

# About Estimating Pareto Distribution Parameters

Assist. Prof. Dr. Nazar M. Al Sarraf<sup>1</sup>, Assist. Prof. Dr. Firas A. Mohana<sup>2</sup>,

Assist. Prof. Ghufraan I. Kamal<sup>3</sup>

<sup>1</sup> Al-Rafidain University Collage, nazaralsarraf@ruc.edu.iq

<sup>2</sup>College of Administration and Economics, Baghdad university, drfirasmohana@gmail.com

<sup>3</sup>College of Administration and Economics, Baghdad university, Ghufranka62@gmail.com

## Abstract

Pareto distribution is used in many economic, financial and social applications. This distribution is used for the study of income and wealth and the study of settlement in cities and villages and the study of the sizes of oil wells as well as in the field of communication through the speed of downloading files from the Internet according to their sizes. This distribution is used in mechanical engineering as one of the distributions of models of failure, stress and durability.

Given the practical importance of this distribution on the one hand, and the scarcity of sources and statistical research that deal with it, this research touched on some statistical characteristics such as derivation of its mathematical function, probability density function, cumulative distribution function, methods of estimating parameters, and the difficulties that researchers may face in dealing with these phenomena. The parameters were estimated in a number of methods, including the Maximum Likelihood (MLE), Ordinary Least Squares (OLS), Moment method (MOM), Relative Least squares (RELS) and Ridge regression (RR). In addition an algorithm has been proposed to improve the estimation parameters for this distribution. MSE was used to determine the best of these methods. Conclusions were presented in the light of this and appropriate proposals were decided upon.

**keywords:** Pareto distribution Maximum Likelihood (MLE), Ordinary Least Squares (OLS), Moment method (MOM), Relative Least squares (RELS), Ridge regression (RR).

## حول تقدير معالم توزيع باريتو

أ.م. غفران إسماعيل كمال  
جامعة بغداد - كلية الإدارة والاقتصاد

أ.م.د. فراس أحمد محمد  
جامعة بغداد - كلية الإدارة والاقتصاد

أ.م.د. نزار الصراف  
كلية الرافدين الجامعة

**المستخلص:** يستخدم توزيع باريتو في العديد من التطبيقات الاقتصادية والمالية والاجتماعية. يستخدم هذا التوزيع لدراسة الدخل والثروة ودراسة الاستيطان في المدن والقرى ودراسة أحجام آبار النفط وكذلك في مجال الاتصال من خلال سرعة تحميل الملفات من الإنترنت حسب أحجامها. يستخدم هذا التوزيع في الهندسة الميكانيكية كأحد توزيعات نماذج الفشل والضغط والمتانة. نظرا للأهمية العملية لهذا التوزيع من جهة، وندرة المصادر والبحوث الإحصائية التي تتناوله، فقد تطرق هذا البحث إلى بعض الخصائص الإحصائية مثل اشتقاق وظيفته الرياضية، ودالة كثافة الاحتمال، ودالة التوزيع التراكمي، وطرق تقدير المعالم والصعوبات التي قد يواجهها الباحثون في التعامل مع هذه الظواهر. تم تقدير المعالم في عدد من الطرق، بما في ذلك الإمكان الأعظم (MLE)، المربعات الصغرى العادية (OLS)، طريقة العزوم (MOM)، المربعات الصغرى النسبية (RELS) وانحدار الحرف (RR) بالإضافة إلى ذلك، تم اقتراح خوارزمية لتحسين معاملات التقدير لهذا التوزيع. تم استخدام MSE لتحديد أفضل هذه الطرق. وقدمت الاستنتاجات في ضوء ذلك وتم اتخاذ القرارات المناسبة.

**الكلمات المفتاحية:** توزيع باريتو، الإمكان الأعظم (MLE)، المربعات الصغرى العادية (OLS)، طريقة العزوم (MOM)، المربعات الصغرى النسبية (RELS)، انحدار الحرف (RR).

### 1. Estimating Parameters of a Two-parameter Pareto Distribution

The two-parameter Pareto distribution can be defined in terms of its cumulative distribution function as:

$$F_X(x) = \begin{cases} 1 - \left(\frac{\alpha}{x}\right)^\beta, & \alpha > 0, \beta > 0 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where:  $\beta$  = shape parameter, and  $\alpha$  = scale parameter.

