On the robustness of an epsilon skew extension for Burr III distribution on the real line

Mehmet Niyazi Çankaya¹,² · Abdullah Yalçınkaya³ · Ömer Altındağ⁴ · Olcay Arslan³

Received: 7 June 2017 / Accepted: 11 December 2018 / Published online: 2 January 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract
Burr III (BIII) distribution is used in a wide variety of fields, such as lifetime data analysis, reliability theory, and financial literature, and suchlike. It is defined on the positive axis and has two shape parameters, say \( c \) and \( k \). These shape parameters make the distribution quite flexible. They also control the tail behaviour of the distribution. In this study, we extend BIII distribution to the real line and also add a skewness parameter, say \( \varepsilon \), with an epsilon skew extension approach. When the parameters \( c \) and \( k \) have a relationship such that \( ck \leq 1 \), it is skew unimodal. Otherwise, it is skew bimodal with the same level of peaks on the negative and positive sides of the real line. Thus, the epsilon skew extension of Burr III (ESBIII) distribution with only three parameters can provide adequate fits for data sets that may have heavy-tailedness, skewness, unimodality or bimodality. A location-scale form of this distribution is also given. Distributional properties are investigated. The maximum likelihood (ML) estimation method for the parameters of ESBIII is considered. The robustness properties of the ML estimators are studied in terms of the boundedness of the influence function. Further, tail behaviour of ESBIII distribution is also examined to explore the robustness of ESBIII distribution against the outliers. The modelling capacity of this distribution is illustrated using two real data examples.

Keywords  Burr III distribution · Epsilon skew extension · Asymmetry · Robustness · Bimodality · Unimodality

1 Introduction
It is widely-accepted that using flexible and heavy-tailed distributions is important to fit data sets for practitioners in applied science. Twelve different distributions, which are quite flexible and which are useful in modelling data sets, have been proposed...
In particular, Burr XII distribution is generally used as a flexible and heavy-tailed distribution. See Ali et al. (2015), Burr (1942), Burr and Cislak (1968), Burr (1973), Rodriguez (1977) and Wingo (1983a) for the applications of Burr XII distribution. However, Burr III (BIII) distribution has a wider skewness and kurtosis region than that of Burr XII distribution, as emphasized in Tadikamalla (1980). BIII distribution has two shape parameters, \( c \) and \( k \), such that probability density function of the distribution becomes L-shaped or unimodal when \( ck \leq 1 \) or \( ck > 1 \), respectively. It is also heavy-tailed distribution. BIII distribution is defined on the positive axis, so it is widely used in reliability theory (Zimmer et al. 1998; Zoraghi et al. 2012; Wang et al. 1996; Wingo 1983b), forestry studies (Gove et al. 2008; Lindsay et al. 1996) and hydrology studies (Hao and Singh 2009). It is also utilized in quality and control engineering (Abd-Elfattah and Alharbey 2012; Azimi and Yaghmaei 2013; Shao et al. 2008).

In statistical applications, distributions defined on the real line are needed as well as the distributions defined on the positive axis. The most celebrated distribution defined on the real line is normal distribution. It is well-known that normal distribution is symmetric and light-tailed. However, in the past, researchers have focused on heavy-tailed and/or skewed distributions in order to model various data sets with heavy tails and/or asymmetry. Recently, many researchers have dealt with generalizing existing distributions to model asymmetric data sets. There are various generalization techniques which extend a symmetric distribution to an asymmetric form. An approach to construct a skewed distribution, which first introduces the skew normal (SN), has been proposed by Azzalini (1985). In addition, a skew \( t \) distribution has been presented by Azzalini and Capitanio (2003) and Jones and Faddy (2003). One can refer to Genton (2004) for Azzalini type skew distributions. A so-called epsilon skew extension method was introduced by Mudholkar and Hutson (2000) to generalize symmetric distributions to asymmetric form with a skewness parameter, say \( \varepsilon \), such that \(|\varepsilon| < 1\). In that particular study, the epsilon skew normal (ESN) distribution was presented. If the skewness parameter \( \varepsilon \) is chosen as zero, it reduces to the normal distribution. Since then, this method has been applied to several well known distributions. An epsilon skew extension of exponential power distribution (ESEP) was also considered by Elsaloukh et al. (2005). The exponential power distribution and its generalized family are proposed by Box and Tiao (1973), Çankaya et al. (2015), Çankaya (2018), Mineo and Ruggieri (2005) and Subbotin (1923). Special cases of ESEP distribution reduce ESN distribution (Mudholkar and Hutson 2000) and the epsilon skew Laplace (ESL) distribution (Elsalloukh 2008). The epsilon skew extension of reflected gamma distribution was taken into account by Abdulah and Elsalloukh (2013). A comprehensive content about asymmetric distributions which includes a wide family of symmetric distributions as a special case can be found in Arellano-Valle et al. (2005). Besides the asymmetry, a property of heavy-tailedness has become indispensable to be able to handle the observations that are distant from the bulk of data. It should be noted that if BIII distribution, which will be extended from positive axis to negative axis, is heavy-tailed, then extended BIII distribution, which will be defined on the real line, will also be heavy-tailed. Skew generalized \( t \) (SGT) distribution as a scale mixture of ESEP distribution was introduced by Arslan and Genç (2009) and investigated for its robustness property. Moreover, ESN distribution proposed by Mudholkar and Hutson...
was generalized, which is lack of modelling heavy-tailed data sets, in order to achieve a heavy-tailed extension via a scale mixture approach. It should be noted that all of the above distributions are skew unimodal distributions. However, one may encounter bimodal data sets in real life problems. Hence, modelling the bimodality is an another challenging problem. For this purpose, one can use finite mixture of distributions. However, some bimodal distributions have been proposed to model a bimodality directly. Certain (recent) examples of bimodal distributions can be found in Andrade and Rathie (2016), Arellano-Valle et al. (2010), Bolfarine et al. (2013), Cooray (2013), Çankaya et al. (2015), Dexter (2015), Donatella and Von Dorp (2013), Genç (2013), Gui (2014), Hassan and El-Bassiouni (2016), Jamalizadeh et al. (2011), Jones (2009), Rathie et al. (2016), Rêgo et al. (2012), Shams and Alamatsaz (2013) and Venegas et al. (2017).

In this study, we first reflect BIII distribution to the real line and consider its epsilon skew extension in order to obtain a quite flexible distribution that can also cope with heavy-tailed and/or skewed empirical distributions on the real line. Hereafter, the proposed distribution will be called an epsilon skew extension for Burr III (ESBIII) distribution. If the skewness parameter $\varepsilon$ is not zero, ESBIII distribution can be a skew unimodal or skew bimodal distribution with same level of peaks around location, depending on the values of the shape parameters $c$ and $k$. The proposed distribution may also be an alternative to unimodal/bimodal symmetric and heavy-tailed distributions when $\varepsilon = 0$. Clearly, ESBIII can provide flexible modelling with its two shapes and one skewness parameters. The location-scale form of ESBIII will become a wider class that allows it to be utilized for estimating the location and scale of data sets efficiently.

The remainder of this paper is set up as follows. In Sect. 2, ESBIII distribution is presented and its distributional properties are investigated. Its location-scale form is also given. In Sect. 3, maximum likelihood (ML) estimation for the parameters of ESBIII distribution are considered. In Sect. 4, the robustness of the ML estimators of the parameters is investigated in terms of boundedness of influence function. We also investigate the robustness property of ESBIII distribution in terms of heavy-tailedness. In Sect. 5, certain real data examples are provided in order to compare the modelling performance of ESBIII over certain alternative candidates with (explicit) analytical expressions for their cumulative distribution functions in the same class. The last section includes conclusions and discussions.

## 2 Epsilon skew extension for Burr III distribution on the real line

Consider a random variable $Z$ and assume that it has a BIII distribution with cumulative distribution function (cdf):

$$G(z) = (1 + z^{-c})^{-k}, \quad z > 0; \ c, k > 0, \quad (2.1)$$

and probability density function (pdf):
\[ g(z) = ckz^{-(c+1)}(1 + z^{-c})^{-(k+1)}, \quad z > 0; \ c, k > 0, \tag{2.2} \]

where \( c > 0 \) and \( k > 0 \) are shape parameters that identify characteristics of the distribution such as dispersion, peakedness, and tail thickness. As noted earlier, BIII distribution is L-shaped when \( ck \leq 1 \) and unimodal when \( ck > 1 \) (Burr 1942).

By extending BIII distribution to the real line via an \( \varepsilon \)-skew approach, ESBIII distribution is obtained. The following theorem establishes the construction of ESBIII by means of a stochastic representation.

