Int. J. Nonlinear Anal. Appl. 14 (2023) 1, 451–458
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2022.6796

The effect of extreme values on Lomax distribution estimators with simulation

Aseel Noori Saleh

Department of Studies, Planning and Follow-up, Ministry of Higher Education and Scientific Research, Iraq

(Communicated by Madjid Eshaghi Gordji)

Abstract

The significant impact of the extreme values on the estimations of the statistical distributions leads to the fact that the adoption of these estimations in the estimation of many statistical functions such as (reliability, survival) and others is not strictly acceptable because of the high difference between the real and estimated values of these functions. The research included comparing several estimation methods for the Lomax distribution parameter according to different values of the sample size, the value of the distribution parameter and the estimation method. The simulation results showed the effect of the estimator values on the sample size, the value of the distribution parameter and the estimation method. Other estimation methods can be relied upon, as well as other statistical distributions to know the effect of extreme values on them.

Keywords: Lomax distribution, Maximum Likelihood Estimation Method, Moment Method, Mean Square Error, Extreme Values, Simulation Experiments
2020 MSC: Primary 90C33, Secondary 26B25

1 Introduction

The Lomax distribution has many applications, especially in the fields of management sciences, economics and applied sciences; It is one of the statistical distributions that are Heavy tail distribution. As a result of the importance of the distribution of Lomax, many kinds of research were presented, including the research presented by Kilany [3] in which the weighted Lomax distribution of the second type was presented with the presentation of the theoretical aspects of the distribution, including the probability density function and the survival function, and the distribution parameters were estimated according to the methods of(moment and maximum likelihood estimation) methods, real data represents patients with bladder cancer, simulation data and real data numerical results showed the ability of the estimation methods to provide estimates of survival functions closer to the real [3]. The research presented by Hassan Alsuhabi et. al in [1], in which the distribution of four parameters consisting of the Lomax distribution and the Whipple distribution was presented and the estimation methods were presented (MLE method, Maximum product spacing and Bayesian estimation) to estimate distribution parameters according to a number of simulation experiments according to sample size and value of different parameters.

*Corresponding author
Email address: asselassel908@gmail.com (Aseel Noori Saleh)

Received: January 2022 *Accepted:* March 2022

The research also included the application of estimation methods on two sets of real data representing Covid-19 patients in the United Kingdom and the United States of America.

The application results showed the ability of the different estimation methods to provide estimates of the distribution parameters as close to true [1].

In the presented research many simulation experiments with various (simulation methods, sample size (including various extreme values) and values of parameter of distribution), compare mean square error for each simulation experiment show that the estimators depend upon (simulation methods, sample size, extreme values and values of parameter of distribution).

2 Problem of Research

The extreme values can affect the estimation methods and this leads to providing estimations far from the true values.

3 Aim of Research

The research aims to show the effect of extreme values on the estimations of the parameters of the Lomax distribution according to (estimation methods, the sample size and the value of the distribution parameter).

4 Properties of the Lomax Distribution

Lomax distribution has the following probability density function (*pdf*) [5]

$$f(x,\alpha,\beta) = \frac{\alpha}{\beta}\,[1+\frac{x}{\beta}]^{-(\alpha+1)} \qquad (4.1)$$
$$0 \le x \le \infty \quad ,\alpha > 0\ ,\beta > 0$$

And the cumulative distribution function (*cdf*)

$$F(x,\alpha,\beta) = 1 - \left[1+\frac{x}{\beta}\right]^{-\alpha}, \qquad (4.2)$$

and the mean will be

$$\mu = \begin{cases} \frac{\beta}{\alpha-1} & \text{for} \quad \alpha > 1 \\ \text{undefined} & \text{for other wise} \end{cases} \qquad (4.3)$$

The median will be

$$\mu_e = \beta\left(\sqrt[\alpha]{2}\ -1\right). \qquad (4.4)$$

The mode is $\mu_0 = 0$, the variance is

$$V = \begin{cases} \frac{\beta^2\alpha}{(\alpha-1)^2(\alpha-2)} & for \quad \alpha > 1 \\ \infty & 1 \le \alpha \le 2 \\ \text{undefined} & \text{for other wise} \end{cases} \qquad (4.5)$$

