JOURNAL OF COMPUTATIONAL INNOVATION AND ANALYTICS
https://e-journal.uum.edu.my/index.php/jcia

How to cite this article:
Nyandaiti, N. W., Mohammed, Y. A., & Esbond, G. I. (2025). A novel probability distribution with robust parametric properties. *Journal of Computational Innovation and Analytics*, 4(2), 100-109. https://doi.org/10.32890/jcia2025.4.2.7

A NOVEL PROBABILITY DISTRIBUTION WITH ROBUST PARAMETRIC PROPERTIES

[1]Na'awurti William Nyandaiti, [2]Yusuf Abbakar Mohammed & [3]Gongsin Isaac Esbond

[1]School of Health Information Management, University of Maiduguri Teaching Hospital, Nigeria
[2,3]Department of Statistics, University of Maiduguri, Nigeria

[1]Corresponding author: nnyandaiti@yahoo.com

ABSTRACT

The continuous evolution of statistical modelling has spurred the development of novel probability distributions to address complex data challenges. This paper introduces the Topp-Leone Epsilon (TopLE) distribution, derived by combining the Topp-Leone and epsilon distributions. The TopLE distribution exhibits attractive parametric properties, including flexible shape and scale parameters, a closed-form quantile function, and efficient parameter estimation using computational tools. Simulation studies demonstrate the distribution's efficiency, consistency, and stability, making it suitable for a wide range of applications in lifetime data analysis. Two comparative applications with four standard distributions have shown the TopLE distribution to be a competitive choice in modelling real-life datasets. This work contributes to the statistical literature by providing a versatile and competitive model for modern data complexities.

Keywords: compound, epsilon, generalized distributions, TopLE distribution, Topp-Leone

INTRODUCTION

The development of probability distributions has been pivotal in statistical modelling, particularly for lifetime datasets, as researchers continually seek improved models to address evolving data complexities. The study of lifetime probability distributions has a rich history, including foundational works such as Aalen et al. (2008), who provided an in-depth analysis of survival and event history data. Early models such as the Weibull distribution (Mudholkar & Srivastava, 1993; Gupta & Kundu, 1999) and its extensions, including the exponentiated Weibull (Mudholkar et al., 1995; Qian, 2012), have been widely applied due to their ability to adapt to patterns of many different systems and interpretability meaningfully. However, these models often fall short in capturing unique data characteristics, motivating the development of new distributions.

Traditional statistical distributions (standard distributions such as normal, Weibull, gamma, among others) often struggle to accurately model the complexities of data, which can exhibit non-linearity, heavy-tailing, and skewness. These limitations can lead to inadequate modelling, inaccurate prediction, and poor decision-making. Some existing distributions, such as the Weibull, lognormal, gamma, and logistic, among others, may not provide sufficient flexibility to capture the underlying patterns in the data. There is a need for a flexible and robust statistical distribution that can effectively model the complexities of data, account for censoring, and provide accurate predictions. While the Topp-Leone and epsilon distributions are independently useful as J-shaped probability distributions (Nadarajah & Kotz, 2003; Dombi et al., 2018), their combination as a compound distribution (probability distributions arising from a combination of two or more probability distributions) has not been explored. Developing a compound distribution could address limitations in existing models and provide a more robust tool for analyzing heavy-tail data and reliability systems.

LITERATURE REVIEW

Recent advancements in the field of probability modelling have introduced generalized families of distributions to enhance flexibility and fit. For instance, Kumaraswamy (1980) proposed a bounded probability density function, which has since been extended to various models (Cordeiro et al., 2010; Gongsin & Saporu, 2019). Alzaatreh et al. (2013) introduced a method for generating continuous distribution families, further broadening the scope of statistical modelling. Similarly, the Topp-Leone family of distributions, as explored by Al-Shomrani et al. (2016), has shown significant applicability in real-world datasets due to its unique properties, such as flexible shapes that describe their ability to model skewed data, symmetric data, and various hazard rate shapes. Extensions such as the Topp-Leone Weibull (Tuoyo et al., 2021) and Topp-Leone Lindley distributions (Nzei & Ekhosuehi, 2020) underscore the relevance of this family in modern probability theory.

The formulation of new probability distributions has often been driven by the need to address specific limitations of standard models. Ahmad et al. (2018) and Gongsin & Saporu (2021) emphasized that new distributions could outperform the traditional probability models by fitting complex datasets better, as evidenced by their work on the Weibull-X and Kumaraswamy-Epsilon distributions, respectively. Recent studies by Dombi et al. (2018) introduced the Epsilon distribution, an innovative model specifically designed for reliability analysis. Similarly, Adegoke et al. (2023) explored the Topp-Leone Inverse Gompertz distribution, providing insights into its statistical properties and applicability, further demonstrating the trend toward developing more competitive probability distributions.

The rise of computational tools has significantly facilitated the study of new distributions. Software such as R (Delignette-Muller & Dutang, 2015) enables efficient numerical computation of properties and model-fitting criteria, allowing for practical application of even the most complex models. This computational advancement has allowed researchers to explore novel families, such as the Beta-Normal (Eugene et al., 2002) and Exponentiated T-X (Alzaghal et al., 2013) distributions, broadening the statistical toolkit for lifetime data analysis.

Given the growing diversity of probability distributions, model selection criteria are crucial in identifying the most appropriate fit. Studies by Gupta et al. (1998) and Tahir & Cordeiro (2016) emphasized the importance of flexibility, efficiency, and interpretability in model development. The emergence of the Topp-Leone Epsilon (TopLE) distribution aligns with these trends, aiming to provide a competitive alternative to existing models. By building upon the theoretical foundations of distributions like the Topp-Leone (Jamal et al., 2019) while incorporating new design principles, and bringing in the epsilon distribution, the TopLE distribution seeks to address gaps in current modelling practices while leveraging advancements in computational capabilities.

This review highlights the continuous evolution of probability distributions and the critical need for innovative models that adapt to modern data challenges of variable shapes. The TopLE distribution represents a significant step in this direction, combining theoretical rigor with practical applicability to enhance the analysis of lifetime datasets.

METHODOLOGY

The Topp-Leone Generator of Probability Distributions

The Topp-Leone distributions were families of probability distributions proposed by Topp and Leone in 1955 (Nadarajah & Kotz, 2003). A variant form of the cumulative distribution function and probability density function, respectively, is given by:

$$F_T(t) = \begin{cases} t^a(2-t)^a & for\ 0 \le t \le 1 \\ 0 & otherwise \end{cases} \tag{1}$$

and

$$f_T(t) = 2at^{a-1}(1-t)(2-t)^{a-1}, 0 \le t < 1 \tag{2}$$

Where the values of t are bounded between 0 and 1, such as proportions, percentages, or probabilities.

