DISTRIBUTION-FREE INTERVAL ESTIMATION OF THE LARGEST α-QUANTILE

BY

M. HASEEB RIZVI and K. M. LAL SAXENA

TECHNICAL REPORT NO. 193

APRIL 21, 1972

THIS RESEARCH WAS SPONSORED BY THE ARMY RESEARCH OFFICE
OFFICE OF NAVAL RESEARCH, AND AIR FORCE OFFICE OF
SCIENTIFIC RESEARCH BY CONTRACT NO.
N00014-67-A-0112-0053 (NR-042-267)

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Department of Statistics
Stanford University, Calif.

2a. REPORT SECURITY CLASSIFICATION

2b. GROUP

3. REPORT TITLE
Distribution-Free Interval Estimation of the Largest α-Quantile

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Technical Report

5. AUTHOR(S) (Last name, first name, initial)
RIZVI, M. Haseeb and SAXENA K. M. Lal

6. REPORT DATE
April 21, 1972

7a. TOTAL NO. OF PAGES
10

7b. NO. OF REFS
4

8a. CONTRACT OR GRANT NO.
N00014-67-A-0112-0053

b. PROJECT NO.
NR-042-267

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)
No. 193

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. AVAILABILITY/LIMITATION NOTICES
Distribution of this document is unlimited

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Statistics & Probability Program
Office of Naval Research Code 436
Arlington, Va.

13. ABSTRACT

A procedure based on order statistics is given for the interval estimation of the largest α-quantile of several continuous distributions. Results on infimum of coverage probability are obtained for one-sided and two-sided intervals. An optimality criterion is proposed and an algorithm is given for two-sided intervals. Large sample approximations are also considered.

DD FORM 1 JAN 64 1473

UNCLASSIFIED
Security Classification

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
distribution-free						
interval estimation						
largest α-quantile						
order statistics						

INSTRUCTIONS

1. **ORIGINATING ACTIVITY:** Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. **REPORT SECURITY CLASSIFICATION:** Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. **GROUP:** Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. **REPORT TITLE:** Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. **DESCRIPTIVE NOTES:** If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. **AUTHOR(S):** Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. **REPORT DATE:** Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. **TOTAL NUMBER OF PAGES:** The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. **NUMBER OF REFERENCES:** Enter the total number of references cited in the report.

8a. **CONTRACT OR GRANT NUMBER:** If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. **PROJECT NUMBER:** Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. **ORIGINATOR'S REPORT NUMBER(S):** Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. **OTHER REPORT NUMBER(S):** If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. **AVAILABILITY/LIMITATION NOTICES:** Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ______________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ______________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ______________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. **SUPPLEMENTARY NOTES:** Use for additional explanatory notes.

12. **SPONSORING MILITARY ACTIVITY:** Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. **ABSTRACT:** Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS), (S), (C), or (U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. **KEY WORDS:** Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

DD FORM 1 JAN 64 1473 (BACK)

Unclassified

Security Classification

DISTRIBUTION-FREE INTERVAL ESTIMATION OF
THE LARGEST α-QUANTILE

by

M. Haseeb Rizvi and K. M. Lal Saxena

TECHNICAL REPORT NO. 193
April 21, 1972

PREPARED UNDER CONTRACT N00014-67-A-0112-0053
(NR-042-267)
OFFICE OF NAVAL RESEARCH

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

DISTRIBUTION-FREE INTERVAL ESTIMATION OF THE LARGEST α-QUANTILE*

by

M. Haseeb Rizvi and K. M. Lal Saxena
Stanford University and University of Nebraska

1. Introduction and Formulation of the Problem

Ordering of several unknown parameters is a problem of wide practical applications. Saxena and Tong [3] and Saxena [2] have recently constructed confidence intervals for the largest location and scale parameters respectively. This paper deals with the distribution-free interval estimation of the largest α-quantile of several continuous distributions.

Consider $k(\geq 1)$ distributions with unknown continuous cdfs F_i, $i=1,\ldots,k$. Let $x_\alpha(F_i)$ denote the unique α-quantile $(0 < \alpha < 1)$ of F_i. If $x_\alpha(F_i)$ is not unique, it can be defined to be so in an obvious manner. Define $\theta = \max_{1\leq i\leq k} x_\alpha(F_i)$. For a specified constant γ, a random interval I is desired such that

$$\inf_\Omega P(\theta \in I) \geq \gamma \tag{1}$$

where Ω denotes the set of all possible k-tuples $(F_1,F_2,\ldots,F_k)$. Such an interval I, based on order statistics of random samples of equal sizes from each F_i, is proposed below.

