Check for updates

The Lambert-Topp-Leone distribution: an alternative for modeling proportion and lifetime data

Juan M. Astorga[1]* and Yuri A. Iriarte[2]*

[1]Departamento de Tecnologías de la Energía, Facultad Tecnológica, Universidad de Atacama, Copiapó, Chile, [2]Departamento de Estadística y Ciencia de Datos, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta, Chile

This article discusses the Lambert-Topp-Leone distribution as a flexible alternative for modeling proportion and lifetime data. By extending the Topp-Leone distribution, the proposed model offers greater flexibility in terms of skewness and kurtosis, making it suitable for a broader range of real-world applications. We examine key properties of the distribution, including its moments and behavior in terms of skewness and kurtosis. Parameter estimation using the maximum likelihood method is also discussed. A Monte Carlo simulation study is conducted to evaluate the performance of the estimators. Finally, to illustrate its practical utility, we apply the Lambert-Topp-Leone distribution to real-world datasets, demonstrating its superior fit for proportion and lifetime data compared to traditional models. The results suggest that this distribution provides a valuable tool for researchers and professionals in fields that require versatile modeling of bounded or positively skewed data.

KEYWORDS

goodness-of-fit, kurtosis, Lambert generator, lifetime data, maximum likelihood estimation, proportion data, skewness

1 Introduction

The Topp-Leone (TL) distribution is a continuous probability distribution often used to model bounded data, especially when a decreasing or unimodal pattern in the frequency of events is observed. It was originally proposed by Topp and Leone [1] as an alternative for modeling failure data. The cumulative distribution function (CDF), the probability density function (PDF), and the quantile function (QF) of the TL distribution are expressed as follows:

$$F(x) = \left(1 - y^2\right)^{\beta}, \tag{1}$$

$$f(x) = \frac{2\beta}{b} y \left(1 - y^2\right)^{\beta-1} \quad \text{and} \tag{2}$$

$$F^{-1}(u) = b\left(1 - \sqrt{1 - u^{\frac{1}{\beta}}}\right), \tag{3}$$

respectively, where $y = 1 - x/b$, $0 < x < b$, $b > 0$, $0 < u < 1$, and $0 < \beta < 1$.

Topp and Leone [1] impose the restriction $0 < \beta < 1$ because their goal was to study a distribution with a J-shaped probability density function. Specifically, they were interested in the case where $f(x) > 0$, $f'(x) < 0$ and $f''(x) > 0$ for all $0 < x < b$, where f' and f'' are

the first and second derivatives of Equation 2, respectively. Currently, studies involving the TL distribution consider $b > 0$, so Equation 2 has a unimodal shape when $b > 1$.

The TL distribution has been studied by various authors. Nadarajah and Kotz [2] derive closed-form analytical expressions for the hazard rate function and moments of this distribution. For example, Ghitany et al. [3] study reliability measures of the TL distribution. Zhou et al. [4] study the distribution of the sum, product, and quotients of TL random variables. Kotz and Seier [5] explore the kurtosis behavior of the TL distribution by using the spread-spread function. Nadarajah [6] studies different distributions with a bathtub hazard rate function, including the TL distribution. Genç [7] studies the moments of the order statistics of the TL distribution.

The TL distribution's analytical simplicity and shape flexibility (J-shaped and unimodal) make it a valuable alternative for fitting bounded data. However, there are occasions when the data exhibit high levels of skewness and kurtosis, which can impair the performance of this distribution in fitting the data. In the literature, various approaches can be found to introduce one or more parameters to an arbitrary baseline distribution, which results in greater flexibility in terms of skewness and kurtosis. See, for example [8–15].

In this context, Iriarte et al. [16] propose the Lambert-F distribution generator, whose CDF is defined by

$$G(x; \alpha, \eta) = 1 - \left[1 - F(x; \eta)\right] \alpha^{F(x;\eta)}, \quad (4)$$

where $F(\cdot; \eta)$ is the CDF of an arbitrary baseline distribution with parameter vector η, $\alpha \in (0, e)$ is an extra shape parameter, and $e \approx 2.718$ is the Euler's number.

The corresponding PDF and QF are given by

$$g(x; \alpha, \eta) = f(x; \eta) \alpha^{F(x;\eta)} \left\{1 - \log(\alpha) \left[1 - F(x; \eta)\right]\right\} \quad \text{and} \quad (5)$$

$$G^{-1}(p; \alpha, \eta) = \begin{cases} F^{-1}\left(\dfrac{W_0\left(\frac{\log(\alpha)(p-1)}{\alpha}\right)}{\log(\alpha)}; \eta\right), & \text{if } \alpha \neq 1, \\ F^{-1}(p), & \text{if } \alpha = 1, \end{cases} \quad (6)$$

respectively, where $p \in (0, 1)$, $F(\cdot; \eta)$, $F^{-1}(\cdot; \eta)$, and $f(\cdot; \eta)$ are the CDF, the QF, and the PDF of the baseline distribution, and $W_0(\cdot)$ is the principal branch of the Lambert W function.

Lambert-F distributions exhibit wider ranges of skewness than the corresponding baseline distributions, making them an important alternative for modeling skewed data. See, for example [17–19].

This study presents a new distribution for bounded data, the Lambert-Topp-Leone (LTL) distribution. It arises by considering a baseline TL distribution in the Lambert-F distribution generator. In consequence, an extension of the TL distribution that includes an additional shape parameter, offering more flexibility in terms of skewness and kurtosis, is defined. The new distribution can serve as an alternative to the beta [20] and Kumaraswamy [21] distributions, or to recently introduced distributions such as the NXLD [22], which are widely used for analyzing proportion data. In the context of lifetime data analysis, this distribution can serve as an alternative to the Weibull [23], Birnbaum-Saunders

[24], generalized exponential [25], and generalized Rayleigh [26] distributions.

The rest of the article is organized as follows: In Section 2, the new distribution is introduced, its behavior in terms of skewness and kurtosis is described, and the hazard rate function is analyzed. Section 3 discusses parameter estimation using the maximum likelihood method and presents a simulation study to evaluate the performance of the estimators. Section 4 presents several real-world applications, including proportion data on household spending allocated to food and lifetime data on fatigue. These applications demonstrate that the new distribution may provide a better fit to the data compared to traditional distributions. Finally, concluding remarks are presented in Section 6.

2 The new distribution

In this section, we introduce the Lambert-Topp-Leone (LTL) random variable and study some of its main properties.

2.1 The Lambert-Topp-Leone random variable

In this section, we introduce the LTL random variable and present its PDF, CDF, and QF.

Proposition 1. *A random variable X follows the Lambert-Topp-Leone (LTL) distribution, with parameters $\alpha \in (0, e)$ and $\beta, b > 0$, denoted as $X \sim \mathrm{LTL}(\alpha, \beta, b)$, if its CDF, QF, and PDF are given, respectively, by*

$$G(x; \alpha, \beta, b) = 1 - \left[1 - \left(1 - y^2\right)^\beta\right] \alpha^{\left(1 - y^2\right)^\beta}, \quad (7)$$

$$G^{-1}(p; \alpha, \beta, b)$$
$$= \begin{cases} b\left[1 - \sqrt{1 - \left[\dfrac{1}{\log(\alpha)} W_0\left(\dfrac{\log(\alpha)(p - 1)}{\alpha}\right)\right]^{\frac{1}{\beta}}}\right], \\ \quad \text{if } \alpha \in (0, 1) \cup (1, e), \\ b\left[1 - \sqrt{1 - p^{\frac{1}{\beta}}}\right], \text{ if } \alpha = 1, \end{cases} \quad (8)$$

$$g(x; \alpha, \beta, b) = \frac{2\beta}{b} y \left(1 - y^2\right)^{\beta - 1} \alpha^{\left(1 - y^2\right)^\beta} \left\{1 - \log(\alpha)\left[1 - \left(1 - y^2\right)^\beta\right]\right\}, \quad (9)$$

where $y = 1 - x/b$, $0 < x < b$, $0 < p < 1$, $b > 0$ and $0 < \alpha < e$.

Proof. The result is a direct consequence of substituting Equations 1–3 into Equations 4–6.

Figure 1 presents some LTL pdf curves for $b = 1$ and different choices of the parameters α and β. In the figure, it can be seen that the LTL PDF can present J-shape, inverted N-shape, and unimodal shape.

