Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 27 (2024), No. 6, pp. 1199–1220
DOI : 10.47974/JSMS-1239

New family of distributions based on the unit modified Burr III distribution

Abraham Bamfo Boakye *
Department of Statistics and Actuarial Science
School of Mathematical Sciences
C. K. Tedam University of Technology and Applied Sciences
Navrongo
Ghana

and

Department of Mathematical Sciences
Faculty of Science
Valley View University
Oyibi
Accra
Ghana

Suleman Nasiru †
Sampson Wiredu §
Department of Statistics and Actuarial Science
School of Mathematical Sciences
C. K. Tedam University of Technology and Applied Sciences
Navrongo
Ghana

Abstract

In recent times, there has been significant attention to the improvements to classical probability distributions in the literature. Because of this attention, many probability generators or family distribution functions have emerged. These generators can control the skewness, kurtosis, and tail weight of datasets effectively. The unit modified Burr III

* *E-mail:* **abbbamfo@gmail.com; (Corresponding Author)**
bamfo.boakye@vvu.edu.gh
† *E-mail:* **sulemanstat@gmail.com;**
snasiru@cktutas.edu.gh
§ *E-mail:* **swiredu@cktutas.edu.gh**

family of distributions—an entirely new family of distributions—has been developed. The mathematical and statistical characteristics of this new family have been derived. Two special distributions, the unit modified Burr III Weibull (UMBIIIW) and unit modified Burr III Chen (UMBIIIC) distributions, have been developed from this family. These distributions' parameters have been estimated using the maximum likelihood estimation (MLE) technique. A simulation method based on Monte-Carlo was employed to investigate the behavior of these estimators. Next, a real-life dataset was used to demonstrate the UMBIIIC distribution's applicability. Finally, an entirely new regression equation, called the log unit modified Burr III Weibull (LUMBIIIW) regression equation, was proposed, and real-life data were used to demonstrate how applicable it is. With the data used in the application illustration, the new regression equation offers the best fit among its competitive models.

Subject Classification: *(2010) 60E05, 62F10, 62E15.*

Keywords: *Unit modified Burr III, Location-scale, Mean residual life, Skewness, Kurtosis.*

1. Introduction

For a long period, probability distribution functions have been extensively used to model data sets in scientific, technological, economic, and business environments. The exponential distribution, for example, is utilized in seismic research, the telephone industry, equipment lifetime, and insurance claims. Other distributions also have their suitability in analyzing datasets of varying properties.

Most classical probability distribution functions are parametric, and as such, they cannot fit data as expected [17]. This has necessitated that probability distribution analysts improve the flexibility of these existing distributions by adding more parameters to them or developing generators or family of distributions to control the characteristics such as skewness, kurtosis, and tail-weight of datasets. Among these family of distributions are [3], [4], [5], [8], [10], [17], [21], [23], and [29].

A useful distribution called the unit Burr III (UBIII) distribution was developed by [19]. They developed this distribution by converting the Burr III distribution, which, is widely known to be flexible and adaptable into a unit form. [16] Proposed the unit modified Burr III (UMBIII) distribution that seems to be an improved form of the UBIII distribution.

This paper focuses on developing a generator, called the unit modified Burr III (UMBIII) family of distributions, a generator of the UMBIII distribution, to enhance the flexibility of some of the well-known distributions. This generator can produce distributions that are both light and heavy-tailed with the ability to control skewness, kurtosis, and variability of datasets. Furthermore, a location-scale regression (LSR) model is proposed.

The proposal of this generator is motivated by the fact that no authority has proposed it, and it is believed that this newly generated family will be able to better control the characteristics of some well-known and some classical distributions and to contribute to the existing literature.

There are ten sections in this paper. The unit modified Burr III (UMBIII) family is introduced in the second section. In the third section, the new family's PDF and its mixture representation (MR) are explained in detail. The statistical properties of the generated family are discussed in the fourth Section. Estimation of the parameters for the UMBIII family is the primary focus of the fifth section. Unique distributions within the UMBIII family are analyzed in the sixth section. Monte-Carlo Simulations are performed in the seventh Section. A Real-life dataset is used in the eighth section to establish the applicability of the newly created UMBIIIC distribution. Section 9 presents a suggested regression equation and the last Section presents the study's conclusions.

2. Unit modified Burr III family of distributions

Consider the random variable T that follows the UMBIII distribution with PDF $r(t) = \alpha\beta t^{-2}\ (t^{-1}-1)^{\beta-1}\ [1+\gamma\ (t^{-1}-1)^{\beta}]^{-(1+\alpha/\gamma)}$, $0<t<1$, where $\alpha, \beta, \gamma > 0$. If the random variable X belongs to the UMBIII family, then the cumulative distribution function (CDF) is

$$F(x) = \int_0^{G(x;\eta)} r(t)dt = \left[1+\gamma\left(\frac{G(x;\eta)}{1-G(x;\eta)}\right)^{-\beta}\right]^{-\alpha/\gamma} ;x \in \mathbb{R}, \tag{1}$$

with the shape parameters $\alpha>0, \beta>0$ and $\gamma>0$, a scale parameter. η is a $p\times 1$ vector made up of parameters. The associated PDF of the UMBIII family is given as

$$f(x) = \alpha\beta g(x;\eta)G(x;\eta)^{-2}(G(x;\eta)^{-1}-1)^{\beta-1}[1+\gamma(G(x;\eta)^{-1}-1)^{\beta}]^{-(1+\alpha/\gamma)}, x \in \mathbb{R}, \tag{2}$$

where $g(x,\eta)$ and $G(x,\eta)$ are the baseline PDF and CDF respectively. The family's failure rate function (HRF) is given as

$$h(x) = \frac{\alpha\beta g(x;\eta)G(x;\eta)^{-2}(G(x;\eta)^{-1}-1)^{\beta-1}[1+\gamma(G(x;\eta)^{-1}-1)^{\beta}]^{-(1+\alpha/\gamma)}}{1-[1+\gamma(G(x;\eta)^{-1}-1)^{\beta}]^{-\alpha/\gamma}}, x \in \mathbb{R}. \tag{3}$$

3. UMBIII family mixture representation

The UMBIII family mixture representation (MR) is crucial for determining its structural properties. Therefore, the distribution and density functions of the UMBIII family are presented in their expanded form here.

Lemma 1: *The CDF for the UMBIII family MR is expressed as follows:*

$$F(x)=\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\omega_{ij}G(x;\eta)^{-\beta i+j}, \tag{4}$$

where $\omega_{ij}=\frac{(-1)^{i+j}\gamma^i}{i!}\left(\frac{\alpha}{\gamma}\right)^{(i)}\binom{\beta i}{j}$.

