International Journal of Informatics and Applied Mathematics
e-ISSN:2667-6990 Vol. 4, No. 2, 43-52

Pareto Randomization of the Scaling Parameter for the Gaussian Distribution

Miguel Felgueiras[1,2], Rui Santos[1,2], and Joao Paulo Martins[2,3]

[1] ESTG, Polytechnic Institute of Leiria, Portugal
[2] CEAUL, Faculdade de Ciências, Universidade de Lisboa, Portugal
[3] Escola Superior de Saúde, P.Porto, Porto, Portugal
mfelg@ipleiria.pt; rui.santos@ipleiria.pt; jom@ess.ipp.pt

Abstract. Gaussian distribution is a common choice when dealing with symmetric data. However, other alternatives must be considered in applications with high tail-weight. One option is the randomization of the scale parameter for the Gaussian distribution, enabling a more flexible model for the tails albeit maintaining symmetry. Although any positive random variable can be used as a random scale parameter, Pareto distribution is a suitable choice in order to increase variance and tail-weight. Therefore, the aim of this work is to study the Pareto randomization of the scale parameter for symmetric distributions, in particular for the Gaussian distribution. Estimation problem is tackled and a simulation study is discussed. Finally, an application concerning the directions chosen by ants after a stimulus is provided. The results reveal that the proposed methodology works well both on simulated and real data.

Keywords: Scale Parameter Randomization · Heavy-tailed Distributions · Parameters Estimation · Gaussian Distribution.

1 Introduction

Common distribution functions F usually have a real location parameter μ and a positive scale parameter σ, that is, $F(x|\mu,\sigma)$. In order to obtain a larger variability, useful when analysing some phenomenons [9], it might be advantageous to randomize the scale parameter. Some practical applications concern Log-returns analysis for financial data [13], natural phenomena like the eggs laid by some species or the direction followed by animals after some stimulus [2, 7, 8], or the energy released by the greater earthquakes [4, 6, 12].
A scale mixture can be seen as the distribution of the random variable (r.v.) $Y = \Theta X$, where X and Θ are independent and absolutely continuous r.v., with support $S_X = \mathbb{R}$ and $S_\Theta =]0,\infty[$. In fact, any positive distribution is a possible choice for Θ. However, to ensure that Y has an higher variance than X, which is our goal, we will restrict the study to the cases where $S_\Theta =]1,\infty[$.
The Pareto distribution with density function

$$f_\Theta(\theta) = \alpha\theta^{-\alpha-1}, \quad \theta \geq 1, \quad \alpha > 0, \tag{1}$$

is a suitable option for Θ distribution for several reasons. First, it fulfils the restriction $S_\Theta =]1,\infty[$. Second, Pareto's distribution is heavy tail-weighted (note that the k-th moment only exists when $k < \alpha$) implying that the r.v. Y has a higher kurtosis than the r.v. X. Third, Pareto's distribution has a simple analytical form, leading to straightforward calculus of Y densities. As we will see later, in several situations it is possible to achieve a closed form expression for the Y distribution. The scale mixture can also be seen as a slash distribution, since

$$Y = \Theta X = \frac{X}{\Theta^{-1}}$$

where $\Theta^{-1} \sim Beta(\alpha, 1)$. When $\alpha = 1$ we get the standard slash distribution since $\Theta^{-1} \sim U(0,1)$. Also, note that $\Theta^{-1} \overset{d}{=} U^{1/\alpha}$ where $U \sim U(0,1)$. This duality was already explored in [5, 13]. There are many examples concerning the application of slash distributions in real datasets, that can also be seen as scale mixtures. For instance, a recent work [1] uses a slash logistic distribution when modelling glass fiber strength and the nickel concentration in soils. The α parameter works as a tuning parameter since when $\alpha < 1$ the mixture has heavier tail-weight than the slash distribution, and when $\alpha > 1$ we have the reverse situation. The probability density function of the scale mixture $Y = \Theta X$ can be written as

$$f_Y(y) = \int_1^\infty \alpha\theta^{-\alpha-2} f_X\left(\tfrac{y}{\theta}\right) d\theta, \tag{2}$$

originating explicit densities based in the incomplete gamma function

$$\gamma(a, y) = \int_0^y t^{a-1} e^{-t} dt. \tag{3}$$

For some distributions of the r.v. X, the mixture densities are presented in Table 1 (cf. [5]).

