A Unit Probabilistic Model for Proportion and Asymmetric Data: Properties and Estimation Techniques with Application to Model Data from SC16 and P3 Algorithms

Mohamed S. Eliwa,1,2 Muhammad Ahsan-ul-Haq,3,4 Afrah Al-Bossly,5 and Mahmoud El-Morshedy6

1Department of Statistics and Operation Research, College of Science, Qassim University, P.O. Box 6644, Buraydah 51482, Saudi Arabia
2Department of Mathematics, Faculty of Science, Mansoura University, Mansoura 35516, Egypt
3College of Statistical and Actuarial Sciences, University of the Punjab, Lahore, Pakistan
4Quality Enhancement Cell, National College of Arts, Lahore, Pakistan
5Department of Mathematics, College of Science and Humanities in Al-Kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia
6Department of Statistics and Computer Science, Faculty of Science, Mansoura University, Mansoura 35516, Egypt

Correspondence should be addressed to Mahmoud El-Morshedy; m.elmorshedy@psau.edu.sa

Received 11 October 2021; Accepted 1 February 2022; Published 11 March 2022

Academic Editor: Dost Muhammad Khan

Copyright © 2022 Mohamed S. Eliwa et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this study, a new one-parameter Log-XLindley distribution is proposed to analyze the proportion data. Some of its statistical and reliability properties, including moments with associated measures, hazard rate function, reversed hazard rate, stress strength reliability, and mean residual life function, are investigated in closed forms which help the researchers for modeling data in a small CPU time. It is found that the density function of the introduced distribution can be used as a statistical tool to model asymmetric data. Moreover, the failure rate function can be utilized to model different types of failures, including increasing, bathtub, and J-shaped. The model parameter is estimated using various estimation approaches to get the best estimator to help us in modeling the real data in a good way with high accuracy. A Monte-Carlo simulation study for different sample sizes is performed to assess the performance of the estimations based on some statistical criteria. Finally, two distinctive data sets from SC16 and P3 algorithms, "estimating unit capacity factors," are analyzed to illustrate the flexibility of the new model.

1. Introduction

Researchers have demonstrated a strong interest in developing new extended distributions by adding shape parameters to baseline distributions during the last decade. The primary goal of this research is to improve the modeling abilities of distributions and provide new opportunities to model various data set features. Unbounded support has received a lot of attention from researchers. However, in many real-life circumstances, such as percentages and proportions, observations can only take values within a limited range [1].

Among the unit distributions, the beta distribution is the most well-known distribution. It is frequently utilized in different fields of research, such as economics, biology, and medical sciences. The major flaw of the beta distribution is that its cumulative distribution function (CDF) cannot be written in a closed explicit form. Thus, the researchers proposed and studied various unit distributions. Among the most useful unit distributions, there are the Johnson distribution [2], Topp–Leone distribution [3], unit-gamma distribution [4], Kumaraswamy distribution [5], log-Lindley distribution [6], unit-logistic distribution [7], unit-Birnbaum–Saunders distribution [8], log-xgamma distribution
[9], unit-Lindley distribution [10], unit-Gompertz distribution [11], unit-inverse Gaussian distribution [12], unit-Burr III distribution [13], log-weighted exponential distribution [14], unit-Weibull distribution [15], unit-Modified Burr III distribution [16], unit-Rayleigh distribution [17], Frechet power function distribution [18], and unit-Burr XIII distribution [19].

The purpose of this work is to introduce a new distribution for modeling data sets on the unit interval. To achieve this aim, the XLindley distribution is used to construct a new model. The proposed distribution is entitled the Log-XLindley distribution with one positive shape parameter. The Log-XLindley distribution offers many benefits over well-known unit interval distributions, including the beta and Kumaraswamy distributions. The Log-XLindley distribution is better due to its simple structure and flexibility via hazard rate. The statistical properties, including moments, skewness, kurtosis, stress strength reliability, and mean residual life, maybe derived in explicit forms.

The paper is organized as follows: In Section 2, we propose the Log-XLindley distribution. The statistical characteristics are derived in Section 3. In Section 4, we derived the reliability properties of the proposed distribution. Different computation techniques are used to estimate the model parameter in Section 5. In Section 6, Monte-Carlo simulation analysis is carried out to assess the parameter estimation techniques' finite sample performance. Two real data sets, “SC16 and P3 algorithms: estimating unit capacity factors,” are analyzed over the interval (0, 1) to show the Log-XLindley distribution’s flexibility in Section 7. Finally, some remarks are reported based on the proposed model in Section 8.

