mathematics

MDPI

Article

A Lindley-Type Distribution for Modeling High-Kurtosis Data

Mario A. Rojas * and Yuri A. Iriarte

Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta 1240000, Chile; yuri.iriarte@uantof.cl
* Correspondence: mario.rojas@uantof.cl

Abstract: This article proposes a heavy-tailed distribution for modeling positive data. The proposal arises with the ratio of independent random variables, specifically, a Lindley distribution divided by a beta distribution. This leads to a three-parameter extension of the Lindley distribution capable of modeling high levels of kurtosis. The main structural properties of the proposed distribution are derived. The skewness and kurtosis behavior of the distribution are described. Parameter estimation is discussed under consideration of the moment and maximum likelihood methods. Finally, in order to avoid the parameter non-identifiability problem, a two-parameter version of the proposed distribution is derived. The usefulness of this special case is illustrated by fitting data in two real scenarios.

Keywords: extended slash Lindley distribution; kurtosis; Lindley distribution; Lindley slash distribution; maximum likelihood estimator; moment estimator; probability density function

MSC: 62E10; 62F10

Citation: Rojas, M.A.; Iriarte, Y.A. A Lindley-Type Distribution for Modeling High-Kurtosis Data . *Mathematics* **2022**, *10*, 2240. https://doi.org/10.3390/math10132240

Academic Editor: Jiancang Zhuang

Received: 26 May 2022
Accepted: 24 June 2022
Published: 26 June 2022

1. Introduction

The Lindley (L) distribution [1,2] is one of the alternatives commonly used to model the lifetime of mechanical units. This distribution has various attractive properties, for example: (1) Although it has a single parameter, its probability density function (pdf) can present both monotonic decreasing and unimodal shapes; (2) like other classical distributions, such as Weibull and gamma, the hazard rate function (hrf) of the L distribution exhibits an increasing shape. However, unlike the Weibull and gamma hrf's, the hrf of the L distribution is different from 0 at time $t = 0$ and converges to a finite quantity (a function of the parameter) as $t \to \infty$; and (3) many of its properties have closed forms, which makes it easy to implement computationally.

A random variable Z has a L distribution with shape parameter $\theta > 0$, denoted as $Z \sim \mathrm{L}(\theta)$, if its pdf is given by

$$f_Z(z;\theta) = \frac{\theta^2}{(1+\theta)}(1+z)e^{-\theta z},\ z > 0. \tag{1}$$

Ghitany et al. [3] carry out a detailed study of the properties of the L distribution, among which it is shown that Equation (1) corresponds to the pdf of a mixture distribution with exponential and gamma components. Although the L distribution can model an important variety of data whose histograms exhibit decreasing or unimodal behavior, performance can be poor when the data exhibit high levels of skewness and/or kurtosis. In this sense, it is possible to find in the literature various extensions of the L distribution capable of capturing high levels of skewness and/or kurtosis. Good examples of this are the weighted-Lindley [4], generalized Lindley [5], power-Lindley [6], quasi-Lindley [7], and two-parameter Lindley [8,9] distributions, which are more flexible than the L distribution, especially in terms of skewness. The three-parameter extensions proposed by

Zakerzadeh and Dolati [10], Shanker et al. [11], and Ekhosuehi and Opone [12] are more flexible than the L distribution in both skewness and kurtosis.

An extension of the L distribution oriented especially to model data with high kurtosis levels is the three-parameter Lindley slash (LS) distribution [13]. A random variable Y follows a LS distribution with scale parameter $\sigma > 0$, shape parameter $\theta > 0$ and kurtosis parameter $\alpha > 0$, denoted as $X \sim \text{LS}(\sigma, \theta, \alpha)$, if it can be represented as

$$Y = \sigma \frac{Z}{U^{1/\alpha}}, \tag{2}$$

where $Z \sim \text{L}(\theta)$ and $U \sim \text{Uniform}(0, 1)$ are independent.

From Equation (2), it follows that the LS distribution (when $\sigma = 1$) converges to the L distribution as $\alpha \to \infty$. The pdf of the LS distribution is given by

$$f_Y(y; \sigma, \theta, \alpha) = \frac{\alpha\theta^2}{\sigma(1+\theta)} \int_0^1 \left(1 + \frac{yu}{\sigma}\right) e^{-\frac{\theta y u}{\sigma}} u^\alpha \, du, \quad y > 0.$$

The LS distribution has heavier tails than the L distribution, so it can be used to fit positive data with outliers.

Iriarte and Rojas [14], assuming in Equation (2) that that Z has a power-Lindley distribution, propose a new three-parameter heavy-tailed extension of the L distribution, the slashed-power Lindley distribution. In the literature, it is possible to find a large amount of information on slash-type distributions, which arise assuming certain distributions for the random variables involved in the Equation (2). For some details of slash-type distributions for positive data, see [15–20], among others.

In this article, we introduce a new heavy-tailed extension of the L distribution, the extended slash Lindley (ESL) distribution. Based on the results of Rojas et al. [21], the proposed distribution arises from Equation (2) considering $\sigma = \alpha = 1$ and that U has a beta distribution with mean $\alpha/(\alpha + \beta)$, $\alpha, \beta > 0$. Thus, Equation (2) gives rise to a new distribution with one shape parameter controlling unimodality and two shape parameters controlling kurtosis.

The fact that the ESL distribution has two parameters controlling kurtosis allows this distribution to be able to capture extremely high levels of kurtosis. However, we observe that this same fact generates a problem of non-identifiability of the parameters. In this sense, we propose a two-parameter special case of the ESL distribution provided with one parameter controlling unimodality and one parameter controlling kurtosis. We emphasize that this special case has heavier tails than the L distribution, and even heavier than the LS distribution, so it can be considered as a viable alternative to model positive data with outliers.

The article is organized as follows. In Section 2, we propose the ESL distribution and derive some of its main structural properties. In Section 3, we discuss parameter estimation considering the moment and maximum likelihood methods. In addition, we carry out a simulation study to evaluate the behavior of the moment and maximum likelihood estimators under the two-parameter ESL distribution. Section 4 presents an application with real data in order to illustrate the usefulness of the proposal. Some final comments are considered in Section 5.

2. ESL Distribution

In this section, we propose the ESL distribution and derive its main structural properties.

2.1. Stochastic Representation

Definition 1. *The random variable X follows the extended slash Lindley distribution, denoted as $X \sim \text{ESL}(\theta, \alpha, \beta)$, if it can be represented as $X = Y/U$, where $Y \sim \text{L}(\theta)$ and $U \sim \text{Beta}(\alpha, \beta)$ are independent random variables.*

From Definition 1, it is observed that: (1) If $\beta = 1$, then U has a power distribution [22] on the interval $(0, 1)$. Consequently, X has a nonscaled ($\sigma = 1$) LS distribution; (2) If $\beta = 1$ and $\alpha \to \infty$, then the distribution of X converges to the L distribution.

The following proposition provides the pdf of the ESL distribution.

Proposition 1. *Let $X \sim \mathrm{ESL}(\theta, \alpha, \beta)$. Then, the pdf of X is given by*

$$f_X(x;\theta,\alpha,\beta) = \frac{\theta^2}{(\theta+1)B(\alpha,\beta)} \int_0^1 (1+xw) w^{\alpha} (1-w)^{\beta-1} e^{-\theta x w}\, dw,\ x > 0, \tag{3}$$

where $\theta, \alpha, \beta > 0$ are shape parameters and $B(a,b) = \int_0^1 u^{a-1}(1-u)^{b-1}\, du$ is the beta function.

Proof. From Definition 1, we use the Jacobian method to compute the pdf of X. Specifically, considering the transformations $X = YU^{-1}$ and $W = U$, we have $Y = XW$ and $U = W$. Thus, the determinant of the Jacobian of the transformation is

$$J = \begin{vmatrix} \frac{\partial Y}{\partial X} & \frac{\partial Y}{\partial W} \\ \frac{\partial U}{\partial X} & \frac{\partial U}{\partial W} \end{vmatrix} = w.$$

Hence, the joint pdf of X and W is given by

$$f_{X,W}(x,w) = f_Y(xw;\theta) f_U(w;\alpha,\beta) w, \quad x > 0,\ 0 < w < 1, \tag{4}$$

where $f_Y(\cdot;\theta)$ denotes the pdf of the L distribution with shape parameter $\theta > 0$ and $f_U(\cdot;\beta,\alpha)$ denotes the pdf of the beta distribution with shape parameters $\alpha, \beta > 0$. Then, the marginal distribution of X is obtained directly by integrating Equation (4), which leads to Equation (3). □

The following result shows that the pdf of the ESL distribution can be written in closed form in terms of the gamma, beta and confluent hypergeometric functions. In this sense, we recall the integral representations of these functions given in Andrews [23]. For $\mathrm{R}(a) > 0$ and $\mathrm{R}(b) > 0$, the gamma and beta functions are given by $\Gamma(a) = \int_0^\infty u^{a-1} e^{-u}\, du$ and $\mathrm{B}(a,b) = \int_0^1 u^{a-1}(1-u)^{b-1}\, du$, respectively. For $\mathrm{R}(b) > \mathrm{R}(a) > 0$, the confluent hypergeometric function of the first kind is given by $\mathrm{M}(a,b,z) = \Gamma(b)/[\Gamma(a)\Gamma(b-a)] \int_0^1 e^{zu} u^{a-1} (1-u)^{b-a-1}\, du$. It should be noted that the gamma, beta and confluent hypergeometric functions are implemented in some computer algebra systems (such as MATLAB) and programming languages (such as R [24]). From these functions, we establish the following.

