SHORT COMMUNICATION

Shifted Mixture Model Using Weibull, Lognormal, and Gamma Distributions

Sarvesh Kumar[1] · **Madhu Jain**[1]

Abstract The objective of this investigation is to provide framework to construct a threefold mixture model and its shifted version using Weibull, lognormal, and gamma distributions. The proposed models are examined by establishing the statistical and reliability indices. The parameter estimation using the maximum likelihood estimation method (MLE) and expectation–maximization has been proposed. The usefulness of the shifted mixture models by fitting them into the actual data set has revealed. The goodness-of-fit tests are used to compare the mixture models for the real-life data. Based on statistical testing, it is established that for small data set, shifted mixture model is the best fitted model in comparison with other single and mixed mixture distributions.

Keywords Shifted mixture distribution · MLE · EM algorithm reliability indices · Goodness of fit

Introduction

In recent years, some research articles in reliability literature have appeared to investigate a variety of new mixture models. Using mixed distributions, appropriate modeling of various complex systems can be done. The prediction of various statistical and reliability indices can be made with the help of mixture distributions, particularly when the hazard rate

function has some critical shapes. Many researchers have studied two- and threefold mixture models for analyzing the reliability indices using real data sets. Jiang and Murthy [1] characterized the twofold Weibull mixture model in terms of the probability density function (PDF) and cumulative density function (CDF) for plotting the curves by fitting to the real data set. Sarhan [2] used some survival functions to study the performance of the mixture model and provided the result for the reliability indices of the system. Ateya [3] investigated a finite mixed mixture of generalized exponential distribution and estimated the parameter values using the EM algorithm based on maximum likelihood estimation (MLE) in a general form of right-censored failure data. Atienza et al. [4] explored the combination of lognormal distributions to examine the congealment performance of the concerned mixture model. Nedjar and Zeghdoudi [5] discussed the mixture model to study its various properties and simulations of gamma Lindley distribution. Nie et al. [6] considered an unbiased estimation of some reliability indices based on a mixture model of two exponential distributions with known mixing weight functions. Kumar et al. [7] used the mixed mixture distribution using Weibull, lognormal, and gamma distributions to estimate the reliability indices using real data sets. Kumar and Jain [8] presented the survival analysis of COVID-19 vaccine by using the mixture distribution.

Maximum likelihood estimation is widely used for the parameter estimation due to their noble statistical properties. However, it is not easy to find a closed-form solution using the MLE method in particular when the numbers of variables are large and some data are missing. The expectation–maximization (EM) algorithm can be used for the parameter estimation in such cases. Martin and Tatjana [9] considered the mixed models for the best validation of left-truncated real-time data using the combination of gamma,

✉ Sarvesh Kumar
skumar@ma.iitr.ac.in

Madhu Jain
madhu.jain@ma.iitr.ac.in

1 Department of Mathematics, Indian Institute of Technology Roorkee, Roorkee 247667, India

lognormal, and Weibull distribution, respectively. Kumar et al. [10] proposed a mixture model to investigate the reliability indices by using the combination of homogeneous and non-homogeneous continuous distributions. For parameter estimation, they have implemented the MLE and EM methods.

The mixture distributions and its shifted version are more flexible in comparison with single distributions and can be well fitted to a wide variety of applications including survival analysis. The proposed shifted model provides adequate fitting of survival function having enormous real-life applications in a variant scenario encountered in biology and medicine, engineering and technology, genetics and healthcare, geology and agriculture, business and marketing, economics and social sciences, etc. In some real-time system, the failure may not occur in the starting as such shifted distribution can be used to represent the survival function of real-time observations. The main objectives to develop mixed mixture model in this article are twofold. Firstly, we introduce the shifted mixed distribution (SMWLG) using Weibull, gamma, and lognormal distributions. Secondly, we will estimate the parameters of the proposed models applying EM methods for the mixed distributions. The estimation of the parameter values by EM methods can be used for the model for which some data are missing. The estimation of eighteen parameters for the shifted mixed distribution (SMWLG) will be of taken into account to study the reliability indices. To demonstrate the usefulness of the underlying shifted mixture model, we shall use some real data set and examine whether it is suitable and effective for data modeling and survival analysis.

Preliminaries on Distributions

Here, we describe specific two-parameter Weibull, lognormal, and gamma distributions for developing the mixture models.

