Austrian Journal of Statistics
July 2022, Volume 51, 60–75.
http://www.ajs.or.at/
doi:10.17713/ajs.v51i4.1338

Exponential Transformed Inverse Rayleigh Distribution: Statistical Properties and Different Methods of Estimation

Proloy Banerjee
Aliah University
Department of Mathematics and Statistics

Shreya Bhunia
Aliah University
Department of Mathematics and Statistics

Abstract

In this article a generalization of the inverse Rayleigh distribution has been addressed by using DUS transformation, named as Exponential Transformed Inverse Rayleigh (ETIR) distribution. Some of the statistical properties of this newly proposed distribution like mode, quantiles, moment, moment generating function, survival and hazard rate function have been studied comprehensively. To estimate the parameter of this distribution, four different estimation procedures, such as maximum likelihood estimation (MLE), maximum product spacing method (MPS), least square method (LSE) and weighted least square method (WLSE) are briefly discussed. Performance of these estimates are compared using extensive simulations. As an application point of view the model superiority is verified through two real datasets.

Keywords: lifetime distribution, probability distribution, hazard rate function, maximum likelihood estimation, maximum product spacing method, simulation.

1. Introduction

Modeling and analyzing lifetime data are vital in the field of applied sciences such as medical, finance, engineering etc. Several lifetime distributions like Exponential, Weibull, Gamma and many more play an important role in this context. The consistency and accuracy of statistical analysis is highly influenced by the undertaken probability model or distribution. Owing to this fact, in recent decades developing new distributions has been a standard concept in statistical theory; this is usually accomplished by adding an additional parameter to the baseline distribution Ali, Khalil, Ijaz, and Saeed (2021).

Gupta, Gupta, and Gupta (1998) proposed the cummulative distribution function(CDF) of a new distribution by introducing a shape parameter ($\alpha > 0$) in the CDF of baseline distribution. Gupta and Kundu (2001), Seenoi, Supapakorn, and Bodhisuwan (2014) etc. further used this generalization techniques to develop more flexibile probability models. Another well known idea for generalizing baseline distribution is to transmute it by using Quadratic rank transmutation map(QRTM), described by Shaw and Buckley (2009). On the basis of QRTM, a number of generalizations have been developed in recent years. For example Transmuted extreme value distribution studied by Aryal and Tsokos (2009), Transmuted Rayleigh distribution by Merovci (2013), Khan and King (2015) developed Transmuted modified inverse Rayleigh distribution, Transmuted log-logistic distribution by Aryal (2013) and many more. By adding two more parameters to a continuous distribution, Cordeiro, Ortega, and da Cunha (2013) suggested a new class of distribution. Kumaraswamy (1980) proposed a different way to obtain a new distribution by using the baseline distribution. The primary goal of this modification of baseline distributions is to model real data with non-constant hazard rate functions.

One thing to keep in mind is that all the generalization method mentioned above incorporates few extra parameter(s) in the original model. That additional parameter(s) in one sense provides more flexibility to the existing distribution in analyzing complex data structure but on the other side, complexity in parameter estimation and others inferential procedure increases. Taken into account these difficulties, few new transformation techniques are proposed by Kumar et al., where any additional parameter(s) is not introduced other than those involved in the baseline distribution. For example, DUS transformation Kumar, Singh, and Singh (2015a), Sine transformation Kumar, Singh, and Singh (2015b), MG transformation Kumar, Singh, and Singh (2016). In all cases Exponential distribution is the considered baseline distribution. Using Weibull distribution M transformation Kumar, Singh, Singh, and Mukherjee (2017) is available in literature.

In the present study, we use DUS transformation to get a new lifetime distribution by incorporating Inverse Rayleigh distribution as original continuous distribution. If the PDF and CDF of a baseline lifetime distribution are g(x) and G(x) respectively, then the PDF of a new lifetime distribution is defined as:

$$f(x) = \frac{1}{e-1} g(x) e^{G(x)} \tag{1}$$

and the corresponding CDF and hazard rate function are as follows:

$$F(x) = \frac{1}{e-1} \left[e^{G(x)} - 1 \right] \tag{2}$$

$$h(x) = \frac{1}{e - e^{G(x)}} g(x) e^{G(x)} \tag{3}$$

respectively.

The uniqueness of this article stems from the fact that we offer a detailed overview of the mathematical and statistical properties for new distribution in the hope that it will be useful in lifetime data analysis. The primary objective of this study lies in two ways. The first is analytical, demonstrating that the studied Exponential Transformed Inverse Rayleigh distribution outperforms few well known distributions in the context of two real datasets. Secondly, for various sample sizes and parameter values we compare the performance of different frequentist estimators of this distribution.

The paper is organized as follows: In section (2), Exponential Transformed Inverse Rayleigh distribution is derived and graphically presented. Survival function and Hazard function are discussed in subsequent subsections. Various mathematical and statistical properties of the new distribution like raw moments, moment generating function, mode, quantiles and order statistics etc. are derived in section (3). From frequentists view point, four methods of parameter estimation are discussed in section (4). Comparison of estimators through simulation procedure has been done in section (5). In section (6), model superiority is studied through two real datasets. Finally, in section (7) we conclude the paper.

2. Exponential transformed inverse Rayleigh distribution

A random variable X is said to follow inverse Rayleigh distribution with scale parameter $\sigma > 0$ if its probability density function(PDF) is given by:

$$g(x;\sigma) = \frac{2\sigma^2}{x^3} e^{-(\frac{\sigma}{x})^2} \; ; \quad x > 0, \quad \sigma > 0 \tag{4}$$

and the corresponding cumulative distribution function(CDF) is

$$G(x;\sigma) = e^{-(\frac{\sigma}{x})^2} \; ; \quad x > 0, \quad \sigma > 0 \tag{5}$$

Now considering the inverse Rayleigh as a baseline distribution mentioned in equations (4) and (5) along with the DUS transformation denoted as in equations (1) and (2), we obtain a new distribution, named as Exponential Transformed Inverse Rayleigh distribution. In further sections, we use the abbreviation ETIR for the newly obtained distribution. The expressions of PDF and the corresponding CDF for ETIR distribution is given below:

$$F(x;\sigma) = \frac{1}{e-1}\left\{e^{e^{-(\frac{\sigma}{x})^2}} - 1\right\} \; ; \quad x > 0 \;\; \sigma > 0 \tag{6}$$

$$f(x;\sigma) = \left(\frac{2\sigma^2}{e-1}\right)\frac{1}{x^3} e^{e^{-(\frac{\sigma}{x})^2}} e^{-(\frac{\sigma}{x})^2} \; ; \quad x > 0 \;\; \sigma > 0 \tag{7}$$

It has to be noted that the ETIR is a extension of the inverse Rayleigh distribution. The

