Journal of Statistical Modelling: Theory and Applications
Vol. 4, No. 1, 2023, pp. 25-44
Yazd University Press 2023

JSMTA

Research Paper

Logit Gudermannian distribution: Properties, regression and applications

Abdulzeid Yen Anafo[1], Selasi Kwaku Ocloo*[1], Hanifatu Napari Mumuni[2], Suleman Nasiru[3]
[1]Department of Mathematical Sciences, University of Mines and Technology, Ghana
[2]Statistics Department, Tamale Technical University, Tamale, Ghana
[3]Department of Statistics and Actuarial Science, School of Mathematical Sciences, C. K. Tedam University of Technology and Applied Sciences, Ghana

Received: May 14, 2023/ Revised: October 08, 2023/ Accepted: December 18, 2023

Abstract: In this paper, we propose a unit distribution called the logit Gudermannian distribution and present various statistical properties of the proposed model. Six parameter estimation methods are explored in the quest to estimate the parameters of the proposed distribution. We determine which estimation methods provide better parameter estimates through simulation studies. The study shows that the logit Gudermannian distribution provides a better fit for the datasets used than other unit distributions. Consequently, the logit Gudermannian distribution is used to develop a parametric regression model for studying the relationship between a unit response variable and other exogenous variables. The new regression model's performance is compared to that of other existing regression models and found to be competitive.

Keywords: Gudermannian distribution; One-parameter distribution; Regression; Unit distribution; Univariate transformation.
Mathematics Subject Classification (2010): 60E05, 62F10, 62PO5.

1 Introduction

In the area of probability distributions, researchers are constantly searching for robust approaches to accurately model complex data. One of such intriguing addition to the

*Corresponding author: slsocloo@gmail.com

field is the Gudermannian distribution (Altun, 2019), a versatile probability distribution that offers unique advantages in capturing the characteristics of diverse datasets. Derived from the Gudermannian function, this distribution exhibits remarkable flexibility, making it a valuable tool in various domains, including statistics, machine learning, and data science. Classical probability distributions, such as the normal or Gaussian distribution, have long been the options available due to their simplicity and well-defined properties. However, their limitations in handling non-linear and skewed data have motivated the exploration of alternative distributions that better capture the inherent complexities present in many real-world datasets. Although the beta distribution has been adopted in modelling unit data, it has some limitations. As a result, several flexible distributions have been proposed as alternatives. Distributions including the unit exponential Pareto distribution proposed by Haj Ahmad et al. (2023), compound class of unit Burr XII model proposed by Zayed et al. (2023) , the bounded truncated Cauchy power exponential proposed by Nasiru et al. (2022) , the unit Burr XII distribution proposed by Korkmaz and Chesneau (2021), the bounded odd inverse Pareto exponential distribution proposed by Nasiru et al. (2021), unit power-logarithmic distribution proposed by Chesneau (2021) , the transmuted unit Rayleigh distribution proposed by Korkmaz et al. (2021), the unit Lindley distribution proposed by Mazucheli et al. (2019), the unit-Rayleigh distribution proposed by Bantan et al. (2020), the unit-Birnbaum-Saunders distribution proposed by Mazucheli et al. (2018), the McDonald arcsine distribution proposed by Cordeiro and Lemonte (2014), Kumaraswamy distribution proposed by Kumaraswamy (1980) and the Topp-Leone distribution proposed by Topp and Leone (2021).

The Logit Gudermannian (LG) distribution emerges as an option, bridging the gap between classical distributions and the need for more robust modeling techniques. The LG distribution is characterized by its ability to handle bounded data, as it maps values from the entire real line onto a finite interval. This property makes it particularly well-suited for modeling variables that naturally exhibit upper and lower bounds, such as probabilities, proportions, or constrained measurements. By leveraging the Gudermannian function, which maps the real line to a bounded range, and the logit transformation, which converts probabilities into unbounded values, this distribution provides a powerful framework for capturing and analyzing a wide range of unit data.

The applications of the LG distribution would be diverse and span multiple disciplines. Its usage would extend beyond classical statistical modeling and find relevance in fields such as actuarial science, risk assessment, time series analysis, and beyond.

In this article, we delve into the LG distribution exploring its mathematical formulation, key statistical properties, and practical applications. We demonstrate its versatility through practical applications and the formulation of the LG parametric regression model.

The following is an outline of how the paper is structured: Section 2 introduces the derivation of the probability density function (PDF), cumulative distribution function (CDF) and failure rate function (FRF). Some statistical properties of the proposed model are derived in Section 3. Section 4 proposes the LG generated family of distributions. Six methods of parameter estimation, the parametric LG regression model, and simulation studies assessing the estimation methods are discussed in Section 5, 6, and 7, respectively. Section 8 applies the proposed model to real datasets to demonstrate its usefulness, and Section 9 concludes the study.

2 Logit Gudermannian distribution

According to Altun (2019), the CDF of the Gudermannian distribution is given by $F_Y(y;\zeta) = \frac{2}{\pi}\arctan(e^{\zeta y})$, $\zeta > 0, y \in \mathbb{R}$ while the PDF is denoted by

$$f_Y(y;\zeta) = \frac{2\zeta e^{\zeta y}}{\pi(1+e^{2\zeta y})}, \quad \zeta > 0, \quad y \in \mathbb{R}.$$

Given that the random variable X is related to Y such that $X = 1/(1+e^{-Y})$. Then by univariate transformation of random variables, the CDF of the LG distribution is given by

$$\begin{aligned} F_X(x;\zeta) &= F_Y\left(-\log\left(\frac{1-x}{x}\right);\zeta\right) \\ &= \frac{2}{\pi}\arctan\left[x^{\zeta}(1-x)^{-\zeta}\right], \quad x \in (0,1), \quad \zeta > 0. \end{aligned}$$

The corresponding PDF and the FRF of the LG distribution are given by

$$\begin{aligned} f_X(x;\zeta) &= \frac{2\zeta x^{\zeta-1}(1-x)^{-\zeta-1}}{\pi[1+x^{2\zeta}(1-x)^{-2\zeta}]}, \quad x \in (0,1) \\ h_X(x;\zeta) &= \frac{2\pi^{-1}\zeta x^{\zeta-1}(1-x)^{-\zeta-1}}{[1+x^{2\zeta}(1-x)^{-2\zeta}][1-2\pi^{-1}\arctan[x^{\zeta}(1-x)^{-\zeta}]]}, \quad x \in (0,1), \end{aligned}$$

respectively.

The shapes of the PDF and FRF are shown in Figures 1 and 2, respectively.

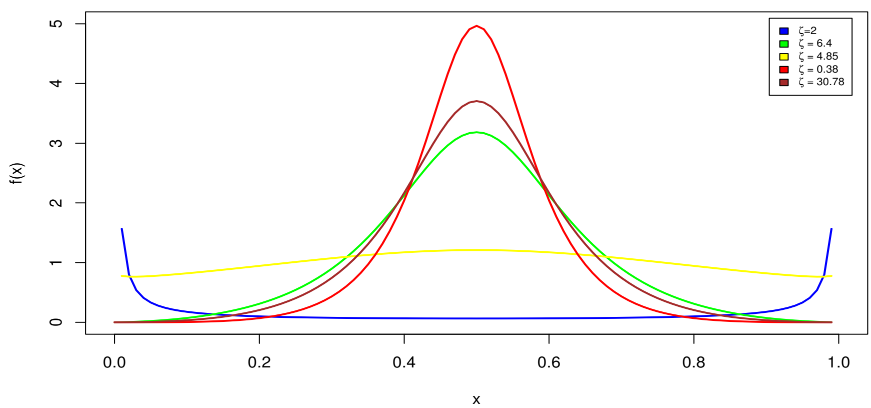


Figure 1: The PDF of the LG distribution.

