

# Finite Mixture Modeling via Skew-Laplace Birnbaum–Saunders Distribution

Mehrdad Naderi<sup>1\*</sup>, Mahdieh Mozafari<sup>2</sup>, Kheirolah Okhli<sup>3</sup>

<sup>1</sup> Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa

<sup>2</sup> Department of Statistics, Faculty of Mathematics and Computing, Higher Education Complex of Bam, Bam, Iran

<sup>3</sup> Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran

## ARTICLE INFO

### Article History

Received 16 Sep 2017

Accepted 19 Oct 2018

### Keywords

Birnbaum–Saunders distribution  
Normal mean-variance mixture  
model

Skew-Laplace distribution

Finite mixture model

ECM algorithm

### 2000 Mathematics Subject

Classification: 62E10, 65C60,  
62E15.

## ABSTRACT

Finite mixture model is a widely acknowledged model-based clustering method for analyzing data. In this paper, a new finite mixture model via an extension of Birnbaum–Saunders distribution is introduced. The new mixture model provide a useful generalization of the heavy-tailed lifetime model since the mixing components cover both skewness and kurtosis. Some properties and characteristics of the model are derived and an expectation and maximization (EM)-type algorithm is developed to compute maximum likelihood estimates. The asymptotic standard errors of the parameter estimates are obtained via offering an information-based approach. Finally, the performance of the methodology is illustrated by considering both simulated and real datasets.

© 2020 The Authors. Published by Atlantis Press SARL.

This is an open access article distributed under the CC BY-NC 4.0 license (<http://creativecommons.org/licenses/by-nc/4.0/>).

## 1. INTRODUCTION

The Birnbaum–Saunders (BS) distribution [1,2] is a positively skewed and unimodal distribution with non-negative support. The BS distribution has recently received considerable attention in the literature of statistics including lifetime, survival and environmental data analysis (see, e.g. [3–6]). An important property of the BS distribution is that it is closely related to the normal distribution by means of a simple stochastic representation. A random variable  $T$  is said to have a BS distribution with the shape and scale parameters  $\alpha$  and  $\beta$ , respectively, if it is generated by

$$T = \frac{\beta}{4} \left[ \alpha Z + \sqrt{(\alpha Z)^2 + 4} \right]^2, \quad (1)$$

where  $Z$  follows a standard normal distribution (denoted by  $N(0, 1)$ ). It can be easily shown that the probability density function (pdf) of  $T$  is

$$f(t; \alpha, \beta) = A(t, \alpha, \beta) \phi(a(t, \alpha, \beta)), \quad t > 0, \alpha > 0, \beta > 0, \quad (2)$$

where  $\phi(\cdot)$  is the pdf of the standard normal distribution,  $a(t, \alpha, \beta) = (\sqrt{t/\beta} - \sqrt{\beta/t})/\alpha$ , and  $A(t, \alpha, \beta)$  is the derivative of  $a(t, \alpha, \beta)$  with respect to  $t$ . A comprehensive review of the BS distribution can be found in [7].

Recently, some generalizations and extensions of the BS distribution have been proposed through replacing the standard normal variable  $Z$  in (1) by other random variables or replacing  $\phi(\cdot)$  in (2) by other pdfs. The main motivation for introducing these extended distributions is to increase the range of skewness and kurtosis and is to make the BS distribution more flexible, since the BS distribution may not fit the data very well [8]. Therefore, the skewed and heavy-tailed family of distributions are used by several researchers to enhance the flexibility of the BS distribution against skewness and outliers. For instance by introducing skew-normal distribution and its extensions such as skew- $t$  distribution [9], the related BS versions of them are proposed (see [10–12] among others).

\*Corresponding author. Email: m.naderi@up.ac.za

Another important family of skewed distributions is the class of normal mean-variance (NMV) mixture models [13]. Concerning the asymmetric properties of the NMV models, Arslan [14] introduced the skew-Laplace (SL) distribution via considering the inverse gamma distribution as a mixing random variable. Arslan also developed the expectation and maximization (EM) algorithm [15] for computing maximum likelihood (ML) estimates.

In reliability studies, data often arise from heterogeneous populations and so they need to be modeled by a mixture of two or more life distributions. The finite mixture of distributions (FM) is a finite convex linear combination of distribution functions which is also useful to approximate complicated probability densities presenting multimodality. A comprehensive survey of the FM models can be found in [16,17]. Because of the usefulness of the FM models in reliability, different kinds of the lifetime mixture models have been recently proposed. Ali [18] introduced a mixture of the inverse Rayleigh model for lifetime study in engineering processes and studied different properties of the proposed model. Balakrishnan *et al.* [19] proposed three different mixture models based on the BS distribution as (a) mixture of two different BS distributions, (b) mixture of a BS distribution and a length-biased version of another BS distribution and (c) mixture of a BS distribution and its length-biased version.

Due to the asymmetric properties and robustness of the SL distribution, the main objective of this paper is to propose a  $g$  components FM model via an extension of the BS distribution based on the SL model. Studying some properties of the new model, the EM-type algorithm is implemented to facilitate the procedure of the ML estimation. It also becomes possible to compute asymptotic standard errors of ML estimation via using the information-based idea [20]. Finally, by a simulation study, we show that the proposed model is robust for analyzing heavy tail lifetime data.

The remainder of the paper is unfolded as follows. In Section 2, we briefly review the univariate SL distribution and some of its properties. We also present the skew-Laplace Birnbaum–Saunders (SLBS) distribution with its characteristics in this section. Section 3 describes the finite mixture model of the proposed SLBS distribution and develops the EM-type algorithm for estimating parameters. In Section 4, we analyze a real data set for the purpose of illustrating the performance of the proposed mixture model. Two simulation studies are carried out to verify the robustness of the new model and to check finite-sample properties of the ML estimates in Section 5. Finally, the paper is closed with a short summary in Section 6.

