5V

(12) LEVEL II

AD

ADA102540

TECHNICAL REPORT ARBRL-TR-02331

QUANTILES OF MULTI-SAMPLE SMIRNOV TYPE STATISTICS

Mark P. Becker
Malcolm S. Taylor

June 1981

DTIC FILE COPY

DTIC ELECTE AUG 7 1981 S D A

US ARMY ARMAMENT RESEARCH AND DEVELOPMENT COMMAND
BALLISTIC RESEARCH LABORATORY
ABERDEEN PROVING GROUND, MARYLAND

81 8 06 045

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: TECHNICAL REPORT ARBRL-TR-02331	2. GOVT ACCESSION NO.: AD-A102540	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): QUANTILES OF MULTI-SAMPLE SMIRNOV TYPE STATISTICS		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Mark P. Becker, Malcolm S. Taylor		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: US Army Ballistic Research Laboratory, ATTN: DRDAR-BLB, Aberdeen Proving Ground, MD 21005		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: RDT&E 1L161102AH43
11. CONTROLLING OFFICE NAME AND ADDRESS: USAARRADCOM Ballistic Research Laboratory, ATTN: DRDAR-BL, Aberdeen Proving Ground, MD 21005		12. REPORT DATE: JUNE 1981
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS (of this report): UNCLASSIFIED
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Birnbaum-Hall Test; Smirnov Type Statistics; Nonparametric Statistics; Hypothesis Testing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Let $X_1^{(i)}, X_2^{(i)}, \ldots, X_{n_i}^{(i)}$, $i = 1,2, \ldots, c$, be samples of c independent random variables $X^{(i)}$ with continuous cdf $F^{(i)}$. To test the hypothesis H_0: $F^{(i)} = F^{(j)}$, $i,j = 1,2, \ldots, c$, Birnbaum and Hall introduced the statistics $D(n_1,n_2, \ldots, n_c)$ and $D^+(n_1,n_2, \ldots, n_c)$ and determined their exact small sample distributions for $c = 2,3$ and $n_i = n_j$. This paper considers exact small

DD FORM 1473, 1 JAN 73 — EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

393471

Item 20. Abstract, Cont'd.

sample distributions for more than three samples ($c > 3$) and unequal sample sizes ($n_i \neq n_j$) and includes representative tables of quantiles under this extension.

TABLE OF CONTENTS

I. INTRODUCTION AND SUMMARY

Let $X_1^{(i)}, X_2^{(i)}, \ldots, X_{n_i}^{(i)}$, $i = 1, 2, \ldots, c$ be samples of c independent random variables $X^{(i)}$ with continuous cumulative distribution functions $F^{(i)}$ and empirical distributions $F^{*(i)}$

$$
\begin{aligned}
F^{*(i)}(x) &= 0 & & x < X_1^{(i)} \\
&= k/n_i & & X_k^{(i)} \leq x < X_{k+1}^{(i)}, \quad 1 \leq k < n_i \\
&= 1 & & X_{n_i}^{(i)} \leq x.
\end{aligned}
$$

We assume, without loss of generality, that $X_1^{(i)} < X_2^{(i)} < \ldots < X_{n_i}^{(i)}$.

To test the hypothesis that the population distribution functions are identical, Birnbaum and Hall[1] introduced the statistics

$$D(n_1, n_2, \ldots, n_c) = \sup_{x,i,j} |F^{*(i)}(x) - F^{*(j)}(x)|, \tag{1}$$

$$D^+(n_1, n_2, \ldots, n_c) = \sup_{x,i<j} [F^{*(i)}(x) - F^{*(j)}(x)] \tag{2}$$

$i,j = 1,2, \ldots, c$ which are generalizations of the well known Kolmogorov-Smirnov two sample statistics $D(m, n)$ and $D^+(m, n)$.