Proof: Using the Taylor series expansion

$$A^{-d}=\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}d^{(i)}(A-1)^i,$$

where $d^{(i)}=d(d+1)(d+2)...(d+i-1)$ is the rising factorial. Hence,

$$\left[1+\gamma\left(\frac{G(x;\eta)}{1-G(x;\eta)}\right)^{-\beta}\right]^{-\alpha/\gamma}=\sum_{i=0}^{\infty}\frac{(-1)^i\gamma^i}{i!}\left(\frac{\alpha}{\gamma}\right)^{(i)}G(x;\eta)^{-\beta i}[1-G(x;\eta)]^{\beta i}.$$

Further applying the binomial expansion yields

$$\left[1+\gamma\left(\frac{G(x;\eta)}{1-G(x;\eta)}\right)^{-\beta}\right]^{-\alpha/\gamma}=\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{(-1)^{i+j}\gamma^i}{i!}\left(\frac{\alpha}{\gamma}\right)^{(i)}\binom{\beta i}{j}G(x;\eta)^{-\beta i+j}.$$

The proof is now complete.

Lemma 2: *The PDF of the UMBIII family is obtained in expanded form as*

$$f(x)=\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\omega_{ij}(j-\beta i)g(x;\eta)G(x;\eta)^{-\beta i+j-1}. \tag{5}$$

Proof: Differentiating the expanded CDF of the UMBIII family makes the proof complete.

4. Statistical properties

We here present the derivations of the UMBIII family's statistical properties.

4.1 *Quantile function*

Quantile functions (QFs) are useful for generating random numbers for simulation and determining the influence of parameters on a distribution's skewness and kurtosis.

Proposition 1: Assume that U is a random variable belonging to the standard uniform distribution with $u \in [0,1]$. The UMBIII family's quantile function (QF) is expressed as

$$x_u = G^{-1}\{[1+\gamma^{-1/\beta}(u^{-\gamma/\alpha}-1)^{1/\beta}]^{-1}\}, u \in [0,1]. \tag{6}$$

Where $G(\cdot)$ denotes the baseline distribution's QF.

Proof: Equating the CDF of the UMBIII family to u,

$$\left[1+\gamma\left(\frac{G(x;\eta)}{1-G(x;\eta)}\right)^{-\beta}\right]^{-\alpha/\gamma} = u.$$

Making $G(x)$ the subject yields

$$G(x) = \left(\tfrac{u^{-\gamma/\alpha}-1}{\gamma}\right)^{-1/\beta}\left[1+\left(\tfrac{u^{-\gamma/\alpha}-1}{\gamma}\right)^{-1/\beta}\right]^{-1}$$

Taking the G inverse of both sides completes the proof.

4.2 *Moments, moment generating functions, and incomplete moments*

The *rth* non-central moments, moment generating functions and incomplete moments (IM) of the UMBIII family are derived here.

4.2.1 Moments

Statistical moments are essential tools for examining the properties of statistical distributions. Moments offer an important understanding of the distributions' center, dispersion, and shape measurements [13].

Proposition 2: The r^{th} non-central moment (NCM), μ'_r, of the UMBIII family is

$$\mu'_r = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\omega_{ij}(j-\beta i)\int_{\mathbb{R}} x^r g(x;\boldsymbol{\eta})G(x;\boldsymbol{\eta})^{-\beta i+j-1}dx, \quad r=1,2,3,... \tag{7}$$

Proof: The NCM, is defined as

$$\mu'_r = E(X^r) = \int_{\mathbb{R}} x^r f(x)dx \tag{8}$$

Substitution of (2) for $f(x)$ in (8) completes the proof.

4.2.2 Moment generating functions

The properties of distribution functions are largely determined by moment generating functions (MGFs). By using MGFs, one can ascertain various moments of a distribution and, as a result, gain additional knowledge about it [13].

Proposition 3: The expression that represents the MGFs, $M_x(t)$, of the UMBIII family is as follows:

$$M_X(t)=\sum_{r=0}^{\infty}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{w_{ij}(j-\beta i)t^r}{r!}\int_{\mathbb{R}} x^r g(x;\boldsymbol{\eta})G(x;\boldsymbol{\eta})^{-\beta i+j-1}dx \ , r=1,2,3,... \tag{9}$$

Proof: It is known from the Taylor series expansion that $e^{tx}=\sum_{r=0}^{\infty}\frac{t^r x^r}{r!}$ and $M_X(t)=\sum_{r=0}^{\infty}\frac{t^r}{r!}\mu_r'$. The proof is now complete by substituting (7) for μ_r'.

4.2.3 Incomplete moments

Measures of inequality, the average remaining lifespan of a random variable's distribution, and deviations from the mean and median all hinge on IM.

Proposition 4: The r^{th} IM, $\varphi_r(t)$, for the UMBIII family is defined as

$$\varphi_r(t)=\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\omega_{ij}(j-\beta i)\int_{-\infty}^{t} x^r g(x;\eta)G(x;\eta)^{-\beta i+j-1}dx, \quad r=1,2,3,... \tag{10}$$

Proof: Define $\varphi_r(t)$ as $\varphi_r(t)=\int_{-\infty}^{t} x^r dF(x)$. Substituting (5) for $dF(x)$ in this definition completes the proof.

4.3 *Mean residual lifetime*

When a component continues to survive until time t, the mean residual life in the context of life testing is the expected remaining lifetime as a function of t [15].

Proposition 5: The MRL, $m_r(t)$, of the UMBIII family, is expressed as

$$m_r(t)=\frac{1}{1-F(t)}\left[\mu-\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\omega_{ij}(j-\beta i)\int_{-\infty}^{t} xg(x;\eta)G(x;\eta)^{-\beta i+j-1}dx\right]-t\ . \tag{11}$$

Proof: $m_r(t)$ of a random variable X is defined as

$$m_r(t) = \frac{1}{1-F(t)}\left[\mu - \int_{-\infty}^{t} xf(x)dx\right] - t\ . \tag{12}$$

where t it time.

Substituting (5) for $f(x)$ in (12) yields the MRL of the UMBIII family.

4.4 *Stress-strength reliability*

In life testing for a component, the number $R = P(Y < X)$, where X and Y are two independent random variables with X denoting the component's stress and Y denoting its strength, to the extent that the component fails if the stress exceeds the strength defines stress-strength reliability (SSR) [27].

