



DIVISION OF MATHEMATICAL SCIENCES
Southern Methodist University,
Dallas, Texas 75275

THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

SAMPLE SIZES FOR APPROXIMATE INDEPENDENCE OF LARGEST
AND SMALLEST ORDER STATISTICS

by

John E. Walsh

Technical Report No. 3
Department of Statistics THEMIS Contract

This document has been approved for public release
and sale; its distribution is unlimited.

Department of Statistics
Southern Methodist University
Dallas, Texas 75222

THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

SAMPLE SIZES FOR APPROXIMATE INDEPENDENCE OF LARGEST
AND SMALLEST ORDER STATISTICS

by

JOHN E. WALSH

Technical Report No. 3
Department of Statistics THEMIS Contract

September 3, 1968

Research sponsored by the Office of Naval Research
Contract N00014-68-A-0515
Project NR 042-260

Reproduction in whole or in part is permitted
for any purpose of the United States Government.

DEPARTMENT OF STATISTICS
Southern Methodist University

SAMPLE SIZES FOR APPROXIMATE INDEPENDENCE OF LARGEST
AND SMALLEST ORDER STATISTICS

John E. Walsh

Southern Methodist University*

ABSTRACT

Let X_n and X_1 be the largest and smallest order statistics, respectively, of a random sample of size n . Quite generally, X_n and X_1 are approximately independent for n sufficiently large. Minimum n for attaining at least specified levels of independence are developed. Level of independence is measured by the maximum difference between the true values of $P(X_1 \leq x_1, X_n \leq x_n)$ and the corresponding values assuming independence of X_n and X_1 . The results are for small maximum differences (say, at most .02) and apply to all possible distributions for the population sampled. The value of minimum n is the smallest allowable n for the continuous case but can be too large otherwise. Minimum n is finite for all nonzero differences.

INTRODUCTION AND RESULTS

The largest and smallest order statistics of a random sample tend to statistical independence as the sample size increases. That is, consider a random sample of size n and let X_n and X_1 be the largest and smallest order statistics, respectively. Also consider

* Research partially supported by NASA Grant NGR 44-007-028
Also associated with ONR Contract N00014-68-A-0515.

$$P(X_1 \leq x_1, X_n \leq x_n) - P(X_1 \leq x_1)P(X_n \leq x_n), \quad (1)$$

which is nonnegative. As $n \rightarrow \infty$, the maximum of this difference (over x_1 and x_n) tends to zero.

Since any n used is finite, there can be interest in how the maximum difference of (1) is affected by n . More specifically, for a given value of the maximum difference, what is the minimum n such that this value is not exceeded? For example, what is the minimum n such that the maximum difference is at most .001? When the maximum difference is small, there is little error in using $P(X_1 \leq x_1)P(X_n \leq x_n)$ as the joint cumulative distribution function (cdf) for X_n and X_1 .

The expression developed for minimum n is based on approximations but is very accurate when the stated maximum difference is small (say, at most .02). This expression provides the smallest permissible value of n when the population sampled is continuous. A smaller value of n could possibly be allowable when the population cdf $F(x)$ is discontinuous, since $F(x_1)$ and/or $F(x_n)$ might not be able to have the values that maximize (1).

Let δ be the specified value for the maximum difference. At most this value occurs if

$$n \geq \frac{-1}{2 \log_e(1 - \delta e^2)} \{1 + [1 - 4 \log_e(1 - \delta e^2)]^{1/2}\}$$

$$\pm (1/\delta)e^{-2} + 1/2 \pm .1353/\delta + .5, (\delta \leq .01).$$

For example, the maximum difference is at most .005 if $n \geq 28$.

These results, which are applicable for all possible $F(x)$, again show that X_n and X_1 tend to independence as $n \rightarrow \infty$. That is, no

matter how small δ is, there are values of n such that the maximum difference is less than δ (say, at most $\delta/2$).

