International Journal of Scientific Advances
ISSN: 2708-7972
Volume: 4 | Issue: 5 | Sep – Oct 2023 Available Online: www.ijscia.com
DOI: 10.51542/ijscia.v4i5.6

Root Mean Square Error of the Maximum Likelihood Estimate of the Parameters of Pareto Distribution

Epimaco A. Cabanlit, Jr.* and Kimberly D. Dela Cruz

Mathematics Department, Mindanao State University, General Santos City, Philippines

*Corresponding author details: Epimaco A. Cabanlit, Jr.; maco_727@yahoo.com

ABSTRACT

Pareto distribution is one of the most important distributions in reliability work. In this paper, estimating the parameters of Pareto Distribution is presented using maximum likelihood estimation for different sample sizes. Computer simulation has been carried out to obtain results. Root mean square errors were computed for several values of the parameters. As the sample size increases, the root means square error of the maximum likelihood decreases.

Keywords: pareto distribution; root mean square error; maximum likelihood estimate; parameters; simulation

INTRODUCTION

The Pareto distribution is given by,

$$f(x) = \frac{\alpha k^{\alpha}}{x^{\alpha+1}}$$

where $x \geq k$, k is the lower bound is a scale parameter of the data and α is the shape parameter. This is sometimes called a power law distribution. It has a property that large numbers are rare, but smaller numbers are more common, and thus it is more common for a person to make a small amount of money versus a large amount of money [1].

This distribution is a good model for personal incomes, the assets of firms, the sizes of cities and the number of firms in various industries. Pareto distribution is a powerful tool for modeling a variety of real-life phenomena in social sciences, and actuarial. It is named after the Italian economist Vilfredo Pareto (1848-1923), who developed the distribution in the 1890's as a way to describe the allocation of wealth in society. He famously observed that 80% of society's wealth was controlled by 20% of its population, a concept now known as the "Pareto Principle" or the "80-20 Rule" (2).

The maximum likelihood estimates of the two parameters are $\hat{k} = min\,\{x_i\}$ and $\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} log\left(\frac{x_i}{\hat{k}_j}\right)}$.

The Root mean square error (RMSE) of an estimator of a population parameter is the square root of the mean square error (MSE). The mean square error is defined as the expected value of the square of the difference between the estimator and the parameter.

In this paper, we used computer simulation to compute the root mean square error of the maximum likelihood estimate of the parameters of Pareto Distribution.

THE ALGORITHM

To compute the root, mean square error, the following are the steps.

(1) Generate n random numbers ξ_i from 0 to 1.

(2) Generate n random variates $x_i = \frac{k}{(1-\xi_i)^{\frac{1}{\alpha}}}$.

(3) Compute $\hat{k_j} = min\,\{x_i\}$ for j=1 to 100.

(4) Compute $\hat{\alpha_j} = \frac{n}{\sum_{i=1}^{n} log\left(\frac{x_i}{\hat{k}_j}\right)}$ for j=1 to 100.

(5) Compute the means of the estimators.

(6) Compute the variances of the estimators.

(7) Compute the root mean square error using the following formulas:

a. $RMSE(k) = \sqrt{\frac{1}{100}\sum_{j=1}^{100}\left(\hat{k}_j - k\right)^2}$

b. $RMSE(\alpha) = \sqrt{\frac{1}{100}\sum_{j=1}^{100}\left(\hat{\alpha}_j - \alpha\right)^2}$

THE RESULTS

As the number of sample size n increases, we can observe the trends of the means of the Maximum Likelihood Estimates of the parameters $k = 5$, $(\alpha = 2, \alpha = 5,$ and $\alpha = 10$ approach to the true values (Tables 1, 2 and 3). This supports that maximum likelihood estimates of the parameters are unbiased estimators, that is $\hat{\theta}$ is an unbiased estimator of θ if $E(\hat{\theta}) = \theta$ [3]. These are pictured by the graphs as seen in Figures 1, 2 and 3.

TABLE 1: Means of the Parameters (k=5, α=2).

