Calhoun: The NPS Institutional Archive

DSpace Repository

Theses and Dissertations | 1. Thesis and Dissertation Collection, all items

1976-03

An Investigation of the Applicability of the Product Limit Estimate to the Statistical Analysis of Sonar Detection Distributions

Route, Ronald Arthur

Monterey, California. Naval Postgraduate School

https://hdl.handle.net/10945/38682

Downloaded from NPS Archive: Calhoun

Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

http://www.nps.edu/library

1976.03
_Route
c. 1

AN INVESTIGATION OF THE APPLICABILITY OF THE PRODUCT LIMIT ESTIMATE TO THE STATISTICAL ANALYSIS OF SONAR DETECTION DISTRIBUTIONS

Ronald Arthur Route

DUDLEY KNOX LIBRARY
RESTRICTED RESOURCES & SERVICES
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 94943-5101

DUDLEY KNOX LIBRARY,
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIF. 93940

NAVAL
POSTGRADUATE
SCHOOL

DUDLEY KNOX LIBRARY.

January 29, 2014

SUBJECT: Change in distribution statement for *An Investigation of the Applicability of the Product Limit Estimate to the Statistical Analysis of Sonar Detection Distributions* – March 1976.

1. Reference: Route, Ronald Arthur. *An Investigation of the Applicability of the Product Limit Estimate to the Statistical Analysis of Sonar Detection Distributions.* Monterey, CA: Naval Postgraduate School, March 1976. UNCLASSIFIED, Distribution limited to U. S. Gov't. agencies only; Test and Evaluation; Mar 76. Other requests for this document must be referred to Superintendent, Naval Postgraduate School, Attn: Code 012. Monterey, Calif. 93940.

2. Upon consultation with NPS faculty, the School has determined that this thesis may be released to the public effective January 13, 2014.

University Librarian
Naval Postgraduate School

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

AN INVESTIGATION OF THE APPLICABILITY OF THE PRODUCT LIMIT ESTIMATE TO THE STATISTICAL ANALYSIS OF SONAR DETECTION DISTRIBUTIONS

by

Ronald Arthur Route

March 1976

Thesis Advisor: D.P.Gaver

~~Distribution Limited to U. S. Government Agencies Only; Test and Evaluation, March 1976. Other requests for this document must be referred to the Naval Postgraduate School, Monterey, California 93940, Code 023.~~

DOWNGRADED
APPROVED FOR PUBLIC RELEASE

T175014

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
An Investigation of the Applicability of the Product Limit Estimate to the Statistical Analysis of Sonar Detection Distributions

5. TYPE OF REPORT & PERIOD COVERED
Master's Thesis; March 1976

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Ronald Arthur Route

8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Naval Postgraduate School
Monterey, California 93940

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
Naval Postgraduate School
Monterey, California 93940

12. REPORT DATE
March 1976

13. NUMBER OF PAGES
58

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)
~~Distribution Limited to U. S. Government Agencies Only; Test and Evaluation, March 1976. Other requests for this document must be referred to the Naval Postgraduate School, Monterey, California 93940, Code 023.~~

DOWNGRADED
APPROVED FOR PUBLIC RELEASE

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)
Product Limit Estimate
PL Estimate
Sonar Detection Distributions

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

The application of the product limit estimate (PL estimate) to the analysis of sonar detection range data was examined using an empirical approach. A sonar detection distribution was assumed to be known and target lateral range distributions corresponding to likely operational scenarios were derived. These distributions were used in a computer simulation to generate sample fleet exercise reconstruction data. For data

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
(Page 1) S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(20. ABSTRACT Continued)

of small sample sizes, the PL estimates of the sonar detection distribution were compared with the known distribution and were found to be reasonably accurate approximations for all tactical situations tested. The PL estimates clearly outperformed the more readily computable, yet always optimistic, reduced sample (or empirical survival function) estimates. Finally, confidence intervals about the estimate were computed by using the jackknife procedure and the apparent optimism of the resulting confidence intervals was noted. Confidence interval baseline data is provided for future comparison with results from the estimates generated from small sample sizes. Further research is indicated to improve the usefulness of the jackknife to this application.

An Investigation of the Applicability of the Product Limit Estimate to the Statistical Analysis of Sonar Detection Distributions

by

Ronald Arthur Route
Lieutenant, United States Navy
B.S., United States Naval Academy, 1971

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
March 1976

Author: Ronald A. Route

Approved by: Donald P. Gaver
Thesis Advisor

Second Reader

David A. Schrady
Chairman, Department of Operations Research and Administrative Sciences

Academic Dean

RRd
R814
c.1

1976.03 Route

DUDLEY KNOX LIBRARY,
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIF. 93940

ABSTRACT

The application of the product limit estimate (PL estimate) to the analysis of sonar detection range data was examined using an empirical approach. A sonar detection distribution was assumed to be known and target lateral range distributions corresponding to likely operational scenarios were derived. These distributions were used in a computer simulation to generate sample fleet exercise reconstruction data. For data of small sample sizes, the PL estimates of the sonar detection distribution were compared with the known distribution and were found to be reasonably accurate approximations for all tactical situations tested. The PL estimates clearly outperformed the more readily computable, yet always optimistic, reduced sample (or empirical survival function) estimates. Finally, confidence intervals about the estimate were computed by using the jackknife procedure and the apparent optimism of the resulting confidence intervals was noted. Confidence interval baseline data is provided for future comparison with results from estimates generated from small sample sizes. Further research is indicated to improve the usefulness of the jackknife to this application.

TABLE OF CONTENTS

I. INTRODUCTION ---------- 8

II. THE PRODUCT LIMIT ESTIMATE ---------- 10

A. THE KAPLAN-MEIER ESTIMATE ---------- 10

B. APPLICATION TO DETECTION RANGE DISTRIBUTIONS - 10

C. DEFINING THE APPLIED PL ESTIMATE ---------- 11

III. TESTING THE PRODUCT LIMIT ESTIMATE ---------- 16

A. PROBLEM ---------- 16

B. PROCEDURE ---------- 16

1. The Assumed Distribution ---------- 17

2. The Lateral Range Distributions ---------- 17

a. Random Search ---------- 19

b. Barrier Search ---------- 19

c. Submarine Stationkeeping ---------- 20

C. SIMULATION AND DATA ---------- 21

D. CONCLUSIONS ---------- 28

IV. CONFIDENCE INTERVALS ---------- 32

A. PROBLEM ---------- 32

B. PROCEDURE ---------- 32

1. Lateral Range Distribution ---------- 33

2. Parameters ---------- 33

3. The Jackknife ---------- 34

4. Intraclass Correlation ---------- 34

C. SIMULATION ---------- 36

D. DISCUSSION AND CONCLUSIONS ---------- 40

APPENDIX A: COMPUTER SIMULATION FOR TESTING EFFECTS OF VARYING TACTICAL SITUATION ---- 42

APPENDIX B: COMPUTER SIMULATION WITH JACK-KNIFED CONFIDENCE INTERVALS ---------- 49

LIST OF REFERENCES ---------- 56

INITIAL DISTRIBUTION LIST ---------- 58

ACKNOWLEDGMENT

The author is grateful for the guidance and assistance given by Dr. Donald P. Gaver throughout the course of this research. Credit is also due to Mr. James G. Kelly of the Naval Underwater Systems Center, Newport Laboratory whose research served as the impetus for this project. Finally, the author wishes to thank the staff of the W.R. Church Computer Center, Naval Postgraduate School for their assistance in support of the computer simulations used in the analysis.

I. INTRODUCTION

Critical to antisubmarine warfare (ASW) tactics and planning and to ASW analyses of all kinds are valid predictors of sonar detection performance. Measures of performance, such as a sonar's detection range probability distribution, are important considerations to the at-sea commander in choosing proper force deployment for a particular ASW situation. In addition, these performance measures serve as inputs to computer simulations and wargames used for tactics development, specifying and testing requirements for new systems, and for training.

Typically, analytically derived probability of detection models are based on the assumption of a detection occurring when the signal excess at the receiver (transducer) exceeds a given threshold value. Hence, probability of detection is equivalent to the probability of the signal excess exceeding the threshold. For a sonar system, signal excess is determined by the sonar equation to be a function of the sonar's figure of merit (a measure of the system's characteristics) and signal transmission loss. Signal excess is usually assumed to have a known component and an unknown component taken to be a random variable with some assumed underlying distribution. In practice, these models are overly optimistic because they depend upon simplifying assumptions which understate the complexities, for example, in the transmission

of sound through a medium such as the ocean. Recent examples are the probability of detection distributions for active sonars generated from models developed by Henry [1] which were shown to be optimistic estimates of actual operational sonar performance by Kelly [2].

It is intuitively appealing to use operational data to derive statistical estimates of a given sonar's probability of detection distribution; the estimate reflects actual sonar performance, not mathematical idealizations of it. Difficult conceptual problems are circumvented, such as how to deal with ping-to-ping correlation or how to model operator response. Finally, using operational data to estimate the detection distribution results in a useful tool for model validation. Kelly has shown that the product limit estimate (PL estimate) can be used to estimate the distribution of sonar detection ranges from fleet exercise data [2] and an estimate of the same form has been recommended for use in Navy operational ASW analyses [3,4]. However, as Kelly points out, methods of computing confidence intervals about the estimate and methods for using the estimate in hypothesis testing have not been derived [2,5]. Some of the PL estimate's intuitive appeal has, therefore, been lost because potential users have no procedure for determining the quality of the resulting estimate.

II. THE PRODUCT LIMIT ESTIMATE

A. THE KAPLAN-MEIER ESTIMATE

In 1958, Kaplan and Meier derived the product limit estimate, a nonparametric (distribution-free) maximum likelihood estimate of lifelength distributions in which censoring of some test values may have occurred [6]. The derivation of the estimate assumed that the lifelength and censoring distributions were independent.

Censoring is said to result when some items of the test sample are either lost to the experiment, or the experiment ends before the items fail. In each case, time of item failure is unknown, or censored, either by item loss or experiment completion. The PL estimate makes use of information contained in the censored event; although failure time is not known, time of failure must exceed the time from the moment at which the item was put on test until either it was lost to the experiment or the experiment ended.

B. APPLICATION TO DETECTION RANGE DISTRIBUTIONS

Similar censoring occurs in the analysis of detection range distributions. Consider a submarine closing on a surface ASW unit, for example a destroyer with active sonar. Let R, a random variable, be the detection range of the destroyer for this encounter. Let a random variable X be the distance to the submarine's closest point of approach

(CPA), or lateral range, to the destroyer. If a definite range law is assumed to hold, then if $R > X$, a detection occurs. But, if $X > R$, a detection does not occur and the true value of R is censored by X. This censoring of detection ranges by the lateral range is the basis for the applicability of the PL estimate.

C. DEFINING THE APPLIED PL ESTIMATE

Kelly has derived a PL estimate, denoted $\hat{G}(r)$, of the cumulative probability of detection distribution or $P[R>r]$ and has shown that it has the same properties as (and is similar in form to) the Kaplan-Meier estimate. The procedure for computing the estimate for n encounters or data points is as follows:

1. For each encounter i, record values of R_i or X_i. X_i is usually available from exercise reconstruction data.

2. Form the Z statistic such that

$$Z_i = \max(R_i, X_i) \tag{1}$$

3. Order the Z statistics such that

$$Z_{(1)} < Z_{(2)} < \ldots < Z_{(n)}$$

4. Define for all $i = 1, \ldots, n$ the random variable

$$D_{(i)} = \begin{cases} 1 & \text{if } Z_{(i)} \text{ is a detection range} \\ 0 & \text{if } Z_{(i)} \text{ is a lateral range only} \end{cases} \tag{2}$$

5. Compute the PL estimate $\hat{G}(r)$ pointwise for $\hat{G}_{(k)}$, $k = 1, \ldots, n$

$$\hat{G}_{(k)} = \frac{D_{(k+1)}}{k+1} + \left(1 - \frac{D_{(k+1)}}{k+1}\right) \hat{G}_{(k+1)}, \quad k = 1, \ldots, n-1$$

$$\hat{G}_{(n)} = 0 \tag{3}$$

Sample data is shown in Table I with the associated estimates graphed in Figure 1. Also shown is the empirical survival function estimate, $1 - \hat{F}(x)$, which Kaplan and Meier call the reduced sample estimate [6]. This estimate is computed from detection event data only and is defined by

