PROGRESSIVELY CENSORING DATA TO ESTIMATE THE SURVIVAL FUNCTION IN THE NEW MIXTURE DISTRIBUTION

Qesma S. Abadi and Iden H.AL-Kanani

Department Of Mathematics /College Of Science For Women, University of Baghdad, Iraq
E-mail: Qesma.Sabar1103a@csu.uobaghdad.edu.iq

Received: 10.11.2019 Revised: 12.12.2019 Accepted: 13.01.2020

Abstract:
In this article, we are interested in deriving and determining the three parameters, two scale parameters and one shape parameter, of new mixture distribution for progressively censored data which is the branch of right censored sample. Then 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 estimation method for progressively censored data based on the Newton-Raphson matrix procedure to find and estimate values of these three parameter by utilizing the real data taken from the National Center For Research and Treatment of Hematology/University of Mustansirya for leukemia diseases. After finding and deriving the estimate of probability density function, estimate survival function and finally estimate the hazard function.

Key word: New Mixture Distribution (NMD), Maximum likelihood estimation method (MLEM), progressively censored data, Hazard function, Newton -Raphson method

INTRODUCTION:
A new proposed distribution is introduced by the researchers Iden and Maysaa, in (2018) [1] who called it "new mixture distribution". The new mixture distribution depends on three parts, the first part is mixed between exponential and standard Weibull distributions. The second part is mixed between exponential and Rayleigh distributions, The third part is mixed between first part and second part .This distribution is characterized by flexible distribution and is based on three parameter(α, β, γ) two of them are scale parameters γ, β and one is a shape parameter α.

A previous study done by Iden and Maysaa depended on the same idea G.M. Cordeiro et al, in(2014) [5] which represents the new method to find the mixed distribution depend on tail of one parameter Exponential distribution and two Weibull distributions.

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

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

The rest of the research is organized as follows: in section two definition and some properties of distribution (NMD) are given, in section three the deriving of point estimation is presented. Finally, in section four contains the results and discussion.

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

The cdf of NMD (α, β, γ) is as follows:

\[ F(t) = 1 - e^{-(2y+β+at^{-α})} \]

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

\[ f(t; α, β, γ) = \begin{cases} (2y + βt + at^{-α})e^{-(2y+β+at^{-α})} & , t > 0 \\ 0 & , o.w \end{cases} \]

This new model have two scale parameters which are denoted by γ, β and one shape parameter α, the parameter space are follows:

\[ Ω = \{(α, β, γ) : γ > 0; β > 0; α > 0\} \]

The rth moment of this distribution was:

\[ m_r = E(t^r) = 2yk(r, γ, β, α) + βk(r + 1, γ, β, α) + k(r + a - 1, γ, β, α) \]

The mean of this distribution is:

\[ μ = E(t) = 2yk(1, γ, β, α) + βk(2, γ, β, α) + ak(3, γ, β, α) \]

The survival of this distribution is:

\[ s(t) = e^{-(2y+β+at^{-α})} \]

The hazard of this distribution is:

\[ h(t) = (2y + βt + at^{-α}) \]

The variance of this distribution is:

\[ \sigma^2 = 2yk(2, γ, β, α) + βk(3, γ, β, α) + ak(a + 1, γ, β, α) + [2yk(1, γ, β, α) + βk(2, γ, β, α) + ak(a, γ, β, α)]^2 \]

Where:

\[ k(r, γ, β, α) = \sum_{n=0}^{∞} \frac{(-2y)^n}{n!}, \psi'(\frac{r + a + 1}{a} - \frac{γ}{2}) \]

\[ k(r + 1, γ, β, α) = \sum_{n=0}^{∞} \frac{(-2y)^n}{n!}, \psi'(\frac{r + a + 2}{a} - \frac{γ}{2}) \]

\[ k(r - a + 1, γ, β, α) = \sum_{n=0}^{∞} \frac{(-2y)^n}{n!}, \psi'(\frac{r - a + a}{a} - \frac{γ}{2}) \]

Journal of critical reviews 527
PARAMETERS ESTIMATION:
Now, we shall show the estimate and derive the unknown parameters of new mixture distribution by using maximum likelihood method for type-one censoring sample.

**Progressively censoring data:**
The common name of this type of censored is progressively censored data or multiple censored data and can be called randomly censored data[6].

Different entering times to the study or experiment and different survival times occur in this type of censoring. The progressively censored samples usually come from the field, because subjects go into service at different times and have different running times when the data are recorded[7].

In most clinical and epidemiologic studies, the period of the study is fixed and the patients enter the study at different times during that period. Some may die before the end of the study, so their exact survival times are known. Others may withdraw before the end of the study and are lost to be followed-up, still the others may be alive at the end of the study[8].

The maximum likelihood Estimator method for progressively censoring data:
The main advantage of maximum likelihood properties method is that it contain the best affinity. That means when the data increases, the estimate converges better and faster with the parameter (parameters) in short, it maximizes the value of the function.

