

New approach for analysis of progressive Type-II censored data from the Pareto distribution

Jung-In Seo^a, Suk-Bok Kang^{1,b}, Ho-Yong Kim^b

^aDepartment of Statistics, Daejeon University, Korea;

^bDepartment of Statistics, Yeungnam University, Gyeongsan, Korea

Abstract

Pareto distribution is important to analyze data in actuarial sciences, reliability, finance, and climatology. In general, unknown parameters of the Pareto distribution are estimated based on the maximum likelihood method that may yield inadequate inference results for small sample sizes and high percent censored data. In this paper, a new approach based on the regression framework is proposed to estimate unknown parameters of the Pareto distribution under the progressive Type-II censoring scheme. The proposed method provides a new regression type estimator that employs the spacings of exponential progressive Type-II censored samples. In addition, the provided estimator is a consistent estimator with superior performance compared to maximum likelihood estimators in terms of the mean squared error and bias. The validity of the proposed method is assessed through Monte Carlo simulations and real data analysis.

Keywords: Pareto distribution, progressive Type-II censored sample, weighted linear regression

1. Introduction

Since its introduction in Cohen (1963), the progressive Type-II censoring scheme has gained considerable popularity and has been extensively studied in, for instance, Balakrishnan *et al.* (2003), Wang (2008), and Seo and Kang (2015). Under this censoring scheme, it is assumed that there are n randomly selected units on a life test. Once the first failure occurs, R_1 units are randomly removed from the $n - 1$ surviving units. Subsequently, following the second observed failure, R_2 units are randomly removed from the $n - 2 - R_1$ surviving units, and the procedure is continued in this manner. Finally, at the time of the m^{th} observed failure, all remaining $R_m = n - m - R_1 - \dots - R_{m-1}$ units are removed from the test. Here, the number of failures m and R_i ($i = 1, \dots, m$) are prefixed. It should be noted that the case $m = n$, where $R_1 = \dots = R_m = 0$ corresponds to the complete sample, whereas the case $R_1 = \dots = R_{m-1} = 0$, $R_m = n - m$ corresponds to a conventional Type-II censoring scheme.

In this study, a new estimation method based on the regression framework is proposed under the progressive Type-II censoring scheme. It focuses on estimations of the unknown parameters in the Pareto distribution. The cumulative distribution function (cdf) and the probability distribution function (pdf) of the random variable X with the Pareto distribution are

$$F(x) = 1 - \left(\frac{\theta}{x}\right)^\lambda,$$
$$f(x) = \lambda\theta^\lambda x^{-(\lambda+1)}, \quad x > \theta, \lambda > 0, \theta > 0,$$

¹ Corresponding author: Department of Statistics, Yeungnam University, 280 Daehak-Ro, Gyeongsan, Gyeongbuk 38541, Korea. E-mail: sbkang@ynu.ac.kr

respectively, where θ is the scale parameter and λ is the shape parameter. Lu and Tao (2007) proposed a new estimation by a weight least square method to estimate unknown parameters of the Pareto distribution. Kim *et al.* (2017) proposed an estimation using a pivotal quantity adopted from the regression framework for obtaining a consistent estimator for the shape parameter in the Pareto distribution. In the present study, these methods are extended to the progressive Type-II censoring scheme, and a consistent estimator of the shape parameter is obtained that do not depend on the nuisance parameter θ and are superior to maximum likelihood estimators (MLEs) in terms of the mean squared error (MSE) and bias.

The paper is organized as follows. In Section 2, existing methods for estimating the unknown parameters of the Pareto distribution are presented. In Section 3, a new approach based on the weighted least squares method is proposed. In Section 4, a Monte Carlo simulation is conducted, and real data analysis is performed to assess the proposed method. Finally, Section 5 concludes the paper.

2. Revisit to classical inference

This section provides the results of existing inferences for the scale parameter θ and the shape parameter λ of the Pareto distribution (Balakrishnan and Aggarwala, 2000). Let $X_{1:m:n} \leq \dots \leq X_{m:m:n}$ be a progressive Type-II censored sample from the Pareto distribution with the censoring scheme (R_1, \dots, R_m) . Then, the corresponding likelihood function is given by

$$L(\theta, \lambda) \propto \lambda^m \theta^{\lambda n} \prod_{i=1}^m x_{i:m:n}^{-\lambda(1+R_{1:m:n})-1}. \quad (2.1)$$

