



## Estimating of Survival Function under Type One Censoring Sample for Mixture Distribution

Qesma S. Abadi

Iden H. AL-Kanani

Department of Mathematics, College of Science for Women, University of Baghdad

[Qesma.Sabar1103a@cs.w.uobaghdad.edu.iq](mailto:Qesma.Sabar1103a@cs.w.uobaghdad.edu.iq)

[Idenalkanani@yahoo.com](mailto:Idenalkanani@yahoo.com)

Article history: Received 31 October,2019, Accepted 16 December 2019, Published in October 2020

Doi: [10.30526/ 33.4.2515](https://doi.org/10.30526/33.4.2515)

### Abstract

In this article, it is interesting to estimate and derive the three parameters which contain two scales parameters and one shape parameter of a new mixture distribution for the singly type one censored data which is the branch of right censored sample. Then to define some special mathematical and statistical properties for this new mixture distribution which is considered one of the continuous distributions characterized by its flexibility. Next, using maximum likelihood estimator method for singly type one censored data based on the Newton-Raphson matrix procedure to find and estimate values of these three parameters by utilizing the real data taken from the National Center for Research and Treatment of Hematology/University of Mustansiriyah for leukemia diseases. After that we find and derive the estimate of probability density function, estimate survival function and finally estimate the hazard function.

**Keyword:** New Mixture Distribution (NMD), Maximum likelihood estimation method (MLEM), Type- one censored sample data, Hazard function.

### 1.Introduction

A new proposed distribution was introduced by the researchers Iden and Maysaa, in (2018) [1] which they called “new mixture distribution”. The new mixture distribution depended on three parts, the first part was a mixture between exponential and standard Weibull distributions, the second part was a mixture between exponential and Rayleigh distributions, and the third part was a mixture between the first part and the second .This



distribution was characterized by a flexible distribution and was based on three parameter  $(\alpha, \beta, \gamma)$  two of them were scale parameters  $\gamma, \beta$  and one shape parameter  $\alpha$ .

In this paper, reference was made to the previous paper, which was based on the same idea G.M. Cordeiro et.al, in(2014)[5] that represented the new method to find the mixed distribution depending on tail of Exponential one parameter distribution and Weibull two parameters distribution.

Suleman.N in (2016)[4] introduced the new distribution depend on the tail of Weibull two parameters distribution and Rayleigh one parameter distribution.

The aim of this paper is to estimate the values of unknown parameters for the new mixed distribution (NMD) based on censored data (type- one censored sample). Then to estimate the death density function, survival function and hazard function.

The rest of the research is organized as follows: section two is devoted definition and some properties of distribution (NMD), section three is for deriving point estimation and finally, section four is made for results and discussion.

**2-Definition and Properties of (NMD)[2,3]:**

**The cdf of NMD  $(\alpha, \beta, \gamma)$  is as follows:**

$$F(t) = 1 - \left[ e^{-(2\gamma t + \frac{\beta}{2}t^2 + t^\alpha)} \right]$$

**Then the probability density function of (NMD) is as follows:**

$$f(t; \alpha, \beta, \gamma) = \begin{cases} (2\gamma + \beta t + \alpha t^{\alpha-1}) e^{-(2\gamma t + \frac{\beta}{2}t^2 + t^\alpha)} & , t > 0 \\ 0 & , 0. w \end{cases}$$

This new model has two scale parameters which are denoted by  $\gamma, \beta$  and one shape parameter  $\alpha$ , the parameter space is follows:

$$\Omega = \{(\gamma, \beta, \alpha); \gamma > 0; \beta > 0; \alpha > 0\}$$

**The rth moment of this distribution is:**

$$\mu'_r = E(t^r) = 2\gamma k(r, \gamma, \beta, \alpha) + \beta k(r + 1, \gamma, \beta, \alpha) + k(r + \alpha - 1, \gamma, \beta, \alpha)$$

**The mean of this distribution is:**

$$\mu'_1 = E(t) = 2\gamma k(1, \gamma, \beta, \alpha) + \beta k(2, \gamma, \beta, \alpha) + \alpha k(3, \gamma, \beta, \alpha)$$

