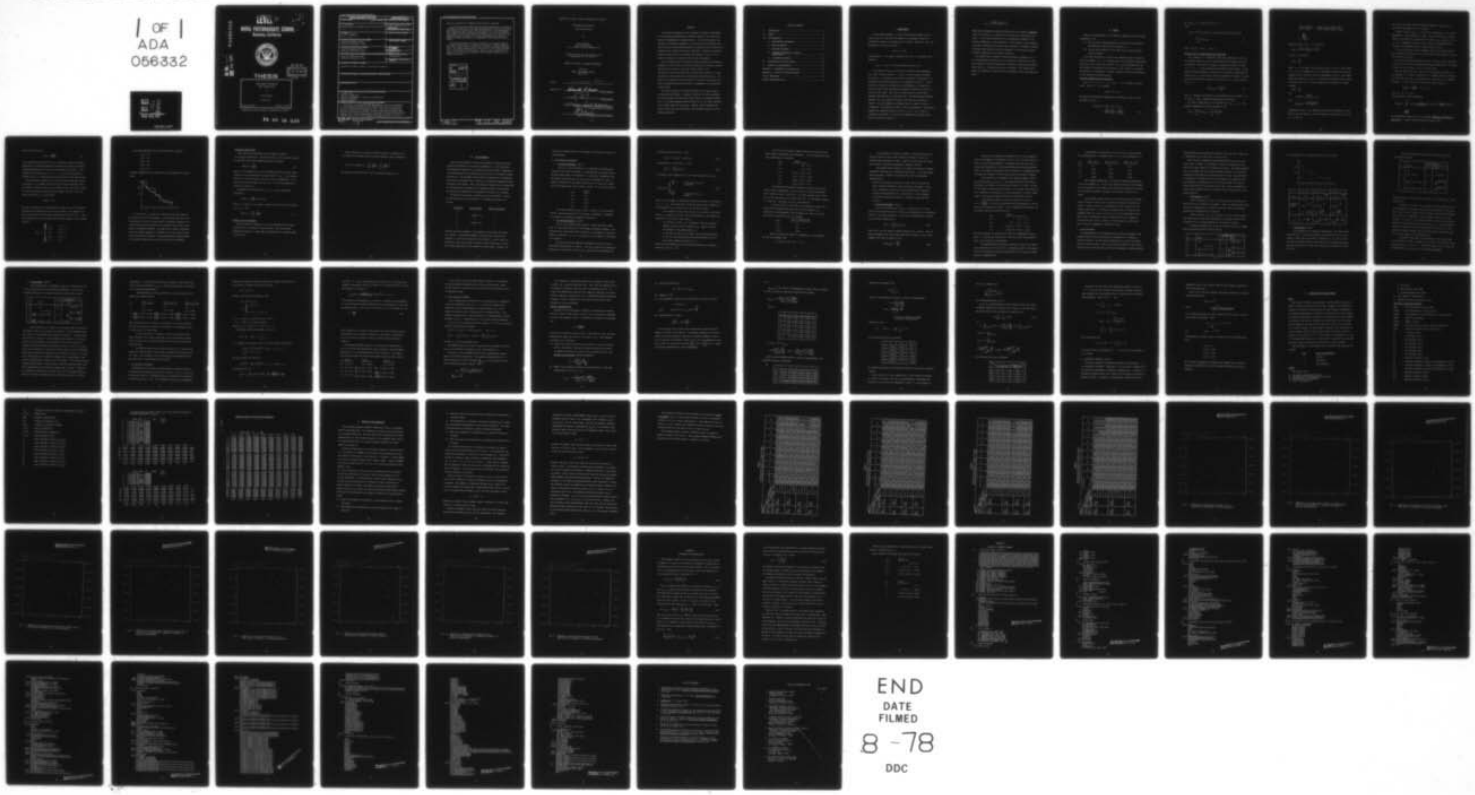
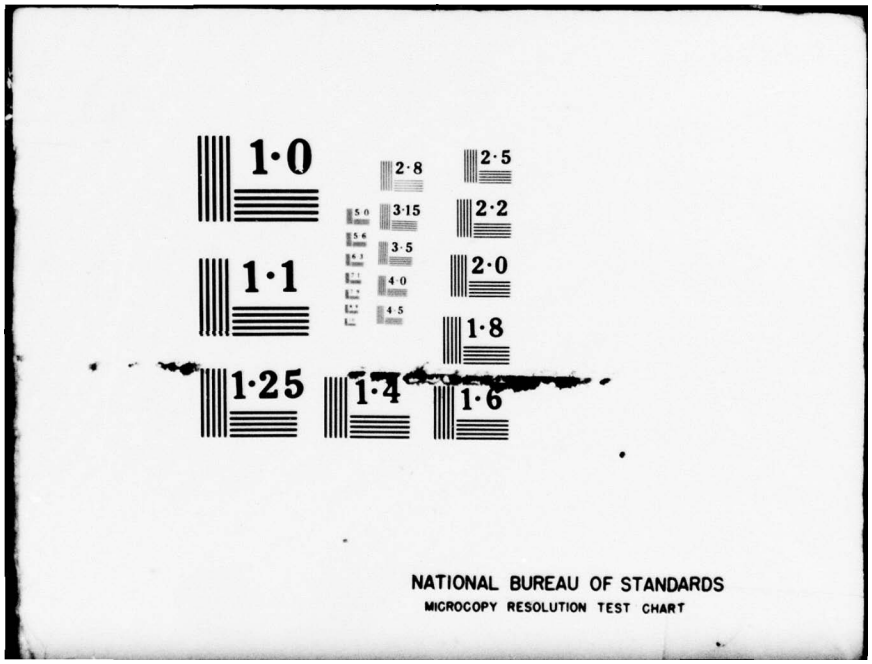
AD-A056 332 NAVAL POSTGRADUATE SCHOOL MONTEREY CALIF F/G 14/4
NONPARAMETRIC ESTIMATION FROM CENSORED DATA.(U)
MAR 78 L W HYUNG

UNCLASSIFIED NL

1 OF 1
ADA
056332



END
DATE
FILMED
8-78
DDC


1·0
1·1
1·25
1·4
1·6
1·8
2·0
2·2
2·5
2·8
3·15
3·5
4·0
4·5
5·0
5·6
NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

AD A056332
AD No.
DDC FILE COPY

LEVEL II

NAVAL POSTGRADUATE SCHOOL

Monterey, California

DDC
RECEIVED
JUL 18 1978
D

THESIS

NONPARAMETRIC ESTIMATION
FROM CENSORED DATA

by

Lee Won Hyung

March 1978

Thesis Advisor: Donald P. Gaver

Approved for public release; distribution unlimited.

78 07 10 020

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)
Nonparametric Estimation from Censored Data

5. TYPE OF REPORT & PERIOD COVERED
Master's Thesis;
March 1978

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Lee Won Hyung

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 1978

13. NUMBER OF PAGES
75

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)
Naval Postgraduate School
Monterey, California 93940

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)
Censored data
Survival probability
Jackknife estimators of logistic transformation
Bayesian estimators
Simulation model

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

For nearly two decades we have witnessed an intensive development of a statistical methodology for assessing length of life and reliability of performance from empirical data. The initial stimulus for research on statistical problems in life testing and reliability came from the need to answer pressing practical questions which could not be treated by the existing statistical techniques. Because life and performance tests are so time consuming and expensive to run, it is a

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EDITION OF 1 NOV 65 IS OBSOLETE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

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

251454

practical necessity to terminate them as soon as possible.

For the statistician this means developing estimation and decision procedure for data, which are severely curtailed in one way or another long before all items on test have actually failed. The estimation is more complicated when the data are truncated, i.e. when the observer loses track of some individuals before death occur. The product limit method of Kaplan and Meier is one way of estimating p(t) when the mechanism causing truncation is independent of the mechanism causing death.

This paper proposes alternative estimators and compares them to the product limit method. A computer simulation is used to generate the times of death and truncation from a variety of assumed distributions. No single estimator gives the best fit to the "true" distribution of death under all situations. However, other estimators are shown to be better than the product limit estimator under all of the assumed situations.

ACCESSION for
NTIS White Section
DDC Buff Section
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
Dist. AVAIL. and/or SPECIAL
A

Approved for public release; distribution unlimited.

Nonparametric Estimation from Censored Data

by

Lee Won Hyung
Major, Korean Army
B.S., Korean Military Academy, 1970

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
March 1978

Author Won Hyung Lee

Approved by: Donald P. Gaver
Thesis Advisor

Second Reader

MG Sovereign (OK Hartman)
Chairman, Department of Operations Research
and Administrative Sciences

Dean of Information and Policy Sciences

ABSTRACT

For nearly two decades we have witnessed an intensive development of a statistical methodology for assessing length of life and reliability of performance from empirical data. The initial stimulus for research on statistical problems in life testing and reliability came from the need to answer pressing practical questions which could not be treated by the existing statistical techniques. Because life and performance tests are so time consuming and expensive to run, it is a practical necessity to terminate them as soon as possible.

For the statistician this means developing estimation and decision procedure for data, which are severely curtailed in one way or another long before all items on test have actually failed. The estimation is more complicated when the data are truncated, i.e. when the observer loses track of some individuals before death occur. The product limit method of Kaplan and Meier is one way of estimating p(t) when the mechanism causing truncation is independent of the mechanism causing death.

This paper proposes alternative estimators and compares them to the product limit method. A computer simulation is used to generate the times of death and truncation from a variety of assumed distributions. No single estimator gives the best fit to the "true" distribution of death under all situations. However, other estimators are shown to be better than the product limit estimator under all of the assumed situations.

TABLE OF CONTENTS

I. INTRODUCTION - 6

II. THEORY - 8

III. THE ESTIMATORS - 16

A. STEP-FUNCTION ESTIMATORS - - - - - - - - - - - - - - - - 17

B. POINT ESTIMATORS - 22

C. BAYESIAN ESTIMATORS - - - - - - - - - - - - - - - - - - 27

D. JACKKNIFE ESTIMATORS OF LOGISTIC TRANSFORMATION - 30

E. A PARAMETER ESTIMATOR - - - - - - - - - - - - - - - - - 36

IV. INSTRUCTIONS FOR USING PROGRAM - - - - - - - - - - - - - - 38

V. RESULTS OF THE SIMULATION - - - - - - - - - - - - - - - - - 43

APPENDIX A. ESTIMATORS FOR GROUPED DATA - - - - - - - - - - - - 60

APPENDIX B. LISTING OF COMPUTER PROGRAM - - - - - - - - - - - - 63

LIST OF REFERENCES - 74

INITIAL DISTRIBUTION LIST - - - - - - - - - - - - - - - - - - - 75

I. INTRODUCTION

Let the random variable T denote the time that elapses until an event occurs; the event may for example be an equipment failure, an individual's death, or the detection of a target. Denote by $p(t)$ the probability of survival to time t,

$$p\{T > t\} = p(t)$$

Picturesquely, T is called a lifetime, and $p(t)$ is a survival probability;

$F(t) = 1 - p(t)$ is the distribution function of T.

In the medical field, one might wish to estimate the probability, $p(t)$ that a patient survives t after a certain surgical procedure for cancer. In electronics, one wishes to estimate the probability of continuous failure-free operation of an equipment for time t. In the military, one might be interested in the probability of conducting a certain mission, under specified environmental conditions, without detection by the enemy. The event of interest may be a human death, equipment malfunctions, or sonar detection. Following Kaplan and Meier, Reference (1), this paper will refer to the event of interest as a "death". The test element in the sample may be a human, a radio, or a submarine. This paper will refer to the test elements as "individuals". Suppose that observed values of T are $t_1, t_2, t_3, \ldots t_N$, so that N lifetimes are observed. In this case an appropriate (unbiased) estimates of survival to time t is

$$\tilde{p}(t) = \frac{\text{number of } t_i\text{'s} > t}{N}$$

Under many circumstances complete lifetimes are not observed; censoring occurs at certains, x_i , beyond which the life of an individual is not known. In such cases construction of an appropriate estimate of the survival probability is more difficult. In this paper various estimates of survival probability are studied when lifetimes are randomly censored. This means that censoring times are assumed to be realizations random variables independent of the actual lifetimes.

The product-limit estimator of Kaplan and Meier, Reference (1), is an accepted method of dealing with the problem of censored data. This paper presents thirteen non-parametric estimators, including the product limit function. Censored data sets are simulated. The thirteen estimators are compared by examining their performance on the simulated data bases.

II. THEORY

There are two approaches to the empirical estimation of the survival probability, $p(t)$:

(1) one may use the observed fraction of survivors at arbitrarily selected times (step function estimator), or

(2) one may focus attention on the times of the observed deaths (point estimator).

The initial discussion is based on the assumption that all observations are complete, i.e., it is assumed that all individuals remain under observation until their time of death. This initial assumption is for the purpose of simplifying the discussion. Then, later in this paper, the discussion is broadened to include incomplete data with observations of both death and censoring events.

Survival Probabilities; No Censoring

Let $0 = t_0 < t_1 < t_2 \ldots < t_i < t_{i+1} < \ldots$ be a sequence of fixed times. Then if T is a lifetime

$$p(t_i) = p\{T > t_i\}$$

and denote the conditional probability of survival to time t_i, given survival to t_{i+1} by

$$p(t_i|t_{i-1}) = p\{T > t_i | T > t_{i-1}\}$$
$$= \frac{p\{T > t_i\}}{p\{T > t_{i-1}\}} = \frac{p(t_i)}{p(t_{i-1})} \qquad (1)$$

If $p(t_{i-1}) = 0$, define $p(t_i|t_{i-1}) = 0$.

Then

$$p(t_i) = p(t_i|t_{i-1})p(t_{i-1}) = p(t_i|t_{i-1})\cdot p(t_{i-1}|t_{i-2})p(t_{i-2})$$

$$= \prod_{j=1}^{i} p(t_j|t_{j-1}) \qquad (2)$$

where $p(t_i|t_o) = p(t_i)$; $p(o) = 1$.

Observations on Uncensored Data at Fixed Times

Let a sample of N individuals come under observation. They are all observed from birth (or the appropriate event defining time zero) until death. With the first approach, preselects a series of times, $0 < t_1 < t_2 < \ldots$ before examining the observed time of death. In the medical follow-up example, one might select the times corresponding to exactly 1,2,3,... years after a surgical procedure for cancer. An estimate of the conditional probability of survival to t_i, given survival to t_{i-1} is

$$\tilde{p}(t_i|t_{i-1}) = \frac{N_i - r_i}{N_i} \qquad (3)$$

With N_i elements were present at the beginning of the interval, i.e., at time t_{i-1}, and r_i elements failed during the interval.

For a set of data which is not censored, $N_i = N_{i-1} - r_{i-1}$. Now replace probabilities by their estimates in (2):

$$\tilde{p}(t_i) = \prod_{j=1}^{i} \tilde{p}(t_j|t_{j-1}) = \prod_{j=1}^{i} \left(\frac{N_j - r_j}{N_j}\right)$$

$$= \left(\frac{N-r_1}{N}\right)\left(\frac{N-r_1-r_2}{N-r_1}\right) \text{ --- } \left(\frac{N-r_1-\ldots-r_{i-1}}{N-r_1-\ldots-r_{i-2}}\right)\left(\frac{N-r_1-\ldots-r_i}{N-r_1-\ldots-r_{i-1}}\right)$$

$$= 1 - \frac{\sum_{j=1}^{i} r_j}{N}$$

Now the estimate $p(t_i)$ is of the form

$$\tilde{p}(t_i) = \frac{N-(r_1 + r_2 + \ldots + r_i)}{N} ,$$

and this is the same as

$$\tilde{p}(t_i) = \frac{S_i}{N}$$

where S_i is the number of the original group, of size N, that survive to t_i. If it is assumed that the N individuals each have the survival probability p(t), and that they die independently, then S_i, the random number that survive to time t_i is binomially distributed, with S_i being a realized value of S_i. Then, considering the estimate as a random variable,

$$\tilde{p}(t_i) = \frac{S_i}{N}$$

and

$$E[\tilde{p}(t_i)] = \frac{N\, p(t_i)}{N} = p(t_i)$$

and

$$Var[\tilde{p}(t_i)] = \frac{p(t_i)(1-p(t_i))}{N}$$

Consequently $\tilde{p}(t_i)$ is an unbiased and consistent estimate of $p(t_i)$. This is true for every t_i, and can be shown to be true for all t_i, i=1, 2,...I, as will.

All of this indicates that the estimate suggested is likely to be a good one if the sample size, N, is large.

Clearly $\tilde{p}(t_i) \leq \tilde{p}(t_{i-1})$. The survival probability, p(t), is thus estimated at a fixed sequency of times. At each time point, t_i being a typical one, there are r_i fewer survivors than at t_{i-1}, where $r_i = 0,1,2,\ldots,N$. Consequently a plot of $\tilde{p}(t_i)$ shows a non-decreasing step function, with downward steps of varying sizes at $t_1, t_2, \ldots$.

If the above times are close together, and if the time of death T, has a density function, then one can anticipate seeing values of r_i that are either zero or unity.

The so-called second approach is really a limiting case of the first, as the time of intervals of measurement decrease indefinitely. Thus when a death (or loss) occurs it is only a single event.

When no losses take place, the case now considered, the time t_i of the ith death is a really a realization of a random variable, denoted by $\underline{t}_i$; this means that $p(\underline{t}_i)$ the probability of surviving $\underline{t}_i$, is a random variable. It can be shown that the expected value of $p(\underline{t}_i)$ is

$$E[p(\underline{t}_i)] = \frac{N-i+1}{N+1} \quad , \quad i=1,2,\ldots,N$$

where $\underline{t}_1 < \underline{t}_2 < \text{---} < \underline{t}_N$.

The derivation involves integrating

$$E[p(\underline{t}_i)] = \int_0^\infty p(t) \cdot \frac{N!}{(i-1)!(N-i+1)!} [1-p(t)]^{i-1} \left(-\frac{dp(t)}{dt}\right) [p(t)]^{N-i}$$

$$= \frac{N-i+1}{N+1}$$

by transformation from p(t) to x; see Cramér, Mathematical Methods of Statistics, H. Cramér, Princeton University Press, 1946.

Thus one is led to use

$$\tilde{p}(t_i) = \frac{N-i+1}{N+1} \qquad (4)$$

as an estimate of the value of $p(t_i)$, t_i being the ith time of death. Expression (4) provides estimator of the survival function at times of observed deaths when there are no losses because of censoring. The estimator at the points t_i: $t_i < t_2 < \text{---} < t_N$, can be connected by straight lines, or a step function with step sizes $1/(N+1)$ may be used.

The estimators of equation (4) give intuitively acceptable results. For example, if the sample consists of only a single individual (N=1), then death is equally likely to occur before or after the time at which the true (but unknown) survival function equals one half. Thus, the result of equation (4) is reasonable:

$$E[\tilde{p}(t_1)] = \frac{1}{2}$$

The point estimates of the second approach always occur at the times of discontinuity forestimates from the first approach. For example, consider a data base (N=4) with deaths observed at times 1,3,4 and 7. The first approach gives the following step function estimate of the survival function:

$$\tilde{p}(t) = \begin{cases} 1.0 & 0 \leq t < 1 \\ 0.75 & 1 \leq t < 3 \\ 0.5 & 3 \leq t < 4 \\ 0.25 & 4 \leq t < 7 \\ 0.0 & t < 7 \end{cases}$$

The second approach gives the following point estimates

$p(o) = 1.0$

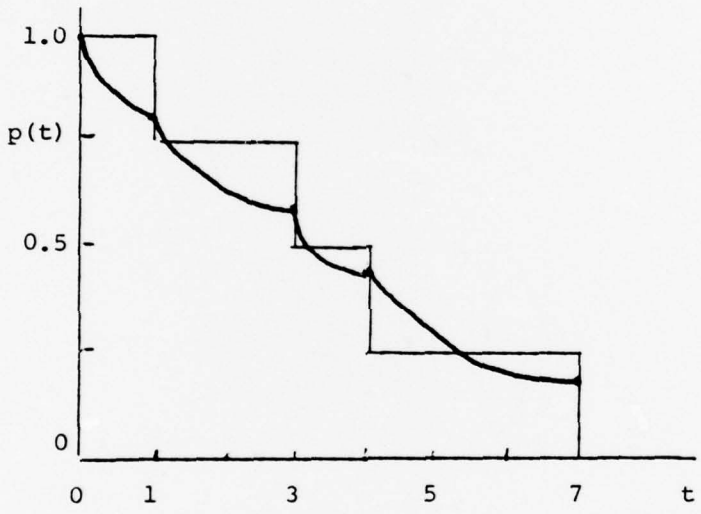
$p(1) = 0.8$

$p(3) = 0.6$

$p(4) = 0.4$

$p(7) = 0.2$

A graphic comparison of the results of the two approaches is given below:



It is difficult to decide how to smooth out the step functions that result from the first approach. By connecting the tops of the "stairsteps," one places an upper bound on reasonable estimates. By connecting the bottom corners of the stairsteps, one places a lower bound on reasonable estimates. One might draw a smooth, decreasing curve that passes through all (or almost all) of the vertical faces of the step-function estimate. The second approach suggests method of selecting a unique point on each of these vertical segments.

