

On the Inverted Beta Generated Slash Distribution: Properties and Application

Abstract

In this paper, we introduce a new distribution called inverted beta generated slash distribution which is the inverted form of the beta generated slash distribution. The beta generated slash distribution is first introduced and then the inverted form of this distribution is established which is named as inverted beta generated slash distribution. The explicit expressions for pdf, cdf, moments, skewness, kurtosis, median are derived and their variation with different values of the parameters are studied. The hazard rate function assumes different shapes depending on the values of the parameters. A few additional properties such as moment generating function, Mills Ratio, Lorenz and Bonferroni curves, order statistics, hazard rate function of the proposed distribution are also explored. The method of maximum likelihood is used to estimate the unknown parameters of this distribution and a simulation study is conducted to check the performance of these estimates. Finally, the proposed distribution is applied to a data set to check the flexibility of the model and the goodness-of-fit of the proposed distribution is compared with three other competing distribution to show its flexibility and advantage particularly in modeling heavy-tailed data sets.

Keywords: Inverted Beta Generated Slash distribution, Slash distribution, Additive property, Estimation, Simulation, Real-life application.

1. Introduction

An inverted distribution, also known as a reversed distribution, is a distribution of the reciprocal of a random variable. If the probability distribution of any random variable X is given, then the distribution of the reciprocal $Y = \frac{1}{X}$ can be easily obtained. If X is a continuous random variable with probability density function $f(x)$ and cumulative density function $F(x)$, then the cumulative distribution function of the reciprocal $G(y)$ is given by:

$$G(Y) = P(Y \leq y) = P\left(X \geq \frac{1}{Y}\right) = 1 - P\left(X < \frac{1}{Y}\right) = 1 - F\left(\frac{1}{Y}\right)$$

The probability density function of Y is obtained as the derivative of cumulative distribution function as shown below:

$$g(y) = \frac{1}{y^2} f\left(\frac{1}{y}\right)$$

The inverted distributions are widely used in many fields such as medical research, econometrics, biological sciences etc. It is also used in financial literature, environmental studies, survival and reliability theory. The statistical literature contains several extensions of inverted distributions, for example, Inverse Weibull distribution by Abd (2013), Inverted Pareto Type I distribution by Abd (2013), Inverted Kumaraswamy distribution by Abd et al. (2017), Inverted Wishart distribution by Dietrich (1997) etc. The inverted beta generated slash distribution is a three-parameter family of continuous probability distributions, which is the distribution of the reciprocal of a variable distributed according to beta generated slash probability law. Due to the importance of inverted distributions, in this paper, we provide an inverse form of the beta generated slash distribution introduced by Bhattacharjee et al. (2024). Slash distribution are defined by Rogers and Tukey (1972) as the ratio of standard normal variate to the uniform random variate which can be stochastically represented as:

$$Y = \frac{X}{U^{\frac{1}{q}}}$$

where X is a standard normal variate and $U \sim Uniform(0,1)$. The shape parameter $q > 0$ controls the kurtosis of the distribution. Since the last decade, various population distributions have been explored with the help of slash construction idea. For example, the slashed versions for the epsilon half-normal by Gui et al. (2013), the logistic distribution by Punathumparambath (2011) and the Birnbaum–Saunders distribution by Gómez et al. (2009) etc. The modified version of the slash distribution has been proposed by Reyes et al. (2013). The beta generated slash distribution (BGS1) is an extension of beta distribution through the slash construction idea and the proposed distribution is the inverted form of this beta generated slash distribution. We shall say that X follows the BGS1 distribution with parameters a, b and q or $X \sim BGS1(a,b,q)$ if it can be stochastically expressed as

$$X = \frac{Z}{U^{\frac{1}{q}}}$$

where $Z \sim beta(a,b)$ and $U \sim Uniform(0,1)$ and are distributed independently of each other.

The paper is organised as follows. Section 2 introduces the density function of the proposed distribution. Expressions for pdf, cdf, various descriptive statistics are derived and behaviour of the curve of the proposed distribution for varying values of the parameters graphically are shown in section 3. The maximum likelihood estimation of the parameters of the distribution are dealt with in section 4. In section 5, some stochastic simulations are performed to illustrate the behaviour of the parameters of the proposed distribution. In section 6, the proposed model is applied to data set on failure times to exhibit the potential of the distribution in modeling real-life data sets. The findings of the paper are finally summarized in section 7.

2. Definition and derivation of the Inverted Beta Generated Slash distribution

In this section, the pdf and cdf of the proposed distribution have been derived. These results have been presented under Theorem 1 as shown below:

Theorem 1. If a random variable $X \sim BGS1(a,b,q)$ then $Y = \frac{1}{X} \sim IBGS1(a,b,q)$, whose probability density function is given by:

$$g(y; a, b, q) = \begin{cases} \frac{qy^{q-1}}{\beta(a,b)}\beta(a+q, b), & 0 \leq y < 1 \\ \frac{qy^{q-1}}{\beta(a,b)}\beta(\frac{1}{y}; a+q, b), & 1 \leq y < \infty \end{cases}$$

And the cumulative distribution function is given by:

$$G(y; a, b, q) = \begin{cases} \frac{\beta(a+q,b)y^q}{\beta(a,b)}, & 0 \leq y < 1 \\ \frac{\beta(\frac{1}{y}, a+q,b)y^q - \beta(a+q,b) - \beta(\frac{1}{y}, a,b)}{\beta(a,b)} + 1, & 1 \leq y < \infty \end{cases}$$

Proof: Let $X \sim BGS1(a, b, q)$ which can be stochastically expressed as:

$$X = \frac{Z}{U^{\frac{1}{q}}}$$

where $Z \sim beta(a,b)$. The pdf of Z is given by:

$$f(z; a, b) = \frac{z^{a-1}(1-z)^{b-1}}{\beta(a, b)}, \quad 0 \leq z \leq 1$$

Suppose

$$W = U \implies Z = XW^{\frac{1}{q}}$$

\therefore

$$\begin{aligned} f_X(x, w) &= f_{x,u}(xw^{\frac{1}{q}}, w)|J| \\ &= \frac{1}{\beta(a, b)} z^{a-1} (1-z)^{b-1} w^{\frac{1}{q}} \\ &= \frac{1}{\beta(a, b)} z^{a-1} w^{\frac{a}{q}} (1-xw^{\frac{1}{q}})^{b-1} \end{aligned}$$

