Research Article

Christophe Chesneau*

A note on an extreme left skewed unit distribution: Theory, modelling and data fitting

<https://doi.org/10.1515/stat-2020-0103>

Received Nov 04, 2020; accepted May 06, 2021

Abstract: In probability and statistics, unit distributions are used to model proportions, rates, and percentages, among other things. This paper is about a new one-parameter unit distribution, whose probability density function is defined by an original ratio of power and logarithmic functions. This function has a wide range of J shapes, some of which are more angular than others. In this sense, the proposed distribution can be thought of as an "extremely left skewed alternative" to the traditional power distribution. We discuss its main characteristics, including other features of the probability density function, some stochastic order results, the closed-form expression of the cumulative distribution function involving special integral functions, the quantile and hazard rate functions, simple expressions for the ordinary moments, skewness, kurtosis, moments generating function, incomplete moments, logarithmic moments and logarithmically weighted moments. Subsequently, a simple example of an application is given by the use of simulated data, with fair comparison to the power model supported by numerical and graphical illustrations. A new modelling strategy beyond the unit domain is also proposed and developed, with an application to a survival times data set.

Keywords: Unit distributions, special integral functions, mathematical inequalities, moments, data analysis.

2020 Mathematics Subject Classification: 60E05, 62E15, 62F10

1 Introduction

In recent years, there has been a significant increase in the development of unit distributions. They work on the modelling of a variety of phenomena involving unit data, such as proportions, probabilities, and percentages, among other things. There are also compositional data to consider (see Aitchison (1982)). The analysis of unit data, in particular, necessitates the development of parametric, semi-parametric, and regression models. Among the most useful unit distributions, there are the famous power (Po) distribution, unit gamma (UG) distribution established in Consul and Jain (1971), log-Lindley (LL) distribution proposed in Gómez-Déniz *et al.* (2014), unit Weibull (UW) distribution developed in Mazucheli *et al.* (2018) and later refined in Mazucheli *et al.* (2020), unit Gompertz (UG) distribution developed in Mazucheli *et al.* (2019), unit Birnbaum-Saunders (UBS) distribution studied in Mazucheli *et al.* (2018), log-xgamma (LXG) distribution established in Altun and Hamedani (2018), unit inverse Gaussian (UIG) distribution introduced in Ghitany *et al.* (2019), unit generalized half-normal (UGHN) distribution proposed in Korkmaz (2020), unit Johnson SU (UJSU) distribution established in Gündüz and Korkmaz (2020), log-weighted exponential (LWE) distribution developed in Altun (2020), unit Rayleigh (UR) distribution studied in Bantan *et al.* (2020), unit modified Burr-III (UMBIII) distribution examined in Haq *et al.* (2020), arcsecant hyperbolic normal (ASHN) distribution proposed in Korkmaz *et al.* (2021), unit Burr-XII (UBXII) distribution developed in Korkmaz and Chesneau (2021), transmuted unit Rayleigh (TUR) distribution studied in Korkmaz *et al.* (2021) and unit half-normal (UHN) distribution created in Bakouch *et al.* (2021).

*Corresponding Author: **Christophe Chesneau:** Normandie Université, UNICAEN, CNRS, LMNO, 14000 Caen, France; Email: christophe.chesneau@unicaen.fr

The majority of these unit distributions are the result of complex mathematical transformations of well-known flexible distributions with larger domains (gamma, Lindley, Weibull, normal, half-normal, among others). Depending on the modelling goals, they have different complexity structures. In this article, beyond the transformation scheme, we propose a new and simple unit distribution combining the following interesting properties: (i) it depends on a single positive parameter, (ii) its probability density function (pdf) is defined as an original ratio of power and logarithmic functions, (iii) its pdf is increasing and can be highly asymmetric on the left, with different types of angular and J forms, which is a relatively uncommon property for a one-parameter unit distribution, (iv) it enjoys strong results in stochastic orders, also involving the Po distribution, (v) its cumulative distribution function (cdf) and hazard rate function (hrf) have a closed-form depending on well-referenced integral functions, (vi) simple expressions exist for diverse moments related quantities, such as ordinary moments, moment generating function and incomplete moments, (vii) simple expressions are found for the logarithmic and logarithmically weighted moments, which is a rare property in a unit distribution, (viii) the behavior of the moments skewness and kurtosis of the distribution are quite manageable and (ix) it has a high degree of applicability and can serve as generator for the creation of new statistical models. Theoretical results, graphics, and numerical works are used to describe these statements in detail. From a theoretical standpoint, the new and Po distributions are contrasted and discussed, revealing some relevant relationships. The new distribution can be referred to as an "extreme left skewed alternative" to the Po distribution because of immediate analytical similarities and a large panel of J shapes in its pdf. Following that, a part is devoted to the inference of the proposed model. Extremely left-skewed data are used to perform a fitting analysis, which yields much better results than the Po model. Furthermore, based on the proposed unit distribution, new general and flexible models are developed and applied to a real data set. Finally, we would like to point out that the article contains a number of inequalities that are used as intermediates in the proofs but may be of interest on their own.

The reminder of the article is as follows. Section 2 defines the new unit distribution through its pdf and cdf, with important properties related to these functions. Section 3 is devoted to various moments related quantities. An example of statistical application is provided in Section 4. A new general modelling strategy is proposed in Section 5, beyond the unit domain. The article concludes in Section 6.

2 The unit power-log distribution

The proposed distribution is presented in this section.

2.1 Probability density function

The following proposition introduces a special function with pdf-like properties.

Proposition 2.1. *The following function has the properties of a pdf:*

$$f_{\alpha}(x) = \frac{1}{\log(1+\alpha)} \frac{x^{\alpha} - 1}{\log(x)}, \quad x \in (0, 1), \quad (1)$$

and $f_{\alpha}(x) = 0$ for $x \notin (0, 1)$, where $\alpha > 0$.

Proof. First, since $\log(1+\alpha) > 0$ for $\alpha > 0$, $x^{\alpha} - 1 < 0$ and $\log(x) < 0$ for $x \in (0, 1)$, it is clear that $f_{\alpha}(x)$ is positive. Also, it is piecewise continuous on $\mathbb{R}$. Now, let us consider the following integral function depending on α :

$$\Phi(\alpha) = \int_0^1 \frac{x^{\alpha} - 1}{\log(x)} dx,$$

with the natural extension $\Phi(\alpha) = 0$ for $\alpha = 0$. By applying the Leibnitz integral rule, we obtain

$$\frac{\partial}{\partial \alpha} \Phi(\alpha) = \frac{\partial}{\partial \alpha} \int_0^1 \frac{x^\alpha - 1}{\log(x)} dx = \int_0^1 \frac{\partial}{\partial \alpha} \frac{x^\alpha - 1}{\log(x)} dx = \int_0^1 x^\alpha dx = \frac{1}{1 + \alpha}.$$

Upon integrating with respect to α , we obtain $\Phi(\alpha) = \log(1 + \alpha) + c$ for a certain constant c . Since $\Phi(0) = 0$, we have $c = 0$, implying that $\Phi(\alpha) = \log(1 + \alpha)$. Therefore,

$$\int_{-\infty}^{+\infty} f_\alpha(x) dx = \int_0^1 f_\alpha(x) dx = \frac{1}{\log(1 + \alpha)} \Phi(\alpha) = 1.$$

This ends the proof of Proposition 2.1. $\square$

The distribution related to the pdf $f_\alpha(x)$ is called the unit power-log (UPL) distribution or $\text{UPL}(\alpha)$ when the parameter α needs to be mentioned. That is, we say that a random variable X follows the UPL distribution over a probability space formally denoted by $(\Omega, \mathcal{F}, \mathbb{P})$ if, for any set A into $\mathbb{R}$, we have $\mathbb{P}(X \in A) = \int_A f_\alpha(x) dx$. The interesting properties of the UPL distribution are the objects of the rest of the study.

Although it is presumed in the sequel that $\alpha > 0$, the case $\alpha \in (-1, 0)$ is not excluded, and will be briefly discussed in a separate subsection (see Subsection 2.5).

First of all, some basic properties of $f_\alpha(x)$ are presented below.

Proposition 2.2.

- When $\alpha \rightarrow 0$, $f_\alpha(x)$ tends to the pdf of the unit uniform distribution.
- $f_\alpha(x)$ is strictly increasing for $x \in (0, 1)$.
- The following asymptotic behavior of $f_\alpha(x)$ at the boundaries holds:

$$f_\alpha(x) \underset{x \rightarrow 0}{\sim} -\frac{1}{\log(1 + \alpha)} \frac{1}{\log(x)} \rightarrow 0, \quad f_\alpha(x) \underset{x \rightarrow 1}{=} \frac{\alpha}{\log(1 + \alpha)}.$$

Proof.