Incomplete observations

When some of the observations are incomplete, equation (4) requires modification. The expected value of the survival function at the time of the first observed death may be written:

$$E[\tilde{p}(t_1)] = \frac{N_1}{N_1+1} \tag{5}$$

Here N_1 is the effective size of the sample during the interval terminated by the time observed for the first death (o,t_1). In the special case of no censoring events, the value of N_1 is unambiguous. It is equal to the initial sample size ($N_1 = N$). In this case equation (5) reduces to equation (4).

Subsequent point estimates for t_2, t_3,... may be calculated iteratively:

$$E[\tilde{p}(t_i)] = \frac{N_i}{N_i+1} \cdot E[p(t_{i-1})]$$

where $t_o = 0$ and N_i is the effective sample size over the time interval (t_{i-1}, t_i). Thus,

$$E[\tilde{p}(t_i)] = \prod_{j=1}^{i} \left(\frac{N_j}{N_j+1}\right) \tag{6}$$

Variance of the estimators

Kaplan and Meier, reference (1), give an expression for the exact calculation of the variance of step functions. They also discuss "Greenwood's formula," a large sample approximation that ignores terms of order $1/N_i^2$.

Herd, reference (2), presence without derivation an expression for the variance of estimates using the second approach (point estimators):

$$V(t_i) = \text{Var}\{E[\tilde{p}(t_i)]\} = \prod_{j=1}^{i} \left(\frac{N_j}{N_j+2}\right) - \prod_{j=1}^{i} \left(\frac{N_j}{N_j+1}\right)^2$$

The notation here follows that for the estimating equation (6).

III. THE ESTIMATORS

This section describes the nine non-parametric estimators and four jackknife estimators of the survival probability. It also describes the parametric estimator for an exponential decay function. Exponential life distributions are the starting point for much of reliability theory and practice. The estimator derived from the exponential is regarded as "par" when the simulated data is based on an underlying exponential decay distribution for deaths. Thus, when deaths are exponentially distributed, the non-parametric estimators may be compared relative to each other, and they may be compared with the parametric estimator as a standard.

A hypothetical data base, consisting of five individuals, is used to illustrate each of the estimators. This sample data base is as follows:

Individual	Time of Death	Time of Truncation
A	1	-
B	Unknown (>2)	2
C	3	-
D	Unknown (>6)	6
E	7	-

The data have been arranged in time sequence of the death and truncation events. In the medical example, the data might indicate that patients A, C and E were observed to die exactly 1, 3 and 7 years, respectively, after their surgery. However, B and D moved away or otherwise became unavailable to the observer at these times. Further, the

cause of the unobservability is unrelated to the patient's health and life expectancy.

A. STEP-FUNCTION ESTIMATORS

1. The First Estimator, "$\tilde{p}_1(t)$"

$\tilde{p}_1(t)$ is a naive estimator; it is expected to perform poorly relative to the other estimators. $\tilde{p}_1$ only depends on the data from individuals whose deaths are observed. It ignores any information from the partial lifetimes noted for the censored observations. $\tilde{p}_1(t)$ is simply the fraction of individuals surviving to at least time t among those individuals whose time of death is known. It is a step function:

t	$p_1(t)$
0-1	1.0
1-3	0.667
3-7	0.333
7-∞	0.00

The naive estimator, $\tilde{p}_1(t)$, takes no account of the successful survival intervals observed for the censored individuals. Therefore it is biased in a downward (pessimistic) direction.

2. The Second Estimator, "$\tilde{p}_2(t)$"

$\tilde{p}_2(t)$ is the product-limit estimate. Kaplan and Meier, reference (1), have shown that this is the maximum likelihood estimator. The observed events, both deaths and truncations, are arranged in increasing order of occurrence: $t_1, t_2, \ldots, t_N$; where N is the number of individuals in the sample.

Let $p(t_i)$ denote the cumulative probability of survival of an individual from time zero to time t_i. Let $p(t|t_i)$ denote the conditional probability of surviving to time $t(> t_i)$, given that the individual has

already survived to time t_i. Then,

$$\tilde{p}_2(t_i) = p_2(t_{i-1}) \cdot p_2(t_i|t_{i-1}) \qquad \text{(E-1)}$$

If we define $t_0 = 0$ and $p(0) = 1$, then

$$\tilde{p}_2(t_i) = \prod_{j=1}^{i} p_2(t_j|t_{j-1}) \qquad \text{(E-2)}$$

The product limit estimator is in the form of equation (E-2) with

$$\tilde{p}_2(t_j|t_{j-1}) = \begin{cases} \frac{N_j}{N_j} = 1 & \text{If the event at } t_j \text{ is truncation} \\ \frac{N_j - 1}{N_j} & \text{If the event at } t_j \text{ is a death} \end{cases} \qquad \text{(E-3)}$$

Here n_j is the number of individuals observed surviving in the interval $t_{j-1} < t < t_j$. This formulation causes the product limit estimator to be insensitive to the exact time of the censoring events.

The estimator is unity from time zero to the time of the first event, t_1, reflecting the fact that all individuals in our example are observed to live until at least time t_1.

- If the event at time t_1 is a truncation, then the estimator remains at unity until at least time t_2. Again, no deaths are observed in the sample before t_2.
- If the event at time t_1 is a death, then the estimator drops to $(N-1)/N$. This drop reflects the observed death of $1/N$ of the survival sample just prior to t_1.

Values of the estimator $\tilde{p}_2$ are calculated iteratively at successive values of $t_i (i=1,2,\ldots,N)$.

The size of the survival sample declines as truncations and deaths remove individuals from observation. For the hypothetical data base listed above, one obtains:

t	$\tilde{p}_2(t)$
0-1	5/5 = 1.0
1-2	4/5 = 0.8
2-3	(4/5) x (3/3) = 0.8
3-6	(4/5) x (2/3) = 0.533
6-7	(8/15) x (1/1) = 0.533
7-∞	(8/15) x (0/1) = 0.0

The product-limit estimator explicitly accounts for the survival of these individuals (up to the time of the last death before each censoring event). Thus, $\tilde{p}_2(t)$ is a step function with a value that is not less than $\tilde{p}_1(t)$ for any value of t. If the sample contains no censoring, then $\tilde{p}_1(t)$ and $\tilde{p}_2(t)$ are identical.

If the last event in the sample is a truncation rather than a death, then the modified data give the following estimate, i.e., individual E had disappeared from the observer at time 6.5 (so that the fact of E's death at time 7 is unknown).

t	$\tilde{p}_2(t)$ - Modified data
0-1	1.0
1-3	0.8
3-6.5	0.533

Since the time of the death for individual E is now unknown, one can only estimate that:

$$0 \leq \tilde{p}_2(t) \leq 0.533 \quad \text{for} \quad t > 6.5$$

If the analyst is willing to assume a functional form for the survival function, then he may calculate the manner in which the estimator $\tilde{p}_2(t)$ decreases to zero. However, the data alone are insufficient when a strictly non-parametric estimator is used.

The product-limit estimator is a useful and intuitively appealing method of dealing with incomplete observations. It has been wider used and studied. However, the product-limit has one disturbing characteristic:

> Most of the biological, physical or other causes of deaths produce a survival probability that continuously decreases in time. It is, therefore, one may be a little uncomfortable estimating the survival probability with a step function. One is tempted to smooth the estimator to make it a monotonic decreasing function of t.

3. <u>The Third Estimator</u>, "$\tilde{p}_3(t)$"

$\tilde{p}_3(t)$ is a modification of $\tilde{p}_2(t)$. Like $\tilde{p}_2(t)$, it is a step function with discrete drops at those times corresponding to the observed deaths in the sample population. It may also be expressed as a product of conditional probabilities:

$$\tilde{p}_3(t_i) = \prod_{j=1}^{i} \tilde{p}_3(t_k | t_{k-1}) \qquad \text{(E-4)}$$

where the t_k are the times of observed deaths and t_o is zero. The conditional probabilities on the right-hand side of Equation (E-4) differ somewhat from those in Equation (E-2):

$$\tilde{p}_3(t_k | t_{k-1}) = \frac{N_k - 1}{N_k} \qquad \text{(E-5)}$$

Equation (E-5) differs from Equation (E-3) in the interpretation of the numbers of individuals at risk. Here, the value of N_k is taken to be the average number of individuals observed surviving in the interval between the (k-1)st observed death and the kth observed death. The number of observed survivors decrease at intermediate times if events are censored, and hence the N_k are not necessary integers.

The value of N_k is regarded as the effective sample size for the interval from t_{k-1} to t_k. In the sample data base shown above, individual B is known to have survived from time 1 to time 2, or half of the interval between the first death at t=1 and the second death at t=3. Therefore, the estimator p_3 treats individual B as half a participant in the interval between the death of individuals A and C.

The effective sample size for this interval is then 3.5 ($n = 3 + \frac{(2-1)}{(3-1)} = 3.5$) (full contributions from individuals C, D and E, plus a half contribution from B). For our hypothetical data base, the following values are calculated for $\tilde{p}_3$:

t	$\tilde{p}_3(t)$		
0-1		5/5	= 1.0
1-3	4.5	x 1.0	= 0.8
3-7	(2.5/3.5)	x 0.8	= 0.571
(7)	(1.75/2.75)	x 0.571	= 0.364

The value of $\tilde{p}_3(t)$ can never be less than the corresponding value of $\tilde{p}_2(t)$. In the special case with no censoring events the estimators $\tilde{p}_1(t)$, $\tilde{p}_2(t)$ and $\tilde{p}_3(t)$ are identical.

One might perturb the data by shifting the time of B's truncation event down to $1+\varepsilon$ or up to $3-\varepsilon$, ε arbitrarily small. The dependence of the estimator $\tilde{p}_3$ upon the exact time of the censoring events may now be demonstrated.

For purposes of illustration, the time of the censoring event for individual $B(t_2)$ is decreased from 2 to 1.1, then increased to 2.9.

t	$\tilde{p}_3(t)$, $t_2 = 2$	$\tilde{p}_3(t)$, $t_2 = 1.1$	$\tilde{p}_3(t)$, $t_2 = 2.9$
0-1	1.0	1.0	1.0
1-3	0.80	0.80	0.80
3-7	0.571	0.538	0.597
(7)	0.364	0.342	0.380

This example demonstrates an intuitively appealing characteristic of the estimator, $\tilde{p}_3$. As the total observed survival time increases for the individuals in our sample (with deaths held constant), the value of the estimating function increases over at least a portion of its range.

We may safely assume that the true survival function eventually tends to zero with time, since no physical or biological system lives forever. However, there are no observations on the survival of individuals beyond time 7. The data only indicate that our step-function estimator drops to a value of .364 at t=7, but the nonparametric estimator gives no information about the survival function's subsequent decline from .364 to zero. However, the data alone are insufficient when a strictly nonparametric estimator is used.

B. POINT ESTIMATOR

As mentioned above, the estimators $\tilde{p}_1$, $\tilde{p}_2$ and $\tilde{p}_3$ are somewhat undesirable because they give step-function estimates for a continuous survival function. The next three estimators $\tilde{p}_4$, $\tilde{p}_5$ and $\tilde{p}_6$ are modification of the first three. Again they provide estimates of the survival function only at those points in time that corresponde to observed deaths.

These estimators are specified by Equations (E-2) and (E-4), except for a substitution of the term (N+1) in place of (N).

Since the point estimators have rigorous definitions at only discrete points in time, it is necessary to offer an interpolation rule. That is, we need a method of "connecting the dots." The method proposed here is to assume that the survival function declines in a piece-wise exponential decay between the discrete points in time. This procedure is equivalent to assuming that the hazard function is essentially constant between a consecutive pair of the discrete times, but that the hazard varies from one time period to the next. Such an assumption is intuitively acceptable unless one suspects violent fluctuations in the hazard function.

1. The Estimator, "$\tilde{p}_4(t)$"

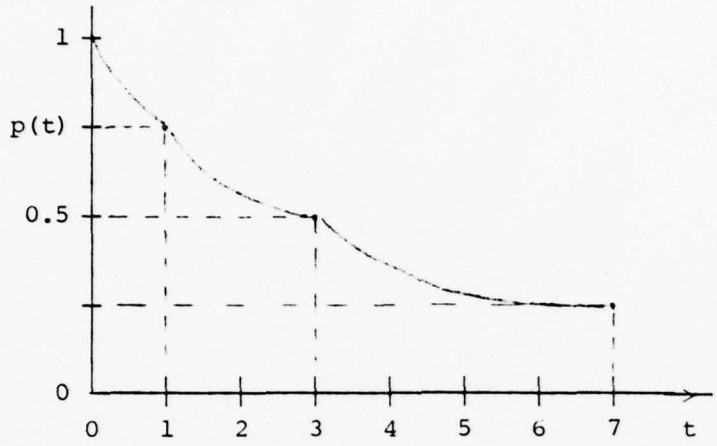
$\tilde{p}_4(t)$ is analogous to $\tilde{p}_1(t)$ in that only those individuals observed to die are included in the sample. These two estimators are naive because they suppress all data from the survival times of individuals terminated from observation by censoring.

These estimates, i.e., $\tilde{p}_1(t)$ and $\tilde{p}_4(t)$, tend to ignore information from the more long-lived individuals in the sample, and they may be expected to give biased estimates of the survival function.

The point estimator $\tilde{p}_4(t)$ gives the following values with sample data base presented earlier in this section.

t	$\tilde{p}_4(t)$	Interpolation t	Interpolation $\tilde{p}_4(t)$
0	1.0	0-1	$e^{t \cdot \ln(3/4)}$
1	$3/4 = 0.75$		
3	$(\frac{2}{3}) \times 0.75 = 0.5$	1-3	$(\frac{3}{4}) e^{\frac{t-1}{2} \cdot \ln(2/3)}$
7	$(\frac{1}{2}) \times 0.5 = 0.25$	3-7	$(\frac{1}{2}) e^{\frac{t-3}{4} \cdot \ln(1/2)}$

The interpolation for connecting the dots are as follows:



t	$0 \leq t \leq 1$	$1 \leq t \leq 3$	$3 \leq t \leq 7$
p(t)	$e^{-t/\tau}$	$p(t-1)\ e^{-\frac{t-1}{\tau}}$	$p(t-1)\ e^{-\frac{t-3}{\tau}}$
interpolation	$t = 1$ $\frac{3}{4} = e^{-\frac{1}{\tau}}$ $\frac{1}{\tau} = -\ln(\frac{3}{4})$	$t = 3$ $\frac{1}{2} = \frac{3}{4} e^{-\frac{2}{\tau}}$ $\frac{1}{\tau} = \frac{-\ln(\frac{2}{3})}{2}$	$t = 7$ $\frac{1}{4} = \frac{1}{2} e^{-\frac{4}{\tau}}$ $\frac{1}{\tau} = \frac{-\ln(\frac{1}{2})}{4}$
p(t)	$e^{t \cdot \ln(\frac{3}{4})}$	$(\frac{3}{4}) e^{\frac{t-1}{2} \cdot \ln(\frac{2}{3})}$	$(\frac{1}{2}) e^{\frac{t-3}{4} \cdot \ln(\frac{1}{2})}$

2. <u>The Estimator</u>, "$\tilde{p}_5(t)$"

The estimator $\tilde{p}_5(t)$ similarly corresponds to the product-limit estimator $\tilde{p}_2(t)$. These two estimators use information from the individuals on whom there are censored observations. $\tilde{p}_5$, like $\tilde{p}_2$, does not exploit information about that portion of the censored observation after the death event (of some other individual) preceding the censoring event.

For our hypothetical data base the following values are calculated for $\tilde{p}_5(t)$:

t	$\tilde{p}_5(t)$	Interpolation	
		t	$\tilde{p}_5(t)$
0	1.0		
1	5/6 = 0.833	0-1	$e^{t \cdot \ln(t/6)}$
3	$(\frac{3}{4})$ x 0.833 = 0.625	1-3	$(\frac{5}{6}) e^{\frac{t-1}{2} \ln(\frac{3}{4})}$
7	$(\frac{1}{2})$ x 0.625 = 0.312	3-7	$(\frac{5}{8}) e^{\frac{t-3}{4} \ln(\frac{1}{2})}$

Whenever censored observations are present, the estimator $\tilde{p}_4(t)$ never exceeds $\tilde{p}_5(t)$.

For $\tilde{p}_5(t)$, the value of N_i is taken to be the number of surviving individuals in the sample just before the observation of the ith death. This value is smaller than the number of surviving individuals just after the (i-1)st death if any truncation events occur in the interval. In fact, N_i is the smallest number of surviving individuals observed at any time during the interval (t_{i-1}, t_i). Thus $\tilde{p}_5$ might be expected to introduce a bias by using values of N_i that are, on the average, too small. However, this bias would be much less severe than the bias anticipated for the estimator $\tilde{p}_4(t)$.

The estimators $\tilde{p}_4$ and $\tilde{p}_5$ are insensitive to the precise times of the censoring events. A change in the time of the censoring event for individual B to $1+\varepsilon$ to $3-\varepsilon$, ε arbitrarily small, does not alter the estimates from $\tilde{p}_4$ and $\tilde{p}_5$ given in the preceding paragraph.

3. The Estimator, "$\tilde{p}_6(t)$"

The estimator $\tilde{p}_6(t)$ corresponds to $\tilde{p}_3(t)$ by accounting for all of the survival time for the truncated observations. For our hypothetical data base, the following values are calculated for $p_6(t)$:

t	$\tilde{p}_6(t)$	Interpolation t	Interpolation $\tilde{p}_6(t)$
0	1.0	0-1	$e^{t \cdot \ln(5/6)}$
1	5/6 = 0.833		$\frac{t-1}{2} \ln (\frac{3.5}{4.5})$
3	$(\frac{3.5}{4.5})$ x 0.833 = 0.648	1-3	$(\frac{5}{6}) e$
7	$(\frac{1.75}{2.75})$ x 0.648 = 0.412	3-7	$\frac{t-3}{4} \ln (\frac{1.75}{2.75})$ $(0.648) e$

The estimator $\tilde{p}_6(t)$ is based on the average number of surviving individuals noted in the various time intervals. These estimators give part credit for individuals whose lifetime is censored in mid-interval. The value of N_i for $\tilde{p}_6(t)$ is an unweighted time average. If the observation of an individual is truncated after 23% of the interval has elapsed, then that individual contributes a value of 0.23 to N_i. Individuals who are observed to survive the entire interval, and the individual whose death terminates the interval each contribute a value of 1.0 to N_i. This interpretation of the effective sample size is approximate if the hazard is approximately constant over the interval. If the hazard function changes markedly within a time interval containing censored events, then this interpretation of the effective sample size is biased. Therefore, the procedure of determining the value of N_i for the estimator $\tilde{p}_6(t)$ is based on the implicit assumption that the survival function is locally

exponential. If the hazard function may be assumed to vary slowly over each of the time intervals (t_{i-1}, t_i) then $\tilde{p}_6$ would appear to be biased on an acceptable approximation.

