Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 76(1) 2022 31–41

On a Class of Bimodal Distributions and their Applications in Modelling Bimodal Error Data

Anjana V., P. Yageen Thomas and Manoj Chacko

University of Kerala, Thiruvananthapuram

Received 06 April 2022; Revised 02 June 2022; Accepted 06 June 2022

SUMMARY

A new family of bimodal distributions is introduced in this paper with an objective of using them for modelling error data sets. A new class of statistics arising from a symmetric distribution is proved to have distributions belonging to the family of the bimodal distributions introduced in this work. The information matrix is derived after addressing the problem of obtaining maximum-likelihood estimates for the parameters of generalized bimodal distribution. A simulation study is conducted to evaluate the properties of maximum likelihood estimators. The applications of the results in building bimodal distributions for some real life data sets are also illustrated.

Keywords: Error data; Symmetric distributions; Bimodal distributions; Maximum likelihood estimate; Ordered density value induced statistics.

1. INTRODUCTION

Symmetry is one class of patterns occurring in nature wherein we could observe near repetition of the pattern, either by reflection or by rotation. The body of most of the multi-cellular organisms exhibit some form of symmetry. Similarly, measurements made on several biological variables follow statistical distributions which are symmetric in form. This makes "Symmetry in Biology" a largely discussed and studied subject of interest. The distribution of errors observed on measurements of orbit of heavenly bodies was observed as normal by Gauss (1857). Recently, Rao and Gupta (1989) have narrated how normal distribution is derived by Hersched's hypothesis on errors. They also described how normal distribution can be derived using Hagen's hypothesis on errors. The third Hagen's hypothesis states that each component of error has an equal chance of being positive or negative. This makes a deduction that the class A of all error models satisfying third Hagen's hypothesis must be symmetrically distributed about zero. It is to be noted that, if X is a random variable with expected value μ, then observations on $X - \mu$ for known value of μ also constitute an error data. If the third hypothesis due to Hagen is seen satisfied on the above data, then in the problem of modelling a distribution to this data, we can limit our search for choosing an appropriate model from the family **A.** Though all models belonging to **A** are symmetrically distributed about zero, they need not have a unique mode. In a recent investigation, the authors come across data sets on errors which have two modes, of which one is positive and other is negative, whereas they are equidistant from the centre. This motivates the authors of this paper to deal with new bimodal distributions and to illustrate their applications to real life problems.

Eisenberger (1964) discussed about a variety of bimodal distributions arising out of a mixture of two normal distributions. Prasad (1954) as well as Sarma *et al.* (1990) discussed about bimodal distributions whose densities are similar to that of mixture of normal distributions. For a discussion on bimodal exponential power distribution see, Hassan and Hijazi (2010) and for details on bimodal skew-symmetric

Corresponding author: P. Yageen Thomas
E-mail address: yageenthomas@gmail.com

normal distribution see, Hassan and El-Bassiouni (2013). However, for many of newly emerging error data sets, the above mentioned bimodal distributions fail to provide the desired level of suitability. This makes it necessary to develop more and more new families of bimodal distributions which are capable of providing better suitability as models to error data sets arising from wide range of conditions. We organize the presentation of results as given below.

We introduced a new family of bimodal distributions that are symmetric about zero and can be used to model error datasets in section 2. A new class of statistics named "Ordered Density Value Induced Statistics" arising from symmetric distributions and follows the generalized bimodal distribution is discussed in section 3. In section 4, the estimation of parameters in the generalized bimodal distribution with a symmetric baseline distribution by the method of maximum likelihood is illustrated. A simulation study to describe the closeness between the estimates used for modelling bimodal distribution with the true values of the parameters is also carried out in section 5. Section 6 is devoted to illustrate the applications of the newly generated bimodal distributions in modelling some real life data sets.

2. ABOUT A GENERALIZED BIMODAL DISTRIBUTION WITH SYMMETRIC PROPERTY

Suppose F(y) is an absolutely continuous cumulative distribution function (cdf) with probability density function (pdf) $f(y)$. Throughout this paper we assume that $f(y)$ is symmetric about zero with the property that $f(y)$ is monotone increasing over (−∞, 0) and monotone decreasing over [0, ∞). We may write \$ to denote the above set of assumptions on the considered distribution. Now for α ≥ 0, β > 0, consider the function

$$f_{\alpha,\beta}(y)=\begin{cases}\dfrac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)}[2F(y)]^{\alpha}[1-2F(y)]^{\beta}f(y), -\infty<y<0,\\[2ex] \dfrac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)}[2(1-F(y))]^{\alpha}[2F(y)-1]^{\beta}f(y), 0\le y<\infty.\end{cases} \quad (1)$$

For convenience, we write

$$f^{(1)}_{\alpha,\beta}(y)=\frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)}[2F(y)]^{\alpha}[1-2F(y)]^{\beta}f(y)$$

and

$$f^{(2)}_{\alpha,\beta}(y)=\frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)}[2(1-F(y))]^{\alpha}[2F(y)-1]^{\beta}f(y).$$

Hence (1) can be also written as

$$f_{\alpha,\beta}(y)=\begin{cases}f^{(1)}_{\alpha,\beta}(y), -\infty<y<0,\\ f^{(2)}_{\alpha,\beta}(y), 0\le y<\infty.\end{cases}$$

Clearly, both $f^{(1)}_{\alpha,\beta}(y)$ and $f^{(2)}_{\alpha,\beta}(y)$ are non-negative in their respective regions. Further, we can easily verify $\int_{-\infty}^{0} f^{(1)}_{\alpha,\beta}(y)dy=\int_{0}^{\infty} f^{(2)}_{\alpha,\beta}(y)dy=\frac{1}{2}$.

This proves that $\int_{-\infty}^{\infty} f_{\alpha,\beta}(y)dy=1$ and hence it follows that $f_{\alpha,\beta}(y)$ is a pdf. From the formulated assumptions \$ on $f(y)$, one can easily verify that $\lim_{y\to-\infty} f_{\alpha,\beta}{}^{(1)}(y)=\lim_{y\to 0} f_{\alpha,\beta}{}^{(1)}(y)=0$ and $f_{\alpha,\beta}{}^{(1)}(y)>0$ for all $-\infty<y<0$. These properties of $f_{\alpha,\beta}{}^{(1)}(y)$ make us to find a mode for $f_{\alpha,\beta}{}^{(1)}(y)$ in the interval $(-\infty,0)$. Similarly, the properties of $f_{\alpha,\beta}{}^{(2)}(y)$ given by $\lim_{y\to 0} f_{\alpha,\beta}{}^{(2)}(y)=\lim_{y\to\infty} f_{\alpha,\beta}{}^{(2)}(y)=0$ and $f_{\alpha,\beta}{}^{(2)}(y)>0$ for the interior points of $[0,\infty)$ make us to find another mode of $f_{\alpha,\beta}(y)$ in the interval $0\le y<\infty$. From (1) we can easily verify that for any $y\in(0,\infty)$, we have $-y\in(-\infty,0)$ and further $f_{\alpha,\beta}(-y)=f_{\alpha,\beta}(y)$. Clearly the above properties of the pdf $f_{\alpha,\beta}(y)$ make us to conclude that $f_{\alpha,\beta}(y)$ is a bimodal symmetric density function generated from the basic distribution with cdf $F(y)$ and pdf $f(y)$. Thus, the probability distribution defined by the pdf $f_{\alpha,\beta}(y)$ as given in (1) is called a generalized bimodal symmetric distribution (GBSD) determined by the baseline symmetric distribution with cdf $F(y)$ and pdf $f(y)$. The proof as given above on $f_{\alpha,\beta}(y)$ as a bimodal distribution makes us to state the following theorem.

