AIMS Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 9(9): 24810–24831.
DOI: 10.3934/math.20241208
Received: 12 June 2024
Revised: 07 August 2024
Accepted: 09 August 2024
Published: 23 August 2024

Research article

A flexible model for bounded data with bathtub shaped hazard rate function and applications

M. R. Irshad[1], **S. Aswathy**[1], **R. Maya**[1], **Amer I. Al-Omari**[2,*] **and Ghadah Alomani**[3]

[1] Department of Statistics, Cochin University of Science and Technology, Cochin 682 022, Kerala, India

[2] Department of Mathematics, Faculty of Science, Al al-Bayt University, Mafraq, Jordan

[3] Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

* **Correspondence:** Email: alomari_amer@yahoo.com.

Abstract: The unit new X-Lindley distribution, which is a novel one-parameter distribution on the unit interval, is presented in this study. It was developed by altering the new X-Lindley distribution using the exponential transformation. This new one-parameter distribution's fundamental features, including moments, incomplete moments, Lorenz and Bonferroni curves, Gini index, mode, extropy, Havrda and Charvat entropy, Rényi entropy, and Tsallis entropy, are explored. Additionally, it has bathtub-shaped hazard rate functions and monotonically increasing hazard rate functions with a single parameter. The new distribution is therefore sufficiently rich to model real data. Also, different estimation methods, such as maximum likelihood, least-squares, and weighted least-squares, are used to estimate the parameters of this model, and using a simulation research, their respective performances are evaluated. Finally, two real-life datasets are used to demonstrate the suggested model's competency.

Keywords: new X-Lindley distribution; failure time; unit new X-Lindley distribution; hazard rate function; parameter estimation; entropy measures; empirical studies
Mathematics Subject Classification: 60E05, 60F05

Abbreviations	Symbols
pdf	Probability density function
cdf	Cumulative distribution function
NXLD	New X-Lindley distribution
UNXLD	Unit new X-Lindley distribution
mrl	Mean residual life
sf	Survival function
hrf	Hazard rate function
$L_F(x)$	Lorenz curve
$B_F(x)$	Bonferroni curve
$Gi(x)$	Gini index
$J(X)$	Extropy of X
$Re(\delta)$	Rényi entropy of X
$Te(\lambda)$	Tsallis entropy of X
$HCe\,(\beta)$	Havrda and Charvat entropy of X
ML	Maximum likelihood
LS	Least-squares
WLS	Weighted least-squares
MLE	ML estimate
LSE	LS estimate
MSE	Mean square error
RMSE	Root mean square error
AIC	Akaike information criterion
AICc	Akaike information criterion corrected
CAIC	Consistent Akaike information criterion
BIC	Bayesian information criterion
KS	Kolmogorov-Smirnov
BD	Beta distribution
TLD	Topp-Leone distribution
KD	Kumaraswamy distribution
ETLD	Exponentiated Topp-Leone distribution
ULD	Unit Lindley distribution
UBD	Unit Burr-III distribution
QQ	Quantile–quantile
P-P	Probability–probability
TTT	Total test time

1. Introduction

The introduction of new distributions allows statisticians and researchers to better model a variety of data types, from symmetric to skewed, and from positive to bi-directional distributions. These distributions provide tools for understanding data variability, estimating parameters, making predictions, and conducting statistical hypothesis testing across diverse fields of study. For instance, percentages, proportions, and other quantities that have limits between 0 and 1 are better represented by bounded distributions. Overall, distributions with bounded support offer practical tools for accurately modeling a wide range of data types and phenomena, supplying information, enabling analysis, and supporting decision-making in a variety of sectors. In the scientific literature, bounded distributions are noticeably scarce compared to unbounded distributions, despite the ubiquity of actual scenarios with bounded data.

The beta distribution, renowned for its versatility and applicability, holds a distinct place in statistical modeling due to its ability to represent data with a range from 0 to 1. The beta distribution was revitalized when Karl Pearson identified it as a member of the Pearson distribution family, specifically as a Pearson Type I distribution. While the beta distribution is a valuable tool, it's not a one-size-fits-all solution. As a result, there are more studies on unit modeling in the literature. The transformation of well-known continuous distributions has typically been used to introduce the newly proposed unit distributions. The benefit of these unit distributions is that they increase the basic distribution's flexibility throughout the unit interval without requiring the addition of new parameters. Such as, the Kumaraswamy Kumaraswamy distribution [1], the log-xgamma distribution [2], the one parameter unit-Lindley distribution [3], the unit-Gompertz distribution [4], the unit Weibull distribution [5], the unit Burr-III distribution [6], the unit-Rayleigh distribution [7], the unit Burr-XII distribution [8], the unit half normal distribution [9], the exponentiated unit Lindley distribution [10], the unit Teissier distribution [11], the unit two parameters Mirra distribution [12] and the two parameters unit Muth distribution [13], etc.

The NXLD proposed by [14], which is obtained as a special case of the new one-parameter polynomial exponential distribution introduced by [15], is used as the baseline distribution in this article to introduce a new bounded distribution, namely the unit new X-Lindley distribution (UNXLD). Here, it is necessary to mention the NXLD. A continuous random variable (rv) X is said to have the NXLD with parameter $\theta > 0$ if its pdf is of the form:

$$f(x;\theta) = \frac{\theta(1+\theta x)e^{-\theta x}}{2}, \quad x > 0. \tag{1.1}$$

The cdf of X is given by

$$F(x;\theta) = 1 - \left(\frac{1}{2}\theta x + 1\right)e^{-\theta x}, \quad x > 0.$$

The NXLD is important for its ability to model positive real data, [14] explored the NXLD's applicability in actuarial science by examining its actuarial properties. Due to its qualities, it is a useful tool for statistical analysis, research, and practical applications in a variety of fields. For numerous real-world datasets, the NXLD demonstrated its uniqueness when compared to other models. Moreover, the Lindley distribution has been applied in numerous fields, such as finance, environmental studies, and medical research. Its ability to handle various types of data makes it a valuable tool in statistical modeling. This NXLD offers a unique combination of features from both the Lindley and exponential distributions. These are the motivations for introducing the NXLD on the unit interval.

Numerous probability distributions are created in the literature by combining, expanding, and changing well-known distributions, which makes it possible to describe lifetime data with a more flexible hrf. The bathtub curve is an idealized representation that helps in understanding and modeling the behavior of systems. In the case of real-life datasets that vary over the positive real line, possess bathtub shaped hrf. [16–18] etc are examples of the same. In the latter case, it was demonstrated that the unit inverse Gaussian distribution and the logit Slash distribution both produce hrfs with a bathtub shape, as shown in [19] and [20], respectively. Due to the log function's inclusion in their density functions, these two distributions are more complex, and as a result, the cdf is not calculated in closed-form. As a result, modeling studies using these two distributions are quite challenging. Another one is the unit Burr-XII distribution [8], which has two parameters, and the Meijer G-function's presence in the formulation of its moments makes it slightly more difficult to understand the model's flexibility.

The Lambert-uniform distribution with only one-parameter, proposed by [21], faces complications in calculating associated properties due to the presence of the principal branch of the Lambert W function and the logarithm function in its pdf. The unit Muth distribution [13] is another two-parameter distribution. Similarly, the definition of moments comprises the gamma function, which causes complexity. The unit generalized half-normal distribution was recently presented by [22], although it lacks closed-form formulas for the cdf and quantile function. Less parameterized models are typically easier to comprehend and interpret. This facilitates understanding of the underlying relationships and implications of the model by researchers, analysts, and stakeholders.

These problems are improved by our newly defined UNXLD. It presents closed-form expression for cdf, hrf, and all its statistical measures, even when a logarithmic function is present in its pdf. More over, it possesses a monotonically increasing and bathtub-shaped hrf with only one parameter. After all, the supremacy of the suggested model is confirmed by fitting a well-known bathtub-shaped dataset that varies over a unit interval with a few other bounded distributions in the unit interval.