The estimator $\tilde{p}_6$, like $\tilde{p}_3$, depends on the precise times of all deaths and censoring events.

t	$\tilde{p}_6(t)$, $t_2 = 2$	$\tilde{p}_6(t)$, $t_2 = 1.1$	$\tilde{p}_6(t)$, $t_2 = 2.9$
0	1.0	1.0	1.0
1	5/6 = 0.833	5/6 = 0.833	5/6 = 0.833
3	$(\frac{3.5}{4.5})$ x 0.833 = 0.648.	$(\frac{3.05}{4.05})$ x 0.833 = 0.628.	$(\frac{3.95}{4.95})$ x 0.833 = 0.665
7	$(\frac{1.75}{2.75})$x 0.648 = 0.412.	$(\frac{1.75}{2.75})$ x 0.628 = 0.399.	$(\frac{1.75}{2.75})$ x 0.665 = 0.423

This illustrates that an increase (or decrease) in the total observed survival time causes an increase (or decrease) in the estimate $\tilde{p}_6$ over at least some of its time range.

If the last event is a censored, and not an observed, death, these estimators also require definition for the time period starting with the time of the last death and ending with the time of the final censoring event.

The method proposed here for $\tilde{p}_4(t)$ and $\tilde{p}_5(t)$ is to continue the exponential function used in the interval terminated by the time of the last death. This procedure can be illustrated with the modified data base used above in the discussion of $\tilde{p}_2$ and $\tilde{p}_3$.

C. THE BAYESIAN ESTIMATORS

Consideration is next given to quasi-Bayesian estimators based on a uniform prior distribution on the unit interval. Let $X_1,\ldots,X_N$ be the true survival times of N individuals which are censored on the right by N follow-up times $Y_1,\ldots,Y_N$. It is assumed that the X_i are independent,

identically distributed random variables with common distribution p(t) and we wish to estimate the survival function

$$p(t) = Pr(X > t)$$

However, we only have available the data,

$$Z_i = \min \{X_i, Y_i\}$$

$$\delta_i = \begin{cases} 1 & \text{if } X_i \leq Y_i \\ 0 & \text{if } X_i > Y_i, \ i=1,\ldots,n \end{cases}$$

If $\delta_i = 0$, then Z_i is called "a loss", and if

$\delta_i = 1$, then Z_i is called "a death".

Then $p_r[\delta_i = 1] = p_r[X_i > t] = p(t)$, $i=1,\ldots,N$.

The maximum likelihood estimator for p(t) is

$$\hat{p}(t) = \frac{s}{N} \quad \text{where} \quad s = \sum_{i=1}^{N} \delta_i$$

is the number of successful tests, s has the binomial distribution.

$$P(S|p) = \binom{N}{s} p^s (1-p)^{N-s}, \ s=0,1,\ldots,N, \ 0 < p < 1$$

$$f_p(p) = 1, \ 0 < p < 1$$

The joint density of s and p is

$$f_{s,p}(s,p) = \binom{N}{s} p^s (1-p)^{N-s}, \ 0 < p < 1, \ s=0,1,\ldots N.$$

The marginal for s is

$$p_s(s) = \int_0^1 \binom{N}{s} p^s (1-p)^{n-s} dp = \binom{N}{s} \cdot \frac{s!(N-s)!}{(N+1)!} = \frac{1}{N+1}$$

for s=0,1,...N. Thus, averaging over the values of p, all of which are assumed to be equally likely, the values of s are equally likely to occur. The posterior for p then is

$$f_{p|S}(p|S) = \frac{\Gamma(N+2)}{\Gamma(s+1)\Gamma(N-s+1)} p^{s}(1-p)^{N-s}, \quad 0 < p < 1 ,$$

a beta density with parameters s+1 and N-s+1. The mean of the posterior is $(s+1)/(N+2)$ and the modal (maximum value) of the posterior is s/N; thus the Bayes estimate of p (given s survivers occur in the sample of N) is

$$p^{*} = \frac{s+1}{N+2} \qquad \text{(C-1)}$$

Then, equation (C-1) yields a step function and also has shown that the uniform prior has the effect of adding two individuals to the population at risk with one dying at time zero and the other essentially immortal.

The Bayesian estimators based on a uniform prior distribution on the unit interval are denoted $\tilde{p}_{11}(t)$, $\tilde{p}_{12}(t)$ and $\tilde{p}_{13}(t)$, that correspond, respectively, to the estimators $\tilde{p}_{1}(t)$, $\tilde{p}_{2}(t)$ and $\tilde{p}_{3}(t)$. The sample data base thus gives the following estimates of the survival function:

t	$\tilde{p}_{11}(t)$	$\tilde{p}_{12}(t)$	$\tilde{p}_{13}(t)$
0-1	4/5 = 0.8	6/7 = 0.857	6/7 = 0.857
1-3	3/5 = 0.6	$(\frac{5}{6})$ x 0.857 = 0.714	$(\frac{5}{6})$ x 0.857 = 0.714
3-7	2/5 = 0.4	$(\frac{3}{4})$ x 0.714 = 0.536	$(\frac{3.5}{4.5})$ x 0.714 = 0.556
(7)	1/5 = 0.2	$(\frac{1}{2})$ x 0.536 = 0.268	$(\frac{1.75}{2.75})$ x 0.556 = 0.354

At the time of the final event (whether a death or a truncation), these step-function estimators drop to some positive value. Again, we have no data to indicate how the survival function proceeds to zero at subsequent times.

D. THE JACKKNIFE ESTIMATOR

We will assume that we observed, or have generated in a simulation, a survival probagility $p(t_j)$, $j=1,\ldots,n$, from various sample sizes. Furthermore we have some parameter or characteristic $p(t_j)$ of the sample size which we wish to estimate with an estimator $\hat{p}(t_j)$. The jackknife estimator $\tilde{p}(t,n)$ described below is an approximately unbiased estimator of $p(t_j)$. A modification of it has other useful properties.

$\tilde{p}_{-i}(t,n-1)$ is the estimator from the sample of n of the X_i's with the ith value deleted from the sample.

$$\tilde{p}_i(t,n) = n\,\tilde{p}(t,n) - (n-1)\,\tilde{p}_{-1}(t,n-1) \qquad i=1,\ldots,n$$

$$\tilde{p}(t,n) = \frac{1}{n}\sum_{i=1}^{n}\tilde{p}_i(t,n) = n\,\tilde{p}(t,n) - \frac{n-1}{n}\sum_{i=1}^{n}\tilde{p}_{-1}(t,n-1)$$

the $\tilde{\tilde{p}}_i(t,n)$, called the PSEUDO-values.

The PSEUDO-values can be used to obtain variance estimates of $\bar{\tilde{p}}(t,n)$ and to set approximate confidence limits, using Student's t.

The idea is that the PSEUDO-values will be approximately independently and normally distributed. The jackknife estimator $\bar{\tilde{p}}(t,n)$ is a sample average so we form an estimate $s^2_{\bar{\tilde{p}}(t,n)}$ of its variance given by the following relationship (Miller, 1974):

$$s^2 = \frac{\Sigma\tilde{p}_i^{\,2}(t,n) - \frac{1}{n}(\Sigma\tilde{p}_i(t,n))^2}{n-1}$$

$$s^2_{\bar{\tilde{p}}(t,n)} = \frac{s^2}{n}$$

This procedure is particularly useful if the number of data points is small, but it must be used with care. Note, that the estimator $\tilde{p}(t,n)$ is designed to eliminate a $\frac{1}{n}$ bias term in the estimator $\tilde{p}(t,n)$. Of course the computational aspects of the complete jackknife can be quite onerous, especially if $\tilde{p}(n)$ were, say, a complicated maximum likelihood estimator. Miller, reference (4) has shown that the product limit estimator is its own jackknife.

Logistic Transformation

Although one can legitimately jackknife the Kaplan-Meier estimate directly, there is some reason to believe that a preliminary transformation will give improved results. Consequently, consider the transformation

$$\ell = \ell n\left(\frac{\tilde{p}(t)}{1-\tilde{p}(t)}\right)$$

and notice that where the range of $\tilde{p}(t)$ is from zero to unity, the above transformation makes the range of ℓ run from $-\infty$ to ∞. The procedure utilized will be as follows.

(A) Compute the overall estimate at a time point t, using all N data points, and using a "continuity" correction that has the effect of removing the effect of a zero in the logarithm (see D.R. Cox, Analysis of Binary Data, Methuen Monograph):

$$\ell_N = \ell n\left(\frac{\tilde{p}_N(t) + \frac{1}{2N}}{1-\tilde{p}_N(t) + \frac{1}{2N}}\right)$$

(B) Compute the ℓ-values by leaving out each data point in turn when computing $\tilde{p}(t)$: for i=1,2,...,N.

$$\ell_{N-1,i} = \ell n\left(\frac{\tilde{p}_{N-1,-i}(t) + \frac{1}{2(N-1)}}{1-\tilde{p}_{N-1,-i}(t) + \frac{1}{2(N-1)}}\right)$$

(C) Form the pseudo-values

$$z_i = N\ell_N - (N-1)\,\ell_{N-1,-i}$$

(D) Compute $\bar{z}$, s_z^2

(E) Put approximate confidence $(1-\alpha)\cdot 100\%$ limits on $E[\ell]$ as follows

$$L \leq E[\ell] \leq H$$

where $$H(L) = \bar{z} +(-)\; t_{1-\alpha}\,(N-1)\sqrt{\frac{s_z^2}{N}}$$

(F) Transform bash to obtain

$$\frac{e^L}{1+e^L}, \text{ and } \frac{e^H}{1+e^H}$$

The true value, $p(t)$, should be enclosed between these levels for roughly $(1-\alpha)\cdot 100\%$ of all samples. The coverage properties of this procedure will now be checked by simulation: successive samples of size N will be selected, the jackknife limits H and L will be computed for each, and a check will be made as to whether $\frac{e^L}{1+e^L} \leq p(t) \leq \frac{e^H}{1+e^H}$ or not. Tables illustrating performance are given subsequently.

Let

$\ell_{N-1,-i}$ is the logistic transformation estimator from the sample n of the X_i's with the ith value deleted from the sample.

$$\ell_{N-1,-i} = \ell n \left(\frac{\tilde{p}_{N-1,-i}(t) + \frac{1}{2(N-1)}}{1-\tilde{p}_{N-1,-i}(t)+\frac{1}{2(N-1)}} \right)$$

$\ell_{N-1,-i}$

t \ i	1	2	3	4	5
t_1	3.04	0.98	0.98	0.98	0.98
t_2	3.04	0.98	0.98	0.98	0.98
t_3	0.63	0	0.98	-0.46	-0.46
t_4	0.63	0	0.98	-0.46	-0.46
t_5		-3.04	-3.04	-3.04	-1.89

$$z_i = N\ell_N - (N-1)\ \ell_{N-1,-i}$$

$$= N\ell n \left(\frac{\tilde{p}_N(t) + \frac{1}{2N}}{1-\tilde{p}_N(t)+\frac{1}{2N}} \right) - (N-1)\ \ell_N \left(\frac{\tilde{p}_{N-1,-i}(t) + \frac{1}{2(N-1)}}{1-\tilde{p}_{N-1,-i}(t)+\frac{1}{2(N-1)}} \right)$$

$z_i(N)$ are called PSEUDO-values of logistic transformation, the following values are calculated:

z_i:

t \ i	1	2	3	4	5
t_1	-6.05	2.198	2.198	2.198	2.198
t_2	-6.05	2.198	2.198	2.198	2.198
t_3	-1.9	0.606	-3.314	2.446	2.446
t_4	-1.9	0.606	-3.314	2.446	2.446
t_5	-3.0626	-3.0626	-3.0626	-3.0626	-7.162

Average of the pseudo-values

$$\bar{z} = \frac{1}{N} \sum_{i=1}^{N} z_i$$

Invert to find jackknife estimator of logistic transformation

$$\bar{z} = \ell n \left(\frac{\tilde{p}(t) + \frac{1}{2N}}{1-\tilde{p}(t)+\frac{1}{2N}} \right)$$

$$\tilde{p}(t) = \frac{(1+\frac{1}{2N}) e^{\bar{z}} - \frac{1}{2N}}{1 + e^{\bar{z}}}$$

called the jackknife estimator of logistic transformation

Variance of the z_i

$$S_z^2 = \text{Var}(z) = \frac{1}{n-1} \sum_{i=1}^{n} z_i - \bar{z}$$

The following values are calculated:

t	$\bar{z}$	$\tilde{p}(t)$	Var J
t_1	0.5484	0.646	13.6
t_2	0.5484	0.646	13.6
t_3	0.0568	0.516	6.727
t_4	0.0568	0.516	6.727
t_5	-3.882	0	3.361

The jackknife estimator for estimating variability and giving confidence interval.

Tukey, reference (3) has suggested that in the jackknife procedure we consider the pseudo values $Z_i(N)$ as approximately independent and identically districuted and consequently, since $\bar{z}$ is an average of

the $Z_i(N)$, proceed as if

$$\frac{N^{\frac{1}{2}}\bar{z} - \ell_N}{\{\frac{1}{N-1}\sum_{i=1}^{n}(z_i - \bar{z})^2\}^{\frac{1}{2}}}$$

has t-distribution with N-1 d.F.

If the z_i are approximately normal variates (Miller has shown) confidence bands for the unknown $\tilde{p}(t)$ are given, as for the mean of any normal variate when estimated from sample size n.

$$\bar{z} \pm \frac{s_z}{\sqrt{n}} t_{1-\alpha/2}(N-1) \qquad \text{(D-1)}$$

i.e.

$$\bar{z} - \frac{s_z}{\sqrt{n}} t_{1-\alpha/2}(N-1) \leq \ell n\left(\frac{\tilde{p}(t) + \frac{1}{2N}}{1-\tilde{p}(t)+\frac{1}{2N}}\right) \leq \bar{z} + \frac{s_z}{\sqrt{n}} t_{1-\alpha/2}(N-1)$$

$$\underline{L}(n) = \bar{z} - \frac{s_z}{\sqrt{N}} t_{1-\alpha/2}$$

$$\bar{L}(n) = \bar{z} + \frac{s_z}{\sqrt{N}} t_{1-\alpha/2}$$

$$\frac{(1+\frac{1}{2N})e^{\underline{L}(N)} - \frac{1}{2N}}{1 + e^{\underline{L}(N)}} \leq \tilde{p}(t) \leq \frac{(1+\frac{1}{2N})e^{\bar{L}(N)} - \frac{1}{2N}}{1 + e^{\bar{L}(N)}}$$

The following values are calculated:

t	4	$t_{1-\alpha/2} = 2.776$
	Lower Int.	Upper Int.
t_1	0	1.0
t_2	0	1.0
t_3	0	1.0
t_4	0	1.0
t_5	0	0.14

The basis for this leap of the imagination seems to be that if $\bar{z} = \bar{X} = \bar{X}_n$ then the procedure for obtaining confidence intervals using equation (D-1) and pseudo-values is the same as the procedure using jackknife. Then if $\bar{X}_N = \bar{z}$ and

$$\bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_i \qquad \text{we have}$$

$$z_i = N \ell_N - (n-1)\, \ell_{N-1,-i}$$

$$= N \bar{X}_N - (N-1) \frac{\{ \sum_{j=1}^{N} X_j \} - X_i}{N-1}$$

$$= \sum_{j=1}^{N} X_j - [\sum_{j=1}^{N} X_j] + X_i = X_i$$

Thus the pseudo value

$$z_i = X_i \quad \text{and} \quad \bar{z} = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X}_n$$

The pseudo values are independent if $\bar{z} = \bar{X}_n$ and they are normal if X_i is normal.

E. PARAMETRIC ESTIMATOR, "$\tilde{p}_7(t)$"

This paper considers one additional estimator, denoted $\tilde{p}_7(t)$. It is a parametric estimator. Therefore, it is not really a competitor to the thirteen non-parametric estimators considered here. In general, a parametric estimator would not be used if the functional form were regarded as unknown. Similarly, a non-parametric estimator would not

normally be used if the survival function were strongly suspected to have a specified form.

$\tilde{p}_7(t)$ is the well known maximum likelihood estimator for the exponential distribution:

$$\tilde{p}_7(t) = e^{-t/\tau}$$

where $$\tau = \frac{\Sigma\, t_i}{\text{number of observed death}}$$

In our sample data base, the total observed survival time is 19, and three deaths are observed. Thus,

$$\Sigma\, t_i = 1 + 2 + 3 + 6 + 7 = 19$$

$$\tau = \frac{19}{3}$$

and $$\tilde{p}_7(t) = e^{-3t/19}$$

Calculations for selected times of interest yield the following estimates:

$$p_7(0) = 1.0$$

$$p_7(1) = 0.854$$

$$p_7(3) = 0.623$$

$$p_7(7) = 0.331$$

The thirteen non-parametric estimators are compared for a variety of generating distributions for both the death mechanism and censoring mechanism.

IV. INSTRUCTIONS FOR USING PROGRAM

INPUT

Each input card bears nine variables. The distribution of time of death is entered in the first set of (five) columns, the censoring distribution is entered in the second set of (ten) columns, a parameter of the censoring distribution is entered in the third set of (ten) columns, the number of replication is entered in the fourth set of (five) columns, the number of the event is entered in the fifth set of (five) columns. For the purpose of all print output used code "0" and "1" in the sixth set of (five) columns, the seed number is entered in the seventh set of (five) columns, after the card giving the time of the last event of a data set, a card with "0" or "1" in the column 50 is inserted, i.e., the "0" indicating more data sets to follow and "1" indicating the last data sets and t value is entered in the ninth set of (eight) columns.

