International Journal of Applied Mathematics
& Statistical Sciences (IJAMSS)
ISSN(P): 2319-3972; ISSN(E): 2319-3980
Vol. 5, Issue 3, Apr - May 2016; 63-82

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

USING THE STANDARD BAYES METHOD AND MODIFIED MOMENT METHOD TO ESTIMATE THE PARAMETERS OF PARETO DISTRIBUTION WITH PRACTICAL APPLICATION

HAZIM MANSOOR GORGEES[1] & ADWEA NAJI ATEWI AL-KHSHALI[2]

[1]Assistant Professor, Department of Mathematics, College of Education for Pure Science Ibn-Alhaithem, University of Baghdad, Iraq

[2]Department of Health Management, Institute of Medical Technology Al-Mansur, Iraq

ABSTRACT

The aim of this paper is to employ some estimation methods, namely, modified moment method and standard Bayes method to estimate the two parameters of the first kind Pareto distribution and comparing the performance of these two methods by using the Mont Carlo simulation technique. Moreover, some Applications related to Pareto distribution, in particular, human population statistics and tropical cyclones in Japan are given. The Chi Square goodness of fit test is presented for each case.

KEYWORDS: Pareto Distribution, Modified Moment Method, Standard Bayes Method, goodness of Fit Test

1. INTRODUCTION

In this paper, the first kind Pareto distribution Par(α,c) is studied Pareto distribution is widely used as a model in many areas of application, for instance, the values of oil reserves in oil fields, sizes of sand particles, areas burnt in forest fires, economics, human population statistics, rainfall, and others. Professor of economics Vilfredo Pareto (1848-1923) was the first who proposed this distribution when he tried to put a comprehensive law to deal with the distribution of income for a particular community[1]. The remaining of this paper is organized as follows: in section 2, basic concepts on Pareto distribution are given, in section 3, two estimation methods are presented. Namely, the modified moment method and standard Bayes method. Monte Carlo results including the estimated values of the parameters as well as the estimation of some properties such as bias, mean, variance, kurtosis, skewness and MSE are presented in section 4. In section 5 we state the Chi Square goodness of fit test. Some real life applications related to Pareto distribution are presented in section 6. In section 7 the discussion of the results and conclusions are given.

2. BASIC CONCEPTS ON PARETO DISTRIBUTION:[2]

A continuous random variable X is said to have Pareto distribution, denoted by X~Par(α,c) if it has Probability density function is:

$$f(x;\alpha,c)=\begin{cases}\dfrac{\alpha c^{\alpha}}{x^{\alpha+1}} & ,x\geq c \quad ,c>0,\alpha>0 \\ 0 & \text{otherwise}\end{cases} \quad \ldots\ldots(1)$$

Where α ,c are the shape and scale parameters respectively.

The greater value can be obtained for the function above, if x=c:

$$f(c) = \frac{\alpha}{c}$$

Pareto distribution depends on two parameters α and c, a wide variety of distribution shapes and scals can be generated from suitable choices of α and c, figures (1) show some Pareto p.d.f's:

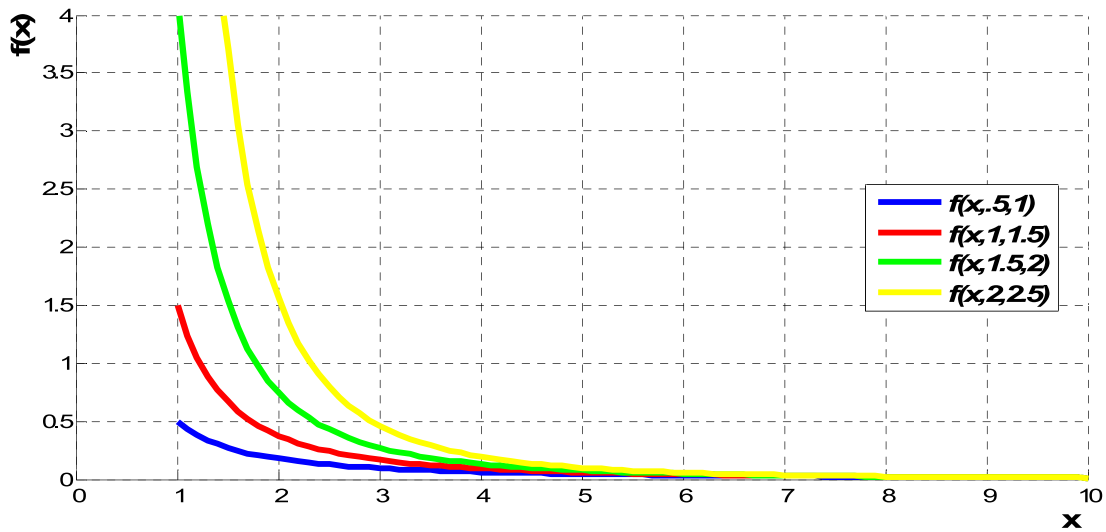

Figure 1: Pareto pdf's

The cumulative distribution function (cdf) of X is:

$$F(x;\alpha,c) = P(X \le x)$$

$$= \int_c^x f(u)du$$

$$F(x;\alpha,c) = 1 - \left(\frac{c}{x}\right)^{\alpha} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

3. ESTIMATION METHODS

3-1 The Modified Moments Method: [3]

Let $x_1, x_2, \ldots, x_n$ be a random sample from $par(\alpha, c)$ and let y_1 be the first order statistic of the sample, then the p.d.f of y_1 is given as:

$$g_1(y_1) = \begin{cases} n\alpha c^{n\alpha} y_1^{-(n\alpha+1)}, & c \le y_1 \infty \\ 0, & e.w. \end{cases}$$

Hence:

$$E(y_1)=\frac{n\alpha c}{(n\alpha-1)}$$

Equating y_1 to $E(y_1)$ in above:

$$y_1=\frac{n\hat{\alpha}\hat{c}}{n\hat{\alpha}-1}$$

Moreover, it can be shown that:

$$\overline{X}=\frac{\hat{\alpha}\hat{c}}{\hat{\alpha}-1}$$

Solving the last two equations yields the following estimators:

$$\hat{\alpha}=\frac{n\overline{X}-y_1}{n(\overline{X}-y_1)} \quad \ldots\ldots (3)$$

$$\hat{c}=\frac{(n\hat{\alpha}-1)y_1}{n\hat{\alpha}} \quad \ldots\ldots (4)$$

3-2 The Standard Bayes Estimator: [4][5][6]

