

## Reliability estimation for the randomly censored pareto distribution

Mahmoud M. Smadi<sup>1</sup>, Saiful Islam Ansari<sup>2</sup>, Ahmed A. Jaradat<sup>3</sup>

<sup>1,3</sup>Department of Mathematics and Statistics, Jordan University of Science and Technology, Jordan

<sup>2</sup>Department of Statistics, University of Tabuk, KSA

<sup>3</sup>Department of Family and Community Medicine, Arabian Gulf University, Bahrain

---

### Article Info

#### Article history:

Received Sep 17, 2018

Revised Mar 27, 2019

Accepted Apr 7, 2019

#### Keywords:

Maximum likelihood estimation

Pareto distribution

Random censoring

Reliability

Survival analysis

---

### ABSTRACT

Widespread applications of random censoring in life testing experiments to estimate reliability of engineering products or systems are available. Different parametric statistical models such as exponential, Rayleigh, Weibull and Maxwell distributions are used under random censoring scheme. In this paper, random censoring under Pareto distribution is considered. The maximum likelihood estimators (MLE's) of the model parameters and survival function were derived along with Fisher information matrix and asymptotic confidence intervals. A simulation study was performed to observe the behavior of the MLE's. The simulation results showed that the bias and MSE were reasonably small in all cases.

Copyright © 2019 Institute of Advanced Engineering and Science.  
All rights reserved.

---

### Corresponding Author:

Mahmoud M. Smadi,  
Department of Mathematics and Statistics,  
Jordan University of Science and Technology,  
Irbid, Jordan.  
Email: smadi@just.edu.jo

---

## 1. INTRODUCTION

In survival analysis and reliability theory, it is difficult to collect life time data for all the subjects under study because of time and cost constraints. The practitioners based on the model and available information use various types of censoring schemes. Conventional Type I and Type II censoring are common censoring schemes which has been used in many situations in survival analysis and reliability engineering. Different censoring schemes are proposed and studied in the literature. Random censoring is used in which the time of censoring is not fixed but taken as random [1]. Another censoring scheme is the progressive and hybrid censoring which has also considered in the literature [2-3].

There is a widespread application of random censoring in life testing experiments and clinical trials where both the survival and censoring times are random, Abu-Taleb et al [4] studied random censoring using exponential for survival and censoring times. Saleem and Aslam [5] studied Bayesian analysis of the Rayleigh survival time assuming random censor time. Danish and Aslam [6] studied Bayesian inference for randomly censored Weibull distribution. Krishna et al. [7] studied estimation in Maxwell distribution with randomly censored data.

The Pareto distribution has been widely used in analysis of life time data from reliability survival and engineering point of view [8]. Sindhu et al. [9] considered Bayesian estimation of the left censored data from the Pareto type II distribution. Asgharzadeh et al. [10] performed estimation and reconstruction based on left censored data from Pareto model. Usta and Gezer [11] studied reliability estimation in Pareto-I distribution based on progressively type II censored sample with binomial removals. Shou [12] studied estimation for the two-parameter Pareto distribution under progressive censoring with uniform removals.

Wu and Chang [13] studied inference in the Pareto distribution based on progressive type II censoring with random removals. Parsi et al. [14] studied simultaneous confidence interval for the parameters of Pareto distribution under progressive censoring.

In this paper the estimation problem for random censoring under Pareto distribution is considered, the survival times  $X_i$ 's and the censoring times  $T_i$ 's follow Pareto distributions. The maximum likelihood estimators for the model parameters and the survival functions are derived along with Fisher information matrix and asymptotic confidence intervals, The estimates of the survival function and asymptotic properties are given along with Fisher information matrix and asymptotic confidence intervals. The results of the simulation study are presented and discussed.