**Theorem 2.1** Let the random variable \( Z \) have a BIII distribution with the shape parameters \( c \) and \( k \), and let \( U \) be a discrete random variable with the following values and the corresponding probabilities:

\[
U = \begin{cases} 
1 + \varepsilon, & P(U = 1 + \varepsilon) = \frac{1+\varepsilon}{2} \\
-(1 - \varepsilon), & P(U = -1 + \varepsilon) = \frac{1-\varepsilon}{2},
\end{cases} \tag{2.3}
\]

where \( \varepsilon \in (-1, 1) \) and \( P(\cdot) \) denotes the probability. Assume that \( Z \) and \( U \) are independent and let \( X = ZU \). Then, the cdf and pdf of the random variable \( X \) are

\[
F(x) = \begin{cases} 
\frac{1-\varepsilon}{2} \left[ 1 - \left( 1 + \left( \frac{-x}{1-\varepsilon} \right)^{-c} \right)^{-k} \right], & x < 0 \\
\frac{1-\varepsilon}{2} + \frac{1+\varepsilon}{2} \left[ 1 + \left( \frac{x}{1+\varepsilon} \right)^{-c} \right]^{-k}, & x > 0,
\end{cases} \tag{2.4}
\]

and

\[
f(x) = \frac{ck}{2} \left( \frac{s(x)x}{1 + s(x)\varepsilon} \right)^{-(c+1)} \left\{ 1 - \left( \frac{s(x)x}{1 + s(x)\varepsilon} \right)^{-c} \right\}^{-(k+1)}, \tag{2.5}
\]

respectively, where \( x \in \mathbb{R} \setminus \{0\} \), \( c, k > 0 \) are shape parameters, \( \varepsilon \in (-1, 1) \) is skewness parameter and \( s(\cdot) \) denotes the signum function.

**Proof** By using the distribution function technique, the cdf of \( X \) for \( x < 0 \) is

\[
F_-(x) = P(ZU \leq x) = \frac{1-\varepsilon}{2} \left\{ 1 - G\left( \frac{-x}{1-\varepsilon} \right) \right\}, \tag{2.6}
\]

and for \( x > 0 \) is

\[
F_+(x) = P(ZU \leq x) = \frac{1-\varepsilon}{2} + \frac{1+\varepsilon}{2} \left\{ G\left( \frac{x}{1+\varepsilon} \right) \right\}. \tag{2.7}
\]

Taking the derivative of the Eq. (2.4) with respect to \( x \) yields the pdf given in the Eq. (2.5). \( \square \)
It should be noted that the zero point is excluded from the domain of the pdf since the pdf has a singularity at zero. The definition of ESBIII distribution is given as follows:

**Definition 2.1** The distribution of the random variable $X$ with pdf in the Eq. (2.5) is called the epsilon skew extension for Burr III distribution with parameters $\varepsilon$, $c$ and $k$, and is denoted in short by $X \sim \text{ESBIII}(\varepsilon, c, k)$.

The location-scale form of ESBIII distribution is given below:

**Definition 2.2** Let $X \sim \text{ESBIII}(\varepsilon, c, k)$. Then the distribution of the random variable $Y = \mu + \sigma X$, $\mu \in \mathbb{R}$, $\sigma > 0$ is called the location-scale form of ESBIII distribution which has the following pdf:

$$f(y) = \frac{ck}{2\sigma} \left( \frac{s(y - \mu) \cdot (y - \mu)}{\sigma(1 + s(y - \mu)\varepsilon)} \right)^{-c} \left\{ 1 + \left( \frac{s(y - \mu) \cdot (y - \mu)}{\sigma(1 + s(y - \mu)\varepsilon)} \right)^{-c} \right\}^{-(k+1)} ,$$

where $y \in \mathbb{R}\setminus\{\mu\}$, $\mu$ and $\sigma$ are location and scale parameters, respectively. Here, we denote the distribution of $Y$ by ESBIII($\mu, \sigma, \varepsilon, c, k$) and write $Y \sim \text{ESBIII}(\mu, \sigma, \varepsilon, c, k)$.

The parameters $\mu$ and $\sigma$ determine the location and scale of the distribution, respectively. The parameter $\varepsilon$ controls the skewness of the distribution. The parameters $c$ and $k$ both control the shape of the density function. It should be noted that it will be symmetric unimodal or symmetric bimodal distribution when $\varepsilon = 0$.

We now discuss some properties of ESBIII distribution given by following subsections.

2.1 Density shapes and flexibility

The pdf of ESBIII distribution may have many desirable forms by choosing specific values of the shape parameters $c$ and $k$. For example, the shape of ESBIII distribution for certain specific values of $c$ and $k$ can be similar to the shapes of ESL and EST distributions. In this context, it is considered that the new distribution can model data sets which cannot be precisely modelled by ESL and EST distributions, as exemplified in Sect. 5. ESBIII distribution has properties of both leptokurtic and platykurtic shape distributions. When $ck \leq 1$, the pdf of ESBIII is unimodal. For $ck \leq 1$, when the difference between $c$ and $k$ is extremely large, the shape of distribution is platykurtic, and when the difference between $c$ and $k$ is extremely small, the shape of distribution is leptokurtic. The distributions proposed by Box and Tiao (1973), Çankaya et al. (2015), Çankaya (2018), Mineo and Ruggieri (2005) and Subbotin (1923) are also in the platykurtic-leptokurtic family. When the values of $ck$ are close to 1 from the left hand side of 1, the distribution has a needle-like peak. When $ck > 1$, the pdf of ESBIII is bimodal.
The new distribution also has a skewness parameter. Therefore, the base lengths of each group in the negative and positive sides of density, or the left and right sides of location, may become different. In Figs. 1, 2 and 3, plots are given to understand the behaviour of ESBIII distribution for different values of parameters. Figures 1, 2 and 3a display the symmetric case and Figs. 1, 2 and 3b represent the role of the skewness parameter $\epsilon$. Note that ESBIII distribution is negatively and positively skewed when $\epsilon < 0$ and $\epsilon > 0$, respectively. The plots of the densities for $\epsilon < 0$ present mirror images of the plots of the densities for $\epsilon > 0$, as given in Figs. 1, 2 and 3b. Therefore, we do not give the plots of the densities for $\epsilon < 0$.

It can be seen from Figs. 1, 2 and 3, the pdf of ESBIII distribution takes different forms, such as unimodal, bimodal and their $\epsilon$-skew forms. Therefore, ESBIII distribution may be a good alternative for modelling various data sets when compared to skew unimodal and skew bimodal distributions used in real data applications.

We observe the bimodality and unimodality of ESBIII distribution. These features can be easily verified by investigating the derivative of the pdf. Let us examine the modes of ESBIII. Consider the pdf of ESBIII($\mu, \sigma, \epsilon, c, k$) distribution. For a unimodal case, that is $ck \leq 1$, the mode of the pdf is expected to be $\mu$. However, the domain of pdf does not include the point $\mu$, since the pdf has singularity at $\mu$, which is a theoretical problem, but which can be ignored in practical applications. For a bimodal case, $ck > 1$, the modes are found by setting the derivative of pdf to zero as $
abla \mu - \sigma (1 - \epsilon) (ck - 1)/(c + 1) ^{1/c}$ and $\mu + \sigma (1 + \epsilon) (ck - 1)/(c + 1) ^{1/c}$.

![Fig. 1](Color online) Examples of unimodal case from ESBIII distribution, $\epsilon = 0$ (symmetric) and $\epsilon = 0.5$ (skewed to positive side). ESBIII can be a member of normal, Laplace and uniform distributions.

---

(c) Springer
Fig. 2 (Color online) Examples of unimodal case from ESBIII distribution, $\epsilon = 0$ (symmetric) and $\epsilon = 0.5$ (skewed to positive side). ESBIII cannot only be a member of Laplace but also it has peakedness which is like needle.

Fig. 3 (Color online) Examples of bimodal case from ESBIII distribution, $\epsilon = 0$ (symmetric) and $\epsilon = 0.5$ (skewed to positive side). There are different levels of peakedness for the different values of the parameters $c$ and $k$. 
2.2 Moments and their properties

The moments of a random variable $X$ distributed as ESBIII($\epsilon, c, k$) are given by the following theorem:

**Theorem 2.2** Let $X \sim \text{ESBIII}(\epsilon, c, k)$. The $r$th, $r > 0$, non-central moments of $X$ are given by:

$$
\mu_r = E(X^r) = kB \left(1 - \frac{r}{c}, \frac{r}{c} + k\right) \left[\frac{(1 + \epsilon)^{r+1} + (-1)^r(1 - \epsilon)^{r+1}}{2}\right], \ c > r,
$$

(2.9)

where $B(\cdot)$ denotes the Beta function.

**Proof** By considering the stochastic representation of ESBIII distribution given in Theorem 2.1 and using the independence of $Z$ and $U$, we can write $E(X^r) = E(Z^r U^r) = E(Z^r)E(U^r)$. It is known that $E(Z^r) = kB \left(1 - \frac{r}{c}, \frac{r}{c} + k\right)$ for $c > r$. Further, the $r$th moment of the random variable $U$ can be easily obtained as $E(U^r) = \left[\frac{(1+\epsilon)^{r+1} + (-1)^r(1-\epsilon)^{r+1}}{2}\right]$. Then, we get Eq. (2.9), which completes the proof. \(\square\)

In Theorem 2.2, the condition $c > r$ is necessary for the existence of the $r$th moment. The above moments are related with the standard case. For the location-scale case, we will have the following moments. Let $Y$ be a random variable distributed as ESBIII($\mu, \sigma, \epsilon, c, k$). The $r$th moment of the random variable $Y$ is given by the following theorem:

**Theorem 2.3** The $r$th moment of the random variable $Y$ distributed as ESBIII ($\mu, \sigma, \epsilon, c, k$) is given by:

$$
\mu'_r = E(Y^r) = \sum_{k=0}^{r} \binom{r}{k} \sigma^k \mu^{r-k} \mu_k, \ c > r,
$$

(2.10)

where $\mu_k = E(X^k)$.

