REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000-000
https://doi.org/00.00000/revstat.v00i0.000

Stochastic Generator Of A New Family Of Lifetime Distributions With Illustration

Authors: Amjad D. Al-Nasser
– Department of Statistics, Faculty of Science, Yarmouk University, Jordan
amjadyu@yahoo.com

Ahmad A. Hanandeh ✉
– Department of Statistics, Faculty of Science, Yarmouk University, Jordan
ahmad.hanandeh@yu.edu.jo

Received: Month 0000 Revised: Month 0000 Accepted: Month 0000

Abstract:

- In this article, a new family of lifetime distributions, referred to as exponential transformation (ET) family, is introduced to help researchers model different types of data sets. Furthermore, the maximum likelihood method was used in estimating the model's parameters. Some structural properties of the distribution have been derived and studied. These are density function, distribution function, and reliability function, hazard rate function, moments, moment generating function, entropies, order statistics, Bonferroni and Lorenz curves. Simulation method was used to investigate the behaviors of the parameters of the proposed distribution; the results showed that the mean square error and standard error for the chosen parameter values decrease as the sample size increases. The proposed distribution was tested on real-life data, the results showed that the ET-exponential distribution performed better than other well-known distributions in modeling data. The results also showed that the distribution can be used as an alternative model in modeling lifetime processes.

Keywords:

- *Lifetime distribution; maximum Likelihood; means square error; parameters; quantile function*

AMS Subject Classification:

- 60E05, 62F10.

✉ Corresponding author

1. INTRODUCTION

With increasing diversity of real-life problems and applications that includes complicated phenomena, there is a growing interest by researchers in developing new lifetime distributions to overcome complicated models. Consequently, significant progress has been made towards constructing numerous classes of new distributions to generate more flexible distributions instead of the classical ones to provide more accurate data modeling. [15] was the first who suggested proposing new distribution by taking baseline distribution, and then [5] introduced a system for generating new distributions by adding a skewing parameter to a symmetric distribution.

The ideas of generating a new class of distributions graduated and can be classified into five schemes; the first one is by using differential equations, the second one is by generating a weighted form of the baseline distribution. The third one concerns adding additional parameter(s) to the baseline distribution. The fourth scheme is to discretize the continuous density function. The last scheme is a distribution transformation scheme that modifies a probability distribution function by forming a stochastic representation of baseline distribution such that the new relationship must satisfy the distribution theory assumptions.

In this article, we are interested in the last scheme to regenerate a new class of lifetime distribution. In the literature, there are several generators proposed based on different mathematical functional relationships. For instance, [19] defined the exponentiated class of distributions by exponentiating a given baseline distribution with a positive parameter. [17] , applied the transformation scheme to the survival function by adding an additional shape parameter to the transformation scheme. [10] , used beta as a generator to develop Beta-G class of distributions. [25] suggested the gamma-G class of distributions. Transmuted family of distributions was developed by [20]; then later [21] proposed the quadratic rank transmutation map, while [7] proposed the Kumaraswamy-generated family.

Recently, [3] defined and studied a new *T-X* family. Logarithmic transformation was proposed by [18] and an extension of this generator was proposed by [4]. Moreover, many used trigonometry functions to provide distribution generators, for instance, [13] used Sine function to develop a new class of distributions while modification of this scheme by using Cosine-Sine (CS) transformation proposed by [6]. Vast modification has been made by many authors to identify a new generator of family of distributions ([14]; [11]; [23]; [2]; [24]).

In similar fashion, this article will propose a new distribution generator based on the exponential function to provide new class of parameter lifetime distribution. Let Y be a non-negative continuous random variable with baseline cumulative distribution function (CDF) $F(y,\theta)$ and probability distribution function (PDF) $f(y,\theta)$; where $\theta \in (0,\infty)$ is real valued represents the distribution parameter, then the stochastic presentation of the proposed CDF for generating a new class of distributions can be defined as:

$$G\left(y,\theta,\alpha\right) = F\left(y,\theta\right) e^{-\alpha\ \overline{F}(y,\theta)} \quad y>0;\ \alpha \geq 0 \tag{1.1}$$

where $\overline{F}(y,\theta) = 1 - F(y,\theta)$. Noting that when $\alpha = 0$, then the proposed distribution is exactly the same as the baseline distribution.

This family will be called as **exponential transformation (ET)** *i.e.*, $ET(y,\theta,\alpha)$. Now, the PDF of ET family can be obtaining by finding the first derivative of Equation 1.1:

$$g\left(y,\theta,\alpha\right) = f\left(y,\theta\right) e^{-\alpha\ \overline{F}(y,\theta)}\ \left(1+\ \alpha\ F\left(y,\theta\right)\right)\ ;\quad y>0;\alpha \geq 0 \tag{1.2}$$

This family can be joined to T-X family by ([3]) as follows: For a general baseline CDF of a continuous probability distribution denoted by $F(y,\theta)$, a new CDF having the form

$$G\left(y,\theta,\alpha\right) = \int_0^{F(y,\theta)} r(t)dt$$

Where $r(t)$ is a PDF defined over $(0,1)$, and $R(t)$ is the associated CDF. Accordingly, the PDF $g(y,\theta,\alpha)$ can be obtained as

$$g\left(y,\theta,\alpha\right) = f\left(y,\theta\right)\ r(F\left(y,\theta\right))$$

If we used $r(t)$ in the following functional form:

$$r\left(t\right) = \left(1+\alpha t\right)\ e^{-\alpha(1-t)}$$

Then, the proposed family could be considered as member of the T-X family. As an illustration of the proposed family, the exponential distribution will be considered as baseline distribution. In this article, the third scheme will be used to generate new class of lifetime distribution.

The remainder of this article proceeds as follows. Section 2 provides some characterizations of the ET family, including shapes of CDF, PDF, reliability measures such as survival and hazard rate. Section 3 is dedicated to the mathematical properties of the ET family such as moments, quantiles, order statistics and entropies. In Section 4, the estimation of parameters is studied. Section 5 offers detailed simulation experiments on model performance and assessment. Section 6 is devoted to studying illustrative examples based on real data. Finally, Section 7 concludes the manuscript with a summary and an eye toward future work to close the paper.

2. CHARACTERIZATIONS

2.1. Asymptotic properties of the CDF and PDF

Suppose that X is a continuous random variable of ET family as given in Equation 1.1, it can be easily seen that this family of distribution satisfies the Kolmogorov axioms of the distribution functions. For instance, it is easily seen that the limit property of $G\,(x,\theta,\ \alpha)$ satisfy the property of CDF:

$$\lim_{x\to\infty} G\,(x,\theta,\alpha) = \lim_{x\to\infty} F\,(x,\theta)\,e^{-\alpha\ \overline{F}(x,\theta)} = 1$$

and

$$\lim_{x\to 0} G\,(x,\theta,\alpha) = \lim_{x\to 0} F\,(x,\theta)\,e^{-\alpha\ \overline{F}(x,\theta)} = 0$$

Hence, the total probability is equal to one. Also, it is monotone right increasing function of x, and $0 \le G\,(x,\theta,\alpha) \le 1;\ \forall\ x$. Therefore, $G\ (x,\theta,\alpha)$ is an absolute continuous distribution function.

Similarly, it is easily can be noted that $g\,(x,\theta,\alpha)$ is non-negative real valued PDF for all x. For instance in the exponential case:

$$\lim_{x\to\infty} g\,(x,\theta,\alpha) = \ 0$$

and

$$\lim_{x\to 0} g\,(x,\theta,\alpha) = \ \theta\ e^{-\alpha}$$

Since both parameters are positive this indicates that $g\,(x,\theta,\alpha)$ is a unimodal distribution. Now, the functional form given in 1.2 satisfied the PDF property:

$$\int_0^\infty g\,(x,\theta,\alpha)\,dx = \int_0^\infty f\,(x,\theta)\,e^{-\alpha\ \overline{F}(x,\theta)}\ \ (1+\alpha\ F\,(x,\theta))\ \ dx$$

To illustrate the usefulness of the new stochastic representation given in Equation 1.1 and the associated PDF given in Equation 1.2, suppose that the baseline distribution is the exponential distribution with mean $\frac{1}{\theta}$, then we have ET-Exp distribution.