Hence, the marginal distribution function of X is given by:

$$f(x) = \begin{cases} f_1(x), & 0 \leq x < 1 \\ f_2(x), & 1 \leq x < \infty \end{cases} \quad (1)$$

where

$$\begin{aligned} f_1(x) &= \frac{x^{a-1}}{\beta(a, b)} \int_0^1 w^{\frac{a}{q}} (1-xw^{\frac{1}{q}})^{b-1} dw \\ &= \frac{q}{\beta(a, b)x^{q+1}} \beta(x; a+q, b) \end{aligned} \quad (2)$$

$\beta(x; a+b, q)$ being the incomplete beta function.

$$\begin{aligned} f_2(x) &= \frac{x^{a-1}}{\beta(a, b)} \int_0^{\frac{1}{x^q}} w^{\frac{a}{q}} (1-xw^{\frac{1}{q}})^{b-1} dw \\ &= \frac{q}{\beta(a, b)x^{q+1}} \beta(a+q, b) \end{aligned} \quad (3)$$

and cdf of X is

$$F(X) = \begin{cases} F_1(X), & 0 \leq y < 1 \\ F_2(X), & 1 \leq y < \infty \end{cases} \quad (4)$$

Where

$$\begin{aligned} F_1(X) &= P(X \leq x) \\ &= \int_0^x \frac{q}{\beta(a, b)x^{q+1}} \beta(x; a+q, b) dx \\ &= \frac{q}{\beta(a, b)} \int_0^x \beta(y; a+q, b) x^{-(q+1)} dy \\ &= \frac{\beta(x; a, b)}{\beta(a, b)} - x^{-q} \frac{\beta(x; a+q, b)}{\beta(a, b)} \end{aligned} \quad (5)$$

$$\begin{aligned} F_2(X) &= P(X \leq x) \\ &= \int_0^1 f_1(x) dx + \int_1^x f_2(x) dx \\ &= 1 - \frac{\beta(a+q, b)}{\beta(a, b)} + \frac{\beta(a+q, b)}{\beta(a, b)} (1-x^{-q}) \end{aligned} \quad (6)$$

Finally, the pdf of $Y = \frac{1}{X}$ is obtained as

$$\begin{aligned}
 g(y) &= I_{(0,1)}(y) \frac{1}{y^2} f\left(\frac{1}{y}\right) + (1 - I_{(0,1)})(y) \frac{1}{y^2} f\left(\frac{1}{y}\right) \\
 &= I_{(0,1)}(y) \frac{1}{y^2} \left[\frac{q}{\beta(a,b) \left(\frac{1}{y}\right)^{q+1}} \beta(a+q,b) \right] \\
 &\quad + (1 - I_{(0,1)})(y) \frac{1}{y^2} \left[\frac{q}{\beta(a,b) \left(\frac{1}{y}\right)^{q+1}} \beta\left(\frac{1}{y}; a+q,b\right) \right] \\
 &= I_{(0,1)}(y) \left[\frac{qy^{q-1}}{\beta(a,b)} \beta(a+q,b) \right] \\
 &\quad + (1 - I_{(0,1)})(y) \left[\frac{qy^{q-1}}{\beta(a,b)} \beta\left(\frac{1}{y}; a+q,b\right) \right] \tag{7}
 \end{aligned}$$

$$I_{(0,1)}(y) = \begin{cases} 1, & \text{if } 0 \leq y < 1 \\ 0, & \text{if } 1 \leq y < \infty \end{cases}$$

Again the cdf of Y is obtained as :

$$G(y) = \begin{cases} G_1(y), & 0 \leq y < 1 \\ G_2(y), & 1 \leq y < \infty \end{cases} \tag{8}$$

Where

$$\begin{aligned}
 G_1(y) &= P(Y \leq y) \\
 &= \int_0^y \frac{qy^{q-1}}{\beta(a,b)} \beta(a+q,b) \\
 &= \frac{\beta(a+q,b)y^q}{\beta(a,b)} \tag{9}
 \end{aligned}$$

$$\begin{aligned}
 G_2(y) &= P(Y \leq y) \\
 &= \int_0^1 g_1(y)dy + \int_1^y g_2(y)dy \\
 &= \frac{\beta\left(\frac{1}{y}, a+q,b\right)y^q - \beta(a+q,b) - \beta\left(\frac{1}{y}, a,b\right)}{\beta(a,b)} + 1 \tag{10}
 \end{aligned}$$

2.1 Location - Scale form of IBGSI Distribution

Another form of IBGSI distribution is the location - scale form. By applying the well known location - scale transformation, we get the location - scale transformed *IBGSL* variate as (refer to Genc et al. (2014) for details)

$$T = \mu + \sigma Y \tag{11}$$

where $Y \sim IBGSI(a,b,q)$, $0 < \mu < \infty$ and $\sigma > 0$. μ and σ are the location and scale parameters respectively. The location- scale form of *IBGSI* distribution has the following pdf:

$$f(t; a, b, q) = \begin{cases} \frac{q(t-\mu)^{(q-1)}}{\beta(a,b)\sigma^q} \beta(a+q,b), & \mu < T < \mu + \sigma \\ \frac{q(t-\mu)^{(q-1)}}{\beta(a,b)\sigma^q} \beta\left(\frac{\sigma}{t-\mu}, a+q,b\right), & \mu + \sigma \leq T < \infty \end{cases} \tag{12}$$

where a, b, q, μ, σ are the parameter vector. We denote it by $T \sim IBGSILS(a, b, q, \mu, \sigma)$.

The density plots of the pdf (7) for some selected values of the parameters are given in figure 1. These plots show the greater flexibility of the newly proposed distribution for different values of the parameters a, b and q .

