Journal of Statistical Theory and Applications
Vol. 19(4), December 2020, pp. 481–486
DOI: https://doi.org/10.2991/jsta.d.201208.001; ISSN 1538-7887
https://www.atlantis-press.com/journals/jsta

Modified Maximum Likelihood Estimations of the Epsilon-Skew-Normal Family

Parichehr Jamshidi [1], Mohsen Maleki [2,*], Zahra Khodadadi [1]

[1] Department of Statistics, Marvdasht Branch, Islamic Azad University, Marvdasht, Iran
[2] Department of Statistics, Faculty of Mathematics and Statistics, University of Isfahan, Isfahan, Iran

ARTICLE INFO

Article History

Received 17 Apr 2018
Accepted 20 Mar 2020

Keywords

Asymmetry
EM-algorithm
Epsilon-skew-normal
Maximum likelihood estimates
Two-piece distributions

2000 Mathematics Subject Classification: 22E46, 53C35, 57S20

ABSTRACT

In this work, maximum likelihood (*ML*) estimations of the epsilon-skew-normal (*ESN*) family are obtained using an *EM*-algorithm to modify the ordinary estimation already used and solve some of its problems within issues. This family can be used for analyzing the asymmetric and near-normal data, so the skewness parameter epsilon is the most important parameter among others. We have shown that the method has better performance compared to the method in G.S. Mudholkar, A.D. Hutson, J. Statist. Plann. Infer. 83 (2000), 291–309, especially in the strong skewness and small samples. Performances of the proposed *ML* estimates are shown via a simulation study and some real datasets under some statistical criteria as a way to illustrate the idea.

1. INTRODUCTION

The importance of the asymmetric distributions in various applications such as meteorology, physics, economics, geology, etc., has been rapidly increasing. Also, the asymmetric distributions which contain famous distributions in the symmetric cases, such as normal distribution, are more important among all others. The *epsilon-skew-normal* (*ESN*) distribution which was introduced by Refs. [1–6], is a flexible family to model the asymmetry data and statistical models. The *ESN* parameters maximum likelihood (*ML*) estimates were found by Refs. [3,7], and their Bayesian estimates were studied by Ref. [8] as well as Ref. [9].

The classical *ML* estimates of this family and its application on some statistical models (as regression model by Ref. [7]; time series model by Ref. [10]; Tobit regression by Ref. [11]) were performed by an especial approach which ordered the data. Applying the variations of each produced segment and three possible likelihood function forms as well as their proposed estimates, the one with higher likelihood values was chosen as the approximation of the *ML* estimates (see e.g. Ref. [12]). But, there exist some problematic issues on this method, e.g., in the strong asymmetry, in the small samples, and the right/left half-normal distribution estimates in which there are not any real distributions. We have focused on an especial mixture of two right/left half-normal distributions which lead to *ESN* family, and have used an EM-algorithm to obtain the *ML* estimates of the model parameters. In addition, we have shown the modifications of the proposed *ML* estimates without the maintained issues (see, e.g., Ref. [13]).

The rest of this paper is organized as follows: Some properties of the *ESN* family and ordinary method of finding the *ML* estimates are considered in Section 2. The new approach of finding the *ML* estimates based on the mixture distributions are provided in Section 3. In Section 4, in order to show the performance of the proposed methodology, some simulation studies are provided which are later used to some real dataset. Finally, the conclusion is given in Section 5.