Corollary 2.1. *Suppose that X is a random variable of ET-Exp, then the CDF and PDF of X are given in Equations 2.1 and 2.2, respectively:*

$$G\,(x,\theta,\alpha) = e^{-\alpha e^{-\theta x}}\left(1-e^{-\theta x}\right);\quad x>0;\theta>0,\alpha\ge 0 \tag{2.1}$$

$$g\,(x,\theta,\alpha) = \theta e^{-\left(\theta x+\alpha e^{-\theta x}\right)}\left(1+\alpha(1-e^{-\theta x})\right);\quad x>0;\theta>0,\alpha\ge 0 \tag{2.2}$$

Figure 1 give a good representation of the new distribution PDF with selected set of parameters, in different cases by assuming both parameters are larger than 1, less than 1 or one of them is less than one and the other parameter is more than one.

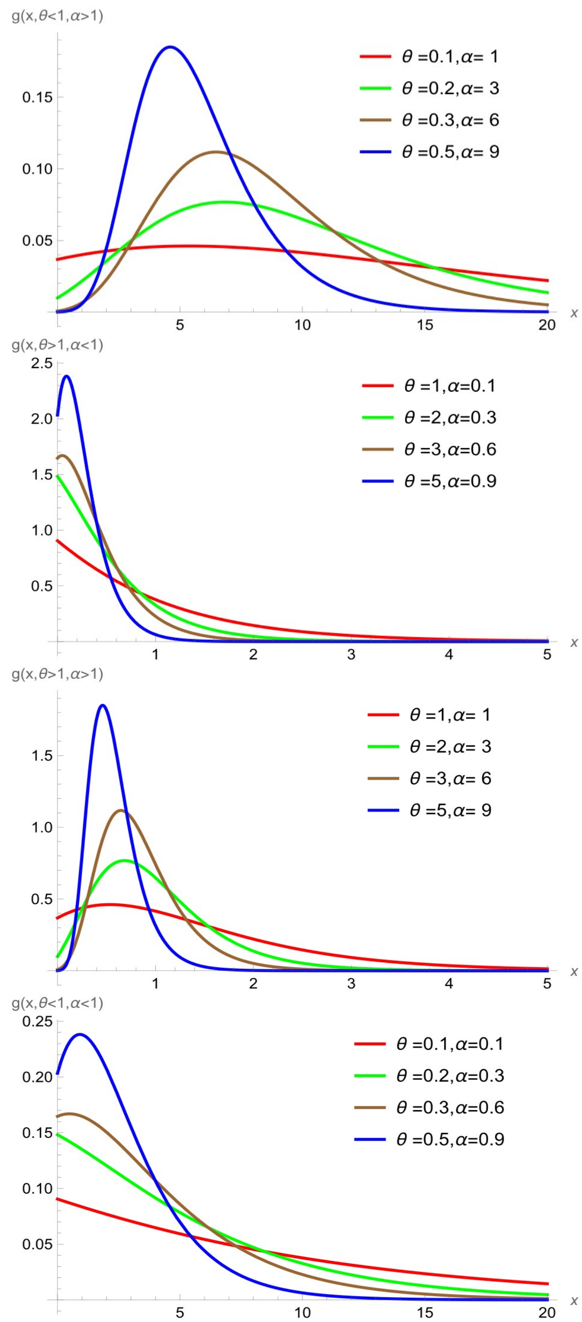


Figure 1: PDF of ET-Exp with selected parameter values

2.2. The CDF and PDF expansion

The following result proposes Taylor series expansions of the exponential function that given in ET family. Accordingly, using the exponential series, we get:

$$e^{-x} = \sum_{j=0}^{\infty} (-1)^j \frac{x^j}{j!}$$

Then the CDF can be written as, respectively:

$$G(x,\theta,\alpha) = F(x,\theta) \sum_{j=0}^{\infty} (-1)^j \frac{\alpha^j}{j!} \overline{F}(x,\theta)^j$$

$$G(x,\theta,\alpha) = F(x,\theta) \sum_{j=0}^{\infty} (-1)^j \frac{\alpha^j}{j!} \{1 - F(x,\theta)\}^j$$

Now, using the power series $(1-z)^m = \sum_{k=1}^{\infty} (-1)^k \binom{m}{k} z^k$, the expansion yields to:

$$G(x,\theta,\alpha) = F(x,\theta) \sum_{j=0}^{\infty} (-1)^j \frac{\alpha^j}{j!} \sum_{k=1}^{\infty} (-1)^k \binom{j}{k} F(x,\theta)^k$$

Which can be simplified to

$$G(x,\theta,\alpha) = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} (-1)^{j+k} \frac{\alpha^j}{j!} \binom{j}{k} F(x,\theta)^{k+1} \tag{2.3}$$

In similar fashion, the PDF can be expanded as follows:

$$g(x,\theta,\alpha) = f(x,\theta)\,(1 + \alpha\, F(x,\theta)) \left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} (-1)^{j+k} \frac{\alpha^j}{j!} \binom{j}{k} F(x,\theta)^k \right) \tag{2.4}$$

Based on Equations 2.3 and 2.4, several mathematical properties of the ET family can be derived for any lifetime distribution.

2.3. Reliability measures of ET family of distributions

Reliability measures are widely used to analyze lifetime models. The most well-known measures are the survival, hazard or faultier rate and cumulative hazard functions; the following theorem presents these measures of the proposed family of distributions.

Theorem 2.1. *Let X be a random variable that follows the ET family of distributions, with PDF and CDF as defined in Equations 1.1 and 1.2, then:*

1. *The survival function is given by*

$$S(x,\theta,\alpha) = 1 - G(x,\theta,\alpha); \qquad x > 0$$
$$= 1 - F(x,\theta)\, e^{-\alpha\, S_F(x,\,\theta)}$$

where $S_F(x,\theta,\alpha)$ is the survival function of the baseline distribution. It is obvious that $\lim_{x\to\infty} S(x,\theta,\alpha) = 0$ and $\lim_{x\to 0} S(x,\theta,\alpha) = 1$.

2. *The hazard function is given by*

$$h(x,\theta,\alpha) = \frac{g(x,\theta,\alpha)}{S(x,\theta,\alpha)} = \frac{f(x,\theta)\, e^{-\alpha S_F(x,\,\theta,\alpha)}\,(1+\alpha\, F(x,\theta))}{1 - F(x,\theta)\, e^{-\alpha S_F(x,\theta,\alpha)}} \tag{2.5}$$

3. *The reversed hazard function is given by*

$$hr(x,\theta,\alpha) = \frac{g(x,\theta,\alpha)}{G(x,\theta,\alpha)} = \frac{f(x,\theta)\, e^{-\alpha S_F(x,\theta,\alpha)}\,(1+\alpha F(x,\theta))}{F(x,\theta)\, e^{-\alpha S_F(x,\theta,\alpha)}} \tag{2.6}$$

Assuming the baseline distribution is the exponential distribution, then Corollary 2.2 is hold.

Corollary 2.2. *Suppose that X is a random variable of ET-Exp, then*

1. *The survival function is given by:*

$$S\left(x,\theta,\alpha\right)=1-\ e^{-\alpha e^{-\theta x}}\left(1-e^{-\theta x}\right)$$

2. *The hazard rate function is given by:*

$$h\left(x,\theta,\alpha\right)=\frac{e^{-\left(\theta x+\alpha e^{-\theta x}\right)}\left(1+\alpha(1-e^{-\theta x})\right)}{1-\ e^{-\alpha e^{-\theta x}}\left(1-e^{-\theta x}\right)}\ \theta$$

3. *The reversed hazard function is given by:*

$$hr\left(x,\theta,\alpha\right)=\frac{e^{-\theta x}\left(1+\alpha(1-e^{-\theta x})\right)}{\left(1-e^{-\theta x}\right)}\ \theta$$

Figure 2 shows comparisons between the hazard rate functions of the baseline distribution and the proposed distribution.

3. MATHEMATICAL PROPERTIES

Some basic mathematical properties such as ordinary moments, quantile function and moment generating function are derived in this section.

