A new extended mixture skew normal distribution, with applications

H. M. Barakat[*], **A. W. Aboutahoun**[†] **and N. N. El-kadar**[‡]

Abstract

One of the most important property of the mixture normal distributions-model is its flexibility to accommodate various types of distribution functions (df's). We show that the mixture of the skew normal distribution and its reverse, after adding a location parameter to the skew normal distribution, and adding the same location parameter with different sign to its reverse is a family of df's that contains all the possible types of df's. Besides, it has a very remarkable wide range of the indices of skewness and kurtosis. Computational techniques using EM-type algorithms are employed for iteratively computing maximum likelihood estimates of the model parameters. Moreover, an application with a body mass index real data set is presented.

Key words: Parametric family; mixture distributions; skew normal distribution; skewness; kurtosis.
Mathematics Subject Classification: 62F20; 62E10; 62F99.

1 Introduction

There has been an increased interest in constructing generalized families of distributions by inserting one or more additional shape parameter to the baseline distri-

[*]Department of Mathematics-Faculty of Science-Zagazig University-Zagazig-Egypt.
[†]Faculty of Science, Alexandria University, Alexandria, Egypt
[‡]Current address: Faculty of Science, Alexandria University, Alexandria, Egypt- Permenant address: Department of Statistics Faculty of Science- Benghazi University-Benghazi-Libya.

bution. Usually, the generalized family is more flexible to analyzing data than its baseline distribution provided that the following conditions are fulfilled:

1. Inclusion a large number of the possible nine types $00, 0+$, $0-, +0$, $++, +-$, $-0, -+$, $--$ of df's.

2. Wide range of the indices of skewness and kurtosis,

where $00, 0+, 0-, +0, ++, +-, -0, -+$ and $--$ are the abbreviations of symmetric and mesokurtic, symmetric and leptokurtic, symmetric and platykurtic, positive skewness and mesokurtic and negative skewness and platykurtic.

Nowadays, there are many generalized families of distributions can be found in the literature. Among those famous generalized families is the Marshall and Olkin family, which is obtained by adding a shape parameter, see Marshall and Olkin (1997). This family has the property that the minimum of a geometric number of independent random variables (rv's) with common distribution in the family has a distribution again in the family. For more general results on the Marshall and Olkin family, the reader is referred to Barakat et al. (2009), Jose (2011), Barreto-Souza et al. (2013) and Cordeiro et al. (2014a).

One of the most known distribution families is the exponentiated family, which is obtained by raising the baseline distribution to a positive power. The genesis of the exponentiated family can be traced back to the first half of the nineteenth century, but the characters of this family was studied in detail for the first time by Box and Tiao (1973) (see, also Nadaraja, 2005). For more detail about this family, see Al-Hussaini and Ahsanullah (2015).

Another important generalized family, which is extensively studied by Jones (2004) and it is considered as a generalization of the exponentiated family is the beta-distribution family. This family is obtained by substituting the independent variable in the incomplete beta function by the baseline distribution. For more general results on the beta-distribution family, see Eugene et al. (2002), Mameli and Musio (2013), Lemonte (2014) and Barakat and Nigm (2014).

A more flexible generalized family than the beta-distribution family is the Kumaraswamy family, which was suggested by Cordeiro and Castro (2011). The Kumaraswamy family is built by replacing the independent variable in the Kumaraswamy distribution by any baseline distribution. The Kumaraswamy distribution is relatively much appreciated in comparison to the beta distribution, and has a simple form which can be unimodal, increasing, decreasing or constant, depending on the parameter values. However, the Kumaraswamy family almost has the same mathematical properties of the beta-distribution family except it does not depend on any special function, which making it more tractable than the beta family. For more general results on this family, the reader is referred to Correa et al. (2012), Cordeiro et al. (2014b) and Barakat et al. (2017b).

One of important generalized families is the Azzalini's family, which is built depending on the skew normal distribution, which is proposed by Azzalini (1985). This family was generalized by adding an extra shape parameter, by Arellano-Valle et al. (2004). For more detail about the Azzalini's family, see Correa et al. (2012), Azzalini and Capitanio (2014) and Popović et al. (2017).

Alzaatreh (2011) suggested a generalized distribution family based on the gamma distribution. Moreover, in 2013 Alzaatreh et al. proposed a general way to extend any baseline distribution to a generalized family that contains its baseline distribution.

It is known that the mixture model, which is a convex combination of two probability density functions (pdf's), is a powerful and flexible tool for modeling complex data for being it combines the properties of the individual pdf's. Besides this advantage, the mixture models are frequently used in many applications (cf. Titterington et al., 1986). In 2015 Barakat suggested a generalized distribution family, which is simpler than the most well-known families for being it is a mixture of a baseline distribution and its reverse (i.e., the distribution of the negative rv), after adding a location parameter to the baseline distribution, and adding the same location parameter with different sign to the reverse of the baseline distribution. By this way the utilized location parameter turns to a shape parameter and we get a two-parameter generalized distribution family. This family has an individual trait, on which it has been built, that if this family contained any distribution, it should contain also its reverse distribution. This trait makes us to name this family, the stable symmetric family.

Barakat (2015) suggested a new generalized family, denoted by ASSN, by taking the standard normal distribution as the baseline distribution of the stable symmetric family. Barakat (2015) showed that the ASSN (additive stable-symmetric normal) family contains all the possible types of df's, except the type 0+, besides it has very remarkable wide range of the indices of skewness and kurtosis. Therefore, it is capable of describing many types of statistical data than many other known families.

Barakat and Khaled (2017) called any generalized family that contains all the possible types of df's (nine types), a full family. Moreover, Barakat and Khaled (2017) suggested two full families, via the mixture of the ASSN family and the standard logistic or Laplace distributions (note that both of these df's are symmetric leptokurtic, i.e., have the type 0+). Barakat et al. (2017a) suggested a more tractable full family with three parameters, denoted MSSN (multiplicative stable-symmetric normal) family via the mixture of the normal distribution $(\nu, 1), \nu \neq 0$ and its reverse, after multiplying and dividing, respectively, to them the same scale parameter. Barakat et al. (2017a) showed that via the theoretical and practical studies the family MSSN is more capable of fitting different types of data than the ASSN family.

In this paper, we suggest another full family with three parameters, denoted ASSA (additive stable-symmetric Azzalini), by using the skew normal distribution instead of the normal distribution $(\nu, 1), \nu \neq 0$, in the definition of ASSN family. We show that the ASSA family is full and it has a very remarkable wide range of the indices of skewness and kurtosis. Moreover, we show that it outperforms both the ASSN and MSSN families.

