Information Sciences Letters
An International Journal

http://dx.doi.org/10.18576/isl/121222

Mathematical and statistical structures of a new generalized probability distribution for fitting different types of datasets

Mahmoud El-Morshedy[1,2,*]

[1]Department of Mathematics, College of Science and Humanities in Al-Kharj, Prince Sattam bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia
[2]Department of Mathematics, Faculty of Science, Mansoura University, Mansoura 35516, Egypt

Received: 5 Oct. 2023, Revised: 6 Nov. 2023, Accepted: 9 Nov. 2023
Published online: 1 Dec. 2023

Abstract: This article presents a novel approach to modeling a heavily tailed continuous distribution known as the logarithmic slash model. This model is built upon the transformation $Y = e^X$, where X follows the slash distribution. Our aim is to introduce a model with desirable characteristics for practical applications, drawing inspiration from the unique features of the logarithmic model. As a result, we develop both univariate and multivariate extensions of the logarithmic slash model and conduct a thorough exploration of its mathematical properties. We employ the maximum likelihood method to estimate the model parameters and conduct simulation studies to assess the biases and mean square errors of these estimators. One of the primary concerns addressed by the logarithmic slash model is its ability to effectively accommodate various types of data. To demonstrate its versatility, we utilize a range of datasets and compare the performance of the logarithmic slash model to a strong competitor in terms of data fit. The results clearly indicate that the logarithmic slash model outperforms its competitor, highlighting its efficacy in handling different types of data.

Keywords: Slash distribution; Multivariate; Zero-inflated data; Moments; Heavy-tailed distribution; Simulation.
2000 AMS Classification: 60E05; 62E10.

1 Introduction

In today's highly advanced era, distribution theory has garnered significant attention due to its ability to offer a wide array of probability models. This surge in interest arises from the growing need for users and applied researchers to effectively address the challenges of formulating models and analyzing real-world datasets. Consequently, there is an urgent demand for the development of practical models to enhance our understanding of real-world phenomena. Notably, advancements in high computing capacity have brought about a profound transformation in the approach to designing probability models, diverging significantly from those proposed before 1997 (refer to Tahir and Cordeiro, [1]). In light of the above, the primary objective of developing or generalizing models is to cater to the contemporary requirements prevalent across various domains, including insurance, engineering, medical sciences, actuarial science, biostatistics, biomathematics, and numerous others. For instance, traditional distributions such as Weibull, Rayleigh, exponential, gamma, Gompertz, and Lindley, commonly used in these fields, exhibit limited properties and lack the flexibility required for comprehensive data modeling. Consequently, heavy-tailed distributions have emerged as a focal point in statistical literature and practical applications, particularly in areas like financial sciences, reliability engineering, and biomedical sciences. In practice, datasets in these domains often exhibit positive skewness, with higher values in the tails and a narrower central distribution compared to a normal distribution. Existing distributions in the literature struggle to adapt to such high-tailed datasets. For instance, the Pareto distribution, commonly employed for representing financial data, falls short in many applications, while the Weibull model performs well for smaller losses but fails to adequately cover significant losses (as discussed in Bhati and Ravi, [2]). Heavy-tailed distributions, on the other hand, offer dependable and accurate models for such

* Corresponding author e-mail: m.elmorshedy@psau.edu.sa

© 2023 NSP
Natural Sciences Publishing Cor.

scenarios. Practitioners have introduced various probability models tailored to capture these characteristics. However, there remains a scarcity of probability distributions that effectively address all of these qualities. Hence, the creation of new models by building upon existing distributions has become a compelling avenue of research, attracting a growing number of researchers. For more details, someone can see Rogers and Tukey [3], Kafadar [4], Shah et al. [5], Morgenthaler [6], El-Morshedy [7], Lange et al. [8], Ahmad et al. [9],Jamshidian [10], Khan et al. [11], Kashid and Kulkarni [12], Reyes et al. [13], Ferede et al. [14], Ghayour et al. [15], Jehhan et al. [16], Ahmed et al. [17], Alizadeh et al. [18], Altun et al. [19], Eldeeb et al. [20], Eliwa and Ahmed [21], Eliwa et al. ([22],[23]), Haj Ahmad et al. [24], Handique et al. [25], among others. Furthermore, in severe event analysis, heavy-tailed distributions are recommended over light and intermediate tail distributions (see Murshed, [26]). Nevertheless, it is crucial to maintain a parsimonious approach when determining the number of parameters to avoid excessive information loss. Typically, an optimal model is characterized by having only three parameters, as elucidated in the work of Johnson et al. [27]. With this principle in mind, we introduce and explore a heavy-tailed distribution with three parameters. In this context, the standard normal distribution $N(0,1)$ is recognized for its density function, which is as follows

$$f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}},\ -\infty<z<\infty. \tag{1}$$

The slash random variable is defined as the ratio of two independent random variables: Assume the standard normal random variable Z be independent of the uniform random variable U on $(0,1)$. Then, the random variable $S=ZU^{-1/q}$ has the standard slash normal (SSN) distribution with the following density

$$\Psi(s;q)=q\int_0^1 t^q f(st)dt,\ -\infty<s<\infty, \tag{2}$$

where $q>0$ is the shape parameter and $f(.)$ denotes the standard normal distribution density function given in (1). Setting $q=1$, the distribution is called the SSN distribution, and it has the following density

$$\Psi(s;1)=\begin{cases}\frac{1}{\sqrt{2\pi}s^2}\left(1-e^{-\frac{s^2}{2}}\right); & \text{if } s\neq 0\\ \frac{1}{2\sqrt{2\pi}} & ;\ \text{otherwise.}\end{cases} \tag{3}$$

The SSN density has heavier tails than those of the normal. The log-normal (LoN) distribution is based on the normal distribution. It describes a variable $W=e^{\mu+\sigma Z}$ where $Z\sim N(0,1)$, symbolically, we write $W\sim Log-N(\mu,\sigma^2)$. It is valid for values of W which are greater than zero. The LoN distribution is used to model continuous random quantities when the distribution is believed to be skewed, such as certain income and lifetime variables. The probability density function (PDF) of the LoN distribution is given by

$$\Phi(w;\mu,\sigma)=\frac{1}{w\sqrt{2\pi}\sigma}e^{-\frac{(\log(w)-\mu)^2}{2\sigma^2}},\ 0<w<\infty, \tag{4}$$

where $-\infty<\mu<\infty$ and $\sigma>0$. The main goal of this article is to offer a novel distribution whose density has longer tails than the LoN distribution based on the transformation $Y=e^X$, where X has the slash distribution with position parameter μ, scale parameter σ, and shape parameter $q>0$. The resulted model can be called log-slash (LoS) distribution. The motivation behind proposing heavy-tailed probability distributions is rooted in the need to better model and understand real-world data that often exhibit characteristics deviating significantly from normal or Gaussian distributions. Several key motivations for considering heavy-tailed distributions include:

– Accurate Data Representation: Many real-world phenomena involve extreme events, outliers, or rare occurrences that cannot be effectively represented by light-tailed distributions like the Gaussian distribution. Heavy-tailed distributions provide a better fit for such data by allowing for the possibility of extreme values.
– Risk Assessment: In fields such as finance, insurance, and environmental science, the accurate modeling of tail events is crucial for risk assessment. Heavy-tailed distributions are essential for capturing tail risk, which is the potential for extreme events to have a significant impact on outcomes.
– Economic and Social Data: Economic and social data often exhibit heavy-tailed behavior due to factors like wealth distribution, income inequality, and social networks. Accurate modeling of these phenomena is vital for economic policy-making, social analysis, and decision-making.
– Natural Phenomena: Natural systems, such as earthquakes, weather patterns, and biological processes, can produce data with heavy tails. Understanding and predicting the tails of these distributions are vital for disaster management, weather forecasting, and epidemiology.

© 2023 NSP
Natural Sciences Publishing Cor.

– Technological Applications: In technology and engineering, the performance and reliability of systems often depend on understanding the tails of distributions. Heavy-tailed distributions are used to model component failures, network traffic, and various other aspects of system behavior.
– Statistical Inference: Heavy-tailed distributions can have a significant impact on statistical inference. Neglecting heavy tails in data analysis can lead to biased estimates, improper confidence intervals, and incorrect hypothesis testing.
– Robustness: Heavy-tailed distributions offer robustness against outliers and unexpected data behavior. They can handle extreme observations without disproportionately affecting the overall model.

In summary, the motivations for proposing heavy-tailed probability distributions stem from the recognition that many real-world phenomena exhibit heavy-tailed behavior. To accurately capture the characteristics of such data, heavy-tailed distributions are essential for a wide range of applications in science, finance, economics, technology, and more. They allow for a more comprehensive understanding of data that goes beyond the assumptions of light-tailed distributions like the Gaussian

The paper's remaining sections are organized as follows: In Section 2, we present the univariate log-slash distribution. Section 3 systematically establishes the key characteristics of the log-slash distribution, supported by comprehensive proofs. In Section 4, we delve into the calculation of moments for the log-slash distribution. Section 5 is dedicated to introducing the multivariate log-slash distribution, along with discussions on specific special cases. The estimation of likelihood is covered in Section 6. Section 7 explores Monte Carlo simulations with varying sample sizes. In Section 8, we analyze three real-world datasets, showcasing the adaptability of the $\mathrm{LoS}(\mu,\sigma,q)$. Finally, Section 9 draws conclusions and presents the key findings derived from this study.

2 Univariate Log-Slash Distribution

As previously mentioned in Section 1, researchers are perpetually in search of fresh heavy-tailed distributions suitable for modeling various types of data. In this section, we leverage the theorem below to introduce a new heavy-tailed distribution, referred to as $\mathrm{LoS}(\mu,\sigma,q)$.

