



Omega ω —Type Probability Models: A Parametric Modification of Probability Distributions

Udochukwu Victor Echebiri¹ · Nosakhare Liberty Osawe¹ ·
Chukwuemeka Thomas Onyia²

Received: 24 August 2023 / Revised: 5 April 2024 / Accepted: 22 April 2024

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

A mathematical approach to developing new distributions is reviewed. The method which composes of integration and the concept of a normalizing constant, allows for primitive interjection of new parameter(s) in an existing distribution to form new model(s), called *Omega-Type* probability models. A probability distribution is proposed from a root model, Lindley distribution, and some properties, such as the series representation of the density and cumulative distribution functions, shape of the density, hazard and survival functions, moments and related measures, quantile function, order statistics, parameter estimation and interval estimate, were studied. Amidst the usual hazard and survival shapes, a constant or uniform trend was realized for the survival function, which projects the possibility of modeling systems that may not terminate over a given period of time. Three different methods of estimation, namely, the Cramer–von Mises estimator, maximum product of the spacing estimator and maximum likelihood estimator, were used. The modified unimodal shape of the proposed distribution is added as a special feature in the improvements made among the Lindley family of distributions. Finally, two real-life datasets were fitted to the new distribution to demonstrate its economic importance.

Keywords Omega-type model · Parametric modification · Lindley family · Survival trend · Data applications

✉ Udochukwu Victor Echebiri
echebiriudochukwu27@gmail.com

¹ Department of Statistics, Faculty of Physical Sciences, University of Benin, Benin, Nigeria

² Department of Statistics, Faculty of Natural Sciences, Enugu State University of Science and Technology, Enugu, Nigeria

1 Introduction

In research analysis, consequential occurrences may not be necessary to obtain data for decision making, where we have access to antecedent outcomes. So, data science, data analysis and data modeling help us to gain meaningful insights in an attempt to obtain viable information. Data science is an interdisciplinary field that intertwines statistical applications, computational and machine learning methods to extrapolate usable information from big data for predictive modeling. In this area, lucrative advances have been made in big data analytics, business and optimization based data mining, internet of things, real-time decision making and artificial Intelligence [1–4]. These data that go through statistical, algorithmic or scientific processes is usually unstructured involving tasks such as data preprocessing and feature engineering; unlike data analysis which deals with data cleaning and visualization for structured data. However, data modeling as a primary objective of probability distribution is the analysis of data by projecting the likelihood of different outcomes using relevant models.

The development of new probability distributions does not entail a lack of appreciation for the existent; rather, it is an attempt to achieve greater effectiveness in data modeling. There are many methodologies for the modification of an existing distribution, as studied by Lai [5]. Among many methods are the transformation method, compounding, convolution, mixture model, skewing, parameterization and the avalanched generalization methods. More emphatically, the generalization method has been used more frequently in recent decades to increase the number of parameters of classical distributions. This approach is conventionally aimed at improving model flexibility by primarily reconstructing for robustness. Moreover, these modifications also have a great impact on the nature of hazard functions of distributions, featuring various real-life representations.

Tahir et al. [6] described a methodology for the development of some of these generalized distributions, given in cumulative distribution function (CDF) order:

$$F(x, \omega) = \int_{R_l}^{\beth(x)} m(t)dt \text{ or } \int_{\beth(x)}^{R_u} m(t)dt. \quad (1)$$

where $\beth(x)$ is the link function, which is derived from the quantile function of a random variable X ; $m(t)$ is the probability density function (PDF) of a dummy random variable T ; and R is the (lower or upper) boundary support. Tahir tabulated the various link functions for the continuous distribution category. Some of the generators or distributions developed in this order are logistic X, Gompertz inverse exponential, Sin G, the Marshal–Olkin transformation with TX, the Gull Alpha Power, the Normal-Power-Logistic, the Exponentiated Odd Log-Logistic Weibull, and several others [6–12].

The addition of parameter(s) is a great economic process harnessed by different methods of probability model development. The motivation of this research stems from the challenge posed by the parameter inclusion strategy of some existing distributions, which allows the incorporation of (extra) parameters by inserting PDF and/or

CDF structures to form higher parametric compound models. However, a more direct procedure can be adopted to efficiently achieve the same result. Hence, our interest in this study is to develop a new approach for increasing the number of parameters in an existing distribution without allowing for model insertions, as found with the generalization method. Second, we propose a probability distribution to exemplify the relevance of the approach. Of course, the proposed model is flexible and can model outcomes that mirror monotonically decreasing and increasing, right-skewed and modified unimodal trends.

The remaining parts of this work are arranged as follows: Sect. 2 of this paper features a direct strategy for parameter inclusion. In Sect. 3, a proposal is made for a new probability model using the insights in the previous section. Section 4 details the study on some of the properties of the proposed distribution. Section 5 describes the applicability of the proposed model through data analysis and/or performance comparison.

2 Omega Type Probability Models

In this development, extra parameters are *premeditatedly* implanted as a function of a variable, directly into an extracted *variable component* from a root model. After the application of a mathematical technique, the derivation of a new PDF is suitable. The procedure composes of the “integration method” and the concept of the “normalizing constant”, as detailed in [13].

To achieve this, we capture the variable components, say $\theta(x, \omega)$, of an existing distribution and treat it as an arbitrary function, where x and ω are the variable and vector of parameters, respectively. The extracted components, however, can be modified by primitively adding some parameters to form new probability distribution(s). This is constructed such that the root distribution $f(x, \omega)$ is obtained as a special case of the new development, as in the generalization method. Now, the choice of the number of parameters to possibly add depends on the structure of the variable in a density function. For example, the variable x in the variable component $\theta(x, \omega) = e^{-ax}$ from the exponential distribution can discretely be changed to x^b . Hence, we obtain $\theta_r(x, \omega) = e^{-ax^b}$ as a new arbitrary function, so long as it retains the propensity to return the initial component e^{-ax} at a given value of b .

Moreover, if we define

$$\theta(x, \omega) = (1 + x)e^{-ax}, \quad (2)$$

from the Lindley distribution, for example; thus, we have increased options for parametric inclusion. In this case, the new possible arbitrary functions become

$$\theta_r(x, \omega) = (1 + x)e^{-ax^b} \text{ or} \quad (3)$$

$$\theta_r(x, \omega) = (1 + x^b)e^{-ax^c}, \quad (4)$$

which are composites of two or three parameters, respectively. Another approach to this inclusion is to *replace a replicated parameter* in the root variable component $\theta(x, \omega)$ with a new parameter. Recall the Kumaraswamy distribution, which has a variable component given by,

$$\theta(x, \omega) = x^{-1+a}(1-x^a)^{-1+b}. \quad (5)$$

The parameter “a” is replicated twice. However, the variable x^a in the second component $(1-x^a)$ can be replaced with x^c which gives us a new three-parameter arbitrary function.

