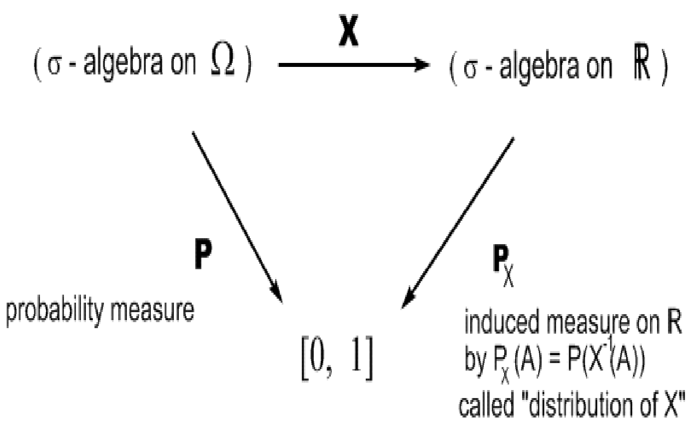
International Recognized Double-Blind Peer Reviewed Multidisciplinary Research Journal
Golden Research Thoughts
ISSN 2231-5063 Impact Factor : 3.4052 (UIF)
Volume - 4 | Issue - 12 | June - 2015 Available online at www.isrj.org

SEQUENTIAL FIXED WIDTH INTERVAL ESTIMATION OF THE MINIMUM OF A RANDOM VARIABLE

H. S. Patil
Department of Statistics, S. B. Zadbuke Mahavidyalaya, Barsi, Maharashtra, India.

Short Profile :
H. S. Patil Department of Statistics, S. B. Zadbuke Mahavidyalaya, Barsi, Maharashtra, India.



ABSTACT:
In this paper we have discussed the standard and modified Meczarski (1985) procedure proposed by Schaalje et. al. (2001) to find the minimum of Weibull density based on simulation study. We propose a sequential procedure based on m.l.e and compare it with these methods. Further we independently compare sequential procedures based on m.l.e and moment estimator with the standard and modified Meczarski procedures for Pareto density and found that the sequential procedure based on m.l.e performs better.

KEYWORDS: Extreme value; Confidence level; Average sample size.

Article Indexed in :
DOAJ Google Scholar DRJI
BASE EBSCO Open J-Gate

1. INTRODUCTION:

Let X_1 , X_2 , …., X_n be independent identically distributed (i.i.d) random variables with distribution function (d.f) F(x). The minimum of a random variable λ is the least upper bound (l.u.b) of {y: F(y) = 0, y → - ∞}. For example, if X follows B(n, p) then λ = l.u.b{0≤ y ≤∞: F(y) = 0} = 0. The problem is to find an interval estimate of λ with desired confidence level 1 – α and desired width d.

Let $X_{n(1)}$ be the first order statistics of current n observations and be used as an estimate of λ. For $P(X_{n(1)} - \lambda \le d) \ge 1 - \alpha$ to hold good we must have,

$$1 - [1 - F(d+ \lambda)]^2 \ge 1 - \alpha$$

which implies $$n \ge \log \alpha / \log[1 - F(d+ \lambda)] \qquad (1.1)$$

Since both F(.) and λ are unknown, there is no fixed sample procedure exists. Hence Meczarski (1985) proposed a sequential procedure to find fixed width interval estimate of λ as: Stop for the first time $n \ge \log \alpha / \log[1 - F_n (d+ X_{n(1)})]$. Thus the stopping rule is given by,

$$N = N(d) = \inf\{n \ge m: n \ge \log \alpha / \log[1 - F_n (d+ X_{n(1)})] \qquad (1.2)$$

where F_n is the empirical d.f based on n samples and m (≥ 2) be an positive integer, the initial sample size. Now from (1.2) we have,

$$P(N \ge n+1) \le P[n \le \log \alpha / \log[1 - F_n (d+ X_{n(1)})]$$

As $n \to \infty$, $F_n(d + X_{n(1)}) \to F(d + \lambda)$ and $\alpha^{1/n} \to 1$. Hence as $n \to \infty$, $P(N > n) = 0$. Thus N is a proper r.v.

Meczarski's Procedure: First take a fixed number of observations, say m (≥ 2) and compute F_m and $N(d) = \log \alpha / \log[1 - F_n (d+ X_{n(1)})]$. If m > N(d), stop sampling otherwise take one more observation and examine the validity of the condition in the rule. Continue this procedure unti,l the condition is satisfied for the first time. Once the sampling is stopped, take $(X_{n(1)} - d, X_{n(1)})$ as the required interval estimate of λ with confidence level 1 – α. Meczarski (1985) has shown that $(X_{n(1)} - d, X_{n(1)})$ is asymptotically (as d→ 0) a (1-α) confidence interval for λ.