## 2. ESTIMATION OF THE PARAMETERS AND ASYMPTOTIC PROPERTIES

Let  $X_1, X_2, \dots, X_n$  be *i.i.d.* positive random variables (survival times) with the unknown survival function  $S_1(x) = P_\theta(X_1 > x)$ . Also, let  $T_1, T_2, \dots, T_n$  be *i.i.d.* positive random variables (censoring times) with the unknown survival function  $S_2(x) = P_\theta(T_1 > t)$ ;  $T$  is called a censoring variable. Assume that all  $X_i$ 's and  $T_i$ 's are independent variables. A randomly censored data set consists of  $n$  *i.i.d.* pairs  $(Y_i, D_i)$ , where  $Y_i = \min(X_i, T_i)$  and  $D_i$  is a binary random variable  $D_i = I(X_i, T_i)$ ;  $i = 1, 2, \dots, n$ ;  $D_i = 1$  if it is observed that  $X_i \leq T_i$  and  $D_i = 0$  if it is observed that  $X_i > T_i$ ;  $Y$  is called a censored survival or failure time random variable. Assuming  $X_i$ , the survival time,  $i=1, \dots, n$ , and  $T_i$ , censoring time, are independent Pareto distributions with probability density functions given by:

$$f_{X_i}(x_i) = \frac{\alpha}{x_i^{\alpha+1}} \quad 1 \leq x_i < \infty \quad (1)$$

and

$$f_{T_i}(t_i) = \frac{\beta}{t_i^{\beta+1}} \quad 1 \leq t_i < \infty \quad (2)$$

The survival function is

$$S(x_s) = P(X > x_s) = \frac{1}{x_s^\alpha} \quad (3)$$

Compared to the exponential distribution, the Pareto tail retain much more probability; and the survival function will decay slowly to zero; and it could be a good model to describe such data with heavy tailed behavior. [If  $(Y_i, D_i)$ ;  $i=1, \dots, n$ ] is observed, then the distribution of the Bernoulli random variable of  $D_i$  is

$$\begin{aligned} P(D_i = 1) &= P(X_i \leq T_i) = \int_1^\infty \int_{x_i}^\infty \frac{\alpha}{x_i^{\alpha+1}} \frac{\beta}{t_i^{\beta+1}} dt_i dx_i \\ &= \frac{\alpha}{\alpha + \beta} \end{aligned} \quad (4)$$

hence

$$P(D_i = 0) = \frac{\beta}{\alpha + \beta}$$

It can be shown that the joint probability density function of  $(Y_i, D_i)$  is given by

$$f_{Y_i, D_i}(y_i, d_i, \alpha, \beta) = (\alpha + \beta) y_i^{-(\alpha+\beta)-1} \left( \frac{\alpha}{\alpha + \beta} \right)^{d_i} \left( \frac{\beta}{\alpha + \beta} \right)^{1-d_i}, \quad y_i > 0, \alpha, \beta > 0 \quad (5)$$

Having in mind that,  $Y_i$  have a Pareto distribution with parameter  $(\alpha+\beta)$ , to derive the maximum likelihood estimators for  $\alpha$  and  $\beta$ , the likelihood function should be used and is given by

$$L(\alpha, \beta | y_1, \dots, y_n, d_1, \dots, d_n) = (\alpha + \beta)^n \left\{ \prod_{i=1}^n y_i^{-(\alpha+\beta)-1} \right\} \left( \frac{\alpha}{\alpha + \beta} \right)^{\sum_{i=1}^n d_i} \left( \frac{\beta}{\alpha + \beta} \right)^{n - \sum_{i=1}^n d_i} \quad (6)$$

$$L(\alpha, \beta | y_1, \dots, y_n, d_1, \dots, d_n) = (\alpha + \beta)^n \left\{ \prod_{i=1}^n y_i^{-(\alpha+\beta)-1} \right\} \left( \frac{\alpha}{\beta} \right)^{\sum_{i=1}^n d_i} \left( \frac{\beta}{\alpha + \beta} \right)^n \quad (7)$$

Taking logarithm of both sides

$$\begin{aligned} L^*(\alpha, \beta | y_1, \dots, y_n, d_1, \dots, d_n) &= \ln(\alpha + \beta)^n + \ln \left\{ \prod_{i=1}^n y_i^{-(\alpha+\beta)-1} \right\} + \ln \left( \frac{\alpha}{\beta} \right)^{\sum_{i=1}^n d_i} + \ln \left( \frac{\beta}{\alpha + \beta} \right)^n \\ &= \ln(\alpha + \beta)^n + \sum_{i=1}^n \ln(y_i^{-(\alpha+\beta)-1}) + (\sum_{i=1}^n d_i) \ln \alpha - (\sum_{i=1}^n d_i) \ln \beta + n \ln \beta - n \ln(\alpha + \beta) \end{aligned} \quad (8)$$