Under the null hypothesis

$$H_o: F^{(i)} = F^{(j)} \quad i,j = 1,2, \ldots, c,$$

the exact small sample distributions of the statistics (1) and (2) were determined by Birnbaum-Hall using difference equations that could be solved recursively and tabled values of

$$P[D(n, n, n) \leq r], \quad P[D(n,n) \leq r], \quad P[D^+(n,n) \leq r]$$

were developed for selected values of n between 1 and 40 and of $r = k/n$, $k = 1,2, \ldots, n$.

---

[1] Z. W. Birnbaum and R. A. Hall, "Small Sample Distributions for Multi-Sample Statistics of the Smirnov Type," Ann. Math. Stat., Vol. 31 (1960), pp. 710-720.

The limitations on the tables were for the most part imposed by the speed and capacity of the computer necessary to solve the difference equations. This paper utilizes some notions from graph theory, and in so doing, provides exact small sample distributions for more than three samples and for unequal sample sizes. Representative tables are included.

II. A GEOMETRICAL PERSPECTIVE

For a geometrical perspective of the problem we parallel in the next two paragraphs the development of Birnbaum-Hall.

The c-dimensional random variable $[F^{*(1)}(x), F^{*(2)}(x), \dots, F^{*(c)}(x)]$, for fixed x, takes on values $(k_1/n_1, k_2/n_2, \dots, k_c/n_c)$, $k_i \in \{0, 1, \dots, n_i\}$; $i = 1, 2, \dots, c$, which lie in the unit hypercube. The transformation $(x_1, x_2, \dots, x_c)$ into $(y_1, y_2, \dots, y_c)$ defined by $y_i = n_i x_i$ maps the hypercube into a c-dimensional hypervolume with sides $(n_1, n_2, \dots, n_c)$ and the random c-vector into points $(k_1, k_2, \dots, k_c)$ with integer valued coordinates.

Under the null hypothesis H_0: $F^{(i)} = F^{(j)}$ $i,j = 1,2, \dots, c$ we may consider the c samples as coming from the same population with N! equally likely ways of drawing the combined sample of size $N = n_1 + n_2 + \dots + n_c$. Now suppose the values of the combined ordered sample are located on the real line, and the null c-vector is used as a counter in the following manner: Starting at a value less than min $\{X_1^{(1)}, X_1^{(2)}, \dots, X_1^{(c)}\}$ with the counter equal $(0, 0, \dots, 0)$ and moving in the direction of increasing magnitude, the k*th* coordinate of the counter is incremented by one whenever a sample value $X_i^{(k)}$ is encountered. In so doing, a 1-1 correspondence between a path in c-space from $(0,0, \dots, 0)$ to $(n_1, n_2, \dots, n_c)$ and the c samples is defined.

The total number of such equally likely paths is the multinomial coefficient $N!/(n_1!n_2! \dots n_c!)$. The critical region for the test will consist of a set of points $R = \{(k_1, k_2, \dots, k_c)\}$ such that any path from $(0,0, \dots, 0)$ to $(n_1, n_2, \dots, n_c)$ passing through R will cause H_0 to be rejected. If we denote the number of paths passing through R as $Q(n_1, n_2, \dots, n_c; R)$ then the probability of an error of the first kind will be

$$\alpha = \frac{Q(n_1,n_2, \ldots ,n_c;R)}{N!/(n_1!n_2! \ldots n_c!)} .$$

III. DETERMINATION OF EXACT SMALL SAMPLE DISTRIBUTION

Determination of the exact small sample distribution of (1) and (2) from this approach involves an enumerative process to determine $Q(n_1,n_2, \ldots ,n_c;R)$. The set R defining the critical region suggests itself as $R = \{(k_1,k_2, \ldots ,k_c) \mid \sup_{i,j} | n_jk_i - n_ik_j| > n_in_jr\}$ and

$R^+ = \{(k_1,k_2, \ldots ,k_c) \mid \sup_{i<j} (n_jk_i - n_ik_j) > n_in_jr\}$ corresponding to

$D(n_1,n_2, \ldots ,n_c) > r$ and $D^+(n_1,n_2, \ldots ,n_c) > r$ respectively. These regions of rejection are again extensions of the Kolmogorov-Smirnov procedure; however, there is no theoretical basis for their limitation to this form.