**The survival of this distribution is:**

$$s(t) = e^{-(2\gamma t + \frac{\beta}{2}t^2 + t^\alpha)} \quad , t > 0$$

**The hazard of this distribution is:**

$$h(t) = (2\gamma + \beta t + \alpha t^{\alpha-1}) \quad , t > 0$$

**The variance of this distribution is:**

$$\sigma_t^2 = 2\gamma k(2, \gamma, \beta, \alpha) + \beta k(3, \gamma, \beta, \alpha) + \alpha k(a + 1, \gamma, \beta, \alpha) + [2\gamma k(1, \gamma, \beta, \alpha) + \beta k(2, \gamma, \beta, \alpha) + \alpha k(a, \gamma, \beta, \alpha)]^2$$

Where:

$$k(r, \gamma, \beta, \alpha) = \frac{1}{\alpha} \sum_{n=0}^{\infty} \frac{(-2\gamma)^n}{n!} \cdot {}_1\psi_0 \left[ \left( \frac{r+n+1}{\alpha}, \frac{2}{\alpha} \right), \frac{-\beta}{2} \right]$$

$$k(r + 1, \gamma, \beta, \alpha) = \frac{1}{\alpha} \sum_{n=0}^{\infty} \frac{(-2\gamma)^n}{n!} \cdot {}_1\psi_0 \left[ \left( \frac{r+n+2}{\alpha}, \frac{2}{\alpha} \right), \frac{-\beta}{2} \right]$$

$$k(r + \alpha - 1, \gamma, \beta, \alpha) = \frac{1}{\alpha} \sum_{n=0}^{\infty} \frac{(-2\gamma)^n}{n!} \cdot {}_1\psi_0 \left[ \left( \frac{r+n+\alpha}{\alpha}, \frac{2}{\alpha} \right), \frac{-\beta}{2} \right]$$

### 3. Parameters Estimation

Now, we shall show the estimate and derive the unknown parameters of the new mixture distribution by using maximum likelihood method for type- one censoring sample.

#### 3.1. Type- One Censoring Data:

It is one of the most important right censoring types .in which the number of failed samples (r) is obtained at the end of a particular experiment or study.

These samples are not specified during the time period of  $t_1 < t_2 < \dots < t_m$  , while from the beginning, the sample size (n) is defined and known to the researcher supervising the experiment and therefore, the number of the remaining samples resulting from the study will be n-r[6].

#### 3.2. The Maximum Likelihood Estimator Method for Type-One Censoring Data:

The main advantage of the maximum likelihood properties method contains the best affinity. That means when the values of the data increase, the estimate converges will be better and faster with the (parameters) in short; it maximizes the value of the function [7].

The likelihood function (joint probability density function) of  $t_1 < t_2 < \dots < t_m$  is as follows:

$$L = \frac{n!}{(n-m)!} \prod_{i=1}^m \{f(t_i; \theta)[s(t_0; \theta)]^{n-m} \} \quad 0 \leq t_1 \leq t_2 \dots \leq t_m \leq t_0 \quad (1)$$

The likelihood function for N.M.D  $(\alpha, \beta, \gamma)$

$$L = \frac{n!}{(n-m)!} \prod_{i=1}^m \left\{ (2\gamma + \beta t_i + \alpha t_i^{\alpha-1}) e^{-(2\gamma t_i + \frac{\beta}{2} t_i^2 + t_i^\alpha)} \left[ e^{-(2\gamma t_0 + \frac{\beta}{2} t_0^2 + t_0^\alpha)} \right]^{n-m} \right\} \quad (2)$$

Let  $\frac{n!}{(n-m)!} = k$

Then we get:

$$(3)L = k \prod_{i=1}^m \left\{ (2\gamma + \beta t_i + \alpha t_i^{\alpha-1}) e^{-(2\gamma t_i + \frac{\beta}{2} t_i^2 + t_i^\alpha)} \left[ e^{-(2\gamma t_0 + \frac{\beta}{2} t_0^2 + t_0^\alpha)} \right]^{n-m} \right\}$$

