A NEW LENGTH BIASED EXPONENTIAL-EXPONENTIAL DISTRIBUTION AND ITS APPLICATIONS TO CANCER DATA

Latha J[1], M Vijayakumar[2], Shibu D S[3]

•

[1]Research Scholar, Department of Statistics,
Annamalai University, Annamalai Nagar, Tamil Nadu, India,
Assistant Professor, Department of Statistics, Christ Nagar College, Maranalloor,
Thiruvananthapuram, Kerala, India.
[2]Department of Statistics, Annamalai University,
Annamalai Nagar, Tamil Nadu, India.
[3]Department of Statistics, University of Kerala,
Thiruvananthapuram, Kerala, India
[1]lathalekshmi.j@gmail.com, [2]mvkumar1975@gmail.com, [3]statshibu@gmail.com

Abstract

The application of weighted distribution is used in many fields of real life such as medicine, reliability, ecology and so on, so we try to contribute in this field through this research. In this paper we propose a new length biased form of the Exponential-Exponential distribution called Length Biased Exponential-Exponential distribution (LBEED) and derive its statistical properties. The statistical properties including the moments, moment generating function, characteristic function, reliability functions and entropy measures are discussed. The method of maximum likelihood and method of moment was employed to estimate the parameter. The utility and adaptability of the proposed model are demonstrated by real-world data sets.

Keywords: exponential-exponential distribution, hazard function, length-biased distribution, maximum likelihood estimation, renyi entropy.

I. Introduction

The exponential distribution is the one that draws the greatest attention from researchers in the literature because of its broad variety of applications. Standard distribution can be understood with the help of weighted distributions, which also give distributions more flexibility. The form of the distribution of recorded observations can be influenced by ascertainment methods, as [5] examined. This is introduced and formulated by [11] in relation to statistical data modelling when standard distributions were found to be in appropriate. In addition to offering methods for expanding distributions for more flexibility in fitting a dataset, weighted distributions can be used to improve understanding. The length biased Rani distribution with survival data was introduced by [15]. The length biased weighted Lomax distribution with outliers was examined by [7]. Weighted distributions are used in many research in the domains of biology, branching processes, and reliability. According to [6], the idea of length biased distribution has several uses in the biomedical field, including family history, illness survival, intermediate events, and the latency period of AIDS caused by blood transfusion.

The weighted exponential families of distributions are introduced with the assistance of [17]. Weighted distributions come in two varieties: size-biased and length. The length biased Burr XII distribution and length-biased weighted exponentiated inverted Weibull distribution were discussed in [3] and [12]. The length-biased exponential distribution for estimating Bayesian reliability is discussed and addressed in [8]. Many writers have used the idea of weighted distributions for various reasons: [10] looked at several general models that result in weighted distributions with weight functions that do not have to be limited by one and researched length biased (size biased) sampling with relevance to wildlife populations and human families. The Length biased quasi-Sujatha distribution with properties and applications to bladder cancer data is addressed by [16]. Recently, [2] examined and studied the length biased powered inverse Rayleigh distribution and its applications. Also [4] proposed a weighted intervened exponential distribution as a lifetime model. A new method for generating families of continuous distributions is presented in [1] and is used in [9] and motivated to find a new distribution which is a length biased form.

The probability density function (PDF) of the Exponential-Exponential distribution (EED) by [1] and [9] is given as

$$f(x) = \theta^2 e^{-\theta^2 x}, x > 0, \theta > 0 \tag{1}$$

Assume X is a non-negative random variable with probability density function (pdf) f (x). Let W(x) be the weight function which is a non-negative function, then the probability density function of the weighted random variable X_w is given by:

$$f_w(x) = \frac{W(x)\, f(x)}{E(w(x))}, x > 0$$

where w (x) be a non-negative weight function and $E(w(x)) = \int w(x)\, f(x)\, dx < \infty$.

The PDF of the length biased distribution is defined by

$$f_L(x; \theta) = \frac{x f(x;\theta)}{E(X)}, -\infty < x < \infty \tag{2}$$

where $\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$, $-\infty < x < \infty$.

II. The Length Biased Exponential-Exponential Distribution (LBEED)

Let X be a random variable with PDF given in (1), respectively, then by using the transformation given in (2), and replacing the expected value as given in the formula; the PDF of the suggested length biased Exponential-Exponential distribution (LBEED) is defined by,

$$f(x;\theta) = x\theta^4 e^{-\theta^2 x}, x > 0, \theta > 0 \tag{3}$$

where θ is the scale parameter.