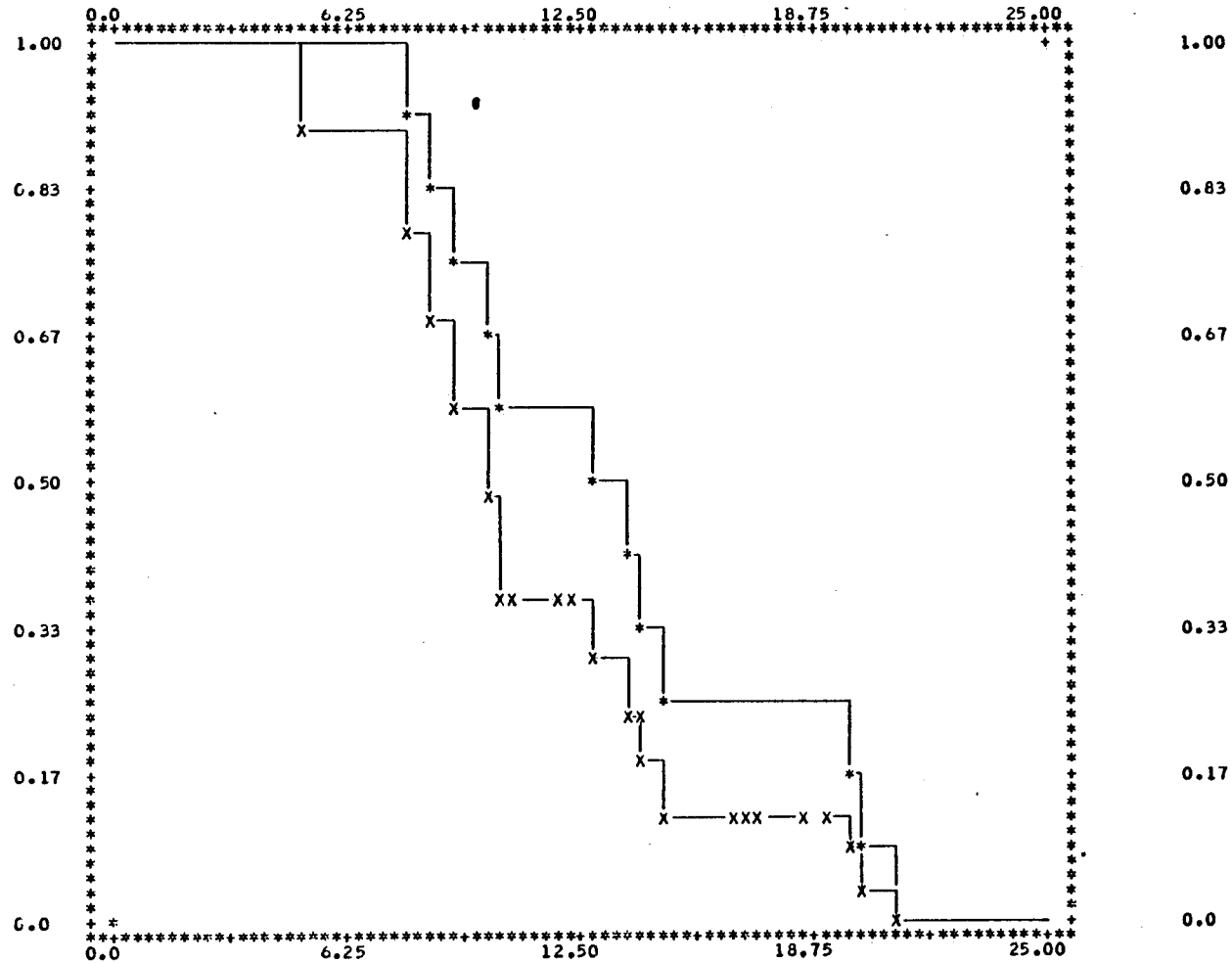
$$1 - \hat{F}(r) = \begin{cases} 1 & \text{for } r < R_{(1)} \\ 1 - \frac{k}{n} & \text{for } R_{(k)} \leq r < R_{(k+1)} \\ 0 & \text{for } r > R_{(n)} \end{cases} \tag{4}$$

This estimate is included for comparison with the PL estimate and because it has properties which allow some tests of statistical inference to be performed (such as the Kolmogorov-Smirnov test) which have not been proven to be applicable to the PL estimate with censored samples. Further, it should be noted that the PL estimate from a sample space of all uncensored data yields the same result as the reduced estimate.

R(I)	X(I)	Z(I)	D(I)
9.04229	6.14829	9.04229	1
13.60812	1.74829	13.60812	1
20.93770	14.39638	20.93770	1
10.01187	9.38966	10.01187	1
12.79815	9.53916	12.79815	1
14.20347	0.31195	14.20347	1
8.46316	2.63672	8.46316	1
19.70718	12.89940	19.70718	1
20.00397	5.65837	20.00397	1
14.57738	8.01824	14.57738	1
10.17032	4.91411	10.17032	1
7.74591	3.63399	7.74591	1

Z(I)	D(I)	GHAT(I)	RS(I)
4.87295	0	0.89513	1.00000
7.74591	1	0.79027	0.91667
8.46316	1	0.68540	0.83333
9.04229	1	0.58053	0.75000
10.01187	1	0.47567	0.66667
10.17032	1	0.37080	0.58333
10.53208	0	0.37080	0.58333
11.92003	0	0.37080	0.58333
12.16581	0	0.37080	0.58333
12.79815	1	0.30089	0.50000
12.83923	0	0.30089	0.50000
13.60812	1	0.23733	0.41667
14.17439	0	0.23733	0.41667
14.20347	1	0.17867	0.33333
14.57738	1	0.12000	0.25000
16.48331	0	0.12000	0.25000
16.73190	0	0.12000	0.25000
16.92357	0	0.12000	0.25000
17.12126	0	0.12000	0.25000
18.40530	0	0.12000	0.25000
18.58986	0	0.12000	0.25000
19.17041	0	0.12000	0.25000
19.70718	1	0.08000	0.16667
20.00397	1	0.04000	0.08333
20.93770	1	0.0	0.0

TABLE I



X-SCALE: "*"= 0.313E 00 UNITS

Y-SCALE: "*"= 0.167E-01 UNITS

ESTIMATE PLOT (X=PRODUCT LIMIT ESTIMATE, *=REDUCED SAMPLE ESTIMATE)

Figure 1

Censored sample values have the effect of depressing the PL estimate from the uncensored or reduced sample estimate. It is this degradation of the estimate for missed detections which makes the PL estimate seem reasonable.

III. TESTING THE PRODUCT LIMIT ESTIMATE

A. PROBLEM

Application of the PL estimate to detection data has been limited in part because statistical information concerning the quality of the estimate is lacking. Actual sonar detection distributions are unknown and are therefore unavailable for comparison with the estimates of the distributions derived from operational data. The effects of differing lateral range censoring distributions are also not known; if certain lateral range distributions perturb the estimates greatly, then exercise data analysis ought to include estimates of the distributions of the lateral ranges. Further, although the PL estimate is a maximum likelihood estimate (MLE), the sample size of operational data is typically small so that large-sample properties of the MLE are not necessarily to be relied upon. The problem, then, is to investigate the behavior of the PL estimate compared to a known detection distribution when the sample size is realistically small.

B. PROCEDURE

The overall procedure used was to assume a known probability of detection distribution for an arbitrary sonar system. Several lateral range distributions were derived based on plausible tactical situations. Encounters

between the submarine and destroyer were simulated on a computer; each encounter was called an event, or exercise event. A number of events constituted an exercise, the data from the exercise being used to construct the PL estimate. The estimate was then compared with the known detection distribution.

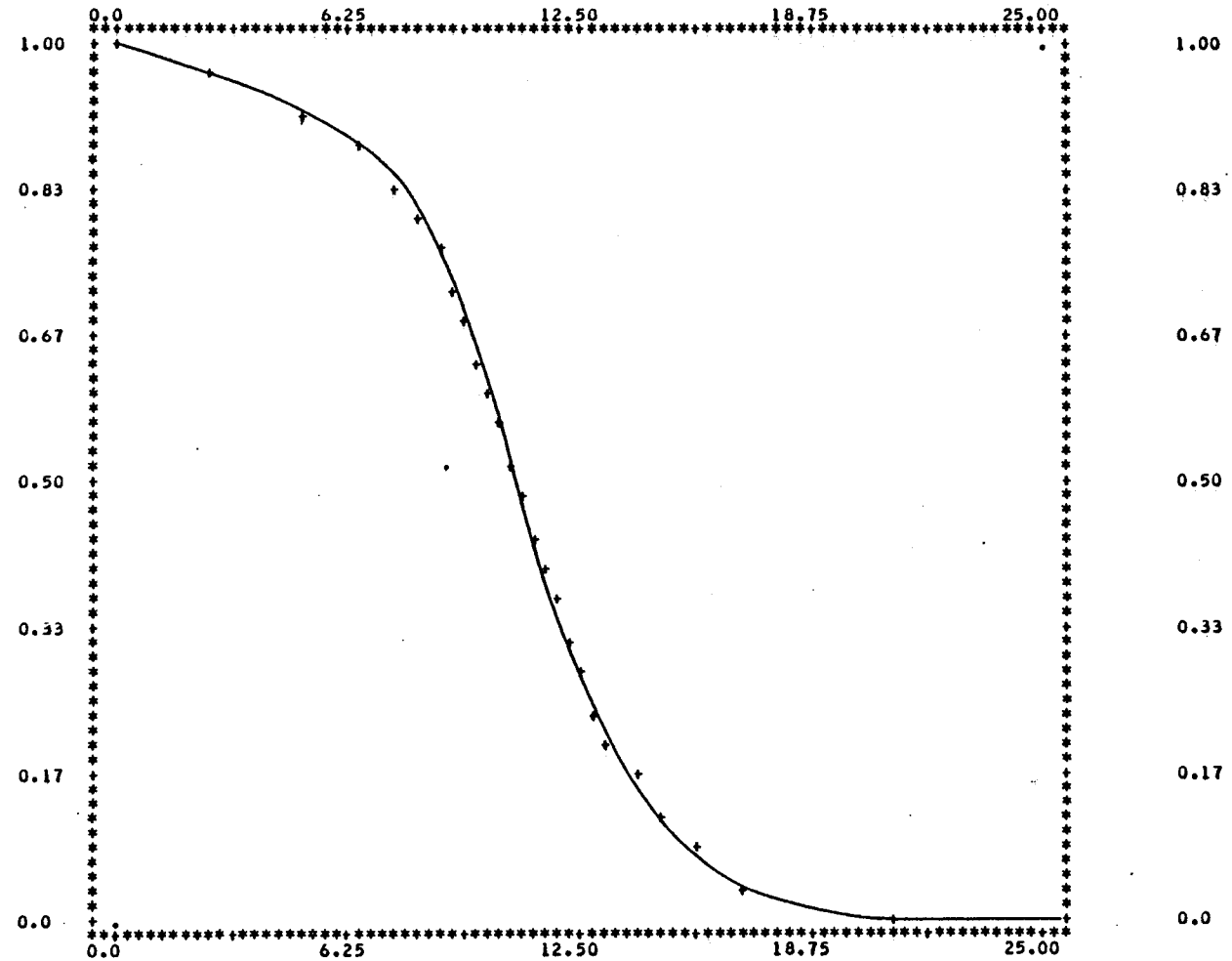
1. The Assumed Distribution

Based on the sonar equation and an assumed log normal distribution of signal excess, Henry [1] developed a family of detection distributions for various sonars operating under several "average" environmental conditions. The detection distribution assumed to be known for the purposes of the simulation is shown in Figure 2 with ranges in kiloyards (kyd) and has the characteristic shape of Henry's distributions.

Since the distribution could not be explicitly modeled or approximated by a well-known distributional form such as the Weibull or normal, twenty-six points were used to approximate the curve and were the inputs to the simulation. The probability integral transform method was used to sample R_i for each event.

2. The Lateral Range Distributions

Lateral range distributions and sampling procedures are given below. Sampling yielded a value of X_i for each event to be paired with corresponding R_i's.



ASSUMED DETECTION DISTRIBUTION

Figure 2

a. Random Search

In this model, the lateral range distribution is assumed to be uniform over the entire feasible range of the sonar, 0-20 kyd. This assumption is common throughout detection theory in describing a random search model. Therefore, the distribution function of X, $F_X(x)$ is

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{20} & 0 \leq x \leq 20 \\ 1 & x > 20 \quad . \end{cases} \tag{5}$$

Sampling is achieved by using the probability integral transform method on the closed-form distribution function such that $X_i = 20(U_i)$ where U is uniformly distributed over the interval [0,1].

b. Barrier Search

Consider the destroyer to be searching along a barrier front of width W, the barrier perpendicular to the submarine's estimated track, as in Figure 3 below.