- By applying the following equivalences: $\log(1 + y) \sim y$ and $e^y \sim 1 + y$ when $y \rightarrow 0$, when $\alpha \rightarrow 0$ and $x \in (0, 1)$, we get

$$f_\alpha(x) = \frac{1}{\log(1 + \alpha)} \frac{e^{\alpha \log(x)} - 1}{\log(x)} \sim \frac{1}{\alpha} \frac{1 + \alpha \log(x) - 1}{\log(x)} = 1.$$

We still have $f_\alpha(x) = 0$ for $x \notin (0, 1)$. Hence, $f_\alpha(x) \rightarrow g(x)$, where $g(x) = 1$ for $x \in (0, 1)$, $g(x) = 0$ for $x \notin (0, 1)$, corresponding to the pdf of the unit uniform distribution.

- We have

$$\frac{\partial}{\partial x} f_\alpha(x) = \frac{1}{\log(1 + \alpha)} \frac{\alpha x^\alpha \log(x) + 1 - x^\alpha}{x(\log(x))^2}.$$

Let us study the sign of the numerator term. In this aim, the following general logarithmic inequality holds:

$$\log(1 + y) \geq \frac{y}{1 + y}, \quad y > -1, \quad (2)$$

where the equality is reached only for $y = 0$. By choosing $y = x^\alpha - 1 \in (-1, 0)$, we obtain

$$\alpha \log(x) = \log(x^\alpha) = \log[1 + (x^\alpha - 1)] > \frac{x^\alpha - 1}{1 + (x^\alpha - 1)} = \frac{x^\alpha - 1}{x^\alpha},$$

implying that $\alpha x^\alpha \log(x) + 1 - x^\alpha > 0$ for $x \in (0, 1)$. Since $x(\log(x))^2 > 0$ for $x \in (0, 1)$, we arrive at $\partial f_\alpha(x) / \partial x > 0$, proving that $f_\alpha(x)$ is strictly increasing.

- When $x \rightarrow 0$, we have

$$f_{\alpha}(x) \sim \frac{1}{\log(1+\alpha)} \frac{0-1}{\log(x)} = -\frac{1}{\log(1+\alpha)} \frac{1}{\log(x)} \rightarrow 0$$

and, when $x \rightarrow 1$, by using $e^y \sim 1+y$ when $y \rightarrow 0$, we get

$$f_{\alpha}(x) = \frac{1}{\log(1+\alpha)} \frac{e^{\alpha \log(x)} - 1}{\log(x)} \sim \frac{1}{\log(1+\alpha)} \frac{1 + \alpha \log(x) - 1}{\log(x)} = \frac{\alpha}{\log(1+\alpha)}.$$

This ends the proof of Proposition 2.2. $\square$

Figure 1 completes Proposition 2.2 through a graphical analysis; it shows the possible shapes of $f_{\alpha}(x)$ for several values of α .

Figure 1: Plots for the pdf $f_{\alpha}(x)$ for various values of α

From Figure 1, we see that $f_{\alpha}(x)$ is near angular for very small α , and increasing in all circumstances, illustrating the findings in Proposition 2.2. Also, we can notice that $f_{\alpha}(x)$ can be concave, such as the green curve, and have tilde shapes or “concave then convex” shapes, such as the pink curve. It is clear that the UPL distribution is mainly left skewed, reaching some extreme behavior in this regard, such as the yellow or gray curves.

Some comparisons between the UPL and Po distributions are now formulated.

- First, we recall that the Po distribution, or $\text{Po}(\alpha)$ distribution, can be defined with the following pdf:

$$k_{\alpha}(x) = (\alpha + 1)x^{\alpha}, \quad x \in (0, 1),$$

and $k_{\alpha}(x) = 0$ for $x \notin (0, 1)$. Hence, the pdf of the UPL distribution is a weighted version of the pdf of the Po distribution; we can write $f_{\alpha}(x) = w_{\alpha}(x)k_{\alpha}(x)$, with $w_{\alpha}(x) = (1 - x^{-\alpha})/[(\alpha + 1)\log(x)\log(1 + \alpha)]$.

- Based on Equation (2), we always have $\alpha / \log(1 + \alpha) = \lim_{x \rightarrow 1} f_\alpha(x) < \lim_{x \rightarrow 1} k_\alpha(x) = \alpha + 1$. However, there is no analytical hierarchy between $f_\alpha(x)$ and $k_\alpha(x)$ for all $x \in (0, 1)$, as confirmed by the curves in Figure 2.

Figure 2: Plots for the difference $f_\alpha(x) - k_\alpha(x)$ for various values of α

From Figure 2, it is clear that the sign of $f_\alpha(x) - k_\alpha(x)$ is not always positive or negative for $x \in (0, 1)$.

- In contrast to the pdf $f_\alpha(x)$, which can present tilde shapes, the pdf $k_\alpha(x)$ is either concave or convex, depending on $\alpha < 1$ and $\alpha > 1$, respectively.

However, some stochastic connections behind $f_\alpha(x)$ and $k_\alpha(x)$ exist and will be presented in the next subsection.

2.2 Relevant stochastic order results

We now present relevant stochastic order results satisfied by the UPL distribution through the use of $f_\alpha(x)$. We adopt the concept of likelihood ratio order as presented in Shaked and Shanthikumar (2007) in the continuous case. That is, let X and Y be two continuous random variables with pdfs $f(x)$ and $g(x)$, respectively, so that $f(x)/g(x)$ decreases (or $g(x)/f(x)$ increases) in x over the union of the domains of X and Y . Then, in the likelihood ratio order, X is smaller than Y .

Proposition 2.3. *Let X be a random variable with the $\text{UPL}(\alpha_1)$ distribution and Y be a random variable with the $\text{UPL}(\alpha_2)$ distribution, with $\alpha_2 \geq \alpha_1$. Then X is smaller than Y in the likelihood ratio order.*

Proof. Proving that X is smaller than Y in the likelihood ratio order is equivalent to prove that the following ratio function is decreasing with respect to x :

$$q_{\alpha_1, \alpha_2}(x) = \frac{f_{\alpha_1}(x)}{f_{\alpha_2}(x)} = \frac{\log(1 + \alpha_2) x^{\alpha_1} - 1}{\log(1 + \alpha_1) x^{\alpha_2} - 1}, \quad x \in (0, 1).$$

Now, we have

$$\frac{\partial}{\partial x} q_{\alpha_1, \alpha_2}(x) = \frac{\log(1 + \alpha_2)}{\log(1 + \alpha_1)} \frac{\alpha_2 x^{\alpha_2-1} (1 - x^{\alpha_1}) - \alpha_1 x^{\alpha_1-1} (1 - x^{\alpha_2})}{(x^{\alpha_2} - 1)^2}. \quad (3)$$

In order to study the sign of this function, let us set $\theta(u) = ux^{u-1}/(1 - x^u)$ for $u > 0$ and $x \in (0, 1)$. Then, we have

$$\frac{\partial}{\partial u} \theta(u) = \frac{x^{u-1}}{(1 - x^u)^2} (1 - x^u + u \log(x)).$$

Owing to the following inequality: $\log(1 + y) < y$ for $y \in (-1, 0)$, applied with $y = x^u - 1 \in (-1, 0)$, we have

$$u \log(x) = \log[1 + (x^u - 1)] < x^u - 1. \quad (4)$$

As a result, $\partial \theta(u)/\partial u < 0$ and $\theta(u)$ is strictly decreasing. That is, for $\alpha_2 \geq \alpha_1$, we have $\theta(\alpha_2) \leq \theta(\alpha_1)$, so $\alpha_2 x^{\alpha_2-1}/(1 - x^{\alpha_2}) \leq \alpha_1 x^{\alpha_1-1}/(1 - x^{\alpha_1})$, which implies that

$$\alpha_2 x^{\alpha_2-1} (1 - x^{\alpha_1}) - \alpha_1 x^{\alpha_1-1} (1 - x^{\alpha_2}) \leq 0.$$

Therefore, based on Equation (3), we have $\partial q_{\alpha_1, \alpha_2}(x)/\partial x \leq 0$, so $q_{\alpha_1, \alpha_2}(x)$ is decreasing. It is established that the desired likelihood ratio order exists. $\square$

The result of Proposition 2.3 is illustrated in Figure 3, where the ratio function $f_{\alpha_1}(x)/f_{\alpha_2}(x)$ is plotted with $\alpha_1 = \alpha$, $\alpha = 0.5, 5$, and $\alpha_2 = \alpha + \epsilon$ with various positive values for ϵ such that $\alpha_2 \geq \alpha_1$.

Figure 3: Plots for the ratio function $f_{\alpha_1}(x)/f_{\alpha_2}(x)$ for (a) $\alpha_1 = 0.5$ and $\alpha_2 = 0.5 + \epsilon$, and (b) $\alpha_1 = 5$ and $\alpha_2 = 5 + \epsilon$, with several positive values of ϵ

From this figure, we see that all the curves are decreasing, supporting the statement of the proposition.

We now link the UPL and Po distributions by using the likelihood ratio order.

Proposition 2.4. Let X be a random variable with the $UPL(\alpha)$ distribution and Y be a random variable with the $Po(\alpha)$ distribution. Then X is smaller than Y in the likelihood ratio order.

