NOTE ON THE PAPER "THE DOUBLE PARETO-LOGNORMAL DISTRIBUTION—A NEW PARAMETRIC MODEL FOR SIZE DISTRIBUTIONS" AND ITS CORRECTION

NEVEN GRBAC AND TIHANA GALINAC GRBAC

Abstract. The double Pareto-lognormal distribution, and the closely related normal-Laplace distribution, are probability distributions with a wide range of applications, which were introduced in the paper [Reed, W. J., & Jorgensen, M. (2004). The double Pareto-lognormal distribution—a new parametric model for size distributions. *Communications in Statistics – Theory and Methods*, *33*(8), 1733–1753] mentioned in the title. The purpose of this paper is to put an end to the confusion regarding the correctness of the formulas for the probability density function and the cumulative distribution function of the double Pareto-lognormal distribution and the normal-Laplace distribution in *loc. cit.*, in view of the correction published in the paper [Amini, Z., & Rabbani, H. (2017). Letter to the editor: Correction to "The Normal-Laplace distribution and its relatives". *Communications in Statistics – Theory and Methods*, *46*(4), 2076–2078]. It is shown that the formulas in the original paper and its correction are in fact equal, so that both are correct, except for a minor typographical error in the formulas for the cumulative distribution functions in the original paper. The source of confusion is the incorrect formula for the Mills ratio in terms of the complementary error function in the correction paper, which makes the impression that the formulas are different.

Date: February 22, 2022.

Key words and phrases. Double Pareto-lognormal distribution, normal-Laplace distribution, probability density function, cumulative distribution function, correction.

1. Introduction

The double Pareto-lognormal distribution and the normal-Laplace distribution are the probability distributions introduced in the paper by Reed and Jorgensen (2004), mentioned in the title of this paper. These distributions proved to be very useful and were already applied in a wide range of applications. It is also important in applications that both probability distributions are naturally generated, so that the parameters of the generative model may be estimated from the context and thus the distribution parameters predicted in advance.

In particular, the double Pareto-lognormal distribution appears as the best choice for modelling the distribution of various phenomena, often related with the size growth. It is very flexible in describing the power-law behavior in both tails, and at the same time the lognormal body of the distribution. The transition between different regimes is smooth, which is one of the reasons for its wide applicability. Some of the applications in which the double Pareto-lognormal distribution exhibits excellent fit to the empirical data are the distribution of earnings and income (Akhundjanov & Toda, 2020; Gabaix, 2009; Reed, 2003; Reed & Jorgensen, 2004), stock price return (Gabaix, 2009; Reed & Jorgensen, 2004), consumption size (Toda, 2017), degree in financial networks (Sun & Chan-Lau, 2017), human settlements size (Gabaix, 1999, 2009; Giesen, Zimmermann, & Suedekum, 2010; González-Val, Ramos, Sanz-Gracia, & Vera-Cabello, 2015; Reed, 2001, 2002; Reed & Jorgensen, 2004), particle size (Reed & Jorgensen, 2004), oil-field size (Reed & Jorgensen, 2004), size of web sites and computer files (Mitzenmacher, 2003; Reed & Jorgensen, 2004), latency in mobile edge computing (Volos, Bando, & Konishi, 2018), node degree of the web browsing session graph (Brown & Doran, 2018), call frequency and duration of mobile phone users (Seshadri et al., 2008), initial mass of stars (Zaninetti, 2017), magnetic field curvature in space plasma (Bandyopadhyay et al., 2020), frequency of oligonucleotides in a genomic sequence (Csűrös, Noé, & Kucherov, 2007), phenotypic variation in single-gene knockouts (Graham, Robb, & Poe, 2012), frequency of protein, metabolite and messenger ribonucleic acids (mRNA) in protein synthesis (Chikashige et al., 2015; Lu & King, 2009), fatigue life of rope wire under different impact loads (Zhiqian & Xun, 2017).

In the recent preprint by Galinac Grbac, Huljenić, and Grbac (2022), we have applied the double Pareto-lognormal distribution to the distribution of software faults among software modules in software systems. It was a great surprise to realize that there is a paper by Amini and Rabbani (2017) with the correction of the formulas by Reed and Jorgensen (2004) for the probability density function and the cumulative distribution function of the normal-Laplace distribution, which would imply that the formulas for the double Pareto-lognormal distribution are incorrect as well. It seemed very unlikely that the distributions with so many applications would contain an error. This is the motivation for digging deeper and writing the present paper.

Careful reading of the original paper (Reed & Jorgensen, 2004), and its correction (Amini & Rabbani, 2017), revealed that they are both correct. In fact, the formulas for the probability density function and the cumulative distribution function of the normal-Laplace distribution of Reed and Jorgensen (2004) and of Amini and Rabbani (2017) are equal. In the original paper they are expressed in terms of the Mills ratio and the probability density function of the standard normal distribution. In the latter the same formulas are expressed in terms of the error function and the complementary error function. However, as we point out in the paper, it is the matter of elementary transformation to show that the two expressions are equal, except for the typographical error in the formula for the cumulative distribution function of Reed and Jorgensen (2004). The error in question is just an incorrect sign. The authors of the cited paper are certainly aware of this minor imprecision, as the correct sign is given in the formula for the cumulative distribution function of the normal-Laplace distribution by Reed (2006).

The source of the confusion regarding the correctness of the probability density function and the cumulative distribution function of the normal-Laplace distribution of Reed and Jorgensen (2004) lies in the incorrect expression for the Mills ratio in terms of the complementary error function, which was used by Amini and Rabbani (2017) to compare the formulas. This error makes the impression that the formulas of Reed and Jorgensen (2004) and of Amini and Rabbani (2017) are different, which is not true, except for the sign mentioned above.

Since the formulas for the probability density function and the cumulative distribution function of the double Pareto-lognormal distribution of Reed and Jorgensen (2004) are derived from those for the normal-Laplace distribution, they are also correct, except for the sign in the formula for the cumulative distribution function. Although the double Pareto-lognormal distribution is not mentioned by Amini and Rabbani (2017), we include the formulas here to avoid any further confusion.

