The Log-Hamza distribution with statistical properties and application

An alternative for distributions having domain (0,1).

Aijaz Ahmad*

•

Department of Mathematics, Bhagwant University, Ajmer, India
aijazahmad4488@gmail.com

Afaq Ahmad

•

Department of Mathematical Science, IUST
baderaafaq@gmail.com

I. H. Dar

•

Department of Statistics, University of Kashmir, Srinagar, India
ishfaqqh@gmail.com

Rajnee Tripathi

•

Department of Mathematics, Bhagwant University, Ajmer, India
rajneetripathi@hotmail.com

Abstract

This work suggests a novel two-parameter distribution known as the log-Hamza distribution, in short (LHD). The significant property of the investigated distribution is that it belongs to the family of distributions that have support (0,1). Several statistical features of the investigated distribution were studied, including moments, moment generating functions, order statistics, and reliability measures. For different parameter values, a graphical representation of the probability density function (pdf) and the cumulative distribution function (CDF) is provided. The distribution's parameters are determined using the well-known maximum likelihood estimation approach. Finally, an application is used to evaluate the effectiveness of the distribution.

Keywords: Log transformation, Hamza distribution, moments, maximum likelihood estimation.

1. Introduction

Statistical distribution is important in modelling many sorts of data from various disciplines of research. Statisticsians have focused their efforts on developing new distributions or generalising current distributions by introducing additional parameters. The major reason for these extensions is to improve the efficiency of these distributions while analysing increasingly complicated data.

The Beta distribution, Kumaraswamy distribution, and Topp-Leone distribution are the most commonly used bound support distributions. Among these distributions, the Beta distribution is the most common and has applications in many fields of study, including bio-science, engineering, economics, and finance. The fundamental disadvantage of the beta distribution is that its cumulative distribution function (c.d.f) comprises a beta function that cannot be written in closed form.The aim of this paper is to introduce a new distribution which is considered an alternative to the family of distributions having support (0,1). To achieve this goal, the Hamza distribution is used to generate a new distribution which is defined on an open interval (0,1) . In this regard, a noteworthy effort has been attempted to limit many continuous distributions in unit intervals, including: Topp-Leone [12], Nadarajah and kotz [10], Cordeiro and castro [3], Gomez-Deniz et al. [4], Mazucheli et al. [7], Ghitany et al.[5], Haq et al.[6], Menezes et al.[8], Rodrigues et al.[11], Aijaz et al.[1].

2. The Log-Hamza Distribution

Suppose a random variable X follow Hamza distribution with probability density function (p.d.f)

$$f(x;\alpha,\beta) = \frac{\beta^6}{\alpha\beta^5+120}\left(\alpha+\frac{\beta}{6}x^6\right)e^{-\beta x} \quad ; \quad x>0,\alpha,\beta>0 \tag{1}$$

The corresponding cumulative distribution function (c.d.f) is given as

$$F(x;\alpha,\beta) = 1-\left[1+\frac{\beta x((\beta x)^5+6(\beta x)^4+30(\beta x)^3+120(\beta x)^2+360(\beta x)+720)}{6(\alpha\beta^5+120)}\right]e^{-\beta x} \quad ; \quad x>0,\alpha,\beta>0 \tag{2}$$

Suppose a random variable $Y=e^{-X} \implies X=-ln(Y)$, then the probability density function (pdf) of Y is given as

$$f(y;\alpha,\beta) = \frac{\beta^6}{\alpha\beta^5+120}\left(\alpha+\frac{\beta}{6}(ln(y))^6\right)y^{\beta-1} \quad ; \quad 0<y<1,\alpha,\beta>0 \tag{3}$$

Figure (1.1) and (1.2) represents some possible shapes of pdf of LHD for different values of parameters