Corollary 1. *Let $X \sim \mathrm{ESL}(\theta, \alpha, \beta)$. Then,*

$$f_X(x;\theta,\alpha,\beta) = \frac{\theta^2 \Gamma(\beta)}{(1+\theta)\mathrm{B}(\alpha,\beta)} \left[r_1(x;\theta,\alpha,\beta) + x r_2(x;\theta,\alpha,\beta) \right], \tag{5}$$

where

$$r_j(x;\theta,\alpha,\beta) = \frac{\Gamma(\alpha+j)\mathrm{M}(\alpha+j, \alpha+\beta+j, -\theta x)}{\Gamma(\alpha+\beta+j)},\ j = 1, 2.$$

Proof. We note that Equation (3) can be written as

$$\begin{aligned} f_X(x;\theta,\alpha,\beta) &= \frac{\theta^2}{(1+\theta)\mathrm{B}(\alpha,\beta)} \left[\int_0^1 w^{\alpha} (1-w)^{\beta-1} e^{-\theta x w}\, dw \right. \\ &\quad \left. + x \int_0^1 w^{\alpha+1} (1-w)^{\beta-1} e^{-\theta x w}\, dw \right], \end{aligned}$$

and the result is obtained by recognizing the confluent hypergeometric function in the previous integrals. □

From Proposition 1 and Corollary 1, we see that

$$\begin{aligned} f_X(x;\theta,\alpha,\beta=1) &= \frac{\alpha\theta^2}{(1+\theta)}\int_0^1 \Big(1+\frac{xu}{\sigma}\Big)e^{-\frac{\theta x u}{\sigma}}u^{\alpha}\,du \\ &= \frac{\alpha\theta^2}{(1+\theta)}\big[r_1(x;\theta,\alpha,1)+xr_2(x;\theta,\alpha,1)\big], \end{aligned}$$

which corresponds to the pdf of the non-scaled version of the LS distribution. Note that the closed form above for the LS pdf is, to our knowledge, not known.

On the other hand, we observe that

$$\lim_{\alpha\to\infty} f_X(x;\theta,\alpha,\beta=1) = \frac{\theta^2}{(1+\theta)}(1+x)e^{-\theta x},$$

which corresponds to the pdf of the L distribution.

Figure 1 shows some curves for the ESL pdf considering different values of θ, α and β. We use the function KUMMERM [25] of the R programming language to compute the confluent hypergeometric function. We provide the R code in Appendix A.

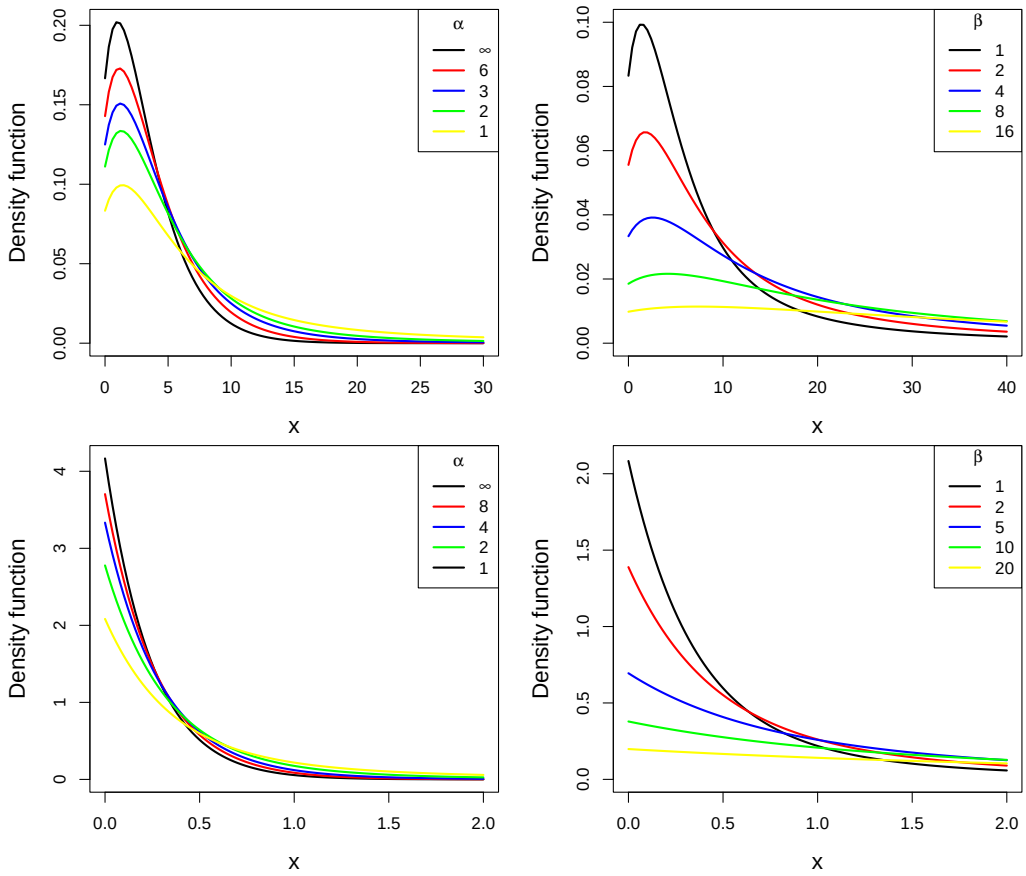

Figure 1. Some pdf curves for the ESL distribution with $\theta = 0.5$ and $\beta = 1$ in the **top left** panel, $\theta = 0.5$ and $\alpha = 1$ in the **top right** panel, $\theta = 5$ and $\beta = 1$ in the **bottom left** panel, and with $\theta = 5$ and $\alpha = 1$ in the **bottom right** panel.

In Figure 1, it can be seen that:

1. The ESL pdf exhibits a unimodal shape when θ is equal to 0.5 (see top panels) and a monotonic decreasing shape when θ is equal to 5 (see bottom panels).
2. If θ and β are fixed (see left panels), it is observed that the unimodal and monotonic shapes of the pdf are preserved with each choice of α. In this case, the different values of α define a set of pdfs with different weights in the tails. The closer to 0 the value of α is, the heavier the tail of the ESL pdf.
3. For fixed values of θ and α (see right panels), something similar is observed. The unimodal and monotonic shapes of the pdf are preserved even when the parameter β varies. Here, different values for β also define a set of pdf's with different weights in the tails. The larger the value of β, the heavier the tail of the ESL pdf.

The above observations suggest that the unimodal and monotone shapes of the ESL pdf depend exclusively on θ, while α and β control the kurtosis of the distribution. More details on the behavior of the kurtosis of the ESL distribution are given in Section 2.3.

2.2. Reliability Analysis

The reliability function (rf) and the hazard rate function (hrf) play a central role in the treatment of lifetime data in reliability studies. If X is a random variable representing the failure time of mechanical units, the rf of X, defined as $R_X(x) := \mathbb{P}(X > x)$, $x > 0$, indicates the probability that mechanical units survive beyond the time x. On the other hand, the hrf of X, defined as $h_X(x) := f_X(x)/R_X(x)$ (where $f_X(\cdot)$ represents the pdf of X), measures the propensity of a mechanical units to fail or die depending on the age it has reached.

In what follows, we concentrate on deriving analytic expressions and some graphical representations for the rf and the hrf of the ESL distribution. For this, we first calculate the cdf of the ESL distribution.

Proposition 2. *Let $X \sim \mathrm{ESL}(\theta, \alpha, \beta)$. Then, the cdf of X is given by*

$$F_X(x;\theta,\alpha,\beta) = F_Y(x;\theta) - \frac{\theta^2}{(1+\theta)B(\alpha,\beta)} E(x;\theta,\alpha,\beta),$$

where $F_Y(\cdot;\cdot)$ is the cdf of the L distribution and

$$E(x;\theta,\alpha,\beta) = \int_0^x (1+v)B\Big(\frac{v}{x};\alpha,\beta\Big)e^{-\theta v}\,dv.$$

such that $B(z;a,b) = \int_0^z u^{a-1}(1-u)^{b-1}\,du$ is the incomplete beta function.

Proof. Considering the change of variable $v = xw$, the pdf of X given in the Equation (3) can be written as $f_X(x;\theta,\alpha,\beta) = \theta^2/[(1+\theta)B(\alpha,\beta)]\int_0^x (1+v)v^\alpha(1-v/x)^{\beta-1}\,dv$. Then, the cdf of X is given by

$$F_X(x;\theta,\alpha,\beta) = \frac{\theta^2}{(1+\theta)B(\alpha,\beta)}\int_0^x \frac{1}{u^{\alpha+1}}\int_0^u (1+v)v^\alpha\Big(1-\frac{v}{u}\Big)^{\beta-1}e^{-\theta v}\,dv\,du.$$

Now, exchanging the order of integration in the previous expression, we obtain that

$$\begin{aligned} F_X(x;\theta,\alpha,\beta) &= \frac{\theta^2}{(1+\theta)B(\alpha,\beta)}\int_0^x \frac{1+v}{v}e^{-\theta v}\int_v^x \Big(\frac{v}{u}\Big)^{\alpha+1}\Big(1-\frac{v}{u}\Big)^{\beta-1}\,du\,dv \\ &= 1 - \frac{(1+\theta+\theta x)e^{-\theta x}}{(1+\theta)} - \frac{\theta^2}{(1+\theta)B(\alpha,\beta)}\int_0^x (1+v)B\Big(\frac{v}{x};\alpha,\beta\Big)e^{-\theta v}\,dv. \end{aligned}$$

Finally, the result is obtained by establishing that $E(x;\theta,\alpha,\beta) = \int_0^x (1+v)B(v/x;\alpha,\beta)$ $e^{-\theta v}\,dv$ and recognizing the cdf of the L distribution in the previous expression. □

Corollary 2. *Let $X \sim \mathrm{ESL}(\theta, \alpha, \beta)$. Then, the rf of X is given by*

$$R_X(x;\theta,\alpha,\beta) = R_Y(x;\theta) + \frac{\theta^2}{(1+\theta)B(\alpha,\beta)}E(x;\theta,\alpha,\beta),$$

and the hrf of X is given by

$$h_X(x;\theta,\alpha,\beta) = \frac{\theta^2\Gamma(\beta)[r_1(x;\theta,\alpha,\beta) + xr_2(x;\theta,\alpha,\beta)]}{(1+\theta)B(\alpha,\beta)R_Y(x;\theta) + \theta^2 E(x;\theta,\alpha,\beta)},$$

where $R_Y(x;\theta) = [(1+\theta+\theta x)/(1+\theta)]e^{-\theta x}$ is the rf of the L distribution.

