On some mixture models based on the Birnbaum-Saunders distribution and associated inference

N. Balakrishnan[a], **Ramesh C. Gupta**[b], **Debasis Kundu**[c], **Víctor Leiva**[d] and **Antonio Sanhueza**[e]

[a] *Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada*

[b] *Department of Mathematics and Statistics, University of Maine, Orono, ME, USA*

[c] *Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Pin 208016, India*

[d] *Departamento de Estadística, CIMFAV, Universidad de Valparaíso, Valparaíso, Chile*

[e] *Departamento de Matemática y Estadística, Universidad de La Frontera, Temuco, Chile*

Abstract

In this paper, we consider three different mixture models based on the Birnbaum-Saunders (BS) distribution, viz., (1) mixture of two different BS distributions, (2) mixture of a BS distribution and a length-biased version of another BS distribution, and (3) mixture of the BS distribution and its length-biased version. For all these models, we study their structural properties as well as the shape of their density and hazard rate functions. For the maximum likelihood estimation of the model parameters, we utilize the EM algorithm, which is described and implemented. For the purpose of illustration, we analyze two data sets related to enzyme and depressive condition problems. In the case of enzyme data, it is shown that Model 1 provides the best fit, while for the depressive condition data, it is shown all three models fit well with Model 3 providing the best fit.

MSC: 65C10; 60E05

Key words: EM algorithm; Fisher information; Goodness-of-fit; Hazard rate function; Inverse Gaussian distribution; Length-biased distributions; Maximum likelihood methods.

---

1 Introduction

Birnbaum and Saunders (1969) proposed a fatigue failure model based on a physical argument derived from the cumulative damage or Miner law. This model is known as the Birnbaum-Saunders (BS) distribution. Desmond (1985) strengthened the physical justification of this model and relaxed some of the assumptions made by Birnbaum and Saunders (1969). The BS model is a positively skewed, unimodal, two-parameter distribution with non-negative support possessing many attractive properties. This model has found several applications in the literature including lifetime, survival and environmental data analysis.

Extensive work has been done on the BS distribution with regard to its properties, inference and applications. A comprehensive treatment on the BS distribution till mid 90's can be found in Johnson, Kotz and Balakrishnan (1995, pp. 651-662). For more recent references on this distribution and some of its generalizations, the readers may refer to Ng, Kundu and Balakrishnan (2003, 2006), Kundu, Kannan and Balakrishnan (2008), Sanhueza, Leiva and Balakrishnan (2008), and Balakrishnan, Leiva, Sanhueza and Vilca (2009).

Since a random variable (RV) following the BS distribution is defined through a standard normal RV, then the probability density function (PDF) and the cumulative distribution function (CDF) of the BS model can be expressed in terms of the standard normal PDF and CDF. As mentioned, the PDF of the BS distribution is unimodal. Kundu, Kannan and Balakrishnan (2008) showed that the hazard rate function (HRF) of this distribution is not monotone and is in fact unimodal for all ranges of the parameter values; see also Gupta and Akman (1995, 1997) and Bebbington, Lai and Zitikis (2008).

Length-biased (LB) – also referred to as size-biased – versions of a distribution have received considerable attention in the literature; see Fisher (1934), Rao (1965), and Cox (1969). The LB distributions are a special case of the weighted distributions, which have been derived from a LB sampling procedure; see Patil (2002). These distributions have found applications in diverse areas such as biometry, ecology, environmental science, and reliability analysis. LB versions of several distributions have been considered. Patil and Rao (1978) provided a table of some basic distributions and their LB versions, while Khattree (1989) presented relationships among variates and their LB versions for some specific distributions. In the context of reliability, this relationship was treated by some other authors; see, for example, Gupta and Keating (1985), Jain (1989), and the references therein. A review of different LB distributions and their applications has been provided by Gupta and Kirmani (1990); one may also refer to Olyede and George (2002) for some more recent works in this direction. Specifically, LB versions of the IG and lognormal distributions can be seen in Sansgiry and Akman (2001) and Gupta and Akman (1995). Recently, Leiva, Sanhueza and Angulo (2009) considered the LB version of the BS (LBS) distribution and illustrated its application in water quality analysis.

In reliability studies, populations often turn out to be heterogeneous simply because there are at least two subpopulations, with one being the standard subpopulation (sometimes called strong) and the other one being the defective subpopulation (weak). For this reason, data arising from such heterogeneous populations need to be modeled by a mixture of two or more life distributions. More details about life distributions can be seen in Johnson, Kotz and Balakrishnan (1995, pp. 639-681) and Marshall and Olkin (2007), and about mixture distributions in McLachlan and Peel (2000).

In this paper, we consider three different two-component mixture models based on the BS and LBS distributions. Model 1 is a mixture of two BS distributions with different sets of parameters, which involves five parameters including one mixing parameter. Model 2 is a mixture of the BS and LBS distributions with different sets of parameters, which is once again a five-parameter model. Finally, Model 3 is a special case of Model 2, which is a mixture of a BS distribution and its LB version, and so consequently it has only three parameters including one mixing parameter. In this paper, we discuss different structural and reliability aspects of these three mixture distributions. It is well-known (see, e.g., Ng, Kundu and Balakrishnan (2003)) that the maximum likelihood (ML) estimators of the parameters of a two-parameter BS distribution can not be obtained in closed-form. Thus, one needs to solve a non-linear equation for the determination of the ML estimates. This is also the case when the parameters of the LBS distribution are estimated by the ML method. For the direct computation of the ML estimates of the model parameters, one needs to solve five-dimensional optimization problems for Models 1 and 2, but this optimization problem is reduced to one three-dimensional for the case of Model 3. To facilitate the associated numerical procedure, we use the expectation and maximization (EM) algorithm proposed by Dempster, Laird and Rubin (1977) for the computation of the ML estimates of all the model parameters through which the required multi-dimensional optimization is solved by a sequence of one-dimensional optimizations. Finally, using the idea of Louis (1982), it also becomes possible to compute the Fisher information matrix, which is useful in constructing asymptotic confidence intervals for the model parameters.

Bhattacharyya and Fries (1982), Desmond (1986), and Jorgensen, Seshadri and Whitmore (1991) noted a relationship between the BS and inverse Gaussian (IG) distributions. They observed that the BS distribution can be obtained as an equally weighted mixture of an IG distribution and its LB (LIG) version (or complementary reciprocal). In the case of Model 1, this relationship can be utilized for computing the ML estimates more efficiently. It is observed that for each E-step of the EM algorithm, the corresponding M-step can be obtained in an explicit form by using the fact that the ML estimators of the parameters of the IG and LIG distributions can be explicitly obtained. This renders the implementation of the EM algorithm to be efficient and effective. In the case of Model 2, while implementing the EM algorithm, it is observed that in each E-step, the corresponding M-step would require two one-dimensional optimizations, which significantly reduces the computational burden. Model 3 is, in fact, a weighted version of the two-parameter BS distribution with a linear weight function. In this last model, the corresponding M-step would require solving only one non-linear equation in each E-step.

The rest of the paper is organized as follows. In Section 2, we briefly describe the BS and LBS distributions. In Section 3, we present the three mixture models and examine some of their properties. In Section 4, we deal with the estimation problem for the three models. In Section 5, we analyze two real data sets for the purpose of illustrating the three mixture models discussed in this work. Finally, in Section 6, we make some concluding comments.

2 Background

2.1 The Birnbaum-Saunders distribution

From now on, if a RV T follows the two-parameter BS distribution, the notation $T \sim \mathrm{BS}(\alpha, \beta)$ is used. In this case, the PDF of T can be written as

$$f_T(t; \alpha, \beta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2\alpha^2}\left[\frac{t}{\beta} + \frac{\beta}{t} - 2\right]\right) \frac{\left[\left\{\frac{\beta}{t}\right\}^{\frac{1}{2}} + \left\{\frac{\beta}{t}\right\}^{\frac{3}{2}}\right]}{2\alpha\beta}, \quad t > 0, \quad (1)$$

where α and β are the shape and scale parameters of the BS distribution, respectively. The corresponding CDF of T is given by

$$F_T(t; \alpha, \beta) = \Phi\left(\frac{1}{\alpha}\left[\left\{\frac{t}{\beta}\right\}^{\frac{1}{2}} - \left\{\frac{\beta}{t}\right\}^{\frac{1}{2}}\right]\right), \quad t > 0, \quad (2)$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution. The quantile function of T is expressed as

$$t(q) = \frac{\beta}{4}\left[\alpha\, z(q) + \sqrt{\{\alpha z(q)\}^2 + 4}\right]^2, \quad 0 < q < 1, \quad (3)$$

where $z(q)$ is the qth quantile of the standard normal distribution. Since $t(0.5) = \beta$, β is also the median of the BS distribution. The HRF of T can be easily obtained from (1) and (2) as

$$h_T(t;\alpha,\beta) = \frac{f_T(t;\alpha,\beta)}{\Phi\left(-\frac{1}{\alpha}\left[\left\{\frac{t}{\beta}\right\}^{\frac{1}{2}} - \left\{\frac{\beta}{t}\right\}^{\frac{1}{2}}\right]\right)}, \quad t > 0. \tag{4}$$

Some basic properties include: (i) $c\,T \sim \mathrm{BS}(\alpha, c\beta)$, with $c > 0$, and (ii) $1/T \sim \mathrm{BS}\,(\alpha, 1/\beta)$. From (3), it is possible to note that $T \stackrel{\mathrm{d}}{=} [\beta/4]\left[\alpha\, Z + \sqrt{\{\alpha Z\}^2 + 4}\right]^2$, where $Z \sim \mathrm{N}(0,1)$, where $\stackrel{\mathrm{d}}{=}$ means equal distribution. Thus, it can be proved that

$$X \stackrel{\mathrm{d}}{=} \frac{1}{2}\left[\left\{\frac{T}{\beta}\right\}^{\frac{1}{2}} - \left\{\frac{\beta}{T}\right\}^{\frac{1}{2}}\right] \sim \mathrm{N}\left(0,\ \frac{\alpha^2}{4}\right), \tag{5}$$

where $\mathrm{N}(a, b^2)$ denotes the normal distribution with mean a and variance b^2, which implies $T = \beta\left[1 + 2X^2 + 2X\{1 + X^2\}^{\frac{1}{2}}\right] \sim \mathrm{BS}(\alpha,\beta)$. Using the transformation given in (5), the rth moment of T can be shown to be

$$\mathrm{E}\,[T^r] = \beta^r \sum_{j=0}^{r} \binom{2r}{2j} \sum_{i=0}^{j} \binom{j}{i} \frac{(2r-2j+2i)!}{2^{r-j+i}(r-j+i)!}\left[\frac{\alpha}{2}\right]^{2r-2j+2i}. \tag{6}$$

In particular, from (6), we have

$$\begin{aligned}
&\mathrm{E}[T] = \beta\left[1 + \frac{\alpha^2}{2}\right], \quad \mathrm{V}[T] = [\alpha\beta]^2\left[1 + \frac{5\alpha^2}{4}\right],\\
&\beta_1[T] = \frac{16\alpha^2[11\alpha^2 + 6]}{[5\alpha^2 + 4]^3}, \quad \text{and} \quad \beta_2[T] = 3 + \frac{6\alpha^2[93\alpha^2 + 40]}{[5\alpha^2 + 4]^3},
\end{aligned} \tag{7}$$

where $\beta_1[T]$ and $\beta_2[T]$ denote the coefficients of skewness and kurtosis of T, respectively. From $\beta_1[T]$ in (7), it may be observed that, as $\alpha \to 0$, the skewness also goes to zero, so that the BS distribution tends in this case to be symmetrical, but it degenerates at the scale parameter β; see Kundu, Kannan and Balakrishnan (2008). If $\beta = 1$, then the mode of the BS distribution can be obtained as the solution of the non-linear equation

$$\frac{x[3x^{\frac{3}{2}} + x^{\frac{1}{2}}]}{1 - x^2} = \frac{1}{\alpha^2}. \tag{8}$$

If $x^*(\alpha)$ is the solution of (8), then this modal value becomes $\beta x^*(\alpha)$. From (8), it is immediate that $x^*(\beta) < 1$, so that the mode of the BS distribution is less than β. Table 1 presents the BS modal values for different values of α. From this table, it is readily seen that x^* is a decreasing function of α.

