THE FILE COPY

(4)

AD-A211 459

ARO Report 89-2

PROCEEDINGS OF THE THIRTY-FOURTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH DEVELOPMENT AND TESTING

This document contains blank pages that were not filmed.

Approved for public release; distribution unlimited.
The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

Sponsored by
The Army Mathematics Steering Committee
on Behalf of

THE CHIEF OF RESEARCH, DEVELOPMENT AND ACQUISITION

DTIC
ELECTE
AUG 15 1989
S D
B

U.S. Army Research Office

Report No. 89-2

July 1989

PROCEEDINGS OF THE THIRTY-FOURTH CONFERENCE ON THE DESIGN OF EXPERIMENTS

Sponsored by the Army Mathematics Steering Committee

HOST

U.S. Army White Sands Missile Range

White Sands Missile Range, New Mexico

HELD AT

New Mexico State University

Las Cruces, New Mexico

19-21 October 1988

Approved for public release; distribution unlimited.
The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, North Carolina

FOREWORD

The Thirty-Fourth Conference on the Design of Experiments in Army Research, Development and Testing was held on 19-21 October 1988 in the auditorium of the Physical Sciences Laboratory on the campus of New Mexico State University, Las Cruces, New Mexico. Mr. John Lockert, Director of the White Sands Missile Range, stated his installation would serve as the host for this meeting. He selected Mr. William Agee to act as the chairperson for local arrangements. The attendees appreciated the quiet and efficient manner in which this gentleman handled the many tasks associated with this event. He is also to be commended for his planning arrangements for a tutorial which was scheduled to be held two days before the start of this conference.

The original format for the Design of Experiments Conferences, which are under the auspices of the Army Mathematics Steering Committee (AMSC), was outlined by the eminent statistician, Professor Samuel S. Wilks, who served as conference chairman until his death. Through these symposia the AMSC hopes to introduce and encourage the use of the latest statistical and design techniques into the research, development and testing conducted by the Army's scientific and engineering personnel. It is believed that this purpose can be best pursued by holding these meetings at various government installations throughout the country.

Members of the program committee were pleased to obtain the services of the following distinguished scientists to speak on topics of interest to Army personnel:

Speaker and Affiliation	Title of Address
Professor Herbert A. David Iowa State University	Some Applications of Order Statistics
Professor Ronald R. Hocking Texas A&M University	Diagnostic Methods - Variance Component Estimation
Professors Donald L. Iglehart and Peter W. Glynn Stanford University	Computational and Statistical Issues in Discrete-Event Simulation
Professor Emanuel Parzen Texas A&M University	Two Sample Functional Statistical Analysis
Professor Edward L. Wegman George Mason University	Parallel Coordinate Density Plots

ccession For
TIS GRA&I ☑
TIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

QUALITY INSPECTED

Four days before the start of the planned two-day tutorial on "Topics in Modern Regression Analysis", its speaker advised Mr. Agee he could not give his planned lectures. Fortunately, Professor Ali Hadi of Cornell University was able, so to speak, to save the day. The attendees were very pleased with Dr. Hadi's interesting and informative tutorial on "Sensitivity Analysis in Linear Regression".

Dr. Marion R. Bryson, Director of the U.S. Army Combat Development Experimentation Center, was the recipient of the eighth Wilks Award for Contributions to Statistical Methodologies in Army Research, Development and Testing. This honor was bestowed on Dr. Bryson for his many significant contributions to the field of statistics. These started by providing statistical consulting while he was on the faculty of Duke University. This era was followed by full-time work devoted to directing analytical studies for the Army. Since then, he has provided overall technical direction to the Army's most modern field test facility. His published works include papers on a wide range of topics of importance to the Army, including methods for scoring casualties, designing field experiments, and inventory control problems.

The AMSC has asked that these proceedings be distributed Army-wide to enable those who could not attend this conference, as well as those that were present, to profit from some of the scientific ideas presented by the speakers. The members of the AMSC are taking this opportunity to thank all the speakers for their interesting presentations and also members of the program committee for their many contributions to this scientific event.

PROGRAM COMMITTEE

Carl Bates
Eugene Dutoit
Doug Tang
Robert Burge
Hugh McCoy
Malcolm Taylor
Henry Tingey
Francis Dressel
Carl Russell
Jerry Thomas

TABLE OF CONTENTS*

Title	Page
Foreword	iii
Table of Contents	v
Program	vii
SOME APPLICATIONS OF ORDER STATISTICS H. A. David	1
MULTI-SAMPLE FUNCTIONAL STATISTICAL DATA ANALYSIS Emanuel Parzen	15
RELIABILITY OF THE M256 CHEMICAL DETECTION KIT David W. Webb and Linda L.C. Moss	27
COMPARISON OF RELIABILITY CONFIDENCE INTERVALS Paul H. Thrasher	33
ENVIRONMENTAL SAMPLING: A CASE STUDY Dennis L. Brandon	73
A GENERALIZED GUMBEL DISTRIBUTION Siegfried H. Lehnigk	77
A GENERALIZATION OF THE EULERIAN NUMBERS WITH A PROBABILISTIC APPLICATION Bernard Harris	79
THE ANALYSIS OF MULTIVARIATE QUALITATIVE DATA USING AN ORDERED CATEGORIAL APPROACH H. B. Tingey, E. A. Morgenthein, and S. M. Free	99
A SMALL SAMPLE POWER STUDY OF THE ANDERSON-DARLING STATISTIC AND A COMPARISON WITH THE KOLMOGOROV AND THE CRAMER-VON MISES STATISTICS Linda L.C. Moss, Malcolm S. Taylor, and Henry B. Tingey	117
NONPARE, A CONSULTATION SYSTEM FOR ANALYSIS OF DATA J. C. Dumer, III, T. P. Hanratty, and M. S. Taylor	173

*This Table of Contents contains only the papers that are published in this technical manual. For a list of all papers presented at the Thirty-Fourth Conference on the Design of Experiments, see the Program of this meeting.

TABLE OF CONTENTS (continued)

Title	Page
NUMERICAL ESTIMATION OF GUMBEL DISTRIBUTION PARAMETERS Charles E. Hall, Jr.	185
EXPERIMENTAL DESIGN AND OPTIMIZATION OF BLACK CHROME SOLAR SELECTIVE COATINGS I. J. Hall and R. B. Pettit	187
DETERMINATION OF DETECTION RANGE OF MONOTONE AND CAMOUFLAGE PATTERNED FIVE-SOLDIER CREW TENTS BY GROUND OBSERVERS George Anitole and Ronald L. Johnson	189
AN EXAMPLE OF CHAIN SAMPLING AS USED IN ACCEPTANCE TESTING Jerry Thomas, Robert L. Umholtz, and William E. Baker	201
SOME NOTES ON VARIABLE SELECTION CRITERIA FOR REGRESSION MODELS (AN OVERVIEW) Eugene F. Dutoit	219
TWO-STAGE TESTING OF COMBAT VEHICLE TIRE SYSTEMS Barry A. Bodt	227
PARALLEL COORDINATE DENSITIES Edward J. Wegman	247
COMPUTATIONAL AND STATISTICAL ISSUES IN DISCRETE-EVENT SIMULATION Peter W. Glynn and Donald L. Iglehart	265
BAYESIAN INFERENCE FOR WEIBULL QUANTILES Mark G. Vangel	281
MAKING FISHER'S EXACT TEST RELEVANT Paul H. Thrasher	331
LIST OF REGISTERED ATTENDEES	359

AGENDA

THE THIRTY-FOURTH CONFERENCE ON THE DESIGN OF EXPERIMENTS
IN ARMY RESEARCH, DEVELOPMENT AND TESTING

19-21 October 1988

Host: White Sands Missile Range

Location: Physical Sciences Laboratory
New Mexico State University
Las Cruces, New Mexico

xxxxx Wednesday, 19 October xxxxx

0815 - 0915 REGISTRATION

0915 - 0930 CALLING THE CONFERENCE TO ORDER

WELCOMING REMARKS:

0930 - 1200 GENERAL SESSION I (Auditorium)

Chairman:

0930 - 1030 KEYNOTE ADDRESS:

SOME APPLICATIONS OF ORDER STATISTICS

H. A. David, Iowa State University

1030 - 1100 BREAK

1100 - 1200 TWO SAMPLE FUNCTIONAL STATISTICAL ANALYSIS

Emanuel Parzen, Texas A&M University

1200 - 1330 LUNCH

1330 - 1500 CLINICAL SESSION A

Chairperson: Carl Bates, U.S. Army Concepts Analysis Agency

Panelists: Bernard Harris, University of Wisconsin-Madison
Robert Launer, University of South Carolina
Emanuel Parzen, Texas A&M University

RELIABILITY OF THE M256 CHEMICAL DETECTION KIT

David W. Webb and Linda L.C. Moss, U.S. Army Ballistic Research Laboratory

COMPARISON OF RELIABILITY CONFIDENCE INTERVALS

Paul H. Thrasher, White Sands Missile Range

1500 - 1530 BREAK

1530 - 1710 COMBINED CLINICAL AND TECHNICAL SESSION

Chairperson: Carl Russell, U.S. Army Operational Test and Evaluation Agency

Panelists: Bernard Harris, University of Wisconsin-Madison
Ronald Hocking, Texas A&M University
Henry Tingey, University of Delaware

ENVIRONMENTAL SAMPLING: A CASE STUDY

Dennis L. Brandon, U.S. Army Engineer Waterways Experiment Station

THE GENERALIZED GUMBEL DISTRIBUTION

Siegfried H. Lehnigk, U.S. Army Missile Command

EULER NUMBERS, EULER-FROBENIUS POLYNOMIALS AND PROBABILITY

Bernard Harris, University of Wisconsin-Madison

xxxxx Thursday, 20 October xxxxx

0800 REGISTRATION

0815 - 1000 TECHNICAL SESSION 1

Chairperson: Oskar M. Essenwanger, U.S. Army Missile Command

REAP-A RADAR ERROR ANALYSIS PROGRAM

William S. Agee and Andrew C. Ellingson, White Sands Missile Range

A METHOD FOR ANALYZING MULTIVARIATE QUALITATIVE DATA USING AN ORDERED CATEGORICAL APPROACH

H.B. Tingey, E. A. Morgenthien, and S. M. Free, University of Delaware

A SMALL SAMPLE POWER STUDY OF THE ANDERSON-DARLING STATISTIC

Linda L.C. Moss, Malcolm S. Taylor, U.S. Army Ballistic Research Laboratory, and Henry B. Tingey, University of Delaware

NONPARE: A CONSULTATION SYSTEM FOR NONPARAMETRIC ANALYSIS OF DATA

Malcolm S. Taylor, John C. Dumer, III and Timothy P. Hanratty, U.S. Army Ballistic Research Laboratory

1000 - 1030 BREAK

1030 - 1200 TECHNICAL SESSION 2

Chairperson: Linda L.C. Moss, U.S. Army Ballistic Research Laboratory

NUMERICAL ESTIMATION OF DISTRIBUTION PARAMETERS

Charles E. Hall, Jr., U.S. Army Missile Command

EXPERIMENTAL DESIGN AND OPTIMIZATION OF BLACK CHROME SOLAR SELECTIVE COATINGS

I. J. Hall and R. B. Pettit, Sandia National Laboratory

MULTI-OBSERVER MULTI-TARGET VISIBILITY PROBABILITIES FOR POISSON SHADOWING PROCESSES IN THE PLANE

M. Yadin and S. Zacks, State University of New York at Binghamton

1200 - 1330 LUNCH

1330 - 1500 APPLICATION SESSION

Chairperson: John Robert Burge, Walter Reed Army Institute of Research

AN EXAMPLE OF CHAIN SAMPLING AS USED IN ACCEPTANCE TESTING

Robert L. Umholtz, Jerry Thomas and William E. Baker, U.S. Army Ballistic Research Laboratory

SOME NOTES ON MODEL SELECTION CRITERIA

Eugene Dutoit, U.S. Army Infantry School

TWO-STAGE TESTING OF COMBAT VEHICLE TIRE SYSTEMS

Barry A. Bodt, U.S. Army Ballistic Research Laboratory

1500 - 1530 BREAK

1530 - 1730 GENERAL SESSION II

Chairperson: Gerald Andersen, U.S. Army Research, Development and Standardization Group (UK)

PARALLEL COORDINATE DENSITY PLOTS

Edvard J. Wegman, George Mason University

COMPUTATIONAL AND STATISTICAL ISSUES IN DISCRETE-EVENT SIMULATION

Donald L. Iglehart and Peter W. Glynn, Stanford University

1830 - 1930 CASH BAR

1930 - 2130 BANQUET AND PRESENTATION OF WILKS AWARD

xxxxx Friday, 21 October xxxxx

0800 REGISTRATION

0815 - 0930 TECHNICAL SESSION 3

Chairperson: Barry A. Bodt, U.S. Army Ballistic Research Laboratory

BAYESIAN INFERENCE FOR WEIBULL QUANTILES

Mark Vangel, U.S. Army Materials Technology Laboratory

A QUALITY ASSURANCE TEST BASED ON P(y<x) CRITERIA

Donald Neal, Trevor Rudalevige and Mark Vangel, U.S. Army Materials Technology Laboratory

MAKING FISHER'S EXACT TEST RELEVANT

Paul H. Thrasher, White Sands Missile Range

0930 - 1000 BREAK

1000 - 1130 GENERAL SESSION III

Chairperson: Douglas B. Tang, Walter Reed Army Institute of Research; Chairman of the AMSC Subcommittee on Probability and Statistics

OPEN MEETING OF THE STATISTICS AND PROBABILITY SUBCOMMITTEE OF THE ARMY MATHEMATICS STEERING COMMITTEE

DIAGNOSTIC METHODS - VARIANCE COMPONENT ESTIMATION

Ronald R. Hocking, Texas A&M University

ADJOURN

PROGRAM COMMITTEE

Carl Bates	Robert Burge	Francis Dressel
Eugene Dutoit	Hugh McCoy	Carl Russell
Doug Tang	Malcolm Taylor	Jerry Thomas
	Henry Tingey	

SOME APPLICATIONS OF ORDER STATISTICS*

H. A. David
Department of Statistics
102D Snedecor Hall
Iowa State University
Ames, IA 50011-1210

ABSTRACT. Suppose that the random variables $X_1,\ldots,X_n$ are arranged in ascending order as $X_{1:n} \leq \ldots \leq X_{n:n}$. Then $X_{r:n}$ is called the r-th order statistic ($r = 1,\ldots,n$). Order statistics, and functions thereof, have been used extensively in such diverse areas as quality control, the estimation of parameters, life testing, data compression, selection procedures, and the study of extreme meteorological phenomena. In this paper we focus on applications of order statistics to (a) estimators that are resistant to outliers, (b) current measures of location and dispersion such as the moving median and the moving range, and (c) some problems in reliability.

1. INTRODUCTION. If the random variables $X_1,\ldots,X_n$ are arranged in ascending order of magnitude and then written as

$$X_{1:n} \leq \ldots \leq X_{r:n} \leq \ldots \leq X_{n:n},$$

we call $X_{r:n}$ the r-th order statistic (OS)($r = 1,\ldots,n$). Usually $X_1,\ldots,X_n$ are assumed to be a random sample from some underlying population.

The subject of order statistics deals with the properties and applications of these ordered random variables and of functions involving them. Examples are the extremes $X_{1:n}$ and $X_{n:n}$, the range $W_n = X_{n:n} - X_{1:n}$, the extreme deviate (from the sample mean) $X_{n:n} - \bar{X}$, and the maximum absolute deviation from the median (MAD) $\max_{i=1,\ldots,n} |X_i - M|$, where the median M equals $X_{(\frac{n+1}{2}):n}$ (n odd) and $\frac{1}{2}(X_{(\frac{n}{2}):n} + X_{(\frac{n}{2}+1):n})$ (n even).

All these statistics have important application. The extremes arise in the statistical study of droughts and floods, as well as in problems of breaking strength and fatigue failure. The range is well known to provide a quick estimator of the population standard deviation σ, whereas MAD is a more recent estimator of σ valuable because of its high resistance to wild observations (outliers). The extreme deviate is a basic tool in the detection of such outliers, large values of $(X_{n:n} - \bar{X})/\sigma$ indicating the presence of an

*Keynote Address, 34th Conference on the Design of Experiments in Army Research, Development and Testing, New Mexico State University, Las Cruces, October 19, 1988. Prepared with support from the U. S. Army Research Office.

standardized bias and mean squared error (MSE) of $L_n(\mathbf{a})$ under (2.1) can be obtained with the help of tables of the first two moments of normal order statistics in the presence of an outlier (David, Kennedy, and Knight, 1977). For example, under (2.1) the standardized bias $E_n(\lambda)$ of $L_n(\mathbf{a})$ is given by

$$\frac{E\, L_n(\mathbf{a}) - \mu}{\sigma} = \Sigma\, a_i E\left(\frac{X_{i:n} - \mu}{\sigma}\right),$$

or

$$b_n(\lambda) = \Sigma\, a_i \alpha_{i:n}(\lambda), \tag{2.3}$$

where $\alpha_{i:n}(\lambda)$ is the expected value of $X_{i:n}$ for $\mu = 0$, $\sigma = 1$. Note that $\alpha_{i:n}(0)$ is just the widely tabulated expected value $\alpha_{i:n}$ of the i-th OS, $Z_{i:n}$, in random samples of n from a standard normal population. Clearly, $\alpha_{i:n}(\lambda)$ is a strictly increasing function of λ. Also, since for $\lambda = \infty$, (2.1) leaves us with a random normal sample of size n - 1 plus an observation at ∞, we have

$$\alpha_{i:n}(\infty) = \alpha_{i:n-1}, \qquad i = 1,\ldots,n-1 \qquad \alpha_{n:n}(\infty) = \infty, \tag{2.4}$$

(and likewise $\alpha_{i:n}(-\infty) = \alpha_{i-1:n}$, $i = 2,\ldots,n$ $\qquad \alpha_{1:n}(-\infty) = -\infty$.

Some results for samples of 10 are shown in Figures 1 and 2, where

$\bar{X}_{10}$ is the sample mean $\frac{1}{10}\sum_{1}^{10} X_{i:10}$

$T_{10}(1)$ is the trimmed mean $\frac{1}{8}\sum_{2}^{9} X_{i:10}$

$W_{10}(2)$ is the Winsorized mean $\frac{1}{10}\left(2X_{3:10} + \sum_{3}^{8} X_{i:10} + 2X_{8:10}\right)$

$T_{10}(4)$ is the median $\frac{1}{2}(X_{5:10} + X_{6:10})$, etc.

The figures are confined to $\lambda \geq 0$ since results for $\lambda < 0$ follow by skew-symmetry in Fig. 1 and by symmetry in Fig. 2.

BIAS $b_n(\lambda)$. Since $a_i > 0\ \forall\, i$ we see from (2.3) that the bias is a strictly increasing function of λ for each of the estimators, and from (2.4) that

$$b_n(\infty) = \Sigma\, a_i\, \alpha_{i:n-1}. \tag{2.5}$$

This gives the numerical values placed on the right of Fig. 1. The jagged graphs are the corresponding "stylized sensitivity curves" (Tukey, 1970; Andrews et al., 1972) obtained by plotting $L_n(\alpha_{1:n-1},\ldots,\alpha_{n-1:n-1},\lambda)$ against λ. In particular, for the median we have

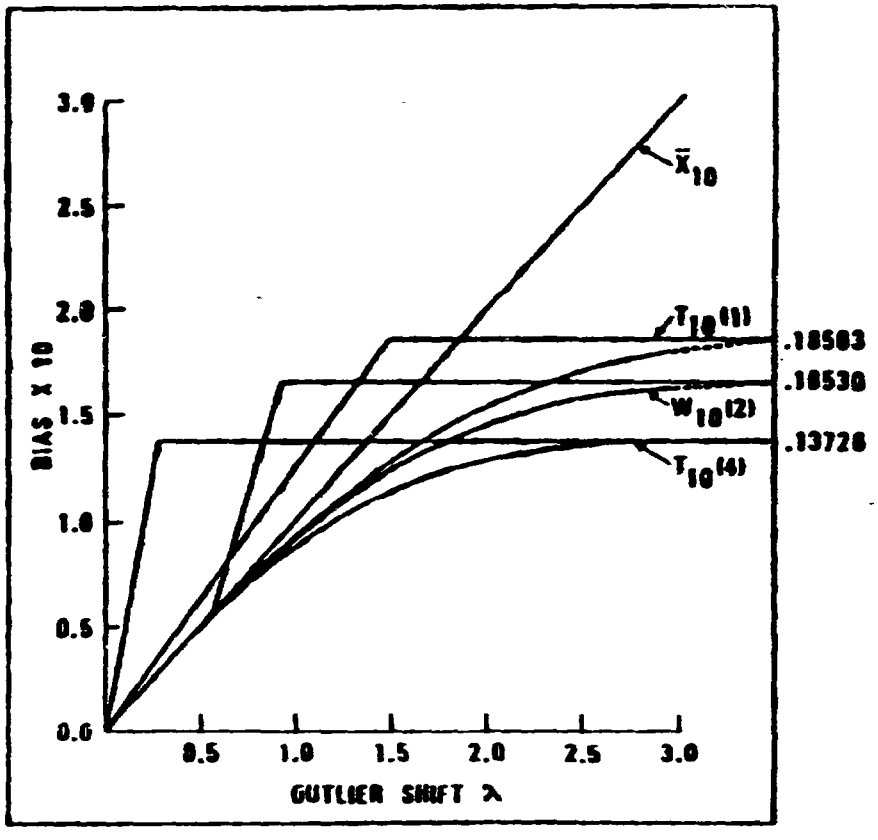


Fig. 1. Bias of various location estimators when $X_i \sim N(0, 1)$, $i = 1, 2, \ldots, 9$ and $X_{10} \sim N(\lambda, 1)$, together with the corresponding sensitivity curves (from David and Shu, 1978).

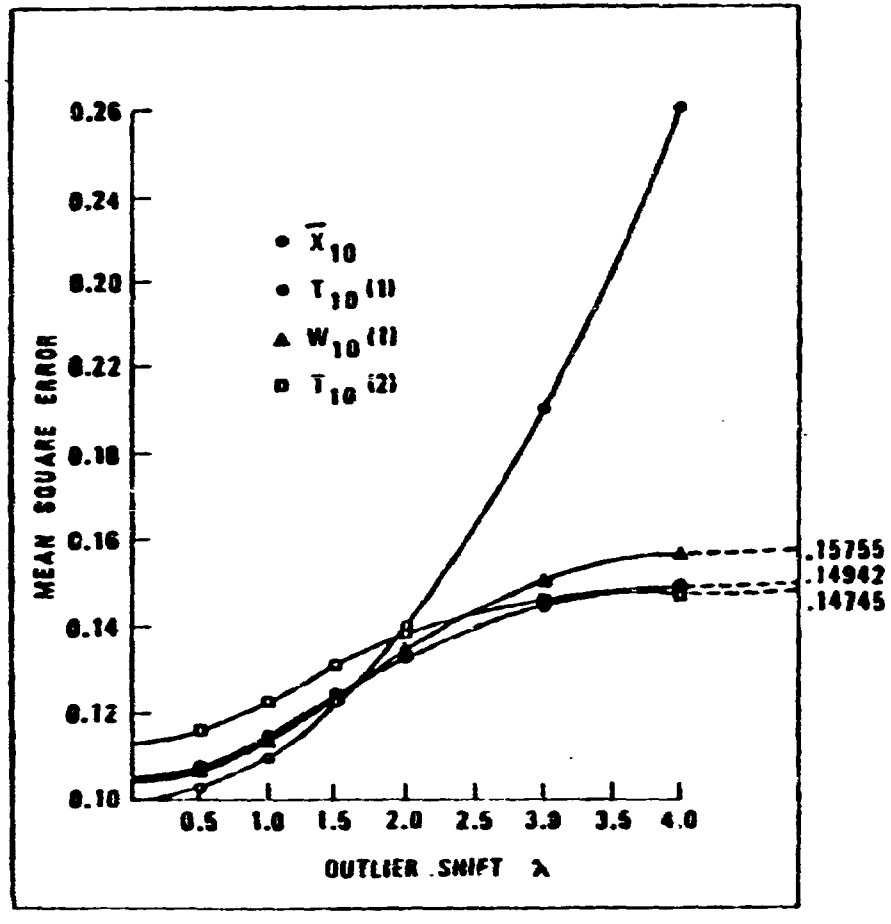


Fig. 2. Mean square error of various MSE-admissible location estimators when $X_i \sim N(0, 1)$, $i = 1, 2, \ldots, 9$ and $X_{10} \sim N(\lambda, 1)$ (from David and Shu, 1978).

$$\text{med}_{10}\,(\alpha_{1:9},\ldots,\alpha_{9:9},\lambda) = \begin{cases} \frac{1}{2}(\alpha_{5:9}+\lambda) = \frac{1}{2}\lambda & 0 < \lambda < \alpha_{6:9} \\ \frac{1}{2}(\alpha_{5:9}+\alpha_{6:9}) = \frac{1}{2}\alpha_{6:9} & \lambda > \alpha_{6:9} \\ = 0.1372\ 6 \end{cases}$$

The last result is the same as given by (2.5). In fact, each of the horizontal lines serves as an asymptote to the corresponding bias function. It is seen that the median performs uniformly best.

MEAN SQUARED ERROR MSE(λ) . No clear-cut results emerge. The sample mean does best for $\lambda < 1.5$ but is quickly outclassed for larger λ. Overall, $T_{10}(1)$ performs best, although the more highly trimmed $T_{10}(2)$ is slightly superior for very large λ.

EXTENSIONS

1. The intuitively appealing result that for symmetric unimodal distributions the median is the least biased among the L_n-estimators can be formally established under (2.1) and also for a class of symmetric non-normal distributions (David and Ghosh, 1985).

2. For $n \leq 20$ appropriately trimmed means still do well in the MSE sense when compared with much more complex estimators, but for λ sufficiently large and n not too small are inferior to the best of the adaptive estimators such as Tukey's biweight (Mosteller and Tukey, 1977, p. 205).

3. An often used alternative outlier model replaces the second line of (2.1) by

$$X_i \sim N(\mu, \tau^2\sigma^2) \qquad \tau^2 > 1$$

For this model location estimators remain unbiased but their variance is increased. Since bias has been sidestepped, only the variance of the estimator needs to be studied (David and Shu, 1978; Rosenberger and Gasko, 1983).

CASE OF SEVERAL EXTREME OUTLIERS. For $q\left(1 \leq q < \frac{1}{2}n\right)$ outliers Rocke (1986) defines as a measure of outlier resistance of an estimator of location T the "expected maximum bias" $D_T(n,q)$ by

$$D_T(n,q) = E\{\sup[\,|T(Z_{1:n-q},\ldots,Z_{n-q:n-q},\lambda_1,\ldots,\lambda_q)|\}, \qquad (2.6)$$

where the supremum is taken over all possible choices of the constants $\lambda_1,\ldots,\lambda_q$ and the Z's are the normal OS. When $T = L_n$, the supremum will evidently occur when the λ's are all $+\infty$ or all $-\infty$. As Rocke points out, by focusing on the worst case of bias one need not specify the usually unknown distribution(s) of the outliers. It suffices to model the good observations which more generally could be from any standardized distribution.

It appears that unwittingly Rocke does not use (2.6) but in fact works with the standardized bias

$$D_T^*(n,q) = E\{T(Z_{1:n-q},\ldots,Z_{n-q:n-q},\infty,\ldots,\infty)\}. \qquad (2.7)$$

If the good observations were independently generated from a unimodal symmetric distribution (mode = maximum), then again the median can be shown to have the least bias $D_T^*(n,q)$ among L_n-statistics (Rocke's proof is incorrect; see the appendix).

3. CURRENT MEASURES OF LOCATION AND DISPERSION

Let X_i be a sequence of independent random variables with cdf $F_i(x)$ $(i = 1,2,\ldots)$. Then $S_n^{(i)} = (X_i,\ldots,X_{i+n-1})$ may be called a moving sample of size n, and $X_{r:n}^{(i)}$, the r-th OS of $S_n^{(i)}$, the moving r-th OS. Moving maxima $(r = n)$ and minima $(r = 1)$ were studied by David (1955) under homogeneity $(F_i(x) = F(x),\ i = 1,2,\ldots,)$ in the course of an investigation of moving ranges $W_n^{(i)} = X_{n:n}^{(i)} - X_{1:n}^{(i)}$ $(i = 1,2,\ldots)$. The latter have a longer history (Grant, 1946), being natural companions to moving averages on quality control charts. Such charts are particularly appropriate when it takes some time to produce a single observation.

Moving medians are robust current measures of location and, like moving averages, smooth the data; see, e.g., Tukey (1977, p. 210). Cleveland and Kleiner (1975) have used the moving midmean, the mean of the central half of the ordered observations in each $S_n^{(i)}$, together with the mean of the top half and the mean of the bottom half, as three moving descriptive statistics indicating both location and dispersion changes in a time series.

Since $S_n^{(i)}$ and $S_n^{(j)}$ involve common random variables iff $d \equiv |i - j| < n$, we see that $X_{r:n}^{(i)}$ and $X_{s:n}^{(j)}$ are independent for $d \geq n$ and dependent otherwise, with $n - d$ rv's in common. To begin with, we assume homogeneity. Then the joint distribution of $S_n^{(i)}$ and $S_n^{(j)}$ will be stationary and will depend only on F(x), n, and d. We therefore consider $S_n^{(1)}$ and $S_n^{(1+d)}$, and more specifically $X_{r:n}^{(1)}$ and $X_{s:n}^{(1+d)}$ $(r,s = 1,\ldots,n)$. Let

$$\pi_{gh}(d) = \Pr\{\text{rank } (X_{r:n}^{(1)}) = g,\ \text{rank } (X_{s:n}^{(1+d)}) = h\}, \qquad (3.1)$$

where rank (Y) denotes the rank of Y in the combined sample $X_1,\ldots,X_{n+d}$. It follows that

$$E(X_{r:n}^{(1)} X_{s:n}^{(1+d)}) = \Sigma_{g,h} \pi_{gh} E(X_{g:n+d} X_{h:n+d}). \qquad (3.2)$$

This permits calculation of $\mathrm{cov}(X_{r:n}^{(1)}, X_{s:n}^{(1+d)})$ in terms of the first two moments of order statistics in samples of n+d from a distribution with cdf F(x), since the π_{gh} can be obtained by combinatorial arguments (David and Rogers, 1983). The joint distribution of $X_{r:n}^{(i)}$ and $X_{s:n}^{(j)}$ has been investigated by Inagaki (1980).

With the help of (3.2) it is possible to evaluate the auto-covariance structure under homogeneity of the moving median and, in fact, of any linear function of the order statistics $\alpha' x_{(\ )} = \alpha_1 x_{1:n} + \ldots + \alpha_n x_{n:n}$. That is, we can find

$$\mathrm{cov}\,(\alpha' x_{(\ )}^{(i)}, \alpha' x_{(\ )}^{(j)})$$

in terms of the first two moments of the OS for sample sizes up to 2n-1 from a distribution with cdf F(x).

Electrical engineers have made extensive use of moving order statistics in digital filters. They view a moving sample as a window on a sequence of signals $x_1, x_2, \ldots$ and speak of median filters when using the moving median to represent the current value of the signal, thereby "filtering out" occasional impulsive disturbances (outliers)(e.g., Arce, Gallagher, and Nodes, 1986). More generally, the median may be replaced by $\alpha' x$ to give order statistic filters (e.g., Bovik and Restrepo, 1987).

For example, suppose that in the automatic smoothing of a basically stationary time series one is prepared to ignore single outliers but wishes to be alerted to a succession of two or more high (or low) values. This calls for use of moving medians in samples of three, since clearly a single outlier will be smoothed out but two successive large values will result in two large medians. The following small example illustrates the situation, where for purposes of comparison we have added the much less informative moving mean $\bar{x}_i$.

x_i	3	1	1	10	2	4	3	9	10	2	1
$x_{2:3}^{(i)}$		1	1	2	4	3	4	9	9	2	
$\bar{x}_i$		$1\frac{2}{3}$	4	$4\frac{1}{3}$	$5\frac{1}{3}$	3	$5\frac{1}{3}$	$7\frac{1}{3}$	7	$4\frac{1}{3}$	

When $X_1, X_2, \ldots$ are not iid, even the distribution of order statistics in

a fixed sample becomes complicated although a fairly elegant expression for the pdf can be written down in terms of permanents (Vaughan and Venables, 1972) if the X's are not identically distributed but still independent. It is easily seen that the moving median and other order statistics will reflect trends except for a smoothing at the ends. Thus for the following sequence, where the upward trend is underlined, we have

x_i	5	2	1	3	4	6	9	12	14	11	7
$x^{(i)}_{2:3}$		2	2	3	4	6	9	12	12	11	
$\bar{x}_i$		$2\frac{2}{3}$	2	$2\frac{2}{3}$	$4\frac{1}{3}$	$6\frac{1}{3}$	9	$11\frac{2}{3}$	$12\frac{1}{3}$	$10\frac{2}{3}$	

For a linear trend given by

$$X_i = i\tau + Z_i, \qquad i = 1,2,\ldots \tag{3.3}$$

where the Z_i are i.i.d., we evidently have

$$\mu^{(j)}_{r:n}(\tau) - \mu^{(i)}_{r:n}(\tau) = (j-i)\tau,$$

with covariances cov $\left(X^{(i)}_{r:n}, X^{(i)}_{s:n}\right)$ $(r,s = 1,\ldots,n)$ independent of i.

Consider now a particular sample $X_1, X_2, \ldots, X_{2m-1}$ $(m = 2,3,\ldots)$ with symmetric unimodal distributions. Then under (3.3), which need hold only for the sample in question, we see that for $\tau > 0$

$$\Pr\{\text{rank } X_i = i\} \text{ increases with } \tau.$$

Thus $X_{r:n}$ will tend to lead the trend, reflect the current state, or lag the trend according as $r \gtrless m$, and will do so increasingly as τ increases; for $\tau < 0$, the results are reversed. However, in contrast to the sample mean, whose variance remains unchanged under a linear trend, the variance of the sample median increases with τ. (I am indebted to Dr. W. J. Kennedy for some computations verifying the latter intuitively obvious result.) Thus the use of the median, under locally linear trend, is appropriate primarily as protection against outliers. In this situation, but under nonlinear trend, Bovik and Naaman (1986) consider the optimal estimation of EX_m by linear functions of order statistics.

4. SOME PROBLEMS IN RELIABILITY

There is a well-known immediate connection between order statistics and the reliability of k-out-of-n systems.

Definition A k-out-of-n system is a system of n components that functions if and only if at least k $(k = 1,\ldots,n)$ of its components function. Series and parallel systems correspond to $k = n$ and $k = 1$.

Let $X_i (i = 1,\ldots,n)$ be the lifetime of the i-th component and $R_i(x) = \Pr\{X_i > x\}$ its reliability at time x (the probability that it will function at time x). Then the reliability of the system S at time x is

$$R_S(x) = \Pr\{X_{n-k+1:n} > x\}.$$

If the X_i are independent (but not necessarily identically distributed) one may write (Sen, 1970; Pledger and Proschan, 1971).

$$R_S(x) = \sum_A \prod_{i=1}^{n} R_i^{\Delta_i}(x) \left[1 - R_i(x)\right]^{1-\Delta_i},$$

where $\Delta_i = 0$ or 1, and A is the region $\sum_{i=1}^{n} \Delta_i > k$. It ca be shown that a series (parallel) system is at most (least) as reliable as the corresponding system of components each having reliability $\bar{R}(x) = \frac{1}{n} \sum_{i=1}^{n} R_i(x)$. An excellent general account, covering also important situations when the X_i are not independent, is given in Barlow and Proschan, (1975).

I will conclude with a problem in reliability, quite different from the above, that was suggested by an enquiry from Malcolm Taylor (see Baker and Taylor, 1981). A fuze contains n detonators, r of which must function within time span t. The ideal requirement r = n may be too demanding in practice and r = n-1 suffices. The n times to detonation, $X_1,\ldots,X_n$, may reasonably be regarded, I was told, as a random sample from a normal population. Let P(r; n,t) be the probability that at least r detonations have occurred in time t.

Now, for a random sample from any continuous distribution with cdf F(x), P(n;n,t) is just

$$\Pr\{X_{n:n} - X_{1:n} \leq t\} = n\int_{-\infty}^{\infty} [F(x+t) - F(x)]^{n-1}\, dF(x),$$

the cdf of the sample range (Hartley, 1942). Let A_1' and A_2 be the events $X_{n-1:n} - X_{1:n} \leq t$ and $X_{n:n} - X_{2:n} \leq t$, respectively. Then,

$$\begin{aligned} P(n-1; n,t) &= \Pr\{A_1' \cup A_2\} \\ &= \Pr\{A_1'\} + \Pr\{A_2\} - \Pr\{A_1'A_2\} \qquad (4.1) \end{aligned}$$

The event A_1' occurs if n-1 or n-2 of $X_1,\ldots,X_n$ fall in the interval $(X_{1:n}, X_{1:n} + t]$ and A_2 if n-2 of the X_i are in $(X_{2:n}, X_{2:n} + t]$. Since A_2 includes the event that n-1 of the X_i are in $(X_{1:n}, X_{1:n} + t]$, we can avoid unnecessary duplication by replacing A_1' in (4.1) by A_1, the event that exactly n-2 of the X_i are in $(X_{1:n}, X_{1:n} + t]$.

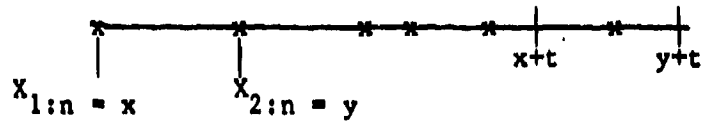
We have immediately, writing $n^{(j)} = n(n-1) \ldots (n-j+1)$, that

$$\Pr\{A_1\} = n^{(2)} \int_{-\infty}^{\infty} [F(x+t) - F(x)]^{n-2} [1 - F(x+t)]\, dF(x)$$

and

$$\Pr\{A_2\} = n^{(2)} \int_{-\infty}^{\infty} [F(x+t) - F(x)]^{n-2}\, F(x) dF(x).$$

The joint occurrence of A_1 and A_2 is illustrated below for $n = 6$. We have



$$\Pr\{A_1 A_2\} = n^{(3)} \int_{-\infty}^{\infty} \int_{x}^{x+t} [F(x+t) - F(y)]^{n-3} [F(y+t) - F(x+t)]\, dF(y) dF(x)$$

From these results $P(n-1; n,t)$ has been tabulated in David and Kinyon (1983) when $F(x) = \Phi(x)$. Note that $P(n-1;n,t)$ may be interpreted as the probability that at least n-1 out of n independent normal $N(\mu,\sigma^2)$ variates are within an interval of length $t\sigma$.

EXAMPLE. As in Baker and Taylor (1981) suppose that $X_1,\ldots,X_n$ are independent normal variates with $\sigma = 10^{-5}$. The entry $P(6;7,3) = 0.9587$ tells us that the probability of at least six detonations out of a possible seven within time span 3σ is 0.9587. By comparison, the probability of seven detonations is only 0.6601, as found from tables of the cdf of the range (Pearson and Hartley, 1970).

David and Kinyon (1983) also give an expression, involving a triple integral, for $P(n-2; n,t)$. It should be noted that $P(r; n,t)$ has received much attention by quite different techniques in the special case when the X_i are independent uniform variates (e.g., Neff and Naus, 1980). From a different viewpoint again, writing

$$P(r; n,t) = \Pr\{\hat{H}_n(\tfrac{r}{n}) \leq t\},$$

where

$$\hat{H}_n(\tfrac{r}{n}) = \min_{i=1,\ldots,n-r+1} (X_{i+r-1:n} - X_{i:n}),$$

we may regard $\hat{H}_n(\frac{r}{n})$ as a measure of dispersion. In fact, $\hat{H}_n(\alpha)$ is the length of the shorth, the shortest α-fraction of the ordered sample (Andrews et al., 1972). It has recently been shown (Grübel, 1988) that $\hat{H}_n(\alpha)$ is asymptotically normal (for fixed α).

APPENDIX

H. A. David and C. C. Yang

Correction to 'Outlier resistance in small samples'

By DAVID M. ROCKE

Biometrika (1986), 73, 175-81

The author does not stay with his own definition of $D_T(n,q)$ but in fact uses

$$D_T(n,q) = E\{T(z_1,\ldots,z_{n-q},\infty,\ldots,\infty)\}.$$

Even with this change the proof of the theorem on p. 176 is in error since the combinatorial term associated with δ_{n-r} should be $\binom{n-q}{n-r}$, not $\binom{n-q}{r}$. However, since $\delta_{n-r} = \delta_{r-q}$, the theorem follows directly from Case 2 of David and Groeneveld (Biometrika (1982), 69, 227-32) and has essentially been proved in P. K. Sen (Ed.) Biostatistics (1985) North-Holland, pp. 309-11.

REFERENCES

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. (1972). Robust Estimates of Locations. Princeton University Press.

Arce, G. R., Gallagher, N. C., and Nodes, T. A. (1986). Median filters: Theory for one- and two-dimensional filters. In: Advances in Computer Vision and Image Processing, Vol. 2, pp. 89-166.

Baker, W. E. and Taylor, M. S. (1981). An order statistic approach to Fuze design. Tech. Rept. ARBRL-TR-02313, U. S. Army Research and Development Command, Ballistic Research Lab, Aberdeen Proving Ground, Md.

Barlow, R. E. and Proschan, F. (1975). Statistical Theory of Reliability and Life Testing: Probability Models. Holt, Rinehart, and Winston, New York.

Barnett, V. and Lewis, T. (1984). Outliers in Statistical Data. 2nd edn. Wiley, New York.

Bovik, A. C. and Naaman, L. (1986). Least-squares signal estimation using order statistic filters. Proc. 20th Ann. Conf. Info. Sci. Syst., pp. 735-39.

Bovik, A. C. and Restrepo, A. (1987). Spectral properties of moving L-estimates of independent data. J. Franklin Inst. 324, 125-37.

Cleveland, W. S. and Kleiner, B. (1975). A graphical technique for enhancing scatterplots with moving statistics. Technometrics 17, 447-54.

David, H. A. (1955). A note on moving ranges. Biometrika 42, 512-15.

David, H. A. (1981). Order Statistics. 2nd edn. Wiley, New York.

David, H. A., Kennedy, W. J., and Knight, R. D. (1977). Means, variances, and covariances of normal order statistics in the presence of an outlier. Selected Tables in Mathematical Statistics. 5, 75-204.

David, H. A. and Kinyon, L. C. (1983). The probability that out of n events at least r(>n-2) occur within time span t. In: Sen, P. K. (Ed.), Contributions to Statistics, pp. 107-13, North-Holland, Amsterdam.

David, H. A. and Rogers, M. P. (1983). Order statistics in overlapping samples, moving order statistics and U-statistics. Biometrika, 70, 245-9.

David, H. A. and Shu, V. S. (1978). Robustness of location estimators in the presence of an outlier. In: David, H. A. (Ed.), Contributions to Survey Sampling and Applied Statistics: Papers in Honor of H. O. Hartley, pp. 235-50, Academic Press, New York.

Galambos, J. (1987). The Asymptotic Theory of Extreme Order Statistics. Krieger, Malabar, Florida.

Grant, E. L. (1946). Statistical Quality Control. McGraw-Hill, New York.

Grubbs, F. E. (1950). Sample criteria for testing outlying observations. Ann. Math. Statist. 21, 27-58.

Grübel, R. (1988). The length of the shorth. Ann. Statist. 16, 619-28.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). Robust Statistics. Wiley, New York.

Inagaki, N. (1980). The distributions of moving order statistics. In: Matusita, K. (Ed.), Recent Developments in Statistical Inference and Data Analysis, pp. 137-42. North-Holland, Amsterdam.

Mosteller, F. and Tukey, J. W. (1977). Data Analysis and Regression. Addison-Wesley, Reading, Massachusetts.

Neff, N. D. and Naus, J. I. (1980). The distribution of the size of the maximum cluster of points on a line. Selected Tables in Mathematical Statistics 6, 1-207.

Pearson, E. S. and Hartley, H. O. Biometrika Tables for Statisticians, Vol. I, 3rd Ed. (with additions). Cambridge University Press.

Pledger, G. and Proschan, F. (1971). Comparisons of order statistics from heterogeneous distributions. In: Rustagi, J. S. (Ed.) Optimizing Methods in Statistics, pp. 89-113. Academic Press, New York.

Rocke, D. M. (1986). Outlier resistance in small samples. Biometrika 73, 175-82.

Rosenberger, J. L. and Gasko, M. (1983). Comparing location estimators: Trimmed means, medians, and trimean. In: Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (Eds.) Understanding Robust and Exploratory Data Analysis, pp. 297-338, Wiley, New York.

Sen, P. K. (1970). A note on order statistics for heterogeneous distributions. Ann. Math. Statist. 41, 2137-9.

Tukey, J. W. (1970). Exploratory Data Analysis. (Limited Preliminary Edition) Addison-Wesley, Reading, Massachusetts.

Tukey, J. W. (1977). Exploratory Data Analysis. Addison-Wesley, Reading, Massachusetts.

MULTI-SAMPLE FUNCTIONAL STATISTICAL DATA ANALYSIS

Emanuel Parzen
Department of Statistics
Texas A&M University
College Station, Texas 77843-3143

ABSTRACT. This paper discusses a functional approach to the problem of comparison of multi-samples (two samples or c samples, where $c \geq 2$). The data consists of c random samples whose probability distributions are to be tested for equality. A diversity of statistics to test equality of c samples are presented in a unified framework with the aim of helping the researcher choose the optimal procedures which provide greatest insight about how the samples differ in their distributions. Concepts discussed are: sample distribution functions; ranks; mid-distribution function; two- sample t test and nonparametric Wilcoxon test; multi-sample analysis of variance and Kruskal Wallis test; Anderson Darling and Cramer von Mises tests; components and linear rank statistics; comparison distribution and comparison density functions, especially for discrete distributions; components with orthogonal polynomial score functions; chi-square tests and their components.

1. INTRODUCTION. We assume that we are observing a variable Y in c cases or samples (corresponding to c treatments or c populations). The samples can be regarded as the value of c variables $Y_1, \ldots, Y_c$ with respective true distribution functions $F_1(y), \ldots, F_c(y)$ and quantile functions $Q_1(u), \ldots, Q_c(u)$. We call $Y_1, \ldots, Y_c$ the conditioned variables (the value of Y in different populations).

The general problem of comparison of conditioned random variables is to model how their distribution functions vary with the value of the conditioning variable $k = 1, \ldots, c$, and in particular to test the hypothesis of homogeneity of distributions:

$$H_0 : F_1 = \ldots = F_c = F$$

The distribution F to which all the others are equal is considered to be the unconditional distribution of Y (which is estimated by the sample distribution of Y in the pooled sample).

2. DATA. The data consists of c random samples

$$Y_k(j), j = 1, \ldots, n_k$$

for $k = 1, \ldots, c$. The pooled sample, of size $N = n_1 + \ldots + n_c$, represents observations of the pooled (or unconditional) variable Y. The c samples are assumed to be independent of each other.

3. SAMPLE DISTRIBUTION FUNCTIONS. The sample distribution functions of the samples are defined (for $-\infty < y < \infty$) by

$$F_k\tilde{}(y) = \text{ fraction } \leq y \text{ among } Y_k(.).$$

The unconditional or pooled sample distribution of Y is denoted

$$F\hat{}(y) = \text{ fraction } \leq y \text{ among } Y_k(.), k = 1, \ldots, c.$$

We use $\hat{}$ to denote a smoother distribution to which we are comparing a more raw distribution which is denoted by a $\tilde{}$. An expectation (mean) computed from a sample is denoted $E\tilde{}$.

Research Supported by the U.S. Army Research Office

4. RANKS, MID-RANKS, AND MID-DISTRIBUTION FUNCTION. Nonparametric statistics use ranks of the observations in the pooled sample; let

$$R_k(t) \text{ denote the rank in the pooled sample of } Y_k(t).$$

One can define $R_k(t) = NF\hat{}(Y_k(t))$.

In defining linear rank statistics one transforms the rank to a number in the open unit interval, usually $R_k(t)/(N+1)$. We recommend $(R_k(t) - .5)/N$. These concepts assume all observations are distinct, and treat ties by using average ranks. We recommend an approach which we call the "mid-rank transform" which transforms $Y_k(t)$ to $P\hat{}(Y_k(t))$, defining the mid-distribution function of the pooled sample Y by

$$P\hat{}(y) = F\hat{}(y) - .5p\hat{}(y).$$

We call

$$p\hat{}(y) = \text{ fraction equal to } y \text{ among pooled sample}$$

the pooled sample probability mass function.

5. SAMPLE MEANS AND VARIANCES. When the random variables are assumed to be normal the test statistics are based on the sample means (for $k = 1, \ldots, c$)

$$Y_k{}^- = E\tilde{}[Y_k] = (1/n_k)\sum_{t=1}^{n_k} Y_k(t).$$

We interpret $Y_k{}^-$ as the sample conditional mean of Y given that it comes from the kth population. The unconditional sample mean of Y is

$$Y^- = E\tilde{}[Y] = p_{.1}Y_1{}^- + \ldots + p_{.c}Y_c{}^-,$$

defining

$$p_{.k} = n_k/N$$

to be the fraction of the pooled sample in the kth sample; we interpret it as the empirical probability that an observation comes from the kth sample.

The unconditional and conditional variances are denoted

$$\text{VAR}\tilde{}[Y] = (1/N)\sum_{k=1}^{c}\sum_{j=1}^{n_k}\{Y_k(j) - Y^-\}^2$$

$$\text{VAR}\tilde{}[Y_k] = (1/n_k)\sum_{j=1}^{n_k}\{Y_k(j) - Y_k{}^-\}^2$$

Note that our divisor is the sample size N or n_k rather than $N - c$ or $n_k - 1$. The latter then arise as factors used to define F statistics.

We define the pooled variance to be the mean conditional variance:

$$\sigma\hat{}^2 = \sum_{k=1}^{c} p_{.k}\ \text{VAR}\tilde{}[Y_k]$$

6. TWO SAMPLE NORMAL T TEST. In the two sample case the statistic to test H_0 is usually stated in a form equivalent to

$$T = \{Y_1^- - Y_2^-\}/\sigma\hat{}\{(N/(N-2))((1/n_1) + (1/n_2))\}^{.5}$$

We believe that one obtains maximum insight (and analogies and extensions) by expressing T in the form which compares Y_1^- with Y^-:

$$T = \{(N-2)p_{.1}/(1-p_{.1})\}^{.5}\{Y_1^- - Y^-\}/\sigma\hat{}$$

The exact distribution of T is $t(N-2)$, t-distribution with $N-2$ degrees of freedom.

7. TWO-SAMPLE NONPARAMETRIC WILCOXON TEST. To define the popuiar Wilcoxon non-parametric statistic to test H_0 we define W_k to be the sum of the n_k ranks of the Y_k values; its mean and variance are given by

$$E[W_k] = n_k(N+1)/2,\ \mathrm{VAR}[W_k] = n_1n_2(N+1)/12$$

The usual definition of the Wilcoxon test statistic is

$$T_k = \{W_k - E[W_k]\}/\{\mathrm{VAR}[W_k]\}^{.5}.$$

The approach we describe in this paper yields as the definition of the nonparametric Wilcoxon test statistic (which can be verified to approximately equal the above definition of T_1, up to a factor $\{1-(1/N)^2\}^{.5}$)

$$T_1 = \{12(N-1)p_{.1}/(1-p_{.1})\}^{.5}(R_1^- - .5\},$$

defining

$$R_1^- = (1/n_1)\sum_{t=1}^{n_1}(R_1(t) - .5)/N$$
$$= (W_1/n_1N) - (1/2N)$$

One reason we prefer this form of expressing non-parametric statistics is because of its relation to mid-ranks;

$$R_k^- = E^-[P\hat{}(Y_k)]$$

One should notice the analogy between our expressions for the parametric test statistic T and the nonparametric test statistic T_1; the former has an exact $t(N-2)$ distribution and the latter has asymptotic distribution Normal$\{0,1\}$.

8. TEST OF EQUALITY OF c SAMPLES NORMAL CASE. The homogeneity of c samples is tested in the parametric normal case by the analysis of variance which starts with a fundamental identity which in our notation is written

$$\mathrm{VAR}^-[Y] = \sum_{k=1}^{c} p_{.k}\{Y_k^- - Y^-\}^2 + \sigma\hat{}^2$$

The F test of the one-way analysis of variance can be expressed as the statistic

$$T^2 = \sum_{k=1}^{c} p_{.k}|T_k|^2,$$
$$= \sum_{k=1}^{c}(1-p_{.k})|TF_k|^2,$$

defining

$$T_k = (N-c)\{Y_k^- - Y^-\}/\hat{\sigma}$$
$$TF_k = \{(N-c)p_{.k}/(1-p_{.k})\}^{.5}\{Y_k^- - Y^-\}/\hat{\sigma}$$

The asymptotic distribution of $T^2/(c-1)$ and TF_k^2 are $F(c-1, N-c)$ and $F(1, N-c)$ respectively.

9. TEST OF EQUALITY OF c SAMPLES NONPARAMETRIC KRUSKAL-WALLIS TEST. The Kruskal-Wallis nonparametric test of homogeneity of c samples can be shown to be

$$TKW^2 = \sum_{k=1}^{c}(1-p_{.k})|TKW_k|^2.$$
$$TKW_k = \{12(N-1)p_{.k}/(1-p_{.k})\}^{.5}\{R_k^- - .5\}$$

The asymptotic distributions of TKW^2 and TKW_k^2 are chi-squared with $c-1$ and 1 degrees of freedom respectively.

10. COMPONENTS. We have represented the analysis of variance test statistic T^2 and the Kruskal-Wallis test statistic TKW^2 as weighted sums of squares of statistics TF_k and TKW_k respectively which we call components, since their values should be explicitly calculated to indicate the source of the significance (if any) of the overall statistics. Other test statistics that can be defined can be shown to correspond to other definitions of components.

11. ANDERSON DARLING AND CRAMER VON MISES TEST STATISTICS. Important among the many test statistics which have been defined to test the equality of distributions are the Anderson-Darling and Cramer-von Mises test statistics. They will be introduced below in terms of representations as weighted sums of squares of suitable components.

12. COMPARISON DISTRIBUTION FUNCTIONS AND COMPARISON DENSITTY FUNCTIONS. We now introduce the key concepts which enable us to unify and choose between the diverse statistics available for comparing several samples. To compare two continuous distributions $F(.)$ and $H(.)$, where H is a true or smooth and F is a model or raw, we define the comparison distribution function

$$D(u) = D(u; H, F) = F(H^{-1}(u))$$

with comparison density

$$d(u) = d(u; H, F) = D'(u) = f(H^{-1}(u))/h(H^{-1}(u)).$$

Under $H_0 : H = F$, $D(u) = u$ and $d(u) = 1$. Thus testing H_0 is equivalent to testing $D(u)$ for uniformity.

Sample distribution functions are discrete. The most novel part of this paper is that we propose to form an estimator $\tilde{D}(u)$ from estimators $\tilde{H}(.)$ and $\tilde{F}(.)$ by using a general definition of $D(.)$ for two discrete distributions $H(.)$ and $F(.)$ with respective probability mass functions p_H and p_F satisfying the condition that the values at which p_H are positive include all the values at which p_F are positive.

13. COMPARISON OF DISCRETE DISTRIBUTIONS. To compare two discrete distributions we define first $d(u)$ and then $D(u)$ as follows:

$$d(u) = d(u; H, F) = p_F(H^{-1}(u))/p_H(H^{-1}(u)),$$
$$D(u) = \int_0^1 d(t)dt.$$

We apply this definition to the discrete sample distributions $F^{\wedge}$ and $F^{\sim}_k$ to obtain

$$d_k^{\sim}(u) = d(u; F^{\wedge}, F_k^{\sim})$$

and its integral $D_k^{\sim}(u)$.

We obtain the following definition of $d_k^{\sim}(u)$ for the c sample testing problem with all values distinct:

$$d_k^{\sim}(u) = N/n_k \text{ if } (R_k(j) - 1)/N < u < R_k(j)/N, j = 1, \ldots, n_k,$$
$$= 0, \quad \text{otherwise.}$$

A component, with score function $J(u)$, is a linear functional

$$T_k^{\sim}(J) = \int_0^1 J(u)d_k^{\sim}(u)du$$

It equals

$$(1/n_k)\sum_{j=1}^{n_k} N \int_{(R_k(j)-1)/N}^{R_k(j)/N} J(u)du$$

which can be approximated by $E^{\sim}[J(P^{\wedge}(Y_k))]$.

14. LINEAR RANK STATISTICS. The concept of a linear rank statistic to compare the equality of c samples does not have a universally accepted definition. One possible definition is

$$T_k^{\sim}(J) = (1/n_k)\sum_{j=1}^{n_k} J((R_k(j) - .5)/N)$$

However we choose the definition of a linear rank statistic as a linear functional of $d_k^{\sim}(u)$, which we call a component; it is approximately equal to the above formula.

We define

$$T_k^{\sim}(J) = ((N-1)\ \mathrm{VAR}[J(U)]p_{\cdot k}/(1-p_{\cdot k}))^{.5} \int_0^1 J(u)\{d_k^{\sim}(u) - 1\}du \qquad (!)$$

where U is Uniform$\{0,1\}$, $E[J(U)] = \int_0^1 J(u)du$,

$$\mathrm{VAR}[J(U)] = \int_0^1 \{J(u) - E[J(U)]\}^2 du.$$

Note that the integral in the definition of $T_k^{\sim}(J)$ equals

$$\int_0^1 J(u)d\{D_k^{\sim}(u) - u\}.$$

The components of the Kruskal-Wallis nonparametric test statistic TKW^2 for testing the equality of c means have score function $J(u) = u - .5$ satisfying

$$E[J(U)] = .5, \ \text{VAR}[J(U)] = 1/12.$$

The components of F test statistic T^2 have score function

$$J(u) = \{Q^{\wedge}(u) - Y^{-}\}/\sigma^{\wedge}$$

where $Q^{\wedge}(u)$ is sample quantile function of the pooled sample Y.

15. GENERAL DISTANCE MEASURES. General measures of the distance of $D^{-}(u)$ from u and of $d^{-}(u)$ from 1 are provided by the integrals from 0 to 1 of

$$\{d^{-}(u) - 1\}^2, \quad \{D^{-}(u) - u\}^2, \quad \{D^{-}(u) - u\}^2/u(1-u), \quad \{d^{\wedge}(u) - 1\}^2$$

where $d^{\wedge}(u)$ is a smooth version of $d^{-}(u)$. We will see that these measures can be decomposed into components which may provide more insight; recall basic components are linear functionals defined by (!)

$$T^{-}(J) = \int_0^1 J(u)d^{-}(u)du.$$

If $\phi_i(u)$, $i = 0, 1, 2, \ldots$, are complete orthonormal functions with $\phi_0 = 1$, then H_0 can be tested by diagnosing the rate of increase (as a function of $m = 1, 2, \ldots$) of

$$\int_0^1 \{d_m(u) - 1\}^2 du = \sum_{i=1}^{m} |T^{-}(\phi_i|^2$$

which measure the distance from 1 of the approximating smooth densities

$$d_{,n}(u) = \sum_{i=1}^{m} T^{-}(\phi_i)\phi_i(u).$$

16. ORTHOGONAL POLYNOMIAL COMPONENTS. Let $p_i(x)$ be Legendre polynomials on (-1,1):

$$p_1(x) = x$$
$$p_2(x) = (3x^2 - 1)/2,$$
$$p_3(x) = (5x^3 - 3x)/2,$$
$$p_4(x) = 35x^4 - 30x^2 + 3.$$

Define Legendre polynomial score functions

$$\phi L_i(u) = (2i + 1)^{.5} p_i(2u - 1).$$

One can show that an Anderson-Darling type statistic, denoted $AD(D^{-})$, can be represented

$$AD(D^{-}) = \int_0^1 \{\{D^{-}(u) - u\}^2/u(1 - u)\}du$$
$$= \sum_{i=1}^{\infty} |T^{-}(\phi L_i)|^2/(i(i+1))$$

Define cosine score functions by

$$\phi C_i(u) = 2^{.5}\cos(i\pi u).$$

One can show that a Cramer–von Mises type statistic, denoted $CM(D^{\sim})$, can be represented

$$CM(D^{\sim}) = \int_0^1 \{D^{\sim}(u) - u\}^2 du$$
$$= \sum_{i=1}^{\infty} |T^{\sim}(\phi C_i)|^2/(i\pi)^2$$

In addition to Legendre polynomial and cosine components we consider Hermite polynomial components corresponding to Hermite polynomial score functions

$$\phi H_i(u) = (i!)^{-.5} H_i(\Phi^{-1}(u))$$

where $H_i(x)$ are the Hermite polynomials:

$$H_1(x) = x,$$
$$H_2(x) = x^2 - 1,$$
$$H_3(x) = x^3 - 3x,$$
$$H_4(x) = x^4 - 6x^2 + 3.$$

17. QUARTILE COMPONENTS AND CHI-SQUARE. Quartile diagnostics of the null hypothesis H_0 are provided by components with quartile "square wave" score functions

$$\begin{aligned} SQ_1(u) &= -2^{.5}, \quad 0 < u < .25, \\ &= 0, \quad .25 < u < .75, \\ &= 2^{.5}, \quad .75 < u < 1; \\ SQ_2(u) &= 1, \quad 0 < u < .25, \\ &= -1, \quad .25 < u < .75, \\ &= 1, \quad .75 < u < 1; \\ SQ_3(u) &= 0 \quad \text{if } 0 < u < .25 \text{ or } .75 < u < 1, \\ &= -2^{.5}, \quad .25 < u < .5, \\ &= 2^{.5}, \quad .5 < u < .75. \end{aligned}$$

A chi-squared portmanteau statistic, which is chi-squared(3), is

$$CQ_k = (N-1)p_{.k}/(1-p_{.k}) \sum_{i=1}^{3} |T^{\sim}(SQ_i)|^2$$
$$= (N-1)p_{.k}/(1-p_{.k}) \int_0^1 \{dQ_k(u) - 1\}^2 du$$

defining the quartile density (for $i = 1, 2, 3, 4$)

$$dQ_k(u) = 4\{D_k^{\sim}(i(.25)) - D_k^{\sim}((i-1).25), (i-1).25 < u < i(.25)$$

A pooled portmanteau chi-squared statistic is

$$CQ = \sum_{k=1}^{c}(1 - p_{.k})CQ_k$$

18. DIVERSE STATISTICS AVAILABLE TO TEST EQUALITY OF c SAMPLES. The problem of statistical inferecence is not that we don't have answers to a given question; usually we have too many answers and we don't know which one to choose. A unified framework may help determine optimum choices. To compare c samples we can compute the following functions and statistics:

1) comparison densities: $d_k\tilde{}(u)$,
2) comparison distributions $D_k\tilde{}(u)$,
3) quartile comparison density $dQ_k(u)$, quartile density chi-square

$$CQ_k = (N-1)p_{.k}/(1-p_{.k})\int_0^1 \{dQ_k(u) - 1\}^2 du.$$

4) non-parametric regression smoothing of $d_k\tilde{}(u)$ using a boundary Epanechnikov kernel, denoted $d_k\hat{}(u)$,
5) Legendre components and chi-squares up to order 4 are defined using definition (!) of $T_k\tilde{}$:

$$TL_k(i) = T_k\tilde{}(\phi L_i)$$

$$CL_k(m) = \sum_{i=1}^{m}|TL_k(i)|^2$$

$$CL(m) = \sum_{k=1}^{c}(1 - p_{.k})CL_k(m)$$

$$AD_k = \sum_{i=1}^{\infty}|TL_k(i)|^2/i(i+1)$$

$$AD = \sum_{k=1}^{c}(1 - p_{.k})AD_k$$

6) Cosine components and chi-squares up to order 4 are defined:

$$TC_k(i) = T_k\tilde{}(\phi C_i)$$

$$CC_k(m) = \sum_{i=1}^{m}|TC_k(i)|^2$$

$$CC(m) = \sum_{k=1}^{c}(1 - p_{.k})CC_k(m)$$

$$CM_k = \sum_{i=1}^{\infty}|TC_k(i)|^2/(i\pi)^2$$

$$CM = \sum_{k=1}^{c}(1 - p_{.k})CM_k$$

7) Hermite components and chi-squares up to order 4 are defined:

$$TH_k(i) = T_k\tilde{}(\phi H_i)$$

$$CH_k(m) = \sum_{i=1}^{m} |TH_k(i)|^2$$

$$CH(m) = \sum_{k=1}^{c} (1 - p_{.k}) CH_k(m)$$

8) density estimators $d_k\hat{}(u)$ computed from components up to order 4,
9) entropy measures with penalty terms which can be used to determine how many components to use in the above test statistics

19. EXAMPLES OF DATA ANALYSIS. The interpretation of the diversity of statistics available is best illustrated by examples.

In order to compare our methods with others available we consider data analysed by Boos (1986) on ratio of assessed value to sale price of residential property in Fitchburg, Mass., 1979. The samples (denoted I, II, III, IV) represent dwellings in the categories single-family, two-family, three-family, four or more families. The sample sizes (54, 43, 31, 28) are proportions .346, .276, .199, .179 of the size 156 of the pooled sample. We compute Legendre, cosine, Hermite components up to order 4 of the 4 samples; they are asymptotically standard normal. We consider components greater than 2 (3) in absolute value to be significant (very significant).

Legendre, cosine, and Hermite components are very significant only for sample I, order 1 (-4.06, -4.22, -3.56 respectively). Legendre components are significant for sample IV, orders 1 and 2 (2.19, 2.31). Cosine components are significant for sample IV, orders I and II (2.36, 2.23) and sample III, order 1 (2.05). Hermite components are significant for sample IV, orders 2 and 3 (2.7 and -2.07).

Conclusions are that the four samples are not homogeneous (have the same distributions). Samples I and IV are significantly different from the pooled sample. Estimators of the comparison density show that sample I is more likely to have lower values than the pooled sample, and sample IV is more likely to have higher values. While all the statistical measures described above have been computed, the insights are provided by the linear rank statistics of orthogonal polynomials rather than by portmanteau statistics of Cramer-von Mises or Anderson-Darling type.

20. CONCLUSIONS. The goal of our recent research (see Parzen (1979), (1983)) on unifying statistical methods (especially using quantile function concepts) has been to help the development of both the theory and practice of statistical data analysis. Our ultimate aim is to make it easier to apply statistical methods by unifying them in ways that increase understanding, and thus enable researchers to more easily choose methods that provide greatest insight for their problem. We believe that if one can think of several ways of looking at a data analysis one should do so. However to relate and compare the answers, and thus arrive at a confident conclusion, a general framework seems to us to be required.

One of the motivations for this paper was to understand two-sample tests of the Anderson-Darling type; they are discussed by Pettitt (1976) and Scholz and Stephens (1987). This paper provides new formulas for these test statistics based on our new definition of sample comparison density functions. Asymptotic distribution theory for rank processes defined by Parzen (1983) is given by Aly, Csorgo, and Horvath (1987); an excellent review of theory for rank processes is given by Shorack and Wellner (1986).

However one can look at k sample Anderson-Darling statistics as a single number formed from combining many test statistics called components. The importance of components is also advocated by Boos (1986), Eubank, La Riccia, and Rosenstein (1987) and Alexander (1989). Insight is greatly increased if instead of basing one's conclusions on the values of single test statistics, one looks at the components and also at graphs of the densities of which the components are linear functionals corresponding to various score functions. The question of which score functions to use can be answered by considering the tail behavior of the distributions that seem to fit the data.

REFERENCES

Alexander, William (1989) "Boundary kernel estimation of the two-sample comparison density function" Texas A&M Department of Statistics Ph. D. thesis.

Aly, E.A.A., M. Csorgo, and L. Horvath (1987) "P-P plots, rank processes, and Chernoff-Savage theorems" in *New Perspectives in Theoretical and Applied Statistics* (ed. M.L. Puri, J.P. Vilaplann, W. Wertz) New York: Wiley 135-156.

Boos, Dennis D. (1986) "Comparing k populations with linear rank statistics" *Journal of the American Statistical Association*, 81, 1018-1025.

Eubank, R.L., V.N. La Riccia, R.B. Rosenstein (1987) "Test statistics derived as components of Pearson's Phi-squared distance measure" *Journal of the American Statistical Association*, 82, 816-825.

Parzen, E. (1979) "Nonparametric statistical data modeling" *Journal of the American Statistical Association*, 74, 105-131.

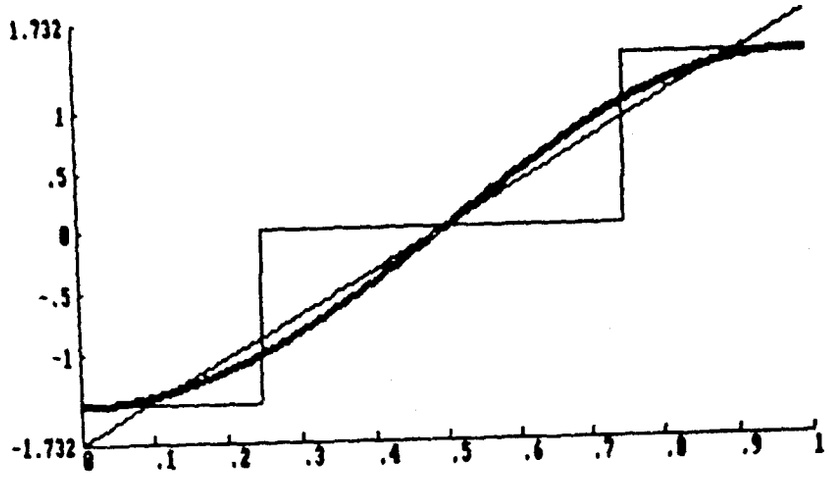
Parzen, E. (1983) "FunStat quantile approach to two-sample statistical data analysis" Texas A&M Institute of Statistics Technical Report A-21 April 1983.

Pettitt, A.N. (1976) "A two-sample Anderson Darling statistic" *Biometrika*, 63, 161-168.

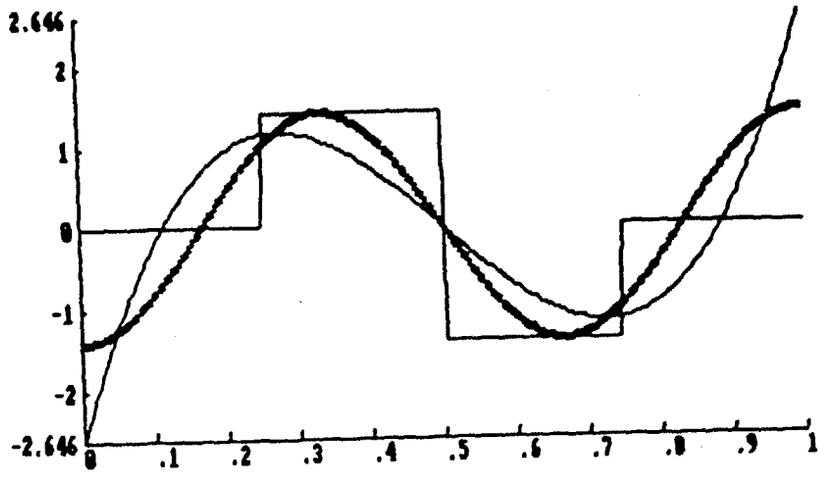
Scholz, F.W, and M.A.Stephens (1987) "k-sample Anderson- Darling tests" *Journal of the American Statistical Association*, 82, 918-924.

Shorack, Galen and Jon Wellner (1986) *Empirical Processes With Applications to Statistics* New York: Wiley.

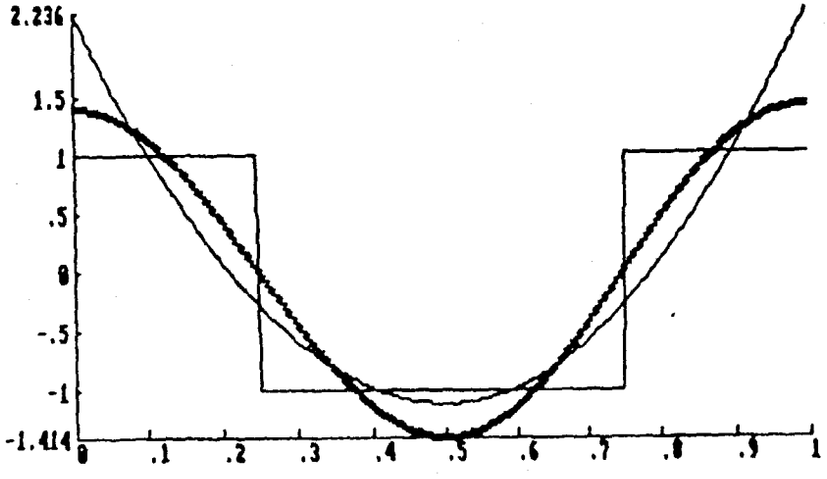
Score functions for location: SQ1 (square wave), phiL1 (light curve), phiC1 (dark curve).



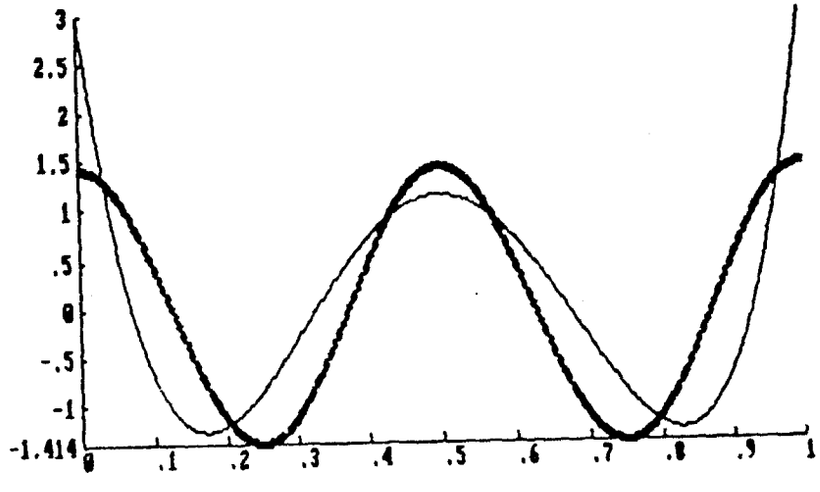
Score functions for skewness: SQ3 (square wave), phiL3 (light curve), phiC3 (dark curve).



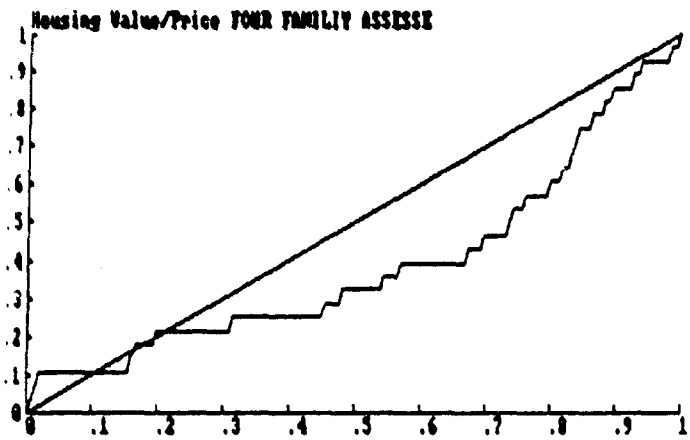
Score functions for scale: SQ2 (square wave), phiL2 (light curve), phiC2 (dark curve).

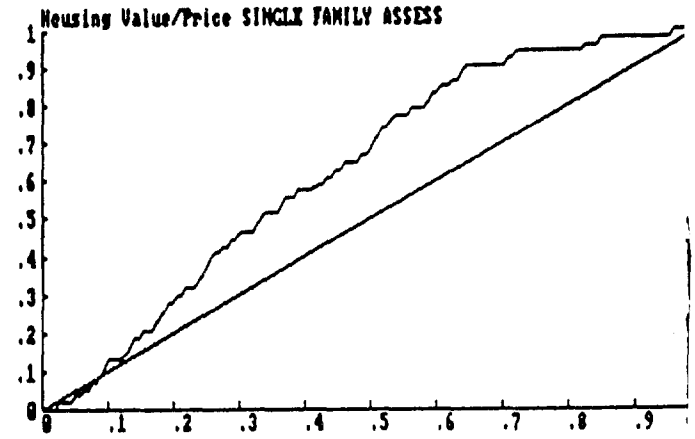


Score functions for kurtosis: phiL1 (light curve), phiC1 (dark curve).

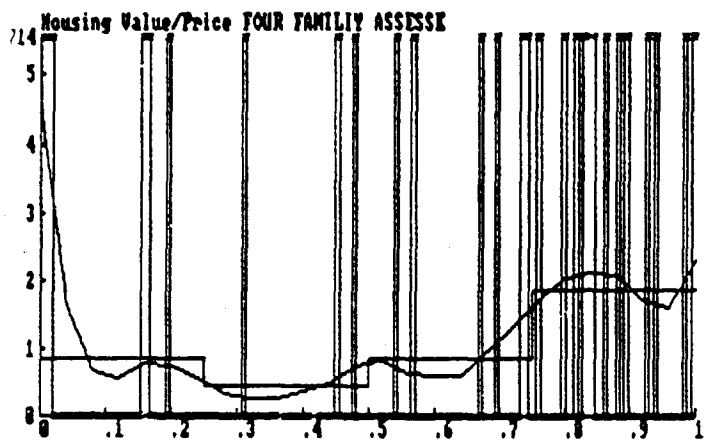


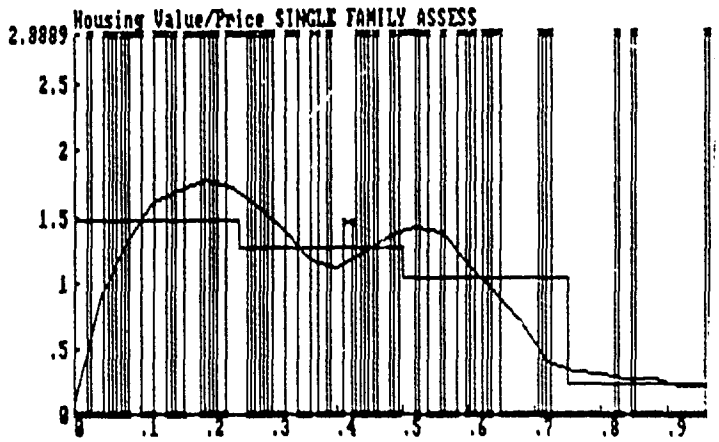
For samples I and IV, sample comparison distribution function $D^{\sim}(u)$



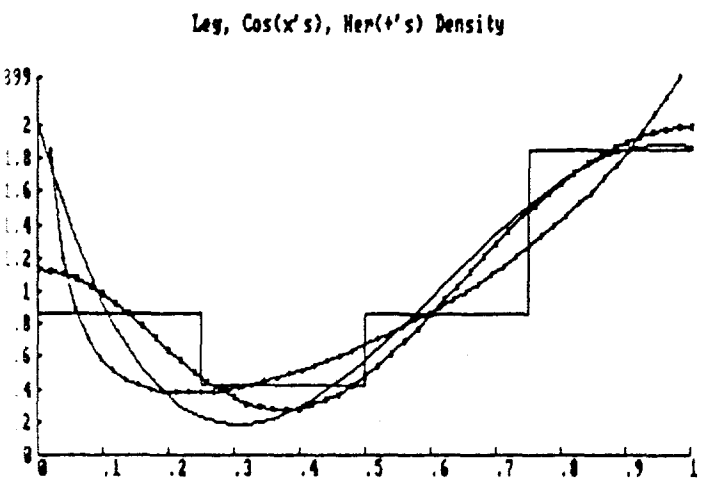


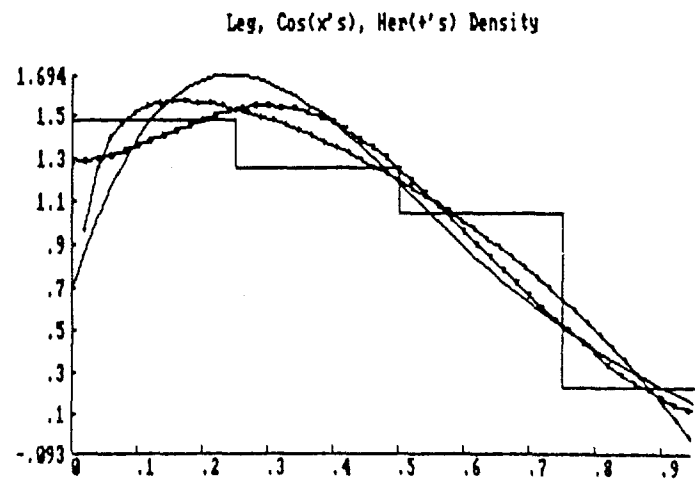
For samples I and IV, sample comparison density $d^{\sim}(u)$, sample quartile density $dQ^{\sim}(u)$ (square wave), nonparametric density estimator $d^{\wedge}(u)$





For samples I and IV, Legendre, cosine, and Hermite orthogonal polynomial estimator of order 4 of the comparison density, denoted $d_4(u)$, compared to sample quartile density $dQ^{\sim}(u)$.





Reliability of the M256 Chemical Detection Kit

David W. Webb & Linda L. C. Moss

U.S. Army Ballistic Research Laboratory

Abstract

The U.S. Army uses the M256 Chemical Detection Kit (CDK) to indicate the presence or absence of certain agents in the battlefield, which is indicated by a color change on the kit. Strength of response is also influenced by the quantity of agent. Lots must meet reliability specifications to be considered "battle-ready". How do we go about collecting and analyzing our data so as to evaluate its reliability? Other problems of interest include quantifying how the agent quantity affects the response and if there are differences between the two manufacturers of the M256 CDK. Consultants at the Ballistic Research Laboratory have employed a dose-response framework to study the reliability problem. We use a binary response (present/not present) and assume a lognormal distribution in arriving at a response curve for each lot. Assessments of our approach and suggestions for alternative approaches are asked of the panel.

Description of Kit

The M256 Chemical Detection Kit (CDK) is used to detect the presence/absence of dangerous concentrations of toxic agents by color-changing chemical reactions. Each CDK contains twelve samplers, which are the actual testing devices. Four types of agents can be detected with the CDK. The tests indicate

a) if it is permissible to remove the protective mask following an agent attack,

b) if agent is present in the air or on surfaces suspected of contamination,

c) if any agent is present after decontamination operations.

The U.S. Army requires that the samplers exhibit at least a 92.5% reliability (with 90% confidence) in responding to agent concentrations at the specification levels. However, the kit should not be so sensitive that soldiers wear their mask at safe levels of concentration, thereby interrupting other battlefield duties.

On the back of each sampler are complete instructions for testing and colored examples of safe and danger responses. After performing the test, a paper test spot is checked for any change of color. The color change will not usually be an exact match with the colors shown on the back of the sampler. This is because the response depends upon the agent quantity. To make matters more complex, when the agent is present the observed response may be nonuniform with a few shades of the danger response showing.

Test Conditions & Restrictions

The lots of kits differ in manufacturer (A or B), age (1-8 years), and storage site (8 sites in the United States and Europe). Not all combinations of these three factors are represented in the design matrix; in fact, the design matrix is very sparse. For example, there was only one lot that was eight years old.

Most lots contain ten or more kits (therefore, 120 or more individual samplers). Some lots contained as many as 1000 kits, while others had as few as one kit.

We are restricted to the number of samplers that may be tested at any time since the test chamber is large enough to hold only six samplers. Another restriction lies in the fact that testing laboratories are only available for the length of time designated in the work contract. This usually is no more than two months.

The test equipment that controls the concentration of agent in the test chamber is very accurate and precise, but it is slow. It may take about an hour to change to a higher concentration. When going from a high to a low concentration, the waiting period may be several hours since the high concentration tends to leave a residual amount of agent in the test chamber.

Our Approach

We have decided to evaluate each agent and the chosen lots separately. From each manufacturer, we have selected one lot from the available age groups. Also, we have tried to choose lots of similar age from the manufacturers so that they can be paired and we can look for general trends. In all, we have chosen fifteen lots ranging in age from 1 to 8 years. Although the sites are in varying climatic areas, most of the warehouses are humidity and temperature controlled; therefore the locations are treated as homogenous. Differences existing between manufacturers are not considered in our initial design, but will be addressed later.

We have taken the route of estimating the reliability of each lot at the specification level of each agent. We have also chosen a dose-response type experiment, where our dose is the agent concentration and the response is safe/danger. For the purpose of determining response, U.S. Army manuals specify a set of nine color chips that progressively range from the "safe" color to the "danger" color. The manual also states a cutoff color for the Bernoulli response. (In most cases, color chips 1-3 correspond to a safe response, while chips 4-9 are considered danger responses.)

We have made the assumption that the response curves follow that of the lognormal cumulative distribution function with unknown mean and standard deviation. The lognormal was selected based on historical precedent, although we note that the log-logistic would have also been a reasonable choice.

To choose the concentration levels at which to run the tests, we have considered several candidate sequential designs. In light of some of our restrictions, however, none of these would be very practical (e.g., Robbins-Monro would have required too much laboratory time).

Instead, we have chosen a two-stage "semi"-fixed design. In the first stage, 11 samplers are tested at seven different levels; one concentration level set at an estimated mean, three concentrations above this estimated mean, and three concentration levels below the estimated mean, each being a multiple of the standard deviation away from the mean. Mean and standard deviation estimates are based on the results of a pretest (which for the purpose of brevity is deleted from this presentation). The multiple of the standard deviation is chosen so that the specification level will be covered by the seven test concentrations.

Stage I

Concentration	Number of Samplers
$\hat{\mu}_1 - 3k\hat{\partial}_1$	1
$\hat{\mu}_1 - 2k\hat{\partial}_1$	1
$\hat{\mu}_1 - k\hat{\partial}_1$	2
$\hat{\mu}_1$	3
$\hat{\mu}_1 + k\hat{\partial}_1$	2
$\hat{\mu}_1 + 2k\hat{\partial}_1$	1
$\hat{\mu}_1 + 3k\hat{\partial}_1$	1
	11

Note: k is chosen so that the seven test concentrations cover the specification level. $\hat{\mu}_1$ and $\hat{\partial}_1$ come from the pretest.

At the conclusion of Stage I, the data are analyzed using the DiDonato-Jarnagin maximum likelihood estimation algorithm to produce new estimates of the parameters, $\hat{\mu}_2$ and $\hat{\partial}_2$. In Stage II, nine more units are tested at five concentration levels; one level set at the new estimated mean, and at two levels above and below this, each now being a multiple of the new standard deviation from the mean.

Stage II

Concentration	Number of Samplers
$\hat{\mu}_2 - 2\hat{\partial}_2$	1
$\hat{\mu}_2 - \hat{\partial}_2$	2
$\hat{\mu}_2$	3
$\hat{\mu}_2 + \hat{\partial}_2$	2
$\hat{\mu}_2 + 2\hat{\partial}_2$	1
	9

At the conclusion of Stage II, the parameter estimates for the lot are re-evaluated using all 20 data points, giving us a final $\hat{\mu}$ and $\hat{\partial}$. With these final estimates, the .925 quantile is estimated by $\hat{\mu} + z_{(.925)}\ \hat{\partial}$.

By taking the variance of the above equation, we get an estimate of the variance of the .925 quantile,

$$\mathrm{Var}(\hat{\mu}) + (z_{(.925)})^2\ \mathrm{Var}(\hat{\partial}) + 2\, z_{(.925)}\, \mathrm{Cov}(\hat{\mu}, \hat{\partial})$$

(The DiDonato-Jarnagin algorithm gives the values of the variances and covariance term.) If the one-sided 90% upper confidence limit of the .925 quantile is less than the specification

concentration, then we can conclude that the lot meets the requirement for that particular agent.

We do not have a statistical technique per se for detecting significant differences between manufacturers or sites. Our "approach" would be to simply look for any obvious trends or differences. To study the age issue, a separate accelerated life test will be conducted at a later date.

Questions

1. Is our approach appropriate for determining an extreme quantile?

2. Can one estimate a quantile when considering more than two possible responses (e.g., the nine color chips)?

3. How might we statistically compare the reliability of the manufacturers (or sites)?

Concluding Remarks

Following our presentation, we heard comments and suggestions from the clinical session panelists and audience. Two major concerns were expressed by several persons. First was uneasinessd towards our assumption of a log-normal distribution. Some respondents felt this to be a potentially dangerous assumption, especially since we are estimating the tail of our distribution. Secondly, some persons questioned our method of estimating the mean of the distribution, and then extrapolating to the .925 quantile. These two problems could lead to some very erroneous conclusions.

In general, the comments we heard confirmed our beliefs that this is a very difficult problem to analyze, in light of the small sample sizes and other laboratory constraints to which the test is subjected. Although no definitive alternative approaches arose from our discussions, some possible attacks that were suggested to us included --

1. Sampling more towards the tails of the distribution.

2. Isotonic regression.

3. Testing at the specification level and employing a general linear model approach with the color chip number corresponding to the color change as the response and age, manufacturer, and storage site as variables.

We would like to thank the panelists and audience for their many suggestions and remarks.

COMPARISON OF RELIABILITY CONFIDENCE INTERVALS

Paul H. Thrasher
Engineering and Analysis Branch
Reliability, Availability, and Maintainability Division
Army Materiel Test and Evaluation Directorate
White Sands Missile Range, New Mexico 88002-5175

ABSTRACT

Some confidence intervals on reliabilities are investigated. Only binomial events are considered. Only the narrowest two-sided and the upper one-sided confidence intervals are calculated. Three methods of estimating the distribution of reliabilities are reviewed and compared. These are the Fisherian approach, the Bayesian approach with the ignorant prior, and the Bayesian approach with the noninformed or noninformative binomial prior. Both the width and location of the confidence intervals differ for these three methods.

INTRODUCTION

Reliability estimates are not as straightforward as might be expected. Measurement of a number of successes x in a sample size n quickly leads to a point estimate of the reliability R equal to x/n. Estimates of confidence intervals are more difficult to obtain however. Two things in addition to the data are needed for confidence interval estimation. First, some function must be used to describe the reliabilities. Second, a method must be selected to locate the confidence interval within the function.

The purpose of this paper is to compare various functions describing reliabilities. For simplicity, all tested items will be assumed dichotomous and independent. That is, the binomial $b(x;n,R)$ is assumed to describe the random variable x if n and R are known. The problem is to select a function for R when x and n are known. The three functions considered here are based on (1) the Fisherian approach and (2) the Bayesian technique using prior distributions of R when (A) R is equally likely to be any value between zero and one and (B) R is unknown numerically but it is known to be a binomial parameter.

To focus attention on the comparison of the confidence intervals from these three functions, the methods used to locate the confidence intervals are restricted in this paper. Only two methods are used in calculations; one is one-sided and the other is two-sided.

Preceding Page Blank

The one-sided confidence interval considered is the upper confidence interval. This is based on the premise that having a reliability too low is much more serious than the reliability being toc high.

The two-sided confidence interval considered is the narrowest possible (Rankin). This is illustrated in Figure 1. It is located by adjusting the confidence limits until (1) the sum of the areas under the tails is α and (2) the functions of these two limits are equal. This correspondence of narrowest interval with equal heights is a geometrical property. It is not based on the choice of the function describing R. It may be demonstrated by (1) starting with the confidence limits of points equal heights, (2) moving the left confidence limit to the right, and (3) noting that the right limit has to be moved further to the right in order to keep the sum of the areas under the tails constant. This is shown in Figure 1 by the dashed lines. A similar argument starts by moving the narrowest confidence limit to the left.

Other possible two-sided confidence limits, not calculated in this paper, are illustrated in Figures 2 and 3. These are the traditional equal-division-of-area-under-the-tails interval and the maximum-likelihood-estimator-in-the-center interval. The first is the easiest to calculate. The second has a symmetric appeal but it is non-existant when the peak of the curve is not at $R=0.5$ and $1-\alpha$ is sufficiently large.

FISHERIAN APPROACH

The traditional Fisherian approach (Mann, Shafer, and Singpurwalla) considers sums of binomial probabilities. This approach yields two Beta functions. The lower confidence limit is obtained from one Beta function; a second function is needed for the upper limit.

Lower Confidence Limit:

The lower $(1-\alpha)100\%$ confidence limit $\underline{R}$ is defined by $P[R>\underline{R}]=1-\alpha$. An alternate expression is $P[R<\underline{R}]=\alpha$. The limit $\underline{R}$ is the largest value of R that makes the data x and n plausible. Plausibility is defined as satisfaction of the degree of confidence $1-\alpha$ of correctly selecting the right $\underline{R}$. The lower 100% confidence limit of R is $\underline{R}=0$ because all values of R satisfy $R\geq 0$. Increasing $\underline{R}$ requires a decrease in $1-\alpha$ or an increase in α. This increase in $\underline{R}$ shifts the binomial distribution of the possible measurements i which resulted in the single measurement x. For the limiting case of $\underline{R}=0$, the

binomial $b(i;n,\underline{R})$ consists of a single spike of unit height at $i=0$. As $\underline{R}$ and α increase, $b(i;n,\underline{R})$ takes a shape illustrated in Figure 4 and described by

$$b(i;n,\underline{R}) = \binom{n}{i} \underline{R}^i (1-\underline{R})^{n-i}$$

where the number of ways of obtaining i successes in n trials if found from by

$$\binom{n}{i} = \frac{n!}{i!\,(n-i)!} = \frac{n(n-1)\ldots(n-i+1)}{i(i-1)\ldots(1)}.$$

The extent of the shifting from the single spike is determined by the data x and n. The value of $\underline{R}$ is determined in two steps. First, $\underline{R}$ is increased until the sum $b(x;n,\underline{R})+b(x+1;n,\underline{R})+\ldots+b(n;n,\underline{R})$ equals the probability α making the confidence relation $P[R<\underline{R}]=\alpha$ or $P[R>\underline{R}]=1-\alpha$ untrue. Second, the continuous variable $\underline{R}$ is decreased infinitesimally making the confidence relation $P[R>\underline{R}]=1-\alpha$ just barely valid. Thus $\underline{R}$ is neither too large or too small to be a $(1-\alpha)100\%$ lower confidence limit on R when

$$\alpha = \sum_{i=x}^{n} \binom{n}{i} \underline{R}^i (1-\underline{R})^{n-i}.$$

The extraction of $\underline{R}$ from this equation can be facilitated by using a Beta function as described in the following paragraph. Before doing that however, it is expedient to note that a measurement of $x=0$ implies that $\underline{R}=0$ for all values of α. This special case isn't algebraically included in the following Beta function. It is adroitly described by an argument based on Figure 4: when $x=0$, $\underline{R}$ has to be 0 to make $b(0;n,\underline{R})=1$ and $b(i;n,\underline{R})=0$ for all $i\neq 0$.

The Beta function of R is

$$f(R) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} R^{a-1} (1-R)^{b-1}$$

where a and b are parameters. Using the equality of the gamma function $\Gamma(j)$ and the factorial $(j-1)!$ when j is an integer yields

$$f(R) = \frac{(a+b-1)!}{(a-1)!(b-1)!} R^{a-1} (1-R)^{b-1}.$$

Postulating that the reliability is described by f(R) and setting the area to the left of $\underline{R}$ at α yields

$$\alpha = \int_0^{\underline{R}} f(R)\, dR.$$

Repeated integrations by part yield

$$\alpha = \sum_{i=a}^{a+b-1} \binom{a+b-1}{i} \underline{R}^i\, (1-\underline{R})^{a+b-1-i}.$$

Comparison of this summation and the summation for α in the previous paragraph yields a=x and a+b-1=n. Thus the parameters in the Beta function for the lower limit $\underline{R}$ are a=x and b=n+1-x.

Upper Confidence Limit:

The upper confidence limit R defined by $P[R<\overline{R}]=1-\alpha$ is obtained from another Beta function. Arguments similar to those in the proceeding section yield the upper Beta function in four steps:

(1) $\overline{R}$ is in the binomial sum

$$\alpha = \sum_{i=0}^{x} \binom{n}{i} \overline{R}^i\, (1-\overline{R})^{n-i},$$

(2) $\overline{R}$ is the lower limit of integration over the second Beta function

$$\alpha = \int_{\overline{R}}^{1} f(R)\, dR,$$

(3) repeated integrations by parts transform this integral to the summation

$$\alpha = \sum_{i=0}^{a'-1} \binom{a'+b'-1}{i} \overline{R}^i\, (1-\overline{R})^{a'+b'-1-i}, \text{ and}$$

(4) the second Beta function parameters are identified by x=a'-1 and n=a'+b'-1 to be a'=x+1 and b'=n-x.

This Beta function does not describe $\overline{R}$ when x=n because $\Gamma(b')=\Gamma(0)=(0-1)!$ is meaningless. For this special case, $\overline{R}=1$ for all α. This may be seen from a binomial distribution symmetric to Figure 4. Using an $\overline{R}$ near 1 and an α containing binomial terms from j=0 to j=x=n, it is easily seen that α is 1 even when $\overline{R}$ is 1. Since $\overline{R}$ continuous, $\overline{R}=1$ for any value of $1-\alpha$.

BAYESIAN APPROACH

The Bayesian approach (Martz and Waller) uses the data x and n to update a prior distribution g(R) describing R to a posterior distribution g(R|x) describing R after x is given. The algebraic relation between these two is based on the equality of the joint density h(x,R) to both the product g(R|x)f(x) and the product f(x|R)g(R). Thus the posterior is found from

$$g(R|x) = f(x|R)\ g(R)\ /\ f(x).$$

This expression is simplified by noting that (1) the conditional density of x given R is

$$f(x|R) = b(x;n,R) = \binom{n}{x} R^x\ (1-R)^{n-x}$$

and (2) the marginal density f(x) from the integral of h(x,R)=f(x|R)g(R) is

$$f(x) = \int_0^1 f(x|R)\ g(R)\ dR = \int_0^1 \binom{n}{x} R^x\ (1-R)^{n-x}\ g(R)\ dR.$$

Thus the general posterior is

$$g(R|x) = \frac{R^x\ (1-R)^{n-x}\ g(R)}{\int_0^1 R^x\ (1-R)^{n-x}\ g(R)\ dR}.$$

Ignorant Prior:

One prior that can be used is the uniform distribution $g(R)=1$ for $0 \leq R \leq 1$ and $g(R)=0$ elsewhere. This is sometimes called the ignorant prior because all values of R between 0 and 1 are equally likely. That is, there is no evidence to favor the selection of any value of R over any other R between 0 and 1.

Use of this prior in the general posterior yields

$$g(R|x) = \frac{R^x (1-R)^{n-x}}{\int_0^1 R^x (1-R)^{n-x} \, dR}.$$

Integration by parts evaluates the denominator. The posterior is thus

$$g(R|x) = \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)} R^{(x+1)-1} (1-R)^{(n-x+1)-1}.$$

This is a Beta function with parameters $a=x+1$ and $b=n-x+1$.

Noninformed Prior:

A second prior that can be used recognizes that the reliability is a binomial parameter but has no information about its value. This is sometimes called the noninformed or noninformative binomial prior.

Every noninformed prior is based on a transformation making the probability density insensitive to the data. For the binomial parameter R in $b(x;n,R)$, it has been empirically found (Box and Tiao) that plots of $K(x,n)b(x;n,\phi)$ versus ϕ yield very similar curves for fixed n and different x's when (1) $K(x,n)$ is determined by numerical integration to make the area under $K(x,n)b(x;n,\phi)$ equal to one and (2) ϕ is given by

$$\phi = \text{Arcsin}(R^{1/2}).$$

Figures 5 and 6 show that for $0<x<n$ these similar curves become nearly equally spaced along the ϕ axis as n is increased. The noninformed argument assumes that all n+1 curves are essentialy equal and equally spaced for all n. This

makes being noninformed about x equivalent to being ignorant about ϕ. The prior assumption that (1) x is unknown but (2) the situation is described by the one of these curves thus leads to a prior distribution of ϕ that is uniform between ϕ=0° and ϕ=90°. The corresponding prior of R may be found from the transformation of variable technique (Freund and Walpole) by applying

$$g(R) = h(\phi) \left| \frac{d\phi}{dR} \right|.$$

Using $h(\phi)$=1 and $\sin(\phi)$=$R^{1/2}$ in this equation yields $g(R)=1/\{2[R(1-R)]^{1/2}\}$.

Use of this binomial noninformed prior in the general posterior yields

$$g(R|x) = \frac{R^{x-1/2} \ (1-R)^{n-x-1/2}}{\int_0^1 R^{(x+1/2)-1} \ (1-R)^{(n-x+1/2)-1} \ dR}.$$

The denominator is recognized as an integral over a Beta function. It is evaluated to be $\Gamma(x+1/2)\Gamma(n-x+1/2)/\Gamma(n+1)$. The posterior is thus found to be a Beta function with a=x+1/2 and b=n-x+1/2.

COMPARISON OF CONFIDENCE INTERVALS

The three methods reviewed in the previous sections have been applied to confidence intervals on reliability. Both two-sided and one-sided intervals have been investigated.

Narrowest Two-Sided Intervals:

Figures 7 through 15 show distributions and narrowest two-sided 80% confidence intervals. Figure 7 illustrates the symmetry about x=n/2. Thus graphs for x<n/2 are not needed to investigate trends. Figure 8 is one example of the destruction of symmetry by making x>n/2. Figure 9 shows that when x=n the symmetry is so completely destroyed that the narrowest two-sided intervals are actually upper one-sided intervals. Figures 10, 11, and 12 and Figures 13, 14, and 15 show the effect of increasing n: for fixed x, the confidence intervals all become narrower but the relationship of the Fisherian, ignorant, and noninformed intervals retains an order.