The Topp-Leone distribution is an attractive generator of probability distributions, in the same class as the Beta and Kumaraswamy distributions (Kumaraswamy, 1980), since it is a continuous random variable. T is uniformly distributed over $(0,1)$, that is $T \sim U(0,1)$, and can be replaced by any arbitrary explicit cumulative distribution function, $G(x) \sim U(0,1)$. The Topp-Leone distribution was first introduced as a generator of probability distributions by Al-Shomrani et al. (2016), long after it was reintroduced by Nadarajah & Kotz (2003). Its form as a generator for a cumulative distribution function, $G(x)$, is given (Al-Shomrani et al., 2016) by:

$$F(x) = \left[1 - \left(\bar{G}(x)\right)^2\right]^a \tag{3}$$

where $\bar{G}(x) = 1 - G(x), x > 0$. The corresponding probability density function is given by Equation (4).

$$f(x) = 2ag(x)\bar{G}(x)\left[1 - \left(\bar{G}(x)\right)^2\right]^{\alpha-1} \quad (4)$$

Many probability distributions have been generated and applied in the literature based on (3) and (4). For example, Topp-Leone Exponential distribution (Al-Shomrani et al., 2016), Topp-Leone Weibull-Lomax distribution (Farrukh et al., 2019), Topp-Leone Lindley distribution (Nzei & Erhosuehi, 2020), Topp-Leone Gompertz distribution (Nzei et al., 2020), Topp-Leone Weibull distribution (Tuoyo et al., 2021), Topp-Leone Inverse Gompertz distribution (Adegoke et al., 2023), among others.

The Epsilon Distribution

The epsilon distribution was derived from the epsilon function by solving the first-order epsilon differential equation (Dombi et al., 2018). The novel distribution has two parameters, λ and δ, respectively specifying the shape and upper bound of the epsilon random variable. It is specified by the cumulative distribution function and probability density function given, respectively, by:

$$G_X(x) = 1 - \left(\frac{x+\delta}{\delta - x}\right)^{-\frac{\lambda\delta}{2}} \quad (5)$$

where $0 < x < \delta, \lambda > 0, \delta > 0$,

and

$$g_X(x) = \lambda \frac{\delta^2}{\delta^2 - x^2}\left(\frac{x+\delta}{\delta - x}\right)^{-\frac{\lambda\delta}{2}} \quad (6)$$

The existence of the threshold parameter, δ, which places an upper bound on the random variable, X, attracts the use of the epsilon distribution in generating new lifetime distributions since all systems have bounded lifetimes, for example, Kumaraswamy-epsilon distribution (Gongsin & Saporu, 2019), beta-epsilon distribution (Gongsin & Saporu, 2021), Weibull-epsilon distribution (Gongsin & Saporu, 2020), exponentiated-epsilon distribution (Gongsin & Saporu, 2019), among others.

The TopLE Distribution

This subsection derives and presents the Topp-Leone Epsilon (TopLE) probability density, cumulative distribution, and quantile functions.

The TopLE Probability Density, Cumulative Distribution, Survival, and Hazard Functions

Substituting the Epsilon cumulative distribution function (5) and probability density function (6) into the Topp-Leone probability density generator function (4), we obtained the TopLE probability function given by:

$$f_X(x) = 2\alpha\lambda \frac{\delta^2}{\delta^2 - x^2}\left(\frac{x+\delta}{\delta - x}\right)^{-\lambda\delta}\left[1 - \left(\frac{x+\delta}{\delta - x}\right)^{-\lambda\delta}\right]^{\alpha-1} \quad (7)$$

Substitute Equation (5) into the Topp-Leone CDF generator (3), the TopLE cumulative distribution function (CDF) was obtained as:

$$F_X(x) = \left[1 - \left(\frac{x+\delta}{\delta - x}\right)^{-\lambda\delta}\right]^{\alpha} \tag{8}$$

Theorem 1

Given that $X \sim TopLE(\alpha, \lambda, \delta)$, then,

$$\int_0^{\delta} f_X(x)\, dx \ = \ 1$$

Proof

It suffices to show that the TopLE probability density function is a true probability density function. That is,

$$\int_0^{\delta} f_X(x)\, dx = \int_0^{\delta} 2\alpha\lambda \frac{\delta^2}{\delta^2 - x^2}\left(\frac{x+\delta}{\delta - x}\right)^{-\lambda\delta}\left[1 - \left(\frac{x+\delta}{\delta - x}\right)^{-\lambda\delta}\right]^{\alpha-1} dx$$

$$= \lim_{x\to\delta}\int_0^{x} 2\alpha\lambda \frac{\delta^2}{\delta^2 - t^2}\left(\frac{t+\delta}{\delta - t}\right)^{-\lambda\delta}\left[1 - \left(\frac{t+\delta}{\delta - t}\right)^{-\lambda\delta}\right]^{\alpha-1} dt$$

$$= \left[1 - \left(\frac{x+\delta}{\delta - x}\right)^{-\lambda\delta}\right]^{\alpha}\Bigg|_{x=\delta}$$

$$= \left[1 - \left(\frac{\delta+\delta}{\delta - \delta}\right)^{-\lambda\delta}\right]^{\alpha}$$

$$= \left[1 - \left(\frac{2\delta}{0}\right)^{-\lambda\delta}\right]^{\alpha}$$

$$= \left[1 - \left(\frac{0}{2\delta}\right)^{\lambda\delta}\right]^{\alpha}$$

$$= [1 - 0]^{\alpha}$$

$$= 1 \qquad \blacksquare$$

Plots of the TopLE probability density function at varying parameter values are given in Figures 1 and 2.