At the end of this introduction, we sketch out the content of the paper. It begins with Section 2 in which the notation is introduced and basic facts recalled. In particular, the inconsistencies in notation of Reed and Jorgensen (2004) and Amini and Rabbani (2017) are pointed out for convenience of the reader. Section 3 is devoted to comparison of the formulas for the probability density function of the normal-Laplace distribution of Reed and Jorgensen (2004) and of Amini and Rabbani (2017). Similarly, the two formulas for the cumulative distribution function of the normal-Laplace distribution are compared in Section 4, in which the formula is carefully derived from the probability density function in order to capture the correct sign. Finally, in Section 5, the formulas for the probability density function and the cumulative distribution function of the double Pareto-lognormal distribution are derived. Although all calculations are elementary, we tend to include more details than usual, in order to put a definitive end to the confusion regarding the correctness of considered formulas.

2. Notation and preliminaries

The notation in the correction by Amini and Rabbani (2017) and the original paper by Reed and Jorgensen (2004) are not fully consistent. Hence, we start by carefully fixing the notation, following Reed and Jorgensen (2004), and pointing out the differences with Amini and Rabbani (2017).

- $N(\nu, \tau^2)$ denotes the normal distribution with mean $\nu \in \mathbb{R}$ and variance τ^2, where $\tau > 0$ is the standard deviation. Amini and Rabbani (2017) denote the same distribution by $N(\nu, \tau)$.
- $\phi(z)$ is the probability density function (pdf) of the standard normal distribution $N(0, 1)$ given by

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}. \tag{1}$$

- $\Phi(z)$ is the cumulative distribution function (cdf) of the standard normal distribution given by

$$\begin{aligned} \Phi(z) &= \int_{-\infty}^{z} \phi(t)\, dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2\, dt}. \end{aligned} \tag{2}$$

- $\Phi^c(z)$ is the complementary cdf of the standard normal distribution given by

$$\begin{aligned} \Phi^c(z) &= \int_{z}^{+\infty} \phi(t)\, dt \\ &= 1 - \Phi(z). \end{aligned} \tag{3}$$

Since $\phi(t)$ is an even function, the change of variables $s = -t$ in the defining integral of Φ^c implies that

$$\Phi^c(z) = \Phi(-z). \tag{4}$$

- $R(z)$ is the Mills ratio of the standard normal distribution given by

$$R(z) = \frac{\Phi^c(z)}{\phi(z)}. \tag{5}$$

- $\mathrm{erf}(z)$ is the Gauss error function given by

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\, dt, \tag{6}$$

This function is used only in the correction by Amini and Rabbani (2017). It is related to $\Phi(z)$ by the equations

$$\Phi(z) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{z}{\sqrt{2}}\right)\right], \tag{7}$$

$$\mathrm{erf}(z) = 2\Phi\left(z\sqrt{2}\right) - 1. \tag{8}$$

- $\mathrm{erfc}(z)$ is the complementary error function given as

$$\mathrm{erfc}(z) = 1 - \mathrm{erf}(z). \tag{9}$$

It is used only by Amini and Rabbani (2017), and related to the complementary cdf $\Phi^c(z)$ by

$$\Phi^c(z) = \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{z}{\sqrt{2}}\right)\right] = \frac{1}{2}\,\mathrm{erfc}\left(\frac{z}{\sqrt{2}}\right), \tag{10}$$

$$\mathrm{erfc}(z) = 2\Phi^c\left(z\sqrt{2}\right). \tag{11}$$

Thus erfc is also related to the Mills ratio as

$$
\begin{aligned}
R(z) &= \frac{\frac{1}{2}\operatorname{erfc}\left(\frac{z}{\sqrt{2}}\right)}{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}} \\
&= \frac{\sqrt{2\pi}}{2}e^{\frac{1}{2}z^2}\operatorname{erfc}\left(\frac{z}{\sqrt{2}}\right).
\end{aligned} \tag{12}
$$

At this point, there is an error in the correction by Amini and Rabbani (2017), because the Mills ratio is incorrectly expressed in terms of erfc. This is the main source of confusion regarding the correctness of the formulas by Reed and Jorgensen (2004).

- $f_W(w)$ is the pdf of the asymmetric Laplace distribution given by the formula

$$
f_W(w) = \begin{cases} \frac{\alpha\beta}{\alpha+\beta}e^{\beta w}, & \text{for } w \leq 0, \\ \frac{\alpha\beta}{\alpha+\beta}e^{-\alpha w}, & \text{for } w > 0, \end{cases} \tag{13}
$$

  with parameters $\alpha > 0$ and $\beta > 0$.

- $NL(\alpha, \beta, \nu, \tau^2)$ denotes the normal-Laplace distribution, with parameters $\alpha > 0$, $\beta > 0$, $\nu \in \mathbb{R}$ and $\tau > 0$, introduced by Reed and Jorgensen (2004). Amini and Rabbani (2017) denote the same distribution by $NL(\alpha, \beta, \nu, \tau)$.

- $g(y)$ denotes the pdf of the normal-Laplace distribution $NL(\alpha, \beta, \nu, \tau^2)$. Amini and Rabbani (2017) denote the same function by $f(y)$, but we stick to the original notation to avoid confusion with the double Pareto-lognormal distribution.

- $G(y)$ denotes the cdf of the normal-Laplace distribution $NL(\alpha, \beta, \nu, \tau^2)$. Amini and Rabbani (2017) denote the same function by $F(y)$, but we stick to the original notation for reasons mentioned above.

- $dPlN(\alpha, \beta, \nu, \tau^2)$ denotes the double Pareto-lognormal distribution, with parameters $\alpha > 0$, $\beta > 0$, $\nu \in \mathbb{R}$ and $\tau > 0$, introduced by Reed and Jorgensen (2004).

- $f(x)$, where $x > 0$, denotes the pdf of the double Pareto-lognormal distribution $dPlN(\alpha, \beta, \nu, \tau^2)$.

- $F(x)$, where $x > 0$, denotes the cdf of the double Pareto-lognormal distribution $dPlN(\alpha, \beta, \nu, \tau^2)$.

