Pak. J. Statist.
2023 Vol. 39(3), 237-258

A FLEXIBLE FAMILY OF DISTRIBUTIONS BASED ON THE ALPHA POWER FAMILY OF DISTRIBUTIONS AND ITS APPLICATION TO SURVIVAL DATA

Veronica Kariuki[1§], **Anthony Wanjoya**[2], **Oscar Ngesa**[3]
and **Mohammed Alqawba**[4]

[1] Department of Mathematics, Pan African Institute of
Basic Science Technology and Innovation, Nairobi, Kenya
[2] Department of Statistics and Actuarial Sciences
Jomo Kenyatta University of Agricultureand Technology
Kiambu, Kenya. Email: awanjoya@fsc.jkuat.ac.ke
[3] Department of Mathematics, Statistics and Physical Science
Taita Taveta University, Voi Kenya. Email: oscanges@ttu.ac.ke
[4] Department of Mathematics, College of Science and Arts
Qassim University, Ar Rass, Saudi Arabia
Email: m.alqawba@qu.edu.sa
[§] Corresponding Author Email: veronica.kariuki@students.jkuat.ac.ke

ABSTRACT

In this paper, we propose a flexible family called the exponentiated alpha-power-G (EAP-G) family. The benefits of the proposed family include its analytical simplicity and its ability to confer flexibility to the baseline distributions in survival analysis. Based on the proposed approach, a three-parameter extension of the exponential distribution called the exponentiated alpha-power exponential (EAPE) distribution is studied in detail. Maximum likelihood is used to estimate the EAPE parameters, and its performance is evaluated via a simulation study. Furthermore, two real-world survival data are used to demonstrate the applicability and examine the flexibility of the proposed distribution. The EAPE distribution is compared to other competing generalizations of the exponential distribution. The real data analysis shows that the proposed model performed better among the competitors and could potentially be very adequate in describing and modeling a wide range of survival data.

KEYWORDS

Exponential distribution; maximum likelihood estimation; alpha-power transformation; exponentiated family; simulation study.

1. INTRODUCTION

Recently, there has been a sharp increase in research making attempts to develop new families of probability distributions and to expand well-known distributions, giving flexible classes that support modeling data in a wide range of areas, including engineering, medical studies, economics, environmental sciences, and finance to mention just a few. Probability distributions have a vast variety of applications in modeling data from a variety of physical phenomena as well as natural processes (Alshanbari et al., 2022). For instance; lifetime distributions have been applied in reliability engineering for the determination of failure rates of equipment, in biological sciences, they have been used to study bacteria count, in medical studies to determine the life expectancy of critically ill patients after treatment, and in actuarial sciences to determine insurance premium among others (Oluyede et al., 2018).

The choice of appropriate distributions to be used on real-life data plays a fundamental role in improving the power, efficiency, and sensitivity of statistical tests. This is so because appropriate distributions lead to a good fit of the data. Therefore, good knowledge of the appropriate distribution to be used for a specific data set is essential. However, the use of classical probability distributions has various limitations: for example, the hazard rate of the exponential distribution is constant, and the hazard rate of the Weibull distribution varies from decreasing, constant, or increasing. In real life, the hazard rate of the available data from different disciplines may take different forms such as bathtub, non-monotonic, or even unimodal (Ijaz et al., 2020). Another example is that distributions such as the normal distribution and the student-t distribution are symmetric while available data may exhibit characteristics with varying degrees of skewness and kurtosis (Ma & Genton, 2004).

An alternative and appropriate approach to overcome these limitations is to modify the existing statistical probability distributions and create more flexible distributions with a better fit to real data (Kilai et al., 2022) and (Aslam et al., 2019). Modifying distributions or families of distributions creates more flexible distributions that provide a more reliable fit to different types of hazard rates. Probability distribution modification may also achieve heavy-tailed distributions used in modeling varied data (Zhao et al., 2021). A tractable cumulative distribution function (CDF) that model data characterized by different skewness and kurtosis levels are also generated. Methods of modifying distributions exist in the literature, some of which include the method of introducing skewness presented by (Azzalini and Valle, 1996), beta generated method by (Eugene et al., 2006), the exponentiated method by (Govind, 1993), Kumaraswamy-G method by (Cordeiro & de Castro, 2011), transformed transformer and exponentiated transformed transformer methods by (Alzaatreh et al., 2013), exponentiated generalized transformed transformer by (Nasiru et al., 2017) and (Kilai et al., 2022) presented an extension of the gull alpha power family of distribution. Additionally, Nassar et al. (2019) proposed the Marshall–Olkin alpha power family.

Mahdavi and Kundu (2017) introduced a method of modifying distributions called the alpha power transformed (APT). This method adds one extra shape parameter to the classical distribution under study, improving its flexibility for modeling real data. The family has been used in studying the flexibility of different baseline distributions. For example, Aldahlan (2020) transformed the log-logistic distribution using APT, Mead et al. (2019) studied the alpha power transformed exponentiated-Weibull distribution. The APT Fréchet distribution is introduced by (Nasiru et al., 2019). (Nassar et al., 2020) studied different estimation methods for the parameters of the alpha-power exponential distribution. Afify et al. (2020) proposed the alpha-power exponentiated-exponential distribution.

However, the method adds one shape parameter that only controls skewness and fails to control kurtosis. To be able to control both measures simultaneously, we propose to introduce another shape parameter to the APT family of distribution. Hence, we propose the so-called EAP-G family which extends the APT family and provides more flexibility to the baseline distributions in survival analysis. The EAP-G family is constructed by using the APT family as a baseline in the exponentiated-G (EG) family. Furthermore, we will study a special member of the proposed family called the EAPE distribution in more detail.

The exponential distribution has gained a vast variety of applications in modeling survival data, especially in modeling lifespan tests and the metrics related to them such as the mean residual life function, the HR, the mean time to failure, and reliability.

One major reason why researchers have been attracted to the exponential distribution is becauseof the perception that it has done outstandingly well in several uses on various survival analyses. This is due to the existence of its closed-form solutions of the cumulative distribution function. Notably, the exponential distribution is the only known continuous distribution characterized by a constant hazard(failure) rate as well as a memory-less property (Piriaei et al., 2020). Due to the availability of various methods of modifying distributions, the exponential distribution has been modified by a numberof researchers to achieve a better fit for real data. Some of these extensions include; the exponentiated exponential by (Nadarajah & Kotz, 2006b), beta-exponential distribution by (Nadarajah & Kotz, 2006a), and the exponentiated generalized alpha power exponential distribution by (ElSherpieny & Almetwally, 2022) and the extended odd Weibull exponential distribution by (Afify & Mohamed, 2020). This study, therefore, proposes to extend the exponential distribution using the exponentiated family of distributions together with the alpha power transformation family of distributions. We call the proposed distribution the exponentiated alpha power exponential distribution.

The cumulative distribution function (CDF) of the Alpha Power family of distributions is defined as:

$$H_{APT}(x;\gamma,\psi) = \begin{cases} \dfrac{\gamma^{G(x;\psi)}-1}{\gamma-1} & \text{if } \gamma>0, \gamma\neq 1 \\ G(x;\psi) & \text{if } \gamma = 1 \end{cases} \tag{1}$$

the probability density function (PDF) is given by:

$$h(x;\gamma,\psi) = \begin{cases} \dfrac{\log\gamma}{\gamma-1} g(x;\psi)\gamma^{G(x;\psi)} & \text{if } \gamma>0, \gamma\neq 1 \\ g(x;\psi) & \text{if } \gamma = 1 \end{cases} \tag{2}$$

where, $\gamma > 0$ with $\gamma \neq 1$ is a shape parameter and ψ is a parameter vector for the baseline distribution. The CDF of the exponentiated-G family of distributions is given by:

$$\mathrm{F}(\mathrm{x};\eta) = [\mathrm{G}(\mathrm{x};\eta)]^{\eta} \tag{3}$$

while the PDF associated with equation 3 is given as:

$$f(x;\eta) = \eta g(x;\eta)[G(x;\eta)]^{\eta-1} \tag{4}$$

where $\eta > 0$ is a shape parameter.