2. THE ESN FAMILY

The *ESN* distribution denoted by $ESN(\theta, \sigma, \varepsilon)$ is a unimodal distribution with mode and location parameter $\theta \in \mathbb{R}$, scale parameter $\sigma \in \mathbb{R}^+$, skewness parameter $\varepsilon \in (-1, 1)$, and probability masses $(1+\varepsilon)/2$ at below the mode and $(1-\varepsilon)/2$ at above the mode, with the following

*Corresponding author. Email: m.maleki.stat@gamil.com

standard density function of $X \sim ESN(0, 1, \varepsilon)$:

$$f_{\text{esn}}(x; 0, 1, \varepsilon) = \begin{cases} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2(1+\varepsilon)^2}\right), & x < 0 \\ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2(1-\varepsilon)^2}\right), & x \geq 0 \end{cases}, \tag{1}$$

Note that, $f_{\text{esn}}(x; \theta, \sigma, \varepsilon) = \frac{1}{\sigma} f_{\text{esn}}\left(\frac{x-\theta}{\sigma}; 0, 1, \varepsilon\right)$, and it has the same range of skewness as the skew-normal distribution investigated by Refs. [14–16]. The standard random variable $X \sim ESN(\theta, \sigma, \varepsilon)$ has the following stochastic representation:

$$X = \theta + \sigma(1-U)(1-\varepsilon)|Z_1| - \sigma U(1+\varepsilon)|Z_2|, \tag{2}$$

where U, Z_1 and Z_2 are independent, for which $P(U=1) = (1+\varepsilon)/2 = 1 - P(U=0)$, and Z_1 and Z_2 are the standard normal distributed.

The mean and variance of the random variable $X \sim ESN(\theta, \sigma, \varepsilon)$ respectively are

$$E(X) = \theta - \frac{4\sigma\varepsilon}{\sqrt{2\pi}}, \ \ \text{Var}(X) = \frac{\sigma^2}{\pi}\left[(3\pi - 8)\varepsilon^2 + \pi\right]. \tag{3}$$

To see more statistical details of the *ESN* distribution, refer to the Refs. [1,3].

To obtain the *ML* estimates of the $X \sim ESN(\theta, \sigma, \varepsilon)$, Ref. [3] as well as Ref. [1] considering the sample $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ and its order statistic of the sample $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$, have assumed that there exists the auxiliary integer k such that the first k-th samples come from the left half-normal and the remaining samples from the right half-normal. Finally, by considering the possible values of $k = 0, n$ (corresponds to right/left half-normal) and other values between these two values which lead to $(n-1)$ half-open intervals in the form of $\left[X_{(j)}, X_{(j+1)}\right]; j = 1, \ldots, (n-1)$, and using the numerical method, they have obtained the $(n+1)$ plausible *ML* estimates and choose the one with the lowest likelihood values as the *ML* estimates of the *ESN* parameters. (See full details of this method and statistical properties of the *ESN* family in Ref. [3].) As previously mentioned, some problematic issues emerge within this method which is later discussed in Section 4. Moreover, we have used an especial method of constructing the *ESN* family with mixture models and applied an *EM*-algorithm to have the modified *ML* estimates of the *ESN* family parameters in the Section 3.

3. ML ESTIMATES OF THE ESN PARAMETERS USING AN EM-ALGORITHM

3.1. ML Estimates

In fact the location-scale *ESN* distribution is the reparameterization of a mixture of left- and right half-normal (*RHN*) densities with special component probabilities as follows:

$$f_{esn}\left(x|\theta, \sigma_1, \sigma_2\right) = 2\pi\phi\left(x|\theta, \sigma_1\right) I_{(-\infty,\theta)}(x) + 2(1-\pi)\phi\left(x|\theta, \sigma_2\right) I_{(\theta,+\infty)}(x), \tag{4}$$

where $\pi = \sigma_1/(\sigma_1 + \sigma_2)$. Note that in this form, the scale parameter σ and skewness parameter recover as in the form of $\sigma = (\sigma_1 + \sigma_2)/2$ and $\varepsilon = (\sigma_1 - \sigma_2)/2\sigma$.