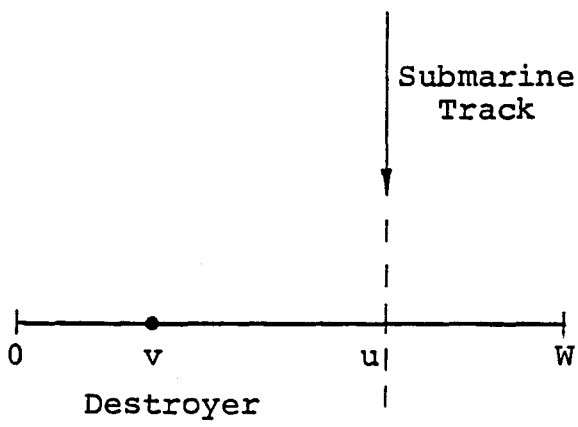


Figure 3

X is then a two-dimensional random variable with distribution function $F_X(x)$. Assuming U and V to be uniformly distributed over [0,W], then

$$F_X(x) = \int_0^\infty P[v-x \le U \le v+x]\ f_V(v)\,dv$$

$$= \int_0^x \int_0^{v+x} f_U(u)\,f_V(v)\,dudv + \int_x^{W-x} \int_{v-x}^{v+x} f_U(u)\,f_V(v)\,dudv$$

$$+ \int_{W-x}^{W} \int_{v-x}^{W} f_U(u)\,f_V(v)\ dudv$$

$$= \int_0^x \int_0^{v+x} \frac{1}{W^2}\,dudv + \int_x^{W-x} \int_{v-x}^{v+x} \frac{1}{W^2}\,dudv$$

$$+ \int_{W-x}^{W} \int_{v-x}^{W} \frac{1}{W^2}\,dudv$$

$$= 1 - \left(\frac{W-x}{W}\right)^2 \qquad (6)$$

Three barriers were considered: W = 25 kyd, W = 50 kyd, and W = 100 kyd. Sampling was done again by the probability integral transform such that $X_i = W(1 - \sqrt{U_i})$ where U is uniformly distributed over the interval [0,1].

c. Submarine Stationkeeping

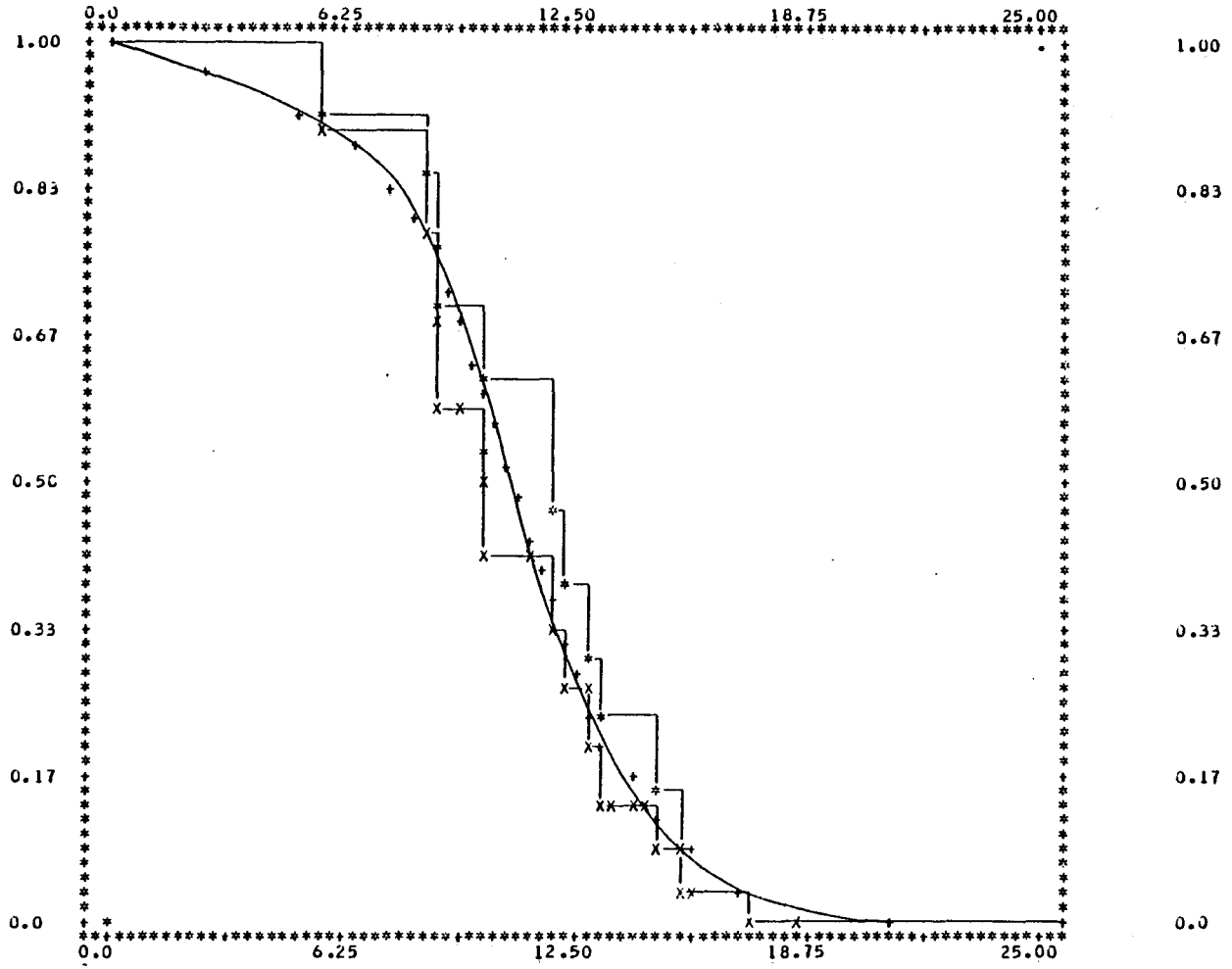
For a submarine closing to maintain station a distance S from the destroyer (to act in an escort role, to trail the destroyer or to close for an attack from distance S),

consider the lateral range distribution to be normal, with mean S and standard deviation of 1 kyd.

Sampling could have been accomplished by using a pre-programmed subroutine, but for speed and to minimize memory requirements, a statistical procedure was used which summed twelve uniform [0,1] random variates to yield a Normal(6,1) variate, then adding S-6 to achieve a mean equal to S. Values of S = 5 kyd and S = 10 kyd were tested. Since higher values for the standard deviation would tend to cause the lateral range distribution to be less peaked or spiked, and most likely perturb the estimate less, only a standard deviation of 1 kyd was tested.

C. SIMULATION AND DATA

An event i was simulated by sampling once from the known distribution to yield R_i and once from the lateral range distribution to yield X_i. Thus, for the ith event, the submarine would close on the destroyer to a range of X_i kyd, while the destroyer's detection range for the event would be R_i kyd. Twenty-five events were taken to be a realistic size for an exercise and the PL estimate and reduced sample estimate of detection range for each exercise were computed. Figures 4, 6 and 8 are the results of three exercises using the lateral range distribution associated with the random search model. Figures 5, 7, and 9 are percent or p-p plots [7] from the same respective cases; if the estimates and the

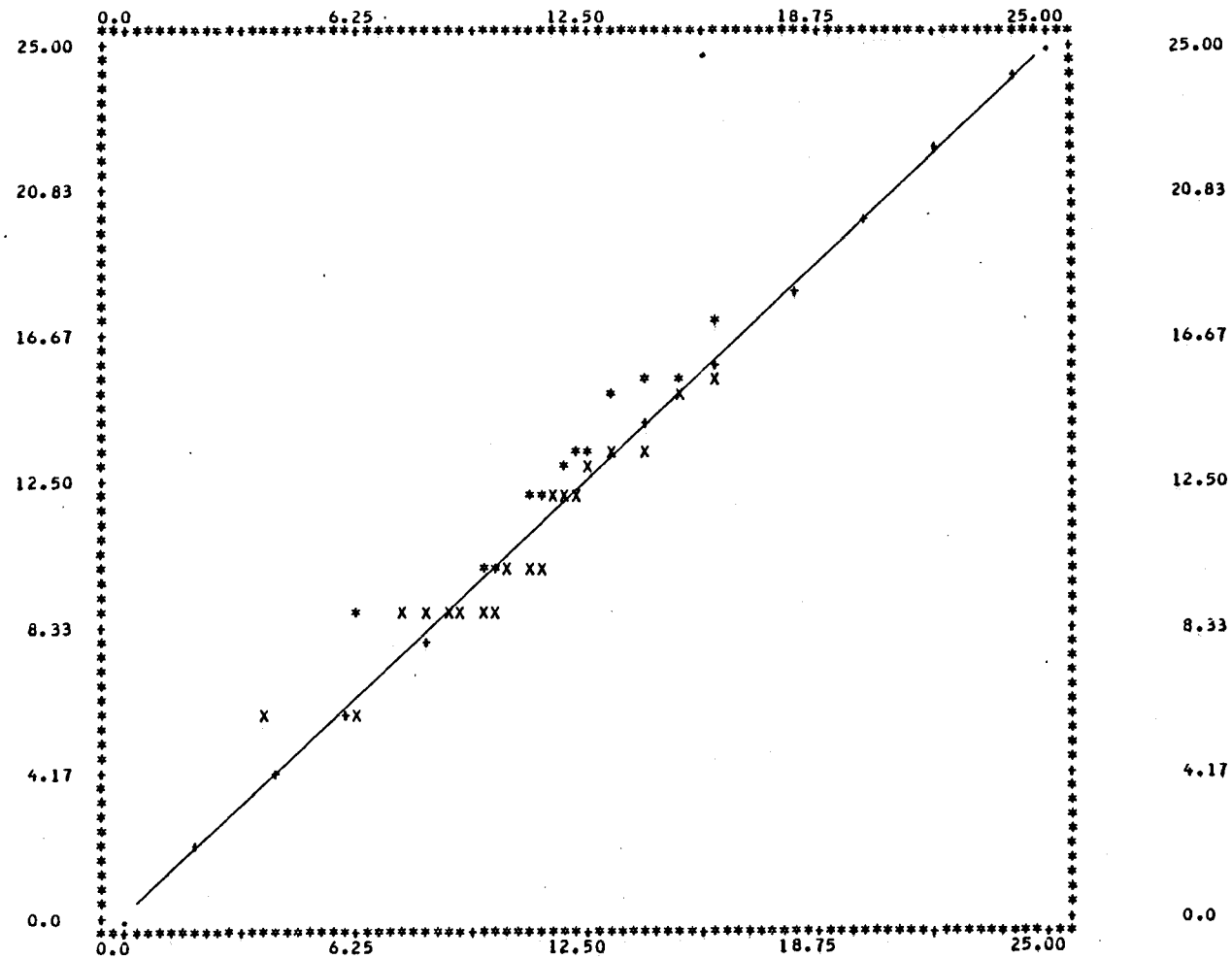


X-SCALE: "*"= 0.313E 00 UNITS

Y-SCALE: "*"= 0.167E-01 UNITS

DISTRIBUTION PLOT (+=ASSUMED DISTRIBUTION, X=PRODUCT LIMIT ESTIMATE, *=REDUCED SAMPLE ESTIMATE)