The remaining part of this paper is described as follows. Section 2 describes the exponentiated alpha power exponential distribution and its properties. The maximum likelihood estimation for the parameters of the proposed distribution is given in section 3. A simulation study is conducted to examine the performance of the estimates in section 4, and section 5 contains an application of the extended exponential distribution to real-world data. Section 6 contains some concluding remarks.

2. THE EAP-G FAMILY

The EAP-G family can be specified by the following CDF.

$$F_{EAP}(x;\gamma,\eta,\psi) = \begin{cases} \left[\dfrac{\gamma^{G(x;\psi)}-1}{\gamma-1}\right]^{\eta} & \text{if } \eta,\gamma > 0, \gamma \neq 1, \\ G(x;\psi) & \text{if } \gamma = 1. \end{cases} \tag{5}$$

The PDF of the EAP-G family reduces to

$$f_{EAP}(x;\gamma,\eta,\psi) = \begin{cases} \eta \dfrac{\log\gamma}{\gamma-1} g(x;\psi)\gamma^{G(x;\psi)} \left[\dfrac{\gamma^{G(x;\psi)}-1}{\gamma-1}\right]^{\eta-1} & \text{if } \eta,\gamma > 0, \gamma \neq 1, \\ g(x;\psi) & \text{if } \gamma = 1. \end{cases} \tag{6}$$

Using the baseline exponential distribution in the EAP-G family, we can define the EAPE distribution.

The CDF of EAPE distribution is defined as

$$F_{EAPE}(x) = \left(\frac{\gamma^{1-e^{-\zeta x}}-1}{\gamma-1}\right)^{\eta}. \tag{7}$$

The corresponding PDF is given by

$$f_{EAPE}(x) = \frac{\eta\zeta log\, \gamma e^{-\zeta x}\gamma^{1-e^{-\zeta x}}}{\gamma-1}\left(\frac{\gamma^{1-e^{-\zeta x}}-1}{\gamma-1}\right)^{\eta-1} \tag{8}$$

where $\gamma > 0$ and $\eta > 0$ are shape parameters, and ζ is a scale parameter. Figure 1 below demonstrates that the PDF of EAPE distribution can be J-shaped, inverted-J, almost symmetric, right-skewed, and unimodal. Figure 2 indicates that the hazard rate function assumes different shapes including increasing, decreasing, and unimodal.

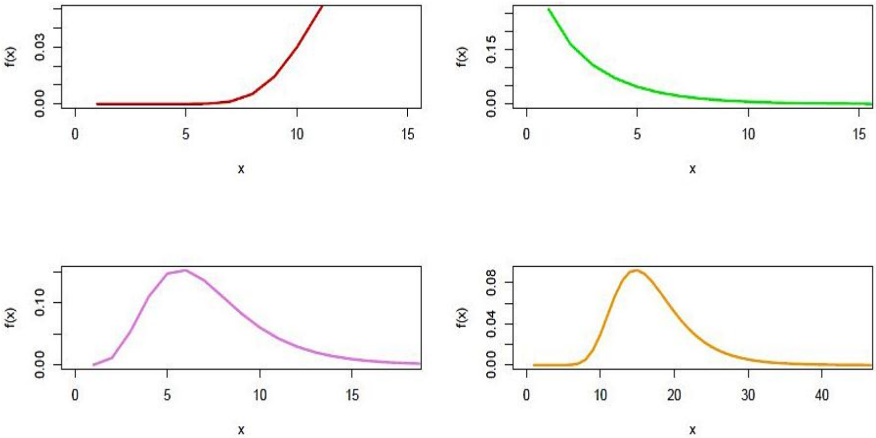


Figure 1: EAPE Distribution PDF Shapes for Different Parameter Combination

The survival function is derived as;

$$S_{EAPE}(x) = 1 - \left(\frac{\gamma^{1-e^{-\zeta x}} - 1}{\gamma - 1}\right)^{\eta}. \tag{9}$$

So, the hazard function is given as

$$h_{EAPE}(x) = \frac{\eta\zeta e^{-\zeta x}(\log\gamma)\gamma^{1-e^{-\zeta x}}\left(\gamma^{1-e^{-\zeta x}} - 1\right)^{\eta-1}}{(\gamma - 1)^{\eta} - \left(\gamma^{1-e^{-\zeta x}} - 1\right)^{\eta}} \tag{10}$$

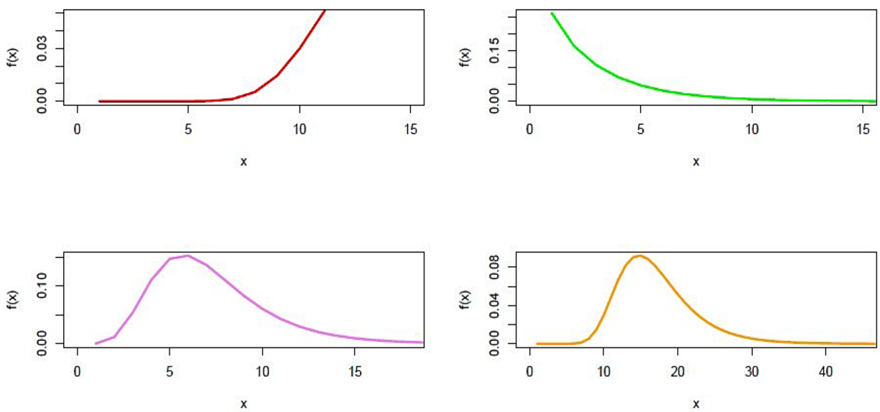


Figure 2: EAPE Distribution Hazard Shapes for Different Parameter Combination.

Table 1
Special Sub-Models of EAPE

γ	η	ζ	Sub-Model
γ	1	ζ	Alpha Power Exponential (APE) distribution
1	η	ζ	Exponentiated Exponential (EE) distribution
1	1	ζ	Exponential (E) distribution

2.1 Important Representation of EAPE

Using power series and binomial expansion, we have that

$$\gamma^{1-e^{-\zeta x}} = \sum_{i=0}^{\infty} \frac{(\log\gamma)^i}{i!}\left(1-e^{-\zeta x}\right)^i,$$

$$\left(\gamma^{1-e^{-\zeta x}}-1\right)^{\eta-1} = \sum_{j=0}^{\infty} \binom{\eta-1}{j}(-1)^j\left(\gamma^{1-e^{-\zeta x}}\right)^{\eta-1-j}$$

$$\left(\gamma^{1-e^{-\zeta x}}\right)^{\eta-1-j} = \sum_{k=0}^{\infty}\left(1-e^{-\zeta x}\right)^k\frac{(\log\gamma)^k}{k!}(\eta-1-j)^k$$

Using the above three relations, EAPE PDF given in equation 6 can be re-written as

$$f_{EAPE}(x) = \sum_{m=0}^{\infty} a_m g_{(m+1)\zeta}(x) \tag{11}$$

where

$$a_m = \sum_{i\cdot j\cdot k}\frac{\eta(-1)^{m+j}}{(\gamma-1)^{\eta}(m+1)}\binom{k+i}{m}\binom{\eta-1}{j}(\eta-1-j)^k\frac{(\log\gamma)^{k+i+1}}{i!\,k!}$$

and $g_{(m+1)\zeta}(x)$ is the exponential probability distribution function with the scale parameter $(m+1)\zeta$. Equation 9 clearly demonstrates that the EAPE PDF can be presented as a linear combination of the exponential density function with a scale parameter given as $(m+1)\zeta$. Thus, some of the structural properties of the EAPE PDF can be derived from the properties of the exponential density function.