From the previous result, we see that

$$\begin{aligned} h_X(x;\theta,\alpha,\beta=1) &= \frac{\theta^2[r_1(x;\theta,\alpha,1) + xr_2(x;\theta,\alpha,1)]}{\alpha(1+\theta)R_Y(x;\theta) + \theta^2 E(x;\theta,\alpha,1)} \quad \text{and} \\ \lim_{\alpha\to\infty} h_X(x;\theta,\alpha,\beta=1) &= \frac{\theta^2(1+x)}{1+\theta+\theta x}, \end{aligned}$$

which correspond to the hrfs of the SL and L distributions, respectively.

We provide an R code for the computation of the cdf given in Proposition 2. Using this code, together with the R code for the ESL pdf presented in the previous section, the ESL hrf can be easily computed. Figure 2 presents some ESL hrf curves considering different choices for θ, α and β. From the figure, the following can be noted:

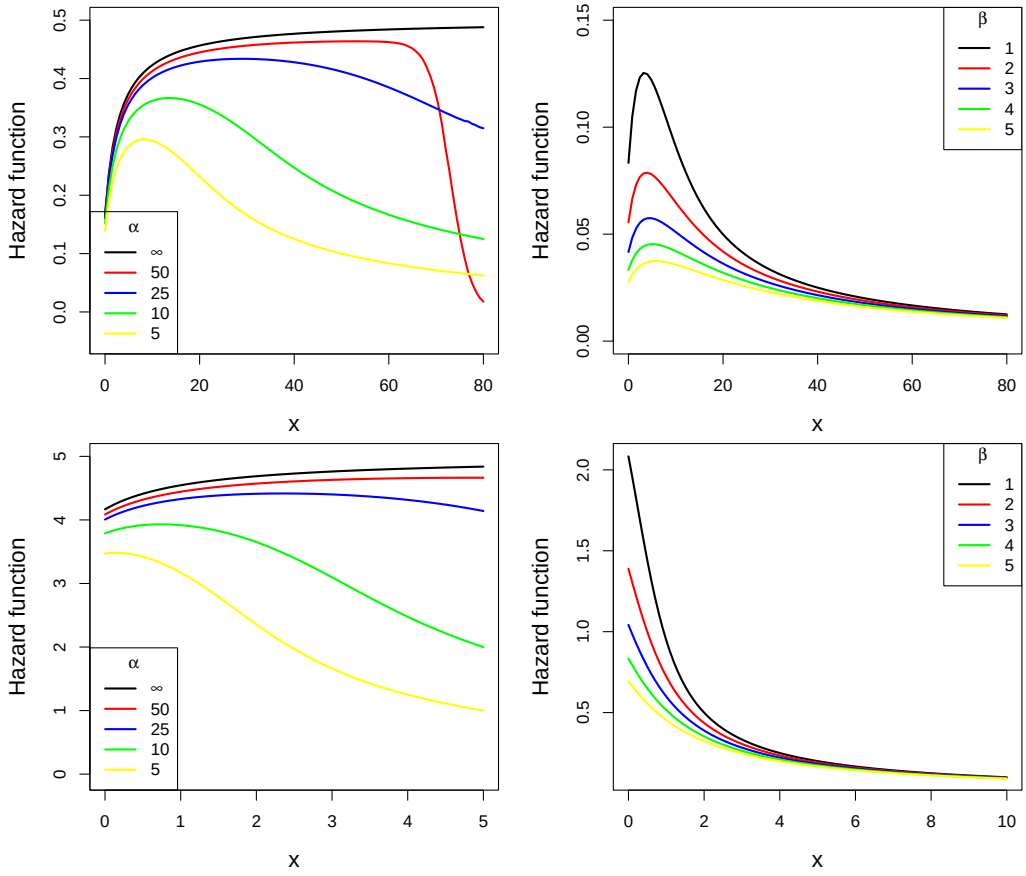

Figure 2. Some hrf curves for the ESL distribution with $\theta = 0.5$ and $\beta = 1$ in the **top left** panel, $\theta = 0.5$ and $\alpha = 1$ in the **top right** panel, $\theta = 5$ and $\beta = 1$ in the **bottom left** panel, and with $\theta = 5$ and $\alpha = 1$ in the **bottom right** panel.

1. The ESL hrf can present unimodal or monotonic decreasing shapes.
2. In the left panels ($\beta = 1$), the unimodal shape of the hrf of the ESL distribution approaches the increasing shape of the hrf of the L distribution when α assumes a sufficiently large value.
3. In the right panels, it is observed that the hrf of the ESL distribution is unimodal when $\theta = 0.5$ and decreasing when $\theta = 5$. It is also observed that the unimodal or increasing forms are preserved for the different choices of the parameters α and β. Thus, the unimodal and monotonic decreasing shape seems to depend on θ.

2.3. Moments and Related Measurements

Proposition 3. *Let $X \sim \mathrm{ESL}(\theta, \alpha, \beta)$. Then, for $r = 1, 2, \ldots$ and $\alpha > r$, it follows that rth raw moment is given by*

$$\mu_r = E(X^r) = \frac{r!(\theta + r + 1)}{\theta^r(\theta + 1)} \prod_{i=1}^{r} \frac{\alpha + \beta - i}{\alpha - i}.$$

Proof. Using the representation given in Definition 1, we have that

$$\mu_r = E(X^r) = E\left(\left(\frac{Y}{U}\right)^r\right) = E(Y^r)E\left(U^{-r}\right).$$

where it follows that $E(U^{-r}) = \prod_{i=1}^{n} \frac{\alpha+\beta-i}{\alpha-i}$, $\alpha > r$ and $E(Y^r) = \frac{r!(\theta+r+1)}{\theta^r(\theta+1)}$ is the rth raw moment of the L distribution. □

Corollary 3. *Let $X \sim \mathrm{ESL}(\theta, \alpha, \beta)$. Then, it follows that*

$$E(X) = \frac{(\theta + 2)(\alpha + \beta - 1)}{\theta(\theta + 1)(\alpha - 1)}, \ \alpha > 1 \ \textit{and}$$

$$Var(X) = \frac{\alpha + \beta - 1}{\theta(\theta + 1)(\alpha - 1)} \left(\frac{2(\theta + 3)(\alpha + \beta - 2)}{\theta(\alpha - 2)} - \theta - 2 \right), \quad \alpha > 2.$$

Corollary 4. *Let $X \sim \mathrm{ESL}(\theta, \alpha, \beta)$. Then, Fisher's skewness ($\sqrt{\beta_1}$) and kurtosis (β_2) coefficients are given by*

$$\sqrt{\beta_1} = \frac{2(\alpha - 2)^{1/2} A}{(\alpha + \beta - 1)^{1/2}(\alpha - 3)B^{3/2}}, \ \alpha > 3, \ \textit{and} \ \beta_2 = \frac{3(\alpha - 2)C}{(\alpha - 3)(\alpha - 4)(\alpha + \beta - 1)B^2}, \ \alpha > 4,$$

where

$$\begin{aligned} A &= 3(\theta + 1)^2(\theta + 4)(\alpha - 1)^2(\alpha + \beta - 2)(\alpha + \beta - 3) \\ &\quad -3(\theta + 1)(\theta + 2)(\theta + 3)(\alpha - 1)(\alpha - 3)(\alpha + \beta - 1)(\alpha + \beta - 2) \\ &\quad +(\theta + 2)^3(\alpha - 2)(\alpha - 3)(\alpha + \beta - 1)^2, \end{aligned}$$

$$B = 2(\theta + 1)(\theta + 3)(\alpha - 1)(\alpha + \beta - 2) - (\theta + 2)^2(\alpha - 2)(\alpha + \beta - 1) \quad \textit{and}$$

$$\begin{aligned} C &= 8(\theta + 1)^3(\theta + 5)(\alpha - 1)^3(\alpha + \beta - 2)(\alpha + \beta - 3)(\alpha + \beta - 4) \\ &\quad -8(\theta + 1)^2(\theta + 2)(\theta + 4)(\alpha - 1)^2(\alpha - 4)(\alpha + \beta - 1)(\alpha + \beta - 2)(\alpha + \beta - 3) \\ &\quad +4(\theta + 1)(\theta + 2)^2(\theta + 3)(\alpha - 1)(\alpha - 3)(\alpha - 4)(\alpha + \beta - 1)^2(\alpha + \beta - 2) \\ &\quad -(\theta + 2)^4(\alpha - 2)(\alpha - 3)(\alpha - 4)(\alpha + \beta - 1)^3. \end{aligned}$$

Figure 3 depicts plots of the Fisher's skewness and kurtosis coefficients of the ESL distribution. These plots show that the coefficients decrease as θ and α decrease and as β increases. Although the coefficients depend on the three parameters, it can be seen that θ has a slight impact on the skewness and kurtosis, while α and β have a great impact, especially on kurtosis.

