ARO Report 83-2

PROCEEDINGS OF THE TWENTY-EIGHTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH DEVELOPMENT AND TESTING

Sponsored by
The Army Mathematics Steering Committee
on Behalf of

THE CHIEF OF RESEARCH, DEVELOPMENT AND ACQUISITION

U. S. Army Research Office

Report No. 83-2

June 1983

PROCEEDINGS OF THE TWENTY-EIGHTH CONFERENCE ON THE DESIGN OF EXPERIMENTS

Sponsored by the Army Mathematics Steering Committee

HOST

U. S. Army Combat Developments Experimentation Command
Fort Ord, California

HELD AT

Hilton Inn Resort
Monterey, California

20-22 October 1982

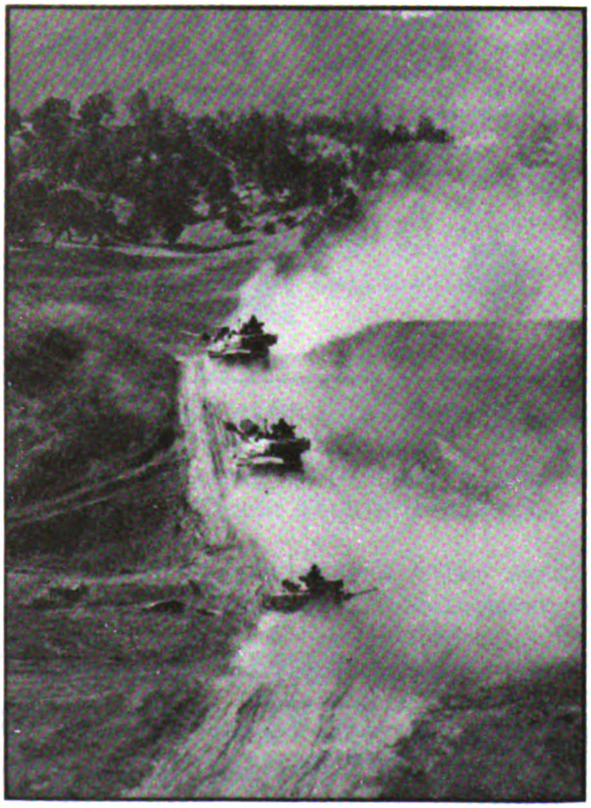
Approved for public release; distribution unlimited. The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

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

FOREWORD

The twenty-Eighth Conference on the Design of Experiments in Army Research, Development and Testing had as its host the U. S. Army Combat Development Experimentation Command, Fort Ord, California. It was held at the Hilton Inn Resort, Monterey, California, on 20-22 October 1982. A brief history of the host installation appeared in a booklet issued the attendees at this meeting. Exerpts from this booklet are reproduced below.

U.S. ARMY COMBAT DEVELOPMENTS EXPERIMENTATION COMMAND



FORT ORD, CALIFORNIA

COMBAT DEVELOPMENTS EXPERIMENTATION COMMAND

The United States Army Combat Developments Experimentation Command was established in 1956 at Fort Ord, Calif., setting the stage for the introduction of a new form of military evaluation . . . the Combat Field Experiment.

The command is more familiar to many people, military and civilian alike, as "CDEC." With its highly sophisticated electronic field laboratory, located at Fort Hunter Liggett, CDEC's military-scientific team has the mission of providing hard factual answers to basic questions about how the Army of the future should be organized, how it should be equipped and how it can best fight!

To provide such answers CDEC has developed a method of evaluation significantly different from anything previously available to the military decision maker. CDEC applies the technique of the scientific field experiment to military systems and problems. Experiments are conducted under conditions simulating as closely as possible those of an actual combat situation.

CDEC began with soldiers armed with stopwatches, compasses, slide rules and clipboards as data collection systems. Today, CDEC has evolved into a high technology command using computers, lasers, intervisibility equipment, an accurate position location system and other sophisticated measurement equipment.

"Vision to Victory" is CDEC's motto. It is a vision with a purpose that centers on the production of hard data through field experimentation and operations research pointing the way to the Army of tomorrow.

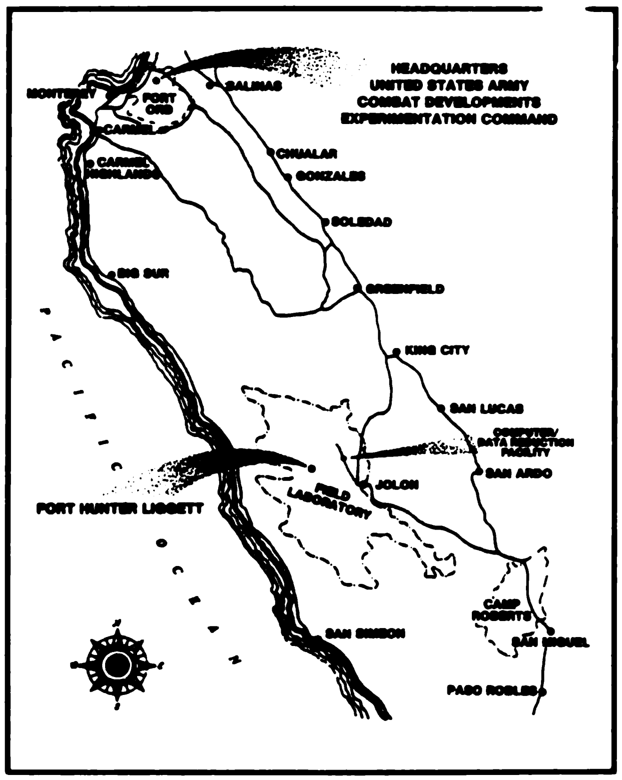
CDEC's experiments involve highly realistic mock battles during which casualties from various types of engagements are taken out of the battle by a computer operating as a high-speed, impartial umpire.

Here, weapons "fire" beams of light instead of live ammunition. Each weapon is equipped with a low-power, eye-safe laser system and each soldier-player or vehicle, such as a tank or aricraft, has several laser detectors which send signals when "hit" by a laser beam. The players and weapons are linked by radio to a central computer.

When a laser is fired, a coded impulse is sent to the computer. If the beam hits any of several laser detectors on the target, its particular coded impulse allows the computer to determine instantly which weapon it came from, its location and whether it was powerful and close enough to destroy the target. These instruments tell players almost instantly if they were "hit" or "missed" during these simulated firefights.

Computers take tanks and other vehicles out of action when hit by simulated fire. When ammo loads are exhausted, these same computerized systems can simply "turn off" the players' weapons.

CDEC



U.S. ARMY COMBAT DEVELOPMENTS EXPERIMENTATION COMMAND

The Army Mathematics Steering Committee (AMSC), the sponsor of the Conferences on the Design of Experiments, appreciates the fact that the Combat Development Experimentation Command was willing, for a second time, to serve as host for these conferences. For both of these meetings, Dr. Marion Bryson, Scientific Advisor, USA CDEC, has served as chairman on local arrangements. His careful planning helped make these two of the most profitable meetings in this series of statistical symposia.

The Program Committee would like to thank Dr. Larry Crow of the U.S. Army Materiel Systems Analysis Agency for organizing the "Special Software Test and Evaluation Session"; and also Mr. Langhorne Withers, U.S. Army Operational Test and Evaluation Agency, for arranging a special session devoted to "Logistic Supportability". The agenda gives information about these two interesting solicited events for this conference. Members of the Program Committee feel they were fortunate in obtaining the following nationally known scientists to give invited addresses at this meeting.

SPEAKER AND AFFILIATION	TITLE OF ADDRESS
Professor Brad Efrom Stanford University	Bootstrap Methods
Professor Leo Breiman University of California-Berkeley	Tools in Data Analysis
Professor David W. Scott Rice University	Nonparametric Bivariate Density Estimation as a Tool for Data Analysis
Professor Nancy R. Mann University of California-Los Angeles	The Influence of W. Edward Deming on the Implementation of Statistical Quality Control--The Early Days and Now

Another event associated with this conference was a tutorial seminar on "Non-Parametric Statistics". It was given, on 18 - 19 October 1982, at the U.S. Army Combat Developments Experimentation Command by Professor William J. Conover, Texas Tech University.

The winner of the second Wilks Award for Contributions to Statistical Methodologies in Army Research, Development and Testing was presented to Professor Bernard Harris of the Mathematics Research Center, University of Wisconsin-Madison, at a banquet held at the Naval Postgraduate School on Wednesday night, 20 October 1982. This honor was bestowed on Dr. Harris for his many contributions to various statistical fields. He has helped Army scientists with many of their design problems, and his advice in conducting these conferences has proved invaluable. He recently developed, together with Dr. Andrew P. Soms, new methodologies and optimality results for the long unsolved problem of confidence bounds for system reliability.

Members of the AMSC feel that it is appropriate to again express their thanks to Mr. Philip G. Rust of Thomasville, Georgia for endowing both of the Wilks

Awards. The first one entitled, "The Samuel S. Wilks Memorial Medal and Award", was initiated in 1964 and is now being administered by the American Statistical Association. The second one, initiated in 1981, is called "The Wilks Award for Contributions to Statistical Methodologies in Army Research, Development and Testing", and is under the auspices of the AMSC. Mr. Rust's generous gifts in memory of his friend, Sam Wilks, will contribute to the welfare of the military services as well as foster statistical science in general.

TABLE OF CONTENTS*

*This table of contents contains only the papers that are published in this technical manual. For a list of all papers presented at the Twenty-Eighth Conference on the Design of Experiments, see the Program of the meeting.

A G E N D A

for the

TWENTY-EIGHTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN

ARMY RESEARCH, DEVELOPMENT AND TESTING

20-22 October 1982

Host: US Army Combat Developments Experimentation Command

Location: Hilton Inn Resort, Monterey, California

***** Wednesday, 20 October *****

0815-0915 REGISTRATION -- Hilton Inn, Second Floor Foyer

0915-0930 CALLING OF THE CONFERENCE TO ORDER (Presidio Room)

Marion R. Bryson, Scientific Advisor, US Army Combat Developments Experimentation Command

WELCOMING REMARKS

BG Grail L. Brookshire, Commanding General, US Army Combat Developments Experimentation Command

0930-1200 GENERAL SESSION I (Presidio Room)

CHAIRMAN - Marion R. Bryson, Scientific Advisor, US Army Combat Developments Experimentation Command

0930-1030 KEYNOTE ADDRESS

BOOTSTRAP METHODS

Brad Efron, Stanford University

1030-1100 BREAK

1100-1200 TOOLS IN DATA ANALYSIS

Leo Breiman, University of California-Berkeley

1200-1330 LUNCH

1330-1500 TECHNICAL SESSION I - GOODNESS OF FIT (Presidio Room)

CHAIRMAN - Oskar Essenwanger, US Army Missile Command

ENTROPY INTERPRETATION OF TESTS FOR NORMALITY BY SHAPIRO-WILK STATISTICS

Emanuel Parzen, Texas A&M University

EXACT PROBABILITY LEVELS FOR MULTI-SAMPLE SMIRNOV-TYPE STATISTICS

William E. Baker and Malcolm S. Taylor, Ballistic Research Laboratory

ANALYZING n SAMPLES OF 2 OBSERVATIONS EACH

J. R. Knaub, Jr., K. S. Druschel, L. M. Grile and G. Petet, US Army Logistics Center

1330-1500 TECHNICAL SESSION II - FIELD DATA ANALYSIS (Vista del Mar Room)

CHAIRMAN - Carl Bates, US Army Concepts Analysis Agency

IN-MINEFIELD EFFECTIVENESS (IME) MEASURE FOR BREACHING VEHICLES

Mark S. Adams, US Army Mobility Equipment Research and Development Command

ROBUST RANGE MEASUREMENT PREPROCESSING

William S. Agee and Robert H. Turner, US Army White Sands Missile Range

HOW GOOD IS GOOD - A FIELD EVALUATION OF CAMOUFLAGE

Ronald L. Johnson and George Anitole, US Army Mobility Equipment Research and Development Command

1330-1500 TECHNICAL SESSION III - RELIABILITY AND FAILURE ANALYSIS (Peninsula Room)

CHAIRMAN - Jerry Thomas, Ballistic Research Laboratory

ESTIMATING MEAN LIFE FROM LIMITED TESTING

Donald W. Rankin, US Army White Sands Missile Range

ANALYSIS OF FERROGRAPHIC ENGINE WEAR DATA USING QUALITY CONTROL TECHNIQUES

Robert L. Launer and Edward A. Saibel, US Army Research Office

AN EFFICIENT METHOD FOR DETERMINING THE "A" AND "B" DESIGN ALLOWABLES

Donald Neal, US Army Materials & Mechanics Research Center

1500-1530 BREAK

1530-1700 TECHNICAL SESSION IV - DESIGN OF EXPERIMENTS (Presidio Room)

CHAIRMAN - William L. Lehmann, Chief Scientist, Scientific Services Support Company, Ft. Ord

CHOICE OF RESPONSE SURFACE DESIGN AND ALPHABETIC OPTIMALITY

G.E.P. Box, Mathematics Research Center, University of Wisconsin-Madison

AMBUSHED BY A LURKING VARIABLE

Barry Bodt and Jerry Thomas, Ballistic Research Laboratory

THE USE OF THE R( )-NOTATION FOR THE EVALUATION OF RESULTS OF PARALLEL LINE ASSAYS

J. Robert Burge, Walter Reed Army Institute of Research

1530-1700 TECHNICAL SESSION V - MULTIVARIATE AND AUTO CORRELATED PROCESSES (Vista del Mar Room)

CHAIRMAN - Gerald M. Powell, US Army Natick R&D Laboratories

THE USE OF BOX-JENKINS METHODOLOGY IN FORECASTING DARCOM'S CENTRAL PROCUREMENT WORKLOAD

Charles A. Correia, US Army Logistics Management Center

WIND VARIABILITY IN THE BOUNDARY LAYER AND ITS ASSOCIATION WITH TURBULENCE RED AND WHITE NOISE

Oskar Essenwanger, US Army Missile Command

COMPARISON OF CEP ESTIMATORS FOR ELLIPTICAL NORMAL ERRORS

Audrey E. Taub and Marlin A. Thomas, Naval Surface Weapons Center

1830 OPEN BAR AND BANQUET (Officers Club, Naval Postgraduate School, Monterey)

*****Thursday, 21 October *****

0830-1030 CLINICAL SESSION A (Presidio Room)

CHAIRMAN - Carl T. Russell, US Army Cold Regions Test Center

PANELISTS - Bernard Harris, Mathematics Research Center
Emanuel Parzen, Texas A&M University

VALIDATION OF MULTIPLE TERRAIN FACTOR VALUE PREDICTIONS USING LIMITED GROUND TRUTH MEASUREMENTS

James H. Robinson and Gerald W. Turnage, US Army Engineer Waterways Experiment Station

A NEED FOR A METHODOLOGY FOR PRIORITIZATION OF MISSION AREA (MA) DEFICIENCIES

Richard T. Maruyama, HQS, US Army Training and Doctrine Command

0830-1030 TECHNICAL SESSION VI - SIMULATION AND SOFTWARE VALIDATION (Vista del Mar Room)

CHAIRMAN - Charles A. Correia, US Army Logistics Management Center

QUADRATURE DESIGNS FOR UNBIASED ESTIMATION OF INTERPOLATION

Andrew F. Siegel and Fanny Zambuto, Princeton University

ESTIMATING THE VARIANCE OF THE LOSS EXCHANGE RATIO

Eugene Dutoit, US Army Infantry School

AN EXAMPLE OF SOFTWARE VALIDATION USING A FACTORIAL DESIGN

Joseph M. Tessmer, Department of Energy

1030-1100 BREAK

1100-1200 GENERAL SESSION II (Presidio Room)

CHAIRMAN - Malcolm Taylor, Ballistic Research Laboratory

NONPARAMETRIC BIVARIATE DENSITY ESTIMATION AS A TOOL FOR DATA ANALYSIS

David W. Scott, Rice University

1200-1330 LUNCH

1330-1700 LOGISTIC SUPPORTABILITY - Special Presentation and Round Table Discussion (Presidio Room)

CHAIRMAN - Walter W. Hollis, Deputy Undersecretary of the Army, Operations Research

Langhorne Withers, US Army Operational Test and Evaluation Agency

1300-1305 INTRODUCTION - Testing Supportability in OT

Langhorne Withers, US Army Operational Test & Evaluation Agency

1305-1310 PANEL WELCOME AND INTRODUCTION

Walter W. Hollis, DUSA/OR
Russell R. Shorey, OASD/MRA&L
MG William K. Hunzeker, USA Log Center
R. L. Ely, Applied Physics Lab, The Johns Hopkins University
BG Kenneth A. Jolemore, DCSLDG/S&M
COL W. B. Moats, AF TEC

1310-1355 OTEA PRESENTATION

Improving the Test and Evaluation of Integrated Logistics Support in OT

Steve French
LTC Larry Davis
Doug McGowan
Bill Dunn

1355-1410 QUESTIONS AND COMMENTS ON OTEA APPROACH

Panel and Attendees

1410-1455 DARCOM PRESENTATION

Supportability-Requirements, Design, Test and Evaluation

Arthur Nordstrom

1455-1515 BREAK

1515-1530 QUESTIONS ON THE DARCOM APPROACH

Panel and Attendees

1530-1615 TRADOC PRESENTATION

Logistic Supportability Testing and Evaluation During OT

Donald Reich

1615-1630 QUESTIONS ON THE TRADOC APPROACH

Panel and Attendees

1630-1700 ROUND TABLE DISCUSSION

Panel

1730-1930 COCKTAIL RECEPTION - Open Bar; Hilton Inn

*****Friday, 22 October *****

0830-1030 CLINICAL SESSION B (Presidio Room)

CHAIRMAN - Langhorne Withers, US Army Operational Test & Evaluation Agency

PANELISTS - G.E.P. Box, Mathematics Research Center
Robert Bechhofer, Cornell University

EFFECTIVENESS ANALYSIS OF SELECTIVE REENLISTMENT BONUS

Carl B. Bates and Breton C. Graham, US Army Concepts Analysis Agency

THE PERVERSITY OF MISSING POINTS IN 2^4 DESIGNS

Carl T. Russell, US Army Cold Regions Test Center

0830-1030 TECHNICAL SESSION VII - DISTRIBUTION THEORY AND LIFE TESTING (Vista del Mar Room)

CHAIRMAN - David F. Cruess, Uniformed Services University of Health Sciences

SENSITIVITY AND ASYMPTOTIC PROPERTIES OF BAYESIAN RELIABILITY ESTIMATES

Bernard Harris and Andrew P. Soms, Mathematics Research Center, University of Wisconsin

AN INFERENTIAL APPROACH TO EXPERIMENTAL DESIGN FOR QUANTAL RESPONSE MODELS

Tom Leonard, Mathematics Research Center, University of Wisconsin

DATA-BASED NONPARAMETRIC ESTIMATION OF THE HAZARD FUNCTION WITH APPLICATIONS TO MODEL DIAGNOSTICS AND EXPLORATORY ANALYSIS

Martin A. Tanner, Mathematics Research Center, University of Wisconsin

1030-1100 BREAK

1100-1215 GENERAL SESSION III (Presidio Room)

CHAIRMAN - Douglas B. Tang, Walter Reed Army Institute of Research and Chairman, AMSC Subcommittee on Probability and Statistics

OPEN MEETING OF THE AMSC SUBCOMMITTEE ON PROBABILITY AND STATISTICS

THE INFLUENCE OF W. EDWARDS DEMING ON THE IMPLEMENTATION OF STATISTICAL QUALITY CONTROL -- THE EARLY DAYS AND NOW

Nancy R. Mann, University of California-Los Angeles

1215-1330 LUNCH

1330-1700 SPECIAL SOFTWARE TEST AND EVALUATION SESSION (Presidio Room)

CHAIRMAN - Larry Crow, US Army Materiel Systems Analysis Agency

1330-1345 NEED FOR MEASURING SOFTWARE RELIABILITY

Ray Bell, US Army Materiel Systems Analysis Activity

1345-1415 NEED AND PROBLEMS ASSOCIATED WITH REPRESENTATIVE SOFTWARE TEST DATA

Steve French, US Army Operational Test & Evaluation Agency

1415-1515 METHODS FOR GENERATING REPRESENTATIVE SOFTWARE TEST DATA, SCORING OF DATA, AND ANALYSIS OF DATA

Larry Crow, James Kniss, Paul Kunselman, and Gregory J. Gibson, US Army Materiel Systems Analysis Activity

1515-1530 BREAK

1530-1600 CHARACTERISTICS AND ANALYSIS OF SOFTWARE FAILURE

Larry Crow, US Army Materiel Systems Analysis Activity and Nozer D. Singpurwalla, George Washington University

1600-1645 PANEL DISCUSSION

CHAIRMAN: James Kniss, US Army Materiel Systems Analysis Activity

1645 ADJOURN

ENTROPY INTERPRETATION OF GOODNESS OF FIT TESTS

Emanuel Parzen
Institute of Statistics
Texas A&M University

ABSTRACT. This paper describes a synthesis of statistical reasoning called FUN.STAT (because it is fun; functional (useful); based on functional analysis; estimates functions; and all graphs are of functions). FUN.STAT has three important components: quantile and density-quantile signatures of populations, entropy and information measures, and functional statistical inference.

A FUN.STAT approach to the problem of identifying the probability distribution $F(x)$ of a random variable X from a random sample is outlined. To identify F_0 in the location-scale parameter model $F(x) = F_0((x-\mu)/\sigma)$, we estimate entropy difference $\Delta = H^0(f) - H(f)$. $H(f)$ is Shannon entropy and $H^0(f) = \log \sigma + H(f_0)$ is entropy of the assumed model (which may maximize entropy). Estimators $\hat{H}_1$, $\hat{H}_2$, $\hat{H}_3$ of $H(f)$ are defined which are respectively fully parametric, fully non-parametric, and parametric-select. Significance levels for $\hat{\Delta}$ are obtained by Monte Carlo methods. The family of parametric-select estimators of Δ may provide optimum tests of F_0 (such as normal or exponential) and estimators of F when one rejects F_0.

KEY WORDS: Entropy-based statistical inference, goodness of fit tests, test for normality, Shapiro-Wilk statistic, quantile, density-quantile, quantile-density, autoregressive density estimator.

1. INTRODUCTION. Let $X_1,\ldots,X_n$ be a random sample of a continuous random variable X with distribution function $F(x) = \Pr[X \le x]$, $-\infty<x<\infty$, and quantile function $Q(u) = F^{-1}(u)$, $0<u<1$. Tests of normality or exponentiality are special cases of a location-scale parameter model, which we denote by the hypothesis

$$H_0: F(x) = F_0\left(\frac{x-\mu}{\sigma}\right), \quad Q(u) = \mu + \sigma\, Q_0(u)$$

where $F_0(x)$ is a specified distribution with quantile function $Q_0(u)$. Table 1 lists F_0 and Q_0 for various standard distributions.

Research supported by the U. S. Army Research Office Grant DAAG 29-80-C-0070.

Table 1. STANDARD DISTRIBUTION FUNCTIONS AND QUANTILE FUNCTIONS

Name	$F_0(x)$	$Q_0(u)$
Normal	$\Phi(x) = \int_{-\infty}^{x} \phi(y)\, dy$, $\phi(x) = (2\pi)^{-\frac{1}{2}} \exp - \frac{1}{2} x^2$	$\Phi^{-1}(u)$
Exponential	$1 - e^{-x}$	$\log (1-u)^{-1}$
Weibull, Quantile shape parameter β	$1 - e^{-x^c}$, $x \geq 0$ $c = \frac{1}{\beta}$	$\left\{\log (1-u)^{-1}\right\}^{\beta}$
Extreme value of minimum	$1 - e^{-e^x}$ $-\infty<x<\infty$	$\log \log (1-u)^{-1}$
Extreme value of maximum	$e^{-e^{-x}}$ $-\infty<x<\infty$	$- \log \log u^{-1}$
Log normal	$\Phi(\log x)$, $x>0$	$\exp \Phi^{-1}(u)$
Logistic	$1-(1+e^x)^{-1}$	$\log \frac{u}{1-u}$

Many statistics have been introduced by statisticians to test the composite (location and scale parameters unspecified) hypothesis of normality. A superior omnibus test of normality (in terms of power) seems to be provided by a test statistic $W = \hat{\sigma}_2/\hat{\sigma}_1$, where $\hat{\sigma}_1$ and $\hat{\sigma}_2$ are scale estimators defined as follows: $\hat{\sigma}_1$ is sample standard deviation, while $\hat{\sigma}_2$ is a linear combination of order statistics estimator of σ. We call W a statistic of Shapiro-Wilk type because it is a variant of a test introduced by Shapiro and Wilk (1965) and Shapiro and Francia (1972).

The question arises: to discover a motivation for the W statistic which explains the source of its power, and to use this insight to extend W to other distributions F_0. In this paper we propose that the power of W can be explained by representing it as an "entropy difference" test statistic. We show that the test statistic for normality introduced by Vasicek (1977) is also an entropy difference statistic, as are test statistics introduced in Parzen (1979).

2. INFORMATION DIVERGENCE AND ENTROPY. To compare two distribution functions F(x) and G(x) with probability densities f(x) and g(x), a useful measure is information divergence, defined by

$$I(f;g) = \int_{-\infty}^{\infty} \{-\log \frac{g(x)}{f(x)}\} f(x)\, dx$$

It can be decomposed into cross-entropy

$$H(f;g) = \int_{-\infty}^{\infty} \{-\log g(x)\} f(x)\, dx$$

and entropy

$$H(f) = H(f;f) = \int_{-\infty}^{\infty} \{-\log f(x)\} f(x)\, dx$$

by the important identity

$$0 \le I(f;g) = H(f;g) - H(f).$$

To estimate entropy it is useful to express it in terms of the quantile density function q(u) and density-quantile function fQ(u) defined by

$$q(u) = Q'(u), \quad fQ(u) = f(Q(u)) = \{q(u)\}^{-1}$$

By making the change of variable u = F(x) one can show that

$$H(f) = \int_0^1 - \log fQ(u)\, du$$

$$= \int_0^1 \log q(u)\, du.$$

Under the hypothesis H_0 that $F(x) = F_0((x-\mu)/\sigma)$, a location-scale model, $q(u) = \sigma q_0(u)$ and

$$H(f) = \log \sigma + H(f_0).$$

3. ENTROPY DIFFERENCE TO TEST GOODNESS OF FIT. To test the hypothesis H_0 we propose to investigate (and eventually establish how to use optimally) test statistics which are entropy-difference statistics

$$\Delta(f) = H^0(f) - H(f)$$

where $H^0(f)$ is a parametric evaluation of the entropy of f, evaluated under the assumption that it obeys H_0, defined by

$$H^0(f) = \log \sigma + H(f_0),$$

while H(f) is a non-parametric evaluation of H(f), usually most conveniently obtained by

$$H(f) = \int_0^1 \log q(u)\, du \quad .$$

To estimate H(f) we have three types of estimators which we call

$\hat{H}_1$ fully parametric estimator,

$\hat{H}_2$ fully non-parametric estimator,

$\hat{H}_3$ smooth or parametric select estimator

Similarly to estimate $H^0(f)$ we have several types of estimators depending on the estimator $\hat{\sigma}_j$ we adopt for σ; thus

$$\hat{H}^0{}_j = \log \hat{\sigma}_j + H(f_0)$$

Three important possibilities for $\hat{\sigma}_j$ are:

$\hat{\sigma}_1$ maximum likelihood estimator,

$\hat{\sigma}_2$ optimal linear combination of order statistics estimator

$\hat{\sigma}_3$ estimator of score deviation $\sigma_3 = \int_0^1 f_0Q_0(u)\, q(u)\, du$.

Under H_0 these estimators are all asymptotically efficient estimators of σ.

While one can conceive of about 9 possible estimators of the entropy difference Δ, we discuss only three estimators which we denote $\hat{\Delta}_{11}$, $\hat{\Delta}_{12}$, and $\hat{\Delta}_{33}$.

4. ENTROPY-DIFFERENCE INTERPRETATION OF SHAPIRO-WILK STATISTIC

To test the hypothesis H_0: X is $N(\mu,\sigma^2)$, a test statistic W of Shapiro-Wilk type is of the form

$$W = \hat{\sigma}_2 \div \hat{\sigma}_1$$

where $\hat{\sigma}_1$ is the sample standard deviation and

$$\hat{\sigma}_2 = \sum_{j=1}^{n} \Phi^{-1} \left(\frac{j-0.5}{n}\right) X_{(j)} \div \left\{ \sum_{j=1}^{n} \left| \Phi^{-1} \left(\frac{j-0.5}{n}\right) \right|^2 \right\}^{\frac{1}{2}}$$

is an asymptotically efficient estimator of σ based on linear combinations of the order statistics $X_{(1)} < \ldots < X_{(n)}$ of the random sample. The first step in the entropy interpretation of W is to consider instead the statistic

$$\hat{\Delta}_{11} = - \log W = \log \hat{\sigma}_1 - \log \hat{\sigma}_2 = \hat{H}_1^o - \hat{H}_1$$

where [with $f_o(x) = \phi(x) = (2\pi)^{-\frac{1}{2}} \exp -(\frac{1}{2}) x^2$, and $H(f_o) = \frac{1}{2} (1 + \log 2\pi)$]

$$\hat{H}_1^o = \log \hat{\sigma}_1 + H(f_o)$$

is an estimator of $H^o(f)$ based on $\hat{\sigma}_1$, and $\hat{H}_1$ is a purely parametric estimator of $H(f)$ based on the parametric estimator $\hat{\sigma}_2$; note $\hat{H}_1 = \hat{H}_2^o$.

Significance levels for the entropy-difference statistic $\hat{\Delta}_{11} = - \log W$ are obtainable from tables of the W statistic [for example, Filliben (1975)]. An example of 5% significance levels (for accepting normality) are

$\hat{\Delta}_{11} \leq 0.05$, for sample size n = 20 ;

$\hat{\Delta}_{11} \leq 0.023$, for sample size n = 50 .

5. ENTROPY-DIFFERENCE INTERPRETATION OF VASICEK STATISTIC

To test the hypothesis H_0: X is $N(\mu,\sigma^2)$ Vasicek (1977) proposes a statistic which is equivalent to

$$\hat{\Delta}_{12} = \hat{H}_1^o - \hat{H}_2$$

where $\hat{H}_1^o$ is an estimator of the parametric evaluation $H^o(f)$ of entropy, and

$\hat{H}_2$ is a fully non-parametric estimator of H(f) based on the gap or leap (of order 2ν) estimator

$$\tilde{q}_\nu\left(\frac{j}{n+1}\right) = \frac{n+1}{2\nu}\{X_{(j+\nu)} - X_{(j-\nu)}\} \quad , \quad j=\nu+1,\ldots,n-\nu$$

of q(j/(n+1)), and

$$\hat{H}_2 = \frac{1}{n-2\nu} \sum_{j=\nu+1}^{n-\nu} \tilde{q}_\nu \left(\frac{j}{n+1}\right)$$

Some significance levels of $\hat{\Delta}_{12}$ are given in Table 2; they are transformations of the significance levels given by Vasicek (1977) and obtained by Monte-Carlo simulation.

6. ENTROPY-DIFFERENCE INTERPRETATION OF PARZEN GOODNESS OF FIT PROCEDURE

To test the general hypothesis H_0: X is $F_0(\frac{x-\mu}{\sigma})$, Parzen (1979) proposes forming raw estimators $\tilde{d}(u)$ of

$$d(u) = \frac{1}{\sigma_0} f_0Q_0(u)\, q(u) \ ,$$

where $\sigma_0 = \int_0^1 f_0Q_0(t)\, q(t)\, dt$. To form $\tilde{d}(u)$ and $\tilde{\sigma}_0$ we replace q(u) by the least smooth gap estimator $\tilde{q}_2$ (u). Smooth estimators $d_m(u)$ of d(u) are formed by the autoregressive method. From estimators of the pseudo-correlations

$$\rho(v) = \int_0^1 e^{2\pi iuv}\, d(u)\, du, \quad v=0,\pm 1,\ldots,\pm m$$

one estimates the coefficients of the autoregressive order m approximator

$$d_m(u) = K_m \left| 1 + \alpha_m(1)\, e^{2\pi iu} + \ldots + \alpha_m(m)\, e^{2\pi ium}\right|^{-2}$$

to d(u). The coefficient K_m plays an important role in entropy calculations since

$$\int_0^1 - \log d_m(u)\, du = -\log K_m$$

can be regarded as an estimator $\quad \hat{\Delta}_{33} = \int_0^1 - \log \hat{d}(u)\, du$ of Δ.

This formula, which we prove below, provides an entropy-difference interpretation of the goodness of fit procedures in Parzen (1979).

To prove this interpretation of Δ_{33}, write

$$- \log d(u) = \log \sigma_0 - \log f_0Q_0(u) - \log q(u)$$

Therefore

$$\int_0^1 - \log d(u)\, du = H^O(f) - H(f)$$

is an entropy-difference.

The autoregressive estimator $\hat{d}_m(u)$ of $d(u)$ provides a parametric select estimator of $q(u)$ by

$$\hat{q}(u) = \tilde{\sigma}_0 \hat{d}_m(u)\, q_0(u)$$

A parametric select estimator of $H(f)$ is

$$\hat{H}_3 = \int_0^1 \log \hat{q}(u)\, du$$

$$= \int_0^1 \log \hat{d}_m(u)\, du + \hat{H}_3^O$$

where

$$\hat{H}_3^O = \log \tilde{\sigma}_0 + H(f_0)$$

is an estimator of $H^O(f)$ based on $\tilde{\sigma}_0$.

The parametric select entropy-difference test statistic $\hat{\Delta}_{33}$ should be denoted $\hat{\Delta}_{33,m}$ because it depends on the order m of the autoregressive estimator $\hat{d}_m(u)$ of $d(u)$. Significance levels of $\hat{\Delta}_{33,m}$ derived by a very approximate Monte Carlo simulation (in the case of testing for normality) are given in Table 2. They show that the parametric select estimators of Δ provide a smooth progression of significance levels from the fully parametric estimators of Δ to the fully non-parametric estimators. In practice, we recommend adaptive determination of the order m by the data, rather than choosing a fixed order m.

It may be useful to use a rough approximation to the 5% significance levels of $\hat{\Delta}_{33,m}$ which is provided by $2m/n$. A criterion for accepting H_0: X is $F_0(\frac{X-\mu}{\sigma})$ is:

$$\hat{\Delta}_{33,m} = - \log \hat{K}_m \leq \frac{2m}{n} \quad , m=1,2,\dots \quad .$$

One rejects H_0 if there exists a value of m for which the Akaike-type criterion

$$AIC(m) = \frac{2m}{n} + \log \hat{K}_m \leq 0 \qquad ;$$

the value of m which minimizes AIC(m) is chosen as an "optimal" value $\hat{m}$. An optimal parametric-select estimator of the true quantile-density function q(u) is

$$\hat{q}_{\hat{m}}(u) = \tilde{\sigma}_0 \, \hat{d}_{\hat{m}}(u) \, q_0(u) \quad .$$

7. CONCLUSION

We believe that the interpretation given in this paper of powerful goodness of fit procedures as entropy-difference statistics provides a striking demonstration of the FUN.STAT synthesis of statistical reasoning. In addition to elegance of the theory, very practical and implementable procedures are obtained.

The parametric select estimators $\hat{\Delta}_{33,m}$ of entropy-difference test statistics for goodness of fit have for m=1 approximately the properties of fully parametric estimators (such as Shapiro-Wilk $\hat{\Delta}_{11}$) and have for large values of m approximately the properties of fully non-parametric estimators (such as Vasicek $\hat{\Delta}_{12}$). Thus it appears the series $\hat{\Delta}_{33,m}$ provide all the test-statistics required. Further the autoregressive approach provides non-parametric estimators of the true distribution when one rejects the null hypothesis H_0.

One may find that a sample passes the goodness of fit procedure for two null hypotheses. An appealing procedure, whose properties remain to be investigated, is to choose that null hypothesis for which $\hat{\Delta}_{33,m}$ is always less than the corresponding statistic for the other hypothesis.

The entropy-difference statistics $\hat{\Delta}_{33,m}$ are implemented in our one-sample univariate data analysis computer program ONESAM. Table 3 lists auto-regressive estimates of entropy-difference when testing for normality data sets in Stigler (1977). An asterisk indicates a data set which is not normal in our judgement.

In Table 2 we report significance levels for $\hat{\Delta}_{12}$ obtained (by Monte Carlo calculations) by Dudewicz and van der Muelen (1981) in the case of testing for uniformity rather than normality.

The closeness of the Dudewicz-van der Muelen levels to the Vasicek levels suggests a conjecture, which remains to be proved, that the entropy-difference statistics have distributions which are approximately the same for all null hypotheses H_0: X is $F_0(\frac{x-\mu}{\sigma})$.

A final noteworthy feature is that the autoregressive method of estimating quantile-density functions and density-quantile functions, introduced in Parzen (1979), can be shown to have a maximum entropy property [compare Parzen (1982)].

Table 2. 5% SIGNIFICANCE LEVELS FOR ENTROPY DIFFERENCE STATISTICS

Accept H_0: X is $N(\mu,\sigma^2)$ for some μ and σ if entropy difference is less than threshold given.

Sample Size n	$\hat{\Delta}_{11}$ Shapiro-Wilk	$\hat{\Delta}_{33,m}$ Autoregressive order m Monte Carlo 5% level (rough approximation 2m/n)					$\hat{\Delta}_{12}$ Vasicek gap estimator $\tilde{q}_\nu(u)$ (Dudewicz-van der Muelen)				
		m=1	m=2	m=3	m=4	m=5	ν=5	ν=4	ν=3	ν=2	ν=1
20	.05	.141	.235	.299	.378	.398		.40	.40	.43	.61
		(.10)	(.20)	(.30)	(.40)	(.50)		(.43	.43	.47	.66)
50	.023	.045	.081	.126	.153	.176	.21	.21	.23		
		(.04)	(.08)	(.12)	(.16)	(.20)	(.22	.22	.24)		

Shapiro-Wilk and Vasicek levels are based on Monte Carlo simulation of normal; Dudewicz-van der Muelen levels are based on Monte Carlo simulation of uniform.

One can conjecture a relation between gap order 2ν and autoregressive order m for the corresponding estimators to have similar distributions and therefore similar significance levels:

$$(2\nu)\ m = n = \text{sample size}$$

To understand what this conjecture is alleging note that for n=20, m=4 is similar to 2ν = 6; for n=50, m=6 is similar to 2ν = 8.

When one uses gap estimators of q(u), and thus of entropy, one has the problem of determining the order 2ν. One can more easily develop criteria for determining the order m of autoregressive estimators of q(u).

Table 3. ANALYSIS OF STIGLER (1977) DATA SETS BY ONESAM PROGRAM

		$\|\tilde{\rho}(\nu)\|^2$			$\hat{\Delta}_{33,m}$			AIC(m)			Opt. Order
Stigler Data Set	Sample Size	ν=1	ν=2	ν=3	m=1	m=2	m=3	m=1	m=2	m=3	$\hat{m}$
1	18	.042	.025	.057	.04	.08	.17	.07	.15	.17	0
*2	17	.193	.030	.042	.21	.27	.34	-.10	-.03	.02	1
3	18	.108	.027	.047	.11	.14	.17	-.00	.08	.16	0
4	21	.057	.159	.041	.06	.20	.21	.04	-.01	.08	2
5	21	.146	.015	.041	.16	.17	.22	-.06	.01	.07	1
6	21	.047	.102	.002	.05	.13	.15	.05	.06	.14	0
7	21	.041	.046	.040	.04	.11	.18	.05	.08	.11	0
8	21	.079	.047	.011	.08	.18	.27	.01	.01	.02	0
*9	20	.285	.235	.124	.34	.42	.42	-.24	-.22	-.12	1
10	20	.027	.059	.045	.03	.09	.15	.07	.11	.15	0
11	26	.046	.006	.033	.05	.06	.11	.03	.09	.12	0
12	20	.107	.001	.023	.11	.13	.13	-.01	.07	.17	1
13	20	.084	.027	.063	.09	.16	.20	.01	.04	.10	0
*14	20	.162	.094	.130	.18	.22	.39	-.08	-.02	-.09	3
15	20	.066	.006	.001	.07	.09	.09	.03	.11	.21	0
*16	20	.080	.056	.093	.08	.17	.44	.01	.03	-.14	3
17	23	.065	.014	.038	.07	.11	.14	.02	.07	.12	0
19	29	.002	.019	.008	.00	.02	.03	.07	.12	.18	0

REFERENCES

Dudwicz, E. J. and Van der Muelen, E. C. (1981). Entropy-Based Tests of Uniformity Journal of the American Statistical Association, 76, 967-974.

Filliben, J. J. (1975). The probability plot correlation coefficient test for normality, Technometrics, 17, 111-117.

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

Parzen, E. (1982). Maximum entropy interpretation of autoregressive spectral densities. Statistics and Probability Letters, 1, 2-6.

Shapiro, S. S. and Francis, R. S. (1972). Approximate analysis of variance test for normality. J. American Statistical Association, 67, 215-216.

Shapiro, S. S. and Wilk, M. B. (1968). An analysis of variance test for normality, Biometrika, 52, 591-611.

Shapiro, S. S., Wilk, M. B. and Chen, H. J. (1968). A comparative study of various tests for normality, J. American Statistical Association, 63, 1343-1372.

Stigler, S. M. (1977) Do robust estimators work with real data, Annals of Statistics, 5, 1055-1098.

Vasicek, O. (1976). A Test for Normality Based on Sample Entropy, Journal of the Royal Statistical Society, B, 38, 54-59.

EXACT PROBABILITY LEVELS FOR MULTI-SAMPLE SMIRNOV-TYPE STATISTICS

William E. Baker
Malcolm S. Taylor
Experimental Design and Analysis Branch
Systems Engineering and Concepts Analysis Division
U.S. Army Ballistic Research Laboratory
Aberdeen Proving Ground, Maryland

ABSTRACT. Let there be given c independent random samples of continuous random variables of size $n_1, n_2, \ldots, n_c$; and denote the observations in the i^{th} sample by $x_{i1}, x_{i2}, \ldots, x_{in_i}$. Suppose it is desired to test the null hypothesis that the samples all come from the same population.

Birnbaum and Hall proposed for this null hypothesis the test statistic

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

for i,j = 1,2, ... ,c, where F_i^* denotes the empirical cumulative distribution function for the i^{th} sample. In their 1960 paper they published the probabilities $P[D(n,n,n) < r]$ for n = 1(1)20(2)40 where r = k/n, k = 1,2, ...,n.

The tables referenced here are an extension of the Birnbaum-Hall Tables, resulting from the examination of a larger number of samples and the consideration of unequal sample sizes. In addition, an application of this work to a problem in ballistics is discussed. Although length precludes the inclusion of the tables in this paper, they are available in a technical report published by the Ballistic Research Laboratory under the same title.

I. INTRODUCTION. Let there be given c independent random samples of continuous random variables of size $n_1, n_2, \ldots ,n_c$, where $n_1 + n_2 + \ldots + n_c = N$; and denote the observations in the i^{th} sample by $x_{i1}, x_{i2}, \ldots, x_{in_i}$. Suppose it is desired to test the null hypothesis that the samples all come from the same population.

Birnbaum and Hall* proposed for this null hypothesis the following two-sided and one-sided test statistics respectively:

*Birnbaum, Z.W. and Hall, R.A., "Small Sample Distributions for Multi-Sample Statistics of the Smirnov Type," The Annals of Mathematical Statistics, Vol. 31, No. 3, pp. 710-720, September 1960.

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

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

$i,j = 1,2, \ldots ,c$, where F_i^* denotes the empirical cdf for the i^{th} sample.

II. COMPUTATIONAL METHODS.

Under the null hypothesis, the c samples may be considered as c successive drawings of $n_1,n_2, \ldots ,n_c$ observations from the same population, with equal probability of each of the N! ways of drawing the ordered sample of size N. The values taken on by the random vector $[F_1^*(x),F_2^*(x), \ldots ,F_c^*(x)]$, for fixed x, after a component-wise transformation $k_i = n_i F_i^*(x)$, establish a one-to-one correspondence between the combined ordered samples and a path in c-space from $(0,0, \ldots ,0)$ to $(n_1,n_2, \ldots ,n_c)$. The number of distinct combined ordered samples (paths) is $N!/(n_1!n_2! \ldots n_c!)$, each of which is equally likely to occur under the null hypothesis.

Introducing the notation

$Q(k_1,k_2, \ldots ,k_c)$ = number of paths from $(0,0, \ldots ,0)$ to $(k_1,k_2, \ldots ,k_c)$,

the following difference equation is established:

$$Q(k_1,k_2, \ldots ,k_c) = Q(k_1-1,k_2, \ldots ,k_c) + Q(k_1,k_2-1, \ldots ,k_c) + \ldots + Q(k_1,k_2, \ldots ,k_c-1)$$

with initial condition $Q(0,0, \ldots ,0) = 1$.

For a given subset $R \subset \mathcal{R} = \{(k_1,k_2, \ldots ,k_c) | 0 \leq k_i \leq n_i, i=1,2, \ldots ,c\}$, let $Q(k_1,k_2, \ldots ,k_c;\bar{R})$ = number of paths from $(0,0, \ldots ,0)$ to $(k_1,k_2, \ldots ,k_c)$ not containing points of R.

And, as before

$$Q(k_1,k_2, \ldots ,k_c;\bar{R}) = Q(k_1-1,k_2, \ldots ,k_c;\bar{R}) + Q(k_1,k_2-1, \ldots ,k_c;\bar{R}) + \ldots + Q(k_1,k_2, \ldots ,k_c-1;\bar{R})$$

with conditions

$$Q(0,0, \ldots ,0) = 1,$$
$$Q(k_1,k_2, \ldots ,k_c;\bar{R}) = 0, \text{ for } (k_1,k_2, \ldots ,k_c) \in R.$$

Under the null hypothesis, the probability that the samples determine a path from $(0,0, \ldots ,0)$ to $(n_1,n_2, \ldots ,n_c)$ which does not encounter any point of R is

$$P_{\bar{R}} = Q(n_1,n_2, \ldots ,n_c;\bar{R})/[N!/(n_1!n_2! \ldots n_c!)].$$

If our decision rule is to reject the null hypothesis whenever the samples determine a path containing points in a given set R, then $1 - P_{\bar{R}}$ is the probability of an error of the first kind.

The regions of rejection for the two-sided and one-sided tests, are respectively

$$D(n_1,n_2, \ldots ,n_c) > r \quad \text{and} \quad D^+(n_1,n_2, \ldots ,n_c) > r,$$

which determine the corresponding sets R, R^+

$$R = \{(k_1,k_2, \ldots ,k_c) \mid \sup_{i,j} | n_jk_i - n_ik_j| > n_in_jr\},$$

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

where $i,j = 1,2, \ldots ,c$.

The tables were computed by evaluating the difference equation for $Q(k_1,k_2, \ldots ,k_c;\bar{R})$.

For sample sizes not included in the tables, the inequality

$$P[D(n_1,n_2, \ldots ,n_c) \leq r] = P[\sup_{x,i,j} | F_i^*(x) - F_j^*(x)| \leq r]$$

$$= 1 - P[\sup_x |F_i^*(x) - F_j^*(x)| > r \text{ for some } i < j]$$

$$\geq 1 - \sum_{i<j\leq \ldots \leq k} \sum \ldots \sum P[D(n_i,n_j, \ldots ,n_k) > r], \; k < c,$$

allows the tabled values for both equal and unequal samples to be used to test the null hypothesis. The test will be conservative, but should prove useful for the range of values of c for which it would likely be applied.

Asymptotic expressions are not available for extension of these tables. As a matter of fact, Hodges* points out that asymptotic expressions advanced for the case $c = 2$ are inaccurate to an extent that their usefulness is questionable.

III. EXACT TABLES. In their paper, Birnbaum and Hall published the probabilities $P[D(n,n,n) < r]$ for $n = 1(1)20(2)40$ and $r = k/n$, $k = 1,3, \ldots ,n$, such that the probabilities for each n range from less than .90 to more than .995.

The tables referenced here are an extension of the Birnbaum-Hall tables. The first table contains the probabilities $P[D(n, \ldots ,n) < r)$ with the number of independent samples of size n ranging from three to seven and the corresponding sample sizes taking on values from sixty for the three-sample case to five for the seven-sample case. Birnbaum and Hall's values are a subset of the first table as indicated in Figure 1.

Perhaps of more importance, the second table contains the quantiles that allow the test to be applied to samples of unequal size, an option presently not available. Here for the useful cases $c = 3$, $2 < n_1,n_2,n_3 < 25$ and $c = 4$, $2 < n_1,n_2,n_3,n_4 < 15$, are tables $P[D(n_1, \ldots ,n_c) < r]$ for $r = 0.1,0.2, \ldots ,1.0$ as represented in Figure 2.

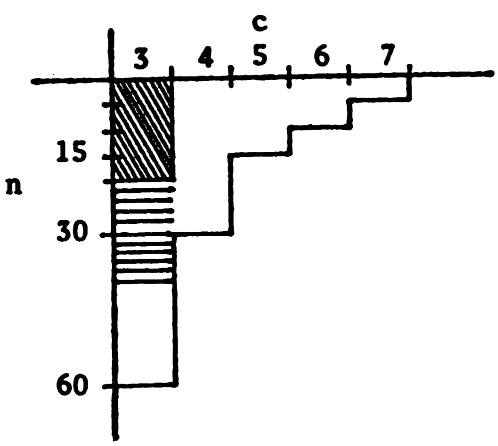


Figure 1

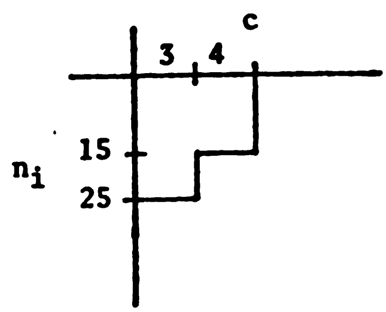


Figure 2

*Hodges, J.L., Jr., "The Significance Probability of the Smirnov Two-Sample Test," Arkiv für Mathematik, Vol. 3, 1957. pp. 469-486.

IV. EXAMPLE. For purpose of illustration, consider four independent random samples, of sizes five, seven, seven, and thirteen respectively.

Sample I	-0.927, -2.243, 0.815, -0.341, -0.250
Sample II	-0.451, -1.516, -1.447, 0.504, 1.645, 0.022, 1.098
Sample III	0.032, -1.772, 1.049, -0.073, -2.053, -1.123, 0.799
Sample IV	0.250, -0.185, -0.028, 0.004, -0.462, -0.032, 0.279, -1.053 0.597, 0.235, 0.510, 1.103, 0.241

The null hypothesis to be tested is that all four samples come from the same distribution. Evaluation of the statistic D(5,7,7,13), ordinarily programmed for machine calculation, can be carried out in the following manner:

1) Pool and order the samples

-2.243,	-2.053,	-1.772,	-1.516,	-1.447,	-1.123,	-1.053,	-0.927
-0.462,	-0.451,	-0.341,	-0.250,	-0.185,	-0.073,	-0.032,	-0.028
0.004,	0.022,	0.032,	0.235,	0.241,	0.250,	0.279,	0.504
0.510,	0.597,	0.799,	0.815,	1.049,	1.098,	1.103,	1.645

2) For every x in the pooled sample, evaluate $F_i^*(x)$, i = 1,2,3,4, and calculate Sup $|F_i^*(x) - F_j^*(x)|$, i,j = 1,2,3,4

x	$F_1^*(x)$	$F_2^*(x)$	$F_3^*(x)$	$F_4^*(x)$	Sup
-2.243	1/5	0	0	0	0.20
-2.053	1/5	0	1/7	0	0.20
-1.772	1/5	0	2/7	0	0.29
-1.516	1/5	1/7	2/7	0	0.29
-1.447	1/5	2/7	2/7	0	0.29
-1.123	1/5	2/7	3/7	0	0.43
-1.053	1/5	2/7	3/7	1/13	0.35
-0.927	2/5	2/7	3/7	1/13	0.32
-0.462	2/5	2/7	3/7	2/13	0.28
-0.451	2/5	3/7	3/7	2/13	0.28
-0.341	3/5	3/7	3/7	2/13	0.45
-0.250	4/5	3/7	3/7	2/13	0.65
-0.185	4/5	3/7	3/7	3/13	0.57
-0.073	4/5	3/7	4/7	3/13	0.57
-0.032	4/5	3/7	4/7	4/13	0.49
-0.028	4/5	3/7	4/7	5/13	0.42
0.004	4/5	3/7	4/7	6/13	0.37
0.022	4/5	4/7	4/7	6/13	0.34
0.032	4/5	4/7	5/7	6/13	0.34
0.235	4/5	4/7	5/7	7/13	0.26
0.241	4/5	4/7	5/7	8/13	0.23
0.250	4/5	4/7	5/7	9/13	0.23
0.279	4/5	4/7	5/7	10/13	0.23
0.504	4/5	5/7	5/7	10/13	0.09
0.510	4/5	5/7	5/7	11/13	0.13
0.597	4/5	5/7	5/7	12/13	0.21
0.799	4/5	5/7	6/7	12/13	0.21
0.815	5/5	5/7	6/7	12/13	0.29
1.049	5/5	5/7	7/7	12/13	0.29
1.098	5/5	6/7	7/7	12/13	0.14
1.103	5/5	6/7	7/7	13/13	0.14
1.645	5/5	7/7	7/7	13/13	0.00

3) Choose the largest number appearing in the final column; that is $D(5,7,7,13) = 0.65$.

The critical level $\hat{\alpha}$ is defined as the smallest significance level at which the null hypothesis would be rejected for the given observations. The second table shows that if the four samples were from the same population, $P[D(5,7,7,13) \leq .6] = 0.701271$ and $P[D(5,7,7,13) \leq .7] = 0.827140$. Therefore, $P[D(5,7,7,13) > .6] = 0.298729$ and $P[D(5,7,7,13) > .7] = 0.172860$. Since, in this example, $D(5,7,7,13) = 0.65$, $\hat{\alpha}$ would be between 0.18 and 0.30.

V. APPLICATION. The test found application in a bomblet study conducted at the Ballistic Research Laboratory. In this study, bomblets were supplied by three different vendors. Fifteen bomblets from one vendor were filled with a high explosive; thirty bomblets supplied by each of the other two vendors contained an inert substance. All bomblets were subjected to the same field test, meant to simulate the dispensing of a bomblet by a type of munition; this resulted in an out-of-round characteristic. A measurement was then taken in order to determine the degree of ovalness. The experimenter wanted to establish whether bomblets from all three vendors, filled with two different materials, achieved approximately the same degree of ovalness.

This is tantamount to determining whether three independent samples come from the same population. Thus, we were able to state a hypothesis, evaluate the two-sided test statistic, compare it with the tabled values, and determine whether or not to reject the hypothesis. The data and analysis follow.

H_o: All three samples of bomblets achieve the same degree of ovalness.

H_1: At least one sample of bomblets achieves a significantly different degree of ovalness.

The data (measurement of ovalness) are as follows:

Vendor I	0.002	0.010	0.036	0.004	0.006
	0.038	0.023	0.003	0.013	0.002
	0.013	0.010	0.013	0.008	0.007
Vendor II	0.007	0.004	0.003	0.010	0.009
	0.005	0.011	0.002	0.011	0.007
	0.010	0.011	0.023	0.009	0.029
	0.024	0.004	0.020	0.011	0.019
	0.004	0.014	0.011	0.008	0.009
	0.014	0.011	0.016	0.007	0.013
Vendor III	0.007	0.001	0.072	0.013	0.001
	0.025	0.007	0.010	0.028	0.004
	0.002	0.014	0.020	0.041	0.018
	0.011	0.010	0.007	0.035	0.001
	0.015	0.008	0.008	0.007	0.007
	0.047	0.015	0.054	0.010	0.020

The test statistic takes on the value D(15,30,30) = 0.2333. Quantiles of D(15,30,30) are not tabulated, and so the computer program which generated the table was rerun to obtain exact values. The results showed that $P[D(15, 30, 30) \leq .2] = 0.148915$ and $P[D(15, 30, 30) \leq .3] = 0.650568$. Therefore, the critical level $\hat{\alpha}$ would be between 0.35 and 0.86.

The data do not suggest rejecting the null hypothesis at a moderate significance level. Therefore, the experimenter concluded that the three bomblet samples achieved approximately the same degree of ovalness.

REFERENCES

Becker, M.P. and Taylor, M.S., Quantiles of Multi-Sample Smirnov Type Statistics, Technical Report ARBRL-TR-02331, Ballistic Research Laboratory, June 1981.

Birnbaum, Z.W. and Hall, R.A., "Small Sample Distributions for Multi-Sample Statistics of the Smirnov Type," The Annals of Mathematical Statistics, Vol. 31, No. 3, September 1960.

Conover, W.J., Practical Nonparametric Statistics, John Wiley & Sons, Inc., New York, 1971.

Hodges, J.L., Jr., "The Significance Probability of the Smirnov Two-Sample Test," Arkiv für Matematik, Vol. 3, 1957.

ANALYZING n SAMPLES OF 2 OBSERVATIONS EACH

J. R. Knaub, Jr., L. M. Grile, G. Petet
U. S. Army Logistics Center, Ft. Lee,VA

ANALYZING n SAMPLES OF 2 OBSERVATIONS EACH

I. Introduction.

When data are scarce, it is common to combine small samples from a number of sources considered to be reasonably similar. When sample sizes are extremely small, testing the assumption of similarity of sources is often only attempted by subjective means. This paper provides a method to add quantitative risk assessments to the study of this assumption, using two observations per sample.

In addition to general compatibility testing of the sources using modified versions of Westenberg's Interquartile Range Test, and the Westenberg-Mood Median Test, a new hypothesis test has been developed to aid in identifying whether one (or more) of the sources of data provides a substantially larger or smaller set of values due to its underlying population.

Because the probabilistic risk assessments provided here address a situation so commonly found in analyzing military operations as well as in test and evaluation, details are provided to simplify the implementation of this methodology. A major goal here has been that power analyses be described in terms meaningful to the user and the decision maker. The new hypothesis test makes use of simulation-aided power analyses.

The tests for general comparability, using modified Westenberg tests, were first introduced in reference 2. The FORTRAN code for these tests is given here in Appendix I. This program involves straight forward binomial probabilities. The first format statement explains the variables and the null and alternative hypotheses. It is written in terms of the interquartile range of the combined sample, but could easily be written in terms of being above or below the median of the combined sample. The null hypothesis (H_0) and the alternative (H_1) are repeated here. "H_0: Each sample has at least 100 x RA percent chance of having one observation inside and one observation outside the interquartile range. H_1: Each sample has at least 100 x RB percent chance of having both observations fall together either inside or outside." The null hypotheses for the Westenberg tests thus indicate a general compatibility among the data sources. If a set of data made up of pairs of observations from a number of sources appear to be reasonably homogeneous as judged by these tests, however, it may still be that one, or perhaps a few pairs of observations may have been drawn from a source very different from the underlying population for the majority of the pairs of observations. Therefore, a new test is needed due to the inadequacy of the modified Westenberg tests to discern such a situation.

In order to determine whether a pair of observations may have values appreciably larger, or smaller than the other observations, the probability of having both members of that pair of observations be among the largest, or smallest in the combined sample should be investigated. This is accomplished here in the Length of Initial Run (LIR) Test. In this test, the larger value in each pair of observations is labeled "A," while the smaller of the observations from that source is labeled "B." If no pair of observations is drawn from an underlying population considerably different from the others (in particular if no pair of observations is drawn from an underlying population whose location is considerably larger, or smaller than the others) then the length of the run of A's in the combined sample before the first B and the length of the run of B's before the first A should not be too short. (Note that run lengths can vary from one to n.)

II. Mathematical Theory of LIR Test.

Under the null hypothesis, that all pairs of observations were drawn from the same or identical populations, the length of the initial run of A's is identically distributed as the length of the initial run of B's. This discussion will only be couched in terms of the initial run of A's.

The probability of having an "A" as the largest value in the combined sample is unity. The probability of having an "A" in the second largest position, under H_0, is the number of observations not included in the sample pair that the first A came from (2n-2), divided by the total remaining number of observations available (2n-1). The probability of having a third A in a row is the previous probability multiplied by the number of observations not included in either sample pair that the first two A's came from (2n-4), divided by the total remaining number of observations available (2n-2). Therefore, the probability (under H_0) of having at least r of the A's before the first B is:

$$\left(\frac{2n-2}{2n-1}\right)\left(\frac{2n-4}{2n-2}\right)\left(\frac{2n-6}{2n-3}\right)\cdots\cdot\left(\frac{2n-2r+2}{2n-r+1}\right)$$

If exactly r of the A's preceed the first B, the probability of this occurrence (under H_0) is the above expression multiplied by the probability that the r+1th largest value is a B. This would be the number of B values whose corresponding A value is among the members of the initial run (r), divided by the total remaining number of observations (2n-r). Simplifying, therefore, the probability (under H_0) of having an initial run of length r is:

$$\frac{(2n-2)!!(2n-r-1)!r}{(2n-2r)!!(2n-1)!}$$, where $k!! = k(k-2)(k-4)\cdots\cdot$ (stopping at 2 if k is even, or 1 if k is odd).

This further simplifies to

$$\frac{(n-1)!2^{n-1}\,(2n-r-1)!r}{(n-r)!2^{n-r}\,(2n-1)!} = \frac{(n-1)!2^{r-1}\,(2n-r-1)!r}{(n-r)!\,(2n-1)!}$$

In general, if N is the number of observations per sample and n is the number of samples, then the probability of an initial run of length r is:

$$\frac{(n-1)!N^{r-1}\,(Nn-r-1)!(N-1)r}{(n-r)!\quad(Nn-1)!}$$

Simulations could be used for alternative hypotheses and for irregular numbers of observations. This paper, however, is concerned with N=2 observations per sample.

Simulations (see Appendix III) were used to determine the relative frequency distribution of initial run lengths under various alternative hypotheses. Each alternative studied assumed one pair of observations to be taken from one underlying population and all others taken from a second underlying population, with a few exceptions for sensitivity study purposes. The accuracy of the simulations was examined in several manners. First, both underlying populations were set identical and the results compared to the frequency distribution for the null hypothesis. Agreement here demonstrated that the closed form solution for the null distribution is correct and also that the simulation was accurate to approximately three significant digits using 20,000 replications for the cases of interest shown in Appendix II, also under H_0. However, under any alternative hypothesis, accuracy of the simulation is degraded due to the fact that the distributional forms which the pairs of observations are being drawn from are not exactly what they have been represented to be. Table I, however, provides a set of chi-square "poorness" of fit tests which show that, in the case investigated there, the distribution is almost exactly as was represented. (Similar results were obtained using other distributions.) Table II is used to demonstrate the small differences in resulting output when inputs are varied to degrees that were found unlikely to actually occur. (Note that the differences found in Table II were of only approximately the same magnitude as in Table I.) From this, it is generally concluded that only two significant digits should be used from the relative and cumulative relative frequency outputs.

In addition to the type of validation shown above, the simulation results were compared to a closed form solution for the probability of a run of length one when one pair of observations is drawn from one distribution and all others from a second distribution. In order for there to be a run of length one, both observations from the same pair must be the two largest (or smallest) observations in the combined sample. Therefore, if p is the probability of a run of length one, and we are investigating the initial run of A's, and only one pair of observations is drawn from one distribution with all others drawn from a second distribution,

$$p = \int_{x=-\infty}^{\infty} g_{2,m,m}(x) \int_{t=x}^{\infty} g_{1,2,1}(t)\,dt\,dx$$

$$+ \frac{1}{m-1} \int_{x=-\infty}^{\infty} g_{1,2,2}(x) \int_{t=x}^{\infty} g_{2,m,m-1}(t)\,dt\,dx \, .$$

TABLE Ia

	True N(12,1)		VS		Random Nos. Generated From A N(12,1)		
CELL #	#OBS	CELL#	#OBS	CELL#	#OBS	CELL#	#OBS
1	370	26	410	51	417	76	405
2	375	27	428	52	402	77	403
3	402	28	382	53	398	78	405
4	369	29	413	54	381	79	360
5	412	30	386	55	447	80	361
6	376	31	400	56	385	81	393
7	407	32	396	57	393	82	388
8	424	33	426	58	375	83	408
9	413	34	386	59	404	84	381
10	381	35	433	60	401	85	388
11	377	36	380	61	419	86	404
12	380	37	407	62	371	87	388
13	394	38	418	63	414	88	408
14	381	39	393	64	400	89	421
15	382	40	407	65	415	90	411
16	364	41	370	66	396	91	422
17	388	42	428	67	415	92	397
18	398	43	411	68	429	93	395
19	423	44	395	69	389	94	398
20	427	45	405	70	424	95	394
21	381	46	347	71	417	96	414
22	398	47	413	72	406	97	429
23	393	48	445	73	439	98	392
24	394	49	415	74	404	99	420
25	370	50	392	75	392	100	417

THE CHI-SQUARE VALUE FOR A N(12,1)
USING 40000 GENERATED RANDOM NUMBERS
TESTED AGAINST A N(12,1)
IS 94.56

TABLE Ib

	True N(12,1)		VS		Random Nos. Generated From A N(12,0.95)		
CELL #	#OBS	CELL#	#OBS	CELL#	#OBS	CELL#	#OBS
1	282	26	385	51	449	76	437
2	295	27	426	52	415	77	390
3	317	28	430	53	418	78	349
4	344	29	403	54	402	79	387
5	343	30	424	55	468	80	410
6	372	31	409	56	395	81	394
7	361	32	424	57	426	82	399
8	380	33	424	58	385	83	393
9	390	34	409	59	435	84	392
10	402	35	433	60	417	85	389
11	378	36	417	61	444	86	401
12	362	37	412	62	387	87	409
13	370	38	445	63	413	88	406
14	396	39	424	64	454	89	394
15	382	40	403	65	399	90	414
16	383	41	426	66	439	91	375
17	362	42	401	67	435	92	368
18	369	43	458	68	416	93	382
19	412	44	399	69	432	94	366
20	436	45	432	70	441	95	383
21	438	46	387	71	424	96	366
22	385	47	418	72	451	97	367
23	388	48	460	73	418	98	335
24	416	49	442	74	398	99	354
25	396	50	415	75	414	100	300

THE CHI-SQUARE VALUE FOR A N(12,0.95)
USING 40000 GENERATED RANDOM NUMBERS
TESTED AGAINST A N(12,1)
IS 302.45

TABLE Ic

True N(12,1)		VS		Random Nos. Generated From A N(12,1.05)			
CELL #	#OBS	CELL#	#OBS	CELL#	#OBS	CELL#	#OBS
1	495	26	418	51	402	76	393
2	443	27	366	52	371	77	384
3	448	28	407	53	388	78	398
4	445	29	378	54	372	79	398
5	412	30	386	55	398	80	359
6	442	31	367	56	396	81	342
7	444	32	427	57	366	82	385
8	428	33	356	58	372	83	401
9	399	34	423	59	370	84	392
10	395	35	352	60	383	85	380
11	396	36	386	61	387	86	398
12	399	37	423	62	400	87	396
13	378	38	385	63	358	88	402
14	384	39	362	64	368	89	422
15	354	40	385	65	422	90	417
16	404	41	365	66	385	91	433
17	411	42	403	67	376	92	433
18	411	43	394	68	421	93	418
19	413	44	387	69	381	94	416
20	383	45	360	70	382	95	427
21	373	46	345	71	404	96	424
22	392	47	392	72	411	97	473
23	392	48	427	73	392	98	479
24	368	49	391	74	431	99	472
25	391	50	380	75	391	100	566

THE CHI-SQUARE VALUE FOR A N(12,1.05)
USING 40000 GENERATED RANDOM NUMBERS
TESTED AGAINST A N(12,1)
IS 277.28

TABLE IIa

Input distributions are N(12.00, 1.00) and N(10.00, 1.00). Number of samples from each distribution is 1 and 9 respectively. The random number seed for this run is 65557.