2 The New Extended Azzalini Distribution-ASSA Family

Azzalini (1985) introduced a skew-normal distribution by adding a shape parameter to the normal df. namely, the pdf of the skew-normal distribution is given by

$$\mathcal{A}(z; \lambda) = 2\phi(z)\Phi(\lambda z), \quad -\infty < \lambda < \infty, -\infty < z < \infty,$$

where ϕ and Φ are the standard normal pdf and df, respectively. The df of the skew-normal distribution is given by

$$A(z;\lambda) = 2\int_{-\infty}^{z}\int_{-\infty}^{\lambda t}\phi(t)\phi(u)dudt.$$

Clearly $\mathcal{A}(x;0) = \phi(x)$. Moreover, $\overline{A}(-x;\lambda) = 1 - A(-x;\lambda) = A(x;-\lambda)$, $A(x;1) = \Phi^2(x)$ (which is the df of the maximum of i.i.d two rv's from Φ) and if $W \sim A(x;\lambda)$, then $W^2 = Z^2$, i.e., $|W| \sim |Z|$, where $Z \sim \Phi(x)$. This implies that the even moments of η and Z are identical. Also, this implies the existence of the odd moments of W. Moreover, Azzalini (1985) derived the moment generating function of W, which is given by

$$M_W(t) = 2\exp(\frac{t^2}{2})\Phi(\delta t),\ \delta = \frac{\lambda}{\sqrt{1+\lambda^2}}. \tag{2.1}$$

Using (2.1), we can easily derive the mean, variance, the third and the fourth central moments, as

$$\left.\begin{array}{l}
\mathrm{E}(W) := \mu_{W} = b\delta, b = \sqrt{\frac{2}{\pi}},\\
\mathrm{E}(W-\mu_{W})^2 = \sigma^2_{W} := C_W^{[2]} = 1-(b\delta)^2,\\
\mathrm{E}(W-\mu_{W})^3 := C_W^{[3]} = (2b^2-1)b\delta^3,\\
\mathrm{E}(W-\mu_{W})^4 := C_W^{[4]} = 3+(4\delta^2-3b^2\delta^2-6)b^2\delta^2.
\end{array}\right\} \tag{2.2}$$

For any $0 \leq \bar{\alpha} = 1-\alpha \leq 1$ and $-\infty < c < \infty$, the suggested new family, denoted by ASSA family, is defined via the mixture of the df $A(x;\lambda)$ and its reverse $\overline{A}(-x;\lambda) = A(x;-\lambda)$, by

$$\begin{aligned}
G(y;\alpha,c,\lambda) &= \alpha A(y+c,\lambda) + \bar{\alpha}\bar{A}(-y+c,\lambda)\\
&= \alpha A(y+c,\lambda) + \bar{\alpha}A(y-c,-\lambda). \qquad (2.3)
\end{aligned}$$

In the sequel, we attach the df G with its rv (say Y) and write $G_Y(y;\alpha,c,\lambda)$ instead of $G(y;\alpha,c,\lambda)$ (especially, when we talk about the moments of $G(y;\alpha,c,\lambda)$). The pdf and the survival function of $G_Y(y;\alpha,c,\lambda)$ are given by

$$g_Y(y;\alpha,c,\lambda) = \alpha\mathcal{A}(y+c,\lambda) + \bar{\alpha}\mathcal{A}(y-c,-\lambda) \tag{2.4}$$

and

$$\begin{aligned}
\bar{G}_Y(y;\alpha,c,\lambda) &= 1 - G_Y(y;\alpha,c,\lambda) = \alpha + \bar{\alpha} - \alpha A(y+c,\lambda) - \bar{\alpha}\bar{A}(-y+c,\lambda)\\
&= \alpha\bar{A}(y+c,\lambda) + \bar{\alpha}A(-y+c,\lambda)\\
&= \alpha\bar{A}(y+c,\lambda) + \bar{\alpha}\bar{A}(y-c,-\lambda).
\end{aligned}$$

The following lemma gives the mean, the variance, the coefficient of skewness and the coefficient of kurtosis of the ASSA family. It is well-known that the aforementioned statistics uniquely determine the vast majority of the known df's.

Lemma 2.1. *For the ASSA family, we have*

1. *The mean is* $\mu_{_Y} = (2\alpha - 1)(b\delta - c)$.

2. *The variance is* $\sigma^2_{_Y} = 1 + 4(b\delta - c)^2\alpha\bar{\alpha} - (b\delta)^2$.

3. *The coefficient of skewness is*

$$\gamma^{[1]}_{_Y} = \frac{C^{[3]}_Y}{\sigma^3_{_Y}} = \frac{(2\alpha - 1)[(2b^2 - 1)b\delta^3 - 8(b\delta - c)^3\alpha\bar{\alpha}]}{(1 + 4(b\delta - c)^2\alpha\bar{\alpha} - (b\delta)^2)^{\frac{3}{2}}}.$$

4. *The coefficient of kurtosis is*

$$\gamma^{[2]}_{_Y} = \frac{C^{[4]}_Y}{\sigma^4_{_Y}} = \frac{3 + (4\delta^2 - 3b^2\delta^2 - 6)b^2\delta^2}{(1 + 4(b\delta - c)^2\alpha\bar{\alpha} - (b\delta)^2)^2}$$

$$+\frac{4(b\delta - c)\alpha\bar{\alpha}[4(b\delta - c)^3(\bar{\alpha}^3 + \alpha^3) + 6(1 - (b\delta)^2)(b\delta - c) + 4(2b^2 - 1)b\delta^3]}{(1 + 4(b\delta - c)^2\alpha\bar{\alpha} - (b\delta)^2)^2}.$$

Proof. The proof follows directly by using the relation (2.2) and applying Theorem 2.1 in Barakat (2015), after routine calculations. □

Remark. It is easy to check that, when $\lambda = 0$ all the statistics $\mu_{_Y}, \sigma^2_{_Y}, \gamma^{[1]}_{_Y}$ and $\gamma^{[2]}_{_Y}$ are equal to those corresponding statistics pertaining the ASSN family.

