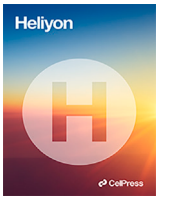
Heliyon 10 (2024) e37242

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

A novel two–parameter unit probability model with properties and applications

Zawar Hussain [a], Farrukh Jamal [a], Abdus Saboor [b,*], Shakaiba Shafiq [a], Arshid Khan [b], Shahida Perveen [b], Fuad A. Awwad [c], Emad A.A. Ismail [c], Musharraf Ali [d]

[a] Department of Statistics, The Islamia University of Bahawalpur, Punjab 63100, Pakistan
[b] Institute of Numerical Sciences, Kohat University of Science & Technology, Kohat, KP, 26000, Pakistan
[c] Department of Quantitative Analysis, College of Business Administration, King Saud University, P.O. Box 71115, Riyadh 11587, Saudi Arabia
[d] Department of Mathematics, G.F. College, Shahjahanpur, Affiliated M.J.P. Rohilkhand University, Bareilly, India

ARTICLE INFO

MSC:
60E05
62E15

Keywords:
NTPUP model
Order statistics
Maximum likelihood method
Simulation analysis

ABSTRACT

This paper develops a novel two-parameter unit probability model which is the generalized form Kumaraswami distribution that exhibits greater flexibility compared to well-known existing distributions, attributed to its distinct hazard and density function shapes. Extensive analysis has been conducted to explore numerous statistical features of the specified distribution, specifically moments, and order statistics providing explicit expressions for these measures. The maximum likelihood estimation is employed to estimate the model parameters and a numerical simulation analysis confirms the consistency of this estimation approach. Furthermore, the applicability of the specified model is demonstrated by considering four real data sets, showcasing its effectiveness in capturing the characteristics of real life data. The proposed model shows promise as a versatile tool for analyzing diverse data sets in a wide range of fields.

1. Introduction

Distribution theory plays a pivotal role in diverse scientific disciplines, guiding the selection of appropriate models to analyze complex data. The statistical literature offers a wide array of distributions tailored for modeling lifetime data, each chosen based on data characteristics and distributional assumptions. For instance, the gamma and the Weibull distributions are often preferred for modeling right-skewed data, while the ubiquitous normal distribution suits bell-shaped data sets spanning the real line. However, these distributions possesses some limitations particularly in their inability to accommodate non-monotonic shapes. To overcome this drawback and enhance the flexibility of the existing distributions, numerous extensions have been developed over time. The pursuit of a more versatile and robust probability distributions stems from the inherent complexity and diversity of real-world data. Researchers continually endeavor and efficiency of data analysis across various scientific and industrial applications. Continuous and discrete distributions can also be compounded to form a new model to improve the quality of existing distributions like G-families is also the generalization technique to form new distributions.

Although many unbounded models have been developed but there is still lack of bounded models for developing models for fitting specific real world scenarios. Unbounded models, as in most cases, are unable to adequately model the data. In situations

* Corresponding author.
E-mail addresses: saboorhangu@gmail.com, dr.abdussaboor@kust.edu.pk (A. Saboor).

https://doi.org/10.1016/j.heliyon.2024.e37242
Received 7 January 2024; Received in revised form 29 August 2024; Accepted 29 August 2024

Available online 3 September 2024
2405-8440/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

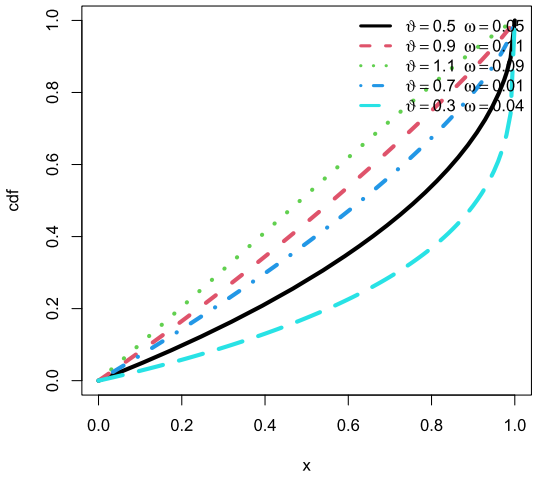


Fig. 1. cdf plots of the NTPUP model for various parametric values.

when boundary conditions, such the proportion of a certain attribute, are unpredictable, it is necessary to represent the occurrences using a continuous distribution with a boundary. The renowned Beta model [15] is the only continuous bounded model that has been extensively reviewed in the literature. Recently, a number of continuous models with bounded domains have been discussed and utilized to model unpredictability of a bounded case in multiple applied areas of research. For instance, the log-Lindley [3], the unit-logistic [5], the unit-gamma [4], the Kumaraswamy unit [6], the Topp-Leone [7], the Simplex [8] and the McDonald's generalized beta type I [15] models.

This manuscript aims to offer a new two–parameter unit probability (NTPUP) model. By employing the Laymen Alternative II (LAII) process, the development of this model is organized methodically as: $F(x:\vartheta) = 1-(1-G(x))^{\vartheta}$, where $F(x:\vartheta)$ belongs to the cdf of the $LAII$ model [1]. The cumulative distribution function (cdf) and the probability density function (pdf) of NTPUP model are respectively defined by:

$$F_{NTPUP}(x;\vartheta,\omega) = 1-(1-x)^{\vartheta}(1+x)^{\omega}, \quad x\in(0,1),\ \vartheta\geq\omega,\ \vartheta\in\Re^{+}, \quad \omega\in\Re, \tag{1}$$

and

$$f_{NTPUP}(x;\vartheta,\omega) = (1-x)^{\vartheta-1}(1+x)^{\omega-1}\left(x(\vartheta+\omega)+\vartheta-\omega\right). \tag{2}$$

The cdf of the NTPUP model is monotonically an increasing function: $\lim_{x\to 0} F_{NTPUP}(x;\vartheta,\omega) = 0$ and $\lim_{x\to 1} F_{NTPUP}(x;\vartheta,\omega) = 1$. Moreover, Fig. 1 makes it clear that the NTPUP model's cdf is strictly an increasing function.

The pdf plots of the NTPUP model for multiple values of parameters are portrayed in Fig. 2. The NTPUP model exposed a variety of shapes be altering the parametric inputs, including the parabolic, j-shaped, bathtub, right-skewed, and reversed j-shaped geometrics that are given in the Fig. 2.

The cdf and pdf defined in Eqns. (1) and (2), respectively can be written in the form of mixture representation as:

$$F_{NTPUP}(x;\vartheta,\omega) = 1-\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\binom{\vartheta}{j}\binom{\omega}{i}(-1)^{j}x^{i+j} \tag{3}$$

and

$$f_{NTPUP}(x;\vartheta,\omega) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\binom{\vartheta-1}{i}\binom{\omega-1}{j}(-1)^{j}x^{i+j}(\vartheta-\omega+x(\vartheta+\omega)). \tag{4}$$

The motivations for the NTPUP model are as follows:

- The NTPUP model is presented as a flexible, bounded distribution that may match a variety of data forms, including non-monotonic patterns such as parabolic and bathtub curves.
- Because it is specified on the interval $[0,1]$, unlike traditional distributions, it is suitable for data that is limited to specific ranges, such proportions or percentages.
- The NTPUP model's parameterizations facilitates reading and modification, increasing its applicability in various scientific and practical sectors where traditional distributions would not be as efficient.
- For a comprehensive statistical description, the NTPUP provides expressions for kurtosis, skewness, entropy and moments.
- A broadly employed method known for its efficiency and strong statistical properties, maximum likelihood estimation (MLE), is utilized to estimate parameters.