Test for the Length of the Initial Run of A's before the first B.

Number of replications: 20000

Length of Run:	Observed Frequency:	Relative Frequency:	Cumulative Frequency:
1	7275	0.363750	0.363750
2	4391	0.219550	0.583300
3	2980	0.149000	0.732300
4	2123	0.106150	0.838450
5	1534	0.076700	0.915150
6	857	0.042850	0.958000
7	517	0.025850	0.983850
8	246	0.012300	0.996150
9	67	0.003350	0.999500
10	10	0.000500	1.000000

TABLE IIb

Input distributions are N(12.00, 0.95) and N(10.00, 1.00). Number of samples from each distribution is 1 and 9 respectively. The random number seed for this run is 65557.

Test for the Length of the Initial Run of A's before the first B.

Number of replications: 20000

Length of Run:	Observed Frequency:	Relative Frequency:	Cumulative Frequency:
1	7406	0.370300	0.370300
2	4480	0.224000	0.594300
3	3018	0.150900	0.745200
4	2134	0.106700	0.851900
5	1419	0.070950	0.922850
6	806	0.040300	0.963150
7	469	0.023450	0.986600
8	203	0.010150	0.996750
9	55	0.002750	0.999500
10	10	0.000500	1.000000

TABLE IIc

Input distributions are N(12.00, 1.05) and N(10.00, 1.00). Number of samples from each distribution is 1 and 9 respectively. The random number seed for this run is 65557.

Test for the Length of the Initial Run of A's before the first B.

Number of replications: 20000

Length of Run:	Observed Frequency:	Relative Frequency:	Cumulative Frequency:
1	7141	0.357050	0.357050
2	4273	0.213650	0.570700
3	2969	0.148450	0.719150
4	2119	0.105950	0.825100
5	1611	0.080550	0.905650
6	958	0.047900	0.953550
7	556	0.027800	0.981350
8	284	0.014200	0.995550
9	78	0.003900	0.999450
10	11	0.000550	1.000000

where $g_{1,2,k}(x)$ is the distribution of the $k^{\underline{th}}$ order statistic out of 2 observations in distribution 1 (distribution 1 is, in general, the distribution of larger location if the initial run of A's is being investigated); and $g_{2,m,k}$ is the distribution of the $k^{\underline{th}}$ order statistic out of m for distribution 2. Since there are n-1 pairs of observations taken from distribution 2, m = 2(n-1) = 2n-2, and therefore:

$$\frac{1}{m-1} = \frac{1}{2n-2-1} = \frac{1}{2n-3}$$

The expression for p is based on the fact that if both observations of a given pair have larger values than any other observation in the combined sample, then the B value associated with the largest A has to be larger than all other 2n-2 values.

Once again, use of this validation technique supported the conclusion that two significant digits should be used in the results.

If $f_1(x)$ is the density function for distribution 1, and $f_2(x)$ for distribution 2, an approximation can be made for p when the number of observations drawn from each distribution are equal, or nearly equal, and very small. This approximation will be very poor unless the assumptions are enforced. In general, however, the calculations are much easier than those in the previous expression. In this case the approximation is as follows:

$$\phi = \int_{x=-\infty}^{\infty} f_2(x) \int_{t=x}^{\infty} f_1(t)dtdx; \quad \Phi = \phi/(1-\phi)$$

$$p \approx \left(\frac{\nu\Phi}{\nu\Phi+2n-\nu}\right)\left(\frac{\Phi}{(\nu-1)\Phi+2n-\nu}\right) + \left(\frac{2n-\nu}{\nu\Phi+2n-\nu}\right)\left(\frac{1}{\nu\Phi+2n-\nu-1}\right)$$

where ν is the number of observations taken from the first distribution. ϕ is the probability that if one observation were drawn from each of the two distributions, the observation from distribution 1 would have a larger value. The approximation is therefore a weighted counting procedure which does not fully account for the shapes of the true distributions of interest.

The tables of Appendix II provide power information for a variety of cases when the number of samples (of size two each) is 5, 10, 20, and 50. The alternative hypotheses could have more than two underlying distributions (up to n) but two are sufficient to illustrate what is being investigated here. The importance of this test is to determine the likelihood of having one (or possibly more than one) pair of observations drawn from an underlying population which is substantially different from the underlying population from which the rest of the observations were drawn.

If sample sizes are large, it is possible to reject H_0 when the truth is close enough to H_0 for practical purposes, unless a specific H_1 is used for a power analysis. When sample sizes are small, as in the case here, one may fail to reject H_0 when the truth is not close to H_0. Power analyses would help by completing the quantification of the problem. Without a power analysis, a hypothesis test is only half completed. Null and alternative hypotheses work in pairs analogously to confidence limits. In the present situation, a power analysis is very important due to the unconventional nature of the problem. (It is interesting to note that when the number of samples [of size two each] increases, the power level at a given significance level remains apparently approximately constant [see Appendiix II].)

This paper makes use of a much neglected application for simulation. Simulation can be used as a check for a closed form solution when the development of such a solution was subtle, in addition to handling situations where a closed form solution is difficult if not impossible.

III. Example.

Suppose that ten processes (or items of equipment, etc.) are to be examined for a certain trait and that it is expected that they will all behave similarly for that trait. Also, suppose that the expense involved in studying those processes is great, or that for some other practical reason, the number of observations per process (item, scenario, etc.) must be kept extremely small. If two observations each are used, a combined sample size of 20 is obtained. Whether or not these samples should be combined would then be open to examination. In addition to any subjective arguments, the modified Westenberg tests and the LIR test should be applied to assist in this examination. Suppose (for use in the LIR test) that the larger value in each pair of observations is labelled "A," and the smaller values labelled "B." Subscripts "1"-"10" could be used to denote which process is being represented. (This will be used in the modified Westenberg tests.) Suppose that when the values for each of these observations are ranked from largest to smallest, the following result is obtained:

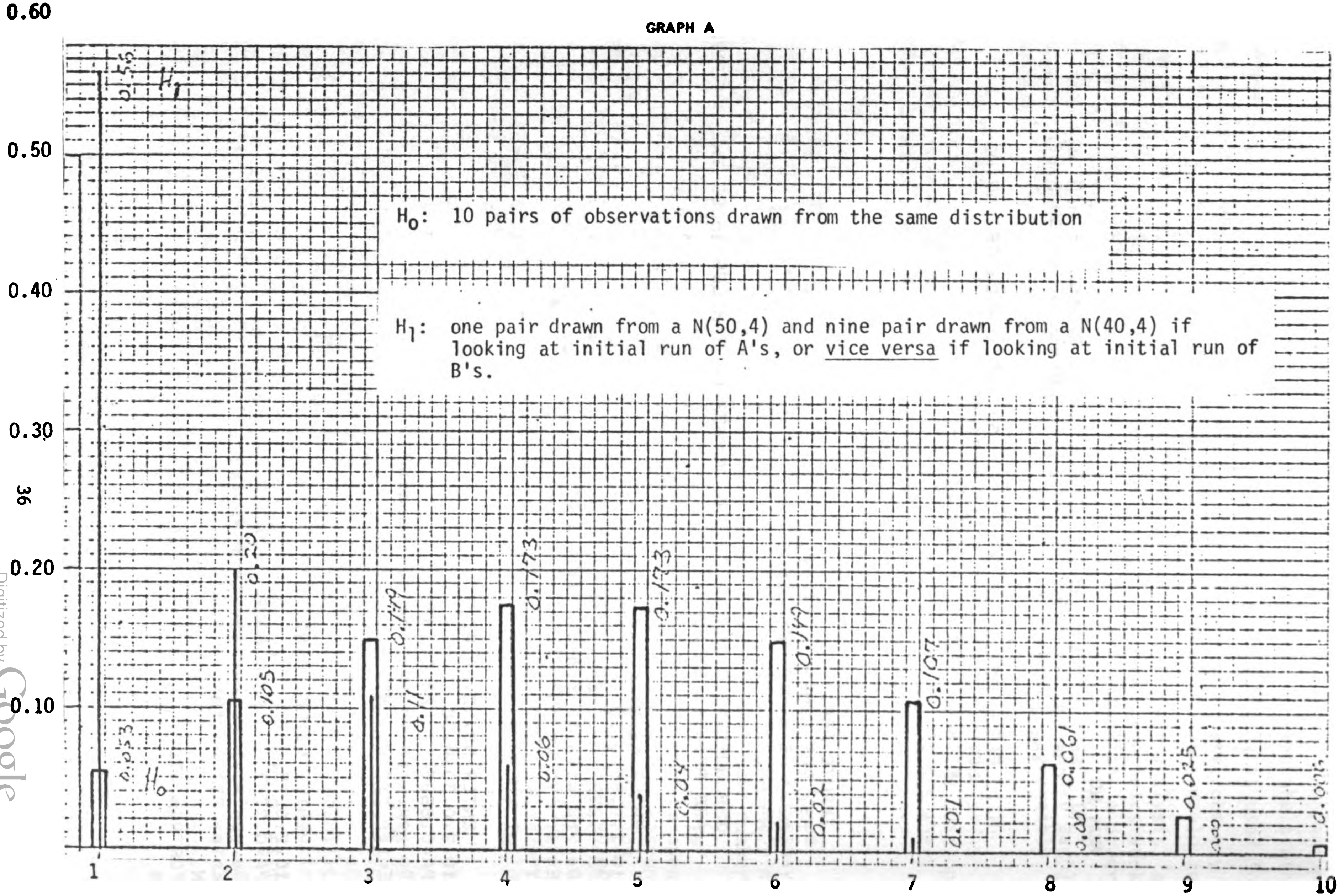
$$A_8\ B_8\ A_7\ A_2\ A_1 \,\vdots\, A_4\ A_5\ A_6\ A_{10}\ A_3 \,\Big|\, B_2\ A_9\ B_7\ B_4\ B_1 \,\vdots\, B_6\ B_{10}\ B_3\ B_9\ B_5$$

The number of times that a pair of observations are both found on the same side of the median, NZ, is 2. Also, the number of times that a pair of observations are both found either inside or outside the interquartile range is NZ=2. Table III shows results taken from the program of Appendix I. PA is the probability level of the test associated with H_0, and PB is the probability level associated with H_1. An examination of this table shows that there is no good reason, based on this ranked data and aside from subjective arguments, to conclude that these samples should not be combined, but actually there is good reason to conclude that such a combination is advisable. However, the LIR test can be used to show that even though general compatibility appears evident, the

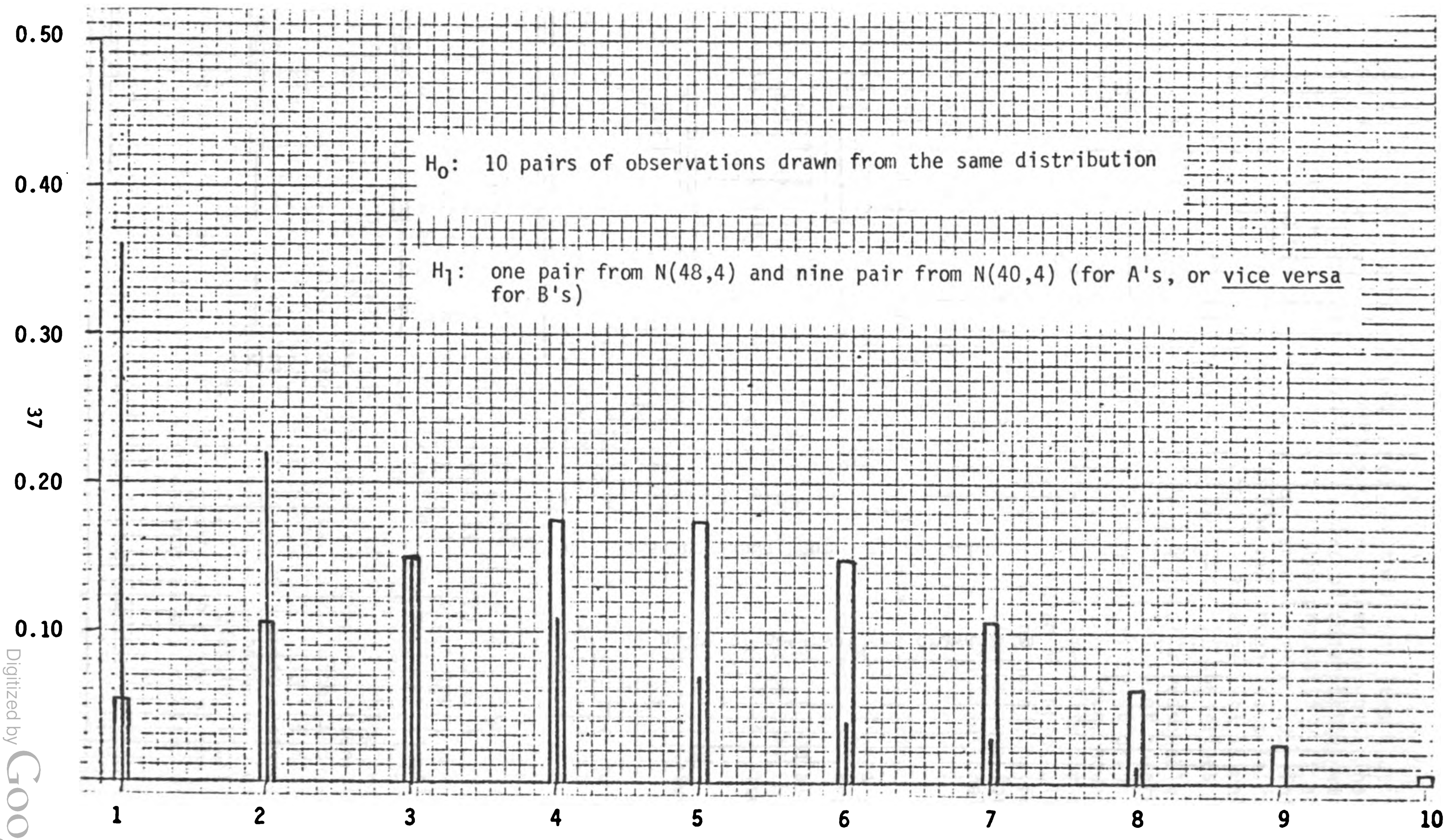
TABLE III

```
RUN WC
  HO:  EACH SAMPLE HAS AT LEAST 100xRA% CHANCE OF HAVING
       ONE OBSERVATION INSIDE AND ONE OBSERVATION OUTSIDE
        THE INTERQUARTILE RANGE.

  H1:  EACH SAMPLE HAS AT LEAST 100xRB% CHANCE OF HAVING
       BOTH OBSERVATIONS FALL TOGETHER
       EITHER INSIDE OR OUTSIDE




  INPUTS ARE:
  NS, THE NUMBER OF SAMPLES
  NZ, THE NUMBER OF ZEROES
  RA AND RB

  THE NUMBER OF ZEROES IS THE NUMBER OF SAMPLES WHOSE TWO OBSERVATIONS
  ARE FOUND TOGETHER


  ENTER NS,NZ,RA,RB
10,2,0.95,0.50
PA=0.2262      PB=0.1875
Do you wish to run the test again?
Enter "Y" or "N".
Y


  ENTER NS,NZ,RA,RB
10,2,0.50,0.50
PA=0.9688      PB=0.1875
Do you wish to run the test again?
Enter "Y" or "N".
Y


  ENTER NS,NZ,RA,RB
10,2,0.75,0.25
PA=0.7627      PB=0.6328
Do you wish to run the test again?
Enter "Y" or "N".
N
FORTRAN STOP
$
```

GRAPH A

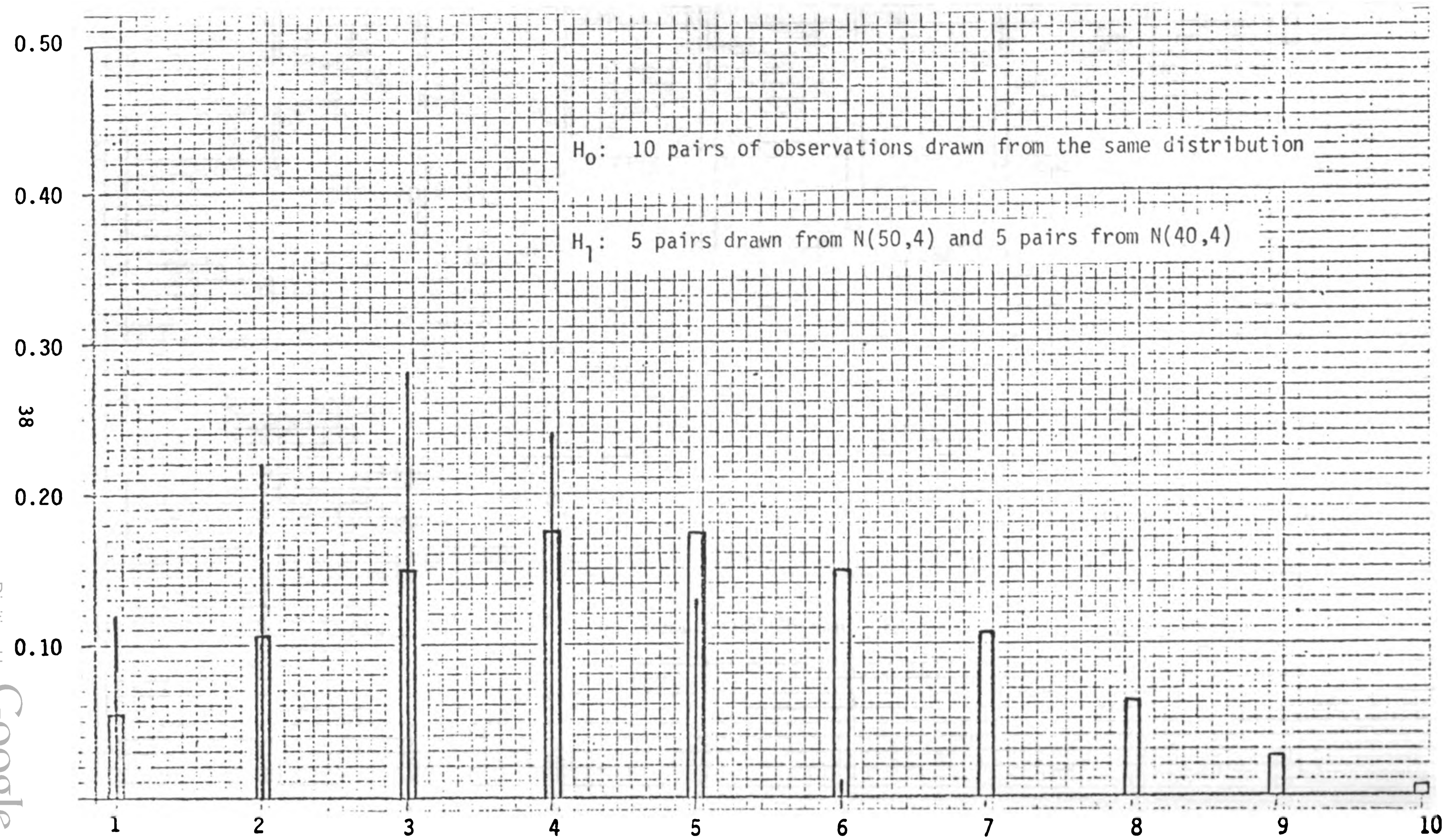

H₀: 10 pairs of observations drawn from the same distribution
H₁: one pair drawn from a N(50,4) and nine pair drawn from a N(40,4) if looking at initial run of A's, or vice versa if looking at initial run of B's.
0.60
0.50
0.40
0.30
0.20
0.10
1
2
3
4
5
6
7
8
9
10
0.55
0.053
0.20
0.105
0.11
0.173
0.06
0.04
0.107
0.01
0.061
0.025

GRAPH B



H_0: 10 pairs of observations drawn from the same distribution
H_1: one pair from N(48,4) and nine pair from N(40,4) (for A's, or vice versa for B's)
0.50
0.40
0.30
0.20
0.10
1
2
3
4
5
6
7
8
9
10

GRAPH C



H_0: 10 pairs of observations drawn from the same distribution
H_1: 5 pairs drawn from N(50,4) and 5 pairs from N(40,4)
0.50
0.40
0.30
0.20
0.10
1
2
3
4
5
6
7
8
9
10

initial run of A's is extremely short (one) and this indicates that sample 8 may provide values too large in comparison to the other values. The LIR test shows that if the null hypothesis (that all pairs of observations were drawn from the same underlying distribution) is true, the probability of an initial run of length one is 0.053 when the number of samples (n) is 10. This is the probability of a Type I error. However, if the null hypothesis is "accepted" here, the probability of a Type II error (for any alternative) is unity since no amount of evidence against H_o would then suffice. If the values of the observations here suggest that graph A is the result of a reasonable alternative hypothesis, then the power against that alternative is of interest and is approximately 0.56. Against the alternative used to arrive at graph B, the power is 0.36, and the power from graph C would be 0.12. Note, however, that the alternative shown here against which the power is lowest is for a situation better investigated by the tests of Appendix I, program WC, originally found in reference 2. Note also that even if all 20 observations here came from the same distribution, that is a small sample to use to determine the form of that distribution. Therefore, other alternatives involving distributional forms other than normality may be needed to complete the analysis.

Suppose B_8 and A_7 were not as shown, but exchanged in position. The values of NZ would remain the same, but the length of the initial run of A's would now be two. If H_o is rejected when the observed run length is 2, the probability of a Type I error is 0.158. Against the first, second, and third alternatives mentioned above, the probability of a Type II error would be 0.44, 0.64, and 0.88, respectively, and the power would be 0.76, 0.58, and 0.34, also respectively.

If a run of length 3 is considered, these figures are, respective to the order given above, 0.307, 0.24, 0.42, 0.66, 0.87, 0.73, and 0.62.

Considering the above, when n=10 it could be deemed reasonable to reject H_o when the run length is 2 or less, and "accept" it when it is 3 or more.

In conclusion, if some of the observations in the example given earlier are shifted in rank it may affect one or more or perhaps none of these tests. Also, if the modified Westenberg tests greatly discourage the combining of the samples, then the LIR test will probably not be very useful. In using the modified Westenberg tests, if it is not possible to divide the observations into groups of equal size (above and below the median and inside and outside of the interquartile range), then apply the tests shifting the observation(s) which are in question from one possible grouping to the other and average the results obtained. Finally, because the Bernoulli trials in the modified Westenberg tests used here are not truly independent, these tests are approximate. Also, some RA and RB values may be inappropriate. For example, in Table III, RA could never be 0.95; however, it is used to illustrate the strong conclusiveness of these particular results. A study of the null hypothesis indicates that for RA = 0.5, the binomial distribution used for this test has thicker "tails" than warranted. (The exact null distribution is described by

$$\Sigma\left\{\frac{(NS!/2^{NZ/2})^2}{(2NS)!/2^{NS}} \cdot \frac{NS!/(NS-NZ)!}{[(NZ/2)!]^2}\right\}.)$$

Also, the true null distribution is skewed to the right, but not appreciably when NS exceeds 50. For smaller numbers of samples, WB 1 from reference 2 is of use. The advantages in using program WC are that understandable alternative hypotheses can be shown for decision making, and a large number of sample pairs can be handled with very little computer time. All that is really needed is a table of the cumulative binomial distribution. This test should be considered "quick and dirty" as a preliminary to the implementation of the LIR test. The LIR test is an exact test and is easily and meaningfully applied.

APPENDIX I

FORTRAN CODE FOR Modified Westenberg Test
(Designed for n pairs of observations)

This program was referred to as "WC" in reference 2. Note that unlike the other modified Westenberg tests of reference 2, which are exact, this one is an approximate test.

```
      CHARACTER*4 ANSWER
      WRITE(6,1100)
1100  FORMAT(5X,'H0:  EACH SAMPLE HAS AT LEAST 100xRA% CHANCE',
     /  ' OF HAVING',/,9X,'ONE OBSERVATION INSIDE AND ONE'
     /  ' OBSERVATION OUTSIDE',/,9X,' THE INTERQUARTILE RANGE.',
     /  //,5X,'H1:  EACH SAMPLE HAS AT LEAST 100xRB% CHANCE',
     /  ' OF HAVING',/,9X,'BOTH OBSERVATIONS FALL TOGETHER',
     /  /,9X,'EITHER INSIDE OR OUTSIDE',//////,5X,'INPUTS ARE: ',
     /  /,5X,'NS, THE NUMBER OF SAMPLES',/,
     /  5X,'NZ, THE NUMBER OF ZEROES',
     /  /,5X,'RA AND RB',//,5X,'THE NUMBER OF ZEROES IS THE NUMBER OF',
     /  ' SAMPLES WHOSE TWO OBSERVATIONS',/,5X,'ARE FOUND TOGETHER')
1000  WRITE(5,999)
 999  FORMAT(//,5X,'ENTER NS,NZ,RA,RB')
      READ(5,*)NS,NZ,RA,RB
      I=NS/2
      J=NZ/2
      PA=0
      PB=0
      K=0
   2  X=I
      Y=K
      W=0.5*I
      IF(Y.GE.W)GOTO 100
   3  P=X
      P=P/(X-Y)
      T=0
   5  T=T+1
      P=P*(X-T)
      P=P/(X-Y-T)
      L=X-Y-T
      IF(L.GT.1)GOTO 5
      IF(K.GT.J)GOTO 6
      PB=PB+P*((RB**K)*((1-RB)**(I-K)))
      K=K+1
      IF(K.LE.J)GOTO 2
      K=K-1
   6  PA=PA+P*(((1-RA)**K)*(RA**(I-K)))
      K=K+1
      IF(K-I)2,2,2000
 100  Y=X-Y
      GOTO 3
2000  WRITE(6,2001)PA,PB
2001  FORMAT(1X,'PA=',F6.4,5X,'PB=',F6.4)
      WRITE(6,2002)
2002  FORMAT(T2,'Do you wish to run the test again?',/,
     /  T2,'Enter "Y" or "N".')
      READ(5,2003)ANSWER
2003  FORMAT(A)
      IF(ANSWER.EQ.'Y')GOTO 1000
2010  STOP
      END
```

APPENDIX II

POWER TABLES FOR LIR TEST

The distributions from which the input is to be drawn for each of the alternative hypotheses are shown here followed by histograms of the relative frequency distributions for these alternatives. In all cases here, both of the distributions from which the samples are drawn are of the same type but with parameters which differ in some respect. The samples could have been drawn from totally different distributional forms, and more than two such distributions could have been used (up to n); however, what is used here is sufficient to demonstrate this test under conditions which illustrate its usefulness.

In the case of the normal distribution, when standard deviations are the same, the symmetric nature makes the relative frequency distribution for the initial run of A's the same as that for the initial run of B's, if the number of pairs of observations from each of the two input distributions is interchanged. This is true in all cases where symmetric input distributions of equal variance are used. Also, when two symmetric input distributions with the same location, but unequal variances are used, this principle applies. Whenever this occurs, the output relative frequency distributions here are written in terms of the initial run of A's.

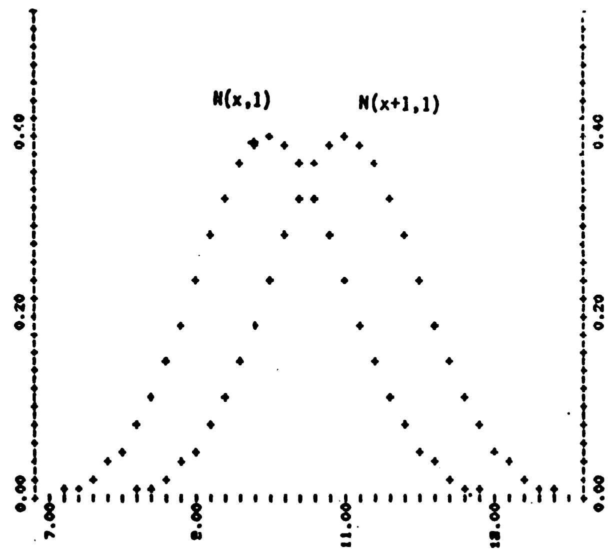
The Church-Harris-Downton (C-H-D) method of testing the probability of motor case rupture in missile testing, see reference 1, makes use of a statistic related to ϕ (shown in Section II here). This statistic is $(\mu_1 - \mu_2)/(\sigma_1^2 + \sigma_2^2)^{1/2}$, where the subscripts "1" and "2" refer to the two input distributions. For any of the tables involving two normal distributions, if μ_1, μ_2, σ_1, and σ_2 are changed such that the above statistic remains constant, then the output relative frequency distributions given here are applicable.

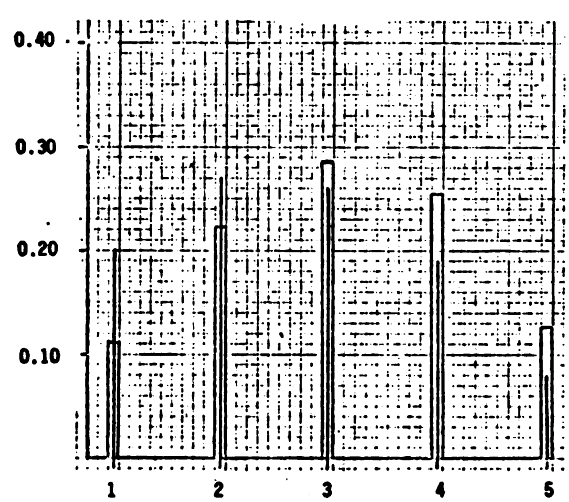
In the case of the gamma input distributions, whenever the β's (scale parameters) are both multiplied by the same factor, the output distributions are still applicable. For triangular input distributions, if all parameters are added to, subtracted from, divided or multiplied by the same number, the output distributions will not be changed.

The tables given here are for two normal distributions, two gamma distributions, two triangular distributions and finally, two beta distributions. The parameters were picked, in many cases, such that the power against the alternatives was approximately 0.5 when the significance level was approximately 0.1 to 0.15. This occurs for H_o, as shown in the following table:

# of samples of size 2 each	run length under H_o	(probability level under H_o)
5	1	~11%
10	2	~15%
20	3	~15%
50	{ 4	~10%
	5	~15%

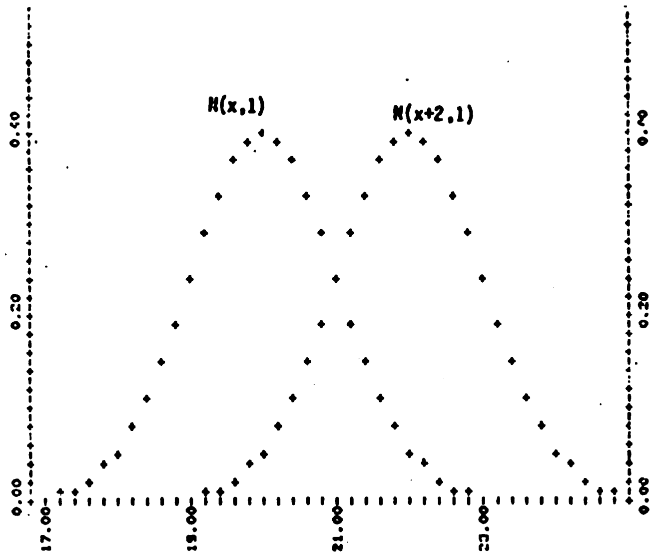
The following tables provide a variety of examples of alternative hypotheses and results obtained using them. It is hoped that this appendix is sufficient to provide a working knowledge of the power of this test to its users. When any specific alternative which the user is interested in investigating does not appear here, and the user does not wish to spend the time to get the programs of Appendix III to run at a convenient facility, it is hoped that the results can be interpolated from results provided here. Note that following each graph of the distributions from which the observations are hypothesized (H_1) to have been drawn, only relative frequency distributions are provided. The hollow bar graphs show the relative frequency distribution under H_o.

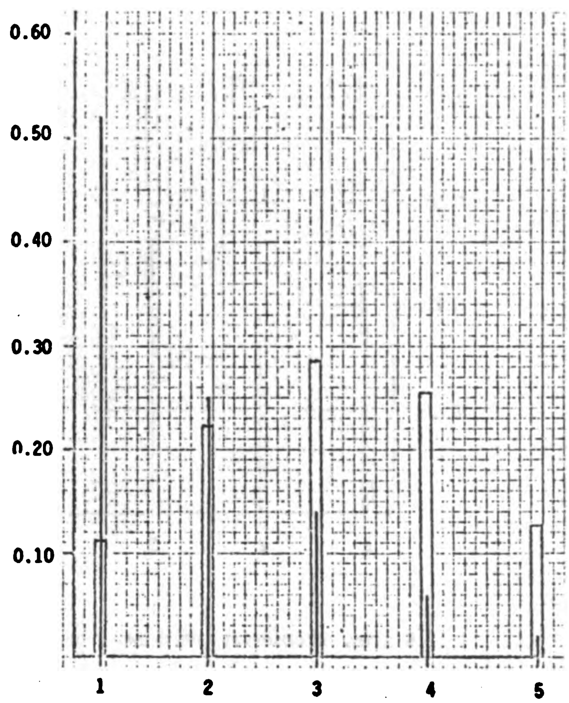




H_0: 5 pairs of observations drawn from the same distribution

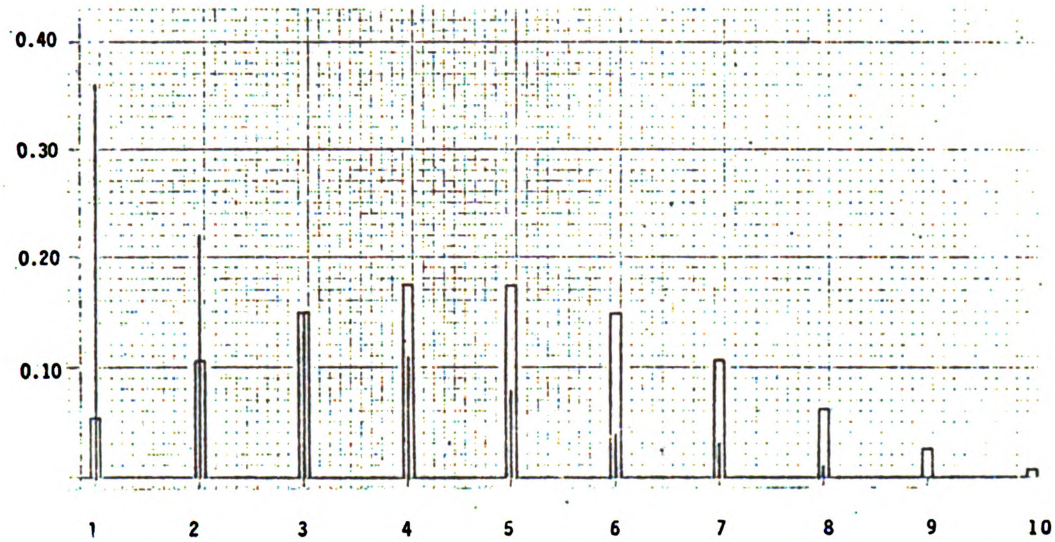
H_1: 4 pairs from N(x,1) and 1 pair from N(x+1,1)





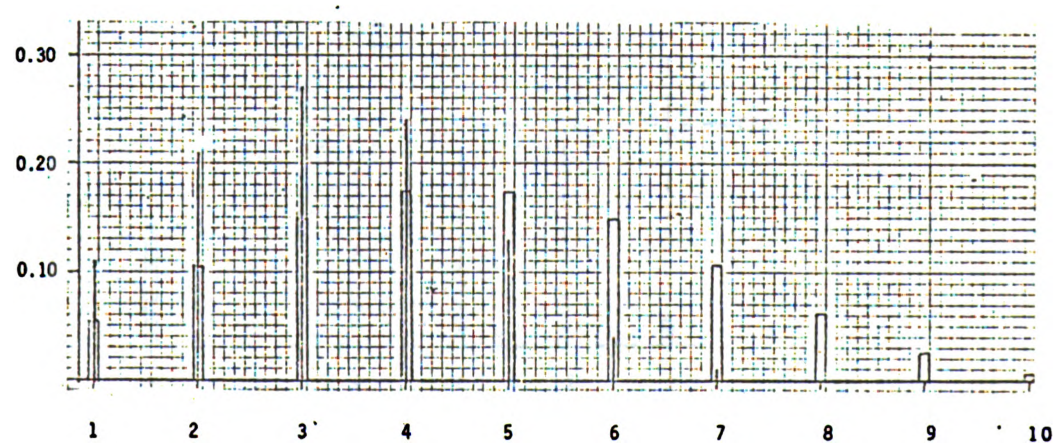
H_0: 5 pairs of observations drawn from the same distribution

H_1: 4 pairs from N(x,1) and 1 pair from N(x+2,1)



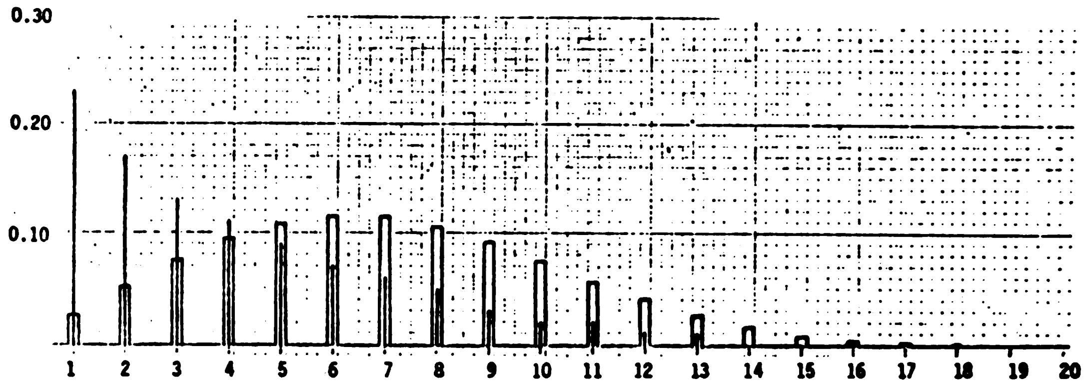
H_0: 10 pairs of observations drawn from the same distribution

H_1: 9 pairs from N(x,1) and 1 pair from N(x+2,1)



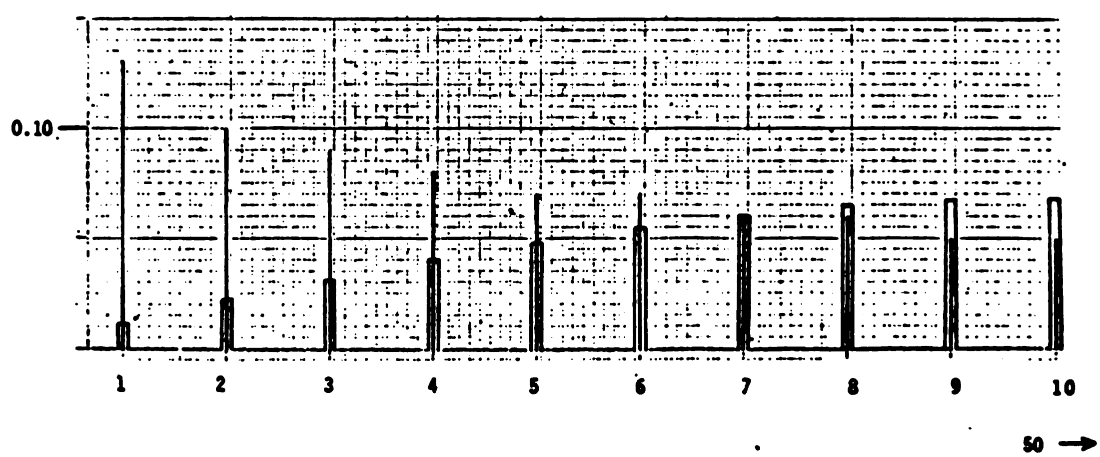
H_0: 10 pairs of observations drawn from the same distribution

H_1: 5 pairs from N(x,1) and 5 pairs from N(x+2,1)



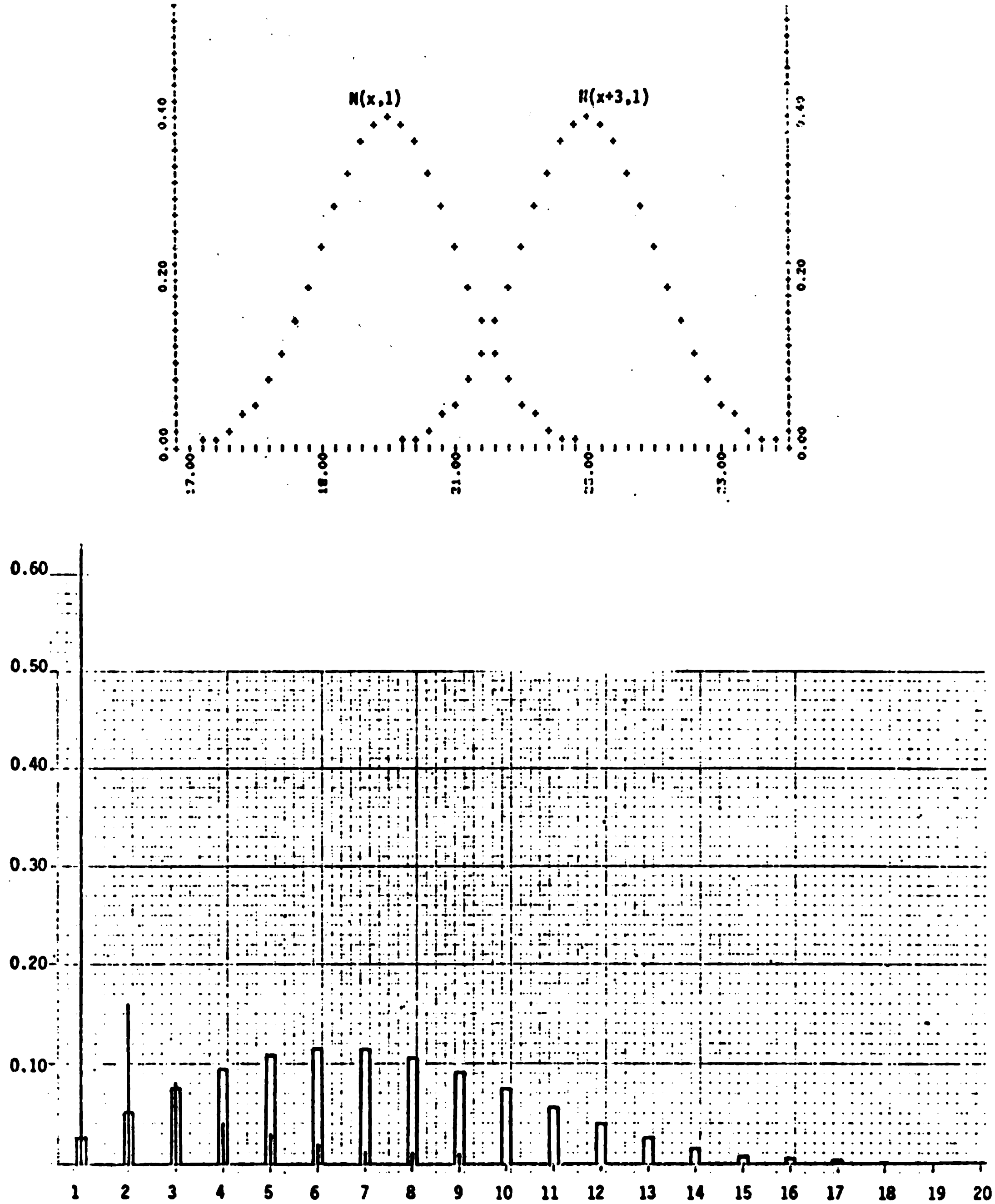
H_0: 20 pairs of observations drawn from the same distribution

H_1: 19 pairs from N(x,1) and 1 pair from N(x+2,1)



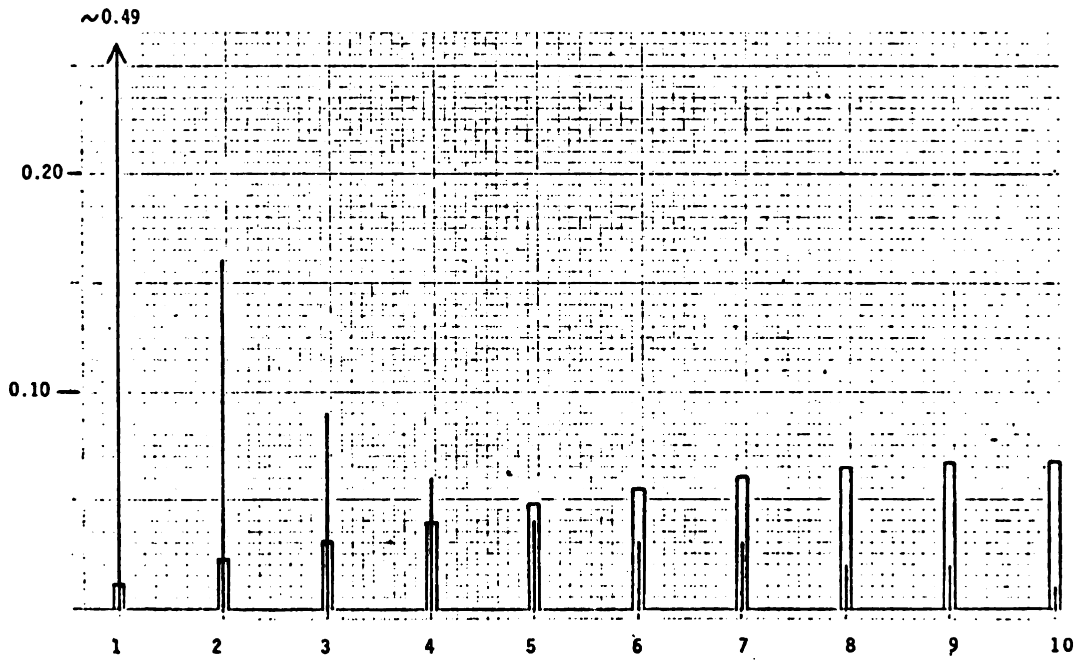
H_0: 50 pairs of observations drawn from the same distribution

H_1: 49 pairs from N(x,1) and 1 pair from N(x+2,1'



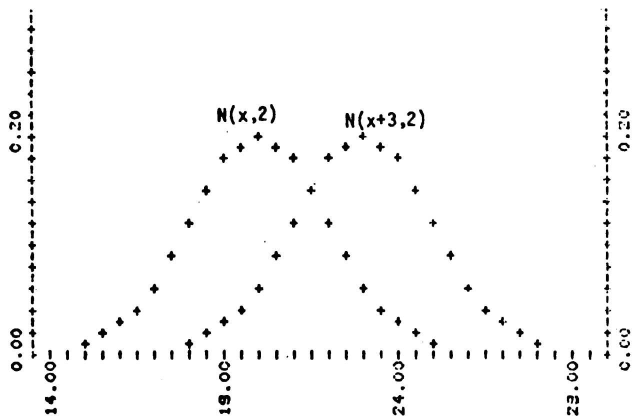
H_0: 20 pairs of observations drawn from the same distribution

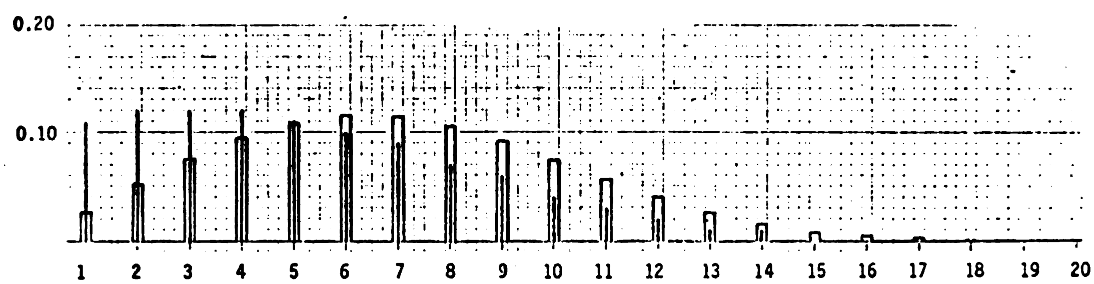
H_1: 19 pairs from N(x,1) and 1 pair from N(x+3,1)



H_0: 50 pairs of observations drawn from the same distribution

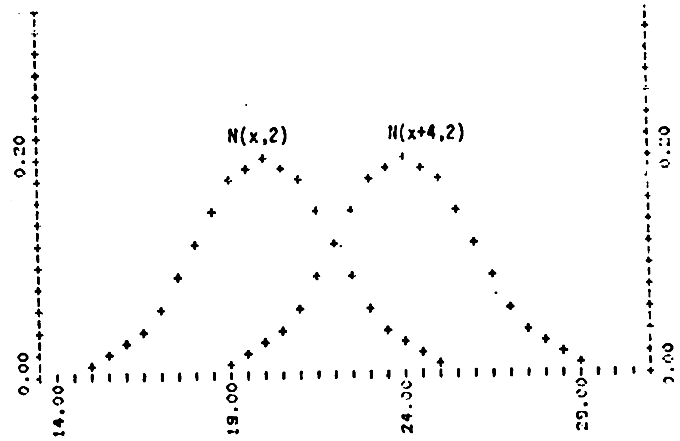
H_1: 49 pairs from N(x,1) and 1 pair from N(x+3,1)

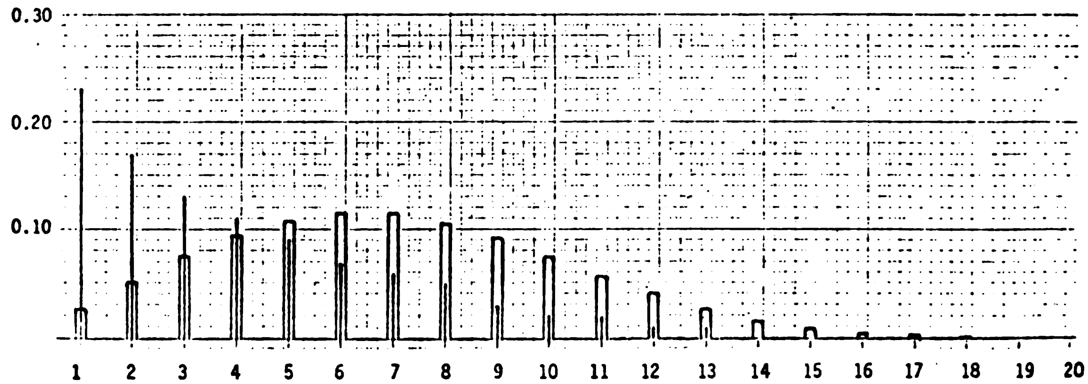




H_0: 20 pairs of observations drawn from the same distribution

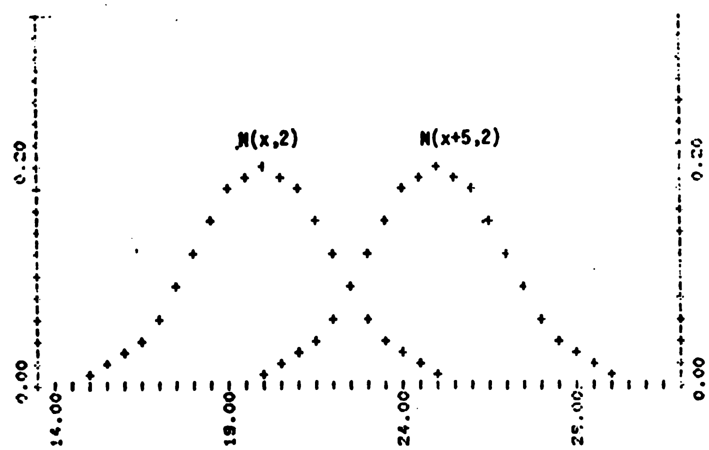
H_1: 19 pairs from N(x,2) and 1 pair from N(x+3,2)

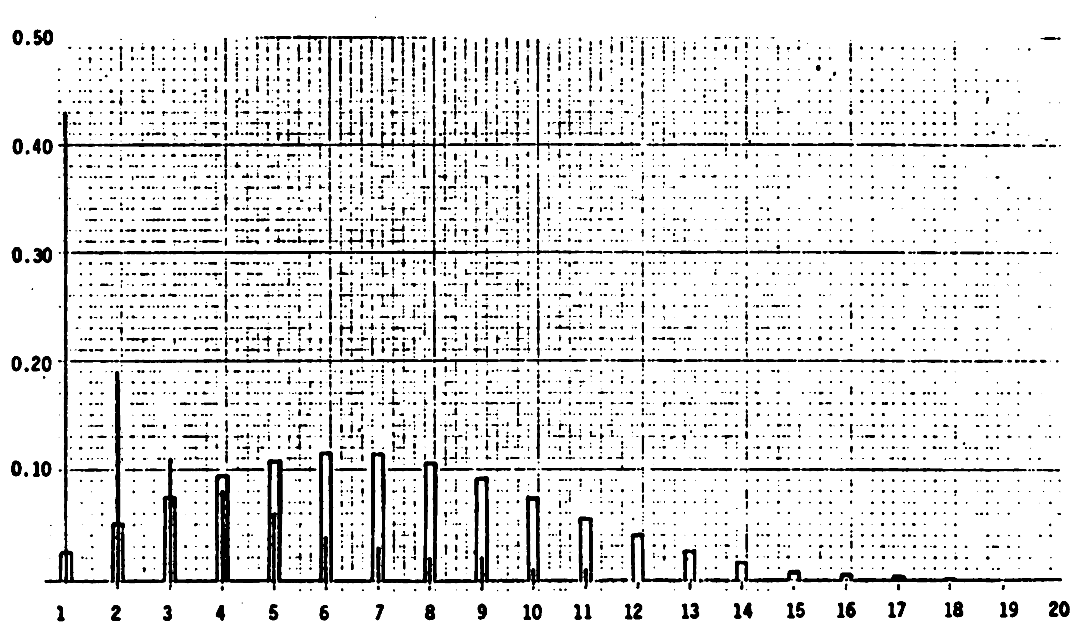




H_0: 20 pairs of observations drawn from the same distribution

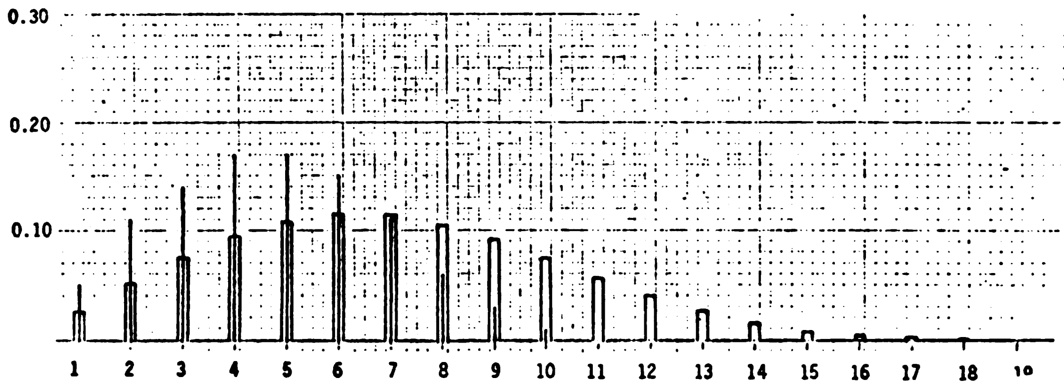
H_1: 19 pairs from N(x,2) and 1 pair from N(x+4,2)





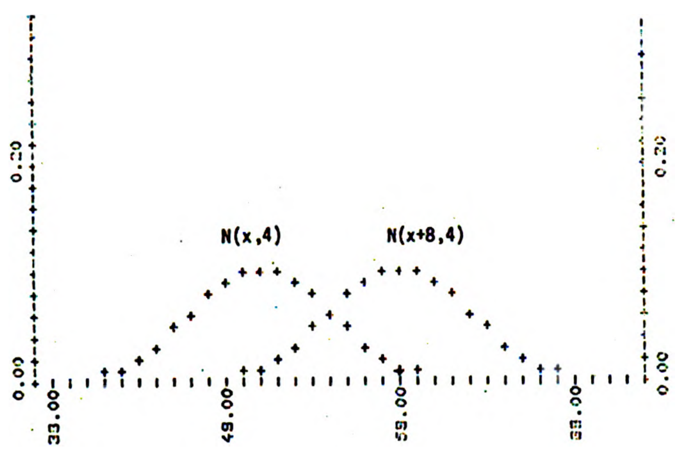
H_0: 20 pairs of observations drawn from the same distribution

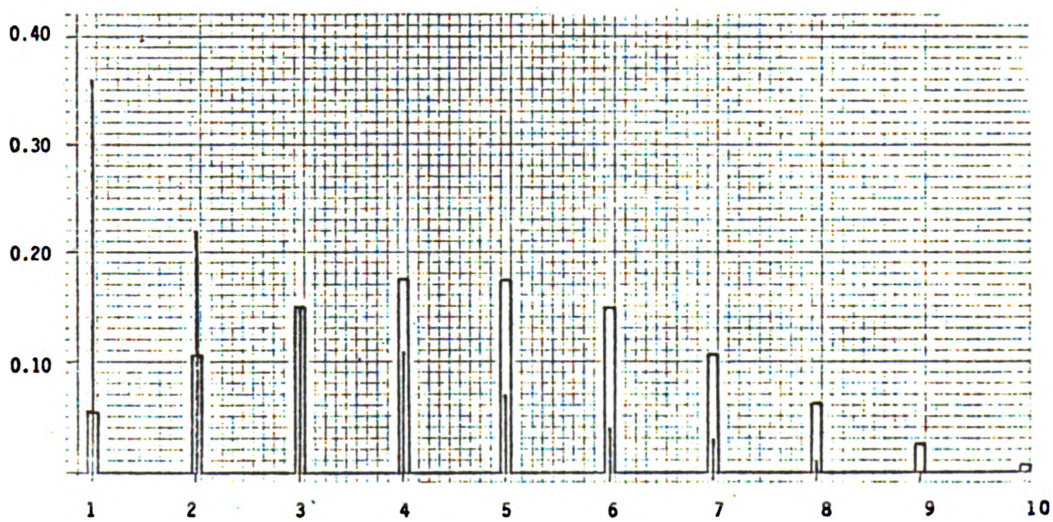
H_1: 19 pairs from N(x,2) and 1 pair from N(x+5,2)



H_0: 20 pairs of observations drawn from the same distribution

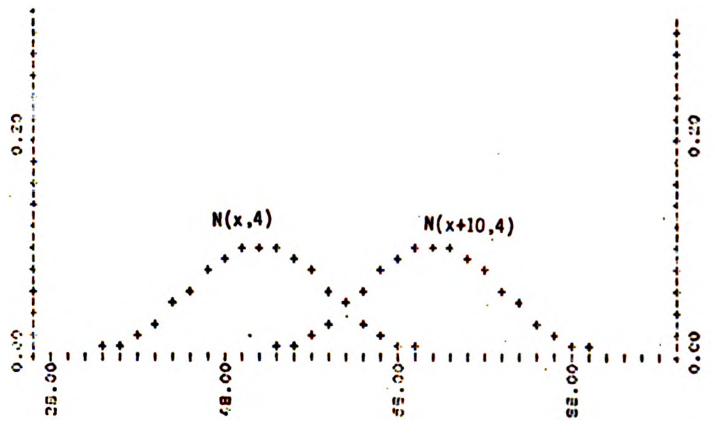
H_1: 10 pairs from N(x,2) and 10 pairs from N(x+5,2)

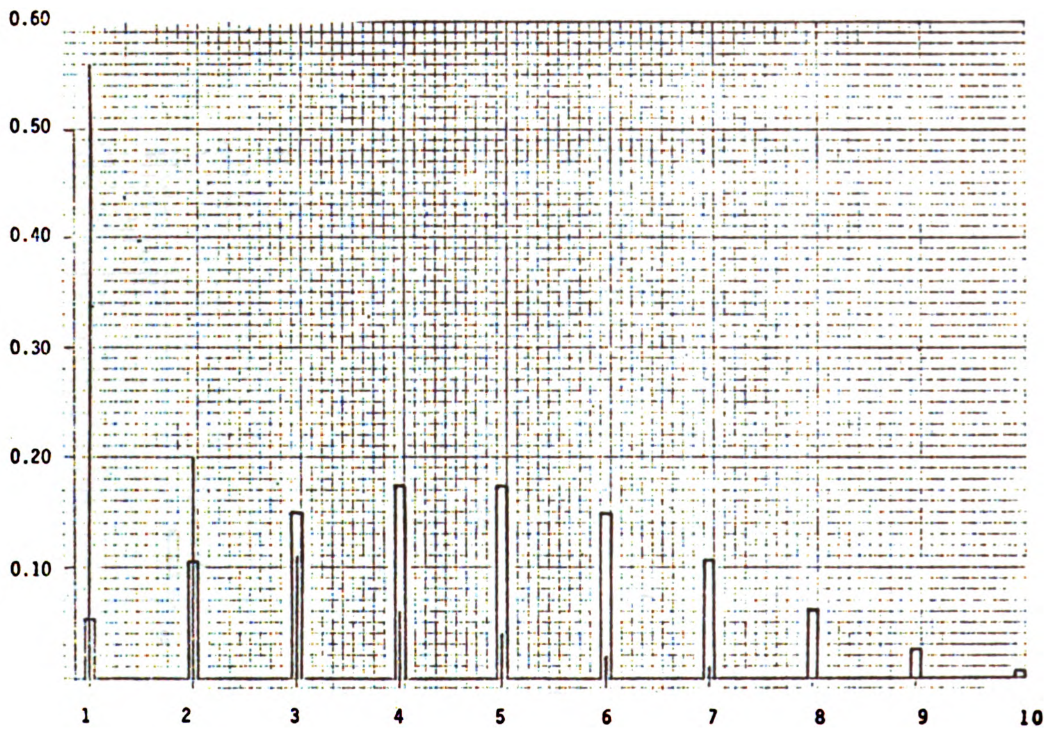




H_0: 10 pairs of observations drawn from the same distribution

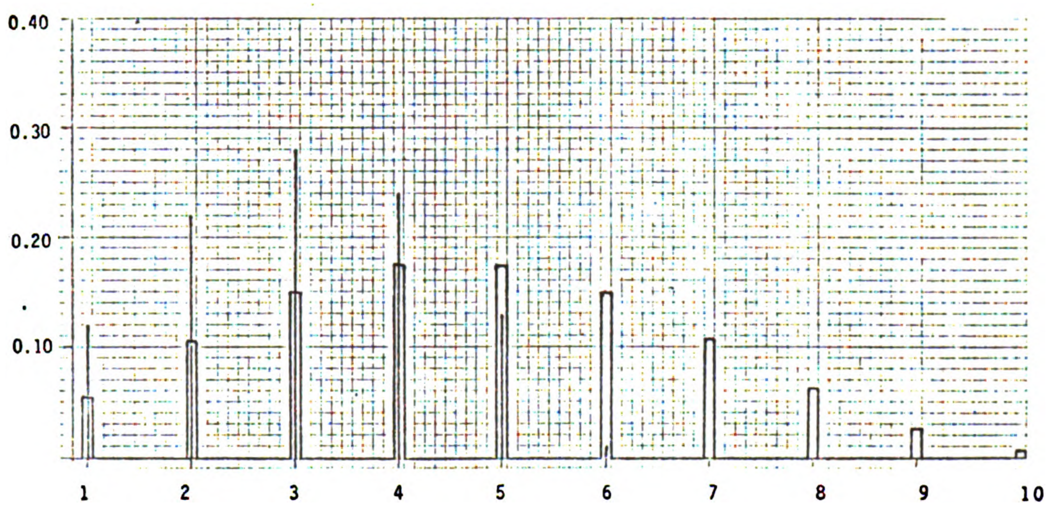
H_1: 9 pairs from N(x,4) and 1 pair from N(x+8,4)