Figure 1

TopLE Probability Density Plots at varying λ Parameter Values

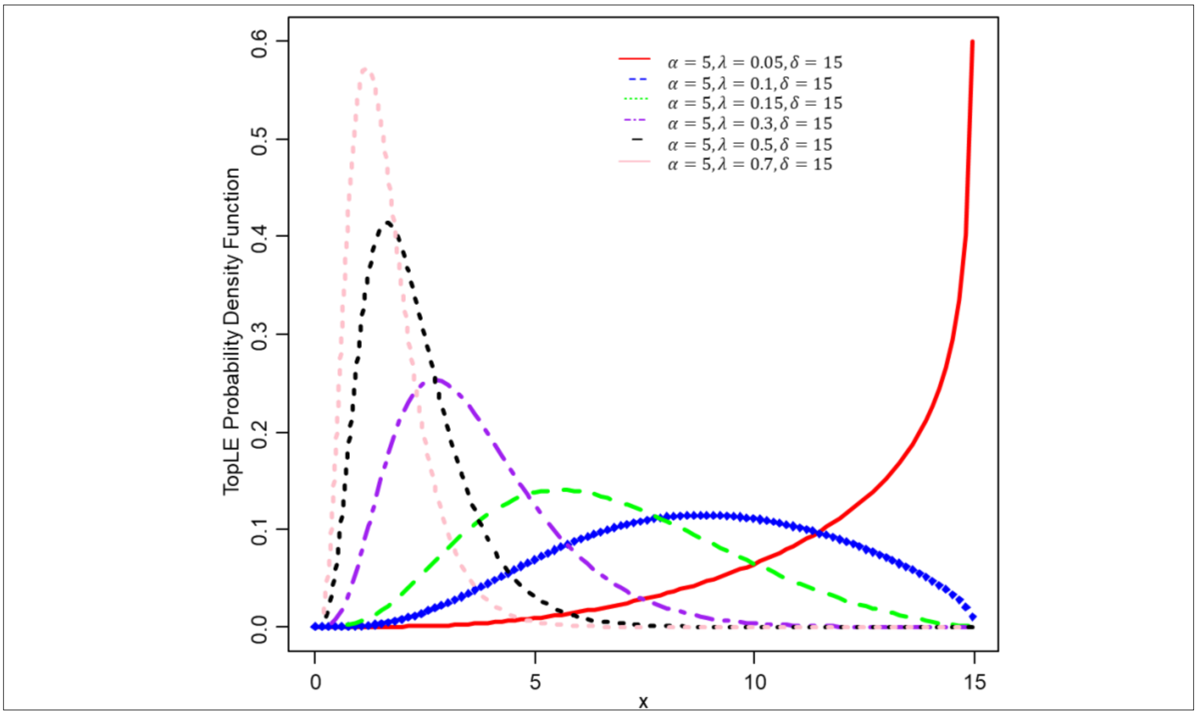
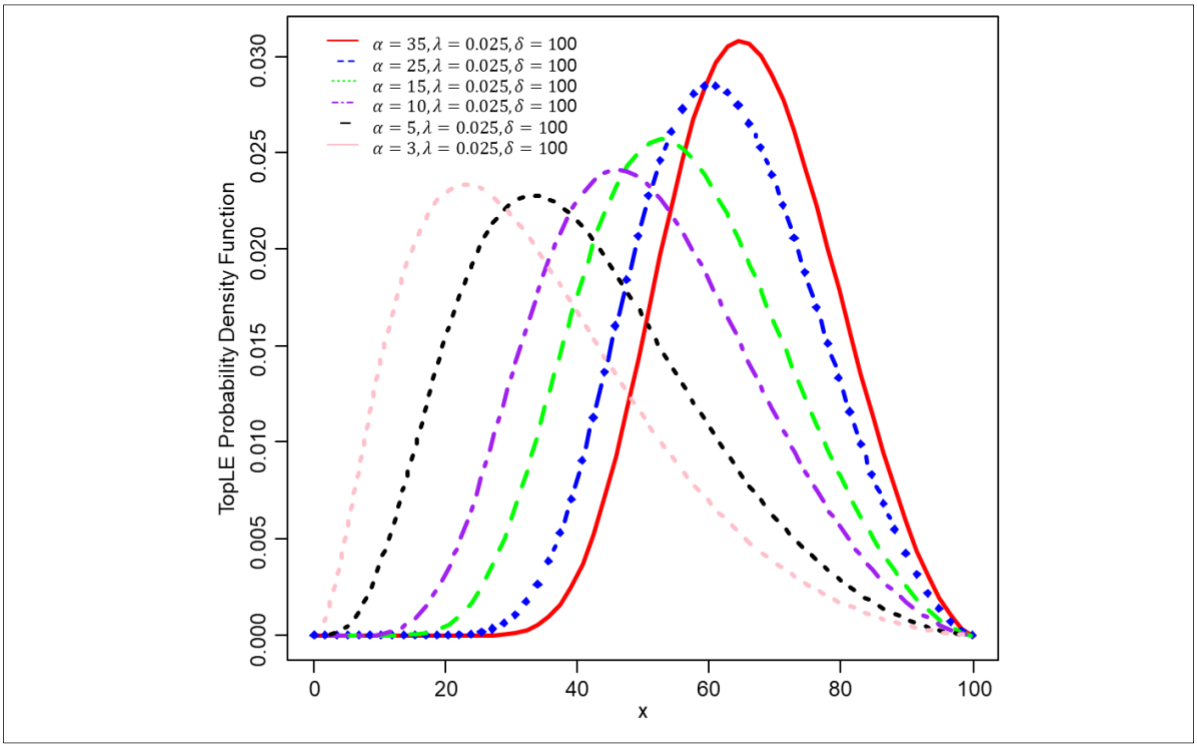


Figure 2

TopLE Probability Density Plots at varying α Parameter Values



From Figure 1 and Figure 2, it can be seen that. λ and α play the role of the shape and scale parameters. The survival function of the TopLE distribution is given by:

$$S_X(x) = 1 - \left[1 - \left(\frac{x+\delta}{\delta - x}\right)^{-\lambda\delta}\right]^{\alpha} \tag{9}$$

The TopLE distribution exhibits a variate generation property as its quantile function was obtained in a closed form. It is given by:

$$Q_X(u) = \delta \frac{\left(1 - u^{\frac{1}{\alpha}}\right)^{-\frac{1}{\lambda\delta}} - 1}{\left(1 - u^{\frac{1}{\alpha}}\right)^{-\frac{1}{\lambda\delta}} + 1} \tag{10}$$

where $\alpha > 0$, $\lambda > 0$, $\delta > 0$, and $u \sim U(0,1)$.

Distribution of Order Statistics of the TopLE Distribution

Let $X_1, X_2, \cdots, X_n$ be random variables from the TopLE distribution. Let $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ be the ordered X_i's so that $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$, then the distribution of the r^{th} order statistic, $X_{(r)}$, is given (David & Nagaraja, 2003) by:

$$\mathrm{f}(x_{(r)}) = \frac{2\alpha\lambda n!}{(r-1)!\,(n-r)!}\frac{\delta^2}{\delta^2 - x^2}\sum_{i=1}^{n-r}\sum_{j=1}^{\alpha(i+r)-1}(-1)^{i+j}\left(\frac{x+\delta}{\delta - x}\right)^{-\lambda\delta(j+1)} \tag{11}$$

From Equation (11), the probability density function of the first order statistic ($r = 1$) of the TopLE random variable, X, is given by:

$$\mathrm{f}_{X_{(1)}}(x) = 2\alpha\lambda n \frac{\delta^2}{\delta^2 - x^2}\sum_{i=1}^{n-1}\sum_{j=1}^{\alpha(i+1)-1}(-1)^{i+j}\left(\frac{x+\delta}{\delta - x}\right)^{-\lambda\delta(j+1)} \tag{12}$$