The rest of the paper is presented as follows: In Section 2, moments, incomplete moments, and other properties are explored together with the introduction of the UNXLD. In Section 4, we explore different estimation approaches, such as ML, LS, and WLS. Additionally, a simulation exercise is conducted to assess the effectiveness of model parameter estimates obtained through these methods. Two datasets are used in Section 5 to clarify the recommended distribution. Finally, Section 6 presents the conclusions.

2. The unit new X-Lindley distribution

The UNXLD is obtained by the exponential transformation, $X = e^{-Y}$, of the NXLD. A rv X is said to follow the UNXLD with parameter $\theta > 0$, if its pdf is of the following form:

$$f(x;\theta) = \frac{\theta}{2} x^{\theta-1}(1 - \theta \log x), \quad x \in (0,1), \ \theta > 0. \tag{2.1}$$

The corresponding cdf is given by

$$F(x;\theta) = x^{\theta} - \frac{\theta x^{\theta} \log x}{2}. \tag{2.2}$$

For any $\theta > 0$, $lim_{x\to 0} f(x;\theta) = 0$ and $lim_{x\to 1} f(x;\theta) = \frac{\theta}{2}$. Figure 1 displays the pdf of the UNXLD for various values of the parameter θ.

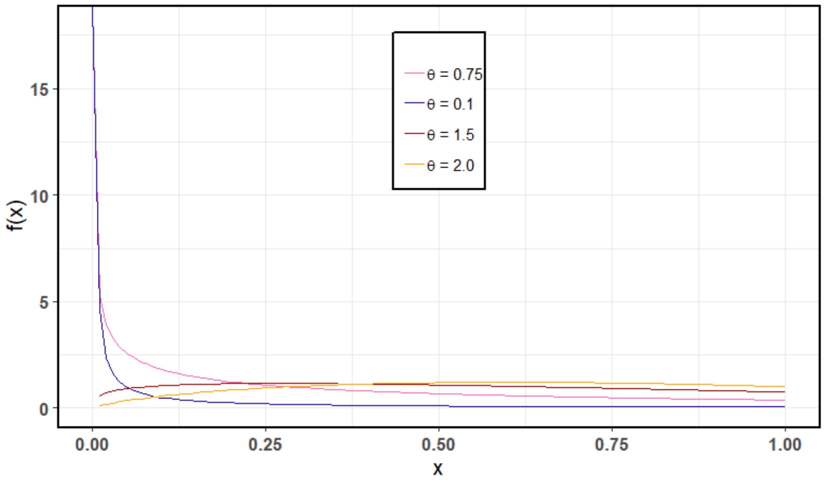


Figure 1. Pdfs of the UNXLD for different values of θ.

Furthermore, from Figure 1, it is clear that it serves as an appropriate probability model for modeling right-skewed measurements within the unit interval.

2.1. Moments

Assume that X is a rv with UNXLD. The k^{th} non-central moment of X for $k = 1, 2, 3...$ is given by

$$E(X^k) = \frac{\theta(k+2\theta)}{2(k+\theta)^2}. \tag{2.3}$$

From (2.3), the mean and variance of UNXLD is obtained as

$$E(X) = \frac{\theta(2\theta+1)}{2(\theta+1)^2}, \tag{2.4}$$

and

$$V(X) = \frac{7\theta^4 + 20\theta^3 + 16\theta^2 + 4\theta}{4(\theta+2)^2(\theta+1)^4}.$$

Additionally, the skewness and kurtosis of UNXLD are given by

$$Skewness(X) = \frac{4(2+\theta)^2(-24+\theta(-130+\theta(-244+\theta(-154+3\theta(13+\theta(22+5\theta))))))^2}{\theta(3+\theta)^4(4+\theta(16+\theta(20+7\theta)))^3},$$

and

$$Kurtosis(X) = \frac{1}{\theta(3+\theta)^2(4+\theta)^2(4+\theta(16+\theta(20+7\theta)))^2} 3(2+\theta)^2(384+\theta(2880 + \theta(8864+\theta(14704+\theta(15272+\theta(12100+\theta(8324+\theta(4099+3\theta(362+370)))))))))).$$

The behavior of the mean, variance, skewness, and kurtosis coefficients for various values of the parameter is graphically depicted in Figures 2 and 3. It is evident that as θ increases, the mean also increases. Additionally, the variance initially increases for certain values of θ, but then it begins to decrease. As for skewness, it increases with θ, indicating a right skew. Additionally, the kurtosis also increases as θ increases.

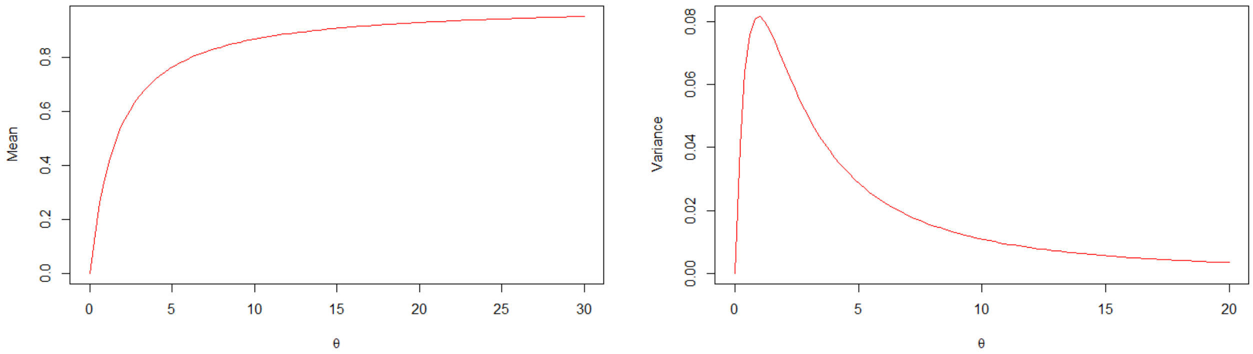


Figure 2. Mean and variance of the UNXLD for different values of θ.

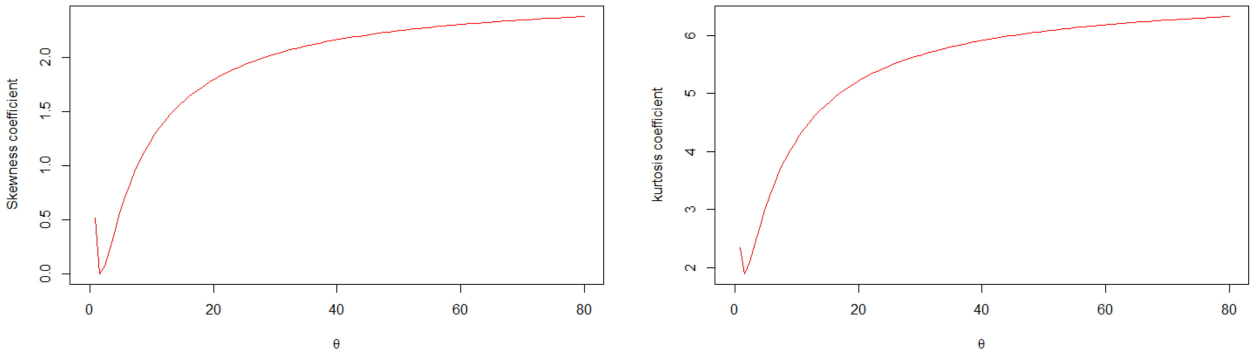


Figure 3. Skewness and kurtosis of the UNXLD for different values of θ.