Table 1
Values of x^* for different choices of α.

α	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
x^*	0.720	0.471	0.337	0.258	0.207	0.172	0.146	0.126	0.110	0.098

We now briefly describe how the ML estimation of the parameters of the BS distribution can be done based on a random sample $T_1, \ldots, T_n$, where $T_i \sim \mathrm{BS}(\alpha, \beta)$, for $i = 1, \ldots, n$. The ML estimator of α, say $\widehat{\alpha}$, is explicitly obtained as

$$\widehat{\alpha} = \left[\frac{s}{\widehat{\beta}} + \frac{\widehat{\beta}}{r} - 2\right]^{\frac{1}{2}}, \tag{9}$$

where $\widehat{\beta}$ is the ML estimator of β. However, it is not possible to find an explicit form for $\widehat{\beta}$. Thus, the ML estimate of β can be obtained as the unique positive root of the equation

$$\beta^2 - \beta[2s + K(\beta)] + r[s + K(\beta)] = 0, \tag{10}$$

where $s = [1/n]\sum_{i=1}^{n} t_i$ and $r = [[1/n]\sum_{i=1}^{n}[1/t_i]]^{-1}$ are the sample arithmetic and harmonic means, respectively, and $K(u) = [[1/n]\sum_{i=1}^{n}[u + t_i]^{-1}]^{-1}$ is the harmonic mean function, for $u \geq 0$, such that $K(0) = r$. Note that (10) can be easily solved by using the fixed-point type equation given by

$$h(\beta) = \beta, \quad \text{where} \quad h(\beta) = \frac{\beta^2 + r[s + K(\beta)]}{2r + K(\beta)}. \tag{11}$$

Then, a simple iterative scheme such as $h(\beta^{(m)}) = \beta^{(m+1)}$ can be used to determine the ML estimate of β based on an initial guess $\beta^{(0)}$, which can be the sample median, for example. Once $\widehat{\beta}$ is obtained as a solution of (10), then the ML estimate of α can be obtained from (9). Engelhardt, Bain and Wright (1981) obtained asymptotic inference for α and β by using the result that

$$\begin{pmatrix}\widehat{\alpha}\\ \widehat{\beta}\end{pmatrix} \quad \dot{\sim} \quad \mathrm{N}\left(\begin{pmatrix}\alpha\\ \beta\end{pmatrix}, \begin{pmatrix}\frac{\alpha^2}{2n} & 0\\ 0 & \frac{\beta^2}{n[0.25+\alpha^{-2}+I(\alpha)]}\end{pmatrix}\right), \tag{12}$$

where $I(\alpha) = 2\int_0^\infty [[1 + g(\alpha x)]^{-1} - 0.5]^2 d\Phi(x)$ and $g(y) = 1 + y^2/2 + y[1 + y^2/4]^{1/2}$. An alternative way for determining the variance-covariance matrix of $\widehat{\boldsymbol{\theta}} = (\widehat{\alpha}, \widehat{\beta})$ is from the expected information matrix, say $J(\boldsymbol{\theta})$, as $\mathrm{Var}[\widehat{\boldsymbol{\theta}}] = J(\boldsymbol{\theta})^{-1}$. Instead of the expected information matrix, one may prefer to use the observed information matrix since it is easier to compute. In the application part of this work, we carried out asymptotic inference on the model parameters based on the observed information matrix; for more details on this approach, see Efron and Hinkley (1978).

2.2 Length-biased distributions

As mentioned earlier, the LB distributions are a special case of the weighted distributions. Suppose Y is a non-negative RV with PDF $f_Y(y)$. The weighted version of Y, say Y_w, with weight function $w(\cdot)$, has a probabilistic model called the weighted distribution. The PDF of Y_w is given by

$$f_{Y_w}(y) = \frac{w(y) f_Y(y)}{\mathrm{E}[w(Y)]}, \quad y > 0. \tag{13}$$

Here, it is assumed that $\mathrm{E}[w(Y)] < \infty$. A particular case of the weighted distributions is obtained when we use the weight function $w(y) = y$ in (13). In this case, Y_w is called the LB (or size-biased) version of Y, denoted by L. The PDF of L is given by $f_L(l) = l\, f_Y(l)/\mu$, for $l > 0$, with $\mu = \mathrm{E}[Y] < \infty$. The distribution of L is the LB version of the distribution of Y. Note that the LB distribution involves the same parameters as in the original model.

2.3 Connection between the BS and IG distributions

Consider the following mixture representation of the BS distribution. Suppose $X_1 \sim \mathrm{IG}(\mu, \lambda)$, i.e., X_1 has an IG distribution with parameters $\mu > 0$ and $\lambda > 0$. The PDF of X_1 is given by

$$f_{X_1}(x; \mu, \lambda) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\lambda}{2} \frac{[x-\mu]^2}{\mu^2 x}\right) \frac{\sqrt{\lambda}}{\sqrt{x^3}}, \quad x > 0. \tag{14}$$

Furthermore, let X_2 be a RV such that $1/X_2 \sim \mathrm{IG}(1/\mu, \lambda/\mu^2)$, which is independent of X_1. Then, consider a new RV given by

$$T = \begin{cases} X_1, & \text{with probability } 1/2, \\ X_2, & \text{with probability } 1/2. \end{cases} \tag{15}$$

Evidently, the PDF of T (being a mixture of X_1 and X_2) is defined as

$$f_T(x;\mu,\lambda) = \frac{1}{2}f_{X_1}(x;\mu,\lambda) + \frac{1}{2}f_{X_2}(x;\mu,\lambda), \quad x > 0, \tag{16}$$

where $f_{X_1}(x;\mu,\lambda)$ and $f_{X_2}(x;\mu,\lambda)$ are the densities of X_1 and X_2, respectively. Here, the PDF of X_2 can be expressed as

$$f_{X_2}(x;\mu,\lambda) = \frac{x f_{X_1}(x;\mu,\lambda)}{\mu}, \quad x > 0, \tag{17}$$

where $\mu = \mathrm{E}[X_1]$, so that X_2 is the LB version of X_1. Therefore, if $\alpha = \sqrt{\mu/\lambda}$ and $\beta = \mu$ in (16), then $T \sim \mathrm{BS}(\alpha,\beta)$. The representation given in (16) is quite useful for developing the EM algorithm. For more details on the IG model, one may refer to Chhikara and Folks (1989), Johnson, Kotz and Balakrishnan (1994, Chapter 15), and Seshadri (1993), and to Balakrishnan, Leiva, Sanhueza and Cabrera (2009) for more recent results about the BS and IG distributions and its mixture.

2.4 The length-biased Birnbaum-Saunders distribution

Let $T \sim \mathrm{BS}(\alpha,\beta)$. Then, the LB version of T, say L, i.e., with LBS distribution, denoted it by $L \sim \mathrm{LBS}(\alpha,\beta)$, has PDF expressed as

$$f_L(l;\alpha,\beta) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2\alpha^2}\left[\frac{l}{\beta}+\frac{\beta}{l}-2\right]\right)\frac{l\left[\left\{\frac{\beta}{l}\right\}^{\frac{1}{2}}+\left\{\frac{\beta}{l}\right\}^{\frac{3}{2}}\right]}{\alpha[\alpha^2+2]\beta^2}, \quad l > 0. \tag{18}$$

The CDF of L is given by

$$\begin{aligned} F_L(l;\alpha,\beta) = \Phi(a_l) + \frac{\alpha^2}{[2+\alpha^2]} \times \Bigg[\exp\left(\frac{2}{\alpha^2}\right)&\left\{-\Phi\left(-\frac{\sqrt{4+\alpha^2 a_l^2}}{\alpha}\right)\right\} \\ &- \phi(a_l)\left\{a_l + \frac{\sqrt{4+\alpha^2 a_l^2}}{\alpha}\right\}\Bigg], \quad l > 0, \end{aligned} \tag{19}$$

where $a_l = a_l(\alpha,\beta) = [1/\alpha][\sqrt{l/\beta} - \sqrt{\beta/l}]$. The HRF of L can be easily obtained from (18) and (19) as

$$h_L(l;\alpha,\beta) = \frac{f_L(l;\alpha,\beta)}{1-F_L(l;\alpha,\beta)}, \quad l > 0. \tag{20}$$

Using the fact that the BS distribution in (1) is unimodal, it can be shown that the LBS distribution in (18) is also unimodal. The mode of the LBS distribution can be obtained as the solution of the cubic equation

$$\left[\frac{x}{\beta}\right]^3 + \left[\frac{x}{\beta}\right]^2 [1-\alpha^2] - \left[\frac{x}{\beta}\right][1+\alpha^2] - 1 = 0. \tag{21}$$

It can be further shown that the HRF of the LBS distribution obtained from (20) is also unimodal for all values of α and β. A basic property of the distribution is $c\,L \sim \mathrm{LBS}(\alpha, c\beta)$, with $c > 0$. The rth moment of L is given by

$$\begin{aligned}\mathrm{E}[L^r] = \frac{\mathrm{E}[T^{r+1}]}{\mathrm{E}[T]} &= \frac{1}{[\alpha^2+2]}\left[\frac{2^{r+1}\Gamma(r+3/2)\alpha^{2r+2}\beta^r}{\sqrt{\pi}}\right. \\ &\left. - \sum_{k=1}^{r+1}[-1]^k\binom{2r-2}{k}\frac{\mathrm{E}[T^{r-k+1}]}{\beta^{1-k}} - \sum_{k=r+2}^{2r+1}[-1]^k\binom{2r+2}{k}\frac{\mathrm{E}[T^{k-r-1}]}{\beta^{k-2r-1}}\right],\end{aligned} \tag{22}$$

where $T \sim \mathrm{BS}(\alpha, \beta)$ and its moments are as given in (6). In particular, from (22), we obtain

$$\begin{aligned}\mathrm{E}[L] &= \beta\;\frac{[2+4\alpha^2+3\alpha^4]}{2+\alpha^2}, \quad \mathrm{V}[T] = [\alpha\,\beta]^2\frac{[4+17\alpha^2+24\alpha^4+6\alpha^6]}{[2+\alpha^2]^2}, \\ \beta_1[L] &= \frac{3\alpha[8+48\alpha^2+95\alpha^4+48\alpha^6+8\alpha^8]}{\sqrt{[2+9\alpha^2+18\alpha^4+15\alpha^6]^3}}, \quad \text{and} \\ \beta_2[L] &= 3 + \frac{3\alpha^2[80+602\alpha^2+1508\alpha^4+1149\alpha^6+384\alpha^8+48\alpha^{10}]}{[4+17\alpha^2+24\alpha^4+6\alpha^6]^2}.\end{aligned} \tag{23}$$

For more details about the LBS distribution, one may refer to Leiva, Sanhueza and Angulo (2009).