The skewness is

$$S_k = \frac{2(\alpha+1)}{(\alpha-1)}\ \sqrt{\frac{\alpha-2}{\alpha}} \qquad \alpha > 3 \qquad (4.6)$$

The kurtosis

$$\frac{6(\alpha^3+\alpha^2-6\alpha-2)}{\alpha(\alpha-3)(\alpha-4)} \qquad \alpha > 4 \qquad (4.7)$$

5 Estimation method

There are a number of estimation methods for the parameters of the Lomax distribution, including:

5.1 the Maximum Likelihood Estimation Method (MLE)

This method depends on the maximum Likelihood function, which assumes a sample of size (n) and each of them has a Lomax distribution according to the parameters (α, β).

The probability density function for each individual will be [2, 8]

$$f(x, \alpha, \beta) = \frac{\alpha}{\beta}\,[1 + \frac{x}{\beta}]^{-(\alpha+1)} \tag{5.1}$$

$$L = \prod_{i=1}^{n} f(x_i, \alpha, \beta)$$

$$\mathrm{L} = \prod_{i=1}^{n} \frac{\alpha}{\beta}\,[1 + \frac{x_i}{\beta}]^{-(\alpha+1)}$$

$$L = [\frac{\alpha}{\beta}]^{n} \prod_{i=1}^{n} \left[1 + \frac{x_i}{\beta}\right]^{-(\alpha+1)}. \tag{5.2}$$

Take the logarithm $(l = Log(L))$

$$l = n\log\left(\frac{\alpha}{\beta}\right) - (\alpha+1)\sum_{i=1}^{n} Log\left(1 + \frac{x_i}{\beta}\right). \tag{5.3}$$

Take the partial derivative of the parameter (α),

$$\frac{\partial L}{\alpha} = \frac{n}{\widehat{\alpha}} - \sum_{i=1}^{n} \log\left(1 + \frac{x_i}{\beta}\right). \tag{5.4}$$

Set the derivative equal to zero

$$\frac{n}{\widehat{\alpha}} - \sum_{i=1}^{n} Log\left(1 + \frac{x_i}{\beta}\right) = 0. \tag{5.5}$$

The estimator will be

$$\widehat{\alpha}_{MLE} = \frac{n}{\sum_{i=1}^{n} Log(1 + \frac{x_i}{\beta})}. \tag{5.6}$$

The partial derivation of the parameter (β)

$$\frac{\partial L}{?\beta} = \frac{n}{\mathrm{A}}\acute{\mathrm{A}} - \frac{n}{\mathrm{B}} - \mathrm{A}\acute{\mathrm{B}} - \acute{\mathrm{A}}\mathrm{B} - \acute{\mathrm{B}} \tag{5.7}$$

with

$$\mathrm{A} = \frac{n}{\mathrm{B}} \qquad , \acute{\mathrm{A}} = \frac{n\acute{\mathrm{B}}}{\mathrm{B}^2}$$

$$\mathrm{B} = \sum_{i=1}^{n} Log\left(1 + \frac{x_i}{\beta}\right) \qquad , \acute{\mathrm{B}} = \frac{-1}{\beta} \sum_{i=1}^{n} (\frac{x_i}{\beta + x_i}) \tag{5.8}$$

Setting ($\frac{\partial L}{?\beta} = 0$) and using Newton-Raphson method, then

$$\beta_{n+1} = \beta_n - \frac{g(\beta)}{g\acute{(}\beta)} \tag{5.9}$$

with

$$g(\beta) = \frac{n}{\text{A}}\acute{\text{A}} - \frac{n}{\text{B}} - \text{A}\acute{\text{B}} - \acute{\text{A}}\text{B} - \acute{\text{B}} \tag{5.10}$$

$$g\acute{(}\beta) = \frac{n\text{A}''\text{A} - \acute{\text{A}}\acute{\text{A}}}{\text{A}^2} + \frac{n}{\beta^2} - \text{AB}'' - 2\acute{\text{A}}\text{B} - \text{A}''\acute{\text{B}} - \text{B}'' \tag{5.11}$$

$$\text{A}'' = -n\frac{\text{B}''\text{B}^2 - 2\text{B}\acute{\text{B}}^2}{[\text{B}^2]^2} \tag{5.12}$$

$$\text{B}'' = \sum_{i=1}^{n} \frac{x_i[2\beta + x_i]}{\beta^2[\beta + x_i]^2} \tag{5.13}$$