2.1 EAPE Quantile Function

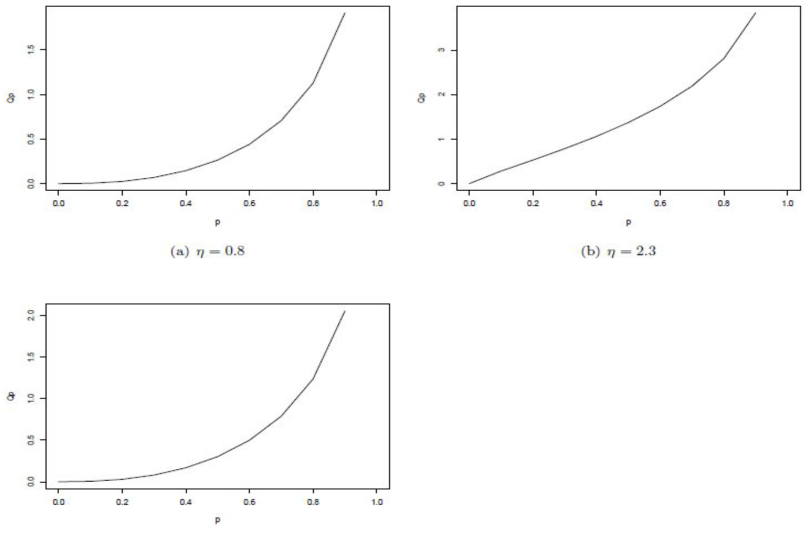
Using inverse transformation, we obtain the quantile function by inverting the EAPE CDF as

$$p = \left(\frac{\gamma^{1-e^{-\zeta x}}-1}{\gamma-1}\right)^{\eta}$$

Letting $x = Q(p)$ we obtain the quantile function as;

$$Q(p) = -\frac{1}{\zeta}\,ln\left[1-\frac{ln\left(p^{1/\eta}(\gamma-1)+1\right)}{ln(\gamma)}\right] \tag{12}$$

If P is a uniform random variable over the interval $(0,1)$, then it follows that the random variable $X = Q(P)$ assumes a PDF given by equation 6. Consequently, the simulation of random variables from EAPE distribution is straightforward using $Q(P)$. The quantile function is also useful in getting quartiles, skewness, and kurtosis.



(a) $\eta = 0.8$ (b) $\eta = 2.3$ (c) $\eta = 6$

Figure 3: Plots of EAPE Quantile for Different Parameter Values η of with Fixed Values of ($\gamma = 1.6, \zeta = 0.4$)

2.2 Skewness and Kurtosis

Galton's skewness (also called Bowley's skewness) is given as;

$$G = \frac{Q\left(\frac{3}{4}\right) + Q\left(\frac{1}{4}\right) - 2Q\left(\frac{1}{2}\right)}{Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right)}.$$

Whereas, Moor's kurtosis as defined by (Jones, 2007) is given as;

$$M = \frac{Q\left(\frac{3}{8}\right) - Q\left(\frac{1}{8}\right) + Q\left(\frac{7}{8}\right) - Q\left(\frac{5}{8}\right)}{Q\left(\frac{6}{8}\right) - Q\left(\frac{2}{8}\right)}.$$

These measures of shape are preferred as they are less sensitive to outliers. They are also known to exist even if distributions have no moments.

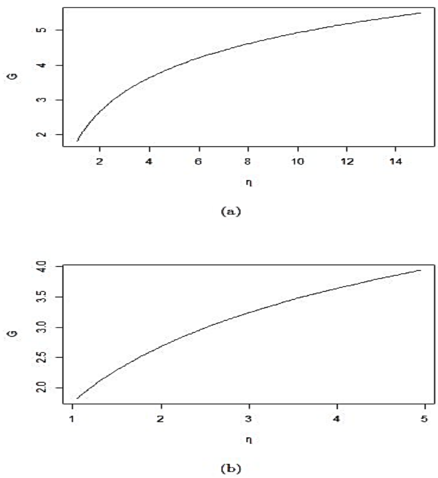


Figure 4: Plots of Galton's Skewness for EAPE Distribution with Fixed Values of ($\gamma = 5, \zeta = 0.4$)

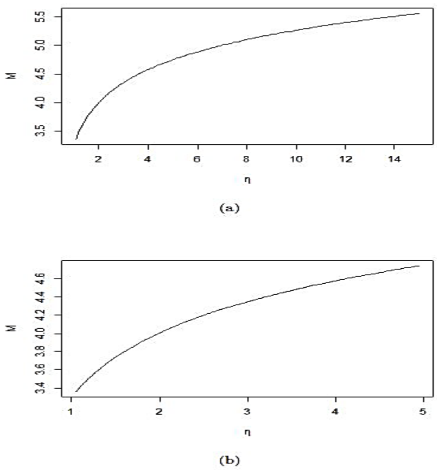


Figure 5: Plots of Moor's Kurtosis for EAPE Distribution with Fixed Values of ($\gamma = 5, \zeta = 0.4$)

2.3 Moments of EAPE Distribution

The r^{th} moment from EAPE PDF is given as;

$$\begin{aligned} U_r' &= \int_0^\infty x^r f_{EAPE}(x)dx \\ &= \int_0^\infty x^r \sum_{m=0}^\infty a_m g_{(m+1)\zeta}(x)dx \\ &= \sum_m a_m \int_0^\infty x^r g_{(m+1)\zeta}(x)dx \\ &= \sum_m a_m \frac{r!}{[(m+1)\zeta]^r} \end{aligned} \tag{13}$$

Table 2 below gives a summary of the first five moments and the variance of EAPE distributionfor some given parameter value combinations.

Table 2
Summary of the First Five Moments and Variance of EAPE Distribution

μ'_r	A	B	C	D	E
μ'_1	2.424	3.111	4.441	0.830	2.057
μ'_2	12.265	17.604	28.359	1.252	4.947
μ'_3	92.882	140.847	239.420	2.671	13.832
μ'_4	934.098	1456.537	2541.109	7.365	44.749
μ'_5	11711.960	18524.470	32741.090	24.980	166.606
σ^2	6.390	7.923	8.639	0.563	0.715

$A: \gamma = 1.6, \eta = 0.8, \zeta = 0.4$
$B: \gamma = 5, \eta = 0.8, \zeta = 0.4$
$C: \gamma = 1.6, \eta = 2.3, \zeta = 0.4$
$D: \gamma = 5, \eta = 0.8, \zeta = 1.5$
$E: \gamma = 5, \eta = 6, \zeta = 1.5$

2.5 Moment Generating Function

EAPE moment generating function (MGF) is given as

$$M_x(t) = \sum\nolimits_{r,m} \frac{a_m t^r}{[(m+1)\zeta]^r} \tag{14}$$

where am is as defined in equation 11.

The MGF defined in equation 14 uniquely determines EAPE distribution. The function can alsobeused in deriving moments of EAPE distribution.

2.6 Order Statistics

The PDF of the j^{th} order statistic from EAPE distribution can be given as;

$$f_j(x) = \frac{n!}{(j-1)!\,(n-j)!} f_{\text{EAPE}}(x)[F_{\text{EAPE}}(x)]^{j-1}[1 - F_{\text{EAPE}}(x)]^{n-j} \tag{15}$$

but $F_{\text{EAPE}}(x) = \sum_q d_q T(x)$

where $dq = \sum_p (-1)^p \binom{\eta}{p} \frac{(\log\gamma)^q}{q!} \frac{(\eta-p)^q}{(\gamma-1)^\eta}$ and $T(x) = \left(1 - e^{-\zeta x}\right)^q$ is the CDF of the exponentiated exponential (EE) distribution that has a shape parameter q and a scale parameter of $\zeta. f_{EAPE}$ is as defined in equation 11. We can, therefore, write the PDF of the j^{th} order statistic as;

$$f_j(x) = \frac{n!}{(j-1)!\,(n-j)!}\sum_m a_m g_{(m+1)\zeta}(x)\left[\sum_q d_q T(x)\right]^{j-1}\left[1-\sum_q d_q T(x)\right]^{n-j} \tag{16}$$

for $j = 1$ we obtain the density of the minimum order statistic as,

$$f_1(x) = n\sum_m a_m g_{m+1)\zeta}(x)\left[1-\sum_q d_q T(x)\right]^{n-1}. \tag{17}$$

Similarly, for $j = n$ we obtain the density of the maximum order statistic as;

$$f_n(x) = n\sum_m a_m g_{(m+1)\zeta}(x)\left[\sum_q d_q T(x)\right].^{n-1} \tag{18}$$

3. PARAMETER ESTIMATION

In this section, the maximum likelihood estimation method was used to estimate the parameters of the distribution, since the method results to estimate values with desirable properties. Such propertiesinclude consistency, efficiency, and minimum variance unbiased among others. The estimates can also be utilized in generating the confidence intervals for the parameters.