Proposition 6: Given that $X_1 \sim UMBIII\ (x;\ \alpha,\ \beta,\ \gamma)$ and $X_2 \sim UMBIII\ (x;\ \alpha,\ \beta,\ \gamma)$, then the statistical definition of the SSR of the UMBIII family is

$$R = \alpha\beta\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} C_{ijk}\int_{\mathbb{R}} g(x;\eta)(G(x;\eta)^{-1}-2)^{k}\, G(x;\eta)^{-\beta i+j-2}dx\,, \tag{13}$$

where $C_{ijk} = \frac{(-1)^{i+j+k}}{i!k!}\gamma^{i}\left(1+\frac{2\alpha}{\gamma}\right)^{(i)}\binom{\beta i}{j}(\beta-1)^{k}$.

Proof: Suppose $X_1 \sim UMBIII\ (x;\alpha,\ \beta,\ \gamma)$ and $X_2 \sim UMBIII\ (x;\alpha,\ \beta,\ \gamma)$, then the SSR is defined statistically as

$$R = P(X_2 < X_1) = \int_{\mathbb{R}} f(x)F(x)dx. \tag{14}$$

Substituting the $F(x)$ and $f(x)$ in (1) and (2) respectively for R results in

$$R = \alpha\beta\int_{\mathbb{R}} g(x;\eta)G(x;\eta)^{-2}(G(x;\eta)^{-1}-1)^{\beta-1}[1+\gamma(G(x;\eta)^{-1}-1)^{\beta}]^{-\left(1+2\alpha/\gamma\right)}dx.$$

Applying algebraic manipulation through the use of the binomial series expansion $(1-z)^{s} = \sum_{j=0}^{\infty}(-1)^{j}\binom{s}{j}z^{j}$, $|z|\le 1$, and the Taylor series expansion $A^{-d} = \sum_{i=0}^{\infty}\frac{(-1)^{i}}{i!}\, d^{(i)}\,(A-1)^{i}$, where $d^{(i)} = d(d+1)\ (d+2)\ \ldots\ (d+i-1)$, $[1+\gamma\ (G(x;\eta)^{-1}-1)^{\beta}]^{-\left(1+\frac{2\alpha}{\gamma}\right)}$ in R can be written as $\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{(-1)^{i+j}\gamma^{i}}{i!}\left(1+\frac{2\alpha}{\gamma}\right)^{(i)}\binom{\beta i}{j}$ $G(x;\boldsymbol{\eta})^{-\beta i+j}$.

Binomial expansion of the expression $(G(x;\eta)^{-1}-1)^{\beta-1}$ in R gives

$$\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}(\beta-1)^k(G(x;\eta)^{-1}-2)^k.$$

The expression for R is consequently given by

$$R=\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{(-1)^{i+j+k}}{i!k!}\gamma^i\binom{\beta i}{j}(\beta-1)^k\int_{\mathbb{R}}g(x;\eta)(G(x;\eta)^{-1}-2)^kG(x;\eta)^{-\beta i+j-2}dx.$$

Hence, the proof is complete.

5. Maximum likelihood estimation for UMBIII family

In this section, the role of MLE is to estimate the unknown parameters of the UMBIII family.

Suppose $x_1,x_2,...,x_n$ is a random sample whose size is a positive integer n. Let the sample be distributed according to the PDF of the UMBIII family with parameters $\alpha,\ \beta,\ \gamma,$ and η. Given the family, the loglikelihood function is

$$\begin{aligned}\ell &= n\log(\alpha\beta)+\sum_{i=1}^{n}\log g(x_i;\eta)-2\sum_{i=1}^{n}\log G(x_i;\eta)+(\beta-1)\sum_{i=1}^{n}\log(1-G(x_i;\eta))\\ &\quad-(\beta-1)\sum_{i=1}^{n}\log G(x_i;\eta)-(1+\alpha/\gamma)\sum_{i=1}^{n}\log\left(1+\gamma\left(\tfrac{1-G(x_i;\eta)}{G(x_i;\eta)}\right)^{\beta}\right). \qquad (15)\end{aligned}$$

We obtain the MLE estimates for the UMBIII family when we maximize the loglikelihood function in (15). We use the package in [6] to maximize the loglikelihood function in (15).

6. Special distributions

The UMBIII family can be used to extend several probability distribution functions to improve their adaptability and flexibility in various applications. In this work, two well-known distributions, the Weibull [28] and Chen [2] and [7] have been extended. The UMBIIIW and the UMBIIIC distributions are the extensions of the Weibull and Chen distributions respectively.

6.1 *UMBIIIW Distribution*

Substitution of the Weibull CDF for $G(x)$ in (1) produces the UMBIIIW distribution as shown in (16). The PDF of the UMBIIIW distribution (17) is obtained by differentiating (16) with respect to x. The CDF of the Weibull distribution $G(x) = 1-e^{-\lambda x^k}$ [28] is the baseline distribution here. The CDF and PDF of the UMBIIIW distribution are respectively

$$F(x)=[1+\gamma((1-e^{-\lambda x^k})^{-1}-1)^{\beta}]^{-\alpha/\gamma},\ x>0,\lambda>0,k>0,\alpha>0,\beta>0,\gamma>0, \tag{16}$$

and

$$f(x)=\alpha\lambda\beta k x^{k-1}e^{-\lambda x^k}(1-e^{-\lambda x^k})^{-2}((1-e^{-\lambda x^k})^{-1}-1)^{\beta-1}$$
$$[1+\gamma((1-e^{-\lambda x^k})^{-1}-1)^{\beta}]^{-(1+\alpha/\gamma)},\ x>0. \tag{17}$$

The corresponding HRF is

$$h(x)=\frac{\alpha\lambda\beta k x^{k-1}e^{-\lambda x^k}(1-e^{-\lambda x^k})^{-2}((1-e^{-\lambda x^k})^{-1}-1)^{\beta-1}[1+\gamma((1-e^{-\lambda x^k})^{-1}-1)^{\beta}]^{-(1+\alpha/\gamma)}}{1-[1+\gamma((1-e^{-\lambda x^k})^{-1}-1)^{\beta}]^{-\alpha/\gamma}},\ x>0. \tag{18}$$

The UMBIIIW distribution's density function is displayed in Figure 1, where interesting features such as decreasing, and right-skew curves are evident. Additionally, Figure 1's hazard rate graph shows a variety of shapes, such as the bathtubs and upside-down bathtubs, in addition to increasing and decreasing failure rates.