**Proof** This is an immediate consequence of the binomial theorem since $Y = \mu + X\sigma$. \(\square\)

The expectation, variance, skewness and kurtosis coefficients of ESBIII($\epsilon, c, k$) can be easily found by utilizing Theorem 2.2. These are given in the following corollaries.

**Corollary 2.1** The expectation and variance of the random variable $X$ are

$$
E(X) = kB \left(1 - \frac{1}{c}, \frac{1}{c} + k\right) 2\epsilon, \ c > 1,
$$

(2.11)

\(\square\) Springer
and

\[ Var(X) = kB \left( 1 - \frac{2}{c}, \frac{2}{c} + k \right) (1 + 3\varepsilon^2) \]

\[ -k^2B^2 \left( 1 - \frac{1}{c}, \frac{1}{c} + k \right) 4\varepsilon^2, \quad c > 2, \quad (2.12) \]

respectively.

**Corollary 2.2** The skewness and kurtosis coefficients of the random variable \( X \) are obtained using the following formulae:

\[ \text{Skewness}(X) = \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3}{(\mu_2 - \mu_1^2)^{3/2}}, \quad c > 3, \quad (2.13) \]

and

\[ \text{Kurtosis}(X) = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_2\mu_1^2 - 3\mu_1^4}{(\mu_2 - \mu_1^2)^2}, \quad c > 4, \quad (2.14) \]

respectively. Here \( \mu_r \) with \( r = 1, 2, 3, 4 \) are obtained using Eq. (2.9).

We can further explore skewness and kurtosis in order to better understand the behaviour of ESBIII distribution. Tables 1 and 2 present the values of skewness and kurtosis for unimodal cases, respectively. Note that the distribution is unimodal when \( ck \leq 1 \). Therefore, we give skewness and kurtosis values when the values of \( ck \) vary from 0 to 1. We take that the shape parameter \( c \) is bigger than 3 and 4 for the existence of skewness and kurtosis values, respectively. It should be noted that when the sign of skewness parameter is positive/negative, the distribution is positively/negatively skewed. For both cases, the kurtosis of distribution is positive. In addition, in order to understand the behaviour of the skewness and kurtosis values of ESBIII distribution

| Table 1 | The skewness values of ESBIII distribution for the different values of parameters \( c \) and \( k \) for the case \( \varepsilon = 0.99 \) |
|---------|-----------------------------------------------------------------|
| \( c \) | \( 0.05 \) | \( 0.10 \) | \( 0.25 \) | \( 0.50 \) | \( 0.75 \) | \( 0.90 \) | \( 0.95 \) | \( 1 \) |
| 3.05    | 78.44  | 60.76  | 48.96  | 47.65  | 49.68  | 51.17  | 51.67  | 52.16  |
| 3.10    | 40.99  | 31.58  | 25.15  | 24.26  | 25.20  | 25.94  | 26.19  | 26.44  |
| 3.50    | 11.04  | 8.22   | 6.06   | 5.45   | 5.50   | 5.62   | 5.66   | 5.71   |
| 4       | 7.31   | 5.29   | 3.64   | 3.03   | 2.93   | 2.95   | 2.97   | 2.98   |
| 5       | 5.47   | 3.84   | 2.41   | 1.77   | 1.57   | 1.52   | 1.51   | 1.51   |
| 10      | 4.26   | 2.88   | 1.58   | 0.86   | 0.53   | 0.40   | 0.36   | 0.33   |
| 50      | 4.01   | 2.68   | 1.40   | 0.65   | 0.27   | 0.11   | 0.06   | 0.01   |
| 100     | 4.00   | 2.67   | 1.39   | 0.65   | 0.26   | 0.10   | 0.05   | 0.00   |
Table 2  The kurtosis values of ESBIII distribution for the different values of parameters $c$ and $k$ for the case $\varepsilon = 0$ and $\varepsilon = 0.99$.

| $c$  | $ck$  | 0.05  | 0.25  | 0.50  | 0.75  | 1    |
|------|-------|-------|-------|-------|-------|------|
| 4.05 | (687.8, 792.9) | (156.3, 295.2) | (90.1, 274.9) | (68.1, 300.7) | (57.2, 337.3) |
| 4.10 | (352.4, 402.6) | (80.1, 146.7)  | (46.2, 134.9)  | (34.9, 146.8)  | (29.3, 164.4)  |
| 4.50 | (84.1, 91.3) | (19.2, 29.0)  | (11.1, 24.5)  | (8.4, 25.7)  | (7.1, 28.2)  |
| 5    | (50.7, 53.0) | (11.6, 15.1)  | (6.7, 11.7)  | (5.1, 11.8)  | (4.3, 12.6)  |
| 10   | (23.9, 23.0) | (5.5, 4.8)  | (3.2, 2.9)  | (2.4, 2.6)  | (2.0, 2.5)  |
| 50   | (20.8, 19.6) | (4.8, 3.8)  | (2.8, 2.2)  | (2.1, 1.83)  | (1.8, 1.82)  |
| 100  | (20.7, 19.5) | (4.7, 3.8)  | (2.7, 2.1)  | (2.1, 1.81)  | (1.8, 1.80)  |

The first and second values in the parenthesis show the kurtosis values for two cases $\varepsilon = 0$ and $\varepsilon = 0.99$, respectively.

Fig. 4 (Color online) The skewness values of ESBIII distribution for the different values of skewness parameter.

under different values of skewness parameter, we show Figs. 4 and 5 for unimodal cases.

From Table 1, it can be seen that the distribution presents greatest asymmetry when the value of $ck$ goes to zero and $c$ tends to 3. Moreover, it tends to be symmetric when the value of $ck$ goes to one and $c$ goes to $\infty$. Additionally, depending on the values of parameters, the skewness values show great variability. Therefore, ESBIII distribution can have great flexibility for modelling data sets that have very wide range of skewness.

Table 2 presents the range of the kurtosis coefficient for ESBIII distribution when the asymmetry parameter $\varepsilon = 0$ and $\varepsilon = 0.99$. We observe from Table 2 that the distribution is leptokurtic when $c$ goes to 4, but it tends to platykurtic distribution when $c$ goes to infinity and the values of $ck$ increase to 1. The high kurtosis value (792.9) shows that the distribution provides enough flexibility for modelling heavy-tailed data sets. On the other hand, the low kurtosis value (1.80) indicates that the distribution also provides enough flexibility for modelling light-tailed data sets.
In Tables 1 and 2, we only provide a positively skewed case for $\varepsilon = 0.99$. For the negatively skewed case $\varepsilon = -0.99$, the skewness values are exactly the same as the values in Table 1, except that their signs are negative for all values of the parameters $c$ and $k$. Similarly, when $\varepsilon = -0.99$, the kurtosis values are exactly the same as the values shown in Table 2. Therefore, we did not give the skewness and kurtosis values for negatively skewed cases. Consequently, we observe from Tables 1 and 2 and Figs. 4 and 5 that ESBIII provides great flexibility for modelling data sets having symmetric, asymmetric, unimodal, bimodal, light-tailed or heavy-tailed distributions.

2.3 Quantile function and random number generation

The quantile function of ESBIII can be easily derived by inverting the distribution function $F$. Therefore, the quantile function $F^{-1}(q)$ for ESBIII($\mu, \sigma, \varepsilon, c, k$) is obtained as:

\[
x_q = F^{-1}(q) = \begin{cases} 
\mu - \sigma (1 - \varepsilon) \left[ \left( 1 - \frac{2q}{1 + \varepsilon} \right)^{-1/k} - 1 \right]^{-1/c}, & q \leq \frac{1 - \varepsilon}{2} \\
\mu + \sigma (1 + \varepsilon) \left[ \left( \frac{2q}{1+\varepsilon} - \frac{1 - \varepsilon}{1+\varepsilon} \right)^{-1/k} - 1 \right]^{-1/c}, & q > \frac{1 - \varepsilon}{2}.
\end{cases}
\]

(2.15)

By using the quantile function, the random numbers from ESBIII($\mu, \sigma, \varepsilon, c, k$) distribution can be easily generated. The steps of the random number generation procedure are given below:

**Step 1:** Set the values of parameters $\mu, \sigma, \varepsilon, c$ and $k$

**Step 2:** Generate a random number $q$ from uniform distribution on the interval $(0, 1)$
**Step 3:** Then, $x_q = F^{-1}(q)$ is a random number from ESBIII($\mu, \sigma, \epsilon, c, k$) for the random number $q$.