Using Equation 8, LTL pseudo-random numbers can be generated straightforwardly by applying the inversion method. To facilitate this, we propose Algorithm 1.

Figure 2 presents the histogram and cumulative distribution function for six sets of LTL pseudo-random numbers (three of
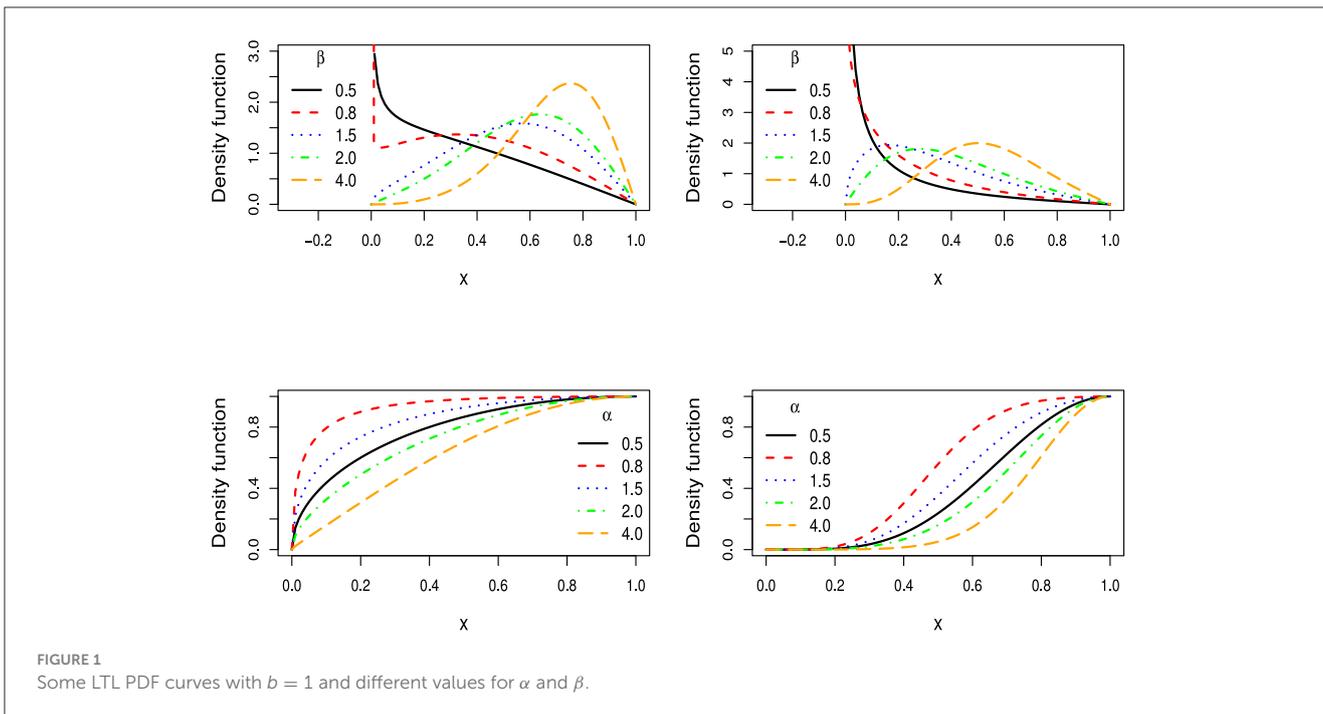
FIGURE 1
Some LTL PDF curves with $b = 1$ and different values for α and β.

```
Require: 0 < α < e, β > 0 and b > 0.
Ensure: A random number X with an LTL distribution
  of parameters α, β and b.
  Generate a random number U from a uniform
  distribution on the interval [0,1].
  Compute X = F_X^{-1}(U; α, β), where F_X^{-1}(·; ·, ·) is as in
  Equation 8.
  return X
```

Algorithm 1. Lambert-Topp-Leone random number generation.

size 50 and three of size 100,000), considering different values for the parameters α, β, and b. R codes [27] for the computation of Equations 7–9 are provided in Appendix 1.1. We use the function lamW::lambertW0() in the R programming language to compute the principal branch of the Lambert function W [28].

2.2 Moments and description of skewness and kurtosis

In this section, we derive the raw moments of the LTL distribution and use this result to describe the skewness and kurtosis behavior of the distribution.

Proposition 2. *Let $X \sim$ LTL(α, β, b). Then, for $r = 1, 2, \ldots$, the rth moment of X is given by*

$$\mathbb{E}(X^r) = b^r a_r(\alpha, \beta), \tag{10}$$

where

$$a_r(\alpha, \beta) = \int_0^1 \left(1 - \sqrt{1 - u^{1/\beta}}\right)^r \alpha^u [1 - \log(\alpha)(1 - u)] \, du.$$

Proof. From Equation 9, and by definition of expectation, we have that

$$\mathbb{E}(X^r) = \int_0^b x^r \frac{2\beta}{b} y \left(1 - y^2\right)^{\beta-1} \alpha^{(1-y^2)^\beta} \left\{1 - \log(\alpha) \left[1 - \left(1 - y^2\right)^\beta\right]\right\} dx.$$

Thus, the result follows by considering the change of variable $u = (1 - y^2)^\beta$.

We employ the integrate() function of the R programming language to compute the function $a_r(\alpha, \beta)$ defined in Equation 10. Accordingly, the int() function is introduced to streamline the computation of $a_r(\alpha, \beta)$. For further details, refer to Appendix 1.1.

Corollary 1. *Let $X \sim$ LTL(α, β). Then, the mean ($\mathbb{E}(X)$), variance ($\mathbb{V}ar(X)$), and skewness (S) and kurtosis (K) coefficients of X are given by*

$$\mathbb{E}(X) = b a_1, \quad \mathbb{V}ar(X) = b^2(a_2 - a_1^2),$$

$$S = \frac{a_3 - 3a_1 a_2 + a_1^3}{\left(a_2 - a_1^2\right)^{3/2}} \quad and \quad K = \frac{a_4 - 4a_1 a_3 + 6a_1^2 a_2 - 3a_1^4}{(a_2 - a_1^2)^2},$$

where $a_j = a_j(\alpha, \beta)$, with $j = 1, 2, 3, 4$, is as in Proposition 2.

Figure 3 presents plots of the skewness and kurtosis coefficients of the LTL distribution. As shown in the figure, the highest values of skewness and kurtosis are obtained for small values of α and β.