3.1. Moments

Some of the most important characteristics (tendency, dispersion, skewness and kurtosis) of a statistical distribution can be studied through moments. Suppose that the moments of $ET(x,\theta,\ \alpha)$ can be obtain by finding the expected value of $k(x)$; where

$$k(x)\ =\begin{cases} x^r, & \text{for moment of order } r \\ e^{tx}, & \text{for moment of generating function} \\ e^{itx}, & \text{for characteristic function} \end{cases}$$

Hence,

$$E\left(k(x)\right)=\int_0^\infty k(x)\ f\left(x,\theta\right)e^{-\alpha\overline{F}(x,\theta)}\ \left(1+\alpha\ F\left(x,\theta\right)\right)dx$$

$$=\int_0^\infty k(x)\ f\left(x,\theta\right)\left(1+\alpha\ F\left(x,\theta\right)\right)\left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}(-1)^{j+k}\frac{\alpha^j}{j!}\ \binom{j}{k}\ F(x,\theta)^k\right)\ dx$$

which is equivalent to the expected value based on the baseline distribution

$$E_F\left(k(x)\ e^{-\alpha\overline{F}(x,\theta)}\ \left(1+\alpha F\left(x,\theta\right)\right)\right)$$

Then the expected value can be obtained using expansion technique or by using integral estimation.

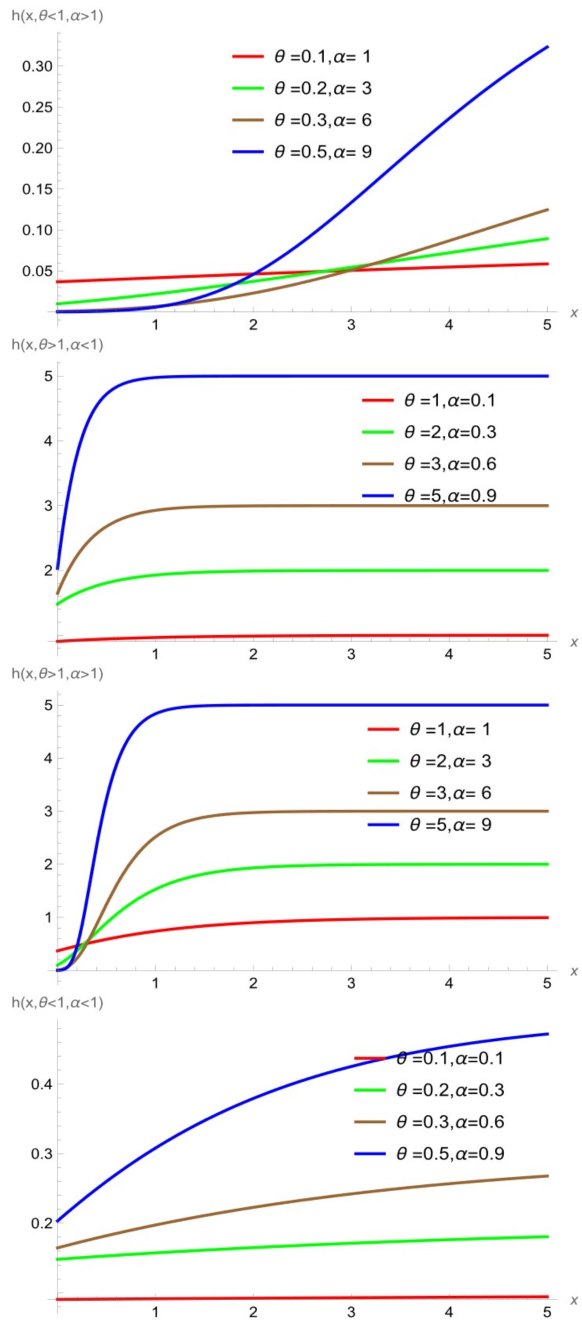


Figure 2: Hazard rate of ET-Exp with selected parameter values

Corollary 3.1. *Suppose that X is a random variable of ET-Exp, then the rth moment is given by:*

$$E\left(x^{r}\right)=\int_{0}^{\infty} x^{r}\ \theta e^{-\left(\theta x+\alpha e^{-\theta x}\right)}\left(1+\alpha\left(1-e^{-\theta x}\right)\right)\ dx$$

$$=\int_{0}^{\infty} x^{r}\ e^{-\theta x}\left(1+\alpha\ \left(1-e^{-\theta x}\right)\right)\left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}(-1)^{j+k}\frac{\alpha^{j}}{j!}\ \binom{j}{k}\ \left(1-e^{-\theta x}\right)^{k}\right)\ dx$$

Using the first fourth moments one can compute numerically the population mean, variance, standard deviation, skewness and kurtosis coefficients for some give parameter's values.

Corollary 3.2. *Suppose that X is a random variable of ET-Exp, then the moment generating function and characteristic function are, respectively, given by:*

$$E\left(e^{tx}\right) = \int_0^\infty e^{tx}\ \theta e^{-\left(\theta x + \alpha e^{-\theta x}\right)}\left(1 + \alpha\left(1 - e^{-\theta x}\right)\right)\ dx$$

$$= \int_0^\infty e^{(t-\theta)x}\left(1 + \alpha\ \left(1 - e^{-\theta x}\right)\right)\left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}(-1)^{j+k}\frac{\alpha^j}{j!}\ \binom{j}{k}\ \left(1 - e^{-\theta x}\right)^k\right)\ dx$$

Similarly,

$$E\left(e^{itx}\right) = \int_0^\infty e^{itx}\ \theta e^{-\left(\theta x + \alpha e^{-\theta x}\right)}\left(1 + \alpha\left(1 - e^{-\theta x}\right)\right)\ dx$$

$$= \int_0^\infty e^{(it-\theta)x}\left(1 + \alpha\ \left(1 - e^{-\theta x}\right)\right)\left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}(-1)^{j+k}\frac{\alpha^j}{j!}\ \binom{j}{k}\ \left(1 - e^{-\theta x}\right)^k\right)\ dx$$

3.2. Quantile function

The quantile function of X, say $F^{-1}(y)$, is given by the inverse function of $F(x)$. Let X follow $ET(x, \theta,\ \alpha)$ family, the quantile function of X is given by:

$$X = Q\left(u\right) = Q_F(u\ e^{-\ \alpha(1-u)}\ ; \theta)$$

where Q_F is the quantile function of the baseline distribution. Therefore, if U follow the standard uniform distribution, then $X = Q(u)$ follows the $ET(x, \theta,\ \alpha)$ family.

Now assuming that our baseline function is exponential, then, after some algebra, it follows that the Quantile function for ET-Exp distribution can be written as:

$$X = -\frac{\log(-\frac{W(u\alpha\ e^{\alpha}) - \alpha}{\alpha})}{\theta}$$

where $W(.)$ is the Lambert W function.

3.3. Order statistics

Let $X_{(1)}, X_{(2)},\ \ldots\ , X_{(n)}$, be the order statistics of a random sample $X_1, X_2\ ,\ \ldots\ , X_n$ from the distribution with PDF $g\left(x\right)$ and CDF $G\left(x\right)$. Then, the PDF of the i^{th} order statistics $X_{(i)}$ is given by:

$$g_{X_{(i)}}\left(x\right) = \frac{n!}{(i-1)!\,(n-i)!}\ g\left(x\right)\left[\ G\left(x\right)\right]^{(i-1)} \times \left[1 - G\left(x\right)\right]^{(n-i)} \tag{3.1}$$

By substituting equations 1.1 and 1.2 into equation 3.1, it follows that the PDF of the i^{th} order statistics $X_{(i)}$ of the ET-family is given by:

$$\begin{aligned} g_{X_{(i)}}\left(x\right) =& \frac{n!}{(i-1)!\,(n-i)!}\ f\left(x,\theta\right)e^{-\alpha S_F(x,\ \theta)}\ \left(1 + \alpha F\left(x,\theta\right)\right)\left[F\left(x,\theta\right)e^{-\alpha S_F(x,\ \theta)}\right]^{(i-1)} \\ &\times\left[1 - F\left(x,\theta\right)e^{-\alpha S_F(x,\ \theta)}\right]^{(n-i)} \end{aligned} \tag{3.2}$$

Assuming the baseline distribution is the exponential distribution, then the equation 3.2 will be:

$$\begin{aligned} g_{X_{(i)}}\left(x\right) =& \frac{n!}{(i-1)!\,(n-i)!}\ \theta e^{-\left(\theta x + \alpha e^{-\theta x}\right)}\left(1 + \alpha(1 - e^{-\theta x})\right)\ \left[e^{-\alpha e^{-\theta x}}\left(1 - e^{-\theta x}\right)\right]^{(i-1)} \\ &\times\left[1 - e^{-\alpha e^{-\theta x}}\left(1 - e^{-\theta x}\right)\right]^{(n-i)} \end{aligned} \tag{3.3}$$