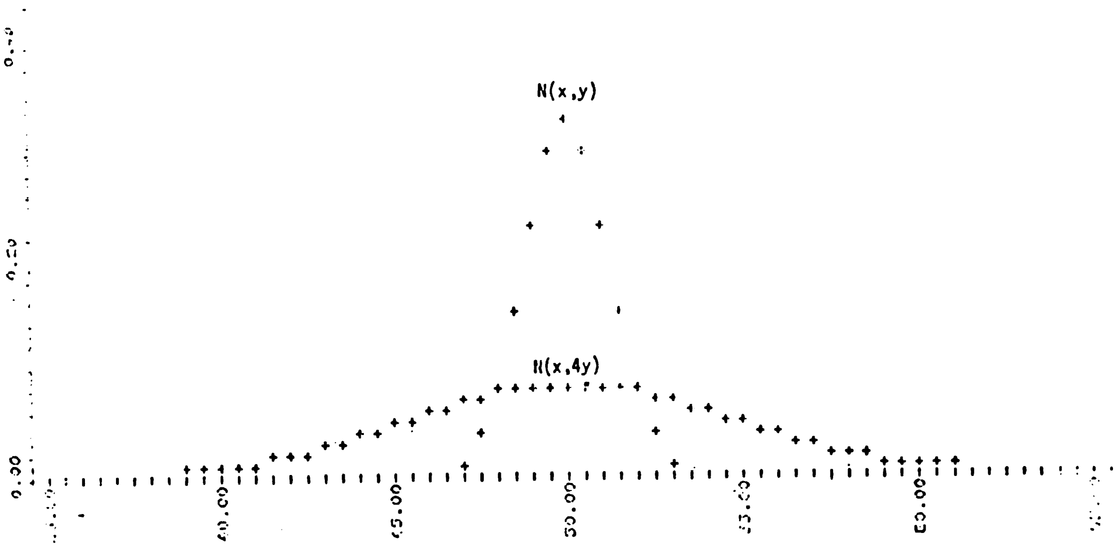
H_0: 10 pairs of observations drawn from the same distribution

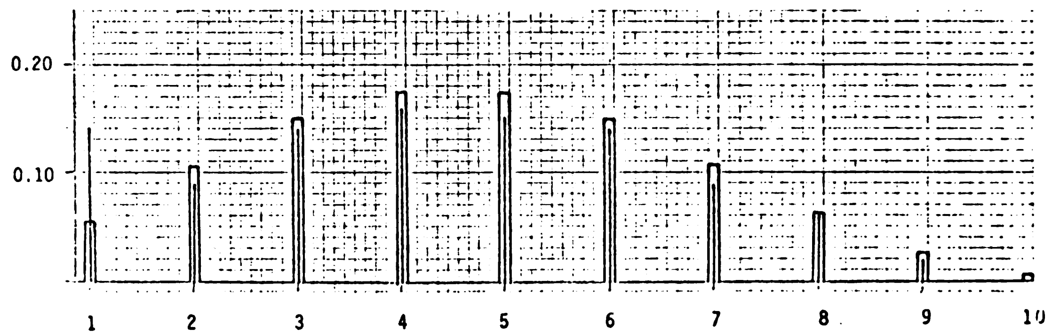
H_1: 9 pairs from N(x,4) and 1 pair from N(x+10,4)



H_0: 10 pairs of observations drawn from the same distribution

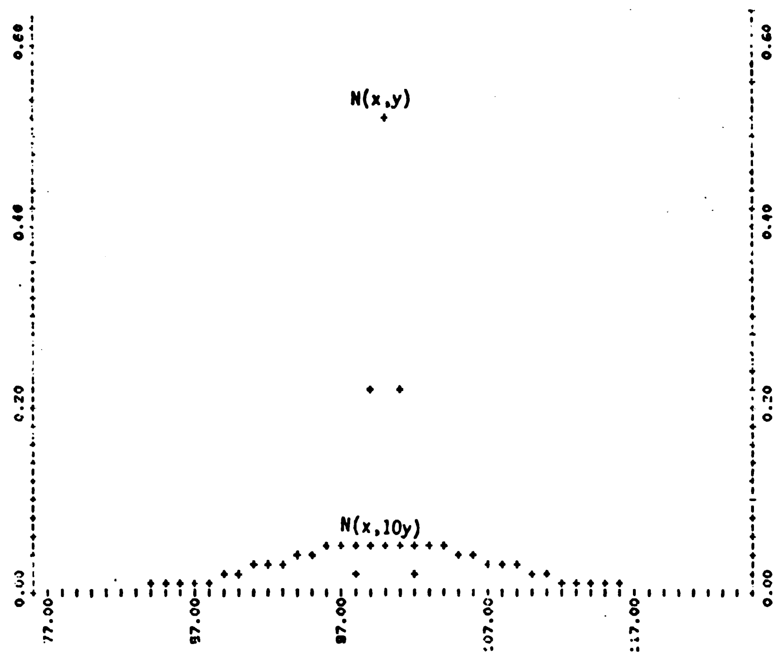
H_1: 5 pairs from N(x,4) and 5 pairs from N(x+10,4)

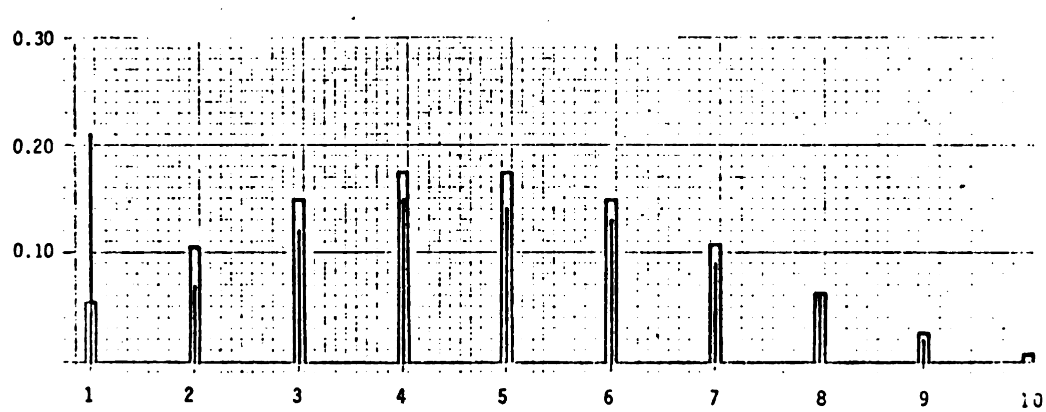




H_0: 10 pairs of observations drawn from the same distribution

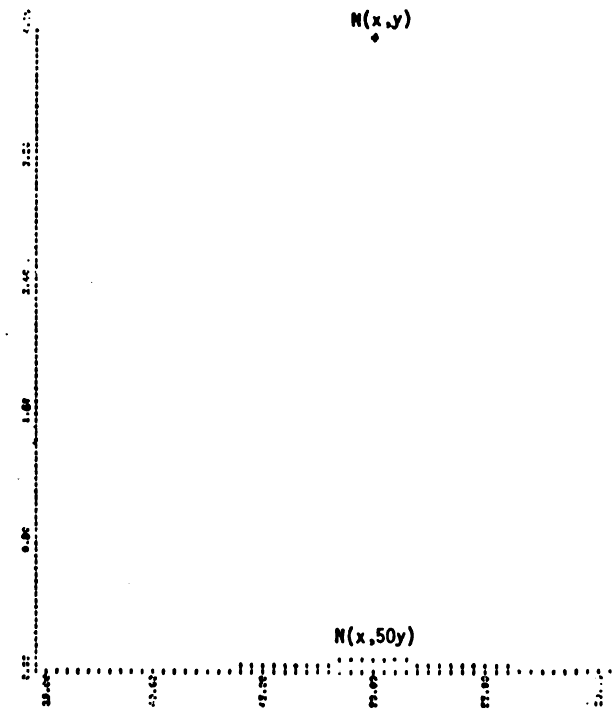
H_1: 9 pairs from N(x,y) and 1 pair from N(x,4y)

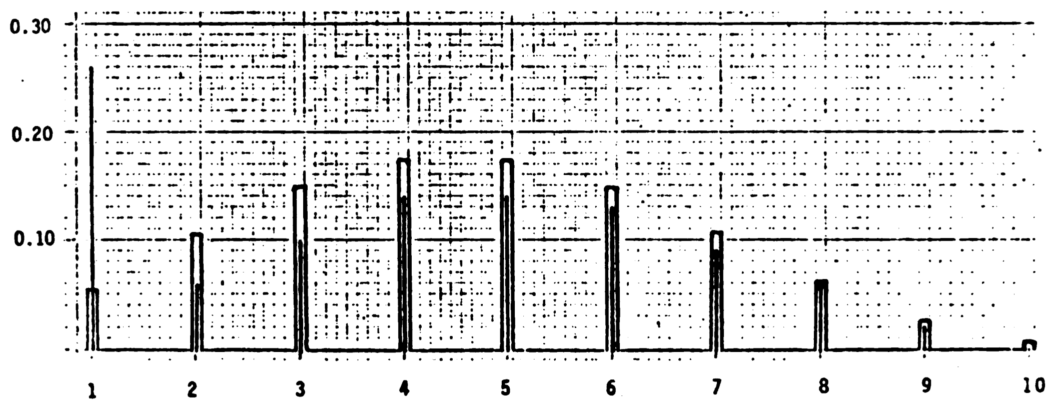




H_0: 10 pairs of observations drawn from the same distribution

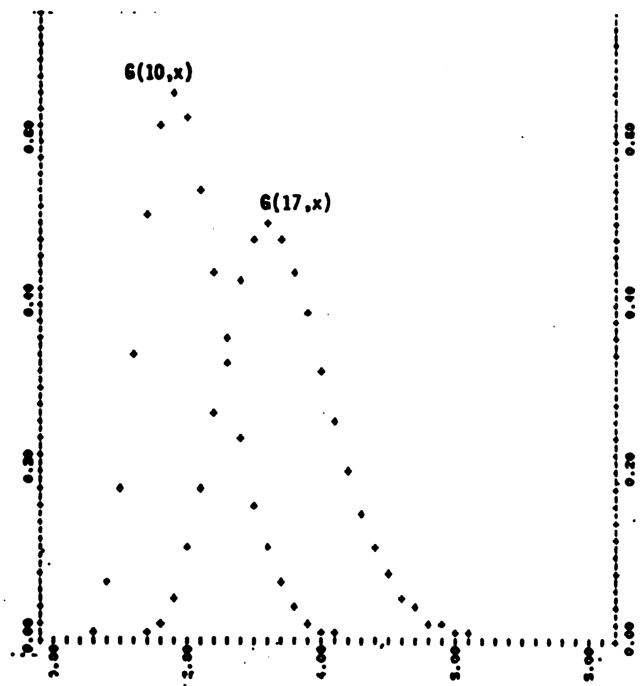
H_1: 9 pairs from N(x,y) and 1 pair from N(x,10y)

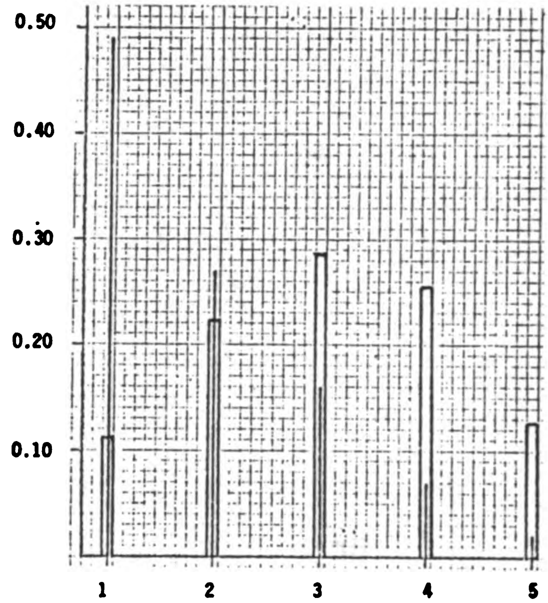




H_0: 10 pairs of observations drawn from the same distribution

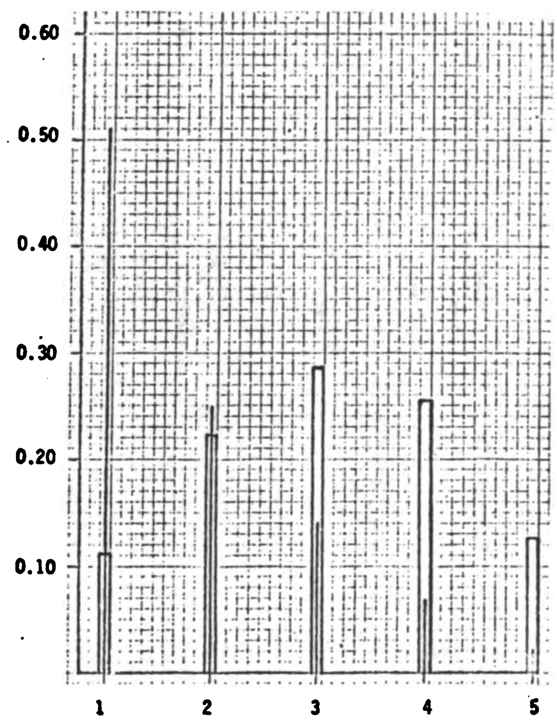
H_1: 9 pairs from N(x,y) and 1 pair from N(x,50Y)


6(10,x)
6(17,x)



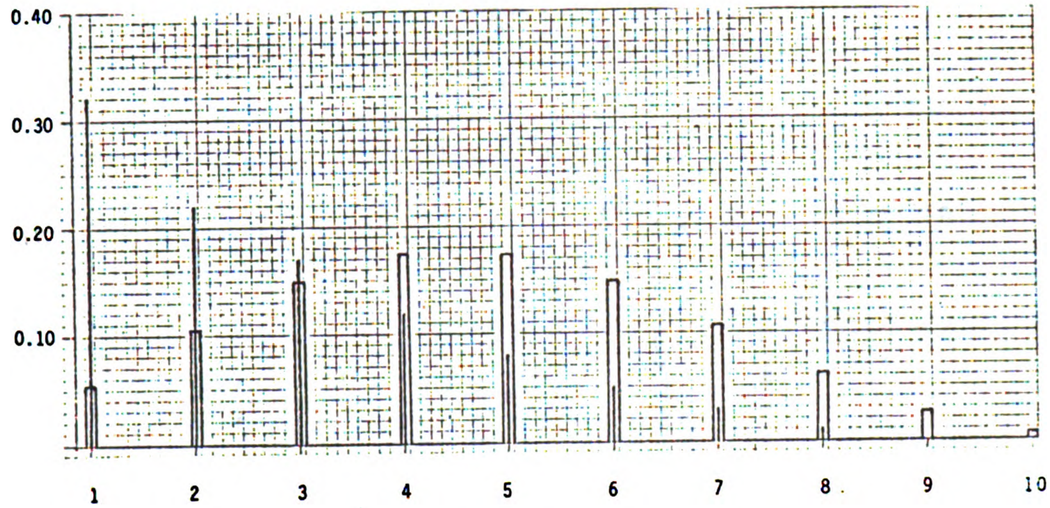
H_0: 5 pairs of observations drawn from the same distribution

H_1: 4 pairs from G(10,x) and 1 pair from G(17,x) (for LIR of A's)



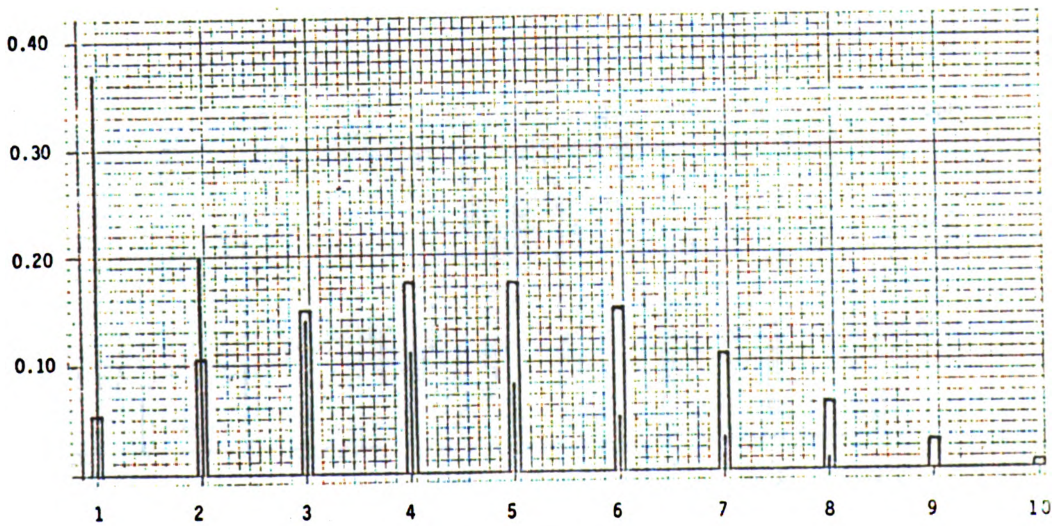
H_0: 5 pairs of observations drawn from the same distribution

: 4 pairs from G(17,x) and 1 pair from G(10,x) (for LIR of B's)



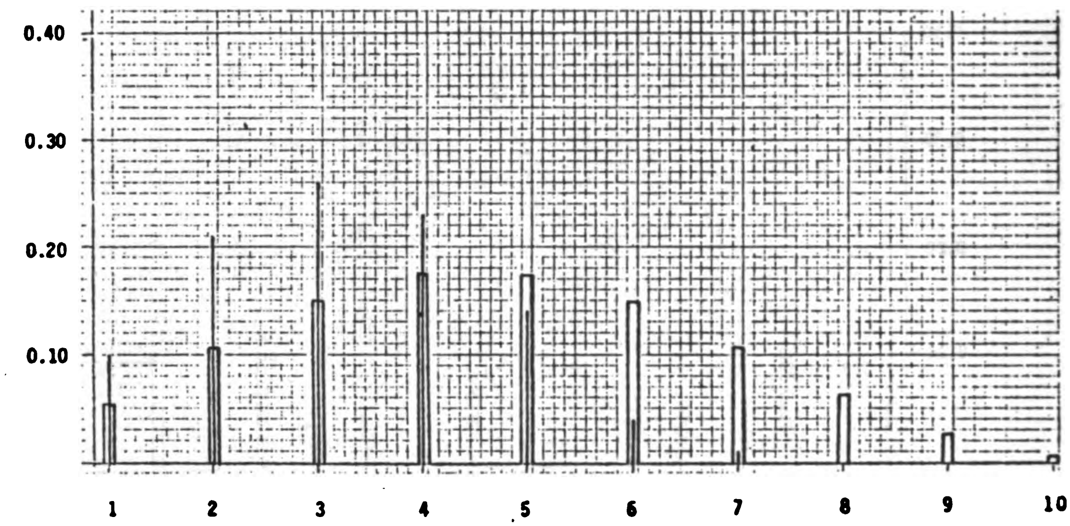
H_0: 10 pairs of observations drawn from the same distribution

H_1: 9 pairs from G(10,x) and 1 pair from G(17,x) (for LIR of A's)



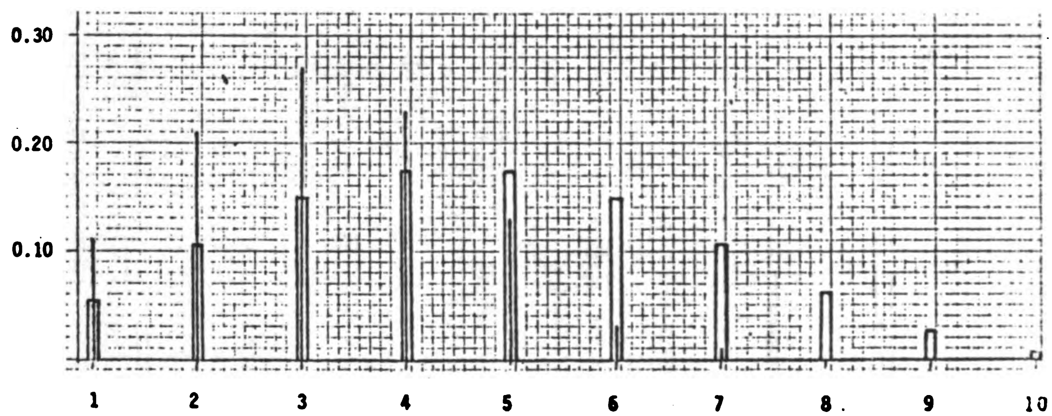
H_0: 10 pairs of observations drawn from the same distribution

H_1: 9 pairs from G(17,x) and 1 pair from G(10,x) (for LIR of B's)



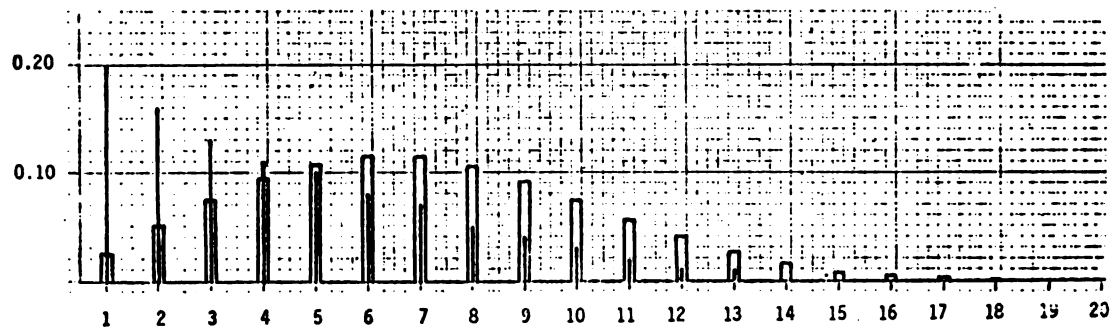
H_0: 10 pairs of observations drawn from the same distribution

H_1: 5 pairs from G(17,x) and 5 pair from G(10,x) (for LIR of A's)



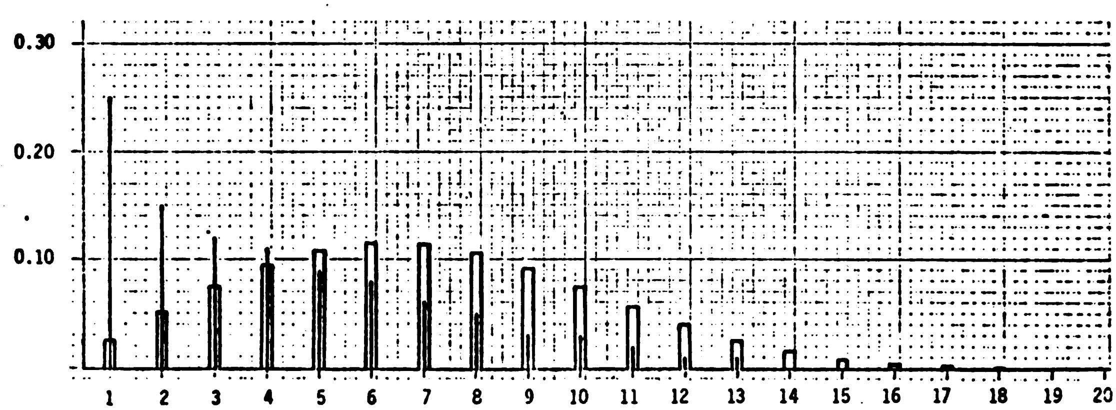
H_0: 10 pairs of observations drawn from the same distribution

H_1: 5 pairs from G(17,x) and 5 pair from G(10,x) (for LIR of B's)



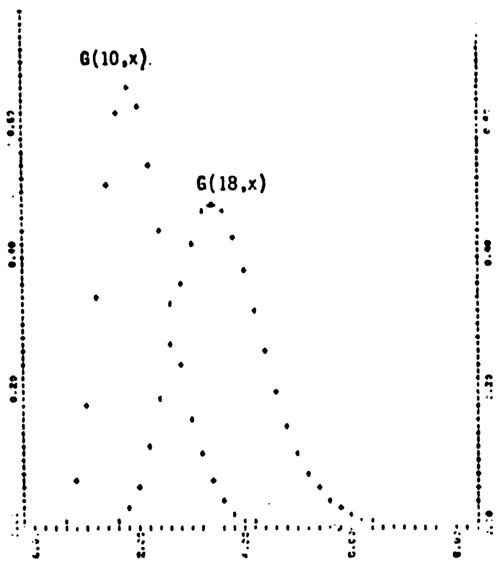
H_0: 20 pairs of observations drawn from the same distribution

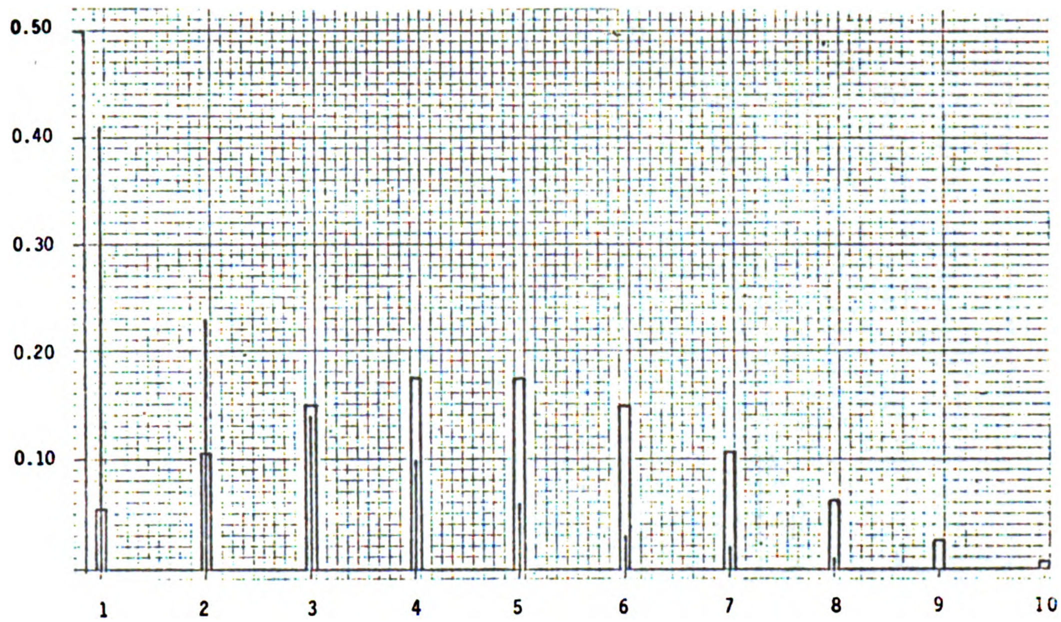
H_1: 19 pairs from G(10,x) and 1 pair from G(17,x) - A's



H_0: 20 pairs of observations drawn from the same distribution

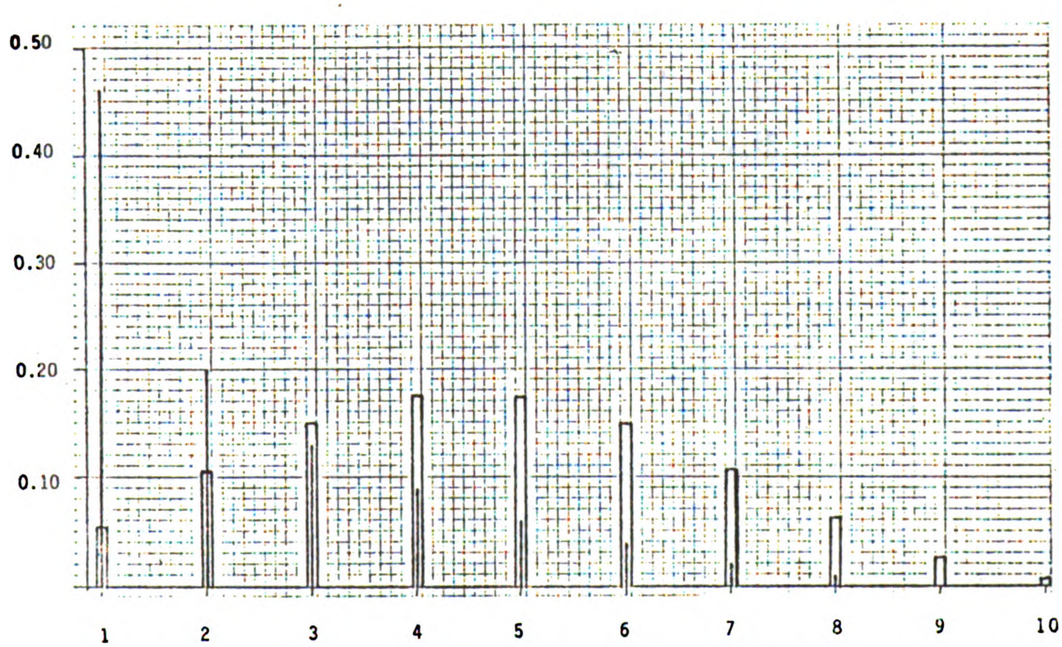
H_1: 19 pairs from G(17,x) and 1 pair from G(10,x) - B's


G(10,x)
G(18,x)



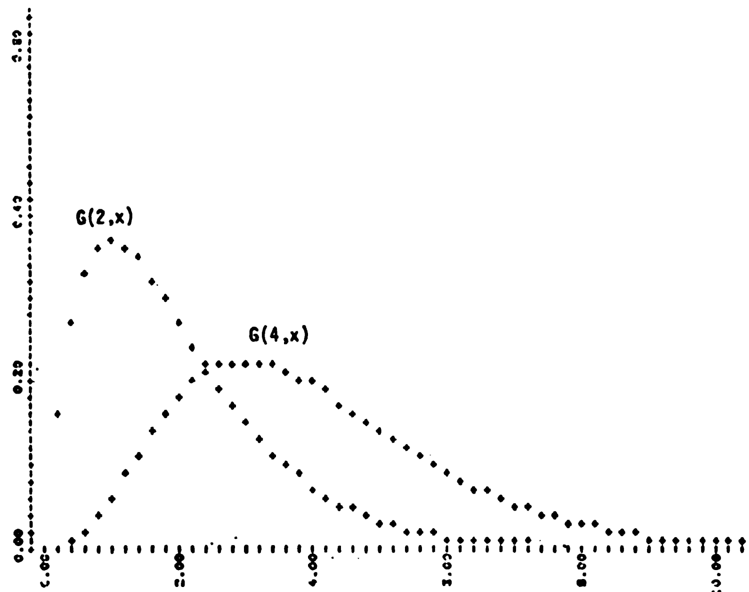
H_0: 10 pairs of observations drawn from the same distribution

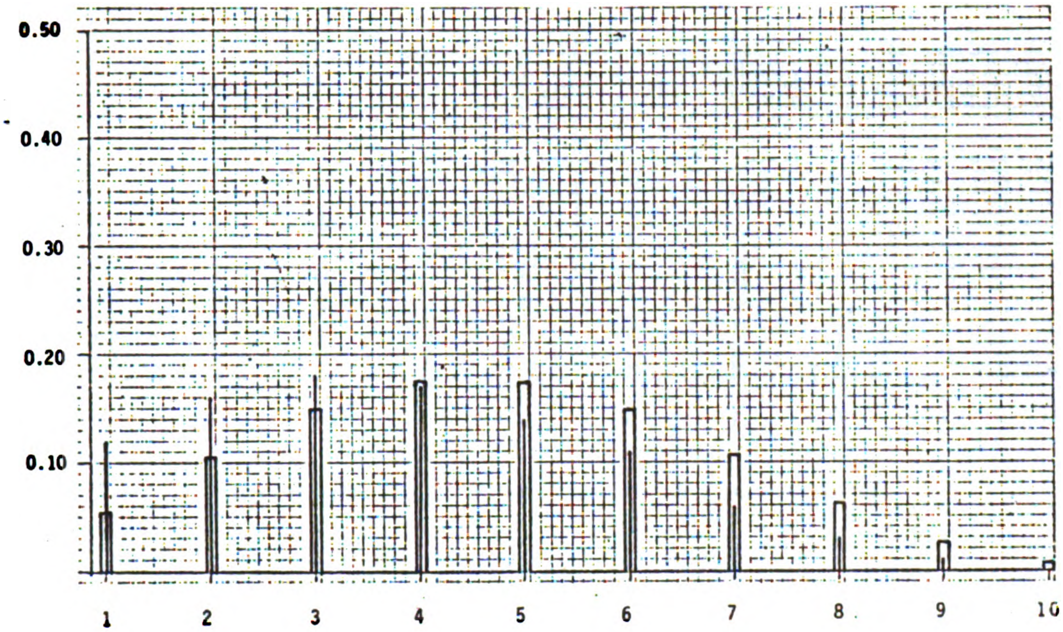
H_1: 9 pairs from G(10,x) and 1 pair from G(18,x) - A's



H_0: 10 pairs of observations drawn from the same distribution

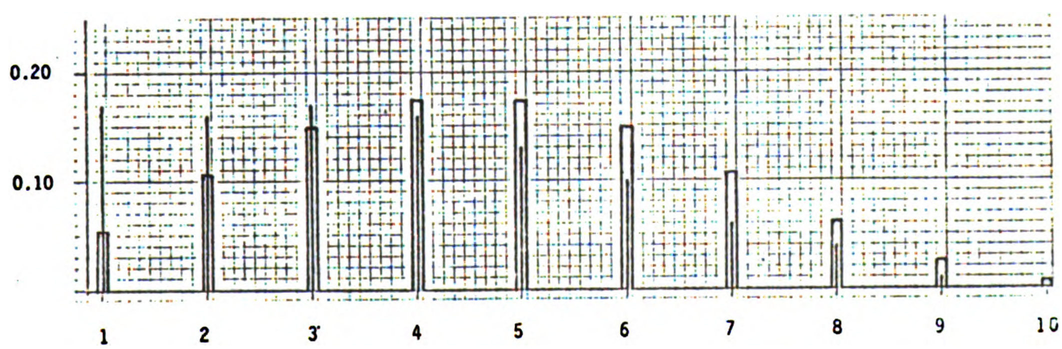
H_1: 9 pairs from G(18,x) and 1 pair from G(10,x) - B's


G(2,x)
G(4,x)



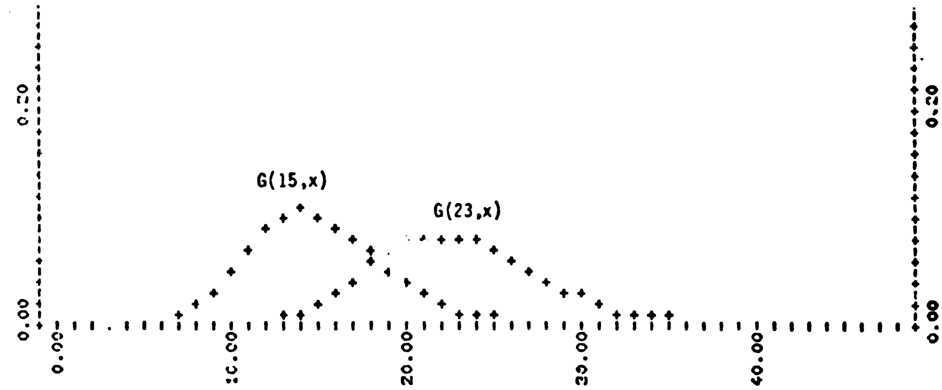
H_0: 10 pairs of observations drawn from the same distribution

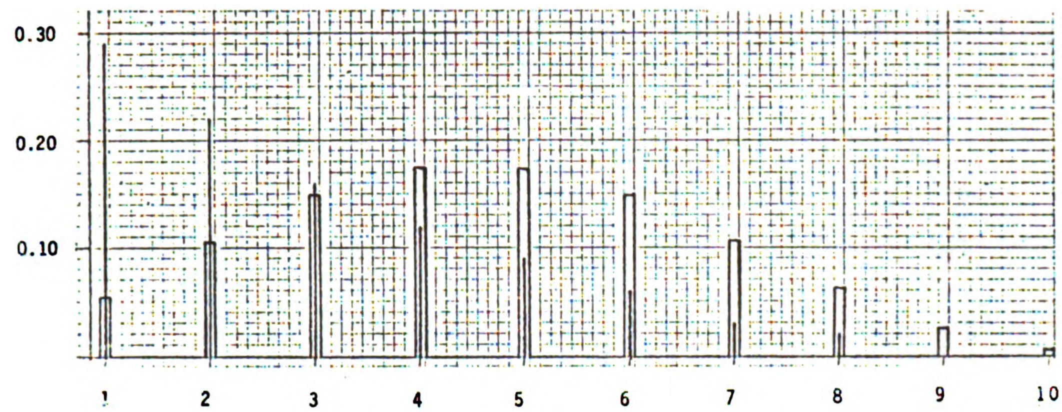
H_1: 9 pairs from G(2,x) and 1 pair from G(4,x) - A's



H_0: 10 pairs of observations drawn from the same distribution

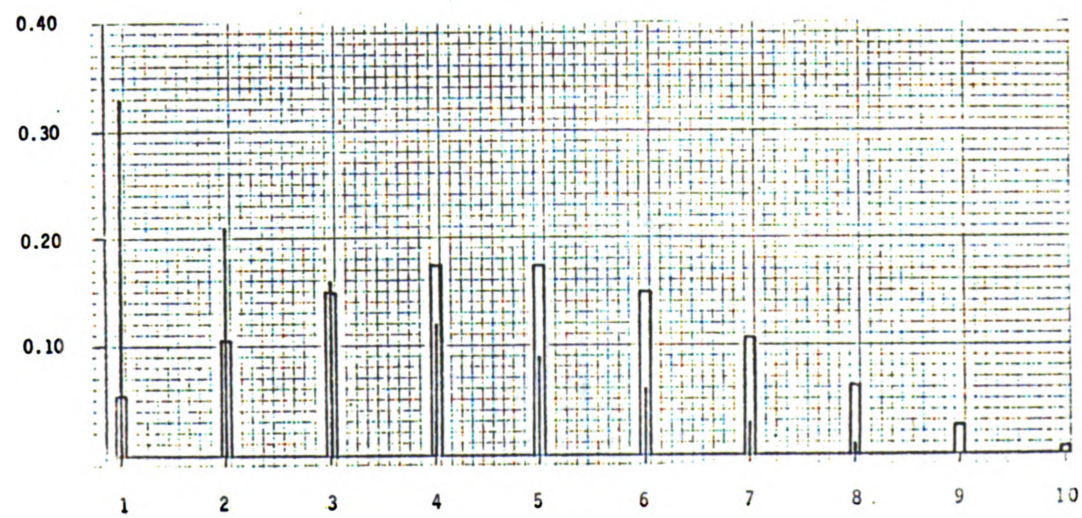
H_1: 9 pairs from G(4,x) and 1 pair from G(2,x) - B's


G(15,x)
G(23,x)
0.00
0.20
0.00
10.00
20.00
30.00
40.00



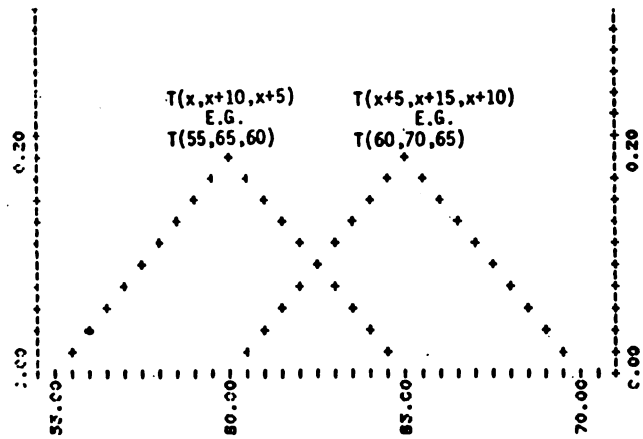
H_0: 10 pairs of observations drawn from the same distribution

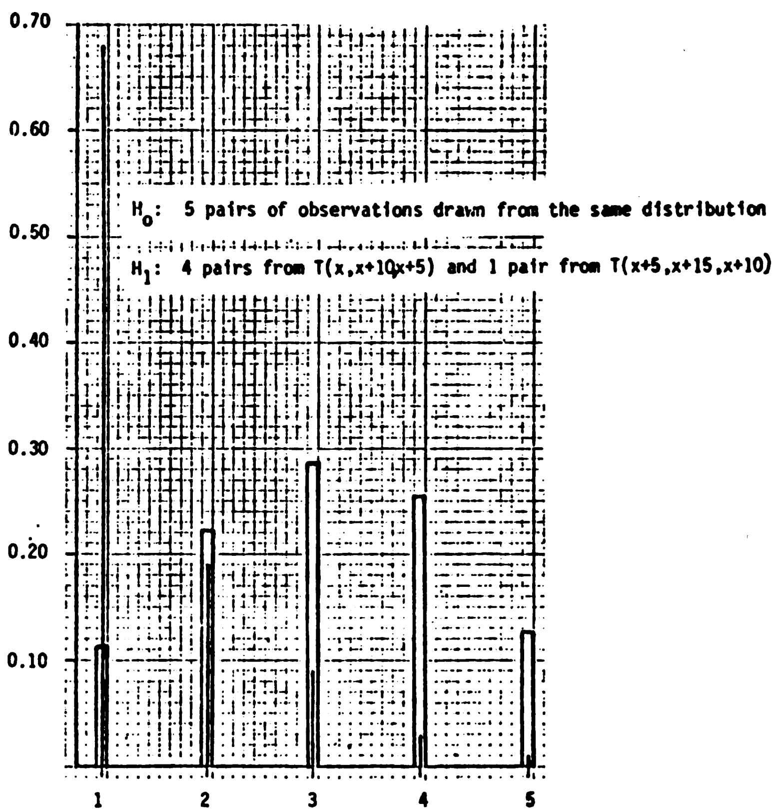
H_1: 9 pairs from G(15,x) and 1 pair from G(23,x) - A's

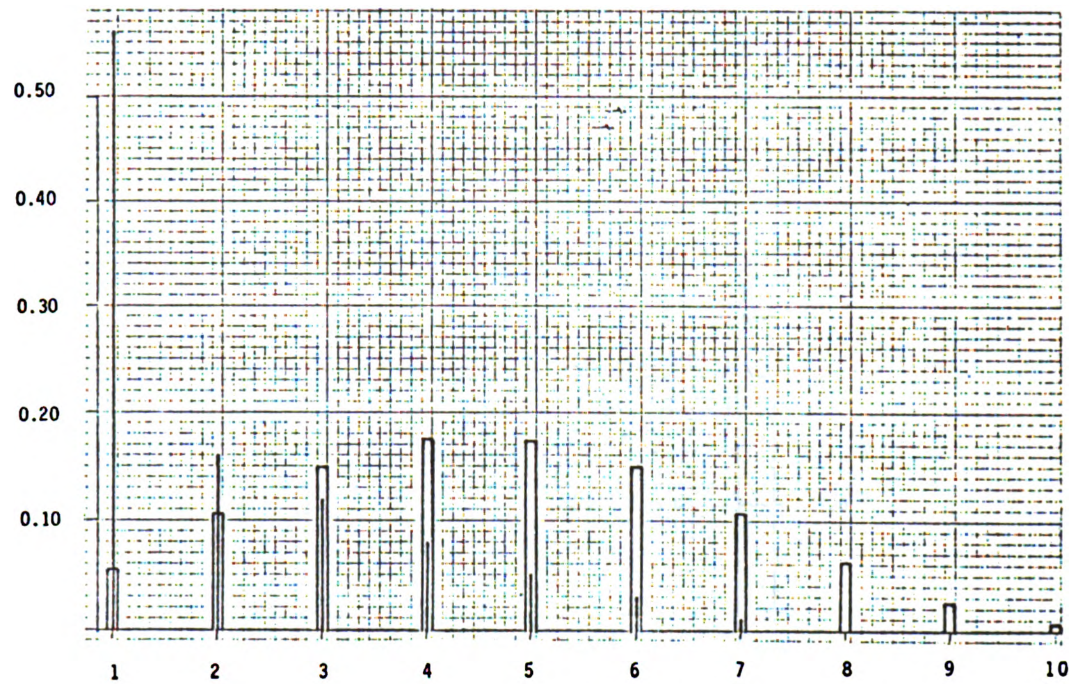


H_0: 10 pairs of observations drawn from the same distribution

H_1: 9 pairs from G(23,x) and 1 pair from G(15,x) - B's

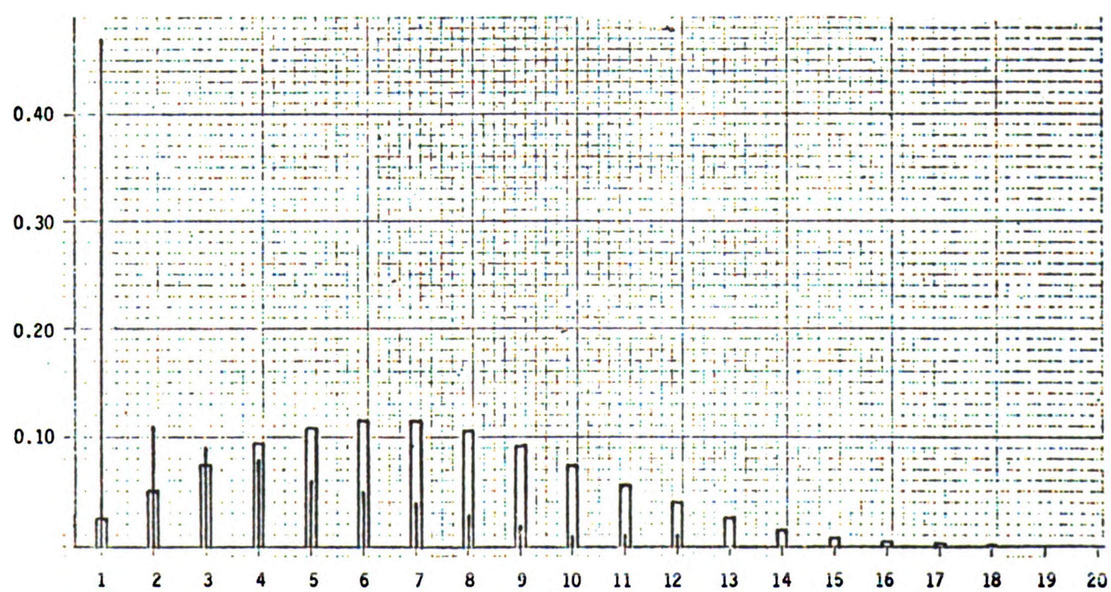

T(x,x+10,x+5)
E.G.
T(55,65,60)
T(x+5,x+15,x+10)
E.G.
T(60,70,65)
0.20
0.00
55.00
60.00
65.00
70.00


0.70
0.60
0.50
0.40
0.30
0.20
0.10
H0: 5 pairs of observations drawn from the same distribution
H1: 4 pairs from T(x,x+10,x+5) and 1 pair from T(x+5,x+15,x+10)
1
2
3
4
5



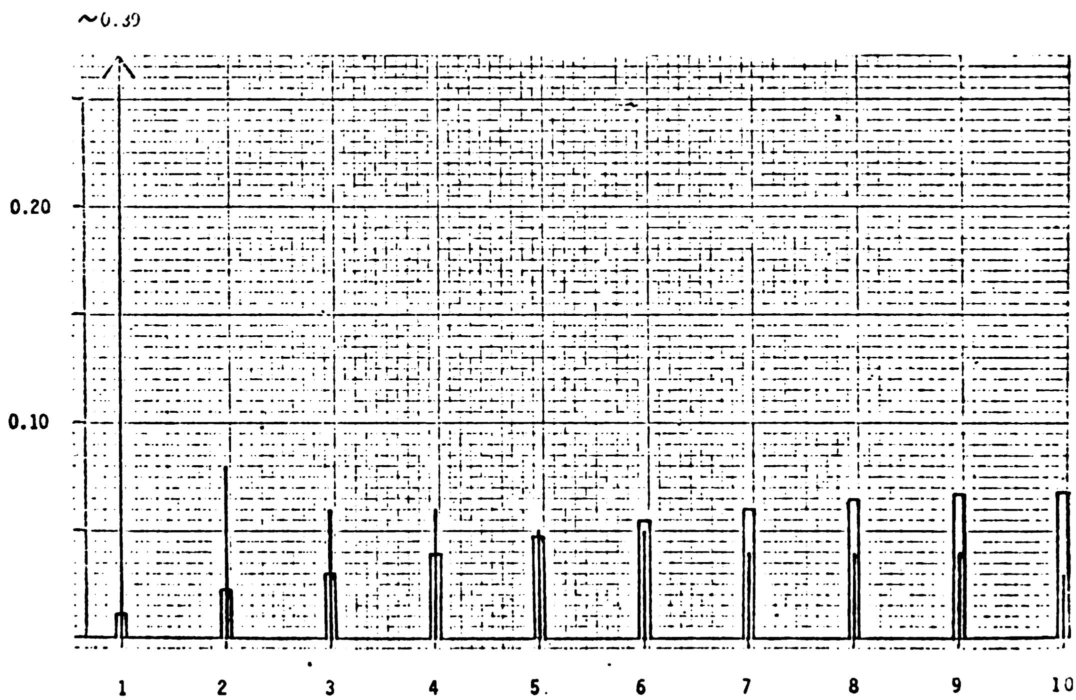
H_0: 10 pairs of observations drawn from the same distribution

H_1: 9 pairs from T(x,x+10,x+5) and 1 pair from T(x+5,x+15,X+10)



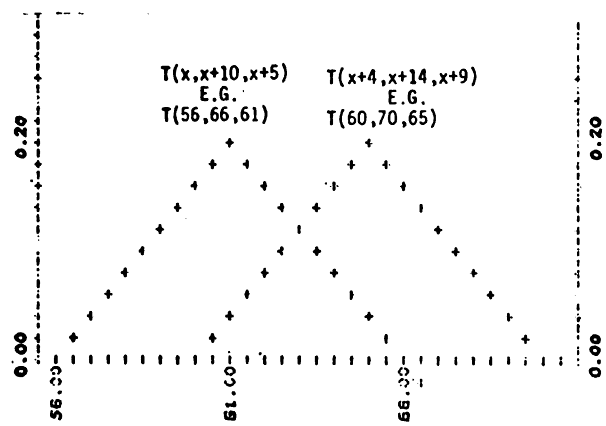
H_0: 20 pairs of observations drawn from the same distribution

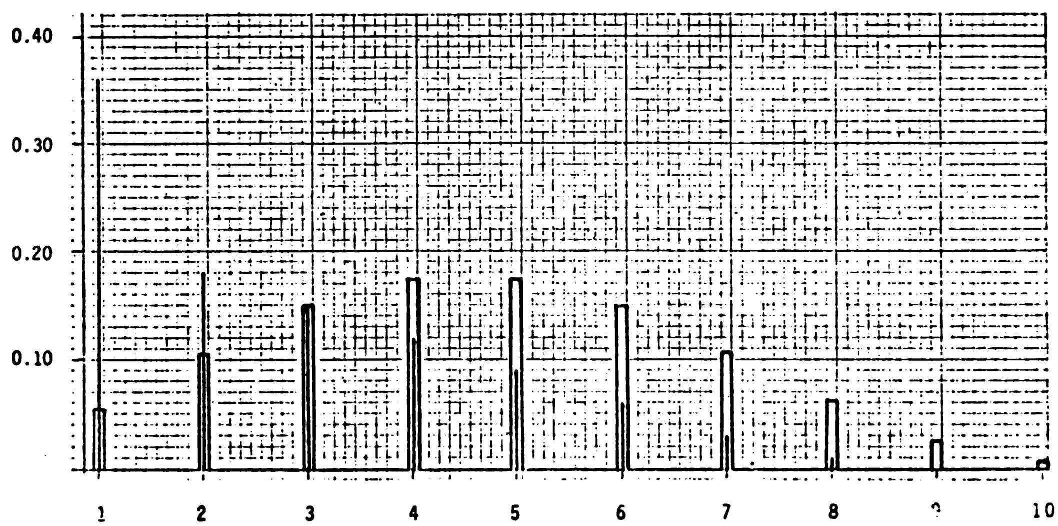
H_1: 19 pairs from T(x,x+10,x+5) and 1 pair from T(x+5,x+15,x+10)



H_0: 50 pairs of observations drawn from the same distribution

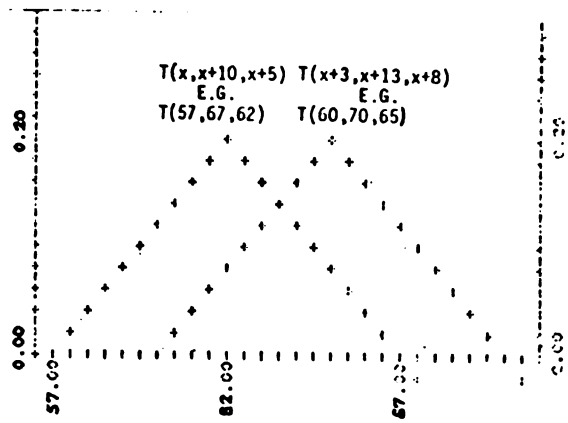
H_1: 49 pairs from T(x,x+10,x+5) and 1 pair from T(x+5,x+15, x+10)

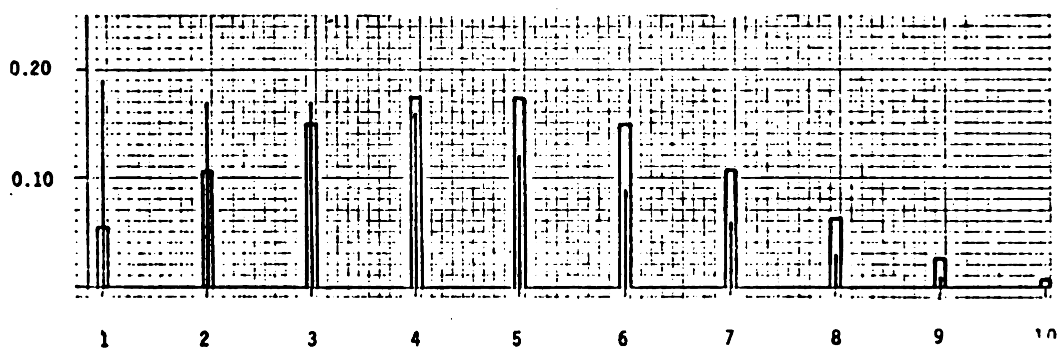




H_0: 10 pairs of observations drawn from the same distribution

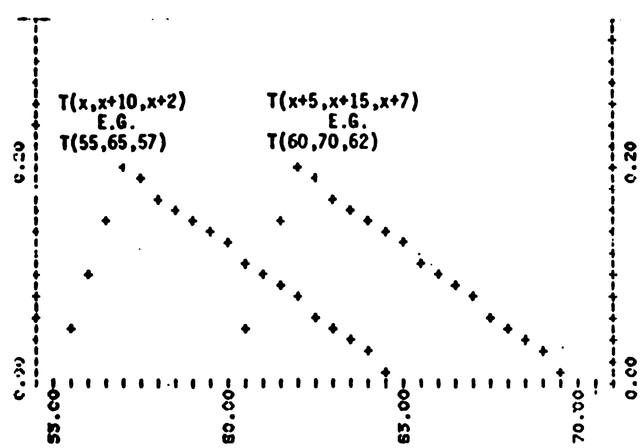
H_1: 9 pairs from T(x,x+10,x+15) and 1 pair from T(x+4,x+14,x+9)

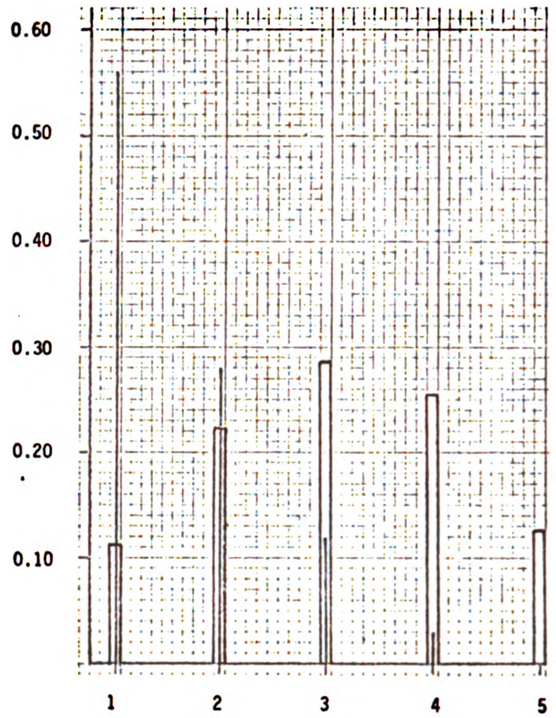




H_0: 10 pairs of observations drawn from the same distribution

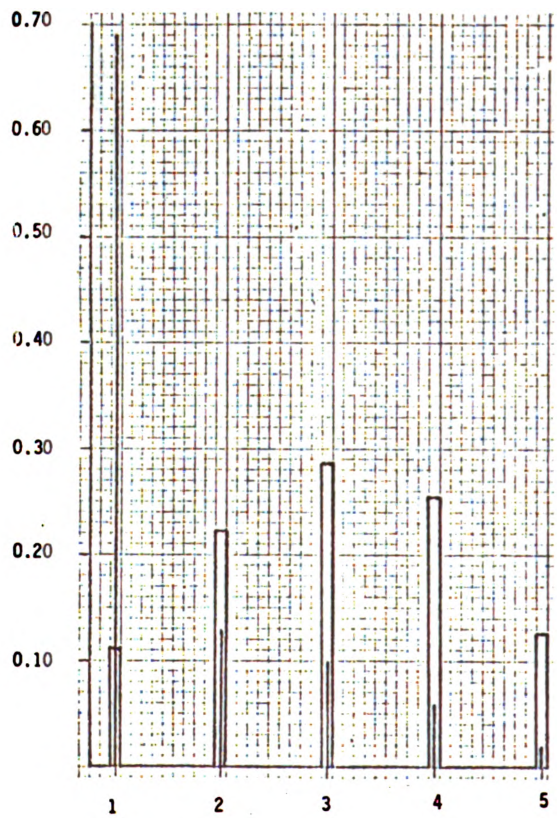
H_1: 9 pairs from T(x,x+10,x+5) and 1 pair from T(x+3,x+13,x+8)


T(x,x+10,x+2)
E.G.
T(55,65,57)
T(x+5,x+15,x+7)
E.G.
T(60,70,62)
0.20
0.00
55.00
60.00
65.00
70.00



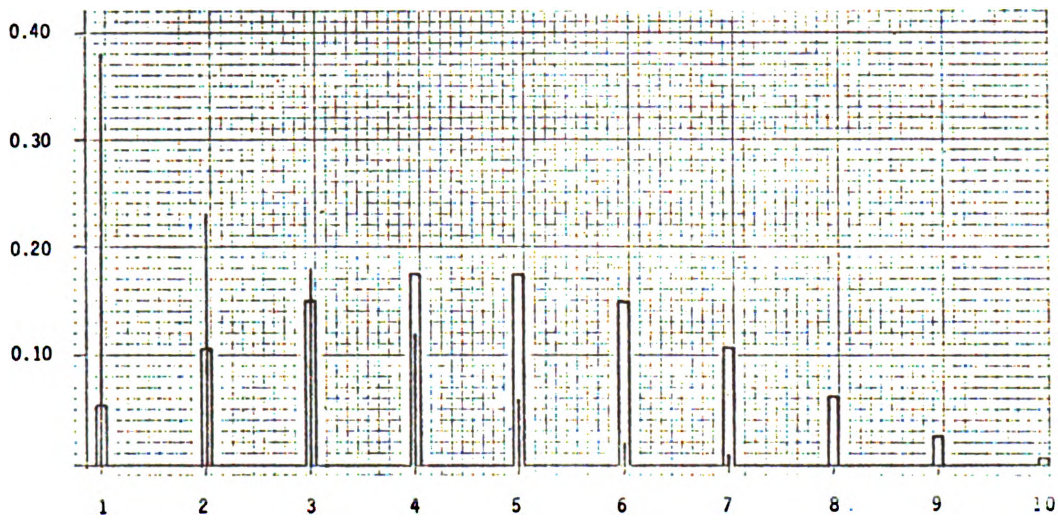
H_o: 5 pairs of observations drawn from the same distribution

H_1: 4 pairs from T(x,x+10,x+2) and 1 pair from T(x+5,x+15,x+7) - A's



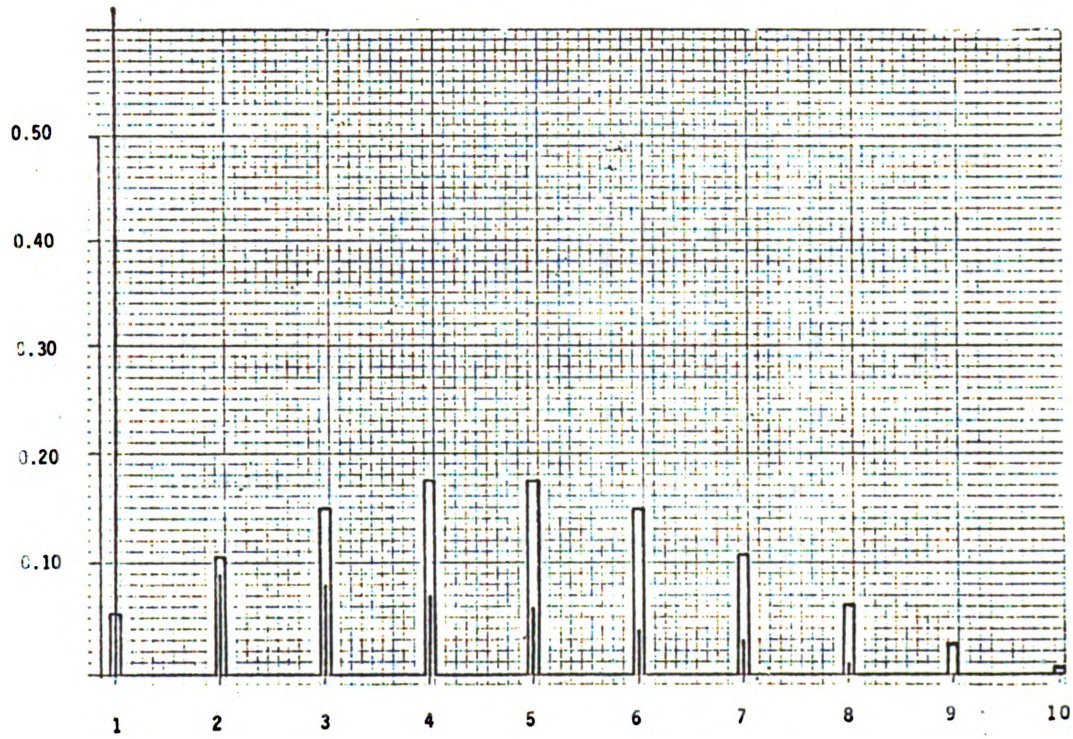
H_o: 5 pairs of observations drawn from the same distribution

H_1: 4 pairs from T(x+5,x+15,x+7) and 1 pair from T(x,x+10,x+2) - B's



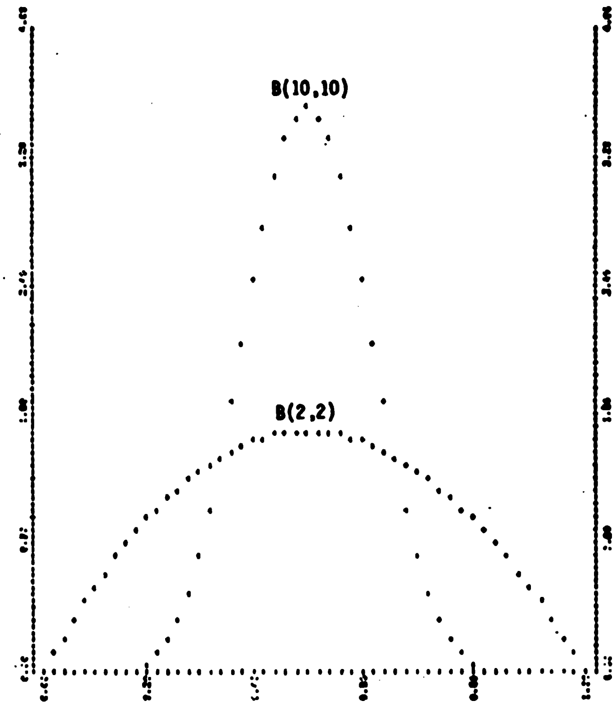
H_0: 10 pairs of observations drawn from the same distribution

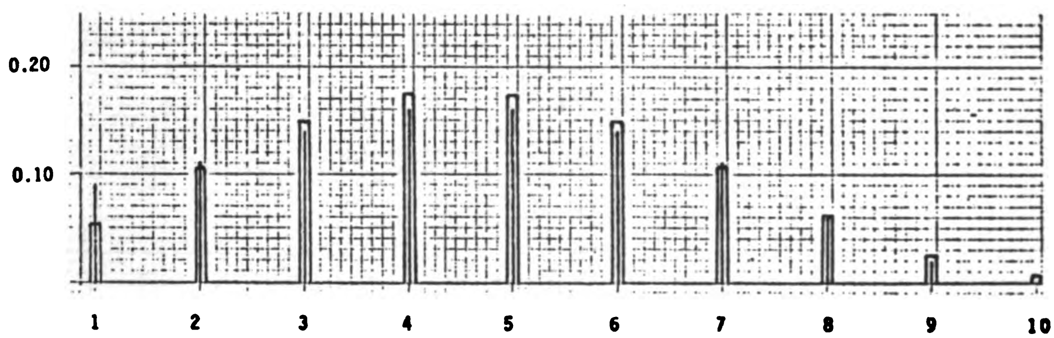
H_1: 9 pairs from T(x,x+10,x+2) and 1 pair from T(x+5,x+15,x+7) - A's



H_0: 10 pairs of observations drawn from the same distribution

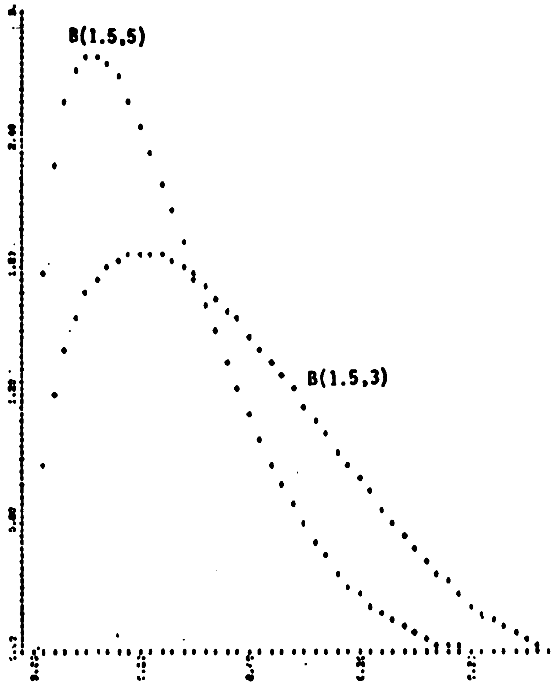
H_1: 9 pairs from T(x+5,x+15,x+7) and 1 pair from T(x,x+10,x+2) - B's