### 2.4 Rényi and Shannon entropies

An entropy for a random variable denotes the variation of uncertainty and it can be used for characterization of probability distributions. A general entropy measure was defined by Rényi (1961). Rényi entropy for a random variable $X$ with pdf $f$ is given by:

$$J_R(\alpha) = \frac{1}{1 - \alpha} \log \left\{ \int_{\mathbb{R}} f^\alpha(x) dx \right\}, \quad \alpha > 0, \quad \alpha \neq 1. \quad (2.16)$$

Rényi entropy is a generalization of Shannon entropy previously introduced by Shannon (1961). Shannon entropy is defined as $E(-\log f(X))$, which corresponds to the limit of Rényi entropy when $\alpha$ goes to 1, that is:

$$E(-\log f(X)) = J_R(1) = \lim_{\alpha \to 1} J_R(\alpha). \quad (2.17)$$

**Theorem 2.4** Let $X \sim ESBIII(\mu, \sigma, \epsilon, c, k)$. Then, Rényi and Shannon entropies of $X$ are

$$J_R(\alpha) = \frac{1}{1 - \alpha} \log \left[ \left( \frac{ck}{2\sigma} \right)^\alpha \frac{2\sigma}{c} B\left( \alpha(k - 1/c) + 1/c, \alpha(1 + 1/c) - 1/c \right) \right], \quad \alpha(k - 1/c) + 1/c > 0, \quad \alpha(1 + 1/c) - 1/c > 0, \quad \alpha > 0, \quad \alpha \neq 1 \quad (2.18)$$

where $\alpha(k - 1/c) + 1/c > 0, \quad \alpha(1 + 1/c) - 1/c > 0, \quad \alpha > 0, \quad \alpha \neq 1$ and

$$J_R(1) = (k + 1)\psi(k + 1) + (1 + 1/c)\gamma + (1/c - k)\psi(k) + \log \left( \frac{2\sigma}{ck} \right), \quad (2.19)$$

respectively. Here $B(\cdot)$ is beta function, $\psi(\cdot)$ is digamma function and $\gamma$ is Euler constant.

The proof of Theorem 2.4 is given in the “Appendix”. Note that the arguments in the beta function must be greater than zero to guarantee the existence of the corresponding integral. Concerning Rényi and Shannon entropies given in Eqs. (2.18) and (2.19), respectively, it can be seen that the location parameter $\mu$ has no impact on the entropies of Rényi and Shannon, as expected. It is interesting to note that the skewness parameter $\epsilon$ also has no impact on Rényi and Shannon entropies. Such a situation, which is also proposed by Arslan and Genç (2009), occurs for the symmetric distribution asymmetrized with the epsilon-skew extension of BIII distribution to the real line.
3 Maximum Likelihood estimators of parameters in ESBIII and an algorithm for computation

Let \( x = (x_1, x_2, \ldots, x_n) \) be a random sample of size \( n \) from an ESBIII distributed population with the parameters \( \mu, \sigma, \varepsilon, c \) and \( k \). The ML estimation method will be preferred to estimate the unknown parameters, owing to its desirable properties. The log-likelihood function is given by:

\[
L(x; \mu, \sigma, \varepsilon, c, k) = n \log(ck) - n \log(2\sigma) - (c + 1) \sum_{i=1}^{n} \log(z_i)
- (k + 1) \sum_{i=1}^{n} \log(1 + z_i^{-c}),
\]

where \( s_i = s(x_i - \mu) \) and \( z_i = s_i \cdot (x_i - \mu)/(\sigma[1 + s_i \cdot \varepsilon]) \).

The ML estimators of the parameters \( \mu, \sigma, \varepsilon, c \) and \( k \) will be obtained as the solution of the following estimating equations:

\[
\frac{\partial L}{\partial \mu} = (c + 1) \sum_{i=1}^{n} (x_i - \mu)^{-1} - c(k + 1) \sum_{i=1}^{n} \frac{(x_i - \mu)^{-1}}{1 + z_i^c} = 0,
\]

\[
\frac{\partial L}{\partial \sigma} = \frac{n c}{\sigma} - c(k + 1) \sum_{i=1}^{n} \frac{\sigma^{-1}}{1 + z_i^c} = 0,
\]

\[
\frac{\partial L}{\partial \varepsilon} = (c + 1) \sum_{i=1}^{n} (s_i + \varepsilon)^{-1} - c(k + 1) \sum_{i=1}^{n} \frac{(s_i + \varepsilon)^{-1}}{1 + z_i^c} = 0,
\]

\[
\frac{\partial L}{\partial c} = \frac{n}{c} - \sum_{i=1}^{n} \log(z_i) + (k + 1) \sum_{i=1}^{n} \frac{\log(z_i)}{1 + z_i^c} = 0,
\]

\[
\frac{\partial L}{\partial k} = \frac{n}{k} - \sum_{i=1}^{n} \log(1 + z_i^{-c}) = 0.
\]

Since it is not possible to solve the Eqs. (3.2)–(3.6) analytically, certain numerical methods should be used to compute the ML estimates of the parameters. These equations have to be solved simultaneously. For this purpose, a number of classical approaches, such as bisection, secant or Newton-Raphson may be used. As an alternative to these classical approaches, one could use numerical optimization techniques that are available on known programming environments, such as MATLAB or MAPLE. We suggest using the \textit{lsqnonlin} function predefined in MATLAB 2016a. Note that the function \textit{lsqnonlin} enables the solving of the non-linear equations with constraints. Being able to place constraints for the search space of parameters in non-linear equations helps to avoid non-defined solutions. It should be emphasized that the function \textit{lsqnonlin} may even fail to achieve desirable solutions as the estimates for the parameters of interests, since the number of equations is relatively high. If
such a situation occurs, a step-by-step solution of the estimating equations given in (3.2)–(3.6) may be carried out. The following algorithm can be used to obtain the solution of the estimating equations:

**Algorithm 1**: The steps for the solution of the estimating equations given in (3.2)–(3.6)

**Step 1**: Define the number of maximum iterations as $\max$ and the tolerance level as $\delta > 0$ to finalize the computation.

**Step 2**: Set $i = 0$

Set the initial values of the parameters $\mu, \sigma, \varepsilon, c$ and $k$ as $\mu^{(0)} = \text{Median}(x), \sigma^{(0)} = \text{MAD}(x) = \text{Median}(|x - \text{Median}(x)|), \varepsilon^{(0)} = 0, c^{(0)} = 1, k^{(0)} = 1$. Therefore, the numerical optimization begins when the other parameters are supplied, as is given by Step 3.

**Step 3**:

- $\hat{\mu}^{(i+1)} := \mathcal{R} \left( \frac{\partial}{\partial \mu} L(x; \mu, \hat{\sigma}^{(i)}, \hat{\varepsilon}^{(i)}, \hat{k}^{(i)}) = 0, \mu_1 \leq \mu \leq \mu_2, \mu_1, \mu_2 \in \mathbb{R} \right)$,
- $\hat{\sigma}^{(i+1)} := \mathcal{R} \left( \frac{\partial}{\partial \sigma} L(x; \sigma, \hat{\mu}^{(i)}, \hat{\varepsilon}^{(i)}, \hat{k}^{(i)}) = 0, \sigma_1 \leq \sigma \leq \sigma_2, \sigma_1, \sigma_2 > 0 \right)$,
- $\hat{\varepsilon}^{(i+1)} := \mathcal{R} \left( \frac{\partial}{\partial \varepsilon} L(x; \varepsilon, \hat{\mu}^{(i)}, \hat{\sigma}^{(i)}, \hat{k}^{(i)}) = 0, -1 < \varepsilon < 1 \right)$,
- $\hat{k}^{(i+1)} := \mathcal{R} \left( \frac{\partial}{\partial k} L(x; k, \hat{\mu}^{(i)}, \hat{\sigma}^{(i)}, \hat{\varepsilon}^{(i)}, \hat{k}^{(i)}) = 0, k_1 \leq k \leq k_2, k_1, k_2 > 0 \right)$,

Set $i \leftarrow i + 1$

**Step 4**: $\hat{\theta} = (\hat{\mu}^{(i)}, \hat{\sigma}^{(i)}, \hat{\varepsilon}^{(i)}, \hat{k}^{(i)})$. If $||\hat{\theta}^{(i+1)} - \hat{\theta}^{(i)}|| > \delta$ and $i < \max$, then go to Step 3; otherwise go to Step 5.

**Step 5**: ML estimates of the parameters are $\hat{\theta} = (\hat{\mu}^{(i+1)}, \hat{\sigma}^{(i+1)}, \hat{\varepsilon}^{(i+1)}, \hat{k}^{(i+1)})$.

Note that $\mathcal{R}$ is a function to enable the root of non-linear equations to be solved numerically via `lsqnonlin` for their corresponding parameters $\mu, \sigma, \varepsilon, c$ and $k$. Here, $\mu_1, \mu_2, \sigma_1, \sigma_2, c_1, c_2, k_1$ and $k_2$ are quantitative values predefined by the user manually. We try different values for the intervals $[c_1, c_2]$ and $[k_1, k_2]$, especially for a case where there may be a convergence to local points in the optimization. It is visually observable from the shape of the distributed data for the intervals $[\mu_1, \mu_2]$ and $[\sigma_1, \sigma_2]$. It should be noted that the different values for the intervals are tried until the smallest values of goodness of fit test statistics and information criteria can be obtained.