Table 1. Some Pareto scale mixtures densities

Distribution	Density	Mixture density
$X \sim Gaussian\,(0,1)$	$f_X\,(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$	$f_Y\,(y) = \frac{\alpha 2^{0.5\alpha-1}\gamma\left(\frac{\alpha+1}{2}, \frac{y^2}{2}\right)}{\sqrt{\pi}\,\lvert y\rvert^{\alpha+1}}, \quad y \neq 0$
$\frac{2^{-\frac{3+\beta}{2}} \exp\left[-0.5\,\lvert x\rvert^{\frac{2}{1+\beta}}\right]}{\Gamma\left(\frac{3+\beta}{2}\right)}, \quad -1 < \beta \leq 1$		$f_Y\,(y) = \frac{\alpha(1+\beta)\gamma\left(\frac{\beta+1}{2}(\alpha+1), 0.5\lvert y\rvert^{\frac{2}{1+\beta}}\right)}{2^{-\alpha\frac{\beta+1}{2}} 4\Gamma\left(\frac{3+\beta}{2}\right)\lvert y\rvert^{\alpha+1}}, \; y \neq 0$
$X \sim Cauchy(0,1)$	$f_X\,(x) = \frac{1}{\pi}\frac{1}{1+x^2}$	$f_Y\,(y) = \frac{\alpha y^{-\alpha-1}}{\pi} \int_0^y \frac{z^\alpha}{1+z^2} dz, \quad y \neq 0$
$X \sim Gama(\beta,1)$	$f_X\,(x) = \frac{1}{\Gamma(\beta)} x^{\beta-1} e^{-x}$	$f_Y\,(y) = \frac{\alpha y^{-\alpha-1}}{\Gamma(\beta)} \gamma\,(\alpha+\beta, y)\,, \quad y > 0$
$X \sim Beta(p,q)$	$f_X\,(x) = \frac{(1-x)^{q-1}}{x^{1-p} B(p,q)}$	$f_Y\,(y) = \begin{cases} \frac{\alpha B\,(p+\alpha, q, y)}{y^{\alpha+1} B(p,q)}, & 0 < y < 1 \\ \frac{\alpha B\,(p+\alpha, q)}{y^{\alpha+1} B(p,q)}, & y \geq 1 \end{cases}$
$X \sim Weibull\,(\beta,1)$	$f_X\,(x) = \beta x^{\beta-1} e^{-x^\beta}$	$f_Y\,(y) = \frac{\alpha\gamma\left(\alpha\beta^{-1}+1, y^\beta\right)}{y^{\alpha+1}}, \quad y > 0$
$X \sim Pareto\,(\beta)$	$f_X\,(x) = \beta x^{-\beta-1}$	$f_Y\,(y) = \begin{cases} \alpha^2 y^{-\alpha-1} \ln y, & \alpha = \beta, \quad y > 0 \\ \frac{\alpha\beta\left(y^{-\alpha-1} - y^{-\beta-1}\right)}{\beta-\alpha}, & \alpha \neq \beta, \quad y > 0 \end{cases}$

2 Moments and Tail-weight

Mixture moments and other properties can be found in [5]. A key result states that mixture moments can only exist when $k < \alpha$. Hence, the necessary condition for the existence of Pareto's moments is also required for the existence of the mixture's moments, even if the mean value of X, $E(X)$, is equal to 0. The most used tail-weight measure, the kurtosis, denoted as β_2, is given by

$$\beta_2\,(Y) = \beta_2\,(X)\,\frac{(\alpha-2)^2}{\alpha\,(\alpha-4)}, \quad \alpha > 4. \tag{4}$$

Assuming that $E(X) = 0$ (otherwise a location transformation can be performed) and that $\beta_2(X)$ exists, then $\beta_2(Y) > \beta_2(X)$. Thus, the mixture can be used to increase the tail-weight of the original X distribution.