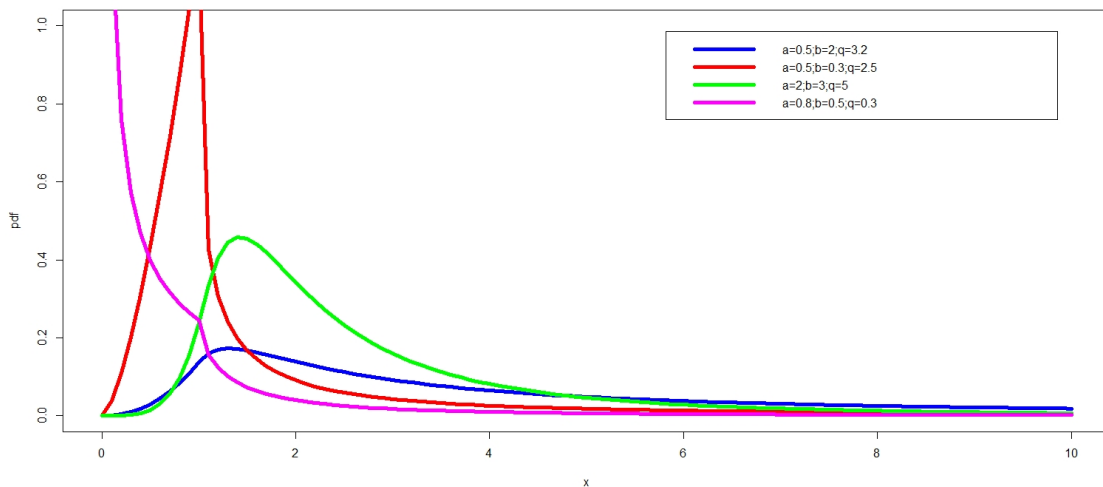


Figure 1 Probability density function plots of the IBGSI(a, b, q) distribution

The cdf plot of *IBGSI* distribution for different values of the parameters are shown in the following figure.

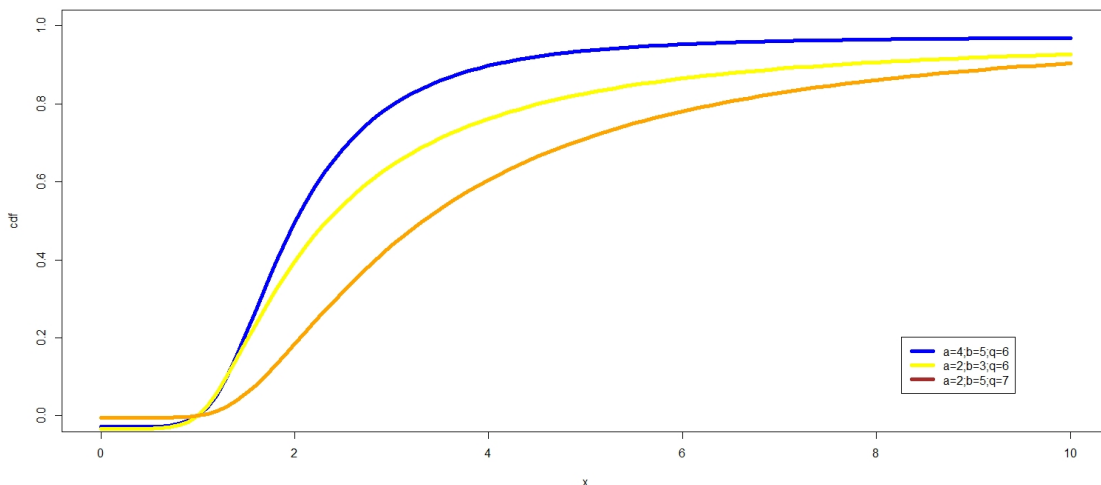


Figure 2 Cumulative distribution function plot of the *IBGSI*(*a,b,q*) distribution

3. Properties of *IBGSI* distribution

In this section, various descriptive statistics such as moments, skewness, kurtosis, median, mode along with some inequality measures and uncertainty measures of *IBGSI* distribution are derived.

3.1 Moments

If $Y \sim IBGSI(a, b, q)$, then the r^{th} raw moment of Y is given by:

$$\mu'_r = E(Y^r) = \int_0^\infty y^r f(y) dy$$

In particular,

$$\begin{aligned} \mu'_1 &= \frac{a+b-1}{(a-1)} \frac{q}{(q+1)}, a > 1 \\ \mu'_2 &= \frac{(a+b-1)(a+b-2)}{(a-1)(a-2)} \frac{q}{(q+2)}, a > 2 \\ \mu'_3 &= \frac{(a+b-1)(a+b-2)(a+b-3)}{(a-1)(a-2)(a-3)} \frac{q}{(q+3)}, a > 3 \\ \mu'_4 &= \frac{(a+b-1)(a+b-2)(a+b-3)(a+b-4)}{(a-1)(a-2)(a-3)(a-4)} \frac{q}{(q+4)}, a > 4 \end{aligned}$$

The measures of skewness and kurtosis, denoted by γ_1 and γ_2 , respectively are defined as

$$\gamma_1 = \frac{\mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1}{(\mu'_2 - \mu'^2_1)^{\frac{3}{2}}}$$

$$\gamma_2 = \frac{\mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2\mu'^2_1 - 3\mu'^4_1}{(\mu'_2 - \mu'^2_1)^2}$$

The skewness and kurtosis values for some selected values of a and b are displayed in table 1

Table 1 Skewness and Kurtosis values for different parameters

(a,b,q)	skewness	kurtosis
(4.5,0.3,1)	0.66	10.69
(4.5,0.6,1)	1.21	17.43
(4.5,0.8,1)	1.35	21.10
(5,0.8,1)	0.92	9.05
(5,0.8,5)	2.90	43.86
(5,0.8,10)	4.15	66.29

From table 1, it is observed that both the skewness and kurtosis increases with an increase in q .

3.2 Median

3.3 Median

The median (M) of any continuous probability distribution is the point that splits the area under the probability density curve into two equal parts. In the case of IBBGS1 distribution, the area under the curve differ between the intervals $[0, 1)$ and $[1, \infty)$. Therefore the median of this distribution will fall within either of these two ranges. The following algorithm are used for calculation of median:

1. Compute $F(1) = \int_0^1 f_1(y)dy$.
2. If $F(1) \geq 0.5$ then the median will lie in $[0, 1)$ and M is obtained by solving the following equation:

$$\int_0^M f_1(y)dy = 0.5$$

$$\implies \frac{\beta(a+q, b) M^q}{\beta(a, b)} = 0.5$$

3. If $F(1) < 0.5$ then the median will lie in $[1, \infty)$ and M is obtained by solving the following equation:

$$\int_0^1 f_1(y)dy + \int_1^M f_2(y)dy = 0.5$$

$$\implies \frac{\beta\left(\frac{1}{M}, a, b\right) - \beta\left(\frac{1}{M}, a+q, b\right) M^q}{\beta(a, b)} = 0.5 \tag{13}$$

The median values for different set of parameters are given in table 2:

Table 2 Median values for different set of parameters

Parameters	Median
(0.9,0.3,2)	0.75938
(1,1.5,2)	0.56557
(2,0.3,0.5)	4.23087
(0.9,0.3,0.5)	9.96179

3.4 Moment Generating Function

The moment generating function of *IBGSL* distribution is given by:

$$\begin{aligned}
 M_Y(t) &= E(e^{ty}) \\
 &= \int_0^1 e^{ty} f_1(y) dy + \int_1^\infty e^{ty} f_2(y) dy \\
 &= 1 + \frac{q}{\beta(a, b)} \sum_{k=1}^{\infty} \frac{t^k}{k!(k+q)} \beta(a-k, b)
 \end{aligned} \tag{14}$$

3.5 Additive Property of *IBGSL* distribution

Theorem 2. *IBGSL* distribution does not satisfy the additive property i.e., if $X \sim \text{IBGSL}(a_1, b_1, q_1)$ and $Y \sim \text{IBGSL}(a_2, b_2, q_2)$, then $(X + Y)$ does not follow the *IBGSL* distribution.