After that, we take the logarithm of the two side of the equation:

$$\ln L = \ln k - (n - m) \left( 2\gamma t_0 + \frac{\beta}{2} t_0^2 + t_0^\alpha \right) + \sum_{i=1}^m \ln(2\gamma + \beta t_i + \alpha t_i^{\alpha-1}) - \sum_{i=1}^m (2\gamma t_i + \frac{\beta}{2} t_i^2 + t_i^\alpha) \dots (4)$$

Deriving the equation (4) for the parameters  $(\alpha, \beta, \gamma)$  we obtain three nonlinear equations respectively as follows:

$$\frac{\partial \ln l}{\partial \alpha} = -(n - m)(t_0^\alpha \ln t_0) + \sum_{i=1}^m \frac{\alpha t_i^{\alpha-1} \ln t_i + t_i^{\alpha-1}}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})} - \sum_{i=1}^m t_i^\alpha \ln t_i \quad (5)$$

$$\frac{\partial \ln l}{\partial \beta} = -(n - m) \left( \frac{1}{2} t_0^2 \right) + \sum_{i=1}^m \frac{t_i}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})} - \sum_{i=1}^m \frac{1}{2} t_i^2 \quad (6)$$

$$\frac{\partial \ln l}{\partial \gamma} = -(n - m)(2t_0) + \sum_{i=1}^m \frac{2}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})} - \sum_{i=1}^m 2t_i \quad (7)$$

The three nonlinear equation (5), (6), (7) are difficult to solve then, using numerical methods such as Newton-Raphson procedure will help to find the estimate values for this parameter

$$\begin{bmatrix} \alpha_{k+1} \\ \beta_{k+1} \\ \gamma_{k+1} \end{bmatrix} = \begin{bmatrix} \alpha_k \\ \beta_k \\ \gamma_k \end{bmatrix} - J^{-1} \begin{bmatrix} f(\alpha) \\ f(\beta) \\ f(\gamma) \end{bmatrix}$$

Let  $\frac{\partial \ln l}{\partial \alpha} = f(\alpha)$ ,  $\frac{\partial \ln l}{\partial \beta} = f(\beta)$ ,  $\frac{\partial \ln l}{\partial \gamma} = f(\gamma)$

$$f(\alpha) = -(n - m)(t_0^\alpha \ln t_0) + \sum_{i=1}^m \frac{\alpha t_i^{\alpha-1} \ln t_i + t_i^{\alpha-1}}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})} - \sum_{i=1}^m t_i^\alpha \ln t_i \tag{8}$$

$$f(\beta) = -(n - m) \left( \frac{1}{2} t_0^2 \right) + \sum_{i=1}^m \frac{t_i}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})} - \sum_{i=1}^m \frac{1}{2} t_i^2 \tag{9}$$

$$f(\gamma) = -(n - m)(2t_0) + \sum_{i=1}^m \frac{2}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})} - \sum_{i=1}^m 2t_i \tag{10}$$

The Jacobin matrix ( $j_k$ ) is defined as the following formula:

$$j_k^{-1} = \begin{bmatrix} \frac{\partial f(\alpha)}{\partial \alpha} & \frac{\partial f(\alpha)}{\partial \beta} & \frac{\partial f(\alpha)}{\partial \gamma} \\ \frac{\partial f(\beta)}{\partial \alpha} & \frac{\partial f(\beta)}{\partial \beta} & \frac{\partial f(\beta)}{\partial \gamma} \\ \frac{\partial f(\gamma)}{\partial \alpha} & \frac{\partial f(\gamma)}{\partial \beta} & \frac{\partial f(\gamma)}{\partial \gamma} \end{bmatrix}^{-1}$$