2. The Log-XLindley Distribution

A random variable $Y$ is said to have the XLindley distribution with parameter shape $\theta > 0$ if its density function can be expressed as

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

The Log-XLindley distribution is derived from the XLindley distribution using the logarithmic transformation of type $X = e^{-Y}$, where $Y$ is the XLindley distribution. The probability density function (PDF) with support $(0, 1)$ is given by

$$f(x) = \frac{\theta^2 (2 + \theta - \log x)x^{\theta-1}}{(1 + \theta)^2}; \quad 0 < x < 1, \quad (2)$$

where $\theta > 0$ is the shape parameter. The corresponding CDF to (2) can be formulated as

$$F(x) = \frac{x^\theta [(1 + \theta)^2 - \theta \log x]}{(1 + \theta)^2}; \quad 0 < x < 1. \quad (3)$$

For the Log-XLindley distribution, the limiting behavior of PDF at lower and upper limits is

$$\lim_{x \to 0^+} f(x) = \begin{cases} \infty, & \text{for } \theta \leq 1, \\ 0, & \text{for } \theta > 1, \end{cases} \quad (4)$$

$$\lim_{x \to 1^-} f(x) = \frac{\theta^2 (2 + \theta)}{(1 + \theta)^2}. \quad (5)$$

Figure 1 illustrates some plots of the Log-XLindley distribution with one specific values of the parameter $\theta$.

It is noted that the density function can be used as a probability tool to discuss and analyze asymmetric data. Moreover, the shape of the density function can be either decreasing or uni-modal, which makes the proposed model can be used effectively in modeling different types of data sets in various fields.

3. Statistical Properties

3.1. Moments and Incomplete Moments with Associated Measures. If the random variables $X$ have the Log-XLindley distribution with parameter $\theta > 0$, then the ordinary moments (OM) can be expressed in an explicit form as follows:

$$\mu'_r = E(X^r) = \int_0^1 x^r f(x)dx$$

$$= \int_0^1 x^r \frac{\theta^2 (2 + \theta - \log x)x^{\theta-1}}{(1 + \theta)^2}dx$$

$$= \frac{\theta^2}{(1 + \theta)^2} \int_0^1 (2 + \theta - \log x)x^{\theta-1}dx$$

$$= \frac{\theta^2}{(1 + \theta)^2} \left\{ \int_0^1 (2 + \theta)x^{\theta-1}dx - \int_0^1 x^{\theta-1}\log x dx \right\}. \quad (6)$$

After simple computations, the OM can be expressed as

$$E(X^r) = \frac{\theta^2 [(2 + r)\theta + 2r + 1]}{(1 + \theta)^2(r + \theta)^2}. \quad (7)$$

The moments of the origin can be obtained by substituting $r = 1, 2, 3, 4$, respectively. The first four moments around the origin can be expressed as

$$\mu'_1 = E(X) = \frac{\theta^2 [\theta^2 + 3\theta + 3]}{(1 + \theta)^4},$$

$$\mu'_2 = E(X^2) = \frac{\theta^2 [\theta^2 + 4\theta + 5]}{(1 + \theta)^4(2 + \theta)},$$

$$\mu'_3 = E(X^3) = \frac{\theta^2 [\theta^2 + 5\theta + 7]}{(1 + \theta)^4(3 + \theta)^2},$$

$$\mu'_4 = E(X^4) = \frac{\theta^2 [\theta^2 + 6\theta + 9]}{(1 + \theta)^4(4 + \theta)^2}. \quad (8)$$
Thus, the variance and index of dispersion of the proposed model can be formulated as

\[
\text{Var}(X) = \frac{\theta^2}{(1 + \theta)^2} \left\{ \frac{\theta^2 + 4\theta + 5}{(2 + \theta)^2} - \frac{\theta^2(\theta^2 + 3\theta + 3)^2}{(1 + \theta)^6} \right\},
\]

(9)

\[
\text{D}(X) = \frac{\theta^2(\theta^2 + 3\theta + 3)(1 + \theta)^4}{\theta^2/(1 + \theta)^2} \left\{ \frac{\theta^2 + 4\theta + 5}{(2 + \theta)^2} - \theta^2(\theta^2 + 3\theta + 3)^2/(1 + \theta)^6 \right\},
\]

(10)

respectively. Based on the \(\text{D}(X)\) measure, the introduced model can be used to model dispersion data. Moreover, the coefficient of skewness and kurtosis can be obtained by using well-known relations via the OM property. If the random variables \(X\) have the Log-XLindley distribution with parameter \(\theta > 0\), then the incomplete moments (ICM) can be expressed as

\[
\phi(X^r) = \int_0^t x^r f(x) \, dx.
\]