Considering the paths in c-space from $(0,0, \ldots ,0)$ to $(n_1,n_2, \ldots ,n_c)$ as simple paths in a finite directed graph, (see e.g. Berge[2]) we have at our disposal a method of enumeration appearing in a paper by Pototchi[3] which seems well suited to computer manipulation. This method, while conceptually analogous to the difference equation evaluated by Birnbaum-Hall, allows us to accomplish the enumeration necessary to evaluate $Q(n_1,n_2, \ldots ,n_c;R)$ and to consider the number of samples $c > 3$ as well as unequal sample sizes in an efficient and economical manner.

IV. TABLES

Table 1 describes the upper tail of the distribution of D, and contains the probabilities $P[D(n_1,n_2,n_3) \leq r]$ for $5 \leq n_1,n_2,n_3 \leq 10$ starting at the largest increment of $r = .1(.1)1.0$ for which the tail probability equals or exceeds .20.

---

[2]C. Berge, The Theory of Graphs and its Applications, John Wiley & Sons, New York, 1962.

[3]A. Pototchi, "A Simple Algorithm for Determining the Number of Paths in a Finite Graph," Economic Cyber Studies Res., Vol. 2 (1967), pp. 81-85.

Table 2 contains the probabilities $P[D(n,n,n,n) \leq r]$ for $n = 2(1)10$ corresponding to values of $r = k/n$ consistent with the differences $|F^{*(i)}(x) - F^{*(j)}(x)|$ which can occur.

The values in Table 1 for the triples (n,n,n), $n = 5(1)10$ correspond to those computed by Birnbaum-Hall. The values in Table 2 for the 4-tuples (n,n,n,n), $n = 5,10$ are consistent with those Monte Carloed by Gardner,[4] et al. Taken together, these provide additional verification of the accuracy of the tables.

---

[4] R. H. Gardner, J. E. Pinder III, and R. S. Wood, "Monte Carlo Estimation of Percentiles for the Multi-Smirnov Test," J. Statist. Comput. Simul., Vol. 10 (1980), pp. 243-249.

TABLE 1. QUANTILES OF THE BIRNBAUM-HALL, OR THREE SAMPLE SMIRNOV, STATISTIC FOR UNEQUAL SAMPLE SIZES.

n_1	n_2	n_3	r	$P[D(n_1,n_2,n_3) \leq r]$	n_1	n_2	n_3	r	$P[D(n_1,n_2,n_3) \leq r]$
5	5	5	.6	.81109	5	6	6	.6	.64253
			.7	.81109				.7	.89712
			.8	.97819				.8	.93216
			.9	.97819				.9	.98982
			1.0	1.00000				1.0	1.00000
5	5	6	.6	.67353	5	6	7	.6	.70972
			.7	.85609				.7	.87336
			.8	.94628				.8	.95008
			.9	.98457				.9	.99238
			1.0	1.00000				1.0	1.00000
5	5	7	.6	.75909	5	6	8	.6	.70123
			.7	.83058				.7	.88876
			.8	.96542				.8	.95968
			.9	.98776				.9	.99372
			1.0	1.00000				1.0	1.00000
5	5	8	.6	.74472	5	6	9	.6	.72578
			.7	.86485				.7	.91119
			.8	.97582				.8	.96508
			.9	.98947				.9	.99446
			1.0	1.00000				1.0	1.00000
5	5	9	.5	.46916	5	6	10	.6	.75753
			.6	.80259				.7	.92490
			.7	.88479				.8	.96825
			.8	.98176				.9	.99490
			.9	.99043				1.0	1.00000
			1.0	1.00000					
5	5	10	.5	.52097	5	7	7	.6	.77668
			.6	.83928				.7	.84966
			.7	.89685				.8	.96547
			.8	.98531				.9	.99462
			.9	.99100				1.0	1.00000
			1.0	1.00000					
					5	7	8	.6	.74794
								.7	.88248
								.8	.97350
								.9	.99578
								1.0	1.00000