The linear representation of the PDF can be used to deduce the statistical properties of the LG distribution.

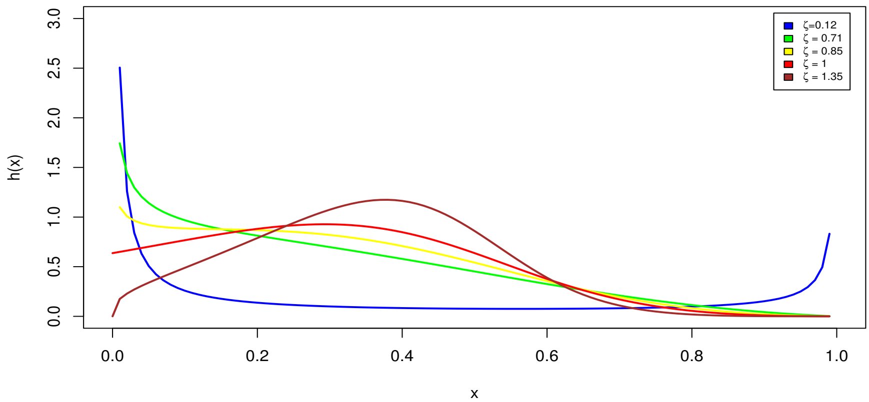


Figure 2: The FRF shapes of the LG distribution.

Lemma 2.1. *The linear representation of the PDF of the LG distribution is given as*

$$f_X(x;\zeta) = \frac{2\zeta}{\pi}\sum_{l=0}^{\infty}\sum_{p=0}^{\infty}(-1)^l\binom{\zeta(2l+1)+p}{p}x^{\zeta(2l+1)+p-1}.$$

Proof. By employing the Taylor series expansion,

$$z^{-\lambda} = \sum_{l=0}^{\infty}\frac{(-1)^l\lambda^{(l)}}{l!}z^l, \quad z>0,$$

where $\lambda^{(l)} = \lambda(\lambda+1)(\lambda+2)\ldots(\lambda+l-1)$ is the rising factorial,

$$\frac{1}{[1+x^{2\zeta}(1-x)^{-2\zeta}]} = \sum_{l=0}^{\infty}(-1)^l x^{2\zeta l}(1-x)^{-2\zeta l}.$$

The PDF can be expressed as

$$f_X(x;\zeta) = \frac{2\zeta}{\pi}\sum_{l=0}^{\infty}(-1)^l x^{\zeta(2l+1)-1}(1-x)^{-[\zeta(2l+1)+1]}.$$

The PDF

$$f_X(x;\zeta) = \frac{2\zeta}{\pi}\sum_{l=0}^{\infty}\sum_{p=0}^{\infty}(-1)^l\binom{\zeta(2l+1)+p}{p}x^{\zeta(2l+1)+p-1},$$

can be obtained by utilizing the generalised binomial expansion

$$(1-v)^{-\omega} = \sum_{p=0}^{\infty}\binom{\omega+p-1}{p}v^p, \quad |v|<1.$$

This completes the proof. □

The quantile function of the LG distribution can be expressed as

$$Q_X(q) = [1 + (\tan[\pi q/2])^{-1/\zeta}]^{-1}, \quad q \in [0, 1].$$

Proposition 2.2. *The CDF of the LG distribution increases in ζ for $x \in (0, 0.5]$ and decreases in ζ for $x \in [0.5, 1]$ given $\zeta > 0$.*

Proof. Differentiating the CDF with respect to ξ , we obtain

$$\frac{\partial F_X(x;\zeta)}{\partial \zeta} = -\frac{2\left(\frac{1-x}{x}\right)^{\zeta}\log\left(\frac{1-x}{x}\right)}{\pi\left[1+\left(\frac{1-x}{x}\right)^{2\zeta}\right]}.$$

Using the fact that $\log\left(\frac{1-x}{x}\right) < 0$ for $x \in [0.5, 1]$ and $\log\left(\frac{1-x}{x}\right) > 0$ for $x \in [0, 0.5)$, the results in Proposition 2.2 is achieved. □

Proposition 2.2 can be used to deduced the stochastic ordering property. $F_X(x;\zeta)$ is positively ordered with respect to ζ for $x \in (0, 0.5]$, thus if $\zeta_1 \le \zeta_2$, then $F_X(x;\zeta_1) \le F_X(x;\zeta_2)$ and negatively ordered with respect to ξ for $x \in [0.5, 1]$, thus if $\zeta_1 \le \zeta_2$, then $F_X(x;\zeta) \ge F_X(x;\zeta_2)$.

3 Statistical properties

In this section, we discuss various statistical properties of the LG distribution, including nth non-central and incomplete moments, moment generating functions, and order statistics.

3.1 Moments and generating functions

Moments play a crucial role in analyzing measures of central tendency, spread, and shape in statistical analysis.

Proposition 3.1. *The nth non-central moment of the LG distribution is*

$$\mu_n^{'} = \frac{2\zeta}{\pi}\sum_{i=0}^{\infty}\sum_{p=0}^{\infty}\frac{(-1)^l(\zeta(2l+1)+p)\Gamma(\zeta(2l+1)+p)}{(n+\zeta(2l+1)+p)(\zeta(2l+1))\Gamma(\zeta(2l+1))p\Gamma(p)}, \quad n = 1, 2, \ldots,$$

where $\Gamma(.)$ is the gamma function.

Proof. The nth non-central moment is defined by $\mu_n^{'} = \int_0^1 x^n dF_x(x;\zeta)$. Hence, substituting the linear expansion of the PDF into the definition we have

$$\mu_n^{'} = \frac{2\zeta}{\pi}\sum_{i=0}^{\infty}\sum_{p=0}^{\infty}(-1)^l\binom{\zeta(2l+1)+p}{p}\int_0^1 x^{n+\zeta(2l+1)+p-1}dx.$$

Simplifying the integral and using the fact that $a\Gamma(a) = \Gamma(a+1)$, we obtain

$$\mu_n^{'} = \frac{2\zeta}{\pi}\sum_{i=0}^{\infty}\sum_{p=0}^{\infty}\frac{(-1)^l(\zeta(2l+1)+p)\Gamma(\zeta(2l+1)+p)}{(n+\zeta(2l+1)+p)(\zeta(2l+1))\Gamma(\zeta(2l+1))p\Gamma(p)}, \quad n = 1, 2, \ldots.$$

