CHILEAN JOURNAL OF STATISTICS

Edited by Víctor Leiva and Carolina Marchant

A free open-access journal indexed by

Volume 13 Number 1
April 2022
ISSN: 0718-7912 (print)
ISSN: 0718-7920 (online)

Published by the
Chilean Statistical Society
SOCHE
SOCIEDAD CHILENA DE ESTADÍSTICA

Aims

The Chilean Journal of Statistics (ChJS) is an official publication of the Chilean Statistical Society (`www.soche.cl`). The ChJS takes the place of *Revista de la Sociedad Chilena de Estadística*, which was published from 1984 to 2000.

The ChJS covers a broad range of topics in statistics, as well as in artificial intelligence, big data, data science, and machine learning, focused mainly on research articles. However, review, survey, and teaching papers, as well as material for statistical discussion, could be also published exceptionally. Each paper published in the ChJS must consider, in addition to its theoretical and/or methodological novelty, simulations for validating its novel theoretical and/or methodological proposal, as well as an illustration/application with real data.

The ChJS editorial board plans to publish one volume per year, with two issues in each volume. On some occasions, certain events or topics may be published in one or more special issues prepared by a guest editor.

Contents

Chilean Journal of Statistics
Volume 13, Number 1 (April 2022), Pages 67–90
DOI: 10.32372/ChJS.13-01-04

Statistical Modeling
Research Paper

On the Topp-Leone log-normal distribution: Properties, modeling, and applications in astronomical and cancer data

Christophe Chesneau[1,*], Muhammed Rasheed Irshad[2], Damodaran Santhamani Shibu[3], Soman Latha Nitin[3], and Radhakumari Maya[4]

[1]Department of Mathematics, Université de Caen Basse-Normandie, Caen, France,
[2]Department of Statistics, Cochin University of Science and Technology, Kerala, India,
[3]Department of Statistics, University College, Kerala, India,
[4]Department of Statistics, University of Kerala, Kerala, India

(Received: 03 January 2022 · Accepted in final form: 16 February 2022)

Abstract

In the realm of astronomy, the two-parameter log-normal distribution has ominous implications. In this article, we propose a new version of the two-parameter log-normal distribution with an application to astronomical data. More precisely, a new modulating parameter is added to the two-parameter log-normal distribution through the use of the Topp-Leone generator of distributions. The moments, quantile function, several reliability measures, and other significant aspects of the proposed distribution are investigated. The maximum likelihood approach and a Bayesian technique are both utilized to estimate the unknown parameters. In addition, we present a parametric regression model and a Bayesian regression method. A simulation study is carried out to assess the long-term performance of the estimators of the distribution parameters. Two real datasets are employed to show the applicability of this new distribution. The efficiency of the newly added parameter is tested by utilizing the likelihood ratio test. The parametric bootstrap approach is also utilized to determine the adequacy of the suggested model for the datasets.

Keywords: Bayesian estimation · bootstrapping · maximum likelihood estimation · regression · simulation.

Mathematics Subject Classification: Primary 60E05 · Secondary 62F15.

1. Introduction

In practice, the two-parameter log-normal (LN) distribution can be used to fit empirical data in a variety of ways. This is especially true in the field of astronomy. Studies and research have established evidence of an LN distributional characteristic for very high energy emission of light curves from galaxies; for more information, see Abdalla et al. (2017). It is also worth noting that the distribution of galaxy density contrast, which is a parameter utilized in galaxy formation to indicate where there are local enhancements in matter

*Corresponding author. Email: christophe.chesneau@unicaen.fr, christophe.chesneau@gmail.com

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

density, is roughly an LN distribution; whether the distribution of mass fluctuation from the Dark Energy Survey, which is derived from weak lensing convergence in a similar way to convex glass lenses, is an LN distribution is less clear. It was first identified by Hubble (1934) that the distribution of galaxies in angular cells on the celestial sphere is well predicted by an LN distribution. Again, recently, Shah et al. (2018) and Shah et al. (2020) elaborately highlighted the considered LN distributional behavior of the gamma-ray (γ-ray) flux distribution on the brightest blazars, which are observed by the Fermi-LAT, a space observatory's large area telescope (LAT) being used to perform γ-ray. For more applications of the LN distribution in the area of astrophysics and cosmology, one can go through the articles by Bernardeau and Kofman (1994), Blasi et al. (1999) and Parravano et al. (2012).

Fundamental distributions occasionally fail to adequately characterize and anticipate the vast majority of real-world datasets resulting from complicated processes. Because the quality of statistical analysis results is strongly dependent on the assumed model, choosing an adaptive model for data analysis is critical. Therefore, more allied distributions must be found in order to obtain better quality and more accurate results. Since the LN distribution has superior importance in the field of astronomy, it is inevitable to derive new generalized versions of the LN distribution, not only for modeling astronimocal data but also for the variety of datasets from other study areas where the LN distribution has the best fit. Note that the LN distribution has been utilized in a range of domains which includes most of the applied areas such as economics, sociology, biology, and meteorology, to name just a few; for more details, see Jobe et al. (1989).

There has recently been a boom in interest in the art of adding parameters to well-known existing distributions in order to obtain diverse forms of hazard rate functions (HRFs) for use in various real-life circumstances, as well as for evaluating data with a high degree of skewness and kurtosis. Several researchers have started to build families of distributions based on conventional distributions or using different methodologies in order to generalize any baseline distribution; for example, see Afify (2017). In this article, using a flexible generalization technique that includes an additional shape parameter, we investigate a novel lifetime distribution that is also a generalized version of the two-parameter LN distribution. The aim is to uncover some of the suggested model's statistical features and apply them to real-world data. The main motivations for developing this lifetime model are: (i) to propose a new flexible version of the LN distribution that can be used, particularly to analyze astronomical data, because the LN distribution has eminent superiority in the field of astronomy, as well as the ability to be applied to a broader class of reliability problems, (ii) to extend both the LN and Topp-Leone distributions, and (iii) to investigate some additional shapes of the HRF.

The remaining sections of the article are structured as follows. Section 2 reveals our distribution methodology. The specification of the new distribution is presented in Section 3. In Section 4, its moments are calculated. The quantile function (QF) and some of its associated measures are obtained in Section 5. The various functions and moments related to the reliability measures are discussed in Section 6. In Section 7, the maximum likelihood (ML) and Bayesian estimation techniques are employed to estimate the unknown parameters of the new model. Also, a parametric bootstrap method of simulation using the ML estimates is presented in Section 8. A parametric regression model associated with the new distribution is defined in Section 9. Again, a Bayesian regression method is presented in Section 10. A simulation study is proposed in Section 11 to analyze the performance of the ML estimators of the parameters. In Section 12, one univariate uncensored real dataset based on an astronomical study, and one censored real dataset based on a cancer study are evaluated to depict the potential of the new distribution over competing distributions. The final concluding remarks are given in Section 13.

2. Construction of the distribution

A simple bounded J-shaped distribution that has attracted various statisticians as an alternative to uniform(0,1) and beta distributions was proposed by Topp and Leone (1955). It is called the Topp-Leone (TL) distribution. The cumulative distribution function (CDF) and probability density function (PDF) of the TL distribution are respectively stated as

$$F_{\mathrm{TL}}(x;\alpha) = [x(2-x)]^{\alpha},$$

and

$$f_{\mathrm{TL}}(x;\alpha) = 2\alpha x^{\alpha-1}(1-x)(2-x)^{\alpha-1}, \quad \alpha > 0,\ x \in (0,1).$$

It is worth mentioning that the TL distribution has a bathtub shaped HRF for all $\alpha < 1$. Later, Sangsanit and Bodhisuwan (2016) introduced the Topp-Leone generalized exponential distribution, using the TL distribution as a generator distribution with application to the maximum stress per cycle and breaking stress of carbon fiber datasets. Now, we consider the method for generating new distributions, called the TX family, proposed by Alzaatreh et al. (2013). The essence of the TX family is presented below. Let X be a continuous baseline random variable with CDF F_X, and T be a continuous generator random variable of a distribution with support on $[a,b]$ and CDF Ψ. Then, the CDF of the TX family is given by

$$F_{\mathrm{TX}}(x) = \Psi[W(F_X(x))], \tag{1}$$

where $W(F_X(x)) \in [a,b]$ is differentiable and monotonically non-decreasing.

Considering the immense applicability of the TL and LN distributions, we propose to apply both the distributions in Equation (1), in which the LN distribution is the baseline and the TL distribution is a generator distribution, and henceforth, we call the resulting distribution the Topp-Leone log-normal (TLLN) distribution, which provides greater versatility in modeling skewed datasets.

We also propose an entirely different method to derive the new distribution. Sharma (2018) proposed a new three-parameter distribution called the Topp-Leone normal (TLN), which is defined on the entire real line and is ideal for modeling increasing HRF data. The CDF of the TLN distribution is expressed as

$$F_{\mathrm{TLN}}(y) = \left\{\Phi\left(\frac{y-\mu}{\sigma}\right)\left[2-\Phi\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^{\alpha}, \quad y,\mu \in \mathbb{R},\ \sigma,\alpha > 0,$$

where Φ is the CDF of the standard normal distribution. Then, the random variable $X = \mathrm{e}^{Y}$ follows the TLLN distribution with parameters α, μ and σ.

3. Definition of the distribution

The definition of the new distribution, as well as several key features, are examined in this section.