TABLE 1. (CONT'D)

n_1	n_2	n_3	r	$P[D(n_1,n_2,n_3) \leq r]$
5	7	9	.6	.78143
			.7	.90138
			.8	.97793
			.9	.99642
			1.0	1.00000
5	7	10	.5	.54453
			.6	.81483
			.7	.91269
			.8	.98047
			.9	.99680
			1.0	1.00000
5	8	8	.6	.74256
			.7	.91251
			.8	.98065
			.9	.99685
			1.0	1.00000
5	8	9	.6	.78949
			.7	.92949
			.8	.98457
			.9	.99745
			1.0	1.00000
5	8	10	.5	.62530
			.6	.81808
			.7	.93951
			.8	.98681
			.9	.99780
			1.0	1.00000
5	9	9	.5	.59437
			.6	.83481
			.7	.94532
			.8	.98821
			.9	.99803
			1.0	1.00000
5	9	10	.5	.65118
			.6	.86208
			.7	.95464
			.8	.99031
			.9	.99836
			1.0	1.00000

n_1	n_2	n_3	r	$P[D(n_1,n_2,n_3) \leq r]$
5	10	10	.5	.70953
			.6	.88851
			.7	.96359
			.8	.99235
			.9	.99870
			1.0	1.00000
6	6	6	.6	.68408
			.7	.93216
			.8	.93216
			.9	.99383
			1.0	1.00000
6	6	7	.6	.74897
			.7	.91138
			.8	.94990
			.9	.99569
			1.0	1.00000
6	6	8	.6	.74664
			.7	.90488
			.8	.95945
			.9	.99662
			1.0	1.00000
6	6	9	.6	.74555
			.7	.92900
			.8	.96485
			.9	.99712
			1.0	1.00000
6	6	10	.6	.78119
			.7	.94384
			.8	.96806
			.9	.99739
			1.0	1.00000
6	7	7	.5	.49330
			.6	.81149
			.7	.88975
			.8	.96517
			.9	.99719
			1.0	1.00000

TABLE 1. (CONT'D)

n_1	n_2	n_3	r	$P[D(n_1,n_2,n_3) \leq r]$
6	7	8	.6	.78637
			.7	.90257
			.8	.97319
			.9	.99792
			1.0	1.00000
6	7	9	.6	.79154
			.7	.92361
			.8	.97765
			.9	.99830
			1.0	1.00000
6	7	10	.5	.57700
			.6	.82881
			.7	.93638
			.8	.98023
			.9	.99851
			1.0	1.00000
6	8	8	.6	.78567
			.7	.91100
			.8	.98035
			.9	.99855
			1.0	1.00000
6	8	9	.5	.61580
			.6	.80505
			.7	.92848
			.8	.98430
			.9	.99888
			1.0	1.00000
6	8	10	.5	.65020
			.6	.83826
			.7	.93879
			.8	.98659
			.9	.99906
			1.0	1.00000
6	9	9	.5	.65335
			.6	.81993
			.7	.94488
			.8	.98798
			.9	.99918
			1.0	1.00000

n_1	n_2	n_3	r	$P[D(n_1,n_2,n_3) \leq r]$
6	9	10	.5	.67974
			.6	.84898
			.7	.95453
			.8	.99011
			.9	.99934
			1.0	1.00000
6	10	10	.5	.70263
			.6	.87756
			.7	.96382
			.8	.99217
			.9	.99950
			1.0	1.00000
7	7	7	.5	.56209
			.6	.86823
			.7	.86823
			.8	.97750
			.9	.99830
			1.0	1.00000
7	7	8	.5	.63985
			.6	.81671
			.7	.89916
			.8	.98371
			.9	.99883
			1.0	1.00000
7	7	9	.5	.61004
			.6	.82936
			.7	.91681
			.8	.98702
			.9	.99909
			1.0	1.00000
7	7	10	.5	.62594
			.6	.86663
			.7	.92731
			.8	.98888
			.9	.99922
			1.0	1.00000
7	8	8	.6	.79392
			.7	.92685
			.8	.98888
			.9	.99924
			1.0	1.00000