The distribution of timeof death and of censoring time used code as follows:

Code	Type of Distribution
1	Uniform
2	Exponential
3	Delta function

OUTPUT

The output lists:

1) the time of each observed failure
2) estimated survival probability at that time
3) the variance of that estimator
4) result of goodness fit

a) mean error

b) mean absolute error (ABS)

c) root-mean-square error (RMS)

5) total number of observed death

6) confidence interval at particular time

Definition of Fortran Variables

NDIE : the distribution of time of death

NTRUNC : the distribution of censoring time

XTRUNC : the parameter of the distribution of censoring time

NREPL : number of replication

NEVENT : number of event

NWRITE : write all output or partial output of simulation

NEND : indicate more data sets or last data set

TN : t statistic value

p_1 : the estimator, $\tilde{p}_1(t)$

p_2 : the estimator, $\tilde{p}_2(t)$

p_3 : the estimator, $\tilde{p}_3(t)$

p_4 : the estimator, $\tilde{p}_4(t)$

p_5 : the estimator, $\tilde{p}_5(t)$

p_6 : the estimator, $\tilde{p}_6(t)$

p_7 : parametric estimator , $\tilde{p}_7(t)$

p_8 : jackknife estimator of logistic transformation of $\tilde{p}_4(t)$

p_9 : jackknife estimator of logistic transformation of $\tilde{p}_5(t)$

p_{10} : jackknife estimator of logistic transformation of $\tilde{p}_6(t)$

p_{11} : Bayesian estimator of $\tilde{p}_1(t)$

p_{12} : Bayesian estimator of $\tilde{p}_2(t)$

p_{13} : Bayesian estimator of $\tilde{p}_3(t)$

p_{14} : jackknife estimator of logistic transformation of $\tilde{p}_2(t)$

SL(I,J) : PSEUDO-value

SBAR : average of pseudo-value

Var : variance of estimator, $\tilde{p}(t)$

Var_J : variance of jackknife estimator

u(I,J) : mean of goodness fit

w(I,J) : absolute mean of goodness fit

s(I,J) : root mean square error

C_1 : upper confidence interval of $p_{14}(t)$

C_2 : lower confidence interval of $p_{14}(t)$

C_3 : upper confidence interval of $p_8(t)$

C_4 : lower confidence interval of $p_8(t)$

C_5 : upper confidence interval of $p_9(t)$

C_6 : lower confidence interval of $p_9(t)$

C_7 : upper confidence interval of $p_{10}(t)$

C_8 : lower confidence interval of $p_{10}(t)$

To compare RMS with product limit ($p_2(t)$) and jackknife estimator of logistic transformation ($p_{14}(t)$)

```
      2     1     4.0000 1000    20     1  505  :  Input
      1     1 0.01275 0.95000 0.93726          :  Output
      1     0 0.01581 0.0     0.00000
      1     1 0.02085 0.89722 0.88376
      1     1 0.02940 0.84444 0.83159
      1     1 0.11925 0.79167 0.78256
      1     1 0.13877 0.73889 0.73445
      1     0 0.17874 0.0     0.05254
      1     0 0.18582 0.0     0.00000
      1     1 0.19034 0.67731 0.68724
      1     1 0.19818 0.61574 0.63201
      1     1 0.27670 0.55417 0.57755
      1     1 0.28060 0.49259 0.52356
      1     0 0.31870 0.0     0.00000
      1     1 0.48075 0.42222 0.47045
      1     1 0.65596 0.35185 0.40828
      1     1 0.70112 0.28148 0.34624
      1     1 0.77889 0.21111 0.28413
      1     1 1.18657 0.14074 0.22201
      1     0 1.25712 0.0     0.05254
      1     1 1.47370 0.0     0.05254

        0.100   0.200   0.300   0.400   0.500   0.600   0.700   0.800   0.900
P1   -0.002   0.000   0.002  -0.000  -0.001  -0.001   0.007   0.028   0.066  MEAN
P2   -0.013  -0.007   0.004   0.015   0.029   0.044   0.065   0.109   0.166  MEAN

P1    0.063   0.071   0.084   0.090   0.097   0.093   0.091   0.077   0.078  ABS
P2    0.057   0.068   0.075   0.080   0.087   0.087   0.094   0.116   0.166  ABS

P1    0.069   0.091   0.105   0.113   0.119   0.115   0.113   0.100   0.102  RMS
P2    0.071   0.085   0.094   0.101   0.108   0.109   0.120   0.140   0.185  RMS

C1    0.985   0.951   0.901   0.841   0.781   0.722   0.681   0.678   0.751  CONF
C2    0.787   0.591   0.459   0.355   0.267   0.185   0.116   0.057   0.005  CONF

      57.726  91.254  97.085  97.959  98.542  99.125  97.959  98.542  97.668  PER

      1000    1000    1000    1000     998     990     951     796     343


      2     2     4.0000 1000    10     1 1509  :  Input
      1     0 0.03688 0.0     0.00000          :  Output
      1     1 0.14510 0.88889 0.90796
      1     0 0.22306 0.0     0.00000
      1     1 0.22401 0.76190 0.78995
      1     0 0.30447 0.0     0.00000
      1     1 0.46359 0.60952 0.67485
      1     1 0.71228 0.45714 0.54479
      1     1 0.79300 0.30476 0.42580
      1     1 1.14695 0.15238 0.36025
      1     1 2.27255 0.0     0.06065

        0.100   0.200   0.300   0.400   0.500   0.600   0.700   0.800   0.900
P1   -0.003  -0.002  -0.005  -0.001   0.006   0.018   0.031   0.062   0.134  MEAN
P2    0.005   0.013   0.020   0.041   0.067   0.100   0.142   0.200   0.274  MEAN

P1    0.073   0.100   0.116   0.129   0.131   0.128   0.115   0.107   0.111  ABS
P2    0.074   0.096   0.105   0.114   0.119   0.128   0.148   0.200   0.274  ABS

P1    0.098   0.130   0.149   0.160   0.165   0.160   0.146   0.142   0.148  RMS
P2    0.093   0.121   0.133   0.142   0.150   0.162   0.182   0.227   0.291  RMS

C1    0.975   0.953   0.922   0.887   0.859   0.844   0.854   0.896   0.966  CONF
C2    0.840   0.640   0.455   0.336   0.242   0.163   0.092   0.033  -0.015  CONF

      52.667  70.933  78.933  91.733  95.733  97.600  97.067  97.067  96.800  PER

      1000    1000    1000     996     984     944     853     656     375
```

Computer output of the fourteen estimators

2 1 4.0000 1000 20 0 505

	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900	
P 1	-0.033	-0.058	-0.079	-0.100	-0.114	-0.121	-0.114	-0.092	-0.049	MEAN
P 2	-0.002	0.000	0.002	-0.000	-0.001	-0.003	0.003	0.018	0.056	MEAN
P 3	-0.001	0.001	0.003	0.002	0.002	0.002	0.009	0.026	0.067	MEAN
P 4	-0.055	-0.073	-0.086	-0.099	-0.106	-0.106	-0.094	-0.070	-0.029	MEAN
P 5	-0.021	-0.015	-0.010	-0.007	-0.003	-0.001	0.007	0.017	0.040	MEAN
P 6	-0.020	-0.014	-0.008	-0.004	0.001	0.006	0.018	0.037	0.077	MEAN
P 7	-0.004	-0.007	-0.008	-0.008	-0.007	-0.005	0.000	0.010	0.033	MEAN
P 8	-0.045	-0.069	-0.085	-0.099	-0.107	-0.109	-0.100	-0.080	-0.039	MEAN
P 9	-0.015	-0.013	-0.009	-0.005	0.001	0.006	0.015	0.025	0.051	MEAN
P10	-0.015	-0.013	-0.008	-0.004	0.002	0.007	0.018	0.032	0.060	MEAN
P11	-0.077	-0.087	-0.092	-0.100	-0.100	-0.095	-0.077	-0.046	0.005	MEAN
P12	-0.038	-0.027	-0.016	-0.008	0.001	0.011	0.028	0.054	0.103	MEAN
P13	-0.038	-0.026	-0.015	-0.006	0.004	0.015	0.033	0.061	0.111	MEAN
P14	0.011	0.014	0.020	0.023	0.029	0.039	0.064	0.117	0.222	MEAN
P 1	0.074	0.100	0.118	0.133	0.142	0.141	0.131	0.104	0.059	ABS
P 2	0.053	0.071	0.084	0.090	0.097	0.095	0.092	0.077	0.070	ABS
P 3	0.052	0.071	0.084	0.090	0.097	0.094	0.092	0.080	0.079	ABS
P 4	0.073	0.098	0.113	0.125	0.131	0.126	0.112	0.085	0.045	ABS
P 5	0.050	0.066	0.078	0.083	0.085	0.087	0.085	0.076	0.066	ABS
P 6	0.050	0.066	0.078	0.083	0.089	0.086	0.083	0.074	0.082	ABS
P 7	0.021	0.039	0.054	0.065	0.073	0.077	0.075	0.065	0.053	ABS
P 8	0.074	0.100	0.114	0.125	0.132	0.129	0.116	0.094	0.060	ABS
P 9	0.053	0.069	0.078	0.082	0.088	0.087	0.088	0.084	0.080	ABS
P10	0.053	0.069	0.078	0.082	0.088	0.086	0.086	0.079	0.074	ABS
P11	0.066	0.105	0.116	0.124	0.125	0.117	0.100	0.066	0.034	ABS
P12	0.056	0.068	0.077	0.082	0.087	0.084	0.084	0.079	0.105	ABS
P13	0.056	0.067	0.077	0.082	0.088	0.085	0.085	0.083	0.111	ABS
P14	0.058	0.073	0.084	0.089	0.096	0.093	0.095	0.123	0.223	ABS
P 1	0.096	0.129	0.150	0.165	0.173	0.168	0.153	0.119	0.068	RMS
P 2	0.065	0.091	0.105	0.113	0.119	0.116	0.114	0.099	0.094	RMS
P 3	0.065	0.091	0.105	0.113	0.119	0.116	0.115	0.103	0.105	RMS
P 4	0.096	0.126	0.144	0.154	0.158	0.150	0.131	0.100	0.055	RMS
P 5	0.065	0.084	0.097	0.103	0.109	0.108	0.106	0.096	0.084	RMS
P 6	0.065	0.084	0.097	0.103	0.109	0.106	0.104	0.096	0.104	RMS
P 7	0.027	0.049	0.068	0.082	0.091	0.095	0.092	0.082	0.071	RMS
P 8	0.096	0.127	0.144	0.154	0.159	0.154	0.128	0.108	0.086	RMS
P 9	0.068	0.087	0.098	0.103	0.109	0.108	0.110	0.107	0.114	RMS
P10	0.068	0.087	0.098	0.103	0.108	0.107	0.107	0.101	0.097	RMS
P11	0.110	0.133	0.146	0.152	0.152	0.139	0.118	0.081	0.042	RMS
P12	0.073	0.087	0.097	0.103	0.107	0.104	0.105	0.102	0.124	RMS
P13	0.073	0.087	0.096	0.103	0.108	0.105	0.108	0.107	0.132	RMS
P14	0.070	0.092	0.107	0.114	0.123	0.118	0.123	0.158	0.258	RMS
C1	0.968	0.932	0.883	0.826	0.768	0.717	0.667	0.741	0.850	CONF
C2	0.761	0.570	0.452	0.357	0.273	0.196	0.129	0.071	0.028	CONF
C3	0.961	0.913	0.850	0.776	0.696	0.611	0.519	0.420	0.318	CONF
C4	0.707	0.486	0.360	0.267	0.191	0.127	0.079	0.045	0.036	CONF
C5	0.967	0.925	0.879	0.820	0.756	0.688	0.623	0.571	0.582	CONF
C6	0.760	0.570	0.454	0.363	0.283	0.209	0.148	0.096	0.054	CONF
C7	0.967	0.930	0.880	0.821	0.757	0.689	0.623	0.571	0.580	CONF
C8	0.760	0.569	0.453	0.361	0.282	0.209	0.148	0.096	0.095	CONF
	1000	1000	1000	1000	1000	1000	994	933	556	

V. RESULTS OF THE SIMULATION

This paragraph presents graphical comparisons of the 13 estimators based on simulated data. Each comparison is based on 1000 replications of a simulated data base. The bias and RMS error (square-root of mean-squared error) of each estimator depends on the parameters that control the simulated data base. No single estimator dominates all others under all conditions.

The bias and RMS errors of the estimators depend on several factors:

(A) The sample size (NEVENT) of individuals under observation at time zero affects the accuracy of the estimators. In general, a larger sample size leads to a better estimate than a smaller sample. Values of NEVENT selected for simulation are 5, 10, 25, and 50 (plus one simulation with NEVENT = 100).

(B) The distribution of times at which the observations are censored (unless the individual dies earlier) affects the performance of the various estimators. This distribution is particularly important in conjunction with the distribution of lifetimes (do most individuals die before censoring is likely?, are deaths and censoring events about equally likely at all times?, are most observations censored before death?). Three types of distributions are assumed to underlie the censoring mechanism:

(1) Some of the samples are generated on the assumption that no censoring occurs.

(2) Some samples are generated from a uniform distribution of times of censoring.

(3) Other data bases are generated from an exponential distribution of censoring times.

(C) The distribution of lifetimes (ignoring the possibility of censoring) also affects the performance of the various estimators. Two types of distributions are assumed to underlie the death mechanism:

(1) Some of the samples are generated from a uniform distribution of lifetimes.

(2) Other data bases are generated from an exponential distribution of lifetimes.

If a uniform distribution of lifetimes is selected, its range is always over the interval from time 0 to time 1. If an exponential distribution is selected, it always has a mean lifetime of 1. The distributions of truncation times (uniform or exponential) have parameters .25, .5, .667, .75, 1, 1.333, 1.5, 2 and 4. A wide variety of samples may be simulated by mxing various pairs of distributions (for censoring times and deaths). Since the time units are arbitrary, the restriction on mean lifetimes is irrelevant.

The true value of the survival function is, p(t), and the form of this function affects the relative performance of the 13 nonparametric estimators. For example, the Bayesian estimator $p_{12}(t)$ tends to be better as measured by square-root of mean-squared error than its counterpart (the product-limit estimator, $p_2(t)$) for the time frame in which

$$.3 < p(t) < .9$$

However, the product limit estimator tends to be better for those times when p(t) is close to zero or unity.

The point estimators, $p_5(t)$ and $p_6(t)$ tend to be better than the product-limit estimator ($p_2(t)$) for all time periods. The jackknife

estimators of logistic transformation ($p_8(t)$, $p_9(t)$, $p_{10}(t)$) of point estimators tends as same as its counterpart point estimators ($p_4(t)$, $p_5(t)$, $p_6(t)$) for all time periods. And also the estimator formed by jackknifing the logistic transformation ($p_{14}(t)$) of the product limit estimator tends to be better than its counterpart product limit ($p_2(t)$) for the time frame in which

$$.1 < p(t) < .7$$

However, the product limit estimator tends to be better for those times when p(t) is close to unity. Point estimators, $p_5(t)$ and $p_6(t)$ tend to be same for the time frame in which

$$0.1 < p_T(t) < 0.9$$

However, the $p_5(t)$ tends to be better for those times when $p_T(t)$ is close to unity. The jackknife procedure may be validated, in an empirical sense, by sampling experiments or computer simulation in the follow-ign manner. First, times of censoring and death are obtained by drawing random numbers from postulated distributions. Second, the jackknifed estimator of the logistic-transformed product-limit estimation is found, and confidence limits are computed by the method of Tukey, reference (3). Since the true value of survival function, p(t), is known, so is the theoretical value of A. The jackknife confidence intervals can be checked for coverage: if $L_\alpha \leq A \leq H_\alpha$ then the particular interval covers, while otherwise (if $A < L_\alpha$ or $H_\alpha < A$) it does not cover. Finally, the above procedure can be repeated many times (say 1000) and the fraction of repetitions which contains the true value of A is recorded. This fraction of the coverage should desirably be close to $(1-\alpha)$, the nominal confidence level.

The jackknife confidence limit procedure can be said to be robust of validity, ref (7), if the actual coverage is close to the nominal coverage, $1-\alpha$, for a various distributions. Such seems to be true for large n ($n \geq 50$). However, the jackknife confidence limits do not cover accurately when the true value of p(t) is close to unity.

The following tables illustrate confidence limits of jackknife method of product limit ($p_2(t)$). Many computer generated graphics are presented on the following pages to complete this section.

Table 1

Simulation Experiments Validating Table

95% Confidence Limits (t value = 2.776)

n=5.

Dist. of Death	Dist. of Censoring	trunc. parameter	distinguish / True Value	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
Exponential	Exponential	2.0	Upper Int.	0.962	0.935	0.917	0.912	0.922	0.945	0.971	0.995	1.0
			Lower Int.	0.709	0.515	0.366	0.259	0.173	0.115	0.068	0.034	0.009
			Coverage	56.735	66.122	75.408	85.714	90.816	96.939	98.980	95.918	96.939
		1.3333	Upper Int.	0.962	0.935	0.918	0.914	0.923	0.942	0.969	0.992	1.0
			Lower Int.	0.741	0.559	0.422	0.313	0.238	0.166	0.10	0.059	0.068
			Coverage	57.209	65.116	72.093	79.070	86.047	93.029	95.023	05.023	86.047
Uniform	Exponential	4.0	Upper Int.	0.955	0.910	0.871	0.846	0.825	0.819	0.827	0.837	0.855
			Lower Int.	0.744	0.552	0.407	0.301	0.217	0.163	0.113	0.078	0.047
			Coverage	50.792	68.358	74.194	85.630	92.669	96.774	92.150	95.308	88.886
		0.6667	Upper Int.	0.957	0.923	0.901	0.893	0.898	0.904	0.911	0.915	0.932
			Lower Int.	0.775	0.618	0.494	0.393	0.324	0.276	0.202	0.166	0.025
			Coverage	54.545	65.545	73.636	82.727	92.273	94.091	86.364	87.27	72.787
Exponential	Uniform	2.0	Upper Int.	0.964	0.931	0.911	0.901	0.906	0.917	0.940	0.963	**
			Lower Int.	0.760	0.558	0.419	0.305	0.231	0.171	0.125	0.122	**
			Coverage	55.887	56.259	72.340	82.496	87.943	91.489	85.668	87.943	**
		1.3333	Upper Int.	0.957	0.924	0.902	0.891	0.897	0.914	0.979	**	**
			Lower Int.	0.757	0.580	0.456	0.380	0.305	0.256	0.247	**	**
			Coverage	45.887	54.255	72.340	85.496	88.15	93.52	96.688	**	**
Uniform	Uniform	4.0	Upper Int.	0.948	0.906	0.862	0.836	0.821	0.816	0.821	0.837	0.863
			Lower Int.	0.718	0.533	0.385	0.291	0.216	0.155	0.112	0.080	0.049
			Coverage	53.117	61.039	68.571	87.987	93.506	96.429	91.558	95.130	89.844
		1.0	Upper Int.	0.953	0.917	0.893	0.882	0.883	0.891	0.907	0.924	0.938
			Lower Int.	0.751	0.60	0.496	0.399	0.318	0.276	0.251	0.131	0.049
			Coverage	58.0	70.0	78.333	78.667	78.667	80.0	70.0	73.333	66.667

Table 2

Simulation Experiments Validating Table

n=10 95% Confidence Limits (t value = 2.262)

Dist. of Death	Dist. of Censoring	trunc. parameter	distinguish	True Value	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
Exponential	Exponential	4.0		Upper Int.	0.975	0.953	0.922	0.887	0.859	0.844	0.854	0.896	0.966
				Lower Int.	0.840	0.640	0.455	0.336	0.242	0.163	0.092	0.033	0.0
				Coverage	52.667	70.933	78.933	91.733	95.733	97.60	97.067	97.067	96.8
		1.0		Upper Int.	0.966	0.930	0.891	0.861	0.851	0.858	0.892	0.938	0.976
				Lower Int.	0.582	0.359	0.241	0.176	0.125	0.984	0.058	0.032	0.015
				Coverage	62.0	87.50	94.643	98.214	100.0	100.0	100.0	98.214	98.214
Uniform	Exponential	1.0		Upper Int.	0.954	0.899	0.843	0.792	0.753	0.730	0.731	0.745	0.760
				Lower Int.	0.612	0.410	0.305	0.250	0.204	0.160	0.120	0.083	0.060
				Coverage	57.798	77.982	89.297	96.330	99.083	98.165	99.083	96.024	76.453
		2.0		Upper Int.	0.956	0.897	0.834	0.772	0.716	0.673	0.646	0.637	0.649
				Lower Int.	0.615	0.391	0.286	0.227	0.138	0.151	0.120	0.089	0.058
				Coverage	53.542	78.125	84.167	94.167	98.333	98.542	99.167	97.917	82.083
Exponential	Uniform	4.0		Upper Int.	0.975	0.948	0.913	0.874	0.840	0.821	0.828	0.865	0.937
				Lower Int.	0.546	0.295	0.165	0.100	0.068	0.048	0.030	0.012	0.003
				Coverage	60.0	86.154	96.923	98.174	100.0	100.0	100.0	100.0	100.0
		2.0		Upper Int.	0.940	0.896	0.855	0.814	0.787	0.785	0.806	0.851	**
				Lower Int.	0.860	0.709	0.578	0.468	0.379	0.308	0.251	0.168	**
				Coverage	57.6	51.2	78.933	91.733	96.0	98.133	97.333	96.8	**
Uniform	Uniform	1.0		Upper Int.	0.949	0.896	0.837	0.794	0.767	0.761	0.779	0.800	0.840
				Lower Int.	0.596	0.423	0.324	0.265	0.225	0.182	0.150	0.130	0.097
				Coverage	59.091	79.545	93.182	93.182	100.0	97.727	97.727	72.727	56.818
		4.0		Upper Int.	0.957	0.902	0.836	0.767	0.703	0.647	0.603	0.583	0.589
				Lower Int.	0.602	0.383	0.272	0.220	0.184	0.153	0.121	0.089	0.055
				Coverage	57.093	79.159	87.566	94.921	99.124	100.0	100.0	98.949	86.165

