Review Of Research
Vol. 3 | Issue. 12 | Sept. 2014
Impact Factor : 2.1002 (UIF)
ISSN:-2249-894X
Available online at www.ror.isrj.org

ORIGINAL ARTICLE

FIXED AREA SEQUENTIAL CONFIDENCE REGION FOR THE PARAMETERS OF UNIFORM (θ_1, θ_2) DISTRIBUTION

H. S. PATIL

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

Abstract: In the literature (1-α)-level sequential fixed width confidence intervals for the parameter θ of U(0, θ) distribution have been obtained. Here the lower limit of the support is known. In this article we propose some sequential procedures to obtain (1-α)-level fixed area confidence regions for the parameters of U(θ_1, θ_2) distribution and their performances are evaluated based on extensive simulation study.

Keywords: Sequential procedures; Triangular region; Average sample number (ASN); Coverage probability; Sample range; Shortest length criteria; **A**n unbiased estimate; Modified procedures; Two stage procedure; Simulation study.

AMS (2000) Subject Classification: Primary 62L12; Secondary 62F25.

1. INTRODUCTION

Let X_1, X_2 …Xn be independent identically distributed (IID) random variables with common probability density function (pdf)

$$f(x, \theta_1, \theta_2) = 1/(\theta_2 - \theta_1), \text{ if } \theta_1 \le x \le \theta_2, \theta_1 < \theta_2$$
$$= 0 \text{ otherwise.} \qquad (1.1)$$

The maximum likelihood estimator (MLE) of (θ_1, θ_2) is ($X_{(1\ n)}$, $X_{(n\ n)}$). Let R = $X_{(n\ n)}$ - $X_{(1\ n)}$ be the sample range. Based on pivotal R/(θ_2 – θ_1), [1] has shown that {R, R/c} is the shortest confidence interval of level (1- α) for (θ_2 – θ_1), where c (< 1) is the solution of equation

$$c^{n-1}(c(n-1) - n) + \alpha = 0, 0 < \alpha <1. \qquad (1.2)$$

[2] has proposed $C(\underset{\sim}{X}) = \{\underset{\sim}{\theta} : R \leq \theta_2 - \theta_1 \leq R/c,\ \theta_1 \leq X_{(1n)} < X_{(n\,n)} \leq \theta_2\}$ as a (1- α) level confidence set for $\underset{\sim}{\theta} = (\theta_1, \theta_2)$), where c is given by (1.2). The area of $C(\underset{\sim}{X})$ is $R^2\,(1-c)^2/2c^2$, which is random. Further it is shown that $C(\underset{\sim}{X})$ is unbiased but not UMA for $\underset{\sim}{\theta}$ and $C(\underset{\sim}{X})$ has largest confidence level in the class of confidence sets of the same Lebesgue measure. Equivalently $C(\underset{\sim}{X})$ has least Lebesgue measure amongst all those having the same confidence level. Based on such a confidence set we propose sequential procedures to obtain (1-α)-level fixed area confidence sets for the parameters of $U(\theta_1, \theta_2)$ distribution and evaluate their performances based on extensive simulation study.

Uniform distribution plays an important role as a statistical model for physical, biological and social phenomena. For example continuous uniform distribution is an appropriate model for i) inter occurrence time of certain atomic processes, ii) time to wait for the service from a very busy sever, iii) errors arising after rounding floating point numbers up to the nearest integer and iv) time to convert analog (like an image or sound signal) to digital form. More over uniform distribution is often used as a non-informative prior in Bayesian inference.

In the following section we obtain some result related to fixed sample size procedure.

2. Preliminaries

Let $X_1, X_2, \ldots X_n$, be n IID $U(\theta_1, \theta_2)$ variables. For d > 0, consider the confidence region

$$CR_n(d) = \{\underset{\sim}{\theta} = (\theta_1, \theta_2): (X_{(1\,n)} - \theta_1) + (\theta_2 - X_{(n\,n)}) < d,\ \theta_1 \leq X_{(1n)} < X_{(n\,n)} \leq \theta_2\}. \quad (2.1)$$

Note that $CR_n(d)$ is a triangular region of area $d^2/2$ and is described below.