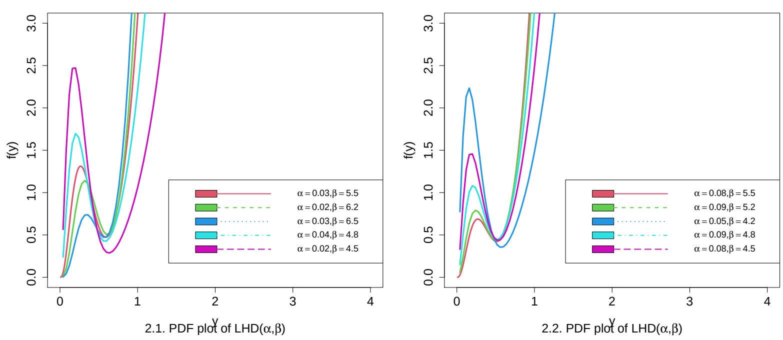

Figure 1

The corresponding cumulative distribution function (cdf) is given by

$$F(y;\alpha,\beta)=\left[1+\frac{((\beta ln(y))^6-6(\beta ln(y))^5+30(\beta ln(y))^4-120(\beta ln(y))^3}{6(\alpha\beta^5+120)}\right.$$
$$\left.+\frac{360(\beta ln(y))^2-720(\beta ln(y)))}{6(\alpha\beta^5+120)}\right]y^\beta \quad ; \quad 0\le y\le 1,\alpha,\beta>0 \tag{4}$$

3. Reliability Measures of Log-Hamza Distribution

This section is focused on researching and developing distinct ageing indicators for the formulated distribution.

3.1. Survival function

Suppose Y be a continuous random variable with cdf $F(y)$.Then its Survival function which is also called reliability function is defined as

$$S(y)=p_r(Y>y)=\int_y^\infty f(y)dy=1-F(y)$$

Therefore, the survival function for log-Hamza distribution is given as

$$S(y;\alpha,\beta)=1-F(y,\alpha,\beta)$$
$$S(y)=1-\left[1+\frac{((\beta ln(y))^6-6(\beta ln(y))^5+30(\beta ln(y))^4-120(\beta ln(y))^3}{6(\alpha\beta^5+120)}\right. \tag{5}$$
$$\left.+\frac{360(\beta ln(y))^2-720(\beta ln(y)))}{6(\alpha\beta^5+120)}\right]y^\beta \quad ; \quad 0\le y\le 1,\alpha,\beta>0$$

3.2. Hazard rate function

The hazard rate function of a random variable y is denoted as

$$h(y;\alpha,\beta)=\frac{f(y,\alpha,\beta)}{S(y,\alpha,\beta)} \tag{6}$$

using equation (3) and (4) in equation (6), then the hazard rate function of log-Hamza distribution is given as

$$h(y)=\frac{\beta^6\left(\alpha+\frac{\beta}{6}(ln(y))^6\right)y^{\beta-1}}{6(\alpha\beta^5+120)-[6(\alpha\beta^5+120)+(A)]\,y^\beta} \quad ; \quad 0<y<1,\alpha,\beta>0$$

where

$$A=(\beta ln(y))^6-6(\beta ln(y))^5+30(\beta ln(y))^4-120(\beta ln(y))^3+360(\beta ln(y))^2-720(\beta ln(y))$$

Figure (3.1) and (3.2) represents some possible shapes of hrf of LHD for different values of parameters

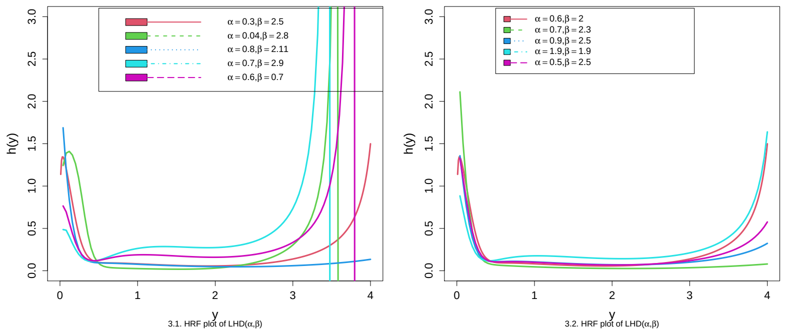

Figure 2

3.3. Cumulative hazard rate function

The cumulative hazard rate function of a random variable y is given as

$$H(y,\alpha,\beta) = -\,ln[\bar{F}(y,\alpha,\beta)] \tag{7}$$

using equation (12) in equation (17), then we obtain cumulative hazard rate function of IWB-III distribution