Table 3

Simulation Experiments Validating Table
95% Confidence Limits (t value = 2.093)

n=25

Dist. of Die	Dist. of Trunc.	Trunc. Parameter	distinguish / True Value	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
Exponential	Exponential	2.0	Upper Int.	0.986	0.948	0.893	0.833	0.774	0.717	0.682	0.798	0.805
			Lower Int.	0.787	0.603	0.477	0.372	0.283	0.199	0.126	0.060	0.0
			Coverage	68.148	93.827	97.778	98.765	98.272	98.272	99.259	98.025	95.309
		1.333	Upper Int.	0.941	0.897	0.848	0.787	0.746	0.702	0.693	0.731	0.826
			Lower Int.	0.778	0.612	0.501	0.411	0.325	0.245	0.173	0.112	0.058
			Coverage	66.260	91.057	97.154	98.374	98.780	95.935	96.341	93.496	80.488
		0.667	Upper Int.	0.967	0.910	0.851	0.794	0.753	0.745	0.785	0.825	0.892
			Lower Int.	0.850	0.696	0.582	0.478	0.384	0.302	0.211	0.159	0.103
			Coverage	67.742	90.323	90.323	93.548	93.548	90.323	87.097	74.194	58.065
Exponential	Uniform	4.0	Upper Int.	0.948	0.911	0.864	0.809	0.748	0.689	0.636	0.614	0.683
			Lower Int.	0.751	0.578	0.469	0.376	0.293	0.218	0.149	0.091	0.045
			Coverage	66.821	93.968	97.681	99.768	98.840	99.536	98.840	98.376	93.039
		1.333	Upper Int.	0.922	0.860	0.792	0.730	0.675	0.643	0.715	**	**
			Lower Int.	0.836	0.708	0.601	0.513	0.431	0.369	0.327	**	**
			Coverage	66.260	91.057	97.154	98.374	98.780	95.935	96.341	**	**
Uniform	Exponential	2.00	Upper Int.	0.951	0.878	0.799	0.714	0.627	0.541	0.455	0.372	0.309
			Lower Int.	0.393	0.299	0.274	0.250	0.221	0.186	0.146	0.100	0.054
			Coverage	85.985	87.121	90.278	91.919	91.919	93.813	95.581	97.854	88.005
		1.00	Upper Int.	0.950	0.877	0.796	0.717	0.638	0.559	0.490	0.446	0.428
			Lower Int.	0.402	0.306	0.275	0.249	0.218	0.181	0.140	0.096	0.056
			Coverage	83.770	86.721	89.344	90.328	93.934	95.738	95.525	98.689	86.866
Uniform	Uniform	2.00	Upper Int.	0.952	0.879	9.797	0.761	0.624	0.541	0.460	0.394	0.347
			Lower Int.	0.406	0.303	0.274	0.250	0.220	0.186	0.146	0.100	0.057
			Coverage	84.253	89.367	89.906	90.675	92.059	94.616	95.577	97.443	86.945

Table 4

Simulation Experiments Validating Table
95% Confidence Limits (t value = 2.010)

n = 50

Dist. of Death	Dist. of Censoring	trunc. parameter	distinguish / True Value	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
Exponential	Uniform	1.0	Upper Int.	0.938	0.856	0.772	0.692	0.626	0.636	**	**	**
			Lower Int.	0.300	0.302	0.290	0.268	0.238	0.215	**	**	**
			Converge	94.650	94.870	95.140	95.520	95.910	96.720	**	**	**
		1.5	Upper Int.	0.950	0.876	0.799	0.717	0.636	0.566	0.540	**	**
			Lower Int.	0.238	0.245	0.244	0.228	0.202	0.168	0.132	**	**
			Converge	94.860	95.120	95.210	95.830	95.960	96.120	96.120	**	**
		3.0	Upper Int.	0.968	0.908	0.839	0.766	0.689	0.612	0.541	0.503	0.625
			Lower Int.	0.121	0.142	0.153	0.149	0.134	0.111	0.082	0.049	0.020
			Converge	94.275	95.237	95.237	95.618	95.820	95.820	96.012	96.947	95.846
Exponential	Exponential	0.6667	Upper Int.	0.959	0.892	0.822	0.753	0.689	0.648	0.670	0.756	0.864
			Lower Int.	0.201	0.199	0.198	0.182	0.155	0.121	0.083	0.046	0.021
			Converge	92.381	94.325	94.875	95.312	95.312	96.012	96.463	96.463	95.211
		1.3333	Upper Int.	0.970	0.913	0.848	0.780	0.711	0.646	0.602	0.635	0.798
			Lower Int.	0.128	0.136	0.144	0.137	0.121	0.097	0.066	0.034	0.009
			Converge	94.568	94.761	95.327	95.327	96.032	96.032	95.882	95.882	94.895
		2.0	Upper Int.	0.974	0.920	0.858	0.791	0.721	0.653	0.594	0.585	0.734
			Lower Int.	0.103	0.114	0.124	0.120	0.107	0.087	0.061	0.030	0.007
			Converge	94.523	94.876	95.038	95.523	95.823	96.720	96.720	95.823	94.028
Uniform	Uniform	1.0	Upper Int.	0.937	0.855	0.769	0.681	0.598	0.520	0.465	0.472	0.599
			Lower Int.	0.318	0.312	0.301	0.277	0.246	0.209	0.165	0.122	0.076
			Converge	94.667	95.222	95.667	95.778	95.778	96.033	96.112	94.778	94.556
		4.0	Upper Int.	0.936	0.853	0.766	0.677	0.585	0.490	0.393	0.295	0.203
			Lower Int.	0.296	0.303	0.295	0.275	0.247	0.211	0.168	0.118	0.058
			Converge	94.826	95.362	96.012	96.723	96.723	97.023	97.023	96.052	94.234

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

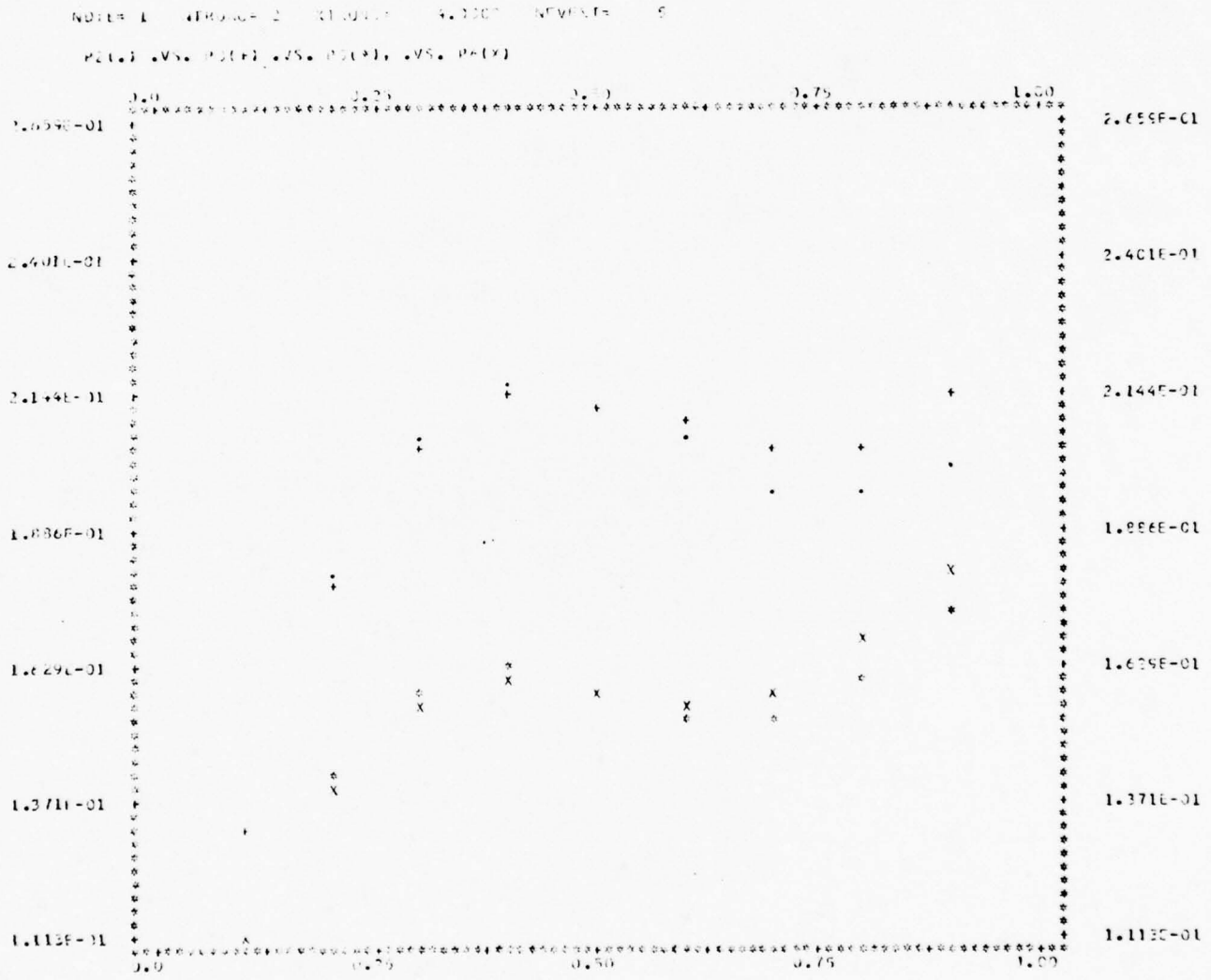


Fig. 1: Comparison of RMS derived from sample size 5.
(Step function estimators vs. point estimators)

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

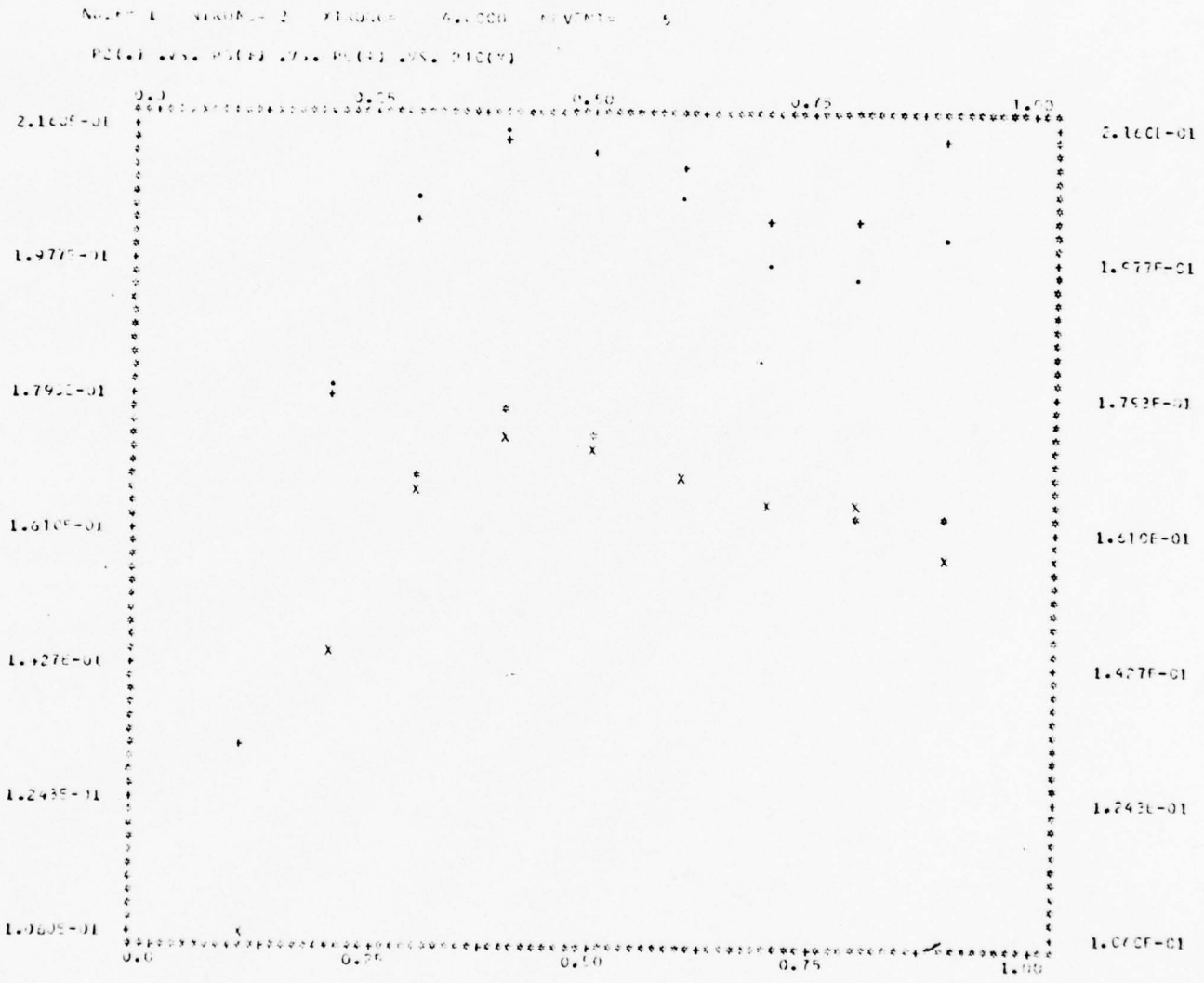


Fig. 2: Comparison of root mean square derived from sample size 5. (Step function estimators vs. jackknife estimators of logistic transformation)

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

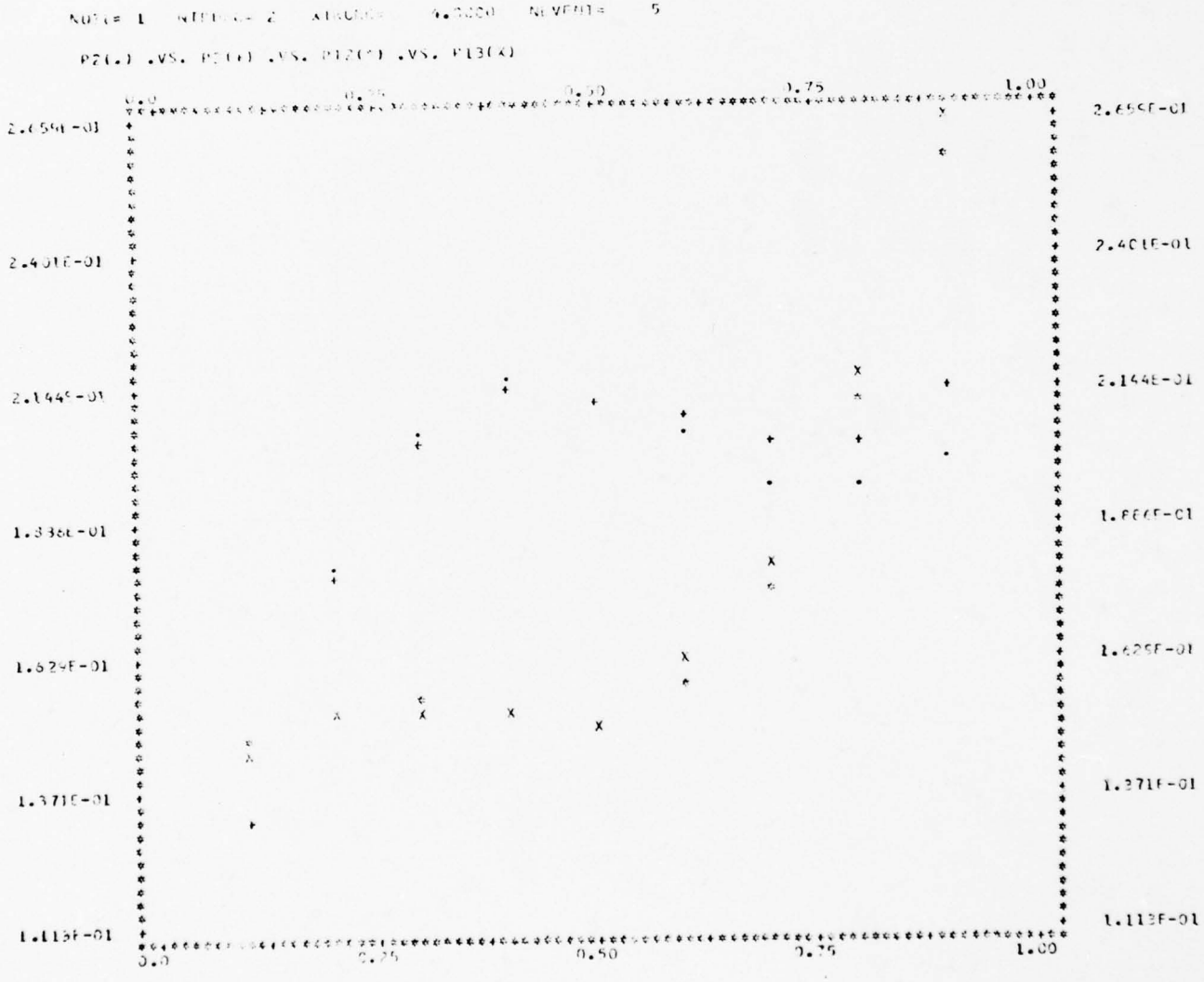


Fig. 3: Comparison of root mean square derived from sample sizes. (Step function estimators vs. Bayesian estimators)

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

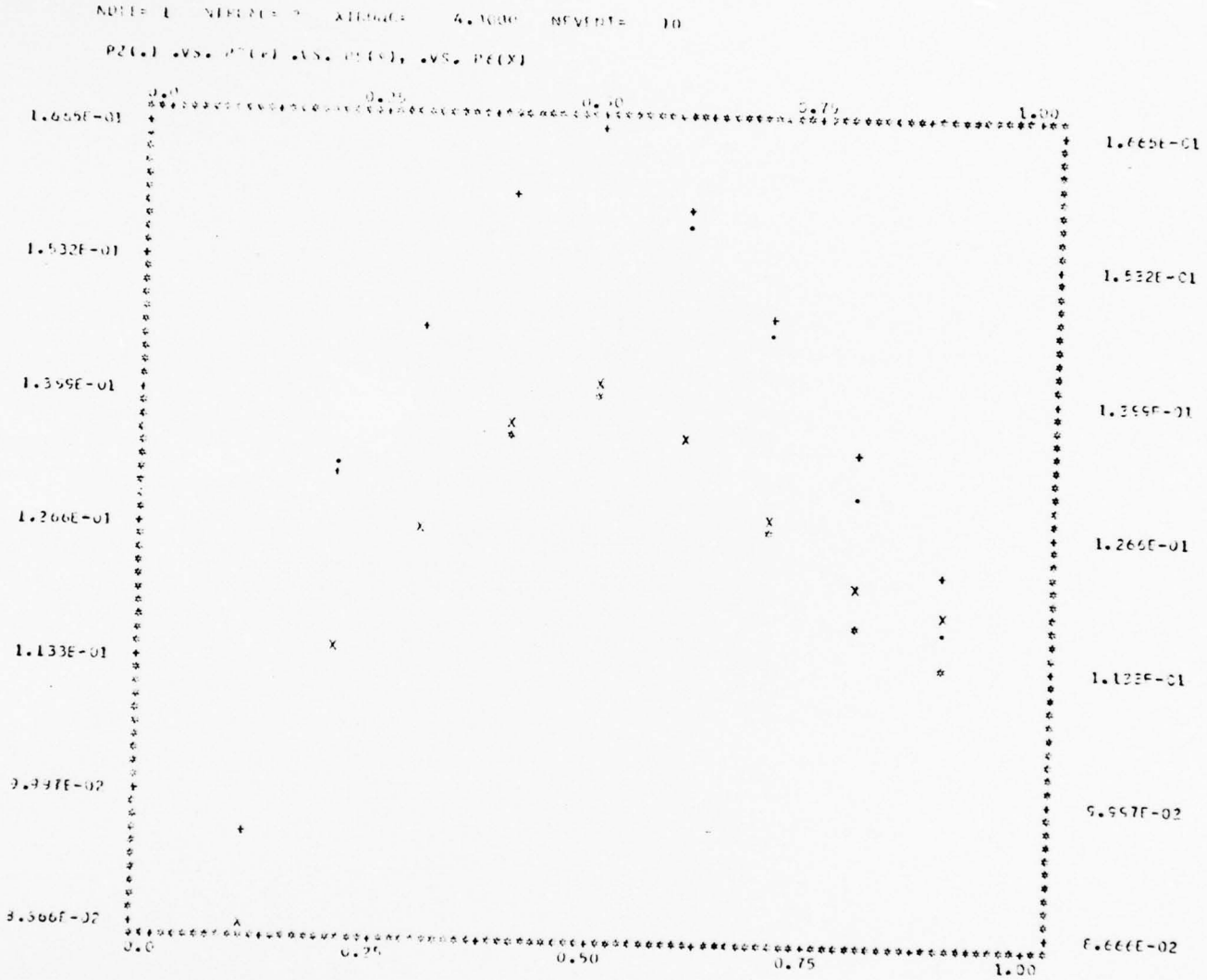


Fig. 4: Comparison of root mean square derived from sample size 10. (Step function estimators vs. point estimators)