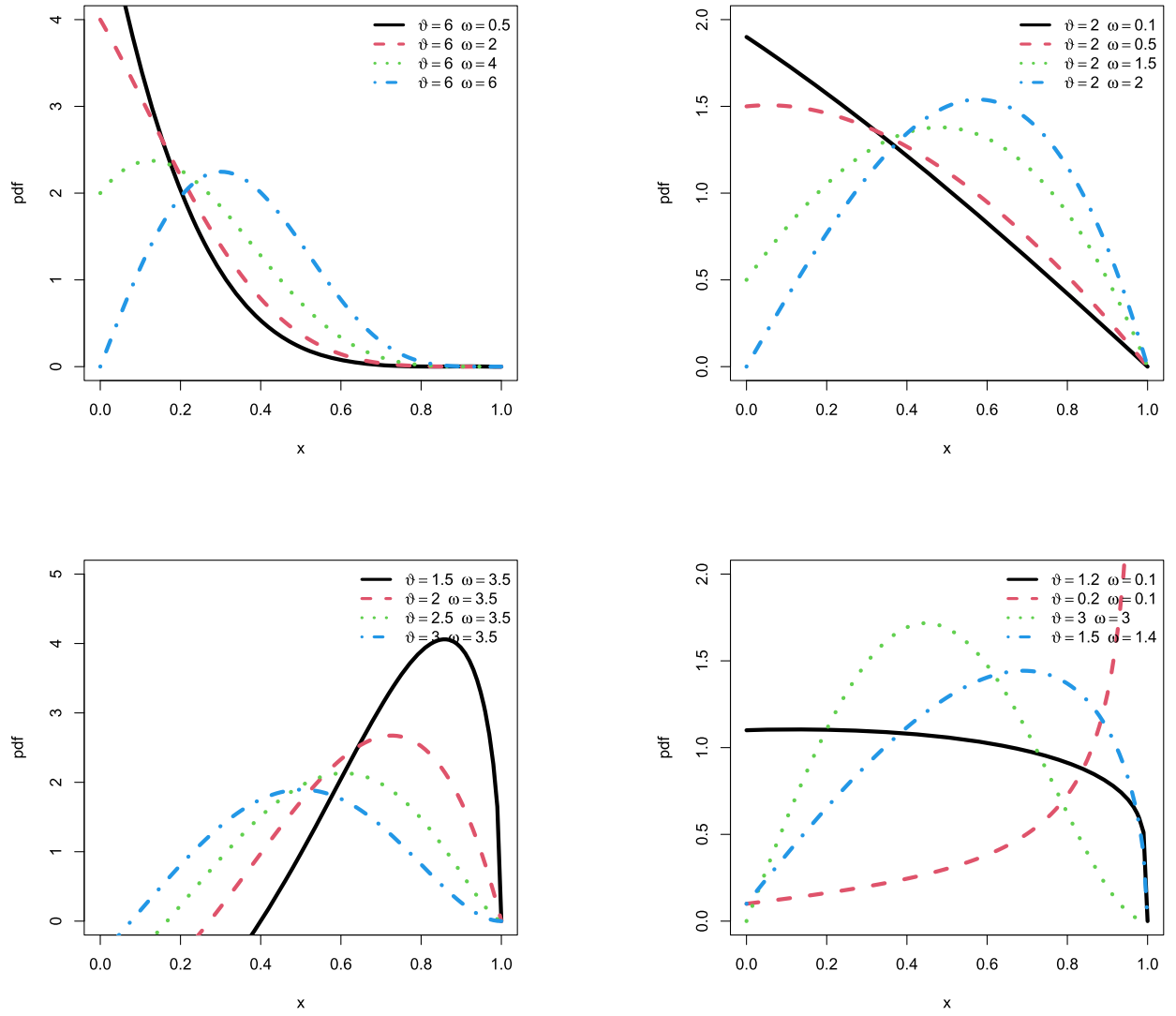


Fig. 2. Plots of pdf of the NTPUP distribution for various parametric values.

- Suited to four distinct real data sets successfully, proving its robustness and suitability for real-world data processing.

The remaining part of the manuscript is ordered as: Some properties of NTPUP model have been derived in Section 2. Estimation of parameters have been obtained and simulation analysis is executed in Section 3. Applications to real world data sets have been done in Section 4. The final concluding remarks of this study have been given in Section 5.

2. Statistical functions

The survival function (SF) denoted by $S(x;\vartheta,\omega)$ is obtained from Eqn. (1) in the following form:

$$S_{NTPUP}(x;\vartheta,\omega) = (1-x)^{\vartheta}(1+x)^{\omega}. \tag{5}$$

The hazard rate function (HRF) describes the instantaneous rate of failure at any given time. The HRF of the NTPUP model denoted as: $H_{NTPUP}(x;\vartheta,\omega) = f(x;\vartheta,\omega)/S(x;\vartheta,\omega)$ is formulated by using Eqn. (2) and Eqn. (5)

$$H_{NTPUP}(x;\vartheta,\omega) = -\frac{\vartheta}{x-1} - \frac{\omega}{x+1}. \tag{6}$$

The rate at which an event occurs given that it has not occurred until a specified time is described by the reversed hazard rate function (RHRF), also referred to as the hazard rate for surviving past a certain time. The RHRF of the NTPUP model denoted as: $h_{NTPUP}(x;\vartheta,\omega) = f(x;\vartheta,\omega)/F(x;\vartheta,\omega)$, is laid out using Eqn. (2) and Eqn. (1) as:

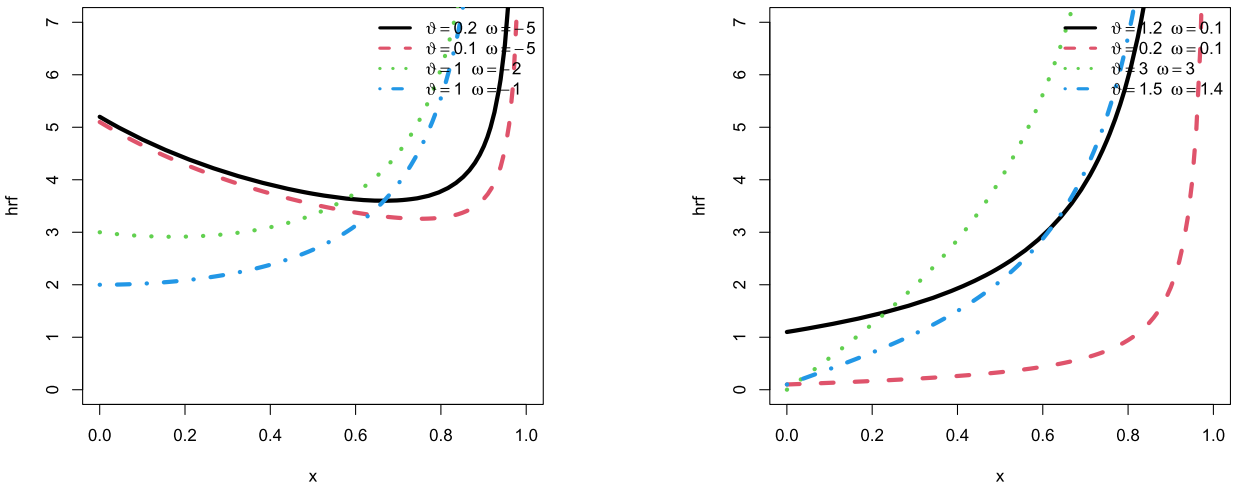


Fig. 3. Plots of HRF of NTPUP model for various parametric values.

$$h_{NTPUP}(x;\vartheta,\omega)=\frac{\vartheta-\omega+x(\vartheta+\omega)}{(1-x)(1+x)((1-x)^{-\vartheta}(1+x)^{-\omega}-1)}. \tag{7}$$

The visualizations of HRF of the NTPUP model are presented in the Fig. 3 depicting the bathtub and increasing forms. The bathtub shape represents early flaws, random failures, and wear-out failures, respectively, with decreasing, constant and increasing failure rates. The aging impacts in the system are highlighted by the increasing shapes, which shows an increasing failure rate over time.

2.1. Moments and moment generating function

Moments like variance and mean, are quantitative metrics that depict the form of a probability distribution. Additionally, the moment generating function (MGF) uses the variable's expected exponential to encode every moment of a random variable. These ideas are crucial because the MGF may be used to precisely identify the distribution and make the computation of moments easier. Moments also provide crucial information about the features of the distribution. The r^{th} moment of the NTPUP model using Eqn. (2) is derived as:

$$E_{NTPUP}(x^r,\vartheta,\omega)=\int_0^1 x^r f_{NTPUP}(x;\vartheta,\omega)dx, \tag{8}$$

putting the value of $f_{NTPUP}(x;\vartheta,\omega)$ from Eqn. (4) we obtained

$$E_{NTPUP}(x^r,\vartheta,\omega)=\int_0^1 x^r\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\binom{\vartheta-1}{i}\binom{\omega-1}{j}(-1)^j x^{i+j}(\vartheta-\omega+x(\vartheta+\omega))dx, \tag{9}$$

after simplification of Eqn. (9), we get

$$E_{NTPUP}(r,\vartheta,\omega)=\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{\binom{\vartheta-1}{i}\binom{\omega-1}{j}(-1)^i(2\vartheta(r+i+j)+3\vartheta+\omega)}{(r+i+j+1)(r+i+j+2)}. \tag{10}$$

By putting $r=1$ in Eqn. (10), we get the first moment

$$E_{NTPUP}(1;\vartheta,\omega)=\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{\binom{\vartheta-1}{i}\binom{\omega-1}{j}(-1)^i(2\vartheta(1+i+j)+3\vartheta+\omega)}{(i+j+2)(i+j+3)}.$$

Similarly the MGF of NTPUP model using Eqn. (2) is obtained as:

$$M_{\vartheta,\omega}(t)=\sum_{i=0}^{\infty}\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{t^i\binom{\vartheta-1}{n}\binom{\omega-1}{m}(-1)^n(2\vartheta(1+n+m)+3\vartheta+\omega)}{i!(n+k+2)(n+k+3)}.$$

Table 1 is furnished using R programming language to compute some probable quantitative values of the first four moments, variance, standard deviation, covariance, skewness and kurtosis of the NTPUP model for some parametric values. Fig. 4 represents the 3D plots of mean and variance of NTPUP model. An illustration of how the NTPUP model's mean value varies with different values of ϑ and

Table 1
Some possible numerical values of the first four moments, variance, standard deviation, covariance, skewness and kurtosis of the NTPUP model for some parameter values.