DERIVATIONS

Let $a = a(x_n)$ and $b = b(x_1)$ be defined by $P(X_n \leq x_n) = e^{-a}$, $P(X_1 \leq x_1) = 1 - e^{-b}$. In the derivations, all values of a and b in the range zero to infinity are considered to be possible (corresponds to the continuous case). Then,

$$F(x_n) = e^{-a/n}, \quad 1 - F(x_1) = e^{-b/n},$$

so that, in general,

$$\begin{aligned} P(X_1 \leq x_1, X_n \leq x_n) &= F(x_n)^n - [F(x_n) - F(x_1)]^n \\ &= e^{-a} - (e^{-a/n} - 1 + e^{-b/n})^n. \end{aligned}$$

If X_n and X_1 are independent,

$$\begin{aligned} P(X_1 \leq x_1, X_n \leq x_n) &= F(x_n)^n - F(x_n)^n [1 - F(x_1)]^n \\ &= e^{-a} - e^{-(a+b)} \end{aligned}$$

Thus, the value of (1), the difference of these two probabilities, can be expressed as

$$e^{-(a+b)} - e^{-a} [1 - e^{a/n} + e^{(a-b)/n}]^n$$

which, by some expansions in terms of $1/n$, equals

$$\begin{aligned} e^{-(a+b)} - e^{-a} \exp[-b - ab/n - ab(a+b)/2n^2 + O(1/n^3)] \\ \approx e^{-(a+b)} \{1 - \exp[-ab/n - ab(a+b)/2n^2]\} \end{aligned}$$

for n sufficiently large (say, $n \geq 8$) and $a + b$ not large. It is to be noted that $a + b \leq -\log_e \delta$ in all cases where the difference is to be at most δ .

This expression is set equal to δ , ($\delta \leq .02$), and the n (not necessarily an integer) yielding this value is determined. Then, this expression for n is maximized with respect to $\underline{a}$ and $\underline{b}$.

First, consider the more crude approximation where terms of order $1/n^2$ are neglected. Then,

$$e^{-(a+b)}(1 - e^{-ab/n}) = \delta$$

so that

$$n \approx -ab/\log_e(1 - \delta e^{a+b})$$

$$\approx (1/\delta)abe^{-(a+b)}.$$

Thus, to this order of approximation, $a = b = 1$ are the maximizing values. That is, the true maximizing values for $\underline{a}$ and $\underline{b}$ should be near unity.

Now consider the approximation where terms of order $1/n^3$ are neglected. This yields the quadratic equation

$$n^2 + nab/\log_e(1 - \delta e^{a+b}) + ab(a+b)/2\log_e(1 - \delta e^{a+b}) = 0,$$

with solution

$$2n = -[ab/\log_e(1 - \delta e^{a+b})]$$

$$\times \{1 + [1 - 2(a+b)(ab)^{-1}\log_e(1 - \delta e^{a+b})]^{1/2}\}.$$

Expansion with respect to δ yields

$$n\delta = abe^{-(a+b)}[1 + (1/2)\delta e^{a+b}]^{-1}[1 + (a+b)(2ab)^{-1}\delta e^{a+b}] + O(\delta^2),$$

so that $\log_e n\delta$ equals

$$\log_e a + \log_e b - a - b - (1/2)\delta e^{a+b} + (a+b)(2ab)^{-1}\delta e^{a+b} + O(\delta^2).$$

This monotonically increasing function of n is maximized with respect to $\underline{a}$ by setting $\partial \log_e n\delta / \partial a$ equal to zero, yielding

$$1/a - 1 - (1/2)\delta e^{a+b} \left[1 + (a+b)/ab - 1/ab + (a+b)/a^2 b \right] + O(\delta^2) = 0.$$

Let the terms of order δ^2 be neglected. Also, since δ is small, the solution for the case where terms of order $1/n^2$ are neglected should be usable in the coefficient of δ . This yields the solution $a = 1$, and a similar analysis yields the solution $b = 1$. Thus, $a = b = 1$ is the maximizing choice (to a good approximation) even when terms of order $1/n^2$ are included. Use of $a = b = 1$, combined with n being an integer, yields the expression stated for determining minimum n for given δ .