The effect of changing x for fixed n is seen to be a change in the order of the Fisherian, ignorant, and noninformed intervals. The Fisherian interval seems to be the widest. For x near n/2, the ignorant interval seems to be narrower than the noninformed interval. For x near n however, the noninformed interval seems to be the narrowest of the three.

Upper One-Sided Intervals:

Figures 16 through 31 show distributions and upper one-sided 90% confidence intervals on reliability. The lower confidence limit appears lower for the Fisherian analysis than for the Bayesian analyses. The Bayesian ignorant and noninformed priors seem to lead to two sets of results. The lower confidence limit appears lower for the noninformed when x is near n/2 but higher for the noninformed when x is near n.

The symmetry of the Beta functions makes the lower confidence limits for x near 0 such that the Fisherian is lowest, the noninformed Bayesian is next lowest, and the ignorant Bayesian is the highest of the three. This is shown in Figures 25 though 28. These figures and Figures 29 through 31 also show that large n leads to fairly close agreement between the three methods.

CONCLUSION

The three methods are all on sound theoretical ground but give different results. No single method provides most logical confidence intervals. The choice between methods has to be based on goals and philosophy. Since the Fisherian method leads to the widest confidence intervals, it is the most conservative approach. Since proponents of the Bayesian method prefer priors which contain more information than the ignorant or noninformed prior, the Bayesian method (without a prior based on previous tests/calculations) does not meet all the goals of analysts with a Bayesian philosophy. Thus the Fisherian method seems to be a good, conservative method for the initial analysis. This initial analysis can provide a prior for a future Bayesian analysis of addition data from a future test.

REFERENCES

Rankin, Donald W., "Estimating Reliability from Small Samples," Proceedings of the Twenty-Second Conference on the Design of Experiments in Army Research, Development, and Testing, Army Research Office

Mann, Nancy R., Schafer, Ray E., and Singpurwalla, Nozer D., Methods for Statistical Analysis of Reliability and Life Data, John Wiley & Sons, New York, 1974

Martz, Harry F., and Waller, Ray A., Bayesian Reliability Analysis, John Wiley & Sons, New York, 1982

Box, George E.P., and Tiao, George C., Bayesian Interference in Statistical Analysis, Addison-Wesley, Reading, Massachusetts, 1973

Freund, John E., and Walpole, Ronald E., Mathematical Statistics, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1980

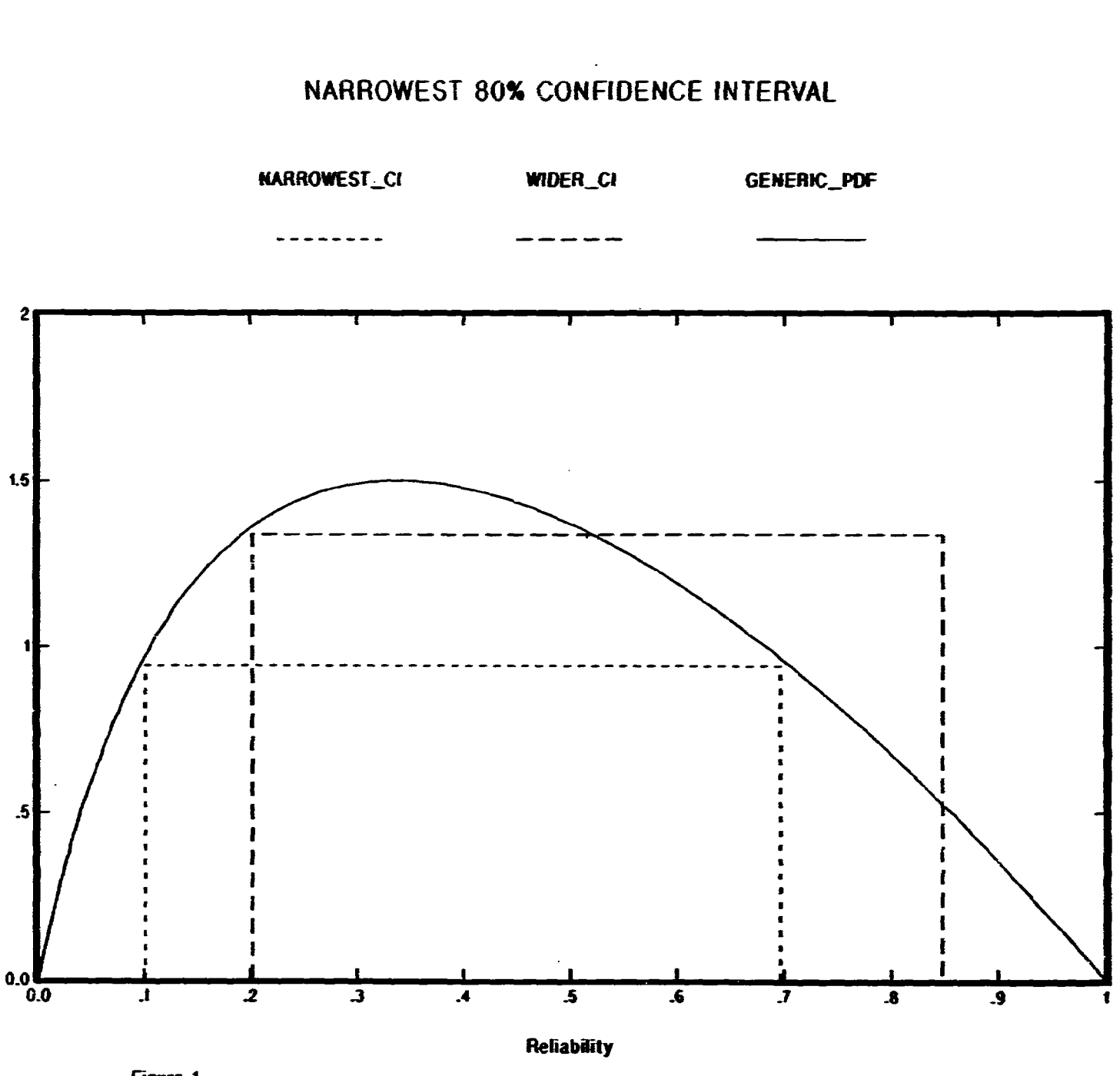

NARROWEST 80% CONFIDENCE INTERVAL
NARROWEST_CI
WIDER_CI
GENERIC_PDF
2
1.5
1
.5
0.0
0.0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
Reliability

Figure 1

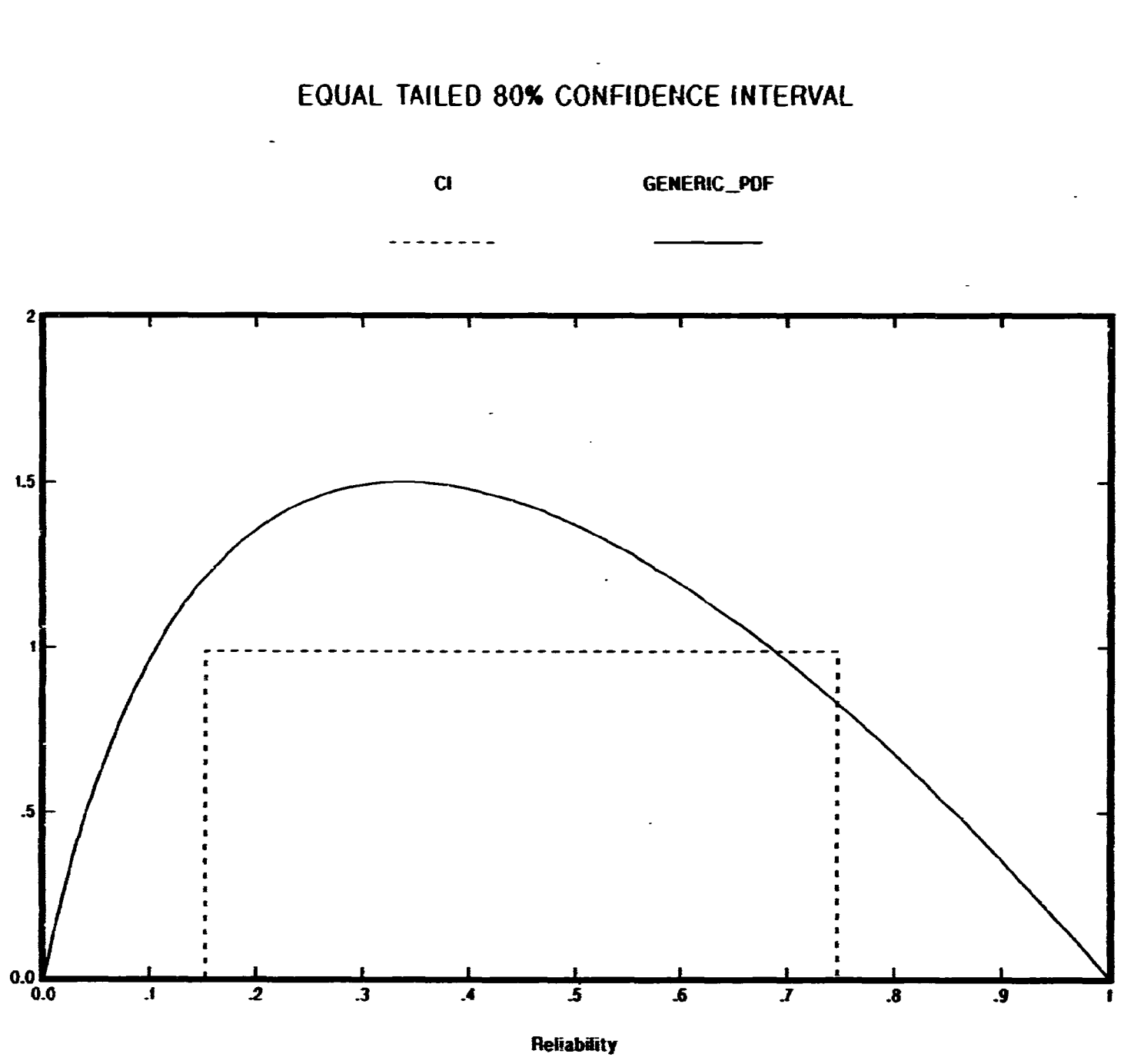

EQUAL TAILED 80% CONFIDENCE INTERVAL
CI
GENERIC_PDF
2
1.5
1
.5
0.0
0.0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
Reliability

Figure 2

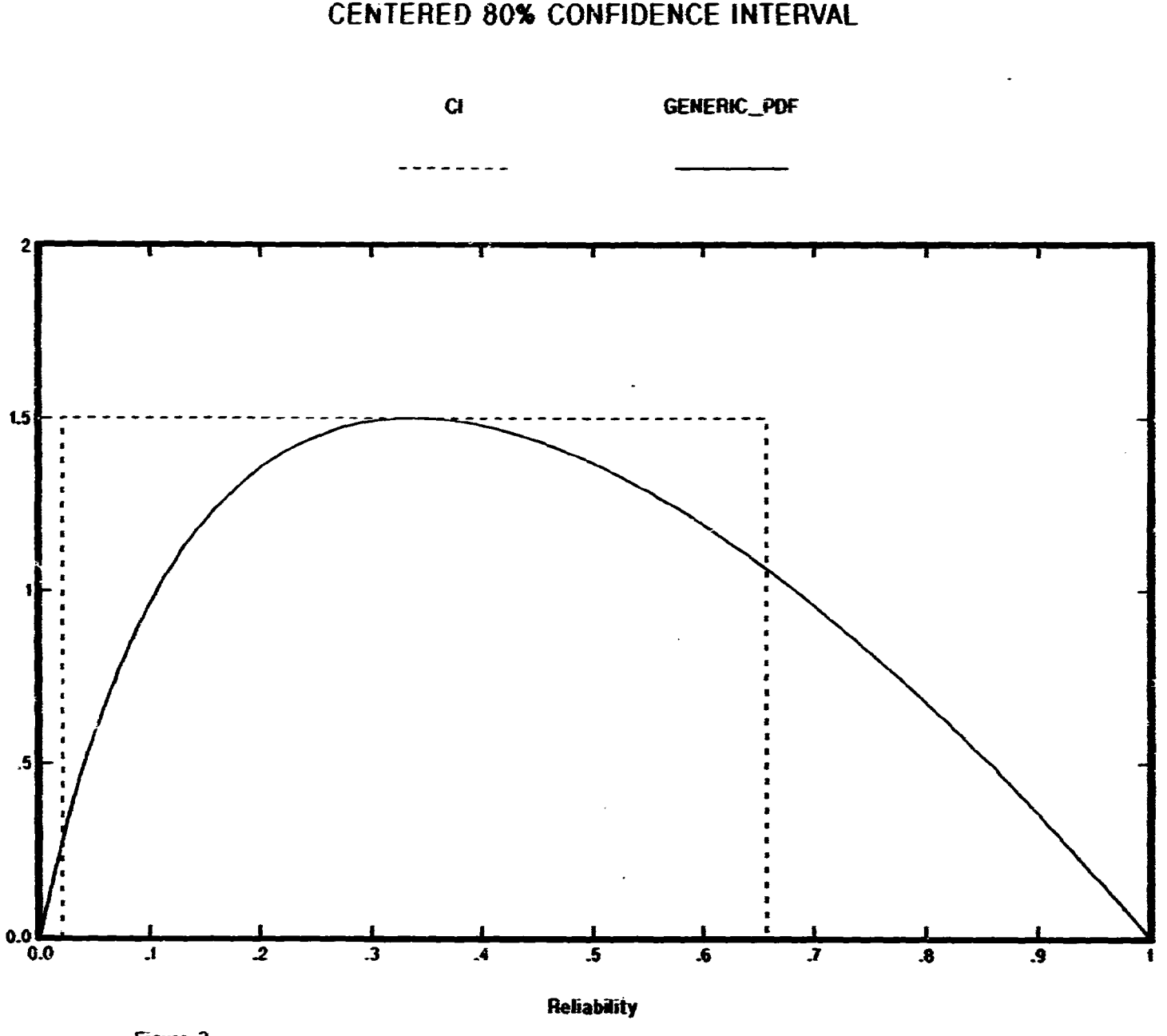

CENTERED 80% CONFIDENCE INTERVAL
CI
GENERIC_PDF
2
1.5
1
.5
0.0
0.0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
Reliability

Figure 3

BINOMIAL pdf FOR THE

RELIABILITY LOWER CONFIDENCE LIMIT

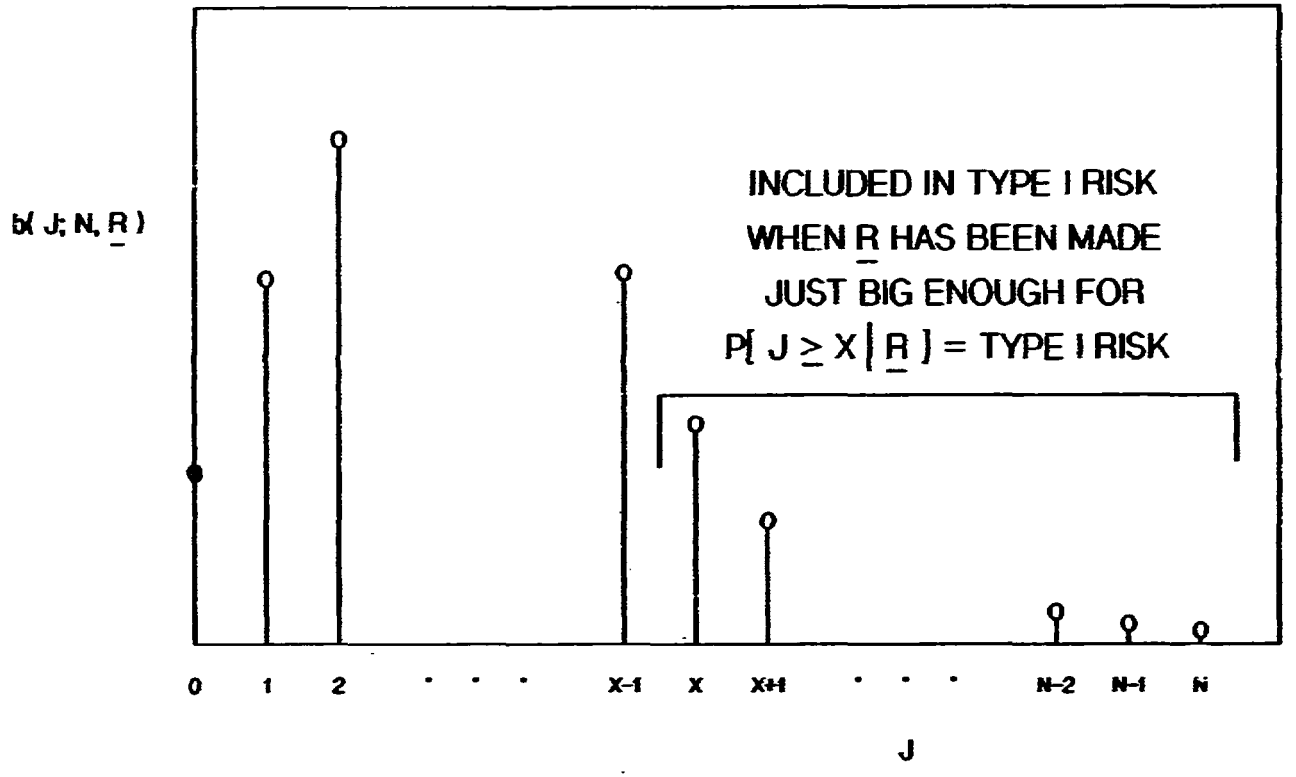


Figure 4

NORMALIZED BINOMIAL LIKELIHOODS

OF RELIABILITY AND TRANSFORMED RELIABILIY FOR N = 4

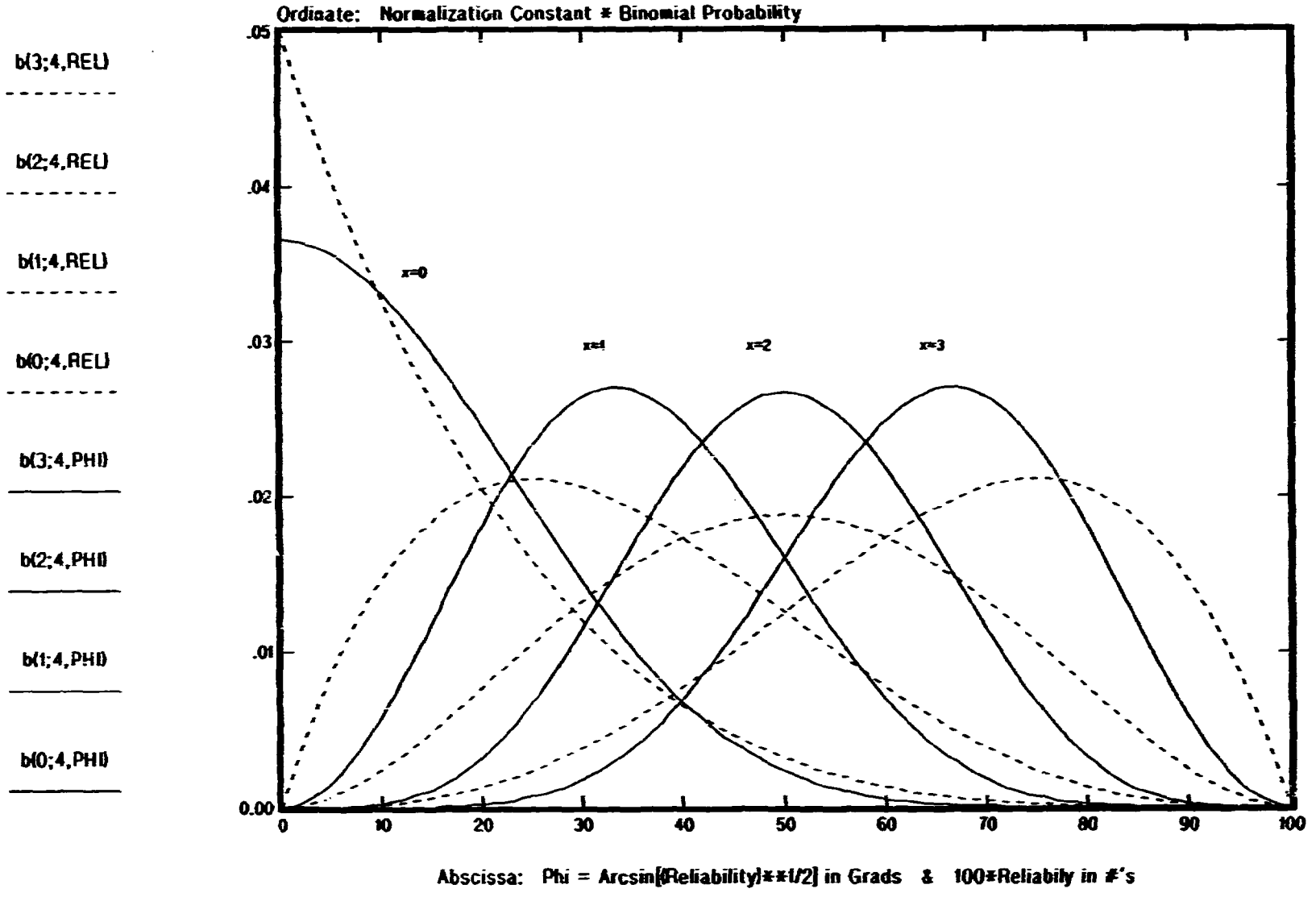


Figure 5

NORMALIZED BINOMIAL LIKELIHOODS

OF TRANSFORMED RELILIABILITY FOR N = 12

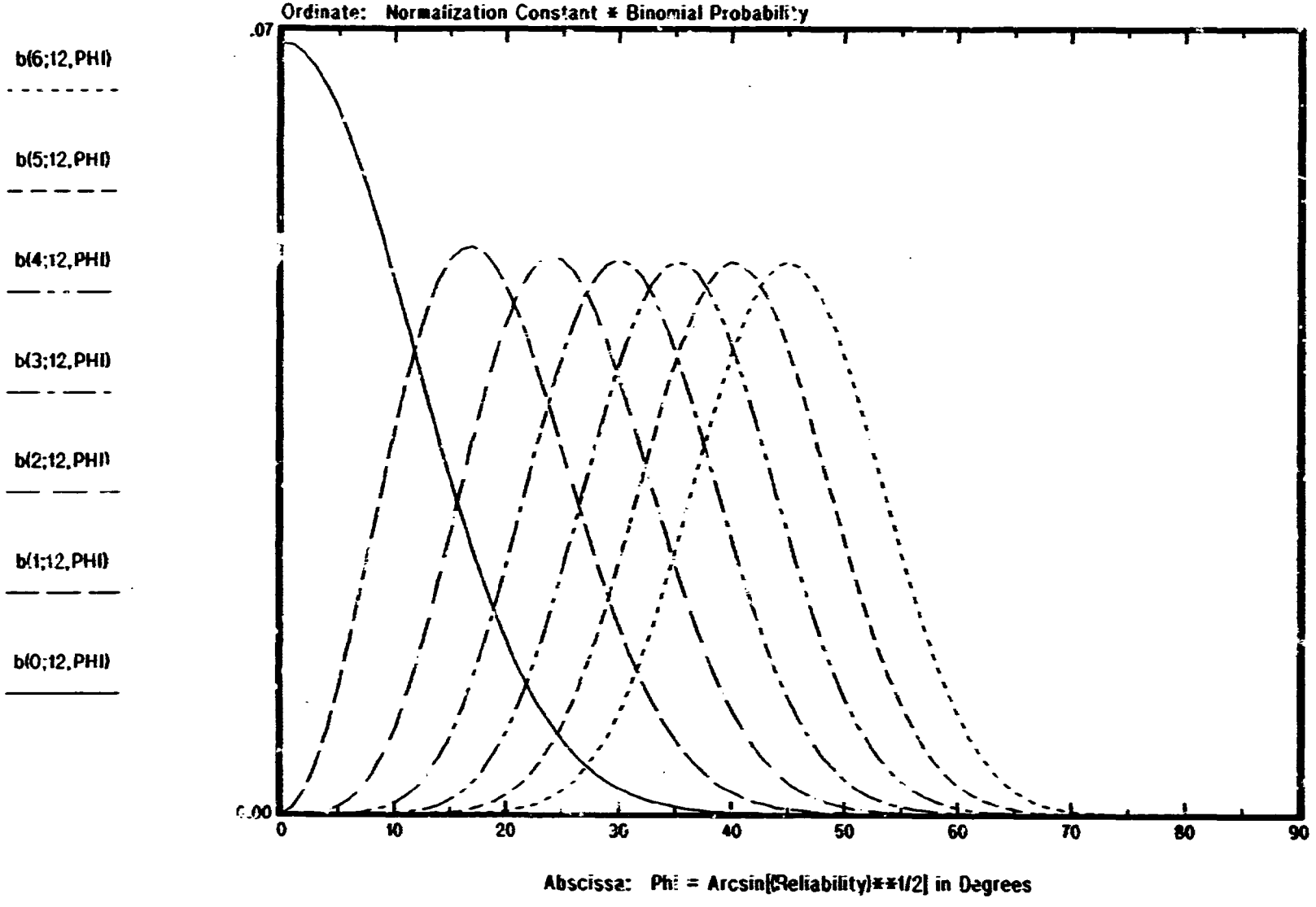


Figure 6

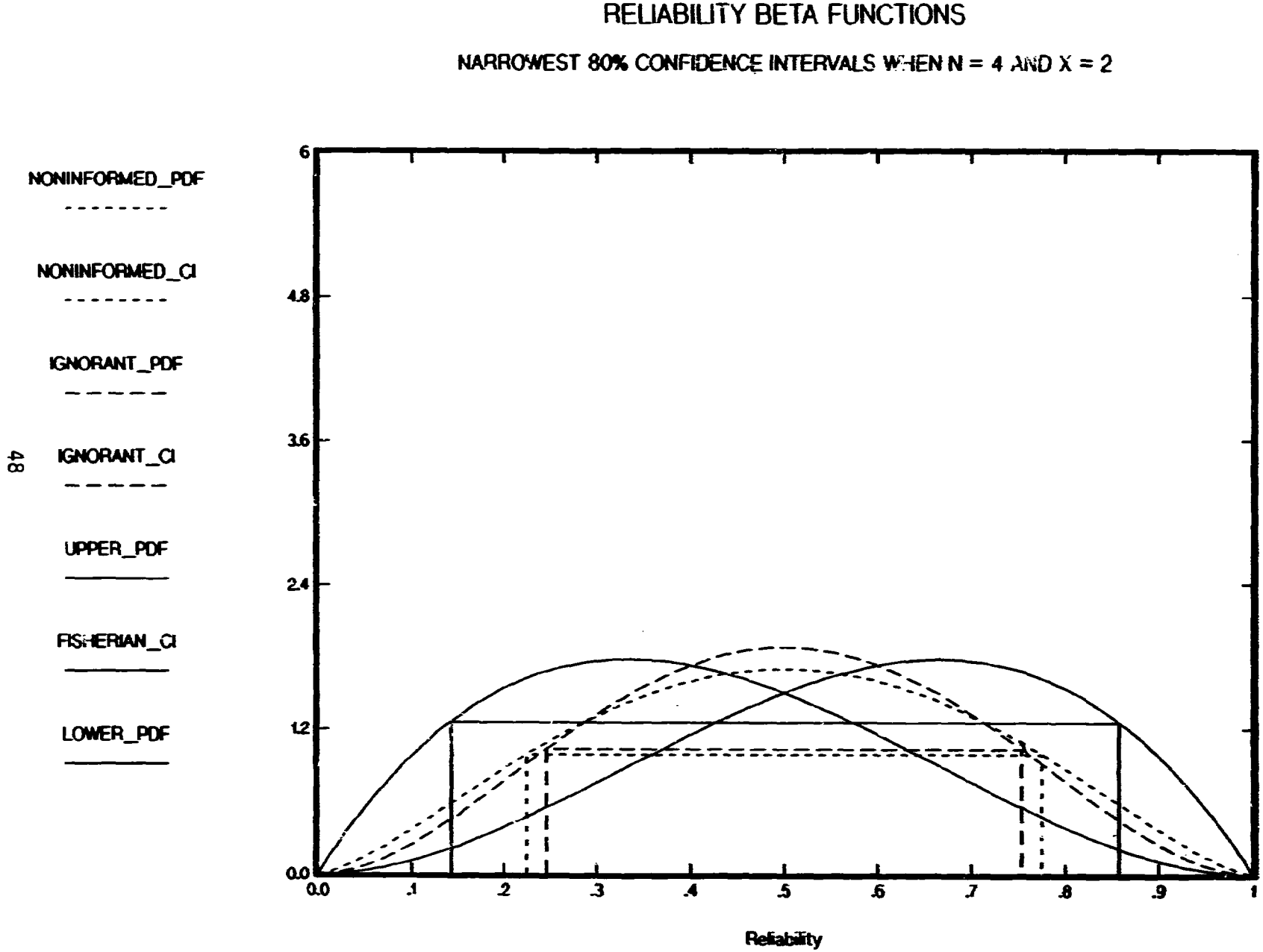

RELIABILITY BETA FUNCTIONS
NARROWEST 80% CONFIDENCE INTERVALS WHEN N = 4 AND X = 2
NONINFORMED_PDF
NONINFORMED_CI
IGNORANT_PDF
IGNORANT_CI
UPPER_PDF
FISHERIAN_CI
LOWER_PDF
6
4.8
3.6
2.4
1.2
0.0
0.0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
Reliability

Figure 7

RELIABILITY BETA FUNCTIONS

NARROWEST 80% CONFIDENCE INTERVALS WHEN N = 4 AND X = 3

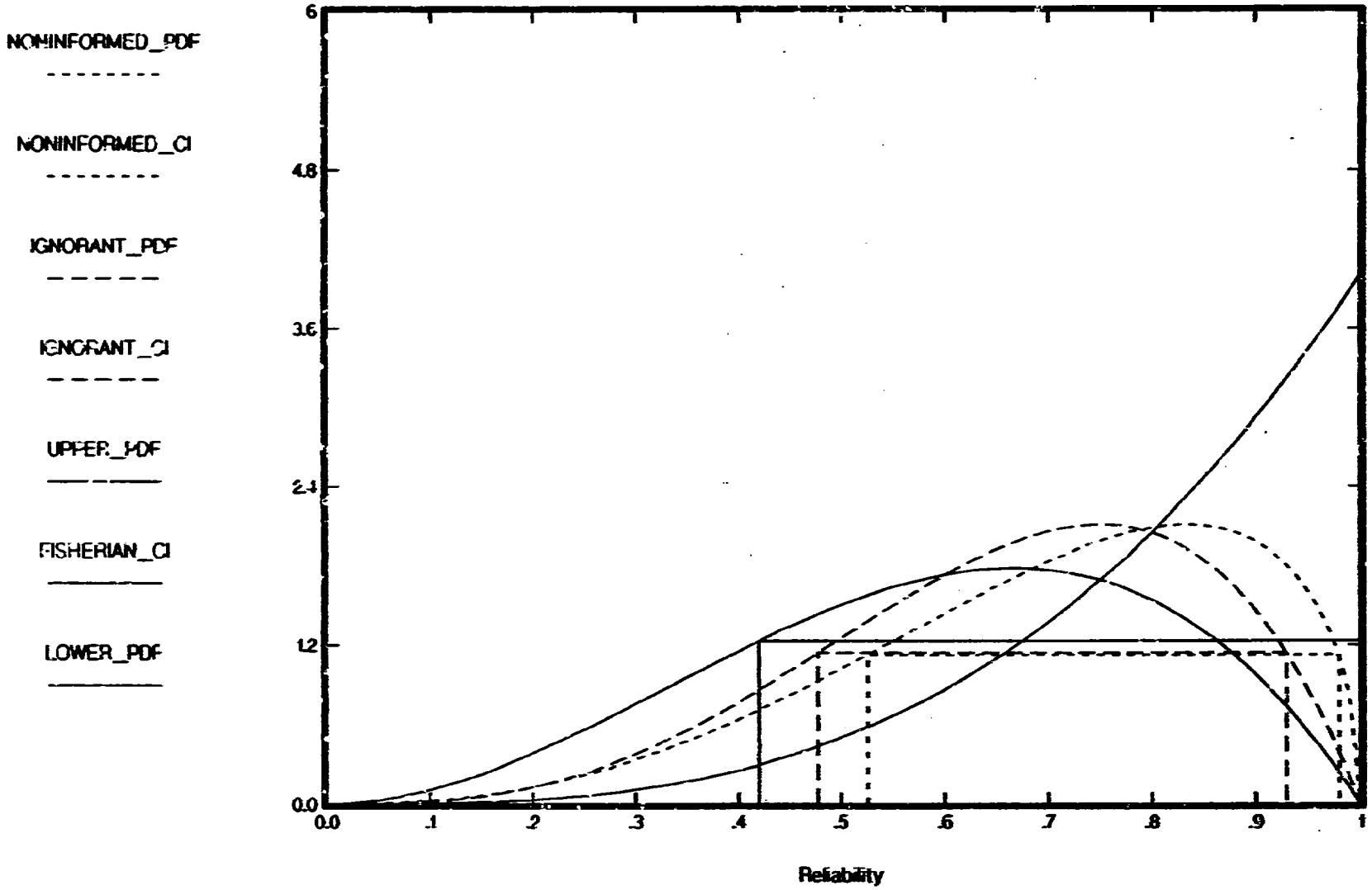


Figure 8

RELIABILITY BETA FUNCTIONS

NARROWEST 80% CONFIDENCE INTERVALS WHEN N = 4 AND X = 4

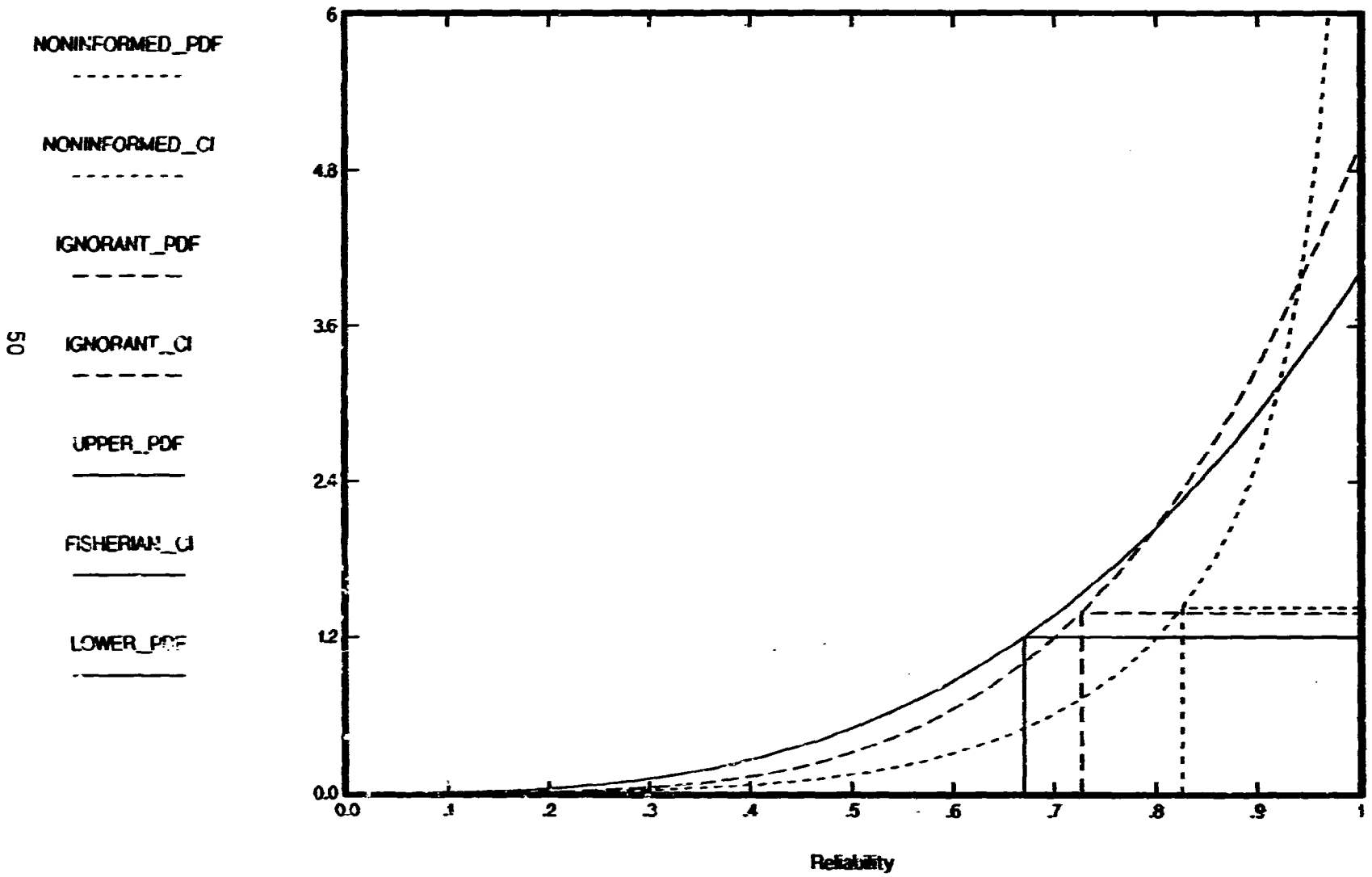


Figure 9

RELIABILITY BETA FUNCTIONS

NARROWEST 80% CONFIDENCE INTERVALS WHEN N = 12 AND X = 7

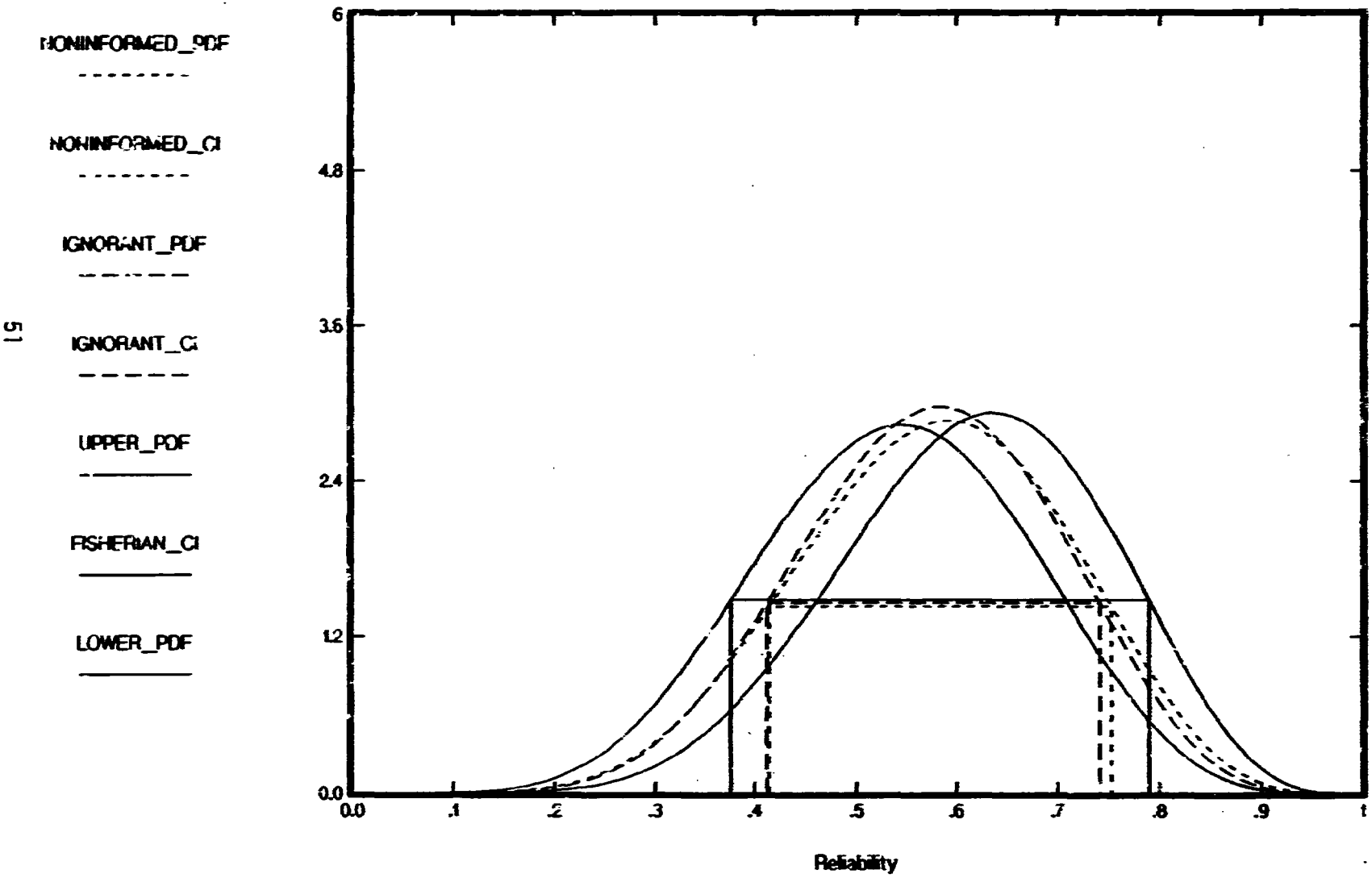


Figure 10

RELIABILITY BETA FUNCTIONS

NARROWEST 80% CONFIDENCE INTERVALS WHEN N = 12 AND X = 9

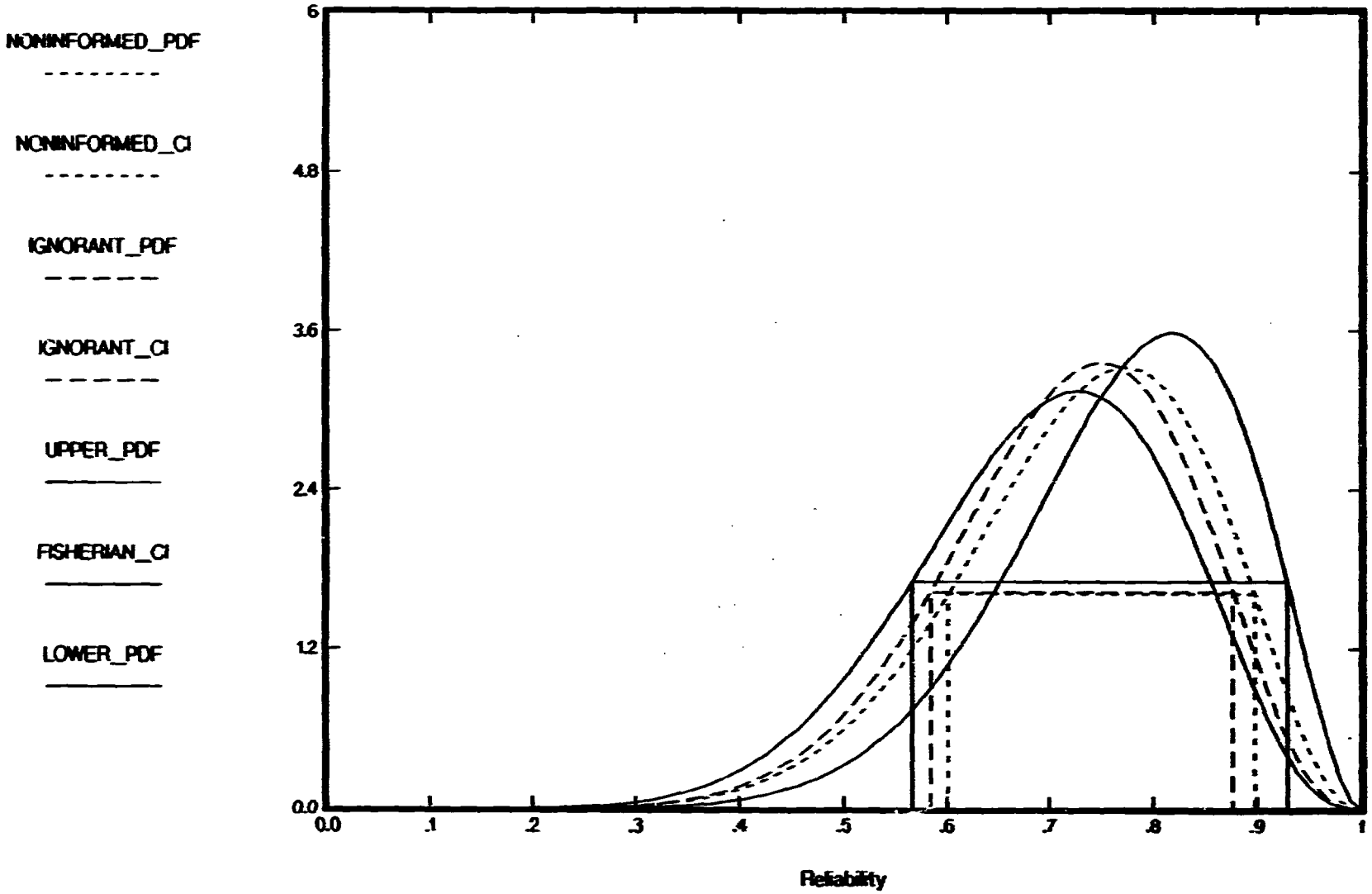


Figure 11

RELIABILITY BETA FUNCTIONS

NARROWEST 80% CONFIDENCE INTERVALS WHEN N = 12 AND X = 11

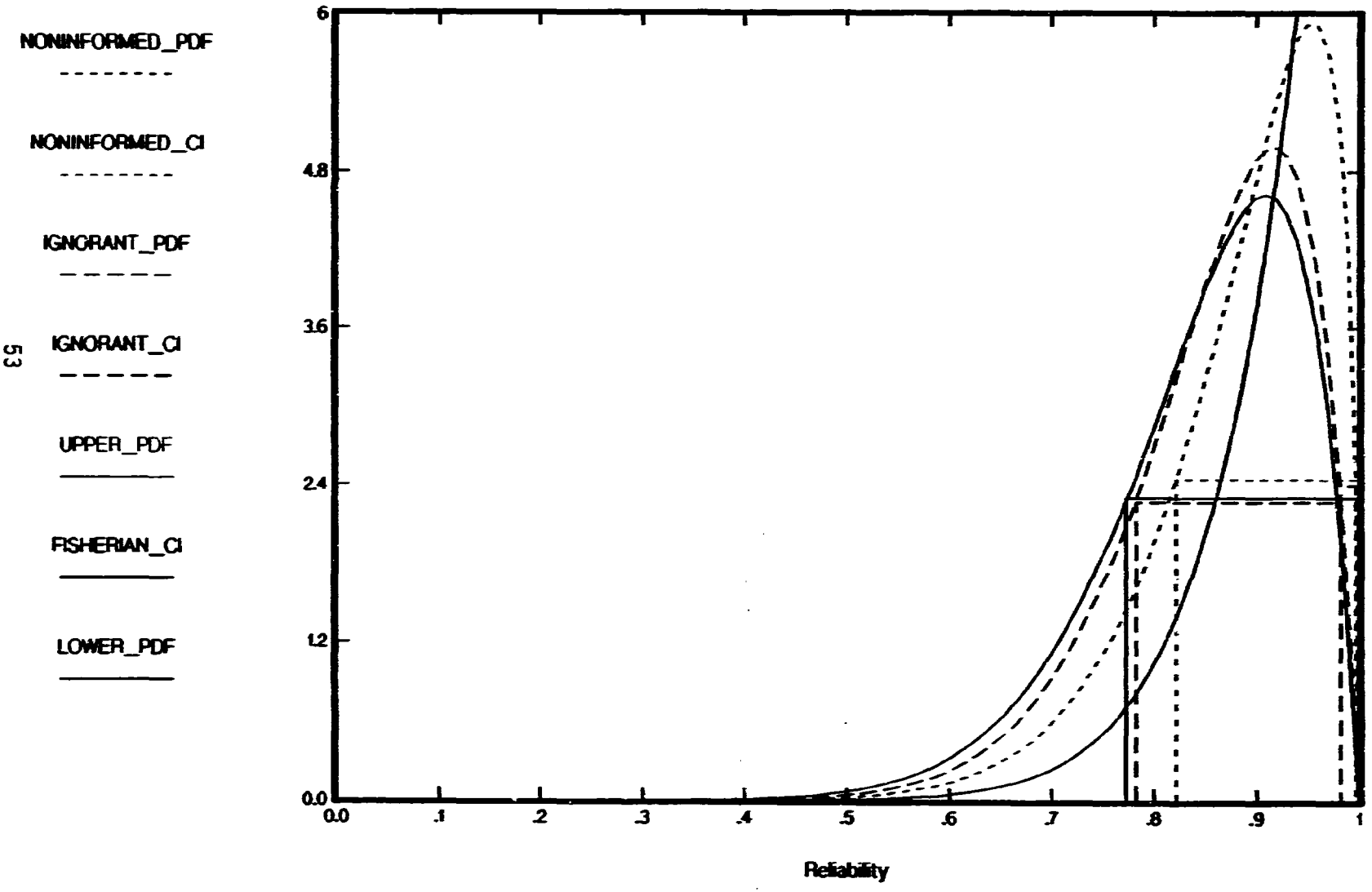


Figure 12

RELIABILITY BETA FUNCTIONS

NARROWEST 80% CONFIDENCE INTERVALS WHEN N = 24 AND X = 14

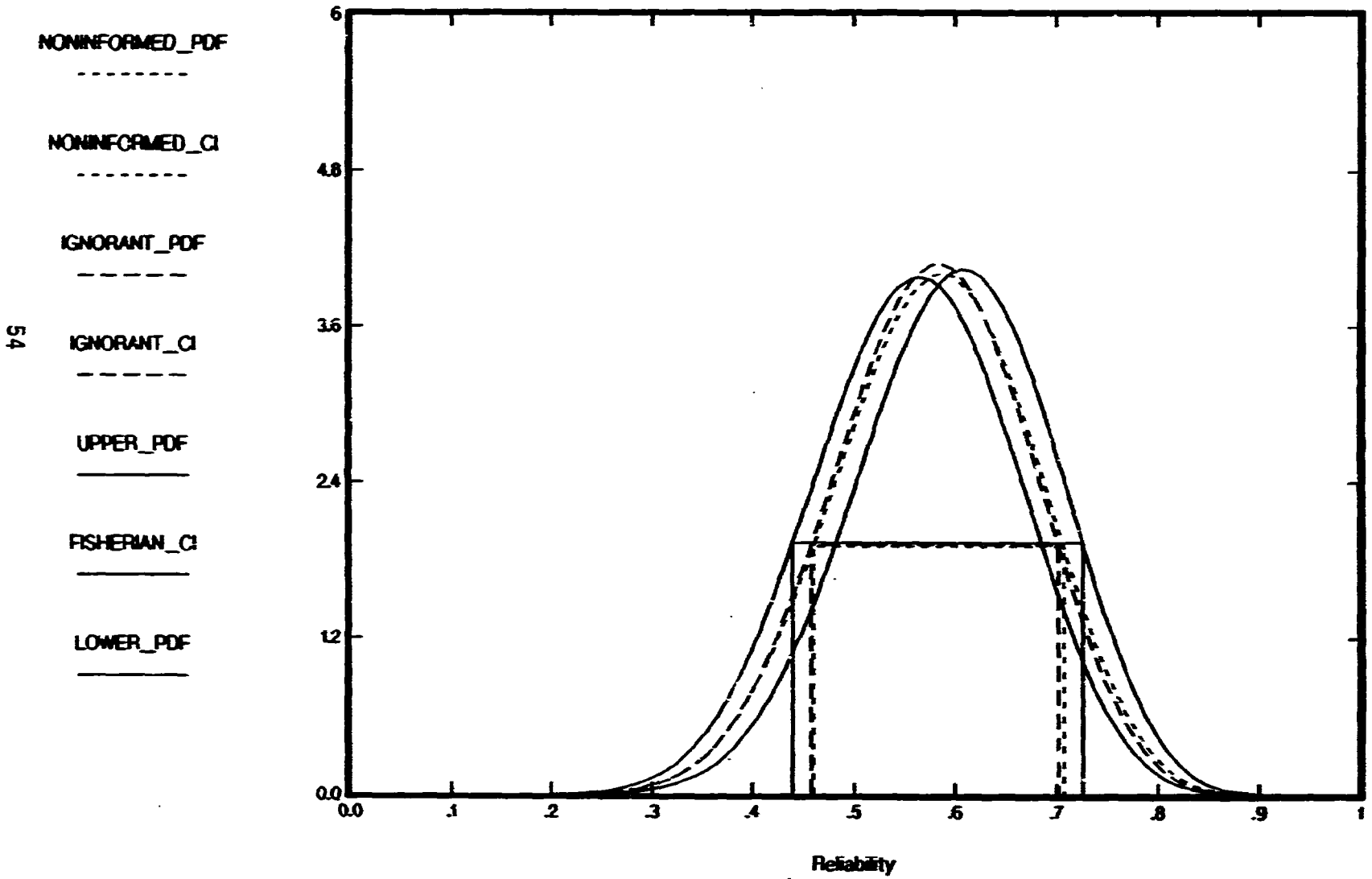


Figure 13

RELIABILITY BETA FUNCTIONS

NARROWEST 80% CONFIDENCE INTERVALS WHEN N = 24 AND X = 18

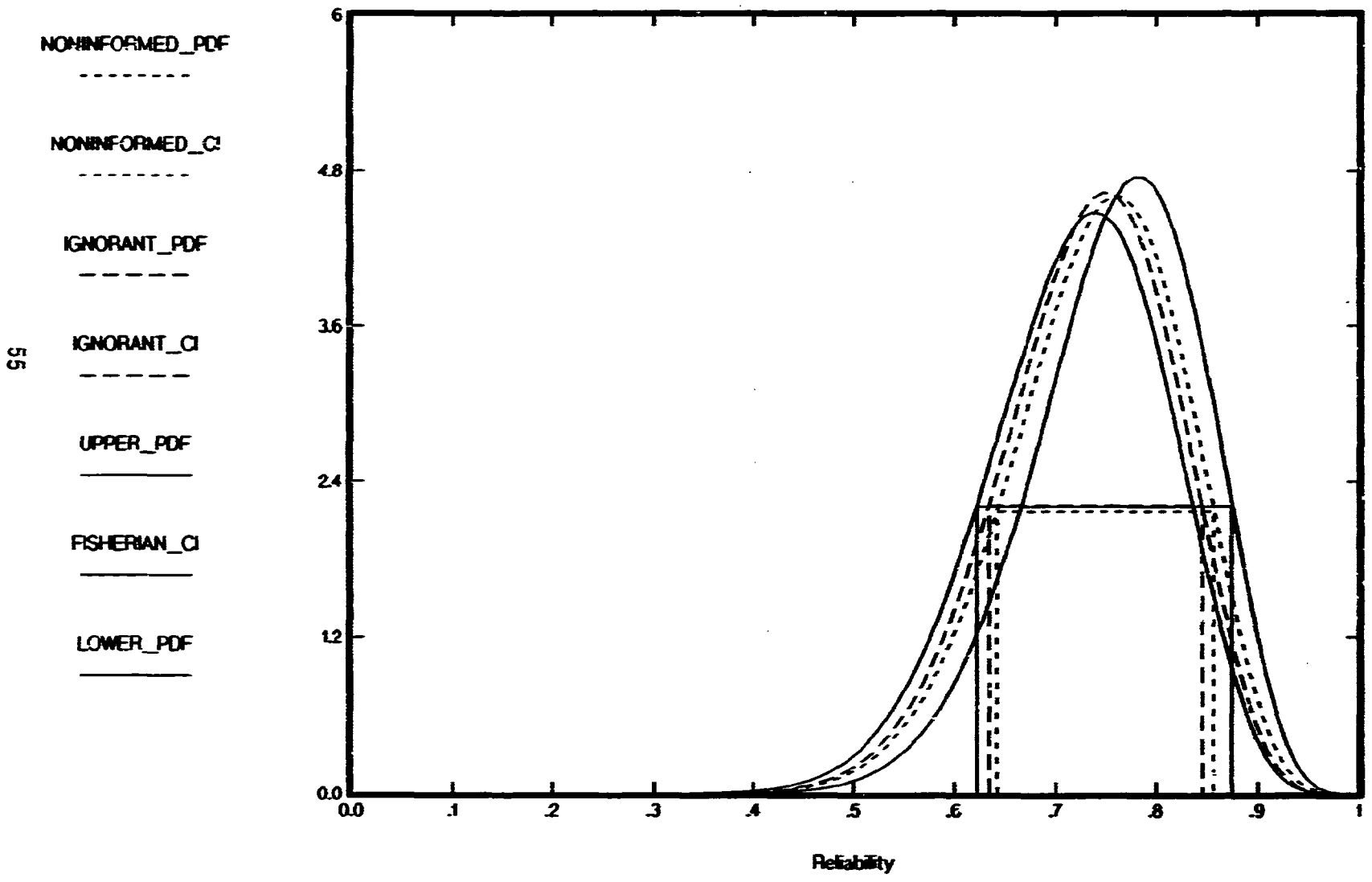


Figure 14

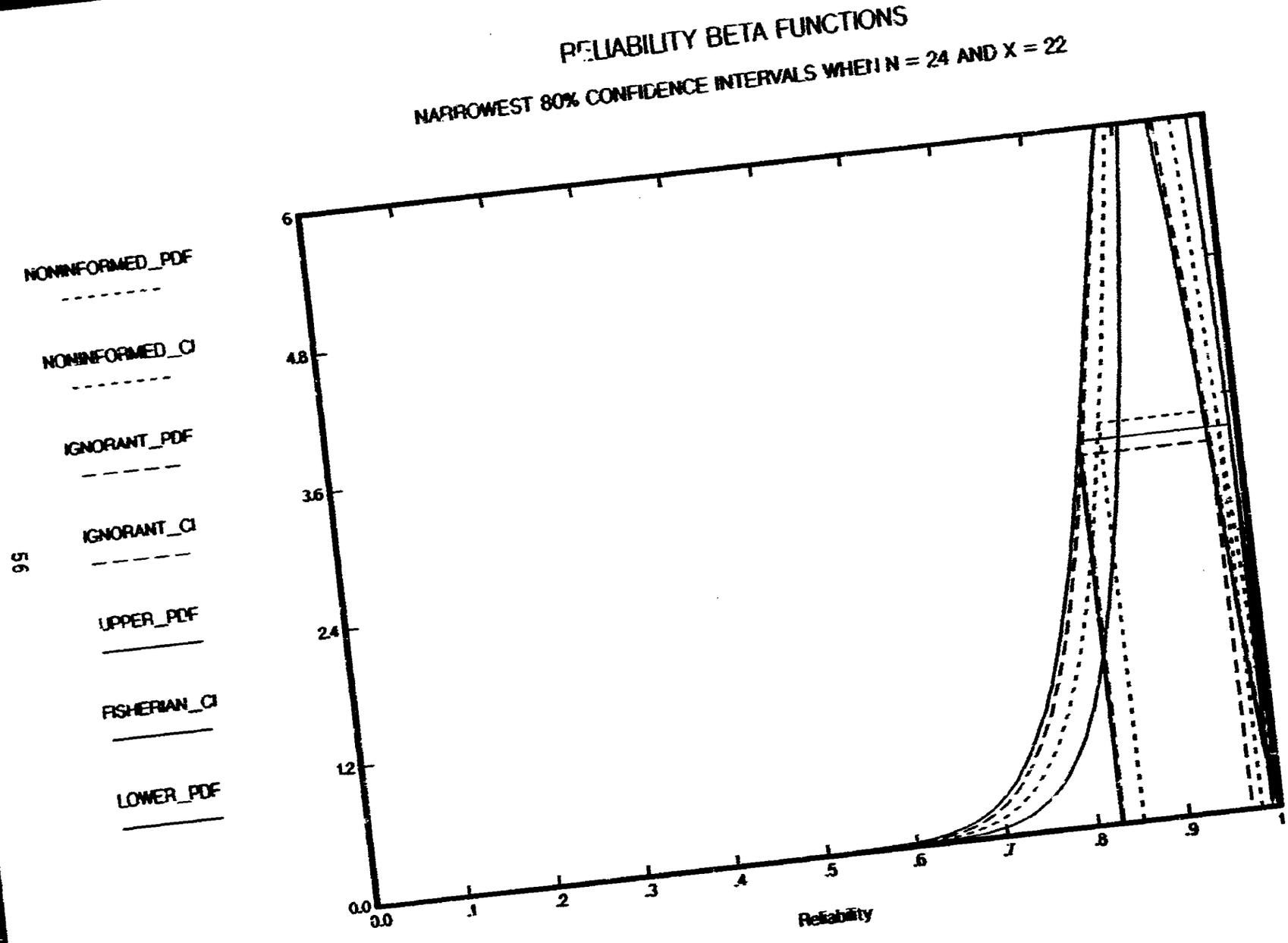

RELIABILITY BETA FUNCTIONS
NARROWEST 80% CONFIDENCE INTERVALS WHEN N = 24 AND X = 22
NONINFORMED_PDF
NONINFORMED_CI
IGNORANT_PDF
IGNORANT_CI
95
UPPER_PDF
FISHERIAN_CI
LOWER_PDF
6
4.8
3.6
2.4
1.2
0.0
0.0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
Reliability

Figure 15

RELIABILITY BETA FUNCTIONS

UPPER 90% CONFIDENCE INTERVALS WHEN N = 5 AND X = 3

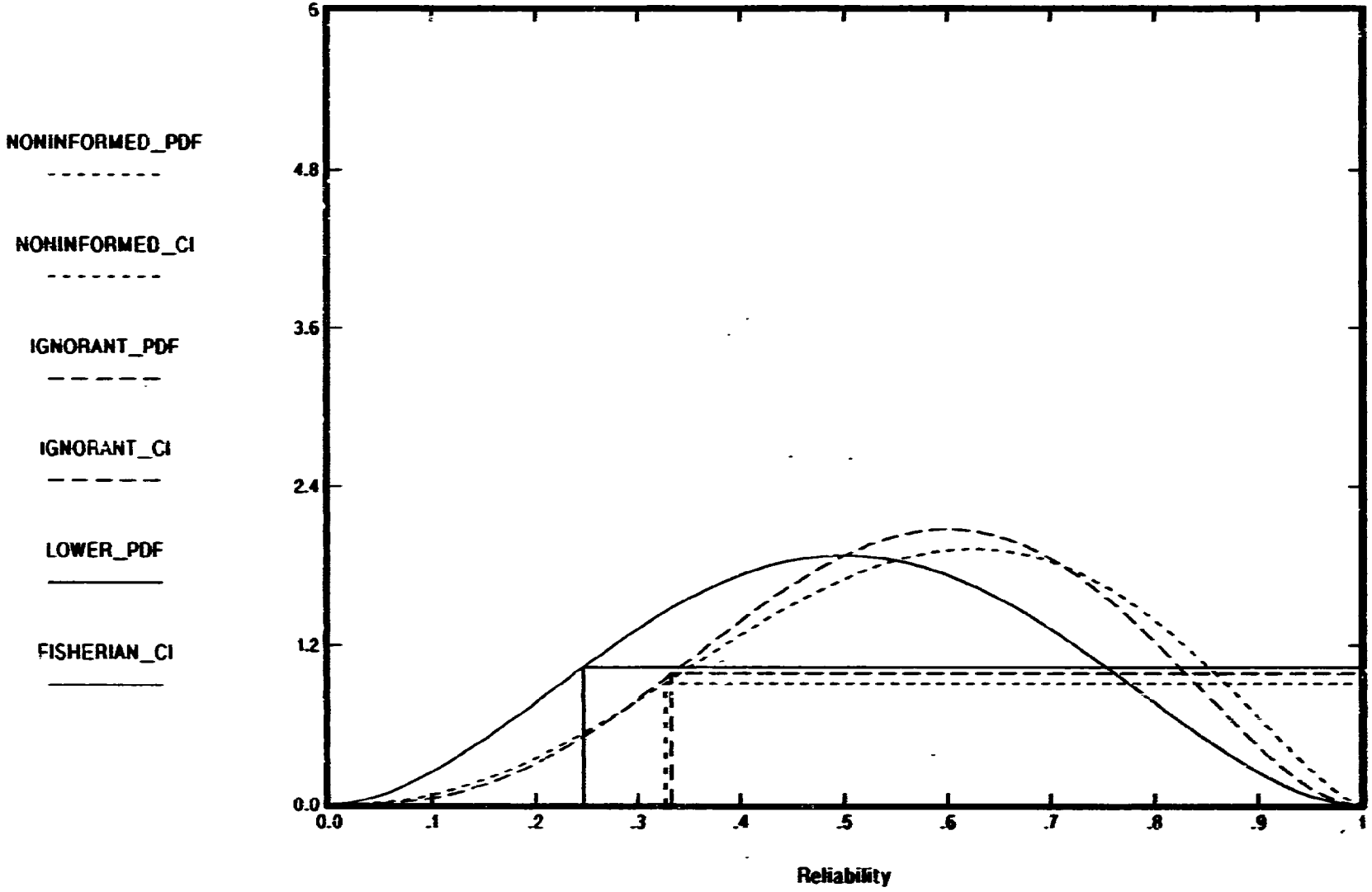


Figure 16

RELIABILITY BETA FUNCTIONS

UPPER 90% CONFIDENCE INTERVALS WHEN N = 5 AND X = 4

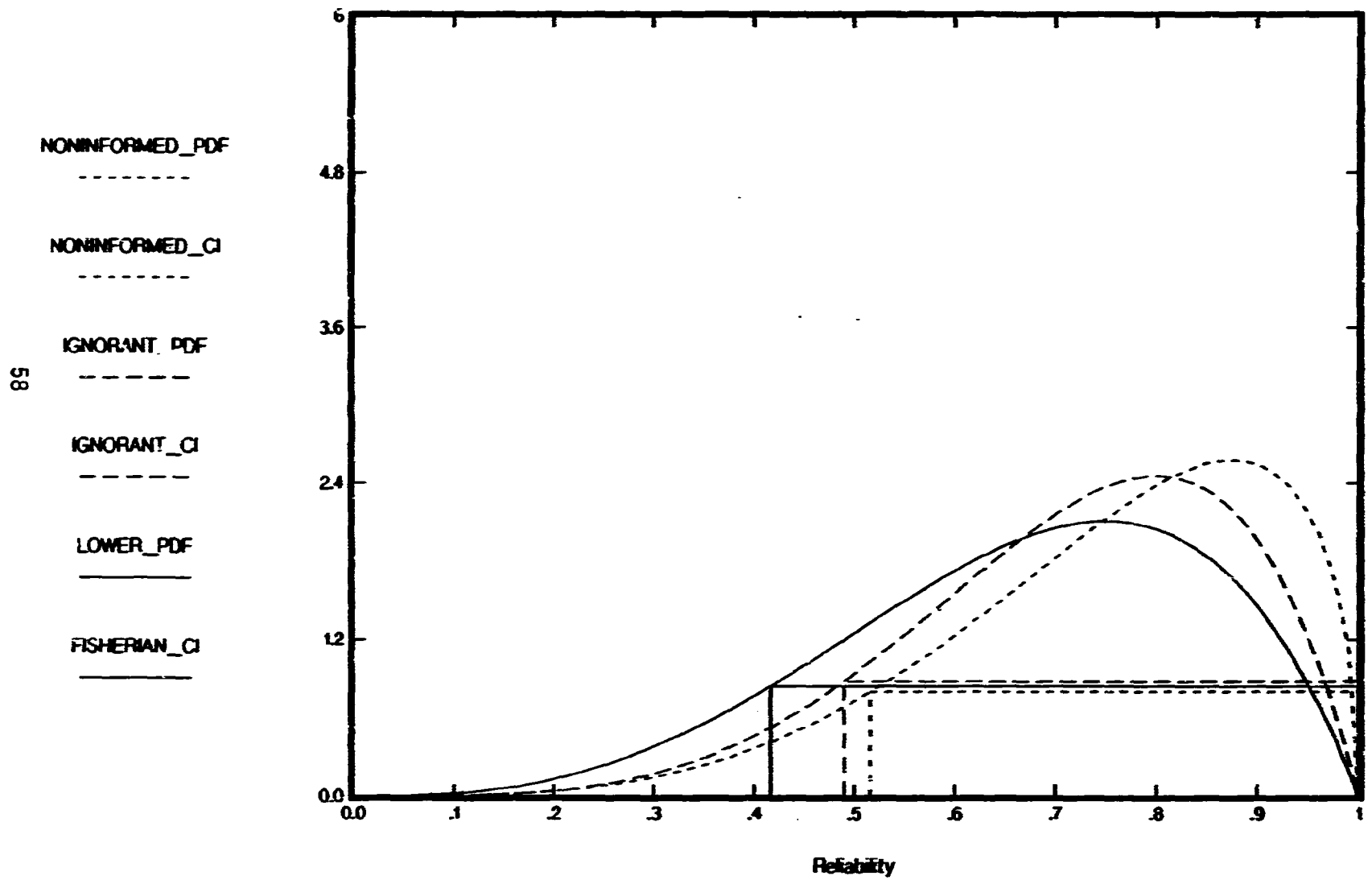


Figure 17

RELIABILITY BETA FUNCTIONS

UPPER 90% CONFIDENCE INTERVALS WHEN N = 5 AND X = 5

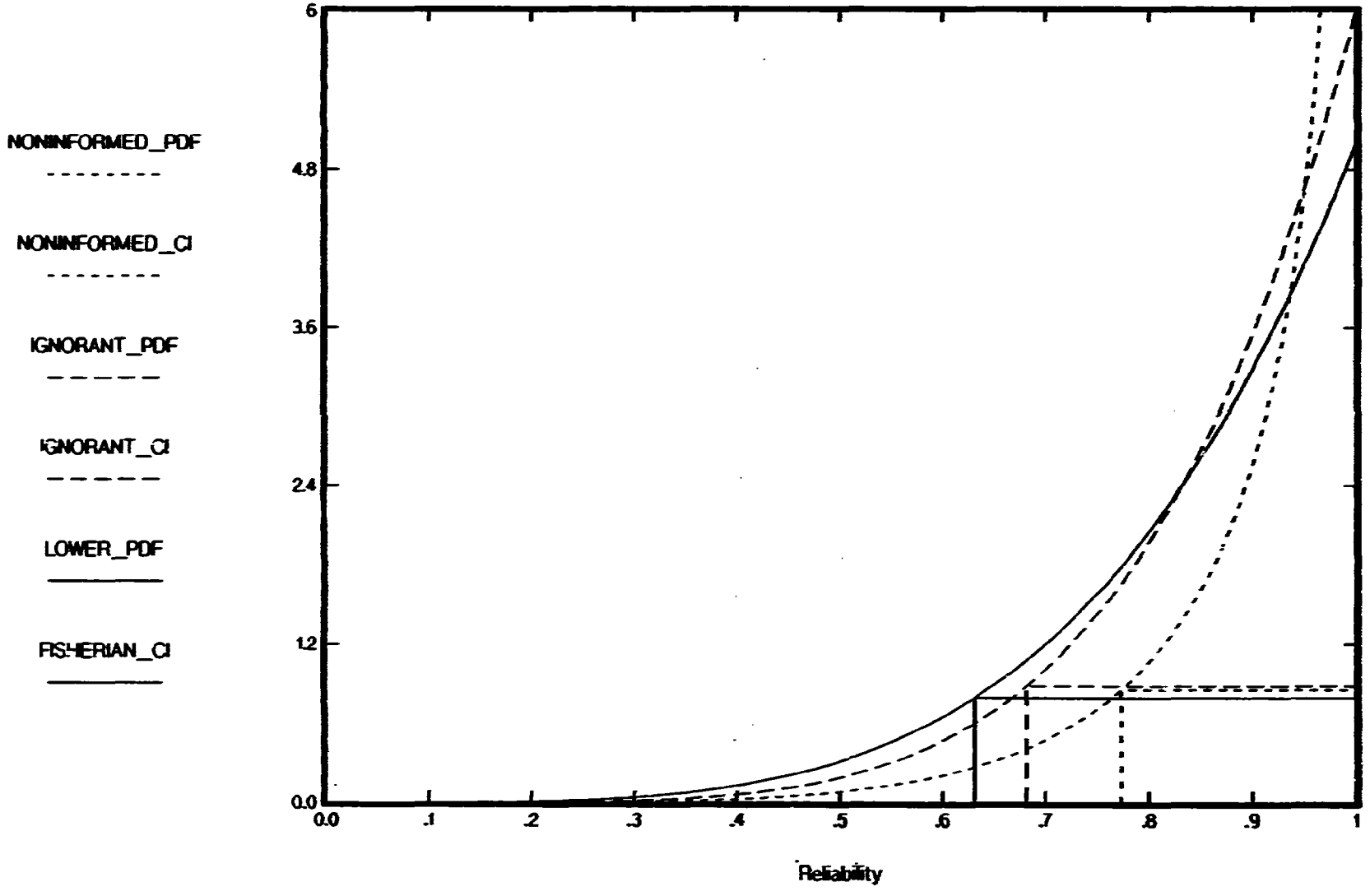


Figure 18

RELIABILITY BETA FUNCTIONS

UPPER 90% CONFIDENCE INTERVALS WHEN N = 15 AND X = 9

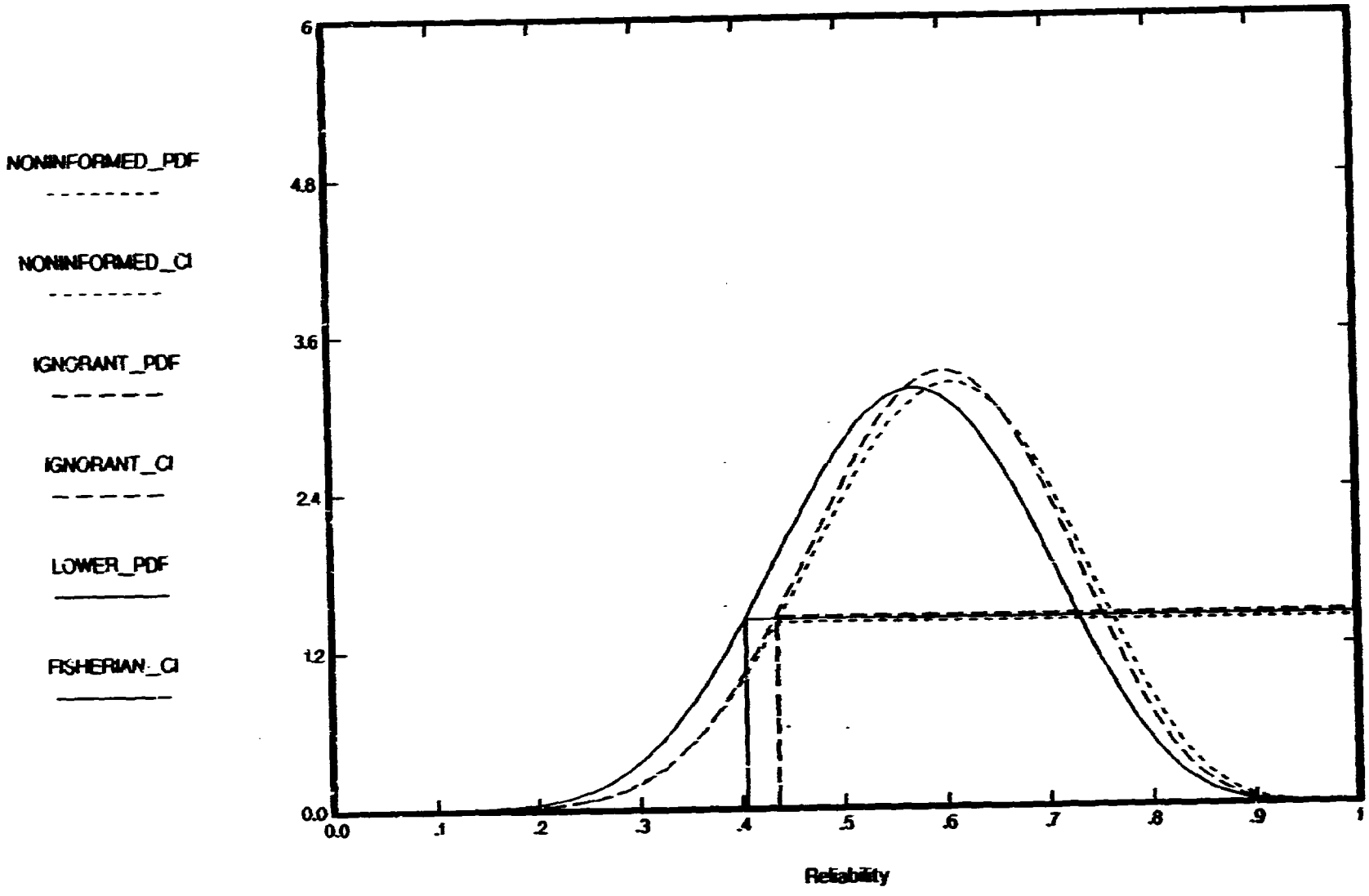


Figure 19

RELIABILITY BETA FUNCTIONS

UPPER 90% CONFIDENCE INTERVALS WHEN N = 15 AND X = 12

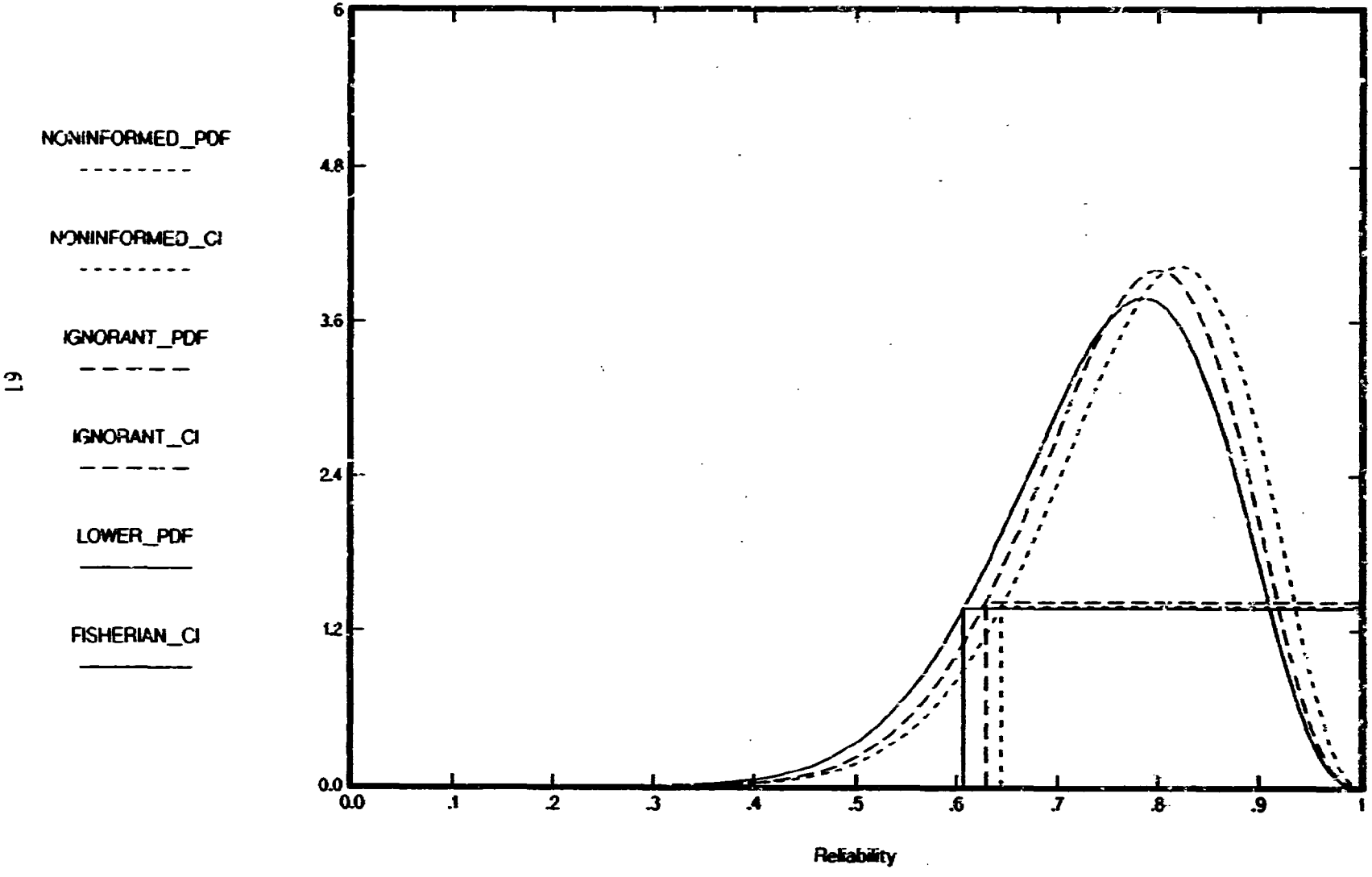


Figure 20

RELIABILITY BETA FUNCTIONS

UPPER 90% CONFIDENCE INTERVALS WHEN N = 15 AND X = 14

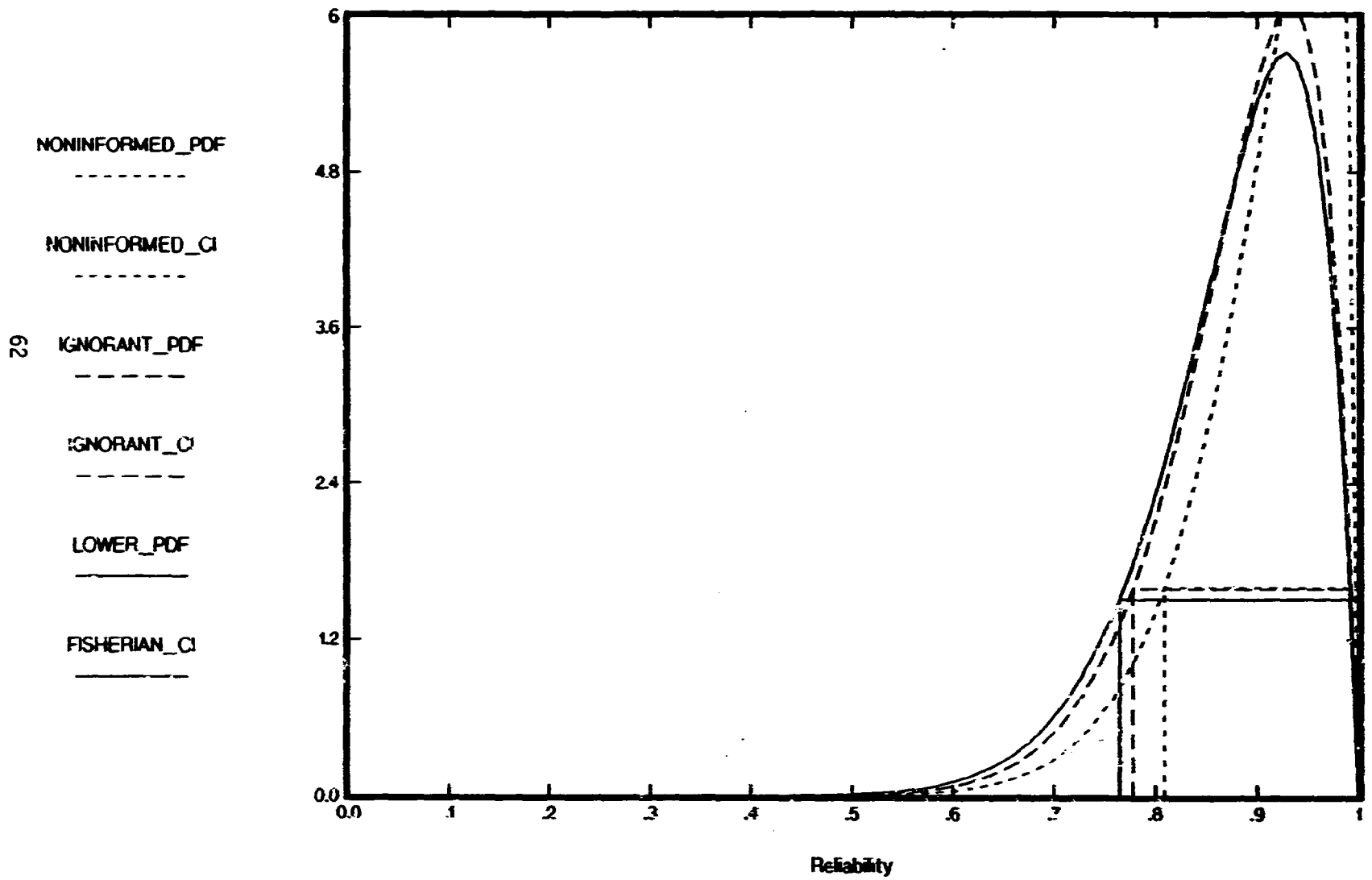


Figure 21

RELIABILITY BETA FUNCTIONS

UPPER 90% CONFIDENCE INTERVALS WHEN N = 30 AND X = 18

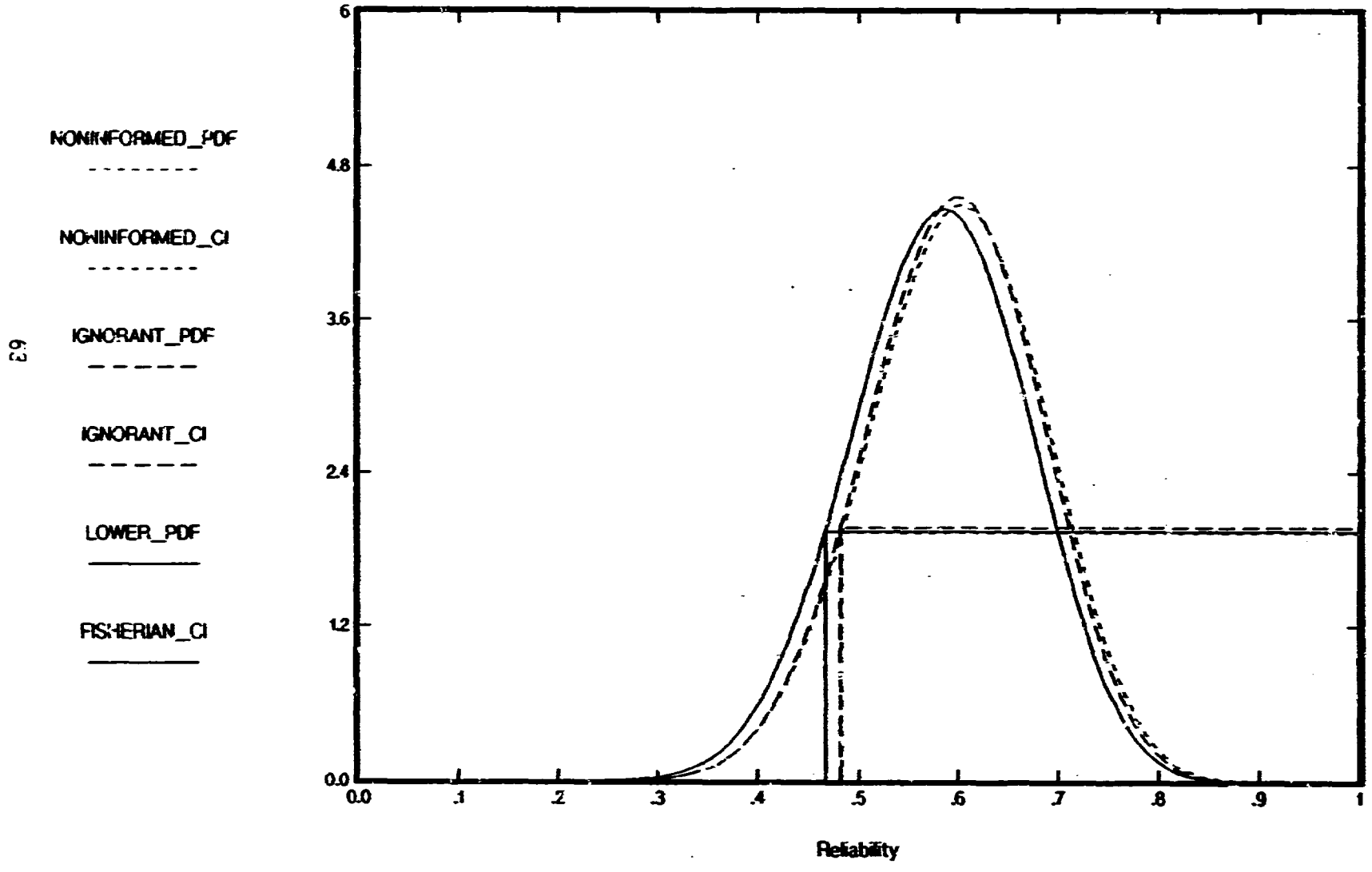


Figure 22

RELIABILITY BETA FUNCTIONS

UPPER 90% CONFIDENCE INTERVALS WHEN N = 30 AND X = 24

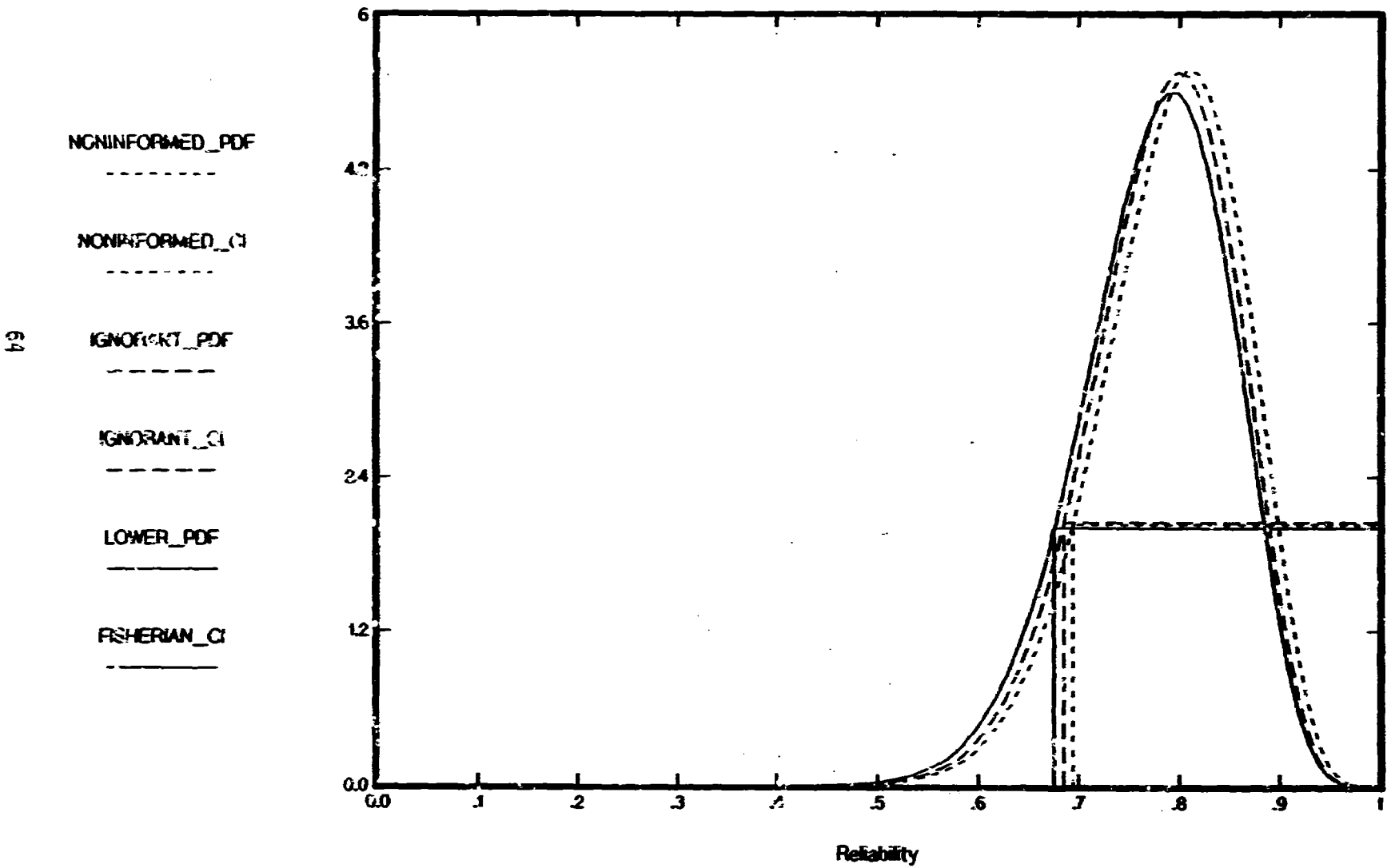


Figure 23

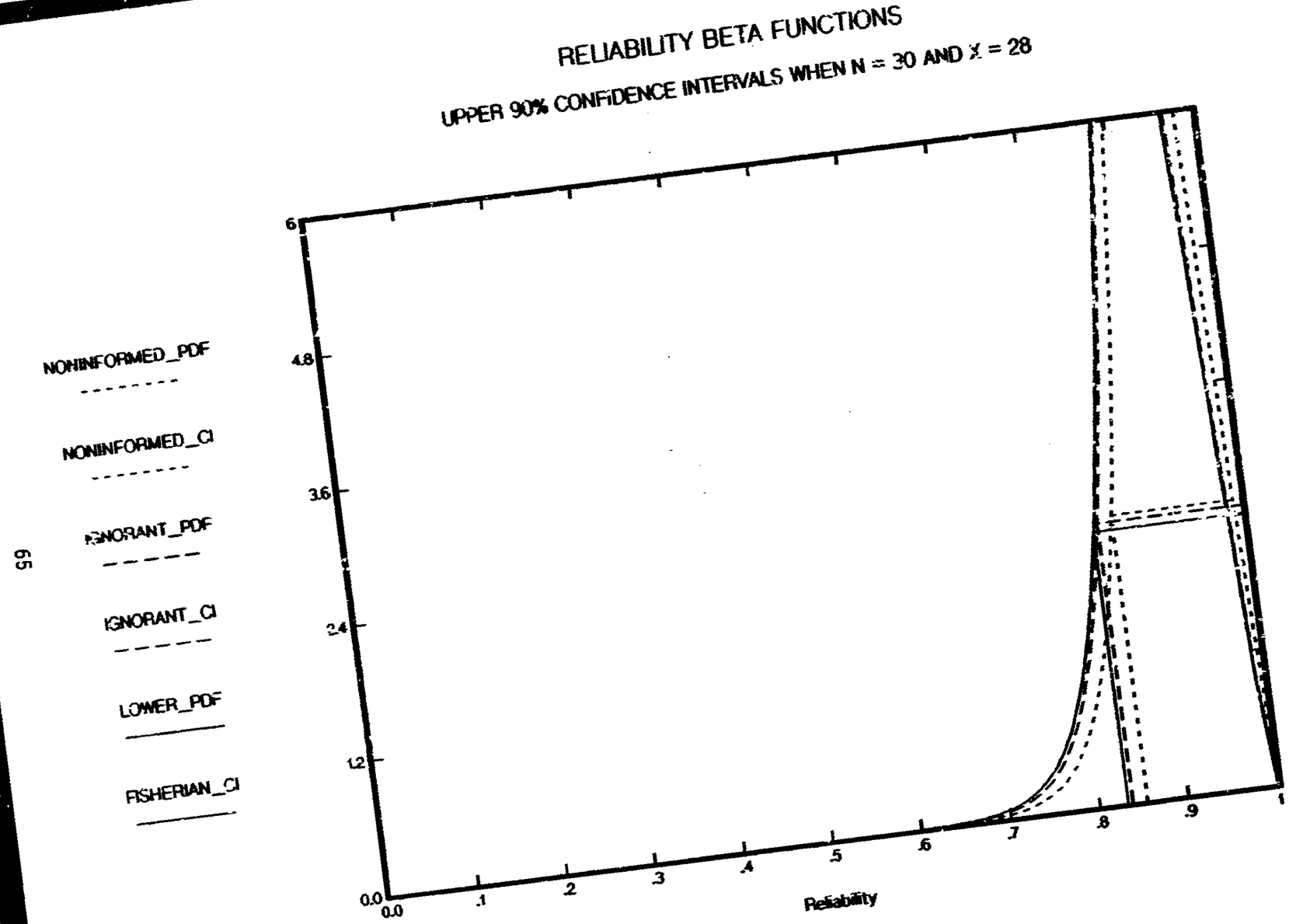

RELIABILITY BETA FUNCTIONS
UPPER 90% CONFIDENCE INTERVALS WHEN N = 30 AND X = 28
NONINFORMED_PDF
NONINFORMED_CI
IGNORANT_PDF
IGNORANT_CI
LOWER_PDF
FISHERIAN_CI
6
4.8
3.6
2.4
1.2
0.0
0.0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
Reliability