We now briefly describe how the ML estimation of the parameters of the LBS distribution can be done based on a random sample $L_1, \ldots, L_n$, where $L_i \sim \mathrm{LBS}(\alpha, \beta)$, for $i = 1, \ldots, n$. The log-likelihood function is given by

$$\begin{aligned}\ell_{\mathrm{LBS}}(\alpha, \beta) = -n\log(\alpha) - 2n\log(\beta) &+ \sum_{i=1}^{n}\log\left(\left[\frac{\beta}{l_i}\right]^{\frac{1}{2}} + \left[\frac{\beta}{l_i}\right]^{\frac{3}{2}}\right) \\ &- \frac{1}{2\alpha^2}\sum_{i=1}^{n}\left[\frac{l_i}{\beta} + \frac{\beta}{l_i} - 2\right] + \sum_{i=1}^{n}\log(l_i) - n\log(\alpha^2+2).\end{aligned} \tag{24}$$

For a given β, the ML estimator of α, say $\widehat{\alpha}(\beta)$, can be obtained as

$$\widehat{\alpha}(\beta) = \frac{1}{\sqrt{6}}\left[\{A(\beta)-2\} + \sqrt{\{A(\beta)-2\}^2 + 24A(\beta)}\right]^{\frac{1}{2}}, \tag{25}$$

where $A(\beta) = [1/n]\sum_{i=1}^{n}[l_i/\beta + \beta/l_i - 2]$. Thus, the ML estimate of β, say $\widehat{\beta}$, can be obtained by first maximizing the profile log-likelihood function of β, viz., $\ell_{\text{LBS}}(\widehat{\alpha}(\beta), \beta)$. Once $\widehat{\beta}$ is numerically computed, the ML estimate of α, say $\widehat{\alpha}$, can be obtained from (25) as $\widehat{\alpha}(\widehat{\beta})$. Note that the maximization of the profile log-likelihood function $\ell_{\text{LBS}}(\widehat{\alpha}(\beta), \beta)$ can be done by solving the non-linear equation

$$-\frac{3}{\beta} + \frac{1}{n\beta}\sum_{i=1}^{n}\frac{2\,[\beta/l_i]^{\frac{3}{2}}}{[\beta/l_i]^{\frac{1}{2}} + [\beta/l_i]^{\frac{3}{2}}} + \frac{s}{\beta^2\widehat{\alpha}^2(\beta)} - \frac{1}{r\widehat{\alpha}^2(\beta)} = 0, \tag{26}$$

where s and r are defined in (10) and $\widehat{\alpha}(\beta)$ in (25). Once again, observe that (26) can be solved by using a fixed-point type equation $h(\beta) = \beta$, where now

$$h(\beta) = \left[\frac{2\beta}{n}\sum_{i=1}^{n}\frac{[\beta/l_i]^{\frac{3}{2}}}{[\beta/l_i]^{\frac{1}{2}} + [\beta/l_i]^{\frac{3}{2}}} + \frac{s}{\widehat{\alpha}^2(\beta)}\right]\left[3 + \frac{\beta}{r\widehat{\alpha}^2(\beta)}\right]^{-1}. \tag{27}$$

A simple iterative scheme, as detailed in Subsection 2.1, can be used to solve (27).

3 Three new mixture models and their properties

3.1 Model 1

A RV Y has a mixture distribution of two different BS models (MTBS) if its PDF is given by

$$f_Y(y) = p f_{T_1}(y; \alpha_1, \beta_1) + [1-p] f_{T_2}(y; \alpha_2, \beta_2), \quad y > 0, \tag{28}$$

where $T_i \sim \text{BS}(\alpha_i, \beta_i)$, with $\alpha_i > 0, \beta_i > 0$, and $0 \leq p \leq 1$, for $i = 1, 2$. The RV Y with PDF in (28) is said to have a MTBS distribution, which is denoted by $Y \sim \text{MTBS}(\alpha_1, \beta_1, \alpha_2, \beta_2, p)$. The CDF of Y is simply

$$F_Y(y) = p F_{T_1}(y; \alpha_1, \beta_1) + [1-p] F_{T_2}(y; \alpha_2, \beta_2), \quad y > 0, \tag{29}$$

where $F_{T_i}(\cdot;\alpha_i,\beta_i)$, for $i=1,2$, is as given in (2). It appears that the PDF in (28) is either unimodal or bimodal and its shape for some choices of $(\alpha_1,\beta_1,\alpha_2,\beta_2,p)$ are presented in Figure 1.

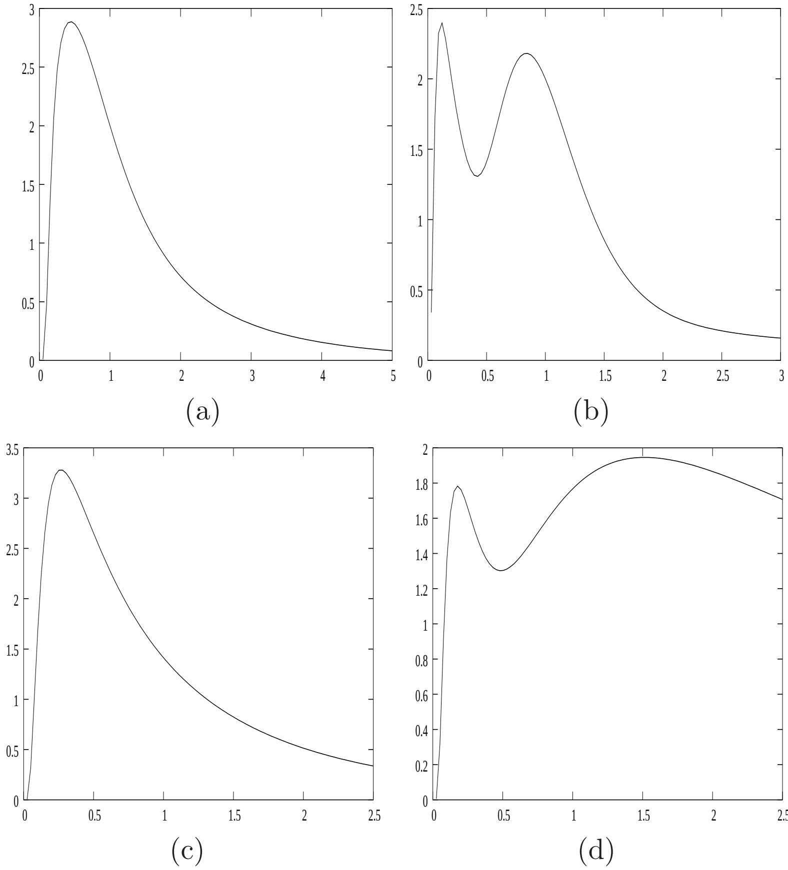

Fig. 1. PDF of the MTBS distribution given in (28) for the choices of $(\alpha_1,\beta_1,\alpha_2,\beta_2,p)$ as (a) (1.0, 1.0, 0.5, 1.0, 0.75), (b) (2.5, 1.0, 0.5, 1.0, 0.25), (c) (1.0, 0.5, 1.0, 1.0, 0.5), and (d) (1.0, 0.5, 0.5, 5.0, 0.25).

The HRF of Y in this case is of the form

$$h_Y(y)=\frac{pf_{T_1}(y;\alpha_1,\beta_1)+[1-p]f_{T_2}(y;\alpha_2,\beta_2)}{1-[pF_{T_1}(y;\alpha_1,\beta_1)+[1-p]F_{T_2}(y;\alpha_2,\beta_2)]},\quad y>0, \tag{30}$$

from which it is clear that it can be expressed as

$$h_Y(y)=w_1(y)h_{T_1}(y)+w_2(y)h_{T_2}(y),\quad y>0, \tag{31}$$

where $h_{T_1}(\cdot)$ and $h_{T_2}(\cdot)$ are HRFs of T_1 and T_2, respectively, and $w_1(\cdot)$ and $w_2(\cdot)$ are non-linear weight functions depending on y. As mentioned, the HRF of the BS distribution is unimodal for all values of α and β. However, due to the complicated form of the HRF of the MTBS distribution, it is difficult to examine theoretically its shape characteristics. Graphically, we have observed that the HRF in (30) is either unimodal or bimodal depending on the values of α_1, β_1, α_2, β_2 and p. For some choices of the parameters, a plot of the HRF is presented in Figure 2.