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)		2a. REPORT SECURITY CLASSIFICATION	
SOUTHERN METHODIST UNIVERSITY		UNCLASSIFIED	
		2b. GROUP	
		UNCLASSIFIED	
3. REPORT TITLE			
Sample Sizes for Approximate Independence of Largest and Smallest Order Statistics			
4. DESCRIPTIVE NOTES (Type of report and inclusive dates)			
Technical Report			
5. AUTHOR(S) (First name, middle initial, last name)			
John E. Walsh			
6. REPORT DATE		7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
September 3, 1968		5	None
8a. CONTRACT OR GRANT NO.		9a. ORIGINATOR'S REPORT NUMBER(S)	
N00014-68-A-0515		3.	
b. PROJECT NO.		9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	
NR 042-260			
c.			
d.			
10. DISTRIBUTION STATEMENT			
No limitations			
11. SUPPLEMENTARY NOTES		12. SPONSORING MILITARY ACTIVITY	
		Office of Naval Research	
13. ABSTRACT			
Let X_n and X_1 be the largest and smallest order statistics, respectively, of a random sample of size n. Quite generally, X_n and X_1 are approximately independent for n sufficiently large. Minimum n for attaining at least specified levels of independence are developed. Level of independence is measured by the maximum difference between the true values of $P(X_1 \leq x_1, X_n \leq x_n)$ and the corresponding values assuming independence of X_n and X_1. The results are for small maximum differences (say, at most .02) and apply to all possible distributions for the population sampled. The value of minimum n is the smallest allowable n for the continuous case but can be too large otherwise. Minimum n is finite for all nonzero differences.			

UNCLASSIFIED DISTRIBUTION LIST
FOR
SIGNAL ANALYSIS STATISTICS

- | | |
|---|---|
| 1 ARCON Corporation
Lakeside Office Park
North Avenue at Route 128
Wakefield, Massachusetts 01880
ATTN: Dr. A. Albert | 1 University of California
Graduate School of Business
Administration
Los Angeles, California 90024
ATTN: Prof. J. Marschak |
| 1 University of California
Operations Research Center
Institute of Engineering Research
Berkeley, California 94720
ATTN: Prof. R. E. Barlow | 1 University of California
Department of Engineering
Los Angeles, California 90024
ATTN: Prof. R. R. O'Neill |
| 1 University of California
Department of Statistics
Berkeley, California 94720
ATTN: Prof. P. J. Bickel | 1 Carnegie-Mellon University
Graduate School of Industrial
Administration
Pittsburgh, Pennsylvania 15213
ATTN: Prof. G. Thompson |
| 1 University of California
Operations Research Center
Institute of Engineering Research
Berkeley, California 94720
ATTN: Prof. D. Gale | 1 Carnegie-Mellon University
Graduate School of Industrial
Administration
Pittsburgh, Pennsylvania 15213
ATTN: Prof. A. C. Stedry |
| 1 University of California
Department of Economics
Berkeley, California 94720
ATTN: Prof. R. Radner | 1 Case Western Reserve University
Cleveland, Ohio 44106
ATTN: Prof. B. V. Dean |
| 1 University of California
Center for Research in
Management Science
Berkeley, California 94720
ATTN: Prof. W. I. Zangwill | 1 Case Western Reserve University
Systems Research Center
Cleveland, Ohio 44106
ATTN: Prof. M. Mesarovic |
| 1 University of California, San Diego
Department of Mathematics
P. O. Box 109
La Jolla, California 92038
ATTN: Prof. M. Rosenblatt | 1 University of Chicago
Statistical Research Center
Chicago, Illinois 60637
ATTN: Prof. W. Kruskal |
| 1 University of California
Management Sciences Research
Project
Los Angeles, California 90024
ATTN: Prof. J. R. Jackson | 1 Columbia University
Department of Industrial
Engineering
New York, New York 10027
ATTN: Prof. C. Derman |