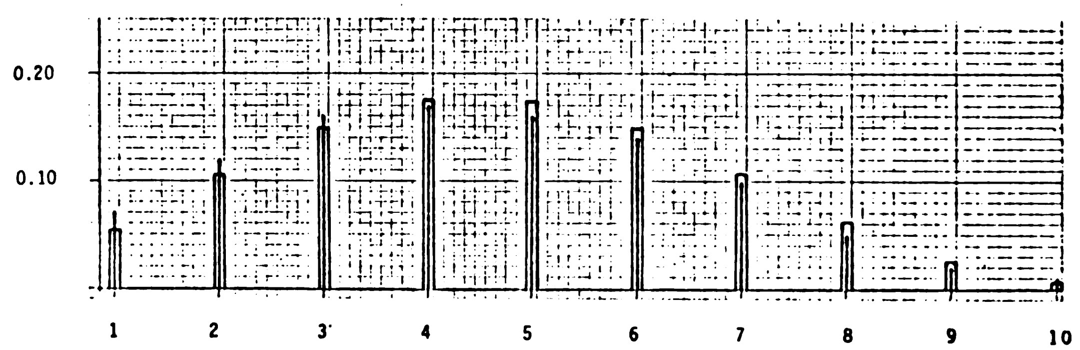




H_0: 10 pairs of observations drawn from the same distribution

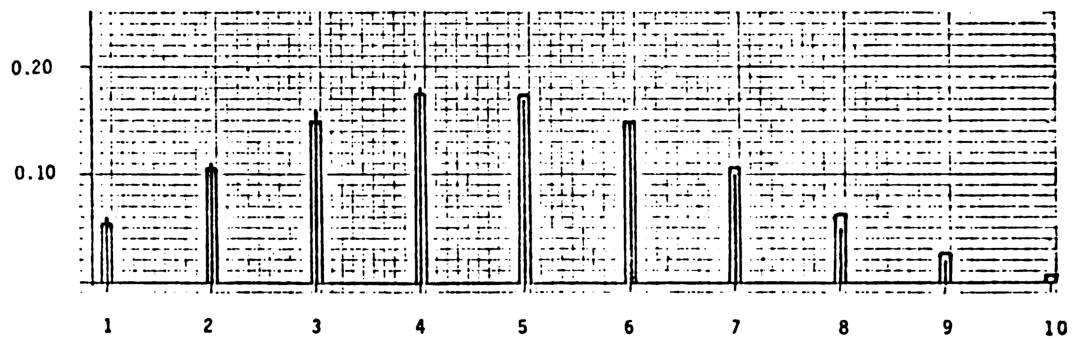
H_1: 9 pairs from B(10,10) and 1 pair from B(2,2)


B(1.5,5)
B(1.5,3)



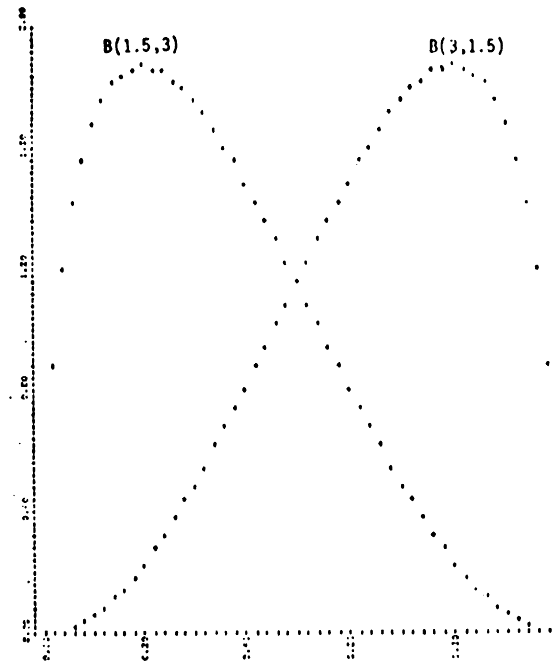
H_0: 10 pairs of observations drawn from the same distribution

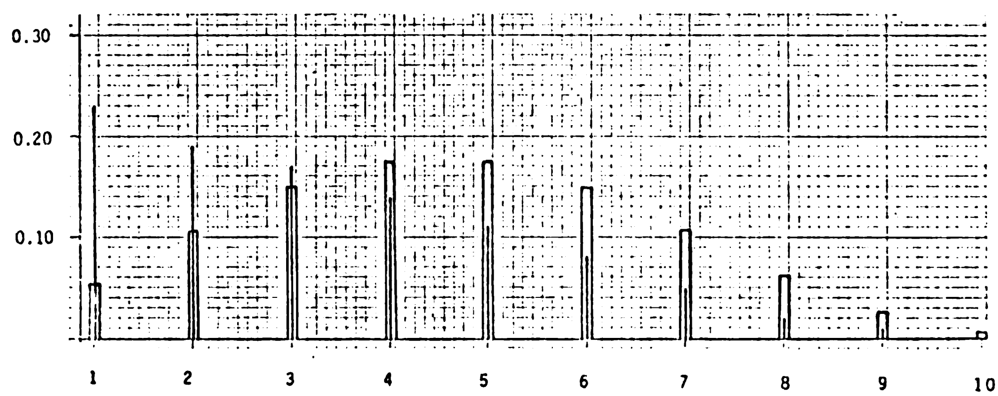
H_1: 9 pairs from B(1.5,5) and 1 pair from B(1.5,3) - A's



H_0: 10 pairs of observations drawn from the same distribution

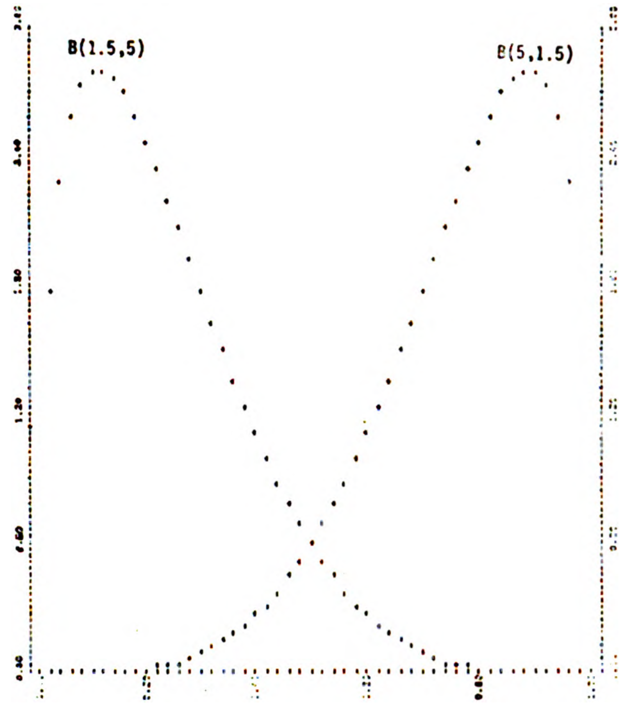
H.: 9 pairs from B(1.5,3) and 1 pair from B(1.5,5) - B's

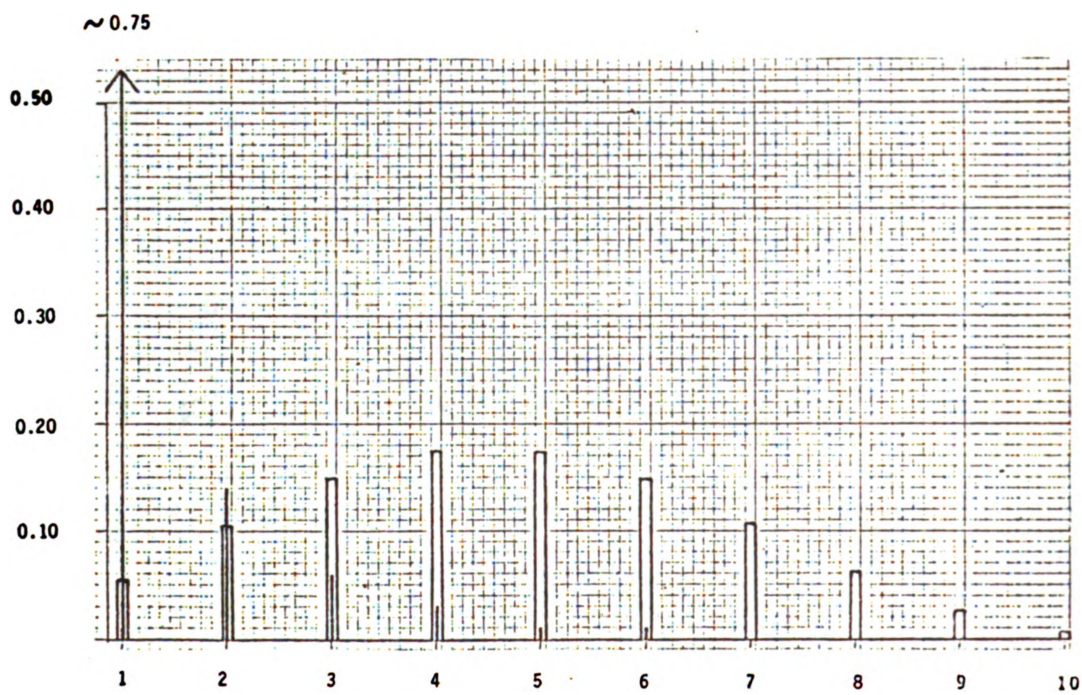




H_0: 10 pairs of observations drawn from the same distribution

H_1: 9 pairs from B(1.5,3) and 1 pair from B(3,1.5)





H_0: 10 pairs of observations drawn from the same distribution

H_1: 9 pairs from B(1.5,5) and 1 pair from B(5,1.5)

APPENDIX III

ALIRT

As stated in the introduction to this paper, if "A" is the larger value and "B" is the smaller value of one pair of observations, then the probability for the Length of Initial Run (LIR) of "A"'s or the probability for the LIR of "B"'s can be used to indicate whether data has been drawn from the same distribution or from two different distributions.

In most cases, determining the probability of a LIR equal to N (where N=1,2,3,...) is impractical using analytical methods.

The purpose of this program is to use simulation to estimate the probability of observing a LIR equal to N for an alternative hypothesis that assumes data has been drawn from two different distributions rather than the same distribution (as the null hypothesis assumes).

LIST OF VARIABLES

Variables used in subprogram INITIALIZE

DIST: Value ,determined by input, which specifies the probability distribution.

IREP: Input value for the total number of replications to be performed.

NFILE: Logical unit assignment based on input value.

NAME: Name of output file (when NFILE=8).

DISP: Disposition of output file at termination of program (when NFILE=8).

ITEM: Value (A or B) designating the type of run test to be performed

LETTER: Specifies the type of probability distribution; N-normal, G-gamma, B-beta, or T-triangular.

Variables used in subprogram SORT

SUM: The total number of samples drawn (NSAMP(1)+NSAMP(2)).

ARUN: Length of the initial run of 'A's before the first 'B'.

BRUN: Length of the initial run of 'B's before the first 'A'.

ARESULT(SUM): Array which stores the number of times each 'ARUN' length occurs during the entire simulation.

BRESULT(SUM): Array which stores the number of times each 'BRUN' length occurs during the entire simulation.

Variables used in subprogram REPORT

PERCENT(SUM): Relative frequency destribution of the LIR (ARESULT(SUM)/IREP; BRESULT(SUM)/IREP).

CUM(SUM): Cumulative frequency distribution of the LIR.

ANSWER: Input value which determines the status of bar graph output of frequency distribution (print:YES or NO).

Variables used in subprogram NORMAL

MU(I): Mean for distribution 1.

SIGMA(I): Standard deviation for distribution I.

X: An independent identically distributed uniform random number generated by the VAX 11/780 subprogram 'RAN'.

T= (2X-1): A test value to determine acceptability of the generated random numbers.

Y: N~(0,1).

X1,X2: N~(MU,SIGMA)'s.

Variables used in subprogram GAMMA

ALPHA(I) I=1,2: Shape parameter.

BETA(I) I=1,2: Scale parameter.

R: An independent identically distributed uniform random number generated by the VAX 11/780 subprogram 'RAN'.

TEST & W: Test values used to determine acceptability of the generated random numbers.

Y: GAM~(ALPHA(I),1) I=1,2.

X(I) I=1,2: GAM~(ALPHA(I),BETA(I)).

Variables used in subprogram BETA

ALPHA(I,J) I,J=1,2: Shape parameters.

R: An independent identically distributed uniform random number generated by the VAX 11/780 subprogram 'RAN'.

TEST & W: Test values used to determine acceptability of the generated random numbers.

Y: GAM~(ALPHA(I),1) I=1,2.

X1(I) I=1,2: GAM~(ALPHA(I),1) I=1,2.

X2(I) I=1,2: BET~(ALPHA(I),ALPHA(J)) I,J=1,2.

Variables used in subprogram TRIANG

G: Minimum value. Location parameter.

H: Maximum value. Max-Min=Scale parameter.

C: Mode. Shape parameter.

R: An independent identically distributed uniform random number generated by the VAX 11/780 subprogram 'RAN'.

X2,X3: TRIANG~(0,1,(mode-min)/(max-min)).

X(I) I=1,2: TRIANG~(min,max,mode).

Variables common to subprograms

NSAMP(I) I=1,2: The number of two-observation samples required for distribution I.

A(SUM) SUM=1,NSAMP(1)+NSAMP(2): Array of greater values of each two-observation sample from the combined NSAMP'S.

B(SUM) SUM=1,NSAMP(1)+NSAMP(2): Array of lesser values of each two-observation sample from the combined NSAMP'S.

COUNT: Counter for current number of replications; simulation terminates when COUNT=IREP.

I1: Initial value required for random number generation.

PROGRAM LIMITATIONS

Subprogram TRIANG: The mode 'C' used to generate the TRIANG(0,1,C), where C=(mode-min)/(max-min), is restricted to 0<C<1.

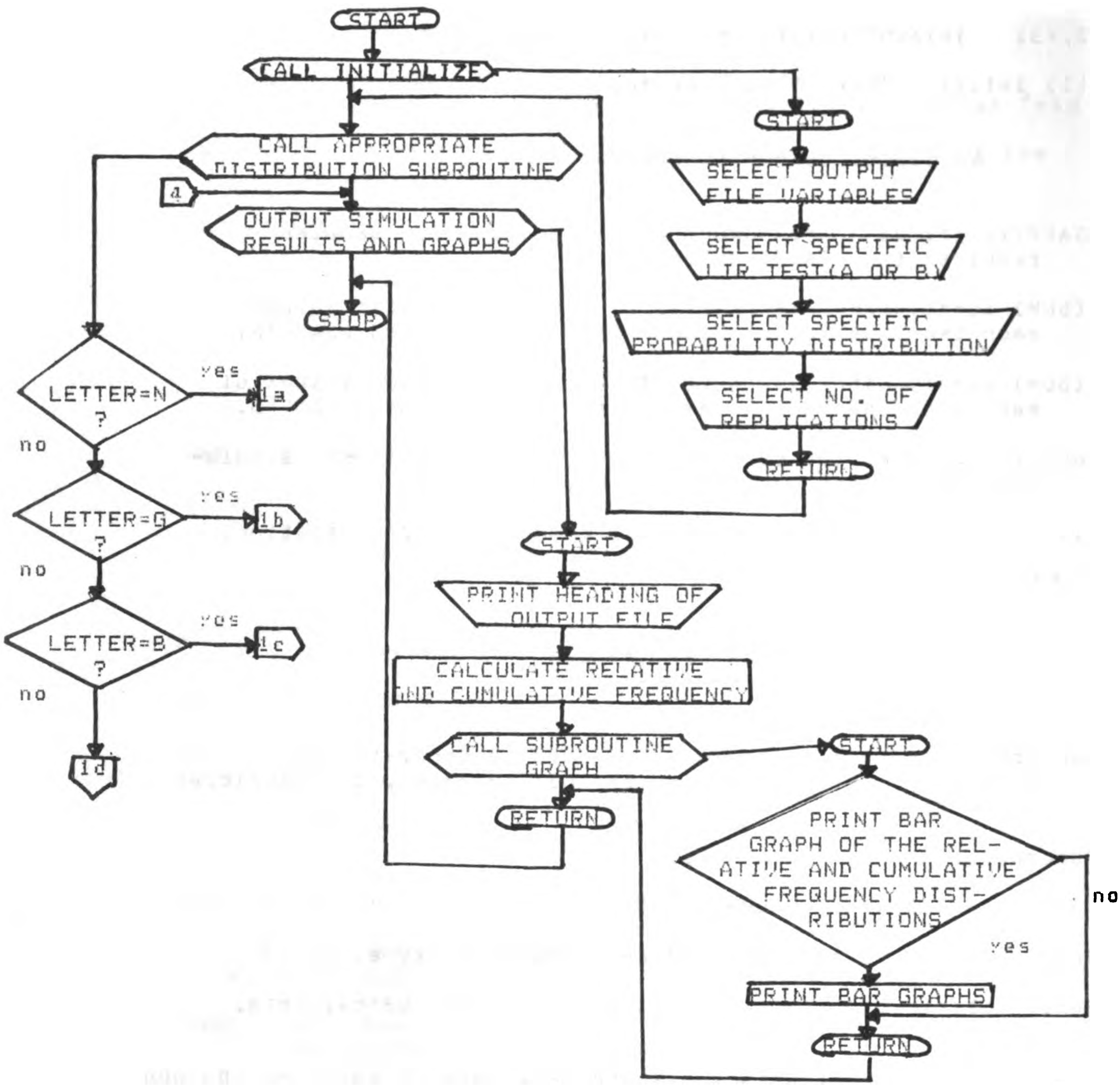
Subprogram GAMMA: ALPHA>1.

Subprogram BETA: ALPHA(I,J)>1 I=1,2.

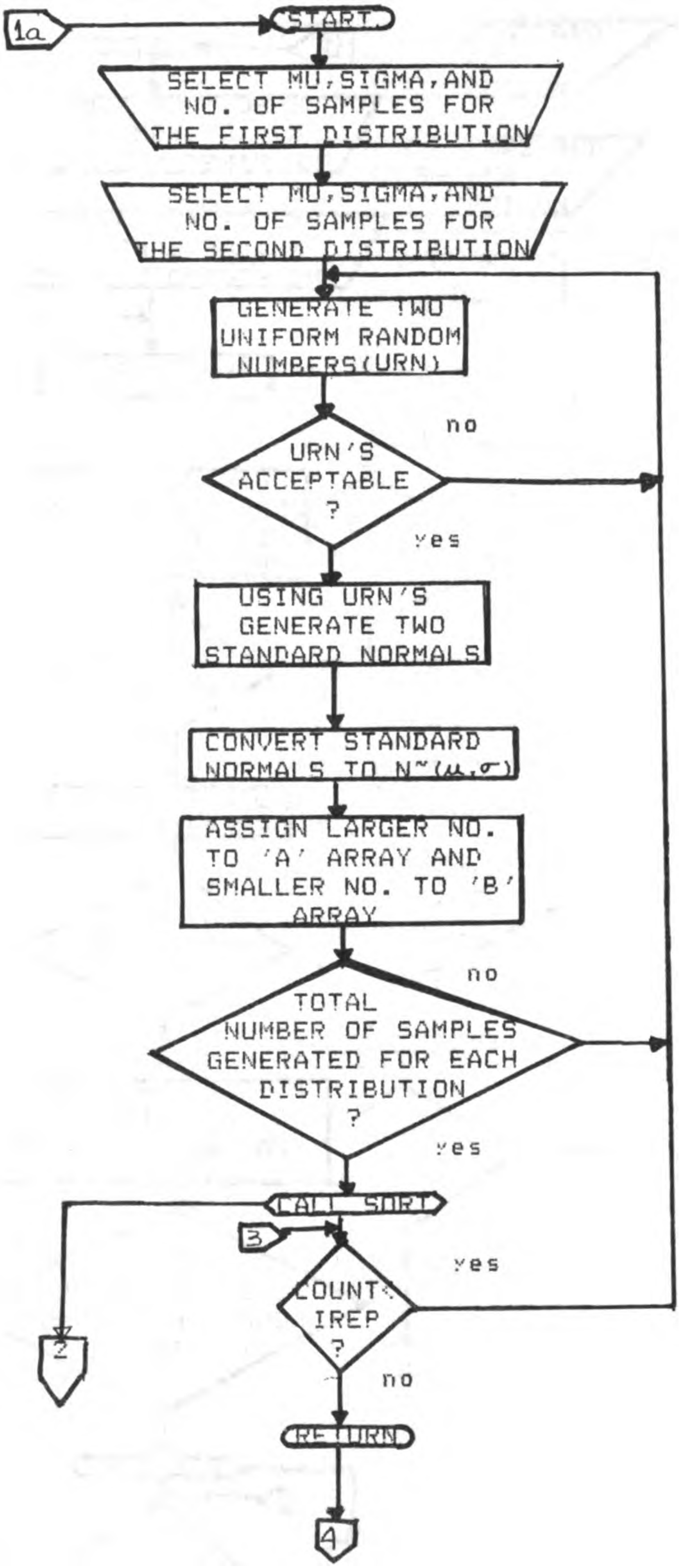
Input distributions must both be of the same type.

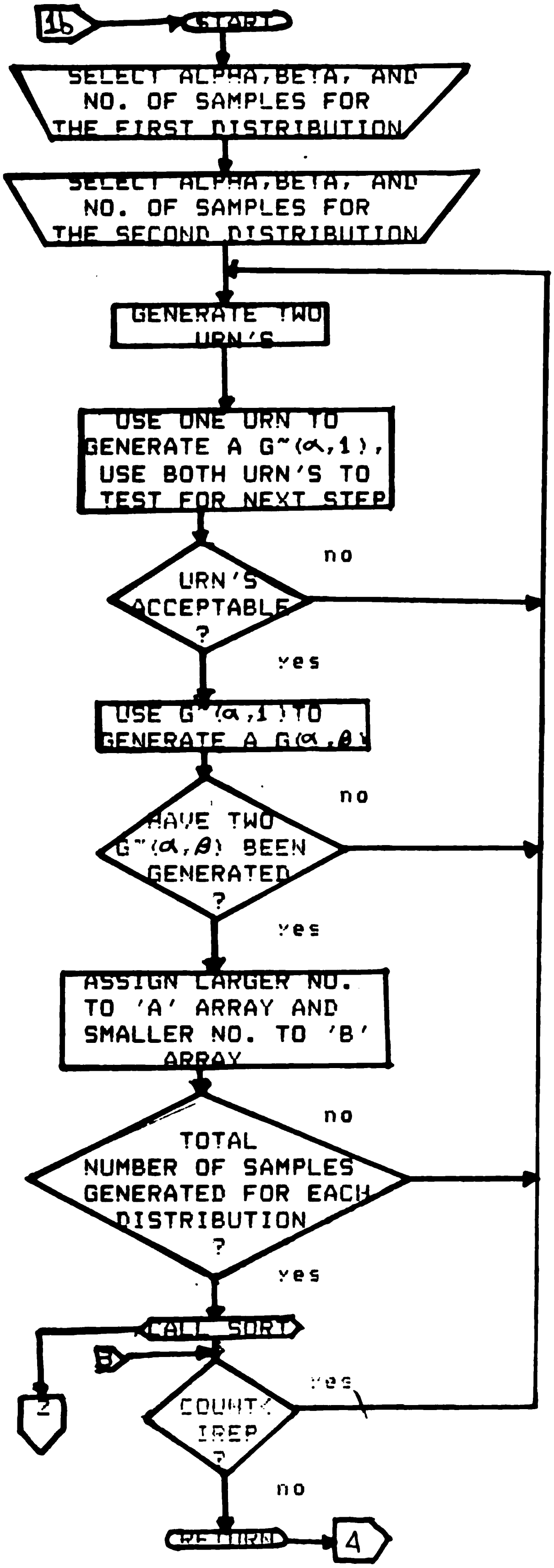
Input distribution type is limited to normal, gamma, beta, or triangular.

The number of replications must be less than or equal to 100,000.

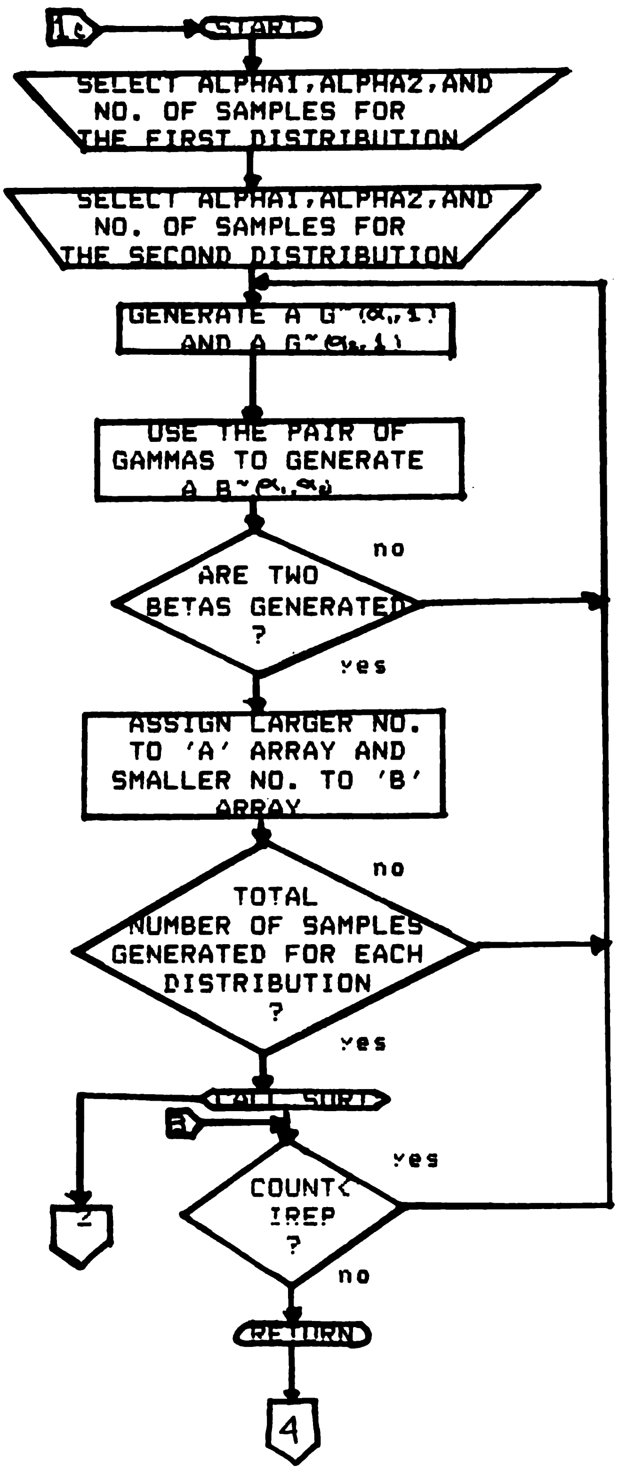


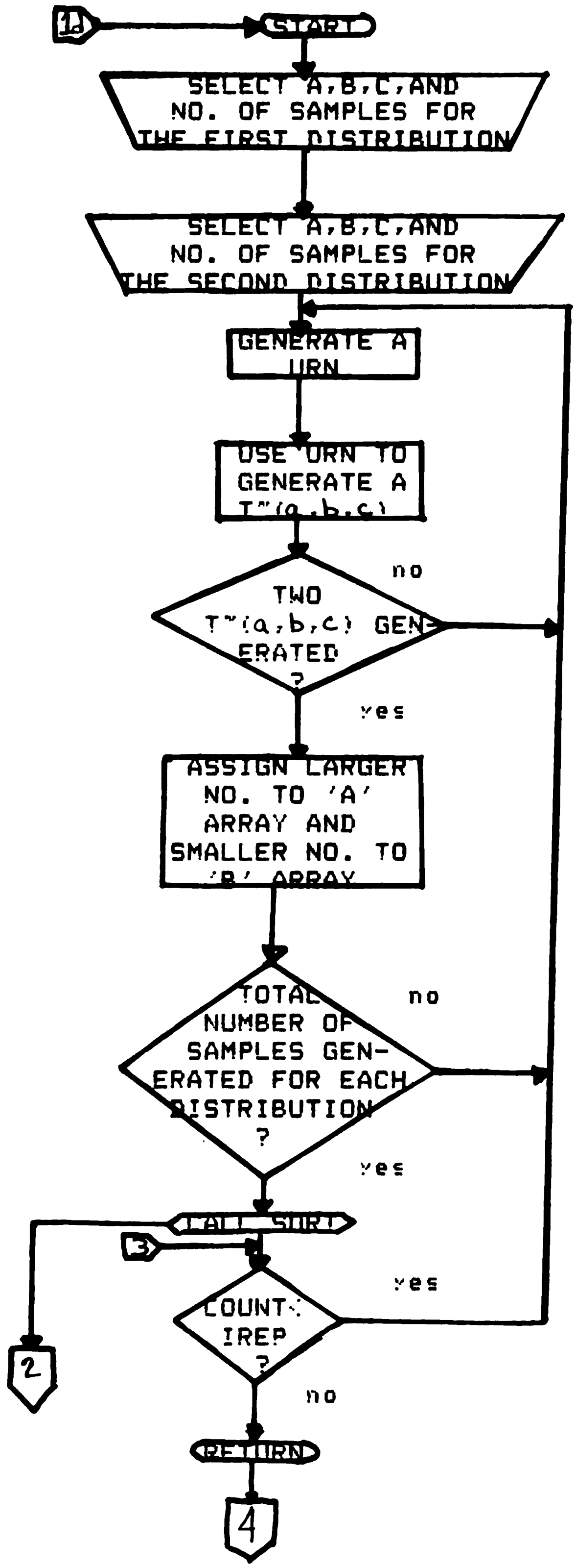
T REPORTS.LIS

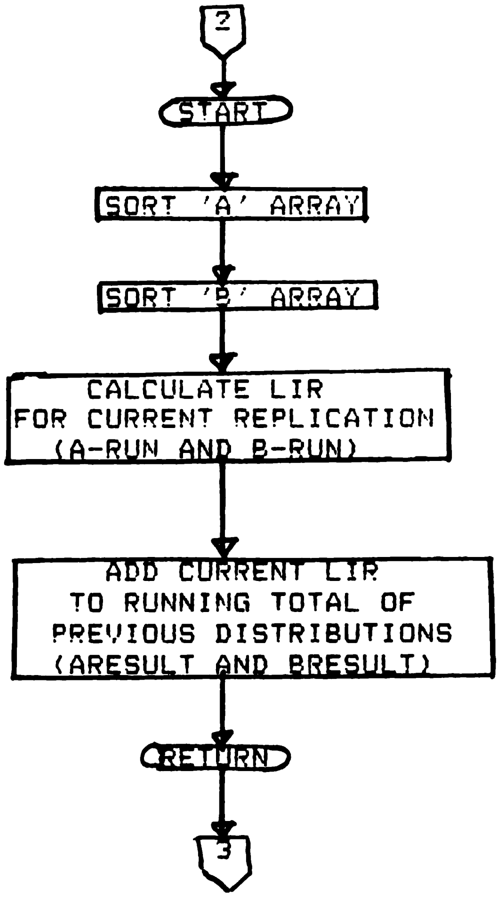

1a
START
SELECT MU,SIGMA,AND NO. OF SAMPLES FOR THE FIRST DISTRIBUTION
SELECT MU,SIGMA,AND NO. OF SAMPLES FOR THE SECOND DISTRIBUTION
GENERATE TWO UNIFORM RANDOM NUMBERS(URN)
URN'S ACCEPTABLE ?
no
yes
USING URN'S GENERATE TWO STANDARD NORMALS
CONVERT STANDARD NORMALS TO N~(μ,σ)
ASSIGN LARGER NO. TO 'A' ARRAY AND SMALLER NO. TO 'B' ARRAY
TOTAL NUMBER OF SAMPLES GENERATED FOR EACH DISTRIBUTION ?
no
yes
CALL SORT
3
COUNT< IREP ?
yes
no
2
RETURN
4



1b
START
SELECT ALPHA,BETA, AND NO. OF SAMPLES FOR THE FIRST DISTRIBUTION
SELECT ALPHA,BETA, AND NO. OF SAMPLES FOR THE SECOND DISTRIBUTION
GENERATE TWO URN'S
USE ONE URN TO GENERATE A G~(α,1), USE BOTH URN'S TO TEST FOR NEXT STEP
URN'S ACCEPTABLE ?
no
yes
USE G~(α,1) TO GENERATE A G(α,β)
HAVE TWO G~(α,β) BEEN GENERATED ?
no
yes
ASSIGN LARGER NO. TO 'A' ARRAY AND SMALLER NO. TO 'B' ARRAY
TOTAL NUMBER OF SAMPLES GENERATED FOR EACH DISTRIBUTION ?
no
yes
CALL SORT
B
2
yes
no
RETURN
4


START
SELECT ALPHA1,ALPHA2,AND NO. OF SAMPLES FOR THE FIRST DISTRIBUTION
SELECT ALPHA1,ALPHA2,AND NO. OF SAMPLES FOR THE SECOND DISTRIBUTION
GENERATE A G~(α₁,1) AND A G~(α₂,1)
USE THE PAIR OF GAMMAS TO GENERATE A B~(α₁,α₂)
ARE TWO BETAS GENERATED ?
no
yes
ASSIGN LARGER NO. TO 'A' ARRAY AND SMALLER NO. TO 'B' ARRAY
TOTAL NUMBER OF SAMPLES GENERATED FOR EACH DISTRIBUTION ?
no
yes
CALL SORT
2
yes
COUNT< IREP ?
no
RETURN
4


1a
START
SELECT A,B,C,AND NO. OF SAMPLES FOR THE FIRST DISTRIBUTION
SELECT A,B,C,AND NO. OF SAMPLES FOR THE SECOND DISTRIBUTION
GENERATE A URN
USE URN TO GENERATE A T~(a,b,c)
TWO T~(a,b,c) GENERATED ?
no
yes
ASSIGN LARGER NO. TO 'A' ARRAY AND SMALLER NO. TO 'B' ARRAY
TOTAL NUMBER OF SAMPLES GENERATED FOR EACH DISTRIBUTION ?
no
yes
CALL SORT
3
COUNT< IREP ?
yes
no
2
RETURN
4


2
START
SORT 'A' ARRAY
SORT 'B' ARRAY
CALCULATE LIR
FOR CURRENT REPLICATION
(A-RUN AND B-RUN)
ADD CURRENT LIR
TO RUNNING TOTAL OF
PREVIOUS DISTRIBUTIONS
(ARESULT AND BRESULT)
RETURN
3

```
*************************MAIN PROGRAM****************************************

      COMMON/ALG/A(50),B(50),SUM
      COMMON/ALL/NFILE,IREP,ITEM
      COMMON/OUT/ARESULT(50),BRESULT(50),PERCENT(50)
      COMMON/FILE/NAME,DISP
      CALL INITIALIZE(DIST)
 19   GOTO(1,2,3,4),DIST
  1   CALL NORMAL
      GOTO 10
  2   CALL GAMMA
      GOTO 10
  3   CALL BETA
      GOTO 10
  4   CALL TRIANG
 10   CALL REPORT
      STOP
      END

*********************************************************************
      *                                                   *
      *  This subroutine inputs the type of probability   *
      *   distribution and the number of replications.    *
      *                                                   *
*********************************************************************

      SUBROUTINE INITIALIZE(DIST)
      COMMON/ALL/NFILE,IREP,ITEM
      COMMON/FILE/NAME,DISP
      CHARACTER*4 LETTER,CHOICE
      CHARACTER*13 NAME,DISP
 35   WRITE(6,31)
 31   FORMAT(T2,'Output may be written to a file and then printed ',
     /'or printed at this terminal.',/,T2,'Enter "F" for file or "T"',
     /  ' for terminal.')
      READ(5,79)CHOICE
      IF(CHOICE.EQ.'F')THEN
       NFILE=8
      ELSE IF(CHOICE.EQ.'T')THEN
       NFILE=6
       GOTO 3
      ELSE
       GOTO 35
      END IF
      WRITE(6,12)
 12   FORMAT(T2,'Enter the name of the output file.')
      READ(5,17)NAME
 17   FORMAT(A)
      WRITE(6,21)
 21   FORMAT(T2,'Select the disposition of the output file.',
     /  /,T2,'Type: KEEP,PRINT,or PRINT/DELETE.')
      READ(6,27)DISP
 27   FORMAT(A)
      OPEN(8,FILE=NAME,STATUS='NEW',DISP=DISP)
  3   WRITE(6,72)
 72   FORMAT(T2,'Which type of test is to be performed?',/,T2,
     /  'Enter "ARUN" if you wish to know the length of the A  ',
     /  'run before the first B.',/,' Enter "BRUN" if you wish the',
```

```
     /  ' length of the B run before the first A.',/,T2,'Enter "BOTH"'
     /  ' if you would like the results of both tests.')
        READ(5,79)TEST
  79    FORMAT(A)
        IF(TEST.EQ.'ARUN')THEN
         ITEM=1
        ELSE IF(TEST.EQ.'BRUN')THEN
         ITEM=2
        ELSE IF(TEST.EQ.'BOTH')THEN
         ITEM=3
        ELSE
         GOTO 3
        END IF
  13    WRITE(6,10)
  10    FORMAT(T2,'Select the type of distribution you would ',
     /  'like.',/,' Type the first letter of the name to make ',
     /  'your input.')
        READ(5,23)LETTER
  23    FORMAT(A)
        IF(LETTER.EQ.'N')THEN
         DIST=1
        ELSE IF(LETTER.EQ.'G')THEN
         DIST=2
        ELSE IF(LETTER.EQ.'B')THEN
         DIST=3
        ELSE IF(LETTER.EQ.'T')THEN
         DIST=4
        ELSE
         GOTO 13
        END IF
 92     WRITE(6,93)
 93     FORMAT(T2,'Enter the number of replications ',
     /  'to be performed.')
 91     READ(5,*)IREP
        IF(IREP.GT.100000)GOTO 92
        RETURN
        END


**********************************************************************
      *                                                      *
      *    This subroutine generates random numbers          *
      *         from the normal distribution.                *
      *                                                      *
**********************************************************************

        SUBROUTINE NORMAL
        DIMENSION V(2),MU(2),SIGMA(2),NSAMP(2)
        COMMON/ALG/A(50),B(50),SUM
        COMMON/ALL/NFILE,IREP,ITEM
        COMMON/FILE/NAME,DISP
        CHARACTER*13 NAME,DISP
        REAL MU
        INTEGER COUNT,I1,SUM
        COUNT=0
        IF(NFILE.EQ.6)GOTO 6
        OPEN(8,FILE=NAME,STATUS='OLD',DISP=DISP)
  6     PRINT 5
  5     FORMAT(1X,'Enter mu, sigma and number of samples',
     /  ' for the',/,T2,
     /  'first normal distribution separated by commas.')
```

```
      READ(5,*)MU(1),SIGMA(1),NSAMP(1)
      WRITE(6,17)
17    FORMAT(T2,'Enter mu, sigma and number of samples',
     /  ' for the',/,T2,
     /  'second normal distribution separated by commas.')
      READ(5,*)MU(2),SIGMA(2),NSAMP(2)
      WRITE(6,19)
19    FORMAT(T2,'Enter a five-digit number for random',
     /  ' number generation.')
      READ(5,*)I1
      WRITE(NFILE,98)MU(1),SIGMA(1),MU(2),SIGMA(2),NSAMP(1),
     /  NSAMP(2),I1
98    FORMAT(T2,'Input distributions are N~(',F5.2,',',F5.2,') and',
     /  ' N~(',F5.2,',',F5.2,').',/,T2,'Number of samples from each ',
     /  'distribution is ',I2,' and ',
     /  I2,' respectively.',/,T2,'The random number seed for this ',
     /  'run is ',I6,'.')
97    COUNT=COUNT+1
      J=0
      SUM=0
100   J=J+1
      N=0
90    T=0
      DO 80 I=1,2
      X=RAN(I1)
      V(I)=2.*X-1.
      W=V(I)*V(I)
      T=T+W
80    CONTINUE
      IF(T.GT.1.)GO TO 90
      Y1=LOG(T)
      Y2=-2.*Y1/T
      Y=SQRT(Y2)
      N=N+1
      SUM=SUM+1
      X1=MU(J)+SIGMA(J)*V(1)*Y
      X2=MU(J)+SIGMA(J)*V(2)*Y
      IF(X1.GT.X2)THEN
      A(SUM)=X1
      B(SUM)=X2
      ELSE
      A(SUM)=X2
      B(SUM)=X1
      ENDIF
      IF(N.LT.NSAMP(J))GO TO 90
      IF(J.LT.2)GO TO 100
      CALL SORT
      IF(COUNT.LT.IREP)GOTO 97
      RETURN
      END
*********************************************************************
      *                                                         *
      *    This subroutine generates random numbers             *
      *          from the gamma distribution.                   *
      *                                                         *
*********************************************************************

      SUBROUTINE GAMMA
      DIMENSION U(2),X(2),ALPHA(2),BETA(2),NSAMP(2)
```

```
      COMMON/ALG/A(50),B(50),SUM
      COMMON/ALL/NFILE,IREP,ITEM
      COMMON/FILE/NAME,DISP
      CHARACTER*13 NAME,DISP
      INTEGER COUNT,I1,SUM
      REAL LTHETA,LNZ
      COUNT=0
      IF(NFILE.EQ.6)GOTO 6
      OPEN(8,FILE=NAME,STATUS='OLD',DISP=DISP)
 6    PRINT 10
10    FORMAT(1X,'Enter alpha, beta and number of samples for',
   /  ' the',/,T2,
   /  'first gamma distribution separated by commas.')
      READ(5,*)ALPHA(1),BETA(1),NSAMP(1)
      IF(ALPHA(1).LE.1)GOTO 6
 8    PRINT 11
11    FORMAT(1X,'Enter alpha, beta and number of samples for',
   /  ' the',/,T2,
   /  'second gamma distribution separated by commas.')
      READ(5,*)ALPHA(2),BETA(2),NSAMP(2)
      IF(ALPHA(2).LE.1)GOTO 8
      PRINT 12
12    FORMAT(T2,'Enter a five-digit number for random',
   /  ' number generation.')
      READ(5,*)I1
      WRITE(NFILE,98)ALPHA(1),BETA(1),ALPHA(2),BETA(2),NSAMP(1),
   /  NSAMP(2),I1
98    FORMAT(T2,'Input distributions are G~(',F5.2,',',F5.2,') and',
   /  ' G~(',F5.2,',',F5.2,').',/,T2,'Number of samples from ',
   /  'each distribution is ',I2,' and ',
   /  I2,' respectively.',/,T2,'The random number seed for this ',
   /  'run is ',I6,'.')
      THETA=4.5
      LTHETA=LOG(4.5)
      D=1.+LTHETA
      FOUR=LOG(4.)
97    COUNT=COUNT+1
      SUM=0
      K=0
60    K=K+1
      G1=2.*ALPHA(K)-1.
      G2=SQRT(G1)
      G=1./G2
      Q=ALPHA(K)+G2
      H=ALPHA(K)-FOUR
      N=0
80    J=0
20    J=J+1
50    DO 30 I=1,2
      R=RAN(I1)
      U(I)=R
30    CONTINUE
      V1=U(1)/(1.-U(1))
      V2=LOG(V1)
      V=G*V2
      Y1=EXP(V)
      Y=ALPHA(K)*Y1
      Z=U(2)*U(1)*U(1)
      W=H+Q*V-Y
      TEST=W+D-THETA*Z
      IF(TEST.GE.0.)GO TO 40
```

```
      LNZ=LOG(Z)
      IF(W.GE.LNZ)GO TO 40
      GO TO 50
40    X(J)=Y*BETA(K)
      IF(J.LT.2)GO TO 20
      SUM=SUM+1
      N=N+1
      IF(X(1).LT.X(2))THEN
        A(SUM)=X(2)
        B(SUM)=X(1)
      ELSE
        A(SUM)=X(1)
        B(SUM)=X(2)
      END IF
      IF(N.LT.NSAMP(K))GO TO 80
      IF(K.LT.2)GO TO 60
      CALL SORT
      IF(COUNT.LT.IREP)GOTO 97
      RETURN
      END

**************************************************************************
          *                                                     *
          *    This subroutine generates random numbers        *
          *          from the beta distribution.               *
          *                                                     *
**************************************************************************

      SUBROUTINE BETA
      DIMENSION X1(2),X2(2),U(2),ALPHA(2,2),NSAMP(2)
      COMMON/ALG/A(50),B(50),SUM
      COMMON/ALL/NFILE,IREP,ITEM
      COMMON/FILE/NAME,DISP
      CHARACTER*13 NAME,DISP
      REAL LTHETA,LNZ
      INTEGER COUNT,I1,N,O,SUM
      COUNT=0
      IF(NFILE.EQ.6)GOTO 3
      OPEN(8,FILE=NAME,STATUS='OLD',DISP=DISP)
3     PRINT 10
10    FORMAT(1X,'Enter alpha1, alpha2 and number of samples ',
     / 'for the',/,T2,
     / 'first beta distribution separated by commas.')
      READ(5,*)ALPHA(1,1),ALPHA(2,1),NSAMP(1)
      PRINT 11
11    FORMAT(1X,'Enter alpha1, alpha2 and number of samples ',
     / 'for the',/,T2,
     / 'second beta distribution separated by commas.')
      READ(5,*)ALPHA(1,2),ALPHA(2,2),NSAMP(2)
      PRINT 12
12    FORMAT(T2,'Enter a five-digit number for random',
     / ' number generation.')
      READ(5,*)I1
      WRITE(NFILE,98)ALPHA(1,1),ALPHA(2,1),ALPHA(1,2),ALPHA(2,2),
     / NSAMP(1),NSAMP(2),I1
98    FORMAT(T2,'Input distributions are B~(',F5.2,',',F5.2,') and',
     / ' B~(',F5.2,','F5.2,').',/,T2,'Number of samples from',
     / ' each distribution is ',I2,' and ',
     / I2,' respectively.',/,T2,'The random number seed for this ',
```

```
     /   'run is ',I6,'.')
         THETA=4.5
         LTHETA=LOG(4.5)
         D=1.+LTHETA
         FOUR=LOG(4.)
   97    COUNT=COUNT+1
         O=0
         SUM=0
         DO 90 N=1,2
         O=O+1
         DO 80 M=1,NSAMP(N)
         SUM=SUM+1
         DO 70 L=1,2
         DO 60 K=1,2
         G1=2.*ALPHA(K,N)-1.
         G2=SQRT(G1)
         G=1./G2
         Q=ALPHA(K,N)+G2
         H=ALPHA(K,N)-FOUR
    50   DO 30 I=1,2
         R=RAN(I1)
         U(I)=R
    30   CONTINUE
         V1=U(1)/(1.-U(1))
         V2=LOG(V1)
         V=G*V2
         Y1=EXP(V)
         Y=ALPHA(K,N)*Y1
         Z=U(2)*U(1)*U(1)
         W=H+Q*V-Y
         TEST=W+D-THETA*Z
         IF(TEST.GE.0.)GO TO 40
         LNZ=LOG(Z)
         IF(W.GE.LNZ)GO TO 40
         GO TO 50
   40    X1(K)=Y
   60    CONTINUE
         X2(L)=X1(1)/(X1(1)+X1(2))
   70    CONTINUE
         IF(X2(1).GT.X2(2))THEN
           A(SUM)=X2(1)
           B(SUM)=X2(2)
         ELSE
           A(SUM)=X2(2)
           B(SUM)=X2(1)
         ENDIF
   80    CONTINUE
   90    CONTINUE
         CALL SORT
         IF(COUNT.LT.IREP)GOTO 97
         RETURN
         END
```

```
**********************************************************************
         *                                                  *
         *   This subroutine generates random numbers       *
         *      from the triangular distribution.           *
         *                                                  *
**********************************************************************
```

```
      SUBROUTINE TRIANG
      DIMENSION X(2),G(2),H(2),C(2),NSAMP(2)
      COMMON/ALG/A(50),B(50),SUM
      COMMON/ALL/NFILE,IREP,ITEM
      COMMON/FILE/NAME,DISP
      CHARACTER*13 NAME,DISP
      INTEGER COUNT,I1,SUM
      COUNT=0
      IF(NFILE.EQ.6)GOTO 9
      OPEN(8,FILE=NAME,STATUS='OLD',DISP=DISP)
 9    PRINT 50
 50   FORMAT(1X,'Enter minimum, maximum, mode and number of ',
     / ' of samples from the',/,T2,
     / 'first triangular distribution separated by commas.')
      READ(5,*)G(1),H(1),C(1),NSAMP(1)
 13   PRINT 51
 51   FORMAT(1X,'Enter minimum, maximum, mode and number of ',
     / ' of samples from the',/,T2,
     / 'second triangular distribution separated by commas.')
      READ(5,*)G(2),H(2),C(2),NSAMP(2)
      IF(G(1).GT.H(1).OR.G(1).GT.C(1).OR.C(1).GT.H(1))THEN
       WRITE(6,7)
 7     FORMAT(T2,'INCORRECT PARAMETERS ON  DISTRIBUTION.',
     / '  TRY AGAIN.')
       GOTO 9
      ELSE IF(G(2).GT.H(2).OR.G(2).GT.C(2).OR.C(2).GT.H(2))THEN
       WRITE(6,7)
       GOTO 13
      ELSE
       PRINT 52
 52    FORMAT(T2,'Enter a five-digit number for random number',
     / ' generation.')
       READ(5,*)I1
      END IF
      WRITE(NFILE,98)G(1),H(1),C(1),G(2),H(2),C(2),NSAMP(1),
     / NSAMP(2),I1
 98   FORMAT(T2,'Input distributions are',
     / ' T~(',F5.2,',',F5.2,',',F5.2,') and',
     / ' T~(',F5.2,',',F5.2,',',F5.2,').',
     / /,T2,'Number of samples from each distribution is ',
     / I2,' and ',I2,' respectively.',
     / /,T2,'The random number seed for this run is ',I6,'.')
 97   COUNT=COUNT+1
      SUM=0
      DO 30 I=1,2
      T=H(I)-G(I)
      COMP=(C(I)-G(I))/T
      COMP1=1.-COMP
      DO 20 L=1,NSAMP(I)
      SUM=SUM+1
      DO 10 J=1,2
      U=RAN(I1)
      IF(U.LE.COMP)THEN
        X1=COMP*U
        X2=SQRT(X1)
        X(J)=G(I)+T*X2
      ELSE
        X1=COMP1*(1.-U)
        X2=SQRT(X1)
        X3=1.-X2
```

```
         X(J)=G(I)+T*X3
       ENDIF
 10    CONTINUE
       IF(X(1).GT.X(2))THEN
         A(SUM)=X(1)
         B(SUM)=X(2)
       ELSE
         A(SUM)=X(2)
         B(SUM)=X(1)
       ENDIF
 20    CONTINUE
 30    CONTINUE
       CALL SORT
       IF(COUNT.LT.IREP)GOTO 97
       RETURN
       END
```

```
************************************************************************
      *                                                               *
      *         This subroutine performs a bubble sort.               *
      *                                                               *
************************************************************************
```

```
       SUBROUTINE SORT
       DIMENSION AA(2),BB(2)
       COMMON/ALG/A(50),B(50),SUM
       COMMON/ALL/NFILE,IREP,ITEM
       COMMON/OUT/ARESULT(50),BRESULT(50),PERCENT(50)
       COMMON/FILE/NAME,DISP
       CHARACTER*13 NAME,DISP
       INTEGER ARUN,BRUN,SUM
 12    INTERA=1
       DO  I=1,(SUM-1)
        IF(A(I).GE.A(I+1))GOTO 25
         AA(2)=A(I)
         AA(1)=A(I+1)
         A(I)=AA(1)
         A(I+1)=AA(2)
         INTERA=0
 25     CONTINUE
       END DO
       IF(INTERA.EQ.0)GOTO 12
  9    INTERB=1
       DO J=1,(SUM-1)
        IF(B(J).GE.B(J+1))GOTO 27
         BB(2)=B(J)
         BB(1)=B(J+1)
         B(J)=BB(1)
         B(J+1)=BB(2)
         INTERB=0
 27     CONTINUE
       END DO
       IF(INTERB.EQ.0)GOTO 9
       ARUN=0
       BRUN=0
       DO I=1,SUM
        IF(A(I).GT.B(1))THEN
         ARUN=ARUN+1
```

```
         ELSE
          GOTO 7
         END IF
        END DO
 7      CONTINUE
        DO I=SUM,1,-1
         IF(B(I).LT.A(SUM))THEN
          BRUN=BRUN+1
         ELSE
          GOTO 18
         END IF
        END DO
 18     CONTINUE

              **************************************
              *   End of replication bookkeeping:  *
              * This part of the subroutine stores *
              * the statistic "# of runs of length *
              *       X" in the array RESULT(X)    *
              **************************************

        DO I=1,SUM
         IF(ARUN.EQ.I)ARESULT(I)=ARESULT(I)+1
         IF(BRUN.EQ.I)BRESULT(I)=BRESULT(I)+1
        END DO
        RETURN
        END

**********************************************************************
       *                                                      *
       *       This subroutine  generates the output of       *
       *                the simulation.                       *
       *                                                      *
**********************************************************************

        SUBROUTINE REPORT
        DIMENSION AA(2),BB(2),CUM(50),INT(50)
        COMMON/ALG/A(50),B(50),SUM
        COMMON/ALL/NFILE,IREP,ITEM
        COMMON/OUT/ARESULT(50),BRESULT(50),PERCENT(50)
        COMMON/FILE/NAME,DISP
        CHARACTER*13 NAME,DISP,ANSWER
        INTEGER SUM
        IF(NFILE.EQ.6)GOTO 11
        OPEN(8,FILE=NAME,STATUS='OLD',DISP=DISP)
 11     IF(ITEM.EQ.1.OR.ITEM.EQ.3)THEN
         ASSIGN 37 TO IOUT
        ELSE
         ASSIGN 39 TO IOUT
        END IF
        WRITE(NFILE,IOUT)
 37     FORMAT(///,T2,'Test for the Length of the Initial Run of',
     /  ' A''s before the first B.',/)
 39     FORMAT(///,T2,'Test for the Length of the Initial Run of ',
     /  'B''s before the first A.',/)
        WRITE(NFILE,13)IREP
 13     FORMAT(T2,'Number of Replications: ',I6,/)
        IF(ITEM.EQ.1.OR.ITEM.EQ.3)THEN
         WRITE(NFILE,42)
 42      FORMAT(T2,'Length ',4X,'Observed  ',4X,
```

```
      /    'Relative   ',4X,'Cumulative')
           WRITE(NFILE,44)
    44     FORMAT(T2,'of Run:',4X,'Frequency:',4X,
      /    'Frequency:',4X,'Frequency:')
           T=IREP
           CUM(0)=0.0
           DO I=1,SUM
            PERCENT(I)=ARESULT(I)/T
            CUM(I)=CUM(I-1)+PERCENT(I)
           END DO
           DO I=1,SUM
            INT(I)=ARESULT(I)
            WRITE(NFILE,17)I,INT(I),PERCENT(I),CUM(I)
    17      FORMAT(T2,I3,T12,I6,T28,F8.6,T42,F8.6)
           END DO
           WRITE(NFILE,62)
    62     FORMAT(T2,/,/,/,/)
           CALL GRAPH
           IF(ITEM.EQ.1)GOTO 57
           ITEM=2
           GOTO 11
          ELSE
           WRITE(NFILE,42)
           WRITE(NFILE,44)
           T=IREP
           CUM(0)=0.0
           DO I=1,SUM
            PERCENT(I)=BRESULT(I)/T
            CUM(I)=CUM(I-1)+PERCENT(I)
           END DO
           DO I=1,SUM
            INT(I)=BRESULT(I)
            WRITE(NFILE,17)I,INT(I),PERCENT(I),CUM(I)
           END DO
           WRITE(NFILE,62)
           CALL GRAPH
          END IF
    57    CONTINUE
          RETURN
          END

*********************************************************************
          *                                                      *
          *     This subroutine generates the relative and       *
          *      cumulative frequency distribution               *
          *                 function graphs.                     *
          *                                                      *
*********************************************************************

          SUBROUTINE GRAPH
          DIMENSION TEST1(50),CUM(50)
          COMMON/ALG/A(50),B(50),SUM
          COMMON/ALL/NFILE,IREP,ITEM
          COMMON/OUT/ARESULT(50),BRESULT(50),PERCENT(50)
          COMMON/FILE/NAME,DISP
          CHARACTER*13 NAME,DISP,ANSWER
          INTEGER SUM
          IF(NFILE.EQ.6)GOTO 13
          OPEN(8,FILE=NAME,STATUS='OLD',DISP=DISP)
```

```
13     IF(ITEM.EQ.2)THEN
        ASSIGN 99 TO L
        ASSIGN 88 TO K
        ASSIGN 84 TO M
       ELSE
        ASSIGN 98 TO L
        ASSIGN 86 TO K
        ASSIGN 82 TO M
       END IF
       WRITE(6,L)
98     FORMAT(T2,'For this test of the length of the run of',
   /   ' A''s before the first B,',/,T2,
   /   'do you wish a bar graph of the relative and',
   /   ' cumulative frequency distributions?',/,T2,
   /   'Enter "YES" or "NO".')
99     FORMAT(T2,'For this test of the length of the run of',
   /   ' B''s before the first A,',/,T2,
   /   'do you wish a bar graph of the relative and',
   /   ' cumulative frequency distributions?',/,T2,
   /   'Enter "YES" or "NO".')
       READ(5,91)ANSWER
91     FORMAT(A)
       IF(ANSWER.EQ.'NO')GOTO 89
       WRITE(NFILE,27)
27     FORMAT('1')
       WRITE(NFILE,K)
86     FORMAT(T5,'Relative frequency graph',
   /   ' for the Length of the Initial Run of',
   /   ' A''s before the first B.',/)
88     FORMAT(T5,'Relative frequency graph',
   /   ' for the Length of the Initial Run of',
   /   ' B''s before the first A.',/)
       WRITE(NFILE,37)
37     FORMAT(T5,'0.00',T10,'.04',T14,'.08',T18,'.12',T22,'.16',
   /T26,'.20',T30,'.24',T34,'.28',T38,'.32',T42,'.36',T46,'.40',
   /T50,'.44',T54,'.48',T58,'.52',T62,'.56',T66,'.60',T70,'.64',
   /T74,'.68',T78,'.72',T82,'.76',T86,'.80',T90,'.84',T94,'.88',
   /T98,'.92',T102,'.96',T106,'1.00')
31     WRITE(NFILE,33)
33     FORMAT(T6,'+-+-+-+-+-+-+-+-+-+-',
   / '+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-',
   / '+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+')
       CUM(0)=0.0
       DO I=1,SUM
        TEST1(I)=(PERCENT(I)+.005)*1000.
        J=TEST1(I)/10.
        IF(J.EQ.0)THEN
         ASSIGN 43 TO IOUT
        ELSE
         ASSIGN 41 TO IOUT
        END IF
        CUM(I)=CUM(I-1)+PERCENT(I)
        WRITE(NFILE,IOUT)I,PERCENT(I)
41      FORMAT('0',T3,I2,T6,'-',<J>(1H+),2X,F8.3)
43      FORMAT('0',T3,I2,T6,'-',2X,F8.3)
       END DO
       WRITE(NFILE,27)
       WRITE(NFILE,M)
82     FORMAT(T5,'Cumulative frequency graph',
   /   ' for the Length of the Initial Run of',
   /   ' A''s before the first B.',/)
```

```
 84     FORMAT(T5,'Cumulative frequency graph',
    /   ' for the Length of the Initial Run of',
    /   ' B''s before the first A.',/)
        WRITE(NFILE,37)
        WRITE(NFILE,33)
        DO I=1,SUM
         TEST1(I)=(CUM(I)+.005)*1000.
         J=TEST1(I)/10.
         IF(J.EQ.0)THEN
          ASSIGN 43 TO IOUT
         ELSE
          ASSIGN 41 TO IOUT
         END IF
         WRITE(NFILE,IOUT)I,CUM(I)
        END DO
        WRITE(NFILE,27)
 89     CONTINUE
        RETURN
        END
```

****

APPENDIX IV

ACKNOWLEDGEMENTS

Thanks are due to Ms. Kimberly S. Druschel, Mr. Ronald L. Fischer, and others at Fort Lee, VA, for support and helpful conversations.

APPENDIX V

REFERENCES

1. Downs and Cox, "The Probability of Motor Case Rupture," presented at the 20th Conference on Design of Experiments in Army R,D&T, October 1974.

2. Knaub, "Design of a Multiple Sample Westenberg Type Test for Small Sample Sizes," presented at the 27th Conference on Design of Experiments in Army R,D&T, October 1981.

3. Law and Kelton, Simulation Modeling and Analysis, McGraw-Hill Inc., 1982.

ADDENDUM

After listening to the presentation of this paper, Dr. W. J. Conover of Texas Tech, commented that perhaps a rank test based on the rank sum of the A's or some other appropriate measure might be used as a more powerful test of overall compatibility among the samples with emphasis on shift of location. In view of the weakness of the modified Westenberg test for the median which is given here, and the generally high power of rank tests, this suggestion seems promising. I would still, however, suggest the LIR test for the purpose for which it is intended: It emphasizes the improvement which can be obtained by exclusion of a particular sample. However, beware of repeated deletions. The probability levels of the test change step-wise with each application.

IN-MINEFIELD EFFECTIVENESS MEASURE FOR BREACHING VEHICLES

Mark S. Adams
Systems Employment Branch, Countermine Laboratory
US Army Mobility Equipment Research and Development Command
Fort Belvoir, Virginia

I INTRODUCTION

The development of realistic models is required to assess the military worth of countermine systems in mine warfare scenarios. Explicit closed form solutions delineating countermine equipment effectiveness are being developed to become modular components of a more complex war game modelling mine warfare.

This report develops a closed solution to measure the effectiveness of armored vehicles proceeding through cleared lanes. An equation is derived to determine the expected number of mines a vehicle will encounter in a scenario. The expected number of mine encounters is used to calculate a measure to compare the value of changes in tactical methods and countermine materiel.

A discussion of the applicability of the effectiveness measure to support mine and countermine studies is also presented. A set of mine warfare situations are formulated as an example of the ease of using the expression derived in this report.

II BACKGROUND

In June 1980, USAES requested MERADCOM to perform analyses determining what marking systems could be used with mine-clearing rollers. MERADCOM tasked HEL who, in turn, subcontracted Armament Systems, Inc. and the final report, "An Investigation of Requirements for Cleared-Lane Marking Systems (CLAMS) for Hasty Breaching of Minefields with Mine-Clearing Rollers," was completed in March 1981. Section 5.0 of this report, "Assessment of the Problems Associated with Traversing and Marking a Minefield," thoroughly discusses doctrine, literature and field test data and reports on requirements of the width of a cleared lane. The requirements established in sources such as FM 90-7 "Obstacles" (which states a 4 meter wide vehicle assault lane may be used for the hasty breach), and, the final report by the USA Armor and Engineering Board of OT II testing of the Mine-Clearing Roller (that notes the width of the lane cleared by one tank with a roller is inadequate to allow safe tracking by other tanks and personnel carriers), are not driven by , or directly related to, mobility mission requirements or effectiveness. Some procedure is necessary to aid translating mission requirements into equipment performance requirements and the converse. With the expression developed here, postulated systems performance functions can generate values of in-minefield effectiveness as the measure of the expected number of mines a vehicle will encounter.

III OBJECTIVE AND SCOPE

The objective of this report is to derive an equation to calculate expected mine encounters of vehicles crossing a minefield, and describe methods in computing and applications of results of the equation to determine in-minefield effectiveness (IME) measures of various systems.

The mathematical scope of the derivation extends to an integral calculus statement:

$$\int_{AREA} (\text{density function})\, d(AREA) = \text{units} \tag{1}$$

For the application here, this translates to: the integral of the density function in mines per square meter over the area swept out by a vehicle passing through a minefield is equal to the number of mines the vehicle will encounter. This would be exactly true if mines were a continuous phenomenon. But since they are point located, or at best disjoint, this equation is an approximation to the expected number of mine encounters. This method does not calculate where a vehicle will encounter a mine, only the expected number of encounters that are found in the area used in the calculation. This expected value is the identical concept to the average, or mean value.

Actual mine location will vary under a host of conditions (mine laying procedure, minefield layout, etc.), so there is a possiblity that, though the expected value of encounters in an intended path is positive, actually transversing the minefield will result in no mine encounters.

Under rather standard assumptions of randomness, the occurrance of mine encounters can be treated as a Poisson process. Arguements to independent and identical distributions are not very serious because, firstly, the calculation derived is an approximation, and secondly, the performance measure relates to the probability of no occurances, so the memoryless criteria of such a process is robust. According to a Poisson distribution, the probability of n occurances, Pr(n), given that the expected number of occurances is λ, is given by equation 2.

$$Pr(n) = \frac{\lambda^n e^{-\lambda}}{n!}, \tag{2}$$

The probability of encountering no mines (n=0) given the expected number of encounters, λ, to be E(N) is

$$Pr(0 \text{ mines}) = e^{-E(N)} \tag{3}$$

This probability is the in-minefield effectiveness measure.

Included in the IME equation derived within is a parameterized function that describes the vehicle path and a probability distribution function of random play about the path. The resulting product of functions provides a flexibility in the quantifiable description of countermining situations. Moreover, the results are immediately and inexpensivly obtained compared to simulations and war games of similar scope.