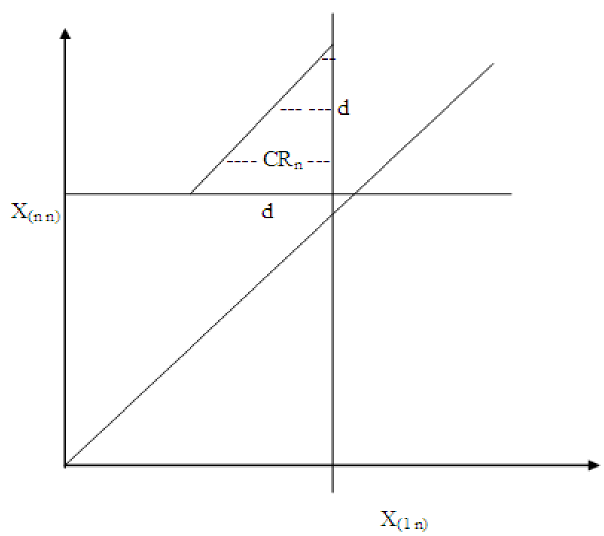

Figure 2.1: Confidence region $CR_n(d)$

For if the confidence region of the form (2.1) and has at most area A_0, it is necessary and sufficient for d to be less than or equal to $d_0 = (2A_0)^{1/2}$. In order to reduce the sample size and/or to increase the coverage probability, in the following we consider $d = (2A_0)^{1/2}$.

Let $R_n = X_{(n\ n)} - X_{(1\ n)}$. The joint distribution of $(X_{(1\ n)}, X_{(n\ n)})$ and of $(X_{(1\ n)}, R_n)$ are respectively given by

$$f(x, y) = n(n-1)\,(y - x)^{n-2} / (\theta_2 - \theta_1)^n \; [\theta_1 \le x < y \le \theta_2]$$

$$= 0 \text{ otherwise}$$

and
$$g(x, r) = n(n-1)\, r^{n-2} / (\theta_2 - \theta_1)^n,\ \theta_1 \le x \le \theta_2,\ 0 < r < \theta_2 - \theta_1,\ r + x < \theta_2$$

$$= 0 \text{ otherwise.}$$

The random variables $X_{(1\ n)}$ and R_n are not independent and the marginal density of R_n is given by

$$g(r) = n(n-1)\, r^{n-2} (\theta_2 - \theta_1 - r) / (\theta_2 - \theta_1)^n \; [r < \theta_2 - \theta_1].$$

The distribution of R_n depends on $\underset{\sim}{\theta}$ only through $\theta_2 - \theta_1 = \delta$ (say). Let $V_n = R_n/\delta$. The pdf of V_n is,

$$g(v) = n(n-1)\, v^{n-2} (1-v),\ 0 < v < 1.$$

That is V_n has beta distribution of first kind with parameters n - 1 and 2. Consider

$$P(\underset{\sim}{\theta} \in CR_n(d)) = P(1 - V_n < d/\delta)$$

$$= P(V_n > 1 - d/\delta) = \begin{cases} 1 & \text{if } d/\delta \ge 1 \\ 1 - (1 - d/\delta)^{n-1}((n-1)d/\delta + 1) & \text{if } d/\delta < 1 \end{cases}.$$

Let $d/\delta < 1$, then $N(d, \delta)$, the least sample size so that $P(\underset{\sim}{\theta} \in CR_n(d)) \ge 1 - \alpha$ is given by

$$N(d, \delta) = \inf\{n\ (\ge 2): (1 - d/\delta)^{n-1}((n-1)d/\delta + 1) < \alpha\}$$

$$= \inf\{n\ (\ge 2): K(n, d, \delta) < \alpha\}, \qquad (2.2)$$

where $K(n, d, \delta) = (1 - d/\delta)^{n-1}((n-1)d/\delta + 1)$. In the following we prove some properties of $K(n, d, \delta)$ and used to obtain sequential procedures

Lemma 2.1: Let $n \ge 2$ and $d > 0$ be a fixed number, for simplicity $K(n, d, \delta)$ be denoted by $K(n, \delta)$. Then

(i) for each δ fixed, $K(n, \delta)$ is decreasing in n and $K(n, \delta)$ tends to 0 as n tends to ∞.

