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

NEVEN GRBAC [1] AND TIHANA GALINAC GRBAC [1]

Dear Editor,

the double Pareto-lognormal distribution, and the closely related normal-Laplace distribution, are probability distributions with a wide range of applications, which were introduced by Reed and Jorgensen (2004) in the paper "The double Pareto-lognormal distribution—a new parametric model for size distributions", mentioned in the title, and published in vol. 33, no. 8 of this journal, pp. 1733–1753. The purpose of this letter 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 "Letter to the editor: Correction to 'The Normal-Laplace distribution and its relatives' ", by Amini and Rabbani (2017), published in vol. 46, no. 4 of this journal, pp. 2076–2078.

In our recent preprint (Galinac Grbac, Huljenić, & 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 letter by Amini and Rabbani (2017) with the correction of the formulas by Reed and Jorgensen (2004). It seemed very unlikely that the distributions with so many applications would contain an error. This is the motivation for digging deeper and writing this letter.

Date: June 8, 2022.

[1] Juraj Dobrila University of Pula, Zagrebačka 30, HR-52100 Pula, Croatia.

E-mail: neven.grbac@unipu.hr tihana.galinac@unipu.hr.

We explain in this letter 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 letter, which makes the impression that the formulas are different. Although Amini and Rabbani (2017) only discuss the normal-Laplace distribution, we include here, for the sake of completeness, the correct formulas for the closely related double Pareto-lognormal distribution.

We first introduce the required notation. Let $\phi(z)$, $\Phi(z)$ and $\Phi^c(z)$ denote the pdf, the cdf and the complementary cdf, respectively, of the standard normal distribution. The Mills ratio $R(z)$ of the standard normal distribution is defined as $R(z) = \frac{\Phi^c(z)}{\phi(z)}$. Let $\operatorname{erf}(z)$ and $\operatorname{erfc}(z)$ denote the Gauß error function and the complementary error function, respectively. The correct formula for $R(z)$ in terms of erfc is

$$R(z) = \frac{\sqrt{2\pi}}{2} e^{\frac{1}{2}z^2} \operatorname{erfc}\left(\frac{z}{\sqrt{2}}\right).$$

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 source of confusion regarding the correctness of the formulas by Reed and Jorgensen (2004).

The formulas for the pdf $g(y)$ and the cdf $G(y)$ of the normal-Laplace distribution $NL(\alpha, \beta, \nu, \tau^2)$, with parameters $\alpha > 0$, $\beta > 0$, $\nu \in \mathbb{R}$ and $\tau > 0$, are given in terms of the Mills ratio, the pdf and the cdf of the standard normal distribution as

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

These are the same as equation (5) and equation (15) of Reed and Jorgensen (2004), except that in equation (15) the sign in the numerator of the second term is incorrect. Note that the correct formula for $G(y)$ is also given in equation (1) of Reed (2006).

The same formulas are expressed in equations (3) and (4) of Amini and Rabbani (2017) in terms of erf and erfc as

$$g(y) = \frac{\alpha\beta}{2(\alpha+\beta)}\left[e^{\frac{1}{2}\alpha(-2y+2\nu+\alpha\tau^2)}\operatorname{erfc}\left(\frac{\alpha\tau}{\sqrt{2}}-\frac{y-\nu}{\tau\sqrt{2}}\right)+e^{\frac{1}{2}\beta(2y-2\nu+\beta\tau^2)}\operatorname{erfc}\left(\frac{\beta\tau}{\sqrt{2}}+\frac{y-\nu}{\tau\sqrt{2}}\right)\right],$$

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

These formulas are exactly equal to the ones by Reed and Jorgensen (2004), once the correct formula for the Mills ratio in terms of erfc is used. The incorrect formula for the Mills ratio in the correction letter makes the false impression that the formulas are different.

Finally, we provide the formulas for the pdf $f(x)$ and the cdf $F(x)$ of the double Pareto-lognormal distribution $dPlN(\alpha,\beta,\nu,\tau^2)$, with parameters $\alpha>0$, $\beta>0$, $\nu\in\mathbb{R}$ and $\tau>0$, in terms of the cdf and the complementary cdf of the standard normal distribution as

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

where $A(\theta, \nu, \tau) = e^{\theta\nu + \frac{\theta^2\tau^2}{2}}$. These are the same as equation (8) and (23) of Reed and Jorgensen (2004), except for the incorrect sign in the square-brackets in equation (23).

In hope that this letter will end the confusion regarding the formulas of Reed and Jorgensen (2004), and be helpful to scholars applying these distributions in their studies, we would be glad to provide a detailed account of the discussed issue upon an e-mail request, as it is too lengthy and elementary to occupy the valuable pages of this journal.

Sincerely Yours,

Neven Grbac and Tihana Galinac Grbac

Acknowledgments

The work discussed in this letter is supported by the Croatian Science Foundation under the project HRZZ-IP-2019-04-4216.

References

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.

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.

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.