The MLE $\hat{\theta}$ is the first order statistic of a progressive Type-II sample because the likelihood function (2.1) is an increasing function of θ . In addition, the MLE of λ is given by

$$\hat{\lambda} = \frac{m}{\sum_{i=1}^m (1 + R_i) \log(X_{i:m:n}) - n \log X_{1:m:n}},$$

which has the inverse gamma distribution with parameters $(m - 1, \lambda m)$. It can be easily shown by the pivotal quantity

$$W(\lambda) = \lambda \sum_{i=1}^m (1 + R_i) \log(X_{i:m:n}) - n \log X_{1:m:n}$$

that has the gamma distribution with parameters $(m - 1, 1)$. Therefore, the MLEs $\hat{\theta}$ and $\hat{\lambda}$ are biased estimators. Alternately, Balakrishnan and Aggarwala (2000) provided estimators of θ and λ , given by

$$\begin{aligned} \hat{\theta}_U &= \left[1 - \frac{m}{n(m-1)\hat{\lambda}} \right] \hat{\theta}, \\ \hat{\lambda}_U &= \frac{(m-2)}{m} \hat{\lambda}, \end{aligned}$$

respectively. Note that both $\hat{\theta}_U$ and $\hat{\lambda}_U$ are unbiased and consistent estimators with the variances

$$\begin{aligned} \text{Var}(\hat{\theta}_U) &= \frac{\theta^2 m}{\lambda n(m-1)(\lambda n - 2)}, \\ \text{Var}(\hat{\lambda}_U) &= \frac{\lambda^2}{m-3}, \end{aligned}$$

respectively.

3. Estimation based on regression framework

In this section, a new approach is proposed to obtain an estimator that is superior to the MLE in terms of MSE and bias.

Let

$$\begin{aligned} Y_{i:m:n} &= -\log [1 - F(x_{i:m:n})] \\ &= \lambda \log \left(\frac{X_{i:m:n}}{\theta} \right), \quad i = 1, \dots, m. \end{aligned}$$

Then, $Y_{1:m:n} \cdots Y_{m:m:n}$ is a progressive Type-II censored sample from the standard exponential distribution. Consider the spacing (Viveros and Balakrishnan, 1994)

$$S_i = (Y_{i:m:n} - Y_{i-1:m:n}) \sum_{j=i}^m (R_j + 1), \quad i = 1, \dots, m \quad (Y_{0:m:n} \equiv 0),$$

which are independent and identically distributed standard exponential random variables. Then,

$$\begin{aligned} D_{i:m:n} &= Y_{i:m:n} - Y_{i-1:m:n} \\ &= \lambda \log \left(\frac{X_{i:m:n}}{X_{i-1:m:n}} \right), \quad i = 2, \dots, m, \end{aligned}$$

have the exponential distribution with the mean

$$E(D_{i:m:n}) = \left[\sum_{j=i}^m (1 + R_j) \right]^{-1},$$

and can lead to the following linear regression model:

$$E(D_{i:m:n}) = \lambda \log \left(\frac{X_{i:m:n}}{X_{i-1:m:n}} \right) + \varepsilon_i, \quad i = 2, \dots, m, \tag{3.1}$$

where ε_i is the error term with $E(\varepsilon_i) = 0$. It should be noted that (3.1) can be considered a simple regression line with no intercept. That is, the model only has the shape parameter λ . Then, an estimator can be obtained that does not depend on the nuisance parameter θ , as:

$$\hat{\lambda}_{w/1} = \frac{\sum_{i=2}^m w_i E(D_{i:m:n}) \log \left(\frac{X_{i:m:n}}{X_{i-1:m:n}} \right)}{\sum_{i=2}^m w_i \left[\log \left(\frac{X_{i:m:n}}{X_{i-1:m:n}} \right) \right]^2}$$

by minimizing the squared distance

$$\sum_{i=2}^m w_i \left[E(D_{i:m:n}) - \lambda \log \left(\frac{X_{i:m:n}}{X_{i-1:m:n}} \right) \right]^2,$$

where w_i is the weight on each data point. However, the problem with this approach is that it is not easy to find the weight w_i that makes $\hat{\lambda}_{w/1}$ an unbiased estimator or a consistent estimator. Instead, the idea of Kim *et al.* (2017) is extended to the progressive Type-II censoring scheme here.