Lemma 2.2. *For the ASSA family, the coefficients of skewness and kurtosis satisfy the following implications:*

$$\gamma^{[1]}_{_Y}\begin{cases} = 0, \quad if\ \alpha = \frac{1}{2} \pm \sqrt{\frac{1}{4} - \frac{(2b^2-1)b\delta^3}{8(b\delta-c)^3}},\ (2b^2 - 1)b\delta^3 \le 2(b\delta - c)^3; \\ \\ > 0, \quad if \begin{cases} \alpha = 1\ ,\ \lambda \ge 1 \quad or \\ c > \xi(\alpha, \bar{\alpha}, b, \delta),\ 1 > \alpha > \frac{1}{2}, \quad or \\ c < \xi(\alpha, \bar{\alpha}, b, \delta),\ 1 > \alpha > 0; \end{cases} \\ \\ < 0, \quad if \begin{cases} \alpha = 0\ ,\ \lambda \ge 1 \quad or \\ c < \xi(\alpha, \bar{\alpha}, b, \delta),\ 1 > \alpha > \frac{1}{2}, \quad or \\ c > \xi(\alpha, \bar{\alpha}, b, \delta),\ \frac{1}{2} > \alpha > 0. \end{cases} \end{cases}$$

and

$$\gamma_Y^{[2]}\begin{cases} =3, \quad if \begin{cases} \lambda=0\ ,\alpha>\frac{1}{2}\ ,c>0,\ or\ \lambda=0,\ \alpha<\frac{1}{2}\ ,c<0,\ or \\ \lambda=0\ ,\alpha>\frac{1}{2}\ ,c<0,\ or\ \lambda=0,\ \alpha<\frac{1}{2}\ ,c>0,\ or \\ \lambda=0,\ \alpha=1,\ or \lambda=0,\ \alpha=0\ ,or\ \lambda=1,\ \alpha=\frac{1}{2},\ c=0,\ \ or \\ \alpha=\pm\sqrt{\frac{1}{4}-\bar{Z}}+\frac{1}{2},\ 1\geq 4\bar{Z},\ 0<\alpha<1, or \\ \alpha=\pm\sqrt{\frac{1}{4}-\underline{Z}}+\frac{1}{2},\ 1\geq 4\underline{Z},\ 0<\alpha<1; \end{cases} \\ <3, \quad if \begin{cases} \alpha\in\left(\sqrt{\frac{1}{4}-\bar{Z}}+\frac{1}{2},\ \sqrt{\frac{1}{4}-\underline{Z}}+\frac{1}{2}\right),\frac{1}{4}>\bar{Z}>\underline{Z},\alpha>\frac{1}{2},\ or \\ \alpha\in\left(\frac{1}{2}-\sqrt{\frac{1}{4}-\underline{Z}},\ \frac{1}{2}-\sqrt{\frac{1}{4}-\bar{Z}}\right),\frac{1}{4}>\bar{Z}>\underline{Z},0<\alpha<\frac{1}{2}; \end{cases} \\ >3, \quad if \begin{cases} \delta\neq 0,\ \alpha=0,\ or\ \delta\neq 0,\ \alpha=1,\ or \\ \alpha\notin\left(\sqrt{\frac{1}{4}-\bar{Z}}+\frac{1}{2}\ ,\sqrt{\frac{1}{4}-\underline{Z}}+\frac{1}{2}\right),\frac{1}{4}>\bar{Z}>\underline{Z},1>\alpha>\frac{1}{2},\ or \\ \alpha\notin\left(\frac{1}{2}-\sqrt{\frac{1}{4}-\underline{Z}},\ \frac{1}{2}-\sqrt{\frac{1}{4}-\bar{Z}}\right),\frac{1}{4}>\bar{Z}>\underline{Z},0<\alpha<\frac{1}{2}; \end{cases} \end{cases}$$

where $\xi(\alpha,\bar{\alpha},b,\delta)=(b-\frac{1}{2}\sqrt[3]{\frac{2b^2-1}{\alpha\bar{\alpha}}})\delta$,

$$\bar{Z}=\frac{((b\delta-c)^3+(2b^2-1)b\delta^3)+\sqrt{((b\delta-c)^3+(2b^2-1)b\delta^3)^2-3(b\delta-c)^2(3b^2-2)b^2\delta^4}}{12(b\delta-c)^3}$$

and

$$\underline{Z}=\frac{((b\delta-c)^3+(2b^2-1)b\delta^3)-\sqrt{((b\delta-c)^3+(2b^2-1)b\delta^3)^2-3(b\delta-c)^2(3b^2-2)b^2\delta^4}}{12(b\delta-c)^3}.$$

Moreover, in the above implications, concerning $\gamma_Y^{[2]}$*, if the condition* $\frac{1}{4}>\bar{Z}>\underline{Z}$ *is changed to* $\frac{1}{4}>\underline{Z}>\bar{Z}$*, we will obtain the corresponding implications by changing every* $\underline{Z}$ *to* $\bar{Z}$ *and vice versa.*

Proof. First, consider the assumption $\gamma_Y^{[1]}=0$, which implies (by using Lemma 2.1) the equation $(2\alpha-1)[(2b^2-1)b\delta^3-8(b\delta-c)^3\alpha\bar{\alpha}=0$. It is easily checking that the last equation has the distinct roots $\alpha=\frac{1}{2}$ and $\alpha=\frac{1}{2}\pm\sqrt{\frac{1}{4}-\frac{(2b^2-1)b\delta^3}{8(b\delta-c)^3}}$, $(2b^2-1)b\delta^3<2(b\delta-c)^3$. Clearly, the aforementioned roots can be uniquely described by the relation $\alpha=\frac{1}{2}\pm\sqrt{\frac{1}{4}-\frac{(2b^2-1)b\delta^3}{8(b\delta-c)^3}}$, $(2b^2-1)b\delta^3\leq 2(b\delta-c)^3$. The proof of the remaining cases for $\gamma_Y^{[1]}$ are obvious. Finally, all the cases concerning $\alpha=0$ or $\alpha=1$, convert the family (2.3) to the skew normal distribution due to Azzalini (1985) or its reverse, respectively.Therefore, all the given implications, which pertain to these cases can directly be proved from the well know properties of the skew normal distribution. On the other hand, for proving the relations satisfied by $\gamma_Y^{[2]} = 3$, consider the equation $\gamma_Y^{[2]} = 3$. By using Lemma 2.1 and after some algebra, we get

$$96(b\delta - c)^4(\alpha\bar{\alpha})^2 - 16(b\delta - c)[(b\delta - c)^3 + (2b^2 - 1)b\delta^3]\alpha\bar{\alpha} + 2(3b^2 - 2)b^2\delta^4 = 0.$$