If $r = n$, then the distribution of the n^{th} order statistic of a random variable, X, characterized by the TopLE distribution is given by:

$$\mathrm{f}_{X_{(n)}}(x) = 2\alpha\lambda n \frac{\delta^2}{\delta^2 - x^2}\sum_{j=1}^{\alpha n-1}(-1)^{j}\left(\frac{x+\delta}{\delta - x}\right)^{-\lambda\delta(j+1)} \tag{13}$$

Parameter Estimation using the fitdistrplus Package

To use the *fitdistrplus* package, the probability density, cumulative distribution, and quantile functions of the TopLE distribution were defined in the R environment as *dtople, ptople,* and *qtople*, respectively. The precedence *d, p,* and *q,* stand for the probability density function, cumulative distribution function and quantile function, respectively. Then the *fitdist( )* function from the *fitdistrplus* package was called via the string code "*tople"*, corresponding to the root name used to specify *dtople, ptople,* and *qtople*.

The syntax used to estimate the parameters of the TopLE distribution based on the default maximum likelihood method was:

$$ftople = fitdist\left(data,'\,tople', start = list(\alpha = \alpha_0, \lambda = \lambda_0, \delta = \delta_0)\right)$$

The initial values, α_0, λ_0 and δ_0, for the distribution parameters were supplied in the argument start, as a named list with initial values for each parameter as they appeared in the *p, d, q* functions. Having supplied the reasonable starting values, best determined by comparison of the histogram of the data with the shape of the TopLE probability density plots in Figure 1 and 2, the *fitdist( )* function was run to obtain the estimates of the parameters of the TopLE distribution for given data values denoted *"data"* in the function.

Simulation

Monte Carlo simulations were carried out using the quantile function (10). For fixed parameter values of the TopLE distribution, the algorithm for the simulation (Gentle, 2005) is given by:

(i) generate p from $U(0, 1)$,
(ii) put p in the quantile function for fixed values of the parameters of the distribution,
(iii) return $x = Q(p)$.

The simulations were replicated 1000 times for varying α, λ and δ and for n = 25, 50, 70, 100, 120, 150, 200, 250, 300, 500, 750, 1000, and 1500. The measures of standard error (se) were computed by:

$$se = \sqrt{\frac{1}{m(m-1)}\sum_{i=1}^{m}\left(\hat{\theta}_i - \bar{\theta}\right)^2} \tag{14}$$

where m is the number of replications, $\hat{\theta}_i$ is the i^{th} replicate the parameter estimate and $\bar{\theta}$ is the mean value of the parameter estimates from the m replicates. The bias of the parameter estimates was computed by:

$$Bias = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{\hat{\theta}_i - \theta}{\theta}\right) \tag{15}$$

ANALYSIS AND RESULTS

Simulation Results

Simulation of the TopLE distribution was conducted using the quantile function given in Equation (10) to determine the efficiency, consistency, and stability of parameter estimation of the TopLE distribution. Two simulations for variants of parameters were carried out, and the parameter estimates were obtained using the *optim* package in R to optimize the TopLE log-likelihood function. 1000 replications were carried out for each sample size given above. The simulation results are represented in Figure 3 and Figure 4.

Table 1

Simulation Results of TopLE Distribution for $\alpha = 5, \lambda = 0.15$ *and* $\delta = 15$

n	$\alpha = 5$ $\hat{\alpha}$ (se)	$bias(\hat{\alpha})$	$\lambda = 0.15$ $\hat{\lambda}$ (se)	$bias(\hat{\lambda})$	$\delta = 15$ $\hat{\delta}$ (se)	$bias(\hat{\delta})$
25	4.52 (1.52)	−0.096	0.146 (0.031)	−0.027	14.43 (2.19)	−0.038
50	5.48 (1.61)	0.096	0.148 (0.024)	−0.013	13.96 (1.11)	−0.069
70	5.13 (1.44)	0.026	0.147 (0.024)	−0.020	14.63 (2.04)	−0.025
100	5.06 (1.14)	0.012	0.148 (0.019)	−0.013	14.65 (1.43)	−0.023
120	5.00 (1.01)	0.000	0.147 (0.017)	−0.020	14.64 (1.21)	−0.024
150	4.97 (0.88)	−0.006	0.147 (0.015)	−0.020	14.64 (0.99)	−0.024
200	4.97 (0.75)	−0.006	0.148 (0.013)	−0.013	14.67 (0.83)	−0.022
250	4.99 (0.67)	−0.002	0.148 (0.011)	−0.013	14.75 (0.73)	−0.017
300	4.99 (0.61)	−0.002	0.149 (0.010)	−0.007	14.76 (0.64)	−0.016
500	5.01 (0.47)	0.002	0.149 (0.008)	−0.007	14.86(0.48)	−0.009
750	4.99 (0.37)	−0.002	0.149 (0.006)	−0.007	14.89 (0.38)	−0.007
1000	4.98 (0.32)	−0.004	0.149 (0.005)	−0.007	14.91 (0.32)	−0.006
1500	4.97 (0.26)	−0.006	0.149 (0.004)	−0.007	14.91 (0.25)	−0.006

Here, efficiency means the ability of the model to estimate parameters with minimal bias and minor standard errors. From Table 1, it can be observed that as n increases, the standard errors (se) for each of the parameters tend to decrease. This suggests that the model's parameter estimates become more efficient with larger sample sizes, which is typical in statistical estimation: larger samples provide more information, leading to more precise estimates. For example, for $\hat{\alpha}$, the standard error decreased from 1.52 when $n = 25$ to 0.26 when $n = 1500$; similarly, the standard error for $\hat{\lambda}$ decreased from 0.031 when n = 25 to 0.004 when $n = 1500$; finally, for $\hat{\delta}$, the standard error dropped from 2.19 to 0.25 over the same range of sample sizes. These observations indicate that the TopLE distribution's parameter estimates become more precise with larger sample sizes, demonstrating its efficiency.

For consistency, we refer to how well the TopLE distribution parameter estimates converge to the actual parameter values as the sample size increases. In this case, the estimates for α, λ, and δ are fairly stable as n increases, suggesting that the model is consistent. These are seen in Figure 2a below. The plots in Figure 2a indicate that the TopLE distribution is consistent, with the estimates for each parameter converging to their actual values as n increases.

Stability can be interpreted as the model's ability to provide estimates close to their original values as n increases, without significant variability in the estimates. The simulation results reveal that the estimates for each parameter, alpha, lambda, and delta, stabilize as the sample size n increases. It can also be observed that the 95% confidence intervals (CIs) for each parameter become narrower as n increases (Figure 2*a*), reflecting the distribution's increasing stability.