$$\theta_r(x, \omega) = x^{-1+a}(1-x^c)^{-1+b} \quad (6)$$

Consequently, by integrating the new arbitrary functions $\theta_r(x, \omega)$ and applying the concept of a *normalizing constant*, a new probability density function can be realized. Currently, this development is not limited to cases of superscripts, as in x^* , but seems more consistent in mathematical structures expressed in chain format. This suggests that we could also integrate parameters in components of this sort:

$$\sin(x) \rightarrow \sin(ax), \text{Log}(x) \rightarrow \text{Log}(bx), \Gamma(x) \rightarrow \Gamma(cx), \text{erf}(x) \rightarrow \text{erf}(ax)$$

and several others. Now, let X and ω be a random variable and a vector of parameters, respectively; then, a probability distribution can be constructed:

$$u(x, \omega) = \theta_r(x, \omega) \left[\int_{-\infty}^{\infty} \theta_r(x, \omega) dx \right]^{-1} \quad (7)$$

for the probability density function, and the cumulative distribution function is

$$U(x, \omega) = \int_{-\infty}^x \left\{ \theta_r(t, \omega) \left[\int_{-\infty}^{\infty} \theta_r(t, \omega) dt \right]^{-1} \right\} dt \quad (8)$$

where $\theta_r(x, \omega)$ is the new variable component modified from an existing distribution as already defined. Of course, the result obtained from $\left[\int_{-\infty}^{\infty} \theta_r(x, \omega) dx \right]^{-1}$ is a constant-parameter function that serves as the normalizing constant. This implies that the product of $\theta_r(x, \omega)$ and the normalizing constant yields a new PDF. In this approach, it should be noted that, consistently, the root distribution is returned as a special case of new development.

For the purposes of nomenclature, we term this construction *Omega ω -Type Probability Models*, where *omega ω* represents the primitive parametric substitution and *Type* denotes the root model from which the arbitrary function was taken.

3 Omega Lindley Distribution

Originally, the Lindley distribution was introduced by Lindley [14], and its properties were studied by Ghitany Atieh and Nadarajah [15]. However, to compensate for the monotonic property of the Lindley distribution, several authors, including [16–19] among others, have proposed at least a generalization of the Lindley distribution. These developments have increased the flexibility of the baseline distribution in different ways, ranging from monotone increasing trends, symmetric trends, left- and right-skewed trends, bathtubs, etc., in terms of density and hazard considerations. However, a real-life observation characterized by a modified unimodal (bathtub + inverted bathtub) trend in both density and hazard function considerations is not yet a feature of the Lindley family of distributions. Hence, we introduce a probability model, which will capture that essence. Considering the exemplified function in (4), we give a construction, and by recalling (7), we obtain the new PDF as

$$\begin{aligned}
 u(x, \omega) &= \left\{ (1+x^b) e^{-ax^c} \right\} \left[\int_{-\infty}^{\infty} (1+x^b) e^{-ax^c} dx \right]^{-1}, \\
 &= \frac{c}{N_{\circ}} a^{1+b/c} (1+x^b) e^{-ax^c}, \quad x > 0, \quad a > 0, \quad b > 0, \quad c > 0. \quad (9)
 \end{aligned}$$

where $N_{\circ} = a^{b/c} \Gamma\left[\frac{1}{c}\right] + \Gamma\left[\frac{1+b}{c}\right]$.

and the corresponding cumulative distribution function is given by

$$U(x, \omega) = \frac{(ax^c) a^{\frac{1+b}{c}} x^{-\frac{1+b}{c}} \varphi_{abc}}{N_{\circ}}, \quad (10)$$

where $\varphi_{abc} = \left\{ \left[(ax^c)^{b/c} \left(\Gamma\left[\frac{1}{c}\right] - \Gamma\left[\frac{1}{c}, ax^c\right] \right) \right] + \left[x^b \Gamma\left(\frac{1+b}{c}\right) \right] - \left[\Gamma\left(\frac{1+b}{c}, ax^c\right) \right] \right\}$.

However, by evaluating $b = c = 1$, (9) and (10) reduce to the Lindley distribution, which is the root model.

4 Mathematical Properties of the Omega Lindley Distribution

In this section, we derive several possible mathematical properties, such as series representation, shape of the density, hazard and survival functions, moments and related measures, order statistics, quantile function and parameter estimation. Notably, algorithms from *Wolfram Mathematica* aided the derivations and numerical analyses in this work.

4.1 Series Representation

The following forms of mathematical (and special) functions can be represented as series:

$$e^{-ax^c} = \sum_{i=0}^{\infty} \frac{(-ax^c)^i}{i!}. \quad (11)$$

$$(1+x^b) = \sum_{j=0}^{\infty} \binom{1}{j} x^{bj} \quad (12)$$

$$\Gamma(\theta, x) = \Gamma(\theta) - \sum_{p=0}^{\infty} \frac{(-1)^p x^{\theta+p}}{p! (\theta+p)} \quad (13)$$

$$\rightarrow \Gamma\left[\frac{1}{c}, ax^c\right] = \Gamma\left(\frac{1}{c}\right) - \sum_{p=0}^{\infty} \frac{(-1)^p (ax^c)^{\frac{1}{c}+p}}{p! \left(\frac{1}{c}+p\right)}. \quad (14)$$

$$\Gamma\left(\frac{1+b}{c}, ax^c\right) = \Gamma\left(\frac{1+b}{c}\right) - \sum_{p=0}^{\infty} \frac{(-1)^p (ax^c)^{\frac{1+b}{c}+p}}{p! \left(\frac{1+b}{c}+p\right)} \quad (15)$$

By implication generally, Eqs. (9) and (10) are represented as.

$$u(x, \omega) = \frac{ca^{b/c}}{N_0} \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \binom{1}{j} \frac{(-ax^c)^i}{i!} x^{bj}. \quad (16)$$

$$U(x, \omega) = \frac{(ax^c)a^{\frac{1+b}{c}} x^{-\frac{1+b}{c}} \varphi_{abc}}{N_0}, \quad (17)$$

where

$$\varphi_{abc} = \left[(ax^c)^{b/c} \left(\Gamma\left[\frac{1}{c}\right] - \left[\Gamma\left(\frac{1}{c}\right) - \sum_{p=0}^{\infty} \frac{(-1)^p (ax^c)^{\frac{1}{c}+p}}{p! \left(\frac{1}{c}+p\right)} \right] \right) \right] + \left[x^b \Gamma\left(\frac{1+b}{c}\right) \right] - \left[\Gamma\left(\frac{1+b}{c}\right) - \sum_{p=0}^{\infty} \frac{(-1)^p (ax^c)^{\frac{1+b}{c}+p}}{p! \left(\frac{1+b}{c}+p\right)} \right].$$

These, in essence, simply reduce the computational complexities that come with having a variable in special functions such as the incomplete gamma function.

4.2 The Shape of the Density, Hazard and Survival Functions of the Omega Lindley Distribution

Here, we examine the various possible shapes of the density function obtained in (9), where the hazard rate function and survival model of the ω – Lindley distribution are derived as

$$H(x, \omega) = \frac{u(x, \omega)}{1 - U(x, \omega)}; S(x, \omega) = 1 - U(x, \omega). \tag{18}$$

The different plots are intended to X-ray the modeling potentials of the proposed distribution.

