Journal of Mathematical Extension
Vol. 15, No. 2, (2021) (2)1-34
URL: https://doi.org/10.30495/JME.2021.1293
ISSN: 1735-8299
Original Research Paper

Mixture of Extended Birnbaum-Saunders Distributions: An Approach via the Mean-Mixture of Normal Models

S. Mahbudi
Science and Research Tehran Branch, Islamic Azad University

A. Jamalizadeh*
Shahid Bahonar University of Kerman

R. Farnoosh
Iran University of Science and Technology

Abstract. The Birnbaum-Saunders (BS) distribution is one of the most considered right-skewed distributions to model failure times for materials subject to lifetime data. In this paper, a new extension of the BS model is initially proposed based on the family of mean-mixtures of normal distributions. Then, we present a new probabilistic mixture model based on the new extended BS distribution for modeling and clustering right-skewed and heavy-tailed data. The maximum likelihood (ML) parameter estimates of the model in question are estimated by employing an expectation-maximization (EM) type algorithm. Moreover, the empirical information matrix is derived by using an information based approach. Simulations and real data analysis illustrate the performance of the proposed methodology.

AMS Subject Classification: 62F10; 62J02.
Keywords and Phrases: Birnbaum-Saunders distribution; Mean-mixtures of normal distributions; Finite mixture model; ECM algorithm.

Received: June 2019; Accepted: January 2020
*Corresponding Author

1 Introduction

Finite mixture (FM) model is one the most considered statistical tools for cluster analysis in dealing with the various datasets in the biological and social sciences. Some recently applications of the FM model can be found biometrics [30], genetics and medicine [37], marketing [45], pattern recognition problems [39], and reliability studies [11], among the others. The probability distribution function (PDF) of a random variable X distributed by the FM model is

$$f(x;\boldsymbol{\Theta}) = \sum_{i=1}^{g} \pi_i f_i(x;\boldsymbol{\theta}_i),$$

where $\boldsymbol{\pi} = (\pi_1, ..., \pi_g)^\top$ is a vector of mixing proportions ($\pi_i \geq 0$ and $\sum_{i=1}^{g} \pi_i = 1$), $f_i(x;\boldsymbol{\theta}_i)$ is the mixing component for $i = 1, ..., g$, and $\boldsymbol{\Theta} = (\pi_1, ..., \pi_{g-1}, \boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_g)$ denotes the parameters set. Details of the FM models can be found in [41, 26, 27, 13]. Recently, the interest of using skew distributions in the FM model has been grown due to their flexibility. For instance, Jamalizadeh [19] proposed the finite mixture of univariate scale-shape mixture of normal distributions (FM-SSMN) and studied some of its characteristics and properties. Wang [44] extended the FM-SSMN distributions in to multivariate version and showed that the new model can provide interesting contour plots. Based on the other class of skew distribution, Naderi et al. [31, 32] introduced the finite mixture of univariate and multivariate normal mean-mixture of Birnbaum-Saunders distribution, respectively.

Although all aforementioned distributions provide straightforward platform for data analysis, they are defined in the real line, $\mathbb{R}$, and using the $\mathbb{R}$ distributions for positive valued (life time) data may leads to boundary bias problem [36, 32]. To cope with these datasets, Ali [3] introduced the FM model based on the inverse Rayleigh distribution. Ali [3] used this model for engineering processes and provided some properties of the proposed model. One can also find the mixture of gamma, exponential, inverse Gaussian and Weibull distributions in Wiper et al. [46, 21, 7, 20].

The Birnbaum-Saunders (BS) distribution [10] is one of most flexible life models. Applications of the BS distribution have been recently used for data analysis can be found in diverse fields such as econometrics Aslam and Kantam [4], engineering Jamalizadeh et al. [18] and environmental analysis Mohammadi et al. [29]. Theoretically, the random variable T generated from the linear transformation

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

is said to follow the BS distribution, where α and β are the shape and scale parameters, respectively, and X has a standard normal distribution, $\mathrm{N}(0, 1)$. Although the main motivation of the BS distribution originally came from the modeling material fatigue Birnbaum and Saunders [10], various extensions of the BS distribution are proposed through the linear representation (1) to accommodate strongly skewed and heavily tailed data. For instance, by replacing the standard normal variable X in (1) with other random variables, Vilca-Labra et al . [43], Khosravi et al. [22] and Hashemi et al. [15], proposed the skew-normal-BS (SN-BS), skew-t-BS (ST-BS) and skew-normal-t distributions, respectively. More recently, [16] also introduced Normal mean-variance Lindley Birnbaum-Saunders distribution as an alternative model for analysing positive financial datasets. Although, these generalized models may have not physical meaning, as the BS distribution, they can be used to fit right-skewed and non-negative datasets.

Recently, Negarestani et al. [34] exploited the definition of restricted skew normal distribution to introduce a class of skewed model which can provide wider range of skewness and kurtosis than the skew-normal and skew-t distributions. Calling the class of mean-mixture of normal (MMN) distribution, Negarestani et al. [34] also studied the properties of new model and illustrated its utility in regression and time series analyses. Owning the interesting properties of MMN model, the main objectives of this paper are as follows. 1) We present a new extension of the BS distribution by considering the MMN distribution as a core model in the representation (1). 2) Some interesting properties of the new model, referred to as the MMN-BS henceforth, are studied. 3) Finally, we also propose a FM model based on the new extended BS distribution for analyzing multi-modal datasets.

The outline of the paper is as follows. In Section 2, we establish the notations and outline some preliminary results. In Section 3 we discuss the main results of the paper and some specification of the MMN-BS model. The finite mixture of MMN-BS distributions along with its parameter estimation via an EM-type algorithm are presented in Section 4. The utility of the proposed model is illustrated in Sections 5 and 6 by considering two real datasets and conducting two simulation studies. Some concluding remarks are finally given in Section 7.

2 Mean-mixtures of normal distribution

Let Z ba a normally distributed random variable with mean zero and variance 1, $\mathrm{N}(0,1)$. Following Negarestani et al. [34], a random variable Y is in the mean-mixture of normal family, $Y \sim \mathrm{MMN}(\mu, \sigma^2, \lambda, \boldsymbol{\nu})$, if it can be written as

$$Y = \mu + \sigma \left(\delta U + (1-\delta^2)^{1/2} X \right),$$

where $\delta = \lambda/\sqrt{1+\lambda^2}$ and U is an arbitrary random variable, independent of X, with cumulative distribution function (CDF) $H(\cdot;\boldsymbol{\nu})$ or probability distribution function (PDF) $h(\cdot;\boldsymbol{\nu})$ which is indexed by a scalar or vector parameter $\boldsymbol{\nu} \in \mathbb{R}^k$. It can be seen that Y has the following hierarchical representation:

$$\begin{aligned} Y|(U=u) &\sim \mathrm{N}(\mu + \sigma\delta u, \sigma^2(1-\delta^2)), \\ U &\sim h(0,1;\boldsymbol{\nu}). \end{aligned} \tag{2}$$

Then, the pdf of $Y \sim \mathrm{MMN}(\mu, \sigma^2, \lambda, \boldsymbol{\nu})$ is given by

$$\begin{aligned} f_{\mathrm{MMN}}(y;\mu,\sigma^2,\lambda,\boldsymbol{\nu}) &= \int_{-\infty}^{+\infty} \phi(y;\mu+\sigma\delta u, \sigma^2(1-\delta^2))\ dH(u;\boldsymbol{\nu}) \\ &= \int_{-\infty}^{+\infty} \phi(y;\mu+\sigma\delta u, \sigma^2(1-\delta^2)) h(u;\boldsymbol{\nu})\ du, \quad y \in \mathbb{R}, \end{aligned} \tag{3}$$

where $\phi(\cdot;\mu,\sigma^2)$ is the PDF of normal distribution with mean μ and variance σ^2. In the following, three spacial cases of the MMN model are introduced.

2.1 Convolution with truncated normal distribution

If U in the hierarchical representation (2) followed by the standard truncated normal distribution lying within the truncated interval $(0,+\infty)$, denoted by $U \sim \mathrm{TN}(0,1;(0,+\infty))$, then the random variable Y has a skew-normal distribution [5], whose PDF can be given as

$$f_{\mathrm{SN}}(y;\mu,\sigma^2,\lambda) = 2\phi(y,\mu,\sigma^2)\Phi\left(\lambda \frac{y-\mu}{\sigma}\right). \tag{4}$$

where $\Phi(\cdot)$ is the CDF of N(0,1). We will use the notation $Y \sim \text{SN}(\mu, \sigma^2, \lambda)$ if Y has PDF (4).

Lemma 2.1. *Suppose $Y \sim SN(\mu, \sigma^2, \lambda)$ and $U \sim TN(0, 1; (0, \infty))$. Then, $U|Y = y \sim TN\big(\mu, (1+\lambda^2)^{-1}; (0,\infty)\big)$, where $\mu = w\lambda/\sqrt{1+\lambda^2}$. Furthermore, for $k = 2, \ldots,$*

$$
\begin{aligned}
E(U^k|Y=y) =& \mu E(U^{k-1}|Y=y) + \frac{k-1}{1+\lambda^2} E(U^{k-2}|Y=y), \\
E(U|Y=y) =& \mu + \frac{\phi(\lambda w)}{\sqrt{1+\lambda^2}\Phi(\lambda w)}.
\end{aligned}
$$

where $w = (y-\mu)/\sigma$.

Proof. Details of proof can be found in [5]. □

2.2 Convolution with exponential distribution

The convected mean-mixture normal of exponential (MMNE) distribution can be obtained by the hierarchical representation (2), if the random variable U has a standard exponential distribution, then, the PDF of Y can be obtained from (3) as

$$
\begin{aligned}
f_{\text{MMNE}}(y; \mu, \sigma^2, \lambda) =& \frac{\sqrt{1+\lambda^2}}{\sigma|\lambda|} \exp\left\{ -\frac{\sqrt{1+\lambda^2}}{\lambda} w + \frac{1}{2\lambda^2} \right\} \\
& \Phi\left(\frac{\lambda\sqrt{1+\lambda^2} w - 1}{|\lambda|} \right), \quad y \in R; \lambda \neq 0
\end{aligned}
$$

where $w = (y-\mu)/\sigma$. In this case, we denote $Y \sim \text{MMNE}(\mu, \sigma^2, \lambda)$.

Lemma 2.2. *If $Y \sim MMNE(\mu, \sigma^2, \lambda)$ and $U \sim E(1)$, Then, $U|Y = y \sim TN\big(\mu, \lambda^{-2}; (0,\infty)\big)$, where $\mu = w\frac{\sqrt{1+\lambda^2}}{\lambda} - \lambda^{-2}$. Furthermore, for $k = 1, 2, \ldots,$*

$$
\begin{aligned}
E(U^k|Y=y) =& \mu E(U^{k-1}|Y=y) + \frac{k-1}{\lambda^2} E(U^{k-2}|Y=y), \\
E(U|Y=y) =& \mu + \frac{\phi(|\lambda|\mu)}{|\lambda|\Phi(|\lambda|\mu)}.
\end{aligned}
$$

Proof. The proof can be found in Negarestani et al. [34]. □

2.3 Convolution with mixture of exponential and half-normal distributions

Here, we assume that the random variable U in (2) follows a mixture of the exponential with mean 2 and the standard half-normal distributions with PDF

$$f_U(u;\nu) = \nu \frac{1}{2} \exp\left\{-\frac{u}{2}\right\} + 2(1-\nu)\phi(u), u > 0, \quad 0 < \nu < 1.$$

The density of Y is then given by

$$f_{\text{MMNEH}}(y;\mu,\sigma^2,\lambda,\nu) = \frac{\nu\sqrt{1+\lambda^2}}{2\sigma|\lambda|} \exp\Big\{ -\frac{\sqrt{1+\lambda^2}}{2\lambda} w + \frac{1}{8\lambda^2}\Big\}$$
$$\Phi\left(\frac{\lambda\sqrt{1+\lambda^2}w - 1}{|\lambda|}\right) + (1-\nu)\frac{2}{\sigma}\phi(w)\Phi(\lambda w), \quad y \in \mathbb{R} \tag{5}$$

where $\mu \in \mathbb{R}$, $\sigma^2 > 0$, and $0 < \nu < 1$. In this case, we denote $Y \sim \text{MMNEH}(\mu,\sigma^2,\lambda,\nu)$.