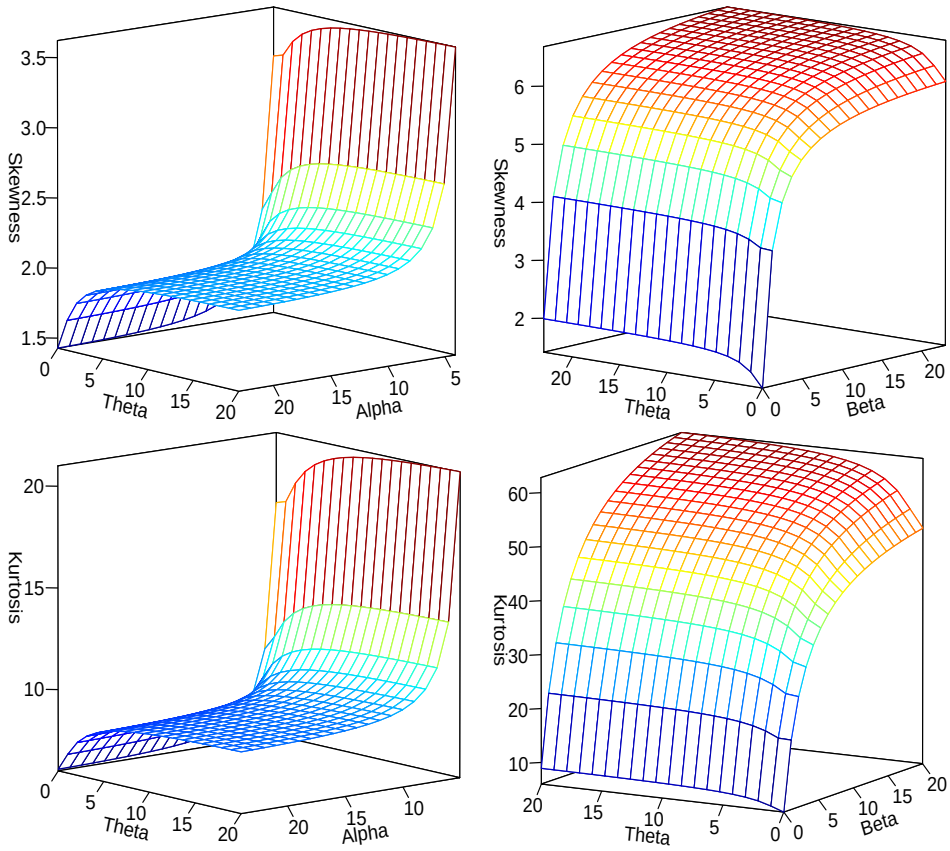

Figure 3. Plot of the skewness and kurtosis coefficients for the ESL (θ, α, β) distribution with $\beta = 1$ in the **left** panels and $\alpha = 1$ in the **right** panels.

3. Parameter Estimation

This section discusses parameter estimation for the ESL distribution considering the moment and maximum likelihood methods. In addition, we carry out a simulation study that illustrates the behavior of the estimators under different sample sizes.

3.1. Moment Estimation

Proposition 4. *Let $X_1, \ldots, X_n$ be a random sample of the random variable $X \sim \text{ESL}(\theta, \alpha, \beta)$. Then, for $\alpha > 3$, the moment estimators $\hat{\theta}_{\text{M}}$, $\hat{\alpha}_{\text{M}}$ and $\hat{\beta}_{\text{M}}$ for θ, α and β are given by*

$$\left[\frac{\overline{X^2}\hat{\theta}_{\text{M}}(\hat{\theta}_{\text{M}}+2)}{2\overline{X}(\hat{\theta}_{\text{M}}+3)} - 1\right][\alpha(\hat{\theta}_{\text{M}}) - 2] - \beta(\hat{\theta}_{\text{M}}) = 0, \tag{6}$$

$$\hat{\alpha}_{\text{M}} = \alpha(\hat{\theta}_{\text{M}}) \quad \textit{and} \quad \hat{\beta}_{\text{M}} = \beta(\hat{\theta}_{\text{M}}),$$

where

$$\alpha(\hat{\theta}_{\text{M}}) = \frac{6[\overline{X}(\hat{\theta}_{\text{M}}^2 + 7\hat{\theta}_{\text{M}} + 12) + \overline{X^2}\hat{\theta}_{\text{M}}(\hat{\theta}_{\text{M}}^2 + 6\hat{\theta}_{\text{M}} + 8) - \overline{X^3}\hat{\theta}_{\text{M}}(\hat{\theta}_{\text{M}}^2 + 6\hat{\theta}_{\text{M}} + 9)]}{\hat{\theta}_{\text{M}}[3\overline{X^2}(\hat{\theta}_{\text{M}}^2 + 6\hat{\theta}_{\text{M}} + 8) - 2\overline{X^3}(\hat{\theta}_{\text{M}}^2 + 6\hat{\theta}_{\text{M}} + 9)]}, \tag{7}$$

$$\beta(\hat{\theta}_{\text{M}}) = \left[\frac{\overline{X^3}\hat{\theta}_{\text{M}}(\hat{\theta}_{\text{M}}+3)}{3\overline{X^2}(\hat{\theta}_{\text{M}}+4)} - 1\right][\alpha(\hat{\theta}_{\text{M}}) - 3]. \tag{8}$$

Proof. The first three equations in the method of moments are given by $\mu_j = \overline{X^j}$, with $j = 1, 2, 3$, where μ_j is as in Proposition 3. Thus, combining these equations, we obtain that

$$\begin{aligned}\beta(\hat{\theta}_{\mathrm{M}}) &= \left[\frac{\overline{X^2}\hat{\theta}_{\mathrm{M}}(\hat{\theta}_{\mathrm{M}}+2)}{2\overline{X}(\hat{\theta}_{\mathrm{M}}+3)} - 1\right]\left[\alpha(\hat{\theta}_{\mathrm{M}}) - 2\right],\\ \beta(\hat{\theta}_{\mathrm{M}}) &= \left[\frac{\overline{X^3}\hat{\theta}_{\mathrm{M}}(\hat{\theta}_{\mathrm{M}}+3)}{3\overline{X^2}(\hat{\theta}_{\mathrm{M}}+4)} - 1\right]\left[\alpha(\hat{\theta}_{\mathrm{M}}) - 3\right],\end{aligned}$$

where the last of these equations corresponds to the result given in Equation (8) and the algebraic manipulation of the first gives rise to Equation (6). Finally, Equation (7) is obtained by equating the previous equations under a suitable algebra. □

3.2. Maximum Likelihood Estimation

Suppose $X_1, \ldots, X_n$ is a random sample of the random variable $X \sim \mathrm{ESL}(\theta, \alpha, \beta)$. Then, from Equation (3), we have that the log-likelihood function can be written as

$$\begin{aligned}\ell(\theta, \alpha, \beta; x_i) &= \log\prod_{i=1}^{n} f_X(x_i; \theta, \alpha, \beta)\\ &= 2n\log(\theta) - n\log(\theta+1) - \log[B(\alpha, \beta)] + \sum_{i=1}^{n}\log[K(x_i)], \qquad (9)\end{aligned}$$

where

$$K(x_i) = \int_0^1 (1 + x_i w) w^{\alpha}(1-w)^{\beta-1} e^{-\theta x_i w}\, dw.$$

The maximum likelihood (ML) estimators $\hat{\theta}_{\mathrm{ML}}$, $\hat{\alpha}_{\mathrm{ML}}$ and $\hat{\beta}_{\mathrm{ML}}$ for θ, α and β can be obtained by taking the partial derivatives of Equation (9) and solving the corresponding system of equations, which is given by the equations

$$\frac{2n}{\theta} + \frac{n}{\theta+1} + \sum_{i=1}^{n}\frac{K_1(x_i)}{K(x_i)} = 0, \qquad (10)$$

$$n\Psi(\alpha) - n\Psi(\alpha+\beta) + \sum_{i=1}^{n}\frac{K_2(x_i)}{K(x_i)} = 0, \qquad (11)$$

$$n\Psi(\beta) - n\Psi(\alpha+\beta) + \sum_{i=1}^{n}\frac{K_3(x_i)}{K(x_i)} = 0, \qquad (12)$$

where

$$\begin{aligned}K_1(x_i) &= -x_i\int_0^1 (1+x_i w) w^{\alpha+1}(1-w)^{\beta-1} e^{-\theta x_i w}\, dw,\\ K_2(x_i) &= \int_0^1 (1+x_i w)\log(w) w^{\alpha}(1-w)^{\beta-1} e^{-\theta x_i w}\, dw,\\ K_3(x_i) &= \int_0^1 (1+x_i w)\log(1-w) w^{\alpha}(1-w)^{\beta-1} e^{-\theta x_i w}\, dw,\end{aligned}$$

such that $\Psi(\cdot)$ is the digamma function.