(ϑ,ω)	m_1	m_2	m_3	m_4	$var(x)$	$S.D$	CV	CS	CK
(1.2,1.2)	0.632	0.455	0.351	0.284	0.055	0.234	0.371	-0.437	2.279
(1.5,1.2)	0.543	0.356	0.257	0.197	0.062	0.249	0.459	-0.199	2.083
(1.2,-1.2)	0.340	0.183	0.120	0.087	0.068	0.260	0.765	0.663	2.415
(1.5,-1.2)	0.306	0.151	0.092	0.063	0.058	0.240	0.785	0.778	2.707
(1.8,1.2)	0.474	0.286	0.194	0.141	0.061	0.248	0.523	0.011	2.031
(1.8,1.5)	0.512	0.319	0.221	0.163	0.058	0.240	0.469	-0.107	2.100
(2.0,1.5)	0.470	0.278	0.185	0.132	0.057	0.239	0.509	0.014	2.081
(2.0,1.8)	0.507	0.310	0.210	0.152	0.053	0.231	0.455	-0.087	2.145
(3.0,1.8)	0.352	0.170	0.096	0.060	0.046	0.215	0.612	0.371	2.311
(3.2,2.0)	0.344	0.163	0.090	0.055	0.044	0.211	0.611	0.383	2.344

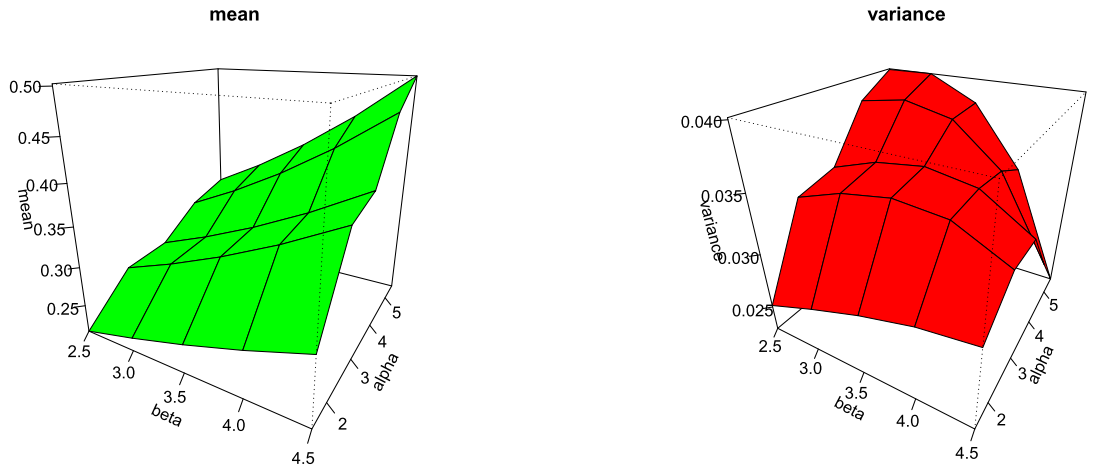


Fig. 4. 3D plots of NTPUP model for various parametric values.

ω can be seen in the 3D plot of the mean. Examining this plot allows one to see how these characteristics interact and relate to one another, as well as how they effect the distribution's central tendency. Plotting's peak and valleys show areas where the mean is substantially impacted, which can help choose parameters for particular uses. In a similar vein, the NTPUP model's dispersion or variability is dependent on the parameters, as demonstrated by the 3D plot of variance in the Fig. 4. Understanding the stability and spread of the distribution under various parameter settings is made easier with the help of the Fig. 4.

2.2. Rényi entropy

Rényi entropy (RE), which generated the idea of the Shannon entropy by introducing a parameter that modifies the weight assigned to various probabilities within the distribution, is a metric of the diversity, unpredictability, or randomness of a probability distribution.

The RE of the NTPUP model is derived as:

$$R_{NTPUP}(p;\vartheta,\omega)=\frac{1}{1-p}\log[\int_0^1((1-x)^{\vartheta-1}(1+x)^{\omega-1}\,(x(\vartheta+\omega)+\vartheta-\omega))^p dx]. \tag{11}$$

Some analytical outcomes of the RE for multiple parametric values are provided in the Table 2. These values shed light on how varying parameter settings affect the NTPUP model's diversity and unpredictability. Identifying patterns and trends, notably an upsurge or decrease in entropy, can help one understand how responsive the model is to changes in the parameters. Fig. 5 showcase the 3D representation of RE for the NTPUP model. By highlighting areas of higher or lower entropy, this graphical representation makes it easier to understand how the entropy varies. Plotting reveals that parameter combinations with higher entropy indicate greater diversity and randomness, whereas lower entropy indicated greater predictability.

Table 2
Numerical values of the RE of NTPUP distribution for some parametric values.

Parameters	(1.2,-1.2)	(1.2,1.2)	(1.5,1.2)	(1.5,1.5)	(2.0,1.8)	(3,2)	(5,3.4)	(5,4)	(5.5,4.5)
Values	-0.3947	-0.3951	-0.568	-0.749	-0.872	-1.809	-0.437	-2.257	-2.658

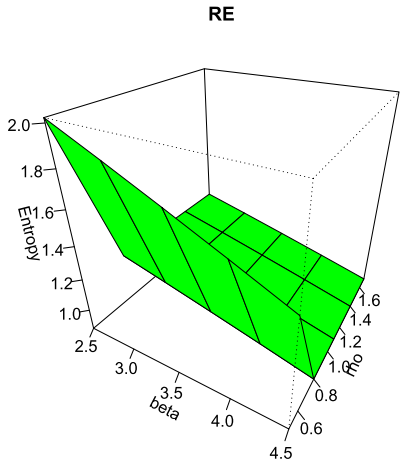


Fig. 5. 3D plot of RE of NTPUP distribution.

2.3. Order statistics

Order statistics (OS), which include the minimum, maximum, and different percentiles, are the sorted values of a random sample that offer crucial details about the sample's variability and distribution.

For the NTPUP model, the pdf of the OS is derived as:

$$f_{i:n}(x,\vartheta,\omega)=\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{n-i}\sum_{m=0}^{\infty}\binom{n-i}{l}\binom{i+l-1}{j}\binom{\omega j+\omega-1}{m}\binom{\vartheta j+\vartheta-1}{k}(-1)^{l+j+k}x^{m+k}(\vartheta-\omega+x(\vartheta+\omega)).$$

2.4. Mean residual life function

Mean residual life (MRL) is the predicted remaining lifetime of a system or component provided that it has endured up to a certain time, offering perspectives on its potential for reliability and longevity. The NTPUP model's MRL function can be formulated as follows:

$$m(t;\vartheta,\omega)=\sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\frac{\binom{\vartheta-1}{k}\binom{\omega-1}{n}(-1)^{k}\left(\frac{(\vartheta+\omega)\left(1+\theta^{n+k+3}\right)}{n+k+3}+\frac{(\vartheta-\omega)\left(1+\theta^{n+k+2}\right)}{n+k+2}\right)}{(1-x)^{\vartheta}(1+x)^{\omega}}.$$

2.5. Mean waiting time function

One of the most important statistical feature which is widely used in queuing theory to gauge efficacy and performance is the mean waiting time (MWT). The MWT predicts that at what length of time one should anticipate to wait for a particular event to happen.

The NTPUP model's MWT is formulated as:

$$\overline{\mu}(x;\vartheta,\omega)=t-\sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\frac{\binom{\vartheta-1}{k}\binom{\omega-1}{n}(-1)^{k}\,t^{n+k+2}\left(\frac{(\vartheta+\omega)t}{n+k+3}+\frac{(\vartheta-\omega)t}{n+k+2}\right)}{1-(1-t)^{\vartheta}(1+t)^{\omega}}.$$

2.6. Lorenz and Bonferoni curve

The Lorenz (L) and Bonferoni (B) curves which are frequently used in the social sciences and economics to evaluate income equity. Policymakers can examine and resolve issues of social equilibrium and economic disparity by using these two useful curves.

The L and B curves, respectively are derived for the NTPUP model in the following form:

$$L(p;\vartheta,\omega)=\frac{1}{\mu}\sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\binom{\vartheta-1}{k}\binom{\omega-1}{n}(-1)^k\,x^{n+k+2}\left(\frac{(\vartheta+\omega)x}{n+k+3}+\frac{(\vartheta-\omega)x}{n+k+2}\right),$$

and

$$B(p;\vartheta,\omega)=\frac{1}{\mu}\sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\frac{\binom{\vartheta-1}{k}\binom{\omega-1}{n}(-1)^k\,x^{n+k+2}\left(\frac{(\vartheta+\omega)x}{n+k+3}+\frac{(\vartheta-\omega)x}{n+k+2}\right)}{1-(1-x)^{\vartheta}(1+x)^{\omega}}.$$

2.7. Gini index

The Gini index (GI) is a vital tool for understanding and addressing economic inequality. It aids in the development of policies aimed at promoting social justice and equitable growth. Using Eqn. (2), the GI for the NTPUP model is formulated as follows: The GI for NTPUP distribution using Eqn. (2) is derived as:

$$G(\theta,\vartheta,\omega)=\frac{1}{\mu}\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}(-1)^n\,\theta^{n+m+1}\frac{\binom{\vartheta}{n}\binom{\omega}{m}-\binom{2\vartheta}{n}\binom{2\omega}{m}}{m+n+1}.$$

3. Estimation and simulation

The NTPUP model's parameters are estimated using the MLE. Consistent circumstances guarantee the presence and uniqueness of the MLEs because the log-likelihood function of the NTPUP model behaves well. This ensured the existence and uniqueness of the MLEs, resulting in accurate and consistent parameter estimation. A simulation analysis is undertaken to ensure the consistency and reliability of the MLEs.