The central moment (μ_n) and the cumulants (κ_n) of the LG random variable are obtained as

$$\mu_n = \sum_{i=0}^{n}(-1)^i\binom{n}{i}\mu_1^{'n}\mu'_{n-i},$$
$$\kappa_n = \mu^{'} - \sum_{i=1}^{n-1}\binom{n-1}{i-1}\kappa_i\mu^{'}_{n-i},$$

respectively, where $\kappa_1 = \mu^{'}_1, \kappa_2 = \mu^{'}_2 - (\mu^{'}_1)^2$, $\kappa_3 = \mu^{'}_3 - 3\mu^{'}_2\mu^{'}_1 + 2(\mu^{'}_1)^3$, $\kappa_4 = \mu^{'}_4 - 4\mu^{'}_3\mu^{'} + 6\mu^{'}_2(\mu^{'}_1)^2 - 3(\mu^{'}_1)^4$ and so on. The coefficient skewness and kurtosis are respectively obtained from the cumulants using $\phi_1 = \kappa_3/\kappa_2^{3/2}$ and $\phi_2 = \kappa_4/\kappa_2^2$. □

Proposition 3.2. *The nth incomplete moment of the LG distribution is given by*

$$\varphi_n(y) = \frac{2\zeta}{\pi}\sum_{l=0}^{\infty}\sum_{p=0}^{\infty}\frac{(-1)^l(\zeta(2l+1)+p)\Gamma(\zeta(2l+1)+p)y^{n+\zeta(2l+1)+p}}{(n+\zeta(2l+1)+p)(\zeta(2l+1))\Gamma(\zeta(2l+1))p\Gamma(p)}, \quad n = 1, 2, \ldots.$$

Proof. The nth incomplete moment is given by $\varphi_n(y) = \int_0^y x^n dF_X(x;\zeta)$. Using the linear expansion of the PDF and simplifying we have

$$\begin{aligned}\varphi_n(y) &= \frac{2\zeta}{\pi}\sum_{l=0}^{\infty}\sum_{p=0}^{\infty}\binom{\zeta(2l+1)+p}{p}\int_0^y x^{n+\zeta(2l+1)+p-1}dx \\ &= \frac{2\zeta}{\pi}\sum_{l=0}^{\infty}\sum_{p=0}^{\infty}\frac{(-1)^l(\zeta(2l+1)+p)\Gamma(\zeta(2l+1)+p)y^{n+\zeta(2l+1)+p}}{(n+\zeta(2l+1)+p)(\zeta(2l+1))\Gamma(\zeta(2l+1))p\Gamma(p)}.\end{aligned}$$

□

Proposition 3.3. *The moment generating function of the LG distribution is given by*

$$M_X(t) = \frac{2\zeta}{\pi}\sum_{n=0}^{\infty}\sum_{l=0}^{\infty}\sum_{p=0}^{\infty}\frac{(-1)^l t^n(\zeta(2l+1)+p)\Gamma(\zeta(2l+1)+p)}{n!(n+\zeta(2l+1)+p)(\zeta(2l+1))\Gamma(\zeta(2l+1))p\Gamma(p)}.$$

Proof. By definition, the moment generating function is given as

$$M_X(t) = E(e^{tX}) = \int_0^1 e^{tx}dF_X(x;\zeta) = \sum_{n=0}^{\infty}\frac{t^n}{n!}\mu^{'}_n.$$

Substituting the nth non-central moment, we obtain the required function. □

3.2 Order statistics

By utilizing order statistics, researchers can approximate the minimum and maximum values, as well as the range, of a given set of data. If $X_1 \le X_2 \le \cdots \le X_n$ are order statistics from the LG distribution, it follows that the PDF of the kth order statistic is given as

$$f_k(x;\zeta) = \Omega_k[F_X(x;\zeta)]^{k-1}[1 - F_X(x;\zeta)]^{n-k}f_X(x;\zeta),$$

where $\Omega_k = \frac{n!}{(k-1)!(n-k)!}$.

Using the binomial series expansion, $(1-z)^{\delta-1} = \sum_{l=0}^{\infty}(-1)^l\binom{\delta-1}{l}z^l, |z| \leq 1$, the PDF of the kth order statistics given by

$$f_k(x) = \Omega_k \sum_{l=0}^{k-1}(-1)^l\binom{k-1}{l}[1-F_X(x;\zeta)]^{n-k+1}f_X(x;\zeta).$$

The PDF of the kth order statistics obtained from the LG distribution is

$$f_k(x) = \frac{\Omega_{k:n}2\zeta x^{\zeta-1}(1-x)^{-\zeta-1}}{\pi[1+x^{2\zeta}(1-x)^{-2\zeta}]}\sum_{l=0}^{k-1}(-1)^j\binom{k-1}{l}\left[1-\frac{2}{\pi}\arctan\left[\left(\frac{1-x}{x}\right)^{-\zeta}\right]\right]^{n-k+l}.$$

The CDF of the minimum order statistic X_1 is derived as

$$\begin{aligned} F_{X_1}(x) &= 1-[1-F_X(x;\zeta)]^n \\ &= 1-\left[1-\frac{2}{\pi}\arctan\left[\left(\frac{1-x}{x}\right)^{-\zeta}\right]\right]^n, \end{aligned}$$

while the CDF of the maximum order statistic Y_n is given by;

$$F_{X_1}(x) = [F_X(x)]^n = 1-\left[\frac{2}{\pi}\arctan\left[\frac{1-x}{x}\right]^{\zeta}\right]^n.$$

4 LG generated family of distributions

The developments of generators for modifying existing distributions have gained much attention recently. This section presents another family of distributions using the LG distribution. This new family is known as the LG-generated (LG-G) family of distributions. Suppose that the random variable follows the LG-G family of distributions. Then the CDF of the new family is obtained as

$$F_Y(y) = \frac{2}{\pi}\arctan\left[\left(\frac{G(y;\varpi)}{1-G(y;\varpi)}\right)^{\zeta}\right], \quad \zeta>0, \quad y\in\mathbb{R}, \quad \varpi\in\mathbb{R}.$$

The corresponding PDF of the family is given by

$$f_y(y) = \frac{2\zeta g(y;\varpi)G(y;\varpi)^{\zeta-1}(1-G(y;\varpi))^{-\zeta-1}}{\pi[1+G(y;\varpi)^{2\zeta}(1-G(y;\varpi))^{-2\zeta}]}, \quad y\in\mathbb{R}, \quad \varpi\in\mathbb{R}.$$

The failure rate function is given by

$$h_Y(y) = \frac{2\zeta g(y;\varpi)G(y;\varpi)^{\zeta-1}(1-G(y;\varpi))^{-\zeta-1}}{\pi[1+G(y;\varpi)^{2\zeta}(1-G(y;\varpi))^{-2\zeta}]\left[1-\frac{2}{\pi}\arctan\left[\left(\frac{G(y;\varpi)}{1-G(y;\varpi)}\right)^{\zeta}\right]\right]}.$$

The details of the LG-G family are beyond the scope of this paper.

5 Parameter estimation methods

The estimation of LG distribution parameters with a focus on six techniques are discussed in this section

5.1 Maximum likelihood estimation

If $x_1, x_2, \ldots x_v$ is a random sample of size v from the LG distribution, the the log-likelihood function for estimating the parameter ζ can be expressed as

$$\begin{aligned}\ell &= v\log\left(\frac{2\zeta}{\pi}\right) + (\zeta-1)\sum_{i=1}^{v}\log(x_i) - (\zeta+1)\sum_{i=1}^{v}\log(1-x_i)\\ &\quad -\sum_{i=1}^{v}\log(1+x_i^{2\zeta}(1-x_i)^{-2\zeta}).\end{aligned}$$

Using MATLAB, MATHEMATICA, or R, we can directly maximize the log-likelihood function to obtain the parameter estimate.