Definition 3.1 Let X be a random variable which follows the TLLN distribution with parameters α, μ and σ. Then, its CDF is given by

$$F(x) = \left\{\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left[2-\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]\right\}^{\alpha}, \tag{2}$$

and its PDF is defined as

$$f(x) = \frac{2\alpha}{\sigma x}\,\phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left[1-\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]\times \\ \left\{\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left[2-\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]\right\}^{\alpha-1}, \tag{3}$$

where $x > 0$, $\mu \in \mathbb{R}$ and $\sigma, \alpha > 0$. Also, ϕ is the PDF of the standard normal distribution.

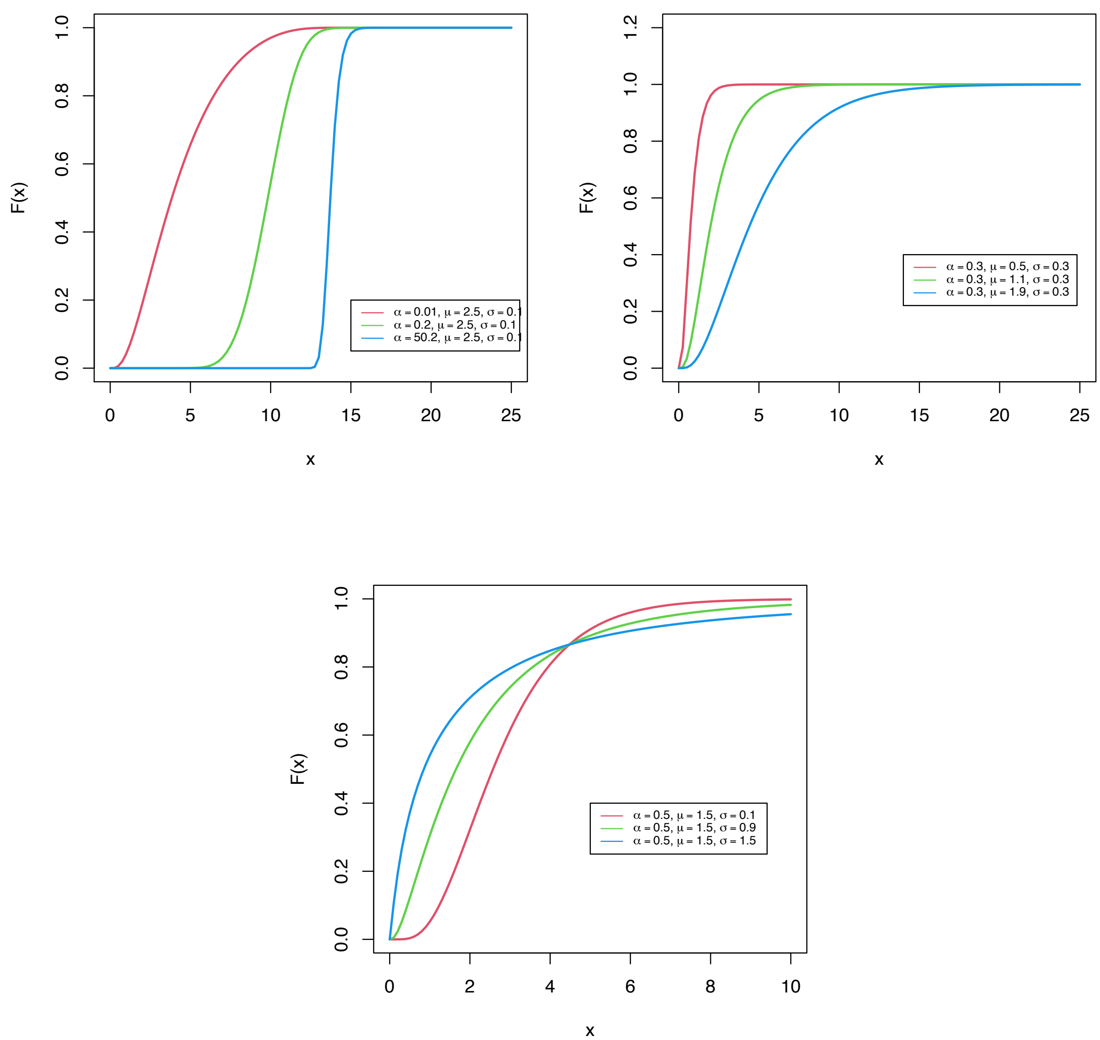

Figure 1. Plots of the CDF of the TLLN distribution.

The plots in Figures 1 and 2 depict the corresponding CDF and PDF of the TLLN distribution. We observe that the PDF may be decreasing and unimodal with a certain flexibility in the mode and tails. It is, however, mainly right-skewed or almost symmetrical. Next, some expansions for the CDF and PDF are provided. It is also interesting to note that the TLLN distribution can be expressed as an infinite sum of exponentiated LN distributions when α is a non-integer or as a finite sum when α is an integer. Indeed, the CDF of the

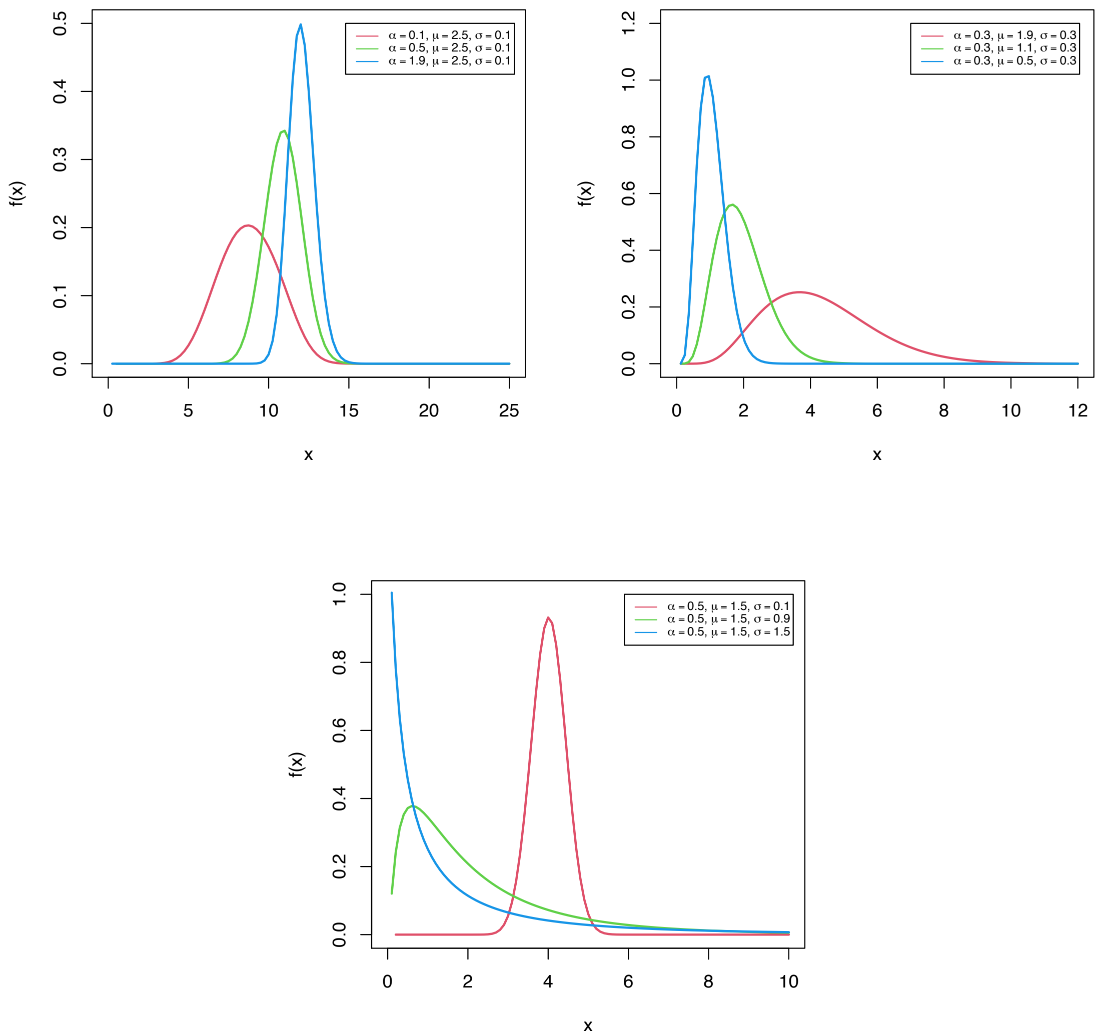

Figure 2. Plots of the PDF of the TLLN distribution.

TLLN distribution in Equation (2) can be simplified as follows:

$$F(x) = \sum_{j=0}^{\infty} \binom{\alpha}{j} (-1)^j 2^{\alpha-j} \left[\Phi \left(\frac{\log(x) - \mu}{\sigma} \right) \right]^{\alpha+j}$$

because of the identity given by

$$(2 - b)^{\alpha} = \sum_{j=0}^{\infty} \binom{\alpha}{j} (-1)^j 2^{\alpha-j} b^j,$$

for $|b| < 2$.