Figure 4

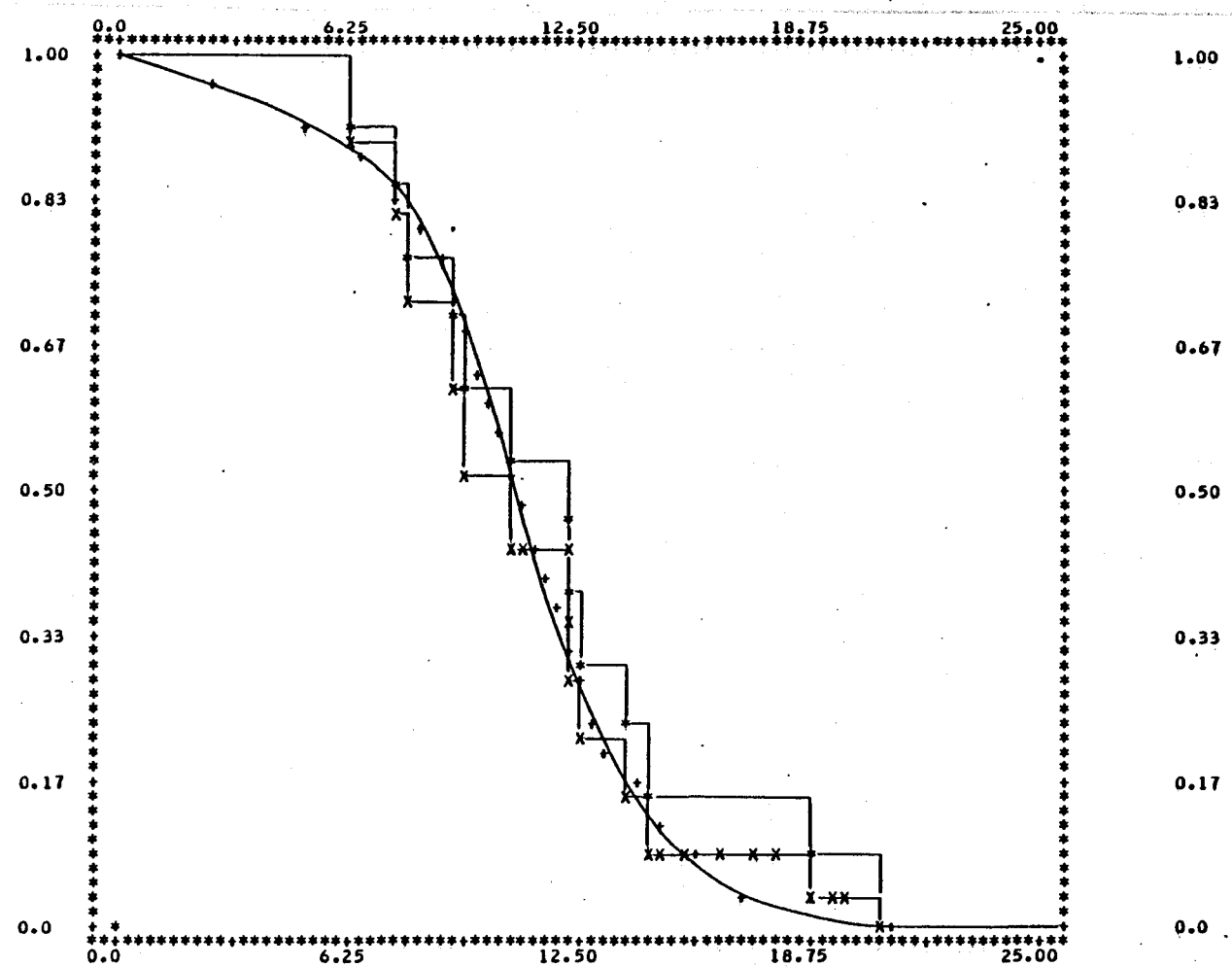


X-SCALE: "*"= 0.313E 00 UNITS

Y-SCALE: "*"= 0.417E 00 UNITS

P-P PLOT (+=REFERENCE, X=PRODUCT LIMIT ESTIMATE, *=REDUCED SAMPLE ESTIMATE)

Figure 5

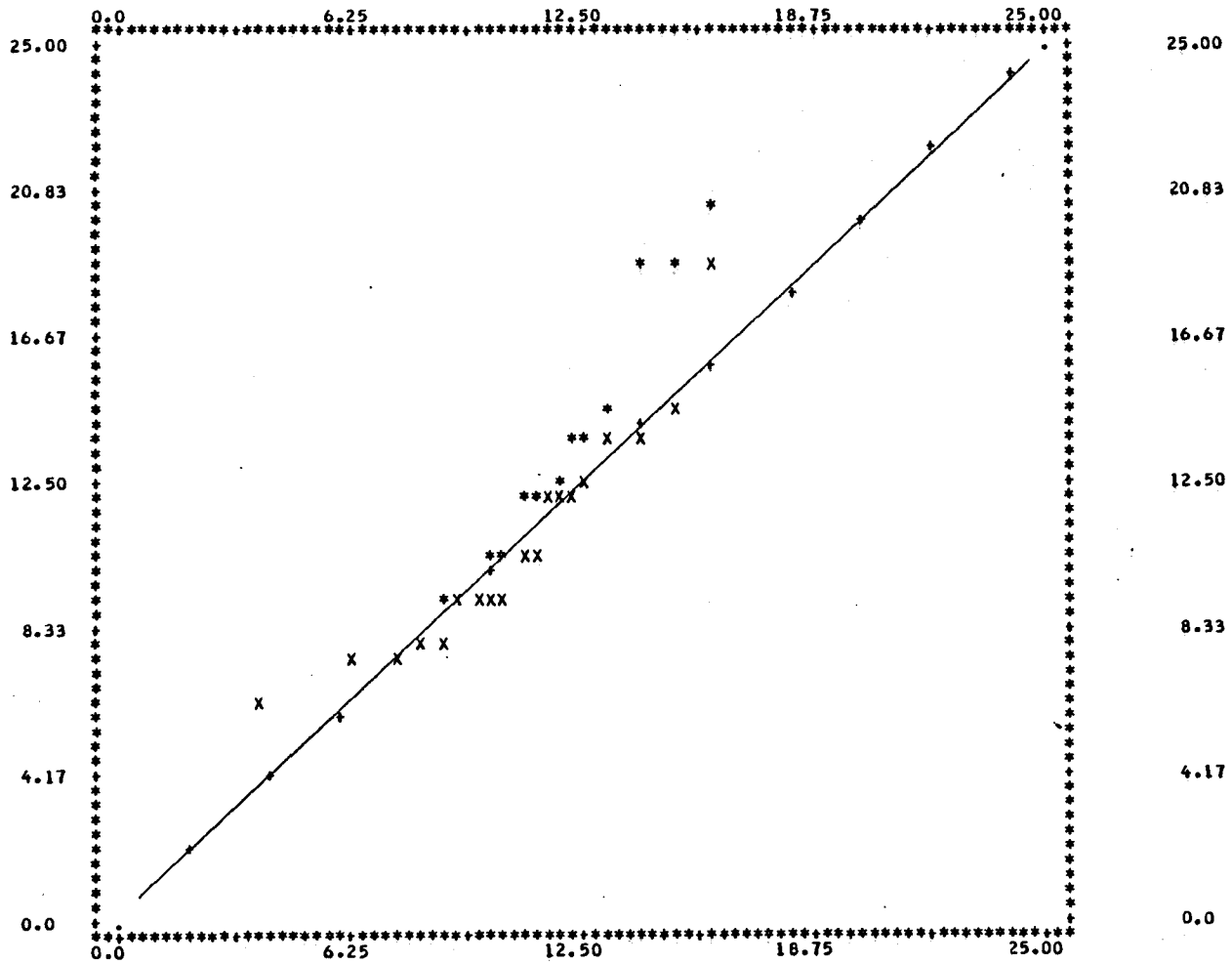


X-SCALE: "*"= 0.313E 00 UNITS

Y-SCALE: "*"= 0.167E-01 UNITS

DISTRIBUTION PLOT (+=ASSUMED DISTRIBUTION, X=PRODUCT LIMIT ESTIMATE, *=REDUCED SAMPLE ESTIMATE)

Figure 6

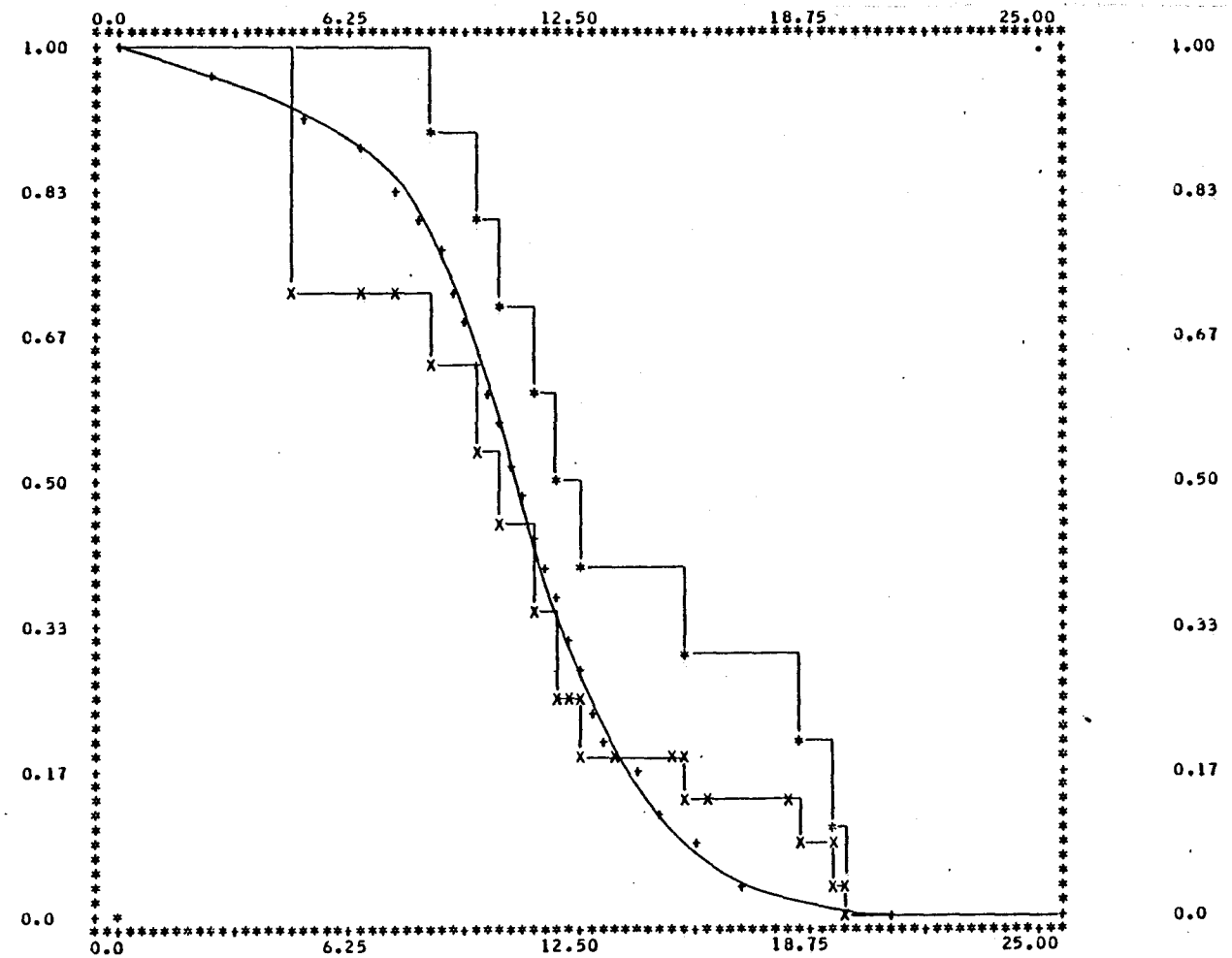


X-SCALE: "*"= 0.313E 00 UNITS

Y-SCALE: "*"= 0.417E 00 UNITS

P-P PLOT (+=REFERENCE, X=PRODUCT LIMIT ESTIMATE, *=EMPIRICAL SURVIVAL FUNCTION)

Figure 7

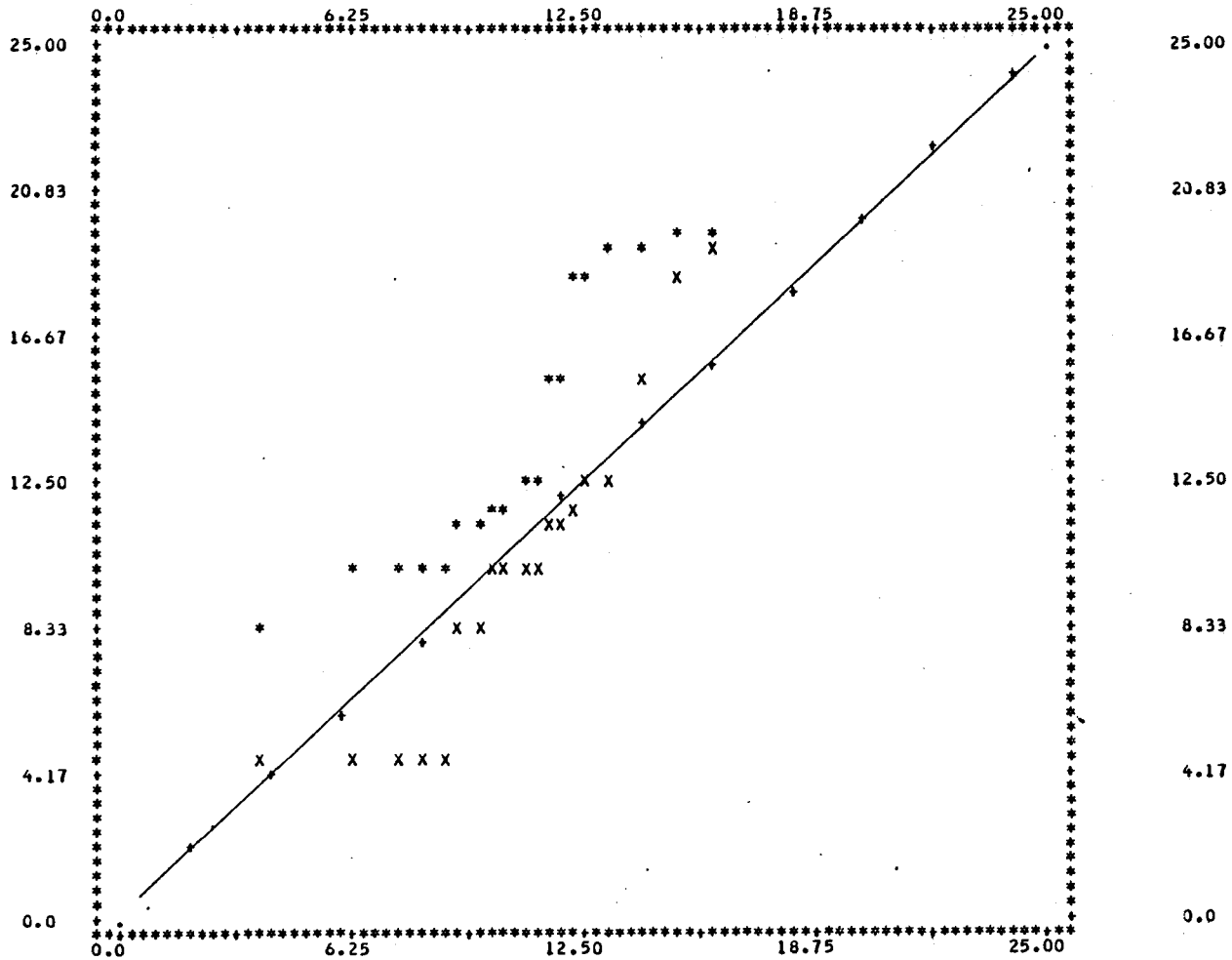


X-SCALE: "+"= 0.313E 00 UNITS

Y-SCALE: "*"= 0.167E-01 UNITS

DISTRIBUTION PLOT (+=ASSUMED DISTRIBUTION, X=PRODUCT LIMIT ESTIMATE, *=REDUCED SAMPLE ESTIMATE)

Figure 8



X-SCALE: "*"= 0.313E 00 UNITS

Y-SCALE: "*"= 0.417E 00 UNITS

P-P PLOT (+=REFERENCE, X=PRODUCT LIMIT ESTIMATE, *=REDUCED SAMPLE ESTIMATE)

Figure 9

known distribution were identically distributed, the p-p plots would appear as straight lines. A reference line is plotted for comparison.