By using an *EM*-algorithm to obtain the *ML* estimates of the *ESN* parameters $\Theta = (\theta, \sigma_1, \sigma_2)^\top$, for each i.i.d. sample $\boldsymbol{X} = (X_1, \ldots, X_n)^\top \sim ESN(\Theta)$, by using auxiliary (latent) variables $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^\top$ (i.e., completed data $\boldsymbol{D} = (\boldsymbol{X}, \boldsymbol{Z})^\top$, where in terms of the components of the mixture (4) can be equivalently represented as

$$\begin{cases} X_i|Z_i = 1 \sim LHN(\theta, \sigma_1) \\ X_i|Z_i = 0 \sim RHN(\theta, \sigma_2) \end{cases}, \quad i = 1, \ldots, n, \tag{5}$$

where *LHN* and *RHN* denotes the left- and right half-normal distribution, respectively and $Z_i \sim Binomial(1, \pi); i = 1, \ldots, n$ is a multinomial (component-label) vector with probability mass function $P(Z_i = z_i) = \pi^{z_i}(1-\pi)^{1-z_i}$, for which $z_i = 0, 1; i = 1, \ldots, n$. So the augmented (completed) log-likelihood function is in the form of

$$\ell\left(\Theta|\boldsymbol{D}\right) = -n\log(\sigma_1 + \sigma_2) - \frac{1}{2}\sum_{i=1}^{n}\left[Z_i\left(\frac{X_i - \theta}{\sigma_1}\right)^2 + (1 - Z_i)\left(\frac{X_i - \theta}{\sigma_2}\right)^2\right], \tag{6}$$

where $\Theta = (\theta, \sigma_1, \sigma_2)^\top$.

The conditional expectation of latent variables is $\hat{z}_i = E\left[Z_i|\hat{\Theta}, x_i\right] = I_{\left(-\infty,\hat{\theta}\right)}(x_i)$. Now, the *E-Step* on the $(\text{k}+1)$ th iteration of the *EM*-algorithm (Ref. [17]) requires the calculation of Q-function, i.e., in the form of $Q(\Theta|D) = E_{\Theta}\left[\ell(\Theta|D)\right]$. So,

E-Step:

$$Q\left(\Theta|\boldsymbol{D}\right) = -n\log\left(\sigma_1+\sigma_2\right) - \frac{1}{2}\sum_{i=1}^{n}\left[\hat{z}_i\left(\frac{x_i-\theta}{\sigma_1}\right)^2 + \left(1-\hat{z}_i\right)\left(\frac{x_i-\theta}{\sigma_2}\right)^2\right]. \tag{7}$$

M-steps:

M-step 1: Update θ by

$$\hat{\theta} = \frac{\sum_{i=1}^{n}\left[\hat{z}_i\sigma_2^2 + \left(1-\hat{z}_i\right)\sigma_1^2\right]X_i}{\sum_{i=1}^{n}\left[\hat{z}_i\sigma_2^2 + \left(1-\hat{z}_i\right)\sigma_1^2\right]}.$$

M-steps 2–3: Update $\sigma_j, j = 1, 2$, by solving the following stressed cubic equation:

$$\sigma_j^3 + p\sigma_j + q = 0; j = 1, 2,$$

where $p = -\frac{1}{n}\sum_{i=1}^{n}\hat{z}_{ij}\left(x_i-\theta\right)^2$ and $q = p\sigma_i$, for which $\hat{z}_{ij} = \hat{z}_i I_{(j=1)} + \left(1-\hat{z}_i\right)I_{(j=2)}$. Note that $p < 0$ and $q < 0$, so the cubic equation has unique just root in the $(0, +\infty)$ interval. The *EM*-algorithm must be iterated so that a sufficient convergence rule is satisfied, e.g. if $||\hat{\Theta}^{(k+1)} - \hat{\Theta}^{(k)} \leq \varepsilon||$ (see Ref. [18]).