Theorem 1. Assume the random variable $X=\mu+\sigma S$ has the slash (S) distribution with location parameter $-\infty<\mu<\infty$, scale parameter $\sigma>0$ and shape parameter $q>0$, where S has the SSN distribution, symbolically, we can write $X\sim S(\mu,\sigma,q)$. If we defined a new random variable $Y=e^X$, then Y has log-S (LoS) distribution, symbolically, we can write $Y\sim LoS(\mu,\sigma,q)$. The PDF of the random variable Y is given by

$$f(y;\mu,\sigma,q)=\frac{q}{\sigma\sqrt{2\pi}y}\int_0^1 e^{-\frac{(\log(y)-\mu)^2}{2\sigma^2}t^2}t^q dt\ ,\ 0<y<\infty. \tag{5}$$

Proof. Since $X\sim S(\mu,\sigma,q)$, then

$$\begin{aligned}\Pr(Y<k)&=\Pr(e^X<k)=\Pr(X<\log(k))\\&=\int_{-\infty}^{\log(k)}\left(\frac{q}{\sigma\sqrt{2\pi}}\int_0^1 e^{-\frac{(x-\mu)^2}{2\sigma^2}t^2}t^q dt\right)dx\ ,\ -\infty<x<\infty.\end{aligned}$$

Apply the transformation $Y=e^X$, then

$$\Pr(Y<k)=\int_{-\infty}^{k}\left(\frac{q}{\sigma y\sqrt{2\pi}}\int_0^1 e^{-\frac{(\log(y)-\mu)^2}{2\sigma^2}t^2}t^q dt\right)dy\ ,\ 0<y<\infty.$$

The last term above, is the cumulative density function of Y, so we have our result.

Remarks

1. Using Maple software package, the PDF of $S(\mu,\sigma,q)$ can be expressed in a closed form as

$$\begin{aligned}f(y)=&\frac{q2^{\frac{q+3}{4}}}{(q+1)(q+3)\sigma y\sqrt{\pi}}\left(\frac{\ln y-\mu}{\sigma}\right)^{-\frac{1+q}{2}}e^{-\left(\frac{\ln y-\mu}{2\sigma}\right)^2}W\left(\frac{q+1}{4},\frac{q+3}{4},\frac{(\ln y-\mu)^2}{2\sigma^2}\right)\\&+\frac{q2^{\frac{q+3}{4}}}{(q+1)\sigma y\sqrt{\pi}}\left(\frac{\ln y-\mu}{\sigma}\right)^{-\frac{5+q}{2}}e^{-\left(\frac{\ln y-\mu}{2\sigma}\right)^2}W\left(\frac{q+5}{4},\frac{q+3}{4},\frac{(\ln y-\mu)^2}{2\sigma^2}\right),\end{aligned}$$

© 2023 NSP
Natural Sciences Publishing Cor.

where $W(.,.,.)$ is a Whittaker function introduced by Whittaker and Edmund [28].

2.Putting $\mu = 0$ and $\sigma = 1$ in (5), the standard form of the univariate LoS distribution can be generated, say $LoS(0,1,q)$. For $q = 1$, the PDF can be formulated as

$$f(y;0,1,1) = \frac{(\frac{1}{y\sqrt{2\pi}} - \Phi(y))}{(\log(y))^2}\ ,\ 0 < y < \infty, \tag{6}$$

where $\Phi(.)$ denotes the standard form of the LoN distribution density function $\Phi(y;0,1)$ given in (4). Figure 1 shows the PDFs for different models.

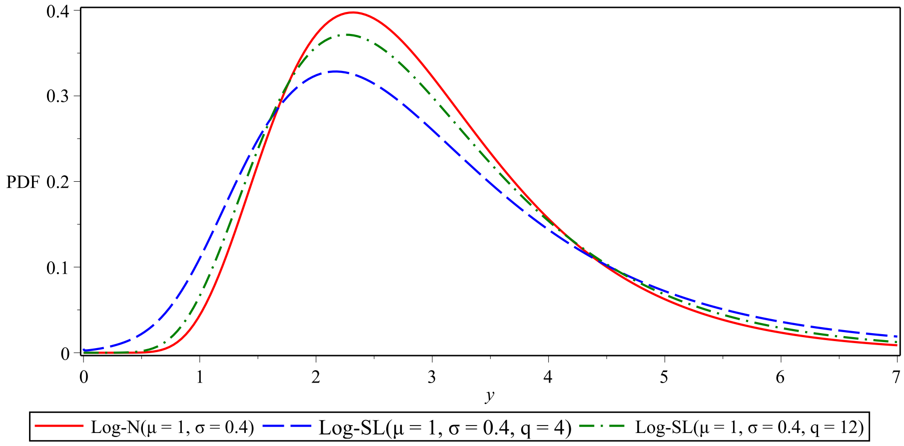

Figure 1. The PDF plots of the LoN and LoS distributions for various values of q.

From Figure 1, one can easily see that, the LoS distribution is heavier in tails than the LoN model. Heavy tails and less peak of the distribution are associated with smaller q. The limiting distribution of $LoS(\mu,\sigma,q)$, as $q \to \infty$ is $LoN(\mu,\sigma)$. Also, one can easily see that, when q increases the curve of $LoS(\mu,\sigma,q)$ approach to the curve of $LoN(\mu,\sigma)$. We can directly proof this result from the definition of $LoS(\mu,\sigma,q)$, since

$$Y = e^X \sim LoS(\mu,\sigma,q) \Longrightarrow X \sim S(\mu,\sigma,q)$$

and

$$Y = e^{Z/U^{1/q}} \Rightarrow Y = e^Z \text{ (at } q \to \infty) \Rightarrow Y \sim Log - N(\mu,\sigma).$$

3 Statistical Properties

Proposition 1. If $Y \sim LoS(\mu,\sigma,q)$, then the random variable $W = aY$ is also LoS distribution, say $W \sim LoS(\log(a) + \mu,\sigma,q)$, where a is a real number.

Proof.

$$\Pr(W < k) = \Pr(aY < k) = \Pr\left(Y < \frac{k}{a}\right)$$
$$= \int_{-\infty}^{\frac{k}{a}} \left(\frac{q}{\sigma\sqrt{2\pi}y}\int_0^1 e^{-\frac{(\log(y)-\mu)^2}{2\sigma^2}t^2} t^q dt\right) dy\ ;\ 0 < y < \infty.$$

Apply the transformation $W = aY$, then

$$\Pr(W < k) = \int_{-\infty}^{k} \left(\frac{aq}{\sigma w\sqrt{2\pi}}\int_0^1 e^{-\frac{(\log(\frac{w}{a})-\mu)^2}{2\sigma^2}t^2} t^q dt\right) \frac{1}{a}dw$$
$$= \int_{-\infty}^{k} \left(\frac{q}{\sigma w\sqrt{2\pi}}\int_0^1 e^{-\frac{(\log(w)-(\log(a)+\mu))^2}{2\sigma^2}t^2} t^q dt\right) dw,$$

where $0 < w < \infty$. The last term above, is the cumulative density function for $LoS(\log(a)+\mu,\sigma,q)$, so we have our result.

Proposition 2. If $Y \sim LoS(\mu,\sigma,q)$, then the random variable $W = \frac{1}{Y}$ is also LoS distribution, say $W \sim LoS(-\mu,\sigma,q)$.

Proof.

$$\begin{aligned}\Pr(W<k) &= \Pr(\frac{1}{Y}<k) = \Pr(Y>\frac{1}{k})\\ &= \int_{\frac{1}{k}}^{\infty}\left(\frac{q}{\sigma\sqrt{2\pi}y}\int_0^1 e^{-\frac{(\log(y)-\mu)^2}{2\sigma^2}t^2}t^q dt\right)dy;\ 0<y<\infty.\end{aligned}$$

Apply the transformation $W = 1/Y$, then

$$\begin{aligned}\Pr(W<k) &= -\int_k^0\left(\frac{wq}{\sigma\sqrt{2\pi}}\int_0^1 e^{-\frac{\left(\log(w^{-1})-\mu\right)^2}{2\sigma^2}t^2}t^q dt\right)\frac{1}{w^2}dw\\ &= \int_0^k\left(\frac{q}{\sigma w\sqrt{2\pi}}\int_0^1 e^{-\frac{(\log(w)+\mu)^2}{2\sigma^2}t^2}t^q dt\right)dw,\end{aligned}$$

where $0 < w < \infty$. The last term above, is the cumulative density function for $LoS(-\mu,\sigma,q)$, so we have our result.

Proposition 3. Assume $Y \sim LoS(\mu,\sigma,q)$, then the random variable $W = Y^a$ is also LoS distribution, say $W \sim LoS(a\mu,a\sigma,q)$.

Proof.

$$\begin{aligned}\Pr(W<k) &= \Pr(Y^a<k) = \Pr\left(Y<k^{1/a}\right)\\ &= \int_{-\infty}^{k^{1/a}}\left(\frac{q}{\sigma\sqrt{2\pi}y}\int_0^1 e^{-\frac{(\log(y)-\mu)^2}{2\sigma^2}t^2}t^q dt\right)dy,\ 0<y<\infty.\end{aligned}$$

Apply the transformation $W = Y^a$, then

$$\begin{aligned}\Pr(W<k) &= \int_0^k\left(\frac{q}{a\sigma w^{1/a}\sqrt{2\pi}}\int_0^1 e^{-\frac{\left(\frac{1}{a}\log(w)-\mu\right)^2}{2\sigma^2}t^2}t^q dt\right)w^{\frac{1}{a}-1}dw\\ &= \int_0^k\left(\frac{q}{a\sigma w\sqrt{2\pi}}\int_0^1 e^{-\frac{(\log(w)-a\mu)^2}{2a^2\sigma^2}t^2}t^q dt\right)dw,\end{aligned}$$

where $0 < w < \infty$. The last term above, is the cumulative density function for $LoS(a\mu,a\sigma,q)$, so we have our result.