Each pairing of known and lateral range distribution was used to simulate 100 exercises of 25 events each. Table II summarizes the results of the simulations for the six lateral range distributions. Averaged over the 100 exercises are the PL estimates and reduced sample estimates at the 90th, 75th, 50th, 25th and 10th percentiles. Figure 10 plots the average residuals between the estimates and the known distribution. Details of the computer simulation may be found in the Appendix.

D. CONCLUSIONS

The simulation results show that for exercises of twenty-five events each, both the PL estimate and reduced sample estimate conform rather well to the known detection distribution; this is particularly true for the percentiles at middle and long ranges. More importantly, the PL estimate does appreciably better on the average than the reduced sample estimate. Examination of all plots similar to Figures 4, 6 and 8 for all exercises (600 total) has shown that, as expected, the reduced sample estimate is always more optimistic than the PL estimate; on the average, this is shown in Table II and Figure 10 as well.

For exercises corresponding to the tactical situations of random search, barrier search for barriers of widths

LATERAL RANGE DISTRIBUTION	ESTIMATE	PERCENTILES (Known Distribution Value)									
		90 (6.15 kyd)		75 (8.70 kyd)		50 (10.85 kyd)		25 (12.60 kyd)		10 (15.05 kyd)	
		Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD
Uniform [0,20]	PL	6.06	2.34	8.26	1.77	10.82	1.00	12.90	1.09	15.04	1.49
	RS	8.24	1.66	9.96	1.06	11.91	1.02	13.96	1.53	16.32	1.90
Barrier (W = 25 kyd)	PL	5.94	2.19	8.53	1.42	10.84	0.90	12.82	1.01	15.06	1.54
	RS	7.68	1.56	9.74	1.03	11.67	0.79	13.63	1.34	16.06	1.66
Barrier (W = 50 kyd)	PL	6.27	2.50	8.26	2.03	10.71	1.21	12.74	1.17	14.95	1.80
	RS	7.90	2.11	9.86	1.46	11.80	1.24	13.75	1.63	16.15	2.36
Barrier (W = 100 kyd)	PL	7.49	2.57	8.28	2.52	10.61	1.81	12.78	1.83	15.39	2.77
	RS	8.49	2.51	10.36	1.79	12.39	1.86	14.40	2.59	15.91	2.84
Normal (5,1)	PL	6.06	1.46	8.64	1.02	10.84	0.82	12.67	0.91	15.13	1.37
	RS	7.65	0.92	9.27	0.74	11.23	0.73	13.14	1.02	15.21	1.41
Normal (10,1)	PL	8.66	0.60	8.93	0.71	10.75	0.91	12.66	0.92	15.13	1.37
	RS	10.36	0.66	11.33	0.56	12.50	0.80	14.38	1.17	16.38	1.81

TABLE II

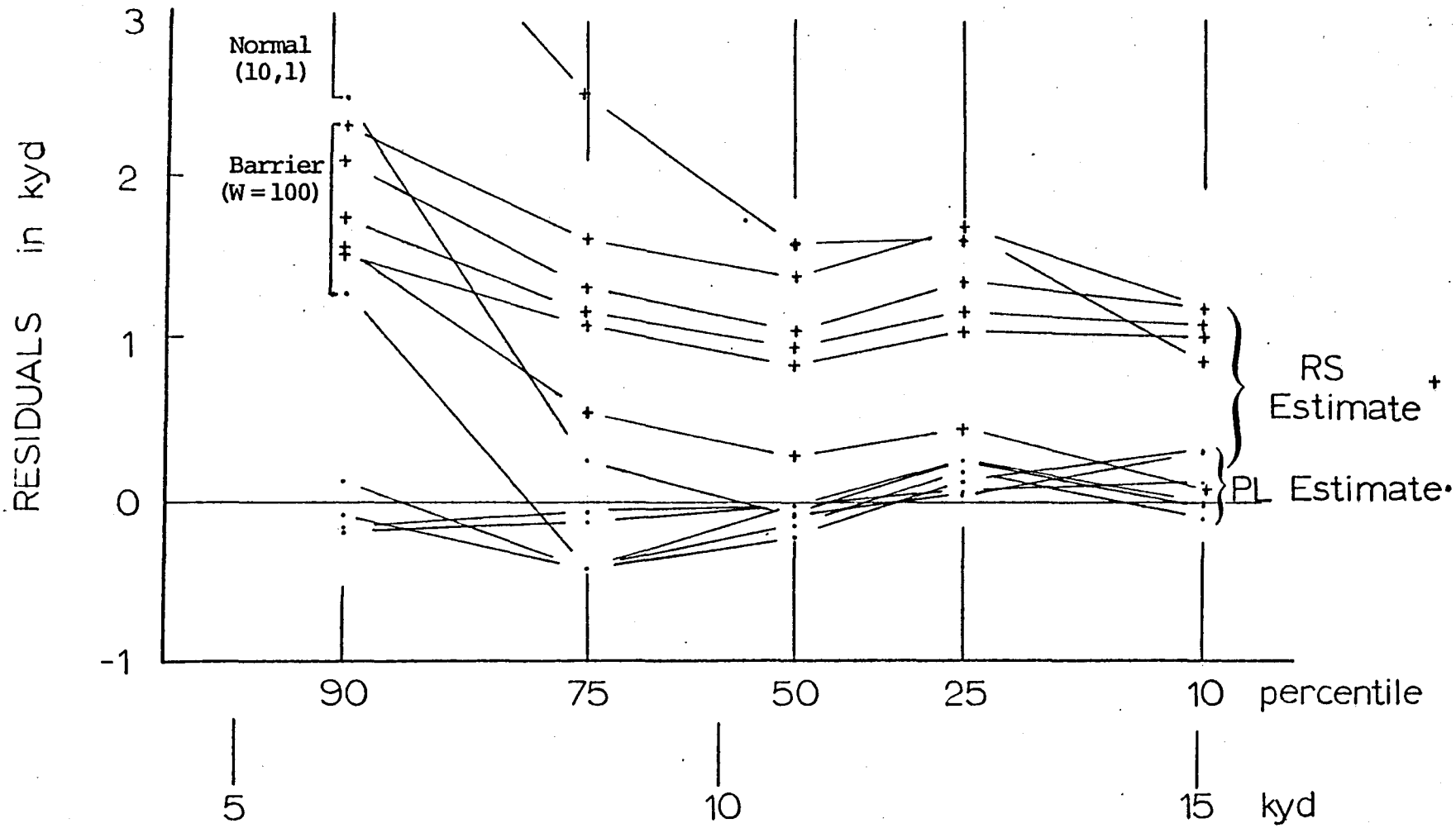

RESIDUALS in kyd
3
2
1
0
-1
Normal
(10,1)
Barrier
(W=100)
RS
Estimate
PL Estimate
90
75
50
25
10
percentile
5
10
15
kyd

Figure 10

25 kyd and 50 kyd and submarine stationkeeping of 5 kyd, the PL estimate appears to work extremely well. The lateral range distributions for these situations are characterized by expected values within the sonar's maximum detection range and variances such that samples tend to be spread throughout the sample space. As the submarine's CPA's become concentrated further from the destroyer, however, the predictive value of the PL estimate appears to decrease at the 90th percentile. For the barrier search with barrier width of 100 kyd and the stationkeeping at 10 kyd, lateral ranges tend to concentrate at greater distances and censoring (degradation of the PL estimate) occurs at longer ranges. Hence, the estimate seems to be somewhat optimistic at short range.

If it were feasible to actually run 100 exercises to estimate a sonar's detection distribution, then the results in Table II show that the PL estimate would do very well; the standard deviations of the percentile estimates would yield acceptably narrow confidence intervals. It should be noted that the data points for each percentile plotted as approximately straight lines on normal probability paper so that the assumption that the data is normally distributed is not a bad one. However, the problem of setting confidence intervals on data from one exercise only is the challenging problem. A solution would be most useful to the analyst.

IV. CONFIDENCE INTERVALS

A. PROBLEM

The nature of sonar detection data acquisition precludes obtaining large sample sizes except in the rare cases for which large-scale aggregation may be appropriate. Because of the time and expense involved in conducting exercises at sea, sample sizes are typically small. Miller [8] has shown that jackknifing the PL estimate produces confidence intervals which are asymptotically valid (i.e. cover the true parameters with stated frequency), yet the asymptotic properties may not be valid for small samples. The problem is to determine the properties of the jackknife when it is used for computing confidence intervals for the PL estimate of the detection range distribution from data from a single exercise.

B. PROCEDURE

Using one lateral range distribution as a sample case, a single exercise of twenty-five encounters was computer simulated. Data was jackknifed according to the procedures described by Mosteller and Tukey [9]. The resulting pseudo-values were used to generate nominal 80%, 95% and 99% confidence intervals about selected percentiles of the estimate. One thousand exercises were simulated, and the number of times the confidence intervals covered the

percentiles of the known distribution was recorded. A simulation was also run with a correction factor inserted to adjust for possible pseudo-value correlation, and the resulting confidence intervals were noted. Finally, the percentage of confidence intervals which actually covered each parameter was compared with the nominal or stated coverage.

1. Lateral Range Distribution

Of the lateral range distributions derived in Section III, the uniform distribution associated with the random search model is probably the one which is applicable to a majority of exercise situations. For this reason, the uniformly distributed lateral range was used to examine the confidence intervals from the jackknife.

2. Parameters

Selected percentiles of the detection range distribution were taken as the parameters to be estimated by the jackknife. This was done because the percentiles have operational appeal; it is typical to want to know that 50% of the time, detection range will exceed 8 kyd (±2 kyd at 95% confidence), say, than to want to know that for a target at 8 kyd or greater, probability of detection is 0.5 (±.15 at the 95% confidence level). In a sense, the two are equivalent in that both can be used to generate confidence bands about the estimate.

3. The Jackknife

The jackknife procedure yielded, for each exercise, pseudo-values which were used to generate confidence intervals. The pseudo-values resulted from PL estimates of the data with some data points omitted.

For example, consider an exercise with twenty-four encounters. Twenty-four pseudo-values for each percentile can be generated by computing the PL estimate twenty-four times with a sample size of twenty-three, leaving out one observed Z_i value each time. Similarly, pseudo-values may be generated from grouped data. As an example, eight pseudo-values can be generated by computing the estimate eight times, omitting three different observations each time. An example of the results of the latter case is shown in Table III.

According to Tukey, the pseudo-values can be assumed to be normally distributed with their mean estimating the percentile and the variances estimated by the usual s^2 statistic. Confidence intervals are obtained by using Student's t in the usual manner; however, one degree of freedom is lost for every duplicate pseudo-value. See Gray and Schucany [10] for asymptotic verification.