Let $X_1, X_2, ..., X_n$ be an independent and identically distributed random sample from EAPE (γ, η, ζ) distribution. Then, the likelihood function is given as the joint distribution as follows;

$$\mathrm{L}(x \mid \eta, \gamma, \zeta) = \prod_{i=1}^{n} f_{EAPE}(x) \tag{19}$$

Substituting the EAPE PDF given in equation 8 into equation 19 yields;

$$L(x \mid \eta, \gamma, \zeta) = \prod_{i=1}^{n} \frac{\eta\zeta(\log\gamma)e^{-\zeta x_i}\gamma^{1-e^{-\zeta x_i}}}{\gamma-1}\left(\frac{\gamma^{1-e^{-\zeta x_i}}}{\gamma-1}\right)^{\eta-1}$$

Hence, the log-likelihood function is defined as;

$$\ell(x;\eta,\gamma,\zeta) = \mathrm{nlog}(\eta\zeta) + \mathrm{nlog}(\log\gamma) + \sum [1-e^{-\zeta x_i}](\log\gamma) - \zeta\sum x_i - n\eta\log(\gamma-1) + (\eta-1)\sum \log\left(\gamma^{1-e^{-\zeta x}}-1\right). \tag{20}$$

$$\frac{\partial \ell}{\partial \gamma} = -\frac{n\eta}{\gamma-1} + \frac{n}{\gamma \log \gamma} + (\eta-1)\sum_{i=1}^{n}\left[\frac{\left(1-e^{-\zeta x_i}\right)\gamma^{-e^{-\zeta x_i}}}{\left(\gamma^{1-e^{-\zeta x_i}}-1\right)}\right] + \sum_{i=1}^{n}\left(\frac{1-e^{-\zeta x_i}}{\gamma}\right)$$

$$\frac{\partial \ell}{\partial \eta} = \frac{n}{\eta} - \text{nlog}(\gamma-1) + \sum_{i=1}^{n}\log\left(\gamma^{1-e^{-\zeta x_i}}-1\right)\frac{\partial \ell}{\partial \zeta}$$

$$= \frac{n}{\zeta} + \sum_{i=1}^{n} x_i e^{-\zeta x_i}(\log \gamma) - \sum_{i=1}^{n} x_i + (\eta-1)\sum_{i=1}^{n}\left[\frac{x_i e^{-\zeta x_i}(\log \gamma)\gamma^{1-e^{-\zeta x_i}}}{\left(\gamma^{1-e^{-\zeta x_i}}-1\right)}\right] \quad (21)$$

Equating the above three score functions to zero results in a system of equations and upon solving them, the maximum likelihood estimates (MLEs) for the parameters of the EAPE distribution may be determined. Since the system of equations cannot be solved analytically, numerical methods are usedin solving them. In this study, we employed the Broyden-Fletcher-Goldfarb-Shanno (BFGS) methodin solving the normal equations to obtain the maximum likelihood estimators for the parameters ofthe studied distributions. The BFGS has an algorithm that was independently introduced by (Broyden, 1970), (Fletcher, 1970), (Goldfarb, 1970), and (Shanno, 1970). The BFGS algorithm is one of the most efficientalgorithms for solving unconstrained optimization problems. It is an iterative technique that begins with an initial guess value Θ_o and an initial hessian matrix H_o to provide a solution to a given function. The following steps are followed so that for $k = (0,1,\dots)$ an approximation point Θ_k and an mXm matrix H_k are obtained on the k^{th} iteration. Given: $\Theta_o, H_o > 0$ and unconstrained optimization problem $\ell(\Theta)$.

1. First, the search direction, also called the quasi-Newton direction is obtained as; $\Delta\Theta_k = -H_k^{-1}\nabla\ell(\Theta_k)$.
2. The step-length d_k is obtained such that it meets certain line search conditions.
3. The next iterate is obtained as; $\Theta_{k+1} = \Theta_k + d_k\Delta\Theta_k$
4. An important feature of the algorithm is the choice of H_k. That is, it must be positive definite and must satisfy the quasi-newton formula given as;
   $H_{k+1}\alpha_k = \gamma_k$
   where, $\alpha_k = d_k\Delta\Theta_k$
   and $\gamma_k = \nabla\ell(\Theta_k + \alpha_k) - \nabla\ell(\Theta_k)$
5. Finally, the matrices H_k are updated by the BFGS formula
   $$H_{k+1} = H_k - \frac{H_k\alpha_k\alpha_k^T H_k}{\alpha_k^T H_k \alpha_k} + \frac{\gamma_k\gamma_k^T}{\alpha_k^T\gamma_k}$$
   where $\Theta = (\gamma, \zeta, \eta)$ is a vector of parameters of the EAPE distribution. For the purposes of determining the standard errors of the estimates of the parameters of EAPE distribution, Fisher's information matrix has been derived as;

$$I(\gamma,\zeta,\eta) = -E\begin{bmatrix} \dfrac{\partial^2\ell}{\partial\eta^2} & \dfrac{\partial^2\ell}{\partial\eta\,\partial\gamma} & \dfrac{\partial^2\ell}{\partial\eta\xi} \\ & \dfrac{\partial^2\ell}{\partial\gamma^2} & \dfrac{\partial^2\ell}{\partial\gamma\,\partial\zeta} \\ & & \dfrac{\partial^2\ell}{\partial\zeta^2}\end{bmatrix}, \tag{21}$$

which is equivalent to the final Hessian matrix obtained in the BFGS algorithm explained above. The elements of the fisher's information matrix for the EAPE distribution are given as

$$\frac{\partial^2\ell}{\partial\gamma^2} = -(\gamma\log\gamma)^2[\log\gamma + 1]$$

$$-(1-\eta)\sum_{i=1}^{n}\left(1-e^{-\zeta x_i}\right)\gamma^{-e^{-\zeta x_i}}\left[\frac{e^{-\zeta x_1}\left(\gamma^{1-e^{-\zeta x}}-1\right)+\gamma^{-e^{-\zeta x}}\left(1-e^{-\zeta x}\right)}{\left(\gamma^{1-e^{-\zeta x}}-1\right)^2}\right]$$

$$-\sum\left(\frac{1-e^{-\zeta x}}{\gamma^2}\right)\frac{\partial^2\ell}{\partial\eta^2} = -\frac{n}{\eta^2}\frac{\partial^2\ell}{\partial\zeta^2} = -\frac{n}{\zeta^2} - \sum x^2 e^{-\zeta}(\log\gamma)$$

$$+(\eta-1)\sum\frac{xe^{-\zeta x}\gamma^{1-e^{-\zeta x}}\left[\left(\gamma^{1-e^{-\zeta x}}-1\right)[1+e^{-\zeta x}(\log\gamma)] - e^{-\zeta x}(\log\gamma)\gamma^{1-e^{-\zeta x}}\right]}{\left(\gamma^{1-e^{-\zeta x}}-1\right)^2}$$

$$\frac{\partial^2\ell}{\partial\gamma\,\partial\eta} = \frac{n}{\gamma-1} + \sum\frac{\left(1-e^{-\zeta x}\right)\gamma^{-e^{-\zeta}}}{\left(\gamma^{1-e^{-\zeta x}}-1\right)}\frac{\partial^2\ell}{\partial\gamma\,\partial\zeta} = (\eta-1)$$

$$\times\sum\frac{xe^{-\zeta x}\gamma^{-e^{-\zeta x}}\left[\left(\gamma^{1-e^{-\zeta z}}-1\right)\left[\left(1-e^{-\zeta z}\right)\log\gamma+1\right] - \left(1-e^{-\zeta z}\right)(\log\gamma)\gamma^{1-e^{-\zeta x}}\right]}{\left(\gamma^{1-e^{-\zeta x}}-1\right)^2}$$

$$+\sum\frac{xe^{-6x}}{\gamma}\frac{\partial^2\ell}{\partial\eta\,\partial\zeta} = \sum\frac{xe^{-\zeta x}(\log\gamma)\gamma^{1-e^{-\zeta x}}}{(\gamma^{1-e-\zeta x}-1)}$$