The graphical plot in Fig. 1 indicates that the density function of the $\omega - Lindley$ distribution exhibits a monotonic increasing and decreasing trend and a right skewed, approximately symmetric and modified unimodal trend (bathtub and inverted bathtub). This is a more robust improvement to the exponential and Lindley distribution, which is characterized by a monotonic decreasing trend only. The respective parameter combinations for each of the density shapes are given thus:

- $a = 1.28, b = 6.48, c = 1 \parallel a = 0.21, b = 3.4, c = 1.2;$
- $a = 1, b = 1.15, c = 1 \parallel a = 1, b = 1, c = 1;$
- $a = 0.01, b = 2.11, c = 0.29;$
- $a = 0.1, b = 0.51, c = 0.57.$

The respective parameter combinations for each of the hazard shapes are given thus:

- $a = 2.43, b = 1, c = 0.25$
- $a = 1, b = 2, c = 0.27$
- $a = 2.26, b = 1, c = 0.24$
- $a = 4.17, b = 2.56, c = 0.49$

The graphical plots in Fig. 2 show that the hazard rate function of the $\omega - Lindley$ distribution exhibits both decreasing and increasing trends: bathtub and modified unimodal shaped properties. In addition, Fig. 3 shows the survival trend, which is

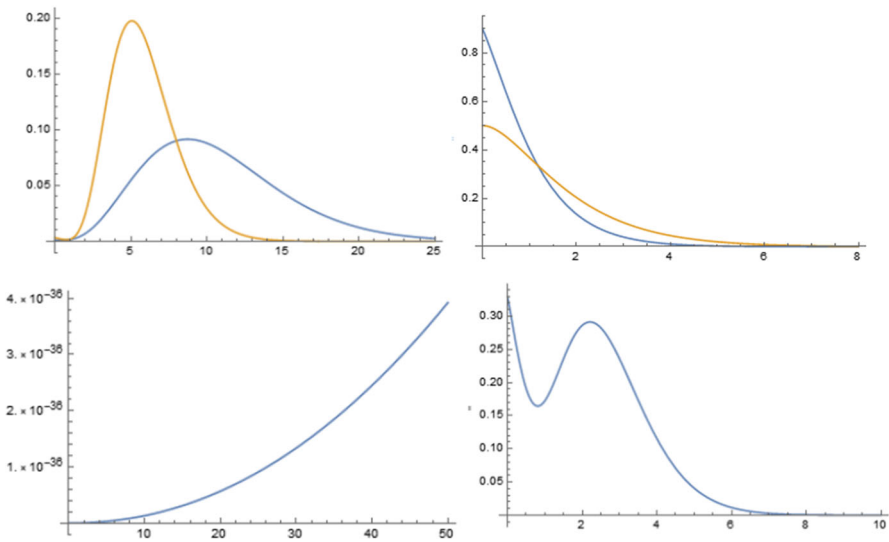


Fig. 1 Density shapes of the $\omega - Lindley$ distribution

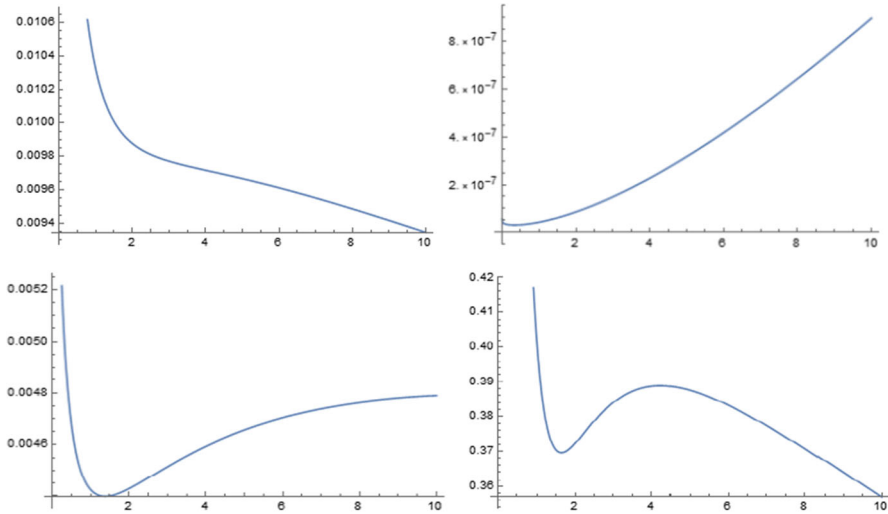


Fig. 2 Hazard plots for the ω – Lindley distribution

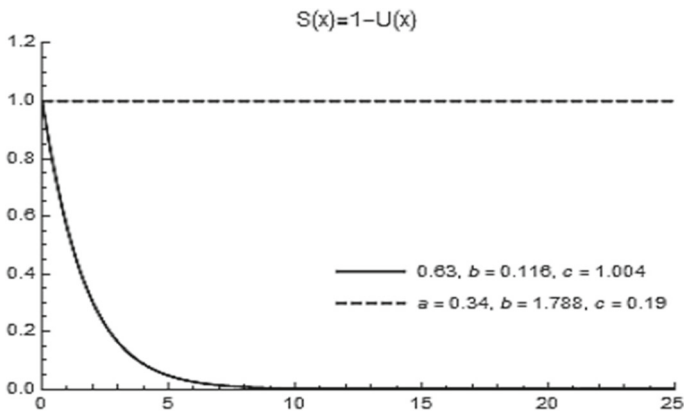


Fig. 3 Survival plots for ω – Lindley distribution

conventionally marked by its decreasing characteristics; however, a constant trend is also obtained. By implication, the proposed distribution would not only model systems or components that will definitely terminate but also one such system that can be sustained over a given time.

4.3 Moments and Related Measures

If X is a random variable from a continuous distribution and $g(x)$ is the density function, then the r th moment about the origin of X is defined by

$$E(X^r) = \int_0^{\infty} x^r g(x) dx = \mu^r. \quad (19)$$

Now, for the ω -Lindley distribution, we recall (9) and obtain

$$\begin{aligned} E(X^r) &= \frac{ca^{1+b/c}}{N_{\circ}} \int_0^{\infty} x^r (1+x^b) e^{-ax^c}, \\ &= \frac{a^{\frac{1+b}{c} - \frac{1+b+r}{c}} (a^{b/c} \Gamma[\frac{1+r}{c}] + \Gamma[\frac{1+b+r}{c}])}{N_{\circ}}. \end{aligned} \quad (20)$$

Of course, the first four r th moments of the ω -Lindley distribution are further obtained at $r = 1, 2, 3$ and 4

$$\mu'_1 = \frac{a^{\frac{1+b}{c} - \frac{2+b}{c}} (a^{b/c} \Gamma[\frac{2}{c}] + \Gamma[\frac{2+b}{c}])}{N_{\circ}} = \mu \mu'_2 = \frac{a^{\frac{1+b}{c} - \frac{3+b}{c}} (a^{b/c} \Gamma[\frac{3}{c}] + \Gamma[\frac{3+b}{c}])}{N_{\circ}}, \quad (22)$$

$$\mu'_3 = \frac{a^{\frac{1+b}{c} - \frac{4+b}{c}} (a^{b/c} \Gamma[\frac{4}{c}] + \Gamma[\frac{4+b}{c}])}{N_{\circ}} \quad \mu'_4 = \frac{a^{\frac{1+b}{c} - \frac{5+b}{c}} (a^{b/c} \Gamma[\frac{5}{c}] + \Gamma[\frac{5+b}{c}])}{N_{\circ}} \quad (23)$$

Therefore, the variance, skewness and kurtosis of the ω -Lindley distribution can be obtained

$$\text{Variance}(\mu_2) = \mu'_2 - \mu^2 = \sigma^2 \text{Skewness}(S_k) = \frac{\mu_3}{(\mu_2)^{3/2}} = \frac{\mu'_3 - 3\mu'_2\mu + 2\mu^3}{(\mu'_2 - \mu^2)^{3/2}} \quad (24)$$

$$\text{Kurtosis}(K_s) = \frac{\mu_4}{(\mu_2)^2} = \frac{\mu'_4 - 4\mu'_3\mu + 6\mu'_2\mu^2 - 3\mu^4}{(\mu'_2 - \mu^2)^2}. \quad (25)$$

Numerical analysis of the theoretical moments of the ω -Lindley distribution for the selected parameter values is shown in Tables 1, 2 and 3.