4. Intraclass Correlation

Little is known about the correlation between pseudo-values, although intuitively, correlation ought to exist to some degree. Consider two pseudo-values from a sample of twenty-four where one observation is omitted from the PL

PSEUDO-VALUES

90	75	50	25	10
9.03026	10.96828	11.42198	13.22559	15.33658
8.92673	10.83189	11.82939	21.17265	15.33658
2.00990	6.20160	11.02980	16.65758	21.57747
2.34384	9.06931	11.39334	13.21452	15.33658
8.73779	10.58301	11.44626	16.94177	21.57747
9.02261	9.20656	10.28065	13.22063	15.33658
8.90831	10.80763	11.56508	16.65758	21.57747
8.96167	10.87790	11.97289	16.95706	15.47432

ESTIMATES

7.24264	9.81827	11.36742	16.00592	17.69412

S/SQRT(N)

1.10635	0.58327	0.18531	0.96921	1.13704

T STATISTIC / (DEGREES OF FREEDOM)

2.36500	2.36500	2.36500	2.44700	4.30300
(7)	(7)	(7)	(6)	(2)

95% CONFIDENCE INTERVALS

4.62612	8.43884	10.92916	13.63427	12.80144
9.85916	11.19770	11.80569	18.37756	22.58679

TABLE III

estimate each time. Each pseudo-value, then, has twenty-two common data points used in computing the estimate and positive correlation between them would seem reasonable.

Gray and Schucany [10], following Walsh [11], show that for correlation ρ between pseudo-values which are assumed to be samples from a multivariate normal population, the correlation factor for Student's t statistic would be

$$\left(\frac{1 + (n-1)\rho}{1 - \rho}\right)^{\frac{1}{2}} \tag{7}$$

They suggest that an adequate model of pseudo-value correlation for n pseudo-values is $\rho = 1/n$. Thus, the correlation factor becomes $[(2n-1)/(n-1)]^{1/2}$.

C. SIMULATION

One thousand exercises of twenty-five events were simulated. Data from each exercise was grouped in five different ways, pseudo-values of the selected percentiles were computed using the jackknife for each grouping, and confidence intervals were computed for each percentile estimate. For groupings requiring only twenty-four data points, the final observation of each exercise (not the last ordered observation) was omitted. The Gray and Schucany correlation correction was applied to the sample of twenty-five pseudo-values. Results were compared with the percentiles of the known distribution and are shown in Tables IV(a) - IV(c). Details of the computer simulation are included in the Appendix.

SAMPLE SIZE	DATA GROUPS FOR JACKKNIFE A. % Coverage B. Average C.I. Width C. Average d.f.		80% CONFIDENCE Percentile 90	75	50	25	10
25	25 Pseudo-values (25 groupings of 24 data points each; omit 1 data point each time)	A	47.4	45.5	37.6	40.9	49.3
		B	4.1	3.7	2.8	2.7	4.4
		C	10.8	10.3	8.9	6.9	4.5
25	25 Pseudo-values (25 of 24; omitting 1 each time) Incl. ρ correction	A	65.7	56.7	50.3	57.1	64.0
		B	5.9	5.2	3.9	4.2	5.6
		C	10.8	10.4	9.1	6.9	4.5
24	8 Pseudo-values (8 of 21; omitting 3 each time)	A	44.2	43.9	47.1	46.5	39.1
		B	3.9	3.9	3.0	3.1	4.4
		C	6.8	6.8	6.7	6.2	4.6
25	5 Pseudo-values (5 of 20; omitting 5 each time)	A	75.2	65.8	60.7	63.4	51.7
		B	10.8	6.2	3.2	3.2	4.8
		C	3.998	3.99	3.99	3.9	3.4
24	4 Pseudo-values (4 of 18; omitting 6 each time)	A	52.6	60.3	64.0	64.6	60.3
		B	4.5	4.4	3.2	3.3	5.2
		C	2.996	2.998	2.994	2.97	2.8
24	3 Pseudo-values (3 of 16; omitting 8 each time)	A	59.0	68.5	69.3	71.9	71.9
		B	5.0	4.8	3.5	3.7	5.6
		C	1.998	1.999	2.0	1.993	1.96

TABLE IV(a)

SAMPLE SIZE	DATA GROUPS FOR JACKKNIFE A. % Coverage B. Avg. C.I. Width C. Average d.f.		95% CONFIDENCE Percentile 90	75	50	25	10
25	25 Pseudo-values	A	70.8	68.8	59.5	66.0	74.4
		B	6.5	6.2	4.9	5.1	9.6
		C	10.8	10.3	8.9	6.9	4.5
25	25 Pseudo-values Incl. ρ correction	A	82.2	79.5	75.0	78.9	78.7
		B	9.2	8.4	6.3	7.2	10.7
		C	10.8	10.4	9.1	6.9	4.5
24	8 Pseudo-values	A	68.2	68.8	72.2	73.6	63.0
		B	6.6	6.5	5.0	5.2	7.8
		C	6.8	6.8	6.7	6.2	4.6
25	5 Pseudo-values	A	95.5	88.4	86.5	86.2	80.6
		B	18.2	10.9	5.8	5.8	9.0
		C	3.998	3.99	3.99	3.9	3.4
24	4 Pseudo-values	A	78.8	87.1	87.8	89.0	85.1
		B	8.3	8.3	6.2	6.5	10.3
		C	2.996	2.998	2.994	2.97	2.8
24	3 Pseudo-values	A	86.5	90.0	91.2	91.3	89.6
		B	10.4	10.4	7.9	8.4	13.2
		C	1.998	1.999	2.0	1.993	1.96

TABLE IV(b)

SAMPLE SIZE	DATA GROUPS FOR JACKKNIFE A. % Coverage B. Avg. C.I. Width C. Average d.f.		99% CONFIDENCE Percentile 90	75	50	25	10
25	25 Pseudo-values	A	80.5	81.8	77.9	80.6	84.3
		B	9.0	8.9	8.1	8.9	20.9
		C	10.8	10.3	8.9	6.9	4.5
25	25 Pseudo-values Incl. ρ correction	A	89.1	88.8	88.1	91.0	87.0
		B	12.5	11.7	9.1	10.8	17.5
		C	10.8	10.4	9.1	6.9	4.5
24	8 Pseudo-values	A	80.2	87.7	88.2	88.5	81.6
		B	9.5	9.6	7.4	7.9	12.7
		C	6.8	6.8	6.7	6.2	4.6
25	5 Pseudo-values	A	98.4	96.7	94.6	94.1	90.3
		B	21.0	16.4	9.5	9.5	15.8
		C	3.998	3.99	3.99	3.9	3.4
24	4 Pseudo-values	A	92.4	96.2	95.6	96.8	92.4
		B	14.0	14.0	11.1	11.7	19.6
		C	2.996	2.998	2.994	2.97	2.8
24	3 Pseudo-values	A	96.7	97.6	96.5	98.1	96.3
		B	21.0	20.1	16.4	18.2	29.8
		C	1.998	1.999	2.0	1.99	1.96

TABLE IV(c)

D. DISCUSSION AND CONCLUSIONS

In every case, the confidence intervals computed from the jackknifed pseudovalues did not cover the known percentile value at the percentage specified, the nominal level. More importantly, the worst under-coverage occurred where confidence intervals were sufficiently narrow to be of real value; better coverage occurred only at the expense of wider confidence intervals. In analyzing 200 exercises in which the 95% confidence intervals did not capture the true percentiles in the five-group (twenty events each) runs, an approximately equal number of misses above and below the actual values were found; therefore, no bias was indicated. Plots of groups of twenty-five pseudo-values on normal probability paper resulted in rather poor straight-line approximations indicating some doubt in the assumption that these pseudo-values could be normally distributed. Robust procedures, chiefly Winsorizing the pseudo-values before computing the confidence intervals, failed; the estimate (the average of the pseudo-values) depended heavily on pseudo-value outliers and accuracy decreased markedly where the outliers were Winsorized.

Use of the correlation correction improved confidence interval coverage, but in this case, was not as adequate a model as Gray and Schucany suggest. A value of ρ could be found by trial and error which would ensure proper coverage, but would require a determination of actual pseudo-value correlation for comparison and justification before it could

be used. Also, it is likely that sufficiently wide confidence intervals would result that their usefulness would be degraded.

As discussed by Gaver [12], the tendency for the jackknifed confidence intervals to cover the true parameter less frequently than specified is not uncommon. For small sample sizes, when the jackknife may be the only tool available, this possible tendency to under-cover must be acknowledged. Resulting confidence intervals should be viewed as likely to be optimistic, although the correlation correction factor is of help in improving the coverage.

APPENDIX A

COMPUTER SIMULATION FOR TESTING EFFECTS OF VARYING TACTICAL SITUATION

The computer simulation is a discrete event digital simulation; each exercise event is a simulation event. The program is written in the FORTRAN IV language, and is grouped into easily identified structural sections shown in the listing on page 45.

The following codes appear in the left margin to assist potential users. Except for the C for "comment", these codes are for documentation purposes only and must not appear in a program to be executed.

CODE	MEANING
C	Comments
D	Distribution: These cards determine the lateral range distribution to be sampled
OUT	Output: These statements control the frequency at which interim results are printed. As written, data from every 20th exercise is printed

One-dimensional arrays are listed in the order in which they appear and are used for the following purposes:

ARRAY NAMES	PURPOSE
XS, YS	Scaling arrays for the XY plot subroutine - contain axis endpoints

PPSUM, PESUM	Contain the sums of each of the five percentiles (90, 75, 50, 25, 10) for all exercises. PPSUM for the PL estimate and PESUM for the reduced sample estimate. Used to find average percentile values at end of program.
PPSSQ, PESSQ	As for PPSUM and PESUM above, except contain the sum of the squares of the percentiles.
TILE	Contains percentile values which may be of interest (5% - 95% in increments of 5%).
XG	Holds values to plot the 45° reference line in the p-p plots.
PK, RK	PK contains the probabilities of the known detection distribution associated with each range RK.
Z	Contains Z_i's (later, $Z_{(i)}$'s) for each exercise.
D, ZHOLD	Holds D_i and Z_i (unordered) values for each exercise.
DELTA	The $D_{(i)}$ array.
GHAT, ECDF	Contains values of PL estimate (GHAT) reduced sample estimate (ECDF) corresponding to ranges in Z array.
DSUM	Accumulates the number of detection events for each exercise.
PCDF, RCDF	Contains probabilities (PCDF) and corresponding ranges (RCDF) of reduced sample estimate.
PTILE, EDTILE	Hold percentiles of PL estimate (PTILE) and reduced sample estimate (EDTILE) corresponding to respective elements of TILE.
P90K a b	Hold selected values of KTILE, PTILE, EDTILE: Part(a) indicates 90, 75, 50, 25, or 10th percentile Part(b) indicates KTILE, PTILE or EDTILE.

AP, AE	Contains average percentile values for PL and reduced sample estimates respectively.
SP, SE	Contains standard deviation of percentile values.

Subroutine RANDOM (ISEED, U, 1) uses ISEED as the seed value for a random number generator and returns one number U which is uniformly distributed over the interval [0,1]. Subroutine OVFLOW suppresses overflow errors caused by RANDOM.

Program output is in two forms:

1. Interim Output

For each selected exercise, output consists of a listing of the $Z_{(i)}$'s, $D_{(i)}$'s and the PL and reduced estimates as in the lower half of Table I, a plot of the assumed distribution and the estimates as in Figure 4, and a p-p plot of the same form as Figure 5.