Proposition 2.1. *The k^{th} incomplete moment at x of a rv X with the UNXLD is for any non-negative integer k and for any value of $x \in (0, 1)$ is given by*

$$m_k(x;\theta) = \frac{\theta^2 x^{k+\theta}[(k+\theta)\log x - 1]}{2(k+\theta)^2}.$$

Proof. According to the definition of the k^{th} incomplete moment, there are

$$\begin{aligned} m_k(x;\theta) &= \int_0^x t^k f(t;\theta)dt \\ &= \int_0^x t^k \frac{\theta}{2} t^{\theta-1}(1-\theta\log t)dt \\ &= \frac{\theta^2 x^{k+\theta}[(k+\theta)\log x - 1]}{2(k+\theta)^2}. \end{aligned}$$

□

2.2. Hazard rate function

The hrf provides insight into the risk of an event happening at a particular time, considering the history of events up to that time.

Using (2.2), the sf and hrf of X are given by

$$s(x;\theta) = 1 - x^{\theta} + \frac{\theta x^{\theta}\log x}{2},$$

and

$$h(x;\theta) = \frac{\theta x^{\theta-1}(1-\theta\log x)}{2\left(1 - x^{\theta} + \frac{\theta x^{\theta}\log x}{2}\right)}. \tag{2.5}$$

Additionally, for any $\theta > 0$, $lim_{x\to 0}h(x;\theta) = 0$ and $lim_{x\to 1}h(x;\theta) = \infty$.

The reversed hrf of X is given by

$$\tau(x;\theta) = \frac{\theta(1-\theta\log x)}{x(2-\theta\log x)}.$$

Figure 4 illustrates the hrf for various values of the parameter θ, providing a general understanding of the different forms of the hrf (2.5). It is evident that the UNXLD can exhibit increasing, and bathtub-shaped hrf.

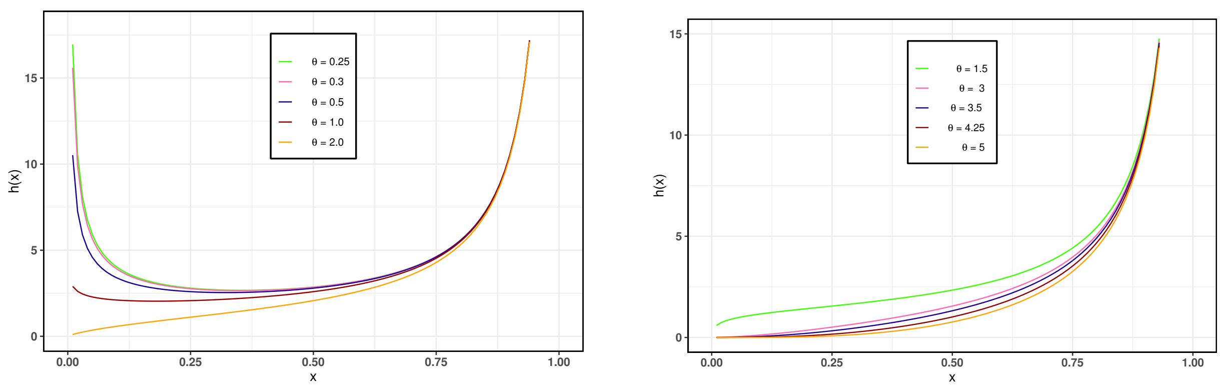


Figure 4. The hrf of the UNXLD for different values of θ.

Similarly, mrl provides the average time a system or component can be expected to continue functioning before reaching a specific condition, which is valuable for making informed decisions about maintenance, replacement, and resource allocation.

Proposition 2.2. *The mrl of a rv having the UNXLD is given by*

$$r(t;\theta) = \frac{1}{s(t;\theta)}\left(1 - t - \frac{1}{\theta+1} + \frac{t^{\theta+1}}{\theta+1} - \frac{\theta t^{\theta+1}\log t}{2(\theta+1)} - \frac{\theta}{2(\theta+1)^2} + \frac{\theta t^{\theta+1}}{2(\theta+1)^2}\right),\quad 0 < t < 1,$$

where $s(x;\theta) = 1 - F(x;\theta)$ *is the sf,* $0 < x < 1$.

Proof. Let X be a rv with a UNXLD with cdf (2.2). The mrl of X can be stated as follows:

$$\begin{aligned} r(t;\theta) &= E(X-t|X>t) \\ &= \frac{1}{s(t;\theta)}\int_t^1 s(x;\theta)dx \\ &= \frac{1}{s(t;\theta)}\Big(1-t-\frac{1}{\theta+1}+\frac{t^{\theta+1}}{\theta+1}-\frac{\theta t^{\theta+1}\log t}{2(\theta+1)}-\frac{\theta}{2(\theta+1)^2}+\frac{\theta t^{\theta+1}}{2(\theta+1)^2}\Big). \end{aligned}$$

□

2.3. Inequality measures and Gini index

Lorenz and Bonferroni curves serve as essential tools for measuring income inequality and have broad applications beyond economics, extending into diverse fields such as reliability analysis, demography, medicine, and insurance. For the UNXLD, we will derive Lorenz and Bonferroni curves in this section.

Proposition 2.3. *The Lorenz and Bonferroni curves and Gini index of the UNXLD are given respectively as follows:*

(1)

$$L_F(x) = \frac{x^{\theta+1}(2\theta+1-(\theta+1)\theta\log x)}{2\theta+1};$$

(2)

$$B_F(x) = \frac{x(2\theta+1-(\theta+1)\theta\log x)}{(2\theta+1)(1-\frac{\theta\log x}{2})};$$

(3)

$$Gi(x) = 1-\frac{\theta(\theta(\theta(16\theta+21)+8)+1)}{(2\theta+1)^4}, \theta > -\frac{1}{2}.$$

Proof. By using the first incomplete moment $m_1(x;\theta)$ and the E(X), one can determine the Lorenz and Bonferroni curves.

(1)

$$L_F(x;\theta) = \frac{m_1(x;\theta)}{E(x)}. \tag{2.6}$$

Now,

$$\begin{aligned} m_{(x;\theta)} &= \int_0^x tf(t;\theta)dt \\ &= \frac{\theta x^{\theta+1}(2\theta+1-(\theta+1)\theta\log x)}{2(\theta+1)^2}. \end{aligned} \tag{2.7}$$

By substituting (2.7), and (2.4) into (2.6), we have

$$L_F(x;\theta) = \frac{x^{\theta+1}(2\theta+1-(\theta+1)\theta\log x)}{2\theta+1}.$$

(2)

$$B_F(x;\theta) = \frac{m_1(x;\theta)}{F(x;\theta)E(X)}. \quad (2.8)$$

By substituting (2.7), (2.4), and (2.2) into (2.8), we have

$$B_F(x) = \frac{x(2\theta + 1 - (\theta + 1)\theta \log x)}{(2\theta + 1)\left(1 - \frac{\theta \log x}{2}\right)}.$$

(3) The proof is obtained directly by using the cdf in Eq (2.2) and the mean in (2.4) of the UNXLD and is given by

$$Gi(x) = 1 - \frac{\int_0^1 (1 - F(x,\theta))^2 \, dx}{\mu}.$$

□

2.4. Mode

Proposition 2.4. *The pdf of the UNXLD has the mode given by*

$$x_M = e^{\frac{1}{(1-\theta)\theta}}.$$

Proof. To find the mode, take the log of the UNXLD pdf as:

$$log(f(x;\theta)) = log(\theta/2) + log(x^{(\theta-1)}) + log(1 - \theta log(x)).$$

Differentiate $log(f(x;\theta))$ with respect to x to obtain:

$$\frac{\partial log(f(x;\theta))}{\partial x} = \frac{\theta - 1}{x} - \frac{\theta}{x(1 - \theta \log(x))}.$$

Equating this equation to zero and solving for x, we obtain

$$x_M = e^{\frac{1}{(1-\theta)\theta}}.$$

□

3. Extropy and entropy measures

Let X be a non-negative absolutely continuous rv with pdf $f(x)$. In this section, the extropy measure suggested by [23] and three entropy measures, include, the Tsallis entropy proposed by [24], the Rényi entropy introduced by [25], and the Havrda and Charvat entropy suggested by [26] for the UNXLD, are presented.