Figure 24

RELIABILITY BETA FUNCTIONS

UPPER 90% CONFIDENCE INTERVALS WHEN N = 60 AND X = 30

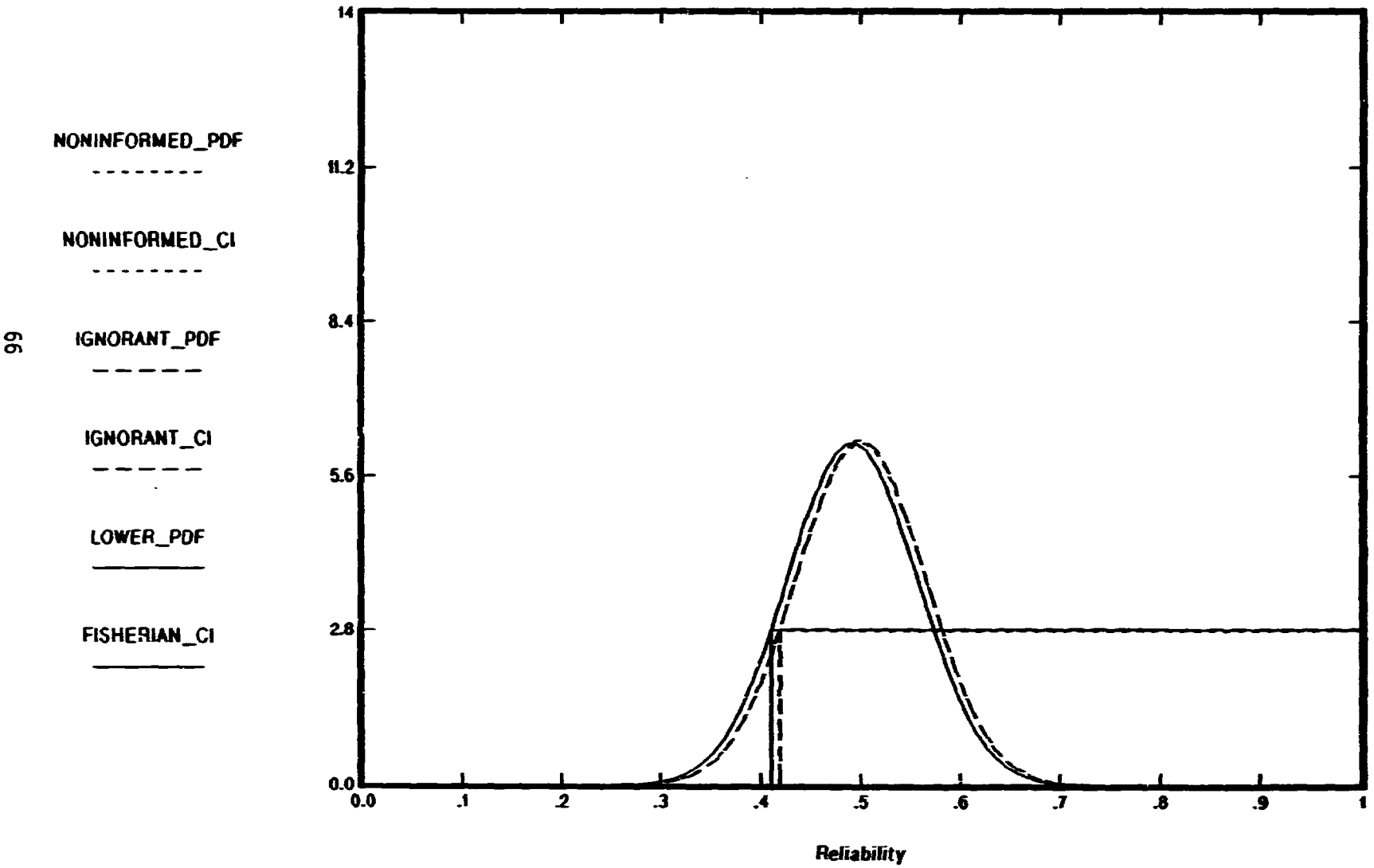


Figure 25

RELIABILITY BETA FUNCTIONS

UPPER 90% CONFIDENCE INTERVALS WHEN N = 60 AND X = 24

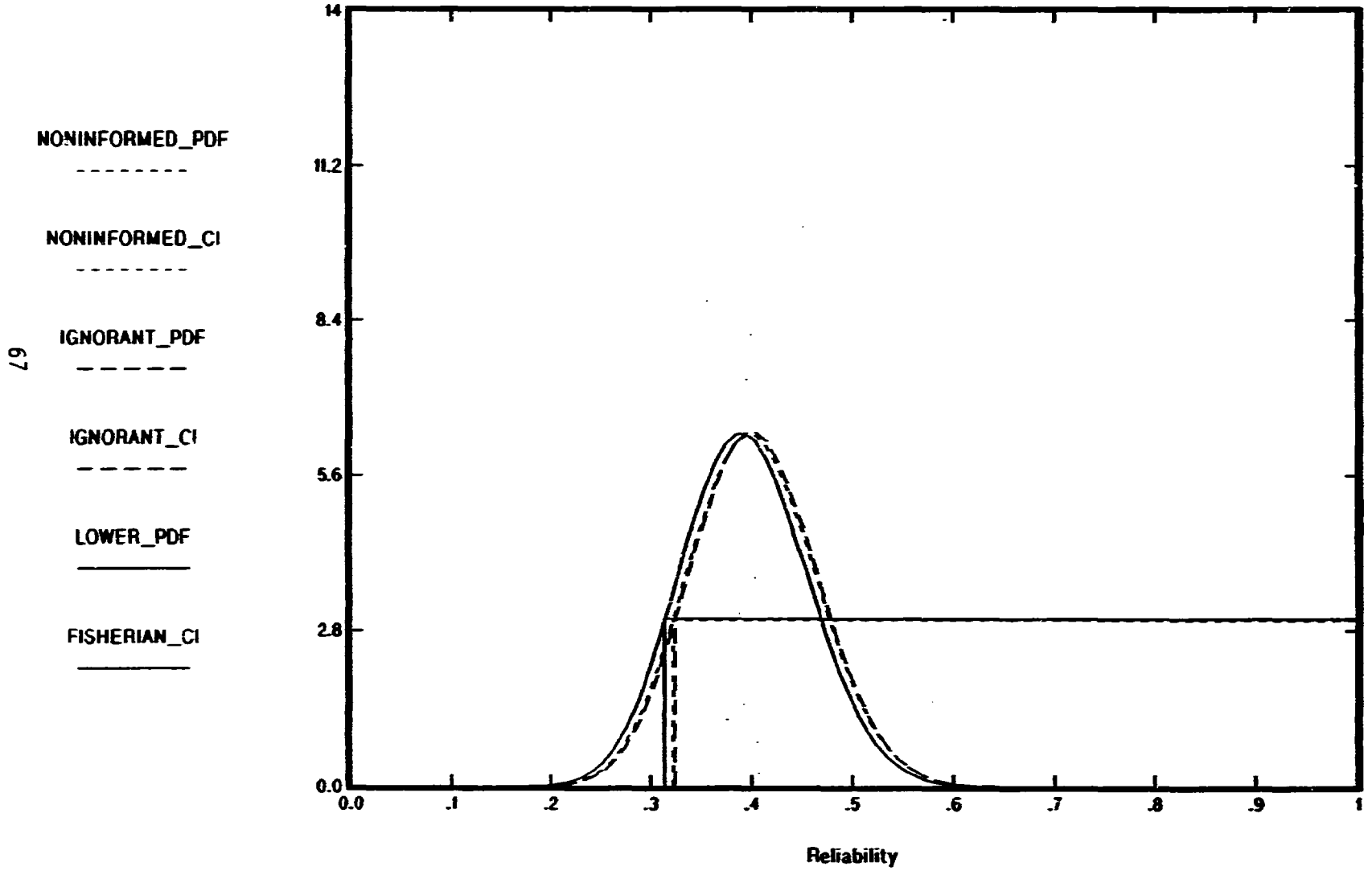


Figure 26

RELIABILITY BETA FUNCTIONS

UPPER 90% CONFIDENCE INTERVALS WHEN N = 60 AND X = 12

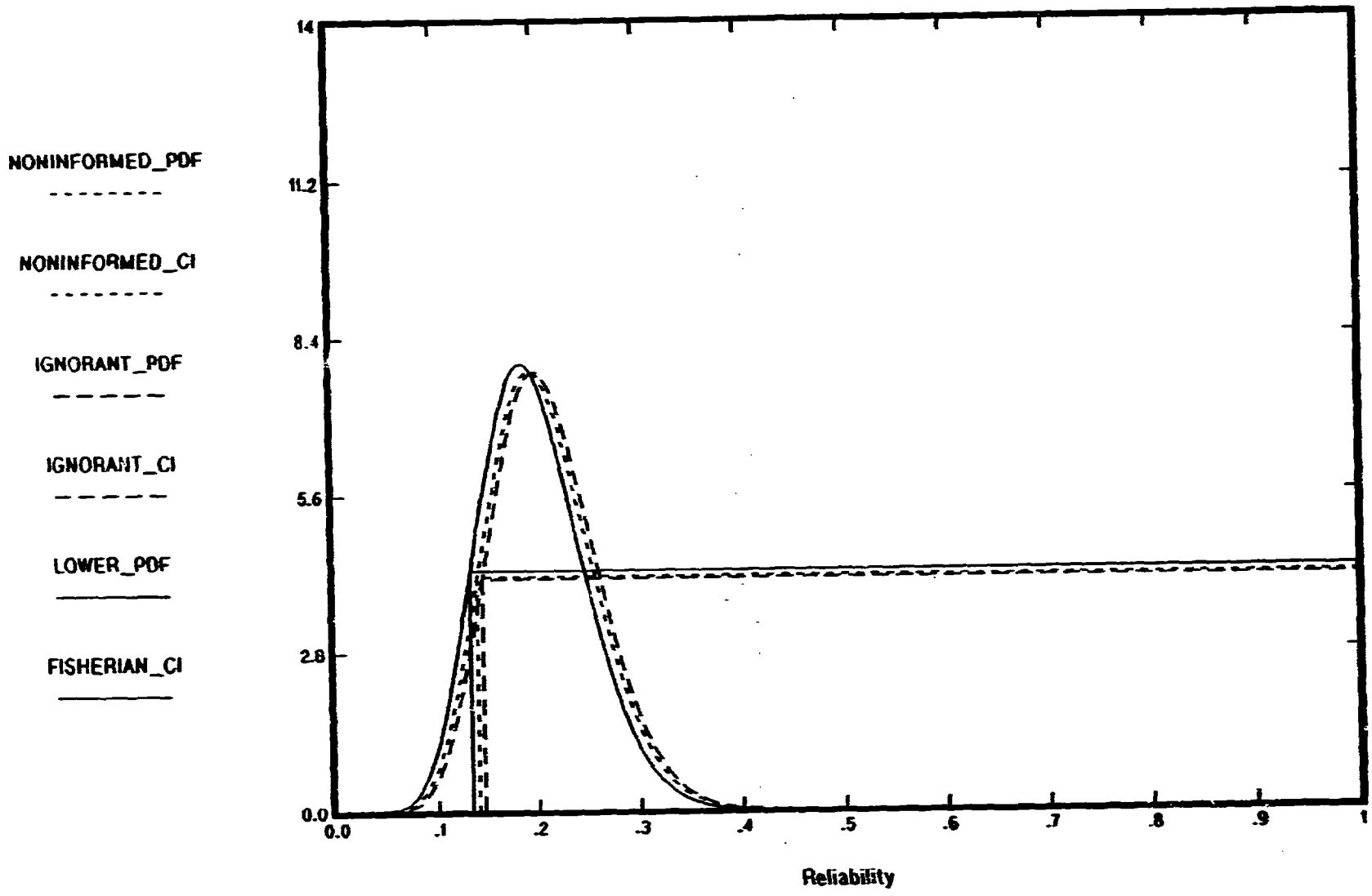


Figure 27

RELIABILITY BETA FUNCTIONS

UPPER 90% CONFIDENCE INTERVALS WHEN N = 60 AND X = 4

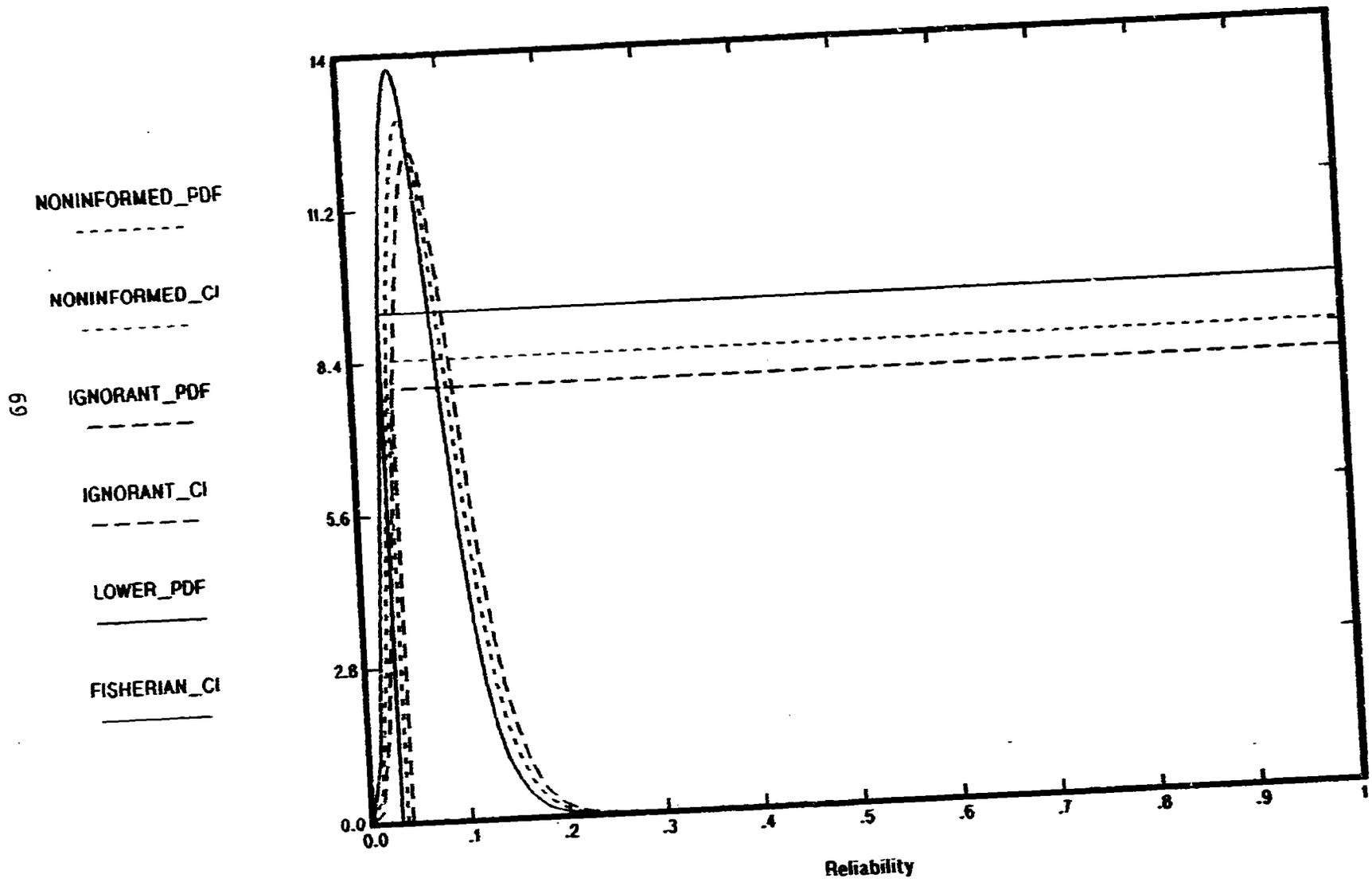


Figure 28

RELIABILITY BETA FUNCTIONS

UPPER 90% CONFIDENCE INTERVALS WHEN N = 60 AND X = 36

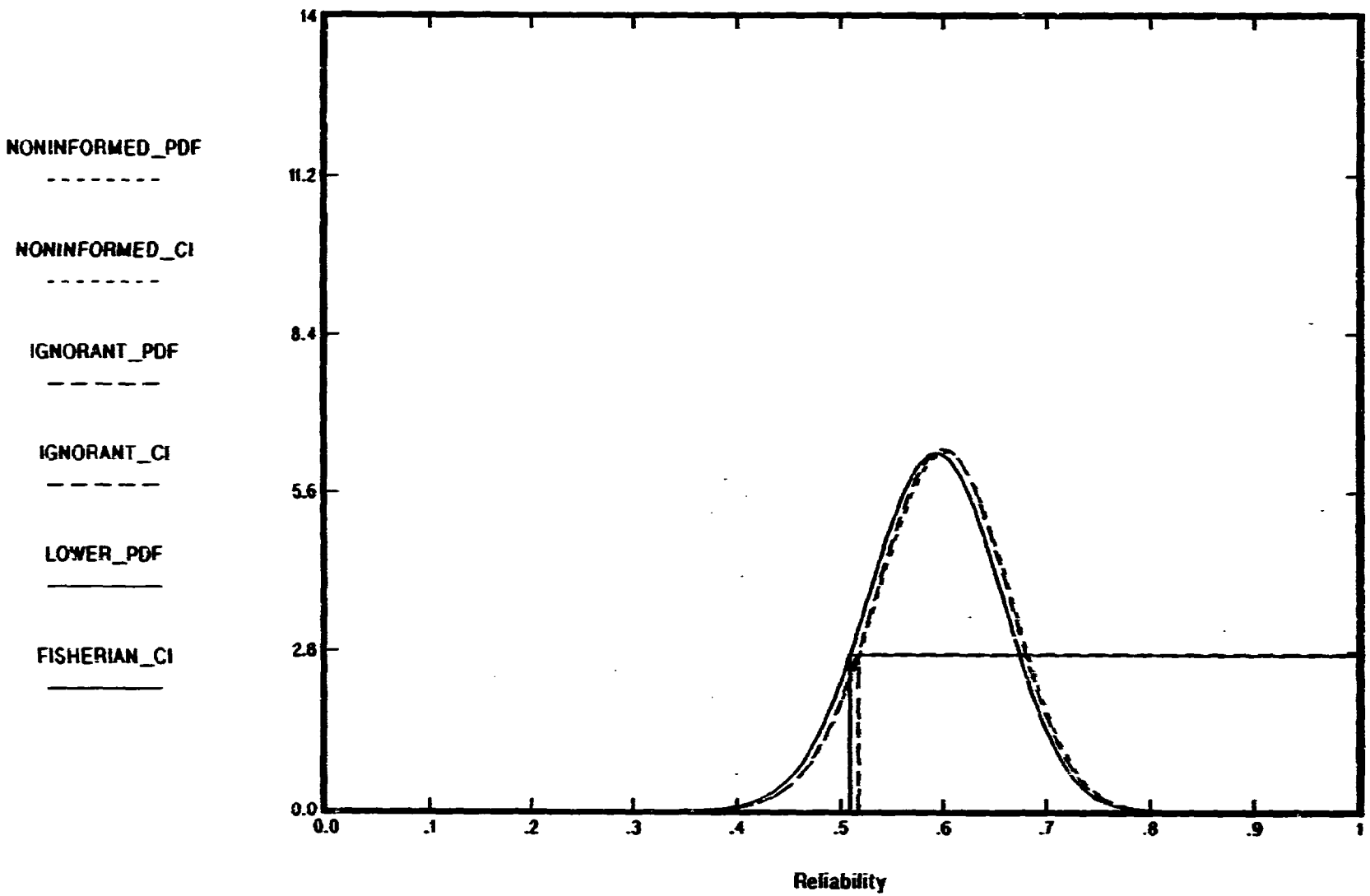


Figure 29

RELIABILITY BETA FUNCTIONS

UPPER 90% CONFIDENCE INTERVALS WHEN N = 60 AND X = 48

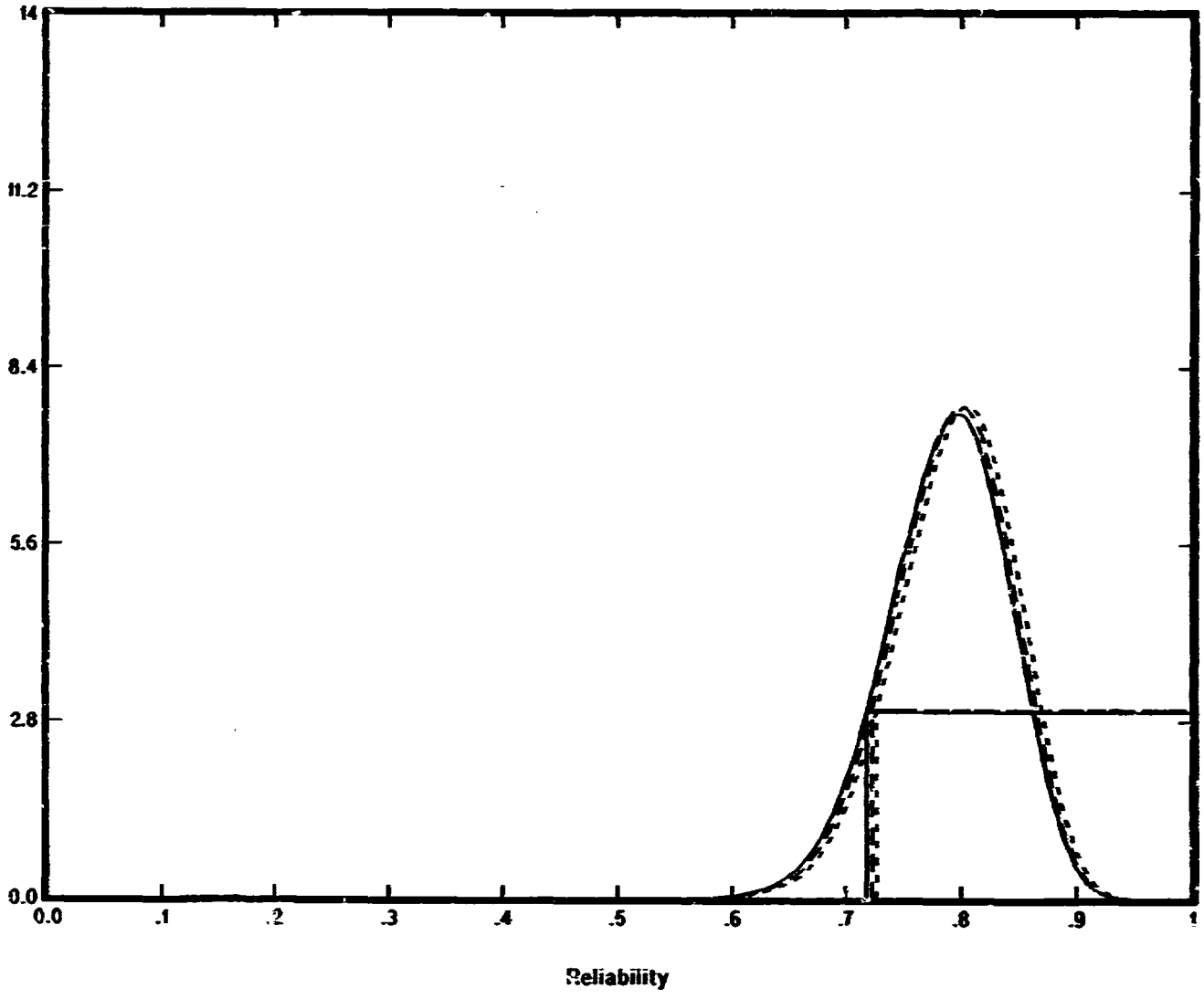


Figure 30

RELIABILITY BETA FUNCTIONS

UPPER 90% CONFIDENCE INTERVALS WHEN N = 60 AND X = 56

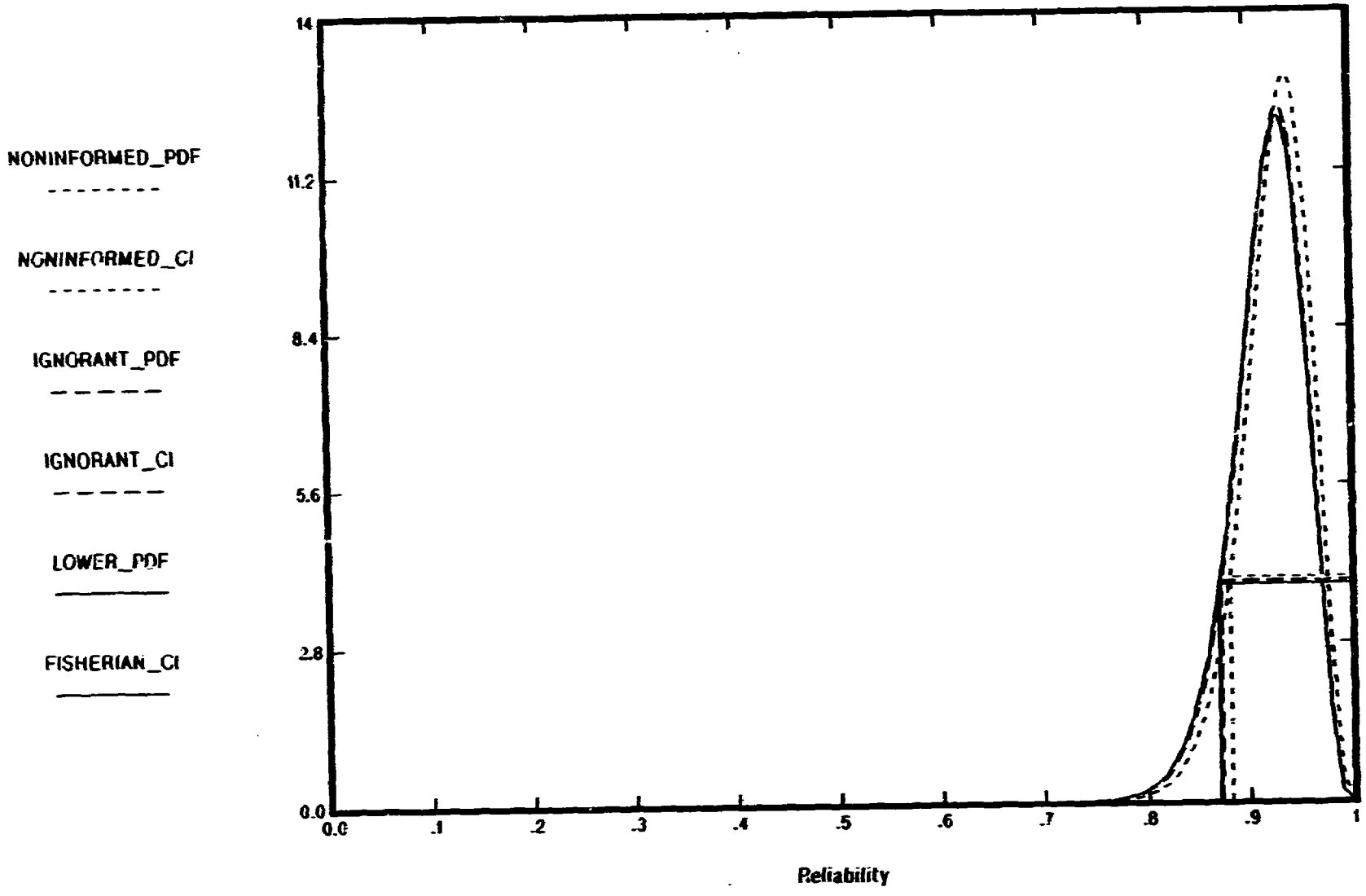


Figure 31

Environmental Sampling: A Case Study

Dennis L. Brandon
US Army Engineer
Waterways Experiment Station
Vicksburg, Mississippi 39180

Abstract Sampling strategies have been developed to accomplish various environmental objectives. The objectives may be: (1) to estimate the average of characteristics in a population; (2) to estimate the variability of characteristics of interest in a population; (3) to decide if characteristics of interest in a population meet certain standards or criteria; (4) to identify the source(s) which caused characteristics in a population to exceed standards. A study designed to achieve objectives 3 and 4 will be presented. Modifications and alternate approaches will also be discussed.

Background Navigable waterways of the United States have and will continue to play a vital role in the nation's development. The Corps, in fulfilling its mission to maintain, improve, and extend these waterways, is responsible for the dredging and disposal of large volumes of sediment each year. Nationwide, the Corps dredges about 230 million cubic yards in maintenance work and about 70 million cubic yards in new dredging annually at a cost of about $450 million. In accomplishing its national dredging and regulatory missions, the Corps has conducted extensive research and development in the field of dredged material management. Federal expenditures on dredged material research, monitoring, and management activities have cumulatively exceeded $100 million. Techniques developed to evaluate contaminant mobility in dredged material can be applied to other contaminated areas. Accordingly, the plant and animal bioassays are two techniques developed to assess the environmental impact of dredged material in wetland and upland disposal environments. These bioassays, surface soil samples, groundwater samples, and additional plant tissues were used to evaluate a contaminated site in western California.

The case study site is approximately 200 acres with both upland and wetland areas (see Figure 1). This site was known to have very high concentrations of metals in surface soils. Major pathways for contaminant mobility are the meandering stream which flows north and the drainage ditches. Also, tidal inundation affects a substantial portion of this site.

The objectives of the study were to: (1) define the extent of the hazardous substance contamination on the site; (2) identify the sources of the hazardous substances detected on the property; (3) evaluate the extent of migration of the hazardous substances on the property; (4) assess the bioavailability, mobility, and toxicity of the hazardous substances detected on the property; (5) evaluate the condition of the wetland and upland habitats on the property. This paper focuses on the use of soil samples to achieve objectives 1 thru 4.

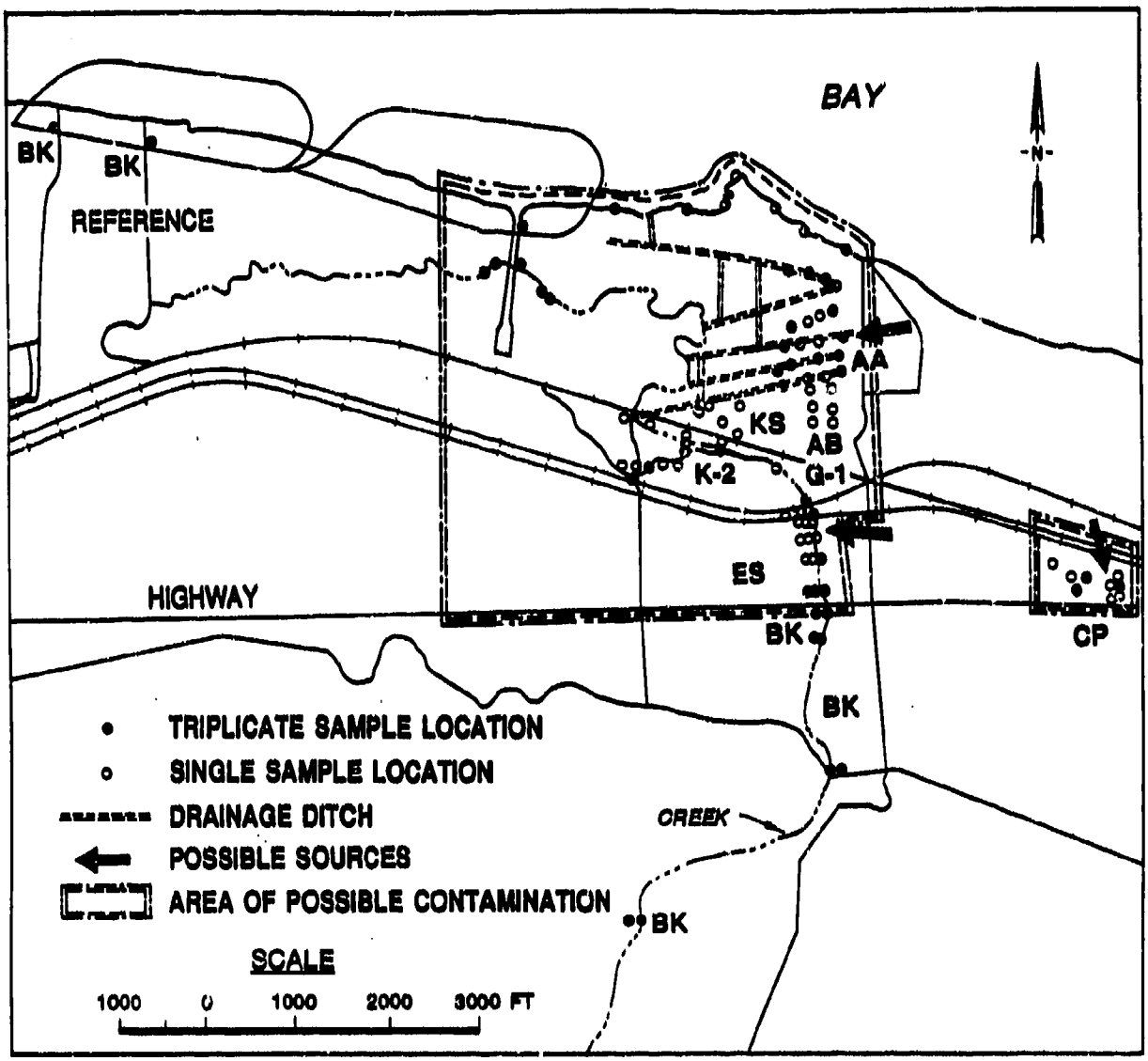


Figure 1. Map of Study Area

SAMPLING PLAN The sampling plan was formulated based on previous soil and water data, historical information, and the potential pathways for contaminant mobility. The sampling locations are shown in Figure 1. Three samples were collected at some locations and one sample was collected at the other locations. The triplicate samples were used in statistical comparisons. This sampling plan reduced the cost of the investigation by allowing a selected number of sample locations to be tested extensively while other sample locations received one-third the cost and effort. A total of 178 samples were collected and analyzed for As, Cd, Cu, Pb, Ni, Se, and Zn.

There is an analogy between the strategy used here and the disposal philosophy of many Corps elements. Most dredging and disposal decisions are made at the local level on a case by case basis. Often, the environmental objective is to prevent further degradation of the disposal area. Therefore, samples are collected at the dredge site and disposal site. A statistical evaluation performed on the chemical analysis of the samples becomes the basis for determining whether degradation will occur. In this study, samples were collected at the remote reference area and an area of degradation (i.e. contamination). Ten triplicate samples were collected in the remote reference area. Twenty-eight triplicate samples were collected in the area of contamination. Locations having a mean concentration of metals in soil, plants, or animals statistically greater than similar data from all remote reference locations were declared contaminated. These concentrations provide a judgemental basis for classifying the 64 single sample locations.

Three sources of contamination were identified from historical information. One additional source was indicated by the soil analysis and later verified with historical information. Sources were thought to be areas with several high metal concentrations in a vicinity and a gradual decrease in metal concentrations as one moves away from this area. The sources found in this study appeared to have released metal in two different forms. One method was to bury or discharge contaminants associated with a solid material in an area. The other source discharged highly contaminated liquids into a stream. Identifying sources was further complicated by the fact that some of the discharges were intermittent and possibly hadn't occurred in several years. This study was successful in identifying sources which discharged contaminants associated with solids. Identifying the source of liquid discharges was more difficult due to seasonal fluctuation of the stream.

The soil analysis was partially successful in achieving objectives 1 thru 4. The extent of contamination from known sources was established and locations requiring further investigation were identified. This plan has been augumented with additional sampling. These samples further delineated the extent of contamination horizontally across the site and vertically down the soil profile. As a result of this study, 26.5 acres were declared contaminated.

A Generalized Gumbel Distribution

Siegfried H. Lehnigk
Research Directorate
Research, Development, and Engineering Center
U.S. Army Missile Command
Redstone Arsenal, AL 35898-5248

A generalized Gumbel (extreme value type I) distribution class is introduced. In addition to the usual shift and scale parameters this new distribution contains an arbitrary positive shape parameter. The classical Gumbel distribution results as special case for shape equal to unity. Microcomputer-based algorithms for estimation of the parameters are presented. They are based on the moment equations and on the logarithmic likelihood function associated with the distribution density. A program diskette for microcomputer use will be made available upon request. A combined paper by this author and Charles E. Hall, Jr., will be published elsewhere.

A Generalization of the Eulerian Numbers with a Probabilistic Application

Bernard Harris

University of Wisconsin, Madison

C. J. Park

San Diego State University

1 Introduction and Historical Summary

In this paper we study a generalization of the Eulerian numbers and a class of polynomials related to them. An interesting application to probability theory is given in Section 3. There we use these extended Eulerian numbers to construct an uncountably infinite family of lattice random variables whose first n moments concide with the first n moments of the sum of $n+1$ uniform random variables. A number of combinatorial identities are also deduced.

The Eulerian numbers are defined by

$$A_{nj} = \sum_{v=0}^{j}(-1)^v \binom{n+1}{v} (j-v)^n, j = 0,1,2,\ldots,n; \quad n = 0,1,2,\ldots. \tag{1}$$

They satisfy the recursion

$$A_{nj} = jA_{n-1,j} + (n-j+1)A_{n-1,j-1} \tag{2}$$

and the Worpitzky [25] relation

$$x^n = \sum_{j=1}^{n} \binom{x+j-1}{n} A_{nj}. \tag{3}$$

Also,

$$A_{nj} = A_{n,n-j+1}; \tag{4}$$

$$\sum_{j=1}^{n} A_{nj} = n!. \tag{5}$$

In addition, they possess a number of combinatorial interpretations which are described below.

Let $X_n = \{1,2,\ldots,n\}$ and let $P_n(k)$ be the number of permutations of the elements of X_n having exactly k increases between adjacent elements, the first element always being counted as an increase.

For $n = 4$, the 24 permutations and the number of increases are given in Table 1.1.

Table 1.1

	Permutation	Number of Increases
1	1234	4
2	1243	3
3	1324	3
4	1342	3
5	1423	3
6	1432	2
7	2134	3
8	2143	2
9	2314	3
10	2341	3
11	2413	3
12	2431	2
13	3124	3
14	3142	2
15	3214	2
16	3241	2
17	3412	3
18	3421	2
19	4123	3
20	4132	2
21	4213	2
22	4231	2
23	4312	2
24	4321	1

As seen from the tabulation, $P_4(1) = 1, P_4(2) = 11, P_4(3) = 11, P_4(4) = 1$, which concides with $A_{4j}, j = 1, 2, 3, 4$.

Let

$$A_n(t) = \sum_{j=1}^{n} A_{nj} t^j. \tag{6}$$

Then

$$\frac{1-t}{1-t\exp[x(1-t)]} = \sum_{n=0}^{\infty} A_n(t) x^n/n!, t \neq 1. \tag{7}$$

The above relations and some of their properties can be found in [8]; the polynomials (6) are also discussed in L. Carlitz [4]. These results may also be found in the expository paper of L. Carlitz [3]. The formulas (1) and (2) are also given in L. v. Schrutka [21].

Désiré André [1] established that A_{nj} is the number of permutations of $\{X_n\}$ with j "elementary inversions". He also established that A_{nj} is the number of circular permutations of $\{X_{n+1}\}$ with j "elementary inversions". The equivalence of these two results with the enumeration of the number of increases in permutations of $\{X_n\}$ can be trivially established.

G. Frobenius [15] studied the polyomonials

$$A_n(x) = \sum_{j=1}^{n} A_{nj} x^j, \tag{8}$$

introduced by Euler, and established many of their properties. In particular, relations with the Bernoulli numbers are given in [15].

In D. P. Roselle [20], the enumeration of permutations by the number of rises, A_{nj}, is related to enumeration by the number of successions, that is, a permutation π of $\{X_n\}$ has a succession if $\pi(i) = i + 1, i = 1, 2, \ldots, n$.

Some number theoretic properties of A_{nj} are given in L. Carlitz and J. Riordan [7] and in L. Carlitz [5].

In this paper, we study a generalization of the Eulerian numbers. A generalization in a different direction was given by E. B. Shanks [22], who apparently did not note a connection of his coefficients with the Eulerian numbers. L. Carlitz [2] noted the relationship of Shank's results to

the Eulerian numbers and obtained representations for these generalized Eulerian numbers using results due to N. Nielsen [17].

F. Poussin [18] considered the enumeration of the number of inversions of permutations of $\{X_n\}$ which end in $j, 1 \leq j \leq n$. This produces a decomposition of the Eulerian numbers. She also introduced a polyomial generating function for these numbers. The sums of these polynomials are the Euler-Frobenius polynomials.

Another deomposition of the Eulerian numbers with a combinatorial interpretation is given by J.F. Dillon and D.P. Roselle [12].

J. Riordan [19] lists many properties of the Eulerian numbers in Exercise 2, page 38-39 and describes the combinatorial interpretation of the Eulerian numbers in terms of triangular permutations (which is equivalent to the elementary inversions described by André [1]). He also gives a brief table of the Eulerian numbers on page 215. See also L. Comtet [10], where generating functions for the Eulerian numbers are given and the Eulerian numbers are obtained by enumerating the number of permutations with a specified number of increases. Many properties of the Eulerian numbers are given as well as their historical origins in terms of sums of powers.

F.N. David and D.E. Barton [11] suggest the use of the Eulerian numbers as a statistical test for the randomness of a sequence of observations in time, employing the probability distribution given by

$$P_{nj} = A_{nj}/n!, \; j = 1, 2, \ldots, n. \tag{9}$$

The generating function (7) is derived and employed to obtain the moments and cumulants of the distribution (9). In particular, David and Barton show that the factorial moments are the generalized Bernoulli numbers. However, David and Barton do not make any identification of these distributions with the Eulerian numbers.

Using probabilistic arguments, Carlitz, Kurtz, Scoville and Stackelberg [6] showed that the Eulerian numbers, when suitably normalized, have an asymptotically standard normal distribution. This was accomplished by representing the distribution P_{nj} as the distribution of a sum of indepen-

dent Bernoulli random variables. S. Tanny [24] demonstrated the asymptotic normality by utilizing the relationship of the Eulerian numbers to the distribution of the sum of independent uniform random variables and applying the central limit theorem.

L. Takács [23] obtained a generalization of the Eulerian numbers which provide the solution to a specific occupancy problem. Namely, let a sequence of labelled boxes be given, the first box labelled 1, the second box 2, and so on. At trial number n distribute l balls randomly in the first n boxes so that the probability that each ball selects a specific box is $1/n$ and the selections are stochastically independent. For $l = 1$, the probability that $j - 1$ boxes are empty after trial number n is $A_{nj}/n!, j = 1, 2, \ldots, n$. Takács' paper contains many references and describes additional combinatorial problems whose solution is related to the Eulerian numbers.

Finally, L. Toscano [25] obtained formulas expressing the Eulerian numbers in terms of Stirling numbers of the second kind.

2 Generalized Eulerian Numbers

We now introduce a generalization of the Eulerian numbers and investigate its properties.

Let δ be an arbritrary real number and let

$$A_{nj}(\delta) = \sum_{v=0}^{j} \binom{n+1}{v} (-1)^v (\delta + j - v)^n, j = 0, 1, \ldots, n; n = 0, 1, 2, \ldots. \tag{10}$$

These polynomials are mentioned in L. Carlitz, D.P. Roselle and R.A. Scoville [8]. As noted there, $A_{nj}(0)$ are the Eulerian numbers. These polynomials are also used by P.S. Dwyer [13] to calculate sample factorial moments. Dwyer does not relate these to the Eulerian numbers.

We begin our analysis with the following theorem:

Theorem 1. Let n and k be non-negative integers and let δ be any real number. Then

$$\mu_{[k]}^{(n)}(\delta) = \frac{1}{n!}\sum_{j=0}^{n}(\delta+j)^{(k)}\sum_{v=0}^{j}\binom{n+1}{v}(-1)^v(\delta+j-v)^n$$

$$= \frac{1}{n!}\sum_{j=0}^{n}(\delta+j)^{(k)}A_{nj}(\delta) \tag{11}$$

is independent of δ for $k = 0, 1, \ldots, n$.

Proof. The following identity (see N. Nielsen, [17], page 28) will be utilized in the proof.

$$(x-\delta)^n = \sum_{j=0}^{n}\binom{x+j-1}{n}\sum_{v=0}^{j}(-1)^v\binom{n+1}{v}(\delta+j-v)^n. \tag{12}$$

Let Δ and E be the operators defined by

$$\Delta(f(x)) = f(x+1) - f(x)$$

and

$$E(f(x)) = f(x+1).$$

Then, it can be shown that (C. Jordan, [16])

$$\Delta^r = \sum_{v=0}^{r}(-1)^v\binom{r}{v}E^{r-v}. \tag{13}$$

In particular, for $r = 0, 1, \ldots, n$,

$$\Delta^r(\delta+j)^{(n)} = \sum_{l=0}^{r}(-1)^l\binom{r}{l}(\delta+j+r-l)^{(n)}$$

$$= n^{(r)}(\delta+j)^{(n-r)}, \tag{14}$$

the last equality follows from elementary properties of the function $(\delta+j)^{(n)}$ (C. Jordan, [16], p. 51). Thus, for $r = 0, 1, \ldots, n$, from (12) we have

$$(\delta+j)^{(n-r)} = \sum_{l=0}^{r}(-1)^l\binom{r}{l}(\delta+j+r-l)^{(n)}/n^{(r)}. \tag{15}$$

Hence,

$$\mu_{[k]}^{(n)}(\delta) = \frac{1}{n!}\sum_{j=0}^{n}(\delta+j)^{(k)}\sum_{v=0}^{j}\binom{n+1}{v}(-1)^v(\delta+j-v)^n$$

$$= \frac{1}{n!}\sum_{j=0}^{n}\sum_{l=0}^{n-k}(-1)^l\binom{n-k}{l}(\delta+j+n-k-l)^{(n)}\sum_{v=0}^{j}\binom{n+1}{v}(-1)^v(\delta+j-v)^n/n^{(n-k)}. \tag{16}$$

Thus, it follows that

$$\mu_{[k]}^{(n)}(\delta) = \sum_{l=0}^{n-k}(-1)^l\binom{n-k}{l}\sum_{j=0}^{n}\binom{\delta+j+n-k-l}{n}\sum_{v=0}^{j}(-1)^v\binom{n+1}{v}(\delta+j-v)^n/n^{(n-k)}. \tag{17}$$

Setting $x = \delta + n - k - l + 1$ in (12) we get

$$(n-k-l+1)^n = \sum_{j=0}^{n}\binom{\delta+n-k-l+j}{n}\sum_{v=0}^{j}(-1)^v\binom{n+1}{v}(\delta+j-v)^n$$

and hence

$$\mu_{[k]}^{(n)}(\delta) = \sum_{l=0}^{n-k}(-1)^l\binom{n-k}{l}(n-k-l+1)^n/n^{(n-k)} \tag{18}$$

and is independent of δ.

In particular, $\mu_{[n]}^{(n)}(\delta) = 1$, $\mu_{[n-1]}^{(n)}(\delta) = (2^n-1)/n$ and $\mu_{[n-2]}^{(n)}(\delta)$ $= (3^n - 2^{n+1}+1)/n(n-1)$, $\mu_{[n-3]}^{(n)}(\delta) = \frac{4^n-3^{n+1}+3(2^n)-1^n}{n(n-1)(n-2)}$.

A brief table of $\mu_{[k]}^{(n)}(\delta)$ for $k = 0,1,2,3$ and $n = 0,1,2,3$ is given in the Appendix to this paper.

The Nielsen identity (12) seems to have been discovered in a somewhat less general context by Paul S. Dwyer [13], who employed it to calculate factorial moments by means of cumulative sums; see also Ch. A. Charalambides [9], who in addition to discussing Dwyer's work also showed that these generalized Eulerian are related to enumeration of compositions of integers.

The following corollary will be subsequently employed.

Corollary. Let n and k be non-negative integers with $k \leq n$. Then

$$\mu_k^{(n)}(\delta) = \frac{1}{n!}\sum_{j=0}^{n}(\delta+j)^k \sum_{v=0}^{j}\binom{n+1}{v}(-1)^v(\delta+j-v)^n \tag{19}$$

is independent of δ.

Proof. We can write

$$(\delta+j)^k = \sum_{r=0}^{k}\beta_{rk}(\delta+j)^{(k)}, \tag{20}$$

where β_{rk} are the Stirling numbers of the second kind. Since the coefficients β_{rk} do not depend on δ, substituting (20) into (19) and interchanging the order of summation, we get

$$\mu_k^{(n)}(\delta) = \sum_{r=0}^{k}\beta_{rk}\mu_{[k]}{}^{(n)}(\delta), \tag{21}$$

which is independent of δ.

Prior to demonstrating that the independence of δ, noted in Theorem 1 and its corollary can not be extended to $k = n+1$, we will need to calculate the derivative of $\mu_k^{(n)}(\delta)$. Thus, we have:

Theorem 2. Let n and k be non-negative integers and let δ be any real number. Then

$$\frac{d}{d\delta}\mu_k^{(n)}(\delta) = k\mu_{k-1}^{(n)}(\delta) - \sum_{r=0}^{k-1}\binom{k}{r}\mu_r^{(n-1)}(\delta). \tag{22}$$

Proof. Since

$$\mu_k^{(n)}(\delta) = \frac{1}{n!}\sum_{j=0}^{n}(\delta+j)^k \sum_{v=0}^{j}\binom{n+1}{v}(-1)^v(\delta+j-v)^n,$$

$$\frac{d\mu_k^{(n)}(\delta)}{d\delta} = \frac{1}{n!}\sum_{j=0}^{n}k(\delta+j)^{k-1}\sum_{v=0}^{j}\binom{n+1}{v}(-1)^v(\delta+j-v)^n$$

$$+\frac{1}{(n-1)!}\sum_{j=0}^{n}(\delta+j)^k\sum_{v=0}^{j}\binom{n+1}{v}(-1)^v(\delta+j-v)^{n-1}.$$

Comparing the first term with (19) and employing the Pascal triangle identity on the second term we get

$$\frac{d\mu_k^{(n)}(\delta)}{d\delta} = k\mu_{k-1}^{(n)}(\delta) + \frac{1}{(n-1)!}\sum_{j=0}^{n}(\delta+j)^k \sum_{v=0}^{j}\left[\binom{n}{v} + \binom{n}{v-1}\right](-1)^v(\delta+j-v)^{n-1}. \tag{23}$$

Further,

$$\frac{1}{(n-1)!}\sum_{j=0}^{n}(\delta+j)^k\sum_{v=0}^{j}\binom{n}{v}(-1)^v(\delta+j-v)^{n-1}$$

$$= \frac{1}{(n-1)!}(\delta+n)^k\sum_{v=0}^{n}\binom{n}{v}(-1)^v(\delta+n-v)^{n-1}$$

$$+ \frac{1}{(n-1)!}\sum_{j=0}^{n-1}(\delta+j)^k\sum_{v=0}^{j}\binom{n}{v}(-1)^v(\delta+j-v)^{n-1}. \tag{24}$$

From (13)

$$\Delta^n(\delta^{n-1}) = \sum_{v=0}^{n}(-1)^v\binom{n}{v}(\delta+n-v)^{n-1} = 0, \tag{25}$$

since it is the nth difference of a polynomial of degree $n-1$. The second term on the right hand side of (24) is $\mu_k^{(n-1)}(\delta)$.

In addition,

$$\frac{1}{(n-1)!}\sum_{j=0}^{n}(\delta+j)^k\sum_{v=0}^{j}\binom{n}{v-1}(-1)^v(\delta+j-v)^{n-1}$$

$$= \frac{1}{(n-1)!}\sum_{j=0}^{n}(\delta+j)^k\sum_{v=0}^{j-1}\binom{n}{v}(-1)^{v+1}(\delta+j-v-1)^{n-1}$$

$$= \frac{1}{(n-1)!}\sum_{j=1}^{n}(\delta+j)^k\sum_{v=0}^{j-1}\binom{n}{v}(-1)^{v+1}(\delta+j-v-1)^{n-1}$$

$$= \frac{1}{(n-1)!}\sum_{j=1}^{n}\Big[\,(\delta-1+j)+1\,\Big]^{k}\sum_{v=0}^{j-1}\binom{n}{v}(-1)^{v+1}(\delta+j-v-1)^{n-1}$$

$$= \frac{1}{(n-1)!}\sum_{r=0}^{k}\binom{k}{r}\sum_{j=1}^{n}(\delta-1+j)^{r}\sum_{v=0}^{j-1}\binom{n}{v}(-1)^{v+1}((\delta-1+j)-v)^{n-1}$$

$$= \frac{1}{(n-1)!}\sum_{r=0}^{k}\binom{k}{r}\sum_{j=0}^{n-1}(\delta+j)^{r}\sum_{v=0}^{j}\binom{n}{v}(-1)^{v+1}(\delta+j-v)^{n-1}$$

$$= -\sum_{r=0}^{k}\binom{k}{r}\mu_{r}^{(n-1)}(\delta). \tag{26}$$

Thus, by (23), (25) and (26) we have shown that

$$\frac{d\mu_{k}^{(n)}(\delta)}{d\delta} = k\mu_{k-1}^{(n)}(\delta) + \mu_{k}^{(n-1)}(\delta) - \sum_{r=0}^{k}\binom{k}{r}\mu_{r}^{(n-1)}(\delta),$$

or

$$\frac{d\mu_{k}^{(n)}(\delta)}{d\delta} = k\mu_{k-1}^{(n)}(\delta) - \sum_{r=0}^{k-1}\binom{k}{r}\mu_{r}^{(n-1)}(\delta), \tag{27}$$

establishing the theorem.

Corollary 1. For $1 \le k \le n$, then

$$k\mu_{k-1}^{(n)}(\delta) = \sum_{r=0}^{k-1}\binom{k}{r}\mu_{r}^{(n-1)}(\delta). \tag{28}$$

Proof. By the Corollary to Theorem 1, $\mu_{k}^{(n)}(\delta)$ is independent of δ for $0 \le k \le n$ and hence $\frac{d\mu_{k}^{(n)}(\delta)}{d\delta} = 0$ for such values of k.

Corollary 2. If $k = n+1$, then

$$\frac{d\mu_{n+1}^{(n)}(\delta)}{d\delta} = c_{n+1,n} - (n+1)\mu_{n}^{(n-1)}(\delta), \tag{29}$$

where $c_{n+1,n}$ is a constant (depending on n, but independent of δ).

Proof. From the Corollary to Theorem 1 and Theorem 2, $(n+1)\mu_n^{(n)}(\delta)$ is independent of δ and all terms in $\sum_{r=0}^{n}\binom{n+1}{r}\mu_r^{(n-1)}(\delta)$ with the exception of $r = n$ are independent of δ. This last term is $(n+1)\mu_n^{(n-1)}(\delta)$.

Corollary 2 can be extended to $k = n+2$ and so forth, but the expressions obtained become more complicated and do not appear to be particularly useful. However, we do make use of Corollary 2 in the next theorem.

Theorem 3. For every n, $\mu_{n+1}^{(n)}(\delta)$ is a polynomial of degree $n+1$ in δ with leading coefficient $(-1)^n$.

Proof. We proceed by induction, using Corollary 2 to Theorem 2.

$$\text{For } n = 0, \mu_1^{(0)}(\delta) = \delta. \text{ Then } \frac{d\mu_2^{(1)}(\delta)}{d\delta} = c_{2,1} - 2\mu_1^{(0)}(\delta) = c_{2,1} - 2\delta.$$

Performing the indicated integration, we have

$$\mu_2^{(1)}(\delta) = c_{2,1}\delta - \delta^2 + d,$$

where d is an unspecified constant.

Assume that the conclusion holds for $n = m$. Then

$$\frac{d\mu_{m+2}^{(m+1)}(\delta)}{d\delta} = c_{m+2,m+1} - (m+2)\mu_{m+1}^{(m)}$$

$$= c_{m+2,m+1} - (m+2)(a_0\delta^{m+1} + \sum_{j=1}^{m+1} a_j\delta^{m+1-j}), \tag{30}$$

where a_0 is $+1$ or -1. Integrating, we get $\mu_{m+2}^{(m+1)}(\delta) = -a_0\delta^{m+2} + P_{m+1}(\delta)$, where $P_{m+1}(\delta)$ is a polynomial of degree $m+1$.

A table of $\mu_n^{(k)}(\delta)$ appears in the Appendix for $n = 5, k = 0, 1, \ldots, 10$ and selected values of δ.

3 Applications to Probability Theory

Let $U_1, U_2, \ldots, U_{n+1}$ be independent random variables uniformly distributed on $(0,1)$. Let $S_{n+1} = \sum_{i=1}^{n+1} U_i$. The distribution of S_{n+1} is well-known and is given by the probability density function

$$f_{S_{n+1}}(x) = \frac{1}{n!}\sum_{\nu=0}^{n+1}\binom{n+1}{\nu}(-1)^{\nu}(x-\nu)_+^n, 0 < x < n+1, \tag{31}$$

(for example, see W. Feller [14].) Where

$$(x-a)_+ = \begin{cases} 0, & x-a<0 \\ x-a, & x-a \geq 0. \end{cases} \tag{32}$$

Write

$$S_{n+1} = [S_{n+1}] + \delta,$$

where $[S_{n+1}]$ denotes the integer part of S_{n+1}. Clearly δ is a continuous random variable and $0 \leq \delta < 1$; $[S_{n+1}]$ is a discrete random variable with carrier set $\{0,1,2,\ldots,n\}$.

The conditional distribution of S_{n+1} given that the fractional part of S_{n+1} is δ is given by

$$P\{S_{n+1} = x | x - [x] = \delta\} = f_{S_{n+1}}(j+\delta) / \sum_{j=0}^{n} f_{S_{n+1}}(j+\delta), \tag{33}$$

where $j + \delta = x, j = 0,1,\ldots,n$; i.e. $j = [x]$.

From (31),

$$f_{S_{n+1}}(j+\delta) = \frac{1}{n!}\sum_{\nu=0}^{n+1}\binom{n+1}{\nu}(-1)^{\nu}(j+\delta-\nu)_+^n.$$

But $j + \delta - \nu \geq 0$ is equivalent to $\nu \leq j$, thus we get

$$f_{S_{n+1}}(j+\delta) = \frac{1}{n!}\sum_{\nu=0}^{j}\binom{n+1}{\nu}(-1)^{\nu}(j+\delta-\nu)^n, \tag{34}$$

which is $A_{nj}(\delta)/n!$. Also,

$$\sum_{j=0}^{n} f_{S_{n+1}}(j+\delta) = 1 = \mu_{[0]}{}^{(n)}(\delta)$$

and thus (34) is a discrete probability distribution with carrier set $\{\delta, 1+\delta, \ldots, n+\delta\}$.

Let $W_{n+1,\delta}$ be the random variable whose distribution is given by (34). We then have the following theorem.

Theorem 4. The moments of order $k = 0, 1, \ldots, n$ of $W_{n+1,\delta}$ coincide with the corresponding moments of S_{n+1}, that is,

$$E\{S_{n+1}^k\} = E\{W_{n+1,\delta}^k\}, k = 0, 1, \ldots, n;\ 0 \leq \delta < 1. \tag{35}$$

Proof. $E\{S_{n+1}^k\} = E_\delta(E\{S_{n+1}^k|\delta\}) = E_\delta\{E\{W_{n+1,\delta}^k\}\}$.

However,

$$E\{W_{n+1,\delta}^k\} = \frac{1}{n!}\sum_{j=0}^{n}(j+\delta)^k \sum_{\upsilon=0}^{j}\binom{n+1}{\upsilon}(-1)^\upsilon(j+\delta-\upsilon)^n,$$

which is independent of δ, by the Corollary to Theorem 1.

A brief table of $W_{n+1,\delta}$ for $n = 5$ is given in the Appendix.

Remark. It is easy to see that the marginal distribution of δ, the fractional part of S_{n+1}, is uniform on $(0, 1)$. An elementary proof follows.

$$P\{\delta \leq \delta^*\} = \sum_{j=0}^{n}\int_j^{j+\delta^*} f_{S_{n+1}}(x)\,dx$$

$$= \sum_{j=0}^{n}\int_0^{\delta^*} f_{S_{n+1}}(j+u)\,du$$

$$= \int_0^{\delta^*}\sum_{j=0}^{n}(f_{S_{n+1}}(j+u))\,du;$$

but $\sum_{j=0}^{n} f_{S_{n+1}}(j+u) = 1$ for every $0 \leq u < 1$. Hence

$$P\{\delta \leq \delta^*\} = \int_0^{\delta^*} du = \delta^*.$$

Finally, we note that $W_{n+1,\delta}$ is asymptotically normally distributed. This is stated in the following theorem.

Theorem 5. As $n \to \infty$, for $0 \leq \delta < 1$, the distribution of

$$\sqrt{\frac{12}{(n+1)}}\left(W_{n+1,\delta} - \left(\frac{n+1}{2} \right) \right) \tag{36}$$

coverges weakly to the standard normal distribution.

Further,

$$\frac{A_{nj}(\delta)}{n!} = \sqrt{\frac{6}{\pi n}} e^{-\frac{6}{n}(j+\delta-(n+1)/2)^2} (1 + O(n^{-\frac{1}{2}})). \tag{37}$$

Proof. Both (36) and (37) are immediate consequences of the representation of $\frac{A_{nj}(\delta)}{n!}$ as the conditional distribution of the sum of $n+1$ independent uniform random variables on $[0, 1)$ given the fractional part of the sum and the central limit theorem.

Appendix

This Appendix is devoted to some tables illustrative of some of the quantities introduced in the body of the paper.

Table A.1

Table of $\mu_{[k]}^{(n)}(\delta), k = 0, 1, 2, 3; n = 0, 1, 2, 3$

$\frac{k}{n}$	0	1	2	3
0	1	δ	$\delta(\delta-1)$	$\delta(\delta-1)(\delta-2)$
1	1	$\frac{n+1}{2}$	$1+\delta-\delta^2-(\frac{n+1}{2})$	$2\delta^3+3\delta^2-\delta-1+n$
2	1	$\frac{n+1}{2}$	$\frac{(n+1)(3n-2)}{12}$	$1+2\delta-2\delta^3-\frac{(n+1)3n}{4}$
3	1	$\frac{n+1}{2}$	$\frac{(n+1)(3n-2}{12}$	$\frac{(n+1)(n-1)(n-2)}{24}$

Table A.2

The Distribution of $W_{n+1,\delta}$, $n = 5$, $\delta = .1, .4, .5, .9$

W	$\delta = .1$	$\delta = .4$	$\delta = .5$	$\delta = .9$
δ	8×10^{-8}	9×10^{-5}	3×10^{-4}	.005
$1 + \delta$	.013	.044	.062	.177
$2 + \delta$	.260	.396	.438	.545
$3 + \delta$	.545	.476	.438	.260
$4 + \delta$	.177	.083	.062	.013
$4 + \delta$	.005	6×10^{-4}	3×15^{-4}	5×10^{-8}

Note the symmetry for $\delta = .5$ and that $\delta = .9$ and $\delta = .1$ are identical when the column for $\delta = .9$ is read going up and $\delta = .1$ is read going down (the entries 8×10^{-8} and 5×10^{-8} differ as a consequence of rounding errors).

Table A.3

$\mu_n^k(\delta), n = 5, k = 0, 1, \ldots, 10; \delta = 0, .1, .3, .5, .7, .9$

	$\delta = 0$	.1	.3	.5	.7	.9
$k = 0$	1	1	1	1	1	1
1	3	3	3	3	3	3
2	9.5	9.5	9.5	9.5	9.5	9.5
3	31.5	31.5	31.5	31.5	31.5	31.5
4	108.7	108.7	108.7	108.7	108.7	108.7
5	388.5	388.5	388.5	388.5	388.5	388.5
6	1432.50	1432.50	1432.53	1432.55	1432.53	1432.50
7	5431.50	5431.51	5432.01	5432.48	5432.31	5431.69
8	2118.7	21117.60	21122.56	21129.77	21129.66	21122.07
9	84010.5	83989.19	84020.48	84096.88	84116.67	84049.80
10	341270.5	341018.48	341121.81	341763.40	342089.16	341628.77

References

[1] André, D., Mémoire sur les inversions élémentaires des permutations, Mem. della Pontifica Accad. Romana dei Nuovo Lincei, 24 (1906), 189-223.

[2] Carlitz, L., Note on a paper of Shanks, Am. Math. Montly, 59 (1952), 239-241.

[3] Carlitz, L., Eulerian numbers and polynomials, Math. Mag. 32 (1959), 247-260.

[4] Carlitz, L., Eulerian numbers and polynomials of higher order, Duke Math. J. 27 (1960), 401-424.

[5] Carlitz, L., A note on the Eulerian numbers, Arch. Math., 14 (1963), 383-390.

[6] Carlitz, L., Kurtz, D.C., Scoville, R., and Stackelberg, O.P., Asymptotic Properties of Eulerian Numbers, Z. Wahrscheinlichkeitstheorie verw. Geb., 23 (1972), 47-54.

[7] Carlitz, L. and Riordan, J., Congruences for Eulerian numbers, Duke Math. J. 20 (1953) 339-344.

[8] Carlitz, L., Roselle, D.P., and Scoville, R.A., Permutations and sequences with repetitions by number of increases, J. Combin. Theory, 1 (1966), 350-374.

[9] Charalambides, Ch.A., On the enumeration of certain compositions and related sequences of numbers, Fibonnacci Quarterly, 20 (1982), 132-146.

[10] Comtet. L., Analyse Combinatoire, Presses Universitaires de France, Paris, 1970.

[11] David, F.N. and Barton, D.E., Combinatorial Chance, Charles Griffin and Company Ltd., London, 1962.

[12] Dillon, J.F. and Roselle, D.P., Eulerian numbers of higher order, Duke Math. J., 35 (1968), 247-256.

[13] Dwyer, P.S., The calculation of moments with the use of cumulative totals, Ann. Math. Statist., 9 (1938), 288-304.

[14] Feller, W., An Introduction to Probability Theory and its Applications, 2nd Ed., Vol. II, John Wiley & Sons, Inc., New York, 1971.

[15] Frobenius, G., Über die Bernoullischen Zahlen und die Eulerschen Polynome, Sitz. Ber. Preuss. Akad. Wiss. (1910), 809-847.

[16] Jordan, K., Calculus of Finite Differences, 2nd Ed., Chelsea Publishing Co., New York (1960).

[17] Nielsen, N. Traité élémentaire des nombres de Bernoulli, Gauthier-Villars et Cie., Paris (1923).

[18] Poussin, F., Sur une propriété arithmétique de certains polynômes associés aux nombres d'Euler, Comptes Rendus Acad. Sci. Paris, 266 (1968), 392-393.

[19] Riordan, J., An Introduction to Combinatorial Analysis, John Wiley & Sons, Inc., New York, 1958.

[20] Roselle, D.P., Permutations by number of rises and successions, Proc. Amer. Math. Soc., 19 (1968), 8-16.

[21] v. Schrutka, Lothar, Eine nueue Einteilung der Permutationen, Math. Annalen., 118 (1941), 246-250.

[22] Shanks, E.B., Iterated sums of powers of the binomial coefficients, Am. Math. Monthly, 58 (1951), 404-407.

[23] Takács, L., A generalization of the Eulerian numbers, Publicationes Mathematicae Debrecen, 26 (1979), 173-181.

[24] Tanny, S., A probabilistic interpretation of Eulerian numbers, Duke Math. J., 40 (1973), 717-722.

[25] Toscano, L., Sulla somma di alcune serie numeriche, Tohoku Math. J., 38 (1933), 332-342.

[26] Worpitzky, J., Studien über die Bernoullischen und Eulerschen Zahlen, Journal für die reine und angewandte Mathematik, 94 (1883), 203-232.

The Analysis of Multivariate Qualitative Data Using an Ordered Categorial Approach

H. B. Tingey
University of Delaware

E. A. Morgenthein
Bristol-Meyers

S. M. Free

ABSTRACT

When the experimental units being classified are sub-sampling units in the study, an ordered categorical procedure cannot be applied directly. Further, the count data obtained which is routinely analyzed by univariate statistical methods, ignores the dependence among the responses. A modification of the method developed by Nair (1986, 1987) is used to derive the scores and indices, which are analyzed by nonparametric AOV. An example from teratogenicity studies is used to illustrate the technique.

Introduction

This problem arises from the consideration of studies where a reproduction safety test must be performed prior to the use of drug, chemical or food additive. The standard protocol in such studies requires that pregnant female subjects (usually rodents) are randomly assigned to one of four treatment groups. The appropriate dosage is administered shortly after the beginning of gestation. When the animals are near term, they are sacrificed and the number of potential offspring are counted. Other data collected are the number of implantation, early and late fetal deaths, number of live offspring and the number of fetuses according to various degrees of increasing severity of malformation. Also data on continuous variables such as fetal weight are collected. It is unclear from the literature which statistical methods are appropriate for the analysis of this type of data.

For continuous measurements one may quickly turn to the analysis of variance. For count data describing the number of fetuses with or without some qualitative outcome, other methods have evolved. A per-fetus analysis using total of early death and total number of implantation in a Fisher exact-test or a chi- squared test of independence may be performed, but this appears to inflate samples sizes and ignores the dependence of observations within litters. A review of per-fetus analysis is given by Haseman and Hogan (1975) who conclude the per letter analysis is more appropriate.

All but one of the proposed methods for per-litter analysis consider a single outcome. The need to include within and among-litter variation negates the use of simple binomial or Poisson models for count data. In the methods which consider several single responses, a problem of family error rate arises. Since the tests are not independent, the nominal family error rate cannot be exactly determined. The multivariate method developed by

Ryttman (1976) relies on the assumption of normality which is violated in the case of fetal deaths. This lack of success, however, does not preclude a multivariate approach. In situations where ranking the categories from mild to severe is possible, ordered categorical models may be applied and the family error problem may be eliminated.

In this paper we obtain a scoring system for various outcomes which produces a severity index for each litter. This index is sensitive to location shifts. The modeling which follows will be based on this index.

The study design prohibits the straight-forward application of ordered categorical procedures because the items (fetuses) are not independent. Thus a scoring procedure allows consideration of the effect of letter size on severity of the response, as a whole, in the litter. Here the sampling unit is the fetus or individual. Three observations should be made; i) results are different per litter than for per fetus, ii) per litter evaluates the proportion of fetuses affected rather than the numbers of affected litters, and iii) observed treatment - control differences is less significant than per-fetus indicates (via simulation).

Univariate Analysis.

The simple analysis is based on litter as the experimental unit. This analysis is carried out using binomial and poisson models. The binomial assumption states that conditional on litter size the number affected is binomial. The analysis is based on transformed data, usually the arc-sine of the observed proportion. The poisson model does not account for litter size as it assumes the mean number affected is the same for all dose groups. The analysis again used a transformation, usually the square root of the observed number. Neither fits the data very well. This may be due to extra binomial or extra poisson variability, as the case may be.

More sophisticated models are reviewed by Haseman and Kupper (1979) include: weighted least squares based on proportion and unequal sample sizes. This approach due to Cochran (1943) requires sample sizes which are too large for this application. Others include, the normal-binomial (Luning et. al. 1966), beta-binomial (Williams 1975), negative binomial (McCaughran and Arnold 1976), correlated - binomial Altham (1978), jackknife Gladen (1979). Several nonparametric procedures have been tried, namely the Mann-Whitney U, the Kruskal-Wallis and the Jouckheere/ Terspstra. Some attempts at multivariate analysis have been tried by Ryttman (1976), log-linear models by Haberman and others (1974) and generalized linear models by McCullagh (1980). All of the latter techniques have distributional assumptions.

Since some of the ordered categorical procedures develop or accept scores for the categories, this approach was pursued. Scores induce relative spacing among the categories. Thus, a mean score may be obtained for each litter. This implies analysis by litter as a sampling unit. We note that CATMOD in SAS allows for scoring, but the scores must be

user specified.

Ipsen (1955) suggested a scoring for Bioassay. Instead of estimating an LD_{50} or ED_{50} based on number of survivors after x days, he ordered the data into categories with the continuum represented by time (days). The scores proposed are such that the variance of the linear regression of mean scores on the log dose is maximized with respect to the total variance. An adjustment is made if the scores do not reflect the ordering of the categories.

Bradley et.al. (1962) scores by maximizing the treatment sum of squares after scaling relative to the error sum of squares. This is an iterative procedure which does not require the assumption of linearity.

Using no distributional assumption, Nair (1986, 1987) suggested some techniques for analyzing ordered categorical data in the field of quality control. He showed the Taguchi statistic for 2 x 2 tables, can be orthogonally decomposed into K - 1 components where K is the number of categories. In the two sample case he showed the first components is equivalent to Wilcoxon's rank test on grouped data. Thus, this components would be sensitive to shifts in the multinomial model. Further, the second components corresponds to Mood's rank test for grouped data, thus is sensitive to scale changes in the 2 x K model.

In the non equiprobable case the correspondence does not apply though the interpretation still holds. This result has been verified using a comparison density approach for the two sample problem by Eubank, LaRiccia and Rosenstein (1987).

The decomposition of Taguchi's accumulation chi-squared (1966, 1974) requires the solution of an eigenvector problem. Nair (1986, 1987) provides the method for deriving two sets of scores. These yield statistics that are approximately equal to those obtained from the orthogonal decomposition, but do not require a rigorous solution. The approximate and exact statistics have comparable power.

When applied to 2 x K tables, the first set of Nair's scores is sensitive to shifts in location of the underlying random variable. It is reasonable to suggest, when applied to litters, these scores yield a continuous index useful for detecting shift. In teratogenicity studies the location shifts of interest would be those that indicate a significant dose-response.

Nair's Method

As already mentioned, the first and second components of the orthogonal decomposition correspond to the Wilcoxon and Mood rank test, respectively.

Wilcoxon tests,

$$H_0 : G(x) = F(x)$$
$$H_1 : G(x) = F(x - \delta)$$

where F, G are two distribution functions.

Mood tests,

$H_0 : G(x) = F(x)$
$H_1 : G(x) = F(x/\theta)$
where θ is a constant.

For more than two treatment groups the first component corresponds to the Kruskal - Wallis statistic for grouped data, and the second to the generalized Mood statistic. In the general case (except equiprobable) case the equalities are no longer exact, but the first two components have good power for detecting location and scale shifts respectively. The focus of this work is on location shifts.

Notation

Observed frequency for the $(i,k)^{th}$ cell $= Y_{ik}$ $\quad i = 1,2$ $\quad j = 1,2,...K$
Column total. $= C_K = Y_{1K} + Y_{2K}$
Row total. $= R_i = R_i = \sum_{k=1}^{K} Y_{ik}$
Cumulative row frequencies $Z_{ik} = \sum_{j=1}^{k} Y_{ij}$
Cumulative column totals. $D_k = \sum_{j=1}^{k} c_j$
The row proportions $= r_i = R_i/N$
The column proportions $= c_k = C_k/N$
Cumulative column proportion up to and including column k $d_k = D_k/N$.

Vector conventions used.

bold lower case letter = a vector
bold upper case letter = a matrix
transpose = ^{t}
a vector raised to a power implies each element is raised to a power
(this is non standard)

Multinomial model. $2 \times K$ case

Two random samples of size R_i $i = 1,2$ are drawn from two multinomial populations.
For each population, the probabilities of the K outcomes are given by

$$P_{ik}\ k\ =\ 1,\ 2,\ \ldots\ K.$$

The cumulative probabilities for population i are given by

$$\tau_{ik}\ =\ \sum_{j=1}^{k} P_{ij}$$

If the K categories are assumed to be ordered the hypothesis is

$$H_0:\ \tau_{1k}\ =\tau_{2k}\quad k=1,\ 2,\ \ldots\ K$$
$$H_1:\ (\tau_{1k}\ -\ \tau_{2k})\leq\ 0 \text{ for all } K$$
(strict inequality for at least one k).

Alternative statistics to Pearson's χ^2 .

Taguchi's statistic $$T_E\ =\ \sum_{k=1}^{K}\ [d_k(1-d_k)]^{-1}\{\sum_{i=l}^{2}\ R_i(Z_{ik}/R_i\ -\ d_k)\}$$

If χ^2_{kP} is the Pearson χ^2 statistic from a 2×2 table where column 1 contains the cumulative frequencies of categories 1 through k and column 2 contains the cumulative frequencies of categories (k+1) through K. Then,

$$T_E = \sum_{k=1}^{K-1}\ \chi^2{}_{kP}$$

where T_E is a "ccs type statistic".

T_E assigns weight $w_k\ =\ [d_k(1-d_k)]^{-1}$ to the kth term in the sum which is equal for each k under H_0.

Nair's Statistic $$T\ =\ \sum_{k=1}^{K-1}\ w_k\ \{\sum_{i=1}^{2}\ R_i(z_{ik}/R_i - d_k)^2$$

The statistics in the class are obtained by the choice of the set $\{w_k\}$ where $w_k > 0$ for $k = 1,\ 2,\ \ldots\ K-1$. The decomposition is carried out conditionally on the marginal proportions. For y_1, $w_k\ k = 1,\ 2,\ \ldots\ K-1$, W is a diagonal matrix. Using the d we form the $(K-1) \times K$ matrix A by;

$$A = \begin{pmatrix} 1-d_1 & -d_1\ldots\ldots & & -d_1 \\ 1-d_2 & 1-d_2 & -d_2\ldots & -d_2 \\ \vdots & & & \\ 1-d_{k-1} & 1-d_{k-1}\ldots\ldots & 1-d_{k-1} & -d_{k-1} \end{pmatrix}$$

Thus T is given by

$$T = y_1' A' W A y_1 / N r_1 r_2 .$$

To express T as a sum of squares in y_1 we need to express

$$A'WA$$

as a product of a diagonal matrix, Q and its transpose. Let Λ be diagonal of order K formed by the colum proportions $\{c_k\}$ and let Γ be the diagonal matrix of order $(k-1)$ containing the eigenvalues of $A'WA\Lambda$ then the decomposition yields

$$A'WA = Q\Gamma Q'$$

where Q contained the eigenvectors of $A'WA\Lambda$ such that

$$\overline{Q} = [1 \mid Q]$$

satisfies

$$\overline{Q}' \Lambda \overline{Q} = I$$

substituting $Q\Gamma Q'$ into T above with

$$u = Q'(N r_1 r_2)^{-\frac{1}{2}}$$

yields

$$T = \sum_{j=1}^{K-1} \tau_j U_j^2$$

where the τ_j's are elements of the vector of eigenvalues, τ, and U_j's are elements of u. Under H_0 the distribution of y_1 conditional on row and column proportions is multiple hypergeometric with

$$
\begin{aligned}
E(Y_{1k}) &= Nr_1c_k \\
Cov(Y_{1k}, Y_{1l}) &= N(1 - 1/N)^{-1}r_1r_2(1 - c_k) \qquad k = 1 \\
&= -N(1 - 1/N)r_1r_2(c_kc_l) \qquad k \neq 1
\end{aligned}
$$

or

$$
\begin{aligned}
E(y_1) &= Nr_1\Lambda 1 \\
cov(y_1 &= N(1 - 1/N)^{-1}r_1r_2\Lambda[(I - 1) \cdot 1^t\Lambda]
\end{aligned}
$$

1 is a $K \times 1$ vector of ones. It follows that

$$
\begin{aligned}
E(u) &= Nr_1(Q^t\Lambda 1) / \sqrt{Nr_1r_2} &= 0 \\
cov(u) &= (1 - 1/N)^{-1}Q^t\Lambda[I - 1\ 1^t\Lambda]Q &= (1 - 1/N)^{-1}I
\end{aligned}
$$

implies the U_j's are uncorrelated with zero means.

Under H_0 it can be shown that the limiting distribution of y_1 converges to the multivariate normal distribution as N goes to infinity. Thus

$$
T = \sum_{j-1}^{K-1} \tau_j {U_j}^2 \xrightarrow{d} \sum_{j=1}^{K-1} \tau_j {\chi^2}_{1,j}
$$

a weighted sum of independent χ^2 random variables, each with 1 *df*

The approximate solution by Nair proposed two sets of statistics which have properties the same as those obtained for the equiprobable case (i.e. $c_k = 1/k$). That is the first component of T_E or $U_{E,1}$ is equivalent to the Wilcoxon test on the $2 \times K$ table and $U_{E,2}$, the second component, is equivalent to Mood's

$$
M = \sum_{k=1}^{K} [k - (K + 1)/2]^2 Y_{1k}.
$$

They do not require the solution of the eigenvalue problem as orthogonal decomposition is not necessary. The first component, all observations in the category are assigned a

score proportional to the midpoint of the category. The second component the scores are quadratic in the midrank. Additionally, each set of scores is adjusted to satisfy orthogonality.

To calculate the scores, let c be of length K with elements being the column proportions. Form

$$B = \begin{pmatrix} .5 & 0..... & & & 0 \\ 1 & .5 & 0.... & & 0 \\ \vdots & & & & \\ 1 & 1 & .5 & 0... & 0 \\ 1 & 1 & 1 & 1 & .5 \end{pmatrix}$$

Let $\tau = Bc$ and $\tau^* = \tau - .5(1)$. Note the τ's are Bross's ridits. The first set of scores is obtained from

$$1 = \tau^* / \sqrt{c^t \tau^{*2}}$$

where τ^{*2} is a vector of squares of elements of τ^*. The second set of scores is obtained in two steps. First let

$$e = 1^t[1 - (c^t 1^3)1] - 1.$$

Then

$$s = e/c^t e^2.$$

The approximate statistics for the $2 \times K$ table are

$$V_1^2 = L_1^2/R_1 + L_2{}^2/R_2$$

where

$$L_i = 1^t y_i \quad 1, 2$$

and

$$V_2{}^2 = S_1{}^2/R_1 + S_2{}^2/R_2$$

where

$$S_i = s^t\, i = 1, 2$$

which are comparable in magnitude and consequently in power to U_1 and U_2 respectively.

We now apply the method:

Conduct of the study and data:

PROTOCOL

1. Sprague-Dawley rat study
2. Herbicide: nitrofen (2, 4-dichloro-4 nitrodipheynl ether)
3. Test compound administered during organogenesis
4. Sacrifice prior to parturition and cesarean-sectioned
5. Record litter and fetal data
6. Administration of compound follows daily dose regimen
7. Treatment groups; control and three dose groups
8. Inseminated females randomly assigned to 4 groups of 24 rats each
9. Dose levels. 6.25, 12.5, 25 mg/kg/day body weight on days 6-15 of gestation.
   Controls: gavage solution w/o test compound

Live fetuses are weighed, sexed and examined for external malformations. They are then sacrificed in order to perform the skeletal and visceral examination. Recorded are the number of corpora luta on each ovary, number of implantations, number of fetuses, and the number of resportions in each utrine horn. Table I displays the data for each rodent and close level.

The following definitions are employed to categorize the fetuses: Dead - Early or late resorption of dead at c-section, malformed-gross visceral or skeletal variation, growth retarded - body weight more than two standard deviation from the mean for the given sex or by a range test. Normal - absence of any of the previous outcomes. Table's II and III summarize the results by number and percent for each dose by category. It should be noted that the differing number of letters is due to nonpregnant females, not toxicity.

The final columm of Table I is the calculated severity index. This index is calculated by multiplying the score for the category by the number of fetuses in the category, summing and dividing by the number of implantations, i.e.,

$$SI = n^t c / n^t 1.$$

Details of the calculations of a severity index are given in the following example.

Consider the following sample data:

Table I

Nitrofen Data - Strague-Dawley Rats

Id	Number of Implantations	Normal	Growth Retared	Malformed	Dead	Severity Index
Dose Group = Control (0.0 mg/kg/day b.w.)						
19	1	1	0	0	0	0.00000
8	4	3	0	0	1	0.68423
11	5	4	1	0	0	0.25139
7	8	7	0	0	1	0.34212
1	12	9	0	3	0	0.49145
16	14	14	0	0	0	0.00000
24	14	11	0	3	0	0.42124
6	15	15	0	0	0	0.00000
9	15	13	0	1	1	0.31352
20	15	12	0	3	0	0.39316
22	15	12	1	2	0	0.34590
2	16	14	1	1	0	0.20142
4	16	16	0	0	0	0.00000
10	16	16	0	0	0	0.00000
12	16	14	2	0	0	0.15712
17	16	16	0	0	0	0.00000
23	16	15	0	0	1	0.12286
3	17	11	0	6	0	0.69382
5	17	10	0	6	1	0.85481
13	17	13	0	3	1	0.50790
15	17	13	0	2	2	0.55326
21	18	13	0	4	1	0.58890
Dose Group = Low (6.25 mg/kg/day b.w.)						
32	1	0	0	0	1	2.73692
28	12	10	0	2	0	0.32763
43	12	9	0	3	0	0.49145
26	14	10	0	4	0	0.56166
31	14	14	0	0	0	0.00000
39	14	11	0	3	0	0.42124
41	14	9	0	5	0	0.70207
47	14	10	0	4	0	0.56166
48	14	14	0	0	0	0.00000
33	15	10	0	5	0	0.65527
38	15	10	0	5	0	0.65527
40	15	13	0	1	1	0.31352
45	15	12	0	3	0	0.39316
25	16	11	0	5	0	0.61432
27	16	12	0	4	0	0.49145
34	16	10	0	6	0	0.73718
35	16	12	0	4	0	0.49145
37	16	9	0	7	0	0.86004
44	16	12	0	4	0	0.49145
46	16	11	0	5	0	0.61432
36	17	7	0	8	2	1.24708

Table I (Cont'd.)

Nitrofen Data - Strague-Dawley Rats

Id	Number of Implantations	Growth Normal	Retared	Malformed	Dead	Severity Index
Dose Group = Control (0.0 mg/kg/day b.w.)						
54	2	0	0	0	2	2.73692
70	3	1	0	2	0	1.31054
59	4	2	0	2	0	0.98290
64	8	5	0	3	0	0.73718
53	11	4	1	5	1	1.25663
55	13	7	0	5	1	0.96661
58	14	7	0	6	1	1.03798
60	14	7	0	5	2	1.09306
65	14	8	0	5	1	0.89757
68	14	10	0	3	1	0.61674
62	15	6	0	6	3	1.33371
67	15	13	0	1	1	0.31352
71	15	8	0	4	3	1.07160
49	16	6	0	10	0	1.22863
69	16	11	0	4	1	0.66251
56	18	15	0	2	1	0.37047
57	18	13	0	5	0	0.54606
72	18	7	0	11	0	1.20133
Dose Group = High (25.0 mg/kg/day b.w.)						
91	2	0	0	1	1	2.35136
80	7	3	0	3	1	1.23348
86	8	6	1	1	0	0.40284
73	10	1	0	9	0	1.76923
77	14	3	0	11	0	1.54456
78	14	3	0	11	0	1.54456
79	14	2	0	12	0	1.68498
83	14	9	0	5	0	0.70207
93	14	1	0	12	1	1.88047
76	15	0	0	15	0	1.96581
84	15	6	0	9	0	1.17949
92	15	7	0	8	0	1.04843
74	16	4	0	11	1	1.52255
87	16	6	0	10	0	1.22863
94	16	8	0	8	0	0.98290
95	16	6	0	10	0	1.22863
96	16	4	0	10	2	1.57075
89	17	6	0	11	0	1.27199
90	17	0	0	11	6	2.23797
75	18	6	0	12	0	1.31054
81	18	6	0	12	0	1.31054
88	19	11	0	7	1	0.86829

Table II

Group	Number of Implantations				
	Normal	Gr.Retarded	Malformed	Dead	Total
Control	252	5	35	8	300
Low	239	1	89	5	334
Mid	130	1	79	18	228
High	98	1	199	13	311

Table III

Group	Percent of Implantations			
	Normal	Gr.Retarded	Malformed	Dead
Control	84.0	1.7	11.7	2.7
Low	71.6	0.3	26.6	1.5
Mid	57.0	0.4	34.6	7.9
High	31.5	0.3	64.0	4.2

	Number of Implantations (Fetuses)				
	Normal	Gr.Retarded	Malformed	Dead	Totals
Control	252	5	35	8	300
Low Dose	239	1	89	5	334
Middle Dose	130	1	79	18	228
High Dose	98	1	199	13	311
Totals	719	8	402	44	1173

Calculate the column proportions:
c_k .61295823 .00682012 .34271100 .03751066

Calculate Bross's ridits (1958) by the formula $\tau_k = (c_0 + c_1 + \ldots + c_{k-1}) + .5c_k$ where $c_0 = 0$:
τ_k .30647912 .61636829 .79113385 .98124468

Now, let $\tau_k^* = \tau_k - .5$:
τ_k^* -.19352088 .11636829 .29113385 .48124468

Calculate the constant $d = [c_1\tau_1^{*2} + c_2\tau_2^{*2} + C_3\tau_3^{*2} + c_4\tau_4^{*2}]^{1/2}$:

$$d = [.61295823(-.19352088)^2 + .00682012(.11636829)^2 + .34271100(.29113385)^2 + .03751066(.48124468)^2]^{1/2}$$
$$= .24654207$$

The vector of scores (Nair, 1986, 1987) is then obtained by $l_k = \tau_k^*/d$:
l_k -0 7849 0.4720 1.1809 1.9520

Shifting the scores so that the score for a normal implantation (fetus) is zero, the final scores are:
l_k^* 0.0000 1.2569 1.9658 2.7369

Then, a litter with 11 implantations of which 4 are classified as normal, 1 as growth retarded, 5 as malformed and 1 dead, would have a severity index of:

$$SI=[0.0000(4)+1.2569(1)+1.9658(5)+2.7369(1)]/11=1.2566$$

This can be interpreted in light of the above scores, i.e., an index near zero would be indicative of a litter with nearly all normal fetuses at cesarean-section and a score near 2.7369 would be indicative of a litter with nearly all fetuses dead at cesarean-section.

Designs for the Analysis of the Severity Index

Five designs were evaluated which assume normality. The one-way classification, a one-way classification using litter size as a covariate, a generalized randomized block using

litter size as a blocking variable, and a weighted analysis using in one case litter size as a weight and in another the square root of litter size. The results are summarized in Table IV in terms of calculated F, associated P values and R^2.

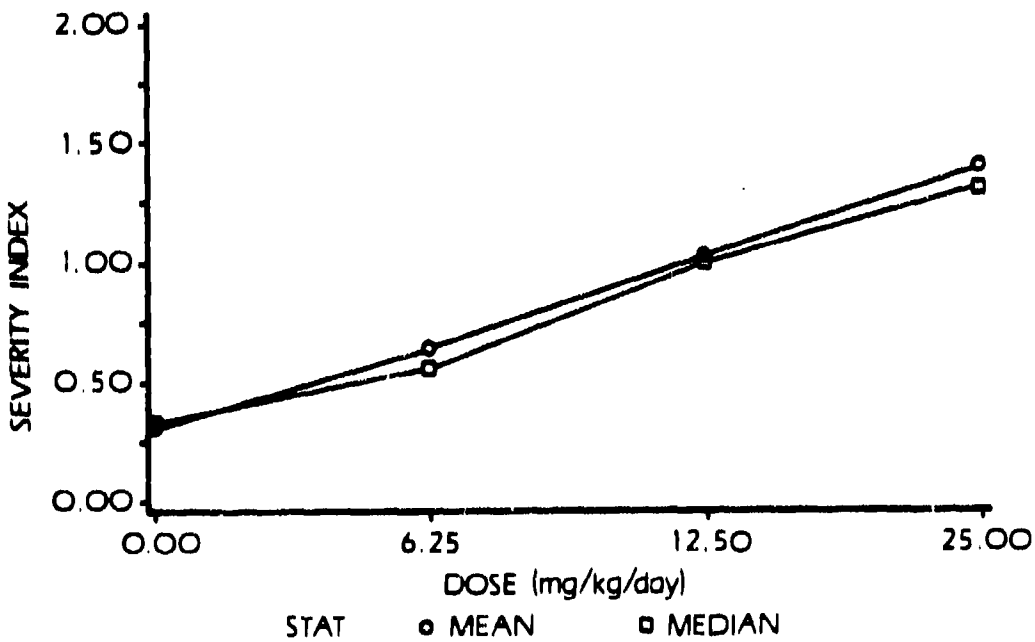
Table IV

	f	F	P	R^2
One way analysis	3,81	22.97	< .0001	.46
Covariance	3,80	25.99	< .0001	.53
Generalized RBD	3,65	20.52	< .0001	.58
Weighted AOV (litter size)	3,81	15.62	< .0001	.37
Weight AOV ($\sqrt{\text{littersize}}$)	3,81	33.11	< .0001	.55

As was expected the covariance and blocking provided an improvement over the one-way classification as measured by R^2. However, the magnitude of the improvement does not seem to warrent the chance of violating the more restrictive assumptions placed on the experiment by those designs. A better alternative, in the parametric case, may be using the square root of litter size as a weight which provides nearly the same value of R^2 as does the blocking design. However, we would prefer the one-way analysis for its simplicity and robustness in application.

The normality assumption on the severity index is quite suspect in many situations. As an alternative, the nowparametric Kruskal-Wallis procedure was carried out. In view of the overwhelming significance of the parametric procedures, this result was not surprising $\chi^2 = 47.75; af = 3\ p < .0001$. Figure 1 compares the linearity of the mean severity index and the median severity index.

Figure 1



Statistical Procedure

The consideration of litter size is not necessary for analysis of the SI's. It is important to note than the SI's are probably not normally distributed, particularly in the control group and at the higher dose levels. The following is suggested for toxicity-teratogenicity studies.

1. If the SI's are reasonably normal, calculate the AOV F-statistic for a one way layout. Use this statistic to test for differences in location.
2. If F is significant, follow with linear constrasts to test for increasing trend.
3. If significant use Dunnett's procedure to compare control mean with each of the treatment means to establish no-effect leve.
4. In the presence of non normality use a similar sequence of nonparametric test. e.g. K - W, Jonckheere/Terpstra, and Dunn's procedure.

SAS code is available which reads litter data, calculates scores, computes SI's and calculated the statistics. The results above have been "tested" by simulation analysis of additional nitrofen studies and two other biological examples. Also, the method detects different dose patterns with equal ability. The K - W test showed consistently higher power than the F-statistic in the simulation studies.

REFERENCES

[1] Altham, P.M.E: (1978). Two generalizations of the binomial distribution. Applied Statistics **27**, 162-167

[2] Bradley, R. A., S. K. Katti and I. J. Coons. (1962). Optimal scaling for ordered categories. Psychometrika **27**, 355-374.

[3] Bross, I.D.J. (1958). How to under ridit analysis. Biometrics **14**, 18-38.

[4] Cochran, W. G. (1943). Analysis of variance for percentages based on unequal numbers. J. American Statistical Association **38**, 287-301.

[5] Eubank, R.L., V.N. LaRiccia and R.B. Rosenstein. (1987). Test statistics derived as components of Pearson's phi-squared distance measure. J. American Statistical Association **82**, 816-825.

[6] Gladen, B. (1979). The use of the jackknife to estimate proportions from toxicological data in the presence of litter effects. J. American StatisticalAssociation **74**, 278-283.

[7] Haberman, S.J. (1974). Log-linear models for frequency tables with ordered classifications. Biometrics **30**, 589-600.

[8] Haseman, J.K. and M.D. Hogan. (1975). Selection of the experimental unit in teratology students. Teratology **12**, 165-171.

[9] Haseman, J.K. and L.L. Kupper. (1979). Analysis of dichotomous response data from certain toxicological experiments. Biometrics **35**, 281-293.

[10] Ipsen, J. (1955). Appropriate scores in bio-assay using death times and survivor symptoms. Biometrics **2**, 465-480.

[11] Luning, K. G., W. Sheridan, K. H. Ytterborn and U. Gullberg. (1966). The relationship between the number of implantations and the rate of intra-uterine death in mice. Mutation Research **3**, 444-451.

[12] McCaughran, D.A. and D.W. Arnold. (1976). Statistical models for numbers of implantation sites and embryonic deaths in mice. Toxicology and Applied Pharmacology **38** , 325-333.

[13] McCullagh, P. (1980). Regression models with ordinal data (with discussion). J. Royal Statistical Society, B **42**, 109-142.

[14] Morgenthien, E.A. (1988). An Ordered Categorical Approach to the Analysis of Qualitative Data From Developmental Toxicity/Teratogenicity Studies. Unpublished Ph.D. Dissertation, University of Delaware.

[15] Nair, V.J. (1986). Testing in industrial experiments with ordered categorical data (with discussion). Technometrics **28**. 283-311.

[16] Nair, V.J. (1987). Chi-squared type tests for ordered alternatives in contingency tables. J. American Statistical Association **82**, 283-291.

[17] Ryttman, H. (1976) A new statistical evaluation of the dominant-lethal mutation test. Mutation Research **38**, 225-238.

[18] Williams, D.A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. Biometrics **31**, 949-952.

A SMALL SAMPLE POWER STUDY OF THE ANDERSON-DARLING STATISTIC AND A COMPARISON WITH THE KOLMOGOROV AND THE CRAMÉR-VON MISES STATISTICS

Linda L. Crawford Moss, US Army Ballistic Research Laboratory
Malcolm S. Taylor, US Army Ballistic Research Laboratory
Henry B. Tingey, University of Delaware

Abstract

The Anderson-Darling goodness-of-fit procedure emphasizes agreement between the data and the hypothesized distribution in the extremes or tails. An improved table of the quantiles of the Anderson-Darling statistic, useful for small sample sizes, was constructed using the Cray-2 supercomputer. The power of the Anderson-Darling test is compared to the Kolmogorov and the Cramér-von Mises tests when the null hypothesis is the normal distribution and the alternative distributions are the Cauchy, the double exponential, and the extreme value distributions.

1. INTRODUCTION

Consider a random sample $X_1, X_2, ..., X_n$ from a population with a continuous distribution function. One method of testing the hypothesis that the n observations come from a population with a specified distribution function $F(x)$ is by a chi-square test. This test requires a subjective partitioning of the real line $\mathbf{R}$ and a comparision of the empirical histogram with the hypothetical histogram. A more objective method, is to compare the empirical distribution function $F_n(x)$ with the hypothetical distribution function $F(x)$. The empirical distribution function based on n observations is defined as $F_n(x) = \frac{k}{n}$ if exactly k observations are less than or equal to x, for $k = 0, 1, ..., n$.

To compare the empirical and hypothetical distribution functions a measure of their difference is required. Addressing this, Anderson and Darling [1952] considered the following metrics in function space:

$$W_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \, \psi[F(x)] \, dF(x) \tag{1.1}$$

and

$$K_n = \sup_{-\infty < x < \infty} \sqrt{n}\,|F_n(x) - F(x)|\sqrt{\psi[F(x)]}\,. \qquad (1.2)$$

Samples producing large values of W_n^2 (or K_n) lead to rejection of the null hypothesis that the population distribution function is F(x). One of the contributions of Anderson and Darling was the incorporation of a non-negative weight function ψ in (1.1) and (1.2). By a suitable choice for ψ, specific ranges of values of the random variable X, corresponding to different regions of the distribution F(x), may be emphasized. For $\psi[F(x)] \equiv 1$, W_n^2 becomes the Cramér-von Mises statistic [Cramér, 1928 and von Mises, 1931] and K_n becomes the Kolmogorov statistic [Kolmogorov, 1933].

The tails of the distribution function will be accentuated in the investigation detailed in this paper; Anderson and Darling suggest using

$$\psi[F(x)] = \frac{1}{F(x)[1 - F(x)]}\,.$$

With this choice for the weighting function, metric (1.1) becomes the basis for the Anderson-Darling statistic.

In Section 2, the Anderson-Darling test statistic is developed; in Section 3, the most accurate tabulation to date of the test statistic is provided. In Section 4, the description and the results of a power study are given in which the Anderson-Darling, the Cramér-von Mises, and the Kolmogorov statistics are compared.

2. THE ANDERSON-DARLING STATISTIC

For a fixed value of the random variable X, say $X = x$, the empirical distribution function $F_n(x)$ is a statistic, since it is a function of the sample values $x_1, x_2, \ldots, x_n$. The distribution of this statistic is established as a lemma.

Lemma (2.1): If $F_n(x)$ is the empirical distribution function corresponding to a random sample $X_1, X_2, \ldots, X_n$ of size n from a distribution $H(\cdot)$, then for a fixed x, $nF_n(x)$ is distributed binomial$(H(x), n)$.

Proof:
$P(nF_n(x) = k) = P(\text{exactly k values } x_i \leq x)$, for $k = 0, 1, \ldots, n$.

Let $Z_i = I_{(-\infty, x]}(X_i)$, where the indicator function I is defined as

$$I_{(-\infty, x]}(X_i) = \begin{cases} 1, & \text{if } -\infty < X_i \leq x \\ 0, & \text{otherwise .} \end{cases}$$

Then ΣZ_i counts the number of sample values $x_i \leq x$.
Here each $Z_i \sim$ Bernoulli$(H(x))$, so $\Sigma Z_i \sim$ binomial$(H(x), n)$.
Therefore,

$$\begin{aligned} P(nF_n(x) = k) &= P(\text{exactly k values } x_i \leq x) \\ &= P(\Sigma Z_i = k) \\ &= \binom{n}{k} H(x)^k \left(1 - H(x)\right)^{n-k} . \blacksquare \end{aligned}$$

From Lemma 2.1,

$$E[F_n(x)] = \frac{1}{n} E\left[nF_n(x)\right] = H(x)$$

and

$$\mathrm{Var}[F_n(x)] = \frac{1}{n^2} \mathrm{Var}\left[nF_n(x)\right] = \frac{1}{n} H(x)\,[1 - H(x)]\,. \tag{2.1}$$

To assist in the determination of a suitable weighting function $\psi(\cdot)$, that is, a function that will weight more heavily values in the tails of the distribution F(x) at the expense of values closer to the median, consider the expectation of the squared discrepancy $E\left[F_n(x) - F(x)\right]^2$. It is important to keep in mind that the value x is fixed, so F(x) is a constant, and the expectation is with respect to the random variable $F_n(x)$ whose distribution was established in Lemma 2.1. Then

$$n\,E\left[F_n(x) - F(x)\right]^2 = n\,E\left[F_n(x) - H(x) + H(x) - F(x)\right]^2$$

$$= n\,E\left[\left\{F_n(x) - H(x)\right\} - \left\{F(x) - H(x)\right\}\right]^2$$

which, after algebraic manipulation (Appendix A) yields the variance and bias^2

$$= n\left[\frac{1}{n}\left\{H(x)\{1 - H(x)\}\right\} + \left\{F(x) - H(x)\right\}^2\right]. \tag{2.2}$$

Under the null hypothesis H_o: $H(x) = F(x) \quad \forall x$, (2.2) becomes

$$n\, E\left[F_n(x) - F(x)\right]^2 = F(x)[1 - F(x)]\,. \tag{2.3}$$

Anderson-Darling chose as a weighting function, $\psi[F(x)] = \dfrac{1}{F(x)\,[\,1 - F(x)]}$. Weighting by the reciprocal of (2.3) takes into consideration the variance of the statistic $F_n(x)$ and also maintains the objective of accentuating values in the tails of $F(x)$.

With this choice of weighting function and without loss of generality assuming $x_1 \leq x_2 \leq \ldots \leq x_n$, let $F(x) = u$, $dF(x) = du$, and $F(x_j) = u_j$. Then the Anderson-Darling test statistic (2.4) can be rewritten as expression (2.5) by expansion and integration (Appendix B).

$$W_n^2 = n \int_{-\infty}^{\infty} \frac{[F_n(x) - F(x)]^2}{F(x)\,[1 - F(x)]} dF(x)\,, \tag{2.4}$$

$$W_n^2 = -\,n - \frac{1}{n} \sum_{j=1}^{n} \left[(2j-1)\ln u_j + (2(n-j)+1)\ln(1-u_j)\right]. \tag{2.5}$$

3. DISTRIBUTION OF THE ANDERSON-DARLING STATISTIC

The asymptotic distribution of W_n^2 was derived by Anderson and Darling [1952]. Lewis [1961] undertook the tabulation of $F(z; n) = P(W_n^2 \leq z)$ for $n = 1, 2, ..., 8$ and for incremental values of z over the interval [0.025, 8.000]. Lewis' table entries were computed using a Monte Carlo procedure to generate an empirical approximation $F_m(z;n)$ to the distribution function $F(z;n)$ based on m samples of size n. At that time, computational restrictions essentially limited the accuracy of the table entries to within 0.00326 of the true value.

Following an analogous procedure based on expression (2.5) and the observation that the U_i are distributed U[0,1] [Feller, 1966], the table appearing in Lewis' paper was recalculated using a Cray-2 supercomputer. A Kolmogorov-type bound [Conover, 1980] was used to construct a 95% confidence band for the distribution function $F(z;n)$.

In general, the width of a $(1 - \alpha)100\%$ confidence band is equal to twice the value of the $(1 - \alpha)100\%$ quantile of the Kolmogorov statistic $K_m = \sup_{-\infty < x < \infty} \sqrt{m}\,|F_m(x) - F(x)|$, where m is the number of sample

values used in the construction of $F_m(x)$. With n fixed, the 95% confidence band can be made arbitrarily small by a suitable choice for m, the number of Monte Carlo samples. The commonly tabled [Miller, 1956] asymptotic approximation for the 95th quantile is $1.358/\sqrt{m}$. However, Harter [1980] suggests using

$$\frac{1.358}{\left[m + \frac{r}{3.5}\right]^{\frac{1}{2}}}, \quad \text{where } r = (m+4)^{\frac{1}{2}} \tag{3.1}$$

for an improved approximation.

Using approximation (3.1) to construct a 95% confidence band with the width not exceeding 0.001, the value for m must be at least 7,375,881. In this simulation, m was chosen to be 7.4 million. Table 1 lists the reconstruction of Lewis' table, now accurate within 0.0005. Again, z ranges from 0.025 to 8.000 and for n = 1, 2, ..., 10. The column labeled "∞" contains the asymptotic values, rounded to four decimal places.