The asymptotic distribution (under regularity conditions) of the ML estimator of $\delta = (\theta, \alpha, \beta)'$ is $N_3(\delta, I(\delta)^{-1})$, where $I(\delta)^{-1}$ is the expected information matrix. Taking into account the structure of Equation (9), we observe that it is not easy to derive the analytical expression of this matrix. Thus, we consider an approximation from the observed information matrix, where the elements of this matrix are computed as minus the second

partial derivatives of the log-likelihood function with respect to each parameter (assessed in the ML estimates). The observed information matrix is given by

$$I(\delta) = \begin{pmatrix} I_{\theta\theta} & I_{\alpha\theta} & I_{\beta\theta} \\ & I_{\alpha\alpha} & I_{\beta\alpha} \\ & & I_{\beta\beta} \end{pmatrix},$$

where

$$\begin{aligned}
I_{\theta\theta} &= -\frac{\partial^2 \ell(\theta,\alpha,\beta;x_i)}{\partial\theta^2} = -\sum_{i=1}^{n}\frac{K_{11}(x_i)}{K(x_i)} + \sum_{i=1}^{n}\frac{K_1^2(x_i)}{K^2(x_i)} + \frac{2n}{\theta^2} - \frac{n}{(\theta+1)^2},\\
I_{\alpha\alpha} &= -\frac{\partial^2 \ell(\theta,\alpha,\beta;x_i)}{\partial\alpha^2} = -\sum_{i=1}^{n}\frac{K_{22}(x_i)}{K(x_i)} + \sum_{i=1}^{n}\frac{K_2^2(x_i)}{K^2(x_i)} + n\Psi_1(\alpha) - n\Psi_1(\alpha+\beta),\\
I_{\beta\beta} &= -\frac{\partial^2 \ell(\theta,\alpha,\beta;x_i)}{\partial\beta^2} = -\sum_{i=1}^{n}\frac{K_{33}(x_i)}{K(x_i)} + \sum_{i=1}^{n}\frac{K_3^2(x_i)}{K^2(x_i)} + n\Psi_1(\beta) - n\Psi_1(\alpha+\beta)\},\\
I_{\alpha\theta} &= -\frac{\partial^2 \ell(\theta,\alpha,\beta;x_i)}{\partial\alpha\partial\theta} = -\sum_{i=1}^{n}\frac{K_{12}(x_i)}{K(x_i)} + \sum_{i=1}^{n}\frac{K_1(x_i)K_2(x_i)}{K^2(x_i)},\\
I_{\beta\theta} &= -\frac{\partial^2 \ell(\theta,\alpha,\beta;x_i)}{\partial\beta\partial\theta} = -\sum_{i=1}^{n}\frac{K_{13}(x_i)}{K(x_i)} + \sum_{i=1}^{n}\frac{K_1(x_i)K_3(x_i)}{K^2(x_i)},\\
I_{\beta\alpha} &= -\frac{\partial^2 \ell(\theta,\alpha,\beta;x_i)}{\partial\beta\partial\alpha} = -\sum_{i=1}^{n}\frac{K_{23}(x_i)}{K(x_i)} + \sum_{i=1}^{n}\frac{K_2(x_i)K_3(x_i)}{K^2(x_i)} + \Psi_1(\alpha+\beta),
\end{aligned}$$

such that

$$\begin{aligned}
K_{11}(x_i) &= x_i^2\int_0^1 (1+x_i w)w^{\alpha+2}(1-w)^{\beta-1}e^{-\theta x_i w}\,dw,\\
K_{22}(x_i) &= \int_0^1 (1+x_i w)\log^2(w)w^{\alpha}(1-w)^{\beta-1}\,e^{-\theta x_i w}\,dw,\\
K_{33}(x_i) &= \int_0^1 (1+x_i w)\log^2(1-w)w^{\alpha}(1-w)^{\beta-1}\,e^{-\theta x_i w}\,dw,\\
K_{12}(x_i) &= -x_i\int_0^1 (1+x_i w)\log(w)w^{\alpha+1}(1-w)^{\beta-1}\,e^{-\theta x_i w}\,dw,\\
K_{13}(x_i) &= -x_i\int_0^1 (1+x_i w)\log(1-w)w^{\alpha+1}(1-w)^{\beta-1}\,e^{-\theta x_i w}\,dw,\\
K_{23}(x_i) &= \int_0^1 (1+x_i w)\log(w)\log(1-w)w^{\alpha}(1-w)^{\beta-1}\,e^{-\theta x_i w}\,dw.
\end{aligned}$$

and $\Psi_1(\cdot)$ is the trigamma function.

Then, approximate $100(1-\varphi)\%$ confidence intervals for θ, α and β can be determined by $\hat{\theta} \pm z_{\varphi/2}s_{\hat{\theta}}$, $\hat{\alpha} \pm z_{\varphi/2}s_{\hat{\alpha}}$ and $\hat{\beta} \pm z_{\varphi/2}s_{\hat{\beta}}$, respectively, where $z_{\varphi/2}$ is the upper $(\varphi/2)$th percentile of the standard normal distribution, and $s_{\hat{\theta}}$, $s_{\hat{\alpha}}$ and $s_{\hat{\beta}}$ are the diagonal elements of the matrix $[I(\delta)]^{-1}$ (assessed in the ML estimates).

3.3. Practical Considerations

From Equation (6), we observe that the moment estimator $\hat{\theta}_{\mathrm{M}}$ of θ does not have an explicit analytical form, so it is necessary to solve Equation (6) numerically to obtain the estimate. We use the function UNIROOT of the R programming language to solve this problem.

From the system of Equations (10)–(12), we see that the ML estimators $\hat{\theta}_{\mathrm{ML}}$, $\hat{\alpha}_{\mathrm{ML}}$ and $\hat{\beta}_{\mathrm{ML}}$ for θ, α and β do not have closed forms. Therefore, the ML estimates must be obtained by numerically solving the system of Equations (10)–(12), which can be quite a complicated problem. In this case, we prefer to obtain the ML estimates by solving the

optimization problem $\max_{(\theta,\alpha,\beta)} \ell(\theta,\alpha,\beta;x_i)$, subject to $\theta,\alpha,\beta>0$, where $\ell(\theta,\alpha,\beta;x_i)$ is as in Equation (9). We solve this problem using the function OPTIM of the R programming language under consideration of the L-BFGS-B algorithm [26]. This algorithm requires the declaration of initial values in the parameter space to initialize the iterative process. Through simulation experiments, we observe that the estimates obtained are strongly influenced by the initial value considered. The first alternative that we considered to overcome this drawback was to use as initial values the moment estimates obtained from the Proposition 4. However, these estimates present an important bias, so they cannot be considered as good initial values. Finally, we note that this drawback occurs due to the existence of two shape parameters that control the kurtosis of the distribution, α and β, which leads to a problem of non-identifiability of the parameters. This problem disappears if there is only one shape parameter that controls the kurtosis and, for this, different solutions can be adopted. In particular, based on the kurtosis behavior described in Section 2.3, we assume $\beta = 1 + 100/\alpha$, obtaining a more parsimonious distribution with extremely heavy tails. In this choice for β, the following can be verified:

1. The L distribution can be obtained as a limit case of the two-parameter ESL distribution once $\beta \to 1$ as $\alpha \to \infty$.
2. Taking into account the kurtosis of the ESL distribution increases as β grows, the two-parameter ESL distribution has very heavy tails since $\beta \to \infty$ as $\alpha \downarrow 0$.
3. In the choice $\beta = 1 + 100/\alpha$, the number 100 allows us to obtain a high value for β when α assumes a value considered small. For example, we get $\beta = 1 + 100/2 = 51$ when $\alpha = 2$. Thus, the representation of the ESL random variable given in Proposition (1) assumes a Beta$(\alpha = 2, \beta = 51)$ distribution for the variable U in the denominator, thus defining a distribution with extremely heavy tails.

For practical purposes, in accordance with the above, we recommend the reader to use the two-parameter version of the ESL distribution.

3.4. Simulation Study

In this section, we carry out a simulation study to evaluate the behavior of the moment and ML estimators assuming that $\beta = 1 + 100/\alpha$. The study is developed considering the following stages:

1. We choose the values 0.5 and 2 for the parameter θ and the values 3, 4 and 5 for the parameter α. In this way, we establish three scenarios where the pdf of the ESL distribution is unimodal (when $\theta = 0.5$) and three other scenarios where the pdf is decreasing (when $\theta = 2$).
2. Under the consideration of the choices above, we generate 1,000 pseudo-random samples of size $n = 100, 200, 300, 400$ and 500 from the two-parameter ESL distribution. Samples are generated as follows:
   (a) Generate Y having a L(θ) distribution.
      i. Generate $S \sim Exponential(\theta)$.
      ii. Generate $W \sim Gamma(2,\theta)$.
      iii. Generate $V \sim Bernoulli\left(\frac{\theta}{1+\theta}\right)$.
      iv. Compute $Y = VS + (1-V)W$.
   (b) Generate X having a ESL(θ,α) distribution.
      i. Generate $U \sim Beta(\alpha, 1 + 100/\alpha)$.
      ii. Compute $X = YU^{-1}$.

   Step (a) is based on the results proposed by Ghitany et al. [3] for the L distribution, while step (b) is based on the representation of the ESL random variable given in Definition 1. An R code is provided in Appendix A.
3. For each sample generated, we obtain the ML estimates by maximizing Equation (9) using the OPTIM function of the R programming language.

Table 1 reports the average estimate (AE), the average of the estimated bias (AB), and the standard deviation (SD) of the estimates obtained in each scenario and each sample size considered. Note that the AEs approach the true values of the parameters and the SDs become small as the sample size increases, an expected result since the ML and moment estimators are consistent.

Table 1. The average estimate (AE), the average of the estimated bias (AB) and the standard deviation (SD) of the estimates obtained in each scenario and sample size considered in the study.