IV DERIVATION

Let $\delta(x,y)$ be the minefield density function over some Cartesian coordinate system; typical units for such a function are mines per meter squared (M/m^2). Take, for example, a situation at FIG 1. In this example, a mechanically emplaced minefield of 3 rows of pressure AT mines with an inter-mine spacing of four meters would require 300 mines. Since the area of this minefield is 20,000 square meters, the overall density of the minefield is $0.015 M/m^2$. An alternate representation of this minefield could partition the mine rows into separate minefields. Using the variables defined in the figure, the expected number of mines, E(N), a vehicle would encounter is $2 \cdot (\text{ote}-\text{ite}) \cdot d \cdot \delta$. The difference, in meters, from the <u>o</u>uter <u>t</u>rack <u>e</u>dge to the <u>i</u>nner <u>t</u>rack <u>e</u>dge (ote-ite) is one track-width. For this example, $E(N)=2\cdot(\text{ote}-\text{ite})\cdot 50\cdot 0.015=1.5\cdot(\text{ote}-\text{ite})$ or 1.5 times a track-width of the vehicle (in meters). This calculation is simply the area swept out by the vehicles tracks times the constant minefield density. In integral form, however,

$$E(N) = \int_0^d \left(\int_{w/2-\text{ote}}^{w/2-\text{ite}} \delta(x,y)\, dx + \int_{w/2+\text{ite}}^{w/2+\text{ote}} \delta(x,y)\, dx \right) dy \qquad (4)$$

Since $\delta(x,y) = 0.015\ M/m^2$ for $0 \le x \le w$, $0 \le y \le d$, we have,

$$E(N) = \int_0^d 0.015 \left(\int_{w/2-\text{ote}}^{w/2-\text{ite}} dx + \int_{w/2+\text{ite}}^{w/2+\text{ote}} dx \right) dy$$

$$= \int_0^d ((0.015\cdot 2)(\text{ote}-\text{ite}))\,dy$$

$$= 0.030(\text{ote}-\text{ite})d$$

$$= (1.5)(\text{ote}-\text{ite}) \qquad (5)$$

Nomenclature for the IME Equation

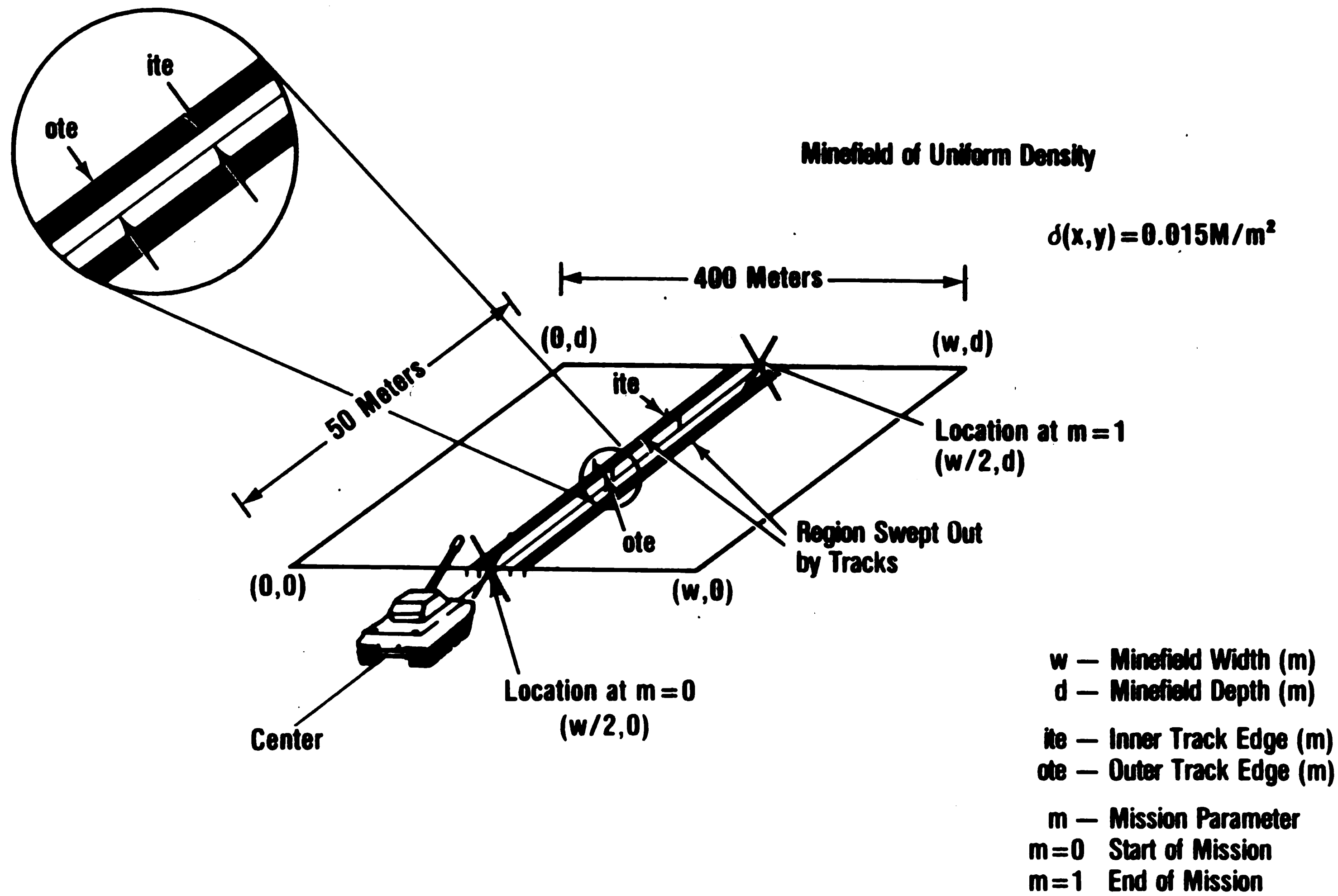


Figure 1. STRAIGHT LINE PATH THROUGH A MINEFIELD

This is the identical result obtained eariler.

Consider now that the vehicle goes through the minefield on something other than a straight line path. Define $P(m) = (X(m),Y(m))$ as the parametrical representation of such a path, where m ranging from 0 to 1 relates to the amount of the mission complete (i.e., m=0 is the start of the mission, m=1 the end). The areas swept out by tracks of a vehicle are no longer simple rectangles, as shown at FIG 2. Incorporating this parametrical path representation, equation (3) becomes

$$E(N) = \int_0^1 \left(\int_{olx}^{ilx} \delta(x,y)(dY/dm)dx + \int_{irx}^{orx} \delta(x,y)(dY/dm)dx \right) dm$$

where olx = outer left-track x = $X(m) - ote/\sqrt{1+s^2(m)}$

ilx = inner left-track x = $X(m) - ite/\sqrt{1+s^2(m)}$

irx = inner right-track x = $X(m) + ite/\sqrt{1+s^2(m)}$

orx = outer right-track x = $X(m) + ote/\sqrt{1+s^2(m)}$

$y = Y(m) + s^2(m)(X(m)-x)$

$$s(m) = \frac{-dX/dm}{dY/dm} \tag{6}$$

Analytically, this integral represents the summation of partitioned rectangles that make up the area traced by the vehicle tracks, as shown in FIG 3.

Thus far, vehicles are restricted to following the prespecified paths perfectly. A probability distribution function of play about the path is inserted to account for effects of vehicles not able to follow precise paths. This play function, $\Omega(z,m)$, is dependent upon distance from the path, z, and the mission parameter measure, m. The final form for the expected number of mine encounters is

$$E(N) = \int_0^1 \int_{-\infty}^{+\infty} \Omega(z,m) \left(\int_{olx}^{ilx} \delta(x,y)(dY/dm)dx + \int_{irx}^{orx} \delta(x,y)(dY/dm)dx \right) dz\, dm$$

Projection of Non-Perpendicular Orientations onto the X-Axis

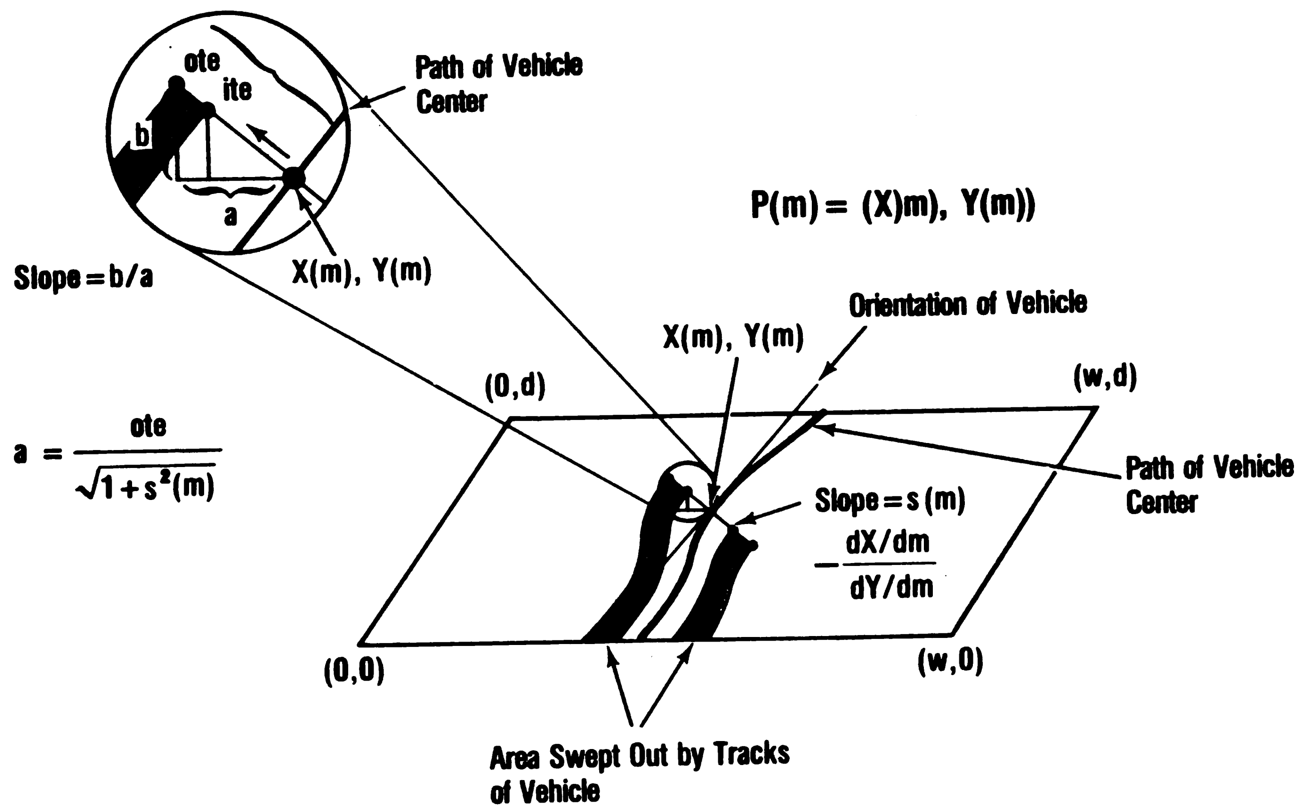


Figure 2. NON-LINEAR PATH THROUGH A MINEFIELD

Partitioning of the Area Integrated by the IME Equation

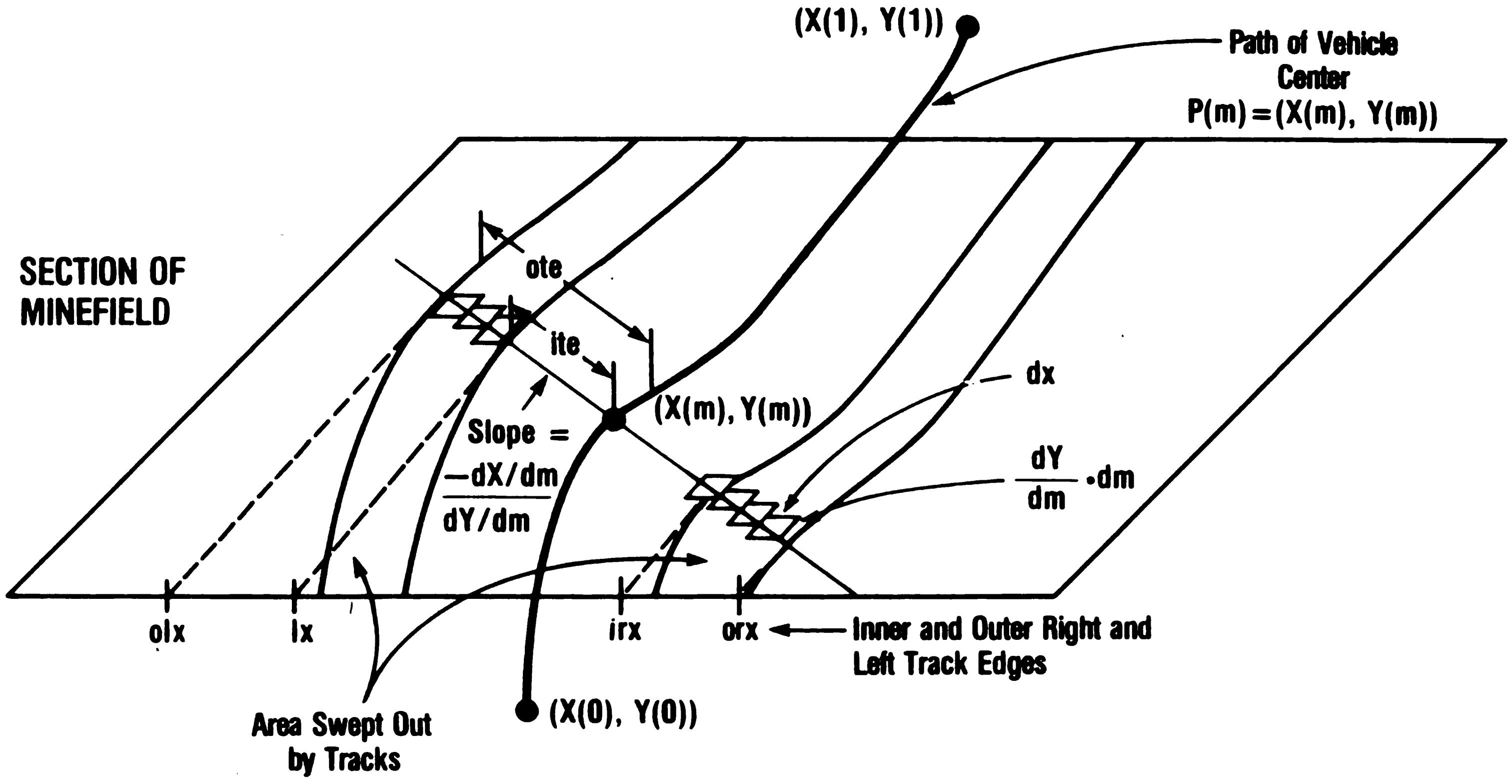


Figure 3. GRAPH OF PARAMETERIZED PATH THROUGH A MINEFIELD

where $olx = X(m) - (ote-z)/\sqrt{1+s^2(m)}$

$ilx = X(m) - (ite-z)/\sqrt{1+s^2(m)}$

$irx = X(m) + (ite+z)/\sqrt{1+s^2(m)}$

$orx = X(m) + (ote+z)/\sqrt{1+s^2(m)}$

y, s(m) as defined in equation (6). (7)

Once E(N) is calculated, the probability of the vehicle traversing the path P(m) with play function $\Omega(z,m)$ through a minefield density $\delta(x,y)$ without encountering a mine is given by equation (3):

$$Pr(0 \text{ mines}) = e^{-E(N)}$$

The Pr(0 mines) is the in-minefield effectiveness (IME) measure for breaching vehicles. Being a probability, the IME measure will always fall between 0 and 1, the latter designating a certainty of no mine encounters. The closer the IME is to one, the better the chances a vehicle will successfully cross the minefield.

V APPLICATION

This section presents sample applications of the IME measure equation. The intent of this section is to demonstrate the flexibility of the model; it is not an exhaustive list of the capability of the procedure. The situations to be studied will dictate the forms of three functions: minefield density, vehicle path, and path play.

Let us choose as a problem measuring the ability of follow-on vehicles to breach a minefield first cleared by a single lead tank equipped with a mine-clearing roller and cleared lane marking system.

A minefield density function to represent this situation must be formulated. Consider the mechanically emplaced minefield of 300 mines in a 400 meter by 50 meter rectangle discussed earlier. Before any neutralization, the density could be taken as a constant $0.015M/m^2$. As the lead tank equipped with a roller passes through the minefield and detonates mines, the density of the minefield is lowered. Assuming perfect capabilities, and a straight line breach perpendicularly through the center of the minefield, the density of the minefield drops to zero within the two rectangles traced out by the signature of the roller banks. A representative minefield density function before and after nuetralization is shown at FIG 4. The after-neutralization minefield density function is used to determine the expected number of mines encountered by follow-on vehicles. The algebraic expression of this density function is

$$\delta(x,y) = \begin{cases} 0.0 & 0.9 \leq |x-200| < 2.0 \\ 0.015 & \text{otherwise, } 0 \leq x \leq 400,\ 0 \leq y \leq 50 \end{cases} \tag{8}$$

Physically, equation (8) models a lead tank that crossed the minefield at mid-front (x=200 of the 400 m minefield), with 2 mine-clearing roller banks 1.1 m wide separated by 1.8 meters.

The second of the three IME functions, the vehicle path, models the attempted path of follow-on vehicles making full advantage of neutralized zones. In this instance, the intended path for such vehicles is to retrace the straight line path of the lead tank guided by some marking system. The parameterized form of this path is

$$P(m) = (X(m), Y(m)) : \begin{array}{l} X(m) = 200 \\ Y(m) = 50 \cdot m \end{array} \tag{9}$$

At the start of the mission, the vehicle is at P(0) which is (200, 0) on the Cartesian system employed. By the end of the mission (m=1), the vehicle has travelled in a straight line to (200, 50); this is the same path as the roller equipped lead tank.

However, due to many conditions, follow-on vehicles cannot exactly duplicate the lead tank path. A family of play functions is postulated and implemented to model the ability of these follow-on vehicles to stay on the intended path. The play function can be interpreted as the capability of the driver of a follow-on vehicle to stay on the intended path based on his skills, training, driving aid devices, and/or marking systems. A perfect path follow-on vehicle could be thought of as one whose play never strays off the intended trace (i.e., the center

Mine-Clearing Roller Effects on a Uniform Density Minefield

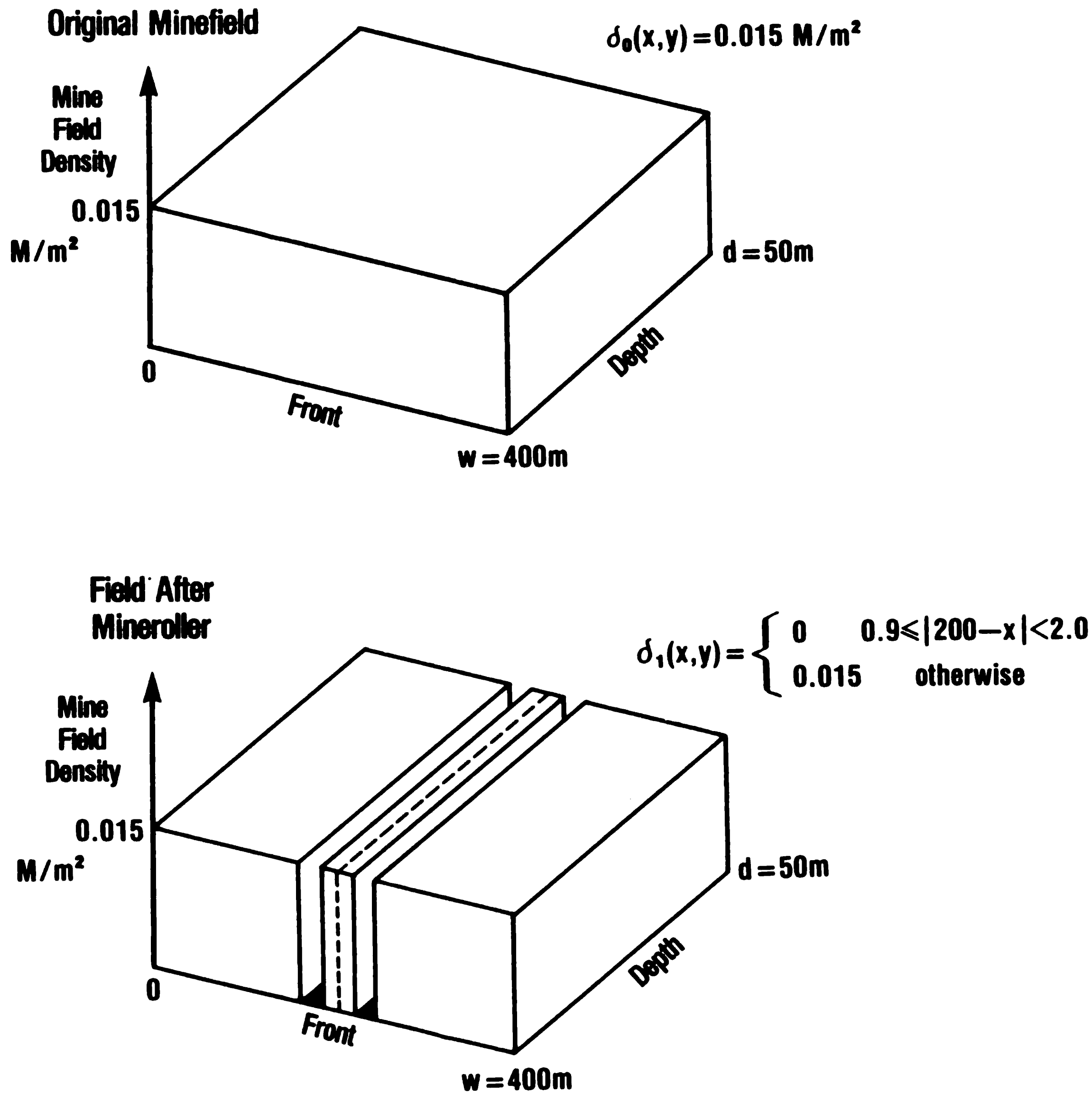


Figure 4. MINEFIELD DENSITY BEFORE AND AFTER NEUTRALIZATION

of the follow-on vehicle exactly retraces the center of the path of the lead vehicle). As conditions drop from less than perfect, follow-on vehicles have higher probabilities to stray away from the intended path. The sample play functions used in this application are at FIG 5.

The discontinuous nature of $\delta(x,y)$ and several of the $\Omega(z,m)$ functions lead to cumbersome arithmetic calculations. To ease this problem, a computer program[1] was written to readily analyze the expressions. Table I is a summary of sample applications of the IME methodology. Column four of this table shows the IME measures for the postulated systems. The effective path width listed in Column five is the sum of the range of play allowed and the width of the follow-on vehicle.

The IME measures change significantly over the set of play functions. For test run 1, the IME is 1.00, meaning 100% chance of crossing the minefield without mine encounter. This is a reasonable result, for this trial is with no play; the follow-on vehicle path exactly matches the lead tank path, and the tracks of the follow-on vehicles will always fall between the bounds of the safe zones cleared by the rollers. Allowing the vehicle to sway $\pm$ 0.5 m from the perfect lane (trial 2) drops the chance of encountering no mines 5 percentage points. Normally distributed play functions perform better than triangularly distributed ones of equal range (trial 3 vs 5 and 4 vs 6) because the normal distributions have a greater central tendency (they hug the line better) than the triangular distributions.

The IME measure indicates that a reduction in play of 1 meter betters the probability of no mine encounters by 0.13 (trial 6 compared with trial 5). The chance of encountering no mines jumps from 72% to 85% when the follow-on vehicle

[1]A listing of this program, written in SIMSCRIPT II.5, is in the Appendix.

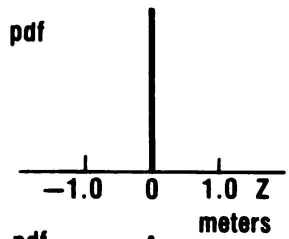
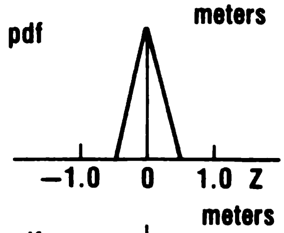
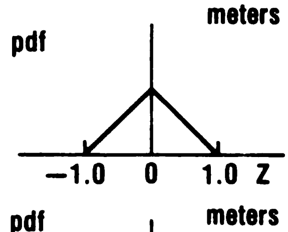
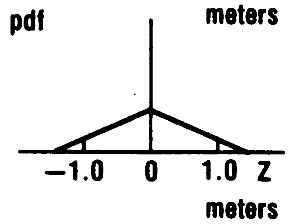
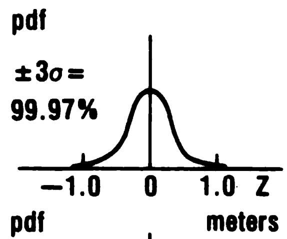
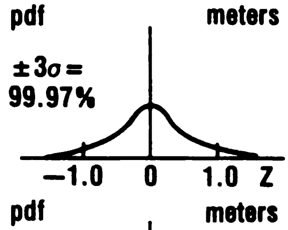
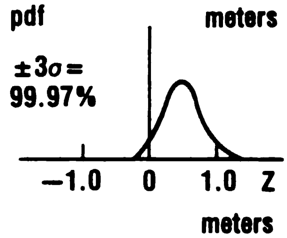
Test	Description	Range (m)	Ω (z,m)	Comment
1	No Play.	0.0		Perfect Tracing of Path.
2	Triangular Distribution	1.0		
3	Triangular Distribution	2.0		
4	Triangular Distribution	3.0		
5	Normal Distribution	2.0		$\mu=0.0$ $\sigma=1/3$
6	Normal Distribution	3.0		$\mu=0.0$ $\sigma=1/2$
7	Normal Distribution	2.0		$\mu=0.5$ $\sigma=1/3$ (Right Bias)

Figure 5. SAMPLE PLAY FUNCTIONS USED IN THE IME APPLICATIONS

Assumptions:

Minefield: 400 × 50m
3 Rows of AT Mines 4.0m Apart

Follow-on Vehicle Track Signature:

Track	Belly	Track	
.7m	2.2m	.7m	3.6 Total Meters

Inner Track Edge = ite = 1.1m
Outer Track Edge = ote = 1.8m

Neutralization: Straight Path Through Minefield Center with Roller

2 Cleared Zones 1.1m Each, 1.8 Meters Apart

Cleared	Un-cleared	Cleared
1.1m	1.8m	1.1m

$$IME = e^{-E(N)}$$

$$E(N) = \int_0^1 \int_{-\infty}^{\infty} \Omega(z,m) \left[\int_{olx}^{ilx} \delta(x,y) \frac{dY}{dm} dx + \int_{irx}^{orx} \delta(x,y) \frac{dY}{dm} dx \right] dz\, dm$$

$$\delta(x,y) = \begin{cases} 0 & 0.9 \leq |x-200| < 2.0 \\ 0.015 & \text{otherwise} \end{cases}$$

Results:

Test Run	$\Omega(z,m)$ Play Function About Path	E(N) = Expected Number of Mine Encounters	IME = Prob. of Encountering Zero Mines	Effective Path Width
1	None	0.000	1.00	3.6m
2	Triangular, Range = 1m	0.054	0.95	4.6m
3	Triangular, Range = 2m	0.256	0.77	5.6m
4	Triangular, Range = 3m	0.440	0.64	6.6m
5	Normal, Range = 2m	0.165	0.85	5.6m
6	Normal, Range = 3m	0.321	0.72	6.6m
7	Normal, Right Bias, Range = 2m	0.473	0.62	5.6m

Table I. SUMMARY OF SAMPLE APPLICATION OF THE IN-MINEFIELD EFFECTIVENESS MEASURE

restricts its deviation about the path a half meter on each side. This translates to a potential benefit due to increased survivability. This benefit can be realized through increased driver skill, improved training, and/or cleared-lane marking system, any approaches to result in less play of follow-on about a clear path.

The following section describes a study done to evaluate marking systems and how IME could have been employed to achieve meaningful results.

VI DISCUSSION

The Concept Evaluation Program of CLAMS (3 December 1981, TRADOC ACN 52725) compared the operational performance of chemiluminescent candles to highway safety flares in marking a breach through a minefield. Trial runs were scored as successful if a vehicle stayed within predetermined path widths 88% of the time during a breach. Measurements were taken as the vehicle passed each marker. Results from this test were non-conclusive. Only 7 of 203 attempts by M60 tanks to negotiate a 4 meter path were successful. The binary nature of the outcome of a trial (labeled success or failure) contributed greatly to the insensitivity of the results of the field experiments. Moreover, the outcome labels had little to do with mission success or failures of vehicles breaching hypothetical minefields as those simulated by the tests. Failure to maintain a four meter path in the test did not directly equate to failure to breach the simulated minefield, and the same is true for success. There was no reference to a real military worth.

The IME equation provides the means to combine minefield density, vehicle track signature and path into a quantitative assessment of military worth. The example results show the gains in terms of higher survivability by achieving narrower vehicle path tracings. Field experiments taken to measure vehicle path functions can be translated into quantitative measures attributable to the military worth of the marking systems through the IME.

V CONCLUSIONS

The IME measure is an easily calculated, yet sensitive indicator of the performance of mine and countermine systems. Elements of mine/countermine systems can be modelled by the various data and function inputs to the IME equation, as listed in Table II. Complex functions can be evaluated with a computer program specifically designed to solve IME equations. The IME measure is a useful index because it translates system performance characteristics of alternative mine/countermine systems into survivability figures. The IME process can quantify benefits of new developments, whether organizational, operational, or materiel in nature. IME measures can also be used as input to higher level, larger scope war games where previous data were randomly generated or estimated.

Examples of other uses of IME are:

a. Mixed mine type minefield effectivness. The different mine type densities and corresponding track or vehicle signatures can initially be separated out and later combined for an aggregate effectiveness measure.

b. Smart mine design parameters. The parameterized path function can be time normalized and probability of mine/vehicle encounter based on duration of exposure as well as area.

c. Countermine systems mix analyses. Single systems and combination can be studied.

d. Wide area countermine systems analyses. Hypothetical countermine system performances can be compared as to how well they neutralize threat minefields for follow on vehicles.

MINE/COUNTERMINE ELEMENT	IME EQUATION COMPONENTS THAT MODEL ELEMENT	
MINE DETECTION	$\delta\ (x,y)$	Minefield density
	$\Omega\ (z,m)$	Play
MINE NEUTRALIZATION	$\delta\ (x,y)$	Minefield density
	$P(m)$	Path
MARKING SYSTEMS	$P(m)$	Path
	$\Omega\ (z,m)$	Play
TERRAIN	$P(m)$	Path
	$\delta\ (z,m)$	Play
FUZING/TANK SIGNATURE INTERACTIONS	ite	inner track edge dimension
	ote	outer track edge dimension
MINE LAYING PATTERN	$\delta\ (x,y)$	Minefield density
TRAINING AND DOCTRINE	$\delta\ (x,y)$	Minefield density
	$P(m)$	Path
	$\Omega\ (z,m)$	Play

TABLE II. IME EQUATION COMPONENTS MODELLING MINE/COUNTERMINE ELEMENTS

ABBREVIATIONS

AT	Antitank
CLAMS	Cleared-Lane Marking System
FAE	Fuel-Air Explosive
HEL	Human Engineering Laboratory
IME	In-Minefield Effectiveness
OT	Operational Testing
USAES	U.S. Army Engineer School
m	meters
M/m^2	mines per meter squared

SYMBOLS

d	depth of minefield (meters)
w	width of minefield (meters)
m	mission parameter $o \leq m \leq 1$
x	position within minefield along the width (meters)
y	position within minefield along the depth (meters)
z	deviation from prescribed path (meters)
ite	inner track edge; the distance from the center of a vehicle to the inner track edge (meters)
ote	outer track edge; the distance from the center of a vehicle to the outer track edge (meters)
ilx	inner left track edge integrand
irx	inner right track edge integrand
olx	outer left track edge integrand
orx	outer right track edge integrand
P(m)	parameterized path function
X(m)	X component of path
Y(m)	Y component of path
dX/dm	first derivative of X(m)
dY/dm	first derivative of Y(m)
s(m)	slope of vehicle orientation at m
E(N)	the expected value of the number of mine encounters
Pr(n)	the probability of n occurances
$\delta(x,y)$	delta, the minefield density function
λ	lambda, the mean
μ	mu, the mean of a normal distribution
σ	sigma, the standard deviation of a normal distribution
$\Omega(z,m)$	omega, the play function
∞	infinity

APPENDIX A

COMPUTER PROGRAM LISTING FOR THE [illegible] EQUATION

CDC 6600 CACI SIMSCRIPT II.5 VERSION /4.5-00/ NOS-BE 1 12/22/81 14.16.05. PA

```
PREAMBLE
  NORMALLY, MODE IS REAL
  DEFINE DELTA, PLAY, SLOPE AS REAL FUNCTIONS WITH 2 ARGUMENTS
  DEFINE DY.DM, X.POS, Y.POS AS REAL FUNCTIONS WITH 1 ARGUMENT

  END ''PREAMBLE
```

CDC 6600 CACI SIMSCRIPT II.5 VERSION /4.5-00/ NOS-BE 1 12/22/81 14.06.25. P

```

 MAIN

   LET SUM=0.0


   READ M, Z, XX, LB, UB, ITE, OTE


   LET DM=1/M
   LET DZ=(UB-LB)/Z
   LET DXX=1/XX


   FOR M=DM/2 TO 1.0 BY DM, DO
      LET X=X.POS(M)
      LET Y=Y.POS(M)
      LET S=SLOPE(X, Y)
      LET DY=DY.DM(M)
      FOR Z=LB+DZ/2 TO UB BY DZ, DO
         LET WEIGHT=PLAY(Z, M)
         LET K=SQRT.F(1+S**2)
         LET OLX=X-(OTE-Z)/K
         LET ILX=X-(ITE-Z)/K
         FOR XX=OLX+DXX/2 TO ILX BY DXX, DO
            LET YY=Y+S*(XX-X)
            ADD DM*DZ*DXX*DY*WEIGHT*DELTA(XX, YY) TO SUM
            LOOP '' ON XX      (LEFT TRACK)

```

```
            LET IRX=X+(ITE+Z)/K
            LET ORX=X+(OTE+Z)/K
            FOR XX=IRX+DXX/2 TO ORX BY DXX, DO
               LET YY=Y+S*(XX-X)
               ADD DM*DZ*DXX*DY*WEIGHT*DELTA(XX, YY) TO SUM
               LOOP ## ON XX      (RIGHT TRACK)

            LOOP ##ON Z          (PLAY)

         LOOP ##ON M             (MISSION)

      PRINT 6 LINES WITH LB, UB, ITE, OTE, SUM THUS
LOWER, UPPER BOUNDS ON VEHICLE PLAY = ***.**, ***.** METERS
       CENTER TO INSIDE TRACK EDGE = **.** METERS
       CENTER TO OUTSIDE TRACK EDGE= **.** METERS

    EXPECTED NUMBER OF MINES ENCOUNTERED = ***.***

      LET PROB=EXP.F(-SUM)
      PRINT 1 LINE WITH PROB THUS
    PROB. OF ENCOUNTERING NO MINES   IS  *.***
      FOR I= 1.0 TO 20.0 BY 1.0, DO
         LET PROB=PROB*SUM/I
         PRINT 1 LINE WITH I, PROB THUS

    PROB. OF ENCOUNTERING ** MINE(S) IS  *.***
         LOOP ##ON I

      END ##MAIN
```

CDC 6600   CACI SIMSCRIPT II.5 VERSION /4.5-00/ NOS-BE 1   12/22/81 14.16.05.   PA

```
     FUNCTION DELTA(A, B)

      IF 0.9 LE ABS.F(A-200.0) LE 2.0
         RETURN WITH 0.0
      ELSE
         RETURN WITH 0.015
      END ##DELTA FUNCTION     (OF MINE DENSITY)
```

```
CDC  6600   CACI SIMSCRIPT II.5 VERSION /4.5-00/  NOS-BE  1            12/22/81  14.16.05.     PAG

       FUNCTION X.POS(A)

         RETURN WITH 200.0
         END ''X.POS FUNCTION
```

```
CDC  6600   CACI SIMSCRIPT II.5 VERSION /4.5-00/  NOS-BE  1            12/22/81  14.16.05.     PAG

       FUNCTION Y.POS(A)

         RETURN WITH 50.0*A
         END ''Y.POS FUNCTION
```



```
CDC  6600   CACI SIMSCRIPT II.5 VERSION /4.5-00/  NOS-BE  1            12/22/81  14.16.05.     PA(

       FUNCTION DY.DM(A)

         RETURN WITH 50.0
         END ''DY.DM FUNCTION
```

```
CDC  6600   CACI SIMSCRIPT II.5 VERSION /4.5-00/  NOS-BE  1            12/22/81  14.16.05.     PA

       FUNCTION SLOPE(A, B)

         RETURN WITH 0.0
         END ''SLOPE FUNCTION
```

```

FUNCTION PLAY(A, B)

   IF ABS.F(A) LE 0.025
      RETURN WITH 1.0
   ELSE
      RETURN WITH 0.0
   END ##PLAY FUNCTION
```

INPUT FILE

```
40   40   40
-1.  1.
1.1
1.8
```

```
LOWER, UPPER BOUNDS ON VEHICLE PLAY =  -1.00,   1.00 METERS
      CENTER TO INSIDE TRACK EDGE =  1.10 METERS
      CENTER TO OUTSIDE TRACK EDGE=  1.80 METERS

   EXPECTED NUMBER OF MINES ENCOUNTERED =   0.

   PROB. OF ENCOUNTERING NO MINES    IS  1.000
   PROB. OF ENCOUNTERING  1 MINE(S) IS  0.
   PROB. OF ENCOUNTERING  2 MINE(S) IS  0.
   PROB. OF ENCOUNTERING  3 MINE(S) IS  0.
   PROB. OF ENCOUNTERING  4 MINE(S) IS  0.
   PROB. OF ENCOUNTERING  5 MINE(S) IS  0.
   PROB. OF ENCOUNTERING  6 MINE(S) IS  0.
   PROB. OF ENCOUNTERING  7 MINE(S) IS  0.
   PROB. OF ENCOUNTERING  8 MINE(S) IS  0.
   PROB. OF ENCOUNTERING  9 MINE(S) IS  0.
   PROB. OF ENCOUNTERING 10 MINE(S) IS  0.
   PROB. OF ENCOUNTERING 11 MINE(S) IS  0.
   PROB. OF ENCOUNTERING 12 MINE(S) IS  0.
   PROB. OF ENCOUNTERING 13 MINE(S) IS  0.
   PROB. OF ENCOUNTERING 14 MINE(S) IS  0.
   PROB. OF ENCOUNTERING 15 MINE(S) IS  0.
   PROB. OF ENCOUNTERING 16 MINE(S) IS  0.
   PROB. OF ENCOUNTERING 17 MINE(S) IS  0.
   PROB. OF ENCOUNTERING 18 MINE(S) IS  0.
   PROB. OF ENCOUNTERING 19 MINE(S) IS  0.
   PROB. OF ENCOUNTERING 20 MINE(S) IS  0.
```

```
FUNCTION PLAY(A, B)

  IF A LE 0.0
     RETURN WITH 2 + 4*A
  ELSE
     RETURN WITH 2 - 4*A
  END ''PLAY FUNCTION
```

INPUT FILE

```
40  40  40
-.5 .5
1.1
1.8
```

```
LOWER, UPPER BOUNDS ON VEHICLE PLAY =   -.50,    .50 METERS
      CENTER TO INSIDE TRACK EDGE =  1.10 METERS
      CENTER TO OUTSIDE TRACK EDGE=  1.80 METERS

  EXPECTED NUMBER OF MINES ENCOUNTERED =    .054

  PROB. OF ENCOUNTERING NO MINES     IS   .947
  PROB. OF ENCOUNTERING  1 MINE(S)   IS   .051
  PROB. OF ENCOUNTERING  2 MINE(S)   IS   .001
  PROB. OF ENCOUNTERING  3 MINE(S)   IS   .000
  PROB. OF ENCOUNTERING  4 MINE(S)   IS   .000
  PROB. OF ENCOUNTERING  5 MINE(S)   IS   .000
  PROB. OF ENCOUNTERING  6 MINE(S)   IS   .000
  PROB. OF ENCOUNTERING  7 MINE(S)   IS   .000
  PROB. OF ENCOUNTERING  8 MINE(S)   IS   .000
  PROB. OF ENCOUNTERING  9 MINE(S)   IS   .000
  PROB. OF ENCOUNTERING 10 MINE(S)   IS   .000
  PROB. OF ENCOUNTERING 11 MINE(S)   IS   .000
  PROB. OF ENCOUNTERING 12 MINE(S)   IS   .000
  PROB. OF ENCOUNTERING 13 MINE(S)   IS   .000
  PROB. OF ENCOUNTERING 14 MINE(S)   IS   .000
  PROB. OF ENCOUNTERING 15 MINE(S)   IS   .000
  PROB. OF ENCOUNTERING 16 MINE(S)   IS   .000
  PROB. OF ENCOUNTERING 17 MINE(S)   IS   .000
  PROB. OF ENCOUNTERING 18 MINE(S)   IS   .000
  PROB. OF ENCOUNTERING 19 MINE(S)   IS   .000
  PROB. OF ENCOUNTERING 20 MINE(S)   IS   .000
```

```
FUNCTION PLAY(A, B)

  IF A LE 0.0
     RETURN WITH 1.0+A
  ELSE
     RETURN WITH 1.0-A
  END ''PLAY FUNCTION
```

INPUT FILE

```
40  40  40
-1.  1.
 1.1
 1.8
```

```
LOWER, UPPER BOUNDS ON VEHICLE PLAY =  -1.00,   1.00 METERS
      CENTER TO INSIDE TRACK EDGE =  1.10 METERS
      CENTER TO OUTSIDE TRACK EDGE=  1.80 METERS

    EXPECTED NUMBER OF MINES ENCOUNTERED =    .256

    PROB. OF ENCOUNTERING NO MINES    IS   .774
    PROB. OF ENCOUNTERING  1 MINE(S) IS   .198
    PROB. OF ENCOUNTERING  2 MINE(S) IS   .025
    PROB. OF ENCOUNTERING  3 MINE(S) IS   .002
    PROB. OF ENCOUNTERING  4 MINE(S) IS   .000
    PROB. OF ENCOUNTERING  5 MINE(S) IS   .000
    PROB. OF ENCOUNTERING  6 MINE(S) IS   .000
    PROB. OF ENCOUNTERING  7 MINE(S) IS   .000
    PROB. OF ENCOUNTERING  8 MINE(S) IS   .000
    PROB. OF ENCOUNTERING  9 MINE(S) IS   .000
    PROB. OF ENCOUNTERING 10 MINE(S) IS   .000
    PROB. OF ENCOUNTERING 11 MINE(S) IS   .000
    PROB. OF ENCOUNTERING 12 MINE(S) IS   .000
    PROB. OF ENCOUNTERING 13 MINE(S) IS   .000
    PROB. OF ENCOUNTERING 14 MINE(S) IS   .000
    PROB. OF ENCOUNTERING 15 MINE(S) IS   .000
    PROB. OF ENCOUNTERING 16 MINE(S) IS   .000
    PROB. OF ENCOUNTERING 17 MINE(S) IS   .000
    PROB. OF ENCOUNTERING 18 MINE(S) IS   .000
    PROB. OF ENCOUNTERING 19 MINE(S) IS   .000
    PROB. OF ENCOUNTERING 20 MINE(S) IS   .000
```

A10

```
FUNCTION PLAY(A, B)

   IF A LE 0.0
      RETURN WITH 2/3 + 4*A/9
   ELSE
      RETURN WITH 2/3 - 4*A/9
   END ##PLAY FUNCTION
```

INPUT FILE

```
 40   40   40
-1.5  1.5
 1.1
 1.8
```

```
LOWER, UPPER BOUNDS ON VEHICLE PLAY =  -1.50,   1.50 METERS
       CENTER TO INSIDE TRACK EDGE =  1.10 METERS
       CENTER TO OUTSIDE TRACK EDGE=  1.80 METERS

    EXPECTED NUMBER CF MINES ENCOUNTERED =    .440

    PROB. OF ENCOUNTERING NO MINES    IS    .644
    PROB. OF ENCOUNTERING  1 MINE(S)  IS    .283
    PROB. OF ENCCUNTERING  2 MINE(S)  IS    .062
    PROB. OF ENCCUNTERING  3 MINE(S)  IS    .009
    PROB. OF ENCCUNTERING  4 MINE(S)  IS    .001
    PROB. OF ENCOUNTERING  5 MINE(S)  IS    .000
    PROB. OF ENCOUNTERING  6 MINE(S)  IS    .000
    PROB. OF ENCOUNTERING  7 MINE(S)  IS    .000
    PROB. OF ENCCUNTERING  8 MINE(S)  IS    .000
    PROB. OF ENCOUNTERING  9 MINE(S)  IS    .000
    PROB. OF ENCOUNTERING 10 MINE(S)  IS    .000
    PROB. OF ENCCUNTERING 11 MINE(S)  IS    .000
    PROB. OF ENCOUNTERING 12 MINE(S)  IS    .000
    PROB. OF ENCOUNTERING 13 MINE(S)  IS    .000
    PROB. OF ENCOUNTERING 14 MINE(S)  IS    .000
    PROB. OF ENCOUNTERING 15 MINE(S)  IS    .000
    PROB. OF ENCOUNTERING 16 MINE(S)  IS    .000
    PROB. OF ENCOUNTERING 17 MINE(S)  IS    .000
    PROB. OF ENCCUNTERING 18 MINE(S)  IS    .000
    PROB. OF ENCOUNTERING 19 MINE(S)  IS    .000
    PROB. OF ENCCUNTERING 20 MINE(S)  IS    .000
```

```
CDC  6600   CACI SIMSCRIPT II.5 VERSION /4.5-00/  NOS-BE  1                  12/24/81  08.47.37.     PA

      FUNCTION PLAY(A, B)

        RETURN WITH 2/SQRT.F(PI.C*2)*EXP.F(-1/2*A**2*4)
        END ''PLAY FUNCTION
```

INPUT FILE

```
40  40  40
-1.  1
 1.1
 1.8
```

```
LOWER, UPPER BOUNDS ON VEHICLE PLAY =  -1.50,   1.50 METERS
      CENTER TO INSIDE TRACK EDGE =  1.10 METERS
      CENTER TO OUTSIDE TRACK EDGE=  1.80 METERS

   EXPECTED NUMBER OF MINES ENCOUNTERED =    .321

   PROB. OF ENCOUNTERING NO MINES    IS   .725
   PROB. OF ENCOUNTERING  1 MINE(S)  IS   .233
   PROB. OF ENCOUNTERING  2 MINE(S)  IS   .037
   PROB. OF ENCOUNTERING  3 MINE(S)  IS   .004
   PROB. OF ENCOUNTERING  4 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING  5 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING  6 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING  7 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING  8 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING  9 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 10 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 11 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 12 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 13 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 14 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 15 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 16 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 17 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 18 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 19 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 20 MINE(S)  IS   .000
```

```
FUNCTION PLAY(A, B)

  RETURN WITH 3/SQRT.F(PI.C*2)*EXP.F(-1/2*(A)**2*9)
  END ''PLAY FUNCTION
```

INPUT FILE

```
 40   40   40
-1.5  1.5
 1.1
 1.8
```

```
LOWER, UPPER BOUNDS ON VEHICLE PLAY =  -1.00,   1.00 METERS
      CENTER TO INSIDE TRACK EDGE =  1.10 METERS
      CENTER TO OUTSIDE TRACK EDGE=  1.80 METERS

   EXPECTED NUMBER OF MINES ENCOUNTERED =    .165

   PROB. OF ENCOUNTERING NO MINES    IS    .848
   PROB. OF ENCOUNTERING  1 MINE(S) IS    .140
   PROB. OF ENCOUNTERING  2 MINE(S) IS    .012
   PROB. OF ENCOUNTERING  3 MINE(S) IS    .001
   PROB. OF ENCOUNTERING  4 MINE(S) IS    .000
   PROB. OF ENCOUNTERING  5 MINE(S) IS    .000
   PROB. OF ENCOUNTERING  6 MINE(S) IS    .000
   PROB. OF ENCOUNTERING  7 MINE(S) IS    .000
   PROB. OF ENCOUNTERING  8 MINE(S) IS    .000
   PROB. OF ENCOUNTERING  9 MINE(S) IS    .000
   PROB. OF ENCOUNTERING 10 MINE(S) IS    .000
   PROB. OF ENCOUNTERING 11 MINE(S) IS    .000
   PROB. OF ENCOUNTERING 12 MINE(S) IS    .000
   PROB. OF ENCOUNTERING 13 MINE(S) IS    .000
   PROB. OF ENCOUNTERING 14 MINE(S) IS    .000
   PROB. OF ENCOUNTERING 15 MINE(S) IS    .000
   PROB. OF ENCOUNTERING 16 MINE(S) IS    .000
   PROB. OF ENCOUNTERING 17 MINE(S) IS    .000
   PROB. OF ENCOUNTERING 18 MINE(S) IS    .000
   PROB. OF ENCOUNTERING 19 MINE(S) IS    .000
   PROB. OF ENCOUNTERING 20 MINE(S) IS    .000
```

A16

```

FUNCTION PLAY(A, B)

  RETURN WITH 3/SQRT.F(PI.C*2)*EXP.F(-1/2*(A-0.5)**2*9)
  END ''PLAY FUNCTION
```

INPUT FILE

```
40  40  40
-.5 1.5
1.1
1.8
```

```
LOWER, UPPER BOUNDS ON VEHICLE PLAY =   -.50,   1.50 METERS
       CENTER TO INSIDE TRACK EDGE =  1.10 METERS
       CENTER TO OUTSIDE TRACK EDGE=  1.80 METERS

   EXPECTED NUMBER OF MINES ENCOUNTERED =    .473

   PROB. OF ENCOUNTERING NO MINES    IS   .623
   PROB. OF ENCOUNTERING  1 MINE(S)  IS   .295
   PROB. OF ENCOUNTERING  2 MINE(S)  IS   .070
   PROB. OF ENCOUNTERING  3 MINE(S)  IS   .011
   PROB. OF ENCOUNTERING  4 MINE(S)  IS   .001
   PROB. OF ENCOUNTERING  5 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING  6 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING  7 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING  8 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING  9 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 10 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 11 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 12 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 13 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 14 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 15 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 16 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 17 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 18 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 19 MINE(S)  IS   .000
   PROB. OF ENCOUNTERING 20 MINE(S)  IS   .000
```

ROBUST RANGE MEASUREMENT PREPROCESSING

William S. Agee and Robert H. Turner
Mathematical Services Branch
Data Sciences Division
National Range Operations Directorate
US Army White Sands Missile Range
White Sands Missile Range, New Mexico

ABSTRACT

The RMS/MTTS instrumentation system located at MacGregor Range is a range measuring, multiple target tracking system. In order to obtain a vehicle trajectory from this system, the range measurement from several receivers are processed by least squares. Since the measured vehicle trajectories are often low altitude, the resulting nonlinear least squares equations are ill-conditioned. In addition, this measurement system is plagued by outliers, sometimes by dense burst of outliers. The combination of ill-conditioning and outliers is lethal and attempts to robustify the nonlinear least squares processing have failed. An alternative is to preprocess each of the range measurement sequences, eliminating the outliers and replacing them if necessary. Each sequence of range measurements is preprocessed by robustly fitting a cubic spline using iteratively reweighted least squares. Due to the nature of spline fitting and the possible dense bursts of outliers, the choice of a good set of initial weights for use in the iteratively reweighted least squares is important to the efficiency of the method. These initial weights are determined using robust, local fitting techniques. Several robust techniques have been tested for this local fitting application. The robust spline preprocessing is illustrated with some especially troublesome data sequences and the relative performance of several robust methods for choosing the initial weights is compared.

INTRODUCTION

The RMS/MTTS instrumentation system located at MacGregor Range is a range measuring, multiple target tracking system. In order to obtain a vehicle trajectory from this system, the range measurements from several receivers are processed by least squares. Because the measured vehicle trajectories of interest are often low altitude and because of the geometry of the receiving stations, the resulting nonlinear least squares equations are often ill-conditioned. In addition, this measurement system is subject to outliers, sometimes dense bursts of outliers. This combination of ill-conditioning and outliers is lethal and our attempts to robustify the nonlinear least squares estimation process have failed. An alternative is to preprocess each of the range measurement sequences, identifying the outliers, and replacing them if necessary. The ill-conditioned least squares problem can then be treated without being troubled by outliers.

Suppose we preprocess the measurement sequence, $R(t_i)$, $i = 1,N$. For typical aircraft trajectories the measurement rate is 10/sec with a time of interest of 40 - 120 sec so that N is often in the range 400 - 1200. The purpose of the preprocessing may be to detect outliers, to precompute measurement variances for future least squares processing, or to synchronize several different discrete measurement sequences. The preprocessing of the range measurement sequence, $R(t_i)$, is done by fitting a cubic spline to the discrete measurements using iteratively reweighted least squares (IRWLS). Specifically, at the k^{th} iteration we minimize,

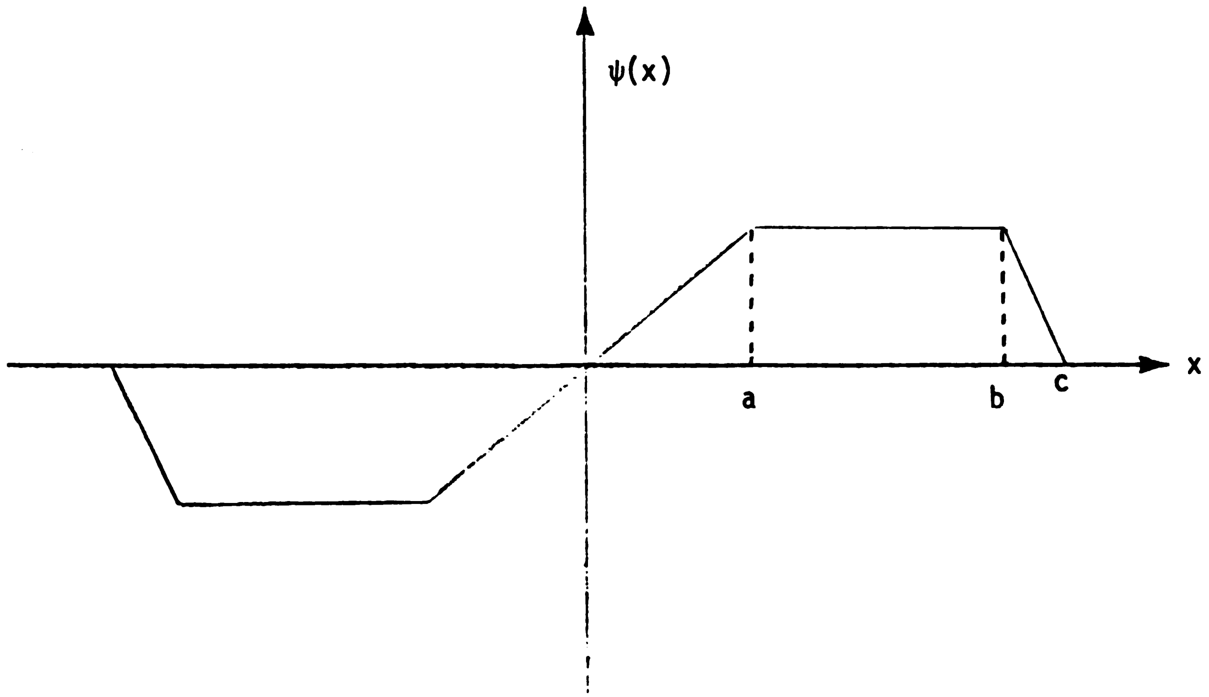
$$\sum_{j=1}^{N} W_j^{(k)} \left(R(t_j) - \sum_i b_i^{(k)} B_i(t_j)\right)^2, \qquad (1)$$

where $B_i(\cdot)$ are the cubic B-splines and $b_i^{(k)}$ are the spline coefficients to be estimated. The weights, $W_j^{(k)}$ are computed from the Hampel ψ-function using the spline fit from the $(k-1)^{\underline{st}}$ iteration.

$$W_j^{(k)} = \frac{\psi\left(\dfrac{R(t_j) - \sum_i b_i^{(k-1)} B_i(t_j)}{S_j^{k-1}}\right)}{\dfrac{R(t_j) - \sum_i b_i^{(k-1)} B_i(t_j)}{S_j^{k-1}}} \tag{2}$$

where

$$\psi(x) = \begin{cases} x & |x| \le a \\ a \cdot \text{sgn}(x) & a < |x| \le b \\ a\left(\dfrac{x-c \ \text{sgn}(x)}{b-c}\right) & b < |x| \le c \\ 0 & |x| > c \end{cases} \tag{3}$$



$S_j^{(k-1)}$ estimates the dispersion in the residual, $R(t_j) - \sum_i b_i^{(k-1)} B_i(t_j)$. The value of $S_j^{(k-1)}$ can be computed either locally or globally from the residuals at the $(k-1)^{st}$ iteration. The dispersion $S_j^{(k-1)}$ is a MAD estimate obtained from

$$S_j^{(k-1)} = \underset{m \varepsilon T_j}{\text{median}} \left| R(t_m) - \sum_i b_i^{(k-1)} B_i(t_m) \right| / .6745 \qquad (4)$$

If the set T_j is in some sense the set of points close to t_j, the estimate $S_j^{(k-1)}$ is local. If the set T_j is the set, $T_j = \{t_m | m = 1,N\}$ the estimate is global. For the present application only the global estimate $S_j^{(k-1)} = S^{(k-1)}$ will be used. If a very long data sequence, say about one hour, a local estimate would probably be preferable to the global estimate.

CHOICE OF KNOTS

Let $\{T_i, i = 1,M\}$ be a set of knot times. These knot times are used to define the cubic B-splines, $B_i(t_j)$. Of most importance in the choice of the knot times is their spacing, which determines the ability of the cubic spline to fit the data. However, for each additional knot time there is one additional spline coefficient to be estimated, thus increasing the computational load. Thus, we want to have as few knots as possible and the rules for their choice simple and yet be able to adequately represent the data. With this simple philosophy for selecting knots we will try to assign a fixed number of data points, NPTO, to each knot interval. The first four knots are placed at the first data. The last knot interval may have more than NPTO points but fewer than 2 · NPTO data points. If there is a large time break in the data, a knot is placed at the beginning and end of the time break. The interval between these two knots has zero data points and the interval immediately preceding the time break may have more than NPTO points but fewer than 2 · NPTO points. If immediately after a time break there is a second time break before NPTO points have been read, the few (less than NPTO) points read between the two time breaks are discarded. If a time break occurs while reading points for the first knot interval, the few (less than NPTO) points are discarded and the first four knots repositioned at the first data time after the time break. If a time break occurs during the last interval, the portion of the last interval contiguous to the previous interval is kept and the remainder of the points in the last interval are discarded. If there are at least NPTO points kept, these points form the last interval. If there are less than NPTO points kept, these points are appended to the previous interval so that the number of knot intervals is reduced by one. The time difference between successive data points which is used to define a time break is named FITBRK. FITBRK is dependent on the sample rate. The time difference between successive data points used to define a time break in the first and last knot intervals is FITBRK/5. This smaller value is used in the first and last interval because it is critical to obtain a good fit in these intervals. The flow chart on the following pages more clearly defines the logic for selecting the knot times. The following define the variables in the flow chart:

NOTS = number of interior knots

KR = number of knot intervals

TT(·) = array of knot times

NCØUN = total data point count

T(·) = array of data times

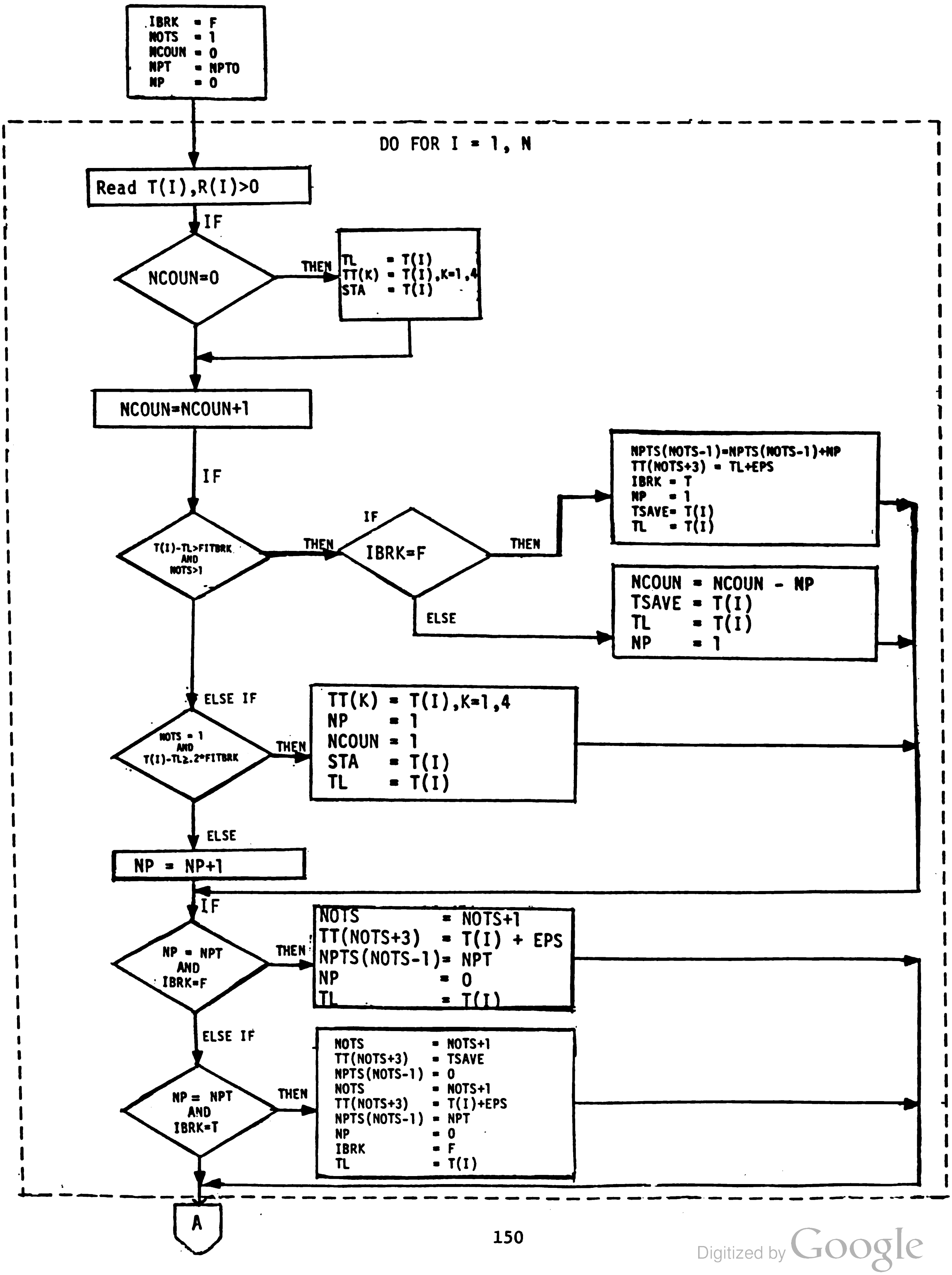
R(·) = array of range measurements

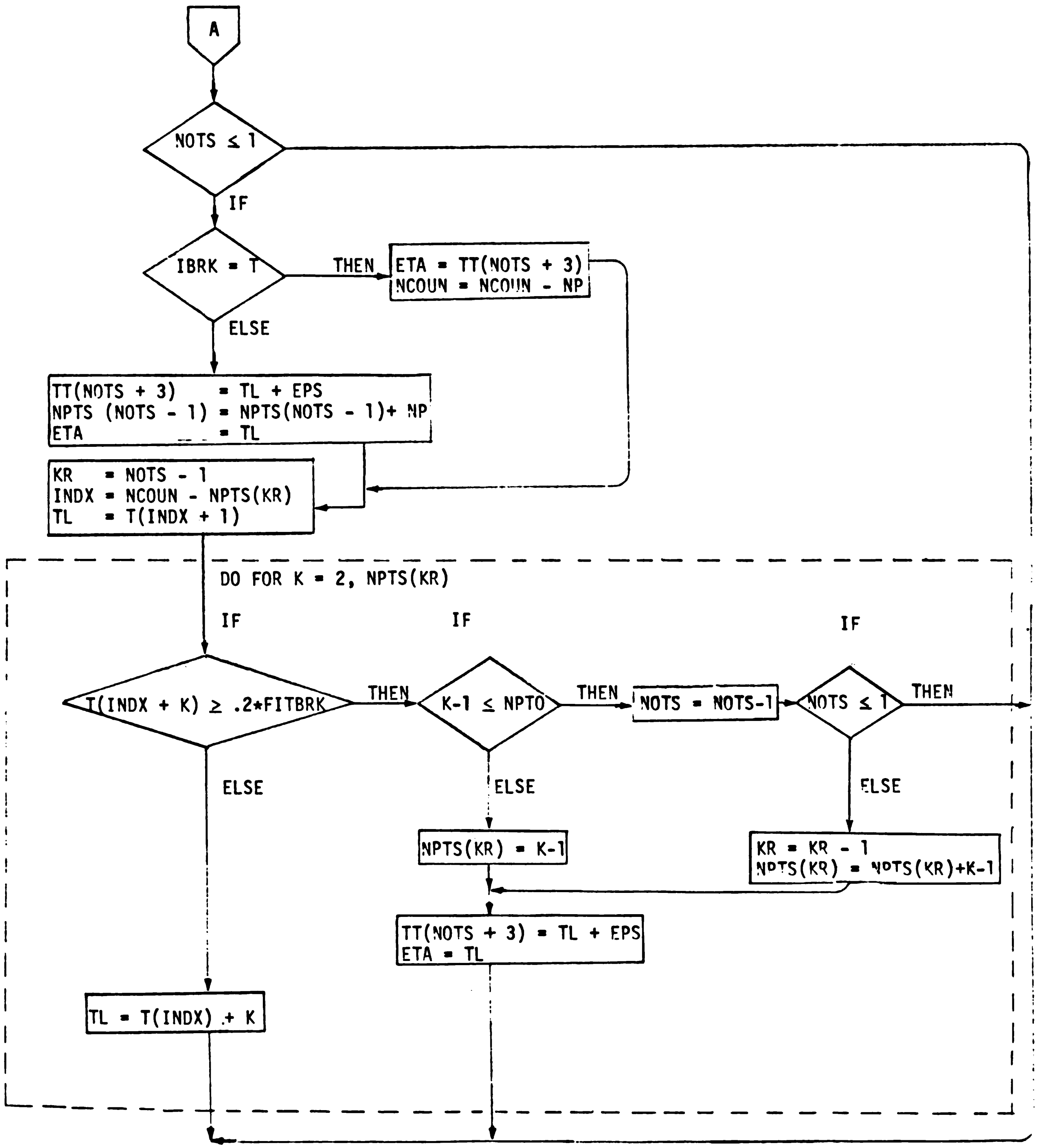
NPTS(·) = array of point counts for knot intervals

IBRK = logical denoting the occurrence of a time break

STA = data start time

ETA = data end time


IBRK = F
NOTS = 1
NCOUN = 0
NPT = NPTO
NP = 0
DO FOR I = 1, N
Read T(I),R(I)>0
IF
NCOUN=0
THEN
TL = T(I)
TT(K) = T(I),K=1,4
STA = T(I)
NCOUN=NCOUN+1
IF
T(I)-TL>FITBRK
AND
NOTS>1
THEN
IF
IBRK=F
THEN
NPTS(NOTS-1)=NPTS(NOTS-1)+NP
TT(NOTS+3) = TL+EPS
IBRK = T
NP = 1
TSAVE= T(I)
TL = T(I)
ELSE
NCOUN = NCOUN - NP
TSAVE = T(I)
TL = T(I)
NP = 1
ELSE IF
NOTS = 1
AND
T(I)-TL≥.2*FITBRK
THEN
TT(K) = T(I),K=1,4
NP = 1
NCOUN = 1
STA = T(I)
TL = T(I)
ELSE
NP = NP+1
IF
NP = NPT
AND
IBRK=F
THEN
NOTS = NOTS+1
TT(NOTS+3) = T(I) + EPS
NPTS(NOTS-1)= NPT
NP = 0
TL = T(I)
ELSE IF
NP = NPT
AND
IBRK=T
THEN
NOTS = NOTS+1
TT(NOTS+3) = TSAVE
NPTS(NOTS-1) = 0
NOTS = NOTS+1
TT(NOTS+3) = T(I)+EPS
NPTS(NOTS-1) = NPT
NP = 0
IBRK = F
TL = T(I)
A


A
NOTS ≤ 1
IF
IBRK = T
THEN
ETA = TT(NOTS + 3)
NCOUN = NCOUN - NP
ELSE
TT(NOTS + 3) = TL + EPS
NPTS (NOTS - 1) = NPTS(NOTS - 1)+ NP
ETA = TL
KR = NOTS - 1
INDX = NCOUN - NPTS(KR)
TL = T(INDX + 1)
DO FOR K = 2, NPTS(KR)
IF
IF
IF
T(INDX + K) ≥ .2*FITBRK
THEN
K-1 ≤ NPTO
THEN
NOTS = NOTS-1
NOTS ≤ 1
THEN
ELSE
ELSE
ELSE
NPTS(KR) = K-1
KR = KR - 1
NPTS(KR) = NPTS(KR)+K-1
TT(NOTS + 3) = TL + EPS
ETA = TL
TL = T(INDX) + K

THE LEAST SQUARES NORMAL EQUATIONS

At the $k^{\underline{th}}$ iteration of the fitting procedure the weighted sum of squares

$$\sum_{j=1}^{N} W_j^{(k)} \left(R(t_j) - \sum_i b_i^{(k)} B_i(t_j)\right)^2 \tag{5}$$

is minimized. The least squares normal equations are obtained by differentiating (5) with respect to $b_i^{(k)}$. The least squares normal equations are

$$\sum_{j=1}^{N} W_j^{(k)} B(t_j) B^T(t_j) \hat{b}^{(k)} = \sum_{j=1}^{N} W_j^{(k)} B(t_j) R(t_j) \tag{6}$$

where $B^T(t_j)$ is the vector of cubic B-splines

$$B^T(t_j) = [B_1(t_j) \; B_2(t_j) \; - - - \; B_n(t_j)] \tag{7}$$

Due to the nature of the B-splines the positive definite matrix on the left of (6) is banded with three bands above and below the main diagonal. To conserve storage the four distinct diagonals of this matrix are stored as columns of a vertical matrix. The dimension of the vector $\hat{b}^{(k)}$ is NOTS + 2 where NOTS is the number of interior knots. The banded least squared normal equations are solved by a banded Cholesky decomposition algorithm. The sums of both sides of (6) are performed sequentially so that all of the ranges and weights are not needed in core simultaneously. The IRWLS can be continued for a fixed number of iterations or until the fit has converged.