3.4. Entropy measure

Entropy of a variable is a measure of variation of the uncertainty and it is widely used in science, e.g., physics and engineering. Here, we focus our attention on two types of entropy, namely Rényi and Tsallis

3.4.1. Rényi Entropy

The Rényi entropy of a random variable X with distribution $g(x)$ of order δ, where $\delta > 0$ and $\delta \neq 1$, can be obtained as follows

$$R(\delta) = \frac{1}{1-\delta} log \left(\int g^{\delta}(x) dx \right) \tag{3.4}$$

By substituting equations 1.2 into equation 3.4 leads to

$$R(\delta) = \frac{1}{1-\delta} log \left(\int \left(f(x,\theta) \, e^{-\alpha \, \overline{F}(x,\theta)} \left(1 + \alpha \, F(x,\theta)\right) \right)^{\delta} dx \right)$$

Moreover, The Rényi entropy for ET-Exp distribution is:

$$R(\delta) = \frac{1}{1-\delta} log \left(\int_0^{\infty} \left(\theta e^{-\left(\theta x + \alpha e^{-\theta x}\right)} \left(1 + \alpha(1 - e^{-\theta x})\right) \right)^{\delta} dx \right)$$

$$= \frac{1}{1-\delta} log \left(\int_0^{\infty} \left(e^{-\theta x} \left(1 + \alpha \left(1 - e^{-\theta x}\right)\right) \left(\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (-1)^{j+k} \frac{\alpha^j}{j!} \binom{j}{k} \left(1 - e^{-\theta x}\right)^k \right) \right)^{\delta} dx \right)$$

3.4.2. Tsallis entropy

The Tsallis entropy of a random variable X with distribution $g(x)$ of order λ, where $\lambda > 0$ and $\lambda \neq 1$, can be obtained as follows

$$S(\lambda) = \frac{1}{1-\lambda} \left(1 - \int g^{\lambda}(x) dx \right) \tag{3.5}$$

By substituting equations 1.2 into equation 3.5 leads to

$$S(\lambda) = \frac{1}{1-\lambda} \left(1 - \int \left(f(x,\theta) \, e^{-\alpha \, \overline{F}(x,\theta)} \left(1 + \alpha \, F(x,\theta)\right) \right)^{\lambda} dx \right)$$

Moreover, The Tsallis entropy for ET-Exp distribution is:

$$S(\lambda) = \frac{1}{1-\lambda} \left(1 - \int_0^{\infty} \left(\theta e^{-\left(\theta x + \alpha e^{-\theta x}\right)} \left(1 + \alpha(1 - e^{-\theta x})\right) \right)^{\lambda} dx \right)$$

$$S(\lambda) = \frac{1}{1-\lambda} \left(1 - \int_0^{\infty} \left(e^{-\theta x} \left(1 + \alpha \left(1 - e^{-\theta x}\right)\right) \left(\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (-1)^{j+k} \frac{\alpha^j}{j!} \binom{j}{k} \left(1 - e^{-\theta x}\right)^k \right) \right)^{\lambda} dx \right)$$

4. PARAMETER ESTIMATION

In this section, estimation of the unknown parameters of the $ET(x,\theta,\alpha)$ family of distributions based on complete samples are determined using method of moment (MOM) and maximum likelihood estimation (MLE) method. Let $x_1, \; x_2, \ldots, \; x_n$ be the observed values from $ET(x,\theta,\alpha)$ family.

4.1. Method of moment

The MOM estimator can be obtained by solving the following equations

$$E_F\left(x\ e^{-\alpha \overline{F}(x,\theta)}\ \left(1+\ \alpha\ F\left(x,\theta\right)\right)\right) = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$E_F\left(x^2\ e^{-\alpha \overline{F}(x,\theta)}\ \left(1+\ \alpha\ F\left(x,\theta\right)\right)\right) = \frac{\sum_{i=1}^{n} {x^2}_i}{n}$$

Using Mathematica, we may replace the first moment of ET-Exp family by:

$$E\left(X\right)=\frac{1}{\alpha\theta}\left(1-e^{-\alpha}+\alpha\left(\int_0^{\alpha}\frac{\sinh\left(t\right)}{t}dt-\int_0^{\alpha}\frac{\cosh\left(t\right)-1}{t}dt\right)\right)$$

$$=\frac{1}{\alpha\theta}\left(1-e^{-\alpha}+\alpha\left(\log(\alpha)-\text{Chi}(\alpha)+\text{Shi}(\alpha)+\gamma\right)\right)$$

While the second moment can be replaced by the following formula:

$$E\left(X^2\right)=\frac{2\left(\alpha^2{}_3F_3\left(1,1,1;2,2,2;-\alpha\right)+\log\left(\alpha\right)+\Gamma\left(0,\alpha\right)+\gamma\right)}{\alpha\theta^2}$$

Where γ is Euler's constant, with numerical value ≈ 0.577216, the incomplete gamma function satisfies

$$\Gamma\left(0,\alpha\right)=\int_{\alpha}^{\infty}\frac{e^{-t}}{t}dt$$

and ${}_3F_3(1,1,1;2,2,2;-\alpha)$ is the generalized hypergeometric function.

4.2. Maximum likelihood estimation method

Using the MLE, the point estimator of the unknown parameter can be obtained by solving the following likelihood function:

$$L=\prod_{i=1}^{n} f\left(x_i,\theta\right)\ e^{-\alpha\overline{F}(x,\theta)}\ \left(1+\alpha\ F\left(x_i,\theta\right)\right)$$

Taking the Log of the likelihood function will simplify the estimation problem:

$$LogL = \left\{\sum_{i=1}^{n} Log\left(f\left(x_i,\theta\right)\right)-\sum_{i=1}^{n}\alpha\overline{F}(x_i,\theta)+\sum_{i=1}^{n}Log(1+\alpha\ F\left(x_i,\theta\right))\right\}$$

Now, we have to find the first order condition:

$$\frac{dLogL}{d\theta} = \sum_{i=1}^{n}\frac{df\left(x_i,\theta\right)/d\theta}{f\left(x_i,\theta\right)}+\sum_{i=1}^{n}\alpha\ f\left(x_i,\theta\right)+\sum_{i=1}^{n}\frac{\alpha\ f\left(x_i,\theta\right)}{1+\alpha\ F\left(x_i,\theta\right)}$$

$$\frac{dLogL}{d\alpha} = -\sum_{i=1}^{n}\overline{F}(x_i,\theta)+\sum_{i=1}^{n}\frac{F\left(x_i,\theta\right)}{1+\alpha\ F\left(x_i,\theta\right)}$$

Then setting each of the first order conditions to zero and using a numerical method we can find the optimal estimator of the unknown parameters.

Similarly, taking the exponential case, the MLE the point estimator of the unknown parameter can be obtained by solving the following likelihood function:

$$L = \prod_{i=1}^{n} \theta e^{-\left(\theta x_i + \alpha e^{-\theta x_i}\right)} \left(1 + \alpha(1 - e^{-\theta x_i})\right)$$

Taking the Log of the likelihood function will simplify the estimation problem:

$$LogL = \left\{ \sum_{i=1}^{n} Log\left(\theta e^{-\theta x_i}\right) - \sum_{i=1}^{n} \alpha e^{-\theta x_i} + \sum_{i=1}^{n} Log(1 + \alpha(1 - e^{-\theta x_i})) \right\}$$

Now, we have to find the first order condition:

$$\frac{dLogL}{d\theta} = \sum_{i=1}^{n}\left(\frac{1}{\theta} - x_i\right) + \sum_{i=1}^{n} \alpha\ \theta e^{-\theta x_i} + \sum_{i=1}^{n} \frac{\alpha\ \theta e^{-\theta x_i}}{1 + \alpha\ (1 - e^{-\theta x_i})}$$

$$\frac{dLogL}{d\alpha} = -\sum_{i=1}^{n} e^{-\theta x} + \sum_{i=1}^{n} \frac{(1 - e^{-\theta x_i})}{1 + \alpha(1 - e^{-\theta x_i})}$$

The non-linear equations above can not be solved analytically, and thus we have used an R-code to solve it analytically on R-software [22].

5. SIMULATION

In this section, we study the performance of ML estimators for different sample sizes, i.e, n= 50, 75, 100, 250, and 400. We have employed the inverse CDF technique for data simulation for ET-Exp distribution using R software. Bias, Variance and MSE for the ET-Exp distribution are observed. As it was expected, Table 1 shows that as the sample size increase, the values of MSE are getting smaller for the parameter estimate.