In the sequel, we show that the formula for the pdf $g(y)$ of the normal-Laplace distribution of Reed and Jorgensen (2004), expressed in terms of the Mills ratio and the pdf of the standard normal distribution, is equal to the formula for $g(y)$ obtained by Amini and Rabbani (2017) in terms of the complementary error function. The confusion arose because the Mills ratio is incorrectly expressed in terms of the complementary error function by Amini and Rabbani (2017). Hence, there is no error in the formula for $g(y)$ of Reed and Jorgensen (2004). The same then holds for the pdf $f(x)$ of the double Pareto-lognormal distribution.

Regarding the cdf $G(y)$ of the normal-Laplace distribution and the cdf $F(x)$ of the double Pareto-lognormal distribution, the formulas of Reed and Jorgensen (2004) are also essentially correct, except for the typographical error, which is the incorrect sign in the second term, as explained below.

3. Calculation of the normal-Laplace pdf

In this section we show that the formulas for the pdf $g(y)$ of the normal-Laplace distribution of Reed and Jorgensen (2004) and of Amini and Rabbani (2017) are in fact equal.

The normal-Laplace distribution is introduced by Reed and Jorgensen (2004) as the probability distribution of the random variable $\widehat{Y}$ given as the final state of an ordinary Brownian motion after an exponentially distributed time with the initial state following the normal distribution. However, it is shown in the Appendix of Reed (2003), that the distribution of the random variable $\widehat{Y}$ can be decomposed as the sum of two independent random variables Z and W, where Z follows the normal distribution $N(\nu, \tau^2)$ and W

the asymmetric Laplace distribution with parameters α and β, defined by its pdf $f_W(w)$ in equation (13). The parameters are determined by the coefficients in the stochastic differential equation of the Brownian motion and its initial condition.

It is well-known that the pdf of the sum of two independent random variables is given by the convolution of their densities. In the notation as above, we have

$$
\begin{aligned}
g(y) &= \phi\left(\frac{y-\nu}{\tau}\right) * f_W(y) \\
&= \int_{-\infty}^{+\infty} \phi\left(\frac{t-\nu}{\tau}\right) f_W(y-t)\, dt \\
&= \int_{-\infty}^{+\infty} \phi\left(\frac{y-t-\nu}{\tau}\right) f_W(t)\, dt
\end{aligned}
\tag{14}
$$

This convolution is expressed in equation (5) of Reed and Jorgensen (2004) in terms of the Mills ratio and the pdf of the standard normal distribution as

$$
g(y) = \frac{\alpha\beta}{\alpha+\beta}\phi\left(\frac{y-\nu}{\tau}\right)\left[R\left(\alpha\tau - \frac{y-\nu}{\tau}\right) + R\left(\beta\tau + \frac{y-\nu}{\tau}\right)\right]. \tag{15}
$$

The same convolution is calculated in the Appendix of Amini and Rabbani (2017) and expressed in equation (3) of *loc. cit.* in terms of the complementary error function as

$$
\begin{aligned}
&g(y) = \\
&\quad \frac{\alpha\beta}{2(\alpha+\beta)}\left[e^{\frac{1}{2}\alpha(-2y+2\nu+\alpha\tau^2)}\operatorname{erfc}\left(\frac{\alpha\tau}{\sqrt{2}} - \frac{y-\nu}{\tau\sqrt{2}}\right) + e^{\frac{1}{2}\beta(2y-2\nu+\beta\tau^2)}\operatorname{erfc}\left(\frac{\beta\tau}{\sqrt{2}} + \frac{y-\nu}{\tau\sqrt{2}}\right)\right].
\end{aligned}
\tag{16}
$$

The calculation is correct and hence we omit it here. The goal is to prove that the two expressions for $g(y)$ are equal. However, we begin with a simple lemma exhibiting an identity of exponential functions.

Lemma 3.1. *The following identity*

$$
e^{\frac{1}{2}\theta(2\xi+\theta\tau^2)} = \frac{\phi\left(\frac{\xi}{\tau}\right)}{\phi\left(\theta\tau + \frac{\xi}{\tau}\right)}
$$

holds for any $\theta, \xi \in \mathbb{R}$ and $\tau > 0$.

Proof. Rewriting the exponent of the exponential function on the left-hand side gives

$$\begin{aligned}\theta(2\xi+\theta\tau^2) &= 2\theta\xi+\theta^2\tau^2\\ &= \frac{\xi^2}{\tau^2}+2\theta\xi+\theta^2\tau^2-\frac{\xi^2}{\tau^2}\\ &= \left(\theta\tau+\frac{\xi}{\tau}\right)^2-\left(\frac{\xi}{\tau}\right)^2.\end{aligned}$$

Hence, the exponential function can be expressed in terms of the pdf of the standard normal distribution as

$$\begin{aligned}e^{\frac{1}{2}\theta(2\xi+\theta\tau^2)} &= e^{\frac{1}{2}\left[\left(\theta\tau+\frac{\xi}{\tau}\right)^2-\left(\frac{\xi}{\tau}\right)^2\right]}\\ &= \frac{e^{-\frac{1}{2}\left(\frac{\xi}{\tau}\right)^2}}{e^{-\frac{1}{2}\left(\theta\tau+\frac{\xi}{\tau}\right)^2}}\\ &= \frac{\phi\left(\frac{\xi}{\tau}\right)}{\phi\left(\theta\tau+\frac{\xi}{\tau}\right)},\end{aligned}$$

as claimed. □

Proposition 3.2. *Equation* (15), *which is equation* (5) *of Reed and Jorgensen (2004), and equation* (16), *which is equation* (3) *of Amini and Rabbani (2017), expressing the pdf* $g(y)$ *of the normal-Laplace distribution, are equal.*

Proof. In equation (11) the complementary error function erfc is expressed in terms of the complementary cdf Φ^c of the standard normal distribution as

$$\operatorname{erfc}(z) = 2\Phi^c\left(z\sqrt{2}\right).$$

Inserting this in equation (16) for the pdf of the normal-Laplace distribution of Amini and Rabbani (2017), we obtain

$$g(y) = \frac{\alpha\beta}{\alpha+\beta}\left[e^{\frac{1}{2}\alpha(-2y+2\nu+\alpha\tau^2)}\Phi^c\left(\alpha\tau-\frac{y-\nu}{\tau}\right)+e^{\frac{1}{2}\beta(2y-2\nu+\beta\tau^2)}\Phi^c\left(\beta\tau+\frac{y-\nu}{\tau}\right)\right]. \tag{17}$$