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

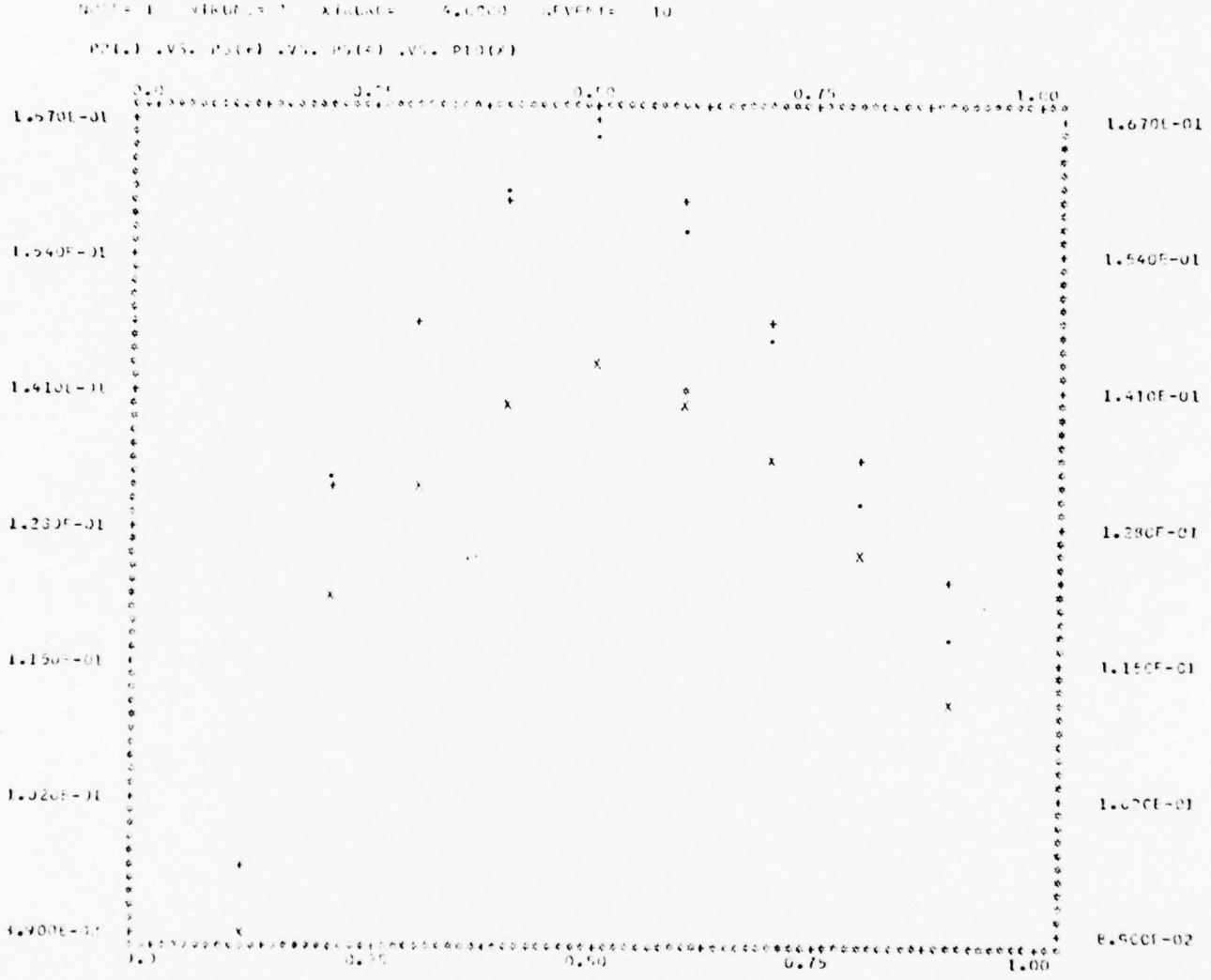


Fig. 5: Comparison of root mean square derived from sample size 10. (Step function estimators vs. jackknife estimators of logistic transformation)

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

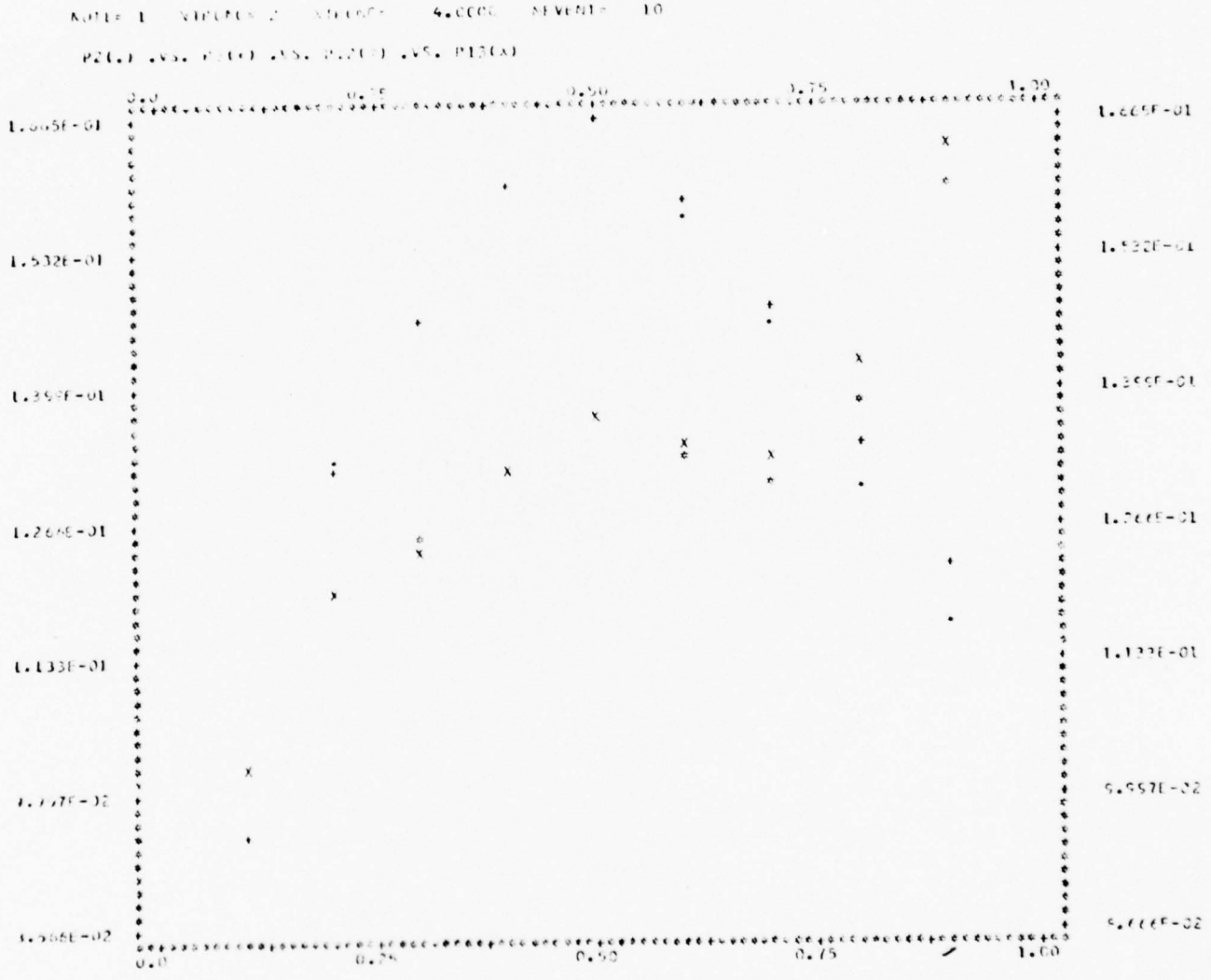


Fig. 6: Comparison of RMS derived from sample size 10.
(Step function estimators vs. Bayesian estimators)

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

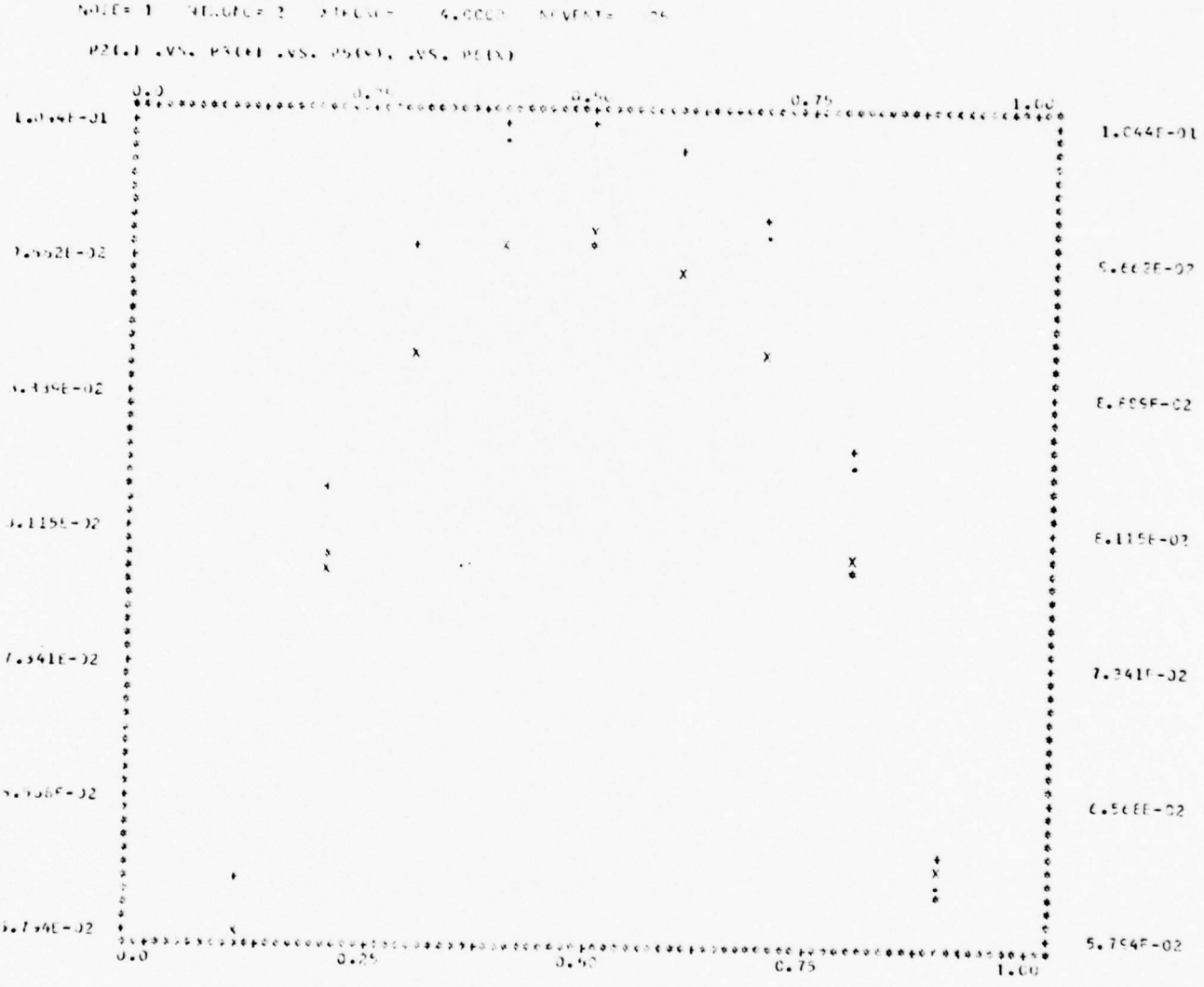


Fig. 7. Comparison of RMS derived from sample size 25.
(Step function estimators vs. point estimators)

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

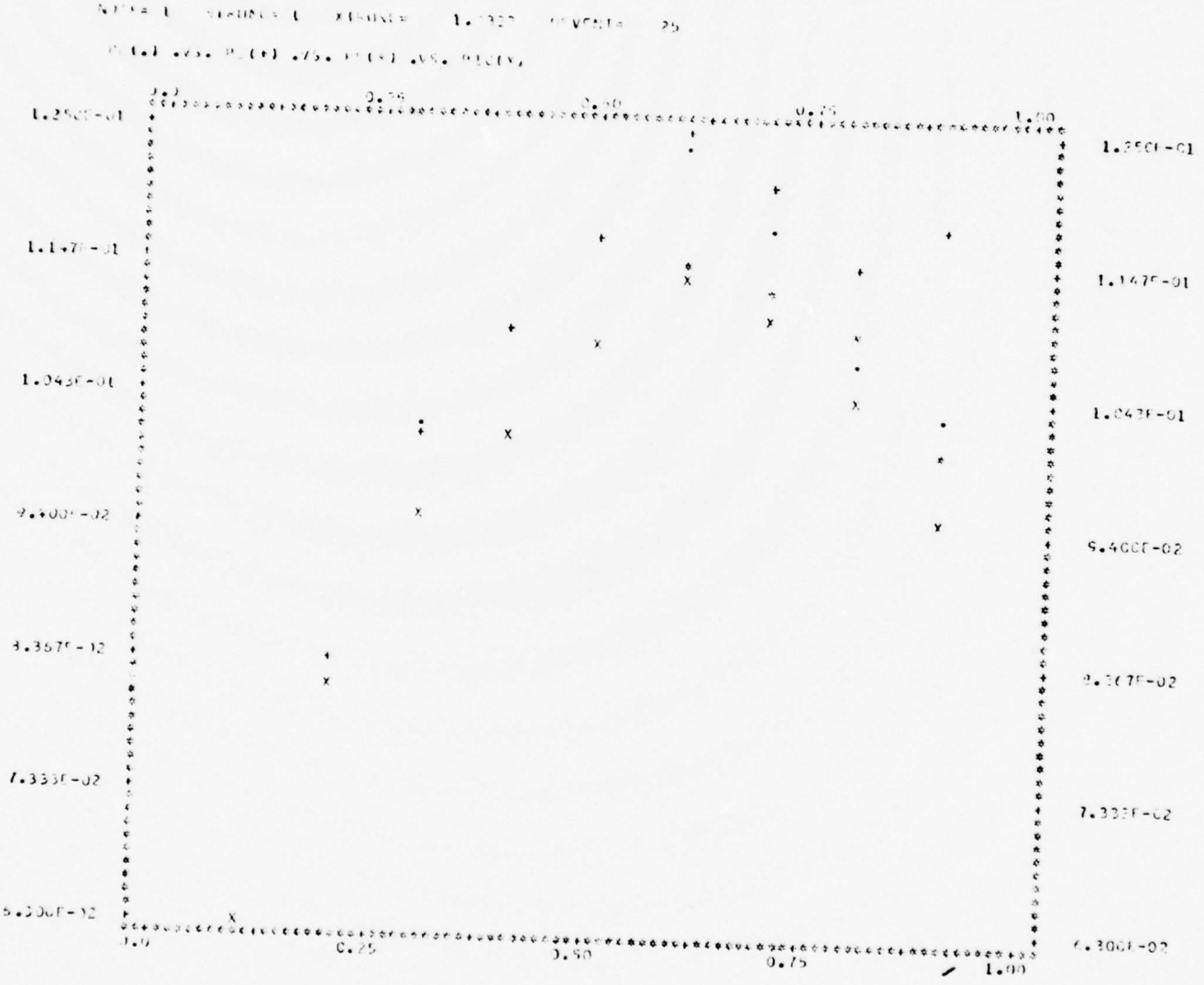


Fig. 8: Comparison of RMS derived from sample size 25.
(Step function estimators vs. jackknife estimators of logistic transformation)

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

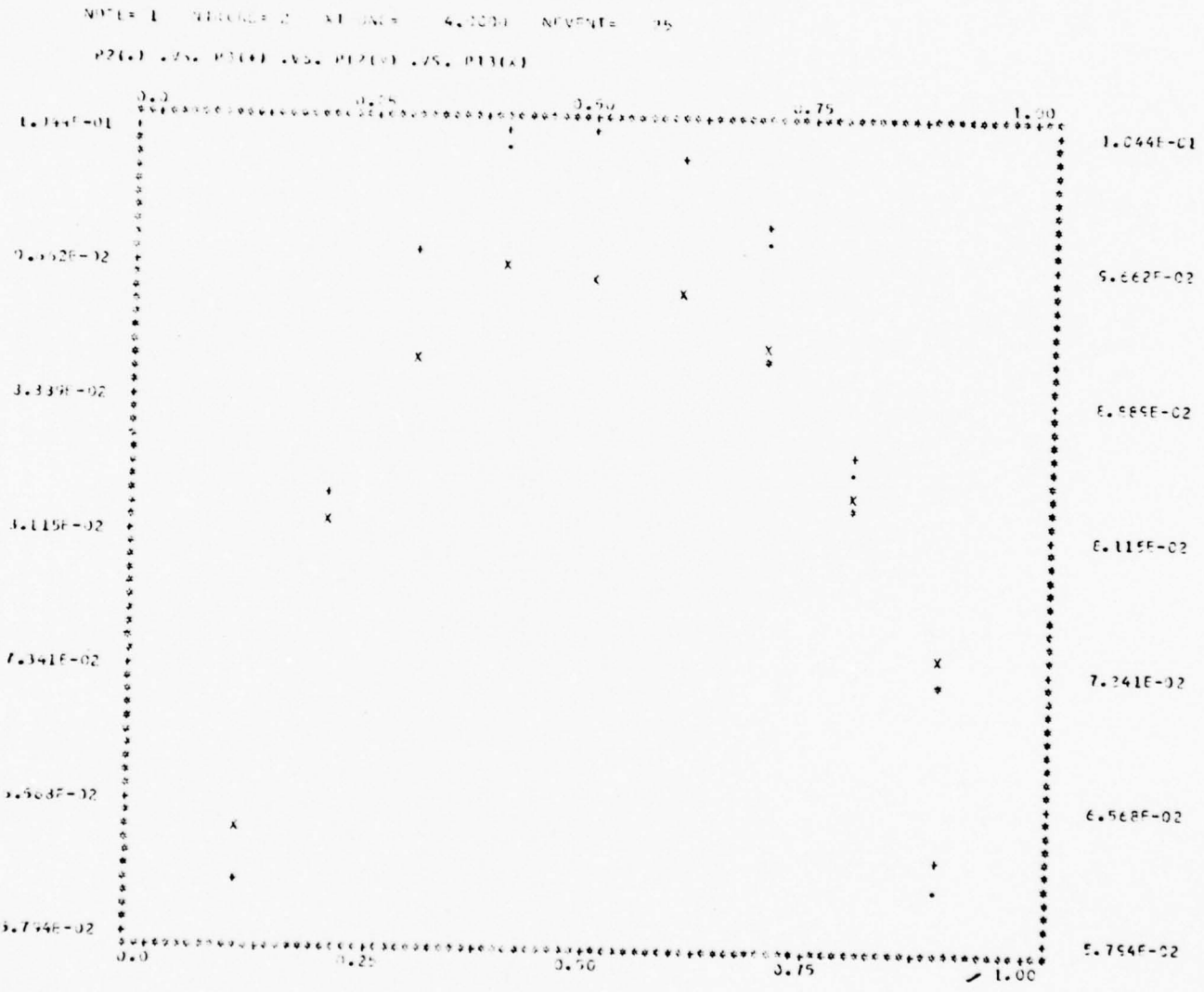


Fig. 9: Comparison of RMS derived from sample size 25.
(Step function estimators vs. Bayesian estimators)

APPENDIX A

ESTIMATORS FOR GROUPED DATA

With grouped censored data the definition of $p(t_i/t_{i-1})$ given by equation (5) does not hold unless the assumption is made that all truncations occur at the end of the time interval. If, on the other hand, it is assumed that all truncations occur at the beginning of Δt_i the equivalent form of equation (5) is

$$p(t_i/t_{i-1}) = \frac{N_i - a_i - r_i}{N_i - a_i} \qquad \text{(G-1)}$$

With N_i elements were presents at beginning of interval, i.e., at time t_{i-1}, r_i elments failed during the interval, and a_i elements truncated from the sample during the interval but prior to failing. As a hypothesis, assume that all aborts occur simultaneously somewhere within the time interval, so that r' failures occur prior to the truncations and time remaining $r_i - r'$ after the truncations. Then

$$p(t_i/t_{i-1}) = \frac{N_i - r'}{N_i} \cdot \frac{N_i - a_i - r_i}{N_i - r_i - a_i} \qquad \text{(G-2)}$$

Thus, the value of $p(t_i/t_{i-1})$ depends on when the truncations occur. It is assumed that this is not known for the grouped data case. Nevertheless, it is possible to place limits on the value of $p(t_i/t_{i-1})$ since equation (G-2) always gives values between those of equation (5) and (G-1). Thus

$$\frac{N_i - a_i - r_i}{N_i - a_i} \leq p(t_i/t_{i-1}) \leq \frac{N_i - r_i}{N_i} \qquad \text{(G-3)}$$

For average sample size approximation, a simpler expression from the point of view of computational ease may be derived by substituting a/2 for a in equation (G-1) giving

$$p(t_i) = \frac{N - \frac{a}{2} - r}{N - \frac{a}{2}} \qquad \text{(G-4)}$$

The equation (G-4) may be thought of as the result of assuming that the average number of elements in the time interval is the number at the beginning decreased by half the number of truncations.

