http://ijopaar.com; 2016 Vol. 3(2); pp. 160-163.

Bayesian Estimation of Pareto Distributionwith Record Values using Lindley's Approximation

Rajeev Pandey
Department of Statistics, University of Lucknow
Lucknow 226007, India
Email: prof.rajeevlu@gmail.com

Date of revised paper submission: 12th August 2016; Date of acceptance: 12st September 2016
Date of publication: 30th September 2016; Impact Factor: 3.498; Quality Factor: 4.39
***First Author / Corresponding Author; Paper ID: A16309**

ABSTRACT

In this study, we looked at parameter estimate based on Pareto distribution's upper record values. For the unknown parameters of this distribution, maximum likelihood and approximate Bayes estimators based on upper record values are derived. On the basis of higher record values, Lindley's approximation is utilized to produce approximate Bayes estimators under the square error loss function.

Keywords: Pareto Distribution, Record Values, Bayesian Estimation, Square error loss function and Lindley's Approximation.

1. Introduction

The theory of record values and its applications are being utilized extensively in data analysis, particularly in the study of the stock market to make predictions about the price of a stock that may be greater or lower than the lastone. Ahsanullah (1995) [1] and Arnold and Balakrishnan (1998) [2] have provided extensive use of record values for various real life situations. A.M. Nigm and H. I. Hamdy 2007 [7] have mentioned that the Pareto Distribution is within the category of distributions with decreasing failure rates. It has been observed by Dupuis, Tsao (2007) [4] and Castillo, Hadi (1997) [5] etc. and many more have been observed that Pareto Distribution is widely applicable in the fields of engineering, biology, medicine and others, and more over this distribution is quiet helpful for the purpose of modeling and analysis life time data. In order to compute the estimates of the unknown parameters of the distribution under the upper record values, thestudy provides Bayes estimates via Lindley's Approximation approach. Let X_1, X_2, X_3, ..., X_n be a series of independent random variables with cdf $F(x)$ and pdf f(x) distributions.

If $Y_j \geq Y_{j-1}$; $j \geq 1$, then X_j is referred to as an upper record and is denoted by $X_{U(j)}$ in the set $Y_n = \max(X_1, X_2, X_3, ..., X_n)$, where $n \geq 1$. Let $X_{U(1)}, X_{U(2)}, X_{U(3)}, ..., X_{U(n)}$ be the first upper record values of size n resulting from a series of independent and identically pareto variable with the probability density function

$$f(x) = \alpha\beta^{\alpha}(x+\beta)^{-(\alpha+1)}; \qquad x \geq 0, \alpha, \beta > 0 \qquad (1.1)$$

and cumulative distribution function

$$F(x) = 1 - \beta^{\alpha}(x+\beta)^{-\alpha}; \qquad x \geq 0, \alpha, \beta > 0 \qquad (1.2)$$

Where β is scale and α is shape parameter.

2. Estimation of Parameters under Maximum Likelihood Estimation (MLE)

Suppose that $\underline{x} = x_{u(1)}, x_{u(2)}, \ldots, x_{u(n)}$ be the first upper record values of size n from Pareto distribution. The likelihood function for observed record $\underline{x}$ given by,

$$l(\alpha, \beta|\underline{x}) = f(x_{u(n)}) \prod_{i=0}^{n-1} \frac{f(x_{u(i)})}{1-F(x_{u(i)})} \tag{2.1}$$

Where f(.) and F(.) are given, respectively by (1.1) and (1.2). Substituting f(.) and F(.) in (1.2), we get

$$l(\alpha, \beta|\underline{x}) = \alpha^n \beta^\alpha (x_{u(n)} + \beta)^{-\alpha} \prod_{i=1}^{n} (x_{u(i)} + \beta)^{-1} \tag{2.2}$$

Log likelihood function may be then written as

$$L(\alpha, \beta|\underline{x}) = log l(\alpha, \beta|\underline{x})$$

i.e, $$L(\alpha, \beta|\underline{x}) = nlog\alpha + \alpha log\beta - \alpha(x_{u(n)} + \beta) - \sum_{i=1}^{n} log(x_{u(i)} + \beta) \tag{2.3}$$

Taking derivatives with respect to α and β of (2.3) and equating them to zero, we obtain the likelihood equations for α and β to be

$$\frac{\partial L(\alpha, \beta|\underline{x})}{\partial \alpha} = \frac{n}{\alpha} + log\beta - log(x_{u(n)} + \beta) \tag{2.4}$$

$$\frac{\partial L(\alpha, \beta|\underline{x})}{\partial \beta} = \frac{\alpha}{\beta} + \frac{\alpha}{(x_{u(n)} + \beta)} - \sum_{i=1}^{n} \frac{1}{(x_{u(i)} + \beta)} \tag{2.5}$$

The equations (2.4) and (2.5) cannot solve analytically for α and β. Therefore, we use R software to solve these equations and find the MLE's of the unknown parameters α and β.