6. APPLICATION

In this Section, we demonstrate the capability of the ET-Exp distribution by fitting the model to four datasets, namely over the Gompertz, Exponential, Lindley, Weibull, and Generalized Exponential (GE) distributions. For theses four datasets, the maximum likelihood procedure is used to estimate the parameters of each distribution. Using the obtained estimates, we get the values of Akaike information criterion (AIC), Bayesian information criterion (BIC) and -log L.

Moreover, we find the Kolmogorov-Smirnov (K-S) statistic with its corresponding P-value (P-Val), and Anderson-Darling (AD) statistics. Basic descriptive statistics are calculated for all datasets, including the five number summary, mean, variance, skewness and kurtosis. The distribution with the lowest AIC, BIC, and -Log L is considered as the most flexible distribution for a given dataset.

Growth hormone data: The first set of data consists of 40 observations represents the estimated time since given growth hormone medication until the children reached the target age in the Programa Hormonal de Secretaria de Saude de Minas Gerais [16]. The dataset was analyzed by [1]. The datasets are: 2.15, 2.20, 2.55, 2.56, 2.63, 2.74, 2.81, 2.90, 3.05, 3.41, 3.43, 3.43, 3.84, 4.16, 4.18, 4.36, 4.42, 4.51,4.60, 4.61, 4.75, 5.03, 5.10, 5.44, 5.90, 5.96, 6.77, 7.82,8.00, 8.16, 8.21, 8.72, 10.40, 13.20, 13.70. Table 2 provides the descriptive statistics for this data set and Table 3 presents the results of MLEs and goodness of fit tests for this data set using each distribution.

Sample Size	$\theta = 0.1$			$\alpha = 3$		
n	**Estimate**	**Bias**	**MSE**	**Estimate**	**Bias**	**MSE**
50	0.10264	-0.00264	0.00020	3.28413	-0.28412	1.26297
75	0.10165	-0.00165	0.00013	3.17688	-0.17688	0.73347
100	0.10132	-0.00132	0.00009	3.13598	-0.13598	0.53885
250	0.10042	-0.00042	0.00004	3.05039	-0.05039	0.18580
400	0.10031	-0.00031	0.00002	3.03191	-0.03191	0.11652
Sample Size	$\theta = 3.1$			$\alpha = 0.2$		
n	**Estimate**	**Bias**	**MSE**	**Estimate**	**Bias**	**MSE**
50	3.35913	-0.25913	0.51010	0.38221	-0.18221	0.20171
75	3.27712	-0.17712	0.32454	0.32181	-0.12181	0.12360
100	3.23269	-0.13269	0.23782	0.29313	-0.09313	0.09281
250	3.15159	-0.05159	0.09936	0.23775	-0.03775	0.03938
400	3.12687	-0.02686	0.06317	0.22022	-0.02022	0.02522
Sample Size	$\theta = 6$			$\alpha = 3$		
n	**Estimate**	**Bias**	**MSE**	**Estimate**	**Bias**	**MSE**
50	6.15306	-0.15306	0.73829	3.27790	-0.27789	1.26813
75	6.09756	-0.09756	0.45751	3.17543	-0.17543	0.73278
100	6.09033	-0.09033	0.34335	3.14626	-0.14626	0.53472
250	6.03441	-0.03441	0.13502	3.05740	-0.05740	0.19413
400	6.01596	-0.01596	0.08147	3.03188	-0.03188	0.11481
Sample Size	$\theta = 0.6$			$\alpha = 0.3$		
n	**Estimate**	**Bias**	**MSE**	**Estimate**	**Bias**	**MSE**
50	0.63933	-0.03933	0.01777	0.45488	-0.15488	0.22516
75	0.62767	-0.02767	0.01160	0.40929	-0.10929	0.14199
100	0.62057	-0.02057	0.00878	0.38447	-0.08447	0.11042
250	0.60632	-0.00632	0.00367	0.32713	-0.02713	0.04553
400	0.60414	-0.00414	0.00242	0.31588	-0.01588	0.03049

Table 1: Estimate, Bias and Mean Square Error of MLEs of parameters α and θ

Parameters	N	**Min**	Q_1	**Median**	Q_3	**Mean**	**Max**	**Skewness**	**Kurtosis**	**Variance**
Data set I	35	2.15	3.23	4.51	6.365	5.306	13.7	1.3706	4.4008	8.4754

Table 2: The descriptive statistics for the growth hormone medication data set.

Distribution	α	θ	**-Log L**	**K-S**	**P-Val**	**A-D**	**P-Val**	**AIC**	**BIC**
Gompertz	0.18	0.50	87.10	0.21	0.10	1.81	0.12	178.20	181.30
Lindley	0.33		87.47	0.25	0.03	2.41	0.055	176.95	178.50
Exponential	0.19		93.41	0.33	0.0008	4.49	0.005	188.81	190.37
ET-Exp	7.18	0.52	79.84	0.11	0.83	0.63	0.62	163.68	166.79
Weibull	1.99	6.03	82.49	0.15	0.45	0.98	0.37	168.98	172.09
GE	6.51	0.48	79.10	0.10	0.86	0.53	0.72	162.20	165.31

Table 3: MLEs and goodness of fit statistics for the growth hormone medication data set.

Flood data: The second set of data has been presented by [9] and acts the maximum flood levels (in million of cubic feet/s) of the Susquehanna River at Harrisburg, Pennsylvania from 1890 to 1969, and its values are: 0.645, 0.613, 0.315, 0.449, 0.297, 0.402, 0.379, 0.423, 0.379, 0.324, 0.269, 0.740, 0.218, 0.412, 0.494, 0.416, 0.338, 0.392, 0.484 and 0.265. Table 4 provides the descriptive statistics for this data set and Table 5 presents the results of MLEs and goodness of fit tests for this data set using each distribution.

Parameters	N	Min	Q_1	Median	Q_3	Mean	Max	Skewness	Kurtosis	Variance
Data set II	20	0.218	0.3217	0.397	0.4577	5.4127	0.74	0.9116	3.368	0.0176

Table 4: The descriptive statistics for the maximum flood levels of the Susquehanna River data set.

Distribution	α	θ	-Log L	K-S	P-Val	A-D	P-Val	AIC	BIC
Gompertz	6.08	0.06	-9.71	0.19	0.47	1.06	0.32	-15.42	-13.43
Lindley	3.02		1.72	0.41	0.002	4.42	0.006	5.42	6.42
Exponential	2.42		2.3	0.42	0.0015	4.66	0.004	6.6	7.59
ET-Exp	33.65	10.01	-14.36	0.11	0.97	0.19	0.99	-24.71	-22.72
Weibull	3.31	0.46	-12.43	0.17	0.63	0.58	0.67	-20.86	-18.87
GE	31.81	9.83	-14.38	0.11	0.97	0.19	0.99	-24.75	-22.76

Table 5: MLEs and goodness of fit statistics for the maximum flood levels of the Susquehanna River data set.

Rock samples data: The third set of data is given by [8] consists of the shape perimeter by squared (area) from measurements of 48 rock samples from a petroleum reservoir. The data are listed as follows: 0.0903296, 0.2036540, 0.2043140, 0.2808870, 0.1976530, 0.3286410, 0.1486220, 0.1623940, 0.2627270, 0.1794550, 0.3266350, 0.2300810, 0.1833120, 0.1509440, 0.2000710, 0.1918020, 0.1541920, 0.4641250, 0.1170630, 0.1481410, 0.1448100, 0.1330830, 0.2760160, 0.4204770, 0.1224170, 0.2285950, 0.1138520, 0.2252140, 0.1769690, 0.2007440, 0.1670450, 0.2316230, 0.2910290, 0.3412730, 0.4387120, 0.2626510, 0.1896510, 0.1725670, 0.2400770, 0.3116460, 0.1635860, 0.1824530, 0.1641270, 0.1534810, 0.1618650, 0.2760160, 0.2538320, 0.2004470

Table 6 presented the descriptive statistics for this data set and Table 7 presents the results of MLEs and goodness of fit tests for this data set using each distribution.