The CDF of the is length biased Exponential-Exponential distribution (LBEED) is given by,

$$F(x;\theta) = 1 - (x\,\theta^2 + 1)e^{-\theta^2 x}, x > 0, \theta > 0 \tag{4}$$

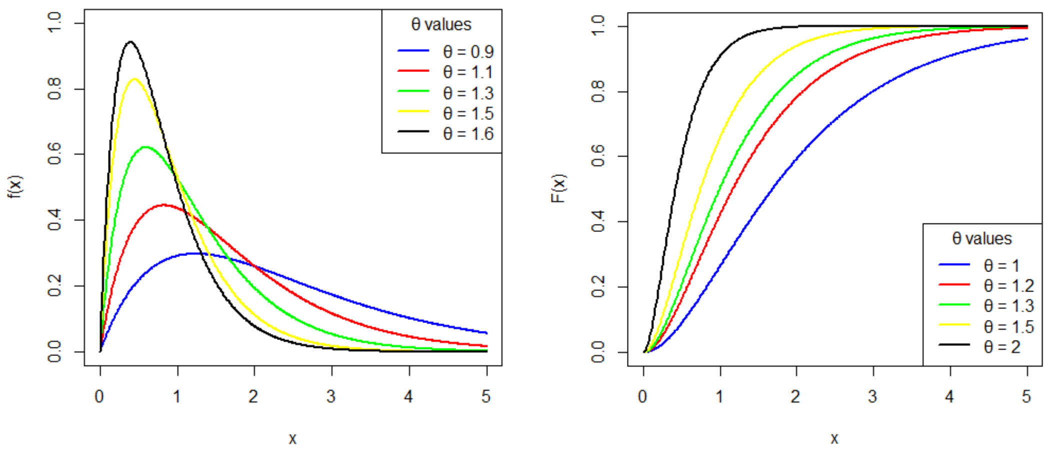

Figure 1: *PDF (left) and CDF (right) plots of LBEED for different values of the parameter θ.*

III. Reliability Analysis

I. Survival Function

This section deals with the survival function and hazard function of the proposed Length-biased Exponential-Exponential distribution.

$$S(x;\theta) = 1 - F(x;\theta)$$

$$S(x;\theta) = (x\,\theta^2 + 1)\, e^{-\theta^2 x}, \quad x > 0,\ \theta > 0 \tag{5}$$

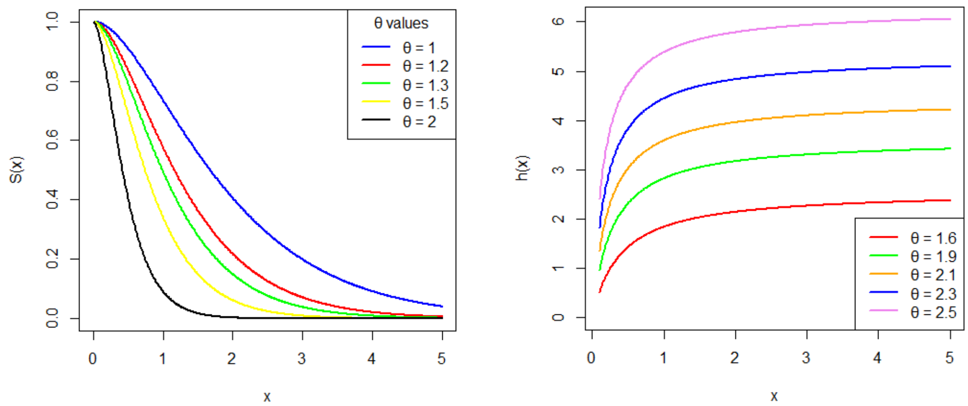

Figure 2: *Survival function (left) and Hazard function (right) of the LBEED for different values of the parameter θ.*

II. Hazard Function

The hazard function or failure rate of the proposed distribution is given by

$$H(x;\theta) = \frac{f(x;\theta)}{S(x;\theta)}$$

$$\mathrm{H}\,(x;\theta) = \frac{x\theta^4}{x\theta^2+1} \tag{6}$$

III. Reverse Hazard Function

The reverse hazard function of the Length-biased Exponential-Exponential distribution is given by,

$$h_r(x) = \frac{f(x;\theta)}{F(x;\theta)}$$

$$h_r(x) = \frac{x\,\theta^4 e^{-\theta^2 x}}{1-(x\,\theta^2+1)e^{-\theta^2 x}} \tag{7}$$

IV. Statistical Properties

The structural properties of length biased exponential- exponential distribution are included in this section.