Parameter		$\hat{\theta}_{\mathrm{M}}$			$\hat{\alpha}_{\mathrm{M}}$			$\hat{\theta}_{\mathrm{ML}}$			$\hat{\alpha}_{\mathrm{ML}}$		
θ	α	**AE**	**SD**	**AB**	**AE**	**SD**	**AB**	**AE**	**SD**	**AB**	**AE**	**SD**	**AB**
						$n = 100$							
0.5	3	0.351	0.147	−0.149	3.965	1.103	0.965	0.495	0.189	−0.005	3.250	0.771	0.250
	4	0.413	0.190	−0.087	5.035	1.602	1.035	0.506	0.201	0.006	4.354	1.206	0.354
	5	0.451	0.216	−0.049	6.197	2.301	1.197	0.515	0.220	0.015	5.490	1.676	0.490
2	3	1.370	0.556	−0.630	4.041	1.123	1.041	2.110	0.897	0.110	3.299	0.895	0.299
	4	1.662	0.835	−0.338	5.213	1.897	1.213	2.210	1.236	0.210	4.422	1.564	0.422
	5	1.829	0.976	−0.171	6.420	2.691	1.420	2.229	1.321	0.229	5.686	2.496	0.686
						$n = 200$							
0.5	3	0.381	0.128	−0.119	3.637	0.681	0.637	0.499	0.134	−0.001	3.113	0.482	0.113
	4	0.438	0.167	−0.062	4.623	1.025	0.623	0.501	0.138	0.001	4.160	0.736	0.160
	5	0.465	0.172	−0.035	5.687	1.409	0.687	0.505	0.149	0.005	5.261	1.100	0.261
2	3	1.508	0.526	−0.492	3.699	0.756	0.699	2.043	0.688	0.043	3.131	0.595	0.131
	4	1.778	0.744	−0.222	4.684	1.160	0.684	2.070	0.755	0.070	4.207	0.940	0.207
	5	1.885	0.837	−0.115	5.840	1.771	0.840	2.084	0.795	0.084	5.382	1.550	0.382
						$n = 300$							
0.5	3	0.396	0.121	−0.104	3.521	0.576	0.521	0.499	0.108	−0.001	3.079	0.383	0.079
	4	0.460	0.166	−0.040	4.432	0.821	0.432	0.501	0.116	0.001	4.077	0.581	0.077
	5	0.478	0.157	−0.022	5.467	1.105	0.467	0.504	0.115	0.004	5.160	0.814	0.160
2	3	1.603	0.512	−0.397	3.529	0.632	0.329	2.024	0.549	0.024	3.089	0.479	0.089
	4	1.836	0.678	−0.164	4.488	0.919	0.488	2.063	0.625	0.063	4.098	0.743	0.098
	5	1.910	0.749	−0.090	5.581	1.270	0.581	2.075	0.667	0.075	5.213	1.118	0.213
						$n = 400$							
0.5	3	0.410	0.114	−0.090	3.424	0.508	0.424	0.500	0.095	0.000	3.045	0.329	0.045
	4	0.460	0.144	−0.040	4.378	0.730	0.378	0.501	0.095	0.001	4.078	0.481	0.078
	5	0.483	0.153	−0.017	5.359	0.946	0.359	0.502	0.104	0.002	5.112	0.699	0.112
2	3	1.653	0.486	−0.347	3.451	0.563	0.451	2.017	0.478	0.017	3.053	0.412	0.053
	4	1.858	0.652	−0.142	4.406	0.794	0.406	2.030	0.516	0.030	4.083	0.625	0.083
	5	1.911	0.643	−0.089	5.470	1.042	0.470	2.033	0.524	0.033	5.170	0.875	0.170
						$n = 500$							
0.5	3	0.418	0.112	−0.082	3.384	0.477	0.384	0.500	0.088	0.000	3.024	0.293	0.024
	4	0.468	0.142	−0.032	4.334	0.716	0.334	0.500	0.091	0.000	4.012	0.473	0.012
	5	0.492	0.138	−0.008	5.251	0.849	0.251	0.501	0.090	0.001	5.050	0.605	0.050
2	3	1.663	0.464	−0.337	3.434	0.520	0.434	2.015	0.426	0.015	3.051	0.364	0.051
	4	1.869	0.601	−0.131	4.361	0.742	0.361	2.028	0.472	0.028	4.080	0.572	0.080
	5	1.962	0.637	−0.038	5.349	0.964	0.349	2.031	0.463	0.031	5.122	0.776	0.122

Comparing the results obtained with each method, we can see that the ML method provides estimates with less bias. It is also observed that the ML estimates for α are more efficient in all the sample sizes considered. Finally, we observe that the SDs of the moment estimates of θ are smaller than the corresponding SDs obtained with the ML method when the sample size is less than or equal to 200. However, in this case, the AEs provided by the moment method are much more distant from the real value of θ than the AEs provided by the ML method. For sample sizes greater than 200, the ML estimates for θ are less biased and more efficient.

4. Real Data Analysis

In this section, by fitting data featuring outliers, we compare the performance of the two-parameter ESL distribution with that of the L, power-Lindley (PL) [6], and LS distributions. The pdfs of the L and LS distributions can be consulted in Section 1, while the pdf of the PL distribution is given by $f(x;\theta,\alpha) = [(\alpha\theta^2)/(\theta+1)](1+x^\alpha)x^{\alpha-1}e^{-\theta x^\alpha}$, with $x > 0$ and $\theta, \alpha > 0$. Note that the PL distribution is one of the most popular extensions to the L distribution, also note that this distribution has the same parameter dimension as the two-parameter ESL distribution.

We evaluate the comparative performance of the fitted distributions using the Akaike Information Criterion (AIC) [27] and the Bayesian Information Criterion (BIC) [28]. On the other hand, we consider the goodness-of-fit tests proposed by Chen and Balakrishnan [29] to assess the quality of fit of the distributions. Specifically, based on the modified Cramer-von Mises (W^*) and Anderson–Darling (A^*) statistics, we test the hypothesis $H_0 : x_1, \ldots, x_n$ is an observed random sample from a continuous distribution $F(x;\delta)$, where $F(\cdot;\cdot)$ is known, but δ must be estimated efficiently. Here, H_0 is rejected at significance level equal to 0.05 if $W^* > 0.126$ and $A^* > 0.752$. Under a significance level equal to 0.10, H_0 is rejected when W^* and A^* are greater than 0.104 and 0.631, respectively.

For the analyzed data, we evidenced the presence of outliers by elaborating the traditional box-plot and the adjusted box-plot [30] for skewed data. Note that the second plot includes a robust measure of skewness in the determination of the whiskers, which helps to avoid misreporting of outliers. We use the adjbox function [31] available in the R programming language for the elaboration of this plot.

4.1. Time Failures Related to Pascal Programming

The data considered in this application were collected from a single-user workstation at the Center for Software Reliability and represents the time failures related to Pascal programming. Some descriptive statistics for these data are as follows: sample size, 104; average, 147.8; standard deviation, 245.095; Fisher's skewness coefficient, 3; Fisher's kurtosis coefficient, 14.579. Figure 4 presents the traditional box-plot and the adjusted box-plot for the time failure data in which the existence of outliers can be observed.

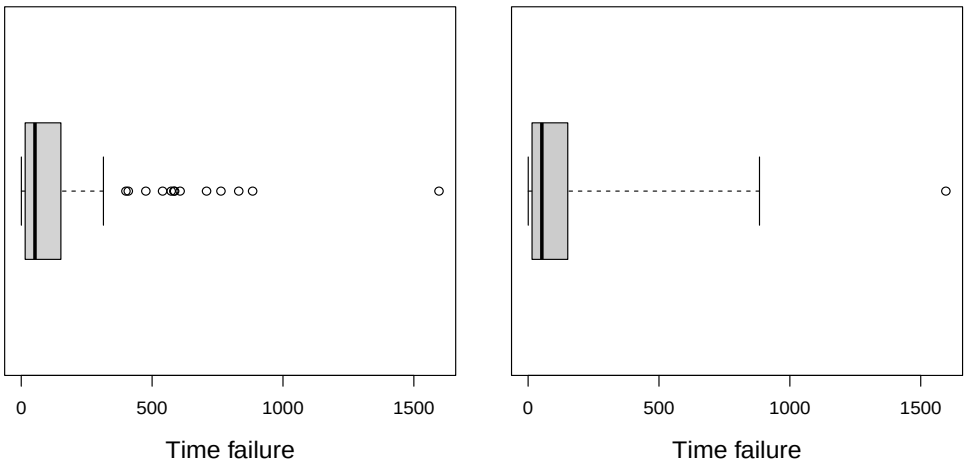

Figure 4. Box-plots for time failure data: Traditional box-plot in the **left** panel and adjusted box-plot in the **right** panel.

Previous studies with these data can be found in Lyu [32] and Astorga et al. [33].

Computing initially the moment estimators under the ESL distribution, we have the following estimates: $\hat{\theta}_{\mathrm{M}} = 0.274$ and $\hat{\alpha}_{\mathrm{M}} = 2.730$. Using the moment estimates as initial values, the ML estimates are computed and presented in Table 2 with the standard errors in parentheses. In addition, we also report the maximum value (ℓ) of the corresponding log-likelihood function for each fitted distribution. Note that the largest ℓ value is reported for the ESL distribution.

Table 2 also reports for each fitted distribution the values associated with the information criteria and with the modified statistics. In the table, based on the A^* statistic, considering a level of significance equal to 0.05, it is observed that only the ESL distribution adequately fits the time failure data. An analogous observation can be considered from the values of both statistics under a level of significance equal to 0.10. On the other hand, also in Table 2, it is observed that the ESL distribution is the one with the lowest AIC and BIC values, suggesting that this distribution should be selected to model the time failure data.

Table 2. The ML estimates (with standard errors in parentheses), the maximum values of the corresponding log-likelihood functions, the values associated with the information criteria (AIC and BIC) and with the modified statistics (W^* and A^*) for each fitted distribution to the time failure data.

Parameter	L	PL	LS	ESL
σ	-	-	2.735×10^4	-
			(1.046×10^4)	
θ	0.013	0.215	1.169×10^3	1.819
	(0.001)	(0.034)	(5.483×10^2)	(0.004)
α	-	0.480	0.799	1.284
		(0.030)	0.148	(0.003)
ℓ	−703.450	−600.236	−602.341	−599.780
AIC	1408.900	1204.472	1210.683	1203.561
BIC	1411.545	1209.761	1218.616	1208.850
W^*	0.353	0.160	0.119	0.084
A^*	2.087	0.978	0.808	0.566

Figure 5 depicts plots of the fitted L, PL, LS and ESL distributions using the ML estimates. Notice that the fitted ESL distribution has heavier tails.