5.2 Ordinary and weighted least squares estimation

If $x_{(1)}, x_{(2)}, \ldots, x_{(v)}$ are the ordered observations from the LG distribution, then the ordinary least squares (LSE) estimate of $\hat{\zeta}$ is obtained by minimising the function

$$LSE = \sum_{i=1}^{v}\left[\left(\frac{4}{\pi}\arctan\left[\left(\frac{1-x_{(i)}}{x_i}\right)^{-\zeta}\right]\right) - \frac{i}{v+1}\right]^2,$$

with respect to ζ.

The weighted least squares (WLSE) estimate of $\hat{\zeta}$ is obtained by minimising the function

$$WLSE = \sum_{i=1}^{v}\frac{(v+1)^2(v+2)}{l(v-i+1)}\left[\left(\frac{4}{\pi}\arctan\left[\left(\frac{1-x_{(i)}}{x_i}\right)^{-\zeta}\right]\right) - \frac{i}{v+1}\right]^2,$$

with respect to ζ.

5.3 Cramér-von Mises estimation

If $x_{(1)}, x_{(2)} \ldots, x_{(v)}$ are the ordered observations based on the LG distribution, then the Cramér-von Mises (CVME) estimate of $\hat{\zeta}$ is obtained by minizing the function:

$$CVME = \frac{1}{12v} + \sum_{i=1}^{v}\left[\left(\frac{4}{\pi}\arctan\left[\left(\frac{1-x_{(i)}}{x_i}\right)^{-\zeta}\right]\right) - \frac{2i-1}{2v}\right]^2,$$

with respect to ζ.

5.4 Anderson-Darling estimation

If $x_{(1)}, x_{(2)} \dots, x_{(v)}$ are the ordered observations based on the LG distribution, then the Anderson-Darling (ADE) estimate is obtained by minimizing the function

$$ADE = -v - \frac{1}{v}\sum_{i=1}^{v}(2i-1)[\log F_X(x_{(i)}|\zeta) + \log(1 - F_X(x_{(i)}|\zeta))],$$

with respect to ζ.

5.5 Percentile estimation

If $x_{(1)}, x_{(2)}, \dots, x_{(v)}$ are the ordered observations from the LG distribution and $q_i = i/(v+1)$ is an unbiased estimator of $F_X(x_{(i)}|\zeta)$, then the PE estimate of the LG distribution's parameter can be derived by minimizing the function

$$PE = \sum_{i=1}^{v}\left[x_{(i)} - \left(1 + \left(\tan\left[\frac{\pi q_i}{2}\right]\right)^{-1/\zeta}\right)^{-1}\right]^2,$$

with respect to ζ.

6 Parametric LG regression

If a response variable follows the LG distribution, a new parametric LG regression (PLGR) model can be obtained by linking the parameter $\zeta > 0$ to a set of covariates using the logarithmic link function. Thus, $\zeta_i = \exp(\boldsymbol{z_i'}\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2 ..., \theta_k)'$ is the vector of the coefficients of the covariates, $\boldsymbol{z_i'} = (1, z_{i1}, z_{i2} ..., z_{ik})$ and $i = 1, 2, 3..., n$. The PDF of the PLGR model is given as

$$f_X(x) = \frac{2\exp(\boldsymbol{z_i'}\boldsymbol{\theta})x^{\exp(\boldsymbol{z_i'}\boldsymbol{\theta})-1}(1-x)^{-\exp(\boldsymbol{z_i'}\boldsymbol{\theta})-1}}{\pi\left[1 + x^{2\exp(\boldsymbol{z_i'}\boldsymbol{\theta})}(1-x)^{-2\exp(\boldsymbol{z_i'}\boldsymbol{\theta})}\right]}.$$

The log-likelihood function is then given by

$$\begin{aligned}\ell &= n\log(2/\pi) + \sum_{i=1}^{n}\log\zeta_i + \sum_{i=1}^{n}(\zeta_i - 1)\log x_i - \sum_{i=1}^{n}(\zeta_i + 1)\log(1 - x_i) \\ &\quad - \sum_{i=1}^{n}\log[1 + x_i^{2\zeta_i}(1 - x_i)^{-2\zeta_i}].\end{aligned}$$

The parameter estimates of the model are derived by maximizing the log-likelihood function directly.

7 Simulation study

The simulation results are presented and interpreted to find the most appropriate estimation method. The simulation studies are carried out using different parameter values of the LG distribution to generate random observations with the help of the quantile function. The simulations are conducted using R software and results presented in Tables 1, 2 and 3 for three set of parameter values.

Table 1, 2 and 3 displays simulated results for the different estimators using the MSEs and ABs. For all the estimation techniques, it can be shown that the ABs and MSEs values generally decrease as the sample size increases. Even though the experiment was performed 1,000 times, the simulation results are comparable, demonstrating the estimators' consistency and efficacy. However, it can be noticed that the MLE method was the best and more efficient in estimating parameter ζ for the LG distribution, with the lowest MSEs and ABs as sample sizes increases compared with the other estimators except at sample size $n = 300$ in Table 1. In addition, the ABs and MSEs of MLE for all parameter values tend to decrease faster compared with the other estimators. Hence, the MLE method was the best for the parameter ζ.

Table 1: Simulated MSE and AB values for $\zeta = 1.2$.

	Method	MLE	LSE	WLSE	CVME	ADE	PE
	MSEs	0.0024	0.0461	0.0481	0.0428	0.0468	0.0154
$n = 30$	ABs	0.0603	0.1615	0.1675	0.1568	0.1654	0.0980
	Sum of Ranks	2	8	12	6	10	4
	MSEs	0.0015	0.0954	0.0141	0.0149	0.0148	0.0842
$n = 80$	ABs	0.0099	0.0152	0.0928	0.0960	0.0951	0.2418
	Sum of Ranks	2	8	5	9	7	11
	MSEs	0.0006	0.0061	0.0060	0.0056	0.0061	0.0059
$n = 200$	ABs	0.0120	0.0619	0.0617	0.0597	0.0619	0.0617
	Sum of Ranks	2	12	8	4	10	6
	MSEs	0.0038	0.0036	0.0039	0.0041	0.0036	0.0041
$n = 300$	ABs	0.0194	0.0486	0.0498	0.0506	0.0478	0.0507
	Sum of Ranks	4	5	8	11	3	11
	MSEs	0.0026	0.0028	0.0030	0.0030	0.0424	0.0032
$n = 400$	ABs	0.0018	0.0052	0.0442	0.0426	0.0029	0.0445
	Sum of Ranks	2	5	9	7	7	11
	MSEs	0.0012	0.0021	0.0357	0.0021	0.0022	0.0019
$n = 550$	ABs	0.0017	0.0363	0.0020	0.0371	0.0380	0.0351
	Sum of Ranks	2	7	8	9	11	5
	MSEs	0.0011	0.0018	0.0018	0.0019	0.0018	0.0019
$n = 650$	ABs	0.0040	0.0336	0.0331	0.0343	0.0340	0.0346[6]
	Sum of Ranks	2	7	4	10	7	12

8 Application of the LG distribution

The LG distribution is fitted to two lifetime datasets to ascertain its applicability. The fit of the LG distribution is compared with five (5) competitive distributions namely; Kumaraswamy (Kumaraswamy, 1980), unit Weibull (Mazucheli et al., 2018), unit-Marshall-Olkin extended exponential (UMOEE) (Ghosh et al., 2019), unit Gompertz (Mazucheli et al., 2019) and Topp-Leone (Topp and Leone, 2021) distributions.