3 Symmetric Mixtures

The main focus of this work is symmetric mixtures (specially Gaussian mixtures), that is, the ones generated by a symmetric r.v. X. Let us consider, without loss of generalization, that $E(X) = 0$ (otherwise perform a location shift). In this situation it is clear that the mixture skewness coefficient, if exists, will be $\beta_1(Y) = 0$. From now on, we will assume that we are dealing with a mixture that can be written as

$$W = \mu + \sigma Y = \mu + \sigma \Theta X, \tag{5}$$

where μ and σ are respectively the location and the scale parameters of the mixture, $\Theta \sim Pareto(\alpha)$ and X is an absolutely continuous random variable, symmetric with $E(X) = 0$.

3.1 Gaussian Mixtures

When $X \sim Gaussian(0, 1)$ the mixture defined in (5) is infinitely divisible [11]. Moreover, if $\mu = 0$ and $\sigma = 1$ in equation (5), the mixture density is (cf. Table 1)

$$f_W(w) = \alpha 2^{0.5\alpha-1} |w|^{-\alpha-1} \pi^{-0.5} \gamma\left(\frac{\alpha+1}{2}, \frac{w^2}{2}\right), \quad w \neq 0. \tag{6}$$

For instance, when $\alpha = 1$ we get

$$f_W(w) = \frac{1 - e^{-\frac{w^2}{2}}}{\sqrt{2\pi} w^2}, \quad w \neq 0, \tag{7}$$

and when $\alpha = 3$

$$f_W(w) = \frac{3\left(2 - \left(2 + w^2\right) e^{-w^2/2}\right)}{\sqrt{2\pi} w^4}, \quad w \neq 0. \tag{8}$$

Figure 1 presents some possible densities for the mixture defined in equation (5), for $X \sim Gaussian(0, 1)$, $\mu = 0$, $\sigma = 10$ and $\alpha \in \{.25, .5, .75, 1, 1.5, 2\}$. Thus, the mixture density is closer to the Gaussian density as α increases, and more far apart as α decreases. It is evident that α parameter works in a similar way as the n parameter in a $t-$Student distributions. However, in this situation, the W distribution is heavier tail-weighted for small values of α. From equation (4), it is straightforward that $\lim_{\alpha \to \infty} \beta_2(W) = \beta_2(X) = 3$.

3.2 Estimation

Maximum likelihood estimator for α is obtained by solving the maximum likelihood equation