FIGURE 2
Histograms and cumulative distributions for six sets of LTL pseudo-random numbers generated with different values of α, β, and b, using sample sizes of $n = 50$ and $n = 100,000$. In each panel, the red line represents the LTL PDF or the LTL CDF, as appropriate.
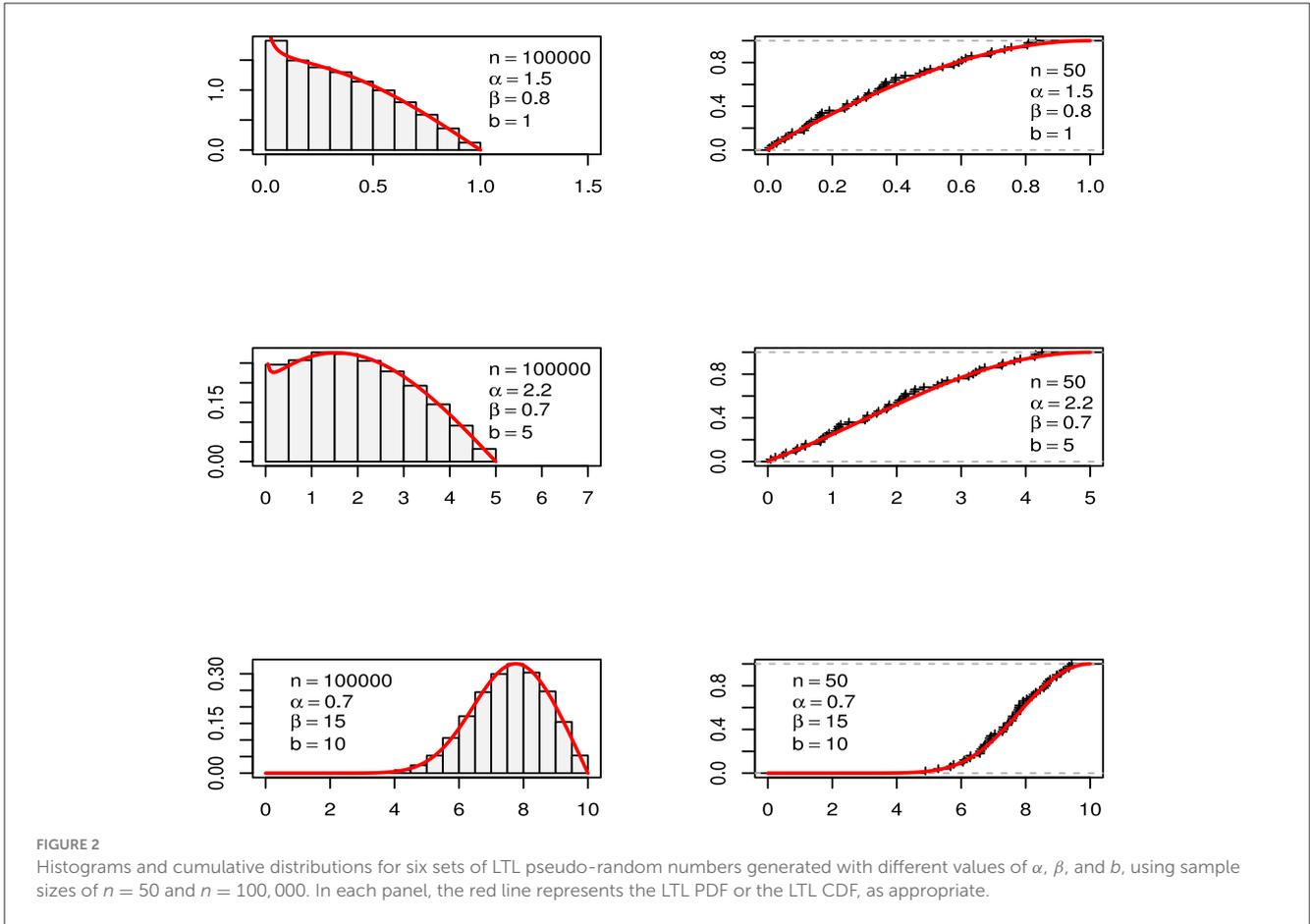
2.3 Hazard rate function

As a direct consequence of Proposition 1, the following corollary proposes the hazard rate function (HRF) of the LTL distribution.

Corollary 2. *Let $X \sim \mathrm{LTL}(\alpha, \beta, b)$. The HRF of X is given by*

$$h_X(x) = \lim_{\triangle x \to 0} \frac{\mathbb{P}(x \leq X < x + \triangle x | X \geq x)}{\triangle x}$$

$$= \frac{2\beta y \left(1 - y^2\right)^{\beta-1} \left\{1 - \log(\alpha)\left[1 - \left(1 - y^2\right)^{\beta}\right]\right\}}{b\left[1 - \left(1 - y^2\right)^{\beta}\right]},$$

where $y = 1 - x/b$, $0 < x < b$, $b > 0$, $0 < \alpha < e$ and $\beta > 0$.

Remark 1. *From Corollary 2, it can be seen that:*

1. If $\alpha = 1$, then

$$h_X(x) = \frac{2\beta}{b} \frac{y\left(1 - y^2\right)^{\beta-1}}{1 - \left(1 - y^2\right)^{\beta}}, \quad y = 1 - \frac{x}{b}, \ 0 < x < b, \ b, \beta > 0,$$

that is the LT hrf.

2. If $h_{\mathrm{TL}}(x)$ represents the TL hrf, then

$$\lim_{x \to 0^+} \frac{h_X(x)}{h_{\mathrm{TL}}(x)} = 1 - \log(\alpha) \quad and \quad \lim_{x \to 1^-} \frac{h_X(x)}{h_{\mathrm{TL}}(x)} = 1.$$

Thus, the LTL hrf can be understood as a modification in the early times of the TL hrf.

Figure 4 presents several LTL hazard rate function (hrf) curves with $b = 5$ and different choices of α and β. As shown in the figure, the LTL hrf can exhibit both the bathtub shape (typical of the TL distribution) and an increasing shape.

The bathtub shape of a HRF characterizes a specific pattern of failure or event rates over time, commonly observed in reliability and survival analysis. This pattern consists of three distinct phases: (i) an initial phase with a decreasing hazard rate, dominated by early failures due to initial defects or weak components; (ii) a middle phase with an approximately constant hazard rate, representing the normal operational life of the system or product; and (iii) a final phase with an increasing hazard rate, attributed to wear-out, aging, or system deterioration.

Among the distributions in the literature that exhibit this shape, notable examples include the exponentiated Weibull [29], generalized gamma [30], beta [20], and Kumaraswamy [21] distributions. In Section 4, these distributions are utilized to assess the comparative performance of the LTL distribution in modeling real-world data.

3 Parameter estimation

In this section, we discuss the maximum likelihood (ML) estimation of the LTL distribution and evaluate its behavior using Monte Carlo simulation.

FIGURE 3
Plots of the skewness and kurtosis coefficients of the LTL distribution.

FIGURE 4
Plots of the hrf for the LTL distribution with- ($b = 5, \beta = 0.5$) in the **left panel** and ($b = 5, \beta = 2.5$) in the **right panel**.
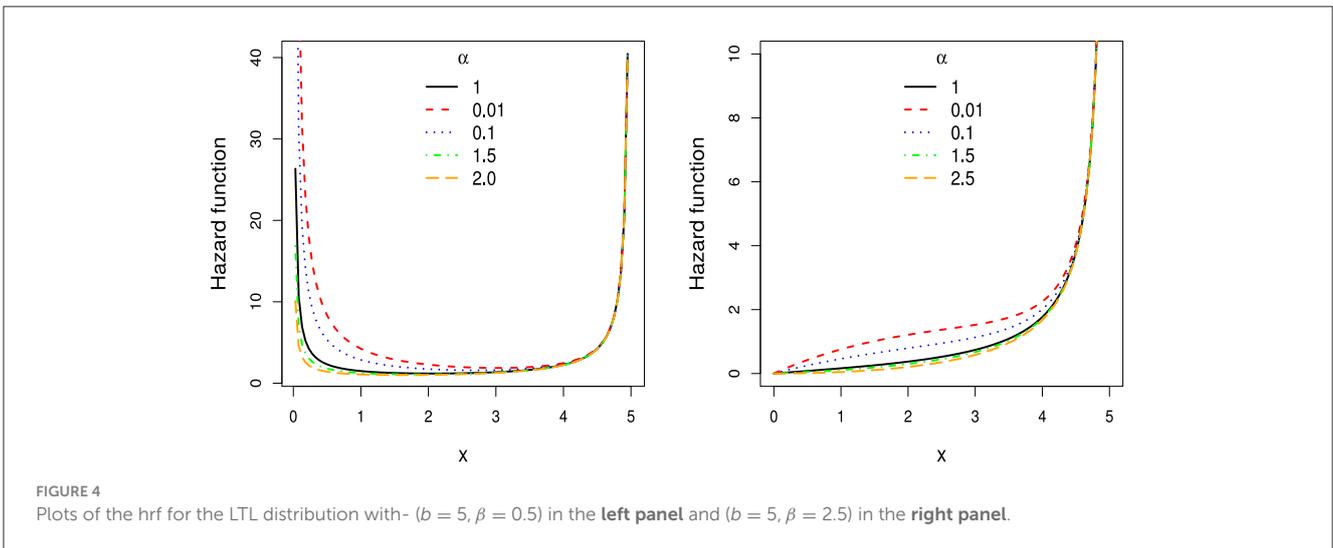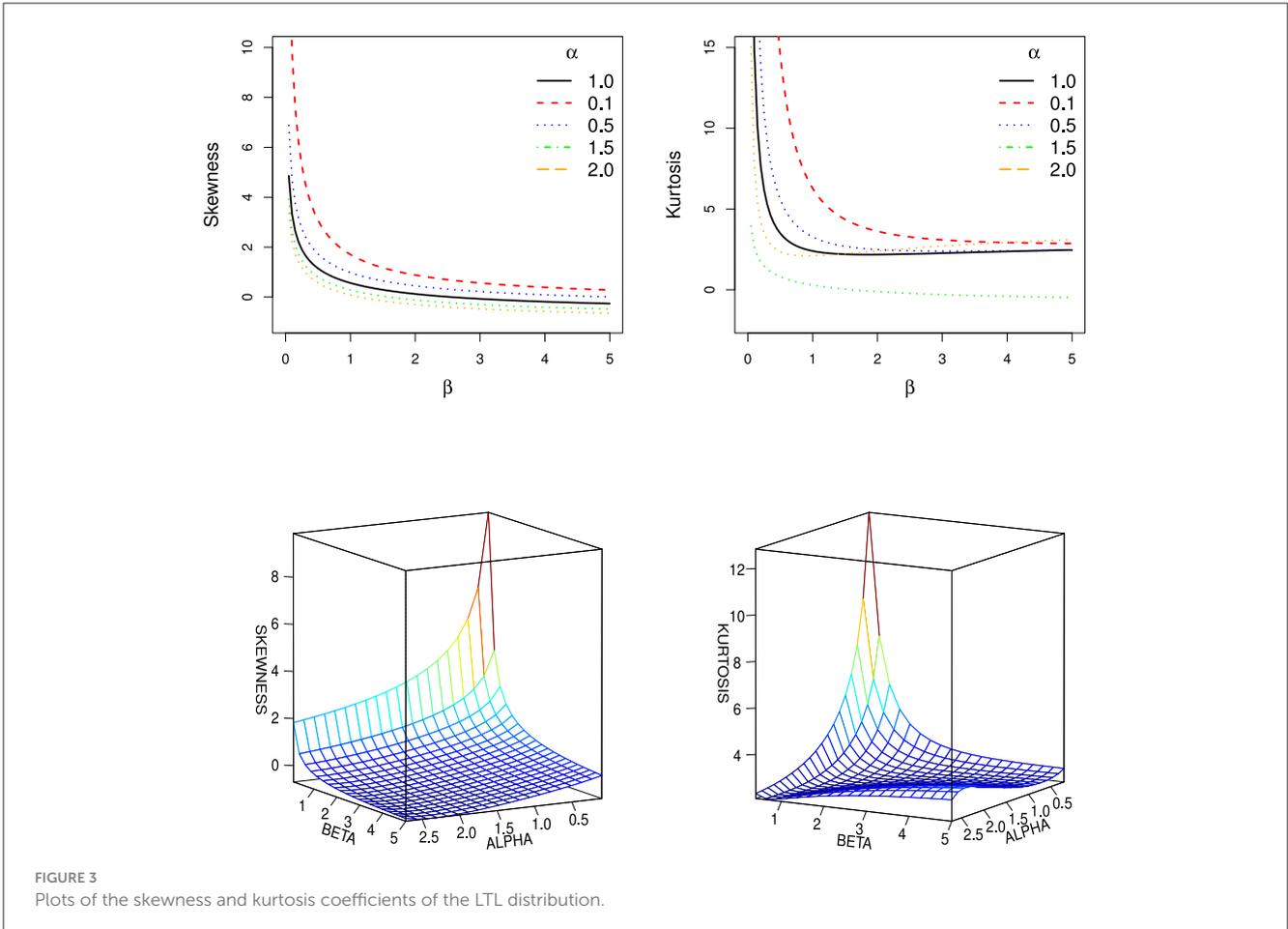
3.1 Maximum likelihood estimation

