Global Journal of Modeling and Intelligent Computing (GJMIC)

Vol. 2, No. 1, January, 2022 ISSN: 2767-1917

A MIXTURE OF HYBRID TYPE DISTRIBUTIONS FOR THE SURVIVAL ANALYSIS OF COVID-19 VACCINE

Sarvesh Kumar, Indian Institute of Technology Roorkee, India (skumar@ma.iitr.ac.in)
Madhu Jain, Indian Institute of Technology Roorkee, India (madhu.jain@ma.iitr.ac.in)

ABSTRACT

This article proposes the estimation approach for a mixture type model to analyse the length of periods between the successive COVID-19 vaccines. The study can be used to predict the parameters used in hybrid type distributions of the vaccination strategy within a pre-determined period. This mixture model provides valuable information for obtaining an accurate estimate of the reliability and hazard rate of the vaccine. We implement a mixture of hybrid type distributions model by taking real-time data of Argentina and Brazil. The weibull, lognormal, and exponential distributions are used to compare reliability indices. The proposed model is validated by fitting it into an accurate data set taken from WHO site. Finally, we provide some graphs and tables by taking illustrations to demonstrate the computational tractability and goodness of fit of the model.

Keyword-: Mixture model, Covid-19 vaccine, Weibull, Lognormal, Exponential Distribution, Reliability Indices, Goodness of Fit.

1. INTRODUCTION

The COVID-19 was first noticed in Wuhan, Hubei province, China, in December 2019 and has been a primary human threat to nationwide health around the globe. These viruses directly concatenate the lungs, causing severe pneumonia and raising the death risk. Many countries have been affected by these diseases; Still, this virus is crucial because infectious disease sharply disperses in the population. This article mainly focuses on forecasting the reliable vaccine duration for covid-19. The mixed distributions of the time intervals between vaccine provides valuable insights about the specific period. For the forecasting purposes, several continuous distributions, such as Weibull, Lognormal, exponential distributions can be used to study the reliability function and survival analyses to predict the reliability indices. There are many real life problems where survival/reliability plays an important role. It is worthwhile to cite the eminent researchers' articles in different contexts concerning reliability function and hazard data. Rade (1989) and Sarhan (1993) considered reliability concepts for studying various mixture models. Rade (2002) proposed the reliability indices for a single, double and triple independent and identical component series and parallel systems. Sarhan (2005) used some survival functions to investigate the performance by determining the system reliability indices.

Many researchers have used two and three-fold mixture models to predict mean time to failure and other reliability indices. The mixture of distributions can be used for the study with a specific period. Jiang and Murthy (1995) characterized the two-fold Weibull mixed model using the parameter estimation on a given data set. Brennan (2006) developed the software mixture models and estimated the parameter values for the model. Peng (2013) established Bayesian estimation for the model to determine the unknown parameters. Ateya [2014) discussed the finite mixture of generalized exponential distributions to estimate the parameter values. Polymenis (2020) characterized the variety of exponential distributions for evaluating the reliability function and hazard rate for covid-19. These models are implemented by establishing the reliability function and hazard rates.. Welling et al. (2020) used Weibull distribution to enable the incubation period's imitation. The main objective of present article is to establish mixed hybrid type distributions and select the best possible model to represent the length of periods between the successive COVID-19 vaccines.

We propose mixture models for validating the distributed vaccines for covid-19. Using the maximum likelihood estimation (MLE), mean square error (MSE), and expectation maximization (EM) algorithm, we estimate the parameters of the mixed mixture model. The weibull, lognormal and exponential distributions are used in the mixture distribution model. We demonstrate that these models are robust for fitting the vaccines data sets. The rest of the paper is organized as follows. Section 2 presents preliminaries on weibull, lognormal, exponential distributions. k- fold mixture model is framed in section 3 to present some reliability indices such as mean time to failure (MTTF0, hazard function, etc. To develop 3-fold mixture distribution model, special distributions viz. weibull, lognormal and exponential distributions are used. Section 4 outlines the different parameter estimation approaches viz., EM, MSE, and MLE methods for estimating the parameters of mixture model. In section 5, the goodness of fit test approach is given. In section 6, numerical illustrations are provided to tests the mixture model on the observed data collected from the WHO site. Finally, in section 7, the scope of investigation done and future perspective are given.