(ii) for each n fixed, K(n, δ) is increasing in δ and increases to 1.

(iii) for each n fixed, K(n, δ) is decreasing in d.

Proof: (i) Let a = d/δ, we note that 0 < a < 1. Consider log K(n, δ) = (n-1) log(1- a) + log((n-1)a +1). By considering n as a real number, we have ∂ log K(n, δ)/∂n = log(1-a) + a/((n-1)a +1). To prove result it is enough to show that log(1-a) < - a/((n-1)a +1), equivalently 1- a < $e^{-a/((n-1)a+1)}$. For 0 <a < 1, we have 1- a < e^{a} and a > a/{(n-1)a +1} for n ≥ 2. Thus 1- a < $e^{-a/((n-1)a+1)+1}$ < $e^{a/((n-1)a+1)+1}$ < e^{a}, which is true. Further by L'Hospital rule, $\lim_{n \to \infty}$ K(n, δ) = $\lim_{n \to \infty}$ a /((1-a) $(1-a)^{-n}$ log(1-a)) = $\lim_{n \to \infty}$ a$(1 - a)^{n}$/((1-a) log(1-a)) = 0. Hence part (i). The graphs of the function K(n, δ) for n ≥ 2, d= 1 and δ > d are given below.

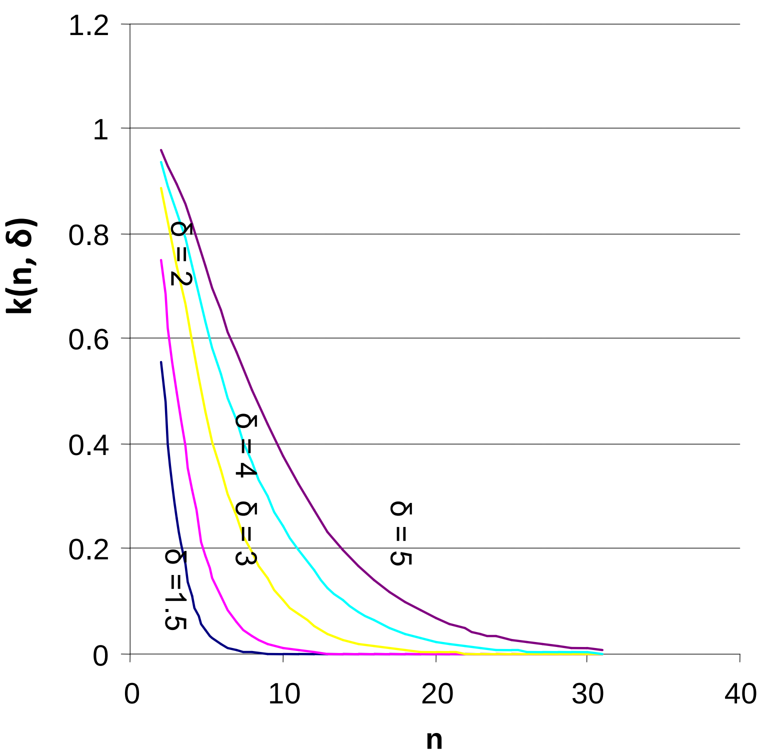

Figure 2.2: Graph of k(n, δ) against n

(ii) Since δ – d > 0, ∂ log K(n, δ) /∂δ = (n-1)d/($δ^2$ – dδ) - (n-1)d/($δ^2$ + (n -1)dδ) = n(n-1)d/(δ(δ-1)(δ – d + nd)) > 0, for each n fixed. Hence K(n, δ) is increasing in δ and it increases to 1.