Table 2: Simulated MSE and AB values for $\zeta = 5.5$.

	Method	MLE	LSE	WLSE	CVME	ADE	PE
	MSEs	1.0080	1.0232	1.0458	1.3664	1.9937	1.0118
$n = 30$	ABs	0.2996	0.7827	0.8102	0.4717	0.7887	0.3007
	Sum of Rank	2	5	7	4	7	3
	MSEs	0.3505	0.3554	0.3615	1.9380	1.2939	0.3543
$n = 80$	ABs	0.1692	0.4709	0.4760	1.4483	0.8884	0.1786
	Sum of Rank	2	6	8	12	10	4
	MSEs	0.1331	0.1378	0.1410	0.3053	0.3398	0.1360
$n = 200$	ABs	0.1404	0.2977	0.2979	0.9053	0.5347	0.1542
	Sum of Rank	2	6	8	10	12	4
	MSEs	0.1296	0.2411	0.2155	0.3292	0.3597	0.1490
$n = 300$	ABs	0.1149	0.3898	0.2325	0.4562	0.2478	0.2314
	Sum of Rank	2	9	6	11	10	4
	MSEs	0.1071	0.3649	0.2727	0.1614	0.2988	0.1081
$n = 400$	ABs	0.1089	0.2035	0.2184	0.3416	0.4362	0.1109
	Sum of Rank	2	8	8	8	11	4
	MSEs	0.0220	0.0458	0.0496	0.1821	0.1734	0.0353
$n = 550$	ABs	0.0919	0.1706	0.1773	0.3411	0.3361	0.3329
	Sum of Rank	2	5	7	12	10	6
	MSEs	0.0210	0.0417	0.0419	0.1379	0.1354	0.1117
$n = 650$	ABs	0.0612	0.1651	0.1648	0.2943	0.2917	0.1090
	Sum of Rank	2	6	6	12	10	6

Table 3: Simulated MSE and AB values for $\zeta = 10.8$.

	Method	MLE	LSE	WLSE	CVME	ADE	PE
	MSE	3.0459	3.2967	3.7686	3.8976	3.3470	3.5036
$n = 30$	AB	1.4262	1.5142	1.4826	1.6600	1.4288	1.4508
	Sum of Ranks	2	7	9	12	5	7
	MSE	0.5494	1.2393	1.1335	3.4347	1.2939	1.2408
$n = 80$	AB	0.4989	0.8841	0.8479	1.4483	0.8884	0.8621
	Sum of Rank	2	7	4	12	10	7
$n = 200$	MSE	0.4021	0.4778	0.5016	1.3053	0.4540	0.4306
	AB	0.4001	0.5437	0.5524	0.9053	0.5347	0.5210
	Sum of Rank	2	8	10	12	6	4
	MSE	0.3001	0.3107	0.3225	0.3292	0.3597	0.3348
$n = 300$	AB	0.4012	0.4339	0.4498	0.4558	0.0478	0.4564
	Sum of Rank	3	5	7	9	7	11
	MSE	0.2019	0.2239	0.2375	0.2614	0.2988	0.2496
$n = 400$	AB	0.3001	10.8266	0.3857	0.4052	0.4362	0.3978
	Sum of Rank	2	8	5	9	11	7
	MSE	0.0012	0.1723	0.1737	0.1821	0.1734	0.1700
$n = 550$	AB	0.2330	0.3304	0.3314	0.3411	0.3361	0.3329
	Sum of Rank	2	5	8	12	9	6
	MSE	0.0010	0.1465	0.1624	0.1379	0.1354	0.1417
$n = 650$	AB	0.2330	0.3037	0.3199	0.2943	0.2917	0.2979
	Sum of Rank	2	10	12	6	4	8

8.1 Trade share data

In this first application, we consider economic growth data taken from Stock and Watson (2003) and used by Hassan et al. (2021). The dataset contains values of trade shares of 61 different countries. Table 4 displays the descriptive statistics of the trade share dataset, mean ($\bar{x}$), median (M), standard deviation (SD), coefficient of skewness (CK) and coefficient of kurtosis (CK). It can be observed that the trade share data are approximately symmetric with CK of 0.0061 and leptokurtic with excess kurtosis of 2.5528. The data also depicts a low standard deviation. Figure 3 shows the boxplot and time total on test (TTT) plots. The TTT plots for the dataset show a concave characteristic that indicates increasing failure intensity. The boxplot depicts slightly skewed to the left with no extreme value.

Table 4: Descriptive statistics of trade share data.

n	Minimum	Maximum	$\bar{x}$	M	SD	CS	CK
61	0.1405	0.9794	0.5142	0.5278	0.1935	0.0061	2.5528

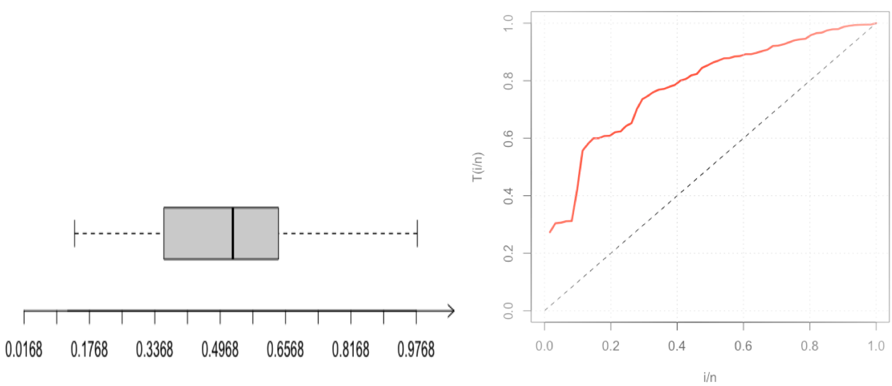


Figure 3: Boxplots (left) and TTT plots (right) of the trade share dataset.

Table 5: Estimates and standard errors connected to model parameters for the trading share dataset.

Distribution	Parameters	Estimates	Standard Error
LG	ζ	1.6292	0.1882
Kumaraswamy	a	2.3294	0.3055
	b	2.7624	0.5549
UMOEE	a	0.5549	4.1115
	b	3.3844	0.5056
Unit Gompertz	a	0.6162	0.2660
	b	1.0921	0.2471
Unit Weibull	a	1.3396	0.1726
	b	1.7345	0.1695
Topp-Leone	a	2.7392	0.3507

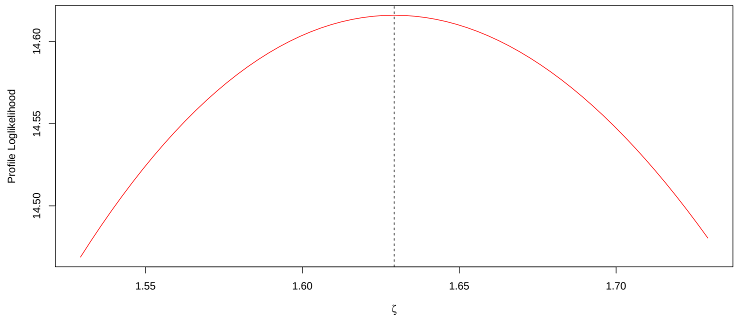


Figure 4: The profile log-likelihood plot for trade share dataset.