In this section, we discuss the maximum likelihood estimation for the LTL distribution, considering both cases where b are known and unknown. The case where b is known is particularly useful in scenarios where proportion data are being fitted, in which case $b = 1$ can be assumed. On the other hand, the case where b

is unknown is useful in contexts that require fitting non-negative data, such as life data.

3.1.1 The case where b is known

Let $X_1, X_2, \ldots, X_n$ be a random sample of $X \sim \text{LTL}(\alpha, \beta, b)$, where b is known. Then, the log-likelihood function associated with

$\theta = (\alpha, \beta)$ is written as

$$\ell(\theta; \mathrm{x}) = n\log(2) + n\log(\beta) - n\log(b)$$
$$+ \sum_{i=1}^{n} \log(y_i) + (\beta - 1) \sum_{i=1}^{n} \log(1 - y_i^2)$$
$$+ \log(\alpha) \sum_{i=1}^{n} (1 - y_i^2)^\beta$$
$$+ \sum_{i=1}^{n} \log\{1 - \log(\alpha)[1 - (1 - y^2)^\beta]\}, \quad (11)$$

where $\mathrm{x} = (x_1, \ldots, x_n)^T$ is the vector of observed values and $y_i = 1 - x_i/b$, for $i = 1, \ldots, n$.

The ML estimator $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ of $\theta = (\alpha, \beta)$ can be obtained by solving the system of equations given by the following score equations:

$$0 = \frac{\partial \ell(\theta; \mathrm{x})}{\partial \alpha} = \sum_{i=1}^{n} (1 - y_i^2)^\beta - \sum_{i=1}^{n} \frac{1 - (1 - y_i^2)^\beta}{1 - \log(\alpha)[1 - (1 - y_i^2)^\beta]},$$

$$0 = \frac{\partial \ell(\theta; \mathrm{x})}{\partial \beta} = \frac{n}{\beta} + \sum_{i=1}^{n} \log(1 - y_i^2)$$
$$+ \log(\alpha) \sum_{i=1}^{n} \log(1 - y_i^2)(1 - y_i^2)^\beta$$
$$+ \log(\alpha) \sum_{i=1}^{n} \frac{\log(1 - y_i^2)(1 - y_i^2)^\beta}{1 - \log(\alpha)[1 - (1 - y_i^2)^\beta]}.$$

Since the root of this system do not have a closed form, the ML estimates for $\theta = (\alpha, \beta)$ have to be obtained using numerical methods.

The standard errors of $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ can be obtained as the square roots of the elements of the diagonal of the matrix

$$\mathrm{K}^{-1}(\hat{\theta}) = \left\{ -\left. \begin{pmatrix} \dfrac{\partial^2 \ell(\theta; \mathrm{x})}{\partial \alpha^2} & \dfrac{\partial^2 \ell(\theta; \mathrm{x})}{\partial \alpha \partial \beta} \\ \dfrac{\partial^2 \ell(\theta; \mathrm{x})}{\partial \beta^2} \end{pmatrix} \right|_{\theta = \hat{\theta}} \right\}^{-1},$$

where

$$\frac{\partial^2 \ell(\theta; \mathrm{x})}{\partial \alpha^2} = -\frac{1}{\alpha^2} \sum_{i=1}^{n} (1 - y_i^2)^\beta$$
$$+ \frac{1}{\alpha^2} \sum_{i=1}^{n} \frac{1 - (1 - y_i^2)^\beta}{1 - \log(\alpha)[1 - (1 - y_i^2)^\beta]}$$
$$- \frac{1}{\alpha^2} \sum_{i=1}^{n} \frac{[1 - (1 - y_i^2)^\beta]^2}{\{1 - \log(\alpha)[1 - (1 - y_i^2)^\beta]\}^2},$$

$$\frac{\partial^2 \ell(\theta; \mathrm{x})}{\partial \alpha \partial \beta} = \frac{1}{\alpha} \sum_{i=1}^{n} (1 - y_i^2)^\beta \log(1 - y_i^2)$$
$$+ \frac{1}{\alpha} \sum_{i=1}^{n} \frac{(1 - y_i^2)^\beta \log(1 - y_i^2)}{1 - \log(\alpha)[1 - (1 - y_i^2)^\beta]}$$

$$+ \frac{\log(\alpha)}{\alpha} \sum_{i=1}^{n} \frac{(1 - y_i^2)^\beta [1 - (1 - y_i^2)^\beta] \log(1 - y_i^2)}{\{1 - \log(\alpha)[1 - (1 - y_i^2)^\beta]\}^2},$$

$$\frac{\partial^2 \ell(\theta; \mathrm{x})}{\partial \beta^2} = -\frac{n}{\beta} + \log(\alpha) \sum_{i=1}^{n} (1 - y_i^2)^\beta \log^2(1 - y_i^2)$$
$$+ \log(\alpha) \sum_{i=1}^{n} \frac{(1 - y_i^2)^\beta \log^2(1 - y_i^2)}{1 - \log(\alpha)[1 - (1 - y_i^2)^\beta]}$$
$$- \log^2(\alpha) \sum_{i=1}^{n} \frac{(1 - y_i^2)^{2\beta} \log^2(1 - y_i^2)}{\{1 - \log(\alpha)[1 - (1 - y_i^2)^\beta]\}^2}.$$

Alternatively, to obtain ML estimates of $\theta = (\alpha, \beta)$, it is possible to solve the problem of maximizing Equation 11 with the help of some numerical optimization routine such as the function `stats:optim()` of the R programming language [27]. In this case, minimizing the negative log-likelihood, this function returns the ML estimates and the numerical Hessian matrix. R codes for the computation of Equation 11 and for its maximization are provided in Appendix 1.1.