INITIAL WEIGHTS

In many situations the IRWLS procedure works successfully when all of the initial weights are set to one, i.e., the iteration is started with an ordinary unweighted least squares solution. We have found that the use of the unweighted least squares start will usually result in convergence of the IRWLS cubic spline to a good fit with outliers correctly identified, but that many fewer iterations are required if a robust choice of initial weights is used. When outliers are present in either the first or last intervals, the choice of initial weights in these intervals is most important.

The initial weights for the robust cubic spline fit are chosen on a localized basis. Let $R(t_i)$, i = 1, NPTS(K) be the range measurements in the $K^{\underline{th}}$ knot interval. To determine the weights $W_i^{(o)}$, i = 1, NPTS(K) in the $K^{\underline{th}}$ interval a linear curve is robustly fitted to the measurements in the interval. Several methods for robustly fitting the linear curve have been tried, including the nested median method of Siegel [1], the method of Theil [2], a modified Theil method, and an M-estimate using a Hampel ψ-function. Most methods performed about equally well on the data sequences tested. The results of some of these tests are given in Appendix A. Because of its simplicity, the modified method of Theil was selected for routine application. This method is described in the following paragraph.

Let $\bar{R}$ be the median of the observations in the $K^{\underline{th}}$ knot interval.

$$\bar{R} = \underset{i = 1, NPTS(K)}{\text{median}} \{R(t_i)\} \tag{8}$$

Let $\bar{t}$ be the time corresponding to $\bar{R}$. The median $\bar{R}$ can be represented as the average of two observations,

$$\bar{R} = (R(t_{m_1}) + R(t_{m_2}))/2 \tag{9}$$

where $m_1 = m_2$ if NPTS(K) is odd. Define the set of slopes $\{S_i\}$

$$S_i = \frac{R(t_i) - \bar{R}}{t_i - \bar{t}} \qquad \begin{array}{l} i = 1, NPTS(K) \\ i \neq m_1, m_2 \end{array} \tag{10}$$

Let $\bar{s}$ be the median of the slopes,

$$\bar{s} = \text{median } \{S_i\} \quad \begin{array}{l} i = 1, NPTS(K) \\ i \neq m_1, m_2 \end{array} \tag{11}$$

Let $\bar{r}_i$ be the residual,

$$\bar{r}_i = R(t_i) - \bar{s}(t_i - \bar{t}), \ i = 1, NPTS(K) \tag{12}$$

Let $\bar{r}$ be the median of these residuals,

$$\bar{r} = \text{median } \{\bar{r}_i\} \quad i = 1, NPTS(K) \tag{13}$$

Now compute the residuals,

$$r_i = \bar{r}_i - \bar{r}, \ i = 1, NPTS(K) \tag{14}$$

The initial weights, $W_i^{(o)}$, are computed from these residuals using a Hampel ψ-function.

$$W_i^{(o)} = \frac{\psi\left(\frac{r_i}{s_k}\right)}{\left(\frac{r_i}{s_k}\right)}, \ i = 1, NPTS(K) \tag{15}$$

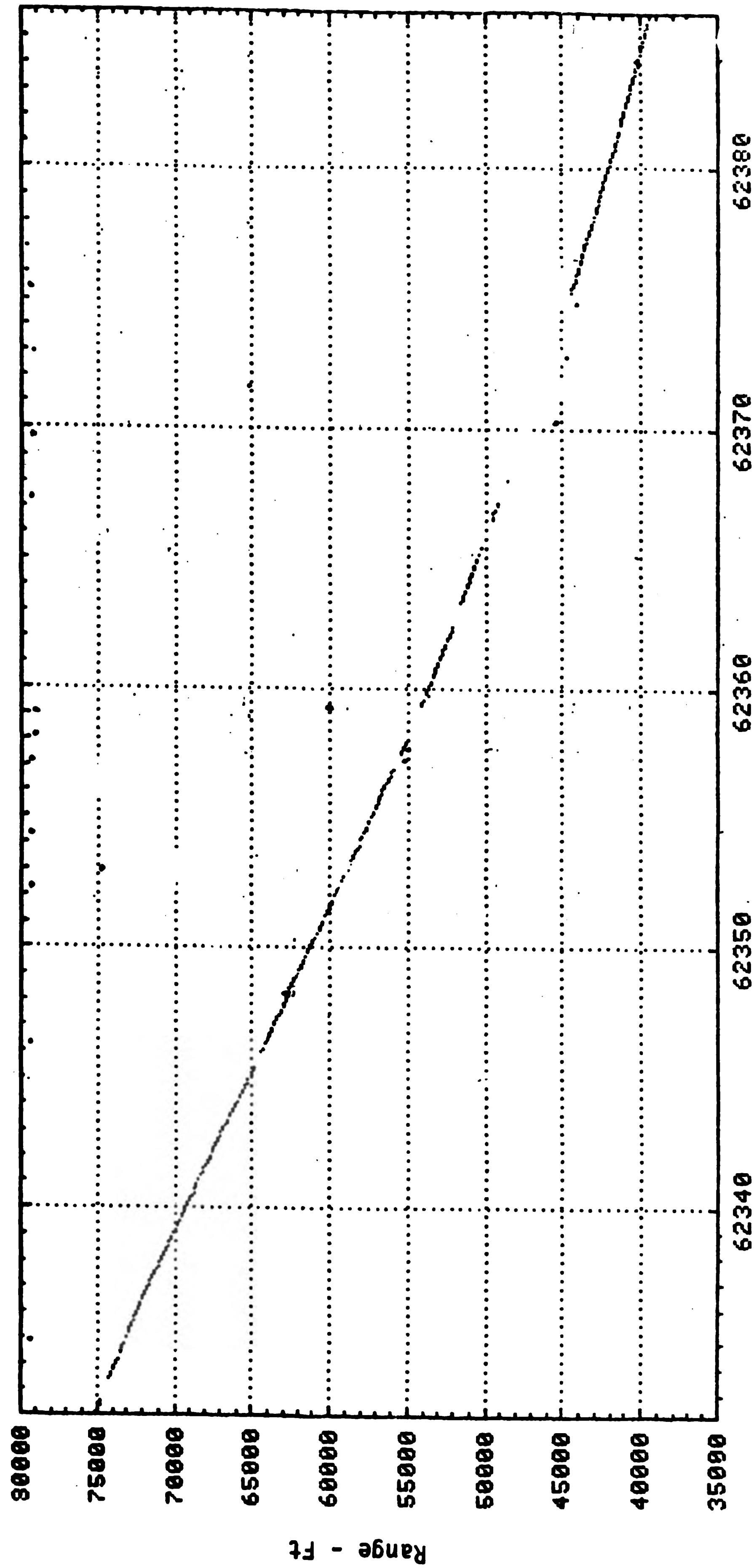
Where s_k is the robust dispersion parameter,

$$s_k = \underset{i}{\text{median}} \{|r_i|\}/.6745 \tag{16}$$

Since the main concern in setting the initial weights is to protect the cubic spline fit from the gross outliers, the break points of the Hampel ψ-function are set at a = 2, b = 3, c = 4.

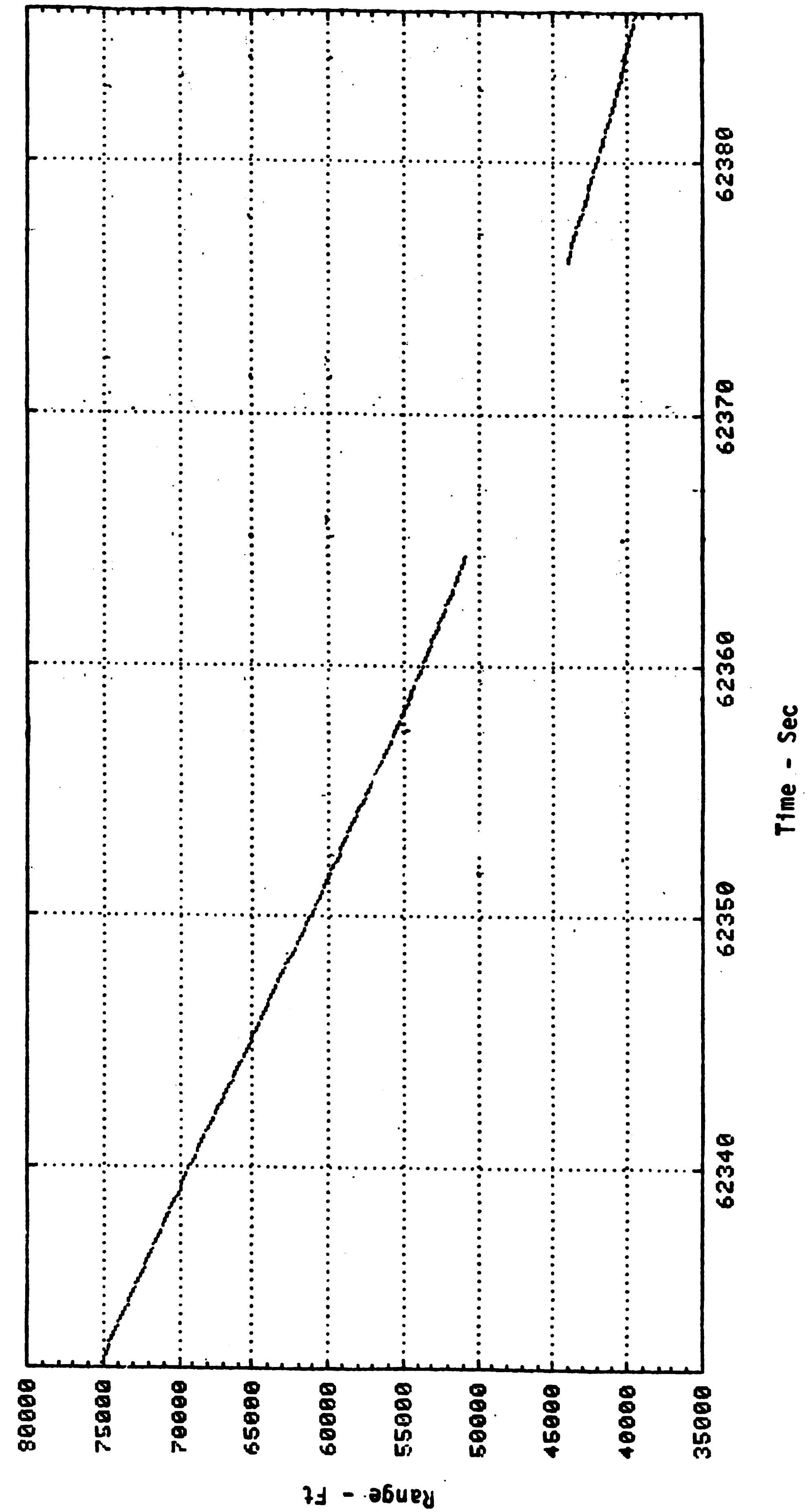
SOME EXAMPLES

Several hundred data passes have been run with the fitting procedure described. Since there is an average of maybe five receivers on each data pass, the robust preprocessing method described has been used on more than one thousand measurement sequences. The method has performed successfully on all of these sequences. Most of these sequences are rather uneventful, having only a few isolated outliers. There have been some sequences which have some rather dense bursts of outliers. These sequences best illustrate the ability of the method described to detect outliers. Fig 1 presents a range measurement sequence and Fig 2 the robust cubic spline fit to this sequence. Note that the outliers in Fig 1, which have been darkened, occur in many sizes. The outliers at the top of the graph were added by hand since they all occurred far off scale at the top. The sequence of Fig 1 has about 10% outliers. All outliers have been successfully detected and removed by the robust spline fit. The measurements in Fig 1 have two dense burst of outliers, one in the interval (62356.6, 62359.5) and another in the interval (62367.2, 62375.4). The measurement sequence in Fig 3 has outlier bursts in the intervals (62358.4, 62362.6), (62374.4, 62376.4), and (62379.7, 62382.4). The sequence in Fig 3 has about 15% outliers. The sequence in Fig 5 has bursts of outliers during the intervals (63117.8, 63122.6) and (63128.2, 63131.5). Any points away from the main curve should be considered outliers in Figs 1, 3, and 5. Note also in Figs 1, 3, and 5 that there are time breaks in the measurement sequences, another important consideration in preprocessing. The cubic spline fit to the sequence of Fig 1 is given in Fig 2. The cubic spline fit to the sequence of Fig 3 is given in Fig 4 and the cubic spline fit to the measurement sequence in Fig 5 is given in Fig 6. The knot intervals in Figs 2, 4, 6 are designed to contain twenty data points. Note that some of the time breaks have been filled with fitted data points. The filling of the time breaks is controlled by the length of the time breaks in relation to the sample rate and the proportion of outliers found in a knot interval. The robust cubic spline preprocessor has deleted all outliers from the measurement sequence, generated measurements during the time breaks as desired, and synchronized different measurement sequences if desired. In addition the measurement variances are available for further processing. The IRWLS cubic spline fit converged in 3 - 4 iterations for the examples displayed. This fairly rapid convergence is dependent on a robust method for choosing good initial weights. Surprisingly, the IRWLS cubic spline iteration for these examples also converges using an unweighted least squares start, but at the expense of more iterations. For the measurement sequences displayed here the IRWLS cubic spline fit converged in 7 - 8 iterations using an unweighted least squares start. Thus, at least in these examples, a good choice of the initial weights results only in a significant improvement in computing efficiency and not in an improvement of fit. Besides a good selection of initial weights, another important choice is the number of data points per knot interval, NPTO. NPTO must be large enough so that is likely that only a fraction, say less than one fourth of the data points in any interval will be outliers. On the other hand, if NPTO is too large, the robust linear curve fit may not be a good enough representation of the variation of the data in the interval.



Range - Ft
80000
75000
70000
65000
60000
55000
50000
45000
40000
35000
Time - Sec
62340
62350
62360
62370
62380

FIGURE 1



Range - Ft
80000
75000
70000
65000
60000
55000
50000
45000
40000
35000
62340
62350
62360
62370
62380
Time - Sec

FIGURE 2

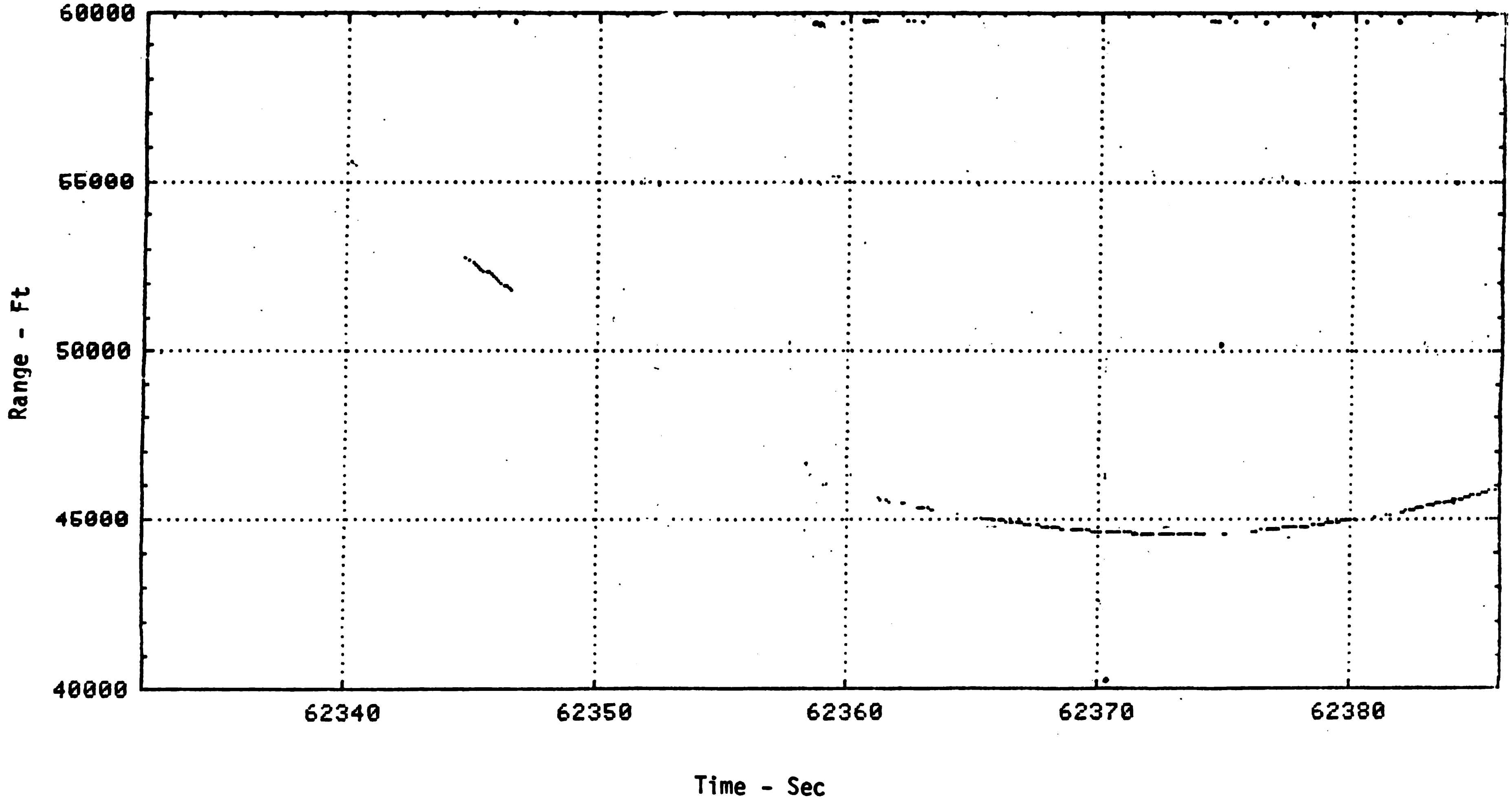
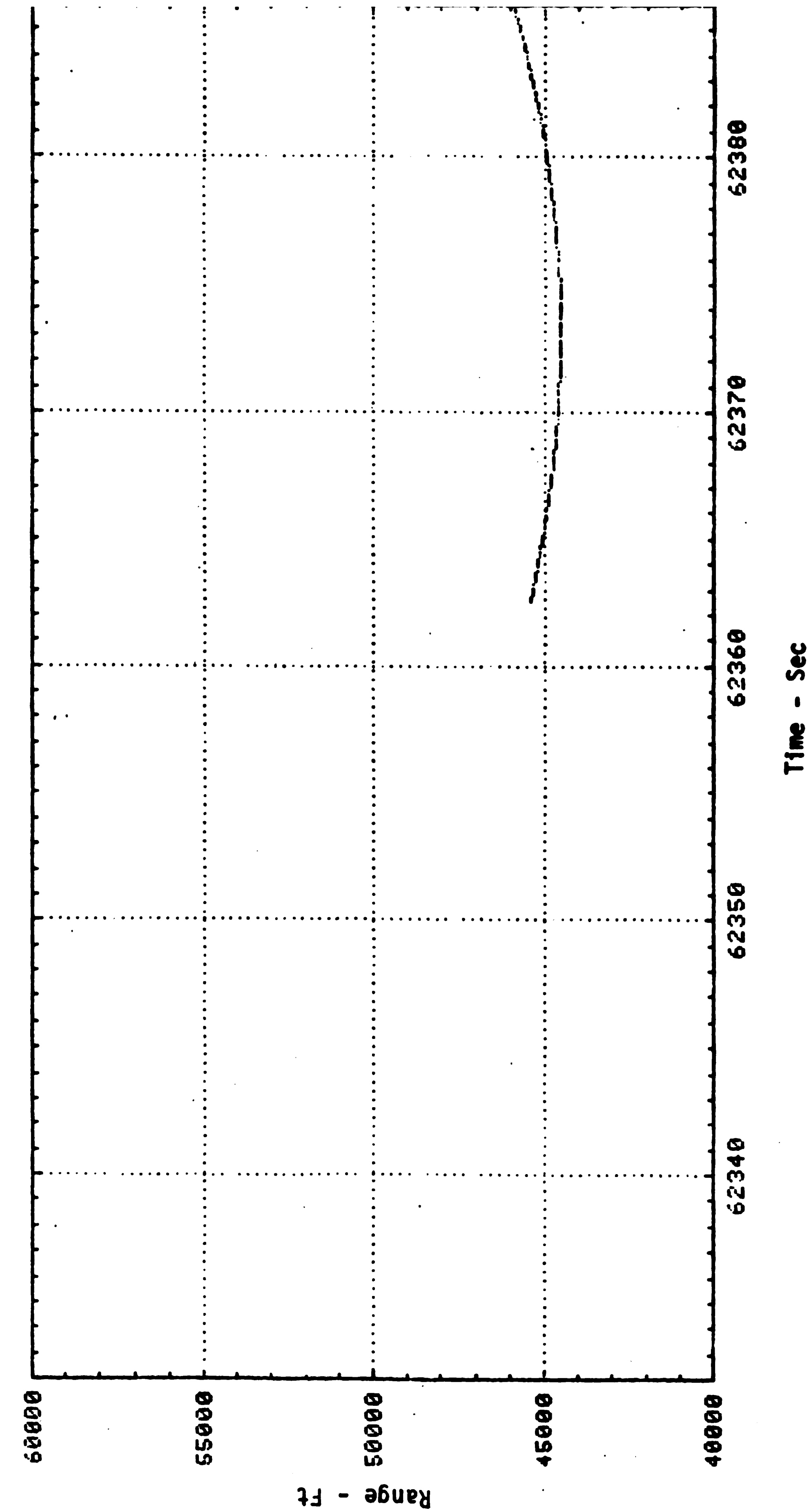

60000
55000
50000
45000
40000
Range - Ft
62340
62350
62360
62370
62380
Time - Sec

FIGURE 3



Range - Ft
60000
55000
50000
45000
40000
Time - Sec
62340
62350
62360
62370
62380

FIGURE 4

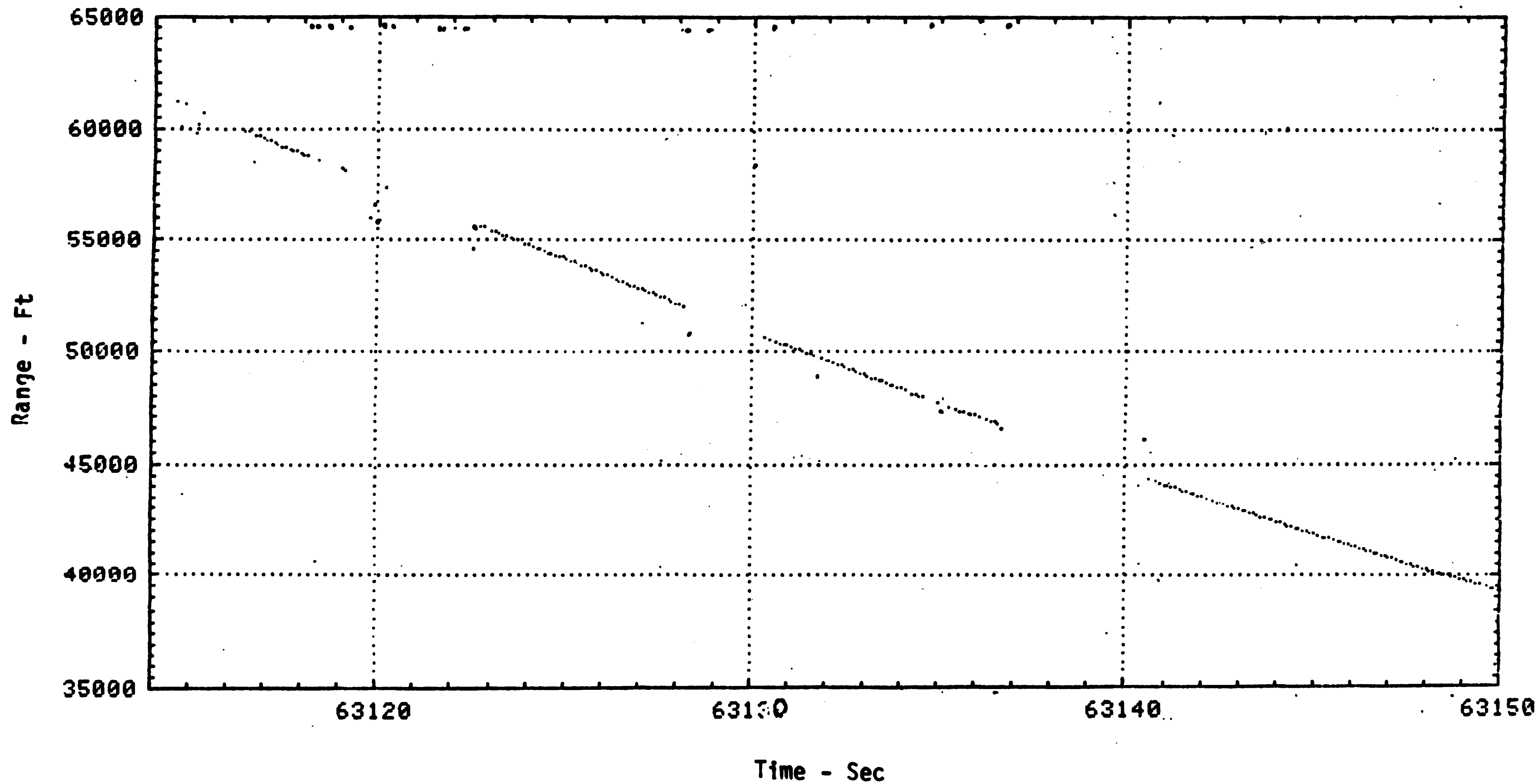

65000
60000
55000
50000
45000
40000
35000
Range - Ft
63120
63130
63140
63150
Time - Sec

FIGURE 5

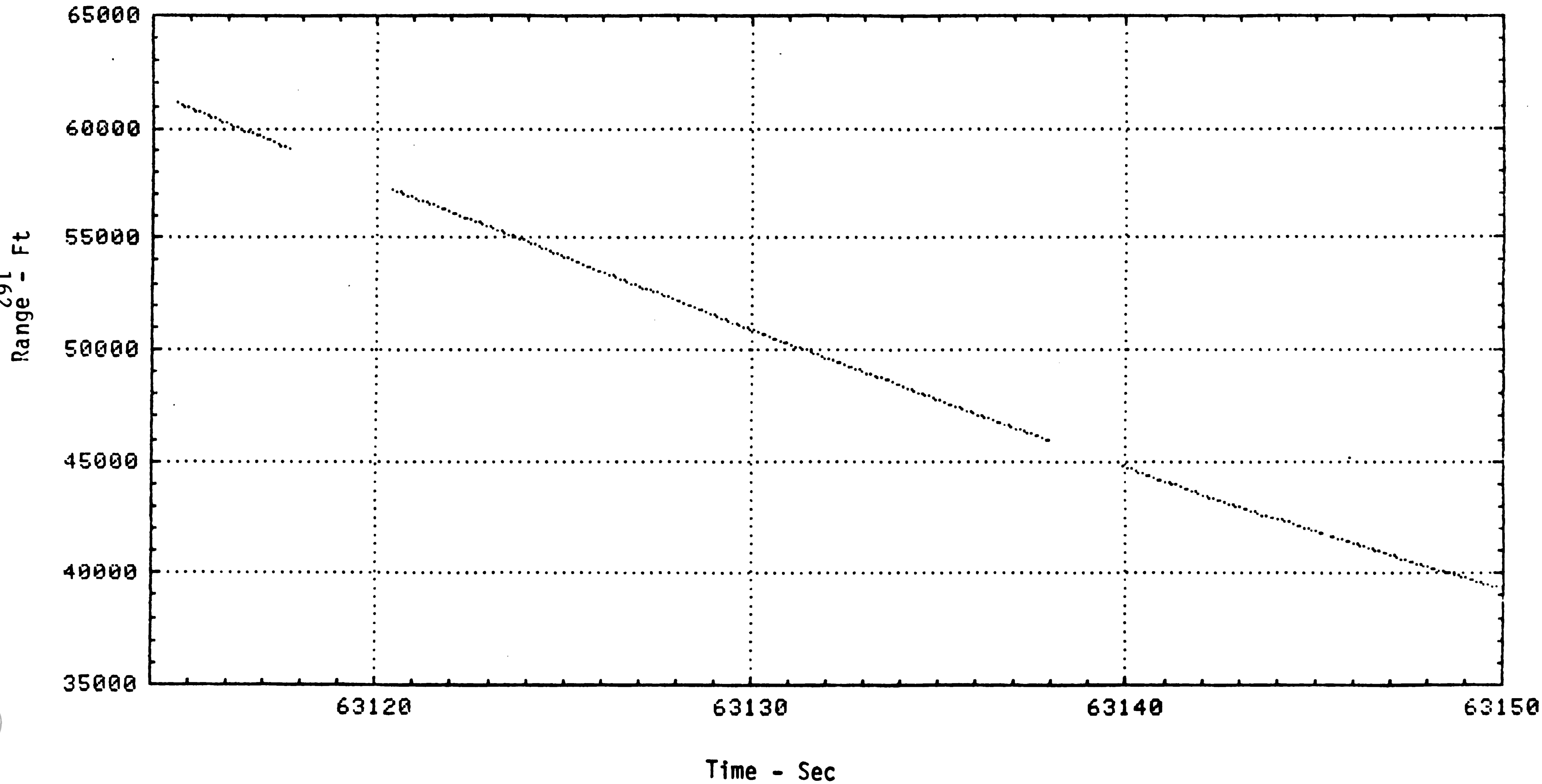

65000
60000
55000
50000
45000
40000
35000
Range - Ft
63120
63130
63140
63150
Time - Sec

FIGURE 6

APPENDIX

This appendix describes several methods of choosing the initial weights for the robust cubic spline preprocessing and compares the results of using these methods on several data sets. Each of these methods robustly fits a linear curve in each of the knot intervals and then computes the initial weights from the curve fit residuals using a Hampel ψ-function. Let $R(t_i)$, t_i = 1, NPTS(k) be the range measurements in the k^{th} knot interval.

Theil Method

Define the slopes s_{ij}

$$S_{ij} = \frac{R(t_j) - R(t_i)}{t_j - t_i} \qquad j>i$$

Let $\bar{s}$ be the median of these slopes,

$$\bar{s} = \underset{\substack{i,j \\ j>i}}{\text{median}} \{s_{ij}\}$$

Define the residuals $\bar{r}_i$,

$$\bar{r}_i = R(t_i) - \bar{s}\, t_i$$

Let $\bar{r}$ be the median of the residuals, $\bar{r}_i$

$$\bar{r} = \underset{i=1,NPTS(k)}{\text{median}} \{\bar{r}_i\}$$

Then the residuals $r_i = \bar{r}_i - \bar{r}$ are used to compute the initial weights with a Hampel ψ-function.

Nested Medians

Nested or repeated medians is a robust regression method recently described by Siegel [1]. Siegel shows that this method has the highest breakdown method of any known method. This method is particularly easy to apply for a linear fit. It is similar to the Theil method and modified Theil method already described.

Define the slopes s_{ij},

$$s_{ij} = \frac{R(t_j) - R(t_i)}{t_j - t_i} \quad i \neq j \tag{A-1}$$

Define $\bar{s}_i$ by

$$\bar{s}_i = \underset{\substack{j=1,NPTS(k)\\ j\neq i}}{\text{median}} \{s_{ij}\} \tag{A-2}$$

and further let $\bar{s}$ be defined by

$$\bar{s} = \underset{i=1,NPTS(k)}{\text{median}} \{\bar{s}_i\} \tag{A-3}$$

Similarly, let a_{ij} be the intercepts

$$a_{ij} = \frac{R(t_i)t_j - R(t_j)t_i}{t_j - t_i} \quad j \neq i \tag{A-4}$$

Define $\bar{a}_i$ as

$$\bar{a}_i = \underset{\substack{j=1,\ NPTS(k)\\ j\neq i}}{\text{median}} \{a_{ij}\} \tag{A-5}$$

and further define $\bar{a}$ by

$$\bar{a} = \underset{i=1,NPTS(k)}{\text{median}} \{\bar{a}_i\} \tag{A-6}$$

Let r_i be the residuals

$$r_i = R(t_i) - \bar{a} - \bar{s}\, t_i, \; i=1, NPTS(k) \tag{A-7}$$

The weights $W_i^{(0)}$ are computed from these residuals using a Hampel ψ-function.

The following data sets were taken from the knot intervals of the data sequences used previously to illustrate the application of the robust range measurement preprocessing. The first data set, shown in Fig A1 is taken from the measurement sequence given in Fig 1. The measurements in this set are from the time interval 62356.6 - 62359.5. The second data set, shown in Fig A2 is taken from the measurement sequence in Fig 5. The measurements in this set are from the time interval 63128.2 - 63131.6. In each of the data sets the weights are calculated from the residuals r_i by

$$W_i^{(0)} = \frac{\psi(r_{i/s})}{(r_{i/s})} \tag{A-8}$$

where $\psi(\cdot)$ is a Hampel ψ-function with breakpoints 2., 3., 4. In both of those data sets there are eight outliers in the sample of twenty. Each of the robust linear methods seem to have no difficulty in identifying the outliers in these data sets.

TIME	RANGE	MODIFIED THEIL RES	NESTED MEDIAN RES	THEIL RES	MODIFIED THEIL WEIGHT	NESTED MEDIAN WEIGHT	THEIL WEIGHT
.6	56083	-8	0	5	1	1	1
.7	56004	-11	-5	0	1	1	1
.8	55938	-2	2	7	1	1	1
.9	57336	1471	1473	1478	0	0	0
1.0	97546	41757	41757	41761	0	0	0
1.1	74022	18308	18307	18311	0	0	0
1.2	55190	-447	-451	-447	0	0	0
1.4	55478	-8	-15	-12	1	1	1
1.5	55400	-11	-20	-16	1	1	1
1.6	55321	-14	-25	-22	1	1	1
1.7	55249	-11	-23	-21	1	1	1
1.8	55170	-14	-29	-26	1	1	1
1.9	55111	2	-14	-12	1	1	1
2.0	54009	-1024	-1042	-1040	0	0	0
2.1	99357	44399	44379	44380	0	0	0
3.1	172631	118428	118389	118389	0	0	0
3.2	59980	5853	5812	5811	0	0	0
3.3	54094	42	0	-1	1	1	1
3.4	54022	46	1	0	1	1	1
3.5	53950	49	3	1	1	1	1

TIME	RANGE	MODIFIED THEIL RES	NESTED MEDIAN RES	THEIL RES	MODIFIED THEIL WEIGHT	NESTED MEDIAN WEIGHT	THEIL WEIGHT
.2	94462	43071	43013	42937	0	0	0
.3	50859	-465	-521	-594	0	0	0
.5	59606	8411	8362	8297	0	0	0
.6	63825	12696	12649	12588	0	0	0
.7	122290	71226	71182	71125	0	0	0
1.0	58359	7491	7456	7410	0	0	0
1.1	79974	29172	29139	29097	0	0	0
1.3	50699	-2	-29	-64	1	1	1
1.4	50584	-21	-45	-77	1	1	1
1.5	169737	119196	119175	119148	0	0	0
1.6	50472	-3	-21	-45	1	1	1
1.7	50387	-22	-38	-58	1	1	1
1.8	50335	-9	-22	-38	1	1	1
1.9	50275	-3	-13	-26	1	1	1
2.0	50223	9	2	-6	1	1	1
2.1	50131	-17	-21	-26	1	1	1
2.2	50085	2	0	0	1	1	1
2.3	50013	-4	-3	0	1	1	1
2.4	49947	-5	-1	6	1	1	1
2.5	49901	14	21	32	1	1	1

FIGURE A2

REFERENCES

1. Siegel, Andrew F., Robust Regression Using Repeated Medians, Princeton Univ., Dept of Statistics Technical Report #172, Series 2, Sept 1980.

2. Theil, H., A Rank-Invariant Method of Linear and Polynomial Regression Analysis, Indag. Math., V12, p85-91, p173-177, p467-482, 1950.

HOW GOOD IS GOOD - A FIELD EVALUATION OF CAMOUFLAGE

GEORGE ANITOLE and RONALD L. JOHNSON

U.S. Army Mobility Equipment Research and Development Command

Fort Belvoir, Virginia 22060

ABSTRACT

In development of prototype camouflage, performance characteristics are determined by field evaluations. The ultimate camouflage being that of a target with no restrictions on time or manpower in its erection or retraction, using standard materials and methods. This ultimate condition is not normally measured in field studies where prototype camouflage is evaluated against a base line target. Percentages of camouflage improvement in detection and identification rates by the prototype over base line conditions are identified. This paper concerns a real field study designed to allow the camoufleur to measure relative improvements of prototype camouflage against both base line and ultimate conditions. The camoufleur then ascertains true values of the prototype camouflage, or How Good is Good.

1.0 INTRODUCTION

The development of camouflage involves many phases and evaluations. In all cases, the developers goal is to produce the best camouflage possible within the restraints of time and manpower. One method of camouflage evaluation is through the conduct of field tests. The resulting data is analyzed, enabling the developer to determine the effectiveness of his product. The problem with this method of evaluation is that while it provides a good estimate of the prototype camouflage performance, as compared against the base line camouflage condition, no information is available for a comparison against the ultimate camouflage condition. The ultimate camouflage condition is defined as having no restrictions on time or manpower in its erection or retraction, using standard materials and methods. Such a multiple comparison would allow the developer to determine the virtues of additional prototype refinement to approach the ultimate camouflage condition. This paper concerns a real field study designed to allow the developer to objectively measure the relative effectiveness of prototype camouflage against both the base line and the ultimate camouflage conditions.

2.0 TEST SITE AND EQUIPMENT

2.1 Test Site

The test site was located at the USAF, Avon Park Bombing Range near Avon Park, Florida (60 miles south of Orlando, Florida). This range had an average elevation of 68 feet above mean sea level, and is flat in topography with mixed Oak, Pine, and Palmetto tree hammocks dispersed in sawgrass. The land is primarily swampy in nature and displays a light green color throughout.

2.2 Test Equipment

The equipment tested consisted of two trailers with a prime mover and were over 40 feet long, 7 feet high and 8 feet wide. The equipment was tactically emplaced along a tree line (one was emplaced outside the treeline and one was partially concealed by the treeline).

3.0 CAMOUFLAGE CONDITIONS

3.1 Pattern Painted (Base Line)

The two test items were pattern painted the tropic color blend in accordance with TC5-200[1/]. This color blend consisted of 45 percent forest green, 45 percent dark green, 5 percent light green and 5 percent black.

3.2 Pattern Painted with the Addition of Prototype Camouflage

The second camouflage condition was achieved by the addition of camouflage kits to the two patterned vehicles. These camouflage kits were constructed of special supports, and modified lightweight net screening.

3.3 Pattern Painted with the Addition of Standard Camouflage Screens and Techniques (Ultimate)

The third camouflage condition was obtained by using the U.S. Army Standard Lightweight Camouflage Screening System deployed over the vehicles. The screens were erected in accordance with TM5-1080-200-10[2]/. Camouflage improvements were made to the screens by tying natural foliage to the support systems, around the screen edges, and protruding through the screens. Oak leaf mulch was dispersed around the edge of the camouflage screens to break up the straight line edge effect. Palmetto fronds and sawgrass clumps were placed over the batten spreaders of the support system to reduce shine and were woven into the camouflage screens to simulate the appearance of natural foliage.

4.0 TEST IMAGERY

The site was photographed using 9 inch strip color, aerial film at scales of 1:5,000 and 1:10,000[3]/ each with 60% forward overlap. The target location was identical for each camouflage condition. The end product was three strips of imagery at each of the scales of 1:10,000 and 1:5,000. The 1:10,000 scale of imagery was 17 frames long while the 1:5,000 scale imagery was 5 frames in length.

5.0 TEST PROCEDURES

The subjects consisted of 99 pairs of operational Image Interpreters (II's). They were instructed on the purpose and tasks to be performed. Each team had three-quarters of an hour to detect targets on one of the strips of imagery scaled 1:10,000. At the end of this time period, the II's were given the corresponding strip of imagery scaled 1:5,000 and a set of equipment keys that they studied in an attempt to identify the two targets of interest. They were allowed 15 minutes to determine an identification. No team of II's viewed more than one camouflage condition.

6.0 RESULTS

The percentages of detection for each of the two test items were determined for each of the three camouflage conditions. A statistical[4]/ analysis of the data revealed that of the two items investigated, the item embedded in the trees indicated no significant differences between the percentages of detection for the three conditions of camouflage identified in Section 3.0. The item not as deeply embedded in the trees was identified by more II's and yielded significant differences between camouflage conditions as follows:

- The pattern painted item was detected significantly ($\alpha = 0.025$) more often than the pattern item with standard camouflage screens. (Ultimate Condition)

- The pattern painted item with prototype camouflage was detected significantly ($\alpha = 0.025$) more often than the pattern painted item with standard camouflage screens. (Ultimate Condition)

The percentages of identification for each of the two test items were determined for each of the three conditions. A statistical analysis of the data revealed that both target items yielded significant differences between camouflage conditions as follows:

o The pattern painted items were identified significantly ($\alpha = 0.025$) more often than the pattern painted items with the addition of prototype camouflage.

o The pattern painted items were identified significantly ($\alpha = 0.025$) more often than the pattern painted items with the addition of camouflage screens and techniques. (Ultimate Conditions)

7.0 DISCUSSION

The results of the study indicated that the design of the experiment was successful in statistically evaluating the base line camouflage condition against both the prototype camouflage condition and the ultimate camouflage condition. A look at the detection data indicates that for the item not embedded in the trees, the prototype camouflage yielded significantly ($\alpha = 0.025$) less detections than the base line camouflage condition. However, it was detected significantly more than the ultimate camouflage condition. This finding tells the camoufleur, that while the prototype camouflage for the item has decreased detections, more development is required to bring it up to speed with the ultimate camouflage condition. However, the trade off of the amount of time and manpower required to decrease the number of detections must be considered. The number of detections for the item embedded in the trees is so low that no further development is necessary.

The experimental design was also successful in statistically evaluating the base line camouflage condition against the prototype and ultimate camouflage conditions for the task of item identification. In this study, both the prototype and ultimate camouflage conditions yielded significantly ($\alpha = 0.025$) less identifications than the base line condition. There was no significant difference between the number of identifications for the prototype and ultimate camouflage conditions. This fact would indicate to the camoufleur that no additional refinement is required to reduce rate of identification.

8.0 SUMMARY

The purpose of this study was to design the field evaluation of a camouflage system in such a manner that the camoufleur could statistically differentiate between the base line, prototype, and ultimate levels of camouflage for both rates of detection and identification. With this information, the camoufleur could determine the feasibility of additional prototype development to approach the effectiveness of the ultimate camouflage condition. The ultimate camouflage condition was defined as having no restrictions on time or manpower in its erection or retraction using standard materials and methods. This study was conducted in Avon Park, Florida. The results from the data, using operational II's showed that the desired experimental discrimination between camouflage levels was achieved.

REFERENCES:

1. U.S. Army Combat Arms Training Board, Camouflage Pattern Painting, TC5-200, 28 August 1975

2. Headquarters, Department of the Army, Operator's Manual Camouflage Screen System, TM-5-1080-200-10, September 1978

3. The Naval Reconnaissance and Technical Support Center, Image Interpretation Handbook, Vol 1, TM-30-245, December 1967

4. Natrella, Mary G., Experimental Statistics, National Bureau of Standards, Handbook 91, U.S. Department of Commerce, Washington, D.C. 1966

ESTIMATING MEAN LIFE FROM LIMITED TESTING

Donald W. Rankin
Army Materiel Test and Evaluation Directorate
US Army White Sands Missile Range
White Sands Missile Range, New Mexico 88002

ABSTRACT. Exact probability formulae are developed, with no restrictive assumptions, for use with tests which produce data of the constant failure rate type. Although universally valid, the formulae are particularly apropos when straitened test circumstances are dictated. Programming suggestions are included.

1. INTRODUCTION. This paper is a sequel to the one entitled *Estimating Reliability from Small Samples* and presented before the twenty-second conference on the Design of Experiments in October 1976 [4].

The Poisson distribution is treated in a manner parallel to that afforded the binomial distribution in the earlier paper.

2. DEFINITION OF *EVENT*. Probability statistics require the identification of a unit commonly called *event* or *trial*. Often this identification is self-evident. Suppose a test consists of drawing a sample of specified size (n, say) from a larger population of similar items, then determining the number of defective items (k) in the sample. It requires no stretch of the imagination to say that drawing that sample of size n constitutes an *event* or *trial* and that the failure ratio k/n is the *result* of that event. It is to be noted that the failure ratio is dimensionless; i.e., k and n are measured in the same units.

Identification is not always so clear-cut. For example, suppose an operator of heavy trucks notices that in the preceding six months, he has experienced 13 major mechanical breakdowns--one every two weeks, on the average. The definition of *failure* is obvious, but what is a *success*? To what do we add k to get n, the sample size? The mathematical answer is that $n \to \infty$. But this is also a useless answer; no realistic test design could require an infinite sample size.

To avoid facing this dilemma, let us arbitrarily define *event* in some convenient unit different from that in which k is expressed. As a consequence, we no longer have a failure ratio. In its place we substitute a failure rate--of k per event. Thus the failure rate depends upon an observed k, but upon a defined *event*.

To return to the truck operator, let us say that examination of the log books reveals a total operating mileage of 267150 for the period in question. This figure (267150 miles) is taken as the definition of *event*. The observed failure rate then becomes

$$\frac{13 \text{ failures}}{267150 \text{ miles}} = 0.0000486667 \text{ failures per mile.}$$

It is sometimes regarded as preferable to express the reciprocal of the failure rate, calling it *mean life*. Thus we would have

$$\frac{267150 \text{ miles}}{13 \text{ failures}} = 20550 \text{ mean miles between failures.}$$

The term *event* can be defined in any of a variety of units--area, volume, weight, time--almost anything that can be measured.

3. POISSON PROBABILITY. Consider the well-known series

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}, \ (k = 0, 1, 2, 3, \ldots) \tag{1}$$

This series converges for all finite values of x, provided only that x remains constant. Multiplying by e^{-x} produces

$$1 = \sum_{k=0}^{\infty} \frac{x^k e^{-x}}{k!} \tag{2}$$

Poisson noted (1837) that if x is a constant failure rate and k is a non-negative integer, the probability of observing exactly k failures during an event is given by the appropriate term of the above expansion; i.e., by

$$p(k) = \frac{x^k e^{-x}}{k!} \tag{3}$$

This last expression, then is a probability function in the discrete variable k. Unfortunately, however, it does not suffice. In most test designs, it will be possible to define *event* arbitrarily and to observe the value of k exactly, but nothing will be known about x. Usually, in fact, x will be the principal value sought. A probability function in x is required.

Now x can take on any non-negative value; i.e., it is a <u>continuous</u> variable within the limits $0 \leq x < \infty$. Necessarily

$$1 = \int_0^{\infty} f(x)\, dx$$

defines f(x) as the required probability function in x, whatever form it may take. With k fixed, the expression

$$\frac{x^k e^{-x}}{\underline{|k}}$$

becomes a density function in x (though not necessarily a probability function). It is necessary to evaluate the definite integral

$$I_k = \int_0^{\infty} \frac{x^k e^{-x}}{\underline{|k}}\, dx$$

Since k is constant, $\underline{|k}$ can be taken outside the integral sign, leaving

$$\underline{|k}\, I_k = \int_0^{\infty} x^k e^{-x}\, dx = \Gamma\,(k + 1)$$

but also, k is an integer, hence $\underline{|k} = \Gamma\,(k + 1)$.

It is seen that $I_k = 1$, and therefore that

$$f(x) = \frac{x^k e^{-x}}{\underline{|k}}$$

is the required probability function in the continuous variable x.

It is helpful to inspect graphs of the probability function

$$f(x) = \frac{x^k e^{-x}}{\underline{|k}} \tag{4}$$

Several are depicted in Figure 1 for various integer values of k. Among features which should be noted are the following:

1) When k = 0, the function degenerates to

$$f(x) = e^{-x} \tag{5}$$

and is most easily treated as a separate case.

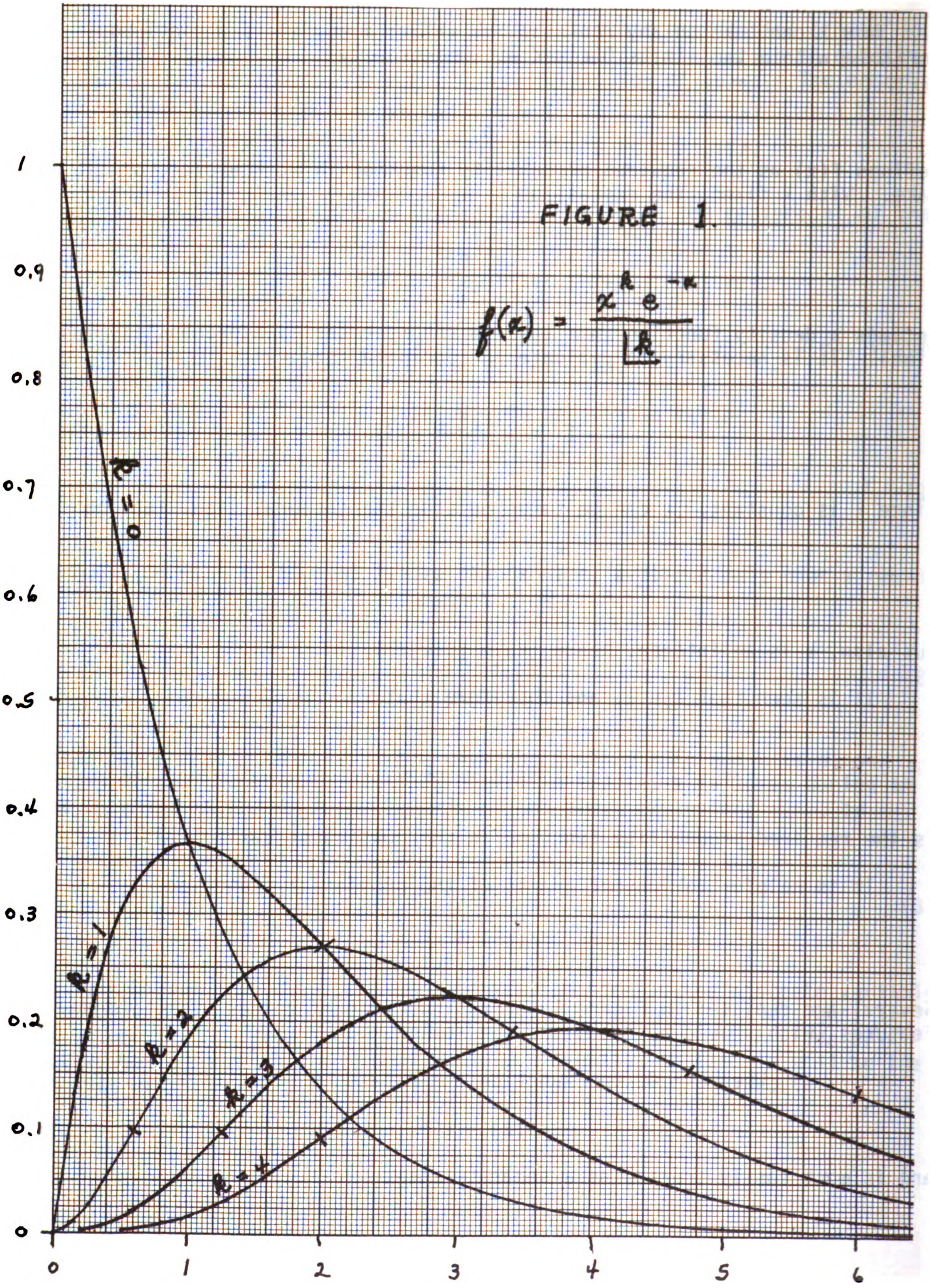


FIGURE 1.
f(x) = x^k e^-x / |k
k = 0
k = 1
k = 2
k = 3
k = 4
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
1
2
3
4
5
6

2) $$f'(x) = \frac{1}{|\underline{k}} \{kx^{k-1} e^{-x} - e^{-x} x^k\} = \frac{x^{k-1} e^{-x}}{|\underline{k}} (k - x) \qquad (6)$$

Thus a maximum occurs when $x = k$.

3) $$f''(x) = \frac{x^{k-2} e^{-x}}{|\underline{k}} = \{k(k-1) - 2kx + x^2\} \qquad (7)$$

A point of inflection is found whenever

$$x^2 - 2kx + k(k - 1) = 0,$$

i.e., when $x = k \pm \sqrt{k}$. For some programming purposes, when $k = 1$, the origin may serve as the missing point of inflection. The slope there is unity.

4) Every curve crosses every other curve exactly once, and in consecutive order.

5) Two consecutive curves intersect at the maximum point of the second, since the only non-trivial solution of

$$\frac{x^k e^{-x}}{|\underline{k}} = \frac{x^{k+1} e^{-x}}{|\underline{k+1}}$$

occurs when $x = k + 1$.

4. <u>TRANSFORMING THE PROBABILITY FUNCTION</u>. If the case $k = 0$ is treated separately, the transformation $w = x/k$ suggests itself. Letting $x = kw$, $dx = kdw$ and it is seen that

$$\int_{kw=o}^{\infty} \frac{(kw)^k e^{-kw}}{|\underline{k}} kdw = 1,$$

since merely employing the transformation will not affect the value of this definite integral. But the probability function <u>in w</u> is

$$g(w) = \frac{ke^{-kw} (kw)^k}{|\underline{k}} \qquad (8)$$

Basically, this transformation rescales the abscissae by $1/k$, and hence the densities (ordinates) by k, thereby preserving area. Several graphs of this function are shown in Figure 2. Notice that every curve has its maximum point at $w = 1$. Also, $\bar{w} \to 1$. Points of inflection occur at $1 \pm 1/\sqrt{k}$.

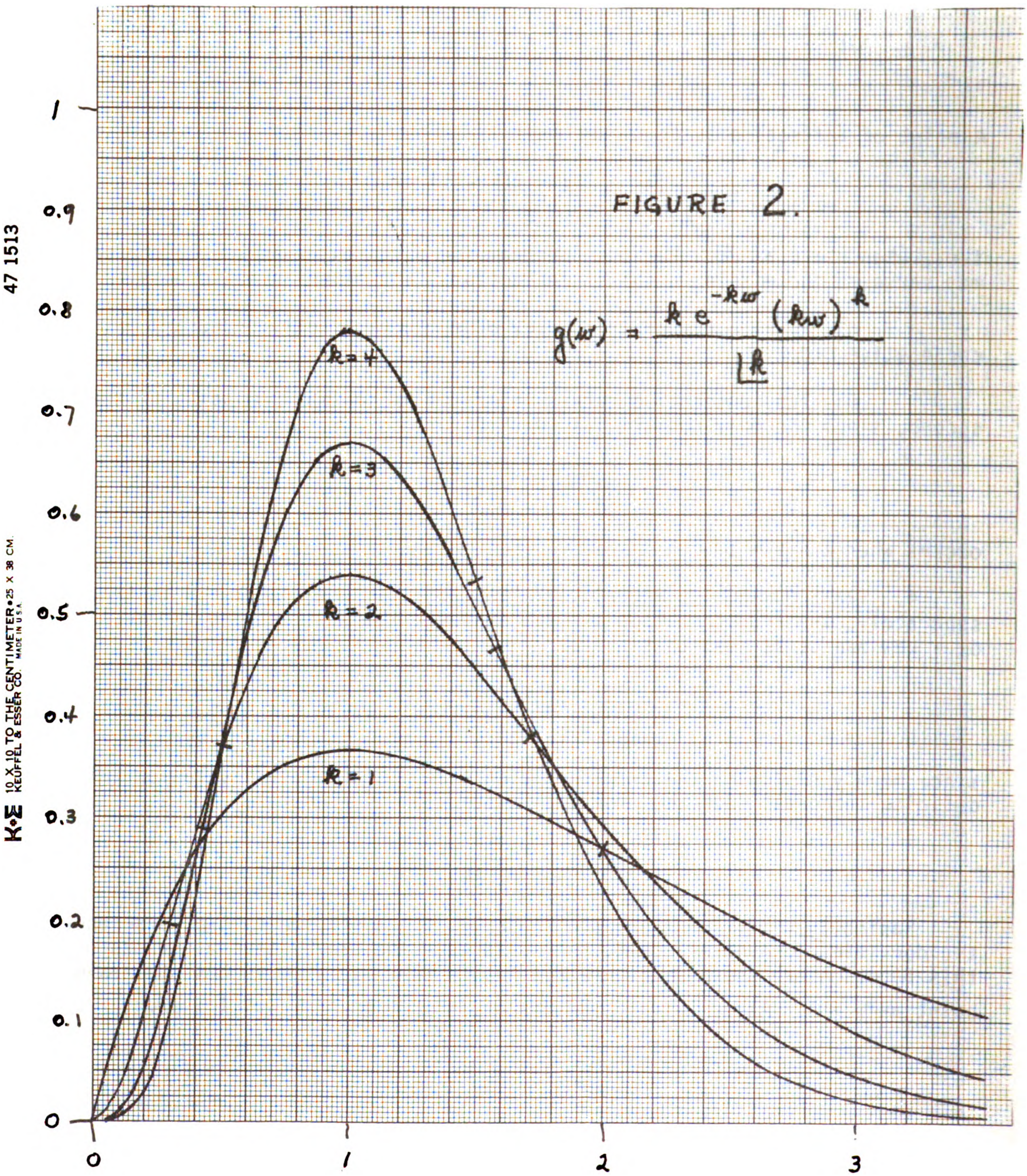


FIGURE 2.
$g(w) = \frac{k e^{-kw} (kw)^k}{\lfloor k}$
k = 4
k = 3
k = 2
k = 1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
1
2
3
K+E 10 X 10 TO THE CENTIMETER • 25 X 38 CM.
KEUFFEL & ESSER CO. MADE IN U.S.A.
47 1513

Although the transformation is useful for studying this family of functions, it matters very little whether levels of confidence are computed from

$$\int_{x=0}^{z} f(x)\, dx \quad \text{or} \quad \int_{w=0}^{z/k} g(w)\, dw.$$

In this paper, the form in f(x) will be used.

5. INTEGRATION BY PARTS. When a function is defined by (or can be described as) a definite integral, very frequently it will be found that repeated integration by parts will produce an expansion suitable for computing. In fact, as in the instance at hand, it may be possible to expand in either ascending or descending factorials (or powers, as the case may be), thereby producing two different expansions, both of which are valid. Usually, one will appear in the familiar form of a power series which converges more rapidly for smaller values of the argument. The other will be the associated asymptotic expansion. If the parameter which appears in the factorial part of the probability function can be restricted to integer values only, the asymptotic expansion becomes finite in length and is an exact expression.

The sought probability integral can be stated

$$P(z) = \int_{x=0}^{z} f(x)\, dx = \int_{x=0}^{z} \frac{x^k e^{-x}}{\underline{|k}}\, dx \tag{9}$$

and gives the probability that x does not exceed the (perhaps arbitrary) value z.

Can the indefinite integral $\int \frac{x^k e^{-x}}{\underline{|k}} dx$ be evaluated by parts, k being a fixed, positive integer?

Let $u = e^{-x}$ and $dv = \frac{x^k}{\underline{|k}} dx$.

Then $du = -e^{-x}dx$ and $v = \frac{x^{k+1}}{\underline{|k+1}}$.

Thus $\int \frac{x^k e^{-x}}{\underline{|k}} dx = \frac{e^{-x} x^{k+1}}{\underline{|k+1}} + \int \frac{e^{-x} x^{k+1}}{\underline{|k+1}} dx$

It is apparent at once that the second integral is like the first, save k has been augmented by unity. It is clear that the process can be reapplied endlessly, yielding

$$\int \frac{x^k e^{-x}}{\underline{|k}} dx = \sum_{i=1}^{\infty} \frac{e^{-x} x^{k+i}}{\underline{|k+i}}, \ (i = 1, 2, 3, \ldots).$$

Passing to the lower limit of the definite integral ($x = 0$), the sum vanishes, since x factors every term. (It may be more correct to say that the sum reduces to the constant of integration.) Thus

$$P(z) = \int_{x=0}^{z} \frac{x^k e^{-x}}{\underline{|k}} dx = \sum_{i=1}^{\infty} \frac{e^{-z} z^{k+i}}{\underline{|k + i}} \tag{10}$$

The term-to-term recurrence ratio is $z/(k + i)$. Since z is constant while $(k + i)$ increases without bound, the series will (eventually) converge for all positive values of z.