The QF for the UMBIIIW is the inverse of F in (16) given by

$$x_u=[-\lambda^{-1}\log(1-((((u^{-\gamma/\alpha}-1)^{\frac{1}{\beta}}\gamma^{-\frac{1}{\beta}})+1))^{-1})]^{\frac{1}{k}},0\le u\le 1. \tag{19}$$

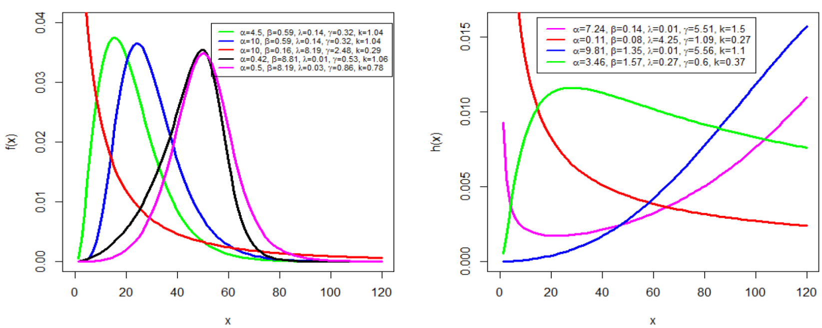

Figure 1

Graphs of PDFs (left) and HRFs (right) of the UMBIIIW distribution

6.2 *UMBIIIC distribution*

The baseline distribution here is Chen [7]. The CDF and PDF are respectively $G(x) = 1 - e^{k(1-e^{x^{\lambda}})}$ and

$$g(x) = \lambda k x^{\lambda-1} e^{(k(1-e^{x^{\lambda}})+x^{\lambda})},\ x > 0,\ \lambda > 0,\ k > 0.$$

The UMBIIIC distribution is

$$F(x) = [1 + \gamma((1 - e^{k(1-e^{x^{\lambda}})})^{-1} - 1)^{\beta}]^{-\alpha/\gamma}, x > 0, k > 0, \lambda > 0, \beta > 0, \alpha > 0, \gamma > 0. \tag{20}$$

The UMBIIIC distribution's corresponding density (PDF) and hazard functions are, respectively

$$f(x) = \alpha\beta\lambda k x^{\lambda-1} e^{(k(1-e^{x^{\lambda}})+x^{\lambda})}(1 - e^{k(1-e^{x^{\lambda}})})^{-2}((1 - e^{k(1-e^{x^{\lambda}})})^{-1} - 1)^{\beta-1}$$
$$[1 + \gamma((1 - e^{k(1-e^{x^{\lambda}})})^{-1} - 1)^{\beta}]^{-(1+\alpha/\gamma)},\ x > 0. \tag{21}$$

and

$$h(x) = \frac{\alpha\beta\lambda k x^{\lambda-1} e^{(k(1-e^{x^{\lambda}})+x^{\lambda})}(1-e^{k(1-e^{x^{\lambda}})})^{-2}((1-e^{k(1-e^{x^{\lambda}})})^{-1}-1)^{\beta-1}[1+\gamma((1-e^{k(1-e^{x^{\lambda}})})^{-1}-1)^{\beta}]^{-(1+\alpha/\gamma)}}{1-[1+\gamma((1-e^{k(1-e^{x^{\lambda}})})^{-1}-1)^{\beta}]^{-\alpha/\gamma}}, x > 0. \tag{22}$$

Interesting features such as decreasing density patterns, right-skew, and symmetry, can be seen in Figure 2's density graph of the UMBIIIC distribution. In addition, several kinds of shapes, including a bathtub-like pattern and an inverted bathtub shape, as well as increasing and decreasing failure rates, are shown in the graph of the HRFs.

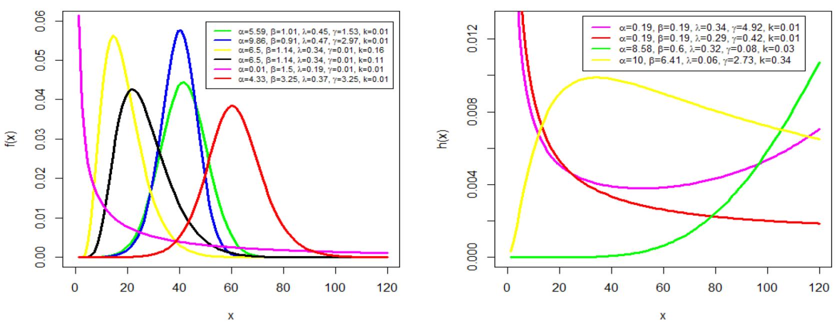

Figure 2

Graphs of PDFs (left) and HRFs (right) of the UMBIIIC distribution

The UMBIIIC's QF is given in (23)

$$x_u = \left[\log(1-(\tfrac{1}{k}\log(1-(((u^{-\gamma/\alpha}-1)^{\frac{1}{\beta}}\gamma^{-\frac{1}{\beta}})+1)^{-1}))) \right]^{\frac{1}{\lambda}},\ 0 \le u \le 1. \tag{23}$$

7. Simulation

The performances of the MLE estimators were validated using Monte-Carlo simulation. Sample sizes of 25, 75, 100, 150, and 200 were used in the experiment. For any given sample size, the experiment was conducted 1500 times. The UMBIIIC distribution was employed in this context for a few chosen parameters. They are:

1. $\alpha = 1.95, \beta = 0.60, \lambda = 1.19, \gamma = 1.29, k = 4.40$ and
2. $\alpha = 2.01, \beta = 0.70, \lambda = 2.50, \gamma = 1.21, k = 4.00.$

Multiple measures were employed to evaluate the parameter estimators for every sample. They consist of the root mean square error (RMSE), average bias (AB), coverage probability (CP), and average width (AW). The experimental results, which are displayed in Table 1, showed that the AW, AB, and RMSE for the UMBIIIC distribution steadily decreased with increasing sample size for all parameter values. Furthermore, the CP values were found to be quite similar to the predicted value of 0.95. These findings highlight the accuracy with which the maximum likelihood approach can estimate the UMBIIIC distribution's parameters

8. Applications

Real-life dataset that was employed to demonstrate the effectiveness of the suggested model (UMBIIIC) can be found in [26]. The performance of the developed models and the existing ones were compared using several accuracy indicators. These indicators included the Bayesian information criterion (BIC), the negative of the loglikelihood (L), the Akaike information criterion (AIC), and the corrected Akaike information criterion (AICc). The goodness-of-fit (GoF) of the models was also assessed using additional measures like the Anderson-Darling (AD), Cramér-von Misses (CM), and Kolmogorov-Smirnov (KS) tests. The best model was determined to be the one with the lowest GoF, lowest information criteria, and highest negative loglikelihood. The *bbmle* package of the R statistical