Taking partial derivatives of log likelihood function with respect to  $\alpha$  and  $\beta$ , equating to zero, and checking the second partial derivatives to be less than zero, the following maximum likelihood estimators can be obtained:

$$\begin{aligned} \hat{\alpha} &= \frac{\sum_{i=1}^n d_i}{\sum_{i=1}^n \ln y_i} \\ \hat{\beta} &= \frac{n - \sum_{i=1}^n d_i}{\sum_{i=1}^n \ln y_i} \end{aligned}$$

The above estimators are ratio estimators. The means and variances of these estimators are difficult to be obtain; thus, and the performance of the estimators will be investigated via simulation. The information matrix is given by:

$$I(\Theta) = \begin{bmatrix} \frac{n}{\alpha(\alpha + \beta)} & 0 \\ 0 & \frac{n}{\beta(\alpha + \beta)} \end{bmatrix}$$

Therefore, the asymptotic variance-covariance matrix, i.e. the inverse of the information can be expressed as:

$$\Sigma = I^{-1}(\Theta) = \begin{bmatrix} \frac{\alpha(\alpha + \beta)}{n} & 0 \\ 0 & \frac{\beta(\alpha + \beta)}{n} \end{bmatrix}$$

The asymptotic distribution of MLE's of the parameters  $\alpha$  and  $\beta$  can be written as

$$(\hat{\alpha} - \alpha, \hat{\beta} - \beta) \approx N_3(0, I^{-1}(\Theta))$$

Since  $I^{-1}(\Theta)$  involve the unknown parameters, they can be replaced by the corresponding ML estimates. Thus, the approximate  $(1-\alpha)$  100% confidence intervals for  $\alpha$  and  $\beta$ , respectively, are

$$\left[ \hat{\alpha} \pm \frac{Z_{\alpha/2}}{\sqrt{\hat{V}_{11}}} \right] \text{ and } \left[ \hat{\beta} \pm \frac{Z_{\alpha/2}}{\sqrt{\hat{V}_{22}}} \right]$$

where  $\hat{V}_{11}$  and  $\hat{V}_{22}$  are the estimated variances.

### 3. ESTIMATION OF THE SURVIVAL FUNCTION AND ASYMPTOTIC PROPERTIES

To obtain the maximum likelihood estimators of the survival functions,  $S_1(x)$  and  $S_2(y)$ . The following analysis can be performed. To fix  $x$  and  $y$  let

$$H(\Theta) = [S_1(x), S_2(y)] = [P(X_1 > x), P(Y_1 > y)]$$

Using invariance property of the MLE, the MLE of  $S_1(x)$  and  $S_2(y)$  can be expressed as:

$$\hat{P}_{\Theta}(X_1 > x) = \int_x^{\infty} \frac{\alpha}{z^{\alpha+1}} dz = x^{-\alpha} \tag{9}$$

and

$$\hat{P}_{\Theta}(Y_1 > y) = \int_y^{\infty} (\alpha + \beta) z^{-(\alpha+\beta)-1} dz = y^{-(\alpha+\beta)} \tag{10}$$

Thus the MLE of  $H(\Theta) = [x^{-\alpha}, y^{-(\alpha+\beta)}]$  can be expressed as:

$$\hat{H}(\Theta) = [x^{-\hat{\alpha}}, y^{-(\hat{\alpha}+\hat{\beta})}]$$

The above estimators are ratio estimators; means and variances of these estimators are difficult to obtain, thus, the performance of the estimators will be investigated via simulation. Under certain regularity conditions, the asymptotic distribution can be given by [14]

$$\sqrt{n}[H(\hat{\Theta}) - H(\Theta)] \rightarrow N(0, \frac{\partial H(\Theta)}{\partial \Theta} \sum (\frac{\partial H(\Theta)}{\partial \Theta})^T)$$