Proof. Let us prove that the following ratio function is decreasing with respect to x :

$$q_\alpha(x) = \frac{f_\alpha(x)}{k_\alpha(x)} = \frac{1}{(\alpha + 1) \log(1 + \alpha)} \frac{x^\alpha - 1}{x^\alpha \log(x)}, \quad x \in (0, 1).$$

Now, we have

$$\frac{\partial}{\partial x} q_\alpha(x) = \frac{1}{(\alpha + 1) \log(1 + \alpha)} \frac{x^{-\alpha-1} [\alpha \log(x) + 1 - x^\alpha]}{(\log(x))^2}.$$

With the same arguments to Equation (4), we have $\alpha \log(x) + 1 - x^\alpha < 0$ for $x \in (0, 1)$, implying that $\partial q_\alpha(x)/\partial x < 0$, so $q_\alpha(x)$ is decreasing. The desired likelihood ratio order is provided. $\square$

Proposition 2.4 has several consequences on other characteristics of the UPL and Po distributions that will be presented along the article.

2.3 Cumulative distribution function

The following proposition presents the cdf of the UPL distribution through special integral functions that are well-referenced in the literature.

Proposition 2.5. The cdf of the UPL distribution is obtained as

$$F_\alpha(x) = \frac{1}{\log(1 + \alpha)} (\text{Ei}[(1 + \alpha) \log(x)] - \text{Li}(x)), \quad x \in (0, 1),$$

$F_\alpha(x) = 0$ for $x \leq 0$ and $F_\alpha(x) = 1$ for $x \geq 1$, where $\text{Ei}(x)$ and $\text{Li}(x)$ denote the well-known exponential and logarithmic integrals, respectively, defined by

$$\text{Ei}(x) = - \int_{-x}^{+\infty} \frac{e^{-t}}{t} dt = \int_{-\infty}^x \frac{e^t}{t} dt, \quad x \in \mathbb{R}^*, \quad \text{Li}(x) = \text{Ei}(\log(x)) = \int_0^x \frac{1}{\log(t)} dt, \quad x \in (0, 1),$$

respectively

Proof. For $x \in (0, 1)$, we have

$$F_\alpha(x) = \int_{-\infty}^x f_\alpha(t) dt = \int_0^x \frac{1}{\log(1 + \alpha)} \frac{t^\alpha - 1}{\log(t)} dt = \frac{1}{\log(1 + \alpha)} \left(\int_0^x \frac{t^\alpha}{\log(t)} dt - \text{Li}(x) \right).$$

Now, by the successive changes of variables $t = e^{-y}$ and $z = (1 + \alpha)y$, we get

$$\int_0^x \frac{t^\alpha}{\log(t)} dt = - \int_{-\log(x)}^{+\infty} \frac{e^{-(1+\alpha)y}}{y} dy = - \int_{-(1+\alpha)\log(x)}^{+\infty} \frac{e^{-z}}{z} dz = \text{Ei}[(1 + \alpha) \log(x)].$$

We obtain the desired result by putting the above equalities together, concluding the proof of Proposition 2.5. $\square$

The numerical values of $\text{Ei}(x)$, $\text{Li}(x)$ and $F_\alpha(x)$ involve the use of standard numerical integration procedures available in every mathematical package. In this study, the software R is used (see R Core Team (2014)), combined with the package `pracma` and the function `expint_Ei`. Since $f_\alpha(x)$ is strictly increasing with respect to x for $x \in (0, 1)$ by Proposition 2.2, $F_\alpha(x)$ is a convex function only.

Figure 4 illustrates the possible shapes of $F_\alpha(x)$ for several values of α .

Figure 4: Plots for the cdf $F_\alpha(x)$ for various values of α

From Figure 4, we see that $F_\alpha(x)$ is convex. Various degrees of convexity are also observed.

From Proposition 2.4, the likelihood ratio order between the UPL and Po distributions implies that, for any $x \in \mathbb{R}$,

$$K_\alpha(x) \leq F_\alpha(x),$$

where $K_\alpha(x)$ denotes the cdf of the Po distribution, that is $K_\alpha(x) = x^{\alpha+1}$ for $x \in (0, 1)$, $K_\alpha(x) = 0$ for $x \leq 0$ and $K_\alpha(x) = 1$ for $x > 1$. See Shaked and Shanthikumar (2007) for more detail on this likelihood order consequence.

2.4 On the quantile and hazard rate functions

The numerical values of the quantiles of the UPL distribution can be derived from the cdf; the quantile x_u is obtained by solving the following non-linear equation with respect to u : $F_\alpha(x_u) = u$, that is

$$\text{Ei}[(1 + \alpha) \log(x_u)] - \text{Li}(x_u) = u \log(1 + \alpha).$$

Any mathematical software can be used to determine x_u for a given value of α .

The pdf and cdf of the UPL distribution allow us to define the hrf as

$$h_\alpha(x) = \frac{f_\alpha(x)}{1 - F_\alpha(x)} = \frac{x^\alpha - 1}{\log(x) [\log(1 + \alpha) - (\text{Ei}[(1 + \alpha) \log(x)] - \text{Li}(x))]}, \quad x \in (0, 1),$$

and $h_\alpha(x) = 0$ for $x \notin (0, 1)$. At the boundaries, $h_\alpha(x)$ has the following asymptotic behavior:

$$h_\alpha(x) \underset{x \rightarrow 0}{\sim} -\frac{1}{\log(1 + \alpha)} \frac{1}{\log(x)} \rightarrow 0, \quad h_\alpha(x) \underset{x \rightarrow 1}{=} +\infty.$$

Due to the complexity of the denominator term, the properties of this function are difficult to identify from an analytical point of view. However, a graphical analysis suggests that it is increasing, with diverse tilde or J shapes, as illustrated in Figure 5.

Figure 5: Plots for the hrf $h_\alpha(x)$ for various values of α

Further detail on the importance of the hrf in survival analysis in cases of bounded domain can be found in Klein and Moeschberger (1997).

2.5 On the case $\alpha \in (-1, 0)$

It should be noted that the pdf of the UPL distribution indicated in Proposition 2.1 is still valid for $\alpha \in (-1, 0)$. Certain properties of the UPL distribution are actually inverted in this situation. In particular, the following results are provable:

- $f_\alpha(x)$ is strictly decreasing for $x \in (0, 1)$, and the following limit at the origin holds:

$$f_\alpha(x) \underset{x \rightarrow 0}{\sim} -\frac{1}{\log(1+\alpha)} \frac{x^\alpha}{\log(x)} \rightarrow +\infty.$$

Also, the extreme left skewed property is no longer available.

- $F_\alpha(x)$ is concave.
- $h_\alpha(x)$ has diverse U shapes, and the following limit at the origin is valid:

$$h_\alpha(x) \underset{x \rightarrow 0}{\sim} -\frac{1}{\log(1+\alpha)} \frac{x^\alpha}{\log(x)} \rightarrow +\infty.$$

Some plots of $f_\alpha(x)$, $F_\alpha(x)$ and $h_\alpha(x)$ with selected negative values of α are presented in Figure 6, illustrating the results above. In addition, the following stochastic order result holds: Let X be a random variable with the $\text{UPL}(\alpha)$ distribution and Y be a random variable with the $\text{Po}(\alpha)$ distribution. Then Y is smaller than X in the likelihood ratio order. The contrary was shown for the case $\alpha > 0$.

More facts on the case $\alpha \in (-1, 0)$ may be the subject of an independent study, which we drop here for brevity.

Figure 6: Plots for the (a) pdf $f_\alpha(x)$, (b) cdf $F_\alpha(x)$ and (c) hrf $h_\alpha(x)$ for various values of $\alpha \in (-1, 0)$

3 Diverse moments

The moments of a distribution are essential to describe various of its characteristics and statistical capacities. Here, diverse moments of the UPL distribution are investigated theoretically and practically.

3.1 Ordinary moments

The expressions of the ordinary moments of the UPL distribution are discussed in the following result.

Proposition 3.1. *Let X be a random variable with the UPL distribution and r be a positive integer. Then, the r th ordinary moment of X is obtained as*

$$m_r = \mathbb{E}(X^r) = \frac{1}{\log(1 + \alpha)} \log \left(1 + \frac{\alpha}{1 + r} \right). \quad (5)$$

Proof. By the transfer formula, we have

$$m_r = \int_{-\infty}^{+\infty} x^r f_\alpha(x) dx = \frac{1}{\log(1 + \alpha)} \int_0^1 x^r \frac{x^\alpha - 1}{\log(x)} dx.$$

Let us now treat the integral term by following the spirit of the proof of Proposition 2.1. We consider the following integral function:

$$\Phi_r(\alpha) = \int_0^1 x^r \frac{x^\alpha - 1}{\log(x)} dx,$$

with $\Phi_r(\alpha) = 0$ for $\alpha = 0$. By virtue of the Leibnitz integral rule, we get

$$\frac{\partial}{\partial \alpha} \Phi_r(\alpha) = \frac{\partial}{\partial \alpha} \int_0^1 x^r \frac{x^\alpha - 1}{\log(x)} dx = \int_0^1 x^r \frac{\partial}{\partial \alpha} \frac{x^\alpha - 1}{\log(x)} dx = \int_0^1 x^{\alpha+r} dx = \frac{1}{1 + \alpha + r}.$$

Upon integrating with respect to α , we obtain $\Phi_r(\alpha) = \log(1 + \alpha + r) + c$ for a certain constant c . Since $\Phi_r(0) = 0$, we have $c = -\log(1 + r)$. Therefore

$$\Phi_r(\alpha) = \log(1 + \alpha + r) - \log(1 + r) = \log \left(1 + \frac{\alpha}{1 + r} \right).$$

By dividing by $\log(1 + \alpha)$, the given formula for m_r is obtained; the proof of Proposition 3.1 ends.