- 1 University of Connecticut
Department of Statistics
Storrs, Connecticut 06268
ATTN: Prof. H. O. Posten
- 1 Cornell University
Department of Industrial
Engineering
Ithaca, New York 14850
ATTN: Prof. R. E. Bechhofer
- 1 Cornell University
Plant Breeding & Biometry
337 Warren Hall
Ithaca, New York 14850
ATTN: Prof. Walter T. Federer
- 1 Cornell University
Department of Mathematics
White Hall
Ithaca, New York 14850
ATTN: Prof. J. Kiefer
- 1 Decision Studies Group
845 Page Mill Road
Palo Alto, California 94304
ATTN: Dr. Stuart Bessler
- 20 Defense Documentation Center
Cameron Station
Alexandria, Virginia 22314
- 1 Defense Logistics Studies
Information Exchange
Army Logistics Management Center
Fort Lee, Virginia 23801
ATTN: W. B. Wichard
- 1 University of Delaware
Department of Mathematics
Newark, Delaware 19711
ATTN: Prof. R. Remage, Jr.
- 1 Directorate of Data Automation
(Hq. USAF)
AFADABB, Washington, D.C. 20330
- 1 Florida State University
Department of Statistics
Tallahassee, Florida 32306
ATTN: Prof. R. A. Bradley
- 1 Florida State University
Department of Statistics
Tallahassee, Florida 32306
ATTN: Prof. I. R. Savage
- 1 University of Florida
Department of Statistics
Gainesville, Florida 32603
ATTN: R. L. Scheaffer
- 1 Logistics Research Project
The George Washington University
707 22nd Street, N.W.
Washington, D. C. 20037
ATTN: Dr. W. H. Marlow
- 1 Harvard University
Department of Economics
Cambridge, Massachusetts 02138
ATTN: Prof. K. J. Arrow
- 1 Harvard University
Department of Statistics
Cambridge, Massachusetts 02139
ATTN: Prof. W. G. Cochran
- 1 Institute for Defense Analyses
Communications Research Division
von Neumann Hall
Princeton, New Jersey 08541
- 1 International Business Machines
Corporation
P. O. Box 218, Lamb Estate
Yorktown Heights, New York 10598
ATTN: Dr. Alan Hoffman
- 1 Israel Institute of Technology
Technion
Haifa, ISRAEL
ATTN: Prof. P. Naor