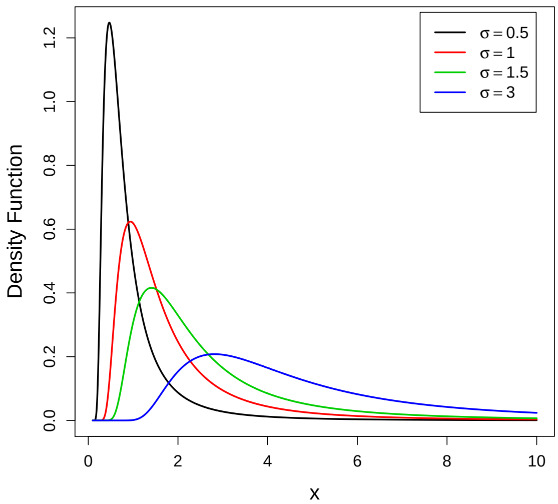


Figure 1: PDF of ETIR distribution for different scale parameter

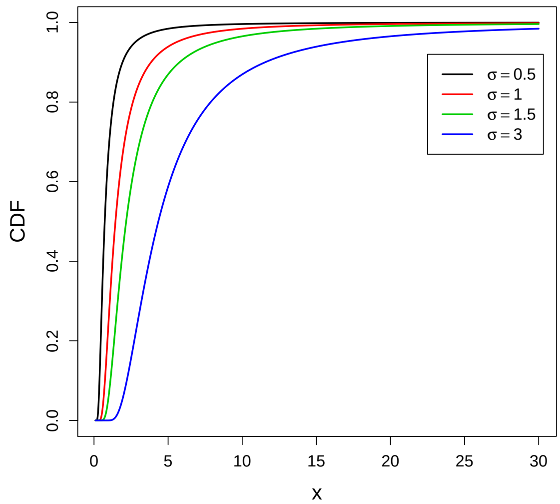


Figure 2: CDF of ETIR distribution for different scale parameter

newly obtained model has given more flexibility to analyze complex datasets. Figure 1 and 2 illustrate the possible shape of the PDF and CDF of ETIR distribution respectively. It is clear from the figures that the ETIR distribution is highly right skewed and useful in modeling positively skewed datasets.

2.1. Survival function and hazard rate function

Two most important inter-related probability measures for the life time distribution are Survival function and Hazard function. Both the measures generally used to describe and model the intrinsic characteristics of several survival data. The survival function is denoted as $S(x) = P(X > x) = 1 - F(x)$; and defined as the probability that an individual or an item is survived at least x unit of time. Similarly, the hazard rate function or failure rate function of an item or an individual at a time point x is propensity to fail in the next short interval of time $[x,\ x + \Delta x]$ given that it has survived upto time x. The hazard rate function, $h(x)$ is given as:

$$h(x) = \Delta_{x\to 0}\lim \frac{P\left(x < X < x + \Delta x | X > x\right)}{\Delta x} = \frac{f(x;\sigma)}{S(x;\sigma)}\ ; \quad S(x;\sigma) > 0$$

The survival and hazard rate function for the ETIR distribution is given by the following expressions (8) and (9) respectively:

$$S(x) = \frac{1}{e-1}\left\{e - e^{e^{-\left(\frac{\sigma}{x}\right)^2}}\right\} \tag{8}$$

$$h(x) = \frac{\frac{2\sigma^2}{x^3}e^{\left(-\frac{\sigma}{x}\right)^2}}{e^{1-e^{-\left(\frac{\sigma}{x}\right)^2}} - 1} \tag{9}$$

In Figure 3, it is seen that the shape of the hazard function curve first increases and then it is started to decrease and finally it is converged to some constant value. In survival analysis, it is well known that those lifetime distributions are very useful whose shape of the hazard rate first increases and then decreases. For an example, the hazard curve of infant mortality rate has similarity with this lifetime distribution. Over a time period, the infant mortality rate increases but after the infants got immunity in their body and also the improvement of medical treatment, the curve of the mortality rate decreases. Kotz and Nadarajah (2000), Rao and Mbwambo (2019) is recommended for further details.

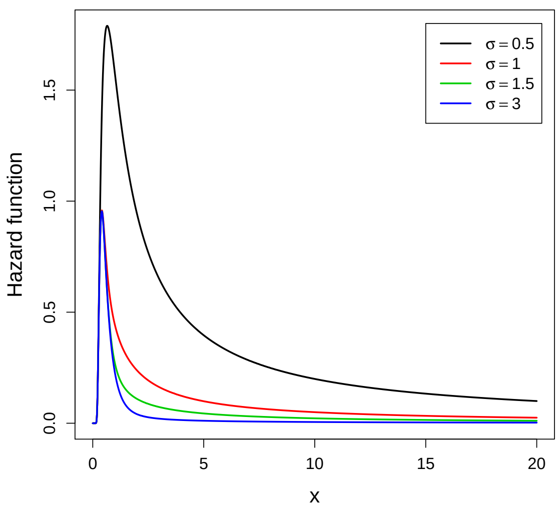


Figure 3: Hazard rate function for different scale parameter

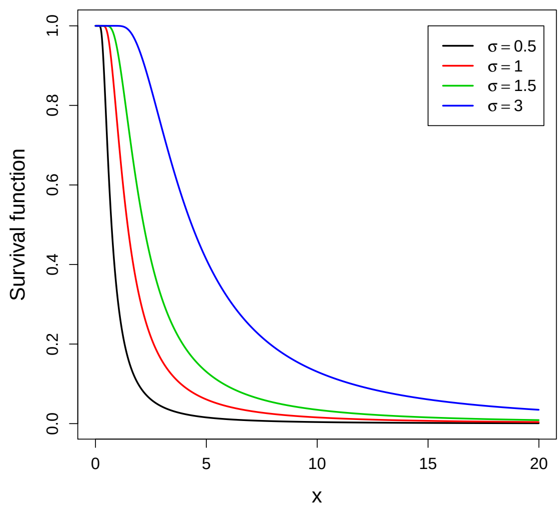


Figure 4: Survival function for different scale parameter

2.2. Shape of ETIR distribution

The shape of the distribution is important since it provides details about the nature of the distribution. Glaser (1980) proposed a theorem to find the shape of the hazard rate mathematically. According to the theorem, $\eta(t) = -\frac{f'(t)}{f(t)}$; where f(t) is continuous and twice differentiable on the interval $(0, \infty)$.