## 2. PRELIMINARIES

A random variable  $X$  is said to have a univariate SL distribution (denoted by  $X \sim SL(\mu, \sigma, \lambda)$ ) if its pdf is given by

$$f_{SL}(x; \mu, \sigma, \lambda) = \frac{1}{2\tau\sigma} \exp \left\{ -\tau \frac{|x - \mu|}{\sigma} + \frac{\lambda(x - \mu)}{\sigma^2} \right\}, \quad x \in \mathbb{R},$$

where  $\sigma > 0$ ,  $\mu$  and  $\lambda \in \mathbb{R}$  are the scale, location and skewness parameters, respectively, and  $\tau = \sqrt{1 + (\lambda/\sigma)^2}$ . It can be easily shown that the SL distribution is generated through the NMV representation. Specifically, let  $Z$  and  $W$  be two independent random variables followed by  $N(0, 1)$  and the exponential distribution with mean 2 ( $Exp(0.5)$ ), respectively. Then, the random variable  $X \sim SL(\mu, \sigma, \lambda)$  admits the stochastic representation

$$X \stackrel{d}{=} \mu + W\lambda + \sigma\sqrt{W}Z. \quad (3)$$

**Proposition 2.1.** *The skewness and kurtosis of  $X \sim SL(\mu, \sigma, \lambda)$  are respectively given by*

$$\gamma_x = \frac{16\lambda^3 + 12\lambda}{(4\lambda^2 + 2)^{1.5}}, \quad \text{and} \quad \kappa_x = \frac{36\lambda^4 + 36\lambda^2 + 6}{(2\lambda^2 + 1)^2} - 3.$$

**Proof.** The results are straightforward by using representation (3). □

Figure 1 displays the curves of skewness and kurtosis of the SL distribution. It can be observed that the SL distribution takes wider ranges of skewness and kurtosis as compared with the skew-normal distribution. In the following, the notation  $f_{SL}(\cdot; \lambda)$  stands for the pdf of the standard SL distribution ( $\mu = 0, \sigma = 1$ ), denoted by  $X \sim SL(\lambda)$ .

Replacing  $Z$  by  $X \sim SL(\lambda)$  in (1), the SLBS distribution with the following pdf can be introduced

$$f_{SLBS}(t; \alpha, \beta, \lambda) = A(t, \alpha, \beta) f_{SL}(a(t, \alpha, \beta); \lambda), \quad t, \alpha, \beta > 0, \quad \lambda \in \mathbb{R}. \quad (4)$$

This model will be denoted by  $SLBS(\alpha, \beta, \lambda)$  henceforth. The stochastic representation of the SLBS distribution can be readily obtained by (1) and (3) as

$$T \stackrel{d}{=} \frac{\beta}{4} \left[ \alpha(W\lambda + \sqrt{W}Z) + \sqrt{(\alpha(W\lambda + \sqrt{W}Z))^2 + 4} \right]^2,$$

where  $Z \sim N(0, 1)$  and  $W \sim Exp(0.5)$ , independently.

**Theorem 2.1.** Some properties of the SLBS distribution are as follows:

1. The density of SLBS distribution tends to the density of Laplace BS distribution, as  $\lambda \rightarrow 0$ .
2. The random variable  $T$  distributed by  $SLBS(\alpha, \beta, \lambda)$  is degenerated to  $\beta$  as  $\alpha$  tends to zero.
3. If  $T \sim SLBS(\alpha, \beta, \lambda)$ , then  $X \stackrel{d}{=} \frac{1}{\alpha} \left[ \sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}} \right] \sim SL(\lambda)$ .
4. Let  $T \sim SLBS(\alpha, \beta, \lambda)$ . It can be easily shown that the cumulative distribution function (cdf) of  $T$  is  $F_{SLBS}(t; \alpha, \beta, \lambda) = F_{GH}(a(t, \alpha, \beta); 1, 0, 1, 0, 1, \lambda)$ , where  $F_{GH}(\cdot; \kappa, \chi, \psi, \mu, \sigma, \lambda)$  represents the cdf of generalized hyperbolic distribution with parameter  $(\kappa, \chi, \psi, \mu, \sigma, \lambda)$ . This leads to readily obtain the hazard rate function of  $T$  as

$$H(t) = \frac{f_{SLBS}(t; \alpha, \beta, \lambda)}{1 - F_{GH}(a(t, \alpha, \beta); 1, 0, 1, 0, 1, \lambda)}. \quad (5)$$

Using the “ghyp” package in the statistical software R and part (4) of Theorem 2.1, the hazard rate function of  $T$  can be computed. Figure 2 presents the hazard function curves of (5) for some parameter values. As a result of the fact that the hazard function of the proposed distribution can be decreasing, increasing, upside-down bathtub, it is evident that the proposed three-parameter distribution is quite flexible and can be used effectively in analyzing data with monotone as well as non-monotone hazard shape which are quite common in reliability and biological studies, for instance.

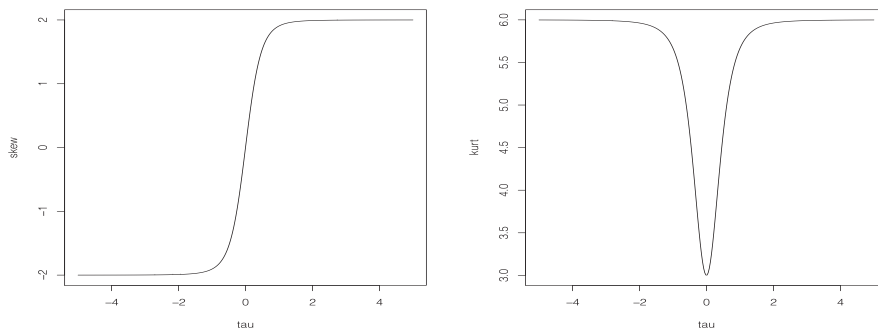
Now in the following theorem, we present a transformational result for the random variable  $T$  with SLBS distribution. This theorem can be useful in proposing a regression model with errors followed by the SLBS distribution.