Proof: The m.g.f. of *IBGSL*(a, b, q) distribution is given by:

$$\begin{aligned}
 M_Y(t) &= E(e^{ty}) \\
 &= \int_0^1 e^{ty} f_1(y) dy + \int_1^\infty e^{ty} f_2(y) dy \\
 &= 1 + \frac{q}{\beta(a, b)} \sum_{k=1}^{\infty} \frac{t^k}{k!(k+q)} \beta(a-k, b)
 \end{aligned}$$

Let $Z = X+Y$ where $X \sim \text{IBGSL}(a_1, b_1, q_1)$ and $Y \sim \text{IBGSL}(a_2, b_2, q_2)$ and are independently distributed of each other. Then the m.g.f. of Z is

$$\begin{aligned}
 M_Z(t) &= M_{X+Y}(t) \\
 &= M_X(t)M_Y(t) \\
 &= \left(1 + \frac{q_1}{\beta(a_1, b_1)} \sum_{k=1}^{\infty} \frac{t^k}{k!(k+q_1)} \beta(a_1-k, b_1) \right) \times \\
 &\quad \left(1 + \frac{q_2}{\beta(a_2, b_2)} \sum_{k=1}^{\infty} \frac{t^k}{k!(k+q_2)} \beta(a_2-k, b_2) \right)
 \end{aligned} \tag{15}$$

which is not the m.g.f. of *IBGSL* distribution.

Thus, $X+Y$ does not follow *IBGSL* distribution or in other words, the *IBGSL* distribution does not satisfy the additive property.

3.6 Mean Deviation about mean

The mean deviation about mean of a population measure the amount of scatter in a population to some extent. For a random variable Y with pdf $g(y)$, cdf $G(Y)$, mean $\mu = E(Y)$, the mean deviation about mean are defined by:

$$\begin{aligned}
 \delta_1(y) &= \int_0^{\infty} |y - \mu|g(y)dy \\
 &= \int_0^{\mu} (\mu - y)g(y)dy + \int_{\mu}^{\infty} (y - \mu)g(y)dy \\
 &= \mu G(\mu) - \int_0^{\mu} yf(y)dy - \mu [1 - G(\mu)] + \int_{\mu}^{\infty} yg(y)dy \\
 &= 2\mu G(\mu) - 2\mu + 2 \int_{\mu}^{\infty} yg(y)dy \\
 &= 2\mu G(\mu) - 2 \int_0^{\mu} yg(y)dy
 \end{aligned} \tag{16}$$

Hence the mean deviation for *IBGSI* (a, b, q) are given by:

$$\begin{aligned}
 \delta_1(y) &= I_{(0,1)}(y) \left[\frac{2q\beta(a+q, b)\mu^q}{\beta(a, b)(1+q)} \left\{ \frac{\beta(a+q, b)}{\beta(a, b)} - \mu \right\} \right] \\
 &\quad + (1 - I_{(0,1)}(y)) \left[\frac{2q}{\beta(a, b)} \left\{ \frac{\beta(a-1, b) - \beta(a+q, b)}{(1-q)} \right\} \right] \\
 &\quad \left\{ \frac{\beta\left(\frac{1}{\mu}, a+q, b\right) - \beta(a+q, b) + \beta(a, b) - \beta\left(\frac{1}{\mu}, a, b\right)}{\beta(a, b)} \right\} \\
 &\quad - \frac{2q}{\beta(a, b)(q+1)} \left\{ \beta\left(\frac{1}{\mu}, a+q, b\right) \mu^{q+1} - \beta(a+q, b) + \beta(a-1, b) \right\}
 \end{aligned} \tag{17}$$

3.7 Mills Ratio

The Mills Ratio is the ratio of complementary cumulative distribution function to the probability density function. Mills ratio can be used in regression analysis to take account of a possible selection bias. Mills Ratio for *IBGSI*(a, b, q) distribution is :

$$\begin{aligned}
 m(y) &= I_{(0,1)}(y) \frac{1 - G_1(y)}{g_1(y)} + (1 - I_{(0,1)}(y)) \frac{1 - G_2(y)}{g_2(y)} \\
 &= I_{(0,1)}(y) \left\{ \frac{1 - \frac{\beta(a+q,b)y^q}{\beta(a,b)}}{\frac{qy^{q-1}}{\beta(a,b)}\beta(a+q,b)} \right\} \\
 &\quad + (1 - I_{(0,1)}(y)) \left\{ \frac{\frac{\beta(a+q,b) - \beta(\frac{1}{y}, a+q,b)y^q + \beta(\frac{1}{y}, a,b)}{\beta(a,b)}}{\frac{qy^{q-1}}{\beta(a,b)}\beta\left(\frac{1}{y}; a+q, b\right)} \right\} \\
 &= I_{(0,1)}(y) \left\{ \frac{\beta(a,b) - \beta(a+q,b)y^q}{qy^{q-1}\beta(a+q,b)} \right\} \\
 &\quad + (1 - I_{(0,1)}(y)) \left\{ \frac{\beta(a+q,b) - \beta(\frac{1}{y}, a+q,b) + \beta(\frac{1}{y}, a,b)}{qy^{(q-1)}\beta(\frac{1}{y}, a+q,b)} \right\} \tag{18}
 \end{aligned}$$