If $\eta'(t) > 0 \quad \forall\, t > 0$, then the hazard rate is increasing and if $\eta'(t) < 0 \quad \forall\, t > 0$, then the hazard rate is decreasing. So, we have

$$\eta(x) = \frac{3}{x} - \frac{2\sigma^2}{x^3}\left(1 + e^{-\left(\frac{\sigma}{x}\right)^2}\right)$$

Therefore,

$$\eta'(x) = -\frac{3}{x^2} - \frac{4(\sigma^2)^2}{x^6} e^{-(\frac{\sigma}{x})^2} + \frac{6\sigma^2}{x^4}\left(1 + e^{-(\frac{\sigma}{x})^2}\right) \tag{10}$$

The last term of equation (10) attains the minimum value zero as $x \to \infty$. Therefore it is clearly seen that $\eta'(x) < 0$ i.e; the distribution has decreasing hazard rate.

2.3. Random number generation

To generate random numbers from the ETIR distribution, the inversion method or the inverse transformation method is considered. The algorithm is to first generate a random number U from $Uniform(0, 1)$ distribution and then the equation $x = F^{-1}(U)$ provides the random number x from the ETIR distribution. Here we have,

$$\frac{e^{e^{-(\frac{\sigma}{x})^2}} - 1}{e - 1} = U$$

Therefore,

$$x = \frac{\sigma}{\sqrt{-ln\left[ln\left(U(e-1)+1\right)\right]}} \tag{11}$$

By using equation (11), for known values of scale parameter σ, we can easily generate random numbers of size n from the ETIR distribution.

3. Statistical properties of ETIR distribution

In this section, some basic and significant statistical and mathematical measures of the Exponential Transformed Inverse Rayleigh(ETIR) distribution such as moments, moment generating function, characteristic function, cumulant generating function, mode, quantile function and order statistics are derived and discussed.

3.1. Raw moments

The r^{th} order raw moment about origin, μ'_r of the proposed distribution with having PDF (7) is obtained as follows:

$$\mu'_r = E(X^r)$$

$$\mu'_r = \frac{\sigma^r}{e-1}\Gamma(1 - \frac{r}{2}) \sum_{k=0}^{\infty} \frac{1}{k!}(k+1)^{\frac{r}{2}-1}; \qquad for \quad r < 2 \tag{12}$$

Now in equation (12) if we put $r = 1$; equation reduces to

$$\mu'_1 = \frac{\sigma\sqrt{\pi}}{e-1} \sum_{k=0}^{\infty} \frac{1}{k!}(k+1)^{-\frac{1}{2}}$$

So, we have the mean μ'_1 of our newly proposed ETIR distribution. From equation (12) it can be seen that, for $r = 2$, μ'_2 becomes undefined. Hence the variance and the higher order raw moments of this distribution cannot be derived as the gamma function is defined only for positive numbers.

3.2. Moment generating function, characteristic function and cumulant generating function

Many interesting characteristics and features of a distribution can be obtained through studying the generating functions of that distribution.

The Moment generating function (MGF) of a random variable X having the proposed distribution is given as follows:

$$M_X(t) = E(e^{tX})$$

$$M_X(t) = \frac{1}{e-1}\sum_{k=0}^{\infty}\frac{1}{k!}\sum_{m=0}^{\infty}\frac{t^m}{m!}\sigma^m(k+1)^{\frac{m}{2}-1}\Gamma\left(1-\frac{m}{2}\right)$$

Similarly, the Characteristic function of X can be found as,

$$\phi_X(t) = E(e^{itX})$$

$$\phi_X(t) = \frac{1}{e-1}\sum_{k=0}^{\infty}\frac{1}{k!}\sum_{m=0}^{\infty}\frac{(it)^m}{m!}\sigma^m(k+1)^{\frac{m}{2}-1}\Gamma\left(1-\frac{m}{2}\right)$$

Where $i=\sqrt{-1}$ denotes imaginary number.

The Cumulant generating function of X is:

$$K_X(t) = ln\left(M_X(t)\right)$$

$$K_X(t) = ln\left[\frac{1}{e-1}\sum_{k=0}^{\infty}\frac{1}{k!}\sum_{m=0}^{\infty}\frac{t^m}{m!}\sigma^m(k+1)^{\frac{m}{2}-1}\Gamma\left(1-\frac{m}{2}\right)\right]$$

3.3. Mode

Mode is the value which have maximum probability area for any distribution. So, the mode for ETIR distribution is the value for which $f(x;\sigma)$ equation (7) is maximum. Hence, mode is the solution of $f'(x;\sigma)=0$; and for which $f''(x;\sigma)<0$.

So, differentiating equation (7) with respect to x and equating to zero, we get

$$\frac{2\sigma^2}{x^3(e-1)}e^{e^{-(\frac{\sigma}{x})^2}}e^{-(\frac{\sigma}{x})^2}\left[\frac{2\sigma^2}{x^3}+\frac{2\sigma^2}{x^3}e^{-(\frac{\sigma}{x})^2}-\frac{3}{x}\right]=0 \tag{13}$$

Clearly, equation (13) cannot be solved analytically. Therefore, one can solve equation (13) numerically by using some numerical iteration techniques, particularly here we prefer Newton-Raphson method.

3.4. Quantile function

To calculate the Quantile function $\Pi(p)$, $0<p<1$ of a random variable X associated with a probability density function (7) we solve the following equation:

$$F\left(\Pi(p)\right) = P(X \le \Pi(p)) = p \tag{14}$$

To find the 1^{st}, 2^{nd} and 3^{rd} quartile, we calculate the above equation for $p=0.25,\ 0.50$ and 0.75 respectively.

Hence, the 2^{nd} quartile or the median of the proposed distribution is:

$$Q_2 = \frac{\sigma}{\sqrt{-ln\left[ln\left(\frac{e+1}{2}\right)\right]}}$$

Similarly, by solving the equation (14) for $p=0.25$ and 0.75, we get 1^{st} and 3^{rd} quartile respectively,

$$Q_1 = \frac{\sigma}{\sqrt{-ln\left[ln\left(\frac{e+3}{4}\right)\right]}} \quad and \quad Q_3 = \frac{\sigma}{\sqrt{-ln\left[ln\left(\frac{3e+1}{4}\right)\right]}}$$

3.5. Order statistics

In reliability theory and quality control testing, order statistic plays an important role to predict time to fail of a certain item by considering few early failures Dey, Raheem, and Mukherjee (2017). Let $X_1 < X_2 < ... < X_n$ be an ordered sample from a continuous distribution with CDF $F_X(x)$ and PDF $f_x(x)$. Then the PDF of r^{th} order statistic $X_{(r)}$ is given by:

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!} f_X(x)(F_X(x))^{r-1} \left[1 - F_X(x)\right]^{n-r}; \quad r = 1, 2, ..., n$$

So, for the ETIR distribution PDF of the r^{th} order statistic is given as:

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!} \frac{2\sigma^2}{x^3(e-1)^n} e^{e^{-(\frac{\sigma}{x})^2}} e^{-(\frac{\sigma}{x})^2} \left(e^{e^{-(\frac{\sigma}{x})^2}} - 1\right)^{r-1} \left(e - e^{e^{-(\frac{\sigma}{x})^2}}\right)^{n-r} \quad (15)$$