2. PRELIMINARIES OF DISTRIBUTIONS

I. Weibull distribution (WD)

Weibull distribution (WD) can be used for the reliability prediction and other applications. Two parameters are mainly used to define this distribution. The probability density function (PDF) of WD with scale parameters "α" and shape parameters "β" is given by

$$f(t,\alpha,\beta)=\frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1}e^{-}\left(\frac{t}{\beta}\right)^{\alpha},\ t>0 \tag{1}$$

Mean time to failure (MTTF) of WD is

$$\text{MTTF}=\beta\Gamma\left(1+\frac{1}{\alpha}\right) \tag{2}$$

where $'\Gamma'$ denotes the gamma function.

Reliability function of WD is given by

$$R(t)=e^{-\left(\frac{t}{\beta}\right)^{\alpha}},\quad t>0 \tag{3}$$

Hazard function of WD is

$$H(t)=\frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1}, \tag{4}$$

II. Lognormal distribution (LD)

The lognormal distribution (LD) is mainly used for the critical survival of units whose failure modes are fatigue-stress. This distribution is mathematically defined in terms of two parameters, namely, mean "μ" and standard deviation "σ". The PDF and MTTF of lognormal distribution are given by

$$f(t,\mu,\sigma^2)=\frac{1}{t\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\log t-\mu}{\sigma}\right)^2},\ t>0 \tag{5}$$

$$\text{MTTF}=e^{\left(\mu+\frac{\sigma^2}{2}\right)} \tag{6}$$

The reliability and hazard functions of LD are

$$R(t)=\left\{1-\varphi\left[\frac{\log t-\mu}{\sigma}\right]\right\} \tag{7}$$

$$H(t)=\frac{\frac{1}{t\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\log t-\mu}{\sigma}\right)^2}}{\left\{1-\varphi\left[\frac{\log t-\mu}{\sigma}\right]\right\}}. \tag{8}$$

III. Exponential distribution (ED)

The exponential distribution (ED) can be used to study the hazard rate, reliability prediction and other applications. This distribution is mathematically defined in term of shape parameter $'\lambda'$ as

$$f(t,\lambda)=\frac{1}{\lambda}e^{-}\left(\frac{t}{\lambda}\right),\quad t>0 \tag{9}$$

$$\text{MTTF}=\lambda \tag{10}$$

$$R(t)=e^{-\left(\frac{t}{\lambda_1}\right)} \tag{11}$$

$$H(t)=\frac{\frac{1}{\lambda}e^{-}\left(\frac{t}{\lambda}\right)}{e^{-\left(\frac{t}{\lambda_1}\right)}} \tag{12}$$

3. MIXED K-FOLD MIXTURE MODELS

Now, we consider the three-component mixture of Weibull, lognormal and exponential (MWLE) distributions for the survival function and its reliability properties**.**

If $X_1, X_2, X_3, \dots, X_k$ be random variables and $f_i(t)$ denotes the PDF of $X_i, i = 1,2,3,\dots,k$ then PDF for the k- Fold mixture model is given by,

$$f(t)=\sum_{i=1}^{k} p_i\, f_i(t) \tag{13}$$

where p_i is mixing weight and $\sum_{i=1}^{k} p_i = 1$. For MWLE

$$f(t) = p_1 \frac{\alpha_1}{\beta_1} e^{-(\frac{t}{\beta_1})^{\alpha_1}} (\frac{t}{\beta_1})^{\alpha_1 - 1} + p_2 \frac{1}{\sigma_1 t\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\log t - \mu_1}{\sigma_1})^2} + p_3 \frac{1}{\lambda_1} \mathrm{e}^{-}(\frac{t}{\lambda_1}) \tag{14}$$

$$\text{MTTF}= p_1\beta_1\Gamma\left(1+\frac{1}{\alpha_1}\right) + p_2 e^{\mu_1+\frac{1}{2}\sigma_1{}^2} + p_3\lambda_1 \tag{15}$$