Let

$$D_{i:m:n}^* = Y_{i:m:n} - Y_{1:m:n}, \quad i = 2, \dots, m.$$

Then, by the same argument, another estimator of λ is obtained as

$$\hat{\lambda}_{wl2} = \frac{\sum_{i=2}^m w_i E(D_{i:m:n}^*) \log\left(\frac{X_{i:m:n}}{X_{1:m:n}}\right)}{\sum_{i=2}^m w_i \left[\log\left(\frac{X_{i:m:n}}{X_{1:m:n}}\right)\right]^2},$$

where

$$E(D_{i:m:n}^*) = \sum_{j=2}^i \left[\sum_{k=j}^m (1 + R_k) \right]^{-1}, \quad 2 \leq i \leq m,$$

by Theorem 7.2.1 in Balakrishnan and Cramer (2014). To find the weight w_i that makes $\hat{\lambda}_{wl2}$ a consistent estimator, the following lemma is required:

Lemma 1. *Let*

$$Q_{i:m:n} = \frac{D_{i:m:n}^* E(D_{i:m:n}^*) - E^2(D_{i:m:n}^*)}{\text{Var}(D_{i:m:n}^*)}, \quad (3.2)$$

$$T_{i:m:n} = \frac{D_{i:m:n}^{*2} - E^2(D_{i:m:n}^*)}{\text{Var}(D_{i:m:n}^*)}, \quad (3.3)$$

$$U_{i:m:n} = \frac{E^2(D_{i:m:n}^*)}{\text{Var}(D_{i:m:n}^*)},$$

where

$$\text{Var}(D_{i:m:n}^*) = \sum_{j=2}^i \left[\sum_{k=j}^m (1 + R_k) \right]^{-2}, \quad 2 \leq i \leq m,$$

by Theorem 7.2.1 in Balakrishnan and Cramer (2014). Then,

- (a) $\frac{1}{m^2} \sum_{i=2}^m Q_{i:m:n}$ converges to zero in probability as $m \rightarrow \infty$.
- (b) $\frac{1}{m^2} \sum_{i=2}^m T_{i:m:n}$ converges to zero in probability as $m \rightarrow \infty$.
- (c) $\frac{1}{m^2} \sum_{i=2}^m U_{i:m:n}$ does not converges to zero as $m \rightarrow \infty$.

Proof: It is clear that

$$E\left(\frac{1}{m^2} \sum_{i=2}^m Q_{i:m:n}\right) = 0 \quad \text{and} \quad E\left(\frac{1}{m^2} \sum_{i=2}^m T_{i:m:n}\right) = \frac{m-1}{m^2}.$$

Then, both (3.2) and (3.3) converge in the mean to 0. This implies convergence in probability, Karr (1993). In addition,

$$\begin{aligned} \frac{1}{m^2} \sum_{i=2}^m U_{i:m:n} &= \frac{1}{m^2} \sum_{i=2}^m \frac{\left\{ \sum_{j=2}^i \left[\sum_{k=j}^m (1 + R_k) \right]^{-1} \right\}^2}{\sum_{j=2}^i \left[\sum_{k=j}^m (1 + R_k) \right]^{-2}} \\ &\geq \frac{1}{m^2} \sum_{i=2}^m \frac{\left\{ \sum_{j=2}^i \left[m - 1 + \sum_{k=j}^m R_k \right]^{-1} \right\}^2}{\sum_{j=2}^i (m - i + 1)^{-2}} \\ &= \frac{1}{m^2} \sum_{i=2}^m (i - 1) \left(\frac{m - i + 1}{m - 1 + \sum_{k=j}^m R_k} \right)^2 \\ &= \frac{1}{m^2 \left(m - 1 + \sum_{k=j}^m R_k \right)^2} \left(m^2 \sum_{j=1}^m j + \sum_{j=1}^m j^3 - 2m \sum_{j=1}^m j^2 \right), \end{aligned}$$

which converges to a constant as $m \rightarrow \infty$. Therefore, it completes the proof. □

Theorem 1. Let $w_i = 1/\text{Var}(D_{i:m:n}^*)$. Then, the estimator $\hat{\lambda}_{wI2}$ is a consistent estimator.

Proof: The estimator $\hat{\lambda}_{wI2}$ can be written as

$$\begin{aligned} \hat{\lambda}_{wI2} &= \lambda \frac{\sum_{i=2}^m D_{i:m:n}^* E(D_{i:m:n}^*) / \text{Var}(D_{i:m:n}^*)}{\sum_{i=2}^m D_{i:m:n}^{*2} / \text{Var}(D_{i:m:n}^*)} \\ &= \lambda \frac{\sum_{i=2}^m Q_{i:m:n} + \sum_{i=2}^m U_{i:m:n}}{\sum_{i=2}^m T_{i:m:n} + \sum_{i=2}^m U_{i:m:n}} \\ &= \lambda \frac{\sum_{i=2}^m Q_{i:m:n} / m^2 + \sum_{i=2}^m U_{i:m:n} / m^2}{\sum_{i=2}^m T_{i:m:n} / m^2 + \sum_{i=2}^m U_{i:m:n} / m^2}, \end{aligned} \tag{3.4}$$

by Lemma 1, the fraction term in (3.4) converge to 1 in probability as $m \rightarrow \infty$, and this completes the proof. □

4. Application

In this section, the proposed estimators are assessed by Monte Carlo simulations; in addition, two real data sets are presented.