I. Moments and related measures

Assume that X is the random variable that follows a length-biased Exponential-Exponential distribution (LBEED), then the rth order moment, that is $E(X^r)$ can be obtained as,

$$\mu_r' = E(X^r) = \int_0^\infty x^r f(x;\theta)\,dx$$

$$\mu_r' = \int_0^\infty x^r\,x\,\theta^4 e^{-\theta^2 x}\,dx$$

$$= \int_0^\infty x^{r+1}\,\theta^4 e^{-\theta^2 x}\,dx$$

Put $\theta^2 x = u,\ \theta^2\,dx = du$

$$\mu_r' = \frac{\theta^2}{\theta^{2(r+1)}}\int_0^\infty u^{r+1}\,e^{-u}$$

$$\mu_r' = \frac{\Gamma(r+2)}{\theta^{2r}} \tag{8}$$

The mean and variance can be obtained by substituting r= 1, 2, 3 and 4

$$\mu_1' = \frac{2}{\theta^2}$$

$$\mu_2' = \frac{6}{\theta^4}$$

$$\mu_3' = \frac{24}{\theta^6}$$

$$\mu_4' = \frac{120}{\theta^8}$$

$$\text{Variance, } \mu_2 = \frac{2}{\theta^4} \tag{9}$$

II. Skewness and Kurtosis

The coefficient of skewness is defined as $\sqrt{\beta_1} = \frac{\mu_3{}^2}{\mu_2{}^{\frac{3}{2}}}$ and is given by $\sqrt{\beta_1} = \frac{\frac{4}{\theta^6}}{\left(\frac{2}{\theta^4}\right)^{\frac{3}{2}}}$.

The coefficient of kurtosis gives the flatness or peakedness of the curve and it is defined as $\beta_2 = \frac{\mu_4}{\mu_2{}^2}$ and is given by, $\beta_2 = \frac{\frac{24}{\theta^8}}{\left(\frac{2}{\theta^4}\right)^2}$

III. Mode

The mode of the length biased exponential-exponential distribution is given by $x = \frac{1}{\theta^2}$.

IV. Moment Generating function and Characteristic function

The moment generating function and Characteristic function of length-biased Exponential-Exponential distribution are included in the section.

$$M_x(t) = E\left(e^{tx}\right)$$

$$= \int_0^\infty e^{tx}\, x\, \theta^4 e^{-\theta^2 x} dx$$

$$M_x(t) = \frac{\theta^4}{(\theta^2 - t)^2}, |t| < \theta^2 \quad (10)$$

Similarly, the Characteristic function is obtained as

$$\phi_t(x) = E\left(e^{itx}\right) = \frac{\theta^4}{(\theta^2 - it)^2} \quad (11)$$

V. Entropy Measures

The entropy of a random variable X is a measure of variation of the uncertainty. It plays important role in various fields, including probability and statistics, physics communication theory and economics by quantifying diversity and unpredictability. A system with higher entropy indicates greater uncertainty and more uniform outcomes.

I. Renyi Entropy

The Renyi entropy is also important in quantum information, where it can be used as a measure of entanglement. For a given probability distribution, Renyi entropy is given by,

$$H_\alpha(f(x,\theta)) = \frac{1}{1-\alpha} \log \int_0^\infty [f(x,\theta)]^\alpha\, dx$$

$$\int_0^\infty [f(x,\theta)]^\alpha\, dx = \int_0^\infty x^\alpha\, \theta^{4\alpha} e^{-\theta^2 x\alpha}\, dx$$

Put $u = x\alpha\theta^2$, $du = \alpha\theta^2 dx$

$$\int_0^\infty [f(x,\theta)]^\alpha \, dx = \frac{\theta^{4\alpha}}{(\alpha\theta^2)^{\alpha+1}} \; \Gamma(\alpha+1)$$

$$H_\alpha(f(x,\theta)) = \frac{1}{1-\alpha} \log\left[\frac{\theta^{4\alpha}}{(\alpha\theta^2)^{\alpha+1}} \; \Gamma(\alpha+1)\right]$$

$$H_\alpha(f(x,\theta)) = \frac{1}{1-\alpha} \left[4\alpha\log\theta + \log\Gamma(\alpha+1) - (\alpha+1)\log\alpha\theta^2\right] \tag{12}$$