Weibull Distribution (WD)

The probability density function (PDF) of WD with shape parameters 'α' and scale parameters 'β' is

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

The mean time to failure (MTTF) of WD is

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

Reliability of WD is given by

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

Hazard function of WD is

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

Lognormal Distribution (LD)

The PDF and MTTF of two-parameter lognormal distribution are given by

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

and

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

Reliability function of LD is

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

where φ is the CDF of normal distribution with *mean* 'μ' *and* standard deviation 'σ,' respectively.

Hazard function of LD is

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

Gamma Distribution (GD)

The PDF and MTTF of two-parameter gamma distributions with shape parameter 'a' and scale parameter 'b' are defined as follows:

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

and

$$\text{MTTF} = ab. \tag{10}$$

Reliability function of GD is

$$R(t) = \left[1 - \frac{1}{\Gamma(a)}\left\{\left(\frac{t}{b}\right)^a \sum_{m=1}^{\infty} \frac{\left(\frac{t}{b}\right)^m}{m!(a+m)}\right\}\right] \tag{11}$$

Hazard function of GD is

$$h(t) = \frac{\frac{t^{a-1}}{\Gamma(a)b^a} e^{-\left(\frac{t}{b}\right)}}{\left\{ 1 - \frac{1}{\Gamma(a)} \left(\frac{t}{b}\right)^a \sum_{m=1}^{\infty} \frac{\left(\frac{t}{b}\right)^m}{m!(a+m)} \right\}} \quad (12)$$

Mixture of Weibull, Lognormal, and Gamma Distributions

Now, we consider the three-component mixture of Weibull, lognormal, and gamma distributions and obtain the survival function and reliability indices.

If $X_1, X_2, and X_3$ be the random variables, and $f_i(t)$ be the PDF of $X_i, i = 1, 2, 3$, respectively, then PDF of threefold model is given by Jayant and Sudha [11]. $f(t) = \sum_{i=1}^{3} p_i f_i(t)$, where p_i is mixing weights and $\sum_{i=1}^{3} p_i = 1$.

To develop threefold mixture of Weibull, lognormal, and gamma distributions (MWLG), we use the shape and scale parameters $(\alpha_1, \beta_1) and (a_1, b_1)$, respectively, for the Weibull and gamma distributions. Let μ and σ be the mean and standard deviation for the normal distribution. Thus, the PDF of MWLG is

$$f_M(t) = p_1 \frac{\alpha_1}{\beta_1} e^{-\left(\frac{t}{\beta_1}\right)^{\alpha_1}} \left(\frac{t}{\beta_1}\right)^{\alpha_1 - 1}$$
$$+ p_2 \frac{1}{\sigma_1 t \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log t - \mu_1}{\sigma_1}\right)^2} + p_3 \frac{t^{a_1-1}}{\Gamma(a_1)b_1^{a_1}} e^{-\left(\frac{t}{b_1}\right)} \quad (13)$$

The MTTF for MWLG model is given by

$$\text{MTTF}_M = p_1 \beta_1 \Gamma\left(1 + \frac{1}{\alpha_1}\right) + p_2 e^{\mu_1 + \frac{1}{2}\sigma_1^2} + p_3 a_1 b_1 \quad (14)$$

$$R_M(t) = p_1 e^{-\left(\frac{t}{\beta_1}\right)^{\alpha_1}} + p_2 \left\{ 1 - \varphi\left(\frac{\log t - \mu_1}{\sigma_1}\right) \right\}$$
$$+ p_3 \left[1 - \frac{1}{\Gamma(a_1)} \left\{ \left(\frac{t}{b_1}\right)^{a_1} \sum_{m=1}^{\infty} \frac{\left(\frac{t}{b_1}\right)^m}{m!(a_1 + m)} \right\} \right] \quad (15)$$

The hazard rate of MWLG model is

$$h_M(t) = \frac{p_1 \frac{\alpha_1}{\beta_1} e^{-\left(\frac{t}{\beta_1}\right)^{\alpha_1}} \left(\frac{t}{\beta_1}\right)^{\alpha_1-1} + p_2 \frac{1}{\sigma_1 t \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log t - \mu_1}{\sigma_1}\right)^2} + p_3 \frac{t^{a_1-1}}{\Gamma(a_1)b_1^{a_1}} e^{-\left(\frac{t}{b_1}\right)}}{p_1 e^{-\left(\frac{t}{\beta_1}\right)^{\alpha_1}} + p_2 \left\{1 - \varphi\left[\frac{\log t - \mu_1}{\sigma_1}\right]\right\} + p_3 \left\{1 - \frac{1}{\Gamma(a_1)}\left(\frac{t}{b_1}\right)^{a_1} \sum_{m=1}^{\infty} \frac{\left(\frac{t}{b_1}\right)^m}{m!(a_1+m)}\right\}} \quad (16)$$