Parameters	N	Min	Q_1	Median	Q_3	Mean	Max	Skewness	Kurtosis	Variance
Data set III	48	0.0903	0.1623	0.1989	0.2627	0.2181	0.4641	1.1693	4.1098	0.00697

Table 6: The descriptive statistics for the rock samples from a petroleum reservoir

Distribution	α	θ	-Log L	K-S	P-Val	A-D	P-Val	AIC	BIC
Gompertz	0.14	8.22	-45.25	0.18	0.10	2.57	0.05	-86.50	-82.75
Lindley	5.31		-25.63	0.38	0.00	9.33	0.00	-49.26	-47.39
Exponential	4.58		-25.09	0.39	0.00	9.57	0.00	-48.18	-46.31
ET-Exp	19.41	16.60	-57.94	0.10	0.71	0.31	0.93	-111.89	-108.15
Weibull	2.75	0.25	-52.74	0.15	0.23	1.23	0.26	-101.48	-97.74
GE	17.84	16.06	-58.10	0.10	0.74	0.27	0.96	-112.14	-108.45

Table 7: MLEs and goodness of fit statistics for the rock samples from a petroleum reservoir data set.

Ball bearings failure time data: The fourth set of data is obtained from [12] represents the number of million revolutions before failure of 23 endurance of deep-groove ball bearings in the life test. These failure times are: 17.88, 28.92, 33.00, 41.52, 42.12, 45.60, 48.40, 51.84, 51.96, 54.12, 55.56, 67.80, 68.64, 68.64, 68.88, 84.12, 93.12, 98.64, 105.12, 105.84, 127.92, 128.04 and 173.40.

Table 8 reveals certain descriptive statistics regarding this data set and 9 provides the results of MLEs and goodness of fit tests for this data set using each distribution.

Parameters	N	Min	Q_1	Median	Q_3	Mean	Max	Skewness	Kurtosis	Variance
Data set IV	23	17.88	47.00	67.80	95.88	72.22	173.40	0.94	3.49	1405.58

Table 8: The descriptive statistics for the ball bearings failure time data

Distribution	α	θ	-Log L	K-S	P-Val	A-D	P-Val	AIC	BIC
Gompertz	0.33	0.02	115.98	0.15	0.65	0.73	0.53	235.96	238.23
Lindley	0.027		115.74	0.19	0.36	0.93	0.34	233.47	234.61
Exponential	0.014		121.43	0.31	0.03	0.31	0.03	244.87	246.00
ET-Exp	5.99	0.035	113.11	0.11	0.93	0.22	0.98	230.22	232.50
Weibull	2.10	81.87	113.69	0.15	0.67	0.32	0.91	231.38	233.65
GE	5.28	0.032	112.98	0.11	0.96	0.19	0.99	229.96	232.23

Table 9: MLEs and goodness of fit statistics for the ball bearings failure time data.

It can be seen that for the four datasets, GE and ET-Exp distributions have the smallest values of the Kolmogorov-Smirnov (largest P-values), Anderson-Darling, AIC and BIC goodness-of-fit tests statistics which indicate that the best fit is provided by the GE and ET-Exp distributions for these data sets. Based on the values of these statistics, we conclude that the GE and ET-Exp distributions provide the best fit in all the data sets examined. In the cases considered, the ET-Exp and GE performed far better than the Gompertz, Lindly and Exponential distributions while the Weibull distribution performed better than Gompertz and Lindly but not as good as ET-Exp and GE.

7. CONCLUDING REMARKS

A new family of lifetime distributions referred to as exponential transformation (ET) with flexible and desirable properties is proposed. Properties of the ET distribution and a sub-distribution were presented. The PDF, CDF, moments, hazard function, reliability and quantile function were presented. Entropy measures including rényi entropy, tsallis entropy for ET distribution were also derived. Estimate of the model parameters via the method of maximum likelihood obtained and applications to illustrate the usefulness of the proposed model to real data given. The applications provided show that ET distribution performs better than other several models in the literature.

ACKNOWLEDGMENTS

We would like to thank the two referees, the Editor and the associate editor for the constructive comments on the paper.

REFERENCES

[1] Alizadeh, A. and Bagheri, S. and Samani, E. B. and Ghobadi, S. and Nadarajah, S. (2018). Exponentiated power lindley power series class of distributions: theory and applications, *Communications in Statistics-Simulation and Computation*, **47**, 9, 2499–2531.

[2] Altun, E. and Korkmaz, M. Ç and El-Morshedy, M. and Eliwa, M. (2021). A new flexible family of continuous distributions: the additive odd-g family, *Mathematics*, **9**, 16, 1837.

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

[4] Aslam, M., Ley, C., Hussain, Z., Shah, S. F. and Asghar, Z. (2020). A new generator for proposing flexible lifetime distributions and its properties, *PloS one*, **15**, 4, e0231908.

[5] Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian journal of statistics*, **12**, 2, 171–178.

[6] Chesneau, C., Bakouch, H., and Hussain, T. (2019). A new class of probability distributions via cosine and sine functions with applications, *Communications in Statistics-Simulation and Computation*, **48**, 8, 2287–2300.

[7] Cordeiro, G. and de Castro, M. (2011). A new family of generalized distributions, *Journal of statistical computation and simulation*, **81**, 7, 883–898.

[8] Cordeiro, G. and dos Santos Brito, R. (2012). The beta power distribution, *Brazilian journal of probability and statistics*, **26**, 1, 88–112.

[9] Dumonceaux, R. and Antle, C. E. (1973). Discrimination between the lognormal and the Weibull distributions, *Technometrics*, **15**, 4, 923–926.

[10] Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications, *Communications in Statistics-Theory and methods*, **31**, 4, 497–512.

[11] Goldoust, M., Rezaei, S., Alizadeh, M. and Nadarajah, S. (2019). The odd log-logistic power series family of distributions: properties and applications, *Statistica*, **79**, 1, 77–107.

[12] Gupta, R. D. and Kundu, D. (1999). Theory & methods: generalized exponential distributions, *Australian & New Zealand Journal of Statistics*, **41**, 2, 173–188.

[13] Kumar, D. and Singh, U. and Singh, S. K. (2015). A new distribution using sine function-its application to bladder cancer patients data, *Journal of Statistics Applications & Probability*, **4**, 3, 417-427.

[14] Kumar, D., Kumar, P., Kumar, P., Singh, S. K. and Singh, U. (2021). PCM transformation: properties and their estimation, *Journal of Reliability and Statistical Studies*, **14**, 2, 373–392.

[15] Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes, *Journal of hydrology*, **46**, 1-2, 79–88.

[16] Lemos de Morais, A. (2009). *A class of generalized beta distributions, Pareto power series and Weibull power series*, Universidade Federal de Pernambuco.

[17] Marshall, A. and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families, *Biometrika*, **84**, 3, 641–652.

[18] Maurya, S., Kaushik, A. and Singh, R., and Singh, S. K. and Singh, U. (2016). A new method of proposing distribution and its application to real data, *Imperial Journal of Interdisciplinary Research*, **2**, 6, 1331–1338.

[19] Mudholkar, G. and Srivastava, D. (1993). Exponentiated weibull family for analyzing bathtub failure-rate data, *IEEE transactions on reliability*, **42**, 2, 299–302.

[20] Shaw, W. and Buckley, I. (2007). The alchemy of probability distributions: Beyond gram-charlier & cornish-fisher expansions, and skew-normal or kurtotic-normal distributions, *Map, Research report, Financial Mathematics Group, King's College*, 1–18.

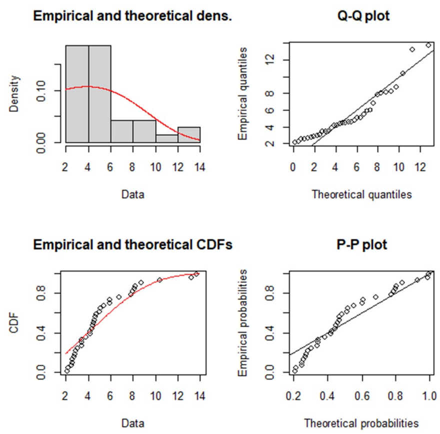
[21] Shaw, W. and Buckley, I. (2009). The alchemy of probability distributions: beyond Gram-Charlier expansions, and a skew-kurtotic-normal distribution from a rank transmutation map, *arXiv preprint arXiv:0901.0434*.

[22] Team, R Core (2021). R: A language and environment for statistical computing (R Version 4.0. 3, R Foundation for Statistical Computing, Vienna, Austria, *Google Scholar There is no corresponding record for this reference*.