Applying Lemma 3.1 with $\theta = \alpha$ and $\xi = -(y-\nu)$ and using the fact that the pdf ϕ of the standard normal distribution is an even function, the first exponential function can be written as

$$e^{\frac{1}{2}\alpha(-2y+2\nu+\alpha\tau^2)} = \frac{\phi\left(\frac{y-\nu}{\tau}\right)}{\phi\left(\alpha\tau - \frac{y-\nu}{\tau}\right)}.$$

For the second exponential function, Lemma 3.1 with $\theta = \beta$ and $\xi = y - \nu$ is applied to obtain

$$e^{\frac{1}{2}\beta(2y-2\nu+\beta\tau^2)} = \frac{\phi\left(\frac{y-\nu}{\tau}\right)}{\phi\left(\beta\tau + \frac{y-\nu}{\tau}\right)}.$$

Inserting these expressions for the exponential functions into $g(y)$, we obtain

$$\begin{aligned} g(y) &= \frac{\alpha\beta}{\alpha+\beta}\left[\frac{\phi\left(\frac{y-\nu}{\tau}\right)}{\phi\left(\alpha\tau - \frac{y-\nu}{\tau}\right)}\Phi^c\left(\alpha\tau - \frac{y-\nu}{\tau}\right) + \frac{\phi\left(\frac{y-\nu}{\tau}\right)}{\phi\left(\beta\tau + \frac{y-\nu}{\tau}\right)}\Phi^c\left(\beta\tau + \frac{y-\nu}{\tau}\right)\right] \\ &= \frac{\alpha\beta}{\alpha+\beta}\phi\left(\frac{y-\nu}{\tau}\right)\left[\frac{\Phi^c\left(\alpha\tau - \frac{y-\nu}{\tau}\right)}{\phi\left(\alpha\tau - \frac{y-\nu}{\tau}\right)} + \frac{\Phi^c\left(\beta\tau + \frac{y-\nu}{\tau}\right)}{\phi\left(\beta\tau + \frac{y-\nu}{\tau}\right)}\right] \\ &= \frac{\alpha\beta}{\alpha+\beta}\phi\left(\frac{y-\nu}{\tau}\right)\left[R\left(\alpha\tau - \frac{y-\nu}{\tau}\right) + R\left(\beta\tau + \frac{y-\nu}{\tau}\right)\right], \end{aligned}$$

which is exactly equation (15) for the pdf of the normal-Laplace distribution of Reed and Jorgensen (2004), as claimed. □

4. Calculation of the normal-Laplace cdf

Since the formulas for the pdf $g(y)$ of the normal-Laplace distribution of Reed and Jorgensen (2004) and of Amini and Rabbani (2017) are equal by Proposition 3.2, it is now clear that the formulas for the cdf $G(y)$ should also be equal. Equation (4) for $G(y)$ of Amini and Rabbani (2017) can be rewritten in terms of the Mills ratio to obtain equation (15) for $G(y)$ of Reed and Jorgensen (2004). However, we include here the calculation of the cdf to show that Reed and Jorgensen (2004) have a typographical error in the formula for $G(y)$. The sign in the numerator of the second term in incorrect. The corrected formula is given in Proposition 4.2 below. It coincides with the correct formula for $G(y)$ given in equation (1) of Reed (2006).

We begin the calculation of the cdf $G(y)$ of the normal-Laplace distribution with a technical lemma required to compute the integral of the pdf $g(y)$.

Lemma 4.1. *Let*

$$I_\theta(y) = \int_{-\infty}^{y} e^{\theta t}\Phi\left(\theta\tau + \frac{t-\nu}{\tau}\right) dt,$$

where $\theta, \nu \in \mathbb{R}$, $\theta \neq 0$, and $\tau > 0$. Then, the following identity

$$I_\theta(y) = \frac{e^{\theta y}}{\theta}\Phi\left(\theta\tau + \frac{y-\nu}{\tau}\right) - \frac{e^{\theta\nu - \frac{\theta^2\tau^2}{2}}}{\theta}\Phi\left(\frac{y-\nu}{\tau}\right)$$

holds for any $\theta, \nu \in \mathbb{R}$, $\theta \neq 0$, and $\tau > 0$.

Proof. This is an elementary calculation. The first step is to make the change of variables

$$s = \frac{t-\nu}{\tau}$$

in the integral $I_\theta(y)$. One obtains

$$I_\theta(y) = \tau e^{\theta\nu}\int_{-\infty}^{\frac{y-\nu}{\tau}} e^{\theta\tau s}\Phi\left(\theta\tau + s\right) ds.$$

Integration by parts of the latter integral, with

$$u = \Phi\left(\theta\tau + s\right),$$

$$dv = e^{\theta\tau s}\, ds,$$

gives

$$I_\theta(y) = \frac{e^{\theta\nu}}{\theta}\left[e^{\theta\tau s}\Phi\left(\theta\tau + s\right)\Bigg|_{-\infty}^{\frac{y-\nu}{\tau}} - \int_{-\infty}^{\frac{y-\nu}{\tau}} e^{\theta\tau s}\phi\left(\theta\tau + s\right) ds \right],$$

where we used the fact that $\Phi'(z) = \phi(z)$ holds by definition.