- 1 The Johns Hopkins University
Department of Statistics
34th & Charles Streets
Baltimore, Maryland 21218
ATTN: Prof. G. S. Watson
- 1 L.T.V. Electrosystems
Greenville, Texas 75401
ATTN: Dr. Oran Brigham, Jr.
- 1 L.T.V. -- Dept. #4-553400
Guidance & Control Development
P. O. Box 6118
Dallas, Texas 75222
ATTN: Dr. P. R. Hendricks
- 1 L.T.V. Research Center
P. O. Box 6144
Dallas, Texas 75222
ATTN: Dr. F. W. Fenter
- 1 The University of Michigan
Department of Mathematics, W.E.
Ann Arbor, Michigan 48104
ATTN: Prof. R. M. Thrall
- 1 Maritime Transportation Research
Board
National Academy of Sciences
2101 Constitution Avenue
Washington, D. C. 20418
ATTN: RADM. J. B. Oren USCG (RET)
- 1 Naval Applied Science Laboratory
Technical Library, Code 222
Flushing & Washington Avenues
Brooklyn, New York 11251
- 1 Library
Naval Electronics Laboratory
Center
San Diego, California 92152
- 1 Naval Ordnance Station
Louisville, Kentucky 40214
- 1 Technical Library
Naval Ordnance Station
Indian Head, Maryland 20640
- 2 Bureau of Naval Personnel
Navy Department
Technical Library
Washington, D. C. 20370
STOP 82
- 1 Library, Code 0212
Naval Postgraduate School
Monterey, California 93940
- 6 Technical Information Center
Naval Research Laboratory
Washington, D. C. 20390
- 1 Naval Research Laboratory - Code 5620
Washington, D. C. 20390
ATTN: Robert D. Misner
- 1 Naval Ship Engineering Center,
Philadelphia
Division Technical Library
Philadelphia, Pennsylvania 19112
- 1 Applied Mathematics Laboratory
Naval Ships Research Development
Center
Washington, D. C. 20007
- 1 Naval Undersea Warfare Center
3202 E. Foothill Boulevard
Pasadena, California 91107
Technical Library
- 1 Department of the Navy
Naval Ordnance Systems Command (MTEA)
Naval Plant Representative
General Dynamics Pomona Division
P. O. Box 2507
1675 West Fifth Avenue
Pomona, California 91766
- 1 University of New Mexico
Department of Statistics
Albuquerque, New Mexico 87106
ATTN: Prof. Julius R. Blum
- 1 New York University
Institute of Mathematical Sciences
New York, New York 10453
ATTN: Prof. W. M. Hirsch

- 1 New York University
Department of Industrial Engineering
and Operations Research
Bronx, New York 10453
ATTN: Prof. J. H. K. Kao
- 1 State University of New York
Department of Statistics
4244 Ridge Lea Road
Amherst, New York 14226
ATTN: R. L. Schmidt
- 1 North Carolina State University
Department of Experimental Statistics
Raleigh, North Carolina 27607
ATTN: Prof. C. P. Quesenberry
- 1 University of North Carolina
Department of Statistics
Chapel Hill, North Carolina 27514
ATTN: Prof. W. L. Smith
- 1 University of North Carolina
Department of Statistics
Chapel Hill, North Carolina 27514
ATTN: Prof. M. R. Leadbetter
- 1 Oak Ridge National Laboratory
Statistics Group
Mathematics Division
P. O. Box Y
Oak Ridge, Tennessee 37831
ATTN: Dr. Donald A. Gardiner
- 1 Director
Office of Naval Research,
Branch Office
1030 East Green Street
Pasadena, California 91101
ATTN: Dr. A. R. Laufer
- 1 Office of Naval Research
San Francisco Area Office
1076 Mission Street
San Francisco, California 94103
- 3 Logistics and Mathematical
Statistics Program
Office of Naval Research
Washington, D. C. 20360
- 1 Director
Office of Naval Research,
Branch Office
219 South Dearborn Street
Chicago, Illinois 60604
ATTN: Dr. A. R. Dawe
- 1 Director
Office of Naval Research,
Branch Office
495 Summer Street
Boston, Massachusetts 02210
ATTN: Dr. A. L. Powell
- 1 Office of Naval Research
New York Area Office
207 West 24th Street
New York, New York 10011
ATTN: Dr. J. Laderman
- 2 Commanding Officer
Office of Naval Research,
Branch Office
BOX 39, FPO
New York, New York 09510
- 1 Office of Naval Research,
Resident Representative
University of Texas
Rooms 2507-08 Main Building
P. O. Box 7938
Austin, Texas 78712
ATTN: Mr. F. M. Lucas
- 1 Systems Analysis Division
Room BC760, Pentagon
Washington, D. C. 20350
ATTN: Mr. A. S. Rhodes, Op-964
- 1 Princeton University
Department of Economics and
Sociology
Princeton, New Jersey 08540
ATTN: Prof. O. Morgenstern
- 1 Princeton University
Department of Mathematics
Princeton, New Jersey 08540
ATTN: Prof. A. W. Tucker