3.1. The MLE

Applying the MLE technique, the unknown parameters of the NTPUP model (2) have estimated, given below:

$$\ln L(\vartheta,\omega)=(\vartheta-1)\sum_{i=1}^{n}\ln(1-x_i)+(\omega-1)\sum_{i=1}^{n}\ln(1+x_i)+\sum_{i=1}^{n}\ln(\vartheta-\omega+x_i(\vartheta+\omega)). \tag{12}$$

Taking derivatives of Eqn. (12) with respect to (w.r.t.) ϑ and putting $\frac{d\ln L(\vartheta,\omega)}{d\vartheta}=0$, we get

$$\sum_{i=1}^{n}\ln(1-x_i)+\sum_{i=1}^{n}\frac{1+x_i}{\vartheta-\omega+x_i(\vartheta+\omega)}=0.$$

Similarly taking derivatives of Eqn. (12) w.r.t. ω and putting $\frac{d\ln L(\vartheta,\omega)}{d\omega}=0$, we obtained

$$\sum_{i=1}^{n}\ln(1+x_i)+\sum_{i=1}^{n}\frac{x_i-1}{\vartheta-\omega+x_i(\vartheta+\omega)}=0.$$

Even though MLEs for fundamental statistical models may frequently be produced in closed form, other models may need to be optimized using numerical techniques. The numerical integration method is used in the R programming language to calculate the MLEs of the NTPUP model since there are not in a closed structure.

3.2. Simulation

A simulation analysis is executed in this part to assess how effectively the MLEs work for the NTPUP model's unknown parameters. The NTPUP mode's cdf is being utilized with the inverse transformation approach. Figs. 6-8 visually represent the analysis results. The steps for the simulation process are as follows:

1. **Step 1 (Initialization):** Choose starting values for the parameters ϑ and ω. The selected initial values are: $I=(\vartheta=2,\omega=1.5)$, $II=(\vartheta=2,\omega=1.7)$, $III=(\vartheta=2,\omega=-1.9)$.
2. **Step 2 (Sample generation):** Obtain n–sized random samples using the NTPUP model. Sample sizes ranging from $n=50$ to $n=200$ are taken into consideration.
3. **Step 3 (Estimation and accuracy measures):** Estimate the MLEs as well as the mean square errors (MSEs) and biases for the NTPUP distribution parameters. These metrics of accuracy are computed using the following formulas

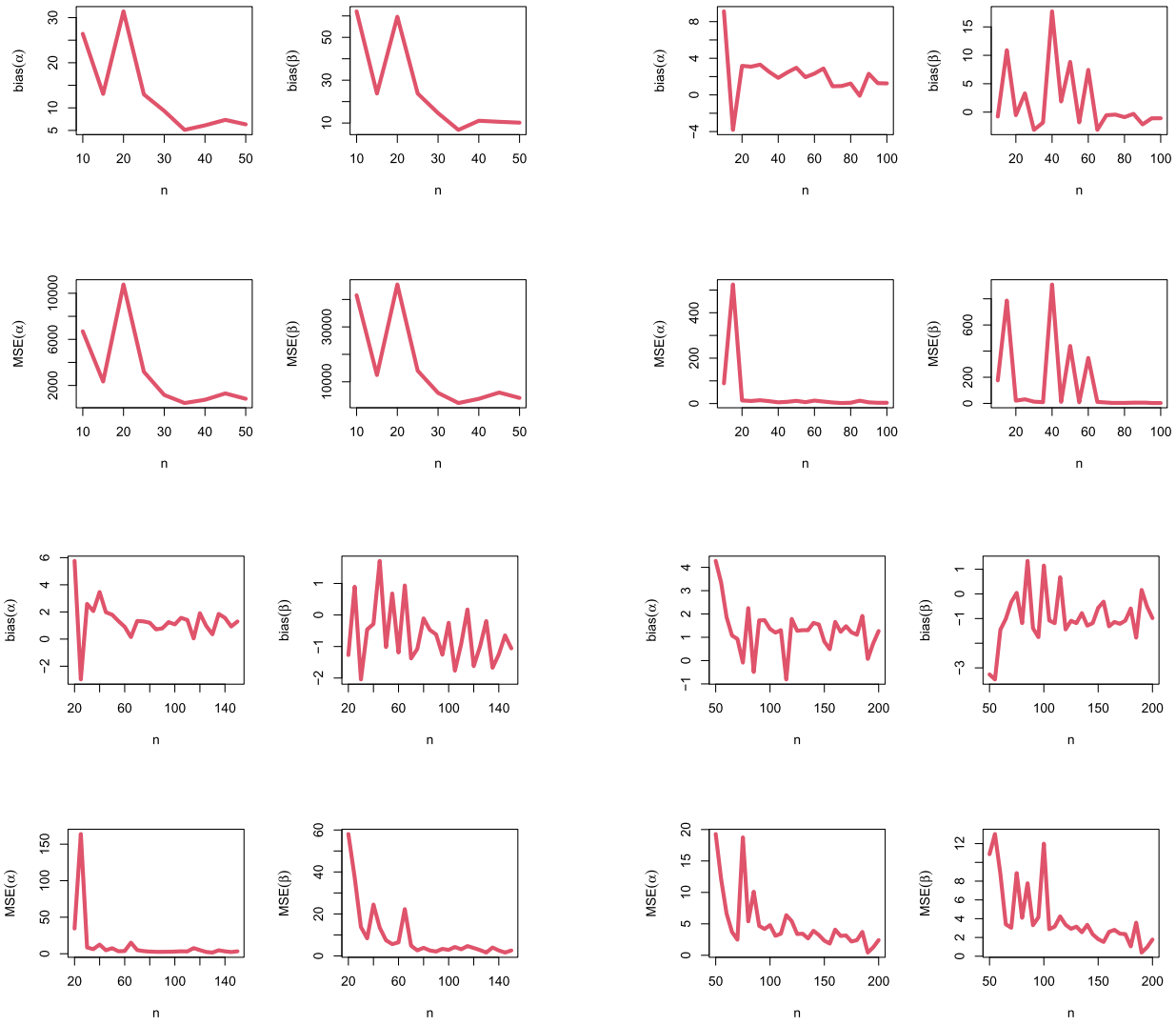


Fig. 6. Simulation of parameters $\vartheta = 2$, $\omega = 1.5$.

$$MSE_c = \frac{1}{8000}\sum_{i=1}^{N}(\hat{c}_i - c)^2,$$

$$Bias_c = \frac{1}{8000}\sum_{i=1}^{N}(\hat{c}_i - c),$$

respectively, for $c = (\vartheta, \omega)$. Where $\hat{c}_i$ reflects the MLE of c as estimated at the i^{th} iteration of the simulation.

4. **Step 4 (Repetition):** Steps 2 and 3 are repeated $N = 8000$ times to guarantee reliable results.

As portrayed in Figs. 6-8, the simulation outcomes are apparent that as the sample size n increases, both the apparent biases and MSEs decrease and approach zero for all parameters ϑ and ω. This highlights the reliability and consistency of the MLEs for the NTPUP model.

4. Applications

The modeling efficiency of the NTPUP model is thoroughly contrasted against several related models, including Burr XII (B12) [12], Kumaraswamy [13], Lomax [14], Beta [15], Burr X (BX) [16], LAII [17] and Power Burr Hatake [18]. To evaluate and contrast the models, we employed well-established statistical measures: namely CVM (Cramér-Von Mises), AD (Anderson-Darling), KS (Kolmogorov-Smirnov) statistics along with their corresponding p-values. Additionally, model selection criterion such as the Akaike