The representation in Figure 3 of the simulation results in Table 1 summarizes the new model's exhibition of high efficiency**,** consistency**, and** stability**.** The precision of the parameter estimates improves as n increases, the estimates converge to their respective actual values, and the model's performance becomes increasingly stable, as evidenced by the shrinking standard errors, bias, and confidence intervals. These factors indicate that the TopLE distribution is robust and reliable for statistical modelling.

Figure 3

Simulation Results of the TopLE Distribution for $\alpha = 5, \lambda = 0.15$ *and* $\delta = 15$

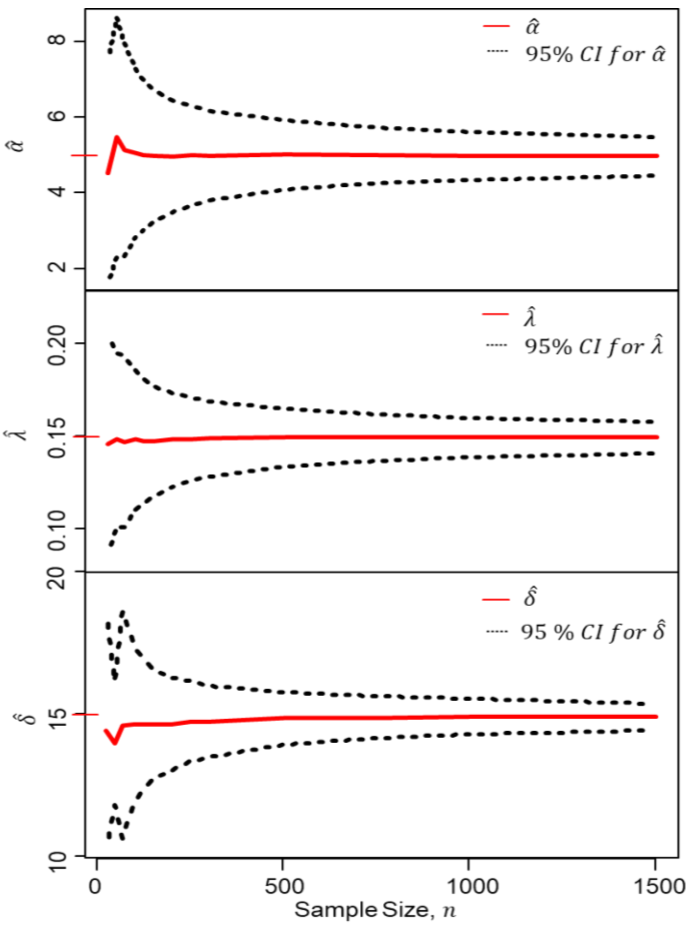
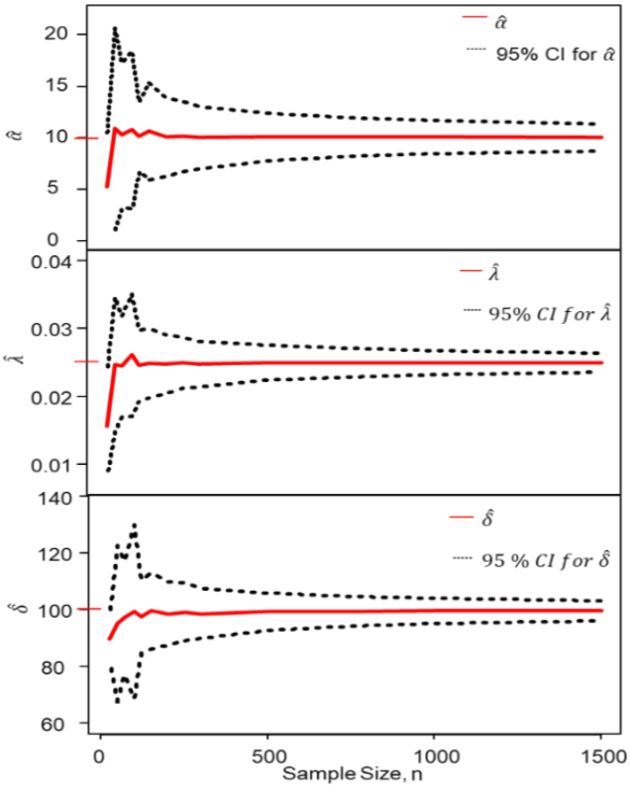


Figure 4

Simulation Results of the TopLE Distribution for $\alpha = 10, \lambda = 0.025$ *and* $\delta = 100$



Again, in Table 2 below, we demonstrated the efficiency**,** consistency, and stability of the TopLE distribution for different sample sizes (n) with actual values of $\alpha = 10$, $\lambda = 0.025$ and $\delta = 100$. From the table, as n increases, the standard errors and biases of $\hat{\alpha}$, $\hat{\lambda}$ and $\hat{\delta}$ decrease, which is characteristic of increased estimation efficiency. These decreasing standard errors and biases indicate that the model becomes more efficient as n increases, offering more precise estimates with larger datasets.

Table 2

Simulation Results of the TopLE Distribution for $\alpha = 10, \lambda = 0.025$ *and* $\delta = 100$

n	$\alpha = 10$ $\hat{\alpha}$ (se)	$bias(\hat{\alpha})$	$\lambda = 0.025$ $\hat{\lambda}$(se)	$bias(\hat{\lambda})$	$\delta = 100$ $\hat{\delta}$ (se)	$bias(\hat{\delta})$
25	5.27 (2.37)	−0.473	0.016 (0.004)	−0.360	89.85 (3.45)	−0.102
50	10.89 (4.98)	0.089	0.025 (0.005)	0.000	95.29 (13.94)	−0.047
70	10.27 (3.57)	0.027	0.024 (0.004)	−0.040	97.26 (10.21)	−0.027
100	10.78 (3.89)	0.078	0.026 (0.005)	0.040	99.63 (15.60)	−0.004
120	10.13 (1.71)	0.013	0.025 (0.003)	0.000	97.72 (6.58)	−0.023

150	10.64 (2.38)	0.064	0.025 (0.003)	0.000	99.80 (6.88)	−0.002
200	10.12 (1.93)	0.012	0.025 (0.002)	0.000	98.70 (5.70)	−0.013
250	10.16 (1.72)	0.016	0.025 (0.002)	0.000	99.30 (5.22)	−0.007
300	10.03 (1.54)	0.003	0.025 (0.002)	0.000	98.75 (4.50)	−0.012
500	10.10 (1.17)	0.010	0.025 (0.001)	0.000	99.45 (3.39)	−0.006
750	10.08 (0.94)	0.008	0.025 (0.001)	0.000	99.64 (2.67)	−0.004
1000	10.09 (0.81)	0.009	0.025 (0.001)	0.000	99.70 (2.28)	−0.003
1500	10.05 (0.65)	0.005	0.025 (0.001)	0.000	99.72 (1.79)	−0.003

Consistency is demonstrated here again in the ability of the model to produce estimates that converge to the actual parameter values as n increases. Examining the estimates for α, λ and δ, we find that the estimates become increasingly stable and closer to the actual values with larger sample sizes. The parameter estimates remain close to their actual values, with only minor deviations. These estimates suggest that the model is consistent, as each tends to converge to its actual value as n increases. The confidence intervals (Figure 4) for each parameter become narrower, indicating that the model's estimates become more precise and stable. The confidence intervals continue to shrink, becoming much tighter as n becomes large. The reduction in the width of these intervals suggests that the TopLE distribution's parameter estimates become more stable as n increases, providing more reliable and precise estimates.