3.2. Model Selection

In this paper we have just *ESN* family but with different numerical *ML* estimate types, therefore we compare different *ESN* distributions based on different estimated parameters (through mentioning different numerical approaches) to better fit on the simulated and real datasets. The Akaike information criteria (*AIC*; Ref. [19]) is in the form of $AIC = 2\text{k} - 2\ell\left(\hat{\Theta}|\boldsymbol{x}\right)$, where $\ell\left(\hat{\Theta}|\boldsymbol{x}\right)$ is the maximized log-likelihood function, the *Kolmogorov–Smirnov* (*K-S*) and *Anderson–Darling* (*A-D*) statistic tests are implemented to choose more suitable models. The one-sample *K-S* statistic is given by $D_n = sup_x|F_n(X) - F(X;\Theta)|$, where sup_x is the *supremum* of the set of distances between the empirical distribution function $F_n(\cdot)$ (Ref. [20]) and target *ESN* distribution function $F(\cdot)$, and the *A-D* statistic is given by $A^2 = -n - S$, where $S = \sum_{i=1}^{n}\frac{2i-1}{n}\left[lnF(X_i;\Theta) - ln\left(1-F\left(X_{n+1-i};\Theta\right)\right)\right]$, for the target distribution $F(\cdot;\Theta)$ with sample $\boldsymbol{X} = \left(X_1, \dots, X_n\right)^\top \sim ESN(\Theta)$. The minimum values of the mentioned criteria choose the more suitable model (and as a result, better parameter estimates).

4. NUMERICAL STUDIES

In this section, we simulate the some strongly and weakly skewed *ESN* samples and use the proposed *ML* estimates (denoted by *Pr-ML*) to evaluate the ordinary *ML* estimates (denoted by Or-*ML*) which correspond to Refs. [1,3]. Then we apply the both *ML* estimation methods to some real datasets. The implementation of the necessary algorithms is based on the *R* software version 3.5.2 with a core i7 760 processor 2.8 GHz, and the relative tolerance of 10^{-3} is used for convergence of the *EM*-algorithms.

4.1. Simulations

In this part, we consider 10,000 samples of size $n = 50$, 100, and 250 from various weak and moderate skewness $\varepsilon = 0.5, -0.5$, and strong skewness $\varepsilon = 0.85, -0.95$, respectively, with standard location-scale parameter values. We recorded the means and standard deviations of the *Pr-ML* and *Or-ML* estimates of parameters in Table 1. The results show the performance of the *Pr-ML* estimates in each sample size.

4.2. Applications

In this section, considering four various real datasets, we show the performance of the proposed *Pr-ML* estimates in applications. All of *ESN* parameters estimates and criteria are given in Table 2, and the fitted *ESN* densities based on the two approaches *Pr-ML* and *Or-ML* estimates are curved on the histograms of the mentioned datasets in Figure 1.

The first dataset is corresponds to the "*Tstop*" component of the "*Bronchiolitis obliterans syndrome after lung transplants*" called "*bosms3*" and available in the "*flexsurv*" package of *R* software. Both *Pr-ML* and *Or-ML* estimates satisfy the purely skewed right half-normal ($\varepsilon = -1$) distributions and the criteria are approximately identical on the dataset (see, e.g., the Table 2 and top-left of the Figure 1).

The second dataset is corresponds to the "weight" component of the "*Weight versus age of chicks on different diets*" called "*ChickWeight*" and it is available in the "*datasets*" package of *R* software. In this case, although the *Or-ML* and *Pr-ML* estimates are close (see, e.g., the Table 2 and top-right of the Figure 1), but all of the criteria prefer the fitted *ESN* distribution based on the *Pr-ML* to *Or-ML* estimates.

Table 1 Mean and standard deviations (SDs) of the 10,000 times *Or-ML* and *Pr-ML* estimates of the ESN distribution.