Smallest order statistic is always the minimum of the sample, i.e; $X_{(1)} = \min\{X_1, X_2, ..., X_n\}$ and the largest order statistic is the maximum of the sample, i.e; $X_{(n)} = \max\{X_1, X_2, ..., X_n\}$. Expressions for the smallest and largest order statistics are obtained by substituting r = 1 and n in equation (15) respectively. The corresponding CDF of the r^{th} order statistic is obtained as:

$$F_{X_{(r)}}(x) = \sum_{i=r}^{n} \binom{n}{i} F_X^i(x)[1 - F_X(x)]^{n-i}$$

$$F_{X_{(r)}}(x) = \sum_{i=r}^{n} \binom{n}{i} \frac{\left(e^{e^{-(\frac{\sigma}{x})^2}} - 1\right)^i \left(e - e^{e^{-(\frac{\sigma}{x})^2}}\right)^{n-i}}{(e-1)^n} \quad (16)$$

4. Estimation of the scale parameter of the ETIR distribution

In Statistics, estimating the unknown parameter(s) for the given sample is an important step towards understanding the probabilistic model fully. Several estimation procedures under classical as well as Bayesian paradigm are available in literature, for more details see Louzada, Ramos, and Perdoná (2016), Dey, Dey, and Kundu (2014), Kundu and Raqab (2005), Mazucheli, Ghitany, and Louzada (2016), Fan (2015) etc. In this study our focus is to estimate the unknown scale parameter σ of ETIR distribution under frequentists approaches. Here we briefly describe four estimation procedures namely Maximum likelihood method (MLE), Maximum product spacings method (MPS), Least square method (LSE) and Weighted least square method (WLSE) respectively.

4.1. Maximum likelihood estimation method

A number of desirable properties for a good estimator such as consistency, asymptotic efficiency, invariance property etc. are satisfied by Maximum likelihood estimation method (MLE). This makes the MLE one of the most frequently used techniques for parameter estimation. Let $x_1, x_2, ..., x_n$ be the sample of size n, drawn from the ETIR distribution with PDF given in equation (7).

The likelihood function for σ is given by:

$$L(\sigma; X) = \left(\frac{2\sigma^2}{e-1}\right)^n \prod_{i=1}^{n} \left(\frac{1}{x_i^3}\right) e^{\sum_{i=1}^{n} e^{-\left(\frac{\sigma}{x_i}\right)^2}} e^{-\sigma^2 \sum_{i=1}^{n} \frac{1}{x_i^2}} \quad (17)$$

Taking logarithm on both sides, the log - likelihood function becomes:

$$lnL(\sigma; X) = nln\left(\frac{2}{e-1}\right) + 2nln\sigma + \sum_{i=1}^{n} ln\left(\frac{1}{x_i^3}\right) + \sum_{i=1}^{n} e^{-\left(\frac{\sigma}{x_i}\right)^2} - \sigma^2 \sum_{i=1}^{n} \frac{1}{x_i^2}$$

Therefore for maximizing the equation (17), we differentiate the above log-likelihood function with respect to σ and equate it to zero and get the following expression:

$$\sigma^2 \sum_{i=1}^{n} e^{-\left(\frac{\sigma}{x_i}\right)^2} \frac{1}{x_i^2} + \sigma^2 \sum_{i=1}^{n} \frac{1}{x_i^2} = n \tag{18}$$

It is obvious that equation (18) is not written explicitly and hence it cannot be solved analytically. So, to obtain $\hat{\sigma}_{MLE}$ we have to solve the above non-linear equation numerically. For this purpose some iteratively based procedures such as the Newton-Raphson algorithm is recommended.

4.2. Maximum product spacing method

Maximum product spacing method(MPS) was introduced by Cheng & Amin(1979, 1983) in the context of the parameter(s) estimation of continuous univariate distributions as an alternative to MLE. Further Ranneby (1984), independently developed the method as an approximation to the Kullback-Leibler measure of information. Invariance property of MPS estimators have been discussed by Coolen and Newby (1991). In the situation where the MLE method fails to provide consistent estimator due to the unbounded likelihood function many authors suggested the use of MPS. A comprehensive study on MPS method has been done by Shah and Gokhale (1993), Shao and Hahn (1999), Ghosh and Jammalamadaka (2001), Wong and Li (2006), Singh, Singh, and Singh (2014) etc.

let us suppose that $x_1 < x_2 < ... < x_n$ be an ordered sample from ETIR with CDF given in equation (6). The spacing $D_i's$ is defined as follows:

$$D_i = F(x_i|\sigma) - F(x_{i-1}|\sigma) \; ; \quad i = 1, 2, ..., n+1$$

Where, $F(x_0|\sigma) = 0$ and $F(x_{n+1}|\sigma) = 1$ such that $\sum_{i=1}^{n+1} D_i = 1$. In MPS method, the estimate of σ maximizes the logarithm of the geometric mean of sample spacing.

$$\begin{aligned}
S_n(\sigma) &= ln\left[\left(\prod_{i=1}^{n+1} D_i\right)^{\frac{1}{n+1}}\right] \\
&= \frac{1}{n+1}\left[lnD_1 + lnD_2 + ... + lnD_{n+1}\right] \\
&= \frac{1}{n+1}\left[lnF(x_1) + \sum_{i=2}^{n} ln\left(F(x_i) - F(x_{i-1})\right) + ln\left(1 - F(x_n)\right)\right] \\
&= \frac{1}{n+1}\left[ln\left(\frac{e^{e^{-\left(\frac{\sigma}{x_1}\right)^2}} - 1}{e-1}\right) + \sum_{i=2}^{n} ln\left(\frac{e^{e^{-\left(\frac{\sigma}{x_i}\right)^2}} - e^{e^{-\left(\frac{\sigma}{x_{i-1}}\right)^2}}}{e-1}\right) + ln\left(\frac{e - e^{e^{-\left(\frac{\sigma}{x_n}\right)^2}}}{e-1}\right)\right] \\
&= \frac{1}{n+1}\left[ln\left(e^{e^{-\left(\frac{\sigma}{x_1}\right)^2}} - 1\right) - ln(e-1) + \sum_{i=2}^{n}\left\{ln\left(e^{e^{-\left(\frac{\sigma}{x_i}\right)^2}} - e^{e^{-\left(\frac{\sigma}{x_{i-1}}\right)^2}}\right) - ln(e-1)\right\}\right. \\
&\quad \left. + ln\left(e - e^{e^{-\left(\frac{\sigma}{x_n}\right)^2}}\right) - ln(e-1)\right]
\end{aligned} \tag{19}$$