TABLE 1. Quantiles of F(z;n) for n = 1 thru 10 and ∞.

	n										
z	1	2	3	4	5	6	7	8	9	10	∞
0.0250	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0750	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.1000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.1250	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0003
0.1500	0.0000	0.0000	0.0000	0.0000	0.0003	0.0006	0.0008	0.0009	0.0010	0.0010	0.0014
0.1750	0.0000	0.0000	0.0000	0.0014	0.0025	0.0029	0.0033	0.0034	0.0036	0.0037	0.0042
0.2000	0.0000	0.0000	0.0022	0.0065	0.0075	0.0081	0.0085	0.0087	0.0089	0.0090	0.0096
0.2250	0.0000	0.0000	0.0120	0.0152	0.0160	0.0166	0.0170	0.0173	0.0174	0.0176	0.0180
0.2500	0.0000	0.0008	0.0260	0.0274	0.0283	0.0288	0.0291	0.0294	0.0294	0.0296	0.0296
0.2750	0.0000	0.0299	0.0431	0.0431	0.0440	0.0444	0.0445	0.0448	0.0447	0.0449	0.0443
0.3000	0.0000	0.0586	0.0625	0.0621	0.0628	0.0629	0.0629	0.0631	0.0628	0.0629	0.0618
0.3250	0.0000	0.0869	0.0840	0.0838	0.0842	0.0839	0.0838	0.0837	0.0833	0.0834	0.0817
0.3500	0.0000	0.1147	0.1071	0.1078	0.1073	0.1068	0.1065	0.1063	0.1058	0.1056	0.1036
0.3750	0.0000	0.1421	0.1318	0.1330	0.1319	0.1311	0.1305	0.1303	0.1296	0.1295	0.1269
0.4000	0.1166	0.1691	0.1579	0.1591	0.1574	0.1564	0.1555	0.1552	0.1547	0.1543	0.1513
0.4250	0.1949	0.1955	0.1853	0.1856	0.1833	0.1821	0.1812	0.1808	0.1803	0.1797	0.1764
0.4500	0.2484	0.2217	0.2136	0.2120	0.2096	0.2083	0.2072	0.2066	0.2061	0.2056	0.2019
0.4750	0.2913	0.2474	0.2420	0.2384	0.2358	0.2344	0.2331	0.2325	0.2321	0.2315	0.2276
0.5000	0.3278	0.2727	0.2693	0.2646	0.2618	0.2603	0.2589	0.2584	0.2579	0.2573	0.2532
0.5250	0.3598	0.2976	0.2955	0.2902	0.2875	0.2859	0.2844	0.2839	0.2834	0.2828	0.2786
0.5500	0.3885	0.3221	0.3209	0.3154	0.3128	0.3111	0.3095	0.3091	0.3085	0.3079	0.3036
0.5750	0.4145	0.3462	0.3455	0.3398	0.3373	0.3356	0.3341	0.3337	0.3329	0.3325	0.3281
0.6000	0.4385	0.3700	0.3693	0.3637	0.3612	0.3595	0.3580	0.3576	0.3569	0.3564	0.3520
0.6250	0.4607	0.3936	0.3922	0.3870	0.3844	0.3827	0.3812	0.3809	0.3803	0.3798	0.3753
0.6500	0.4814	0.4166	0.4143	0.4096	0.4069	0.4052	0.4038	0.4035	0.4029	0.4024	0.3980
0.6750	0.5007	0.4394	0.4357	0.4314	0.4286	0.4270	0.4256	0.4254	0.4248	0.4244	0.4199
0.7000	0.5188	0.4618	0.4563	0.4524	0.4495	0.4481	0.4467	0.4464	0.4459	0.4456	0.4412

TABLE 1. continued

	n										
z	1	2	3	4	5	6	7	8	9	10	∞
0.7500	0.5522	0.5056	0.4953	0.4923	0.4893	0.4879	0.4867	0.4864	0.4859	0.4856	0.4815
0.8000	0.5821	0.5458	0.5317	0.5289	0.5261	0.5249	0.5238	0.5235	0.5231	0.5228	0.5190
0.8500	0.6091	0.5794	0.5652	0.5627	0.5601	0.5590	0.5581	0.5580	0.5574	0.5573	0.5537
0.9000	0.6337	0.6090	0.5962	0.5938	0.5916	0.5905	0.5897	0.5897	0.5891	0.5891	0.5858
0.9500	0.6563	0.6358	0.6247	0.6225	0.6206	0.6196	0.6190	0.6188	0.6184	0.6183	0.6154
1.0000	0.6771	0.6601	0.6511	0.6489	0.6473	0.6464	0.6458	0.6457	0.6453	0.6454	0.6427
1.0500	0.6963	0.6823	0.6754	0.6732	0.6718	0.6712	0.6706	0.6704	0.6701	0.6703	0.6680
1.1000	0.7141	0.7028	0.6977	0.6958	0.6944	0.6939	0.6934	0.6933	0.6931	0.6932	0.6912
1.1500	0.7306	0.7217	0.7182	0.7164	0.7153	0.7149	0.7145	0.7144	0.7142	0.7143	0.7127
1.2000	0.7459	0.7392	0.7367	0.7356	0.7345	0.7342	0.7339	0.7339	0.7337	0.7338	0.7324
1.2500	0.7603	0.7555	0.7538	0.7531	0.7522	0.7520	0.7518	0.7518	0.7517	0.7518	0.7508
1.3000	0.7737	0.7704	0.7696	0.7693	0.7686	0.7684	0.7683	0.7683	0.7684	0.7684	0.7677
1.3500	0.7862	0.7844	0.7842	0.7844	0.7838	0.7837	0.7836	0.7836	0.7837	0.7838	0.7833
1.4000	0.7980	0.7975	0.7977	0.7982	0.7978	0.7977	0.7978	0.7977	0.7979	0.7980	0.7978
1.4500	0.8090	0.8097	0.8103	0.8110	0.8108	0.8109	0.8109	0.8108	0.8110	0.8112	0.8111
1.5000	0.8194	0.8210	0.8221	0.8229	0.8228	0.8230	0.8230	0.8230	0.8232	0.8235	0.8235
1.5500	0.8291	0.8316	0.8330	0.8340	0.8340	0.8342	0.8343	0.8343	0.8345	0.8348	0.8350
1.6000	0.8382	0.8416	0.8431	0.8442	0.8444	0.8447	0.8448	0.8448	0.8450	0.8453	0.8457
1.6500	0.8469	0.8508	0.8526	0.8538	0.8541	0.8544	0.8546	0.8546	0.8549	0.8550	0.8556
1.7000	0.8550	0.8595	0.8615	0.8627	0.8631	0.8634	0.8636	0.8636	0.8640	0.8642	0.8648
1.7500	0.8626	0.8677	0.8698	0.8709	0.8715	0.8718	0.8720	0.8720	0.8725	0.8727	0.8734
1.8000	0.8698	0.8754	0.8775	0.8787	0.8793	0.8797	0.8798	0.8799	0.8803	0.8805	0.8814
1.8500	0.8767	0.8826	0.8847	0.8860	0.8866	0.8870	0.8872	0.8872	0.8877	0.8879	0.8888
1.9000	0.8830	0.8893	0.8915	0.8928	0.8934	0.8938	0.8940	0.8941	0.8945	0.8947	0.8957
1.9500	0.8891	0.8956	0.8979	0.8991	0.8997	0.9001	0.9004	0.9006	0.9008	0.9011	0.9021
2.0000	0.8948	0.9015	0.9038	0.9051	0.9058	0.9061	0.9064	0.9066	0.9068	0.9071	0.9082
2.0500	0.9002	0.9070	0.9093	0.9107	0.9113	0.9117	0.9120	0.9121	0.9124	0.9127	0.9138
2.1000	0.9053	0.9122	0.9146	0.9159	0.9165	0.9169	0.9172	0.9174	0.9177	0.9179	0.9190
2.1500	0.9102	0.9171	0.9195	0.9208	0.9214	0.9218	0.9221	0.9223	0.9226	0.9228	0.9239
2.2000	0.9148	0.9218	0.9241	0.9254	0.9259	0.9264	0.9266	0.9268	0.9272	0.9273	0.9285

TABLE 1. continued

	n										
z	1	2	3	4	5	6	7	8	9	10	∞
2.2500	0.9191	0.9262	0.9284	0.9297	0.9302	0.9307	0.9309	0.9311	0.9315	0.9316	0.9328
2.3000	0.9232	0.9303	0.9325	0.9337	0.9343	0.9347	0.9349	0.9351	0.9355	0.9356	0.9368
2.3500	0.9271	0.9342	0.9363	0.9375	0.9381	0.9385	0.9387	0.9389	0.9393	0.9394	0.9405
2.4000	0.9308	0.9378	0.9399	0.9411	0.9416	0.9420	0.9423	0.9425	0.9428	0.9429	0.9441
2.4500	0.9342	0.9412	0.9433	0.9444	0.9449	0.9454	0.9456	0.9458	0.9461	0.9463	0.9474
2.5000	0.9375	0.9445	0.9465	0.9476	0.9480	0.9485	0.9487	0.9489	0.9492	0.9494	0.9504
2.5500	0.9407	0.9475	0.9494	0.9505	0.9510	0.9514	0.9516	0.9519	0.9521	0.9523	0.9534
2.6000	0.9436	0.9504	0.9523	0.9532	0.9537	0.9542	0.9544	0.9546	0.9548	0.9550	0.9561
2.6500	0.9465	0.9531	0.9549	0.9558	0.9563	0.9567	0.9570	0.9572	0.9574	0.9576	0.9586
2.7000	0.9491	0.9556	0.9574	0.9583	0.9588	0.9592	0.9594	0.9596	0.9599	0.9600	0.9610
2.7500	0.9517	0.9580	0.9598	0.9606	0.9611	0.9614	0.9617	0.9619	0.9622	0.9623	0.9633
2.8000	0.9541	0.9602	0.9620	0.9628	0.9633	0.9636	0.9638	0.9640	0.9643	0.9644	0.9654
2.8500	0.9564	0.9624	0.9641	0.9648	0.9653	0.9657	0.9659	0.9661	0.9663	0.9665	0.9674
2.9000	0.9586	0.9644	0.9661	0.9668	0.9672	0.9676	0.9678	0.9680	0.9682	0.9684	0.9692
2.9500	0.9607	0.9663	0.9679	0.9686	0.9690	0.9694	0.9696	0.9698	0.9700	0.9701	0.9710
3.0000	0.9626	0.9681	0.9696	0.9703	0.9708	0.9711	0.9713	0.9714	0.9717	0.9718	0.9726
3.0500	0.9645	0.9697	0.9713	0.9719	0.9724	0.9727	0.9729	0.9730	0.9733	0.9734	0.9742
3.1000	0.9663	0.9713	0.9728	0.9735	0.9739	0.9742	0.9744	0.9745	0.9748	0.9749	0.9756
3.1500	0.9679	0.9729	0.9743	0.9749	0.9753	0.9757	0.9758	0.9760	0.9762	0.9763	0.9770
3.2000	0.9695	0.9743	0.9757	0.9763	0.9767	0.9770	0.9771	0.9773	0.9775	0.9776	0.9783
3.2500	0.9711	0.9756	0.9770	0.9776	0.9780	0.9783	0.9784	0.9786	0.9787	0.9788	0.9795
3.3000	0.9725	0.9769	0.9782	0.9788	0.9792	0.9794	0.9796	0.9797	0.9799	0.9800	0.9807
3.3500	0.9738	0.9781	0.9794	0.9799	0.9803	0.9806	0.9807	0.9809	0.9810	0.9811	0.9818
3.4000	0.9751	0.9792	0.9805	0.9810	0.9814	0.9816	0.9817	0.9819	0.9821	0.9822	0.9828
3.4500	0.9764	0.9803	0.9815	0.9820	0.9824	0.9826	0.9827	0.9829	0.9830	0.9831	0.9837
3.5000	0.9775	0.9814	0.9825	0.9830	0.9833	0.9836	0.9837	0.9839	0.9840	0.9840	0.9846
3.5500	0.9786	0.9823	0.9834	0.9839	0.9842	0.9844	0.9846	0.9847	0.9849	0.9849	0.9855
3.6000	0.9797	0.9833	0.9843	0.9848	0.9851	0.9853	0.9854	0.9856	0.9857	0.9857	0.9863
3.6500	0.9807	0.9841	0.9851	0.9856	0.9859	0.9861	0.9862	0.9864	0.9865	0.9865	0.9870
3.7000	0.9816	0.9849	0.9859	0.9863	0.9867	0.9868	0.9869	0.9871	0.9872	0.9872	0.9878

TABLE 1. continued

	n										
z	1	2	3	4	5	6	7	8	9	10	∞
3.7500	0.9825	0.9857	0.9867	0.9871	0.9874	0.9875	0.9877	0.9878	0.9879	0.9879	0.9884
3.8000	0.9834	0.9865	0.9874	0.9878	0.9881	0.9882	0.9883	0.9885	0.9886	0.9886	0.9891
3.8500	0.9842	0.9872	0.9880	0.9884	0.9887	0.9888	0.9889	0.9891	0.9892	0.9892	0.9897
3.9000	0.9850	0.9878	0.9887	0.9890	0.9893	0.9894	0.9895	0.9897	0.9898	0.9898	0.9902
3.9500	0.9858	0.9885	0.9893	0.9896	0.9899	0.9900	0.9901	0.9902	0.9903	0.9903	0.9908
4.0000	0.9864	0.9890	0.9898	0.9902	0.9904	0.9905	0.9906	0.9908	0.9909	0.9908	0.9913
4.0500	0.9871	0.9896	0.9904	0.9907	0.9909	0.9910	0.9911	0.9913	0.9913	0.9913	0.9917
4.1000	0.9877	0.9901	0.9909	0.9912	0.9914	0.9915	0.9916	0.9917	0.9918	0.9918	0.9922
4.1500	0.9883	0.9906	0.9914	0.9917	0.9919	0.9919	0.9920	0.9922	0.9922	0.9922	0.9926
4.2000	0.9889	0.9911	0.9918	0.9921	0.9923	0.9924	0.9925	0.9926	0.9927	0.9927	0.9930
4.2500	0.9895	0.9916	0.9922	0.9925	0.9927	0.9928	0.9929	0.9930	0.9931	0.9931	0.9934
4.3000	0.9900	0.9920	0.9926	0.9929	0.9931	0.9931	0.9932	0.9934	0.9934	0.9934	0.9938
4.3500	0.9905	0.9924	0.9930	0.9933	0.9935	0.9935	0.9936	0.9937	0.9938	0.9938	0.9941
4.4000	0.9910	0.9928	0.9934	0.9936	0.9938	0.9938	0.9939	0.9941	0.9941	0.9941	0.9944
4.5000	0.9918	0.9935	0.9941	0.9943	0.9945	0.9945	0.9946	0.9947	0.9948	0.9947	0.9950
4.6000	0.9926	0.9941	0.9946	0.9949	0.9950	0.9951	0.9951	0.9952	0.9953	0.9953	0.9955
4.7000	0.9933	0.9947	0.9952	0.9954	0.9956	0.9956	0.9956	0.9957	0.9958	0.9958	0.9960
4.8000	0.9939	0.9952	0.9957	0.9959	0.9960	0.9960	0.9961	0.9962	0.9962	0.9962	0.9964
4.9000	0.9945	0.9957	0.9961	0.9963	0.9964	0.9964	0.9965	0.9966	0.9966	0.9966	0.9968
5.0000	0.9950	0.9961	0.9965	0.9967	0.9968	0.9968	0.9968	0.9969	0.9969	0.9969	0.9971
5.5000	0.9970	0.9977	0.9979	0.9980	0.9981	0.9981	0.9982	0.9982	0.9982	0.9982	0.9983
6.0000	0.9982	0.9986	0.9988	0.9988	0.9989	0.9989	0.9989	0.9989	0.9990	0.9990	0.9990
7.0000	0.9993	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
8.0000	0.9997	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

4. POWER STUDY

The power of the Anderson-Darling test was compared with two other goodness-of-fit procedures based on the empirical distribution function: the Kolmogorov and the Cramér-von Mises tests. The Kolmogorov statistic introduced in Section 1 as metric (1.2) with weighting function $\psi\,[F(x)] = 1$ becomes

$$K_n = \sup_{-\infty < x < \infty} \sqrt{n}\,|F_n(x) - F(x)|\,. \tag{4.1}$$

For an ordered sample $x_1 \leq x_2 \leq \cdots \leq x_n$ and $F(x_i) = u_i$, (4.1) may be evaluated using $D = \max(D^+, D^-)$, where

$$D^+ = \max_i \left(\frac{i}{n} - u_i\right) \qquad \text{and}$$

$$D^- = \max_i \left(u_i - \frac{i-1}{n}\right).$$

The Cramér-von Mises statistic, defined as

$$W_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2\, dF(x)$$

can be reduced to (4.2) for ease of computation (Appendix C);

$$W_n^2 = \sum_{i=1}^{n} \left[u_i - \frac{2i-1}{2n} \right]^2 + \frac{1}{12n} \quad . \tag{4.2}$$

In the power study, two cases were considered. Case 1 corresponds to the situation in which the parameters of the hypothesized distribution are completely specified. Case 2 corresponds to the situation in which the parameters are not specified and must be estimated from the sample data.

For both case 1 and 2, the null hypothesis is

H_o: A random sample $X_1, X_2, \ldots, X_n$ comes from a normal population

or

H_o: $H(x) = F(x)$, where $F(x) \sim N(\mu, \sigma^2)$.

As alternative hypotheses, the Cauchy, double exponential, and extreme value distributions were chosen, each with location parameter the same as the null hypothesis. This provided a heavy-tailed, light-tailed, and skewed distribution, respectively, against which the power of the three goodness-of-fit tests are compared.

The power functions do not exist in closed form; they are approximated empirically via a Monte Carlo simulation. To determine a point on the power curve, a large number of samples of size n was generated from a specific distribution serving as the alternative hypothesis. The number of times the null hypothesis was rejected at a specific level of significance was recorded. The ratio of the number of rejections, Y, to the total number of

samples generated, N, provides an estimate, $\hat{p} = Y/N$, of the probability of rejecting the null hypothesis when it should be rejected (power). The value $\hat{p}$ determines a point on the power curve corresponding to a specific sample size n, a specific significance level α, and a specific alternative hypothesis.

To determine the number of samples of size n required for a sufficiently accurate estimate of $\hat{p}$, a nonparametric technique was employed. Since the counter Y is distributed binomial($\cdot$; p,N) where the parameter p is the true but unknown power, and since an approximate confidence interval for p can be constructed [Conover, 1980] using

$$1-\alpha \simeq P\left[\frac{Y}{N} - z_{1-\frac{\alpha}{2}}\left\{\frac{Y}{N}\left(1-\frac{Y}{N}\right)/N\right\}^{\frac{1}{2}} < p < \frac{Y}{N} + z_{1-\frac{\alpha}{2}}\left\{\frac{Y}{N}\left(1-\frac{Y}{N}\right)/N\right\}^{\frac{1}{2}}\right], \quad (4.3)$$

samples of size n continued to be generated from the alternative distribution until the confidence interval for p given in (4.3) was sufficiently small.

The confidence interval coefficient $1 - \alpha$ was chosen to be 0.975 and the confidence interval width not to exceed 0.025. Then the confidence limits (4.3) were successively evaluated until the interval width was satisfied. Considering a "worst-case" scenario in which $p = 1/2$ and the variance of its estimate $\hat{p}$ is greatest, equating $Y/N = 1/2$ in (4.3) suggests that samples of magnitude 8037 might be required. A minimum value for N of 100 was imposed to prevent premature termination of the procedure.

4.1. Case 1: Distribution Parameters Specified.

The power study for case 1 specified the parameters of the hypothesized distribution as N(0,1). The results of the study are summarized in Figures 1 — 12. For each of the three distributions serving as an alternative hypothesis, samples of size n = 5, 10, 15, 20 were chosen for study and, as previously mentioned, the location parameters of both the null and alternative hypotheses coincided. The scale parameter for the alternative hypothesis were values from 0.025 to 3.000 in increments of 0.025.

The level of significance for the study was 0.05. The critical value for each test was determined from tables in Conover [1980] for the Kolmogorov test, Stephens and Maag [1968] for the Cramér-von Mises test, and Table 1 in Section 3 of this paper for the Anderson-Darling test.

The Anderson-Darling test demonstrated overall superiority for the sample sizes and hypotheses chosen for this study. This is perhaps to be anticipated in view of the emphasis on agreement in the tails by the Anderson-Darling procedure, but the magnitude of difference over the Kolmogorov and Cramér-von Mises tests is impressive.

The power curves corresponding to n = 10, 15, 20 are distinguished by their characteristic of decreasing to a global minimum before becoming

monotone increasing. An explanation of this feature is suggested by consideration of Figures 13 – 15 in which the distribution functions of the N(0,1) and Cauchy (0,ς) are compared. There it is seen (Figure 14) that corresponding to $\varsigma = 0.50$ the two distribution functions are similar; an increase (decrease) in the scale parameter ς causes the tails of the distributions to become more distinct. Values in a neighborhood of $\varsigma \simeq 0.50$ marked the global minimum throughout the study.

4.2. Case 2: Distribution Parameters Estimated.

The Anderson-Darling, Kolmogorov, and Cramér-von Mises goodness-of-fit tests were developed for use in case 1 where distribution parameters are specified, and so precludes their use in the more likely situation where parameters must be estimated. In practice, these procedures are sometimes used anyway with the caveat that the tests are likely to be conservative. Stephens [1974] provides adjustments to the test statistics that enable the tests to be used to test the assumption H_o: $H(x) = F(x)$, where $F(x) \sim N(\mu, \sigma^2)$ and the population parameters are estimated from the data.

The results of the power study for case 2, are summarized in Figures 16 – 27. As in case 1, the sample sizes are n = 5, 10, 15, and 20, and

the level of significance is 0.05. Both location and scale parameters coincide; the scale parameter are values from 0.025 to 3.000 in increments of 0.025.

The power plots are horizontal, demonstrating that power does not change with scale parameter and provides empirical support for Stephens' transformations. Power increases with increasing sample size, as would be expected. When both location and scale parameters agree, all three tests are competitive for the sample sizes and alternative distributions chosen for this study.

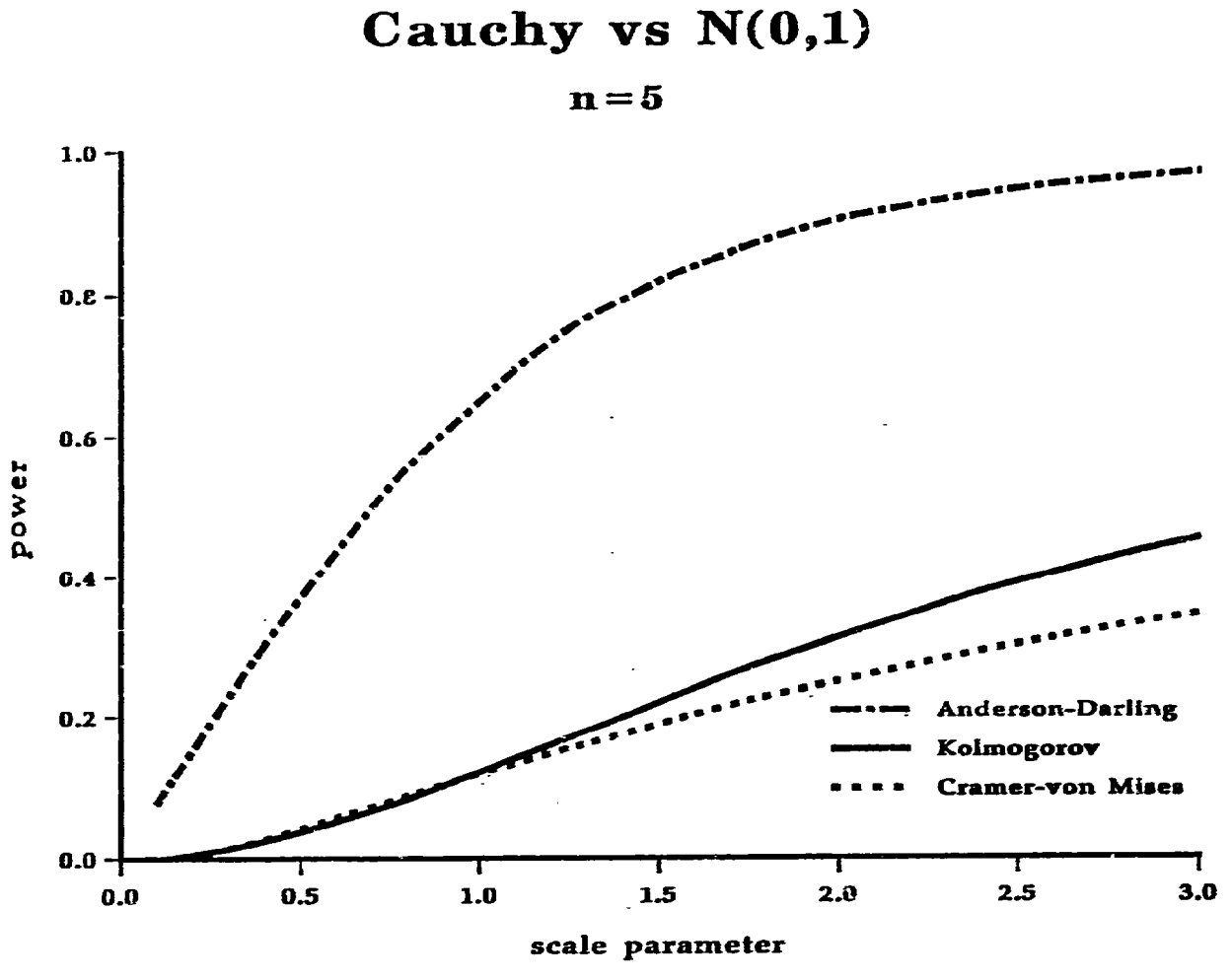


FIGURE 1. Power Curve for Cauchy Generated Data, n = 5

FIGURE 2 Power Curve for Cauchy Generated Data, n = 10

FIGURE 3. Power Curve for Cauchy Generated Data, n = 15

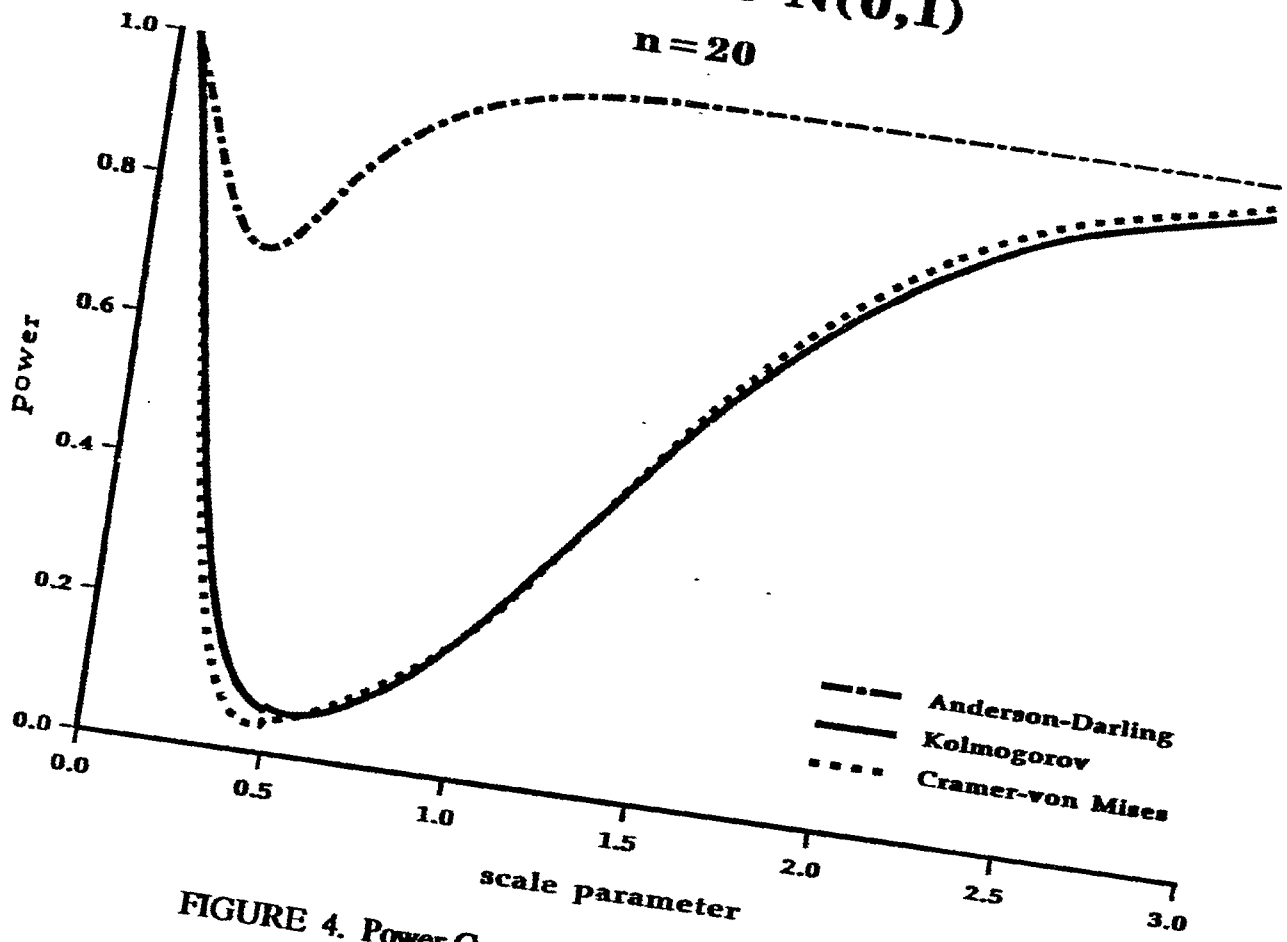


FIGURE 4. Power Curve for Cauchy Generated Data, n = 20

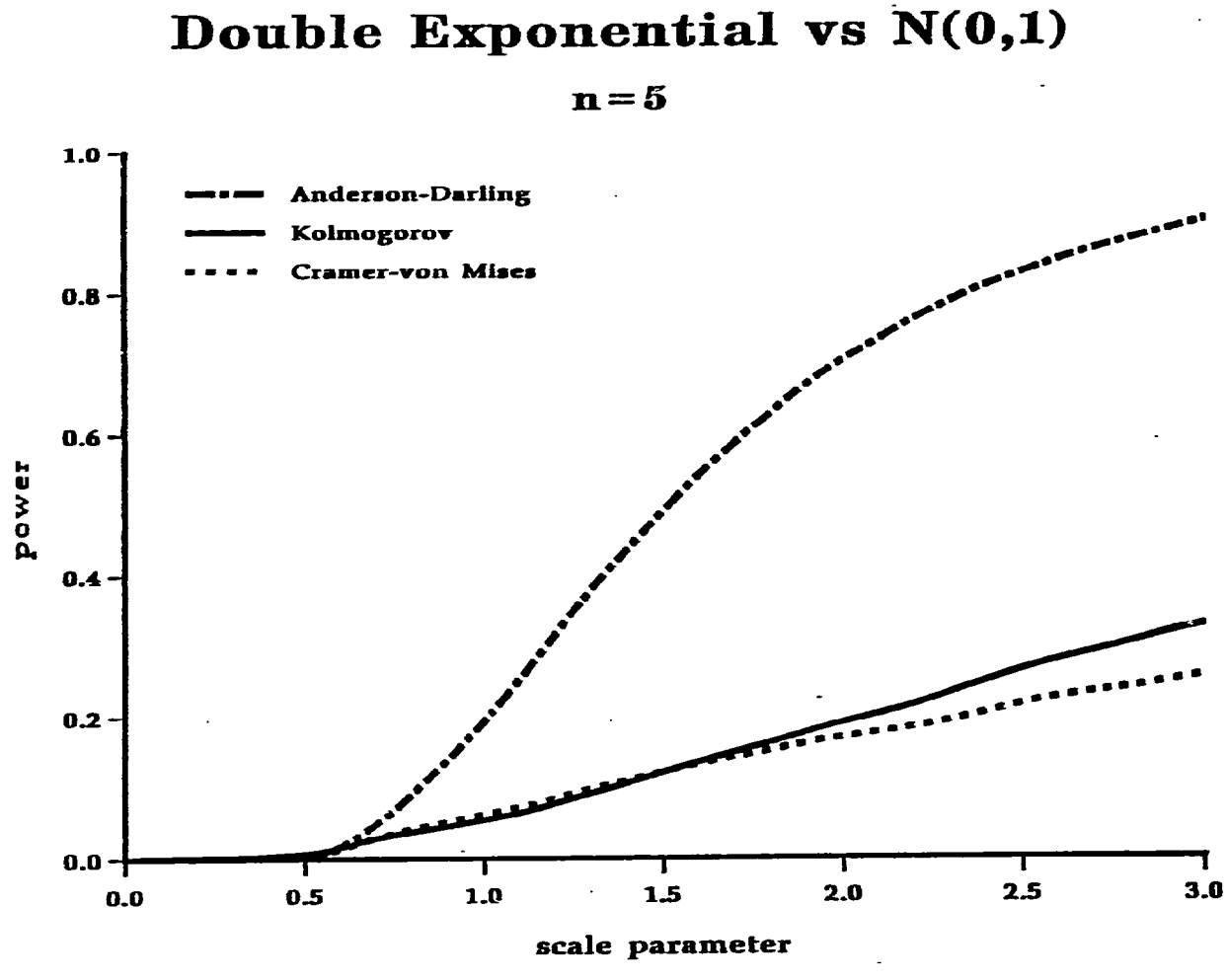


FIGURE 5. Power Curve for Double Exponential Generated Data, n = 5

FIGURE 6. Power Curve for Double Exponential Generated Data, n = 10

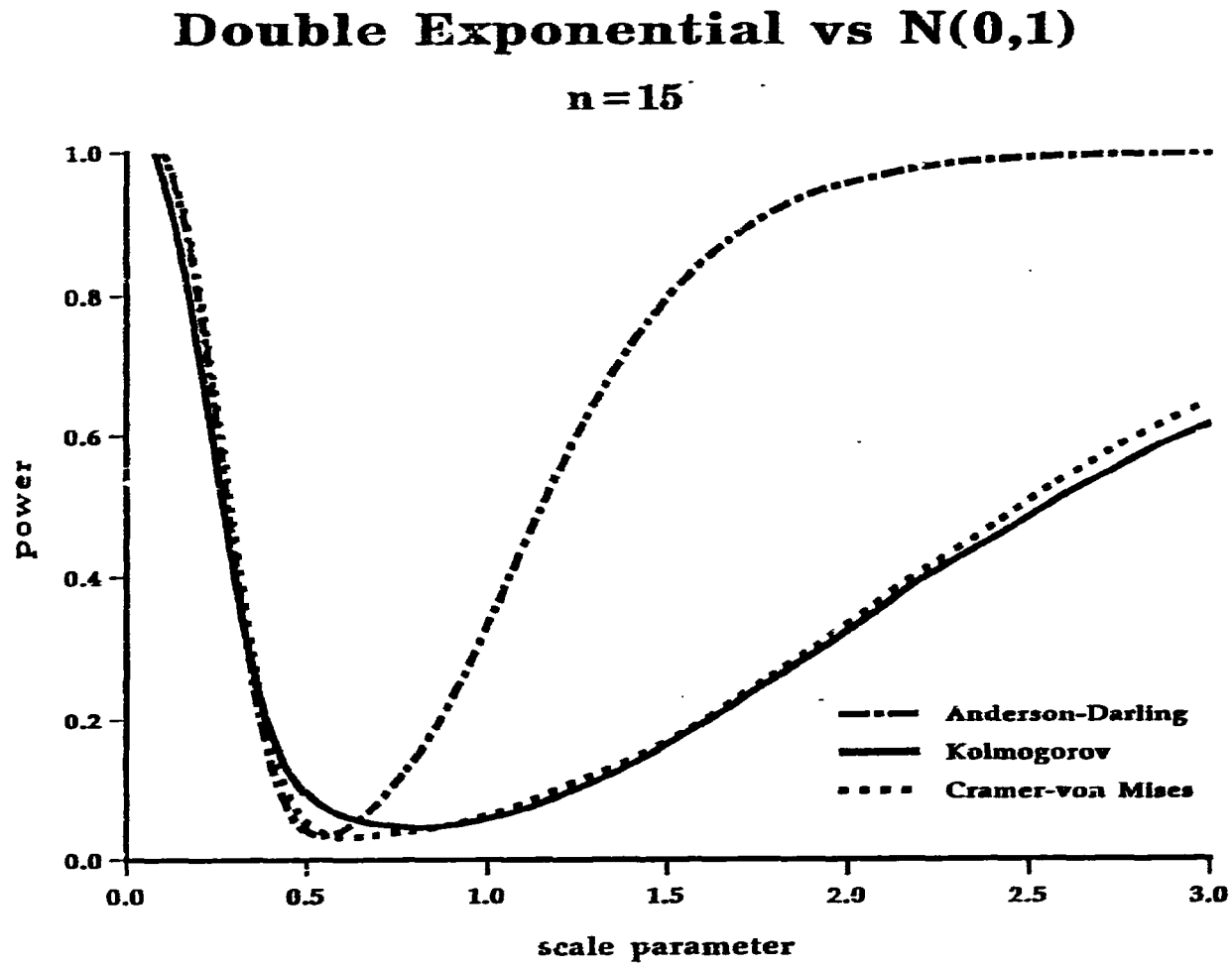


FIGURE 7. Power Curve for Double Exponential Generated Data, n = 15

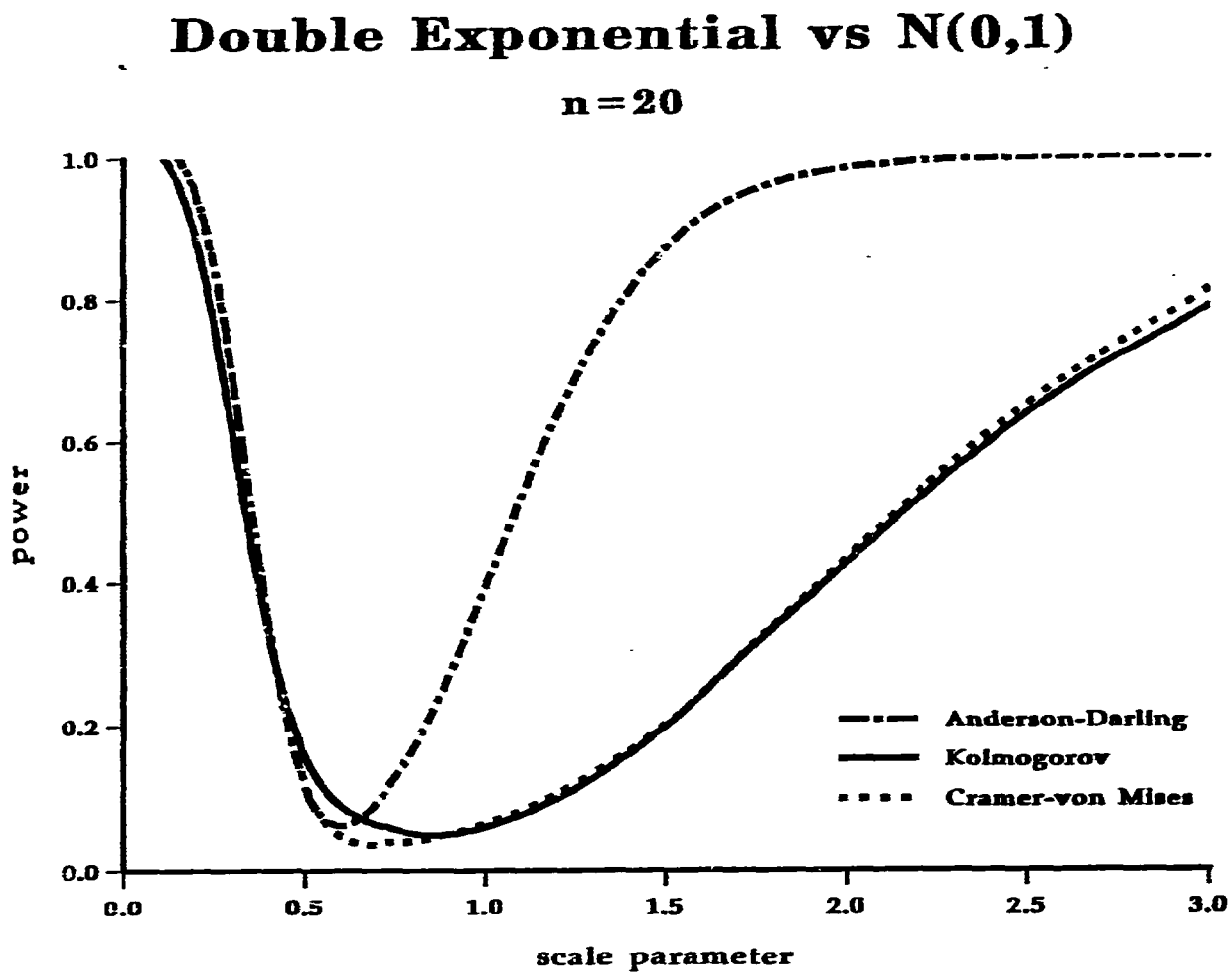


FIGURE 8. Power Curve for Double Exponential Generated Data, n = 20

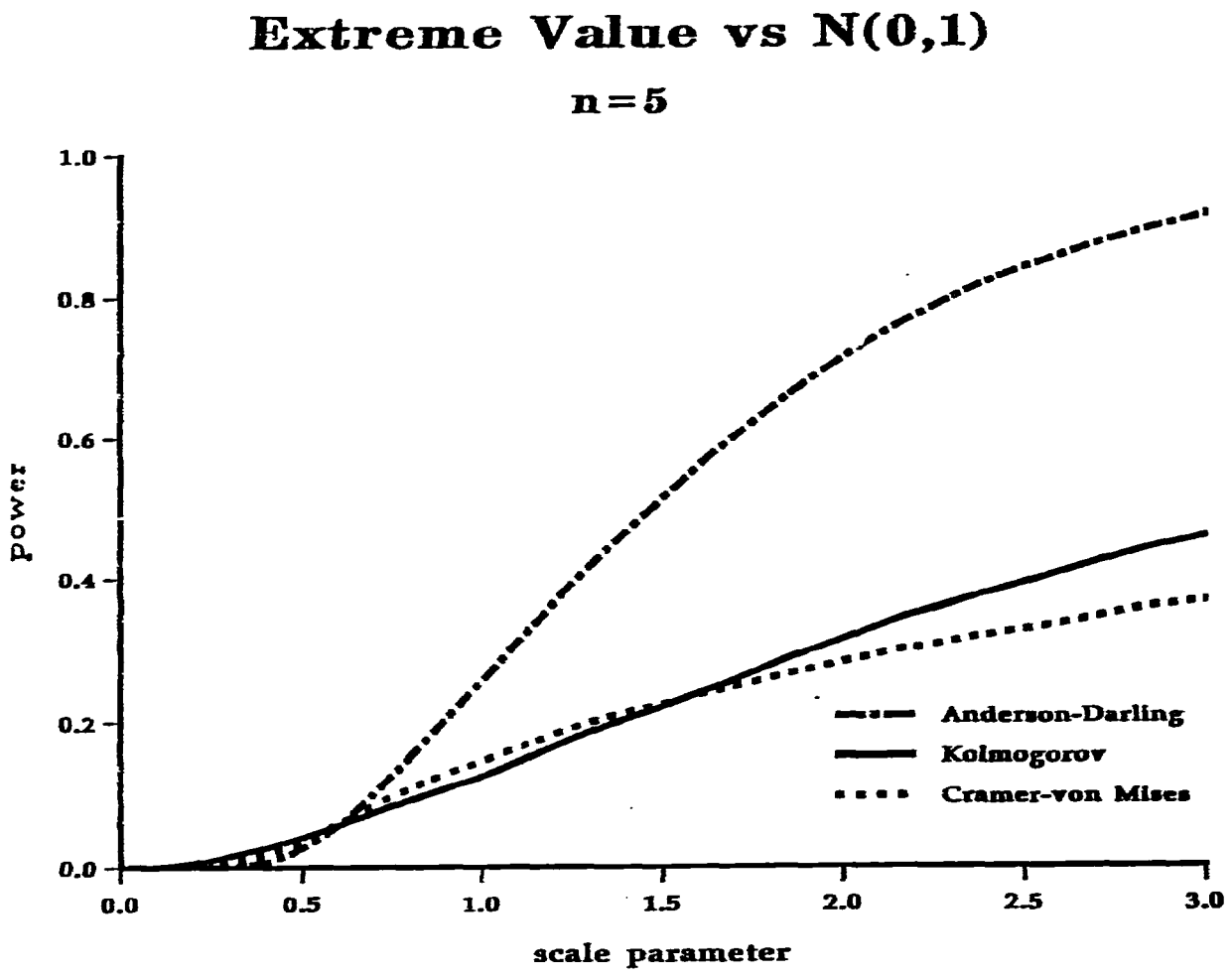


FIGURE 9. Power Curve for Extreme Value Generated Data, n = 5

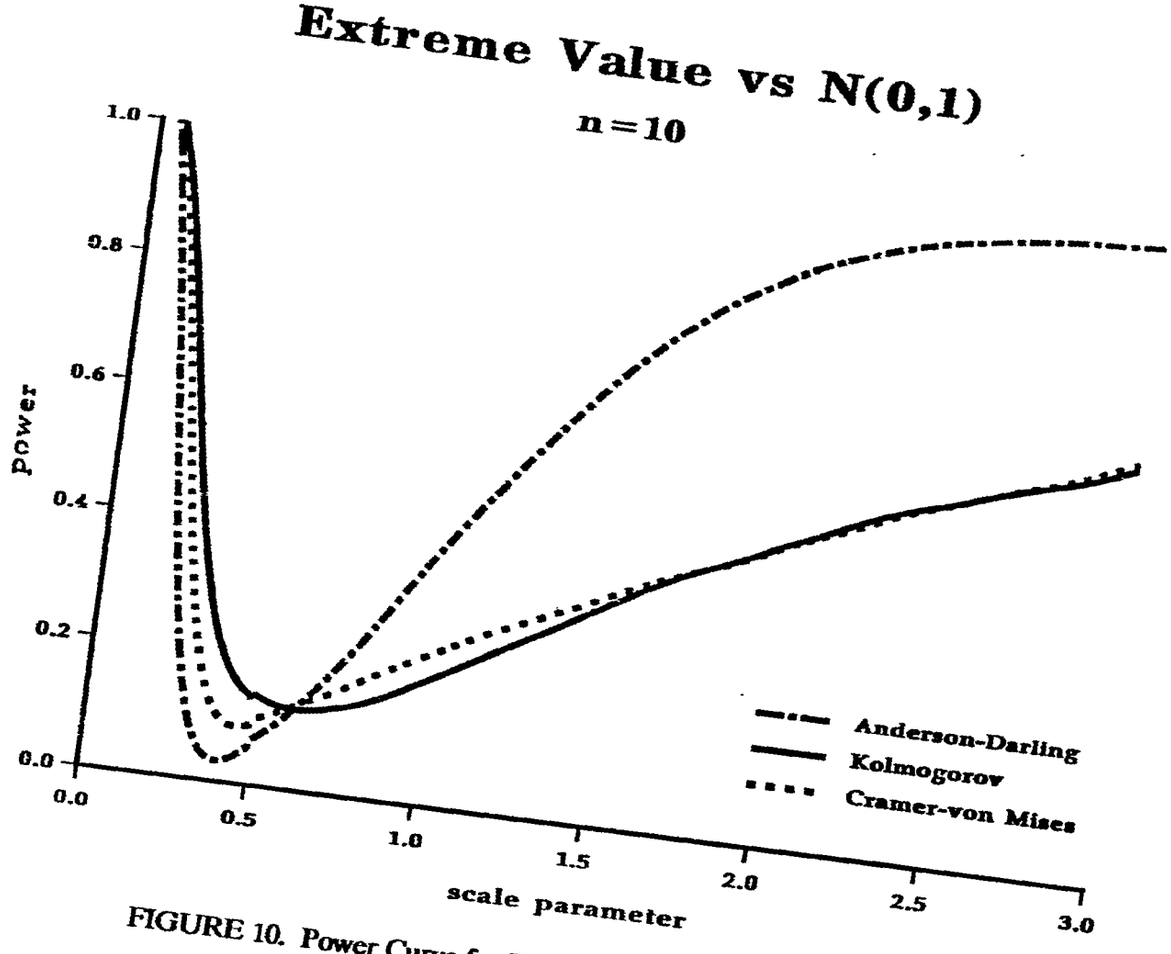


FIGURE 10. Power Curve for Extreme Value Generated Data, n = 10

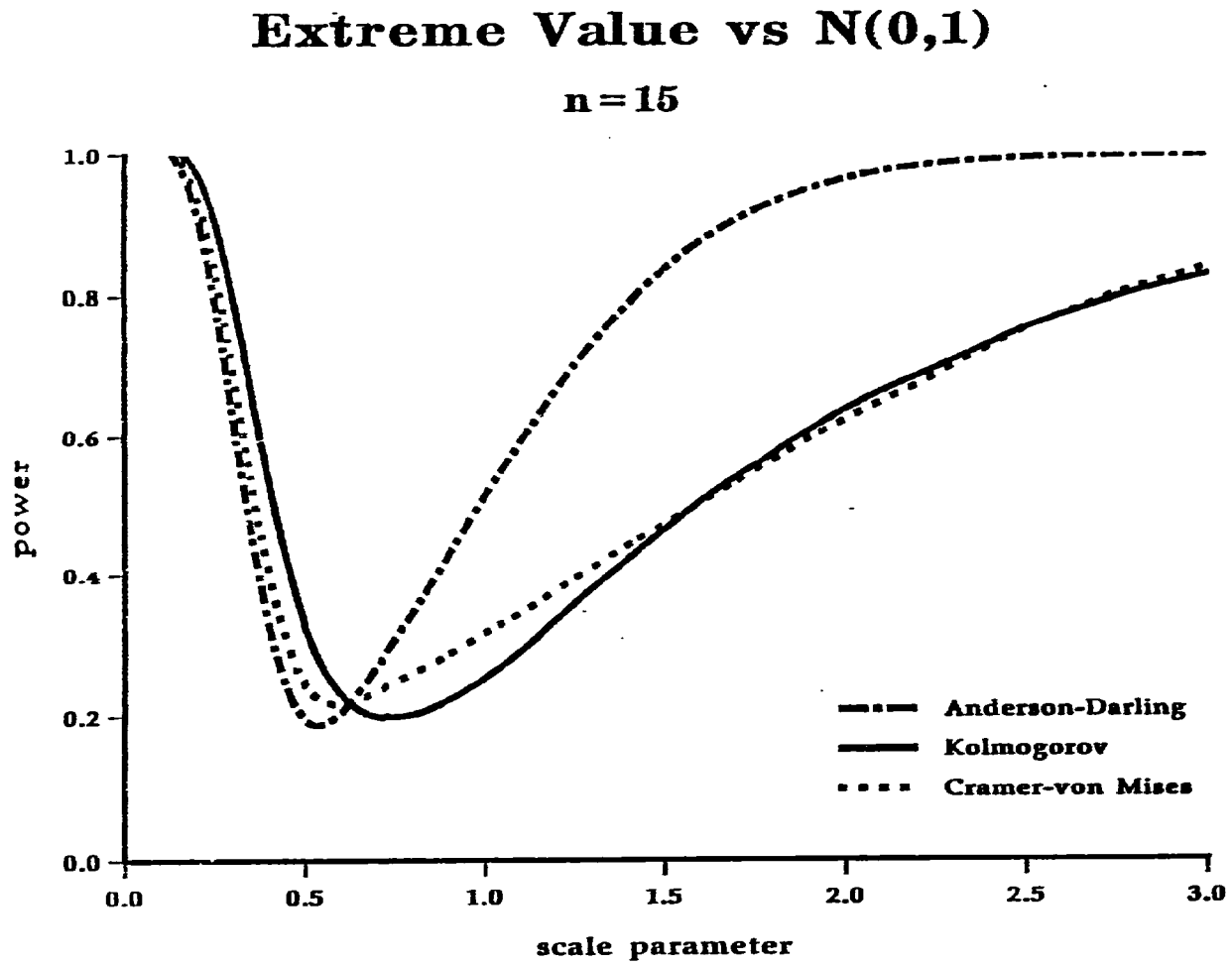


FIGURE 11. Power Curve for Extreme Value Generated Data, n = 15

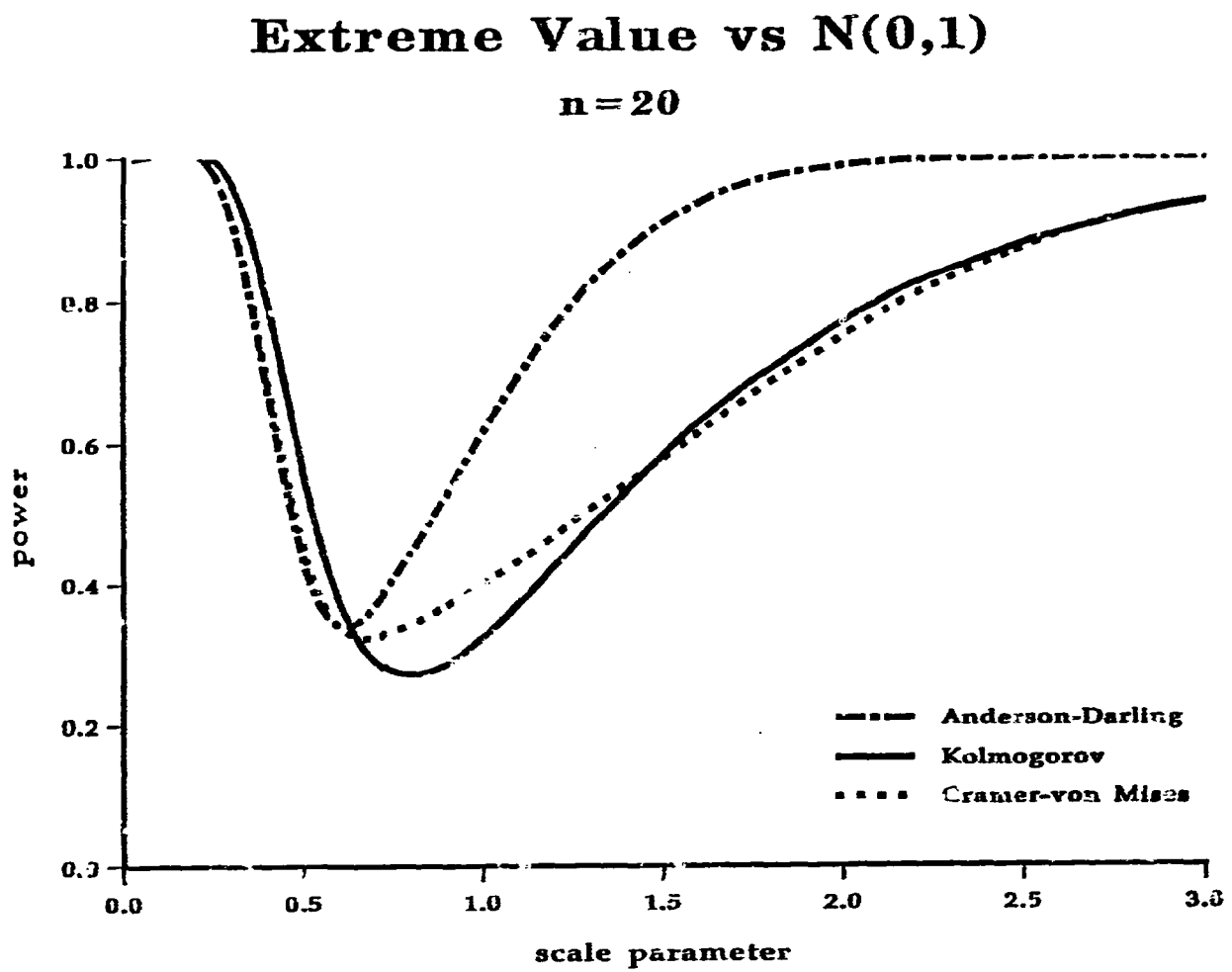


FIGURE 12. Power Curve for Extreme Value Generated Data, n = 20

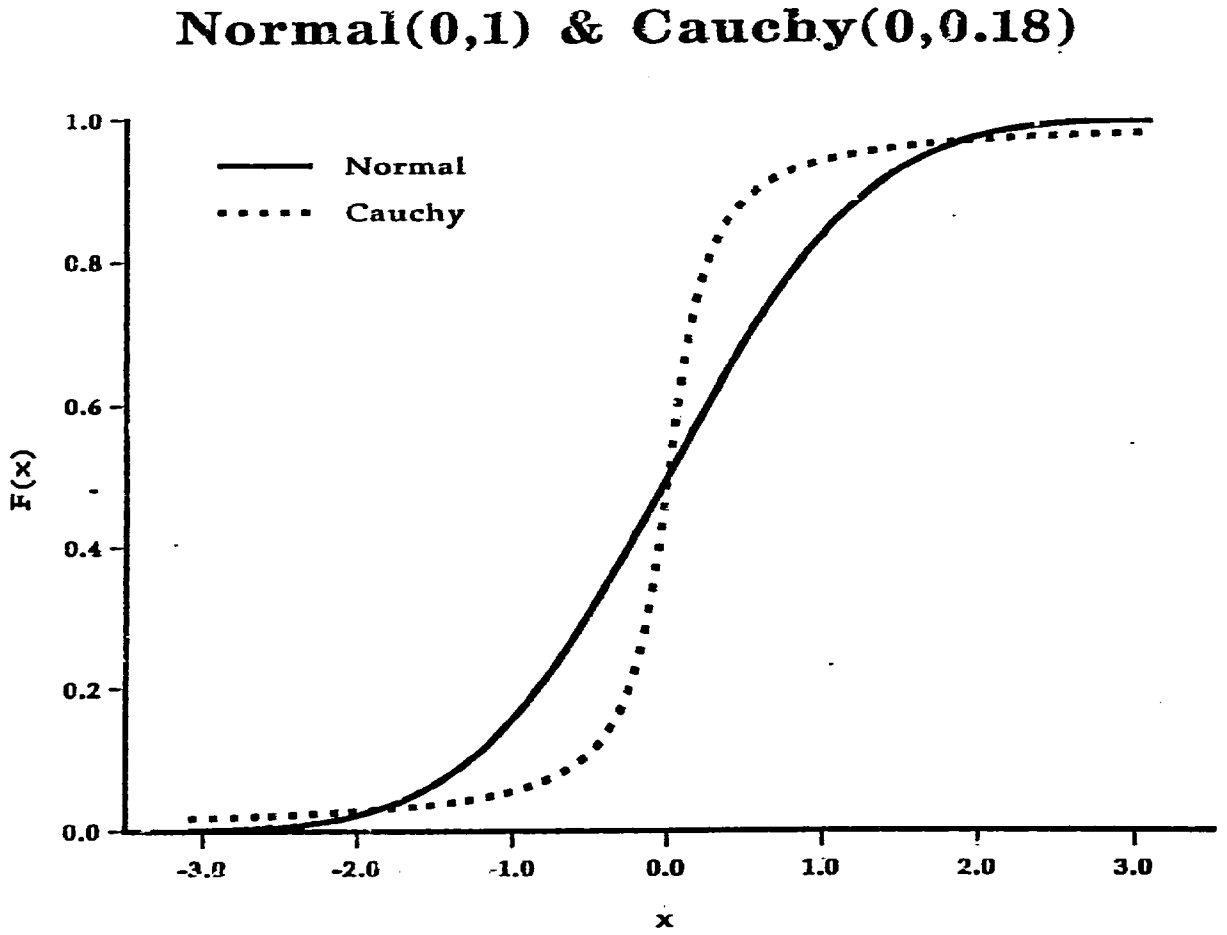


FIGURE 13. Cumulative Distribution Curves, Normal(0,1) & Cauchy(0,0.18)

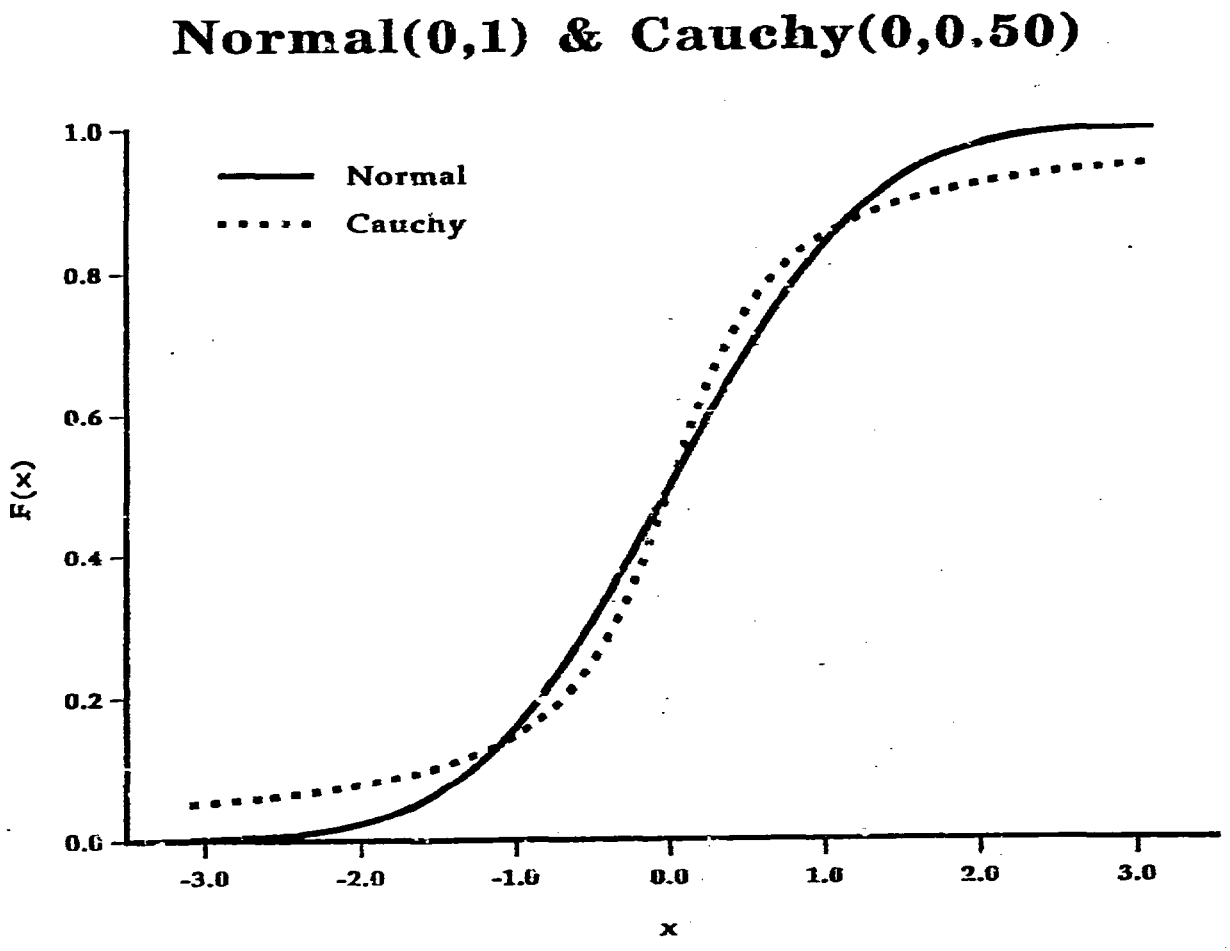


FIGURE 14. Cumulative Distribution Curves, Normal(0,1) & Cauchy(0,0.50)

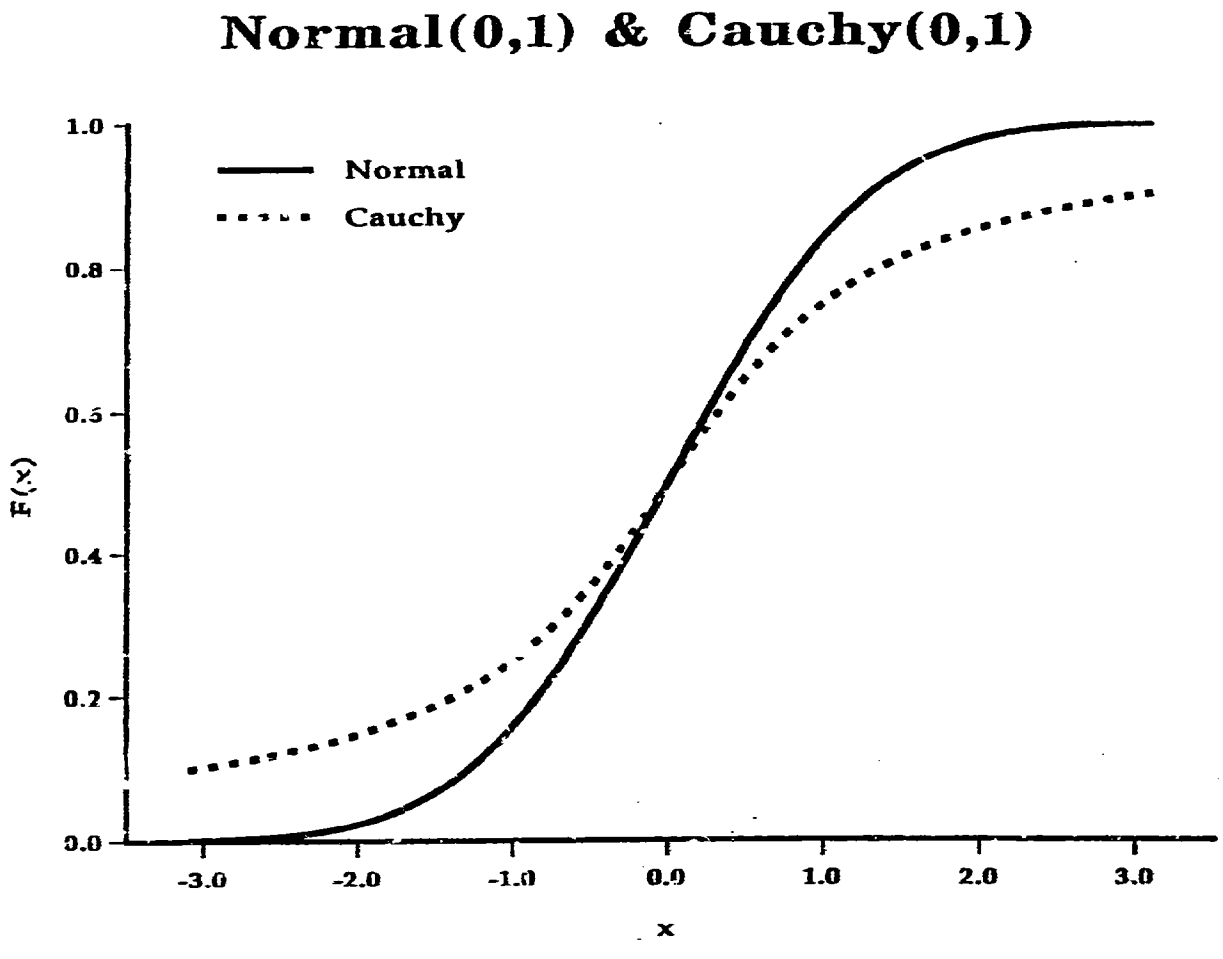


FIGURE 15. Cumulative Distribution Curves, Normal(0,1) & Cauchy(0,1)

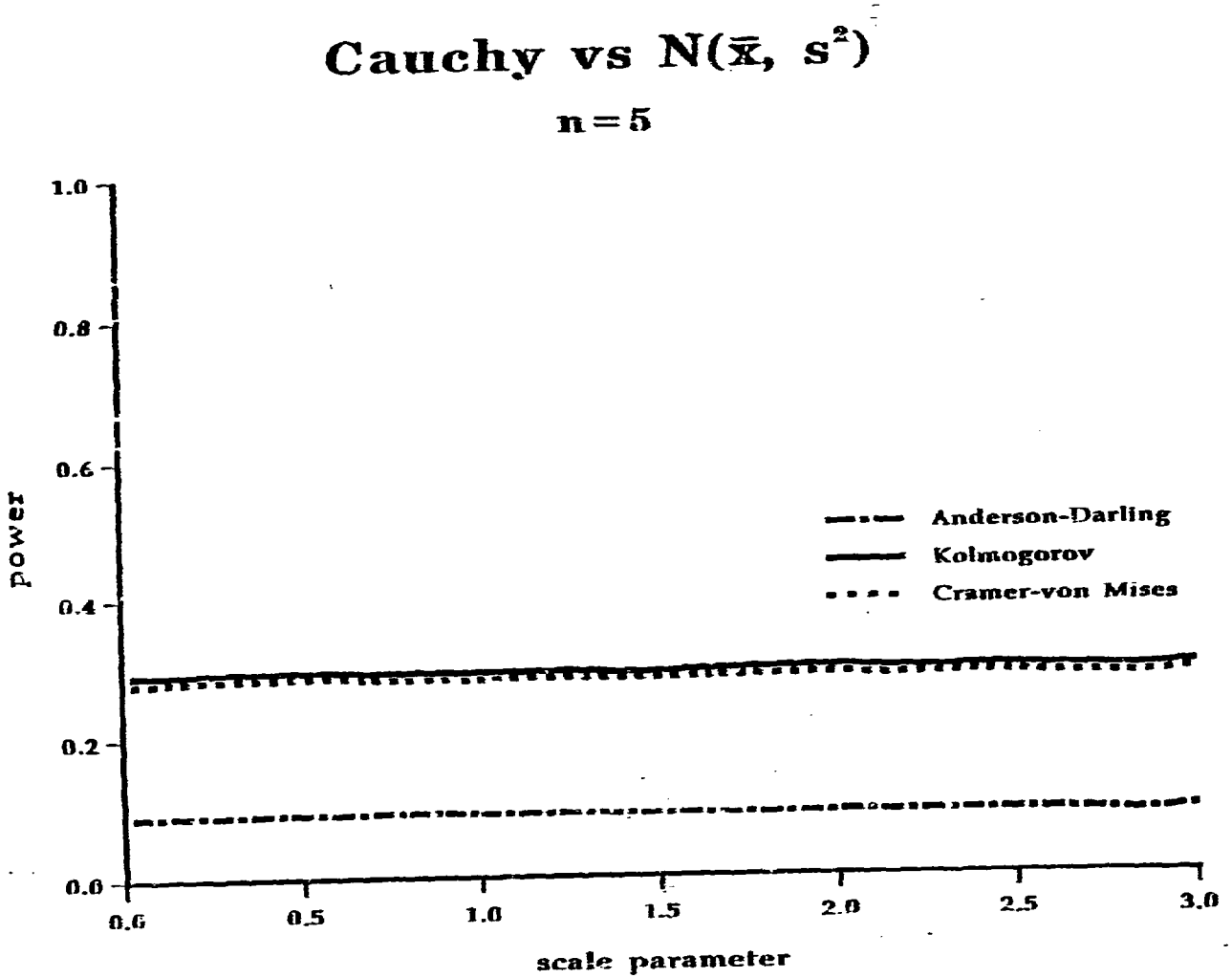


FIGURE 16. Power Curve for Cauchy Generated Data, n = 5

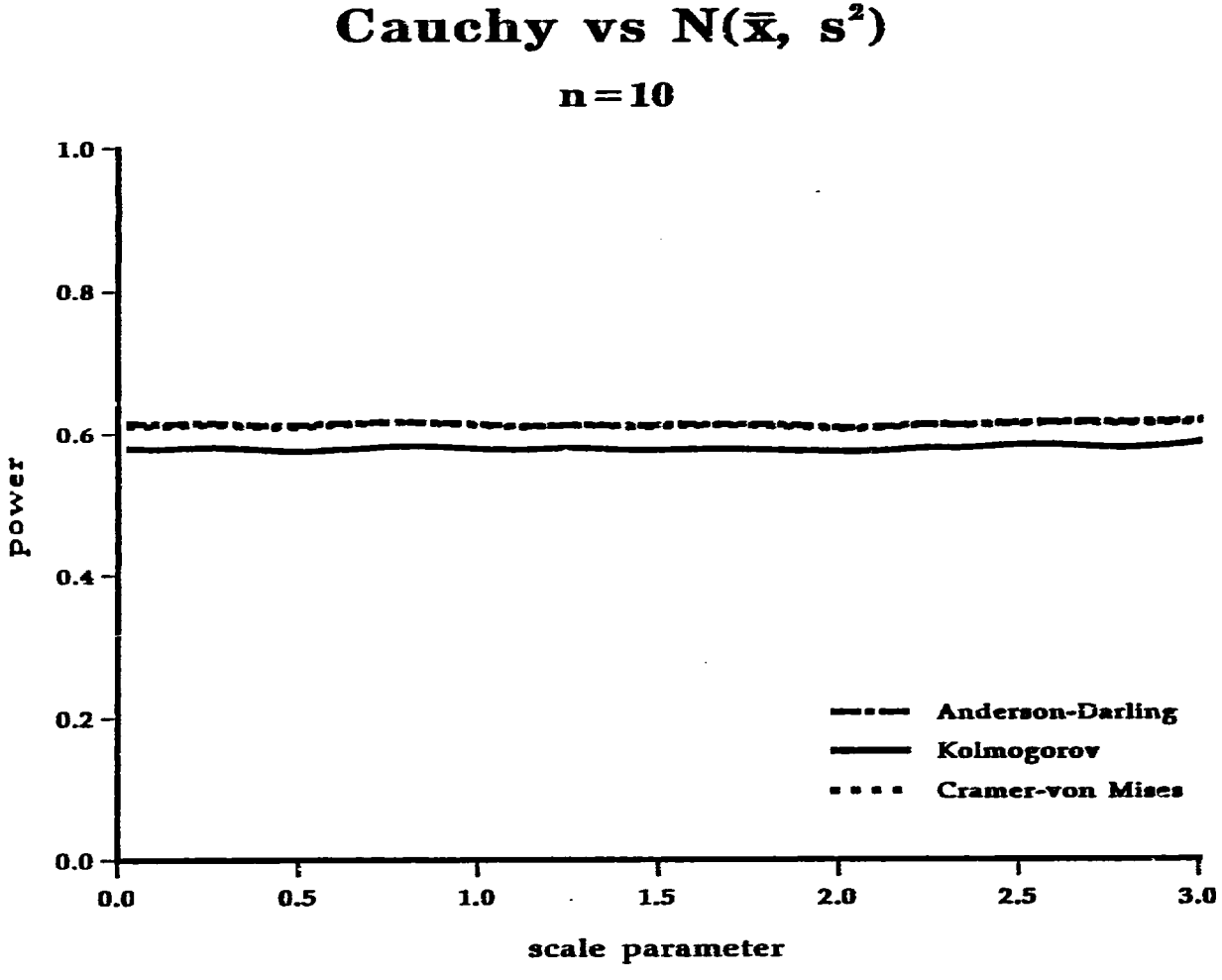


FIGURE 17. Power Curve for Cauchy Generated Data, n = 10

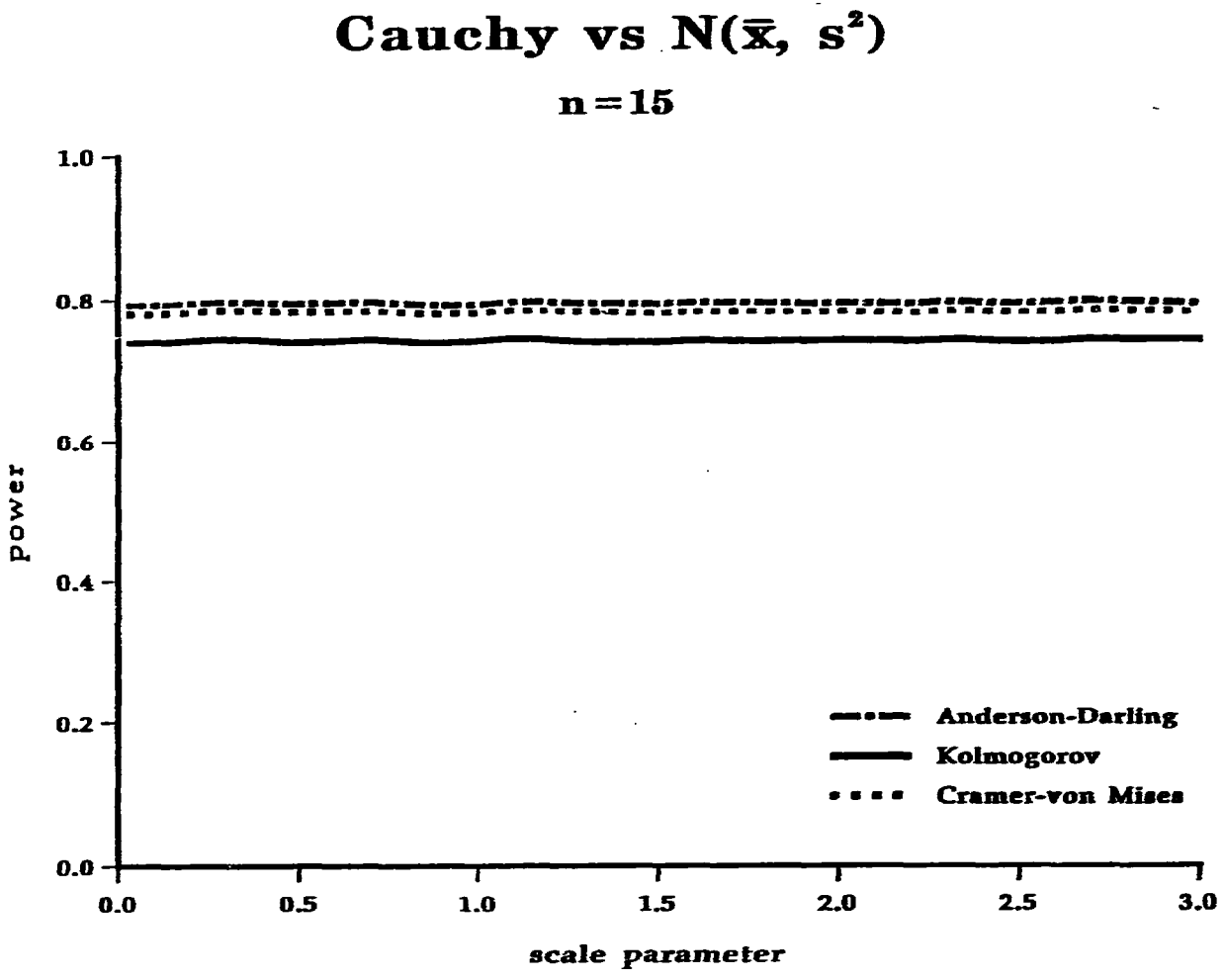


FIGURE 18. Power Curve for Cauchy Generated Data, n = 15

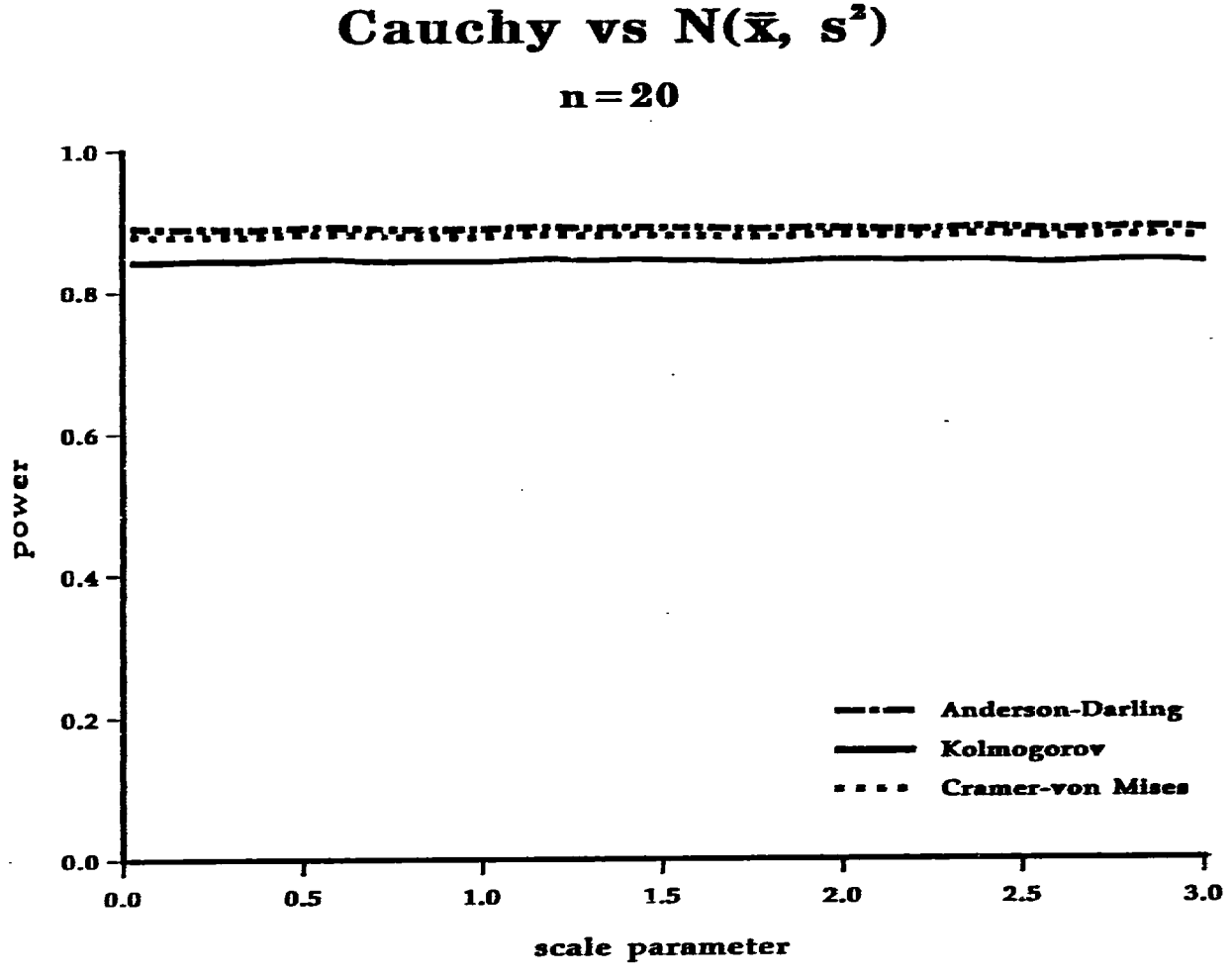


FIGURE 19. Power Curve for Cauchy Generated Data, n = 20

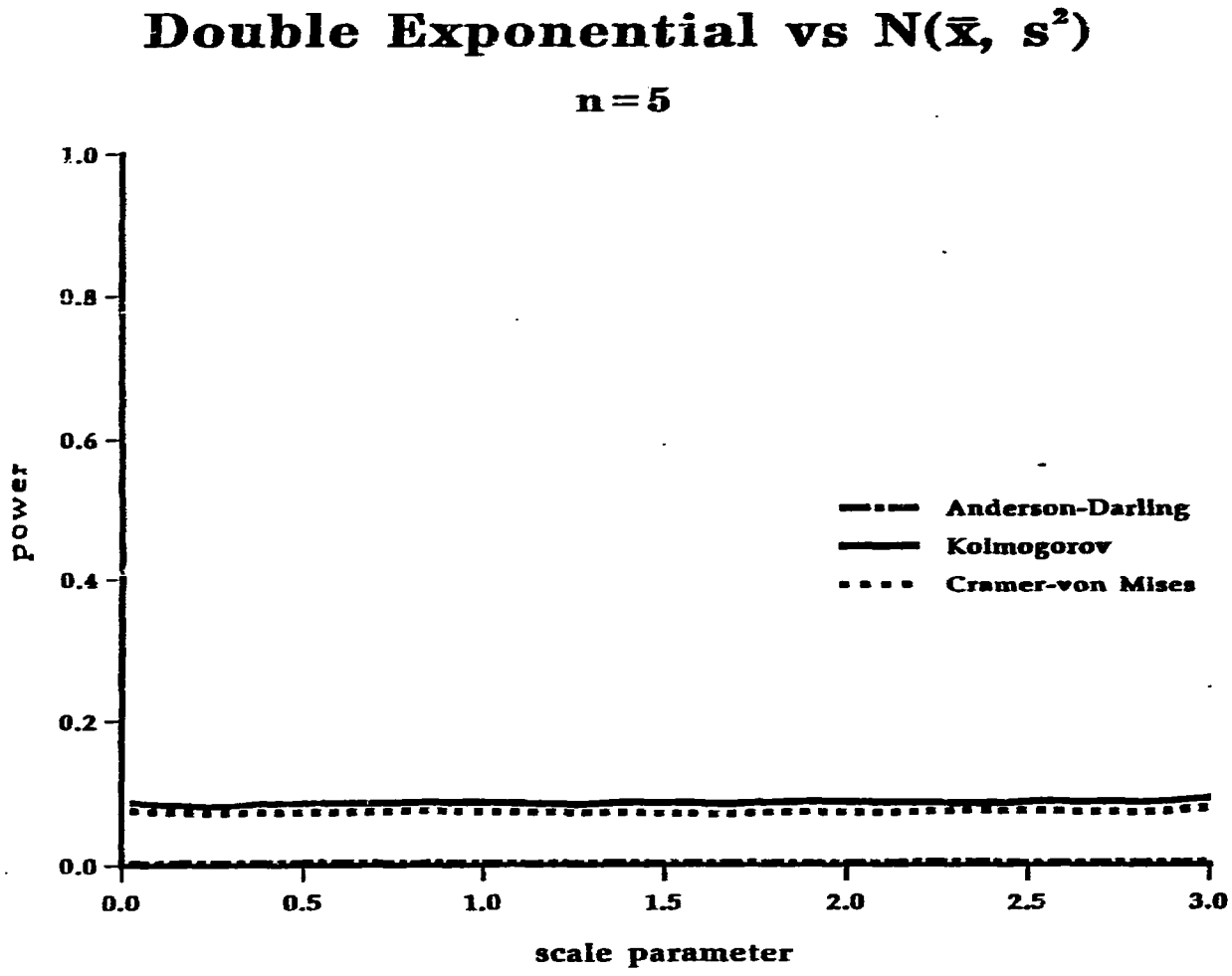


FIGURE 20. Power Curve for Double Exponential Generated Data, n = 5

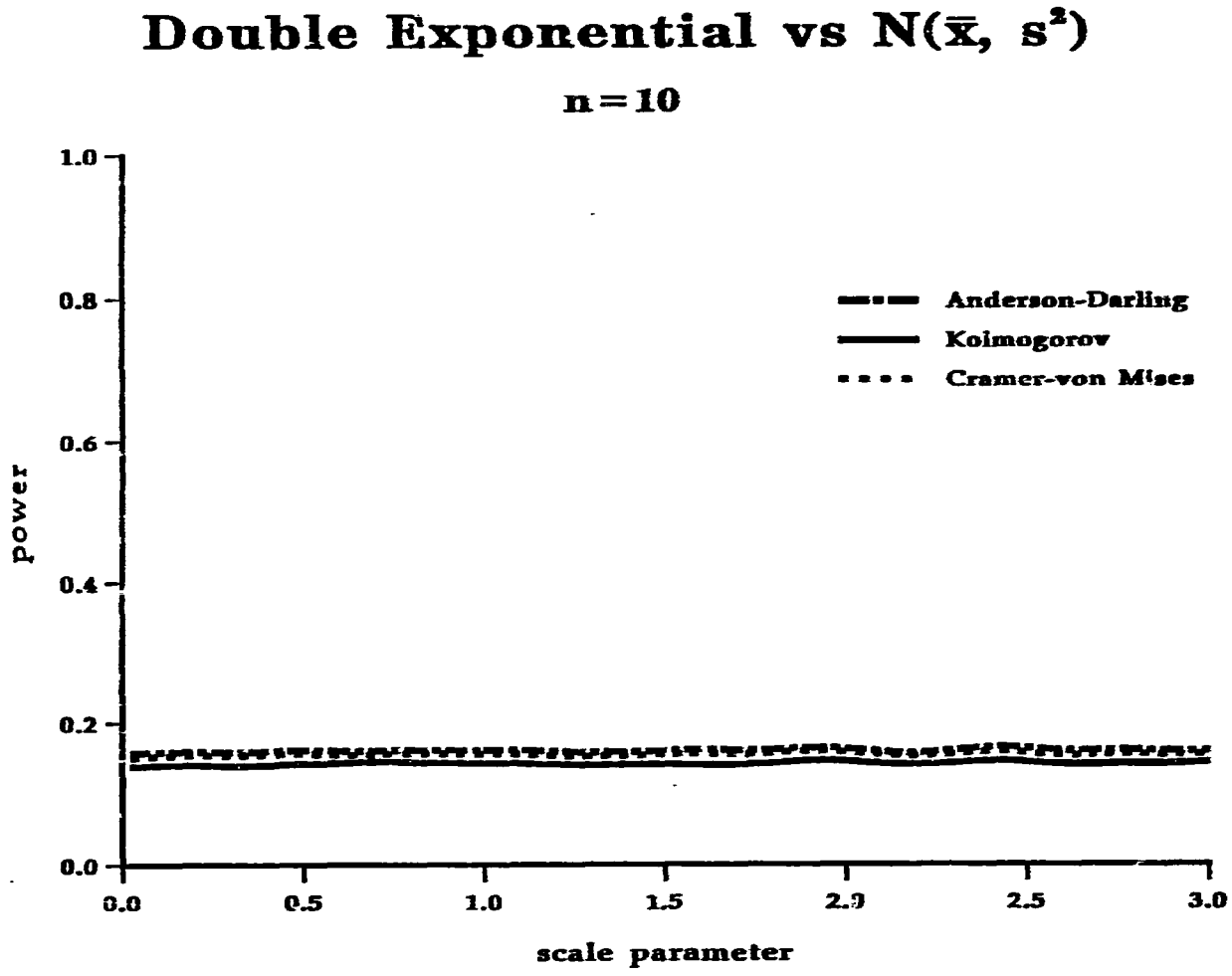


FIGURE 21. Power Curve for Double Exponential Generated Data, n = 10

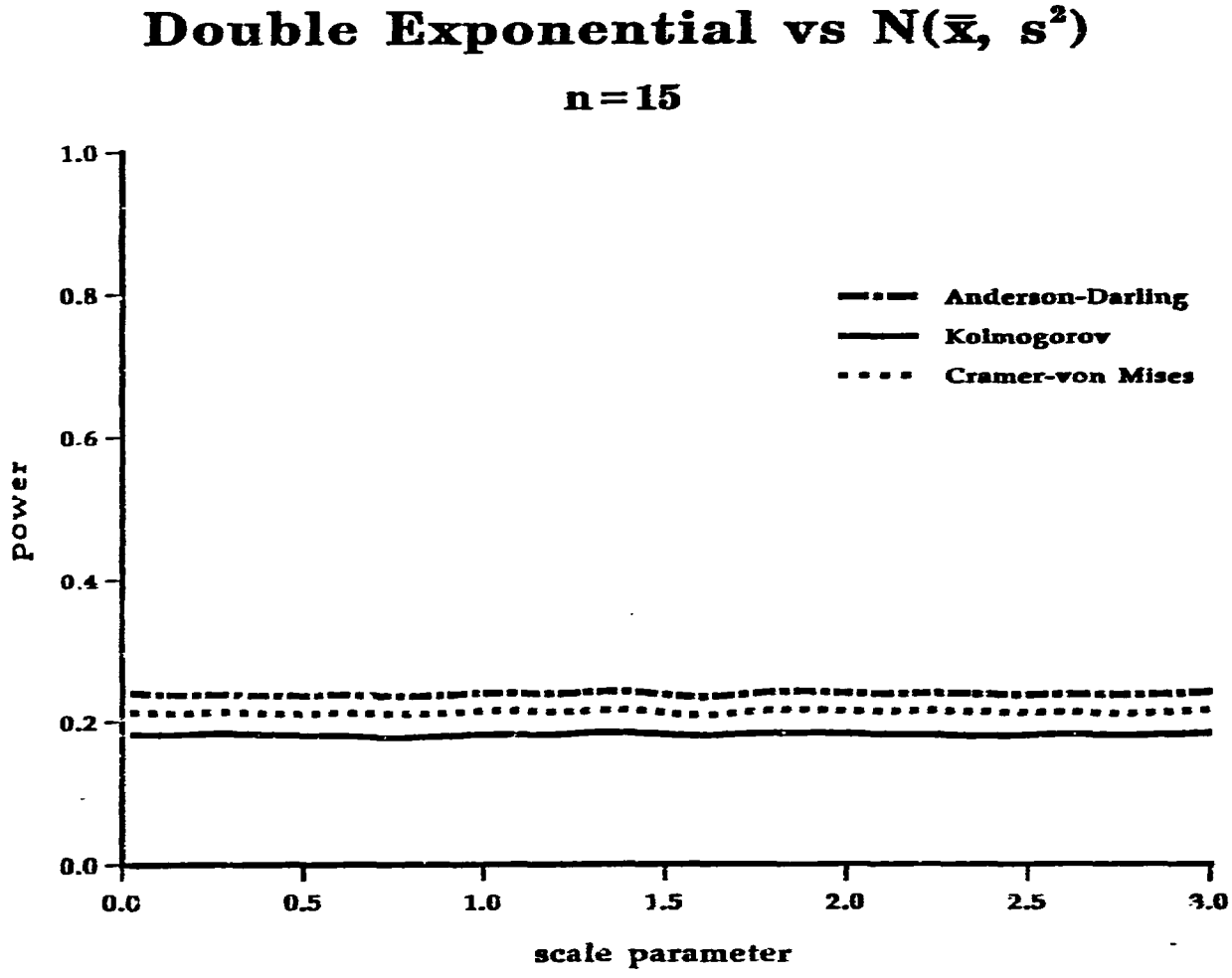


FIGURE 22. Power Curve for Double Exponential Generated Data, n = 15

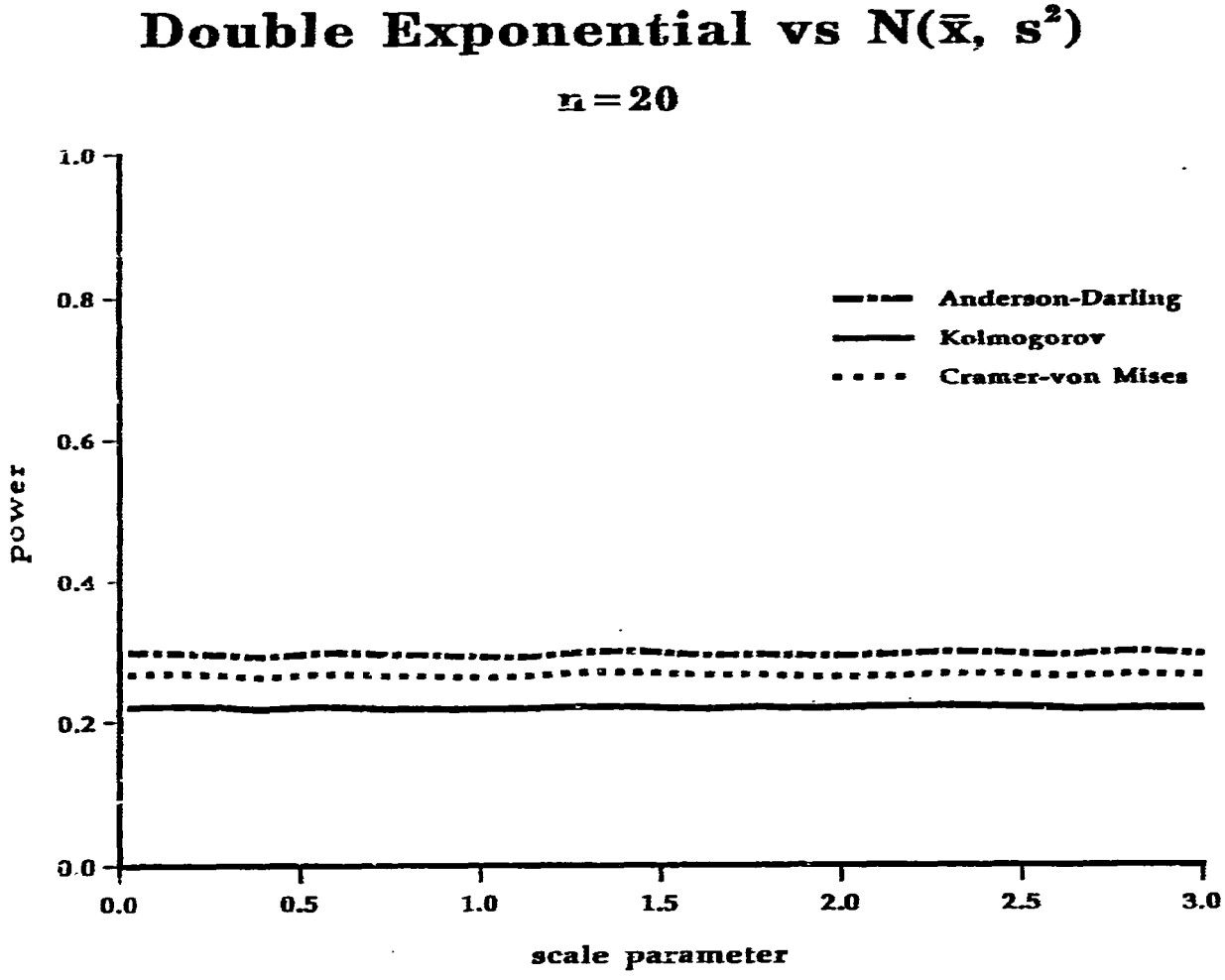


FIGURE 23. Power Curve for Double Exponential Generated Data, n = 20

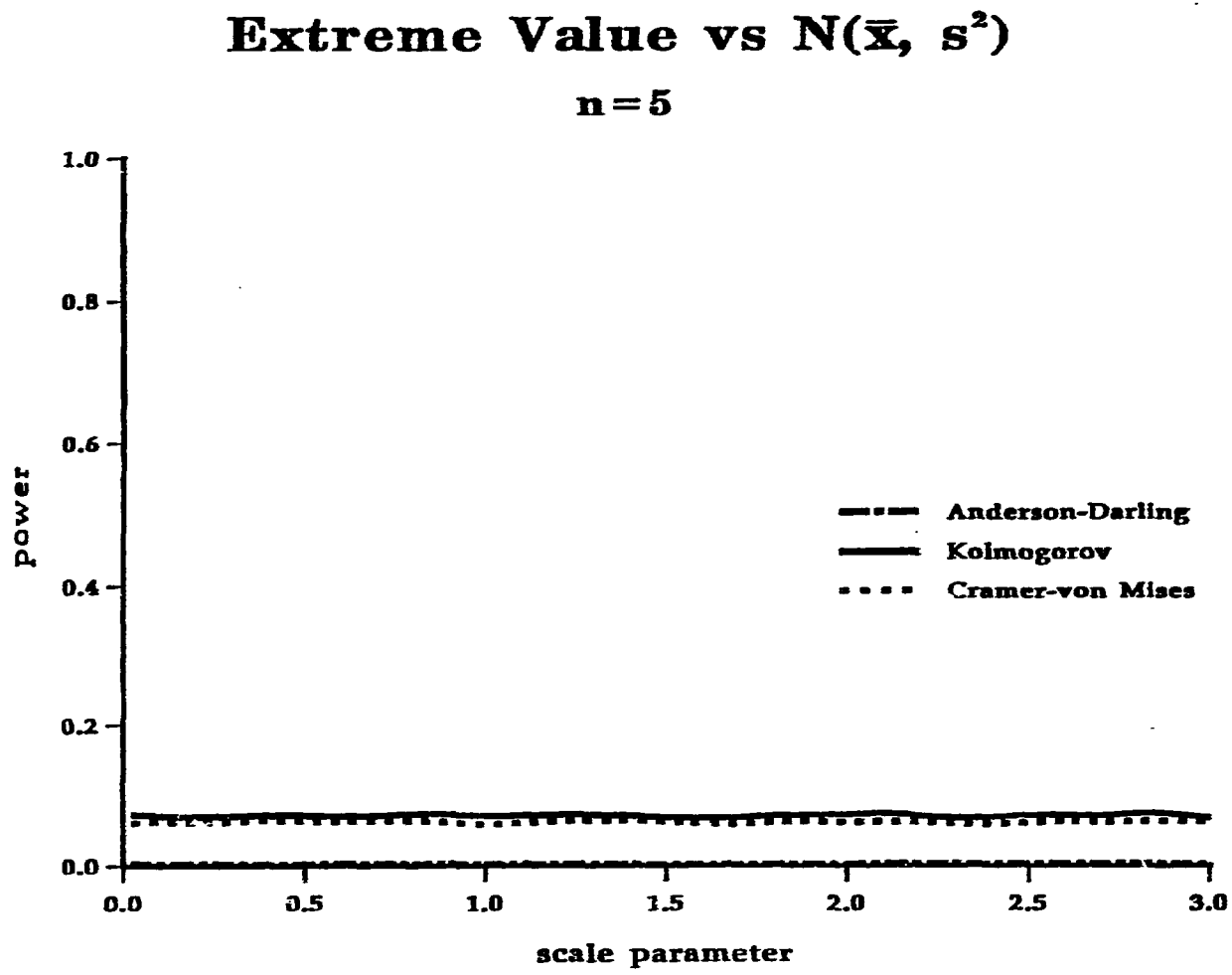


FIGURE 24. Power Curve for Extreme Value Generated Data, n = 5

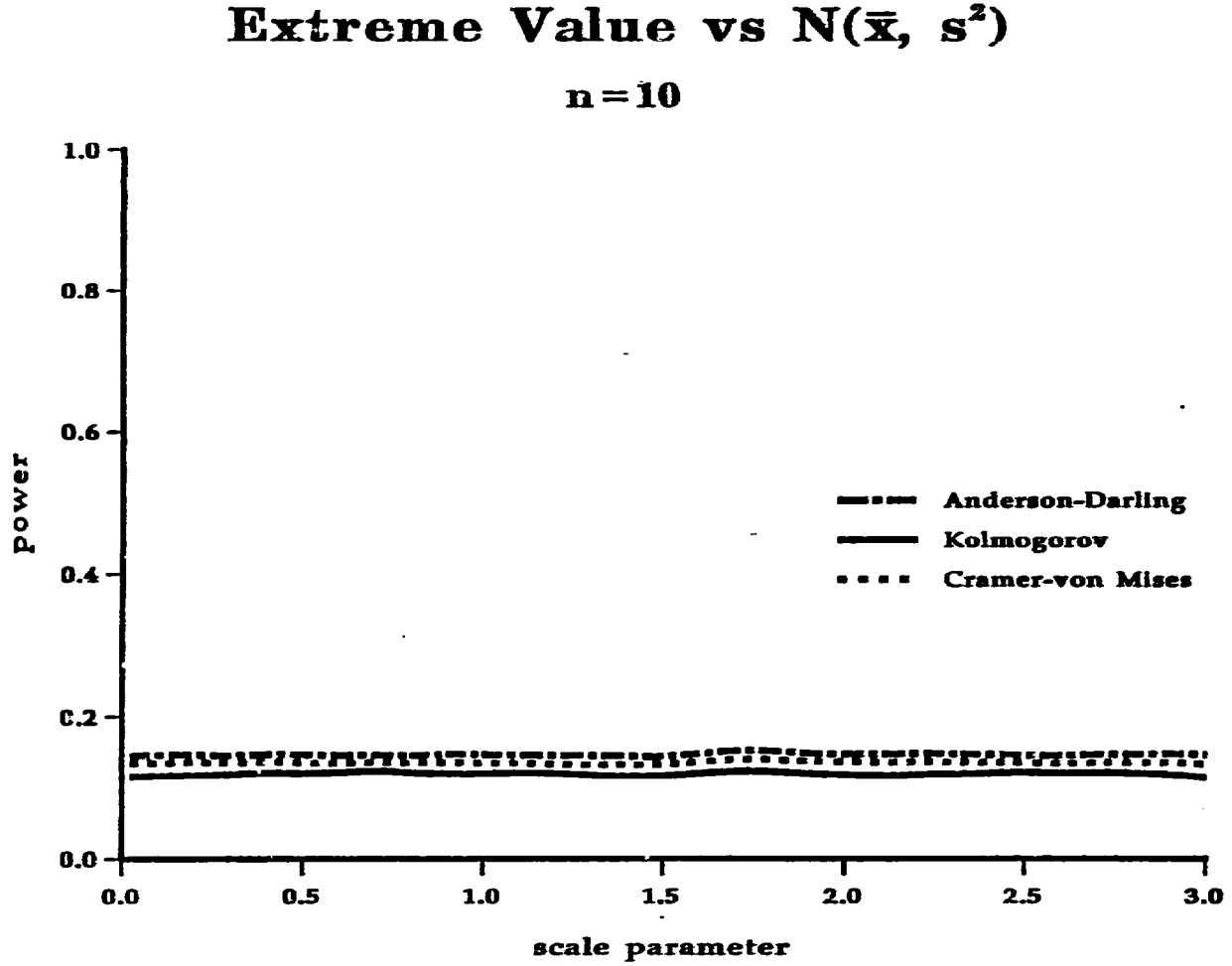


FIGURE 25. Power Curve for Extreme Value Generated Data, n = 10

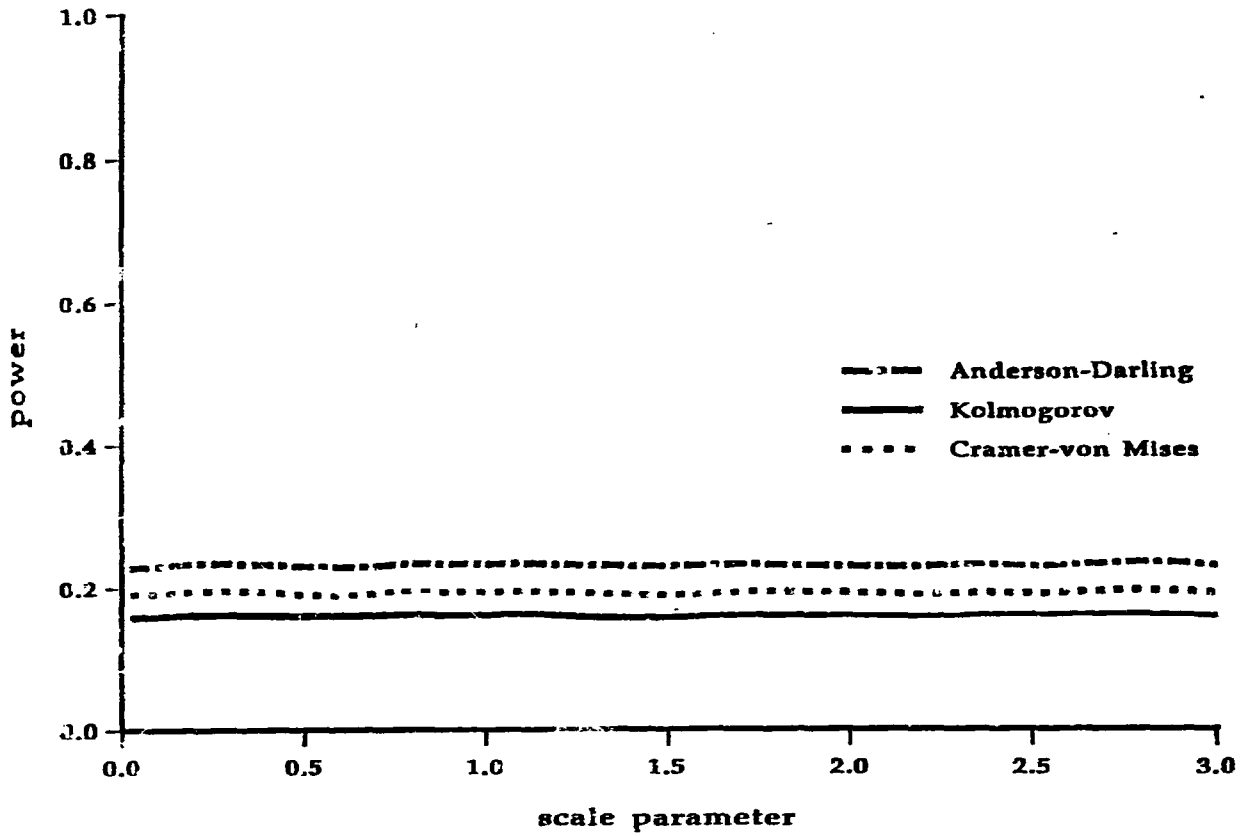


FIGURE 26. Power Curve for Extreme Value Generated Data, n = 15

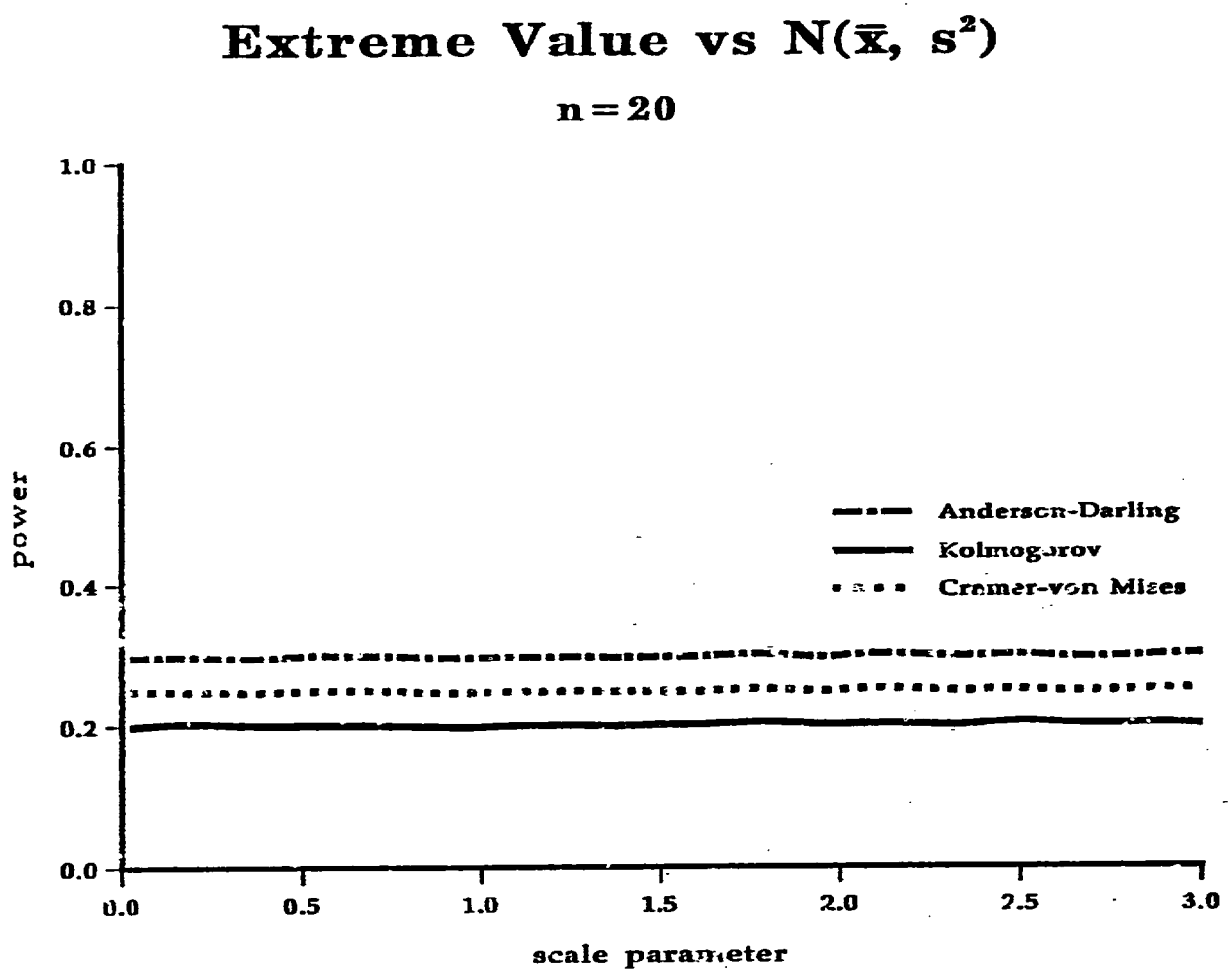


FIGURE 27. Power Curve for Extreme Value Generated Data, n = 20

REFERENCES

Anderson, T. W. and D. A. Darling. "Asymptotic Theory of Certain 'Goodness of Fit' Criteria Based on Stochastic Process," *Annals of Mathematical Statistics.* 1952, Vol. 23, pp. 193-212.