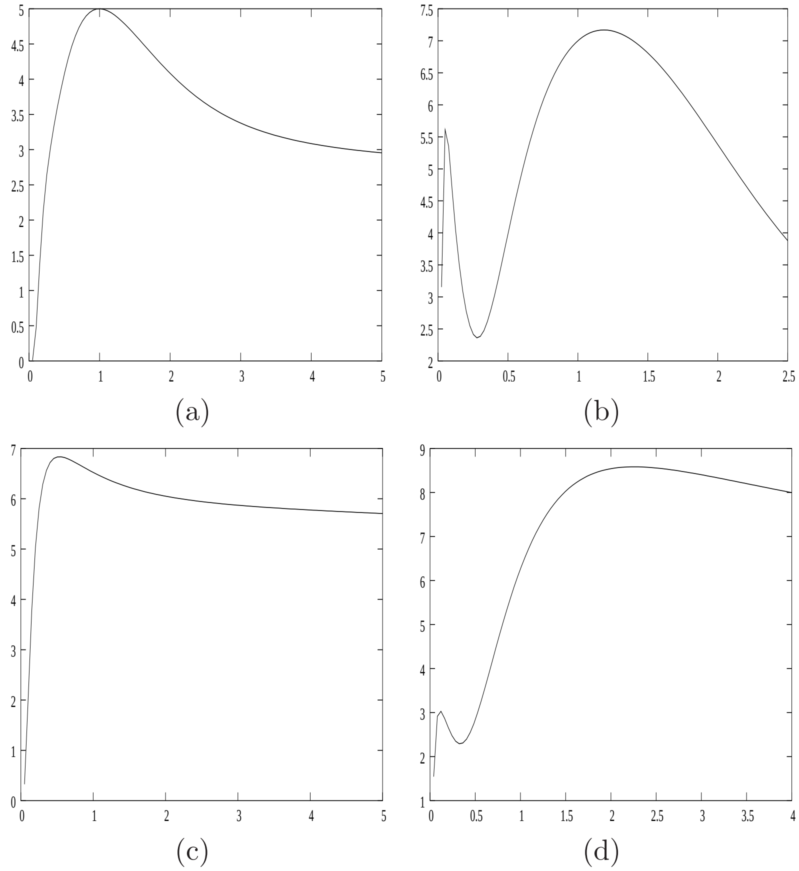

Fig. 2. HRF of the MTBS distribution given in (30) for the choices of $(\alpha_1, \beta_1, \alpha_2, \beta_2, p)$ as (a) (1.0, 1.0, 0.5, 1.0, 0.75), (b) (2.5, 1.0, 0.5, 1.0, 0.25), (c) (1.0, 0.5, 1.0, 1.0, 0.5), and (d) (1.0, 0.5, 0.25, 4.0, 0.75).

For Y having the PDF in (28), the first two raw moments are given by

$$\mathrm{E}[Y] = p\beta_1\left[1+\frac{1}{2}\alpha_1^2\right]+[1-p]\beta_2\left[1+\frac{1}{2}\alpha_2^2\right] \quad \text{and} \tag{32}$$

$$\mathrm{E}[Y^2] = \frac{3}{2}p\beta_1^2\left[\alpha_1^4+2\alpha_1^2+\frac{2}{3}\right]+\frac{3}{2}[1-p]\beta_2^2\left[\alpha_2^4+2\alpha_2^2+\frac{2}{3}\right]. \tag{33}$$

Moreover, it is evident that if Y has PDF as in (28), then $1/Y$ has its PDF as

$$f_{1/Y}(y) = pf_{1/T_1}(y;\alpha_1,\beta_1)+[1-p]f_{1/T_2}(y;\alpha_2,\beta_2), \quad y>0, \tag{34}$$

where $1/T_i \sim \mathrm{BS}(\alpha_i, 1/\beta_i)$, for $i=1,2$. Therefore, $1/Y$ also has a MTBS distribution.

3.2 Model 2

A RV Y has a mixture distribution of BS and LBS models (MBSLBS) if its PDF is given by

$$f_Y(y) = p f_T(y; \alpha_1, \beta_1) + [1-p] f_L(y; \alpha_2, \beta_2), \quad y > 0, \tag{35}$$

where $T \sim \mathrm{BS}(\alpha_1, \beta_1)$ and $L \sim \mathrm{LBS}(\alpha_2, \beta_2)$, with $0 \leq p \leq 1$ being the mixing parameter. The RV Y with PDF in (35) is said to have a MBSLBS distribution, which is denoted by $Y \sim \mathrm{MBSLBS}(\alpha_1, \beta_1, \alpha_2, \beta_2, p)$. The CDF of Y is then

$$F_Y(y) = p F_T(y; \alpha_1, \beta_1) + [1-p] F_L(y; \alpha_2, \beta_2), \quad y > 0, \tag{36}$$

where $F_T(\cdot)$ and $F_L(\cdot)$ are respectively the CDFs of the BS and LBS distributions as given in (2) and (19), respectively. By graphical plots, it has been observed that the PDF of the MBSLBS distribution in (35) is either unimodal or bimodal and its shapes for some choices of the parameters $(\alpha_1, \beta_1, \alpha_2, \beta_2, \rho)$ are presented in Figure 3.

As in the case of Model 1, the HRF of the MBSLBS distribution can also be expressed as a weighted mixture of the HRFs of the BS and LBS distributions, with the weight functions depending on y. The HRF of the MBSLBS distribution can take on different shapes and its plot for some choices of the parameters are presented in Figure 4.

Since the rth moment of the LBS distribution can be expressed in terms of the moments of the BS distribution, the moments of the MBSLBS distribution can be expressed purely in terms of the moments of the BS distribution. Though Model 2 has five parameters and thus provides great flexibility in modeling, we now consider Model 3 which is a special case of Model 2 when $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$, as this simpler form provides adequate fit in some cases (as in the case of depressive condition data in Section 5).

3.3 Model 3

Here, we concentrate on the special case of the MBSLBS distribution when the two distributions share the same parameter values. Such a three-parameter MBSLBS distribution has its PDF as

$$f_Y(y; \alpha, \beta) = [1-p]\ f_T(y; \alpha, \beta) + p\, f_L(y; \alpha, \beta), \quad y > 0, \tag{37}$$

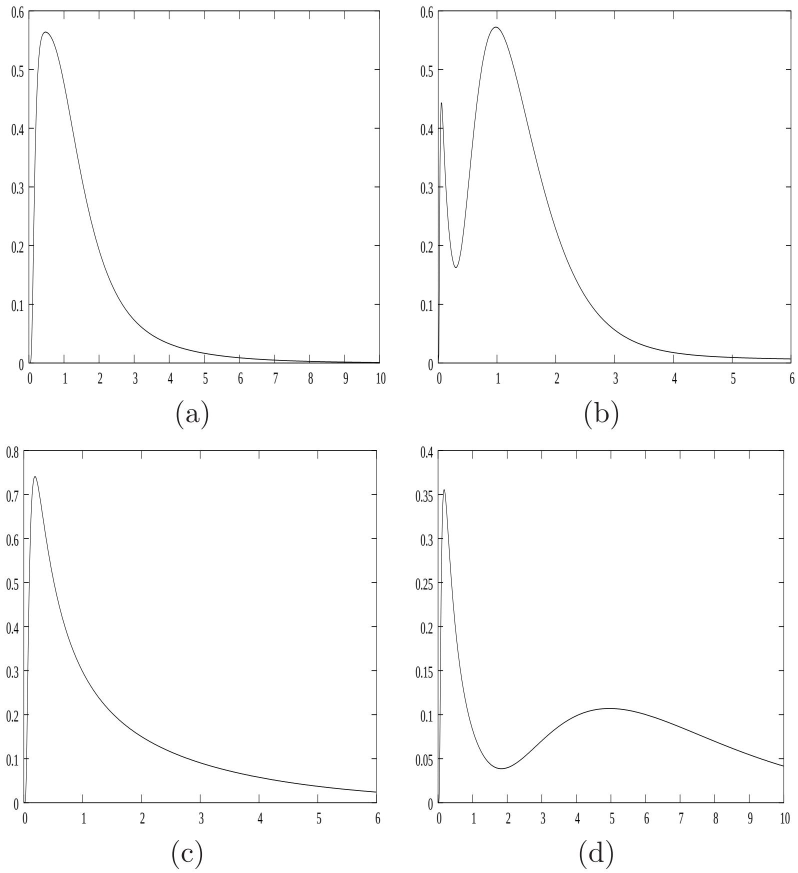

Fig. 3. PDF of the MBSLBS distribution in given (35) for the choices of $(\alpha_1, \beta_1, \alpha_2, \beta_2, p)$ as (a) (1.0, 1.0, 0.5, 1.0. 0.75), (b) (2.5, 1.0, 0.5, 1.0, 0.25), (c) (1.0, 0.5, 1.0, 1.0, 0.5), and (d) (1.0, 0.5, 0.5, 5.0, 0.25).

where the variates T and L are as defined in Model 2. The RV Y with PDF in (37) is said to have a reduced MBSLBS (RMBSLBS) distribution, which is denoted by $Y \sim \text{RMBSLBS}(\alpha, \beta, p)$. Noting that (37) is a weighted version of the BS model with a linear weight function, it can be shown that the PDF of Y is unimodal for all values of the parameters $\alpha > 0$ and $\beta > 0$. The parameter β only modifies the scale, while the parameter α affects the asymmetry and kurtosis of the distribution. The mixing parameter p modifies both scale and kurtosis. The CDF, PDF and HRF of this model can all be easily obtained from those of Model 2. Using the fact that the HRF of the BS distribution is unimodal, it can be shown that the HRF of the RMBSLBS distribution with PDF given in (37) is always unimodal.

If $Y \sim \text{RMBSLBS}(\alpha, \beta, p)$, the following properties can be shown to hold:

(P1) $c\,Y \sim \text{LBS}(\alpha, c\,\beta, p)$, with $c > 0$;
(P2) The RV

$$U = \frac{1}{\alpha^2}\left[\frac{Y}{\beta} + \frac{\beta}{Y} - 2\right]$$

is a mixture of two gamma distributions with mixing parameter w, i.e., the

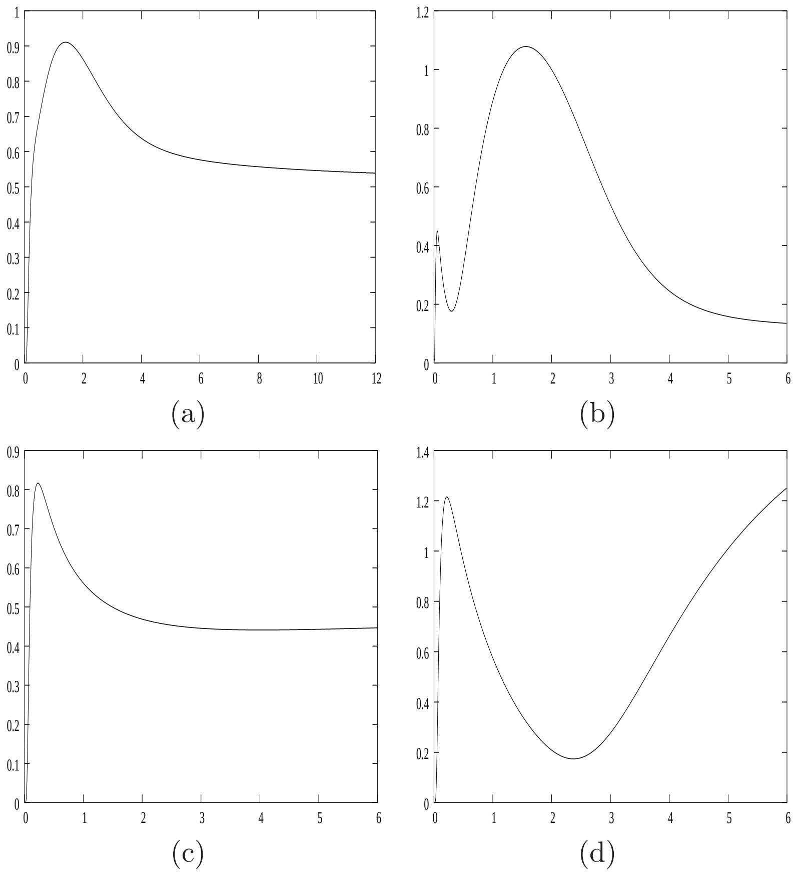

Fig. 4. HF of the MBSLBS distribution obained from (35) and (36) for the choices of $(\alpha_1, \beta_1, \alpha_2, \beta_2, p)$ as (a) (1.0, 1.0, 0.5, 1.0. 0.75), (b) (2.5, 1.0, 0.5, 1.0, 0.25), (c) (1.0, 0.5, 1.0, 1.0, 0.5), and (d) (1.0, 0.5, 0.5, 5.0, 0.25).

