An Alternative Method for Estimating the Parameters of Log-Cauchy Distribution

MOHAMMAD A. AMLEH[1*], AHMAD AL-NATOOR[2], BAHA' ABUGHAZALEH[2],
[1]Department of Mathematics, Faculty of Science, Zarqa University, Zarqa, JORDAN
[2]Department of Mathematics, Faculty of Science, Isra University, Amman, JORDAN
*Corresponding author

Abstract: - In this paper, we discuss the estimation problem of the location and scale parameters of the log-Cauchy distribution, a member of the super heavy-tailed distributions family. We consider several methods of estimation, including a new percentile method, maximum likelihood estimation and robust estimators. A Monte Carlo simulation experiment is conducted to compare the proposed estimation methods. Further, we illustrate the estimation methods via a real life example.

Key-Words: - Heavy-tailed distributions; Log-Cauchy distribution; Percentile estimators, Maximum likelihood estimators; Robust estimators

Received: October 27, 2022. Revised: December 22, 2022. Accepted: January 19, 2023. Published: February 23, 2023.

1 Introduction

Heavy-tailed distributions play an essential role in the study of rare events, as there can be cases such that the probability of extremely large observations cannot be ignored. Therefore, phenomena describing the occurrence of extreme values with a high relative probability can be modeled by these types of distributions. In many different fields such as computer science, networks, communications, economics, and finance, it is common to find examples of heavy tail datasets. For more details on the heavy-tailed distributions, one may refer to [1] – [5].

The log-Cauchy distribution is one of the heavy-tailed distributions. It is considered a special case of the generalized beta distribution of type II. In fact, it is a special case of the log-student distribution. It can be transformed from Cauchy distribution as follows. If X is a Cauchy distribution, then $Y = e^X$ has a log-Cauchy distribution. Because the expected value and variance of the Cauchy distribution are not defined, the Cauchy distribution is sometimes referred to as the primary example of a pathological distribution. In fact, the moment generating function of the Cauchy distribution has not been defined, see for example, [6] and [7]. Therefore, it is a matter of priority that these exotic properties are achieved in the log-Cauchy distribution.

In the literature concerning the problem of estimating parameters, several estimation techniques have been proposed for the Cauchy distribution, see for example, [8] – [11]. However, it is noteworthy that no attention was paid to estimating the parameters of the log-Cauchy distribution. In this paper, we consider the estimation problem of the parameters of the log-Cauchy distribution. We present a robust alternative method for estimating the required parameters. Two additional methods are presented for comparison purposes. A simulation study is conducted to compare the effectiveness of the estimation techniques and a real dataset is considered for illustrative purposes.

2 log-Cauchy Distribution

The log-Cauchy distribution is sometimes seen as an example of a "super heavy-tailed" distributions, because its tail is considered heavier than the Pareto-type heavy tail. Its tail is decaying logarithmically, see [1]. The probability density function (pdf) of the log-Cauchy distribution is defined as:

$$f(y,\theta,\lambda) = \frac{1}{\pi y}\left(\frac{\lambda}{\lambda^2 + (\ln y - \theta)^2}\right), y > 0, (1)$$

where $\theta \epsilon \mathbb{R}$ is the location parameter and $\lambda > 0$ represents the scale parameter. The cumulative distribution function (cdf) of the log-Cauchy distribution is given by

$$F(y,\theta,\lambda)=\frac{1}{2}+\frac{1}{\pi}tan^{-1}\left(\frac{\ln y-\theta}{\lambda}\right), y>0, \quad (2)$$

The survival and hazard rate functions of the log-Cauchy distribution are expressed as:

$$S(y,\theta,\lambda)=-\frac{1}{2}+\frac{1}{\pi}tan^{-1}\left(\frac{ln\, y-\theta}{\lambda}\right), y>0,$$

and

$$h(y,\theta,\lambda)=\frac{\lambda}{y(\lambda^2+(ln\, y-\theta)^2)\left(tan^{-1}\left(\frac{ln\, y-\theta}{\lambda}\right)-\frac{\pi}{2}\right)}, y>0,$$

respectively.

3 Estimating the Parameters of the log-Cauchy Distribution

In this section, we discuss the problem of estimating the location and scale parameters of the log-Cauchy distribution. Three methods of estimation are presented.

3.1 Percentiles Estimators

Percentiles play an essential role in statistical inference. They are used recently in estimating parameters, see [12]. The technique is similar to the moment method estimation, but instead of equating population moments to the sample moments, it is based on equating population percentiles to the sample percentiles and then solving the obtained equations simultaneously.

In this context, we propose new percentile estimators based on the popular quartiles as follows. Assume that $Y_1, Y_2, \dots, Y_n$ is a random sample of size n from log-Cauchy distribution. The quantile function of the log-Cauchy distribution can be written as:

$$q(t)=e^{\theta+\lambda\tan\left(\pi\left(t-\frac{1}{2}\right)\right)}, 0<t<1, \quad (3)$$

Therefore, if Y is a log-Cauchy distributed, based on Eq. (3), the median Y is given by:

$$M(Y)=e^{\theta}. \quad (4)$$

The lower and upper quartiles of Y are given as:

$$Q_1(Y)=e^{\theta-\lambda}, \quad (5)$$

and

$$Q_3(Y)=e^{\theta+\lambda}, \quad (6)$$

respectively. The new approach is based on equating the population quartiles in Eqs. (4) to (6) to the sample quartiles as follows:

$$e^{\theta}=Y_m$$

$$e^{\theta-\lambda}=Y_{q_1} \quad (7)$$

$$e^{\theta+\lambda}=Y_{q_3},$$

where Y_m is the sample median, Y_{q_1} and Y_{q_3} represent the sample lower and upper quartiles, respectively. Solving the system of equations in (7) gives:

$$\hat{\theta}_{PE}=\log(Y_m), \quad (8)$$

and

$$\hat{\lambda}_{PE}=\frac{1}{2}\log\left(\frac{Y_{q_3}}{Y_{q_1}}\right). \quad (9)$$

Eqs. (8) and (9) are the percentiles estimators required to estimate the parameters θ and λ, respectively.

3.2 Maximum Likelihood Estimation

If $Y_1, Y_2, \dots, Y_n$ is a random sample of size n taken from a log-Cauchy distribution, then the likelihood function, based on this sample, is given by:

$$L(\theta,\lambda)=\frac{\lambda^n}{\pi^n}\prod_{i=1}^{n}\left(\frac{1}{y_i(\lambda^2+(\log y_i-\theta)^2)}\right). \quad (10)$$

The associated log-likelihood function may be expressed as:

$$l(\theta,\lambda)=nlog\,\lambda-\sum_{i=1}^{n}\log y_i-\sum_{i=1}^{n}\log(\lambda^2+(\log y_i-\theta)^2). \quad (11)$$

The maximum likelihood estimators (MLEs) of θ and λ can be obtained by maximizing the log-likelihood function in Eq. (11). Consequently, the likelihood equations are given by:

$$\frac{\partial l}{\partial\theta}=\sum_{i=1}^{n}\frac{2(\log y_i-\theta)}{\lambda^2+(\log y_i-\theta)^2}=0. \quad (12)$$

$$\frac{\partial l}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} \frac{2\lambda}{\lambda^2 + (\log y_i - \theta)^2} = 0. \quad (13)$$

The estimation process, through Eqs. (12) and (13), cannot be obtained in closed form. Accordingly, equations (12) and (13) will be solved simultaneously using a numerical technique such as the Newton-Raphson method. The resulting MLEs of the parameters θ and λ will be denoted by $\hat{\theta}_{MLE}, \hat{\lambda}_{MLE}$, respectively.

3.3 Robust Estimators

The data used to estimate the parameters of heavy-tailed distributions are generally contaminated with extreme values, which are called outliers. Outliers may be considered in related applications as rare events. By contrast, observations that are not extreme are called inliers. An estimator is said to be robust when these outliers do not affect the estimates of the parameters, more details on robust estimators can be found in [13] and [14]. Now, as mentioned above, the log-Cauchy distribution is one of the heavy-tailed distributions. Therefore, robust estimators of the parameters are proposed in the literature. Hence, based on a random sample $Y_1, Y_2, \ldots, Y_n$ taken from log-Cauchy distribution, Olive [15] suggested the following robust estimators:

$$\begin{aligned} \hat{\theta}_{RE} &= median(\log Y_1, \log Y_2, \ldots, \log Y_n) \\ \hat{\lambda}_{RE} &= MAD(\log Y_1, \log Y_2, \ldots, \log Y_n), \end{aligned} \quad (14)$$

where $MAD(x_1, x_2, \ldots, x_n)$ represents the median absolute deviation of x_i about the median of $x_i, i = 1, 2, \ldots, n$. Clearly, if the sample size n is odd, it can be shown that: $\hat{\theta}_{RE} = \hat{\theta}_{PE}$.

4 Simulation and Real Example

Now, in this section, we perform a simulation experiment for computing the estimates of the parameters θ and λ based on the techniques discussed in Section 3. A real dataset is considered to illustrate the estimation methods.

4.1. Simulation Experiment

Here, we conduct an intensive Monte Carlo simulation experiment for evaluating the suggested estimators. The performance of the considered estimators is measured in terms of the bias and the mean square error (MSE) of the estimators, which are expressed for any estimator $\hat{\alpha}$ of a parameter α as:

$$bias(\hat{\alpha}) = \frac{1}{m} \sum_{j=1}^{m} \left(\hat{\alpha}_j - \alpha\right),$$

and

$$bias(\hat{\alpha}) = \frac{1}{m} \sum_{j=1}^{m} \left(\hat{\alpha}_j - \alpha\right)^2,$$

respectively. Here, the Monte Carlo simulation is conducted based on different sample sizes and parameter values. For this, we generate random samples of log- Cauchy distribution by considering the following schemes:

Scheme 1: $\theta = 0.5, \lambda = 0.75$
Scheme 2: $\theta = 2, \lambda = 1.5$
Scheme 3: $\theta = 3, \lambda = 2$

Samples from log-Cauchy distribution were randomly generated under these schemes with 1000 replications of the simulation process. Using these random samples, estimation biases and MSEs of the estimators are obtained. The results are presented in Tables 1 to 3.

Based on these tables, we observe the following remarks. The biases of the three estimators are generally small, which indicates the good performance of the estimators in this sense. By considering the MSE as an optimal criterion, it has been observed that PEs are highly competitive to MLEs and outperform REs in most of the considered cases. It can be seen that as n increases, the MSEs for all estimators decrease. Moreover, it is noticeable that the MSEs of the three estimators are close to each other.

Table 1: **Scheme 1**: ,
Biases & MSEs of the estimates of

n	PE		MLE		RE	
	Bias	MSE	Bias	MSE	Bias	MSE
10	0.0716	0.1569	0.0027	0.1679	0.0096	0.1937
20	-0.0067	0.0923	-0.0175	0.0797	-0.0126	0.0902
30	0.0327	0.0447	0.0241	0.0337	0.0310	0.0446
50	-0.0046	0.0239	-0.0070	0.0203	-0.0051	0.0241
75	0.0110	0.0138	0.0026	0.0121	0.0110	0.0138
100	0.0083	0.0120	0.0061	0.0107	0.0081	0.0120
Biases & MSEs of the estimates of						
n	PE		MLE		RE	
	Bias	MSE	Bias	MSE	Bias	MSE
10	0.0391	0.2763	-0.0208	0.1386	0.0511	0.1976
20	0.0366	0.0798	0.0005	0.0631	0.0564	0.0941
30	0.0201	0.0428	0.0112	0.0383	0.0444	0.0443
50	-0.0075	0.0282	-0.0108	0.0248	-0.0033	0.0311
75	-0.0040	0.0183	0.0009	0.0167	0.0103	0.0219
100	-0.0201	0.0119	-0.0213	0.0104	-0.0166	0.0124

Table 2: Scheme 2: ,
Biases & MSEs of the estimates of

n	PE		MLE		RE	
	Bias	MSE	Bias	MSE	Bias	MSE
10	0.1291	0.6887	0.0809	0.6768	0.0610	0.6765
20	-0.0052	0.3092	0.0024	0.2702	-0.0246	0.3125
30	0.0918	0.241	0.0599	0.2079	0.0860	0.239
50	-0.0049	0.0933	-0.0037	0.0655	-0.0068	0.0931
75	0.1140	0.0738	0.1053	0.069	0.1140	0.0738
100	-0.0526	0.0562	-0.0431	0.0514	-0.0532	0.0563
Biases & MSEs of the estimates of						
n	PE		MLE		RE	
	Bias	MSE	Bias	MSE	Bias	MSE
10	0.0216	0.7123	0.0044	0.7244	0.1780	1.115
20	0.0624	0.3398	0.0034	0.2715	0.0598	0.3343
30	0.1142	0.2322	0.0749	0.1856	0.1258	0.2513
50	0.0447	0.0882	0.0174	0.0789	0.0627	0.0971
75	-0.0633	0.0495	-0.0433	0.0426	-0.0503	0.0728
100	-0.0769	0.0421	-0.1014	0.0430	-0.0740	0.0425

Table 3: Scheme 3:,.

n	PE		MLE		RE	
	Biases & MSEs of the estimates of					
	Bias	MSE	Bias	MSE	Bias	MSE
10	-0.1953	1.3399	-0.2864	1.4012	-0.3018	1.4823
20	-0.0095	0.5562	-0.0001	0.4847	-0.0321	0.5661
30	0.091	0.2606	0.012	0.2673	0.0815	0.2556
50	-0.0563	0.2028	-0.0743	0.1167	-0.0607	0.2035
75	-0.1618	0.1751	-0.1565	0.1648	-0.1618	0.1751
100	0.0456	0.1181	0.0192	0.0867	0.0433	0.1178
	Biases & MSEs of the estimates of					
n	PE		MLE		RE	
	Bias	MSE	Bias	MSE	Bias	MSE
10	-0.0409	0.6044	-0.0566	0.6292	0.1491	0.7862
20	0.0807	0.5808	-0.0355	0.3555	0.1077	0.519
30	0.0838	0.2574	0.0979	0.2086	0.1405	0.2962
50	0.1055	0.1462	0.1169	0.1984	0.1584	0.2115
75	0.0649	0.1137	0.0073	0.0900	0.0591	0.1259
100	-0.1071	0.0887	-0.1113	0.0686	-0.0688	0.0972

4.2. Real Example

To clarify the estimation methods discussed in this paper, we consider an example of real data that was used by Alzaatreh [16]. It accounts for 157 of the national consumer price index in Brazil. The data are shown in Table 4. Since the log-Cauchy distribution is defined on positive real numbers, for each sample point $x_i, i = 1, 2, \ldots, 157$; we take the exponential value e^{x_i}. To test the goodness- of-fit of the transformed data to the log-Cauchy distribution, Kolmogorov-Smirnov (K-S) test is used. The K-S statistic of the distance between the empirical distribution and the fitted one, based on estimates $\hat{\theta}_{MLE}, \hat{\lambda}_{MLE}$, is 0.12057 and the corresponding p-value is 0.2042. Therefore, it is appropriate to fit the transformed data using the log-Cauchy distribution. To see the accuracy of the log-Cauchy distribution under the estimation methods presented in this study, the true CDF of the data is plotted in Fig. 1, along with the estimated CDFs. The obtained estimates of the parameters θ and λ based on the considered methods are displayed in Table 5. It can be observed that the values of the estimates are close to each other.

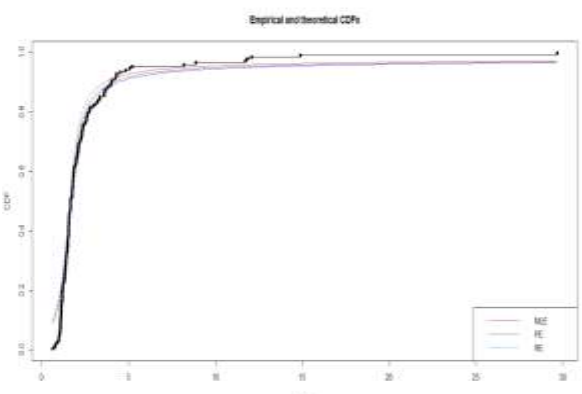

Fig. 1: The empirical cdf (dots) and the estimated cdfs based on the three methods.

Table 4: National index of the prices of consumers in Brazil

0.69	0.13	0.17	2.47	0.57	0.55
0.97	0.12	0.50	2.70	0.84	0.74
0.43	0.27	0.73	3.39	0.48	0.07
0.30	0.23	0.50	1.57	0.49	0.05
0.25	0.38	0.40	0.83	0.77	0.47
0.59	0.40	0.41	0.86	0.55	1.28
0.32	0.54	0.57	1.15	0.29	1.29
0.31	0.58	0.39	0.61	0.16	0.65
0.26	0.15	0.53	0.09	0.43	0.42
0.26	0.00	0.54	0.65	1.21	2.46
-0.18	0.49	0.18	0.34	0.29	2.18
0.11	0.54	0.35	0.38	0.71	2.10
-0.31	0.85	0.11	0.02	1.46	2.49
-0.49	0.57	0.60	0.50	1.65	0.15
1.62	1.44	0.37	0.62	1.39	0.72
0.44	0.03	0.39	0.31	0.30	0.45
0.42	-0.11	0.82	1.07	-0.05	1.01
0.49	0.70	0.18	0.74	0.09	0.29
0.62	0.91	0.04	1.29	0.13	0.10
0.42	0.73	-0.06	0.94	0.05	-0.03
0.43	0.44	0.99	0.44	0.61	0.81
0.16	0.57	1.38	0.79	0.74	0.33
-0.02	0.86	1.37	1.11	0.94	1.28
0.11	0.44	1.46	0.60	0.96	0.93
-0.07	0.17	0.68	1.20	1.51	1.17
-0.28	0.15	0.45	1.33	1.40	1.02
0.39					

5 Conclusion

In this study, we have considered the estimation problem of the parameters of log-Cauchy distribution, one of the super heavy-tailed distributions. Three estimators are addressed including, a new alternative estimator based on percentiles, maximum likelihood estimator and a robust estimator.

We have compared the performance of the estimators using a Monte Carlo simulation experiment in terms of the biases and MSEs for different sample sizes and parameter values. The alternative estimators PEs are recommended as they are computationally attractive and have good performances based on the bias and MSE criteria.

In this context, the proposed alternative method can be used for estimating parameter for a wide range of statistical distributions, which may be discussed in future studies.

Table 5: estimators of θ and λ

Method	$\hat{\theta}$	$\hat{\lambda}$
Percentile Method (PE)	0.500	0.295
MLE	0.477	0.263
Robust Estimation (RE)	0.500	0.320

References:

[1] Alves, I. F., Haan L., and Neves C., A test procedure for detecting super-heavy tails. *Journal of Statistical Planning and Inference*, 139(2), 2009, p 213-227.

[2] Bondesson, Lennart. On the Lévy measure of the log-normal and the log-Cauchy distributions. *Methodology and Computing in Applied Probability* 4(3), 2002, p 243-256.

[3] Holger D., On smooth statistical tail functionals, *Scandinavian Journal of Statistics*, 25(1), 1998, p 187-210.

[4] Amleh, M. A., and Raqab, M. Z., Inference in simple step-stress accelerated life tests for Type-II censoring Lomax data. *Journal of Statistical Theory and Applications*, 20(2), 2021, 364-379.

[5] Amleh M. A., Baha' Abughazaleh, Ahmad Al-Natoor, Conformable fractional Lomax probability distribution, *J. Math. Comput. Sci.*, 12 (2022), Article ID: 130.

[6] Johnson, N. L., Kotz S. and Balakrishnan N., *Continuous univariate distributions*, volume 1.John Wiley & sons, 1994.

[7] Casella, G., and Berger, R. L. *Statistical inference*, 2nd edition, Pacific Grove, CA, Duxbury, 2002.

[8] Barnett, V. D. Order statistics estimators of the location of the Cauchy distribution, *Journal of the American Statistical Association*, 61(316) 1966, p 1205-1218.

[9] Ferguson, Thomas S., Maximum likelihood estimates of the parameters of the Cauchy distribution for samples of size 3 and 4, *Journal of the American Statistical Association*, 73(361), 1978, p 211-213.

[10] Bai, Z. D., and J. C. Fu. On the maximum- likelihood estimator for the location parameter of a Cauchy distribution, *Canadian Journal of Statistics*, 15(2), 1987, p 137-146.

[11] Zhang, Jin. A highly efficient L-estimator for the location parameter of the Cauchy distribution, *Computational statistics*, 25(1), 2010, p 97-105.

[12] Bhatti, Sajjad Haider, et al. Efficient estimation of Pareto model: Some modified percentile estimators, *PLOS ONE*, 13(5), 2018, e0196456.
[13] Huber, P. J. *Robust statistics.* In International encyclopedia of statistical science, Springer, Berlin, Heidelberg, 2011.
[14] Al-Athari F. A. and Amleh M. A., A comparison between least trimmed of squares and MM-estimation in linear regression using simulation technique, *in Proceedings of the International Arab Conference on Mathematics and Computations*, Zarqa University, Zarqa, Jordan, May 2016
[15] Olive, D. J., *Applied robust statistics*, Southern Illinois University, 2008.
[16] Alzaatreh, A., An alternative to the Cauchy distribution, *MethodsX*, 6, 2019, p: 938-952

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors completed all aspects of the study through joint work by proposing the new method and comparing it with the previous methods, in addition to work out the simulation experiments and the real data analysis using R language. The final version has been read and approved by all authors.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.