The likelihood function (joint probability density function) of \( t_1 < t_2 < \cdots < t_n \) is as follows:

\[
L = \prod_{i=1}^{n} \mathbb{P}(t_i; \theta) = \prod_{i=1}^{n} \left[ f(t_i; \theta) \delta_i \right]^{1-\delta_i}
\]

When \( C_i \) be the maximum time for which the \( i \)th units or patients can be observed and the survival times of the \( i \)th units or patients, \( t_i \) is unknown

\[
\delta_i = \begin{cases} 
1 & \text{if the } i \text{th units or patients fail or die} \\
0 & \text{if the } i \text{th units or patients fail or die}
\end{cases}
\]

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

\[
L = \prod_{i=1}^{n} \left[ (2y + \beta_1 + \alpha t_1^{-\alpha}) e^{-(2y + \beta_1 + \alpha t_1^{-\alpha})} \right]^{\delta_i} e^{-\sum_{i=1}^{n} (y(t_i - c_i))^{1-\delta_i}}
\]

Let \( \prod_{i=1}^{n} = k \)

Then we get:

\[
L = k \prod_{i=1}^{n} \left[ (2y + \beta_1 + \alpha t_1^{-\alpha}) e^{-(2y + \beta_1 + \alpha t_1^{-\alpha})} e^{-\sum_{i=1}^{n} (y(t_i - c_i))^{1-\delta_i}}
\]

\[
L = k \sum_{i=1}^{n} (2y + \beta_1 + \alpha t_1^{-\alpha}) \delta_i e^{-\sum_{i=1}^{n} (y(t_i - c_i))^{1-\delta_i}}
\]

After taking the logarithm of the two side of the equation we get:

\[
\ln L = n \ln k + \sum_{i=1}^{n} \delta_i \ln (2y + \beta_1 + \alpha t_1^{-\alpha}) - \sum_{i=1}^{n} \left( y(t_i - c_i) + \frac{\alpha}{2} \delta_i t_i + \frac{\gamma}{2} \delta_i t_i^2 \right)
\]

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

\[\frac{\partial \ln L}{\partial \alpha} = \sum_{i=1}^{n} \delta_i \left( \alpha t_i^{-\alpha-1} \ln t_i - \sum_{i=1}^{n} (1 - \delta_i) c_i^{-\alpha} \right)\]

\[\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{n} \delta_i \left( \beta t_i^{-\alpha} \ln t_i - \sum_{i=1}^{n} (1 - \delta_i) c_i^{-\alpha} \right)\]

\[\frac{\partial \ln L}{\partial \gamma} = \sum_{i=1}^{n} \delta_i \left( \gamma t_i^{-\alpha} \ln t_i - \sum_{i=1}^{n} (1 - \delta_i) c_i^{-\alpha} \right)\]

The three nonlinear equation (5), (6), (7) are difficult to solve, then numerical methods such as Newton-Raphson procedure is used to find the estimate values for these parameters.
\[
\frac{\partial f(\beta)}{\partial \beta} = \frac{\delta_i (a^\gamma t_i^\beta + b t_i^\gamma)}{(2y + \beta t_i + a t_i^\alpha)^2} - \sum_{i=1}^{n} \delta_i i(2y + \beta t_i + a t_i^\alpha)^2
\]

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

\[
f(\gamma) = \sum_{i=1}^{n} \frac{\delta_i (a^\alpha t_i^\beta + b t_i^\gamma)}{(2y + \beta t_i + a t_i^\alpha)^2} - \sum_{i=1}^{n} \frac{2\delta_i i(2y + \beta t_i + a t_i^\alpha)^2}{2y + \beta t_i + a t_i^\alpha}
\]

Then we stopped the Newten-Raphson method by using the following formula:

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

Finally, we estimate the death density function \( f(t) \), survival function \( \delta(t) \) and hazard function \( h(t) \) where:

\[
f(t) = (2\varphi + \beta t + \alpha t^\alpha - 1) e^{-\left(2\varphi + \beta t + \alpha t^\alpha\right)}
\]

\[
\delta(t) = e^{-\left(2\varphi + \beta t + \alpha t^\alpha\right)} h(t) = (2\varphi + \beta t + \alpha t^\alpha - 1)
\]

RESULTS AND DISCUSSION:

This study was based on samples and real data taken from the National Center For Research and Treatment of Hematology/University of Mustansiriya of Chronic Lymphocytic Leukemia which is one type of leukemia cancers that effects people aged from 60-80 years old which is symbolized as (CILL). The duration of the study was six months, equivalent to (184) days. The number of patients who participate in the study was (42) patients, (29) of them left the study and we did not know their fate because we were not able to follow up and (11) of them died during the study period.