The cumulative hazard rate function of MWLG model is

$$H_M(t) = \int_0^t h_M(x)\mathrm{d}x \quad (17)$$

Reversed hazard rate function is

$$(\text{RH})_M(t) = f_M(t)/F_M(t) \quad (18)$$

Mills ratio of MWLG model is

$$\text{MR}_M(t) = R_M(t)/f_M(t) \quad (19)$$

Shifted Mixture Model

Let $Y_1, Y_2, and Y_3$ be the random variables representing the shifted Weibull, lognormal, and gamma distributions (SMWLG) with location parameters $\tau_1, \tau_2, and \tau_3$, respectively. The shifted mixture of Weibull, lognormal, and gamma (SMWLG) distributions is formed by using $g_1(t), g_2(t), and g_3(t)$ as PDF of Weibull, lognormal, and gamma distributions, and the respective shifted PDF are denoted by $g_4(t), g_5(t), and g_6(t)$. The scale and shape parameters for Weibull (gamma) distribution are denoted by $\alpha_2 (a_2)$ and $\beta_2 (b_2)$, respectively. Furthermore, for SMWLG, we denote the mean and standard deviation of normal distribution by μ_2 and σ_2, respectively. Thus,

$$f_{\text{SM}}(t) = q_1 g_1(t) + q_2 g_2(t) + q_3 g_3(t) + q_4 g_4(t - \tau_1)$$
$$+ q_5 g_5(t - \tau_2) + q_6 g_6(t - \tau_3) \tau_1, \tau_2, \tau_3 > 0 \quad (20)$$

Here q_i is mixing weight and $\sum_{i=1}^{6} q_i = 1$.

Now, we get

$$f_{\text{SM}}(t) = q_1 \frac{\alpha_2}{\beta_2} \left(\frac{t}{\beta_2}\right)^{\alpha_2 - 1} e^{-\left(\frac{t}{\beta_2}\right)^{\alpha_2}}$$
$$+ q_2 \frac{1}{\sigma_2 t \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log t - \mu_2}{\sigma_2}\right)^2} + q_3 \frac{t^{a_2-1}}{\Gamma(a_2)b_2^{a_2}} e^{-\left(\frac{t}{b_2}\right)}$$
$$+ q_4 \frac{\alpha_3}{\beta_3} \left(\frac{t - \tau_1}{\beta_3}\right)^{\alpha_3 - 1} e^{-\left(\frac{t - \tau_1}{\beta_3}\right)^{\alpha_3}}$$
$$+ q_5 \frac{1}{\sigma_3 (t - \tau_2)\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log(t - \tau_2) - \mu_3}{\sigma_3}\right)^2}$$
$$+ q_6 \frac{(t - \tau_3)^{a_3-1}}{\Gamma(a_3)b_3^{a_3}} e^{-\left(\frac{t - \tau_3}{b_3}\right)} \quad (21)$$

MTTF for SMWLG is

$$
\begin{aligned}
\mathrm{MTTF}_{\mathrm{SM}} =& q_1\beta_2\Gamma\left(1 + \frac{1}{\alpha_2}\right) + q_2 e^{\mu_2 + \frac{1}{2}\sigma_2^2} \\
& + q_3 a_2 b_2 + q_4\left(\beta_3\Gamma\left(1 + \frac{1}{\alpha_3}\right) + \tau_1\right) \\
& + q_5 e^{\mu_3(1-\tau_2) + \frac{1}{2}\sigma_3^2(1-\tau_2)^2} + q_6(a_3 b_3 + \tau_3)
\end{aligned}
\tag{22}
$$