3.1.2 The case where b is unknown

Let $X_1, X_2, \ldots, X_n$ be a random sample of $X \sim \mathrm{LTL}(\alpha, \beta, b)$, where α, β, and b are unknown. In this case, the ML estimator for b is $\hat{b} = T(\mathrm{x}) = \max(\mathrm{x})$, where $\mathrm{x} = (x_1, x_2, \ldots, x_n)^T$ is the vector of observed values.

It should be noted that the PDF of $T(\mathrm{X}) = \max(\mathrm{X})$ is $g_{T(\mathrm{X})}(x) = ng(x)[G(x)]^{n-1}$, where $G(x) = G(x; \alpha, \beta, b)$ and $g(x) = g(x; \alpha, \beta, b)$ are as in Equations 7, 9, respectively. So, the PDF of $T(\mathrm{X})$ can be written as

$$g_{T(\mathrm{X})}(x) = \frac{2n\beta}{b} y (1 - y^2)^{\beta - 1} \alpha^{(1 - y^2)^\beta}$$
$$\left\{ 1 - \log(\alpha) \left[1 - (1 - y^2)^\beta \right] \right\}$$
$$\times \left\{ 1 - \left[1 - (1 - y^2)^\beta \right] \alpha^{(1 - y^2)^\beta} \right\}^{n-1},$$

where $0 < x < b, 0 < \alpha < e, \beta > 0, b > 0$, and $n \in \mathbb{N}$. Thus, the rth raw moment of $T(\mathrm{X})$ is given by

$$\mu_r = \mathbb{E}[T(\mathrm{X})^r] = \int_0^b x^r g_{T(\mathrm{X})}(x)\, dx, \quad r = 1, 2, 3, \ldots$$

Then, the variance of $T(\mathrm{X})$ is $V[T(\mathrm{X})] = \mu_2 - \mu_1^2$, so the standard error of $\hat{b} = T(\mathrm{x}) = \max(\mathrm{x})$ is

$$\mathrm{se}(\hat{b}) = \sqrt{V[T(\mathrm{x})]} = \sqrt{\mu_2 - \mu_1^2}. \quad (12)$$

We compute Equation 12 in the R programming language, see Appendix 1.1 for code details.

Once the ML estimate of b has been calculated, the ML estimates of α and β can be calculated considering b as known, as illustrated in Section 3.1.1.

3.2 Simulation studies

This section presents the results of two simulation studies designed to evaluate the performance of the ML estimators for

α, β, b	Features of LTL distribution
2.0, 0.5, 1.0	Bounded to the interval (0, 1) and with J-shaped PDF
2.0, 0.5, 5.0	Bounded to the interval (0, 5) and with J-shaped PDF
0.1, 2.0, 1.0	Bounded to the interval (0, 1) and with unimodal PDF
0.1, 2.0, 5.0	Bounded to the interval (0, 5) and with unimodal PDF

the LTL distribution parameters and assess the overall fitting performance in the presence of outliers.

3.2.1 Performance evaluation of parameter estimators

In this section, a simulation study is carried out to evaluate the performance of ML estimators for the parameters of the LTL distribution. We generated 1,000 random samples from the LTL distribution for each of the sample sizes $n = 100, 200, 300, 500, 1,000$, and 2,000. The pseudo-random numbers were generated using the inversion method discussed in Section 2.1, and we considered the simulation scenarios presented in Table 1.

For each generated sample, we compute the ML estimator for the parameters of the LTL distribution, considering the results of Section 3.1.2. Table 2 reports the average estimates and standard deviations obtained in each simulation scenario. From the table, it can be seen that the average estimates approach the true values of the parameters and the standard deviations decrease to 0 as the sample size increases, which demonstrates the consistency of the estimators.

3.2.2 Performance evaluation of the LTL distribution under outlier observations

The objective of this section is to evaluate the performance of the LTL distribution in the presence of outlier observations. Specifically, we analyze the distribution's robustness and fitting accuracy through simulations under different scenarios and sample sizes.

We consider two distinct simulation scenarios for generating pseudo-random numbers from the LTL distribution. In Scenario 1, the parameters are ($\alpha = 2, \beta = 3, b = 5$), where the PDF is unimodal and exhibits a Fisher kurtosis value of 2.706 (PDF with light tails). In Scenario 2, the parameters are ($\alpha = 0.1, \beta = 0.5, b = 5$), where the PDF is decreasing and exhibits a kurtosis value of 14.524 (PDF with heavy tails). In both scenarios, 1,000 pseudo-random samples were generated from the LTL distribution for sample sizes $n = 100, 200, 300, 400$, and 500.

For each generated sample, outliers were identified using the interquartile range (IQR) criterion. Specifically, an observation was considered an outlier if it fell outside the range (i, s), where $i = q_1 - 1.5(q_3 - q_1)$, $s = q_3 + 1.5(q_3 - q_1)$, and q_1 and q_2 represent the first and third sample quartiles, respectively.

To evaluate the performance of the LTL distribution, the Kolmogorov–Smirnov (KS) goodness-of-fit test was applied to each sample. The test assesses the agreement between the empirical distribution and the theoretical LTL distribution. A sample was

deemed appropriately fitted if the KS test yielded a p-value greater than 0.05. The ML estimates of the LTL distribution parameters were obtained based on the results from Section 3.1.2.

For each set of 1,000 simulated samples, the following metrics were calculated:

1. Outlier presence: the percentage of samples that exhibited at least one outlier based on the IQR criterion.
2. Success percentage: the percentage of samples appropriately fitted with the LTL distribution (p-value>0.05 in the Kolmogorov–Smirnov (KS) test [31]).

These metrics provide insights into the LTL distribution's capability to handle outlier observations and its overall fitting performance across different simulation scenarios and sample sizes.

Table 3 presents the results for each set of 1,000 simulated samples. It can be observed that in Scenario 1, the predominance of outliers is lower compared to Scenario 2, primarily due to differences in the weight of the PDF tails in each scenario. Notably, the success rates exceed 92%, indicating that the presence of outliers does not adversely affect the performance of the LTL distribution. Interestingly, the success rate increases as the proportion of samples containing outliers grows.

4 Data analysis

In this section, we illustrate the utility of the LTL distribution through two application examples in different real-world settings. In the first example, we employ the LTL distribution to fit a proportion data set, comparing its performance to the beta (B) [32] and Kumaraswamy (K) [21] distributions, which are two of the earliest alternatives used to fit this type of data. In the second example, we employ the LTL distribution to fit a fatigue life data set, where we compare its performance to six lifetime distributions: the Weibull (W) [23], generalized exponential (GE) [25], Birnbaum-Saunders (BS) [24], generalized Rayleigh (GR) [26], exponentiated Weibull (EW) [29], generalized gamma (GG) [30], Rayleigh (R) [20], and exponential (E) [20] distributions. The PDFs of these distributions are presented in Appendix 1.2.

We employ the Kolmogorov–Smirnov (KS) and Cramer-von Mises (CvM) goodness-of-fit tests to assess the quality of the fits. For this, we use the `stats::ks.test()` and `goftest::cvm.test()` functions of the R programming language. In addition, we employ the excess mass test proposed by Ameijeiras-Alonso et al. [33] to show the unimodality of the data considered in each application example. More specifically, we test the hypothesis H_0 (the data has exactly one mode) versus the alternative hypothesis H_1 (the data has at least two modes). For this, we use the function `multimode::modetest` [34] in the R programming language.

4.1 Family spending on food

Family spending on food is an important component of the family budget, and its amount can vary significantly by factors such as geographic location, family size, and dietary preferences, to name a few. In industry, knowledge about household food spending

TABLE 2 Averages and standard deviations of the ML estimates obtained in each simulation scenario under the different sample sizes.