(iii) Similar to case (ii). □

The graphs of the function K(n, δ) as a function of δ, when d = 1 and different n are given below.

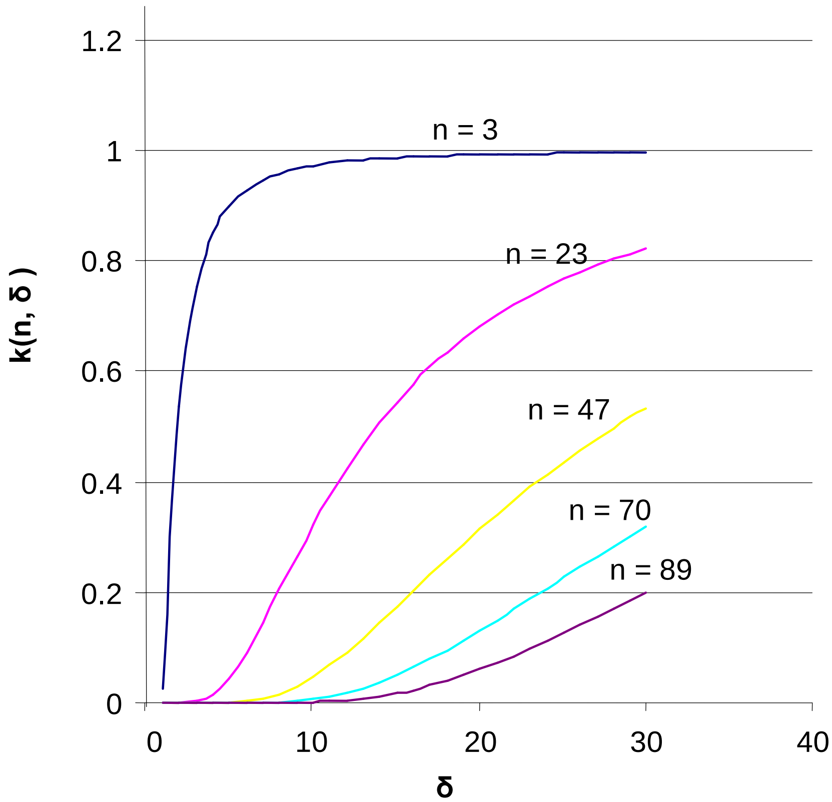

Figure 2.3: Graph of k(n, δ) against δ

We note that the least sample size n for which $P(\underset{\sim}{\theta} \in CR_n(d)) \geq 1 - \alpha$ (equivalently $K(n, \delta) < \alpha$) depend on δ, which is unknown. In the following we propose some sequential stopping rules N of (1- α)- level fixed area and propose the confidence region C(X) for $\underset{\sim}{\theta} = (\theta_1, \theta_2)$ given by,

$$C_N(X) = \{\underset{\sim}{\theta} : R_N < \theta_2 - \theta_1 \leq R_N + d, \theta_1 \leq X_{(1\,N)} < X_{(N\,N)} \leq \theta_2\}.$$

3. Some sequential procedures to find confidence region of fixed area for $\underset{-}{\theta} = (\theta_1, \theta_2)$

In the following, by using Lemma 2.1, we propose stopping rules; based on the lower bound of δ, an unbiased estimate of δ, the shortest length criteria and a two-stage procedure.

Further by extensive simulation we examine for the attainability of required coverage probability (1- α).

