Journal of Al-Qadisiyah for computer science and mathematics
Vol.5 No.1 year 2013

Page123-129

Estimation of Weibull distribution Parameters with Complete Data Using T.O Method

By: Lecturer

Jubran A.A.K.

Department of Statistics and Information

College of Computer Science and Mathematics

University of Al-Qadisiya

E-mail: jubalaa98@yahoo.com

Recived : 2 /10 /2013 Accepted: 4 /11 / 2013

Abstract:

In this paper we present the comprehensive analysis for complete data of failure, and deal with the estimating the parameters of Weibull distribution. The Maximum likelihood and Least Square techniques were taken and compared with our method Term Omission (T.O). For illustration purpose, we obtained the results on sets of real data of computer hard disk failure.As well as we used simulation to enhance the comparison between the methods under study.

Keywords: *Weibull Distribution, Maximum likelihood, Least Square, Terms Omission*.

1. Introduction

Weibull distribution is one of the most important distributions used that has a paramount importance in the scientific studies that rely determine the life time in the application of reliability, as well as in various areas of failure.

Some of probability density functions f(t) may be used as a model of life time distribution that defined over the range of time, $0<t<\infty$. Another useful function is cumulative distribution F(t) of Weibull distribution, has been used as a median rank regression (MRR).

In this paper, we compared three methods (Maximum Likelihood, Least Square, Term Omission) to estimate Weibull Distribution Parameters from the data of 16 computer hard disk drive failures which recorded the number of failures occurring within fixed time using Mean Square Error (MSE)

2. Weibull Distribution

The probability density function of two-parameter has the form

$$f_w(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^{\beta}} \qquad 0 < t \quad \dots (1)$$

The cumulative Weibull distribution function is given by: [8]

$$F(t)=1-e^{-\left(\frac{t}{\eta}\right)^{\beta}} \quad \ldots (2)$$

where β>0 is a shape parameter, and η>0 is a scale parameter.

To estimate *F(t)* we may use one of the following methods presented in Table 1 where n is number of data points.[1]

Table 1. Methods for estimating $F(t_i)$.

Method	$F(t_i)$
Mean Rank	$\frac{i}{n+1}$
Median Rank	$\frac{i-0.3}{n+0.4}$
Symmetrical CDF	$\frac{i-0.5}{n}$

3. Methods of Analysis:

In this paper, we will study three methods (Maximum Likelihood, Least Square, Term Omission), to compare between them, choosing the best estimator of Weibull Distribution Parameters from the data of 16 computer hard disk drive failures.

3.1 Maximum Likelihood Method (MLM)

In maximum likelihood estimator, the corresponding likelihood equations need to be solved numerically. In this paper, we propose to estimate parameters of Weibull distribution as follows:

Let us assume that the lifetime *t* of a product follows the Weibull distribution W(β,η) with probability density function

$$f_w(t)=\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}e^{-\left(\frac{t}{\eta}\right)^{\beta}} \qquad 0<t$$

Now we tried to determine the maximum likelihood estimator of the parameters β and η,

$$\frac{\partial \ln L}{\partial \eta}=-\frac{n\beta}{\eta}+\frac{\beta}{\eta^{\beta+1}}\sum t_i^{\beta}=0 \quad \ldots (3)$$

$$\frac{\partial \ln L}{\partial \beta}=\frac{n}{\beta}-n\ln\eta+\sum \ln(t_i)+\frac{\ln\eta}{\eta^{\beta}}\sum t_i^{\beta}-\frac{1}{\eta^{\beta}}\sum t_i^{\beta}\ln(t_i)=0 \quad \ldots (4)$$

Then we can educe the estimator of η as follows:

$$\hat{\eta}=\left(\frac{\sum t_i^{\beta}}{n}\right)^{1/\beta}$$

and for β.[2]

$$\frac{1}{\beta}-\frac{1}{\sum t_i^{\beta}}\sum t_i^{\beta}\ln(t_i)+\frac{\sum \ln(t_i)}{n}=0 \quad \ldots (5)$$

which is usually solved numerically by Newten-Raphson method, which can be written in the form

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

3.2 Least Square Method (LSM)

The other method technique we shall study is known as the Least Square Method. We assume there is a linear relation between the two variables.