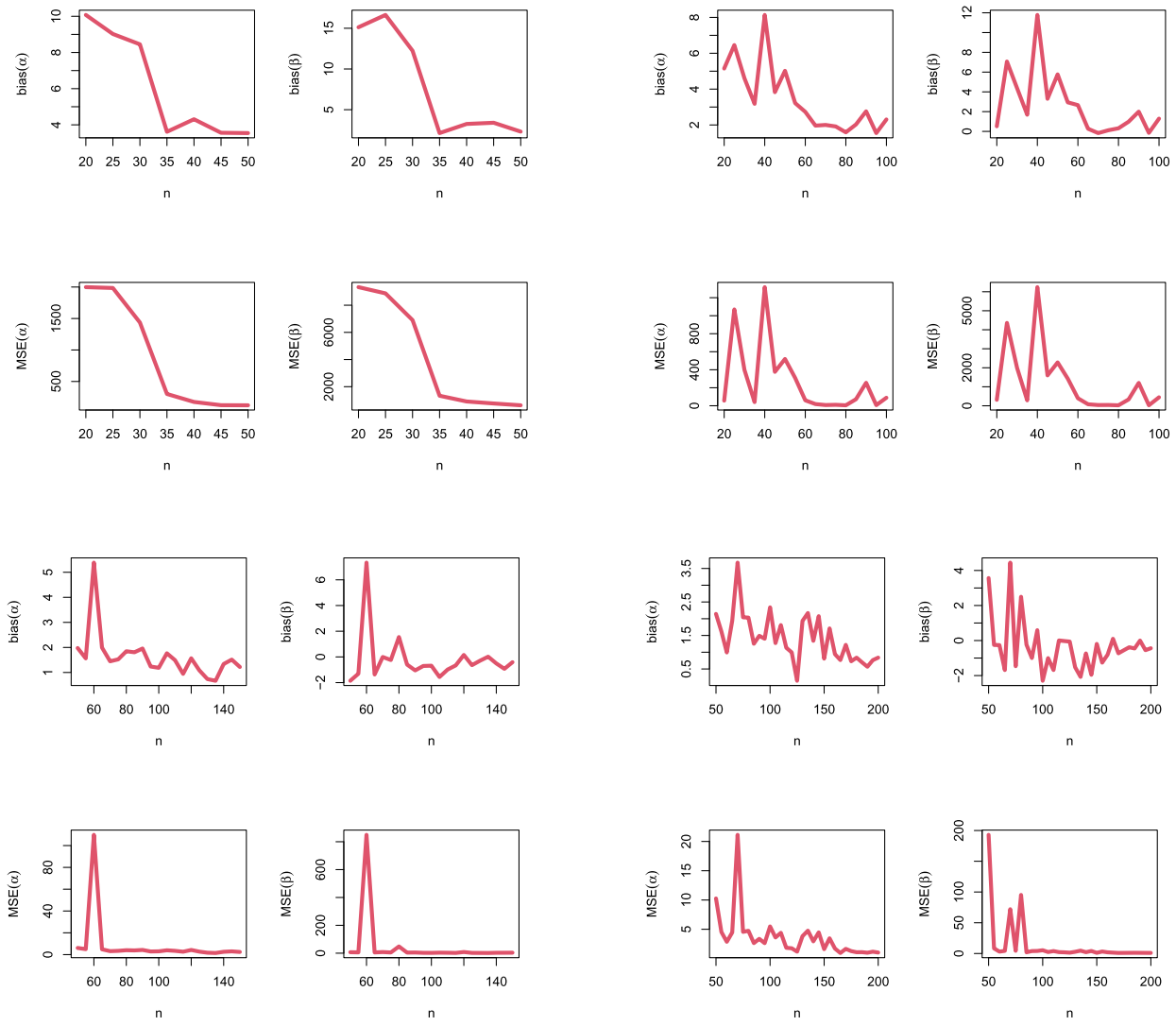


Fig. 7. Simulation of parameters $\vartheta = 2$, $\omega = 1.7$.

Information Criterion (AIC), Bayesian information criterion (BIC), corrected Akaike Information Criterion (CAIC) and Hannan-Quinn Information Criterion (HQIC) are also estimated, delivering an in-depth assessment of the model's efficacy.

- **Data set I:** The set of data I is a Covid-19 data spans 172 days, from March first to August twentieth, 2020, and pertains to Italy and Spain [9].
- **Data set II:** The set of data II is derived from [10] based on 63 observations of 1.5 cm glass fiber strengths that were first collected by UK National Physical Laboratory personnel.
- **Data set III:** The data set III is the entire milk production from the first birth of 107 Sindi-race cows [11].
- **Data set IV:** The fourth data set (bathtub shaped data), originally analyzed by Caramanis et al. [2], focuses on estimating unit capacity factors through a comparative study of two distinct algorithms: SC16 and P3.

The following visual and computational outcomes are depicted by utilizing all the aforementioned data sets 1, 2, 3, and 4, respectively.

- Tables 3, 5, 7 and 9 deliver the MLEs and their consequent standard errors (S.Es) for the NTPUP model together with several competitive models. Important information about the accuracy and consistency of the parameter estimates for each model under consideration can potentially be found in these tables.

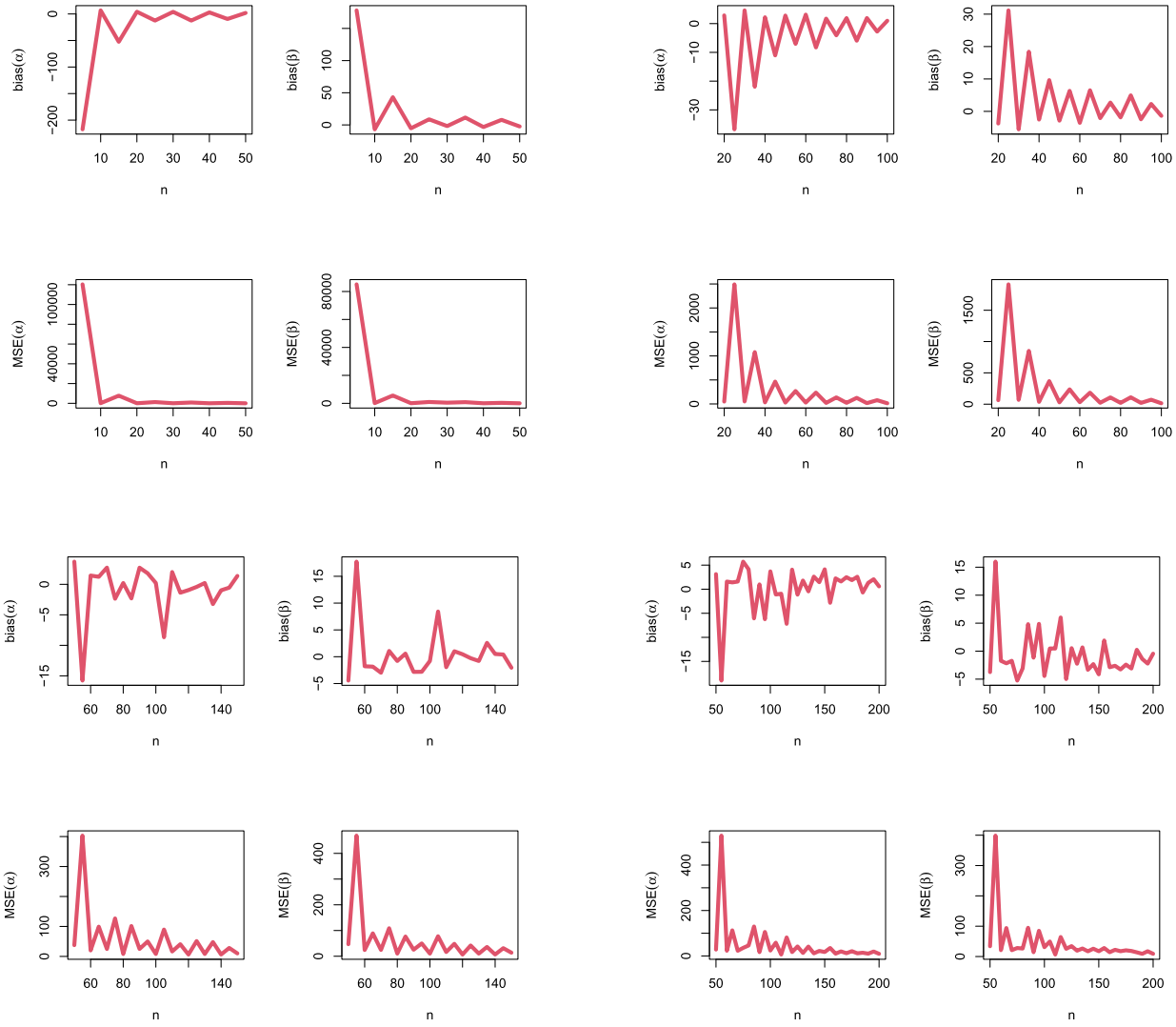


Fig. 8. Simulation of parameters $\vartheta = 2$, $\omega = -1.9$.

- The accuracy metrics estimated for all the data sets are displayed in the Tables 4, 6, 8 and 10, respectively. These tables disclosed that for every data set, the NTPUP model consistently performed superior than the competing models. A better fit to the data in indicated by lower values in the accuracy metrics, which demonstrate the higher performance.
- Variance covariance matrices for the NTPUP model have been shown in the matrices (13), (15), (17) and (19). For all the aforementioned data sets, these matrices offer a thorough analysis of the variability of the predicted parameters and the extent of their association. The correlation coefficients between pairs are shown in the correlation matrices (14), (16), (18) and (20), which provide further information on the relationship between the parameters in the NTPUP model.
- For the related data sets, the box plots and total time on test (TTT) plots are displayed in the Figs. 9, 11, 13 and 15, respectively. These figures provided more support for contrasting the results of the different models, which help to determine the distribution and variability of the data.
- For all the data sets, the pdf, cdf, probability-probability (PP) and quantile-quantile (QQ) plots are shown in Figs. 10, 12, 14 and 16, respectively. These graphics provide a through visual evaluation of the goodness-of-fit of the NTPUP model in comparison to the other models. The PP and QQ plots demonstrate the model's ability to match the data distribution's overall shape, while the pdf and cdf plots illustrate how effectively the model represents the overall data.

The aforementioned points disclosed that in comparison to the other models studies, the NTPUP model maintains its applicability and efficacy due to its superior performance for the designated data sets, as supported by both computational results and visual evidence.

Table 3
MLEs and SEs (in parentheses) for data set I.

Model	MLEs and SEs (in parentheses)	
NTPUP	23.252	19.782
(ϑ, ω)	(3.755)	(4.284)
LAII	6.913	-
(ϑ)	(0.526)	-
Lom	8.371	-
(k)	(0.636)	-
Ku	1.369	13.761
(ϑ, ω)	(0.094)	(2.528)
Beta	9.951	1.472
(ϑ, ω)	(1.124)	(0.144)
PBH	8.752	1.459
(ϑ, ω)	(1.442)	(0.088)
B12	1.474	16.416
(ϑ, ω)	(0.089)	(2.770)
BX	0.582	5.322
(ϑ, ω)	(0.053)	(0.310)

Table 4
The Value, AIC, CAIC, BIC, HQIC, W^*, A^*, K-S, (p-value) values for data set I.