By carefully observing Tables 1, 2 and 3, we see that the ω -Lindley distribution is right skewed ($S_k > 0$), approximately symmetric ($S_k \approx 0$), leptokurtic ($K_s > 3$), platykurtic ($K_s < 3$) and mesokurtic ($K_s \approx 3$). These numerical characteristics are backup points that support the graphical illustration of the density function, as shown in Fig. 1.

Table 1 Moments of ω – Lindley distribution at $c = 1$

μ_r^*	$a = 1.2$ $b = 0.5$	$a = 2.5$ $b = 5$	$a = 2$ $b = 5.095$
μ'_1	1.0197	1.5027	2.5511
μ'_2	1.9324	3.8485	8.8019
μ'_3	5.3162	12.028	35.377
μ'_4	19.136	42.956	160.51
σ^2	0.8926	1.5904	2.2938
S_k	1.8088	0.4268	0.3511
K_s	7.8627	2.9665	3.0670

Table 2 Moments of ω – Lindley distribution at $c = 1.5$

μ_r^*	$a = 0.5$ $b = 1$	$a = 0.1$ $b = 6.5$	$a = 3.25$ $b = 0.75$
μ'_1	1.1275	3.5664	0.5111
μ'_2	2.4450	33.068	0.2861
μ'_3	6.7750	353.611	0.2105
μ'_4	22.3698	4249.4	0.1877
σ^2	1.1737	20.3494	0.0249
S_k	1.0786	0.9862	9.8864
K_s	4.0764	3.0024	1.7046

Table 3 Moments of ω – Lindley distribution at $c = 5$

μ_r^*	$a = 1.3$ $b = 4$	$a = 1$ $b = 4.5$	$a = 3.25$ $b = 0.75$
μ'_1	0.642	0.5622	1.0775
μ'_2	0.4541	0.4286	0.5995
μ'_3	0.3685	0.3740	0.3819
μ'_4	0.3249	0.3540	0.265
σ^2	0.0419	0.1125	– 0.5615
S_k	2.6959	0.1726	9.8864
K_s	– 4.593	2.0585	– 3.9604

4.4 Order Statistics

Let X_1, X_2, \dots, X_n be a random sample of size n from the ω – Lindley distribution. Let $X_1 < X_2 < \dots < X_n$ denote the corresponding order statistics. The density and distribution functions of the k th-order statistics $Y = X_k$ are given by:

$$f_Y(y) = \frac{n!}{(k-1)!(n-k)!} f(y) F^{k-1}(y) \{1 - F(y)\}^{n-k}, \quad (26)$$

$$\text{However, } \{1 - F(y)\}^{n-k} = \sum_{l=0}^{\infty} \binom{n-k}{l} (-1)^l [F(y)]^l$$

$$\begin{aligned} \therefore f_Y(y) &= \frac{n! ca^{1+b/c}}{(k-1)!(n-k)!} \left\{ (1+y^b) e^{-ay^c} \right\} N_o^{-2} \left\{ (ay^c) a^{\frac{1+b}{c}} y^{-\frac{1+b}{c}} \varphi_{abc} \right\}^{k-1} \\ &\quad \left\{ \sum_{l=0}^{\infty} \binom{n-k}{l} (-1)^l \left[\frac{(ay^c) a^{\frac{1+b}{c}} y^{-\frac{1+b}{c}} \varphi_{abc}}{N_o} \right]^l \right\}. \end{aligned} \quad (27)$$

$$\text{and } F_Y(y) = \sum_{j=k}^n \binom{n}{j} F^j(y) \{1 - F(y)\}^{n-j} \quad (28)$$

$$= F_Y(y) = \sum_{j=k}^n \sum_{l=0}^{n-j} \binom{n}{j} \binom{n-j}{l} (-1)^l F^{j+l}(y), \quad (29)$$

$$= \sum_{j=k}^n \sum_{l=0}^{\infty} \binom{n}{j} \binom{n-j}{l} (-1)^l \left[\frac{(ay^c) a^{\frac{1+b}{c}} y^{-\frac{1+b}{c}} \varphi_{abc}}{N_o} \right]^{j+l}. \quad (30)$$

By implication, the PDF of minimum order statistics is obtained by evaluating $j = k = 1$ in (27) to obtain:

$$f_{1:n} = ca^{1+b/c} \left\{ (1+y^b) e^{-ay^c} \right\} N_o^{-2} \left\{ \sum_{l=0}^{\infty} \binom{n-1}{l} (-1)^l \left[\frac{(ay^c) a^{\frac{1+b}{c}} y^{-\frac{1+b}{c}} \varphi_{abc}}{N_o} \right]^l \right\}. \quad (31)$$

where the corresponding PDF of the maximum order statistics is obtained at $j = k = n$ in (27):

$$\begin{aligned} f_{n:n} &= ca^{1+b/c} \left\{ (1+y^b) e^{-ay^c} \right\} N_o^{-2} \left\{ (ay^c) a^{\frac{1+b}{c}} y^{-\frac{1+b}{c}} \varphi_{abc} \right\}^{n-1} \\ &\quad \left\{ \sum_{l=0}^{\infty} \binom{0}{l} (-1)^l \left[\frac{(ay^c) a^{\frac{1+b}{c}} y^{-\frac{1+b}{c}} \varphi_{abc}}{N_o} \right]^l \right\}. \end{aligned} \quad (32)$$

4.5 The Quantile Function of the Omega Lindley Distribution

The expression $Q_x(p) = U^{-1}(p)$ gives the quantile function of the ω - Lindley distribution, where $0 < p < 1$ and the function U is the cumulative distribution

function. Now, recalling (10), the p^{th} quantile function is obtained by solving

$$U(x, \omega) = p. \quad (33)$$

Structurally, it may not be feasible to obtain the inverse of x as in $Q_x(p) = U^{-1}(p)$ due to the gamma components in the CDF; hence, we leave the expression as a non-linear equation; thus,

$$U(x, \omega) - p = 0. \quad (34)$$

Moreover, the median of the $\omega - Lindley$ distribution can be obtained by evaluating $p = 0.5$ in (34). The derivation of the quantile function of the proposed distribution is intentionally skipped due to the special functions in the CDF. Hence, we use algorithms in Wolfram Mathematica to generate random numbers.