5.2 the Moment Estimation (ME)

The moment sample will be [5]

$$m_1 = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x} \tag{5.14}$$

$$m_2 = \frac{\sum_{i=1}^{n} x_i^2}{n} \tag{5.15}$$

and

$$E(X) = \frac{\beta}{\alpha - 1} \tag{5.16}$$

$$E\left(X^2\right) = \frac{2\beta^2}{(\alpha-1)^2(\alpha-2)}. \tag{5.17}$$

By tacking ($E(X) = m_1$) and ($E\left(X^2\right) = m_2$) we get

$$\widehat{\alpha}_{MOM} = \frac{2n\overline{x}^2}{\sum_{i=1}^{n} x_i^2} + 2 \tag{5.18}$$

$$\widehat{\beta}_{MOM} = \overline{x}\left(\widehat{\alpha}_{MOM} - 1\right) \tag{5.19}$$

5.3 the Shrinkage Estimation (SE)

The search for the estimator of the contraction is by finding a match between two or more estimators so that the new estimator is closer to the distribution parameter than the first or second method or both together, and so on.

The shrinkage estimator is based on the following general formula [9]

$$\widehat{\alpha}_{SE} = p\widehat{\alpha}_{MLE} + (1-p)\,\widehat{\alpha}_{MOM} \tag{5.20}$$

$$\widehat{\beta}_{SE} = p\widehat{\beta}_{MLE} + (1-p)\,\widehat{\beta}_{MOM} \tag{5.21}$$

$$0 \le p \le 1. \tag{5.22}$$

It is clear that when ($p = 0$) the ($\widehat{\alpha}_{SE}$) estimator will be equal to ($\widehat{\alpha}_{MOM}$), and if ($p = 1$) then the ($\widehat{\alpha}_{SE}$) estimator will be equal to ($\widehat{\alpha}_{MLE}$), when ($p = 0$) the ($\widehat{\beta}_{SE}$) estimator will be equal to ($\widehat{\beta}_{MOM}$), and if ($p = 1$) then the ($\widehat{\beta}_{SE}$) estimator will be equal to ($\widehat{\beta}_{MLE}$).

To get the value of (p) that leads to obtaining the optimum shrinkage estimator, the Newton-Raphson method is followed and successive substitutions are made to obtain the shrinkage estimator.

6 Simulation Experiments

A number of simulation experiments were carried out according to each Sample size ($n = 30, 50, 100$) The value of the first parameter ($\alpha = 0.25, 0.50, 0.75$) The value of the second parameter ($\beta = 1, 1.5, 2$) As for pollution, it was made to be (4.3) items for a sample size ($n = 30$) and (4.5) when ($n = 50$) and (5.3) vocabulary at ($n = 100$) so that the polluted individual generated by using randomized function within the range $[L_R\ , U_R]$ [4, 6] such that

$$L_R = \overline{x} + 4S \tag{6.1}$$

$$U_R = \overline{x} + 7S \tag{6.2}$$

with
($\overline{x}$ *and* S) represent the (*Sample mean and slandered deviation respectively*).

The generation function of random sample distributed as(Lomax Distribution) with (α, β). By using the following cumulative distribution function for (Lomax distribution)

$$F\left(x, \alpha, \beta\right) = 1 - \ [1 + \frac{x}{\beta}]^{-\alpha}.$$

Put (R) instead of ($F\left(x, \alpha, \beta\right)$) so that ($R$) represents the random generation function of the uniform distribution within the period [0, 1] then

$$R = 1 - \ \left[1 + \frac{x}{\beta}\right]^{-\alpha} \tag{6.3}$$

Then

$$x = \beta\left[\left[\frac{1}{1 - R}\right]^{\frac{1}{\alpha}} - 1\right] \tag{6.4}$$

For comparing estimators stage each simulation experiment was repeated (1000) times and the mean square error will be [7]

$$MSE_{\widehat{\theta}} = \frac{\sum_{j=1}^{iter}\left[\widehat{\theta} - \theta\right]^2}{iter} \tag{6.5}$$

such that

$$\widehat{\theta}\ \textit{represent the estimator},\ \theta\ \ \textit{represent the real value}.$$