Which is square symmetric matrix now deriving each function  $f(\alpha)$ ,  $f(\beta)$ ,  $f(\gamma)$  for the parameters  $\alpha, \beta, \gamma$  as follows:

$$f(\alpha) = -(n - m)(t_0^\alpha \ln t_0) + \sum_{i=1}^m \frac{\alpha t_i^{\alpha-1} \ln t_i + t_i^{\alpha-1}}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})} - \sum_{i=1}^m t_i^\alpha \ln t_i$$

$$\frac{\partial f(\alpha)}{\partial \alpha} = -(n - m)t_0^\alpha (\ln t_0)^2 + \sum_{i=1}^m \frac{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})[\alpha t_i^{\alpha-1} (\ln t_i)^2 + t_i^{\alpha-1} \ln t_i] - (\alpha t_i^{\alpha-1} \ln t_i + t_i^{\alpha-1})^2}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})^2} -$$

$$\sum_{i=1}^m t_i^\alpha (\ln t_i)^2$$

$$\frac{\partial f(\alpha)}{\partial \beta} = \sum_{i=1}^m \frac{-(\alpha t_i^{\alpha-1} \ln t_i + t_i^{\alpha-1}) \cdot t_i}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})^2} = - \sum_{i=1}^m \frac{t_i^\alpha (\alpha \ln t_i + 1)}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})^2}$$

$$\frac{\partial f(\alpha)}{\partial \gamma} = \sum_{i=1}^m \frac{-2(\alpha t_i^{\alpha-1} \ln t_i + t_i^{\alpha-1})}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})^2} = - \sum_{i=1}^m \frac{2t_i^{\alpha-1} (\alpha \ln t_i + 1)}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})^2}$$

Similarly the function  $f(\beta)$  deriving for the parameter  $\alpha, \beta, \gamma$  is as follows:

$$f(\beta) = -(n - m) \left( \frac{1}{2} t_0^2 \right) + \sum_{i=1}^m \frac{t_i}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})} - \sum_{i=1}^m \frac{1}{2} t_i^2$$

$$\frac{\partial f(\beta)}{\partial \alpha} = \sum_{i=1}^m \frac{-(\alpha t_i^{\alpha-1} \ln t_i + t_i^{\alpha-1}) t_i}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})^2} = - \sum_{i=1}^m \frac{t_i^\alpha (\alpha \ln t_i + 1)}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})^2}$$

$$\frac{\partial f(\beta)}{\partial \beta} = \sum_{i=1}^m \frac{-t_i^2}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})^2} = - \sum_{i=1}^m \frac{t_i^2}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})^2}$$

$$\frac{\partial f(\beta)}{\partial \gamma} = \sum_{i=1}^m \frac{-2t_i}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})^2} = - \sum_{i=1}^m \frac{2t_i}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})^2}$$

For the same deriving the function for the parameters  $\alpha, \beta, \gamma$  is as follows:

$$f(\gamma) = -(n - m)(2t_0) + \sum_{i=1}^m \frac{2}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})} - \sum_{i=1}^m 2t_i$$

$$\frac{\partial f(\gamma)}{\partial \alpha} = \sum_{i=1}^m \frac{-2(\alpha t_i^{\alpha-1} \ln t_i + t_i^{\alpha-1})}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})^2} = - \sum_{i=1}^m \frac{2t_i^{\alpha-1} (\alpha \ln t_i + 1)}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})^2}$$

$$\frac{\partial f(\gamma)}{\partial \beta} = \sum_{i=1}^m \frac{-2t_i}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})^2} = - \sum_{i=1}^m \frac{2t_i}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})^2}$$

$$\frac{\partial f(\gamma)}{\partial \gamma} = \sum_{i=1}^m \frac{-4}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})^2} = - \sum_{i=1}^m \frac{4}{(2\gamma + \beta t_i + \alpha t_i^{\alpha-1})^2}$$

Finally we stopped the Newton-Raphson method by using the following formula:

$$\begin{bmatrix} \alpha_{k+1} - \alpha_k \\ \beta_{k+1} - \beta_k \\ \gamma_{k+1} - \gamma_k \end{bmatrix} \leq \begin{bmatrix} \epsilon_\alpha \\ \epsilon_\beta \\ \epsilon_\gamma \end{bmatrix}$$