By definition to obtain the density function (probability density function), we take the partial derivative of the cumulative density function with respect to  $x$  and obtain:

$$f(x) = \begin{cases} \frac{\alpha^\beta \beta}{x^{\beta+1}}, & 0 < \alpha \leq x. \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

#### 1.1. Method of Moments (MOM)

The  $k$ th moment of the Pareto distribution is given as:

$$E(X^k) = \int_{\alpha}^{\infty} x^k f(x) dx$$

In order to obtain the estimate of  $\alpha$  from a sample of  $n$  observations, we recall that the

probability of an observation greater than  $x$  is  $\left(\frac{\alpha}{x}\right)^\beta$ . Thus, the probability that all  $n$  sample

values  $x_1, \dots, x_n$  are greater than  $x$  is  $\left(\frac{\alpha}{x}\right)^{n\beta}$ . This is, therefore also the probability that the lowest sample value is greater than  $x$ .

Thus, the c.d.f of the lowest sample is

$$F_X(x) = 1 - \left(\frac{\alpha}{x}\right)^{n\beta} \tag{3}$$

we obtain method of moment's estimates as:

$$\therefore \hat{\alpha} = \frac{(n\beta - 1)x_0}{n\beta} \tag{4}$$

$$\hat{\beta} = \frac{n\bar{x} - x_0}{n(\bar{x} - x_0)} \tag{5}$$

Where  $x_0$  is the minimum value and  $\bar{x}$  is the mean.

#### 1.2. Method of Maximum Likelihood (MLE)

Given that  $X_1, \dots, X_n$  are random variables that follow the Pareto distribution, the likelihood function denoted by  $L = L(x, \alpha, \beta)$  for the sample is:

$$L = L(x, \alpha, \beta) = \frac{\alpha^{n\beta} \beta^n}{\left(\prod_{i=1}^n x_i\right)^{\beta+1}} \quad (6)$$

By taking the log of the likelihood function:

$$\ln L = n \ln \beta + n \beta \ln \alpha - (\beta + 1) \sum_{i=1}^n \ln x_i \quad (7)$$

To obtain the estimate for each parameter, we differentiate equation (7) w.r.t. each parameter and equate it to zero.

$$\frac{\partial \ln L}{\partial \beta} = \frac{n}{\beta} + n \ln \alpha - \sum_{i=1}^n \ln x_i = 0 \quad (8)$$

Simplifying the equation and making  $\beta$  the subject formula, we have:

$$\hat{\beta} = \frac{n}{\sum_{i=1}^n \ln x_i - n \ln \alpha} = \frac{n}{\sum_{i=1}^n (\ln x_i - \ln \alpha)} = \frac{n}{\sum_{i=1}^n \ln \left(\frac{x_i}{\alpha}\right)} \quad (9)$$

Since the likelihood function (L) is not bounded with respect to  $\alpha$ , a maximum likelihood estimate cannot be obtained for  $\alpha$  by differentiating L w.r.t.  $\alpha$ . since  $\alpha$  is the lower bound, we may maximize L subject to the constraint:

$$\hat{\alpha} \leq \min x_i \quad (10)$$

Thus, we see that L is maximized w.r.t.  $\alpha$  subject to (3.10) when

$$\hat{\alpha} = \min x_i \quad (11)$$

### 1.3. Least Squares Method (LSE)

For the estimation of probability distribution parameters, the least squares method (LSM) is extensively used in reliability engineering and mathematics problems.