Now, note that

$$[\Phi(.)]^{\alpha+j} = [1-(1-\Phi(.))]^{\alpha+j} = \sum_{k=0}^{\infty}\binom{\alpha+j}{k}(-1)^k[1-\Phi(.)]^k$$
$$= \sum_{k=0}^{\infty}\sum_{r=0}^{k}\binom{\alpha+j}{k}\binom{k}{r}(-1)^{k+r}[\Phi(.)]^r.$$

As a result, the CDF of the TLLN distribution takes the form

$$F(x) = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{r=0}^{k} W_{j,k,r}(\alpha)\left[\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]^r,$$

where

$$W_{j,k,r}(\alpha) = \binom{\alpha}{j}\binom{\alpha+j}{k}\binom{k}{r}(-1)^{j+k+r}2^{\alpha-j}.$$

Thus, the TLLN distribution can be expressed as the infinite sum of exponentiated LN distributions indexed by the power parameter r. If the parameter α is an integer, the TLLN distribution can be expressed as the finite sum of exponentiated LN distributions given as

$$F(x) = \sum_{j=0}^{\alpha}\sum_{k=0}^{\alpha+j}\sum_{r=0}^{k} W_{j,k,r}(\alpha)\left[\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]^r.$$

Again, applying the series expansion in Equation (3), the PDF of the TLLN distribution can be written as

$$f(x) = \frac{2\alpha}{\sigma x}\,\phi\left(\frac{\log(x)-\mu}{\sigma}\right)\sum_{j=0}^{\infty}\binom{\alpha-1}{j}(-1)^j\left[1-\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]^{2j+1},$$

or

$$f(x) = \sum_{j=0}^{\infty}\sum_{k=0}^{2j+1} b_\alpha(j,k)\psi_k(x), \tag{4}$$

where

$$b_\alpha(j,k) = 2\binom{\alpha-1}{j}\binom{2j+1}{k}\frac{(-1)^{j+k}}{k+1} \tag{5}$$

and

$$\psi_k(x) = \frac{k+1}{\sigma x}\phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left[\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]^k$$

is the PDF of the exponentiated LN distribution with power parameter $k+1$.

Remark 1 If x is fixed, the PDFs of the TLLN and LN distributions correspond when

$$\alpha = \frac{\log\left[\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]}{\log\left\{\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left[2-\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]\right\}}.$$

The proof is straightforward and omitted for the sake of brevity.

LEMMA 3.2 For $\alpha = 1$, the CDF of the TLLN distribution in Equation (2) becomes the CDF of the transmuted LN distribution with transmuted parameter equals to 1.

Proof: To begin with, a retrospective on the transmuted distributions is necessary. Shaw and Buckley (2009) introduced a new family of distributions called transmuted distributions, and the general expression of its CDF is

$$F_T(x) = (1+\lambda)G(x) - \lambda[G(x)]^2, \quad |\lambda| \leq 1,$$

where G is the baseline CDF and λ is called the transmuted parameter. Thus, the CDF of the TLLN distribution can be written as

$$F(x) = 2\Phi\left(\frac{\log(x)-\mu}{\sigma}\right) - \left[\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]^2,$$

which is also the CDF of the transmuted LN distribution with $\lambda = 1$. It is worth mentioning that the transmuted LN distribution is not discussed in the available literature. Hence, one may study its properties and applications in detail.

4. MOMENTS

In this section, we derive the expression for the raw moments of the TLLN distribution. Let m be a positive integer and X be a random variable following the TLLN distribution. The mth raw moment of the TLLN distribution is then calculated using Equation (4) as

$$\mu'_m = \mathrm{E}(X^m) = \sum_{j=0}^{\infty}\sum_{k=0}^{2j+1}(k+1)b_\alpha(j,k)\mu'_{m,k}, \tag{6}$$

where

$$\mu'_{m,k} = \int_0^\infty \frac{x^{m-1}}{\sigma}\phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left[\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]^k \mathrm{d}x$$

is the probability-weighted moment of the LN distribution. In other words, the raw moments of the TLLN distribution can be written as the weighted sum of the probability-weighted moments of the LN distribution.

Remark 1 If α is an integer, then the mth raw moment of the TLLN distribution is directly derived from Equation (6), and it is stated as

$$\mu'_m = \mathrm{E}(X^m) = \sum_{j=0}^{\alpha-1}\sum_{k=0}^{2j+1}(k+1)b_\alpha(j,k)\mu'_{m,k}.$$

5. Quantile function and associated measures

In this section, we derive an explicit expression for the QF of the TLLN distribution as well as several of its associated measures.

Theorem 5.1 Let $p \in (0,1)$. Then, the pth quantile of the TLLN distribution is given by

$$Q_p = F^{-1}(p) = \exp\left[\mu + \sigma\Phi^{-1}\left(1 - \sqrt{1 - p^{1/\alpha}}\right)\right], \tag{7}$$

where Φ^{-1} is the QF of a standard normal distribution.

Proof: For the TLLN distribution, Q_p is the solution of the following equation:

$$\left\{\Phi\left(\frac{\log(Q_p) - \mu}{\sigma}\right)\left[2 - \Phi\left(\frac{\log(Q_p) - \mu}{\sigma}\right)\right]\right\}^{\alpha} = p$$
$$\Rightarrow 2\Phi\left(\frac{\log(Q_p) - \mu}{\sigma}\right) - \left[\Phi\left(\frac{\log(Q_p) - \mu}{\sigma}\right)\right]^2 = p^{1/\alpha}. \tag{8}$$

On simplifications, since $p \in (0,1)$, Equation (8) reduces to

$$\left[1 - \Phi\left(\frac{\log(Q_p) - \mu}{\sigma}\right)\right]^2 = 1 - p^{1/\alpha}$$

$$\Rightarrow \frac{\log(Q_p) - \mu}{\sigma} = \Phi^{-1}\left(1 - \sqrt{1 - p^{1/\alpha}}\right)$$

$$\Rightarrow Q_p = \exp\left[\mu + \sigma\Phi^{-1}\left(1 - \sqrt{1 - p^{1/\alpha}}\right)\right].$$

Remark 1 Since Φ^{-1} is the QF of the standard normal distribution, Q_p in Equation (7) also gets the form

$$Q_p = \exp\left[\mu + \sigma\sqrt{2}\mathrm{erf}^{-1}\left(1 - 2\sqrt{1 - p^{1/\alpha}}\right)\right], \tag{9}$$

where erf^{-1} is the inverse error function.

By putting $p = 1/2$ in Equation (9), we get the median of the TLLN distribution, and it is given by

$$\mathrm{M} = Q_{0.5} = \exp\left[\mu + \sigma\sqrt{2}\mathrm{erf}^{-1}\left(1 - 2\sqrt{1 - \left(\frac{1}{2}\right)^{1/\alpha}}\right)\right].$$

Equation (9) delivers the first and third quartiles of the TLLN distribution ($Q_{0.25}$ and $Q_{0.75}$) for $p = 1/4$ and $p = 3/4$, respectively.

6. Reliability measures

We derive the expressions for various reliability measures in this section. The HRF of the TLLN distribution is expressed by

$$h(x) = \frac{f(x)}{S(x)},$$

where $S(x) = 1 - F(x)$ is the survival function of the TLLN distribution given by

$$S(x) = 1 - \left\{\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left[2 - \Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]\right\}^{\alpha}.$$

Thus, the desired HRF gets the following form:

$$h(x) = \frac{2\alpha\phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left[1-\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]\left\{\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left[2-\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]\right\}^{\alpha-1}}{\sigma x\left[1-\left\{\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left[2-\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]\right\}^{\alpha}\right]}.$$

Also, plots in Figure 3 refer to the shapes of the HRF and show that the TLLN distribution possesses increasing, decreasing, and upside-down bathtub shapes. Also, as seen in Figure 3, the distribution has a new decreasing-increasing-decreasing shape that we call the inverted N-shaped HRF, as well as a special shape that starts with a flat region and continues with an increasing-decreasing shape that we call the constant-increasing-decreasing shaped HRF.

Let r be a positive integer and X be a random variable following the TLLN distribution. Then, the rth conditional moment of the TLLN distribution is stated as

$$\begin{aligned}\mathrm{E}(X^r|X>t) &= \frac{1}{S(t)}\int_t^{\infty} x^r f(x)\mathrm{d}x \\ &= \frac{1}{S(t)}\sum_{j=0}^{\infty}\sum_{k=0}^{2j+1}\left(\frac{k+1}{\sigma}\right) b_{\alpha}(j,k) I_1(r,k), \qquad (10)\end{aligned}$$

where $b_\alpha(j,k)$ is given in Equation (5) and $I_1(r,k)$ is formulated as

$$I_1(r,k) = \int_t^{\infty} x^{r-1}\phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left[\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]^k \mathrm{d}x. \qquad (11)$$

It is worth mentioning that the rapid aging of a component requires low vitality, whereas high vitality implies relatively slow aging during the given time period.