$$H(y,\alpha,\beta) = -ln\left[1-\left(1+\frac{((\beta ln(y))^6-6(\beta ln(y))^5+30(\beta ln(y))^4-120(\beta ln(y))^3}{6(\alpha\beta^5+120)} + \frac{360(\beta ln(y))^2-720(\beta ln(y)))}{6(\alpha\beta^5+120)}\right)y^{\beta}\right] \tag{8}$$

3.4. Mean residual function

The mean residual lifetime is the predicted residual life or the average completion period of the constituent after it has exceeded a certain duration y. It is extremely significant in reliability investigations.
Mean residual function of random y variable can be obtained as

$$m(y;\alpha,\beta) = \frac{1}{S(y,\alpha,\beta)}\int_y^1 tf(t,\alpha,\beta)dt - y$$
$$= \frac{1}{S(y,\alpha,\beta)}\frac{\beta^6}{(\alpha\beta^5+120)}\int_y^1\left(\alpha+\frac{\beta}{6}(ln(t))^6\right)t^{\beta-1}dt - y$$

Making substitution $ln(t) = -z$, sothat $0 \le z \le -ln(y)$, we have

$$m(y;\alpha,\beta) = \frac{1}{S(y,\alpha,\beta)}\frac{\beta^6}{(\alpha\beta^5+120)}\int_0^{-ln(y)}\left(\alpha+\frac{\beta}{6}z^6\right)e^{-\beta z}dz - y$$

After solving the integral, we get

$$m(y;\alpha,\beta) = \frac{1}{S(y,\alpha,\beta)}\frac{\beta^5}{(\alpha\beta^5+120)}\left\{\alpha(1-y^{\beta})+\frac{1}{6\beta^5}\gamma\left(5,ln(y^{-\beta})\right)\right\} - y$$

Where $\gamma(a,x) = \int_0^x u^{a-1}e^{-u}du$ denotes lower incomplete gamma function

4. Statistical Properties Of log-Hamza Distribution

This section is devoted to derive and examine disttinct properties of log-Hamza

4.1. Moments

Let y denotes a random variable, then the r^{th} moment of log-Hamza is denoted as μ_r' and is given by

$$\mu_r' = E(y^r) = \int_0^1 y^r f(y,\alpha,\beta)dy$$
$$= \frac{\beta^6}{\alpha\beta^5 + 120} \int_0^1 y^{r+\beta-1} \left(\alpha + \frac{\beta}{6}(ln(y))^6\right) dy$$

Making substitution $ln(y) = -z$, so that $0 < z < \infty$, we have

$$\mu_r' = \frac{\beta^6}{\alpha\beta^5 + 120} \int_0^\infty \left(\alpha + \frac{\beta}{6}z^6\right) e^{-(\beta+r)z} dz$$

After solving the integral, we have

$$\mu_r' = \frac{\beta^6[(\beta+r)^6 + 120\beta]}{(\alpha\beta^5 + 120)(\beta+r)^7}$$

The first four raw moments of log-Hamza distribution are given as.

$$\mu_1' = \frac{\beta^6[(\beta+1)^6 + 120\beta]}{(\alpha\beta^5 + 120)(\beta+1)^7} \qquad \mu_2' = \frac{\beta^6[(\beta+2)^6 + 120\beta]}{(\alpha\beta^5 + 120)(\beta+2)^7}$$
$$\mu_3' = \frac{\beta^6[(\beta+3)^6 + 120\beta]}{(\alpha\beta^5 + 120)(\beta+3)^7} \qquad \mu_4' = \frac{\beta^6[(\beta+4)^6 + 120\beta]}{(\alpha\beta^5 + 120)(\beta+4)^7}$$