## 4 Robustness

In this section, we will explore the robustness of the ML estimators in terms of influence function and tail behaviour of ESBIII distribution.

### 4.1 Robustness of the ML estimators

Influence function is one of the measures to assess the robustness of an estimator, and estimators with bounded influence function are considered to have good local robustness property. Therefore, for this reason, we will find the influence function of the ML estimators for the parameters of ESBIII distribution to evaluate their robustness capability. In general, the influence function (IF) of the ML estimators is a linear
transformation of the related score vector. Concerning ESBII distribution, IF can be
defined as follows:

Let $\theta = (\mu, \sigma, \varepsilon, c, k)$ be a parameter vector, $f$ be pdf of ESBIII, $\hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{\varepsilon}, \hat{c}, \hat{k})$ be ML estimators of parameters $\theta$, $\Psi = (\psi_\mu, \psi_\sigma, \psi_\varepsilon, \psi_c, \psi_k)$ be a vector for score functions $\psi_\mu, \psi_\sigma, \psi_\varepsilon, \psi_c$ and $\psi_k$. Then, a score vector of parameters $\theta$ is

$$\Psi(x; \theta) = \frac{\partial}{\partial \theta} \left( -\log f(x; \theta) \right),$$  

and the IF for the ML estimators $\hat{\theta}$ is

$$IF(x; \hat{\theta}) = -\left( E \left[ \frac{\partial}{\partial \theta} \Psi(x; \theta) \right] \right)^{-1} \Psi(x; \theta)|_{\theta = \hat{\theta}},$$

where the expectation is taken over the underlying distribution ESBIII. Clearly, the boundedness of the influence function will depend on the boundedness of the elements of the score vector $\Psi(x; \hat{\theta})$. If the elements of $\Psi(x; \hat{\theta})$ are bounded when $x$ goes to $\pm \infty$, then the corresponding $IF(x; \hat{\theta})$ will also become bounded (Hampel et al. 1986; Huber 1984).

Now consider the individual score functions corresponding to the parameters $\mu, \sigma, \varepsilon, c$ and $k$ of ESBIII. By taking the partial derivatives of $-\log f(x; \theta)$ with respect to the related parameters, the corresponding score functions are obtained as follows:

$$\psi_\mu(x) = \frac{c + 1}{x} - \frac{c(k + 1)z^{-c}}{x(1 + z^{-c})},$$

$$\psi_\sigma(x) = c - \frac{c(k + 1)z^{-c}}{1 + z^{-c}},$$

$$\psi_\varepsilon(x) = \frac{(c + 1)z}{x} - \frac{c(k + 1)z^{-c+1}}{x(1 + z^{-c})},$$

$$\psi_c(x) = \frac{1}{c} - \log(z) + \frac{(k + 1)z^{-c+1} \log(z)}{1 + z^{-c}},$$

$$\psi_k(x) = \frac{1}{k} - \log(1 + z^{-c}),$$

where $z = \frac{s(x)x}{1 + s(x)x}$.

We examine whether or not these score functions are bounded when $x$ goes to $\pm \infty$. Therefore, we will imply that ML estimators are robust if the score functions are bounded. For ESBIII, these limits are $\lim_{x \to \pm \infty} \psi_\mu(x) = 0$, $\lim_{x \to \pm \infty} \psi_\sigma(x) = c$, $\lim_{x \to \pm \infty} \psi_\varepsilon(x) = -\infty$, $\lim_{x \to \pm \infty} \psi_k(x) = 1/k$ and $\lim_{x \to \pm \infty} \psi_c(x) = \pm \frac{c+1}{1+c}$. Here, the score function $\psi_c(x)$ is only unbounded. If the parameter $c$ is assumed to be fixed and finite, then the score functions $\psi_\sigma(x)$, $\psi_\varepsilon(x)$ and $\psi_c(x)$ become bounded. It is concluded that the ML estimators of location ($\mu$), scale ($\sigma$), skewness ($\varepsilon$) and shape ($k$) parameters will be locally robust when the other shape parameter $c$ remains as
fixed and finite, since it is sufficient to guarantee the boundedness of IF for ML estimators. Two shape parameters in the distribution proposed by Arslan and Genç (2009) must remain as fixed and finite to guarantee the robustness of the ML estimators for location, scale and skewness parameters in their distributions. However, for ESBIII distribution, estimating the one shape parameter, which is $k$, along with location, scale and skewness parameters does not cause to non-robustness for ML estimators. This may be regarded as an advantage over rival distributions, such as epsilon-skew generalized $T$ distribution from Arslan and Genç (2009) and Gómez et al. (2007) and its special forms. Estimating the shape parameter $c$ along with the remainder parameters $\mu$, $\sigma$, $\varepsilon$ and $k$ causes unboundedness of the corresponding IF. Therefore, the ML estimators lose their robustness to the outlying observations. However, if the data set modelled with ESBIII distribution contains relatively few observations which are far from the bulk of data, the ML estimators of all the parameters of ESBIII may remain to be robust even if the $\psi_c$ is unbounded, because the unboundedness of $\psi_c$ is based on the theory. However, for practical purposes, the local robustness of all ML estimators can be guaranteed.

We draw plots of the score functions for certain specific values of parameters as shown in Fig. 6. These plots demonstrate the behaviour of score functions when $x$ goes to infinity. It is noted that the score functions are not defined at point zero, since the pdf has a singularity at zero.
4.2 Tail behaviour of ESBIII

In order to investigate the tail behaviour of ESBIII distribution, we first give preliminaries regarding the property of heavy-tailedness. Note that both of right and left tails of the distribution have to be investigated separately, since ESBIII distribution is defined on the real line. However, investigating only one tail will be sufficient for an evaluation of the tail behaviour of ESBIII, because it is a reflection of (positive definite) BIII distribution to the negative side of the real line. Therefore, we will only consider the right (positive) tail behaviour. The definitions of heavy-tailedness and necessary tools are given in this manner.

**Definition 4.1** A distribution function $F$ is said to be (right) heavy-tailed if, and only if, the moment generating function related with $F$ is infinite; that is,

$$
\int_{\mathbb{R}} \exp(tx) F(dx) = \infty, \quad \text{for all } t > 0.
$$

Sometimes, the condition in Definition 4.1 may be intractable. Then, the alternative theorem is used to indicate the heavy tailedness of ESBIII.

**Definition 4.2** Let $F$ be a distribution function. Then, $R = -\ln(1 - F)$ is called as the hazard function related with $F$. If the hazard function $R$ is differentiable, then its derivative $r(x) = \frac{d}{dx} R(x)$ is called as the hazard rate.

**Theorem 4.1** Let $F$ and $R$ be a distribution function and the hazard function related with $F$, respectively. Then the following assertions are equivalent:

(i) $F$ is heavy-tailed.

(ii) $\lim_{x \to \infty} R(x)/x = 0$.

If the function $r$ has a continuous hazard rate $r$ on $(0, \infty)$, then (ii) in Theorem 4.1 reduces to $\lim_{x \to \infty} r(x) = 0$ (Embrechts et al. 1997; Markovich 2007).

The hazard rate $r$ of BIII distribution is

$$
 r_{BIII}(x) = \frac{ck(1 + x^{-c})^{-(k+1)}}{x^{c+1}(1 - (1 + x^{-c})^{-k})}.
$$

Let us find $\lim_{x \to \infty} r_{BIII}(x)$. In Eq. (4.9), it should be noted that the parts $(1 + x^{-c})^{-(k+1)}$ and $(1 + x^{-c})^{-k}$ are alike cdf of BIII distribution. Since these parts behave as a distribution function, the maximum value will go to $1$ for $(1 + x^{-c})^{-(k+1)}$ and $0$ for $1 - (1 + x^{-c})^{-k}$ as $x \to \infty$. Then, the denominator of $r_{BIII}$ goes to infinity as $x \to \infty$ even if $1 - (1 + x^{-c})^{-k}$ goes to zero, because the magnitude of $x^{c+1}$ is greater than that of $1 - (1 + x^{-c})^{-k}$. Therefore, the function $r_{BIII}$ goes to zero when $x \to \infty$. Consequently, BIII is heavy-tailed distribution.
The hazard rate $r$ of ESBIII for $x \geq 0$ is

$$r_{ESBIII}(x) = \frac{ck(1 + \varepsilon)^c \left[ 1 + (1 + \varepsilon)x^{-c} \right]^{-(k+1)}}{x^{c+1} \left( 1 - \left[ 1 + (1 + \varepsilon)x^{-c} \right]^{-k} \right)}.$$  

(4.10)

Let us find $\lim_{x \to \infty} r_{ESBIII}(x)$. Since the Eq. (4.10) is a rescaled version of the Eq. (4.9) with $(1 + \varepsilon)^c$, $\lim_{x \to \infty} r_{ESBIII}(x) = 0$. Therefore, ESBIII for $x \geq 0$ is right heavy tailed distribution. Similarly, the hazard rate $r$ for $x < 0$, as a mirror or reflection of BIII to the negative side of the real line, goes to zero as $x \to -\infty$. Therefore, ESBIII for $x < 0$ is left heavy-tailed distribution. Consequently, ESBIII is heavy-tailed distribution.