---

*The second author's work was supported in part by University of Nebraska Research Council.

2. Proposed Procedure and Its Probability of Coverage

Independent random samples of common size n are taken from each of the k distributions. Let $Y_{r,i}$ denote the r^{th} order statistic from F_i and $Y_r = \max_{1 \leq i \leq k} Y_{r,i}$ for $r=1,\ldots,n$; define $Y_0 = -\infty$ and $Y_{n+1} = +\infty$. For $s < t$, consider the random interval $I_0 = (Y_s, Y_t)$ and assert that $\theta \in I_0$, where s and t are chosen so as to satisfy (1).

With $G_r(p)$ denoting the incomplete beta function

$$(2) \qquad G_r(p) = r\binom{n}{r}\int_0^p u^{r-1}(1-u)^{n-r}du = \sum_{j=r}^{n}\binom{n}{j}p^j(1-p)^{n-j} ,$$

the cdf of $Y_{r,i}$ is given by $G_r(F_i(y))$. We will adopt the convention that $G_0(\cdot) \equiv 1$ and $G_{n+1}(\cdot) \equiv 0$. The probability of coverage of θ by I_0 is then

$$(3) \quad P(\theta \in I_0) = P(Y_s \leq \theta) - P(Y_t \leq \theta)$$

$$= P(Y_{s,i} \leq \theta,\ i=1,\ldots,k) - P(Y_{t,i} \leq \theta,\ i=1,\ldots,k)$$

$$= \prod_{i=1}^{k} G_s(F_i(\theta)) - \prod_{i=1}^{k} G_t(F_i(\theta)) .$$

We know that $F_i(\theta) \geq \alpha$ for $i=1,\ldots,k$ with equality for at least one i. Hence without any loss of generality we assume that $F_k(\theta) = \alpha$. Thus (3) becomes

(4) $$P(\theta \epsilon I_o) = G_s(\alpha) \prod_{i=1}^{k-1} G_s(F_i(\theta)) - G_t(\alpha) \prod_{i=1}^{k-1} G_t(F_i(\theta)) .$$

3. Minimization of the Probability Coverage

For one-sided random intervals the minimization over Ω of $P(\theta \epsilon I_o)$ is given by Theorem 1. The proof of the theorem follows from considerations of (4) and noting that $G_o(\cdot) \equiv 1$, $G_{n+1}(\cdot) \equiv 0$ and $\alpha \leq F_i(\theta) \leq 1$ for each i. The details are omitted.

Theorem 1

(a) For $s > 0$, $t = n + 1$, that is, with $I_o = (Y_s, \infty)$,

(5) $$\inf_{\Omega} P(\theta \epsilon I_o) = G_s^k(\alpha);$$

and (b) for $s = 0$, $t < n + 1$, that is, with $I_o = (-\infty, Y_t)$,

(6) $$\inf_{\Omega} P(\theta \epsilon I_o) = 1 - G_t(\alpha) .$$

Note that for any $s < t$, $G_s(x) > G_t(x)$. Therefore in (a) of Theorem 1, we choose s to be the largest integer such that the right side of (5) exceeds γ of requirement (1) with $0 < \gamma < \{1-(1-\alpha)^n\}^k$. Further in (b) of Theorem 1, we choose t to be the smallest integer such that the right side of (6) exceeds γ with $0 < \gamma < 1 - \alpha^n$. It is clear from the above upper bounds on γ that, for fixed k and α, I_o can satisfy (1) for any value of γ between 0 and 1, provided n is taken large enough.

Next we consider the minimization over Ω of $P(\theta \epsilon I_o)$ for two-sided random intervals.