3.8 Order Statistics

Consider a random sample y_1, y_2, \dots, y_n of size n drawn from $IBGSI(a, b, q)$ distribution. Further, let $y_{(1)} < y_{(2)} < \dots < y_{(n)}$ denote the order statistics corresponding to this sample. Then the probability density function of the k^{th} order statistic is given by

$$f_{(k)}(y) = \frac{n!}{(k-1)!(n-k)!} [F(y)]^{k-1} [1 - F(y)]^{n-k} f(y)$$

Hence the density of k^{th} order statistic for $IBGSI(a, b, q)$ distribution is

$$\begin{aligned}
 g_{(k)}(y) &= I_{(0,1)}(y) \left[\frac{n!}{(k-1)!(n-k)!} \left\{ \frac{\beta(a+q,b)y^q}{\beta(a,b)} \right\}^{k-1} \left\{ 1 - \frac{\beta(a+q,b)y^q}{\beta(a,b)} \right\}^{n-k} \frac{qy^{q-1}}{\beta(a,b)}\beta(a+q,b) \right] \\
 &\quad + (1 - I_{(0,1)}(y)) \left[\left\{ \frac{\beta(\frac{1}{y}, a+q,b)y^q - \beta(a+q,b) - \beta(\frac{1}{y}, a,b)}{\beta(a,b)} \right\}^{k-1} \right. \\
 &\quad \left. \left\{ \beta(a+q,b) + \beta(\frac{1}{y}, a,b) - \frac{1}{y}, a+q,b)y^q\beta(a,b) \right\}^{n-k} \frac{qy^{q-1}}{\beta(a,b)}\beta\left(\frac{1}{y}; a+q, b\right) \right] \tag{19}
 \end{aligned}$$

In particular, the p.d.f of the smallest order statistic $y_{(1)}$ is

$$\begin{aligned}
 g_{(1)}(y) &= I_{(0,1)}(y) \left[n \left\{ 1 - \frac{\beta(a+q,b)y^q}{\beta(a,b)} \right\}^{n-1} \frac{qy^{q-1}}{\beta(a,b)}\beta(a+q,b) \right] \\
 &\quad + (1 - I_{(0,1)}(y)) \left[\left\{ \frac{\beta(a+q,b) + \beta(\frac{1}{y}, a,b) - \beta(\frac{1}{y}, a+q,b)y^q}{\beta(a,b)} \right\}^{n-1} \frac{qy^{q-1}}{\beta(a,b)}\beta\left(\frac{1}{y}; a+q, b\right) \right] \tag{20}
 \end{aligned}$$

The pdf of the largest order statistic $y_{(n)}$ is

$$\begin{aligned}
 g_{(n)}(y) &= I_{(0,1)}(y) \left[n \left\{ \frac{\beta(a+q,b)y^q}{\beta(a,b)} \right\}^{n-1} \frac{qy^{q-1}}{\beta(a,b)}\beta(a+q,b) \right] \\
 &\quad + (1 - I_{(0,1)}(y)) \left[\left\{ \frac{\beta(\frac{1}{y}, a+q,b)y^q - \beta(a+q,b) - \beta(\frac{1}{y}, a,b)}{\beta(a,b)} \right\}^{n-1} \frac{qy^{q-1}}{\beta(a,b)}\beta\left(\frac{1}{y}; a+q, b\right) \right] \tag{21}
 \end{aligned}$$

3.9 Lorenz and Bonferroni Curve

The Bonferroni and Lorenz Curve are the most used tools in income inequality measurement. These two curves are widely used in the field of reliability, demography, medicine and insurance. The Bonferroni and Lorenz curves are defined as:

$$L(G(y)) = I_{(0,1)}(y) \left[\frac{1}{\mu} \int_0^y t g_1(t) dt \right] + (1 - I_{(0,1)}(y)) \left[\frac{1}{\mu} \int_1^y t g_2(t) dt \right] \quad (22)$$

$$\begin{aligned} B(G(y)) &= I_{(0,1)}(y) \left[\frac{1}{\mu G_1(y)} \int_0^y t g_1(t) dt \right] + (1 - I_{(0,1)}(y)) \left[\frac{1}{\mu G_2(y)} \int_1^y t g_2(t) dt \right] \\ &= I_{(0,1)}(y) \left[\frac{L(G_1(y))}{G_1(y)} \right] + (1 - I_{(0,1)}(y)) \left[\frac{L(G_2(y))}{G_2(y)} \right] \end{aligned} \quad (23)$$

After simplifications,

$$\begin{aligned} L(G(y)) &= I_{(0,1)}(y) \left[\frac{\beta(a+q, b)(a-1)y^{q+1}}{\beta(a, b)(a+b-1)} \right] \\ &- (1 - I_{(0,1)}(y)) \left[\frac{(a-1) \left\{ y^{q+1} \beta\left(\frac{1}{y}, a+q, b\right) - \beta(a+q, b) + \beta(a-1, b) \right\}}{\beta(a, b)(a+b-1)} \right] \end{aligned} \quad (24)$$

$$\begin{aligned} B(G(y)) &= I_{(0,1)}(y) \left[\frac{y(a-1)}{a+b-1} \right] \\ &+ (1 - I_{(0,1)}(y)) \left[\frac{(a-1) \left\{ y^{q+1} \beta\left(\frac{1}{y}, a+q, b\right) - \beta(a+q, b) + \beta(a-1, b) \right\}}{(a+b-1) \left\{ \beta\left(\frac{1}{y}, a+q, b\right) - \beta(a+q, b) - \beta\left(\frac{1}{y}, a, b\right) + \beta(a, b) \right\}} \right] \end{aligned} \quad (25)$$

3.10 Hazard Rate Function

The hazard rate function is a very important tool in understanding about the failure mechanism of a lifetime distribution. Hazard rate function can be used to postulate life distributions in the presence of several competing risk factors. It measures the instantaneous rate at which a system or component is likely to fail, given that it has survived up to a certain time. The hazard rate function of $IBGSI(a, b, q)$ distribution is obtained by using the following formula:

$$\begin{aligned} h(y) &= I_{(0,1)}(y) \frac{g_1(y)}{1 - G_1(y)} + (1 - I_{(0,1)}(y)) \frac{g_2(y)}{1 - G_2(y)} \\ &= I_{(0,1)}(y) \left[\frac{qy^{q-1} \beta(a+q, b)}{\beta(a, b) - y^q \beta(a+q, b)} \right] \\ &+ (1 - I_{(0,1)}(y)) \left[\frac{qy^{q-1} \beta\left(\frac{1}{y}, a+q, b\right)}{\beta(a+q, b) + \beta\left(\frac{1}{y}, a, b\right) - \beta\left(\frac{1}{y}, a+q, b\right)} \right] \end{aligned} \quad (26)$$

The hrf plot of $IBGSI$ distribution for different values of parameters, is plotted in Figure 3.