Observe that the first summand tends to zero as $s \to -\infty$, because $\Phi(\theta\tau + s)$ decays to zero as fast as e^{s^2}, while the exponential function $e^{\theta\tau s}$ grows to infinity only as fast as

e^s in the case of $\theta > 0$ and decays to zero in the case of $\theta < 0$. In the second summand, we use the formula (1) for $\phi(\theta\tau + s)$ and the definition (2) of the cdf $\Phi(z)$. Thus, the integral $I_\theta(y)$ becomes

$$
\begin{aligned}
I_\theta(y) &= \frac{e^{\theta\nu}}{\theta}\left[e^{\theta(y-\nu)}\Phi\left(\theta\tau + \frac{y-\nu}{\tau}\right) - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{y-\nu}{\tau}} e^{\theta\tau s}e^{-\frac{1}{2}(\theta\tau+s)^2}\,ds\right] \\
&= \frac{e^{\theta y}}{\theta}\Phi\left(\theta\tau + \frac{y-\nu}{\tau}\right) - \frac{e^{\theta\nu}e^{-\frac{\theta^2\tau^2}{2}}}{\theta\sqrt{2\pi}}\int_{-\infty}^{\frac{y-\nu}{\tau}} e^{-\frac{1}{2}s^2}\,ds \\
&= \frac{e^{\theta y}}{\theta}\Phi\left(\theta\tau + \frac{y-\nu}{\tau}\right) - \frac{e^{\theta\nu-\frac{\theta^2\tau^2}{2}}}{\theta}\Phi\left(\frac{y-\nu}{\tau}\right),
\end{aligned}
$$

as required. □

Proposition 4.2. *The cdf G(y) of the normal-Laplace distribution is given by the formula*

$$
G(y) = \Phi\left(\frac{y-\nu}{\tau}\right) - \phi\left(\frac{y-\nu}{\tau}\right)\frac{\beta R\left(\alpha\tau - \frac{y-\nu}{\tau}\right) - \alpha R\left(\beta\tau + \frac{y-\nu}{\tau}\right)}{\alpha+\beta},
$$

which is the same as equation (15) *of Reed and Jorgensen (2004), except for the sign in the numerator of the second term.*

Proof. By definition the cdf $G(y)$ is given as the integral

$$
G(y) = \int_{-\infty}^{y} g(t)\,dt,
$$

where $g(t)$ is the pdf of the normal-Laplace distribution. The formula (17) for $g(y)$ in terms of the exponential function and the complementary cdf of the standard normal distribution is the most convenient for the calculation of the integral. We have

$$
\begin{aligned}
&G(y) = \qquad (18) \\
&\frac{\alpha\beta}{\alpha+\beta}\left[e^{\alpha\nu+\frac{\alpha^2\tau^2}{2}}\int_{-\infty}^{y} e^{-\alpha t}\Phi^c\left(\alpha\tau - \frac{t-\nu}{\tau}\right)dt + e^{-\beta\nu+\frac{\beta^2\tau^2}{2}}\int_{-\infty}^{y} e^{\beta t}\Phi^c\left(\beta\tau + \frac{t-\nu}{\tau}\right)dt\right],
\end{aligned}
$$

and we must compute the two integrals. The idea is to reduce them to the integral $I_\theta(y)$ calculated in Lemma 4.1.

Let I and J denote the first and the second integral, respectively. Using the property (4) of the cdf $\Phi(z)$, and applying Lemma 4.1 with $\theta = -\alpha$, we rewrite I as

$$
\begin{aligned}
I &= \int_{-\infty}^{y} e^{-\alpha t}\Phi^c\left(\alpha\tau - \frac{t-\nu}{\tau}\right) dt \\
&= \int_{-\infty}^{y} e^{-\alpha t}\Phi\left(-\alpha\tau + \frac{t-\nu}{\tau}\right) dt \\
&= I_{-\alpha}(y) \\
&= -\frac{e^{-\alpha y}}{\alpha}\Phi\left(-\alpha\tau + \frac{y-\nu}{\tau}\right) + \frac{e^{-\alpha\nu-\frac{\alpha^2\tau^2}{2}}}{\alpha}\Phi\left(\frac{y-\nu}{\tau}\right) \\
&= \frac{e^{-\alpha\nu-\frac{\alpha^2\tau^2}{2}}}{\alpha}\Phi\left(\frac{y-\nu}{\tau}\right) - \frac{e^{-\alpha y}}{\alpha}\Phi^c\left(\alpha\tau - \frac{y-\nu}{\tau}\right),
\end{aligned}
$$

and using the definition (3) of the complementary ccdf $\Phi^c(z)$, and applying again Lemma 4.1 with $\theta = \beta$, we rewrite J as

$$
\begin{aligned}
J &= \int_{-\infty}^{y} e^{\beta t}\Phi^c\left(\beta\tau + \frac{t-\nu}{\tau}\right) dt \\
&= \int_{-\infty}^{y} e^{\beta t}\left[1 - \Phi\left(\beta\tau + \frac{t-\nu}{\tau}\right)\right] dt \\
&= \int_{-\infty}^{y} e^{\beta t}\, dt - \int_{-\infty}^{y} e^{\beta t}\Phi\left(\beta\tau + \frac{t-\nu}{\tau}\right) dt \\
&= \left.\frac{e^{\beta t}}{\beta}\right|_{-\infty}^{y} - I_{\beta}(y) \\
&= \frac{e^{\beta y}}{\beta} - \left[\frac{e^{\beta y}}{\beta}\Phi\left(\beta\tau + \frac{y-\nu}{\tau}\right) - \frac{e^{\beta\nu-\frac{\beta^2\tau^2}{2}}}{\beta}\Phi\left(\frac{y-\nu}{\tau}\right)\right] \\
&= \frac{e^{\beta\nu-\frac{\beta^2\tau^2}{2}}}{\beta}\Phi\left(\frac{y-\nu}{\tau}\right) + \frac{e^{\beta y}}{\beta}\left[1 - \Phi\left(\beta\tau + \frac{y-\nu}{\tau}\right)\right] \\
&= \frac{e^{\beta\nu-\frac{\beta^2\tau^2}{2}}}{\beta}\Phi\left(\frac{y-\nu}{\tau}\right) + \frac{e^{\beta y}}{\beta}\Phi^c\left(\beta\tau + \frac{y-\nu}{\tau}\right).
\end{aligned}
$$