Proposition 3.1. *If $X \sim UNXLD$, then the extropy of X is defined as:*

$$J(X) = \frac{\theta^2(2(3 - 5\theta)\theta - 1)}{8(2\theta - 1)^3}. \quad (3.1)$$

Proof. Using the pdf of the UNXLD in Eq (2.1) and the definition of extropy $J(X) = -\frac{1}{2}\int_0^1 f^2(x)\,dx$, the proof is given by

$$\begin{aligned}
J(X) &= -\frac{1}{2}\int_0^1 f^2(x)dx \\
&= -\frac{1}{2}\int_0^1 \frac{1}{4}\theta^2 x^{2\theta-2}\left((\theta\log(x))^2 - 2\theta\log(x) + 1\right)dx \\
&= -\frac{1}{2}\left(\int_0^1 \frac{\theta^2}{4}x^{2\theta-2}dx - \int_0^1 \frac{\theta^3}{2}x^{2\theta-2}\log(x)dx + \int_0^1 \frac{\theta^4}{4}x^{2\theta-2}\log^2(x)dx\right) \\
&= -\frac{1}{2}\left(\frac{\theta^4}{2(2\theta-1)^3} + \frac{\theta^3}{2(1-2\theta)^2} + \frac{\theta^2}{4(2\theta-1)}\right) \\
&= \frac{\theta^2(2(3-5\theta)\theta-1)}{8(2\theta-1)^3}.
\end{aligned}$$

□

Proposition 3.2. *If $X \sim UNXLD$, then the Rényi entropy of X is defined as:*

$$Re(\delta) = \frac{1}{1-\delta}\log\left(\frac{\theta^{2\delta}e^{\frac{\delta(\theta-1)+1}{\theta}}}{2^{\delta}(\delta(\theta-1)+1)^{\delta+1}}\Gamma\left(\delta+1, \frac{\delta(\theta-1)+1}{\theta}\right)\right), \delta > 0, \delta \neq 1. \tag{3.2}$$

Proof. Using the pdf of the UNXLD in Eq (2.1) and the definition of Rényi entropy $Re(\delta) = \frac{1}{1-\delta}\log\left(\int_0^\infty f^\delta(x)dx\right)$, $\delta > 0$, $\delta \neq 1$, the proof can be obtained. □

Proposition 3.3. *If $X \sim UNXLD$, then the Tsallis entropy of X is defined as:*

$$Te(\lambda) = \frac{1}{\lambda-1}\left(1 - \frac{\theta^{2\lambda}e^{\frac{(\theta-1)\lambda+1}{\theta}}}{2^{\lambda}((\theta-1)\lambda+1)^{\lambda+1}}\Gamma\left(\lambda+1, \frac{(\theta-1)\lambda+1}{\theta}\right)\right), \lambda > 0, \lambda \neq 1. \tag{3.3}$$

Proof. Using the pdf of the UNXLD in Eq (2.1) and the definition of Tsallis entropy $Re(\lambda) = \frac{1}{\lambda-1}\left(1 - \int_0^\infty f^\lambda(x)\,dx\right)$, $\lambda > 0$, $\lambda \neq 1$, the proof can be obtained. □

Proposition 3.4. *If $X \sim UNXLD$, then for $\beta > 0, \beta \neq 1$ the Havrda and Charvat entropy of X is defined as:*

$$HCe(\beta) = \frac{1}{2^{1-\beta}-1}\left(\frac{\theta^{2\beta}e^{\frac{\beta(\theta-1)+1}{\theta}}}{2^{\beta}(\beta(\theta-1)+1)^{\beta+1}}\Gamma\left(\beta+1, \frac{\beta(\theta-1)+1}{\theta}\right) - 1\right). \tag{3.4}$$

Proof. Using the pdf of the UNXLD in Eq (2.1) and the definition of Havrda and Charvat entropy $HCe(\beta) = \frac{1}{2^{1-\beta}-1}\left(\int_0^\infty f^\beta(x)dx - 1\right)$, $\beta > 0, \beta \neq 1$, the proof can be obtained. □

4. Estimation of parameters

The estimation of the UNXLD parameter is covered in this section. The ML approach is described in Subsection 4.1. The LS and WLS approaches are described in Subsection 4.2. Section 4.3 compares the effectiveness of these techniques using a Monte Carlo simulation analysis.

4.1. Maximum likelihood estimation

Let $X_1, X_2, ..., X_n$ be a random sample of size n taken from the UNXLD with parameter θ. $x_1, x_2, ..., x_n$ are the observed values. Then the likelihood function is given by

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

Then the derivative of the log-likelihood function is given by

$$\frac{\partial}{\partial\theta}\log L(x;\theta) = \frac{n}{\theta} + \sum_{i=1}^{n}\log x_i - \sum_{i=1}^{n}\frac{\log x_i}{1-\theta\log x_i}. \tag{4.1}$$

The MLE of θ is obtained by maximizing the $\log L(x;\theta)$ with respect to θ. Which is done by solving the equation $\frac{\partial}{\partial\theta}\log L(x;\theta) = 0$. Due to the difficulty in solving this, we can use the optim function in R to obtain the MLE numerically.

4.2. Ordinary and weighted least-squares estimation

Let $X_1, X_2, ..., X_n$ be a random sample taken from the UNXLD with parameter θ. Let $X_{1:n}, X_{2:n}, ..., X_{n:n}$ be the order statistics and are denoted by $x_{1:n}, x_{2:n}, ..., x_{n:n}$ the ordered observed data. Let us set

$$R(\theta) = \sum_{i=1}^{n}\left[F(x_{i:n};\theta) - \frac{i}{n+1}\right]^2.$$

Then, the LSE of θ, say $\hat{\theta}_{LS}$, is acquired by minimizing $R(\theta)$ with respect to θ. Practically the LSE is obtained by solving

$$\frac{\partial}{\partial\theta}R(\theta) = 2\sum_{i=1}^{n}\left[F(x_{i:n};\theta) - \frac{i}{n+1}\right]D(x_{i:n},\theta) = 0,$$

where

$$\begin{aligned} D(x_{i:n},\theta) &= \frac{\partial}{\partial\theta}F(x_{i:n};\theta) \\ &= x_{i:n}^{\theta}\log x_{i:n}\left(\frac{1-\theta\log x_{i:n}}{2}\right). \end{aligned}$$

Similarly, the WLS estimate (WLSE) of θ, say $\hat{\theta}_{WLS}$, is acquired by minimizing the non-linear function

$$W(\theta) = \sum_{i=1}^{n}\frac{(n+1)^2(n+2)}{i(n-i+1)}\left[F(x_{i:n};\theta) - \frac{i}{n+1}\right]^2,$$

with respect to θ, that is acquired by solving

$$\frac{\partial}{\partial\theta}W(\theta) = 2\sum_{i=1}^{n}\frac{(n+1)^2(n+2)}{i(n-i+1)}\left[F(x_{i:n};\theta) - \frac{i}{n+1}\right]D(x_{i:n},\theta) = 0.$$

The LSE and WLSE can be evaluated numerically using the optim function in R.

4.3. Simulation study

In the simulation study, the Monte Carlo simulation was carried out to demonstrate the model's efficiency using several estimating techniques, including ML, LS, and WLS. For $N = 10000$ samples of sizes $n = 50, 75, 200, 300, 500$, and 1000, the estimates for the true values of the parameters were determined. These random samples from the UNXLD are generated by applying the inverse cdf to uniformly distributed random numbers. The following quantities were computed:

- Mean of the estimates : $Mean(\eta) = \frac{1}{N} \sum_{i=1}^{N} \eta_i$;
- Average bias of the estimates, $Bias(\eta) = \frac{1}{N} \sum_{i=1}^{N}(\eta_i - \theta)$;
- Mean square error of the estimates, $MSE(\eta) = \frac{1}{N} \sum_{i=1}^{N}(\eta_i - \theta)^2$;
- Root mean square error of the estimates, RMSE$(\eta) = \frac{1}{N} \sum_{i=1}^{N} \frac{|\eta_i - \theta|}{\theta}$,

where $\eta \in (\hat{\theta}_{ML}, \hat{\theta}_{LS}, \hat{\theta}_{WLS})$, i denotes the number of the sample. The simulation results for the ML, LS, and WLS estimation methods are displayed in Tables 1–3. From these tables, it can be concluded that, for MLEs, there is a noticeable decline in both absolute bias and MSE as the sample size increases. Consequently, the performance of MLE proves to be consistently reliable. Figure 5 shows a graphical comparison of the MSEs derived using the three approaches.