5 Real data application

Two real data sets will be used to illustrate the modelling capability and the robustness property of the proposed distribution. We will also compare distributions that are in the same class, as well as (explicit) analytical expressions for their cdfs. Unimodal and bimodal distributions are considered. One of these is a bimodal extension of generalized gamma called BEGG($\alpha$, $\beta$, $\delta_0$, $\delta_1$, $\eta$, $\varepsilon$, $\mu$, $\sigma$) in Çankaya et al. (2015). The special case of BEGG is the epsilon skew Laplace (ESL). The special cases of BEGG can be found in Çankaya et al. (2015) and the generalized form of BEGG in Çankaya (2018). The generalized $T$ in the sense of epsilon skew form called ESGT($p$, $q$, $\varepsilon$, $\mu$, $\sigma$) in Arslan and Genç (2009), the alpha-skew Laplace called ASL($\alpha$, $\mu$, $\sigma$) in Shams and Alamatsaz (2013), the exponentiated sinh Cauchy called ESC($\lambda$, $\beta$, $\mu$, $\sigma$) in Cooray (2013) and Rathie-Swamee called RS($a$, $b$, $p$, $\mu$, $\sigma$) in Swamee and Rathie (2007) are used to model the data sets. The cdfs of BEGG, ESGT and ESL depend on the special function, which is a deficiency of cdf for computation when they are compared by cdf with explicit analytical expressions. Note that BEGG, ASL, ESC and RS can have the bimodal shape for the correctly chosen values of their parameters. ESGT and ESL have the unimodal shape for all the values of their parameters.

Using an (explicit) analytical expression for cdf is important to avoid problems due to any numerical integration error while getting cdf from pdf. Since we used these kinds of functions, it should be noted that consulting the goodness of fit test (GOFT) statistics and information criteria (IC) can be a good indicator for examining their fitting performance and reasonable comparison among distributions. However, it is known that IC are more accurate than GOFT statistics, because GOFT statistics are based on a hypothetical approach. In other words, they depend on the value of test statistics. In the case of IC, such as Akaike IC (AIC), corrected Akaike IC (cAIC) and Bayesian IC (BIC), they cover all data information. It is also important to note that IC are sensitive to penalized terms $2k$ from $AIC = -2 \log(L) + 2k$, $2k \frac{n}{n-k-1}$ from $cAIC = -2 \log(L) + 2k \frac{n}{n-k-1}$ and $k \log(n)$ from $BIC = -2 \log(L) + k \log(n)$. Here, $k$ and $n$ are the number of parameters estimated and the sample size, respectively (Burnham and
Table 3  Parameter estimates of the different distributions for the microarray data set

| Distributions          | \( \hat{\mu} \) | \( \hat{\sigma} \) | \( \hat{\epsilon} \) |
|------------------------|-----------------|-----------------|-----------------|
| ESBIII \((c = 2.3826, k = 0.7786)\) | – 0.0061 | 0.0770 | 0.0533 |
| ESL \((\alpha = 1)\) | 0.0130 | 0.0770 | 0.0484 |
| ESGT \((p = 1, q = 3/4)\) | 0.0440 | 0.0549 | – 0.2773 |
| BEGG \((\hat{\alpha}, \hat{\beta}, \hat{\delta}_0, \hat{\delta}_1, \hat{\eta})\) | 0.0105 | 0.0701 | 0.2301 |
| ASL \((\hat{\lambda} = 0.0781)\) | 0.0193 | 0.1022 | – |
| ESC \((\hat{\lambda} = 2.3591, \hat{\beta} = 0.944)\) | 0.0051 | 0.1522 | – |
| RS \((\hat{\alpha} = 2.2233, \hat{\beta} = 2.8148, \hat{\rho} = 0.4536)\) | 0.0022 | 0.2563 | – |

For BEGG distribution; \(\hat{\alpha} = 0.9824, \hat{\beta} = 1, \hat{\delta}_0 = 0.0001, \hat{\delta}_1 = 0.4622, \hat{\eta} = 0.7277\)

Table 4  IC and GOFT statistics of distributions with ML estimates: microarray data set

|       | ESBIII | ESL | ESGT | BEGG | ASL | ESC | RS |
|-------|--------|-----|------|------|-----|-----|-----|
| IC    | – 432.49 | – 352.16 | – 108.00 | – 82.51 | – 173.25 | – 158.42 | – 195.24 |
| cAIC  | – 431.95 | – 351.95 | – 107.79 | – 81.19 | – 173.03 | – 158.07 | – 194.71 |
| BIC   | – 418.64 | – 343.84 | – 99.69 | – 60.34 | – 164.93 | – 147.34 | – 181.39 |
| GOFT  |        |      |      |      |      |      |     |
| KS    | 0.0550  | 0.1145 | 0.1222 | 0.0890 | 0.0730 | 0.0807 | 0.0443 |
| AD    | 0.4517  | 3.0152 | 2.2421 | 7.3964 | 1.3028 | 1.6524 | 0.2374 |
| CVM   | 0.0514  | 0.4597 | 0.2492 | 0.1660 | 0.1453 | 0.1992 | 0.0282 |

Anderson 2002). Due to this sensitivity, GOFT statistics, such as Kolmogorov-Smirnov (KS), Anderson-Darling (AD) and Cramér von Mises (CVM) are also computed.

Tables 3, 4, 5 and 6 provide the ML estimates of parameters in corresponding distributions used for fitting data sets, AIC, cAIC and BIC from IC and KS, AD and CVM from GOFT statistics. In addition, Figs. 7 and 8 illustrate the fitting performance of distributions.

Example 1  The univariate cases of the data sets in cDNA microarray were modelled by pdfs in Arslan (2009a, b). These data sets analyzed were also considered by Acıtaş et al. (2013) and Purdom and Holmes (2005). cDNA microarray data sets known as AT-matrix: 1375 genes × 118 drugs, which are available at the web site in Prommier et al. (2018), are considered. In these data sets, one column of gene-drug correlation data, which is known as “sid 487493, erythrocyte band 7 integral membrane protein [5,3:aa045112]”, will be modelled by means of distributions.

From Table 4, when the distributions are ordered according to modelling capacity based on IC from lowest to highest, ESBIII, ESL, RS, ASL, ESC, ESGT and BEGG distributions are given in order for microarray data.

Example 2  A Martin Marietta data set was analysed by Acıtaş et al. (2013), Arslan (2009a, b) and Arslan and Genç (2009). We focus on fitting the dependent variable
Table 5  Parameter estimates of the different distributions for the Martin Marietta data set

| Distributions | \( \hat{\mu} \) | \( \hat{\sigma} \) | \( \hat{\varepsilon} \) |
|---------------|---------------|---------------|---------------|
| ESBIII \((\hat{c} = 2.1241, \hat{k} = 0.7190)\) | −0.0005 | 0.0652 | 0.0564 |
| ESL \((\alpha = 1)\) | −0.0050 | 0.0652 | −0.1434 |
| ESGT \((p = 2.5, q = 0.45)\) | −0.0713 | 0.0739 | 0.6569 |
| BEGG \((\hat{\alpha}, \hat{\beta}, \hat{\delta}_0, \hat{\delta}_1, \hat{\eta})\) | 0.0201 | 0.0615 | 0.0301 |
| ASL \((\hat{\lambda} = 0.0001)\) | 0.0188 | 0.1018 | − |
| ESC \((\hat{\lambda} = 2.5921, \hat{\beta} = 0.9860)\) | 0.0139 | 0.1534 | − |
| RS \((\hat{\alpha} = 2.4635, \hat{\beta} = 2.0296, \hat{\rho} = 0.0165)\) | 0.0035 | 0.2490 | − |

For BEGG distribution; \( \hat{\alpha} = 1.0824, \hat{\beta} = 1.02, \hat{\delta}_0 = 0.2528, \hat{\delta}_1 = 0.009, \hat{\eta} = 0.7105 \)

Table 6  IC and GOFT statistics of distributions with ML estimates: Martin Marietta data set

|            | ESBIII | ESL | ESGT | BEGG | ASL | ESC | RS |
|------------|--------|-----|------|------|-----|-----|-----|
| IC         |        |     |      |      |     |     |     |
| AIC        | −242.74 | −195.23 | −103.97 | −11.57 | −90.43 | −88.42 | −93.71 |
| cAIC       | −241.62 | −194.81 | −103.54 | −8.74  | −90.01 | −87.70 | −92.60 |
| BIC        | −232.26 | −188.95 | −97.68  | 5.19   | −84.15 | −80.05 | −83.24 |
| GOFT       |        |     |      |      |     |     |     |
| KS         | 0.0952  | 0.1435 | 0.1611 | 0.1523 | 0.1234 | 0.1103 | 0.1102 |
| AD         | 0.8099  | 2.3994 | 1.6696 | 3.5113 | 2.0052 | 1.4574 | 0.7477 |
| CVM        | 0.1272  | 0.3861 | 0.2344 | 0.4537 | 0.3129 | 0.2172 | 0.0756 |

in the regression model. This variable is defined to be the excess rate of the Martin Marietta company. It will be modelled by distributions.