Sample size	Scenario			Results					
n	α	β	b	$\overline{\hat{\alpha}}$	SD $(\hat{\alpha})$	$\overline{\hat{\beta}}$	SD $(\hat{\beta})$	$\overline{\hat{b}}$	SD $(\hat{b})$
100	0.5	4.0	1.0	0.793	0.345	3.832	1.669	0.941	0.031
200	0.5	4.0	1.0	0.714	0.218	3.895	0.433	0.958	0.022
300	0.5	4.0	1.0	0.673	0.161	3.909	0.344	0.965	0.019
500	0.5	4.0	1.0	0.626	0.115	3.938	0.251	0.972	0.014
1,000	0.5	4.0	1.0	0.592	0.082	3.952	0.181	0.980	0.010
2,000	0.5	4.0	1.0	0.561	0.053	3.961	0.124	0.986	0.007
100	0.5	4.0	5.0	0.792	0.323	3.847	0.617	4.716	0.146
200	0.5	4.0	5.0	0.712	0.208	3.894	0.436	4.787	0.111
300	0.5	4.0	5.0	0.676	0.158	3.919	0.333	4.814	0.097
500	0.5	4.0	5.0	0.628	0.115	3.940	0.262	4.859	0.075
1,000	0.5	4.0	5.0	0.586	0.078	3.958	0.178	4.901	0.050
2,000	0.5	4.0	5.0	0.560	0.055	3.969	0.124	4.930	0.037
100	0.1	2.0	1.0	0.267	0.227	1.984	0.246	0.857	0.064
200	0.1	2.0	1.0	0.228	0.137	1.991	0.163	0.882	0.056
300	0.1	2.0	1.0	0.204	0.111	1.996	0.136	0.894	0.052
500	0.1	2.0	1.0	0.172	0.058	1.997	0.095	0.920	0.042
1,000	0.1	2.0	1.0	0.147	0.033	1.997	0.063	0.943	0.029
2,000	0.1	2.0	1.0	0.136	0.018	1.998	0.047	0.955	0.019
100	0.1	2.0	5.0	0.274	0.237	1.982	0.256	4.263	0.304
200	0.1	2.0	5.0	0.233	0.161	1.990	0.179	4.394	0.285
300	0.1	2.0	5.0	0.211	0.137	1.990	0.141	4.473	0.269
500	0.1	2.0	5.0	0.175	0.060	1.992	0.089	4.592	0.221
1,000	0.1	2.0	5.0	0.148	0.033	1.992	0.062	4.708	0.147
2,000	0.1	2.0	5.0	0.136	0.018	1.993	0.048	4.775	0.097

helps companies make informed decisions about products, pricing, marketing, and distribution strategies. This allows them to better meet the needs of consumers and remain competitive in the market.

In this section, we analyze a dataset of 1,519 observations on the proportion of the British household budget allocated to food. These data are available as BudgetUK (wfood) in the Ecdat package of the R programming language [35]. A descriptive analysis reveals that the minimum and maximum values are 0.0571 and 0.7890, with Fisher skewness and kurtosis values of 0.147 and 2.899, respectively. In addition, the excess mass test yields a p-value of 0.376, indicating that, at conventional significance levels, the hypothesis that the frequency distribution of these data is unimodal is not rejected. Consequently, the LTL distribution with $b = 1$ appears to be a viable alternative for modeling these data.

Table 4 reports the ML estimates for the parameters of the LTL, B, K, and TL distributions fitted to the data on the proportion of the shared household budget spent on food. The results obtained in the KS and CvM goodness-of-fit tests are also presented in Table 4 for each distribution. Note that the LTL distribution exhibits a larger p-value, indicating that this distribution performs better in fitting the data.

TABLE 3 Percentages of outlier presence and success rates obtained from 1,000 simulated samples for each scenario and sample size considered.

Scenario	n	Presence of outliers (%)	Success rate (%)
1	100	37.2	93.8
	200	53.7	94.7
	300	66.9	95.0
	400	75.7	95.4
	500	80.5	95.5
2	100	98.4	99.6
	200	99.6	99.7
	300	100	99.9
	400	100	100
	500	100	100

By computing probabilities using the LTL distribution, we observe that, in the analyzed population, a family is likely to allocate at most 25%, 50%, and 75% of their shared

TABLE 4 Maximum likelihood (ML) estimates, observed statistics, and *p*-values obtained from goodness-of-fit tests for distributions fitted to the proportion of household expenditure allocated to food.

Distribution	Estimate		KS test		CvM test	
	$\widehat{\alpha}$	$\widehat{\beta}$	Statistic	*p*-value	Statistic	*p*-value
LTL	1.619×10^{-4} (7.550×10^{-5})	4.889 (0.102)	0.015	0.875	0.035	0.953
B	6.910 (0.245)	12.490 (0.451)	0.026	0.226	0.141	0.415
K	3.632 (0.075)	28.468 (1.935)	0.023	0.374	0.221	0.228
TL	1.707 (0.043)	–	0.300	0.000	45.306	0.000

FIGURE 5
Histogram of the proportion of household shared expenditure on food along with the fitted PDFs.

FIGURE 6
Q–Q plot of the fatigue life data.
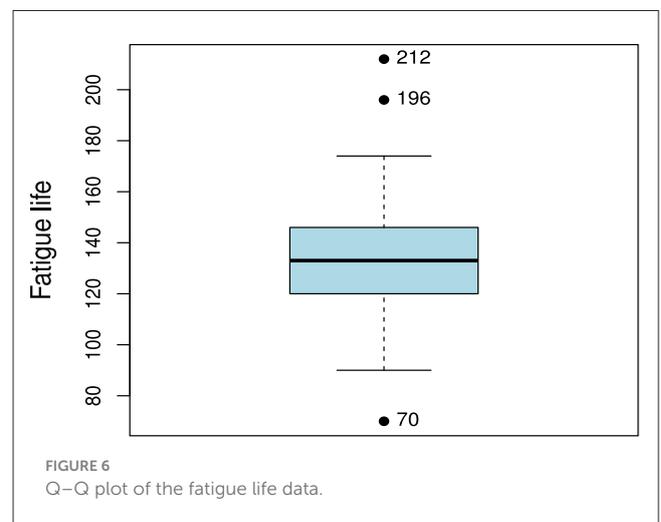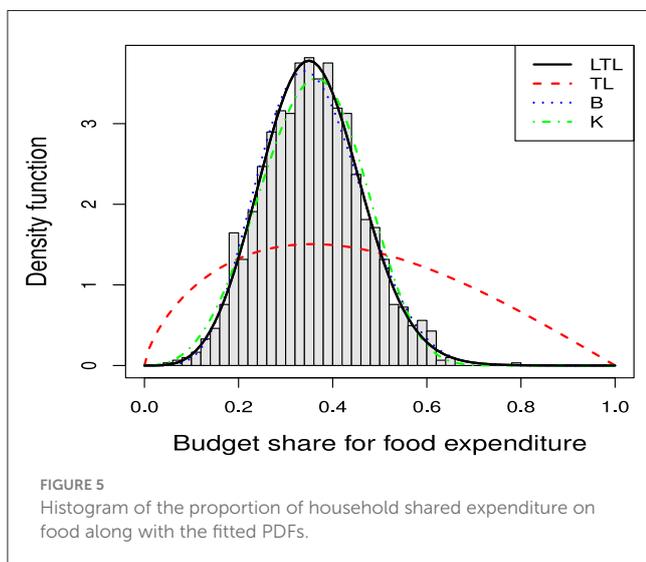
family budget to food with probabilities of 0.1572, 0.9110, and 0.9995, respectively.

Figure 5 presents the histogram of the proportion of household expenditure allocated to food, along with the PDFs of the fitted distributions. The figure illustrates that the LTL PDF aligns more closely with the empirical frequency values, particularly around the mode.

4.2 Fatigue life data

Analyzing fatigue life data is crucial in engineering, materials, and data science applied to predicting the failure time of components and structures subjected to load cycles. These data help to understand how a material's performance degrades over time due to repetitive loading, which is essential for the safety and efficiency of structures such as bridges, aircraft, automobiles, and industrial machinery.

We consider 101 observations on the fatigue life of 6061-T6 aluminum coupons cut parallel to the direction of rolling and oscillated at 18 cycles per second. These data were originally

analyzed in Birnbaum and Saunders [24] and are currently available under the name "fatigue" in the bibs package of the R programming language.

A descriptive analysis of the data indicates that the minimum and maximum observed values are 70 and 212, respectively, with Fisher skewness and kurtosis values of 0.3 and 4.1. In addition, the excess mass test yields a *p*-value of 0.686, suggesting that the hypothesis of unimodality for the frequency distribution of the fatigue life data is not rejected at conventional significance levels. Consequently, the LTL distribution with an unspecified parameter *b* appears to be a viable alternative for modeling these data.