II. Tsallis Entropy

A generalization of Boltzmann-Gibbs (B-G) statistical mechanics initiated by Tsallis has focused a great deal to attention. This generalization of B-G statistics was proposed firstly by introducing the mathematical expression of Tsallis entropy [14] for a continuous random variable is defined as follows:

$$S_\lambda = \frac{1}{\lambda-1}\left[1 - \int_0^\infty f(x)^\lambda \, dx\right]$$

$$S_\lambda = \frac{1}{\lambda-1}\left[1 - \int_0^\infty \left(x\,\theta^4\, e^{-x\theta^2}\right)^\lambda dx\right]$$

on simplification we get,

$$S_\lambda = \frac{1}{\lambda-1}\left[1 - \frac{\Gamma(\lambda+1)}{\lambda^{\lambda+1}\theta^2}\right] \tag{13}$$

VI. Bonferroni and Lorenz Curves

The Bonferroni and Lorenz curves have applications not only in economics to study income and poverty, but also in other fields like reliability, demography, insurance and medical science. The Bonferroni curve of the LBEED is defined as:

$$B(p) = \frac{1}{p\mu}\int_0^q x\, f(x)dx = \frac{1}{p\mu}\left[\mu - \int_q^\infty x\, f(x)dx\right]$$

$$B(p) = \frac{1}{p}\left[1 - \frac{1}{2}\left[e^{-\theta^2 q}\left[(\theta^2 q)^2 + 2(\theta^2 q) + 2\right]\right]\right] \tag{14}$$

The Lorenz curve of the LBEED is defined as:

$$L(p) = \frac{1}{\mu}\int_0^q x\, f(x)dx = \frac{1}{\mu}\left[\mu - \int_q^\infty x\, f(x)dx\right]$$

$$L(p) = \frac{\theta^2}{2}\left[\frac{2}{\theta^2} - e^{-\theta^2 q}\left[(\theta^2 q)^2 + 2(\theta^2 q) + 2\right]\right] \tag{15}$$

VII. Fisher's Information

The Fisher information determine the amount of information contained in an observable random variable X about an unknown distribution parameter upon which the probability of X depends. The Fisher's information is defined as,

$$I\ (\theta) = -E\left(\frac{d^2}{d\theta^2}\log\big(f(x)\big)\right)$$

The Fisher's information of Length biased exponential-exponential distribution is given by $I(\theta) = \frac{4}{\theta^2}$, as θ increases the Fisher's value decreases.

VIII. Order Statistics

Order statistics make their appearance in many statistical theory and practice. We know that if $X_{(1)}$, $X_{(2)}$,…, $X_{(n)}$ denotes the order statistics of a random sample X_1, X_2,…, X_n from a continuous population with cdf $F_x(x)$ and pdf $f_x(x)$, then the pdf of r^{th} order statistics $X_{(r)}$ is given by,

$$f_{x(r)}(x) = \frac{n!}{(r-1)!\,(n-r)!}\,.\,f_x(\mathrm{x}).\,[F_x(\mathrm{x})]^{r-1}\,[1-\ F_x(\mathrm{x})]^{n-r}$$

Using (3) and (4) in equation (16), we will obtain the probability density function of rth order statistics of length biased exponential-exponential distribution as

$$f_{x(r)}(x) = \frac{n!}{(r-1)!(n-r)!}\,x\,\theta^4 e^{-\theta^2 x}\left[1-(x\,\theta^2+1)e^{-\theta^2 x}\right]^{r-1}\left[(x\,\theta^2+1)e^{-\theta^2 x}\right]^{n-r} \tag{16}$$

The probability density function of first order statistic $X_{(1)}$ of length biased exponential-exponential distribution can be obtained as

$$f_{x(1)}(x) = n\,x\,\theta^4 e^{-\theta^2 x}\left[(x\,\theta^2+1)e^{-\theta^2 x}\right]^{n-1} \tag{17}$$

Similarly, the probability density function of higher order statistic $X_{(n)}$ of length biased exponential-exponential distribution can be obtained as

$$f_{x(n)}(x) = n\,x\,\theta^4 e^{-\theta^2 x}\left[1-(x\,\theta^2+1)e^{-\theta^2 x}\right]^{n-1} \tag{18}$$

IX. Parameter Estimation

In this section we estimate the parameter of length biased exponential-exponential distribution using two different methods such as maximum likelihood and method of moments.