Alternative proof. The proof of Proposition 3.1 can be performed via series expansions. Indeed, it follows from the series expansion of the exponential function, the Lebesgue dominated convergence theorem, the following formula:

$$\int_0^1 x^v (\log(x))^s dx = (-1)^s \frac{s!}{(v+1)^{s+1}}, \quad (6)$$

where s denotes an integer and v a real number such that $v > -1$, which is a particular case of (Gradshteyn and Ryzhik, 2007, Equation 4.2726), and the series expansion of the logarithmic function, that

$$\begin{aligned} m_r &= \frac{1}{\log(1 + \alpha)} \int_0^1 x^r \frac{e^{\alpha \log(x)} - 1}{\log(x)} dx = \frac{1}{\log(1 + \alpha)} \int_0^1 x^r \left[\sum_{k=1}^{+\infty} \frac{\alpha^k}{k!} (\log(x))^{k-1} \right] dx \\ &= \frac{1}{\log(1 + \alpha)} \sum_{k=1}^{+\infty} \frac{\alpha^k}{k!} \int_0^1 x^r (\log(x))^{k-1} dx = \frac{1}{\log(1 + \alpha)} \sum_{k=1}^{+\infty} \frac{\alpha^k}{k!} (-1)^{k-1} \frac{(k-1)!}{(r+1)^k} \\ &= \frac{1}{\log(1 + \alpha)} \sum_{k=1}^{+\infty} \frac{(-1)^{k-1}}{k} \left(\frac{\alpha}{r+1} \right)^k = \frac{1}{\log(1 + \alpha)} \log \left(1 + \frac{\alpha}{1+r} \right). \end{aligned}$$

For the final step, it is assumed that $\alpha < 1 + r$, which is a technical constraint that the first proof does not have. $\square$

It is worth noting that Proposition 3.1 holds for every real number r such that $r > -1$.

The following result completes Proposition 3.1; it is about the monotonicity of m_r with respect to α and r .

Proposition 3.2. *By considering m_r defined in Proposition 3.1 as a function of α and r , the following results hold:*

- m_r is a strictly increasing function with respect to α ,
- m_r is a strictly decreasing function with respect to r .

Proof.

- Based on the expression of Equation (5), after some developments, we have

$$\frac{\partial}{\partial \alpha} m_r = \frac{1}{\log(1 + \alpha)^2} \left[\frac{1}{r + \alpha + 1} \log(1 + \alpha) - \frac{1}{1 + \alpha} \log \left(1 + \frac{\alpha}{1+r} \right) \right]. \quad (7)$$

In order to study the sign of this function, let us set $\omega(x) = (1 + x) \log(1 + x)/x$. Then we have

$$\frac{\partial}{\partial x} \omega(x) = \frac{x - \log(1 + x)}{x^2}.$$

Based on the following general logarithmic inequality: $\log(1 + y) < y$ for $y > 0$, we have $\partial \omega(x)/\partial x > 0$, implying that $\omega(x)$ is a strictly increasing function. In particular, for $x < y$, we have $\omega(x) < \omega(y)$, which is equivalent to

$$\frac{\log(1 + x)}{\log(1 + y)} < \frac{x}{1 + x} \frac{1 + y}{y}.$$

By taking $x = \alpha/(1 + r)$ and $y = \alpha$ such that $x < y$, we get

$$\frac{\log(1 + \alpha/(1 + r))}{\log(1 + \alpha)} < \frac{\alpha/(1 + r)}{1 + \alpha/(1 + r)} \frac{1 + \alpha}{\alpha} = \frac{1 + \alpha}{1 + r + \alpha},$$

so

$$\frac{1}{r + \alpha + 1} \log(1 + \alpha) - \frac{1}{1 + \alpha} \log \left(1 + \frac{\alpha}{1+r} \right) > 0.$$

This inequality combined with Equation (7) gives $\partial m_r / \partial \alpha > 0$, proving that m_r is a strictly increasing function with respect to α .

– The second point is more immediate; we have

$$\frac{\partial}{\partial r} m_r = -\frac{1}{\log(1+\alpha)} \frac{\alpha}{(1+r)(1+\alpha+r)} < 0,$$

implying that m_r is a strictly decreasing function with respect to r . In fact, this property holds for the r th ordinary moment for any unit distribution in general.

The proof of Proposition 3.2 ends. $\square$

Based on Proposition 3.1, the mean and variance of X are obtained as

$$m = m_1 = \frac{1}{\log(1+\alpha)} \log\left(1 + \frac{\alpha}{2}\right)$$

and

$$\sigma^2 = \mathbb{V}(X) = m_2 - m_1^2 = \frac{1}{\log(1+\alpha)} \log\left(1 + \frac{\alpha}{3}\right) - \frac{1}{(\log(1+\alpha))^2} \left[\log\left(1 + \frac{\alpha}{2}\right)\right]^2,$$

respectively. The skewness and kurtosis coefficients of X are given as

$$\gamma_1 = \frac{m_3 - 3m\sigma^2 - m^3}{\sigma^3}, \quad \beta_4 = \frac{m_4 - 4m_3m + 6m_2m^2 - 3m^4}{\sigma^4},$$

respectively. Table 1 presents some numerical values for these measures by taking several values of α .

Table 1: Numerical values for the first four ordinary moments, skewness and kurtosis coefficients of the UPL distribution for selected values of α

α	m_1	m_2	m_3	m_4	σ^2	γ_1	β_4
0.01	0.501	0.334	0.251	0.201	0.083	−0.004	1.801
0.1	0.512	0.344	0.259	0.208	0.082	−0.041	1.814
0.5	0.550	0.380	0.290	0.235	0.077	−0.178	1.887
1.5	0.611	0.443	0.348	0.286	0.070	−0.415	2.119
3	0.661	0.500	0.404	0.339	0.063	−0.646	2.463
8	0.732	0.591	0.500	0.435	0.055	−1.046	3.298
100	0.852	0.766	0.706	0.66	0.040	−1.925	6.269
200000	0.943	0.910	0.886	0.868	0.020	−3.539	16.479

Table 1 shows that the UPL distribution is mostly left skewed, with a lot of kurtosis flexibility. Proposition 3.2 is also illustrated for the considered values; it is clear that m_r increases as α increases, and m_r decreases as r increases. Also, from this table, we remark that σ^2 and γ_1 decrease as α increases, and β_4 increases as α increases, for the considered values of α .

From Proposition 2.4, the likelihood ratio order between the UPL and Po distributions implies that, for any integer r ,

$$m_r \leq m_r^o,$$

where m_r^o denotes the r th ordinary moment of a random variable with the Po distribution. We refer the reader to Shaked and Shanthikumar (2007) for the details on this consequence.

The following proposition is about the moment generating function of the UPL distribution.

Proposition 3.3. *Let X be a random variable with the UPL distribution. Then, the moment generating function of X is obtained as*

$$M(t) = \mathbb{E}(e^{tX}) = \frac{1}{\log(1+\alpha)} \sum_{k=0}^{+\infty} \frac{t^k}{k!} \log\left(1 + \frac{\alpha}{1+k}\right), \quad t \in \mathbb{R}.$$

Proof. The proof is an immediate consequence of the following formula:

$$\mathbb{E}(e^{tX}) = \mathbb{E}\left(\sum_{k=0}^{+\infty} \frac{t^k}{k!} X^k\right) = \sum_{k=0}^{+\infty} \frac{t^k}{k!} m_k,$$

combined with Proposition 3.1. $\square$

Basically, the r th ordinary moment can be re-found by the following relation: $m_r = \partial^r M(t) / \partial t^r |_{t=0}$. Also, the r th cumulant of X is given by $\kappa_r = \partial^r \{\log M(t)\} / \partial t^r |_{t=0}$. In particular, we have $\kappa_1 = m_1$, κ_2 is the second central moment of X and κ_3 is the third one.

3.2 Incomplete moments

The incomplete moment of the UPL distribution is now considered.