The reliability function is

$$
\begin{aligned}
R_{\mathrm{SM}}(t) =& q_1 e^{-\left(\frac{t}{\beta_2}\right)^{\alpha_2}} + q_2\left\{1 - \varphi\left[\frac{\log t - \mu_2}{\sigma_2}\right]\right\} \\
& + q_3\left\{1 - \frac{1}{\Gamma(a_2)}\left(\frac{t}{b_2}\right)^{a_2}\sum_{m=1}^{\infty}\frac{\left(\frac{t}{b_2}\right)^m}{m!(a_2 + m)}\right\} + q_4 e^{-\left(\frac{t-\tau_1}{\beta_3}\right)^{\alpha_3}} \\
& + q_5\left\{1 - \varphi\left[\frac{\log(t - \tau_2) - \mu_3}{\sigma_3}\right]\right\} \\
& + q_6\left\{1 - \frac{1}{\Gamma(a_3)}\left(\frac{t - \tau_3}{b_3}\right)^{a_3}\sum_{m=1}^{\infty}\frac{\left(\frac{t-\tau_3}{b_3}\right)^m}{m!(a_3 + m)}\right\}
\end{aligned}
\tag{23}
$$

The hazard function is obtained using

$$
h_{\mathrm{SM}}(t) = f_{\mathrm{SM}}(t)/R_{\mathrm{SM}}(t) \tag{24}
$$

The cumulative hazard rate of SMWLG model is

$$
H_{\mathrm{SM}}(t) = \int_0^t h_{\mathrm{SM}}(x)\mathrm{d}x \tag{25}
$$

Reversed hazard rate function is

$$
(\mathrm{RH})_{\mathrm{SM}}(t) = f_{\mathrm{SM}}(t)/F_{\mathrm{SM}}(t) \tag{26}
$$

The Mills ratio of SMWLG model is

$$
\mathrm{MR}_{\mathrm{SM}}(t) = R_{\mathrm{SM}}(t)/f_{\mathrm{SM}}(t) \tag{27}
$$

Remarks When $\tau_1 = \tau_3 = \tau_3 = 0$, the shifted mixture model (SMWLG) deduces to mixture model (MWLG) composed of Weibull, lognormal, and gamma distributions.

Parameter Estimation

For parameter estimation, we use the EM method cf. Zhang and Barnes [12] based on a random sample of size m. The estimation of the eighteen parameters (θ) of SMWLG given in Eq. (21) is performed as follows:

The likelihood function corresponding to Eq. (21) is given by

$$
\begin{aligned}
L(\boldsymbol{\theta}) = \prod_{j=1}^m [& q_1 g_1(t_j, \boldsymbol{\theta}_1) + q_2 g_2(t_j, \boldsymbol{\theta}_2) + q_3 g_3(t_j, \boldsymbol{\theta}_3) \\
& + q_4 g_4(t_j, \boldsymbol{\theta}_4) + q_5 g_5(t_j, \boldsymbol{\theta}_5) + q_6 g_6(t_j, \boldsymbol{\theta}_6)]
\end{aligned}
\tag{28}
$$

The MLE equation yields

$$
\frac{\partial \log L(\boldsymbol{\theta})}{\partial \theta_i} = 0, i = 0, 1, 2, 3, 4, 5, 6, \tag{29}
$$

where $\boldsymbol{\theta}_0 = (q_1, q_2, q_3, q_4, q_5, q_6)$,
$\boldsymbol{\theta}_1 = (\alpha_2, \beta_2), \boldsymbol{\theta}_2 = (\mu_2, \sigma_2), \boldsymbol{\theta}_3 = (a_2, b_2)$,
$\boldsymbol{\theta}_4 = (\alpha_3, \beta_3), \boldsymbol{\theta}_5 = (\mu_3, \sigma_3)$, and $\boldsymbol{\theta}_6 = (a_3, b_3)$.
Let us assume that $t_j, j = 1, 2, 3, \ldots, m$ represents a collection of 'm' incomplete data and y_1, y_2, y_3 be the missing data, where $y_{rj} = y_r(t_j) = 1$, if the data value belongs to r^{th} missing data and otherwise 0 for $r = 1, 2, 3, 4, 5, 6$, and $j = 1, 2, \ldots, m$. Now, we apply EM to the shifted mixture model and find the 'y_r' which are the missing values.