3. Estimation of parameters under Lindley's Approximation

Assume the following gamma prior densities for and for model (1.1).

$$\pi_1(\alpha|p, q) = \frac{q^p}{\Gamma p} \alpha^{p-1} exp(-q\alpha) \qquad (\alpha \geq 0) \tag{3.1}$$

$$\pi_2(\beta|r, s) = \frac{s^r}{\Gamma r} \beta^{r-1} exp(-s\beta) \qquad (\beta \geq 0) \qquad .(3.2)$$

The joint prior density of α and β may be written as

$$\pi(\alpha, \beta) = \pi_1(\alpha|p, q)\pi_2(\beta|r, s)$$

$$\pi(\alpha, \beta) = \frac{q^p s^r}{\Gamma p \Gamma r} \alpha^{p-1} \beta^{r-1} exp(-q\alpha - s\beta) \tag{3.3}$$

Based on the likelihood function of the observed sample is same as (2.4) and the joint prior in (3.3), the joint posterior density of α and β given the data

$$\pi^*(\alpha, \beta|\underline{x}) = \frac{l(\alpha, \beta|\underline{x})\pi(\alpha, \beta)}{\int_0^\infty \int_0^\infty l(\alpha, \beta|\underline{x})\pi(\alpha, \beta) d\alpha d\beta} \tag{3.4}$$

Therefore, the Bayes estimate of any function of α and β say (α, β), under square error loss function is

$$\tilde{g}(\alpha, \beta) = E_{\alpha,\beta|data}[g(\alpha, \beta)] = \frac{\int_0^\infty \int_0^\infty g(\alpha, \beta) l(\alpha, \beta|\underline{x})\pi(\alpha, \beta) d\alpha d\beta}{\int_0^\infty \int_0^\infty l(\alpha, \beta|\underline{x})\pi(\alpha, \beta) d\alpha d\beta} \tag{3.5}$$

Now, we utilize the Lindley's Approximation technique from the posterior distributions and then compute the Bayes Estimator of g (α, β) under the squared error loss (SEL) function. The ratio of two integrals provided by (3.5) cannot be achieved in a closed form; therefore, we use Lindley's approximation. By multiplying the likelihood by the joint prior, the equation for the joint posterior up to proportionality may be desired as

$$\pi^*(\alpha,\beta|\underline{x}) \propto \alpha^{n+p-1}\beta^{n+r-1}\exp\left[-\left(q\alpha+s\beta-\beta\log\left(1-\exp\left(-x_{u(n)}^{\alpha}\right)\right)\right)\right]$$

$$+\prod_{i=1}^{n-1}\frac{x_{u(i)}^{\alpha-1}\exp(-x_{u(i)}^{\alpha})}{1-\exp(-x_{u(i)}^{\alpha})} \tag{3.6}$$

The posterior distribution of the supplied Eq. (3.6) cannot be analytically reduced to well-known distributions, making it impossible to sample directly using conventional methods. However, the plot of the posterior distribution indicates that it is comparable to the normal distribution. The selection of the hyper parameters (p, q, r and s) that brings (3.6) close to the proposal distribution. We compute the Bayes estimates

4. Lindley's Approximation

We consider the Lindley's approximation method to obtain the Bayes estimates of the parameters, which includes the posterior expectation is expressible in the form of ratio of integral as follow:

$$I(x)=E(\alpha,\beta|\underline{x})=\frac{\int u(\alpha,\beta)e^{L(\alpha,\beta)+G(\alpha,\beta)}d(\alpha,\beta)}{e^{L(\alpha,\beta)+G(\alpha,\beta)}d(\alpha,\beta)} \tag{4.1}$$