Finally, we estimate the probability death density function  $\hat{f}(t)$ , survival function  $\hat{s}(t)$  and hazard function  $\hat{h}(t)$  where:

$$\begin{aligned} \hat{f}(t) &= (2\hat{\gamma} + \hat{\beta}t + \hat{\alpha}t^{\hat{\alpha}-1})e^{-\left(2\hat{\gamma}t + \frac{\hat{\beta}}{2}t^2 + t^{\hat{\alpha}}\right)} \\ \hat{s}(t) &= e^{-\left(2\hat{\gamma}t + \frac{\hat{\beta}}{2}t^2 + t^{\hat{\alpha}}\right)} \\ \hat{h}(t) &= (2\hat{\gamma} + \hat{\beta}t + \hat{\alpha}t^{\hat{\alpha}-1}) \end{aligned}$$

#### 4. Results and Discussion

This study is based on samples and real data taken from the National Center For Research and Treatment Of Hematology/University of Mustansiriyah of chronic lymphocytic leukemia which is one type of leukemia cancers that effects people from aged 60-80 years which is symbolized (CLL). The duration of the study was six months, equivalent to (184) days. The number of patients who entered the study was (64) patients, (29) of them left the study and we did not know their fate because we were not able to follow up them and (11) of them died during the study period.

Now, if we know the distribution of this sample (data), we use chi-square test goodness of fit which is determined whether the data (sample) corresponded with the mixture distribution on through the spss program. The null and alternative hypotheses for chi-square test are:

$H_0$  : the data are distributed as a mixture distribution

$H_1$  : the data are not distributed as a mixture distribution

Then, calculate chi-square=18.14 < tabulate chi-square=18.31

with level of significant (0.05) and degree of freedom equal (10),

That means the data are distributed as new mixture distribution NMD

By using the initial values for parameters which are taken from (Maysaa and Iden)[1]

$\alpha_0 = 1$ ,  $\beta_0 = 0.01$ ,  $\gamma_0 = 0.004$

After that we utilizing matlab programing we get the following estimated parameters values:

$\hat{\alpha} = 0.09956$ ,  $\hat{\beta} = 0.00151$ ,  $\hat{\gamma} = 0.0028$

Then we evaluating the results for probability death density function, next evaluating the survival

function and evaluating the hazard rate function, then we putt in the following table.

Table (1-1): Estimated values of functions  $\hat{f}(t)$ ,  $\hat{s}(t)$ ,  $\hat{h}(t)$  by MLE method

Failure Time	$\hat{f}(t)$	$\hat{s}(t)$	$\hat{h}(t)$
49	0.00234762805704734596	0.02842736024291173129	0.08258339982984242389
51	0.00205444195382150316	0.02402926645266854301	0.08549748939977938584
53	0.0017854862339620879	0.02019342471617567619	0.08841918887249708214
57	0.0013211472758740978	0.01401267382325285954	0.09428231132317976755
62	0.00087272428840775696	0.00858627088390184751	0.10164183033684654057
63	0.00079923193843544844	0.007750715000492989	0.10311718833483269584
65	0.00066694380862687189	0.00628771429662956728	0.10607094679609995568
68	0.00050211198698322505	0.00454364869879524079	0.11050854066168478889
71	0.00037243676210255601	0.00323988858895136166	0.11495357074086943566
78	0.0001751429319184233	0.00139723621318390798	0.12534955096770778722
82	0.00010980438378349287	0.00083626824009491583	0.13130282667560130698

From the following table we note that:

1. Showing that the value of  $f(t; \alpha, \beta, \gamma)$  is decreasing when the life times are increasing.
2. Showing that the value of  $s(t; \alpha, \beta, \gamma)$  is decreasing when the life times are increasing.
3. Showing that the value of  $h(t; \alpha, \beta, \gamma)$  is increasing when the life times are increasing.