Inserting the expressions for I and J in equation (18) for $G(y)$, we obtain

$$\begin{aligned}
G(y) &= \frac{\alpha\beta}{\alpha+\beta}\left\{e^{\alpha\nu+\frac{\alpha^2\tau^2}{2}}\left[\frac{e^{-\alpha\nu-\frac{\alpha^2\tau^2}{2}}}{\alpha}\Phi\left(\frac{y-\nu}{\tau}\right)-\frac{e^{-\alpha y}}{\alpha}\Phi^c\left(\alpha\tau-\frac{y-\nu}{\tau}\right)\right]\right.\\
&\qquad\left.+\,e^{-\beta\nu+\frac{\beta^2\tau^2}{2}}\left[\frac{e^{\beta\nu-\frac{\beta^2\tau^2}{2}}}{\beta}\Phi\left(\frac{y-\nu}{\tau}\right)+\frac{e^{\beta y}}{\beta}\Phi^c\left(\beta\tau+\frac{y-\nu}{\tau}\right)\right]\right\}\\
&= \frac{\beta}{\alpha+\beta}\Phi\left(\frac{y-\nu}{\tau}\right)-\frac{\beta}{\alpha+\beta}e^{\frac{1}{2}\alpha(-2y+2\nu+\alpha\tau^2)}\Phi^c\left(\alpha\tau-\frac{y-\nu}{\tau}\right)\\
&\qquad+\frac{\alpha}{\alpha+\beta}\Phi\left(\frac{y-\nu}{\tau}\right)+\frac{\alpha}{\alpha+\beta}e^{\frac{1}{2}\beta(2y-2\nu+\beta\tau^2)}\Phi^c\left(\beta\tau+\frac{y-\nu}{\tau}\right)\\
&= \Phi\left(\frac{y-\nu}{\tau}\right)-\frac{\beta e^{\frac{1}{2}\alpha(-2y+2\nu+\alpha\tau^2)}\Phi^c\left(\alpha\tau-\frac{y-\nu}{\tau}\right)-\alpha e^{\frac{1}{2}\beta(2y-2\nu+\beta\tau^2)}\Phi^c\left(\beta\tau+\frac{y-\nu}{\tau}\right)}{\alpha+\beta}.
\end{aligned} \tag{19}$$

The exponential functions in the numerator can be expressed in terms of the pdf $\phi(z)$ of the standard normal distribution using Lemma 3.1 in the same way as in the proof of Proposition 3.2. Thus,

$$\begin{aligned}
G(y) &= \Phi\left(\frac{y-\nu}{\tau}\right)-\frac{\beta\frac{\phi\left(\frac{y-\nu}{\tau}\right)}{\phi\left(\alpha\tau-\frac{y-\nu}{\tau}\right)}\Phi^c\left(\alpha\tau-\frac{y-\nu}{\tau}\right)-\alpha\frac{\phi\left(\frac{y-\nu}{\tau}\right)}{\phi\left(\beta\tau+\frac{y-\nu}{\tau}\right)}\Phi^c\left(\beta\tau+\frac{y-\nu}{\tau}\right)}{\alpha+\beta}\\
&= \Phi\left(\frac{y-\nu}{\tau}\right)-\phi\left(\frac{y-\nu}{\tau}\right)\frac{\beta R\left(\alpha\tau-\frac{y-\nu}{\tau}\right)-\alpha R\left(\beta\tau+\frac{y-\nu}{\tau}\right)}{\alpha+\beta},
\end{aligned}$$

as claimed. $\square$

In equation (4) of Amini and Rabbani (2017), the cdf $G(y)$ of the normal-Laplace distribution is expressed in terms of the error function erf and the complementary error function erfc as follows

$$\begin{aligned}
G(y) &= \frac{1}{2(\alpha+\beta)}\left[\alpha+\beta-2\beta e^{\frac{1}{2}\alpha(-2y+2\nu+\alpha\tau^2)}+(\alpha+\beta)\operatorname{erf}\left(\frac{y-\nu}{\tau\sqrt{2}}\right)\right.\\
&\quad\left.+\,\beta e^{\frac{1}{2}\alpha(-2y+2\nu+\alpha\tau^2)}\operatorname{erfc}\left(-\frac{\alpha\tau}{\sqrt{2}}+\frac{y-\nu}{\tau\sqrt{2}}\right)+\alpha e^{\frac{1}{2}\beta(2y-2\nu+\beta\tau^2)}\operatorname{erfc}\left(\frac{\beta\tau}{\sqrt{2}}+\frac{y-\nu}{\tau\sqrt{2}}\right)\right].
\end{aligned} \tag{20}$$

We now show that this formula is equal to the one obtained in Proposition 4.2.

Proposition 4.3. *The formula in Proposition 4.2, which is equation* (15) *of Reed and Jorgensen (2004) up to the sign in the numerator of the second term, and equation* (20)*, which is equation* (4) *of Amini and Rabbani (2017), expressing the cdf $G(y)$ of the normal-Laplace distribution, are equal.*

Proof. We use the relation of $\mathrm{erf}(z)$ and $\mathrm{erfc}(z)$ to the cdf $\Phi(z)$ and complementary cdf $\Phi^c(z)$ of the standard normal distribution given in equations (8) and (11). Associating together the first two summands and the fourth one, and putting on the same denominator the remaining summands, equation (20) can be written as

$$
\begin{aligned}
&G(y) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{y-\nu}{\tau\sqrt{2}}\right)\right] \\
&- \frac{2\beta e^{\frac{1}{2}\alpha(-2y+2\nu+\alpha\tau^2)} - \beta e^{\frac{1}{2}\alpha(-2y+2\nu+\alpha\tau^2)}\,\mathrm{erfc}\left(-\frac{\alpha\tau}{\sqrt{2}} + \frac{y-\nu}{\tau\sqrt{2}}\right) - \alpha e^{\frac{1}{2}\beta(2y-2\nu+\beta\tau^2)}\,\mathrm{erfc}\left(\frac{\beta\tau}{\sqrt{2}} + \frac{y-\nu}{\tau\sqrt{2}}\right)}{2(\alpha+\beta)} \\
&= \Phi\left(\frac{y-\nu}{\tau}\right) \\
&- \frac{\beta e^{\frac{1}{2}\alpha(-2y+2\nu+\alpha\tau^2)} - \beta e^{\frac{1}{2}\alpha(-2y+2\nu+\alpha\tau^2)}\Phi^c\left(-\alpha\tau + \frac{y-\nu}{\tau}\right) - \alpha e^{\frac{1}{2}\beta(2y-2\nu+\beta\tau^2)}\Phi^c\left(\beta\tau + \frac{y-\nu}{\tau}\right)}{\alpha+\beta}.
\end{aligned}
$$