7 Experimental Results

After executing the simulation experiments, the following tables and figures were obtained

From table (1) and Figure (1) we can see that the estimator for the first parameter was affected by (sample size, parameter value and the estimation method) and the best estimation method with minimum mean square error value for each experiment was (MLE with (5) times), (MOM with (18) times) and (SE with (4) times)

From table (2) and Figure (2) we can see that the estimator for the first parameter was affected by (sample size, parameter value and the estimation method) and the best estimation method with minimum mean square error value for each experiment was (MLE with (11) times), (MOM with (13) times) and (SE with (3) times)

8 Suggestions and Conclusions

The simulation results from tables and graphs that formed many Suggestions and Conclusions which are:

1- There is an effect of outliers on the results of the estimators and their mean square error.

Table 1: The Estimators and The mean Square Error with each estimation method for the first parameter

β	α	n	The Estimator			The Mean Square Error			*Best*
			MLE	MOM	SE	MLE	MOM	SE	
0.5	0.25	20	0.2341889	0.2643635	0.2417242	0.0057223	0.0053928	0.0071103	2
		50	0.1879797	0.2514573	0.2249263	0.0143941	0.0016084	0.0080155	2
		100	0.247132	0.2569520	0.2571440	0.0081360	0.0045919	0.0059362	2
	0.5	20	0.3340700	0.5130598	0.42200	0.033353	0.0081658	0.0245789	2
		50	0.5897039	0.5091909	0.5460527	0.0148274	0.0005969	0.006002	2
		100	0.4239087	0.5093739	0.4689650	0.0064598	0.001835	0.0109141	2
	0.75	20	0.7355191	0.757760	0.7495014	0.0074942	0.0036474	0.0045816	2
		50	0.7936008	0.7545883	0.7730369	0.0102597	0.0081877	0.0084724	2
		100	0.5374718	0.7511752	0.6391405	0.0541696	0.0067567	0.0260570	2
1	0.25	20	0.31760	0.264433	0.2866781	6.119E-05	0.00573	0.00700	1
		50	0.2455842	0.2514415	0.2486601	0.0059174	0.0092914	0.0053774	3
		100	0.1764893	0.2576576	0.2150409	0.0091360	0.0039928	0.0056447	2
	0.5	20	0.659230	0.5016849	0.5856067	0.0293631	0.0073982	0.0213377	2
		50	0.3669060	0.5031323	0.4302963	0.020614	0.0086607	0.0119056	2
		100	0.5271763	0.5082220	0.5182500	0.0042609	0.0052408	0.008483	1
	0.75	20	0.5848818	0.7521454	0.6694737	0.0366200	0.0051017	0.0208145	2
		50	0.7512596	0.7567662	0.7548171	0.0060343	0.0005256	0.0012788	2
		100	0.7230489	0.755869	0.7331859	0.0080809	0.008904	0.0038194	3
1.5	0.25	20	0.2041172	0.2574050	0.2355762	0.0045197	0.0054457	0.0087335	1
		50	0.2702878	0.2541997	0.2680779	0.0021915	0.0090346	0.0095849	1
		100	0.1937548	0.2504231	0.2214374	0.0080182	0.0079057	0.0077132	3
	0.5	20	0.5164819	0.5053133	0.5074642	0.007493	0.0054062	0.0018135	3
		50	0.3730227	0.5092749	0.4378021	0.0273640	0.0091286	0.0180852	2
		100	0.3566345	0.5021821	0.4290277	0.0251185	0.0081606	0.0198153	2
	0.75	20	0.9769343	0.7504786	0.8615377	0.0581288	0.0039960	0.0276584	2
		50	0.8101540	0.7572458	0.7769053	0.00691	0.0080383	0.0100038	1
		100	0.565394	0.7544671	0.6610177	0.0419244	0.0002927	0.0229737	2

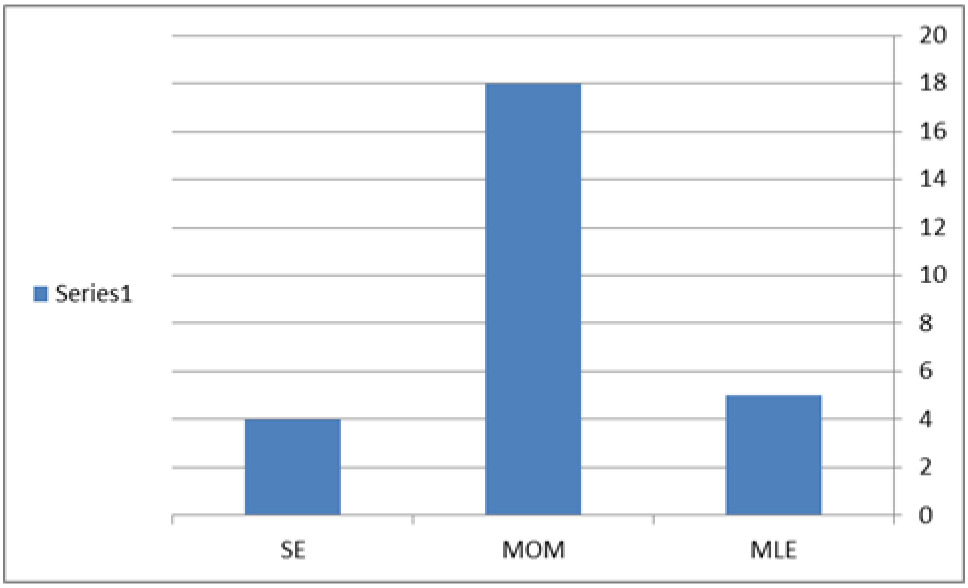

Figure 1: the number of best estimation method for the first parameter

2- The estimation method is affected by both the sample size and the value of the distribution parameter.

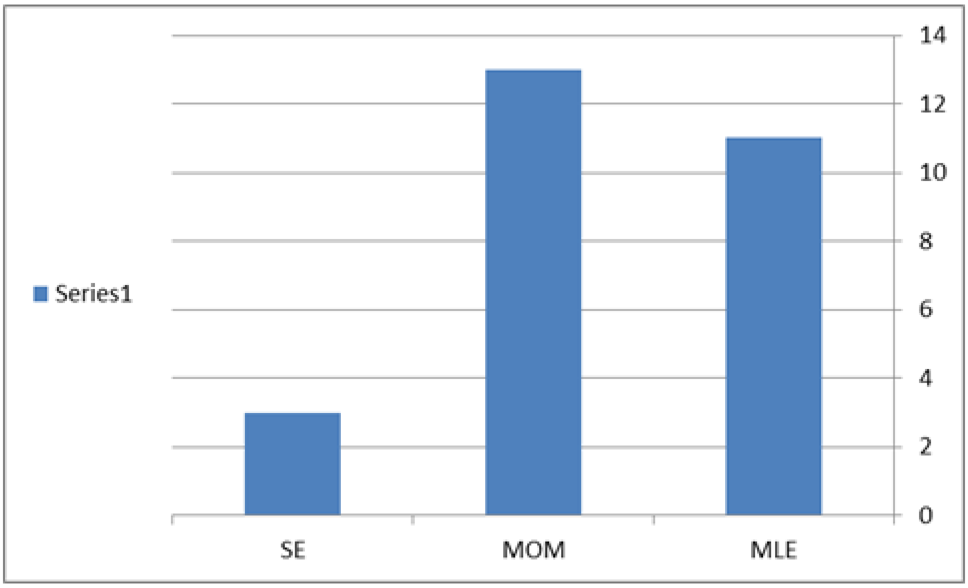

Figure 2: the number of best estimation method for the second parameter

Table 2: The Estimators and The mean Square Error with each estimation method for the second parameter

β	α	n	The Estimator			The Mean Square Error			*Best*
			MLE	*MOM*	*SE*	*MLE*	*MOM*	*SE*	
0.5	0.25	20	0.49947813	0.51052398	0.504600527	0.006391489	0.006705596	0.00689845	1
		50	0.50137554	0.50948745	0.500534899	0.002590575	0.009085542	0.00541473	1
		100	0.50195546	0.50827818	0.500111919	0.004640617	0.007434782	0.00300806	3
	0.5	20	0.97576093	0.503566094	0.73289599	0.225776353	0.006446238	0.117327	2
		50	0.50097653	0.502448097	0.506606038	0.003399028	0.003319373	0.00952708	2
		100	0.50689822	0.50170319	0.508087398	0.005183525	8.72373E-06	0.00395131	2
	0.75	20	0.49985591	0.50713987	0.505574243	0.006346542	0.00867194	0.00728590	1
		50	0.49881494	0.50507901	0.49635292	0.009146881	0.005175571	0.00640989	2
		100	0.50072097	0.5015010	0.504129848	0.009934879	0.007248776	0.00826575	2
1	0.25	20	1.02314618	1.00613235	1.01165807	0.00098146	0.00802677	0.00603792	1
		50	1.00603976	1.00386773	1.007309543	0.001885594	0.006581762	0.00866836	1
		100	1.00025717	1.00054563	1.003646793	0.002840718	0.001174519	0.00946679	2
	0.5	20	1.00054358	1.01441350	1.003159975	0.008333692	0.007515976	0.00080285	3
		50	1.00520582	1.00779505	1.000866818	0.003742823	0.009575358	0.00756505	1
		100	1.44459847	1.00949390	1.222826205	0.196803119	0.009444949	0.10521063	2
	0.75	20	2.58811153	1.00418398	1.798893221	2.523116688	0.002003011	1.26390213	2
		50	1.00317286	1.00132015	0.999996882	0.008122294	0.003780885	0.00620529	2
		100	0.99959168	1.00212701	1.004733941	0.002222872	0.001569693	0.00372872	2
1.5	0.25	20	1.50923065	1.50637617	1.508013535	0.004448633	0.005701448	0.00856913	1
		50	1.50707706	1.50280906	1.505960677	0.00140227	0.004198614	0.00060753	3
		100	1.50399894	1.50088803	1.505579116	0.009129761	0.001075018	0.00867405	2
	0.5	20	1.51087667	1.51253385	1.511527938	0.002300309	0.004215082	0.00475540	1
		50	1.50980843	1.508822917	1.501153177	0.001963781	0.008532203	0.0037783	1
		100	1.5013108	1.507817803	1.509123942	0.005716012	0.004730745	0.00788014	2
	0.75	20	1.48472747	1.50414319	1.496758653	0.004427539	0.007507681	0.00940950	1
		50	1.50551278	1.50319558	1.500615845	0.009287262	0.005724735	0.00747705	2
		100	1.5046343	1.50165015	1.503080086	0.002643947	0.009913488	0.00819521	1

3- The best estimate method is (mom) with a success rate of (%57) percent.

4- The parameters of the Lomax distribution can be estimated by using the Bayesian methods and robust method to compare the results.

5- Parameters of other distributions (such as Gumbel and Fréchet distributions) can be estimated, with the samples including outliers.

References

[1] H. Alsuhabi, I Alkhairy, E.M. Almetwally, H.M. Almongy, A.M. Gemeay, E. Hafez, R.A. Aldallal and M. Sabry, *A superior extension for the Lomax distribution with application to Covid-19 infections real data*, Alexandria Engin. J. **61** (2022), no. 12, 11077–11090.

[2] D.E. Giles, H. Feng and R.T. Godwin. *On the bias of the maximum likelihood estimator for the two-parameter Lomax distribution*, Commun. Statist. Theory Meth. **42** (2013), no. , 1934–1950.

[3] N. Kilany, *Weighted lomax distribution*, SpringerPlus **5** (2016), no. 1, 1–18.

[4] J.P. Kleijnen, *Design and analysis of simulation experiments*, International Workshop on Simulation, Springer, 2015.

[5] J.A. Labban, *On 2-parameter estimation of Lomax distribution*, J. Phys.: Conf. Ser. **1294** (2019), no. 3, 032018.

[6] M. Parkinson, *The extreme value method for estimating the variance of the rate of return*, J. Bus. **53** (1980), no. 1, 61–65.

[7] N.N. Prasad and J.N. Rao, *The estimation of the mean squared error of small-area estimators*, J. Amer. Statist. Assoc. **85** (1990), no. 409, 163–71.

[8] H.M. Salem, *The exponentiated Lomax distribution: Different estimation methods*, Amer. J. Appl. Math. Statist. **2** (2014), no. 6, 364–368.

[9] A.N. Salman and A.M. Hamad, *On estimation of the stress–strength reliability based on lomax distribution*, IOP Conf. Ser.: Materials Sci. Engin. 2019.