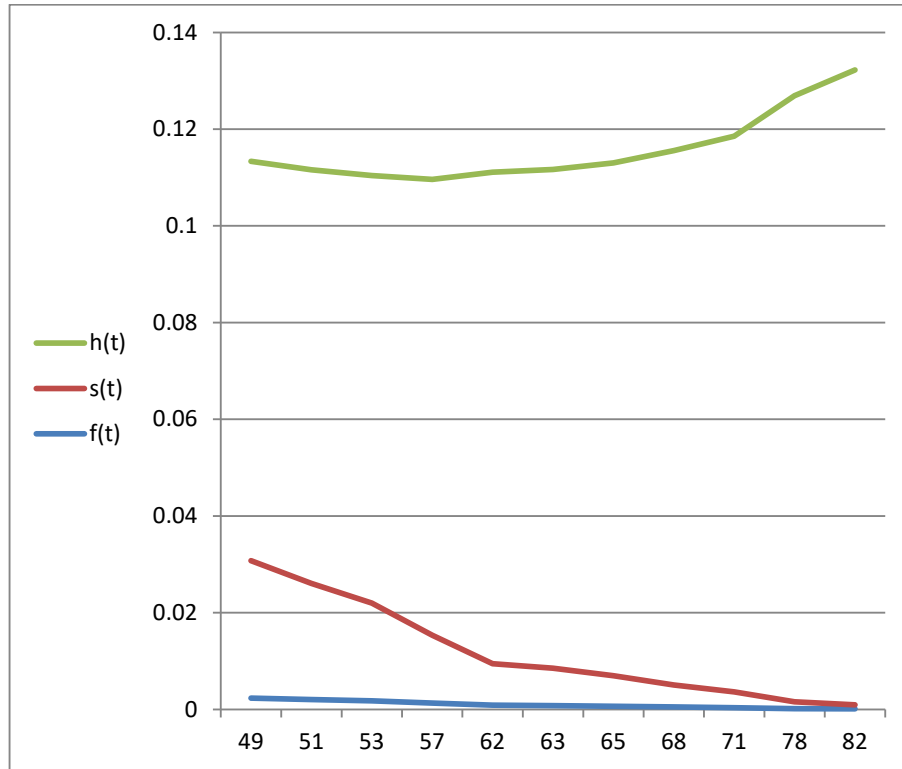


Figure 1. The  $f(t)$ ,  $s(t)$ ,  $h(t)$  respect to  $t$

## 5. Conclusions

In this paper, we can derive the parameters of a new mixture distribution which contains three parameters: two of them are scales and one of them is shape for singly type one censoring samples. Then we estimate the values of these parameters by using maximum likelihood estimator method. After that we find and estimate the values of probability death density function  $\hat{f}(t)$ , survival function  $\hat{s}(t)$  and hazard function  $\hat{h}(t)$ . We note that each of both of the  $f(t; \alpha, \beta, \gamma)$  and  $s(t; \alpha, \beta, \gamma)$  is decreasing when the life times are increasing while the  $h(t; \alpha, \beta, \gamma)$  is increasing when the life times are increasing. Therefore, we can show the behavior of three probability functions in Figure (1).

## References

1. Mohammed, M. J.; Hussein, I.H. Some Estimation Methods for New Mixture Distribution with Simulation and Application [c]. Iop Conference Series: *Materials Science and Engineering*, **2019**, 571, IOP Publishing,
2. Mohammed, M. J. ; Hussein, I.H. Study of New Mixture Distribution[a]., Int., Conf., *journal of engineering and apply Science ICEMASP in Turkey*. **2018**, 14,7,566-573ISSN:1816-949X
3. Mohammed, M. J. ;Hussein, I.H. comparison between new mixture distribution and many other distributions[b] . *jour adv research in dynamical and control systems*, **2018**, 10. ISSN :1943-023X
4. Nasiru, S. Serial Weibull Rayleigh distribution: theory and application. *International Journal of Computing Science and Mathematics*, **2016**, 7,3,239-244.
5. Cordeiro, G. M., et al. The exponential–Weibull lifetime distribution. *Journal of Statistical Computation and simulation* , **2014**, 84,12, 2592-2606.
6. Cox, D. R. Analysis of survival data; Chapman and Hall/CRC;1 st ed; Crcpress,**2018**,1351466615.
7. Dai, H. ; Wang, H. Analysis for Time-to-Event Data under Censoring and Truncation,1 st ed. *Acadmic press*,**2016**,0128054808