----------. "A Test of Goodness of Fit," *Journal of the American Statistical Association.* 1954, Vol. 49, pp. 765-769.

Cogan, Edward J. and Robert Z. Norman. *Handbood of Calculus, Difference and Differential Equations.* Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1958.

Conover, W. J. *Practical Nonparametric Statistics,* 2ed. New York: John Wiley & Sons, 1980.

Cramér, Harald. "On the composition of elementary errors," *Skandinavisk Aktuarietidskrift,* 1928, Vol. 11, pp. 13-74, 141-180.

Feller, William. *An Introduction to Probability Theory and its Application,* Volume II. New York: John Wiley and Sons, Inc., 1966.

Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Edited by Milton Abramowitz and Irene A. Stegun, National Bureau of Standards, Applied Mathematics Series 55, June 1964.

Harter, H. Leon. "Modified Asymptotic Formulas for Critical Values of the Kolmogorov Test Statistic," *The American Statistician.* May 1980, Vol. 34, No. 2, pp. 110-111.

Johnson, Norman L. and Samuel Kotz. *Continuous Univariate Distributions - 1, Distributions in Statistics.* New York: John Wiley & Sons. 1970.

Jolley, L. B. W. *Summation of Series.* 2nd Revised Edition. New York: Dover Publications, Inc., 1961.

Kernighan, Brain W. and Dennis M. Ritchie. *The C Programming Language.* New Jersey: Prentice-Hall, Inc., 1978.

Kolmogorov, A. N. "Sulla determinazione empirica delle leggi di probabilita," *Giornale dell' Istituto Italiano degli Attuari*, 1933, Vol. 4, pp 83-91.

Lewis, Peter A. W. "Distribution of the Anderson-Darling Statistic," *The Annals of Mathematical Statistics*, 1961, Vol. 32, pp. 1118-1124.

Miller, R. G., Jr. "Table of percentage points of Kolmogorov statistics," *Journal of the American Statistical Association*, 1956, Vol. 51, pp. 111-112.

Rubinstein, Reuven Y. *Simulation and the Monte Carlo Method.* New York: John Wiley & Sons, 1981.

Stephens, M. A. "EDF Statistics for Goodness of Fit and Some Comparisons," *Journal of the American Statistical Association.* September 1974, Vol. 69, No. 347, pp. 730-737.

----------. "Use of the Kolmogorov-Smirov, Cramér-Von Mises and Related Statistics without Extensive Tables," *Royal Statistical Society, Series B.* 1970, Vol. 32, No. 1, pp. 115-122.

---------- and Urs R. Maag. "Further percentage points for W_N^2," *Biometrika.* 1968, Vol. 55, pp. 428-430.

von Mises, R. *Wahrscheinlichkeitsrechnung und Ihre Anwendung in der Statistik und Theoretischen Physik.*, 1931, F. Deutioke, Leipzig.

APPENDIX A: EXPECTION OF SQUARED DISCREPANCY BETWEEN AN EMPIRICAL DISTRIBUTION FUNCTION AND A SPECIFIED DISTRIBUTION FUNCTION

To help find a suitable weighting function, $\psi[F(x)]$, lets look at $n\,E\left[F_n(x) - F(x)\right]^2$.

$$
\begin{aligned}
n\,E\left[F_n(x) - F(x)\right]^2 &= n\,E\left[F_n(x) - H(x) + H(x) - F(x)\right]^2 \\
&= n\,E\left[\left\{F_n(x) - H(x)\right\} - \left\{F(x) - H(x)\right\}\right]^2 \\
&= n\,E\left[\left\{F_n(x) - H(x)\right\}^2 - 2\left\{F_n(x) - H(x)\right\}\left\{F(x) - H(x)\right\}\right. \\
&\qquad \left. + \left\{F(x) - H(x)\right\}^2\right] \\
&= n\,E\left[F_n(x) - H(x)\right]^2 - 2\left\{F(x) - H(x)\right\}E\left[F_n(x) - H(x)\right] \\
&\qquad + \left\{F(x) - H(x)\right\}^2 \\
&= n\,E\left[F_n(x) - H(x)\right]^2 + \left\{F(x) - H(x)\right\}^2
\end{aligned}
$$

$$= n\,E\left[F_n(x)^2 - 2\,F_n(x)H(x) + H^2(x)\right] + \left\{F(x) - H(x)\right\}^2$$

$$= n\,E\left[F_n(x)^2\right] - 2\,H(x)\,E\left[F_n(x)\right] + H^2(x) + \left\{F(x) - H(x)\right\}^2$$

$$= n\left[\frac{1}{n}\,H(x)\{1 - H(x)\} + H^2(x) - 2\,H^2(x) + H^2(x) + \left\{F(x) - H(x)\right\}^2\right]$$

$$n\,E\left[F_n(x) - F(x)\right]^2 = n\left[\frac{1}{n}\left\{H(x)\{1 - H(x)\}\right\} + \left\{F(x) - H(x)\right\}^2\right]$$

Under the null hypothesis, i.e., H_o: $H(x) = F(x)$,

$$n\,E\left[F_n(x) - F(x)\right]^2 = F(x)[1 - F(x)].$$

APPENDIX B: EXPANSION AND INTEGRATION OF THE ANDERSON-DARLING STATISTIC

$$W_n^2 = n \int_{-\infty}^{\infty} \frac{[F_n(x) - F(x)]^2}{F(x)[1 - F(x)]}\, dF(x)$$

$$= n \left\{ \int_{-\infty}^{x_1} \frac{[F_n(x) - F(x)]^2}{F(x)[1 - F(x)]}\, dF(x) + \sum_{k=1}^{n-1} \int_{x_k}^{x_{k+1}} \frac{[F_n(x) - F(x)]^2}{F(x)[1 - F(x)]}\, dF(x) + \int_{x_n}^{\infty} \frac{[F_n(x) - F(x)]^2}{F(x)[1 - F(x)]}\, dF(x) \right\}$$

Let $F(x) = u$
$dF(x) = du$
$F(x_i) = u_i$

$$= n \left\{ \int_0^{u_1} \frac{[0 - u]^2}{u(1 - u)}\, du + \sum_{k=1}^{n-1} \int_{u_k}^{u_{k+1}} \frac{\left[\frac{k}{n} - u\right]^2}{u(1 - u)}\, du + \int_{u_n}^{1} \frac{(1 - u)^2}{u(1 - u)}\, du \right\}$$

$$= n \left\{ -u_1 - \ln(1 - u_1) + \sum_{k=1}^{n-1} \int_{u_k}^{u_{k+1}} \frac{\left[\frac{k^2}{n^2} - \frac{2k}{n} u + u^2\right]}{u(1 - u)}\, du + \left[-1 + u_n - \ln u_n\right] \right\}$$

$$= n \left\{ - u_1 - \ln (1 - u_1) + \sum_{k=1}^{n-1} \left[\frac{k^2}{n^2} \int_{u_k}^{u_{k+1}} \frac{du}{u (1 - u)} \right] \right.$$

$$+ \sum_{k=1}^{n-1} \left[- \frac{2k}{n} \int_{u_k}^{u_{k+1}} \frac{u}{u (1 - u)} \, du \right]$$

$$+ \sum_{k=1}^{n-1} \left[\int_{u_k}^{u_{k+1}} \frac{u^2}{u (1 - u)} \, du \right]$$

$$\left. + \left[-1 + u_n - \ln u_n \right] \right\}$$

$$= n \left\{ - u_1 - \ln (1 - u_1) \right.$$

$$+ \sum_{k=1}^{n-1} \frac{k^2}{n^2} \left[\ln (u_{k+1}) - \ln (1 - u_{k+1}) - \ln u_k + \ln (1 - u_k) \right]$$

$$+ \sum_{k=1}^{n-1} \frac{2k}{n} \left[\ln (1 - u_{k+1}) - \ln (1 - u_k) \right]$$

$$+ \sum_{k=1}^{n-1} \left[- u_{k+1} - \ln (1 - u_{k+1}) + u_k + \ln (1 - u_k) \right]$$

$$\left. + \left[- 1 + u_n - \ln u_n \right] \right\}$$

$$= n \left\{ - u_1 - \ln (1 - u_1) + \sum_{k=1}^{n-1} - \left(\frac{2k-1}{n^2} \right) \left[\ln u_k - \ln (1 - u_k) \right] \right.$$

$$+ \left(\frac{n-1}{n} \right)^2 \left[\ln u_n - \ln (1 - u_n) \right]$$

$$+ \sum_{k=1}^{n-1} \left[- \frac{2}{n} \ln (1 - u_k) \right]$$

$$+ \left(\frac{2n-2}{n} \right) \ln (1 - u_n)$$

$$+ u_1 + \ln (1 - u_1) - u_n - \ln (1 - u_n)$$

$$\left. + \left[-1 + u_n - \ln u_n \right] \right\}$$

$$= n \left\{ -1 - \sum_{k=1}^{n} \left[\left(\frac{2k-1}{n^2} \right) \ln u_k + \left(\frac{2(n-k)+1}{n^2} \right) \ln (1 - u_k) \right] \right\}$$

$$= -n - \frac{1}{n} \sum_{k=1}^{n} \left[(2k-1) \ln u_k + (2(n-k)+1) \ln (1 - u_k) \right] .$$

APPENDIX C: DERIVATION OF THE CRAMÉR-VON MISES STATISTIC

For an ordered sample $x_1 \leq x_2 \leq \cdots \leq x_n$ the empirical distribution function is defined as

$$F_n(x) = \begin{cases} 0 & x < x_1 \\ \dfrac{k}{n} & \text{for} \quad x_k \leq x < x_{k+1} \\ 1 & x_n \leq x \end{cases} .$$

The Cramér-von Mises statistic may be written

$$n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \, dF(x)$$

$$= n \left\{ \int_{-\infty}^{x_1} [F_n(x) - F(x)]^2 \, dF(x) + \sum_{k=1}^{n-1} \int_{x_k}^{x_{k+1}} [F_n(x) - F(x)]^2 \, dF(x) \right.$$

$$\left. + \int_{x_n}^{\infty} [F_n(x) - F(x)]^2 \, dF(x) \right\}$$

Let $F(x) = u$
$dF(x) = du$
$F(x_i) = u_i$

Then

$$n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 dF(x)$$

$$= n \left\{ \int_0^{u_1} \left[0 - u\right]^2 du + \sum_{k=1}^{n-1} \int_{u_k}^{u_{k+1}} \left[\frac{k}{n} - u\right]^2 du + \int_{u_n}^{1} \left[1 - u\right]^2 du \right\}$$

$$= n \left\{ \frac{1}{3} u_1^3 + \sum_{k=1}^{n-1} \left[\frac{k^2}{n^2} u - \frac{k}{n} u^2 + \frac{1}{3} u^3\right]_{u_k}^{u_{k+1}} + \left[\frac{1}{3} - u_n + u_n^2 - \frac{1}{3} u_n^3\right] \right.$$

$$= n \left\{ \frac{1}{3} u_1^3 + \frac{1}{n^2} \sum_{k=1}^{n-1} k^2 (u_{k+1} - u_k) - \frac{1}{n} \sum_{k=1}^{n-1} k \, (u_{k+1}^2 - u_k^2) \right.$$

$$\left. + \frac{1}{3} \sum_{k=1}^{n-1} (u_{k+1}^3 - u_k^3) + \left[\frac{1}{3} - u_n + u_n^2 - \frac{1}{3} u_n^3\right] \right.$$

$$= n \left\{ \frac{1}{3} u_1^3 + \frac{1}{n^2} \left[\sum_{k=1}^{n-1} \left(- (2k - 1) u_k\right) + (n - 1)^2 u_n\right] \right.$$

$$- \frac{1}{n} \left[- \sum_{k=1}^{n-1} (u_k^2) + (n - 1) u_n^2 \right]$$

$$\left. \left. + \frac{1}{3} u_n^3 - \frac{1}{3} u_1^3 + \left[\frac{1}{3} - u_n + u_n^2 - \frac{1}{3} u_n^3\right] \right\} \right.$$

$$= n \left\{ \sum_{k=1}^{n-1} \left[\frac{1}{n} u_k^2 - \left(\frac{2k - 1}{n^2}\right) u_k\right] + \frac{1}{3} + \frac{1}{n} u_n^2 - \left(\frac{2n - 1}{n^2}\right) u_n \right\}$$

$$= \sum_{k=1}^{n} \left[u_k^2 - \left(\frac{2k - 1}{n}\right) u_k\right] + \frac{n}{3}$$

Completing the square,

$$= \sum_{k=1}^{n} \left[u_k - \left(\frac{2k-1}{2n} \right) \right]^2 + \frac{n}{3} - \sum_{k=1}^{n} \left(\frac{2k-1}{2n} \right)^2$$

$$= \sum_{k=1}^{n} \left[u_k - \left(\frac{2k-1}{2n} \right) \right]^2 + \frac{n}{3} - \frac{1}{4n^2} \left(1^2 + 3^2 + \cdots + (2n-1)^2 \right)$$

$$= \sum_{k=1}^{n} \left[u_k - \left(\frac{2k-1}{2n} \right) \right]^2 + \frac{n}{3} - \frac{1}{4n^2} \left(\frac{(n)(4n^2-1)}{3} \right)$$

$$= \sum_{k=1}^{n} \left[u_k - \left(\frac{2k-1}{2n} \right) \right]^2 + \frac{1}{12n} .$$

Nonpare, a Consultation System for Analysis of Data

J.C. Dumer III
T.P. Hanratty
M.S. Taylor

US Army Ballistic Research Laboratory
ATTN: SLCBR-SE-P
Aberdeen Proving Ground, MD 21005-5066

Abstract. Nonpare, a consultation system for analysis of data using nonparametric statistical procedures, is under active development. It is intended to serve as an intelligent interface that will act as a guide, an instructor, and an interpreter to a body of statistical software. Nonpare exists as a prototype, with a limited release planned in 1989 for field testing.

1. Introduction

Statistical software packages, to large extent, accept any properly configured data set and proceed to process it. Few if any checks are made to ensure the adequacy of the data and the suitability of the analysis, and little is done to provide an explanation or interpretation of the results. This requires a great deal from the user. Declining computation costs, together with increased availability of computers and proliferation of statistical software, has further enhanced the opportunity for faulty data analysis. Application of expert system techniques from artificial intelligence to produce more cognizant software is one approach to reversing this unfortunate trend.

In 1985, a workshop sponsored by AT&T Bell Laboratories brought together many of the active investigators in artificial intelligence and statistics and was the genesis of a book by the same title edited by Gale [1]. This reference is in essence the proceedings of the workshop; but the papers given there, some with extensive bibliographies, provide the most complete centrally-located account of research in this topic to date.

This report details an effort underway at the US Army Ballistic Research Laboratory (BRL) to develop a consultation system for analysis of data using nonparametric statistical procedures. The system, called Nonpare, is intended to serve as an intelligent interface that will act as a guide, an instructor, and an interpreter to a body of statistical software. The system is currently a prototype, with a first release planned for 1989 for field testing.

2. Nonpare

Nonparametric statistics is too large an area to hope to encompass at once, especially if the entire field of mathematical statistics is partitioned into parametric and nonparametric procedures. The common-sense approach to construction of consultation systems suggests limiting the domain of application, but nonparametric statistics has qualities that make it strongly appealing.

Nonparametric data analysis is characterized chiefly by the absence of restrictive distribution assumptions – notably freedom from dependence on the normal (Gaussian) distribution. Many nonparametric statistical procedures are exact rather than approximate for small data sets, and they are the only confirmatory procedures which can be used to analyze data collected on a nominal or an ordinal scale of measurement. For these and other compelling reasons advanced, for example, by Conover, [2] Hollander and Wolfe, [3] and Lehmann, [4] nonparametric procedures find use in a wide variety of disciplines.

2.1 The System Structure

Nonpare uses Genie, an expert system shell developed at the BRL, [5] to provide a frame-based production system with forward and backward inferencing as well as an explanation facility that allows the user to interrogate the system – what hypotheses are being entertained, what rules are being verified, what facts are in evidence. Genie was chosen over commercial expert system shells for the research and development of Nonpare because of its accessibility for modification.

Nonpare, shown schematically in figure 1, consists of three subsystems in addition to Genie.

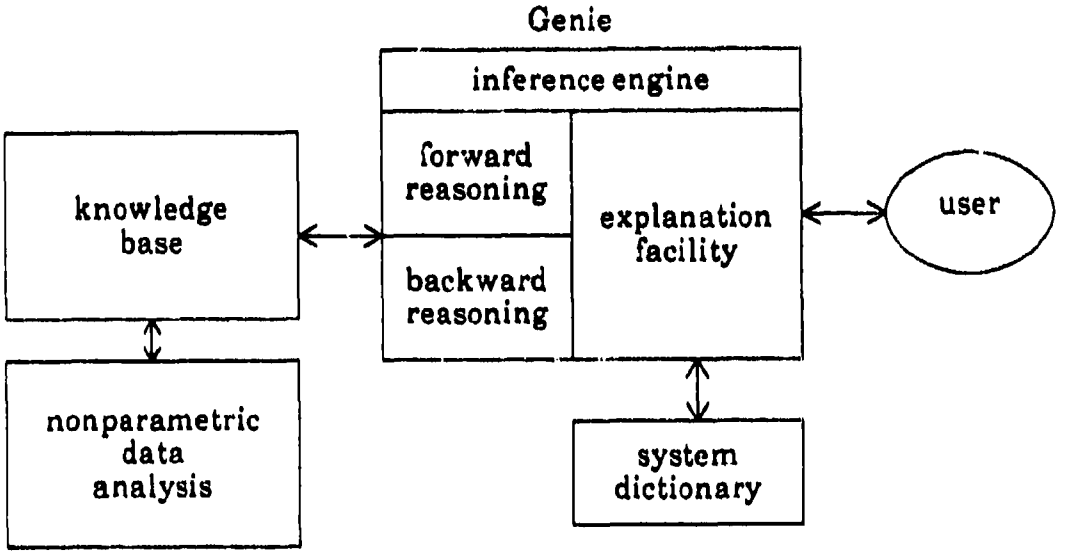


Figure 1. Nonpare system overview.

The system dictionary is a facility whose purpose is to provide on-line explanation of statistical jargon that may appear during the interactive dialog between Nonpare and the user. Expert domain knowledge, codified in English-like rules, resides in the knowledge base. Once an appropriate procedure(s) has been identified, the data are analyzed and the results explained by the nonparametric data analysis component. Graphics is used to summarize the data and enhance the explanation. In total, the user is led within system limitations to an appropriate statistical procedure through an interactive process in which the user is questioned and can in turn question the consultation system. Nonpare is written in Interlisp-D and currently runs on Xerox 1100 Series Lisp machines.

3. An Illustrative Session

Following the dictum of American educator John Dewey (1859-1952) that "We learn by doing," a detailed session with Nonpare follows, in which the main system features are illustrated.

Example 3.1

Suppose that a ballistician needs to assess the effectiveness of a newly designed kinetic energy penetrator against a specific armor plate. In particular, the experimenter would like to establish whether the probability of perforation exceeds .80 a level already attained with existing technology. Fourteen rounds are fired, and [p]erforation and [n]onperforation recorded to obtain: n, p, p, p, n, p, p, p, p, n, p, p,

p, p. Is the Pr{perforation} > .80 ?

(A diversion here. Searching for a statistical procedure with a set of data already collected is precisely how *not* to proceed. The purpose for collecting the data should first be established, and then the statistical tools available to support this purpose determined. Then the collection and analysis of data can proceed in an informed manner. Lamentably, the methodology-search scenario is enacted over and over again; so this example is not too contrived.)

It should be apparent from the onset that the question regarding Pr{perforation} > .80 can never be answered unequivocally yes or no, but only with some degree of qualification.

Nonpare presently has nineteen distinct data analysis procedures at its disposal; the number continues to increase. No assumptions have been made about their frequency of use; one procedure has not been declared most likely to be exercised, a second procedure next most likely, and so on, since the base of potential users is so broad. For the user, this means that any procedure is a likely starting point, as in this session, the dialog of which begins in figure 2. In the remainder of this section, the conventions that boldface denotes system prompts and brackets contain user input will be adopted. An occasional system response may be italicized but should not be confusing within the context of its appearance.

```
Do you have a sample X1, ..., Xn? y
Are you interested in whether the data conform to a specified
distribution? n
Are you interested in the probability of occurrence of a particular
category or event? y
Enter the name of the category of interest -> perforation
Are the n trials producing the values X1, ..., Xn independent? ^
```

Figure 2. Beginning dialog with Nonpare.

The session begins with a question about the configuration of the data.

Do you have a sample X_1, ..., X_n ? The data, n, ..., p, look like X_1, ..., X_n; respond [y]es.

Are you interested in whether the data conform to a specified distribution ? Nonpare is investigating a possible goodness-of-fit situation. A statistician, anticipating an approach to this problem, might find a [y]es response is appropriate here. A nonstatistician, for whom this portion of the system is designed, and who is interested in whether Pr{perforation} > .80, should respond [n]o, as indicated.

Are you interested in the probability of occurrence of a particular category or event ? [y]es. The user is interested in the probability of occurrence of a perforation.

Enter the name of the category of interest. [perforation]. Domain-dependent terminology is being introduced.

Are the n trials producing the values X_1, ..., X_n independent ? Suppose the user is unsure of the technical implications of the term "independent." An acceptable response is [What is independent]– as shown in figure 3.

```
Are the n trials producing the values X1, ..., Xn independent? what is in
dependent

Independence relates to freedom from external influence or control- here,
  the reference is to measurements (data) being free to assume values
without regard to other measurements that may be made.

Are the n trials producing the values X1, ..., Xn independent? y
Does each trial have the same probability p of producing the perforation
? y
Are you interested in considering whether the probability of occurrence
of the perforation equals or is bounded by some specified value p*? ^
```

Figure 3. A call to the system dictionary.

Independence relates to freedom from external influence or control – here, the reference is to measurements (data) being free to assume values without regard to other measurements that may be made.

This illustrates a dilemma for the subject area specialist. It may be impossible to rigorously define a term without reliance upon other terms that are equally obscure to a user with only a modest statistical background. This is the case here, where independence is bound to basic concepts of probability theory. Nonpare's response conveys the notion, but regrettably not the substance, of independence. More work is needed here. For now, assume the experimenter has collected a set of independent data.

Are the n trials producing the values X_1, ..., X_n independent ? [y]es.

Does each trial have the same probability p of producing the perforation ? [y]es. Notice that Nonpare is now using language the user provided, when it talks about probability of perforation.

Are you interested in considering whether the probability of occurrence of the perforation equals or is bounded by some specified value p* ? [y]es. The user is interested in the inequality Pr{perforation} > .80. After a [y]es response, the system suggests a possible approach, shown in figure 4.

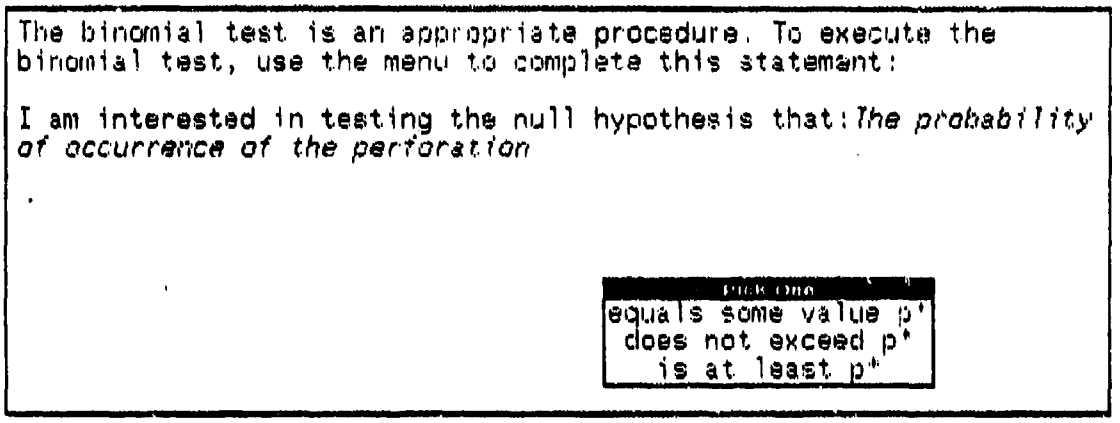


Figure 4. A call to the nonparametric data analysis subsystem.

The menu allows the user to select either a two-sided or one-sided test of hypothesis and is a potential source of error. Beginning statistics students, not realizing that a null (or empty) hypothesis is chosen to be rejected, might mistakenly choose **is at least p*** at this juncture. Here again, *some* level of statistical competence is required. Selecting the hypothesis **does not exceed p*** from the menu using a mouse, the user obtains for confirmation (figure 5) the statement:

I am interested in testing the null hypothesis that: ***The probability of occurrence of the perforation does not exceed p*.***

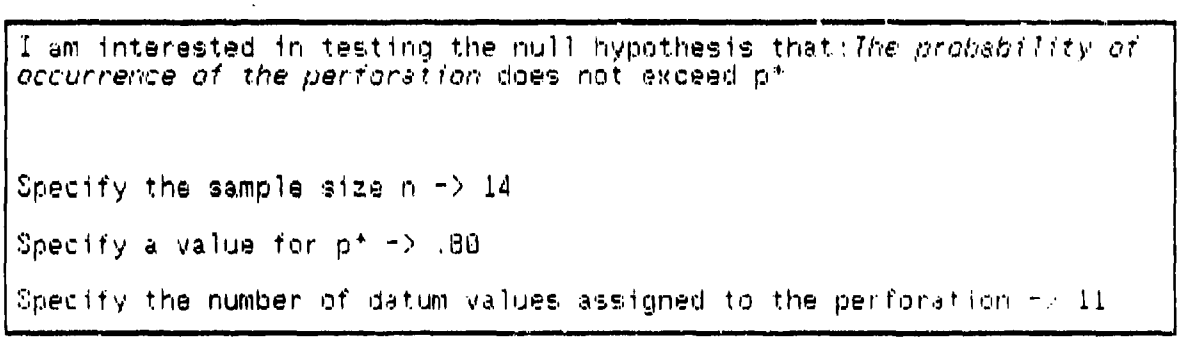


Figure 5. Hypothesis confirmation and input parameter declaration.

Specify the sample size n. [14]

Specify a value for p*. [.80]

Specify the number of datum values assigned to the perforation. [11]

The first two "Specify ... " commands determine the appropriate binomial distribution; the third determines the size of the critical region for the statistical procedure, which is explained in figure 7, following the system-generated histogram shown in figure 6.

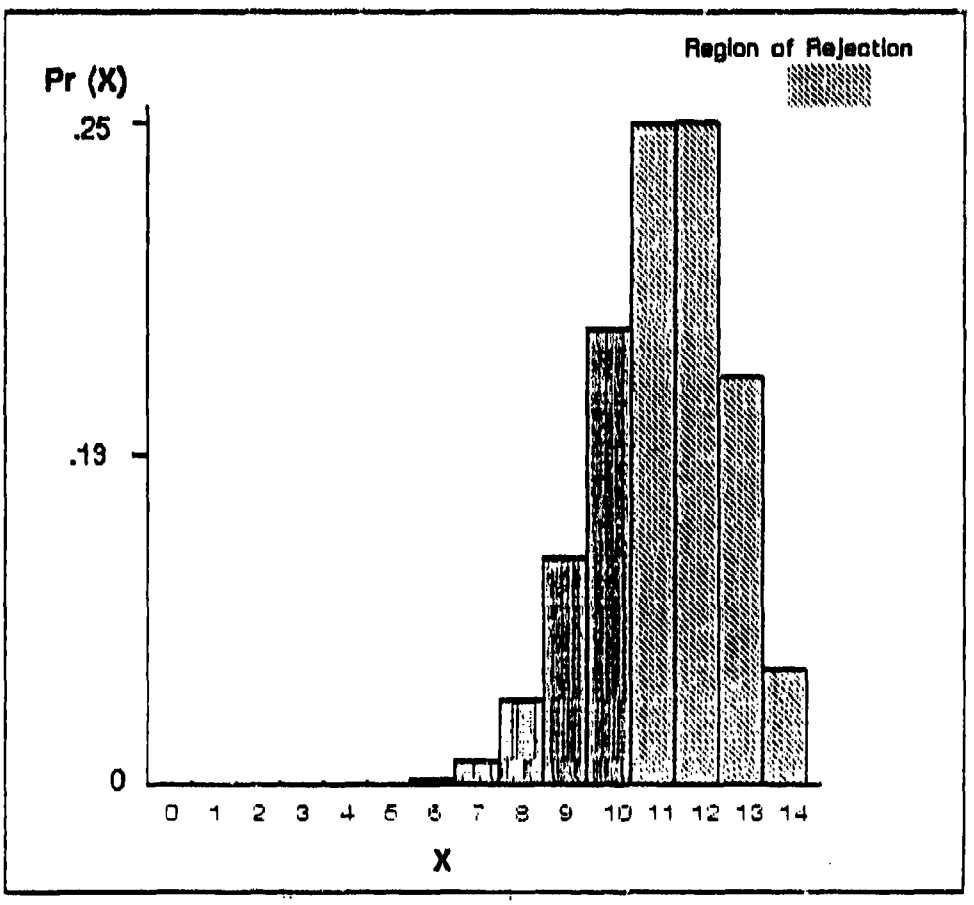


Figure 6. Statistical graph summary.

The histogram displays the probability of observing exactly n (n=0, ..., 14) armor perforations in fourteen shots if the true (but unknown) Pr{perforation} = .80. A statistician will readily assimilate this graph. If the user merely looks at it as a plot involving n rounds in which the light gray region, corresponding to $n \geq 11$, holds some special significance, and it provides some reassurance regarding the unseen computations, it will have served its purpose here. Figure 7, which appears on the terminal simultaneously, explains that

The critical level of this test, corresponding to the light gray region, is .69

This means that if you reject the hypothesis (*The probability of occurrence of the perforation does not exceed .8*) **you do so with a .69 probability of being in error.**

```
The critical level of this test, corresponding to the light gray region,
is .69

This means that if you reject the hypothesis (The probability of
occurrence of the perforation does not exceed .8) you do so with a .69
probability of being in error.


Would you like to run the binomial procedure again? n
```

Figure 7. Explanation and interpretation of results.

Since the investigation began with the assumption (null hypothesis) that the Pr{perforation} $\leq$ .80, the evidence collected–eleven perforations, three nonperforations–is not sufficient to support abandonment of that assumption. A probability of being in error of .69 is more than a reasonable person would be willing to assume. And so, the response to the original question, Is the Pr{perforation} > .80 ? is a qualified no, the qualification being expressed through invocation of the critical level.

Would you like to run the binomial procedure again ?

At this juncture, an experimenter might well be asking a number of "What if ..." questions. "What if I had been able to afford three more firings?" or, "What if I had observed one more perforation?" and so on. A response of [y]es here allows the user to exercise the binomial procedure directly, without having to respond again to all the preliminary questions. A [n]o response is given, but this is an excellent place to use Nonpare's tutorial capabilities to study the sensitivity of the binomial procedure to modification of parameter values or slight changes in the data.

Are you interested in determining an interval in which the probability p of occurrence of the perforation lies?

The foregoing analysis suggests that an assertion that the probability of perforation lies in the interval (.80, 1] cannot be made. What interval might be expected to capture this unknown parameter? A response of [y]es causes this question to be answered, first graphically, as in figure 8, and then verbally, as in figure 9.

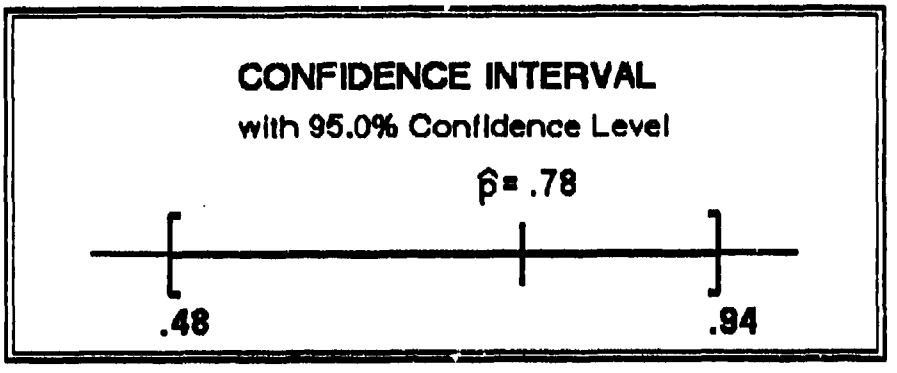


Figure 8. Display for a 95% confidence interval.

Figure 8 shows that the Pr{perforation}, whose estimate based on the fourteen firings is $\hat{p}$ = .78, lies within the interval [.48, .94] with a high level of confidence. This interval is so broad one can see why the assertion that Pr{perforation} > .80 is ill-advised. The formal interpretation of the confidence interval is given as

The probability of occurrence of the perforation is contained in the interval [.48, .94] with an a priori probability .95 .

```
Are you interested in determining an interval in which the probability p
of occurrence of the perforation lies? y

The probability of occurrence of the perforation is contained in the
interval [.48,.94] with an a priori probability .95.

Would you like a confidence level other than .95 ? n
```

Figure 9. Explanation and interpretation of the confidence interval.

Would you like a confidence level other than .95 ? [no]. The 95% confidence level was prechosen. A [y]es response allows the user to control the confidence level. The session is terminated with a [n]o response, shown in figure 9.

At the conclusion of the session the inference engine displays a fact solution tree for all the intermediate decisions leading to the final conclusion. Buttoning with a mouse any node of the fact tree produces the logic leading to that location. In figure 10, fact11 was buttoned, and the corresponding trace is displayed beneath the fact tree. These are features of the inference engine rather than Nonpare, but they are valuable as diagnostics to the developer and provide some measure of reassurance to the user.

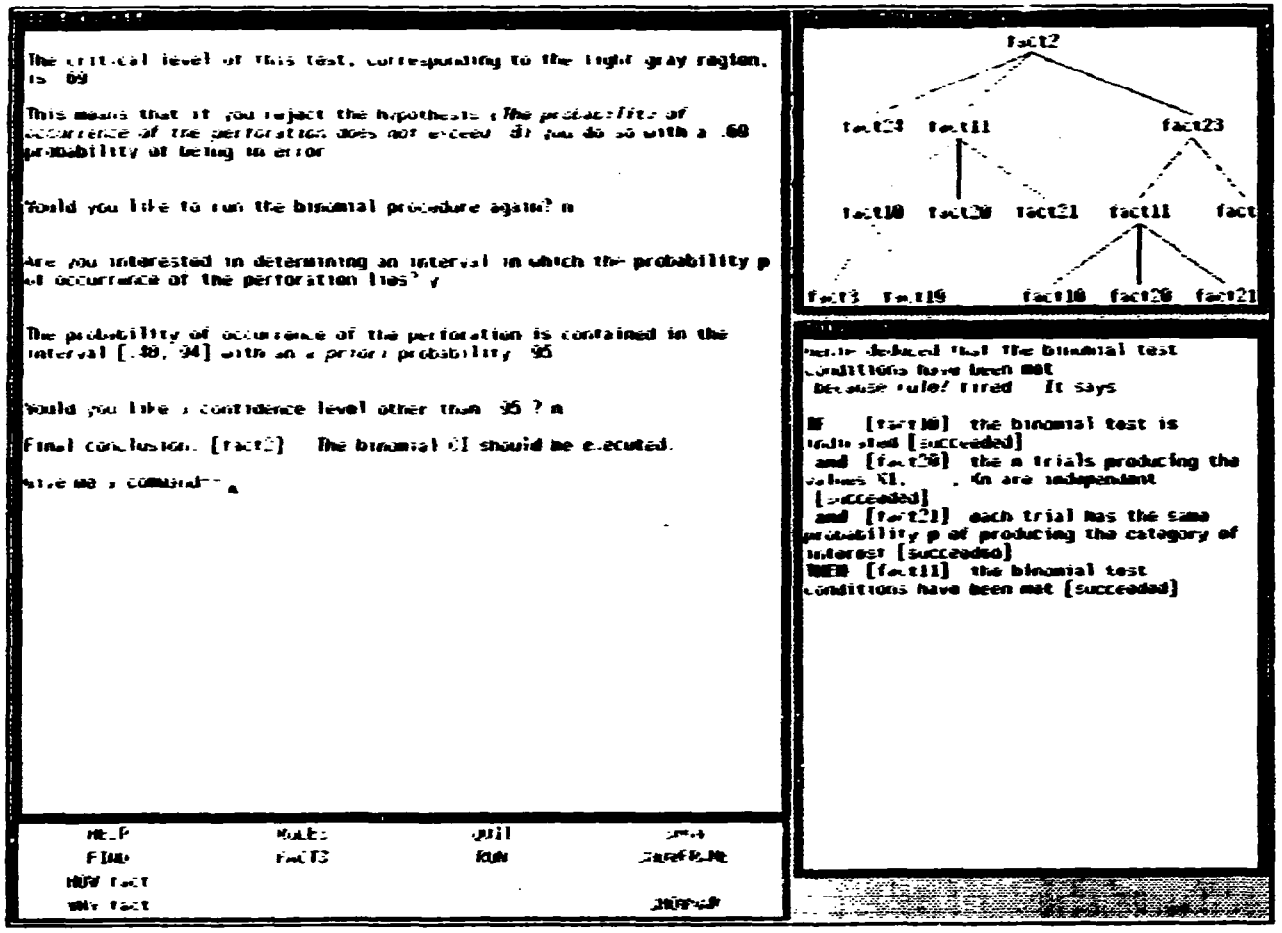


Figure 10. Session conclusion

4. Conclusions

Nonpare, a consultation system for analysis of data using nonparametric statistical procedures, has been described; and most of its operational features have been illustrated. The essence of the system is the rule-based interface with accompanying software for data analysis and the interpretation of the ensuing computations. Nonpare is under active development, but its feasibility as an operational system has been established. Enlargement of the rule-base and the addition of more statistical procedures is clearly indicated before it can approach its potential. Not surprisingly, tangential problems in basic research have been spawned by this effort. A first release is planned for 1989 for field testing.

References

[1] W.A. Gale, Ed., *Artificial Intelligence and Statistics* (Addison-Wesley, 1986).

[2] W.J. Conover, *Practical Nonparametric Statistics* (John Wiley, 1980).

[3] M. Hollander and D.A. Wolfe, *Nonparametric Statistical Methods* (John Wiley, 1973).

[4] E.L. Lehmann, *Nonparametrics* (Holden-Day, 1975).

[5] F.S. Brundick, et.al.,Genie: An inference engine with applications to vulnerability analysis, Technical Report BRL-TR-2739, US Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD (1986).

Numerical Estimation of Gumbel Distribution Parameters

Charles E. Hall, Jr.
Research Directorate
Research, Development, and Engineering Center
U.S. Army Missile Command
Redstone Arsenal, AL 35898-5248

ABSTRACT. The parameters which maximize the log-likelihood function of the Gumbel distribution were estimated by two different methods. A derivative approach was used, which calculated the intersection of the zeros of the implicit functions obtained from the derivatives of the log-likelihood function. A direct maximization was also performed. Both methods yielded positive results.

EXPERIMENTAL DESIGN AND OPTIMIZATION OF BLACK CHROME SOLAR SELECTIVE COATINGS

I. J. Hall and R. B. Pettit
Sandia National Laboratories
Albuquerque, NM 87185

ABSTRACT. Some years ago Sandia Laboratories was given the task of investigating selective coatings for solar applications. Early experimental results, which were based on one variable at the time experiments, produced acceptable coatings in the laboratory. However, when full scaled parts were coated by commercial electroplaters, the coatings quickly degraded when heated in air. At this point a systematic approach using a fractional factorial design was used to determine both the effects and interactions between several variables, including the bath composition(four variables), current density, plating time, substrate, and bath temperature. Response surface for the optical properties of the coatings were constructed for both the as-plated and the thermally aged samples. These response surfaces were then used to specify ranges for the bath compositions, and other plating parameters, that provided coatings with optimum thermal stability. With proper control of the plating variables, selective coatings were obtained that should maintain high solar absorptance values during years of operational at 300°C in air.

1. INTRODUCTION. Two variables are of interest to selective coating investigators, namely, absorptance (α) and emittance (ϵ). Good selective coatings have high α's and low ϵ's. In our investigations we concentrated on making α as large as possible and settling for the corresponding ϵ if it was not "too big". The independent variables that effected α and ϵ divided themselves into two groups (bath variables and plating variables) in such a way that a split-plot experimental design would have been appropriate. The bath variables would have been associated with the whole plots and the plating variables with the sub-plots. The bath variables were chromic acid, trivalent chromium, addition agent and iron. and the plating variables were plating time, current density, bath temperature, bath agitation and substrate. For a specified combination of bath variables a entire set of experiments were possible for the plating variables as in a split-plot design. Because of many constraints we did run the experiment as a split-plot design. The dependent variable readings (α's) were obtained by coating a substrate and then measuring the absorptance with a Beckman Model DK_2A spectroreflectometer. Readings were obtained for the substrate both as-plated and as-aged. The as-aged readings were obtained after the specimens were heated in a 450°C furnace for 40 hours while the as-plated readings were taken before they were subjected to any extreme environments. The aged readings were the most important because we were concerned about the thermal stability of the coatings, i.e. would coatings not degrade at high temperature for extended time periods. The experimentation was done in three phases that are briefly described below.

2. Experimentation. Based on previous experience, we decided that the bath variables were most important and thus we concentrated most of our efforts on investigating these variables. The plating variables were set at nominal values. We used standard response surface methodology to guide us in the experimentation. (See Box, Hunter, and Hunter, "Statistics for Experimenters", Chapter 15, 1978) The first phase consisted of running a 1/2 replicate of a 2^4 factorial experiment on the four bath variables. The experimentation was done in a rather limited range of the factor space. The results of this experiment were used to determine a path of steepest ascent (Phase two). Three more experiments were done along this line of steepest ascent. These experiments would normally indicate a region in the bath variable space that would produce larger α values. In our case however all the coatings turned grey after a short time in the furnace - a highly undesirable result. The most valuable information from these three bath experiments was that a "cliff" existed in the response surface. Because of time limitations we did not repeat the experiments along the steepest ascent line. Based on a combination of engineering judgement and factorial design methodology, several more baths were mixed and the α's measured on the coated substrates (Phase three). A total of eighteen baths were mixed and the results from these baths were used to estimate a quadratic surface - i.e. α was written as a function of a second degree polynomial in the four bath variables and the variable coefficients were estimated using a backward stepwise statistical package. The final regression equation had 11 terms including the constant term with an $R^2 = 0.96$. Several graphs were drawn based on this equation that allowed us to map out an acceptable region in the bath variable space. This space was very near the "cliff" in the response surface. A limited number of experiments also were done involving the plating variables for a fixed bath. Based on these experiments we were able to specify ranges for the plating variables as well.

3. Summary. Using response surface methodology we were able to determine the variables and the range of variables that produced stable selective coatings. The procedures developed in the laboratory were subsequently implemented in a production environment with excellent results. The close interaction between the statistician and the experimenter led to a satisfactory solution with a rather limited number of experiments.

DETERMINATION OF DETECTION RANGE OF MONOTONE AND CAMOUFLAGE PATTERNED FIVE-SOLDIER CREW TENTS BY GROUND OBSERVERS

George Anitole and Ronald L. Johnson
U. S. Army Belvoir Research, Development
And Engineering Center
Fort Belvoir, Virginia 22060-5606

Christopher J. Neubert
U. S. Army Materiel Command
Alexandria, Virginia 22333-0001

ABSTRACT

Field evaluations have determined that camouflage patterns reduce detectability ranges for uniforms and vehicles in woodland environments. This study identified the effects of three patterned and two monotoned Five-Soldier Crew Tents using detection ranges and number of false detections as determined by ground observers. The distance of correct detections were recorded along with the number of false detections. An analysis of variance for the detection ranges and number of false detections was performed. The Duncan's Multiple-Range Test was used to determine significant differences ($\alpha = 0.05$) in groups of tents. From this data, it was determined that the three patterned Five-Soldier Crew Tents were more difficult to detect than the two monotone tents.

1.0 SECTION I - INTRODUCTION

Several years ago, the U.S. Army decided that camouflage patterns have a definite advantage when used on uniforms and vehicles in woodland environments. This had led to a similar consideration for tents, since the current U.S. Army tents are solid (i.e., monotone) color. Tents present a large, relatively smooth form, making them conspicuous targets. The use of patterns to break up this signature could increase camouflage effectiveness. However, before such a judgement could be made, a field test was planned to determine the relative merits of various patterns versus monotones in a woodland background. The Natick RD&E Center fabricated three patterned tents and two monotone tents for evaluation. In consultation with Belvoir, the patterned tents were fabricated in the standard four-color uniform pattern, one in the standard pattern size and the other two in progressively larger expanded patterns. The two monotone tents were in colors Forest Green and Green 483 (483 being the textile equivalent of paint color Green 383). A test plan[1/] was developed by Belvoir at the request and funding of Natick, and the field test was conducted by Belvoir at Ft. Devens, Massachusetts, in the summer of 1987. This report describes the test and its results.

2.0 SECTION II - EXPERIMENTAL DESIGN

2.1 Test Targets

Five, Five-Soldier Crew Tents were supplied by Natick for this study in the following patterns and colors:

- Tent A - Standard size four-color uniform pattern repeated every 27.25 inches
- Tent B - Forest Green
- Tent C - Expanded four-color uniform pattern repeated every 36 inches
- Tent D - Expanded four-color uniform pattern repeated every 50 inches
- Tent E - Green 483

2.2 Test Sites

The study was conducted at the Turner Drop Zone, Ft. Devens, Massachusetts, a large cleared tract of land surrounded by a mix of coniferous and deciduous trees resembling a central European forest background. Two test sites were selected. Site #1 was located on the western end of the drop zone, so that the morning sun shone directly upon the test tent. Site #2 was located on the eastern edge of the drop zone, so that the afternoon sun shone directly upon the test tent. An observation path, starting at the opposite end of the drop zone from the test tent location, was laid out for each site. Each path followed zig-zag, random length directions toward its test site, and afforded a continuous line of sight to its respective test tent location. The paths were within a 30° to 40° cone from the target tents, and were surveyed and marked at approximately 50-meter intervals using random letter markers. For Site #2, the distance between markers after the first 15 markers was about 25 meters along the path. A night evaluation involving other camouflage targets led to this procedural change. The markers and distances from the tents are shown in Tables 1 and 2.

Table 1

Distances of Markers to Tents for Site #1

ALPHABET MARKER	DISTANCE IN METERS ALONG PATH FROM STARTING POINT TO TENT	ALPHABET MARKER	DISTANCE IN METERS ALONG PATH FROM STARTING POINT TO TENT
S	1,162.64	S'	464.78
Y	1,128.57	Y'	446.74
Q	1,094.00	Q'	428.17
L	1,049.93	L'	413.48
F	1,008.07	F'	398.48
P	978.31	P'	383.34
E	947.02	E'	364.64
K	902.75	K'	346.27
A	858.10	A'	334.46
T	817.81	T'	322.69
V	778.91	V'	308.59
B	750.15	B'	289.59
M	709.76	M'	281.60
U	674.87	U'	269.08
H	702.65	H'	253.16
Z	677.99	Z'	235.50
R	648.46	R'	217.81
N	613.35	N'	199.60
X	602.56	X'	178.93
I	594.57	I'	158.76
D	578.05	D'	141.15
C	561.16	C'	120.05
O	541.70	O'	102.34
J	525.33	J'	85.37
G	505.62	G'	62.81
W	483.64	W'	41.84

Table 2

Distances of Markers to Tents for Site #2

ALPHABET MARKER	DISTANCE IN METERS ALONG PATH FROM STARTING POINT TO TENT	ALPHABET MARKER	DISTANCE IN METERS ALONG PATH FROM STARTING POINT TO TENT
F	1,205.36	A	653.34
W	1,168.83	Z	613.20
U	1,130.58	E	574.09
Q	1,086.03	P	540.30
C	1,048.10	H	513.10
R	1,006.15	K	496.46
V	982.00	S	475.57
O	974.13	F'	459.10
M	942.37	W'	417.71
I	901.58	U'	379.40
B	869.75	Q'	338.25
J	858.01	C'	296.90
L	851.64	R'	278.53
X	841.26	V'	258.20
G	803.95	O'	220.73
D	764.09	I'	180.87
Y	723.46	B'	143.94
T	695.32	J'	111.00
N	673.60	L'	69.76

2.3 Test Subjects

A total of 153 enlisted soldiers from Ft. Devens served as ground observers. All personnel had at least 20/30 corrected vision and normal color perception. A minimum of 30 observers were used for each test tent, about evenly split between test sites. Each observer was used only once.

2.4 Data Generation

The test procedure was to determine the detection distances of the five tents involved by searching for them while traveling along the predetermined measured paths. Each ground observer started at the beginning of the observation path, i.e., marker S for Site #1 and marker F for Site #2. The observer rode in the back of an open 5/4-ton truck accompanied by a data collector. The truck traveled down the observation path at a very slow speed, about 3-5 mph. The observer was instructed to look for military targets in all directions except directly to his rear. When a possible target was detected, the observer informed the data collector and pointed

to the target. The truck was immediately stopped, and the data collector sighted the apparent target. If the sighting was correct, i.e., the Five-Soldier Crew Tent, the data collector recorded the alphabetical marker nearest the truck. If the detection was not correct, the false detection was recorded, and the data collector informed the observer to continue looking. The truck proceeded down the observation path. This search process was repeated until the correct target (tent) was located.

The tents were rotated between the two test sites on a daily basis, until all tents had been observed by at least 15 observers at each site. (This number of observers allows the use of parametric statistics which have a good opportunity to yield absolute conclusions). Their orientations with respect to the sun were kept constant at both test sites. The Five-Soldier Crew Tent was positioned so that a full side was facing the direction of observer approach.

3.0 SECTION III-RESULTS

3.1 Range of Detection

Tables 3, 4, and 5 show the detection data for the Five-Soldier Crew Tents. Table 3 gives the mean detection range in meters for each tent, and its associated 95 percent confidence interval. Table 4 shows the analysis of variance[2/] performed upon the data of Table 3 to determine if there were significant differences in the detection ranges, i.e., if pattern and color had an effect upon detection range. Table 5 indicates which tent patterns and solid colors differed significantly from each other in this respect. Figure 1 is a graphic display of the detection ranges of Table 3.

Table 3

Mean Detection Ranges (Meters) and 95 Percent Confidence Intervals.

TENT	N	MEAN	STANDARD ERROR	95 PERCENT CONFIDENCE INTERVAL	
				LOWER LIMIT	UPPER LIMIT
A	31	327.54	127.75	280.68	374.40
B	30	427.71	173.74	362.83	492.58
C	32	351.17	129.42	304.51	397.83
D	30	387.12	161.79	326.76	447.59
E	30	674.88	214.94	594.62	755.14

Table 4

Analysis of Variance for Tent Detection
Across Five Levels of Color Variation

SOURCE	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARE	F-TEST	SIG LEVEL
TENT COLOR	4	2,377,907.968	594,476.9927	22.0083	0.00*
ERROR	148	3,983,214.260	26,913.6099		
TOTAL	152	6,361,122.228			

*Significant at α less than 0.001 level.

Table 4 indicates that there are significant differences in the ability of the ground observers to detect the Five-Soldier Crew Tents in different four-color patterns and solid colors

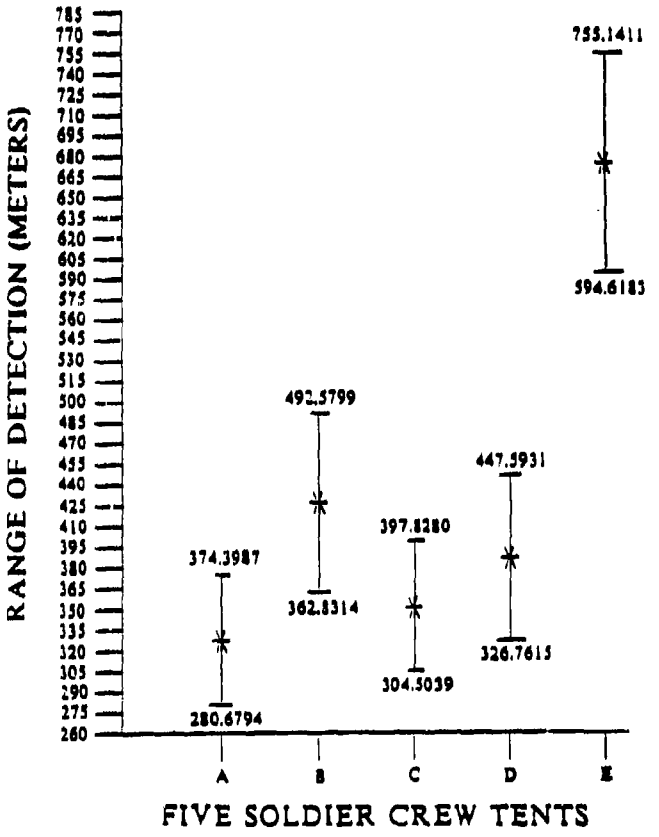


Figure 1. Mean Ranges of Detection and 95 Percent Confidence Intervals for Five-Soldier Crew Tents

Table 5

Duncan's Multiple-Range Test (Range of Detection)

SUBSET 1		SUBSET 2		SUBSET 3	
GROUP	MEAN	GROUP	MEAN	GROUP	MEAN
A	327.54	C	351.17	E	674.88
C	351.17	D	387.12		
D	387.12	B	427.71		

The harmonic mean group size is 30.58. The subsets are significant at $\alpha = 0.05$

The Duncan's Multiple-Range test separates a set of significantly different means into subsets of homogeneous means. One of the assumptions is that each random sample is of equal size. Since this was not true, the harmonic mean of the group was used as the group size. As seen above, the range of detection was the shortest for tents A, C, and D and these tents do not differ significantly from each other ($\alpha = 0.05$). Tent E had the longest mean range of detection and is significantly ($\alpha = 0.05$) different from the other 4 tents in this respect.

3.2 False Detections

The number of false detections is defined as the number of times a target other than the test target is detected by an observer. In this study such detections are rocks, trees, shadows, etc. These detections, as a rule, are a function of how hard it is to detect the test target. The more difficult the detection task, the greater the number of false detections. Tables 6, 7, and 8 show the false detection data. Table 6 gives the mean false detection value, and its associated 95 percent confidence interval, for each of the Five-Soldier Crew Tents. Table 7 contains the analysis of variance performed upon the data of Table 6 to determine if there were significant differences in the rate of false detections. Table 8 indicates which tent patterns and colors had significant rates of false detection.

Table 6

Mean False Detection Rates and 95 Percent Confidence Intervals

				95 Percent Confidence Interval	
Tent	N	Mean	Standard Error	Lower Limit	Upper Limit
A	31	4.87	3.27	3.67	6.07
B	30	3.53	2.53	2.59	4.48
C	32	3.38	1.96	2.67	4.08
D	30	3.87	2.76	2.83	4.90
E	30	2.50	1.91	1.79	3.21

Table 7

Analysis of Variance for Rates of False Detection across Five Levels of Color Variance

SOURCE	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARE	F-TEST	LEVEL
TENT COLOR	4	90.086	22.521	3.50	0.009
ERROR	148	953.417	6.44		
TOTAL	152	1043.503			

Significant at less than 0.01 level.

Table 7 indicates that there are significant differences in the rates of false detection for the Five-Soldier Crew Tents.

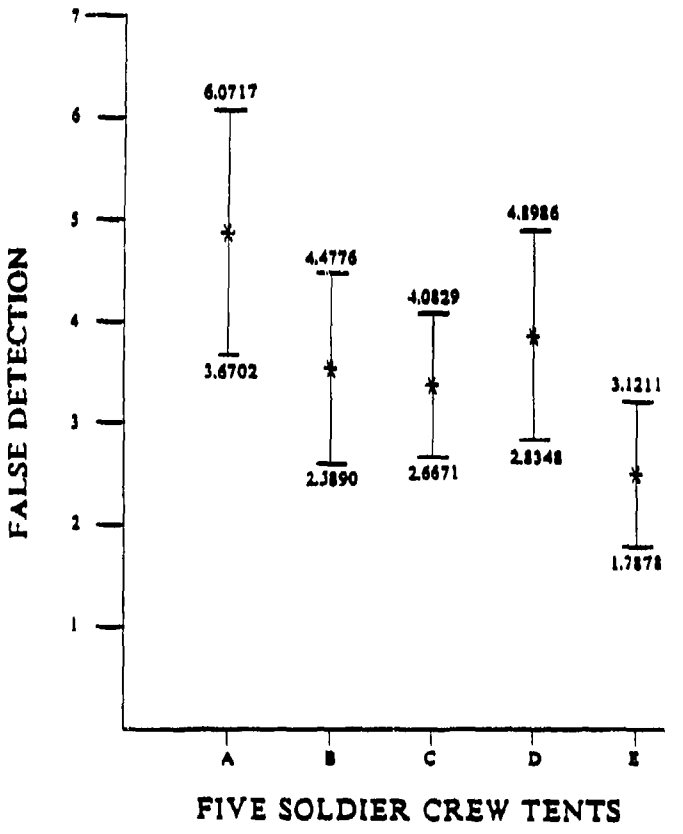


Figure 2. Mean Rates of False Detection and 95 Percent Confidence Intervals For Five-Soldier Crew Tents

Table 8

Duncan's Multiple-Range Test
(Rates of False Detection)

SUBSET 1		SUBSET 2	
GROUP	MEAN	GROUP	MEAN
E	2.50	B	3.53
C	3.38	D	3.87
B	3.53	A	4.87
D	3.87		

Harmonic mean group size is 30.58.

The rates of false detection for tent groups E, C, B, and D, and B, D, and A were not significantly different ($\alpha = 0.05$). However subset 1 is significantly different from subset 2.

4.0 SECTION IV - DISCUSSION

The Duncan's Multiple-Range Test (Table 5) shows that the group of Five-Soldier Crew Tents A, C, and D had the shortest detection range. Tent A is the standard size woodland uniform four-color pattern, while Tents C and D are expansions of this pattern. The pattern at Tent A is repeated every 27.25 inches, the pattern for Tent C is repeated every 36 inches, and the pattern for Tent D is repeated every 50 inches. Tents C, D, and B are significantly different from each other. Tent B is solid color, Forest Green. Tent E, which is not solid color Green 483, had the longest mean detection range (674.89 meters), and this is significantly ($\alpha = 0.05$) longer than any of the other means for the Five-Soldier Crew Tents. Thus, it can be concluded that the patterned tents are harder to detect from ground observation, but that the pattern should not be expanded beyond the repeat of every 36 inches. The human eye is probably resolving the larger pattern repeated every 50 inches as being different from the tree and bush background (the color brown, in particular, becomes distinguishable from the woodland background when overexpanded).

When working with detection ranges, the question of field data stability is always paramount to the amount of weight that can be given to the test conclusions. One of the best methods to determine data stability is through a test-retest procedure. Field studies are very expensive and time consuming, so this data is very rare. We do have such an opportunity to examine this type of data for the Turner Drop Zone. A ground evaluation of camouflage nets was conducted in the summers of 1985[3/] and 1987[4/]. The net sites and test procedures were identical to the sites and test procedures in which the Five-Soldier Crew Tents were evaluated. In both net studies, the standard camouflage net was evaluated. In 1985 this net had a mean detection range of 411.75 meters, while in 1987 the mean detection range was 414.41 meters. This difference in mean detection range is only 2.66 meters. From these results, it is inferred that the mean detec-

tion ranges for the Five-Soldier Crew Tents are stable, and solid conclusions about their camouflage effectiveness can be made.

The analysis of false detections seen in Table 8 and Figure 2 also lends credence to the belief that the Five-Soldier Crew Tent A had the best performance as to camouflage effectiveness, with Tent E the worst performance. In the following discussion of false detections in Section 3.2, it would be expected that Tent A, being the hardest to find, would have the most false detections, and Tent E the least number of false detections. This is exactly what occurred, with Tent A having a mean false detection rate of 4.87, and Tent E a mean false detection rate of 2.50. Duncan's Multiple-Range Test (Table 8) shows that the two rates of false detection differ significantly ($\alpha = 0.05$) from each other. The false detection rates of tents B, C, and D are not in the expected ordinal position. The expected order, based upon mean range of detection, would be B, D, and C, while the true order of rates of false detection is C, B, and D. However, a check of Tables 5 and 8 shows that these tents are not significantly different from each other either for range of detection or for rate of false detection. Thus, from a statistical view, these three tents are considered to have the same ordinal position.

5.0 SECTION V-SUMMARY AND CONCLUSIONS

Five, Five-Soldier Crew Tents were evaluated by ground observers to determine their camouflage effectiveness as measured by the mean detection range and the mean rate of false detection. These tents were in the following four-color camouflage patterns and solid colors:

- Tent A - Standard size four-color uniform pattern repeated every 27.25 inches.
- Tent B - Forest Green
- Tent C - Expanded four-color uniform pattern repeated every 36 inches
- Tent D - Expanded four-color uniform pattern repeated every 50 inches
- Tent E - Green 483

A minimum of 30 ground observers per Five-Soldier Crew Tent were driven toward each of two sites on marked observation trails in the back of an open 5/4-ton truck. The observers were looking for military targets, and they informed the data collector when they thought they saw one. If the detection was correct, the closest alphabetic ground marker to the truck was recorded. From this letter, the distance to the tent from the truck was determined. If the detection was not correct, i.e., false detection, it was noted on the data sheet. The ground observer then continued the search, with the truck traveling down the observation path until the test target was seen. An analysis of the resulting data provided the following conclusions:

A. Five-Soldier Crew Tent A was the most camouflage effective, with the lowest mean range of detection and highest rate of false detections.

B. Four-color pattern Five-Soldier Crew Tents are more camouflage effective than solid colors.

C. The expanded four-color pattern, repeated every 50 inches, is too large to be effective in denying detection. (The color brown becomes distinguishable from the woodland background when overexpanded).

D. The solid colors Green 483 and standard Forest Green should not be used.

E. The mean range of detection data appears to be very stable. A test-retest field study using identical sites and test procedures in the summers of 1985 and 1987 involving the standard camouflage net yielded mean detection ranges of 411.75 and 414.41 meters respectively.

REFERENCES

1. Anitole, George and Johnson, Ronald, Unpublished Outline Test Plan, Evaluation of Camouflage Tents, U.S. Army Belvoir Research, Development and Engineering Center, Fort Belvoir, VA, 1987.

2. Natrella, Mary G., Experimental Statistics, National Bureau of Standards Handbook 91, U.S. Department of Commerce, Washington, D.C., 1966.

3. Anitole, George, and Johnson, Ronald, Statistical Evaluation of Woodland Camouflage Nets by Ground Observers, U.S. Army Belvoir Research, Development and Engineering Center, Fort Belvoir, VA, August 1986.

4. Anitole, George, and Johnson, Ronald, Evaluation of Woodland Camouflage Nets by Ground Observers, U.S. Army Belvoir Research, Development and Engineering Center, Fort Belvoir, VA, 1988.

AN EXAMPLE OF CHAIN SAMPLING AS USED IN ACCEPTANCE TESTING

JERRY THOMAS
ROBERT L. UMHOLTZ
WILLIAM E. BAKER

PROBABILITY & STATISTICS BRANCH
SYSTEM ENGINEERING & CONCEPTS ANALYSIS DIVISION
US ARMY BALLISTIC RESEARCH LABORATORY
ABERDEEN PROVING GROUND, MD 21005-5066

ABSTRACT

The Probability and Statistics Branch of the Ballistic Research Laboratory was asked to develop a procedure of acceptance testing for armor packages. Because the available sample sizes were extremely small, we were unable to identify a sampling plan directly applicable to this problem. Accordingly, we have devised a new procedure by adapting an existing technique, known as chain sampling, to both the attribute portion (structural integrity) and the variable portion (penetration depth) of the acceptance testing process. Operating characteristic curves and power curves are presented for this procedure, and suggestions are made concerning the simultaneous use of quality control charts.

I. INTRODUCTION

In most cases a consumer's decision concerning whether or not to accept a manufactured product is based on an examination of a sample from that product. When General Mills introduces a new pre-sweetened breakfast cereal, they spend millions of dollars in advertisement costs with the hope that the consumer will sample it. Here, the consumer considers the entire supply of this new cereal as a single manufactured lot, to be accepted or rejected. Product acceptance, in this case, corresponds to the consumer purchasing more boxes of the new cereal.

This is merely an everyday example of what is known as acceptance sampling, that is, various techniques which allow for discrimination between an acceptable product and an unacceptable one. Sampling may be based on an attribute criterion, a variable criterion, or some combination of these. In our example the consumer may judge the sweetness of the cereal as satisfactory or excessive (attribute), or he may measure the time in milk before the cereal becomes soggy (variable). Sampling by attributes is a dichotomous situation in that, based on a particular attribute, each item is either defective or non-defective; rejection occurs if there is a high percentage of defectives in the sample. Sampling by variables establishes an acceptable level of a particular variable, and rejection occurs if its sample value crosses the acceptable threshold. Of course, in our example of a box of cereal, the sample size was one. Generally, this will not be the case; but occasionally, for one reason or another, the consumer is forced to make a decision based upon a very small sample size.

Because decisions are made from samples, there is some risk of error, either the error of accepting a bad product or the error of rejecting a good product. The amount of protection desired against such risks can be specified. The Acceptable Process Level (APL) is a high-quality level that should be accepted $100(1-\alpha)\%$ of the time; α is thus defined to be the producer's risk. The Rejectable Process Level (RPL) is a low-quality level that should be accepted only $100(\beta)\%$ of the time; β is thus defined to be the consumer's risk. Unfortunately, these error factors vary inversely; that is, as the consumer's risk grows, the producer's risk diminishes and vice versa. The Operating Characteristic (OC) curve is an important part of any acceptance sampling plan, since it provides a graphical display of the probability of accepting a product versus the value of the particular parameter being inspected. The OC curve is a function of APL, RPL, α, and β, as well as sample size. Given a particular acceptance sampling plan, the OC curve depicts the associated error risks and demonstrates the relationship among all of the variables.

The US Army Ballistic Research Laboratory (BRL) has developed acceptance sampling plans for armor packages. These plans were briefed to the Project Manager M1A1 on 14 April 1988 at Aberdeen Proving Ground, Maryland. Their general structures were accepted with the guidance that the processes would be officially adopted pending some refinements.

II. CHAIN SAMPLING

Numerous sampling techniques exist, each with special properties that make it applicable to particular situations. Sampling plans reviewed in the literature required sample sizes much larger than those feasible for armor testing. In our case extremely small sample sizes were warranted due to the expense of both the armor and the testing procedure, augmented by the destructive nature of the test itself. Accordingly, we have devised a new procedure by adapting an existing technique, chain sampling, for use in this project.

Chain sampling is particularly appropriate for small samples because it uses information over the past history of production lots. Even with small samples, it is possible to accept a marginal lot provided that a given number of lots immediately preceding (i.e., the chain) were acceptable. When a consumer uses an expendable product such as the breakfast cereal in our previous example, he utilizes chain sampling in his decision of whether or not to subsequently purchase the same product. If the first or second box he buys is unacceptable, he will probably discard the product forever. However, if the tenth box is unacceptable, he might continue with one more purchase of the same cereal taking into consideration its past history of nine boxes of acceptable quality.

An advantage of chain sampling is its automatic incorporation of reduced or tightened inspection procedures when applicable. That is, as quality remains acceptable over a period of time and our confidence grows, the sample size is reduced (or, more accurately, samples are taken less frequently). If quality becomes marginal, inspection is tightened by taking samples more frequently. When quality diminishes to the point where a production lot must be rejected, the production process is stopped and necessary adjustments and corrections are made. At that point a new chain begins and continues as before.

Certain assumptions must be made before chain sampling is considered as a sampling technique. In particular, production should be a steady, continuous process in which lots are tested in the order of their production. Also, there should be confidence in the supplier to the extent that lots are expected to be of essentially the same quality. Generally, a fixed sample size will be maintained with the investigator taking more or fewer samples as tightened or reduced inspection is dictated.

III. ACCEPTANCE SAMPLING PLAN

The armor packages tested at the BRL consist of a right side and a left side, which are designated as one set. One month's production is considered to be a production lot. Every month we continue testing one set at a time until a decision can be made about that production lot. For a given set, one shot is fired into each side; and, if spacing on the target permits, a second shot follows. In each of the first three months, a total of at least four shots is required in order to make a decision concerning that month's production. This provides additional confidence during the early stages of the plan. There are two portions of the acceptance sampling plan. The first is structural integrity, handled using attribute methods; the second is depth of penetration of a particular round fired into the armor, handled using variable techniques. For both portions, decisions concerning a production lot should be based upon the data from all available shots on that lot.

A combined chain sampling plan was proposed. The maximum length of the chain was fixed at eight, meaning that after the chain has been established, we will consider the current set along with the seven immediately preceding. While the chain is growing, there is an area between the criteria for acceptance and rejection in which we can make no decision. At least one set will be tested each month; but if no decision can be made, tightened inspection will dictate the examination of additional sets, possibly up to a maximum of eight. Table 1 shows the relationships among months, sets, and shots for this particular procedure. Note that the maximum number of sets and, hence, the maximum number of shots decrease over time as the chain is being formed. Following the third month and the concurrent drop in the minimum number of shots, then when the chain is at its full length (definitely by the eighth month), one set and at most four shots are all that is required in order to make a decision for each subsequent production lot.

A rejection in either the structural integrity or the penetration depth will result in overall rejection of the production lot. In that case production is stopped, adjustments and corrections are made, and testing resumes with the construction of a new chain. If neither measure results in a rejection but at least one falls within the no-decision region, another set should be examined and both categories re-evaluated using the additional data.

A. Acceptance Sampling by Attributes

Projectiles are fired at these packages, which are then inspected for structural integrity. With attribute sampling, only two outcomes are possible. The structural integrity is assessed to be either defective or non-defective, regardless of the number of shots. Any decision to either accept or reject a lot is based on the number of defective plates in the sample being considered.

Chain sampling is employed in this attribute sampling plan. Results from the most recent eight sets influence decisions regarding a lot. A lot can be either accepted or rejected at any time (except for one case discussed in the next paragraph). In the early stages of sampling there is also an area in between acceptance and rejection where no decision is rendered immediately but sampling is continued. After a chain reaches its full length of eight sets, a decision to accept or reject is made immediately.

In the sampling plan, a safeguard is built in to prevent rejection of a good lot after only one set. If there are no defectives in the first set, the lot is accepted. Otherwise, no decision is made. Subsequently, rejection would occur only when there were three or more failures in the most recent eight sets.

Table 2 shows the decision rules for a chain building to a maximum length of eight. The OC curves for this plan are depicted in Figure 1. It shows that for a chain at full length, the probability of accepting a lot whose true defective rate is 5% is equal to 0.96, while the probability of accepting a lot whose true defective rate is 10% is equal to 0.79. Power curves for the plan are depicted in Figure 2. For a chain at full length, the probability of rejecting a lot from a process whose true defective rate is 5% is equal to 0.04, while the probability of rejecting a lot whose true defective rate is 10% is equal to 0.21. (Note, if these probabilities are deemed to be unsatisfactory, a different plan providing more satisfactory levels could be developed by varying the maximum chain length or modifying the decision rules).

TABLE 1. Relationships Among Variables in Chain Sampling Procedure

Month	Required Sets		Required Shots	
	Minimum	Maximum	Minimum	Maximum
1	1	8	4	32
2	1	7	4	28
3	1	6	4	24
4	1	5	2	20
5	1	4	2	16
6	1	3	2	12
7	1	2	2	8
8	1	1	2	4
9	1	1	2	4
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
k	1	1	2	4

TABLE 2. Decision Rules for Acceptance Sampling by Attributes.

SET NUMBER	DECISION RULES: ACCEPT	REJECT	NO DECISION
1	$f_1 = 0$	----	$f_1 \geq 1$
2	$\sum_{i=1}^{2} f_i = 0$	$\sum_{i=1}^{2} f_i \geq 3$	$1 \leq \sum_{i=1}^{2} f_i \leq 2$
.			
.			
.			
5	$\sum_{i=1}^{5} f_i = 0$	$\sum_{i=1}^{5} f_i \geq 3$	$1 \leq \sum_{i=1}^{5} f_i \leq 2$
6	$\sum_{i=1}^{6} f_i \leq 1$	$\sum_{i=1}^{6} f_i \geq 3$	$\sum_{i=1}^{6} f_i = 2$
7	$\sum_{i=1}^{7} f_i \leq 1$	$\sum_{i=1}^{7} f_i \geq 3$	$\sum_{i=1}^{7} f_i = 2$
8	$\sum_{i=1}^{8} f_i \leq 2$	$\sum_{i=1}^{8} f_i \geq 3$	----
9	$\sum_{i=2}^{9} f_i \leq 2$	$\sum_{i=2}^{9} f_i \geq 3$	----
.			
.			
.			
k	$\sum_{i=k-7}^{k} f_i \leq 2$	$\sum_{i=k-7}^{k} f_i \geq 3$	----

f_i = number of failures in set i

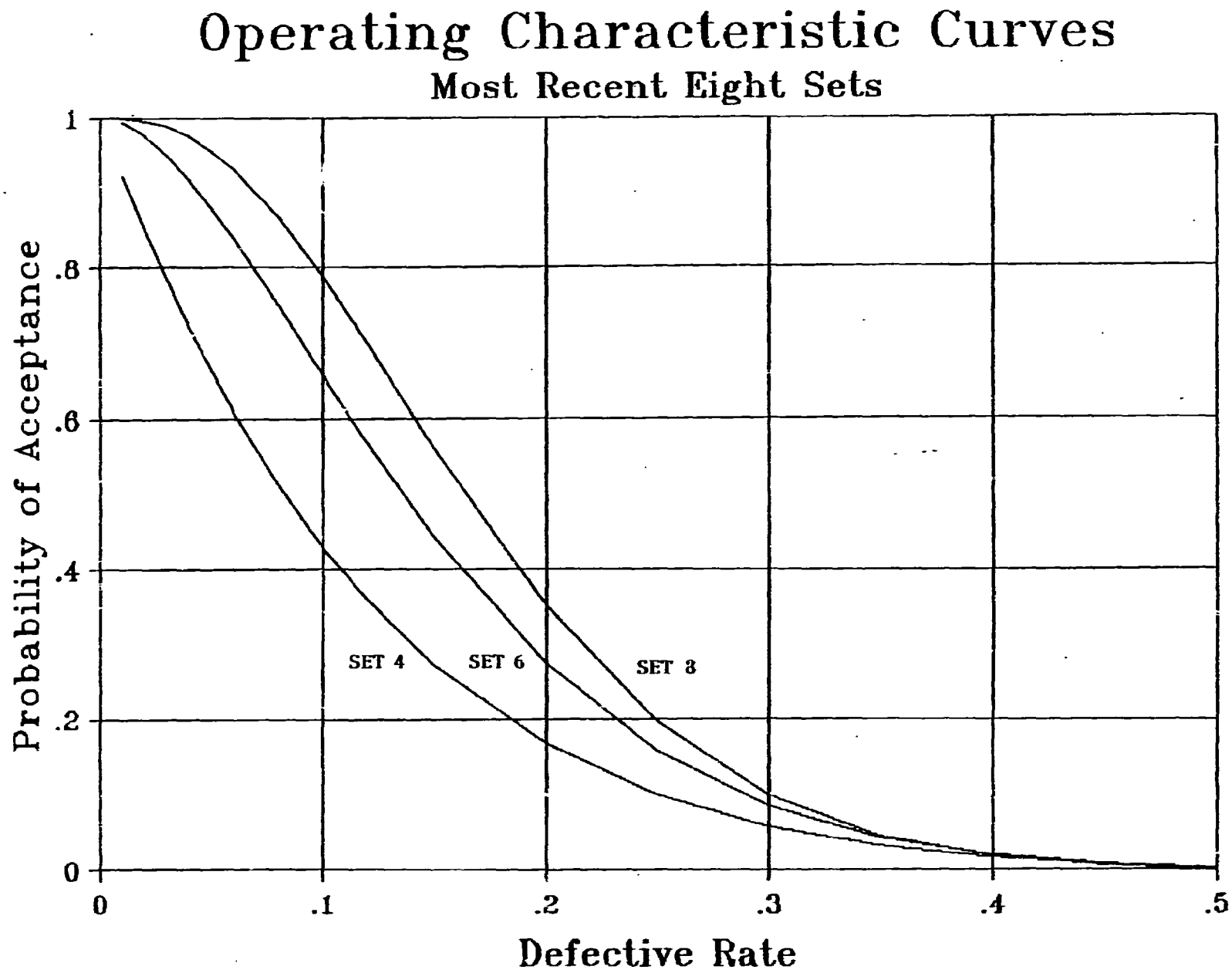


FIGURE 1. OC Curves for Acceptance Sampling by Attributes.

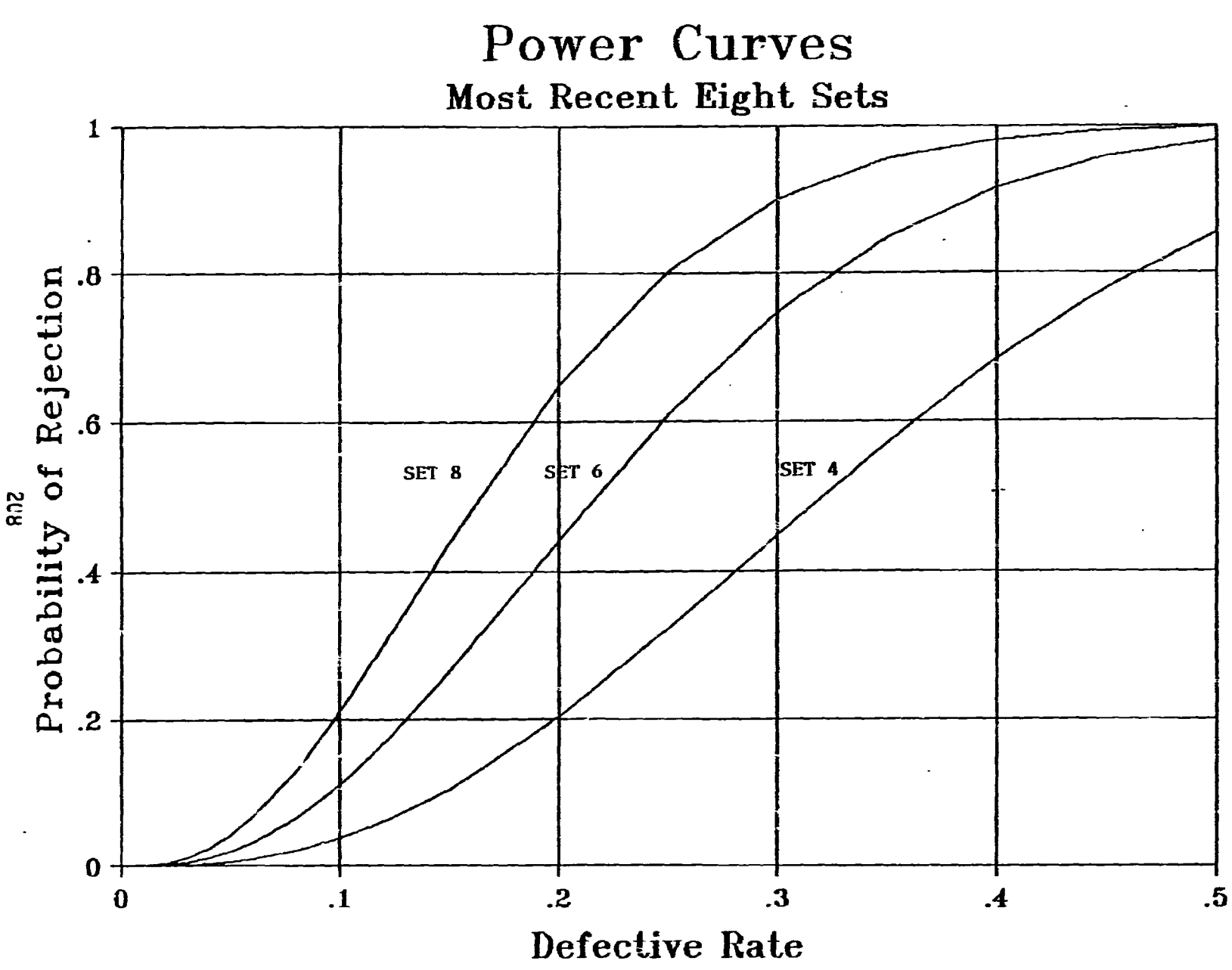


FIGURE 2. Power Curves for Acceptance Sampling by Attributes.

B. Acceptance Sampling by Variables

When primary interest is in a process level rather than a percent defective, sampling by variables is the proper procedure. For the armor packages, depth of penetration for a particular munition was the process level of interest. When variable sampling plans are established, two major assumptions must be satisfied: first, the distribution of the variable of interest must be known; and second, a good estimate of its standard deviation must be available.

In our particular problem there were 22 baseline shots from which we were to determine a distribution and estimate its standard deviation, as well as establish acceptable and rejectable process levels (APL & RPL). The 22 shots had a mean ($\bar{X}_b$) of 5mm with a standard deviation (s_b) of 30mm. The data had been transformed, allowing for both positive and negative penetration values. When plotted, the data appeared normal; and, indeed, the hypothesis of normality could not be rejected using statistical goodness-of-fit tests. The APL was established at 20mm (1/2 baseline standard deviation from the baseline mean) and the RPL was set at 80mm (2 1/2 baseline standard deviations from the baseline mean). α, the probability of rejecting at the APL, was set at 0.05; and β, the probability of accepting at the RPL, was allowed to vary with the sample size -- for a sample of size four, β would equal 0.10.

As in the attribute case, a set consists of a right side and a left side. For each set an attempt will be made to fire a second round into each side. Because this might not always be possible, due primarily to discrepancies between the aim point and the hit location, each set can result in either two, three, or four data points, depending on whether or not both shots on each side are considered to be good hits. It is important that during the first three months, while the chain is being formed, at least four shots are available upon which to make a decision. Table 3 outlines the decision rules for the variable sampling plan. Like the attribute sampling plan, it incorporates chain sampling with a maximum length of eight sets. The plan will not reject based on the first sample, and it has a region of no decision until the chain reaches its full length. In this table, $\bar{X}$ represents the mean penetration depth for all shots currently considered, s represents the standard deviation of this sample, n is the total number of shots used in computing $\bar{X}$, and $t_{.95}$ represents the 95th percentile of the t-distribution for the appropriate degrees of freedom (n-1). Thus, n can vary from 2 to 32 depending upon the length of the chain and the number of shots available on each side of the armor package.

Because n varies so widely, any one of many OC curves may be applicable. Figure 3 shows these curves for sample sizes 2, 32, and many integers in between. The abscissa value, D, represents a multiple of s_b from $\bar{X}_b$; thus, the numbers in parentheses are the penetration depths in millimeters. Note that for all n, the probability of accepting at the APL is 0.95 ($1-\alpha$). Because the probability of accepting at the RPL is too high for $n=2$ and $n=3$, the procedure will not allow lot acceptance at these small sample sizes (see Table 3). Table 4 provides the values for the t-statistic for ($1-\alpha$)-levels of 0.99 and 0.95 and degrees of freedom from 3 to 31.

Power curves show the probability of rejecting a particular lot. Generally, they are nothing more than the complement of OC curves. However, for our procedure this is not the case, since there is a region of no decision. Figure 4 shows the power curves for this variable sampling procedure. Basically, there are two sets of curves -- the first two pertaining to $\alpha = 0.05$ and the next three pertaining to $\alpha = 0.01$. Note from Table 3 that in order to reject

TABLE 3. Decision Rules for Acceptance Sampling by Variables.

DECISION RULES			
SET NUMBER	ACCEPT *	REJECT *	NO DECISION
1 ($n < 4$)	---	---	ALL
1 ($n = 4$)	$\frac{\bar{X} - APL}{s/\sqrt{n}} \leq t_{.95}$	---	$\frac{\bar{X} - APL}{s/\sqrt{n}} > t_{.95}$
2 (combine with 1)	$\frac{\bar{X} - APL}{s/\sqrt{n}} \leq t_{.95}$	---	$\frac{\bar{X} - APL}{s/\sqrt{n}} > t_{.95}$
3 (combine with 1,2)	$\frac{\bar{X} - APL}{s/\sqrt{n}} \leq t_{.95}$	$\frac{\bar{X} - APL}{s/\sqrt{n}} > t_{.99}$	$t_{.99} \geq \frac{\bar{X} - APL}{s/\sqrt{n}} > t_{.95}$
. . .			
7 (combine with 1-6)	$\frac{\bar{X} - APL}{s/\sqrt{n}} \leq t_{.95}$	$\frac{\bar{X} - APL}{s/\sqrt{n}} > t_{.99}$	$t_{.99} \geq \frac{\bar{X} - APL}{s/\sqrt{n}} > t_{.95}$
8 (combine with 1-7)	$\frac{\bar{X} - APL}{s/\sqrt{n}} \leq t_{.95}$	$\frac{\bar{X} - APL}{s/\sqrt{n}} > t_{.95}$	---
9 (combine with 2-8)	$\frac{\bar{X} - APL}{s/\sqrt{n}} \leq t_{.95}$	$\frac{\bar{X} - APL}{s/\sqrt{n}} > t_{.95}$	---
. . .			
k (combine with (k-7) - (k-1))	$\frac{\bar{X} - APL}{s/\sqrt{n}} \leq t_{.95}$	$\frac{\bar{X} - APL}{s/\sqrt{n}} > t_{.95}$	---

* At least four shots are required in each of the first three months; otherwise, regard as "No Decision".

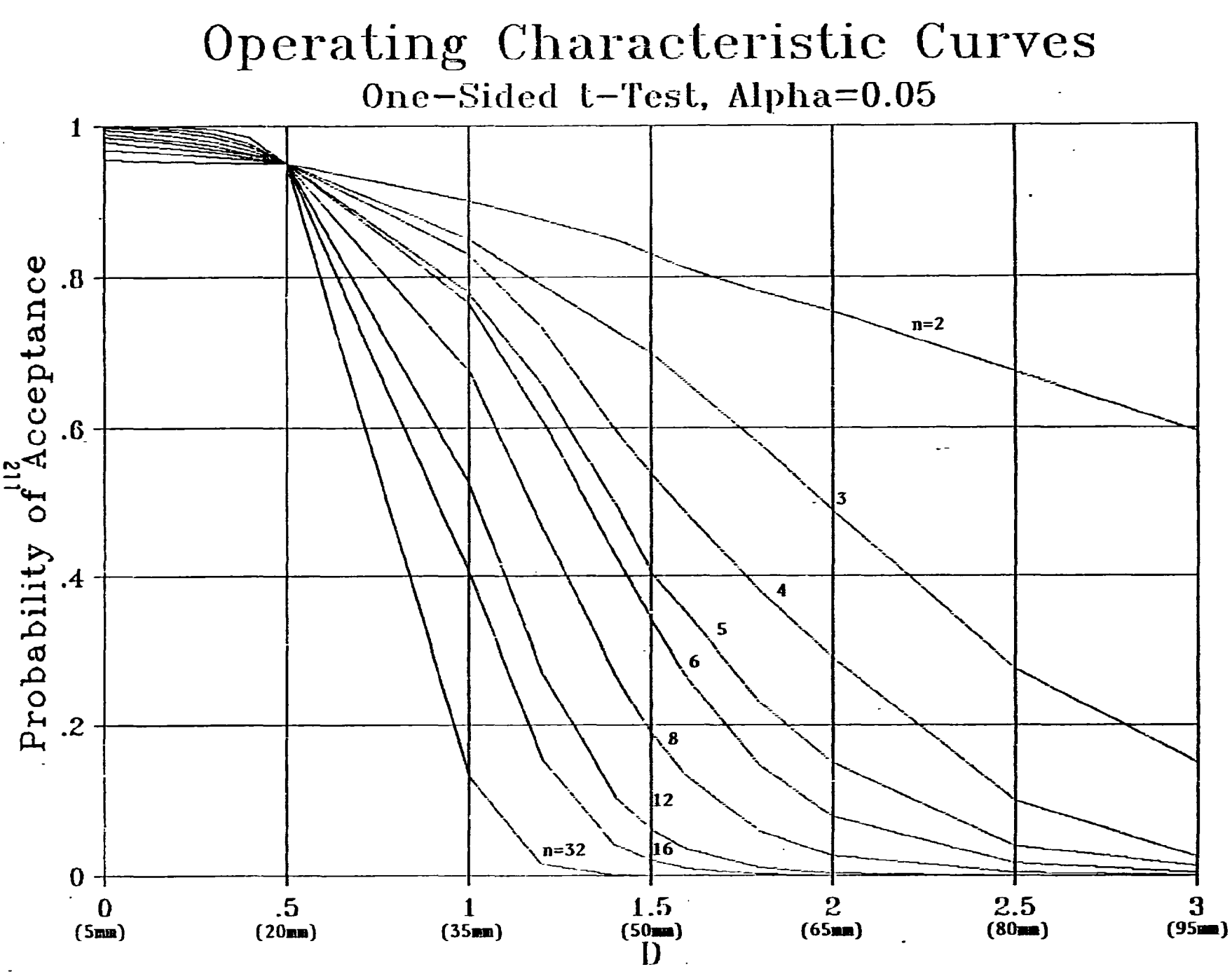


Figure 3. OC Curves for Acceptance Sampling by Variables.

TABLE 4. Values of the Cumulative t-Statistic

Degrees of Freedom (n-1)	(1-α) - level 0.95	0.99
3	2.35	4.54
4	2.13	3.75
5	2.02	3.37
6	1.94	3.14
7	1.90	3.00
8	1.86	2.90
9	1.83	2.82
10	1.81	2.76
11	1.80	2.72
12	1.78	2.68
13	1.77	2.65
14	1.76	2.62
15	1.75	2.60
16	1.75	2.58
17	1.74	2.57
18	1.73	2.55
19	1.73	2.54
20	1.73	2.53
21	1.72	2.52
22	1.72	2.51
23	1.71	2.50
24	1.71	2.49
25	1.71	2.49
26	1.71	2.48
27	1.70	2.47
28	1.70	2.47
29	1.70	2.46
30	1.70	2.46
31	1.70	2.45

*This table is abridged from Tables of the Probability Integral of the Central t-Distribution by R.E. Mioduski, BRL Technical Note #1570, August 1965.

