AUT Journal of Mathematics and Computing

AUT J. Math. Comput., 2(1) (2021) 53-60
DOI: 10.22060/ajmc.2020.18499.1037

Estimation of the parameter of Lévy distribution using ranked set sampling

Sahar Dorniani*[a], Adel Mohammadpour[b], Nader Nematollahi[c]

[a] *Department of Statistics, Roudehen Branch, Islamic Azad University, Tehran, Iran*
[b] *Department of Mathematics and Computer Science, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran*
[c] *Department of Statistics, Allameh Tabataba'i University, Tehran, Iran*

ABSTRACT: Ranked set sampling is a statistical technique for data collection that generally leads to more efficient estimators than competitors based on simple random sampling. In this paper, we consider estimation of scale parameter of Lévy distribution using a ranked set sample. We derive the best linear unbiased estimator and its variance, based on a ranked set sample. Also we compare numerically, variance of this estimator with mean square error of the maximum likelihood, a median based estimator and an estimator based on Laplace transform. It turns out that the best linear unbiased estimator based on ranked set sampling is more efficient than other mentioned estimators.

Review History:

Received:27 May 2020
Revised:02 December 2020
Accepted:20 December 2020
Available Online:01 February 2021

Keywords:

Best linear unbiased estimator
Lévy distribution
Ranked set sampling
Scale parameter

1. Introduction

Lévy distribution is an interesting distribution and play a very important rule in modeling many types of physical and economic systems. For example, hitting times for a Brownian motion yielding a Lévy distribution ([17]). In economics, Lévy distribution is appropriate for modeling financial data ([14]). For other applications of Lévy distribution, see [12] and references therein.

If $\Phi(x)$ denotes the standard normal distribution function, then

$$F(x,\gamma) = 2\Big\{1 - \Phi\Big(\sqrt{\tfrac{\gamma}{x}}\Big)\Big\}, \quad x > 0, \tag{1}$$

defines a distribution function with probability density function (pdf)

$$f(x,\gamma) = \sqrt{\frac{\gamma}{2\pi}}x^{-\frac{3}{2}}\exp(-\frac{\gamma}{2x}), \quad x > 0, \tag{2}$$

where the scale parameter γ is a positive real number. This distribution is known as Lévy distribution with parameter γ and if the random variable X has Lévy distribution with scale parameter γ, it is denoted by $X \sim$Lévy(γ).

Lévy distribution is a heavy tail and does not have any moment. Therefore, to estimate the scale parameter γ by a random sample $X_1, \ldots, X_n$ from this distribution, we cannot find an unbiased estimator based on a linear combination of $X_1, \ldots, X_n$. In this paper, we show that the moments of some order statistics of this distribution exist, and hence, one can use a linear combination of these order statistics to estimate the parameter of Lévy distribution.

*Corresponding author.
E-mail addresses: sdorniani@yahoo.com, adel@aut.ac.ir, na_nemat@yahoo.com

For estimation of the unknown parameter of a population, many sampling methods are suggested in the literature. An efficient sampling procedure that can be viewed as a generalization of the Simple Random Sampling (SRS) is Ranked Set Sampling (RSS). This sampling method can be used in situations where the sampling units may be difficult to measure, time-consuming, or costly, but ranking sampling units (e.g., by visual inspection) may be relatively easy. For example, in psychology, biology, medical, ecological and agricultural studies, a ranking of experimental or sampling units is possible without actually measuring them ([18]).

The RSS is based on some order statistics of the SRS and it is well-known that the RSS method is more efficient to estimate parameters than SRS method.

For estimating the scale parameter γ of Lévy distribution, [1] purposed the Maximum Likelihood (ML) estimator, a median-based (Q) estimator and an estimator based on Laplace Transform (LT). In this paper, we purpose the Best Linear Unbiased Estimator (BLUE) of γ based on a RSS. We derive the BLUE and its variance and compare the variance of BLUE with Mean Square Error (MSE) of ML, Q and LT estimators through a simulation study. It is shown that the BLUE is more efficient than mentioned estimators. To this end, in Section 2, we described RSS method. In Section 3, we derive the BLUE estimator. Finally, in Section 4, we compare the proposed estimator with the above-mentioned estimators of the scale parameter.