The asymptotic variance-covariance matrix of $\hat{\gamma}, \hat{\zeta}, \hat{\eta}$ is given as;

$$\text{var}-\text{cov}(\hat{\gamma},\hat{\zeta},\hat{\eta}) = [I(\gamma,\zeta,\eta)]^{-1}$$

where the variance of the MLEs $\hat{\gamma}, \hat{\zeta}$ and $\hat{\eta}$ are obtained as the elements of the leading diagonal of the variance-covariance matrix.

4. SIMULATION STUDY

In this section, we conducted a Monte Carlo simulation study to be able to examine the behavior ofthe MLEs for the parameters of EAPE distribution. Using the quantile function of the EAPE PDF defined in equation 10, random samples of sizes $n = (50,100,150,\ldots,500)$ from EAPE distribution were generated. The simulation was repeated $N = 1000$ times for each sample size and simulation results were obtained for different parameter combinations as set A: $\gamma = 1.6$, $\eta = 0.7$, $\zeta = 0.8$, and set B: $\gamma = 3.6$, $\eta = 0.72$, $\zeta = 1.07$. The study utilized nlminb() function in R for the purposes of numerical evaluation of EAPE MLES performance with the "BFGS" as the argument method. Further, we calculated the average bias (AB) and the root mean square error (RMSE) of the MLEs using the following formulas respectively and were then examined.

$$\mathrm{AB}(\hat{\Theta}_i) = \frac{1}{N}\sum_{i=1}^{N} (\hat{\Theta}_i - \Theta),$$

and

$$\mathrm{RMSE}(\hat{\Theta}_i) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (\hat{\Theta}_i - \Theta)^2}.$$

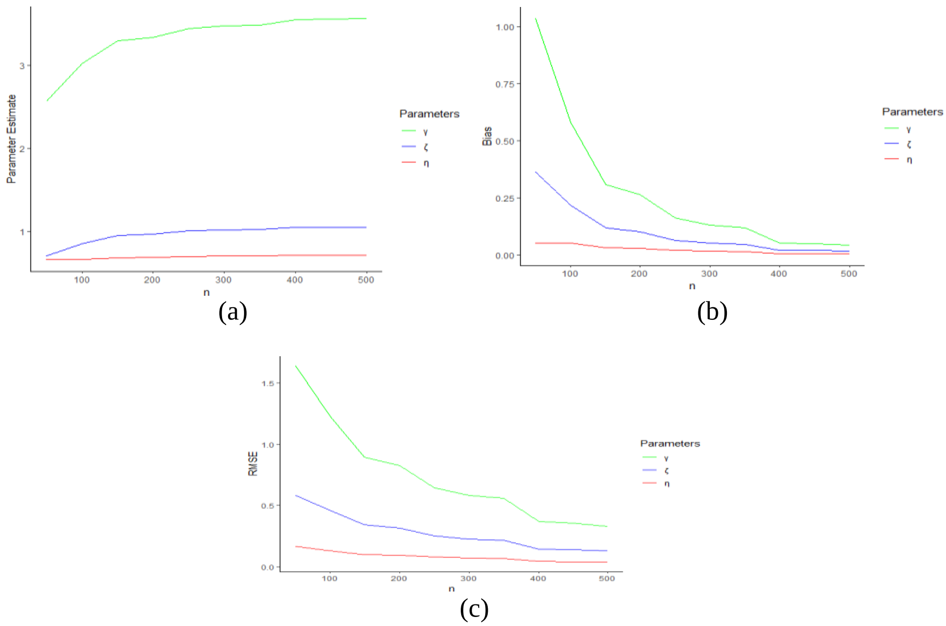
where, $\Theta = (\gamma, \eta, \zeta)$

Table 3
Estimated MLE Values, Average Bias and Root Mean Square Errors for the EAPE Distribution for Set A: $\gamma = 1.4, \zeta = 0.8, \eta = 0.7$

n	Parameter Estimate			Average Bias			RMSE		
	$\hat{\gamma}$	$\hat{\zeta}$	$\hat{\eta}$	$\hat{\gamma}$	$\hat{\zeta}$	$\hat{\eta}$	$\hat{\gamma}$	$\hat{\zeta}$	$\hat{\eta}$
50	1.295	0.62	0.659	-0.105	-0.180	-0.041	0.205	0.355	0.125
100	1.345	0.698	0.666	-0.055	-0.102	-0.034	0.149	0.274	0.102
150	1.366	0.737	0.676	-0.034	-0.063	-0.024	0.117	0.218	0.085
200	1.384	0.769	0.689	-0.016	-0.031	-0.011	0.081	0.152	0.059
250	1.383	0.769	0.688	-0.017	-0.031	-0.012	0.082	0.154	0.059
300	1.388	0.778	0.691	-0.012	-0.022	-0.009	0.068	0.129	0.052
350	1.392	0.784	0.693	-0.008	-0.016	-0.007	0.058	0.110	0.046
400	1.394	0.789	0.695	-0.006	-0.011	-0.005	0.049	0.093	0.040
450	1.398	0.796	0.698	-0.002	-0.004	-0.002	0.028	0.054	0.0230
500	1.400	0.797	0.700	-0.002	-0.002	-0.001	0.019	0.022	0.016

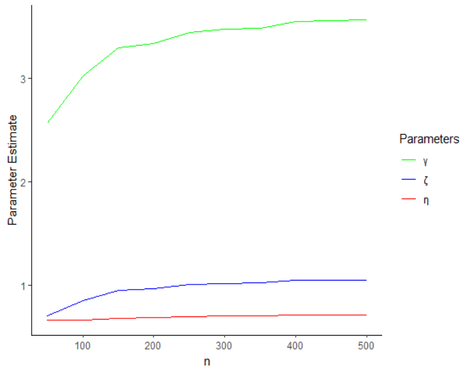
Table 4
Estimated MLE Values, Average Bias and Root Mean Square Errors for the EAPE Distributionfor Set B: $\gamma = 3.6, \zeta = 1.07, \eta = 0.72$

n	Parameter Estimate			Average Bias			RMSE		
	$\hat{\gamma}$	$\hat{\zeta}$	$\hat{\eta}$	$\hat{\gamma}$	$\hat{\zeta}$	$\hat{\eta}$	$\hat{\gamma}$	$\hat{\zeta}$	$\hat{\eta}$
50	2.568	0.708	0.670	-1.032	-0.362	-0.050	1.638	0.58	0.167
100	3.020	0.853	0.668	-0.580	-0.217	-0.052	1.228	0.460	0.129
150	3.293	0.953	0.688	-0.307	-0.117	-0.032	0.893	0.341	0.100
200	3.337	0.970	0.692	-0.263	-0.100	-0.028	0.826	0.316	0.092
250	3.439	1.007	0.701	-0.161	-0.063	-0.019	0.647	0.251	0.079
300	3.470	1.020	0.705	-0.130	-0.05	-0.015	0.581	0.225	0.070
350	3.480	1.024	0.706	-0.120	-0.046	-0.014	0.558	0.216	0.067
400	3.548	1.050	0.714	-0.052	-0.020	-0.006	0.368	0.143	0.044
450	3.551	1.051	0.714	-0.049	-0.019	-0.006	0.358	0.138	0.042
500	3.558	1.054	0.715	-0.042	-0.016	-0.005	0.329	0.128	0.040