**Theorem 2.2.** Let  $T \sim SLBS(\alpha, \beta, \lambda)$ . The random variable  $Y = \log(T)$  has a following pdf:

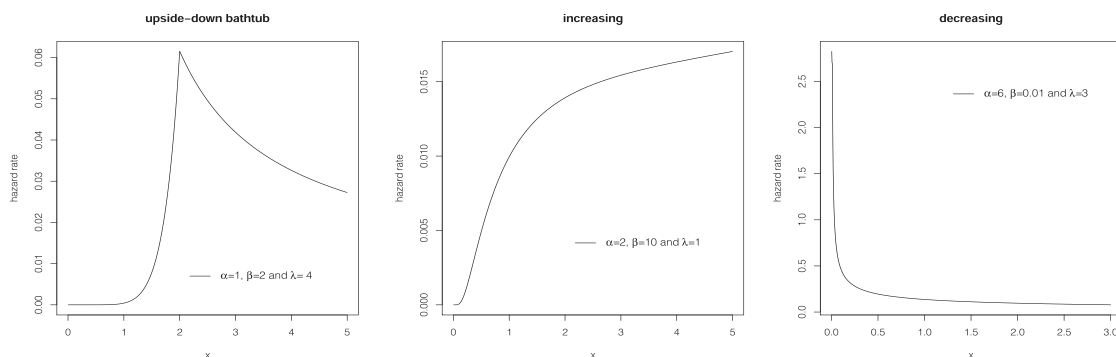
$$f_Y(y; \alpha, \gamma, \lambda, \tau) = U(y, \alpha, \gamma) f_{SL}(u(y, \alpha, \gamma); \lambda),$$

where  $\gamma = \log(\beta)$ ,  $U(y, \alpha, \gamma) = 2 \cosh\left(\frac{y - \gamma}{2}\right) / \alpha$  and  $u(y, \alpha, \gamma) = \sinh\left(\frac{y - \gamma}{2}\right) / \alpha$ .

**Remark 2.1.** As a measure of the uncertainty or randomness of a system, the entropy plays an important role in many sciences. The entropy has been used in the various situations and its numerous measures have been studied and compared in the literature. For a random variable



**Figure 1** | The skewness and kurtosis plots of the skew-Laplace (SL) distribution.



**Figure 2** | The hazard rate plots of the skew-Laplace Birnbaum-Saunders (SLBS) distribution for some parameter values.

$X$ , the Rényi entropy which is important in quantum information, and Shannon entropy which plays a similar role as the kurtosis measure in comparing the shapes of various densities and measuring heaviness of tails, are respectively defined by

$$\mathfrak{R} = (1 - \delta)^{-1} \log \left( \int f(x)^\delta dx \right), \quad \text{and} \quad \mathfrak{S} = -E[\log(f(x))],$$

where  $\delta > 0$ ,  $\delta \neq 1$  and  $f(x)$  represents the pdf of  $X$ . The Rényi and Shannon entropies of  $T \sim \text{SLBS}(\alpha, \beta, \lambda)$  can be expressed as

$$\begin{aligned} \mathfrak{R} &= \frac{-\delta}{1-\delta} (\log(2) - \delta \log(1 + \lambda^2)/2) + (1 - \delta)^{-1} \\ &\quad \times \log \left( \int_0^\infty A(t; \alpha, \beta)^\delta \exp\{-\tau \delta |a(t; \alpha, \beta)| + \lambda \delta a(t; \alpha, \beta)\} dt \right), \\ \mathfrak{S} &= -\log(4\alpha\tau\sqrt{\beta}) + 2\lambda^2 + \tau - 1.5E[\log(T)] + E[\log(T + \beta)]. \end{aligned}$$

There is no a closed-form expression for the integral and expectations. Hence, Theorems 2.1 and 2.2 are useful to compute them numerically.

Consequently, we establish the following proposition and theorem, which are useful for obtaining the complete log-likelihood function and for the calculation of some conditional expectations involved in the proposed EM-type algorithm discussed in the next section.

**Proposition 2.2.** *Let  $T$  be the random variable following the  $\text{SLBS}(\alpha, \beta, \lambda)$  distribution. The hierarchical representation of  $T$  is*

$$\begin{aligned} T | W = w &\sim \text{EBS}(\alpha\sqrt{w}, \beta, 2, -\sqrt{w}\lambda), \\ W &\sim \text{Exp}(0.5), \end{aligned}$$

where extended Birnbaum-Saunders distribution (EBS) denotes the extended BS distribution [21].

**Theorem 2.3.** *Let  $W \sim \text{Exp}(0.5)$  and  $T \sim \text{SLBS}(\alpha, \beta, \lambda)$ . For every  $t \in \mathbb{R}$ , the conditional distribution of  $W$  given  $T = t$  is the generalized inverse Gaussian (GIG) distribution,  $\text{GIG}(0.5, a^2(t, \alpha, \beta), 1 + \lambda^2)$ . Moreover,*

$$E[W^r | T = t] = \left( \frac{a(t, \alpha, \beta)}{\sqrt{1 + \lambda^2}} \right)^r R(\sqrt{1 + \lambda^2} a(t, \alpha, \beta), 0.5, r), \quad r = \pm 1, \pm 2, \dots,$$

where  $R(c, a, b) = K_{a+b}(c)/K_a(c)$ ,  $K_\kappa(\cdot)$  denotes the modified Bessel function of the third kind with index  $\kappa$ .

**Proof.** By Bayes' rule, the conditional distribution is simply obtained. Moreover, the conditional expectation can be derived by using properties of the GIG distribution and the Bessel function, see [22].  $\square$

### 3. FINITE MIXTURE OF THE SLBS DISTRIBUTIONS

Consider  $n$  independent, random variables  $T_1, \dots, T_n$ , which are taken from a mixture of SLBS (Mix-SLBS) distributions. The pdf of  $g$ -component Mix-SLBS model is

$$f(t_j; \Theta) = \sum_{i=1}^g \pi_i f_{\text{SLBS}}(t_j; \theta_i), \quad j = 1, 2, \dots, n, \quad (6)$$

where  $\pi_i$ 's are mixing proportions subject to  $\sum_{i=1}^g \pi_i = 1$ ,  $f_{\text{SLBS}}(\cdot; \theta_i)$  is the density of SLBS distribution (4) with  $\theta_i = (\alpha_i, \beta_i, \lambda_i)$  and  $\Theta = (\pi_1, \dots, \pi_{g-1}, \theta_1, \dots, \theta_g)$ . By observing data  $\mathbf{t} = (t_1, \dots, t_n)^\top$ , the observed log-likelihood function can be obtained as

$$\ell(\Theta | \mathbf{t}) = \sum_{j=1}^n \log \left( \sum_{i=1}^g \pi_i f_{\text{SLBS}}(t_j; \theta_i) \right). \quad (7)$$