Reliability of MWLE model is

$$R(t)= p_1 e^{-(\frac{t}{\beta_1})^{\alpha_1}} + p_2\left\{1-\varphi\left[\frac{\log t-\mu_1}{\sigma_1}\right]\right\} + p_3 e^{-(\frac{t}{\lambda_1})}, \tag{16}$$

where φ is a cumulative density function of the lognormal distribution.

The hazard rate of (MWLE) is

$$H(t)=\frac{p_1\frac{\alpha_1}{\beta_1}e^{-(\frac{t}{\beta_1})^{\alpha_1}}(\frac{t}{\beta_1})^{\alpha_1-1} + p_2\frac{1}{\sigma_1 t\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{\log t-\mu_1}{\sigma_1})^2} + p_3\frac{1}{\lambda_1}\mathrm{e}^{-}(\frac{t}{\lambda_1})}{p_1 e^{-(\frac{t}{\beta_1})^{\alpha_1}} + (p_2)\left\{1-\varphi\left[\frac{\log t-\mu_1}{\sigma_1}\right]\right\} + p_3 e^{-(\frac{t}{\lambda_1})}} \tag{17}$$

4. PARAMETER ESTIMATION

Appling some modified parameter estimation approaches such as maximum likelihood estimation (MLE) and expectation-maximization (EM) methods, we estimate the parameter values for the distributions of mixture models (Marshall and Olkin, (1967)).

EM method for MWLE

For the mixture model based on a random sample, we use the EM and some other method for the estimation of the eight parameter values of MWLE. Now

$$\frac{\partial L}{\partial \boldsymbol{\theta}} = 0 \tag{18}$$

where $L=\prod_{j=1}^{m} f(t_j;\boldsymbol{\theta})$, $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8)$ (19)

$\theta_i = p_i$ for i=1,2,3.

$\theta_4 = \alpha, \theta_5 = \beta, \theta_6 = \mu, \theta_7 = \sigma, \theta_8 = \lambda.$

The likelihood function corresponding f (t) given in equation (12) is

$$L(\boldsymbol{\theta}) = \prod_{j=1}^{m}[p_1 f_1(t_j, \alpha_1, \beta_1) + p_2 f_2(t_j, \mu_1, \sigma_1) + p_3 f_3(t_j, \lambda_1)], \tag{20}$$

where $t_1, t_2, t_3, \dots, t_m$ are the collection of observations of m incomplete data and x_1, x_2, x_3 be a set of missing observations, where $x_{rj} = x_r(t_j) = 1$, if the observation belong to r^{th} component and otherwise 0 for r=1, 2, 3 and $j = 1, 2, \dots, m$. Now, we apply EM to the mixture model and find the $\boldsymbol{x_r}$, a missing value. The EM consists of mainly two steps, the first step is E-step and the second M-step.

Here x's variable is treated as missing observation in the Expectation step. Thus

$x_r = (x_{1j}, x_{2j}, x_{3j})$, are calculated by using of the conditional expectation E $(x_{lj} | t_j)$.