Note that for m=1, the condition $n \ge \log \alpha / \log[1 - F(d+ \lambda)]$ is satisfied, since the right hand side of inequality becomes 0. However with m =1, $P[\lambda \in (X_1 - d, X_1] = P(X_1 - d < \lambda) = F(\lambda +d)$ and it reduces to FSS procedure of size 1. Hence the condition F(λ +d) ≥ 1- α cannot be confirmed. So we take m ≥ 2 as initial sample by size.

Modified Meczarski's Procedure: The modified form of Meczarski procedure proposed by Schaalje et. al. (2001) is as below.

Take sa mples until the standard Meczarski's stopping rule is satisfied for the first time. Let n = n_0 be the sample size at this stage. Now take an additional number of samples of size, say k_1 and recalculate N(d). Then the modified procedure stops if $n_0 + k_1 - N(d) \geq c_1$ where c_1 is chosen to be sufficiently large so that the Meczarski's stopping rule has really been satisfied rather than stopping due to instability of F_n. If the modified stopping rule is not satisfied, take an additional k_2 samples and recalculate N(d) and stop if $n_0 + k_1 + k_2 - N(d) \geq c_2$ **(** **>** c_1). This p process is repeated p times if no stopping rule occurred otherwise stop sampling. The selection of k_i, p and c_i is to be such that the procedure can detect whether instability of F_n is affecting the procedure without affecting much in the sample size reduction.

2. Numerical Example and Performance Evaluation: Schallje et. al. (2001) have studied the performance of the proposed modified procedure by simulation method in case of Weibull density by taking p =10, and for i = 1, 2, ..., 10, k_i = 10 and c_i = 9, 17, 24, 30, 35, 39, 42, 44, 48, 50.

The Weibull density is $f(x) = \beta x^{\beta-1} \exp(-(x/h)^{\beta}) / h^{\beta}$, where x > 0, h (> 0) is the scale parameter and β (> 0) is the shape parameter. Note that Weibull density has a minimum λ = 0. Let h =1 in the following.

Since the Weibull density have different variation due to change in β, for comparison purpose d is specified in terms of the proportion of inter-deciles range $D_9 - D_1$, where D_9 and D_1 are 9^{th} and 1^{st} deciles respectively. Since $f(x) = \beta x^{\beta-1} \exp(-x^{\beta})$ implies $P[X^{\beta} > t] = e^{-t}$ implies $P[X > t^{1/\beta}] = e^{-t}$. Thus $P[X \leq t^{1/\beta}] = 1 - e^{-t}$. So we have, $D_9 = P[X \leq t^{1/\beta}] = 0.9$ implies $1 - e^{-t} = 0.9$ implies t = log(10) and hence $D_9 = [\log(10)]^{1/\beta}$. Similarly $D_1 = [\log(10/9)]^{1/\beta}$.

Since $\hat{F}(x) = 1 - \exp(-x^{\hat{\beta}})$, the stopping rule for sequential procedure based on m.l.e using (1.1) is given by, Stop for the first n such that;

$$n \geq \log\alpha / \log[1 - 1 + (\exp(-(d + x_{n(1)}^{\hat{\beta}}))]$$

$$n \geq \log\alpha / [- (d + x_{n(1)}^{\hat{\beta}})]$$

$$n \geq \log(1/\alpha) / [(d + x_{n(1)}^{\hat{\beta}})] \tag{2.1}$$

where $\hat{\beta}$ is an m.l.e of β, obtained from the equation

$$n/\beta + \sum_{i=1}^{n} \log(x_i) - \sum_{i=1}^{n} (x_i^{\beta} \log x_i) = 0$$

The simulation study consists of generating a random sample form Weibull density and checked to see whether the true minimum 0 is within ($X_{n(1)} – d$, $X_{n(1)}$). The proportion of times 0 included in the interval ($X_{n(1)} – d$, $X_{n(1)}$) out of 5000 runs estimates the actual level of confidence and also the average sample size required to terminate is calculated for different combinations of α, β, d and m = 2 and for all procedures. The results are tabulated in Table 2.1.

Table 2.1: The confidence levels and average sample size of standard and modified Meczarski and m.l.e procedures.

				Mec. Proc.		Mod. Mec. Proc.		M L E Procedure		
B	d	α	F-Size	S_{α}	M-Size	MS_{α}	MM-size	Est. B	$MLES_{\alpha}$	MLE-Size
0.5	0.1	0.05	5	0.0682	4	0	18	0.7606	0.1868	4
0.5	0.1	0.1	4	0.0838	3	0	16	0.8147	0.2144	3
0.5	0.2	0.05	3	0.0582	3	0	15	0.8716	0.1362	2
0.5	0.2	0.1	3	0.0476	2	0	13	0.9201	0.1222	2
1	0.1	0.05	14	0.3016	7	0.00024	36	1.14156	0.3290	8
1	0.2	0.05	7	0.2422	4	0.0004	24	1.7015	0.3278	4
1	0.2	0.1	6	0.2484	3	0.001	21	1.8397	0.3516	3

Remark 2.1: Though the average sample sizes required are nearly same for both m.l.e procedure and standard Meczarski (1985) procedure but the later one attains better confidence level. Note that both procedures do not achieve the desired confidence level 1 – α.