The matrix of partial derivatives

$$\frac{\partial}{\partial \Theta} H(\Theta) = \begin{bmatrix} Ax^{-\alpha} & By^{-(\alpha+\beta)} \\ 0 & By^{-(\alpha+\beta)} \end{bmatrix}$$

where  $A = -\frac{1}{\log x}$  and  $B = -\frac{1}{\log y}$

Thus the asymptotic variance-covariance matrix of  $\hat{H}(\Theta)$  is given by

$$\begin{bmatrix} \frac{\alpha + \beta}{n} (A^2 \alpha x^{-2\alpha} + B^2 \beta y^{-2(\alpha+\beta)}) & B^2 \beta (\alpha + \beta) y^{-2(\alpha+\beta)} \\ B^2 \beta (\alpha + \beta) y^{-2(\alpha+\beta)} & B^2 \beta (\alpha + \beta) y^{-2(\alpha+\beta)} \end{bmatrix}$$

**4. SIMULATION**

A Monte Carlo simulation was used to investigate the performance of the maximum likelihood estimators of  $\alpha$ ,  $\beta$ , and the survival function  $P(X_1 > x_s)$  developed in the previous sections. For the survival function the value of  $x_s$ , was considered as the mean of the survival time. The simulation study was conducted using different combinations of parameter values and sample sizes of 50,150, and 300. The simulation results were based on 1000 replicates. The means and root mean square errors (RMSE) of the maximum likelihood estimators were calculated.

With combination of different settings of  $\alpha$ ,  $\beta$  and sample size  $n$ , in every case we the following *algorithm* was used to obtain the simulation results:

Step 1: Generate  $x_1, x_2, \dots, x_n$  from Pareto( $\alpha$ )

Step 2: Generate  $T_1, T_2, \dots, T_n$  from Pareto( $\beta$ )

Step 3: Set  $Y_i = \min(X_i, T_i)$ ,  $i = 1, 2, \dots, n$

Step 4: For  $i=1, 2, \dots, n$ , set

$$D_i = \begin{cases} 1 & \text{if } X_i \leq T_i \\ 0 & \text{if } X_i > T_i \end{cases}$$

Step 5: Set the maximum likelihood estimates as follows

$$\hat{\alpha} = \frac{\sum_{i=1}^n d_i}{\sum_{i=1}^n \ln y_i} \quad \hat{\beta} = \frac{n - \sum_{i=1}^n d_i}{\sum_{i=1}^n \ln y_i}$$

S6: Repeat steps S1-S5 for 1000 iterations, compute the mean and mean square error of the maximum likelihood estimates for  $\alpha$  and  $\beta$ . The results are displayed in Tables 1-3. The following remarks can be drawn based on these results:

- The results indicate that the bias and MSE's were reasonably good in all cases.
- As the shape of the parameters  $\alpha$  and  $\beta$  increases (curves decays faster), the RMSE of the maximum likelihood estimators of the parameters  $\alpha$  and  $\beta$  increases.
- As the survival times curve decreases faster ( $\alpha$  increases) and censoring times curve decays slower ( $\beta$  decreases), the RMSE decreases. As survival times curve slowly decreases ( $\alpha$  decreases) and censoring times curve decays faster ( $\beta$  increases), the RMSE increases.
- As expected the biase and RMSE decrease as the sample sizes increases.

Means and RMSE of MLE'S of  $\alpha$  and  $\beta$ ,  $P(X_1 > x_s)$ ,  $\hat{P}(X_1 > x_s)$ ,  $n=50, n=150, n=300$  as shown in Tables 1, 2 and 3.

Table 1. Means and RMSE of MLE'S of  $\alpha$  and  $\beta$ ,  $P(X_1 > x_s)$ ,  $\hat{P}(X_1 > x_s)$ ,  $n=50$

$\alpha$	$\hat{\alpha}$	RMSE	$\beta$	$\hat{\beta}$	RMSE	$P(X_1 > x_s)$	$\hat{P}(X_1 > x_s)$	RMSE
2	2.02535	0.41439	2	2.04536	0.42040	0.25000	0.25532	0.06860
2	2.02742	0.49390	4	4.08861	0.73840	0.25000	0.25906	0.08334
4	4.06462	0.71462	2	2.04346	0.31641	0.31641	0.31690	0.06187
4	4.05071	0.82878	4	4.09072	0.84080	0.31641	0.32032	0.07190
4	4.03601	0.91022	6	6.14759	1.66677	0.31641	0.32350	0.08030
4	4.05484	0.98780	8	8.17723	1.47679	0.31641	0.32353	0.08665
6	6.10108	1.33143	4	4.07434	0.93154	0.33490	0.33552	0.06586
8	8.12923	1.42924	4	4.08692	1.02753	0.34361	0.34366	0.06245
6	6.07607	1.24312	6	6.12608	1.26120	0.33490	0.33843	0.07235
6	6.05401	1.36534	9	9.22139	1.75015	0.33490	0.34154	0.08071
6	6.08226	1.48170	12	12.26583	2.21519	0.33490	0.34148	0.08706
9	9.15162	1.69715	6	6.11149	1.39731	0.34644	0.34686	0.06604
12	12.19384	2.14386	6	6.17038	1.54130	0.35200	0.35192	0.06256