2. Final Output

A table is generated showing the five percentile (90, 75, 50, 25 and 10) values generated by the estimates for each of 100 exercises. The averages and standard deviations for each percentile over the 100 exercises are also printed.

```
      CALL OVFLOW
      ISEED=490214
      INTEGER D(25),DELTA(25),DSUM
      REAL RK(26),PK(26),Z(25),ZHOLD(25),GHAT(25),XS(2),
     1PCDF(25),RCDF(25),XG(12),TILE(19),KTILE(19),PTILE(19),
     2P90K(100),P90P(100),P90E(100),P75K(100),P75P(100),
     3P50K(100),P50P(100),P50E(100),P25K(100),P25P(100),
     4P10K(100),P10P(100),P10E(100),PPSUM(5),PESUM(5),
     5AP(5),AE(5),SP(5),SE(5),ECDF(25),YS(2),EDTILE(19),
     6P75E(100),P25E(100),PPSSQ(5),PESSQ(5)
      XS(1) = 0.0
      XS(2) = 25.0001
      YS(1) = 0.0
      YS(2) = 1.0001
      DO 5 I = 1,12
      XG(I) = 2. * I
    5 CONTINUE
      DO 10 I = 1,5
      PPSUM(I) = 0.0
      PESUM(I) = 0.0
      PPSSQ(I) = 0.0
      PESSQ(I) = 0.0
   10 CONTINUE
C  SET PERCENTILES REQUIRED
      DO 15 I = 1,19
      TILE(I) = I * .05
   15 CONTINUE
C  READ IN KNOWN PROBABILITY OF DETECTION DISTRIBUTION
      READ(5,200)(PK(I),RK(I),I=1,26)
  200 FORMAT(2F10.5)
      DO 95 M = 1,100
C  GENERATE RANDOM VARIATES
      DO 25 I = 1,25
D     CALL RANDOM(ISEED,U,1)
D     X = 20. * U
      CALL RANDOM(ISEED,U,1)
      DO 20 J = 1,26
      IF(U.GT.PK(J)) GO TO 20
      R = RK(J) + (PK(J)-U)*(RK(J-1)-RK(J))/(PK(J)-PK(J-1))
      GO TO 21
   20 CONTINUE
   21 IF(R.GE.X) GO TO 22
      Z(I) = X
      D(I) = 0
      GO TO 25
   22 Z(I) = R
      D(I) = 1
   25 CONTINUE
C  COPY Z ARRAY INTO Z HOLD, D ARRAY INTO DELTA
      DO 30 I = 1,25
      ZHOLD(I) = Z(I)
      DELTA(I) = D(I)
      ECDF(I) = 0.0
   30 CONTINUE
C  SORT Z ARRAY
      CALL PXSORT(Z,1,25)
C  MATCH DELTAS TO SORTED ARRAY
      DO 40 I = 1,25
      DO 35 J = 1,25
      IF(Z(I).NE.ZHOLD(J)) GO TO 35
```

```
      DELTA(I) = D(J)
      GO TO 40
   35 CONTINUE
   40 CONTINUE
C  COMPUTE PRODUCT LIMIT ESTIMATE
      GHAT(25) = 0.0
      DO 45 J = 1,24
      K = 26 - J
      L = 25 - J
      XI = FLOAT(DELTA(K))/FLOAT(K)
      GHAT(L) = XI + (1. - XI)*GHAT(K)
   45 CONTINUE
C  COMPUTE ECDF FROM DETECTION EVENTS ONLY
      DSUM = 0
      DO 50 I = 1,25
      RCDF(I) = 0.0
      PCDF(I) = 0.0
      DSUM = DSUM + DELTA(I)
   50 CONTINUE
      K = 1
      DO 60 I = 1,DSUM
      DO 55 J = K,25
      IF(DELTA(J).EQ.0) GO TO 51
      RCDF(I) = Z(J)
      PCDF(I) = 1. - FLOAT(I)/FLOAT(DSUM)
      K = K + 1
      GO TO 60
   51 K = K+1
   55 CONTINUE
   60 CONTINUE
C  COMPUTE SELECTED PERCENTILES
      DO 80 I = 1,19
      X = TILE(I)
      DO 65 K = 1,26
      J = 27 - K
      IF(X.LT.PK(J)) GO TO 65
      IF(PK(J)-PK(J-1))61,61,62
   61 KTILE(I) = 21.0
      GO TO 66
   62 A = PK(J)-PK(J-1)
      KTILE(I)=RK(J)+(RK(J-1)-RK(J))*(PK(J)-X)/A
      GO TO 66
   65 CONTINUE
   66 DO 70 J = 1,25
      IF(X.GT.GHAT(J)) GO TO 71
   70 CONTINUE
   71 PTILE(I) = Z(J)
      DO 75 J = 1,25
      IF(X.GT.PCDF(J)) GO TO 76
   75 CONTINUE
   76 EDTILE(I) = RCDF(J)
   80 CONTINUE
      P90K(M) = KTILE(18)
      P75K(M) = KTILE(15)
      P50K(M) = KTILE(10)
      P25K(M) = KTILE(5)
      P10K(M) = KTILE(2)
      P90P(M) = PTILE(18)
      P75P(M) = PTILE(15)
      P50P(M) = PTILE(10)
      P25P(M) = PTILE(5)
      P10P(M) = PTILE(2)
      P90E(M) = EDTILE(18)
      P75E(M) = EDTILE(15)
      P50E(M) = EDTILE(10)
      P25E(M) = EDTILE(5)
```

```
      P10E(M) = EDTILE(2)
      PPSUM(1) = PPSUM(1) + P90P(M)
      PPSUM(2) = PPSUM(2) + P75P(M)
      PPSUM(3) = PPSUM(3) + P50P(M)
      PPSUM(4) = PPSUM(4) + P25P(M)
      PPSUM(5) = PPSUM(5) + P10P(M)
      PESUM(1) = PESUM(1) + P90E(M)
      PESUM(2) = PESUM(2) + P75E(M)
      PESUM(3) = PESUM(3) + P50E(M)
      PESUM(4) = PESUM(4) + P25E(M)
      PESUM(5) = PESUM(5) + P10E(M)
      PPSSQ(1) = PPSSQ(1) + (P90P(M)**2)
      PPSSQ(2) = PPSSQ(2) + (P75P(M)**2)
      PPSSQ(3) = PPSSQ(3) + (P50P(M)**2)
      PPSSQ(4) = PPSSQ(4) + (P25P(M)**2)
      PPSSQ(5) = PPSSQ(5) + (P10P(M)**2)
      PESSQ(1) = PESSQ(1) + (P90E(M)**2)
      PESSQ(2) = PESSQ(2) + (P75E(M)**2)
      PESSQ(3) = PESSQ(3) + (P50E(M)**2)
      PESSQ(4) = PESSQ(4) + (P25E(M)**2)
      PESSQ(5) = PESSQ(5) + (P10E(M)**2)
C  PRINT AND PLOT SELECTED INTERIM RESULTS
OUT   ICHECK = M/20
OUT   IF(FLOAT(M)/20.- ICHECK) 95,81,95
   81 WRITE(6,250)
  250 FORMAT(1H1)
      CALL PLOTP(XS,YS,2,1)
      CALL PLOTP(RK,PK,26,2)
      CALL PLOTP(RCDF,PCDF,25,2)
      CALL PLOTP(Z,GHAT,25,3)
      WRITE(6,300)
  300 FORMAT(1X//8X,'DISTRIBUTION PLOT  (+=ASSUMED DISTRIBUT
     1ION, X=PRODUCT LIMIT ESTIMATE, *=REDUCED SAMPLE ESTIMA
     2TE)'//)
      WRITE(6,325)
  325 FORMAT(1H1)
      CALL PLOTP(XS,XS,2,1)
      CALL PLOTP(XG,XG,12,2)
      CALL PLOTP(KTILE,EDTILE,19,2)
      CALL PLOTP(KTILE,PTILE,19,3)
      WRITE(6,350)
  350 FORMAT(1X//8X,'P-P PLOT  (+=REFERENCE, X=PRODUCT LIMIT
     1 ESTIMATE, *=REDUCED SAMPLE ESTIMATE)'//)
  400 FORMAT(1H1/,19X,'Z(I)',4X,'DELTA(I)',3X,'GHAT(I)',6X,
     1'RS(I)'//)
      N = 1
      DO 85 I = 1,25
      IF(DELTA(I).NE.0) GO TO 86
      ECDF(I) = 1.0
   85 CONTINUE
   86 DO 90 K = I,25
      IF(DELTA(K).EQ.0) GO TO 91
      ECDF(K) = PCDF(N)
      N = N + 1
      GO TO 90
   91 L = N - 1
      ECDF(K) = PCDF(L)
   90 CONTINUE
      WRITE(6,400)
  400 FORMAT(1H0,18X,'Z(I)',4X,'D(I)',3X,'GHAT(I)',6X,
     1'RS(I)'//)
      WRITE(6,450)(Z(I),DELTA(I),GHAT(I),ECDF(I),I=1,25)
  450 FORMAT(15X,F10.5,4X,I1,4X,F7.5,2X,F10.5)
   95 CONTINUE
C  PRINT FINAL RESULTS
      WRITE(6,500)
  500 FORMAT(1H1,20X,'P90',21X,'P75',21X,'P50',21X,'P25',
     121X,'P10'/)
```

```
      DC 105 I = 1,100
      WRITE(6,550) P90K(I),P90P(I),P90E(I),P75K(I),P75P(I),
     1P75E(I),P50K(I),P50P(I),P50E(I),P25K(I),P25P(I),
     2P25E(I),P10K(I),P10P(I),P10E(I)
  105 CONTINUE
  550 FORMAT(13X,5(3(F5.2,2X),3X))
      DC 100 I = 1,5
      AP(I) = PPSUM(I) / 100.
      AE(I) = PESUM(I) / 100.
      SP(I) = SQRT((PPSSQ(I) - (PPSUM(I)**2)/100.)/99.)
      SE(I) = SQRT((PESSQ(I) - (PESUM(I)**2)/100.)/99.)
  100 CCNTINUE
      WRITE(6,600) (AP(I),AE(I),I = 1,5)
  600 FORMAT(1X//14X,'AVG',3X,4(F5.2,2X,F5.2,12X),F5.2,2X,
     1F5.2)
      WRITE(6,650)(SP(J),SE(J),J=1,5)
  650 FCRMAT(14X,'S.D.',2X,4(F5.2,2X,F5.2,12X),F5.2,2X,
     1F5.2)
      STOP
      END
```

DATA

KNCWN DISTRIBUTION

0.00	21.0
0.04	17.0
0.08	15.7
0.12	14.8
0.16	14.0
0.20	13.1
0.24	12.7
0.28	12.4
0.32	12.1
0.36	11.9
0.40	11.6
0.44	11.3
0.48	11.0
0.52	10.6
0.56	10.3
0.60	10.0
0.64	9.8
0.68	9.4
0.72	9.0
0.76	8.6
0.8	8.0
0.84	7.5
0.88	6.6
0.92	5.0
0.96	2.5
1.00	0.0

APPENDIX B

COMPUTER SIMULATION WITH JACKKNIFED CONFIDENCE INTERVALS

The structure of this simulation is similar to the one described in Appendix A. Differences are discussed below.

Margin codes:

CODE	MEANING
C	Comments
N	These statements control the number of pseudo-values which result
O	These cards control the number of observations omitted from each jackknifed estimate
I	Statements which control which observations are included in each jackknife
NEX	Controls the number of exercises to be simulated
COR	Statements to which correlation correction factors can be added

ARRAY NAMES	PURPOSE
FREQ	Accumulates the number of times the confidence intervals cover each known percentile
AVDF	Contains the average number of degrees of freedom for each percentile estimate
CDIST	Contains the average confidence interval width for each percentile

ZSM, GSM	Contains values of the PL estimate (GSM) at corresponding ranges (ZSM) after the estimate has been smoothed
Z1, D1	Contains values of $Z_{(i)}$ and $D_{(i)}$ after observations have been omitted for the jackknife
THETA	The pseudo-values
THE0, THEI	The values used in computing the pseudovalue: THE0 contains the percentile value when all $Z_{(i)}$'s are used in the estimate, THEI the value when observations are omitted.
SUM, SQSUM	Accumulates the sums and sums of squares of the pseudo-values
CL, CU	Lower and upper confidence intervals for each percentile
DF	Contains the degrees of freedom used to estimate each percentile for each exercise