Then

$$\widetilde{x_{rj}} = E(x_{rj} | t_j) = \frac{p_r f_r(t_j, \boldsymbol{\theta}_r)}{\sum_{r=1}^{3} p_r f_r(t_j, \boldsymbol{\theta}_r)} \qquad r = 1, 2, 3. \tag{21}$$

The function E $(x_{rj} | t_j)$ is calculated in the E-step and M-step of the EM, under the normalize condition $p_1 + p_2 + p_3 = 1$. Then, we evaluate all-missing parameters and also do the estimation of parameters. For the complete- data, log-likelihood function is given by

$W = \log L(\boldsymbol{\theta})$

$$W = \sum_{j=1}^{m}\sum_{l=1}^{3} x_{rj} \log f_r(t_j, \boldsymbol{\theta}), \tag{22}$$

Partially differentiating equation (22) with respect to parameters, and after some algebraic manipulation, we get

$$\widetilde{p_r} = \sum_{j=1}^{m} \widetilde{x_{rj}}/n, \ r = 1, 2, 3. \tag{23}$$

$$\gamma = {\beta_1}^{\alpha_1} = \frac{\sum_{j=1}^{m} \widetilde{x_{1j}} {t_j}^{\alpha_1}}{\sum_{j=1}^{m} \widetilde{x_{1j}}} \tag{24}$$

$$\sum_{j=1}^{m} \widetilde{x_{1j}} - \sum_{j=1}^{m} \frac{1}{\gamma} \widetilde{x_{1j}} {t_j}^{\alpha_1} \{ \alpha_1 log\ t_j - \log \gamma \} + \widetilde{x_{1j}} \{ \alpha_1 \log t_j - \log \gamma \} = 0 \tag{25}$$

$$\mu_1 = \frac{\sum_{j=1}^{m} \widetilde{x_{2j}} \log t_j}{\sum_{j=1}^{m} \widetilde{x_{2j}}} \tag{26}$$

$${\sigma_1}^2 = \frac{\sum_{j=1}^{m} \widetilde{x_{2j}} (\log t_j - \mu_1)^2}{\sum_{j=1}^{m} \widetilde{x_{2j}}} \tag{27}$$

$$\lambda_1 = \frac{\sum_{j=1}^{m} \widetilde{x_{3j}} t_j}{\sum_{j=1}^{m} \widetilde{x_{3j}}} \tag{28}$$

Now, to implement the EM algorithm for the proposed mixed model, we start with initial value $\boldsymbol{\theta}^0$for the W given in (22) which is a linear function of the unobserved data x related to the concerned problem. Then, we calculate the parameter values by E- step and M-step. The M-step consists of substituting these $\widetilde{x_1}$ values for x_i in equations (23) to (28) and then evaluating the parameter values through E- step and M-step and iterate until convergence has achieved.

5. GOODNESS OF FIT (GOF) TESTS

These type of tests helps us to decide the best-fitted model on real data set. The goodness of fit test can be applied using statistical methods such Akaike's information criterion (AIC) and Bayesian information criterion (BIC) tests for the best fit of the mixture/ hybrid model by taking the survival data. We consider the mixture models described in the previous section to examine the goodness of fit.

In the present study, both AIC and BIC tests are written in the form [-2logL + UV], where L is the likelihood function, and V is the count of parameters. We consider U equal to 2 for AIC, and log(n) for BIC. To choose the best fitting mixed model of different types of distributions, these tests have been employed. The best fit has the smallest value of AIC.

6. NUMERICAL EXPERIMENTS

By taking the Realtime data from WHO site (https://covid19.who.int/table), the parameter estimation has been done and parameter values as well as AIC and BIC values are summarized in Tables 1 and 3. The reliability indices are calculated for the vaccinated people in Argentina and Brazil respectively, for the doses per hundred people. For vaccinated people in Argentina and Brazil respectively, Tables 2 and 4 present the mean time to the failure (MTTF) of hybrid type mixture models.

Illustration 1.

In this case, Argentina's vaccination data are taken and corresponding CDF and reliability function of fully vaccinated people in Argentina during Dec 30 2020 to Feb 03 2021 are displayed in figs 1 and 2. The data used from WHO site (https://covid19.who.int/table) shows the number of vaccine doses given to people vaccinated per hundred. The comparison of different models has been done for the reliability and cumulative distribution functions (CDF) of vaccinated people in Argentina.