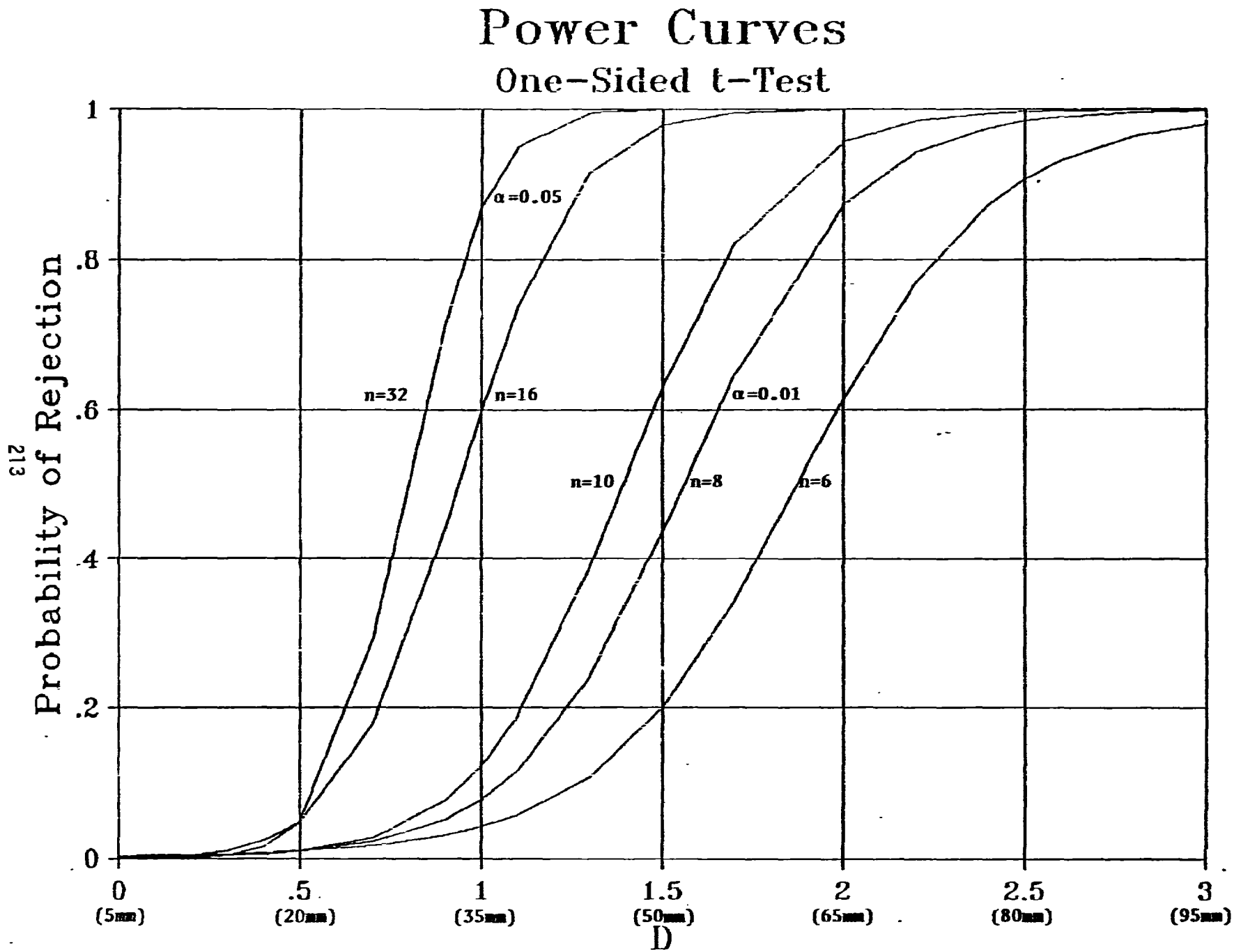


Figure 4. Power Curves for Acceptance Sampling by Variables.

before the chain is at its maximum length, we use the smaller α-level, and Figure 4 shows some possible sample sizes for $\alpha = 0.01$. If we reject at an α-level of 0.05, our sample size must be somewhere between 16 and 32; and these curves are also shown in Figure 4. Generally, the power curves are of more interest to the producer than the OC curves, since they highlight the producer's risk.

C. Quality Control Charts

Variations in the manufacturing process are either random or assignable. A process is "in control" when only random variations are present. Assignable variations, if uncorrected, may eventually result in rejection of a manufactured lot. However, they can often be identified through the use of quality control charts.

A quality control chart is a graphical comparison of test data with some previously computed control limits. The most common quality control chart is the Shewhart chart, named for its originator, Dr. Walter A. Shewhart. Figure 5 is a Shewhart control chart for mean penetration depth, the variable of interest in our armor package acceptance sampling plan. The APL is the central line with an upper control limit equal to the RPL, two baseline standard deviations away from the APL. If we were concerned about extremely low penetration depths, we would incorporate a lower control limit as well. Assuming a normal distribution with parameters equal to those of the baseline data implies that if only random variations are present, 99.38% of the time the mean penetration depth of the sample will fall below the upper control limit. This leaves a false-alarm frequency of less than 1% (0.62%) - so low that this control limit seems to be a reasonable threshold to distinguish between random variations and assignable variations.

The mean penetration depth is plotted for consecutive sets of armor plate. If, over a period of time, we see a drifting toward the control limit, the process can be examined and adjusted. This might possibly eliminate some future rejection of an entire lot.

A similar chart should be constructed for the range of penetration depth within the sample, to insure that the variability of the armor packages is not increasing. A third chart for structural integrity, the attribute of interest in our acceptance sampling plan, would also be useful. In each case appropriate upper control limits must be established.

Over the years alternative quality control charts have emerged, each with their own set of advantages and disadvantages. One of the most popular has been the cumulative sum control chart (cusum chart). Here, decisions are made based on all the data rather than just the last sample. An advantage of the cusum chart is that it often displays sudden and persistent changes in the process mean more readily (that is, with fewer samples and less expense) than a comparable Shewhart chart. However, control limits are somewhat less intuitive and, therefore, more difficult to establish. Somewhere in between the Shewhart chart and the cusum chart are quality control charts that use some, but not all, of the past data. Many of these techniques incorporate a weighting factor, providing more weight to the most recent data.

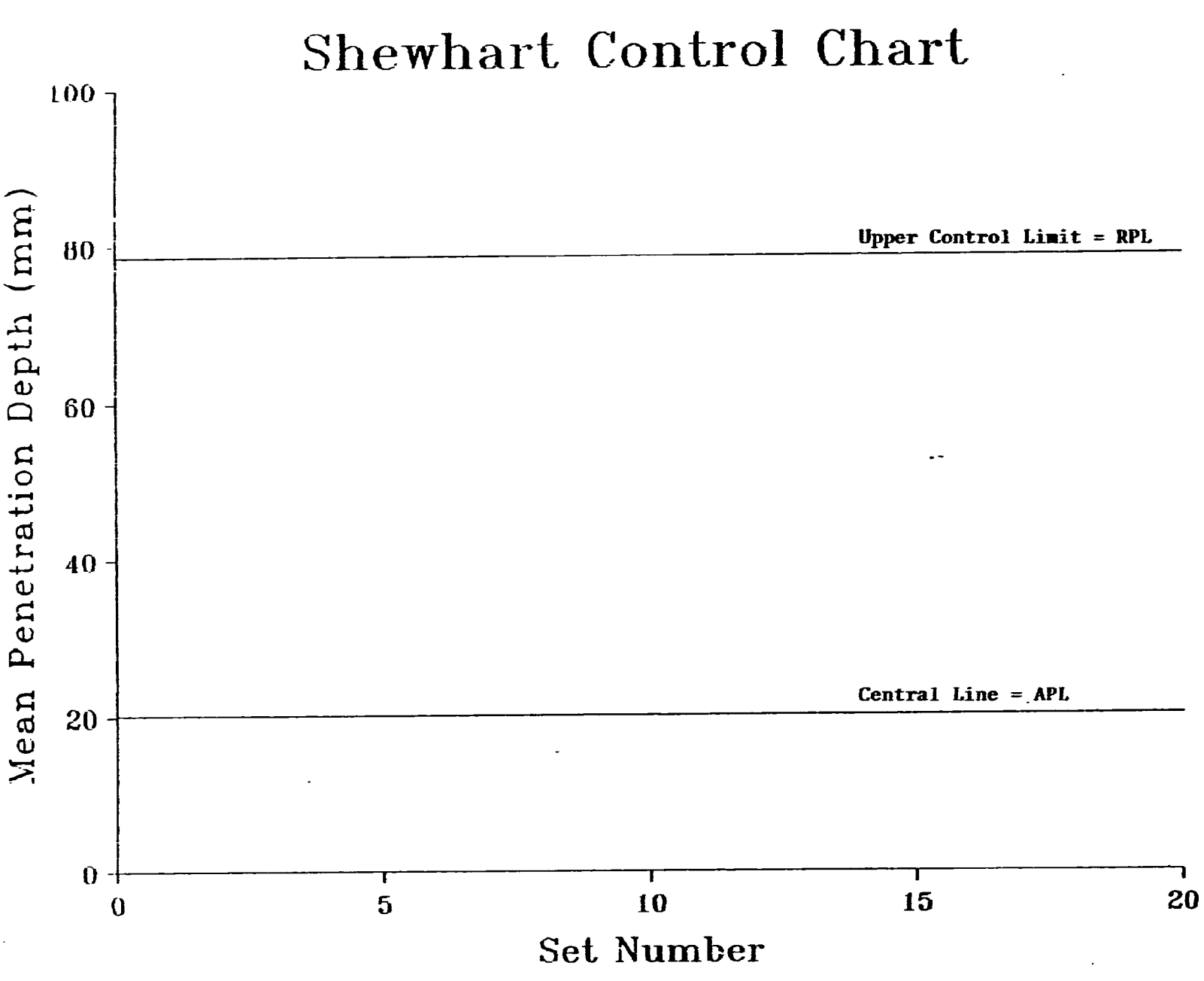


FIGURE 5. Quality Control Chart for Mean Penetration Depth.

It is important that some type of quality control charts be represented in the acceptance sampling plan. They are relatively easy to maintain and might provide early warning signs which could be beneficial to both the producer and the consumer.

IV. CONCLUSIONS

Generally, it is not feasible for a consumer to inspect every item from a production lot that he might want to purchase. A judicious choice of a lot acceptance plan will allow him to sample the production lot and determine with a pre-established level of confidence whether or not it meets his specifications. Chain sampling is a particular method of lot acceptance sampling used when sample sizes are small. It utilizes the most recent lot information to provide more confidence in the decision.

In testing armor packages for acceptance by the US Army, chain sampling provides a logical method, since destructive testing dictates small sample sizes. A technique involving both structural integrity (attribute sampling) and penetration depth (variable sampling) has been proposed. One set of armor packages is represented by both a left side and a right side. The procedure allows for accepting the production lot (one month's production) after examining just one set. It allows for rejecting the production lot only after testing at least two sets. There is a region of no decision; but after the chain has reached its maximum length of eight sets, a decision must be rendered.

Operating characteristic curves and power curves provide the probability of accepting and rejecting lots given a percent structurally defective (attributes) and given a mean penetration depth (variables).

In addition to the acceptance sampling plans, control charts should be used for both the attribute and variable parameters. These charts display sample results for particular parameters such as percent defective, mean penetration depth, and variability of penetration depth. The data might be presented as individual sample points or as sums over a preceding number of samples. By continually examining the control charts, we can see when one of the parameters is drifting toward the rejection region, enabling the producer to make adjustments and, possibly, preventing rejection of an entire lot of armor plate.

The proposed lot acceptance plan was briefed to the Project Manager M1A1 on 14 April 1988 at Aberdeen Proving Ground, Maryland. It was approved and will be adopted subject to any refinements agreed upon by both the US Army Ballistic Research Laboratory and the Project Manager.

BIBLIOGRAPHY

Crow, Edwin L., et. al., Statistics Manual, with Examples Taken from Ordnance Development, Dover Publications, Inc., 1960.

Duncan, Acheson J., Quality Control and Industrial Statistics, Richard N. Irwin, Inc., 1974.

Grant, Eugene L. & Leavenworth, Richard S., Statistical Quality Control, McGraw Hill Book Company, 1980.

Juran, J.M., Editor, Quality Control Handbook, McGraw Hill Book Company, 1974.

Mioduski, R.E., Tables of the Probability Integral of the Central t-Distribution, BRL Technical Note #1570, August 1965.

Schilling, Edward G., Acceptance Sampling in Quality Control, Marcel Dekker, Inc., 1982.

Thomas, D.W., Chairman, Statistical Quality Control Handbook, Western Electric Company, Inc., 1956.

SOME NOTES ON VARIABLE SELECTION CRITERIA FOR REGRESSION MODELS (AN OVERVIEW)

Eugene F. Dutoit
U.S. Army Infantry School
Fort Benning, Georgia

Abstract. There are several decision rules for determining when to enter additional independent variables into linear multiple regression. Three of these are: (1) examining the incremental significance in the multiple correlation coefficients, (2) Mallows' C_p statistic to determine the best combination of independent variables, and (3) considering the changes in magnitude of the standard error of estimate. This paper will examine some of the interrelationships between the three methods cited above. These relationships will be applied to a data set and the results presented.

Acknowledgement. The author wishes to thank Professors Ron Hocking and Emanuel Parzen for their comments and suggestions. It is this spirit of freely shared ideas that makes these "Design of Experiments Conferences" valuable to Army statisticians.

1. The Problem.

Given experimental data in the form:

Y	X_1	X_2	X_3	. . .	X_q
Y_1	X_{11}	X_{12}	X_{13}	. . .	X_{1q}
Y_2	X_{21}	X_{22}	X_{23}	. . .	X_{2q}
Y_3	X_{31}	X_{32}	X_{33}	. . .	X_{3q}
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
Y_n	X_{n1}	X_{n2}	X_{n3}	. . .	X_{nq}

where $(X_1, X_2, X_3, \ldots, X_q)$ are candidate independent variables (that make sense according to some theoretical bases) and Y is the dependant variable. The researcher wants to form some model

$$Y' = a + \sum_{i=1}^{k} b_i X_i \qquad (1)$$

where $k \leq q$. This would indicate the model (equation 1) consisted of the best set of candidate independent variables. This paper will provide an overview of the following measures and criteria in order to shed some light on this problem.

2. The Multiple Correlation Coefficient. (R^2)

a. Incremental Significance. The well known test for the incremental significance in R^2 by adding an additional independent variable X_i into equation (1) is:

$$F = \frac{(R^2_{y.1,2,\ldots,k_1} - R^2_{y.1,2,\ldots,k_2}) / (k_1 - k_2)}{(1 - R^2_{y.1,2,\ldots,k_1}) / (N - K_1 - 1)} \quad (2)$$

where k_1 = number of independent variables for larger R^2
k_2 = number of independent variables for smaller R^2
$R^2_{y.1,2,\ldots,k_1}$ = larger R^2
$R^2_{y.1,2,\ldots,k_2}$ = smaller R^2
N = number of cases

The test follows an F distribution with degrees of freedom equal to $(k_1 - k_2)$, $(N - k_1 - 1)$.

b. Adjusted R^2. As independent variables are added to equation (1), the value of R^2 will also increase. This increase may be small (i.e., statistically not significant). In order to account for this mathematical increase in R^2, the so-called shrinkage formula is used to calculate an adjusted R^2 as:

$$R^2 \text{ adj} = R^2_k - \frac{k\,(1 - R^2_k)}{N - k - 1} \quad (3)$$

where k = number of independent variables in regression.

N = number of cases.

3. Mallows' Cp Statistic. Myers (reference 1) presents the Cp statistic in the following form:

$$Cp = p + \frac{(S^2 - \hat{\sigma}^2)(N - p)}{\hat{\sigma}^2} \quad (4)$$

where $p = k + 1$

$\hat{\sigma}^2$ = estimated variance of the complete model (i.e.; all independent variables included).

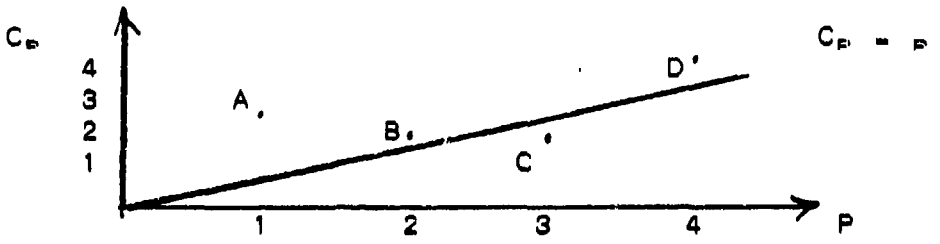
S^2 = estimated variance of the candidate (subset) model.

N = number of cases.

$\hat{\sigma}^2$ and S^2 are obtained as the residual mean squares from the regression ANOVA.

The following interpretation is based on the discussion from the Myers (reference 1) text:

Figure 1



Reference to equation (4) shows that if $S^2 < \hat{\sigma}^2$, the plot of Cp will fall below the line Cp = p. The above inequality is desirable for it states that the variation about some subset regression model is less than the variation about the full model. Only point C in the above diagram meets this condition. This concept will be discussed in the following paragraph concerning the standard error of estimate about regression. It should be noted that if $(S^2 = \hat{\sigma}^2)$, then equation (4) becomes Cp = p. This is always the case for the full model. An alternative format for Cp is given by Daniel and Wood (reference 2) as:

$$Cp = \frac{RSS_k}{\hat{\sigma}^2} - [N - 2p] \qquad (5)$$

where p = k + 1 (as before)

RSS_k = residual sum of squares with k independent variable (p parameters)

$\hat{\sigma}^2$ = residual mean square of the complete model (as before).

It can be shown (not here) that equation (5) is equivalent to equation (4).

3A. Another Alternative Form for Cp. Given q independent variables, the total regression model is:

$$\hat{Y} = a + b_1X_1 + b_2X_2 + \ldots + b_qX_q \qquad (6)$$

The regression ANOVA table is presented below as Figure 2.

Figure 2
Regression ANOVA - Full Model

Source	DF	SS	MS
Explained	q	$(N-1)(S^2_y)(R^2_q)$	
Residual	$\frac{N-q-1}{N-1}$	$[N-q-1]\ S^2_{y\|x_q}$	$S^2_{y\|x_q}$

The model with k independent variables: where k < q is:

$$\hat{y} = a + b_1X_1 + b_2X_2 + \ldots b_kX_k \qquad (7)$$

The regression ANOVA table is given in Figure 3.

Figure 3
Regression ANOVA - Subset Model

Source	DF	SS	MS
Explained	k	$(N-1)(S^2_y)(R^2_k)$	
Residual	$\frac{N-k-1}{N-1}$	$[N-k-1]\ S^2_{y\|x_k}$	$S^2_{y\|x_k}$

where

$$S^2y|X_k = \frac{N-1}{N-k-1}(S^2_y)(1-R^2_k)$$

$$S^2y|X_q = \frac{N-1}{N-q-1}(S^2_y)(1-R^2_q) \qquad (8)$$

Referring to the Myers format of Cp (equation 4) and substituting equations (8) for S^2 and $\hat{\sigma}^2$:

$$Cp = (k+1) + \frac{(S^2y|X_k - S^2Y|X_q)(N-k-1)}{S^2y_k}$$

Further substitution [reference equations (8)] and simple algebra yield another format for the Cp test:

$$Cp = \frac{(N-q-1)(1-R^2_k)}{(1-R^2_q)} - (N-2k-2) \qquad (9)$$

Note that in various forms the Cp test can be expressed as a function of R^2, S^2y X, N, q, k. This leads to another independent decision method, namely the standard error of regression (Sy|X).

4. $\underline{Sy|X_k}$. When performing step-wise regression, the value of $S^2Y|X$ usually gets smaller as independent variables are added to regression. In other words

$$S^2Y|X_k > S^2Y|X_{k+1} \text{ (usually)}. \qquad (10)$$

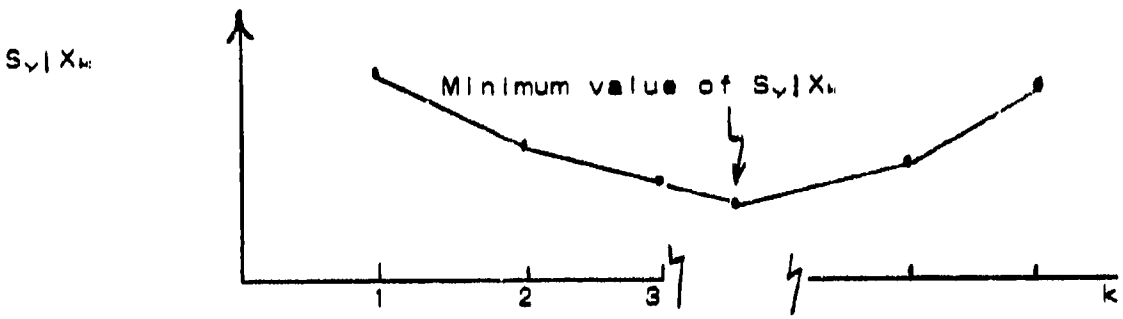
However, this is not always the case. The ratio is actually

$$\frac{S^2Y|X_k}{S^2Y|X_{k+1}} \begin{matrix} > \\ < \end{matrix} 1 \qquad (11)$$

Note that equation (4) allows for cases where some subset model has less variance (S^2) than the variance for the complete

model ($\hat{\sigma}^2$). In this case the Cp plot is below the line (i.e.; point C in Figure 1). This can be expressed in Figure 4:

Figure 4



The minimum value in Figure 4 corresponds to point C in Figure 1. The subset of regression that has minimum variance would be the best predictor of the dependent variable Y. The ratio described by equation (11) can be rewritten to gain some insight to the process. From equations (8) the following expressions can be inferred:

$$S^2Y|X_k = \frac{N-1}{N-k-1} \quad (S^2_Y) \quad (1-R^2_k) \qquad (12)$$

$$S^2Y|X_{k+1} = \frac{N-1}{N-k-2} \quad (S^2Y) \quad (1-R^2_{k+1})$$

The ratio in inequality (11) now becomes:

$$\frac{S^2Y|X_k}{S^2Y|X_{k+1}} = \frac{(N-k-2)}{(N-k-1)} \cdot \frac{1-R^2k}{1-R^2_{k+1}} \qquad (13)$$

where $\frac{N-k-2}{N-k-1} < 1$

and $\frac{1-R^2_k}{1-R^2_{k+1}} > 1$

Therefore, the value of the ratio cited as inequality (11) will depend on the magnitudes of the ratios shown above. Note that equation (3) (R^2_{adj}) contains a shrinkage factor with terms ($1-R^2_k$) and ($N-k-1$). These terms are also contained in equation (13). Intuitively it appears that the adjusted correlation coefficient (R^2_{adj}) should be a maximum for the subset of regression where $S^2Y|X$ is a minimum. The value of Cp should also be minimum for the same subset of independent variables.

5. Example. The following example was taken from Myers (reference 1). It is found on page 110, Table 4-1. The example uses sales data for asphalt shingles obtained from (N=15) districts. The variables considered in this example are:

X_1 = Number of promotional accounts.
X_2 = Number of active accounts.
X_3 = Number of competing brands.
X_4 = District potential.
Y = Sales in thousands of dollars.

The results of a step wise regression are given below:

Figure 5
Stepwise Results

STEP	VARIABLE	SY | X*	ADJ R^2 *	p	C_p *
1	3	49.99	.60935	2	1227.1
2	2,3	6.67	.99303	3	11.4
3	1, 2, 3	4.98	.99612	4	3.4
4	1, 2, 3, 4	5.12	.99590	5	5

Notice that the last step (number 4) has a Cp value equal to p (5). This is always true for the full model. Also note that step 3 is the best subset regression. It is this step (variables X_1, X_2, X_3,) where the values of Sy|x and cp are minimum and adjusted R^2 is maximum.

The results of all variable cases is presented in Figure 6. The combination shown in step 11 is the best subset regression. It is the same combination of optimum values of R^2adj, Sy|x and Cp (variables X_1, X_2, X_3,).

Figure 6
All Cases

STEP	VARIABLE	Sy | x *	R^2	ADJ R^2 *	p	C_p *
1	1	82.49	.01200	-.06400	2	3361
2	2	58.63	.50101	.46263	2	1692
3	3	49.99	.63725	.60935	2	1227.1
4	4	79.05	.09284	.02306	2	3085.1
5	1,2	60.53	.50900	.42716	3	1666.8
6	1,3	51.71	.64169	.58197	3	1213.9
7	1,4	81.26	.11509	-.03240	3	3011.2
8	2,3	6.67	.99403	.99303	3	11.4
9	2,4	60.82	.50422	.42159	3	1683.1
10	3,4	48.83	.68051	.62726	3	1081.4
11	1,2,3	4.98	.99695	.99612	4	3.4
12	1,2,4	62.90	.51395	.38139	4	1651.9
13	1,3,4	50.27	.68959	.60493	4	1052.4
14	2,3,4	6.97	.99404	.99241	4	13.3
15	1,2,3,4	5.12	.99707	.99590	5	5

6. Summary. This paper has examined several methods of determining when to enter additional independent variables into linear multiple regression in order to form a optimum subset from all the candidate variables.

The interrelationships between Cp, Sy|x and adjusted R^2 are studied. These three indicators appear to provide the same information in the model selection decision process. Although they all lead to the same decision regarding the subset regression selection, each measure provides a different perception on the subject.

References

Myers, R; Classical and Modern Regression with Applications. Duxbury Press. Boston, MA; 1986

Daniel, C; Wood, F; Fitting Equations to Data. Wiley, New York, 1971

TWO-STAGE TESTING OF COMBAT VEHICLE TIRE SYSTEMS

Barry A. Bodt

US Army Ballistic Research Laboratory
Aberdeen Proving Ground, Maryland 21005-5066

ABSTRACT

An effort is underway to enhance the battlefield survivability of combat vehicle tire systems. The Army is currently investigating several new tire technologies with regard to their ability to function after battlefield degradation. The tires, in a run-flat condition, must support some vehicle mobility according to that vehicle's mission profile. The immediate objective of this program is choosing, for further research, the most promising among the new tire technologies. The presenter has been tasked to develop an appropriate test plan.

Sound experimental strategy, for this or any study, must be accompanied by a clear understanding of the problem(s) to be resolved. A list of question areas worth exploring to help gain this understanding is suggested by Hahn (Technometrics, 1984) as part of more general guidelines. The presenter demonstrates their usefulness to that end in the above mentioned tire program. The test plan and the process by which it evolved is discussed.

TWO-STAGE TESTING OF COMBAT VEHICLE TIRE SYSTEMS

1. INTRODUCTION

An effort is underway to enhance the battlefield survivability of combat vehicle tire systems. The impetus for current investigations dates back to a 1979 issue paper, submitted to DA by the US Training and Doctrine Command (TRADOC). In 1985 the Tank Automotive Command (TACOM) established a tire task force, the need for which was supported by the results of a 1984 independent evaluation of one tire system performed by the Operational Test and Evaluation Agency (OTEA). OTEA observed that when the run-flat tires for the High Mobility Multi-Purpose Wheeled Vehicle (HMMWV) were run flat for 30 miles, the tires became unserviceable and had to be replaced. The objective of the TACOM Tire Task Force is to identify a survivable tire system (STS) technological replacement which demonstrates acceptable battlefield survivability. A two-phase approach (operational and technical) has been adopted to screen available STS technologies in search of candidates for more intense research and development. The operational phase, considering the standard and seven new STS technologies, was completed by the Combat Developments Experimentation Center (CDEC) in 1987. The technical phase, the focus of this paper, is being conducted by the Vulnerability Lethality Division (VLD) of the Ballistic Research Laboratory (BRL) according to the test plan developed by the Probability and Statistics Branch (PSB) of the Systems Engineering and Concepts Analysis Division (SECAD) of the BRL.

This paper is intended to accomplish two tasks. The first is to discuss the test plan that has been adopted for the technical testing phase -- not in great detail but sufficiently to demonstrate the degree to which experimental objectives are satisfied. As part of the discus-

sion it is shown how, for example, tire performance specifications, factors thought to influence performance, and physical and budgetary constraints are incorporated in the test strategy. The second is to illustrate the usefulness of well-defined consulting guidelines in extracting the necessary information from experimenters. Any sound experimental strategy must be accompanied by a clear understanding of the problem to be resolved, but information essential to that understanding is often difficult to obtain. The fragmented manner in which information is passed from client to consultant inhibits a cogent assimilation of facts needed for efficient problem solving. Hahn (1984) suggests imposing the structure of question area guidelines (see Figure 1) both to help sort the information coming in and to direct consultation sessions down new promising paths.

The remainder of the paper is organized as follows. In Section 2 the problem and test plan are developed, punctuated by Hahn's guidelines. It is hoped that this presentation will both give fair treatment to the Army concern as well as illustrate a reasonable approach to consultation. In Section 3 a brief critique of the test plan's strengths and weaknesses is given, followed by some closing comments.

2. EVALUATION OF THE TEST PLAN

Problem information is divulged in this section according to Hahn's guidelines, and that constitutes our presentation of his technique. We seek only to show how encompassing those question areas are by developing in full the Army's problem through their use. In the text, italicized words and phrases refer back to guidelines in Figure 1. The guidelines have been juggled to allow for a logical presentation and the order shown in Figure 1 corresponds, with few exceptions, to that of this section. This is simply a matter of convenience and not a claim

1. The objectives of the experiment.

2. The variables to be held constant and how this will be accomplished (as well as those that are to be varied).

3. The uncontrolled variables - what they are and which ones are measurable.

4. The response variables and how they will be measured.

5. Special considerations which indirectly impose experimental constraints.

6. The budged size of the experiment and the deadlines that must be met.

7. Conditions within the experimental region where the expected outcome is known; the anticipated performance is expected to be inferior, especially for programs where an optimum is sought; and experimentation is impossible or unsafe.

8. Past test data and, especially, any information about different types of repeatability.

9. The desirability and opportunities for running the experiment in stages.

10. The anticipated complexity of the relationship between the experimental variables and the response variables and any anticipated interactions.

11. The procedures for running a test, including the ease with which each of the variables can be changed from one run to the next.

12. The details of the physical set-up.

Figure 1. Important Question Areas for Statisticians to Address.

for an ideal sequence. In fact, each consulting session is likely not only to naturally gravitate toward different orders but also to move around from area to area, possibly returning several times to some.

2.1. Understanding the Problem

Let us begin by considering *objectives*. We consider two types; military and experimental. The military objective is that HMMWV tires remain serviceable when degraded through battlefield exposure to small caliber munitions and shell fragments. Serviceable means that the tire exhibits performance consistent with the standards specified in the NATO-Finabel 20 A 5 1956 NATO Test Procedure. Summarized expectations set forth therein say that the combat tire must possess (as nearly as possible) the same over the road performance as the classic radial tire in terms of maximum vehicle speed and lateral and longitudinal traction and stability. After degradation, normal military performance of the vehicle is still required when no more than two tires (one drive and one steering) are damaged. The experimental objective is to screen six, including the standard, tire systems with the purpose of selecting a subset for further research, development, and the eventual upgrading of combat tires. The selection criteria must be driven by the military objectives summarized above.

Question areas 2-4 in Figure 1 each concern *variables*. It is in the identification and classification of these variables that the experimental strategy begins to take form. In Table 1 the most important ones are given. Care is taken to initially classify them as candidates for response, design or nuisance variables and to subclassify them for each of the last two categories. The scale of measurement is also noted. A short definition of each of these variables is given in the appendix. Because the variables listed in Table 1 represent only those

Table 1. Relevant Variables.

Relevant Variables	Variable Classification						Measurement Scale		
		Design			Nuisance				
	Response	Fixed	Random	# Levels	Controlled	Uncontrolled	nominal	ordinal	interval
terrain				3	√		√		
tire position		√		3			√		
vehicle				4	√		√		
driver					√		√		
tire motion		√		2			√		
threat obliquity		√		3					√
tire technology		√		6			√		
fragment shape		√					√		
fragment size		√							√
fragment velocity					√				√

Table 1. Relevant Variables. (Continued)

Relevant Variables	Variable Classification						Measurement Scale		
		Design			Nuisance				
	Response	Fixed	Random	# Levels	Controlled	Uncontrolled	nominal	ordinal	interval
road temperature						✓			✓
tire pressure		✓		2					✓
vehicle load					✓				✓
driving speeds					✓				✓
distance to tire					✓				✓
round					✓		✓		
miles until failure	✓								✓
delivery method		✓		4			✓		
# shots on tire		✓		2					
subjective assessments	✓							✓	

considered essential, all must be incorporated in the experimental strategy. We briefly discuss several of them here so that the reader may gain a sense of the complexity of the problem.

The logical starting point for discussion is with tire technology, for it is the selection from among these prototypes that is the objective of this experiment. Six manufacturer offerings, including the standard, are to be considered, but there are basically only four technologies. When combat tires are exposed to small caliber munitions and shell fragments they will surely tear, puncture, or in some other way be damaged so as to induce partial or complete deflation. Then in order for military objectives to be satisfied, the survivable tire will either successfully negate this damage or be structurally capable of supporting vehicle mobility without benefit of full tire pressure. Taking the first tact, the sealant tire systems contain chemical compounds which are intended to flow to a source of air loss, solidify, and thereby negate the threat damage. Run-flats take the second tact and are able to support the vehicle with a metal or plastic insert which acts in the tires stead when the tire is deflated. Self-supporting tires are so named because molded into the tread is a rigid fiber glass band, designed to carry the tire's full load in the absence of tire pressure. Solid urethane tires circumvent the problem by containing no air to be lost, but they do so at the cost of additional weight, inhibiting vehicle mobility (Drelling et al., 1987).

A limitation of the CDEC exercise is that tire degradation from shell fragments is not considered. Interest in the more irregular punctures and tears caused by the shell fragment threat is a consideration in involving the BRL in the technical phase of experimentation. To make inferences about tire performance after fragment damage the consensus is that either live shells should be detonated near standing tires or the expected resulting fragments should

be simulated. *A special consideration* in long range plans is that an acceptance test consistent with current testing be developed. Due to the repeatability requirements inherent in acceptance testing the shell-detonation approach was dropped in favor of fragment simulation

This decision led to variables involving fragment shape, size, and velocity. Due to *budget and time constraints* it appears unreasonable to select many values for each and then proceed in a factorial manner when incorporating them in the design. Rather we option to create two fragment profiles, each representative of a distinct threat. Avoiding great detail, a standard BRL fragment design is specified for shape. Velocity and mass are determined as follows. Each are a function of the distance between the shell at detonation and the tire. The distance selected corresponds to a 50% mobility kill for the vehicle according to models accessed by the Army Material Systems Analysis Activity (AMSAA). *Avoiding the experimental region where the expected outcome is known,* we do not consider distances so close that the personnel are not likely to survive. The median velocity and mass among computer simulated fragments possessing an appropriate trajectory then serve as representative values for these characteristics. Trial firings suggest some deviations in these choices so that resulting damage seemed similar to actual fragment damage previously observed.

Other factors of keen interest include the terrain traveled and tire position, motion, and pressure. The mission profile for the HMMWV dictates that it shall be able to travel primary roads, secondary roads, and cross country. Further it suggests that in a characteristic mission those three terrains might comprise 30%, 30%, and 40%, respectively, of the total mileage covered Tire position refers to its designation as a drive tire or a drive and steering tire; the HMMWV is 4-wheel drive. In addition to this one-at-a-time damage, recall that the NATO Finabel standards require acceptable performance when two tires on the vehicle are

damaged. When attacked, the HMMWV may be moving or at rest. Proponents of the sealant technology claim that if the tire is in motion when punctured, then the sealant mechanism will be more effective in finding and repairing the damage. *Past test data* indicates that tire pressure may influence the type of puncture, that is, clean or ragged. Manufacturer recommended high and low pressures for each tire will be considered.

The *special consideration* that this experiment complement the CDEC exercise fixed two important test definitions. TACOM decided that the response would remain defined as miles until failure. Failure occurs when either the tire begins to come apart when in use or the operator must slow to less than 50% of normal operating speed in order to maintain control. Under a rigid value for normal speed, failure could depend on the size and strength of the operator. We propose to account for that by establishing a profile on operators (actually driving teams) in their normal operation of the vehicle. The 50% rule is then based on normal team performance. Driving teams are established to avoid failure due to fatigue. *Past test data* reveals that some degraded tires remain serviceable after 100 continuous miles of operation. In order to avoid truncated data, the test course is extended to 125 continuous miles, but at the additional cost of trial time. It is felt that if two operators are allowed to rotate after each 25 mile lap, then fatigue will not enter into the failure determination.

2.2 Test Plan

The test plan will be implemented in *stages*. A fairly large number of experimental conditions define the experiment outlined in Section 2.1. To examine each of these conditions in a factorial manner will require more resources than the *experimental budget* will allow; for all but the standard tire no more than 30 prototypes will be made available. Moreover, recall

that the principal objective of this study is to facilitate comparison among tires. Placing too much emphasis (sample allocation) on ancillary issues may partially obscure (weaken conclusions regarding) the main experimental focus. For these reasons, resource limitations and emphasis, we choose to run the experiment in two primary stages.

The division of testing meets the above concerns. In stage 1 all the experimental conditions are incorporated in the design as factors or fixed test procedures. Only the standard HMMWV tire is considered in stage 1. The purpose of this stage is two-fold. First the various test conditions may be examined. It is hoped that some will prove unnecessary for inclusion in stage 2, thereby increasing the experimental information per sampled test condition. Second, test procedures may be smoothed. Field test exercises nearly always present unexpected problems, often resulting in samples which must be invalidated for the analysis. Here we run only the risk of wasting some more plentiful, standard tires instead of the scarce new prototypes. In stage 2 the prototypes will be examined by an experienced testing group under the conditions remaining after stage 1. Since the complete details will not be available until stage 1 is concluded, we defer further discussion of stage 2 to future papers. In the remainder of this paper stage 1 testing serves as the main focus.

Stage 1 will be run as a 1/2 replication of a $4x2^4$ factorial design, requiring 32 observations. The design factors, each discussed in Section 2.1, are listed in Table 2. The 4 levels for threat include a 7.62mm round fired at 45° and 90° obliquity on the sidewall, a small fragment simulator, and a large fragment simulator. Note that only 2 tire position levels, drive or steering, are considered. The case in which two tires are damaged, requiring twice as many samples, is handled only in a limited sense. Imbedded in the stage 1 factorial design are four treatment combinations having two damaged tires which arise from a 1/2 replicate of a 2^4

Table 2. 1/2 Replication of $4 \times 2^4 = 32$ Observations + 4 Observations for 2 Tires Damaged.

Relevant Variables	Variable Classification						Measurement Scale		
		Design			Nuisance				
	Response	Fixed	Random	# Levels	Controlled	Uncontrolled	nominal	ordinal	interval
miles until failure	✓								✓
threat		✓		4			✓		
tire pressure		✓		2			✓		
tire motion		✓		2			✓		
tire position		✓		2			✓		
driving team		✓		2			✓		

design. The remaining 4 observations are already included in the principal stage 1 design. The other three factors are handled as previously noted. This design allows hypothesis tests on all main effects and on most first-order interactions. Due to the *anticipated complexity of the relationship among variables* some first-order interactions can be sacrificed to the experimental error formed by the remaining second, and above, order interactions. The remaining variables are addressed in stage 1 as suggested in Figure 2.

Randomization for the stage 1 design is complete except in the case of driving team, vehicle, and terrain. In consideration of the *procedures for running a test and the ease in which variables can be changed* and the *details of the physical set up* some compromises were made. The complete randomization of the driving teams is not possible because both teams are to be used simultaneously. The first driver in the rotation for each team was randomized. As indicted in Figure 2, four vehicles are used but are not included as test factors. To mitigate their effect on the outcome, they have been selected according to age and state of repair and have been partially randomized over the design. Also noted in Figure 2, the three terrains mentioned in Section 2.1 comprise the test track. The course layout attempts to mix or randomize the terrains so that not all the mileage for one type will be traveled before the next is encountered.

3. CRITIQUE OF APPROACH

In this section we address the primary advantages and disadvantages of the test plan interpreted in terms of the stated military and experimental objectives and follow with some comments pertaining to the consulting technique employed. Beginning with the military objectives, all of the variables considered important by TACOM or the NATO Finable Stan-

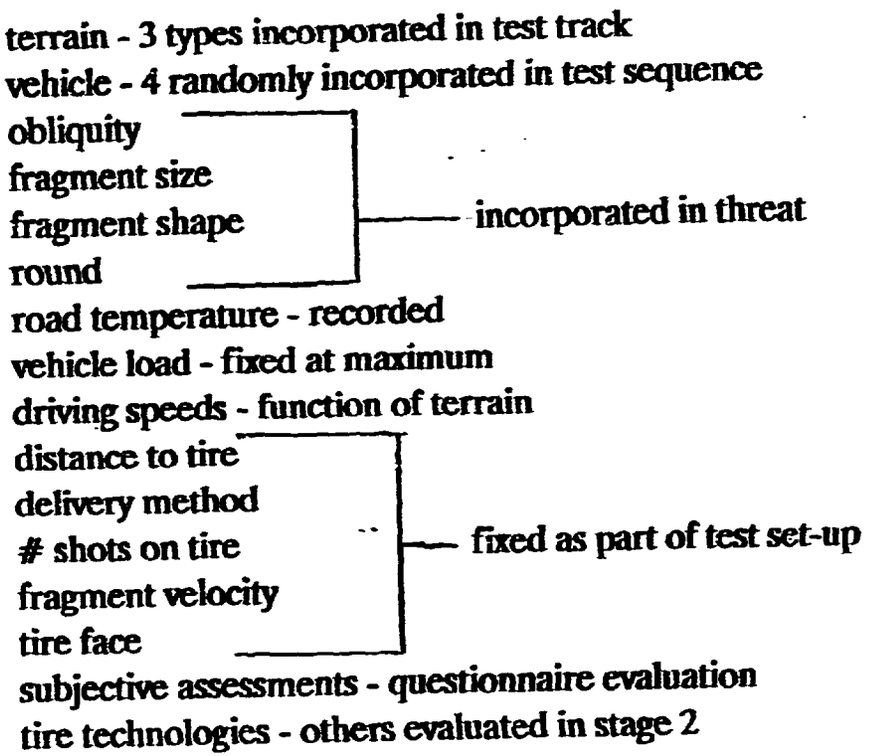


Figure 2. Incorporating Remaining Variables in the Design.

dards are included in the test plan in a manner suitable to TACOM. Sometimes this requires compromise, such as in the use of terrain. Terrain is considered only through its inclusion in the test course in proportions consistent with the HMMWV mission profile. For some other variables the military interests are clarified in the test plan. For example, normal operating speeds in the failure definition are more sensibly tied to the normal performance of individual driving teams. Also, efforts to handle the fragment threat result in a reasonable fragment simulation procedure which may be used in follow-on acceptance testing.

With regard to experimental objectives, the selection of STS prototypes for further research and development follows directly from analysis of the second testing stage. Further, the stage 1 plan imposes an analyzable design structure on a complex problem providing for the testing of all important hypotheses. In addition, running the experiment in stages has the emphasis and resource advantages mentioned in Section 2.2. However, the test plan has several weaknesses. By examining the standard tire only in stage 1, comparisons between it and other STS prototypes are hindered. Experimental error is an issue since complete randomization is not possible and since some pooling of low-order interactions into the error term may be necessary. Choice of an error term for the imbedded test of the two-tire effect is far from straight-forward, particularly since four of the eight observations must be used in the analysis twice. Finally, we had to take some liberties in the combination of variables to form factors so that a design would be possible with the available samples.

As to consulting, we cannot prove the usefulness of Hahn's guidelines, but we hope that the illustration is convincing. Surely, the information can be obtained through other methods, but the imposed structure of this approach facilitates a very comprehensive investigation. In the end, all methods must be judged by the experimental strategies which they help to

develop, but their performance is hopelessly confounded with the skills of the consultant using them. Of course the purposes of those strategies are to meet both application objectives and satisfy statistical theory. Whether this strategy satisfies those purposes, and if not, whether fault lies with the consultant, the approach, or the problem are questions left for the reader to decide.

REFERENCES

Drelling, J.S., Pietzyk, S., Schrag, H. (1987), "Survivable Tire Test," CDEC-TR-87-014.

Hahn, G. J. (1984), "Experimental Design in the Complex World," Technometrics, 26, 19-31.

Kempthorne, O. (1952), Design and Analysis of Experiments, New York; John Wiley & Sons.

APPENDIX

terrain - the three driving surfaces listed as primary road, secondary road, and cross country. Each will induce different tire stress and all are included in the vehicle's mission profile.

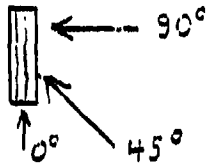
tire position - placement of the damaged tires. In testing, the right front, right rear, or both may be degraded.

vehicle - High Mobility Multi-purpose Wheeled Vehicle (HMMWV). This vehicle's tire system is the program's focus. The individual HMMWV effect is an issue.

driver - operator of the vehicle. The influence of different drivers should be accounted for.

tire motion - the state of the tire, either static or rolling.

threat obliquity -



tire technology - six Survivable Tire Systems to be compared.

fragment shape - partially determines the nature of the tear or puncture.

fragment size - partially determines the nature of the tear or puncture.

fragment velocity - partially determines the nature of the tear or puncture.

road temperature - effects the tire vulnerability. Previous testing revealed reduced susceptibility to puncture with low tire pressures.

vehicle load - influences tire stress.

driving speed - influences tire stress.

distance to tire - refers to distance traveled by fragment or small caliber munition.

round - threat munition(s) to be used.

miles until failure - response.

delivery method - fragment propelling methods considered.

# shots on tire	-	the number of punctures to be made in each tire to obtain its degraded state.
subjective assessments	-	comments solicited from drivers regarding handling of vehicle when tires are in normal or degraded mode.

Parallel Coordinate Densities[1]

Edward J. Wegman
Center for Computational Statistics
242 Science-Technology Building
George Mason University
Fairfax, VA 22030

1. *Introduction.* The classic scatter diagram is a fundamental tool in the construction of a model for data. It allows the eye to detect such structures in data as linear or nonlinear features, clustering, outliers and the like. Unfortunately, scatter diagrams do not generalise readily beyond three dimensions. For this reason, the problem of visually representing multivariate data is a difficult, largely unsolved one. The principal difficulty, of course, is the fact that while a data vector may be arbitrarily high dimensional, say n, Cartesian scatter plots may only easily be done in two dimensions and, with computer graphics and more effort, in three dimensions. Alternative multidimensional representations have been proposed by several authors including Chernoff (1973), Fienberg (1979), Cleveland and McGill (1984a) and Carr et al. (1986).

An important technique based on the use of motion is the computer-based kinematic display yielding the illusion of three dimensional scatter diagrams. This technique was pioneered by Friedman and Tukey (1973) and is now available in commercial software packages (Donohoe's MacSpin and Velleman's Data Desk). Coupled with easy data manipulation, the kinematic display techniques have spawned the exploitation of such methods as projection pursuit (Friedman and Tukey, 1974) and the grand tour (Asimov, 1985). Clearly, projection-based techniques lead to important insights concerning data. Nonetheless, one must be cautious in making inferences about high dimensional data structures based on projection methods alone. It would be highly desireable to have a simultaneous representation of all coordinates of a data vector especially if the representation treated all components in a similar manner. The cause of the failure of the standard Cartesian coordinate representation is the requirement for orthogonal coordinate axes. In a 3-dimensional world, it is difficult to represent more than three orthogonal coordinate axes. We propose to give up the orthogonality requirement and replace the standard Cartesian axes with a set of n parallel axes.

2. *Parallel Coordinates.* We propose as a multivariate data analysis tool the following representation. In place of a scheme trying to preserve orthogonality of the n-dimensional coordinate axes, draw them as parallel. A vector $(x_1, x_2, \ldots, x_n)$ is plotted by plotting x_1 on axis 1, x_2 on axis 2 and so on through x_n on axis n. The points plotted in this manner are joined by a broken line. Figure 2.1 illustrates two points (one solid, one dashed) plotted in parallel coordinate representation. In this illustration, the two points agree in the fourth coordinate. The principal advantage of this plotting device is clear. Each vector $(x_1, x_2, \ldots, x_n)$ is represented in a planar diagram so that each vector component has essentially the same representation.

The parallel coordinates proposal has its roots in a number of sources. Griffen (1958) considers a 2-dimensional parallel coordinate type device as a method for graphically computing the Kendall tau correlation coefficient. Hartigan (1975) describes the "profiles algorithm" which he describes as "histograms on each variable connected between variables by identifying cases." Although he does not recommend drawing all profiles, a profile diagram with all profiles plotted is a parallel coordinate plot. There is however far more mathematical structure, particularly high dimensional structure, to the parallel coordinate diagram than Hartigan exploits. Inselberg (1985) originated the parallel coordinate

[1]This research was sponsored by the Army Research Office, Contract DAAL03-87-K-0087

representation as a device for computational geometry. His 1985 paper is the culmination of a series of technical reports dating from 1981. Finally we note that Diaconis and Friedman (1983) discuss the so-called M and N plots. Their special case of a 1 and 1 plot is a parallel coordinate plot in two dimensions. Indeed, the 1 and 1 plot is sometimes called a before-and-after plot and has a much older history. The fundamental theme of this paper is that the transformation from Cartesian coordinates to parallel coordinates is a highly structured mathematical transformation, hence, maps mathematical objects into mathematical objects. Certain of these can be given highly useful statistical interpretations so that this representation becomes a highly useful data analysis tool.

3. *Parallel Coordinate Geometry.* The parallel coordinate representation enjoys some elegant duality properties with the usual Cartesian orthogonal coordinate representation. Consider a line $\mathcal{L}$ in the Cartesian coordinate plane given by $\mathcal{L}$: $y=mx+b$ and consider two points lying on that line, say $(a, ma+b)$ and $(c, mc+b)$. For simplicity of computation we consider the xy Cartesian axes mapped into the xy parallel axes as described in Figure 3.1. We superimpose a Cartesian coordinate axes t,u on the xy parallel axes so that the y parallel axis has the equation $u=1$. The point $(a, ma+b)$ in the xy Cartesian system maps into the line joining $(a, 0)$ to $(ma+b, 1)$ in the tu coordinate axes. Similarly, $(c, mc+b)$ maps into the line joining $(c, 0)$ to $(mc+b, 1)$. It is a straightforward computation to show that these two lines intersect at a point (in the tu plane) given by $\overline{\mathcal{L}}$: $(b(1-m)^{-1}, (1-m)^{-1})$. Notice that this point in the parallel coordinate plot depends only on m and b the parameters of the original line in the Cartesian plot. Thus $\overline{\mathcal{L}}$ is the dual of $\mathcal{L}$ and we have the interesting duality result that points in Cartesian coordinates map into lines in parallel coordinates while lines in Cartesian coordinates map into points in parallel coordinates.

For $0 < (1-m)^{-1} < 1$, m is negative and the intersection occurs between the parallel coordinate axes. For $m=-1$, the intersection is exactly midway. A ready statistical interpretation can be given. For highly negatively correlated pairs, the dual line segments in parallel coordinates will tend to cross near a single point between the two parallel coordinate axes. The scale of one of the variables may be transformed in such a way that the intersection occurs midway between the two parallel coordinate axes in which case the slope of the linear relationship is negative one.

In the case that $(1-m)^{-1}<0$ or $(1-m)^{-1}>1$, m is positive and the intersection occurs external to the region between the two parallel axes. In the special case $m=1$, this formulation breaks down. However, it is clear that the point pairs are $(a, a+b)$ and $(c, c+b)$. The dual lines to these points are the lines in parallel coordinate space with slope b^{-1} and intercepts $-ab^{-1}$ and $-cb^{-1}$ respectively. Thus the duals of these lines in parallel coordinate space are parallel lines with slope b^{-1}. We thus append the ideal points to the parallel coordinate plane to obtain a projective plane. These parallel lines intersect at the ideal point in direction b^{-1}. In the statistical setting, we have the following interpretation. For highly positively correlated data, we will tend to have lines not intersecting between the parallel coordinate axes. By suitable linear rescaling of one of the variables, the lines may be made approximately parallel in direction with slope b^{-1}. In this case the slope of the linear relationship between the rescaled variables is one. See Figures 3.2 for an illustration of large positive and large negative correlations. Of course, nonlinear relationships will not respond to simple linear rescaling. However, by suitable nonlinear transformations, it should be possible to transform to linearity. The point-line, line-point duality seen in the transformation from Cartesian to parallel coordinates extends to conic sections. An instructive computation involves computing in the parallel coordinate space the image of an ellipse which turns out to be a general hyperbolic form. For purposes of conserving space we do not provide the details here.

It should be noted, however, that the solution to this computation is not a locus of points, but a locus of lines, a line conic. The envelope of this line conic is a point conic. In the case of this computation, the point conic in the original Cartesian coordinate plane is an ellipse, the image in the parallel coordinate plane is as we have just seen a line hyperbola with a point hyperbola as envelope. Indeed, it is true that a conic will always map into a conic and, in particular, an ellipse will always map into a hyperbola. The converse is not true. Depending on the details, a hyperbola may map into an ellipse, a parabola or another hyperbola. A fuller discussion of projective transformations of conics is given by Dimsdale (1984). Inselberg (1985) generalizes this notion into parallel coordinates resulting

in what he calls hstars.

We mentioned the duality between points and lines and conics and conics. It is worthwhile to point out two other nice dualities. Rotations in Cartesian coordinates become translations in parallel coordinates and vice versa. Perhaps more interesting from a statistical point of view is that points of inflection in Cartesian space become cusps in parallel coordinate space and vice versa. Thus the relatively hard-to-detect inflection point property of a function becomes the notably more easy to detect cusp in the parallel coordinate representation. Inselberg (1985) discusses these properties in detail.

4. *Further Statistical Interpretations.* Since ellipses map into hyperbolas, we can have an easy template for diagnosing uncorrelated data pairs. Consider Figure 3.2. With a completely uncorrelated data set, we would expect the 2-dimensional scatter diagram to fill substantially a circumscribing circle. As illustrated in Figure 3.2, the parallel coordinate plot would approximate a figure with a hyperbolic envelope. As the correlation approaches negative one, the hyperbolic envelope would deepen so that in the limit we would have a pencil of lines, what we like to call the cross-over effect. As the correlation approaches positive one, the hyperbolic envelope would widen with fewer and fewer cross-overs so that in the limit we would have parallel lines. Thus correlation structure can be diagnosed from the parallel coordinate plot. As noted earlier, Griffen (1958) used this as a graphical device for computing the Kendall tau.

Griffen, in fact, attributes the graphical device to Holmes (1928) which predates Kendall's discussion. The computational formula is

$$r = 1 - \frac{4X}{n(n-1)}$$

where X is the number of intersections resulting by connecting the two rankings of each member by lines, one ranking having been put in natural order. While the original formulation was framed in terms of ranks for both x and y axes, it is clear that the number of crossings is invariant to any monotone increasing transformation of either x or y, the ranks being one such transformation. Because of this scale invariance, one would expect rank-based statistics to have an intimate relationship to parallel coordinates.

It is clear that if there is a perfect positive linear relationship with no crossings, then $X = 0$ and $r = 1$. Similarly, if there is a perfect negative linear relationship, Figure 3.2 is again appropriate and we have a pencil of lines. Since every line meets every other line, the number of intersections is $\binom{n}{2}$ so that

$$r = 1 - \frac{4\binom{n}{2}}{n(n-1)} = -1.$$

It should be further noted that clustering is easily diagnosed using the parallel coordinate representation.

So far we have focused primarily on pairwise parallel coordinate relationships. The idea however is that we can, so to speak, stack these diagrams and represent all n dimensions simultaneously. Figure 4.1 thus illustrates 6-dimensional Gaussian uncorrelated data plotted in parallel coordinates. A 6-dimensional ellipsoid would have a similar general shape but with hyperbolas of different depths. This data is deep ocean acoustic noise and is illustrative of what might be expected.

Figure 4.2 is illustrative of some data structures one might see in a five-dimensional data set. First it should be noted that the plots along any given axis represent dot diagrams (a refinement of the histograms of Hartigan), hence convey graphically the one-dimensional marginal distributions. In this illustration, the first axis is meant to have an approximately normal distribution shape while axis two the shape of the negative of a χ^2. As discussed above, the pairwise comparisons can be made. Figure

4.2 illustrates a number of instances of linear (both negative and positve), nonlinear and clustering situations. Indeed, it is clear that there is a 3-dimensional cluster along coordinates 3, 4 and 4.

Consider also the appearance of a mode in parallel coordinates. The mode is, intuitively speaking, the location of the most intense concentration of probability. Hence, in a sampling situation it will be the location of the most intense concentration of observations. Since observations are represented by broken line segments, the mode in parallel coordinates will be represented by the most intense bundle of broken line paths in the parallel coordinate diagram. Roughly speaking, we should look for the most intense flow through the diagram. In Figure 4.2, such a flow begins near the center of coordinate axis one and finishes on the left-hand side of axis five.

Figure 4.2 thus illustrates some data analysis features of the parallel coordinate representation including the ability to diagnose one-dimensional features (marginal densities), two-dimensional features (correlations and nonlinear structures), three-dimensional features (clustering) and a five-dimensional feature (the mode). In the next section of this paper we consider a real data set which will be illustrative of some additional capabilities.

5. *An Auto Data Example.* We illustrate parallel cooordinates as an exploratory analysis tool on data about 86 1980 model year automobiles. They consist of price, miles per gallon, gear ratio, weight and cubic inch displacement. For n = 5, 3 presentations are needed to present all pairwise permutations. Figures 5.1, 5.2 and 5.3 are these three presentations. In Figure 5.1, perhaps the most striking feature is the cross-over effect evident in the relationship between gear ratio and weight. This suggests a negative correlation. Indeed, this is reasonable since a heavy car would tend to have a large engine providing considerable torque thus requiring a lower gear ratio. Conversely, a light car would tend to have a small engine providing small amounts of torque thus requiring a higher gear ratio.

Consider as well the relationship between weight and cubic inch displacement. In this diagram we have a considerable amount of approximate parallelism (relatively few crossings) suggesting positive correlation. This is a graphic representation of the fact that big cars tend to have big engines, a fact most are prepared to believe. Quite striking however is the negative slope going from low weight to moderate cubic inch displacement. This is clearly an outlier which is unusual in neither variable but in their joint relationship. The same observation is highlighted in Figure 5.2.

The relationship between miles per gallon and price is also perhaps worthy of comment. The left-hand side shows an approximate hyperbolic boundary while the right-hand side clearly illustrates the cross-over effect. This suggests for inexpensive cars or poor mileage cars there is relatively little correlation. However, costly cars almost always get relatively poor mileage while good gas mileage cars are almost always relatively inexpensive.

Turning to Figure 5.2, the relationship between gear ratio and miles per gallon is instructive. This diagram is suggestive of two classes. Notice that there are a number of observations represented by line segments tilted slightly to the right of vertical (high positive slope) and a somewhat larger number with a negative slope of about -1. Within each of these two classes we have approximate parallelism. This suggests that the relationship between gear ratios and miles per gallon is approximately linear, a believable conjecture since low gears = big engines = poor mileage while high gears = small engines = good mileage. What is intriguing, however, is that there seems to be really two distinct classes of automobiles each exhibiting a linear relationship, but with different linear relationships within each class.

Indeed in Figure 5.3, the third permutation, we are able to highlight this separation into two classes in a truly 5-dimensional sense. The shaded region in Figure 5.3 describes a class of vehicles with relatively poor gas mileage, relatively heavy, relatively inexpensive, relatively large engines and relatively low gear ratios. Figure 5.4 is a repeat of this graphic but with different shading highlighting a class of vehicles with relatively good gas mileage, relatively light weight, relatively inexpensive, relatively small engines and relatively high gear ratios. In 1980, these two characterizations describe respectively domestic automobiles and imported automobiles.

6. *Graphical Extensions of Parallel Coordinate Plots.* The basic parallel coordinate idea

suggests some additional plotting devices. We call these respectively the Parallel Coordinate Density Plots, Relative Slope Plots and Color Histograms. These are extensions of the basic idea of parallel coordinates, but structured to exploit additional features or to convey certain information more easily.

6.1 Parallel Coodinate Density Plots. While the basic parallel coordinate plot is a useful device itself, like the conventional scatter diagram, it suffers from heavy overplotting with large data sets. In order to get around this problem, we use a parallel coordinate density plot which is computed as follows. Our algorithm is based on the Scott (1985) notion of average shifted histogram (ASH) but adapted to the parallel coordinate context. As with an ordinary two dimensional histogram, we decide on appropriate rectangular bins. A potential difficulty arises because a line segment representing a point may appear in two or more bins in the same horizontal slice. Obviously if we have k n-dimensional observations, we would like to form a histogram based on k entries. However, since the line segment could appear in two or more bins in a horizontal slice, the count for any given horizontal slice is at least k and may be bigger. Moreover, every horizontal slice may not have the same count. To get around this, we convert line segments to points by intersecting each line segment with a horizontal line passing through the middle of the bin. This gives us an exact count of k for each horizontal slice. We construct an ASH for each horizontal slice (typically averaging 5 histograms to form our ASH). We have used contours to represent the two-dimensional density although gray scale shading could be used in a display with sufficient bit-plane memory. An example of an parallel coordinate density plot is given in Figure 6.1. Parallel coodinate density plots have the advantage of being graphical representations of data sets which are simultaneously high dimensional and very large.

6.2 Relative Slope Plots. We have already seen that parallel line segments in a parallel coordinate plot correspond to high positive correlation (linear relationship). As in our automobile example, it is possible for two or more sets of linear relationships to exist simultaneously. In an ordinary parallel coordinate plot, we see these as sets of parallel lines with distinct slopes. The work of Cleveland and McGill (1984b) suggests that comparison of slopes (angles) is a relatively inaccurate judgement task and that it is much easier to compare magnitudes on the same scale. The relative slope plot is motivated by this. In an n-dimensional relative slope plot there are $n-1$ parallel axes, each corresponding to a pair of axes, say x_i and x_j, with x_j regarded as the lower of the two coordinate axes. For each observation, the slope of the line segment between the pair of axes is plotted as a magnitude between -1 and $+1$. The maximum positive slope is coded as $+1$, the minimum negative slope as -1 and a slope of ∞ as 0. The magnitude is calculated as $\cos \eta$ where η is the angle between the x_j axis and the line segment corresponding to the observation. Each individual observation in the relative slope plot corresponds to a vertical section through the axis system. An example of a relative slope plot is given in Figure 6.2. Notice that since slopes are coded as heights, simply laying a straightedge will allow us to discover sets of linear relationships within the pair of variables x_i and x_j.

6.3 Color Histograms. The basic set-up for the color histogram is similar to the relative slope plots. For an n-dimensional data set, there are n parallel axes. A vertical section through the diagram corresponds to an observation. The idea is to code the magnitude of an observation along a given axis by a color bin, the colors being chosen to form a color gradient. We typically choose 8 to 15 colors. The diagram is drawn by choosing an axis, say x_k, and sorting the observations in ascending order. Along this axis, we see blocks of color arranged according to the color gradient with the width of the block being proportional to the number of observations falling into the color bin. The observations on the other axes are arranged in the order corresponding to the x_k axis and color coded according to their magnitude. Of course, if the same color gradient shows up say on the x_m axis as on the x_k, then we know x_k is positively "correlated" with x_m. If the color gradient is reversed, we know the "correlation" is negative. We used the phrase "correlation" advisedly since in fact if the color gradient is the same but the color block sizes are different, the relationship is nonlinear. Of course if the x_m axis shows color speckle, there is no "correlation" and x_k is unrelated to x_m. An example of a color histogram is given in Figure 6.3 (for purposes of reproduction here it is really a gray-scale histogram).

7. *Implementations and Experiences.* Our parallel coordinates data analysis software has been implemented in two forms, one a PASCAL program operating on the IBM RT under the AIX operating system. This code allows for up to four simultaneous windows and offers simultaneous

display of parallel coordinates and scatter diagram displays. It offers highlighting, zooming and other similar features and also allows the possibility of nonlinear rescaling of each axis. It incorporates axes permutations and also includes Parallel Coordinate Density Plots, Relative Slope Plots and Color Histograms.

Our second implementation is under development in PASCAL for MS-DOS machines and includes similar features. In addition, it has a mouse-driven painting capability and can do real-time rotation of 3-dimensional scatterplots. Both programs use EGA graphics standards, with the second also using VGA or Hercules monochrome standards.

We regard the parallel coordinate representation as a device complementary to scatterplots. A major advantage of the parallel coordinate representation over the scatterplot matrix is the linkage provided by connecting points on the axes. This linkage is difficult to duplicate in the scatterplot matrix. Because of the projective line-point duality, the structures seen in a scatterplot can also be seen in a parallel coordinate plot. Moreover, the work of Cleveland and McGill (1984b) suggests that it is easier and more accurate to compare observations on a common scale. The parallel coordinate plot and the derivatives of it de facto have a common scale and so for example a sense of variability and central tendency among the variables are easier to grasp visually in parallel coordinates when compared with the scatterplot matrix. On the other hand, one might interpret all the ink generated by the lines as a significant disadvantage of the parallel coordinate plot. Our experience on this is mixed. Certainly for large data sets on hard copy this is a problem. When viewed on an interactive graphics screen particularly a high resolution screen, we have often found that individual points in a scatterplot can get lost because they are simply not bright enough. That does not happen in a parallel coordinate plot. However, if many points are plotted in monochrome, it is hard to distinguish between points. We have gotten around this problem by plotting distinct points in different colors. In an EGA implementation, this means 16 colors. This is surprisingly effective in separating points. In one experiment, we plotted 5000 5-dimensional random vectors using 16 colors, and inspite of total overplotting, we were still able to see some structure. In data sets of somewhat smaller scale, we have implement a scintillation technique. With this technique, when there is overplotting we cause the screen view to scintillate between the colors representing the overplotted points. The speed of scintillation is is proportional to the number of points overplotted and by carefully tracing colors, one can follow an individual point through the entire diagram.

We have found painting to be an extraordinarily effective technique in parallel coordinates. We have a painting scheme that not only paints all lines within a given rectangular area, but also all line lying between to slope constraints. This is very effective in separating clusters. We also use invisible paint to eliminate observation points from the data set temporarily. This is a natural way of doing a subset selection.

References

Asimov, Daniel (1985), "The grand tour: a tool for viewing multidimensional data," *SIAM J. Scient. Statist. Comput.*, 6, 128-143.

Carr, D. B., Nicholson, W. L., Littlefield, R., Hall, D. L. (1986), "Interactive color display methods for multivariate data," in *Statistical Image Processing and Graphics*, (Wegman, E. and DePriest, D., eds.), New York: Marcel Dekker, Inc.

Chernoff, H. (1973), "Using faces to represent points in k-dimensional space," *J. Am. Statist. Assoc.*, 68, 361-368.

Cleveland, W. S. and McGill, R. (1984a), "The many faces of the scatterplot," *J. Am. Statist. Assoc.*, 79, 807-822.

Cleveland, W. S. and McGill, R. (1984b), "Graphical perception: theory, experimentation, and application to the development of graphical methods," *J. Am. Statist. Assoc.*, 79, 531-554

Diaconis, P. and Friedman, J. (1983), "M and N plots," in *Recent Advances in Statistics*, 425-447, New York: Academic Press, Inc.

Dimsdale, B. (1984), "Conic transformations and projectivities," IBM Los Angeles Scientific Center Report #6320-2753.

Fienberg, S. (1979), "Graphical methods in statistics," *Am. Statistician*, 33, 165-178.

Friedman, J. and Tukey, J. W. (1973), "PRIM-9" a film produced by Stanford Linear Accelerator Center, Stanford, CA Bin 88 Productions, April, 1973.

Friedman, J. and Tukey, J. W. (1974), "A projection pursuit algorithm for exploratory data analysis," *IEEE Trans. Comput.*, *C*-23, 881-889.

Griffen, H. D. (1958), "Graphic computation of tau as a coefficient of disarray," *J. Am. Statist. Assoc.*, 53, 441-447.

Hartigan, John A. (1975), *Clustering Algorithms*, New York: John Wiley and Sons, Inc.

Holmes, S. D. (1928), "Appendix B: a graphical method for estimating R for small groups," 391-394 in *Educational Psychology* (Peter Sandiford, auth.), New York: Longmans, Green and Co.

Inselberg, A. (1985), "The plane with parallel coordinates," *The Visual Computer*, 1, 69-91.

Scott, D. W. (1985), "Average shifted histograms: effective nonparametric density estimators in several dimensions," *Ann. Statist.*, 13, 1024-1040.

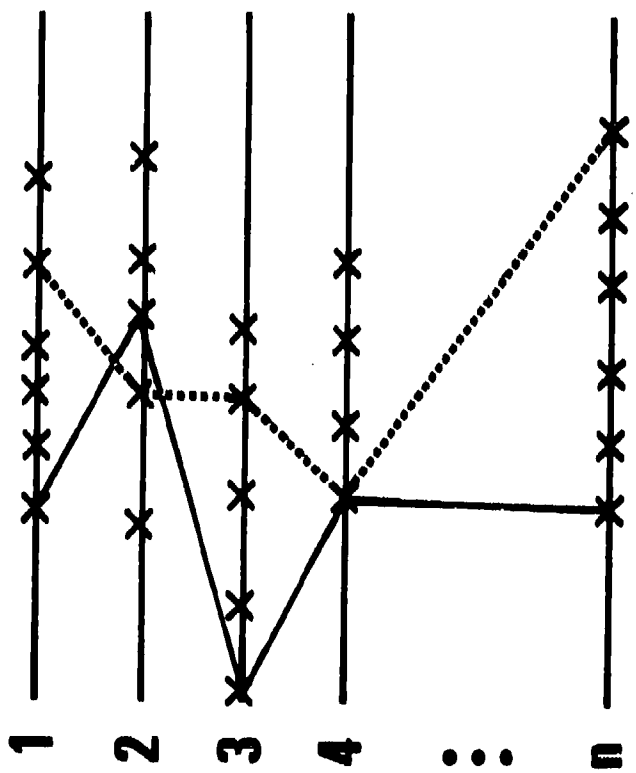


Figure 2.1 Parallel coordinate representation of two n-dimensional points.

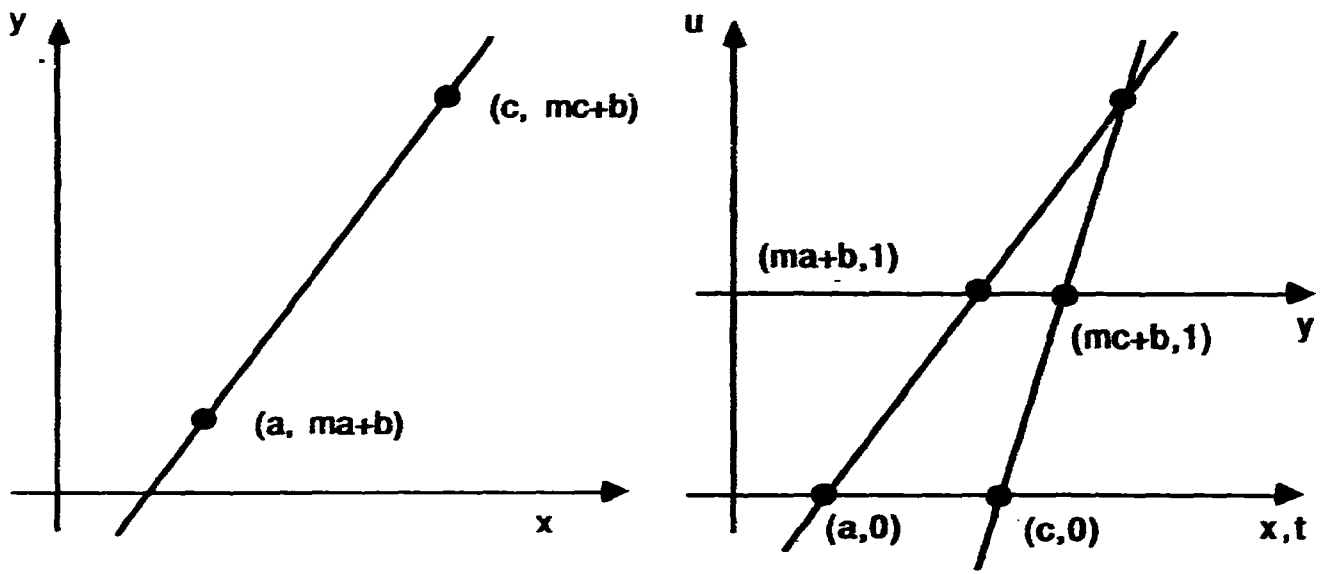


Figure 3.1 Cartesian and parallel coordinate plots of two points. The tu Cartesian coordinate system is superimposed on the xy parallel coordinate system.

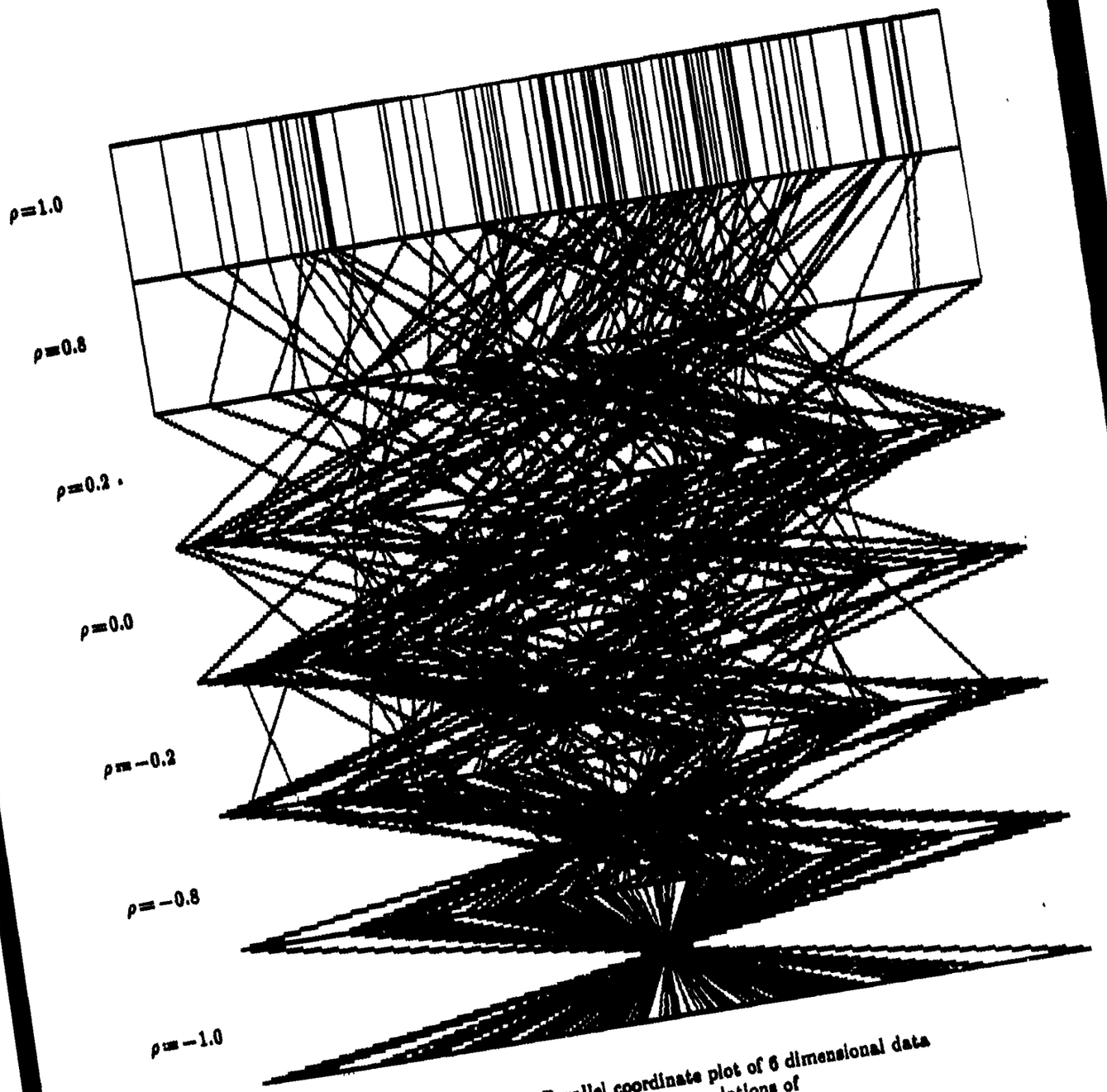


Figure 3.2 Parallel coordinate plot of 6 dimensional data illustrating correlations of $\rho = 1, .8, .2, 0, -.2, -.8$ and -1.

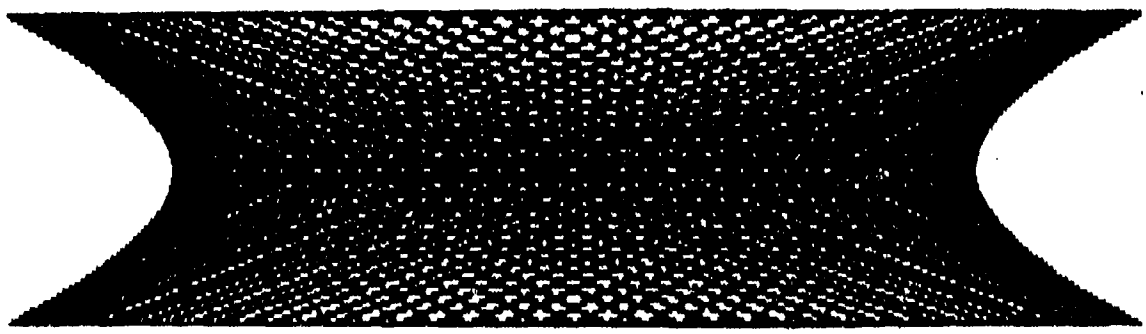


Figure 4.1a Parallel coordinate plot of a circle.

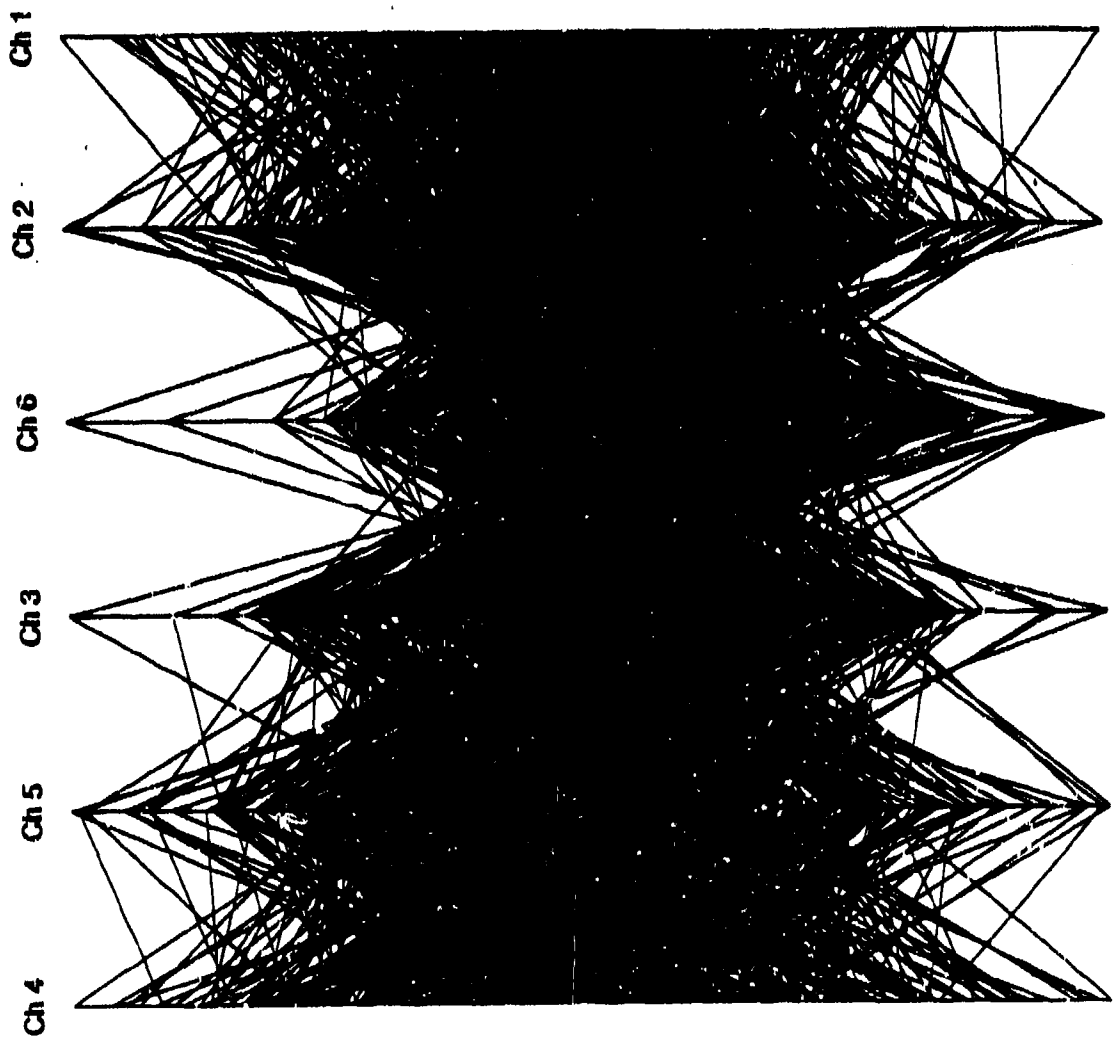


Figure 4.1b Parallel coordinate plot of 6 channel sonar data. The data is uncorrelated Gaussian noise. The second coordinate represents a relatively remote hydrophone and has a somewhat different mean. Notice the approximate hyperbolic shape.

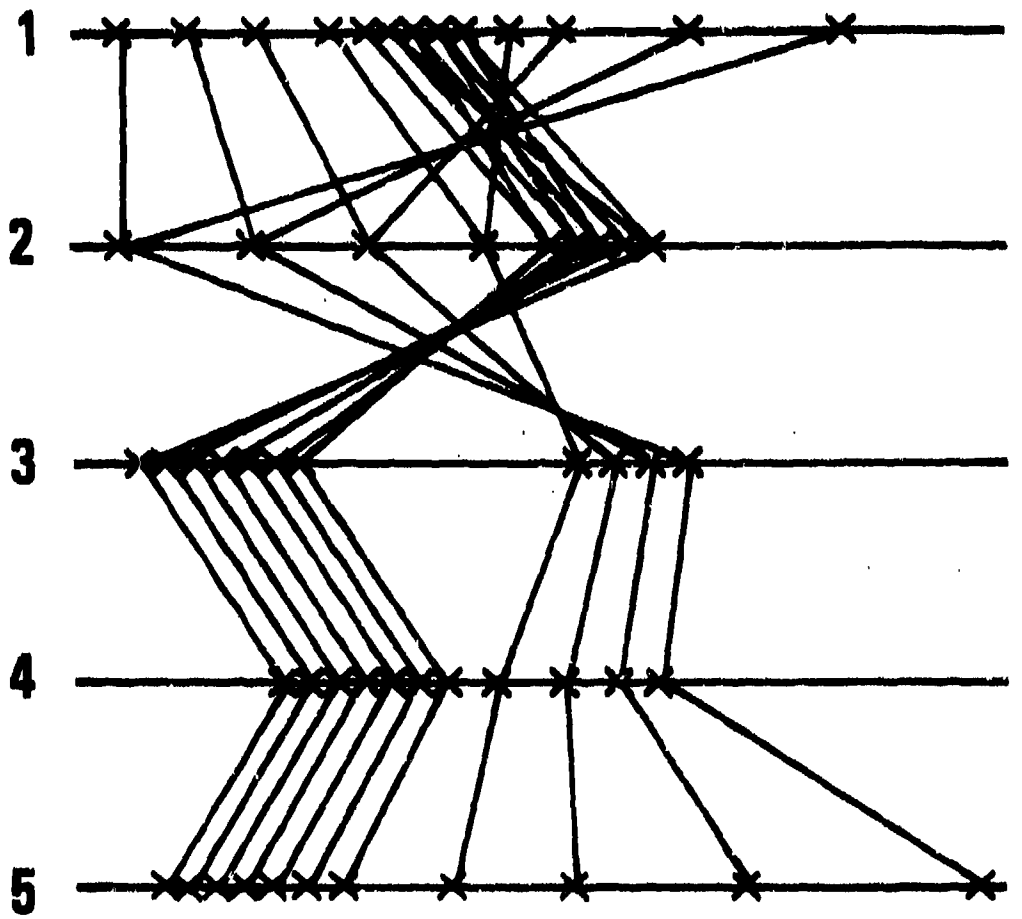


Figure 4.2 A five dimensional scatter diagram in parallel coordinates illustrating marginal densities, correlations, three dimensional clustering and a five dimensional mode.

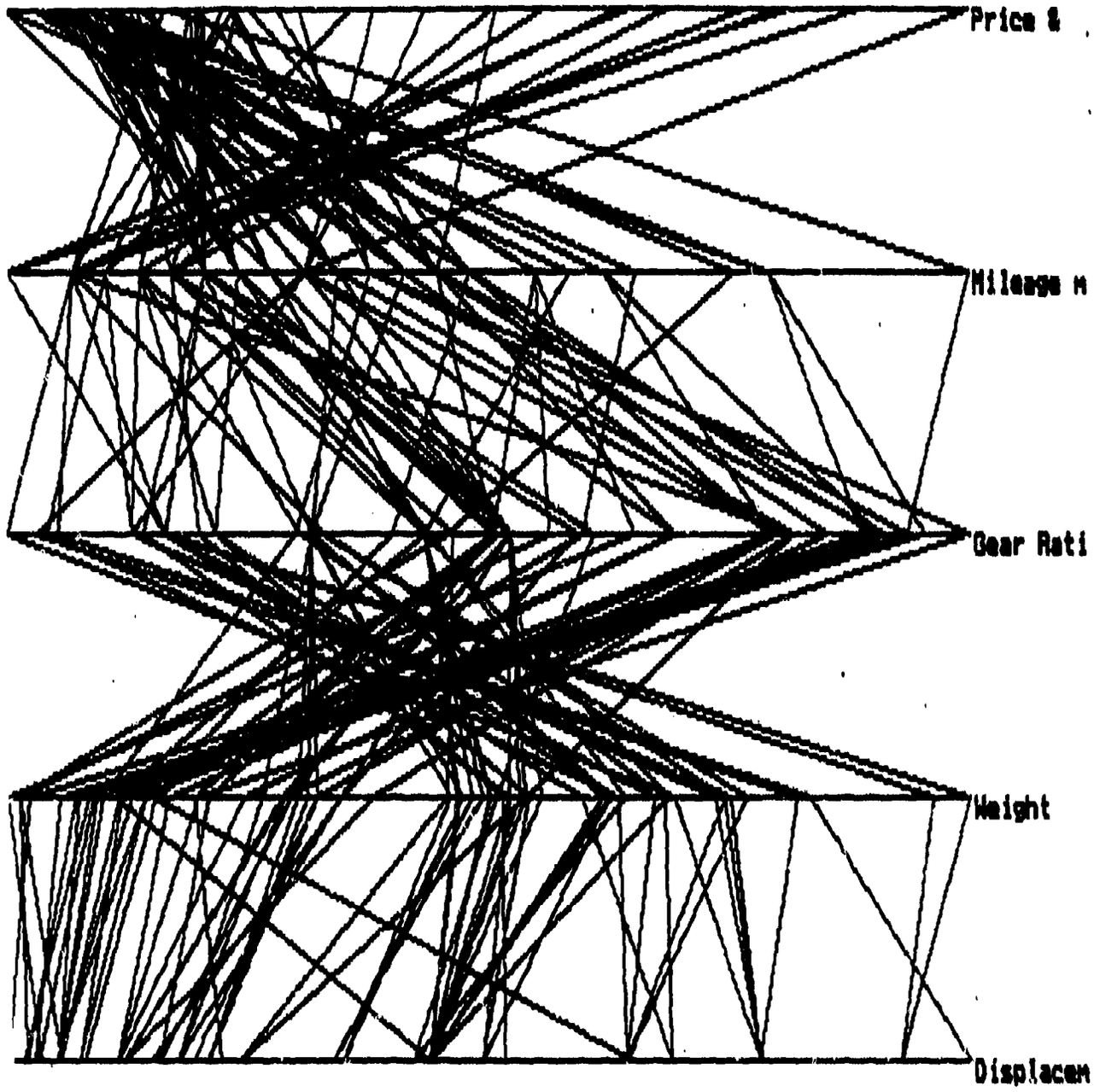


Figure 5.1 A parallel coordinate plot in five dimensions of automobile data. Note the negative correlation between gear ratios and weight.

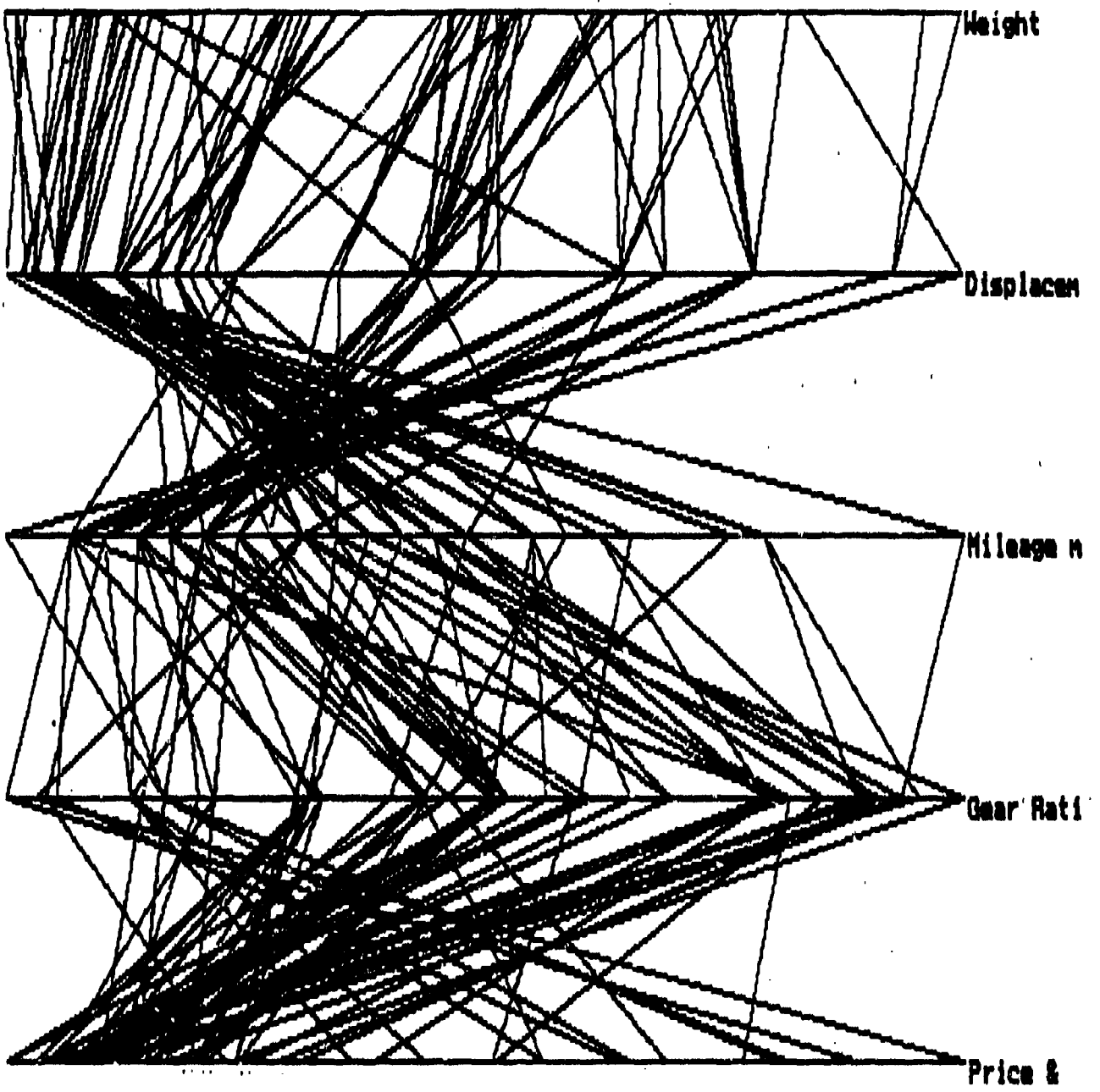


Figure 5.2 The second permutation of the five dimensional presentation of the automobile data. Notice the two classes of linear relations gear ratio and miles per gallon.

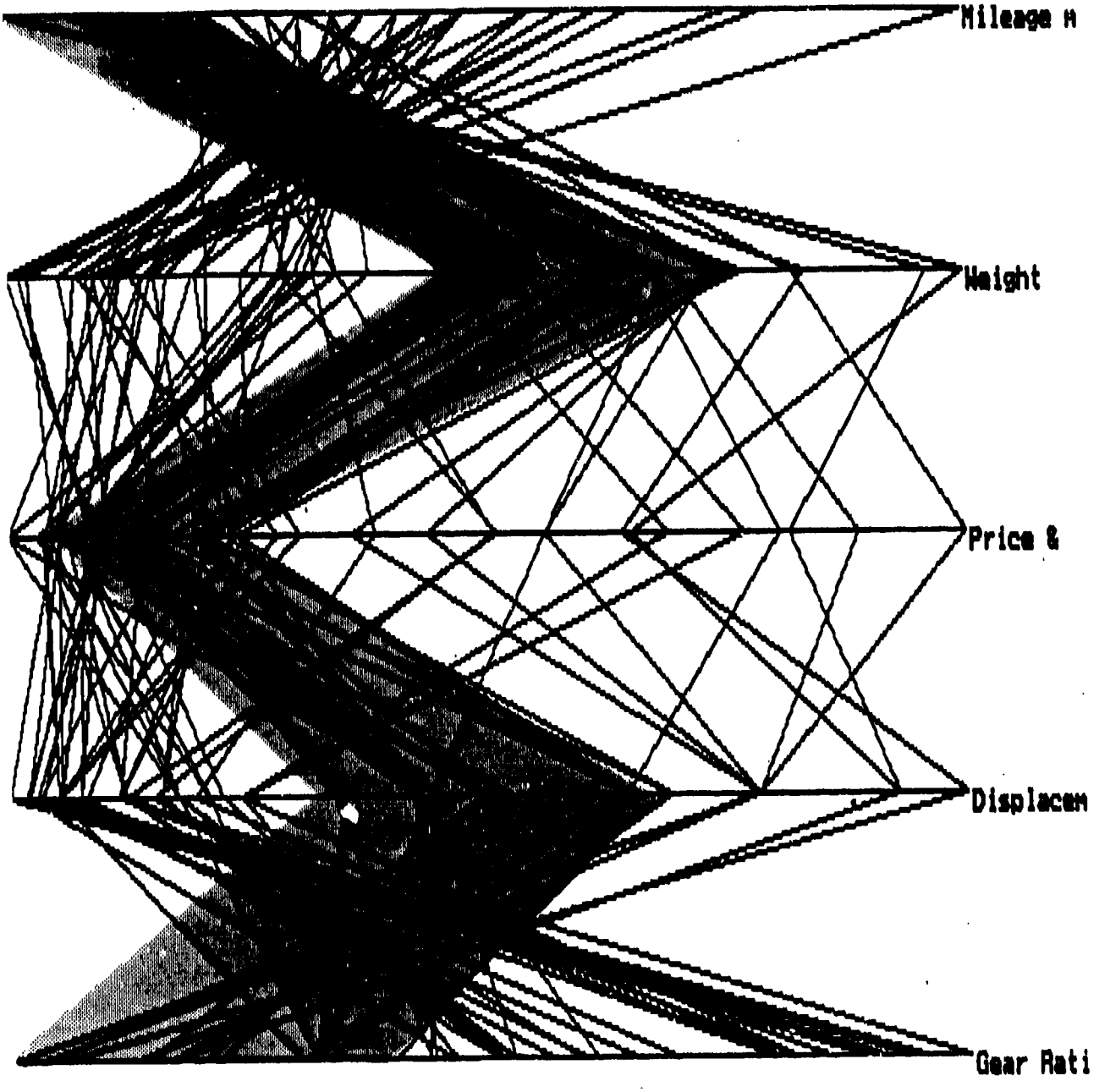


Figure 5.3 **The third permutation of the five dimensional automobile data. Note the highlighting of the domestic automobile group.**

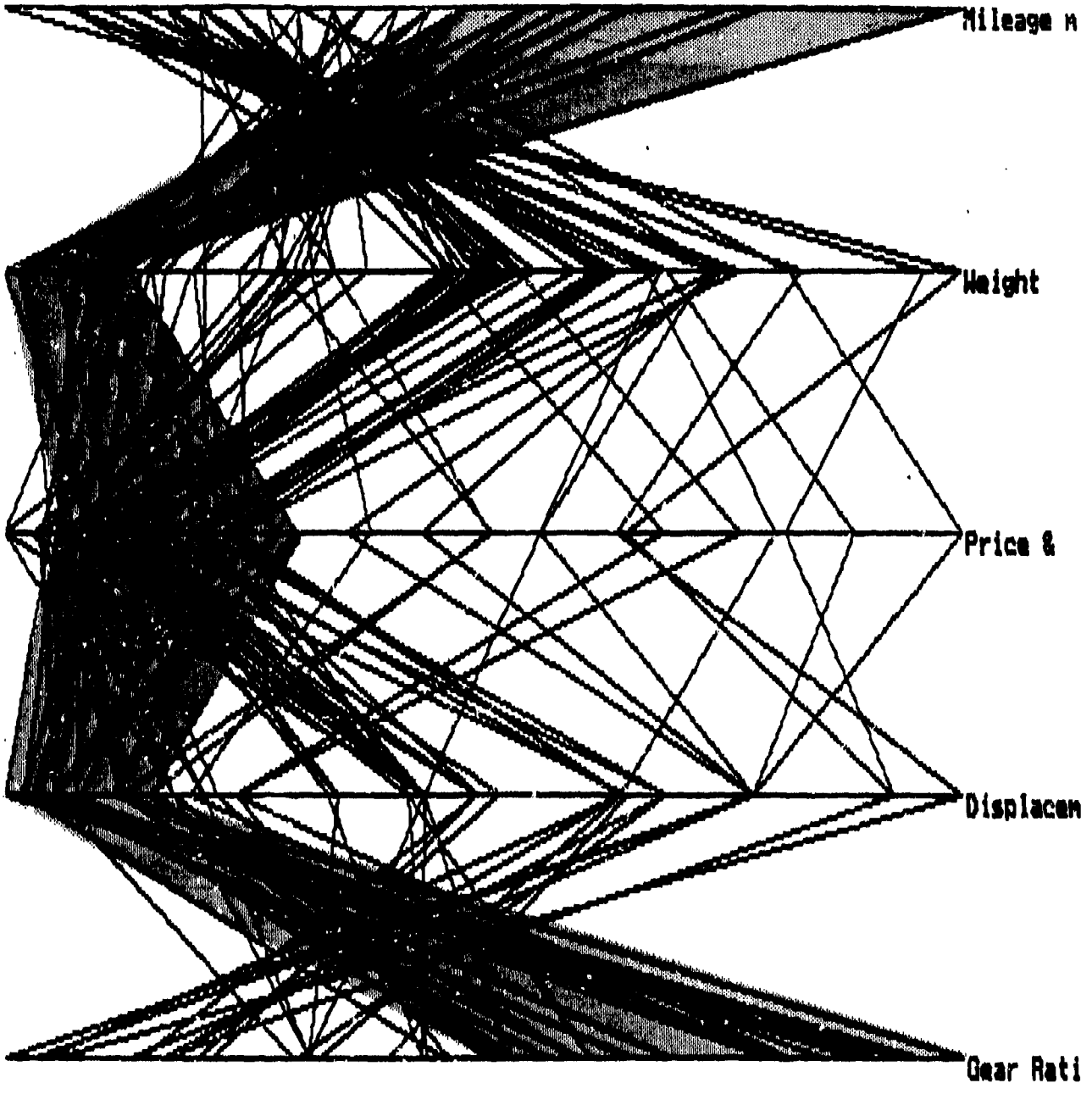


Figure 5.4 **The third permutation showing highlighting of the imported automobile group.**

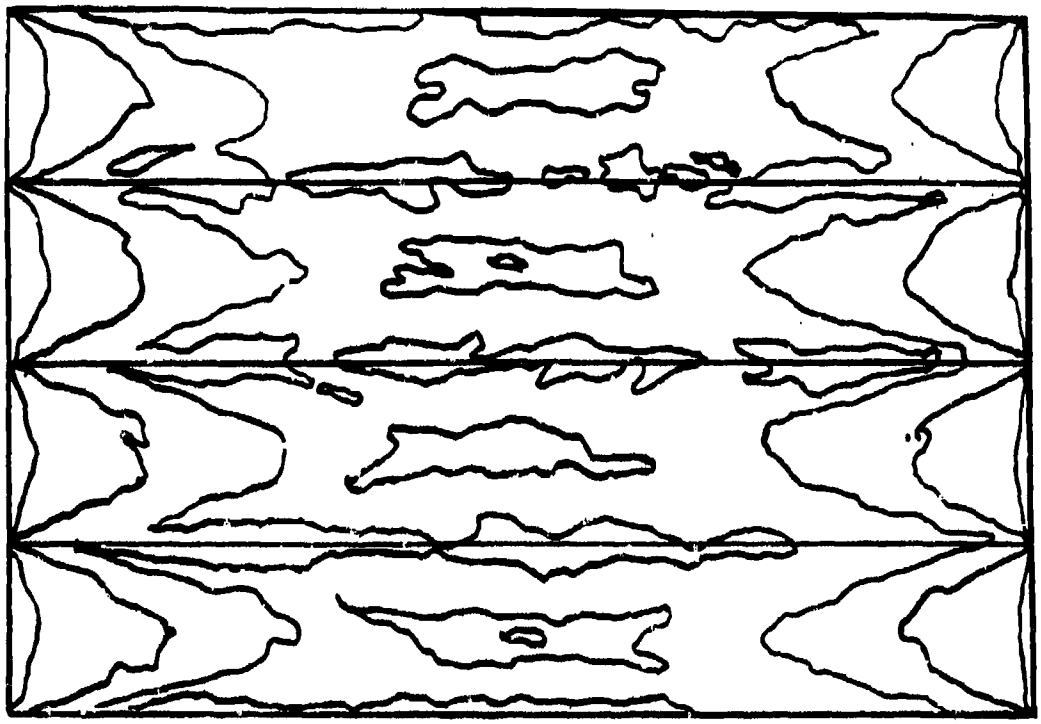


Figure 6.1 **Parallel coordinate density plot of 5000 uniform random variables. This plot has five contour levels 5%, 25%, 50 %, 75% and 95%.**

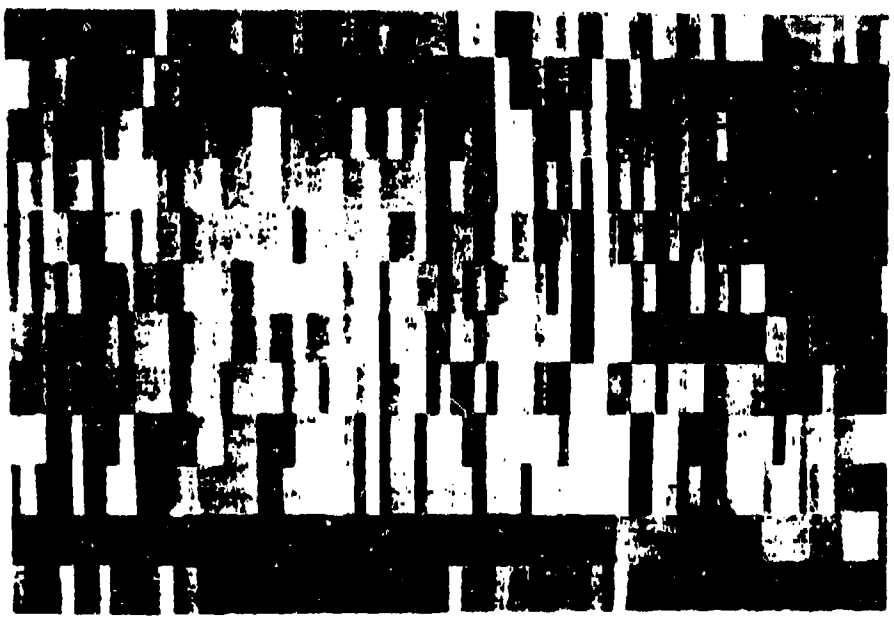


Figure 6.3 **Color histogram of 13 dimensional automobile data. This plot is show in grey scale for purposes of reproduction.**

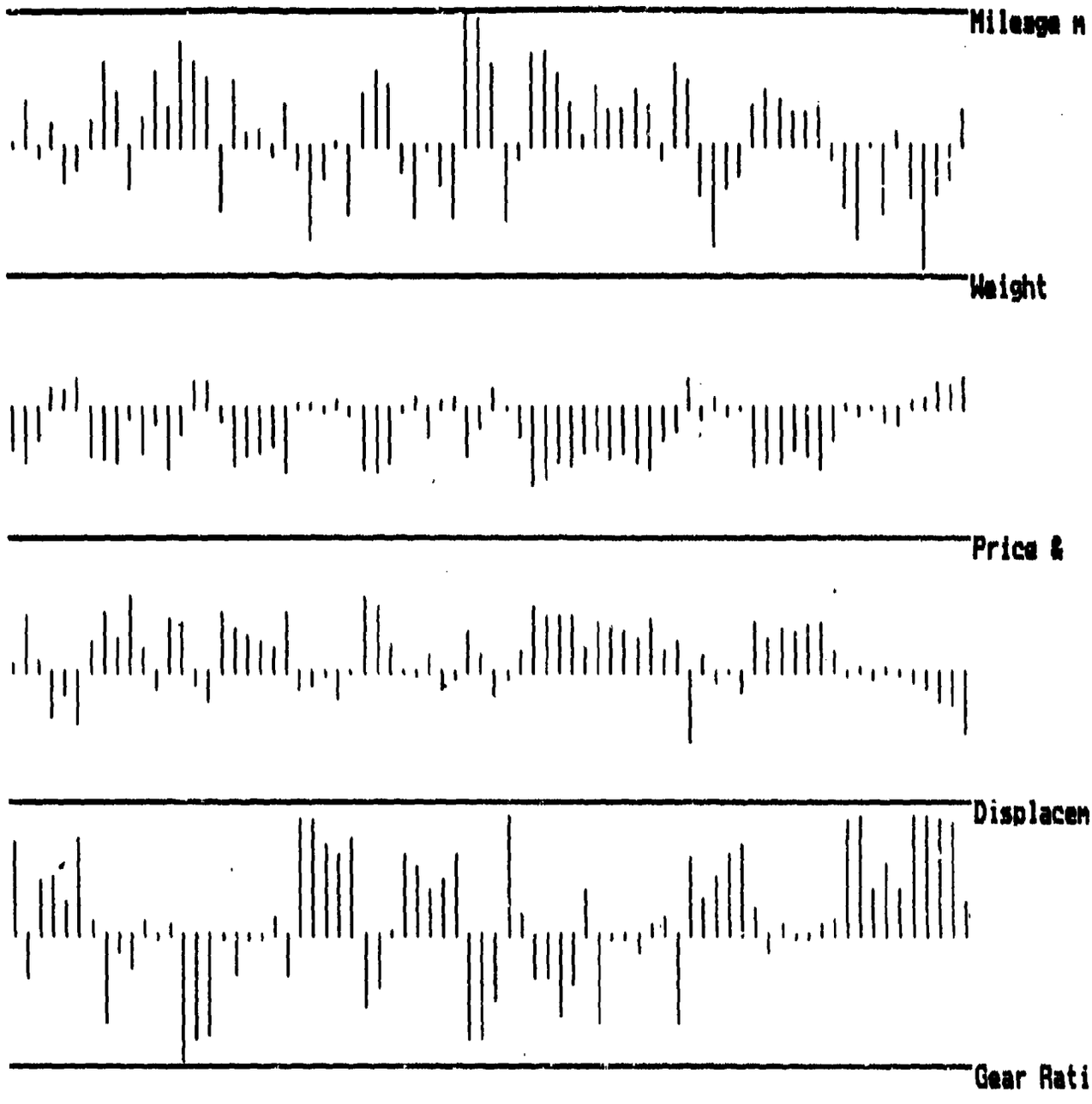


Figure 6.2 **Relative slope plot of five dimensional automobile data. Data presented in the same order as in Figure 7.4**

COMPUTATIONAL AND STATISTICAL ISSUES IN DISCRETE-EVENT SIMULATION

Peter W. Glynn

and

Donald L. Iglehart

Department of Operations Research
Stanford University
Stanford, CA 94305

Abstract

Discrete-event simulation is one of the most important techniques available for studying complex stochastic systems. In this paper we review the principal methods available for analyzing both the transient and steady-state simulation problems in sequential and parallel computing environments. Next we discuss several of the variance reduction methods designed to make simulations run more efficiently. Finally, a short discussion is given of the methods available to study system optimization using simulation.