Using equation (4) and definition (3), the first two summands in the numerator can be transformed into

$$
\begin{aligned}
\beta e^{\frac{1}{2}\alpha(-2y+2\nu+\alpha\tau^2)} &- \beta e^{\frac{1}{2}\alpha(-2y+2\nu+\alpha\tau^2)}\Phi^c\left(-\alpha\tau + \frac{y-\nu}{\tau}\right) \\
&= \beta e^{\frac{1}{2}\alpha(-2y+2\nu+\alpha\tau^2)}\left[1 - \Phi\left(\alpha\tau - \frac{y-\nu}{\tau}\right)\right] \\
&= \beta e^{\frac{1}{2}\alpha(-2y+2\nu+\alpha\tau^2)}\Phi^c\left(\alpha\tau - \frac{y-\nu}{\tau}\right),
\end{aligned}
$$

so that the formula for $G(y)$ becomes

$$
G(y) = \Phi\left(\frac{y-\nu}{\tau}\right) - \frac{\beta e^{\frac{1}{2}\alpha(-2y+2\nu+\alpha\tau^2)}\Phi^c\left(\alpha\tau - \frac{y-\nu}{\tau}\right) - \alpha e^{\frac{1}{2}\beta(2y-2\nu+\beta\tau^2)}\Phi^c\left(\beta\tau + \frac{y-\nu}{\tau}\right)}{\alpha+\beta},
$$

which is the same formula as in the proof of Proposition 4.2. It remains to apply Lemma 3.1 to the exponential functions in the numerator to get the formula for $G(y)$ given in Proposition 4.2. Thus, the two formulas are equal, as claimed. $\square$

5. The pdf and cdf of the double Pareto-lognormal distribution

For completeness and to avoid any further confusion, we derive here the formulas for the pdf $f(x)$ and cdf $F(x)$ of the double Pareto-lognormal distribution. These are not mentioned in the correction by Amini and Rabbani (2017), but as they are determined by the pdf $g(y)$ and cdf $G(y)$ of the normal-Laplace distribution, we decided to include them. As in the case of the normal-Laplace distribution, equation (8) for the pdf and equation (23) for the cdf of the double Pareto-lognormal distribution of Reed and Jorgensen (2004) are correct, except for the sign in the square-brackets of (23) which is corrected below.

The double Pareto-lognormal distribution is introduced by Reed and Jorgensen (2004) as the probability distribution of the random variable $\widehat{X}$, given as the final state of the geometric Brownian motion after an exponentially distributed time, with the initial state following the lognormal distribution. As explained by Reed and Jorgensen (2004), the random variable $\widehat{X}$ is related to the random variable $\widehat{Y}$ defined in Section 3 as

$$\widehat{Y} = \log \widehat{X}.$$

Hence, the random variable $\widehat{X}$ is obtained as the function of the random variable $\widehat{Y}$, so that the pdf $f(x)$ of the double Pareto-lognormal distribution is given in terms of the pdf $g(y)$ of the normal-Laplace distribution by

$$\begin{aligned} f(x) &= g(y) \left| \frac{dy}{dx} \right| \qquad (21) \\ &= \frac{1}{x} g(\log x), \end{aligned}$$

where $y = \log x$.

The formula for $g(y)$, which is the most convenient for the calculation of $f(x)$, is given by equation (17) in the proof of Proposition 4.2. Inserting $y = \log x$ in the exponential functions gives

$$e^{\frac{1}{2}\alpha(-2y+2\nu+\alpha\tau^2)} = e^{-\alpha\log x}e^{\alpha\nu+\frac{\alpha^2\tau^2}{2}}$$

$$= A(\alpha,\nu,\tau)x^{-\alpha},$$

$$e^{\frac{1}{2}\beta(2y-2\nu+\beta\tau^2)} = e^{\beta\log x}e^{-\beta\nu+\frac{\beta^2\tau^2}{2}}$$

$$= A(-\beta,\nu,\tau)x^{\beta},$$

where

$$A(\theta,\nu,\tau) = e^{\theta\nu+\frac{\theta^2\tau^2}{2}}.$$

Note that there is a minor typesetting error in equation (9) of Reed and Jorgensen (2004) defining $A(\theta,\nu,\tau)$, in which α should be replaced by θ. Then, inserting these in equation (21), with $g(y)$ expressed as in equation (17), and using equation (4), the pdf $f(x)$ equals

$$\begin{aligned} f(x) &= \frac{\alpha\beta}{\alpha+\beta}\left[A(\alpha,\nu,\tau)x^{-\alpha-1}\Phi^c\left(\alpha\tau - \frac{\log x-\nu}{\tau}\right)\right. \\ &\quad \left. + A(-\beta,\nu,\tau)x^{\beta-1}\Phi^c\left(\beta\tau + \frac{\log x-\nu}{\tau}\right)\right] \\ &= \frac{\alpha\beta}{\alpha+\beta}\left[A(\alpha,\nu,\tau)x^{-\alpha-1}\Phi\left(\frac{\log x-\nu-\alpha\tau^2}{\tau}\right)\right. \\ &\quad \left. + A(-\beta,\nu,\tau)x^{\beta-1}\Phi^c\left(\frac{\log x-\nu+\beta\tau^2}{\tau}\right)\right], \end{aligned} \tag{22}$$

which is equal to equation (8) of Reed and Jorgensen (2004).

Instead of integrating $f(x)$, we use the relation between $\widehat{X}$ and $\widehat{Y}$ to determine the cdf $F(x)$ of the double Pareto-lognormal distribution. It is given by

$$F(x) = G(\log x).$$

Note a minor typographical error just above equation (23) of Reed and Jorgensen (2004), where e^x should be replaced with $\log x$ as here. The formula for $G(y)$, which is the most convenient for computing $F(x)$, is given by equation (19) in the proof of Proposition 4.2. Inserting $y = \log x$ in that equation, using expressions for exponential functions obtained above, and equation (4), we obtain

$$
\begin{aligned}
&F(x) = \\
&\quad \Phi\left(\frac{\log x - \nu}{\tau}\right) - \frac{\beta A(\alpha,\nu,\tau)x^{-\alpha}\Phi^c\left(\alpha\tau - \frac{\log x - \nu}{\tau}\right) - \alpha A(-\beta,\nu,\tau)x^{\beta}\Phi^c\left(\beta\tau + \frac{\log x - \nu}{\tau}\right)}{\alpha+\beta} \\
&\quad = \Phi\left(\frac{\log x - \nu}{\tau}\right) - \frac{1}{\alpha+\beta}\left[\beta A(\alpha,\nu,\tau)x^{-\alpha}\Phi\left(\frac{\log x - \nu - \alpha\tau^2}{\tau}\right)\right. \\
&\qquad\qquad \left. - \alpha A(-\beta,\nu,\tau)x^{\beta}\Phi^c\left(\frac{\log x - \nu + \beta\tau^2}{\tau}\right)\right],
\end{aligned}
\tag{23}
$$

which is equal to equation (23) of Reed and Jorgensen (2004), except for the sign in the square-brackets.