From Table 6, when the distributions are ordered according to modelling capacity based on IC and GOFT statistics from lowest to highest, ESBIII, ESL, ESGT, ASL, RS, ESC and BEGG distributions are given in order for the Martin Marietta data.

From Tables and Figures, it can be seen that ESBIII distribution fits better than other distributions; ESL, ESGT, BEGG, ASL, ESC and RS distributions. It can be seen that the parameters \( c \) and \( k \) produce more flexibility when ESBIII distribution is compared with other distributions. The shape parameters \( c \) and \( k \) of ESBIII produce flexible behaviour for both peakedness and tail thickness. Moreover, the interspace between peakedness and tail of ESBIII may also be flexible due to two shape parameters \( c \) and \( k \). Since ESBIII is in a class of heavy-tailed distributions, it can model the data at tail. The bimodality property of ESBIII distribution plays an important role in fitting the data set as well. RS is a competitive distribution when compared with ESBIII distribution having same number of parameters. The bimodality in the Martin Marietta data set could not be fitted by RS distribution, which shows a deficiency of RS. It should also be noted that the analytical expression of RS does not include a skewness parameter for modelling unequal probabilities around location.
Using the fixed values of parameters $p$ and $q$ as tuning constants in robustness is preferred by Arslan and Genç (2009) and Lucas (1997), because the ML estimators of parameters are not robust due to fact that their IF is unbounded. In our case, since the score function for the parameter $c$ is infinite, the IF of the ML estimators of the parameters will be unbounded. We prefer to estimate this, because it can be clearly seen that IC show the competence of our flexible distribution. Using the robustness of ML estimators for the parameters of distribution, IC and/or GOFT statistics representing the modelling capacity of the proposed distribution is suggested. For real data examples, examining whether or not IC are small will be a good indicator with which to approve the modelling capacity of the proposed distribution. Since GOFT statistics and IC give satisfactory results, the parameter $c$ was also estimated. In fact, the IF of the ML estimators for the parameters $\mu$, $\sigma$, $c$, $k$ and $\epsilon$ is unbounded. However, there is an another point that should be taken into account, which is that the modelling capacity of the proposed model is very high when IC are considered. If the parameter $c$ is considered to be fixed, its estimated value would be the best choice.
We observe that the degree of interaction of the shape parameters $c$ and $k$ with ESBIII may be lower than that of the parameters $p$ and $q$ with ESGT. In this sense, it is also noted that the tail parameter $q$ leads to a Laplace distribution for $p = 2$ in ESGT (Arslan and Genç 2009). However, the shape parameters $c$ and $k$ with ESBIII are flexible when compared with ESGT.

6 Conclusions

An extension of BIII distribution to the real line has been proposed. The main goal for introducing the new distribution is that BIII distribution has a heavy tailedness (robustness) property. An extension to the real line already preserves the heavy tailedness property, because ESBIII distribution is only a reflection of BIII distribution defined on the positive side to the negative side. A skewness parameter is also added to the extended BIII to develop the distribution to a skew distribution. The skewness
parameter gives flexibility for modelling data sets that have different tail behaviours on the positive and the negative sides of the real line; that is, on both sides of the location. In addition, the inherited shape parameters $c$ and $k$ provide extra flexibility for modeling data sets that may have unimodal, bimodal, leptokurtic or platykurtic empirical distributions. The values of the skewness and kurtosis coefficients for certain values of the shape parameters also confirm the remarkable flexibility of the newly proposed distribution. Therefore, ESBIII distribution can cope with skewness, heavy-tailedness and bimodality all together by means of only its three parameters. We have studied the robustness of the ML estimators for the parameters of ESBIII distribution and observed that the ML estimators are locally robust as long as the shape parameter $c$ remains as fixed and known. We have also explored the heavy-tailedness of ESBIII distribution and have shown that ESBIII is a heavy-tailed distribution. Some properties, such as the Rényi and Shannon entropies, the $r$th moment and modes, of ESBIII distribution have been provided. Finally, two real data examples were given to illustrate modelling capacity of ESBIII distribution. The results of the real data examples have shown that ESBIII distribution has noticeable superiority over the similar type of distributions for modeling data sets that have bimodality and skewness.

For a future study, comparison could be made between the flexibility performance of Azzalini type skewness of extended Burr III distribution and the epsilon skew extension proposed in this study. In an ongoing paper, Shannon, Tsallis, Rényi and hybrid entropies of distributions (Çankaya and Korbel 2017; Jizba and Korbel 2016) will be used as tools to test what much the degree of heavy-tailedness can be produced by such distributions in the class of heavy-tailedness.

Acknowledgements We thank the anonymous reviewers for their careful reading of our manuscript and their many insightful comments and suggestions. We would like to say our special thanks to the Editorial Board Members. We thank the native speaker who supported us for representation in the formal organization of this paper.

Appendix

Proof of Theorem 2.4

$$\int_{\mathbb{R}} f_{\alpha}(x) \, dx = \int_{-\infty}^{\mu} f_{\alpha}^{-}(x) \, dx + \int_{\mu}^{\infty} f_{\alpha}^{+}(x) \, dx,$$

where

$$f_{-}(x) = \frac{ck}{2\sigma} \left( \frac{-x - \mu}{\sigma(1 - \epsilon)} \right)^{-c+1} \left\{ 1 + \frac{-x - \mu}{\sigma(1 - \epsilon)} \right\}^{-c} \left( \frac{1}{\sigma(1 - \epsilon)} \right)^{-k+1},$$

$$f_{+}(x) = \frac{ck}{2\sigma} \left( \frac{x - \mu}{\sigma(1 + \epsilon)} \right)^{-c+1} \left\{ 1 + \frac{x - \mu}{\sigma(1 + \epsilon)} \right\}^{-c} \left( \frac{1}{\sigma(1 + \epsilon)} \right)^{-k+1}. $$

Now consider the integral $\int_{\mu}^{\infty} f_{\alpha}^{+}(x) \, dx$. This integral is analytically obtained by applying the transformations, $y = \frac{x - \mu}{\sigma(1 + \epsilon)}$, $u = y^{-c}$, $v = 1 + u$ and $z = 1/v$, consecutively. These steps are given as follows.
\[
\int_{-\infty}^{\infty} f_{\alpha}^{\alpha}(x) \, dx = \int_{-\infty}^{\infty} \left[ \frac{ck}{2\sigma} \left( \frac{x - \mu}{\sigma(1 + \varepsilon)} \right)^{-(c+1)} \left\{ 1 + \left( \frac{x - \mu}{\sigma(1 + \varepsilon)} \right)^{-c} \right\}^{-(k+1)} \right] \, dx
\]

where \( \alpha(k - 1/c) + 1/c > 0, \alpha(1 + 1/c) - 1/c > 0 \). Similarly, the other integral is obtained as

\[
\int_{-\infty}^{\infty} f_{\alpha}^{\alpha}(x) \, dx = \left( \frac{ck}{2\sigma} \right) \frac{\sigma(1 - \varepsilon)}{c} B\left( \alpha(k - 1/c) + 1/c, \alpha(1 + 1/c) - 1/c \right),
\]

The Shannon entropy is obtained by finding \( \lim_{\alpha \to 1} J_{\mathbb{R}}(\alpha) \). One can apply L’Hôpital rule to find this limit. Thus,

\[
\lim_{\alpha \to 1} J_{\mathbb{R}}(\alpha) = (k + 1)\psi(k + 1) + (1 + 1/c)\gamma + (1/c - k)\psi(k) + \log \left( \frac{2\sigma}{ck} \right),
\]

which is the Shannon entropy. Here \( B(\cdot) \) is beta function, \( \psi(\cdot) \) is digamma function and \( \gamma \) is Euler constant.

\[ \square \]

References

Abd-Elfattah AM, Alharbey AH (2012) Bayesian estimation for Burr distribution type III based on trimmed samples. ISRN Appl Math 250393:1–18
Abdulah E, Elsalloukh H (2013) Analyzing skewed data with the epsilon skew gamma distribution. J Stat Appl Prob 2(3):195–202
Açıtaş S, Kasap P, Şenoğlu B, Arslan O (2013) One-step M-estimators: Jones and Faddy’s skewed t-distribution. J Appl Stat 40(7):1–15
Ali A, Hasnain SA, Ahmad M (2015) Modified Burr III distribution: properties and applications. Pak J Stat 31(6):697–708