Bayes style in estimation depends on the assumption that the Prior information available on the parameter to be estimated, can be formulated as the probability density function which is called the Prior probability density function (Prior P.d.f), and Bayes estimator depends on the Posterior probability density function (Posterior. P.d.f):

$$h(\theta \mid x_1,\ldots,x_n)\propto L(x_1,\ldots,x_n;\theta)J(\theta)$$

Where, $L(x_1,\ldots,x_n;\theta)$ is the Likelihood function

$J(\theta)$: Prior P.d.f for θ

$$h(\theta \mid x_1,\ldots,x_n)=kL(x_1,\ldots,x_n;\theta)J(\theta)$$

$$k^{-1}=\int_{\forall\theta} L(x_1,\ldots,x_n;\theta)J(\theta)d\theta \quad ;k\ is\ constant$$

$$\hat{\theta}_B=E(\theta / x_1,\ldots,x_n)$$

For the case of Pareto Distribution and by using non informative Prior (p.d.f):

According to Jeffrey method we have:

$$J_1(\alpha)\propto\frac{1}{\alpha^s} \quad ,\alpha>0$$

$$J_2(c) \propto \frac{1}{c^r} \qquad , c > 0$$

$$J(\alpha, c) \propto J_1(\alpha) \cdot J_2(c)$$

Where, each of $J_1(\alpha)$, $J_2(c)$ is the prior p.d.f for α, c respective ly

$$J(\alpha, c) \propto \frac{1}{\alpha^s c^r}$$

Let $x_1, x_2 \ldots, x_n$ be a random sample from Par(α,c), then the likelihood function is:

$$L(x_1, x_2, \ldots, x_n; \alpha, c) = \underset{i=1}{\overset{n}{\pi}} f(x_i / \alpha, c)$$

$$L(x_1, x_2, \ldots, x_n; \alpha, c) = \underset{i=1}{\overset{n}{\pi}} \frac{\alpha c^{\alpha}}{x_i^{\alpha+1}}$$

we get:

$$h(\alpha, c / x_1, \ldots, x_n) \propto \frac{1}{\alpha^s c^r} \cdot \frac{\alpha^n c^{n\alpha}}{\left(\underset{i=1}{\overset{n}{\pi}} x_i \right)^{\alpha+1}}$$

$$h(\alpha, c / x_1, \ldots, x_n) \propto \alpha^{n-s} c^{n\alpha - r} \exp\left[-(\alpha+1) \sum_{i=1}^{n} \ln x_i \right]$$

Let r=1

$$h(\alpha, c / x_1, \ldots, x_n) = k \ \alpha^{n-s} c^{n\alpha-1} \exp\left[-(\alpha+1) \sum_{i=1}^{n} \ln x_i \right]$$

$$k^{-1} = \int_0^{\infty} \int_0^{x_{(1)}} \alpha^{n-s} c^{n\alpha-1} \exp\left[-(\alpha+1) \sum_{i=1}^{n} \ln x_i \right] dc d\alpha$$

$$= \exp\left(-\sum_{i=1}^{n} \ln x_i \right) \int_0^{\infty} \left[\frac{c^{n\alpha}}{n\alpha} \right]_0^{x_{(1)}} \alpha^{n-s} \exp\left(-\alpha \sum_{i=1}^{n} \ln x_i \right) d\alpha$$

$$= \exp\left(-\sum_{i=1}^{n} \ln x_i \right) \int_0^{\infty} \alpha^{n-s} \frac{x_{(1)}^{n\alpha}}{n\alpha} \exp\left(-\alpha \sum_{i=1}^{n} \ln x_i \right) d\alpha$$

$$= \frac{\exp\left(-\sum_{i=1}^{n} \ln x_i \right)}{n} \int_0^{\infty} \alpha^{n-s-1} \exp\left[-\alpha \left(\sum_{i=1}^{n} \ln x_i - n \ln x_{(1)} \right) \right] d\alpha$$

$$=\frac{\exp\left(-\sum_{i=1}^{n}\ln x_i\right)}{n}\int_0^\infty \alpha^{n-s-1}\exp\left[-\alpha\sum_{i=1}^{n}\ln\left(\frac{x_i}{x_{(1)}}\right)\right]d\alpha$$

$$k^{-1}=\frac{\exp\left(-\sum_{i=1}^{n}\ln x_i\right)}{n}\;\frac{\Gamma(n-s)}{\left[\sum_{i=1}^{n}\ln\left(\frac{x_i}{x_{(1)}}\right)\right]^{n-1}}$$

$$k=\frac{n}{\exp\left(-\sum_{i=1}^{n}\ln x_i\right)}\;\frac{\left[\sum_{i=1}^{n}\ln\left(\frac{x_i}{x_{(1)}}\right)\right]^{n-1}}{\Gamma(n-s)}$$

$$h(\alpha,c\mid X_1,\dots,X_n)=\frac{n\alpha^{n-s}c^{n\alpha-1}\exp\left(-\alpha\sum_{i=1}^{n}\ln x_i\right)}{\Gamma(n-s)\left[\sum_{i=1}^{n}\ln\left(\frac{x_i}{x_{(1)}}\right)\right]^{-(n-s)}},\quad 0<c<x_{(1)}<x<\infty,\ \alpha>0$$

$$\hat{c}=x_{(1)}=\min\{X_1,X_2,\dots,X_n\}\cdots\cdots(5)$$

$$M(\alpha\mid X_1,\dots,X_n)=\int_c h(\alpha,c\mid X_1,\dots,X_n)dx$$

$$M(\alpha\mid X_1,\dots,X_n)=\frac{\alpha^{n-s-1}e^{-\alpha\sum_{i=1}^{n}\ln(x_i)}}{)(n-s)\left[\sum_{i=1}^{n}\ln\left(\frac{x_i}{c}\right)\right]^{n-s}}$$

Suppose the Loss Function is Squared loss function then Standard estimator for shape parameter α is mean of posterior distribution:

$$\hat{\alpha}=E\left(\frac{\alpha}{x}\right)=\frac{n-s}{\sum_{i=1}^{n}\ln\left(\frac{x_i}{c}\right)}$$

If s=0 then the standard Bayes estimator $\hat{\alpha}$ is the MLE, and if s=2 then Bayes estimator $\hat{\alpha}$ is minimum variance unbiased estimator (MVUE) of α

$$\hat{\alpha}=\frac{n-2}{\sum_{i=1}^{n}\ln\left(\frac{x_i}{c}\right)}\cdots\cdots(6)$$

4. MONTE CARLO RESULTS

In this section, a large scale of samples generated from the Pareto distribution (par(3,1)) with size n(m); n=size of sample and m=Iterations, 5(1000), 10(1000), 20(1000), 30(1000), 50(1000), 75(1000), 100(1000), 250(1000), 500(1000) and 1000(1000), These samples are used to estimate $(\hat{\alpha},\hat{c})$ of the Pareto distribution by using two method of estimator, Modified moments method and Standard Bayes method, and estimation some properties such us bias, mean, variance, kurtosis, skewness and MSE.