Here, $\boldsymbol{y_{rj}} = (y_{1j}, y_{2j}, y_{3j}, y_{4j}, y_{5j}, y_{6j})$ are evaluated by using of the conditional expectation $E(y_{rj}|t_j)$. Now,

$$
\widetilde{y}_{rj} = E(y_{rj}|t_j) = \frac{q_r g_r(t_j, \boldsymbol{\theta}_r)}{\sum_{r=1}^6 q_r g_r(t_j, \boldsymbol{\theta}_r)} \quad, r = 1, 2, 3, 4, 5, 6.
\tag{30}
$$

The functions $E(y_{rj}|t_j)$ are evaluated in the E-step and M-step. Now, we evaluate the parameters (θ) and respective parameter values. For SMWLG model, log-likelihood function log L (θ) is defined by Zhang and Gui [13].

$$
\log \mathrm{L}(\boldsymbol{\theta}) = \sum_{j=1}^m \sum_{i=1}^6 \widetilde{y}_{ji}\log g_i(t_j, \boldsymbol{\theta}_i) \tag{31}
$$

Partially differentiating Eq. (31) with respect to all parameters, we get

$$
\widetilde{q}_r = \sum_{j=1}^m \widetilde{y}_{rj}/m \quad r = 1, 2, 3, 4, 5, 6. \tag{32}
$$

$$
\sum_{j=1}^m \widetilde{y}_{1j} - \sum_{j=1}^m \frac{1}{\gamma}\widetilde{y}_{1j}t_j^{\alpha_2}\{\alpha_2 \log t_j - \log \gamma\} + \widetilde{y}_{1j}\{\alpha_2 \log t_j - \log \gamma\} = 0 \tag{33}
$$

$$
\gamma = \beta_2^{\alpha_2} = \frac{\sum_{j=1}^m \widetilde{y}_{1j}t_j^{\alpha_2}}{\sum_{j=1}^m \widetilde{y}_{1j}} \tag{34}
$$

$$\sum_{j=1}^{m} \widetilde{y_{2j}} - \sum_{j=1}^{m} \frac{1}{\gamma} \widetilde{y_{2j}} (t_j - \tau_1)^{\alpha_3} \{\alpha_3 log(t_j - \tau_1) - \log \gamma\}$$
$$+ \widetilde{y_{2j}} \{\alpha_3 \log(t_j - \tau_1) - \log \gamma\} = 0 \qquad (35)$$

$$\gamma = \beta_3^{\alpha_3} = \frac{\sum_{j=1}^{m} \widetilde{y_{2j}} (t_j - \tau_1)^{\alpha_3}}{\sum_{j=1}^{m} \widetilde{y_{2j}}} \qquad (36)$$

$$\mu_2 = \frac{\sum_{j=1}^{m} \widetilde{y_{3j}} \log t_j}{\sum_{j=1}^{m} \widetilde{y_{3j}}} \qquad (37)$$

$$\sigma_2^2 = \frac{\sum_{j=1}^{m} \widetilde{y_{3j}} (\log t_j - \mu_2)^2}{\sum_{j=1}^{m} \widetilde{y_{3j}}} \qquad (38)$$

$$\mu_3 = \frac{\sum_{j=1}^{m} \widetilde{y_{4j}} \log(t_j - \tau_2)}{\sum_{j=1}^{m} \widetilde{y_{4j}}} \qquad (39)$$

$$\sigma_3^2 = \frac{\sum_{j=1}^{m} \widetilde{y_{4j}} (\log(t_j - \tau_2) - \mu_2)^2}{\sum_{j=1}^{m} \widetilde{y_{4j}}} \qquad (40)$$

$$b_2 = \frac{\sum_{j=1}^{m} (\breve{y}_{5j}) t_j}{a_2 \sum_{j=1}^{m} (\breve{y}_{5j})} \qquad (41)$$

$$b_3 = \frac{\sum_{j=1}^{m} (\breve{y}_{6j})(t_j - \tau_2)}{a_3 \sum_{j=1}^{m} (\breve{y}_{6j})} \qquad (42)$$

$$\frac{\Gamma'(a_2)}{\Gamma(a_{2)}} \sum_{j=1}^{m} \breve{y}_{5j} + \sum_{j=1}^{m} \breve{y}_{5j} \{\log \delta_2 - \log a_2\} - \sum_{j=1}^{m} \breve{y}_{5j} \log t_j = 0 \qquad (43)$$

$$\frac{\Gamma'(a_3)}{\Gamma(a_3)} \sum_{j=1}^{m} \breve{y}_{6j} + \sum_{j=1}^{m} \breve{y}_{6j} \{\log \delta_3 - \log a_3\} - \sum_{j=1}^{m} \breve{y}_{6j} \log(t_j - \tau_3) = 0 \qquad (44)$$

where $\delta_2 = \frac{\sum_{j=1}^{m} \breve{y}_{5j} t_j}{\sum_{j=1}^{m} \breve{y}_{5j}}$ and $\delta_3 = \frac{\sum_{j=1}^{m} \breve{y}_{6j}(t_j - \tau_3)}{\sum_{j=1}^{m} \breve{y}_{6j}}$. Here dash '$\prime$' denotes the derivative w.r.t. parameters a_2 and a_3 in Eqs. (43) and (44), respectively.