The ML estimate of the parameters involved in (7) can be obtained directly. However, maximization of (7) is complicated. Another approach for getting ML estimator is the EM algorithm. In this approach, the idea is to solve tractable complete log-likelihood problems repeatedly instead of solving a difficult incomplete log-likelihood problem. For applying this approach to Mix-SLBS model, it is convenient to construct a log-likelihood function by introducing a set of allocation variables  $Z_j = (Z_{1j}, \dots, Z_{gj})$  for  $j = 1, \dots, n$ , taking  $Z_{ij} = 1$  if  $y_j$  belongs to

the  $i$ th component and  $Z_{ij} = 0$  otherwise. This implies that  $\mathbf{Z}_j$  independently follows a multinomial distribution with one trial and probabilities  $(\pi_1, \dots, \pi_g)$ , denoted by  $\mathbf{Z}_j \sim M(1; \pi_1, \dots, \pi_g)$ . It also follows from Proposition 2.2 that the hierarchical formulation of (6) can be represented by

$$\begin{aligned} T_{ij} | (W = w_j, Z_{ij} = 1) &\sim EBS(\alpha_i \sqrt{w_j}, \beta_i, 2, -\lambda_i \sqrt{w_j}), \\ W_j | Z_{ij} = 1 &\sim \text{Exp}(0.5), \\ \mathbf{Z}_j &\sim M(1; \pi_1, \dots, \pi_g). \end{aligned}$$

Therefore, the complete data log-likelihood function for  $\Theta$  associated with the observed variable  $\mathbf{t}$  and hidden variables  $\mathbf{w} = (w_1, \dots, w_n)^\top$  and  $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_n)^\top$ , omitting additive constants, is

$$\ell_c(\Theta | \mathbf{t}, \mathbf{w}, \mathbf{z}) = \sum_{j=1}^n \sum_{i=1}^g z_{ij} \left\{ \log(\pi_i \alpha_i) + \log\left(\frac{t_j + \beta_i}{\sqrt{\beta_i}}\right) - \frac{w_j^{-1}}{2\alpha_i^2} \delta(t_j, \beta_i) - \frac{\lambda_i^2}{2} w_j + \frac{\lambda_i}{\alpha_i} \xi(t_j, \beta_i) \right\}, \quad (8)$$

where  $\delta(t, \beta) = (t/\beta - \beta/t - 2)$ , and  $\xi(t, \beta) = a(t, 1, \beta)$ .

### 3.1. Parameter Estimation via ECM Algorithm

To compute the ML estimate of the unknown parameters involved in (8), we adopt the Expectation Conditional Maximization (ECM) algorithm [23]. The ECM algorithm is a variant of the EM algorithm with the maximization (M) step of EM replaced by a sequence of computationally simpler conditional maximization (CM) steps. The ECM algorithm for ML estimation of the Mix-SLBS proceeds as follows:

- E-step: At the iteration  $k$ , we compute the so-called  $Q$ -function, defined as the expected value of complete data log-likelihood (8) with  $\Theta$  evaluated at  $\hat{\Theta}^{(k)}$

$$Q(\Theta | \hat{\Theta}^{(k)}) = \sum_{j=1}^n \sum_{i=1}^g \hat{z}_{ij}^{(k)} \left\{ \log(\pi_i \alpha_i) + \log\left(\frac{t_j + \beta_i}{\sqrt{\beta_i}}\right) - \frac{\hat{u}_{ij}^{(k)}}{2\alpha_i^2} \delta(t_j, \beta_i) - \frac{\lambda_i^2}{2} \hat{w}_{ij}^{(k)} + \frac{\lambda_i}{\alpha_i} \xi(t_j, \beta_i) \right\}, \quad (9)$$

where  $\hat{u}_{ij}^{(k)} = E[W_j^{-1} | t_j, \hat{\theta}_i^{(k)}]$  and  $\hat{w}_{ij}^{(k)} = E[W_j | t_j, \hat{\theta}_i^{(k)}]$  calculated by using (7) in Theorem 2.3.

- CM-steps: Put  $n_i = \sum_{j=1}^n \hat{z}_{ij}^{(k)}$ ,  $A_i = \sum_{j=1}^n \hat{z}_{ij}^{(k)} \hat{w}_{ij}^{(k)}$ ,  $B_i = \sum_{j=1}^n \hat{z}_{ij}^{(k)} \hat{u}_{ij}^{(k)}$ ,  $R_i = \sum_{j=1}^n \hat{z}_{ij}^{(k)} \hat{u}_{ij}^{(k)} t_j$  and  $S_i = \sum_{j=1}^n \hat{z}_{ij}^{(k)} \hat{w}_{ij}^{(k)} / t_j$  and update  $\hat{\Theta}^{(k)}$  by maximizing (9) over  $\Theta$ . This leads to the following CM estimators:

$$\begin{aligned} \hat{\pi}_i^{(k+1)} &= \frac{n_i}{n}, \quad \hat{\lambda}_i^{(k+1)} = \frac{1}{A_i} \sum_{j=1}^n \hat{z}_{ij}^{(k)} a(t_j; \hat{\alpha}_i^{(k)}, \hat{\beta}_i^{(k)}), \\ \hat{\alpha}_i^{2(k+1)} &= \frac{R_i}{n_i \hat{\beta}_i^{(k)}} + \frac{\hat{\beta}_i^{(k)} S_i}{n_i} - 2 \frac{B_i}{n_i} - \frac{1}{A_i n_i} \left( \sum_{j=1}^n \hat{z}_{ij}^{(k)} \xi(t_j, \hat{\beta}_i^{(k)}) \right)^2, \\ \hat{\beta}_i^{(k+1)} &= \arg \max_{\beta} \ell_{obs}(\beta; \hat{\alpha}_i^{(k+1)}, \hat{\lambda}_i^{(k+1)}), \end{aligned}$$

where

$$\ell_{obs}(\beta; \hat{\alpha}_i^{(k+1)}, \hat{\lambda}_i^{(k+1)}) = \sum_{j=1}^n \hat{z}_{ij}^{(k)} \left\{ \log\left(\frac{t_j + \beta_i}{\sqrt{\beta_i}}\right) - \frac{\hat{u}_{ij}^{(k)}}{2\hat{\alpha}_i^{2(k+1)}} \delta(t_j, \beta_i) + \frac{\hat{\lambda}_i^{(k+1)}}{\hat{\alpha}_i^{(k+1)}} \xi(t_j, \beta_i) \right\}.$$