Theorem 2.1. Suppose *F(x)* is the cdf satisfying the set of assumptions \$, then for real constants α ≥ 0, β > 0 the following function

$$f_{\alpha,\beta}(y)=\begin{cases}\dfrac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)}[2F(y)]^{\alpha}[1-2F(y)]^{\beta}f(y), -\infty<y<0,\\ \dfrac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)}[2(1-F(y))]^{\alpha}[2F(y)-1]^{\beta}f(y), 0\le y<\infty,\end{cases} \quad (2)$$

represents the pdf of a bimodal symmetric distribution in which *F(x)* is the cdf and *f (x)* is the corresponding pdf of the baseline distribution used for defining the bimodal density.

Note 2.1. One can verify that (1) is a pdf even for −1 < α < 0 and −1 < β ≤ 0. But in those ranges of α and β, $f_{\alpha,\beta}(y)$ need not be bimodal. For example if −1 < β ≤ 0, then all functions $[2F(y)]^{\alpha}$, $[1-2F(y)]^{\beta}$ and $f(y)$ in the range [−∞, 0) are individually increasing so that the function $f^{(1)}_{\alpha,\beta}(y)$ formed through the product of three such functions is again increasing. Hence no mode exists for $f_{\alpha,\beta}(y)$ in the interval (−∞, 0). Similarly, the condition −1 < α < 0 also does not guarantee for the existence for two modes for $f_{\alpha,\beta}(y)$. It is further trivial to note that $f_{\alpha,\beta}(y)=f(y)$ for α = β = 0, which is only unimodal.

There is extensive literature available in generating new distributions from a given baseline distribution. Some varieties of such distributions are: (1) Marshall and Olkin (1997) family of distributions, (2) T−X family of distributions as defined in Alzaatreh *et al.* (2013), (3) T-transmuted X family of distributions as defined in Jayakumar and Babu (2017) and so on. A family of distributions defined from any baseline distribution whose pdf is similar in form to the pdf of the distribution defined in (1) is called beta generalized family of distributions. In particular if *F(x)* and $f(x)$ are the cdf and pdf respectively of a baseline distribution, then for a>0, b>0 the pdf $g_{a,b}(x)$ of the beta generalized distribution is given by

$$g_{a,b}(x)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}[F(x)]^{a-1}[1-F(x)]^{b-1}f(x), -\infty<x<\infty.$$

Some recently discussed distributions belonging to beta generalized family of distributions are beta-logistic distribution (see, Thomas and Priya, 2015) and symmetric beta-Cauchy distribution (see, Thomas and Priya, 2016) and so on. More specifically any beta generalized distribution with a symmetric baseline distribution having identically same value for the shape parameters a and b is again a symmetric distribution.

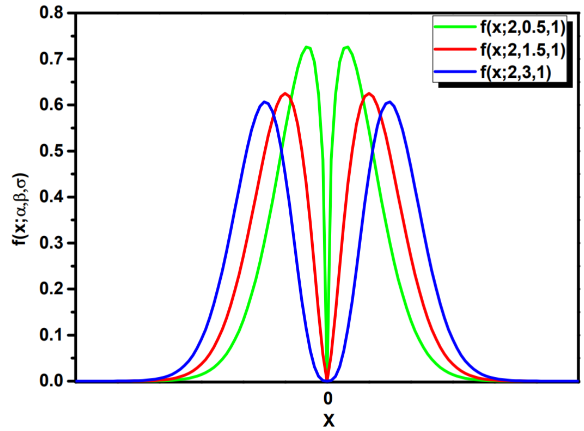

Fig. 1. Generalized bimodal densities with standard normal as the baseline distribution for parameter values: α =2, σ=1 and β =0.5, 1.5 and 3

All these symmetric distributions can serve as baseline distributions to define different varieties of GBSD.

Thus, unlike the existing bimodal distributions available in the present literature which are usable as error models, the GBSD as given in (3) is so extensive, that diverse and broad types of error data sets arising from many real life problems are capable of being modelled by GBSD with comparatively good fitness measures.

We can easily observe that a more general form of bimodal distribution incorporating a scale parameter in (1) is given by the pdf

$$f(x;\alpha,\beta,\sigma)=\begin{cases}\dfrac{\Gamma(\alpha+\beta+2)}{\sigma\Gamma(\alpha+1)\Gamma(\beta+1)}[2F(\tfrac{x}{\sigma})]^{\alpha}[1-2F(\tfrac{x}{\sigma})]^{\beta}f(\tfrac{x}{\sigma}), -\infty<x<0,\\ \dfrac{\Gamma(\alpha+\beta+2)}{\sigma\Gamma(\alpha+1)\Gamma(\beta+1)}[2(1-F(\tfrac{x}{\sigma}))]^{\alpha}[2F(\tfrac{x}{\sigma})-1]^{\beta}f(\tfrac{x}{\sigma}), 0\le x<\infty,\end{cases} \quad (3)$$

where $\alpha\ge 0$ and $\beta>0$ are the shape parameters and σ is a scale parameter. Clearly, $f(x;\alpha,\beta,\sigma)$ also belongs to the family **A** of error distributions corresponding to the baseline distribution with cdf $F(\frac{x}{\sigma})$ and pdf $\frac{1}{\sigma}f(\frac{x}{\sigma})$.

Corresponding to the following baseline distribution with pdf

$$f_1(x,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}, -\infty<x<\infty, (Normal\ \ distribution), \quad (4)$$

for convenience we may also write $f_1(x,\sigma)=\frac{1}{\sigma}g_1(\frac{x}{\sigma})=N(x,\sigma)$. Similarly, if we have

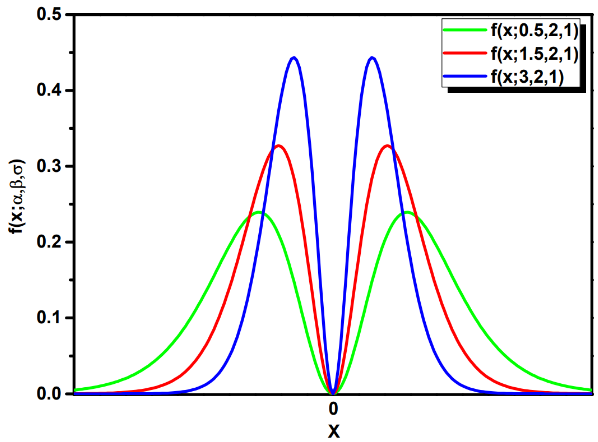

Fig. 2. Generalized bimodal densities with logistic as the baseline distribution for parameter values: β=2, σ=1 and α=0.5, 1.5 and 3

$$f_2(x,\sigma)=\frac{1}{\sigma}\frac{e^{-\frac{x}{\sigma}}}{(1+e^{-\frac{x}{\sigma}})^2},-\infty<x<\infty,\ (Logistic\ \ distribution), \tag{5}$$

then the above density may also be written as $f_2(x,\sigma)=\frac{1}{\sigma}g_2(\frac{x}{\sigma})=L(x,\sigma)$. For the baseline distribution with pdf

$$f_3(x,\sigma)=\frac{1}{\pi\sigma}\frac{1}{1+(\frac{x}{\sigma})^2},-\infty<x<\infty\ (Cauchy\ \ distribution), \tag{6}$$

an equivalent representation is $f_3(x,\sigma)=\frac{1}{\sigma}g_3(\frac{x}{\sigma})=C(x,\sigma)$. For some specified values of α and β with σ=1, the graphical representation of the respective bimodal distributions as defined in (3) generated using MATHEMATICA (ver. 11.3) software are given in figures 1, 2 and 3 respectively.