4. Estimation

In this section, we discuss the maximum likelihood method of estimation for the unknown model parameters of $IBGSI(a, b, q)$. Let y_1, y_2, \dots, y_n be a random sample of size n from $IBGSI(a, b, q)$ distri-

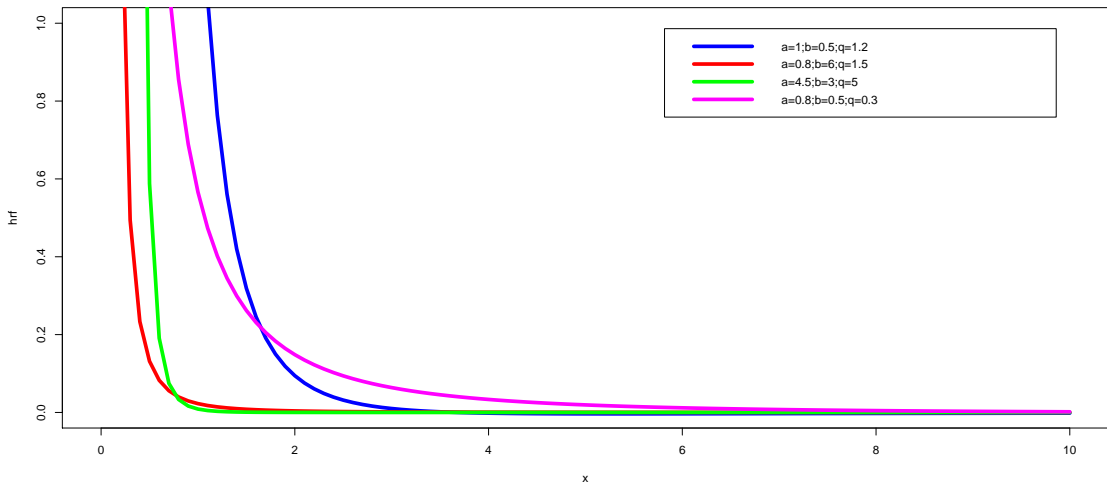


Figure 3 Hazard rate function plots of the IBSI(a,b,q) distribution

bution. Then the log - likelihood function is obtained as:

$$L(a, b, q, \mathbf{y}) = \prod_{i=1}^n f(y_i, a, b, q) \\ = L_1(a, b, q, \mathbf{y}) * L_2(a, b, q, \mathbf{y})$$

$$L_1(a, b, q, \mathbf{y}) = \prod_{i=1}^n g_1^{I_{(0,1)}(y_i)} \\ = g_1^{\sum_{i=1}^n I_{(0,1)}(y_i)}$$

$$\log L_1(a, b, q, \mathbf{y}) = \sum_{i=1}^n I_{(0,1)}(y_i) \left[\log q + \log \beta(a + q, b) - \log \beta(a, b) + (q + 1) \log y_i \right] \quad (27)$$

Again,

$$L_2(a, b, q, \mathbf{y}) = \prod_{i=1}^n g_1^{1-I_{(0,1)}(y_i)} \\ = g_2^{\sum_{i=1}^n \{1-I_{(0,1)}(y_i)\}} \\ = g_2^{\{n-\sum_{i=1}^n I_{(0,1)}(y_i)\}}$$

$$\log L_2(a, b, q, \mathbf{y}) = \sum_{i=1}^n (n - I_{(0,1)}(y_i)) \left[\log q + \log \beta \left(\frac{1}{y}, a + q, b \right) - \log \beta(a, b) + (q + 1) \log y_i \right] \quad (28)$$

$$\begin{aligned} \log L(a, b, q, \mathbf{y}) &= \sum_{i=1}^n I_{(0,1)}(y_i) \left[\log q + \log \beta(a + q, b) - \log \beta(a, b) + (q + 1) \log y_i \right] \\ &+ \sum_{i=1}^n (n - I_{(0,1)}(y_i)) \left[\log q + \log \beta\left(\frac{1}{y_i}, a + q, b\right) - \log \beta(a, b) + (q + 1) \log y_i \right] \end{aligned} \quad (29)$$

The maximum likelihood estimates (MLE) of the parameters are computed by solving the maximum likelihood equations, which are given by

$$\begin{aligned} \frac{\partial \log L}{\partial a} &\implies \sum_{i=1}^n I_{(0,1)}(y_i) \left[\frac{1}{\beta(a + q, b)} \frac{d}{da} \beta(a + q, b) - \{\psi_0(a) - \psi_0(a + b)\} \right] \\ &+ (n - I_{(0,1)}(y_i)) \left[\frac{1}{\beta\left(\frac{1}{y_i}, a, b\right)} \frac{d}{da} \beta\left(\frac{1}{y_i}, a, b\right) - \{\psi_0(a) - \psi_0(a + b)\} \right] = 0 \end{aligned} \quad (30)$$

$$\begin{aligned} \frac{\partial \log L}{\partial b} &\implies \sum_{i=1}^n I_{(0,1)}(y_i) \left[\frac{1}{\beta(a + q, b)} \frac{d}{db} \beta(a + q, b) - \{\psi_0(b) - \psi_0(a + b)\} \right] \\ &+ (n - I_{(0,1)}(y_i)) \left[\frac{1}{\beta\left(\frac{1}{y_i}, a, b\right)} \frac{d}{db} \beta\left(\frac{1}{y_i}, a, b\right) - \{\psi_0(b) - \psi_0(a + b)\} \right] = 0 \end{aligned} \quad (31)$$

$$\begin{aligned} \frac{\partial \log L}{\partial q} &\implies \sum_{i=1}^n I_{(0,1)}(y_i) \left[\frac{1}{q} + \frac{1}{\beta(a + q, b)} \frac{d}{dq} \beta(a + q, b) + \log y_i \right] \\ &+ (n - I_{(0,1)}(y_i)) \left[\frac{1}{q} + \frac{1}{\beta\left(\frac{1}{y_i}, a + q, b\right)} \frac{d}{dq} \beta\left(\frac{1}{y_i}, a + q, b\right) + \log y_i \right] = 0 \end{aligned} \quad (32)$$

The above maximum likelihood equations are not in closed form and so, they are difficult to be solved analytically. Hence, we shall use a suitable numerical technique to solve the above equations for a , b and q . Here all the calculations have been carried out using the R software version 3.6.3. The `maxLik` package is used to obtain the maximum likelihood estimates of the parameters, the `rootSolve` package is used to generate random variables from *IBGSI* distribution and `zipfR` package is used to evaluate the incomplete beta function.