The above procedure is iterated until a suitable convergence rule is satisfied. To avoid an indication of lack of progress of the algorithm [24], we recommend adopting the Aitken acceleration [25] method as the stopping criterion. To apply this approach, the asymptotic estimate of the log-likelihood [16] is computed by

$$\ell_{\infty}(\hat{\Theta}^{(k+1)}) = \ell(\hat{\Theta}^{(k+1)}) + \frac{1}{1 - a^{(k)}} \{ \ell(\hat{\Theta}^{(k+1)}) - \ell(\hat{\Theta}^{(k)}) \}.$$

where the Aitken acceleration factor is  $a^{(k)} = \left( \ell(\hat{\Theta}^{(k+1)}) - \ell(\hat{\Theta}^{(k)}) \right) / \left( \ell(\hat{\Theta}^{(k)}) - \ell(\hat{\Theta}^{(k-1)}) \right)$ . Therefore, the ECM algorithm can be considered to have reached convergence if  $\ell_{\infty}(\hat{\Theta}^{(k+1)}) - \ell(\hat{\Theta}^{(k)}) < \varepsilon$ , [26]. This paper considers the tolerance  $\varepsilon = 10^{-5}$ .

### 3.2. Estimation of Standard Errors

To compute the asymptotic covariance of the ML estimates  $\hat{\Theta}$ , we employ the information-based method suggested by Meilijson [27]. Formally, the empirical information matrix is defined as

$$I_e(\Theta | \mathbf{t}) = \sum_{j=1}^n \mathbf{s}(t_j | \Theta) \mathbf{s}^T(t_j | \Theta) - \frac{1}{n} \mathbf{S}(\mathbf{t} | \Theta) \mathbf{S}^T(\mathbf{t} | \Theta), \quad (10)$$

where  $\mathbf{S}(\mathbf{t} | \Theta) = \sum_{j=1}^n \mathbf{s}(t_j | \Theta)$  and  $\mathbf{s}(t_j | \Theta)$  are individual scores which can be determined from the result of [20] as  $\mathbf{s}(t_j | \Theta) = \partial f(t_j | \Theta) / \partial \Theta = E[\partial \ell_c(\Theta | t_j, w_j, z_j) / \partial \Theta | t_j, \Theta]$ . Substituting the ML estimates  $\hat{\Theta} = (\hat{p}_1, \dots, \hat{p}_{g-1}, \hat{\alpha}_1, \dots, \hat{\alpha}_g, \hat{\beta}_1, \dots, \hat{\beta}_g, \hat{\lambda}_1, \dots, \hat{\lambda}_g)$  into (10) gives

$$I_e(\Theta | \mathbf{t}) = \sum_{j=1}^n \hat{\mathbf{s}}_j \hat{\mathbf{s}}_j^T, \quad (11)$$

where  $\hat{\mathbf{s}}_j = (\hat{s}_{j,p_1}, \dots, \hat{s}_{j,p_{g-1}}, \hat{s}_{j,\alpha_1}, \dots, \hat{s}_{j,\alpha_g}, \hat{s}_{j,\beta_1}, \dots, \hat{s}_{j,\beta_g}, \hat{s}_{j,\lambda_1}, \dots, \hat{s}_{j,\lambda_g})$ . The explicit expressions for the elements of  $\hat{\mathbf{s}}_j$  are summarized below

$$\begin{aligned} \hat{s}_{j,p_r} &= \frac{\hat{z}_{rj}}{\hat{p}_r} - \frac{\hat{z}_{gj}}{\hat{p}_g}, & \hat{s}_{j,\lambda_r} &= \hat{z}_{rj} (-\hat{\lambda}_r \hat{w}_{rj} + a(t_j; \hat{\alpha}_r, \hat{\beta}_r)), \\ \hat{s}_{j,\alpha_r} &= \hat{z}_{rj} \left( \frac{-1}{\hat{\alpha}_r} + \frac{\hat{u}_{rj}}{\hat{\alpha}_r^3} \left( \frac{t_j}{\hat{\beta}_r} + \frac{\hat{\beta}_r}{t_j} - 2 \right) - \frac{\hat{\lambda}_r}{\hat{\alpha}_r^2} \xi(t_j, \hat{\beta}_r) \right), \\ \hat{s}_{j,\beta_r} &= \hat{z}_{rj} \left( \frac{-1}{2\hat{\beta}_r} - \frac{\hat{u}_{rj}}{2\hat{\alpha}_r^2} \left( \frac{1}{t_j} - \frac{t_j}{\hat{\beta}_r^2} \right) + \frac{1}{t_j + \hat{\beta}_r} + \frac{\hat{\lambda}_r}{2\hat{\beta}_r \hat{\alpha}_r} \left( \sqrt{t_j / \hat{\beta}_r} + \sqrt{\hat{\beta}_r / t_j} \right) \right). \end{aligned}$$

As a result, the standard errors of  $\hat{\Theta}$  are obtained as the square roots of the diagonal elements of the inverse of (11).