(11)

\[
\phi(X^r) = \int_0^t x^r \frac{x^2}{(1 + \theta)^2} (2 + \theta - \log x)x^{\theta-1} \, dx
\]

\[
= \frac{\theta^2}{(1 + \theta)^2} \int_0^t (2 + \theta - \log x)x^{\theta-1} \, dx
\]

\[
= \frac{\theta^2}{(1 + \theta)^2} \left\{ \int_0^t (2 + \theta)x^{\theta-1} \, dx - \int_0^t x^{\theta-1}\log x \, dx \right\}
\]

\[
= \frac{\theta^2}{(1 + \theta)^2} \left\{ \frac{(2 + \theta)x^{\theta}}{\theta} - \frac{t^{\theta+1}}{(r + \theta)} + \frac{t^{\theta-1}}{(r + \theta)^2} \right\}.
\]

(12)
Using Equation (2), we get
After simple algebra, the ICM can be formulated as
\[
\psi(X) = \frac{\theta^2 (2 + (2 + r - \log x) - \log x)}{1 + \theta^2 - \theta \log x}; \quad 0 < x < 1.
\]
(13)

The ICM can be used to measure inequality, including income quintiles, the Lorenz curve, Pietra, and Gini measures of inequality, among others.

4. Reliability Measures

4.1. Hazard (Reversed) Rate, Cumulative Hazard Function, and Mills Ratio. If the random variable \(X\) has the Log-XLindley distribution, then the survival function (SF) and its hazard rate can be written, respectively, as follows:

\[
S(x) = 1 - \frac{x^\theta (1 + \theta)^2 - \theta \log x}{(1 + \theta)^2}; \quad 0 < x < 1,
\]
(14)

\[
h(x) = \frac{\theta^2 (2 + \theta - \log x)x^{\theta - 1}}{(1 + \theta)^2 - x^\theta (1 + \theta)^2 - \theta \log x}; \quad 0 < x < 1.
\]
(15)

Mathematically, the hazard rate function (HRF) of the proposed model is bathtub-shaped at \(\theta < 1\), whereas the HRF can take J- and increasing-shaped at \(\theta = 1\) and \(\theta > 1\), respectively. Figure 2 shows some plots of the HRF based on some determined values of the parameter \(\theta\).

Regarding Figure 2, it is noted that the HRF can take different shapes, including bathtub, J-shaped, and increasing. Thus, the HRF of the proposed model can be used to discuss various types of data in different fields. The reversed hazard rate is

\[
r(x) = \frac{\theta^2 (2 + \theta - \log x)}{x[1 + (\theta)^2 - \theta \log x]}; \quad 0 < x < 1.
\]
(16)

The cumulative hazard and Mills ratio can be expressed, respectively, as

\[
H(x) = -\log \left(\frac{(1 + \theta)^2 - x^\theta (1 + \theta)^2 - \theta \log x}{1 + \theta}\right); \quad 0 < x < 1,
\]
(17)

\[
M(x) = \frac{(1 + \theta)^2 - x^\theta (1 + \theta)^2 - \theta \log x}{\theta^2 (2 + \theta - \log x)x^{\theta - 1}}; \quad 0 < x < 1.
\]
(18)

4.2. Stress-Strength Reliability (SSR)

Theorem 1. If the random variables \(Y_1\) and \(Y_2\) have the Log-XLindley distribution with parameters \(\theta > 0\) and \(\alpha > 0\), respectively, then the SSR can be formulated in an explicit form as
\textit{Proof.} The SSR can be derived from the following relation:
\begin{equation}
\text{SSR} = P(Y_1 < Y_2)
= \int_0^1 f_1(x)F_2(x)dx.
\end{equation}

Using (2) and (3), the SSR can be written as
\begin{equation}
\text{SSR} \equiv \frac{\theta^2}{(1 + \theta)^2 (1 + \alpha)^2} \int_0^1 x^{\alpha + \theta - 1} \left[ (1 + \alpha)^2 - \alpha \log x \right] (2 + \theta - \log x)dx,
\end{equation}

After integration and simplification, the expression of the SSR can be expressed as (19).

\section*{4.3. Mean Residual Life (MRL).} If the random variable \( X \) has the Log-XLindley distribution with parameter \( \theta > 0 \), then the MRL can be expressed in an explicit form as follows:
\begin{equation}
m(t) = \frac{1}{1 - t^\theta (1 + \theta)^{-2} \left[ (1 + \theta)^2 - \theta \log t \right]} \left[ \int_t^1 x^\theta \left[ (1 + \theta)^2 - \theta \log x \right] \frac{dx}{(1 + \theta)^2} \right],
\end{equation}

After simple modification, the MRL can be expressed as
\begin{equation}
m(t) = \frac{(1 - t) - (1 - t^{\theta+1}/\theta + 1) + \theta/(1 + \theta)^2 \left[ t^{\theta+1} \log t/(\theta + 1) - 1 - t^{\theta+1}/(1 + \theta)^2 \right]}{1 - t^\theta (1 + \theta)^{-2} \left[ (1 + \theta)^2 - \theta \log t \right]}.
\end{equation}

Mathematically, the MRL of the new model is unimodal-shaped at \( \theta < 1 \), whereas the MRL can take inverse J- and decreasing-shaped at \( \theta = 1 \) and \( \theta > 1 \), respectively.

