Volume : 5 | Issue : 11 | November 2016 • ISSN No 2277 - 8179 | IF : 3.508 | IC Value : 69.48

A Sequential Approach for the Point Estimation of a Pareto Shape Parameter

Statistics

KEYWORDS: Sequential procedure, Minimum risk point estimation, Second-order approximation, Stopping rule.

Sandeep Charak Department of Statistics University of Jammu Jammu-180006

Rahul Gupta Department of Statistics University of Jammu Jammu-180006

ABSTRACT *The Pareto distribution and its generalization is an important probability distribution and plays an important role in many areas of Economics and Econometrics. Study of its scale and the shape parameters has always been a challenge for researchers. In this paper we consider the problem of minimum risk point estimation of shape parameter of pareto distribution. We prove the failure of the fixed sample size procedures to handle the estimation problem. Purely sequential procedure is developed to tackle the situation and the second-order approximations are obtained for the proposed sequential procedure.*

1. Introduction : Wang (1973) proposed Sequential estimation of the scale parameter of the pareto distribution. Ghosh and Wackerley (1986) proposed fixed estimation of the location parameter of a Pareto distribution. Further work on Pareto distribution is given by Mukhopadhyay and Ekwo (1987a, b) and Mukhopadhyay (1987) who considered sequential point estimation problems for the scale parameter of a Pareto distribution. Under a very general loss structure, they derived several asymptotic results regarding the associated "risk" and "regret" functions. They also considered the problem of constructing a fixed-ratio confidence interval for the scale parameter and proposed various sampling techniques to achieve the intended goal. Castillo and Daoudi (2009) and De Zea Bermudez and Kotz, S. (2010), focused on new methods for Parametric estimation of the generalized Pareto distribution .Nadarajah and Ali(2008) applied Pareto random variables for hydrological modeling. In the present note we consider the problem of point estimation of the shape parameter of a Pareto distribution. The failure of fixed sample size procedure to deal with the estimation problem is established in the next section and a purely sequential procedure is developed in section 3, besides the second-order approximations are obtained for the proposed sequential procedure.

2. The Setting of the Estimation Problem and Failure of the Fixed Sample size Procedure

Let us consider a sequence $\{X_i\}, i = 1,2, \ldots$ of independent random variables from a first kind of Pareto Distribution $f(X;\mu,\sigma) = \sigma^{-1}\mu^{-1/\sigma}X^{-1/\sigma-1}$; $X \geq \mu > 0, \sigma > 0$, where μ and σ are respectively, the unknown scale and shape parameters. Given an observed random sample $X_1, \ldots, X_n$ of size n(≥2), for $X_{n(1)} = \min(X_1, \ldots, X_n)$, Let $u_{n(1)} = \log X_{n(1)}$ and $\hat{\sigma}_n = (n-1)\sum_{i=1}^{n} \log\{X_i/X_{n(1)}\}$.

Let $(\log\mu, \sigma) \in \Re \times \Re^+$, where $\Re$ and $\Re^+$ denote, respectively, the 1-dimensional Euclidean space and the positive-half of the real line. Let $u_{n(1)} = \log\hat{\mu}(X_1, \ldots, X_n)$ and $\hat{\sigma}_n = \hat{\sigma}(X_1, \ldots, X_n)$ be the estimators of $\log\mu$ and σ respectively, satisfying the following assumptions:

(A_1) : $n\sigma^{-1}[(u_{n(1)} - \log\mu)] \sim \chi^2_{(2)}$,

where $\chi^2_{(r)}$ denotes a chi-square random variable with 2 degrees of freedom.

(A_2) : For all $n \geq 2$, $u_{n(1)}$ and $\hat{\sigma}_n$ are stochastically independent.

(A_3) : $q(n-1)\hat{\sigma}_n/\sigma = \sum_{j=1}^{n-1} Z_j^{(2)}$, where $Z_j^{(2)} \sim \chi^2_{(2)}$.

Our problem is the point estimation of σ by $\hat{\sigma}_n$. Let the loss incurred in estimating σ by $\hat{\sigma}_n$ be squared error plus linear cost of sampling, that is,

$$L(\sigma,\hat{\sigma}_n) = A(\hat{\sigma}_n - \sigma)^2 + cn, \quad (2.1)$$

where A (>0) is the known weight and c (>0) is the known cost per unit sample observation. Using (A_3), the risk corresponding to the loss function (2.1) is

$$R_n(c) = \frac{2A\sigma^2}{2(n-1)} + cn. \quad (2.2)$$

The value of $n = n_0$ minimizing the risk (2.2) comes out to be

$$n_0 = 1 + (2A/2c)^{1/2}\sigma, \quad (2.3)$$

and substituting $n = n_0$ in (2.2), the associated minimum risk is

$$R_{n_0}(c) = c(2n_o - 1) \quad (2.4)$$

Now, since n_0 depends on σ, in the absence of any knowledge about σ, no fixed sample size procedure minimizes the risk for all values of σ.