## 4. REAL DATA ANALYSIS

In this section, we use Enzyme data to illustrate the performance of the Mix-SLBS model to fit the data. These data, originally analyzed by Bechtel *et al.* [28], correspond to the enzymatic activity in the blood and represent the metabolism of carcinogenic substances among 245 unrelated individuals. The enzymatic activity is quantified by the molar ratio between two metabolites of caffeine. Bechtel *et al.* concluded that a mixture of two skewed distributions is suitable for analyzing these Enzyme data. Recently, Balakrishnan *et al.* [19] used these data to compare three kinds of BS mixture model. Their proposed model were (a) mixture of two BS distribution (Mix-BS), (b) Mixture of Length-biased BS and BS distributions (Mix-LBBS) and (c) Mixture of Length-biased BS and BS distributions with the same parameters (Mix-LBSBS). Here, the Akaike Information Criterion (AIC) [29] as well as the Bayesian Information Criterion (BIC) [30] are computed to identify the best selected model for comparing our proposed model with these three models. The AIC and BIC are formulated by  $m c_n - 2\ell(\hat{\Theta})$ , where  $\ell(\hat{\Theta})$  is the actual log-likelihood,  $m$  is the number of free parameters that have to be estimated under the model and the penalty term  $c_n$  is a convenient sequence of positive numbers. We have  $c_n = 2$  for AIC and  $c_n = \log(n)$  for BIC. Also, the Kolmogorov–Smirnov (KS) test is carried out as a measure of Goodness of fit test. The KS statistic is a distance between the empirical cdf and the estimated theoretical cdf for the models. The  $p$ -value of KS test can be used as a similarity assessment of the experimental data against the fitted distribution.

We fit Mix-SLBS model for different number of components  $g$ , and find that the best  $g = 2$  based on AIC and BIC criteria. The results of the study are summarized in Table 1. Results based on AIC and BIC show that the Mix-SLBS provides a highly improved fit of the data over three other competitors. The results of Table 1 can also be comparable with table 5 in Benites *et al.* [31] where they fits Mix-BS distributions with different number of components. Moreover, the Mix-SLBS model yields quite smaller standard errors for the estimated parameters which shows that the Mix-SLBS distributions allows to produce more precise estimates for this data example.

On the other hand, the  $p$ -value of the KS test for Mix-SLBS distributions is significantly greater than the Mix-BS, Mix-LBBS and Mix-LBSBS models which strongly confirms that the Enzyme data follow the Mix-SLBS distributions.

Finally, we provide the probability-probability (PP)-plot of the two best fitted models in Figure 3. The PP-plot shows that the Mix-SLBS model adapts to the shape of the histogram very accurately.

## 5. SIMULATION STUDY

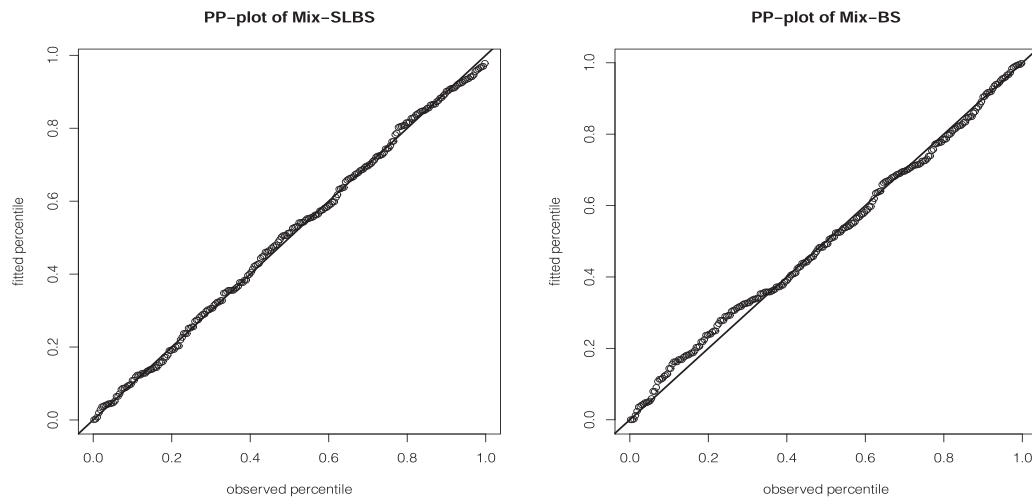
In this section, we conduct two simulation studies in order to examine the performance of the proposed method. The first study shows that the underlying Mix-SLBS model is robust in the ability to cluster heterogeneous lifetime data in the presence of outliers. In the second simulation study, we investigate if the ML estimates obtained using the proposed ECM algorithm can provide good asymptotic properties. All simulation studies are done in the statistical software R.

**Table 1** | Parameter estimates with corresponding standard errors of the Enzyme data set.

Parameter	Mix-SLBS		Mix-BS		Mix-LBBS		Mix-LBSBS	
	MLE	SE	MLE	SE	MLE	SE	MLE	SE
$\pi$	0.636	0.007	0.629	0.031	0.450	0.028	0.417	0.026
$\alpha_1$	0.399	0.001	0.533	0.032	0.365	0.034	1.038	0.084
$\beta_1$	0.185	0.012	0.175	0.007	0.171	0.007	0.216	0.012
$\lambda_1$	-0.029	0.001	—	—	—	—	—	—
$\alpha_2$	0.210	0.016	0.319	0.025	1.274	0.114	1.038	0.084
$\beta_2$	1.003	0.039	1.274	0.043	0.213	0.044	0.216	0.012
$\lambda_2$	0.587	0.003	—	—	—	—	—	—
$\ell(\hat{\Theta})$	<b>-46.394</b>		-59.168		-71.091		-115.899	
AIC	<b>106.787</b>		128.336		152.182		237.798	
BIC	<b>131.296</b>		145.842		169.688		248.302	
KS	<b>0.039</b>		0.053		0.111		0.151	
p-value	<b>0.832</b>		0.507		0.005		< 0.001	

BS, Birnbaum–Saunders; LBBS, Length-biased Birnbaum–Saunders; LBSBS, Length-biased Birnbaum–Saunders and Birnbaum–Saunders distributions with the same parameters; MLE, maximum likelihood estimation; SE, standard error; SLBS, skew-Laplace Birnbaum–Saunders.

The bold numbers are related to the best model coordinating to the each model comparison measure.

**Figure 3** | The probability-probability (PP)-plots of two best distributions.

## 5.1. Robustness of the Model

In the first experiment, the ability of the Mix-SLBS model in clustering observations (allocating them into groups of observations that are similar in some sense) is verified. Although, we know that each data point belongs to one of  $g$  heterogeneous populations, we cannot find any discriminating point between them. Through modeling via mixture models, we can cluster the data in terms of the estimated (posterior) probability of belonging to a given group. Our main idea is to use the flexibility of these models in life time data and to extend the ability of them by including possible higher skewness and kurtosis of the related components.