Table 2. Means and RMSE of MLE'S of  $\alpha$  and  $\beta$ ,  $P(X_1 > x_s)$ ,  $\hat{P}(X_1 > x_s)$ , n=150

$\alpha$	$\hat{\alpha}$	RMSE	$\beta$	$\hat{\beta}$	RMSE	$P(X_1 > x_s)$	$\hat{P}(X_1 > x_s)$	RMSE
2	2.01499	0.23961	2	2.01413	0.23135	0.25000	0.25076	0.04059
2	2.01753	0.28949	4	4.02014	0.40678	0.25000	0.25186	0.04918
4	4.02558	0.41826	2	2.02200	0.28626	0.31641	0.31633	0.03733
4	4.02999	0.47922	4	4.02826	0.46271	0.31641	0.31662	0.04662
4	4.03125	0.53591	6	6.03428	0.64446	0.31641	0.31752	0.04807
4	4.03507	0.57899	8	8.04029	0.81757	0.31641	0.31750	0.05160
6	6.03479	0.65988	4	4.04445	0.51701	0.33490	0.33490	0.03951
8	8.05115	0.83652	4	4.04400	0.57353	0.34361	0.34337	0.03765
6	6.04498	0.71883	6	6.04239	0.69406	0.33490	0.33497	0.04294
6	6.04687	0.80387	9	9.05144	0.96668	0.33490	0.33557	0.04835
6	6.05261	0.86840	12	12.06042	1.22635	0.33490	0.33579	0.05192
9	6.05261	0.86840	6	6.06660	0.77551	0.34644	0.34662	0.03961
12	12.07673	1.25479	6	6.06600	0.86000	0.35200	0.35171	0.03771

Table 3. Means and RMSE of MLE'S of  $\alpha$  and  $\beta$ ,  $P(X_1 > x_s)$ ,  $\hat{P}(X_1 > x_s)$ , n=300

$\alpha$	$\hat{\alpha}$	RMSE	$\beta$	$\hat{\beta}$	RMSE	$P(X_1 > x_s)$	$\hat{P}(X_1 > x_s)$	RMSE
2	2.00122	0.16211	2	2.01420	0.17441	0.25000	0.25135	0.02780
2	2.00160	0.19295	4	4.02854	0.30462	0.25000	0.25192	0.03300
4	4.00474	0.28276	2	2.01401	0.20940	0.31641	0.31701	0.02553
4	4.00244	0.32421	4	4.02840	0.34881	0.31641	0.31755	0.02921
4	4.00807	0.35491	6	6.03764	0.47956	0.31641	0.31730	0.03189
4	4.00319	0.38589	8	8.05708	0.60924	0.31641	0.31804	0.03466
6	6.00618	0.44772	4	4.02682	0.38529	0.33490	0.33562	0.02703
8	8.00949	0.56551	4	4.02802	0.41801	0.34360	0.34414	0.02575
6	6.00365	0.48632	6	6.04260	0.52322	0.33490	0.33598	0.02940
6	6.01210	0.53237	9	9.05646	0.71934	0.33490	0.33572	0.03209
6	6.00479	0.57884	12	12.08562	0.91386	0.33490	0.33644	0.03488
9	9.00927	0.67150	6	6.04023	0.57794	0.34644	0.34644	0.34731
12	12.01424	0.84827	6	6.04203	0.62821	0.35200	0.35251	0.02570

## 5. CONCLUSION

The MLE's for model parameters and survival functions were derived along with the Fisher information matrix and asymptotic confidence intervals. A simulation study verified the behavior of the MLE's. The results showed that the bias and MSE were reasonably small in all cases. It is shown that as the shape parameters for Pareto distributions increases (i.e. curves decays faster), the RMSE of the MLE's increases. As the survival times curve decreases faster and censoring times curve decays slower, the RMSE decreases and vice versa.