Table 1. Simulation results: MLEs, bias, MSE, and RMSE.

	$\theta = 0.6$				$\theta = 1.0$			
n	MLE	Bias	MSE	RMSE	MLE	Bias	MSE	RMSE
50	0.53725	0.06275	0.00394	0.10458	0.92859	0.07141	0.00510	0.07141
75	0.57799	0.02201	0.00048	0.03668	0.96221	0.03779	0.00143	0.03779
200	0.56584	0.03416	0.00117	0.05693	0.95166	0.04834	0.00234	0.04834
300	0.58957	0.01043	0.00011	0.01739	0.98068	0.01932	0.00037	0.01932
500	0.62133	0.02133	0.00046	0.03556	1.04315	0.04315	0.00186	0.04315
1000	0.60301	0.00301	0.00001	0.00502	1.00761	0.00761	0.00006	0.00761
	$\theta = 1.5$				$\theta = 2.1$			
n	MLE	Bias	MSE	RMSE	MLE	Bias	MSE	RMSE
50	1.39245	0.10755	0.01157	0.07170	1.94940	0.15060	0.02268	0.07171
75	1.44324	0.05676	0.00322	0.03784	2.02057	0.07943	0.00631	0.03782
200	1.42729	0.07271	0.00529	0.04848	1.99820	0.10180	0.01036	0.04848
300	1.47088	0.02912	0.00085	0.01942	2.05923	0.04077	0.00166	0.01941
500	1.56445	0.06445	0.00415	0.04297	2.19025	0.09025	0.00814	0.04297
1000	1.51125	0.01125	0.00013	0.00750	2.11575	0.01575	0.00025	0.00750

Table 2. Simulation results: LSEs, bias, MSE, and RMSE.

	$\theta = 0.6$				$\theta = 1.0$			
n	LSE	Bias	MSE	RMSE	LSE	Bias	MSE	RMSE
50	0.58156	0.02189	0.00034	0.03073	0.96916	0.03695	0.00095	0.03084
75	0.56095	0.03777	0.00153	0.06509	0.93485	0.06326	0.00425	0.06515
200	0.55214	0.03869	0.00229	0.07977	0.92020	0.06476	0.00637	0.07980
300	0.58601	0.01329	0.00020	0.02331	0.97663	0.02252	0.00055	0.02337
500	0.62936	0.03077	0.00086	0.04893	1.04890	0.05116	0.00239	0.04890
1000	0.59583	0.00176	0.00002	0.00696	0.99300	0.00315	0.00005	0.00700
	$\theta = 1.5$				$\theta = 2.1$			
n	LSE	Bias	MSE	RMSE	LSE	Bias	MSE	RMSE
50	1.45373	0.05543	0.00214	0.03084	2.03526	0.07757	0.00419	0.03083
75	1.40227	0.09490	0.00955	0.06515	1.96319	0.13284	0.01872	0.06515
200	1.38028	0.09719	0.01433	0.07981	1.93241	0.13605	0.02809	0.07981
300	1.46495	0.03379	0.00123	0.02337	2.05093	0.04731	0.00241	0.02337
500	1.57335	0.07672	0.00538	0.04890	2.20270	0.10743	0.01055	0.04890
1000	1.48950	0.00474	0.00011	0.00700	2.08531	0.00662	0.00022	0.00699

Table 3. Simulation results: WLSEs, bias, MSE, and RMSE.

	$\theta = 0.6$				$\theta = 1.0$			
n	WLSE	Bias	MSE	RMSE	WLSE	Bias	MSE	RMSE
50	0.57811	0.02189	0.00048	0.03649	0.96305	0.03695	0.00137	0.03695
75	0.56223	0.03777	0.00143	0.06296	0.93674	0.06326	0.00400	0.06326
200	0.56131	0.03869	0.00150	0.06448	0.93524	0.06476	0.00419	0.06476
300	0.58671	0.01329	0.00018	0.02216	0.97748	0.02252	0.00051	0.02252
500	0.63077	0.03077	0.00095	0.05129	1.05116	0.05116	0.00262	0.05116
1000	0.59824	0.00176	0.00000	0.00294	0.99685	0.00315	0.00001	0.00315
	$\theta = 1.5$				$\theta = 2.1$			
n	WLSE	Bias	MSE	RMSE	WLSE	Bias	MSE	RMSE
50	1.44457	0.05543	0.00307	0.03695	2.02243	0.07757	0.00602	0.03694
75	1.40510	0.09490	0.00901	0.06326	1.96716	0.13284	0.01765	0.06326
200	1.40281	0.09719	0.00945	0.06479	1.96395	0.13605	0.01851	0.06478
300	1.46621	0.03379	0.00114	0.02253	2.05269	0.04731	0.00224	0.02253
500	1.57672	0.07672	0.00589	0.05114	2.20743	0.10743	0.01154	0.05115
1000	1.49526	0.00474	0.00002	0.00316	2.09338	0.00662	0.00004	0.00315

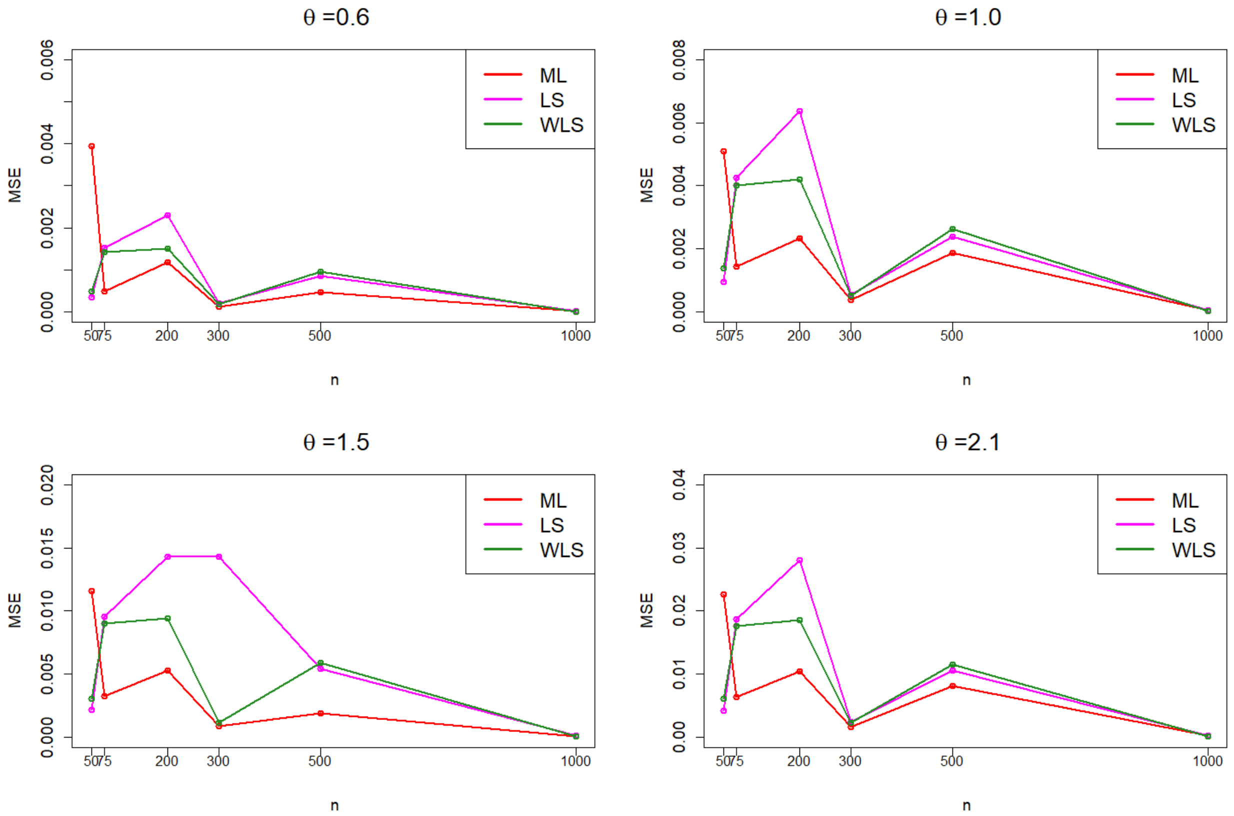


Figure 5. Graphical comparison of the MSEs obtained from ML, LS, and WLS estimation methods for different values of the parameter.