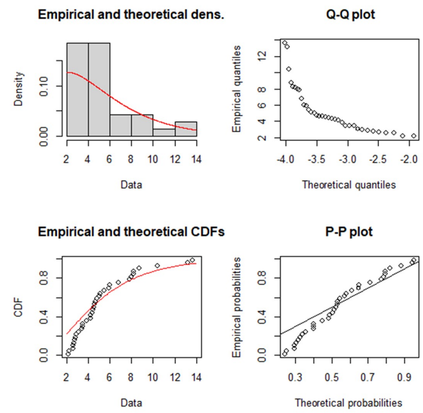
[23] Zaidi, S., Sobhi, M., El-Morshedy, M. and Afify, A. (2021). A new generalized family of distributions: Properties and applications, *Aims Math*, **6**, 1, 456–476.

[24] Zichuan, M., Hussain, S., Iftikhar, A., Ilyas, M., Ahmad, Z., Khan, D. M. and Manzoor, S. (2020). A new extended-family of distributions: properties and applications, *Computational and Mathematical Methods in Medicine*, **2020**, 4650520.

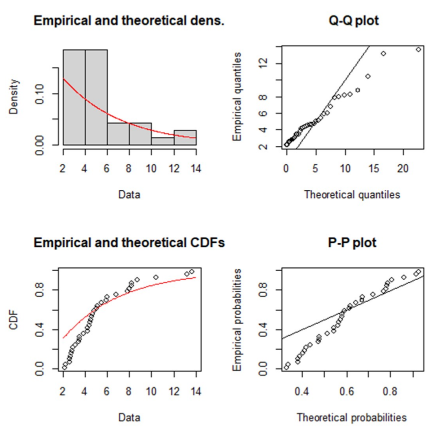
[25] Zografos, K. and Balakrishnan, N. (2009). On families of beta-and generalized gamma-generated distributions and associated inference, *Statistical methodology*, **6**, 4, 344–362.



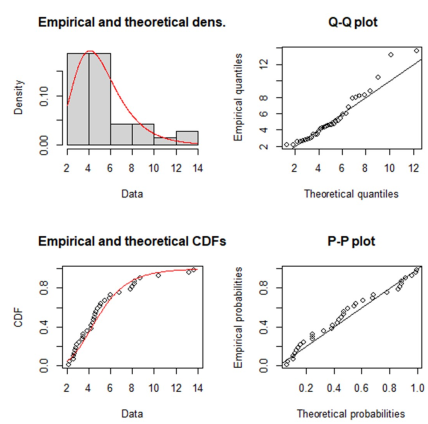
(a) Gompertz



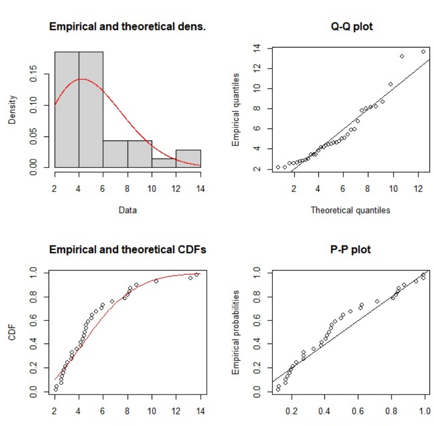
(b) Lindly



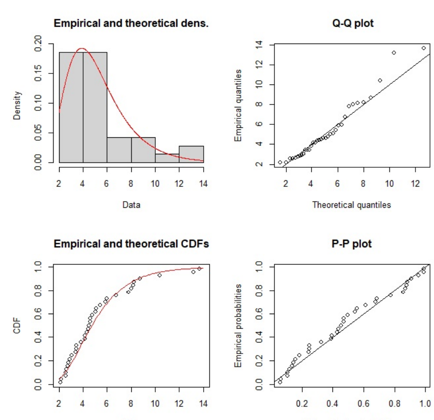
(c) Exponential



(d) ET-Exponential

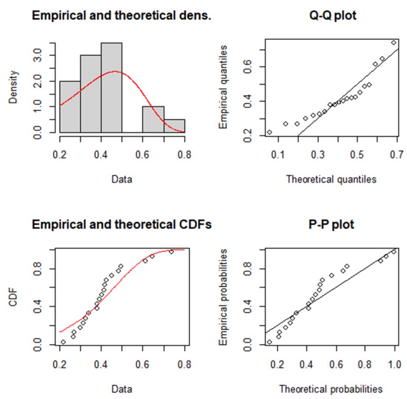


(e) Weibull

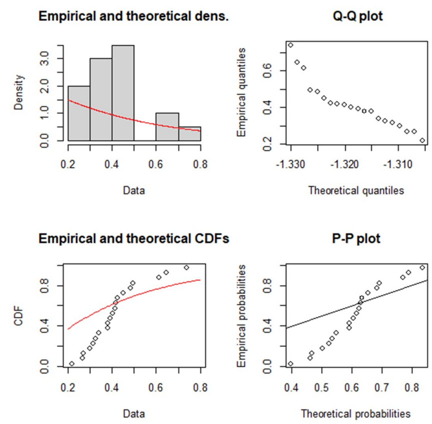


(f) Generalized Exponential

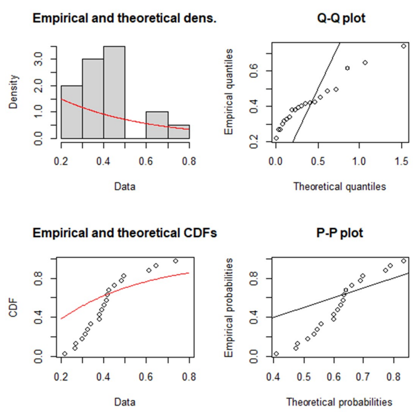
Figure 3: The empirical and theoretical PDFs, empirical and theoretical CDFs, Q-Q plots and p-p plot for (a) Gompertz, (b) Lindly, (c) Exponential, and (d) ET-Exponential, (e) Weibull, and (f) Generalized-Exponential for the growth hormone medication dataset



