1}^{r} (x_i - \overline{x})(y_i - \overline{y})} \qquad (18)$$

$$\hat{\alpha}_{yx} = \left[-\frac{(\overline{x} - \overline{y})}{\hat{\beta}_{yx}} \right] \qquad (19)$$

with $\overline{x} = \sum_{i=1}^{r} \frac{x_i}{r}$ and $\overline{y} = \sum_{i=1}^{r} \frac{y_i}{r}$

This is applicable to both failure and censored data.

$\hat{\alpha}$ and $\hat{\beta}$ are obtained from LSY and LSX by substituting (13) into (16), (17), (18) and (19).

If failure then n = r

Bernard's median rank estimator is:

$$F(t_i) = \frac{i - 0.3}{n + 0.4} \tag{20}$$

With n = number of sample used and also i = 1, 2, …, n is the ordered number of items.

It is imperative to state here that because the censoring was random, Mean Order Number was used to determine the ranks for the failure data.

Simulation study: In trying to illustrate and compare the methods as described above, a random sample of size, n = 25, 50 and 100 with 10 and 20% random censoring were generated from the Weibull distribution to take care of small, medium and large data sets. The scale parameter was chosen to be 0.8 and 1.5 and the shape parameter 0.5 and 1.2. These were replicated 1000 times and the parameters were estimated using the methods above. The comparisons were based on values from Mean Square Error (MSE) and Mean Percentage Error (MPE).

where,

$$MSE(\hat{\theta}) = (1000)^{-1} \sum_{r=1}^{1000} (\hat{\theta}^r - \theta)^2$$

and

$$MPE(\hat{\theta}) = (1000)^{-1} \sum_{r=1}^{1000} \frac{|\hat{\theta}^r - \theta|}{\theta}$$

In the tables below, the MSE and MPE of the estimators in the parentheses are from the 10% censoring whiles those not in parentheses are from the 20% censoring.

CONCLUSION

In this study consideration is given to three methods in estimating the parameters of the Weibull distribution under Random Type-I censoring. The following conclusions are observed.

Maximum Likelihood Estimator with reference to Table 1 and 2 is dominant in estimating the scale parameter in both the 10 and 20% censoring. We also observe that when the scale and the shape parameters are 0.8 and 0.5, respectively with an increase in sample size in both Table 1 and 2, the Least Square Estimator on X has a small MSE and MPE, an indicative of a good estimator. In terms of percentage, MLE had about 91.7% over the other estimators with the 20% censored data but 75% with the 10% censored. It is important to state that the MSE and MPE of the MLE are not that flamboyant over Least Square Estimator on X so as to coin it as being disadvantageous in estimating the scale parameter.

Table 1: MSE estimated values of α and β

n	α	β	MLE $\hat{\alpha}$	$\hat{\beta}$	LSY $\hat{\alpha}$	$\hat{\beta}$	LSX $\hat{\alpha}$	$\hat{\beta}$
25	0.8	0.5	0.07860	0.02270	0.14600	0.00050	0.10700	0.00210
			(0.0129)	(0.0144)	(0.0297)	(0.0014)	(0.0133)	(0.0008)
50	0.8	0.5	0.08530	0.00730	0.12900	0.00040	0.08590	0.00090
			(0.0126)	(0.0042)	(0.0227)	(0.0002)	(0.0129)	(0.0010)
100	0.8	0.5	0.08160	0.00520	0.10600	0.00090	0.08070	0.00190
			(0.0167)	(0.0026)	(0.0193)	(0.0003)	(0.0150)	(0.0015)
25	0.8	1.2	0.00760	0.08520	0.01650	0.01790	0.01050	0.00010
			(0.0014)	(0.0561)	(0.0041)	(0.0317)	(0.0014)	(0.0009)
50	0.8	1.2	0.00660	0.03010	0.01150	0.00770	0.00770	0.00090
			(0.0009)	(0.0135)	(0.0029)	(0.0102)	(0.0015)	(0.0004)
100	0.8	1.2	0.00590	0.02470	0.00990	0.00150	0.00770	0.00090
			(0.0014)	(0.0069)	(0.0025)	(0.0029)	(0.0015)	(0.0005)
25	1.5	0.5	0.31700	0.01910	0.66400	0.00280	0.39920	0.00030
			(0.0428)	(0.0106)	(0.1170)	(0.0027)	(0.0538)	(0.0004)
50	1.5	0.5	0.32000	0.00620	0.47100	0.00070	0.34930	0.00030
			(0.0454)	(0.0029)	(0.0739)	(0.0006)	(0.0399)	(0.0005)
100	1.5	0.5	0.31200	0.00570	0.38710	0.00030	0.31340	0.00170
			(0.0525)	(0.0019)	(0.0737)	(0.0001)	(0.0533)	(0.0009)
25	1.5	1.2	0.02670	0.08520	0.05800	0.01810	0.03680	0.00040
			(0.0048)	(0.0561)	(0.0107)	(0.0118)	(0.0043)	(0.0033)
50	1.5	1.2	0.02310	0.03010	0.04030	0.00770	0.02710	0.00020
			(0.0055)	(0.0216)	(0.0134)	(0.0067)	(0.0072)	(0.0001)
100	1.5	1.2	0.02060	0.02470	0.03480	0.00150	0.02720	0.00090
			(0.0046)	(0.0092)	(0.0077)	(0.0023)	(0.0051)	(0.0007)