Where,

u(α, β) =is a function of α and β only

L(α, β) = Log-likelihood function

G(α, β) = Log of joint prior density According to D.V. Lindley,

if ML estimates of the parameters are available and n is sufficiently large then the above ratio of the integral can be approximated and the approximate Bayes estimator of α under squared error loss (SEL) is,

$$\hat{\alpha}_S^L = \hat{\alpha}+0.5\left[2\left(\hat{u}_\alpha\,\hat{p}_\beta\hat{\sigma}_{\alpha\beta}+2\hat{u}_\alpha\,\hat{p}_\beta\hat{\sigma}_{\alpha\alpha}\right)\right]+0.5[(\hat{u}_\alpha\hat{\sigma}_{\alpha\beta})(\hat{L}_{\beta\beta\beta}\hat{\sigma}_{\alpha\beta}+\hat{L}_{\beta\alpha\beta}\hat{\sigma}_{\beta\alpha})$$

$$+(\hat{u}_\alpha\hat{\sigma}_{\alpha\alpha})(\hat{L}_{\beta\alpha\alpha}\hat{\sigma}_{\beta\alpha}+\hat{L}_{\alpha\alpha\alpha}\hat{\sigma}_{\alpha\alpha})] \tag{4.2}$$

and similarly, the Bayes estimates for β under SELF is,

$$\hat{\beta}_S^L = \hat{\beta}+0.5\left[2\left(\hat{u}_\beta\,\hat{p}_\beta\hat{\sigma}_{\beta\beta}+2\hat{u}_\beta\,\hat{p}_\alpha\hat{\sigma}_{\beta\alpha}\right)\right]+0.5[(\hat{u}_\beta\hat{\sigma}_{\beta\beta})(\hat{L}_{\beta\beta\beta}\hat{\sigma}_{\beta\beta}+\hat{L}_{\beta\alpha\beta}\hat{\sigma}_{\beta\alpha})$$

$$+(\hat{u}_\beta\hat{\sigma}_{\alpha\beta})(\hat{L}_{\beta\alpha\alpha}\hat{\sigma}_{\beta\alpha}+\hat{L}_{\alpha\alpha\alpha}\hat{\sigma}_{\alpha\alpha})] \tag{4.3}$$

5. Application to Real Data

To provide an example of the inferential techniques developed in the sections above. We selected actual data that V. Choulakian and M. A. Stephens 2012 [9] had also utilized. The data represent the Wheaton River in Carcross, Yukon Territory, Canada, exceedances of flood maxima (in m3/s). The statistics are excesses for the years 1958 through 1984. The data are given below 1.7, 2.2, 14.4, 20.6, 39, 64 Based on these seven upper record values, we compute the approximate MLEs and Bayes estimates of α and β using Lindley's Approximation. Table 1: Estimates of α and β obtained by MLE and Lindley' Approximation

Method	**Parameter**	Estimates
MLE	α	2.3935
	β	5.6822
Lindley's Approximation	α	1.5070
	β	1.6915

5. Conclusion

In this study, we address the Bayes estimate of the Pareto distribution's unknown parameters when the data are higher record values. Under the suppositions of squared error loss functions, we present the Bayes estimators and assume the gamma priors on the unknown parameters. The Bayes estimators can be produced via numerical integration; however, they cannot be obtained in explicit forms. Because of this, we use Lindley's Approximation approach. We observe the estimates based on Lindey's are better than Maximum likelihood estimates.

6. Acknowledgement

We are extremely grateful to Prof. R.K. Singh for their invaluable assistance and support.

7. Funding Source

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

8. Decelerations of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

REFERENCES

[1] M. Ahsanullah, Record Statistics, NOVA Science Publishers Inc., Hunting ton, New York, 1995.

[2] B. C. Arnold, N. Balakrishnan and H. N. Nagaraja, Records, John Wiley and Sons, INC, 1998.

[3] S. Rezaei, R. Tahmasbi and M. Mahmoodi, Estimation of P[Yi ≥ X] for generalized Pareto distribution, J. Statist. Plann. Inference 140 (2010), 480-494.

[4] D.J. Dupuis and M. Tsao, A hybrid estimator for generalized pareto and extreme-value distributions, Communications in Statistics - Theory and Methods (2007).

[5] E. Castillo and A. S. Hadi, Fitting the Generalized Pareto Distribution to Data, Journal of the American Statistical Association (1997).

[6] A.M. Nigm and H. I. Hamdy, Bayesian prediction bounds for the pareto lifetime model,Communications in Statistics - Theory and Methods (2007).

[7] B. C. ARNOLD and S. JAMES PRES, Bayesian Estimation and Prediction for Pareto Data, Journal of the American Statistical Association (1989).

[8] V. Choulakian and M. A. Stephens, Goodness-of-Fit Tests for the Generalized Pareto Distribution, Technometrics (2012).