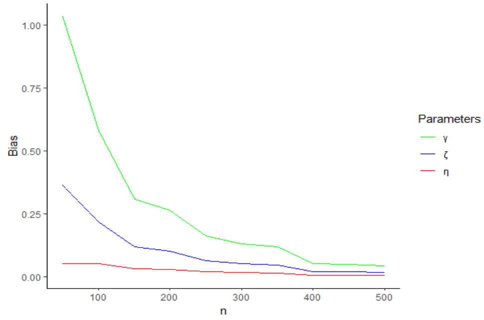


(a) (b) (c)

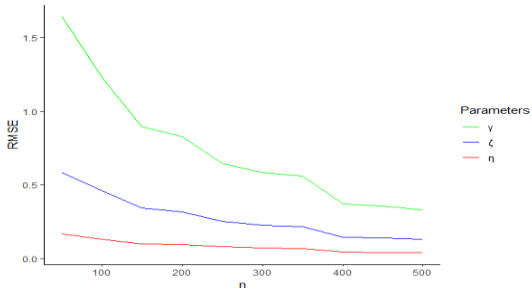
Figure 6: Graphical Representation of Simulation Results in Table 3 MLE (a), Absolute Bias (b) and RMSE (c)



(a)



(b)



(c)

Figure 7: Graphical Representation of Simulation Results in Table 4 MLE (a), Absolute Bias (b) and RMSE (c)

A summary of the findings from the simulation study is given in Table 3, and Table 4 while the graphical representation of these results is provided in Figures 6 and 7. Based on the results in the simulation tables and graphs, it is self-evident that the MLEs are effective in estimating unknown parameters and that the resulting estimates are relatively stable and close to the actual true values. Furthermore, as the sample sizes increases, the AB and RMSEs decrease and so do the associated absolute biases.

5. REAL WORLD DATA APPLICATIONS

For the purpose of illustrating flexibility as well as the importance of the EAPE distribution, two real-world survival data sets were considered. The study compared the goodness-of-fit test measures and also the information criterion measures of EAPE distribution with those of some well-known competing distributions. The test measure used is the Kolmogorov–Smirnov (K–S) test statistic withits corresponding p-value. While the criterion considered are the Akaike information criterion (AIC), and the Bayesian information criterion (BIC). The smaller the value of these statistics, the better the model fits the data.

Table 5
Kevlar 49/Epoxy Strands (with Pressure at 90%) Failure Times Data Set

0.01	0.01	0.02	0.02	0.02	0.03	0.03	0.04	0.05	0.06	0.07
0.07	0.08	0.09	0.09	0.10	0.10	0.11	0.11	0.12	0.13	0.18
0.19	0.20	0.23	0.24	0.24	0.29	0.34	0.35	0.36	0.38	0.40
0.42	0.43	0.52	0.54	0.56	0.60	0.60	0.63	0.65	0.67	0.68
0.72	0.72	0.72	0.73	0.79	0.79	0.80	0.80	0.83	0.85	0.90
0.92	0.95	0.99	1.00	1.01	1.02	1.03	1.05	1.10	1.10	1.11
1.15	1.18	1.20	1.29	1.31	1.33	1.34	1.40	1.43	1.45	1.50
1.51	1.52	1.53	1.54	1.54	1.55	1.58	1.60	1.63	1.64	1.80
1.80	1.81	2.02	2.05	2.14	2.17	2.33	3.03	3.03	3.34	4.20
4.69	7.89									

We begin with the exploration of the data and provide a summary of descriptive statistics in Table

The plot of total time on test (TTT), the histogram, and the box plot are also provided in Figure 8. The TTT plot helps researchers in identifying the nature of the hazard rate function (Aarset, 1987). As depictedin the figure, the TTT plot indicates that the first data set is characterized by a failure rate that is modified bathtub in shape. Also, the values of skewness and kurtosis given in Table 6 indicate thatthe data is positively skewed and is also leptokurtic. These features are both confirmed by the shapeof the histogram.

The maximum likelihood estimates for the Kevlar data were also determined and Table 7 gives the MLEs and their respective standard errors for the parameters of all the models considered. Accordingto the standard error test, at a given significance level, a parameter is said to be significant if its standard error is less than half the estimate. It is thus, self-evident that at the 5% significance level, most of the parameters of the fitted distributions were significant.

For model fit and comparison with some competing distributions, the negative log-likelihood values, goodness-of-fit test statistic measures, and values for the information criterion are provided in Table 8. As depicted from these values, the EAPE distribution has the lowest measures, hence, provides abetter fit than those of the considered competing distributions. Figure 9 displays the visual representation ofthe estimated probability density function (PDF) and estimated CDF for the competing models.

Table 6
Summary of Descriptive Statistics for Kevlar Data

Statistic	**Minimum**	**Mean**	**Median**	**Mode**	**Variance**	**Skewness**	**Kurtosis**	**Maximum**
Value	0.010	1.025	0.800	0.500	1.253	**3.002**	**13.709**	7.890

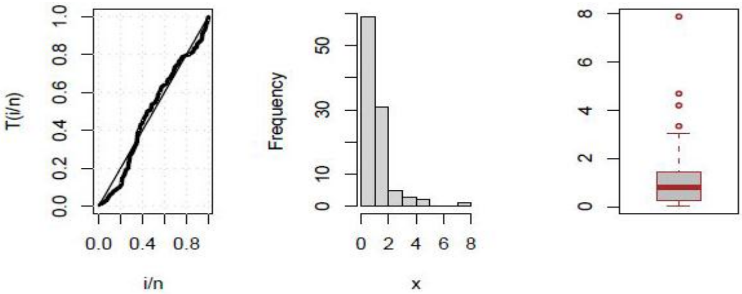
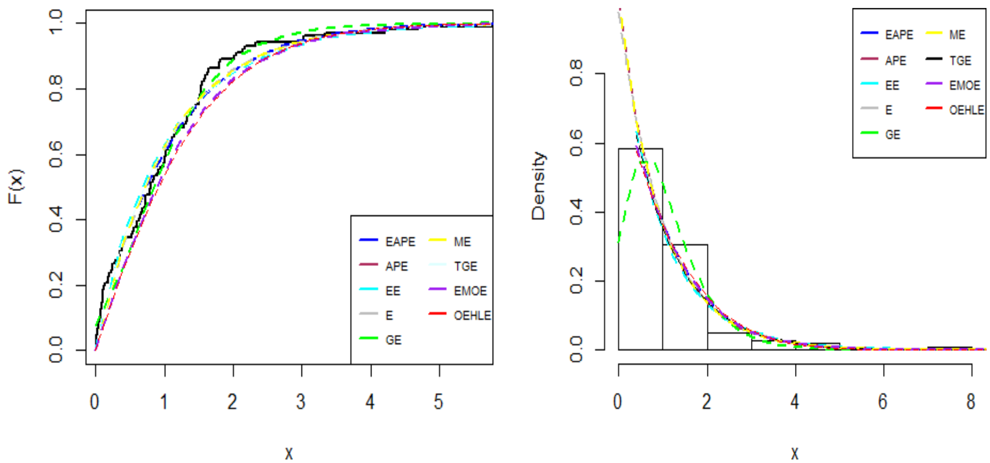


Figure 8: TTT plot, Histogram, and Box-Plot for Kevlar Data

Table 7
The MLEs and their Respective Standard Errors for Different Models for Kevlar data