I. Based on lower bound of δ:

In the following we propose a sequential stopping rule N_1 that depends on R_n, an almost sure sharp lower bound for δ. The rule N_1 is essentially is obtained by replacing K(n, d, δ) by K(n, d, R_n). Define

$$N_1 = \inf \{n\ (\geq 2): K(n, d, R_n) < \alpha\}.$$

$$= \inf \{n\ (\geq 2):(1- d/R_n))^{n-1}(1+(n-1)dR_n) < \alpha\}= N(d, R). \qquad (3.1)$$

For d fixed, let n(δ, α) be the least positive number satisfying (2.2) (that is exact minimum sample size). Since $K(n, \delta) = (1- d/\delta)^{n-1}((n-1)d/\delta +1)$ is increasing in δ and $R_n < \delta$ almost surely (a.s.), we have $N_1 \leq n(\delta, \alpha)$ a.s.. Thus N_1 is bounded and hence it is a proper stopping random variable.

It is difficult to obtain an expression for $P(\underset{\sim}{\theta} \in C_{N1}(X))$. However as $N_1 \leq n(\delta, \alpha)$ a.s., the coverage probability $P(\underset{\sim}{\theta} \in C_{N1}(X))$ will not be larger than (1- α). In the following we carry out the simulation study with 30,000 iterations with α = 0.05, d =1, $\theta_1 = 0$ and for various values of θ_2. The simulated average sample number (ASN) and the coverage probability (COV) of the rule N_1 are obtained in Table 3.1. It is observed that coverage probability (COV) of the rule N_1 is not always exceeding (1- α). In the following we propose a modified rule

$$M_1 = N(fd, R), \quad 0< f < 1,$$

where f is a suitable fraction. By the definition of N_1 and Lemma 2.1, we have $N(d_1, .) > N(d_2, .)$ for $d_1 < d_2 \leq 1$. Thus we have $M_1 \geq N_1$**.** RR However we $M_1 \leq n(\delta/f, \alpha)$ a.s and hence M_1 is also a proper stopping random variable. RR Let f = 1- kα. For M_1 to be as least as possible, f has to be as large as possible. Hence we choose k as a least positive number that might depend on α but not on $\underset{\sim}{\theta}$ and confidence region based on rule M_1 attains desired coverage probability. With α = 0.05, it is observed that, for k < 10, rule M_1 does not attain 95% coverage and for k ≥ 10, rule M_1 attain 95% coverage. For k =10, the simulated values of ASN (M-ASN) and coverage probability (M-COV) of the rule M_1 for different values of θ_2 are tabulated in Table 3.1. Though M-ASN is larger than ASN, coverage probability is attained for the rule M_1. With 30,000 iterations, Table 3.1 gives minimum sample size, simulated values of sample size's and the coverage probabilities of the rules N_1 and M_1 for α = 0.05.

II. Based on unbiased estimate of δ: We know that $U_n = (n+1)R_n/(n-1)$ is an unbiased estimate of δ. Let $N_2 = N(d, U_n)$ and $M_2 = N(fd, U_n)$. Note that $M_2 \geq N_2$ Since $U_n < R_n < \delta$ almost surely (a.s.), we have $N_2 \leq n(\delta, \alpha$ and $M_2 \leq n(\delta/f, \alpha)$.) a.s. we have N_2 , and M_2 are proper stopping random variables. Following similar study as in case I**,** an appropriate value of k is 11 for α = 0.05. With f = 1 - 11α, we carry out the simulation study and the results are tabulated in Table 3.1.

III. Based on shortest length criteria: Based on [1] shortest confidence interval, we propose a purely sequential stopping rule $N_3(d)$ such that

$$N_3 = \inf\{n \geq 2: R_n(1-c_n) < d\} \tag{3.2}$$

where c_n is given by (1.2). So we take $M_3 = \inf\{n \geq 2: R_n(1-c_n) < fd\}$. Since $c_n < 1$, $R_n < d < \delta$ almost surely (a.s.) and we have $N_3 \leq M_3 \leq n(\delta/f, \alpha)$ a.s. Thus N_3 and M_3 are proper stopping random variable. An appropriate value of k is 7.5 for α = 0.05. With f = 1 - 7.5α simulated results are tabulated in Table 3.1.