Lemma 2.3. *Let $Y \sim MMNEH(\mu,\sigma^2,\lambda,\nu)$ and U has PDF (5). Then, the conditional PDF of U, given $Y = y$ is*

$$f_{U|Y=y}(u) = \pi(\boldsymbol{y})\frac{\phi\left(u;\mu_1,\lambda^{-2}\right)}{\Phi(|\lambda|\mu_1)} + (1-\pi(\boldsymbol{y}))\frac{\phi\left(u;\mu_2,(1+\lambda^2)^{-1}\right)}{\Phi(\lambda z)}$$

where $\mu_1 = (\lambda\sqrt{1+\lambda^2}w - 1/2)/\lambda^2$, and $\mu_2 = w\lambda/\sqrt{1+\lambda^2}$,

$$\pi(y) = \frac{\sqrt{1+\lambda^2}\nu}{2\sigma|\lambda| f_{MMNEH}(y;\mu,\sigma^2,\lambda,\nu)} \exp\left\{-\frac{\sqrt{1+\lambda^2}}{2\lambda}w + \frac{1}{8\lambda^2}\right\}\Phi(|\lambda|\mu_1).$$

Furthermore, for any $y \in \mathbb{R}$, and $k = 1, 2, \ldots,$

$$E\left(U^k|Y=y\right) = \pi(y)E\left(V_1^k\right) + (1-\pi(y))E(V_2^k),$$

where $V_1 \sim TN\big(\mu_1, \lambda^{-2}; (0,\infty)\big)$, $V_2 \sim TN\big(\mu_2, (1+\lambda^2)^{-1}; (0,\infty)\big)$ *and*

$$\begin{aligned} E(V_1) =& \mu_1 + \frac{\phi\left(|\lambda|\mu_1\right)}{|\lambda|\Phi\left(|\lambda|\mu_1\right)}, \\ E(V_1^k) &= \mu_1 E(V_1^{k-1}|Y=y) + \frac{k-1}{\lambda^2} E(V_1^{k-2}|Y=y), \quad k \geq 2, \\ E(V_2) =& \mu_2 + \frac{\phi\left(\lambda z\right)}{(1+\lambda^2)\Phi\left(\lambda z\right)}, \\ E(V_2^k) &= \mu_2 E(V_2^{k-1}|Y=y) + \frac{k-1}{1+\lambda^2} E(V_2^{k-2}|Y=y), \quad k \geq 2. \end{aligned}$$

Proof. The proof can be found in Negarestani et al. [34]. □

3 The mean-mixtures of normal-Birnbaum-Saunders distribution

Definition 3.1. A positive random variable T is said to have a MMN-BS distribution if T has a linear relation with the MMN model as

$$T = \frac{\beta}{4}\left[\alpha Y + \sqrt{(\alpha Y)^2 + 4}\right]^2 \tag{6}$$

where $Y \sim \text{MMN}(0, 1, \lambda, \boldsymbol{\nu})$. The PDF and the corresponding CDF of T can be presented by

$$\begin{aligned} f_{\text{MMN-BS}}(t; \alpha, \beta, \lambda, \boldsymbol{\nu}) &= f_{\text{MMN}}(a(t,\alpha,\beta); 0, 1, \lambda, \boldsymbol{\nu}) A(t,\alpha,\beta), \\ F_{\text{MMN-BS}}(t; \alpha, \beta, \lambda, \boldsymbol{\nu}) &= F_{\text{MMN}}(a(t,\alpha,\beta); 0, 1, \lambda, \boldsymbol{\nu})), \qquad t > 0, \end{aligned} \tag{7}$$

where $F_{\text{MMN}}(\cdot)$ is the CDF of the standard ($\mu = 0, \sigma^2 = 1$) MMN distribution and

$$a(t,\alpha,\beta) = \frac{1}{\alpha}\left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right) \quad \text{and} \quad A(t,\alpha,\beta) = \frac{t+\beta}{2\alpha\sqrt{t^3\beta}}.$$

The notation $T \sim \text{MMN-BS}(\alpha, \beta, \lambda, \boldsymbol{\nu})$ is used henceforth if T has PDF (7). Fig. 1 shows a graphical illustration of the PDF (7) for two special cases of MMN model and for $\beta = 1$ and different setting of parameters. It can be observe that the MMN-BS distribution is an asymmetric and positively

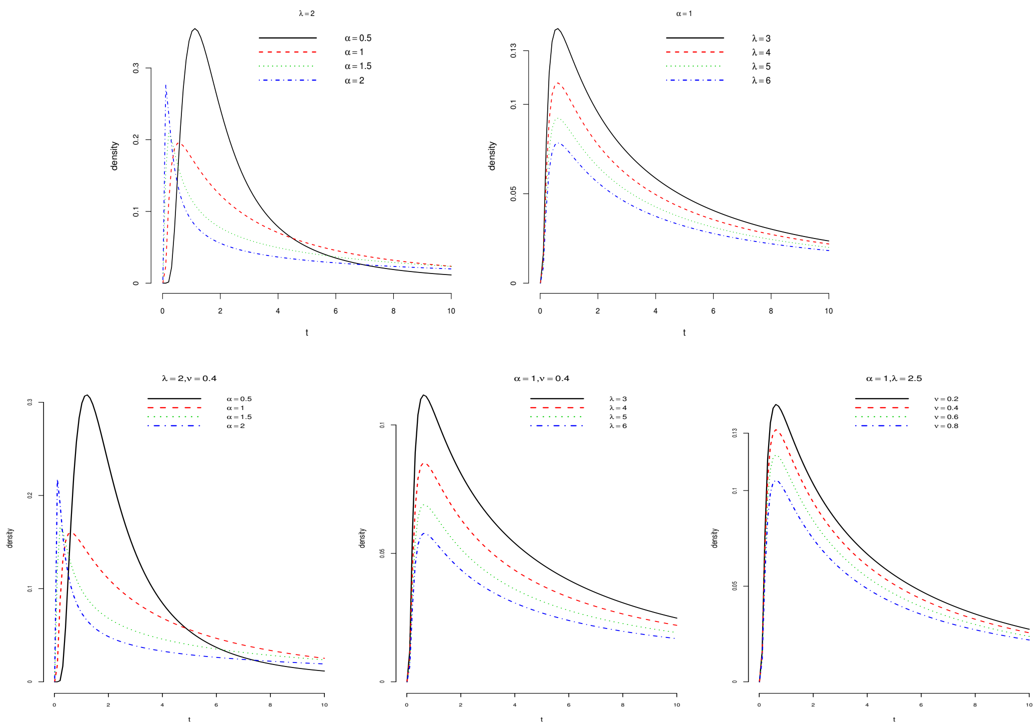

Figure 1: The density plots of the MMNE-BS (up) and MMNEH-BS (down) distribution for various values of parameters with $\beta = 1$.

skewed distribution and can provide diverse degrees of skewness and kurtosis which enable us to utilize it in order to model positive data. It is also clear that the parameters λ and ν have substantial effects on its skewness and kurtosis of the SN-BS (see [42]; for detail SN-BS), mean-mixture normal of exponential-BS (MMNE-BS) and mean-mixture normal of exponential-half-normal BS (MMNEH-BS) distributions.

To investigate the effects of shape parameters on the skewness and kurtosis, the skewness and kurtosis of T can be obtained respectively as

$$\gamma_t = \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3}{(\mu_2 - \mu_1^2)^{1.5}} \quad \text{and} \quad \kappa_t = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4}{(\mu_2 - \mu_1^2)^2},$$

where $\mu_r = E(T^r)$ for $r = 1, 2, 3, 4$. The closed form of μ_r are provided in Appendix **A**. Table 1 and 2 give the numerical value of γ_t and κ_t for the MMNE-BS and MMNEH-BS distributions with different sets of parameter values. It can be observed from these Tables that the MMN-BS family of distributions can takes wider ranges of skewness and kurtosis as compared with the BS, SN-BS and ST-BS distributions.

Table 1: Value of skewness and kurtosis based on moments of the MMNEBS(α, β, λ) distribution when $\beta = 1$.

	$\|\lambda\| = 0.10$				$\|\lambda\| = 0.25$			
	γ_t		κ_t		γ_t		κ_t	
α	$-\lambda$	λ	$-\lambda$	λ	$-\lambda$	λ	$-\lambda$	λ
0.40	1.1991	1.1739	2.3673	2.2545	1.2242	1.2309	2.4770	2.5575
0.50	1.4840	1.4395	3.5983	3.3573	1.5279	1.4946	3.8303	3.7146
0.75	2.1207	2.0102	7.1815	6.3808	2.2209	2.0413	7.9199	6.7108
1.00	2.6216	2.4367	10.7375	9.1701	2.7814	2.4301	12.1558	9.2837
1.25	2.9933	2.7413	13.7563	11.4118	3.2057	2.6969	15.8653	11.2403
1.50	3.2638	2.9570	16.1422	13.1159	3.5188	2.8804	18.8588	12.6644
2.00	3.6044	3.2197	19.3161	15.2516	3.9184	3.0796	22.9355	14.0077
	$\|\lambda\| = 0.50$				$\|\lambda\| = 0.75$			
0.40	1.2357	1.6756	2.5515	5.8180	1.2242	2.5360	2.5349	14.6000
0.50	1.5755	1.9941	4.0992	8.0149	1.6018	2.9379	4.2600	18.9943
0.75	2.3671	2.6047	9.0438	13.0219	2.4903	3.6134	10.0387	27.0506
1.00	3.0268	2.9922	14.4702	16.6274	3.2450	3.9014	16.6750	28.9705
1.25	3.5377	3.2198	19.4101	18.5227	3.8377	3.8318	22.8930	24.9290
1.50	3.9200	3.3034	23.4986	18.3337	4.2849	3.5326	28.1303	18.9983
2.00	4.4139	3.0925	29.2055	13.7152	4.8665	2.8173	35.5539	10.1241
	$\|\lambda\| = 1$				$\|\lambda\| = 2$			
0.40	1.2114	3.4686	2.4907	26.7954	1.2449	5.3426	2.4583	53.4120
0.50	1.6236	3.8938	4.3810	32.7632	1.7498	5.1274	4.9410	44.0425
0.75	2.6010	4.4305	10.9645	38.2974	2.9948	3.9147	14.4840	20.9382
1.00	3.4437	4.2781	18.8071	30.8258	4.1258	2.9923	27.0290	10.7827
1.25	4.1125	3.7694	26.3225	21.0925	5.0500	2.4275	39.7635	6.3796
1.50	4.6201	3.2364	32.7356	14.0882	5.7615	2.0726	50.9809	4.1822
2.00	5.2827	2.4468	41.9230	6.8264	6.6968	1.6871	67.4229	2.2166

Proposition 3.2. *The stochastic representation of the MMN-BS distribution is*

$$T = \frac{\beta}{4}\left[\alpha(X + \lambda U) + \sqrt{(\alpha(X + \lambda U))^2 + 4}\right]^2,$$

Table 2: Value of skewness and kurtosis based on moments of the MMNEHBS$(\alpha, \beta, \lambda, \nu)$ distribution when $\beta = 1$.