4.6 Parameter Estimation

In this section, we present methods for estimating the parameters of the $\omega - Lindley$ distribution, which include the Cramer–von Mises estimator, the maximum product of spacing estimators and the maximum likelihood estimator.

4.6.1 Cramer–Von Mises (CVM) Estimator

Macdonald [20] proposed CVM as an estimation method with the objective of minimizing the following function:

$$V_c = (12n)^{-1} + \sum_{i=1}^n \left[F(x_{(i)}) - \frac{2i-1}{2n} \right]^2. \quad (35)$$

Now, by substituting the CDF of the $\omega - Lindley$ distribution $U(x, \omega)$ as in (10) into (35), we obtain

$$V_c(a, b, c) = (12n)^{-1} + \sum_{i=1}^n \left[A \left(ax_{(i)}^c x_{(i)}^{-\frac{1+b}{c}} \varphi_{abc} \right) - \frac{2i-1}{2n} \right]^2, \quad (36)$$

where $A = \frac{a^{\frac{1+b}{c}}}{N_0}$.

Using a software algorithm, (36) can be minimized to obtain the estimates of the CVM, or we can maximize it by solving

$$\frac{\partial V_c(a, b, c)}{\partial a} = 0; \quad \frac{\partial V_c(a, b, c)}{\partial b} = 0; \quad \frac{\partial V_c(a, b, c)}{\partial c} = 0. \quad (37)$$

4.6.2 Maximum Product of Spacing (MPS) Estimators

This method was originally used by Cheng and Amin [21] to estimate the parameters of lognormal distributions and is given as

$$M_p = \frac{1}{n+1} \sum_{i=1}^{n+1} \log R_i, \quad (38)$$

where $R_i = F(x_{(i)}) - F(x_{(i-1)})$ and $F(x_{(0)}) = 0$; $F(x_{(n+1)}) = 1$.

To obtain the function in (38) for the $\omega - Lindley$ distribution, we substitute the CDF $U(x, \omega)$ in $F(x_{(*)})$

$$M_p(a, b, c) = \frac{1}{n+1} \sum_{i=1}^{n+1} \log[U(x_{(i)}) - U(x_{(i-1)})], \quad (39)$$

$$= \frac{A}{n+1} \sum_{i=1}^{n+1} \log \left[\left((ax_{(i)}^c) x_{(i)}^{-\frac{1+b}{c}} \varphi_{abc}^{x_{(i)}} \right) - \left((ax_{(i-1)}^c) x_{(i-1)}^{-\frac{1+b}{c}} \varphi_{abc}^{x_{(i-1)}} \right) \right] \quad (40)$$

The MPS can be estimated by maximizing (40) at $\frac{\partial M_p(a, b, c)}{\partial \omega} = 0$, where

$$\frac{\partial M_p(a, b, c)}{\partial a} = \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{\partial}{\partial a} \{ \log[U(x_{(i)}) - U(x_{(i-1)})] \}. \quad (41)$$

$$\frac{\partial M_p(a, b, c)}{\partial b} = \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{\partial}{\partial b} \{ \log[U(x_{(i)}) - U(x_{(i-1)})] \}. \quad (42)$$

$$\frac{\partial M_p(a, b, c)}{\partial c} = \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{\partial}{\partial c} \{ \log[U(x_{(i)}) - U(x_{(i-1)})] \}. \quad (43)$$

4.6.3 Maximum Likelihood Estimator (MLE)

Myung [22] highlighted the objective of MLE, which is to obtain the parameter values of a model that maximizes the likelihood function over the spatial dimension of the parameters.

Let $x_i, i = 1, 2, \dots, n$ be a vector of observations from the $\omega - Lindley$ distribution; then, the log-likelihood function is defined by

$$\ln L = \sum_{i=1}^n \log f(x_i; \omega) = \sum_{i=1}^n \log \left[\frac{ca^{1+b/c} (1+x^b) e^{-ax^c}}{a^{b/c} \Gamma[\frac{1}{c}] + \Gamma[\frac{1+b}{c}]} \right], \quad (44)$$

$$= \log \left[\frac{ca^{1+b/c}}{a^{b/c} \Gamma[\frac{1}{c}] + \Gamma[\frac{1+b}{c}]} \right]^n + \sum_{i=1}^n \log(1+x^b) - a \sum_{i=1}^n x_i^c, \quad (45)$$

$$= n \log c + n \left(1 + \frac{b}{c}\right) \log a - n \left(\frac{b}{c}\right) \log a - \log \Gamma\left(\frac{1}{c}\right) \\ - \log \Gamma\left(\frac{1+b}{c}\right) + \sum_{i=1}^n \log(1+x^b) - a \sum_{i=1}^n x_i^c. \quad (46)$$

The score function for (46) is defined by

$$\frac{\partial \ln L}{\partial a} = n \left(1 + \frac{b}{c}\right) \left(\frac{1}{a}\right) - n \left(\frac{b}{c}\right) \left(\frac{1}{a}\right) - \sum_{i=1}^n x_i^c, \quad (47)$$

$$\frac{n}{a} - \sum_{i=1}^n x_i^c = 0.$$

$$\frac{\partial \ln L}{\partial b} = n \left(\frac{1}{c}\right) \log a - n \left(\frac{1}{c}\right) \log a - \frac{P_g(0, \frac{1+b}{c})}{c} + \sum_{i=1}^n \frac{x^b \log x}{(1+x^b)}, \quad (48)$$

$$\sum_{i=1}^n \frac{x_i^b \log x}{(1+x_i^b)} - \frac{P_g(0, \frac{1+b}{c})}{c} = 0.$$

$$\frac{\partial \ln L}{\partial c} = \frac{n}{c} - n \left(\frac{b}{c^2}\right) \log a + n \left(\frac{b}{c^2}\right) \log a + \frac{P_g(0, \frac{1}{c})}{c^2} \\ + \frac{(1+b)P_g(0, \frac{1+b}{c})}{c^2} - a \sum_{i=1}^n x_i^c \log(x_i), \quad (49)$$

$$\frac{n}{c} + \frac{P_g(0, \frac{1}{c})}{c^2} + \frac{(1+b)P_g(0, \frac{1+b}{c})}{c^2} - a \sum_{i=1}^n x_i^c \log(x_i) = 0.$$

where $P_g(*, *)$ is poly-gamma.

It can be clearly seen that (47, 48 and 49) is in close form. Therefore, a numerical analysis such as the Newton-Raphson iterative method, which is a root finding algorithm, can be used to obtain the MLEs of \hat{a} , \hat{b} and \hat{c} . This scheme is given by

$$\hat{\omega} = \omega - H^{-1}(\omega)S(\omega). \quad (50)$$

where $S(\omega)$ is the score function and $H^{-1}(\omega)$ is the second derivative of the log-likelihood function termed the Hessian matrix. Refer to [23, 24] for a detailed study.