\section*{5. Estimation Techniques}

\subsection*{5.1. Maximum Likelihood Estimator.} Let \( x_1, x_2, \ldots, x_n \) be a random sample from the Log-XLindley distribution; then, the log-likelihood function is given by
\begin{equation}
I(\theta) = 2n \log \theta - 2n \log(1 + \theta) + (\theta - 1) \sum_{i=1}^n \log x_i + \sum_{i=1}^n \log(2 + \theta - \log x_i).
\end{equation}

Taking the partial derivative to (25) with respect to the parameter \( \theta \), the following equation is obtained:
\begin{equation}
\frac{\partial l(\theta)}{\partial \theta} = \frac{2n}{\theta} - \frac{2n}{(1 + \theta)} + \sum_{i=1}^n \frac{\log x_i + \frac{1}{2} \sum_{i=1}^n \frac{1}{2 + \theta - \log x_i}}{(2 + \theta - \log x_i)}.
\end{equation}

The maximum likelihood estimator of \( \theta \) is derived by solving the nonlinear equation \( \partial l(\theta)/\partial \theta = 0 \). The resulted equation cannot be solved without using a numerical approach like the Newton–Raphson method.

\subsection*{5.2. Least Squares (LS) and Weighted Least Squares (WLS) Estimators.} Let \( x_1 < x_2 < x_3 < \cdots < x_n \) be an ordered sample of size \( n \) from the Log-XLindley distribution. Then, the LS
estimator (LSE) of the Log-XLindley parameter can be derived by minimizing
\[
\text{LSE}(\theta) = \sum_{i=1}^{n} \left( F(x_{(i:n)}|\theta) - \frac{i}{n+1} \right)^2,
\]
with respect to the parameter \(\theta\). The WLS estimate (WLSE) of \(\theta\), say \(\hat{\theta}_{\text{WLSE}}\), can be determined by minimizing
\[
\text{WLSE}(\alpha) = \sum_{i=1}^{n} \left( \frac{(n+1)^2(n+2)}{i(n-i+1)} \left[ F(x_{(i:n)}|\theta) - \frac{i}{n+1} \right] \right)^2,
\]
with respect to \(\alpha\).

5.3. Anderson–Darling (AD) and Right Tail Anderson–Darling (RAD) Estimators. The AD estimator (ADE) is a minimum distance-based estimator. It can be obtained by minimizing
\[
\text{ADE}(\theta) = -n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1) \left[ \log(F(x_{(i)}|\theta)) \right] + \log(1 - F(x_{(i)}|\theta)),
\]
with respect to the parameter \(\theta\), whereas the RAD estimator (RADE) of the model parameter can be derived by minimizing
\[
\text{RADE}(\theta) = \frac{n}{2} - \frac{1}{2} \sum_{i=1}^{n} F(x_{(i:n)}|\theta)
- \frac{1}{n} \sum_{i=1}^{n} (2i - 1) \left[ \log(1 - F(x_{(n+1-i:n)}|\theta)) \right],
\]
with respect to the parameter \(\theta\).

5.4. Cramer–Von Mises Estimator (CVME). The CVME is a minimum distance-based estimator. The CVME of the Log-XLindley distribution can be obtained by minimizing
\[
\text{CVME}(\theta) = \frac{1}{12n} + \sum_{i=1}^{n} \left[ \log(F(x_{(i)}|\theta)) - \frac{2i-1}{2n} \right]^2,
\]
with respect to the parameter \(\theta\).

5.5. Maximum Product of Spacing Estimator (MPSE). For \(r = 1, 2, 3, \ldots, n+1\), let \(K_r(\theta) = F(x_{(r)}|\theta) - F(x_{(r-1)}|\theta)\) be the uniform spacing of a random sample from the Log-XLindley distribution, and \(F(x_{(r)}|\theta) = 0\), \(F(x_{(n+1)}|\theta) = 1\), and \(\sum_{r=1}^{n+1} K_r(\theta) = 1\). The MPSE of parameter \(\theta\), say \(\hat{\theta}_{\text{MPSE}}\), can be derived by maximizing the geometric mean of the spacings:
\[
W(\theta) = \left[ \prod_{r=1}^{n+1} K_r(\theta) \right]^{1/(n+1)}
\]
with respect to the parameter \(\theta\).