Now, we use the chi-square test from goodness of fit which is one of agreeable test for model which is taken from restrict sample of study.Then if we determine whether the data (sample) corresponded with the mixture distribution on through the SPSS program. The null and alternative hypotheses for chi-square tests is:

\[
H_0 : \text{the data are distributed as a mixture distribution}
\]

\[
H_1 : \text{the data are not distributed as a mixture distribution}
\]

Then calculate chi-square=18.14 <tabulate chi-square=18.31 with level of significant (0.05) and degree of freedom equals (10), that means the data is distributed as new mixture distribution NMD

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

\[
a_0 = 1 , \beta_0 = 0.01 , \gamma_0 = 0.004
\]

After that, utilizing Matlab programeing we got the following estimated parameters values:

\[
\hat{\alpha} = 0.0022 , \hat{\beta} = 0.000067 , \hat{\gamma} = 0.0342
\]

Then evaluating the results for probability death density function, evaluating the results survival function and evaluating the results hazard rate function, then we put in the following table.

![Figure (1-1) Estimated probability function curves of progressively censored data](image)

From table (1-1) we can make the following comments:

1- that the value of \( f(t; a, \beta, \gamma) \) is decreasing when the life time is increasing.

2- the value of \( s(t; a, \beta, \gamma) \) is decreasing when the life time is increasing.

Journal of critical reviews
3- that the value of \( h(t; a, \beta, \gamma) \) is increasing when the life time is increasing.

CONCLUSIONS

In this paper we can derive the parameters of new mixture distribution which contains three parameters, two of them are scales and one of them is shape for progressively censoring data. Then we estimate the values of these parameters by using maximum likelihood estimator method. After that finding and estimating the values of probability death density function \( f(t) \), survival function \( \hat{s}(t) \) and hazard function \( \hat{h}(t) \). We note that each of both of the probability density function and the survival function is decreasing when the values of lifetime are increasing while the hazard function is increasing when the values of lifetime are increasing. Therefore, we can show that the behavior of three probability functions in figure(1-1).

Finally, the research presented a minimum database that provides the use of the new mixture distribution with some statistical methods and real data according to quantitative and qualitative measures that help in the study of a certain phenomenon of a simple societal sample in order to obtain some statistical results from the study population and thus is a planning tool to improve the quality of statistical information based on this distribution and progressively censoring data.

REFERENCES:

1. Mohammed, M. J. and Hussein I. H; "Some Estimation Methods for New Mixture Distribution with Simulation and Application [C]." Iop Conference Series: Materials Science and Engineering IOP Publishing. vol51, 2019.
2. Mohammed, M. J. and Hussein I. H; "Study of New Mixture Distribution[a]" Int., Conf. of Engineering Medicine and Applied Sciences ICEMASP in Turky. Vol14, pp:7566-7573,2018.
3. Mohammed, M. J. and Hussein I. H; "comparison between new mixture distribution and many other distributions[b]." jour adv research in dynamical and control systems;2018.
4. Nasiru, S; "Serial Weibull Rayleigh distribution: theory and application." International Journal of Computing Science and Mathematics 7(3): 239-244;2016.
5. Cordeiro, G. M., et al.; "The exponential-Weibull lifetime distribution." Journal of Statistical Computation and simulation 84(12): 2592-2606;2014.
6. Klein, J. P. and Moeschberger M. L; " Survival analysis: techniques forensored and truncated data", Springer Science & Business Media; vol 542;2005.
7. Tableman, M. and Kim J. S; "Survival analysis using S: analysis of time-to-event data", Chapman and Hall/CRC;2003.
8. Lee, E. T. and Wang, J; "Statistical methods for survival data analysis", John Wiley & Sons, vol 535;2003.
9. Khalili, M., Alimohammadi, N, Narges, K. "Iranian nurses' ethical challenges in controlling children's fever(2018) International Journal of Pharmaceutical Research, 10 (1), pp. 337-340. htpls://www.scopus.com/inward/record.uri?eid=2-s2.0-85056993661&partnerID=40&md5=54626f18e11e4563e892cdaa8be6d3
10. Pathirage Kamal Perera. "Traditional medicine-based therapies for cancer management." Systematic Reviews in Pharmacy 10.1 (2019): 90-92. Print. doi:10.5530/srp.2019.1.15

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

| Failure Time | \( f(t) \)       | \( \hat{s}(t) \)   | \( \hat{h}(t) \)   |
|-------------|------------------|-------------------|-------------------|
| 49          | 0.000845567637   | 0.011788482736    | 0.071728284026    |
| 51          | 0.000733818079   | 0.010211701239    | 0.071860512011    |
| 53          | 0.000636667801   | 0.008843483662    | 0.0719928735920   |
| 57          | 0.00047866391    | 0.006627180108    | 0.072257941327    |
| 62          | 0.000334917675   | 0.004613838873    | 0.072589807520    |
| 63          | 0.000311742179   | 0.004290645618    | 0.072656240387    |
| 65          | 0.000270036962   | 0.003709851422    | 0.072789158416    |
| 68          | 0.000217593611   | 0.002981197728    | 0.072988654669    |
| 71          | 0.000175229237   | 0.002394225443    | 0.073188277864    |
| 78          | 0.000105477257   | 0.001432054947    | 0.073654476767    |
| 82          | 0.000078803509   | 0.001066049062    | 0.07392109637     |

| 530 |