Table 5 shows the MLEs of the parameters together with standard errors of the distributions. Using the profile of the log-likelihood function in Figure 4 we ascertain the estimated parameter as a true maxima. From Table 6, the LG distribution provides the best fit as the LG distribution has the highest log-likelihood (L) value, lowest Akaike information criterion (AIC), Bayesian information criterion (BIC) and consistent AIC (AICc) together with the highest p-values of the Kolmogorov-Smirnov (KS), Cramér-von Mises (CV) and Anderson-Darling (AD) tests. Consequently, we consider the LG distribution an appropriate alternative to the other competing models for the given data. The appropriateness and applicability of the fitted models are visualised in Figure 5. The fitted models' estimated PDF and CDF suggests that the LG model offers the best fit.

Table 6: Selection criteria for the trade share dataset.

Distribution	L	AIC	AICc	BIC	AD	CV	KS
LG	14.616	-27.232	-27.164	-25.121	0.385	0.056	0.066
					(0.863)	(0.844)	(0.940)
Kumaraswamy	13.622	-23.243	-23.036	-19.021	0.4121	0.056	0.069
					(0.836)	(0.842)	(0.914)
UMOEE	13.811	-23.622	-23.415	-19.401	0.545	0.058	0.068
					(0.700)	(0.832)	(0.926)
Unit Gompertz	10.878	-17.751	-17.544	-13.530	1.447	0.208	0.110
					(0.190)	(0.253)	(0.424)
Unit Weibull	14.242	-24.484	-24.277	-20.262	22.009	4.995	22.009
					(3.7×10^{-10})	(2.7×10^{-16})	(9.3×10^{-6})
Topp-Leone	13.918	-25.836	-25.768	-23.725	0.593	0.088	0.089
					(0.654)	(0.649)	(0.726)

8.2 Milk production data

The milk production dataset contains transformed total milk yield of 107 cows belonging to the SINDI breed during their first lactation cycle. The data was used by Yousof

(a) (b)

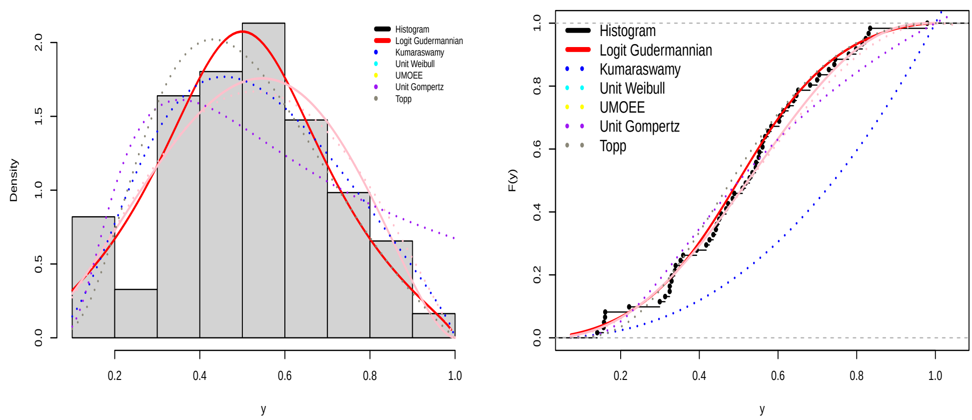


Figure 5: Fitted PDFs (a) and fitted CDFs (b) of the trade share dataset.

et al. (2017), Nasiru et al. (2021) and Bhatti et al. (2021). From Table 7, the total milk production dataset is negatively skewed (-0.3353) with a positive excess kurtosis (2.6861).

Table 7: Descriptive statistics of Milk production dataset.

n	Minimum	Maximum	$\bar{x}$	M	SD	CS	CK
107	0.0168	0.8781	0.4689	0.4741	0.1920	-0.3353	2.6861

From Figure 6, it can be observed that the TTT curve is concave with increasing failure rate while boxplot also indicates approximately symmetric data.

The estimates of the parameter(s) of the fitted models with their corresponding standard errors are displayed in Table 8.

Table 8: Estimates and standard errors associated with the model parameters for milk production dataset.

Distribution	Parameters	Estimates	Standard Error
LG	ζ	1.6242	0.1418
Kumaraswamy	a	2.3294	0.3055
	b	2.7624	0.5549
UMOEE	a	9.2403	4.1115
	b	3.3844	0.5056
Unit Gompertz	a	2.1193	0.8683
	b	0.3878	0.1145
Unit Weibull	a	0.9846	0.1015
	b	1.5620	0.1036
Topp-Leone	a	2.0802	0.2011

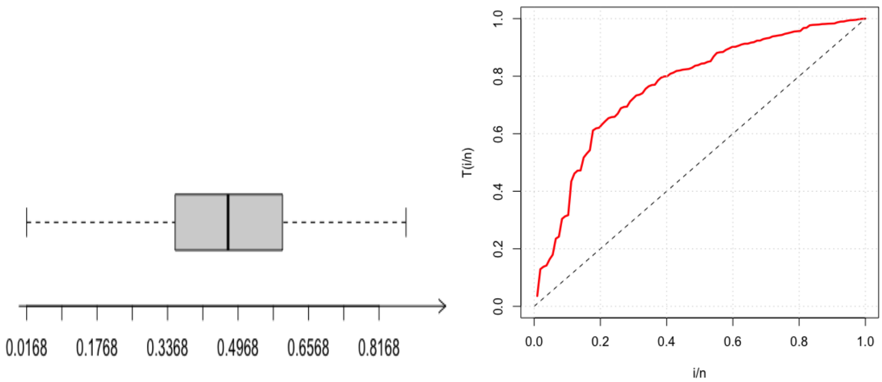


Figure 6: Boxplots (left) and TTT plot (right) of the Milk production dataset.

Using the profile of the log-likelihood function in Figure 7 we ascertain the estimated parameter as a true maxima.

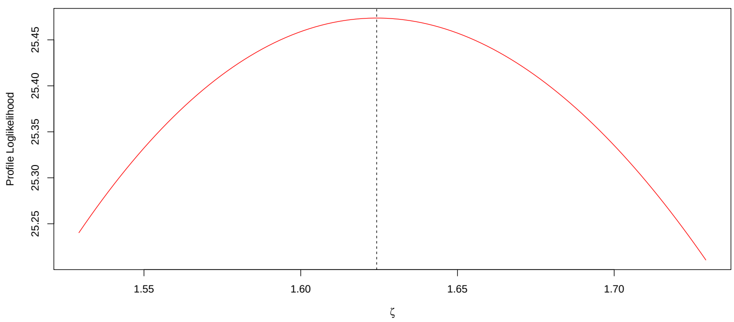


Figure 7: The profile log-likelihood plot for the milk production dataset.