Proposition 4. If $Y \sim LoS(\mu,\sigma,q)$, then the random variable $W = Y + c$ is said to have a shifted LoS model with support $y \in (c,\infty)$. The PDFof W can be expressed as

$$f(w;\mu,\sigma,q) = \frac{q}{\sigma\sqrt{2\pi}(w-c)}\int_0^1 e^{-\frac{(\log(w-c)-\mu)^2}{2\sigma^2}t^2}t^q dt;\ c<w<\infty.$$

Proof: The proof is easy, so we omitted it.

© 2023 NSP
Natural Sciences Publishing Cor.

4 Moments

The cth moment of the random variable $S = ZU^{-1/q}$ is given by Wang and Genton [29] as a form

$$E(S^c) = \begin{cases} 0 & \text{if } c \text{ is odd} \\ \frac{[(c-1)(c-3)....3.1]q}{q-c} & \text{if } c \text{ is even} \end{cases}, q > c. \tag{7}$$

Thus, the cth moment of the of the random variable $X = \mu + \sigma S \sim S(\mu, \sigma, q)$, is given as

$$E(X^c) = E([\mu + \sigma S]^c) = \sum_{i=0}^{c} \binom{c}{i} \mu^{c-i} \sigma^c E(S^c), \tag{8}$$

where $E(S^c)$ given in (7). For more detail about $E(S^c)$ and $E(X^c)$ (see Wang and Genton, [29]). The rth moment of the $LoS(\mu, \sigma, q)$ can be derived in the following Proposition.

Proposition 5. The rth moment of the random variable $Y \sim LoS(\mu, \sigma, q)$ can be formulated as

$$E(Y^r) = \sum_{k=0}^{\infty} \frac{r^k}{k!} \left(\sum_{c=0}^{k} \binom{k}{c} \sigma^c \mu^{k-c} E(S^c) \right); \; r = 1, 2, ... , \tag{9}$$

where $E(S^c)$ is the cth moment of the of the random variable $X \sim S(\mu, \sigma, q)$, given in (7).

Proof. From the definition of the random variable $Y = e^X$, one can easily get

$$E(Y^r) = E\left(e^{rX}\right) = M_X(r); \; r = 1, 2, 3, ... ,$$

where $M_X(r)$ is the moment generating function of the random variable $X = \mu + \sigma S \sim S(\mu, \sigma, q)$, then

$$E(Y^r) = M_X(r) = \sum_{k=0}^{\infty} \frac{r^k}{k!} E(X^k). \tag{10}$$

Substituting from (8) in (10), we get (9) which complete the proof.

Remarks

1. Putting $\mu = 0$ and $\sigma = 1$ in Equation (9), the rth moment of the standard LoS distribution can be derived as

$$E(Y^r) = \sum_{k=0}^{\infty} \frac{r^k}{k!} E(S^r); \; r = 1, 2, ...$$

2. Using Maple software package, the skewness and kurtosis of the $LoS(\mu, \sigma, q)$ at $\mu = 0.3$ and $\sigma = 0.8$ for different values of q can be given in Table 1.

Table 1. The skewness and kurtosis measures.

Measure $\downarrow$ $q \rightarrow$	1	2	3	4	5	6	7	8	9	10
Skewness	1.365	1.129	1.029	0.968	0.734	0.661	0.509	0.506	0.499	0.487
Kurtosis	2.998	2.963	2.954	2.863	2.657	2.651	2.649	2.641	2.598	2.579

As we can see, the proposed model can be used to discuss positively skewed data with playkurtic shapes.

© 2023 NSP
Natural Sciences Publishing Cor.

5 Multivariate LoS Distribution

In numerous practical situations, one encounters bivariate or multivariate data. Consequently, several researchers have endeavored to propose distributions tailored to handle the diverse forms of such data. Within this section, we delve into the probability density function (PDF) of the multivariate LoS distribution. In this sequence, we denote $k-$ dimensional multivariate normal distribution with mean vector $\mu \in \mathbb{R}^k$ and covariance matrix $\Sigma \in \mathbb{R}^{k\times k}$ by $N_k(\mu,\Sigma)$, the PDF can be proposed as

$$\phi_k(\mathbf{x};\boldsymbol{\mu},\Sigma) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} e^{-\frac{(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}}, \quad \mathbf{x}\in\mathbb{R}^k. \tag{11}$$

Tarmast [30] defined the multivariate LoN distribution as the distribution of the random vector $Y = [Y_1,Y_2,....,Y_k] \in \mathbb{R}^k$. By using the transformation $Y_i = \exp(X_i)$, where $X_i = [X_1,X_2,....,X_k]$ be a $k-$component random vector having a multivariate normal distribution with a mean vector $\mu \in \mathbb{R}^k$ and covariance matrix $\Sigma \in \mathbb{R}^{k\times k}$. The PDF of the random vector Y is given by

$$g_k(\mathbf{y};\boldsymbol{\mu},\Sigma) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \prod_{i=1}^{k} y_i^{-1} e^{-\frac{(\log(\mathbf{y})-\boldsymbol{\mu})^T\Sigma^{-1}(\log(\mathbf{y})-\boldsymbol{\mu})}{2}}, \quad 0<y_i<\infty, \tag{12}$$

where $\log(y) = [\log(y_1),\log(y_2),....,\log(y_k)]$ is a $k-$component column vector. Wang and Genton [29] defined the multivariate slash distribution as a random vector

$$\mathbf{X} = \mu + \overset{1/2}{\Sigma}\mathbf{Z}U^{-1/q}, \tag{13}$$

where $Z \sim N_k(0,I_k)$ is independent of $U \sim U(0,1)$. The PDF of the random vector X in (13) is given by

$$\Psi_k(\mathbf{x};\mu,\boldsymbol{\Sigma},q) = q\int_0^1 t^{q+k-1}\phi_k(\mathbf{x}t;\mu t,\boldsymbol{\Sigma})dt, \quad \mathbf{x}\in\mathbb{R}^k. \tag{14}$$

where $\phi_k(.)$ denotes the $k-$ dimensional multivariate normal distribution density function given in (11). When $\mu = 0$ and $\Sigma = I_k$, X in (13) has a standard form of a multivariate slash distribution, symbolically we write $SL_k(0,I_k,q)$. For more details about the multivariate slash distribution (see, Wang and Genton, [29]).

Definition 1. Assume $X = [X_1,X_2,...,X_k]$ be a $k-$component random vector having multivariate slash distribution with location parameter μ, positive definite scale matrix parameter Σ and tail parameter $q > 0$. Now, we use the transformation $Y_i = \exp(X_i)$ and define a $k-$component random vector $Y = [Y_1,Y_2,...,Y_k]$. The density of Y is multivariate LoS distribution, symbolically, we can write $Y \sim LoS_k(\mu,\Sigma,q)$, and it is easily shown to be

$$f_k(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},q) = q\prod_{i=1}^{k}(y_i)^{-1}\int_0^1 t^{q+k-1}\phi_k(t\log(\mathbf{y});\boldsymbol{\mu}t,\boldsymbol{\Sigma})dt, \quad 0<y_i<\infty. \tag{15}$$

where $\log(y) = [\log(y_1),\log(y_2),....,\log(y_k)]$ is a $k-$component column vector and $\phi_k(.)$ denotes the $k-$ dimensional multivariate normal distribution density function given in (11). The standard form of a multivariate LoS distribution, say $LoS_k(0,I_k,q)$ can be derived from (15) when $\mu = 0$ and $\Sigma = I_k$.

Special Cases

1. If $q \to \infty$, the PDF in (15) tends to the PDF of the multivariate LoN distribution given in (12).
2. If $k = 1$, the PDF in (15) tends to the PDF of the univariate LoS distribution given in(5).
3. If $k = 2$ in (15), then we obtain the bivariate LoS distribution and its PDF is given by

$$f_2(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},q) = q\sum_{i=1}^{2} y_i \int_0^1 t^{q+1}\phi_k(t\log(\mathbf{y});\boldsymbol{\mu}t,\boldsymbol{\Sigma})dt, \quad 0<y_1,y_2<\infty, \tag{16}$$

where $\log(y) = [\log(y_1),\log(y_2)]$ and $q > 0$.

© 2023 NSP
Natural Sciences Publishing Cor.

6 Univariate and Multivariate Extensions: Maximum Likelihood Estimation

There are typically several methods at one's disposal for estimating unknown parameters. One of the most widely recognized and frequently employed techniques is the maximum likelihood estimator (MLE). Estimators derived through this method possess valuable properties and can be employed to establish confidence intervals and perform various other statistical tests. In this case, we use the MLE method to estimate the $LoS(\mu,\sigma,q)$ parameters. Let $y_1,...,y_n$ be a data set modelled at the location scale by the $LoS(\mu,\sigma,q)$ distribution. In this Section, we evaluate the MLE of the model parameters from complete samples. The log-likelihood (L) function is represented by

$$L(\mu,\sigma,q) = n\log\left[\frac{q}{\sqrt{2\pi}}\right] - n\log(\sigma) - \sum_{i=1}^{n}\log(y_i) + \sum_{i=1}^{n}\log\int_0^1 e^{-\frac{(\log(y_i)-\mu)^2}{2\sigma^2}t^2} t^q dt. \tag{17}$$

On taking partial derivatives of the L function with respect to μ,σ and q and equating the derivatives to 0, we get

$$\sum_{i=1}^{n}\left(\frac{(\log(y_i)-\widehat{\mu})}{\widehat{\sigma}^2}\int_0^1 e^{-\frac{(\log(y_i)-\widehat{\mu})^2}{2\widehat{\sigma}^2}t^2} t^{\widehat{q}+2}dt\right)\left(\int_0^1 e^{-\frac{-(\log(y_i)-\widehat{\mu})^2}{2\widehat{\sigma}^2}t^2} t^{\widehat{q}}dt\right)^{-1} = 0, \tag{18}$$

$$-\frac{n}{\widehat{\sigma}} + \sum_{i=1}^{n}\left(\int_0^1 e^{-\frac{(\log(y_i)-\widehat{\mu})^2}{2\widehat{\sigma}^2}t^2} t^{\widehat{q}}dt\right)^{-1}\left(\frac{(\log(y_i)-\widehat{\mu})}{\widehat{\sigma}^3}\int_0^1 e^{-\frac{(\log(y_i)-\widehat{\mu})^2}{2\widehat{\sigma}^2}t^2} t^{q+2}dt\right) = 0 \tag{19}$$

and

$$\frac{n}{\widehat{q}} + \sum_{i=1}^{n}\left(\int_0^1 e^{-\frac{(\log(y_i)-\widehat{\mu})^2}{2\widehat{\sigma}^2}t^2} t^{\widehat{q}}dt\right)^{-1}\left(\widehat{q}\int_0^1 e^{-\frac{(\log(y_i)-\widehat{\mu})^2}{2\widehat{\sigma}^2}t^2} t^{\widehat{q}-1}dt\right) = 0. \tag{20}$$

The system of non-linear equations can be solved numerically using Maple software to get $\widehat{\mu}_{MLE}, \widehat{\sigma}_{MLE}$ and $\widehat{q}_{MLE}$. The $\widehat{\mu}_{MLE}$ and $\widehat{\sigma}_{MLE}$ can be given by

$$\widehat{\mu} = \frac{\sum_{i=1}^{n}\omega_i\log(y_i)}{\sum_{i=1}^{n}\omega_i} \text{ and } \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\omega_i\left(\log(y_i)-\widehat{\mu}\right)^2, \tag{21}$$

where

$$\omega_i(s) = \left(\int_0^1 e^{-\frac{(st)^2}{2}} t^q dt\right)^{-1}\left(\int_0^1 e^{-\frac{(st)^2}{2}} t^{q+2}dt\right);\ s = |\log(y_i)-\widehat{\mu}|/\widehat{\sigma}. \tag{22}$$

If the random vector having $LoS_k(\mu,\Sigma,q)$, then the L function can be expressed as

$$L(\boldsymbol{\mu},\boldsymbol{\Sigma},q) = n\log[q] - \sum_{j=1}^{k}\sum_{i=1}^{n}\log(y_{ji}) + \sum_{i=1}^{n}\log\int_0^1 t^{q+k-1}\phi_k(t\log(\mathbf{y});\boldsymbol{\mu}t,\boldsymbol{\Sigma})dt. \tag{23}$$

7 Simulation

In this section, we have conducted a simulation study to evaluate how the performance of Maximum Likelihood Estimators (MLEs) varies with sample size n. We focus on estimating the unknown parameters of the $LoS(\mu,\sigma,q)$ distribution. Additionally, we explore both univariate and multivariate scenarios, as outlined below:

7.1 Univariate case

In this segment, we investigate the performance of the MLEs for $LoS(\mu,\sigma,q)$ distribution with respect to sample size n. The evaluation is based on a simulation study: Generate 1000 samples of size n from $LoS(\mu,\sigma,q)$ distribution; calculate the MLEs for the 1000 samples, say $(\widehat{\mu}_i,\widehat{\sigma}_i,\widehat{q}_i)$ for $i=1,2,....,1000$; compute the biases and mean squared errors (MSE) where $\text{Bias}(n) = \frac{1}{1000}\sum_{i=1}^{1000}(\overset{\wedge}{\theta}_i-\theta)$ and $\text{MSE}(n) = \frac{1}{10500}\sum_{i=1}^{1000}(\overset{\wedge}{\theta}_i-\theta)^2$. Repeat these steps for $n=20,50,200,300,500$ with $LoS(0.6,0.7,0.8)$ and $LoS(0.5,0.7,1.9)$. Table 2 show how the biases and MSEs decrease with respect to the change of the sample size.

© 2023 NSP
Natural Sciences Publishing Cor.

Table 2. Simulation results for the LoS distribution parameters.

		$LoS(0.6,0.7,0.8)$			$LoS(0.5,0.7,1.9)$		
n	Parameter	Bias	MSE	MRE	Bias	MSE	MRE
20	μ	0.23846363	0.13654294	0.15386441	0.38967159	0.34283456	0.37762549
	σ	0.43894531	0.28468708	0.32745513	0.53027643	0.32046535	0.43966416
	q	0.53846815	0.34095681	0.42057641	0.28468149	0.17745321	0.23098462
50	μ	0.19364710	0.07485341	0.12056497	0.29548252	0.25497632	0.31745780
	σ	0.37743052	0.20437865	0.26635480	0.37458309	0.21835573	0.33096645
	q	0.41486820	0.25004682	0.33484651	0.23264571	0.13946830	0.20947675
100	μ	0.13394634	0.00332835	0.07486344	0.21994689	0.18620835	0.23087541
	σ	0.27496391	0.14993655	0.18454731	0.27749510	0.17759462	0.21047634
	q	0.29474640	0.18554751	0.21764505	0.17745381	0.10377642	0.14404639
200	μ	0.08452854	0.00047527	0.00450764	0.15548363	0.12885649	0.15937728
	σ	0.18354951	0.08846836	0.12548615	0.18703062	0.12056783	0.14009527
	q	0.18846405	0.11946542	0.13846655	0.12995642	0.02284681	0.10443771
300	μ	0.02437549	0.00008947	0.00086534	0.11376074	0.03938614	0.13927465
	σ	0.11047531	0.00443845	0.02275381	0.08846581	0.07949974	0.05548621
	q	0.12846554	0.03054682	0.05547538	0.05548360	0.00544375	0.01048654
500	μ	0.00043785	0.00000038	0.00008165	0.00520483	0.00275493	0.07745921
	σ	0.00774953	0.00069465	0.00438461	0.00365489	0.00054037	0.00284558
	q	0.02945781	0.00054792	0.00659462	0.00194693	0.00060947	0.00026486

Table 3. Simulation results for the multivariate LoS distribution parameters.

		$LoS_2(0.6,0.9,0.8,1.2,2.6)$			$LoS_2(1.5,1.7,1.9,1.3,1.8)$		
n	Parameter	Bias	MSE	MRE	Bias	MSE	MRE
20	μ_1	0.88462482	0.72365495	0.93196825	0.59231695	0.43196309	0.51030289
	μ_2	0.63484650	0.58996285	0.65320395	0.45039557	0.34028741	0.41036948
	σ_1	0.34072075	0.25312585	0.28531957	0.46329871	0.31039857	0.35202894
	σ_2	0.53076381	0.41096582	0.49306552	0.35031258	0.26736985	0.32098257
	q	0.34745095	0.29832484	0.31023982	0.27036281	0.24302897	0.22993647
50	μ_1	0.62548504	0.53012846	0.72139568	0.41395348	0.36032695	0.39203298
	μ_2	0.47740719	0.44231089	0.51309559	0.32087169	0.28031894	0.30298567
	σ_1	0.28464997	0.21202395	0.26387128	0.31032985	0.27039698	0.28810394
	σ_2	0.41946945	0.34295348	0.38663294	0.26032698	0.21093874	0.23019859
	q	0.24364841	0.20713907	0.22836428	0.21039087	0.18503298	0.16620874
100	μ_1	0.41046327	0.37103956	0.49326954	0.29312847	0.25032997	0.27719385
	μ_2	0.30947689	0.25531925	0.32039852	0.20139655	0.17032039	0.18803298
	σ_1	0.21548504	0.15397356	0.19832585	0.19032987	0.16930284	0.17032697
	σ_2	0.27493469	0.25319758	0.24196574	0.18893257	0.15023967	0.17712962
	q	0.19464965	0.14429307	0.16328549	0.15509689	0.12039592	0.10883017
200	μ_1	0.28496609	0.23185468	0.24963285	0.15530379	0.13302896	0.14330935
	μ_2	0.19406563	0.13320284	0.20123689	0.14302698	0.11082396	0.13392854
	σ_1	0.13984560	0.11933597	0.13395452	0.12037028	0.10223987	0.11096384
	σ_2	0.19453985	0.17302398	0.15563929	0.12029856	0.08823749	0.11930195
	q	0.13338850	0.07193028	0.10268874	0.10023987	0.06632987	0.04238749
300	μ_1	0.13493651	0.10320482	0.15329564	0.10239524	0.07829345	0.08203964
	μ_2	0.08946582	0.00931398	0.07125695	0.08231965	0.06620841	0.05507496
	σ_1	0.08756823	0.04239698	0.07032958	0.07035913	0.03018557	0.05930328
	σ_2	0.11263476	0.08732968	0.10218958	0.08259637	0.00370971	0.02296547
	q	0.07358551	0.00632987	0.05526987	0.00296385	0.00877129	0.00530148
500	μ_1	0.08533601	0.01289458	0.09654855	0.00128939	0.00054036	0.00089355
	μ_2	0.00328458	0.00029344	0.00212858	0.00020398	0.00080289	0.00062641
	σ_1	0.00077496	0.00029713	0.00052972	0.00050397	0.00035967	0.00070395
	σ_2	0.00883574	0.00073182	0.00725896	0.00083036	0.00012957	0.00059344
	q	0.00294672	0.00001395	0.00094128	0.00028996	0.00007496	0.0002398

© 2023 NSP
Natural Sciences Publishing Cor.

7.2 Multivariate case

In this Section, a simulation is discussed for the bivariate LoS model as a special case of the multivariate formula. Thus, we investigate the performance of the MLEs for the $LoS_2(\mu_1,\sigma_1,\mu_2,\sigma_2,q)$ distribution with respect to sample size n. The evaluation is based on a simulation study: Generate 1000 samples of size n from $Log-SL_2(\mu_1,\sigma_1,\mu_2,\sigma_2,q)$ distribution; calculate the MLEs for the 1000 samples, say $(\widehat{\mu}_{1i},\widehat{\sigma}_{1i},\widehat{\mu}_{2i},\widehat{\sigma}_{2i},\widehat{q}_i)$ for $i=1,2,....,1000$; compute the biases and MSE; repeat these steps for $n=20,50,100,200,500$ with $LoS_2(0.6,0.9,0.8,1.2,2.6)$ and $LoS_2(1.5,1.7,\ 1.9,1.3,1.8)$. Table 3 show how the biases and MSEs decrease with respect to the change of the sample size.

The simulation results for both univariate and multivariate cases are depicted in Table 2 and Table 3, respectively. These tables present the biases and mean square errors (MSEs) of the estimated parameters for the $LoS(\mu,\sigma,q)$ and $LoS_2(\mu_1,\sigma_1,\mu_2,\sigma_2,q)$ distributions. The results showcased in these tables demonstrate that the estimates for the parameters of both the $LoS(\mu,\sigma,q)$ and $LoS_2(\mu_1,\sigma_1,\mu_2,\sigma_2,q)$ distributions perform admirably. They exhibit minimal bias and commendable MSEs across all parameter configurations. Furthermore, as the sample size increases, the biases steadily approach zero, indicating that the estimates behave as asymptotically unbiased estimators. Additionally, the MSEs decrease with larger sample sizes, signifying that these estimators consistently provide accurate estimates for the $LoS(\mu,\sigma,q)$ and $LoS_2(\mu_1,\sigma_1,\mu_2,\sigma_2,q)$ parameters.

8 Real Data Analysis

In this section, we illustrate the empirical importance of the $Log-SL(\mu,\sigma,q)$ distribution using four applications to real data with making a comparison with well-known models such as exponential (E) and log-normal (Log-N) distributions. The fitted distributions are compared using some criteria namely, the maximized log-likelihood (-L), Akaike Information Criterion (AIC), correct Akaike information criterion (CAIC), bayesian information Criterion (BIC), Hannan-Quinn information criterion (HQIC) and Kolmogorov-Smirnov (KS) test and its p-value. For more details about this criteria see, Farooq et al. [31].

8.1 Data set (I)

The dataset labeled as (I) contains information on the survival times in months for 38 patients who succumbed to cervical cancer. You can find the data at the following link: www.ssc.ca/documents/case_studies/2002/cervical˜e.html. The specific survival times are listed below: 5.26, 6.64, 8.38, 9.80, 11.08, 11.18, 12.56, 12.66, 13.45, 14.14, 17.46, 17.52, 20.91, 21.67, 23.18, 25.74, 25.78, 32.55, 34.13, 37.55, 38.07, 38.70, 39.85, 41.88, 50.83, 51.16, 53.98, 55.96, 57.11, 62.50, 66.08, 67.82, 67.86, 70.55, 78.05, 82.78, 96.13, 100.67. A variety of non-parametric plots depicting Data Set I are showcased in Figure 2. These results can be found in Table 4.

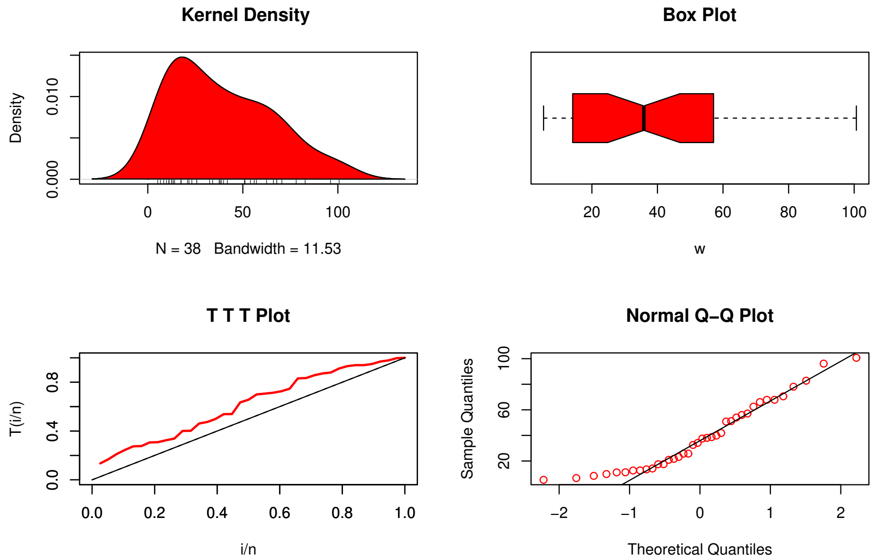

Figure 2. Non-parametric plots for dataset I.

© 2023 NSP
Natural Sciences Publishing Cor.

Table 4. The MLE, -L, AIC, BIC, CAIC, HQIC and KS (P-Value) values.

Model	$\hat{q}$	$\hat{\mu}$	$\hat{\sigma}$	**-L**	**AIC**	**BIC**	**CAIC**	**HQIC**	**KS(P-Value)**
$\mathbf{E}(q)$	38.991	$--$	$--$	177.21	356.41	358.05	356.52	356.99	0.112(0.154)
$\mathbf{Log-N}(\mu,\sigma)$	$--$	3.390	0.630	176.26	356.53	359.80	356.87	357.69	0.124(0.132)
$\mathbf{Log-SL}(\mu,\sigma,q)$	10.934	3.393	0.697	174.11	354.22	359.13	354.93	355.97	0.082(0.263)

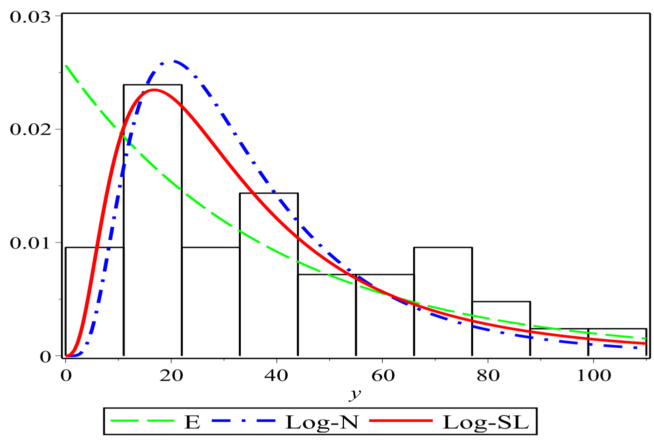

Figure 3. The estimated PDFs for data set I.

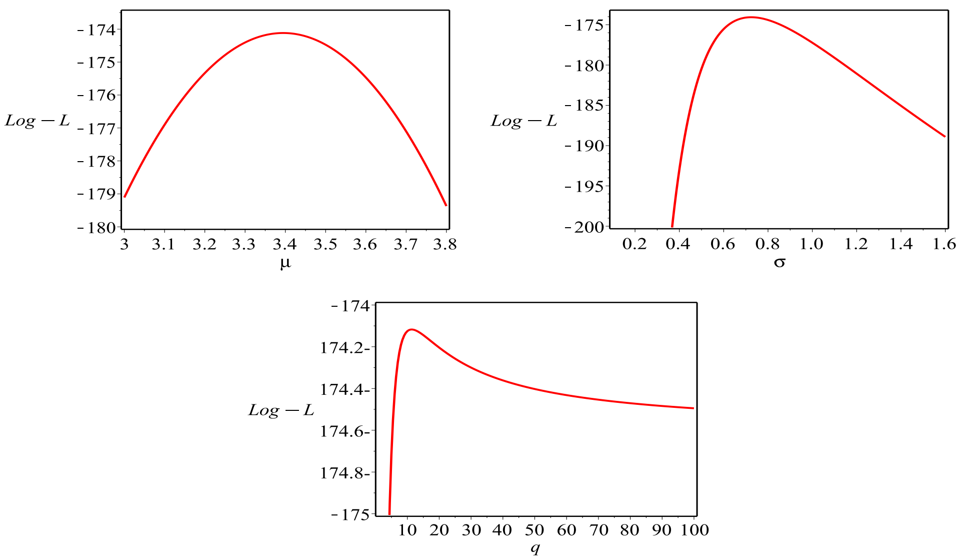

Figure 4. The profile of the log-likelihood for data set I.

Table 4 reveals that among the various tested distributions, the $Log-SL(\mu,\sigma,q)$ distribution stands out as the most suitable choice for fitting this data. This is evident as the $Log-SL(\mu,\sigma,q)$ distribution exhibits the lowest values for -L, AIC, CAIC, BIC, and HQIC. Additionally, Figures 3 and 4 provide insights into the estimated PDFs and the log-likelihood profile for each estimator based on the actual dataset. Figure 4 corroborates that MLE provides a distinct and consistent solution for all estimators.

8.2 Data set (II)

Data set (II) represents the annual flood discharge rates of the Floyd River, measured in cubic feet per second (ft^3/s) spanning from 1935 to 1973. This dataset was originally reported by Akinsete et al. [32]. It includes the following values:

© 2023 NSP
Natural Sciences Publishing Cor.

1460, 4050, 3570, 2060, 1300, 1390, 1720, 6280, 1360, 7440, 5320, 1400, 3240, 2710, 4520, 4840, 8320, 13900, 71500, 6250, 2260, 318, 1330, 970, 1920, 15100, 2870, 20600, 3810, 726, 7500, 7170, 2000, 829, 17300, 4740, 13400, 2940, 5660. Some non-parametric plots for Data Set II are shown in Figure 5.

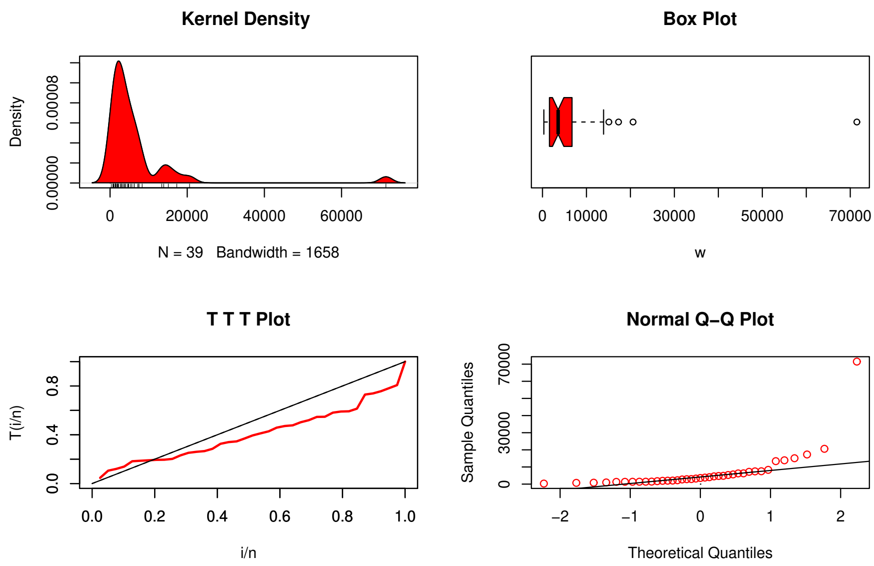

Figure 5. Non-parametric plots for data set II.

The results of the data fitting process can be found in Table 5.

Table 5. The MLE, -L, AIC, BIC, CAIC, HQIC and KS (P-Value) values.

Model	$\hat{q}$	$\hat{\mu}$	$\hat{\sigma}$	**-L**	**AIC**	**BIC**	**CAIC**	**HQIC**	**KS(P-Value)**
$\mathbf{E}(q)$	3.3×10^{-4}	$--$	$--$	399.57	797.13	795.47	797.02	796.53	0.162(0.001)
$\mathbf{Log-N}(\mu,\sigma)$	$--$	5.949	4.282	418.34	832.68	829.35	832.35	831.48	0.182(0.001)
$\mathbf{Log-SL}(\mu,\sigma,q)$	9.691	8.175	0.938	376.65	747.29	742.29	746.60	745.49	0.087(0.785)

Table 5 provides clear evidence that, among all the distributions tested, the $Log-SL(\mu,\sigma,q)$ distribution proves to be the most suitable for fitting this data. This conclusion is supported by the fact that the $Log-SL(\mu,\sigma,q)$ distribution exhibits the lowest values for -L, AIC, CAIC, BIC, and HQIC. Furthermore, Figures 6 and 7 offer valuable insights into the estimated PDFs and the profile of log-likelihood for each estimator, based on the actual data set II.

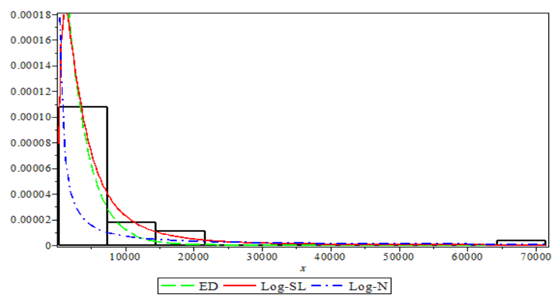

Figure 6. The estimated PDFs for data set II.

© 2023 NSP
Natural Sciences Publishing Cor.

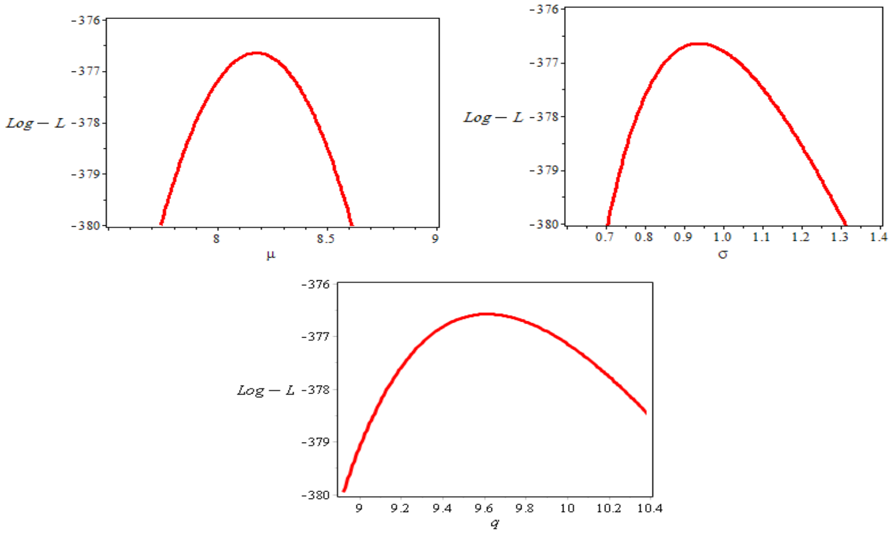

Figure 7. The profile of the log-likelihood for data set II.

Figure 7 demonstrates that MLE provides a singular and consistent solution for all estimators.

8.3 Data set (III)

Data set (III) consists of annual maximum flow values (measured in m^3/s) recorded at Kinrara, Spey, spanning the years 1952 to 1982. There are a total of 31 records in this dataset, as reported by Ahmad et al. [33]. The recorded values are as follows: 89.8, 109.1, 202.2, 146.3, 212.3, 116.7, 109.1, 80.7, 127.4, 138.8, 283.5, 85.6, 105.5, 118.0, 387.8, 80.7, 165.7, 111.6, 134.4, 131.5, 102.0, 104.3, 242.5, 214.8, 144.6, 114.2, 98.3, 102.8, 104.3, 196.2, 143.7. Figure 8 displays several non-parametric plots representing Data Set III.

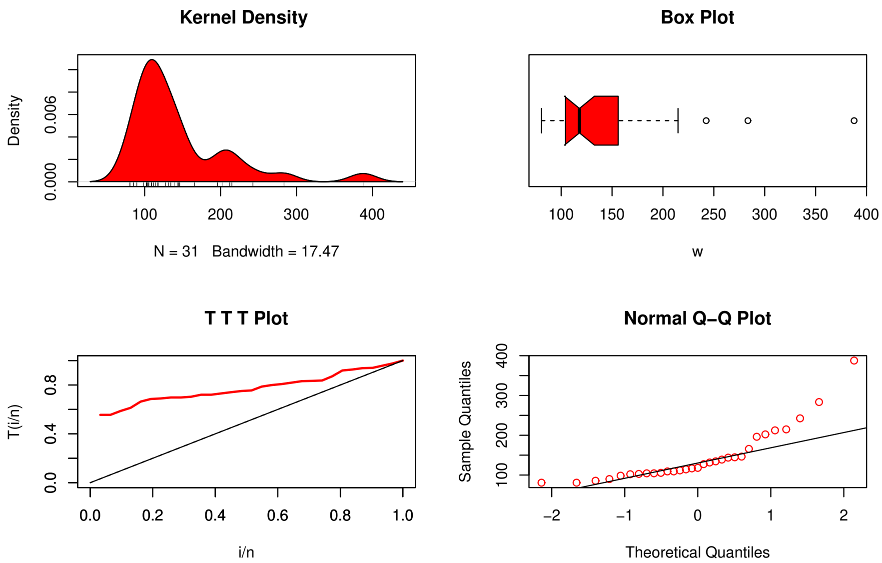

Figure 8. Non-parametric plots for data set III.

The results of the data fitting process can be found in Table 6.

Table 6. The MLE, -L, AIC, BIC, CAIC, HQIC and KS (P-Value) values.

Model	$\hat{q}$	$\hat{\mu}$	$\hat{\sigma}$	**-L**	**AIC**	**BIC**	**CAIC**	**HQIC**	**KS(P-Value)**
$\mathbf{E}(q)$	0.47×10^{-2}	$--$	$--$	187.29	372.57	371.14	372.43	372.10	0.213(0.000)
$\mathbf{Log-N}(\mu,\sigma)$	$--$	5.032	1.787	199.16	372.57	371.14	372.43	372.10	0.182(0.000)
$\mathbf{Log-SL}(\mu,\sigma,q)$	13.991	4.908	0.371	165.53	325.06	320.76	324.18	323.66	0.122(0.058)

Table 6 clearly indicates that, among the various distributions tested, the $Log-SL(\mu,\sigma,q)$ distribution stands out as the most appropriate choice for fitting this data. This conclusion is supported by the fact that the $Log-SL(\mu,\sigma,q)$ distribution exhibits the lowest values for -L, AIC, CAIC, BIC, and HQIC. Additionally, Figures 9 and 10 provide valuable insights into the estimated Probability Density Functions (PDFs) and the profile of log-likelihood for each estimator, based on the actual data set III.

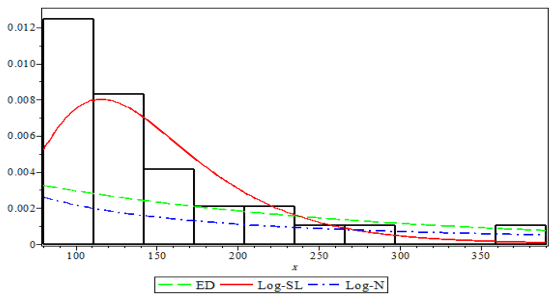

Figure 9. The estimated PDFs for data set III.

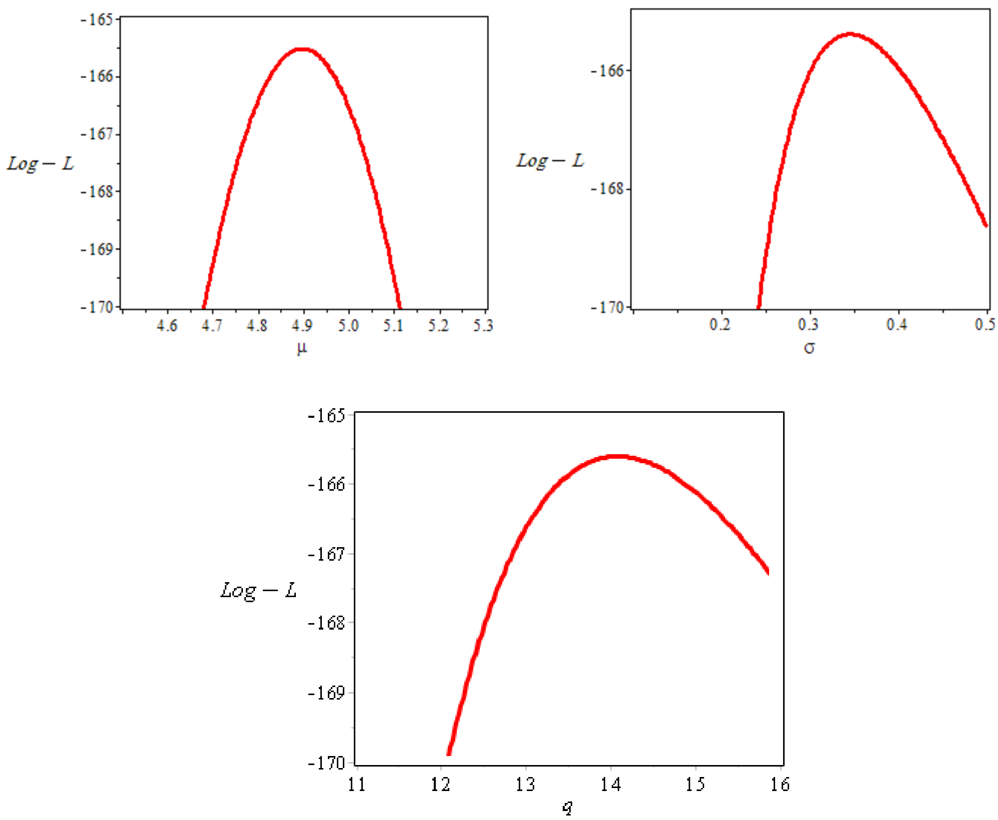

Figure 10. The profile of the log-likelihood for data set III.

Figure 10 clearly demonstrates that MLE yields a distinct and singular solution for each estimator.

© 2023 NSP
Natural Sciences Publishing Cor.

8.4 Data set (IV)

Data set (IV) represents the maximum amount of rainfall in millimeters for the Pakistani city of Kalat. This dataset spans a 30-year period from 1981 to 2010 and comprises 30 values of maximum rainfall. The data was originally reported by Ahmad et al. [33]. The recorded values are as follows: 32.77, 58.65, 60.71, 64.01, 42.6, 75.8, 88.6, 90.1, 97.9, 105.6, 73.1, 76.6, 78.5, 58.3, 122.5, 57.8, 546, 125.8, 50.5, 45.9, 21.7, 45.5, 38, 75.4, 168.2, 72.9, 95.8, 133.4, 71.9, 28. Figure 11 exhibits a collection of non-parametric plots illustrating Data Set III.

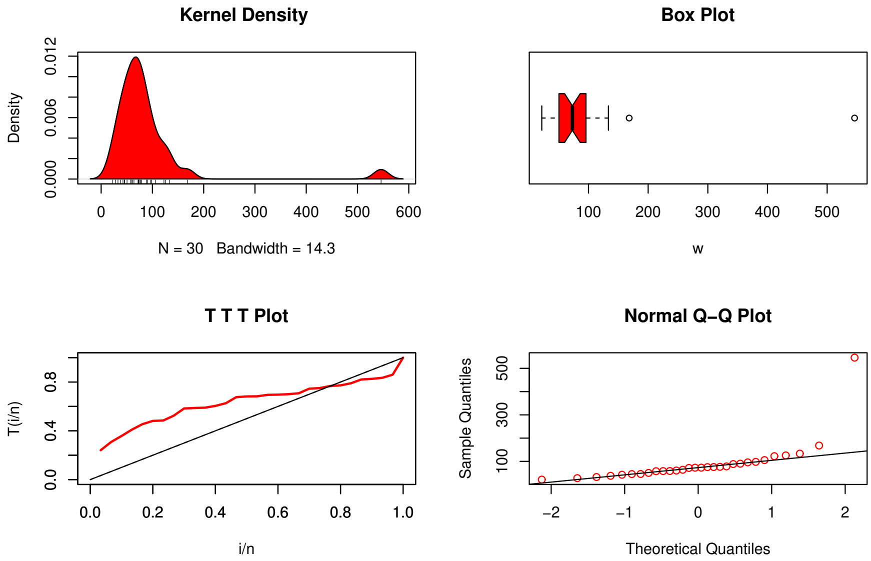

Figure 11. Non-parametric plots for data set IV.

The results of the data fitting process can be found in Table 7.

Table 7. The MLE, -L, AIC, BIC, CAIC, HQIC and KS (P-Value) values.

Model	$\hat{q}$	$\hat{\mu}$	$\hat{\sigma}$	**-L**	**AIC**	**BIC**	**CAIC**	**HQIC**	**KS (P-Value)**
$\mathbf{E}(q)$	0.013	$--$	$--$	165.42	328.83	327.43	328.69	328.39	0.138(0.221)
$\mathbf{Log-N}(\mu,\sigma)$	$--$	4.129	1.074	162.86	321.71	318.91	321.27	320.81	0.129(0.352)
$\mathbf{Log-SL}(\mu,\sigma,q)$	4.931	4.198	0.407	153.88	301.76	297.55	300.83	300.41	0.117(0.834)

Table 7 clearly demonstrates that, among the various distributions examined, the $Log-SL(\mu,\sigma,q)$ distribution is the most suitable choice for fitting this data. This conclusion is supported by the fact that the $Log-SL(\mu,\sigma,q)$ distribution displays the lowest values for -L, AIC, CAIC, BIC, and HQIC. Furthermore, Figures 12 and 13 provide valuable insights into the estimated PDFs and the profile of log-likelihood for each estimator, based on the actual data set IV.

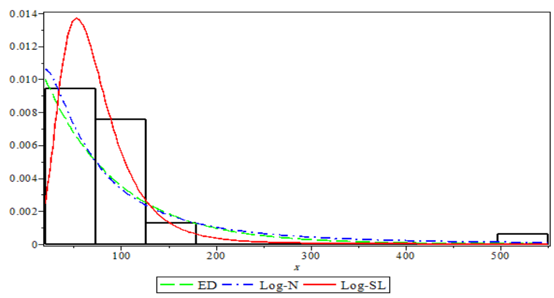

Figure 12. The estimated PDFs for data set IV.

© 2023 NSP
Natural Sciences Publishing Cor.

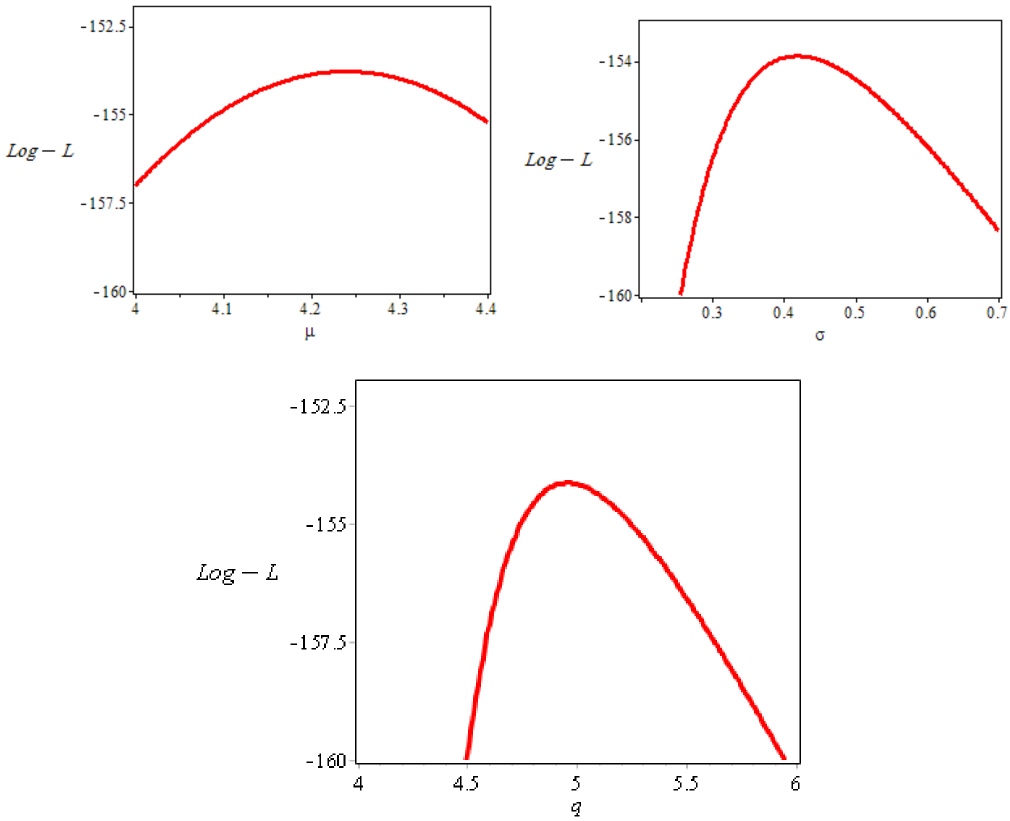

Figure 13. The profile of the log-likelihood for data set VI.

Figure 13 makes it evident that MLE yields a distinct and singular solution for each estimator.

9 Conclusion

An innovative category of both univariate and multivariate log-slash distributions was introduced. This new class was envisioned to offer significant advantages when dealing with datasets characterized by longer tails than those accommodated by conventional distributions like the log-normal distribution and other well-established options in the field of statistical literature. Within this newly introduced category, various distribution properties were thoroughly examined, and comprehensive density calculations were provided, allowing the deduction of a multitude of statistical characteristics. The maximum likelihood estimation method was employed for parameter estimation. Additionally, a comprehensive simulation study was conducted to evaluate the performance of parameter estimation, considering both bias and mean squared errors. Beyond the realm of simulations, a meticulous analysis of three real-world datasets was carried out, showing strong alignment with the proposed model. In summary, the findings indicated that the log-slash distribution outperformed existing heavy-tailed distributions, making it a more suitable choice for effectively modeling a wide range of real-world data spanning various sectors and industries.

Acknowledgement:

The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2023/01/25641).

References

[1] M. H. Tahir, G. M. Cordeiro, Compounding of distributions: a survey and new generalized classes, Journal of Statistical Distributions and Applications, 3, 1-35, (2016).

© 2023 NSP
Natural Sciences Publishing Cor.

[2] D. Bhati, S. Ravi, On generalized log-Moyal distribution: a new heavy tailed size distribution, Insurance: Mathematics and Economics, 79, 247-259, (2018).
[3] W. H. Rogers, J. W. Tukey, Understanding some long-tailed symmetrical distributions, Statistica Neerlandica, 26(3), 211-226, (1972).
[4] K. Kafadar, A biweight approach to the one-sample problem, Journal of the American Statistical Association, 77(378), 416-424, (1982).
[5] Z. Shah, A. Ali, M. Hamraz, D. M. Khan, Z. Khan, M. EL-Morshedy, Z. Almaspoor, A new member of TX family with applications in different sectors, Journal of Mathematics, 2022.
[6] S. Morgenthaler, Robust confidence intervals for a location parameter: the configural approach, Journal of the American Statistical Association, 81(394), 518-525, (1986).
[7] M. El-Morshedy, A Discrete Linear-Exponential Model: Synthesis and Analysis with Inference to Model Extreme Count Data, Axioms, 11(10), 531, (2022).
[8] K. L. Lange, R. J. Little, J. M. Taylor, Robust statistical modeling using the t distribution, Journal of the American Statistical Association, 84(408), 881-896, (1989).
[9] Z. Ahmad, Z. Almaspoor, F. Khan, M. El-Morshedy, On predictive modeling using a new flexible Weibull distribution and machine learning approach: Analyzing the COVID-19 data, Mathematics, 10(11), 1792.
[10] M. Jamshidian, A note on parameter and standard error estimation in adaptive robust regression, Journal of statistical computation and simulation, 71(1), 11-27, (2001).
[11] A. Khan, A. Ali, M. Ijaz, M. Azeem, M. El-Morshedy, An Exponential Ratio Type Estimator of the Population Mean in The Presence of Non-Response Using Double Sampling, (2023).
[12] D. N. Kashid, S. R. Kulkarni, Subset selection in multiple linear regression with heavy tailed error distribution, Journal of Statistical Computation and Simulation, 73(11), 791-805, (2003).
[13] J. Reyes, H. W. Gmez, H. Bolfarine, Modified slash distribution, Statistics, 47(5), 929-941,(2013).
[14] M. M. Ferede, S. Mwalili, G. Dagne, S. Karanja, W. Hailu, M. El-Morshedy, A. Al-Bossly, A Semiparametric Bayesian Joint Modelling of Skewed Longitudinal and Competing Risks Failure Time Data: With Application to Chronic Kidney Disease, Mathematics, 10(24), 4816, (2022).
[15] Z. G. Moradi, M. Arashi, O. L. A. Y.Arslan, A. Iranmanesh, A, A new family of multivariate slash distributions, Communications in Statistics-Theory and Methods, 46(7), 3264-3275, (2017).
[16] J. A. Almamy, M. Ibrahim, M. S. Eliwa, S. Al-mualim, H. M.Yousof, The two-parameter odd Lindley Weibull lifetime model with properties and applications, Int. J. Stat. Probab, 7(2), 57-68, (2018).
[17] E. A. Ahmed, Z. Ali Alhussain, M. M. Salah, H. Haj Ahmed, M. S. Eliwa, Inference of progressively type-II censored competing risks data from Chen distribution with an application, Journal of Applied Statistics, 47(13-15), 2492-2524, (2020).
[18] M. Alizadeh, A. Z. Afify, M. S. Eliwa, S. Ali, The odd log-logistic Lindley-G family of distributions: properties, Bayesian and non-Bayesian estimation with applications, Computational statistics, 35, 281-308, (2020).
[19] E. Altun, H. Alqifari, M. S. Eliwa, A novel approach for zero-inflated count regression model: Zero-inflated Poisson generalized-Lindley linear model with applications, AIMS Mathematics, 8(10), 23272-23290, (2023).
[20] A. S. Eldeeb, M. Ahsan-ul-Haq, M. S. Eliwa, A discrete Ramos-Louzada distribution for asymmetric and over-dispersed data with leptokurtic-shaped: Properties and various estimation techniques with inference, AIMS Mathematics, 7(2), 1726-1741, (2022).
[21] M. S. Eliwa, E. A. Ahmed, Reliability analysis of constant partially accelerated life tests under progressive first failure type-II censored data from Lomax model: EM and MCMC algorithms, AIMS Math, 1, 29-60, (2023).
[22] M. S. Eliwa, Z. Ahmad, M. H. Alabdulhadi, F. Khan, A New Beta Power Generator for Continuous Random Variable: Features and Inference to Model Asymmetric Data, (2023).
[23] M. S. Eliwa, B. T. Alhumaidan, R. N. Alqefari, A discrete mixed distribution: Statistical and reliability properties with applications to model COVID-19 data in various countries, Mathematical biosciences and engineering: MBE, 20(5), 7859-7881, (2023).
[24] H. Haj Ahmad, M. M. Salah, M. S. Eliwa, Z. Ali Alhussain, E. M. Almetwally, E. A. Ahmed, Bayesian and non-Bayesian inference under adaptive type-II progressive censored sample with exponentiated power Lindley distribution, Journal of Applied Statistics, 49(12), 2981-3001, (2022).
[25] L. Handique, S. Chakraborty, The Poisson Transmuted-G Family of Distributions: Its Properties and Applications, Pak.j.stat.oper.res. DOI: http://dx.doi.org/10.18187/pjsor.v17i1.3647, (2020).
[26] M. S. Murshed,Statistical analysis to extreme values with applications to hydrologic events. Unpublished PhD (Statistics) Thesis, Department of Statistics Graduate School of Chonnam National University, Korea, (2011).
[27] N. L. Johnson, S. Kotz, N. Balakrishnan, Continuous univariate distributions, volume 2 (Vol. 289), John wiley & sons, (1995).
[28] E. Whittaker, T. Edmund, An expression of certain known functions as generalized hypergeometric functions, American Mathematical Society, 10, 125-134, (1904).
[29] J. Wang, M. G. Genton, The multivariate skew-slash distribution, Journal of Statistical Planning and Inference, 136(1), 209-220, (2006).
[30] G. Tarmast, Multivariate log-normal distribution, International Statistical Institute: Seoul 53rd Session, 210, (2001).
[31] M. Farooq, M. Ijaz, M. Atif, T. Abushal, M. El-Morshedy, Statistical analysis of Covid-19 mortality rate via probability distributions, PLoS One, 17(10), e0274133, (2022).
[32] A. Akinsete, F. Famoye, C. Lee, The beta-Pareto distribution, Statistics, 42(6), 547-563, (2008).
[33] M. I. Ahmad, C. D. Sinclair, A. Werritty, Log-logistic flood frequency analysis, Journal of Hydrology, 98(3-4), 205-224, (1988).

© 2023 NSP
Natural Sciences Publishing Cor.