Given that the cumulative density function of the Pareto distribution is given as:

$$F_X(x) = 1 - \left(\frac{\alpha}{x}\right)^\beta \quad (12)$$

the method of least squares estimates is given as

$$\hat{\beta}_1 = \frac{\sum \ln x_i \sum \ln(1 - F_X(x)) - n \sum (\ln x_i) (\ln(1 - F_X(x)))}{n \sum (\ln x_i)^2 - \left(\sum \ln(x_i)\right)^2} \quad (13)$$

$$\hat{\beta}_0 = \frac{1}{n} \sum \ln(1 - F_X(x)) - \frac{\hat{\beta}_1}{n} \sum \ln(x_i) \quad (14)$$

Where

$$Y_i = \ln(1 - F_X(x)); \quad \hat{\beta}_0 = \beta \ln \alpha;$$

$$\hat{\beta}_1 = -\beta; \quad X_i = \ln x_i.$$

**1.4 Relative Least Squares Method (R.L.S.M)**

The relative least squares estimators of a and d can be obtained by minimizing the sum of squares of the relative residuals, Pablo and Bruce (1992), w.r.t. a and d as follows

$$S = \sum_{i=1}^n \left( \frac{y_i - a - dx_i}{y_i} \right)^2 \tag{15}$$

$$S = \sum_{i=1}^n (1 - aw_i - dz_i)^2 \tag{16}$$

Where

$$w_i = \frac{1}{y_i} , \quad z_i = \frac{x_i}{y_i}$$

$$\sum_{i=1}^n w_i = a \sum_{i=1}^n w_i^2 + d \sum_{i=1}^n w_i z_i$$

Differentiating w.r.t, a and d then equate to zero

$$\sum_{i=1}^n z_i = a \sum_{i=1}^n w_i z_i + d \sum_{i=1}^n z_i^2$$

(17)

After simplification, we get

$$a = \frac{\sum_{i=1}^n w_i z_i \sum_{i=1}^n z_i - \sum_{i=1}^n w_i \sum_{i=1}^n z_i^2}{(\sum_{i=1}^n w_i z_i)^2 - \sum_{i=1}^n z_i^2 \sum_{i=1}^n w_i^2}$$

$$d = \frac{\sum_{i=1}^n w_i z_i \sum_{i=1}^n w_i - \sum_{i=1}^n z_i \sum_{i=1}^n w_i^2}{(\sum_{i=1}^n w_i z_i)^2 - \sum_{i=1}^n z_i^2 \sum_{i=1}^n w_i^2}$$

Where

$$w_i = \frac{1}{y_i} , \quad z_i = \frac{x_i}{y_i}$$

Also

$$y_i = \log t_i \quad \text{and} \quad x_i = \log(F(t_i))$$

$$a = \frac{\sum_{i=1}^n \left( \frac{1}{\log t_i} \right) \left( \frac{\log(F(t_i))}{\log t_i} \right) \sum_{i=1}^n \left( \frac{\log(F(t_i))}{\log t_i} \right) - \sum_{i=1}^n \left( \frac{1}{\log t_i} \right) \sum_{i=1}^n \left( \frac{\log(F(t_i))}{\log t_i} \right)^2}{\left( \sum_{i=1}^n \left( \frac{1}{\log t_i} \right) \left( \frac{\log(F(t_i))}{\log t_i} \right) \right)^2 - \sum_{i=1}^n \left( \frac{\log(F(t_i))}{\log t_i} \right)^2 \sum_{i=1}^n \left( \frac{1}{\log t_i} \right)^2}$$

$$d = \frac{\sum_{i=1}^n \left( \frac{1}{\log t_i} \right) \left( \frac{\log(F(t_i))}{\log t_i} \right) \sum_{i=1}^n \left( \frac{1}{\log t_i} \right) - \sum_{i=1}^n \left( \frac{1}{\log t_i} \right)^2 \sum_{i=1}^n \left( \frac{\log(F(t_i))}{\log t_i} \right)}{\left( \sum_{i=1}^n \left( \frac{1}{\log t_i} \right) \left( \frac{\log(F(t_i))}{\log t_i} \right) \right)^2 - \sum_{i=1}^n \left( \frac{\log(F(t_i))}{\log t_i} \right)^2 \sum_{i=1}^n \left( \frac{1}{\log t_i} \right)^2}$$