		$\nu = 0.2$				$\nu = 0.5$			
		γ_t		κ_t		γ_t		κ_t	
$\|\lambda\|$	α	$-\lambda$	λ	$-\lambda$	λ	$-\lambda$	λ	$-\lambda$	λ
0.10	0.40	1.1983	1.1829	2.3645	2.3047	1.2048	1.1940	2.3922	2.3593
	0.50	1.4838	1.4499	3.5971	3.4262	1.4948	1.4601	3.6540	3.4872
	0.75	2.1218	2.0230	7.1885	6.4956	2.1461	2.0263	7.3636	6.5318
	1.00	2.6239	2.4506	10.7561	9.3197	2.6621	2.4437	11.0865	9.2848
	1.25	2.9967	2.7555	13.7866	11.5840	3.0471	2.7384	14.2725	11.4625
	1.50	3.2680	2.9711	16.1826	13.3019	3.3282	2.9454	16.8041	13.0986
	2.00	3.6097	3.2319	19.3720	15.4170	3.6834	3.1916	20.1931	15.0381
0.25	0.40	1.2108	1.4398	2.4423	4.4559	1.2108	1.6176	2.4531	5.6512
	0.50	1.5174	1.7567	3.7926	6.4960	1.5308	1.9527	3.8685	8.0144
	0.75	2.2154	2.4204	7.8897	11.8634	2.2639	2.6282	8.2488	13.8267
	1.00	2.7793	2.8911	12.1458	16.4337	2.8616	3.0831	12.8895	18.3884
	1.25	3.2064	3.2027	15.8784	19.4875	3.3174	3.3673	17.0119	21.1087
	1.50	3.5216	3.3778	18.8941	20.5395	3.6552	3.5003	20.3697	21.4939
	2.00	3.9240	3.3825	23.0070	17.8196	4.0880	3.3697	24.9853	17.0840
0.50	0.40	1.2008	3.6618	2.4960	40.2918	1.1675	3.8523	2.3913	35.8592
	0.50	1.5512	4.4101	4.0404	55.2556	1.5465	4.4392	4.0360	45.7668
	0.75	2.3579	5.6390	9.0244	78.2803	2.4206	5.2712	9.5176	57.5466
	1.00	3.0275	5.8260	14.5288	70.7490	3.1512	5.2266	15.7400	48.7422
	1.25	3.5464	5.3214	19.5613	51.6264	3.7202	4.6844	21.5195	34.6440
	1.50	3.9351	4.6295	23.7398	35.4424	4.1477	4.0603	26.3608	23.8162
	2.00	4.4383	3.4747	29.5951	17.1771	4.7019	3.0776	33.1924	12.0015
1.00	0.40	1.1911	9.2654	2.5535	177.5548	1.1331	6.9628	2.2944	92.0197
	0.50	1.6227	9.2593	4.4887	157.4775	1.6038	6.7867	4.3741	78.3611
	0.75	2.6307	7.4555	11.3219	84.4980	2.7103	5.3399	12.0072	40.0730
	1.00	3.5003	5.7516	19.5641	46.7636	3.6720	4.1654	21.5092	22.1659
	1.25	4.1948	4.5486	27.5481	28.4080	4.4426	3.4094	30.8606	14.0110
	1.50	4.7249	3.6769	34.4185	18.1179	5.0311	2.8939	38.9664	9.6378
	1.75	5.4210	2.5934	44.3478	8.1153	5.8025	2.2307	50.7295	5.1432

where $X \sim N(0,1)$ and U have PDF $h(u;\boldsymbol{\nu})$, independently.

Proof. The proposition can be readily obtained throughout (1) and (6). □

Theorem 3.3. *Some properties of the MMN-BS distribution are as follows:*

1. *The MMN-BS distribution contains the ordinary BS distribution as $\lambda \rightarrow$*

0.

2. *The random variable T distributed by MMN-BS$(\alpha, \beta, \lambda, \boldsymbol{\nu})$ is degenerated to β as α tends to zero.*

3. *If $T \sim$ MMN-BS$(\alpha, \beta, \lambda, \boldsymbol{\nu})$, then*

$$Y = \frac{1}{\alpha}\left[\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}}\right] \sim MMN(0, 1, \lambda, \boldsymbol{\nu}).$$

4. *Let $T \sim$ MMN-BS$(\alpha, \beta, \lambda, \boldsymbol{\nu})$. It can be easily shown that the hazard rate function of T is*

$$H(t) = \frac{f_{\textit{MMN-BS}}(t; \alpha, \beta, \lambda, \boldsymbol{\nu})}{1 - F_{MMN}(a(t, \alpha, \beta); 0, 1, \lambda, \boldsymbol{\nu})}.$$

Theorem 3.4. *Let EBS Stands for the extended Birnbaum-Saunders distribution [24]. Then, the hierarchical representation of $T \sim$ MMN-BS$(\alpha, \beta, \lambda, \boldsymbol{\nu})$ is given as*

$$\begin{aligned} T|U = u &\sim EBS(\alpha\sqrt{1-\delta^2}, \beta, -\frac{\delta u}{\sqrt{1-\delta^2}}), \\ U &\sim h(u; \boldsymbol{\nu}). \end{aligned}$$

Proof. The proof is completed by Bayes' rule and some mathematical work. □

Proposition 3.5. *Let $U \sim TN(0, 1; (0, \infty))$ and $T \sim$ SN-BS(α, β, λ) with PDF*

$$f_{\textit{SN-BS}}(t; \alpha, \beta, \lambda) = f_{SN}(a(t, \alpha, \beta); 0, 1, \lambda)A(t; \alpha, \beta), \quad t > 0.$$

Then, $U|T = t \sim TN\left(\mu', (1+\lambda^2)^{-1}; (0, \infty)\right)$, where $\mu' = a(t, \alpha, \beta)\lambda \big/ \sqrt{1+\lambda^2}$. Moreover, for $k = 1, 2, \ldots,$

$$E(U^k|T = t) = \mu' E(U^{k-1}|T = t) + \frac{k-1}{1+\lambda^2}E(U^{k-2}|T = t),$$

where

$$E(U|T = t) = \mu' + \frac{\phi\left(\lambda a(t, \alpha, \beta)\right)}{\sqrt{1+\lambda^2}\Phi\left(\lambda a(t, \alpha, \beta)\right)}.$$

Proof. Details of proof can be found in [42]. □

Theorem 3.6. *Let $U \sim E(1)$ and $T \sim$ MMNE-BS(α, β, λ) with PDF*

$$f_{MMNE\text{-}BS}(t; \alpha, \beta, \lambda) = f_{MMNE}(a(t, \alpha, \beta); 0, 1, \lambda)A(t, \alpha, \beta), \quad t > 0. \tag{8}$$

Then, $U|T = t \sim TN\big(\mu', \lambda^{-2}; (0, \infty)\big)$, where $\mu' = a(t, \alpha, \beta)\frac{\sqrt{1+\lambda^2}}{\lambda} - \lambda^{-2}$. Moreover, for $k = 2, \dots,$

$$E(U^k|T = t) = \mu' E(U^{k-1}|T = t) + \frac{k-1}{\lambda^2} E(U^{k-2}|T = t),$$

where

$$E(U|T = t) = \mu' + \frac{\phi\left(|\lambda|\mu'\right)}{|\lambda|\Phi\left(|\lambda|\mu'\right)}.$$

Proof. From Lemma 2.2, we have $T|U = u \sim \mathrm{EBS}(\alpha\sqrt{1-\delta^2}, \beta, -\frac{\delta u}{\sqrt{1-\delta^2}})$. Using (8) and some algebraic factorization, the conditional pdf can be obtained by applying Baye's rule as

$$\begin{aligned} f(u|t) =& \frac{f(t,u)}{f(t)} = \frac{f(t|u)f(u)}{f(t)} \\ =& \frac{\dfrac{A(t,\alpha,\beta)}{\sqrt{2\pi(1-\delta^2)}} \exp\left\{-\dfrac{1}{2(1-\delta^2)}\left(a(t,\alpha,\beta) - \frac{\lambda}{\sqrt{1+\lambda^2}}u\right)^2\right\} \mathrm{e}^{-u}}{A(t,\alpha,\beta)\frac{\sqrt{1+\lambda^2}}{|\lambda|}\exp\left\{-\frac{\sqrt{1+\lambda^2}}{\lambda}a(t,\alpha,\beta) + \frac{1}{2\lambda^2}\right\}\Phi\left(\frac{\lambda\sqrt{1+\lambda^2}a(t,\alpha,\beta)-1}{|\lambda|}\right)} \end{aligned}$$

After some algebraic manipulations, the resulting conditional distribution of U given $T = t$ is given by

$$f(u|t) = \frac{|\lambda| \exp\left\{-\dfrac{\lambda^2}{2}\left(u - \mu'\right)^2\right\}}{\sqrt{2\pi}\Phi\left(|\lambda|\mu'\right)}$$

Thus, the conditional distribution of U given $T = t$ is $\mathrm{TN}\left(\mu', \lambda^{-2}; (0, \infty)\right)$. Based on some particular moments of the truncared normal distribution have tractable forms, we have

$$E(U^k|T = t) = \mu' E(U^{k-1}|T = t) + \frac{k-1}{\lambda^2} E(U^{k-2}|T = t), \quad k = 2, 3, \dots,$$

where

$$E(U|T=t) = \mu' + \frac{\phi\left(|\lambda|\mu'\right)}{|\lambda|\Phi\left(|\lambda|\mu'\right)}.$$

□

Theorem 3.7. *Let U has PDF (5) and $T \sim$ MMNEH-BS$(\alpha, \beta, \lambda, \nu)$ with PDF*

$$f_{\textit{MMNEH-BS}}(t;\alpha,\beta,\lambda,\nu) = f_{\textit{MMNEH}}(a(t,\alpha,\beta);0,1,\lambda,\nu)A(t,\alpha,\beta) \quad t>0.$$

Then, the PDF of conditional distribution $U|T=t$ is

$$f_{U|T=t}(u) = \pi(t)\frac{\phi\left(u;\mu_1',\lambda^{-2}\right)}{\Phi(|\lambda|\mu_1')} + (1-\pi(t))\frac{\phi\left(u;\mu_2',\frac{1}{1+\lambda^2}\right)}{\Phi(\lambda a(t,\alpha,\beta)},$$

where $\mu_1' = (\lambda\sqrt{1+\lambda^2}a(t,\alpha,\beta) - 1/2)/\lambda^2$, and $\mu_2' = \lambda a(t,\alpha,\beta)/\sqrt{1+\lambda^2}$,

$$\pi(t) = \frac{\sqrt{1+\lambda^2}\nu\Phi(|\lambda|\mu_1')}{2|\lambda|f_{\textit{MMNEH}}(a(t,\alpha,\beta);0,1,\lambda,\nu)}\exp\left\{-\frac{\sqrt{1+\lambda^2}}{2\lambda}a(t,\alpha,\beta) + \frac{1}{8\lambda^2}\right\}.$$

Furthermore, for any $t \in \mathbb{R}^+$, and $k = 1,2,\ldots,$

$$E\left(U^k|T=t\right) = \pi(y)E\left(V_1^k\right) + (1-\pi(y))E(V_2^k),$$

where $V_1 \sim TN\left(\mu_1',\lambda^{-2};(0,\infty)\right)$, $V_2 \sim TN\left(\mu_2',(1+\lambda^2)^{-1};(0,\infty)\right)$ and

$$\begin{aligned}
E(V_1) =& \mu_1' + \frac{\phi\left(|\lambda|\mu_1'\right)}{|\lambda|\Phi\left(|\lambda|\mu_1'\right)},\\
E(V_1^k) =& \mu_1'E(V_1^{k-1}|T=t) + \frac{k-1}{\lambda^2}E(V_1^{k-2}|T=t), \quad k\geq 2,\\
E(V_2) =& \mu_2' + \frac{\phi\left(\lambda a(t,\alpha,\beta)\right)}{(1+\lambda^2)\Phi\left(\lambda a(t,\alpha,\beta)\right)},\\
E(V_2^k) =& \mu_2'E(V_2^{k-1}|T=t) + \frac{k-1}{1+\lambda^2}E(V_2^{k-2}|T=t), \quad k\geq 2.
\end{aligned}$$