Figure 6 presents the quantile-quantile plot (Q–Q plot) of the fatigue life data, revealing the presence of outliers, two of which are located at the upper end of the data range. It is important to highlight that outliers can adversely affect the modeling process using the LTL distribution as the parameter *b* is estimated as the maximum value of the sample. This estimation may unnecessarily expand the support of the distribution and potentially impact its fit to the remaining data. This raises a critical question: can the LTL distribution still provide an appropriate fit for the fatigue life data despite the presence of these outlier observations?

TABLE 5 Maximum likelihood (ML) estimates, observed statistics, and *p*-values obtained from goodness-of-fit tests for distributions fitted to the fatigue life data.

Distribution	Estimate			KS test		CvM test	
	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{b}$	Statistic	*p*-value	Statistic	*p*-value
LTL	0.027 (0.007)	12.964 (1.538)	212 (8.769)	0.060	0.850	0.043	0.917
EW	3.078 (0.742)	104.335 (15.991)	5.338 (3.325)	0.065	0.783	0.051	0.868
GG	16.901 (8.735)	52.945 (52.612)	2.175 (1.158)	0.061	0.840	0.052	0.864
W	6.073 (0.422)	143.164 (2.485)	–	0.098	0.275	0.180	0.311
GE	281.263 (106.094)	21.771 (1.534)	–	0.108	0.186	0.224	0.226
BS	0.170 (0.011)	131.818 (2.226)	–	0.084	0.459	0.086	0.660
GR	17.522 (4.074)	73.088 (73.088)	–	0.087	0.426	0.099	0.592
R	–	95.908 (4.775)	–	0.378	0.000	4.488	0.000
E	–	133.726 (13.305)	–	0.492	0.000	7.010	0.000

Table 5 reports the ML estimates for the parameters of the LTL, W, GE, BS, GR, EW, GG, R, and E distributions fitted to the fatigue life data. The results obtained in the KS and CvM goodness-of-fit tests are also presented in Table 5 for each distribution. Note that the LTL distribution exhibits a larger *p*-value, indicating that this distribution performs better in fitting the data.

Figure 7 (left panel) presents the histogram of the fatigue life data, along with the PDFs of the fitted distributions. The figure illustrates that the LTL PDF aligns more closely with the empirical frequency values, particularly around the mode. In the right panel of Figure 7, the HRFs of the distributions fitted to the fatigue life data are presented, where a similar behavior can be observed for the LTL, W, EW, and GG distributions, but differing significantly for a fatigue life value >50.

Regarding the earlier question about the presence of outlier observations, the results demonstrate that the LTL distribution performs robustly in modeling the fatigue life data, even under the influence of the outliers. Notably, its performance surpasses that of several established lifetime distributions from the literature, highlighting its flexibility in this challenging scenario.

5 Discussion and further research

As demonstrated throughout this article, the new LTL distribution exhibits several strengths, which are summarized as follows.

The LTL distribution inherits the shapes of the PDF and the HRF from the baseline TL distribution. However, the additional parameter α, introduced by the Lambert-F generator, enhances its flexibility, allowing the LTL distribution to exhibit wider ranges of skewness and kurtosis. This makes it a valuable alternative in scenarios where empirical skewness and/or kurtosis levels cannot be adequately captured by the TL distribution.

The new LTL distribution is defined by three parameters: two shape parameters (α and β) and one support-bounding parameter (b). When $b = 1$, the LTL distribution becomes a suitable alternative for modeling proportion data. In this context, it serves as a natural competitor to the widely used beta (B) and Kumaraswamy (K) distributions. This study demonstrates this by fitting the LTL distribution to a real-world dataset, showing that it can outperform the B and K distributions in modeling the data.

When b is unknown, the LTL distribution is suitable for modeling positive data, such as lifetime data. In this context, it serves as a competitor to well-known lifetime distributions, including the exponentiated Weibull (EW) and generalized gamma (GG) distributions. This study demonstrates this by fitting the LTL distribution to a fatigue failure dataset, showing that it can outperform the EW and GG distributions in modeling the data.

A disadvantage of the new LTL distribution is that, unlike the baseline TL distribution, properties such as the PDF, HRF, and moments have a more complex analytical structure. While properties such as the PDF and HRF are manageable, its moments require numerical integration for their computation. To address this difficulty, we have provided R code for computing these properties, which is accessible to readers.

Regarding the estimation of the support-bounding parameter of the LTL distribution, estimating it as the maximum value of the sample can pose challenges. The presence of outliers may artificially inflate the support, introducing biases in the

FIGURE 7
(Left) Histogram of fatigue life data along with fitted PDFs. **(Right)** HRFs for the distributions fitted to the fatigue life data.
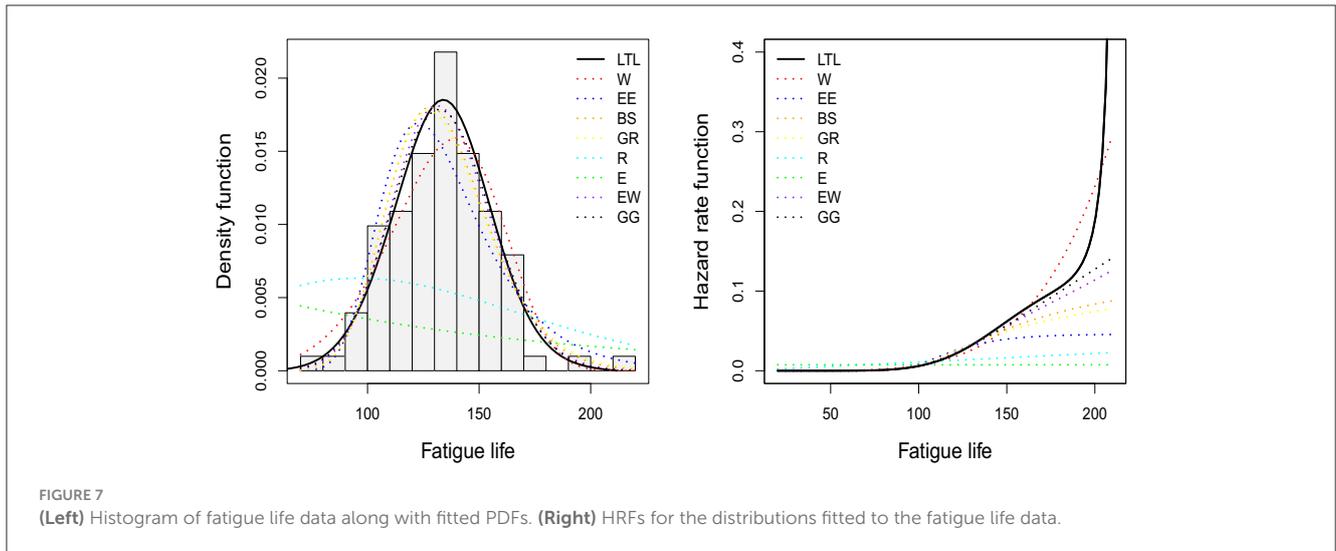
model fit, hindering an accurate representation of the data structure, and compromising the stability and precision of the estimates, particularly in samples with high dispersion. In this study, Monte Carlo simulations have shown that the LTL distribution performs well in the presence of outliers when the samples are generated from an LTL distribution. This suggests that real-world scenarios may exist where the LTL distribution effectively models data with outliers, adequately capturing empirical properties such as unimodality and skewness. This is precisely what is illustrated in Section 4.2.

As ideas for future research, we propose the following:

First, it would be worthwhile to explore alternative estimation methods for the parameters of the LTL distribution. In particular, the results from Section 2.2 could be utilized to derive moment estimators and compare their performance with maximum likelihood estimators.

Second, starting from the LTL quantile function, a quantile regression framework can be formulated. By considering $\theta = G^{-1}(p; \alpha, \beta, b)$, where $G^{-1}(\cdot; \cdot, \cdot, \cdot)$ is defined as in Equation 8, we can express

$$\alpha = \left(\frac{p-1}{\left[1 - \left(1 - \frac{\theta}{b} \right)^2 \right]^\beta} \right)^{\frac{1}{\left[1 - \left(1 - \frac{\theta}{b} \right)^2 \right]^\beta + 1}}.$$

The PDF in Equation 9 can then be re-parameterized in terms of the p-th quantile (with p fixed). This re-parameterized PDF provides the foundation for a regression model to quantify the relationship between a set of explanatory variables and the p-th quantile of a positive response variable.