From (2.3), we conclude that due to dependence of 'optimal' fixed sample size solutions n_0 on σ, the fixed sample size procedure fail to give solution to the estimation problem. To tackle this problem of estimation, we propose a procedure based on variable sample size in the next section.

3. Sequential Procedure for the Point Estimation of shape parameter of Pareto Distribution

Take $m \geq 2$ as the initial sample size. Then the stopping time $N \equiv N(C)$ corresponding to sequential procedure is defined by

$$N = inf\left[n \geq m : n \geq 1 + (2A/2c)^{1/2}\hat{\sigma}_n\right]. \quad (3.1)$$

After stopping, we estimate σ by $\hat{\sigma}_N$, incurring the risk

$$R_N(c) = AE[(\hat{\sigma}_N - \sigma)^2] + cE(N). \quad (3.2)$$

Following Starr and Woodroofe (1969), we define the 'regret' of sequential procedures as

$$R_g(c) = R_N(c) - R_{n_0}(c). \quad (3.3)$$

Now we state and prove the main theorem of this section providing second-order approximations for expected sample size and 'regret' associated with the sequential procedures.

Theorem 3.1: For the sequential procedures defined at (3.1) and all $m = 3$, as $c \to 0$

$$E(N) = n_0 + v + 1/3 + o(1), \quad (3.4)$$

and

$$R_g(c) = c + o(1), \quad (3.5)$$

where υ is specified

Proof : Result (3.4) can be easily proved along the lines of Theorem 3 of Chaturvedi, Tiwari, and Pandey(1992).

Denoting by

$$S_N = \sum_{j=1}^{N}\left\{{}^{1}/_{2}\, Z_j^{(2)}\right\},\ n_{00} = n_0 - 1\ and\ T = N - 1$$

we can write

$$A(\hat{\sigma}_N - \sigma)^2 = (A\sigma^2/T^2)(S_T - T)^2$$
$$= (2c/2)(S_T - T)^2 + (2c/2)(n_{00}^2/T^2 - 1)(S_T - T)^2. \quad (3.6)$$

From (3.6) and Wald's lemma for cumulative sums,

$$E[L(\sigma, \hat{\sigma}_N)] = 2cE(T) + c + E(\upsilon_T),\ \text{say},$$

where $\upsilon_T = (2c/2)(S_T - T)^2(n_{00}^2/T^2 - 1)$. (3.7)

Let us write

$$\upsilon_T = (2c/2)(n_{00}^2/T^2)(1 - T^2/n_{00}^2)^2(S_T - T)^2$$
$$+(cq/2)(1 - T^2/n_{00}^2)(S_T - T)^2$$
$$= \text{I} + \text{II},\ \text{say} \quad (3.8)$$

The stopping rule (3.1) can be rewritten as

$$N = \inf\,[n \geq m :\ S_n \leq (n-1)^2/(n_{00})].$$

Denoting by $R_c = (T^2/n_{00}) - S_T$, we can write

$$1 - T^2/n_{00}^2 = -(T^{-1}/n_{00}^{-1})(S_T - T + R_c). \quad (3.9)$$

Utilizing (3.9) and the result $T/n_{00} \xrightarrow{a.s.} 1$ as $c \longrightarrow 0$, we get

$$\text{I} = (2c/2)(n_{00}^2/T^2)\left\{-(2/n_{00})(S_T - T) + o_p(c)\right\}^2(S_T - T)^2$$
$$= (4c)[(S_T - T)^2/n_{00}^2] + o_p(c). \quad (3.10)$$

Since $S_T \leq T^2/n_{00}$, we have

$$(S_T - T)^4/(n_{00}^2) \leq (T/n_{00})^4[(T - n_{00})/(n_{00})^{1/2}]^4. \quad (3.11)$$

It follows from lemma 2.1 and Theorem 2.3, respectively, of Woodroofe (1977) that $(T/n_{00})^4$ and $[(T - n_{00})/(n_{00})^{1/2}]^4$ are uniformly integrable for all m >3. Moreover, Theorem 1 of Anscombe (1952) leads us to the result that $(S_T - T)/(n_{00})^{1/2} \xrightarrow{L} N(0, 1)$ as $c \to 0$. Using these results and (3.11), we obtain from (3.10) that, for all m >3, as $c \to 0$,

$$E(\text{I}) = 12c. \quad (3.12)$$

Furthermore,

$$\text{II} = -(2c/n_{00})(S_T - T)^3 - (2c/n_{00})R_c(S_T - T)^2$$
$$+(2c/2)(1 - T/n_{00})^2(S_T - T)^2$$
$$= -\text{II}_1 - \text{II}_2 + \text{II}_3,\ \text{say} \quad (3.13)$$