It is easily verifying that the preceding quadratic equation in $\alpha\bar{\alpha}$ has the following roots

$$\alpha\bar{\alpha} = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A},$$

where $A = 96(b\delta - c)^4$, $B = -16(b\delta - c)((b\delta - c)^3 + (2b^2 - 1)b\delta^3)$ and $C = 2(3b^2 - 2)b^2\delta^4$. Therefore, we have the two roots $\alpha\bar{\alpha} = \frac{1}{4} - (\alpha - \frac{1}{2})^2 = \bar{Z}$ and $\alpha\bar{\alpha} = \frac{1}{4} - (\alpha - \frac{1}{2})^2 = \underline{Z}$. This in turns implies that $\gamma_Y^{[2]} = 3$, if $\alpha = \pm\sqrt{\frac{1}{4} - \bar{Z}} + \frac{1}{2}$, $\alpha = \pm\sqrt{\frac{1}{4} - \underline{Z}} + \frac{1}{2}$, provided that $0 \leq \alpha \leq 1$. Bearing in mind, when $\alpha > \frac{1}{2}$, the only two roots $\alpha = \sqrt{\frac{1}{4} - \bar{Z}} + \frac{1}{2}$ and $\alpha = \sqrt{\frac{1}{4} - \underline{Z}} + \frac{1}{2}$ can be obtained, while when $\alpha < \frac{1}{2}$, the other two roots can be obtained, the proof of the remaining cases for $\gamma_Y^{[2]}$, which are concerning to these roots, are obvious. Finally, all the cases concerning $\lambda = 0$, convert the family (2.3) to the SSN family. Therefore, all the given implications, which pertain to these cases can directly be proved from the results of Barakat (2015). $\square$

Theorem 2.1. *The family ASSA is full. Moreover, it has a very remarkable wide range of the indices of skewness and kurtosis, comparing with the skew normal distribution.*

Proof. The proof is directly follows from Table 1, in which we selected some values of $\gamma_Y^{[1]}$ and $\gamma_Y^{[2]}$ that covered all possible types of df's. The selections of the different values of the parameters α, c and λ were done in the light of Lemma 2.2, while the computing the corresponding values of $\gamma_Y^{[1]}$ and $\gamma_Y^{[2]}$ were done in the light of Lemma 2.1. Moreover, we constructed a simple code by Matlab 8.2 on laptop Intel 1.8GHZ processor.

Table 1: Some selected values of $\gamma_Y^{[1]}$ and $\gamma_Y^{[2]}$ that covered all possible types of df's

Types	$0 \le \alpha \le 1$	$-\infty < c < \infty$	$-\infty < \lambda < \infty$	$\gamma_Y^{[1]}$	$\gamma_Y^{[2]}$	σ_Y^2	μ_Y
00	0	-5	0	0	3	1	-5
00	0.5	0	1	0	3	1	0
0+	0.5	0.1	1	0	3.0981	0.8972	0
0+	0.5	0.3	2	0	3.6233	0.6618	0
0−	0.5	-5	1	0	1.0870	31.6419	0
0−	0.5	1	1	0	2.7659	0.8716	0
+0	0.211324865	-2	0	0.8771	3	3.6667	-1.1547
+0	0.788675134	2	0	0.8771	3	3.6667	-1.1547
++	0.1	-5	1	2.4380	7.5435	11.8273	-4.4513
++	0.6	0.5	1	0.027	3.1014	0.6856	0.0128
+−	0.4	-5	1	0.3945	1.2498	30.4035	-1.1128
+−	0.6	1	2	0.0936	2.6625	0.5694	-0.0573
−0	0.788675134	-2	0	-0.8771	3	3.6667	1.1547
−0	0.211324865	2	0	-0.8771	3	3.6667	1.1547
−+	0.9	-4	1	-2.3378	7.3028	8.1811	3.6514
−+	0.4	0.5	1	-0.027	3.1014	0.6856	-0.0128
−−	0.6	-5	1	-0.3945	1.2498	30.4035	1.1128
−−	0.3	1	1	-0.1121	2.8365	0.8412	0.1743

Table 1 does not only show that the ASSA family contains all the possible types of df's, i.e it is a full family, but it possesses very remarkable wide range of the indices of skewness and kurtosis. For example, $\gamma_Y^{[2]}$ reached to 7.5435, while the maximum value of the coefficient of kurtosis of Azzalini's family is 3.869. Actually, since the ASSA family is belonging to the family ASSN, Table 2 in Barakat (2015) shows that $\gamma_Y^{[2]}$ reached to the 13.585, when $c = \pm 10$, $\alpha = 0.05$ and $\lambda = 0$. On the other hand, $-2.3378 \le \gamma_Y^{[1]} \le 2.4380$, while for the skew normal distribution, we have $-0.995 \le$ the coefficient of skewness ≤ 0.995. Above all this, due to the results of Barakat (2015), the value of both skewness and kurtosis of the ASSA family may increase infinitely with increasing the value of some parameters. Finally, since for the skew normal distribution, we have $3 \le$ the coefficient of kurtosis ≤ 3.869 (cf., Azzalini, 1985), then the skew normal distribution contains at most $\{00, 0+, +0, ++, -0, -+\}$ types of df's. □

We include the graph of the pdf, defined by (2.4), of the ASSA family for some selected values of $\gamma_Y^{[1]}$ and $\gamma_Y^{[2]}$ that covered all possible types of df's. These graphs (Figures 1-3) shows that this family is extremely rich and prolific.