Proposition 3.4. *Let X be a random variable with the UPL distribution, r be a positive integer and $t \in (0, 1)$. Then, the r th incomplete moment of X at t is obtained as*

$$m_r(t) = \mathbb{E}(X^r I(X \leq t)) = \frac{1}{\log(1 + \alpha)} [\text{Ei}((1 + r + \alpha) \log(t)) - \text{Ei}((1 + r) \log(t))],$$

where $I(\mathfrak{A})$ denotes the indicator function over an event $\mathfrak{A}$.

Proof. By the transfer formula, we have

$$m_r(t) = \int_{-\infty}^t x^r f_\alpha(x) dx = \frac{1}{\log(1 + \alpha)} \int_0^t x^r \frac{x^\alpha - 1}{\log(x)} dx.$$

Let us now treat the integral term as in the proof of Proposition 2.5. We consider the following incomplete integral function:

$$\Phi_r(t, \alpha) = \int_0^t x^r \frac{x^\alpha - 1}{\log(x)} dx,$$

with $\Phi_r(t, \alpha) = 0$ for $\alpha = 0$. By virtue of the Leibnitz integral rule, we get

$$\frac{\partial}{\partial \alpha} \Phi_r(t, \alpha) = \frac{\partial}{\partial \alpha} \int_0^t x^r \frac{x^\alpha - 1}{\log(x)} dx = \int_0^t x^r \frac{\partial}{\partial \alpha} \frac{x^\alpha - 1}{\log(x)} dx = \int_0^t x^{\alpha+r} dx = \frac{t^{1+\alpha+r}}{1 + \alpha + r}.$$

Now, by applying the change of variables $v = 1 + u + r$ then $w = v \log(t)$, we arrive at

$$\int_0^\alpha \frac{t^{1+u+r}}{1 + u + r} du = \int_{1+r}^{1+r+\alpha} \frac{e^{v \log(t)}}{v} dv = \int_{(1+r) \log(t)}^{(1+r+\alpha) \log(t)} \frac{e^w}{w} dw = \text{Ei}((1 + r + \alpha) \log(t)) - \text{Ei}((1 + r) \log(t)).$$

Therefore

$$\Phi_r(t, \alpha) = \text{Ei}((1 + r + \alpha) \log(t)) - \text{Ei}((1 + r) \log(t)) + c,$$

for a certain constant c . Since $\Phi_r(t, 0) = 0$, we get $c = 0$. By putting the above equalities together, we have

$$m_r(t) = \frac{1}{\log(1 + \alpha)} [\text{Ei}((1 + r + \alpha) \log(t)) - \text{Ei}((1 + r) \log(t))].$$

The proof of Proposition 3.4 ends. $\square$

Based on Proposition 3.4, one may re-find the cdf and ordinary moments of the UPL distribution by the following relations: $F_\alpha(x) = m_0(x)$ and $m_r = \lim_{t \rightarrow 1} m_r(t)$, respectively. Also, we can define the normalized incomplete moment given by

$$\psi_r(t) = \frac{m_r(t)}{m_r} = \frac{1}{\log(1 + \alpha/(1+r))} [\text{Ei}((1+r+\alpha)\log(t)) - \text{Ei}((1+r)\log(t))], \quad t \in (0, 1).$$

In full generality, the normalized incomplete moments given as $\psi_0(t)$ and $\psi_1(t)$ are involved in many applications and provide interesting distributional information. Also, they allow to define the famous Gini coefficient defined in our setting as

$$G_\alpha = \int_0^1 f_\alpha(x) \left[\frac{x}{m} \psi_0(x) - \psi_1(x) \right] dx = \frac{1}{\log(1 + \alpha)} \int_0^1 \frac{x^\alpha - 1}{\log(x)} \left[\frac{x}{m} \psi_0(x) - \psi_1(x) \right] dx.$$

Other important measures can be defined in a similar way, such as the Lorenz curve, Bonferroni curve and Pietra measure. In this regard, we refer to Butler and McDonald (1989) and, in a more general setting, Cordeiro *et al.* (2020).

3.3 Logarithmic moments

Logarithmic moments of a distribution occur naturally in various probability quantities such as entropy and concentration inequalities. Here, we provide the expressions of the logarithmic moments of the UPL distribution.

Proposition 3.5. *Let X be a random variable with the UPL distribution and r be a positive integer. Then, the r th logarithmic moment of X is obtained as*

$$m_r^b = \mathbb{E}((\log(X))^r) = \frac{1}{\log(1 + \alpha)} (-1)^r (r-1)! \frac{(\alpha+1)^r - 1}{(\alpha+1)^r}.$$

Proof. By the transfer formula and Equation (6), we obtain

$$\begin{aligned} m_r^b &= \int_{-\infty}^{+\infty} (\log(x))^r f_\alpha(x) dx = \frac{1}{\log(1 + \alpha)} \int_0^1 (\log(x))^r \frac{x^\alpha - 1}{\log(x)} dx \\ &= \frac{1}{\log(1 + \alpha)} \left(\int_0^1 x^\alpha (\log(x))^{r-1} dx - \int_0^1 (\log(x))^{r-1} dx \right) = \frac{1}{\log(1 + \alpha)} \left((-1)^{r-1} \frac{(r-1)!}{(\alpha+1)^r} - (-1)^{r-1} (r-1)! \right) \\ &= \frac{1}{\log(1 + \alpha)} (-1)^r (r-1)! \frac{(\alpha+1)^r - 1}{(\alpha+1)^r}. \end{aligned}$$

The proof of Proposition 3.5 comes to an end. □

From this result, the mean, variance, skewness and kurtosis of $Y = \log(X)$ can be expressed and studied.

3.4 Logarithmically weighted moments

As another interesting property, the logarithmically weighted moments of the UPL distribution have simple analytical expressions. They are presented below.

Proposition 3.6. *Let X be a random variable with the UPL distribution and r be a positive integer. Then, the r th logarithmically weighted moment of X is obtained as*

$$m_r^\dagger = \mathbb{E}(X^r \log(X)) = -\frac{\alpha}{\log(1 + \alpha)} \frac{1}{(r + \alpha + 1)(r + 1)}.$$

Proof. By the transfer formula and classical integration techniques, we get

$$\begin{aligned} m_r^+ &= \int_{-\infty}^{+\infty} x^r \log(x) f_\alpha(x) dx = \frac{1}{\log(1+\alpha)} \int_0^1 x^r \log(x) \frac{x^\alpha - 1}{\log(x)} dx = \frac{1}{\log(1+\alpha)} \left(\int_0^1 x^{r+\alpha} dx - \int_0^1 x^r dx \right) \\ &= \frac{1}{\log(1+\alpha)} \left(\frac{1}{r+\alpha+1} - \frac{1}{r+1} \right) = -\frac{\alpha}{\log(1+\alpha)} \frac{1}{(r+\alpha+1)(r+1)}. \end{aligned}$$

The proof of Proposition 3.6 ends. $\square$

To our knowledge, the simple expression of the logarithmically weighted moments remains a rare property for a unit distribution.

Another remark is that, from Proposition 3.6, for any positive integer r , we have

$$\frac{m_{r+1}^+}{m_r^+} = \frac{(r+1)(r+\alpha+1)}{(r+2)(r+\alpha+2)},$$

implying that

$$\alpha = \frac{(1-2-r)(r+1)m_r^+ + (r+2)^2 m_{r+1}^+}{(r+1)m_r^+ - (r+2)m_{r+1}^+}.$$

In particular, by substituting $r = 0$, we have the following simple relation:

$$\alpha = \frac{4m_1^+ - m_0^+}{m_0^+ - 2m_1^+}.$$

Therefore, the parameter α is fully determined by the logarithmically weighted moments. This property can be useful for the estimation of α in a statistical setting via a moment estimation method-like.

4 Application

In this section, we show how the UPL distribution can be used in a statistical context.

4.1 Estimation

The inference of the UPL model can be made by using the maximum likelihood method. The book of Casella and Berger (1990) contains a detailed description of the method as well as its statistical benefits.

Let $x_1, \dots, x_n$ be n independent values from a random variable with the UPL(α) distribution, implying that $x_i \in (0, 1)$ for $i = 1, \dots, n$, among others. It is supposed that α is unknown and must be estimated via $x_1, \dots, x_n$. Then, we estimate α by the maximum likelihood estimate (MLE) $\hat{\alpha}$ defined as

$$\hat{\alpha} = \operatorname{argmax}_{\alpha \in (0, +\infty)} \ell(\alpha),$$

where $\ell(\alpha)$ denotes the log-likelihood function with respect to α . Mathematically, we can write

$$\ell(\alpha) = \sum_{i=1}^n \log[f_\alpha(x_i)] = -n \log[\log(1+\alpha)] - \sum_{i=1}^n \log(-\log(x_i)) + \sum_{i=1}^n \log(1-x_i^\alpha). \quad (8)$$

The estimate $\hat{\alpha}$ also satisfies the following non-linear equation:

$$\frac{\partial}{\partial \alpha} \ell(\hat{\alpha}) = -\frac{n}{(1+\hat{\alpha}) \log(1+\hat{\alpha})} + \sum_{i=1}^n \frac{x_i^{\hat{\alpha}} \log(x_i)}{x_i^{\hat{\alpha}} - 1} = 0.$$

Then, by applying the well-known theory on the maximum likelihood method, for n large enough, the distribution of the random estimator behind $\hat{\alpha}$ can be approximated by the normal distribution with mean α and variance $V = (-\partial^2 \ell(\hat{\alpha})/\partial \alpha^2)^{-1}$, that is

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

The knowledge of this limiting distribution is useful to construct important statistical objects, such as asymptotic confidence intervals and likelihood tests. On the other hand, based on $\hat{\alpha}$, the estimations of the unknown pdf $f_{\alpha}(x)$ and cdf $F_{\alpha}(x)$ are given by $\hat{f}(x) = f_{\hat{\alpha}}(x)$ and $\hat{F}(x) = F_{\hat{\alpha}}(x)$, respectively. These two estimated functions will be investigated below with an example of data analysis.