4.2. Moment generating function

suppose Y denotes a random variable follows log-Hamza distribution. Then the moment generating function of the distribution denoted by $M_Y(t)$is given

$$M_Y(t) = E(e^{ty}) = \int_0^1 e^{ty} f(y;\alpha,\beta)dy$$
$$= \int_0^1 \left(1 + ty + \frac{(ty)^2}{2!} + \frac{(ty)^3}{3!} + ...\right) f(y;\alpha,\beta)dy$$
$$= \sum_{r=0}^{\infty} \frac{t^r}{r!} \int_0^\infty y^r f(y;\alpha,\beta)dy$$
$$= \sum_{r=0}^{\infty} \frac{t^r}{r!} E(y^r)$$
$$= \sum_{r=0}^{\infty} \frac{t^r}{r!} \frac{\beta^6[(\beta+r)^6 + 120\beta]}{(\alpha\beta^5 + 120)(\beta+r)^7}$$

The characteristics function of the log-Hamza distribution denoted as $\phi_Y(t)$ can be yeild by replacing $t = it$ where $i = \sqrt{-1}$

$$\phi_Y(t) = \sum_{r=0}^{\infty} \frac{(it)^r}{r!} \frac{\beta^6[(\beta+r)^6 + 120\beta]}{(\alpha\beta^5 + 120)(\beta+r)^7}$$

4.3. Incomplete moments

The general expression for incomplete moments is given as

$$T(t) = \int_0^t y^r f(y;\alpha,\beta)dy$$
$$= \frac{\beta^6}{\alpha\beta^5 + 120} \int_0^t y^{r+\beta-1} \left(\alpha + \frac{\beta}{6}(ln(y))^6\right) dy$$

Making substitution $ln(y) = -z$, so that $-lnt(t) \leq z \leq \infty$, we have

$$=\frac{\beta^6}{\alpha\beta^5+120}\int_{-ln(t)}^{\infty}\left(\alpha+\frac{\beta}{6}z^6\right)e^{-(r+\beta)z}dz$$

After solving the integral, we get

$$T(t) =\frac{\beta^6}{\alpha\beta^5+120}\left(\frac{t^{\beta+r}}{\beta+r}+\frac{\beta}{6(\beta+r)}\Gamma\left(5, ln(t^{-(\beta+r)})\right)\right)$$

where $\Gamma(a,x) = \int_x^{\infty} u^{a-1}e^{-u}du$ denotes the upper incomplete gamma function.

5. Order Statistics of Log-Hamza Distribution

Let us suppose $Y_1, Y_2, ..., Y_n$ be random samples of size n from log-Hamza distribution with pdf $f(y)$ and cdf F(y). Then the probability density function of the k^{th} order statistics is given as

$$f_Y(k) =\frac{n!}{(k-1)!(n-1)!}f(y)\left[F(y)\right]^{k-1}\left[1-F(y)\right]^{n-1} \tag{9}$$

Using equation (3) and (4) in equation (10), we have

$$\begin{aligned} f_Y(k) =&\frac{n!}{(k-1)!(n-1)!}\frac{\beta^6}{\alpha\beta^5+120}\left(\alpha+\frac{\beta}{6}(ln(y))^6\right)y^{\beta-1}\\ &\times\left[\left[1+\frac{((\beta ln(y))^6-6(\beta ln(y))^5+30(\beta ln(y))^4-120(\beta ln(y))^3+360(\beta ln(y))^2-720(\beta ln(y)))}{6(\alpha\beta^5+120)}\right]y^{\beta}\right]^{k-1}\\ &\times\left[1-\left[1+\frac{((\beta ln(y))^6-6(\beta ln(y))^5+30(\beta ln(y))^4-120(\beta ln(y))^3+360(\beta ln(y))^2-720(\beta ln(y)))}{6(\alpha\beta^5+120)}\right]y^{\beta}\right]^{n-k} \end{aligned}$$