$$\frac{d}{d\alpha} \ln L(\alpha|\mathbf{w}) = 0$$

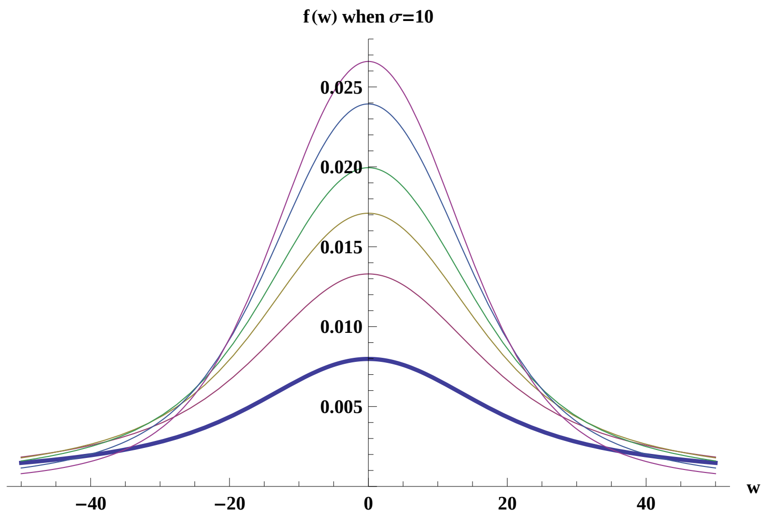

Fig. 1. Densities for the mixtures models where, from lower $f(0)$ $(\alpha = 0.25)$ to higher $f(0)$ $(\alpha = 2)$, $\alpha \in \{.25, .5, .75, 1, 1.5, 2\}$.

in order to α, where

$$L(\alpha|\mathbf{w}) = \prod_{i=1}^{n} f_W(w_i)$$

is the likelihood function. From equation (6),

$$L(\alpha|\mathbf{w}) = \prod_{i=1}^{n} \alpha 2^{0.5\alpha-1} |w_i|^{-\alpha-1} \pi^{-0.5} \gamma\left(\frac{\alpha+1}{2}, \frac{w_i^2}{2}\right)$$

and therefore the equation to be solved is

$$\frac{d}{d\alpha} \ln\left(\prod_{i=1}^{n} \alpha 2^{0.5\alpha-1} |w_i|^{-\alpha-1} \pi^{-0.5} \gamma\left(\frac{\alpha+1}{2}, \frac{w_i^2}{2}\right)\right) = 0$$

that doesn't have an explicit solution. Numerical analysis techniques can be applied to obtain point estimates for a particular dataset. Nevertheless, in this work we will use other methods that allow to obtain explicit estimators that can be used for any dataset.

Estimation of μ Note that the mixture median, denoted by Me, is defined as $P(W < \mathrm{Me}) = 0.5$. Therefore, $\mathrm{Me} = \mu$ and an estimator for μ can be

$$\widehat{\mu} = me, \tag{9}$$

where me represents the sample median. Using moments estimator, and assuming $\alpha > 1$, $E(W) = \mu + \sigma E(\Theta) E(X) = \mu$ and another estimator for μ can be

$$\widehat{\mu} = \overline{X}, \tag{10}$$

where $\overline{X}$ represents the sample mean.

Estimation of σ and α Recall that the k-th moment of the mixture can only exist when $k < \alpha$, independently of the distribution of X. Therefore, as we are estimating two more parameters, we must have $\alpha > 3$ to use the moments method when estimating (μ, σ, α), or at least $\alpha > 2$ if μ is estimated using equation (9). This is a penalizing restriction to the model, because our main purpose is to increase variance and tail-weight, which is achieved for small values of α, mainly $\alpha \in (0, 2]$. However, it is possible to apply a data transformation that will allow the estimation of (μ, σ, α) by the moments method. The transformed r.v. W' is defined as in [3] by

$$W' = \ln|W - \mu| = \ln\sigma + \ln\Theta + \ln|X| = \left(\sigma' + \Theta'\right) + X' = \Theta'' + X'. \tag{11}$$

W' moments exist for the majority of distributions (excluding when X is a super heavy tailed distribution, like the log-Pareto). Now, it is important to check the distributions of Θ'' and X'. Θ'' has a shifted exponential distribution, that is

$$\Theta'' \sim Exp\left(\sigma', \alpha\right), \tag{12}$$

where

$$f_{\Theta''}(x) = \alpha e^{-\alpha\left(x-\sigma'\right)}$$

for $x > \sigma'$. Thence, Θ'' moments are

$$\begin{cases} E\left(\Theta'\right) = \sigma' + \dfrac{1}{\alpha} = \ln\sigma + \dfrac{1}{\alpha} \\ Var\left(\Theta'\right) = \dfrac{1}{\alpha^2}. \end{cases} \tag{13}$$