4.2 Example of data analysis

In order to highlight the interest of the UPL model, we deal with an arbitrary data set containing values of extreme left skewed nature. Thus, we generate 50 values of the random variable $X = 1/(1 + Y)$, where Y follows the Pareto distribution specifies by the following pdf:

$$p_{\eta, \theta}(x) = \theta \eta^{\theta} x^{-(\theta+1)}, \quad x \geq \eta,$$

and $p_{\eta, \theta}(x) = 0$ for $x < \eta$, with $\eta = 0.05$ and $\theta = 0.5$. The obtained data set is presented in Table 2.

Table 2: Values of the considered data set

0.623583091	0.074261315	0.420327650	0.014112652	0.880501445
0.938857354	0.854895021	0.881253230	0.928523376	0.666946194
0.350775729	0.927768965	0.712322580	0.600072612	0.569236629
0.882636474	0.952243330	0.927191343	0.945447466	0.267240836
0.951241606	0.933752113	0.585076574	0.440368246	0.781216314
0.944088730	0.492602556	0.579657631	0.873236156	0.929911951
0.949586793	0.566040824	0.619315200	0.873005115	0.941575655
0.941012073	0.002507848	0.936963292	0.901013344	0.869417092
0.859097389	0.663434523	0.418354290	0.706919488	0.925758938
0.868640916	0.914940971	0.939788880	0.923778525	0.919160226

With these data, based on the methodology described in Subsection 4.1 for the UPL model and the use of the function `nlminb` of the software R, the MLE of α is obtained as $\hat{\alpha} = 9.965334$. Also, with this model,

- the estimated pdf underlying the data is given as

$$\hat{f}(x) = f_{\hat{\alpha}}(x) = \frac{1}{\log(1 + \hat{\alpha})} \frac{x^{\hat{\alpha}} - 1}{\log(x)} = \frac{1}{\log(10.965334)} \frac{x^{9.965334} - 1}{\log(x)}, \quad x \in (0, 1).$$

- based on Proposition 2.5, the estimated cdf underlying the data is given as

$$\begin{aligned} \hat{F}(x) = F_{\hat{\alpha}}(x) &= \frac{1}{\log(1 + \hat{\alpha})} (\text{Ei}[(1 + \hat{\alpha}) \log(x)] - \text{Li}(x)) \\ &= \frac{1}{\log(10.965334)} (\text{Ei}[10.965334 \log(x)] - \text{Li}(x)) \quad x \in (0, 1). \end{aligned}$$

In addition, from $\hat{\alpha}$ and the log-likelihood function defined in Equation (8), we may derive the following well-established criteria: Akaike information criterion (AIC), corrected Akaike information criterion (AICc) and Bayesian information criterion (BIC) defined as

$$\text{AIC} = -2\ell(\hat{\alpha}) + 2k, \quad \text{AICc} = \text{AIC} + \frac{2k(k+1)}{n-k-1}, \quad \text{BIC} = -2\ell(\hat{\alpha}) + k \log(n),$$

respectively, where k denotes the numbers of parameters and n denotes the number of data. Here, we have $k = 1$, $n = 50$ and, after calculus, $\ell(\hat{\alpha}) = 17.01861$. Hence $\text{AIC} = -32.03721$, $\text{AICc} = -31.95388$ and $\text{BIC} = -30.12519$. As model comparison, we consider the Po model with parameter α , with the same estimation strategy. We thus obtain $\hat{\alpha} = 0.8852462$, and, according to the Po model,

- the estimated pdf underlying the data is given as

$$\hat{k}(x) = k_{\hat{\alpha}}(x) = (\hat{\alpha} + 1)x^{\hat{\alpha}} = 1.8852462x^{0.8852462}, \quad x \in (0, 1),$$

- the estimated cdf underlying the data is given as

$$\hat{K}(x) = K_{\hat{\alpha}}(x) = x^{\hat{\alpha}+1} = x^{1.8852462}, \quad x \in (0, 1).$$

We also have $\ell(\hat{\alpha}) = 8.224661$, $k = 1$ and $n = 50$, and we get $\text{AIC} = -14.44932$, $\text{AICc} = -14.36599$ and $\text{BIC} = -12.53730$. Since it has the smallest AIC, AICc and BIC, the UPL model is considered as the best.

Figure 7 plots the curves of $\hat{f}(x)$ in red and $\hat{k}(x)$ in blue over the histogram of the data, and Figure 8 plots the curves of $\hat{F}(x)$ in red and $\hat{K}(x)$ in blue over the empirical cdf of the data.

Figure 7: Plots for the estimated pdfs of the UPL and Po models over the histogram of the data

From Figures 7 and 8, it is evident that the best fit of the data is given by the UPL model; the Po model has missed the concentration of the data in the neighborhood of 1. This illustrates that the UPL model is more suitable for the fit of extreme left skewed data in comparison to the former Po model. This justifies the importance of the UPL model in this regard.

The UPL distribution can be used to define more general models, with different domains, as described in the next section.

Figure 8: Plots for the estimated cdfs of the UPL and Po models over the empirical cdf of the data

5 A note on the UPL-G models

Following the spirit of Cordeiro *et al.* (2020), the UPL distribution can be used to generate a plethora of new distributions by applying the compounding scheme. This direction of work is developed below.

5.1 UPL-G family

Let us consider a parental distribution with cdf $G_{\xi}(x)$ and pdf $g_{\xi}(x)$, where ξ symbolized a vector of parameters. Then, based on Equation (1), we define the corresponding UPL-G distribution by the following pdf:

$$f_{\alpha,\xi}(x) = g_{\xi}(x)f_{\alpha}[G_{\xi}(x)], \quad x \in \mathbb{R}.$$

That is, in an expanded form, we have

$$f_{\alpha,\xi}(x) = \frac{1}{\log(1+\alpha)} g_{\xi}(x) \frac{G_{\xi}(x)^{\alpha} - 1}{\log[G_{\xi}(x)]}, \quad x \in \mathbb{R}.$$

Naturally, the support of the related UPL-G distribution coincides with the one of the parental distribution. To our knowledge, such power-log distributions have not been the object of study in the existing literature. Also, based on Proposition 2.5 and the compounding scheme, the corresponding cdf is obtained as

$$F_{\alpha,\xi}(x) = \frac{1}{\log(1+\alpha)} \left(\text{Ei}\{(1+\alpha) \log[G_{\xi}(x)]\} - \text{Li}[G_{\xi}(x)] \right), \quad x \in \mathbb{R}.$$

Examples of UPL-G distributions are given in Table 3 by their pdfs.

An example of the application of a such UPL-G distribution is presented below.