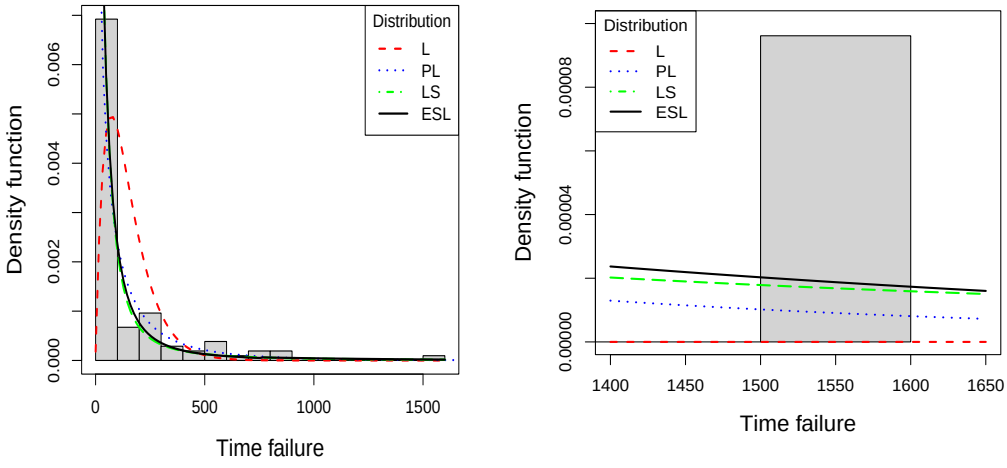

Figure 5. Histogram for the time failure data and curve of the pdf of each fitted distribution.

4.2. State Personal Income Data

We consider a set of annual observations on state personal income (total, nominal, in millions of dollars) recorded in the 48 continental United States. Stock and Watson [34] analyze these data in a study on the demand for tobacco. Some descriptive statistics for this data are as follows: sample size, 104; average, 99.879; standard deviation, 120.541; Fisher's skewness coefficient, 2.817; Fisher's kurtosis coefficient, 13.275. We provide this dataset in Appendix B. Figure 6 presents the traditional box-plot and the adjusted box-plot for the state personal income data in which the existence of outliers can be observed.

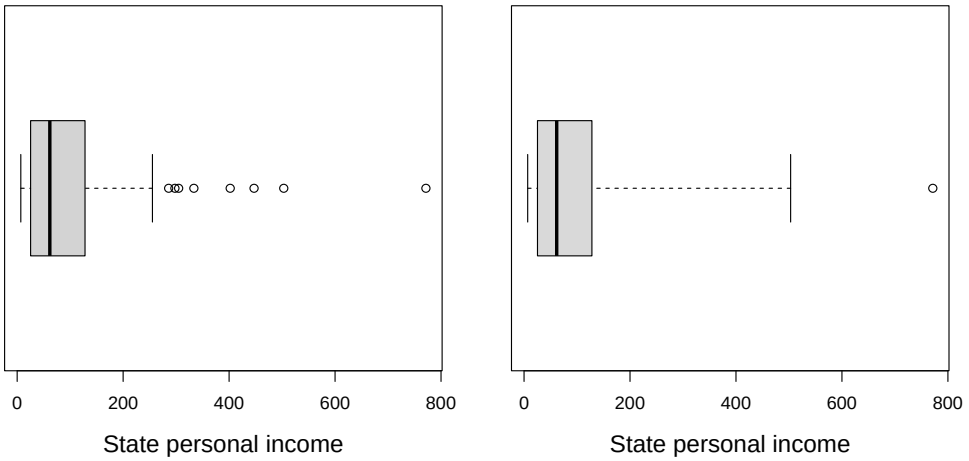

Figure 6. Box-plots for state personal income data: traditional box-plot in the **left** panel and adjusted box-plot in the **right** panel.

With these data, we obtain the following moment estimates: $\hat{\theta}_M = 0.207$ and $\hat{\alpha}_M = 3.702$. Table 3 reports the ML estimates for the parameters of each distribution fitted to the state personal income data. In addition, the maximum values of the corresponding log-likelihood functions and the values associated with the information criteria and the modified statistics are reported. In this table, based on the statistics modified under a level of significance equal to 0.05, it can be seen that the L distribution is the only distribution that does not properly fit the state personal income data. On the other hand, the ESL distribution is the one with the lowest AIC and BIC values, which suggests that this distribution should be selected to fit the state personal income data.

Table 3. The ML estimates (with standard errors in parentheses), the maximum values of the corresponding log-likelihood functions, the values associated with the information criteria (AIC and BIC) and with the modified statistics (W^* and A^*) for each fitted distribution to the state personal income data.

Parameter	L	PL	LS	ESL
σ	-	-	4.260×10^4	-
			(7.528×10^3)	
θ	0.019	0.083	682.380	0.312
	(0.001)	(0.019)	192.712	(0.089)
α	-	0.705	2.614	3.033
		(0.046)	1.147	(0.484)
ℓ	-553.170	-536.240	-536.596	-534.902
AIC	1108.341	1076.480	1079.193	1073.804
BIC	1110.906	1081.609	1086.886	1078.933
W^*	0.140	0.084	0.074	0.051
A^*	1.003	0.653	0.579	0.432

Figure 7 depicts plots of the fitted L, PL, LS and ESL distributions using the ML estimates. Notice that the fitted ESL distribution has heavier tails.

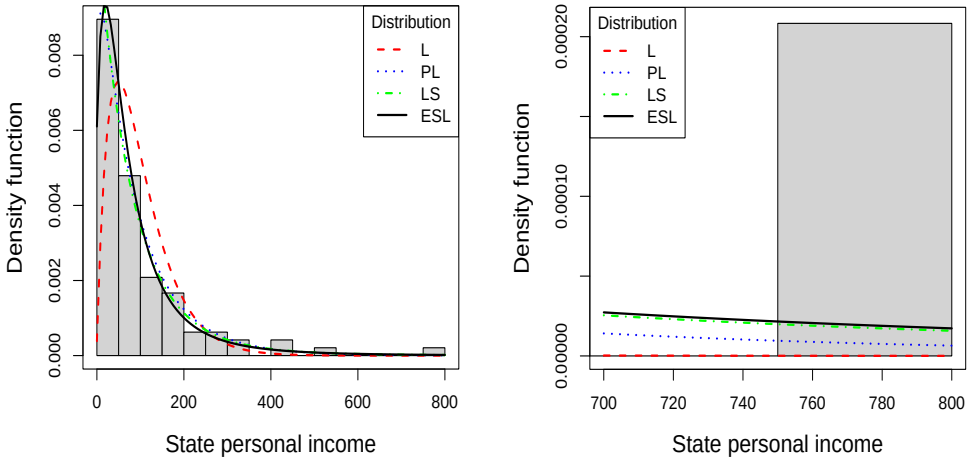

Figure 7. Histogram for the state personal income data and curve of the pdf of each fitted distribution.

5. Concluding Remarks

This article proposes a heavy-tailed extension of the Lindley distribution, the extended slash Lindley (ESL) distribution. The proposal arises as the ratio of independent random variables, specifically, a Lindley distribution divided by a beta distribution. In this way, a three-parameter distribution with heavier tails than the classical Lindley distribution is obtained.

The main structural properties such as the pdf, cdf, rf and hrf are derived. The pdf is written in closed form in terms of the confluent hypergeometric function, which facilitates its computational implementation. Both the pdf and the hrf can exhibit monotonically decreasing or unimodal shapes. In addition, closed expressions are derived for the raw moments of the ESL distribution, which are then used to describe the behavior of the Fisher's skewness and kurtosis coefficients. We note that two of the parameters of the ESL distribution (α and β) control the kurtosis of the distribution, while the remaining parameter (θ) controls the unimodal or monotonic shape of the pdf.

Although the presence of two shape parameters controlling the kurtosis allows the ESL distribution to be able to capture high levels of kurtosis, we observe that this generates a problem of non-identifiability of the parameters, which was evidenced by simulation experiments. This problem disappears when there is a single parameter that controls the kurtosis, so several solutions to this problem can be adopted. In particular, we propose the special case obtained by considering $\beta = 1 + 100/\alpha$, which results in a two-parameter version capable of capturing high levels of kurtosis. We verify that the pdf and the hrf of this special case, as in the three-parameter version, can have unimodal and monotonic decreasing shapes. We also note that the classical Lindley distribution can be derived as a limit case of this special case.

We discuss parameter estimation considering moment and ML methods. It is noted that both methods require the use of numerical procedures. For the two-parameter ESL distribution, through simulation experiments, we observe that the ML method provides estimates with less bias than the estimates obtained with the moment method.

Finally, when fitting data that present a high level of kurtosis, we observe that the two-parameter ESL distribution can present a better performance than the classical Lindley distribution and some of its known extensions.

Author Contributions: Conceptualization, M.A.R. and Y.A.I.; Formal analysis, M.A.R. and Y.A.I.; Investigation, M.A.R. and Y.A.I.; Methodology, M.A.R. and Y.A.I.; Supervision, M.A.R. and Y.A.I.; Validation, M.A.R. and Y.A.I.; Writing—original draft, M.A.R. and Y.A.I.; Writing—review & editing, M.A.R. and Y.A.I. All authors have read and agreed to the published version of the manuscript.

Funding: The investigation of Mario A. Rojas and Yuri A. Iriarte was supported by the internal project SEMILLERO UA-2022 (Chile).

Acknowledgments: The authors would like to thank the editor and the anonymous referees for their comments and suggestions, which significantly improved our manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

L	Lindley
LS	Lindley Slash
ESL	Extended Slash Lindley
PL	Power Lindley
ML	Maximum Likelihood
pdf	probability density function
cdf	cumulative distribution function
rf	reliability function
hrf	hazard rate function

Appendix A

This appendix provides the R codes for the computation of the pdf, the cdf, and the generation of pseudorandom numbers for the ESL distribution.