Checking now X' distribution, note that its density function is given by

$$f_X(x) = e^{x - \frac{e^2 x}{2}} + \sqrt{\frac{2}{\pi}}x, \tag{14}$$

with $x \in \mathbb{R}$. X' moments can be found using standard formulas and equation (14),

$$\begin{cases} E\left(X'\right) = -\dfrac{\gamma + \ln 2}{2} \\ Var\left(X'\right) = \dfrac{\pi^2}{8} \end{cases} \tag{15}$$

where γ represents the Euler constant, $\gamma \approx 0.5772$. Finally, for W' moments

$$\begin{cases} E\left(W'\right) = E\left(\Theta'\right) + E\left(X'\right) = \ln\sigma + \dfrac{1}{\alpha} - \dfrac{\gamma + \ln 2}{2} \\ Var\left(W'\right) = Var\left(\Theta'\right) + Var\left(X'\right) = \dfrac{1}{\alpha^2} + \dfrac{\pi^2}{8} \end{cases} \tag{16}$$

which can be write as

$$\begin{cases} \ln\sigma = E\left(W'\right) - \dfrac{1}{\alpha} + \dfrac{\gamma + \ln 2}{2} \\ \alpha^2 = \left[Var\left(W'\right) - \dfrac{\pi^2}{8}\right]^{-1} \end{cases} \tag{17}$$

where $Var(W')$ represent W' variance. The moments estimators will be

$$\begin{cases} \widehat{\alpha} = \left[S'^2 - \dfrac{\pi^2}{8}\right]^{-1/2} \\ \widehat{\sigma} = \exp\left[\overline{W'} - \widehat{\alpha}^{-1} + \dfrac{\gamma + \ln 2}{2}\right] \end{cases} \tag{18}$$

where $\overline{W'}$ and S'^2 represent the sample mean and the sample variance, respectively.

4 Simulation study

The performance of the estimators defined on equations (9) and (18) was verified through a simulation study.

Table 2. Simulation results of μ, σ and α estimation (10^6 replicas) and samples with dimension $n = 100$.

	$n = 100$					
	$\mu = 10,\ \sigma = 1$			$\mu = 10,\ \sigma = 10$		
α	Mean	SD	MSE	Mean	SD	MSE
	μ estimation					
0.25	10.00	.6757	.4566	9.998	6.762	45.72
0.50	10.00	.3829	.1466	10.00	3.820	14.59
0.75	10.00	.2938	.0863	10.00	2.940	8.642
1.00	10.00	.2509	.0629	10.00	2.511	6.303
1.50	10.00	.2083	.0434	10.00	2.084	4.343
2.00	10.00	.1873	.0351	10.00	1.873	3.508
	σ estimation					
0.25	1.083	.4594	.2180	10.83	4.588	21.74
0.50	.9788	.2878	.0833	9.787	2.879	8.332
0.75	.9514	.2752	.0781	9.515	2.756	7.834
1.00	.9423	.2986	.0925	9.426	2.980	9.213
1.50	.9323	.3323	.1150	9.330	3.324	11.50
2.00	.8951	.3088	.1064	8.951	3.087	10.63
	α estimation					
0.25	.2557	.0361	.0013	.2557	.0361	.0013
0.50	.5038	.0798	.0064	.5038	.0798	.0064
0.75	.7473	.1623	.0264	.7475	.1600	.0256
1.00	1.012	.5789	.3353	1.012	.6950	.4832
1.50	1.719	2.671	7.185	1.720	2.558	6.589
2.00	2.287	5.126	26.36	2.278	4.422	19.64

In this study, each simulated sample had dimension $n = 100$ (Table 2) or $n = 1000$ (Table 3), and the parameters values were $\mu = 10$, $\sigma \in \{1, 10\}$ and $w \in \{0.25, 0.50, 0.75, 1, 1.5, 2\}$. For each combination 10^6 replicas were simulated and the parameters were estimated

Table 3. Simulation results for μ, σ and α estimation (10^6 replicas) and samples with dimension $n = 1000$.