4.7 Interval Estimation

Furthermore, we present the asymptotic confidence intervals for the parameters of the $\omega - Lindley$ distribution. In this attempt, the Fisher information matrix is adopted with the goal of measuring the amount of information that a random vector of variables carries about unknown parameters of a probability model that forecasts for $X_i, i = 1, 2, 3, \dots n$. The information matrix is given by

$$I(\omega_k) = -E[H(\omega_k)] = -E \begin{bmatrix} \frac{\partial^2 l}{\partial a^2} & \frac{\partial^2 l}{\partial a \partial b} & \frac{\partial^2 l}{\partial a \partial c} \\ \frac{\partial^2 l}{\partial a \partial b} & \frac{\partial^2 l}{\partial b^2} & \frac{\partial^2 l}{\partial b \partial c} \\ \frac{\partial^2 l}{\partial a \partial c} & \frac{\partial^2 l}{\partial b \partial c} & \frac{\partial^2 l}{\partial c^2} \end{bmatrix}, \omega = (a, b, c)^T. \tag{51}$$

Under the normality condition $\hat{\omega} \sim N(\omega, I^{-1}(\omega))$, is obtained by inverting (51). Consequently, the interval estimate is obtained by the approximation of $(1 - \alpha)100$ confidence intervals constructed for the parameters a, b and c :

$\hat{a} \pm Z_{\alpha/2} \sqrt{var(\hat{a})}; \hat{b} \pm Z_{\alpha/2} \sqrt{var(\hat{b})}; \hat{c} \pm Z_{\alpha/2} \sqrt{var(\hat{c})}$. where $var(\omega)$ are the variances of the parameters, which are given by the respective diagonal elements of the variance–covariance matrix $I^{-1}(\omega_k)$, and $Z_{\alpha/2}$ is the upper percentile of the standard normal distribution.

5 Real Data Analysis

The analyses of two real-life datasets are detailed in this section to demonstrate the applicability of the proposed distribution. Ghitany [15] presented a dataset that measures the waiting time (in minutes) of 100 bank customers, whereas the second dataset represents the remission time of 128 bladder cancer patients, as reported in [25]. The omega $\omega - Lindley$ distribution is fitted to these data alongside other Lindley families of distributions:

- The three-parameter generalized Lindley distribution (TPGLD) was described in [18]:

$$g(x) = \frac{ac^2 e^{-cx^a} x^{-1+a} (b + x^a)}{1 + bc}, x > 0, a, b, c > 0. \tag{52}$$

- The new generalized two parameter Lindley distribution (NG2PLD) was described in [17]:

$$g(x) = \frac{a^2 e^{-ax} \left(1 + \frac{a^{-2+b} x^{-1+b}}{\Gamma[b]} \right)}{1 + a}, x > 0, a, b > 0. \tag{53}$$

Table 4 Parameters for the comparative analysis of dataset 1

Models	a	b	c
$\omega - Lindley$	0.21	3.4	1.2
TPGLD	1.48	0.0001	0.33
NG2PLD	0.25	1.69	
PLD	1.18	0.46	
LD	0.34		

Table 5 Inferential results of the proposed model and selected Lindley families of distribution (data 1)

Models	$\mathcal{A}_D(P_{value})$	C_{vM} (P_{value})	χ^2 (P_{value})
$\omega - Lindley$	14.0609 (6.2×10^{-6})	1.6761 (6.9×10^{-5})	68.22 (6.9×10^{-10})
TPGLD	303.825 (0.)	16.4226 (0.)	441.52 (6.0×10^{-87})
NG2PLD	19.0733 (6.0×10^{-6})	3.3672 (1.2×10^{-8})	43.78 (1.7×10^{-5})
PLD	190.976 (0.)	16.9311 (0.)	474.86 (4.93×10^{-8})
LD	37.8204 (0.)	5.9551 (2.53×10^{-14})	90.32 (4.2×10^{-14})

Table 6 Parameters for the comparative analysis of data 2

Models	a	b	c
$\omega - Lindley$	0.55	0.12	0.46
TPGLD	0.38	0.97	1.27
N2PLD	0.17	0.08	
PLD	0.25	0.5	
LD	0.5		

The power Lindley distribution (PLD) was described in [16]:

$$g(x) = \frac{ab^2 e^{-bx^a} (1+x^a)x^{a-1}}{1+b}, \quad x > 0, a, b > 0. \quad (54)$$

Table 7 Inferential results of the proposed model and selected Lindley families of distribution (data 2)

Models	$4_D(P_{value})$	$C_{v\mathcal{M}}$ (P_{value})	χ^2 (P_{value})
$\omega - Lindley$	32.5098 (0.)	7.0544 (0.)	95.3438 (1.3×10^{-14})
TPGLD	76.6006 (0.)	15.796 (0.)	190.938 (9.9×10^{-34})
N2PLD	181.491 (0.)	28.8656 (0.)	556.688 (1.3×10^{-110})
PLD	47.8123 (0.)	10.5599 (0.)	324.375 (1.8×10^{-61})
LD	85.2727 (0.)	9.9772 (0.)	216.75 (4.9×10^{-39})

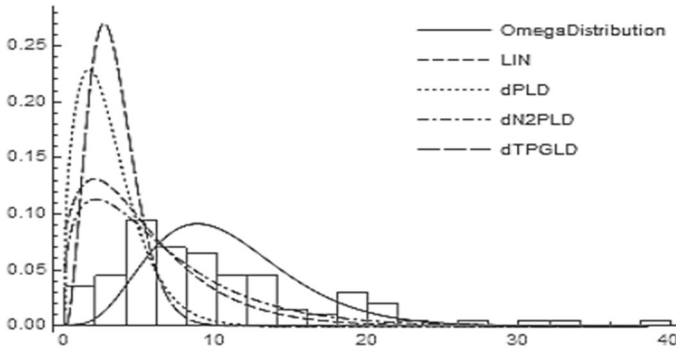


Fig. 4 Density fit plots for dataset 1

The Lindley distribution (LD) was described in [14]:

$$g(x) = \frac{a^2 e^{-ax}(1+x)}{1+a}, \quad x > 0, a > 0. \tag{55}$$

The measurement criteria used for performance comparison in this work were the Cramer–von Mises ($C_{v\mathcal{M}}$), Anderson Darling (4_D), and Pearson chi-square (χ^2) test statistics and their corresponding P values. These automatically selected statistics are