The representation in Figure 4 demonstrates that the TopLE distribution exhibits high efficiency, consistency, and stability. As n increases, the standard errors decrease, confidence intervals become narrower, and the estimates converge to their respective actual values. These findings indicate that the model performs well in estimating the parameters. α, λ and δ, especially with larger sample sizes, and can be considered a robust tool for statistical modelling.

Applications of the TopLE Distribution

Two real-life datasets were modelled using the new probability distribution. These are, the fatigue life (to the nearest thousand cycles) of 67 Specimens of Alloy T7987 that failed before having accumulated 300 thousand cycles of testing obtained from Meeker & Escobar (1998, pp 149); and the monthly records of tax revenue of Egypt between January 2006 and November 2010 (1000 million Egyptian pounds), obtained from Klakattawi (2019). The TopLE distribution was fitted to these datasets with four other standard distributions for comparison. The parameter estimates for each data is carried out using the *fitdistrplus* package in R. The hypotheses for the Kolmogorov-Smirnov (KS) goodness of fit test are given as:

H_0: maximum absolute distance between the empirical and fitted cumulative distribution functions is zero, that is $F(x) = F_n(x)$

H_1: The maximum absolute distance between the empirical and fitted cumulative distribution functions is not zero, that is, $F(x) \neq F_n(x)$.

The fit results for the 67 Specimens of Alloy T7987 are presented in Table 3, while those of the Egyptian Tax data are presented in Table 4.

Table 3

Fitting Fatigue Life of 67 Specimens of Alloy T7987

Distribution	Parameter	Estimates (se)	LL (AIC)	KS	Remark
TopLE	$\hat{\alpha}$	21.1538 (8.2161)	-348.19 (702.37)	0.0561*	Good fit
	$\hat{\lambda}$	0.0098 (0.0014)			
	$\hat{\delta}$	345.9872 (44.4051)			
Normal	$\hat{\mu}$	166.0746 (5.6656)	-352.13 (708.26)	0.0835*	Good fit
	$\hat{\sigma}$	46.3746 (4.0062)			
Weibull	$shape$	3.7278 (0.3328)	-353.29 (710.58)	0.0972*	Good fit
	$scale$	183.6085 (6.3871)			
Gamma	$shape$	13.5346 (2.3041)	-348.64 (701.27)	0.0538*	Good fit
	$rate$	0.0815 (0.0141)			
Logistic	$location$	162.2845 (5.5188)	-352.07 (708.15)	0.0674*	Good fit
	$scale$	25.9900 (2.6560)			

Note: KS = Kolmogorov-Smirnov statistic measures the maximum absolute distance between the empirical and fitted cumulative distribution function plots. * compares to the critical value of 0.1662 at 5% level of significance

Table 4

Fitting Egyptian Monthly Tax Revenue Data

distribution	parameter	Estimates (se)	LL (AIC)	KS	Remark
TopLE	$\hat{\alpha}$	5.5198 (1.4321)	-191.20 (388.41)	0.1230*	Good fit
	$\hat{\lambda}$	0.0894 (0.0117)			
	$\hat{\delta}$	46158.25 (1236.83)			
Normal	$\hat{\mu}$	13.4729 (1.0401)	-206.32 (416.65)	0.1882*	Poor fit
	$\hat{\sigma}$	7.9892 (0.7355)			
Weibull	$shape$	1.8376 (0.1709)	-197.29 (398.57)	0.1422*	Good fit
	$scale$	15.2884 (1.1517)			
Gamma	$shape$	3.6696 (0.6472)	-193.07 (390.15)	0.1327*	Good fit
	$rate$	0.2724 (0.0515)			
Logistic	$location$	12.1967 (0.9065)	-202.61 (409.23)	0.1317*	Good fit
	$scale$	4.0671 (0.4514)			

Note: KS = Kolmogorov-Smirnov statistic measures the maximum absolute distance between the empirical and fitted cumulative distribution function plots. * compares to the critical value of 0.1771 at 5% level of significance

Table 3 shows that the gamma distribution was the best model in providing a good fit to the specimens of the Alloy T7987 dataset from the perspective of the Kolmogorov-Smirnov goodness-of-fit test.

However, the TopLE did fairly well compared to the gamma distribution, but much better than the other three standard distributions. Figure 5 further demonstrates the non-existence of the differences in the fit of the gamma and the TopLE distributions to the data. It can then be said that the TopLE distribution is a competitive contender in describing the probabilistic characteristics of the Alloy T7987 data generation processes.

The Egyptian Tax dataset's fit results also demonstrate the TopLE distribution's capabilities in describing real-life data generation processes. It provided the best fit based on all the fit metrics (KS, AIC, and LL) as shown in Table 4. The density plots of all the fitted distributions on the histogram of the sample dataset, presented in Figure 6, also show the TopLE distribution's ability to describe the characteristics of the Egyptian Tax data generation processes.