In Tables 9, we observe that the LG distribution can be considered as the best model for the Milk production dataset as the values of AIC, AICc and BIC statistics are smaller for the proposed distribution. Also, the p-values of the AD, CV and KS are highest for the proposed distribution.

Figure 8 suggests that LG distribution provides competitive fit for the milk production dataset compared to the competing models.

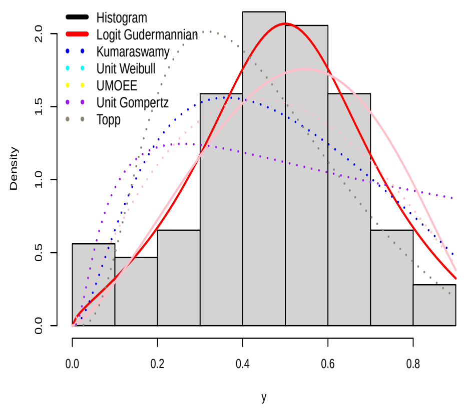
8.3 Percentage of estimated eligible uninsured individuals

The performance of the parametric LG regression is assessed using real dataset. The LG regression model is fitted and compared with the parametric unit Monsef regression (PUMR) model formulated from the unit Monsef model (El-Monsef et al., 2016), the

Table 9: Selection criteria and goodness-of-fit test values for the Milk production dataset.

Distribution	L	AIC	AICc	BIC	AD	CV	KS
LG	25.417	-48.947	-48.909	-46.275	1.418 (0.197)	0.162 (0.356)	0.073 (0.624)
Kumaraswamy	25.395	-46.789	-46.674	-41.444	3.718 (0.012)	0.580 (0.025)	0.076 (0.563)
UMOEE	15.026	-26.097	-25.982	-20.751	2.574 (0.045)	0.333 (0.110)	0.105 (0.190)
Unit Gompertz	5.489	-6.977	-6.862	-6.632	5.796 (0.001)	0.972 (0.003)	0.184 (0.002)
Unit Weibull	16.921	-29.842	-29.727	-24.497	19.926 (5.6×10^{-6})	4.231 (2.2×10^{-16})	0.3379 (4.8×10^{-14})
Topp-Leone	21.526	-41.052	-41.014	-38.380	4.327 (0.006)	0.726 (0.001)	0.179 (0.035)

(a)



(b)

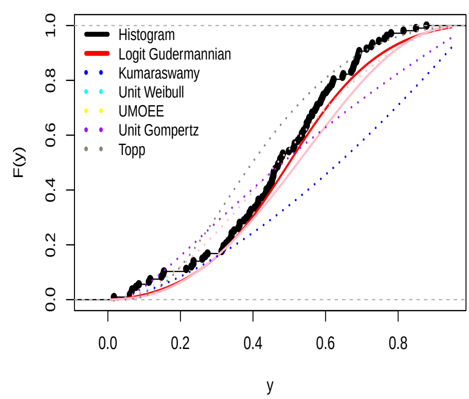


Figure 8: Fitted PDFs (a) and fitted CDFs (b) for the milk production dataset.

parametric unit Lindley regression (PULR) model formulated from the unit Lindley model (Mazucheli et al., 2019) and the parametric Topp-Leone regression (PTLR) model formulated from the Topp-Leone model (Topp and Leone, 2021). The PUMR, PULR and PTLR are formulated by linking their respective parameters to a set of covariates using a logarithmic link function. The PDF of the PUMR, PULR and PTLR are given respectively as

$$f_X(x) = \frac{(\exp(z_i'\theta))^3 \exp\left(\frac{x}{x-1}\exp(z_i'\theta)\right)}{(x-1)^4[2+\exp(z_i'\theta)(2+\exp(z_i'\theta))]},$$

$$f_X(x) = \frac{(\exp(z_i^{'}\theta))^2 \exp\left(\frac{x}{x-1}\exp(z_i^{'}\theta)\right)}{(1-x)^3(1+\exp(z_i^{'}\theta))},$$
$$f_X(x) = 2\exp(z_i^{'}\theta)x^{\exp(z_i^{'}\theta)-1}(1-x)(2-x)^{\exp(z_i^{'}\theta)-1}.$$

The data was retrieved from https://data.world/johnsnowlabs/percentage-of-estimated-eligible-uninsured-people. The interest is in the proportion of estimated eligible uninsured individuals in ten (10) states in the USA with family income at or below 138% of the federal poverty level (fpl) and associated factors such as proportion of uninsured males (z_1) , proportion of uninsured females (z_2) and proportion of uninsured people in families with at least one working adult aged 18 to 64 (z_3). Table 10 shows the estimates of the PLGR, PUMR, PULR and PTLR models. The results obtained shows that the PLGR model provided the best fit to the dataset. Substituting the parameter estimates obtained into the PLGR model we obtain

$$fpl = 6.2539 - 7.7151z_1 - 5.6538z_2 + 2.4005z_3.$$

Table 10: Parameter estimates of the fitted models with corresponding selection criteria.

Model	Parameter	Parameter Estimates	P-values	
PLGR	θ_0	6.254	6.946×10^{-9}	$2\ell = -15.998$
	θ_1	-7.715	0.001	$AIC = -7.998$
	θ_2	-5.654	0.019	$BIC = -6.788$
	θ_3	2.401	0.269	
PUMR	θ_0	5.385	1.043×10^{-7}	$2\ell = -12.419$
	θ_1	-7.511	1.824×10^{-8}	$AIC = -4.419$
	θ_2	-6.721	1.461×10^{-5}	$BIC = -3.209$
	θ_3	3.106	0.130	
PULR	θ_0	4.791	1.499×10^{-5}	$2\ell = 44.907$
	θ_1	-7.689	6.449×10^{-9}	$AIC = 52.907$
	θ_2	-6.741	2.198×10^{-5}	$BIC = 54.118$
	θ_3	4.536	0.040	
PTLR	θ_0	8.928	9.632×10^{-10}	$2\ell = -12.179$
	θ_1	-3.281	0.147	$AIC = -4.179$
	θ_2	-4.578	0.061	$BIC = -2.969$
	θ_3	-4.99	0.0914	

The predictive ability of the fitted models is measured using the Cox-Snell residuals. Figure 9 shows the probability-probability (P-P) plots of the Cox-Snell residuals compared with the standard exponential distribution. Compared with other fitted models, the PLGR model's residuals are considerably closer to the diagonal line, suggesting a better fit for the dataset. The goodness-of-fit statistics of the residuals in Table 11 used in the model diagnostics show that the PLGR model provides a better fit to the dataset. The log-likelihood plots of the regression parameters are presented in Figure 10. The profile confirm the estimated parameters as real maxima.