TABLE 1. (CONT'D)

n_1	n_2	n_3	r	$P[D(n_1,n_2,n_3) \leq r]$
7	8	9	.5	.64705
			.6	.82084
			.7	.94230
			.8	.99158
			.9	.99944
			1.0	1.00000
7	8	10	.5	.68971
			.6	.85147
			.7	.95131
			.8	.99306
			.9	.99955
			1.0	1.00000
7	9	9	.5	.66124
			.6	.84333
			.7	.95640
			.8	.99394
			.9	.99962
			1.0	1.00000
7	9	10	.5	.69235
			.6	.87042
			.7	.96455
			.8	.99523
			.9	.99970
			1.0	1.00000
7	10	10	.5	.72037
			.6	.89658
			.7	.97223
			.8	.99641
			.9	.99979
			1.0	1.00000
8	8	8	.6	.79392
			.7	.95029
			.8	.99292
			.9	.99954
			1.0	1.00000
8	8	9	.5	.67466
			.6	.83749
			.7	.96289
			.8	.99494
			.9	.99968
			1.0	1.00000
8	8	10	.5	.74735
			.6	.86362
			.7	.96997
			.8	.99601
			.9	.99976
			1.0	1.00000
8	9	9	.5	.65584
			.6	.87789
			.7	.97370
			.8	.99659
			.9	.99980
			1.0	1.00000
8	9	10	.5	.71259
			.6	.90160
			.7	.97966
			.8	.99745
			.9	.99985
			1.0	1.00000
8	10	10	.5	.76924
			.6	.92368
			.7	.98495
			.8	.99819
			.9	.99990
			1.0	1.00000
9	9	9	.5	.71542
			.6	.91350
			.7	.98248
			.8	.99785
			.9	.99988
			1.0	1.00000
9	9	10	.5	.76978
			.6	.93369
			.7	.98714
			.8	.99848
			.9	.99992
			1.0	1.00000
9	10	10	.4	.53864
			.5	.82180
			.6	.95150
			.7	.99104
			.8	.99898
			.9	.99995
			1.0	1.00000

TABLE 1. (CONT'D)

n_1	n_2	n_3	r	$P[D(n_1,n_2,n_3) \leq r]$
10	10	10	.4	.63715
			.5	.86930
			.6	.96645
			.7	.99411
			.8	.99936
			.9	.99997
			1.0	1.00000

TABLE 2. QUANTILES OF A FOUR SAMPLE SMIRNOV STATISTIC FOR EQUAL SAMPLE SIZES.

n_1	n_2	n_3	n_4	r	$P[D(n_1,n_2,n_3,n_4) \leq r]$
2	2	2	2	.5	.22857
				1.0	1.00000
3	3	3	3	.333	.03740
				.666	.64286
				1.000	1.00000
4	4	4	4	.25	.00526
				.50	.35707
				.75	.87219
				1.00	1.00000
5	5	5	5	.2	.00068
				.4	.18184
				.6	.69490
				.8	.95982
				1.0	1.00000
6	6	6	6	.500	.52300
				.666	.88076
				.833	.98824
				1.000	1.00000
7	7	7	7	.571	.77875
				.714	.95842
				.857	.99670
				1.000	1.00000

n_1	n_2	n_3	n_4	r	$P[D(n_1,n_2,n_3,n_4) \leq r]$
8	8	8	8	.500	.66901
				.625	.91095
				.750	.98651
				.875	.99910
				1.000	1.00000
9	9	9	9	.444	.56199
				.555	.84982
				.666	.96732
				.777	.99574
				.888	.99976
				1.000	1.00000
10	10	10	10	.5	.78000
				.6	.93879
				.7	.98874
				.8	.99875
				.9	.99994
				1.0	1.00000