Also

$$a = \log(\beta) \quad \text{and} \quad d = \frac{1}{\gamma}$$

$$\hat{\beta} = \text{Antilog} \left\{ \frac{\sum_{i=1}^n \left( \frac{1}{\log t_i} \right) \left( \frac{\log(F(t_i))}{\log t_i} \right) \sum_{i=1}^n \left( \frac{\log(F(t_i))}{\log t_i} \right) - \sum_{i=1}^n \left( \frac{1}{\log t_i} \right) \sum_{i=1}^n \left( \frac{\log(F(t_i))}{\log t_i} \right)^2}{\left( \sum_{i=1}^n \left( \frac{1}{\log t_i} \right) \left( \frac{\log(F(t_i))}{\log t_i} \right) \right)^2 - \sum_{i=1}^n \left( \frac{\log(F(t_i))}{\log t_i} \right)^2 \sum_{i=1}^n \left( \frac{1}{\log t_i} \right)^2} \right\}$$

And

$$\hat{\gamma} = \frac{\left( \sum_{i=1}^n \left( \frac{1}{\log t_i} \right) \left( \frac{\log(F(t_i))}{\log t_i} \right) \right)^2 - \sum_{i=1}^n \left( \frac{\log(F(t_i))}{\log t_i} \right)^2 \sum_{i=1}^n \left( \frac{1}{\log t_i} \right)^2}{\sum_{i=1}^n \left( \frac{1}{\log t_i} \right) \left( \frac{\log(F(t_i))}{\log t_i} \right) \sum_{i=1}^n \left( \frac{1}{\log t_i} \right) - \sum_{i=1}^n \left( \frac{1}{\log t_i} \right)^2 \sum_{i=1}^n \left( \frac{\log(F(t_i))}{\log t_i} \right)}$$

### 1.5 The jackknife method

The jackknife and bootstrap are nonparametric computer-intensive techniques for estimating (e.g.) standard errors of the estimated parameters. The jackknife procedure consists of taking subsamples of the original sample of n independent observations by omitting a single observation at a time. Thus, each subsample consists of n - 1 observations formed by deleting a different observation from the sample. Parameter estimates are then calculated from these subsamples. Standard errors are determined from the variability across the n sets of parameter estimates. A more detailed description of the jackknife method proceeds as follows:

Let  $\hat{\theta}$  be the vector of parameter estimates obtained by MLE from the sample observations  $\{y_1, \dots, y_n\}$ . Divide the sample into g subgroups (at random if  $g < n$ ) of size k. Then from each subgroup, re-estimate  $\theta_j$  from the remaining  $(g - 1)k$  observations. This provides the g partial estimates  $\hat{\theta}^{(-j)}$ ,  $j = 1, \dots, g$ . Form the pseudo-values (the jackknife replications)

$$\theta_{*j} = g\hat{\theta} - (g - 1)\hat{\theta}^{(-j)}.$$

The jackknife estimate of  $\theta$  is the average of the jackknife replications  $\theta_{*j}$ , that is

$$\theta_{(J)} = \frac{1}{g} \sum_{j=1}^g \theta_{*j} = g\hat{\theta} - (g - 1)\tilde{\theta};$$

Where

$$\tilde{\theta} = \frac{1}{g} \sum_{j=1}^g \hat{\theta}^{(-j)}.$$

The corresponding estimated covariance matrix is

$$\hat{C}_J(\theta) = \frac{g-1}{g} \sum_{j=1}^g (\theta_{*j} - \theta_{(J)}) (\theta_{*j} - \theta_{(J)})' \tag{18}$$

### 1.6 Ridge Regression Method (RR)

Like the method of LSE, the ridge regression estimates of  $\beta_0$  and  $\beta_1$  can be obtained by minimizing the error sum of square for the model:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

subject to the single constraint that  $\phi = \beta_0^2 + \beta_1^2$ , where  $\phi$  is a finite positive constraint. Using the method of Lagrange's multiplier, we obtain:

$$\hat{\beta}_1 = \frac{(n+\lambda) \sum X_i Y_i - \sum X_i \sum Y_i}{(n+\lambda)(\lambda + \sum X_i^2) - (\sum x_i)^2} \tag{19}$$

$$\hat{\beta}_0 = \frac{\sum X_i \sum X_i Y_i - \sum Y_i (\lambda + \sum X_i^2)}{(\sum x_i)^2 - (n+\lambda)(\lambda + \sum X_i^2)} \tag{20}$$