Parameters	ML	n = 50		n = 100		n = 250	
		Mean	SD	Mean	SD	Mean	SD
θ (0)	*Or-ML*	−0.1026	0.1637	0.0837	0.0936	−0.0657	0.0746
	Pr-ML	−0.1017	0.1726	0.0866	0.0894	−0.0719	0.0563
σ (1)	*Or-ML*	1.0258	0.0403	1.0137	0.0304	0.9846	0.0307
	Pr-ML	1.0144	0.0397	1.0003	0.0294	0.9930	0.0308
ε (0.05)	*Or-ML*	0.0618	0.0106	0.0589	0.0095	0.0440	0.0082
	Pr-ML	0.0585	0.0098	0.0558	0.0094	0.0528	0.0080
θ (0)	*Or-ML*	0.1134	0.2016	0.1037	0.1073	0.0589	0.0374
	Pr-ML	0.1076	0.2113	0.0937	0.0783	0.0593	0.0412
σ (1)	*Or-ML*	1.0312	0.0553	1.0207	0.0494	1.0201	0.0365
	Pr-ML	1.0189	0.0388	1.0054	0.0307	1.0036	0.0340
ε (−0.5)	*Or-ML*	−0.5303	0.0128	−0.5203	0.0097	−0.5176	0.0066
	Pr-ML	−0.5274	0.0078	−0.5148	0.0071	−0.5104	0.0061
θ (0)	*Or-ML*	0.1981	0.1803	−0.1037	0.1006	0.0579	0.0793
	Pr-ML	0.1805	0.1891	−0.0937	0.0954	0.0667	0.0534
σ (1)	*Or-ML*	1.0537	0.0683	1.0365	0.0546	1.0311	0.0397
	Pr-ML	1.0203	0.0442	1.0112	0.0410	1.0103	0.0385
ε (0.85)	*Or-ML*	0.9204	0.0983	0.9048	0.0784	0.8910	0.0719
	Pr-ML	0.8399	0.0068	0.8401	0.0065	0.8411	0.0059
θ (0)	*Or-ML*	0.1907	0.1887	0.1117	0.0936	0.0794	0.0864
	Pr-ML	0.1870	0.1804	0.0946	0.0911	0.0642	0.0570
σ (1)	*Or-ML*	1.0497	0.0702	1.0405	0.0528	1.0352	0.0401
	Pr-ML	1.0286	0.0513	1.0201	0.0465	1.0112	0.0389
ε (−0.95)	*Or-ML*	−0.9987	0.0057	−0.9902	0.0102	−0.9893	0.0096
	Pr-ML	0.9601	0.0113	−0.9578	0.0094	−0.9523	0.0075

ESN, epsilon-skew-normal; ML, maximum likelihood; Pr-ML, proposed maximum likelihood; Or-ML, ordinary maximum likelihood.

Table 2 The *Or-ML* and *Pr-ML* estimates of the fitted *ESN* distributions on four real datasets.

Data	ML	Parameter			Criteria		
		θ	σ	ε	*AIC*	*K-S*	*A-D*
1st Data	Or-ML	0.0246	1.8938	−1.0000	2108.798	0.0417	1.4202
	Pr-ML	0.0246	1.8939	−1.0000	2108.841	0.0416	1.4169
2nd Data	Or-ML	43.9999	55.2043	−0.9384	6317.510	0.1047	12.8735
	Pr-ML	40.9727	55.5884	−0.9569	6291.083	0.0945	8.3241
3rd Data	Or-ML	0.9790	0.2994	−0.0578	3643.910	0.0551	62.1247
	Pr-ML	1.1461	0.2919	0.3249	3015.756	0.0343	13.3147
4th Data	Or-ML	4.4999	0.2726	−0.6772	4167.713	0.3357	592.7232
	Pr-ML	4.2438	0.3771	−0.6155	893.682	0.0796	4.3623

ESN, epsilon-skew-normal; ML, maximum likelihood; Pr-ML, proposed maximum likelihood; Or-ML, ordinary maximum likelihood; K-S, Kolmogorov–Smirnov; A-D, Anderson–Darling.