DISTRIBUTION LIST

No. of Copies	Organization
12	Commander Defense Technical Info Center ATTN: DDC-DDA Cameron Station Alexandria, VA 22314
1	Commander US Army Materiel Development and Readiness Command ATTN: DRCDMD-ST 5001 Eisenhower Avenue Alexandria, VA 22333
2	Commander US Army Armament Research and Development Command ATTN: DRDAR-TSS (2 cys) Dover, NJ 07801
1	Commander US Army Armament Materiel Readiness Command ATTN: DRSAR-LEP-L, Tech Lib Rock Island, IL 61299
1	Director US Army ARRADCOM Benet Weapons Laboratory ATTN: DRDAR-LCB-TL Watervliet, NY 12189
1	Commander US Army Aviation Research and Development Command ATTN: DRSAV-E P.O. Box 209 St. Louis, MO 61366
1	Director US Army Air Mobility Research and Development Laboratory Ames Research Center Moffett Field, CA 94035
1	Commander US Army Communications Rsch and Development Command ATTN: DRDCO-PPA-SA Fort Monmouth, NJ 07703

No. of Copies	Organization
1	Commander US Army Electronics Research and Development Command Technical Support Activity ATTN: DELSD-L Fort Monmouth, NJ 07703
1	Commander US Army Missile Command ATTN: DRSMI-R Redstone Arsenal, AL 35809
1	Commander US Army Missile Command ATTN: DRSMI-YDL Redstone Arsenal, AL 35809
1	Commander US Army Tank Automotive Rsch and Development Command ATTN: DRDTA-UL Warren, MI 48090
1	Director US Army TRADOC Systems Analysis Activity ATTN: ATAA-SL, Tech Lib White Sands Missile Range NM 88002
1	Commander US Army Research Office ATTN: DRXRO-MA, Dr. Robert Launer P.O. Box 12211 Research Triangle Park, NC 27709
1	Commander US Army Combat Development Experimentation Command ATTN: ATEC-SA, Dr. Marion R. Bryson Fort Ord, CA 93941
1	Commander US Army Operational Test and Evaluation Agency ATTN: CSTE-STS, Mr. Fred C. McCoy 5600 Columbia Pike Falls Church, VA 22041

DISTRIBUTION LIST

No. of Copies	Organization
	Aberdeen Proving Ground
	Dir, USAMSAA ATTN: DRXSY-D DRXSY-MP, H. Cohen
	Cdr, USATECOM ATTN: DRSTE-TO-F DRSTE-AD, Dr. Ted Trybul
	Dir, USACSL, Bldg. E3516, EA ATTN: DRDAR-CLB-PA

USER EVALUATION OF REPORT

Please take a few minutes to answer the questions below; tear out this sheet, fold as indicated, staple or tape closed, and place in the mail. Your comments will provide us with information for improving future reports.

1. BRL Report Number ______________________________

2. Does this report satisfy a need? (Comment on purpose, related project, or other area of interest for which report will be used.)

____

____

____

3. How, specifically, is the report being used? (Information source, design data or procedure, management procedure, source of ideas, etc.) ______________________________

____

____

4. Has the information in this report led to any quantitative savings as far as man-hours/contract dollars saved, operating costs avoided, efficiencies achieved, etc.? If so, please elaborate.

____

____

5. General Comments (Indicate what you think should be changed to make this report and future reports of this type more responsive to your needs, more usable, improve readability, etc.) ______________________________

____

____

____

6. If you would like to be contacted by the personnel who prepared this report to raise specific questions or discuss the topic, please fill in the following information.

Name: ______________________________

Telephone Number: ______________________________

Organization Address: ______________________________

____

____

FOLD HERE

Director
US Army Ballistic Research Laboratory
Aberdeen Proving Ground, MD 21005

NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

BUSINESS REPLY MAIL
FIRST CLASS PERMIT NO 12062 WASHINGTON, DC
POSTAGE WILL BE PAID BY DEPARTMENT OF THE ARMY

Director
US Army Ballistic Research Laboratory
ATTN: DRDAR-TSB
Aberdeen Proving Ground, MD 21005

FOLD HERE