Now differentiating (19) with respect to σ and equating it with zero we get,

$$\frac{e^{e^{-\left(\frac{\sigma}{x_1}\right)^2}} e^{-\left(\frac{\sigma}{x_1}\right)^2} \frac{1}{x_1^2}}{1 - e^{e^{-\left(\frac{\sigma}{x_1}\right)^2}}} + \sum_{i=2}^{n}\left\{\frac{e^{e^{-\left(\frac{\sigma}{x_{i-1}}\right)^2}} e^{-\left(\frac{\sigma}{x_{i-1}}\right)^2} \frac{1}{x_{i-1}^2} - e^{e^{-\left(\frac{\sigma}{x_i}\right)^2}} e^{-\left(\frac{\sigma}{x_i}\right)^2} \frac{1}{x_i^2}}{e^{e^{-\left(\frac{\sigma}{x_i}\right)^2}} - e^{e^{-\left(\frac{\sigma}{x_{i-1}}\right)^2}}}\right\} + \frac{e^{e^{-\left(\frac{\sigma}{x_n}\right)^2}} e^{-\left(\frac{\sigma}{x_n}\right)^2} \frac{1}{x_n^2}}{e - e^{e^{-\left(\frac{\sigma}{x_n}\right)^2}}} = 0 \tag{20}$$

Since this is a non-linear equation, not having a closed form solution, it cannot be solved analytically. We prefer some iterative procedure such as Newton-Raphson method to estimate the parameter.

4.3. Least square and weighted least square method

Swain, Venkatraman, and Wilson (1988) proposed Least square estimaton (LSE) and Weighted least square estimaton (WLSE) procedure to estimate the parameters of Beta distributions. They described the procedures involving a continuous target distribution function $F_X(.)$ from which random sample $\{X_j : 1 \leq j \leq n\}$ with the corresponding order statistics $\left\{X_{(j)} : 1 \leq j \leq n\right\}$ are drawn. let $x_1 < x_2 < ... < x_n$ be an ordered sample then,

$$E\left(F(x_j)\right) = \frac{j}{n+1}; \qquad j = 1, 2, ..., n$$

The LSE of the unknown parameter σ of ETIR distribution is obtained by minimizing S with respect to σ where,

$$\begin{aligned} S &= \sum_{j=1}^{n} \left[F(x_j) - \frac{j}{n+1}\right]^2 \\ &= \sum_{j=1}^{n} \left[\frac{e^{e^{-\left(\frac{\sigma}{x_j}\right)^2}} - 1}{e-1} - \frac{j}{n+1}\right]^2 \end{aligned} \tag{21}$$

We differentiate the equation (21) with respect to σ and equate to zero to get the least square estimator $\hat{\sigma}_{LSE}$.

$$\sum_{j=1}^{n} \left[\frac{e^{e^{-\left(\frac{\sigma}{x_j}\right)^2}} - 1}{e-1} - \frac{j}{n+1}\right] e^{e^{-\left(\frac{\sigma}{x_j}\right)^2}} e^{-\left(\frac{\sigma}{x_j}\right)^2} \frac{1}{x_j^2} = 0 \tag{22}$$

The above equation is non-linear. Hence we solve it numerically. The weighted least square estimator (WLSE) of the unknown parameter can be obtained by minimizing

$$S = \sum_{j=1}^{n} w_j \left[F(x_j) - E(F(x_j))\right]^2$$

with respect to σ, where w_j be the weight function at the j^{th} point. Therefore, the weighted least square estimator of σ say $\hat{\sigma}_{WLSE}$ can be obtained by minimizing

$$S = \sum_{j=1}^{n} \frac{(n+1)^2(n+2)}{(n-j+1)} \left(\frac{e^{e^{-\left(\frac{\sigma}{x_j}\right)^2}} - 1}{e-1} - \frac{j}{n+1}\right)^2 \tag{23}$$

Differentiating (23) with respect to σ and equating it to zero, we get the weighted least squares estimate $\hat{\sigma}_{WLSE}$.

$$\sum_{j=1}^{n} \frac{(n+1)^2(n+2)}{(n-j+1)} \left(\frac{e^{e^{-\left(\frac{\sigma}{x_j}\right)^2}} - 1}{e-1} - \frac{j}{n+1}\right) e^{e^{-\left(\frac{\sigma}{x_j}\right)^2}} e^{-\left(\frac{\sigma}{x_j}\right)^2} \frac{1}{x_j^2} = 0$$

Since the above non-linear equation cannot be solved analytically, using Newton-Raphson method for its numerical solution is recommended.

5. Numerical illustrations

In this section, to study the performance of the estimator $\hat{\sigma}$, we consider four different methods of parameter estimation MLE, MPS, LSE and WLSE. We choose $\sigma = 1$, 1.5, 2, 2.5, 3 and generate random samples of sizes n = 10, 25, 50, 75, 100 from Exponential Transformed Inverse Rayleigh distribution (ETIR) by using inverse transformation method and we repeat this process for K =20000 times. We calculate the estimated value for K times under the considered estimation methods and the average value of the estimator is obtained by using the following formula.

$$\hat{\sigma} = \frac{1}{K}\sum_{i=1}^{K}\hat{\sigma}_i$$

To evaluate the performance of the estimate, we calculate the Mean Square Error (MSE) of that estimator $\hat{\sigma}$ using the formula

$$MSE\left(\hat{\sigma}\right) = \frac{1}{K}\sum_{i=1}^{K}\left(\hat{\sigma}_i - \sigma\right)^2$$

The MSEs of the estimator under the four different methods of parameter estimation is tabulated in Table 1 with an increasing order of sample size. It is observed that for $\sigma = 1$, as we increase the sample size, MSEs of the estimator are gradually decreasing and in case of $\sigma = 1.5$, all the MSEs under four different estimation method are decreasing and almost converge to 0. For $\sigma = 2$, the MSEs of the estimator under Least square method(LSE) is not in a decreasing order with the increment of sample size but for $n = 100$ it converges to the almost same value at where the other MSEs of the estimator converges. For the remaining value of σ, i.e., $\sigma = 2.5$ & 3, the least square method does not perform well as the MSEs are in increasing order as we increase the sample size. Except the LSE, the other methods of estimation MLE, MPS and WLSE performs quite well as the MSEs converge to almost zero. Moreover, from the simulation study it is observed that, based on the MSE criterion, both MLE and MPS method are efficient to estimate the unknown parameter of ETIR distribution and these methods are always recommended for parameter estimation.

To understand it graphically, we represent two graphs for $\sigma = 1$ and 1.5 in Figure (5) displaying the performance of four estimation methods in term of MSE for an increasing order of sample size. It is clearly seen that for $\sigma = 1.5$, with the increment of sample size all the MSEs under the four estimation processes are started to decrease and finally meet to the convergence point zero.