Table 2: MPE estimated values of α and β

n	α	β	MLE $\hat{\alpha}$	MLE $\hat{\beta}$	LSY $\hat{\alpha}$	LSY $\hat{\beta}$	LSX $\hat{\alpha}$	LSX $\hat{\beta}$
25	0.8	0.5	0.35130	0.04760	0.47750	0.01360	0.40940	0.02920
			(0.1421)	(0.0760)	(0.2162)	(0.0234)	(0.1442)	(0.0140)
50	0.8	0.5	0.36540	0.05420	0.45030	0.01260	0.36780	0.01860
			(0.1413)	(0.0408)	(0.1887)	(0.0044)	(0.1426)	(0.0202)
100	0.8	0.5	0.35720	0.04560	0.40740	0.01920	0.35560	0.02780
			(0.1620)	(0.0324)	(0.1747)	(0.0096)	(0.1531)	(0.0244)
25	0.8	1.2	0.10900	0.07690	0.16120	0.03530	0.12840	0.00260
			(0.0463)	(0.0624)	(0.0796)	(0.0469)	(0.0470)	(0.0066)
50	0.8	1.2	0.10120	0.04580	0.13440	0.02320	0.10900	0.00180
			(0.0390)	(0.0306)	(0.0674)	(0.0260)	(0.0488)	(0.0010)
100	0.8	1.2	0.09580	0.04140	0.12410	0.01030	0.10950	0.00780
			(0.0465)	(0.0219)	(0.0625)	(0.0143)	(0.0489)	(0.0033)
25	1.5	0.5	0.37520	0.08740	0.54330	0.03340	0.42120	0.00620
			(0.1380)	(0.0652)	(0.2282)	(0.0328)	(0.1557)	(0.0064)
50	1.5	0.5	0.37710	0.04960	0.45810	0.01660	0.39400	0.00110
			(0.1425)	(0.0346)	(0.1800)	(0.0154)	(0.1333)	(0.0096)
100	1.5	0.5	0.37260	0.04760	0.41600	0.01020	0.37300	0.02620
			(0.1530)	(0.0280)	(0.1800)	(0.0038)	(0.1542)	(0.0188)
25	1.5	1.2	0.10920	0.00770	0.16110	0.03540	0.12840	0.00260
			(0.0463)	(0.0624)	(0.0691)	(0.0286)	(0.0439)	(0.0151)
50	1.5	1.2	0.10120	0.04580	0.13470	0.02320	0.10900	0.00180
			(0.0494)	(0.0388)	(0.0773)	(0.0216)	(0.0565)	(0.0013)
100	1.5	1.2	0.09570	0.04140	0.12440	0.01030	0.10900	0.00780
			(0.0451)	(0.0253)	(0.0585)	(0.0126)	(0.0478)	(0.0032)

In the case of the shape parameter, LSX is about 66.7% better with both the 10 and 20% censored data than that of LSY which is about 37.3% for both. MLE seem not to be a good estimator for the shape parameter when compared with the other two estimators.

We may therefore conclude that, in estimating the parameters of the Weibull distribution with 10 and 20% random censoring, it would be rewarding if Maximm Likelihood Estimator is use to estimate the scale parameter whiles Least Square Estimator on X is use for the shape parameter.

REFERENCES

Al-Kanani, H.I. and A.J. Shaima, 2011. Estimate survival function for the brain cancer disease by using three parameters weibull distribution. J. Basrah Res. Sci., pp: 80.

Berger, R.W. and K. Lawrence, 1974. Estimating weibull parameters by linear and nonlinear regression. Technometrics, 16(4): 617-619.

Horst, R., 2009. The Weibull Distribution Handbook. CRC Press, Boca Raton, pp: 784, ISBN: 1420087436.

Panani, H. and A. Saeid, 2011. Estimation of the weibull distribution based on type- ii censored samples. Appl. Math. Sci., 52(5): 2549-2558.

Saunders, C.S., 2007. Reliability, Life Testing and the Prediction of Service Lives: For Engineers and Scientists. Springer, New York, pp: 307, ISBN: 0387325220.

Stefano, P., E. Biganzoli and P. Boracchi, 2007. Review of the Maximum Likelihood Functions for Right Censored Data: A New Elementary Derivative. COBRA Preprint Series, Retrieved from: http://biostats.bepress.com/cobra/art21/.

Zhang, L.F., M. Xie and L.C. Tang, 2006. Bias correction for the least squares estimator of weibull shape parameter with complete and censored data. Reliab. Eng. Syst. Safe., 91: 930-939.

Zhang, L.F., M. Xie and L.C. Tang, 2007. A study of two estimation approaches for parameters of weibull distribution based on WPP. Reliab. Eng. Syst. Safe., 92: 360-368.