Number of samples (n)	$k = 5$	$\alpha = 2$
5	5.55	3.347
10	5.266	2.597
30	5.087	2.146
50	5.051	2.061
100	5.022	2.034
500	5.077	2.013
1000	5.003	1.98423

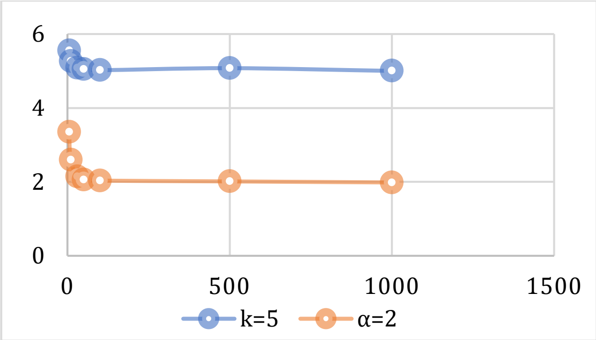

FIGURE 1: Graph of the means of the parameters k=5, α=2.

TABLE 2: Means of the Parameters ($k = 5, \alpha = 5$).

Number of samples (n)	$k = 5$	$a = 5$
5	5.1988	7.8101
10	5.1079	6.3232
30	5.0371	5.3566
50	5.0194	5.1649
100	5.01	5.1399
500	5.0019	5.043
1000	5.0011	4.8866

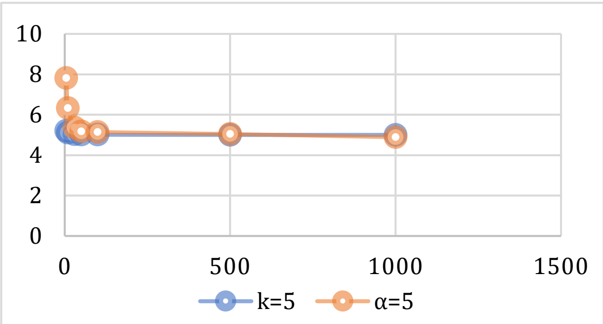

FIGURE 2: Graph of the means of the parameters k=5, α=5.

TABLE 3: Means of the Parameters ($k = 5, \alpha = 10$).

Number of samples (n)	$k = 5$	$a = 10$
5	5.102	16.27
10	5.045	13.02
30	5.015	10.6
50	5.009	10.54
100	5.005	10.2
500	5.001	10.09
1000	5	10.01

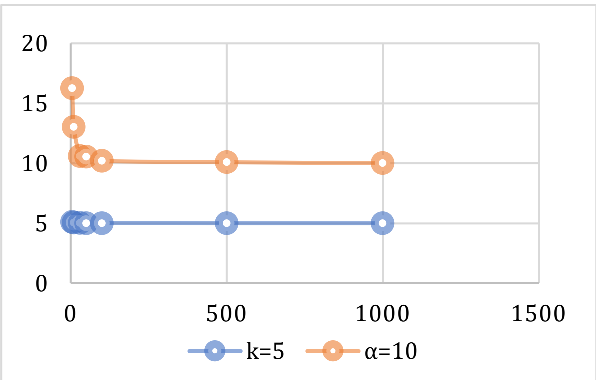

FIGURE 3: Graph of the means of the parameters k=5, α=10.

Similarly, the variances for $k = 5$, ($\alpha = 2, \alpha = 5, \alpha = 10$) converge to zero as the sample sizes increase (Tables 3, 4 and 5). This can be appreciated by viewing the graphs as depicted in Figures 3, 4 and 5. With these results, we can say that the estimators are consistent. Estimators are consistent if $\lim_{n\to\infty} \text{Var}(\hat{\theta}) = 0$ [3].

TABLE 4: Variances of the Parameters ($k = 5, \alpha = 2$).