Table 3: Some UPL-G distributions based on standard parental distributions

UPL-G	Name	Domain	$G_{\xi}(x)$	$g_{\xi}(x)$	(α, ξ)	Corresponding pdf $f_{\alpha, \xi}(x)$
UPLU	Uniform	$(0, \frac{1}{b})$	bx	b	(α, b)	$\frac{b}{\log(1+\alpha)} \frac{b^{\alpha} x^{\alpha} - 1}{\log(b) + \log(x)}$
UPLK	Kum.	$(0, 1)$	$1 - (1 - x^a)^b$	$abx^{a-1}(1 - x^a)^{b-1}$	(α, a, b)	$\frac{ab}{\log(1+\alpha)} x^{a-1}(1 - x^a)^{b-1} \frac{[1 - (1 - x^a)^b]^{\alpha} - 1}{\log[1 - (1 - x^a)^b]}$
UPLE	Exp.	$(0, +\infty)$	$1 - e^{-\theta x}$	$\theta e^{-\theta x}$	(α, θ)	$\frac{\theta}{\log(1+\alpha)} e^{-\theta x} \frac{(1 - e^{-\theta x})^{\alpha} - 1}{\log(1 - e^{-\theta x})}$
UPLW	Weibull	$(0, +\infty)$	$1 - e^{-\theta x^{\lambda}}$	$\theta \lambda x^{\lambda-1} e^{-\theta x^{\lambda}}$	$(\alpha, \lambda, \theta)$	$\frac{\theta \lambda}{\log(1+\alpha)} x^{\lambda-1} e^{-\theta x^{\lambda}} \frac{(1 - e^{-\theta x^{\lambda}})^{\alpha} - 1}{\log(1 - e^{-\theta x^{\lambda}})}$
UPLLom	Lomax	$(0, +\infty)$	$1 - (1 + \rho x)^{-\theta}$	$\rho \theta (1 + \rho x)^{-(\theta+1)}$	(α, ρ, θ)	$\frac{\rho \theta}{\log(1+\alpha)} (1 + \rho x)^{-(\theta+1)} \frac{[1 - (1 + \rho x)^{-\theta}]^{\alpha} - 1}{\log[1 - (1 + \rho x)^{-\theta}]}$
UPLGu	Gumbel	$\mathbb{R}$	$\exp(-e^{-bx})$	$b \exp(-bx - e^{-bx})$	(α, b)	$\frac{b}{\log(1+\alpha)} \exp(-e^{-bx}) [1 - \exp(-ae^{-bx})]$
UPLLog	Logistic	$\mathbb{R}$	$\frac{1}{1 + e^{-bx}}$	$b \frac{e^{-bx}}{(1 + e^{-bx})^2}$	(α, b)	$\frac{b}{\log(1+\alpha)} \frac{e^{-bx}}{(1 + e^{-bx})^2} \frac{1 - (1 + e^{-bx})^{-\alpha}}{\log(1 + e^{-bx})}$
UPLC	Cauchy	$\mathbb{R}$	$\frac{1}{\pi} \arctan(bx) + \frac{1}{2}$	$\frac{b}{\pi} \frac{1}{1 + x^2 b^2}$	(α, b)	$\frac{b}{\pi \log(1+\alpha)} \frac{1}{1 + x^2 b^2} \frac{[\arctan(bx)/\pi + 1/2]^{\alpha} - 1}{\log[\arctan(bx)/\pi + 1/2]}$

5.2 The UPLE model

We now provide an application to the UPLE distribution, or $UPLE(\alpha, \theta)$ distribution, as introduced in Table 3, and defined with the following pdf:

$$f_{\alpha, \theta}(x) = \frac{\theta}{\log(1+\alpha)} e^{-\theta x} \frac{(1 - e^{-\theta x})^{\alpha} - 1}{\log(1 - e^{-\theta x})}, \quad x > 0,$$

and $f_{\alpha, \theta}(x) = 0$ for $x \leq 0$, with $\alpha > 0$ and $\theta > 0$. The corresponding cdf is obtained as

$$F_{\alpha, \theta}(x) = \frac{1}{\log(1+\alpha)} \left(\text{Ei}[(1+\alpha) \log(1 - e^{-\theta x})] - \text{Li}(1 - e^{-\theta x}) \right), \quad x > 0,$$

and $F_{\alpha, \theta}(x) = 0$ for $x \leq 0$. As a result of the power-logarithmic mechanism, the UPLE distribution defines a new two-parameter extension of the exponential distribution. In a typical fit scenario, the parameters α and θ are assumed to be unknown, and we need to estimate them by using data. This parametric estimation can be performed by the maximum likelihood approach as described below. Let $x_1, \dots, x_n$ be n independent values from a random variable X with the $UPLE(\alpha, \theta)$ distribution, representing the observed data. Then, the MLEs $\hat{\alpha}$ and $\hat{\theta}$ are defined by

$$(\hat{\alpha}, \hat{\theta}) = \arg\max_{(\alpha, \theta) \in (0, +\infty)^2} \ell(\alpha, \theta),$$

where $\ell(\alpha, \theta)$ denotes the log-likelihood function with respect to α and θ . Mathematically, $\ell(\alpha, \theta)$ is given by

$$\begin{aligned} \ell(\alpha, \theta) &= \sum_{i=1}^n \log[f_{\alpha, \theta}(x_i)] = -n \log[\log(1+\alpha)] + n \log(\theta) - \theta \sum_{i=1}^n x_i + \sum_{i=1}^n \log[1 - (1 - e^{-\theta x_i})^{\alpha}] \\ &\quad - \sum_{i=1}^n \log[-\log(1 - e^{-\theta x_i})]. \end{aligned}$$

The theory and practice of ensuring the effectiveness of MLEs are well known, and we refer to Casella and Berger (1990) in this regard once more.

5.3 Example of data analysis

We now consider the survival times data set presented in Lee (1992). This data set contains 121 survival times of patients with breast cancer, which are given in Table 4.

Table 4: Values of the survival times data set

0.3	0.3	4.0	5.0	5.6	6.2	6.3	6.6	6.8	7.4	7.5	8.4
8.4	10.3	11.0	11.8	12.2	12.3	13.5	14.4	14.4	14.8	15.5	15.7
16.2	16.3	16.5	16.8	17.2	17.3	17.5	17.9	19.8	20.4	20.9	21.0
21.0	21.1	23.0	23.4	23.6	24.0	24.0	27.9	28.2	29.1	30.0	31.0
31.0	32.0	35.0	35.0	37.0	37.0	37.0	38.0	38.0	38.0	39.0	39.0
40.0	40.0	40.0	41.0	41.0	41.0	42.0	43.0	43.0	43.0	44.0	45.0
45.0	46.0	46.0	47.0	48.0	49.0	51.0	51.0	51.0	52.0	54.0	55.0
56.0	57.0	58.0	59.0	60.0	60.0	60.0	61.0	62.0	65.0	65.0	67.0
67.0	68.0	69.0	78.0	80.0	83.0	88.0	89.0	90.0	93.0	96.0	103.0
105.0	109.0	109.0	111.0	115.0	117.0	125.0	126.0	127.0	129.0	129.0	139.0
154.0											

This data set was also examined in Ramos *et al.* (2013) and Nassar *et al.* (2018), among other places.

Then, based on the UPLE model and the use of the function `nlminb` of the software R, the MLEs of α and θ are obtained as $\hat{\alpha} = 2.22157398$ and $\hat{\theta} = 0.03072415$, respectively. Also,

- the estimated pdf underlying the data is given as

$$\begin{aligned}\hat{f}(x) &= f_{\hat{\alpha}, \hat{\theta}}(x) = \frac{\hat{\theta}}{\log(1 + \hat{\alpha})} e^{-\hat{\theta}x} \frac{(1 - e^{-\hat{\theta}x})^{\hat{\alpha}} - 1}{\log(1 - e^{-\hat{\theta}x})} \\ &= \frac{0.03072415}{\log(3.22157398)} e^{-0.03072415x} \frac{(1 - e^{-0.03072415x})^{2.22157398} - 1}{\log(1 - e^{-0.03072415x})}, \quad x > 0,\end{aligned}$$

- the estimated cdf underlying the data is given as

$$\begin{aligned}\hat{F}(x) &= F_{\hat{\alpha}, \hat{\theta}}(x) = \frac{1}{\log(1 + \hat{\alpha})} \left(\text{Ei}[(1 + \hat{\alpha}) \log(1 - e^{-\hat{\theta}x})] - \text{Li}(1 - e^{-\hat{\theta}x}) \right) \\ &= \frac{1}{\log(3.22157398)} \left(\text{Ei}[(3.22157398) \log(1 - e^{-0.03072415x})] - \text{Li}(1 - e^{-0.03072415x}) \right), \quad x > 0.\end{aligned}$$

As usual, from $\hat{\alpha}$ and $\hat{\theta}$, we may derive the AIC, AICc and BIC defined as

$$\text{AIC} = -2\ell(\hat{\alpha}, \hat{\theta}) + 2k, \quad \text{AICc} = \text{AIC} + \frac{2k(k+1)}{n-k-1}, \quad \text{BIC} = -2\ell(\hat{\alpha}, \hat{\theta}) + k \log(n).$$

Here, we have $k = 2$, $n = 121$ and $\ell(\hat{\alpha}, \hat{\theta}) = -579.2112$, and we obtain $\text{AIC} = 1162.422$, $\text{AICc} = 1162.524$ and $\text{BIC} = 1168.014$.

As model comparison, we consider the PoE model with parameters α and θ defined by the following pdf:

$$k_{\alpha, \beta}(x) = (\alpha + 1)\theta e^{-\theta x} (1 - e^{-\theta x})^{\alpha}, \quad x > 0,$$

and $k_{\alpha, \beta}(x) = 0$ for $x \leq 0$. The associated cdf is

$$K_{\alpha, \beta}(x) = (1 - e^{-\theta x})^{\alpha+1}, \quad x > 0,$$

and $K_{\alpha, \beta}(x) = 0$ for $x \leq 0$. More details on the PoE distribution, also known as exponentiated exponential distribution, can be found in Gupta and Kundu (2001).