To implement the EM algorithm for the proposed mixed model SMWLG, we start with the initial value of θ for the logL(θ) used in (31). Here, logL(θ) is a linear function of the unobserved data 'y' for the concerned problem. Now, we calculate the parameter values by E-step and M-step involved in EM method. The M-step consists of substituting these $\widetilde{y}_{rj}$ values for in Eqs. (32)–(44), and again, we evaluate parameter values through E-step and M-step and iterate until convergence has been achieved. The value for parameter estimation and testing indices using EM is given in Table 1. The goodness-of-fit test can be applied using statistical methods. Now, we compute Akaike's information criterion (AIC), consistent Akaike's information criterion (CAIC), and Bayesian information criterion (BIC) tests for best fit of models by taking the survival data. The least value in the CAIC column of Table 1 is displayed in bold letters.

In the present study, AIC, BIC, and CAIC tests are written in the form [-2logL + UV], where L is the likelihood function, and V is the count of parameters. U is equal to 2 for AIC and log(n) for BIC. For small n, the corrected version of AIC is CAIC = AIC + 2 V*(V + 1)/(n-V-1). To choose the best fitting mixed distribution for the competing mixture models, CAIC test is more appropriate when sample size is small, and the best fit has the minimum CAIC value [14].

Numerical Simulation

For validating the proposed mixture models, we consider the real data set and perform the statistical tests for the goodness of fit. We consider the data of ith (i = 1,2,3, …,12) failures (t_i) for a system until the time of 12th failures as given below

Table 1 Parameter estimation values and statistical indices for real-time data using by EM algorithm

Mixture model	Estimate value	LL	AIC	CAIC	BIC
Weibull	$\alpha = 1.0625, \beta = 84.9063$	− 65.0035	134.0071	135.3404	134.9769
Lognormal	$\mu = 3.8649, \sigma = 1.2488$	− 66.0728	136.1457	137.479	137.1155
MWL	$\mu_0 = 3.7277, \sigma_0 = 1.2356, \alpha_0 = 1.0838, \beta_0 = 90.3207, p = .5221$	− 68.9812	149.9624	166.7624	152.7818
Gamma	$a = 1.0162, b = 81.6768$	− 65.0226	134.0451	135.3784	135.0149
MWLG	$\mu_1 = 3.7277, \sigma_1 = 1.2356, \alpha_1 = 1.0838\ \beta_1 = 90.3207, a_1 = 1.0313,$ $b_1 = 83.3136, p_1 = 0.3425, p_2 = 0.3163, p_3 = 0.3412$	− 65.1890	148.378	238.378	152.7421
SMWLG	$\mu_2 = 3.6959, \sigma_2 = 1.2605, \alpha_2 = 1.0938, \beta_2 = 89.6811,$ $a_2 = 0.9325, b_2 = 90.4873, \mu_3 = 3.7876, \sigma_3 = 1.2232, \alpha_3 = 0.9325,$ $\beta_3 = 83.5153, a_3 = 0.9663, b_3 = 85.9120, q_1 = 0.1711, q_2 = 0.1581,$ $q_3 = 0.1707, q_4 = 0.1737, q_5 = 0.1519, q_6 = 0.1745$	− 65.1569	166.3138	**68.5995**	175.0422

cf. Rigdon and Basu [15] $t_i = 3, 9, 20, 25, 41, 50, 69, 91, 128, 151, 182, 227$.

According to the coagulation of different mixture models, we consider the SMWLG for the estimation of parameters on the real data sets. The values of goodness-of-fit indicators AIC, CAIC, and BIC are computed for all the models. The corresponding parameter estimation and statistical indices are presented in Table 1 for the data set. For small size sample, the lowest value of CAIC demonstrates that the respective mixture model could be chosen as the best model is comparison to other mixture models. For the given data, we notice that the mixture model SMWLG performs better than mixture model MWLG in terms of goodness of fit due to the fact that the indicator CAIC is lowest for this model.

For the different models, the trends of PDF, CDF, R(t), and h(t) are shown in Fig. 1a–d. Figure 2a and b display the reversed hazard rate function and MR, respectively.