Number of samples (n)	$k = 5$	$a = 2$
5	0.389	4.822
10	0.079	1.604
30	0.006	0.154
50	0.003	0.093
100	0.001	0.047
500	6.00E-05	0.012251
1000	9.00E-06	0.003887

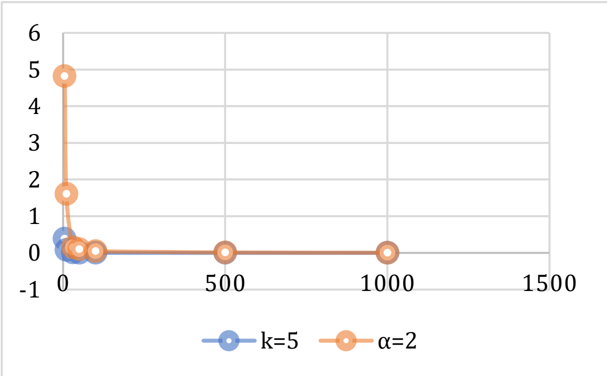

FIGURE 4: Graph of the variances of the parameters k=5, α=2.

TABLE 5: Variances of the Parameters ($k = 5, \alpha = 5$).

Number of samples (n)	$k = 5$	$a = 5$
5	0.035269	21.40974
10	0.015394	4.318826
30	0.001485	1.271762
50	0.003	0.590717
100	9.03E-05	0.23629
500	3.35E-06	0.055196
1000	1.24E-06	0.021706

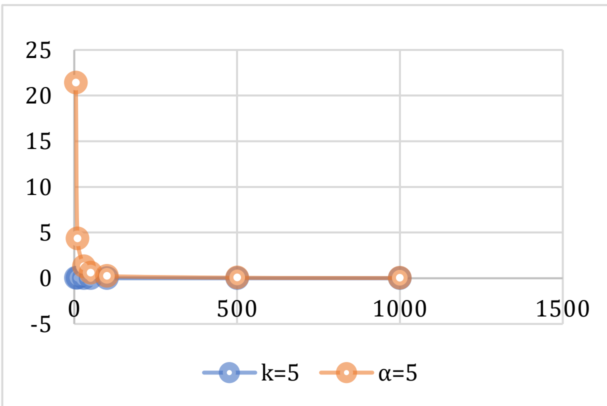

FIGURE 5: Graph of the variances of the parameters k=5, α=5.

TABLE 6: Variances of the Parameters ($k = 5, \alpha = 10$).

Number of samples (n)	$k = 5$	$\alpha = 10$
5	0.00966	139.776
10	0.00206	21.0265
30	0.00024	3.89196
50	0.0001	2.07936
100	2.50E-05	0.9174
500	1.10E-06	0.33507
1000	1.30E-07	0.08682

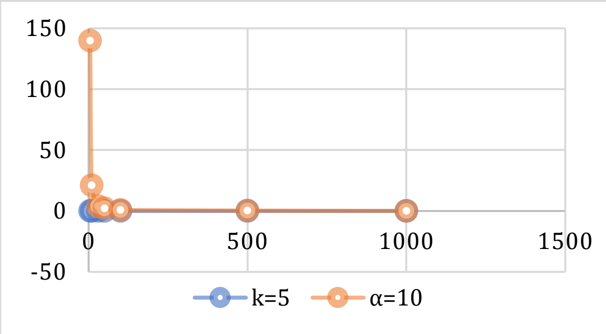

FIGURE 6: Graph of the variances of the parameters k=5, α=10.

Accordingly, the root means square errors of the maximum likelihood estimates $k = 5, (\alpha = 2, \alpha = 5, \alpha = 10)$ converge to zero (Tables 7, 8 and 9). Graphs of these convergences are pictured in Figures 7, 8 and 9 respectively.

TABLE 7: Root Mean Square Error of the Parameters ($k = 5, \alpha = 2$).