Records are usually available to provide a fairly precise time the death events. In the medical example, the exact time of death is usually recorded in medical records required by law. In the equipment lifetesting example, the time of malfunction or failure is usually known very precisely if the results are catastrophic; and maintenance records give a reasonably precise time even if the failure is not critical to a larger system. In the military example, the event of interest is usually a sensor detection or some other action that is routinely recorded in a log book.

Equaiton (G-4) is a modification to the product-limit estimator, p_2, when the times of truncation are known only in grouped form. Herd, reference (2), suggests a similar modification to estimators using the second approach (p_5 or p_6) with aggregated truncation data. Illustrate results for this method based on the sample data base of the main test are given below. Here, of course, we do not know that individual B dropped out of observation at time 2 and that individual D dropped out at time 6. We know only that the two truncations occurred in the interval (1,3) and (3,7), respectively.

Product limit's modification is denoted by $\tilde{p}_2'(t)$ and Herd's modification is denoted by $\tilde{p}_5'(t)$.

Their results on the sample data base are as follows.

t	$\tilde{p}_2'(t)$
0-1	5/5 = 1.0
1-3	4/5 x 1.0 = 0.8
3-7	2.5/3.5 x 0.8 = 0.571
(7)	0.5/1.5 x 0.571 = 0.190

t	$\tilde{p}_5'(t)$
0	1.0
1	5/6 x 1.0 = 0.833
3	3.5/4.5 x 0.833 = 0.648
7	1.5/2.5 x 0.648 = 0.389

APPENDIX B

LISTING OF COMPUTER PROGRAM

```
C       TRUNCATED DATA PROGRAM
C
        DIMENSION NN(9),D(14),T(1000),IT(1000),P1(1000),P2(100
       *0),P3(1000),P4(1000),P5(1000),P6(1000),PJA4(1000),PJA5
       *(1000),PJA6(1000),TJ(1000),ITJ(1000),PZ(1000),S(14,9),
       *U(14,9),W(14,9),P11(1000),P12(1000),P13(1000),SL1(50,5
       *0),SL2(50,50),SL3(50,50),SBAR1(50),SBAR2(50),SBAR3(50)
       *PJ4(50,50),PJ5(50,50),PJ6(50,50),PJ2(50,50),PZ2(1000),
      )*SL4(50,50),SBAR4(50),PJA2(1000),C(14,9),RMS1(50),RMS2(
       *(50),RMS3(50),RMS4(50),UINT1(50),UINT2(50),UINT4(50),U
       *INT3(50),RINT1(50),RINT2(50),RINT3(50),RINT4(50),CF(8)
       *VARJ2(50),VARJ4(50),VARJ5(50),VARJ6(50),UP1(50),UP2(50
       *),UP3(50),UP4(50),RO1(50),RO2(50),RO3(50),RO4(50)
        CALL CVFLOW
C
     1  FORMAT (I5,I10,F10.4,5I5,F8.3)
     2  FORMAT(1X,'NDIE ERROR')
     3  FORMAT(1X,'NTRUNC ERROR')
     4  FORMAT(1X,'NREPL ERROR')
     5  FORMAT(1X,'NEVENT ERROR')
     6  FORMAT(1X,'NWRITE ERROR')
     7  FORMAT (3X,9F8.3)
     8  FORMAT (1X,2I5,14F8.5)
     9  FORMAT (1X,'P',I2,9F8.3,3X,'MEAN')
    10  FORMAT(' ')
    11  FORMAT('1')
    12  FORMAT(1X,I5,10F10.5)
    13  FORMAT (1X,          2I5,F10.4,4I5)
    14  FORMAT (1X,'P',I2,9F8.3,3X,'ABS')
    15  FORMAT (6X,9(I4,4X))
    16  FORMAT (1X,'P',I2,9F8.3,3X,'RMS')
    18  FORMAT (1X,2I5)
C
C       READ INPUTS AND SET INITIAL VALUES
C
        A=0.01
    25  READ(5,1)NDIE,NTRUNC,XTRUNC,NREPL,NEVENT,NWRITE,ISEED,
       *NEND,TN
        JEVENT=NEVENT-1
        WRITE(6,13)NDIE,NTRUNC,XTRUNC,NREPL,NEVENT,NWRITE,ISEE
       *D
        WRITE (6,10)
        NP=14
        SRH=SQRT(.5)
        DO 26 I=1,1000
        T(I)=0.0
        P1(I)=0.0
        P2(I)=0.0
        P3(I)=0.0
        P4(I)=0.0
        P5(I)=0.0
        P6(I)=0.0
        TJ(I)=0.0
        P11(I)=0.0
        P12(I)=0.0
        P13(I)=0.0
    26  PZ(I)=0.0
C
C       TEST INPUTS
C
        IF (NDIE.LT.1) GOTO 105
        IF (NDIE.GT.2) GOTO 105
        IF (NTRUNC.LT.1) GO TO 103
        IF (NTRUNC.GT.3) GOTO 103
        IF (NREPL.LT.1) GO TO 104
        IF (NREPL.GT.1000)GOTO 104
        IF(NEVENT.LT.2) GOTO 102
        IF (NEVENT.LE.1000) GOTO 200
C
C       ERROR MESSAGES
C
  102   WRITE (6,5)
```

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

```
      STOP
  103 WRITE (6,3)
      STOP
  104 WRITE (6,4)
      STOP
  105 WRITE (6,2)
  108 STOP
C
CC    START MAIN CALCULATION
C
  200 DO 250 J=1,9
      NN(J)=0
      DO 250 I=1,NP
      S(I,J)=0
      U(I,J)=0
      C(I,J)=0.0
  250 W(I,J)=0
      DO 4999 IREPL=1,NREPL
      NDI=0
      DO 999 IEVENT=1,NEVENT
C
CC    CREATE TTRUNC()
C
      CALL RANDOM(ISEED,TTR,1)
      GOTO (300,350,400),NTRUNC
      GOTO 103
  300 TTR=TTR*XTRUNC
      GOTO 500
  350 TTR=-XTRUNC*ALOG(TTR)
      GOTO 500
  400 TTR            =XTRUNC
C
CC    CREATE TDIE()
C
  500 CALL RANDOM(ISEED,TDI,1)
      GOTO (800,700),NDIE
      GOTO 102
  700 TDI=-ALOG(TDI)
C
CC    DETERMINE SMALLER OF TDIE() AND TTRUNC()
C
  800 IF (TDI.LE.TTR) GOTO 810
      TT=TTR
      ITT=0
      GOTO 850
  810 TT=TDI
      NDI=NDI+1
      ITT=1
  850 T(IEVENT)=TT
      IT(IEVENT)=ITT
C
CC    ORDER DATA
C
  870 IF (IEVENT.EQ.1) GOTO 999
      II=IEVENT-1
      DO 890 I=1,II
      IF (TT.GT.T(I)) GOTO 890
      III=II-I+1
      DO 880 J=1,III
      JJ=IEVENT-J
      IT(JJ+1)=IT(JJ)
  880 T(JJ+1)=T(JJ)
      IT(I)=ITT
      T(I)=TT
      GOTO 999
  890 CONTINUE
C
  999 CONTINUE
      TN=T(NEVENT)
      T7=0
      DII=NDI+2.
      IF (NDI.GT.0) GOTO 8122
```

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

```
      P4(NEVENT)=SRH
      P5(NEVENT)=SRH
      TT=0
      DO 8121 J=1,NEVENT
 8121 TT=TT+T(J)
      DN=TT/T(NEVENT)
      P6(NEVENT)=SQRT(DN/(DN+1.))
      GOTO 1111
C
C     CALCULATE P1() AND P4() AND P11() VECTORS AND P7 DATA
C
 8122 N=NDI
      J=0
      II=0
      III=0
      DO 2199 I=1,NEVENT
      T7=T7+T(I)
      IF(IT(I).EQ.1) GOTO 2150
      J=1
      GOTO 2199
 2150 P1(I)=FLOAT(N-1)/FLOAT(NDI)
      P4(I)=FLOAT(N)/FLOAT(NDI+1)
      P11(I)=FLOAT(N)/DII
      TTT=T(I)
      III=II
      II=I
      J=0
      N=N-1
 2199 CONTINUE
      T7=-NDI/T7
      IF (J.EQ.0) GOTO 2221
      TII=T(II)
      DTI=T(III)-TII
      IF (NDI.GT.1) GOTO2200
      TAU=-TII/ALOG(P4(II))
      TTAU=-TII/ALOG(P11(II))
      GOTO 2210
 2200 TAU= DTI/ALOG(P4(II)/P4(III))
      TTAU=(T(III)-TII)/ALOG(P11(II)/P11(III))
 2210 DT=TN-TII
      IF (DT.GT.150*TAU) TAU=DT/150
      IF(DT.GT.150*TTAU)TTAU=DT/150.
      P11(NEVENT)=P11(II)*EXP(-DT/TTAU)
      P4(NEVENT)=P4(II)*EXP(-DT/TAU)
 2221 IF(NDI.NE.NEVENT) GO TO 2225
      DO 2224 I=1,NEVENT
      P2(I)=P1(I)
      P3(I)=P1(I)
      P12(I)=P11(I)
      P13(I)=P11(I)
      P5(I)=P4(I)
 2224 P6(I)=P4(I)
      GOTO 1111
C
C     CALCULATE P2() AND P5() AND P12() VECTORS
C
 2225 PP=1.
      N=NEVENT
      PPP=FLOAT(N+1)/FLOAT(N+2)
      P=1.
      J=0
      DO 2399 I=1,NEVENT
      IF  (IT(I).EQ.1) GOTO 2350
      J=1
      GOTO 2399
 2350 PP=PP*FLOAT(N-1)/FLOAT(N)
      P=P*FLOAT(N)/FLOAT(N+1)
      PPP=PPP*FLOAT(N)/FLOAT(N+1)
      P2(I)=PP
      P5(I)=P
      P12(I)=PPP
      J=0
```

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

```
 2399 N=N-1
      IF (J.EQ.0) GOTO2425
      IF (NDI.GT.1) GOTO 2400
      TAU=-TII/ALOG(P5(II))
      TTAU=-TII/ALOG(P12(II))
      GOTO2410
 2400 TAU=DTI/ALOG(P5(II)/P5(III))
      TTAU=DTI/ALOG(P12(II)/P12(III))
 2410 IF (DT.GT.150*TAU) TAU=DT/150
      P5(NEVENT)=P*EXP(-DT/TAU)
      IF(DT.GT.150*TTAU)TTAU=DT/150.
      P12(NEVENT)=PPP*EXP(-DT/TTAU)
C
C     CALCULATE P3() AND P6() AND P13() VECTORS
C
 2425 PP=1.
      N=NEVENT
      P=1.
      PPP=FLOAT(N+1)/FLOAT(N+2)
      J=0
      TT=0
      TTT=0
      DO 2599 I=1,NEVENT
      IF (IT(I).EQ.1) GOTO 2550
      J=J+1
      TT=TT+T(I)-TTT
      GOTO 2599
 2550 DN=N
      IF (TT.NE.0) DN=DN+TT/(T(I)-TTT)
      PP=PP*(DN-1)/DN
      P=P*DN/(DN+1)
      PPP=PPP*DN/(DN+1.)
      P3(I)=PP
      P6(I)=P
      P13(I)=PPP
      J=0
      TT=0
      TTT=T(I)
 2599 N=N-1
      IF(J.EQ.0)GO TO 1111
      DTT=TN-TTT
      IF(DTT.GE.1E-70)GOTO 3005
      DN=.5*(J+1)
      GOTO 3010
 3005 DN=TT/DTT
 3010 P6(NEVENT)=P*SQRT(DN/(DN+1.))
      IF(NDI.GT.1)GO TO 1997
      TTAU=-TII/ALOG(P13(II))
      GOTO 1998
 1997 TTAU=DTI/ALOG(P13(II)/P13(III))
 1998 IF(DT.GT.150*TTAU)TTAU=DT/150.
      P13(NEVENT)=PPP*EXP(-DT/TTAU)
C
C SET UP A LOOP FOR ALL JACKKNIFE CALLS (PJ4,PJ5,PJ6)
C
 1111 DO 1000 I=1,NEVENT
      DO 1011 J=1,NEVENT
      PJ2(I,J)=0.0
      PJ4(I,J)=0.0
      PJ5(I,J)=0.0
      PJ6(I,J)=0.0
      SL1(I,J)=0.0
      SL2(I,J)=0.0
      SL3(I,J)=0.0
      SL4(I,J)=0.0
 1011 CONTINUE
      PJA2(I)=0.0
      PJA4(I)=0.0
      PJA5(I)=0.0
      PJA6(I)=0.0
      SBAR1(I)=0.0
      SBAR2(I)=0.0
```

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

```
      SBAR3(I)=0.0
      SBAR4(I)=0.0
      PZ2(I)=0.0
      VARJ2(I)=0.0
      VARJ4(I)=0.0
      VARJ5(I)=0.0
      VARJ6(I)=0.0
 1000 CONTINUE
      DO 1001 I=1,NEVENT
C
C  MOVES DATA INTO TJ() AND ITJ() VECTORS
C
      K=1
      JNEXT=0
      JBEFOR=0
      JAFTER=0
 1002 IF(K.EQ.I)GO TO 7003
      TJ(K)=T(K)
      ITJ(K)=IT(K)
      IF(IT(K).EQ.0)GO TO 7001
      JNEXT=JBEFOR
      JBEFOR=K
 7001 K=K+1
      GO TO  1002
 7002 JAFTER=JBEFOR
      JBEFOR=JNEXT
      GO TO 1010
 7003 IF(I.GT.JEVENT) GO TO 7002
 1003 IF(K.GT.JEVENT)GO TO 1004
      TJ(K)=T(K+1)
      ITJ(K)=IT(K+1)
      IF(ITJ(K).EQ.0)GO TO 4002
      IF(JAFTER.EQ.0)JAFTER=K
 4002 K=K+1
      GO TO 1003
 1004 IF(JAFTER.EQ.0)JAFTER=JEVENT
 1010 NDIJ=NDI-IT(I)
C
C    CHECK IF ZERO DEATHS
C
      IF(NDIJ.EQ.0)GOTO 1001
      N=NDIJ
      P=1.
      J=0
      II=0
      III=0
      GO TO 1014
C
C  CALCULATE PJ4()VECTORS
C
 1014 DO 1016 IJ=1,JEVENT
      IF(ITJ(IJ).EQ.1)GO TO 1015
      J=1
      GO TO 1016
 1015 P=FLOAT(N)/FLOAT(NDIJ+1)
      PZ(IJ)=P
      III=II
      II=IJ
      J=0
      N=N-1
 1016 CONTINUE
      IF(J.EQ.0)GO TO 1019
      TII=TJ(II)
      DTI=TJ(III)-TII
      IF(NDIJ.GT.1)GO TO 1017
      TAU=-TII/ALOG(PZ(II))
      GO TO 1018
 1017 TAU=DTI/ALOG(PZ(II)/PZ(III))
 1018 DT=TJ(JEVENT)-TII
      IF(DT.GT.150.*TAU)TAU=DT/150.
      PZ(JEVENT)=P*EXP(-DT/TAU)
 1019 K=1
```

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

```
 1020 IF(K.EQ.I)GO TO 2021
      PJ4(K,I)=ALOG((PZ(K)+A)/(1-PZ(K)+A))
      K=K+1
      GO TO 1020
 2021 IF(K.EQ.NEVENT)GO TO 5021
      IF(IT(I).EQ.0)GO TO 1025
 5021 TJA=TJ(JAFTER)
      IF(JBEFOR.NE.0)GO TO 1022
      PX=ALOG(PZ(JAFTER))*T(I)/TJA
      PZZ=0.0
      IF (PX.LT.-150) GOTO 7025
      PZZ=EXP(PX)
 7025 PJ4(I,I)=ALOG((PZZ+A)/(1-PZZ+A))
      GO TO 1025
 1022 TJB=TJ(JBEFOR)
      DTJ=TJA-TJB
      DTB=T(I)-TJB
      PX=DTB*ALOG(PZ(JAFTER)/PZ(JBEFOR))/DTJ
      PZZ=0.0
      IF (PX.LT.-150) GOTO 7026
      PZZ=PZ(JBEFOR)*EXP(PX)
 7026 PJ4(I,I)=ALOG((PZZ+A)/(1-PZZ+A))
 1025 IF(K.GT.JEVENT)GOTO 3027
      PJ4(K+1,I)=ALOG((PZ(K)+A)/(1-PZ(K)+A))
      K=K+1
      GO TO 1025
 3027 IF(NDI.NE.NEVENT)GOTO 1026
      DO 3028 K=1,NEVENT
      DO 3028 L=1,NEVENT
      PJ5(K,L)=PJ4(K,L)
 3028 PJ6(K,L)=PJ4(K,L)
      GOTO 1001
C
C  CALCULATE PJ5() VECTORS
C
 1026 N=JEVENT
      P=1.
      PP=1.
      J=0
      DO 1028 IJ=1,JEVENT
      IF(ITJ(IJ).EQ.1)GO TO 1027
      J=1
      GO TO 1028
 1027 P=P*FLOAT(N)/FLOAT(N+1)
      PP=PP*FLOAT(N-1)/FLOAT(N)
      PZ(IJ)=P
      PZ2(IJ)=PP
      J=0
 1028 N=N-1
      IF(J.EQ.0)GO TO 1031
      IF(NDIJ.GT.1)GO TO 1029
      TAU=-TII/ALOG(PZ(II))
      GO TO 1030
 1029 TAU=DTI/ALOG(PZ(II)/PZ(III))
 1030 IF(DT.GT.150*TAU)TAU=DT/150.
      PZ(JEVENT)=P*EXP(-DT/TAU)
 1031 K=1
 1032 IF(K.EQ.I)GO TO 2033
      PJ5(K,I)=ALOG((PZ(K)+A)/(1-PZ(K)+A))
      PJ2(K,I)=ALOG((PZ2(K)+A)/(1-PZ2(K)+A))
      K=K+1
      GO TO 1032
 2033 IF(K.EQ.NEVENT)GO TO 5033
      IF(IT(I).EQ.0)GO TO 1136
 5033 IF(JBEFOR.NE.0)GO TO 1135
      PX=T(I)*ALOG(PZ(JAFTER))/TJA
      PZZ=0.0
      IF (PX.LT.-150) GOTO 7027
      PZZ=EXP(PX)
 7027 PJ5(I,I)=ALOG((PZZ+A)/(1-PZZ+A))
      GO TO 1136
 1135 PX=DTB*ALOG(PZ(JAFTER)/PZ(JBEFOR))/DTJ
```