- | | |
|---|--|
| 1 Princeton University
Department of Mathematics
Princeton, New Jersey 08540
ATTN: Prof. J. W. Tukey | 1 Stanford University
Department of Operations Research
Stanford, California 94305
ATTN: Prof. G. B. Dantzig |
| 1 Purdue University
Division of Mathematical Sciences
Lafayette, Indiana 47907
ATTN: Prof. S. S. Gupta | 1 Stanford University
Department of Operations Research
Stanford, California 94305
ATTN: Prof. F. S. Hillier |
| 1 Purdue University
Division of Mathematical Sciences
Lafayette, Indiana 47907
ATTN: Prof. H. Rubin | 1 Stanford University
Department of Operations Research
Stanford, California 94305
ATTN: Dr. D. L. Iglehart |
| 1 Purdue University
Graduate School of Industrial
Administration
Lafayette, Indiana 47907
ATTN: Prof. Andrew Whinston | 1 Stanford University
Department of Mathematics
Stanford, California 94305
ATTN: Prof. S. Karlin |
| 1 Rutgers - The State University
Statistics Center
New Brunswick, New Jersey 08903
ATTN: Prof. H. F. Dodge | 1 Stanford University
Department of Statistics
Stanford, California 94305
ATTN: Prof. G. J. Lieberman |
| 1 University of Sheffield
Department of Probability
and Statistics
Sheffield 10, ENGLAND
ATTN: Prof. J. Gani | 1 Stanford University
Department of Statistics
Stanford, California 94305
ATTN: Prof. E. Parzen |
| 1 Southern Methodist University
Department of Statistics
Dallas, Texas 75222
ATTN: Prof. D. B. Owen | 1 Stanford University
Department of Statistics
Stanford, California 94305
ATTN: Prof. H. Solomon |
| 1 Stanford University
Department of Statistics
Stanford, California 94305
ATTN: Prof. T. W. Anderson | 1 Stanford University
Department of Operations Research
Stanford, California 94305
ATTN: Prof. A. F. Veinott, Jr. |
| 1 Stanford University
Department of Statistics
Stanford, California 94305
ATTN: Prof. H. Chernoff | 1 Texas A&M Foundation
Institute of Statistics
College Station, Texas 77843
ATTN: Prof. H. O. Hartley |

- 1 Texas Instruments, Inc.
Central Research Laboratory
M.S. 132 - P. O. Box 5936
Dallas, Texas 75222
ATTN: Dr. Bernard H. List
- 1 Texas Instruments, Inc.
North Building Library
M.S. 211 - P. O. Box 6015
Dallas, Texas 75222
- 1 University of Texas
Department of Mathematics
Austin, Texas 78712
ATTN: Dr. A. Charnes
- 1 Tracor, Inc.
1701 Guadalupe
Austin, Texas 78701
ATTN: Dr. W. C. Moyer
- 1 Unitech, Inc.
2209 Manor Road
Austin, Texas 78722
ATTN: Mr. Hugh Batey
- 1 U. S. Naval Postgraduate School
Department of Operations Research
and Economics
Monterey, California 93940
ATTN: Prof. C. R. Jones
- 1 Commander
U. S. Naval Weapons Laboratory (MAL)
Dahlgren, Virginia 22448
- 1 University of Washington
Department of Mathematics
Seattle, Washington 98105
ATTN: Prof. Z. W. Birnbaum
- 1 University of Wisconsin
Department of Statistics
Madison, Wisconsin 53706
ATTN: Prof. G. E. P. Box
- 1 Yale University
Department of Statistics
New Haven, Connecticut 06520
ATTN: Prof. F. J. Anscombe
- 1 Yale University
Department of Statistics
New Haven, Connecticut 06520
ATTN: Prof. L. J. Savage
- 1 Cowles Commission for Research
in Economics
Yale University
New Haven, Connecticut 06520
ATTN: Prof. Martin Shubik