The pdf of the first order statistics Y_1 of log-Hamza distribution is given by

$$\begin{aligned} f_Y(1) =&n\frac{\beta^6}{\alpha\beta^5+120}\left(\alpha+\frac{\beta}{6}(ln(y))^6\right)y^{\beta-1}\\ &\times\left[1-\left[1+\frac{((\beta ln(y))^6-6(\beta ln(y))^5+30(\beta ln(y))^4-120(\beta ln(y))^3+360(\beta ln(y))^2-720(\beta ln(y)))}{6(\alpha\beta^5+120)}\right]y^{\beta}\right]^{n-1} \end{aligned}$$

The pdf of the n^{th} order statistics Y_n of log-Hamza distribution is given by

$$\begin{aligned} f_Y(1) =&n\frac{\beta^6}{\alpha\beta^5+120}\left(\alpha+\frac{\beta}{6}(ln(y))^6\right)y^{\beta-1}\\ &\times\left[\left[1+\frac{((\beta ln(y))^6-6(\beta ln(y))^5+30(\beta ln(y))^4-120(\beta ln(y))^3+360(\beta ln(y))^2-720(\beta ln(y)))}{6(\alpha\beta^5+120)}\right]y^{\beta}\right]^{n-1} \end{aligned}$$

6. Maximum Likelihood Estimation of log-Hamza Distribution

Let the random samples $y_1, y_2, y_3, ..., y_n$ are drawn from log-Hamza distribution. The likelihood function of n observations is given as

$$L=\prod_{i=1}^{n}\frac{\beta^6}{\alpha\beta^5+120}\left(\alpha+\frac{\beta}{6}(ln(y))^6\right)y^{\beta-1}$$

The log-likelihood function is given as

$$l =6nlog(\beta)-nlog(\alpha\beta^6+120)+\sum_{i=1}^{n}log(\alpha+\frac{\beta}{6}(log(y_i))^6)+(\beta-1)\sum_{i=1}^{n}logy_i \tag{10}$$

The partial derivatives of the log-likelihood function with respect to α and β are given as

$$\frac{\partial l}{\partial \alpha} = \frac{-n\beta^5}{(\alpha\beta^5+120)} + \sum_{i=1}^{n} \frac{6}{(6\alpha+\beta(ln(y_i))^6)} \tag{11}$$

$$\frac{\partial l}{\partial \beta} = \frac{6n}{\beta} - \frac{5n\alpha\beta^4}{(\alpha\beta^5+120)} + \sum_{i=1}^{n} \frac{(ln(y_i))^6}{6\alpha+\beta(ln(y_i))^6} + \sum_{i=1}^{n} log(y_i) \tag{12}$$

For interval estimation and hypothesis tests on the model parameters, an information matrix is required. The 2 by 2 observed matrix is

$$I(\xi) = \frac{-1}{n}\begin{bmatrix} E\left(\frac{\partial^2 logl}{\partial\alpha^2}\right) & E\left(\frac{\partial^2 logl}{\partial\alpha\partial\beta}\right) \\ E\left(\frac{\partial^2 logl}{\partial\beta\partial\alpha}\right) & E\left(\frac{\partial^2 logl}{\partial\beta^2}\right) \end{bmatrix}$$

The elements of above information matrix can be obtain by differentiating equations (12) and (13) again partially . Under standard regularity conditions when $n \to \infty$ the distribution of $\hat{\xi}$ can be approximated by a multivariate normal $N(0, I(\hat{\xi})^{-1})$ distribution to construct approximate confidence interval for the parameters. Hence the approximate $100(1-\psi)\%$ confidence interval for α and β are respectively given by

$$\hat{\alpha} \pm Z_{\frac{\psi}{2}}\sqrt{I_{\alpha\alpha}^{-1}(\hat{\xi})} and \quad \hat{\beta} \pm Z_{\frac{\psi}{2}}\sqrt{I_{\beta\beta}^{-1}(\hat{\xi})}$$