For $r = 1$, Equation (10) gives the vitality function of the TLLN distribution, which is

$$\begin{aligned}V(t) = \mathrm{E}(X|X>t) &= \frac{1}{S(t)}\int_t^{\infty} x f(x)\mathrm{d}x \\ &= \frac{1}{S(t)}\sum_{j=0}^{\infty}\sum_{k=0}^{2j+1}\left(\frac{k+1}{\sigma}\right) b_{\alpha}(j,k) I_1(1,k), \qquad (12)\end{aligned}$$

where $I_1(1,k)$ is obtained by putting $r = 1$ in Equation (11), and is given by

$$I_1(1,k) = \int_t^{\infty} \phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left[\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]^k \mathrm{d}x.$$

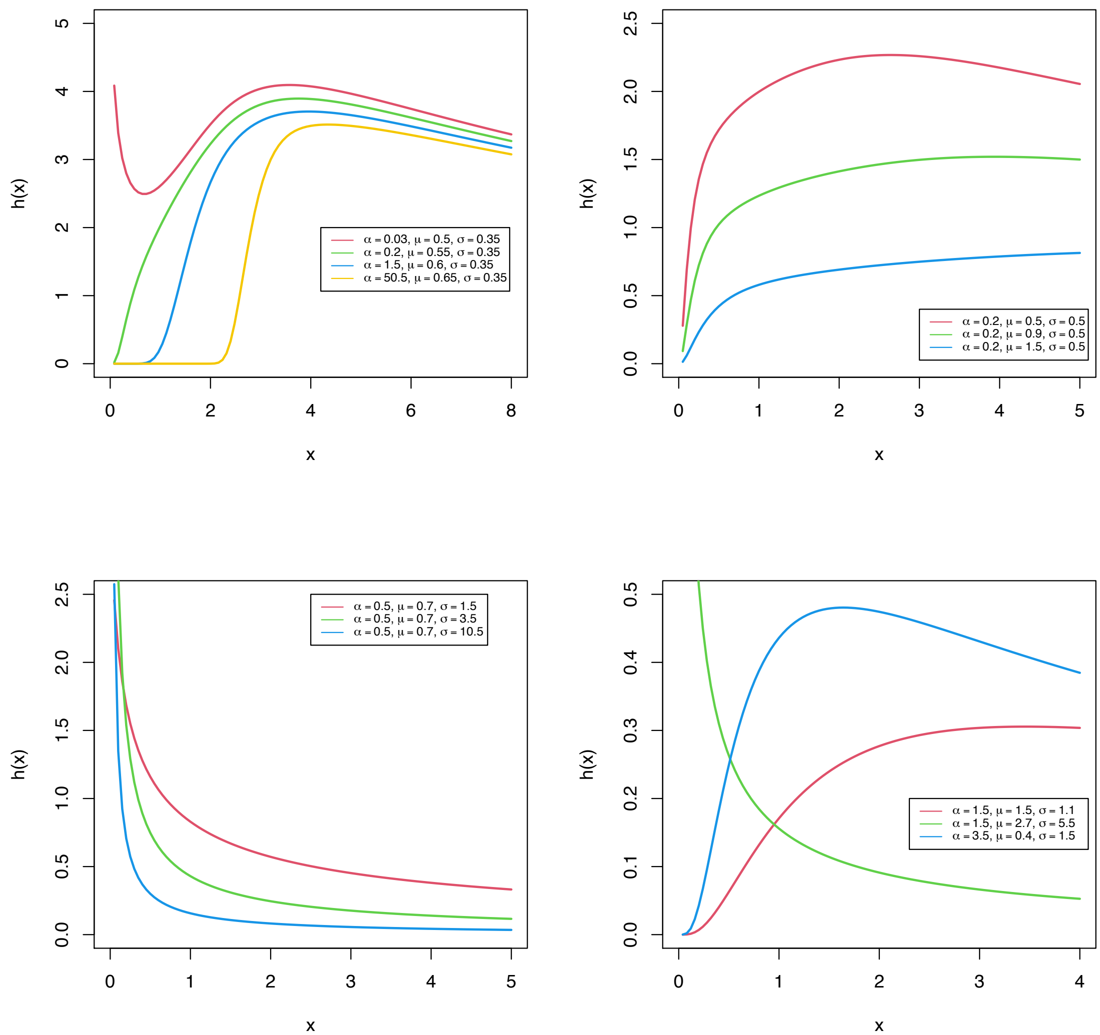

Figure 3. Plots of the HRF of the TLLN distribution.

If X is a random variable representing a component's lifetime, then $\log(G(t)) = \mathrm{E}(\log(X)|X > t)$ represents the ideal geometric mean of the lifetimes of components that have survived up to time t. The geometric vitality function of the TLLN distribution is stated as

$$\log(G(t)) = \frac{1}{S(t)} \sum_{j=0}^{\infty} \sum_{k=0}^{2j+1} (k+1) b_\alpha(j,k) I_2(k),$$

where $I_2(k)$ can be expressed as

$$I_2(k) = \int_t^{\infty} \frac{\log(x)}{\sigma x} \phi\left(\frac{\log(x)-\mu}{\sigma}\right) \left[\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]^k \mathrm{d}x.$$

The concept of residual life is of special interest in reliability theory. It measures the amount of time a unit has left after reaching the age of t. The rth order moment of the residual life of the TLLN distribution is defined as

$$\begin{aligned}\mu_r(t) = \mathrm{E}\left[(X-t)^r|X>t\right] &= \frac{1}{S(t)}\int_t^{\infty}(x-t)^r f(x)\mathrm{d}x \\ &= \frac{1}{S(t)}\sum_{i=0}^{r}\binom{r}{i}(-1)^{r-i}t^{r-i}I_1(i,k),\end{aligned}$$

where $I_1(i,k)$ is given in Equation (11). Now, by taking $r=1$, we get the expression for the mean residual life (MRL) function, which also gets the form

$$\mu_1(t) = V(t) - t,$$

where $V(t)$ is given in Equation (12). Similarly, the second moment of the residual lifetime of the TLLN distribution is stated as

$$\mu_2(t) = \frac{1}{S(t)}\sum_{j=0}^{\infty}\sum_{k=0}^{2j+1}\left(\frac{k+1}{\sigma}\right)b_\alpha(j,k)I_1(2,k) - \frac{2tV(t)}{S(t)} + t^2,$$

where $I_1(2,k)$ is defined as

$$I_1(2,k) = \int_t^{\infty} x\phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left[\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]^k \mathrm{d}x.$$

Thus, the variance of the residual life function of the TLLN distribution can be obtained using $\mu_1(t)$ and $\mu_2(t)$. The rth order moment of the reversed residual life of the TLLN distribution is formulated as

$$\begin{aligned}m_r(t) = \mathrm{E}\left[(t-X)^r|X\leq t\right] &= \frac{1}{F(t)}\int_0^{t}(t-x)^r f(x)\mathrm{d}x \\ &= \frac{1}{F(t)}\sum_{i=0}^{r}\binom{r}{i}(-1)^{i}t^{r-i}\left[\mu_i' - I_1(i,k)\right], \qquad (13)\end{aligned}$$

where $I_1(i,k)$ is given in Equation (11). Now, the mean $m_1(t)$ and second moment $m_2(t)$ of the reversed residual life of the TLLN distribution can be obtained by setting $r=1$ and $r=2$, respectively, in Equation (13). Again, using $m_1(t)$ and $m_2(t)$, one can obtain the variance of the reversed residual life function of the distribution.

7. Estimation of the parameters

In this section, we discuss how to estimate the parameters of the TLLN distribution utilizing two well-known methods, namely the maximum likelihood (ML) method and the Bayesian method. Next, we consider the ML estimation for the TLLN model parameters α, μ and σ. Let $X_1, \ldots, X_n$ represent a random sample from the TLLN distribution, and $x_1, \ldots, x_n$ represent the observed values. Then, the log-likelihood function can then be written in the following form:

$$\begin{aligned}\mathcal{L}_n &= \sum_{i=1}^{n} \log[f(x_i)] \\ &= n\log(2) + n\log(\alpha) - n\log(\sigma) - \sum_{i=1}^{n}\log(x_i) + \sum_{i=1}^{n}\log\left[\phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)\right] \\ &+ \sum_{i=1}^{n}\log\left[1-\Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)\right] + (\alpha-1)\sum_{i=1}^{n}\log\left[\Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)\right] \\ &+ (\alpha-1)\sum_{i=1}^{n}\log\left[2-\Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)\right].\end{aligned}$$

The ML estimates of (α, μ, σ) are $(\widehat{\alpha}, \widehat{\mu}, \widehat{\sigma}) = \text{argmax}_{(\alpha,\mu,\sigma)}\mathcal{L}_n$. We can formulate them by using nonlinear log-likelihood equations. First, the score function associated with the log-likelihood function is

$$\boldsymbol{U} = \left(\frac{\partial \mathcal{L}_n}{\partial \alpha}, \frac{\partial \mathcal{L}_n}{\partial \mu}, \frac{\partial \mathcal{L}_n}{\partial \sigma}\right)^{\top}.$$

The associated nonlinear log-likelihood equations are $\boldsymbol{U} = (0,0,0)^{\top}$, that is,

$$\frac{n}{\alpha} + \sum_{i=1}^{n}\log\left[\Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)\right] + \sum_{i=1}^{n}\log\left[2-\Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)\right] = 0, \tag{14}$$

$$\begin{aligned}\sum_{i=1}^{n}\frac{\log(x_i)-\mu}{\sigma^2} + \frac{1}{\sigma}\sum_{i=1}^{n}\frac{\phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)}{1-\Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)} - \frac{\alpha-1}{\sigma}\sum_{i=1}^{n}\frac{\phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)}{\Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)} \\ + \frac{\alpha-1}{\sigma}\sum_{i=1}^{n}\frac{\phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)}{2-\Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)} = 0\end{aligned} \tag{15}$$

and

$$\begin{aligned}-\frac{n}{\sigma} + \sum_{i=1}^{n}\frac{(\log(x_i)-\mu)^2}{\sigma^3} + \sum_{i=1}^{n}\frac{\left(\frac{\log(x_i)-\mu}{\sigma^2}\right)\phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)}{1-\Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)} \\ - \frac{\alpha-1}{\sigma^2}\sum_{i=1}^{n}\frac{(\log(x_i)-\mu)\phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)}{\Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)} \\ + \frac{\alpha-1}{\sigma^2}\sum_{i=1}^{n}\frac{(\log(x_i)-\mu)\phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)}{2-\Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)} = 0.\end{aligned} \tag{16}$$

Solving the nonlinear Equations (14), (15) and (16) synergistically, one can obtain the ML estimates. For known μ and σ, the ML estimate of α is given by

$$\widehat{\alpha} = -\frac{n}{\sum_{i=1}^{n} \log\left[\Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)\right] + \sum_{i=1}^{n} \log\left[2-\Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)\right]}.$$