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

```
      PZZ=0.0
      IF (PX.LT.-150) GOTO 7028
      PZZ=PZ(JBEFCR)*EXP(PX)
 7028 PJ5(I,I)=ALOG((PZZ+A)/(1-PZZ+A))
 1136 IF(K.GT.JEVENT)GO TO 1036
      PJ5(K+1,I)=ALOG((PZ(K)+A)/(1-PZ(K)+A))
      PJ2(K+1,I)=ALOG((PZ2(K)+A)/(1-PZ2(K)+A))
      K=K+1
      GO TO 1136
C
C  CALCULATE PJ6() VECTORS
C
 1036 N=JEVENT
      P=1.
      J=0
      TT=0
      TTT=0
      DO 1038 IJ=1,JEVENT
      IF(ITJ(IJ).EQ.I)GO TO 1037
      J=1
      TT=TT+TJ(IJ)-TTT
      GO TO 1038
 1037 DN=N
      IF(TT.NE.0)DN=DN+TT/(TJ(IJ)-TTT)
      P=P*DN/(DN+1.)
      PZ(IJ)=P
      J=0
      TT=0
      TTT=TJ(IJ)
 1038 N=N-1
      IF(J.EQ.0)GO TO 1041
      DT=TJ(JEVENT)-TTT
      IF(DT.GT.1E-7)GO TO 1039
      DN=.5*(J+1)
      GO TO 1040
 1039 DN=TT/DT
 1040 PZ(JEVENT)=P*SQRT(DN/(DN+1.))
 1041 K=1
 1042 IF(K.EQ.I)GO TO 2043
      PJ6(K,I)=ALOG((PZ(K)+A)/(1-PZ(K)+A))
      K=K+1
      GO TO 1042
 2043 IF(K.EQ.NEVENT)GO TO 5043
      IF(IT(I).EQ.0)GO TO 1146
 5043 IF(JBEFCR.NE.0)GO TO 1045
      PX=T(I)*ALOG(PZ(JAFTER))/TJA
      PZZ=0.0
      IF (PX.LT.-150) GOTO 7029
      PZZ=EXP(PX)
 7029 PJ6(I,I)=ALOG((PZZ+A)/(1-PZZ+A))
      GO TO 1146
 1045 PX=DTB*ALOG(PZ(JAFTER)/PZ(JBEFCR))/DTJ
      PZZ=0.0
      IF (PX.LT.-150) GOTO 7030
      PZZ=PZ(JBEFCR)*EXP(PX)
 7030 PJ6(I,I)=ALOG((PZZ+A)/(1-PZZ+A))
 1146 IF(K.GT.JEVENT)GOTO 1001
      PJ6(K+1,I)=ALOG((PZ(K)+A)/(1-PZ(K)+A))
      K=K+1
      GO TO 1146
 1001 CONTINUE
      DO 8888 I=1,NEVENT
      DO 8811 J=1,NEVENT
      IF(I.EQ.J)GOTO 8811
      SL1(I,J)=FLOAT(NEVENT)*ALOG((P4(I)+A)/(1-P4(I)+A))-FLO
     *AT(NEVENT-1)*PJ4(I,J)
      SL2(I,J)=FLOAT(NEVENT)*ALOG((P5(I)+A)/(1-P5(I)+A))-FLO
     *AT(NEVENT-1)*PJ5(I,J)
      SL3(I,J)=FLOAT(NEVENT)*ALOG((P6(I)+A)/(1-P6(I)+A))-FLO
     *AT(NEVENT-1)*PJ6(I,J)
      SL4(I,J)=FLOAT(NEVENT)*ALOG((P2(I)+A)/(1-P2(I)+A))-FLO
     *AT(NEVENT-1)*PJ2(I,J)
```

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

```
8811 CONTINUE
8888 CONTINUE
     DO 9999 I=1,NEVENT
     DO 9911 J=1,NEVENT
     SBAR1(I)=SL1(I,J)/NEVENT+SBAR1(I)
     SBAR2(I)=SL2(I,J)/NEVENT+SBAR2(I)
     SBAR3(I)=SL3(I,J)/NEVENT+SBAR3(I)
     SBAR4(I)=SL4(I,J)/NEVENT+SBAR4(I)
9911 CONTINUE
     IF(SBAR1(I).LT.-180.)SBAR1(I)=-180.
     IF(SBAR1(I).GT.174.)SBAR1(I)=174.
     IF(SBAR2(I).LT.-180.)SBAR2(I)=-180.
     IF(SBAR2(I).GT.174.)SBAR2(I)=174.
     IF(SBAR3(I).LT.-180.)SBAR3(I)=-180.
     IF(SBAR3(I).GT.174.)SBAR3(I)=174.
     IF(SBAR4(I).LT.-180.)SBAR4(I)=-180.
     IF(SBAR4(I).GT.174.)SBAR4(I)=174.
     EL1=EXP(SBAR1(I))
     EL2=EXP(SBAR2(I))
     EL3=EXP(SBAR3(I))
     EL4=EXP(SBAR4(I))
     PJA4(I)=EL1/(1.+EL1)
     PJA5(I)=EL2/(1.+EL2)
     PJA6(I)=EL3/(1.+EL3)
     PJA2(I)=EL4/(1.+EL4)
9999 CONTINUE
     DO 8999 I=1,NEVENT
     DO 8911 J=1,NEVENT
     IF(SL4(I,J).EQ.0.0)GOTO 3111
     VARJ2(I)=(SL4(I,J)-SBAR4(I))**2/FLOAT(NEVENT-1)+VARJ2(
    *I)
3111 IF(SL1(I,J).EQ.0.0)GOTO 3112
     VARJ4(I)=(SL1(I,J)-SBAR1(I))**2/FLOAT(NEVENT-1)+VARJ4(
    *I)
3112 IF(SL2(I,J).EQ.0.0)GOTO 3113
     VARJ5(I)=(SL2(I,J)-SBAR2(I))**2/FLOAT(NEVENT-1)+VARJ5(
    *I)
3113 IF(SL3(I,J).EQ.0.0)GOTO 8911
     VARJ6(I)=(SL3(I,J)-SBAR3(I))**2/FLOAT(NEVENT-1)+VARJ6(
    *I)
8911 CONTINUE
     RMS1(I)=SQRT(VARJ2(I)/FLOAT(NEVENT))
     RMS2(I)=SQRT(VARJ4(I)/FLOAT(NEVENT))
     RMS3(I)=SQRT(VARJ5(I)/FLOAT(NEVENT))
     RMS4(I)=SQRT(VARJ6(I)/FLOAT(NEVENT))
     UP1(I)=SBAR4(I)+RMS1(I)*TN
     RO1(I)=SBAR4(I)-RMS1(I)*TN
     UP2(I)=SBAR1(I)+RMS2(I)*TN
     RO2(I)=SBAR1(I)-RMS2(I)*TN
     UP3(I)=SBAR2(I)+RMS3(I)*TN
     RO3(I)=SBAR2(I)-RMS3(I)*TN
     UP4(I)=SBAR3(I)+RMS4(I)*TN
     RO4(I)=SBAR3(I)-RMS4(I)*TN
     IF(UP1(I).GT.174.)UP1(I)=174.
     IF(UP1(I).LT.-100.)UP1(I)=-100.
     IF(RO1(I).GT.174.)RO1(I)=174.
     IF(RO1(I).LT.-100.)RO1(I)=-100.
     IF(UP2(I).GT.174.)UP2(I)=174.
     IF(UP2(I).LT.-100.)UP2(I)=-100.
     IF(RO2(I).GT.174.)RO2(I)=174.
     IF(RO2(I).LT.-100.)RO2(I)=-100.
     IF(UP3(I).GT.174.)UP3(I)=174.
     IF(UP3(I).LT.-100.)UP3(I)=-100.
     IF(RO3(I).GT.174.)RO3(I)=174.
     IF(RO3(I).LT.-100.)RO3(I)=-100.
     IF(UP4(I).GT.174.)UP4(I)=174.
     IF(UP4(I).LT.-100.)UP4(I)=-100.
     IF(RO4(I).GT.174.)RO4(I)=174.
     IF(RO4(I).LT.-100.)RO4(I)=-100.
     UINT1(I)=1./(1.+1./EXP(UP1(I)))
     RINT1(I)=1./(1.+1./EXP(RO1(I)))
     UINT2(I)=1./(1.+1./EXP(UP2(I)))
```

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

```
      RINT2(I)=1./(1.+1./EXP(RO2(I)))
      UINT3(I)=1./(1.+1/EXP(UP3(I)))
      RINT3(I)=1./(1.+1./EXP(RO3(I)))
      UINT4(I)=1./(1.+1./EXP(UP4(I)))
      RINT4(I)=1./(1.+1./EXP(RO4(I)))
 8999 CONTINUE
C
C     PRINT OUTPUT
C
 3050 IF (NWRITE.EQ.0) GOTO 3500
      WRITE(6,8)(IREPL,IT(I),T(I),P1(I),P2(I),P3(I),P4(I),P5
     *(I),P6(I),PJA2(I),PJA4(I),PJA5(I),PJA6(I),P11(I),P12(I
     *),P13(I),I=1,NEVENT)
      WRITE (6,10)
C
C     REDUCE VECTORS
C
 3500 K=1
      DO 4000 I=1,NEVENT
      IF (IT(I).EQ.1) GOTO 3900
      IF (I.NE.NEVENT) GOTO 4000
 3900 T(K)=T(I)
      P1(K)=P1(I)
      P2(K)=P2(I)
      P3(K)=P3(I)
      P4(K)=P4(I)
      P5(K)=P5(I)
      P6(K)=P6(I)
      PJA2(K)=PJA2(I)
      PJA4(K)=PJA4(I)
      PJA5(K)=PJA5(I)
      PJA6(K)=PJA6(I)
      P11(K)=P11(I)
      P12(K)=P12(I)
      P13(K)=P13(I)
      UINT1(K)=UINT1(I)
      RINT1(K)=RINT1(I)
      UINT2(K)=UINT2(I)
      RINT2(K)=RINT2(I)
      UINT3(K)=UINT3(I)
      RINT3(K)=RINT3(I)
      UINT4(K)=UINT4(I)
      RINT4(K)=RINT4(I)
      K=K+1
 4000 CONTINUE
C
C     CALCULATE DIFFERENCES, MEAN, RMS, AND MEAN ABS
C
      K=1
      TTT=0
      CI=-T(1)
      PP1=1
      PP2=1
      PP3=1
      PP4=1
      PP5=1
      PP6=1
      PP8=1
      PP9=1
      PP10=1
      PP11=(NCI+1.)/CII
      PP12=(NEVENT+1.)/(NEVENT+2.)
      PP13=PP12
      PP14=1.
      PC4=P4(1)
      PC5=P5(1)
      PC6=P6(1)
      PC8=PJA4(1)
      PC9=PJA5(1)
      PC10=PJA6(1)
      CONF1=1.
      CONF2=1.
```

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

```
      CCNF3=1.
      CCNF4=1.
      CCNF5=1.
      CCNF6=1.
      CCNF7=1.
      CCNF8=1.
      T4=DT/ALOG(PQ4)
      T5=DT/ALOG(PQ5)
      T6=DT/ALOG(PQ6)
      T8=DT/ALOG(PQ8)
      T9=DT/ALOG(PQ9)
      T10=DT/ALOG(PQ10)
      DO 4998     J=1,9
      TT=.1*J
      P=1-TT
      IF (NDIE.EQ.2)    TT=-ALOG(P)
      IF (TT.GT.TN) GOTO 4999
      NN(J)=NN(J)+1
4100  IF (TT.LE.T(K))   GOTO 4200
      PP1=P1(K)
      PP2=P2(K)
      PP3=P3(K)
      PP4=PQ4
      PP5=PQ5
      PP6=PQ6
      PP8=PQ8
      PP9=PQ9
      PP10=PQ10
      PP11=P11(K)
      PP12=P12(K)
      PP13=P13(K)
      PP14=PJA2(K)
      CCNF1=UINT1(K)
      CCNF2=RINT1(K)
      CCNF3=UINT2(K)
      CCNF4=RINT2(K)
      CCNF5=UINT3(K)
      CCNF6=RINT3(K)
      CCNF7=UINT4(K)
      CCNF8=RINT4(K)
      TTT=T(K)
      K=K+1
      PQ4=P4(K)
      PQ5=P5(K)
      PQ6=P6(K)
      PQ8=PJA4(K)
      PQ9=PJA5(K)
      PQ10=PJA6(K)
      DT=T(K)-TTT
      TT1=ALOG(PP8/PQ8)
      TT2=ALOG(PP9/PQ9)
      TT3=ALOG(PP10/PQ10)
      IF(PQ8.EQ.0.0.OR.TT1.EQ.0.0)TT1=ALOG((PP8+A)/(1-PQ8+A))
      IF(PQ9.EQ.0.0.OR.TT2.EQ.0.0)TT2=ALOG((PP9+A)/(1-PQ9+A))
      IF(PQ10.EQ.0.0.OR.TT3.EQ.0.0)TT3=ALOG((PP10+A)/(1-PQ10+A))
      T4=DT/ALOG(PP4/PQ4)
      T5=DT/ALOG(PP5/PQ5)
      T6=DT/ALOG(PP6/PQ6)
      T8=DT/TT1
      T9=DT/TT2
      T10=DT/TT3
      GOTO 4100
4200  D(1)=PP1-P
      D(2)=PP2-P
      D(3)=PP3-P
      DTT=TTT-TT
      D(4)=PP4*EXP(DTT/T4)-P
      D(5)=PP5*EXP(DTT/T5)-P
      D(6)=PP6*EXP(DTT/T6)-P
      D(7)=EXP(TT*T7)-P
      D(8)=PP8*EXP(DTT/T8)-P
      D(9)=PP9*EXP(DTT/T9)-P
```

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

```
      D(10)=PP10*EXP(DTT/T10)-P
      D(11)=PP11-P
      D(12)=PP12-P
      D(13)=PP13-P
      D(14)=PP14-P
      CF(1)=CCNF1
      CF(2)=CCNF2
      CF(3)=CCNF3
      CF(4)=CCNF4
      CF(5)=CCNF5
      CF(6)=CCNF6
      CF(7)=CCNF7
      CF(8)=CCNF8
      DO 4911 I=1,8
      CC=CF(I)
      C(I,J)=C(I,J)+CC
      IF(NWRITE.EQ.0)GOTO 4911
      WRITE(6,10)
      WRITE(6,12)I,P,(C(I,L),L=1,9)
 4911 CONTINUE
      DO 4997  I=1, NP
      DD=D(I)
      U(I,J)=U(I,J)+DD
      W(I,J)=W(I,J)+ABS(DD)
      S(I,J)=S(I,J)+DD*DD
      IF (NWRITE.EQ.0) GOTO 4997
      WRITE (6,10)
      WRITE (6,12) I,P ,(U(I,L),L=1,9)
      WRITE (6,12) I,P ,(W(I,L),L=1,9)
      WRITE (6,12) I,P ,(S(I,L),L=1,9)
 4997 CONTINUE
 4998 CONTINUE
C
 4999 CONTINUE
C
C     PRINT SUMMARY STATISTICS
C
      DO 5100 J=1,9
 5100 D(J)=.1*J
      WRITE (6,7) (D(J),J=1,9)
      WRITE (6,10)
      DO 5300 J=1,9
      IF (NN(J).GT.0) GOTO 5150
      DO 5125 I=1,NP
      S(I,J)=1E9
      U(I,J)=1E9
      C(I,J)=1E9
 5125 W(I,J)=1E9
      GOTO 5300
 5150 XJ=1./NN(J)
 5200 DO 5250 I=1,NP
      S(I,J)=SQRT(S(I,J)*XJ)
      U(I,J)=U(I,J)*XJ
      C(I,J)=C(I,J)*XJ
 5250 W(I,J)=W(I,J)*XJ
 5300 CONTINUE
      WRITE (6,9) (I,(U(I,J),J=1,9),I=1,NP)
      WRITE (6,10)
      WRITE (6,14)(I,(W(I,J),J=1,9),I=1,NP)
      WRITE (6,10)
      WRITE (6,16)(I,(S(I,J),J=1,9),I=1,NP)
      WRITE (6,10)
      WRITE(6,20)(I,(C(I,J),J=1,9),I=1,8)
   20 FORMAT(1X,'C',I1,9F8.3,3X,'CONF')
      WRITE(6,15)
      WRITE (6,15) (NN(J),J=1,9)
      IF (NEND.NE.0) GOTO 108
      WRITE (6,11)
      GOTO 25
      END
```

THIS PAGE IS BEST QUALITY PRACTICABLE
FROM COPY FURNISHED TO DDC

LIST OF REFERENCES

1. "Nonparametric Estimation from Incomplete Observation," E. L. Kaplan and Paul Meier, American Statistical Association Journal, June 1958.

2. "Electronic Reliability," G. R. Herd, ARINC Monograph No. 3, 1 May 1956.

3. Biometrics, J. W. Tukey, 1951.

4. "Jackknife Censored Data," Rupert G. Miller, Jr., Technical Report, Stanford University, 1976.

5. "Estimating Cumulative Probability from Aggregated Truncated Data," H. Spitz, Research Contribution No. 56, Center for Naval Analysis, 6 July 1964.

6. "Fortran Program to Estimate Cumulative Survival Probability and Variance," Robert L. Hubbard, OEG Internal Research Memorandum 28, 31 October 1962.

7. "Modeling and Estimating System Availability," Donald P. Gaver, NPS 55-77-4, January 1977.

8. "Nonparametric Bayesian Estimation of Survival Curves from Incomplete Observations," V. Susarla and J. Van Ryzin, Journal of the American Statistical Association, December 1906.

9. "Confidence Interval Estimation of Survival Probabilities for Censored Data," David R. Thomas and Gary L. Grunkemeir, Journal of the American Statistical Association, December 1975.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93940	2
3.	Department Chairman, Code 55 Department of Operations Research and Administrative Sciences Naval Postgraduate School Monterey, California 93940	2
4.	Professor Donald P. Gaver, Code 55Gv Department of Operations Research and Administrative Sciences Naval Postgraduate School Monterey, California 93940	5
5.	Assoc. Professor R. Neagle Forrest, Code 55 Department of Operations Research and Administrative Sciences Naval Postgraduate School Monterey, California 93940	2
6.	Dr. Robert L. Hubbard Center for Naval Analyses 1401 Wilson Blvd Arlington, Virginia 22209	1
7.	Mr. Richard B. Halley Portland State University P. O. Box 751 Portland, Oregon 97207	2
8.	Maj. Lee Won Hyung, Korean Army OR Office, G-3, R.O.K. Army HQ. Republic of Korea	5