where

$$\frac{\partial^2 l}{\partial \alpha^2} = -\sum_{i=1}^{n} \frac{36}{(6\alpha+\beta(ln(y_i))^6)^2}$$

$$\frac{\partial^2 l}{\partial \beta^2} = \frac{-6n}{\beta^2} - \frac{5n\alpha\beta^3(120-\alpha\beta^5)}{(\alpha\beta^5+120)^2} - \sum_{i=1}^{n} \frac{(ln(y_i))^6}{(6\alpha+\beta(ln(y_i))^6)^2}$$

$$\frac{\partial^2 l}{\partial \alpha\partial\beta} = \frac{\partial^2 l}{\partial \beta\partial\alpha} = \frac{600n\beta^4}{(\alpha\beta^5+120)^2} + \sum_{i=1}^{n} \frac{6(ln(y_i))^6}{(6\alpha+\beta(ln(y_i))^6)^2}$$

7. Data Aanalysis

This subsection evaluates a real-world data set to demonstrate the log-Hamza distribution's applicability and effectiveness. The log-Hamza distribution (LHD) adaptability is determined by comparing its efficacy to that of other analogous distributions such as beta distribution (BD), Kumaraswamy distribution (KSD) and Topp-Leone distribution (TLD).
To compare the versatility of the explored distribution, we consider the criteria like AIC (Akaike information criterion), CAIC (Consistent Akaike information criterion), BIC (Bayesian information criterion) and HQIC (Hannan-Quinn information criterion). Distribution having lesser AIC, CAIC, BIC and HQIC values is considered better.

$$AIC = -2l + 2p, \qquad AICC = -2l + 2pm/\,(m-p-1), \qquad BIC = -2l + p(log(m))$$

$$HQIC = -2l + 2plog(log(m)) \qquad K.S = max_{1\le j\le m}\left(F(x_j) - \frac{j-1}{m}, \frac{j}{m} - F(x_j)\right)$$

Where $'l'$ denotes the log-likelihood function,'p'is the number of parameters and'm'is the sample size.

Data set: The followig obser vation are due to Caramanis et al [2] and Mazmumdar and Gaver [9], where they compare the two distinct algorithms called SC16 and P3 for estimating unit capacity factors. The values resulted from the algorith SC16 are
0.853, 0.759, 0.866, 0.809, 0.717, 0.544, 0.492, 0.403, 0.344, 0.213, 0.116, 0.116, 0.092, 0.070, 0.059, 0.048, 0.036, 0.029, 0.021, 0.014, 0.011, 0.008, 0.006.
The ML estimates with corresponding standard errors in parenthesis of the unknown parameters

Table 1: *Descriptive statistics for data set*

Min.	Max.	Ist Qu.	Med.	Mean	3rd Qu.	kurt.	Skew.
0.0060	0.8660	0.0325	0.1160	0.2881	0.5180	1.9741	0.7676

Table 2: *The ML Estimates (standard error in parenthesis) for data set*

Model	$\hat{\alpha}$	$\hat{\beta}$
LHD	1.9503	2.0969
	(1.5513)	(0.2355)
BD	0.4869	1.1679
	(0.1208)	(0.3577)
KSD	0.5043	0.0242
	(1.1862)	(0.3264)
TLD	0.5943	....
	(0.1239)	....

Table 3: *Comparison criterion and goodness-of-fit statistics for data set*

Model	-2l	AIC	AICC	BIC	HQIC	K.S statistic	p-value
LHD	-25.551	-21.551	-20.951	-19.280	-20.980	0.1034	0.9663
BD	-19.214	-15.214	-14.614	-12.943	-14.643	0.183	0.4202
KSD	-19.341	-15.341	-14.741	-13.070	-14.770	0.178	0.4526
TLD	-16.230	-14.230	-14.039	-13.094	-13.944	0.168	0.5273

are presented in Table 2 and the comparison statistics, AIC, BIC, CAIC, HQIC and the goodness-of-fit statistic for the data set are displayed in Table 3.