The third dataset is corresponds to the "*Yearly Treering Data*" called "*treering*" and it is available in the "datasets" package of R software. In this case, all of the criteria strongly prefer the fitted *ESN* distribution based on the *Pr-ML* to *Or-ML* estimates (see, e.g., the Table 2 and bottom-left of the Figure 1).

The fourth dataset is corresponds to the "*mag*" component of the "*Locations of Earthquakes off Fiji*" called "*quakes*" and it is available in the "*datasets*" package of R software. In this case, also all of the criteria strongly prefer the fitted *ESN* distribution based on the *Pr-ML* to *Or-ML* estimates (see, e.g., the Table 2 and bottom-right of the Figure 1).

5. CONCLUSION

We have proposed and implemented an *EM*-type algorithm to estimate the well-known *ESN* family parameters by applying the special stochastic representation. The proposed estimation methodology has better *ESN* distribution fitting, especially in the strong skewness. The

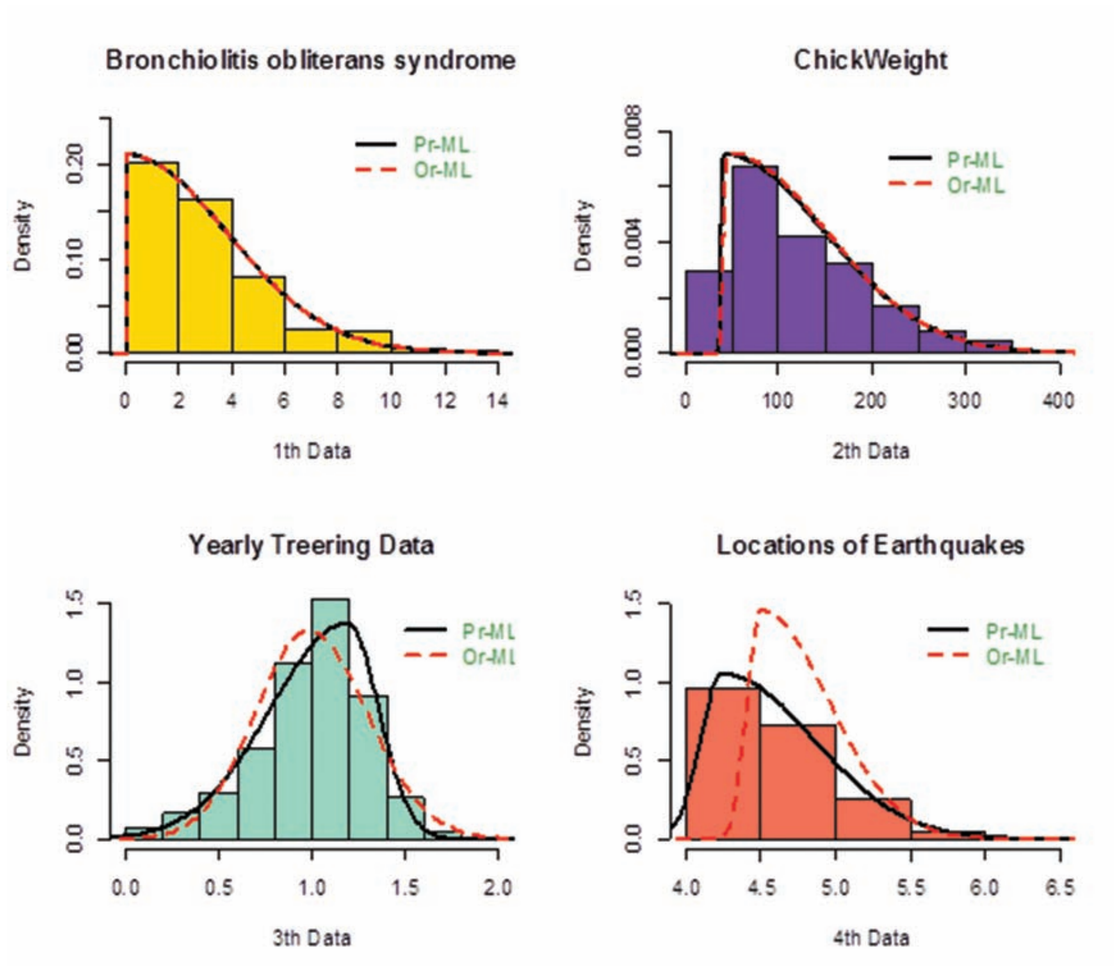

Figure 1 Histograms of the real datasets with the curved fitted epsilon-skew-normal (*ESN*) densities based on the proposed maximum likelihood (*Pr-ML*) and ordinary maximum likelihood (*Or-ML*) estimates.

performance of the proposed methodology is illustrated using the simulation studies and four real datasets. The performances of the *ESN* family have shown on many statistical models to cover the asymmetry, e.g., Refs. [3,7,10]. In fact, this methodology can be affronted on them to modify their parameter estimations.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this paper.

AUTHORS' CONTRIBUTIONS

All authors have read and agreed to the published version of the manuscript.

FUNDING STATEMENT

There is no funding of this paper.

ACKNOWLEDGMENTS