Distribution	ℓ	AIC	CAIC	BIC	HQIC	W^*	A^*	K-S	p-value
NTPUP	-197.691	–391.382	-391.314	-385.075	-388.82	0.1829	1.0480	0.0688	0.3851
LAII	-186.5082	-371.02	-370.99	-367.86	-369.74	0.421	2.299	0.151	0.0008
Lom	-173.91	-345.83	-345.80	-342.67	-344.55	0.649	3.58	0.182,	0.0254
Ku	-195.2897	-386.57	-386.508	-380.27	-384.0209	0.3581	1.9625	0.398	0.254
Beta	-193.69	-383.03	-383.03	-376.799	-380.54	0.418	2.2824	0.10451	0.0457
PBH	-193.69	-383.38	-383.31	-377.07	-380.822	0.3920	2.48	0.09381	0.09643
B12	-194.177	-384.38	-384.28	-378.048	-381.79	0.3761	2.0599	0.0926	0.1028
BX	-194.58	-391.18	-391.11	-384.87	-388.627	0.2212	1.2315	0.08	0.22

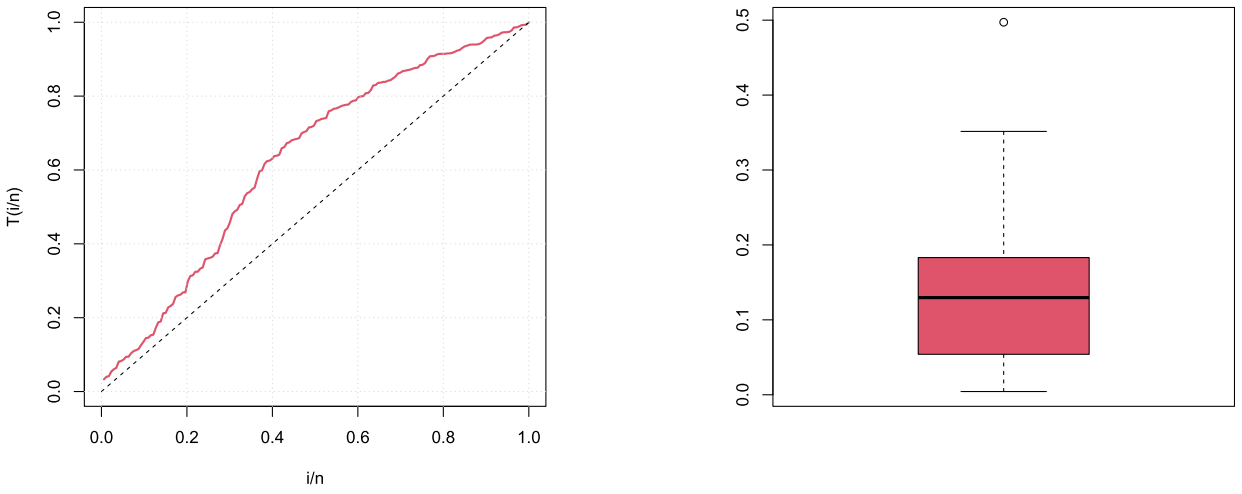


Fig. 9. TTT, and box plot for COVID 19 data in Italy EMW distribution.

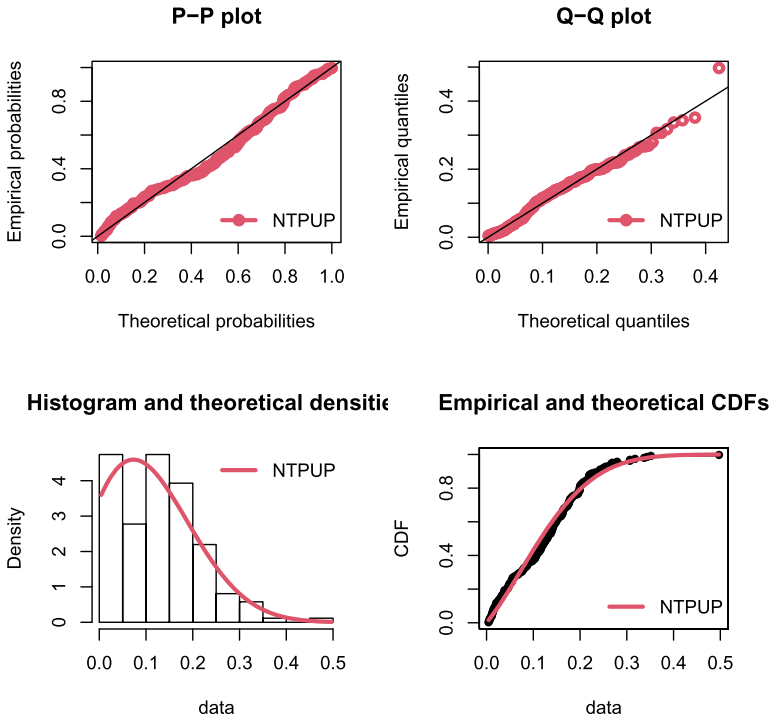


Fig. 10. Comparison plot of different pdfs, cdfs, PP and QQ for COVID 19 data in Italy.

Table 5
MLEs and SEs (in parentheses) for data set II.

Model	MLEs and SEs (in parentheses)	
NTPUP	1.360	-0.619
(ϑ, ω)	(0.585)	(1.090)
LAII	1.662385	-
(ϑ)	(0.585)	-
Lom	3.404	-
(k)	(0.622)	-
Ku	0.963	1.608
(ω, ϑ)	(0.202)	(0.413)
Beta	1.620	0.967
(ϑ, ω)	(0.410)	(0.224)
PBH	1.943	1.382
(ϑ, ω)	(0.423)	(0.199)
B12	1.325	3.404
(ϑ, ω)	(0.197)	(0.749)
BX	0.530	1.780
(ϑ, ω)	(0.114)	(0.256)

Var-Cov Matrix for Data set 1

$$\begin{pmatrix} 14.04220 & 15.87825 \\ 15.87825 & 18.26760 \end{pmatrix} \tag{13}$$

and correlation matrix for Data set 1

$$\begin{pmatrix} 1.0000000 & 0.9913896 \\ 0.9913896 & 1.0000000 \end{pmatrix} \tag{14}$$

Table 6
The Value, AIC, CAIC, BIC, HQIC, W^*, A^*, K-S(p-value) values for data set II.

Distribution	ℓ	AIC	CAIC	BIC	HQIC	W^*	A^*	K-S	p-value
NTPUP	-3.455	-2.910	-2.465	-0.107	-2.013	0.015	0.1249	0.051	1
LAII	-3.294	-4.588	-4.445	-3.187	-4.140	0.0184	0.156	0.074	0.992
Lom	2.061	6.122	6.265	7.523	6.570	0.068	0.422	0.149	0.466
Ku	-3.311	-2.622	-2.178	0.180	-1.726	0.0182	0.154	0.082	0.978
Beta	-3.305	-2.610	-2.166	0.1923	-1.714	0.0184	0.156	0.067	0.998
PBH	-1.510	0.978	1.423	3.781	1.88	0.041	0.27	0.093	0.93
B12	-1.752	0.497	0.9413	3.299	1.39	0.036	0.236	0.086	0.965
BX	-2.234	-0.4689	-0.0244	2.334	0.428	0.02443	0.177	0.066	0.998

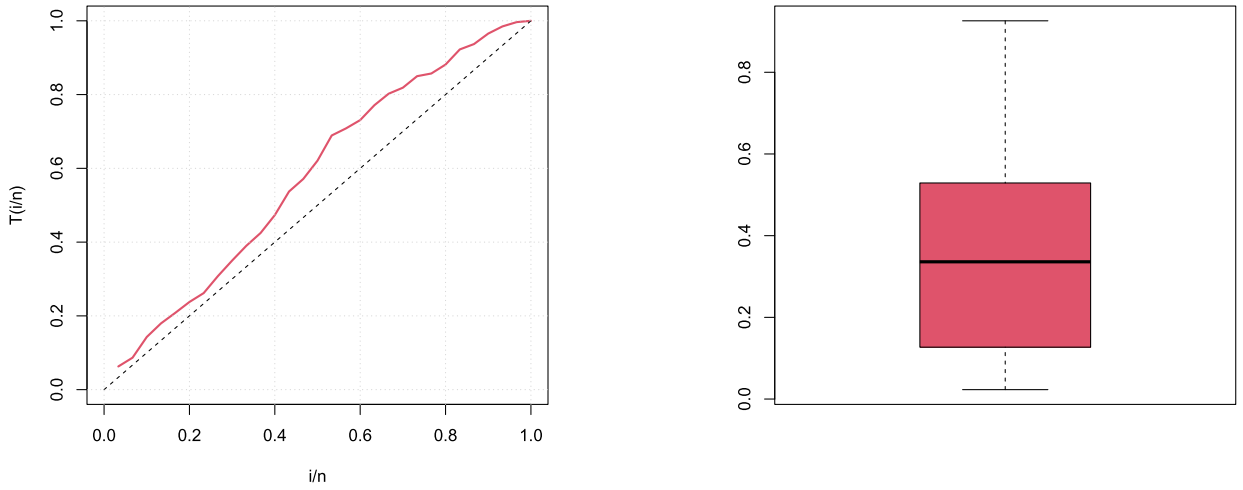


Fig. 11. TTT, and box plot for the NTPUP model.