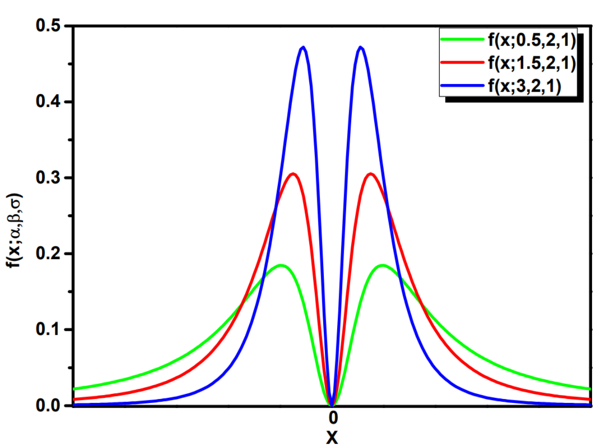

Fig. 3. Generalized bimodal densities with Cauchy as the baseline distribution for parameter values: β=2, σ=1 and α=0.5, 1.5 and 3

3. SOME STATISTICS WHICH ARE DISTRIBUTED AS GBSD

Suppose $F(x)$ is an absolutely continuous cdf having a pdf $f(x)$ which is symmetric about zero. Let $f(x)$ be such that $f(x)$ is monotone increasing and decreasing in the intervals (−∞, 0) and [0, ∞) respectively. We assume that for any random sample $X_1, X_2, ..., X_n$ of observations drawn from the distribution with pdf $f(x)$, an ordered arrangement of the density values $f(X_1), f(X_2),..., f(X_n)$ can be made without any ambiguity. Some examples of pdfs satisfying the above assumptions are the pdfs defined in (4), (5) and (6) for known or unknown values of σ. We may call the assumptions stated above by the symbol \$. Under the conditions \$, if we arrange the observations of the sample as $X_{1!n}, X_{2!n}, \cdots, X_{n!n}$ such that $f(X_{1!n}) \le f(X_{2!n}) \le \cdots \le f(X_{n!n})$, then this newly arranged observations are known as "Ordered Density Value Induced" (ODVI) statistics. In particular $X_{r!n}$ is called the r[th] ODVI statistic. Now we can prove the following theorem.

Theorem 3.1. Let $X_{1!n}, X_{2!n}, \cdots, X_{n!n}$ be the ODVI statistics of a random sample of size n drawn from a distribution satisfying the conditions *\$* with pdf $f(x)$ and cdf *F(x).* Then for positive integer *r* such that $1 \le r \le n$, the pdf $f_{r!n}(x)$ *of* $X_{r!n}$ is given by

$$f_{r!n}(x)=\begin{cases} d(n,r)[2F(x)]^{r-1}[1-2F(x)]^{n-r}f(x), & -\infty<x<0,\\ d(n,r)[2(1-F(x))]^{r-1}[2F(x)-1]^{n-r}f(x), & 0\le x<\infty,\end{cases}$$

where $d(n,r)=\dfrac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)}$. (7)

Proof: By the definition of the pdf of $X_{r!n}$, we have

$$f_{r!n}(x)=\lim_{\Delta x\to 0}\frac{P(x<X_{r!n}\le x+\Delta x)}{\Delta x}. \tag{8}$$

But the event $x < X_{r!n} \le x + \Delta x$ for the case when $x < 0$ makes a partition of **R**, the real line into three sets $E_1 = (-\infty, x]\cup(-x, \infty)$, $E_2 = (x, x + \Delta x]$, $E_3 = (x + \Delta x, -x]$ and assign on them r−1, 1, n−r observations, respectively. The probability that an observation lies in the above partition sets in the limit as Δx → 0 are 2F(x), $f(x)dx$ and 1−2F(x) respectively. Then using multinomial probability law on the numerator of (8) and applying limit as Δx → 0, we obtain

$$f_{r!n}(x) = \frac{n!}{(r-1)!(n-r)!}[2F(x)]^{r-1}[1-2F(x)]^{n-r} f(x), -\infty < x < 0. \tag{9}$$

Similar portraying of (8) for $0 \le x < \infty$ and applying limit as $\Delta x \to 0$, lead us to obtain

$$f_{r!n}(x) = \frac{n!}{(r-1)!(n-r)!}[2(1-F(x))]^{r-1}[2F(x)-1]^{n-r} f(x), 0 \le x < \infty. \tag{10}$$

Clearly (9) and (10) together establishes the theorem.

If we put $\alpha = r-1$ and $\beta = n-r$ for $r \ne n$, then the pdfs $f_{r!n}(x)$ for $r = 1,2,...,n-1$ are all members of GBSD as given in (1). Clearly, for $r = n$, $f_{r!n}(x)$ becomes unimodal. Thus, we conclude that there is close connection between GBSD and the sampling distributions of ODVI statistics.

4. ESTIMATION OF PARAMETERS OF GBSD

Let $X_1, X_2, ..., X_n$ be a random sample of size n drawn from the distribution with pdf $f(x; \alpha, \beta, \sigma)$ which is as given in (3). Then writing the associated likelihood and dealing with the likelihood equations for getting the maximum likelihood estimates (MLE's) of α, β and σ is somewhat different from familiar likelihoods as the form of the density of the observations which are negative is different from those of positive observations. However, in the following theorem we establish that the MLE's of α, β and σ can be determined just by attempting on the MLE's of the parameters α, β and σ of the folded form of GBSD about zero.

Theorem 4.1. *Let* $X_1, X_2, ..., X_n$ be a random sample of size n drawn from the GBSD with pdf $f(x; \alpha, \beta, \sigma)$ defined in (3). *Let* $Z_i = |X_i|, i = 1, 2, ..., n.$ Then the MLE's of the parameters *α, β and σ* involved in (3) are the same as the MLE's of those parameters based on independent random variables $Z_1, Z_2, ..., Z_n$ each distributed identically as the distribution with pdf obtained by folding *f (x; α, β, σ)* about *x = 0.*

Proof. Let X be a random variable which has a GBSD with pdf *f (x; α, β, σ)* as defined in *(3). Let*

$$f(x;\alpha,\beta,\sigma) = \begin{cases} f^{(1)}(x;\alpha,\beta,\sigma), -\infty < x < 0, \\ f^{(2)}(x;\alpha,\beta,\sigma), 0 \le x < \infty, \end{cases} \tag{11}$$

where

$$f^{(1)}(x;\alpha,\beta,\sigma) = \frac{\Gamma(\alpha+\beta+2)}{\sigma\Gamma(\alpha+1)\Gamma(\beta+1)}[2F(\tfrac{x}{\sigma})]^{\alpha}[1-2F(\tfrac{x}{\sigma})]^{\beta} f(\tfrac{x}{\sigma}), -\infty < x < 0$$

and

$$f^{(2)}(x;\alpha,\beta,\sigma) = \frac{\Gamma(\alpha+\beta+2)}{\sigma\Gamma(\alpha+1)\Gamma(\beta+1)}[2(1-F(\tfrac{x}{\sigma}))]^{\alpha}[2F(\tfrac{x}{\sigma})-1]^{\beta} f(\tfrac{x}{\sigma}), 0 \le x < \infty.$$