Now let $u = \frac{x^k}{\underline{|k}}$ and $dv = e^{-x}dx$. Then

$$du = \frac{kx^{k-1}}{\underline{|k}} dx = \frac{x^{k-1}}{\underline{|k - 1}} dx \text{ and } v = -e^{-x}, \text{ whence}$$

$$\int \frac{x^k \; e^{-x}}{\underline{|k}} dx = -e^{-x} \frac{x^k}{\underline{|k}} + \int \frac{x^{k-1} \; e^{-x}}{\underline{|k - 1}} dx.$$

Noting that $-e^{-x}$ will factor every term, we can write the result in the form

$$\int \frac{x^k e^{-x}}{\underline{|k}} dx = -e^{-x} \left\{ \frac{x^k}{\underline{|k}} + \frac{x^{k-1}}{\underline{|k - 1}} + \; . \; . \; . + \frac{x^2}{2} + x + 1 \right\}$$

At the lower limit ($x = 0$), the right-hand member becomes

$$\lim_{x \to 0} \; -e^{-x} \left\{ \left(\frac{x^0}{\underline{|0}} \right) \right\} = -1 \left(\frac{1}{1} \right) = -1.$$

The definite integral thus is given by

$$P(z) = \int_{x=0}^{z} \frac{x^k e^{-x}}{\underline{|k}} dx = 1 - e^{-z} \left\{ \frac{z^k}{\underline{|k}} + \frac{z^{k-1}}{\underline{|k - 1}} + \; . \; . \; . \; z + 1 \right\} \tag{11}$$

Are the two solutions equivalent? Is it true that

$$\sum_{i=1}^{\infty} \frac{e^{-z} \; z^{k+i}}{\underline{|k + i}} = 1 - e^{-z} \left\{ 1 + z + \frac{z^2}{2} + \; . \; . \; . + \frac{z^{k-1}}{\underline{|k - 1}} + \frac{z^k}{\underline{|k}} \right\} \qquad ?$$

Multiplying by e^z and transposing, it is seen that

$$\left\{1 + z + \ . \ . \ . \ + \frac{z^k}{\underline{|k}}\right\} + \sum_{i=1}^{\infty} \frac{z^{k+i}}{\underline{|k + i}} = e^z$$

is the well-known Maclaurin series for e^z. Therefore the two solutions are indeed equivalent.

It is a fact that if the upper limit of integration be taken at the maximum ($w = 1$; i.e., $z = k$), the level of confidence will always be less than 1/2 and hence of little statistical interest. (See Table 1.) However, the argument $z = k$ has an important use of a different sort. It enables us to select a series for computing whose terms are known to decrease monotonically. This results in worthwhile economy for larger values of k. There are two cases to consider.

First: Let $0 < z < k$. The series

$$P(z) = \sum_{i=1}^{\infty} \frac{e^{-z}z^{k+i}}{\underline{|k + i}} \qquad (10)$$

is chosen for use. Obviously, the term-to-term recurrence ratio is given by $z/(k + i)$. Under the stated conditions, this is always less than unity.

Second: Let $z > k$. The formula

$$P(z) = 1 - e^{-z} \left\{\frac{z^k}{\underline{|k}} + \frac{z^{k-1}}{\underline{|k - 1}} + \ . \ . \ . \ + \frac{z^2}{2} + z + 1\right\} \qquad (11)$$

is used. The recurrence ratio is

$$\frac{k + 1 - i}{z}, \ (i = 1, 2, 3, \ . \ . \ ., k)$$

which again is less than unity. For large values of k, the interior series can be summed as though it were an infinite series, thus achieving a laudable saving in the number of terms required.

6. COMPUTING A LEVEL OF CONFIDENCE ($z > k$). The value of z may be derived from any source, or it may be arbitrarily specified. The proper formula, as we have seen, is

$$P(z) = 1 - e^{-z} \left\{\frac{z^k}{\underline{|k}} + \frac{z^{k-1}}{\underline{|k - 1}} + \ . \ . \ . \ + z + 1\right\} \qquad (11)$$

TABLE 1

CONFIDENCE LEVEL AT MAXIMUM ORDINATE

k	$\int_0^k f(x)\,dx$
0	0.000000
1	0.264241
2	0.323324
3	0.352768
4	0.371163
5	0.384039
6	0.393697
7	0.401286
8	0.407453
9	0.412592
10	0.416960
12	0.424035
15	0.431910
20	0.440907
30	0.451648
50	0.462483
100	0.473438
200	0.481206
400	0.486706
1000	0.491591

When k is small (k < 12, say), the resulting finite expression submits easily to direct computation. But when k is very large, two difficulties arise.

First: The number of terms becomes excessive. If the series is summed as though it were an infinite series--i.e., the relative size of each new term is observed--the process can be truncated when additional terms no longer affect the result in the computer.

Second: Large factorials will overflow the computer. To circumvent this, the first term of the series is computed by logarithms. Stirling's formula (k > 11) is given by

$$\ln_e \underline{|k} = 0.91893\ 85332 + \left(k + \frac{1}{2}\right) \ln_e k$$

$$- k + \frac{1}{12k} \left\{1 - \frac{1}{30k^2} \left(1 - \frac{2}{7k^2}\right)\right\} . \qquad (12)$$

The first term is (disregarding sign) $\frac{e^{-z}\, z^k}{\underline{|k}}$; hence, its logarithm will be $k \ln_e z - z - \ln_e \underline{|k}$, which should not cause overflow within the range of useful numbers.

7. COMPUTING z WHEN A LEVEL OF CONFIDENCE IS SPECIFIED. (L = P(z) > 0.5) No new formula is available for the inverse. Instead, successive approximations $z_0, z_1, z_2, \ldots$ are computed until a steady state is reached. Newton's method serves very well. See the discussion in [4] pp. 279-280.

For any z_i, compute $P(z_i)$, $f(z_i)$ and $f'(z_i)$. The required incremental area is of course $P(z) - P(z_i)$. We approximate this area with a trapezoid of width Δz whose ordinates are $f(z_i)$ and $f(z_i) + \Delta z f'(z_i)$. We have seen earlier that the first term of the wanted series for $P(z_i)$ is

$$\frac{e^{-z_i}\, z_i^k}{\underline{|k}}$$

$$\text{Also } f(z_i) = \frac{e^{-z_i}\, z_i^k}{\underline{|k}};$$

$$\text{and } f'(z_i) = \left\{\frac{k}{z_i} - 1\right\} f(z_i).$$

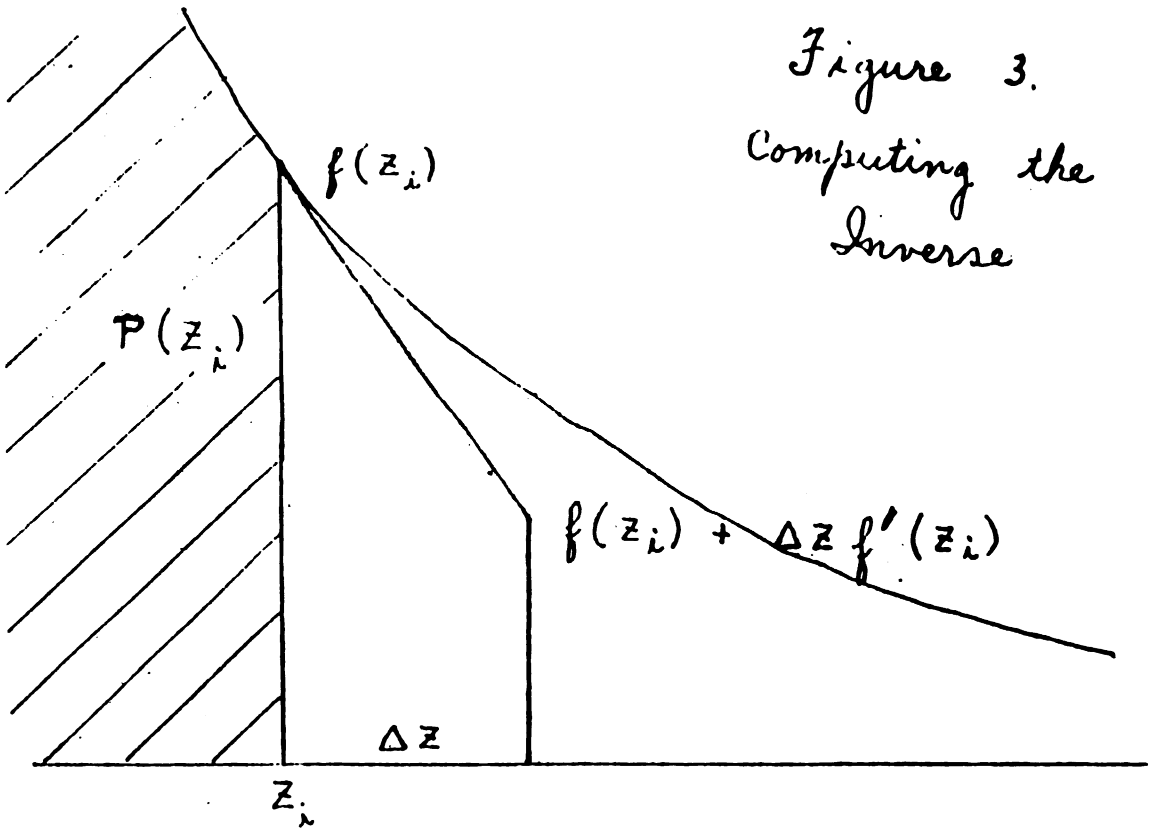


Figure 3.
Computing the Inverse

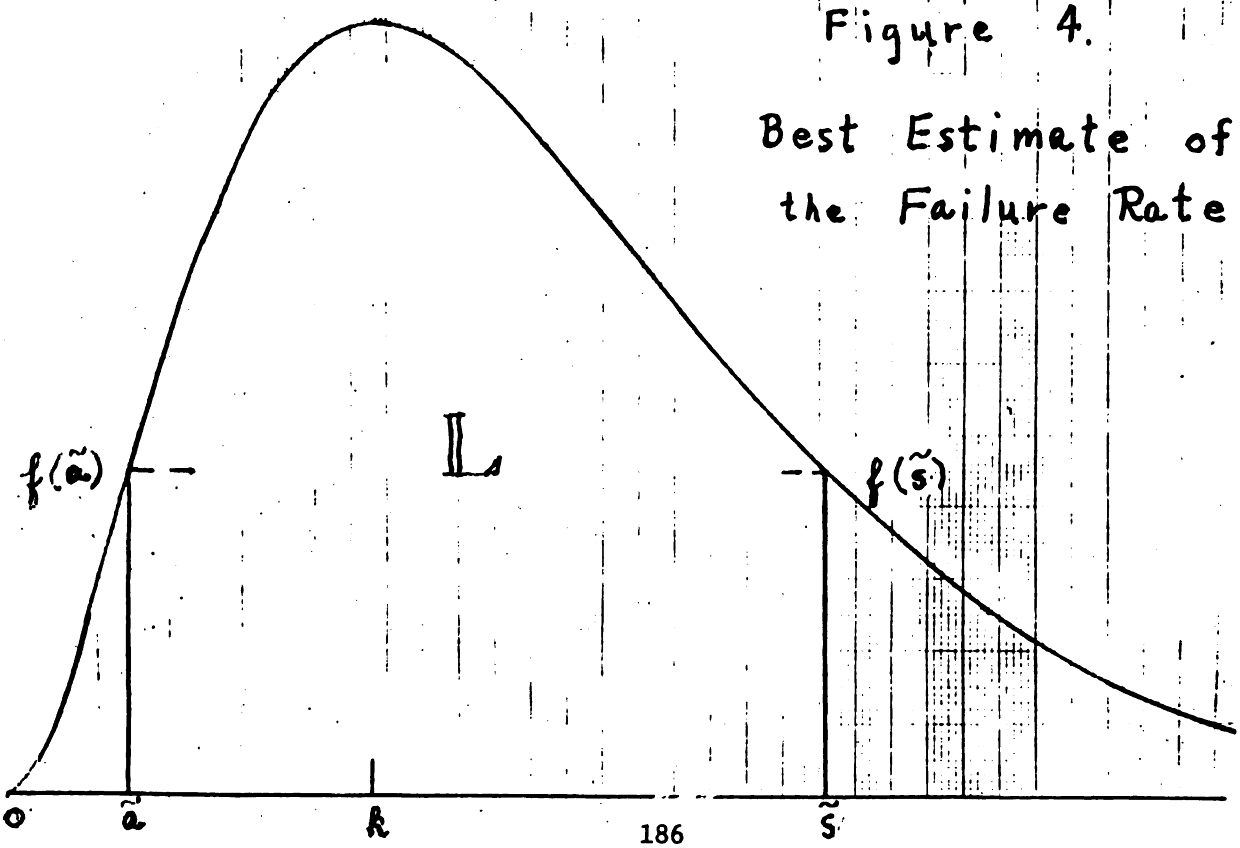


Figure 4.
Best Estimate of the Failure Rate

The approximating trapezoid is given by

$$P(z) - P(z_i) \approx \Delta z \left\{f(z_i) + \frac{\Delta z}{2} f'(z_i)\right\}$$

which can be solved for Δz.

$$\Delta z \approx \frac{-f(z_i) \pm \sqrt{[f(z_i)]^2 + 2f'(z_i)\,[P(z) - P(z_i)]}}{f'(z_i)}$$

Since ultimately $\Delta z \to o$, it is apparent that the positive square root yields the true solution. Noting that

$$\frac{f(z_i)}{f'(z_i)} = \frac{z_i}{k - z_i}$$

the formula can be simplified to

$$\Delta z \approx \frac{z_i}{z_i - k} - \sqrt{\left(\frac{z_i}{z_i - k}\right)^2 + \frac{2[P(z) - P(z_i)]}{f'(z_i)}} \qquad (13)$$

The process is stable when started from the right-hand point of inflection; i.e.,

$$z_0 = k + \sqrt{k} \qquad (14)$$

8. <u>THE BEST ESTIMATE OF THE FAILURE RATE.</u>* For a specified level of confidence L, the general solution of the probability integral is

$$L = \int_a^s f(x)\, dx.$$

There are, of course, an unlimited number of solution pairs (a, s) which satisfy this equation. Up to this point, we have concerned ourselves with the case a = o. This form properly is used to test for compliance with an imposed standard.

Sometimes, however, that standard is absent, unrealistic, or even erroneous. But it is still required to make a meaningful statement about the failure rate. In this situation, the *Best Estimate* is recommended. Essentially, that solution pair (a, s) is chosen which minimizes the difference $|s - a|$.

---

*See [4] pp. 267-270.

Values of a and s thus determined are designated by a tilde ($\tilde{a}$, $\tilde{s}$).

Some properties of the Best Estimate of the Failure Rate are:

a. $\tilde{s} - \tilde{a}$ is minimum, by definition.

b. The limits of integration lie on opposite sides of the maximum; i.e., $\tilde{a} < k < \tilde{s}$.

c. The ordinates at $\tilde{a}$ and $\tilde{s}$ are equal; i.e, $f(\tilde{a}) = f(\tilde{s})$.

d. The solution is unique.

There are several steps in the solution.

Step One. For any s_i, compute $f(s_i)$, $f'(s_i)$, $P(s_i)$.

(To begin, set $s_0 = k + \sqrt{k}$.)

Step Two. For each s_i, solve for the value $a < k$ such that $f(a) = f(s_i)$.

For any a_j, compute $f(a_j)$ and $f'(a_j)$.

Then

$$\Delta a = \frac{f(s_i) - f(a_j)}{f'(a_j)}. \tag{15}$$

The process is repeated until f(a) and hence a is found to the desired accuracy. This value of f(a) is then associated with $f(s_i)$ by appending the subscript i. (The subscript j is dropped, being no longer necessary.) For every new value of s_i, the a-process is begun afresh by setting $a_j = k - \sqrt{k}$.

Step Three. The value $f(a_i) = f(s_i)$ having been found, compute $P(a_i)$. (The values for a_i and $f'(a_i)$ will already have been computed.) The desired incremental area is $L - P(s_i) + P(a_i)$

Step Four. The incremental area always will appear in two separate parts. The ratio of these areas can be estimated quite closely by the slopes. Thus

$$\left(L - P(b_i) + P(a_i)\right)\left(\frac{f'(a_i)}{f'(a_i) - f'(s_i)}\right)$$

will appear on the right. It is convenient to express the ratio in terms of the ordinates.

$$\frac{f'(a_i)}{f'(a_i) - f'(s_i)} = \frac{\left(\frac{k - a_i}{a_i}\right) f(a_i)}{\left(\frac{k - a_i}{a_i}\right) f(a_i) - \left(\frac{k - s_i}{s_i}\right) f(s_i)}$$

But since $f(a_i) = f(s_i)$, this value can be cancelled from numerator and denominator, leaving

$$\frac{f'(a_i)}{f'(a_i) - f'(s_i)} = \frac{s_i(k - a_i)}{k(s_i - a_i)}. \tag{16}$$

Thus a suitable approximating trapezoid is given by

$$\left(L - P(s_i) + P(a_i)\right) \left(\frac{s_i(k - a_i)}{k(s_i - a_i)}\right) \approx \Delta s \; \{f(s_i) + \frac{\Delta s}{2} f'(s_i)\} \tag{17}$$

which can be solved for Δs by the method of Section 7, above.

9. EXPRESSING RESULTS IN TERMS OF MEAN LIFE. It should be noted that the methods developed in this paper are virtually independent of the definition of *Event*. (*Event* often will be synonymous with *Duration of Test*.) Suitable values of $\tilde{a}$ and $\tilde{s}$ (or z, as the case may be) having been found, it is apparent that they should be expressed in the units *failures per event*. If at this point the definition of *event* is imposed, the results can be expressed in *failures per mile* or *failures per hour* or whatever.

Now the simple reciprocal converts to *mean life*. It should be remembered that taking the reciprocal reverses the sense of inequality signs.

APPENDIX A

CHI-SQUARE AND OTHER POISSON-RELATED FUNCTIONS

Let us define the following special functions:

Incomplete exponential function:

$$e_n(x) = 1 + x + \frac{x^2}{\underline{|2}} + \ldots + \frac{x^n}{\underline{|n}}.$$

The series consists of n + 1 terms.

Gamma function:

$$\Gamma(x) = \int_0^\infty e^{-t}\, t^{x-1}\, dt \qquad (x > o)$$

$$\text{Thus } \Gamma(x + 1) = \int_0^\infty e^{-t}\, t^{x}\, dt.$$

Incomplete gamma function:

$$\gamma(x,z) = \int_0^z e^{-t}\, t^{x-1}\, dt \qquad (x > o)$$

and, of course, $o < z < \infty$.

Prym's function:

$$\Gamma(x,z) = \int_z^\infty e^{-t}\, t^{x-1}\, dt \qquad (z > o)$$

Immediately it is seen that

$$\gamma(x,z) + \Gamma(x,z) = \Gamma(x)$$

and that dividing both sides of this equation by $\Gamma(x)$ will produce a probability relationship.

Thus we can state

$$P(x,z) = \frac{\gamma(x,z)}{\Gamma(x)} = 1 - \frac{\Gamma(x,z)}{\Gamma(x)}.$$

Now for any particular problem, x and hence $\Gamma(x)$ will remain fixed. In terms of Prym's function we can write

$$P(x,z) = 1 - \frac{1}{\Gamma(x)} \int_z^\infty e^{-t} t^{x-1} dt.$$

It is easy to develop $\Gamma(x,z)$, using repeated integration by parts.* It is found that

$$\Gamma(x,z) \sim e^{-z} z^{x-1} \sum_{s=0}^{\infty} \frac{(x-1)(x-2)\ldots(x-s)}{z^s}$$

is a valid asymptotic expansion for fixed x and large z.

When x is an integer, the series terminates.

When x is not an integer, the terms of the series alternate in sign after $s > x$. The series diverges after $s > x + z$.

Let us replace x with k in the formulae in order that x can be employed as a variable of integration. Thus the formulae restated appear as follows:

$$P(k,z) = 1 - \frac{\Gamma(k,z)}{\Gamma(k)} = 1 - \frac{1}{e^z} [e_{k-1}(z)]$$

$$P(k,z) = 1 - \frac{e^{-z} z^{k-1}}{\Gamma(k)} \sum_{s=0}^{k-1} \frac{(k-1)(k-2)\ldots(k-s)}{z^s}$$

When k is a positive integer.

This case of k being a positive integer was studied at length in its application to sampling distributions by Helmert (1876) and K. Pearson (1900). Thus arose the statistics of the χ^2 distribution. The exponent 2 in χ^2 has little significance beyond ensuring that the parameter is non-negative.

*See [3] p. 66.

The χ^2 probability function* is defined by:

$$P(\chi^2|\nu) = \frac{1}{2^{\nu/2}\,\Gamma(\nu/2)} \int_0^{\chi^2} (t)^{(\nu/2)-1}\, e^{-t/2}\, dt$$

$$Q(\chi^2|\nu) = \frac{1}{2^{\nu/2}\,\Gamma(\nu/2)} \int_{\chi^2}^{\infty} (t)^{(\nu/2)-1}\, e^{-t/2}\, dt$$

$$P(\chi^2|\nu) + Q(\chi^2|\nu) = 1$$

Comparing this to the earlier-derived

$$P(k,z) = \frac{\gamma(k,z)}{\Gamma(k)} = \frac{1}{\Gamma(k)} \int_{x=0}^{z} e^{-x}\, x^{k-1}\, dx,$$

it is seen that the only differences are in the scaling of the parameters. For let $\nu = 2k$. Then

$$P(\chi^2|2k) = \frac{1}{2^k\Gamma(k)} \int_0^{\chi^2} t^{k-1}\, e^{-t/2}\, dt$$

$$= \frac{1}{\Gamma(k)} \int_0^{\chi^2} \left(\frac{t}{2}\right)^{k-1} \left(\frac{1}{2}\, e^{-t/2}\right) dt$$

Now let $t = 2x$, from which $dt = 2dx$. Replacing the variable of integration,

$$P(\chi^2|2k) = \frac{1}{\Gamma(x)} \int_{2x=0}^{2x=\chi^2} x^{k-1}\, e^{-x}\, dx$$

and it is seen that $\chi^2 = 2z$ properly scales the limit of integration.

When $\nu = 2k$ is an ODD integer, two things happen. $\Gamma(k)$ contains the factor $\sqrt{\pi}$ and $\sum_{s=0}^{\infty} \frac{(k-1)(k-2)\ldots(k-s)}{z^s}$ does not terminate. The behavior (accuracy) of the asymptotic expansion near $z \approx k - s$ must be investigated.

*See [1] 26.4 page 940.

APPENDIX B

PROGRAM PLANNING - POISSON

1. INTRODUCTION. As a general rule, the only variable of observation will be k, the number of failures. The variable of integration will be x, with z one of its extreme values (limits of integration).

It is necessary to define *event* in some suitable unit (time, distance, mass, volume, etc.); e.g., *event* = 4240 hours. *Event* often is synonymous with *Duration of Test*.

Many formulae of interest are greatly simplified if expressed as functions of f(x) or of f(z). Thus

$$f(x) = \frac{e^{-x} x^{k}}{\underline{|k}}$$

$$f'(x) = \left(\frac{k}{x} - 1\right) f(x)$$

$$f''(x) = \left(\frac{k^2 - k}{x^2} - \frac{2k}{x} + 1\right) f(x)$$

$$P(z) = \sum_{i=1}^{\infty} \frac{e^{-z} z^{k+i}}{\underline{|k + i}} \qquad (i = 1, 2, 3, \ldots)$$

$$= T_1 + T_2 + T_3 + \ldots + T_i + \ldots$$

$$T_1 = \frac{z}{k + 1} f(z) \text{ and}$$

$$T_j = \frac{z}{k + j} T_{j-1}$$

Also, $1 - P(z) = T_o + T_1 + T_2 + \ldots + T_i + \ldots$

$$T_o = f(z)$$

$$T_{j+1} = \frac{k - j}{z} T_j$$

This latter series terminates when $k = j$.

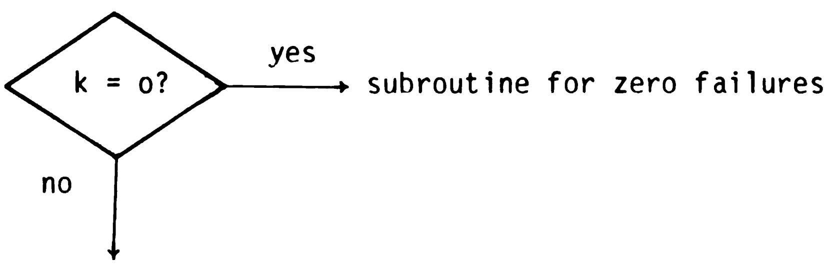
For large values of z, compute f(z) by logarithms.

$$\ln_e f(z) = k \ln_e z - z - \ln_e \underline{|k}$$

Stirling's formula for $\ln_e \underline{|k}$ is useful here. If k does not change, it need be computed but once.

2. COMPUTING L (z specified). Equations (4), (5), (9), (10) and (11).

Enter data



Compute $\ln_e \underline{|k}$. (If $k > 15$, use Stirling's formula.)

Compute $k \ln_e z - z$.

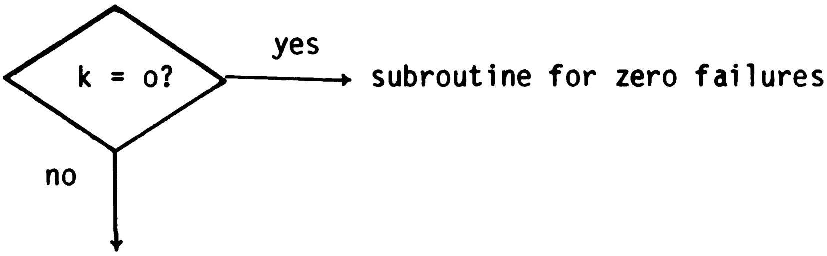
Compute $f(z) = \dfrac{z^k e^{-z}}{\underline{|k}} = T_o$

Compute $L = \int_0^z f(x)\, dx$ from one of the methods in the previous paragraph.

$\left(\text{If } k > 15 \text{ \underline{and} } z < k, \text{ use } L = \sum_{i=1}^{\infty} \dfrac{e^{-z} z^{k+i}}{\underline{|k + i}}\right)$

3. COMPUTING z (L specified). To the above, add equations (6), (7) and (13).

Enter data



Subsequent portion of method assumes $L > \frac{1}{2}$.

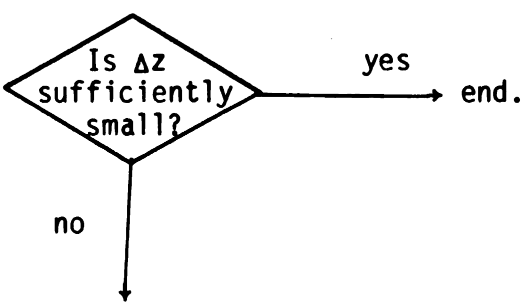
Assign $z_0 = k + \sqrt{k}$

Label [1]

Apply method of paragraph 2 above to compute $L_0 = \int_0^{z_0} f(x)\,dx$

Compute $f'(z_0) = \left(\frac{k}{z_0} - 1\right) f(z_0)$ and

$$\Delta z \approx \frac{z_i}{z_{i-k}} - \sqrt{\left(\frac{z_i}{z_i - k}\right)^2 + \frac{2(L - L_i)}{f'(z_i)}}$$



Assign $z_{i+1} = z_i + \Delta z$

Return to Label [1].

4. COMPUTING *BEST ESTIMATE OF THE FAILURE RATE* (L specified.)

Enter data

$k = o?$ — yes → Employ paragraph 3, above.

no ↓

Assign $s_o = k + \sqrt{k}$

Label [2]

Compute $\int_o^{s_i} f(x)\, dx$ by method of paragraph 2 above.

Compute $f'(s_i) = \left(\frac{k}{s_i} - 1\right) f(s_i)$

$k = 1?$ — yes → assign $a_o = f(s_i)$

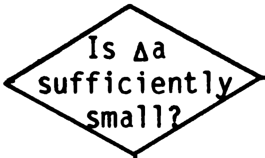
no → assign $a_o = k - \sqrt{k}$

Label [3]

Compute $f(a_j)$ and $f'(a_j)$

Compute $\Delta a = \dfrac{f(s_i) - f(a_j)}{f'(a_j)}$

Compute and store $a_{j+1} = a_j + \Delta a$



no → Return to Label [3]

yes ↓

Compute $\int_0^{a_i} f(x)\,dx$ by methods of paragraph 2 above.

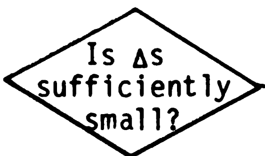
$\left(\text{If } k > 15,\ \int_0^{a} f(x)\,dx = \sum_{i=1}^{\infty} \frac{e^{-a}\,a^{k+i}}{(k+i)!}\right)$

The needed increment of area is

$$A = L - \int_0^{s_i} f(x)\,dx + \int_0^{a_i} f(x)\,dx$$

The approximating trapezoid yields (momentarily dropping subscripts for convenience)

$$\Delta s = \frac{s}{k - s}\left\{1 - \sqrt{1 + \frac{2(k - a)\,(k - s)}{k(s - a)} \cdot \frac{A}{f(s)}}\right\}$$



yes → end.

no ↓

$s_{i+1} = s_i + \Delta s$

Return to Label [2].

BIBLIOGRAPHY

[1] Abramowitz, M. and Stegun, I. A., Eds., HANDBOOK OF MATHEMATICAL FUNCTIONS; Dover Publications, Inc., New York, 1972.

[2] Boole, G., THE CALCULUS OF FINITE DIFFERENCES; 5th Ed., Chelsea Publishing Co., New York, 1970.

[3] Olver, F. W. J., ASYMPTOTICS AND SPECIAL FUNCTIONS; Academic Press, New York, 1974.

[4] Rankin, D. W., ESTIMATING RELIABILITY FROM SMALL SAMPLES; in Proc. 22nd Conf. on D.O.E. in Army R.D.&T., Department of Defense, 1977.

[5] Whittaker, E. T. and Watson G. N., A COURSE OF MODERN ANALYSIS; 4th Ed., Cambridge University Press, 1963.

An Efficient Method for Determining the "A" and "B" Design Allowables

DONALD NEAL AND LUCIANO SPIRIDGLIOZZI
Army Materials and Mechanics Research Center
Watertown, Massachusetts, USA

ABSTRACT

Suggested statistical procedures for obtaining material "A" and "B" allowables from both complete and censored samples are outlined in this paper. The allowables represent a value determined from a specified probability of survival with a 95% confidence in the assertion. The survival probabilities are .99 for the "A" allowables and .90 for the "B" allowables. Both parametric and non-parametric statistical models are evaluated with respect to their desirability in obtaining the allowables. Exploratory data analysis procedures are introduced in order to determine acceptable distribution functions for representing the data in addition to recognizing outliers (bad data) or multi-modality. It is demonstrated from a variety of materials test data that allowable determinations require prior application of exploratory data analysis procedures in order to assure acceptable results. The analysis also provides a process for recognizing either poor testing procedures or inferior material processing.

The two parameter Weibull, normal, lognormal distribution functions are the proposed statistical models for computing the allowables (when non-parametric methods are not applicable). They will usually provide an acceptable range of possible allowable values. The Informative Quantile Function is applied to the test data in order to select the function that best represents the data. In determining the allowables, the desirability of the Weibull function application is shown when limited number of probability ranked data values are available in the primary region (lower ranked numbers) of interest. The required conservatism in this region is satisfied while also satisfying criteria for acceptability of the data representation. The existence of multi-modality or gross outliers in the data set, will in some instances introduce excessively conservative estimates of the allowables when the Weibull function is applied. If the multi-modality case is a reality then a suggested procedures using the Penalized Likelihood Method is used in conjunction with Cramer Rao lower bound estimate for the 95% confidence values.

Extensive tables for samples sizes (5(1)100) have been developed for computing the allowables using the Weibull function. Use of the Monte Carlo Method in conjunction with maximum likelihood (ML) relations describes the procedure used in obtaining the necessary values in determining the allowables. A simple computer code has been made available so that ML estimate of Weibull parameter can be determined thereby resulting in direct computation of the allowables.

In order to demonstrate the desirability of the method, allowables have been determined for Kevlar, Graphite, and Glass composite materials subjected to shear, tensile, and compressive loads. Most of the test data was obtained from the MIL-HDBK-17 (USA Army Materials and Mechanics Research Center) project for composite material applications in aircraft structures.

Introduction

The work described in this report is part of a continuing effort to provide statistical procedures for determining the "A" and "B" allowables for the MIL-HDBK-17 (Handbook for Composite Material in Aircraft Application). The preparation of this handbook is a prime mission of the Army Materials and Mechanics Research Center, Reliability Mechanics and Standardization Division. The current statistical procedures used in the MIL-HDBK-5 is not considered applicable in the determination of the allowables for composite materials.

The selection of an adequate statistical model (parametric or non-parametric) for representing material strength data can result in either a conjectural approach or a costly test program. Often times the normal distribution will be selected since the allowables can be readily determined from tables. Most conventional tests for determining model acceptability will rarely reject the normal function. Unfortunately, selection of the normal distribution can result in erroneous allowable estimates due to extrapolation beyond the lower ordered test results. A quote from Hahn and Shapito [1] which says, "Although many models might appear appropriate within the range of the data, there might well be in error in range for which predictions are desired," adequately summarizes this important issue. This complicates the issue with respect to allowable computations, therefore possibly requiring an extreme value distribution representation which will compensate for the uncertainities in the lower tail region.

The non-parametric method is not realistic, for example, if an "A" allowable were needed for a specific material, the 300 tests would be required. This could result in an extremely costly test program, in that, control of environment and the manufacturing process of composite material must be precise.

In order to address the allowable computation problem more rigorously, the authors have examined the relative merits of 4 distinct distribution functions including the Penalized Likelihood Method [2] for multi-modal case. In most instances the 2 parameter Weibull function is recommended. This extreme value function will usually provide acceptable estimates of the allowables. Either precise, or slightly conservative estimates will be obtained. According the Freudenthal and Gumbel [3], the use of the Weibull distribution to represent the distribution of the breaking strength of materials has been justified by using extreme value theory. In order to recognize the most desirable function, the authors examined the Root Mean Sequence error (function vs. ranked test data) in addition to the Informative Quantile (IQ) Function [4] plots of the data. The IQ results provided an excellent description of test data in terms of a specified distribution function.

In applying the Weibull distribution it is important to recognize data with outliers in vicinity of higher ordered values in addition to multi-modal behavior. Data contaminated in this manner will usually reduce accuracy in the Weibull allowable computation. The Quantile Box Plot [5] was used in determining outliers and multi-modality. This method proved to be more reliable then the conventional robust procedures [6, 7, 8,] currently being suggested for determining outliers. If there is not a rationale for removing outliers then they should remain in the data set, otherwise erroneous estimates of the allowables will result. In the multi-modal case, careful examination of test procedures and material processing should be made prior to acceptance of this phenomena. It is possible that in testing certain composite materials, bimodal behavior could occur. In exploring some recent test results from Kevlar, Graphite and Glass composites in addition to ceramic materials resulted in occasional bimodality behavior. In most cases errors in testing or materials conditioning and processing have accounted for this situation.

The following robust method for applying exploratory procedures in the examination of outliers was used primary as verification of the Quantile Box Plot results. The methods singular advantage is that visual inspection is not necessary in recognition of the outliers. The disadvantage results from arbitrariness in selection of scale and the tuning constant. In some instances where a large amount of skewness or a small data sets exists, then the Quantile Box Plot will be dispersion.

Robustness Method

The outliers are determined in a formal manner by applying a robust method involving application of the ML estimation where the residuals are weighted in a systematic manner. The computed weights describe the relative importance of the data points. For example, a zero weight should indicate exclusion of a point. The removal of outliers (bad data) will essentially define robust data. The robust procedures applied in this paper involves using both the M-estimating technique of Huber [6] and Andrews [7]. Initially the Huber technique is applied in order to determine a robust location parameter (weighted mean). The Andrew's function is then applied using location parameter estimated from the Huber result. It should be noted that this robust method requires a uni-modal distribution of the data, therefore initial application of the Quantile Box Plot should be made inorder to establish unimodality.

The Huber m-estimation technique which involves defining the likelihood funtion

$$L(\theta) = \prod_{i=1}^{N} f(X_i-\theta), \quad -\infty < \theta < \infty \qquad (1)$$

where f is a contaminated normal distribution,

X_i = data,
θ = location parameter and
N = sample size

by maximizing log L (θ) such that

$$\Sigma\psi(X_i-\theta) = 0, \qquad (2)$$

where $\psi = f'/f$

then the solution of (2) is ML estimates of θ designated as $\hat{\theta}$. In order to represent ψ in scale invariant form, equation (2) can be rewritten as

$$\sum \psi\left(\frac{(X_i - \theta)}{d}\right) = 0$$

(3)

with d equal to the estimate of scale. The scale is often defined as

$d = \text{median} \mid X_i - \text{median}(X_i) \mid /.6745$
or simple M.A.D./.6745 (4)

This estimate is considerable more robust than using the complete samples which could result in poor respresentation of the actual scale.

By solving

$$\sum_{i=1}^{N} W_i(X_i - \theta) = 0$$

(5)

where
$$W_i = \psi\left(\frac{X_i - \theta}{d}\right) \Big/ \left(\frac{X_i - \theta}{d}\right)$$

$$\psi = \begin{cases} r & |r| \le c_1 \\ c_1 \text{ sign } (r) & |r| > c_1 \end{cases} ,$$

c_1 is defined as the tuning constant and

$$r = \left(\frac{X_i - \dot{\theta}}{d}\right)$$

An iterative process is then used in the solution of (5) such that when the differences in W_i become negligible therefore providing the necessary criteria for an acceptable solution for the θ and W_i values. For c_1 = 1.345 the Huber's ψ function provides a 95% efficiency.

With estimate of $\hat{\theta}$ determined from the solution of (2) the iteration is continued where the ψ function is now defined as

$$\psi(r) = \begin{cases} c_1 \sin(r/c_1), & |r| \le \Pi c_1 \\ 0, & |r| > \Pi c_1 \end{cases}$$

(6)

This new function is called the Andrew's wave equation. In order to obtain the desired robust data for this ψ function, the tuning was adjusted to c_1 = 1.345 and the scale defined as in equation (4).

It should be noted that Andrew's function was selected for its ability to describe outliers as data with essentially zero weights.

Quantile Box Plot

A general description of the Quantile Box Plot is shown in Figure 1. Where the quantile function is defined as

$$Q\ (u) = F^{-1}\ (u),\ 0 \leq u \leq 1 \qquad (7)$$

that is, if the random variable x with distribution function given by F(x), then the root of F(x) = u, 0 ≤ u ≤ 1 is the p^{th} quantile of F (x). From the ordered statistic $x_1 \leq x_2 \leq . . . , x_n$, Q is defined as piece-wise linear function with interval (0,1) divided into 2n subintervals. Therefore representing Q as

$$Q\left(\frac{2j-1}{2n}\right) = x_j\ ,\quad j = 1, 2, \ldots, n. \qquad (8)$$

In order in interpolate

$$u \in \left(\frac{2j-1}{2n},\ \frac{2j+1}{2n}\right)$$

$$Q(u) = n\ \left(u - \frac{2j-1}{2n}\right)\ x_{j+1} + n\ \left(\frac{2j+1}{2n} - u\ \right) x_j, \qquad (9)$$

where n equals the sample size.

The box boundaries are defined as

Q (.25) to Q (.75)
Q (.125) to Q (.875)
Q (.0625) to Q (.9375)

The Quantile function Q (u) is useful for detesting the presence of outliers, modes and the existince of two populations. Flat slots in Q(u) indicate modes. Sharp rises in Q(u) for u near 0 or 1 suggest outliers; sharp rises in Q(u) within the boxes indicate the existence of two (or more) populations. The obvious bimodality shown in figure 2 is represented by the Quantile Box Plot displayed in figure 3. In figure 4 (lower ordered value) the gross outlier is suggested by the extended vertical line at lower left region of graph in figure 5. The results shown in figures (2, 3, 4, 5) are not representative of typical data sets. In many instance multi-modality and outliers are not obvious from routine examination of the data.

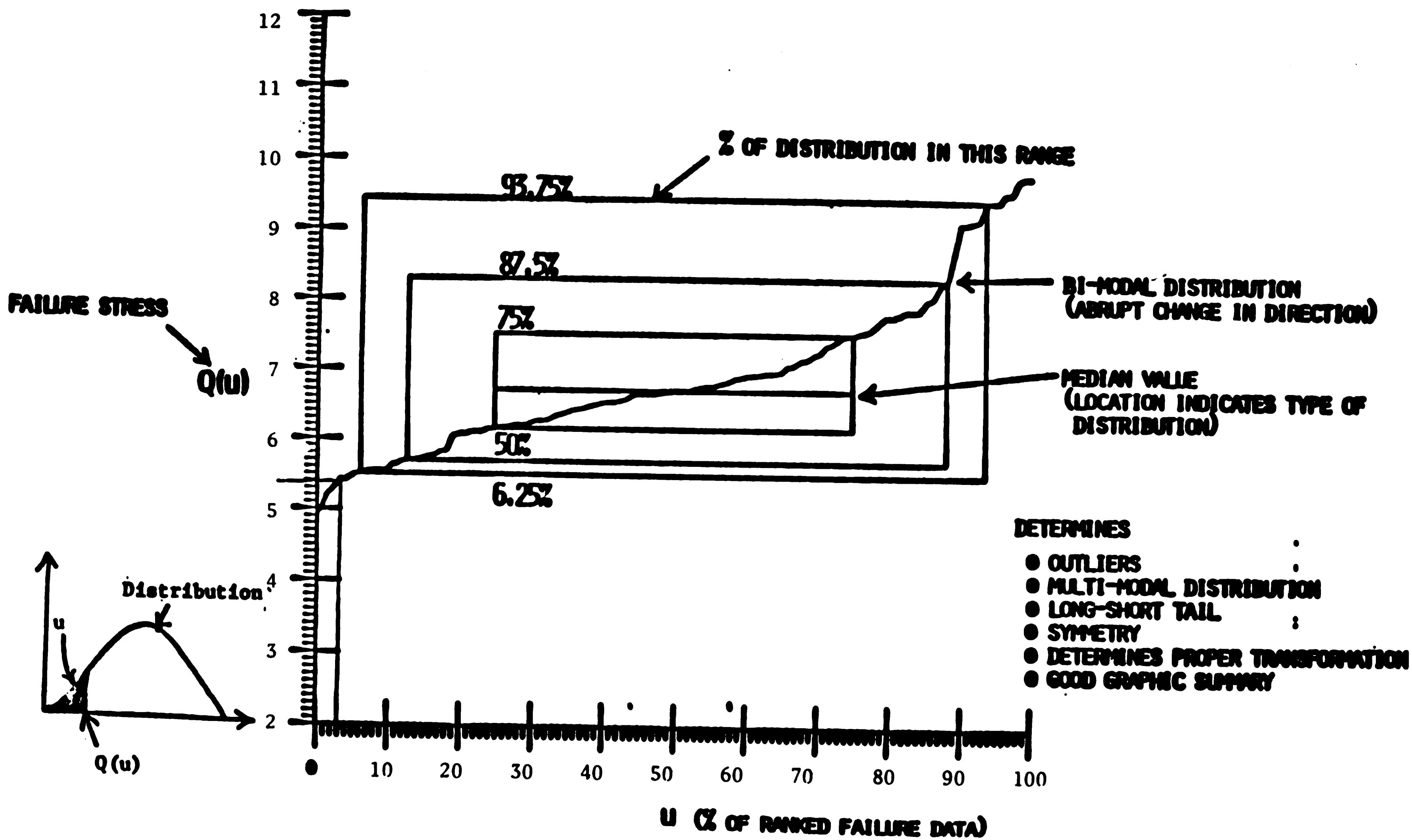


Figure 1 Quantile Box Plot

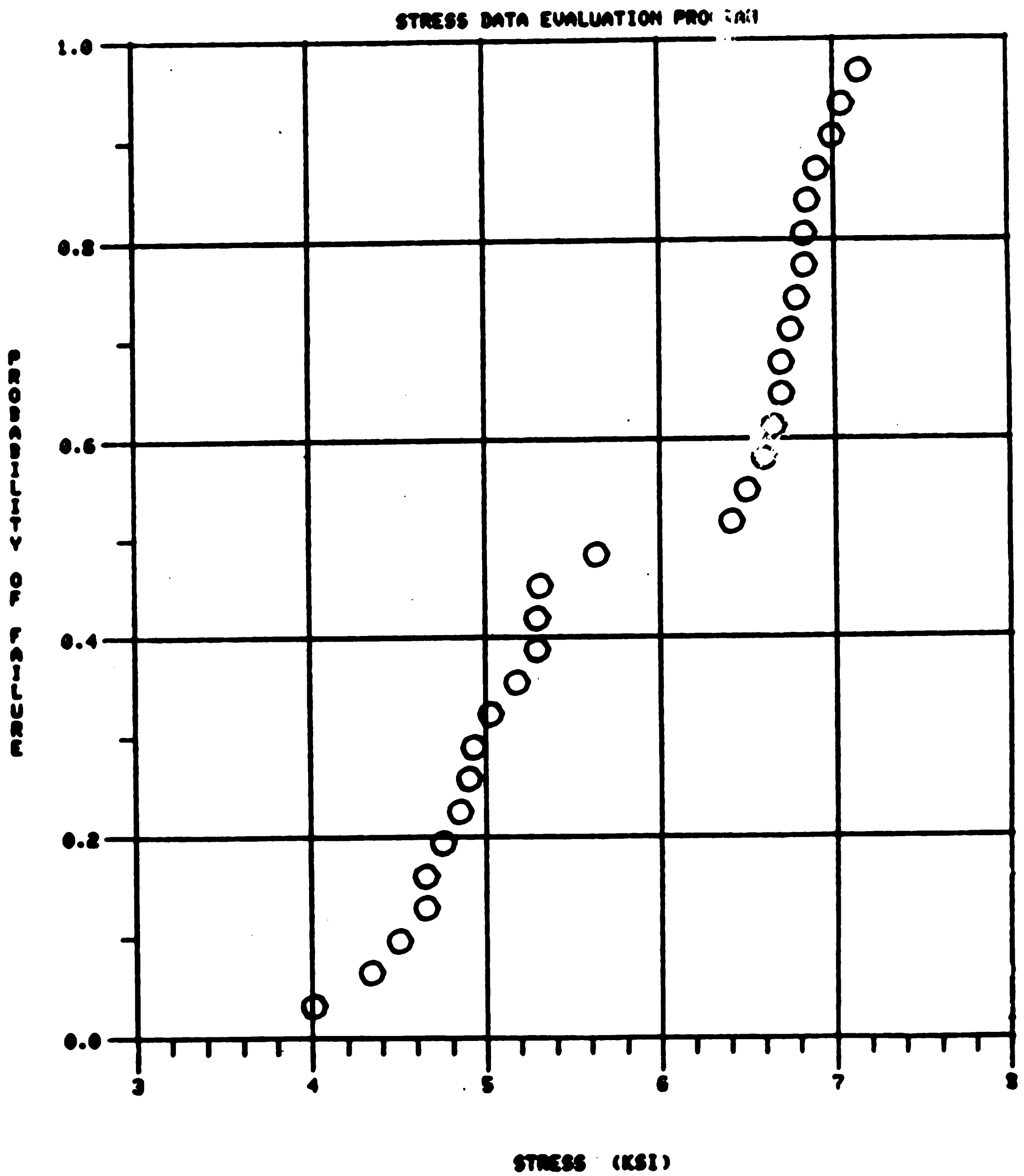


Figure 2 Probability of Failure vs Strength

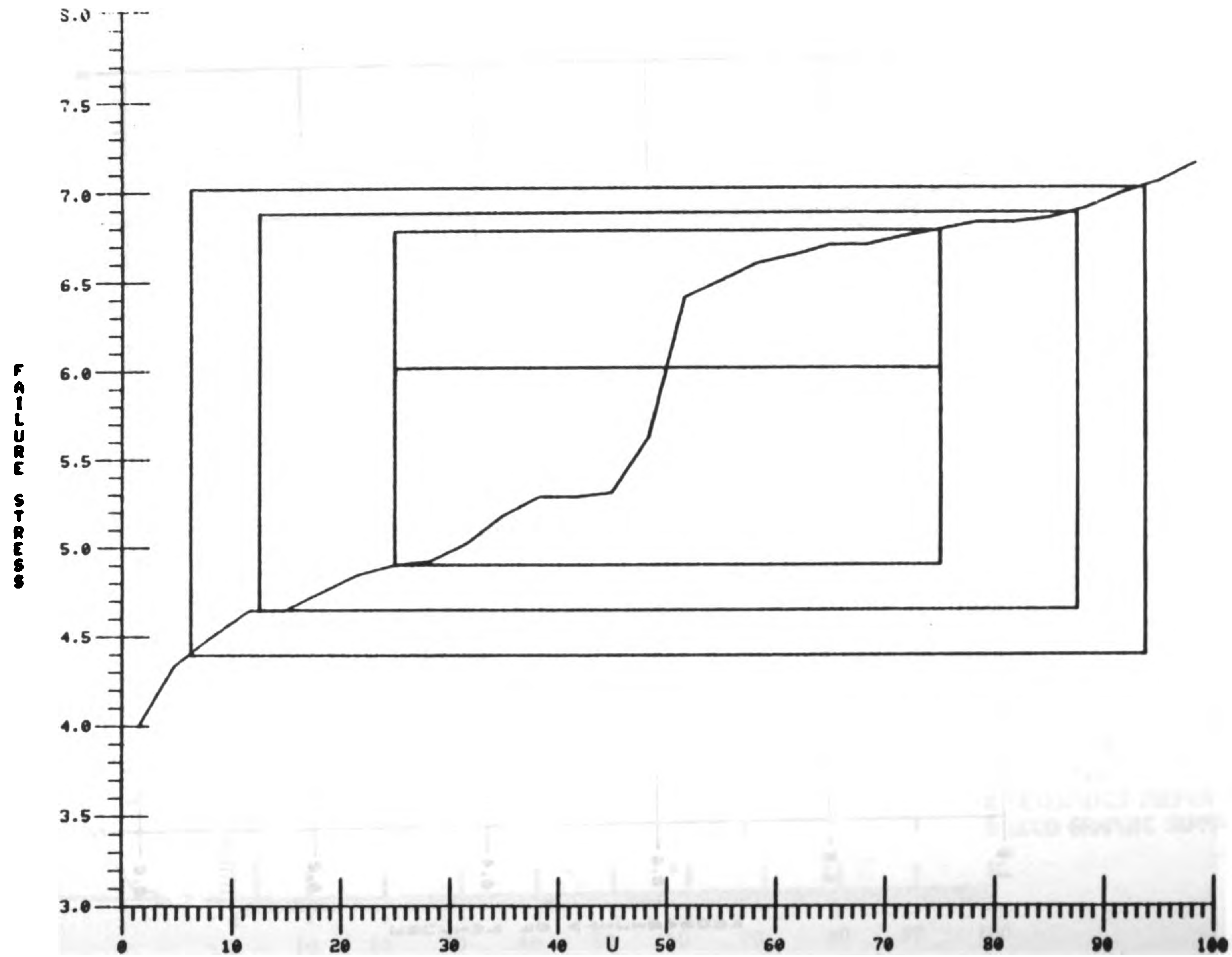
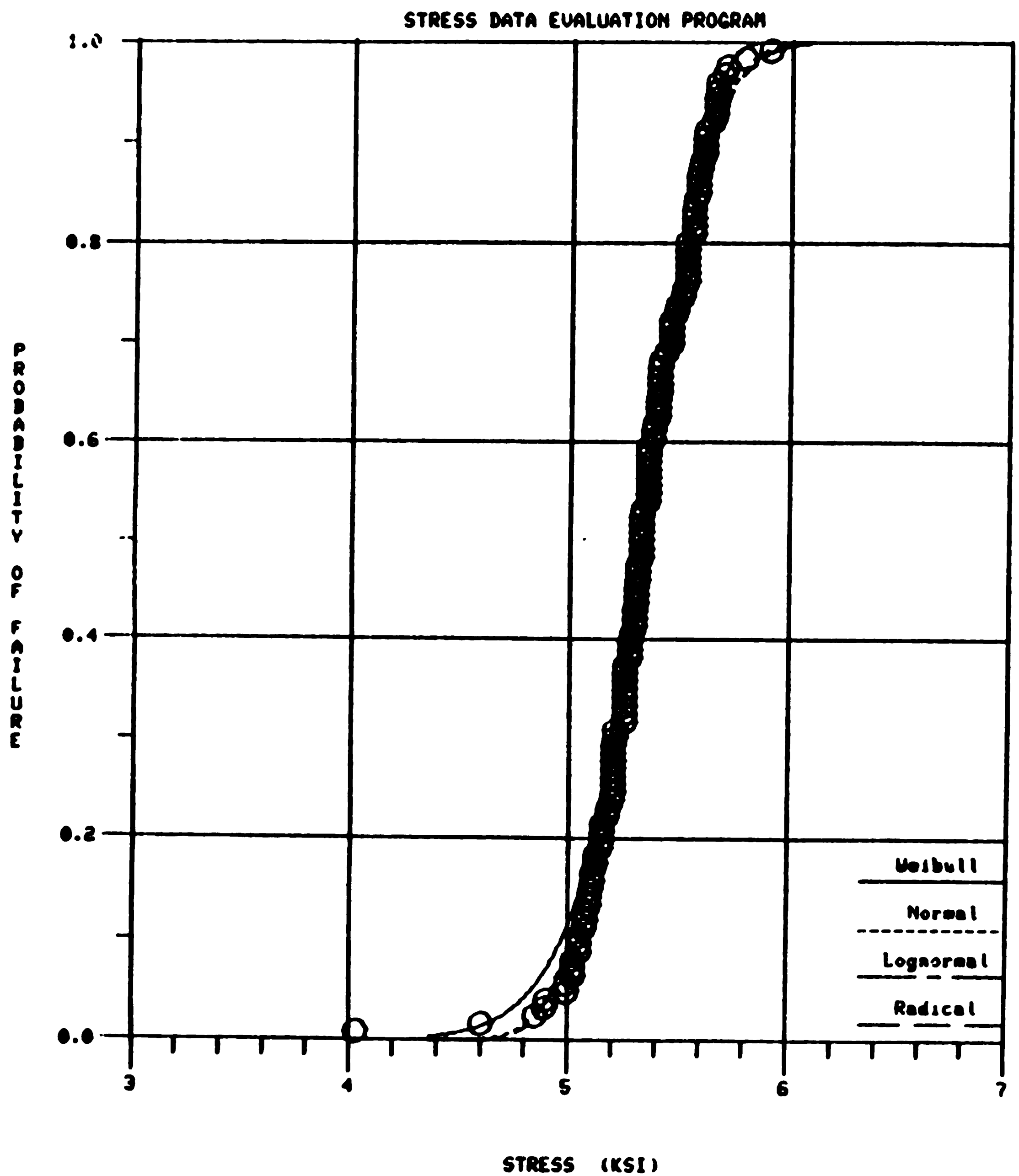


Figure 3 Quantile Box Plot



RMS ERROR

	Normal	Lognormal	Weibull	Radical
R1	.0308	.0329	.0393	.0318
R2	.0302	.0323	.0389	.0318
R3	.0293	.0315	.0382	.0317

		90% CONFIDENCE INTERVAL	
MEAN (KSI)=	5.320	5.283	5.357
S DEV (KSI)=	.242	.219	.270

WEIBULL PARAMETERS

		90% CONFIDENCE INTERVAL	
SLOPE M=	25.680	22.887	28.909
CHAR VALUE=	5.426	5.393	5.459
99% ORIGIN=	4.544	4.427	(95% CL)
ORIGIN=	.000		

RADICAL PARAMETERS

A=	4.212	EXP B (N)=	12.000
B=	2.426	EXP(R)=	12.000
C=	-1.226		
SIGI=	2.985	SIGF=	6.637

NON-PARAMETRI SOLN: A AND B ALLOWABLES

OPTIMAL:A(99,95,N=300)B(90,95,N= 30)
GIVEN DATA:
A(99,72,N=125)B(90,88,N=125)

WEIBULL:

DESIGN A= 4.427
DESIGN B= 4.901

NORMAL:

DESIGN A= 4.681
DESIGN B= 4.958

LOGNORMAL:

DESIGN A= 4.692
DESIGN B= 4.952

Figure 4 Statistical Data Evaluation of Kelvar Material (Tension Test)

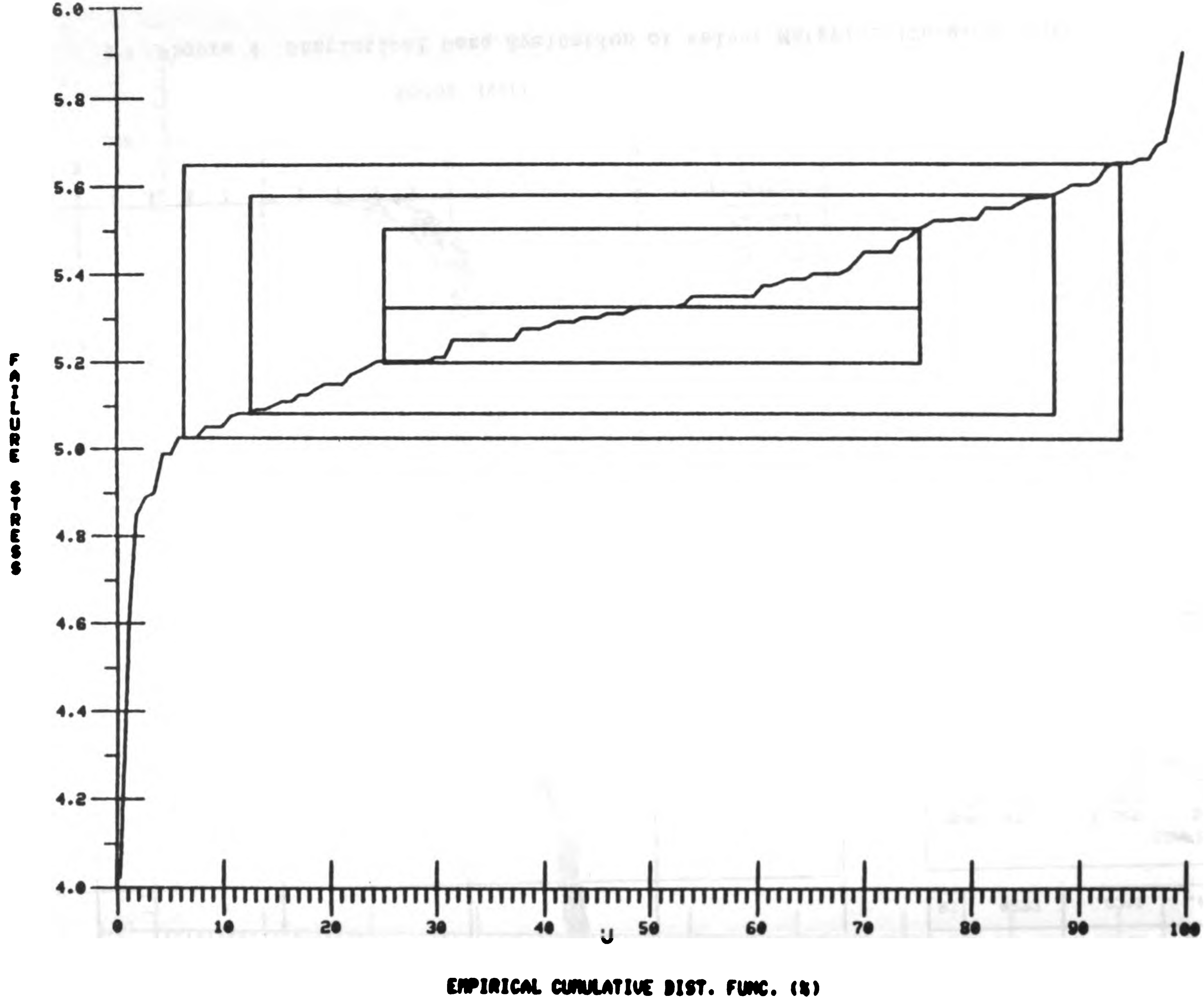


Figure 5 Quantile Box Plot - Kevlar Material

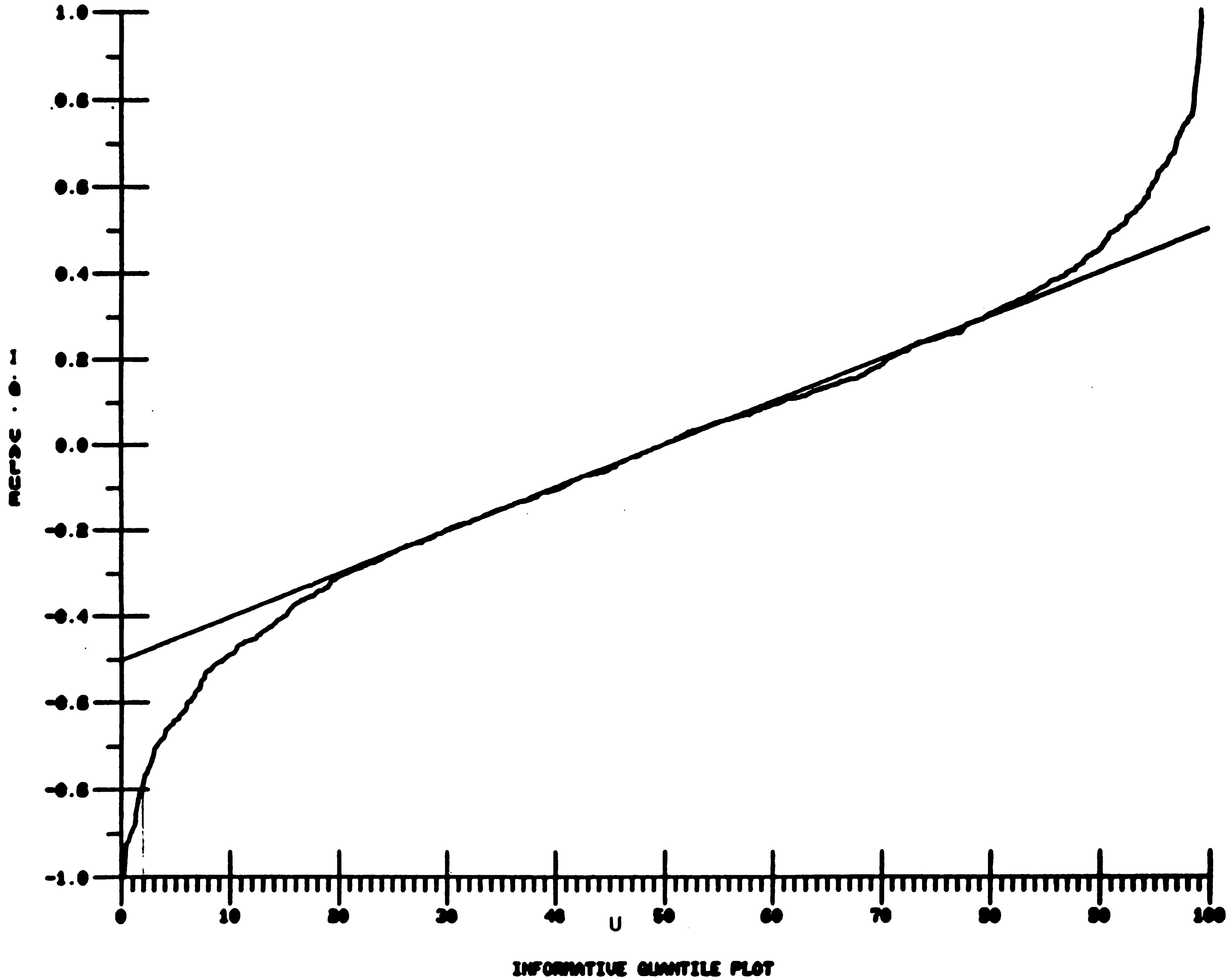


Figure 6 Normal Distribution

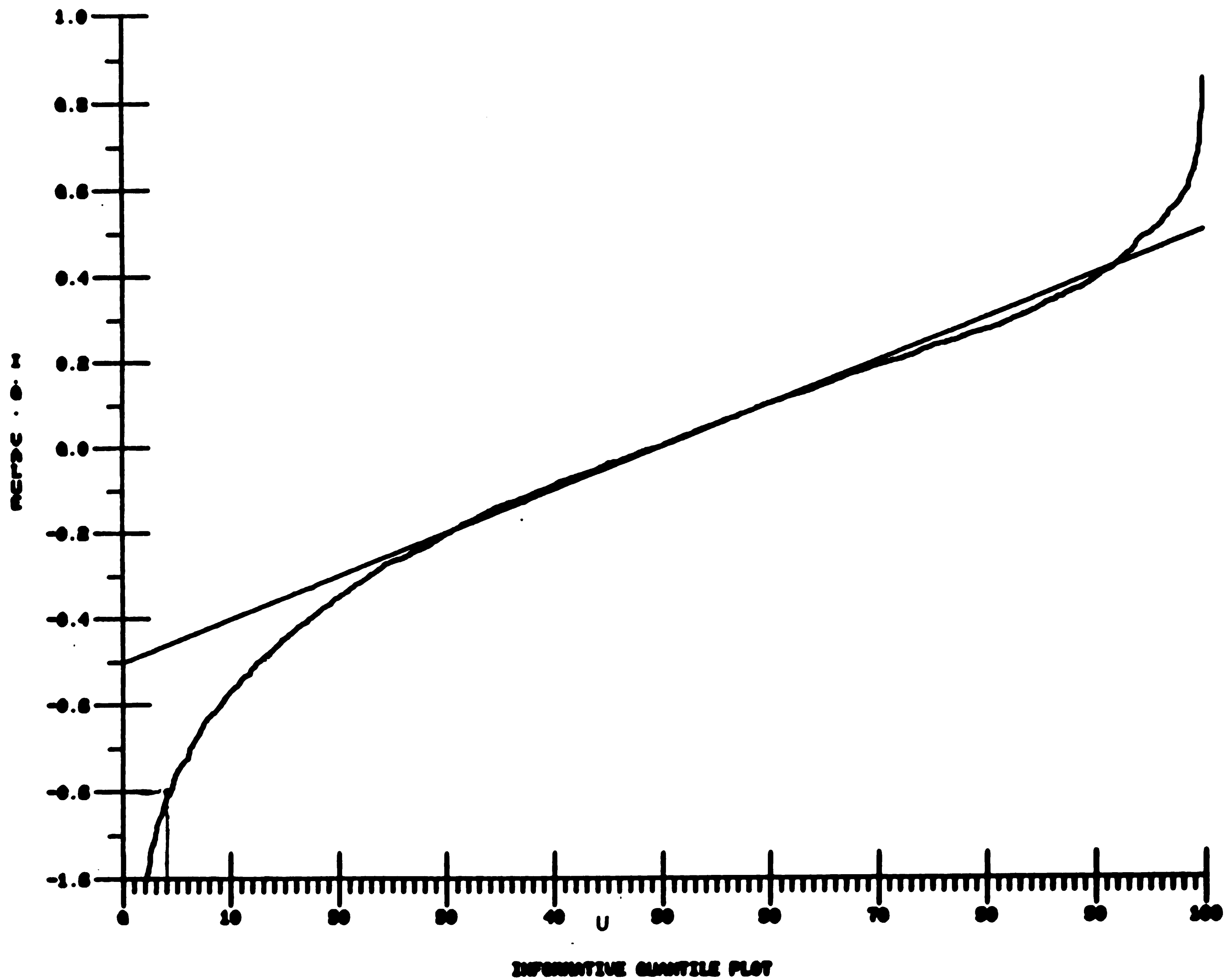


Figure 7 Weibull Distribution

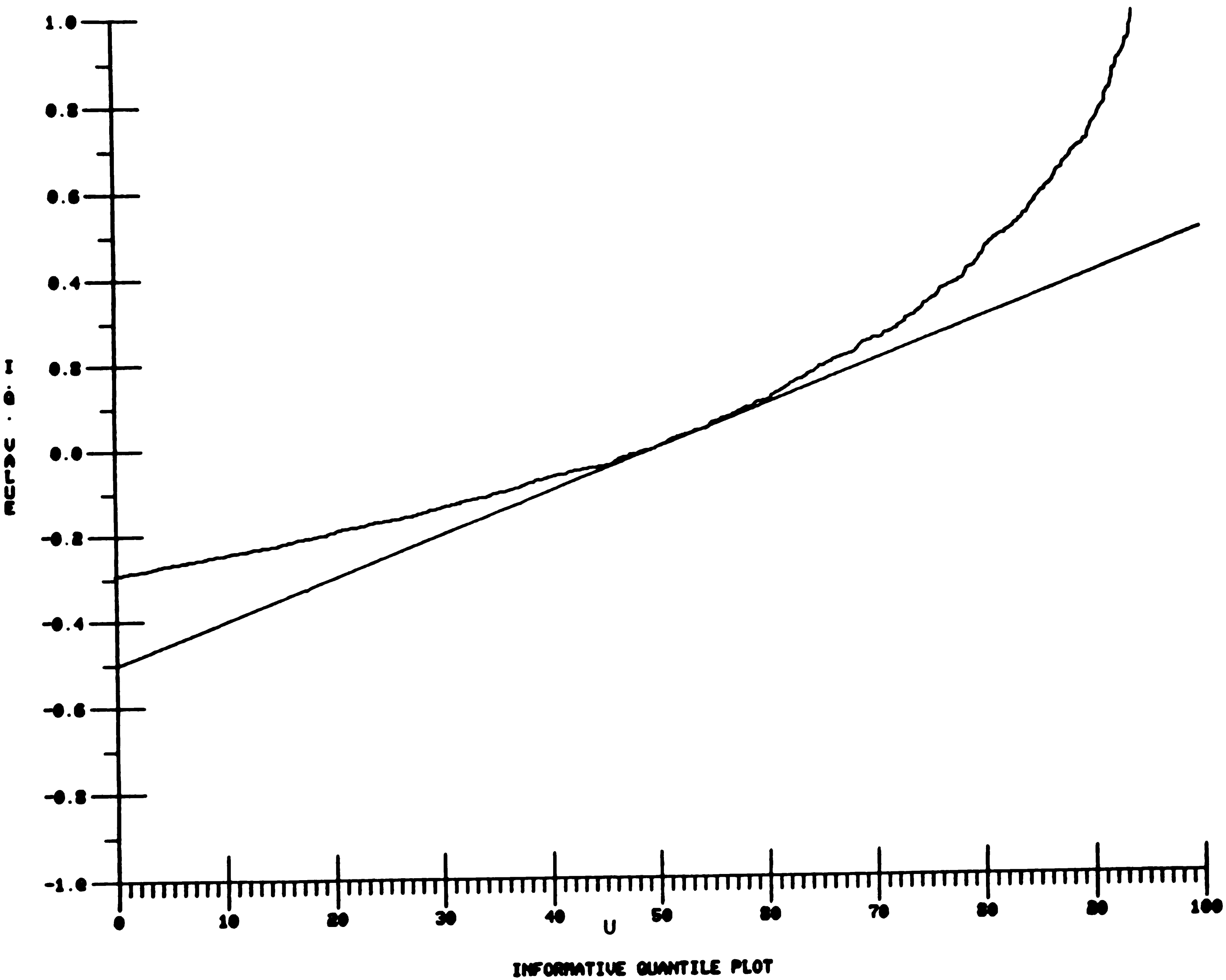


Figure 8 Exponential Distribution

Informative Quantile (IQ) Function

The objective of the IQ function [4] is to identify familar distribution such as normal or Weibull to which the statistical ranked data belongs. This method provides an accurate and simple approach to the problem of identifying which function should represent the data. With the identification completed then more elaborate procedures are recommended inorder to provide verification of the assumed model. If one of the conventional distribution cannot be established as an acceptable model then the Maximum Penelized Likelihood (MPL) approach is suggested. The details of MPL will be discussed later in the text. Applications of the IQ method to a random selection of small sample sizes (N < 50) from large sample of 300 has resulted in an accurate identification of the parent population distribution. Similar results were also obtained from larger samples of 1000 with samples of 25 and 50.