The Least Squares Estimates are:

$$\hat{\beta} = -\hat{\beta}_1 = \frac{\sum \ln x_i \sum \ln(1 - F_X(x)) - (n+\lambda) \sum \ln x_i (\ln(1 - F_X(x)))}{(n+\lambda)(\lambda + \sum \ln x_i^2) - (\sum \ln x_i)^2} \tag{21}$$

$$\hat{\beta}_0 = \frac{1}{n+\lambda} \sum Y_i + \frac{\hat{\beta}_1}{n+\lambda} \sum \ln x_i \tag{22}$$

where:

$$\lambda = \frac{p\sigma^2}{\beta' \beta};$$

$p$  = The number of parameters of the distribution

$\beta' \beta$  = The covariance matrix

(Rasheed and Ahter, 2011).

## 1.7 The Proposed method

This method is based on the search for the best neighborhood by a higher error limit and lower target area for landmark estimates. Which we can summarize with the following algorithm.

- 1- The distribution parameters are estimated using the suggested methods
- 2- The best three methods are selected for parameter estimation(a,b)
- 3- All possible combinations of parameters estimates are formed
- 4- Specify a possible solution area for each parameter with an error limit of 0.05 above and below each parameter

$$\hat{\theta}_i - 0.05 \leq \theta_i \leq \hat{\theta}_i + 0.05$$

@The estimated parameter (a or b) is one of the methods used in this paper

- 5- 1000 acceptable solution points are generated for each solution area per parameter followed by generating combinations of these possible solutions
- 6- Each time a new model is generated, the previous models are compared with the MSE comparison scale

Table (1): represents the MSE values and the number of times the priority of the methods used according to the sample size of the model a = 0.2 and b = 2

	n=10		n=50		n=100	
Method	MSE	frq	MSE	frq	MSE	frq
MLE	0.000107	9	1.4E-05	6	7.57E-06	12
Moment	0.008734	0	0.008036	0	0.007446	0
Ols	0.000176	7	5.3E-05	3	2.9E-05	3
RELS	9.41E-05	10	1.69E-05	2	9.84E-06	13
RR	0.000256	0	0.000261	0	0.000249	0
jn	0.000105	6	1.48E-05	6	7.56E-06	3
semigen	1.17E-07	468	1.21E-08	483	1.1E-08	470
Best method	semigen	semigen	semigen	semigen	semigen	semigen

Note that the best method at the sample size n = 10 was the proposed method semigen where it had the smallest mse = 1.17E-0.7 and was the best 468 times out of 500 iterations, 93.6% of all the experiments that were saved here followed by the method RELS B MSE = 9.41E-05 was repeated 10 times, which achieved a test strength of 2%. At the sample size n = 50, the semigen method was also carrying less MSE = 1.21E-8, with a frequency of 96.5%, followed by MLE method with a frequency of 1.2% and finally at sample size n = 100 was the lowest MSE = 1.1E-8 and the semigen with 94% frequency followed by jn where MSE = 7.56E-6 and the second highest preference was RELS 2.6%.

Table (2): represents the values of MSE and the number of times the priority of the methods used according to the sample size of the model  $a = 0.4$  and  $b = 3$

	n=10		n=50		n=100	
Method	MSE	frq	MSE	frq	MSE	frq
MLE	0.000304	4	3.72E-05	2	1.99E-05	5
Moment	0.003784	0	0.003265	0	0.003114	0
Ols	0.000516	1	0.000144	0	7.92E-05	2
RELS	0.000257	5	4.47E-05	2	2.6E-05	2
RR	0.001021	0	0.001008	0	0.000977	0
jn	0.000352	3	3.85E-05	2	1.98E-05	2
semigen	2.45E-07	487	7.76E-09	495	6.06E-09	490
Best method	semigen	semigen	semigen	semigen	semigen	semigen

Note that the best method when the sample size  $n = 10$  was the proposed method semigen where it had the smallest mse = 2.45E-07 and was the best 487 times out of 500 repetitions, or 97.4% of each of the experiments that were saved here followed by the method RELS B MSE = 0.000257 was repeated 5 times, which achieved a test strength of 1%. At the sample size  $n = 50$ , the semigen method was also carrying less MSE = 7.76E-09, with a 99% frequency followed by jn method with a frequency of 0.2%. MSE = 6.06E-09 and the semigen with 98% frequency followed by jn where MSE = 1.98E-05 and the second highest preference was MLE with 1%.