Figure 5

Distributions' Fit to Alloy T7987 Data

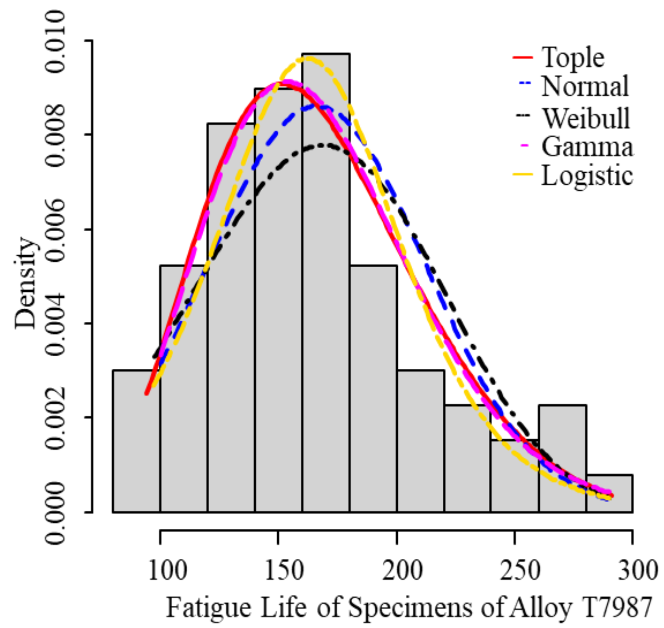
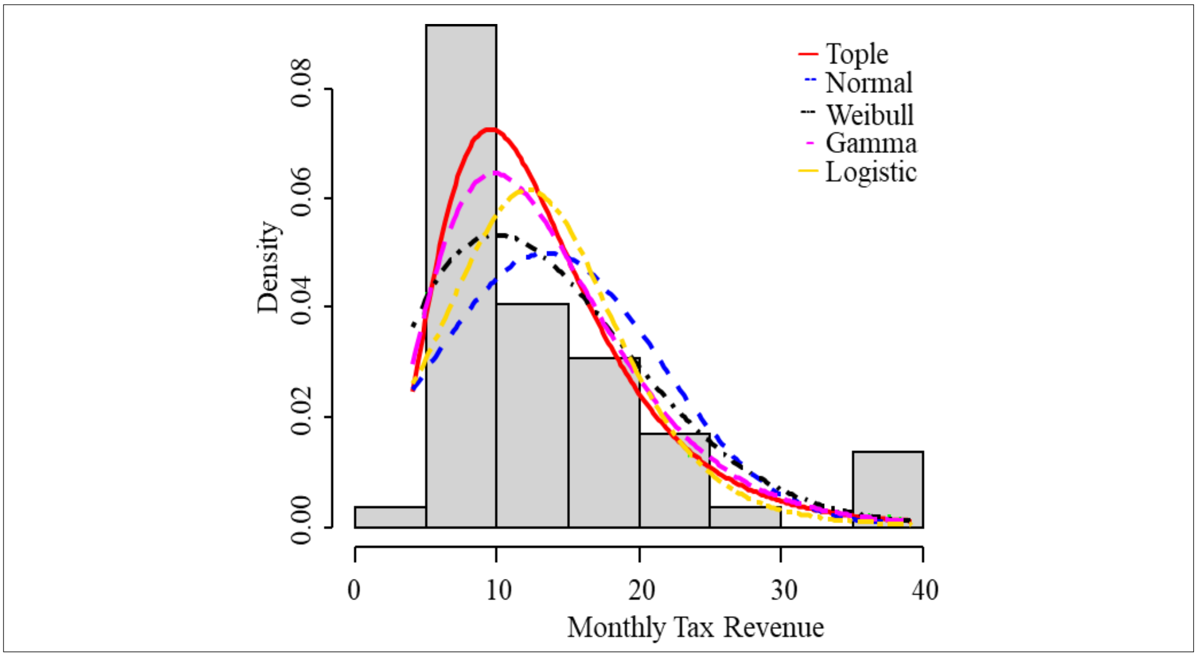


Figure 6

Distributions' Fit to Monthly Tax Revenue Data



Discussion

The development of the TopLE distribution demonstrates the benefits of synthesizing established probability distributions to create new, versatile models. Combining the Topp-Leone and epsilon distributions, both of which are J-shaped (Nadarajah & Kotz, 2003; Dombi et al., 2018), the TopLE model produced favourable properties such as flexibility in capturing diverse data characteristics and ease of parameter estimation. The derived probability density, cumulative distribution, and quantile functions showcase their potential for applicability in real-world scenarios.

Simulation results validate the robustness of the TopLE distribution. As sample size increases, parameter estimates converge to their actual values with diminishing standard errors, biases, and confidence intervals, confirming the model's efficiency, consistency, and stability. This aligns with theoretical expectations and underscores the distribution's practicality for statistical modelling. Moreover, the availability of closed-form solutions enhances its computational efficiency, allowing for straightforward implementation in statistical software like R.

The TopLE distribution's potential applications span various fields, particularly in modelling lifetime and reliability data, where traditional distributions may fall short; for examples, the upper bound parameter, δ, for the random variable characterized by the new distribution depicts a real-life scenario of finite systems' existence. Future research could explore extensions of the TopLE distribution, such as multivariate adaptations, and investigate its performance in specific domains like engineering, healthcare, and environmental studies. The TopLE distribution's comparative application showed its potential as an alternative when considering probabilistic models to characterize dynamic systems. The distribution outperformed the traditional distributions in the datasets, which were fitted by all metrics of good fit.

CONCLUSION

This study presents the TopLE distribution as a novel and competitive addition to the family of probability distributions. Its derivation from the Topp-Leone and epsilon distributions yields a model with significant theoretical and practical advantages, including flexibility, efficient parameter estimation, and robustness across sample sizes. Simulation results highlight the model's efficiency, consistency, and stability, making it a reliable choice for analyzing lifetime datasets. The TopLE distribution is a powerful tool for researchers and practitioners, offering theoretical rigor and practical utility. Further exploration of its applications in parametric survival analysis and extensions to multivariate form will continue to enrich the field of statistical modelling.

ACKNOWLEDGEMENT

We want to extend our sincere gratitude to the reviewers for their insightful comments and constructive feedback, which significantly improved the quality of this research. We also appreciate the editorial team's expertise and guidance throughout the publication process. This research was made possible by the generous sponsorship of the University of Maiduguri Teaching Hospital (UMTH).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Aalen, O., Borgan, O., & Gjessing, H. (2008). *Survival and Event History Analysis: A Process Point of View*. Springer-Verlag.

Adegoke M. T., Oladoja, O. M., Bashiru, O, S., Mustapha, A. A., Aderupatan, D. E., & Nzei, L. C. (2023). Topp-Leone Inverse Gompertz Distribution: properties and different estimation techniques and application, *Pakistan Journal of Statistics,* 39 (4),2023

Ahmad, Z., Elgarhy, M., & Hamedani, G. G. (2018). A new Weibull-X family of distributions: properties, characterizations, and applications. *Journal of Statistical Distributions and Applications*, 5(1), 5.

Al-Shomrani A., Arif O., Shawky A., Hanif S., & Shahbaz M. Q. (2016). Topp–Leone Family of Distributions: Some Properties and Application, *Pakistan Journal of Statistics and Operations Research*. Vol.XII No.3 2016 pp443-451

Aly, E. E., & Benkherouf, L. (2011).A new family of distributions based on probability generating function. *Sankhya B* 73(1).