The cumulative Weibull distribution function is given by:

$$F(t) = 1 - e^{-\left(\frac{t}{\eta}\right)^{\beta}}$$

To come up with a relation between CDF and the two parameters β, η of Weibull distribution, we take the double logarithmic transformation of the CDF.

$$1 - F(t) = e^{-\left(\frac{t}{\eta}\right)^{\beta}}$$

$$\frac{1}{1 - F(t)} = e^{\left(\frac{t}{\eta}\right)^{\beta}}$$

$$\ln\left[\frac{1}{1 - F(t)}\right] = \left(\frac{t}{\eta}\right)^{\beta}$$

$$\ln t = \frac{1}{\beta}\ln\left\{\ln\left[\frac{1}{1 - F(t)}\right]\right\} + \ln\eta \qquad \dots (6)$$

Equation (6) can be written as $y = bx + a$ [9]

where

$$x = \ln\left\{\ln\left[\frac{1}{1 - F(t)}\right]\right\} \ , \ y = \ln(t) \ , \ b = \frac{1}{\beta} \ , \ a = \ln\eta$$

by linear regression formula. [3]

$$b = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}$$

or

$$\hat{\beta} = \frac{n\sum x_i^2 - \left(\sum x_i\right)^2}{n\sum x_i y_i - \sum x_i \sum y_i} \qquad \dots (7)$$

and

$$\hat{\eta} = \exp\left(\frac{\sum_{i=1}^{n} y_i}{n} - \frac{\sum_{i=1}^{n} x_i}{n\hat{\beta}}\right) \qquad \dots (8)$$

and $F(t_i)$ can be estimated by using Benard's formula, $\frac{i - 0.3}{n + 0.4}$ which is a good approximation to the median rank estimator.[9]

3.3 Term Omission Method (TOM)

The last method technique we discuss is Term Omission Method (TOM), by using median rank to estimate $F(t_i)$, with the following procedures: [5,6,7]

1- Recall Cumulative Function of Weibull distribution for any two values of t's.

$t_i \qquad F(t_i) = 1 - e^{-\left(\frac{t_i}{\eta}\right)^{\beta}}$

$t_j \qquad F(t_j) = 1 - e^{-\left(\frac{t_j}{\eta}\right)^{\beta}}$

2- Subtract (1) from each and multiply with (-1) to obtain

$t_i \qquad e^{-\left(\frac{t_i}{\eta}\right)^{\beta}}$

$t_j \qquad e^{-\left(\frac{t_j}{\eta}\right)^{\beta}}$

3- Taking logarithm to obtain

$t_i \qquad -\left(\frac{t_i}{\eta}\right)^{\beta}$

$t_j \qquad -\left(\frac{t_j}{\eta}\right)^{\beta}$

4- Multiple each by (-1) to obtain

$t_i \qquad \frac{t_i^{\beta}}{\eta^{\beta}}$

$t_j \qquad \frac{t_j^{\beta}}{\eta^{\beta}}$

and let $k = \left(\frac{t_i}{\eta}\right)^{\beta}$ to use it later for estimating the η (scale parameter of Weibull distribution)

5- Divide the second on the first and to obtain

$$\left(\frac{t_i}{t_j}\right)^{\beta}$$

6- Again taking logarithm to obtain

$$\beta\left[\ln t_i - \ln t_j\right]$$

7- Finally, as divided by ($\ln t_i - \ln t_j$), we can estimate β (shape parameter)

Summarization of above procedures, the estimator of β is

$$\beta^{\ell} = \frac{\ln\left[\frac{\ln[1-F(t_i)]}{\ln[1-F(t_j)]}\right]}{\ln t_i - \ln t_j} \quad , \quad 1 \le \ell \le n-1 \quad \ldots(9)$$

and

$$\hat{\beta} = \underset{1\le \ell \le n-1}{Min}\left[\sum_{i=1}^{n} F\{(t_i;\beta^{\ell}) - F(t_i;\beta)\}^2\right] \ \ldots (10)$$

Once $\hat{\beta}$ is obtained, then $\hat{\eta}$ can easily be obtained. Recall

$$k = \left(\frac{t_i}{\eta}\right)^{\beta}$$

Hence

$$\hat{\eta} = \frac{t_i}{\sqrt[\beta]{k}}$$

4. Numerical Example

In order to illustrate and compare the three analytical methods MLM, LSM and TOM using MSE (Mean Square Error) which can be calculated by the following equation

$$MSE = \frac{\sum_{i=1}^{n}\{\hat{F}(t_i) - F(t_i)\}^2}{n}$$

where $\hat{F}(t_i) = 1 - e^{-\left(\frac{t_i}{\hat{\eta}}\right)^{\hat{\beta}}}$ and $F(t_i) = \frac{i-0.3}{n+0.4}$ with 50% rank.

we shall introduce an example with complete data of 16 Computer Hard Disk Failure which it summarized in Table 2 using median rank (Table 1)

Table 2. Analysis of Hard Disk Failure Data. [9]

I	Time in hour	F(t)	i	Time in hour	F(t)
1	7	0.042683	9	380	0.530488
2	12	0.103659	10	388	0.591463
3	49	0.164634	11	437	0.652439
4	140	0.22561	12	472	0.713415
5	235	0.286585	13	493	0.77439
6	260	0.347561	14	524	0.835366
7	320	0.408537	15	529	0.896341
8	320	0.469512	16	592	0.957317

5. Results

The comparison of the three methods is based on values from Mean Square Error (MSE) and Mean Absolute Error (MAE).[4] The following table represents the comparison between the methods (MLE, LSM, TOM) for two-parameters Weibull distribution and choose the best method according to MSE (Mean Square Error) with equation below:

$$MSE = \frac{\sum_{i=1}^{n}\{\hat{F}(t_i) - F(t_i)\}^2}{n}, \quad MAE = \frac{\sum_{i=1}^{n}\left|\hat{F}(t_i) - F(t_i)\right|}{n}$$

with n= number of sample used.