We would like to express our very great appreciation to editor and reviewer(s) for their valuable and constructive suggestions during the planning and development of this research work.

REFERENCES

1. R.B. Arellano-Valle, H. Gómez, F.A. Quintana, J. Stat. Plann. Infer. 128 (2005), 427–443.
2. B. Moravveji, Z. Khodadadi, M. Maleki, Iran. J. Sci. Technol. Trance. A. Sci. 43 (2018), 991–1001.
3. G.S. Mudholkar, A.D. Hutson, J. Stat. Plann. Infer. 83 (2000), 291–309.
4. M. Maleki, M.R. Mahmoudi, Commun. Stat. Theor. Meth. 46 (2017), 12356–12369.
5. M. Maleki, J.E. Contreras-Reyes, M.R. Mahmoudi, Axioms. 8 (2019), 38.
6. M. Maleki, Z. Barkhordar, Z. Khodadadi, D. Wraith, J. Stat. Comput. Simul. 89 (2019), 2765–2781.
7. A. Hutson, J. Appl. Stat. 31 (2004), 673–683.
8. F.J. Rubio, M.F.G. Steel, Bayesian Anal. 9 (2014), 1–22.
9. M. Maleki, A.R. Nematollahi, Commun. Stat. Theor. Meth. 46 (2017), 7546–7561.
10. P. Bondon, J. Mult. Anal. 100 (2009), 1761–1776.
11. T.L. Mashtare, A.D. Hutson, Commun. Stat. Theor. Meth. 40 (2011), 408–423.
12. H.W. Gómez, F.J. Torres, H. Bolfarine, Commun. Stat. Theor. Meth. 36 (2007), 73–81.
13. A. Hoseinzadeh, M. Maleki, Z. Khodadadi, J.E. Contreras-Reyes, J. Comput. Appl. Math. 349 (2019), 132–141.
14. A. Azzalini, Scand. J. Stat. 12 (1985), 171–178. https://www.jstor.org/stable/4615982
15. A. Azzalini, Statistica. 46 (1986), 199–208.
16. B. Liseo, The skew-normal class of densities: Inferential aspects from a bayesian viewpoint (in Italian). Statistica. 50 (1990), 71–82.
17. A.P. Dempster, N.M. Laird, D.B. Rubin, J. R. Stat. Soc. B. 39 (1977), 1–22.
18. X. Meng, D.B. Rubin, Biometrika. 80 (1993), 267–278.
19. H. Akaike, IEEE Trans. Autom. Control. 19 (1974), 716–723.
20. R.B. D'Agostino, M.A. Stephens, Goodness-of-Fit Techniques. Marcel Dekker, New York, 1986.