5. Application

5.1. Methodology

In this section, two real-life datasets are used to demonstrate the advantages of the UNXLD. After the data were fitted with the UNXLD, the outcomes were compared to those offered by other probability distributions specified in the unit interval based on the AIC, AICc, CAIC, and BIC together with the MLE. In addition, the KS test and the corresponding $p-$value were used to assess the models' goodness-of-fit, which are listed below

- Akaike information criterion:

$$AIC = 2k - 2\log l;$$

- Akaike information criterion corrected:

$$AICc = AIC + \frac{2k(k+1)}{n-k-1};$$

- Consistent Akaike information criterion:

$$CAIC = -2\log l + k(\log n + 1);$$

- Bayesian information criterion:

$$BIC = k \log n - 2 \log l.$$

Here, $\log l$ denotes the estimated value of the maximum log-likelihood, k denotes the number of parameters and n denotes the number of observations

For comparison, the following probability distributions were taken into account: BD, TLD proposed by [27], KD suggested by [28], ETLD introduced by [29], and UBD suggested by [6].

The pdf of the compared distributions that are used to compare with the UNXLD on the unit interval $(0, 1)$.

- Beta distribution:

$$f(x; \alpha; \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1 - x)^{\beta-1}, x \in (0, 1), \alpha > 0, \beta > 0.$$

- Topp-Leone distribution:

$$f(x; \alpha) = 2\alpha(1 - x)x^{\alpha-1}(2 - x)^{\alpha-1},\ \alpha > 0.$$

- Kumaraswamy distribution:

$$f(x; \alpha; \beta) = \alpha\beta x^{\alpha-1}(1 - x^{\alpha})^{\beta-1},\ \alpha > 0, \beta > 0.$$

- Exponentiated Topp-Leone distribution:

$$f(x; \alpha; \beta) = 2\alpha\beta(1 - x)(x(2 - x))^{\alpha-1}(1 - x^{\alpha}(2 - x)^{\alpha})^{\beta-1},\ \alpha > 0, \beta > 0.$$

- Unit Burr-II distribution:

$$f(x; \alpha; \beta) = \alpha\beta x^{-1}(-\log x)^{\beta-1}(1 + (-\log x)^{\beta})^{-\alpha-1}, x \in (0, 1),\ \alpha > 0, \beta > 0.$$

5.2. Data description and inference

The first dataset here considered relates to a comparison of the SC 16 and P3 unit capacity factor estimation techniques. [30] and [31] has already been investigated this dataset. The data are displayed in Table 4.

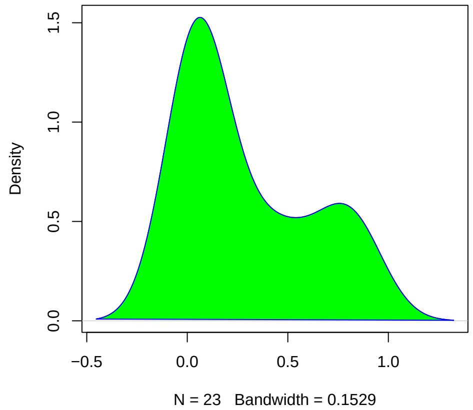
Table 4. The observations of the dataset 1.

0.853	0.759	0.866	0.809	0.717	0.544	0.492	0.403
0.344	0.213	0.116	0.116	0.092	0.070	0.059	0.048
0.036	0.029	0.021	0.014	0.011	0.008	0.006	

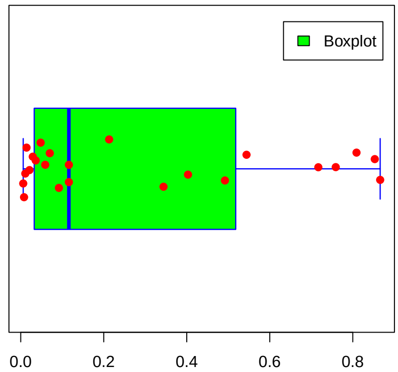
The second set of data comprises the initial 58 observations documenting the time of failure for Kevlar 49/epoxy strands tested at a stress level of 90%. This data was used by [32] and is presented in Table 5. Figures 6–9 show the density plot, QQ plot, PP plot, box plot, estimated pdfs, and TTT plot proposed by [33] may be used for obtaining empirical behavior of the hrf for both datasets considered here.

Table 5. The observations of the dataset 2.

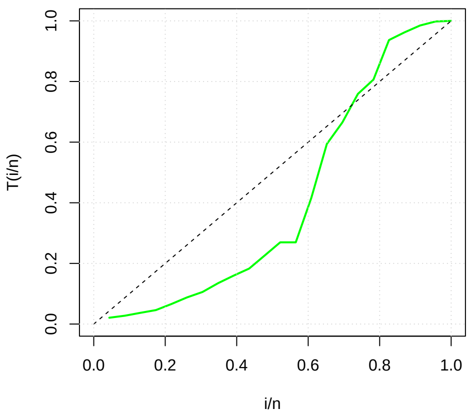
0.01	0.01	0.02	0.02	0.02	0.03	0.03	0.04	0.05
0.06	0.07	0.07	0.08	0.09	0.09	0.1	0.1	0.11
0.11	0.12	0.13	0.18	0.19	0.2	0.23	0.24	0.24
0.29	0.34	0.35	0.36	0.38	0.4	0.42	0.43	0.52
0.54	0.56	0.6	0.6	0.63	0.65	0.67	0.68	0.72
0.72	0.72	0.73	0.79	0.79	0.8	0.8	0.83	0.85
0.9	0.92	0.95	0.99	0.01				



(a) Density

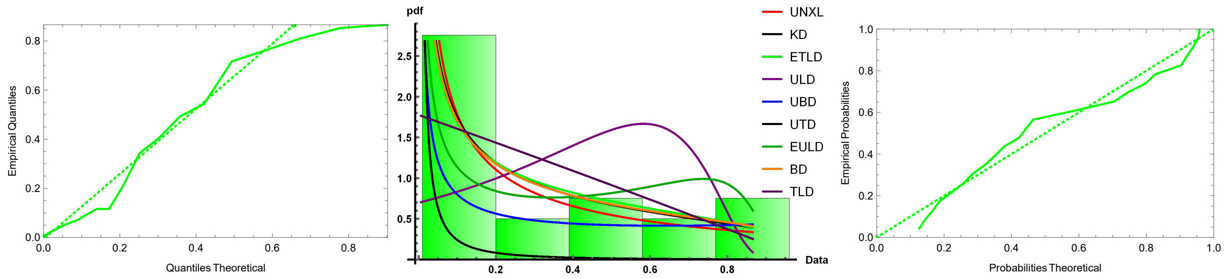


(b) Box



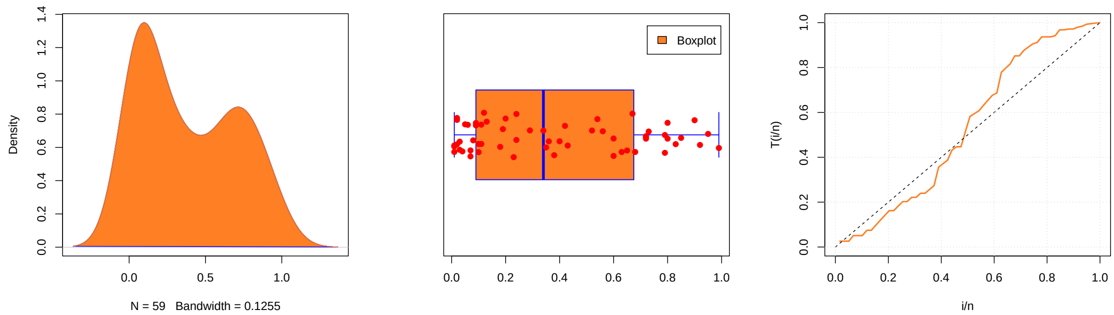
(c) TTT

Figure 6. The density, box, and TTT plots based on the first dataset.