PDF of U can be expressed as

$$f_U(u) = [1 - w]f_{U_1}(u) + w f_{U_2}(u), \quad u > 0,$$

where $w = p\alpha^2/[\alpha^2 + 2]$, $U_1 \sim \text{Gamma}(1/2, 2)$ with $\mathrm{E}[U_1] = 1$, and $U_2 \sim \text{Gamma}(3/2, 2)$ with $\mathrm{E}[U_2] = 3$. Hence, $\mathrm{E}[U] = 1 + 2w$.

4 Estimation via EM algorithm

4.1 Model 1

In this subsection, we discuss the ML estimation of the parameters of Model 1 with PDF in (28), assuming that all the parameters are different. For ease in notation, we denote $p_1 = p$ and $p_2 = 1 - p$. Now, we assume a random sample of size n from the PDF in (28), denoted by $\{Y_1, \ldots, Y_n\}$. For facilitating the

EM algorithm, we first use (16) and rewrite the PDF in (28) as

$$f_Y(y) = \frac{1}{2}\sum_{j=1}^{2} p_j f_{X_1}(y;\mu_j,\lambda_j) + \frac{1}{2}\sum_{j=1}^{2} p_j f_{X_2}(y;\mu_j,\lambda_j), \quad t>0, \tag{38}$$

where $f_{X_1}(\cdot)$ and $f_{X_2}(\cdot)$ are as defined in (14) and (17), respectively, with $\mu_j = \beta_j$ and $\lambda_j = \beta_j/\alpha_j^2$, for $j = 1, 2$.

It is well known that a mixture model can be treated as a missing value problem and so the EM algorithm can be used to efficiently compute the ML estimates of the model parameters; see McLachlan and Peel (2000, pp. 47-50). For this purpose, we now assume that the complete observations are as follows. For Y being the RV with PDF as in (38), we then define an associated random vector $W = (U_1, V_1, U_2, V_2)$ as follows. Here, each U_j and V_j can take on values 0 or 1, with $\sum_{j=1}^{2}[U_j + V_j] = 1$, where $\mathrm{P}(U_j = 1) = \mathrm{P}(V_j = 1) = p_j/2$, for $j = 1, 2$. Moreover, $Y|\,(U_j = 1)$ and $Y|\,(V_j = 1)$ have densities $f_{X_1}(\cdot;\ \mu_j,\lambda_j)$ and $f_{X_2}(\cdot;\ \mu_j,\lambda_j)$, respectively, for $j = 1, 2$. These facts yield the joint distribution of (Y, W). Now, if we have the complete observations (y_i, w_i), where $w_i = (u_{i1}, v_{i1}, u_{i2}, v_{i2})$, for $i = 1, \ldots, n$, then with $\boldsymbol{\theta} = (\mu_1, \lambda_1, p_1, \mu_2, \lambda_2, p_2)$ being the parameter vector, the complete data log-likelihood function is given by

$$\begin{aligned}\ell^{(c)}(\boldsymbol{\theta}|y,w) = c &+ \sum_{j=1}^{2}[u_{\cdot j} + v_{\cdot j}]\log(p_j) + \sum_{i=1}^{n}\sum_{j=1}^{2} u_{ij}\log\left(f_{X_1}(y_i;\mu_j,\lambda_j)\right)\\ &+ \sum_{i=1}^{n}\sum_{j=1}^{2} v_{ij}\log\left(f_{X_2}(y_i;\mu_j,\lambda_j)\right),\end{aligned} \tag{39}$$

where $u_{\cdot j} = \sum_{i=1}^{n} u_{ij}$, $v_{\cdot j} = \sum_{i=1}^{n} v_{ij}$, and c is a constant independent of the parameter vector $\boldsymbol{\theta}$. Therefore, for the complete data case, the ML estimators of the unknown parameters can be obtained as follows. For p_j, we have

$$\widehat{p}_j = \frac{u_{\cdot j} + v_{\cdot j}}{n}, \tag{40}$$

and for μ_j and λ_j, we must maximize

$$\sum_{i=1}^{n} u_{ij}\log(f_{X_1}(y_i;\mu,\lambda)) + \sum_{i=1}^{n} v_{ij}\log(f_{X_2}(y_i;\mu,\lambda)) \tag{41}$$

with respect to μ and λ, respectively. The function in (41) can be shown to be

$$[u_{\cdot j} + v_{\cdot j}]\frac{1}{2}\log(\lambda) - \lambda\left[\sum_{i=1}^{n}[u_{ij} + v_{ij}]\frac{[y_i - \mu]^2}{2\mu_j^2 y_i}\right] - v_{\cdot j}\log(\mu), \tag{42}$$

and so the ML estimators of μ_j and λ_j can be obtained by maximizing (42) with respect to μ and λ, respectively. After some algebraic calculations, it can be seen that these ML estimators are given by

$$\widehat{\mu}_j = \frac{B_j[A_j - B_j] + \sqrt{[B_j\{A_j - B_j\}]^2 + A_j C_j D_j [2B_j - A_j]}}{D_j A_j} \quad \text{and} \tag{43}$$

$$\widehat{\lambda}_j = \frac{[u_{\cdot j} + v_{\cdot j}]\widehat{\mu}_j^2}{\sum_{i=1}^{n}[u_{ij} + v_{ij}][y_i - \mu_j]^2/y_i}, \tag{44}$$

where $A_j = v_{\cdot j}$, $B_j = [1/2][u_{\cdot j} + v_{\cdot j}]$, $C_j = [1/2]\sum_{i=1}^{n} y_i[u_{ij} + v_{ij}]$, and $D_j = [1/2]\sum_{i=1}^{n}[u_{ij} + v_{ij}]/y_i$. Thus, in the case of complete data, the ML estimators of the parameters can all be obtained explicitly. Now, we are ready to implement the EM algorithm. The existence of explicit expressions of the ML estimators simplifies the implementation of the EM algorithm as well as make it computationally efficient.

Suppose, at the mth stage of the EM algorithm, the estimate of the unknown parameter $\boldsymbol{\theta}$ is $\boldsymbol{\theta}^{(m)} = (\mu_1^{(m)}, \lambda_1^{(m)}, p_1^{(m)}, \mu_2^{(m)}, \lambda_2^{(m)}, p_2^{(m)})$. We now describe how the next iterate $\boldsymbol{\theta}^{(m+1)}$ can be obtained.

At the E-step of the EM algorithm, the pseudo log-likelihood function is formed by replacing the missing values with their expectation. Since in this case u_{ij}'s and v_{ij}'s are missing, the pseudo log-likelihood function at the mth stage is obtained from (39) by replacing u_{ij} and v_{ij} with $a_{ij}^{(m)}$ and $b_{ij}^{(m)}$, respectively, where

$$a_{ij}^{(m)} = \mathrm{E}[U_{ij}|y, \boldsymbol{\theta}^{(m)}] \quad \text{and} \quad b_{ij}^{(m)} = \mathrm{E}[V_{ij}|y, \boldsymbol{\theta}^{(m)}]. \tag{45}$$

Note that, here, $a_{ij}^{(m)}$ and $b_{ij}^{(m)}$ are the usual posterior probabilities associated with the ith observation, which are given by

$$a_{ij}^{(m)} = \frac{p_j^{(m)} f_{X_1}(y_i; \mu_j^{(m)}, \lambda_j^{(m)})}{\sum_{l=1}^{k} p_l^{(m)} f_{X_1}(y_i; \mu_l^{(m)}, \lambda_l^{(m)}) + \sum_{l=1}^{k} p_l^{(m)} f_{X_2}(y_i; \mu_l^{(m)}, \lambda_l^{(m)})} \text{ and} \tag{46}$$

$$b_{ij}^{(m)} = \frac{p_j^{(m)} f_{X_2}(y_i; \mu_j^{(m)}, \lambda_j^{(m)})}{\sum_{l=1}^{k} p_l^{(m)} f_{X_1}(y_i; \mu_l^{(m)}, \lambda_l^{(m)}) + \sum_{l=1}^{k} p_l^{(m)} f_{X_2}(y_i; \mu_l^{(m)}, \lambda_l^{(m)})}, \tag{47}$$

respectively; see McLachlan and Peel (2000, pp. 48-49). Consequently, the pseudo log-likelihood function at the mth stage can be expressed as

$$\ell^{(\mathrm{ps})}(\boldsymbol{\theta}|y,\boldsymbol{\theta}^{(m)}) = \sum_{i=1}^{n}\sum_{j=1}^{2} a_{ij}^{(m)} \log\left(\frac{p_j}{2} f_{X_1}(y_i;\mu_j,\lambda_j)\right) + \sum_{i=1}^{n}\sum_{j=1}^{2} b_{ij}^{(m)} \log\left(\frac{p_j}{2} f_{X_2}(y_i;\mu_j,\lambda_j)\right). \tag{48}$$

The M-step of the EM algorithm involves the maximization of (48) with respect to the unknown parameters. Let us now denote $a_{\cdot j}^{(m)} = \sum_{i=1}^{n} a_{ij}^{(m)}$, $b_{\cdot j}^{(m)} = \sum_{i=1}^{n} b_{ij}^{(m)}$, $A_j^{(m)} = b_{\cdot j}^{(m)}$, $B_j^{(m)} = [1/2][a_{\cdot j}^{(m)} + b_{\cdot j}^{(m)}]$, $C_j^{(m)} = [1/2]\sum_{i=1}^{n} y_i[a_{ij}^{(m)} + b_{ij}^{(m)}]$, and $D_j^{(m)} = [1/2]\sum_{i=1}^{n}[a_{ij}^{(m)} + b_{ij}^{(m)}]/y_i$. Then, the elements of the vector $\boldsymbol{\theta}^{(m+1)}$ can be obtained as

$$p_j^{(m+1)} = \frac{a_{\cdot j}^{(m)} + b_{\cdot j}^{(m)}}{n}, \tag{49}$$

$$\mu_j^{(m+1)} = \frac{[\{B_j^{(m)}\}^2 - A_j^{(m)} B_j^{(m)}]}{D_j^{(m)}[B_j^{(m)} - A_j^{(m)}]} \tag{50}$$

$$+ \frac{\sqrt{[\{B_j^{(m)}\}^2 - A_j^{(m)} B_j^{(m)}]^2 + A_j^{(m)} C_j^{(m)} D_j^{(m)}[B_j^{(m)} - A_j^{(m)}]}}{D_j^{(m)}[B_j^{(m)} - A_j^{(m)}]}, \tag{51}$$

$$\lambda_j^{(m+1)} = \frac{[a_{\cdot j}^{(m)} + b_{\cdot j}^{(m)}][\mu_j^{(m+1}]^2}{\sum_{i=1}^{n}[a_{ij}^{(m)} + b_{ij}^{(m)}][y_i - \mu_j^{(m+1)}]^2/y_i}. \tag{52}$$