2. Ranked Set Sampling

In real life sampling situation where the measurement of the variable of interest from the experimental units is costly or time consuming but the ranking of sample items related to the variable can be easily done by judgment without actual measurement, the RSS method can be used and is highly beneficial and much superior to the standard SRS. The concept of RSS was used first time in [8], to estimate the population mean of pasture yields in agricultural experimentation. [16] and [3] provided a mathematical foundation for RSS. The properties of RSS in a variety of statistical procedures have been investigated in the literature. For more details, see [2].

The original RSS can be described as follows. A random sample of size n units is drown from a population. The units are ranked by judgment or any cost-free or little-cost method, and only the unit ranked the smallest is quantified. Then another random sample of size n units is drawn and ranked, and only the unit ranked the second smallest is quantified. The procedure replicates until the unit ranked the largest in the nth random sample of size n is quantified. In this way, we obtain a total of n measurement units, one from each sample. Let X_{ij} , $i, j = 1, \dots, n$ be n independent random sample of size n and $X_{(ij)}$ be the jth order statistic of the ith random sample, $i, j = 1, \dots, n$, then the RSS is $X_{(11)}, X_{(22)}, \dots, X_{(nn)}$ and the whole sample can be presented as shown in Table 1.

Table 1: Display of n^2 random sample for constructing RSS

$\mathbf{X_{(11)}}$	$X_{(12)}$	$\cdots$	$X_{(1n)}$
$X_{(21)}$	$\mathbf{X_{(22)}}$	$\cdots$	$X_{(2n)}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
$X_{(n1)}$	$X_{(n2)}$	$\cdots$	$\mathbf{X_{(nn)}}$

Note that $X_{(11)}, \dots, X_{(nn)}$ are independent and $X_{(ii)}$ is distributed as the ith order statistics of a random sample of size n.

In literature, RSS is used to estimate the population parameters in both nonparametric and parametric inferences. In parametric inference, [10] in the estimation of location parameter of Cauchy distribution, [4] in the estimation of parameters of two-parameter Weibull and extreme-value distributions, [6, 7] in estimation of parameters of two-parameter exponential and logistic distributions, [5] in the estimation of parameters of mixture models under RSS and [15] in the estimation of parameters of normal and exponential distributions, used RSS method to find ordinary RSS estimator and/or BLUE based on RSS. For a review of literature on parametric estimation based on RSS, see [11] and references therein.

3. Estimation Based on a Ranked Set Sample

In this section, we use linear combinations of order statistics for estimating scale parameter of the Lévy distribution using RSS. Let $X_{(11)}, X_{(22)}, \dots, X_{(nn)}$ be a RSS drown from Lévy distribution with scale parameter γ. We know that if $X \sim$ Lévy(γ), then $Z = \frac{X}{\gamma} \sim$ Lévy(1). So, $X_{(ii)} = \gamma Z_{(ii)}$ where $Z_{(11)}, \ldots, Z_{(nn)}$ is a RSS from Z. Therefore,

$$E(X_{(ii)}) = \gamma c_{(i)}, \quad \mathtt{var}(X_{(ii)}) = \gamma^2 d_{(i)},$$

where $c_{(i)}$ and $d_{(i)}$ are the mean and variance of $Z_{(ii)}$ which can be calculated as follows:

$$d_{(i)} = e_{(i)} - c^2_{(i)} \tag{3}$$

and

$$\begin{aligned} c_{(i)} &= C. \int_0^\infty z(2\pi z^3)^{-\frac{1}{2}} \exp(-\frac{1}{2z})\big[2(1-\Phi(\frac{1}{\sqrt{z}}))\big]^{i-1}\big[1-2(1-\Phi(\frac{1}{\sqrt{z}}))\big]^{n-i}dz, \\ e_{(i)} &= C. \int_0^\infty z^2(2\pi z^3)^{-\frac{1}{2}} \exp(-\frac{1}{2z})\big[2(1-\Phi(\frac{1}{\sqrt{z}}))\big]^{i-1}\big[1-2(1-\Phi(\frac{1}{\sqrt{z}}))\big]^{n-i}dz, \end{aligned} \tag{4}$$

where $C = \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)}$, $\Gamma(\cdot)$ is a gamma function and $\Phi(\cdot)$ is the standard normal distribution function. It should be noted that some of the moments of order statistics $X_{(11)}, X_{(22)}, \ldots, X_{(nn)}$ of Lévy distribution does not exist. Therefore, we are limited ourselves to use order statistics for which their moments exist. These order statistics can be determined in the following lemma.

Lemma 1. *Let m be a real number, $X_1, \ldots, X_n$ be a random sample from Lévy distribution with scale parameter $\gamma > 0$, and $X_{1:n}, \ldots, X_{n:n}$ be its corresponding order statistics. In order that $E(X^m_{i:n})$ exists it is necessary and sufficient that $i < n + 1 - 2m$.*
Proof. See the Appendix.

Take $m = 2$, then from Lemma 1 and the fact that the bounds of i must be an integer, variance of $X_{(ii)}$ exists if and only if $1 \le i \le n-4$.

For finding the BLUE of γ, we consider a linear combination of $X_{(ii)}$ as $\gamma^* = \sum_{i=1}^{n-4} k_i X_{(ii)}$. In the following theorem, we find the BLUE of γ among the estimators of the form γ^*.

Theorem 1. *Let X be a random variable with pdf (2) and $X_{(11)}, \ldots, X_{(nn)}$ be a RSS from this distribution. Then the BLUE of γ is given by*

$$\tilde{\gamma}_{BLUE} = \frac{\sum\limits_{i=1}^{n-4} \frac{c_{(i)}}{d_{(i)}} X_{(ii)}}{\sum\limits_{i=1}^{n-4} \frac{c^2_{(i)}}{d_{(i)}}} \tag{5}$$

with variance

$$\mathtt{var}(\tilde{\gamma}_{BLUE}) = \frac{\gamma^2}{\sum\limits_{i=1}^{n-4} \frac{c^2_{(i)}}{d_{(i)}}} \tag{6}$$

where $c_{(i)}$ and $d_{(i)}$ are defined in (4) and (3), respectively.
Proof. We consider the general form of a linear combination of $X_{(ii)}$'s as $\gamma^* = \sum_{i=1}^{n-4} k_i X_{(ii)}$ where k_i's are constants and minimizing the variance of γ^* under unbiasedness condition: $\sum\limits_{i=1}^{n-4} k_i c_{(i)} = 1$ and $c_{(i)}$ is defined in (4). To determine the minimizing value of $k_i, i = 1, \ldots, n-4$, we use the Lagrange method, i.e.,

$$\begin{aligned} &\Lambda = \mathtt{var}\Big(\sum_{i=1}^{n-4} k_i X_{(ii)}\Big) + \lambda\Big(\sum_{i=1}^{n-4} k_i c_{(i)} - 1\Big), \\ &\frac{\partial \Lambda}{\partial k_i} = 0 \qquad \Rightarrow \qquad 2\gamma^2 k_i d_{(i)} + \lambda c_{(i)} = 0, \\ &\frac{\partial \Lambda}{\partial \lambda} = 0 \qquad \Rightarrow \qquad \sum_{i=1}^{n-4} k_i c_{(i)} = 1. \end{aligned}$$

Solving the above system of equations, we get

$$\lambda = \frac{-2\gamma^2}{\sum_{i=1}^{n-4} \frac{c_{(i)}^2}{d_{(i)}}},$$

and

$$k_i = \frac{-\lambda c_{(i)}}{2\gamma^2 d_{(i)}} = \frac{c_{(i)}}{d_{(i)} \sum_{i=1}^{n-4} \frac{c_{(i)}^2}{d_{(i)}}},$$

which gives the $\tilde{\gamma}_{\text{BLUE}}$ in (5). Note that it is easy to show that the above solution is the minimum point of Λ.