4.1. Simulation study

The estimators discussed in Sections 2 and 3 are compared in terms of MSE and bias. Unlike estimators $\hat{\lambda}$ and $\hat{\lambda}_U$, the exact mean and variance of the estimator $\hat{\lambda}_{wI2}$ cannot be expressed in a closed form. Therefore, its MSE and bias are obtained over 10,000 replications. Progressive Type-II censored samples were generated from the Pareto distribution with $\lambda = 0.5, 1.5$ and $\theta = 1$ using the algorithm in Balakrishnan and Sandhu (1995). The MSEs and biases of other estimators were obtained using their mean and variance (Table 1). For notational simplicity, the scheme $(0, 0, \dots, n - m)$ is denoted by $((m - 1) * 0, n - m)$. For instance, $(3 * 0, 2)$ denotes the progressive censoring scheme $(0, 0, 0, 2)$.

Table 1: MSEs (biases) for estimators of θ and λ

θ	λ	n	m	Censoring scheme	$\hat{\lambda}$	$\hat{\lambda}_U$	$\hat{\lambda}_{w/2}$	$\hat{\theta}$	$\hat{\theta}_U$	
0.5	20	20	20	(20*0)	0.021 (0.056)	0.015 (0.000)	0.017 (0.002)	0.028 (0.111)	0.013 (0.000)	
				(2*0, 1, 0, 2, 0, 2, 2*0, 5)	0.071 (0.125)	0.036 (0.000)	0.045 (0.027)	0.028 (0.111)	0.014 (0.000)	
		30	20	15	(30*0)	0.012 (0.036)	0.009 (0.000)	0.011 (0.000)	0.011 (0.071)	0.005 (0.000)
					(9*0, 10, 10*0)	0.021 (0.056)	0.015 (0.000)	0.016 (-0.005)	0.011 (0.071)	0.005 (0.000)
		40	30	20	(5, 6*0, 10, 7*0)	0.034 (0.077)	0.021 (0.000)	0.023 (-0.006)	0.011 (0.071)	0.005 (0.000)
					(40*0)	0.008 (0.026)	0.007 (0.000)	0.008 (0.000)	0.006 (0.053)	0.003 (0.000)
	1	40	30	20	(10*0, 5, 7*0, 3, 10*0, 2)	0.012 (0.036)	0.009 (0.000)	0.011 (0.001)	0.006 (0.053)	0.003 (0.000)
					(8*0, 2*10, 10*0)	0.021 (0.056)	0.015 (0.000)	0.016 (-0.009)	0.006 (0.053)	0.003 (0.000)
			20	20	10	(20*0)	0.191 (0.167)	0.132 (0.000)	0.152 (0.007)	0.002 (0.034)
		(2*0, 1, 0, 2, 0, 2, 2*0, 5)				0.643 (0.375)	0.321 (0.000)	0.405 (0.081)	0.002 (0.034)	0.001 (0.000)
		30		20	15	(30*0)	0.107 (0.107)	0.083 (0.000)	0.095 (0.001)	0.001 (0.023)
			(9*0, 10, 10*0)			0.191 (0.167)	0.132 (0.000)	0.148 (-0.014)	0.001 (0.023)	0.001 (0.000)
40	30		20	(5, 6*0, 10, 7*0)	0.303 (0.231)	0.188 (0.000)	0.204 (-0.017)	0.001 (0.023)	0.001 (0.000)	
		(40*0)		0.074 (0.079)	0.061 (0.000)	0.072 (-0.001)	0.001 (0.017)	0.000 (0.000)		
	20	30	15	(10*0, 5, 7*0, 3, 10*0, 2)	0.107 (0.107)	0.083 (0.000)	0.095 (0.002)	0.001 (0.017)	0.000 (0.000)	
(8*0, 2*10, 10*0)				0.191 (0.167)	0.132 (0.000)	0.147 (-0.028)	0.001 (0.017)	0.000 (0.000)		

MSE = mean squared error.