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

Alzaghal, A., Famoye, F., & Lee, C. (2013). Exponentiated T-X family of distributions with some applications. *International Journal of Statistics and Probability*, 2(3), 31

Barrios, R., & Dios, F. (2012). Exponentiated Weibull distribution family under aperture averaging for Gaussian beam waves. *Optics Express*, 20:, 13055–13064

Cakmakyapan, S., Ozel, G., El Gebaly, Y. M. H., & Hamedani, G. G. (2018). The Kumaraswamy Marshall-Olkin Log-Logistic Distribution with Application. *Journal of Statistical Theory and Applications*, 17(1), 59-76.

Choudhury, A. (2005). A simple derivation of moments of the exponentiated Weibull distribution. *Metrika*, 62: 17–22.

Cordeiro, G. M., Ortega, E. M., & Nadarajah, S. (2010). The Kumaraswamy Weibull distribution with application to failure data. *Journal of the Franklin Institute*, 347(8), 1399–1429.

David, H. A., & Nagaraja, H. N. (2003). Order Statistics, 3rd edn. John Wiley & Sons Inc., New Jersey (2003)

Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R Package for Fitting Distributions, J*ournal of Statistical Software,* February 2015, Volume 64, Issue 4.

Debnath, L., & Basu, K. (2015). A Short History of Probability Theory and its Application, *International Journal of Mathematical Education in Science and Technology*, 46(1), 13–39. https://www.learntechlib.org/p/168053/

Dombi, J., Jonas, T., & Toth, Z. (2018). The Epsilon Probability Distribution and its Application in Reliability Theory. *Acta Polytechnica Hungarica* 15(1),197 – 216.

Eugene, N., Lee, C., & Famoye, F. (2002). Beta-normal distribution and its applications. *Communication in Statistics: Theory and Methods*, 31: 497–512.

Gentle, E.J. (2005). Random Number Generation and the Monte Carlo Methods, Second Edition (2005), Springer

Gongsin, I. E., & Saporu, F. W. O. (2019). On the Construction of Kumaraswamy-Epsilon Distribution with Applications, *International Journal of Science and Research (IJSR)* Volume 8 Issue 11, November 2019, ISSN: 2319–7064.

Gongsin, I. E., & Saporu, F. W. O. (2021). A beta–epsilon distribution: properties and applications, *Afrika Matematika* 32:1573 1583.

Gongsin, I. E., & Saporu, F. W. O. (2020). The Weibull Epsilon distribution: properties and applications, *International Journal of Mathematical and Computational Methods,* volume 5, 2020, pp 80-88.

Gongsin, I. E., & Saporu, F. W. O. (2019). The Exponentiated-Epsilon distribution: its properties and Applications, *International Journal of Science and Research (IJSR),* Volume 8 Issue 12, 2019, pp 1024 -1028.

Gupta, R. C., Gupta, R. D., & Gupta, P. L. (1998). Modelling failure time data by Lehmann alternatives. *Communications in Statistics: Theory and Methods*, 27: 887–904.

Gupta, R. D., & Kundu, D. (1999). Generalized exponential distributions. *Australian & New Zealand Journal of Statistics*, 41: 173–188.

Gupta, R. D., & Kundu, D. (2001). Exponentiated exponential distribution: an alternative to gamma and Weibull distributions. *Biometrical Journal*, 43: 117–130. 83

Gupta, R. D., & Kundu, D. (2002). Generalized exponential distribution: Statistical inferences. *Journal of Statistical Theory and Applications*, 1: 101–118

Jamal, F., Reyad, H. M., Nasir, M. A., Chesneau, C. Shah, M. A. A., & Ahmed, S. O. (2019). Topp-Leone Weibul-Lomax Distribution: Properties, Regression Model and Application. HAL Id: hal-02270561. https://hal.science/hal-02270561

Jones, M. C. (2004). Families of distributions arising from distributions of order statistics. *TEST*, 13: 1–43.

Klakattawi, H. S. (2019). The Weibull-Gamma Distribution: Properties and Applications, *Entropy*, 21, 438, pp 1 – 15.

Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, 46: 79–88.

Lehmann, E. L. (1953). The power of rank tests. *Annals of Mathematical Statistics,* 24: 23–43.

Mudholkar, G., & Srivastava, D. (1993). Exponentiated Weibull family for analyzing bathtub failure-real data. *IEEE Transaction on Reliability*, 42: 299–302.

Mudholkar, G. S., & Hutson, A.D. (1996). The exponentiated Weibull family: Some properties and a flood data application. *Communications in Statistics-Theory and Methods*, 25: 3059–3083.

Mudholkar, G. S., Srivastava, D. K., & Freimer, M. (1995). The exponentiated Weibull family: a reanalysis of the bus-motor-failure data. *Technometrics*, 37: 436–445

Nadarajah, S., & Kotz, S. (2003). Moments of some J-shaped distributions. *Journal of Applied Statistics*, 30:3, 311-317.

Nadarajah, S., Cordeiro, G. M., & Ortega, E. M. (2015). The Zografos– Balakrishnan–G Family of Distributions: Mathematical Properties and Applications 85 Mathematical properties and applications. *Communications in Statistics – Theory and Methods*, 1: 18–215.

Nzei, L. C., & Ekhosuehi, N. (2020). Topp-Leone Lindley Distribution, *Abacus* (Mathematics Science Series) Vol. 47, No. 1, December 2020

Oguntunde, P. E., Adejumo, A., & Owoloko, E. A. (2017). Application of Kumaraswamy Inverse Exponential Distribution to Real Lifetime Data. *International Journal of Applied Mathematics and Statistics*.

Qian, L. (2012). The Fisher information matrix for the three-parameter exponentiated Weibull distribution under type II censoring. *Statistical Methodology*, 9: 320–329.

Ristic, M. M., & Balakrishnan, N. (2012). The gamma–exponentiated exponential distribution. *Journal of Statistical Computation and Simulation*, 8: 1191–1206.

Tahir, M., & Cordeiro, G. M. (2016). Compounding of distributions: a survey and new generalized classes. *Journal of Statistical Distributions and Applications*, 3: 13.

Tahir, M., & Nadarajah, S. (2015). Parameter induction in continuous univariate distributions: Well-established G-classes. *Anais da Academia Brasileira de Ciências,* 87: 539–568.

Tuoyo D. O., Opone F. C., & Ekhosuehi N. (2021). The Topp-Leone Weibull Distribution: Its Properties and Application. *Earthline Journal of Mathematical Sciences*, Volume 7, Number 2, 2021, Pages 381-401.

Zografos, K., & Balakrishnan, N. (2009). On families of beta- and generalized gamma generated distributions and associated inference. *Statistical Methodology*, 6: 344.