Var-Cov Matrix for Data set 2

$$\begin{pmatrix} 0.3442864 & 0.549222 \\ 0.5492220 & 1.193075 \end{pmatrix} \tag{15}$$

and correlation matrix for Data set 2

$$\begin{pmatrix} 1.0000000 & 0.8569471 \\ 0.8569471 & 1.0000000 \end{pmatrix} \tag{16}$$

Var-Cov Matrix for Data set 3

$$\begin{pmatrix} 0.1040755 & 0.1211564 \\ 0.1211564 & 0.1544256 \end{pmatrix} \tag{17}$$

and correlation matrix for Data set 3

$$\begin{pmatrix} 1.0000000 & 0.9556799 \\ 0.9556799 & 1.0000000 \end{pmatrix} \tag{18}$$

Var-Cov Matrix for Data set 4

$$\begin{pmatrix} 0.02601207 & 0.04597288 \\ 0.04597288 & 0.24998186 \end{pmatrix} \tag{19}$$

and correlation matrix for Data set 4

$$\begin{pmatrix} 1.0000000 & 0.5701117 \\ 0.5701117 & 1.0000000 \end{pmatrix} \tag{20}$$

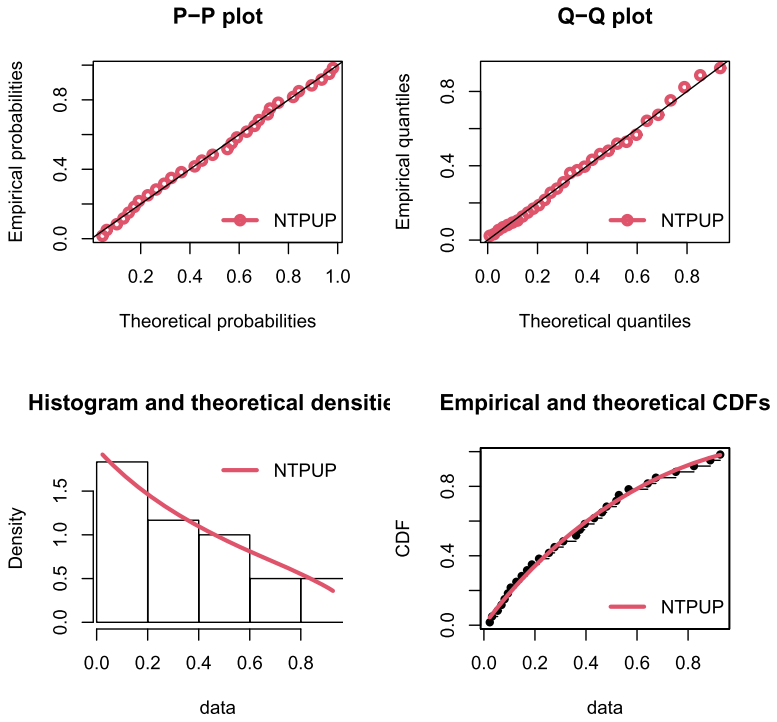


Fig. 12. Comparison plot of different pdfs, cdfs, PP and QQ for COVID-19 data.

Table 7
MLEs and SEs (in parentheses) for data set III.

Model	MLEs and SEs (in parentheses)	
NTPUP	0.5304	2.927
(ϑ, ω)	(0.1145)	(0.2563)
LAII	1.418118	-
(ϑ)	(0.137)	-
Lom	2.662	-
(k)	(0.257)	-
Beta	2.829	2.4125
(ϑ, ω)	(0.3744)	(0.3144)
PBH	3.0044	2.693
(ϑ, ω)	(0.4263)	(0.2106)
B12	2.601	5.382
(ϑ, ω)	(0.2099)	(0.7799)
BX	1.257	-2.116
(ϑ, ω)	(0.1605)	(0.122)

Table 8
The Value, AIC, BIC, W^*, A^*, for data set III.

Distribution	ℓ	AIC	CAIC	BIC	HQIC	W^*	A^*	K-S	P-value
NTPUP	-25.086	-46.17	-46.056	-40.83	-44.00	0.151	0.988	0.050624	1
LAII	-5.830	-9.661	-9.622	-6.988	-8.577	0.204	1.298	0.074038	0.9924
Lom	42.394	86.788	86.826	89.461	87.872	0.838	4.994	0.14981	0.4665
Beta	-23.777	-43.554	-43.4391	-38.209	-41.39	0.2083	1.326	0.066919	0.9979
PBH	-20.217	-36.435	-36.3195	-31.089	-34.27	0.256	1.680	0.093776	0.9323
B12	-21.347	-38.695	-38.58	-33.35	-36.528	0.232	1.523	0.086185	0.965
BX	-18.155	-32.311	-32.196	-26.965	-30.144	0.339	2.173	0.066022	0.9983

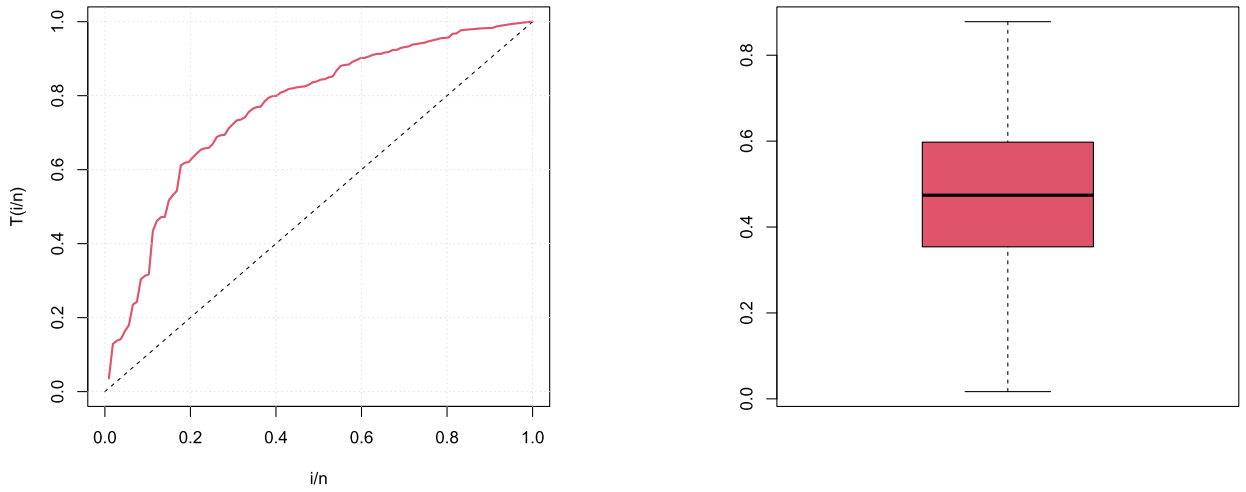


Fig. 13. TTT, and box plot for sindi race data for the NTPUP model.

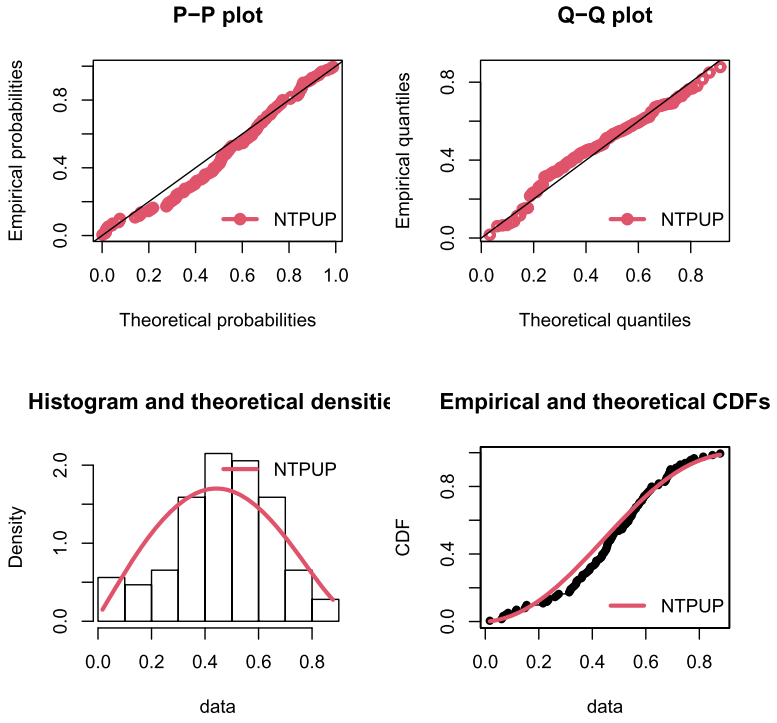


Fig. 14. Comparison plot of different pdfs, cdfs, PP and QQ for sindi race data.

5. Conclusion

This research has introduced a ground breaking two-parameter unit probability distribution named as the NTPUP distribution that distinguishes itself through its exceptional flexibility, offering a unique perspective on modeling bounded interval data. A comprehensive analysis has been carried out, among the various statistical properties of this novel distribution. An adoption of the maximum likelihood estimation method has proven to be a reliable approach for estimating the model parameters. The numerical simulation affirmed the consistency and trustworthiness of this estimation, enhancing the confidence in the practical usability of the proposed distribution. The real-world applicability of this distribution have been demonstrated through its successful implementation in analyzing four distinct real data-sets. This practical validation underscores its versatility and effectiveness in capturing the intricate characteristics of real-life data across a broad spectrum of fields. Specifically, wind speed and air density [19] may be the two primary areas of attention for applied practitioners as they tend to vary from one region to another. The flexibility of the proposed model might be an interesting area for them to explore further in future.