Figure 1: ASSA family

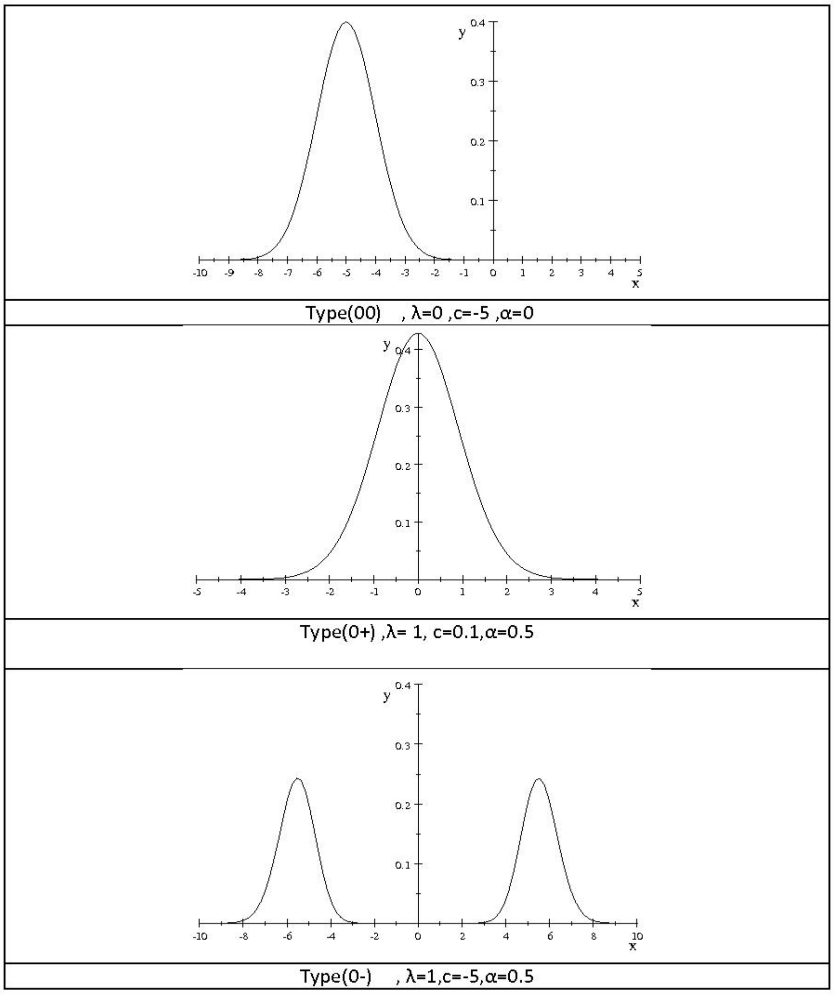


Figure 2: ASSA family

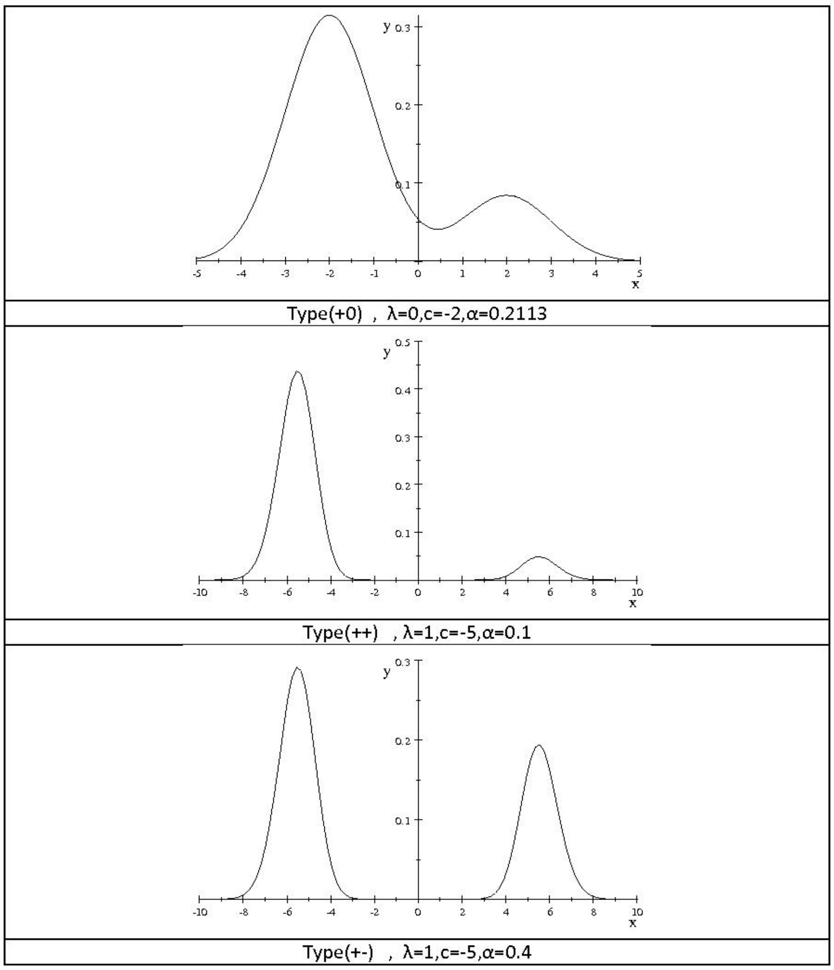
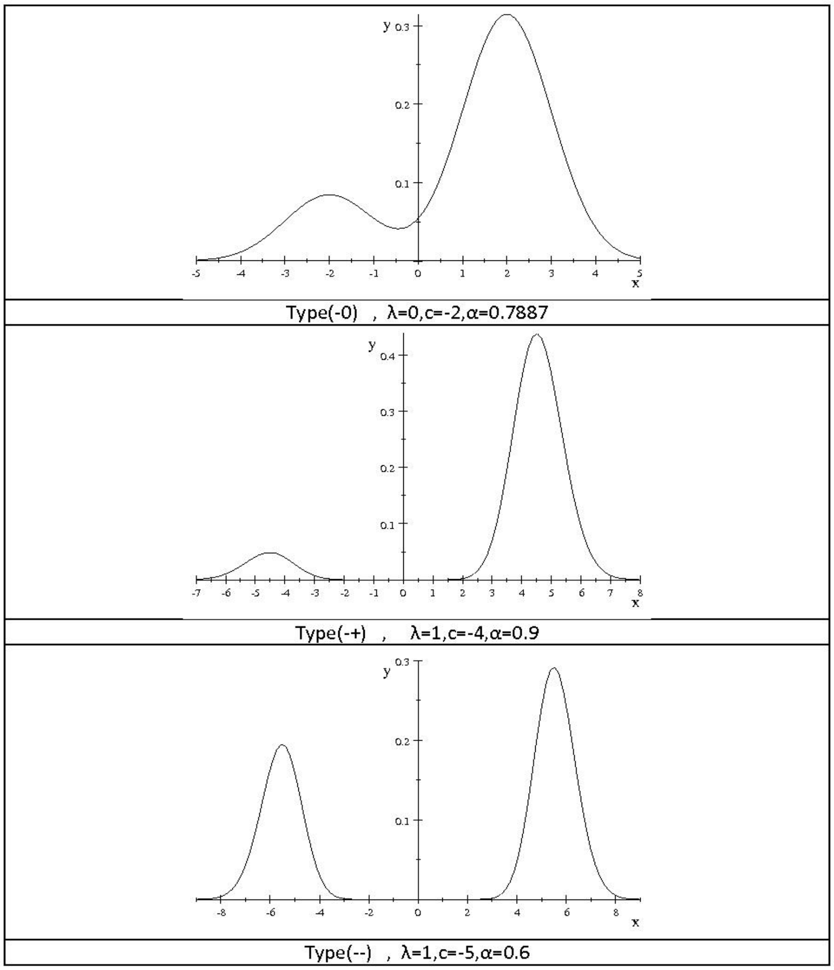