6. Monte-Carlo Simulation

To assess the performance of the estimators listed in the previous section, we conducted a comprehensive simulation study. We used the Log-XLindley distribution to
generate samples with \( n = 10, 25, 50, 100, 150, \) and 200 and then calculated the average values (AVEs) of the estimators to get the mean square errors (MSEs), average absolute biases (ABBs), and mean relative errors (MREs) for \( \theta = 0.5, 1.0, 1.5, 2.0, \) and 5.0. The ABBs, MREs, and MSEs are given by

| \( n \) | Measure | MLE | LSE | WLE | CVME | ADE | RADE | MPSE |
|---|---|---|---|---|---|---|---|---|
| 10 | AVEs | 1.0837 | 1.0606 | 1.0591 | 1.0682 | 1.0498 | 1.1083 | 0.9891 |
| 25 | | 1.0282 | 1.0207 | 1.0166 | 1.0224 | 1.0163 | 1.0404 | 0.9802 |
| 50 | | 1.0139 | 1.0141 | 1.0098 | 1.0117 | 1.0109 | 1.0154 | 0.9871 |
| 100 | | 1.0058 | 1.0061 | 1.0054 | 1.0073 | 1.0034 | 1.0102 | 0.9896 |
| 150 | | 1.0041 | 1.0046 | 1.0038 | 1.0056 | 1.0024 | 1.0055 | 0.9914 |
| 200 | | 1.0042 | 1.0024 | 1.0019 | 1.0032 | 1.0022 | 1.0049 | 0.9941 |

| \( n \) | AABs | 0.0837 | 0.0606 | 0.0591 | 0.0682 | 0.0498 | 0.1083 | 0.0109 |
|---|---|---|---|---|---|---|---|
| 10 | 0.0282 | 0.0207 | 0.0166 | 0.0224 | 0.0163 | 0.0404 | 0.0198 |
| 25 | 0.0139 | 0.0141 | 0.0098 | 0.0117 | 0.0109 | 0.0154 | 0.0129 |
| 50 | 0.0058 | 0.0061 | 0.0054 | 0.0073 | 0.0034 | 0.0102 | 0.0014 |
| 100 | 0.0041 | 0.0046 | 0.0038 | 0.0056 | 0.0024 | 0.0055 | 0.0086 |
| 150 | 0.0042 | 0.0024 | 0.0019 | 0.0032 | 0.0022 | 0.0049 | 0.0059 |
| 200 | | | | | | |

| \( n \) | MREs | 0.0837 | 0.0606 | 0.0591 | 0.0682 | 0.0498 | 0.1083 | 0.0109 |
|---|---|---|---|---|---|---|---|
| 10 | 0.0282 | 0.0207 | 0.0166 | 0.0224 | 0.0163 | 0.0404 | 0.0198 |
| 25 | 0.0139 | 0.0141 | 0.0098 | 0.0117 | 0.0109 | 0.0154 | 0.0129 |
| 50 | 0.0058 | 0.0061 | 0.0054 | 0.0073 | 0.0034 | 0.0102 | 0.0014 |
| 100 | 0.0041 | 0.0046 | 0.0038 | 0.0056 | 0.0024 | 0.0055 | 0.0086 |
| 150 | 0.0042 | 0.0024 | 0.0019 | 0.0032 | 0.0022 | 0.0049 | 0.0059 |
| 200 | | | | | | |

| \( n \) | MSEs | 0.1053 | 0.1386 | 0.1348 | 0.1289 | 0.0935 | 0.1907 | 0.0722 |
|---|---|---|---|---|---|---|---|
| 10 | 0.0517 | 0.0402 | 0.0326 | 0.0355 | 0.0334 | 0.0520 | 0.0271 |
| 25 | 0.0261 | 0.0180 | 0.0161 | 0.0174 | 0.0158 | 0.0218 | 0.0332 |
| 50 | 0.0063 | 0.0085 | 0.0077 | 0.0084 | 0.0071 | 0.0102 | 0.0062 |
| 100 | 0.0042 | 0.0057 | 0.0050 | 0.0055 | 0.0049 | 0.0066 | 0.0042 |
| 150 | 0.0033 | 0.0041 | 0.0037 | 0.0040 | 0.0037 | 0.0050 | 0.0032 |

| \( n \) | MREs | 0.0937 | 0.0839 | 0.0606 | 0.0854 | 0.0555 | 0.1308 | 0.0128 |
|---|---|---|---|---|---|---|---|
| 10 | 0.0344 | 0.0249 | 0.0239 | 0.0257 | 0.0205 | 0.0460 | 0.0197 |
| 25 | 0.0174 | 0.0085 | 0.0103 | 0.0134 | 0.0145 | 0.0210 | 0.0123 |
| 50 | 0.0093 | 0.0051 | 0.0062 | 0.0071 | 0.0062 | 0.0116 | 0.0094 |
| 100 | 0.0061 | 0.0030 | 0.0040 | 0.0036 | 0.0046 | 0.0084 | 0.0070 |
| 150 | 0.0047 | 0.0018 | 0.0034 | 0.0041 | 0.0025 | 0.0063 | 0.0064 |
| 200 | | | | | | |