Table 1: The estimated parameters, AIC and BIC values for the mixture models of vaccinated people in Argentina			
Mixture Model	**Estimate values**	**AIC**	**BIC**
Weibull and Lognormal	$\mu_1 = -0.9526, \sigma_1 = 0.7639$ $\alpha_1 = 3.0625, \beta_1 = 0.6416$ $p_1 = 0.5895, p_2 = 0.4105$	11.4100	16.0455
Lognormal and Exponential	$\mu_{12} = -0.6961, \sigma_{12} = 0.5379$ $\lambda_{12} = 1.9037, p_{12} = 0.5253$ $p_{22} = 0.4747$	13.0470	19.9569
Weibull and Exponential	$\alpha_{13} = 4.8125, \beta_{13} = 0.6586$ $\lambda_{13} = 2.2821, p_{13} = 0.6061$ $p_{23} = 0.3939$	6.2309	6.3248
Weibull, Lognormal and Exponential	$\mu_{14} = -0.8102, \sigma_{14} = 0.6359$ $\alpha_{14} = 5.0625, \beta_{14} = 0.6726$ $\lambda_{14} = 2.2553, p_{14} = 0.4241$ $p_{24} = 0.2970, p_{34} = 0.2789$	11.1348	17.3152

Table 2: Estimation of the mean time to failure (MTTF) of mixture models of vaccinated people in Argentina					
Models	Weibull and Lognormal	Lognormal and Exponential	Weibull and Exponential	Weibull, Lognormal and Exponential	Real data
Means	0.5501	0.5520	0.5394	0.5456	0.5394

Illustration 2.

In this case, the data of Brazilian people vaccinated per hundred during Jan 29, 2021 to March 05, 2021 are taken from WHO site (https://covid19.who.int/table).The CDF and PDF for different mixture models and are shown in figures 3 and 4**.**

Table 3: The estimated parameter values, AIC and BIC values for the models of vaccinated people in Brazil			
Mixture Model	**Estimate value**	**AIC**	**BIC**
Weibull and Lognormal	$\mu_1 = -1.7165, \sigma_1 = 1.4538$ $\alpha_1 = 1.4186, \beta_1 = 0.5567$ $p_1 = 0.6158, p_2 = 0.3842$	16.5360	19.4455
Lognormal and Exponential	$\mu_{12} = -1.5647, \sigma_{12} = 1.3676$ $\lambda_{12} = 2.0016, p_{12} = 0.3870$ $p_{22} = 0.6130$	19.5319	21.9564
Weibull and Exponential	$\alpha_{13} = 1.3125, \beta_{13} = 0.5046$ $\lambda_{13} = 2.3202, p_{13} = 0.5052$ $p_{23} = 0.4948$	14.2607	16.6853
Weibull, Lognormal and Exponential	$\mu_{14} = -1.7163, \sigma_{14} = 1.4481$ $\alpha_{14} = 1.4375, \beta_{14} = 0.5446$ $\lambda_{14} = 2.1395, p_{14} = 0.3843$ $p_{24} = 0.2464, p_{34} = 0.3693$	20.6585	24.5378

Table 4. Estimation of the mean time to failure (MTTF) of mixture models of vaccinated people in Brazil					
Models	Weibull and Lognormal	Lognormal and Exponential	Weibull and Exponential	Weibull, Lognormal and Exponential	Real data
Means	0.5105	0.5125	0.4482	0.4890	0.5492

7. CONCLUSIONS

The composite distribution has some advantages, and can be fitted to the specific data. It would be advantageous to apply the mixed distributions to study the vaccination program of disease covid-19 and thus resolve the health issues. This study has investigated the mixed composition of Weibull, lognormal and exponential distributions to study the reliability indices of vaccinated people against COVID-19. The parameter estimation and goodness of fit are used. It is seen that the proposed hybrid model provides a better fit for the real-time data set and can be further used to decrease the death rates. The proposed model provides insights for a high reliable vaccine by the parameter estimation of the survival function