Figure 3: ASSA family



Study of the effect of the parameter λ on $\gamma_Y^{[1]}$ and $\gamma_Y^{[2]}$

For studying the effect of the shape parameter λ on the skewness and the kurtosis of the ASSA family in each of the possible types of df's, we fix the two other parameters α, c and check the value of $\gamma_Y^{[1]}$ and $\gamma_Y^{[2]}$, when the parameter λ varies. Of course all the types, for which the parameter λ should be constant are excluded from this study. Namely, in the type $+0$, we have $\lambda = 0, \alpha > \frac{1}{2}, c > 0$, or $\lambda = 0, \alpha < \frac{1}{2}, c < 0$ (cf. Barakat, 2015). Also, the type -0 is obtained, when $\lambda = 0, \alpha > \frac{1}{2}, c < 0$, or $\lambda = 0, \alpha < \frac{1}{2}, c > 0$ (cf. Barakat, 2015). Besides the two types $+0$ and -0, the type 00 should be excluded, since both the skewness and the kurtosis have fixed values 0 and 3, respectively. This study is summarized in Table 2. Table 2 reveals a very interesting property of the ASSA family, that both values of the skewness and the kurtosis increase (or decrease) with increasing λ if the sign of the corresponding type is + (or −). This property is remarkable useful in maneuvering the skewness and kurtosis via the parameter λ.

Table 2: The effect of the parameter λ on $\gamma_Y^{[1]}$ and $\gamma_Y^{[2]}$

Types	α	c	λ	$\gamma_Y^{[1]}$	$\gamma_Y^{[2]}$	Types	α	c	λ	$\gamma_Y^{[1]}$	$\gamma_Y^{[2]}$
0+	0.5	0.3	1	0	3.1778	+−	0.6	3	1	0.3455	1.5206
0+	0.5	0.3	10	0	4.1096	+−	0.6	3	10	0.3680	1.3585
0+	0.5	0.3	20	0	4.1325	+−	0.6	3	20	0.3684	1.3551
0+	0.5	0.3	50	0	4.1390	+−	0.6	3	50	0.3686	1.3541
0+	0.5	0.3	100	0	4.1399	+−	0.6	3	100	0.3686	1.3540
0−	0.5	5	1	0	1.1285	−+	0.4	0.5	1	-0.0270	3.1014
0−	0.5	5	10	0	1.0703	−+	0.4	0.5	10	-0.1090	4.6710
0−	0.5	5	20	0	1.0693	−+	0.4	0.5	20	-0.1108	4.7251
0−	0.5	5	50	0	1.0690	−+	0.4	0.5	50	-0.1113	4.7407
0−	0.5	5	100	0	1.0689	−+	0.4	0.5	100	-0.1113	4.7429
++	0.6	0.5	1	0.0270	3.1014	−−	0.4	3	1	-0.3455	1.5206
++	0.6	0.5	10	0.1090	4.6710	−−	0.4	3	10	-0.3680	1.3585
++	0.6	0.5	20	0.1108	4.7251	−−	0.4	3	20	-0.3684	1.3551
++	0.6	0.5	50	0.1113	4.7407	−−	0.4	3	50	-0.3686	1.3541
++	0.6	0.5	100	0.1113	4.7429	−−	0.4	3	100	-0.3686	1.3540

3 Statistical Inference for the Parameters of the ASSA Family

in this section, we implement a simulation study with different sample sizes and different parameter values of the location-scale ASSA family (the standard ASSA family is defined by (2.3)). Namely,

$$G^{\langle ASSA\rangle}(z;\alpha,c,\lambda,\mu,\sigma)=\alpha A(\frac{z-c_1}{\sigma},\lambda)+\bar{\alpha}A(\frac{z-c_2}{\sigma},-\lambda), \tag{3.1}$$

where

$$\left.\begin{aligned} &c_1=\mu-\sigma c,\\ &c_2=\mu+\sigma c, \sigma>0, 0\leqslant\alpha\leqslant 1.\end{aligned}\right\} \tag{3.2}$$

Our focus in this study will be on the main two shape parameters c and λ. We compute the maximum likelihood estimates (MLE) of the parameters by using the simulated samples and the EM algorithm. The EM algorithm is an iterative method for approximating the maximum of a likelihood function. In particular, the EM algorithm is applied when we have unobserved latent variables. A typical example is that the mixture of df's (as in this paper), for more details see McLachlan and Peel (2000). It is worth mentioning that the EM algorithm is guaranteed to monotonically converge to local optima under mild continuity conditions (cf. Wu, 1983).

In our study, we were careful that the study covers all the possible types of df's. Actually, we select the second case of each type in Table 1 (with given values of α, c, λ). Moreover, for each case, we choose appropriate values of the two locations parameters c_1 and c_2, in the model (3.1) (e.g, in the second case of the first type we should have $-\infty<c_1=c_2<\infty$). Finally, by using the two relations defined by (3.2), we compute the values of the remaining parameters μ and σ. On the other hand, we could now generate three groups of random samples each of size 100, 1000 and 2000 from the family (3.1). Furthermore, we repeat this simulation 10 time for each group. For every replication in each group, we compute the MLE of the parameters $c_1, c_2, \sigma, \lambda$ and again by using the two relation defined by (3.2), we compute the estimate of c. Moreover, we compute the average estimate values for each parameter and group (each average estimate is computed for the 10 replications of each group). Finally, we attache these averages (as the estimates of the parameters) with their mean square errors (MSE). All the computation were implemented by using the package mixsmsn in the R package. The summary of this study is shown in Table 3, where the MLE estimate (average estimate) of the parameters c, σ and λ are denoted by $\hat{c}$, $\hat{\sigma}$ and $\hat{\lambda}$.