Model	$\hat{\gamma}$	$\hat{\zeta}$	$\hat{\eta}$
EAPE	**3.627(1.490)**	**1.065(0.188)**	**0.719(0.167)**
APE	8.841(2.614)	0.934(0.201)	-
EE	-0.866(0.110)	0.888(0.120)	-
E	-	0.976(0.097)	-
GE	-0.820(0.040)	0.685(0.055)	7.639(1.087)
ME	2.892(0.067)	2.183(0.006)	0.931(0.201)
TGE	0.779(0.170)	0.955(0.136)	0.285(0.360)
EMOE	1.092(0.240)	1.843(1.099)	0.745(0.142)
OEHLE	0.716(0.086)	46.806(44.701)	0.023(0.021)



(a) (b)

Figure 9: Fitted PDFs (a), Empirical CDF, and Fitted CDFs (b) for Kevlar Data

Table 8
The -log-likelihood Values, Information Criterion, and Goodness-of-Fit Test Measures (and their corresponding p-values) for Kevlar Data

Model	ℓ	AIC	BIC	$K-S$	$p-value$
EAPE	**-102.260**	**210.420**	**212.365**	**0.070**	**0.705**
APE	-103.450	210.899	216.130	0.085	0.464
EE	-102.820	210.640	214.870	0.089	0.404
E	-103.479	218.959	219.574	0.089	0.403
GE	-120.471	246.943	254.788	0.107	0.200
ME	-103.447	212.893	220.738	0.084	0.472
TGE	-102.550	211.100	218.945	0.076	0.598
EMOE	-102.298	210.596	218.441	0.120	0.109
OEHLE	-102.664	211.328	219.174	0.125	0.087

Data Set 2: The second data set represents the number of daily deaths because of COVID-19 in China between January 23 to March 28 2020 obtained from https://www.worldometers.info/coronavirus/country/c

A summary of descriptive statistics is given in Table 9 and the value of the kurtosis suggests that it is platykurtic while skewness indicates right-skewed data. The plot of total time on test (TTT), the histogram, and the box plot are also provided in Figure 10. As depicted in the figure, the TTT plotindicates that the second data set is characterized by a failure rate that is modified bathtub in shape. The maximum likelihood estimates for the China Covid-19 data were also determined and Table 10 gives the MLEs and their respective standard errors for the parameters of all the models considered.

It is self-evident that at the 5% significance level, most of the parameters of the fitted distributions were significant.

For model fit and comparison with some competing distributions, the negative log-likelihood values, goodness-of-fit test statistic measures, and values for the information criterion are provided in Table 11. As depicted from these values, the EAPE distribution has the lowest measures, hence, provides abetter fit than those of the considered competing distributions. Figure 11 displays the visual representationof the estimated probability density function (PDF) and estimated CDF for the competing models.

Table 9
Summary Statistics for China Covid-19 Data

Statistic	Minimum	Mean	Median	Mode	Variance	Skewness	Kurtosis	Maximum
Value	0.320	1.675	1.470	1.500	1.001	**1.087**	**1.207**	4.750

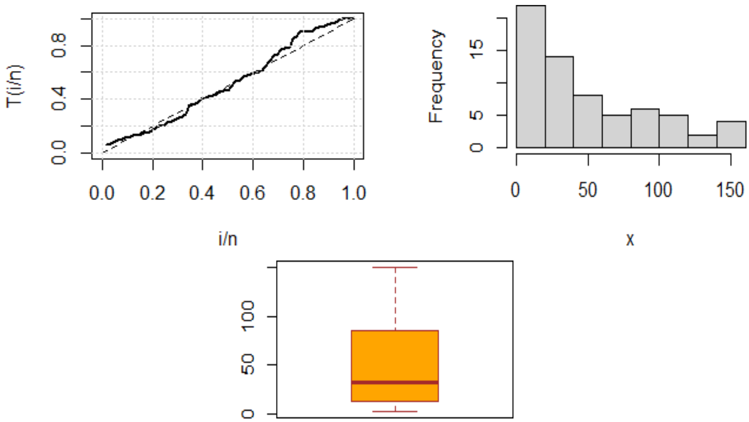


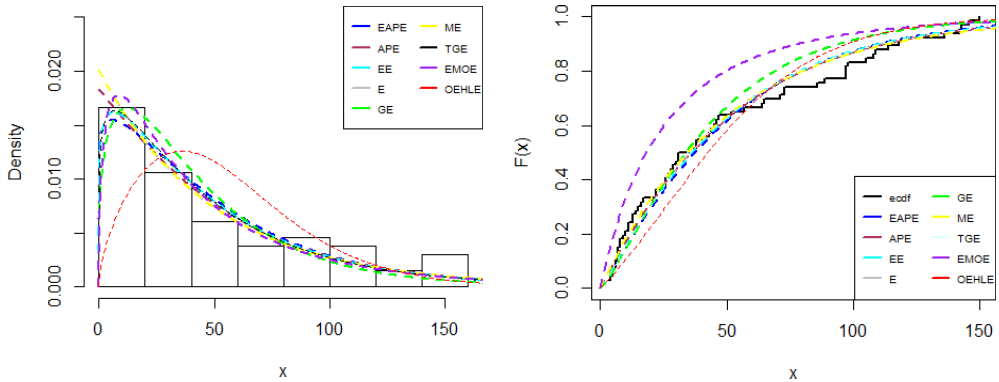
Figure 10: TTT Plot, Histogram, and Box-Plot for China Covid-19 Data

Table 10
The MLEs and their Respective Standard Errors for Different Models for China Covid-19 Data

Model	$\hat{\gamma}$	$\hat{\zeta}$	$\hat{\eta}$
EAPE	$\mathbf{1.500}(\mathbf{0.562})$	$\mathbf{0.023}(\mathbf{0.003})$	$\mathbf{1.086}(\mathbf{0.106})$
APE	$1.403(1.040)$	$0.022(0.004)$	$-$
EE	$-$	$1.148(0.189)$	$0.022(0.003)$
E	$-$	$0.020(0.002)$	$-$
GE	$-0.662(0.024)$	$39.985(3.769)$	$1.429(0.217)$
ME	$4.591(47.170)$	$4.645(34.170)$	$0.020(0.002)$
TGE	$01.168(0.217)$	$0.022(0.004)$	$0.066(0.378)$
EMOE	$0.019(0.006)$	$0.564(0.539)$	$1.438(0.619)$
EOHLE	$1.57(0.314)$	$13.584(2.114)$	$0.002(0.000)$

Table 11
The -Log-Likelihood Values, Information Criterion, and Goodness-of-Fit Test Measures (and their corresponding p-values) for China Covid-19 Data

Model	ℓ	**AIC**	**BIC**	$K-S$	$p-value$
EAPE	**-322.684**	**650.367**	**652.936**	**0.079**	**0.789**
APE	-323.760	651.521	655.900	0.088	0.691
EE	-323.513	651.027	655.406	0.092	0.634
E	-333.853	659.705	661.895	0.095	0.525
GE	-323.113	652.226	658.790	0.159	0.071
ME	-323.854	653.707	660.276	0.085	0.723
TGEE	-323.497	652.995	659.564	0.091	0.631
EMOE	-323.320	652.639	659.208	0.232	0.002
EOHLE	-324.429	654.859	661.428	0.126	0.242



(a) (b)

Figure 11: Fitted PDFs (a), Empirical CDF and Fitted CDFs (b) for China Covid-19 Data

6. CONCLUSION

This study has developed and studied the EAPE distribution, which extends the alpha power exponential distribution, and successfully derived its basic statistical properties. The properties derived include; the probability distribution function, cumulative distribution function, survival function, hazard rate function, quantile function, moments and moment generating function, and order statistics. The method of maximum likelihood estimation approach has been applied to estimate the parameters of EAPE distribution, and the average bias together with the RMSE has been used to assessthe performance of the MLEs. The RMSE was in agreement with the asymptotic theory, while the AB approached zero as the sample size increased. We can, thus, conclude that the maximum likelihood estimators for the parameters of EAPE distribution are consistent. Through an application to real survival data, this study has shown that the new distribution did provide a good fit. Notably, the data set was right skewed and leptokurtic and was characterized by a modified bathtub hazard rate. In addition, the proposed distribution was compared with some competing distributions including the exponentiated Weibull, the exponentiated gull alpha power exponential, the exponentiated generalized alpha power exponential distribution, and the Kumaraswamy generalized exponentiated exponential distribution. The proposed EAPE distribution outperformed the well-known consideredcompeting distributions. In conclusion, the proposed model can be used as an alternative distributionto its competing distributions for modeling survival data exhibiting different shapes of the hazard ratefunction.