Let Z = |X|. Then the distribution of Z is known as half-GBSD or folded GBSD about x = 0. In this case, the pdf of Z is given by

$$g(z;\alpha,\beta,\sigma) = 2f^{(2)}(x;\alpha,\beta,\sigma), \quad x \ge 0. \tag{12}$$

Now from the given observations $X_1, ..., X_n$, set apart the negative values. Let there be n_1 negative values which are denoted $X_{11}, X_{12}, ..., X_{1n_1}$. Similarly, set apart the non-negative observations as $X_{21}, X_{22}, ..., X_{2n_2}$ such that $n = n_1 + n_2$. Then the likelihood is written as

$$L(\alpha,\beta,\sigma) = \{\prod_{i=1}^{n_1} f^{(1)}(x_{1i};\alpha,\beta,\sigma)\}.\{\prod_{j=1}^{n_2} f^{(2)}(x_{2j};\alpha,\beta,\sigma)\}. \tag{13}$$

Since $f(x;\alpha,\beta,\sigma)$ is the pdf which is symmetric about zero, we have $f^{(1)}(x_{1i};\alpha,\beta,\sigma) = f^{(2)}(-x_{1i};\alpha,\beta,\sigma) = f^{(2)}(|x_{1i}|;\alpha,\beta,\sigma)$ as x_{1i} is negative for i = 1, 2, ..., n_1. Further for all observations which are non-negative, we have $f^{(2)}(x_{2j};\alpha,\beta,\sigma) = f^{(2)}(|x_{2j}|;\alpha,\beta,\sigma), j = 1,2,...,n_2$.

Thus on putting $z_i = |x_i|$, we can write

$$L(\alpha,\beta,\sigma) = \prod_{i=1}^{n} f^{(2)}(|x_i|;\alpha,\beta,\sigma) = \prod_{i=1}^{n} f^{(2)}(z_i;\alpha,\beta,\sigma). \tag{14}$$

It is well known that, if $X_1, X_2, ..., X_n$ is a random sample of size n arising from the symmetric density $f(x_i; \alpha, \beta, \sigma)$, then $Z_i = |X_i|, i = 1, 2, ..., n$ can be regarded as a random sample of size n drawn from the half-GBSD with pdf given by $g(z;\alpha,\beta,\sigma) = 2f^{(2)}(z;\alpha,\beta,\sigma)$ for z ≥ 0. Let L_1(α, β, σ) denote the likelihood based on the independent random variables $Z_1, Z_2, ..., Z_n$. Then we have

$$L_1 = 2^n \prod_{i=1} f^{(2)}(z_i;\alpha,\beta,\sigma). \tag{15}$$

Clearly (14) and (15) attains maximum at same values of α, β and σ. This proves the theorem.

Note 4.1. For more theoretical discussion in support of the above theorem one may also see, Thomas and Anjana (2021).

As a consequence of the above theorem, we need to proceed in the following way to determine the MLE's of α, β and σ involved in GBSD with pdf $f(x;\, \alpha,\, \beta,\, \sigma)$. If $X_1, X_2, ..., X_n$ are the given sample of observations from $f(x;\alpha,\beta,\sigma)$, then we write $Z_i = |X_i|,\ i = 1,\ 2,\ ...,\ n.$ Now from theorem 3.1, we have to maximize the likelihood

$$L = \left(\tfrac{1}{\sigma}\right)^n \frac{2^{n\alpha}\{\Gamma(\alpha+\beta+2)\}^n}{\{\Gamma(\alpha+1)\}^n\{\Gamma(\beta+1)\}^n} \prod_{i=1}^{n} \Big[1 - F(\tfrac{z_i}{\sigma})\Big]^{\alpha} \prod_{i=1}^{n} \Big[2F(\tfrac{z_i}{\sigma})-1\Big]^{\beta} \prod_{i=1}^{n} f(\tfrac{z_i}{\sigma}),$$

$$\log\ L = Constant - n\log\sigma + n\alpha\log 2 + n\log(\Gamma(\alpha+\beta+2)) - n\log(\Gamma(\alpha+1)) - n\log(\Gamma(\beta+1)) + \alpha\sum_{i=1}^{n}\log\Big[(1 - F(\tfrac{z_i}{\sigma}))\Big] + \beta\sum_{i=1}^{n}\log\Big[2F(\tfrac{z_i}{\sigma})-1\Big] + \sum_{i=1}^{n}\log f(\tfrac{z_i}{\sigma}).$$

The maximum likelihood (ML) equations are

$$n[\Psi(\alpha+\beta+2)-\Psi(\alpha+1)] + \sum_{i=1}^{n}\log\Big[2(1-F(\tfrac{z_i}{\sigma}))\Big] = 0, \tag{16}$$

$$n[\Psi(\alpha+\beta+2)-\Psi(\beta+1)] + \sum_{i=1}^{n}\log\Big[2F(\tfrac{z_i}{\sigma})-1\Big] = 0, \tag{17}$$

$$\sigma - \frac{\alpha}{n}\sum_{i=1}^{n}\frac{z_i f(\frac{z_i}{\sigma})}{(1-F(\frac{z_i}{\sigma}))} + \frac{2\beta}{n}\sum_{i=1}^{n}\frac{z_i f(\frac{z_i}{\sigma})}{(2F(\frac{z_i}{\sigma})-1)} + \frac{1}{n}\sum_{i=1}^{n}\frac{z_i f'(\frac{z_i}{\sigma})}{f(\frac{z_i}{\sigma})} = 0, \tag{18}$$

where $\Psi(.)$ is the usual di-gamma function defined by $\Psi(t) = \dfrac{\frac{d}{dt}\Gamma(t)}{\Gamma(t)}$. With respect to any cdf $F(\frac{x}{\sigma})$ and $f(\frac{x}{\sigma})$, the above equations can be solved for the MLE's $\hat{\alpha}, \hat{\beta}$ and $\hat{\sigma}$ of α, β and σ respectively. As explicit solutions are not readily available for $\hat{\alpha}, \hat{\beta}$ and $\hat{\sigma}$, they are solved by Newton-Raphson numerical method.