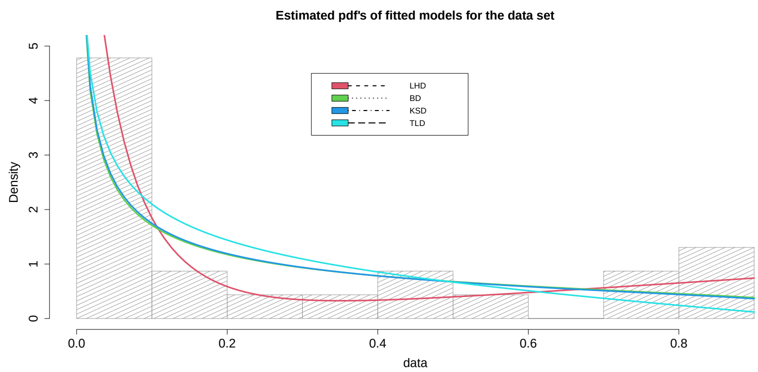

Figure 3

It is observed from table 3 that LHD provides best fit than other competative models based on the measures of statistics, AIC, BIC, AICC, HQIC and K-S statistic. Along with p-values of each model.

8. Conclusion

This study proposed a new two parameters distribution known as Log-Hamza distribution which is defined on unit interval and is used for modelling the real life data. Several structural properties of the proposed distribution including moments, moment generating function, order statistics and reliability measures has been discussed. The parameters of the distribution are estimated by famous method of maximum likelihood estimation. Finally the efficiency of the distribution is examined through an application when compared with Beta distribution, Kumaraswamy distribution and Topp-Leone distribution.

References

[1] A. Aijaz, M.Jallal, S.Q. Ain Ul and R.Tripathi. The Hamza distribution with statistical properties and applications. *Asian journal of probability and statistics*, *8 (2020)*, 28-42.

[2] M. Caramanis,J.Stremel, W. Fleck and S. Daneil. Probabilistic production costing: an investigation of alternative algorithms. *Internation journal of electrical power and energy system,5(2),(1983)*, 75-86.

[3] G.M. Corderio and M. de Castro. A new family of generalized distribution. *Journal of statistical computation and simulation,81(7)* (2011),75-86.

[4] Gomez-Deniz, M.A.Sordo, E. Calderin-Ojeda. The Log-Lindley distribution as an alternative to beta regression model with applications in insurance.. *Insurance: Mathematics and Economics,54 (2014)*, 49-57.

[5] M.E. Ghitnay, J. Mazucheli. A.F.B. Menezes and F. Alqallaf. The unit-Inverse Gaussian distribution: A new alternative to two-parameter distribution on the unit interval. *Commum. Stat. theory methods*, *48 (2019)*, 3423-3438.

[6] M.A Haq,S. Hashmi, K. Aidi, P.L. Ramos and F. Louzada. Unit modified Burr-III distribution: Estimation, characterization and validation test. *Ann. data sci.*, *87(15) (2020)*.

[7] J. Mazucheli, A.F.B. Menezes, L.B Fernandes, R.P. de Oliveria and M.E. Ghitney. The unit-Weibull distribution and associated inference. *J.Appl. probab. Stat*, *13 (2018)*,1-22.

[8] A.F.B. Menezes, J. Mazucheli and S. Dey. The unit-Gompertz distribution with applications. *Statistica*, *79 (2019)*,25-43.

[9] M. Mazumdar and D.P. Gaver. On the computation of power-generating system reliability indexes. *Tecnometrics*, *26(2) (2019)*,173-185.

[10] S. Nadarajah and S. Kotz. Moments of some J-shaped distribution. *journal of applied statistics*, *30(3) (2003)*,311-317.

[11] J. Rodrigues, J.L. Bazan and A.K. Aflexible procedure for formulating probability distribution on the unit-interval with applications. *Commun. stat. Theory methods*, *49 (2020)*, 738-754.

[12] C.W. Topp and F.C. Leone. A family of J-Shaped frequency function. *journal of the American statistical association*, *50(269) (1955)*, 209-219.