The asymptotic confidence intervals (CIs) for the parameters α, μ and σ are now executed. On taking the second partial derivatives of Equations (14), (15) and (16) taken at the ML estimates, the observed Hessian matrix of the TLLN distribution can be obtained, and it is given by

$$\widehat{H} = \left.\begin{pmatrix} \frac{\partial^2 \mathcal{L}_n}{\partial \alpha^2} & \frac{\partial^2 \mathcal{L}_n}{\partial \alpha \partial \mu} & \frac{\partial^2 \mathcal{L}_n}{\partial \alpha \partial \sigma} \\ \frac{\partial^2 \mathcal{L}_n}{\partial \mu \partial \alpha} & \frac{\partial^2 \mathcal{L}_n}{\partial \mu^2} & \frac{\partial^2 \mathcal{L}_n}{\partial \mu \partial \sigma} \\ \frac{\partial^2 \mathcal{L}_n}{\partial \sigma \partial \alpha} & \frac{\partial^2 \mathcal{L}_n}{\partial \sigma \partial \mu} & \frac{\partial^2 \mathcal{L}_n}{\partial \sigma^2} \end{pmatrix}\right|_{(\alpha,\mu,\sigma)=(\widehat{\alpha},\widehat{\mu},\widehat{\sigma})}.$$

Now, the observed Fisher's information matrix $\widehat{J}$ is obtained as $\widehat{J} = -\widehat{H}$. The inverse of this matrix provides the variance-covariance matrix of the ML estimators, which can be written as

$$\widehat{\boldsymbol{\Sigma}} = \widehat{J}^{-1} = \begin{pmatrix} \widehat{\boldsymbol{\Sigma}}_{11} & \widehat{\boldsymbol{\Sigma}}_{12} & \widehat{\boldsymbol{\Sigma}}_{13} \\ \widehat{\boldsymbol{\Sigma}}_{21} & \widehat{\boldsymbol{\Sigma}}_{22} & \widehat{\boldsymbol{\Sigma}}_{23} \\ \widehat{\boldsymbol{\Sigma}}_{31} & \widehat{\boldsymbol{\Sigma}}_{32} & \widehat{\boldsymbol{\Sigma}}_{33} \end{pmatrix},$$

and $\widehat{\boldsymbol{\Sigma}}_{ij} = \widehat{\boldsymbol{\Sigma}}_{ji}$ for $i \neq j = 1, 2, 3$. The asymptotically normal distribution of the ML estimators has been thoroughly established. The random version of $\widehat{\Theta} = (\widehat{\alpha}, \widehat{\mu}, \widehat{\sigma})$ follows the multivariate normal distribution $N_3(\Theta, \widehat{\Sigma})$, where $\Theta = (\alpha, \mu, \sigma)$. Thus, we obtain $100 \times (1-\delta)\%$ asymptotic CIs of the parameters using the following formulae:

$$I_\alpha = \left[\widehat{\alpha} \mp \upsilon_{\delta/2}\sqrt{\widehat{\boldsymbol{\Sigma}}_{11}}\right], \quad I_\mu = \left[\widehat{\mu} \mp \upsilon_{\delta/2}\sqrt{\widehat{\boldsymbol{\Sigma}}_{22}}\right], \quad I_\sigma = \left[\widehat{\sigma} \mp \upsilon_{\delta/2}\sqrt{\widehat{\boldsymbol{\Sigma}}_{33}}\right],$$

where υ_δ is the upper δth percentile of the standard normal distribution. Next, we perform the Bayesian analysis for the TLLN model parameters. To do so, each parameter must have a prior distribution. We employ two types of priors for this: the half-Cauchy (HC) and the classical normal (N) priors. The PDF of the HC distribution with scale parameter a is defined as

$$f_{\mathrm{HC}}(x) = \frac{2a}{\pi(x^2+a^2)}, \quad x > 0, \; a > 0. \tag{17}$$

The HC distribution is known to have no mean or variance. Meanwhile, its mode is equal to 0. Since the PDF of the HC distribution is virtually flat but not totally flat at scale value equals 25, which verges on acquiring adequate information for the numerical approximation algorithm to continue looking at the target posterior distribution, the HC distribution with scale parameter equals 25 is recommended as a non-informative prior. According to Gelman and Hill (2006), the uniform distribution, or whether more information is required, is a superior alternative to the HC distribution. As a result, for the parameters α and σ, the HC distribution with a scale parameter equaling 25 was chosen as a non-informative prior distribution in this study. Thus, we set the prior distributions of the parameters to be

$$\mu \sim \mathrm{N}(0, 1000), \quad \alpha, \sigma \sim \mathrm{HC}(25). \tag{18}$$

Thus, using Equation (18), we obtain the joint posterior PDF as given by

$$\pi(\mu, \alpha, \sigma | x) \propto L_n \times \pi(\mu) \times \pi(\alpha) \times \pi(\sigma), \tag{19}$$

where L_n is the likelihood function of the TLLN distribution. From Equation (19), it is obvious that there is no analytical solution to find out the Bayesian estimates. Thus, we use a remarkable method of simulation, namely the Metropolis-Hastings (MH) algorithm of the Markov Chain Monte Carlo (MCMC) method. Upadhyay et al. (2001) provides a thorough description of the MCMC approach.

8. Bootstrap confidence intervals

In this section, we utilize the parametric bootstrap method to approximate the distribution of the ML estimators of the TLLN model parameters. Then, we can employ the bootstrap distribution to estimate each parameter's CIs for the fitted TLLN distribution. Let $\widehat{\Xi}$ be the ML estimate of Ξ, where $\Xi \in (\alpha, \mu, \sigma)$, using a given dataset $\{x_1, x_2, \ldots, x_n\}$. The bootstrap is a method to estimate the distribution of the statistic $\widehat{\Xi}$ by getting a random sample $\Xi_1^*, \Xi_2^*, \ldots, \Xi_B^*$ for Ξ based on B random samples that are drawn with replacement from the original data $x_1, x_2, \ldots, x_n$. Thus, the bootstrap sample $\Xi_1^*, \Xi_2^*, \ldots, \Xi_B^*$ can be used to construct bootstrap CIs for α, μ and σ.

Thus, using the following formulae, we calculate the $100 \times (1-\delta)\%$ bootstrap CIs for the parameters:

$$\mathfrak{I}_\alpha = \left[\widehat{\alpha} \mp z_{\delta/2}\widehat{\mathrm{SE}}_{\alpha,\mathrm{boot}}\right], \quad \mathfrak{I}_\mu = \left[\widehat{\mu} \mp z_{\delta/2}\widehat{\mathrm{SE}}_{\mu,\mathrm{boot}}\right], \quad \mathfrak{I}_\sigma = \left[\widehat{\sigma} \mp z_{\delta/2}\widehat{\mathrm{SE}}_{\sigma,\mathrm{boot}}\right].$$

where z_δ denotes the δth percentile of the bootstrap sample, SE is the standard error, and, for $\Xi \in \{\alpha, \mu, \sigma\}$,

$$\widehat{\mathrm{SE}}_{\Xi,\mathrm{boot}} = \sqrt{\frac{1}{B}\sum_{b=1}^{B}\left(\Xi_b^* - \frac{1}{B}\sum_{b=1}^{B}\Xi_b^*\right)^2}.$$

9. TLLN regression model

In this section, we define a regression model based on the TLLN distribution, called the TLLN regression model.

To begin, consider a random variable X following the TLLN distribution with PDF given in Equation (3), as well as the random variable Y defined by $Y = \log(X)$. Then, Y has the following PDF:

$$f_Y(y) = \frac{2\alpha}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)\left[1 - \Phi\left(\frac{y-\mu}{\sigma}\right)\right]\left\{\Phi\left(\frac{y-\mu}{\sigma}\right)\left[2 - \Phi\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^{\alpha-1},$$
$$y \in \mathbb{R}, \ \mu \in \mathbb{R}, \ \alpha, \sigma > 0. \tag{20}$$

We refer to Equation (20) as the log-Topp-Leone log-normal (log-TLLN) distribution, or otherwise, the Topp-Leone normal (TLN) distribution, which is given by Sharma (2018). In this setting, the standardized random variable $Z = (Y - \mu)/\sigma$ has the PDF given by

$$f_Z(z) = 2\alpha\phi(z)\left[1 - \Phi(z)\right]\left\{\Phi(z)\left[2 - \Phi(z)\right]\right\}^{\alpha-1}. \tag{21}$$

Now, linear location-scale regression model linking the response variable y_i and the explanatory variable vector $\boldsymbol{v}_i^\top = (v_{i1}, \ldots, v_{ip})$, is obtained as

$$y_i = \mu_i + \sigma z_i, \quad i = 1, \ldots, n, \tag{22}$$

where z_i is the random error component that has the PDF in Equation (21), $\mu_i = \mathrm{v}_i^\top \tau$ is the location parameter of y_i, where $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_p)^\top$, α and σ are unknown parameters. The location parameter vector $\mu = (\mu_1, \ldots, \mu_n)^\top$ is represented by a linear model $\mu = \boldsymbol{V}\boldsymbol{\tau}$, where $\boldsymbol{V} = (V_1, \ldots, V_n)^\top$ is a known model matrix. Ultimately, in this study, we propose the TLLN regression model from Equation (22) and it is given by

$$x_i = \mathrm{e}^{y_i} = \mathrm{e}^{\mu_i + \sigma z_i}, \quad i = 1, \ldots, n. \tag{23}$$

Consider a sample $(x_1, \boldsymbol{v}_1), \ldots, (x_n, \boldsymbol{v}_n)$ of n independent observations. Conventional likelihood estimation techniques can be applied here. Now, for the vector of parameters $\boldsymbol{\psi} = (\boldsymbol{\tau}^\top, \alpha, \sigma)^\top$ from model (23), the total log-likelihood function for right censored has the form

$$l(\boldsymbol{\psi}) = \sum_{i=1}^{n} \delta_i \log\left[f(x_i)\right] + \sum_{i=1}^{n} (1 - \delta_i) \log\left[S(x_i)\right],$$

with $\delta_i = 1$, if survival (uncensored) and $\delta_i = 0$, if not (censored). We recall that, for $i = 1, \ldots, n$, $f(x_i)$ and $S(x_i)$ are the PDF and survival function of the TLLN distribution taken at x_i, respectively.