Keywords: stochastic simulation, output analysis, variance reduction, parallel computation, and system optimization.

1. Introduction.

Computer simulation of complex stochastic systems is an important technique for evaluating system performance. The starting point for this method is to formulate the time varying behavior of the system as a basic stochastic process $Y \equiv \{Y(t) : t \geq 0\}$, where $Y(\cdot)$ may be vector-valued. [Discrete time processes can also be handled.] Next a computer program is written to generate sample realizations of Y. Simulation output is then obtained by running this program. Our discussion in this paper is centered on the analysis of this simulation output. The goal being to develop sound probabilistic and statistical methods for estimating system performance.

Two principal problems arise: the transient simulation problem and the steady-state simulation problem. Let T denote a stopping time and $X \equiv h\{Y(t) : 0 \leq t \leq T\}$, where h is a given real-valued function. The transient problem is to estimate $\alpha \equiv E\{X\}$. Examples of α include the following:

$$\begin{aligned} \alpha &= E\{f(Y(t_0))\}, \\ &= E\left\{\frac{1}{t_0}\int_0^{t_0} f(Y(s))ds\right\}, \end{aligned}$$

and

$$\alpha = P\{Y \text{ does not enter } A \text{ before } t_0\}.$$

Here t_0 is a fixed time (> 0), f is a given real-valued function, and A is a given subset of the state-space of Y. The transient problem is relevant for systems running for a limited (but possibly random) length of time that cannot be expected to reach a steady-state. Our goal here is to provide both point and interval estimates for α.

For the steady-state problem we assume the Y process is asymptotically stationary in the sense that

$$\frac{1}{t}\int_0^t f(Y(s))ds \Rightarrow \alpha$$

as $t \to \infty$. Here $\Rightarrow$ denotes weak convergence and f is a given real-valued function defined on the state-space of Y. The easiest example to think about here is an irreducible, positive recurrent, continuous time Markov chain. In this case $Y(t) \Rightarrow Y$ as $t \to \infty$ and $\alpha \equiv E\{f(Y)\}$. Examples of α in this case include the following:

$$\alpha = E\{Y^k\} \quad \text{(when } Y \text{ is real-valued)},$$

$$\alpha = P\{Y \in A\},$$

and

$$\alpha = E\{c(Y)\},$$

where c is a given cost function. Again as in the transient case, we wish to construct both point and interval estimates for α.

2. Transient Problem.

Assume we have a computational budget of t time units with which to simulate the process Y and estimate $\alpha \equiv E\{X\}$, as defined in Section 1. In a sequential computing environment we would generate independent, identically distributed (iid) copies

$$(X_1, \tau_1), (X_2, \tau_2), \cdots,$$

where the X_i's are copies of X and τ_i is the computer time required to generate X_i. Let $N(t)$ denote the number of copies of X generated in time t; this is just the renewal process associated with the iid τ_i's. A natural point estimator for α is

$$\hat{X}_{N(t)} \equiv \begin{cases} \frac{1}{N(t)} \sum_{i=1}^{N(t)} X_i & , \ N(t) > 0 \\ 0 & , \ N(t) = 0. \end{cases}$$

The standard asymptotic results for $\hat{X}_{N(t)}$ are the strong law of large numbers (SLLN) and the central limit theorem (CLT).

STRONG LAW OF LARGE NUMBERS. If $E\{\tau_1\} < \infty$ and $E\{|\ X_1\ |\} < \infty$, then $\hat{X}_{N(t)} \to \alpha$ *a.s.* as $t \to \infty$.

CENTRAL LIMIT THEOREM. If $E\{\tau_1\} < \infty$ and $\mathrm{var}\{X_1\} < \infty$, then

$$t^{1/2}[\hat{X}_{N(t)} - \alpha] \Rightarrow (E\{\tau_1\} \cdot \mathrm{var}\{X_1\})^{1/2} \cdot N(0,1),$$

where $N(0,1)$ is a mean zero, variance one normal random variable. The SLLN follows from the SLLN for iid summands and the SLLN for renewal processes. The CLT result can be found in BILLINGSLEY (1968), Section 17.

From the SLLN we see that $\bar{X}_{N(t)}$ is a strongly consistent point estimator for α. Thus for large t we would use $\bar{X}_{N(t)}$ as our point estimate. On the other hand, the CLT can be used in the standard manner to construct a confidence interval for α. Here the constant $E\{\tau_1\} \cdot var\{X_i\}$ appearing in the CLT would have to be estimated.

Suppose now that we are in a parallel computing environment with p independent processors. Now we wish to estimate α for a fixed t as $p \to \infty$. On the p processors we generate iid copies of (X, τ):

$$\begin{array}{llll} (X_{11}, \tau_{11}), (X_{12}, \tau_{12}) & , \cdots , & (X_{1N_1(t)}, \tau_{1N_1(t)}) \\ (X_{21}, \tau_{21}), (X_{22}, \tau_{22}) & , \cdots , & (X_{2N_2(t)}, \tau_{2N_2(t)}) \\ & \vdots & \\ (X_{p1}, \tau_{p1}), (X_{p2}, \tau_{p2}) & , \cdots , & (X_{pN_p(t)}, \tau_{p,N_p(t)}). \end{array}$$

A number of estimators have been proposed for estimating $\alpha = E\{X\}$. The most natural estimator to consider first is that obtained by averaging the realizations of X across each processor and then averaging these sample means. This leads to

$$\alpha_1(p,t) \equiv \frac{1}{p} \sum_{i=1}^{p} \bar{X}^{(i)}_{N_i(t)},$$

where

$$\bar{X}^{(i)}_{N_i(t)} \equiv \begin{cases} \frac{1}{N_i(t)} \sum_{j=1}^{N_i(t)} X_{ij} & , \ N_i(t) > 0 \\ 0 & , \ N_i(t) = 0. \end{cases}$$

Here the processing ends on all processors at time $T_p = t$. If $E\{\tau_1\} < \infty$ and $E\{| X |\} < \infty$, then for all $t > 0$

$$\alpha_1(p,t) \to E\{\bar{X}_{N(t)}\} = E\{X \cdot 1_{\{\tau \le t\}}\} \quad a.s.$$

as $p \to \infty$. Here 1_A is the indicator function of the set A. Unfortunately, $E\{X\} \neq E\{X \cdot 1_{\{\tau \le t\}}\}$ and so $\alpha_1(p,t)$ is not strongly consistent for α as $p \to \infty$.

The next estimator for α was proposed by HEIDELBERGER (1987). For this estimator we let all processors complete the replication in process at time t. The estimator is

$$\alpha_2(p,t) \equiv \frac{\sum_{i=1}^{p} \sum_{j=1}^{N_i(t)+1} X_{ij}}{\sum_{i=1}^{p} [N_i(t) + 1]}.$$

Here all processors complete by time

$$T_p = \max_{1 \le i \le p} [\tau_{i1} + \tau_{i2} + \cdots + \tau_{iN_i(t)+1}].$$

Unfortunately, $T_p \to +\infty$ *a.s.* as $p \to \infty$. However, $\alpha_2(p,t)$ is strongly consistent for α. To see this, note that if $E\{|X|\} < \infty$ and $P\{\tau > 0\} > 0$, then as $p \to \infty$

$$\alpha_2(p,t) \to \frac{E\left\{\sum_{j=1}^{N_i(t)+1} X_{ij}\right\}}{E\{N_i(t)+1\}} = E\{X\} \qquad a.s.$$

The equality about is simply Wald's equation. Finally, since $\alpha_2(p,t)$ is a ratio estimator, a CLT is also available from which a confidence interval can be constructed.

The last estimator we consider was proposed by HEIDELBERGER and GLYNN (1987). Here we set

$$\alpha_3(p,t) = \frac{1}{p} \sum_{i=1}^{p} \tilde{X}^{(i)}_{N_i(t)},$$

where

$$\tilde{X}^{(i)}_{N_i(t)} = X^{(i)}_{N_i(t)} + X_{i1} 1_{\{\tau_{i1} > t\}}.$$

Given $N(t) \ge 1$, Heidelberger and Glynn show that the pairs of random variables (X_1, τ_1), $\ldots, (X_{N(t)}, \tau_{N(t)})$ are exchangeable. Using this fact, they prove that $E\{\tilde{X}^{(i)}_{N_i(t)}\} = E\{X_1\}$. Since the $\tilde{X}^{(i)}_{N_i(t)}$'s are iid, we see that $\alpha_3(t)$ is strongly consistent for $\alpha = E\{X_1\}$. Since the summands in $\alpha_3(p,t)$ are iid, the standard CLT holds (under appropriate variance assumptions) and can be used to develop a confidence interval for α. Note that the definition of $\tilde{X}^{(i)}_{N_i(t)}$ requires the ith processor to complete the replication in process at time t, if no observations have been completed by time t; i.e., $\tau_{i1} > t$. Thus the completion time for all p processors is given by

$$T_p = \max_{1 \le i \le p} \{\max(t, \tau_{i1})\}.$$

While $T_p \to \infty$ *a.s.* as $p \to \infty$ (if $P\{\tau_{i1} > t\} > 0$), T_p goes to infinity at a much slower rate than is the case for $\alpha_2(p,t)$. They also show that the following CLT holds:

$$t^{1/2}[\tilde{X}_{N(t)} - \alpha] \Rightarrow \sigma E^{1/2}\{\tau_1\} \cdot N(0,1)$$

as $t \to \infty$, where we assume $0 < \sigma^2 = var\{X_1\} < \infty$ and $0 < E\{\tau_1\} < \infty$. Thus $\bar{X}_{N(t)}$ can also be used in a sequential environment to estimate α.

3. Steady-State Problem.

The steady-state estimation problem is considerably more difficult than the transient estimation problem. This difficulty stems from the following considerations: (i) need to estimate long-run system behavior from a finite length simulation run; (ii) an initial bias (or transient) usually is present since the process being simulated is non-stationary; and (iii) strong autocorrelations are usually present in the process being simulated. While classical statistical methods can often be used for the transient estimation problem, these methods generally fail for the steady-state estimation problem for the reasons mentioned above.

Assume our simulation output process is $Y \equiv \{Y(t) : t \geq 0\}$ and for a given real-valued function f

$$\alpha(t) \equiv \frac{1}{t}\int_0^t f[Y(s)]ds \Rightarrow \alpha. \tag{1}$$

As stated above, we wish to construct point and interval estimators for α. In addition to (1), many methods also assume that a positive constant σ exists such that the following CLT holds:

$$\sqrt{t}[\alpha(t) - \alpha] \Rightarrow \sigma \cdot N(0,1) \tag{2}$$

as $t \to \infty$. From (1) and (2) we can construct a point estimate and confidence interval for α provided we can estimate σ. Estimating σ is generally the hardest problem.

A variety of methods have been developed to address the steady-state estimation problem. In Figure 1 we have given a break-down of these methods. Most of the methods are single replicate methods, since multiple replicate methods tend to be inefficient because of the initial bias problem.

Here we only consider single replicate methods. These methods are of two types: those that consistently estimate σ and those in which σ is cancelled out.

For consistent estimation of σ, we need a process $\{s(t) : t \geq 0\}$ such that $s(t) \Rightarrow \sigma$.

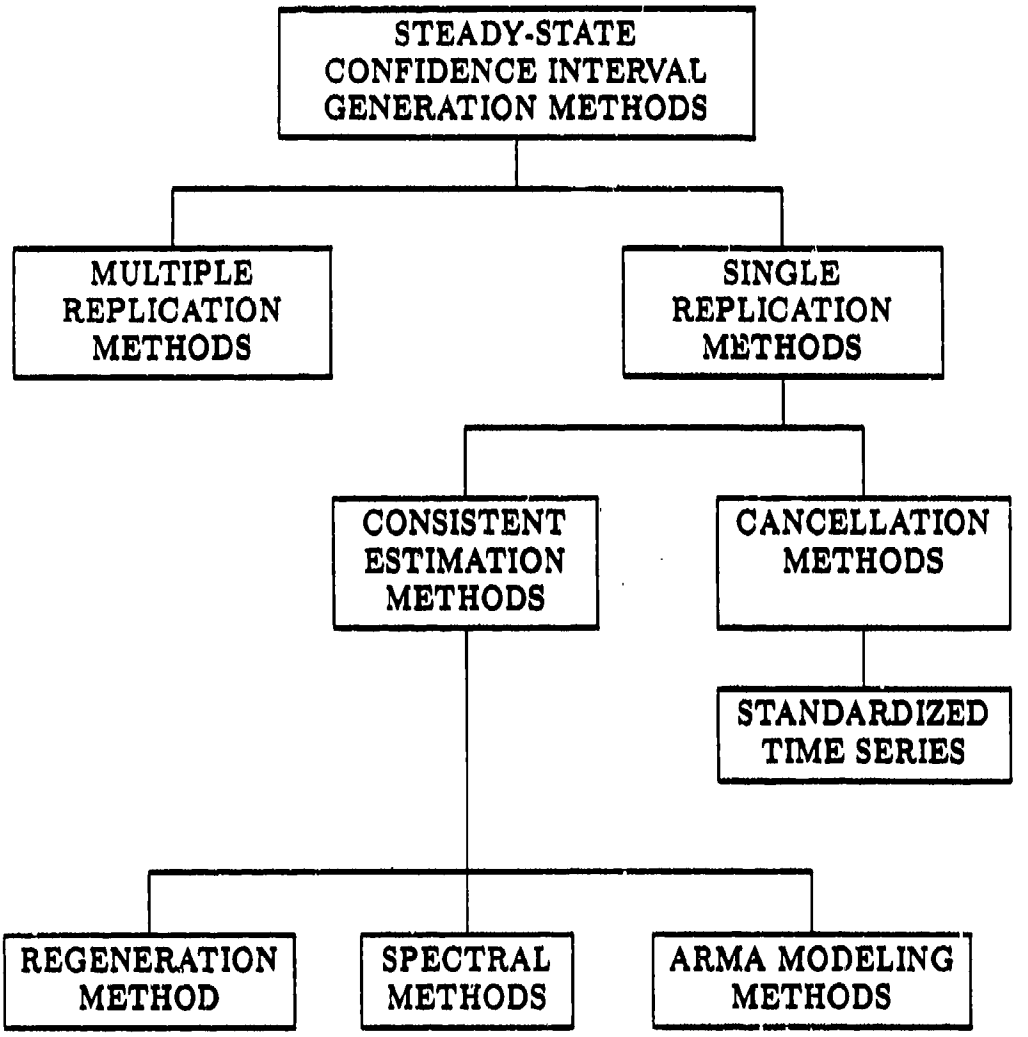

STEADY-STATE CONFIDENCE INTERVAL GENERATION METHODS
MULTIPLE REPLICATION METHODS
SINGLE REPLICATION METHODS
CONSISTENT ESTIMATION METHODS
CANCELLATION METHODS
STANDARDIZED TIME SERIES
REGENERATION METHOD
SPECTRAL METHODS
ARMA MODELING METHODS

Figure 1

In which case (2) leads to a $100(1-\delta)$ % confidence interval for α given by

$$[\alpha(t) - z(1-\delta/2)s(t)/t^{1/2}, \alpha(t) + z(1-\delta/2)s(t)/t^{1/2}],$$

where $\Phi(z(1-\delta/2)) = 1-\delta/2$ and Φ is the standard normal distribution function.

On the other hand, the canceling out methods require a non-vanishing process $\{Z(t): t \geq 0\}$ such that

$$[t^{1/2}(\alpha(t) - \alpha), Z(t)] \Rightarrow [\sigma N(0,1), \sigma Z]$$

as $t \to \infty$. Then using the continuous mapping theorem (cf., BILLINGSLEY (1968), p. 30) we have

$$t^{1/2}(\alpha(t) - \alpha)/Z(t) \Rightarrow N(0,1)/Z \tag{3}$$

as $t \to \infty$. Note from (3) that σ has been cancelled out in a manner reminiscent of the t-statistic.

First we discuss one of the methods in which σ is consistently estimated, namely, the regenerative method; see IGLEHART (1978) for a discussion of this method plus other background material. Here we assume that the simulation output process Y is a regenerative process. We are given a real-valued function f and wish to estimate $\alpha(f) \equiv E\{f(Y)\}$, where $Y(t) \Rightarrow Y$ as $t \to \infty$. Again it is convenient to think of Y as an irreducible, positive recurrent, continuous time Markov chain. Let $T(0) = 0, T_1, T_2, \ldots$ be the regeneration times for Y and set $\tau_i = T_i - T_{i-1}, i \geq 1$. The τ_i's are the lengths of the regenerative cycles. Next define the areas under the Y process in the kth regenerative cycle by

$$Y_k(f) = \int_{\tau_{k-1}}^{\tau_k} f[Y(s)]ds.$$

The following basic facts provide the foundation for the regenerative method:

(i) the pairs $\{(Y_k(f), \tau_k): k \geq 1\}$ are iid;

(ii) if $E\{|f(Y)|\} < \infty$, then $\alpha(f) = E\{Y_1(f)\}/E\{\tau_1\}$.

The regenerative method can be developed on either the intrinsic time scale (t) or on the random time scale (n) corresponding to the number of regenerative cycles simulated. On the intrinsic time scale our point estimate for α is given by

$$\alpha(t, f) \equiv \frac{1}{t}\int_0^t f(Y(s))ds,$$

where t is the length of time the simulation is run. On the random time scale our point is given by

$$\alpha_n(f) \equiv \bar{Y}_n(f)/\bar{\tau}_n,$$

where $\bar{Y}_n(f)$ (respectively, $\bar{\tau}_n$) is the sample mean of $Y_1(f), ..., Y_n(f)$ ($\tau_1, ..., \tau_n$). Here the Y process is simulated to the completion of n regenerative cycles. Using the basic facts (i) and (ii) above, it can be shown that both $\alpha(t, f)$ and $\alpha_n(f)$ are strongly consistent for $\alpha(f)$ as t and n respectively tend to infinity. Next we define $Z_k \equiv Y_k(f) - \alpha(f)\tau_k$ and assume that $var\{Z_k\} \equiv \sigma^2 < \infty$. Then it can be shown that the following two CLT's hold as $t \to \infty$ and $n \to \infty$:

$$t^{1/2}[\alpha(t, f) - \alpha(f)] \Rightarrow (\sigma/E^{1/2}\{\tau_1\})N(0, 1),$$

and

$$n^{1/2}[\alpha_n(f) - \alpha(f)] \Rightarrow (\sigma/E\{\tau_1\})N(0, 1).$$

These two CLT's can then be used to construct confidence intervals for $\alpha(f)$ provided both σ^2 and $E\{\tau_1\}$ can be estimated. The mean $E\{\tau_1\}$ is easily estimated by $\bar{\tau}_n$ and σ^2 can be estimated from its definition in terms of $Y_1(f)$ and τ_1.

Next we turn to a discussion of the principal method available for canceling out σ. This is the method of standardized time series developed by SCHRUBEN (1983). Our discussion is based on the paper GLYNN and IGLEHART (1989) and uses some results from weak convergence theory; see BILLINGSLEY (1968) for background on this theory. From our output process Y we form the random elements of $C[0, 1]$, the space of real-valued continuous functions on the interval $[0, 1]$, given by

$$\bar{Y}_n(t) \equiv \frac{1}{n}\int_0^{nt} Y(s)ds$$

and

$$X_n(t) \equiv n^{1/2}[\bar{Y}_n(t) - \alpha t],$$

where $0 \leq t \leq 1$ and $n \geq 1$. Now we make the basic assumption that a finite, positive constant σ exists such that

$$X_n \Rightarrow \sigma B \text{ as } n \to \infty, \tag{4}$$

where B is standard Brownian motion. This assumption holds for a wide class of output processes. To find the scaling process $\{Z(t) : t \geq 0\}$ consider the class $\mathcal{M}$ of functions $g : C[0,1] \to \mathbb{R}$ such that

(i) $g(\alpha x) = \alpha g(x)$ for all $\alpha > 0$ and $x \in C[0,1]$;

(ii) $g(B) > 0$ with probability one;

(iii) $g(x + \beta k) = g(x)$ for all real β and $x \in C[0,1]$, where $k(t) = t$;

(iv) $P\{B \in D(g)\} = 0$, where $D(g)$ is the set of discontinuities of g.

The process

$$S_n(t) \equiv \frac{\bar{Y}_n(t) - \alpha t}{g(\bar{Y}_n)}, \quad 0 \leq t \leq 1,$$

is called a standardized time series. Using weak convergence arguments it is easy to show from (4) that

$$S_n(1) \Rightarrow B(1)/g(B) \tag{5}$$

as $n \to \infty$. Unfolding this CLT we have the following $100(1-\delta)\%$ confidence interval for α:

$$[\bar{Y}_n(1) - z(1-\delta/2)g(\bar{Y}_n), \bar{Y}_n(1) + z(\delta/2)g(\bar{Y}_n)],$$

where $P\{B(1)/g(B) \leq z(\alpha)\} = \alpha$ for $0 \leq \alpha \leq 1$. Thus each $g \in \mathcal{M}$ gives rise to a confidence interval for α provided we can find the distribution of $B(1)/g(B)$. Fortunately, this can be done for a number of interesting g functions.

One of the g functions leads to the batch means method, perhaps the most popular method for steady-state simulation. We conclude our discussion of the method of standardized time series by displaying this special g function. To this end we first define the Brownian bridge mapping $\Gamma : C[0,1] \to C[0,1]$ as

$$(\Gamma x)(t) = x(t) - tx(1), \quad x \in C[0,1], \quad 0 \leq t \leq 1$$

Now think of partitioning our original output process Y into $m \geq 2$ intervals of equal length and define the mapping $b_m : C[0,1] \to \mathbb{R}$ by

$$b_m(x) = \left[\left(\frac{m}{m-1}\right)\sum_{i=1}^{m}(x(i/m) - x((i-1)/m)^2\right]^{1/2},$$

for $x \in C[0,1]$. Finally, the g function of interest is $g_m = b_m \circ \Gamma$. To see that g_m corresponds to the batch means method we observe that

$$g_m(\bar{Y}_n) = m^{-1/2}\left[\frac{1}{m-1}\sum_{i=1}^{m}\left(Z_i(n) - \frac{1}{m}\sum_{j=1}^{m} Z_j(n)\right)^2\right]^{1/2},$$

where

$$Z_i(n) = \int_{(i-1)n/m}^{in/m} Y(x)dx/(n/m)$$

is the ith batch mean of the process $\{Y(t) : 0 \le t \le n\}$. Specializing (5) to the function g_m we see that

$$[\frac{1}{m}\sum_{i=1}^{m} Z_i(n) - \alpha]/g_m(\bar{Y}_n) \Rightarrow t_{m-1}$$

as $n \to \infty$, where t_{m-1} is a Student's-t random variable with $m-1$ degrees of freedom. This follows from the fact that $B(1)/g_m(B)$ is distributed as t_{m-1} since B has independent normal increments. For other examples of functions $g \in \mathcal{M}$ for which the distribution of $B(1)/g(B)$ is known see GLYNN and IGLEHART (1989).

4. Variance Reduction Techniques.

Once a basic method is developed to produce point estimates and confidence intervals for a parameter of interest, we turn our attention to making these methods more efficient. Over the years a dozen or more techniques have been proposed to improve simulation efficiency. Good references for many of these techniques are BRATLEY, FOX, and SCHRAGE (1987), WILSON (1984). Here we have elected to outline three of these techniques.

As we have seen in Sections 2 and 3, confidence intervals for parameters being estimated are generally constructed from an associated CLT. Each CLT has an intrinsic variance constant, say, σ_1^2. The idea for many variance reduction techniques (VRT's) is to modify the original estimate in such a way as to yield a new CLT with a variance constant $\sigma_2^2 < \sigma_1^2$. This will, of course, lead to confidence intervals of shorter length, or alternatively, confidence intervals of the same length from a shorter simulation run. Frequently

VRT's are based on some analytic knowledge or structural properties of the process being simulated.

The first VRT we discuss is known as importance sampling. This idea was first developed in conjunction with the estimation of $E\{h(X)\} \equiv \alpha$, where h is a known real-valued function and X a random variable with density, say, f. Instead of sampling X from f, we sample X from a density g which has been selected to be large in the regions that are "most important", namely, where $|f|$ is largest. Then we estimate α by the sample mean of $h(X)f(X)/g(X)$; see HAMMERSLEY and HANDSCOMB (1964).

This same basic idea can be carried forward to the estimation of parameters associated with stochastic processes. We generate the process with a new probabilistic structure and estimate a modified parameter to produce an estimate of the original quantity of interest. The example we consider here is the $M/M/1$ queue with arrival rate λ, service rate μ, and traffic intensity $\rho \equiv \lambda/\mu < 1$. Let V denote the stationary virtual waiting time and consider estimating the quantity $\alpha \equiv P\{V > u\}$ for large u. When ρ is less than one, the virtual waiting time process has a negative drift and an impenetrable barrier at zero. Thus the chance of the process getting above a large u is small, and a long simulation would be required to accurately estimate α. The idea used here in importance sampling is to generate a so-called conjugate process obtained by reversing the roles of λ and μ. For the conjugate process the traffic intensity is greater than one, and the estimation problem becomes much easier. ASMUSSEN (1985) reports efficiency increases on the order of a factor of 3 to a factor of 400 over straight regenerative simulation depending on the values of ρ and u. In general, importance sampling can yield very significant variance reductions. Further work along these lines can be found in SIEGMUND (1976), GLYNN and IGLEHART (1989), SHAHABUDDIN et al. (1988), and WALRAND (1987).

The second VRT we discuss is known as indirect estimation. Assume we are interested in estimating $\alpha \equiv E\{X\}$, but happen to know that $E\{Y\} = aE\{X\} + b$ where a and b are known. Sometimes it happens that a CLT associated with the estimation of $E\{Y\}$ will have a smaller variance constant associated with it than does the CLT for estimating $E\{X\}$. In this case we would prefer to estimate $E\{Y\}$ and we use the affine transformation above to yield an estimate for $E\{X\}$. This idea has proved to be useful in queuing simulations where

the affine transformation is a result of Little's Law. In general, variance reductions realized using this method are not dramatic, being usually less than a factor of 2. For further results along these lines, see LAW (1975) and GLYNN and WHITT (1986). While the affine transformation works in queuing theory, it is conceivable that other transformations might arise in different contexts.

The third and final VRT we discuss here is known as discrete time conversion. Suppose that $X = \{X(t) : t \geq 0\}$ is an irreducible, positive recurrent, continuous time Markov chain (CTMC). Then $X(t) \Rightarrow X$ as $t \to \infty$ and we may be interested in estimating $\alpha \equiv E\{f(X)\}$, where f is a given real-valued function. As we have discussed above, the regenerative method can be used to estimate α. A CTMC has two sources of randomness: the embedded discrete time jump chain and the exponential holding times in the successive states visited. The discrete time conversion method eliminates the randomness due to the holding times by replacing them by their expected values. It has been shown that this leads to a variance reduction when estimating α. Also, as an added side benefit computer time is saved since the exponential holding times no longer need to be generated. Gains in efficiency for this method can be substantial. Further discussion of this idea can be found in HORDIJK, IGLEHART, and SCHASSBERGER (1976), and FOX and GLYNN (1986).

5. System Optimization Using Simulation.

Consider a family of stochastic systems indexed by a parameter θ (perhaps vector-valued). Suppose $\alpha(\theta)$ is our performance criterion for system θ. Our concern here is to find that system, say θ_0, which optimizes the value of α. For a complex system it is frequently impossible to evaluate α analytically. Simulation may be the most attractive alternative. We could naively simulate the systems at a sequence of parameter settings $\theta_1, \theta_2, \cdots, \theta_k$ and select setting that optimizes $\alpha(\theta_i)$. In general this would not be very efficient, since k would have to be quite large. A better way would be to estimate the gradient of α and use this estimate to establish a search direction. Then stochastic approximation and ideas from non-linear programming could be used to optimize α.

Two general methods have been proposed to estimate gradients: the likelihood ratio method and the infinitesimal perturbation method. We will discuss both methods briefly.

Suppose $X = \{X_n : n \geq 0\}$ is a discrete time Markov chain (DTMC) and that the cost of running system θ for $n+1$ steps is $g(\theta, X_0, \ldots, X_n)$. The expected cost of running system θ is then given by

$$\alpha(\theta) \equiv E_\theta\{g(\theta, X_0, \ldots, X_n)\}, \tag{6}$$

where E_θ is expectation relative to the probability measure $P(\theta)$ associated with system θ. If $E_\theta\{\cdot\}$ were independent of θ, we would simply simulate iid replicates of $\nabla g(\theta, X_0, \ldots, X_n)$. By introducing the likelihood function $L(\theta, X_0, \ldots, X_n)$ it is possible to write $\alpha(\theta)$ as

$$\alpha(\theta) = E_{\theta_0}\{g(\theta, X_0, \cdots, X_n)L(\theta, X_0, \cdots, X_n)\}$$

for a fixed value of θ_0. Then we can write

$$\nabla\alpha(\theta) = E_{\theta_0}\{\nabla g(\theta, X_0, \cdots, X_n)L(\theta, X_0, \cdots, X_n)\},$$

where the interchange of ∇ and E_{θ_0} must be justified. A similar approach can be developed to estimate the gradient of a performance criterion for a steady-state simulation. For an overview of this approach see GLYNN (1987), and REIMAN and WEISS (1986).

The second method which has been proposed for estimating gradients is called the infinitesimal perturbation analysis (IPA) method. In this method a derivative, with respect to an input parameter, of a simulation sample path is computed. For example, we might be interested in estimating the mean stationary waiting time for a queueing system as well as its derivative with respect to the mean service time. Since we are taking a derivative of the sample path inside an expectation operator, the interchange of expectation and differentiation must be justified in order to produce an estimate for the gradient $\nabla\alpha(\theta)$, say. The IPA method assumes that if the change in the input parameter, θ, is small enough, then the times at which events occur get shifted slightly, but their order does not change. It has been shown that the IPA method yields strongly consistent estimates for the performance gradient in a variety of queueing contexts; see HEIDELBERGER, CAO, ZAZANIS, and SURI (1988) for details on the IPA method and a listing of queueing problems for which the technique works.

REFERENCES

ASMUSSEN, S. (1985). Conjugate processes and the simulation of ruin-problems. **Stoch. Proc. Appl.** 20, 213–229.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. John Wiley and Sons, New York.

BRATLEY, P., FOX, B., and SCHRAGE, L. (1987). *A Guide to Simulation*. 2nd Ed. Springer-Verlag, New York.

FOX, B. and GLYNN, p. (1986). Discrete-time conversion for simulating semi-Markov processes. **Operations Research Letters** 5, 191–196.

GLYNN, P. and WHITT, W. (1989). Indirect estimation via $L = \lambda W$. **Operations Research** 37, 82–103.

GLYNN, p. (1987). Likelihood ratio gradient estimation: an overview. **Proceedings of the 1987 Winter Simulation Conference**, 366–375.

GLYNN, P. and HEIDELBERGER, P. (1987). Bias properties of budget constrained Monte Carlo simulations, I: estimating a mean. Technical Report, Department of Operations Research, Stanford University.

GLYNN, P. and IGLEHART, D. (1989). Simulation output analysis using standardized time series. To appear in **Math. of Operations Res.**

GLYNN, P. and IGLEHART, D. (1989). Importance sampling for stochastic simulations. To appear in **Management Sci.**

HAMMERSLEY, J. and HANDSCOMB, D. (1964). *Monte Carlo Methods*. Methuen, London.

HEIDELBERGER, P. (1987). Discrete event simulations and parallel processing: statistical properties. IBM Reserach Report RC 12733. Yorktown Heights, New York.

HEIDELBERGER, P., CAO, X-R, ZAZANIS, M. and SURI, R. (1988). Convergence properties of infinitesimal perturbation analysis estimates. **Management Sci.** 34, 1281–1302.

HORDIJK, A. IGLEHART, D. and SCHASSBERGER, R. (1976). Discrete-time methods for simulating continuous time Markov chains. **Adv. Appl. Prob.** 8, 772–788.

IGLEHART, D. (1978). The Regenerative method for simulation analysis. In *Current*

Trends in Programming Methodology - Software Modeling. (K. M. Chandy and R. T. Yeh, editors). Prentice-Hall, Englewood Cliffs, NJ, 52–71.

LAW, A. (1975). Efficient estimators for simulated queueing systems. **Management Sci. 22**, 30–41.

REIMAN, M. and WEISS, A. (1986). Sensitivity analysis via likelihood ratios. **Proceedings of the 1986 Winter Simulation Conference**, 285–289.

SHAHABUDDIN, P., NICOLA, V., HEIDELBERGER, P., GOYAL, A., and GLYNN, P. (1988). Variance reduction in mean time to failure simulations. **Proceedings of the 1988 Winter Simulation Conference**, 491–499.

SIEGMUND, D. (1976). Importance sampling in the Monte Carlo study of sequential tests. **Ann. Statist. 4**, 673–684.

WALRAND, J. (1987). Quick simulation of rare event in queueing networks. **Proceedings of the Second International Workshop on Applied Mathematics and Performance/Reliability Models of Computer/Communication Systems.** G. Iazeolla, P. J. Courtois, and O. J. Boxina (eds.). North Holland Publishing Co., Amsterdam, 275–286.

WILSON, J. (1984). Variance reduction techniques for digital simulation. **Amer. J. Math. Management Sci. 4**, 277–312.

Bayesian Inference for Weibull Quantiles

Mark G. Vangel

U.S. Army Materials Technology Laboratory
Watertown MA 02172-0001

The posterior distribution of a two parameter Weibull quantile for a noninformative prior may be obtained exactly (Bogdanoff and Pierce, 1973), although the necessary numerical integration detracts from the usefulness of this result. Credible intervals for this posterior have an alternative frequentist interpretation in terms of conditional tolerance limits (Lawless, 1975).

An approximation to the Lawless procedure was proposed by DiCiccio (1987). This approximation does not involve numerical integration and is of order $O_p(n^{-3/2})$; apparently it is adequate even for samples as small as ten.

The focus of this paper is on the use of DiCiccio's result for the routine calculation of Weibull quantile posteriors. Even a nonbayesian may find the posterior cdf's useful since they provide an easy graphical means for obtaining accurate tolerance limits.

Examples from strength data for composite materials are presented and a specific application of importance to aircraft design is discussed.

References

1. Bogdanoff, D.A. and Pierce, D.A. (1973). JASA, 68, 659.
2. DiCiccio, T.J. (1987). Technometrics, 29, 33.
3. Lawless, J.F. (1975). Technomterics, 17, 255.

1. Introduction

The Weibull model is widely used to represent failure data in engineering applications. One reason is because the Weibull distribution is the limiting distribution of the suitably normalized minimum of a sample of positive iid random variables under quite general conditions (Barlow and Proschan, 1975, ch. 6). The model is therefore appropriate for the strength of a system composed of a string of many links where the strength of the links are iid and the system fails when the weakest link fails (Bury, 1975, ch. 16). An example of a physical system which can be modeled in this way is the strength of a brittle fiber in tension. Another reason why the Weibull model is used is that the distribution is very flexible and consequently it often fits data well.

Inference for the Weibull distribution (or, equivalently, for the extreme value distribution, which is the distribution of the logarithm of a Weibull random variable) is complicated by the fact that the Weibull is not in the exponential family, and consequently the minimal sufficient statistics are the entire sample. Also, although MLE's are easily obtained iteratively, the distributions of the MLE's and pivotals based on the MLE's can not be obtained in closed form. The same is true of linear estimators of the Weibull parameters.

At least three approaches to Weibull inference have been taken. The first is to tabulate approximate quantiles of the pivotals obtained by Monte Carlo. From these tables one can obtain confidence intervals on parameters as well as confidence intervals on quantiles (tolerance limits) for complete samples (Thoman, Bain and Antle 1969, 1970). A problem arises for incomplete samples, since tables must be prepared by simulation for each censoring configuration. The tables which have been prepared (Billman, Antle and Bain, 1972) are inadequate. A second appraoch is to approximate the distribution of the pivotals (e.g. Lawless and Mann, 1976). These approximations are empirical and consequently they are not very satisfactory from a theoretical point of view.

Finally we reach the third approach, which is the focus of this paper. For any location-scale family (e.g. the extreme value family) and any equivariant estimators of the parameters (e.g. MLE's) the distribution of certain pivotals can by obtained exactly if one conditions on the ancillary statistics. From these pivotals one can get exact conditional confidence bounds and tolerance limits for any sample size. The method is applicable to both complete and Type II censored samples (i.e., samples for which only the r smallest order statistics are observed) and requires no tables. Since the intervals have exact conditional confidence, it follows that they are also exact unconditionally. In addition, this method has the advantage of making use of all of the information with respect to the parameters which is in the data (the parameter estimates are in general not sufficient statistics), though for the Weibull model this does not appear to be a practical concern (Lawless, 1973). This conditional approach is apparently due to Lawless, who introduced it in (Lawless 1972). An exposition of the procedure appears in (Lawless, 1982), which is also useful as a guide to the literature.

If one choses an appropriate noninformative prior distribution for the parameters of a location-scale family, then the posterior distribution either of the parameters or of a quantile conditional on the ancillaries are formally identical to frequentist confidence and one-sided tolerance limits respectively.

Bayesian and frequentist terminology may thus be interchanged freely and I will do so in this paper. This is particularly valuable when discussing tolerance limits, which have a frequentist interpretation which is difficult for nonstatisticians to understand. A posterior cdf of a quantile, however, is immediately understood intuitively. Such a cdf can by used to obtain graphically arbitrary one sided and approximate two sided conditional tolerance limits since for the cases discussed herein these intervals coincide with noninformative prior Bayesian credible intervals.

The main disadvantage of this conditional approach is that it is computationally intensive. Many numerical integrations must be performed for

each dataset as one iteratively approximates the confidence limit.

One goal of this project has been to implement the Lawless procedure for the extreme value distribution in a 'robust' FORTRAN program which can be used with little user interaction. Another goal has been to investigate a recent approximation to the conditional procedure (DiCiccio, 1987) which is accurate to $O_p(n^{-3/2})$. This approximation makes the the calculation of posterior distributions feasible. A FORTRAN program to calculate and plot the posterior distribution of Weibull quantiles which makes use of the DiCiccio result is discussed. The results of a small simulation to assess the accuracy of the approximation are presented, though little effort was spent on the simulation since the order of convergence in probability has been determined.

2. The Weibull Distribution

The cdf of the Weibull distribution is

$$F(x;\ \alpha,\ \beta) = 1 - e^{-(x/\beta)^{\alpha}}$$

where β is a scale parameter and α a shape parameter. Maximum likelihood estimation is straightforward. The following equation is solved by Newton-Raphson for α:

$$\left(\Sigma^{*} x_i^{\alpha} \log x_i\right)\left(\Sigma^{*} x_i^{\alpha}\right)^{-1} - 1/\alpha = 1/r \,\Sigma \log x_i$$

where $x_1 \le x_2 \le \ldots \le x_r$ are the order statistics, $n \ge r$ is the sample size, and

$$\Sigma^{*} w_i = \sum_{i=1}^{r} w_i + (n - r)\, w_r .$$

A FORTRAN subroutine 'WEIMLE' for performing these calculations is given in the appendix.

3. The Extreme Value Distribution

Let X be distributed Weibull with shape α and scale β. The distribution of

$$Y = \log(X)$$

is

$$H(y;\, u,\, b) = G((y-u)/b) = \exp\,(-\exp\,((y-u)/b))$$

where

$$b = 1/\alpha \text{ and } u = \log \beta$$

are scale and location parameters respectively. The location-scale family H(y; u, b) is called the extreme value distribution. Results for the extreme value distribution are easily interpreted in terms of the Weibull distribution, and vice versa.

4. Conditional Inference for Location-Scale Families

The presentation below follows Lawless (1982). The distribution H(y; u, b) is taken to be the extreme value distribution as in the previous section. The parameter estimates $\hat{u}$ and $\hat{b}$ may be taken to be MLE's, but the results hold for any equivariant estimators -- that is any statistics $\tilde{u}$ and $\tilde{b}$ which satisfy

$$\tilde{u}(dy_1+c,\, \ldots,\, dy_r+c) = d\,\tilde{u}(y_1,\, \ldots,\, y_r) + c$$

$$\tilde{b}(dy_1+c,\, \ldots,\, dy_r+c) = d\,\tilde{b}(y_1,\, \ldots,\, y_r)$$

for any c and any $d > 0$. The maximum likelihood estimates are readily seen to be equivariant.

Let the sample size be n and, to allow for Type II censoring, let $r \leq n$ be the number of data values. Denote the density of $G(\cdot)$ by

$$G'((y-u)/b) = 1/b\, g((y-u)/b)$$

First we demonstrate that the following random variables are pivotal; that is they have probability distributions which do not depend on the parameters:

$$Z_1 = (\hat{u} - u)/\hat{b} \qquad Z_2 = \hat{b}/b \qquad Z_3 = (\hat{u} - u)/b.$$

Let y_i $i \leq r$ be the order statistics of a random sample from $H(y)$. Consider the random variables

$$w_i = (y_i - u)/b .$$

The w_i are the order statistics of a random sample from $G(\cdot)$ and hence are obviously pivotal. Since the estimator $\hat{u}$ is assumed to be equivariant we have that

$$\hat{u}(w_1, \ldots, w_r) = \hat{u}((y_1-u)/b, \ldots, (y_r-u)/b) =$$

$$1/b\ (\hat{u}(y_1, \ldots, y_r) - u) = (\hat{u}-u)/b = Z_3 .$$

Hence Z_3 is pivotal. Similarly, Z_2 is a pivot since

$$\hat{b}(w_1, \ldots, w_r) = \hat{b}((y_1-u)/b, \ldots, (y_r-u)/b) = \hat{b}(y_1, \ldots, y_r)/b = Z_2$$

Finally,

$$Z_1 = (\hat{u}-u)/\hat{b} = ((\hat{u}-u)/b)\ (b/\hat{b}) = Z_3/Z_2$$

The quantities $a_i = (x_i - \hat{u})/\hat{b}$ are immediately seen to be ancillary since the $\{a_i\}$ are a random sample from $G(\cdot)$, where $G(\cdot)$ is a completely known distribution. Only r-2 of these ancillary statistics are independent since

$$\hat{u}(a_1, \ldots, a_r) = (\hat{u}(y_1, \ldots, y_r) - \hat{u})/\hat{b} = 0$$

and

$$\hat{b}(a_1, \ldots, a_r) = \hat{b}(y_1, \ldots, y_r)/\hat{b} = 1.$$

The fundamental result upon which conditional inference from a frequentist perspective is based is that the joint pdf of $(Z_1, Z_2, a_1, \ldots, a_{r-2})$ is of the form

$$h(z_1, z_2, \underline{a}) = k(\underline{a}, r, n)\, z_2^{r-1} \left(\prod_{i=1}^{r} g(a_i + z_1 z_2) \right) [\, G(a_r z_2 + z_1 z_2)]^{n-r}$$

where $k(\underline{a}, r, n)$ is a function of $a_1, a_2, \ldots, a_{r-2}$ *only*. The pdf of (Z_1, Z_2) given $\underline{a}$ is of the same form as h above except that the normalizing constant is different.

The proof is straightforward. Begin with the joint pdf of $(y_1, \ldots, y_r)$ and make the change of variables

$$y_i = \hat{b}\, a_i + \hat{u}.$$

The Jacobian of this transformation is a constant given $\underline{a}$. A second change of variables

$$\hat{u} = Z_1\, \hat{b} + u = bZ_1 Z_2 + u$$

$$\hat{b} = Z_2 b$$

gives the desired result.

5. Confidence Intervals for Extreme Value Quantiles

Using the pivotal density derived in the previous section, it is not difficult to obtain exact confidence intervals on quantiles of the extreme value (or equivalently, the Weibull) distribution. Toward this end, we determine the distribution for the scale parameter pivotal Z_2. This result is of interest in it's own right since it leads to confidence intervals on the extreme value scale (or Weibull shape) parameter. To get the density of Z_2, merely integrate out Z_1 from the joint pdf given in the previous section, giving

$$h(z_2 \mid \underline{a}) = k(\underline{a}) \exp(\Sigma a_i z_2)\, z_2^{r-1} / [\Sigma^* \exp(a_i z_2)]^r .$$

Next, make the change of variables

$$Z_p = Z_1 - w_p/Z_2 \text{ and } Z_2 = Z_2$$

where

$$w_p = \ln(-\ln(1-p)).$$

The joint density of Z_2 and Z_p is

$$f(z_p, z_2 \mid \underline{a}) \propto z_2^{r-1} \exp(\Sigma (a_i z_2 + z_p z_2 + w_p)$$

$$\cdot \exp(-\Sigma^* \exp(a_i z_2 + z_p z_2 + w_p)).$$

The cdf of Z_p can be expressed in terms of the density of Z_p and Z_2 as

$$P(Z_p \le t) = \int_0^\infty dz_2 \int_0^t dz_p\, f(z_p, z_2 \mid \underline{a}).$$

Change variables again, this time letting

$$y = \exp(z_p z_2 + w_p)\, \Sigma^* \exp(a_i z_2) \text{ and } z_2 = z_2 .$$

The double integral can now be written as a single integral by recognizing that the integral over y is the incomplete gamma function:

$$P(Z_p \leq t) = \int_0^\infty dz_2\ h(z_2 \mid \underline{a})\ I(\exp(tz_2 + w_p)\ \Sigma^* \exp(a_1 z_2),\ r)$$

where

$$I(s,\ r) = (1/\Gamma(r)) \int_0^s x^{r-1}\ e^{-x}\ dx.$$

Since the pth quantile of an extreme value random variable is

$$x_p = u + w_p b,$$

the pivotal Z_p can be expressed as

$$Z_p = (\hat{u} - x_p)\ /\hat{b}.$$

The probability distribution of Z_p can therefore be used to obtain exact conditional confidence intervals on extreme value quantiles. One first obtains the constant of integration $k(\underline{a})$ numerically. Next, the $P(Z_p \leq t)$ is evaluated numerically for several choices of t until the quantile of the distribution of Z_p is determined to the desired accuracy. Finally, the confidence bound on x_p is trivially obtained by pivoting.

6. A Bayesian Interpretation

Independently of Lawless, Bogdanoff and Pierce (1973) arrived at results identical fo those outlined above from a Bayesian point of view. Bayesian results are much easier to explain to nonstatisticians. This is particularly true for the problem that I'm prinarily interested in, confidence intervals on a quantile, and the advantages of Bayesian motivation for a particular application will be discussed in a later section.

Let (y_i) be the order statistics of a Type II censored sample of size $r \leq n$ from an extreme value distribution. The usual joint noninformative prior for the location parameter (u) and scale parameter (b) of a location-scale family is:

$$\pi(u) \propto \text{constant} \qquad \pi(\log(b)) \propto \text{constant}$$

$$\pi(u, b) \propto 1/b.$$

Using the expression for the extreme value pdf given in a previous section, the corresponding posterior distribution is seen to be

$$\pi(u,b \mid \underline{y}) \propto b^{-(r+1)} \exp\left(\sum_{i=1}^{r} (y_i - u)/b\right)$$

$$\cdot \exp\left(-\Sigma^{*} \exp\left((y_i - u)/b\right)\right).$$

The location parameter is readily integrated out giving

$$\pi(b \mid \underline{y}) \propto (1/b)^{r} \exp\left(\sum_{i=1}^{r} y_i/b\right) / [\Sigma^{*} \exp(y_i/b)]^{r}.$$

The normalizing constant is determined by numerical integration. Inference based on this result will be shown in the next section to be formally equivalent to Weibull inference using the pivotal for the shape parameter.

Let $\psi(u, b)$ be any scalar function of the parameters about which inference is to be made. Assume that $\psi(u, \cdot)$ is monotonically increasing in u for fixed b. If a function can be found which satisfies this condition piecewise, then the following results may still be applied to each monotonic section of the function. Some useful choices for ψ are

$\psi(u, b) = u$ (location parameter)

$\psi(u, b) = u + \log(-\log(1-p))\, b$ (pth quantile)

$$\psi(u, b) = \exp(-\exp((t-u)/b))) \qquad \text{(reliability at time t)}$$

$$\psi(u, b) = u - \gamma b \qquad \text{(mean; } \gamma \text{ is Euler's constant)}$$

Define the inverse function $\eta(s, b)$ by means of the relation

$$\psi[\eta(s, b), b] = s.$$

The posterior cdf of ψ can be expressed as

$$P(\psi(u, b) \le s \mid \underline{y}) = \int_0^\infty P(u \le \eta(s, b) \mid b, \underline{y})\, \pi(b \mid \underline{y})\, db$$

It is easy to show that the conditional distribution of $\lambda = \exp(-u/b)$ given b is the following gamma distribution:

$$\pi(\lambda \mid b, \underline{y}) = (1/\Gamma(r))\, [\Sigma^* \exp(y_i/b))\, \lambda^{r-1} \exp(-\lambda\, \Sigma^* \exp(y_i/b)).$$

Simple algebra also shows that

$$P(u \le \eta(s, b) \mid b, \underline{y}) = P(\lambda \ge \exp(-\eta(s, b)/b) \mid b, \underline{y}).$$

Combining these results, we have finally that

$$P(\psi(u,b) \le s \mid \underline{y}) =$$

$$\int_0^\infty (1 - I(\exp(-\eta(s,b)/b)\, \Sigma^* \exp(y_i/b), r))\, \pi(b \mid \underline{y})\, db$$

where $I(\theta, r)$ denotes the incomplete gamma function

$$I(\theta,r) = (1/\Gamma(r)) \int_0^\theta x^{r-1} \exp(-x)\, dx.$$

For a confidence bound on the pth quantile x_p,

$$\eta(s,b) = s - w_p b$$

and

$$P(x_p \leq s) \propto \int_0^\infty (1 - I(\exp(w_p - s/b\ \Sigma^* \exp(y_i/b)), r)\ \pi(b \mid \underline{y})\ db.$$

The fact that for inference about quantiles and about the shape (or, in terms of the extreme value distribution, scale) parameter the Bayesian approach is equivalent to the Lawless conditional approach will be demonstrated next.

7. Formal Equivalence of Bayesian and Frequentist Results

First we demonstrate that posterior intervals for the scale parameter (b) have an exact frequency interpretation. Let b_1 be such that the posterior probability that b is greater than b_1 is γ. Since

$$P(b \geq b_1) = P(\hat{b}/b \leq \hat{b}/b_1) = \gamma,$$

we make the change of variable

$$z = \hat{b}/b$$

and substitute for the y_i in terms of the a_i to get

$$\gamma = \int_{b_1}^\infty \pi(b \mid \underline{y})\ db = \int_{b_1}^\infty \exp(\Sigma y_i/b)\ b^{-r} / [\Sigma^* \exp(y_i/b)]^r\ db$$

$$\propto \int_0^{\hat{b}/b_1} \exp(\Sigma a_i z)/[\Sigma^* \exp(a_i z)]^r\ z^{r-2}\ dz$$

$$= \int_0^{\hat{b}/b_1} \exp(\Sigma a_i z)/[\Sigma^* \exp(a_i z)]^r \, z^{r-2} \, dz$$

$$= \int_0^{\hat{b}/b_1} h(z \mid \underline{a}) \, dz.$$

To see the equivalence of the results for a quantile, we make the substitution

$$t = (\hat{u} - s)/\hat{b}$$

and note that

$$\exp(w_p - s/b \, (\Sigma^* \exp(y_i/b))) =$$

$$\exp(w_p + t\hat{b}/b - \hat{u}/b \, (\Sigma^* \exp(a_i \hat{b}/b + \hat{u}/b))) =$$

$$\exp(w_p + t\hat{b}/b \, (\Sigma^* \exp(a_i \hat{b}/b))).$$

The change of variable $z = \hat{b}/b$ gives the desired result.

8. The Log Generalized Gamma Distribution

The probability distribution of a generalized gamma random variable T is

$$F_T(x; \alpha, \beta, k) = (\alpha/\Gamma(k)) \, (x^{\alpha k - 1}/\beta^{\alpha k}) \exp(-(x/\beta)^{\alpha}).$$

Details on inference for this family may be found in (Farewell and Prentice, 1977) and (Lawless, 1980). Note that the case $k = 1$ corresponds to the Weibull distribution.

If T has a generalized gamma distribution, then $Y = \log(T)$ can be written in the form $\mu - \sigma W$ where

$$\mu = \log(\beta) + \log(k)/\alpha,$$

$$\sigma = 1/(\alpha k^{1/2})$$

and W has the probability density

$$f_W(w; k) = k^{k-1/2}/\Gamma(k) \exp(k^{1/2}w - \exp(w/k^{1/2})).$$

Y is said to have a log generalized gamma distribution.

By varying k, one obtains a family of location-scale dsitributions ranging from the normal ($k = \infty$) to the extreme value ($k = 1$). Although we will restrict attention to the case $k = 1$, it is straightforward to adapt both the frequentist and the Bayesian approaches to arbitrary fixed k and even to certain regression situations (Jones et. al., 1980).

9. Approximate Inference for the Log Generalized Gamma Distribution

Let $\tilde{\mu}$ and $\tilde{\sigma}$ denote maximum likelihood estimates of μ and σ subject to the constraint

$$y_{p0} = \tilde{\mu} + w_p\tilde{\sigma},$$

that is, the MLE of the pth quantile is required to equal y_{p0}. If $\hat{\mu}$ and $\hat{\sigma}$ denote the unconstrained MLE's, and if $L(\mu, \sigma)$ denotes the log of the log generalized gamma likelihood, then the asymptotic distribution of the statistic

$$V(y_{p0}) = -2[L(\tilde{\mu}, \tilde{\sigma}) - L(\hat{\mu}, \hat{\sigma})]$$

is χ^2 with one degree of freedom. Lawless (1982, sec. 4.2) suggests that inference based on $V(y_{p0})$ is acceptable for moderate to large samples but that the approximation may be inadequate for small samples.

DiCiccio (1987) has applied general techniques of Barndorf-Nielson (1986) in order to develop a computationally inexpensive modification to the signed square root of $V(y_{p0})$ which yields a likelihood ratio based approximation suitable even for quite small sample sizes. The numerical integration required by the exact methods is only troublesome for moderate to large samples; so the approximation is actually of questionable use over the range of sample sizes for which it is inaccurate.

I will not reproduce the details of the DiCiccio approximation here for two reasons. The most important of these is that only the results are presented in (DiCiccio 1987) and to repeat these results without having studied their derivation would serve no purpose. A second reason is that although the approximation is inexpensive to compute, the formulas are messy, and to reproduce them here is to invite typographical errors. Interested readers should refer to (DiCiccio 1987) and to the FORTRAN implementation as subroutine LAWAPX in the appendix.

10. The Accuracy of the Approximation

The DiCiccio approximation can be shown to be accurate to $O_p(n^{-3/2})$ (DiCiccio, 1987, p. 37), so an extensive simulation study of accuracy is unnecessary. The results of a very small such study are presented in Table 1. Samples of sizes ranging from 10 to 30 were taken from Weibull populations with different shape parameters. Both the Lawless and the DiCiccio methods were used to calculate 95 percent lower confidence limits on the tenth percentiles of the Weibull populations and the mean and standard deviation of the percent different between the Lawless result and the DiCiccio approximation were calculated for 100 replicates for each case. One would expect that the approximation error should be a rapidly decreasing function of n, and this is observed to be the case. The quality of the approximation is also seen to be a function of the shape parameter of the population. Halving the shape parameter (from 10 to 5) approximatly doubles the mean percentage error uniformly over sample sizes. Also, the approximation error actually appears to be a function of the number of uncensored values rather

than the overall sample size, which is not surprising. Overall, the DiCiccio result appears to be satisfactory for samples of 10 or more uncensored values, and remarkably good for samples of 30 or more observed values. This conclusion is based partly on the small simulation presented here and partly on experimenting with various cases of real and simulated data.

11. An Application: Composite Material Basis Values

A criterion used both by aircraft designers when choosing a material for a specific application and by the Federal Aviation Administration when certifying a new material for a structural aircraft application is the *material basis value*. A 'B-basis value' is defined to be a lower 95 percent confidence limit on the tenth percentile of the strength distribution of a material and an 'A-basis value' is a 95 percent lower confidence limit on the first percentile. The reason for these tolerance limits, which have been used in the industry for decades, is that a designer is primarily interested in the lower tail of the strength distribution. In order to design a reliable structure, he would like to estimate the stress level at which a material is 90 percent or 99 reliable. A tolerance limit is an attempt to estimate these quantiles in a conservative way. Such conservatism is particularly necessary for advanced composite materials, which typically have relatively high strength variability. Also, advanced materials are generally expensive to manufacture and test, resulting in small sample sizes.

The work presented here has been motivated by a need for improved methodology for calculating basis values and for communicating lower tail quantile information to the engineer. Typically, the engineer who routinely calculates and interprets these numbers has little appreciation for the rather convoluted frequency arguments behind tolerance limits. The long run proportion of times a statistic calculated from successive samples of size n from a hypothetical population is greater than a certain quantile of that population is of little help to the statistically naive. The simple statement that the tenth (first) percentile is greater than the B-basis (A-basis)

value with 95 percent probability is much more direct and intuitive. Also, the Bayesian approach presents all of the information in the data about the lower tail quantile of interest, which is what should be the ultimate concern of the engineer anyway. The fact that the tolerance limit is only a convenient summary statistic of this distribution becomes clear when the user is presented with the entire posterior and shown how to determine arbitrary tolerance limits graphically.

Table 2 presents B-basis value calculations for a graphite fiber/epoxy material made by four fabricators. Note the agreement between the DiCiccio and the Lawless calculations. Figure 1 consists of the four tenth percentile posteriors. Not only do two of the fabricators have nearly the same B-basis value, they also have virtually identical quantile posteriors. Several questions immediately come to mind: Why did the other two manufacturers produce substantially weaker material?; Are other lower tail quantile posteriors for the two 'similar' fabricators as close together?; etc. Examining the posterior rather than a summary statistic of the posterior leads to insight into the data that might not otherwise be apparent. Figure 2 demonstrates that the B-basis value can be retrieved graphically.

Table 3, Figure 3 amd Figure 4 present corresponding results for another material: woven Kevlar fibers in an epoxy matrix. These data show much less fabricator-to-fabricator variability than do the graphite/epoxy data. This can readily be seen from the tolerance limit calculations. The fact that there is essentially no evidence in the data to suggest that the fabricators differ with respect to the tenth percentiles of their strength distributions is made particularly clear by the 'overlapping' posteriors of this quantile.

12. Conclusion

This paper reviews two results related to conditional inference in location-scale families, emphasizing inference on Weibull quantiles. These methods are due to Lawless (1972) and Bogdanoff and Pierce (1973). For the case of inference on quantiles both procedures are equivalent, though the

former is motivated by frequency considerations, while the latter is derived from a Bayesian point of view. The recent work of DiCiccio (1987) greately reduces the computational burden of both methods with little loss of accuracy.

The advantages of the Bayesian interpretation, at least for inference on quantiles, has been demonstrated by means of an example from an engineering application.

13. References

Barndorff-Nielsen O. E. (1986). "Inference on full or partial parameters, based on standardized signed log likelihood ratio", *Biometrika*, 73, 307-322.

Barlow, R. E. and F. Proschan (1975). *Statistical Theory of Reliability and Life Testing*, New York: Holt, Rinehart and Winston.

Billmann B., C. Antle and L. J. Bain (1972). "Statistical inference from censored Weibull samples", *Technometrics*, 14, 831-840.

Bogdanoff, D. and D. A. Pierce (1973). "Bayes-Fiducial inference for the Weibull distribution", *J. Am. Stat. Assoc.*, 68, 659-664.

Bury, K. (1975), *Statistical Models in Applied Science*, New York: John Wiley and Sons.

DiCiccio, T. J. (1987). "Approximate inference for the generalized gamma distribution ", *Technometrics*, 29, 33-40.

Farewell V. T. and Prentice R. L. (1977). "A study of distributional shape in life testing", *Technometrics*, 19, 69-75.

Jones, R., F. Scholtz, M. Ossiander and G. Shorack (1985). "Tolerance bounds for log gamma regression models", *Technometrics* ,27, 109-118.

Lawless, J. F. (1972). "Confidence Interval Estimation for the parameters of the Weibull distribution", *Utilitas Mathematicae*, 2, 71-87.

Lawless, J. F. (1973). "Conditional inference for the parameters of the Weibull distribution", *J. Am. Stat. Assoc.*, 69, 665-668.

Lawless, J. F. (1980). "Inference in the generalized gamma and log gamma distributions", *Technometrics*, 22, 409-419.

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*, New York: John Wiley and Sons.

Lawless, J. F. and N. R. Mann (1976). "Tests for homogeneity for extreme value scale parameters", *Commun. Stat.*, A5, 389-405.

Thoman, D. R., L. J. Bain and C. E. Antle (1969). "Inference on the parameters of the Weibull dsitribution", *Technometrics*, 11, 805-816.

Thoman, D. R., L. J. Bain and C. E. Antle (1970). "Reliability and tolerance limits in the Weibull distribution", *Technometrics*, 12, 363-371.

TABLE 1

Accuracy of the DiCiccio approximation

A simulation of 95% lower confidence bounds on the 10th percentile using the Weibull distribution with 100 replicates per case was performed. The results are summarized below:

n	r	Shape	Scale	At	Stan. error of mean
10	10	10	1	1.09	.029
20	20	10	1	.380	.0086
10	10	5	1	2.20	.059
20	20	5	1	.761	.017
30	20	5	1	.752	.016

TABLE 2

Carbon fiber / Epoxy specimen tensile strength data
95% LCB on 10th percentile

Fabricator	n	Estimates (KSI)	
		Lawless	DiCiccio
A	48	244.1	244.3
B	36	271.4	271.6
C	33	228.2	228.5
D	25	269.5	269.8

TABLE 3

Kevlar fabric / epoxy specimen tensile strength data
95% LCB on 10th percentile

Fabricator	n	Estimates (KSI)	
		Lawless	DiCiccio
A	23	77.58	77.64
B	18	76.36	76.50
C	30	77.18	77.23
D	10	78.45	78.64

Figure 1

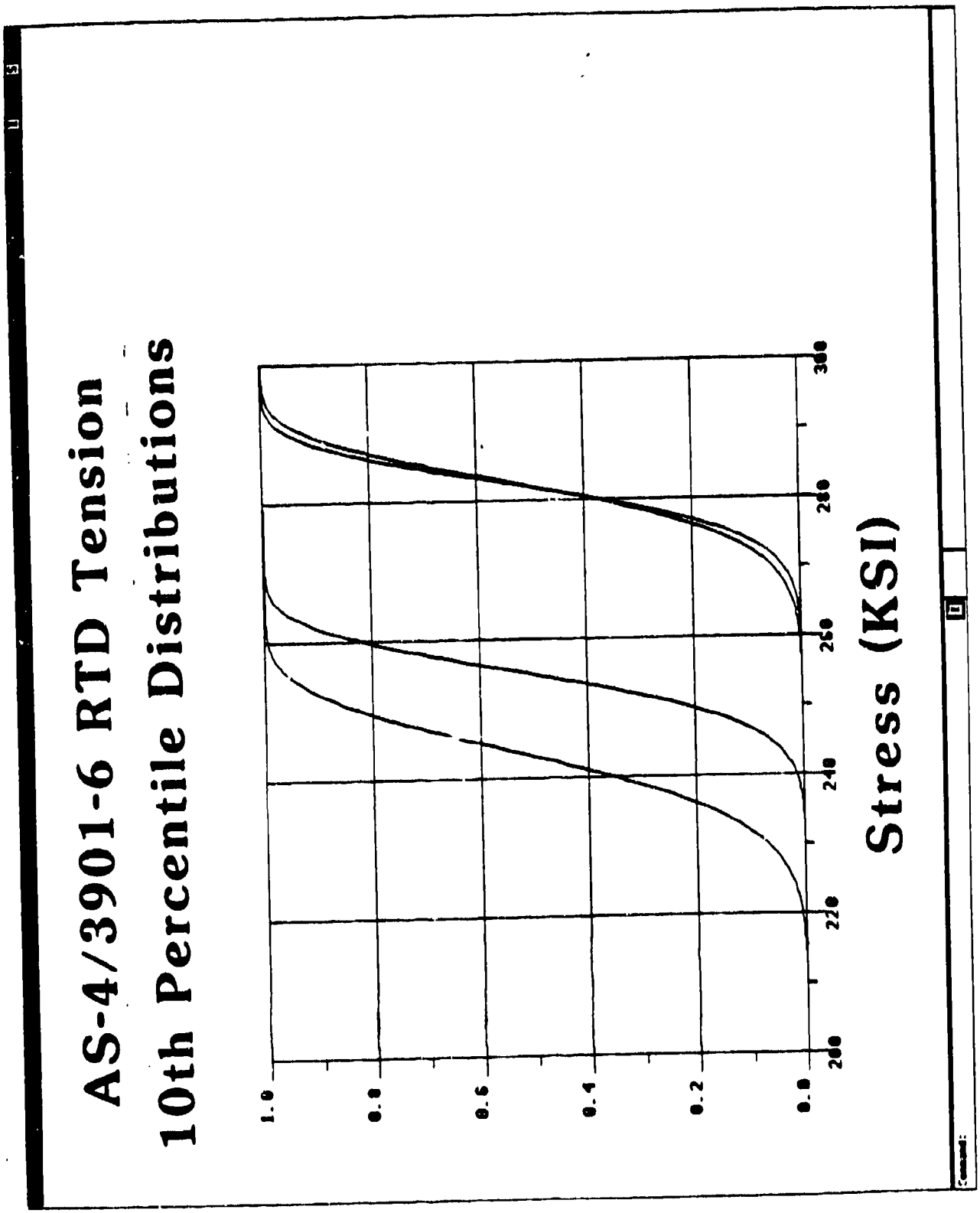

AS-4/3901-6 RTD Tension
10th Percentile Distributions
1.0
0.8
0.6
0.4
0.2
0.0
200
220
240
260
280
Stress (KSI)

Figure 2

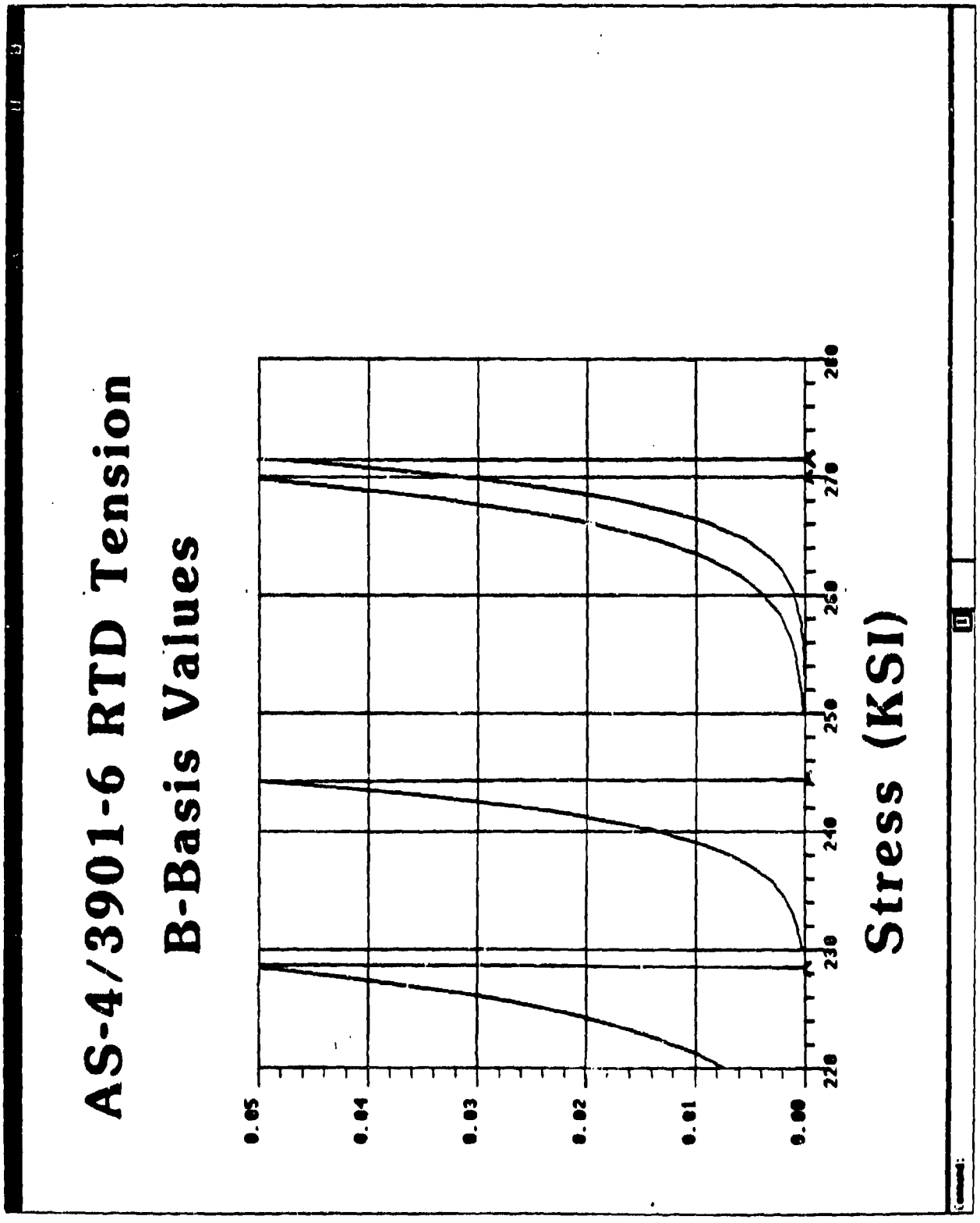

AS-4/3901-6 RTD Tension
B-Basis Values
Stress (KSI)
0.05
0.04
0.03
0.02
0.01
0.00
220
230
240
250
260
270
280

Figure 3

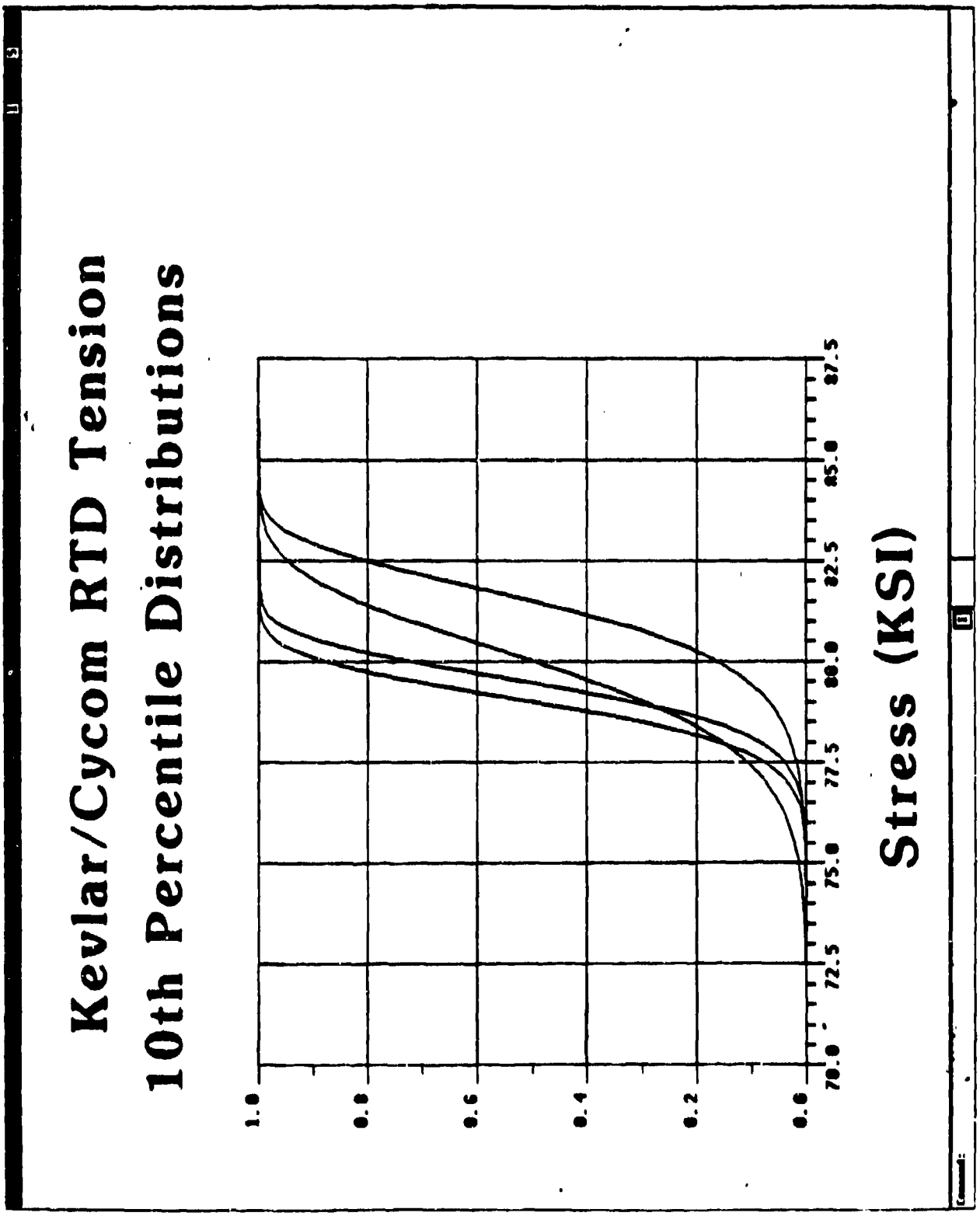
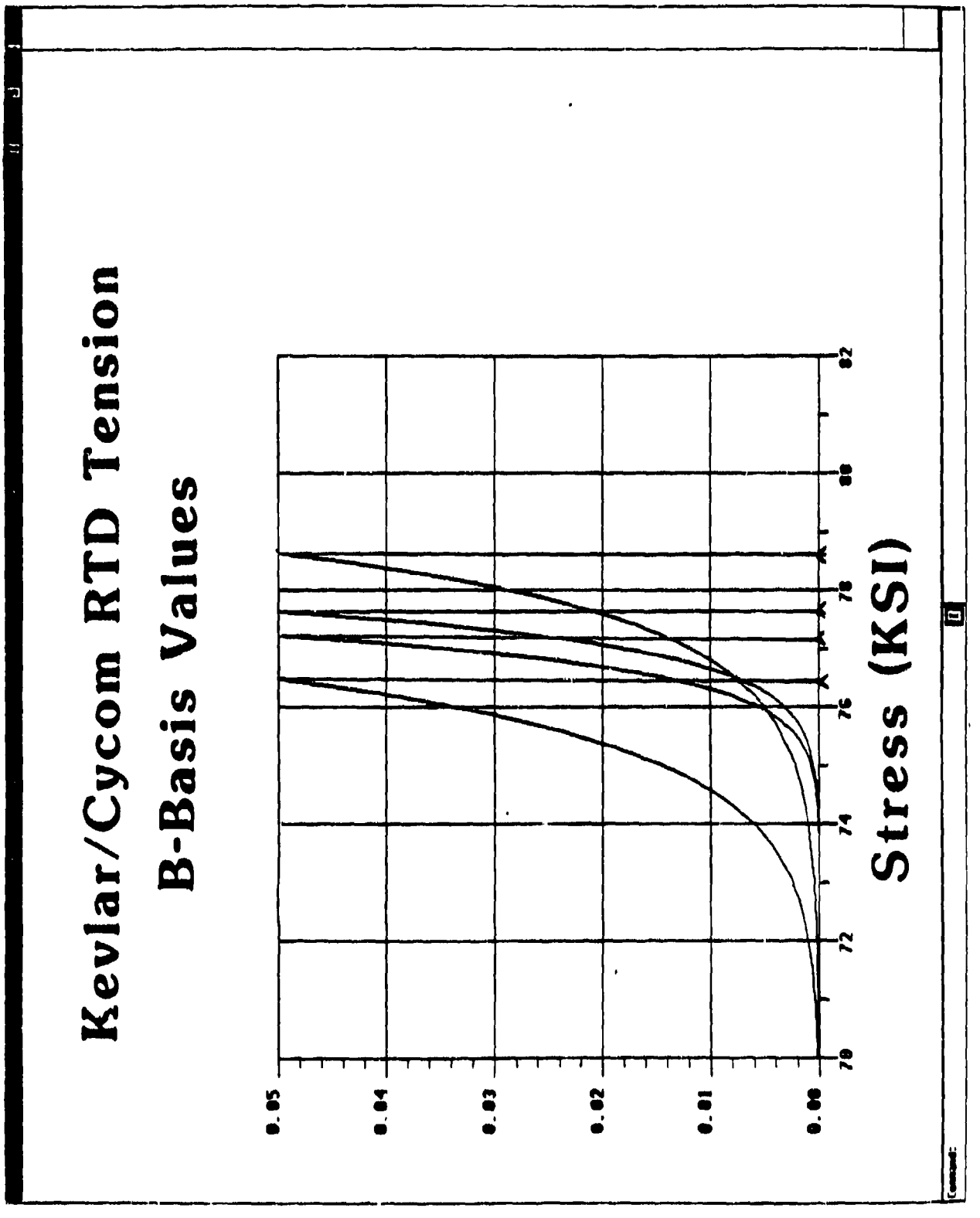

Kevlar/Cycom RTD Tension
10th Percentile Distributions
1.0
0.8
0.6
0.4
0.2
0.0
70.0
72.5
75.0
77.5
80.0
82.5
85.0
87.5
Stress (KSI)

Figure 4


Kevlar/Cycom RTD Tension
B-Basis Values
Stress (KSI)
0.05
0.04
0.03
0.02
0.01
0.00
70
72
74
76
78
80
82

Appendix: FORTRAN Listings

The following programs were developed on an Alliant FX/8 and should run with little modification on any 32 bit machine. However, the software has not been tested to the point where it can be considered error free. The programs are provides as a guide to an individual wishing to implement the algorithms discussed in this paper.

```
      program laupgm
c
c      Mark Vangel, 5/20/88
c
c         Program to implement Lawless' procedure for
c       conditional confidence intervals on quantiles
c       for a location/scale family.  The family chosen
c       here is extreme value.  Data may be Type II censored.
c       Note that conditioning on the ancillaries gives the
c       equivalent of an HPD region for a noninformative
c       prior.
c
      implicit double precision (a-h, o-z)
      parameter     (imax = 500)
      character*12   flenme
      dimension      x(imax)
c
      common /dat/ x
      common /ca/  cnorm, suma, n,  k
      common /cb/  p,     gam,  wb, t
      common /cd/  tol
      data one /1.d0/
c
      data coarse, fine, eps /1.d-2, 1.d-7, 1.d-5/
c
c      -- Output unit number and filename
      write (6, *) 'Output unit number ?'
      read  (5, *)  iout
      if (iout .ne. 6) then
         write (6, *) 'Filename ?'
         read  (5,'(a12)') flenme
         open (unit=iout, file=flenme, status='new')
      end if
c
c      -- This program is tested with random data.
c         It can also be used for data from a file.
c         The first record of the input file has the
c         sample size and the number of uncensored values.
      write (6, *) 'Enter 1 for data from file '
      write (6, *) '      0 for random data.'
      read  (5, *)  idat
      if (idat .eq. 1) then
         write (6, *) 'Filename ?'
         read  (5, '(a12)') flenme
         open (unit=iout+1, file=flenme, status='old')
         read (iout+1, *) n, k
c
c      -- Note: the first field on each remaining record is a
c         batch indicator not used for this program.
         do 10 i=1, k
            read (iout+1, *) dummy, x (i)
10          continue
            call dsvrgn (k, x, x)
      else
         write (6, *) 'Seed ?'
         read  (5, * )  iseed
         write (6, *) 'Weibull shape and scale ?'
         read  (5, *)  shp, scl
         write (6, *) 'Sample size ?'
         read  (5, *)  n
         write (6, *) 'Number uncensored ?'
         read  (5, *)  k
```

```
c
c       -- Get the pseudo-random sample
          call rnset  (iseed)
          call drnwib (n, shp, x)
          call dsvrgn (n, x, x)
          do 20 i=1, k
             x (i) = scl *x (i)
20         continue
        end if
c
c       -- Get the Weibull MLE's
        call wnrmle (eshp, escl, n, k, x, eps, iter, 100)
c
c       -- Extreme value location (xmu) and scale (xsg) estimates
        xmu = dlog (escl)
        xsg = one /eshp
c
c       -- Write out what we have so far.
        write (iout, *)
        write (iout, *) 'The Lawless conditional procedure'
        write (iout, *)
        if (idat .ne. 1) then
        write (iout, *) 'Seed                      : ', iseed
        end if
        write (iout, *) 'Sample size               : ', n
        write (iout, *) 'Number uncensored         : ', k
        if (idat .ne. 1) then
        write (iout, *) 'Weibull shape, scale      : ', shp, scl
        end if
        write (iout, *) 'Weibull MLEs              : ', eshp, escl
        write (iout, *) 'Extreme value location    : ', xmu
        write (iout, *) 'Extreme value scale       : ', xsg
        write (iout, *)
        write (iout, *) 'Weibull data :'
        write (iout, 100) (x(i), i=1, k)
100     format (f10.4)
c
c       -- Now, calculate the tolerance limit using both the
c          exact method and DiCiccio's approximation.
c
        write (6, *)
        write (6, *) 'Lower confidence bound calculation'
        write (6, *) 'Quantile ?'
        read  (5, *)  p
        write (6, *) 'Confidence coefficient ?'
        read  (5, *)  gam
c
c       -- Lawless conditional procedure (Technometrics, 1975)
        call lawles (x, n, k, p, gam, etol)
c
c       -- DiCiccio's approximation (Technometrics, 1987)
        call lawapx (x, n, k, p, gam, atol)
c
c       -- Write out the tolerance limit results
        write (iout, *)
        write (iout, *) 'Lower confidence bound on a quantile'
        write (iout, *) '     (Extreme Value and Weibull)'
        write (iout, *) 'Probability               : ', p
        write (iout, *) 'Confidence                : ', gam
        write (iout, *) 'Extreme value quantile    : ', xp
        write (iout, *) 'Lawless tolerance limits  : ', etol, log(etol
        write (iout, *) 'Approximations            : ', atol, log(atol
```

```
c         write (*,*) '1=cdf, 2=pdf ?'
c         read  (*,*)  itype
          itype = 1
          write (*,*) 'Min and max for abscissa ?'
          read  (*,*)  amin, amax
          write (*,*) 'Min and max for ordinate [0,0 for default] ?'
          read  (*,*)  amin, amax
          call initt (960)
          call binitt
          call comset (ibasex(11), amin)
          call comset (ibasex(12), amax)
          if (amax .ne. zero) then
             call comset (ibasey(11), amin)
             call comset (ibasey(12), amax)
          end if
c
          if (itype .eq. 1) then
              call npts  (nplot *nq)
              call check (quant, cdf)
              call npts  (nq)
              call dsplay (quant, cdf)
              do 30 i=1, nplot-1
                call cplot (quant (i*nq+1), cdf (i*nq+1))
 0            continue
          else if (itype .eq. 2) then
              call npts    (nplot *nq)
              call check   (quant, dens)
              call npts    (nq)
              call dsplay (quant, dens)
              do 40 i=1, nplot-1
                call cplot (quant (i*nq+1), dens (i*nq+1))
40            continue
          end if
          call movabs (0, 1000)
          call anmode
          read (*,*)
          go to 2
       end if
c

       write (*,*) 'Quit ?'
       read  (*,'(a1)') ans
       if (ans .ne. 'y') go to 2
       call finitt (0, 700)
c
       stop
       end
```

```
      program petplt
c
c       Mark Vangel, July 1983
c
c          Program to calculate and plot the posterior of a
c       percentile for the Weibull model.  This program calls
c       subroutines from the Tektronix Plot 10 library.
c
      character*20 filenme
      character*1  ans
      dimension x (1000), cdf (5000), dens(5000), quant (5000),
     $          ipoint (100)
      data zero, one /0.d0, 1.d0/
c
      ipoint (1) = 0
      nq         = 100
      nplot      = 0
c
c       -- Loop over all files.
      write (*,*) '0=brief output, 1=complete output ?'
      read  (*,*)  ibrief
1     continue
         write (*,*) 'Filename ?'
         read  (*,'(a20)') filenme
         if (filenme .eq. ' ') go to 2
         nplot = nplot +1
         open (unit=10, file=filenme)
         read (10,*) nsamp, nobs
         do 10 i=1, nobs
            read (10,*) idum, x (ipoint (nplot) +i)
10       continue
         close (10)
         ipoint (nplot+1) = ipoint (nplot) +nobs
c
c       -- Calculate the posterior for a specified quantile
         write (*,*) 'Percentile ?'
         read  (*,*)  p
         if (p .gt. one) p = p/100
         write (*,*) 'Range of values for posterior ?'
         read  (*,*) qmin, qmax
         dq = (qmax-qmin)/(nq-1.)
         write (*,*) ' Posterior  of ',100*p,' percentile '
         write (*,*) '     from ',qmin, ' to ', qmax
         write (*,*)
         do 20 i=1, nq
            idx           = (nplot -1)*nq +i
            quant (idx)   = (i-1) *dq +qmin
            call lswapx (x (ipoint(nplot) +1), nsamp, nobs,
     $           quant (idx) , p, cdf (idx))
            if (ibrief .eq. 1) then
               write (*,*) quant (idx), cdf (idx)
            end if
20       continue
      go to 1
c
c       -- Now plot the results
2     continue
      write (*,*) 'Plots ?'
      read  (*,'(a1)') ans
      if (ans .eq. 'y') then
c
c       -- Density calculations not yet implemented.
```

```
c         write (*,*) '1=cdf, 2=pdf ?'
c         read  (*,*)  itype
          itype = 1
          write (*,*) 'Min and max for abscissa ?'
          read  (*,*)  qmin, qmax
          write (*,*) 'Min and max for ordinate [0,0 for default] ?'
          read  (*,*)  omin, omax
          call initt (960)
          call binitt
          call comset (ibasex(11), qmin)
          call comset (ibasex(12), qmax)
          if (omax .ne. zero) then
             call comset (ibasey(11), omin)
             call comset (ibasey(12), omax)
          end if
c
          if (itype .eq. 1) then
              call npts  (nplot*nq)
              call check (quant, cdf)
              call npts  (nq)
              call display (quant, cdf)
              do 30 i=1, nplot-1
                call cplot (quant (i*nq+1), cdf (i*nq+1))
 0            continue
          else if (itype .eq. 2) then
              call npts    (nplot*nq)
              call check   (quant, dens)
              call npts    (nq)
              call display (quant, dens)
              do 40 i=1, nplot-1
                call cplot (quant (i*nq+1), dens (i*nq+1))
40            continue
          end if
          call movabs (0, 1000)
          call anmode
          read (*,*)
          go to 2
       end if
c
       write (*,*) 'Quit ?'
       read  (*,'(a1)') ans
       if (ans .ne. 'y') go to 2
       call finitt (0, 700)
c
       stop
       end
```

```
      program lawsim
c
c      Mark Vangel, 10/14/88
c
c         Program to test by simulation an approximation of
c      DiCiccio (1987) to the Lawless conditional procedure.
c
      implicit double precision (a-h, o-z)
      parameter     (imax = 500)
      character*12   flenme
      dimension      x(imax)
c
      common /dat/ x
      common /ca/  cnorm, suma, n,  k
      common /cb/  p,     gam,  wb, t
      common /cd/  tol
c
      data coarse, fine /1.d-2, 1.d-7/
c
c      -- Output unit number and filename
      write (6, *) 'Output unit number ?'
      read  (5, *)  iout
      if (iout .ne. 6) then
         write (6, *) 'Filename ?'
         read  (5,'(a12)') flenme
         open (unit=iout, file=flenme, status='new')
      end if
c
c      -- Get simulation parameters
      write (6, *) 'Seed ?'
      read  (5, * )  iseed
      write (6, *) 'Weibull shape and scale ?'
      read  (5, *)  shp, scl
      write (6, *) 'Sample size ?'
      read  (5, *)  n
      write (6, *) 'Number uncensored ?'
      read  (5, *)  k
      write (6, *) 'Number of replicates ?'
      read  (5, *)  nsim
      write (6, *) 'Quantile ?'
      read  (5, *)  p
      write (6, *) 'Confidence coefficient ?'
      read  (5, *)  gam
c
c      -- Write out what we have so far.
      write (iout, *)
      write (iout, *) 'The Lawless conditional procedure'
      write (iout, *)
      write (iout, *) 'Seed                        : ', iseed
      write (iout, *) 'Number of replicates        : ', nsim
      write (iout, *) 'Sample size                 : ', n
      write (iout, *) 'Number uncensored           : ', k
      write (iout, *) 'Quantile                    : ', p
      write (iout, *) 'Confidence coefficient      : ', gam
      write (iout, *) 'Weibull shape, scale        : ', shp, scl
      write (iout, *)
c
c      -- Loop over the number of replicates
      call rnset  (iseed)
      s  = 0.d0
      ss = 0.d0
      do 99 i=1, nsim
```

```
c
c       -- Get the pseudo-random sample
          call drnwib (n, shu, k)
          call dsvrgn (n, x, k)
          do 20 i=1, k
             x (i) = scl *x (i)
20        continue
c
c       -- Lawless conditional procedure (Technometrics, 1975)
          call lawles (x, n, k, n, gam, etol)

c
c       -- DiCiccio's approximation (Technometrics, 1987)
          call lawapx (x, n, k, n, gam, atol)

c
c       -- Measure approximation in terms of percent error
          dpct = (atol -etol) /etol *100.

c
c       -- Write out the tolerance limit results
          write (iout, *) 'Lawless , DiCiccio, delta =  ', i ',
     $              etol, atol, dpct
          s  = s +dpct
          ss = ss +dpct *dpct
99      continue
c
c       -- Mean and stan. dev. of % error
        s = s /nsim
        ss = sqrt ((ss - nsim *s *s) /(nsim *(nsim -1.)))
        write (iout, *)
        write (iout, *) 'Average % error                    ', s
        write (iout, *) 'Standard deviation of mean % error ', ss

c
        stop
        end
```

```
      subroutine lawles (x, na, ka, pa, gama, etol)
c
c       Mark Vangel, May 1988
c
c          subroutine 'lawles' calculates one sided lower tolerance limits
c       for the Weibull model using the Lawless conditional procedure.
c       This routine performs 'exact' calculations by numerical
c       integration.  For even moderately large samples, the Pidiccic
c       approximation used in subroutine 'lawapx' is very accurate
c       and computationally less troublesome than the method used here.
c
c       x       --  Data                                          (Input)
c       n       --  Total sample size                             (Input)
c       k       --  Number of (uncensored) observations           (Input)
c       p       --  Probability associated with quantile xp       (Input)
c       g       --  Confidence level for lcb on xp                (Input)
c       etol    --  'Exact' lower tolerance limit                 (Output)
c
c
c       Ref : Lawless, Technometrics 1975
c
      implicit double precision (a-h, o-z)
      dimension x (1)
c
      common /par/ eshp, escl, xmu, xsg
      common /ca/  cnorm, suma, n,  k
      common /cb/  p,     gam,  wp, t
      common /cd/  tol
c
      external xntgnd, xconf
      data coarse, eps, fine /1.d-2, 1.d-5, 1.d-7/
      data zero, one, cl, ch /0.d0, 1.d0, .5d0, 1.2d0/
c
c       -- Put stuff in common
      n   = na
      k   = ka
      p   = pa
      gam = gama
c
c       -- Get the Weibull MLE's
      call wnrmle (eshp, escl, n, k, x, eps, iter, 100)
c
c       -- Tolerance limit factor range.  This range is broad
c          enough for virtually any practical confidence coefficient.
c          'iconf' provides a first approximation to the tolerance
c           limit.  Note that 'iconf' returns lower tolerance limits.
      gam    = one -gam
      call iconf (x, n, k, p, gam, tollmt)
      gam    = one -gam
      tollmt = log ((escl /tollmt) **eshp)
      zl     = cl  *tollmt
      zh     = ch  *tollmt
      write (*,*) ' Lawles : First guess = ', log (tollmt)
c
c       -- Now go to log scale.
      do 10 i=1, k
         x (i) = log (x (i))
10    continue
c
c       -- Extreme value location (xmu) and scale (xsg) estimates
      xmu = log (escl)
      xsg = one /eshp
```

```
c
c       -- We need the ancillaries and their sum
        suma = zero
        do 20 i=1, k
           suma  = suma +(x (i) -xmu) /xsg
20      continue
c
c       -- Next we obtain the constant of integration.
c          Use an adaptive quadrature routine to integrate
c           XNTGND on (0, infinity).
        cnorm  = one
        abserr = zero
        relerr = eps
        call dadagi (xntgnd, zero, 1, abserr, relerr, xk, err)
c
c       -- The normalizing constant
        cnorm = one /xk
        write (6, *)
        write (6, *) 'The normalizing constant is  ', cnorm
c
c       -- pth quantile for standard extreme value distribution
        wp = log (-log (one -p))
c
c       -- Estimated pth quantile for data
        zp = xmu +xsg *wp

c
c       -- Brent's algorithm to find tolerance limit factor such
c          that the integral of XNTGND from (0, infinity) equals 'part'.
c          The first pass has a large error tolerance to save time.
c          The second pass uses the final tolerance.
c
        maxit  = 100
        abserr = zero
        relerr = coarse
        tol    = coarse
        call dzbren (xconf, abserr, relerr, zl, zh, maxit)
c
c       -- Move in for the kill.
        maxit  = 100
        zl     = .95d0  *zh
        zh     = 1.05d0 *zh
        relerr = fine
        tol    = fine
        call dzbren (xconf, abserr, relerr, zl, zh, maxit)
        tolfct = zh
c
c       -- Calculate the tolerance limit
        etol = exp (xmu +zh *xsg)
c
c       -- Restore the data
        do 30 i=1, k
           x (i) = exp (x (i))
30      continue
c
        return
        end
```

```
      subroutine lawapx (x, nsampa, nobsa, pa, gama, atol)
c
c      Mark Vangel, July 1988
c
c         An excellent approximation to the one-sided Weibull
c      conditional  tolerance limits of Lawless (1975).
c
c      x      --  Data                                          (Input)
c      nsamp  --  Total sample size                             (Input)
c      nobs   --  Number of (uncensored) observations           (Input)
c      p      --  Probability associated with quantile xp      (Input)
c      g      --  Confidence level for lcb on xp               (Input)
c      atol   --  Approximate lower tolerance limit            (Output)
c
c      Ref :  Lawless, Technometrics 1975
c             Dicicclo, Technometrics 1987
c
      implicit double precision (a-h, o-z)
      dimension  x (1)
      common /par/ ushp, uscl, xmu, xsg
      common /conf/  p, gam, nsamp, nobs
      external aconfi
      data maxit /100/, eps /1.d-5/
      data coarse, fine /1.d-2, 1.d-7/
      data zero, one /0.d0, 1.d0/
      data cl, ch, dl, dh /.8d0, 1.2d0, .95d0, 1.05d0/
c
c      -- Put stuff in common
      p     =  pa
      gam   =  one -gama
      nobs  =  nobsa
      nsamp =  nsampa
c
c      -- Get the MLE's of the Weibull parameters
      call wnrmle (ushp, uscl, nsamp, nobs, x, eps, iter, maxit)
c
c      -- Tolerance limit factor range.  This range is broad
c         enough for virtually any practical confidence coefficient.
c         'iconf' provides a first approximation to the tolerance
c          limit.
      call iconf (x, nsamp, nobs, p, gam, tollmt)
      zl     = cl  *tollmt
      zh     = ch  *tollmt
      write (*,*) ' Lawapx : First guess = ', tollmt
c
c      -- Brent's algorithm to find tolerance limit factor at
c         the desired confidence level. The first pass has a
c         large error tolerance to save time. The second pass
c         uses the final tolerance.
c
      maxit  = 100
      abserr = zero
      relerr = coarse
      tol    = coarse
      call dzbren (aconfi, abserr, relerr, zl, zh, maxit)
c
c      -- Move in for the kill.
      maxit  = 100
      zl     = dl  *zh
```

```
      zh     = dh  *zh
      relerr = fine
      tol    = fine
      call dzbren (acenfi, abserr, relerr, zl, zh, maxit)
      atol = zh
c
      return
      end
```

```
      subroutine iconf (x, nsamp, nobs, p, g, tollmt)
c
c      Mark Vangel, October 1988
c
c      A non-iterative first  approximation to the Lawless conditional
c     procedure (or, alternatively, to the posterior of a quantile under
c     a flat prior). The routine is written for two paramter Weibull
c     analysis, but extension to the generalized log gamma family is
c     straightforward. This routine returns the estimated confidence
c     limit for a provided probability level and confidence (hence
c     the 'i' ~ for inverse ~ in the routine name).  This routine is
c     approximately inverse to 'aconf'.  It provides the same result
c     as 'lawapx' but with a less accurate approximation.
c
c    Ref : Diciccio, T.J., Technometrics 1987 p.33
c
c       x      --  Data                                          (Input)
c       nsamp  --  Total sample size                             (Input)
c       nobs   --  Number of (uncensored) observations           (Input)
c       g      --  Confidence level for lcb on xp               (Input)
c       p      --  Probability associated with quantile xp      (Input)
c       tollmt --  Upper tolerance limit                        (Output)
c
      implicit double precision (a-h, o-z)
      dimension x (1), s (4)
      data eps, maxit /1.d-5, 100/
      data zero, one, half, thhalf, fvhalf
     $           /0.d0, 1.d0, 0.5d0, 1.5d0, 2.5d0/
c
c      -- Get the Weibull MLE's
      call wnrmle (ushp, uscl, nsamp, nobs, x, eps, iter, ier)
c
c      -- Transform to extreme value distribution.
      uloc = log (uscl)
      uscl =  one /ushp
c
c      -- Calculate the derivatives of the log likelihood at
c          the MLE.
      wp = log (-log (one -p))
      do 30 j=1, 4
         s(j) = zero
         do 40 i=1, nobs
            z    = (log(x(i)) -uloc) /uscl
            t    = ((z -wp) **j) *exp (z)
            s(j) = s(j) +t
40       continue
         s(j) = s(j) +(nsamp -nobs) *t
30    continue
c
      d20 = -nobs
      d11 = -s(1)
      d02 = -(nobs +s(2))
      c30 = nobs
      d21 = 2*nobs +s(1)
      d12 = 3*s(1) +s(2)
      d03 = nobs +3*s(2) +s(3)
      d40 = -nobs
      d31 = -(3*nobs +s(1))
      d22 = -(4*nobs +5*s(1) +s(2))
      d13 = -(7*s(1) +6*s(2) +s(3))
      d04 = -nobs -(7*s(2) +6*s(3) +s(4))
```

```
c       -- The approximate mean and variance of r, the signed
c          square root of -2 times likelihood ratio, is calculated in
c          terms of the dij.
        v11 = -d02 /(d20 *d02 -d11 *d11)
c
        a   = (v11 **half *(-d12/d02 +d11*d03 /d02**2)) /2
        b   = (v11       *(-d22/d02 +(d21*d03 +2*d13*d11 +d12**2) /d02**2
     $     -(4*d12*d11*d03 +d04*d11**2) /d02**3
     $     +2*(d11*d03 /d02**2) **2)) /4
        c   = (v11 **thhalf *(d30 -3*d21*d11 /d02 +3*d12 *(d11 /d02) **2
     $     -d03 *(d11/d02) **3)) /6
        d   = (v11 **2 *(d40 -(4*d31*d11 +3*d21**2) /d02
     $     +(6*d22*d11**2 +12*d21*d12*d11) /d02**2
     $     -(6*d21*d03*d11**2 +4*d13*d11**2 +12*(d12*d11)**2) /d02**3
     $     +(12*d12*d03*d11**3 +d04*d11**4) /d02**4
     $     -3*d11**4 *d03**2 /d02**5)) /24
c
        xmu = a +2*c
        xsg = sqrt (one +2*(b +3*d +a*c) +11*c**2)
c
c       -- Diciccio, equation 6.
        zg      =  dnorin (g)
        rg      = xmu  +xsg  *zg
        yp      = uloc +uscl *wn
        tollmt = yp +uscl *(sqrt (v11) *(rg +c*rg **2 +
     $           (fvhalf *c **2 +d) *rg **3))
c
c       -- Go to Weibull scale.
        tollmt = exp (tollmt)
c
        return
        end
```

```
      double precision function xntgnd (z)
c
c          Proportional to the pdf of a certain pivotal quantity.
c     The normalizing constant, 'cnorm', depends on the
c     data and must be obtained by a preliminary numerical
c     integration.  Once 'cnorm' is known, XNTGND becomes a
c     pdf and is used by YNTGND during the primary numerical
c     integration to get the tolerance limit.
c
c     Note : 'cnorm' must be initialized (to 1) before the
c            preliminary integration.  Following that, 'cnorm'
c            can be assigned the value which makes XNTGND a pdf.
c
      implicit double precision (a-h, o-z)
      dimension x(1)
c
c      -- 's1' is used by YNTGND, hence the common block.
      common /dat/   x
      common /cc/    s1
      common /ca/    cnorm, sumx, n, k
      common /par/ eshn, escl, xmu, xsc
      data zero, one /0.d0, 1.d0/
c
      pwr = one /float (k-1)
      s1  = zero
      t   = one
      do 10 i=1, k
         if (i .eq. k) t = dble (n -k +1)
         a = (x (i) -xmu) /xsc
         s1 = s1 +t *dexp (a *z **pwr)
10    continue
      xntgnd = (s1 /dble (k)) **(-k) **pwr
      xntgnd = xntgnd *exp (sumx *(z **pwr -one))
      xntgnd = xntgnd *cnorm
      return
      end
```

```
      double precision function xconf (ta)
c
c         Subroutine to determine the probability of
c       the pth quantile being less than T(t), where T is the
c       tolerance limit factor.
c
      implicit double precision (a-h, o-z)
      common /cb/ p, gam, wp, t
      common /cd/ tol
      external yntgnd
      data zero, one /0.d0, 1.d0/
c
c       -- 't' is to be found so that the integral of YNTGND from
c          (0, infinity) equals 'gam'.
      t      = ta
      abserr = tol
      relerr = zero
      call dqdagi (yntgnd, zero, 1, aoserr, relerr, xi, err)
      xconf = xi -gam
c
c       -- This is just some terminal output to amuse the user while
c          he's waiting for results.
      write (*,*) ' Tolerance factor: confidence = ', t, xi
      return
      end
```

```
      double precision function yntgnd (z)
c
c        Function to calculate the integral for
c     determining the confidence.
c
      implicit double precision (a-h, o-z)
      common /cb/  p, gam, wp, t
      common /cd/  cnorm, sum, n, k
      common /cc/  s
      data zero, one /0.d0, 1.d0/
c
      pwr    = one /float (k-1)
c
c       -- The first factor is the function which was integrated
c          to get 'cnorm'.
      yntgnd = xntgnd (z)
c
c       -- The second factor is the gamma cdf evaluated at a poin
c          which depends on s, a quantity calculated in XNTGND.
      uplim  = s *dexp (wp *t *z **pwr)
      q      = one -dgamdf (float (k), uplim)
      yntgnd = yntgnd *q
      return
      end
```