I. Maximum Likelihood Estimation

In this section, the parameter estimation of length biased exponential-exponential distribution using maximum likelihood method is included. Let X_1, X_2, …, X_n be a random sample of size n from the length biased exponential-exponential distribution, then the likelihood function is given by,

$$L(x,\theta) = \prod_{i=1}^{n} f(x,\theta)$$

$$L(x,\theta) = \theta^4 \prod_{i=1}^{n} x\, e^{-\theta^2 x}$$

The log likelihood function is given by

$$\log L(x,\theta) = 4n \log\theta - \theta^2 \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \log x \quad (19)$$

Differentiating the equation (20) partially w.r.to θ and equating to zero, we get the following system of equation

$$\frac{4n}{\theta} - 2\theta \sum_{i=1}^{n} x_i = 0$$

The maximum likelihood estimator $\hat{\theta}$ of θ is given by

$$\hat{\theta} = \sqrt{\frac{2n}{\sum_{i=1}^{n} x_i}}.$$

II. Method of Moments

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the length biased exponential-exponential distribution. Equating population mean to the corresponding sample mean, the moment estimator $\hat{\theta}$ of θ of LBEED is given by $\hat{\theta} = \sqrt{\frac{2}{\bar{x}}}$.

X. Applications

The real-life application of length biased exponential-exponential distribution is included in this section and show that the length biased exponential-exponential distribution (LBEED) can be a better model than the exponential-exponential distribution (EED) and length biased exponential distribution (LBED). The data set is given below.

In order to compare the performance, length biased exponential- exponential distribution with exponential-exponential distribution. We are using the Criterion values like AIC (Akaike information criterion), AICC (corrected Akaike information criterion), BIC (Bayesian information criterion) and HQIC (Hannan-Quinn Information Criterion). The better distribution corresponds to lesser AIC, AICC and BIC values. The formulae for calculation of AIC, BIC, AICC and HQIC are;

AIC = $2k - 2\log L$, AICC = $AIC + \frac{2k\,(k+1)}{n-\,k-1}$, BIC = $k \log n - 2\log L$ and

HQIC = $2k \log\ (\log(n)) - 2\log L$

where n is the sample size, K is the number of parameters, and L denotes the likelihood function. Any probability model having smaller value of AIC, BIC and $-LogL$ being the best model to fit the data set.

Data Set 1: The following data set represents the survival times (in months) of 32 patients of melanoma studied by [13].

3.25, 3.5, 4.75, 4.75, 5, 5.25, 5.75, 5.75, 6.25, 6.5, 6.5, 6.75, 6.75, 7.78, 8, 8.5, 8.5, 9.25, 9.5, 9.5, 10, 11.5, 12.5, 13.25, 13.5, 14.25, 14.5, 14.75, 15, 16.25, 16.25, 16.5.

Table 1: *The measures, AIC, CAIC, BIC, and HQIC for the melanoma (Skin Cancer) data*

Distribution	MLE	-Log L	AIC	BIC	AICC	HQIC
LBEED	θ = 0.4619	94.47	190.93	192.40	191.06	191.42
EED	θ = 0.3266	103.621	209.241	210.707	209.375	209.727
LBED	θ = 9.376	106.827	215.654	217.120	215.787	216.140

Data Set 2: The data set reported by Efron24 represent the survival times of a group of 44 patients suffering from Head and Neck cancer disease and treated using a combination of radiotherapy and chemotherapy (RT+CT).
12.20, 23.56, 23.7, 25.9, 31.98, 37, 41.35, 47.38, 55.46, 58.36, 63.47 ,68.46, 78.3, 74.5, 81.43, 84, 92, 94, 110, 112, 119, 127, 130, 133, 140, 146, 155, 159, 173, 179, 194, 195, 209, 249, 281, 319, 339, 432, 469, 519, 633, 725, 817, 1776.