6 Final comment

This study introduces the Lambert-Topp-Leone (LTL) distribution as a flexible model for bounded and positive skewed data, particularly suitable for proportion and lifetime data.

The LTL distribution extends the Topp-Leone distribution by incorporating an additional shape parameter, thereby enhancing its flexibility to accommodate a wider range of skewness and kurtosis. Our findings suggest that the LTL distribution provides an improved fit over traditional models, as demonstrated through real-world applications involving household spending proportions and fatigue life data. The simulation studies further confirm the consistency and robustness of maximum likelihood estimators for the LTL parameters, adding to its practical utility. A second simulation study has shown that the LTL distribution performs robustly in the presence of outliers.

Given its adaptability, the LTL distribution holds promise for diverse applications beyond the examples explored here, potentially benefiting fields that require accurate modeling of constrained or positively skewed datasets. Future research could focus on exploring Bayesian estimation methods, assessing the performance of the distribution with diverse data types, and developing quantile regression frameworks based on the new distribution to expand its applicability.

Data availability statement

Publicly available datasets were analyzed in this study. The data are available in the R programming language, and details on their location are provided in the article.

Author contributions

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fams.2025.1527833/full#supplementary-material

References

1. Topp CW, Leone FC. A family of J-shaped frequency functions. *J Am Stat Assoc*. (1955) 50:209–19. doi: 10.1080/01621459.1955.10501259

2. Nadarajah S, Kotz S. Moments of some J-shaped distributions. *J Appl Stat*. (2003) 30:311–7. doi: 10.1080/0266476022000030084

3. Ghitany M, Kotz S, Xie M. On some reliability measures and their stochastic orderings for the Topp-Leone distribution. *J Appl Stat*. (2005) 32:715–22. doi: 10.1080/02664760500079613

4. Zhou M, Yang D, Wang Y, Nadarajah S. Some J-shaped distributions: sums, products and ratios. In: *RAMS'06. Annual Reliability and Maintainability Symposium, 2006*. IWashington, DC: EEE (2006), p. 175–81. doi: 10.1109/RAMS.2006.1677371

5. Kotz S, Seier E. Kurtosis of the Topp-Leone distributions. *Interstat*. (2007) 1:1–15.

6. Nadarajah S. Bathtub-shaped failure rate functions. *Qual Quant*. (2009) 43:855–63. doi: 10.1007/s11135-007-9152-9

7. Genç AĪ. Moments of order statistics of Topp-Leone distribution. *Stat Pap*. (2012) 53:117–31. doi: 10.1007/s00362-010-0320-y

8. Marshall AW, Olkin I. A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika*. (1997) 84:641–52. doi: 10.1093/biomet/84.3.641

9. Eugene N, Lee C, Famoye F. Beta-normal distribution and its applications. *Commun Stat-Theory Methods*. (2002) 31:497–512. doi: 10.1081/STA-120003130

10. Zografos K, Balakrishnan N. On families of beta-and generalized gamma-generated distributions and associated inference. *Stat Methodol*. (2009) 6:344–62. doi: 10.1016/j.stamet.2008.12.003

11. Shaw WT, Buckley IR. The alchemy of probability distributions: beyond Gram-Charlier expansions, and a skew-kurtotic-normal distribution from a rank transmutation map. *arXiv*. (2009). [Preprint]. arXiv:0901:0434. doi: 10.48550/arXiv.0901.0434

12. Cordeiro GM, de Castro M. A new family of generalized distributions. *J Stat Comput Simul*. (2011) 81:883–98. doi: 10.1080/00949650903530745

13. Alzaatreh A, Lee C, Famoye F. A new method for generating families of continuous distributions. *Metron*. (2013) 71:63–79. doi: 10.1007/s40300-013-0007-y

14. Mahdavi A, Kundu D. A new method for generating distributions with an application to exponential distribution. *Commun Stat Theory Methods*. (2017) 46:6543–57. doi: 10.1080/03610926.2015.1130839

15. Atchadé MN, Agbahide AA, Otodji T, Bogninou MJ, Moussa Djibril A. A new shifted Lomax-X family of distributions: properties and applications to actuarial and financial data. *Comput J Math Stat Sci*. (2025) 4:41–71. doi: 10.21608/cjmss.2024.307114.1066

16. Iriarte YA, de Castro M, Gómez HW. The Lambert-F distributions class: an alternative family for positive data analysis. *Mathematics*. (2020) 8:1398. doi: 10.3390/math8091398

17. Iriarte YA, de Castro M, Gómez HW. An alternative one-parameter distribution for bounded data modeling generated from the Lambert transformation. *Symmetry*. (2021) 13:1190. doi: 10.3390/sym13071190

18. Iriarte YA, de Castro M, Gómez HW. A unimodal/bimodal skew/symmetric distribution generated from Lambert's transformation. *Symmetry*. (2021) 13:269. doi: 10.3390/sym13020269

19. Varela H, Rojas MA, Reyes J, Iriarte YA. An alternative Lambert-type distribution for bounded data. *Mathematics*. (2023) 11:667. doi: 10.3390/math11030667

20. Johnson NL, Kotz S, Balakrishnan N. *Continuous Univariate Distributions, Vol. 1*, 2nd Edn. New York, NY: Wiley (1994).

21. Kumaraswamy P. A generalized probability density function for double-bounded random processes. *J. Hydrol*. (1980) 46:79–88. doi: 10.1016/0022-1694(80)90036-0

22. Irshad M, Aswathy S, Maya R, Al-Omari AI, Alomani G. A flexible model for bounded data with bathtub shaped hazard rate function and applications. *AIMS Math*. (2024) 9:24810–31. doi: 10.3934/math.20241208

23. Weihull W. A statistical distribution function of wide applicability. *J Appl Mech*. (1951) 18:290–3. doi: 10.1115/1.4010337

24. Birnbaum ZW, Saunders SC. A new family of life distributions. *J Appl Probab*. (1969) 6:319–27. doi: 10.1017/S0021900200032848

25. Gupta RD, Kundu D. Exponentiated exponential family: an alternative to gamma and Weibull distributions. *Biom J*. (2001) 43:117–30. doi: 10.1002/1521-4036(200102)43:1<117::AID-BIMJ117>3.0.CO;2-R

26. Surles J, Padgett W. Inference for reliability and stress-strength for a scaled Burr Type X distribution. *Lifetime Data Anal*. (2001) 7:187–200. doi: 10.1023/A:1011352923990

27. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna (2023). Available at: https://www.R-project.org/ (accessed February 8, 2025).

28. Adler A. *lamW: Lambert-W Function*. R package version 2.2.3. (2015). Available at: https://CRAN.R-project.org/package=lamW (accessed February 8, 2025).

29. Mudholkar GS, Hutson AD. The exponentiated Weibull family: some properties and a flood data application. *Commun Stat-Theory Methods*. (1996) 25:3059–83. doi: 10.1080/03610929608831886

30. Stacy EW. A generalization of the gamma distribution. *Ann Math Stat*. (1962) 33:1187–92. doi: 10.1214/aoms/1177704481

31. Faraway J, Marsaglia G, Marsaglia J, Baddeley A. *goftest: Classical Goodness-of-Fit Tests for Univariate Distributions. R package version 1.2-3*. (2021). Available at: https://CRAN.R-project.org/package=goftest (accessed February 8, 2025).

32. Johnson NL, Kotz S, Balakrishnan N. *Continuous Univariate Distributions, Volume 2*, 2nd Edn. New York, NY: Wiley (1995).

33. Ameijeiras-Alonso J, Crujeiras RM, Rodríguez-Casal A. Mode testing, critical bandwidth and excess mass. *Test*. (2019) 28:900–19. doi: 10.1007/s11749-018-0611-5

34. Ameijeiras-Alonso J, Crujeiras RM, Rodríguez-Casal A. multimode: an R package for mode assessment. *J Stat Softw*. (2021) 97:1–32. doi: 10.18637/jss.v097.i09

35. Croissant Y, Graves S. *Ecdat: Data Sets for Econometrics*. R package version 0.4-2. (2022). Available at: https://CRAN.R-project.org/package=Ecdat (accessed February 8, 2025).