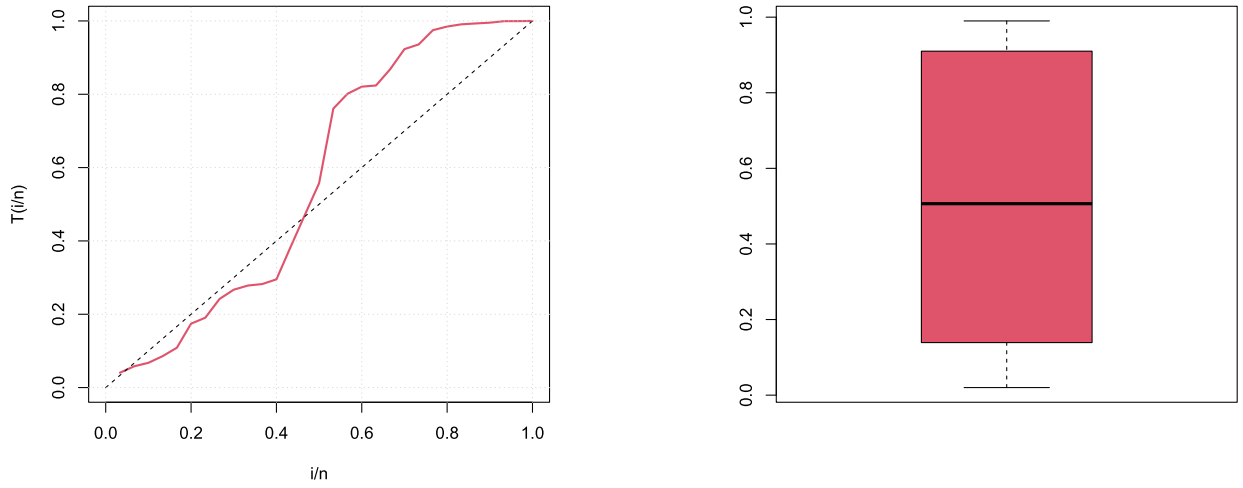


Fig. 15. TTT, and box plot for bathtub data for the NTPUP model.

Table 9
MLEs and SEs (in parentheses) for data set 4.

Model	MLEs and SEs (in parentheses)	
NTPUP	-0.340	1.051
(ϑ, ω)	(0.514)	(0.260)
Lom	2.860	-
(k)	(0.404)	-
Beta	0.604	0.858
(ϑ, ω)	(0.106)	(0.164)
B12	2.140	0.949
(ϑ, ω)	(0.3261)	(0.120)
BX	1.245	0.352
(ϑ, ω)	(0.162)	(0.0560)

Table 10
The Value, AIC, BIC, W, A, for data set 4.

Distribution	ℓ	AIC	CAIC	BIC	HQIC	W	A	K-S	p-value
NTPUP	-0.960	-2.08	-2.334	-5.903	-3.54	0.269	1.791	0.142	0.264
Lom	14.923	31.85	31.93	33.76	32.574	0.581	3.456	0.214	0.020
Beta	5.518	7.036	6.780	7.211	5.58	0.277	1.868	0.149	0.216
B12	10.68	25.357	25.61	29.18	26.81	0.498	3.018	0.194	0.048
BX	4.245	12.8	12.49	16.31	13.95	0.378	2.38	0.202	0.034

CRediT authorship contribution statement

Zawar Hussain: Visualization, Conceptualization. **Farrukh Jamal:** Project administration, Software, Resources. **Abdus Saboor:** Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Writing – review & editing. **Shakaiba Shafiq:** Formal analysis, Investigation, Validation, Writing – original draft. **Arshid Khan:** Data curation, Formal analysis, Methodology. **Shahida Perveen:** Methodology, Writing – original draft. **Fuad A. Awwad:** Funding acquisition. **Emad A.A. Ismail:** Funding acquisition. **Musharraf Ali:** Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Fuad A. Awwad reports a relationship with King Saud University College of Business Administration that includes: employment and funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

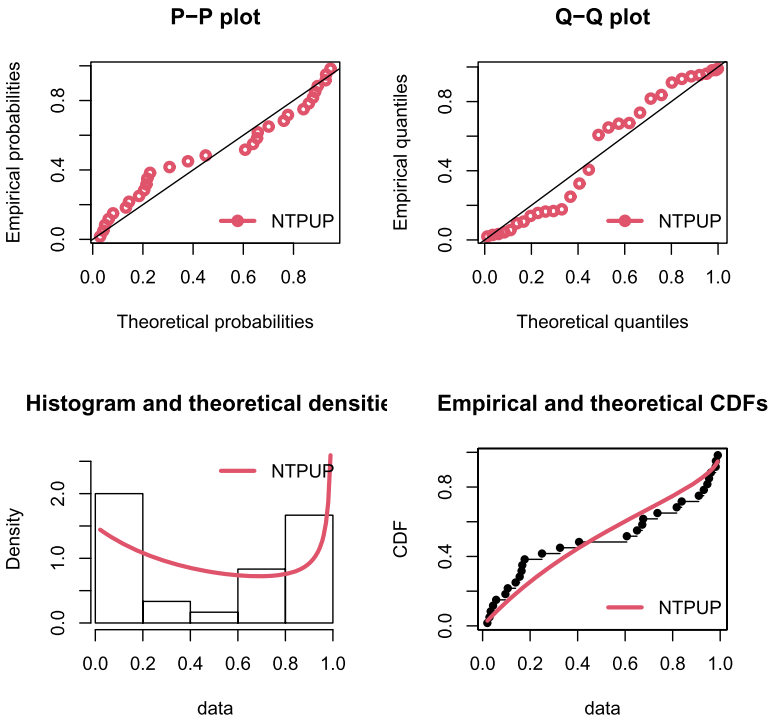


Fig. 16. Comparison plot of different pdf, cdf, PP and QQ for bathtub data.

Data availability

The data presented in this study are available with references.

Acknowledgement

Researchers Supporting Project Number (RSPD2024R1060), King Saud University, Riyadh, Saudi Arabia

References

[1] M.H. Tahir, S. Nadarajah, Parameter induction in continuous univariate distributions: well-established G families, An. Acad. Bras. Ciênc. 87 (2015) 539–568.
[2] M. Caramanis, J. Stremel, W. Fleck, S. Daniel, Probabilistic production costing: an investigation of alternative algorithms, Int. J. Electr. Power Energy Syst. 5 (2) (1983) 75–86.
[3] E. Gómez-Déniz, M.A. Sordo, E. Calderín-Ojeda, The Log–Lindley distribution as an alternative to the beta regression model with applications in insurance, Insur. Math. Econ. 54 (2014) 49–57.
[4] J. Mazucheli, A.F.B. Menezes, S. Dey, Improved maximum-likelihood estimators for the parameters of the unit-gamma distribution, Commun. Stat., Theory Methods 47 (15) (2018) 3767–3778.
[5] A. Grassia, On a family of distributions with argument between 0 and 1 obtained by transformation of the gamma and derived compound distributions, Aust. J. Stat. 19 (2) (1977) 108–114.
[6] G.M. Cordeiro, M. de Castro, A new family of generalized distributions, J. Stat. Comput. Simul. 81 (7) (2011) 883–898.
[7] C.W. Topp, F.C. Leone, A family of J-shaped frequency functions, J. Am. Stat. Assoc. 50 (269) (1955) 209–219.
[8] R. Kieschnick, B.D. McCullough, Regression analysis of variates observed on (0, 1): percentages, proportions and fractions, Stat. Model. 3 (3) (2003) 193–213.
[9] H.M. Alshanbari, O.H. Odhah, E.M. Almetwally, E. Hussam, M. Kilai, A.A.H. El-Bagoury, Novel type I half logistic Burr-Weibull distribution: application to COVID-19 data, Comput. Math. Methods Med. 2022 (2022).
[10] H. Elgohari, H. Yousof, New extension of Weibull distribution: copula, mathematical properties and data modeling, Stat. Optim. Inf. Comput. 8 (4) (2020) 972–993.
[11] M. Muhammad, H.M. Alshanbari, A.R. Alanzi, L. Liu, W. Sami, C. Chesneau, F. Jamal, A new generator of probability models: the exponentiated sine-G family for lifetime studies, Entropy 23 (11) (2021) 1394.
[12] Irving W. Burr, Cumulative frequency functions, Ann. Math. Stat. 13 (2) (1942) 215–232.
[13] M.C. Jones, Kumaraswamy's distribution: a beta-type distribution with some tractability advantages, Stat. Methodol. 6 (1) (2009) 70–81.
[14] Kenneth S. Lomax, Business failures: another example of the analysis of failure data, J. Am. Stat. Assoc. (1954) 847–852.
[15] S. Ferrari, F. Cribari-Neto, Beta regression for modelling rates and proportions, J. Appl. Stat. 31 (7) (2004) 799–815.
[16] Y. Ishihara, H. Kato, K. Nakano, M. Onizuka, Y. Sasaki, Toward BX-Based Architecture for Controlling and Sharing Distributed Data, IEEE, 2019, pp. 1–5.
[17] Douglas L. Kane, William S. Reeburgh, Introduction to special section: land-air-ice interactions (LAII) flux study, J. Geophys. Res. 103 (D22) (1998) 28913–28915.
[18] S. Gamsjaeger, R. Mendelsohn, A. Boskey, S. Gourion-Arsiquaud, K. Klaushofer, E.P. Paschalis, Vibrational spectroscopic imaging for the evaluation of matrix and mineral chemistry, Curr. Osteoporos. Rep. 12 (2014) 454–464.
[19] J. Liu, Z. Yan, A circular-linear probabilistic model based on nonparametric copula with applications to directional wind energy assessment, Entropy 26 (2024) 487, https://doi.org/10.3390/e26060487.