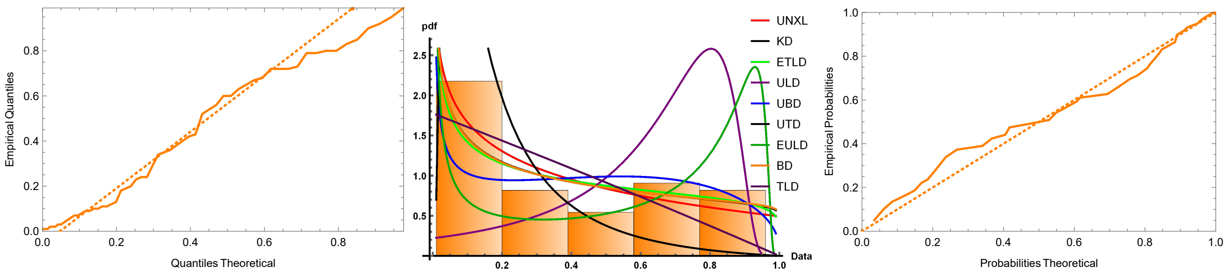
(a) QQ (b) Est.pdfs (c) PP

Figure 7. The QQ, PP, and estimated pdfs based on the first dataset.



(a) Density (b) Box (c) TTT

Figure 8. The density, box, histogram and TTT plots based on the second dataset.



(a) QQ (b) Est.pdfs (c) PP

Figure 9. The QQ, PP, and estimated pdfs based on the second dataset.

The results for dataset 1 and dataset 2 are summarized in Tables 6 and 7. These tables specifically provide the log-likelihood value, MLEs, measures AIC, AICc, CAIC, BIC, and the value of the KS statistic together with the corresponding $p-$value for each fitted model. The UNXLD offers a better fit with the smallest values for the AIC, AICc, CAIC, and BIC criteria, as well as the smallest value for the KS statistic and the largest $p-$value, as can be observed.

Table 6. The MLE and goodness-of-fit results of the dataset 1.

Model	$\log L$	ML estimates	S.Es	AIC	AICc	CAIC	BIC	KS statistic	$p-value$
UNXLD	9.9915	$\hat{\theta}=0.6824$	0.1244	-17.9830	-17.7925	-15.8475	-16.8475	0.1663	0.5483
UBD	5.8728	$\hat{\alpha}=0.7848$ $\hat{\beta}=1.5049$	0.1919 0.2951	-7.745593	-7.145593	-3.474604	-5.474604	0.2394	0.1433
ETLD	9.3913	$\hat{\alpha}=0.4624$ $\hat{\beta}=0.6806$	0.1389 0.1717	-14.7827	-14.1827	-10.5117	-12.5117	0.1933	0.3562
TLD	8.1151	$\hat{\alpha}=0.5943$	0.1239	-14.2303	-14.0398	-12.0948	-13.0948	0.1690	0.5272
BD	9.6075	$\hat{\alpha}=0.4869$ $\hat{\beta}=1.1679$	0.1208 0.3578	-15.2149	-14.6149	-10.9439	-12.9439	0.1836	0.4203
KD	9.6708	$\hat{\alpha}=0.5044$ $\hat{\beta}=1.1862$	0.1288 0.3265	-15.3416	-14.7416	-11.0706	-13.0706	0.1790	0.4529

Table 7. The MLE and goodness-of-fit results of the dataset 2.

Model	$\log L$	ML estimates	S.Es	AIC	AICc	CAIC	BIC	KS statistic	$p-value$
UNXLD	5.7959	$\hat{\theta}=0.9861$	0.1138	-9.5917	-9.5203	-6.5313	-7.5313	0.1013	0.5911
UBD	0.4600	$\hat{\alpha}=1.1262$ $\hat{\beta}=1.3741$	0.1581 0.1571	3.0799	3.2981	9.2008	7.2008	0.1304	0.2778
ETLD	5.4371	$\hat{\alpha}=0.6450$ $\hat{\beta}=0.5639$	0.1305 0.0879	-6.8742	-6.6560	-0.7533	-2.7533	0.1107	0.4765
BD	5.6714	$\hat{\alpha}=0.6777$ $\hat{\beta}=1.0412$	0.1107 0.1873	-7.3427	-7.1245	-1.2218	-3.2218	0.1038	0.5592
KD	5.6824	$\hat{\alpha}=0.6825$ $\hat{\beta}=1.0472$	0.1145 0.1794	-7.3648	-7.1467	-1.2439	-3.2439	0.1036	0.5628

6. Conclusions

By using an exponential transformation, we established a bounded form of the NXLD in this study called the UNXLD. Certain important distributional properties, such as the behavior of the pdf and hrf, along with some tractable statistical properties such as moments, incomplete moments, mode, and quantile function, are proposed. Moreover, all the statistical measures are in closed form. The ML approach, LS, and WLS were used to estimate the model parameters, and simulation studies were used to test the estimates. The UNXLD can be thought of as a good contender in distributions in unit interval, when the dominance of the proposed model has been demonstrated using two real datasets. In future work, it would be intriguing to identify the quantile function and extend the research to develop quantile regression models.

Author contributions

M. R. Irshad, S. Aswathy, R. May: Conceptualization, methodology, software, validation, formal analysis, writing-original draft preparation, writing-review and editing, visualization; Amer I. Al-Omari, Ghadah Alomani: Conceptualization, validation, formal analysis, writing-original draft preparation, writing-review and editing, visualization. All authors have read and approved the final version of the manuscript for publication.

Acknowledgments

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R226), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Availability of data and material

The data supporting the findings of this study are available within the article.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

There is no conflict of interest with the publication of this work.

References

1. E. S. A. El-Sherpieny, M. A. Ahmed, On the Kumaraswamy Kumaraswamy distribution, *Int. J. Basic Appl. Sci.*, **3** (2014), 372–381. https://doi.org/10.14419/ijbas.v3i4.3182
2. E. Altun, G. G. Hamedani, The log-xgamma distribution with inference and application, *J. Soc. Fr. Stat.*, **159** (2018), 40–55.
3. J. Mazucheli, S. R. Bapat, A. F. B. Menezes, A new one-parameter unit-Lindley distribution, *Chilean J. Stat.*, **11** (2020), 53–67.
4. J. Mazucheli, A. F. Menezes, S. Dey, Unit-Gompertz distribution with applications, *Statistica*, **79** (2019), 25–43. https://doi.org/10.6092/issn.1973-2201/8497
5. J. Mazucheli, A. F. B. Menezes, L. B. Fernandes, R. P. De Oliveira, M. E. Ghitany, The unit-Weibull distribution as an alternative to the Kumaraswamy distribution for the modeling of quantiles conditional on covariates, *J. Appl. Stat.*, **47** (2020), 954–974. https://doi.org/10.1080/02664763.2019.1657813
6. K. Modi, V. Gill, Unit Burr-III distribution with application, *J. Stat. Manag. Syst.*, **23** (2020), 579–592. https://doi.org/10.1080/09720510.2019.1646503