4.2 Model 2

In this subsection, we develop the ML estimation of the parameters of Model 2 with PDF in (35). Let us assume a random sample of size n from the PDF in (35), denoted by $\{Y_1,\ldots,Y_n\}$. Based on this random sample, we wish to estimate the five model parameters, viz., $\alpha_1,\beta_1,\alpha_2,\beta_2,p$. It is evident that if we want to maximize the log-likelihood function directly with respect to the unknown parameters, we would need to solve a five-dimensional optimization problem, which would be computationally quite involved. On the other hand, it has already been mentioned that the computation of the ML estimates for both BS and LBS distributions involves solving only one non-linear equation each. We now treat this problem as a standard missing value problem. The complete observations are of the form (Y,Δ), where Δ is a binary random variable taking on the values 0 or 1. If $\Delta = 1$, it means that the corresponding Y arises from the BS distribution and if $\Delta = 0$ it means that Y arises from the LBS distribution. Moreover, $\mathrm{P}(\Delta = 1) = p$ and $\mathrm{P}(\Delta = 0) = 1 - p$. If the complete observations, say $\{(y_1,\delta_1),\ldots,(y_n,\delta_n)\}$, are available, then the complete data log-likelihood function is given by

$$
\begin{aligned}
\ell^{(c)}(\boldsymbol{\theta}|y,\delta) = &\log(p)\sum_{i=1}^{n}\delta_i + [\log(1-p)]\sum_{i=1}^{n}[1-\delta_i] \\
&+\sum_{i=1}^{n}\delta_i \log\left(f_T(y_i;\alpha_1,\beta_1)\right) + \sum_{i=1}^{n}[1-\delta_i]\log\left(f_L(y_i;\alpha_2,\beta_2)\right),
\end{aligned}
\tag{53}
$$

where now $\boldsymbol{\theta} = (\alpha_1, \beta_1, \alpha_2, \beta_2, p)$ and (y, δ) denotes the complete data vector. The ML estimator of $\boldsymbol{\theta}$ can be obtained by maximizing (53) with respect to its elements. First, we obtain the ML estimator of p based on the complete sample as

$$
\widehat{p}^{(c)} = \frac{1}{n}\sum_{i=1}^{n}\delta_i = \bar{\delta}. \tag{54}
$$

The ML estimators of α_1 and β_1 based on the complete sample, say $\widehat{\alpha}_1^{(c)}$ and $\widehat{\beta}_1^{(c)}$, can be obtained by maximizing $\sum_{i=1}^{n}\delta_i \log\left(f_T(y_i;\alpha_1,\beta_1)\right)$ with respect to α_1 and β_1, respectively. It may be easily seen that $\widehat{\beta}_1^{(c)}$ can be obtained as the solution of $h^{(c)}(\beta) = \beta$, where

$$
\begin{aligned}
h^{(c)}(\beta) &= \left[\beta\bar{\delta} + \frac{2}{n}\sum_{i=1}^{n}\frac{\beta\delta_i\left\{\frac{\beta}{y_i}\right\}^{\frac{3}{2}}}{\left\{\frac{\beta}{y_i}\right\}^{\frac{1}{2}} + \left\{\frac{\beta}{y_i}\right\}^{\frac{3}{2}}} + \frac{\sum_{i=1}^{n}\delta_i y_i}{n\{\widehat{\alpha}_1^{(c)}(\beta)\}^2}\right]\left[2\bar{\delta} + \frac{\beta}{n\{\widehat{\alpha}_1^{(c)}(\beta)\}^2}\sum_{i=1}^{n}\frac{\delta_i}{y_i}\right]^{-1}, \\
\widehat{\alpha}_1^{(c)}(\beta) &= \left[\frac{1}{n\bar{\delta}}\sum_{i=1}^{n}\delta_i\left\{\frac{y_i}{\beta} + \frac{\beta}{y_i} - 2\right\}\right]^{\frac{1}{2}}.
\end{aligned}
\tag{55}
$$

We can use an iterative scheme similar to the one as suggested in Subsection 2.1 for computing $\widehat{\beta}_1^{(c)}$ from (55). Once $\widehat{\beta}_1^{(c)}$ is obtained, we can readily compute $\widehat{\alpha}_1^{(c)} = \widehat{\alpha}_1^{(c)}(\widehat{\beta}_1^{(c)})$.

Similarly, the ML estimates of α_2 and β_2 based on the complete sample, say $\widehat{\alpha}_2^{(c)}$ and $\widehat{\beta}_2^{(c)}$, can be obtained by maximizing $\sum_{i=1}^{n}[1-\delta_i]\log\left(f_L(y_i;\alpha_2,\beta_2)\right)$ with respect to α_2 and β_2, respectively. In this case, $\widehat{\beta}_2^{(c)}$ can be computed as the solution of $h^{(c)}(\beta) = \beta$, where

$$
\begin{aligned}
h^{(c)}(\beta) = &\left[\frac{2\beta}{n}\sum_{i=1}^{n}\frac{\{1-\delta_i\}\left\{\frac{\beta}{y_i}\right\}^{\frac{3}{2}}}{\left\{\frac{\beta}{y_i}\right\}^{\frac{1}{2}} + \left\{\frac{\beta}{y_i}\right\}^{\frac{3}{2}}} + \frac{1}{n\{\widehat{\alpha}_2^{(c)}(\beta)\}^2}\sum_{i=1}^{n}\{1-\delta_i\}y_i\right] \\
&\times\left[3\{1-\bar{\delta}\} + \frac{\beta}{n\{\widehat{\alpha}_2^{(c)}(\beta)\}^2}\sum_{i=1}^{n}\frac{1-\delta_i}{y_i}\right]^{-1},
\end{aligned}
\tag{56}
$$

$$
\widehat{\alpha}_2^{(c)}(\beta) = \frac{1}{\sqrt{6[1-\bar{\delta}]}}\left[\{\widetilde{A}(\beta) - 2\} + \sqrt{\{\widetilde{A}(\beta)-2\}^2 + 24\bar{\delta}\widetilde{A}(\beta)}\right]^{\frac{1}{2}}, \tag{57}
$$

$$\widetilde{A}(\beta) = \frac{1}{n}\sum_{i=1}^{n}[1-\delta_i]\left[\frac{y_i}{\beta}+\frac{\beta}{y_i}-1\right]. \tag{58}$$

By employing an iterative scheme similar to the one as suggested earlier for Model 1, we can compute $\widehat{\beta}_2^{(\mathrm{c})}$ from (56). Once $\widehat{\beta}_2^{(\mathrm{c})}$ is obtained, we can readily compute $\widehat{\alpha}_2^{(\mathrm{c})} = \widehat{\alpha}_2^{(\mathrm{c})}(\widehat{\beta}_2^{(\mathrm{c})})$.

Now, we are ready to implement the EM algorithm for the computation of the ML estimators of the parameters of the MBSLBS model. Suppose, at the mth stage of the EM algorithm, the estimate of the parameter vector $\boldsymbol{\theta}$ is $\boldsymbol{\theta}^{(m)} = (\alpha_1^{(m)}, \beta_1^{(m)}, \alpha_2^{(m)}, \beta_2^{(m)}, p^{(m)})$. Then, at this stage of the EM algorithm, the E-step can be formed by replacing the missing values with their expectation in the complete log-likelihood function, thus forming the pseudo log-likelihood function. Since δ_i is missing, we replace it by $\delta_i^{(m)}$, where

$$\delta_i^{(m)} = \frac{p^{(m)} f_T(y_i; \alpha_1^{(m)}, \beta_1^{(m)})}{p^{(m)} f_T(y_i; \alpha_1^{(m)}, \beta_1^{(m)}) + [1-p^{(m)}] f_L(y_i; \alpha_2^{(m)}, \beta_2^{(m)})}, \tag{59}$$

which is the conditional posterior expectation of Δ_i given the observation and $\boldsymbol{\theta}^{(m)}$; see McLachlan and Peel (2000). Once the pseudo log-likelihood function is formed, at the M-step, we maximize it to obtain $\boldsymbol{\theta}^{(m+1)}$. Therefore, $\alpha_1^{(m+1)}$, $\beta_1^{(m+1)}$, $\alpha_2^{(m+1)}$, $\beta_2^{(m+1)}$, $p^{(m+1)}$ are obtained as $\widehat{\alpha}_1^{(\mathrm{c})}$, $\widehat{\beta}_1^{(\mathrm{c})}$, $\widehat{\alpha}_2^{(\mathrm{c})}$, $\widehat{\beta}_2^{(\mathrm{c})}$, $\widehat{p}^{(\mathrm{c})}$, by replacing δ_i with $\delta_i^{(m)}$, where $\widehat{\alpha}_1^{(\mathrm{c})}$, $\widehat{\beta}_1^{(\mathrm{c})}$, $\widehat{\alpha}_2^{(\mathrm{c})}$, $\widehat{\beta}_2^{(\mathrm{c})}$, $\widehat{p}^{(\mathrm{c})}$ are as defined earlier in this subsection.