Table 3. The estimate value of parameters that was found by LSM, MLE, MOM and T.O.M for Weibull distribution with β scale parameter and shape parameter α=1.

Method	β	η	MSE	MAE
TOM	2.17610663	426.0425962	0.00538024	0.056791487
MLE	1.367928	344.5511	0.011035795	0.091990963
LSM	0.917862	366.5093	0.012952	0.095612

6. Simulation with Computational Results

In this research, the comparison between the methods is our primary goal, namely, MLE, LSM, TOM. Where we generate a random samples of different sizes with known parameters, namely 20,50,100. The total deviation is what we used for the purpose of comparison for each method as follows

$$TD = \left|\frac{\hat{\beta}-\beta}{\beta}\right| + \left|\frac{\hat{\eta}-\eta}{\eta}\right| \quad \ldots (11)$$

where β and η are the known parameters, and $\hat{\beta}$ and $\hat{\eta}$ are the estimated parameters in each method. And the results are placed in the table 4.

Table 4. The Comparison between MLE, LSM, TOM with different sample sizes for five decimals.

N	β	η	Sample size	MLE			LSM			TOM		
				β	η	TD	β	η	TD	β	η	TD
1	1	10	20	1.40719	12.36816	0.644	0.9599	9.77674	0.06243	1.02948	9.7765	0.05183
2			50	1.02243	9.74107	0.04832	0.98467	9.78753	0.03658	0.98642	9.96356	0.01722
3			100	0.98901	11.41588	0.15258	0.97311	9.81788	0.0451	0.9788	9.81581	0.03962
4	3.5	25	20	4.34018	25.06561	0.24268	3.47839	24.8461	0.01233	3.4921	24.8505	0.00824
5			50	3.34109	25.97074	0.08423	3.42649	24.88115	0.02576	3.49516	24.94043	0.00377
6			100	3.42776	24.14101	0.055	3.4275	24.8747	0.02573	3.45984	24.89887	0.01552
7	1.5	200	20	1.64436	219.48635	0.19367	1.48653	197.67111	0.02062	1.50106	198.49437	0.00823
8			50	1.23748	150.76069	0.42121	1.44298	199.06342	0.0427	1.48352	198.66948	0.01764
9			100	1.53232	211.66359	0.07987	1.47368	195.38124	0.04064	1.50097	197.08974	0.0152

10			20	18.46156	247.30758	0.08769	19.25052	249.88396	0.03794	19.82637	249.79894	0.00949
11	20	250	50	24.15989	249.47186	0.21011	19.81746	249.76773	0.01006	20.06735	249.74747	0.00438
12			100	17.87796	248.04275	0.11393	18.8088	249.84854	0.06017	19.88782	249.82436	0.00631

7. Conclusion

We presented analytical methods for estimating scale parameter of two-parameter Weibull distribution (TOM, MLE, and LSM). It has been shown from the computational results that TOM is the best estimate method for two parameters Weibull distribution with complete data and with simulation random samples.

References:

[1] Al-Fawzan M. A., “Methods for Estimating the Parameters of the Weibull Distribution”, King Abdulaziz City for Science and Technology, p. 4-10, 2000.

[2] Balakrishnan N. and Kateri M., "On the maximum likelihood estimation of parameters of Weibull distribution based on complete and censored data", Elsevier, 2008.

[3] Bhattacharya P., "Weibull Distribution for Estimating the Parameters", Kolaghat India, 2011.

[4] Dp-Hun Lee, "Comparing the inverse parameter estimation approach with pedo-transfer function method for estimating soil hydraulic conductivity", Geosciences Journal, Vol. (9), Issue (3), PP. 269-276, 2005.

[5] Labban J.A., “Modified and Simplified Method for Finite Difference (Divided)”. Al-Qadisiya Journal for Science, Vol. 10, No. 2, pp. 244-251, 2005.

[6] Labban J.A., “Approximation of Scale Parameter of Inverted Gamma Distribution by TOM Modified”. Vol. (1), No. (1), PP. -8, 2011.

[7] Labban J.A., "Estimation of Scale Parameter of Weibull distribution that has Exponential Family using T.O.M", Al-Qadisiya Journal for Science, Vol. 5, No. 1, PP. 83 - 91, 2013.

[8] Lei Y., "Evaluation of three methods for estimating the Weibull distribution parameters of Chinese pine (Pinus tabulaeformis)", Journal Of Forest Science, Vol. 54, No. 12, PP. 566–571, 2008.

[9] Pasha G.R et al., "Empirical Analysis of The Weibull Distribution for Failure Data, Journal of Statistics, Vol. 13, No. 1. PP. 33-45, 2006.