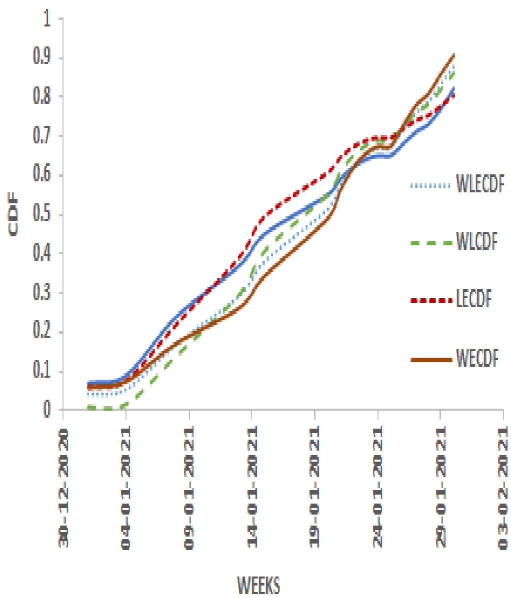

Figure 1: CDF of mixture models for real data of Fully vaccinated people in Argentina during Dec 30 2020 to Feb 03 2021

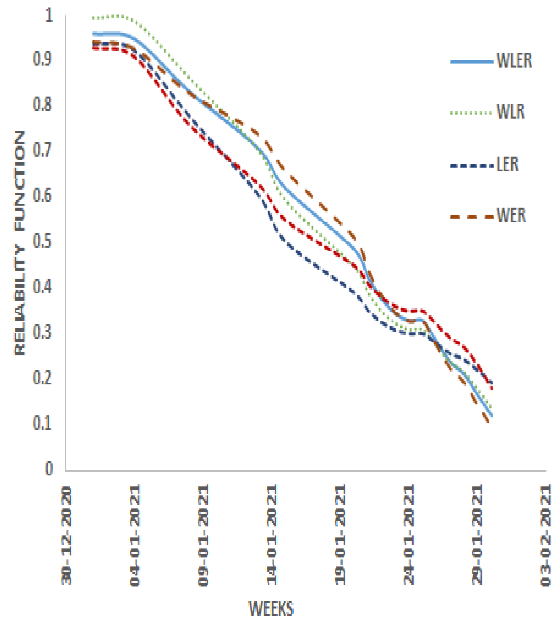

Figure 2: Reliability of mixture models for real data of Fully vaccinated people in Argentina during Dec 30 2020 to Feb 03 2021

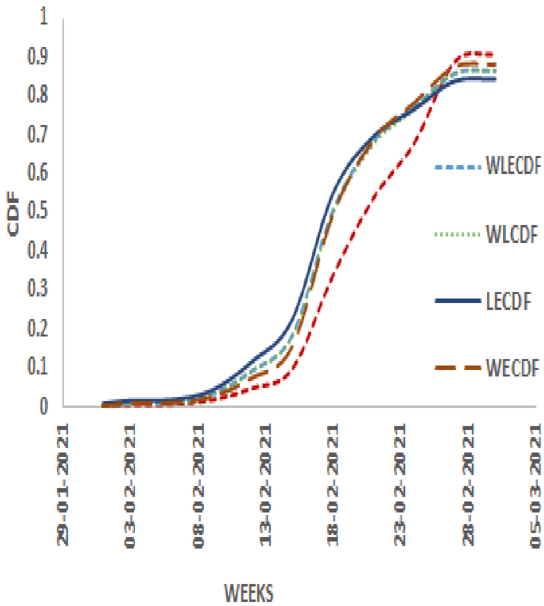

Figure 3: CDF of mixture models for real data of Fully vaccinated people in Brazil during Jan 29 2021 to March 05 2021