10. Bayesian regression method

The Bayesian technique has been shown to be particularly effective in analyzing survival models in many practical circumstances. Hence, in this section, we look at how the Bayesian approach fits the regression model based on the TLLN distribution when prior pieces of information about the parameters are taken into account. As a result, we use a simulation method in this part for Bayesian analysis of this model. Now, to perform a Bayesian analysis, one should adopt prior distributions for the parameters. Here, as described previously, we utilize the HC and N priors. The PDF of the HC distribution with a as the scale parameter is given in Equation (17). Now, we write the right censored likelihood function as

$$L = \prod_{i=1}^{n} \left[f(x_i)\right]^{\delta_i} \left[S(x_i)\right]^{1-\delta_i}, \tag{24}$$

with $\delta_i = 1$, if survival (uncensored) and $\delta_i = 0$, if not (censored).

$$\mu = V\tau \tag{25}$$

as a linear combination of explanatory variables. Thus, we set the prior distributions of the parameters to be

$$\tau_j \sim \mathrm{N}(0, 1000), \quad j = 1, \ldots, J, \quad \alpha, \sigma \sim \mathrm{HC}(25). \tag{26}$$

Now, using Equations (24), (25) and (26), the joint posterior PDF is obtained as

$$\pi(\tau, \alpha, \sigma | x, V) \propto L(x|V, \tau, \alpha, \sigma) \times \pi(\tau) \times \pi(\alpha) \times \pi(\sigma). \tag{27}$$

From Equation (27), it is clear that the analytical solution is not possible to find out the Bayesian estimates. As a result, we employ the simulation approach, specifically the MH algorithm of the MCMC method.

11. Performance of the estimators using simulations

In this section, we conduct simulation experiments to assess the long-run performance of the ML estimators of the TLLN model parameters for some finite sample sizes. We simulated datasets of sizes $n = 60$, 100, and 250 from the TLLN distribution for the parameter values $\alpha = 0.5, \mu = 0.9, \sigma = 0.6$ and iterated each sample 500 times. The average bias and MSE for all replications in the relevant sample sizes are then computed. That is, the analysis computes the values using the given formulae.

Table 1. Estimates, average bias and MSE values of ML estimators from simulations of the TLLN distribution.

Parameters	Sample Size	Estimates	Bias	MSE
α	60	1.7496	1.2496	67.9030
	100	1.0114	0.5114	6.3232
	250	0.5792	0.0792	0.1677
μ	60	0.8240	-0.0760	0.3754
	100	0.8491	-0.0509	0.2180
	250	0.8873	-0.0127	0.0504
σ	60	0.5520	-0.0480	0.1350
	100	0.5811	-0.0189	0.0849
	250	0.5916	-0.0084	0.0229

- Average bias of the simulated estimates $= \frac{1}{500} \sum_{i=1}^{500} (\widehat{\Xi}_i - \Xi)$,
- Average MSE of the simulated estimates $= \frac{1}{500} \sum_{i=1}^{500} (\widehat{\Xi}_i - \Xi)^2$,

where $\widehat{\Xi}_i$ represents the estimate of $\Xi \in \{\alpha, \mu, \sigma\}$ at the ith replication. The results are reported in Table 1. It can be concluded that the MSEs of all the estimates decrease with increasing sample size. This shows the consistency of the estimates.

12. Applications and empirical study

This section consists of demonstrating the empirical importance of the TLLN distribution. We use a real dataset from the area of astronomy to compare the data modeling ability of the

TLLN distribution with other competitive distributions. We employ the RStudio software for numerical evaluations of these datasets. The descriptive measures, which include sample size (n), mean (M), median (Md), variance (Var), skewness (Sk), kurtosis (Ku), minimum (min) and maximum (max) values of the dataset, are given in Table 2.

Table 2. Descriptive statistics of the astronomical dataset.

Statistic	n	M	Md	Var	Sk	Ku	min	max
Values	360	14.458	14.54	1.427	-0.395	0.344	10.749	18.052

To show the potential advantage of the TLLN distribution, the following distributions are considered for comparison.

- The two-parameter LN distribution with PDF

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left[-\frac{(\log(x)-\mu)^2}{2\sigma^2}\right], \quad x>0,\ \mu\in\mathbb{R},\ \sigma>0.$$

- The exponentiated LN (ELN) distribution with PDF

$$f(x) = \frac{\alpha}{x\sigma}\phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left[\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]^{\alpha-1}, \ x>0,\ \mu\in\mathbb{R},\ \alpha,\sigma>0.$$

- The generalized half-normal (GHN) distribution (see Cooray and Ananda, 2008) with PDF

$$f(x) = \sqrt{\frac{2}{\pi}}\left(\frac{\alpha}{x}\right)\left(\frac{x}{\sigma}\right)^{\alpha}\exp\left\{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^{2\alpha}\right\}, \quad x,\alpha,\sigma>0.$$

- The exponentiated exponential distribution (EED) with PDF

$$f(x) = \alpha\sigma\left(1-\mathrm{e}^{-\sigma x}\right)^{\alpha-1}\mathrm{e}^{-\sigma x}, \quad x,\alpha,\sigma>0.$$

- The new generalized Lindley distribution (NGLD) (see Elbatal et al., 2013) with PDF

$$f(x) = \frac{\mathrm{e}^{-\mu x}}{1+\mu}\left(\frac{\mu^{\alpha+1}x^{\alpha-1}}{\Gamma(\alpha)}+\frac{\mu^{\sigma}x^{\sigma-1}}{\Gamma(\sigma)}\right), \quad x>0,\ \alpha,\mu,\sigma>0,$$

  where $\Gamma(\alpha)$ denotes the standard gamma function.
- The gamma distribution.

For the numerical optimization, we maximize the log-likelihood function to find the ML estimates. For fixing a lower and upper bound for each parameter, the numerical optimization technique L-BFGS-B in `fitdistrplus` package of `R` is used. For more information and detailed examples of this package, one should go through the link `https://CRAN.R-project.org/package=fitdistrplus`.

The following statistical tools are utilized in order to compare the competitive models with the proposed models: negative log-likelihood ($-\log(L)$), Kolmogorov-Smirnov (KS), Cramér-von Mises (W^*), Anderson-Darling (A^*) statistics, Akaike information criterion (AIC) and Bayesian information criterion (BIC) values.

We also investigate the empirical HRF for the astronomical dataset using the idea of a total time on test (TTT) plot. It is a graphical representation being used to distinguish between several types of aging as displayed in the HRF shapes. On the mathematical aspect, the TTT plot is drawn by plotting

$$T\left(\frac{i}{n}\right) = \frac{\sum_{r=1}^{i} x_{r:n} + (n-i)x_{i:n}}{\sum_{r=1}^{n} x_{r:n}}.$$

against i/n, where $i = 1, \ldots, n$ and $x_{1:n}, x_{2:n}, \ldots, x_{n,n}$ are the order statistics of the sample. We also present other important graphs, which consist of the empirical CDF and quantile-quantile (Q-Q) plots for the dataset. We utilize the magnitudes of the near-infrared K-band distribution of 360 globular cluster luminosities in Messier 31 (M31), our nearby Andromeda Galaxy, as an astronomical dataset. The data are from Nantais et al. (2006), and the samples are described in detail in Appendix C.3 of Feigelson and Babu (2012), as well as in the R package `astrodatR`. Note that, the K-band is an atmospheric transmission window in infrared astronomy, referring to an area of the infrared spectrum where atmospheric gases absorb relatively little terrestrial heat radiation. Furthermore, globular clusters are densely packed groups of 10^4 to 10^6 ancient stars packed into a dense, roughly spherical shape that is structurally unique from the general population of stars. Astronomers can use them to determine the age of the universe or to locate the Galactic Center by studying them. The TTT plot in Figure 4 indicates that this dataset has an increasing HRF shape, which is also a characteristic of the TLLN model.

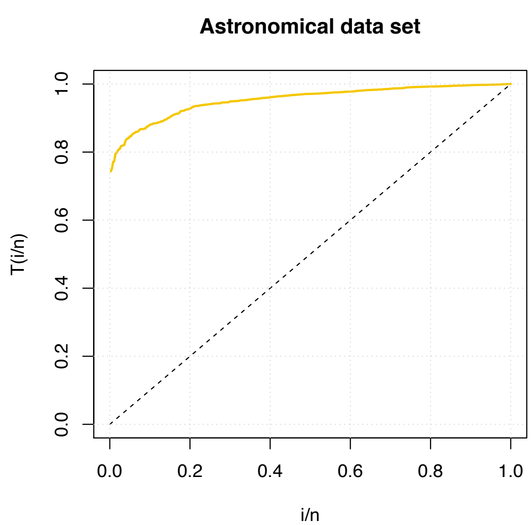

Figure 4. The TTT plot of astronomical dataset.

Next, we present results for the univariate dataset.

Table 3. Astronomical dataset: ML estimates and GOF statistics results.