Table (3): represents the values of MSE and the number of times the priority of the methods used according to the sample size of the model  $a = 0.6$  and  $b = 4$

	n=10		n=50		n=100	
Method	MSE	frq	MSE	frq	MSE	frq
MLE	0.000476	1	5.6E-05	1	2.98E-05	3
Moment	0.001915	0	0.001207	0	0.00112	0
Ols	0.000818	2	0.000221	0	0.000122	0
RELS	0.000395	4	6.76E-05	0	3.91E-05	2
RR	0.002082	0	0.00205	0	0.002009	0
jn	0.00054	3	5.78E-05	1	2.97E-05	1
semigen	1E-06	490	5E-09	498	5E-09	495
Best method	semigen	semigen	semigen	semigen	semigen	semigen

Note that the best method when the sample size  $n = 10$  was the proposed method semigen where it had the smallest mse = 1E-06 and was the best for 490 times out of 500 repetitions, or 98% of all experiments carried out here followed by the method RELS B MSE = 0.000395 It was repeated 4 times, which achieved a test strength of 0.8%. At the sample size  $n = 50$ , the semigen method was also carrying less MSE = 5E-09, with a repeat rate of 99.6%, followed by the MLE method with a frequency of 0.2%. 5e-09 = 99% semigen, followed by jn where MSE = 2.97E-05 and the second highest preference was MLE 0.6

## Conclusions

- 1- The proposed method is superior to the other methods used for all sample sizes and models
- 2- from the MSE at sample size  $n = 10$ , for all models, the second best estimation method was RELS
- 3- At the sample size  $n = 50$ , according to the MSE scale, MLE was the second best estimate method.
- 4- At sample size  $n = 100$ , jeknif was the second best method according to MSE and all models
- 5- All results in terms of the number of times the method outweighed the rest of the methods were identical to the previous conclusions other than the sample size 100 The MLE method was superior to jeknif in terms of how often it appeared as the best method.

## References

- [1] Aban IB, Meerschaert MM, Panorska AK. Parameter estimation for the truncated Pareto distribution. *J Am Stat Assoc.* 2006;101: 270–277
- [2] Afify EE. Estimation of parameters for Pareto distribution Faculty of Eng. Shibeen El Kom Menoufia University. Working Paper; 2003.
- [3] Arnold, B. C. (2015). *Pareto Distributions*. 2nd edition, CRC Press, Taylor and Francis Group
- [4] Atkinson, A. B. (2017). Pareto and the upper tail of the income distribution in the UK: 1799 to the present. *Economica* 84, 129–156
- [5] Bourguignon M, Silva RB, Zea LM, Cordeiro GM. The kumaraswamy Pareto distribution. *J Stat Theory Appl.* 2013;12: 129–144
- [6] Clark DR. A Note on the Upper-truncated Pareto distribution. *Casualty Actuarial Society E-Forum*, Winter. 2013. pp. 1–22.
- [7] Dhaene, G., Jochmans, K., Thuysbaert, B., 2006. Jackknife bias reduction for nonlinear dynamic panel data models with fixed effects, presented at the Econometric Society European Meeting 2006.
- [8] Kang S-B, Cho Y-S. Estimation of the parameters in a Pareto distribution by jackknife and bootstrap method. *J Inf Optim Sci.* 1997; 289–300
- [9] Kiviet, J., Phillips, G., 2002. Improved coefficient and variance estimation in stable first-order dynamic regression models, unpublished.
- [10] Kulldorff G, Vannman K. Estimation of the location and scale parameters of a Pareto distribution by linear functions of order statistics. *J Am Stat Assoc.* 1973;68: 218–22
- [11] Newman MEJ. Power laws, Pareto distributions and Zipf's law. *Contemp Phys.* 2005;46: 323–351.
- [12] Sankaran PG, Nair MT. On finite mixture of Pareto distributions. *Calcutta Stat Assoc Bull.* 2005;57: 225–226.
- [13] Quandt RE. Old and new methods of estimation and the Pareto distribution. *Metrika.* Springer; 1966;10: 55–82.