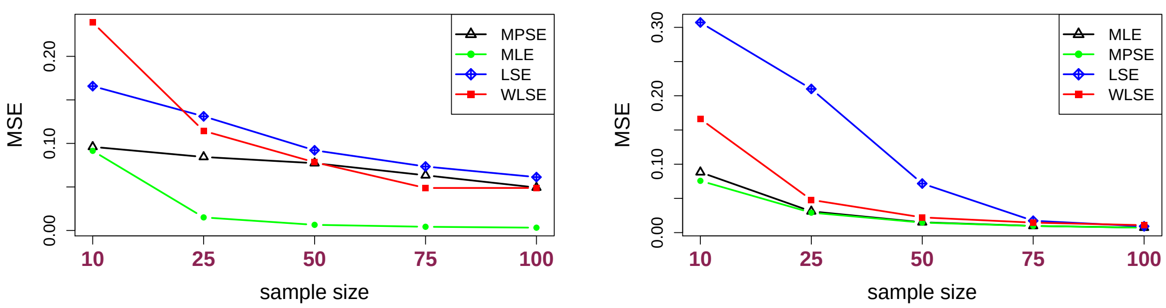
6. Real data applications

In this section, two real datasets are used to access the utility of the proposed ETIR distribution. Moreover, for comparison purpose Exponentiated inverse Rayleigh(EIR), Transmuted inverse Rayleigh(TIR), Inverse Rayleigh(IR) and Rayleigh distribution are also fit with the same datasets. We measure the log-likelihood of the fitted models based on numerically obtained MLE using nlm funcition in R (version 3.6.1) R Core Team (2019) to determine the efficiency of the candidate distributions that best fit the results. Model selection criteria like Akaike Information Criterion(AIC), Bayesian Information Criterion(BIC), Corrected Akaike Information Criterion(AIC_C) and K - S test values are derived. AIC, BIC andAIC_C are defined as follows:

$AIC = 2 * k - 2 * ln\hat{L}$; $\quad BIC = k * ln(n) - 2 * ln\hat{L}$ and $AIC_C = AIC + \frac{2k^2+2k}{n-k-1}$

Table 1: Average estimates of σ with the associated MSE (in parenthesis)

Parameter value	Sample sizes (n)	$\hat{\sigma}_{MLE}$	$\hat{\sigma}_{MPS}$	$\hat{\sigma}_{LSE}$	$\hat{\sigma}_{WLSE}$
$\sigma = 1$	10	1.01372 (0.09604)	0.95087 (0.09149)	0.96432 (0.16577)	0.93089 (0.23889)
	25	0.98316 (0.08444)	0.98260 (0.01504)	0.95313 (0.13110)	0.95988 (0.11441)
	50	0.97301 (0.07731)	0.98686 (0.00651)	0.96240 (0.09210)	0.96757 (0.07839)
	75	0.97585 (0.06339)	0.98940 (0.00427)	0.96884 (0.07349)	0.97977 (0.04882)
	100	0.98070 (0.04932)	0.99192 (0.00324)	0.97601 (0.06129)	0.98280 (0.04887)
$\sigma = 1.5$	10	1.56187 (0.08824)	1.46831 (0.07569)	1.46803 (0.30712)	1.52016 (0.16608)
	25	1.52555 (0.03105)	1.47344 (0.02916)	1.45636 (0.21001)	1.50792 (0.04784)
	50	1.51394 (0.01510)	1.48173 (0.01463)	1.49006 (0.07171)	1.50394 (0.02223)
	75	1.50861 (0.00980)	1.48472 (0.00967)	1.50280 (0.01734)	1.50119 (0.01460)
	100	1.50686 (0.00726)	1.48756 (0.00721)	1.50400 (0.00921)	1.50183 (0.01086)
$\sigma = 2$	10	2.08467 (0.15839)	1.95935 (0.13588)	1.86780 (0.91582)	1.97601 (0.51037)
	25	2.03245 (0.05441)	1.96299 (0.05130)	1.49031 (2.17342)	1.99251 (0.15245)
	50	2.01738 (0.02618)	1.97449 (0.02551)	1.79166 (0.90069)	1.99904 (0.05788)
	75	2.00949 (0.01725)	1.97758 (0.01716)	1.55791 (1.80712)	2.00045 (0.02649)
	100	2.00915 (0.01285)	1.98341 (0.01274)	1.98200 (0.12599)	2.00269 (0.02123)
$\sigma = 2.5$	10	2.60501 (0.24869)	2.44839 (0.21337)	2.44392 (0.89479)	2.39587 (1.19262)
	25	2.54475 (0.08602)	2.45769 (0.08028)	1.68292 (4.31577)	2.38214 (0.79657)
	50	2.52098 (0.04084)	2.46728 (0.03987)	1.95616 (2.82495)	2.42709 (0.44462)
	75	2.51451 (0.02741)	2.47462 (0.02702)	1.51038 (5.01870)	2.44694 (0.31778)
	100	2.51006 (0.02000)	2.47794 (0.01981)	1.55876 (4.96483)	2.46992 (0.17900)
$\sigma = 3$	10	3.12960 (0.35664)	2.94176 (0.30454)	3.02525 (0.72458)	2.79968 (2.21861)
	25	3.05305 (0.12352)	2.94872 (0.11552)	1.97056 (6.49095)	2.84492 (1.20642)
	50	3.02607 (0.05879)	2.96172 (0.05724)	1.78378 (7.45189)	2.94086 (0.47513)
	75	3.01920 (0.03998)	2.97139 (0.03926)	2.22207 (4.77670)	2.98676 (0.16408)
	100	3.01624 (0.02927)	2.97766 (0.02900)	1.72277 (7.64770)	3.00038 (0.07597)



MSE for $\sigma = 1$ MSE for $\sigma = 1.5$

Figure 5: MSE of $\hat{\sigma}$ under the four different estimation methods with the variation of sample size n

where n is sample size, k is the number of parameters and $\hat{L}$ is maximum likelihood for the considered distribution. AIC, BIC, AIC_C and the K-S test statistic with lower values indicate a better fit of distributions.

Dataset 1 The data consists of thirty successive March precipitation (in inches) observations originally was given by Hinkley (1977).
0.77, 1.74, 0.81, 1.2, 1.95, 1.2, 0.47, 1.43, 3.37, 2.2, 3, 3.09, 1.51, 2.1, 0.52, 1.62, 1.31, 0.32, 0.59, 0.81, 2.81, 1.87, 1.18, 1.35, 4.75, 2.48, 0.96, 1.89, 0.9, 2.05

Table 2: Criterion for model comparison

Distributions	Parameter(s) Estimate	-2logL	AIC	BIC	AIC_C	K-S test value
ETIR	$\hat{\sigma} = 0.8293$	84.0526	86.0526	87.4538	86.1955	0.1560
EIR	$\hat{\sigma} = 0.8287$; $\hat{\alpha} = 0.7314$	86.4023	90.4023	93.2047	90.8468	0.1650
IR	$\hat{\sigma} = 0.9267$	88.2730	90.2730	91.6742	90.4159	0.2063
TIR	$\hat{\sigma} = 0.9267$; $\hat{\lambda} = 0.5000$	96.8780	100.8781	103.6805	101.3225	0.3305
Rayleigh	$\hat{\sigma} = 2.0009$	91.2660	93.2660	94.6672	93.4089	0.3463

Dataset 2: In this dataset, survival times (in days) of guinea pigs injected with different doses of tubercle bacilli is given. The dataset is originally reported by Bjebkedal (1960) and futher Kundu and Howlader (2010) and Singh, Singh, and Kumar (2013) used this dataset in Bayesian context.