Proof. Based on Theorm 3.6, the conditional pdf can be obtained by applying Baye's rule as

$$\begin{aligned}
f(u|t) =& \frac{f(t,u)}{f(t)} = \frac{f(t|u)f(u)}{f(t)} \\
=& \frac{\frac{A(t,\alpha,\beta)}{\sqrt{2\pi(1-\delta^2)}}\exp\left\{-\frac{1}{2(1-\delta^2)}\left(a(t,\alpha,\beta)-\frac{\lambda}{\sqrt{1+\lambda^2}}u\right)^2\right\}}{A(t,\alpha,\beta)f_{\text{MMNEH}}(a(t,\alpha,\beta);0,1,\lambda,\nu)} \\
&\times\left(\nu\frac{1}{2}\exp\left\{-\frac{u}{2}\right\}+2(1-\nu)\phi(u)\right) \\
=& \frac{\frac{\frac{\nu}{2}}{\sqrt{2\pi(1-\delta^2)}}\exp\left\{-\frac{1}{2(1-\delta^2)}\left(a(t,\alpha,\beta)-\frac{\lambda}{\sqrt{1+\lambda^2}}u\right)^2\right\}\mathrm{e}^{-\frac{u}{2}}}{f_{\text{MMNEH}}(a(t,\alpha,\beta);0,1,\lambda,\nu)} \\
&+\frac{\frac{2(1-\nu)}{\sqrt{2\pi(1-\delta^2)}}\exp\left\{-\frac{1}{2(1-\delta^2)}\left(a(t,\alpha,\beta)-\frac{\lambda}{\sqrt{1+\lambda^2}}u\right)^2\right\}\phi(u)}{f_{\text{MMNEH}}(a(t,\alpha,\beta);0,1,\lambda,\nu)}
\end{aligned}$$

After some algebraic manipulations, the resulting conditional distribution of U given $T=t$ is given by

$$f(u|t) = \pi(t)\frac{\phi\left(u;\mu_1',\lambda^{-2}\right)}{\Phi(|\lambda|\mu_1')} + (1-\pi(t))\frac{\phi\left(u;\mu_2',\frac{1}{1+\lambda^2}\right)}{\Phi(\lambda a(t,\alpha,\beta)}.$$

Thus, the conditional distribution of U given $T=t$ is mixture of two truncated normal with distributions $V_1 \sim \text{TN}\big(\mu_1',\lambda^{-2};\ (0,\infty)\big)$ and $V_2 \sim \text{TN}\big(\mu_2',(1+\lambda^2)^{-1};\ (0,\infty)\big)$, respectively, and mixing parameter $\pi(t)$. Based on some particular moments of the truncared normal distribution, the proof of conditional expectation is straightforward. □

The above theorems are used in obtaining the complete log-likelihood and conditional expectation for employing EM-type algorithm.

4 Finite mixture of mean-mixtures of normal Birnbaum-Saunders

In this section, the maximum likelihood (ML) estimate of the finite mixture of MMN-BS (FM-MMN-BS) distributions is obtained by implementing an

EM-type algorithm Dempster et al. ([12]. For the sake of notation, let $\boldsymbol{T} = (T_1, \ldots, T_n)$ be a vector of independent random samples identically arises from a FM-MMN-BS distributions. Then, the pdf of T_j for $j = 1, 2, \ldots, n$ is

$$f(t_j, \boldsymbol{\Theta}) = \sum_{i=1}^{g} \pi_i f_{\text{MMN-BS}}(t_j; \boldsymbol{\theta}_i), \qquad \pi_i \geq 0, \ \sum_{i=1}^{g} \pi_i = 1, \tag{9}$$

where $\boldsymbol{\Theta} = (\boldsymbol{\pi}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_g)$ with $\boldsymbol{\theta}_i = (\alpha_i, \beta_i, \lambda_i, \nu_i)$ and $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_{g-1})$. Define a set of latent component indicators $\boldsymbol{Z}_j = (Z_{1j}, \ldots, Z_{gj})^\top$ for each $j = 1, \ldots, n$, where $Z_{rj} = 1$ if T_j arises from the component r and otherwise zero. Therefore, it is convenient to assume Z_j is followed by a multinomial distribution with 1 trial and cell probabilities $\pi_1, \ldots, \pi_g$, denoted by $Z_j \sim \mathrm{M}(1; \pi_1, \ldots, \pi_g)$. This setting leads to obtain the hierarchical representation of (9) as

$$\begin{aligned} T_j | U_j, Z_{ij} = 1 &\sim \mathrm{EBS}(\alpha_i \sqrt{1-\delta_i^2}, \beta_i, -\frac{\delta_i u_j}{\sqrt{1-\delta_i^2}}), \\ U_j &\sim H(u; \boldsymbol{\nu}_i), \\ Z_{ij} &\sim M(1; \pi_i, \ldots, \pi_g), \end{aligned} \tag{10}$$

where $\delta_i = \lambda_i / \sqrt{1+\lambda_i^2}$. Considering the observed data $\boldsymbol{t} = (t_1, \ldots, t_n)^\top$ and latent variables $\boldsymbol{u} = (u_1, \ldots, u_n)$ and $\boldsymbol{Z} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)$, the hierarchical representation (10) results the complete-data log-likelihood function of $\boldsymbol{\Theta}$, ignoring constant values, as

$$\begin{aligned} \ell_c(\boldsymbol{\Theta} | \boldsymbol{t}, \boldsymbol{u}, \boldsymbol{Z}) = \sum_{j=1}^{n} \sum_{i=1}^{g} Z_{ij} \Bigg\{ &\log \pi_i - \log \left(\alpha_i \sqrt{1-\delta_i^2} \right) \\ &+ \log \left(\frac{t_j + \beta_i}{\sqrt{\beta_i}} \right) - \frac{1}{2(1-\delta_i^2)} (a(t_j, \alpha_i, \beta_i) - \delta_i u_j)^2 + \log h(u_j; \boldsymbol{\nu}_i) \Bigg\}. \end{aligned} \tag{11}$$

To obtain the ML estimate of parameters involved in (11), the expectation conditional maximization (ECM; Meng and Rubin [28]) algorithm is used. The ECM algorithm simplify the estimation procedure by breaking the maximization step into several conditional maximization (CM) steps and preserves convergence properties of the EM approach. The ECM algorithm for ML estimation of the FM-MMN-BS distributions proceeds as follows:

- Initialization: Set a reasonable starting values for $\mathbf{\Theta}$, as $\mathbf{\Theta}^{(k)}$, for the number of iteration $k = 0$. In our data analysis, the following procedure for automatically generating $\mathbf{\Theta}^{(0)}$ is exploited:

  1. Partition the data via the K-means algorithm and set $\hat{z}_{ij}^{(0)}$ as the resulting allocation membership. Then, for each cluster i, set $\hat{\pi}_i^{(0)} = \sum_{j=1}^{n} \hat{z}_{ij}^{(0)}/n$.
  2. The initial value of shapes and scales parameters $\hat{\alpha}_i^{(0)}$ and $\hat{\beta}_i^{(0)}$ can be created by the modified moment estimates proposed by Ng et al. [35] for the ith cluster.
  3. The initial skewness $\hat{\lambda}_i^{(0)}$'s to be zero and relatively $\hat{\nu}_i^{(0)} = 0.5$.

- Expectation (E) step: In iteration $k+1$, compute the so-called Q-function, defined as the expected value of the complete-data log-likelihood (11) with respect to the conditional distribution of $\boldsymbol{U}, Z$ given the observed data $\boldsymbol{t}$ and $\hat{\mathbf{\Theta}}^{(k)}$. Here, we need $\hat{u}_{1ij}^{(k)} = E(U_j|t_j, Z_{ij} = 1, \hat{\mathbf{\Theta}}^{(k)})$, $\hat{u}_{2ij}^{(k)} = E(U_j^2|t_j, Z_{ij} = 1, \hat{\mathbf{\Theta}}^{(k)})$, $\hat{\Psi}_{ij}^{(k)} = E\left[\log h(u_j; \boldsymbol{\nu})|t_j, Z_{ij} = 1, \hat{\mathbf{\Theta}}^{(k)}\right]$, and the posterior probability of t_j belong to the ith component of the mixture as

$$\hat{z}_{ij}^{(k)} = E(Z_{ij}|t_j, \hat{\Theta}^{(k)}) = \frac{\hat{\pi}_i^{(k)} f_{\text{MMN-BS}}(t_j, \hat{\theta}_i^{(k)})}{f(t_j, \hat{\mathbf{\Theta}}^{(k)})}.$$

  These results in the Q-function written as

$$Q(\mathbf{\Theta}|\hat{\mathbf{\Theta}}^{(k)}) = \sum_{j=1}^{n}\sum_{i=1}^{G} \hat{z}_{ij}\Big\{ \log \pi_i - \log(\alpha_i) - \frac{1}{2}\log(1-\delta_i^2) + \log(\frac{t_j+\beta_i}{\sqrt{\beta_i}}) - \frac{1}{2(1-\delta_i^2)}(a^2(t_j;\alpha_i,\beta_i) - 2a(t_j,\alpha_i,\beta_i)\delta_i\hat{u}_{1ij} + \delta_i^2\hat{u}_{2ij}) + \hat{\Psi}_{ij}^{(k)}\Big\}.$$

- CM-steps: Maximizing Q-function with respect to the unknown parameters leads to the following CM steps.

CM1: Calculate $\hat{\pi}_i^{(k+1)} = \hat{n}_i^{(k)}/n$ where $\hat{n}_i^{(k)} = \sum_{j=1}^{n} \hat{z}_{ij}^{(k)}$.

CM2: Update $\hat{\alpha}_i^{(k)}$ and $\hat{\delta}_i^{(k)}$ by

$$\hat{\alpha^2}_i^{(k+1)} = \frac{\sum_{j=1}^n \hat{z}_{ij}^{(k)}\ \eta^2(t_j,\hat{\beta}_i^{(k)})}{\sum_{j=1}^n \hat{z}_{ij}^{(k)}} + \left[1-\frac{\sum_{j=1}^n \hat{z}_{ij}^{(k)}\hat{u}_{2ij}^{(k)}}{\sum_{j=1}^n \hat{z}_{ij}^{(k)}}\right] \left[\frac{\sum_{j=1}^n \hat{z}_{ij}^{(k)}\hat{u}_{1ij}^{(k)}\ \eta(t_j,\hat{\beta}_i^{(k)})}{\sum_{j=1}^n \hat{z}_{ij}^{(k)}\hat{u}_{2ij}^{(k)}}\right]^2,$$

$$\hat{\delta}_i^{(k+1)} = \frac{\sum_{j=1}^n \hat{z}_{ij}^{(k)}\hat{u}_{1ij}^{(k)}\ \eta(t_j,\hat{\beta}_i^{(k)})}{\hat{\alpha}_i^{(k)}\sum_{j=1}^n \hat{z}_{ij}^{(k)}\hat{u}_{2ij}^{(k)}},$$

where $\eta(t_j,\hat{\beta}_i^{(k)}) = \sqrt{t_j/\hat{\beta}_i^{(k)}} - \sqrt{\hat{\beta}_i^{(k)}/t_j}$. Consequently, $\hat{\lambda}_i^{(k+1)} = \hat{\delta}_i^{(k+1)}/\sqrt{1-\hat{\delta}_i^{2(k+1)}}$.