Concluding Remarks

This study has presented a comparative study of threefold mixture model with the corresponding shifted mixture models by taking the combination of Weibull, lognormal, and gamma distributions. To validate the proposed model for the
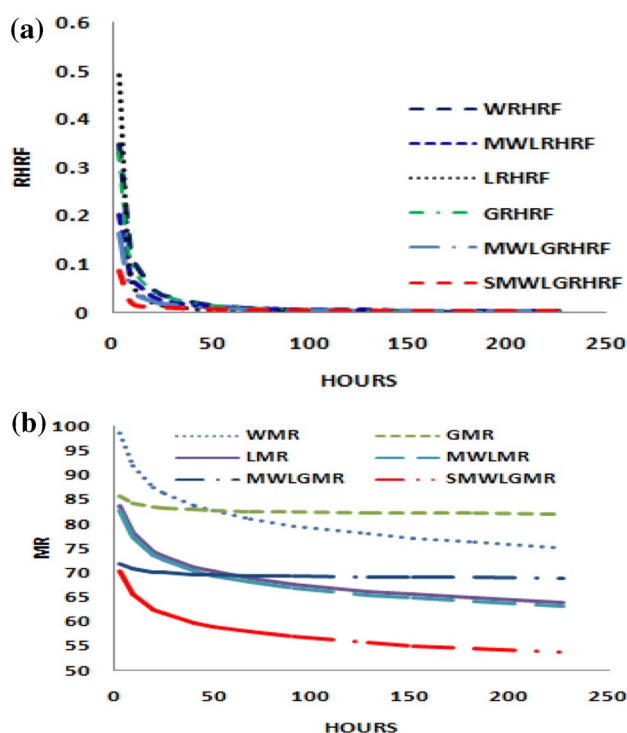

Fig. 2 **a** RHRF for real data set and **b** MR for real data set
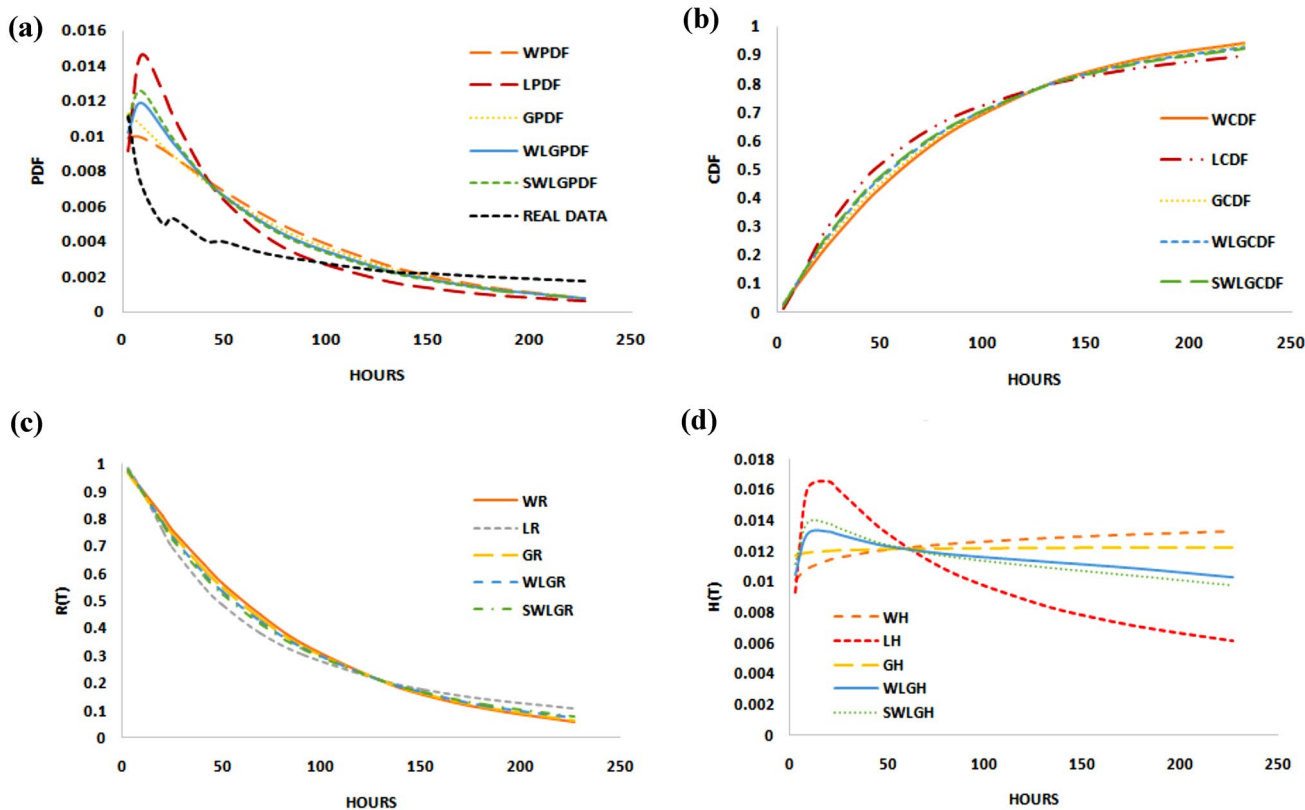