The asymptotic variance-covariance matrix of the estimates $\hat{\alpha}, \hat{\beta}$ and $\hat{\sigma}$ can be obtained by taking the inverse of the Fisher information matrix. For large n, it is appropriate that we approximate the expected values of the second-order derivatives of logarithms of likelihood function by just replacing the parameters in (16) to (18) by their respective estimates (see, Cohen 1965). Thus, the Hessian matrix can be written as,

$$I_n(\hat{\xi}) = \begin{bmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{bmatrix}$$

where I_{ij} is the second order partial derivative of log-likelihood function with respect to i^{th} and j^{th} components of parameter vector $\xi = (\alpha, \beta, \sigma)'$ and computed at $\hat{\xi} = (\hat{\alpha}, \hat{\beta}, \hat{\sigma})'$. Specifically,

$$I_{11} = -\frac{\partial^2 logL}{\partial\alpha^2} / (\hat{\alpha},\hat{\beta},\hat{\sigma}) = -n(\Psi'(\hat{\alpha}+\hat{\beta}+2) - \Psi'(\hat{\alpha}+1)), \tag{19}$$

$$I_{22} = -\frac{\partial^2 logL}{\partial\beta^2} / (\hat{\alpha},\hat{\beta},\hat{\sigma}) = -n(\Psi'(\hat{\alpha}+\hat{\beta}+2) - \Psi'(\hat{\beta}+1)), \tag{20}$$

$$I_{12} = -\frac{\partial^2 logL}{\partial\alpha\partial\beta} / (\hat{\alpha},\hat{\beta},\hat{\sigma}) = -n(\Psi'(\hat{\alpha}+\hat{\beta}+2)), \tag{21}$$

$$I_{13} = -\frac{\partial^2 logL}{\partial\alpha\partial\sigma} / (\hat{\alpha},\hat{\beta},\hat{\sigma}) = -\sum_{i=1}^{n}\frac{z_i f(\frac{z_i}{\hat{\sigma}})}{\hat{\sigma}^2(1-F(\frac{z_i}{\hat{\sigma}}))}, \tag{22}$$

$$I_{23} = -\frac{\partial^2 logL}{\partial\beta\partial\sigma} / (\hat{\alpha},\hat{\beta},\hat{\sigma}) = \sum_{i=1}^{n}\frac{2z_i f(\frac{z_i}{\hat{\sigma}})}{\hat{\sigma}^2(2F(\frac{z_i}{\hat{\sigma}})-1)}, \tag{23}$$

$$I_{33} = -\frac{\partial^2 logL}{\partial\sigma^2} / (\hat{\alpha},\hat{\beta},\hat{\sigma}) = -\frac{n}{\hat{\sigma}^2} + \sum_{i=1}^{n}\frac{z_i^2 f(\frac{z_i}{\hat{\sigma}})^2}{\hat{\sigma}^4}\Big[\frac{\hat{\alpha}}{(1-F(\frac{z_i}{\hat{\sigma}}))^2} + \frac{4\hat{\beta}}{(2F(\frac{z_i}{\hat{\sigma}})-1)^2}\Big] + \sum_{i=1}^{n}\frac{z_i^2 f'(\frac{z_i}{\hat{\sigma}})}{\hat{\sigma}^4}\Big[\frac{\hat{\alpha}}{(1-F(\frac{z_i}{\hat{\sigma}}))} - \frac{2\hat{\beta}}{(2F(\frac{z_i}{\hat{\sigma}})-1)}\Big] + 2\sum_{i=1}^{n}\frac{z_i f(\frac{z_i}{\hat{\sigma}})}{\hat{\sigma}^3}\Big[\frac{\hat{\alpha}}{(1-F(\frac{z_i}{\hat{\sigma}}))} - \frac{2\hat{\beta}}{(2F(\frac{z_i}{\hat{\sigma}})-1)}\Big] - \sum_{i=1}^{n}\frac{z_i f'(\frac{z_i}{\hat{\sigma}})}{\hat{\sigma}^3 f(\frac{z_i}{\hat{\sigma}})}\Big[\frac{z_i f'(\frac{z_i}{\hat{\sigma}})}{\hat{\sigma} f(\frac{z_i}{\hat{\sigma}})} - 2\Big] - \sum_{i=1}^{n}\frac{z_i^2 f''(\frac{z_i}{\hat{\sigma}})}{\hat{\sigma}^4 f(\frac{z_i}{\hat{\sigma}})}. \tag{24}$$

From the asymptotic properties of ML estimators under regularity conditions and by using the multivariate central limit theorem, we state the asymptotic normality results of maximum likelihood estimators of α, β and σ as given below.

$$\sqrt{n}(\hat{\xi}-\xi) \sim N_3(0, I_n^{-1}(\xi)) \quad as \quad n \to \infty.$$

Since the parameters are unknown, $I_n^{-1}(\xi)$ is estimated by $I_n^{-1}(\hat{\xi})$. The asymptotic normality result stated above can be used to obtain the asymptotic confidence intervals for the parameters of GBSD.

5. SIMULATION STUDY

A simulation study is conducted to assess the properties of the maximum likelihood estimators of the parameters of the newly proposed bimodal distribution (GBSD) with pdf given in (3). To illustrate the closeness of the estimated values of α, β and σ with their true values we have used MATHEMATICA (ver 11.3) software to simulate 100 independent observations from the generalized bimodal distribution (GBSD) with normal and logistic distributions as baseline distributions.

We have simulated 100 independent observations from the generalized bimodal distribution with normal (N(x,σ)) baseline distribution denoted by GBSD-N(x, σ) for each of the following choice of parameters: (1) α = 1.5, β = 1.5, σ = 1 and (2) α = 2, β = 3, σ = 4, obtained the maximum likelihood estimates $\hat{\alpha}, \hat{\beta}$ and $\hat{\sigma}$ for each of the two cases, using the method described in section 4 and repeated it 100 times. The mean of the estimates and RMSE are given in Table 1. Again, we have similarly simulated 300 independent observations from GBSD-N(x,σ) for the already assumed values of the parameters and then estimated $\hat{\alpha}, \hat{\beta}$ and $\hat{\sigma}$. This process is again repeated 100 times and the mean of the estimates and their RMSE are also given in Table 1. We noticed that the estimates of the parameters are very close to the assumed values of the parameters of GBSD-N(x,σ).

Similarly, we have also carried out the same schemes of simulation for the generalized bimodal distribution with logistic(L(x,σ)) baseline distribution denoted by GBSD-L(x,σ) for n=100 and 300 and each repeated 100 times and for the assumed values of parameters. The means of the estimates with the RMSE are also given in Table 1. From the table, we observed that the estimates obtained are very close to the assumed values of the parameters. The table also shows that the biases and RMSEs of the estimates of the model parameters diminish as the sample size increases.

Table 1. Estimated parameters and RMSE's obtained in the simulation study

Baseline Distribution	Sample Size	Actual Parameters			Estimated Parameters (RMSE)		
		α	β	σ	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$
Normal	100	1.5	1.5	1	1.4926	1.4912	0.9647
					(0.0346)	(0.0288)	(0.0977)
	300	1.5	1.5	1	1.4961	1.4927	0.9759
					(0.0239)	(0.0263)	(0.0965)
	100	2	3	4	1.9853	2.9812	3.9352
					(0.0254)	(0.0289)	(0.0728)
	300	2	3	4	1.9869	2.9852	3.9542
					(0.0212)	(0.0257)	(0.0653)
Logistic	100	1.5	1.5	1	1.4824	1.4851	0.9752
					(0.0427)	(0.0323)	(0.0989)
	300	1.5	1.5	1	1.4895	1.4886	0.9867
					(0.0359)	(0.0256)	(0.0973)
	100	2	3	4	1.9654	2.8721	3.8512
					(0.0418)	(0.0675)	(0.0942)
	300	2	3	4	1.9775	2.9982	3.9612
					(0.0385)	(0.0552)	(0.0872)

6. REAL LIFE APPLICATIONS

In this section we consider two data sets: (i) the annual soybean productivity data (kg/ha) of some important municipalities in Parana, Brazil and (ii) the data on seasonal rainfall of Kerala as available in Kothawale and Rajeevan (2017). In both of the above data sets, we have obtained the deviations of each observation from the respective median of the data sets and those deviations are reproduced as such in Appendix-I. Clearly, each data set exhibit two peaks on their histograms. We now illustrate the modelling on those data sets by bimodal distributions which we developed in this paper.

6.1 Soybean Productivity Data

Duarte *et al.* (2018) used the annual soybean production data in Kilogram per hectare in a municipality area of Parana, Brazil, which were recorded during 1975 to 2013 to construct a suitable distribution of yields so as to apply it to a crop insurance problem. As two peakedness of the histograms is observed in the data, a bimodal distribution seems to be an ideal model to contain the data. We take the data as such and observe the value of the sample median m as 2.5535. We have transformed now the original data into the deviations of those observations from the median m = 2.5535. The transformed data set is provided as the 1st data set and is presented as Data 1 in Appendix-I.

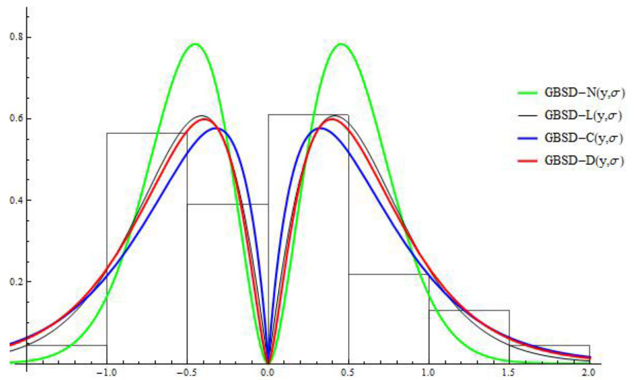

Fig. 4. Histograms of the transformed annual production of soybeans data with fitted density curve of generalized bimodal distributions with N(y, σ), L(y, σ), C(y, σ) and D(y, σ) baseline distributions

From the transformed data as given in 1 of Appendix-I, one can observe that there are exactly 23 observations with positive and the same number of observations with negative sign. To test the null hypothesis H_0 that the median of the population from which the above data (data in 1 of Appendix-I)

Table 2. Estimated parameters and model comparison statistics obtained on modelling the transformed soybean production data by GBSD models with N(y,σ), L(y,σ), C(y,σ) and D(y,σ) baseline distributions

Model	Estimated Parameters	K-S distance between EDF and fitted cdf (p- value)	AIC	BIC
GBSD-N(y,σ)	$\hat{\sigma}=0.6421$ $\hat{\alpha}=1.0960$ $\hat{\beta}=1.7820$	0.1125 (0.5795)	131.603	137.089
GBSD-L(y,σ)	$\hat{\sigma}=0.6071$ $\hat{\alpha}=1.6961$ $\hat{\beta}=1.0673$	0.0962 (0.7515)	110.999	116.485
GBSD-C(y,σ)	$\hat{\sigma}=4.5797$ $\hat{\alpha}=19.5903$ $\hat{\beta}=0.9329$	0.0849 (0.8667)	108.499	113.985
GBSD-D(y,σ)	$\hat{\sigma}=24.4716$ $\hat{\alpha}=95.1934$ $\hat{\beta}=1.5673$	0.0989 (0.7340)	116.738	122.224

is obtained is equal to zero against the alternative H_1 : It is $\neq 0$ by assuming n as large, we consider the test statistic of a sign test (see, Rohatgi and Saleh, 2015),

$$Z=\left|\frac{23-\frac{n}{2}}{\sqrt{\frac{n}{4}}}\right|.$$

Since n=46, we have Z=0. For a two sided test the p-value is 1. Similarly, if we use Wilcoxon signed-rank test, then we rank the absolute values of the observations in 1 of Appendix-I and write T+ and T− as the sum of the ranks of positive and negative observations respectively. If $T=Min(T+,T-)$, then $E(T)=\frac{n(n+1)}{4}$, $Var(T)=\frac{n(n+1)(2n+1)}{24}$. Then assuming n large, we use the statistic $Z=\left|\frac{T-E(T)}{\sqrt{V(T)}}\right|$ and in this case we have T=533 and Z=0.0878 with a p value 0.9346. Thus from both of the above tests, we have reasons to believe that the data set given in 1 of Appendix-I arises from a distribution belonging to **A**.

Suppose the pdf of the baseline distribution is $f_1(y,\sigma)=\frac{1}{\sigma}g_1(\frac{y}{\sigma})$ as defined in (4) with the corresponding cdf $G_1(\frac{y}{\sigma})$, then on replacing $f(y)$ and F(y) in (2) by $\frac{1}{\sigma}g_1(\frac{y}{\sigma})$ and $G_1(\frac{y}{\sigma})$ respectively, then we obtain the pdf of GBSD with normal distribution (say N (y, σ)) as the baseline distribution and is denoted as GBSD-N(y,σ). Similarly on replacing $f(y)$ by $\frac{1}{\sigma}g_2(\frac{y}{\sigma})$ and F(y) by the cdf $G_2(\frac{y}{\sigma})$ derived from $f_2(y,\sigma)$ in the equation (2), we obtain the pdf of GBSD with logistics baseline distribution (say L(y,σ)) and is denoted as GBSD-L(y,σ). Likewise on using Cauchy distribution (say C(y,σ)) with $\frac{1}{\sigma}g_3(\frac{y}{\sigma})$ defined in (6) and its cdf $G_3(\frac{y}{\sigma})$ in (2), we obtain the pdf of GBSD with Cauchy distribution as the baseline distribution and is denoted by GBSD-C(y,σ). Another well known distribution belonging to the family **A**. of distributions is double exponential distribution (denoted by D(y,σ)) with pdf

$$f_4(y,\sigma)=\frac{1}{2\sigma}e^{-\left|\frac{y}{\sigma}\right|},-\infty<y<\infty. \quad (25)$$

We may also write the above density as $f_4(y,\sigma)=\frac{1}{\sigma}g_4(\frac{y}{\sigma})$. If we use D(y,σ) with pdf $\frac{1}{\sigma}g_4(\frac{y}{\sigma})$ and cdf $G_4(\frac{y}{\sigma})$ in (2), then we obtain the pdf of GBSD with D(y,σ) as the baseline distribution and is denoted by GBSD-D(y,σ). We have used the data given in 1 of Appendix-I to construct the models GBSD-N(y,σ), GBSD-L(y,σ), GBSD-C(y,σ) and GBSD-D(y,σ) by applying maximum likelihood method of estimation to estimate the parameters α, β and σ involved in each of the above models. The estimates $\hat{\alpha},\hat{\beta}$ and $\hat{\sigma}$ of each of the GBSD models are presented in Table 2.

Also, we have obtained the K-S distance between empirical distribution function (EDF) and the cdfs of the constructed GBSD models together with p-values and are given in Table 2. The AIC and BIC values corresponding to each of the fitted model as well have been computed and given in Table 2. The graph of densities of the constructed models GBSD- N(y,σ), GBSD-L(y,σ), GBSD-C(y,σ) and GBSD-D(y,σ) are displayed in Fig. 4 along with the histograms of the transformed annual production Soybean data. From Table 2 and Fig. 4, we conclude that among the four bimodal distributions constructed GBSD-C(y,σ) is the best bimodal distribution that can be picked up to represent the data given in 1 of Appendix-I. Thus, the most suitable bimodal density to represent the soybean production data is given below

$$f_3(y;\hat{\alpha},\hat{\beta},\hat{\sigma})=\begin{cases} d(\hat{\alpha},\hat{\beta})\frac{1}{\hat{\sigma}}[2G_3(\frac{y}{\hat{\sigma}})]^{\hat{\alpha}}[1-2G_3(\frac{y}{\hat{\sigma}})]^{\hat{\beta}}g_3(\frac{y}{\hat{\sigma}}), -\infty<y<0, \\ d(\hat{\alpha},\hat{\beta})\frac{1}{\hat{\sigma}}[2(1-G_3(\frac{y}{\hat{\sigma}}))]^{\hat{\alpha}}[2G_3(\frac{y}{\hat{\sigma}})-1]^{\hat{\beta}}g_3(\frac{y}{\hat{\sigma}}), 0\le y<\infty, \end{cases} \quad (26)$$

where $d(\hat{\alpha},\hat{\beta})=\frac{\Gamma(\hat{\alpha}+\hat{\beta}+2)}{\Gamma(\hat{\alpha}+1)\Gamma(\hat{\beta}+1)}$, $\hat{\alpha}=19.5903$, $\hat{\beta}=0.9329$ and $\hat{\sigma}=4.5797$. If the original data from which data given in 1 of Appendix-I is constructed, then the model which is suitable for that data maybe obtained from (26) by replacing y in it by $x-2.5535$ where the range of the first part of the density is $-\infty<x<2.5535$ and that of the second part is $2.5535\le x<\infty$.

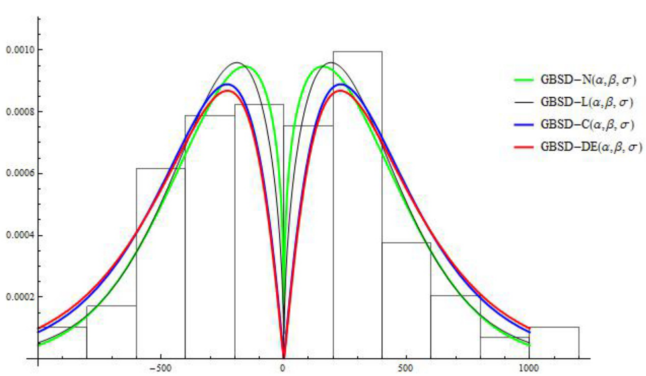

Fig. 5. Histograms of the transformed annual rainfall data with fitted density curve of generalized bimodal distributions with N(y, σ), L(y, σ), C(y, σ) and D(y, σ) baseline distributions

6.2 Annual Rainfall in Kerala Subdivision (1871-2016)

Researchers, Policy makers and Government agencies are very interested in the information of regional and sub-divisional rainfall in India. Statistical characteristics and long-term variability of Annual rainfall data (in millimeter) for the period from 1871 to 2016 of Kerala reported in Kothawale and Rajeevan (2017). If we use the histogram representation of this data, then we observe two peaks and hence we suggest a bimodal distribution for modelling it. Now we take the Indian Institute of Meteorology data as reported in Kothawale and Rajeevan (2017) and find the value of the sample median m as 2801.6. As did in section 6.1, here again we transform the original data into deviation of the observations from the median $m=2801.6$. The transformed data is presented as dataset 2 of Appendix-I.

From the transformed data as given in 2 of Appendix-I, one can observe that there are exactly 73 observations with positive and the same number of observations with negative sign. To test the null

Table 3. Estimated parameters and model comparison statistics obtained on modelling the transformed annual rainfall data by GBSD models with N(y,σ), L(y,σ), C(y,σ) and D(y,σ) baseline distributions

Model	Estimated Parameters	K-S distance between EDF and fitted cdf (p- value)	AIC	BIC
GBSD-N(y,σ)	$\hat{\sigma}=574.9670$ $\hat{\alpha}=0.9047$ $\hat{\beta}=0.3174$	0.0377 (0.9805)	2176.39	2185.34
GBSD-L(y,σ)	$\hat{\sigma}=299.9760$ $\hat{\alpha}=0.7380$ $\hat{\beta}=0.3883$	0.0442 (0.9257)	2177.01	2185.96
GBSD-C(y,σ)	$\hat{\sigma}=2235.36$ $\hat{\alpha}=14.8809$ $\hat{\beta}=1.0736$	0.0781 (0.3192)	2196.93	2205.96
GBSD-D(y,σ)	$\hat{\sigma}=2737.3969$ $\hat{\alpha}=12.1650$ $\hat{\beta}=1.1508$	0.0891 (0.2940)	2195.88	2204.84

hypothesis H_0 that the median of the population from which the above data (data in 2 of Appendix- I) is obtained is equal to zero against the alternative H_1: It is $\neq 0$ by assuming n as large, we consider the test statistic of a sign test (see, Rohatgi and Saleh, 2015),

$$Z = \left| \frac{73 - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \right|$$

Since n=146, we have Z=0. For a two sided test the p-value is 1.

Similarly, if we use Wilcoxon signed-rank test, then we rank the absolute values of the observations in 2 of Appendix-I and write T+ and T- as the sum of the ranks of positive and negative observations respectively. If T=Min (T+, T-), then $E(T) = \frac{n(n+1)}{4}$ and $Var(T) = \frac{n(n+1)(2n+1)}{24}$. Then assuming n large we use the statistic $Z = \left| \frac{T - E(T)}{\sqrt{V(T)}} \right|$ and in this case we have T=5485 and Z=0.2335 with a p value 0.8154. Thus from both of the above tests, we have reasons to believe that the data set given in 2 of Appendix-I arises from a distribution belonging to **A** .

We have used the data given in 2 of Appendix-I to construct the models GBSD-N(y,σ), GBSD-L(y,σ), GBSD-C(y,σ) and GBSD-D(y,σ) by applying maximum likelihood method of estimation to estimate the parameters α, β and σ involved in each of the above models. The estimates $\hat{\alpha}, \hat{\beta}$ and $\hat{\sigma}$ of each of the GBSD models are presented in Table 3.

Also, we have obtained the K-S distance between empirical distribution function (EDF) and the cdfs of the constructed GBSD models together with p-values and are given in Table 3. The AIC and BIC values corresponding to each of the fitted model as well have been computed and given in Table 3. The graph of densities of the constructed models GBSD- N(y,σ), GBSD-L(y,σ), GBSD-C(y,σ) and GBSD-D(y,σ) are displayed in Fig. 5 along with the histograms of the transformed annual rainfall data. From Table 3 and Fig. 5, we conclude that among the four bimodal distributions constructed GBSD-N(y,σ) is the best bimodal distribution that can be picked up to represent the data given in 2 of Appendix-I. Thus the most suitable bimodal density to represent the annual rainfall data is given below

$$f_1(y;\hat{\alpha},\hat{\beta},\hat{\sigma}) = \begin{cases} d(\hat{\alpha},\hat{\beta})\frac{1}{\hat{\sigma}}[2G_1(\frac{y}{\hat{\sigma}})]^{\hat{\alpha}}[1-2G_1(\frac{y}{\hat{\sigma}})]^{\hat{\beta}} g_1(\frac{y}{\hat{\sigma}}), & -\infty < y < 0, \\ d(\hat{\alpha},\hat{\beta})\frac{1}{\hat{\sigma}}[2(1-G_1(\frac{y}{\hat{\sigma}}))]^{\hat{\alpha}}[2G_1(\frac{y}{\hat{\sigma}})-1]^{\hat{\beta}} g_1(\frac{y}{\hat{\sigma}}), & 0 \le y < \infty, \end{cases} \quad (27)$$

where $d(\hat{\alpha},\hat{\beta}) = \frac{\Gamma(\hat{\alpha}+\hat{\beta}+2)}{\Gamma(\hat{\alpha}+1)\Gamma(\hat{\beta}+1)}$, $\hat{\alpha} = 0.9047$, $\hat{\beta} = 0.3174$ and $\hat{\sigma} = 574.9670$. If the original data from which data given in 2 of Appendix-I is constructed then the model which is suitable for that data may be obtained from (27) by replacing y in it by $x - 2801.6$ where the range of the first part of the density is $-\infty < x < 2801.6$ and that of the second part is $2801.6 \le x < \infty$.

7. CONCLUSIONS

There are several error datasets collected from real life problems appears to have two modes. Existing bimodal distributions fail to represent those data sets as reasonable models. In this paper, an extensively large class of bimodal distributions corresponding to each of the symmetric baseline distribution is explored to model such error data sets. A new statistic based on the density value computed at the value of an observation and its rank among those values computed for all observations is named as ODVI statistics is also introduced. The relationship between the sampling distributions of ODVI statistics and GBSD is further established. Maximum likelihood estimation and inference on the parameters of GBSD have been discussed. Application of the GBSD to model real data sets is also illustrated to show the importance of this model. The application of GBSD with different baseline distributions to model more agricultural, medical and economic data will be carried out in our future investigations. Also, there is large scope for conducting theoretical as well as applied studies on GBS distribution and ODVI statistics.

ACKNOWLEDGMENTS

The anonymous reviewer is very much thanked for many of his helpful comments.

REFERENCES

Alzaatreh, A., Lee, C.and Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron,* **71(1)**, 63-79.

Cohen, A.C. (1965). Maximum likelihood estimation in the Weibull distribution based on complete and on censored samples. *Technometrics,* **7(4)**, 579-588.

Duarte, G.V., Braga, A., Miquelluti, D.L. and Ozaki, V.A. (2018). Modeling of soybean yield using symmetric, asymmetric and bimodal distributions: implications for crop insurance. *Journal of Applied Statistics*, **45(11)**, 1920-1937.

Eisenberger, I. (1964). Genesis of bimodal distributions. *Technometrics*, **6(4)**, 357-363.

Gauss, C.F. (1857). *Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections: A Translation of Gauss's" Theoria Motus." With an Appendix.* Little, Brown.

Hassan, M.Y. and El-Bassiouni, M.Y. (2013). Modelling Poisson marked point processes using bivariate mixture transition distributions. *Journal of Statistical Computation and Simulation,* **83(8)**, 1440-1452.

Hassan, M.Y. and Hijazi, R.H. (2010). A bimodal exponential power distribution. *Pakistan Journal of Statistics*, **26(2)**, 379-396.

Jayakumar, K. and Babu, M.G. (2017). T-transmuted x family of distributions. *Statistica*, **77**, 251-276.

Kothawale, D.R. and Rajeevan, M. (2017). Monthly, Seasonal, Annual Rainfall Time Series for All-India, Homogeneous Regions, Meteorological Subdivisions: 1871-2016. *Technical report, Indian Institute of Tropical Meteorology, Table No. 36, P.161*, Pune, India .

Marshall, A.W. and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families*. Biometrika*, **84(3)**, 641-652.

Prasad, A. (1954). Bimodal distributions derived from normal distribution. *Sankhya,* **14**, 369-374.

Rao, C.R. and Gupta, S.D. (1989). *Selected Papers of C R Rao*, Volume 5. Taylor & Francis, New York.

Rohatgi, V.K. and Saleh, A.K. (2015). *An Introduction to Probability and Statistics*. John Wiley & Sons, New York.

Sarma, P.V.S., Rao, K.S.S. and Rao, R.P. (1990). On a family of bimodal distributions.*Sankhya, Series B,* **52**, 287-292.

Thomas, P.Y. and Anjana, V. (2021). Estimation of the scale parameter of a family of distributions using a newly derived minimal sufficient statistic. *Communications in Statistics- Theory and Methods,doi:10.1080/03610926.2021.1884721.*

Thomas, P.Y. and Priya, R.S. (2015). On a less cumbersome method of estimation of parameters of type III generalized logistic distribution by order statistics. *Statistica,* **75(3)**, 291-312.

Thomas, P.Y. and Priya, R.S. (2016). Symmetric beta-Cauchy distribution and estimation of parameters using order statistics. *Calcutta Statistical Association Bulletin*, **68(1-2)**, 111–134.

APPENDIX-I

Data Set-1 and Data Set-2 used for the real life application in this study is given here.

1. Transformed soybean productivity data

-0.4535, -0.5965, -0.3565, -1.1035, -0.9135, -0.8045, -0.8035, -0.9425, -0.9615, -0.8535 , -0.6335, -0.6935, -0.6725, -0.3335, -0.6535, -0.6035, -0.6635, -0.2535, -0.0535, 0.0335, 0.2365, 0.1765, -0.0335, 0.2265, 0.1265, -0.0345, 0.4755, 0.2715, 0.4465, 0.2465, 0.0465, 0.4165, 0.4465, 0.2955, -0.0555, 0.6495, 0.8905, 0.4955, 0.9575, 0.9755, 0.9455, 1.0465, 1.5665, 1.2565, 1.2465, -0.4535

2. Transformed annual rainfall data

389.35, 78.95, -63.05, 544.75, -412.35, -815.25, 202.35, 915.35, -129.35, -369.35, -945.85,684.45, -81.65, -474.65, 329.25, -456.35, -363.05, 188.35, -25.35, -648.35, -48.25, 334.85,-267.15, -434.85, -510.35, 33.25, 420.45, -434.55, -670.15, -130.75, 140.45, 254.25, 232.65,145.45, -317.75, -349.25, 615.05, -102.85, 62.55, -69.55, -209.05, 609.15, -329.05, 124.35,359.65, 64.35, 9.35, -227.45, 248.35, 631.65, -143.35, 538.35, 397.05, 1143.35, 265.95, 150.15,73.05, -450.05, 502.95, 120.25, 241.15, 458.55, 1082.45, -493.95, -466.85, 241.35, 9.85, -300.85,93.45, 300.05, 232.05, 174.85, 638.65, -384.45, -391.65, 840.95, 233.85, 168.85, 153.25, 338.35,-195.15, -488.35, -317.15, 123.25, 356.25, -44.65, 276.95, -97.65, 498.05, 526.65, 1105.65,342.65, -300.35, -220.15, -583.35, -413.85, -106.55, 497.55, -85.05, -64.05, 257.05, 30.35,-394.35, 144.25, 792.35, -629.55, 352.15, 367.95, -102.25, -9.35, 582.75, -412.35, -285.55,-268.05, -272.05, -654.45, -494.85, -175.55, -286.45, -195.55, 212.45, 349.15, -142.65, 566.95,-72.65, -230.55, 505.75, 269.15, 172.65, -582.55,-172.65,-201.55,-489.85, -81.85,-500.85,354.25,326.65,-406.05, -217.55, 283.45, -17.15, -723.15, 386.95, 166.65, -243.95, -964.15