REFERENCES

1. Aarset, M.V. (1987). How to Identify a Bathtub Hazard Rate. *IEEE Transactions on Reliability*, R-36(1), 106-108.
2. Afify, A.Z. and Mohamed, O.A. (2020). A new three-parameter exponential distribution with variable shapes for the hazard rate: Estimation and applications. *Mathematics*, 8(1), 135.

3. Afify, A.Z., Gemeay, A.M. and Ibrahim, N.A. (2020). The heavy-tailed exponential distribution: risk measures, estimation, and application to actuarial data. *Mathematics*, 8(8), 1276.
4. Al-Aqtash, R., Lee, C. and Famoye, F. (2014). Gumbel-weibull distribution: Properties and applications. *Journal of Modern Applied Statistical Methods*, 13(2), 201-225.
5. Aldahlan, M.A. (2020). Alpha Power Transformed Log-Logistic Distribution with Application to Breaking Stress Data. *Advances in Mathematical Physics*, Article ID 2193787, https://doi.org/10.1155/2020/2193787
6. Aldahlan, M.A.D. and Afify, A.Z. (2020). The odd exponentiated half-logistic exponential distribution: Estimation methods and application to engineering data. *Mathematics*, 8(10), 1-26.
7. Alshanbari, H.M., Gemeay, A.M., Hosni El-Bagoury, A.A.A., Khosa, S.K., Hafez, E.H. and Muse, A.H. (2022). A novel extension of Fréchet distribution: Application on real data and simulation. *Alexandria Engineering Journal*, 61(10), 7917-7938.
8. Alzaatreh, A., Lee, C. and Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, 71(1), 63-79.
9. Aslam, M., Hussain, Z. and Asghar, Z. (2019). Modified generalized marshall - olkin family of distributions. *International Journal of Advanced Statistics and Probability*, 7(1), 18-27.
10. Azzalini, A. and Valle, A.D. (1996). The multivariate skew-normal distribution. *Biometrika*, 83(4), 715-726.
11. Broyden, C.G. (1970). A new method of solving nonlinear simultaneous equations. *Computational Journal*, 12, 94-99.
12. Cordeiro, G.M. and de Castro, M. (2011). A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, 81(7), 883-898.
13. ElSherpieny, E.S.A. and Almetwally, E.M. (2022). The Exponentiated Generalized Alpha Power Exponential Distribution: Properties and Applications. *Pakistan Journal of Statistics and Operation Research*, 18(2), 349-367.
14. Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics-Theory and methods*, 31(4), 497-512.
15. Fletcher, R. (1970). A class of methods for nonlinear programming with termination and convergence properties. *Integer and Nonlinear Programming*, pp. 157-173, Amsterdam,
16. Goldfarb, D. (1970). A Family of Variable-Metric Methods Derived by Variational Means. *Mathematics of Computation*, 24(109), 23.
17. Ijaz, M., Mashwani, W.K. and Belhaouari, S.B. (2020). A novel family of lifetime distribution with applications to real and simulated data. *PLoS ONE*, 15(10 October), 1-15.
18. Jones, J.H. (2007). for the for the. *Scientific American*, *57*(October), 613-658.
19. Khan, M.S., King, R. and Hudson, I.L. (2017). Transmuted generalized exponential distribution: A generalization of the exponential distribution with applications to survival data. *Communications in Statistics: Simulation and Computation*, 46(6), 4377-4398.
20. Kilai, M., Waititu, G.A., Kibira, W.A., Alshanbari, H.M. and El-Morshedy, M. (2022). A new generalization of Gull Alpha Power Family of distributions with application to modeling COVID-19 mortality rates. *Results in Physics*, *36*(February), 105339. https://doi.org/10.1016/j.rinp.2022.105339

21. Kilai, M., Waititu, G.A., Kibira, W.A., El-raouf, M.M.A. and Abushal, T.A. (2022). Results in Physics A new versatile modification of the Rayleigh distribution for modeling COVID-19 mortality rates. *Results in Physics, 35*(January), 105260.
22. Ma, Y. and Genton, M.G. (2004). *Flexible Class of Skew-Symmetric Distributions*. 31(1), 459-468.
23. Mahdavi, A. and Kundu, D. (2017). A new method for generating distributions with an application to exponential distribution. *Communications in Statistics - Theory and Methods*, 46(13), 6543-6557.
24. Mead, M.E., Cordeiro, G.M., Afify, A.Z. and Al-Mofleh, H. (2019). The alpha power transformation family: Properties and applications. *Pakistan Journal of Statistics and Operation Research*, 15(3), 525-545.
25. Modi, K., Kumar, D. and Singh, Y. (2020). A new family of distribution with application on two real datasets on survival problem. *Science and Technology Asia*, 25(1), 1-10.
26. Mudholkar, G.S. and Srivastava, D.K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability*, 42(2), 299-302.
27. Nadarajah, S. and Kotz, S. (2006a). The beta exponential distribution. *Reliability Engineering and System Safety*, 91(6), 689-697.
28. Nadarajah, S. and Kotz, S. (2006b). The exponentiated type distributions. *Acta Applicandae Mathematicae*, 92(2), 97-111.
29. Nasiru, S., Mwita, P.N. and Ngesa, O. (2017). Exponentiated Generalized Family of Distributions. *Journal of Statistical and Econometric Methods*, 6(4), 1-17.
30. Nasiru, S., Mwita, P.N. and Ngesa, O. (2019). Alpha power transformed Fréchet distribution. *Applied Mathematics and Information Sciences*, 13(1), 129-141.
31. Nassar, M., Afify, A.Z. and Shakhatreh, M.K. (2020). Estimation methods of alpha power exponential distribution with applications to engineering and medical data. *Pakistan Journal of Statistics and Operation Research*, 16(1), 149-166.
32. Nassar, M., Kumar, D., Dey, S., Cordeiro, G.M. and Afify, A.Z. (2019). The Marshall-Olkin alpha power family of distributions with applications. *Journal of Computational and Applied Mathematics*, 351, 41-53.
33. Oluyede, B.O., Bindele, H.F., Makubate, B. and Huang, S. (2018). A new generalized log-logistic and modified Weibull distribution with applications. *International Journal of Statistics and Probability*, 7(3), 72-93.
34. Pasari, S. and Dikshit, O. (2014). Three-parameter generalized exponential distribution in earthquake recurrence interval estimation. *Natural Hazards*, 73(2), 639-656.
35. Piriaei, H., Yari, G. and Farnoosh, R. (2020). E-Bayesian estimations for the cumulative hazard rate and mean residual life based on exponential distribution and record data. *Journal of Statistical Computation and Simulation*, 90(2), 271-290.
36. Shanno, D.F. (1970). Conditioning of Quasi-Newton Methods for Function Minimization. *Mathematics of Computation*, 24(111), 647.
37. Tharu, R.P., Pahari, S., Sedhai, G.P. and Dhungana, G.P. (2021). Exponentiated Marshall-Olkin Exponential Distribution: Application of COVID-19 Second Wave in Nepal. *Nepal Journal of Mathematical Sciences*, 2(2), 43-56.
38. Zhao, J., Ahmad, Z., Mahmoudi, E., Hafez, E.H. and Mohie El-Din, M.M. (2021). A new class of heavy-tailed distributions: modeling and simulating actuarial measures. *Complexity*, 2021(iii), 1-18.