In table (1), used equation (3), (4), (5) and (6) by using Matlab R2013b we get a simulation of the estimator $(\hat{\alpha},\hat{c})$, are displayed in table (1):

Table 1: Parameters Estimation

Sample size n(m)	*Estimation of* $(\hat{\alpha},\hat{c})$	
	M.M.M	**B.M**
5(1000)	(4.1729, 1.0077)	(2.9343, 1.0713)
10(1000)	(3.4438, 1.0015)	(2.9240, 1.0346)
20(1000)	(3.2797, 1.0000)	(3.0196, 1.0164)
30(1000)	(3.1681, 0.9995)	(3.0063, 1.0106)
50(1000)	(3.1150, 1.0004)	(3.0079, 1.0070)
75(1000)	(3.0870, 0.9997)	(3.0128, 1.0041)
100(1000)	(3.0540, 0.9999)	(2.9944, 1.0032)
250(1000)	(3.0341, 1.0001)	(3.0094, 1.0014)
500(1000)	(3.0057, 1.0000)	(2.9965, 1.0007)
1000(1000)	(3.0065, 1.0000)	(3.0007, 1.0003)

Table(1) show that methods M.M.M and B.M give a good agreement between the true values $\alpha = 3$ and $c = 1$ with the estimators $(\hat{\alpha})$ and $(\hat{c})$ for all sample sizes, but the M.M.M give small values for c and B.M the better To estimation α for all sample sizes. The following are some Pareto figures where par(3,1) for different sample sizes:

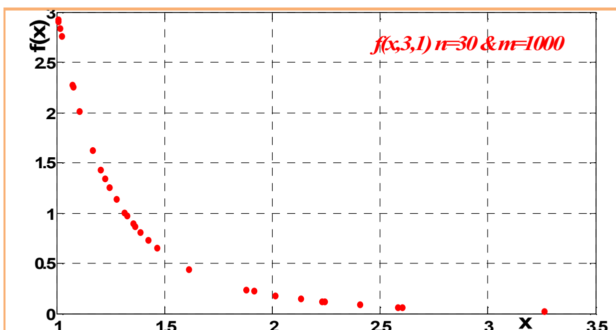

Figure (3.1): Values of p.d.f when n=30 and m=1000

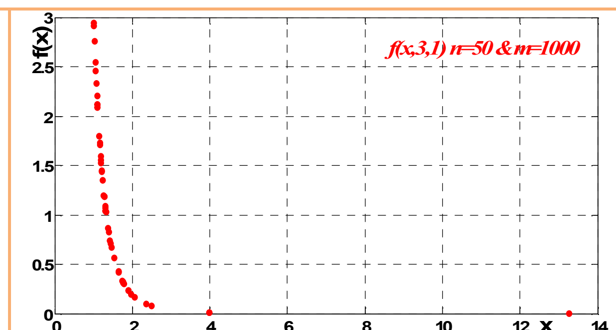

Figure (3.2): Values of p.d.f when n=50 and m=1000

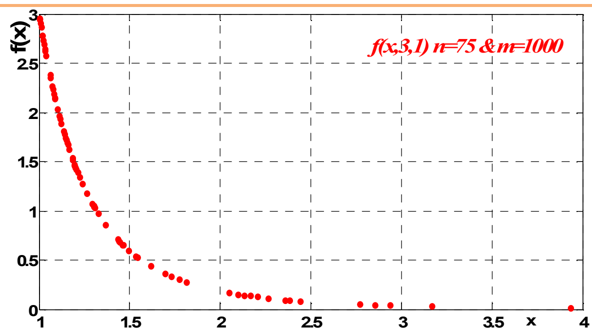

Figure (3.3): Values of p.d.f when n=75 and m=1000

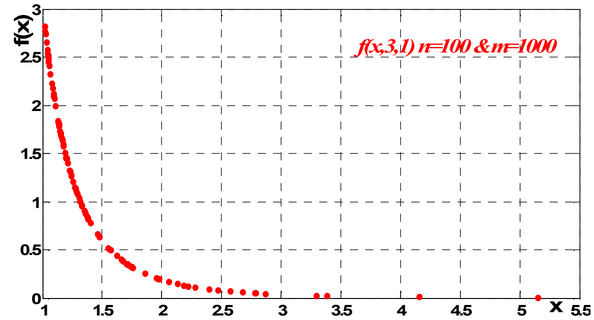

Figure (3.4): Values of p.d.f when n=100 and m=1000

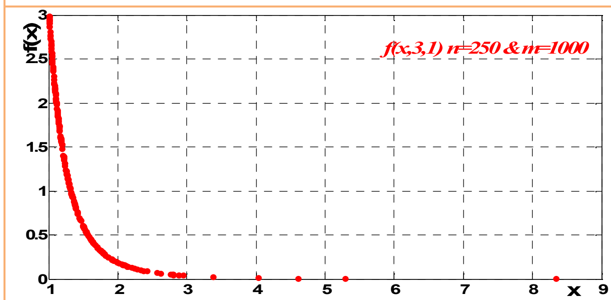

Figure (3.5): Values of p.d.f when n=250 and m=1000

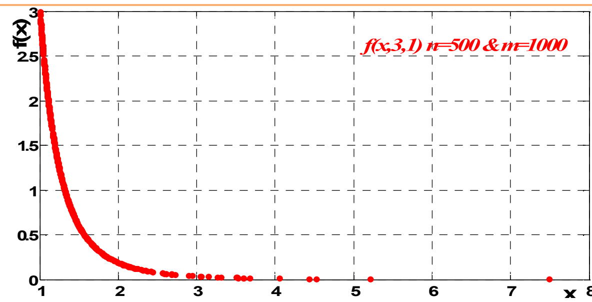

Figure (3.6): Values of p.d.f when n=500 and m=1000

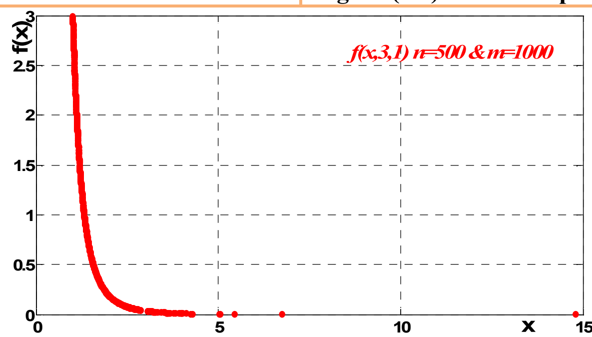

Figure (3.7): Values of p.d.f when n=1000 and m=1000

The biases of estimators $\hat{\alpha}$ and $\hat{c}$ for Pareto distribution which can be found by:

$$\text{bias}\,(\hat{\alpha}) = E(\hat{\alpha}) - \alpha$$

$$\text{bias}\,(\hat{c}) = E(\hat{c}) - c$$

are displayed in tables (2) and (3):

Table (2): Bias of Estimator $(\hat{\alpha})$

Sample size n(m)	*Bias of Estimation* $(\hat{\alpha})$	
	M.M.M	**B.M**
5(1000)	1.1729	-0.0657
10(1000)	0.4438	-0.0760
20(1000)	0.2797	0.0196
30(1000)	0.1681	0.0063
50(1000)	0.1150	0.0079

75(1000)	0.0870	0.0128
100(1000)	0.0540	-0.0056
250(1000)	0.0341	0.0094
500(1000)	0.0057	-0.0035
1000(1000)	0.0065	6.6286e-04

Table 3: Bias of Estimator $(\hat{c})$

Sample size n(m)	***Bias of Estimation $(\hat{c})$***	
	M.M.M	**B.M**
5(1000)	0.0077	0.0713
10(1000)	0.0015	0.0346
20(1000)	4.3803e-05	0.0164
30(1000)	-4.7243e-04	0.0106
50(1000)	4.3246e-04	0.0070
75(1000)	-2.6800e-04	0.0041
100(1000)	-7.5317e-05	0.0032
250(1000)	7.3703e-05	0.0014
500(1000)	2.8117e-05	6.9569e-04
1000(1000)	-9.2540e-06	3.2393e-04

Table (4): show the variance of estimator $(\hat{\alpha})$ which can be obtained by using the equation:

$$Var(\hat{\alpha}) = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\alpha}_i - \overline{\hat{\alpha}}\right)^2$$

Table 4: Variance of Estimator $(\hat{\alpha})$

Sample size n(m)	***Variance of Estimation $(\hat{\alpha})$***	
	M.M.M	**B.M**
5(1000)	6.3794	3.6320
10(1000)	1.3521	1.0715
20(1000)	0.5802	0.5058
30(1000)	0.4114	0.3687
50(1000)	0.2199	0.1938
75(1000)	0.1579	0.1399
100(1000)	0.1023	0.0921
250(1000)	0.0412	0.0345
500(1000)	0.0227	0.0173
1000(1000)	0.0123	0.0098

Table (4) show that the variance values of the estimator $(\hat{\alpha})$ by using the B.M are smaller than the variance values of M.M.M.

Table (5): shows the variance of estimator $(\hat{c})$ which can be obtained by using the equation:

$$Var(\hat{c}) = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{c}_i - \bar{\hat{c}}\right)^2$$

Table 5: Variance of Estimator $(\hat{c})$

Sample size n	***Variance of Estimation $(\hat{c})$***	
	M.M.M	**B.M**
5(1000)	0.0054	0.0052
10(1000)	0.0013	0.0012
20(1000)	2.7613e-04	2.6987e-04
30(1000)	1.2335e-04	1.2494e-04
50(1000)	5.1961e-05	5.1353e-05
75(1000)	1.8262e-05	1.8186e-05
100(1000)	9.9469e-06	9.8536e-06
250(1000)	1.9483e-06	1.9391e-06
500(1000)	4.8348e-07	4.8337e-07
1000(1000)	1.0994e-07	1.0966e-07

Table (5): show that the variance values of the estimator $(\hat{c})$ by using the M.M.M are greater than the variance values of B.M.

The m.s.e of estimator $(\hat{\alpha})$ and $(\hat{c})$ can be obtained by:

$$m.s.e(\hat{\alpha}) = Var[\hat{\alpha}] + [bias(\hat{\alpha})]^2$$

$$m.s.e(\hat{c}) = Var(\hat{c}) + [bias(\hat{c})]^2$$

Tables (6) and (7) show m.s.e of $(\hat{\alpha})$ and $(\hat{c})$ by the two estimations methods.

Table 6: Mean Square Error of Estimator $(\hat{\alpha})$

Sample size n(m)	***Mean Square error of Estimator $(\hat{\alpha})$***	
	M.M.M	**B.M**
5(1000)	7.7551	3.6363
10(1000)	1.5490	1.0773
20(1000)	0.6584	0.5062
30(1000)	0.4397	0.3688
50(1000)	0.2332	0.1938
75(1000)	0.1655	0.1400
100(1000)	0.1052	0.0921
250(1000)	0.0423	0.0346
500(1000)	0.0227	0.0174
1000(1000)	0.0124	0.0098

For all sample sizes in table (6) the m.s,e of $\hat{\alpha}$ given by B.M is less than other method results.

Table 7: Mean Square Error of Estimator $(\hat{c})$

Sample size n(m)	Mean Square Error of Estimator $(\hat{c})$	
	M.M.M	B.M
5(1000)	0.0054	0.0103
10(1000)	0.0013	0.0024
20(1000)	2.7613e-04	5.3748e-04
30(1000)	1.2357e-04	2.3677e-04
50(1000)	5.2148e-05	1.0097e-04
75(1000)	1.8334e-05	3.5333e-05
100(1000)	9.9525e-06	2.0394e-05
250(1000)	1.9538e-06	3.8988e-06
500(1000)	4.8427e-07	9.6735e-07
1000(1000)	1.1003e-07	2.1458e-07

For all sample sizes in table (7) the m.s.e of $\hat{c}$ given by M.M.M is Less than B.M results.

The skewness of estimators $(\hat{\alpha})$ and $(\hat{c})$ which can be obtained by:

$$Sekwenss(\hat{\alpha}) = \frac{\frac{1}{n}\left[\sum_{i=1}^{n}(\alpha_i)^3 - 3\hat{\alpha}\sum_{i=1}^{n}(\alpha_i)^2 + 3\hat{\alpha}^2\sum_{i=1}^{n}(\alpha_i) - \hat{\alpha}^3\right]}{(\sigma^2)^{3/2}}$$

$$Sekwenss(\hat{c}) = \frac{\frac{1}{n}\left[\sum_{i=1}^{n}(c_i)^3 - 3\hat{c}\sum_{i=1}^{n}(c_i)^2 + 3\hat{c}^2\sum_{i=1}^{n}(c_i) - \hat{c}^3\right]}{(\sigma^2)^{3/2}}$$

Are shown in tables (8) and (9):

Table 8: Skewness of Estimator $(\hat{\alpha})$

Sample size n(m)	Skewness of Estimator $(\hat{\alpha})$	
	M.M.M	B.M
5(1000)	1.8293e+03	1.6368e+03
10(1000)	2.7416e+03	2.4061e+03
20(1000)	4.0300e+03	3.8694e+03
30(1000)	4.0474e+03	4.0784e+03
50(1000)	5.8659e+03	6.3890e+03
75(1000)	6.2495e+03	6.9717e+03
100(1000)	8.7039e+03	9.5994e+03
250(1000)	1.3363e+04	1.7009e+04
500(1000)	1.5914e+04	2.3534e+04
1000(1000)	1.9849e+04	2.8032e+04

Table 9: Skewness of Estimator $(\hat{c})$

Sample size n(m)	Skewness of Estimator $(\hat{c})$	
	M.M.M	B.M
5(1000)	5.1883e+05	6.4987e+05
10(1000)	2.2238e+06	2.6022e+06
20(1000)	1.0887e+07	1.1829e+07
30(1000)	2.4273e+07	2.4610e+07
50(1000)	5.3412e+07	5.5448e+07
75(1000)	1.7054e+08	1.7389e+08
100(1000)	3.1837e+08	3.2614e+08
250(1000)	1.4861e+09	1.4861e+09
500(1000)	5.9439e+09	5.9577e+09
1000(1000)	2.7404e+10	2.7538e+10

Tables (8) and (9) show that the two methods give a very small skewness to left and to the right which indicate that the estimators $(\hat{\alpha})$ and $(\hat{c})$ approach rapidly to normal distribution.

The kurtosis of estimators $(\hat{\alpha})$ and $(\hat{c})$ which can be obtained by:

$$\text{kurtosis}(\hat{\alpha}) = \frac{\frac{1}{n}\left[\sum_{i=1}^{n}(\alpha_i)^4 - 4\hat{\alpha}\sum_{i=1}^{n}(\alpha_i)^3 + 6\hat{\alpha}^2\sum_{i=1}^{n}(\alpha_i) - 4\hat{\alpha}^3\sum_{i=1}^{n}(\alpha_i)\right]}{(\sigma^2)^2} - 3$$

$$\text{kurtosis}(\hat{c}) = \frac{\frac{1}{n}\left[\sum_{i=1}^{n}(c_i)^4 - 4\hat{c}\sum_{i=1}^{n}(c_i)^3 + 6\hat{c}^2\sum_{i=1}^{n}(c_i) - 4\hat{c}^3\sum_{i=1}^{n}(c_i)\right]}{(\sigma^2)^2} - 3$$

Are shown in Tables (10) and (11):

Table 10: Kurtosis of Estimator $(\hat{\alpha})$

Sample size n(m)	Kurtosis of Estimator $(\hat{\alpha})$	
	M.M.M	B.M
5(1000)	-1.6462e+03	1.1048e+03
10(1000)	-4.5059e+04	-3.5588e+04
20(1000)	-9.4167e+04	-8.6615e+04
30(1000)	-1.0597e+05	-1.0489e+05
50(1000)	-2.0270e+05	-2.2380e+05
75(1000)	-2.5027e+05	-2.8645e+05
100(1000)	-4.2438e+05	-4.7937e+05
250(1000)	-1.0100e+06	-1.3876e+06
500(1000)	-1.5968e+06	-2.6849e+06
1000(1000)	-2.6974e+06	-4.2693e+06

Table 11: Kurtosis of Estimator $(\hat{c})$

Sample size n(m)	Kurtosis of Estimator $(\hat{c})$	
	M.M.M	B.M
5(1000)	-7.6843e+06	-1.3735e+07
10(1000)	-6.3665e+07	-9.3231e+07
20(1000)	-6.5712e+08	-8.0445e+08
30(1000)	-2.1821e+09	-2.3687e+09
50(1000)	-7.4419e+09	-8.1295e+09
75(1000)	-3.9878e+10	-4.2003e+10
100(1000)	-1.0100e+11	-1.0637e+11
250(1000)	-1.0545e+12	-1.0787e+12
500(1000)	-8.5586e+12	-8.6195e+12
1000(1000)	-8.2726e+13	-8.3434e+13

5. CHI-SQUARE GOODNESS-OF-FIT TEST:[5]

Pearson Chi-Square goodness of fit test can be used for testing the null hypothesis involving discrete as well as continuous distribution, the Pearson chi-square test uses the density function of the random variable X.

Let $x_1, x_2, \ldots, x_n$ be a random sample from a random variable X with probability density function f(x). We wish to test the null hypothesis

$$H_o : X \sim \mathrm{f}(\mathrm{x})$$

Against:

$$H_1 : X \nsim \mathrm{f}(\mathrm{x})$$

For this we gusset the n observations into k mutually exclusive cells and we let p_i be the probability that one x_i of the sample is fall in cell i, Y_i be the observed number of the cell i, and np_i be the expected number of the cell i, i=1,2,…,n. Since the sample size is n, the number of observations expected to fall in the ith interval is equal to np_i. Then:

$$Q = \sum_{i=1}^{n} \frac{(Y_I - np_i)^2}{np_i} \quad \ldots\ldots (7)$$

has an approximate chi-square distribution with k −1 degrees of freedom.

We reject the null hypotheses Ho if $Q \geq \chi^2_{1-\alpha}(k-1)$ where $\chi^2_{1-\alpha}(k-1)$ is the theoretical value of Chi-Square distribution with $1-\alpha$ level of sig. and (k-1) degrees of freedom.

6. SOME REAL LIFE APPLICATIONS RELATED OF PARETO DISTRIBUTION:

6-1 Application (1): Human Population Statistics

6-1-1 Sweden

In this application we use the population statistic for Sweden, each number appears in table (12) below represent the number of peoples for each Sweden cities at 2015 according to the population estimates, we regard each number as a random variables follows a Pareto distribution, so we write a computer program in Matlab R2013b to calculate $(\hat{\alpha},\hat{c})$ and using Chi-square test to find the values Q given by equation (7), where n=62.

Table 12: Represents the Number of People in Each Cities in Sweden in 2015*

2172770	205162	161938	126841	94077	72259	62172	47494	42392
980347	182315	153952	125113	92132	71213	59089	46024	39826
631873	180221	151025	123464	88690	65492	58855	45451	38927
277441	177396	136167	109125	87476	64936	58669	45180	38565
268207	174306	134669	109073	85342	63848	57391	44215	38016
248558	170848	131661	103979	80282	63428	55576	43922	37417
206161	165153	127376	94845	79316	62176	47758	43425	

*The website of the Central Bureau of Statistics(SCB) Sweden (http://www.scb.se/en_/)

i- The Modified Moments Method:

For this method, the estimation of $(\hat{\alpha},\hat{c})$ are:

$\hat{\alpha}=1.3031$ $\hat{c}=369541.308$

The value of the test statistic for this variables is Q=10.19 with 5 deg. of freedom. This value of Q is compared with the tabulated critical values of a given sig. level of Chi-square test as shown in table (13)

Table 13: Represents the Critical Values of a Given sig. Level of Chi-Square Test and Decisions for Sweden

Chi-Squared					
Deg. of freedom Statistic P-Value	5 10.19 0.07004				
Significance Level	0.2	0.1	0.05	0.02	0.01
Critical Value	7.2893	9.2364	11.07	13.388	15.086
Decision	Reject	Reject	Accept	Accept	Accept

ii- The Standard Bayes Method

For this method, the estimation of $(\hat{\alpha},\hat{c})$ are:

$\hat{\alpha}=0.9984$ $\hat{c}=37417$

The value of the test statistic for this variables is Q=14.279 with 5 deg. of freedom. This value of Q is compared with the tabulated critical values of a given sig. level of Chi-square test as shown in table (14)

Table 14: Represents the Critical Values of a Given Sig. Level of Chi-Square Test and Decisions for Sweden

Chi-Squared					
Deg. of freedom Statistic P-Value	5 14.279 0.01393				
Significance Level	0.2	0.1	0.05	0.02	0.01
Critical Value	7.2893	9.2364	11.07	13.388	15.086
Decision	Reject	Reject	Reject	Reject	Accept

The following chart shows the values of the p.d.f. at the values of random variables $X_1, X_2, \ldots, X_{62}$ that are given in table (12).

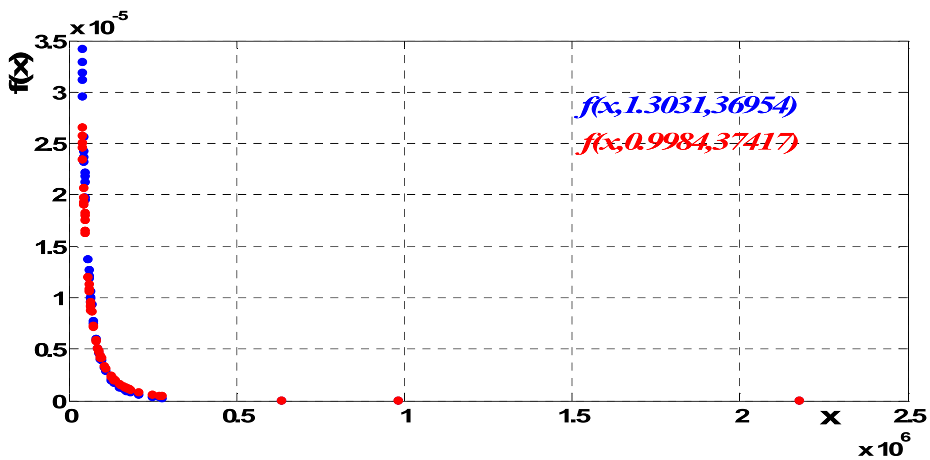

Figure 9: The Values of p.d.f at Specific Values of Random Variables

6-1-2 Maryland:

For this application, each number in table (15)represent the number of peoples for each Maryland state and its Counties at 2015 according to the population estimators, we regard each number as a random variables follow a Pareto distribution, where n=24.

Table 15: Represents the Number of People in Each Cities in Maryland State in 2015.*

559600	252000	245600	91650	74650	33900	50150	33250	102950
832050	309050	1036000	157100	30100	103600	39100	26900	52900
168550	625000	900350	113900	151200	20600			

*the source: united States Census Bureau (U.S. and World population Clock) http://www.census.gov/popclock/

i- The Modified Moments Method:

For this method, the estimation of $(\hat{\alpha}, \hat{c})$ are:

$$\hat{\alpha} = 1.0859 \qquad \hat{c} = 19810$$

The value of the test statistic for this variables is Q=12.945 with 1 deg. of freedom. This value of Q is compared with the tabulated critical values of a given sig. level of Chi-square test as shown in table (16)

Table 16: Represents the Critical Values of a Given Sig. Level of Chi-Square Test and Decisions for Maryland State

Chi-Squared					
Deg. of freedom Statistic P-Value	1 12.945 3.2083E-4				
Significance Level	0.2	0.1	0.05	0.02	0.01
Critical Value	1.6424	2.7055	3.8415	5.4119	6.6349
Decision	Reject	Reject	Reject	Reject	Reject

iv- The Standard Bayes Method:

For this method, the estimation of $(\hat{\alpha},\hat{c})$ are:

$$\hat{\alpha} = 0.503 \qquad \hat{c} = 20600$$

The value of the test statistic for this variables is Q=1.0217 with 2 deg. of freedom. This value of Q is compared with the tabulated critical values of a given sig. level of Chi-square test as shown in table (17)

Table 17: Represents the Critical Values of a Given Sig. Level of Chi-Square Test and Decisions for Maryland State

Chi-Squared					
Deg. of freedom	2				
Statistic	1.0217				
P-Value	0.6				
Significance Level	0.2	0.1	0.05	0.02	0.01
Critical Value	3.2189	4.6052	5.9915	7.824	9.2103
Decision	Accept	Accept	Accept	Accept	Accept

The following chart shows the values of the p.d.f. at the values of random variables $X_1, X_2, \ldots, X_{24}$ that are given in table (15).

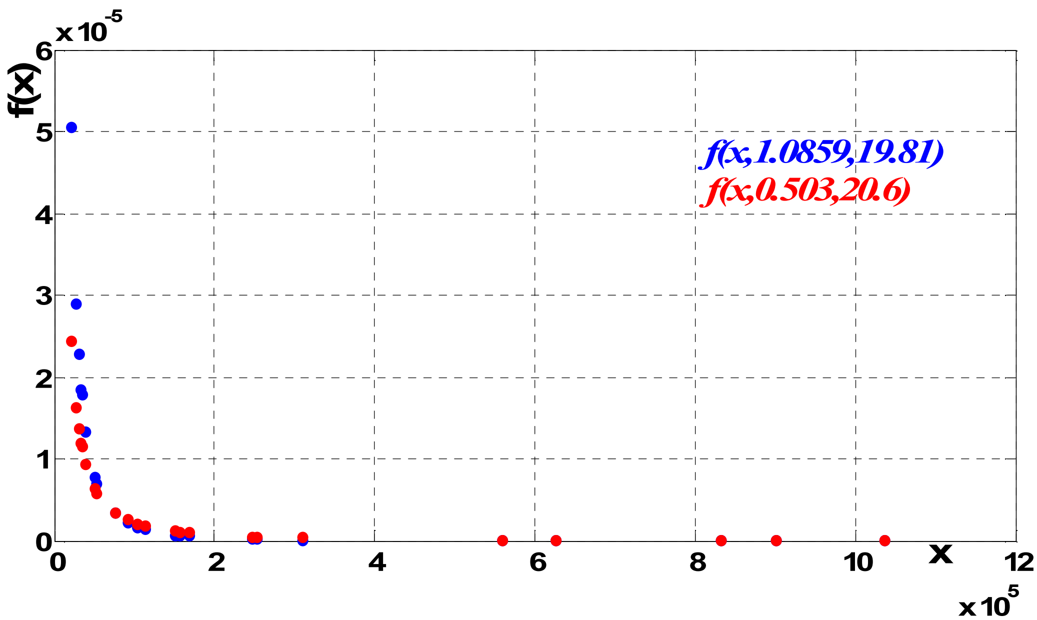

Figure 10: The Values of p.d.f at Determined Values of Random Variables

6-1-3 Alabama:

For this application, each number in table (18) represent the number of peoples for each Alabama state and its Counties at 2015 according to the population estimators, we regard each number as a random variables follow a Pareto distribution, where n=67.

Table 18: Represents the Number of People in Each Cities and Counties in Alabama State *

56223	118324	15354	81996	16832	53171	10629	22487	61337
204543	34637	53269	51229	32157	662177	19246	233033	88886
27182	26756	54720	41463	27284	14421	355475	123102	213585
23367	44236	12886	71996	8722	94572	19819	10084	13411
58466	13426	11232	83426	15807	33234	30549	21327	83262
10711	25765	38505	38281	17760	157737	96108	34108	41656
20347	13817	14196	105019	107711	93509	420180	23185	205652
66163	16510	11154	24354					

*the source: united States Census Bureau (U.S. and World population Clock) http://www.census.gov/popclock/

i- The Modified Moments Method:

For this method, the estimation of $(\hat{\alpha}\ ,\ \hat{c})$ are:

$$\hat{\alpha} = 1.1324 \qquad \hat{c} = 8607$$

The value of the test statistic for this variables is Q=47.662 with 5 deg. of freedom. This value of Q is compared with the tabulated critical values of a given sig. level of Chi-square test as shown in table (19)

Table 19: Represents the Critical Values of a Given Sig. Level of Chi-Square Test and Decisions for Alabama State

Chi-Squared					
Deg. of freedom Statistic P-Value	5 47.662 4.1638E-9				
Significance Level	0.2	0.1	0.05	0.02	0.01
Critical Value	7.2893	9.2364	11.07	13.388	15.086
Decision	Reject	Reject	Reject	Reject	Reject

iv- The Standard Bayes method:

For this method, the estimation of $(\hat{\alpha}, \hat{c})$ are:

$$\hat{\alpha} = 0.6237 \qquad \hat{c} = 8722$$

The value of the test statistic for this variables is Q=13.751 with 5 deg. of freedom. This value of Q is compared with the tabulated critical values of a given sig. level of Chi-square test as shown in table (20)

Table 20: Represents the Critical Values of a Given Sig. Level of Chi-Square Test and Decisions for Alabama State

Chi-Squared					
Deg. of freedom Statistic P-Value	5 13.751 0.01727				
Significance Level	0.2	0.1	0.05	0.02	0.01
Critical Value	7.2893	9.2364	11.07	13.388	15.086
Decision	Reject	Reject	Reject	Reject	Accept

The following chart shows the values of the p.d.f. at the values of random variables $X_1, X_2, \ldots, X_{67}$ that are given in table (18).

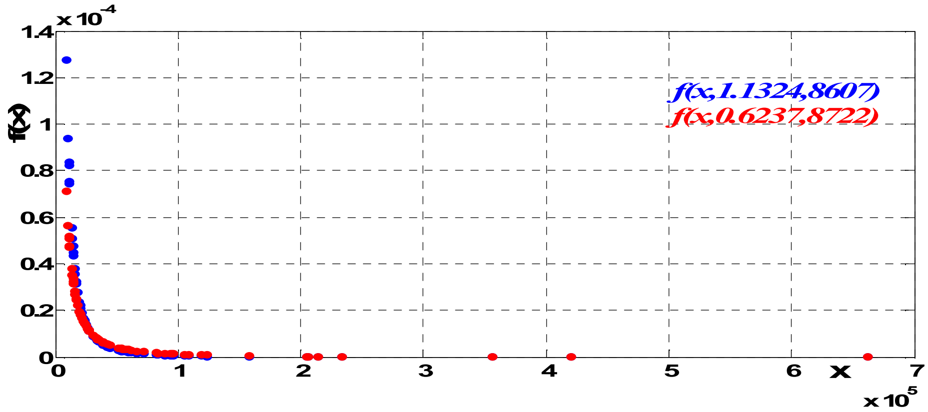

Figure 11: The Values of p.d.f at Determined Values of Random Variables

6-2 Application (2): Tropical Cyclones:

In this application we consider the tropical cyclones reaching TS (tropical storms) intensity or higher in Japan in 2014 where each random variable is tropical cyclones with assuming that these variables have a Pareto distribution, so we write a computer program in Matlab R2013b to calculate $(\hat{\alpha},\hat{c})$ and using Chi-square test to find the values Q by equation (7) and The comparison between them.

For this application the r.v. is the list of tropical cyclones reaching TS intensity or higher in 2014 in Japan as shown in table (21), where n=22.

Table 21: Represents the Number Tropical Cyclones in Japan in 2014*.

35	35	40	70	60	45	115	45	115
35	50	100	55	75	50	110	95	40
65	40	90	110					

* the source: Annual Report on the Activities of the RSMC Tokyo - Typhoon Center 2014, Japan Meteorological Agency (2015)

i- The Modified Moments Method:

For this method, the estimation of $(\hat{\alpha},\hat{c})$ are:

$\hat{\alpha} = 2.0426 \qquad \hat{c} = 34.221$

The value of the test statistic for this variables is Q=2.8081 with 2 deg. of freedom. This value of Q is compared with the tabulated critical values of a given sig. level of Chi-square test as shown in table (22)

Table 22: Represents the Critical Values of a Given Sig. Level of Chi-Square Test and Decisions for Tropical Cyclones

Chi-Squared					
Deg. of freedom Statistic P-Value	2 2.8081 0.2456				
Significance Level	0.2	0.1	0.05	0.02	0.01
Critical Value	3.2189	4.6052	5.9915	7.824	9.2103
Decision	Accept	Accept	Accept	Accept	Accept

ii- The Standard Bayes Method:

For this method, the estimation of $(\hat{\alpha}, \hat{c})$ are:

$\hat{\alpha} = 1.6202 \qquad \hat{c} = 35$

The value of the test statistic for this variables is Q=0.49268 with 2 deg. of freedom. This value of Q is compared with the tabulated critical values of a given sig. level of Chi-square test as shown in table (23)

Table 23: Represents the Critical Values of a Given Sig. Level of Chi-Square Test and Decisions for Tropical Cyclones

Chi-Squared					
Deg. of freedom	2				
Statistic	0.49268				
P-Value	0.78165				
Significance Level	0.2	0.1	0.05	0.02	0.01
Critical Value	3.2189	4.6052	5.9915	7.824	9.2103
Decision	Accept	Accept	Accept	Accept	Accept

The following chart shows the values of the p.d.f. at the values of random variables $X_1, X_2, \ldots, X_{22}$ that are given in table (21).

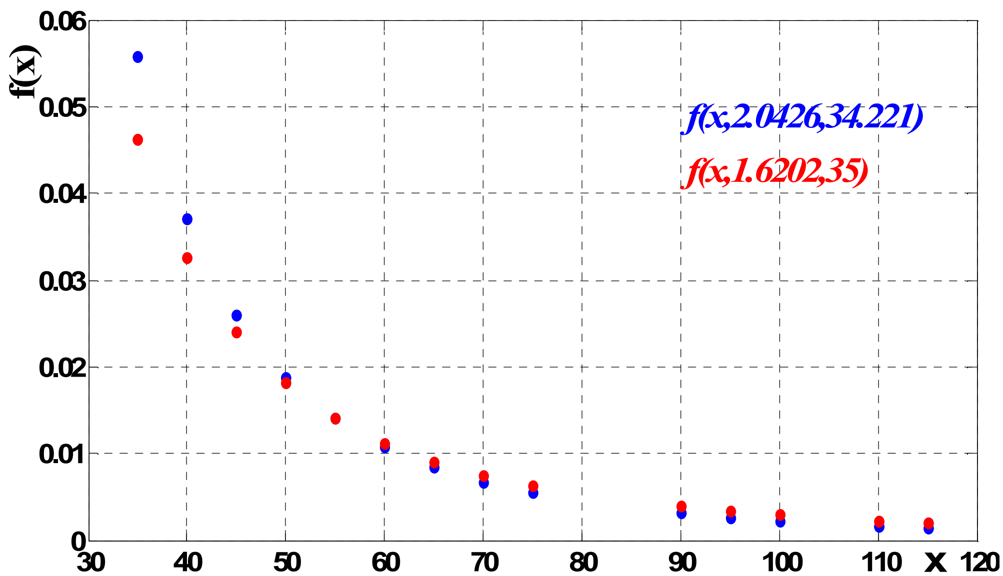

Figure 12: The Values of p.d.f at Specific Values of Random Variables

7. CONCLUSIONS

From our theoretical and practical study, the following conclusions were pointed out:

- B.M yield a small value of MSE comparing with M.M.M.
- The bias of the estimators decrease as the sample size increase.

- The fit of tropical cyclones coincide with Pareto distribution for each level of significant, while, the fit of Sweden population coincide with Pareto distribution only for some levels of significant.
- The goodness of fit test shows that the data of the population of Maryland state in U.S.A follows a Pareto distribution for each level of significant when the standard Bayes methods of estimation is used while the data does not appear to follow a Pareto distribution at any level of significant when the modified moment method of estimation is used.
- The data of Alabama state follow a Pareto distribution only at 0.01 level of significant and the standard Bayes method is used.
- The standard Bayes method of estimation is better than a modified moment in a sense of giving reasonable results for goodness of fit test.
- The procedures studied in this paper can be extended to include the bivariate and generalized Pareto distribution.

REFERENCES

1. L. J. Norman, K. Samuel, and N. Balakrishnan,(1994), Continuous Univariate Distributions,Volume 1, second Edition, Wiley.
2. Mood A.M., Graybill F.A. and Boess D.C. (1974) " Introduction to the theory of statistics", Third Edition, McGraw-Hill, London.
3. Johnson ,N.L & Kotz ,S. (1970) :Theory Of Point Estimation Wads Worth And Brooks Pacific Grove California.
4. Sinha, S. K. and Kate, B. K. (1980). "Life Testing and Reliability" Wiley Eastern limited
5. Sahoo, P.(2013), Probability And Mathematical Statistics University of Louisville, Louisville, KY 40292 USA.
6. Hodge. B.C. (1997), Estimating Pareto's Constant and Gini Coefficient of the Pareto Distribution, A thesis submitted in Partial Fulfillment of the Requirement of the Degree of Science in Mathematics, Department of Mathematical Science University of Nevada, Las Vegas, May 1997.