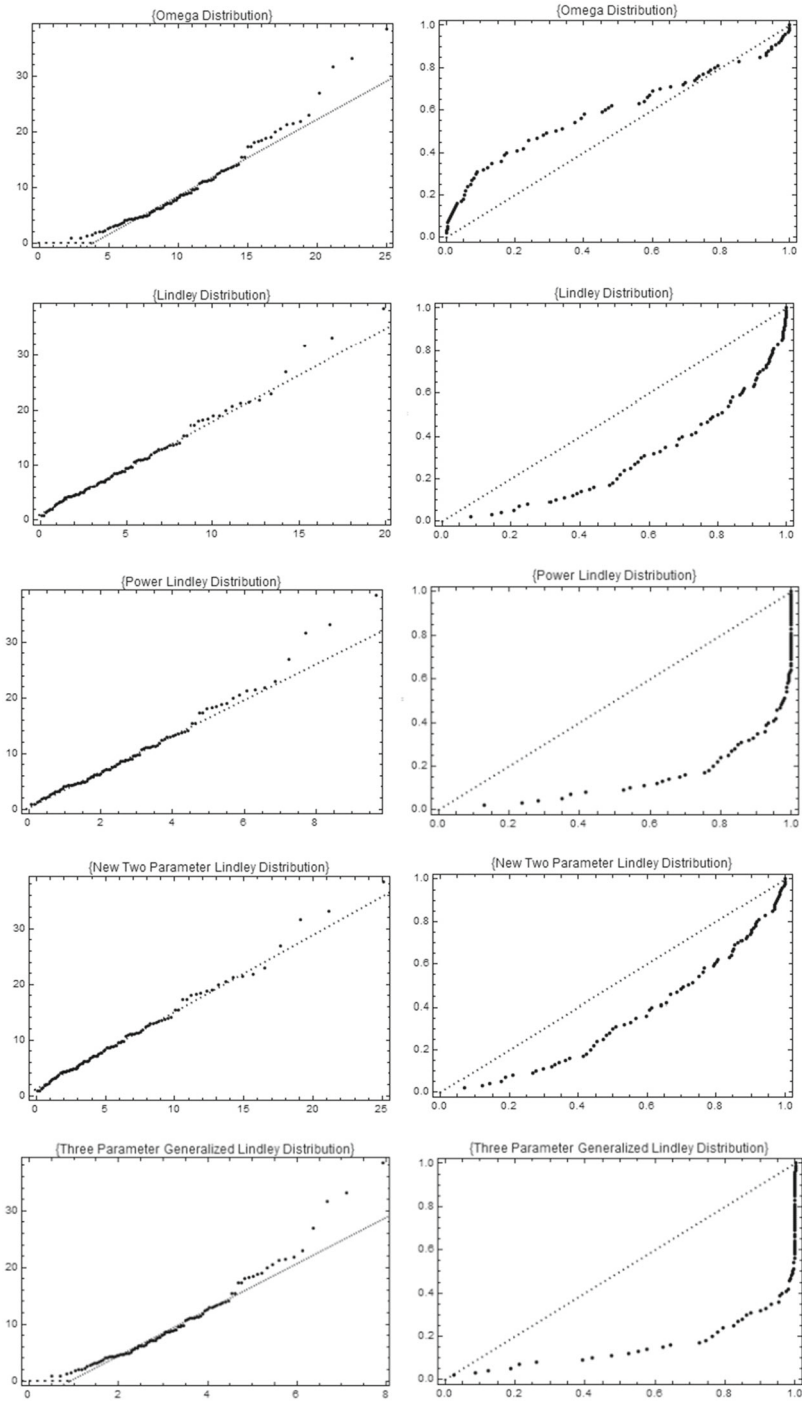


Fig. 5 Fitted quantile (q-q) plot [left] and probability (p-p) plot [right] for data 1

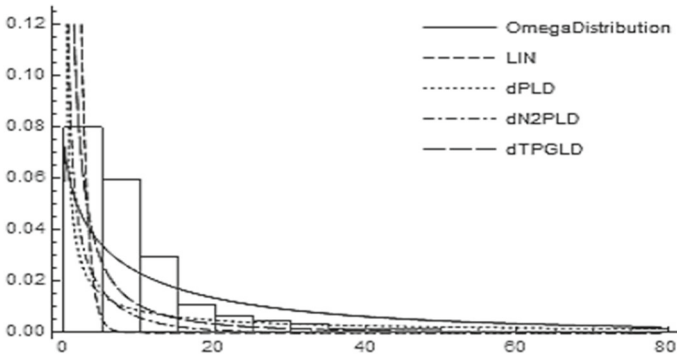


Fig. 6 Density fit plots for dataset 2

based on the following:

$$C_{v\mathcal{M}} \sim \text{Expectation} \left\{ \left[\hat{F}(x) - F(x) \right]^2 \right\}$$

$4_D \sim \text{Expectation} \left\{ \frac{[\hat{F}(x) - F(x)]^2}{F(x)(1 - F(x))} \right\}$, $\chi^2 \sim \sum_{i=1}^n \frac{(o_i - e_i)^2}{e_i}$, where $F(x)$ and $\hat{F}(x)$ are the CDFs of the distributions and the empirical CDFs of the data, respectively, and o_i and e_i are the observed and expected frequencies, respectively. The parameter estimates and analytical results are summarized in Tables 4, 5, 6 and 7, and Figs. 4, 5, 6 and 7 are the supporting diagrammatical analogy of the empirical inference.

Judging from the inferential measures, which project comparative goodness based on the least empirical status of the three automatically selected test criteria, it can clearly be observed from Tables 5 and 7 that the proposed distribution is a better fit for the datasets. This is also proven diagrammatically according to Figs. 4 and 6, which represent the density fit plots. Further support for the results can be found in Figs. 5 and 7, which include quantile and probability (q-q and p-p) plots.

6 Concluding Remarks

In this research, a more efficient approach to increasing the number of parameters in an existing distribution apart from the generalization method is presented, and by applying this concept, a new distribution is proposed. Some mathematical properties of the proposed distribution, including series representations of the PDF and CDF, density, hazard and survival shapes, moments and related measures, quantile functions, parameter estimations, interval estimates and order statistics, are derived and/or realized. Due to the special functions in the CDF, both the hazard and survival functions are rendered in a closed-form structure. As a result, they yield both monotonic and nonmonotonic trends, which is a great economy in data modeling. The survival