## REFERENCES

- [1] L. M. Leemis, "Reliability, probabilistic models and statistical methods," PrenticeHall, Englewood Cliffs, New Jersey, 1995.
- [2] N. Balakrishnan and R. Aggarwala, "Progressive Censoring: Theory, Methods and Applications," Birkhauser, Boston, 2000.
- [3] N. Balakrishnan, D. Debasis Kundu, "Hybrid censoring: Models, inferential results and applications," Computational Statistics & Data Analysis, 57 (1), 166-209, 2013.
- [4] A. A. Abu-Taleb, and M. M. Smadi, "Asymptotic efficiencies of the survival functions estimators for the exponential distribution," International Mathematical Forum, No. 38, 1861-1869, 2006.
- [5] M. Saleem and M. Aslam, "On Bayesian analysis of the Rayleigh survival time assuming the random censor time," Pak. J. Statistics, Vol. 25, No. 2, 71-82, 2009.
- [6] M. Y. Danish and M. Aslam, "Bayesian inference for the randomly censored Weibull distribution," Journal of Statistical Computing and Simulation, Vol. 84, No. 1, 215-230, 2014.
- [7] H. Krishna, Vivekanand, K. Kumar, "Estimation in Maxwell distribution with randomly censored data," Journal of Statistical Computing and Simulation, Vol. 85, No. 17, 3560-3578, 2015.
- [8] N. L. Johnson., S. Kotz, N. Balakrishnan, "Continuous Univariate Distributions," Vol. 1 2nd ed. John Wiley Balakrishnan & Sons, New York, 1994.
- [9] T. N. Sindhu, M. Asalam., and A. Shafiq, "Analysis of the left censored data from the Pareto type II distribution," Caspian Journal of Applied Sciences Research, 2(7), 53-62, 2013.
- [10] M. Asgharzadeh, M. Mohammadpour M., Z. M. Ganji, "Estimation and reconstruction based on left censored data from Pareto model," JIRSS, Vol. 13, No.2, 151-175, 2014.

- [11] U Iihan and H. Gezer, "Reliability estimation in Pareto-I distribution based on progressively type II censored sample with binomial removals," *Journal of scientific research and development*, Vol. 2, No. 12, 108-113, 2015.
- [12] W. Shuo, "Estimation for the two-parameter Pareto distribution under progressive censoring with uniform removals," *Journal of statistical computing and simulation*, Vol. 73(2), 125-134, 2008.
- [13] Wu S. and Chang C., "Inference in the Pareo distribution based on progressive type II censoring with random removals," *Journal of applied statistics*, Vol. 30, No. 2, 163-172, 2003.
- [14] S. Parsi., M. Ganjali., and S. Farsipou., "Simultaneous confidence interval for the parameters of Pareto distribution under progressive censoring," *Communication in Statistics-Theory and Methods*, 39: 94-106, 2010.

## BIOGRAPHIES OF AUTHORS



**Mahmoud M. Smadi**, he recieved his B.S and M.S degrees in statistics from Yarmouk University in 1983 and 1986, respectevy. He recieved his Ph.D in statistics from Colorado State University in 1997. Since 1997 he worked for the Department of Mathematics & Statistics at Jordan University of Science & Technology, Jordan. He is currently associate professor at Department of Mathematics & Statistics at Jordan University of Science & Technology. His research interests include survival and reliability analysis, biostatistics and generelized distributions.



**Saiful Islam Ansari** received his M.Phil. and Ph.D. in 2008 and 2011 respectively from Aligarh Muslim University, Aligarh, India. Currently he is working as Assistant Professor in the department of Statistics at University of Tabuk, Tabuk, KSA. His research areas of interest are Stochastic Programming, Reliability Optimization, Distribution Theory, Statistical computing etc.



**Ahmed A. Jaradat**, he received his B.S and M.S degrees in statistics from Yarmouk University in 1984 and 1986, respectively. He received his Ph.D. in statistics from North Dakota State University, the USA in 1996. Currently, he is working as Associate Professor at the Department of Family and Community Medicine at Arabian Gulf University, Bahrain. His research interests include Biostatistics, Survival and reliability analysis.