Proceeding as for I, it can be shown that II_2 is uniformly integrable for all $m > 2$. From Theorem 2.1 of Woodroofe (1977), since R_c and $(S_T\text{-}T)^2$ are asymptotically independent, the mean of the symptotic distribution of R_c is υ and $(S_T - T)/(n_{00})^{1/2} \xrightarrow{L} N(0, 1)$ as $c \to 0$, we get for all m > 2, as $c \to 0$,

$$E(\text{II}_2) = 2c\upsilon. \quad (3.14)$$

Once again, using the asymptotic independence of $(\text{T} - n_{00})^2/(n_{00})$ $(S_T - T)^2/(n_{00})$ and the results that they are uniformly integrable for all m > 2, with each one distributed as $X^2_{(1)}$, we get

$$E(\text{II}_3) = 2c. \quad (3.15)$$

It follows from the Theorem 8 of Chow, Robbins and Teicher (1965) that

$$\text{E}(S_T - T)^3 = 4E(T) + 3E[T(S_T - T)],$$

and hence,

$$E(\text{II}_1) = (2c/n_{00})[4E(T) + 3E\{T(S_T - T)\}] \quad (3.16)$$

We have,

$$(T/n_{00})(S_T - T) = (T - n_{00})$$
$$+(n_{00})^{-1}(T^2 - n_{00}^2)(T/n_{00} - 1) - (T/n_{00})R_c$$
$$= \text{II}_{11} + \text{II}_{12} - \text{II}_{13}, \quad (3.17)$$

From (3.4), for all m > 2, as $c \to 0$

$$E(\text{II}_{11}) = \upsilon - 2 + o(1) \quad (3.18)$$

and as $c \to 0$,

$$E(\text{II}_{13}) = \upsilon. \quad (3.19)$$

Moreover, since

$$\text{II}_{12} = n_{00}{}^{-1}(T + n_{00})\{(T + n_{00})^2/(n_{00})\},$$

using the results that $(T/n_{00}) \xrightarrow{a.s.} 1$ as $c \to 0$,

$(T - n_{00})/(n_{00})^{1/2} \xrightarrow{L} N(0, 1)$ and it is uniformly integrable for all m > 2, one gets,

$$\text{E}(\text{II}_{12}) = 2. \quad (3.20)$$

Making substitutions from (3.18), (3.19) and (3.20) in (3.17) and using (3.4), we obtain from (3.16), for all m > 2, as $c \to 0$,

$$\text{E}(\text{II}_1) = (8c/2n_{00})[n_{00} + \nu - 2 + o(1)]$$
$$+6c[o(1)]. \quad (3.21)$$

From (3.13), (3.14), (3.15) and (3.21), for all m > 2, as $c \to 0$,

$$\text{E}(\text{II}) = -7c - 2c\nu + o(1). \quad (3.22)$$

We conclude from (3.8), (3.12) and (3.22) that, for all m > 3, as $c \to 0$,

$$\text{E}(\nu_T) = 5c - 2c\nu + o(1) \quad (3.23)$$

Result (3.5) now follows on making substitutions from (3.4) and (3.23) in (3.7).

References

1. Anscombe, F.J (1952): Large sample theory of sequential estimation. Proc.Cambridge Philos.Soc., 48, 600-607.
2. Castillo, J. and Daoudi, J. (2009), Estimation of the generalized Pareto distribution. Statistics and Probability Letters, 79, 684 – 688
3. Chaturvedi, A., Tiwari, N. and Pandey, S.K. (1992) : Second-Order approximations to a class of sequential point estimation procedures. Metron.Vol. 50-n-12, 161-177.
4. Chow, Y.S., Robbins, H. and Teicher, H. (1965) : Moments of randomly stopped sums. Ann. Math. Statist., 36, 789-799.
5. De Zea Bermudez , P. and Kotz, S. (2010), "Parameter estimation of the generalized Pareto distribution—Part I," Journal of Statistical Planning and Inference, Volume 140, Issue 6 (2010), Pages 1353-1373.
6. Ghosh, M. and Wackerley, D. (1986): Fixed estimation of the location parameter of a Pareto distribution. Cal. Statist. Assoc. Bull., 35, 67-75
7. Mukhopadhyay, N. and Ekwo, M. E. (1987a): A note on minimum risk point estimation of the shape parameter of a Pareto distribution. Cal. Statist. Assoc. Bull., 36, 67-78. 10.
8. Mukhopadhyay, N. and Ekwo, M.E.(1987b): Sequential Estimation problems for the Scale Parameter of a Pareto Distribution. Scand. Actuarial Journal , Vol.1987,1-2, 83-103.
9. Mukhopadhyay, N. (1987): Sequential estimation problems for the scale parameter of a Pareto distribution. Scand, Actuar. Jour., 1987, 83- 103.
10. Nadarajah S, Ali MM (2008) Pareto random variables for hydrological modeling. Water Resour Manag 22: 1381–1393.
11. Starr, N. and Woodroofe, M. (1969): Remarks on sequential point estimation. Proc. Nat. Acad. Sci., 63,285-288.
12. Swanepoel, J.W.H. and Van Wyk, J.W.J. (1982) : Fixed width confidence intervals for the location parameter of an exponential distribution.
13. Wang, Y.H.(1973): Sequential Estimation of the Scale Parameter of the Pareto Distribution.Commun.Statist. Theor.Meth.,A2,145-154.
14. Woodroofe, M. (1977): Second-order approximations for sequential point and interval estimation. Ann.Statist.,5, 984-995.