	$n = 1000$					
	$\mu = 10,\ \sigma = 1$			$\mu = 10,\ \sigma = 10$		
α	Mean	SD	MSE	Mean	SD	MSE
	μ estimation					
0.25	10.00	.1994	.0397	10.00	1.991	3.964
0.50	10.00	.1191	.0141	10.00	1.189	1.414
0.75	10.00	.0925	.0086	10.00	.9253	.8563
1.00	10.00	.0793	.0063	10.00	.7920	.6268
1.50	10.00	.0660	.0044	10.00	.6603	.4360
2.00	10.00	.0595	.0035	10.00	.5938	.3527
	σ estimation					
0.25	1.002	.1375	.0189	10.02	1.375	1.890
0.50	.9878	.0896	.0082	9.882	.8951	.8152
0.75	.9820	.0846	.0075	9.819	.8466	.7493
1.00	.9778	.0893	.0085	9.777	.8927	.8468
1.50	.9719	.1071	.0123	9.719	1.072	1.227
2.00	.9704	.1312	.0181	9.704	1.313	1.812
	α estimation					
0.25	.2503	.0114	.0000	.2503	.0114	.0000
0.50	.4985	.0246	.0006	.4986	.0246	.0006
0.75	.7437	.0438	.0020	.7436	.0438	.0020
1.00	.9853	.0749	.0058	.9853	.0748	.0058
1.50	1.464	.2024	.0423	1.464	.2028	.0424
2.00	2.000	1.257	1.579	1.999	.9792	.9589

using equations (9) and (18). The results can be seen in Table 2 and Table 3.
When analysing Table 2, μ estimative seems clearly unbiased, with larger standard deviation (SD) and mean square error (MSE) when $\sigma = 10$ and when α decreases. This was expected because the distribution is heavier tail-weighted when α is small, and variance increases with σ. As for σ estimation, there is a slight bias (for $\alpha \geq 0.5$ it is always negative) but with almost constant levels of SD and MSE. Finally, for α estimation, the bias is also low, but with larger SD and MSE for higher values of α. This is natural because the SD and MSE increase with the parameter. The results displayed in Table 3 are quite similar to the ones observed in Table 2. Bias, SD and MSE decrease as n increases. In general, the performance of the estimators is quite good and, therefore, the introduced procedure can be applied in practical issues.

5 Modelling the directions chosen by ants

To exemplify the Gaussian mixture model, this subsection present an "almost symmetrical" data set modelling. In the *R-package circular - fisher B7 data set* are available the directions chosen by 100 ants in response to an evenly illuminated black target. If the ants react positively to the stimulus, as expected, they should go directly to the the target in an 180 degrees movement. This data set is also available in [7] and has its origins in [8]. Some more or less complex models were proposed by several authors [2, 8, 10] to deal with this kind of problem, but often the adjusted models reveal some low quality fit. Since the data is rounded to 10 grades, we summed a random uniform number to each observation, to avoid excessive probabilities accumulations points that would lead to a small p-value when using the K-S test (similar to the procedure indicated by [10]). We used as estimators for (μ, σ, α) the ones indicated in equations (9) and (18). The obtained estimates were $\hat{\mu} = 181.98$, $\hat{\sigma} = 18.056$ and $\hat{\alpha} = 1.265$. The K-S test yielded p-value= 0.6560, implying that the fitted model should not be rejected. When analysing Figure 2, the model seems well fitted to the data, despite a slight S shape in the qq-plot.