A16

```
      double precision function aconf1 (t)
c
      implicit double precision (a-h, o-z)
c
      dimension x(1)
      common /dat/   x
      common /conf/  p, gam, nsamp, nobs
c
      call aconf (x, nsamp, nobs, t, p, prob)
      aconf1 = prob -gam
c
      return
      end
```

A17

```
      subroutine aconf (x, nsamp, nobs, q, p, conf)
c
c      Mark Vangel, July 1988
c
c      An approximation to the Lawless conditional procedure (or,
c   alternatively, to the posterior of a quantile under a flat prior).
c   The routine is written for two parameter Weibull analysis, but
c   extension to the generalized log gamma family is straightforward.
c
c   Ref : Piciccio, T.J., Technometrics 1987 p.??
c
c      x      --  Data                                          (Input)
c      nsamp  --  Total sample size                             (Input)
c      nobs   --  Number of (uncensored) observations           (Input)
c      q      --  Value for which P (xp =< q) is desired        (Input)
c      p      --  Probability associated with quantile xp       (Input)
c      conf   --  P (xp =< q)                                   (Output)
c
      implicit double precision (a-h, o-z)
      dimension x (1), s (4)
      common /par/ ushpw, uscle, xmuev, xsdev
      data eps, maxit /1.d-5, 100/
      data zero, one, half, thalf
     &    /0.d0, 1.d0, .5d0, 1.5d0/
c
c      -- Get the constrained MLE's, where q is constrained to
c         be the pth quantile.
      call cwnrml (p, q, cshp, cscl, nsamp, nobs, x, eps, iter, maxit)
c
c      -- Transform to extreme value distribution.
      uloc = log (uscle)
      cloc = log (cscl)
      uscl = one /ushpw
      cscl = one /cshp
c
c      -- Calculate the signed square root of -2 times the log
c         of the likelihood ratio
      wp  = log ( -log (one -p))
      cl  = zero
      ul  = zero
      do 20 i=1, nobs
         t   = log (x (i))
         cl  = cl +(t -cloc)/cscl -exp ((t -cloc)/cscl)
         ul  = ul +(t -uloc)/uscl -exp ((t -uloc)/uscl)
20    continue
c
      cl = cl -(nsamp -nobs) *exp ((t -cloc)/cscl) -nobs *log (cscl)
      ul = ul -(nsamp -nobs) *exp ((t -uloc)/uscl) -nobs *log (uscl)
c
      xlr = -2 *(cl -ul)
      sgn = log(q) -uloc -wp *uscl
      sgn = abs (sgn) /sgn
      r   = sgn *sqrt (abs (xlr))
c
c      -- Now calculate the derivatives of the log likelihood at
c         the MLE.
      do 30 j=1, 4
         s(j) = zero
         do 40 i=1, nobs
            z    = (log(x(i)) -uloc) /uscl
            t    = ((z -wp) **j) *exp (z)
            s(j) = s(j) +t
```

```
40        continue
          s(j) = s(j) +(nsamp -nobs) *t
30    continue
c
      d20 = -nobs
      d11 = -s(1)
      d02 = -(nobs +s(2))
      d30 = nobs
      d21 = 2*nobs +s(1)
      d12 = 3*s(1) +s(2)
      d03 = nobs +3*s(2) +s(3)
      d40 = -nobs
      d31 = -(3*nobs +s(1))
      d22 = -(4*nobs +5*s(1) +s(2))
      d13 = -(7*s(1) +4*s(2) +s(3))
      d04 = -nobs -(7*s(2) +4*s(3) +s(4))
c
c      -- The approximate mean and variance of r is calculated in
c         terms of the dij.
      v11 = -d02 /(d20 *d02 -d11 **11)
c
      a   = (v11 **half *(-d12/d02 +d11*d03 /d02**2)) /2
      b   = (v11       *(-d22/d02 +(2*d11**2 +d02*d20) *d12/d02**2
     $     -(4*d12*d11*d03 +d04*d11**2) /d02**3
     $     +2*(d11*d03 /d02**2) **2)) /4
      c   = (v11 **three*half *(d03 -3*d21*d11 /d02 +3*d11 *(-11 /d02 *d
     $     -d03 *(d11/d02 **2)) /4
      d   = (v11 **2 *(d40 -(4*d31*d11 +3*d21**2) /d02
     $     +(6*d22*d11**2 +12*d21*d12*d11) /d02**2
     $     -(6*d11*d03*d11**2 +4*d12*d03*d11**2 +12*(d12**2)*d11) /d02**3
     $     +(12*d12*d02*d11**2 +d04*d11**2) /d02**4
     $     -2*d11**4 +d12**2 /d02**3)) /24
c
      amu = a +2*c
      sig = sqrt (one +2*(b +3*d +amu) +11*c**2)
c
c      -- (r -amu)/sig is approximately distributed N(0, 1).
      r     = (r -amu)/sig
      conf = dnordf (r)
c
      return
      end
```

```
      SUBROUTINE WNRMLE  (A, B, NSAMP, NOBS, X, AEPS, ITER, MAXIT)
C
C     MARK VANGEL, MAY 1986
C
C     ESTIMATE WEIBULL PARAMETERS BY MAXIMUM LIKELIHOOD
C           A  --  WEIBULL SHAPE PARAMETER (RETURNED)
C           B  --  WEIBULL SCALE PARAMETER (RETURNED)
C       NSAMP  --  SAMPLE SIZE
C       NOBS   --  NUMBER OF OBSERVATIONS (UNCENSORED)
C                  (NSAMP-NOBS) VALUES ARE TYPE II CENSORED
C           X  --  DATA VALUES
C       AEPS   --  ABSOLUTE CONVERGENCE TOLERANCE (INPUT)
C       ITER   --  NUMBER OF ITERATIONS REQUIRED  (RETURNED)
C       MAXIT  --  MAXIMUM NUMBER OF ITERATIONS   (INPUT)
C
      IMPLICIT DOUBLE PRECISION (A-H, O-Z)
      DIMENSION X (1), CVTBL (25), SHPTBL (25)
C
C     -- COEFFICIENTS OF VARIATION FOR MOMENT ESTIMATE OF
C        SHAPE PARAMETER
      DATA NTBL   /25/
      DATA CVTBL  /1.     , .52272, .36345, .28054, .22905,
     $             .19377, .16802, .14837, .13286, .12031,
     $             .10944, .10022, .09239, .08737, .08174,
     $             .07667, .07251, .06862, .06513, .06197,
     $             .04177, .03150, .02529, .01694, .01277/
      DATA SHPTBL /1.,2.,3.,4.,5.,6.,7.,8.,9.,10.,11.,12.,13.,
     $             14.,15.,16.,17.,18.,19.,20.,30.,40.,50.,75.,100./
      DATA ZERO, ONE, TOLMIN /0.D0, 1.D0, 1.D-5/
C
C     -- SCALE THE DATA
      SCALE = X (1)
      DO 10 I=1, NOBS
         IF (X (I) .GT. SCALE) SCALE = X (I)
10    CONTINUE
      DO 20 I=1, NOBS
         X (I)    = X (I) / SCALE
20    CONTINUE
C
      XSAMP  = DBLE (NOBS)
      RXSAMP = ONE / XSAMP
      EPS    = AEPS
      IF (EPS .EQ. ZERO) EPS = TOLMIN
C
C     -- CALCULATE SUM, SUM OF LOGS, AND SUM OF SQUARES OF DATA
      S1 = ZERO
      S  = ZERO
      S5 = ZERO
      DO 30 I=1, NOBS
         S1 = S1 + LOG (X (I))
         S  = S  + X (I)
         S5 = S5 + X (I) ** (I)
30    CONTINUE
      S1 = S1 / XSAMP
C
C      -- USE THE COEFFICIENT OF VARIATION TO GET METHOD OF
C         MOMENTS ESTIMATE OF THE SHAPE PARAMETER
      CV = SQRT (NOBS *S5 -S **S) /
      IF (CV .GE. CVTBL (1)) THE
         A = SHPTBL (1)
      ELSE IF (CV .LE. CVTBL (NTBL)) THE
         A = SHPTBL (NTBL)
```

```
      ELSE
         DO 40 I=2, NTBL
            IF (CVTBL (I) .LE. CV) THEN
              A = SHPTBL (I-1) + (SHPTBL (I) -SHPTBL (I-1))*
     $            (CV -CVTBL (I)) /(CVTBL (I-1) -CVTBL (I))
              GO TO 50
            END IF
40       CONTINUE
50       CONTINUE
      END IF
C
C     -- LOOP UNTILL CONVERGENCE OR ITERATION LIMIT EXCEEDED
      ITER = 0
1     CONTINUE
C
C     -- CALCULATE SUMS FIRST
         S2 = ZERO
         S3 = ZERO
         S4 = ZERO
         DO 60 I=1, NOBS
            T1 = X (I) ** A
            T2 = T1 * LOG (X (I))
            T3 = T2 * LOG (X (I))
            S2 = S2 + T1
            S3 = S3 + T2
            S4 = S4 + T3
60       CONTINUE
         S2 = S2 + T1 *(NSAMP -NOBS)
         S3 = S3 + T2 *(NSAMP -NOBS)
         S4 = S4 + T3 *(NSAMP -NOBS)
C
C     -- CALCULATE NEW SHAPE ESTIMATE
      T1 = LOG / A
      T2 = S3 / S2
      T3 = S4 / S2
      A1 = A + (T1 + T1 - T2) / (T1*T1 + T1 - T3*T3)
C
C     -- QUIT WHEN CONVERGED
      IF (ABS (A1 - A) .LE. EPS .OR. ITER .GT.  MAXIT) G. TO 2.
         ITER = ITER + 1
         A    = A1
      GO TO 1
C
C     -- CALCULATE SCALE PARAMETER FROM KNOWN SHAPE PARAMETER
90    CONTINUE
      BSUM = ZERO
      DO 70 I = 1, NOBS
         BSUM = BSUM +X (I) ** A
70    CONTINUE
      PSUM = BSUM + FLOAT(NSAMP -NOBS) *X (.0 E) **A
      B = (NSAMP* BSUM) ** (ONE /A)
C
C     -- UNSCALE DATA
      B = B * SCALE
      DO 80 I=1, NOBS
         X (I) = X (I) * SCALE
80    CONTINUE
      RETURN
      END
```

A21

```
      SUBROUTINE CWNRML  (P,XP, A,B, NSAMP,NOPS, X, AEPS, ITER,MAXIT
C
C
C      MARK VANGEL, JULY 1988
C
C     ESTIMATE WEIBULL PARAMETERS BY MAXIMUM LIKELIHOOD WITH THE
C     CONSTRAINT THAT THE ESTIMATED PTH QUANTILE AT THE MLE
C     IS EQUAL TO XP.
C
C           P  --  PROBABILITY LEVEL OF QUANTILE
C          XP  --  PTH QUANTILE
C           A  --  WEIBULL SHAPE PARAMETER
C           B  --  WEIBULL SCALE PARAMETER
C       NSAMP  --  SAMPLE SIZE
C        NOPS  --  NUMBER OF OBSERVATIONS (UNCENSORED)
C                  (NSAMP-NOPS) VALUES ARE TYPE II CENSORED
C           X  --  DATA VALUES
C        AEPS  --  ABSOLUTE CONVERGENCE TOLERANCE (INPUT)
C        ITER  --  NUMBER OF ITERATIONS REQUIRED  (RETURNED)
C       MAXIT  --  MAXIMUM NUMBER OF ITERATIONS   (INPUT)
C
C
      IMPLICIT DOUBLE PRECISION (A-H, O-Z)
      DIMENSION X (1), CVTBL (25), SHPTBL (25)
C
C      -- COEFFICIENTS OF VARIATION FOR MOMENT ESTIMATE OF
C         SHAPE PARAMETER
      DATA NTBL   /25/
      DATA CVTBL  /1.     , .52772, .36345, .28054, .22906,
     $             .19377, .16902, .14937, .13360, .12071,
     $             .10944, .10122, .09377, .08737, .08197,
     $             .07727, .07251, .06902, .06517, .06174,
     $             .04177, .03150, .02529, .01464, .01277/
      DATA SHPTBL /1.,2.,3.,4.,5.,6.,7.,8.,9.,10.,11.,12.,13.,
     $             14.,15.,16.,17.,18.,19.,20.,30.,40.,50.,75.,100./
      DATA ZERO, ONE, TOLMIN /0.0D0, 1.0D0, 1.0D-5/
C
C     -- SCALE THE DATA
      SCALE = X(1)
      DO 10 I=1, NOPS
         IF (X (I) .GT. SCALE) SCALE = X (I)
10    CONTINUE
      DO 20 I=1, NOPS
         X (I)    = X (I) /SCALE
20    CONTINUE
C
      EPS    = AEPS
      XP     = AXP
      XP     = XP /SCALE
C     C      = -LOG (ONE -P)
      IF (EPS .EQ. ZERO) EPS = TOLMIN
C
C     -- CALCULATE SUM, SUM OF LOGS, AND SUM OF SQUARES OF DATA
      S1 = ZERO
      S  = ZERO
      SS = ZERO
      DO 30 I=1, NOPS
         S1 = S1 + LOG (X (I) /XP)
         S  = S  + X (I)
         SS = SS + X (I) **Y (I)
30    CONTINUE
C
```

```
C       -- USE THE COEFFICIENT OF VARIATION TO GET METHOD OF
C          MOMENTS ESTIMATE OF THE SHAPE PARAMETER
        CV = SQRT (NOBS *SS -S *S) /S
        IF (CV .GE. CVTBL (1)) THEN
           A = SHPTBL (1)
        ELSE IF (CV .LE. CVTBL (NTBL)) THEN
           A = SHPTBL (NTBL)
        ELSE
           DO 40 I=2, NTBL
              IF (CVTBL (I) .LE. CV) THEN
                 A = SHPTBL (I-1) + (SHPTBL (I) -SHPTBL (I-1))*
     $               (CV -CVTBL (I)) /(CVTBL (I-1) -CVTBL (I))
                 GO TO 50
              END IF
40         CONTINUE
50         CONTINUE
        END IF
C
C       -- LOOP UNTILL CONVERGENCE OR ITERATION LIMIT EXCEEDED
        ITER = 0
1       CONTINUE
C
C       -- CALCULATE SUMS FIRST
           S2 = ZERO
           S3 = ZERO
           S4 = ZERO
           DO 60 I=1, NOBS
              T  = C  *(X (I) /XP)** A
              T2 = T  * LOG (X (I) /XP)
              T3 = T2 * LOG (X (I) /XP)
              S2 = S2 + T2
              S3 = S3 + T3
60         CONTINUE
           S2 = S2 + T2 *(NSAMP -NOBS)
           S3 = S3 + T3 *(NSAMP -NOBS)
C
C       -- CALCULATE NEW SHAPE ESTIMATE
        G1 = NOBS / A
        G2 = NOBS / A**2
        A1 = A + (G1 + S1 - S2) / (G2 + S3)
C
C       -- QUIT WHEN CONVERGED
        IF (ABS (A1 - A) .LE. ERR .OR. ITER .GT. MAXIT) GO TO 9
           ITER = ITER + 1
           A    = A1
        GO TO 1
C
C       -- CALCULATE SCALE PARAMETER FROM KNOWN SHAPE PARAMETER
90      CONTINUE
        C  = XP /C **(ONE /A) *SCALE
C
C       -- UNSCALE DATA
        DO 80 I=1, NOBS
           X (I) = X (I) * SCALE
80      CONTINUE
        RETURN
        END
```

MAKING FISHER'S EXACT TEST RELEVANT

Paul H. Thrasher
Engineering Branch
Reliability, Availability, and Maintainability Division
Army Materiel Test and Evaluation Directorate
White Sands Missile Range, New Mexico 88002-5175

ABSTRACT

The Fisher-Irwin Exact Method is made relevant by including q-values in the analysis. Q-values are post-test Type II risks. They provide information which complements the Type I risk provided by the p-value. Reporting both the p-value and relevant q-values enables managers to base decisions on both types of risks. For references on q-values, see four papers by Thrasher in the Proceedings of the Thirtieth through Thirty-Third Conferences on the Design of Experiments in the Army Research, Development, and Testing, U.S. Army Research Office, 4300 South Miami Boulevard, Research Triangle Park, North Carolina 27709-2211.

Q-values are normally calculated using the same algorithm used to find pre-test Type II errors. The q-value calculation inputs are normally (1) the p-value instead of the pre-test Type I risk, (2) the sample size actually used instead of the sample size planned, and (3) the same relevant values of the parameter considered in the pre-test Type II risk calculation. Since the Fisher-Irwin Exact Method doesn't historically have a design stage, there is no pre-test algorithm available for modification. This paper develops the necessary algorithm by extending the p-value calculation based on the binomial rather than the hypergeometric distribution.

The q-value equations are developed and their mathematical properties are examined. Computer programming methods are discussed. Examples are provided for sample sizes both (1) small enough that a hand-held calculator can be used and (2) large enough to require a digital computer. Numerical results are interpreted from the viewpoint of a manager that must balance non-zero Type I and Type II risks.

INTRODUCTION AND OBJECTIVE

The Fisher-Irwin Exact Method is a quick and straightforward technique of comparing two samples of dichotomous items. The normally reported statistic from this test is the p-value. The p-value is the probability of being wrong in marginally rejecting a null hypothesis that the two samples are from one population. In practice, managers conclude that the two samples are from different populations if they believe the p-value is sufficiently low.

This method of analysis has not gained universal acceptance. The reluctance to use this method may well be due to unbalanced reporting of information. The p-value is used to report the Type-I error. This error is sometimes called the producer's error, the contractor's error, or the error of concern for the advocates of maintaining the status quo. Traditionally the

method has not reported information about the Type-II error. This error may be called consumer's error, the government's error, or the error of concern for the advocates of change.

The Fisher-Irwin Exact Method can provide relevant information about the Type-II error. This additional information results from calculating and reporting q-values. Q-values are the probabilities of being wrong in marginally failing to reject the null hypothesis when the two samples are from different populations. Since the two populations may differ in different ways, there is a q-value for each pair of separate populations. Managers can use a q-value, for a relevant pair of unequal populations, as evidence for concluding that the two samples are from those different populations. They reach this conclusion if they believe a relevant q-value is sufficiently high.

This paper provides equations for and examples of calculating (1) the p-value and (2) q-values for the Fisher-Irwin Exact Method using a one-sided analysis. This one-tailed analysis is used to reject a single population in favor of two populations that differ in the direction indicated by the data.

This paper also discusses a digital computer program. This program has been written to (1) handle the necessary voluminous calculations for large sample sizes, (2) retain the analyst's identification of the two measurement samples and the two mutually exclusive and exhaustive categories, and (3) provide an report from which a manager can decide if future actions should be based on one or two populations.

The Fisher-Irwin Exact Method may be implemented in different ways. At the cost of redundancy, this paper uses more than one approach to illustrate different viewpoints.

P-VALUE CALCULATION

The data for the Fisher-Irwin Exact Method, often called Fisher's Exact Test, consists of four numbers. They and their sums are normally arranged in a square array. The following array has double entries to illustrate both the general situation and a specific example:

	Category 1 = Success:	Category 2 = Failure:	Sum:
Sample I = Development:	r=19	n-r=2	n=21
Sample II = Production:	R-r=12	(N-n)-(R-r)=3	N-n=15
Total:	R=31	N-R=5	N=36

Since the choices of Samples I and II and of Categories 1 and 2 are both arbitrary, there are four possible ways the data can be arranged. The ambiguity has been removed from the above table by naming the samples and categories to make (1) $n \geq (N-n)$ and (2) $r/n \geq (R-r)/(N-n)$.

There are two methods of calculating the p-value. The best known uses the hypergeometric distribution. The second uses the binomial distribution. Both are described and illustrated in pages 195-203 of Bradley, James, V., Distribution-Free Statistical Tests, Prentice-Hall, Inc., 1968.

Hypergeometric Approach:

The hypergeometric approach is based on a population of N items which is split into two samples of sizes n and N-n. The null hypothesis is that the difference between the R and N-R items of the two categories did not influence the sample selection. The probability of obtaining the data is the ratio of (1) the number of ways items from one category can be chosen for the two samples to (2) the number of ways items of this category can be chosen from the total population. Thus

$$P[\text{Obtaining the Data}] = \frac{{}_{n}C_{r}\ {}_{N-n}C_{R-r}}{{}_{N}C_{R}}$$

where ${}_{i}C_{j}$ is the number of ways of choosing j items from i items. Both ${}_{i}C_{j}$ and ${}_{i}C_{i-j}$ are found from the following relation of factorials:

$${}_{i}C_{j} = {}_{i}C_{i-j} = \frac{i!}{j!\ (i-j)!}$$

The p-value is the probability of obtaining the data or more extreme partitions of the N-R items of Category 2. (More explicitly, the p-value for the one-sided or one-tailed test is the probability that the partition of the N-R items will be as unbalanced as the data in the direction that the data suggests.) For the specific example, the p-value is

$$\text{p-value} = \frac{{}_{21}C_{2}\ {}_{15}C_{3}}{{}_{36}C_{5}} + \frac{{}_{21}C_{1}\ {}_{15}C_{4}}{{}_{36}C_{5}} + \frac{{}_{21}C_{0}\ {}_{15}C_{5}}{{}_{36}C_{5}}$$

$$= \frac{\frac{21!}{2!\ 19!}\ \frac{15!}{3!\ 12!}}{\frac{36!}{5!\ 31!}} + \frac{\frac{21!}{1!\ 20!}\ \frac{15!}{4!\ 11!}}{\frac{36!}{5!\ 31!}} + \frac{\frac{21!}{0!\ 21!}\ \frac{15!}{5!\ 10!}}{\frac{36!}{5!\ 31!}}$$

$$= .253 + .076 + .008 = .34$$

A formal expression for the hypergeometric approach may be written in two ways. Considering the possible distribution of the N-R items of Category 2 yields one of the two following equations:

if (1) n-r is less than expected because $n-r < n[(N-R)/N]$ and (2) m_{I2} is the minimum possible number of items in Sample I from Category 2,

$$\text{p-value} = \sum_{i=m_{I2}}^{n-r} {}_nC_i \; {}_{N-n}C_{N-R-i} \; / \; {}_NC_{N-R};$$

if (1) n-r is more than expected because $n-r > n[(N-R)/N]$ and (2) M_{I2} is the maximum possible number of items in Sample I from Category 2,

$$\text{p-value} = \sum_{i=n-r}^{M_{I2}} {}_nC_i \; {}_{N-n}C_{N-R-i} \; / \; {}_NC_{N-R}.$$

By considering the distribution of the R items of Category 1 instead of the N-R items of Category 2, this formal expression is written using two other equations:

if (1) r is less than expected because $r < n[R/N]$ and (2) m_{I1} is the minimum possible number of items in Sample I from Category 1,

$$\text{p-value} = \sum_{i=m_{I1}}^{r} {}_nC_i \; {}_{N-n}C_{R-i} \; / \; {}_NC_R;$$

if (1) r is more than expected because $r > n[R/N]$ and (2) M_{I1} is the maximum possible number of items in Sample I from Category 1,

$$\text{p-value} = \sum_{i=r}^{M_{I1}} {}_nC_i \; {}_{N-n}C_{R-i} \; / \; {}_NC_R.$$

Binomial Approach:

The binomial approach is based on an infinite population from which two independent samples are taken. Although the binomial parameter of the distribution of Category 1 items may be estimated by R/N, it is really unknown. Fortunately for the p-value calculation, this parameter will cancel from the equations regardless of its value. Denoting this binomial parameter by p allows the probabilities of obtaining the two samples to be written as

$$P[r;n,p \text{ in Sample I}] = {}_nC_r \; p^r \; (1-p)^{n-r} \text{ and}$$

$$P[R-r;N-n,p \text{ in Sample II}] = {}_{N-n}C_{R-r} \; p^{R-r} \; (1-p)^{(N-n)-(R-r)}.$$

With the restriction that these two samples are independent, the probability of obtaining both samples is the product of the two above equations; this reduces to

$$P[\text{Obtaining both Samples}] = {}_nC_r \; {}_{N-n}C_{R-r} \; p^R \; (1-p)^{N-R}.$$

The probability of obtaining one big sample size of N with R p-type observations is

$$P[R;N,p \text{ in Combined Sample}] = {}_NC_R \; p^R \; (1-p)^{N-R}.$$

Finally, the conditional probability of obtaining the two samples given that the combined sample has been obtained is

$$P[(r \text{ in } n) \text{ and } (R-r \text{ in } N-n) \mid (R \text{ in } N)] = \frac{P[\text{Both Samples}]}{P[\text{Combined Sample}]}.$$

This equation is expressed in terms of the data by division of the two previous equations. The result is

$$P[\text{Obtaining the Data}] = \frac{{}_nC_r \; {}_{N-n}C_{R-r} \; p^R \; (1-p)^{N-R}}{{}_NC_R \; p^R (1-p)^{N-R}} = \frac{{}_nC_r \; {}_{N-n}C_{R-r}}{{}_NC_R}.$$

This is the same equation as was obtained using the hypergeometric approach and the rest of the calculation of the p-value proceeds identically.

The binomial equations in the above paragraph may be illustrated and clarified by using (1) the data from this discussion's specific example and (2) the point estimate of p given by R/N=31/36=.861. The result is

$$P[19;21,.861 \text{ in Sample I}] = {}_{21}C_{19} \; (.861)^{19} \; (1-.861)^{21-19} = .236,$$

$$P[12;15,.861 \text{ in Sample II}] = {}_{15}C_{12} \; (.861)^{12} \; (1-.861)^{15-12} = .203,$$

$$P[\text{Obtaining both Samples}] = {}_{21}C_{19} \; {}_{15}C_{12} \; (.861)^{31} \; (1-.861)^{36-31}$$
$$= .0479 = (.236)(.203),$$

$$P[31;36,.861 \text{ in Combined Sample}] = {}_{36}C_{31}\,(.861)^{31}\,(1-.861)^{36-31}$$

$$= .189, \text{ and by dividing equations}$$

$$P[(19 \text{ in } 21) \text{ and } (12 \text{ in } 15) \mid (31 \text{ in } 36)] = .0479/.189 = .253.$$

The value of .253 is obviously the same intermediate result as was obtained in the hypergeometric approach. In fact, any value of p yields .253.

It is illustrative and useful to obtain P[31;36,.861 in Combined Sample] without the assumption that all 36 items were selected from one population. This is done by using the facts that (1) Sample I and Sample II were obtained independently and (2) the 31 items in Category 1 could have been distributed between the two samples in different ways. The calculation is summarized in the following table. The starred row of this table corresponds to the data and three intermediate results from in the preceeding paragraph.

R	r'	R-r'	P[r';21,.861]	P[R-r';15,.861]	P[r';21,.861]P[R-r';15,.861]	
31	21	10	.0433	.0348	.00151	
31	20	11	.147	.0981	.0144	
31	19	12	.236	.203	.0479	*
31	18	13	.242	.290	.0700	
31	17	14	.175	.257	.0450	
31	16	15	.0961	.106	.0102	
					.189	

The last column contains the probabilities of the different ways that the 31 items can be distributed. The sum of this column is the probability of having 31 items from Category 1 in the two samples. The value of .189 obviously agrees with the shorter calculation in the preceeding paragraph. A modified version of the longer calculation of this paragraph will be needed in the calculation of q-values.

Before calculating q-values, it is illustrative and useful to obtain the p-value from the table of the preceeding paragraph. Note first that the .0479 in the last column of the starred row corresponding to the data agrees with P[Obtaining both Samples] from the short binomial calculation of two paragraphs ago. Note second that the entries above .0479 in the table correspond to probabilities of obtaining more unlikely partitions than the data. Using these two facts yields

$$\text{p-value} = P[\text{Rejecting} \mid \text{Rejection Should Not Occur}]$$

$$= \frac{.00151}{.189} + \frac{.0144}{.189} + \frac{.0479}{.189}$$

$$= .34.$$

This result of .34 does not depend on the number used for p. This may be seen by (1) calculating another table using any p other than 31/36=.861 and (2) summing the probabilities of obtaining partitions as extreme as the data.

This last method of calculating the p-value emphasizes that the data are viewed marginally. The data are viewed as unbalanced just enough to warrant rejection of the single-population hypothesis.

Q-VALUE CALCULATION

Q-values, like the p-value, consider the data to be just sufficiently unbalanced to warrent rejection of the single-population hypothesis. While the p-value is the probability of getting results at least as unbalanced as the data, q-values are the probabilities of more balanced results.

For Fisher's Exact Test, q-values cannot be calculated by using the hypergeometric approach. All q-values are conditional probabilities with the condition being that two different populations provided the two samples. Thus q-values must be calculated by using the binomial approach with different binomial parameters, p_I and p_{II}, for the populations of the two samples. If desired, these two parameters may be replaced with p_I and k where $k=p_I/p_{II}$. Most q-value calculations are functions of only one parameter and lend themselves to a two dimensional power curve representation. For the Fisher-Irwin Exact Method, there are two parameters so the representation must take the form of a three dimensional power surface. Any specific point on this surface does not exhibit as much information as the entire surface. To be specific in the following calculation however, p_I and p_{II} will be taken as the point estimates from the data. That is, the following q-value calculation will address the error of concluding that $p=p_I=p_{II}$ when actually $p_I=r/n=19/21=.905$, $p_{II}=(R-r)/(N-n)=12/15=.800$, and $k=p_I/p_{II}=.905/.800=1.13$. This addresses the intuitive concern of "making the mistake of ignoring what the data's trying to tell us."

Since a q-value for any specific p_I and p_{II} is the probability of falsely retaining the assumption that one p describes all items, a q-value is one minus the probability of rejecting the assumption of a single population when there are two populations described by p_I and p_{II}. This can be calculated from entries in a table of probabilities for all possible values of r consistent with N, n, and R.

The remainder of this section considers two approaches to the q-value equation, shows that the two results are equivalent, and discusses some general mathematical properties.

Category 1 Approach:

For the specific example in this discussion, one table used to calculate a q-value for p_I=19/21 and p_{II}=12/15 is

R	r'	R-r'	P[r';21,.905]	P[R-r';15,.800]	P[r';21,.905]P[R-r';15,.800]
31	21	10	.122	.103	.0126
31	20	11	.270	.188	.0507
31	19	12	.284	.250	.0711
31	18	13	.190	.231	.0438
31	17	14	.0898	.132	.0119
31	16	15	.0321	.0352	.00113
					.191

The sum of .191 on the lower right represents the probability of obtaining a total of 31 items from Category 1 from the two samples. This probability can be divided into entries in the right hand column to find the conditional probabilities of obtaining possible numbers of items from Category 1 in Samples I and II. Taking the data and more extreme divisions of the R items from category 1 as evidence of rejection, the q-value is found from

$$1 - \text{q-value} = P[\text{Rejecting} \mid \text{Rejection Should Occur}]$$

$$= \frac{.0711}{.191} + \frac{.0507}{.191} + \frac{.0126}{.191}$$

$$= .70.$$

Using a more conventional approach, the q-value can be found from less extreme divisions of the R items from Category 1 to be

$$\text{q-value} = P[\text{Failing to Reject} \mid \text{Rejection Should Occur}]$$

$$= \frac{.0438}{.191} + \frac{.0119}{.191} + \frac{.00113}{.191}$$

$$= .30.$$

This procedure may be stated formally with the following equation:

$$\text{q-value} = \frac{\sum_{r'=\min(r')}^{r-1} P[r';n,p_I]P[R-r';N-n,p_{II}]}{\sum_{r'=\min(r')}^{\max(r')} P[r';n,p_I]P[R-r';N-n,p_{II}]}$$

where min(r') and max(r') are the minimum and maximum values of r' allowed by the constraints imposed by fixed values of N, n, and R. Increasing of r' is limited by the total size of both Sample I and Category 1. That is, r' must simultaneously satisfy $r' \leq n$ and $r' \leq R$. Thus the upper limit on the above sum is

$$\max(r') = \min(n,R).$$

Decreasing of r' is limited by the requirement that two measurements must be non-negative. Possible measurements of Category 1 items in Sample I and Category 2 items in Sample II lead to $r' \geq 0$ and $(N-n)-(R-r') \geq 0 \Rightarrow r' \geq n+R-N$. Considering the other two possible measurements leads to four other conditions: $R-r' \leq R$, $R-r' \leq N-n$, $n-r' \leq n$, and $n-r' \leq N-R$. These four conditions are equivalent to the first two. Thus the lower limit of the above sum is

$$\min(r') = \max(0,n+R-N).$$

The final form of the Category 1 equation is thus

$$\text{q-value} = \frac{\sum_{r'=\max(0,n+R-N)}^{r-1} P[r';n,p_I]P[R-r';N-n,p_{II}]}{\sum_{r'=\max(0,n+R-N)}^{\min(n,R)} P[r';n,p_I]P[R-r';N-n,p_{II}]}.$$

Category 2 Approach:

Using Category 2 instead of Category 1 leads to the following table:

N-R	n-r'	(N-n)-(R-r')	P[n-r';21;2/21]	P[(N-n)-(R-r');15,3/15]	P[Both]
5	0	5	.122	.103	.0126
5	1	4	.270	.188	.0507
5	2	3	.284	.250	.0711
5	3	2	.190	.231	.0438
5	4	1	.0898	.132	.0119
5	5	0	.0321	.0352	.00113
					.191

The numbers in the right three columns are the same as in the previous table used in the Category 1 approach. The calculation proceeds as before but the indices in the summation appear differently:

$$1 - \text{q-value} = \frac{\sum_{n-r'=0}^{2} P[n-r';21,.0952]P[(N-n)-(R-r');15,.200]}{\sum_{n-r'=0}^{5} P[n-r';21,.0952]P[(N-n)-(R-r');15,.200]}$$

$$= \frac{.0126}{.191} + \frac{.0507}{.191} + \frac{.0711}{.191} = .70 \text{ or}$$

$$\text{q-value} = \frac{\sum_{n-r'=3}^{5} P[n-r';21,.0952]P[(N-n)-(R-r');15,.200]}{\sum_{n-r'=0}^{5} P[n-r';21,.0952]P[(N-n)-(R-r');15,.200]}$$

$$= \frac{.0438}{.191} + \frac{.0119}{.191} + \frac{.00113}{.191} = .30.$$

This procedure may be stated formally with the following equation:

$$\text{q-value} = \frac{\sum_{n-r'=n-r+1}^{\max(n-r')} P[n-r';n,1-p_I]P[(N-n)-(R-r');N-n,1-p_{II}]}{\sum_{\min(n-r')}^{\max(n-r')} P[n-r';n,1-p_I]P[(N-n)-(R-r');N-n,1-p_{II}]}$$

where $\min(n-r')$ and $\max(n-r')$ are the minimum and maximum possible measurements of Category 2 items in Sample I. By arguments similar to those used in the Category 1 approach, $n-r' \leq n$ and $n-r' \leq N-R$ imply that

$$\max(n-r') = \min(n, N-R)$$

and $n-r \geq 0$ and $R-r' \geq 0 \Rightarrow n-r' \geq n-R$ imply that

$$\min(n-r') = \max(0, n-R).$$

The use and interpretation of this result must of course be done in conjunction with $n \geq (N-n)$ and $r/n \geq (R-r)/(N-n)$. Interpretation must recognize that the arrangement of the data in a standard format may or may not select Category 1 as the category of primary interest and/or Sample I as the first sample drawn and/or tested.

Equivalency of Methods:

Although the equations from the Category 1 and Category 2 approaches appear quite different, they are equivalent. By using the binomial probability relation $P[i;m,p] = P[m-i;m,1-p]$, the Category 2 equation may be rewritten as

$$\text{q-value} = \frac{\sum_{n-r'=n-r+1}^{\min(n,N-R)} P[r';n,p_I]P[R-r';N-n,p_{II}]}{\sum_{n-r'=\max(0,n-R)}^{\min(n,N-R)} P[r';n,p_I]P[R-r';N-n,p_{II}]}.$$

The limits on the summations may be rewritten by using $\min(m,M) = -\max(-m,-M)$, $k+\max(m,M) = \max(k+m,k+M)$, and $k+\min(m,M) = \min(k+m,k+M)$ to yield

$$\text{q-value} = \frac{\sum_{r'=r-1}^{\max(0,n+R-N)} P[r';n,p_I]P[R-r';N-n,p_{II}]}{\sum_{r'=\min(n,R)}^{\max(0,n+R-N)} P[r';n,p_I]P[R-r';N-n,p_{II}]}.$$

By reversing the summation limits to correspond to the normal practice of summing from low to high indices, this equation becomes the Category 1 result.

Range, Sum with P-Value, and Symmetry:

The q-value, like any other probability, is bounded by zero and one. This is verified for the Category 1 equation by splitting the sum in the denominator into two sums which first sum from $r'=\min(r')=\max(0,n+R-N)$ to $r'=r-1$ and then summing from $r'=r$ to $r'=\max(r')=\min(n,R)$. Dividing both numerator and denominator by the sum in the numerator then yields

$$\text{q-value} = \frac{1}{1 + \dfrac{\sum_{r'=r}^{\min(n,R)} P[r';n,p_I]P[R-r';N-n,p_{II}]}{\sum_{r'=\max(0,n+R-N)}^{r-1} P[r',n,p_I]P[R-r',N-n,p_{II}]}}.$$

Since this equation's ratio of the two sums is never negative but it will be infinite if the data yields $r=\max(0,n+R-N)$, q-value > 0. Since this ratio can be 0 by choosing p_I and p_{II} equal to 0 or 1, q-value < 1.

When p_I equals p_{II}, the q-value is one minus the p-value. This occurs because the possible values of r' are divided into two mutually exclusive and exhaustive sets. One set contains possible measurements as unlikely or more unlikely than r. The other contains values of r' more likely than r. These two sets identify conditional probabilities that are summed to find the p-value and q-value. The p-value summation uses the unlikely set with both p_I and p_{II} equated to any common probability. The q-value summation uses the likely set with any values of p_I and p_{II}. The mutual exclusiveness and exhaustiveness of the two sets require that p-value + q-value = 1 when $p_I=p_{II}$.

For the Fisher-Irwin Exact Method, the q-value is symmetric about the off diagonal in a plot of p_I versus p_{II}. That is, symmetry is expressed by

$$\text{q-value}(N,n,R,r,p_I,p_{II}) = \text{q-value}(N,n,R,r,1-p_{II},1-p_I).$$

This may be seen by applying the binomial equation $P[i;m,p] = {}_mC_i\, p^i\, (1-p)^{m-i}$ to the Category 1 equation in a series of three equations:

$$\text{q-value} = \frac{\sum_{r'=\max(0,n+R-N)}^{r-1} P[r';n,p_I]P[R-r';N-n,p_{II}]}{\sum_{r'=\max(0,n+R-N)}^{\min(n,R)} P[r';n,p_I]P[R-r',N-n,p_{II}]}$$

$$= \frac{\sum_{r'=\max(0,n+R-N)}^{r-1} {}_nC_{r'}\ p_I^{r'}\ (1-p_I)^{n-r'}\ {}_{N-n}C_{R-r'}\ p_{II}^{R-r'}\ (1-p_{II})^{N-n-R+r'}}{\sum_{r'=\max(0,n+R-N)}^{\min(n,R)} {}_nC_{r'}\ p_I^{r'}\ (1-p_I)^{n-r'}\ {}_{N-n}C_{R-r'}\ p_{II}^{R-r'}\ (1-p_{II})^{N-n-R+r'}}$$

$$= \frac{(\text{No } r' \text{ Factor}) \sum_{r'=\max(0,n+R-N)}^{r-1} {}_nC_{r'}\ {}_{N-n}C_{R-r'}\ p_I^{r'}\ (1-p_I)^{-r'}\ p_{II}^{-r'}\ (1-p_{II})^{r'}}{(\text{No } r' \text{ Factor}) \sum_{r'=\max(0,n+R-N)}^{r-1} {}_nC_{r'}\ {}_{N-n}C_{R-r'}\ p_I^{r'}\ (1-p_I)^{-r'}\ p_{II}^{-r'}\ (1-p_{II})^{r'}}$$

where the (No r' Factor) is $(1-p_I)^n\ p_{II}^R\ (1-p_{II})^{N-n-R}$. This constant has been factored from each term of the sum over r'. After canceling this factor from the equation, the symmetry is evident because substituting $1-p_{II}$ for p_I and $1-p_I$ for p_{II} yields the same equation.

The q-value for the Fisher-Irwin Exact Method also is symmetric in n and R. Applying ${}_iC_j = i!/j!(i-j)!$ to the above equation and canceling n! and (N-n)! yields

$$\text{q-value} = \frac{\sum_{r'=\max(0,n+R-N)}^{r-1} \frac{1}{r'!\ (n-r')!\ (R-r')!\ (N-n-R+r')!} \left[\frac{p_I\ (1-p_{II})}{(1-p_I)\ p_{II}}\right]^{r'}}{\sum_{r'=\max(0,n+R-N)}^{\min(n,R)} \frac{1}{r'!\ (n-r')!\ (R-r')!\ (N-n-R+r')!} \left[\frac{p_I\ (1-p_{II})}{(1-p_I)\ p_{II}}\right]^{r'}}$$

Substitution of n for R and R for n yields the same equation. Thus symmetry is expressed by

$$\text{q-value}(N,n,R,r,p_I,p_{II}) = \text{q-value}(N,R,n,r,p_I,p_{II}).$$

This equation reflects the mathematical arbitrariness in identifying samples and categories. The samples and categories normally are distinguished physically; but they are interchangeable mathematically.

RECAPITULATION AND INTERPRETATION

The p-value and a relevant q-value can provide influencing factors for management. If the p-value is lower than the risk allowed for the proponent of a single population, management is inclined toward the decision that two populations exist. If a relevant q-value is higher than the risk that the proponent of two populations is willing to take, management is also inclined toward the decision that two relevant populations exist. On the other hand, a high p-value or low relevant q-value inclines management toward the decision that there is one population.

Management will quite often be influenced by factors other than the p-value and a relevant q-value. A subjective decision-making process will naturally be used to consider all factors. The extremity of the lowness or highness of the p-value and a q-value provides the subjective weight for these two factors.

If management cannot determine threshold risks to indicate two populations when the p-value is below and a q-value is above these thresholds, an alternate approach is to compare the p-value and a q-value. Management can set a threshold ratio of Type II to Type I risks and compare a ratio of q-value/p-value to this threshold. Two populations are then indicated if a ratio of q-value/p-value is too high. In a subjective decision-making process considering many factors, the extremity of a q-value/p-value ratio provides the subjective weight of the Fisher's Exact Test factor.

Management should determine which two populations are relevant. Factors other than the data may suggest specific populations. Management should consider a q-value for each and every pair of relevant populations. If the analyst is not provided with the p_I and p_{II} for any relevantly different populations, the report to management should include a table of q-values for a wide range of p_I and p_{II}.

For the primary example in this discussion, .34 is the p-value and .30 is a q-value for the two populations suggested by the data. If these two populations with p_I=.9 and p_{II}=.8 are relevantly different, the two risks of .34 and .30 provide the basis for action. If the existence of these two populations is considered as positive, .34 is the probability of making a false positive decision. Similarly, considering the existence of only one population as negative implies that .30 is the probability of making a false negative decision.

If .34 and .30 are believed sufficiently low and high for probabilities of false positives and negatives respectively, future action is based on the existence of two populations with with p_I and p_{II} estimated by .9 and .8. If .34 and .30 are believed sufficiently high and low, future action is based on a single population.

For this example, .30/.34=1/1.1=.9 is a ratio of q-value/p-value. Subject to the relevancy of p_I=.9 and p_{II}=.8, 1/1.1=.9 is the ratio of risks of making false negative and false positive decisions. Future action is based on two populations if 1/1.1=.9 is believed sufficiently high. Similarly, future action is based a single population if 1/1.1=.9 is believed sufficiently low.

If the p-value and a q-value provide conflicting or indeterminate indications that are unresolvable, the immediate future action is to do additional testing. Additional testing should provide more definitive information by yielding either a low p-value and a high q-value or a high p-value and a low q-value. Naturally increasing the sample sizes may not yield a proportional increase in all the data; but if additional testing actually doubled all the data in this paper's example, the results would be .18 for the p-value, .35 for a q-value, and .35/.18=2 for a q-value/p-value ratio corresponding to p_I=.9 and p_{II}=.8. This possible decrease in the p-value, increase in a q-value, and increase in a ratio of q-value/p-value would increase the tendency to base future actions on two populations.

COMPUTING METHODS AND RESULTS

A digital computer program has been written in Pascal/3000 to facilitate the p-value and q-value analysis of the Fisher-Irwin Exact Method.

Two related manipulations are useful in extending the range of data which yields q-values without computer overflows or underflows. The equation for the q-value can be rewritten as

$$\text{q-value} = \frac{\sum_{r'=\max(0,n+R-N)}^{r-1} \frac{C}{r'!\,(n-r')!\,(R-r')!\,(N-n-R+r')!} \left[\frac{p_I\,(1-p_{II})}{(1-p_I)\,p_{II}}\right]^{r'}}{\sum_{r'=\max(0,n+R-N)}^{\min(n,R)} \frac{C}{r'!\,(n-r')!\,(R-r')!\,(N-n-R+r')!} \left[\frac{p_I\,(1-p_{II})}{(1-p_I)\,p_{II}}\right]^{r'}}$$

where C is any constant. The computer program can assign C with a value which hinders the summed terms from exceeding the computer's working range. To make this assignment without overflowing or underflowing the computer, each term must be considered as

$$\frac{C}{r'!\,(n-r')!\,(R-r')!\,(N-n-R+r')!} \cdot \left[\frac{p_I\,(1-p_{II})}{(1-p_I)\,p_{II}}\right]^{r'} = \exp[\ln(\text{TERMS})]$$

nere the expression ln(TERMS) in the exponential is

$$\ln(\text{TERMS}) = \ln[C] + r'[\ln(p_I) + \ln(1-p_{II}) - \ln(1-p_I) - \ln(p_{II})] - \ln[r'!] - \ln[(n-r')!] - \ln[(R-r')!] - \ln[(N-n-R+r')!].$$

The constant C can be selected to keep ln(TERMS) within the computer's range for x in exp(x). (e.g. -176 to 176). For any value of r', this selection can then be used to force C exp(ln(TERMS)) into the computer's operating range (e.g., $8.6(10)^{-78}$ to $1.15\ (10)^{77}$). Naturally this programming technique is successful only if r' doesn't change too much in the summation between max(0,n+R-N) and min(n,R).

The range of computer calculations for the p-value can be extended by using logarithms. One useful form of the p-value equation is

$$\text{p-value} = \sum_{i=v}^{w} \exp[\ln({}_nC_x) + \ln({}_{N-n}C_y) - \ln({}_NC_z)]$$

where the factors v, w, x, y, and z are dependent on r, R, and N according to the following table:

	If R<(N-R)		If R>(N-R)	
Factor	If R<n	If R>n	If (N-R)<(N-n)	If (N-R)>(N-n)
v	r	r	(N-n)-(k-r)	(N-n)-(R-r)
w	R	n	N-R	N-n
x	i	i	N-R-i	N-R-i
y	R-i	k-i	i	i
z	R	R	N-R	N-R

The computer program operates from a terminal. At the start of the program, the user selects either the terminal screen or a printer for the program output. The information in Figure 1 then appears on the screen. This provides the user with a brief summary of the analysis and asks the user which four independent variables will be entered. Figure 2 provides an example of the terminal screen after the output has been directed to a printer and the user has chosen to enter N, n, R, and r. Figures 3a and 3b contain the output for this input. Correspondingly, Figures 4 and 5 show input and output when the user has entered r, n-r, R-r, and (N-n)-(R-r). Finally Figure 6 shows an example with both input and output on the terminal screen. For this example, the input produces a standard table in a different order than the data.

Some results from the outputs in these figures (and similar computer executions) are compiled in the following table. All possible measurements of r for N=36, n=21, and R=31 are included. The tabulated q-values are referenced to θ and ϕ instead of p_I and p_{II}. This is necessary because the computer places the data in a standardly ordered table sometimes making $p_I=\theta$ and $p_{II}=\phi$ and sometimes resulting in $p_I=\phi$ and $p_{II}=\theta$.

		point-estimates to replace .861		q-value for				
r	p-value	θ	ϕ	point estimates	θ=.8 ϕ=.9	θ=.85 ϕ=.87	θ=.87 ϕ=.85	θ=.9 ϕ=.8
16	.054	.762	1.000	.000	.791	.925	.962	.993
17	.292	.810	.933	.249	.381	.644	.766	.924
18	.663	.857	.867	.306	.098	.274	.407	.682
19	.337	.905	.800	.297	.902	.726	.593	.318
20	.084	.952	.733	.241	.988	.941	.884	.690
21	.008	1.000	.667	.000	.999	.995	.987	.941

This table emphasizes that management needs to determine relevantly different populations instead of just considering the point-estimates suggested by the data. As expected, the extreme r measurements of 16 and 21 lead to low p-values indicating two populations. The two populations indicated however, are not those suggested by the point-estimates of θ and ϕ. (θ and ϕ are the binomial probabilities that describe the two populations; they replace the common-population point-estimate of 31/36=.861.) The q-values for those point-estimates are identically zero. Although every manager is free to determine how high a q-value needs to be for a two population decision, these are low by any standard. Management must realize that no amount of testing can prove that anything is either completely perfect or worthless. Instead, more reasonable values of θ and ϕ must be considered. If θ=.8 and ϕ=.9 are considered for r=16 (or θ=.9 and ϕ=.8 for r=21), the q-value of .791 (or .941) is quite high. Even higher q-values are obtained when θ=.85 and ϕ=.87 are considered for r=16 (or θ=.87 and ϕ=.85 for r=21). This reflects the fact that it's easier to say two things are different if they don't have to be very different. Considering possible populations for which θ and ϕ are on opposite sides of the common-population point-estimate of 31/36=.861 from the point-estimates (e.g. considering θ=.9 and ϕ=.8 for r=16 or θ=.8 and ϕ=.9 for r=21) leads to very high q-values. This reflects the compound fact that (1) obtaining data biased in opposite direction from two existing populations is extremely unlikely so (2) the existence of such data strongly implies more than one population. Less extreme but similar results are obtained for r measurements of 17 and 20. An r measurement of 18 indicates one population unless management is concerned about very extreme alternate populations. Finally an r measurement of 19 indicates two populations if management is careful about what those two populations are.

The dependence of both the p-value and q-value on the number of measurements is illustrated by the following example. Four measurements are assumed to yield values of N, n, R, and r given by $\{N_1,N_2,N_3,N_4\} = \{20,40,80,160\}$, $\{n_1,n_2,n_3,n_4\} = \{10,20,40,80\}$, $\{R_1,R_2,R_3,R_4\} = \{17,34,68,136\}$, and $\{r_1,r_2,r_3,r_4\} = \{9,18,36,72\}$. The second, third and fourth measurements are just multiples of the first. All four of these hypothetical measurements provide point estimates of θ and ϕ of 9/10=.9 and (17-9)/10=.8 respectively. The interior of the following table contains sets of q-values from four executions of the computer program. Each set has the q-value for the smallest sample size first and the largest last.

$\theta \backslash \phi$	.76	.80	.84
.86	{.279,.365,.447,.526}	{.352,.472,.595,.721}	{.444,.599,.752,.885}
.90	{.180,.217,.232,.221}	{.237,.303,.356,.397}	{.316,.421,.525,.633}
.94	{.084,.082,.059,.027}	{.118,.127,.111,.074}	{.169,.201,.209,.192}

The corresponding set of p-values is {.500,.331,.174,.060}. Note (A) that increasing the sample size decreases the p-value and increases the q-value for θ=.9 and ϕ=.8. Thus increasing the sample size, if the data remains proportionate, increases the justification for deciding that two population yielded the two samples. Note also (B) that the p-value has a more pronounced change than the q-values. Thus the p-value is more sensitive than the q-values. A unusually high q-value thus has at least as much significance as an unusually low p-value. Note finally (C) that increasing the sample size when θ=.90 & ϕ=.76, θ=.94 & ϕ=.76, θ=.94 & ϕ=.80, and θ=.94 & ϕ=.84 eventually leads to a decrease in the q-value. This corresponds to universal measurement implying exact results. (This large-measurement effect does not occur for θ=.86 & ϕ=.76, θ=.86 & ϕ=.80, θ=.86 & ϕ=.84, and θ=.90 & ϕ=.84 because they are on the opposite side of the point estimate from the extreme identified by the alternate hypothesis.)

The program is designed so the user can keep track of a sample of prominence and a category of interest. This enables the user to enter and analyze management's relevantly different populations. For example, consider the hypothetical case analyzed in Figures 7a and 7b. Suppose that a field-fired missile is being developed. Enough tests have been made on the initial design to obtain 107 hits and 14 misses. A set of shorter missile fins in proposed to make the field-assembly faster. A short series of tests on the short-fin version yields 11 hits and 3 misses. The short-fin test is prominent in the mind of the missile designer; the short fins should not be used if they significantly degrade the missile's performance. Figures 7a and 7b contain the input and terminal-screen output of an analysis using the Fisher-Irwin Exact Method. The first entry into the computer, 11, identifies the short-fin test as the sample of prominence and hits as the category of interest. The p-value, .249, is somewhat low but the advocates of fast assembly with short fins might claim that .249 is not close enough to zero to warrant the conclusion that short fins have degraded the missile's accuracy.

The q-value for ϕ=11/14=.786 and θ=107/121=.884 is .375. That is slightly higher than the p-value but it might not be large enough justify not using the short fins. If management sets the desired requirement at θ=.900 and decrees that ϕ=.850 is an unacceptable accuracy rate, the last table on Figure 7b provides a basis for decision. The q-value for θ=.900 and ϕ=.850 is .525. Since this is twice the p-value, management has a fairly strong basis for not using the short fins. If management leaves the desired requirement at .900 and raises the unacceptable level to .890, the q-value increases to .706. The argument for rejecting the short fins is thus quite strong if .890 is really an unacceptable accuracy rate.

SUMMARY

The p-value and q-value analysis of the Fisher-Irwin Exact Method has been developed. The p-value equation has been derived using two techniques: hypergeometric and binomial. The binomial technique has been extended to yield a q-value equation. This equation has been derived from two sources: possible category one measurements and possible category two measurements. This q-value equation has been shown to possess mathematical symmetry. The q-value for $p_I = p_{II}$ has been shown to equal one minus the p-value; this was predestined for the Fisher-Irwin Exact Method because it is a general property of the p-value and q-values. A computer program has been written. This makes the analysis practical. Analysts can perform voluminous calculations without approximations. Managers can consider the relative sizes and importance of the p-value and relevant q-values. Managers can iecide if the two samples are from one population or from two populations differing either (1) from the combined point estimate of the population or (2) according to (A) a desired population or standard and (B) an unacceptable population. Computer generated reports have been provided for communication between analysts and managers. The development of the p-value and q-value analysis of the Fisher-Irwin Exact Method has reached the stage of implementation.

CONCLUSION

The analyst has a responsibility to report all information influencing the decision. This information should be in a form that can be understood and used by the decision-maker. Reporting the p-value and relevant q-values satisfies both of these conditions. The p-value and q-values provide the decision-maker with estimates of the risks of making wrong decisions. This makes Fisher's Exact Test relevant.

Greetings! Welcome to a computerized Fisher-Irwin Exact Test Analysis.

Two independent samples are initially assumed to be from a single population. This assumption is rejected and the two samples are considered to represent two statistically different populations if management reaches two conclusions:
1) The p-value is deemed sufficiently low and
2) A q-value for relevantly different populations is deemed sufficiently high.

The q-value is the probability of falsely deciding that two populations exist. A q-value for two relevantly different populations is the probability of falsely deciding that those two populations are one population.

This computerized analysis does a one sided test in the direction indicated by the data. It requires four numerical inputs determining nine numbers:

	Category One (e.g. Success):	Category Two (e.g. Failure):	Sum:
Sample One:	r	n-r	n
Sample Two:	R-r	(N-n)-(R-r)	N-n
Total:	R	N-R	N

Data may be entered in two ways. The theoretical-statistician approach uses "N, n, R, & r". The reliability-engineer uses "r, n-r, R-r, & (N-n)-R-r)".

ENTER RELIABILITY-ENGINEER/THEORETICAL-STATISTICIAN APPROACH "R/S"

Figure 1. Terminal screen at program initiation.

```
ENTER RELIABILITY-ENGINEER/THEORETICAL-STATISTICIAN APPROACH "R/S"  s

ENTER SIZE OF POPULATION "N" 36
ENTER SIZE OF SAMPLE OF PROMINENCE "n" 21
ENTER # OF ITEMS FROM CATEGORY OF INTEREST IN POPULATION "R" 31
ENTER # OF ITEMS FROM CATEGORY OF INTEREST IN SAMPLE OF PROMINENCE "r" 19

ENTER "T" FOR TABLE OF Q-VALUES, "ANYTHING ELSE" TO SKIP TABLE   t

ENTER "C" FOR CLOSE LOOK AT Q-VALUE TABLE IN DATA SUGGESTED REGION,
"ANYTHING ELSE" TO SKIP    c

ENTER "M" FOR MANAGEMENT INDICATED BINOMIAL PARAMETERS TO REPLACE
THE DATA INDICATED PARAMETERS OF THE CATEGORY OF INTEREST IN THE
SAMPLE OF PROMINENCE (I.E. REPLACE θ =  19 /  21 = 0.905) AND THE
SAMPLE OF NON-PROMINENCE (I.E.  REPLACE φ =  12 /  15 = 0.800),
"ANYTHING ELSE" TO SKIP    m

ENTER "θ"    .9
ENTER "φ"    .8

ENTER "M" FOR MANAGEMENT INDICATED BINOMIAL PARAMETERS TO REPLACE
THE DATA INDICATED PARAMETERS OF THE CATEGORY OF INTEREST IN THE
SAMPLE OF PROMINENCE (I.E. REPLACE θ =  19 /  21 = 0.905) AND THE
SAMPLE OF NON-PROMINENCE (I.E. REPLACE φ =  12 /  15 = 0.800),
ANYTHING ELSE" TO SKIP    m

ENTER "θ"    .87
ENTER "φ"    .85

ENTER "M" FOR MANAGEMENT INDICATED BINOMIAL PARAMETERS TO REPLACE
THE DATA INDICATED PARAMETERS OF THE CATEGORY OF INTEREST IN THE
SAMPLE OF PROMINENCE (I.E. REPLACE θ =  19 /  21 = 0.905) AND THE
SAMPLE OF NON-PROMINENCE (I.E. REPLACE φ =  12 /  15 = 0.800),
"ANYTHING ELSE" TO SKIP    m

ENTER "θ"    1
ENTER "φ"    .7

ENTER "M" FOR MANAGEMENT INDICATED BINOMIAL PARAMETERS TO REPLACE
THE DATA INDICATED PARAMETERS OF THE CATEGORY OF INTEREST IN THE
SAMPLE OF PROMINENCE (I.E. REPLACE θ =  19 /  21 = 0.905) AND THE
SAMPLE OF NON-PROMINENCE (I.E. REPLACE φ =  12 /  15 = 0.800),
"ANYTHING ELSE" TO SKIP    skip

END OF PROGRAM
```

Figure 2. Sample of theoretical-statistician input.

ANALYSIS OF FISHER'S EXACT TEST

In the following standardly ordered table, the sample of prominence and category of interest are identified by the user as Sample I and Category 1.

	Category 1:	Category 2:	
Sample I:	r = 19	n - r = 2	n = 21
Sample II:	R - r = 12	(N -n) - (R - r) = 3	N - n = 15
	R = 31	N - R = 5	N = 36

For this data, the post-test risk of a Type I error is p-value = 0.337.

For this data's two suggested binomial parameters of the category of interest (i.e. θ = 19 / 21 = 0.905 and ϕ = 12 / 15 = 0.800), the post-test risk of a Type II error is q-value = 0.297.

For other binomial parameters of the category of interest, q-values may be estimated from the following table:

$\theta \setminus \phi$	0.050	0.150	0.250	0.350	0.450	0.550	0.650	0.750	0.850	0.950
0.050	0.663	0.956	0.990	0.997	0.999	1.000	1.000	1.000	1.000	1.000
0.150	0.182	0.663	0.866	0.946	0.978	0.992	0.997	0.999	1.000	1.000
0.250	0.058	0.389	0.663	0.827	0.917	0.963	0.986	0.996	0.999	1.000
0.350	0.020	0.210	0.456	0.663	0.809	0.903	0.958	0.986	0.997	1.000
0.450	0.007	0.105	0.285	0.483	0.663	0.804	0.903	0.963	0.992	1.000
0.550	0.003	0.048	0.158	0.314	0.491	0.663	0.809	0.917	0.978	0.999
0.650	0.001	0.019	0.075	0.174	0.314	0.483	0.663	0.827	0.946	0.997
0.750	0.000	0.006	0.027	0.075	0.158	0.285	0.456	0.663	0.866	0.990
0.850	0.000	0.001	0.006	0.019	0.048	0.105	0.210	0.389	0.663	0.956
0.950	0.000	0.000	0.000	0.001	0.003	0.007	0.020	0.058	0.182	0.663

For binomial parameters of the category of interest near those indicated by the data, q-values may be estimated from the following table:

$\theta \setminus \phi$	0.700	0.720	0.740	0.760	0.780	0.800	0.820	0.840	0.860	0.880	0.900
0.805	0.418	0.460	0.504	0.551	0.600	0.650	0.702	0.753	0.803	0.852	0.896
0.825	0.362	0.402	0.446	0.493	0.543	0.595	0.649	0.705	0.761	0.816	0.869
0.845	0.304	0.342	0.384	0.429	0.479	0.532	0.589	0.648	0.710	0.772	0.833
0.865	0.245	0.280	0.318	0.361	0.409	0.461	0.518	0.580	0.646	0.716	0.786
0.885	0.187	0.217	0.251	0.289	0.333	0.382	0.438	0.500	0.569	0.644	0.724
0.905	0.132	0.156	0.183	0.215	0.253	0.297	0.348	0.407	0.476	0.553	0.641
0.925	0.083	0.100	0.120	0.144	0.173	0.208	0.251	0.302	0.365	0.440	0.530
0.945	0.043	0.052	0.064	0.079	0.098	0.122	0.152	0.190	0.240	0.304	0.386
0.965	0.015	0.018	0.023	0.030	0.038	0.049	0.064	0.085	0.113	0.153	0.211
0.985	0.002	0.002	0.003	0.004	0.005	0.007	0.009	0.013	0.019	0.028	0.044

Figure 3a. First half of printer output from Figure 2 input.

For binomial parameters of the category of interest near those indicated by management, q-values may be estimated from the following table:

θ \ φ	0.700	0.720	0.740	0.760	0.780	0.800	0.820	0.840	0.860	0.880	0.900
0.800	0.431	0.473	0.518	0.564	0.613	0.663	0.713	0.763	0.812	0.859	0.902
0.820	0.375	0.416	0.460	0.507	0.557	0.609	0.663	0.717	0.772	0.826	0.876
0.840	0.318	0.357	0.399	0.445	0.495	0.548	0.604	0.663	0.723	0.784	0.842
0.860	0.259	0.295	0.334	0.378	0.426	0.479	0.536	0.597	0.663	0.730	0.798
0.880	0.201	0.232	0.267	0.306	0.351	0.402	0.458	0.520	0.589	0.663	0.740
0.900	0.145	0.170	0.199	0.233	0.272	0.318	0.370	0.431	0.499	0.577	0.663
0.920	0.094	0.112	0.134	0.160	0.192	0.229	0.274	0.328	0.393	0.469	0.559
0.940	0.051	0.062	0.076	0.093	0.115	0.141	0.175	0.217	0.271	0.338	0.424
0.960	0.020	0.025	0.031	0.040	0.050	0.064	0.083	0.108	0.142	0.189	0.255
0.980	0.003	0.004	0.006	0.007	0.010	0.013	0.018	0.025	0.035	0.052	0.078
1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

For binomial parameters of the category of interest near those indicated by management, q-values may be estimated from the following table:

θ \ φ	0.750	0.770	0.790	0.810	0.830	0.850	0.870	0.890	0.910	0.930	0.950
0.770	0.617	0.663	0.708	0.753	0.797	0.839	0.879	0.914	0.945	0.969	0.987
0.790	0.568	0.615	0.663	0.711	0.759	0.807	0.852	0.893	0.930	0.960	0.982
0.810	0.513	0.561	0.611	0.663	0.715	0.767	0.819	0.867	0.911	0.948	0.976
0.830	0.453	0.501	0.553	0.606	0.663	0.720	0.777	0.834	0.886	0.932	0.968
0.850	0.389	0.436	0.487	0.542	0.601	0.663	0.726	0.790	0.853	0.909	0.956
0.870	0.321	0.365	0.415	0.469	0.529	0.593	0.663	0.735	0.808	0.877	0.937
0.890	0.251	0.290	0.336	0.387	0.445	0.511	0.583	0.663	0.746	0.831	0.909
0.910	0.181	0.213	0.252	0.297	0.351	0.413	0.486	0.569	0.663	0.764	0.865
0.930	0.114	0.138	0.168	0.204	0.248	0.302	0.368	0.449	0.547	0.663	0.791
0.950	0.058	0.072	0.090	0.113	0.143	0.182	0.233	0.301	0.391	0.510	0.663
0.970	0.018	0.023	0.030	0.040	0.053	0.071	0.098	0.137	0.196	0.288	0.434

For binomial parameters of the category of interest near those indicated by management, q-values may be estimated from the following table:

θ \ φ	0.600	0.620	0.640	0.660	0.680	0.700	0.720	0.740	0.760	0.780	0.800
0.900	0.064	0.076	0.090	0.105	0.124	0.145	0.170	0.199	0.233	0.272	0.318
0.920	0.039	0.046	0.055	0.066	0.079	0.094	0.112	0.134	0.160	0.192	0.229
0.940	0.019	0.023	0.028	0.035	0.042	0.051	0.062	0.076	0.093	0.115	0.141
0.960	0.007	0.008	0.010	0.013	0.016	0.020	0.025	0.031	0.040	0.050	0.064
0.980	0.001	0.001	0.002	0.002	0.003	0.003	0.004	0.006	0.007	0.010	0.013
1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Figure 3b. Second half of printer output from Figure 2 input.

```
ENTER RELIABILITY-ENGINEER/THEORETICAL-STATISTICIAN APPROACH "R/S"  r

ENTER "TEST ONE NUMBER OF SUCCESSES" 20
ENTER "TEST ONE NUMBER OF FAILURES" 1
ENTER "TEST TWO NUMBER OF SUCCESSES" 11
ENTER "TEST TWO NUMBER OF FAILURES" 4

ENTER "T" FOR TABLE OF Q-VALUES, "ANYTHING ELSE" TO SKIP TABLE    skip

ENTER "C" FOR CLOSE LOOK AT Q-VALUE TABLE IN DATA SUGGESTED REGION,
"ANYTHING ELSE" TO SKIP    skip

ENTER "M" FOR MANAGEMENT INDICATED BINOMIAL PARAMETERS TO REPLACE
THE DATA INDICATED PARAMETERS OF THE CATEGORY OF INTEREST IN THE
SAMPLE OF PROMINENCE (I.E. REPLACE θ = 20 / 21 = 0.952) AND THE
SAMPLE OF NON-PROMINENCE (I.E. REPLACE φ = 11 / 15 = 0.733),
"ANYTHING ELSE" TO SKIP    m

ENTER "θ"    .9
ENTER "φ"    .8

ENTER "M" FOR MANAGEMENT INDICATED BINOMIAL PARAMETERS TO REPLACE
THE DATA INDICATED PARAMETERS OF THE CATEGORY OF INTEREST IN THE
SAMPLE OF PROMINENCE (I.E. REPLACE θ = 20 / 21 = 0.952) AND THE
SAMPLE OF NON-PROMINENCE (I.E. REPLACE φ = 11 / 15 = 0.733),
"ANYTHING ELSE" TO SKIP    m

ENTER "θ"    1
ENTER "φ"    .7

ENTER "M" FOR MANAGEMENT INDICATED BINOMIAL PARAMETERS TO REPLACE
THE DATA INDICATED PARAMETERS OF THE CATEGORY OF INTEREST IN THE
SAMPLE OF PROMINENCE (I.E. REPLACE θ = 20 / 21 = 0.952) AND THE
SAMPLE OF NON-PROMINENCE (I.E. REPLACE φ = 11 / 15 = 0.733),
"ANYTHING ELSE TO SKIP    skip

END OF PROGRAM
```

Figure 4. Sample of reliability-engineer input.

ANALYSIS OF FISHER'S EXACT TEST

Although Test One, Test Two, Successes, and Failures may be interchanged several ways mathematically, they have physical identities. To utilize these identities, [A] Test One (i.e. the test with 20 Successes and 1 Failure) is taken as the sample of prominence (i.e. it is considered physically more important than Test Two) and [B] Successes define the category of interest (i.e. the most natural description of a test result is considered to be Success instead of Failure).

In the following standardly ordered table, the sample of prominence and category of interest are identified by the user as Sample I and Category 1.

	Category 1:	Category 2:	
Sample I:	r = 20	n - r = 1	n = 21
Sample II:	R - r = 11	(N - n) - (R - r) = 4	N - n = 15
	R = 31	N - R = 5	N = 36

For this data, the post-test risk of a Type I error is p-value = 0.084.

For this data's two suggested binomial parameters of the category of interest (i.e. θ = 20 / 21 = 0.952 and ϕ = 11 / 15 = 0.733), the post-test risk of a Type II error is q-value = 0.241.

For binomial parameters of the category of interest near those indicated by management, q-values may be estimated from the following table:

$\theta \setminus \phi$	0.700	0.720	0.740	0.760	0.780	0.800	0.820	0.840	0.860	0.880	0.900
0.800	0.785	0.815	0.843	0.869	0.894	0.916	0.936	0.953	0.968	0.979	0.988
0.820	0.741	0.774	0.806	0.837	0.865	0.892	0.916	0.937	0.956	0.971	0.983
0.840	0.690	0.726	0.761	0.795	0.829	0.860	0.889	0.916	0.939	0.959	0.975
0.860	0.628	0.667	0.705	0.744	0.782	0.818	0.853	0.886	0.916	0.942	0.964
0.880	0.556	0.596	0.637	0.679	0.721	0.763	0.804	0.844	0.882	0.916	0.945
0.900	0.471	0.511	0.553	0.597	0.643	0.690	0.737	0.785	0.832	0.876	0.916
0.920	0.374	0.411	0.452	0.496	0.543	0.592	0.645	0.700	0.756	0.812	0.867
0.940	0.265	0.297	0.333	0.372	0.416	0.465	0.518	0.577	0.641	0.709	0.780
0.960	0.152	0.174	0.199	0.229	0.263	0.302	0.349	0.403	0.466	0.539	0.623
0.980	0.050	0.059	0.070	0.083	0.099	0.118	0.143	0.174	0.214	0.267	0.337
1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

For binomial parameters of the category of interest near those indicated by management, q-values may be estimated from the following table:

$\theta \setminus \phi$	0.600	0.620	0.640	0.660	0.680	0.700	0.720	0.740	0.760	0.780	0.800
0.900	0.302	0.332	0.363	0.397	0.433	0.471	0.511	0.553	0.597	0.643	0.690
0.920	0.225	0.250	0.277	0.307	0.339	0.374	0.411	0.452	0.496	0.543	0.592
0.940	0.149	0.168	0.188	0.211	0.237	0.265	0.297	0.333	0.372	0.416	0.465
0.960	0.079	0.090	0.102	0.117	0.133	0.152	0.174	0.199	0.229	0.263	0.302
0.980	0.024	0.027	0.032	0.037	0.043	0.050	0.059	0.070	0.083	0.099	0.118
1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Figure 5. Printer output from Figure 4 input.

```
ENTER RELIABILITY-ENGINEER/THEORETICAL-STATISTICIAN APPROACH "R/S"  s

ENTER SIZE OF POPULATION "N" 36
ENTER SIZE OF SAMPLE OF PROMINENCE "n" 21
ENTER # OF ITEMS FROM CATEGORY OF INTEREST IN POPULATION "R" 31
ENTER # OF ITEMS FROM CATEGORY OF INTEREST IN SAMPLE OF PROMINENCE "r" 18

                    ANALYSIS OF FISHER'S EXACT TEST

     In the following standardly ordered table, the sample of prominence and
category of interest are identified by the user as Sample I and Category 2.

                Category 1:              Category 2:
Sample I:           r =  3                 n - r  =  18          n =  21
Sample II:  R - r =  2       (N -n) - (R - r) =  13       N - n =  15
                    R =  5                 N - R  =  31          N =  36

For this data, the post-test risk of a Type I error is p-value = 0.663.

For this data's two suggested binomial parameters of the category of interest
(i.e. θ =  18 /  21 = 0.857 and φ =  13 /  15 = 0.867), the post-test risk
of a Type II error is q-value = 0.306.

ENTER "T" FOR TABLE OF Q-VALUES, "ANYTHING ELSE" TO SKIP TABLE    skip

ENTER "C" FOR CLOSE LOOK AT Q-VALUE TABLE IN DATA SUGGESTED REGION,
"ANYTHING ELSE" TO SKIP    skip

ENTER "M" FOR MANAGEMENT INDICATED BINOMIAL PARAMETERS TO REPLACE
THE DATA INDICATED PARAMETERS OF THE CATEGORY OF INTEREST IN THE
SAMPLE OF PROMINENCE (I.E. REPLACE θ =  18 /  21 = 0.857) AND THE
SAMPLE OF NON-PROMINENCE (I.E. REPLACE φ =  13 /  15 = 0.867),
"ANYTHING ELSE" TO SKIP    m

ENTER "θ"    .8
ENTER "φ"    .9

For binomial parameters of the category of interest near those indicated by
management, q-values may be estimated from the following table:

θ \ φ    0.700 0.720 0.740 0.760 0.780 0.800 0.820 0.840 0.860 0.880 0.900

0.800    0.158 0.184 0.215 0.250 0.291 0.337 0.391 0.452 0.521 0.598 0.682
0.820    0.127 0.149 0.176 0.207 0.244 0.287 0.337 0.396 0.464 0.542 0.630
0.840    0.098 0.117 0.139 0.166 0.198 0.237 0.283 0.337 0.403 0.480 0.569
0.860    0.073 0.088 0.106 0.128 0.155 0.188 0.228 0.277 0.337 0.411 0.501
0.880    0.051 0.062 0.076 0.093 0.114 0.141 0.174 0.216 0.270 0.337 0.423
0.900    0.033 0.041 0.050 0.063 0.078 0.098 0.124 0.158 0.202 0.260 0.337
0.920    0.019 0.024 0.030 0.037 0.048 0.061 0.079 0.103 0.136 0.182 0.246
0.940    0.009 0.011 0.014 0.019 0.024 0.032 0.042 0.057 0.077 0.108 0.154
0.960    0.003 0.004 0.005 0.007 0.009 0.012 0.016 0.022 0.032 0.047 0.071
0.980    0.000 0.001 0.001 0.001 0.001 0.002 0.003 0.004 0.006 0.009 0.015
1.000    0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
```

Figure 6. Input and screen output in case that alters order in table.

```
ENTER RELIABILITY-ENGINEER/THEORETICAL-STATISTICIAN APPROACH   "R/S"    r

ENTER "TEST ONE NUMBER OF SUCCESSES"  11
ENTER "TEST ONE NUMBER OF FAILURES"  3
ENTER "TEST TWO NUMBER OF SUCCESSES"  107
ENTER "TEST TWO NUMBER OF FAILURES"  14
```

ANALYSIS OF FISHER'S EXACT TEST

Although Test One, Test Two, Successes, and Failures may be interchanged several ways mathematically, they have physical identities. To utilize these identities, [A] Test One (i.e. the test with 11 Successes and 3 Failures) is taken as the sample of prominince (i.e. it is considered physically more important than Test Two) and [B] Successes define the category of interest (i.e. the most natural description of a test result is considered to be Success instead of Failure).

In the following standardly ordered table, the sample of prominence and category of interest are identified by the user as Sample II and Category 1.

	Category 1:	Category 2:	
Sample I:	r = 107	n - r = 14	n = 121
Sample II:	R - r = 11	(N - n) - (R - r) = 3	N - n = 14
	R = 118	N - R = 17	N = 135

For this data, the post-test risk of a Type I error is p-value = 0.249.

For this data's two suggested binomial parameters of the category of interest (i.e. ϕ = 11 / 14 = 0.786 and θ = 107 / 121 = 0.884), the post-test risk of a Type II error is q-value = 0.375.

```
ENTER "T" FOR TABLE OF Q-VALUES, "ANYTHING ELSE" TO SKIP TABLE    t
```

For other binomial parameters of the category of interest, q-values may be estimated from the following table:

θ \ ϕ	0.050	0.150	0.250	0.350	0.450	0.550	0.650	0.750	0.850	0.950
0.050	0.751	0.981	0.996	0.999	1.000	1.000	1.000	1.000	1.000	1.000
0.150	0.154	0.751	0.926	0.976	0.992	0.997	0.999	1.000	1.000	1.000
0.250	0.022	0.429	0.751	0.897	0.959	0.985	0.995	0.999	1.000	1.000
0.350	0.003	0.190	0.516	0.751	0.884	0.951	0.982	0.995	0.999	1.000
0.450	0.000	0.064	0.289	0.549	0.751	0.880	0.951	0.985	0.997	1.000
0.550	0.000	0.016	0.124	0.329	0.559	0.751	0.884	0.959	0.992	1.000
0.650	0.000	0.002	0.035	0.144	0.329	0.549	0.751	0.897	0.976	0.999
0.750	0.000	0.000	0.005	0.035	0.124	0.289	0.516	0.751	0.926	0.996
0.850	0.000	0.000	0.000	0.002	0.016	0.064	0.190	0.429	0.751	0.981
0.950	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.022	0.154	0.751

Figure 7a. First half of reliability-engineer input and screen output for hypothetical missile modification analysis.

```
ENTER "C" FOR CLOSE LOOK AT Q-VALUETABLE IN DATA SUGGESTED REGION,
"ANYTHING ELSE" TO SKIP    c

For binomial parameters of the category of interest near those indicated by
the data, q-values may be estimated from the following table:

θ \ φ    0.686 0.706 0.726 0.746 0.766 0.786 0.806 0.826 0.846 0.866 0.886

0.784    0.499 0.550 0.603 0.655 0.706 0.755 0.801 0.844 0.882 0.916 0.943
0.804    0.430 0.483 0.537 0.593 0.648 0.702 0.755 0.805 0.850 0.891 0.926
0.824    0.358 0.409 0.464 0.521 0.580 0.639 0.698 0.755 0.809 0.859 0.902
0.844    0.283 0.331 0.384 0.441 0.501 0.564 0.628 0.693 0.756 0.815 0.869
0.864    0.208 0.251 0.299 0.353 0.412 0.476 0.544 0.614 0.686 0.756 0.823
0.884    0.138 0.173 0.213 0.260 0.314 0.375 0.443 0.517 0.595 0.676 0.757
0.904    0.079 0.103 0.132 0.168 0.212 0.265 0.327 0.399 0.480 0.569 0.663
0.924    0.035 0.048 0.065 0.088 0.118 0.156 0.205 0.265 0.339 0.427 0.530
0.944    0.010 0.014 0.021 0.031 0.045 0.065 0.093 0.131 0.184 0.255 0.349
0.964    0.001 0.002 0.003 0.005 0.007 0.012 0.020 0.033 0.053 0.087 0.140
0.984    0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.001 0.002 0.003 0.008

ENTER "M" FOR MANAGEMENT INDICATED BINOMIAL PARAMETERS TO REPLACE
THE DATA INDICATED PARAMETERS OF THE CATEGORY OF INTEREST IN THE
SAMPLE OF PROMINENCE (I.E. REPLACE φ = 11 / 14 = 0.786) AND THE
SAMPLE OF NON-PROMINENCE (I.E. REPLACE θ = 107 / 121 = 0.884),
"ANYTHING ELSE" TO SKIP    m

ENTER "φ"    .85
ENTER "θ"    .9

For binomial parameters of the category of interest near those indicated by
management, q-values may be estimated from the following table:

θ \ φ    0.750 0.770 0.790 0.810 0.830 0.850 0.870 0.890 0.910 0.930 0.950

0.800    0.618 0.673 0.726 0.776 0.824 0.867 0.905 0.937 0.962 0.980 0.992
0.820    0.550 0.608 0.666 0.724 0.778 0.830 0.876 0.916 0.948 0.973 0.989
0.840    0.471 0.532 0.595 0.658 0.721 0.781 0.837 0.887 0.929 0.962 0.984
0.860    0.384 0.445 0.510 0.578 0.648 0.717 0.785 0.847 0.901 0.945 0.976
0.880    0.291 0.348 0.411 0.481 0.556 0.634 0.712 0.789 0.860 0.919 0.964
0.900    0.197 0.245 0.302 0.368 0.442 0.525 0.614 0.706 0.796 0.877 0.942
0.920    0.110 0.145 0.189 0.243 0.310 0.389 0.482 0.585 0.696 0.805 0.902
0.940    0.044 0.063 0.088 0.123 0.170 0.232 0.313 0.415 0.539 0.679 0.822
0.960    0.009 0.014 0.022 0.034 0.054 0.084 0.131 0.201 0.306 0.454 0.646
0.980    0.000 0.000 0.001 0.002 0.003 0.006 0.013 0.026 0.056 0.122 0.265
1.000    0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

ENTER "M" FOR MANAGEMENT INDICATED BINOMIAL PARAMETERS TO REPLACE
THE DATA INDICATED PARAMETERS OF THE CATEGORY OF INTEREST IN THE
SAMPLE OF PROMINENCE (I.E. REPLACE φ = 11 / 14 = 0.786) AND THE
SAMPLE OF NON-PROMENENCE (I.E. REPLACE θ = 107 / 121 = 0.884),
"ANYTHING ELSE" TO SKIP    skip

END OF PROGRAM
```

Figure 7b. Second half of hypothetical missile modification analysis.

ATTENDANCE LIST
FOR THE
TUTORIAL SEMINAR IN LINEAR REGRESSION
17-18 OCTOBER 1988
AND THE
THIRTY FOURTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY
RESEARCH, DEVELOPMENT AND TESTING
19-21 OCTOBER 1988

TUTORIAL (T)
CONFERENCE (C)

C LAUNER , Robert L.
Department of Statistics
University of South Carolina
Columbia, SC 29208 (803) 777-7800

C THRASHER, Paul H.
Reliability Division
TE-RE
White Sands Missile Range
New Mexico 88002 (505) 678-6177

T,C WOLFF, Weston C.
Flight Safety
NR-CF
White Sands Missile Range
New Mexico 88002 (505) 678-2205

T,C BLANCO, Abel J.
Atmospheric Sciences Lab
SLCAS-AE-E
White Sands Missile Range
New Mexico 88002 (505) 678-3924

T,C HALL, Irving J.
Div. 6415
Sandia National Labs
PO Box 5800
Albuquerque, New Mexico 87185 (505) 844-4208

C, TESTERMAN, Deloris
STEYP-MT-TA-A
Yuma Proving Ground (AV) 899-3251

C CONOVER, W.J.
College of Bus. Admin
Texas Tech University
Lubbock TX 79409 (806) 742-1546

C DAVID, H.A.
Iowa State University
Department of Statistics
ISU, Ames, IA 50011 (515) 294-7749

T,C DOYLE, Mary
USA TROSCOM
Product Assurance
4300 Goodfellow
St. Louis, MO 63120-1798 (314) 263-9468

C HOCKING, Ron
Texas A&M
Department of Statistics
College Station, TX 77843 (409) 845-3151

C ZACKS, S.
SUNY
Binghamton Center
Department of Math Sciences
Suny, Binghamton NY 13901 (607) 777-2619

T,C BRANDON, Dennis L.
U.S. Army Corps of Engineers
Waterways Experiment Station
P.O. Box 631
Vicksburg, MS 39180-0631 (601) 634-2807

T,C STRATTON, Willard F.
USAMC Materiel Readiness
Support Activity
Lexington, KY 40511-5101 (606) 293-4174

C CHANDRA, Jagdish
U.S. Army Research Office
Mathematical Sciences Division
P.O. Box 12211
Research Triangle Park
NC 27709 (919) 549-0641

C ESSENWANGER, Dr Oskar M.
Research Directorate
U.S. Army Missile Command
ATTN: AMSMI-RD-RE-AP
Redstone Arsenal, AL 35898-5248 (AV) 876-4872

T,C ABRAHIM, Mohd
Mathematical Department
New Mexico State University

T,C BANDAS, Sadin
Mathematical Department
New Mexico State University

T HOVICK, Chris
Mathematical Department
New Mexico State University

T DALE, Richard H.
Data Sciences Division
STEWS-NR-AM
White Sands Missile Range, NM 88002

T GADNEY, George
Data Sciences Division
STEWS-NR-AD
White Sands Missile Range, NM 88002

T,C MCLAUGHLIN, Dale R.
Data Sciences Division
STEWS-NR-AM
White Sands Missile Range, NM 88002

T,C COUNCIL, Konrad K.
Data Sciences Division
STEWS-NR-AM
White Sands Missile Range, NM 88002

T ZERR, Ronald
AMXTM-CW-WS
White Sands Missile Range, NM 88002

T,C CHAMBERS, Charles E.
Data Sciences Division
STEWS-NR-MA
White Sands Missile Range, NM 88002

T,C BURGE, J. Robert
Walter Reed Army Institute of Research
Washington, DC 20307 (202) 576-3151

T,C TANG, Douglas
Walter Reed
Army Institute of Research
Washington, DC 20307-5100 (202) 576-7212

T,C GREGORY, Gavin
UTEP
Math Department
UTEP, El Paso, TX (915) 747-5761

T,C AVARA, Elton P.
U.S. Army Atmospheric
Sciences Laboratory
ATIN: SLCAS-AE-E (AVARA)
White Sands Missile Range
N.M 88002-5501 (505) 678-1570

C WOODS, Anthony K.
USATROSCOM, System Analysis Office
Modeling & Techniques Div.
4300 Goodfellow Blvd.
St. Louis MO 63120-1798 (314) 263-2926

T,C MOSS, Linda L.C.
Ballistics Research Laboratory
U.S. Army Ballistic Research Lab
Aberdeen Proving Ground
MD 21005-5066 (301) 278-6832

T,C THOMAS, Jerry
Ballistics Research Laboratory
SLCBR-SE-P
Aberdeen Proving Ground
MD 21005-5066 (301) 278-6728

T,C BODT, Barry A.
BRL
SLCBR-SE-P
Aberdeen Proving Ground
MD 21005-5066 (301) 278-6646

T,C UMHOLTZ, Robert L.
U.S. Army Ballistic
Research Lab
SLCBR-SE-P
Aberdeen Proving Ground
MD 21005-5066 (301) 278-6832

T,C WEBB, David W.
Ballistic Research Laboratory
ATIN: SLCBR-SE-P
Aberdeen Proving Ground
MD 21005-5066 (301) 298-6646

T,C TAYLOR, Malcom S.
Ballistic Research Laboratory
ATIN: SLCBR-SE-P
Aberdeen Proving Ground
MD 21005-5066 (301) 298-6646

T QUINZI, Tony
TRAC
White Sands Missile Range
NM 88002 (505) 678-4356

T,C RUSSELL, Carl T.
OTEA
ATIN: CSTE-CA
5600 Columbia Pike
Falls Church, VA 22041-5115 (AV) 289-2305

C HALL, Charles E. Jr.
U.S. Army Missile Command
ATIN: AMSMI-RD-RE-OP/C.E. Hall
Redstone Arsenal AL 35898-5248 (205) 876-3934

C LEHNIGK, Siegfried H.
U.S. Army Missile Command
ATIN: AMSMI-RD-RE-OP/S. H. Lehnigh
Restone Arsenal, AL 35898-5248 (205) 876-3526

T,C GREEN, Robert E.
STEWS-ID-P
White Sands Missile Range,
New Mexico 88002 (505) 678-2291

T,C DUTOIT, Gene
U.S. Army Infantry School (404) 545-3165

T,C GRIMES, Dr. Fred M.
TEXCOM Combined
Arms Test Directorate
Ft Hood, TX 76544-5065 (AV) 738-9614

T,C SWINGLE, Donald M.
New Mexico Research Institute
Box 160
Las Cruces, NM 88004 (505) 522-5197

C BATES, Carl B.
U.S. Army Concepts Analysis Agency
8120 Woodmont Ave
Bethesda, MD 20814-2797 (AV) 295-0163

C IGLEHART, Donald L.
Stanford University
Department of Operations Research
Stanford, CA 94305-4022 (415) 723-0850

T,C VISNAW, Vernon V.
US Army Combat Systems Activity
Aberdeen Proving Ground MD
21005-5059 (301) 278-7503

C SOLNOKY, P.
Counsellor, Defense R&D.
Canadian Embassy
2450 Massachusetts Ave NW
Washington, DC 20008 (202) 483-5505

T,C DRESSEL, Francis
U.S. Army Research Office
PO Box 12211
Research Triangle Park
North Carolina 27709-2211 (919) 549-0641

C BRYSON, Dr. Marion R.
TEXCOM Experimentation Center
Ford Ord, CA 93941-7000 (408) 242-4414

T,C TINGEY, Henry B.
University of Delaware
Dept Math Sciences
Newark DE 19716 (302) 451-8034

T,C BISSINGER, Bafney
Penn State University
Hershey Foods
281 W. Main St
Middletown, PA 17057 (717) 944-0649

C GONZALEZ, Ramiro
TRAC
White Sands Missile Range, NM 88001

C SHUSTER, Eugene
Mathematical Department
University of Texas -El Paso
El Paso, Texas

C ROJO, Javier
Mathematical Department
University of Texas - El Paso
El Paso, Texas

C CHENG, Cheng
Mathematical Department
University of Texas - El Paso
El Paso, Texas

C LIU, Yan
Mathematical Department
University of Texas - El Paso
El Paso, Texas

C POLLARD, Charles
RAM Division
White Sands Missile Range, New Mexico 88002

C KAIGH, Bill
Mathematical Department
University of Texas - El Paso
El Paso, Texas

C PARZEN, Emanuel
Department of Statistics
Texas A&M University
College Station, Texas 77843

T,C COHEN, Herb
U.S. Army Materiel Systems Analysis Activity
Aberdeen Proving Ground, MD 21005-5066

C HARMON, Joe
U.S. Army Airborne Board
Fort Bragg, NC 28307 (AV) 236-5115

T,C CASTILLO, Cesar
Data Sciences Division
STEWS-NR-AM
White Sands Missile Range, NM 88002

T,C DALTON, Oren N.
Data Sciences Division
STEWS-NR-AM
White Sands Missile Range, NM 88002

T CATHERSON, Janet
PENTASTAR/FWD
White Sands Missile Range, NM 88002

T,C COHEN, Herb
AMSAA
Aberdeen Proving Ground MD 21005

DJANG, Phillip
TRAC
White Sands Missile Range, NM 88001

T ACKER, Clay D.
Data Sciences Division
STEWS-NR-AM
White Sands Missile Range, NM 88002

T,C ROGERS, Gerald
Math Department
New Mexico State University

C COX, Paul
2930 Huntington Drive
Las Cruces, New Mexico 88001

T,C CULPEPPER, Gideon
Las Cruces, NM

T,C PAGE, Woodrow
1125 Larry Drive
Las Cruces, New Mexico 88001

C ANDERSEN, GERALD

C VANGEL, Mark
U.S. Army Materials Technology Lab
Watertown Arsenal

C,T HARRIS, Bernard
University of Wisconsi, Madison

C WEST, Larry
TECOM Headquarters
Aberdeen Proving Ground, MD 21005-5066

C WEGMAN, Edward
George Mason University

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188
Exp. Date: Jun 30, 1986

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) ARO Report 89-2	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Army Research Office 6b. OFFICE SYMBOL (If applicable) SLCRO-ARO	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) P.O. Box 12211 Research Triangle Park, NC 27709	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AMSC on behalf of ASO (ADA) 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Proceedings of the Thirty-Fourth Conference on the Design of Experiments in Army Research, Development and Testing

12. PERSONAL AUTHOR(S)

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM Jan 88 TO Feb 89	1989 July	367

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

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

This is a technical report of the Thirty-Fourth Conference on the Design of Experiments in Army Research, Development and Testing. It contains most of the papers presented at this meeting. These articles treat various Army statistical and design problems.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Francis G. Dressel	22b TELEPHONE (Include Area Code) (919) 549-0641, Ext. 324 22c. OFFICE SYMBOL SLCRO-MA

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED