

ISSN 2411-3514 ONLINE ISSN 1997-2490PRINTED

Vol. 23, No04, pp. 21 –24, Year 2018

## ESTIMATING PARETO DISTRIBUTION PARAMETERS OF TYPE I

Instructor:HADIR H SHUBBAR

Directorate of Education El-Diwaniyah sed.hhhshb@gmail.com

DOI: 10.29350/jops.2018.23.4.908

#### **ABSTRACT:**

Some of methods like (Moment Method, Maximum Likelihood Method, Term Omission Method) are selected in this paper to estimate the parameters of Pareto (Type I) using simulation of different sample sizes (n=50, 100, 200). A Comparison made between these methods to determine the best method using mean square error criterion.

KEYWORDS: Pareto Distribution, the instinct, weighting mark, projection frontier, the error square.

#### **1. INTRODUCTION**

Pareto distribution is considered of the distributions that statisticians and economists are interested in studying because it relates to economic matters, such as the study of wealth and income distribution as well as traffic, hard drive crash rate for the calculator and the value of oil reserves in oil fields and other applications. (Fatima&Roohi, 2015)

In our research we will consider some methods of estimating Pareto distribution parameters (Type I) which Moment, Maximum Likelihood, Term Omission, data were generated using simulations for a number of samples 200, 100, 50=n and for optional values of Pareto distribution parameters (Type I).

The above methods have been compared using the Mean Squares Error (MSE) known as the following:

$$MSE = \frac{\sum_{i=1}^{n} {\{\hat{F}(t_i) - F(t_i)\}}^2}{n}$$

### 2. PARETO DISTRIBUTION

Pareto distribution is one of the continuous distributions and has several types and the first type will be chosen...

**Definition 1**: It is said to random variable X has a Pareto distribution if the probability density function is known as the following

$$f(x) = \frac{\alpha \beta^{\alpha}}{x^{\alpha + 1}} \qquad \qquad x \ge \beta \qquad \dots (1)$$

Whereas  $\alpha > 0$  represents the measurement parameter and  $\beta > 0$  represents the figure parameter.

**Definition 2**: If X is a random variable has a distribution of type I, then the cumulative mark of the random variable is defined as follow:

$$F(x) = 1 - \left(\frac{\beta}{x}\right)^{\alpha} \qquad x \ge \beta \qquad \qquad \dots (2)$$

We can generate data of Pareto distribution (Type I) from the uniform distribution over the interval (0, 1) using the following equation(Dalpatadu& Singh, 2015)

$$x = \frac{\beta}{\left(1 - F(x)\right)^{1/\alpha}}$$

For the purpose of estimating the distribution parameters we first have to find the estimated values of the cumulative mark F(x) using one of the methods shown in the table below:

Table 1. Estimation Methods of F(x) (Nwobi1 &Ugomma, 2014)

cumulative mark $F(t_i)$	Method	
$\frac{i}{n+1}$	Mean rank	
$\frac{i-0.3}{n+0.4}$	Median rank	
$\frac{i-0.5}{n}$	Symmetry CDF	

#### **3. ESTIMATION METHODS**

In our research, we will highlight on three ways, which we mentioned previously, to becompared to choose the best according to certain criteria.

#### **3-1 Moments Method**

The moments method is summarized through the forming equations in the same number of distribution parameters, these equations result from the equalization of the sample with the constraint of the community, and by solving these equations, we obtain the estimates of the distribution parameters.



**ISSN 2411-3514 ONLINE** ISSN 1997-2490PRINTED

Vol. 23, No04, pp. 21 -24, Year 2018

The community's reluctance to distribute Pareto is: (Para &Jan, 2018)

$$\mu'_{r} = \int_{\beta}^{\infty} x^{r} f(x) dx = \int_{\beta}^{\infty} x^{r} \frac{\alpha \beta^{\alpha}}{x^{\alpha+1}} dx = \alpha \beta^{\alpha} \left[ \frac{-\beta^{r-\alpha}}{r-\alpha} \right], \quad r=1, 2$$

Whereas r represents the number of distribution parameters.

Sample moment for distribution is:

 $M_r' = \frac{1}{n} \sum x_i'$ 

So the community moment be the first, is

$$\mu_1' = \frac{\alpha\beta}{\alpha - 1} \qquad \dots (3)$$

The first sample moments is

$$M_i' = \frac{1}{n} \sum x_i \qquad \dots (4)$$

The second community moments is

$$\mu_2' = \frac{\alpha \beta^2}{\alpha - 2} \qquad \dots (5)$$

The second sample moment is

$$M'_{2} = \frac{1}{n} \sum x_{i}^{2} \qquad \dots (6)$$

Andthrough equalize (3) with (4), we can get the first equation, and as well as equality (5) with (6), we get the second equation, to have two equations as follow:

$$\frac{\alpha\beta}{\alpha-1} = \frac{1}{n} \sum x_i \qquad \dots (7)$$

$$\frac{\alpha\beta^2}{\alpha-2} = \frac{1}{n}\sum x_i^2 \qquad \dots (8)$$

By solving equations (7) and (8) we obtain the Pareto distribution estimations as follow:

$$\hat{\alpha} = 1 + \sqrt{\frac{1 + \bar{x}^2}{s^2}}$$
,  $\hat{\beta} = \frac{\bar{x}(\alpha - 1)}{\alpha}$ ...(9)

Whereas the  $\bar{x}$  represents the arithmetic mean of the sample Represents the sample variation

## 3-2 Maximum Likelihood Method

Let X<sub>1</sub>, X<sub>2</sub>..., X<sub>n lea</sub> random sample of Pareto distribution, the weighting function for this sample is as follows: (Bee; Benedetti &Espa, 2013)

$$L = \prod_{i=1}^{n} f(x) = \prod_{i=1}^{n} \frac{\alpha \beta^{\alpha}}{x_{i}^{\alpha+1}} = \frac{\alpha^{n} \beta^{n\alpha}}{\prod_{i=1}^{n} x_{i}^{\alpha+1}} \qquad \dots (10)$$

By taking the normal logarithm of both sides, we get

$$LnL = n\ln\alpha - n\alpha\ln\beta - (\alpha+1)\sum\ln x_i \qquad \dots (11)$$

By taking the first derivation for  $\alpha$  for equation (11) and equal to zero we obtain:

$$\frac{n}{\alpha} - n\ln\beta - \sum\ln x_i = 0$$

After simplification we obtain the:

$$\hat{\alpha} = \frac{n}{\sum \ln\left(\frac{x_i}{\hat{\beta}}\right)} \qquad \dots (12)$$

Note that if we derive equation (11) for  $\beta$ , we cannot find an estimated value for parameter  $\beta$ , in which case we can adopt the minimum value of x, that is:

$$\hat{\beta} = \min(x_i) \quad \dots (13)$$

2

### 3-3 Term Omission Method

The TOMwas introduced by (Labban, 2005), using the cumulative function F(x) according to the following steps

$$x_{i} \qquad F(x_{i}) = 1 - \left(\frac{\beta}{x_{i}}\right)^{a}$$
$$x_{j} \qquad F(x_{j}) = 1 - \left(\frac{\beta}{x_{j}}\right)^{a}$$

We subtract each amount of 1 to obtain

B- We divide the second value by the first value to obtain

$$k = \frac{\left(\frac{\beta}{x_j}\right)^{\alpha}}{\left(\frac{\beta}{x_i}\right)^{\alpha}} = \frac{x_i^{\alpha}}{x_j^{\alpha}} = \left(\frac{x_i}{x_j}\right)^{\alpha}$$

Whereas k represents the results of dividing the second value on the first value



Vol. 23, No04, pp. 21 -24, Year 2018

C- Take the normal logarithm of the last equation to obtain

$$\ln k = \alpha \ln \left(\frac{x_i}{x_j}\right) \Longrightarrow \alpha = \frac{\ln k}{\ln \left(\frac{x_i}{x_j}\right)}$$

D- Write the estimated value of parameter  $\alpha$  as follows

$$\hat{\alpha}^{\ell} = \frac{\ln k}{\ln\left(\frac{x_{\ell}}{x_{j}}\right)} \qquad , \quad 1 \le \ell \le n-1$$

The estimated value  $\beta$  can also be obtained through the following relationship:

$$\hat{\beta}^{\ell} = x_i (1 - F(x_i))^{\nu_{\hat{\alpha}^{\ell}}}$$
,  $1 \le \ell \le n - 1$ 

If we assume that  $\theta = \{\alpha, \beta\}$  represents the estimated value of the two parameters, it will be the best estimate of Pareto distribution parameters is

$$\theta = \min_{1 \le i \le n-1} \left[ \sum_{i=1}^{n} \left\{ F(x_i; \alpha^i, \beta^i) - F(x_i; \alpha, \beta) \right\}^2 \right] \qquad \dots (14)$$

### **4- RESULTS**

For the purpose of clarifying and comparing the previous methods a different sample will be generated n=50, 100, using simulation and selected values for parameters  $\alpha$ =3· $\beta$ =1, and all methods under study will be compared using the standard of MSE.

$$MSE = \frac{\sum_{i=1}^{n} {\{\hat{F}(t_i) - F(t_i)\}^2}}{n}$$

Where  $\hat{F}(x_i)$  represents the cumulative function values according to the estimated values of the parameters, and  $F(x_i)$  is the estimated value of the cumulative function using the broker rang formula knowledge in form  $F(x_i) = \frac{i - 0.3}{n + 0.4}$ , the following are the tables of the neutron of the two parameters using

tables of the results of the estimation of the two parameters using the three methods under study for different values of sample n=50,100,200 and different values for the two parameters.

Table 2. Pareto distribution parameters estimations according to the studied methods for different samples 200, 100, 50=n and for selected values of parameters  $3=\alpha$ ,  $1=\beta$ 

MSE	ß	α	n	Method
0.124619	1.084736	3.913409		MM
0.126468	1.007225	3.122462	50	MLE
0.001365	0.999836	3.030392		TOM
0.100437	1.06437	3.653876		MM
0.101373	1.003067	3.070221	100	MLE
0.000649	0.999568	3.042329		TOM
0.092813	1.052867	3.49797	200	MM
0.093275	1.001605	3.029803	200	MLE

0.0002981.000213.009366TOMTable 3. Pareto distribution parameters estimators according to<br/>the studied methods of different samples n=50,100,200 and for<br/>selected values of parameters  $\alpha$ =5  $\beta$ =10

MSE	ß	α	n	Method
0.076689	10.11839	5.548958		MM
0.078567	10.0369	5.181931	50	MLE
0.001404	9.982824	4.986237		TOM
0.074213	10.02848	5.257742		MM
0.075104	10.01919	5.138934	100	MLE
0.000618	9.995394	5.076369		TOM
0.072497	9.99183	5.08059		MM
0.07295	10.01022	5.043346	200	MLE
0.000304	9.998912	5.007314		TOM

Table 3: Pareto distribution parameters estimators according to the studied methods of different samples n=50,100,200 and for selected values of parameters  $\alpha=3$   $\beta=4$ 

MSE	ß	α	n	Method
0.082435	3.079026	4.664562		MM
0.084284	3.015978	4.193836	50	MLE
0.001375	2.999164	4.075419		TOM
0.073174	3.042093	4.295594		MM
0.074094	3.007227	4.044268	100	MLE
0.000631	2.998776	3.989911		TOM
0.074789	3.014572	4.18744		MM
0.075268	3.003654	4.056583	200	MLE
0.000303	2.999263	4.03198		TOM

#### **5- DISCUSSION**

The results shown in the previous tables can be inferred by comparing the three methods (Moments Method MM, Maximum Likelihood Method MLE, Term Omission MethodTOM) to estimate Pareto distribution parameters(Type I) that:

1- Preference to the term omission method TOM of the other methods under study according to Mean Squared Error (MSE).

2- The Mean Square Error for all methods is not steady unlike the TOM which is decreasing by increasing the sample size.

### REFERENCES

- Bee, M., Benedetti, RandEspa, G. (2013). On maximum likelihood estimation of a Pareto mixture, *Springer Computational Statistics*, 28(1), 161-178.
- Dalpatadu, R. J.andSingh , A. K. (2015). Estimation of Parameters of the Pareto Distribution Using a Minimization Technique.*International Journal of Applied Science and Technology*, 5(4), 8-10.
- Fatima, A/andRoohi , A. (2015). Transmuted Exponentiated Pareto-I Distribution. *Pakistan Journal of Statistics*, 32(1), 63-80.
- Labban J.A. (2005). Modified and Simplified Method for Finite Difference (Divided). *Al-Qadisiya Journal for Science*, 10(2), 244-251.



ISSN 2411-3514 ONLINE ISSN 1997-2490PRINTED

Vol. 23, No04, pp. 21-24, Year 2018

- Nwobi1, F. N. andUgomma, C. A. (2014). A Comparison of Methods for the Estimation of Weibull Distribution Parameters.*Metodološkizvezki*, 11(1), 65-78.
- Para, B. A. & Jan, T.R. (2018). On Three Parameter Weighted Pareto Type II Distribution: Properties and Applications in Medical Sciences, *Applied Mathematics and Information Science Letter* 6(1), 13-26.