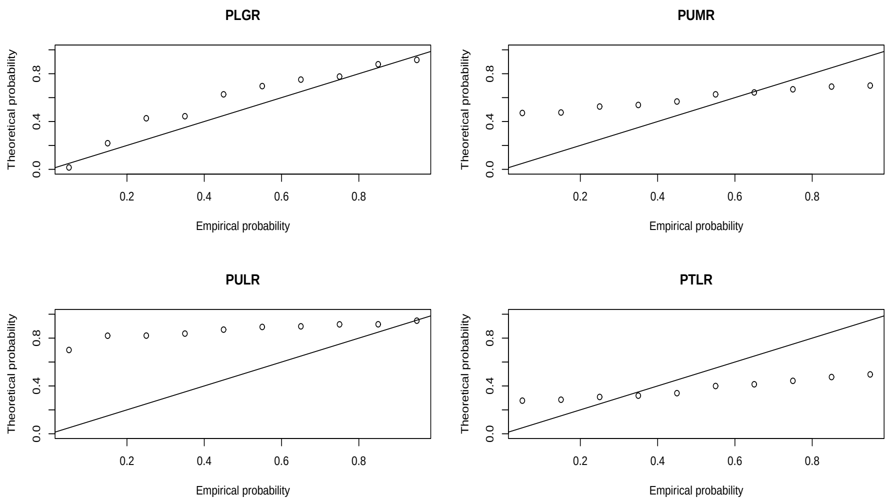


Figure 9: P-P plot of Cox-Snell residuals.

Table 11: Model diagnostics results.

Model	KS statistic	AD statistic	CV statistic
PLGR	0.2274	0.6329	0.1204
PUMR	0.4714	2.5433	0.5165
PULR	0.7202	9.3904	1.8343
PTLR	0.5035	2.9229	0.6140

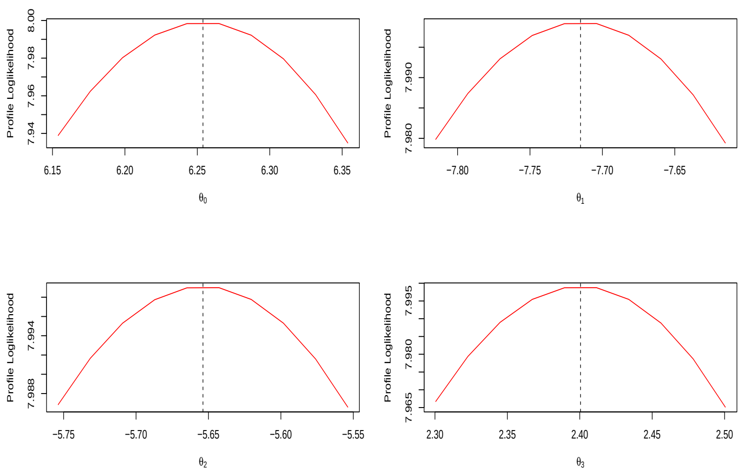


Figure 10: The log-likelihood profiles for the regression parameters.

9 Conclusions

The LG distribution and its properties are analyzed in this study. The LG distribution is found to be suitable for modeling datasets with approximately symmetric and bathtub-shaped PDF plots, as well as bathtub and upside-down bathtub-shaped failure rate functions. Six estimation methods were adopted to find the parameter estimates and a simulation study conducted to assess their consistency. The LG distribution was compared to five other distributions and was found to provide the best fit to two lifetime datasets. Using the LG distribution, a parametric LG regression model was proposed. When the proposed regression model was applied to a real-life dataset it provided the best fit than some other existing regression models.

Acknowledgement

The authors want to say thank you to the anonymous reviewers for their helpful comments that improved the work.

References

Altun, E. (2019). The generalized Gudermannian distribution: inference and volatility modelling. *Statistics*, **53**(2):364–386.

Bantan, R.A., Chesneau, C., Jamal, F., Elgarhy, M., Tahir, M.H., Ali, A., Zubair, M. and Anam, S. (2020). Some new facts about the unit-Rayleigh distribution with applications. *Mathematics*, **8**(11):1954.

Bhatti, F.A., Ali, A., Hamedani, G., Korkmaz, M.Ç. and Ahmad, M. (2021). The unit generalized log Burr XII distribution: Properties and application. *AIMS Mathematics*, **6**(9):10222–10252.

Chesneau, C. (2021). Study of a unit power-logarithmic distribution. *Open Journal of Mathematical Sciences*, **5**(1):218–235.

Cordeiro, G.M. and Lemonte, A.J. (2014). The McDonald arcsine distribution: A new model to proportional data. *Statistics*, **48**(1):182–199.

El-Monsef, A., Sohsah, N.M. and Hassanein, W.A. (2021). Unit Monsef distribution with regression model. *Asian Journal of Probability and Statistics*, **15**(4):330–340.

Ghosh, I., Dey, S. and Kumar, D. (2019). Bounded MO extended exponential distribution with applications. *Stochastics and Quality Control*, **34**(1):35–51.

Haj Ahmad, H., Almetwally, E.M., Elgarhy, M. and Ramadan, D.A. (2023). On unit exponential pareto distribution for modeling the recovery rate of COVID-19. *Processes*, **11**(1):232.

Hassan, A.S., Fayomi, A., Algarni, A. and Almetwally, E.M. (2022). Bayesian and non-Bayesian inference for unit-exponentiated half-logistic distribution with data analysis. *Applied Sciences*, **12**(21):11253.

Korkmaz, M.Ç. and Chesneau, C. (2021). On the unit Burr-XII distribution with the quantile regression modeling and applications. *Computational and Applied Mathematics*, **40**(1):29.

Korkmaz, M.Ç., Chesneau, C. and Korkmaz, Z.S. (2021). Transmuted unit Rayleigh quantile regression model: Alternative to beta and Kumaraswamy quantile regression models. *University Politehnica of Bucharest Scientific Bulletin Series Applied Mathematics and Physics*, **83**(3):149–158.

Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of hydrology*, **46**(1-2):79–88.

Mazucheli, J., Menezes, A.F.B. and Chakraborty, S. (2019). On the one parameter unit-Lindley distribution and its associated regression model for proportion data. *Journal of Applied Statistics*, **46**(4):700–714.

Mazucheli, J., Menezes, A.F. and Dey, S. (2018). The unit-Birnbaum-Saunders distribution with applications. *Chilean Journal of Statistics*, **9**(1):47–57.

Mazucheli, J., Menezes, A.F. and Dey, S. (2019). Unit-Gompertz distribution with applications. *Statistica*, **79**(1):25–43.

Nasiru, S., Abubakari, A.G. and Angbing, I.D. (2021). Bounded odd inverse pareto exponential distribution: Properties, estimation, and regression. *International Journal of Mathematics and Mathematical Sciences*, **2021**:9955657.

Nasiru, S., Abubakari, A.G. and Chesneau, C. (2022). New lifetime distribution for modeling data on the unit interval: Properties, Applications and Quantile regression. *Mathematical and Computational Applications*, **27**(6), 105.

Stock, J.H. and Watson, M.W. (2003). *Introduction to Econometrics.* New York: Prentice Hall.

Topp, C.W. and Leone, F.C. (1955). A family of J-shaped frequency functions. *Journal of the American Statistical Association*, **50**(269):209–219.

Yousof, H.M., Alizadeh, M., Jahanshahi, S.M.A., Ramires, T.G., Ghosh, I. and Hamedani, G.G. (2017). The transmuted Topp-Leone G family of distributions: theory, characterizations and applications. *Journal of Data Science*, **15**(4):723–740.

Zayed, M.A., Hassan, A.S., Almetwally, E.M., Aboalkhair, A.M., Al-Nefaie, A.H. and Almongy, H.M. (2023). A compound class of unit Burr XII model: Theory, Estimation, Fuzzy, and Application. *Scientific Programming*, **2023**:4509889.