We generated 500 samples from a mixture of two BS and two SLBS densities with the parameter values at  $S1 : p = 0.6, \alpha = (2, 1), \beta = (2, 1)$  and  $\lambda = (2, -2)$  and  $S2 : p = 0.6, \alpha = (1, 1), \beta = (1, 1)$  and  $\lambda = (3, 0)$ . For each artificial sample, we also proceeded clustering considering the known true classification. Then, the related mixture models were fitted in order to obtain the estimate of the posterior probability that an observation  $t_i$  belongs to the  $j$ th component of the mixture,  $\hat{z}_{ij}$ . Finally for each replication  $v, v = 1, \dots, 500$ , we obtain the number of correct allocations divided by the sample size,  $r_v$ .

Table 2 shows the mean value of  $r_v$ , i.e.,  $(1/500) \sum_{v=1}^{500} r_v$ . We observe from Table 2 that the rate of correct allocations in both mixture models is increased by increasing sample size when samples are taken from Mix-BS model. But, comparing with the results for the Mix-BS model, it can be seen that modeling using the Mix-SLBS distribution represents a substantial improvement in the clustering when the samples are taken from the Mix-SLBS model. This is happened due to heavier tail of the SLBS distribution. So, the Mix-SLBS has a better performance in clustering data with heavier tail which shows its robustness to discrepant observations.



**Table 2** | Mean right allocations rates for fitted models.

True Model	Sample Size	S1: Fitted Model		S2: Fitted Model	
		Mix-BS	Mix-SLBS	Mix-BS	Mix-SLBS
Mix-BS	100	0.8134	0.7861	0.7998	0.7460
	500	0.8231	0.8350	0.8067	0.7984
	1000	0.8358	0.8345	0.8109	0.8201
Mix-SLBS	100	0.5189	0.9722	0.6018	0.8874
	500	0.4477	0.9761	0.5929	0.9146
	1000	0.4449	0.9776	0.5901	0.9191

BS, Birnbaum–Saunders; SLBS, skew-Laplace Birnbaum–Saunders.

**Table 3** | Bias and mean squared errors for EM estimates of simulated data.

Measure	Parameter	S1				S2				S3			
		Sample Size				Sample Size				Sample Size			
		100	250	500	1000	100	250	500	1000	100	250	500	1000
R.Bias	$p$	0.237	0.034	0.024	0.017	0.211	0.097	0.046	0.026	0.030	0.020	0.022	0.001
	$\alpha_1$	0.450	0.372	0.284	0.108	0.445	0.397	0.308	0.134	0.296	0.176	0.098	0.038
	$\alpha_2$	0.182	0.134	0.110	0.094	0.195	0.114	0.054	0.009	0.084	0.066	0.062	0.007
	$\beta_1$	0.098	0.089	0.069	0.043	0.984	0.569	0.327	0.209	0.247	0.209	0.150	0.095
	$\beta_2$	0.456	0.369	0.301	0.198	0.365	0.259	0.186	0.097	0.080	0.025	0.008	0.002
	$\lambda_1$	0.557	0.457	0.311	0.167	0.399	0.279	0.153	0.068	0.527	0.473	0.335	0.196
	$\lambda_2$	0.209	0.094	0.045	0.013	0.188	0.078	0.026	0.009	0.219	0.208	0.189	0.163
MSE	$p$	0.050	0.027	0.012	0.005	0.059	0.034	0.017	0.008	0.058	0.043	0.031	0.024
	$\alpha_1$	0.053	0.034	0.026	0.015	0.444	0.219	0.134	0.086	0.256	0.185	0.099	0.010
	$\alpha_2$	0.156	0.089	0.055	0.019	0.201	0.152	0.091	0.032	0.113	0.066	0.043	0.009
	$\beta_1$	0.009	0.009	0.008	0.006	1.628	1.079	0.793	0.258	0.062	0.040	0.015	0.001
	$\beta_2$	0.331	0.218	0.161	0.083	0.242	0.173	0.109	0.045	0.049	0.023	0.014	0.008
	$\lambda_1$	1.654	1.202	0.985	0.384	0.887	0.496	0.273	0.129	0.691	0.467	0.323	0.165
	$\lambda_2$	1.029	0.504	0.321	0.176	0.835	0.638	0.356	0.192	0.612	0.412	0.361	0.136

EM, expectation and maximization; MSE, mean squared error.

## 5.2. Asymptotic Properties

We carry out a second simulation study to check the finite-sample performances of the ML estimates obtained by using the ECM algorithm. To show the effect of parameters on the estimation procedure, we consider three set of the parameter values  $\Theta = (p, \alpha_1, \beta_1, \lambda_1, \alpha_2, \beta_2, \lambda_2)$  as follows:

$$S1)\Theta = (0.7, 0.25, 1, 1, 1, 2, 2), \quad S2)\Theta = (0.7, 1, 0.25, 2, 1, 1, 2), \quad \text{and} \quad S3)\Theta = (0.7, 1, 1, 0.5, 1, 1, 2).$$

As suggested in [32], the parameters were taken to produce highly skewed and heavy-tailed distributions. In each replication of 500 trials, we generate data of size  $n = 100, 500, 1000, 5000$ , from the Mix-SLBS model. To investigate the estimation accuracies, the bias and the mean squared error (MSE) of overall samples are computed as:

$$\text{R.Bias} = \frac{1}{500} \sum_{i=1}^{500} \left| \frac{\hat{\theta}_i - \theta}{\theta} \right|, \quad \text{and} \quad \text{MSE} = \frac{1}{500} \sum_{i=1}^{500} (\hat{\theta}_i - \theta)^2,$$

where  $\hat{\theta}_i$  is the estimation of specific parameter  $\theta$  when the data is sample  $i$ .

Table 3 presents the results of the simulation. It is evident that both magnitudes bias and MSE of the estimators decrease and tend to toward zero when the sample size increases. It shows that the ML estimates obtained via the ECM algorithm are empirically consistent.