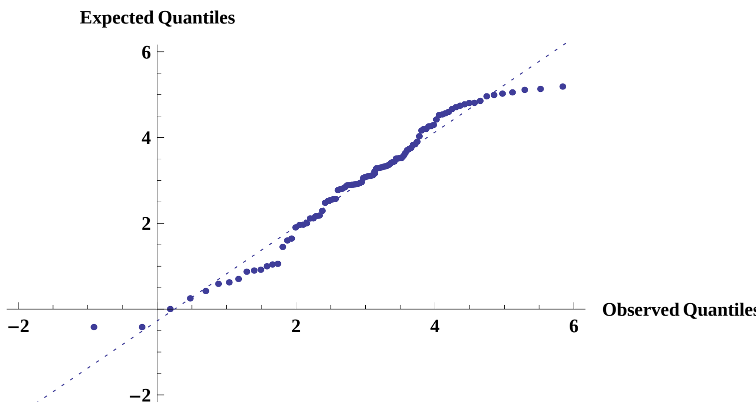

Fig. 2. qq-plot for the fitted Gaussian mixture.

6 Conclusions

Modelling heavy-tailed data is always a difficult process, because standard methods tend to give more importance to the central part of the data and less to the most extreme observations. For symmetrical data, we proposed a model that uses a Gaussian distribution as the starting point, but with a random scale parameter based on the Pareto distribution. To increase flexibility, fixed location and scale parameters were also included leading to a three parameter model. The estimation procedure based on median and moments was validated under simulation, and latter applied to a real data set. The achieved results are quite encouraging as the proposed estimators are easy to apply and revealed good fit to both simulated and real data

7 Discussion and Future Work

Despite the good results that were achieved with the proposed estimation method, it would be a step forward to develop more robust estimation methods. Also, instead of using Gaussian distribution as the starting point, other distributions could be considered, regardless of its symmetry. Ideally, a general estimation method that is independent from the starting distribution should be obtained and evaluated on simulated data. Finally, the development of confidence intervals for the mixture parameters would be a plus.

References

1. Castillo, J.: The extended slash distribution of the sum of two independent logistic random variables. Communications in Statistics - Theory and Methods. https://doi.org/10.1080/03610926.2021.1888123 (2021).
2. Coolen, F.: On Nonparametric Predictive Inference and Objective Bayesianism. Journal of Logic **15**, 21–47 (2006).
3. Felgueiras, M., Santos, R.: Exploring Pareto scale mixtures. AIP Conference Proceedings **1479**, 1121 (2012).
4. Felgueiras, M.: Explaining the seismic moment of large earthquakes by heavy and extremely heavy tailed models. International Journal on Geomathematics **3**(2), 209–222 (2012).
5. Felgueiras, M.: Pareto Scale Mixtures, in Studies in Theoretical and Applied Statistics, Springer, Berlin (2013).
6. Felgueiras, M., Santos, R., Martins, J.: Pareto Models for the Energy Released in Earthquakes. WSEAS Transactions on Power Systems **15**, 94–102 (2020).
7. Fisher, N.: Statistical analysis of circular data. Cambridge (1993).
8. Jander, R.: Die optische Richtangsorientierung der roten Waldameise (Formica rufa. L.). Zeitschrift fur vergMehende Physiologie, **40**, 162–238 (1957).
9. Johnson, N., Kotz, S., Kemp, A.: Univariate Discrete Distributions, Wiley, (1992).
10. Jones, M., Pewsey, A.: A Family of Symmetric Distributions on the Circle. Journal of the American Statistical Association **100**(472), 1422–1428 (2005).
11. Kelker, D.: Infinite divisibility and variance mixtures of the normal distribution. The Annals of Mathematical Statistics **42**, 802–808 (1971).
12. Pisarenko, V., Sornette, A., Sornette, D., Rodkin, M.: Characterization of the tail of the distribution of earthquake magnitudes by combining the GEV and GPD descriptions of Extreme Value Theory. Pure and Applied Geophysics **171**, 1599–1624 (2014)
13. Zörnig, P.: On Generalized Slash Distributions: Representation by Hypergeometric Functions. Stats **2-26**, 371–388 (2019).