12, 15, 22, 24, 24, 32, 32, 33, 34, 38, 38, 43, 44, 48, 52, 53, 54, 54, 55, 56, 57, 58, 58, 59, 60, 60, 60, 60, 61, 62, 63, 65, 65, 67, 68, 70, 70, 72, 73, 75, 76, 76, 81, 83, 84, 85, 87, 91, 95, 96, 98, 99, 109, 110, 121, 127, 129, 131, 143, 146, 146, 175, 175, 211, 233, 258, 258, 263, 297, 341, 341, 376.

Table 3: Criterion for model comparison

Distributions	Parameter(s) Estimate	-2logL	AIC	BIC	AIC_C	K-S test value
ETIR	$\hat{\sigma} = 41.8304$	800.1490	802.1490	804.4256	802.2061	0.1920
EIR	$\hat{\sigma} = 39.0607$; $\hat{\alpha}$ =0.6162	801.8299	805.8299	810.3833	806.0038	0.1961
IR	$\hat{\sigma} = 46.7748$	813.4721	815.4721	817.7488	815.5293	0.2369
TIR	$\hat{\sigma} = 46.7748$; $\hat{\lambda} = 0.5000$	837.3260	841.3260	845.8794	841.5000	0.3608
Rayleigh	$\hat{\sigma} = 90.6991$	816.5921	818.5921	820.8688	818.6492	0.2873

7. Conclusion

Exponential Transformed Inverse Rayleigh(ETIR) distribution has been introduced in this study using DUS transformation. Flexibility of this transformed model is greater than that of the original distribution. Probability density plot along with corresponding distribution function, hazard function and survival function plot have been sketched for selected values of parameter. Both the density plot and the hazard plot indicates the distribution is useful in fitting highly skewed positive datasets. Nature of the hazard function relates with the real phenomenon also. Some of the basic statistical properties such as moments, moment generating function, quantile functions, mode, median along with PDF and CDF of r^{th} order statistic are computed for the proposed distribution. It has been seen that the distribution posses only first moment.

The unknown model parameter is estimated using four estimation procedures. To compare these approaches, we conducted a comprehensive simulation analysis. We have compared estimators with respect to mean square error (MSE). The simulation result shows that the maximum likelihood estimation (MLE), maximum product spacing method (MPS) and Weighted least square method (WLSE) are consistent and performs well in terms of MSE. As the sample size increases the MSE value decreases and almost converges to zero for the three estimation method except Least square method(LSE). Furthermore, both MLE and MPS method provide low MSE value and hence these methods are efficient and consistent to estimate the unknown parameter of ETIR distribution.

The model usefulness is illustrated using two real datasets through maximum likelihood approach. ETIR distribution is fitted well among other well known distribution in terms of AIC, BIC, AIC_C and K-S test value. For modeling survival data as well as positively skewed data this distribution have a significant importance. It is hoped that, for further research, this distribution may contribute in the area of reliability and survival analysis.

References

Ali M, Khalil A, Ijaz M, Saeed N (2021). "Alpha-Power Exponentiated Inverse Rayleigh Distribution and Its Applications to Real and Simulated Data." *PLoS ONE*, **16**(1), e0245253. doi:10.1371/journal.pone.0245253.

Aryal GR (2013). "Transmuted Log-logistic Distribution." *Journal of Statistics Applications and Probability*, **2**(1), 11–20. doi:10.12785/jsap/020102.

Aryal GR, Tsokos CP (2009). "On the Transmuted Extreme Value Distribution with Application." *Nonlinear Analysis*, **71**(12), e1401—-e1407. doi:10.1016/j.na.2009.01.168.

Bjebkedal T (1960). "Acquisition of Resistance in Guinea Pigs Infected with Different Doses of Virulent Tubercle Bacilli." *American Journal of Epidemmiology*, **72**(1), 130–148. doi:10.1093/oxfordjournals.aje.a120129.

Cheng RCH, Amin NAK (1979). "Maximum Product of Spacings Estimation with Applications to the Lognormal Distribution." *Technical report, Department of Mathematics, University of Wales.*

Cheng RCH, Amin NAK (1983). "Estimating Parameters in Continuous Univariate Distributions with a Shifted Origin." *Journal of the Royal Statistical Society: Series B*, **45**(3), 394–403. URL https://www.jstor.org/stable/2345411.

Coolen FPA, Newby MJ (1991). "A Note on the Use of the Product of Spacings in Bayesian Inference." *Kwantitatieve Methoden*, **37**, 19 – 32. URL https://www.vvsor.nl/wp-content/uploads/2020/06/KM1991037002.

Cordeiro GM, Ortega EMM, da Cunha DCC (2013). "The Exponentiated Generalized Class of Distributions." *Journal of Data Science*, **11**, 1–27. doi:10.6339/JDS.201301_11(1).0001.

Dey S, Dey T, Kundu D (2014). "Two-Parameter Rayleigh Distribution: Different Methods of Estimation." *American Journal of Mathematical and Management Sciences*, **33**(1), 55–74. doi:10.1080/01966324.2013.878676.

Dey S, Raheem E, Mukherjee S (2017). "Statistical Properties and Different Methods of Estimation of Transmuted Rayleigh Distribution." *Revista Colombiana de Estadística*, **40**(1), 165–203. URL 10.15446/rce.v40n1.56153.

Fan G (2015). "Bayes Estimation for Inverse Rayleigh Model under Different Loss Functions." *Research Journal of Applied Sciences, Engineering and Technology*, **9**(12), 1115–1118. doi:10.19026/rjaset.9.2605.

Ghosh K, Jammalamadaka SR (2001). "A General Estimation Method using Spacings." *Journal of Statistical Planning and Inference*, **93**, 71–82. doi:10.1016/S0378-3758(00)00160-9.

Glaser RE (1980). "Bathtub and Related Failure Rate Characterizations." *Journal of the American Statistical Association*, **75**(371), 667–672. doi:10.1080/01621459.1980.10477530.