The results in Table 3 show that with increasing the sample size the estimated values become more close to the true values. Moreover, based on the MSE, the three best estimates (in order) are obtained for the types $0-, ++$ and $--$, while the two worst estimates are appeared in the types $+0$ and -0.

Table 3: Simulation study

Types	n	c	$\hat{c}$	$mse(\hat{c})$	σ	$\hat{\sigma}$	$mse(\hat{\sigma})$	λ	$\hat{\lambda}$	$mse(\hat{\lambda})$
00	100	0	0.4715	0.7434433	1	0.7145	0.4328897	1	1.0788	0.7248256
00	1000	0	0.0503	0.06133107	1	0.9837	0.1192741	1	1.3909	0.4472703
00	2000	0	0.0417	0.05686036	1	0.9819	0.07319085	1	1.3523	0.3812229
0+	100	0.3	0.51764	0.5318964	5	5.258	2.623392	2	2.7411	2.002503
0+	1000	0.3	0.34517	0.0 7444808	5	4.6721	0.6313025	2	1.6669	0.504511
0+	2000	0.3	0.39861	0.13619725	5	4.5561	0.6060958	2	1.6208	0.4800517
0−	100	1	1.3401	0.7253889	1	1.098	0.5807814	1	1.8068	1.879988
0−	1000	1	0.9881	0.1833502	1	0.9987	0.1135099	1	0.9929	0.1493362
0−	2000	1	0.964	0.06959023	1	1.0451	0.08126561	1	0.9366	0.0811024
+0	100	2	1.2456	0.8597237	2	3.4741	1.895237	0	3.666	6.881123
+0	1000	2	1.3874	0.6755031	2	2.8791	0.9088954	0	0.9634	0.9661656
+0	2000	2	1.2784	0.7412305	2	2.853	0.8617856	0	1.1332	1.286403
++	100	0.5	0.6373	0.3643604	3	2.8175	1.018498	1	1.433	1.331798
++	1000	0.5	0.5104444	0.07299467	3	3.002889	0.3760384	1	0.9428889	0.08778129
++	2000	0.5	0.4919	0.03552323	3	3.0947	0.2041228	1	0.985	0.04049938
+−	100	1	1.31442	0.6374752	2	1.8064	0.3730177	2	1.8888	0.53229
+−	1000	1	0.1156	0.1664039	2	1.8561	0.2221389	2	1.8211	0.3490534
+−	2000	1	1.06855	0.1556551	2	1.9413	0.2244756	2	1.8937	0.3476488
−0	100	2	1.6484	1.03394	2	3.0432	1.725001	0	3.3027	6.440269
−0	1000	2	1.2957	0.7397465	2	3.0409	1.077557	0	1.0233	1.037666
−0	2000	2	1.3482	0.7096037	2	2.9046	0.9530054	0	0.9535	0.9587976
−+	100	0.5	0.6841	0.4786648	2	2.1959	1.02785	1	1.0145	0.5422168
−+	1000	0.5	0.43637	0.08594303	2	2.2127	0.2975623	1	0.9646	0.09372086
−+	2000	0.5	0.51521	0.07344622	2	1.9922	0.1913573	1	0.9319	0.1082603
−−	100	1	1.03821	0.3648316	2	2.3823	0.8808493	1	2.118	2.073973
−−	1000	1	1.04436	0.1635625	2	2.0302	0.1924734	1	1.0228	0.1343868
−−	2000	1	1.04514	0.1330574	2	1.9822	0.05387021	1	0.9504	0.07408509

4 Application

We have seen in Section 3, that at least theoretically that the proposed ASSA family is capable of fitting a wide spectrum of real world data set for being it contains all the possible types of data (i.e., it is a full family). Moreover, it possesses a very wide range of the indices of skewness and kurtosis. In this section we will confirm its outperforming than the two competitors families ASSN and MSSN via an example of a real data. The standard ASSN and MSSN families are respectively defined by $G^{+}(x;\alpha,c) = \alpha\Phi(x+c) + \overline{\alpha}\Phi(x-c)$ (cf. Barakat, 2015) and $G^{\star}(x;\alpha,c,\nu) = \alpha\Phi(\frac{x}{c}-\nu) + \bar{\alpha}\Phi(cx+\nu),\ \nu \neq 0$ (cf. Barakat, 2017a).

Example 4.1. In this example, we use the location-scale ASSA, ASSN and ASSM families, which are respectively defined by (3.1),

$$G^{\langle ASSN\rangle}(x;\alpha,c,\mu,\sigma) = \alpha\Phi(\frac{x-c_1}{\sigma_1}) + \bar{\alpha}\Phi(\frac{x-c_2}{\sigma_2}), \tag{4.1}$$

where $\sigma_1 = \sigma_2 = \sigma$, $c_1 = \mu - c$, $c_2 = \mu + c$ $0 \leqslant \alpha \leqslant 1$, and

$$G^{\langle MSSN\rangle}(x;\alpha,c,\nu,\mu) = \alpha\Phi(\frac{x-c_1}{\sigma_1}) + \bar{\alpha}\Phi(\frac{x-c_2}{\sigma_2}), \tag{4.2}$$

where $c = \sigma_1 = \frac{1}{\sigma_2}$, $c_1 = \mu + c\nu$, $c_2 = \mu - \frac{\nu}{c}$, $0 \leqslant \alpha \leqslant 1$.

We fit the families (3.1), (4.1) and (4.2) to the body mass index data “BMI ”, which is incorporated in the package mixsmsn in the R-Package. The body mass index data set was collected for men aged between 18 to 80 years old. It came from the national health and nutrition examination survey, which was made by the National Center for Health Statistics (NCHS) of the Center for Disease Control (CDC) in the USA. Actually, with the increase of chronic diseases around the USA, attention was attracted to the obesity problem in the past few years. It is known that people with obesity have higher chances of developing chronic diseases. To quantify overweight and obesity, the BMI, which is defined as the ratio of body weight in kilograms and body height in squared meters, was selected as the standard measure, where people with high BMI (> 25) are considered to have overweight and people with BMI > 30 are considered to be obese. The summary statistics for this data set is given in Table 4.

We compared the performances of the three families defined in (3.1), (4.1) and (4.2) in fitting the given data sets by using Akaike information criterion (AIC)

(Akaike, 1973). This criterion is based on the likelihood value of the model, the number of observations and the number of parameters thereof. The computations of the estimates of the parameters and the values of AIC criterion were carried out by using the mix.print() function in the mixsmsn Package in the R-Package (see Table 5). Table 5 shows that the ASSA family has the best performance.