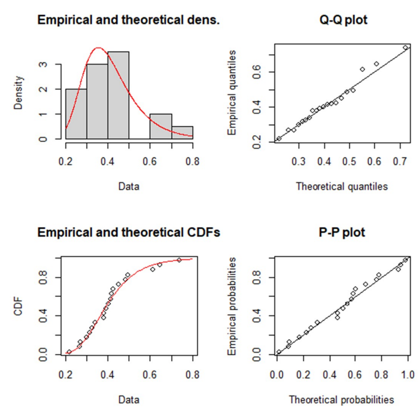
(a) Gompertz



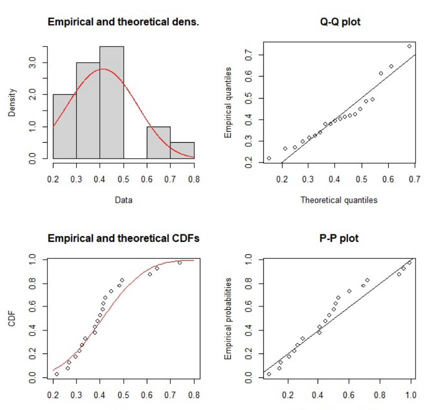
(b) Lindly



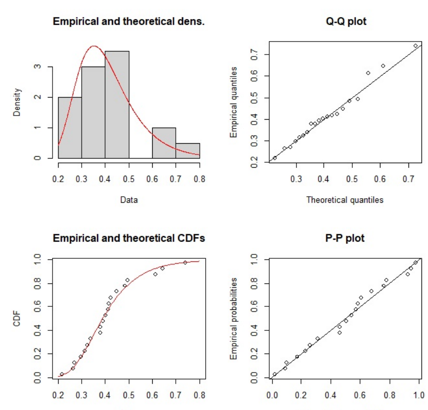
(c) Exponential



(d) ET-Exponential

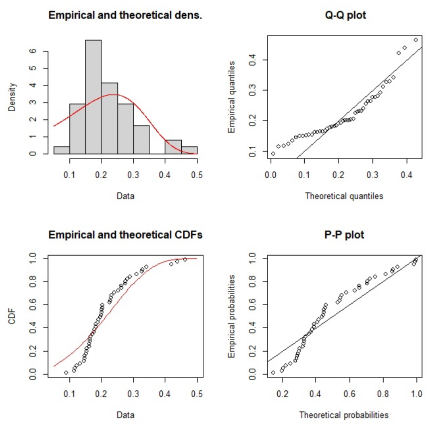


(e) Weibull

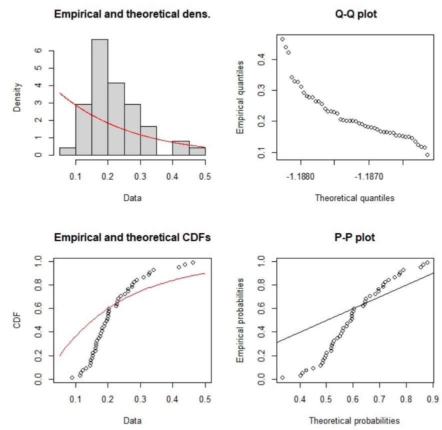


(f) Generalized Exponential

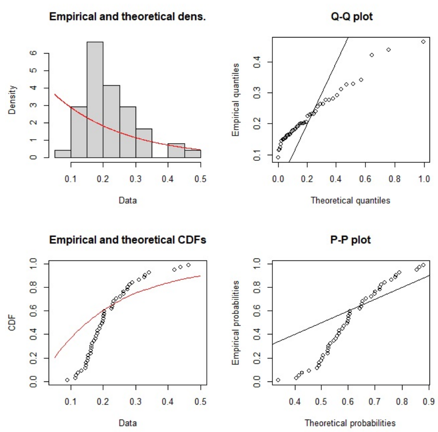
Figure 4: The empirical and theoretical PDFs, empirical and theoretical CDFs, Q-Q plots and p-p plot for (a) Gompertz, (b) Lindly, (c) Exponential, (d) ET-Exponential, (e) Weibull, and (f) Generalized-Exponential for the flood dataset



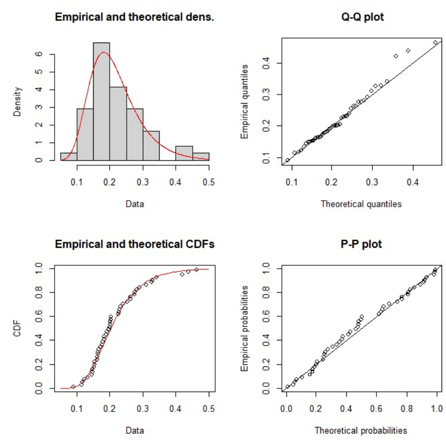
(a) Gompertz



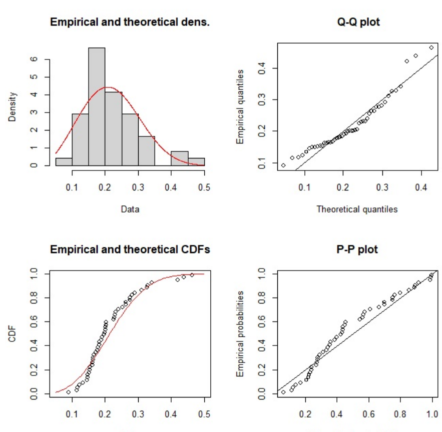
(b) Lindly



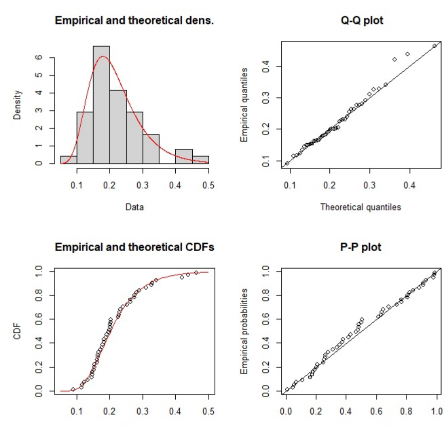
(c) Exponential



(d) ET-Exponential

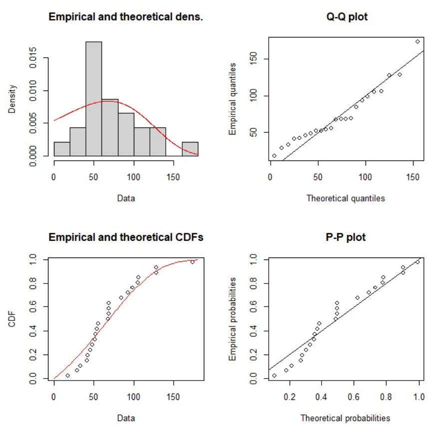


(e) Weibull

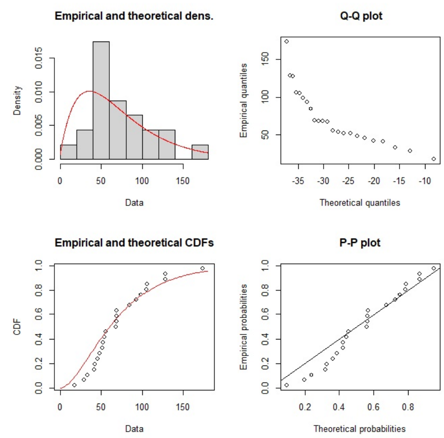


(f) Generalized Exponential

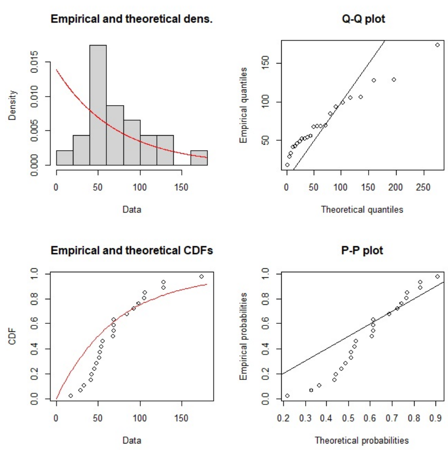
Figure 5: The empirical and theoretical PDFs, empirical and theoretical CDFs, Q-Q plots and p-p plot for (a) Gpmpertz, (b) Lindly, (c) Exponential, (d) ET-Exponential, (e) Weibull, and (f) Generalized-Exponential for the rock sample dataset



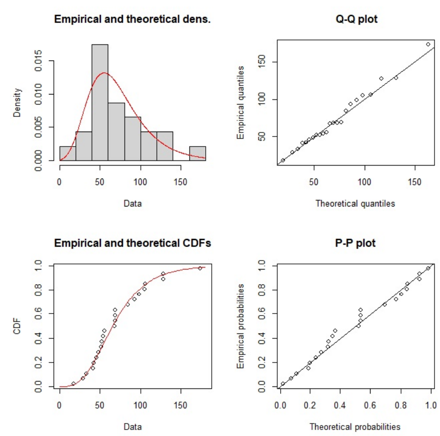
(a) Gompertz



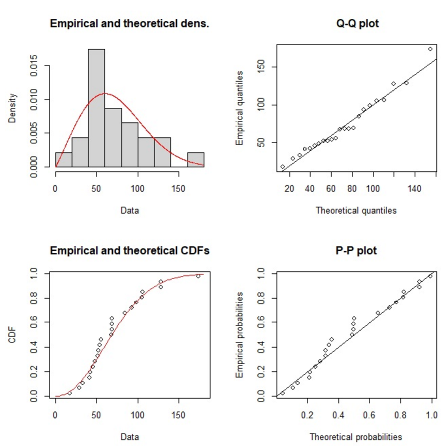
(b) Lindly



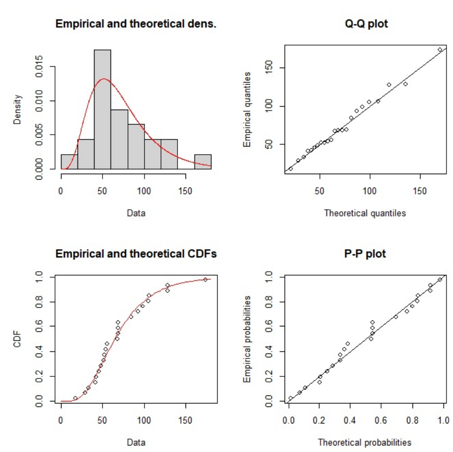
(c) Exponential



(d) ET-Exponential



(e) Weibull



(f) Generalized Exponential

Figure 6: The empirical and theoretical PDFs, empirical and theoretical CDFs, Q-Q plots and p-p plot for (a) Gpmpertz, (b) Lindly, (c) Exponential, (d) ET-Exponential, (e) Weibull, and (f) Generalized-Exponential for the ball bearings failure time data