Springer
Andrade BB, Rathie PN (2016) Fitting asymmetric bimodal data with selected distributions. J Stat Comput Simul 86(16):3205–3224
Arellano-Valle RB, Gomez HW, Quintana FA (2005) Statistical inference for a general class of asymmetric distributions. J Stat Plan Inference 128(2):427–443
Arellano-Valle RB, Cortés MA, Gómez HW (2010) An extension of the epsilon-skew-normal distribution. Commun Stat Theory Methods 39(3):912–922
Arslan O (2009a) Maximum likelihood parameter estimation for the multivariate skew slash distribution. Stat Probab Lett 79(20):2158–2165
Arslan O (2009b) An alternative multivariate skew Laplace distribution: properties and estimation. Stat Pap 51(4):865–887
Arslan O, Genç AI (2009) The skew generalized t distribution as the scale mixture of a skew exponential power distribution and its applications in robust estimation. Statistics 43(5):481–498
Azimi R, Yaghmaei F (2013) Bayesian estimation for the burr type III distribution under type II Doubly censored data. Int J Adv Stat Probab 1(1):1–3
Azzalini A (1985) A class of distributions which includes the normal ones. Scand J Stat 12(2):171–178
Azzalini A, Capitanio A (2003) Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. J R Stat Soc Ser B 65(2):367–389
Bolfarine H, Martinéz-Flórez G, Salinas HS (2013) Bimodal symmetric-asymmetric power-normal families. Commun Stat Theory Methods 42(2):259–276
Box GEP, Tiao GC (1973) Bayesian inference in statistical analysis. Addison-Wesley, Reading
Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. Springer, New York
Burr IW (1942) Cumulative frequency functions. Ann Math Stat 13(2):215–232
Burr IW (1973) Parameters for a general system of distributions to match a grid of $\alpha_3$ and $\alpha_4$. Commun Stat Theory Methods 2(1):1–21
Burr IW, Cislak PJ (1968) On a general system of distributions I. Its curve-shape characteristics II. The sample median. J Am Stat Assoc 63(322):627–635
Çankaya MN (2018) Asymmetric bimodal exponential power distribution on the real line. Entropy 20(23):1–19
Çankaya MN, Korbel J (2017) On statistical properties of Jizba–Arimitsu hybrid entropy. Physica A 475:1–10
Çankaya MN, Bulut YM, Doğru FZ, Arslan O (2015) A bimodal extension of the generalized gamma distribution. Rev Colomb Estad 38(2):353–370
Cooray K (2013) Exponentiated sinh Cauchy distribution with applications. Commun Stat Theory Methods 42(21):3838–3852
Dexter OC (2015) Some skew-symmetric distributions which include the bimodal ones. Commun Stat Theory Methods 44(3):554–563
Donatella V, Van Dorp JR (2013) On a bounded bimodal two-sided distribution fitted to the old-Faithful geyser data. J Appl Stat 40(9):1965–1978
Elsalloukh H (2008) The epsilon-skew Laplace distribution. In: The proceedings of the American statistical association, Biometrics Section, Denver Colorado
Elsalloukh H, Guardiola JH, Young M (2005) The epsilon-skew exponential power distribution family. Far East J Theor Stat 17(1):97–107
Embrechts P, Kluppelberg C, Mikosch T (1997) Modelling extremal events for insurance and finance. Springer, Heidelberg
Genç AI (2013) A skew extension of the slash distribution via beta-normal distribution. Stat Pap 54(2):427–442
Genton MG (2004) Skew-elliptical distributions and their applications: a journey beyond normality. Chapman & Hall/CRC, London
Gómez HW, Torres FJ, Bolfarine H (2007) Large-sample inference for the epsilon-skew-t distribution. Commun Stat Theory Methods 36(1):73–81
Gove JH, Ducey MJ, Leak WB, Zhang L (2008) Rotated sigmoid structures in managed uneven-aged northern hardwork stands: a look at the Burr Type III distribution. Foresty 81(2):161–176
Gui W (2014) A generalization of the slashed distribution via alpha skew normal distribution. Stat Methods Appl 23(1):1–17
Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA (1986) Robust statistics: the approach based on influence functions. Wiley, New York
Hao Z, Singh VP (2009) Entropy-based parameter estimation for extended three-parameter Burr III distribution for low-flow frequency analysis. Trans ASABE 52(4):1193–1202
Hassan MY, El-Bassioni MY (2016) Bimodal skew-symmetric normal distribution. Commun Stat Theory Methods 45(5):1527–1541
Huber PJ (1984) Finite sample breakdown of M- and P-estimators. Ann Stat 12(1):119–126
Jamalizadeh A, Arabpour AR, Balakrishnan N (2011) A generalized skew two-piece skew-normal distribution. Stat Pap 52(2):431–446
Jizba P, Korbel J (2016) On q-non-extensive statistics with non-Tsallisian entropy. Physica A 444:808–827
Jones MC (2009) Kumaraswamy’s distribution: a beta-type distribution with some tractability advantages. Stat Methodol 6:70–81
Jones MC, Faddy MJ (2003) A skew extension of the t-distribution, with applications. J R Stat Soc Ser B 65:159–175
Lindsay SR, Wood GR, Woollons RC (1996) Modelling the diameter distribution of forest stands using the Burr distribution. J Appl Stat 23(6):609–620
Lucas A (1997) Robustness of the student t based M-estimator. Commun Stat Theory Methods 26(5):1165–1182
Markovich N (2007) Nonparametric analysis of univariate heavy-tailed data: research and practice, 2nd edn. Wiley, New York
Mineo AM, Ruggieri M (2005) A software tool for the exponential power distribution: the normalp package. J Stat Softw 12(4):1–24
Mudholkar GS, Hutson AD (2000) The epsilon-skew-normal distribution for analyzing near-normal data. J Stat Plan Inference 83(2):291–309
Prommier Y, Reinhold W, Sunshine M, Varma S (2018) Genomics and bioinformatics group. http://discover.nci.nih.gov. Accessed 10 May 2018
Purdom E, Holmes SP (2005) Error distribution for gene expression data. Stat Appl Genet Mol Biol 4(1):2194–6302
Rathie PN, Silva P, Olinto G (2016) Applications of skew models using generalized logistic distribution. Axioms 5(10):1–26
Rêgo LC, Cintra RJ, Cordeiro GM (2012) On some properties of the beta normal distribution. Commun Stat Theory Methods 41(20):3722–3738
Rényi A (1961) On measures of entropy and information. ISRN Applied Mathematics. Hungarian Academy of Sciences, Budapest Hungary, Budapest
Rodriguez RN (1977) A guide to the Burr type XII distributions. Biometrika 64(1):129–134
Shams HS, Alamatsaz MH (2013) Alpha–Skew–Laplace distribution. Stat Probab Lett 83(3):774–782
Shannon CE (1961) Two-way communication channels. In: Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, volume 1: contributions to the theory of statistics, The Regents of the University of California
Shao Q, Chen YD, Zhang L (2008) An extension of three-parameter Burr III distribution for low-flow frequency analysis. Comput Stat Data Anal 52(3):1304–1314
Subbotin MT (1923) On the law of frequency of errors. Mat Sb 31:296–301
Swamee PK, Rathie PN (2007) Invertible alternatives to normal and lognormal distributions. J Hydrol Eng 12(2):218–221
Tadikamalla PR (1980) A look at the Burr and related distributions. Int Stat Rev 48(3):337–344
Venegas O, Hugo SS, Gallardo DI, Bolfarine H, Gómez HW (2017) Bimodality based on the generalized skew-normal distribution. J Stat Comput Simul 88(1):156–181
Wang FK, Keats JB, Zimmer WJ (1996) Maximum likelihood estimation of the Burr XII parameters with censored and uncensored data. Microelectron Reliab 36(3):359–362
Wingo DR (1983a) Maximum Likelihood Methods for fitting the Burr type XII distribution of life test data. Bio Met J 25(1):77–84
Wingo DR (1983b) Estimating the location of the Cauchy distribution by numerical global optimization. Commun Stat-Simul C 12(2):201–212
Zimmer WJ, Keats JB, Wang FK (1998) The Burr XII distribution in reliability analysis. J Qual Technol 30(4):386–394
Zoraghi N, Abbasi B, Niaki STA, Abdi M (2012) Estimating the four parameters of the Burr III distribution using a hybrid method of variable neighborhood search and iterated local search algorithms. Appl Math Comput 218(19):9664–9675
Publisher's Note  Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Mehmet Niyazi Çankaya\textsuperscript{1,2} · Abdullah Yalçinkaya\textsuperscript{3} · Ömer Altındağ\textsuperscript{4} · Olcay Arslan\textsuperscript{3}

Mehmet Niyazi Çankaya
mehmet.cankaya@usak.edu.tr

Ömer Altındağ
omer.altindag@bilecik.edu.tr

Olcay Arslan
oarslan@ankara.edu.tr

\textsuperscript{1} Applied Sciences School, Department of International Trading, Uşak, Turkey
\textsuperscript{2} Department of Statistics, Faculty of Arts and Sciences, Uşak University, Uşak, Turkey
\textsuperscript{3} Department of Statistics, Faculty of Science, Ankara University, Ankara, Turkey
\textsuperscript{4} Department of Statistics, Faculty of Arts and Sciences, Bilecik Şeyh Edebali University, Bilecik, Turkey