Appendix A.1. First Option for the Computation of the ESL pdf

This code computes Equation (3) using the INTEGRATE function.

```
> dESL <- function(x,theta,alpha,beta){
+   n <- length(x)
+   f <- numeric(n)
+   for (i in 1:n){
+     f[i] <- integrate(function(w,x,t,a,b){
+                 t^2/((t+1)*beta(a,b))*(1+x*w)*w^a*(1-w)^(b-1)*exp(-t*x*w)
+                 },
+               lower=0,upper=1,x=x[i],t=theta,a=alpha,b=beta)$value
+   }
+   return(f)
+ }
```

Appendix A.2. Second Option for the Computation of the ESL pdf

This code computes Equation (5) using the KUMMERM function.

```
> library(fAsianOptions)
> fj <- function(x,theta,alpha,beta,j){
+   gamma(alpha+j)/gamma(alpha+beta+j)*Re(kummerM(x=-theta*x,
+     a=alpha+j,b=alpha+beta+j))
+ }
> dESL <- function(x,theta,alpha,beta){
+   theta^2*gamma(beta)/((theta+1)*beta(alpha,beta))*
+     (fj(x,theta,alpha,beta,1)+x*fj(x,theta,alpha,beta,2))
+ }
```

Appendix A.3. First Option for the Computation of the ESL cdf

This code computes the ESL cdf using the code provided in Appendix A.1.

```
> pESL <- function(x,theta,alpha,beta){
+   n <- length(x)
+   f <- numeric(n)
```

```
+   for(i in 1:n){
+     f[i] <- integrate(dESL,lower=0,upper=x[i],theta=theta,
+                 alpha=alpha,beta=beta)$value
+   }
+ return(f)
+ }
```

Appendix A.4. Second Option for the Computation of the ESL cdf

This code computes the cdf provided in Proposition 2.

```
> funE <- function(x,theta,alpha,beta){
+   n <- length(x)
+   f <- numeric(n)
+   for(i in 1:n){
+     f[i] <- integrate(function(x,v,t,a,b)(1+v)*pbeta(v/x,a,b)*exp(-t*v),
+          lower=0,upper=x[i],x=x[i],t=theta,a=alpha,b=beta)$value
+   }
+ return(theta^2/(theta+1)*f)
+ }
> pESL <- function(x,theta,alpha,beta){
+   1-exp(-theta*x)*(1+theta+theta*x)/(1+theta)-funE(x,theta,alpha,beta)
+ }
```

Appendix A.5. Code to Generate a Pseudorandom Sample from the ESL Distribution

```
>  rESL <- function(n,theta,alpha,beta){
+    s <- rexp(n,theta)
+    w <- rgamma(n,2,theta)
+    v <- rbinom(n,1,theta/(1+theta))
+    u <- rbeta(n,alpha,beta)
+      (v*s+(1-v)*w)/u
+  }
>  x <- rESL(1000,0.01,6,5)
```

Appendix B

State Personal Income Data

Table A1. Set of 96 annual observations on state personal income (total, nominal, in millions of dollars) recorded in the 48 continental United States.

10.293195	11.577261	12.243384	12.448607	14.229156	14.454129
14.575292	14.581495	15.767469	16.296835	17.258916	18.237436
19.462380	20.852964	21.778072	22.868920	23.786644	25.045934
25.678534	26.210736	28.649564	31.716160	32.611268	34.784360
36.205164	36.293064	37.278220	37.902896	38.536176	39.377292
42.703144	43.395580	43.956936	45.995496	46.014968	46.241956
49.466672	53.431900	56.626672	57.749668	60.063368	60.170928
63.152360	63.333300	64.846548	65.732720	69.341920	71.209312
71.751616	72.050072	74.079712	74.851664	78.364336	79.104656
83.903280	84.572688	87.361632	88.870496	92.946544	98.328688
104.315120	113.216856	114.259984	115.959680	117.639672	126.525008
129.680832	133.549208	133.728040	135.115456	153.455776	157.633568
159.800448	161.441792	166.919248	170.033840	170.051568	176.786352
231.003152	231.594240	233.208576	255.312928	285.923232	297.728512
304.767456	333.525344	402.096768	447.102816	503.163328	771.470144

References

1. Lindley, D.V. Fiducial distributions and Bayes' theorem. *J. R. Stat. Soc.* **1958**, *20*, 102–107. [CrossRef]
2. Lindley, D.V. *Introduction to Probability and Statistics from a Bayesian Viewpoint, Part 2, Inference*; Cambridge University Press: New York, NY, USA, 1965.
3. Ghitany, M.E.; Atieh, B.; Nadarajah, S. Lindley distribution and its application. *Math. Comput. Simul.* **2008**, *78*, 493–506. [CrossRef]
4. Ghitany, M.E.; Alqallaf, F.; Al-Mutairi, D.K.; Husain, H.A. A two-parameter weighted Lindley distribution and its applications to survival data. *Math. Comput. Simul.* **2011**, *81*, 1190–1201. [CrossRef]
5. Nadarajah, S.; Bakouch, H.S.; Tahmasbi, R. A generalized Lindley distribution. *Sankhya B* **2011**, *73*, 331–359. [CrossRef]
6. Ghitany, M.E.; Al-Mutairi, D.K.; Balakrishnan, N.; Al-Enezi, L.J. Power Lindley distribution and associated inference. *Comput. Stat. Data. Anal.* **2013**, *64*, 20–33. [CrossRef]
7. Shanker, R.; Mishra, A. A quasi Lindley distribution. *Afr. J. Math. Comput. Sci. Res.* **2013**, *6*, 64–71.
8. Shanker, R.; Sharma, S.; Shanker, R. A two-parameter Lindley distribution for modeling waiting and survival times data. *Appl. Math.* **2013**, *4*, 363–368. [CrossRef]
9. Shanker, R.; Kamlesh, K.K.; Fesshaye, H. A two parameter Lindley distribution: Its properties and applications. *Biostat. Biom. Open Access J.* **2017**, *1*, 85–90.
10. Zakerzadeh, H.; Dolati, A. Generalized Lindley distribution. *J. Math. Ext.* **2009**, *3*, 13–25.
11. Shanker, R.; Shukla, K.K.; Shanker, R.; Tekie, A.L. A three-parameter Lindley distribution. *Am. J. Math. Stat.* **2017**, *7*, 15–26.
12. Ekhosuehi, N.; Opone, F. A three parameter generalized Lindley distribution: Properties and application. *Statistica* **2018**, *78*, 233–249.
13. Gui, W. Statistical properties and applications of the Lindley slash distribution. *J. Appl. Statist. Sci.* **2012**, *20*, 283.
14. Iriarte, Y.A.; Rojas, M.A. Slashed power-Lindley distribution. *Commun. Stat.-Theory Methods* **2019**, *48*, 1709–1720. [CrossRef]
15. Gómez, H.W.; Quintana, F.A.; Torres, F.J. A new family of slash-distributions with elliptical contours. *Stat. Probab. Lett.* **2007**, *77*, 717–725. [CrossRef]
16. Gui, W. A generalization of the slash half normal distribution: Properties and inferences. *J. Stat. Theory Pract.* **2014**, *8*, 283–296. [CrossRef]
17. Iriarte, Y.A.; Gómez, H.W.; Varela, H.; Bolfarine, H. Slashed Rayleigh distribution. *Rev. Colomb. Estad.* **2015**, *38*, 31–44. [CrossRef]
18. Iriarte, Y.A.; Varela, H.; Gómez, H.J.; Gómez, H.W. A gamma-type distribution with applications. *Symmetry* **2020**, *12*, 870. [CrossRef]
19. Gómez, H.J.; Gallardo, D.I.; Santoro, K.I. Slash truncation positive normal distribution and its estimation based on the EM algorithm. *Symmetry* **2021**, *13*, 2164. [CrossRef]
20. Reyes, J.; Barranco-Chamorro, I.; Gallardo, D.I.; Gómez, H.W. Generalized modified slash Birnbaum-Saunders distribution. *Symmetry* **2018**, *10*, 724. [CrossRef]
21. Rojas, M.A.; Bolfarine, H.; Gómez, H.W. An extension of the slash-elliptical distribution. *SORT* **2014**, *38*, 215–230.
22. Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*, 2nd ed.; John Wile & Sons; Hoboken, NJ, USA, 1995; Volume 2, ISBN 0-471-58494-0.
23. Andrews, L.C. *Special Functions of Mathematics for Engineers*, 2nd ed.; SPIE—The International Society for Optical Engineering: Bellingham, WA, USA, 1998.
24. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021.
25. Diethelm W.; Tobias S. *fAsianOptions: Rmetrics-EBM and Asian Option Valuation*; R Package Version 3042.82; R Foundation for Statistical Computing: Vienna, Austria, 2017.
26. Byrd, R.H.; Lu, P.; Nocedal, J.; Zhu, C. A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* **1995**, *16*, 1190–1208. [CrossRef]
27. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **1974**, *19*, 716–723. [CrossRef]
28. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]
29. Chen, G.; Balakrishnan, N. A general purpose approximate goodness-of-fit test. *J. Qual. Technol.* **1995**, *27*, 154–161. [CrossRef]
30. Hubert, M.; Vandervieren, E. An adjusted boxplot for skewed distributions. *Comput. Stat. Data. Anal.* **2008**, *52*, 5186–5201. [CrossRef]
31. Maechler M.; Rousseeuw P.; Croux C.; Todorov V.; Ruckstuhl A.; Salibian-Barrera M.; Verbeke T.; Koller M.; Conceicao E.L.T.; di Palma, M.A. *Robustbase: Basic Robust Statistics*; R Package Version 0.95-0; R Foundation for Statistical Computing: Vienna, Austria, 2022.
32. Lyu, M.R. *Handbook of Software Reliability Engineering*; IEEE Computer Society Press: Los Alamitos, CA, USA, 1996.
33. Astorga, J.M.; Iriarte, Y.A.; Gómez, H.W.; Bolfarine, H. Modified slashed generalized exponential distribution. *Commun. Stat.-Theory Methods.* **2020**, *49*, 4603–4617. [CrossRef]
34. Stock, J.H.; Watson, M.W. *Introduction to Econometrics*, 2nd ed.; Addison Wesley: Boston, MA, USA, 2007.