4.3 Model 3

In this subsection, we develop the ML estimation of the parameters of Model 3 with PDF in (37). Based on a random sample of size n from the PDF in (37), denoted by $\{Y_1, \ldots, Y_n\}$, we wish to estimate the parameter vector $\boldsymbol{\theta}$ now composed by $\boldsymbol{\theta} = (\alpha, \beta, p)$. In this case as well, we define Δ exactly in the same way as in Model 2. Thus, the complete data log-likelihood becomes

$$\begin{aligned}\ell^{(\mathrm{c})}(\boldsymbol{\theta}|y,\delta) = &-n\log(\alpha) - n\log(\alpha^2+2) - \frac{1}{2\alpha^2}\sum_{i=1}^{n}\left[\frac{y_i}{\beta}+\frac{\beta}{y_i}-2\right] \\ &- \left[2n - \sum_{i=1}^{n}\delta_i\right]\log(\beta) + \sum_{i=1}^{n}\log\left(\left[\frac{\beta}{y_i}\right]^{\frac{1}{2}} + \left[\frac{\beta}{y_i}\right]^{\frac{3}{2}}\right) \\ &+ \sum_{i=1}^{n}\delta_i\log(p) + \sum_{i=1}^{n}[1-\delta_i]\log(1-p).\end{aligned} \tag{60}$$

The ML estimator of p for the complete data sample, say $\widehat{p}^{(\mathrm{c})}$, is the same as before, i.e., $\widehat{p}^{(\mathrm{c})} = \frac{1}{n}\sum_{i=1}^{n}\delta_i$. For fixed β, the ML estimator of α for the complete data set, say $\widehat{\alpha}^{(\mathrm{c})}(\beta)$, is the same as in (25). The estimator of β for the complete sample, say $\widehat{\beta}^{(\mathrm{c})}$, can then be obtained by maximizing the profile complete log-likelihood function $\ell^{(\mathrm{c})}(\widehat{\alpha}^{(\mathrm{c})}(\beta), \beta, \widehat{p})$ with respect to β. Note that the maximization of the profile log-likelihood function $\ell_{\mathrm{LBS}}(\widehat{\alpha}^{(\mathrm{c})}(\beta), \beta)$ can be obtained by solving the non-linear equation

$$-\frac{3-\widehat{p}^{(\mathrm{c})}}{\beta} + \frac{1}{n\beta}\sum_{i=1}^{n}\frac{2\left[\frac{\beta}{y_i}\right]^{\frac{3}{2}}}{\left[\frac{\beta}{y_i}\right]^{\frac{1}{2}} + \left[\frac{\beta}{y_i}\right]^{\frac{3}{2}}} + \frac{s}{\beta^2[\widehat{\alpha}^{(\mathrm{c})}(\beta)]^2} - \frac{1}{r[\widehat{\alpha}^{(\mathrm{c})}(\beta)]^2} = 0, \quad (61)$$

where s and r are as defined earlier in (10). It is of interest to observe that (61) can also be solved by solving a fixed-point type equation $\beta = q(\beta)$, where

$$q(\beta) = \left[\frac{2\beta}{n}\sum_{i=1}^{n}\frac{\left\{\frac{\beta}{y_i}\right\}^{\frac{3}{2}}}{\left\{\frac{\beta}{y_i}\right\}^{\frac{1}{2}} + \left\{\frac{\beta}{y_i}\right\}^{\frac{3}{2}}} + \frac{s}{\{\widehat{\alpha}^{(\mathrm{c})}(\beta)\}^2}\right]\left[\frac{\beta}{r\{\widehat{\alpha}^{(\mathrm{c})}(\beta)\}^2} + \{3-\widehat{p}^{(\mathrm{c})}\}\right]^{-1}. (62)$$

Simple iterative scheme, such as the one outlined for Models 1 and 2, can be used to solve the fixed-point type equation in (62). Once $\widehat{\beta}^{(\mathrm{c})}$ is obtained, the ML estimate of α for complete data set can be obtained as $\widehat{\alpha}^{(\mathrm{c})} = \widehat{\alpha}^{(\mathrm{c})}(\widehat{\beta}^{(\mathrm{c})})$, where $\widehat{\alpha}^{(\mathrm{c})}(\beta)$ is the same as defined in (25). Now, we are ready to implement the EM algorithm for the computation of the ML estimates of the parameters of the RMBSLBS model. Suppose, at the mth stage of the EM algorithm, the estimates of the parameter vector $\boldsymbol{\theta}$ is $\boldsymbol{\theta}^{(m)} = (\alpha^{(m)}, \beta^{(m)}, p^{(m)})$. At the mth stage of the EM algorithm, the E-step is formed by replacing the missing value of δ_i with its expectation in the complete data log-likelihood function, thus forming the pseudo log-likelihood function. In this case, the missing δ_i is replaced by its conditional expectation $\delta_i^{(m)}$ given by

$$\delta_i^{(m)} = \frac{p^{(m)} f_T(y_i; \alpha^{(m)}, \beta^{(m)})}{p^{(m)} f_T(y_i; \alpha^{(m)}, \beta^{(m)}) + [1-p^{(m)}] f_L(y_i; \alpha^{(m)}, \beta^{(m)})}. \quad (63)$$

Once the pseudo log-likelihood function is formed at the M-step, we maximize it to obtain $\boldsymbol{\theta}^{(m+1)}$. Hence, $\alpha^{(m+1)}$, $\beta^{(m+1)}$, $p^{(m+1)}$ are obtained as $\widehat{\alpha}^{(\mathrm{c})}$, $\widehat{\beta}^{(\mathrm{c})}$, $\widehat{p}^{(\mathrm{c})}$, by replacing δ_i with $\delta_i^{(m)}$, where $\widehat{\alpha}^{(\mathrm{c})}$, $\widehat{\beta}^{(\mathrm{c})}$, $\widehat{p}^{(\mathrm{c})}$ are as defined earlier in this subsection.

5 Analysis of two data sets

In this section, for the purpose of illustration, we analyze two data sets by using the three mixture models proposed in the preceding sections.

5.1 Enzyme data

These data are available at `http://www.stats.bris.ac.uk/~peter/mixdata` and correspond to enzymatic activity in the blood. The data set represents the metabolism of carcinogenic substances among 245 unrelated individuals. These data have been analyzed earlier by Bechtel et al. (1993). They observed that a mixture of two skewed distributions is suitable for analyzing enzyme data. We first present the histogram of the data in Figure 5, from which it is possible to observe that the underlying distribution can be a mixture of two distributions. We fit here the three proposed mixture models.

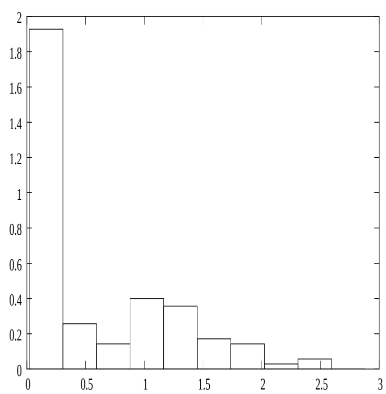

Fig. 5. Histogram of the enzyme data set.

Model 1 In order to fit the MTBS distribution to enzyme data, we need an initial guess of the parameters to start the EM algorithm. For this purpose, we use the method of Finch, Mendell and Thode (1989), which is as follows. Choose p_1 uniformly on $(0, 1)$, say $p_1^{(0)}$. Order the observations as $y_{(1)}, \ldots, y_{(n)}$. Now, based on the subsets $\{y_{(1)}, \ldots, y_{(m)}\}$ and $\{y_{(m+1)}, \ldots, y_{(n)}\}$, estimate (α_1, β_1) and (α_2, β_2), respectively, where m is the integer part of $np_1^{(0)}$. For the initial guess of the parameters α_j and β_j, for $j = 1, 2$, we utilize the modified moment estimates of α and β proposed by Ng, Kundu and Balakrishnan (2003), which are given explicitly as

$$\widetilde{\alpha}_j = \left[2\left\{\sqrt{\frac{s_j}{r_j}} - 1\right\}\right]^{\frac{1}{2}} \quad \text{and} \quad \widetilde{\beta}_j = [s_j r_j]^{\frac{1}{2}}, \quad j = 1, 2,$$

where

$$s_1 = \frac{1}{m}\sum_{i=1}^{m} y_{(i)}, \quad s_2 = \frac{1}{n-m}\sum_{i=m+1}^{n} y_{(i)},$$

$$r_1 = \left[\frac{1}{m}\sum_{i=1}^{m}\frac{1}{y_{(i)}}\right]^{-1}, \quad \text{and} \quad r_2 = \left[\frac{1}{n-m}\sum_{i=m+1}^{n}\frac{1}{y_{(i)}}\right]^{-1}.$$

For different choices of p_1, we perform the EM algorithm and, if the relative absolute difference of the log-likelihood values between two consecutive EM steps is less than 10^{-4}, we stop the EM algorithm. Through this approach, we determine the ML estimates of the model parameters as $\widehat{\alpha}_1 = 0.5325$, $\widehat{\beta}_1 = 0.1747$, $\widehat{\alpha}_2 = 0.3187$, $\widehat{\beta}_2 = 1.2736$, $\widehat{p} = 0.6289$, and the maximized log-likelihood (MLL) value to be -59.1681. The corresponding approximate 95% confidence intervals are found to be $(0.3011, 0.7639)$, $(0.0906, 0.2588)$, $(0.1762, 0.4612)$, $(0.7047, 1.8425)$, and $(0.3978, 0.8600)$, respectively.

Model 2 In order to fit the MBSLBS distribution to enzyme data, we employ the following method to obtain initial guess values for the unknown parameters. We choose some guess value of p, say $0 < p^{(0)} < 1$, and then for each data point we assign group 1 (BS) with probability $p^{(0)}$ and group 2 (LBS) with probability $1-p^{(0)}$. Next, based on the group 1 observations, we estimate the parameters α_1 and β_1 and those are the guess values of α_1 and β_1. Similarly, based on the group 2 observations, we obtain the guess values of the parameters α_2 and β_2. Using these initial guess values and for different choices of p, we perform the EM algorithm and adopted the same stopping criterion as in Model 1. The ML estimates so determined for this model are $\widehat{\alpha}_1 = 0.3652$, $\widehat{\beta}_1 = 0.1705$, $\widehat{\alpha}_2 = 1.2738$, $\widehat{\beta}_2 = 0.2125$, $\widehat{p} = 0.4496$, and the MLL value is -71.0908. The corresponding approximate 95% confidence intervals are found to be $(0.1974, 0.5331)$, $(0.0894, 0.2516)$, $(0.6824, 1.8652)$, $(0.1124, 0.3126)$, and $(0.2599, 0.6393)$, respectively.

Model 3 In order to fit the RMBSLBS distribution to enzyme data, the initial estimates for the parameters were determined in the same way as in Model 2. First, we obtain the initial estimates of α_1 and α_2 and then took the initial estimate of α as the average of these two estimates. Similarly, we also obtain the initial estimate of β and adopted the same stopping criterion as stated above. The ML estimates so determined for this model are $\widehat{\alpha} = 1.0375$, $\widehat{\beta} = 0.2163$, $\widehat{p} = 0.4166$, and the MLL value is -115.899. The corresponding approximate 95% confidence intervals are found to be $(0.5767; 1.4923)$, $(0.1252; 0.3074)$, and $(0.2255; 0.6077)$, respectively.

Model checking The natural question that arises now is how good are the fits of these models to the enzyme data set. For this purpose, we calculate the Kolmogorov-Smirnov (KS) distance between the empirical CDF and the estimated theoretical CDF for the three models. The KS distances for Models 1, 2 and 3 turn out to be 0.053, 0.111 and 0.151, with the associated p-values as 0.507, 0.005, and < 0.0001, respectively. The empirical survival function (ESF) and the fitted survival function (FSF) for the three models are presented in Figure 6. Based on the KS distances and the associated p-values, we conclude that Model 1 provides the best fit to the enzyme data.