Number of samples (n)	$k = 5$	$\alpha = 2$
5	0.83	2.567
10	0.385	1.395
30	0.114	0.398
50	0.073	0.31
100	0.046	0.223
500	0.01	0.111
1000	0.00424	0.06401

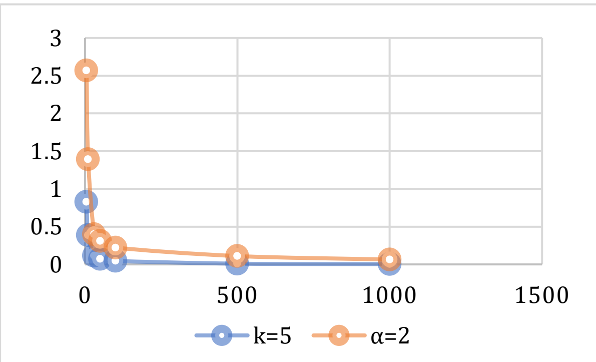

FIGURE 7: Graph of the RMSE of the parameters k=5, α=2.

TABLE 8: Root Mean Square Error of the Parameters ($k = 5, \alpha = 5$).

Number of samples (n)	$k = 5$	$\alpha = 5$
5	0.2728	5.4059
10	0.164	2.492
30	0.0534	1.2367
50	0.026	0.8419
100	0.0138	0.5461
500	0.0026	0.429
1000	0.0016	0.3262

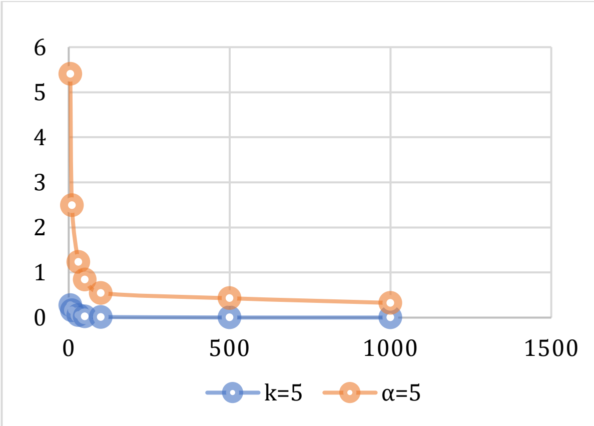

FIGURE 8: Graph of the RMSE of the parameters k=5, α=5.

TABLE 9: Root Mean Square Error of the Parameters ($k = 5, \alpha = 10$).

Number of samples (n)	$k = 5$	$\alpha = 2$
5	0.141	13.33
10	0.064	5.417
30	0.022	2.054
50	0.014	1.532
100	0.007	0.973
500	0.001	0.825
1000	5.00E-04	0.576

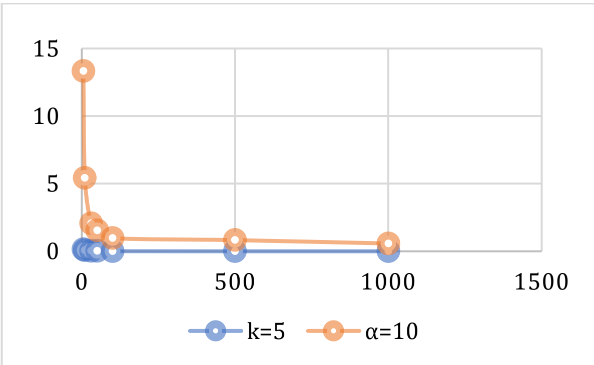

FIGURE 9: Graph of the RMSE of the parameters k=5, α=10.

CONCLUDING REMARKS

Simulation results show that the root mean square errors of the maximum likelihood estimates of the parameters decrease to zero, as we increase the sample size. The Mean Absolute Error of the maximum likelihood estimate of the parameters is a potential study, and so with the other methods of estimating the parameters of the Pareto Distribution. It is suggested to conduct comparative evaluation. This is very interesting since the Root Mean Square Error and the Mean Absolute Error are defined differently, we should expect the results to be different [4].

REFERENCES

[1] Metcalf, L and Casey, W. Pareto Distribution. Cybersecurity and Applied Mathematics. Elsevier Inc, 2016.

[2] Haley, C. Explaining the 80-20 Rule with the Pareto Distribution. Berkeley D Lab. March 15, 2022.

[3] Kokoska, S and Nevison, C. Statistical Tables and Formulae. Springer-Verlag, New York. 1989.

[4] Chai, T. and Draxler, R. R. Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. Geoscientific Model Development, 7, 1247-1250. 2014.