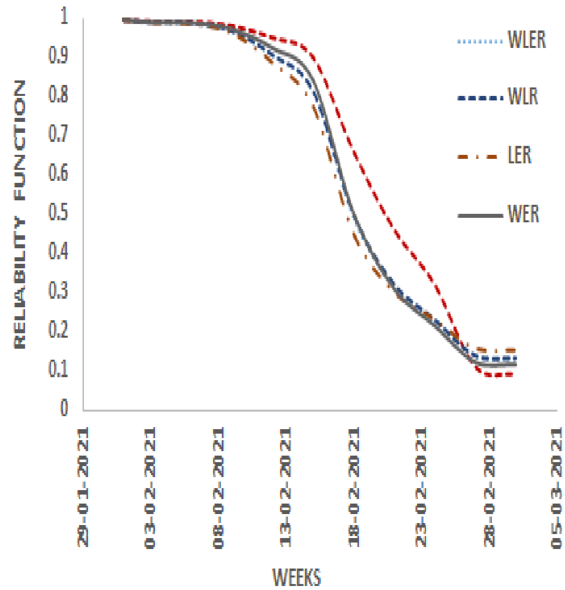

Figure 4: Reliability of mixture models for real data of Fully vaccinated people in Brazil during Jan 29 2021 to March 05 2021

REFERENCES

Arshad, M.Z., Iqbal, M.Z. & Ahmad, M. (2018). Transmuted exponentiated moment Pareto distribution, Open Journal of Statistics 8, 939–961.

Ateya, S. F. (2014). Maximum likelihood estimation under a finite mixture of generalized exponential distributions based on censored data. Statistical Papers 55, 311-325.

Elmahdy, E. E. & Abdallah, W. A. (2013). A new approach for parameter estimation of finite Weibull mixture distribution for reliability modeling. Applied Mathematical Modelling 37, 1800- 1810.

Elmahdy, E. E. (2015). A new approach for Weibull modeling for reliability life data analysis. Applied Mathematics & Computation 250, 708–720.

Jiang, R. & Murthy, DNP. (1995). Modeling failure data by mixture of two Weibull distribution: A graphical approach. IEEE Transactions of Reliability 44(5), 477-487.

Marshall, A. & Olkin, W. I. (1967). A multivariate exponential distribution. Journal of the American Statistical Association 62, 30-44.

Nie, K. Sinha, B. K. & Hedayat, A. S. (2017). Unbiased estimation of reliability function from a mixture of two exponential distributions based on a single observation. Statistics and Probability Letters 127, 7-13.

Polymenis, A. (2020). An application of a mixture of exponential distributions for assessing hazard rates from COVID-19. Journal of Population Therapeutics. & Clinical. Pharmacology 27, 58-63.

Peng, X. & Yan, Z. (2013). Bayesian estimation for generalized exponential distribution based on progressive Type- 1 interval censoring. Acta Mathematica Applicatae Sinica (English) 29(2), 391-402.

Rade, L. (1993). Reliability equivalence. Microelectron Reliability 33, 323-325.

Ruhi, S. S. & Karim, M. R. (2015). Mixture models for analyzing product reliability data case study. Springer Plus 4 634.

Sarhan, A. M. (2002). Reliability equivalence with a basic series / parallel system. Applied Mathematics and Computation 132, 115-133.

Sarhan, A. M. (2005). Reliability equivalence factors of a parallel system. Reliability Engineering & System Safety 87, 405-411.

Sahoo, P. (2008). Probability and Mathematical Statistics. University of Louisville, Louisville, KY USA 712 40292.

Shi, Y. Zhenzhou, L. & Huang, Z. (2020). Time-dependent reliability-based design optimization with probabilistic and interval uncertainties. Applied Mathematical Modelling 80, 268-289.

Zhang, Z. & Gui, W. (2019). Statistical inference of reliability of generalized Rayleigh distribution under progressively type. Journal of Computational and Applied Mathematics 361, 295-312.