Theorem 2 (two-sided intervals)

For $0 < s < t < n + 1$,

$$(7) \qquad \inf_{\Omega} P(\theta \in I_o) = \min(G_s(\alpha) - G_t(\alpha),\ G_s^k(\alpha) - G_t^k(\alpha)) .$$

Proof

Since $P(\theta \in I_o)$, given by (4), involves F_i's evaluated at θ (constant) and $F_i(\theta) \geq \alpha$ for $i=1,\ldots,k-1$, we can write $F_i(\theta) = \alpha + \delta_i$, where $0 \leq \delta_i \leq 1 - \alpha$. This enables us to reparametrize (4) as a function of the δ_i's . Consequently the problem of minimization of (4) over $\Omega = \{(F_1,\ldots,F_k): F_i$ is continuous for each $i\}$ is reduced to its minimization over $\{(\delta_1,\ldots,\delta_{k-1}): 0 \leq \delta_i \leq 1 - \alpha,\ i=1,\ldots,k-1)\}$. We have

$$(8) \qquad P(\theta \in I_o) = G_s(\alpha) \prod_{i=1}^{k-1} G_s(\alpha+\delta_i) - G_t(\alpha) \prod_{i=1}^{k-1} G_t(\alpha+\delta_i)$$

$$= J(\delta_1,\ldots,\delta_{k-1}),\ \text{say} .$$

For some j, fix $\delta_1,\ldots,\delta_{j-1},\delta_{j+1},\ldots,\delta_{k-1}$ and consider $\partial J/\partial\delta_j$. Using (2), we define

$$g_r(p) = \frac{d}{dp} G_r(p) = r\binom{n}{r}p^{r-1}(1-p)^{n-r}, \quad 0 \leq p \leq 1,$$

and observe that $g_t(p)/g_s(p)$ is increasing in p for $t > s$.

Let

$$A = \prod_{\substack{i=1 \\ i\neq j}}^{k-1} G_s(\alpha+\delta_i), \quad B = \prod_{\substack{i=1 \\ i\neq j}}^{k-1} G_t(\alpha+\delta_i), \quad A \geq B .$$

Then from (8) we obtain

$$\frac{\partial J}{\partial \delta_j} = AG_s(\alpha)g_s(\alpha+\delta_j)[1 - \frac{BG_t(\alpha)g_t(\alpha+\delta_j)}{AG_s(\alpha)g_s(\alpha+\delta_j)}] .$$

Since $g_t(\alpha+\delta_j)/g_s(\alpha+\delta_j)$ is increasing in δ_j, it follows that the expression inside the brackets is decreasing in δ_j. Hence we conclude that $\partial J/\partial\delta_j$ has either the same sign at every value of δ_j or at most one change of sign from positive to negative and consequently $\inf_{\delta j} J$ is either at $\delta_j = 0$ or at $\delta_j = 1 - \alpha$. This conclusion is valid for every other j. Therefore infimum of J is achieved when a certain number m of δ_j's are zero and the rest equal to $1 - \alpha$. Define

(9) $$H_m = H_m(s,t) = G_s^{m+1}(\alpha) - G_t^{m+1}(\alpha), \quad 0 \leq m \leq \infty .$$

Differentiating H_m with respect to m, and noting that $G_s(\alpha) > G_t(\alpha)$, it is seen that $\partial H_m/\partial m$ either has the same sign for every value of m or has at most one change of sign from positive to negative. Hence

(10) $$\min_{m=0,1,\ldots,k-1} H_m = \min(H_0, H_{k-1}) .$$

This proves the theorem.

Note that for fixed k and α, I_o can satisfy (1) provided γ lies between 0 and

$$\min_{r=1,k} \left([1-(1-\alpha)^n]^r - \alpha^{nr} \right) = p(k,\alpha,n), \quad \text{say} .$$

Clearly, $p(k,\alpha,n)$ can be made arbitrarily close to 1 by taking n large enough.

4. An Optimum Two-Sided Random Interval

For specified k,α,γ and n, the choice of integers s and t, such that the two-sided random interval $I_o = (Y_s,Y_t)$ satisfies (1), may not be unique unless some criterion for an optimum choice is introduced. For the case $k=1$, Wilks [4] proposed an optimality criterion that requires the ranks s (of Y_s) and t (of Y_t) to be as close together as possible. Extending this criterion to $k \geq 1$, we would be interested in choosing s and t so that the rank-difference $(t-s)$ is minimized for a preassigned γ. For this purpose we present the following algorithm. Let c be a positive integer less than n and consider $I_o = (Y_s,Y_{s+c})$. Denote by $Q(s,c)$ the infimum of $P(Y_s < \theta < Y_{s+c})$ over Ω as given by (7). For every fixed c, let $s_o(c)$ be that value of s for which

$$Q(s_o(c),c) = \max_{1 \leq s \leq n-c} Q(s,c) .$$

Now choose the smallest c, call it c_o, such that

$$Q(s_o(c_o), c_o) \geq \gamma .$$

Then the optimum choice of the random interval satisfying the requirement (1) is (Y_s, Y_{s+c}) with $c = c_o$ and $s = s_o(c_o)$.