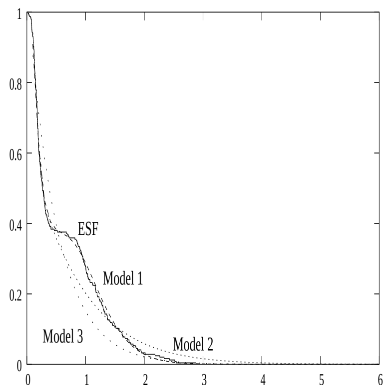

Fig. 6. ESF and FSF of the three mixture models for the enzyme data.

5.2 Depressive condition data

The scale "general rating of affective symptoms for pre-schoolers" (GRASP) measures behavioral and emotional problems of children, who can be classified with depressive condition or not according to this scale. A study conducted by Dr. Nelson Araneda from the University of La Frontera, Temuco, Chile, and the authors, in a city located at the southern part of Chile resulted in real data corresponding to the scores of the GRASP scale of children, which are as follows (frequency in parentheses and none when it is one): 19(16), 20(15), 21(14), 22(9), 23(12), 24(10), 25(6), 26(9), 27(8), 28(5), 29(6), 30(4), 31(3), 32(4), 33, 34, 35(4), 36(2), 37(2), 39, 42, 44. After subtracting the value 16 from each observation, we fit the three mixture models for these data. The histogram of depressive condition data is presented in Figure 7.

By fitting the three mixture models to depressive condition data and using the same methods as applied to enzyme data, the ML estimates of the model parameters, MLL values and KS distances with their associated p-values for all three models are determined. These results are presented in Table 2. We have also plotted the ESF and the FSF for all three models in Figure 8. From the p-values, we conclude that all three models provide good fit to depressive condition data. However, it is of interest to observe that Model 3 yields the shortest KS distance and the largest p-value. Therefore, based on the KS distance, we conclude that Model 3 provides the best fit to these data.

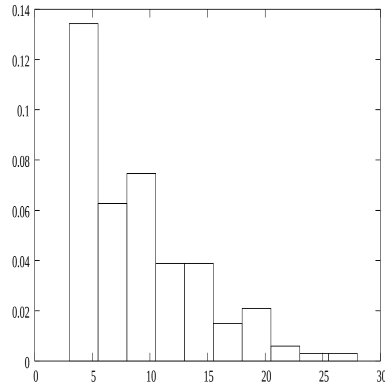

Fig. 7. Histogram of the depressive condition data.

Table 2
ML estimates, MLL and KS distances with their associated p-values for the three mixture models fitted to depressive condition data.

	Estimates					Goodness-of-fit		
Model	α_1	β_1	α_2	β_2	p	MLL	KS	p-value
1	0.5792	7.3003	0.1349	20.2561	0.9631	−387.522	0.091	0.218
2	0.6031	7.5789	61.4061	0.0010	0.9966	−388.094	0.092	0.206
3	0.5976	5.6647	0.5976	5.6647	0.1756	−388.726	0.085	0.283

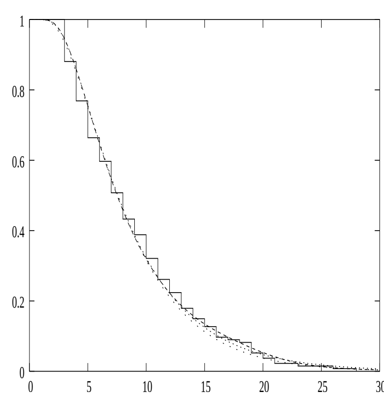

Fig. 8. ESF and FSF of the three mixture models for the depressive condition data.

6 Concluding comments

In this paper, we have considered three different mixture models based on the Birnbaum-Saunders and length-biased Birnbaum-Saunders distributions, and discussed some of their structural aspects. Since the ML estimation by direct maximization of the log-likelihood function is numerically quite involved, we have proposed the EM algorithm for this purpose in order to simplify this estimation procedure. We have analyzed two data sets for the purpose of illustration and it has been demonstrated that the proposed mixture models as well as the suggested EM algorithm work very well. The shapes of the densities and hazard rate functions are important indicators for data analysis as well as for model selection. For the three mixture models, we have used graphical tools to examine their shapes. An analytical examination of the form and features of these functions, for the proposed mixture models, is of interest and remains open. For Model 3, however, it can be shown that the densities and hazard rate functions are both unimodal.

Acknowledgments

Research work of V. Leiva and A. Sanhueza were partially supported by FONDECYT 1080326 and 1090265 grants from Chile.

References

Balakrishnan, N., Leiva, V., Sanhueza, A., Cabrera, E. 2009. Mixture inverse Gaussian distribution and its transformations, moments and applications. Statistics 43, 91-104.

Balakrishnan, N., Leiva, V., Sanhueza, A., Vilca, F. 2009. Estimation in the Birnbaum-Saunders distribution based on scale-mixture of normals and the EM-algorithm. Stat. Oper. Res. Trans. 33, 171-192.

Bebbington, M., Lai, C-D, Zitikis, R., 2008. A proof of the shape of the Birnbaum-Saunders hazard rate function. Mathem. Scientist 33, 49-56.

Bhattacharyya, G.K, Fries, A., 1982. Fatigue failure models: Birnbaum–Saunders vs. inverse Gaussian. IEEE Trans. Rel. 31, 439-440.

Birnbaum, Z.W., Saunders, S.C., 1969. A new family of life distributions. J. Appl. Prob.6, 319-327.

Bechtel, Y.C., Bonaiti-Pellie, C., Poisson, N., Magnette, J., Bechtel, P.R., 1993. A population and family study of n-acetyltransferase using caffeine urinary metabolites. Clin. Pharmacol. Therapeut. 54, 134-141.

Chhikara, R.S., Folks, J.L., 1989. The Inverse Gaussian Distribution. Marcel Dekker, New York.

Cox, D.R., 1969. Some sampling problems in technology. In New Developments in Survey Sampling (Johnson, N.L., Smith, H. Jr., Eds.), 506-527, Wiley, New York.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc.. Ser. B 39, 1-38.

Desmond, A.F., 1985. Stochastic models of failure in random environment. Can. J. Stat. 13, 171-183.

Desmond, A.F., 1986. On the relationship between two fatigue-life models. IEEE Trans. Rel. 35, 167-169.

Efron, B., Hinkley, D., 1978, Assessing the accuracy of the maximum likelihood estimator: observed vs. expected Fisher information. Biometrika 65, 457-487.

Engelhardt, M., Bain, L.J., Wright, F.T., 1981. Inferences on the parameters of the Birnbaum-Saunders fatigue life distribution based on maximum likelihood estimation. Technometrics 23, 251-256.

Fisher, R.A., 1934. The effects of methods of ascertainment upon the estimation of frequencies. Ann. Eugenics 6, 1325.

Finch, S.J., Mendell, N.R., Thode, Jr. H.C., 1989. Probabilistic measures of adequacy of a numerical search for a global maximum. J. Amer. Stat. Soc.84, 1020-1023.

Gupta, R.C., Akman, O., 1995. On the reliability studies of a weighted inverse Gaussian distribution. J. Stat. Plan. Infer. 48, 69-83.

Gupta, R.C., Akman, O., 1997. Estimation of critical points in the mixture of inverse Gaussian model. Stat. Papers 38, 445-452.

Gupta, R.C., Keating, J.P., 1985. Relations for reliability measures under length biased sampling. Scand. J. Stat. 13, 49-56.

Gupta, R.C., Kirmani, S.N.U.A., 1990. The role of weighted distributions in stochastic modeling. Comm. Stat. Theor. Meth. 19, 3147-3162.

Jain, K., Singh, H., Bagai, I., 1989. Relations for reliability measures of weighted distributions. Comm. Stat. Theor. Meth. 18, 4393-4412.

Johnson, N.L., Kotz, S., Balakrishnan, N., 1994. Continuous Univariate Distributions — Vol. 1. 2nd edition. Wiley, New York.

Johnson, N.L., Kotz, S., Balakrishnan, N., 1995. Continuous Univariate Distributions — Vol. 2. 2nd edition. Wiley, New York.

Jorgensen, B., 1987. Some properties of the exponential dispersion model. J. R. Stat. Soc.. Ser. B 49, 127-162.

Jorgensen, B., Seshadri, V., Whitmore, G.A., 1991. On the mixture of the inverse Gaussian distribution with its complementary reciprocal. Scand. J. Stat. 18, 77-89.

Khattree, R., 1989. Characterization of inverse Gaussian and gamma distributions through their length biased versions. IEEE Trans. Rel. 38, 610-611.

Kundu, D., Kannan, N., Balakrishnan, N., 2008. On the hazard function of Birnbaum-Saunders distribution and associated inference. Comp. Stat. Data Anal. 52, 2692-2702.

Leiva, V., Sanhueza, A., Angulo, J.M., 2009. A length-biased version of the

Birnbaum-Saunders distribution with application in water quality. Stoch. Environ. Res. Risk Anal. 23, 299-307.

Louis, T.A., 1982. Finding the observed information matrix when using the EM algorithm. J. R. Stat. Soc.. Ser. B 44, 226-233.

Marshall, A.W., Olkin, I., 2007. Life Distributions. Springer, New York.

McLachlan, G.J., Peel, D. 2000. Finite Mixture Models. Wiley, New York.

Ng, H.K.T., Kundu, D., Balakrishnan, N., 2003. Modified moment estimation for the two-parameter Birnbaum-Saunders distribution. Comp. Stat. Data Anal. 43, 283-298.

Ng, H.K.T., Kundu, D., Balakrishnan, N., 2006. Point and interval estimations for the two-parameter Birnbaum-Saunders distribution based on type-II censored samples. Comp. Stat. Data Anal. 50, 3222-3242.

Olyede, B.O., George, E.O., 2002. On stochastic inequalities and comparisons of reliability measures for weighted distributions. Math. Prob. Eng. 8, 1-13.

Patil, G.P., 2002. Weighted distributions. In Encyclopedia of Environmetrics, vol. 4, pp. 2369 - 2377. Wiley, New York..

Patil, G.P., Rao, C.R., 1978. Weighted distributions and size-biased sampling with applications to wildlife populations and human families. Biometrics 34, 179-189.

Rao, C.R., 1965. On discrete distributions arising out of methods of ascertainment. In Classical and Contagious Discrete Distributions (Patil, G.P. ed.). Pergamon Press and Statistical Publishing Society, Calcutta, pp. 320-332. (Also in Sankhya A, 27, 311-324)

Sanhueza, A., Leiva, V., Balakrishnan, N., 2008. The generalized Birnbaum-Saunders distribution and its theory, methodology and application. Comm. Stat. Theor. Meth. 37, 645-670.

Sansgiry, P.S., Akman, O., 2001. Reliability estimation via length-biased transformation. Comm. Stat. Theor. Meth. 30, 2473-2479.

Seshadri, V., 1993. The Inverse Gaussian Distribution: A Case Study in Exponential Families. Oxford University Press, Oxford, England.