Model	ML estimate	-log(L)	AIC	BIC	KS	W^*	A^*
TLLN	$\widehat{\alpha} = 0.2694$	571.7256	1149.451	1161.110	0.0621	0.2083	1.2120
	$\widehat{\mu} = 2.7796$						
	$\widehat{\sigma} = 0.0601$						
LN	$\widehat{\mu} = 2.6677$	582.5224	1169.045	1176.817	0.0774	0.5396	3.2814
	$\widehat{\sigma} = 0.0847$						
ELN	$\widehat{\alpha} = 0.1070$	573.2549	1152.510	1164.168	0.0657	0.2517	1.4640
	$\widehat{\mu} = 2.7826$						
	$\widehat{\sigma} = 0.0371$						
GHN	$\widehat{\mu} = 9.8248$	592.761	1189.522	1197.294	0.0753	0.6212	4.1150
	$\widehat{\sigma} = 15.2179$						
EXPPL	$\widehat{\alpha} = 1.8821$	605.6499	1217.300	1228.958	0.1096	1.2230	7.3301
	$\widehat{\mu} = 136.4388$						
	$\widehat{\sigma} = 0.0491$						
EED	$\widehat{\alpha} = 44847.58$	624.776	1253.552	1261.324	0.1293	1.7963	10.8203
	$\widehat{\sigma} = 0.7740$						
NGLD	$\widehat{\alpha} = 137.9676$	579.4283	1164.857	1176.515	0.0754	0.4665	2.8083
	$\widehat{\mu} = 9.5441$						
	$\widehat{\sigma} = 138.2696$						
Gamma	$\widehat{\alpha} = 142.2090$	579.3487	1162.697	1170.470	0.0722	0.4406	2.7088
	$\widehat{\sigma} = 9.8355$						

Table 3 displays the ML estimates and goodness-of-fit (GOF) statistics of the distributions corresponding to the astronomical dataset. The TLLN distribution's GOF statistics values are smaller than those of the other compared distributions. The empirical CDF and Q-Q plots for the dataset are given in Figure 5. The proposed distribution gives acceptable shaped curves for those empirical and fitted functions. As a result, we conclude that the TLLN distribution is superior to the other compared distributions for the astronomical dataset.

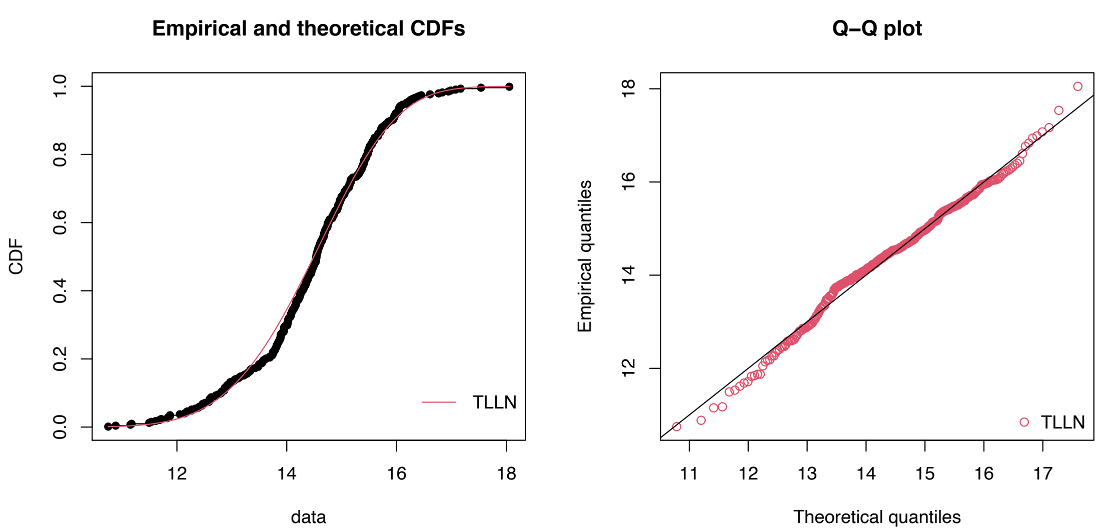

Figure 5. Empirical plots on the astronomical dataset.

Now, the Hessian matrix corresponding to the astronomical dataset is obtained as

$$\widehat{\boldsymbol{H}} = \begin{pmatrix} 4960.117 & 21162.32 & -35812.03 \\ 13376.538 & 57566.52 & -72436.86 \\ -35812.031 & -72436.86 & 296223.69 \end{pmatrix},$$

and the corresponding estimated variance-covariance matrix is

$$\widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 0.0105 & -0.0012 & 9.72 \times 10^{-04} \\ -0.0012 & 0.0002 & -1.07 \times 10^{-04} \\ 0.0009 & -0.0001 & 9.49 \times 10^{-05} \end{pmatrix}.$$

Table 4 provides the 95% asymptotic CIs for the TLLN model parameters.

Table 4. The 95% asymptotic CIs of the TLLN model parameters based on the astronomical dataset.

Parameter	Lower	Upper
α	0.0685	0.4703
μ	2.7543	2.8049
σ	0.0410	0.0792

Here, we focus on estimating the parameters of the TLLN distribution using the Bayesian procedure based on the above-discussed univariate astronomical dataset. In the context of Bayesian estimation, the analysis is performed using the MH algorithm of the MCMC method with 1000 iterations. For comparing Bayes estimates with the ML estimates, both of them for the TLLN model parameters for the real dataset are given in Table 5. The numerical computations for Bayesian estimation are done using the `LaplacesDemon` package of the `R` software, which provides a comprehensive environment for Bayesian inference. For more detailed information and examples regarding this package, one should go through the link https://cran.r-project.org/package=LaplacesDemon.

Table 5. ML and Bayes estimates of the TLLN model parameters on the astronomical dataset.

Parameter	ML	Bayes
α	0.2694	0.2811
μ	2.7796	2.7791
σ	0.0601	0.0602

Using the previously discussed astronomical dataset, we construct the 95% bootstrap for the parameters α, μ, and σ using the computed ML estimates. Based on the TLLN distribution, we simulate 1001 samples of the same size as the real dataset, with true values of the parameters chosen as the ML estimate of the respective parameters. For each sample obtained, we compute the ML estimates $\widehat{\alpha}_b^*$, $\widehat{\mu}_b^*$ and $\widehat{\sigma}_b^*$, for $b \in \{1, \ldots, 1001\}$. Table 6 displays the median and 95% bootstrap CI for the parameters α, μ and σ of the dataset. Examining the joint distribution of the bootstrapped values in a matrix of scatter plots to assess the potential structural correlation among the parameters is also noteworthy. Figure 6 displays the matrix scatterplots of the bootstrapped values of the TLLN model parameters, which depict the joint uncertainty distribution of the fitted parameters.

Table 6. The median and 95% bootstrap CI for the TLLN model parameters on the astronomical dataset.

Parameter	Median	Bootstrap CI
α	0.2695	(0.0878, 0.6941)
μ	2.7792	(2.7415, 2.8087)
σ	0.0609	(0.0373, 0.0881)

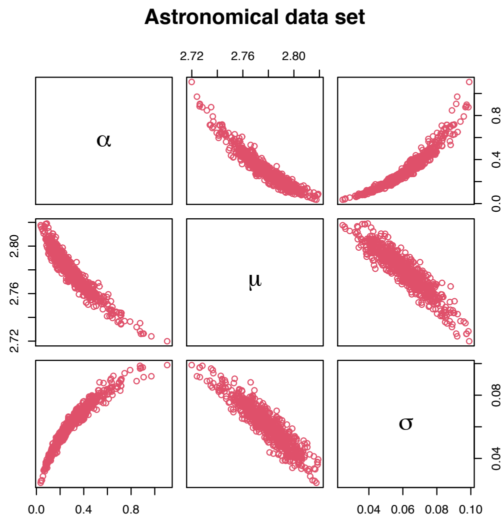

Figure 6. Matrix scatter plot on bootstrapped values of the TLLN model parameters due to the astronomical dataset.

We also utilize the likelihood ratio (LR) test for comparing the TLLN distribution, which has an additional parameter α with the LN and ELN distributions based on the above-discussed univariate astronomical dataset. The LR statistic for comparing the nested models H_0: LN and H_0: ELN against H_1: TLLN is given by

$$\mathrm{LR} = -2\log\left(\frac{\text{likelihood under the null hypothesis}}{\text{likelihood in the whole parameter space}}\right).$$

It is well-known that the random version of this statistic asymptotically follows a chi-square distribution with d degrees of freedom, d being equal to the number of additional parameters in the TLLN model. By using this result and standard statistical tables, we can obtain critical values for the LR test statistics for the given astronomical dataset. Table 7 includes the LR statistics and corresponding p-values for both the datasets. Given the values of test statistics and their associated p-values, we reject the null hypothesis for the above-discussed astronomical dataset and conclude that the TLLN distribution provides a significantly better representation than the LN and ELN distributions.

Table 7. Likelihood ratio statistics and their p-values on the astronomy dataset.

	LR	p-value
TLLN versus LN	21.594	3.37×10^{-06}
TLLN versus ELN	3.0586	2.2×10^{-16}

Now, we use a real, censored dataset based on the prognosis for women with breast cancer. Breast cancer is one of the most common forms of cancer in women. This lifetime dataset was carried out at the Middlesex Hospital, and documented in Leathem and Brooks (1987) and discussed by Collett (2015) which refers to the survival time in months of women who had received a simple or radical mastectomy to treat a tumor of Grade II, III or IV, between January 1969 and December 1971.