Fig. 1 **a** Probability density function (PDF), **b**: CDF for real data set, **c** reliability function $R(t)$, and **d** hazard rate function for real data set

real-time data, we have performed the parameter estimation using MLE and EM methods. The goodness-of-fit tests used to test the threefold mixture model and respective threefold shifted model distributions facilitate the choice of the best mixed model. The data analysis of the shifted mixture distribution in comparison with classical mixed statistical distributions can facilitate a meaningful fit to survival function in life time data analysis. The proposed model showed that the survival mixture model is flexible and maintains the feature of pure classical survival models and facilitates better option to model heterogeneous survival data.

Based on numerical experiment, we can conclude that shifted mixture model is best one in terms of CAIC value. The investigation can be further modified for three-parameter distributions or using memory-based properties. The usefulness of the underlying shifted mixture model which involves the location parameters lies in the fact that it outperforms in comparison with simple single and mixture models for survival analysis. The three-parameter finite mixture distributions (FMDs) can be developed in the future due to its rich flexibility but parameter estimation may become tedious job for the same.

Declarations

Conflict of interests   The authors declare that they have no conflict of interest.

References

1. Jiang R, Murthy DNP (1995) Modeling failure data by mixture of two Weibull distributions: a graphical approach. IEEE Trans Reliab 44(5):477–487
2. Sarhan AM (2005) Reliability equivalence factors of a parallel system. Reliab Eng Syst Saf 87:405–411
3. Ateya SF (2014) Maximum likelihood estimation under a finite mixture of generalized exponential distributions based on censored data. Stat Pap 55:311–325
4. Atienza N, Garcia J, Munoz-Pichardo J, Villa R (2014) Applications of mixture distributions in modulation of length of hospital stay. Stat Med 27(9):1403–1420
5. Nedjar S, Zeghdoudi H (2016) on gamma Lindley distribution: properties and simulations. J Comput Appl Math 298:167–174
6. Nie K, Sinha KB, Hedayat AS (2017) Unbiased estimation of reliability function from a mixture of two exponential distributions based on a single observation. Stat Probab Lett 127:7–13
7. Kumar S, Jain M, Gangopadhay A (2021) A mixture model of the mixed distributions to analyze some reliability indices and survival data. In: Proceedings - 26th international conference on reliability and quality in design (RQD), international society of science and applied technologies (ISSAT), Florida USA, pp 245–250
8. Kumar S, Jain M (2022) A mixture of hybrid type distributions for the survival analysis of COVID-19 vaccine. Glob J Model Intell Comput 2(1):11–18
9. Martin B, Tatjana M (2019) on modeling left- truncated loss data using mixture of distributions. Math Econ 85:35–46
10. Kumar S, Jain M, Gangopadhay A (2022) Reliability indices using mixture distribution homogeneous and non-homogeneous continuous distribution. J Life Cycle Reliab Saf Eng 11:302–312
11. Jayant VD, Sudha GP (2005) Life time data statistical models and methods. Ser Qual Reliab Eng Stat 11:71–77
12. Zhang X, Barnes S (2019) Lognormal based mixture models for robust fitting of hospital length of stay distributions. Oper Res Health Care 22:100–184
13. Zhang Z, Gui W (2019) Statistical inference of reliability of generalized Rayleigh distribution under progressively type. J Comput Appl Math 361:295–312
14. Nohut S (2020) Three-parameter (3P) Weibull distribution for characterization of strength of ceramics showing R-Curve behavior. Ceram Int 47:2270–2279
15. Rigdon SE, Basu AP (2000) Statistical methods for the reliability of repairable systems. Wiley, Hoboken