5. Simulation

In this section, generation of random numbers from *IBGSI*(a, b, q) distribution is discussed. For different values of a, b and q , we generate random samples of size 30, 100, 300, 800 and 1000 from *IBGSI*(a, b, q). Finally, the average values of bias and mean squared error (MSE) of these estimates are calculated by using the Monte Carlo approximation technique, taking $N = 1,000$ replicates. The algorithm used in this simulation study is shown below:

1. Simulate $X \sim BGSi(a, b, q)$
2. Compute $Y = \frac{1}{X}$

Y thus generated is a random number from the $IBGSI(a,b,q)$ distribution. To calculate the average bias and MSE of the likelihood estimates, we use the formulae as shown below :

Let the true value of the parameter a be a^* and estimate be \hat{a} . Then the bias and mean square error (MSE) of \hat{a} in estimating a is given by:

$$Bias(\hat{a}) = \frac{1}{N} \sum_{i=1}^N (\hat{a}_i - a^*) \tag{33}$$

$$MSE(\hat{a}) = \frac{1}{N} \sum_{i=1}^N (\hat{a}_i - a^*)^2 \tag{34}$$

where N is the number of replications and \hat{a}_i is the MLE of \hat{a} obtained in the i^{th} replicate. Similarly, the bias and MSE of b and q are calculated. It is well known that an estimate is consistent if the bias and MSE decreases (approaches to zero) with an increase in the sample size. Table 3 shows the results of the simulation studies. It is seen that the parameters are well estimated and the bias and MSE of all the estimators approaches towards zero with an increase in the sample size. Hence, the estimates of the parameters can be believed to be consistent.

Table 3 Average bias and RMSE for the estimates of $IBGSI$ distribution.

parameters	n	\hat{a}		\hat{b}		\hat{q}	
		Bias(\hat{a})	RMSE(\hat{a})	Bias(\hat{b})	RMSE(\hat{b})	Bias(\hat{q})	RMSE(\hat{q})
(a=0.5, b=0.8, q=2.5)	30	0.00650	0.00118	0.00507	0.00092	-0.00180	0.00033
	100	0.00308	0.00030	0.00135	0.00013	-0.00187	0.00019
	300	0.00307	0.00018	0.00108	0.00005	-0.00191	0.00011
	500	0.00305	0.00014	0.00095	0.00004	-0.00193	0.00008
	800	0.00242	0.00007	0.00099	0.00003	-0.00196	0.00006
	1000	0.00180	0.00004	0.00033	0.00001	-0.00198	0.00004
(a=1.5, b=3, q=5)	30	0.03401	0.00621	0.00057	0.00034	-0.00419	0.00076
	100	0.00448	0.00044	-0.00276	0.00027	-0.00418	0.00041
	300	0.00236	0.00013	-0.00282	0.00016	-0.00416	0.00024
	500	0.00192	0.00007	-0.00286	0.00012	-0.00415	0.00018
	800	0.00174	0.00006	-0.00290	0.00010	-0.00414	0.00014
	1000	0.00110	0.00003	-0.00293	0.00001	-0.00413	0.00013

6. Application

To show the flexibility of the proposed distribution over some existing distributions in modeling heavy - tailed data we apply these distributions to a real life data set. Data set representing vinyl chloride data from clean up gradient monitoring wells in mg/l used by Bhaumik et al. (2009) has been considered . The data set comprises of the observations:

5.1, 1.2, 1.3, 0.6, 0.5, 2.4, 0.5, 1.1, 8.0, 0.8, 0.4, 0.6, 0.9, 0.4, 2.0, 0.5, 5.3, 3.2, 2.7, 2.9, 2.5, 2.3, 1.0, 0.2, 0.1, 0.1, 1.8, 0.9, 2.0, 4.0, 6.8, 1.2, 0.4, 0.2.

The histogram of the data set exhibits a right skewed behavior, which may be aptly modelled by the proposed distribution. Using this data set, we compare the $IBGSI$ distribution with Beta Generated Slash distribution (BGS), Inverted Weibull (IW) distribution and Inverted Gamma (IG) distribution. In order to compare the distributions we calculate the log - likelihood, Akaike Information Criterion (AIC), the Corrected Akaike Information Criterion (AICC) and Kolmogorov-Smirnov (K-S) statistic. The model with minimum AIC, AICC and KS statistic is chosen as the best model to fit the data.

Table 4 Estimated parameters, AIC, AICC and KS values of $IBGSI(a,b,q)$ and other competing distributions fitted to the data set

Distribution	MLE	log-likelihood	AIC	AICC	K-S statistic
IBGSI(a,b,q)	$\hat{a}=2.09857$				
	$\hat{b}=3.797134$	-56.765	119.530	120.189	0.106
	$\hat{q}=0.9000848$				
BGSI(a,b,q)	$\hat{a}=1.22731$				
	$\hat{b}=1.150665$	-57.795	121.253	121.124	0.7745
	$\hat{q}=1.012099$				
IW	$\hat{a}=0.8804067$				
	$\hat{b}=0.6539517$	-58.62659	121.591	122.316	0.6256
IG	$\hat{a}=0.9002228$				
	$\hat{b}=0.5153479$	-2008.242	4020.484	4021.129	0.771

The values of log - likelihood, AIC, AICC and KS statistic are shown in table 4

From the table 4, it has been seen that the $IBGSI$ distribution has maximum likelihood and minimum AIC, AICC and K-S statistic. Hence the $IBGSI$ distribution performs well than the other competing distribution. Furthermore, Figure 4 and Figure 5 shows the histogram of the data sets and the empirical cdf versus fitted cdfs for the data representing vinyl chloride from clean up gradient monitoring wells in mg/l. This figure confirms that the goodness-of-fit of BGSL distribution with respect to all the fitted distributions.

7. Concluding remarks

This paper introduces the inverted beta generated slash distribution having three parameters, which is an inverted form of beta generated slash distribution. The various distributional aspects such as moments, skewness, kurtosis, median, moment generating function, mean deviation, mills ratio, order statistics, Lorenz and Bonferroni curves are studied. The method of maximum likelihood is used to estimate the parameters and a simulation study is performed to study the finite sample behaviour of the ML estimates. The MLE's are found to be consistent and precise in estimating the true value of the parameters. To show the application of the proposed distribution, it is applied to a dataset representing vinyl chloride from clean up gradient monitoring wells in mg/l. The fit of the proposed distribution is compared with beta generated slash distribution (BGSI), Inverted Weibull (IW) distribution and Inverted Gamma (IG) distribution using log - likelihood measure, AIC, AICC and K-S statistic. It is observed that the $IBGSI$ distribution is a better fit to the data as compared to the others. Thus it can be concluded that the proposed distribution is more flexible and has advantage in modeling right skewed heavy - tailed datasets occurring in $[0, \infty)$ or any subset of it.