Remark 2.2: The modified Meczarski procedure suggested by Schaalje et. al. (2001) is more acceptable than both m.l.e procedure and standard Meczarski (1985) procedure because it approximately achieves or exceeds the desired confidence level 1 – α though the average sample size is more.

3. Estimation of the Minimum of a Pareto Density.

The Pareto density $f(x) = \theta/ x^{\theta+1}$, where x >1 and θ > 0, a shape parameter. Note that this density has minimum λ = 1.

Since $\hat{F}(x) = 1 - 1/x^{\hat{\theta}}$, the stopping rule for sequential procedure based on the parameter estimation using (1.1) is given by,

$$n \geq \log\alpha / \log[(X_{n(1)} + d)^{-\hat{\theta}}]$$

$$n \geq \log(1/\alpha) / [\hat{\theta} \log(X_{n(1)} + d)] \qquad (3.1)$$

By the method of moments we have, $\hat{\theta} = \bar{X} / (\bar{X} - 1)$ and by the method of m.l.e we have, $\hat{\theta} = n / (\sum \log(x_i))$. Also $D_9 = 10^{1/\theta}$ and $D_1 = (10/9)^{1/\theta}$.

Here the simulation study is carried out as explained in case of Weibull density to compare the standard and modified Meczarisk procedure with sequential procedure based on moment estimator and m.l.e for different θ and d. The results are tabulated in Table 3.1.

Remark 3.1: Modified Meczarisk procedure attains the desired confidence level than the standard Meczarisk procedure though the average sample size is more.

Remark 3.2: Both modified Meczarisk procedure and sequential procedure based on m.l.e achieves the desired confidence level. But the later one require much less average sample size.

Remark 3.3: Though the sequential procedure based on moment estimator require smallest average sample size than any other procedures but does not achieve the desired confidence level.

Remark: From the above simulation studies, it is clear that the modified Meczarisk procedure suggested by Schaalje et. al. (2001) is useful and informative in case of Weibull density. However it does not perform better than the sequential procedure based on m.l.e in case of Pareto density.

Table 3.1: The confidence levels and average sample size of standard and modified Meczarski and sequential procedures based on m.l.e and moment estimator.

				Mec. Procedure		Mod. procedure		ME procedure		MLE procedure	
β	d	α	F-Size	M-S_α	M-Size	MM – S_α	MM-Size	ME S_α	ME Size	MLE S_α	MLE Size
0.5	0.1	0.001	6	0.0158	4	0.001	23	0.024	2	0.0002	5
o.5	0.1	0.01	4	0.0174	3	0.0008	18	0.0942	2	0.0004	3
0.5	0.2	0.001	5	0.0104	3	0.0014	19	0.0142	2	0.0002	4
0.5	0.2	0.01	4	0.0096	3	0.001	16	0.0492	2	0.0004	3
0.5	0.3	0.001	5	0.0044	3	0.0014	18	0.0144	2	0.0004	4

0.5	0.3	0.01	3	0.0072	2	0.0008	14	0.0334	2	0	3
0.5	0.4	0.001	4	0.0064	3	0.0008	17	0.0264	2	0.0004	3
0.5	0.4	0.01	3	0.0038	2	0.0008	14	0.0254	2	0	2
0.5	0.5	0.001	4	0.0026	2	0.0008	16	0.0162	2	0	3
0.5	0.5	0.01	3	0.003	2	0.0006	13	0.0214	2	0.0002	2
1	0.1	0.001	11	0.0632	7	0.0028	32	0.028	7	0.0006	10
1	0.1	0.01	8	0.0714	5	0.0022	26	0.0158	4	0.0012	6
1	0.2	0.001	7	0.0384	4	0.002	25	0.002	4	0.0004	6
1	0.2	0.01	5	0.0398	3	0.0012	20	0.0098	3	0.0002	4
1	0.3	0.001	6	0.0242	3	0.002	22	0.0002	3	0.0006	5
1	0.3	0.01	4	0.0218	3	0.0016	17	0.0094	2	0.0002	3
1	0.4	0.001	5	0.0136	3	0.0012	20	0.0008	3	0.0004	4
1	0.4	0.01	4	0.0142	2	0.0014	15	0.009	2	0.0004	3
1	0.5	0.001	5	0.0104	3	0.0004	18	0.0008	2	0	4
1	0.5	0.01	3	0.0102	2	0.0008	15	0.0066	2	0.0002	3

References

1) Meczarski M. (1985): Sequential Estimation of the Minimum of a Random Variable. *Statistical Probability Letter*, Vol. 3: 63-64.
2) Schaalje G. B., Johnson M. R. and Bodily C. H. (2001). A note on a Sequential Method for estimating the Minimum of a Random Variable. *Communication in Statistics (Simulation)*, Vol. 30, 1: 91-98.