| \( n \) | MSEs | 0.2624 | 0.4250 | 0.3263 | 0.3608 | 0.2614 | 0.6702 | 0.1906 |
|---|---|---|---|---|---|---|---|
| 10 | 0.0580 | 0.0977 | 0.0906 | 0.0965 | 0.0883 | 0.1331 | 0.0702 |
| 25 | 0.0355 | 0.0438 | 0.0390 | 0.0468 | 0.0406 | 0.0582 | 0.0325 |
| 50 | 0.0172 | 0.0215 | 0.0203 | 0.0217 | 0.0193 | 0.0284 | 0.0160 |
| 100 | 0.0110 | 0.0139 | 0.0130 | 0.0137 | 0.0129 | 0.0174 | 0.0110 |
| 150 | 0.0084 | 0.0102 | 0.0095 | 0.0102 | 0.0092 | 0.0126 | 0.0079 |
Table 4: Simulation results of Log-XLindley distribution at $\theta = 2.0$.

| $n$ | Measure | AVEs | AABs | MREs | MSEs |
|-----|----------|------|------|-------|------|
| 10  |          | 2.1597 | 0.1597 | 0.0799 | 3.8178 |
|     |          | 2.1371 | 0.1371 | 0.0686 | 5.2855 |
| 25  |          | 2.0730 | 0.2085 | 0.0285 | 1.4701 |
| 50  |          | 2.0363 | 0.0571 | 0.0144 | 0.0733 |
| 100 |          | 2.0166 | 0.0288 | 0.0149 | 0.0556 |
| 150 |          | 2.0101 | 0.1117 | 0.0066 | 0.0219 |
| 200 |          | 2.0077 | 0.0068 | 0.0046 | 0.0157 |

Table 5: Simulation results of Log-XLindley distribution at $\theta = 5.0$.

| $n$ | Measure | AVEs | AABs | MREs | MSEs |
|-----|----------|------|------|-------|------|
| 10  |          | 5.5013 | 0.5013 | 0.0306 | 3.8178 |
|     |          | 5.3804 | 0.3804 | 0.0203 | 5.2855 |
| 25  |          | 5.1948 | 0.1948 | 0.0103 | 1.4701 |
| 50  |          | 5.0916 | 0.0916 | 0.0148 | 0.0733 |
| 100 |          | 5.0458 | 0.0458 | 0.0280 | 0.0208 |
| 150 |          | 5.0306 | 0.0306 | 0.0283 | 0.0283 |
| 200 |          | 5.0262 | 0.0262 | 0.0203 | 0.0203 |

Table 6: Some descriptive measures of considered datasets.

| Algorithm | $n$ | Min. | Q1 | Mean | Median | Q3 | Var | Skewness | Kurtosis | Max. |
|-----------|-----|------|----|------|--------|----|-----|-----------|----------|------|
| SC16      | 23  | 0.006 | 0.033 | 0.288 | 0.116 | 0.518 | 0.101 | 0.768 | 1.974 | 0.866 |
| P3        | 22  | 0.010 | 0.047 | 0.304 | 0.118 | 0.544 | 0.101 | 0.711 | 1.884 | 0.874 |
Table 7: The MLEs and standard errors (in parenthesis) for SC16 data.

| Model                 | Parameter estimates (standard errors) | $-l$     | $P$ value (KS) |
|-----------------------|---------------------------------------|----------|----------------|
| Kumaraswamy($\theta$, $\beta$) | 0.50439 (0.12878) 1.18618 (0.32645) | 9.6708   | 0.4529 (0.1790) |
| beta($\theta$, $\beta$)       | 0.48698 (0.12082) 1.16805 (0.35781) | 9.6075   | 0.4202 (0.1836) |
| Log – XLindley($\theta$)      | 0.62512 (0.09916) —                | 9.9019   | 0.5007 (0.1725) |

Table 8: MLEs and standard errors (in parenthesis) for P3 data.

| Models                | Parameter estimates (standard errors) | $-l$     | $P$ value (KS) |
|-----------------------|---------------------------------------|----------|----------------|
| Kumaraswamy($\theta$, $\beta$) | 0.57179 (0.14777) 1.23051 (0.34829) | 6.8436   | 0.3650 (0.1963) |
| beta($\theta$, $\beta$)       | 0.55393 (0.14233) 1.21974 (0.37577) | 6.7819   | 0.3414 (0.2002) |
| Log – XLindley($\theta$)      | 0.68064 (0.11128) —                | 6.9881   | 0.3939 (0.1817) |
Weranthesimulation5000timestoderivethesemetricsfromthepriorvaluesforallestimationapproaches. ThefindingsinTables1–5werereportedutilizingtheRsoftware’soptim-CGfunction. Thefindingsshowthatasthesample size $n$ increased, the AVEs became closer to the real values of $\theta$. Furthermore, when $n$ increases, the ABBs, MREs, and MSEs for all estimators decreased. This proves that the previous estimation techniques work quite well in estimating the model parameter.