Table 1

Simulation result for the distribution of UMBIIIC

Parameter	n	*I*				*II*			
		AB	**RMSE**	**CP**	**AW**	**AB**	**RMSE**	**CP**	**AW**
α	25	0.3851	1.8415	0.806	20.5897	0.7836	1.4551	0.9693	29.0989
	75	0.2987	1.4094	0.8107	9.8525	0.6187	1.2664	0.9253	12.4503
	100	0.2966	1.2806	0.8653	8.3306	0.5574	1.222	0.9141	10.4213
	150	0.2502	1.04	0.876	6.6613	0.4837	1.2139	0.8893	8.8678
	200	0.2278	0.9359	0.896	5.4776	0.3413	1.1025	0.886	6.4144
β	25	0.4721	1.3534	0.9987	9.2735	0.344	0.944	0.9999	8.5779
	75	0.0392	0.2598	0.9887	4.1382	0.1056	0.4226	0.9999	4.0292
	100	0.0227	0.2393	0.9993	3.5408	0.0885	0.3687	0.9987	3.5219
	150	0.0046	0.2168	0.9993	3.0899	0.0776	0.3108	0.9999	3.0979
	200	0.0052	0.2011	0.9999	2.6514	0.0727	0.3019	0.9993	2.5141
λ	25	1.2871	1.8124	0.9947	12.7657	0.9875	1.3552	0.996	22.4967
	75	0.4666	0.9334	0.9107	4.2521	0.8858	1.2036	0.998	14.5573
	100	0.3061	0.6912	0.9873	3.1192	0.7733	1.1307	0.9807	11.9172
	150	0.2619	0.5706	0.9907	2.4515	0.6602	1.0423	0.996	8.7981
	200	0.2019	0.4535	0.99	0.9333	0.6262	0.9917	0.994	7.6912
γ	25	0.8564	2.1701	0.8353	20.5897	0.6334	1.8979	0.9907	29.0989
	75	0.1085	1.4445	0.9107	9.8525	0.1918	1.1503	0.98	12.4503
	100	0.0492	1.2813	0.9253	8.3306	0.1386	1.0548	0.9807	10.4213
	150	0.0073	1.0553	0.938	6.661	0.1082	0.9403	0.9747	8.8678
	200	0.0062	0.9468	0.9573	5.4776	0.1026	0.7705	0.9733	6.4144

software's mle2 subroutine was used to maximize the models' log-likelihood function to estimate the parameters [6].

In Table 3, all the parameter estimates for the UMBIIIC, GUM, NAK, GINW, and GoIE are significant at a 5% level. For EINW, only the estimate of β is significant, and for TCPW, the estimates of α and λ are significant but that of β is insignificant at a 5% significant level.

8.1 *Application of data*

The dataset's source [26], consists of 51 observations of glass fiber strengths measuring 1.5 cm collected from the National Physical Laboratory (NPL) in England. The values for the smallest and largest observations were 0.55 and 2.24 respectively. The standard deviation was 0.327 and the mean was 1.442. The skewness and kurtosis coefficients were -0.644 and 3.803, in that order. This suggests that the dataset is heavy-tailed and negatively skewed.

The UMBIIIC distribution was compared to the following distributions for the strengths of the 1.5 cm glass fiber dataset: Gumbel (GUM) [14], Exponentiated Inverse Flexible Weibull Extension (EINFW) [12], Nakagami (NAK) [22], Generalized Inverse Weibull Distribution (GINW) [11], Truncated Cauchy Power Weibull (TCPW) [1], and Gompertz Inverse Exponential (GoIE) [25] [table 3].

The developed UMBIIIC distribution showed the best performance when compared with the other competing distributions, as shown in Table 2. This is because, across all measures considered, the UMBIIIC distribution had the lowest values.

The comparison of the fitted cumulative distribution functions (CDFs) of different candidate distributions with the empirical CDF is shown in Figure 3. The graphs unequivocally show that the empirical CDF of the glass data is closely approximated by the fitted CDF of the UMBIIIC distribution. This strong similarity suggests that the UMBIIIC distribution appropriately represents the underlying distribution of the data and offers a good fit for the observed data.

Table 2

Measures of GoF and information criteria for dataset

	-L	AIC	BIC	CAIC	AD	CM	KS
UMBIIIC	10.0971	30.1940	39.8530	31.5280	1.1021	0.1804	0.1528
GUM	23.6490	51.0990	54.9630	51.3490	3.4578	0.6524	0.2173
EINFW	21.7700	48.9540	54.9470	49.0970	3.0544	0.5761	0.2096
NAK	16.7390	37.4790	41.3420	37.5260	2.4190	0.4571	0.2114
GINW	35.7630	77.5270	83.3220	78.0370	1.2148	1.0010	0.2527
TCPW	13.7431	31.4860	39.9199	33.9150	1.6450	0.289	0.1720
GOIE	13.5720	33.404	38.9390	33.6550	1.5710	0.2820	0.1711

Although none of the density curves of the UMBIIIC distribution in Figure 2 is left-skewed, the superior performance of the UMBIIIC distribution when a left-skewed dataset from the NPL is applied to it indicates that the UMBIIIC can also model datasets that are left-skewed. This is an indication of the high flexibility of the distribution.

Table 3

Parameter estimates, standard errors and p-values for dataset

Model	Parameter	Estimate	SE	p-value
UMBIIIC	$\hat{\alpha}$	0.1205	0.0302	< 0.0001
	$\hat{\beta}$	16.2072	0.0023	< 0.0001
	$\hat{\lambda}$	0.4551	0.0902	< 0.0001
	$\hat{\gamma}$	0.4090	0.0106	< 0.0001
	$\hat{k}$	0.2612	0.0238	< 0.0001
GUM	$\hat{\alpha}$	1.2699	0.0543	< 0.0001
	$\hat{\beta}$	0.3641	0.0353	< 0.0001
EINFW	$\hat{\alpha}$	-0.5828	0.3049	0.0559
	$\hat{\beta}$	3.1608	0.3422	< 0.0001
	$\hat{\lambda}$	92.1540	58.1557	0.1131
NAK	$\hat{\alpha}$	2.1834	0.1418	< 0.0001
	$\hat{\beta}$	4.6465	0.8891	< 0.0001
GINW	$\hat{\alpha}$	0.6477	0.0401	< 0.0001
	$\hat{\beta}$	2.9058	0.2665	< 0.0001
	$\hat{\lambda}$	6.0576	0.0015	< 0.0001
	$\hat{\alpha}$	0.8000	0.3400	0.0184
TCPW	$\hat{\beta}$	0.0400	0.0500	0.4300
	$\hat{\lambda}$	6.3000	1.7200	< 0.0001
GoIE	$\hat{\alpha}$	0.6477	0.0286	0.0210
	$\hat{\beta}$	2.9059	0.1147	< 0.0001
	$\hat{\lambda}$	6.0576	0.2718	0.0015