IV. Two stage procedure: [2] has proposed a two stage procedure (T) as below.

(i) Take a random sample of size m and obtain R_m and $c_m(\alpha_1)$ given by (1.2) which in turn implies $P(R_m < \delta < R_m/c_m) \geq 1-\alpha_1$. If $(1/c_m(\alpha_1) - 1) < d$, stop and take $(1-\alpha_1)$-level confidence region for $\underset{\sim}{\theta} = (\theta_1, \theta_2)$ as $(R_m, R_m + d)$ otherwise go to second stage by taking $R_m/c_m(\alpha_1)$ as an estimate for δ (in fact upper bound).

(ii) Take an independent random sample $N(R_m, c_m(\alpha_1), d, \alpha_2)$ such that $N_4(d) = \inf\{n \geq 2: (1- c_m(\alpha_1)d/R_m)^{n-1}((n-1)c_m(\alpha_1)d/R_m +1) < \alpha_2\}$

Then take $(1-\alpha)$-level confidence region for $\underset{\sim}{\theta} = (\theta_1, \theta_2)$ as $(R_N, R_N + d)$. Note that $(1-\alpha_1)(1-\alpha_2) = (1-\alpha)$ **a**nd the two stage procedure T can be shown to be closed (refer [2]). With 30,000 iterations, the simulated sample size (T-ASN) and coverage probability (T-COV) for $\theta_1 = 0$, d $=1$, m $=5$, α_1= 0.02, α = 0.05 and different values of θ_2 are tabulated below.

Table 3.1: Exact minimum sample size, Simulated ASN, Coverage Probabilities of the rules N_i's and M_i's and the two stage procedure T

Rules	θ_2	1.1	5.1	10.1	15.1	19.1	100	400
	$n(\delta, \alpha)$	3	23	47	70	89	473	1896
N_1	ASN	22.26	18.73	43.50	67.62	86.53	470.62	1893.09
	COV.	1	0.7916	0.9034	0.9261	0.9273	0.9438	0.9482
M_1	M-ASN	22.31	44.73	92.83	139.81	177.78	945.45	3791.81
	M-COV	1	0.9961	0.9989	0.9988	0.9989	0.9991	0.9992
N_2	ASN	5.89	22.03	46.17	70.03	88.99	472.91	1896.11

	COV.	1	0.8844	0.9283	0.9373	0.9498	0.9485	0.9489
M_2	M-ASN	9.27	52.13	104.91	157.72	199.79	1052.73	4215.47
	M-COV.	1	0.9956	0.9978	0.9991	0.9986	0.9995	0.9995
N_3	ASN	6.35	25.20	49.03	72.71	91.69	475.58	1898.96
	COV.	0.9371	0.9312	0.9421	0.9471	0.9475	0.9485	0.9491
M_3	M-ASN	9.24	39.76	77.76	115.71	146.13	760.15	3019.92
	M-COV.	0.9524	0.9858	0.9921	0.9935	0.9939	0.9951	0.9971
T	T-ASN	17.56	71.07	137.44	204.70	258.09	1336.04	5330.09
	T-COV	0.9997	0.9970	0.9967	0.9973	0.9978	0.9973	0.9973

Remark 3.1: The least values of k for different α and the modified rules M_i's are given in table 3.2.

Table 3.2: Minimum values of k.

Rules	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
M_1	47	10	4.5
M_2	60	11	5
M_3	40	7.5	4

Remark 3.2: From tables 3.1-3.4, it is clear that the procedure M_3 based on shortest length criterion gives desired coverage with smaller ASN as compared to other modified and two stage procedures. This is true for all α. Hence we recommend sequential procedure M_3. Further the simulation study overall reveals that the coverage probability increases with increase in $\theta_2 - \theta_1$.

REFERENCES

[1] Ferentinos K. K. *Shortest Confidence Intervals for the Families of Distributions Involving Truncation Parameters.* The American Statistician, 44(1990), 2, pp 167-168.

[2] Shirke D. T. *A Study of the Confidence Regions of the Parameters.* Ph. D. Thesis (Unpublished), Shivaji University, Kolhapur, 1993.