Table 8 summarizes the TLLN regression model as a result of the censored dataset, including estimates of all parameters, negative log-likelihood ($-l(\boldsymbol{\psi})$) and value of AIC. Here, we utilize the `optim` function of the `R` software for the numerical evaluations.

Table 8. Summaries for the TLLN regression model from the breast cancer dataset.

Parameter	τ_0	τ_1	α	σ	$-l(\boldsymbol{\psi})$	AIC
Estimates	0.6372	-1.2392	40.3536	4.3415	154.2923	316.5846

Table 9 represents the summary of 1000 times iterated simulated results, due to the censored dataset using the MH algorithm of the MCMC method, which includes the posterior mean, standard deviation (SD), Monte Carlo standard error (MCSE), effective sample size due to autocorrelation (ESS), 95% CI and the posterior median. Next, we use the `LaplacesDemon` package of `R` for the numerical evaluations.

Table 9. Summaries for the TLLN Bayesian regression model from the breast cancer dataset.

Parameter	Mean	SD	MCSE	ESS	95% CI	Median
τ_0	2.9068	0.6033	0.2791	9.3047	(1.7993, 3.8864)	3.0101
τ_1	-1.1779	0.5730	0.1699	17.7204	(-1.9877, -0.2168)	-1.2055
α	11.4851	2.6373	1.1827	9.2548	(7.8012, 16.7872)	11.3755
σ	3.8347	0.6524	0.2513	8.6492	(2.7314, 5.3484)	3.6765

13. Conclusions, limitations, and future research

We suggest a new distribution, which is a generalized version of the log-normal distribution, mainly to investigate data in the field of astronomy in this research, but it can also be used to match cancer datasets in biological aspects. We call it the Topp-Leone log-normal distribution. We explore several of its mathematical and statistical aspects. On the theoretical aspect, we provide specific expressions for the moments, quantile function, and various reliability measures. The different shapes of the hazard rate function are discussed. In terms of inference, the model parameters are estimated by using Bayesian estimation and the method of maximum likelihood, and also, the observed information matrix is presented. Furthermore, we adopt the parametric bootstrap technique to obtain confidence intervals for the model parameters. More importantly, we introduce a parametric regression model and a Bayesian regression method based on the new distribution. The usefulness of our methodology is illustrated by two applications of real datasets, one related to the astronomical study and the other to censored cancer data, using goodness-of-fit tests. The novel model consistently outperforms previous models in the literature in terms of fitting. We anticipate that the proposed model will find a larger range of applications in the modeling of positive real-world datasets, including not only astronomy but also biology, physics, engineering, survival analysis, hydrology, economics, and other fields.

The possible limitations of the proposed distribution include the impossibility of modeling phenomena with possible negative values or presenting a bimodal nature. The construction of quantile regression models and bivariate variants of the TLLN distribution are two further possible directions for this research. Additional significant improvements and investigations are required for this study, which we will put into future research.

Author contributions Conceptualization, C.C., M.R.I., D.S.S., S.L.N., R.M.; methodology, C.C., M.R.I., D.S.S., S.L.N., R.M.; software, XX; validation, C.C., M.R.I., D.S.S., S.L.N., R.M.; formal analysis, C.C., M.R.I., D.S.S., S.L.N., R.M.; investigation, C.C., M.R.I., D.S.S., S.L.N., R.M.; writing-original draft preparation, C.C., M.R.I., D.S.S., S.L.N., R.M.; writing-review and editing, C.C., M.R.I., D.S.S., S.L.N., R.M. All authors have read and agreed the published version of the paper.

Acknowledgements The authors would express their gratefulness for the constructive criticism of the Editors-in-Chief and the anonymous reviewers which helped to considerably improve the quality of the paper.

Funding The authors received no financial support for the research, authorship, and/or publication of this article.

Conflicts of interest The authors declare no conflict of interest.

References

Abdalla, H. et al., 2017. Characterizing the γ-ray long-term variability of PKS 2155−304 with H.E.S.S. and Fermi-Lat. Astronomy and Astrophysics, 598, A39.

Afify, A.Z., Altun, E., Alizadeh, M., Ozel, G., and Hamedani, G.G., 2017. The odd exponentiated half-logistic-G family: Properties, characterizations and applications. Chilean Journal of Statistics, 8, 65–91.

Alzaatreh, A., Lee, C., and Famoye, F., 2013. A new method for generating families of continuous distributions. Metron, 71, 63–79.

Bernardeau, F. and Kofman, L., 1994. Properties of the cosmological density distribution function. The Astrophysical Journal, 443, 479.

Blasi, P., Burles, S., and Olinto, a., 1999. Cosmological magnetic field limits in an inhomogeneous universe. The Astrophysical Journal Letters, 514, L79–L82.

Collett, D., 2015. Modelling Survival Data in Medical Research. Chapman and Hall, New York, USA.

Cooray, K. and Ananda, M.M.A., 2008. A generalization of the half-normal distribution with applications to lifetime data. Communications in Statistics: Theory and Methods, 37, 1323–1337.

Elbatal, I., Merovci, F., and Elgarhy, M., 2013. A new generalized Lindley distribution. Mathematical Theory and Modeling, 3, 30–47.

Feigelson, E. and Babu, G.J., 2012. In Modern Statistical Methods for Astronomy: With R Applications. Cambridge University Press, Cambridge, UK.

Gelman, A. and Hill, J., 2006. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, Cambridge, UK.

Hubble, E., 1934. The distribution of extra-galactic nebulae. Astrophysical Journal, 79, 8.

Jobe, J., Crow, E., and Shimizu, K., 1989. Lognormal distributions: Theory and applications. Technometrics, 31, 392.

Leathem, A. and Brooks, S., 1987. Predictive value of lectin binding on breast-cancer recurrence and survival. The Lancet, 329, 1054–1056.

Nantais, J.B., Huchra, J.P., Barmby, P., Olsen, K.A.G., and Jarrett, T.H., 2006. Nearby spiral globular cluster systems. I. Luminosity functions. The Astronomical Journal, 131, 1416–1425.

Parravano, A., Sánchez, N., and Alfaro, E.J., 2012. The dependence of prestellar core mass distributions on the structure of the parental cloud. Astrophysical Journal, 754, 150.

Sangsanit, Y. and Bodhisuwan, W., 2016. The Topp-Leone generator of distributions: Properties and inferences. Songklanakarin Journal of Science and Technology, 38, 537–548.

Shah, Z., Mankuzhiyil, N., Sinha, A., Misra, R., Sahayanathan, S., and Iqbal, N., 2018. Lognormal flux distribution of bright fermi blazars. Research in Astronomy and Astrophysics, 18, 141.

Shah, Z., Misra, R., and Sinha, A., 2020. On the determination of lognormal flux distributions for astrophysical systems. Monthly Notices of the Royal Astronomical Society, 496, 3348–3357.

Sharma, V., 2018. Topp-Leone normal distribution with application to increasing failure rate data. Journal of Statistical Computation and Simulation, 88, 1782–1803.

Shaw, W. and Buckley, I., 2009. The alchemy of probability distributions: beyond Gram-Charlier expansions, and a skew-kurtotic-normal distribution from a rank transmutation map. arXiv:0901.0434.

Topp, C. and Leone, F., 1955. A family of J-shaped frequency functions. Journal of The American Statistical Association, 50, 209–219.

Upadhyay, S. K., Vasishta, N., and Smith, A. F. M., 2001. Bayes inference in life testing and reliability via Markov chain monte Carlo simulation. Sankhya A, 63, 15–40.

Information for authors

The editorial board of the Chilean Journal of Statistics (ChJS) is seeking papers, which will be refereed. We encourage the authors to submit a PDF electronic version of the manuscript in a free format to the Editors-in-Chief of the ChJS (E-mail: chilean.journal.of.statistics@gmail.com). Submitted manuscripts must be written in English and contain the name and affiliation of each author followed by a leading abstract and keywords. The authors must include a"cover letter" presenting their manuscript and mentioning: "We confirm that this manuscript has been read and approved by all named authors. In addition, we declare that the manuscript is original and it is not being published or submitted for publication elsewhere".

Preparation of accepted manuscripts

Manuscripts accepted in the ChJS must be prepared in Latex using the ChJS format. The Latex template and ChJS class files for preparation of accepted manuscripts are available at http://soche.cl/chjs/files/ChJS.zip. Such as its submitted version, manuscripts accepted in the ChJS must be written in English and contain the name and affiliation of each author, followed by a leading abstract and keywords, but now mathematics subject classification (primary and secondary) are required. AMS classification is available at http://www.ams.org/mathscinet/msc/. Sections must be numbered 1, 2, etc., where Section 1 is the introduction part. References must be collected at the end of the manuscript in alphabetical order as in the following examples:

Arellano-Valle, R., 1994. Elliptical Distributions: Properties, Inference and Applications in Regression Models. Unpublished Ph.D. Thesis. Department of Statistics, University of São Paulo, Brazil.

Cook, R.D., 1997. Local influence. In Kotz, S., Read, C.B., and Banks, D.L. (Eds.), Encyclopedia of Statistical Sciences, Vol. 1., Wiley, New York, pp. 380-385.

Rukhin, A.L., 2009. Identities for negative moments of quadratic forms in normal variables. Statistics and Probability Letters, 79, 1004-1007.

Stein, M.L., 1999. Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.

Tsay, R.S., Peña, D., and Pankratz, A.E., 2000. Outliers in multivariate time series. Biometrika, 87, 789-804.

References in the text must be given by the author's name and year of publication, e.g., Gelfand and Smith (1990). In the case of more than two authors, the citation must be written as Tsay et al. (2000).

Copyright