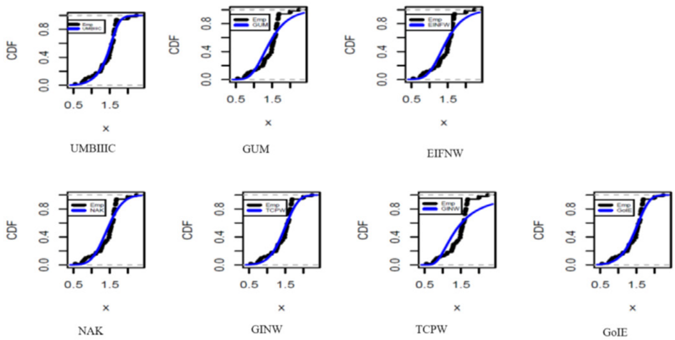

Figure 3

CDF graphs of empirical and fitted distributions for dataset

9. Log-UMBIIIW regression model

We use the transformation $Y = \log_e(X)$, where X and Y are random variables to obtain the log-UMBIII Weibull (LUMBIIIW) distribution. Expressing X in terms of Y, the transformation gives $x = e^y$ and $\frac{dx}{dy} = e^y$. As $x \to +\infty$, $y \to +\infty$. It can also be shown that $f_Y(y) = \left|\frac{dx}{dy}\right| f_X(y)$.

Following the transformation, the LUMBIIIW density function, $f_Y(y)$ is expressed as

$$f_Y(y) = |e^y|\, \alpha\lambda\beta k e^{y(k-1)}(1-e^{-\lambda e^{yk}})^{-2}((1-e^{-\lambda e^{yk}})^{-1}-1)^{\beta-1} \\ [1+\gamma((1-e^{-\lambda e^{yk}})^{-1}-1)^{\beta}]^{-(1+\alpha/\gamma)},\ y \in \mathbb{R}. \quad (24)$$

Simplifying $f_y(y)$ further gives

$$f_Y(y) = \alpha\lambda\beta k e^{yk}(1-e^{-\lambda e^{yk}})^{-2}((1-e^{-\lambda e^{yk}})^{-1}-1)^{\beta-1} \\ [1+\gamma((1-e^{-\lambda e^{yk}})^{-1}-1)^{\beta}]^{-(1+\alpha/\gamma)},\ y \in \mathbb{R}. \quad (25)$$

Considering re-parametrization where $\lambda = e^{(-\mu/\sigma)}$ and $k = 1/\sigma$, the resulting PDF is

$$f_Y(y) = \alpha\beta\frac{1}{\sigma}e^{-\mu/\sigma}e^{y/\sigma}(1-e^{-e^{-\mu/\sigma}e^{y/\sigma}})^{-2}((1-e^{-e^{-\mu/\sigma}e^{y/\sigma}})^{-1}-1)^{\beta-1} \\ \times[1+\gamma((1-e^{-e^{-\mu/\sigma}e^{y/\sigma}})^{-1}-1)^{\beta}]^{-(1+\alpha/\gamma)}, \quad (26)$$

where $\sigma > 0$, $\alpha > 0$, $\gamma > 0$, $\beta > 0$, $\mu \in \mathbb{R}$ and $y \in \mathbb{R}$.

Hence, equation (26) can further be simplified as

$$f_Y(y;\alpha,\gamma,\beta,\mu,\sigma)=\frac{\alpha\beta}{\sigma}e^{\left(\frac{y-\mu}{\sigma}\right)}(1-e^{-e^{\left(\frac{y-\mu}{\sigma}\right)}})^{-2}((1-e^{-e^{\left(\frac{y-\mu}{\sigma}\right)}})^{-1}-1)^{\beta-1}$$
$$[1+\gamma((1-e^{-e^{\left(\frac{y-\mu}{\sigma}\right)}})^{-1}-1)^{\beta}]^{-(1+\alpha/\gamma)}. \quad (27)$$

The PDF, $f_Y(y)$ can consequently be rewritten as

$$f_Y(y)=\frac{\alpha\beta}{\sigma}\exp\left(\frac{y-\mu}{\sigma}\right)\left(1-\exp\left(-\exp\left(\frac{y-\mu}{\sigma}\right)\right)\right)^{-2}\left(\left(1-\exp\left(-\exp\left(\frac{y-\mu}{\sigma}\right)\right)\right)^{-1}-1\right)^{\beta-1}$$
$$\times\left[1+\gamma\left(\left(1-\exp\left(-\exp\left(\frac{y-\mu}{\sigma}\right)\right)\right)^{-1}-1\right)^{\beta}\right]^{-(1+\alpha/\gamma)}, \quad (28)$$

where $\mu\in\mathbb{R},\ \sigma>0,\ \alpha>0,\ \gamma>0,\ \beta>0.$

If the CDF of the LUMBIIIW distribution is expressed as $F_y(y)=P(Y\le y)$, with $Y=\log_e(X)$ in mind we obtain $F_Y(y)=P(X\le e^y)$.

We therefore express the CDF of the LUMBIIIW distribution as

$$F_Y(y)=[1+\gamma((1-e^{-\lambda e^{yk}})^{-1}-1)^{\beta}]^{-\alpha/\gamma},y\in\mathbb{R}. \quad (29)$$

Using the re-parameterization $\lambda=e^{(-\mu/\sigma)}$ and $k={}^{1}\!/_{\sigma}$, the resulting CDF of the LUMBIIIW distribution is given as

$$F_y(y)=[1+\gamma((1-e^{-e^{\left(\frac{y-\mu}{\sigma}\right)}})^{-1}-1)^{\beta}]^{-\alpha/\gamma},y\in\mathbb{R}. \quad (30)$$

The LUMBIIIW distribution's survival function, $S_Y(y)$, can consequently be expressed as

$$S_Y(y)=1-[1+\gamma((1-e^{-e^{\left(\frac{y-\mu}{\sigma}\right)}})^{-1}-1)^{\beta}]^{-\alpha/\gamma},y\in\mathbb{R}. \quad (31)$$

Considering $z=(y-\mu)/\sigma$,

$$f_Z(y)=\frac{\alpha\beta}{\sigma}\exp(z)(1-\exp(-\exp(z)))^{-2}((1-\exp(-\exp(z)))^{-1}-1)^{\beta-1}$$
$$\times[1+\gamma((1-\exp(-\exp(z)))^{-1}-1)^{\beta}]^{-(1+\alpha/\gamma)}. \quad (32)$$

We present an entirely new LSR model where the structure of this new LSR model hinges on the LUMBIIIW density.

$$y_i=\mathbf{q}_\mathbf{i}^\mathbf{T}\mathbf{B}+\sigma z_{i,}\ i=1,\ldots,n, \quad (33)$$

in which σz is the random error that follows the density in (32), and $\mathbf{q}_\mathbf{i}^\mathbf{T}$ is the independent variable vector. Underlying the location parameter, $\mu_i=\mathbf{q}_i^T\mathbf{B}$, are the regression parameters, represented by $\mathbf{B}=(B_1,\ldots,B_p)^T$.