4. Comparison

For estimating the scale parameter of Lévy distribution, [1] purposed three estimators and compared them by a short simulation study. Based on the pdf (2) (which is the pdf used by [1] with $\theta^2 = \gamma$) these estimators are

-Maximum Likelihood (ML) estimator

$$\tilde{\gamma}_{\text{ML}} = \frac{n}{\sum_{i=1}^{n} X_i}.$$

-Median-based (Q) estimator

$$\tilde{\gamma}_{\text{Q}} = \hat{m}[\Phi^{-1}(0.75)]^2,$$

where $\hat{m}$ is the sample median.

-Laplace Transform (LT) estimator

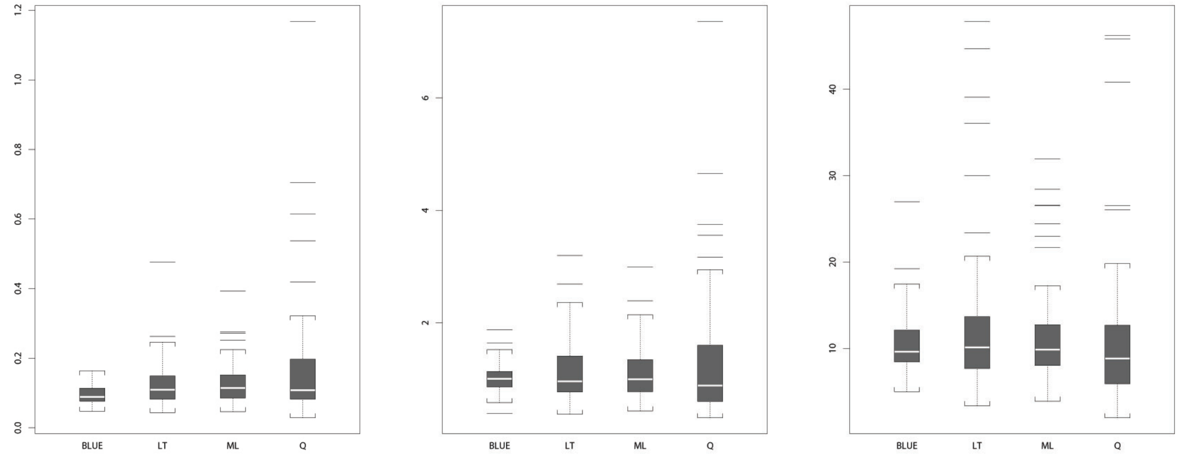
$$\tilde{\gamma}_{\text{LT}} = \frac{1}{2t}[-\ln(\Psi_n(t))]^2$$

where $\Psi_n(t) = \frac{1}{n}\sum_{i=1}^{n} \exp(-tX_i)$ is the empirical Laplace transform and t determines as the unique solution to the equation $\Psi_n(\tilde{t}) = 0.0658$.

They showed that ML and LT estimators are not very different from one another but the Q estimator is less precise. See [1] for more details.

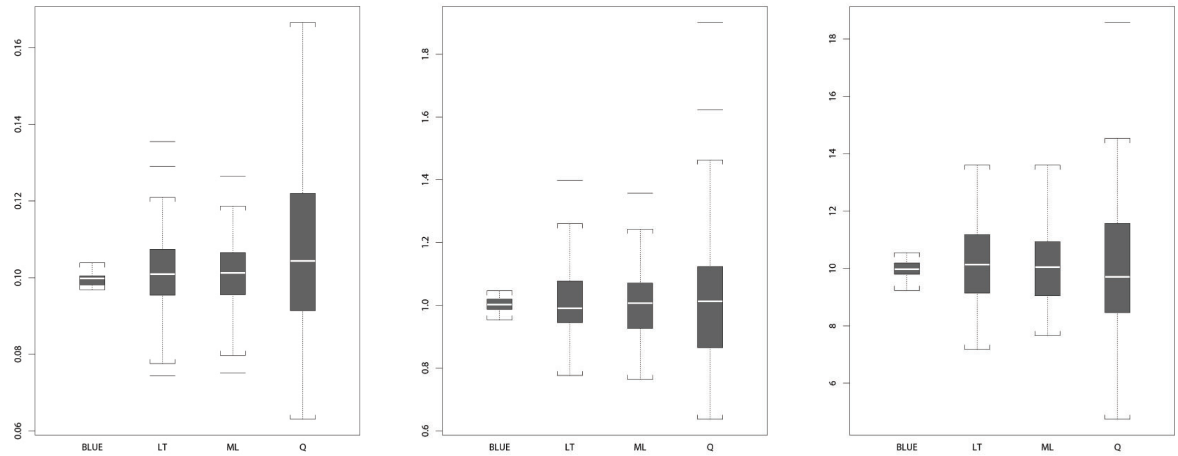
We carry out a simulation study to assess the finite sample performance of the proposed estimator (BLUE) to the estimators provided by [1]. These estimators compared by two methods. First, boxplots of 50 simulated BLUE, LT, ML and Q estimators, for scale parameters $\gamma = 0.1, 1, 10$ and sample sizes $n = 10, 100$, are plotted in Figures 1 and 2. It can be seen that for $n = 10$ and for all values of γ, BLUE is better than the other estimators, especially in small values of γ. When the sample size increase, BLUE has better performance than them.

Next, we compare the MSE of the mentioned estimators for different values of n, $n = 10(1)50$. Graphs of $(\text{MSE}/n)^{0.5}$ for BLUE as well as ML, Q and LT estimators as a function of n, are plotted in Figure 3. We note that MSE of BLUE is equal to its variance which is given by (6), but MSEs of other estimators are approximated by sample MSE through a simulation study (number of iterations is 1000). It can be seen from Figure 3, the BLUE is better than other estimators and its MSE is less than others for different sample sizes. Also, from Figures 1-3, it can be seen that ML and LT estimators are not very different from one another and are more precise than Q estimator, which is indicated by [1].



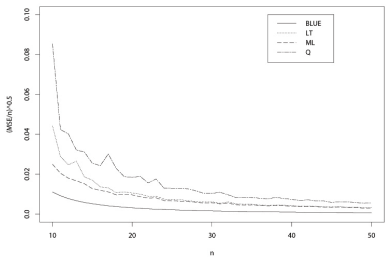
(a) $\gamma = 0.1$ (b) $\gamma = 1$ (c) $\gamma = 10$

Figure 1: Boxplots of BLUE, LT, ML and Q estimators based on 50 simulated data sets of size $n = 10$.

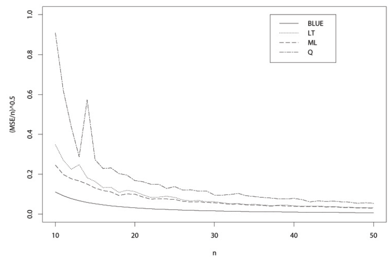


(a) $\gamma = 0.1$ (b) $\gamma = 1$ (c) $\gamma = 10$

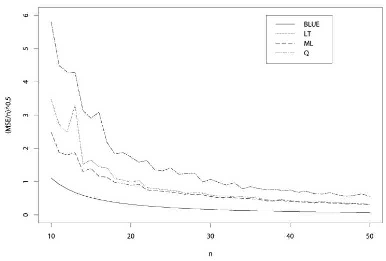
Figure 2: Boxplots of BLUE, LT, ML and Q estimators based on 50 simulated data sets of size $n = 100$.



(a) $\gamma = 0.1$



(b) $\gamma = 1$



(c) $\gamma = 10$

Figure 3: MSEs of BLUE and sample MSE of LT, ML and Q estimators of scale parameter γ with respect to sample size n for $\gamma = 0.1, 1, 10$ in top, middle and bottom, respectively.

Appendix

To prove the Lemma 1, we need the following result.

Let $X \sim$ Lévy(γ) with $\gamma > 0$. Then

$$\lim_{\lambda\to\infty} \lambda^{0.5} P(X > \lambda) = \sqrt{\frac{2\gamma}{\pi}}. \tag{A-1}$$

Proof.

$$\lim_{\lambda\to\infty} \lambda^{0.5} P(X > \lambda) = \lim_{\lambda\to\infty} \lambda^{0.5}(1 - F(\lambda)) = \lim_{\lambda\to\infty} \lambda^{0.5}\Big[1 - 2\Big\{1 - \Phi(\sqrt{\frac{\gamma}{\lambda}})\Big\}\Big].$$

Using Hopital's rule and the fact $\Phi(0) = 0.5$, the result is followed.

Proof of Lemma 1. We use the following equation ([13], [9])

$$P(X_{(k)} > \lambda) = \sum_{j=n-k+1}^{n} (-1)^{j-(n-k+1)} \binom{j-1}{n-k} \sum_{1\le i_1<i_2<\cdots<i_j\le n} P(X_{i_1} > \lambda, \ldots, X_{i_j} > \lambda).$$

Since $X_{i_1}, \ldots, X_{i_j}$ are i.i.d. Lévy(γ), so

$$\begin{aligned}
&\lim_{\lambda\to\infty} \lambda^{0.5(n-k+1)} P(X_{(k)} > \lambda) \\
&= \lim_{\lambda\to\infty} \sum_{j=n-k+1}^{n} (-1)^{j-(n-k+1)} \binom{j-1}{n-k} \lambda^{0.5(n-k+1)} [P(X > \lambda)]^j \sum_{1\le i_1<i_2<\cdots<i_j\le n} 1 \\
&= \lim_{\lambda\to\infty} \sum_{j=n-k+1}^{n} (-1)^{j-(n-k+1)} \binom{j-1}{n-k} \binom{n}{j} \lambda^{0.5(n-k+1)} [P(X > \lambda)]^j \\
&= \binom{n}{n-k+1} \lim_{\lambda\to\infty} [\lambda^{0.5} P(X > \lambda)]^{n-k+1} \\
&\qquad + \sum_{j=n-k+2}^{n} (-1)^{j-(n-k+1)} \binom{j-1}{n-k} \binom{n}{j}
\end{aligned}$$

$$\times \left\{ \lim_{\lambda\to\infty} [\lambda^{0.5} P(X>\lambda)]^{n-k+1} [P(X>\lambda)]^{j-(n-k+1)} \right\} \tag{A-2}$$

From (A-1), $\lim_{\lambda\to\infty}[\lambda^{0.5}P(X>\lambda)]^{n-k+1} = \left(\sqrt{\frac{2\gamma}{\pi}}\right)^{n-k+1}$ and from (1), $\lim_{\lambda\to\infty}[P(X>\lambda)]^{j-(n-k+1)} = 0$ for $n-k+2 \le j \le n$. So, from (A-2) we have

$$\lim_{\lambda\to\infty} \lambda^{0.5(n-k+1)} P(X_k > \lambda) = \binom{n}{n-k+1} \left(\sqrt{\frac{2\gamma}{\pi}}\right)^{n-k+1}.$$

Hence

$$\begin{aligned}
\frac{1}{m} E|X_{(k)}|^m &= \frac{1}{m} \int_0^\infty P(X_{(k)}^m > \upsilon) d\upsilon \\
&= \frac{1}{m} \int_0^\epsilon P(X_{(k)}^m > \upsilon) d\upsilon + \frac{1}{m} \int_\epsilon^\infty P(X_{(k)}^m > \upsilon) d\upsilon \\
&= \int_0^{\epsilon^{\frac{1}{m}}} \lambda^{m-1} P(X_{(k)} > \lambda) d\lambda + \int_{\epsilon^{\frac{1}{m}}}^\infty \lambda^{m-1} P(X_{(k)} > \lambda) d\lambda.
\end{aligned}$$

The first integral is finite and the second is finite if and only if $k < n+1-2m$, which is complete the proof.