7. Data Analysis: Data of SC16 and P3 Algorithms

In this section, we consider two datasets to show the applicability and flexibility of the introduced distributions over famous distributions. Here, we compare the Log-XLindley model with some competitive models like the Kumaraswamy (Kw) and beta (B) distributions. The PDF of the competitive models can be formulated, respectively, as follows:

$$f_{\text{Kw}}(x; \theta, \beta) = \theta \beta x^{\theta - 1} (1 - x^{\theta})^{\beta - 1}, \quad x \in (0, 1),$$

$$f_{\text{B}}(x; \theta, \beta) = B(\theta, \beta) x^{\theta - 1} (1 - x)^{\beta - 1}, \quad x \in (0, 1).$$
Figure 6: The empirical plots for the P3 data.

Figure 7: Profile log-likelihood plots for both data sets.
Some criteria like the Kolmogorov–Smirnov (KS) test with its $P$ value have been used to get the best model among all the tested distributions. The following data are from [20, 21], where they compare the two different algorithms called SC16 and P3 for estimating unit capacity factors. The observations resulting from the algorithm SC16 are 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, and 0.006. However, the observations resulting from the second algorithm named P3 are 0.853, 0.759, 0.874, 0.800, 0.716, 0.557, 0.503, 0.399, 0.334, 0.207, 0.118, 0.118, 0.097, 0.078, 0.067, 0.056, 0.044, 0.036, 0.026, 0.019, 0.014, and 0.010. Some descriptive measures about these data sets are presented in Table 6.

Regarding Table 6, both datasets are asymmetric “positively skewed” with platykurtic-shaped data. Moreover, the datasets are under dispersion where the value of the mean is less than the various values. Information about the failure rate can be helpful in the selection of the appropriate model. A gadget known as the total time on test (TTT) plot [22] can be used for this purpose. If the shape of the TTT plot is straight diagonal, the hazard is constant. The TTT plot has a convex shape for decreasing hazards and a concave shape for increasing hazards. The bathtub-shaped hazard is obtained when first is convex and then concave. The total test time (TTT) graphs for data sets I and II are shown in Figure 3. The hazard curves of both datasets are bathtub-shaped. Figure 4 shows the boxplots for data sets I and II, respectively. Therefore, Log-XLindley distribution can be a good choice to model these data sets.

The MLEs of the considered models along with their standard errors are given in Tables 7 and 8 for data sets I and II, respectively, with goodness-of-fit measures. In this paper, a flexible one-parameter Log-XLindley distribution has been proposed to analyze and discuss the proportion and asymmetric data. Some distributional properties have been derived in explicit forms. It was found that the hazard rate function can be applied to model different types of failures including increasing, bathtub, and J-shaped. The model parameter has been estimated using various estimation approaches to get the best estimator for data. A Monte-Carlo simulation study for different sample sizes has been performed to assess the performance of the estimations based on some statistical criteria. Finally, two distinctive data sets from SC16 and P3 algorithms have been analyzed to illustrate the flexibility of the new model, and it was found that the proposed distribution proved a remarkable superiority when compared to the competitive models.

### 8. Conclusion

In this paper, a flexible one-parameter Log-XLindley distribution has been proposed to analyze and discuss the proportion and asymmetric data. Some distributional properties have been derived in explicit forms. It was found that the hazard rate function can be applied to model different types of failures including increasing, bathtub, and J-shaped. The model parameter has been estimated using various estimation approaches to get the best estimator for data. A Monte-Carlo simulation study for different sample sizes has been performed to assess the performance of the estimations based on some statistical criteria. Finally, two distinctive data sets from SC16 and P3 algorithms have been analyzed to illustrate the flexibility of the new model, and it was found that the proposed distribution proved a remarkable superiority when compared to the competitive models.

### Data Availability

Data are included in the manuscript.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Supplementary Materials

The R code used in the study can be assessed from the attached file. (Supplementary Materials)

### References

[1] L. E. Papke and J. M. Wooldridge, “Econometric methods for fractional response variables with an application to 401 (k) plan participation rates,” *Journal of Applied Econometrics*, vol. 11, no. 6, pp. 619–632, 1996.

[2] N. L. Johnson, “Systems of frequency curves generated by methods of translation,” *Biometrika*, vol. 36, no. 1/2, pp. 149–176, 1949.