The LUMBIIIW regression model is given as

$$y_i = B_0 + B_1 q_{i1} + B_2 q_{i2} + B_3 q_{i3} + \sigma z_i, \quad i = 1, \ldots, n. \tag{34}$$

The parameters of the LUMBIIIW regression equation were estimated by means of the MLE estimation technique. The dependent variable, represented by y_i in (34), is distributed as LUMBIIIW. $\min\{\log_e(x), \log_e(c)\}$ defines y_i. x represents life-time and c, censoring times. The individuals with the log-censoring and log-lifetime, respectively, are indicated by the two sets of observations, F and C, that are subsequently created. With reference to the definition, the LUMBIIIW regression model's loglikelihood function is

$$\begin{aligned}
\ell(\tau) = r\left(\frac{\alpha\beta}{\sigma}\right) + \sum_{i\in F} z_i - 2\sum_{i\in F} \log(1-\exp(-\exp(z_i))) + (\beta-1) \\
\sum_{i\in F} \log((1-\exp(-\exp(z_i)))^{-1} - 1) \\
-(1+\alpha/\gamma)\sum_{i\in F} \log[1+\gamma((1-\exp(-\exp(z_i)))^{-1}-1)^{\beta}] \\
+\sum_{i\in C} \log(1-[1+\gamma((1-\exp(-\exp(z_i)))^{-1}-1)^{\beta}]^{(-\alpha/\gamma)})
\end{aligned} \tag{35}$$

The number of uncensored observations is denoted by r, and the parameter vector $z_i = (y_i - \mathbf{q}_i^{\mathbf{T}}\mathbf{B})/\sigma$ is represented as $\tau = (\alpha, \beta, \gamma, \sigma, B^T)$. By maximizing the loglikelihood function in (35) the maximum likelihood estimators for the LUMBIIIW regression model are obtained.

We used a real-life dataset to test the developed LUMBIIIW regression model. 76 kidney patients who use portable dialysis equipment have their time to infection recurrence listed in the dataset. This dataset can be downloaded from the publication by [18] or accessed through the R software's survival package [20]. The log-generalized inverse-Weibull (LGIW) regression model [11] and the log-Lomax-Weibull (LLW) regression model [9] were the two existing models compared to evaluate the performance of the LUMBIIIW regression model. The kidney dataset had a roughly 24% censoring rate. This dependent variable was the time to infection recurrence (y_i). Patient age (q_{i1}), was one of the independent variables. Other variables were frailty estimate of patients (q_{i2}) and sex of patients (q_{i3}).

Table 4 displays these competing models' parameter estimates and GoFs. The table shows that the LUMBIIIW regression model is the best competitor since it has the lowest GoF values out of all the competing models. At a 5% significance level, it was also found that every parameter for the variables under investigation was positive and significant.

The table's results also show that as people age, their risk of developing kidney infections rises. The risk of kidney infection also rises with an increase in frailty estimate. Furthermore, compared to male patients, female patients had a higher risk of kidney infection.

The kidney infection dataset was employed to evaluate the appropriateness of the LUMBIIIW regression model. To do this the Cox-Snell residuals (CSRs) were estimated. If a model adequately fits the data, the CSRs of the sample should exhibit an exponential distribution whose parameter is unity [24].

The developed LUMBIIIW regression model provides a satisfactory fit to the dataset, as demonstrated by the Probability-Probability (P-P) plot of the CSRs, where the plotted points are aligned to the diagonal line as shown in Figure 4. Table 4 presents the GoF measures and their corresponding p-values (in parentheses) that support this finding.

Table 4

Regression parameter estimates and GoF measures

Model	Parameters	Estimate	SE	p-value	GoF
LUMBIIIW	B_0	11917.000	7.15×10^{-7}	< 0.0001	AIC=1472.4270
	B_1	0.4720	3.37×10^{-5}	< 0.0001	
	B_2	71974.0000	1.53×10^{-6}	< 0.0001	BIC=1491.0730
	B_3	17701.0000	4.19×10^{-6}	< 0.0001	
	α	0.0110	1.55×10^{-2}	0.4900	AIC_C=1474.5760
	β	0.6680	2.94×10^{-1}	0.0232	
	γ	0.0320	3.08×10^{-2}	0.2948	
	σ	11236.0000	1.56×10^{-5}	< 0.0001	
LGIW	B_0	1101000.0000	1.05×10^{-7}	< 0.0001	AIC=1765.5330
	B_1	-23064.0000	4.00×10^{-6}	< 0.0001	
	B_2	9000.0000	1.30×10^{-7}	< 0.0001	BIC=1779.5180
	B_3	1.0000	7.34×10^{-8}	< 0.0001	
	α	0.9730	1.16×10^{-1}	< 0.0001	AIC_C=1767.6820
	σ	1100000.0000	2.22×10^{-8}	< 0.0001	

Contd...