For moderate values of n, say $n \leq 50$, the above algorithm can be carried out in an easy manner using the incomplete beta function tables or the more readily available binomial tables for smaller values of n in view of (2). For example, when $k=4$, $\alpha=0.5$, $\gamma=0.90$, and $n=25$, we obtain $s=8$ and $t=17$. In this illustration, it is interesting to note that even for $k=1,2,3$ and the same values of α, γ and n, we obtain $s=8$ and $t=17$.

5. Large Sample Approximations

For large n, using normal approximation to binomial in (2), with $\Phi(\cdot)$ denoting the standard normal cdf, we obtain

$$G_r(\alpha) \approx \Phi((-r+n\alpha)/(n\alpha(1-\alpha))^{1/2}) . \tag{11}$$

For one-sided intervals of Theorem 1, using (11), in the case (a) we take s to be the largest integer such that

$$s \leq n\alpha + (n\alpha(1-\alpha))^{1/2}\Phi^{-1}(1-\gamma^{1/k}),$$

and in the case (b) we take t to be the smallest integer such that

$$t \geq n\alpha + (n\alpha(1-\alpha))^{1/2}\Phi^{-1}(\gamma) .$$

For the two-sided optimum random interval described in Section 4, using (11), we have

$$Q(s,c) = \min\{\Phi(d-x) - \Phi(-x),\ \Phi^k(d-x) - \Phi^k(-x)\} \quad (12)$$

where

$$d = c(n\alpha(1-\alpha))^{-1/2}, \quad x = (n\alpha - s)(n\alpha(1-\alpha))^{-1/2} .$$

Dudewicz and Tong [1] show that for any given d, the value of x, say x_0, that maximizes the right side of (12) is either the value for which the two terms within braces of (12) are equal or the value which maximizes the second of these two terms. Table 1 of [1] gives x_0 for $k=5(1)10(2)14$ and $d=1(0.1)4$. For $k=1$, obviously $x_0 = d/2$; also for $k=2$, $x_0 = \frac{d}{2}$ as shown in [1]. Thus for $k=1,2$ and n large, from (12) it is seen that s and t will be equidistant from $n\alpha$ on either side. Table 2 of [1] gives coverage probability (12) evaluated at x_0 of Table 1. We consider once again the example of Section 4 to illustrate the use of the tables of [1] for adapting our algorithm to large sample sizes. For $k=4$, $\alpha=0.5$, $\gamma=0.90$, and $n=25$, Table 2 gives $d=3.4$. Using $d=3.4$ we obtain from Table 1, $x_0 = 1.452$. Now from (12) with $x = x_0$, we obtain $s=9.$[illegible] and $c=8.$[illegible]. These values of s and c are then the optimum values $s_0(c_0)$ and c_0 respectively of Section 4. Since these optimum values have to be integers, we round them off as $s_0(c_0) = 8$ and $c_0 = 9$. Thus in this example we obtain the random interval obtained previously.

The goodness of the large sample approximation considered above is directly related to the well known convergence of binomial to normal

For commonly used values of α it is felt that this approximation is adequate for sample sizes larger than 50.

In conclusion it should be pointed out that the problem of interval estimation of the smallest α-quantile $\theta' = \min_{1 \le i \le k} x_\alpha(F_i)$ can be handled in a manner analogous to the discussion of this paper by considering the random interval (Y'_s, Y'_t), $s < t$, where $Y'_r = \min_{1 \le i \le k} Y_{r,i}$.

References

[1] Dudewicz, E. J. and Tong, Y. L., "Optimal confidence intervals for the largest location parameter," Proceedings of the Symposium on Decision Theory, Purdue University (1971), to appear.

[2] Saxena, K. M. Lal, "Interval estimation of the largest variance of k normal populations", Journal of the American Statistical Association, 66 (1971), 408-410.

[3] Saxena, K. M. Lal and Tong, Y. L., "Interval estimation of the largest mean of k normal populations with known variances," Journal of the American Statistical Association, 64 (1969), [illegible].

[4] Wilks, S. S., Mathematical Statistics, John Wiley and Sons, New York, 1962.