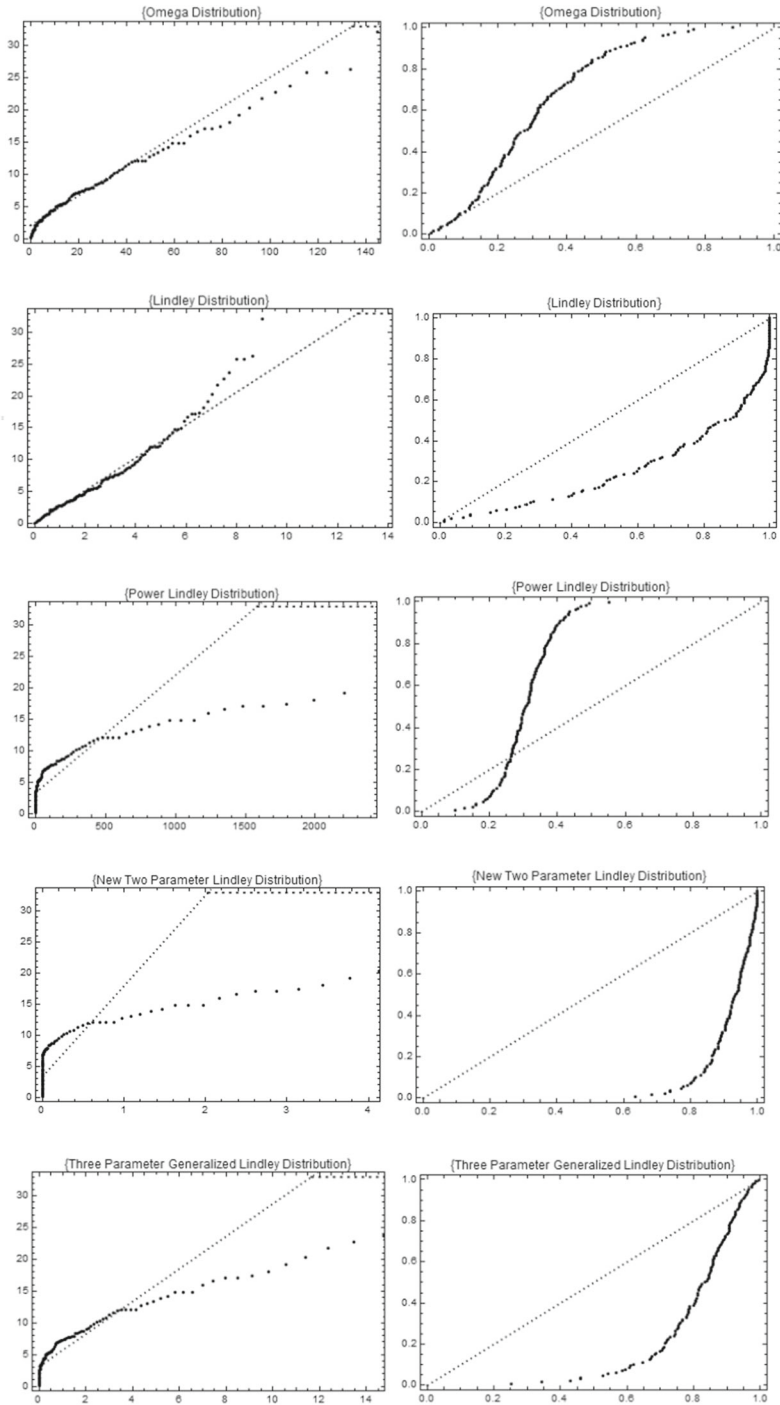


Fig. 7 Fitted quantile (q-q) plot [left] and probability (p-p) plot [right] for dataset 2

function uniquely features a uniform trend, which implies that the proposed distribution can model systems or components whose lifespan is sustained over a given range of time. The parameter estimation is theoretically studied under three different methods, including the Cramer Von Mise, maximum product of spacing and maximum likelihood approaches. Finally, the proposed model is fitted to two real-life datasets to assess its flexibility. Based on the obtained empirical and diagrammatical results, the proposed distribution can better forecast the datasets than several other rival lifetime models. The proposed parameter inclusion strategy is expected to be adopted in the field of mathematical statistics, and the proposed model has wider applications in areas consistent with survival data analysis.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40745-024-00539-y>.

Author Contributions The authors confirm sole responsibility for the following: study conception and design, data coding, analysis and interpretation of results, and manuscript preparation.

Funding No funding was received for conducting this study.

Data availability The dataset used in this study are available to the public under a Creative Common lincence at <https://doi.org/10.14281/18241.20>.

Code availability The coding used in this study are provided in the supplementary section.

Declarations

Conflict of interest Not applicable.

Ethical statement All of the followed procedures were in accordance with the ethical and scientific standards. This article does not contain any studies with human participants by the author.

References

1. Shi Y (2022) *Advances in big data analytics: theory algorithm and practice*. Springer, Singapore
2. Olson DL, Shi Y (2007) *Introduction to business data mining*. McGraw-Hill/Irwin, New York
3. Shi Y, Tian YJ, Kou G, Peng Y, Li JP (2011) *Optimization based data mining: theory and applications*. Springer, Berlin
4. Tien JM (2017) Internet of things, real-time decision making, and artificial intelligence. *Ann Data Sci* 4(2):149–178
5. Lai CD (2013) Constructions and applications of lifetime distributions. *Appl Stoch Model Bus Ind* 29(2):127–140
6. Tahir MH, Cordeiro GM, Alzaatreh A, Mansoor M, Zubair M (2016) The logistic-X family of distributions and its applications. *Commun Stat Theory Methods* 45(24):7326–7349
7. Oguntunde PE, Khaleel MA, Adejumo AO, Okagbue HI, Opanuga AA, Owolabi FO (2018) The Gompertz inverse exponential (GoIE) distribution with applications. *Cogent math Stat* 5(1):1507122
8. Souza L, Junior W, De Brito C, Chesneau C, Ferreira T, Soares L (2019) On the Sin-G class of distributions: theory, model and application. *J Math Model* 7(3):357–379
9. Klakattawi H, Alsulami D, Elaal MA, Dey S, Baharith L (2022) A new generalized family of distributions based on combining Marshal–Olkin transformation with TX family. *PLoS ONE* 17(2):e0263673
10. Kilai M, Waititu GA, Kibira WA, Alshanbari HM, El-Morshedy M (2022) A new generalization of Gull alpha power family of distributions with application to modeling COVID-19 mortality rates. *Results Phys* 36:105339

11. Ekum MI, Adamu MO, Akarawak EE (2022) Normal-power-logistic distribution: properties and application in generalized linear model. *J Indian Soc Probab Stat* 24:1–32
12. Rodrigues GM, Ortega EM, Cordeiro GM, Vila R (2023) Quantile regression with a new exponentiated odd log-logistic weibull distribution. *Mathematics* 11(6):1518
13. Feller W (1968) *An introduction to probability theory and its applications*, 3rd edn. Wiley, Hoboken
14. Lindley DV (1958) Fiducial distributions and Bayes' theorem. *J R Stat Soc Ser B Methodol* 20(1):102–107
15. Ghitany ME, Atieh B, Nadarajah S (2008) Lindley distribution and its application. *Math Comput Simul* 78(4):493–506
16. Ghitany M, Al-Mutairi D, Balakrishnan N, Al-Enezi I (2013) Power Lindley distribution and associated inference. *Comput Stat Data Anal* 64:20–33
17. Ekhosuehi N, Opone FC, Odobaire F (2018) A new generalized two-parameter lindley distribution (NG2PLD). *J Data Sci* 16(3):459–466
18. Ekhosuehi N, Opone F (2018) A three parameter generalized lindley distribution: its properties and application. *Statistica (Bologna)* 78(3):233–249
19. Opone F, Ekhosuehi N, Omosigbo S (2022) Topp–Leone power lindley distribution (TLPLD): Its properties and application. *Sankhya A* 84:597–608
20. Macdonald PDM (1971) Comments and queries comment on “an estimation procedure for mixtures of distributions” by Choi and Bulgren. *J R Stat Soc Ser B Methodol* 33(2):326–329
21. Cheng RCH, Amin NAK (1983) Estimating parameters in continuous univariate distributions with a shifted origin. *J R Stat Soc Ser B Methodol* 45(3):394–403
22. Myung IJ (2003) Tutorial on maximum likelihood estimation. *J Math Psychol* 47(1):90–100
23. Akram S, Ann QU (2015) Newton raphson method. *Int J Sci Eng Res* 6(7):1748–1752
24. Bayat M, Koushki MM, Ghadimi AA, Tostado-Véliz M, Jurado F (2022) Comprehensive enhanced Newton Raphson approach for power flow analysis in droop-controlled islanded AC microgrids. *Int J Electr Power Energy Syst* 143:108493
25. Lee ET, Wang J (2003) *Statistical methods for survival data analysis*. Wiley, London

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.