LLW	B_0	452110.0000	3.15×10^{-9}	< 0.0001	AIC=1547.9220
	B_1	1926.9000	1.28×10^{-7}	< 0.0001	
	B_2	89985.0000	2.86×10^{-9}	< 0.0001	BIC=1564.2370
	B_3	162220.0000	1.27×10^{-9}	< 0.0001	
	α	49.0440	6.53×10^{-6}	< 0.0001	AIC_C=1549.5690
	β	0.0490	6.48×10^{-4}	< 0.0001	
	σ	100010.0000	2.13×10^{-8}	< 0.0001	

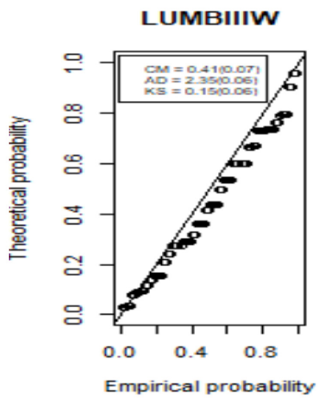

Figure 4

Graph of CSRs of fitted regression model

10. Conclusions

The UMBIII family of distributions, a newly created family of distributions, is proposed. The new family of distributions' mathematical and statistical characteristics, have been derived. The unit modified Burr III baseline CDF was used to develop two unique families of distributions, the UMBIIIW and UMBIIIC, along with the quantile functions that go with them. According to simulation results, for all parameter values in the UMBIIIC distribution, the AW, AB, and RMSE were decreasing as sample size increased. In addition, the CPs closely matched the nominal value of 0.95. This demonstrates how well the maximum likelihood approach estimates the parameters of the UMBIIIC distribution. When the developed models were applied to real-life datasets, it was shown that UMBIIIC outperformed the other models in terms of information criteria and GoF statistics. A step further, the LUMBIIIW regression model was developed.

The developed regression model performed better for the given dataset than both the LLW and LGIW regression models. The suitability of the developed regression model was evaluated using the CSRs.

References

[1] Aldahlan, M. A., Jamal, F., Chesneau, C., Elgarhy, M., and Elbatal, I. The truncated cauchy power family of distributions with inference and applications. *Entropy*, 22(3), 1– 24 (2020).

[2] Anzagra, L., Sarpong, S., and Nasiru, S. Chen-G class of distributions. *Cogent Mathematics and Statistics*, **7**(1), 1721401 (2020).

[3] Arifa, S., Yab, M. Z., and Ali, A. The modified Burr III G family of distributions. *Journal of Data Science*, 15(1), 41-60 (2017).

[4] Bhatti, F. A., Hamedani, G. G., Korkmaz, M. C., Cordeiro, G. M., Yousof, H. M., and Ahmad, M. On Burr III Marshal Olkin family: development, properties, characterizations and applications. *Journal of Statistical Distributions and Applications*, 6, 1-21 (2019).

[5] Bourguignon, M., Silva, R. B., and Cordeiro, G. M. The Weibull-G family of probability distributions. *Journal of data science*, 12(1), 53-68 (2014).

[6] Bolker, M. Ben. Tools for general maximum likelihood estimation. R development (2014).

[7] Chen, Z. A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statistics and Probability Letters*, 49(2), 155–161 (2000).

[8] Chipepa, F., Oluyede, B. O., and Peter, P. O. The Burr III-Topp-Leone-G family of distributions with applications. Heliyon (2021).

[9] Cordeiro, G. M., Ortega, E. M. M., Popović, B. V., and Pescim, R. R. The Lomax generator of distributions: Properties, minification process and regression model. *Applied Mathematics and Computation*, 247, 465–486 (2014).

[10] Cordeiro, G. M., Silva, G. O., and Ortega, E. M. M. The Beta Extended WeibullFamily. *Journal of Probability and Statistical Science*, 10(1), 15–40 (2012).

[11] de Gusmão, F. R. S., Ortega, E. M. M., and Cordeiro, G. M. The generalized inverse Weibull distribution. *Statistical Papers*, 52(3), 591–619 (2011).

[12] El-Morshedy, M., El-Bassiouny, A. H., and El-Gohary, A. Exponentiated Inverse Flexible Weibull Extension Distribution. *Journal of Statistics Applications and Probability*, 6(1), 169–183 (2017).

[13] Goyal, A. Complete Guide to Moment Generating Functions in Statistics For Data Science. Analytics Vidhya (2023).

[14] Gumbel. *Analysis of empirical bivariate extremal distributions*. 15(4), 414–422 (1944).

[15] Gupta, R. C., and Bradley, D. M. Representing the mean residual life in terms of the failure rate. Mathematical and Computer Modelling (2003).

[16] Haq, M. A. ul, Hashmi, S., Aidi, K., Ramos, P. L., and Louzada, F. Unit Modified Burr-III Distribution: Estimation, Characterizations and Validation Test. *Annals of Data Science* (2020).

[17] Hussein, M., Elsayed, H., and Cordeiro, G. M. A New Family of Continuous Distributions: Properties and Estimation. *Symmetry*, 14(2), 276 (2022).

[18] McGilchrist, C., and Aisbett, C. Regression with frailty in survival analysis. *Biometrics*, 47, 461–466 (1991).

[19] Modi, K., and Gill, V. Unit Burr-III distribution with application. *Journal of Statistics and Management Systems*, 23(3), 579-592 (2020).

[20] Moore, D. F. Applied survival analysis using R (Vol. 473). Cham: Springer (2016).

[21] Nadarajah, S., Cordeiro, G. M., and Ortega, E. M. M. The Zografos-Balakrishnan-G family of distributions: Mathematical properties and applications. *Communications in Statistics - Theory and Methods*, 44(1), 186–215 (2009).

[22] Nakagami M. The m-Distribution—A General Formula of Intensity Distribution of Rapid Fading. In *Statistical Methods in Radio Wave Propagation* (Issue 2). Pergamon Press Inc. (1960).

[23] Nasiru, S., Abubakari, A. G., and Abonongo, J. Unit Nadarajah-Haghighi Generated Family of Distributions: Properties and Applications. *Sankhya A*, 1–27 (2020).

[24] Nasiru, S., Abubakari, A. G., and Angbing, I. D. Bounded Odd Inverse Pareto Exponential Distribution: Properties, Estimation, and Regression. *International Journal of Mathematics and Mathematical Sciences* (2021).

[25] Oguntunde, P. E., Khaleel, M. A., Adejumo, A. O., Okagbue, H. I., Opanuga, A. A., and Owolabi, F. O. The Gompertz Inverse Exponential (GoIE) distribution with applications. *Cogent Mathematics and Statistics*, 5(1), 1507122 (2018).

[26] Smith, R. L., and Naylor, J. C. Statistics of the three-parameter Weibull distribution. *Annals of Operations Research*, 9(1), 577–587 (1987).

[27] Temraz, N. S. Y. Inference on the stress strength reliability with exponentiated generalized Marshall Olkin-G distribution. PLOS ONE (2023).

[28] Weibull, W. Statistical theories of the strength of materials. *Royal British Institute of Engineering*, 51, 189–206 (1939).

[29] Zubair, A. The Zubair-G Family of Distributions: Properties and Applications. *Annals of Data Science*, 7(2), 195–208 (2018).

Received June, 2023