## 6. CONCLUSIONS

In this paper, we have dealt with a new finite mixture model based on a new extension of BS distribution, called the Mix-SLBS. We have presented a convenient hierarchical representation and developed the ECM algorithm to obtain the ML estimates of the parameters. Numerical results suggest that the proposed Mix-SLBS distributions is well suited to the experimental data and can be more robust and flexible against outliers as compared with the 3 mixture competitors, proposed by Balakrishnan *et al.* [19].

Some possible directions of the current work remain that deserve further attention. For instance, the linear and non-linear regression model based on the SLBS distribution can be presented by using Theorem 2.2 [33,34]. As another extension of this work, it is worthwhile to generalize BS distribution in more general case by considering GIG distribution instead of  $Exp(0.5)$ . It is also interesting to introduce multivariate SLBS distribution and setting finite mixture of multivariate SLBS distributions [35].

## CONFLICT OF INTEREST

No potential conflict of interest was reported by the authors.

## AUTHORS' CONTRIBUTIONS

M. Naderi developed main methodological parts of the paper whereas M. Mozafari and K. Okhli implemented the methodology in the R program and applied them to the real and simulated data examples.

## ACKNOWLEDGMENTS

We are grateful to the Editor-in-Chief, anonymous referees for their comments, which greatly improved this work. M. Naderi is also appreciate the support by the National Research Foundation, South Africa (Reference: CPRRI60403161466 351 Grant No. 105840, SARChI Research Chair- UID: 71199, and STATOMET).

## REFERENCES

1. Z.W. Birnbaum, S.C. Saunders, *J. Appl. Probab.* 6 (1969), 319–327.
2. Z.W. Birnbaum, S.C. Saunders, *J. Appl. Probab.* 6 (1969), 328–347.
3. Y. Li, A. Xu, *J. Stat. Comput. Simul.* 86 (2016), 1673–1685.
4. C. Marchant, V. Leiva, F.J.A. Cysneiros, J.F. Vivanco, *J. Appl. Stat.* 43 (2016), 2829–2849.
5. G. Martínez-Flórez, H. Bolfarine, H.W. Gómez, *Statistics.* 48 (2014), 896–912.
6. K. Mohammadi, O. Alavi, J.G. McGowan, *Energy Convers. Manag.* 143 (2017), 109–122.
7. N. Balakrishnan, D. Kundu, *Appl. Stoch. Model. Bus. Ind.* 35 (2018), 4–49.
8. N. Balakrishnan, V. Leiva, A. Sanhueza, F. Vilca, *Sort.* 33 (2009), 171–192.
9. S. Nadarajah, S. Kotz, *Stat. Probab. Lett.* 65 (2003), 269–277.
10. A. Jamalizadeh, F. Hashemi, M. Naderi, *Appl. Stoch. Model. Bus. Ind.* 35 (2019), 82–89.
11. D. Kundu, N. Balakrishnan, A. Jamalizadeh, *J. Multivariate Anal.* 101 (2010), 113–125.
12. F. Vilca-Labrador, V. Leiva-Sánchez, *Comm. Stat. Theor. Methods.* 35 (2006), 229–244.
13. A. McNeil, R. Frey, P. Embrechts, *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press, New Jersey, 2005.
14. O. Arslan, *Stat. Pap.* 51 (2010), 865–887.
15. A.P. Dempster, N.M. Laird, D.B. Rubin, *J. R. Stat. Soc. Ser. B. Stat. Methodol.* 39 (1977), 1–38.
16. D. Böhning, E. Dietz, R. Schaub, P. Schlattmann, B.G. Lindsay, *Ann. Inst. Stat. Math.* 46 (1994), 373–388.
17. G. McLachlan, D. Peel, *Finite Mixture Models*, John Wiley and Sons, New York, NY, USA, 2004.
18. S. Ali, *Appl. Math. Model.* 39 (2015), 515–530.
19. N. Balakrishnan, R.C. Gupta, D. Kundu, V. Leiva, A. Sanhueza, *J. Stat. Plan. Inference.* 141 (2011), 2175–2190.
20. T.A. Louis, *J. R. Stat. Soc. Ser. B. Stat. Methodol.* 44 (1982), 226–232.
21. V. Leiva, F. Vilca, N. Balakrishnan, A. Sanhueza, *Comm. Stat. Theor. Methods.* 39 (2010), 426–443.
22. B. Jørgensen, *Statistical properties of the generalized inverse Gaussian distribution*, Springer Science and Business Media, New York, NY, USA, 2012.
23. X.L. Meng, D.B. Rubin, *Biometrika.* 80 (1993), 267–278.
24. P.D. McNicholas, T.B. Murphy, A.F. McDaid, D. Frost, *Comput. Stat. Data Anal.* 54 (2010), 711–723.
25. A.C. Aitken, *Proc. R. Soc. Edinb.* 46 (1926), 289–305.
26. B.G. Lindsay, in *NSF-CBMS Regional Conference Series in Probability and Statistics*, Alexandria, Virginia, 1995, pp. i–163.

27. I. Meilijson, *J. R. Stat. Soc. Ser. B. Stat. Methodol.* 51 (1989), 127–138.
28. Y.C. Bechtel, C. Bonaiti-Pellie, N. Poisson, J. Magnette, P.R. Bechtel, *Clin. Pharmacol. Ther.* 54 (1993) 134–141.
29. H. Akaike, *IEEE Trans. Automat. Control.* 19 (1974), 716–723.
30. G. Schwarz, *Ann. Stat.* 6 (1978), 461–464.
31. L. Benites, R. Maehara, F. Vilca, F. Marmolejo-Ramos, Finite mixture of Birnbaum-Saunders distributions using the k-bumps algorithm, arXiv preprint arXiv:1708.00476, 2017.
32. H. J. Ho, S. Pyne and T. I. Lin, *Stat. Comput.* 22 (2012), 287–299.
33. L. Hernando Vanegas, G.A. Paula, *J. Stat. Comput. Simul.* 86 (2016), 1709–1735.
34. A.J. Lemonte, *J. Stat. Comput. Simul.* 86 (2016), 611–622.
35. M. Khosravi, D. Kundu, A. Jamalizadeh, *Stat. Methodol.* 23 (2015), 1–17.