Table 4: summary statistics for the BMI data set

minimum	maximum	median	mean	variance	skewness	kurtosis
14.86	64.16	26.89	28.19	56.2279	0.7135727	3.294964

Table 5: Fitting compare between the ASSN , MSSN and ASSA families

Family	Parameter estimations	log-likelihood estimate	AIC
ASSN	$\hat{c}_1 = 32.663$		
	$\hat{c}_2 = 21.429$	-9771.748	19551.5
	$\hat{\sigma} = 40.39$		
MSSN	$\hat{c}_1 = 32.663$		
	$\hat{c}_2 = 21.429$	-7659.919	15329.84
	$\hat{\sigma}_1 = 40.39 = \frac{1}{\hat{\sigma}_2}$		
ASSA	$\hat{c}_1 = 20.136$		
	$\hat{c}_2 = 28.489$	-7206.371	14422.74
	$\hat{\sigma} = 8.418$		
	$\hat{\lambda} = 1.070$		

References

[1] Al-Hussaini, E. K. and Ahsanullah, M. (2015). Exponentiated distributions. Atlantis Press, Paris. France.

[2] Akaike, H. (1973). Information theory and the maximum likelihood principle. In: B. N. Petrov and F. Csaki (eds.), Second International Symposium on Information Theory, Akademiai Kiado, Budapest: 267-281.

[3] Alzaatreh, A. (2011). A new method for generating families of continuous distributions. Ph.D. thesis, Central Michigan University Mount Pleasant, Michigan.

[4] Alzaatreh, A., Famoye, F. and Lee, C. (2013). A new method for generating families of continuous distributions. Metron, 71: 63-79.

[5] Arellano-Valle, R. B., Gomez, H. W. and Quintana, F. A. (2004). A new class of skew-normal distribution. Commun. Statist.- Theory Meth., 33(7): 1465-1480.

[6] Azzalini, A. (1985). A class of distributions which includes the normal ones. Scand. J. Statist., 12: 171-178.

[7] Azzalini, A. and Capitanio, A. (2014). The skew-normal and related families. New York: Cambridge University Press.

[8] Barakat, H. M. (2015). A new method for adding two parameters to a family of distributions with application to the normal and exponential families. Statist. Meth. & Appl. (SMA), 24(3): 359-372.

[9] Barakat, H. M. and Khaled, O. M. (2017). Towards the establishment of a family of distributions that best fits any data set. Commun. Statist.-Comput. Simul., Doi:org/10.1080/03610918.2016.1197245.

[10] Barakat , H. M. and Nigm, E. M. (2014). Asymptotic distributions of order statistics and record values arising from the class of beta-generated distributions. Statistics, 48(5), 1005-1012.

[11] Barakat, H. M., Aboutahounand, A. W. and El-kadar, N. N. (2017a). On some generalized families arising from mixture normal distribution with applications, submitted.

[12] Barakat, H. M., Ghitany, M. E. and Al-Hussaini, E. K. (2009). Asymptotic distributions of order statistics and record values under the Marshall-Olkin parameterization operation. Commun. Statist. Theory Meth., 38: 2267-2273.

[13] Barakat, H. M., Nigm, E. M. and Harpy, M. H. (2017b). Limit theorems of order statistics and record values from the gamma and Kumaraswamy-generated-distributions. Bull. Malaysian Math. Soc. (second series), Doi:10.1007/s40840-016-0356-9.

[14] Barreto-Souza, W., Lemonte, A.J. and Cordeiro, G. M. (2013). General results for Marshall and Olkin's family of distributions. An Acad. Bras. Cienc., 85: 3-21.

[15] Box, G. E. P. and Tiao, G. C. (1973). Bayesian Inference in Statistical Analysis. Addision-Wesley, Reading, MA.

[16] Cordeiro, G. M. and Castro, M. (2011). A new family of generalized distributions. J. Statist. Comput. & Simul., 81(7): 883-898.

[17] Cordeiro, G. M., Lemonte, A. J. and Ortiga, E. E. M. (2014a). The Marshall-Olkin family of distributions: Mathematical properties and new models. J. Statat, Theor. Pract., 8: 343-366.

[18] Cordeiro, G. M., Ortiga, E. E. M. and Silva, G. O. (2014b). The Kumaraswamy modified Weibull distribution: Theory and applications. J. Stat. Comput. Simul., 84: 1387-1411.

[19] Correa M. A., Nogueira D. A. and Ferreira E. B. (2012). Kumaraswamy normal and Azzalini's skew normal modeling asymmetry. Sigmae, 1: 65-83.

[20] Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. Commun. Statist.- Theory and Meth., 31: 497-512.

[21] Jones, M. C. (2004). Families of distributions arising from distributions of order statistics. Test, 13(1): 1-43.

[22] Jose, K. K. (2011). Marshall-Olkin family of distributions and their applications in reliability theory, time series modeling and stress strength analysis. Proc. ISI 58th World Statist. Congr. Int. Stat. Inst., 21st-26th August 2011, Dublin: 3918-3923.

[23] Lemonte, A. J. (2014). The beta log-logistic distribution. Braz. J. Probab. Statist., 28: 313-332.

[24] Mameli, V. and Musio, M. (2013). A Generalization of the Skew-Normal Distribution: The Beta skew-normal. Commun. Statist.-Theory and Meth., 42(12): 2229-2244.

[25] Marshall, A. W. and Olkin, I. (1997). A new method for adding a prameter to a family of distributions with application to the exponential and Weibull families. Biometrika, 84: 641-652.

[26] McLachlan, G. and Peel, D. (2000). Finite mixture models. Wiley Series in Probab. and Statist. John Wiley New York.

[27] Nadarajah, S. (2005). A generalized normal distribution. J. Appl. Statist., 32(7): 685-694.

[28] Popović, B. V., Cordeiro, G. M. and Pascoa, M. A. R. (2017). A new extended mixture normal distribution. Math. Commun., 22: 53-73.

[29] Titterington, D. M., Smith, A. F. M. and Makov, U. E. (1986). Statistical analysis of finite mixture distributions. New York: Wiley.

[30] Wu, C. F. (1983). On the convergence properties of the EM algorithm. Ann. Statist., 11(1): 95-103.