By using the maximum likelihood estimation method, we get $\hat{\alpha} = 0.51679438$ and $\hat{\theta} = 0.02774019$, from which we derive

- the estimated pdf given by

$$\begin{aligned}\hat{k}(x) &= k_{\hat{\alpha}, \hat{\theta}}(x) = (\hat{\alpha} + 1)\hat{\theta}e^{-\hat{\theta}x}(1 - e^{-\hat{\theta}x})^{\hat{\alpha}} \\ &= 0.04207616e^{-0.02774019x}(1 - e^{-0.02774019x})^{0.51679438}, \quad x > 0,\end{aligned}$$

- the estimated cdf obtained as

$$\hat{K}(x) = K_{\hat{\alpha}, \hat{\theta}}(x) = (1 - e^{-\hat{\theta}x})^{\hat{\alpha}+1} = (1 - e^{-0.02774019x})^{1.51679438}, \quad x > 0.$$

Also, after calculus, we have $\ell(\hat{\alpha}, \hat{\theta}) = -580.0936$, and we get $\text{AIC} = 1164.187$, $\text{AICc} = 1164.289$ and $\text{BIC} = 1169.779$. Since it has the smallest AIC, AICc and BIC, the UPLE model is the best. Also, one can remark that it has the smallest AIC, AICc and BIC in comparison to those of the three-parameter models considered in Section 7 of Ramos *et al.* (2013) (see (Ramos *et al.*, 2013, three last columns of Table 1)).

Figure 9 plots the curves of $\hat{f}(x)$ in red and $\hat{k}(x)$ in blue over the histogram of the data. Furthermore, the curves of $\hat{F}(x)$ in red and $\hat{K}(x)$ in blue are plotted over the empirical ecdf of the data in Figure 10.

In view of Figures 9 and 10, the best fit is attributed to the UPLE model, showing that the UPL-G scheme can outperform some classical estimation schemes.

Figure 9: Plots for the estimated pdfs of the UPLE and PoE models over the histogram of the survival times data

Figure 10: Plots for the estimated cdfs of the UPLE and PoE models over the empirical cdf of the survival times data

6 Conclusion

In this article, we have focused on a new unit distribution called the unit power-logarithmic (UPL) distribution, which is defined by an original power-log pdf. Several of its important characteristics, such as likelihood ratio order results, an extreme asymmetry to the left, closed-form expressions for the cdf, hrf, ordinary moments, skewness, kurtosis, moments generating function, incomplete moments, logarithmic moments and logarithmically weighted moments have been determined using analytical and graphical tools. On several important characteristics, we compared the UPL with the so-called power distributions, demonstrating some non-trivial relationships. A data-handling application is provided. Then, with an application to a well-known survival data set, a more general modelling strategy based on the UPL distribution is discussed.

With an extensive theoretical study and some statistical basics, this article lays the groundwork for the UPL distribution. Some future directions of research include: (i) an in-depth analysis of the hrf, (ii) additional stochastic order properties, such as mean residual life order, dispersive order, and right-spread order, and (ii) the construction of new regression models (linear or quantile) or bivariate extensions for multivariate analysis.

Acknowledgement: The author wishes to express his gratitude to the anonymous reviewers for their valuable comments, corrections and suggestions.

Conflict of interest: The author states no conflict of interest.

References

- Aitchison, J. (1982), "The statistical analysis of compositional data", *Journal of the Royal Statistical Society Series B (Methodological)*, 44(2): 139-177.
- Altun, E. (2020), "The log-weighted exponential regression model: alternative to the beta regression model", *Communications in Statistics-Theory and Methods*, DOI: 10.1080/03610926.2019.1664586.
- Altun, E. and Hamedani, G. G. (2018), "The log-xgamma distribution with inference and application", *Journal de la Société Française de Statistique*, 159(3): 40-55.
- Bakouch, H. S., Nik, A. S., Asgharzadeh, A. and Salinas, H. S. (2021), "A flexible probability model for proportion data: Unit-half-normal distribution", *Communications in Statistics: Case Studies, Data Analysis and Applications*, DOI:10.1080/23737484.2021.1882355
- Bantan, R. A. R., Chesneau, C., Jamal, F., Elgarhy, M., Tahir, M. H., Aqib, A., Zubair, M. and Anam, S. (2020), "Some new facts about the unit-Rayleigh distribution with applications", *Mathematics*, 8(11): 1-23.
- Butler, R. J. and McDonald, J. B. (1989), "Using incomplete moments to measure inequality", *Journal of Econometrics*, 42(1): 109-119.
- Casella, G. and Berger, R. L. (1990), "Statistical Inference", Brooks/Cole Publishing Company: Bel Air, CA, USA.
- Consul, P. C. and Jain, G. C. (1971), "On the log-gamma distribution and its properties", *Statistical Papers*, 12: 100-106.
- Cordeiro, G. M., Silva, R. B. and Nascimento, A. D. C. (2020), "Recent Advances in Lifetime and Reliability Models", Bentham Sciences Publishers, Sharjah, UAE.
- Ghitany, M. E., Mazucheli, J., Menezes, A. F. B. and Alqallaf, F. (2019), "The unit-inverse Gaussian distribution: A new alternative to two-parameter distributions on the unit interval", *Communications in Statistics-Theory and Methods*, 48(14): 3423-3438.
- Gómez-Déniz, E., Sordo, M. A. and Calderín-Ojeda, E. (2014), "The log-Lindley distribution as an alternative to the beta regression model with applications in insurance", *Insurance: Mathematics and Economics*, 54: 49-57.
- Gradshteyn, I. S. and Ryzhik, I. M. (2007), "Table of Integrals, Series, and Products, Seventh Edition", Edition Jeffrey, A. & Zwillinger D. Academic Press, Burlington, MA.
- Gündüz, S. and Korkmaz, M. Ç. (2020), "A new unit distribution based on the unbounded Johnson distribution rule: The unit Johnson SU distribution", *Pakistan Journal of Statistics and Operation Research*, 16(3): 471-490.
- Gupta, R. D. and Kundu, D. (2001), "Exponentiated-exponential family: an alternative to gamma and Weibull distributions", *Biometrical Journal*, 43(1): 117-130.
- Haq, M. A., Hashmi, S., Aidi, K., Ramos, P. L. and Louzada, F. (2020), "Unit modified Burr-III distribution: Estimation, characterizations and validation test", *Annals of Data Science*, to appear. <https://doi.org/10.1007/s40745-020-00298-6>
- Klein, J. P. and Moeschberger, M. L. (1997), "Survival Analysis: Techniques for Censored and Truncated Data", Springer-Verlag, New York.
- Korkmaz, M. Ç. (2020), "The unit generalized half normal distribution: A new bounded distribution with inference and application", *University Politehnica of Bucharest Scientific Bulletin-Series A-Applied Mathematics and Physics*, 82(2): 133-140.
- Korkmaz, M. Ç., Chesneau, C. and Korkmaz, Z. S. (2021), "On the arcsecant hyperbolic normal distribution. Properties, quantile regression modeling and applications", *Symmetry*, 13(1): 117, 1-24.
- Korkmaz, M. Ç. and Chesneau, C. (2021), "On the unit Burr-XII distribution with the quantile regression modeling and applications", *Computational and Applied Mathematics*, 40, Article number: 29: 1-26.
- Korkmaz, M. Ç., Chesneau, C. and Korkmaz, Z. S. (2021), "Transmuted unit Rayleigh quantile regression model: alternative to beta and Kumaraswamy quantile regression models", *University Politehnica of Bucharest Scientific Bulletin-Series A-Applied Mathematics and Physics*, to appear.
- Lee, E. T. (1992), "Statistical Methods for Survival Data Analysis", John Wiley, New York.
- Mazucheli, J., Menezes, A. F. and Dey, S. (2018), "The unit-Birnbaum-Saunders distribution with applications", *Chilean Journal of Statistics* 9(1): 47-57.
- Mazucheli, J., Menezes, A. F. and Dey, S. (2019), "Unit-Gompertz distribution with applications", *Statistica*, 79(1): 25-43.
- Mazucheli, J., Menezes, A. F. B. and Ghitany, M. E. (2018), "The unit-Weibull distribution and associated inference", *Journal of Applied Probability and Statistics*, 13(2): 1-22.
- Mazucheli, J., Menezes, A. F. B., Fernandes, L. B., de Oliveira, R. P. and Ghitany, M. E. (2020), "The unit-Weibull distribution as an alternative to the Kumaraswamy distribution for the modeling of quantiles conditional on covariates", *Journal of Applied Statistics*, 47(6): 954-974.
- Nassar, M., Alzaatreh, A., Abo-Kasem, O., Mead, M. and Mansoor, M. (2018), "A new family of generalized distributions based on alpha power transformation with application to cancer data", *Annals of Data Science*, 5: 421-436.
- R Core Team (2014), "R: A language and environment for statistical computing", R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Ramos, M. W. A., Cordeiro, G. M., Marinho, P. R. D., Dias, C. R. B. and Hamedani, G. G. (2013), "The Zografos-Balakrishnan log-logistic distribution: Properties and applications", *Journal of Statistical Theory and Applications*, 12(3): 225-244.
- Shaked, M. and Shanthikumar, J. G. (2007), "Stochastic Orders", Wiley, New York, NY, USA.