CM3: For the fixed $\hat{\alpha}_i^{(k+1)}$ and $\hat{\delta}_i^{(k+1)}$, update $\hat{\beta}_i^{(k+1)}$ using $\hat{\beta}_i^{(k+1)} = \arg\max_{\beta_i} \ell_\beta(\hat{\Theta}^{(k)})$, where

$$\ell_\beta(\hat{\Theta}^{(k)}) = \sum_{j=1}^n \hat{z}_{ij}\Bigg\{\log\pi_i - \log(\alpha_i) - \frac{1}{2}\log(1-\delta_i^2) + \log(\frac{t_j+\beta_i}{\sqrt{\beta_i}}) - \frac{1}{2(1-\delta_i^2)}(a^2(t_j;\alpha_i,\beta_i) - 2a(t_j,\alpha_i,\beta_i)\delta_i\hat{u}_{1ij} + \delta_i^2\hat{u}_{2ij}) + \hat{\Psi}_{ij}^{(k)}\Bigg\}.$$

CM3: The update of $\boldsymbol{\nu}_i$ depends on the distribution of latent variable U and can be obtained from $\hat{\boldsymbol{\nu}}_i = \arg\max_{\boldsymbol{\nu}} \sum_{j=1}^n \hat{z}_{ij}\hat{\Psi}_{ij}^{(k)}$.

4.1 Some practical implementation aspects

Remark 4.1. For the spacial cases considered in Section 3, the closed form of the $\hat{u}_{1ij}^{(k)}$ and $\hat{u}_{2ij}^{(k)}$ can be obtained by Proposition 3.5, and Theorems 3.6 and 3.7. We also note both SN-BS and MMNE-BS distributions do not have extra parameter $\boldsymbol{\nu}$, however, for the MMNEH-BS the explicit expression for $\Psi_{ij}^{(k)}$ is difficult to obtain. Thus, we recommend to use the expectation conditional maximization either (ECME; Liu and Rubin [25]) algorithm by maximizing a simpler constrained log-likelihood function constituted on the basis of $(\boldsymbol{t},\boldsymbol{Z})$, in order to update $\hat{\nu}_i^{(k)}$. This yield to replace the following CML estimate of $\boldsymbol{\nu}_i$

$$\hat{\nu}_i^{(k+1)} = \arg\max_{\nu_i} \left\{ \sum_{j=1}^{n} \hat{z}_{ij} \log f_{\text{MMNEH-BS}} \left(t_j; \hat{\alpha}_i^{(k+1)}, \hat{\beta}_i^{(k+1)}, \hat{\lambda}_i^{(k+1)}, \nu_i \right) \right\}.$$

Remark 4.2. Stopping rule. The above E- and CM-steps of the ECM algorithm are iterated until either the number of iterations exceeds the limit ($K_{\max}$) or a convergence rule is satisfied. In our data analysis, Aitken acceleration Aitken [1] is used as a stopping criterion. Based on Aitken approach, the algorithm is considered to have converged if the increment, $\ell_\infty^{(k+1)} - \ell^{(k)}$, is less than a prescribed tolerance, ϵ, where the asymptotic estimate of the log-likelihood is given as

$$\ell_\infty^{(k+1)} = \frac{\ell^{(k+2)}\ell^{(k)} - {\ell^{(k+1)}}^2}{\ell^{(k+2)} - 2\ell^{(k+1)} + \ell^{(k)}},$$

and $\ell^{(k)} = \sum_{j=1}^{j} \log \sum_{i=1}^{g} \pi_i f_{\text{MMN-BS}}(t_j; \boldsymbol{\theta}_i)$ is a maximized log-likelihood at kth iteration. In experimental study, we choose $K_{\max} = 5000$ and $\epsilon = 10^{-6}$.

Remark 4.3. Model selection and goodness of fit test. The most commonly used Akaike Information Criterion (AIC; Akaike [2]) and the Bayesian Information Criterion (BIC; Schwarz [38]) are considered as the model comparison measures. These two criteria can be formulated by $mc_n - 2\ell(\hat{\boldsymbol{\Theta}})$, where m is the number of free parameters and c_n denotes the penalty term that is $c_n = 2$ for AIC and $c_n = \log(n)$ for BIC. In general, the smaller the values of these statistics, the better the fit to the data. We also apply the Kolmogorov-Smirnov (KS;Smirnov [40]) test to assess the goodness-of-fit of the fitted distributions. The KS test, defined theoretically as a distance between the empirical CDF and the estimated theoretical CDF for the model, is a measure to understand how well the theoretical distribution fit the empirical data.

4.2 Observed information-based standard errors

Following Basford et al. [8], we obtain the observed information matrix for obtaining the standard error of ML estimate of parameters. Theoretically, the outer product of the gradient (score) vectors defined as a relatively simple way to approximate the observed information matrix, is given by

$$\hat{I}_e = \sum_{j=1}^{n} \hat{\boldsymbol{s}}_j \hat{\boldsymbol{s}}_j^T, \tag{12}$$

where for the complete-data log-likelihood of the individual observation, $\ell_c(\boldsymbol{\Theta} \mid t_j, u_j, \boldsymbol{z}_j)$,

$$\hat{\boldsymbol{s}}_j = \frac{\partial f(t_j; \boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}} = E\left[\frac{\partial \ell_{cj}(\boldsymbol{\Theta} \mid t_j, u_j, \boldsymbol{z}_j)}{\partial \boldsymbol{\Theta}}\Big| t_j\right], \qquad j = 1, \ldots, n.$$

For the FM-MMN-BS distributions, we have

$$\begin{aligned}\ell_{cj}(\boldsymbol{\Theta}|t_j, u_j, \boldsymbol{Z}_j) = \sum_{i=1}^{G} Z_{ij} \Big\{ & \log \pi_i - \log(\alpha_i \sqrt{1-\delta_i^2}) + \log(\frac{t_j + \beta_i}{\sqrt{\beta_i}}) \\ & - \frac{1}{2(1-\delta_i^2)}(a(t_j, \alpha_i, \beta_i) - \delta_i u_j)^2 + \log h(u_j; \boldsymbol{\nu}_i) \Big\}. \end{aligned} \tag{13}$$

Thus, each gradient vector $\hat{\boldsymbol{s}}_j = (\hat{\boldsymbol{s}}_{j,\pi_1}, \ldots, \hat{\boldsymbol{s}}_{j,\pi_{G-1}}, \hat{\boldsymbol{s}}_{j,\alpha_1}, \ldots, \hat{\boldsymbol{s}}_{j,\alpha_G}, \hat{\boldsymbol{s}}_{j,\beta_1}, \ldots \hat{\boldsymbol{s}}_{j,\beta_G}, \hat{\boldsymbol{s}}_{j,\lambda_1}, \ldots, \hat{\boldsymbol{s}}_{j,\lambda_G}, \hat{\boldsymbol{s}}_{j,\nu_1}, \ldots, \hat{\boldsymbol{s}}_{j,\nu_G})$ consists the following elements

$$\begin{aligned}
\hat{s}_{j,\pi_i} =& \frac{\hat{z}_{ij}}{\hat{\pi}_i} - \frac{\hat{z}_{Gj}}{\hat{\pi}_G}, \\
\hat{\boldsymbol{s}}_{j,\alpha_i} =& \hat{z}_{ij}\left(\frac{-1}{\hat{\alpha}_i} + \frac{1}{\hat{\alpha}_i^3(1-\hat{\delta}_i^2)}\Big(\frac{t_j}{\hat{\beta}_r} + \frac{\hat{\beta}_r}{t_j} - 2\Big) - \frac{\hat{\delta}_i}{\hat{\alpha}_i^2(1-\hat{\delta}_i^2)}\eta\big(t_j, \hat{\beta}_i\big)\hat{u}_{1ij}\right), \\
\hat{\boldsymbol{s}}_{j,\lambda_i} =& \hat{z}_{ij}\left(\frac{\hat{\lambda}_i}{1+\hat{\lambda}_i^2} - \hat{\lambda}_i a^2(t_j, \hat{\alpha}_i, \hat{\beta}_i) - \hat{\lambda}_i \hat{u}_{2ij} - \frac{1}{(1+\hat{\lambda}_i^2)^{1.5}}\hat{u}_{1ij} a(t_j, \hat{\alpha}_i, \hat{\beta}_i)\right), \\
\hat{\boldsymbol{s}}_{j,\beta_i} =& \hat{z}_{ij}\Big(\frac{-1}{2\hat{\beta}_i} - \frac{1}{2\hat{\alpha}_i^2(1-\hat{\delta}_i^2)}\Big(\frac{1}{t_j} - \frac{t_j}{\hat{\beta}_i^2}\Big) \\
& + \frac{1}{t_j + \hat{\beta}_i} + \frac{\hat{\delta}_i \hat{u}_{1ij}}{2\hat{\beta}_i \hat{\alpha}_i (1-\hat{\delta}_i^2)}\big(\sqrt{t_j/\hat{\beta}_i} + \sqrt{\hat{\beta}_i/t_j}\big)\Big), \\
\hat{s}_{j,\boldsymbol{\nu}_i} =& E\left(\frac{\partial \log f(u; \boldsymbol{\nu})}{\partial \boldsymbol{\nu}} \mid t_j, Z_{ij} = 1, \hat{\boldsymbol{\Theta}}\right).
\end{aligned}$$

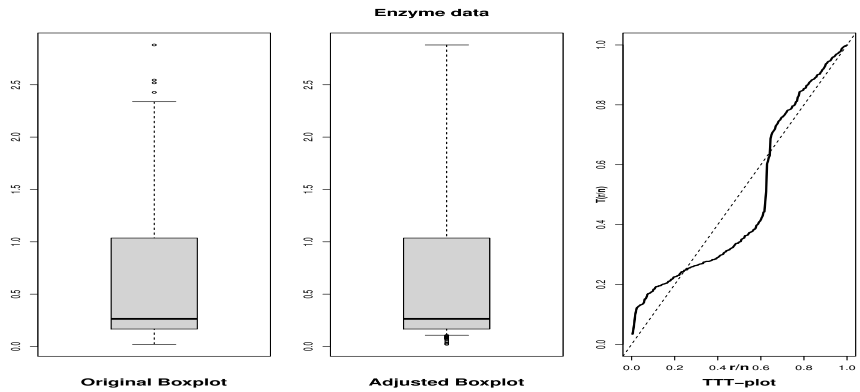

Figure 2: The standard box plot, adjusted box-plot and TTT plot for Enzyme data.

where $\hat{z}_{ij}$, $\hat{u}_{1ij}$ and $\hat{u}_{2ij}$ are the conditional expectations evaluated at $\hat{\boldsymbol{\Theta}}$. As a result, the standard error of parameters are obtained as the square roots of the diagonal elements of the inverse of (12).

5 Real Data Analysis

5.1 Enzyme data

In order to illustrate the utility of the proposed FM-MMN-BS distributions to the real dataset, the Enzyme data is considered. The Enzyme data analyzed previously by Bechtel [9] is related to the enzymatic activity in the blood. Each point of the 245 observations represent the metabolism of carcinogenic substances. Bechtel [9] concluded that the mixture of two right-skewed distributions is suitable for analyzing these Enzyme data that can be seen from the TTT plot presented in figure 2. Moreover, the adjusted box-plot indicate drown in figure 2 that some atypical observations are available on the left tail. These motivate us to fit two-component FM of Weibull (FM-Weibull), FM of gamma (FM-gamma), FM of BS (FM-BS), and mixture of Length-biased BS and BS distributions (LBBS) and mixture of Length-biased BS and BS distributions with the same parameters (LBSBS) proposed in Balakrishnan et al. [6], and tree subclasses of FM-MMN-BS distributions.

Table 3: ML estimates with their standard error and KS distances with their associated p-values for the considered mixture models fitted to the Enzyme dataset.

	FM-Weibull		FM-gamma		FM-BS		LBBS		LBSBS		FM-MMNE-BS		FM-MMNEH-BS		FM-SN-BS	
parameter	MLE	SE	MLE	SE	MLE	SE	MLE	SE	MLE	SE	MLE	SE	MLE	SE	MLE	SE
π	0.444	0.013	0.349	0.011	0.629	0.031	0.450	0.028	0.417	0.026	0.624	0.009	0.624	0.012	0.627	0.021
α_1	1.288	0.020	0.674	0.025	0.533	0.032	0.365	0.034	1.038	0.084	0.327	0.014	0.310	0.011	0.573	0.062
β_1	1.145	0.296	0.022	0.012	0.175	0.007	0.171	0.007	0.216	0.012	0.258	0.013	0.257	0.010	0.140	0.017
λ_1	–	–	–	–	–		–		–		-1.227	0.172	-1.095	0.186	0.533	0.091
α_2	2.681	0.193	0.039	0.016	0.319	0.025	1.274	0.114	1.038	0.084	0.242	0.009	0.227	0.017	0.476	0.050
β_2	0.184	0.039	8.182	0.039	1.274	0.043	0.213	0.044	0.216	0.012	1.005	0.035	1.006	0.044	0.901	0.105
λ_2	–	–	–	–	–		–		–		0.953	0.130	0.808	0.188	1.229	0.773
ν	–	–	–	–	–		–		–		–	–	0.364	0.015	–	
$\ell_{\max}$	-71.741		-76.735		-59.168		-71.091		-115.899		**-43.220**		-44.674		-47.370	
AIC	153.482		163.47		128.336		152.182		237.798		**100.41**		105.34		108.74	
BIC	170.988		180.976		145.842		169.688		248.302		**124.92**		133.35		133.248	
KS	0.117		0.125		0.053		0.111		0.151		**0.036**		0.047		0.043	
p-value	0.004		0.003		0.507		0.005		< 0.001		**0.923**		0.650		0.742	

By applying the EM-type algorithm to the considered model, we obtain ML estimates, maximized log-likelihood values ($\ell_{\max}$) and corresponding AIC and BIC. Results summarized in Table 3 show that the FM-MMN-BS distributions provides a highly improved fit to the data over the others. It can be seen that the subclasses of FM-MMNBS models yields quite smaller standard errors for the ML parameter estimates over the other distributions. This means that the FM-MMN-BS distributions allows to produce more precise estimates for this data example. Moreover, the results of KS test depicted in Table 3 reveals that the p-value of the FM-MMNE-BS model is significantly grater than the FM-Weibull, FM-gamma, FM-BS, LBBS and LBSBS, FM-MMNEH-BS and FM-SN-BS models, which strongly suggests that the Enzyme data follow a mixture of FM-MMNE-BS distributions. This outperformance of FM-MMNE-BS distributions can be observe form figure 3 which present graphical visualization of the fitted densities and the PP-plots of the three best fitted models.

5.2 South Pole data

In the second real data example, the monthly average carbon dioxide readings gathered by the Earth System Research Laboratory of the U.S. National Oceanic and Atmospheric Administration is used. The dataset that is available in the `Stat2Data` package of R, is collected from 1988 to 2016 at the South Pole and originally contains five variables average carbon dioxide, Years, Month, Atmospheric carbon dioxide level (CO2) and Time interval.

We fit the FM-BS, FM-gamma and FM-weibull, and three sub-model of the FM-MMN-BS distribution to the data by ranging $g = 1$ to 3. Table 4

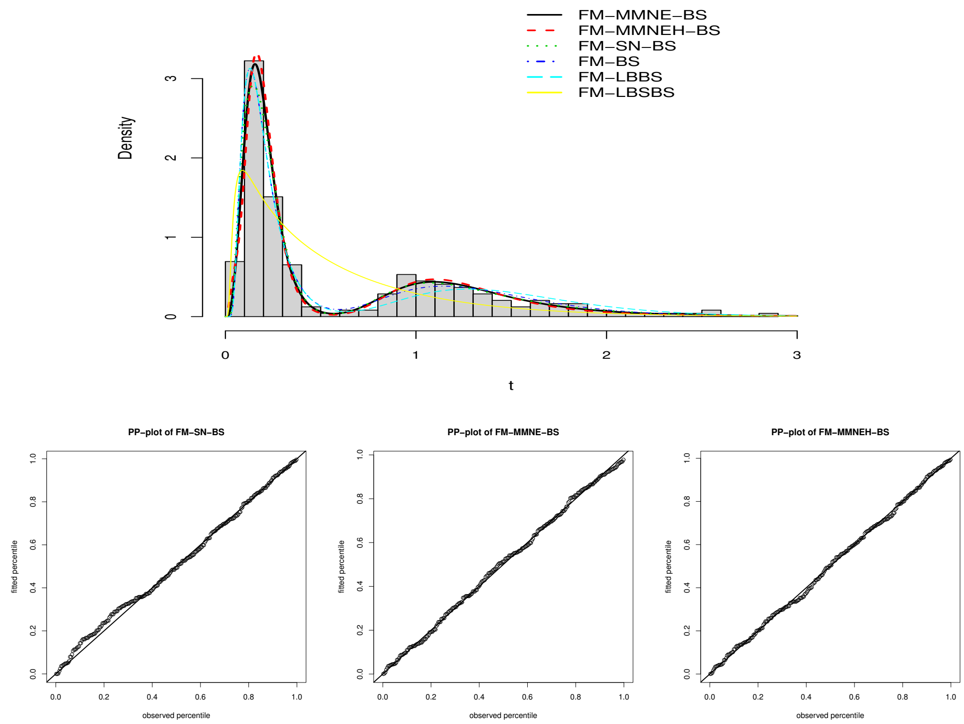

Figure 3: Histogram of the Enzyme data overlaid with six fitted two component mixture densities, and *pp*-plot of the three best models.

shows the $\ell_{\max}$, the number of free parameters (m), AIC and the BIC values. It is observed form the AIC and BIC that the two-component FM-MMNEH-BS distributions outperforms the other models. Table 5 reports the parameter estimates of the best chosen models along with their standard errors. Result of KS test strongly suggests that the considered data follow a mixture of FM-MMNEH-BS distributions. This outperformance of the FM-MMNEH-BS distributions can be observe form the histogram of data and the relative PP-plots for the three best fitted models in figure 4.

Table 4: Estimation performance of models fitted to the South Pole data.

Model	g	m	$\ell_{\max}$	AIC	BIC
FM-gamma	1	2	-1488.52	2981.07	2988.79
	2	5	-1453.12	2916.24	2935.50
	3	8	-1412.55	2841.10	2871.92
FM-weibull	1	2	-1441.86	2887.72	2895.42
	2	5	-1430.11	2870.22	2889.48
	3	8	-1408.41	2832.82	2863.64
FM-BS	1	2	-1442.89	2889.78	2897.49
	2	5	-1399.70	2809.41	2828.67
	3	8	-1389.19	2794.39	2823.20
FM-SN-BS	1	3	-1439.36	2884.72	2896.28
	2	7	-1397.58	2809.15	2836.12
	3	11	-1386.89	2795.79	2838.16
FM-MMNE-BS	1	3	-1439.36	2884.72	2896.28
	2	7	-1389.72	2793.45	2820.42
	3	11	-1383.48	2788.96	2831.33
FM-MMNEH-BS	1	4	-1438.73	2885.46	2900.87
	2	9	-1382.17	**2782.34**	**2817.01**
	3	14	-1382.41	2792.82	2846.75

6 Simulation Study

6.1 Finite sample properties of ML estimates

The first simulation experiment is conducted aiming at verifying finite sample properties of ML estimates. In each 500 trails, artificial samples from three-component FM-SN-BS, FM-MMNE-BS and FM-MMNEH-BS distributions are generated through applying the stochastic representation in (6). For each model four sample sizes n 100, 200, 500 and 1000 is considered. The true parameters are reported in Tables 7 and 8. For each synthetic data set of the FM-SN-BS, FM-MMNE-BS and FM-MMNEH-BS models, the corresponding model is fitted using the ECM algorithm and the parameter estimates are obtained. Then, the average values, standard deviations (Std), absolute bias

Table 5: ML parameter estimates with their standard error and the KS distances together with its corresponding p-values for the four considered mixture models fitted to the South Pole data for $g = 2$.

	FM-BS		FM-MMNE-BS		FM-MMNEH-BS		FM-SN-BS	
parameter	MLE	SE	MLE	SE	MLE	SE	MLE	SE
π	0.713	0.140	0.557	0.086	0.612	0.104	0.718	0.081
α_1	0.032	0.012	0.028	0.009	0.037	0.011	0.033	0.011
β_1	378.809	21.860	381.904	10.278	391.709	18.291	378.721	15.819
λ_1	–	–	0.445	0.084	0.425	0.095	-0.980	0.133
α_2	0.010	0.001	0.052	0.006	0.049	0.010	0.010	0.002
β_2	354.793	17.420	350.498	9.783	356.807	19.580	354.764	12.746
λ_2	–	–	5.034	0.981	4.942	1.088	-0.118	0.082
ν_1	–	–	–	–	0.626	0.096	–	
ν_2	–	–	–	–	0.603	0.110	–	
KS	0.062		0.035		**0.032**		0.055	
p-value	0.412		0.940		**0.980**		0.469	

(AB) and the mean squared error (MSE) of ML estimates are computed, where

$$\text{AB} = \frac{1}{500}\sum_{j=1}^{500}\left|\hat{\theta}^{(j)} - \theta_{true}\right| \qquad \text{and} \qquad \text{MSE} = \frac{1}{500}\sum_{j=1}^{500}\left(\hat{\theta}^{(j)} - \theta_{true}\right)^2,$$

in which $\hat{\theta}^{(j)}$ is the ML estimate of θ_{true} obtained from the j-th replicate. The numerical results are reported in Tables 6, 7 and 8. It can be observed that these three Tables that the mean of parameter estimates are very closed to the true values and the increase of sample size leads to have small value of Std. As can be expected, the AB and MSE values approach zero as the sample size n increases and tends to zero, showing empirically the asymptotic unbiasedness and the consistency of the ML estimates obtained via the ECM algorithm.

6.2 Comparison of fitting and clustering performance

In this simulation study, we suppose that X in representation (1) is followed by the normal inverse Gaussian (NIG) distribution and generate three-component mixture data form it. It is noted that random sample from the NIG distribution with parameter $(\mu, \sigma^2, \lambda, \chi, \psi)$ can be generated from

$$\mu + W\lambda + \sqrt{W}Z,$$

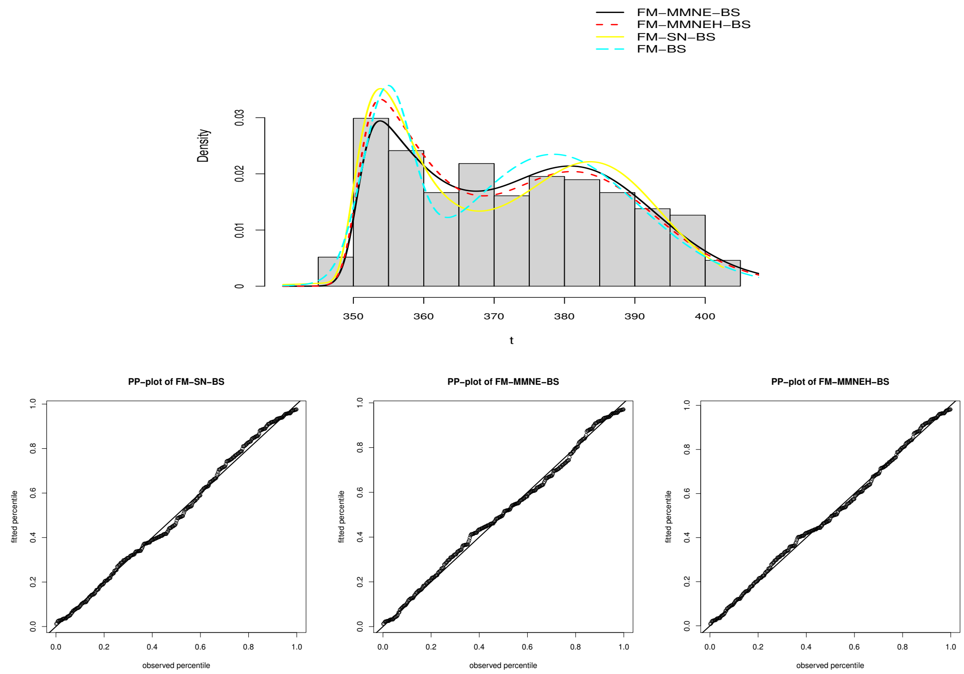

Figure 4: Histogram of the South Pole data overlaid with four fitted two-component mixture densities and *pp*-plot of the three best fitted models.

where $Z \sim \mathrm{N}(0, \sigma^2)$ and W, independently of Z, is followed by the generalized inverse Gaussian (GIG) distribution with parameter $(-0.5, \chi, \psi)$. Details on GIG distribution can be found in Good [14]. The NIG model can provide a reasonable platform for generating asymmetric data with the desired level of skewness and leptokurtosis. By setting $\mu = 0$ and $\sigma = 1$, the three-component mixture data is generated by using the presumed parameters

$$\pi_1 = 2/7, \quad \pi_2 = 2/7, \quad \pi_3 = 3/7, \quad \psi_1 = 5, \quad \psi_2 = 7, \quad \psi_3 = 5, \quad \chi_1 = 4,$$
$$\chi_2 = 8, \ \ \chi_3 = 6, \ \ \alpha_1 = 0.5, \quad \alpha_2 = 1, \quad \alpha_3 = 2.5, \quad \beta_1 = 2, \quad \beta_2 = 2, \ \ \beta_3 = 1.$$

In each replication, we fit the proposed FM-Weibull, FM-gamma, FM-SN-BS, FM-MMNE-BS, FM-MMNEH-BS and FM-BS models to the generated data and obtain AIC and BIC as the model performance criteria and adjusted

Table 6: Mean, Std, AB and MSE for EM estimates over 500 samples from the FM-SN-BS model (true parameter in pretenses).

n	Measure	$\alpha_1(1)$	$\alpha_2(1)$	$\alpha_3(3)$	$\beta_1(2)$	$\beta_2(4)$	$\beta_3(3)$	$\lambda_1(2.6)$	$\lambda_2(1.4)$	$\lambda_3(1.8)$	$\pi_1(0.4)$	$\pi_2(0.3)$
100	Mean	0.9578	0.8991	2.8772	1.9978	3.9797	2.9547	2.5725	1.3299	1.7715	0.4627	0.2543
	Std	0.2127	0.3963	0.5981	0.1980	0.3805	0.5690	0.0920	0.1528	0.3386	0.1820	0.1570
	AB	0.0422	0.1009	0.1228	0.0022	0.0203	0.0453	0.0275	0.3999	0.7715	0.1627	0.1505
	MSE	0.0461	0.1641	0.3656	0.0384	0.1423	0.3194	0.0090	0.1828	0.7076	0.1569	0.1341
200	Mean	0.9759	0.9460	2.9457	2.0168	4.0353	3.0493	2.5790	1.4032	1.7558	0.4558	0.2677
	Std	0.1605	0.2787	0.4228	0.1215	0.2421	0.3818	0.0657	0.1127	0.2627	0.1652	0.1410
	AB	0.0241	0.0540	0.0543	0.0168	0.0353	0.0493	0.0210	0.4032	0.7558	0.1558	0.1392
	MSE	0.0258	0.0790	0.1781	0.0147	0.0587	0.1453	0.0047	0.1750	0.6388	0.1396	0.1256
500	Mean	0.9962	0.9939	3.0028	2.0106	4.0191	2.0383	2.5937	1.4102	1.7577	0.4353	0.2806
	Std	0.0935	0.1628	0.2353	0.0825	0.1529	0.2274	0.0383	0.0620	0.1545	0.1463	0.1295
	AB	0.0038	0.0061	0.0028	0.0106	0.0191	0.0383	0.0063	0.4102	0.7577	0.1133	0.1351
	MSE	0.0086	0.0260	0.0543	0.0068	0.0233	0.0522	0.0015	0.1721	0.5975	0.1295	0.1207
1000	Mean	0.9935	0.9858	2.9841	2.0057	4.0190	2.0369	2.5958	1.4080	1.7879	0.4227	0.2889
	Std	0.0598	0.1041	0.1601	0.0561	0.1035	0.1548	0.0311	0.0430	0.0946	0.1265	0.1124
	AB	0.0065	0.0142	0.0159	0.0127	0.0190	0.0369	0.0042	0.4080	0.7879	0.0927	0.1268
	MSE	0.0036	0.0108	0.0254	0.0032	0.0109	0.0248	0.0012	0.1682	0.6295	0.1066	0.1149

rank index (AIR; Hubert and Arabie [17]) as a clustering performance measure. Table 9 summarizes the fitting results averaged over 300 trials and the average ARI values. From the table, the FM-MMNE-BS distribution provides the best overall fit in terms of AIC or BIC and an improved classification accuracy (ARI=0.856 and =0.824).

7 Conclusion

This paper has introduced a new extension of the BS distribution as well as its finite mixture model, called FM-MMN-BS distributions. We present the hierarchical stochastic representation of the FM-MMN-BS distribution for implementing a feasible and effective ECM algorithm to obtain the ML estimate of parameters. The asymptotic information matrix is also derived by offering an information-based approach. Numerical results illustrated in Section 5 indicate that the FM-MMN-BS model can be well suited to the experimental data. By conducting two simulation studies, the finite sample properties of the ML estimates as well as the ability of the FM-MMN-BS distributions for clustering heterogeneous right-skewed and heavy tails data are examined. Numerical results of simulation 2 suggest that the proposed FM-

Table 7: Mean, Std, AB and MSE for EM estimates over 500 samples from the FM-MMNE-BS model (true parameter in pretenses).

n	Measure	$\alpha_1(1)$	$\alpha_2(1)$	$\alpha_3(3)$	$\beta_1(2)$	$\beta_2(4)$	$\beta_3(3)$	$\lambda_1(2.6)$	$\lambda_2(1.4)$	$\lambda_3(1.8)$	$\pi_1(0.4)$	$\pi_2(0.3)$
100	Mean	0.9734	0.9354	2.9051	2.0212	3.9813	2.9704	2.5767	1.4216	1.7505	0.4482	0.2619
	Std	0.2144	0.4104	0.6061	0.1953	0.3618	0.5479	0.0924	0.2404	0.3198	0.1933	0.1751
	AB	0.0266	0.0646	0.0949	0.0212	0.0197	0.0596	0.0233	0.0316	0.0705	0.1805	0.1693
	MSE	0.0462	0.1710	0.3728	0.0378	0.1300	0.2974	0.0090	0.1808	0.0950	0.2017	0.1632
200	Mean	0.9798	0.9558	2.9358	2.0143	4.0198	3.0357	2.5800	1.4170	1.7531	0.4374	0.2710
	Std	0.1594	0.2876	0.4292	0.1410	0.2587	0.3933	0.0623	0.2037	0.2457	0.1828	0.1517
	AB	0.0202	0.0442	0.0642	0.0143	0.0168	0.0457	0.0200	0.0270	0.0531	0.1632	0.1522
	MSE	0.0256	0.0839	0.1865	0.0197	0.0665	0.1544	0.0042	0.1763	0.0669	0.1721	0.1457
500	Mean	0.9829	0.9886	2.9742	2.0106	4.0149	3.0381	2.5929	1.4155	1.7728	0.4220	0.2770
	Std	0.0947	0.1707	0.2527	0.0824	0.1549	0.2302	0.0385	0.1910	0.1403	0.1733	0.1462
	AB	0.0071	0.0114	0.0358	0.0106	0.0149	0.0381	0.0071	0.0155	0.0328	0.1498	0.1378
	MSE	0.0089	0.0290	0.0635	0.0068	0.0240	0.0539	0.0015	0.1682	0.0368	0.1425	0.1390
1000	Mean	0.9948	0.9893	2.9887	2.0094	4.0103	3.0328	2.5934	1.4105	1.7818	0.4183	0.2835
	Std	0.0639	0.1141	0.1794	0.0596	0.1139	0.1709	0.0305	0.1476	0.0988	0.1404	0.1294
	AB	0.0062	0.0097	0.0213	0.0094	0.0123	0.0328	0.0066	0.0105	0.0218	0.1353	0.1263
	MSE	0.0040	0.0129	0.0319	0.0036	0.0131	0.0300	0.0010	0.1407	0.0209	0.1377	0.1362

Table 8: Mean, Std, AB and MSE for EM estimates over 500 samples from the FM-MMNEH-BS model (true parameter in pretenses).

n	Measure	$\alpha_1(1)$	$\alpha_2(1)$	$\alpha_3(3)$	$\beta_1(2)$	$\beta_2(4)$	$\beta_3(3)$	$\lambda_1(2.6)$	$\lambda_2(1.4)$	$\lambda_3(1.8)$	$\nu(0.4)$	$\pi_1(0.4)$	$\pi_2(0.3)$
100	Mean	1.2682	1.5464	3.7745	2.5537	4.5902	3.6550	2.5335	1.3551	1.7466	0.4734	0.4557	0.2471
	Std	0.4082	0.8144	1.2181	0.2510	0.5081	0.7444	0.1076	0.1554	0.3167	0.2443	0.1709	0.1570
	AB	0.0382	0.0664	0.0745	0.0537	0.0402	0.0550	0.0335	0.0951	0.0966	0.2034	0.1862	0.1557
	MSE	0.0472	0.1263	0.2247	0.0690	0.1443	0.0878	0.0215	0.0800	0.0662	0.0219	0.1933	0.1736
200	Mean	1.2320	1.4670	3.7008	2.5092	4.3098	3.4845	2.5592	1.3710	1.7517	0.4626	0.4306	0.2681
	Std	0.4409	0.8678	1.3165	0.2019	0.3973	0.5883	0.0718	0.0966	0.1955	0.2420	0.1691	0.1514
	AB	0.0320	0.0570	0.0608	0.0392	0.0298	0.0445	0.0208	0.0610	0.0817	0.1126	0.1770	0.1347
	MSE	0.0364	0.0840	0.1280	0.0531	0.0881	0.0604	0.0151	0.0701	0.0489	0.0182	0.1765	0.1685
500	Mean	1.1065	1.2169	3.3266	2.3207	4.2337	3.3585	2.5674	1.3743	1.7608	0.4543	0.4231	0.2713
	Std	0.1622	0.3213	0.4794	0.1277	0.2549	0.3721	0.0404	0.0656	0.1402	0.1225	0.1556	0.1344
	AB	0.0265	0.0369	0.0466	0.0207	0.0137	0.0385	0.0126	0.0543	0.0568	0.0943	0.1563	0.1258
	MSE	0.0294	0.0493	0.0844	0.0373	0.0429	0.0362	0.0116	0.0597	0.0244	0.0132	0.1502	0.1411
1000	Mean	1.0953	1.1874	3.2788	2.1190	4.1425	3.2696	2.5796	1.3855	1.7695	0.4223	0.4188	0.2865
	Std	0.1175	0.2335	0.3509	0.0949	0.1815	0.2660	0.0316	0.0478	0.0951	0.0955	0.1330	0.1275
	AB	0.0153	0.0274	0.0308	0.0120	0.0125	0.0296	0.0104	0.0355	0.0295	0.0623	0.1230	0.1081
	MSE	0.0178	0.0291	0.0597	0.0283	0.0294	0.0137	0.0110	0.0287	0.0182	0.0118	0.1232	0.1293

Table 9: Performance of various BS type models fitted in simulation 2. (m is the number of free parameters)

Model	$\ell_{\max}$	m	AIC	BIC	ARI
FM-gamma	-503.70	8	1023.40	1057.12	0.682
FM-Weibull	-491.13	8	998.26	1031.98	0.726
FM-BS	-462.33	8	940.66	974.38	0.788
FM-SN-BS	-441.39	11	904.78	951.14	0.802
FM-MMNE-BS	-430.15	11	882.30	928.66	0.856
FM-MMNEH-BS	-428.73	14	885.46	944.46	0.824

MMNE-BS and FM-MMNEH-BS models can outperform the well established alternatives in providing better density estimation and an improvement in the clustering.

Appendix A

Let $T \sim$ MMN-BS$(\alpha, \beta, \lambda, \nu)$ and $Y \sim$ MMN$(0, 1, \lambda, \nu)$. In order to calculate skewness and kurtosis of T, by (6) and simple mathematical work, we have

$$
\begin{aligned}
E(T) =& \frac{1}{2}\beta\alpha^2 E(Y^2) + 1 + \frac{1}{2}\alpha\beta V_1,\\
E(T^2) =& \frac{1}{2}\beta^2\alpha^4 E(Y^4) + 1 + \alpha^2\beta(1+\beta)E(Y^2) + \alpha\beta V_1 + \frac{1}{2}\alpha^3\beta^2 V_3,\\
E(T^3) =& 1 + \frac{1}{2}\beta^3\alpha^6 E(Y^6) + \frac{1}{2}\alpha^5\beta^3 V_5 + \frac{3}{2}\alpha^4\beta^2(\beta+1)E(Y^4)\\
& + \frac{3}{2}\alpha^3\beta^2(\beta+1)V_3 + 3\alpha^2\beta(\beta+\frac{3}{2})E(Y^2) + \frac{3}{2}\alpha\beta V_1,\\
E(T^4) =& 1 + \frac{1}{4}\beta^4\alpha^8 E(Y^8) + \frac{1}{2}\beta^4\alpha^7 V_7 + \beta^3\alpha^6(\frac{3}{2}+\beta)E(Y^6)\\
& + \beta^3\alpha^5(2+\beta)V_5 + \beta^2\alpha^4(3+4\beta+\beta^2)E(Y^4)\\
& + 2\alpha^3\beta^2(1+\beta)V_3 + 6\alpha^2\beta^2 E(Y^2),
\end{aligned}
$$

where $V_r = E(Y^r\sqrt{\alpha^2Y^2+4})$, for $r = 1,3,5,7$ which are calculated numerically. Furthermore, since $Y|U = u \sim N(\delta u, 1-\delta^2)$, we have

$$
\begin{aligned}
E(Y^2) =& \delta^2 E(U^2) + 1 - \delta^2, \\
E(Y^4) =& \delta^4 E(U^4) + 6\delta^2 E(U^3) + 3(1-\delta^2)^2, \\
E(Y^6) =& \delta^6 E(U^6) + 15\delta^4 E(U^5) + 45\delta^2 E(U^4) + 15(1-\delta^2)^3, \\
E(Y^8) =& \delta^8 E(U^8) + 28\delta^6 E(U^7) + 210\delta^4 E(U^6) + 420\delta^2 E(U^5) + 105(1-\delta^2)^4,
\end{aligned}
$$

where $E(U^r)$ is obtain by $U \sim \text{TN}(0,1;(0,\infty))$, $U \sim E(1)$ and U has PDF (5) for SN-BS, MMNE-BS and MMNEH-BS, respectively.

References

[1] A.C. Aitken, On Bernoulli's numerical solution of algebraic equations, *Proceedings of the Royal Society of Edinburgh*, 46 (1926), 289–305.

[2] H. Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19 (1974), 716–723.

[3] S. Ali, Mixture of the inverse Rayleigh distribution: Properties and estimation in a Bayesian framework. *Applied Mathematical Modelling*, 39 (2015), 515–530.

[4] M. Aslam, and R.R.L. Kantam, Economic reliability acceptance sampling based on truncated life tests in the Birnbaum-Saunders distribution. *Pakistan Journal of Statistics and Operation Research*, 24 (2008), 269-276.

[5] A. Azzalini, A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12 (1985), 171–178.

[6] N. Balakrishnan, R.C. Gupta, D. Kundu, V. Leiva, and A. Sanhueza, On some mixture models based on the Birnbaum-Saunders distribution and associated inference. *Journal of Statistical Planning and Inference*, 141 (2011), 2175–2190.

[7] N. Balakrishnan, V. Leiva, A. Sanhueza, and E. Cabrera, Mixture inverse Gaussian distributions and its transformations, moments and applications. *Statistics: A Journal of Theoretical and Applied Statistics*, 43 (2009), 91–104

[8] K.E. Basford, D.R. Greenway, G.J. McLachlan, and D. Peel, Standard errors of fitted means under normal mixture, *Computational Statistics* , 12 (1997), 1–17.

[9] Y.C. Bechtel, C. Bonaiti-Pellie, N. Poisson, J. Magnette, and P.R. Bechtel, A population and family study of N-acetyltransferase using caffeine urinary metabolites *Clinical pharmacology and therapeutics*, 54 (1993)., 134–141.

[10] Z. W. Birnbaum, S .C.Saunders, A new family of life distributions. *Journal of Applied Probability*, 6 (1969), 319–327.

[11] T. Bučar, M. Nagode, M. Fajdiga, Reliability approximation using finite Weibull mixture distributions, *Reliability Engineering & System Safety*, 84 (2004), 241–251.

[12] A. P. Dempster, N. M. Laird, and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 39 (1977), 1–38.

[13] S. Fruhwirth-Schnatter, *Finite Mixture and Markov Switching Models.* Springer, New York (2006).

[14] I. J.Good, The population frequencies of species and the estimation of population parameters, Biometrika, 40 (1953), 237–260.

[15] F. Hashemi, V. Amirzadeh, and A.Jamalizadeh, An extension of the Birnbaum-Saunders distribution based on Skew-Normal-t distribution. *Journal of Statistical Research of Iran*, 12 (2015), 1–37.

[16] F. Hashemi, M. Naderi, and A. Jamalizadeh, Normal mean-variance Lindley Birnbaum-Saunders distribution. *Statistics and Its Interface*, 12 (2019), 585–597.

[17] L. Hubert, and P. Arabie, Comparing partitions, *Journal of Classification*, 2 (1985), 193–218.

[18] A. Jamalizadeh, F. Hashemi, and M. Naderi, Discussion of "Birnbaum-Saunders distribution: A review of models, analysis, and applications". *Applied Stochastic Models in Business and Industry*, 35 (2019), 82–89.

[19] A. Jamalizadeh, and T.I. Lin, A general class of scale-shape mixtures of skew-normal distributions: properties and estimation. *Computational Statistics*, 32 (2017), 451–474.

[20] N. Jewell, Mixtures of exponential distributions. *The Annals of Statistics*, 10 (1982), 479–484.

[21] R. Jiang, and D.N.P. Murthy, Mixture of Weibull distributions-parametric characterization of failure rate function. *Applied Stochastic Models and Data Analysis*, 14 (1988), 47–65.

[22] M. Khosravi, V. Leiva, A. Jamalizadeh, and E. Porcu, On a nonlinear Birnbaum-Saunders model based on a bivariate construction and its characteristics. *Communications in Statistics-Theory and Methods*, 45 (2015), 772–793.

[23] D. Kundu, N. Balakrishnan, and A. Jamalizadeh, Bivariate Birnbaum-Saunders distribution and associated inference. *Journal of Multivariate Analysis*, 101 (2010), 113–125.

[24] V. Leiva, F. Vilca, N. Balakrishnan, and A. Sanhueza, A skewed sinh-normal distribution and its properties and application to air pollution, *Communications in Statistics - Theory and Methods*, 39 (2010), 426–443.

[25] C. Liu, and D. B. Rubin, The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence, *Biometrika*, 81 (1994), 633–648.

[26] G.J. McLachlan, and K.E. Basford, *Mixture models: Inference and application to clustering*. Dekker, New York (1988).

[27] G.J. McLachlan, and D. Peel, *Finite Mixture Models*. Wiley, New York (2000).

[28] X.L. Meng, and D.B. Rubin, Maximum likelihood estimation via the ECM algorithm: A general framework, *Biometrika*, 80 (1993), 267–278.

[29] K. Mohammadi, O. Alavi, and J.G. McGowan, Use of Birnbaum-Saunders distribution for estimating wind speed and wind power probability distributions: A review. *Energy Conversion and Management*, 143 (2017), 109–122.

[30] B. Muthén, and K. Shedden, Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55 (1999), 463–469.

[31] M. Naderi, A. Arabpour, and A. Jamalizadeh, On the finite mixture modeling via normal mean-variance Birnbaum-Saunders distribution. *JIRSS-Journal of the Iranian Statistical Society*, 16 (2017), 33–51.

[32] M. Naderi, W.L. Hung, T.I. Lin, and A. Jamalizadeh, A novel mixture model using the multivariate normal mean–variance mixture of Birnbaum–Saunders distributions and its application to extrasolar planets. *Journal of Multivariate Analysis*, 171 (2019), 126–138.

[33] M .Naderi, M. Mozafari, Kh. Okhli, Finite mixture modeling via skew-laplace Birnbaum-Saunders distribution. *Journal of Statistical Theory and Applications*, 19 (2020), 49–58.

[34] H.Negarestani, A. Jamalizadeh, S. Shafiei, N. Balakrishnan, Mean mixtures of normal distributions: properties, inference and application, *Metrica*, 82 (2019), 501–528.

[35] H.K.T. Ng, D. Kundu, and N. Balakrishnan, Modified moment estimation for the two parameter Birnbaum-Saunders distribution. *Computational Statistics and Data Analysis*, 43 (2003), 283–298.

[36] A.Punzo, A. Mazza, and A. Maruotti, Fitting insurance and economic data with outliers: a flexible approach based on finite mixtures of contaminated gamma distributions. *Journal of Applied Statistics*, 45 (2018), 2563–2584.

[37] P. Schlattmann, *Medical Applications of Finite Mixture Models*. Springer (2009).

[38] G. Schwarz, Estimating the dimension of a model. *The Annals of Statistics*, 6 (1978), 461–464.

[39] S.Shoham, Robust clustering by deterministic agglomeration EM of mixtures of multivariate t distributions. *Pattern Recognition*, 35 (2002), 1127–1142.

[40] N. Smirnov, Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics*, 19 (1948), 279–281.

[41] D. M. Titterington, A.F. Smith, and U.E. Markov, *Statistical Analysis of Finite Mixture Distributions*, Wiley, New York (1985).

[42] F. Vilca, L. Santana, V. Leiva, and N. Balakrishnan, Estimation of extreme percentiles in Birnbaum–Saunders distributions, *Computational Statistics & Data Analysis*, 55 (2011), 1665–1678

[43] F. Vilca-Labra, and V. Leiva-Sánchez, A New Fatigue Life Model Based on the Family of Skew-Elliptical Distributions, *Communications in Statistics-Theory and Methods*, 35 (2006), 229–244.

[44] W.L. Wang, A. Jamalizadeh, and T.I. Lin, Finite mixtures of multivariate scale-shape mixtures of skew-normal distributions. *Statistical Papers*, 61 (2020), 2643–2670.

[45] M. Wedel, and S.W. Desarbo, Market Segment Derivation and Profiling Via a Finite Mixture Model Framework. *Pattern Recognition*, 13 (2002), 17–25.

[46] M. Wiper, D.R. Insua, and F. Ruggeri, Mixtures of gamma distributions with applications. *Journal of Computational and Graphical Statistics*, 10 (2001), 440–454.

Shahrbanoo Mahbudi
Graduated Ph.D of Statistics
Department of Statistics
Science and Research Branch, Islamic Azad University
Tehran, Iran
E-mail: shahrbanoo_mahbudi@yahoo.com

Ahad Jamalizadeh
Professor of Statistics
Department of Statistics
Shahid Bahonar University of Kerman
Kerman, Iran
E-mail: a.jamalizadeh@uk.ac.ir

Rahman Farnoosh
Professor of Applied Statistics
Department of Applied Mathematics
Iran University of Science and Technology
Tehran, Iran
E-mail: rfarnoosh@iust.ac.ir