Table 2: Estimates for real data

	$\hat{\lambda}$	$\hat{\lambda}_U$	$\hat{\lambda}_{w/2}$	$\hat{\theta}$	$\hat{\theta}_U$
Device	0.17350	0.13012	0.09001	0.00980	0.00657
Business	2.16083	1.72867	1.79410	1.01000	0.97538

Table 1 shows that the proposed estimator $\hat{\lambda}_{w/2}$ is more efficient than the MLE $\hat{\lambda}$ in terms of MSE and bias, while it has somewhat higher values of the MSE and bias, compared with $\hat{\lambda}_U$. As expected, the estimator $\hat{\theta}_U$ is more efficient than the MLE $\hat{\theta}$ in terms of the MSE and bias. In addition, the MSE of all estimators decrease with an increase in the sample size n as well as with a decrease in the size $n - m$ of the unobserved (censored) data set for a fixed sample size n . It should be noted that the consistent estimator $\hat{\lambda}_{w/2}$ does not depend on the nuisance parameter θ as well as more efficient than the MLE $\hat{\lambda}$ in terms of the MSE and bias.

4.2. Real data

Two real data sets (device lifetime and business failure) are considered in Fernandez (2014). Wu *et al.* (2007) generated a progressive Type-II censored sample

$$0.0098, 0.0376, 0.0661, 0.0849, 0.1112, 0.1447, 0.1904, 0.2463$$

with the censoring scheme (1, 0, 2, 0, 3, 2, 0, 4) from the device lifetime data. The business failure data in Nigm and Hamdy (1987) represents the time (in years) for which a business operates until failure. A sample of fifteen businesses was used. Fernandez (2014) used a progressive Type-II censored sample

$$1.01, 1.05, 1.08, 1.14, 1.28, 1.30, 1.33, 1.43, 1.59, 1.62$$

with the censoring scheme (9*0, 5) from the business failure data. Here, progressive Type-II censored samples were used to obtain the estimates discussed in Sections 2 and 3 (Table 2).

5. Conclusions

A new approach was proposed to estimate the unknown parameters of the Pareto distribution under the progressive Type-II censoring scheme in the regression framework; subsequently, it was proved

that the proposed estimator, $\hat{\lambda}_{w/2}$ is consistent. In addition, the estimator $\hat{\lambda}_{w/2}$ does not depend on the nuisance parameter θ as well as satisfactory in terms of MSE and bias. The proposed approach can be applied to other censoring schemes and life-time distributions despite showing a somewhat lower performance than the estimator $\hat{\lambda}_U$ in simulation study; however, it can be a very good alternative if an unbiased estimator cannot be obtained from the maximum likelihood or pivotal-based methods.

References

- Balakrishnan N and Aggarwala R (2000). *Progressive Censoring: Theory, Methods, and Applications*, Birkhäuser, Boston.
- Balakrishnan N and Cramer E (2014). *The Art of Progressive Censoring*, Springer, New York.
- Balakrishnan N, Kannan N, Lin CT, and Ng HKT (2003). Point and interval estimation for Gaussian distribution, based on progressively Type-II censored samples, *IEEE Transactions on Reliability*, **52**, 90–95.
- Balakrishnan N and Sandhu RA (1995). A simple simulational algorithm for generating progressive Type-II censored samples, *American Statistical Association*, **49**, 229–230.
- Cohen AC (1963). Progressively censored sample in life testing, *Technometrics*, **5**, 327–339.
- Fernandez AJ (2014). Computing optimal confidence sets for Pareto models under progressive censoring, *Journal of Computational and Applied Mathematics*, **258**, 168–180.
- Karr AF (1993). *Probability*, Springer-Verlag, New York.
- Kim JHT, Ahn S, and Ahn S (2017). Parameter estimation of the Pareto distribution using a pivotal quantity, *Journal of the Korean Statistical Society*, **46**, 438–450.
- Lu HL and Tao SH (2007). The estimation of Pareto distribution by a weighted least square method, *Quality & Quantity*, **41**, 913–926.
- Nigm AM and Hamdy HI (1987). Bayesian prediction bounds for the Pareto lifetime model, *Communications in Statistics - Theory and Methods*, **16**, 1761–1772.
- Seo JI and Kang SB (2015). Pivotal inference for the scaled half logistic distribution based on progressively Type-II censored samples, *Statistics & Probability Letters*, **104**, 109–116.
- Viveros R and Balakrishnan N (1994). Interval estimation of parameters of life from progressively censored data, *Technometrics*, **36**, 84–91.
- Wang B (2008). Goodness-of-fit test for the exponential distribution based on progressively Type-II censored sample, *Journal of Statistical Computation and Simulation*, **78**, 125–132.
- Wu JW, Wu SF, and Yu CM (2007). One-sample Bayesian predictive interval of future ordered observations for the Pareto distribution, *Quality & Quantity*, **41**, 251–263.