For each exercise, output includes a table similar to Table III. After all exercises have been simulated, confidence interval coverage frequency, average confidence interval widths, and average degree of freedom used to determine the t statistic for each percentile are listed. For large numbers of exercises, printouts of the pseudo-value tables can be suppressed.

```
      CALL CVFLOW
      ISEED = 530408
ND1   INTEGER D(25),DELTA(25),DF(5),FREQ(5),D1(21)
      REAL RK(26),PK(26),Z(25),ZHOLD(25),GHAT(25),TILE(5),
     1THETA(5,25),THEO(5),THEI(5,25),SUM(5),ZSM(52),GSM(52),
NZ1  2SQSUM(5),SSTAT(5),T(24),CL(5),CU(5),Z1(21),AVDF(5),
     3KTILE(5),CDIST(5)

C  READ IN KNOWN PROBABILITY OF DETECTION DISTRIBUTION
C  AND PERCENTILES OF INTEREST

      READ(5,100)(TILE(I),KTILE(I),I=1,5)
  100 FORMAT(2F10.5)
      READ(5,100)(PK(I),RK(I),I=1,26)
      READ(5,200)(T(I),I=1,24)
  200 FORMAT(F10.5)
      DO 5 I = 1,5
      FREQ(I) = 0
      AVDF(I) = 0.0
      CDIST(I) = 0.0
    5 CONTINUE
NEX   DO 175 N1 = 1,5

C  GENERATE RANDOM VARIATES

      DO 15 I = 1,25
      CALL RANDOM(ISEED,U,1)
      X = 20. * U
      CALL RANDOM(ISEED,U,1)
      DO 10 J = 1,26
      IF(U.GT.PK(J)) GO TO 10
      R = RK(J) + (PK(J)-U)*(RK(J-1)-RK(J))/(PK(J)-PK(J-1))
      GO TO 11
   10 CONTINUE
   11 IF(R.GE.X) GO TO 12
      Z(I) = X
      D(I) = 0
      GO TO 15
   12 Z(I) = R
      D(I) = 1
   15 CONTINUE

C  COPY Z ARRAY INTO Z HOLD, D ARRAY INTO DELTA

      DO 20 I = 1,25
      ZHOLD(I) = Z(I)
      DELTA(I) = D(I)
   20 CONTINUE

C  SORT Z ARRAY

      CALL PXSORT(Z,1,25)

C  MATCH DELTAS TO SORTED ARRAY

      DO 30 I = 1,25
      DO 25 J = 1,25
      IF(Z(I).NE.ZHOLD(J)) GO TO 25
      DELTA(I) = D(J)
      GO TO 30
   25 CONTINUE
   30 CONTINUE

C  COMPUTE PRODUCT LIMIT ESTIMATE FOR SELECTED PERCENTILES

      GHAT(25) = 0.0
      DO 35 J = 1,24
      K = 26 - J
      L = 25 - J
      XI = FLOAT(DELTA(K))/FLOAT(K)
      GHAT(L) = XI + (1. - XI)*GHAT(K)
   35 CONTINUE
```

```
C       SMOOTHING THE ESTIMATE

        ZSM(1) = 0.0
        GSM(1) = 1.0
        ZSM(2) = Z(1)/2.
        GSM(2) = 1.0
        ZSM(3) = Z(1)
        GSM(3) = (1. + GHAT(1))/2.
        DO 40 I = 1,24
        J1 = (2*I)+2
        J2 = J1 + 1
        J = I + 1
        ZSM(J1) = (Z(I)+Z(J))/2.
        GSM(J1) = GHAT(I)
        ZSM(J2) = Z(J)
        GSM(J2) = (GHAT(I)+GHAT(J))/2.
   40   CONTINUE
        ZSM(52) = Z(25) + (ZSM(51)-ZSM(50))
        GSM(52) = 0.0

C       FIND SELECTED PERCENTILES

        DO 50 J = 1,5
        DO 45 K = 1,52
        IF(GSM(K).GT.TILE(J)) GO TO 45
        A = (TILE(J)-GSM(K))*(ZSM(K)-ZSM(K-1))
        THEO(J) = ZSM(K) - A/(GSM(K-1)-GSM(K))
        GO TO 50
   45   CONTINUE
   50   CONTINUE

C  COMPUTE PRODUCT LIMIT ESTIMATE FOR SELECTED PERCENTILES
C  WITH 'INCOMPLETE' DATA

N       DO 95 M = 1,8

C,O     OMIT 3 ELEMENTS PER RUN FROM DATA SET

O       L = (3*M)+1
O       K = L-3
I       DO 55 I = 1,21
        IF(I.EQ.K) GO TO 56
        Z1(I) = ZHOLD(I)
   55   CONTINUE
I  56   DO 60 J = I,21
I       IF(J.GT.21) GO TO 61
        Z1(J) = ZHOLD(L)
        L = L+1
   60   CONTINUE

C       SORT Z1 ARRAY AND MATCH DELTAS

I  61   CALL PXSORT(Z1,1,21)
I       DO 70 I = 1,21
        DO 65 J = 1,25
        IF(Z1(I).NE.ZHOLD(J)) GO TO 65
        D1(I) = D(J)
   65   CONTINUE
   70   CONTINUE

C       COMPUTE THE PRODUCT LIMIT ESTIMATE ('INCOMPLETE' DATA)

I       GHAT(21) = 0.0
I       DO 75 J = 1,20
I       K = 22 - J
I       L = 21 - J
        XI = FLOAT(D1(K))/FLOAT(K)
        GHAT(L) = XI + (1.-XI)*GHAT(K)
   75   CONTINUE

C       SMOOTHING THE ESTIMATE
```

```
      ZSM(2) = Z1(1)/2.
      ZSM(3) = Z1(1)
      GSM(3) = (1. + GHAT(1))/2.
I     DO 80 I = 1,20
      J1 = (2*I)+2
      J2 = J1+1
      J = I + 1
      ZSM(J1) = (Z1(I)+Z1(J))/2.
      GSM(J1) = GHAT(I)
      ZSM(J2) = Z1(J)
      GSM(J2) = (GHAT(I)+GHAT(J))/2.
   80 CONTINUE
I     ZSM(44) = Z1(21) + (ZSM(43)-ZSM(42))
I     GSM(44) = 0.0

C     FIND SELECTED PERCENTILES

      DO 90 J = 1,5
I     DO 85 K = 1,44
      IF(GSM(K).GT.TILE(J)) GO TO 85
      A = (TILE(J)-GSM(K))*(ZSM(K)-ZSM(K-1))
      THEI(J,M) = ZSM(K) - A/(GSM(K-1)-GSM(K))
      GO TO 90
   85 CONTINUE
   90 CONTINUE
   95 CONTINUE

C  PERFORM JACKKNIFE

      DO 105 I = 1,5
      DF(I) = 0
      SUM(I) = 0.0
      SQSUM(I) = 0.0
  105 CONTINUE
      DO 125 I = 1,5
N     DO 115 J = 1,8
N     THETA(I,J) = 8.*THEO(I) - 7.*THEI(I,J)

C     GATHER STATISTICS

      SUM(I) = SUM(I) + THETA(I,J)
      SQSUM(I) = SQSUM(I) + THETA(I,J)**2
  115 CONTINUE
  125 CONTINUE

C     COMPUTE DEGREES OF FREEDOM

      DO 155 I = 1,5
N     DO 145 J = 1,8
      K = J+1
N     DO 135 L = K,8
      IF(THETA(I,J).EQ.THETA(I,L)) GO TO 145
  135 CONTINUE
      DF(I) = DF(I) + 1
  145 CONTINUE
      DF(I) = DF(I) - 1
  155 CONTINUE

C  COMPUTE 95% CONFIDENCE INTERVALS

      DO 165 I = 1,5
N     GHAT(I) = SUM(I)/8.
N     SSTAT(I) = SQRT((SQSUM(I)-(SUM(I)**2)/8.)/7.)/SQRT(8.)
COR   CL(I) = GHAT(I) - T(DF(I))*SSTAT(I)
      IF(CL(I).LT.0.0) CL(I) = 0.0
COR   CU(I) = GHAT(I) + T(DF(I))*SSTAT(I)
      AVDF(I) = AVDF(I) + DF(I)
      CDIST(I) = CDIST(I) + CU(I) - CL(I)
      IF((CL(I).GT.KTILE(I)).OR.(KTILE(I).GT.CU(I)))GOTO 165
      FREQ(I) = FREQ(I) + 1
  165 CONTINUE
```

```
C   PRINT RESULTS
      WRITE(6,300)
  300 FORMAT(1H1////,36X,'PSEUDO-VALUES'//,21X,'90',8X,'75',
     18X,'50',8X,'25',8X,'10'/)
N     WRITE(6,350)((THETA(I,J),I=1,5),J=1,8)
  350 FORMAT(1H ,15X,5F10.5)
      WRITE(6,400)
  400 FORMAT(1H0,37X,'ESTIMATES')
      WRITE(6,450)(GHAT(I),I=1,5)
  450 FORMAT(1H0,15X,5F10.5/)
      WRITE(6,500)
  500 FORMAT(1H0,37X,'S/SQRT(N)')
      WRITE(6,450)(SSTAT(I),I=1,5)
      WRITE(6,550)
  550 FORMAT(1H0,23X,'T STATISTIC  /  (DEGREES OF FREEDOM)')
      WRITE(6,450)(T(DF(I)),I=1,5)
      WRITE(6,650)(DF(I),I=1,5)
  650 FORMAT(1H ,15X,5(4X,'(',I2,')',2X))
      WRITE(6,700)
  700 FORMAT(1H0,29X,'95% CONFIDENCE INTERVALS'/)
      WRITE(6,350)(CL(I),I=1,5)
      WRITE(6,350)(CU(I),I=1,5)
  175 CONTINUE
      WRITE(6,850)(FREQ(I),I=1,5)
  850 FORMAT(1H0,37X,'FREQUENCY'//,15X,5(4X,I3,3X))
      DO 185 I = 1,5
      CDIST(I) = CDIST(I)/FLOAT(N1)
      AVDF(I) = AVDF(I)/FLOAT(N1)
  185 CONTINUE
      WRITE(6,900)(CDIST(I),I=1,5)
  900 FORMAT(1H0,32X,'AVERAGE C.I. WIDTH'//,15X,5F10.5)
      WRITE(6,950)(AVDF(I),I=1,5)
  950 FORMAT(1H0,28X,'AVERAGE DEGREES OF FREEDOM'//,15X,
     15F10.5)
      STOP
      END
```

DATA

PERCENTILES OF INTEREST

0.90	6.15
0.75	8.7
0.50	10.85
0.25	12.6
0.10	15.05

KNOWN DISTRIBUTION

0.00	21.0
0.04	17.0
0.08	15.7
0.12	14.8
0.16	14.0
0.20	13.1
0.24	12.7
0.28	12.4
0.32	12.1
0.36	11.9
0.40	11.6
0.44	11.3
0.48	11.0
0.52	10.6
0.56	10.3
0.60	10.0
0.64	9.8
0.68	9.4
0.72	9.0

0.76	8.6
0.8	8.0
0.84	7.5
0.88	6.6
0.92	5.0
0.96	2.5
1.00	0.0

VALUES OF T STATISTIC (95%, 2-SIDED TEST)

12.706
4.303
3.182
2.776
2.571
2.447
2.365
2.306
2.262
2.228
2.201
2.179
2.160
2.145
2.131
2.120
2.110
2.101
2.093
2.086
2.080
2.074
2.069
2.064

LIST OF REFERENCES

1. Operations Research, Inc. Technical Report 798, Data Base for ASW Analyses (U), by G.H. Henry, Secret, December 1973.

2. Naval Underwater Systems Center Technical Memorandum PA2-4362-75, Statistical Analysis of Sonar Detection Ranges (U), by J.G. Kelly, Confidential, 5 August 1975.

3. Destroyer Development Group Technical Report 15-73, Destroyer Analysis Notebook (U), Confidential, 31 August 1973.

4. Submarine Development Group Two Report 1-73, Submarine Analysis Notebook (U), Confidential, September 1973.

5. Fisher, L. and Kanarek, P., "Presenting Censored Survival Data When Censoring and Survival Times May Not Be Independent," Reliability and Biometry - Statistical Analysis of Lifelength, edited by F. Proschan and R. Serfling, Society for Industrial and Applied Mathematics, 1974.

6. Kaplan, E.L. and Meier, P., "Nonparametric Estimation from Incomplete Observations," Journal of the American Statistical Association, v. 53, June 1958.

7. Wilk, M.B. and Gnanadesikan, R., "Probability Plotting Methods for the Analysis of Data," Biometrika, v. 55, 1, 1968.

8. Stanford University Department of Biostatics Technical Report No. 14, Jackknifing Censored Data, by R.G. Miller, 9 October 1975.

9. Mosteller, F., and Tukey, J., "Data Analysis, Including Statistics," Chapter 10 in The Handbook of Social Psychology, Vol. II, 2nd Edition, Addison-Wesley, 1968.

10. Gray, H.L. and Schucany, W.R., The Generalized Jackknife Statistic, Marcel Dekker, Inc., 95 Madison Ave., New York, N.Y. 1972.

11. Walsh, J.E., "Concerning the Effect of Intraclass Correlation on Certain Significance Tests," Annals of Mathematical Statistics, Vol. 18, 1947.

12. Gaver, D.P., Methods for Assessing Variability, with Emphasis on Simulation Data Interpretation, paper presented at International Symposium on Unvertainty in Hydrologic and Water Resource Systems, University of Arizona, Tucson, December 1972.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3.	Department Chairman, Code 55 Operations Research and Administrative Sciences Department Naval Postgraduate School Monterey, California 93940	1
4.	Professor Donald P. Gaver, Code 55Gv Operations Research and Administrative Sciences Department Naval Postgraduate School Monterey, California 93940	4
5.	Assoc. Professor R. Neagle Forrest, Code 55Fo Operations Research and Administrative Sciences Department Naval Postgraduate School Monterey, California 93940	1
6.	Commander-in-Chief, Pacific Attn: Review and Analysis Office Box 13, J021 FPO, San Francisco, California 96610	1
7.	Mr. James G. Kelly, Code PA2 Naval Underwater Systems Center Newport, Rhode Island 02840	1
8.	LT Ronald A. Route, USN 4301 Reed Street Wheat Ridge, Colorado 80033	1

~~Distribution Limited to U. S. Government Agencies Only; Test and Evaluation; March 1976. Other requests for this document must be referred to the Naval Postgraduate School, Monterey, California 93940, Code 023.~~

DOWNGRADED
APPROVED FOR PUBLIC RELEASE

thesR814
An investigation of the applic[redacted] of
[redacted]2768 001 00001 1
DUDLEY KNOX LIBRARY

DUDLEY KNOX LIBRARY - RESEARCH REPORTS
5 6853 01060220 4