Acknowledgments

This work is supported by the Croatian Science Foundation under the project HRZZ-IP-2019-04-4216. The second named author acknowledges the support of the Croatian Science Foundation under the project HRZZ-IP-2018-01-3628.

Disclosure statement

The authors report there are no competing interests to declare.

References

Akhundjanov, S. B., & Toda, A. A. (2020). Is Gibrat's 'Economic Inequality' lognormal? *Empirical Economics*, *59*, 2071–2091.

Amini, Z., & Rabbani, H. (2017). Letter to the editor: Correction to 'The normal-Laplace distribution and its relatives'. *Communications in Statistics - Theory and Methods*, *46*(4), 2076-2078.

Bandyopadhyay, R., Yang, Y., Matthaeus, W. H., Chasapis, A., Parashar, T. N., Russell, C. T., . . . Burch, J. L. (2020). In situ measurement of curvature of magnetic field in turbulent space plasmas: A statistical study. *The Astrophysical Journal*, *893*(1), Art. no. L25 (9pp).

Brown, K., & Doran, D. (2018). Contrasting web robot and human behaviors with network models. *Journal of Communications*, *13*(8), 473-481.

Chikashige, Y., Arakawa, S., Leibnitz, K., Tsutsumi, C., Mori, C., Osakada, H., . . . Hiraoka, Y. (2015). Cellular economy in fission yeast cells continuously cultured with limited nitrogen resources. *Scientific Reports*, *5*, Art. no. 15617 (14pp).

Csűrös, M., Noé, L., & Kucherov, G. (2007). Reconsidering the significance of genomic word frequencies. *Trends in Genetics*, *23*(11), 543-546.

Gabaix, X. (1999). Zipf's law for cities: An explanation. *The Quarterly Journal of Economics*, *114*(3), 739-767.

Gabaix, X. (2009). Power laws in economics and finance. *Annual Review of Economics*, *1*(1), 255-294.

Galinac Grbac, T., Huljenić, D., & Grbac, N. (2022). The Reed–Jorgensen double Pareto-lognormal distribution as the probability distribution of faults in software systems. Preprint.

Giesen, K., Zimmermann, A., & Suedekum, J. (2010). The size distribution across all cities – double Pareto lognormal strikes. *Journal of Urban Economics*, *68*(2), 129-137.

González-Val, R., Ramos, A., Sanz-Gracia, F., & Vera-Cabello, M. (2015). Size distributions for all cities: Which one is best? *Papers in Regional Science*, *94*(1), 177-196.

Graham, J. H., Robb, D. T., & Poe, A. R. (2012). Random phenotypic variation of yeast (*Saccharomyces cerevisiae*) single-gene knockouts fits a double Pareto-lognormal distribution. *PLoS ONE*, *7*(11), Art. no. e48964 (6pp).

Lu, C., & King, R. D. (2009). An investigation into the population abundance distribution of mRNAs, proteins, and metabolites in biological systems. *Bioinformatics*, *25*(16), 2020-2027.

Mitzenmacher, M. (2003). Dynamic models for file sizes and double Pareto distributions. *Internet Mathematics*, *1*(3), 305 – 333.

Reed, W. J. (2001). The Pareto, Zipf and other power laws. *Economics Letters*, *74*(1), 15-19.

Reed, W. J. (2002). On the rank-size distribution for human settlements. *Journal of Regional Science*, *42*(1), 1-17.

Reed, W. J. (2003). The Pareto law of incomes–an explanation and an extension. *Physica A: Statistical Mechanics and its Applications*, *319*, 469–486.

Reed, W. J. (2006). The normal-Laplace distribution and its relatives. In N. Balakrishnan, J. M. Sarabia, & E. Castillo (Eds.), *Advances in distribution theory, order statistics, and inference* (pp. 61–74). Boston: Birkhäuser.

Reed, W. J., & Jorgensen, M. (2004). The double Pareto-lognormal distribution–a new parametric model for size distributions. *Communications in Statistics – Theory and Methods*, *33*(8), 1733–1753.

Seshadri, M., Machiraju, S., Sridharan, A., Bolot, J., Faloutsos, C., & Leskove, J. (2008). Mobile call graphs: Beyond power-law and lognormal distributions. In *Proceedings* *of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (p. 596–604). New York, NY, USA: Association for Computing Machinery.

Sun, A. J., & Chan-Lau, J. A. (2017). Financial networks and interconnectedness in an advanced emerging market economy. *Quantitative Finance*, *17*(12), 1833-1858.

Toda, A. A. (2017). A note on the size distribution of consumption: more double Pareto than lognormal. *Macroeconomic Dynamics*, *21*(6), 1508–1518.

Volos, H., Bando, T., & Konishi, K. (2018). Latency modeling for mobile edge computing using LTE measurements. In *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)* (p. 1-5).

Zaninetti, L. (2017). A left and right truncated lognormal distribution for the stars. *Advances in Astrophysics*, *2*(3), 197-213.

Zhiqian, R., & Xun, C. (2017). Research on fatigue life of steel wire ropes under impact loads based on double Pareto lognormal distribution. *Advances in Mechanical Engineering*, *9*(8), 1-7.

Neven Grbac, Juraj Dobrila University of Pula, Faculty of Engineering, Zagrebačka 30, HR-52100 Pula, Croatia

Email address: `neven.grbac@unipu.hr`

Tihana Galinac Grbac, Juraj Dobrila University of Pula, Faculty of Engineering, Zagrebačka 30, HR-52100 Pula, Croatia

Email address: `tihana.galinac@unipu.hr`