Gupta RC, Gupta PL, Gupta RD (1998). "Modeling Failure Time Data by Lehman Alternatives." *Communications in Statistics - Theory and Methods*, **27**(4), 887–904. doi:10.1080/03610929808832134.

Gupta RD, Kundu D (2001). "Exponentiated Exponential Family: An Alternative to Gamma and Weibull Distributions." *Biometrical Journal*, **43**(1), 117–130. URL https://doi.org/10.1002/1521-4036(200102)43:1<117::AID-BIMJ117>3.0.CO;2-R.

Hinkley D (1977). "On Quick Choice of Power Transformation." *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **26**(1), 67–69. doi:https://doi.org/10.2307/2346869.

Khan MS, King R (2015). "Transmuted Modified Inverse Rayleigh Distribution." *Austrian Journal of Statistics*, **44**, 17–29. doi:10.17713/ajs.v44i3.21.

Kotz S, Nadarajah S (2000). *Extreme Value Distributions: Theory and Applications*. Imperial College Press, London. URL https://doi.org/10.1142/p191.

Kumar D, Singh U, Singh SK (2015a). "A Method of Proposing New Distribution and Its Application to Bladder Cancer Patients Data." *Journal of Statistics Applications and Probability Letters*, **2**(3), 235–245. URL http://dx.doi.org/10.12785/jsapl/020306.

Kumar D, Singh U, Singh SK (2015b). "A New Distribution Using Sine Function- Its Application to Bladder Cancer Patients Data." *Journal of Statistics Applications and Probability*, **4**(3), 417–427. URL http://dx.doi.org/10.12785/jsap/040309.

Kumar D, Singh U, Singh SK (2016). "Life Time Distributions: Derived from Some Minimum Guarantee Distribution." *Sohag Journal of Mathematics*, **4**(1), 7–11. URL http://dx.doi.org/10.18576/sjm/040102.

Kumar D, Singh U, Singh SK, Mukherjee S (2017). "The New Probability Distribution: An Aspect to a Life Time Distribution." *Mathematical Science Letters*, **6**(1), 35–42. doi:http://dx.doi.org/10.18576/msl/060106.

Kumaraswamy P (1980). "A Generalized Probability Density Function for Double Bounded Random Process." *Journal of Hydrology*, **46**, 79–88. URL https://doi.org/10.1016/0022-1694(80)90036-0.

Kundu D, Howlader H (2010). "Bayesian Inference and Prediction of the Inverse Weibull Distribution for Type-II Censored Data." *Computational Statistics and Data Analysis*, **54**(6), 1547–1558. URL https://doi.org/10.1016/j.csda.2010.01.003.

Kundu D, Raqab MZ (2005). "Generalized Rayleigh Distribution: Different Methods of Estimation." *Computational Statistics and Data Analysis*, **49**, 187–200. URL https://doi.org/10.1016/j.csda.2004.05.008.

Louzada F, Ramos PL, Perdoná GSC (2016). "Different Estimation Procedures for the Parameters of the Extended Exponential Geometric Distribution for Medical Data." *Computational and Mathematical Methods in Medicine*, **2016**. URL https://doi.org/10.1155/2016/8727951.

Mazucheli J, Ghitany ME, Louzada F (2016). "Comparisons of Ten Estimation Methods for the Parameters of Marshall-Olkin Extended Exponential Distribution." *Communications in Statistics - Simulation and Computation*, **46**(7), 5627–5645. URL https://doi.org/10.1080/03610918.2016.1171345.

Merovci F (2013). "Transmuted Rayleigh Distribution." *Austrian Journal Of Statistics*, **42**(1), 21–31. URL https://doi.org/10.17713/ajs.v42i1.163.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ranneby B (1984). "The Maximum Spacing Method. An Estimation Method Related to the Maximum Likelihood Method." *Scandinavian Journal of Statistics*, **11**(2), 93–112. URL https://www.jstor.org/stable/4615946.

Rao GS, Mbwambo S (2019). "Exponentiated Inverse Rayleigh Distribution and an Application to Coating Weights of Iron Sheets Data." *Journal of Probability and Statistics, Hindawi*, **2019**, 1–13. URL https://doi.org/10.1155/2019/7519429.

Seenoi P, Supapakorn T, Bodhisuwan W (2014). "The Length-Biased Exponentiated Inverted Weibull Distribution." *International Journal of Pure and Applied Mathematics*, **92**(2), 191–206. URL http://dx.doi.org/10.12732/ijpam.v92i2.5.

Shah A, Gokhale DV (1993). "On Maximum Product of Spagings (MPS) Estimation for Burr XII Distributions." *Communications in Statistics - Simulation and Computation*, **22**(3), 615–641. URL https://doi.org/10.1080/03610919308813112.

Shao Y, Hahn MG (1999). "Strong Consistency of the Maximum Product of Spacings Estimates with Applications in Nonparametrics and in Estimation of Unimodal Densities." *Annals of the Institute of Statistical Mathematics*, **51**(1), 31–49. doi:10.1023/A:1003827017345.

Shaw WT, Buckley IRC (2009). "The Alchemy of Probability Distributions: Beyond Gram-Charlier Expansions, and a Skew-Kurtotic-Normal Distribution from a Rank Transmutation Map." *arXiv preprint arXiv:0901.0434*. URL https://arxiv.org/abs/0901.0434.

Singh SK, Singh U, Kumar D (2013). "Bayesian Estimation of Parameters of Inverse Weibull Distribution." *Journal of Applied Statistics*, **40**(7), 1597–1607. URL https://doi.org/10.1080/02664763.2013.789492.

Singh U, Singh SK, Singh RK (2014). "Product Spacings as an Alternative to Likelihood for Bayesian Inferences." *Journal of Statistics Applications and Probability*, **3**(2), 179–188. doi:10.12785/jsap/030208.

Swain JJ, Venkatraman S, Wilson JR (1988). "Least-Squares Estimation of Distribution Functions in Johnson's Translation System." *Journal of Statistical Computation and Simulation*, **29**(4), 271–297. URL https://doi.org/10.1080/00949658808811068.

Wong TST, Li WK (2006). "A Note on the Estimation of Extreme Value Distributions using Maximum Product of Spacings." *IMS Lecture Notes-Monograph Series*, **52**, 272–283. URL https://doi.org/10.1214/074921706000001102.

Affiliation:

Proloy Banerjee
Department of Mathematics and Statistics
Aliah University
Action Area II, Plot No.- IIA/27,
Newtown, Kolkata, West Bengal, India-700160
E-mail: proloy.stat@gmail.com

Shreya Bhunia
Department of Mathematics and Statistics
Aliah University
Action Area II, Plot No.- IIA/27,
Newtown, Kolkata, West Bengal, India-700160
E-mail: shreyabhunia.stat@gmail.com

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 51 *Submitted:* 2021-04-10
July 2022 *Accepted:* 2021-10-11