Table 2: *The measures, AIC, CAIC, BIC, and HQIC for the Head and Neck cancer data*

Distribution	MLE	-Log L	AIC	BIC	AICC	HQIC
LBEED	θ = 0.0945	54.9225	111.8449	111.4884	111.9401	110.2765
EED	θ = 0.0669	147.3185	296.637	296.2805	296.7322	295.0686
LBED	θ= 53.7	296.19	594.38	596.14	594.48	594.93

AIC, BIC, AICC, -Log L and HIQC values are lower in the length biased exponential-exponential distribution than in the exponential-exponential distribution, according to Table 1 and Table 2. Our conclusion is that the length biased exponential-exponential distribution provides a better match than the exponential-exponential distribution.

XI. Conclusion

In the present paper, the new distribution called length biased exponential-exponential distribution has been derived. The subject distribution is generated by using the length biased technique and taking the one parameter exponential-exponential distribution as the base distribution. Its various statistical properties including its mean, variance, survival function, hazard rate function, reverse hazard rate function, moment generating function, characteristic function, order statistics, entropies, Bonferroni and Lorenz curves have been discussed. Its parameters have also been estimated through the method of maximum likelihood estimation and method of moments has been observed. The proposed model is very flexible because its density function has many shapes; it can be right skewness, decreasing, and unimodal. Finally, a newly proposed length biased exponential-exponential has been investigated with real life data set to examine its ability and superiority.

References

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

[2] Mustafa, A. and Khan, M. I. (2022). The length-biased powered inverse Rayleigh distribution with applications. *J. Appl. Math. & Informatics*, 40(1-2), 1-13.

[3] Abd-Elfattah, A. M., Mahdy, M. and Ismail, G. (2021). Length biased Burr- XII distribution: properties and application. *International Journal of Sciences: Basic and Applied Research (IJSBAR),* 56(1), 222-244.

[4] Bhat, V. A. and Pundir, S. (2023): Weighted intervened exponential distribution as a lifetime distribution, *Reliability: Theory & Applications*, 2(73), 253-266.

[5] Fisher R. A., The effects of methods of ascertainment upon the estimation of frequencies (1934). *Annals of Eugenics*.6:13-25.

[6] Gupta, R.C. and Kirmani, S. (1990). The role of weighted distributions in stochastic modelling, *Communications in Statistics, Theory and Methods*. 19, 3147- 3162.

[7] Karimi, H. and Nasiri, P. (2018). Estimation parameter for length biased weighted Lomax distributions in the presence of outliers. *Mathematical and Computational Applications*, doi:10.3390/mca23010009.

[8] Mathew, J. and George, S. (2020). Length biased exponential distribution as a reliability model: a Bayesian approach, *Reliability: Theory & Applications*, 3(58):84-91.

[9] Ogunwale, O.D., Ayeni, T.M. (2022) The New Exponential-Exponential Distribution: Theory and Properties. *Quest Journals- Journal of Research in Applied Mathematics*, 8(8) :13-19.

[10] Patil, G. P. and Rao, C. R. (1978). Weighted distributions and Size biased sampling with applications to wildlife populations and human families. *Biometrics*, 34, 179-189.

[11] Rao C. R., (1965) On discrete distributions arising out of method of ascertainment, in classical and Contagious Discrete, G.P. Patiled; *Pergamum Press and Statistical publishing Society*, Calcutta;. p. 320-332.

[12] Saghir, A., Tazeem, S. and Ahmad, I. (2016). The length-biased weighted exponentiated inverted Weibull distribution. *Cogent Mathematics*, 3(1), 1-18.

[13] Susarla V, Vanryzin JV. Empirical Bayes estimation of a distribution (Survival) function from right Censored observations. *The Annals of Statistics*, 1978. doi:10.1214/aos/ 1176344249, 6(4), 740-75.

[14] Tsallis, C. (1988), Possible generalization of boltzmann-gibbs statistics. *Journal of Statistical Physics*, 52, 479-487

[15] Vijayakumar, M., Anwar Shameem, Z. A. and Punathumparambath, B. (2020). Length biased Rani distribution with survival data analysis. *High Technology Letters*, 26(6), 559-573.

[16] Subramanian, C. and Shenbagaraja, R. (2020). Length biased quasi-Sujatha distribution with properties and applications to bladder cancer data. *Journal of Information and Computational Science*, 10(1), 493- 506.

[17] Zubair Ahmad, Gholamhossien Hamedani, and Mohammed Elgarhy (2019). The Weighted Exponentiated Family of Distributions: Properties, Applications and Characterizations. *JIRSS* Vol. 19, No. 01, pp 209-228