7. R. A. R. Bantan, C. Chesneau, J. Farrukh, M. Elgarhy, M. H. Tahir, A. Ali, et al., Some new facts about the unit-Rayleigh distribution with applications, *Mathematics*, **8** (2020), 1954. https://doi.org/10.3390/math8111954

8. M. Ç. Korkmaz, C. Chesneau, On the unit Burr-XII distribution with the quantile regression modeling and applications, *Comput. Appl. Math.*, **40** (2021), 29. https://doi.org/10.1007/s40314-021-01418-5

9. H. S. Bakouch, A. S. Nik, A. Asgharzadeh, H. S. Salinas, A flexible probability model for proportion data: Unit-half-normal distribution, *Commun. Stat. Case Stud. Data Anal. Appl.*, **7** (2021), 271–288. https://doi.org/10.1080/23737484.2021.1882355

10. M. Irshad, V. Dcruz, R. Maya, The exponentiated unit Lindley distribution: Properties and applications, *Ricerche Mat.*, **73** (2021), 1121–1143. https://doi.org/10.1007/s11587-021-00663-4

11. A. Krishna, R. Maya, C. Chesneau, M. R. Irshad, The unit Teissier distribution and its applications, *Math. Comput. Appl.*, **27** (2022), 12. http://dx.doi.org/10.3390/mca27010012

12. A. I. Al-Omari, A. R. Alanzi, S. S. Alshqaq, The unit two parameters Mirra distribution: Reliability analysis, properties, estimation and applications, *Alex. Eng. J.*, **92** (2024), 238–253. https://doi.org/10.1016/j.aej.2024.02.063

13. R. Maya, P. Jodra, M. Irshad, A. Krishna, The unit Muth distribution: statistical properties and applications, *Ricerche Mat.*, **73** (2022), 1843–1866. https://doi.org/10.1007/s11587-022-00703-7

14. N. Khodja, A. M. Gemeay, H. Zeghdoudi, K. Karakaya, A. M. Alshangiti, M. E. Bakr, et al., Modeling voltage real data set by a new version of Lindley distribution, *IEEE Access*, **11** (2023), 67220–67229. https://doi.org/10.1109/ACCESS.2023.3287926

15. A. Beghriche, H. Zeghdoudi, V. Raman, S. Chouia, New polynomial exponential distribution: Properties and applications, *Statist. Transit.*, **23** (2022), 95–112. https://doi.org/10.2478/stattrans-2022-0032

16. A. A. Abd EL-Baset, M. Ghazal, Exponentiated additive Weibull distribution, *Reliab. Eng. Syst. Saf.,* **193** (2020), 106663. http://dx.doi.org/10.1016/j.ress.2019.106663

17. M. Irshad, D. Shibu, R. Maya, V. Dcruz, Binominal mixture Lindley distribution: Properties and applications, *J. Indian Soc. Probab. Stat.*, **21** (2020), 437–469. https://doi.org/10.1007/s41096-020-00090-y

18. A. Khalil, M. Ijaz, K. Ali, W. K. Mashwani, M. Shafiq, P. Kumam, et al., A novel flexible additive Weibull distribution with real-life applications, *Comm. Statist. Theory Methods*, **50** (2021), 1557–1572. https://doi.org/10.1080/03610926.2020.1732658

19. M. E. Ghitany, J. Mazucheli, A. F. B. Menezes, F. Alqallaf, The unit-inverse Gaussian distribution: A new alternative to two-parameter distributions on the unit interval, *Commun. Stat. Theory Methods*, **48** (2019), 3423–3438. https://doi.org/10.1080/03610926.2018.1476717

20. M. Ç. Korkmaz, A new heavy-tailed distribution defined on the bounded interval: The logit Slash distribution and its application, *J. Appl. Stat.*, **47** (2020), 2097–2119. https://doi.org/10.1080/02664763.2019.1704701

21. Y. A. Iriarte, M. de Castro, H. W. Gómez, An alternative one-parameter distribution for bounded data modeling generated from the Lambert transformation, *Symmetry*, **13** (2021), 1190. http://dx.doi.org/10.3390/sym13071190

22. J. Mazucheli, M. Ç. Korkmaz, A. F. B. Menezes, V. Leiva, The unit generalized half-normal quantile regression model: Formulation, estimation, diagnostics, and numerical applications, *Soft Comput.*, **27** (2023), 279–295. http://dx.doi.org/10.1007/s00500-022-07278-3

23. F. Lad, G. Sanfilippo, G. Agro, Extropy: Complementary dual of entropy, *Statist. Sci.*, **30** (2015), 40–58. http://doi.org/10.1214/14-STS430

24. C. Tsallis, Possible generalization of Boltzmann-Gibbs statistics, *J. Stat. Phys.*, **52** (1988), 479–487. http://dx.doi.org/10.1007/BF01016429

25. A. Rényi, On measures of entropy and information, *Berkeley Symp. Math. Statist. Prob.*, **4.1** (1961), 547–562.

26. J. Havrda, F. Charvat, Quantification method of classification processes, concept of structural α-entropy, *Kybernetika*, **3** (1967), 30–35.

27. A. Al-Shomrani, O. Arif, A. Shawky, S. Hanif, M. Q. Shahbaz, Topp-Leone family of distributions: Some properties and application, *Pak. J. Stat. Oper. Res.*, **12** (2016), 443–451. http://dx.doi.org/10.18187/pjsor.v12i3.1458

28. P. Kumaraswamy, A generalized probability density function for double-bounded random processes, *J. Hydrol.*, **46** (1980), 79–88. https://doi.org/10.1016/0022-1694(80)90036-0

29. A. Pourdarvish, S. M. T. K. Mirmostafaee, K. Naderi, The exponentiated Topp-Leone distribution: Properties and application, *J. Appl. Environ. Biol. Sci.*, **5** (2015), 251–256.

30. M. Caramanis, J. Stremel, W. Fleck, S. Daniel, Probabilistic production costing: An investigation of alternative algorithms, *Int. J. Elec. Power Energy Syst.*, **5** (1983), 75–86. https://doi.org/10.1016/0142-0615(83)90011-X

31. M. Mazumdar, D. P. Gaver, On the computation of power-generating system reliability indexes, *Technometrics*, **26** (1984), 173–185. https://doi.org/10.1080/00401706.1984.10487942

32. P. Sudsila, A. Thongteeraparp, S. Aryuyuen, W. Bodhisuwan, The generalized distributions on the unit interval based on the T-Topp-Leone family of distributions, *Trends Sci.*, **19** (2022), 6186–6186. https://doi.org/10.48048/tis.2022.6186

33. M. V. Aarset, How to identify bathtub hazard rate, *IEEE Trans. Reliab.*, **36** (1987), 106–108. https://doi.org/10.1109/TR.1987.5222310

A. Appendix

The R-code for the empirical study of UNXLD is given below.

```
library(fitdistrplus)
data<-NULL
n<-length(data)
n
### PDF ###
```

```
duxl <- function(x,theta){
((theta/2)*(x^(theta-1))*(1-(theta*log(x))))
}

### CDF ###
puxl<- function(q,theta){
((q^theta)-((theta*(q^theta)*(log(q)))/(2)))
}
prefit(x, "uxl", "mle", list(theta=initial), lower=c(0), upper=c(Inf))

est <- fitdist(x,"uxl", start=list(theta=initial),optim.method="Nelder-Mead")
est
summary(est)
ks.test(x,"puxl",est$estimate)
logl=est$loglik
AIC=2*1-2*logl
AIC
BIC=-2*logl+1*(log(length(x)))
BIC
k=number of parameter
CAIC=-2*logl+k*(log(length(x))+1)
CAIC
AICC=AIC+(2*k*(k+1)/(length(x)-k-1))
AICC
```

AIMS Press

© 2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)