[3] C. W. Topp and F. C. Leone, “A family of J-shaped frequency functions,” *Journal of the American Statistical Association*, vol. 50, no. 269, pp. 209–219, 1955.

[4] A. Grassia, “On a family of distributions with argument between 0 and 1 obtained by transformation of the gamma and derived compound distributions,” *Australian Journal of Statistics*, vol. 19, no. 2, pp. 108–114, 1977.

[5] P. Kumaraswamy, “A generalized probability density function for double-bounded random processes,” *Journal of Hydrology*, vol. 46, no. 1–2, pp. 79–88, 1980.
[6] E. Gómez-Déniz, M. A. Sordo, and E. Calderin-Ojeda, “The Log-Lindley distribution as an alternative to the beta regression model with applications in insurance,” *Insurance: Mathematics and Economics*, vol. 54, no. 1, pp. 49–57, 2014.

[7] A. F. B. Menezes, J. Mazucheli, and S. Dey, “The unit-logistic distribution: different methods of estimation,” *Pesquisa Operacional*, vol. 38, no. 3, pp. 555–578, 2018.

[8] J. Mazucheli, A. F. B. Menezes, A. F. B. Menezes, S. Dey, A. F. B. Menezes, and S. Dey, “The unit-Birnbaum-Saunders distribution with applications,” *Chilean Journal of Statistics*, vol. 9, no. 1, pp. 47–57, 2018, http://www.soche.cl/chjs/volumes/09/01/Mazucheli_etal(2018).pdf.

[9] E. Altun, “The log-xgamma distribution with inference and application,” *Journal de la Societe de Statistique de Paris*, vol. 159, no. 3, pp. 40–55, 2018.

[10] J. Mazucheli, A. F. B. Menezes, and S. Chakraborty, “On the one parameter unit-Lindley distribution and its associated regression model for proportion data,” *Journal of Applied Statistics*, vol. 46, no. 4, pp. 700–714, 2019.

[11] J. Mazucheli, A. F. Maringa, and S. Dey, “Unit-Gompertz distribution with applications,” *Statistica*, vol. 79, no. 1, pp. 25–43, 2019.

[12] M. E. Ghitany, J. Mazucheli, A. F. B. Menezes, and F. Alqallaf, “The unit-inverse gaussian distribution: a new alternative to two-parameter distributions on the unit interval,” *Communications in Statistics - Theory and Methods*, vol. 48, no. 14, pp. 3423–3438, 2019.

[13] K. Modi and V. Gill, “Unit burr-III distribution with application,” *Journal of Statistics & Management Systems*, vol. 23, pp. 1–14, 2019.

[14] E. Altun, “The log-weighted exponential regression model: alternative to the beta regression model,” *Communications in Statistics - Theory and Methods*, vol. 50, no. 10, pp. 2306–2321, 2019.

[15] J. Mazucheli, A. F. B. Menezes, L. B. Fernandes, R. P. de Oliveira, and M. E. Ghitany, “The unit-Weibull distribution as an alternative to the Kumaraswamy distribution for the modeling of quantiles conditional on covariates,” *Journal of Applied Statistics*, vol. 47, no. 6, pp. 954–974, 2020.

[16] M. Ahsan-ul-Haq, S. Hashmi, K. Aidi, P. L. Ramos, and F. Louzada, “Unit modified burr-III distribution: estimation, characterizations and validation test,” *Annals of Data Science*, vol. 99, 2020.

[17] R. A. R. Bantan, C. Chesneau, F. Jamal et al., “Some new facts about the unit-Rayleigh distribution with applications,” *Mathematics*, vol. 8, no. 11, pp. 1–23, 2020.

[18] M. Ahsan-ul-Haq, M. Albassam, M. Aslam, and S. Hashmi, “Statistical inferences on odd fréchet power function distribution,” *Journal of Reliability and Statistical Studies*, vol. 14, no. 1, pp. 141–172, 2021.

[19] M. Korkmaz and C. Chesneau, “On the unit Burr-XII distribution with the quantile regression modeling and applications,” *Computational and Applied Mathematics*, vol. 40, no. 1, pp. 1–26, 2021.

[20] M. Caramanis, J. Strevel, W. Fleck, and S. Daniel, “Probabilistic production costing: an investigation of alternative algorithms,” *International Journal of Electrical Power & Energy Systems*, vol. 5, no. 2, pp. 75–86, 1983.

[21] M. Mazumdar and D. P. Gaver, “On the computation of power-generating system reliability indexes,” *Technometrics*, vol. 26, no. 2, pp. 173–185, 1984.

[22] M. V. Aarset, “How to identify a bathtub hazard rate,” *IEEE Transactions on Reliability*, vol. R-36, no. 1, pp. 106–108, 1987.