References

[1] P. Besbeas, B. J. T. Morgan, Efficient and robust estimation for the one-sided stable distribution of index $\frac{1}{2}$. Statist. Probab. Lett., 66 (2004) 251-257.

[2] Z. Chen, Z. Bai, B. K. Sinha, Ranked set sampling: Theory and applications, Springer, New York, 2004.

[3] T. R. Dell, J. L. Clutter, Ranked set sampling theory with order statistics background, Biometrics 28 (1972) 545-555.

[4] H. Fei, B. K. Sinha, Z. Wu, Estimation of parameters in two-parameter Weibull and extreme-value distributions using ranked set sample, J. Statist. Res. 28 (1994), 149-161.

[5] A. Hatefi, M. Jafari Jozani, D. Ziou, Estimation and classification for finite Mixture models under ranked set sampling, Statistica Sinica 24 (2014) 675-698.

[6] K. Lam, B. K. Sinha, Z. Wu, Estimation of parameters in a two-parameter exponential distribution using ranked set sample, Ann. Inst. Statist. Math., 46 (1994) 723-736.

[7] K. Lam, B. K. Sinha, Z. Wu, Estimation of location and scale parameters of a Logistic distribution using ranked set sample, In: Nagaraja, Sen, Morrison, (Eds), Papers in Honor of Herbert A. David. (1995) 187-197.

[8] G. A. McIntyre, A method for unbiased selective sampling using ranked-set sampling, Austral. J. Agric. Res. 3 (1952) 385-390.

[9] M. Mohammadi, M. Mohammadpour, Existence of order statistics moments of α-stable distributions, Preprint, 2010.

[10] N. Ni Chuiv,B. K. Sinha, Z. Wu, Estimation of the location parameter of a Cauchy distribution using a ranked set sample, Technical Report, University of Maryland Baltimore County, 1994.

[11] N. Ni Chuiv, B. K. Sinha, On some aspects of ranked set sampling in parametric estimation, In: Balakrishnan, Rao (Eds), Handbook of Statistics, 17 (1998) 337-377.

[12] A. Piryatinska, Inference for Lévy models and their application in medicine and statistical physics, Ph.D. Thesis, Department of statistics, Case Western Reserve University, Cleveland, OH, 2005.

[13] G. Samorodnitsky, Extrema of skewed stable process, Stochastic Process. Appl. 30 (1988) 17-39.

[14] E. Scalas, K. Kim, The art of fitting financial time series with Lévy stable distributions, 2006, http://mpra.ub.uni-muenchen.de/336/.

[15] B. K. Sinha, S. Purkayastha, On some aspects of ranked set sampling for estimation of Normal and Exponential parameters, Statist. Decisions 14 (1996) 223-240.

[16] K. Takahasi, K. Wakimato, On unbiased estimates of the population mean based on the sample stratified by means of ordering, Ann. Inst. Statist. Math. 20 (1968) 1-31.

[17] V. V. Uchaikin, V. M. Zolotarev, Chance and Stability: stable distributions and their applications, Utrecht: VSP Press, 1999.

[18] G. Zheng, R. Modarres, A robust estimate of the correlation coefficient for bivariate normal distribution using ranked set, J. Statist. Plann. Inference 136 (2006) 298-309.

Please cite this article using:

Sahar Dorniani, Adel Mohammadpour, Nader Nematollahi, Estimation of the parameter of Lévy distribution using ranked set sampling, AUT J. Math. Comput., 2(1) (2021) 53-60
DOI: 10.22060/ajmc.2020.18499.1037