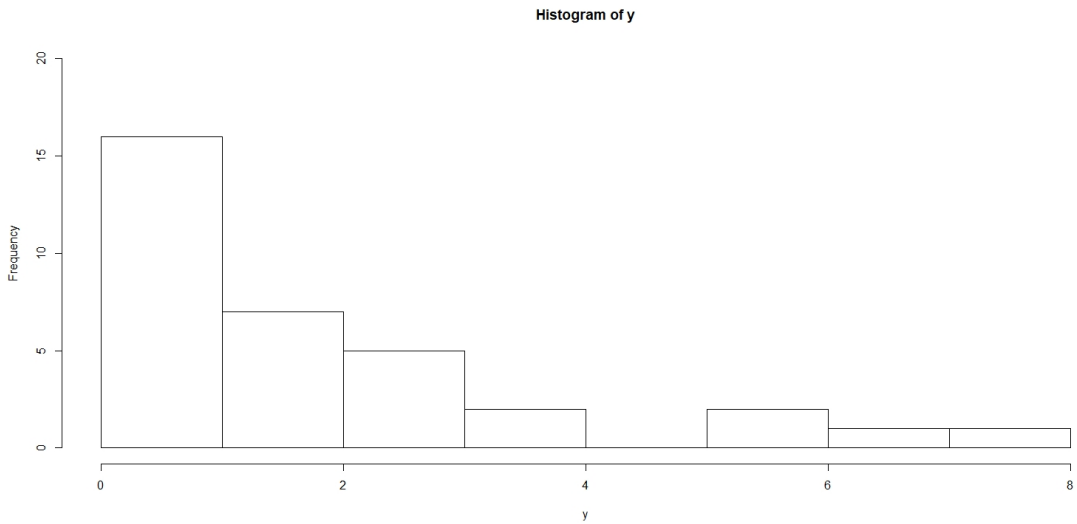


Figure 4 Histogram of Vinyl Chloride data

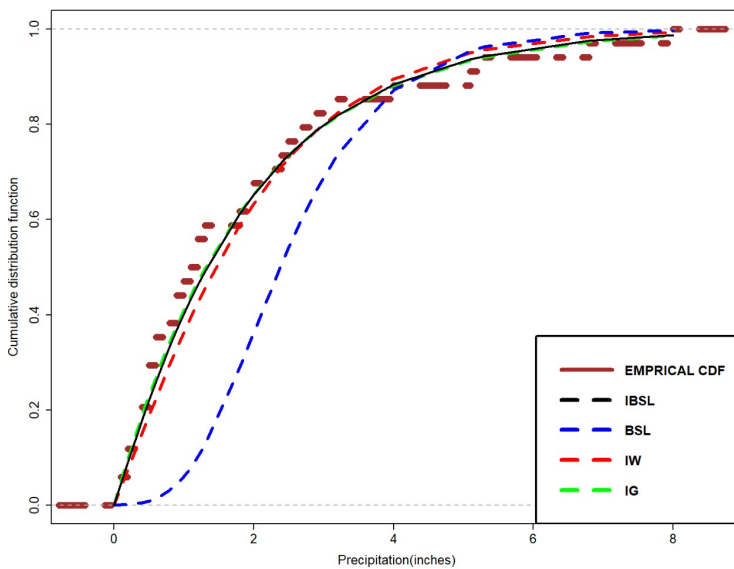


Figure 5 CDF plot of the observed data and fitted distributions

References

Abd Al Fattah AM, El- Helbawy AA and Al Dayian GR. Inverted Kumaraswamy Distribution: Properties and Estimation. Pakistan Journal of Statistics. 2017; 33:37-61.

Abd EL-Kader RI. A general class of some inverted distributions.Ph.D. Thesis, AL-Azhar University; 2013.

Abd EL- Kader RI, Al Dayian GR and Al Gendy SA. Inverted Pareto Type I distribution: properties

- and estimation. *Journal of Faculty of Commerce. Al - Azhar University, Girl's branch.* 2003; 21:19-40.
- Bhattacharjee S and Borah N. Beta generated slash distribution: derivation, properties and application to lifetime data. *Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics.* 2024;73:641-63.
- Bhaumik DK, Kapur K and Gibbons RD. Testing parameters of a gamma distribution for small samples. *Technometrics.* 2009; 51:326-334.
- Conway DA. Farlie-Gumbel-Morgenstern Distributions. *Encyclopedia of Statistical Sciences.* 1983; 28-31
- Dietrich VR. On Moments of the Inverted Wishart Distribution. *A journal of Theoretical and Applied Statistics.* 1997; 30.
- Genc, Ali Alper, Mustafa ÇK and Coskun K. The beta Moyal-slash distribution. *Journal of Selcuk University Natural and Applied Science.* 2014;3(4) :88-104.
- Gómez HW, Olivares-Pacheco JF, Bolfarine H. An extension of the generalized Birnbaum–Saunders distribution. *Statistics & Probability Letters.* 2009; 79(3):331-338.
- Gui W, Chen PH, Wu H. An epsilon half normal slash distribution and its applications to nonnegative measurements, 2013.
- Hassan NJ, Nasar AH and Hadad JM. Distribution of the ratio and product of two independent Weibull and Lindley random variables. *Journal of Probability and Statistics.* 2019; 20.
- Kolev N and Paiva D. Copula-based regression models: A survey. *Journal of statistical planning and inference.* 2009; 139:3847-3856.
- Patton A. Copula methods for forecasting multivariate time series. *Handbook of Economic Forecasting.* 2013; 2:899–960.
- Punathumparambath B. A new family of skewed slash distributions generated by the normal kernel. *Statistica.* 2011; 71(3):345-53.
- Reyes J, Gómez HW and Bolfarine H. Modified slash distribution. *Statistics.* 2013; 47:929-941.
- Rogers WH and Tukey JW. Understanding some long-tailed symmetrical distributions, *Statistics Netherlands.* 1972; 26(3):211-226.
- Shaked M and Shanthikumar JG. *Stochastic Orders and their applications.* Academic Press, New York, 1994
- Shannon CE. A Mathematical Theory of Communication. *Bell System Technical Journal.* 1948; 27:379-423.