The Informative Quantile Function is simply defined as

$$IQ(u) = \frac{Q(u) - Q(.5)}{2\,[Q(.75) - Q(.25)]}, \quad (10)$$

where Q(u) was previously defined in equation (8). An example of IQ vs U plot for normal distribution is shown in figure (6). Note, at U = .02 the corresponding IQ should be approximately -.8. The straight line joined at IQ of -.5 and .5, represents a uniform distribution. This is introduced in order to provide for an easier identification of the unknown distribution. In figure (7) the Weibull distribution is identified, where U = .04 and IQ = -.8. Figure (8) describes the form necessary for identiftying the exponential distribution. By generating a set of IQ's and U's from equation (10) and plotting these according to the figures, identification of the proper models can be made.

Normal Distribution Function

$$F_x(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}\nu^2} \, d\nu$$

$$\text{where } z = \frac{x - m_x}{\sigma_x}$$

and x, m_x and σ_x are strength data, mean value and standard deviation respectively. $F_x(z)$ can be simply and accurately evaluated using the following polynomial representation,

$$F_x(z) = 1 + \frac{1}{2}(1 + d_1 z + d_2 z^2 + d_3 z^3 + d_4 z^4 + d_5 z^5 + d_6 z^6)^{-16} + E(z)$$

where $E(z) < 1.5 \times 10^{-7}$ (11)

and d_1 = .049867847

d_2 = .0211410061

d_3 = .0032776263

d_4 = .0000380036

d_5 = .0000488906

d_6 = .000005383

Lognormal

The lognormal 2 parameter distribution function is evaluated by using Equation (11), where the maximum likelihood method estimates the mean (m_x) and standard deviation (σ_x),

where $m_x = \frac{1}{N} \sum L_n(X_i)$

and $$\hat{\sigma}_x = \left\{ \frac{\sum L_n (X_i)^2 - N\left(\sum L_n (X_i)/n\right)^2}{N} \right\}^{1/2}$$ (12)

X_i = data values, N - sample size

The unbiased estimate of σ_x is $\left(\left(N/(N-1)\right) \sigma_x^2\right)^{1/2}$

By defining $X_1 = L_n X$, then

$$Z = \frac{X_1 - m_x}{\sigma_x}$$

Weibull Distribution Function

The ML method is applied in order to obtain the two parameters of the Weibull function

$$f(x) = \frac{m}{\mu} \left(\frac{x}{\mu}\right)^{m-1} \exp\left[-\left(\frac{x}{\mu}\right)^m\right]$$ (13)

The method requires defining the likelihood function[9]

$$L = N! \prod_{i=1}^{N} \left\{ \frac{m}{\mu} \left(\frac{X_i}{\mu}\right)^{m-1} \exp\left[-\left(\frac{X_i}{\mu}\right)\right]^{m-1} \right\} \qquad (14)$$

where X_i = data,

M, μ = shape and mormalizing parameter and

N = sample size,

By solving the following log likelihood equations

$$\frac{\partial L_n L}{\partial \mu} = 0 \quad \text{and}$$

$$\frac{\partial L_n L}{\partial m} = 0 \qquad (15)$$

determines and $\hat{m}$ and $\hat{\mu}$ values.

Equation (15) must be solved in an iterative manner by using the computer code listed in Appendix A. The unbiased m and μ and their corresponding confidence intervals are obtained from Tables by [10].

It can be shown that:

$$\sigma_A \text{ or } \sigma_B = \hat{\mu}\left(\ln\left(\frac{1}{P_s^*}\right)\right)^{1/\hat{m}} \qquad (16)$$

where σ_A or σ_B = the allowable, depending on P^*_s.

P^*_s = tolerance limit on P^*_s (probability of survival) determined from application of Monte Carlo method. In Tables 1 and 2 the results for A and B allowable and P^*_s computation is tabulated. Parametric determination of three parameter and censored data requires a more elaborate analysis. These procedures will not be outlined in this text. The (^) represents a biased estimate.

Non-parametric method. Non-parametic procedures [11] are usually more desirable than parametric ones, since they provide the exact probabilities. In the parametric case, the reliance is on an assumed distribution function which provides extrapolated results for the probability of survival values. The penalty for applying the non-parametric method is the need for relatively large amounts of data (e.g., 29 values for the "B" allowable and 300 for the "A" allowable). The lowest ranked value describes the corresponding allowable. In the case where 100 data values are available, the sixth lowest ranked data value determines the "B" allowable. The use of 100 values in obtaining the "B" allowables prevents any erroneous estimates if lowest ordered strength values are incorrect. Table 3 shows which ranked data value should be used for a particular sample size. The importance of using sample sizes greater than the required 29 for the "B" allowable is shown in figure 9. For example, if all data is included, the allowable will be "4"; however, this could be erroneous if data value 4 was an outlier. By removing the outlier, the allowable is 4.6.

TABLE 1 P_S* VS. SAMPLE SIZE N FOR "A" ALLOWABLES

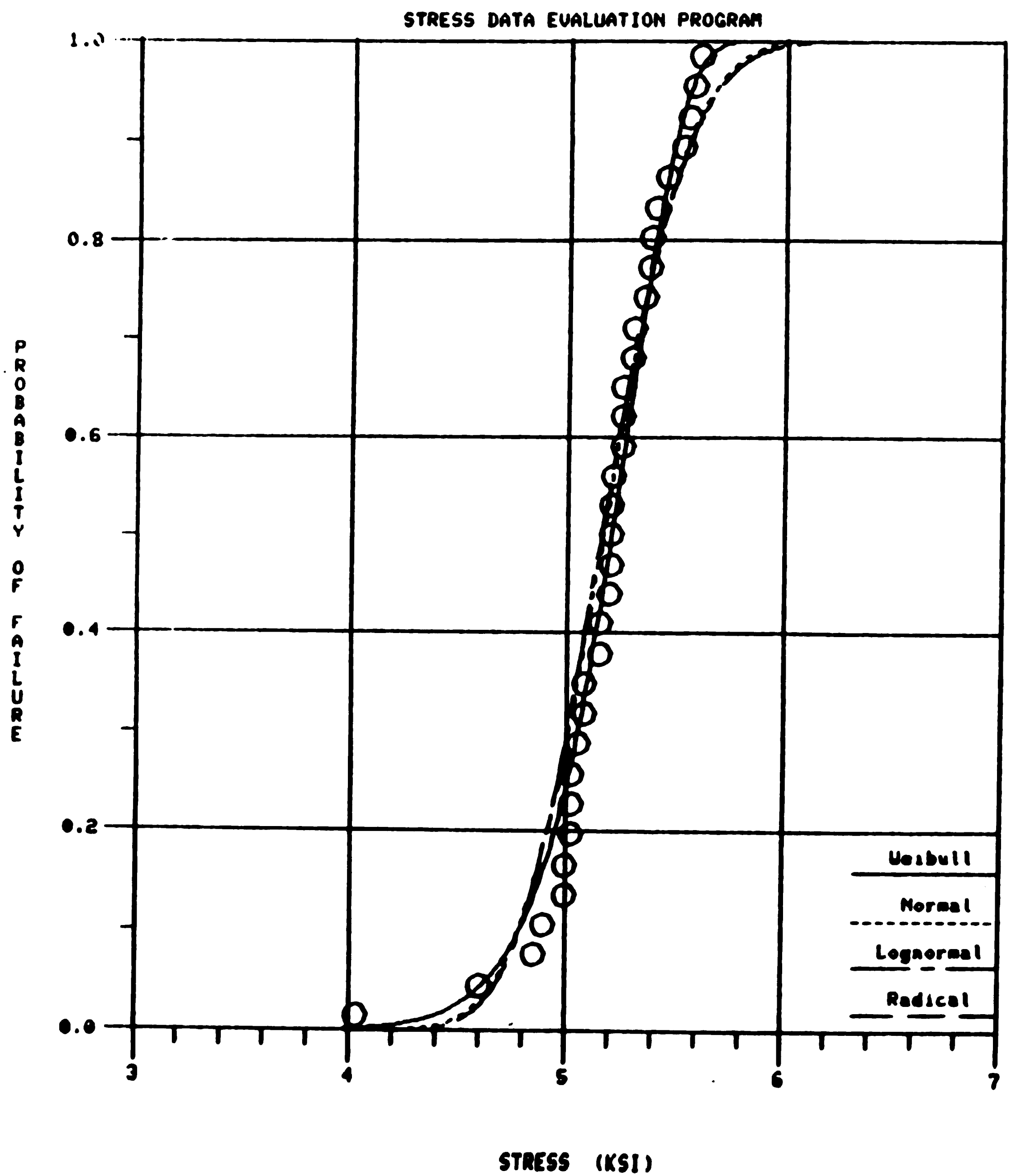
N	P_S*	N	P_S*	N	P_S*	N	P_S*
5	.999999	29	.998214	53	.996816	77	.995970
6	.999992	30	.998130	54	.996776	78	.995940
7	.999972	31	.998048	55	.996736	79	.995911
8	.999930	32	.997968	56	.996697	80	.995883
9	.999859	33	.997891	57	.996657	81	.995855
10	.999780	34	.997816	58	.996619	82	.995827
11	.999713	35	.997743	59	.996580	83	.995800
12	.999650	36	.997673	60	.996543	84	.995773
13	.999579	37	.997606	61	.996505	85	.995747
14	.999500	38	.997541	62	.996468	86	.995721
15	.999420	39	.997479	63	.996432	87	.995695
16	.999340	40	.997420	64	.996396	88	.995670
17	.999256	41	.997363	65	.996360	89	.995645
18	.999160	42	.997309	66	.996325	90	.995620
19	.999050	43	.997257	67	.996290	91	.995596
20	.998940	44	.997207	68	.996256	92	.995572
21	.998843	45	.997159	69	.996222	93	.995548
22	.998760	46	.997113	70	.996189	94	.995525
23	.998684	47	.997068	71	.996156	95	.995502
24	.998613	48	.997025	72	.996124	96	.995479
25	.998540	49	.996982	73	.996092	97	.995456
26	.998463	50	.996940	74	.996061	98	.995434
27	.998382	51	.996898	75	.996030	99	.995412
28	.998298	52	.996857	76	.996000	100	.995390

TABLE 2 P_S* VS. SAMPLE SIZE N FOR "B" ALLOWABLES

5.0	.99880	29.0	.95991	53.0	.94499	77.0	.93884
6.0	.99783	30.0	.95930	54.0	.94468	78.0	.93861
7.0	.99651	31.0	.95861	55.0	.94437	79.0	.93838
8.0	.99450	32.0	.95785	56.0	.94407	80.0	.93815
9.0	.99170	33.0	.95703	57.0	.94377	81.0	.93792
10.0	.98900	34.0	.95617	58.0	.94349	82.0	.93769
11.0	.98716	35.0	.95528	59.0	.94321	83.0	.93747
12.0	.98540	36.0	.95438	60.0	.94293	84.0	.93724
13.0	.98287	37.0	.95349	61.0	.94267	85.0	.93702
14.0	.97992	38.0	.95261	62.0	.94241	86.0	.93680
15.0	.97720	39.0	.95178	63.0	.94215	87.0	.93658
16.0	.97517	40.0	.95100	64.0	.94190	88.0	.93636
17.0	.97358	41.0	.95029	65.0	.94165	89.0	.93614
18.0	.97200	42.0	.94963	66.0	.94140	90.0	.93593
19.0	.97013	43.0	.94904	67.0	.94116	91.0	.93572
20.0	.96820	44.0	.94850	68.0	.94092	92.0	.93552
21.0	.96653	45.0	.94800	69.0	.94069	93.0	.93531
22.0	.96514	46.0	.94755	70.0	.94045	94.0	.93511
23.0	.96400	47.0	.94712	71.0	.94022	95.0	.93492
24.0	.96306	48.0	.94673	72.0	.93999	96.0	.93472
25.0	.96228	49.0	.94636	73.0	.93976	97.0	.93454
26.0	.96161	50.0	.94600	74.0	.93953	98.0	.93435
27.0	.96102	51.0	.94566	75.0	.93930	99.0	.93417
28.0	.96047	52.0	.94532	76.0	.93907	100.0	.93400

TABLE 3. Ranks, r, of observation, n, for an unknown distribution having the probability and confidence of A and B values.

A Basis		B Basis		B Basis		B Basis	
n	r	n	r	n	r	n	r
300	1	29	1	321	24	1269	110
480	2	46	2	345	26	1376	120
630	3	61	3	368	28	1483	130
780	4	76	4	391	30	1590	140
920	5	89	5	413	32	1696	150
1050	6	103	6	436	34	1803	160
1190	7	116	7	459	36	1909	170
1320	8	129	8	481	38	2015	180
1450	9	142	9	504	40	2120	190
1570	10	154	10	560	45	2230	200
1700	11	167	11	615	50	2330	210
1820	12	179	12	671	55	2430	220
1950	13	191	13	726	60	2530	230
2070	14	203	14	781	65	2630	240
2190	15	215	15	836	70	2730	250
2310	16	227	16	890	75	2830	260
2430	17	239	17	945	80	2930	270
2550	18	251	18	999	85	3000	277
2670	19	263	19	1053	90		
2790	20	275	20	1107	95		
2910	21	298	22	1161	100		



RMS ERROR

	Normal	Lognormal	Weibull	Radical
R1	.0627	.0674	.0420	.0499
R2	.0605	.0653	.0404	.0510
R3	.0574	.0624	.0384	.0525

		90% CONFIDENCE INTERVAL	
MEAN (KSI)=	5.166	5.076	5.256
S DEV (KSI)=	.301	.251	.379

WEIBULL PARAMETERS

		90% CONFIDENCE INTERVAL	
SLOPE M=	22.450	17.829	28.341
CHAR VALUE=	5.288	5.217	5.361
99% ORIGIN=	4.345	4.065	(95% CL)
ORIGIN=	.000		

RADICAL PARAMETERS

A=	2.621	EXP B (N)=	12.000
B=	3.504	EXP(R)=	12.000
C=	-.750		
SIGI=	1.871	SIGF=	6.125

NON-PARAMETRI SOLN: A AND B ALLOWABLES

OPTIMAL:A(99,95,N=300)B(90,95,N= 30)

GIVEN DATA:

A(99,28,N= 33)B(90,97,N= 33)

WEIBULL:

DESIGN A= 4.065
DESIGN B= 4.627

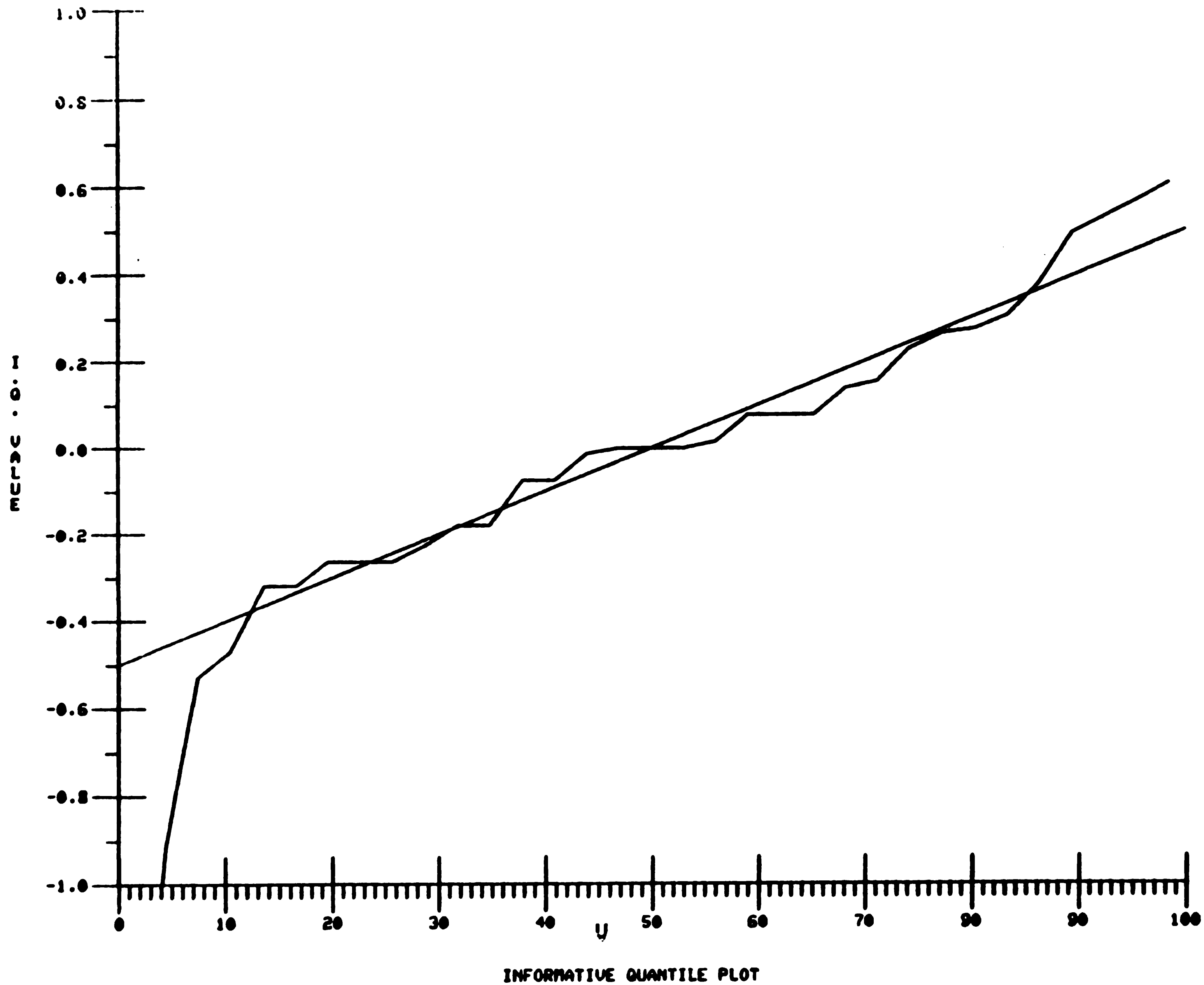
NORMAL:

DESIGN A= 4.257
DESIGN B= 4.640

LOGNORMAL:

DESIGN A= 4.280
DESIGN B= 4.629

Figure 9 Statistical Evaluation of Kevlar Material (Hexcel)



1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
I.Q. VALUE
0
10
20
30
40
50
60
70
80
90
100
U
INFORMATIVE QUANTILE PLOT

Figure 9a

Multi-Modality. If the bimodality displayed in figures 1, 2, and 3 is a reality, not the result of processing or testing errors then the current available parametric procedures usually will not provide acceptable representation of the data. A non-parametric method [2] is suggested having excellent approximation properties, for estimating an unknown probability density function from a random sample $X_1, \ldots, X_N$

The estimator optimizes a criterion function which combines the maximum likelihood principle and a penalty term for smooth (i.e., not bumpy) behavior. The criterion function is a discrete approximation to

$$L(f) = \prod_{i=1}^{N} x(i) \exp\left[-\alpha \int \left|\frac{d^2 f(t)}{dt^2}\right|^2 dt\right] \quad (17)$$

where f(t) is any probability density function; that is, f(t) is nonnegative and intergrates to 1. Equivalently, we may maximize the Ln[L(f)] which separates into two terms - a log likelihood plus a log penalty term.

The penalty term contains an unknown positive constant α which determines the amount of smoothness in the resulting estimator. Values of α that are "too small" result in bumpy estimates while α "too large" oversmooths. In practice, use an α as small as possible without introducing excessive bumps. Several values of α that differ by factors of ten should be tried and graphically displayed and compared to a histogram or parametric assumption.

Numerical integration determines the cumulative density values (probability of survival) for the prescribe percentile, 90 or 99 depending on the desired allowable.

Confidence limit on these estimates say $R_{.90}$ or $R_{.99}$ can be determine from the Cramer-Rao lower bound which determine the variance on R. The confidence limit L is determined iteratively from

$$L_i = R - U_\gamma\left[V(L_i - 1)\right]^{1/2}, \; i = 2, 3\ldots \quad (18)$$

where U_γ is the γ percentage point (95%) of the normal distribution. Initially,

$$L_1 = R - U_\gamma[V(R)]^{1/2}$$

Where,

$$V(R) = R^2 (\ln R)^2 \left\{1.109 - .514 \ln(-\ln R) + .608 [\ln(-\ln R)]^2\right\}/n \quad (19)$$

Subsequent interation in equation 18 requires substitution of L_i for R in equation 19.

Pooling of Data. In obtaining the allowables, test data should be obtained from a number manufacturers (e.g. composite materials from various aircraft industry representatives). All test data should be pooled inorder to obtain an allowable consistent with a general population of that specific materials strength values. If a significant difference exist among the manufacturer then an investigation should be made regarding the cause of this situation. The tests recommended for determining significant differences are the conventional t or Mann-Whitney Test, and two non-parametric tests for the K - sample case. The Kruskal-Wallis [12] multi-sample test for identical populations is applied such that H is corrected for ties. The null hypothesis (identical populations) is rejected at the 2% level. The other distribution free test, is the Jonckheere's[12] K sample trends test against the ordered alternative. Where ties are removed by applying a randomization process. The null hypothesis of randomness is rejected at the 2% level, that is, acceptance of difference in populations at this level.

Statistical Evaluation of Data (Graphical Display). In Figure 4, a plot of probability failure (P_f) versus failure stress (see marked circles) of empirical failure data were shown. The P_f values were determined from the R_j ranking. The four candidate functions are listed on the graph with their corresponding line form representations. In addition to obtaining a visual inspection of best fit, the RMS error provides a quantitative evaluation for the three rankings and the density functions. The sample evaluation for the three rankings and the density functions. The sample means and standard deviations are tabulated with their corresponding 90% confidence interval. The Weibull shape and normalizing (Char. value) parameters are tabulated with confidence intervals. The 99% origin represents one percent probability of failure with 95% lower confidence limit for that number. This representation is the A allowable. If one wants to increase or decrease origin percentage, it can be done by applying methods described in (2). The word "origin" will equal zero if two parameter Weibull function was considered, otherwise, three parameter Weibull function was used. In the radical parameter tabulation, A, B and C are coefficients obtained from a least squared fit routine. Exp B (N) and Exp C (R) are the corresponding exponents determined in the fitting of the data. Sig I and Sig F are the two cut-off points. That is, the smallest and largest projected values determined by the function.

In applying the robust procedures, it is important to have a rationale for ignoring the determined outlier, otherwise, erroneous estimates of survival probability computations could result. The robust scheme can be applied to data from relatively small size specimens, where errors in machining, testing, etc., greatly effect strength determination. The authors have noted considerable improvement in ceramic material failure predictions of large specimens (13 in^3) from knowledge of small specimen (.03125 in^3) strength results when the outliers are removed from the original data. Non-parametrics solutions, are applied to the code (see Figure 4), can provide information regarding confidence levels for A and B allowables, with respect to the number of data points. The code can be altered to include other allowables by using the simple relationships outlined in (11). In the last three boxes the allowable estimates are tabulated for the Weibull, normal and lognormal functions.

Results and Discussion

In figure 4, the results of tension tests on Kevlar composite material (Hexcel Co.) are shown, they are similar for all three functions. This is an ideal situation, since selecting allowables from any of the functions will provide acceptable results. In figure 9, the results from another manufacturer using Hexcel material are shown. The RMs error indicates the Weibull funtion would be an acceptable representation of data. In figure 9a, a IQ plot of data from figure 9 test results also indicates data should be represented by a Weibull function. The results shown in figure 10, describe the existence of an outlier at highest ordered value. See figure 11 display of Box Plot for verification of outlier value (26.5). It is obvious from figure 10 that Weibull function does not represent lower tail region particularly well. The resultant design "B" allowable of 10.9 determined from Weibull computation differs from non-parameter solution by 2.1 a 19% difference. The "A" allowable result, Weibull vs. normal is different by 2.1. Since the non-parametric result is an accurate measure then the Weibull function has produced an error of approximately 19% for at least the "B" allowable and possible more for the "A" allowable. In figure 12, the outlier has been removed from the data displayed in figure 10. Note the substantial increase in the Weibull shape parameter from 5.66 to 7.22 by removing one outlier (highest ordered value). These results indicate need for exploring data prior to applying functional representation. If the outlier actually exists, then allowable obtained from normal distribution should be considered. Determination of acceptable Weibull parameters from the ML method depends on the absence of outliers at highest ordered values. Since the vulnerability of the ML method has been exposed in the above example, it suggested that the Best Linear Unbiased Estimate procedure also be used in order to determine the Weibull parameters, thereby providing flexibility in the selection of parameter estimating procedures.

Figure 13 describes the statistical results from compression tests on Kevlar material obtained from the three manufacturers. In figure 14, the results from pooling original three manufacturers data, with a forth manufacturer (submitted data at a later date). An approximate 6% reduction in allowable estimate with addition of the forth data set. The are two fundamental issues involved one is the need for a substantial number of manufacturers participating in the allowables computation program and secondly the possible reliance on an extreme value type (Weibull) distribution in order to introduce conservatism when accounting for the uncertainties existing from a limited number of pooled samples in representing the population.

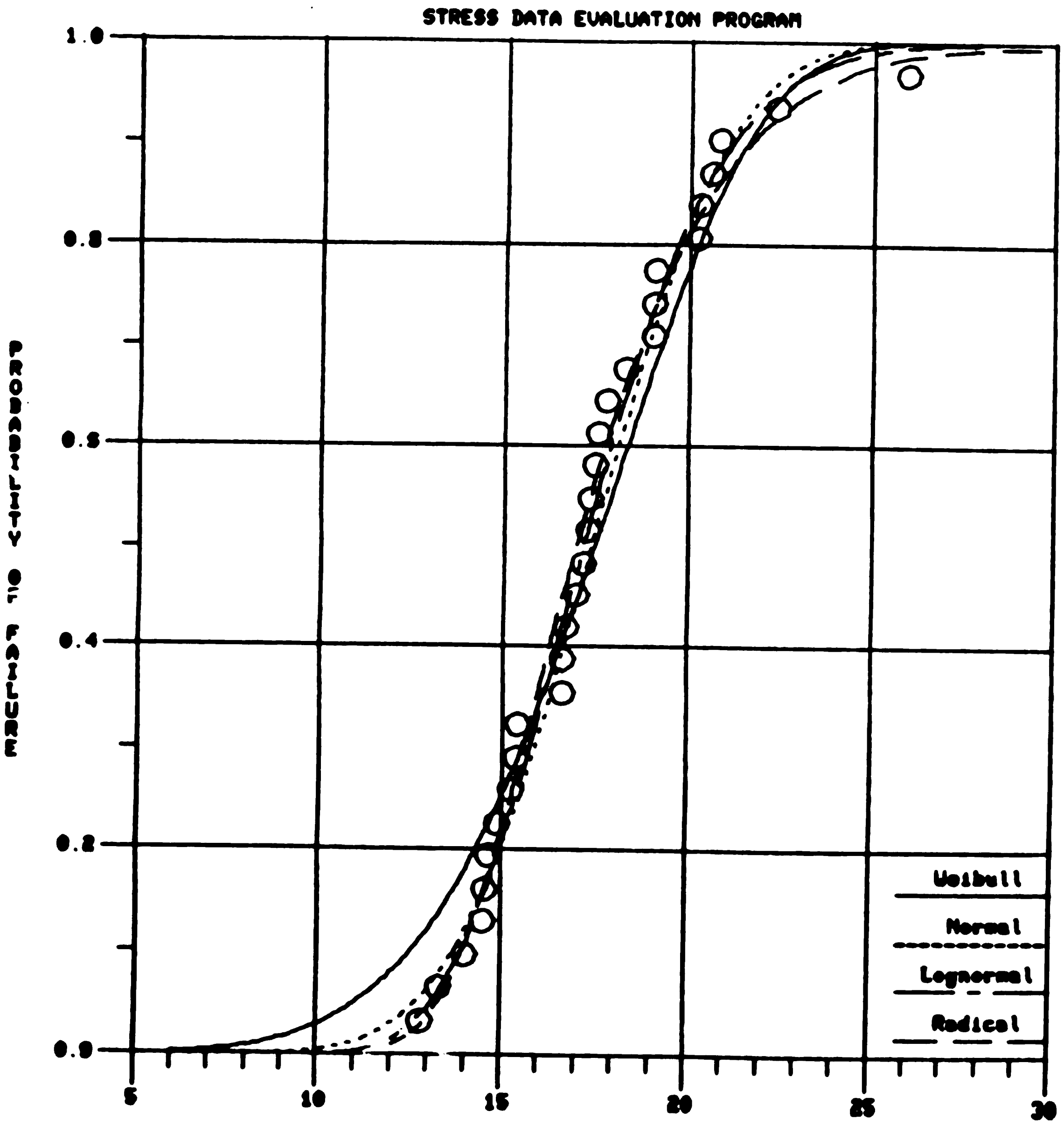
The Kruskal-Wallis test for determining K - sample difference indicated the forth monufacturer data differed significantly from the other samples at .01 levels. At present, the MIL-HDBK-17 recommends pooling all samples unless a rationale has been established for removal of sample.

Even though, homogeneous data is not available for determining allowables, the committee considers it more important to represent the difference among manufacturers, particularily so for composite materials.

In figure 15, the higher ordered values appeared to be from a different mode of failure. The Box Plot example shown in figure 1 is a display of this data which essentially confirms the existence of bimodality. It should be noted in figure 15 that none of the candidate functions adequately represent the ranked data. The allowables are too conservative, particularly for the Weibull and normal representative. A suggested alternative in determining allowables for this data is the Maximum Penalized Likelihood Method[2] described in the text. The results are shown in this figure are an excellent representation of the ranked data including the sizable bump. The allowables of 5.04 ksi for the "A" and 5.51 ksi for the "B". The "B" allowable agrees within .5% of the non-parametric solution. At present, the authors consider this the most acceptable method for determining allowables where multimodality exists in the data. Examination of other methods, such as considering upper mode as censored data or application of the mixed Weibull distributions to the data, proved to be inadequate. The later method may have merit if selected percentiles of the distribution are matched with the corresponding ranked values in a manner that guaranteed a good fit. In figure 16 another example of bimodality is shown. In this case, representation of Weibull shape parameter value of 6.75 is extremely low. The results from MPL method are also shown in the figure including tabulation of the allowables. Manufacturer of this material has recently indicated that lower mode data was incorrectly added to upper mode data due to different autoclaves used in processing of material.

Figure 17 shows the results from Tension test on Kevlar material (Cycom Co.). The allowable computations differ by at most 7% for the three functional representations. Allowables determined from normal computation would be selected for this material and test.

Figure 18, shows data evaluation of Composite graphite material. The existance of bimodality displayed in the sample was a common occurance among several of the graphite test samples.



RMS ERROR

	Normal	Lognormal	Weibull	Radical
R1	.0434	.0329	.0640	.0336
R2	.0424	.0321	.0613	.0331
R3	.0416	.0316	.0575	.0329

		90% CONFIDENCE INTERVAL	
MEAN (KSI)=	17.381	16.473	18.289 [?]
S DEV (KSI)=	2.877	2.416	3.745

WEIBULL PARAMETERS

		90% CONFIDENCE INTERVAL	
SLOPE M=	5.663	4.446	7.231
CHAR VALUE=	18.610	17.596 [?]	19.708 [?]
99% ORIGIN=	6.570	6.454	(95% CL)
ORIGIN=	.000		

RADICAL PARAMETERS

A=	39.478	EXP S (N)=	18.000
B=	11.329	EXP(R)=	18.000
C=	-39.189		
SIGI=	4.289	SIGF=	50.807

NON-PARAMETRI SOLN: A AND B ALLOWABLES
OPTIMAL: A(95.95,N=300)B(90.95,N= 30)
GIVEN DATA:
A(95.95,N= 30)B(95.95,N= 30)

WEIBULL:

DESIGN A= 6.454
DESIGN B= 10.807

NORMAL:

DESIGN A= 8.585 [?]
DESIGN B= 12.289

LOGNORMAL:

DESIGN A= 10.520 [?]
DESIGN B= 12.861 [?]

Figure 10 Statistical Data Evaluation Bend Test Results (Ceramics)

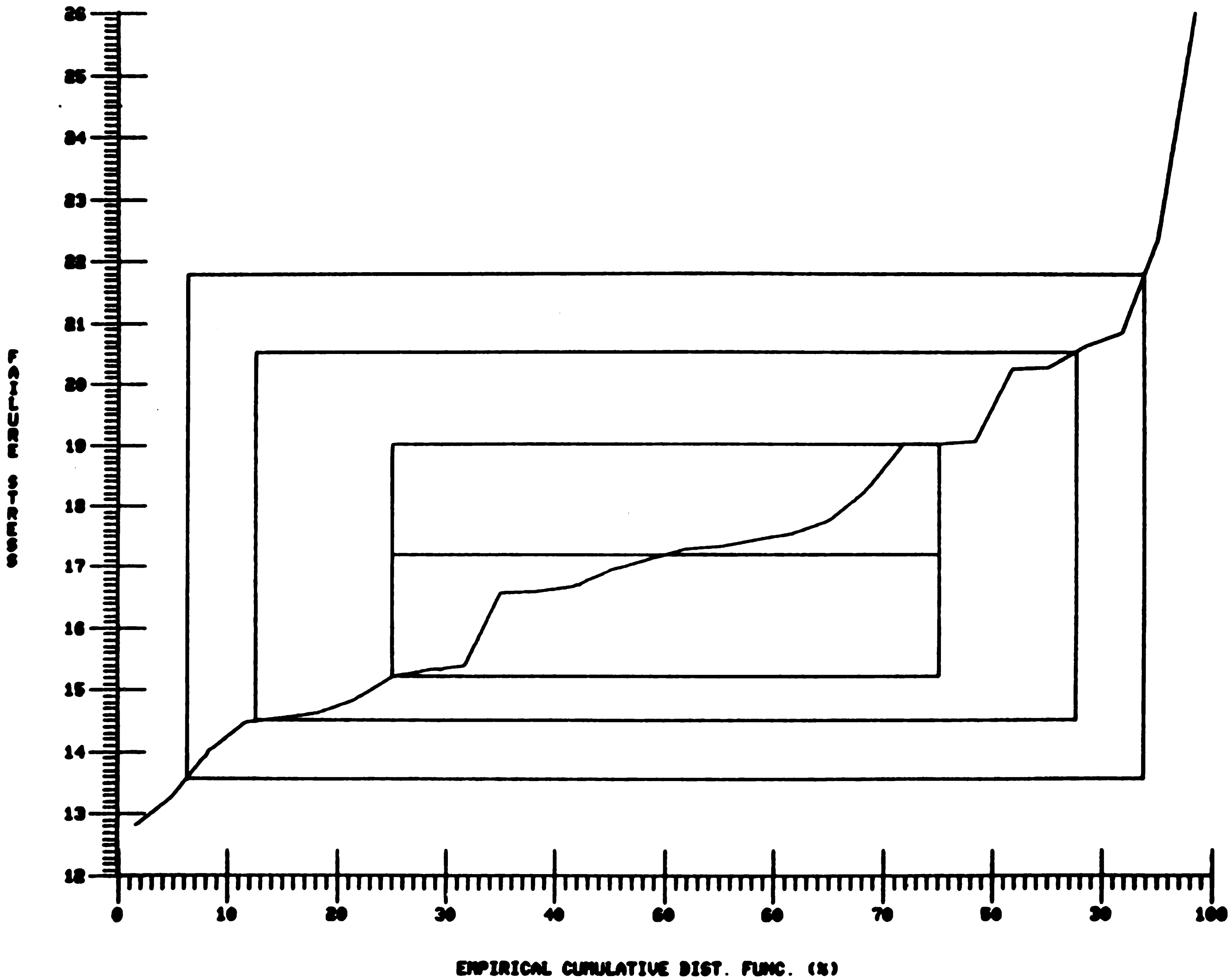
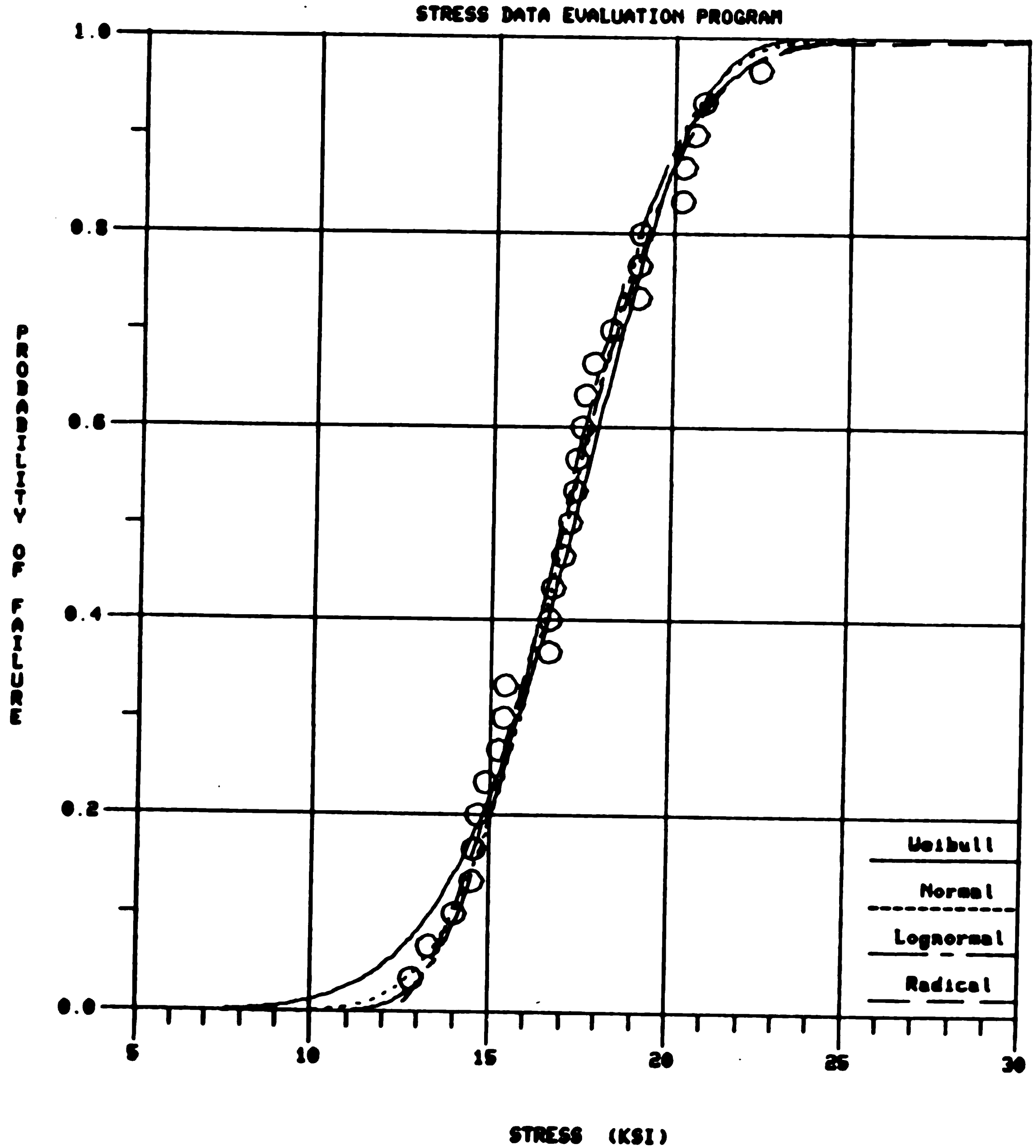


Figure 11 Quantile Box Plot - Bend Test Results



RMS ERROR

	Normal	Lognormal	Weibull	Radical
R1	.0364	.0357	.0448	.0389
R2	.0371	.0366	.0430	.0383
R3	.0387	.0366	.0425	.0384

		90% CONFIDENCE INTERVAL	
MEAN (KSI)=	17.085	16.308	17.663
S DEV (KSI)=	2.419	2.026	3.227

WEIBULL PARAMETERS

		90% CONFIDENCE INTERVAL	
SLOPE M=	7.318	5.718	9.387
CHAR VALUE=	18.135	17.347	18.000
99% ORIGIN=	9.962	7.956	(95% CL)
ORIGIN=	.000		

RADICAL PARAMETERS

A=	29.587	EXP B (N)=	2.000
B=	5.667	EXP(R)=	12.000
C=	-17.595		
SIGI=	11.993	SIGF=	35.255

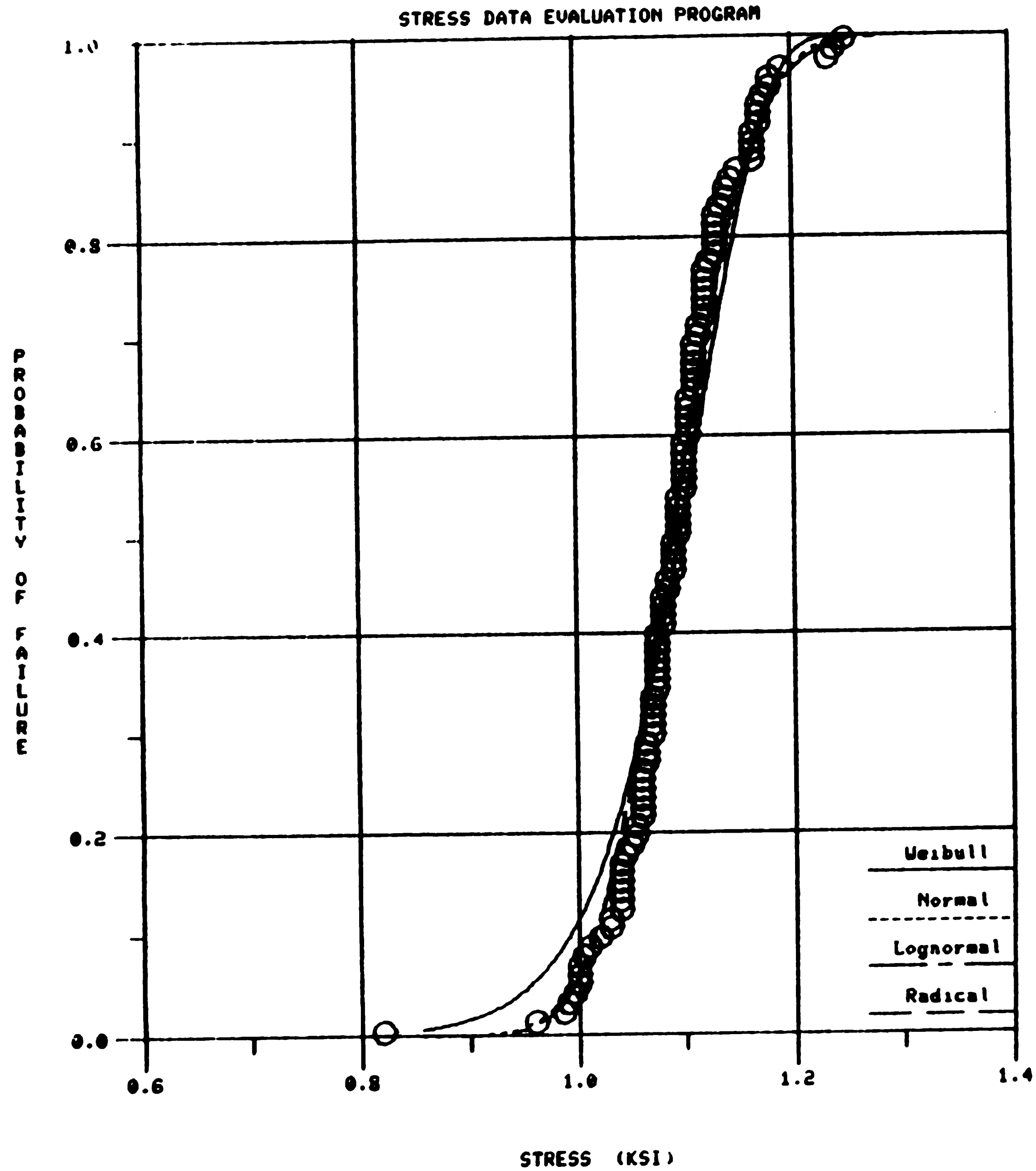
NON-PARAMETRI SOLN: A AND B ALLOWABLES
OPTIMAL: A(99.95,N=300)B(90.95,N= 30)
GIVEN DATA:
A(99.25,N= 29)B(90.95,N= 29)

WEIBULL:
DESIGN A= 7.956
DESIGN B= 11.961

NORMAL:
DESIGN A= 9.634
DESIGN B= 12.761

LOGNORMAL:
DESIGN A= 10.932
DESIGN B= 13.132

Figure 12 Statistical Data Evaluation - Robust Bend Test Results



RMS ERROR

	Normal	Lognormal	Weibull	Radical
R1	.0392	.0403	.0611	.0332
R2	.0385	.0396	.0603	.0342
R3	.0375	.0386	.0592	.0344

		90% CONFIDENCE INTERVAL	
MEAN (KSI)=	1.091	1.081	1.100
S DEV (KSI)=	.059	.053	.067

WEIBULL PARAMETERS

		90% CONFIDENCE INTERVAL	
SLOPE M=	18.963	16.800	21.484
CHAR VALUE=	1.118	1.108	1.128
99% ORIGIN=	.880	.846	(95% CL)
ORIGIN=	.000		

RADICAL PARAMETERS

A=	.972	EXP B (N)=	12.000
B=	.500	EXP(R)=	12.000
C=	-.372		
SIGI=	.600	SIGF=	1.472

NON-PARAMETRI SOLN: A AND B ALLOWABLES

OPTIMAL:A(99,95,N=300)B(90,95,N= 30)
GIVEN DATA:
A(99,67,N=109)B(90,88,N=109)

WEIBULL:

DESIGN A= .846
DESIGN B= .973

NORMAL:

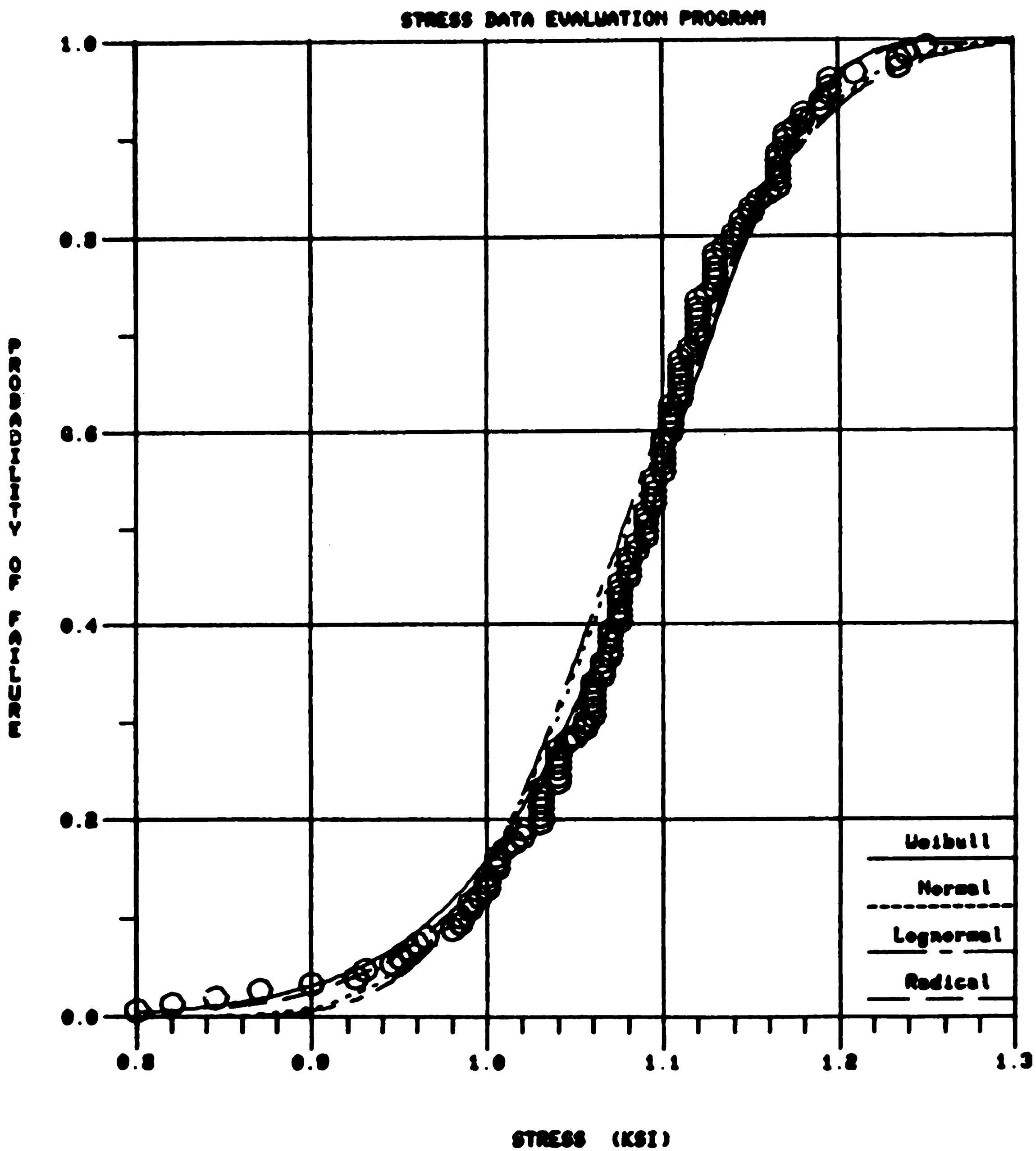
DESIGN A= .932
DESIGN B= 1.001

LOGNORMAL:

DESIGN A= .939
DESIGN B= 1.001

HLC0 HBC0 HAC0

Figure 13 Statistical Data Evaluation Kevlar (Compression Test) Partial Sample



RMS ERROR

	Normal	Lognormal	Weibull	Radical
R1	.0392	.0461	.0305	.0254
R2	.0388	.0458	.0301	.0258
R3	.0383	.0454	.0294	.0266

		90% CONFIDENCE INTERVAL	
MEAN (KSI)=	1.000	1.069	1.091
S DEV (KSI)=	.078	.072	.087

WEIBULL PARAMETERS

		90% CONFIDENCE INTERVAL	
SLOPE M=	16.026	14.313	18.078
CHAR VALUE=	1.115	1.104	1.126
99% ORIGIN=	.839	.069	(95% CL)
ORIGIN=	.000		

RADICAL PARAMETERS

A=	.643	EXP B (N)=	12.000
B=	.826	EXP(R)=	12.000
C=	-.353	SIGF=	1.470
SIGI=	.290		

NON-PARAMETRI SOLN: A AND B ALLOWABLES

OPTIMAL: A(99.95,N=300)B(90.95,N= 30)

GIVEN DATA:

A(99.77,N=146)B(90.22,N=146)

WEIBULL:

DESIGN A= .868
DESIGN B= .948

NORMAL:

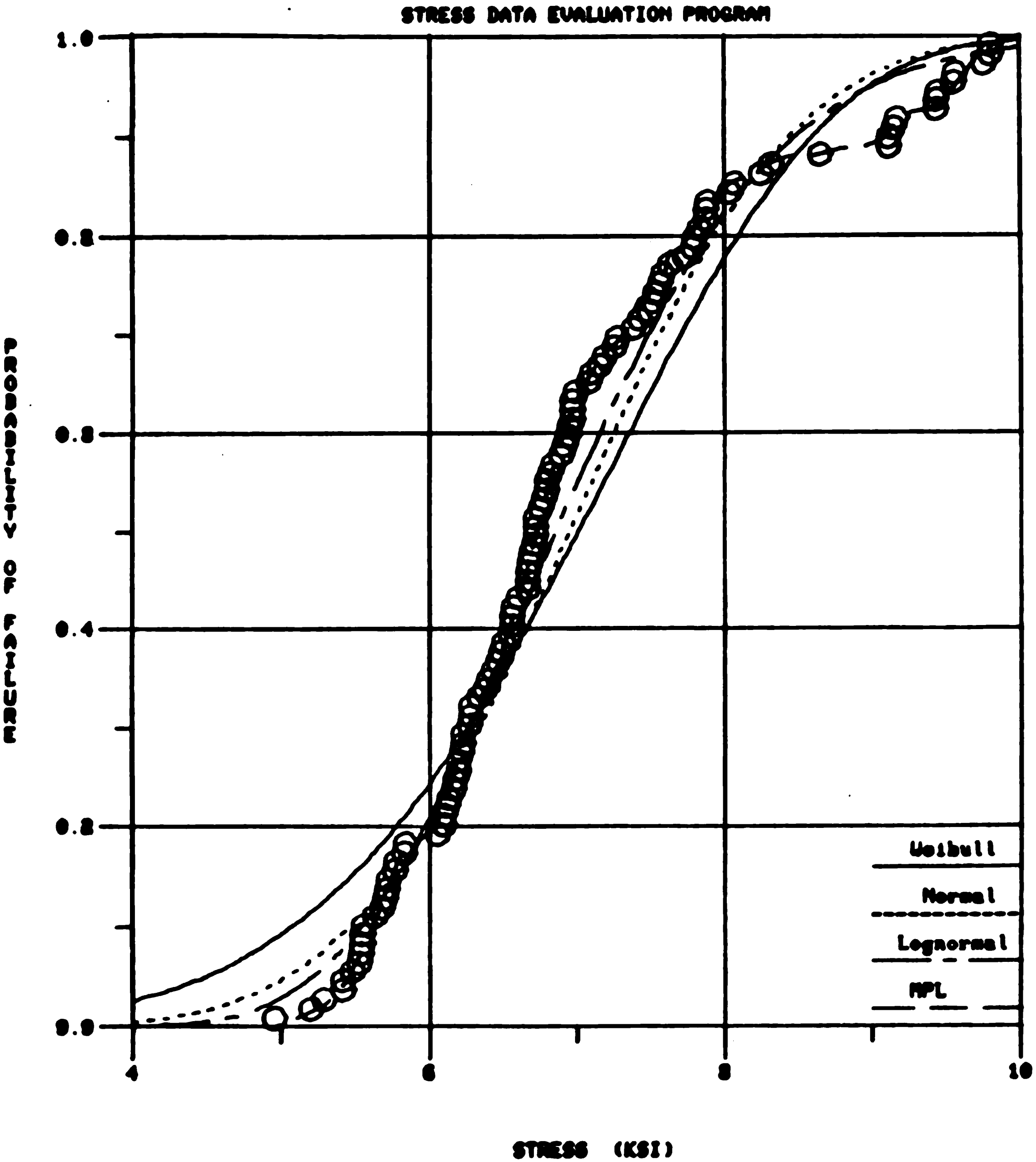
DESIGN A= .874
DESIGN B= .963

LOGNORMAL:

DESIGN A= .883
DESIGN B= .883

HLC0 + HBC0 + HAC0 + HCC0

Figure 14 Statistical Data Evaluation Kevlar (Compression Test) Complete Sample



RMS ERROR

	Normal	Lognormal	Weibull	MPL
R1	.0589	.0401	.0760	.0125
R2	.0587	.0389	.0760	.0122
R3	.0584	.0389	.0762	.0128

		90% CONFIDENCE INTERVAL	
MEAN (KSI)=	6.952	6.762	7.143
S DEV (KSI)=	1.122	1.046	1.310

WEIBULL PARAMETERS

		90% CONFIDENCE INTERVAL	
SLOPE M=	5.858	5.187	6.646
CHAR VALUE=	7.459	7.251	7.678
99% ORIGIN=	3.433	3.030	(95% CL)
ORIGIN=	.000		

MPL(A&B): 5.040 5.522

NON-PARAMETRI SOLN: A AND B ALLOWABLES

OPTIMAL: A(99.95,N=300)B(90.95,N= 30)
GIVEN DATA:
A(99.66,N=108)B(90.88,N=108)

WEIBULL:

DESIGN A=	3.030
DESIGN B=	4.758

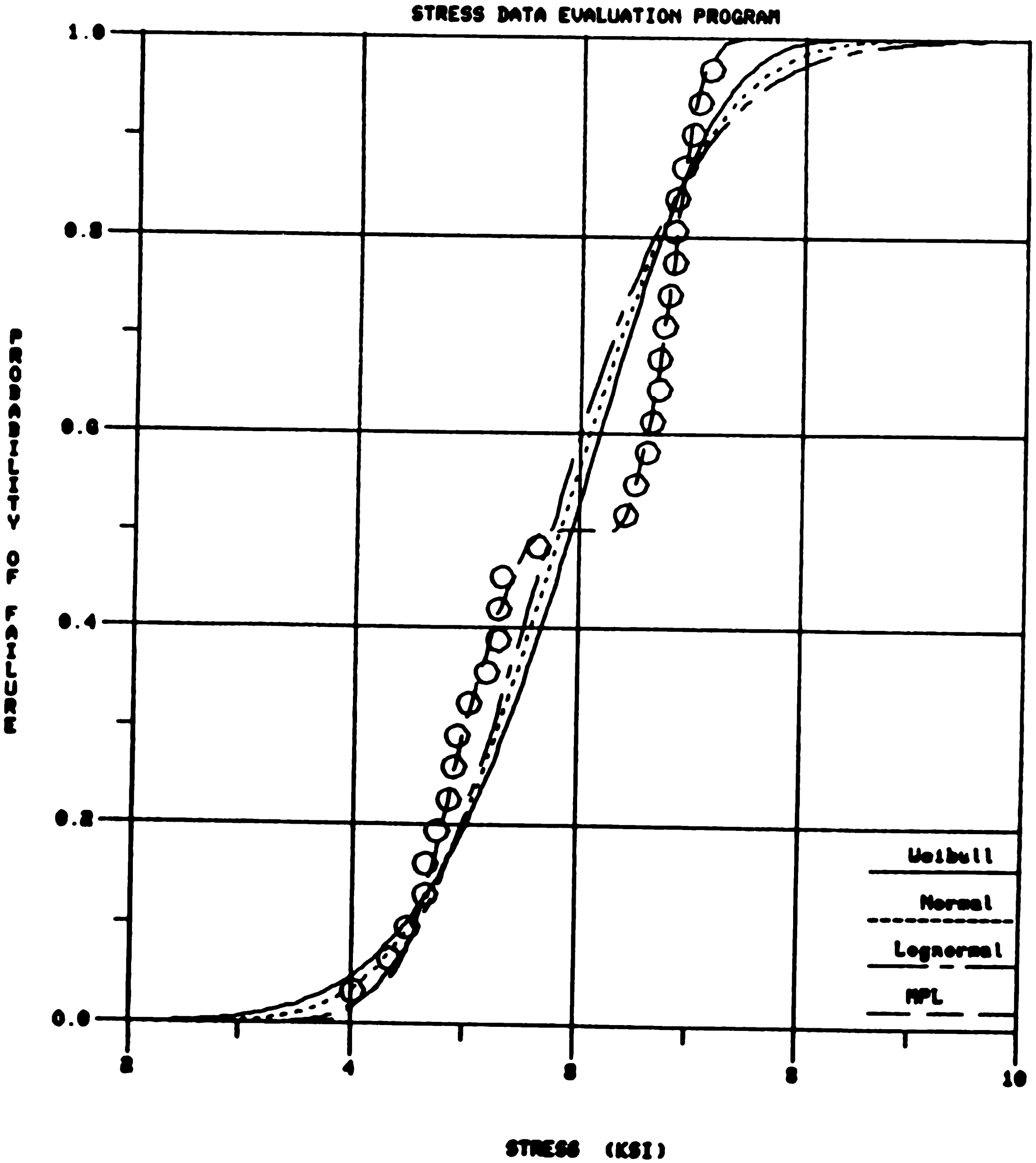
NORMAL:

DESIGN A=	3.852
DESIGN B=	5.190

LOGNORMAL:

DESIGN A=	4.477
DESIGN B=	5.383

Figure 15 Statistical Ranked Data and MPL Results



RMS ERROR

	Normal	Lognormal	Weibull	MPL
R1	.1000	.1026	.0997	.0187
R2	.1008	.1038	.1085	.0178
R3	.1022	.1953	.1019	.0180

		90% CONFIDENCE INTERVAL	
MEAN (KSI)=	5.834	5.513	6.154
S DEV (KSI)=	1.016	.853	1.382

WEIBULL PARAMETERS

		90% CONFIDENCE INTERVAL	
SLOPE M=	6.747	5.296	8.616
CHAR VALUE=	6.255	5.966	6.561
99% ORIGIN=	3.262	8.571	(95% CL)
ORIGIN=	.000		

MPL(A&B): 3.789 4.296

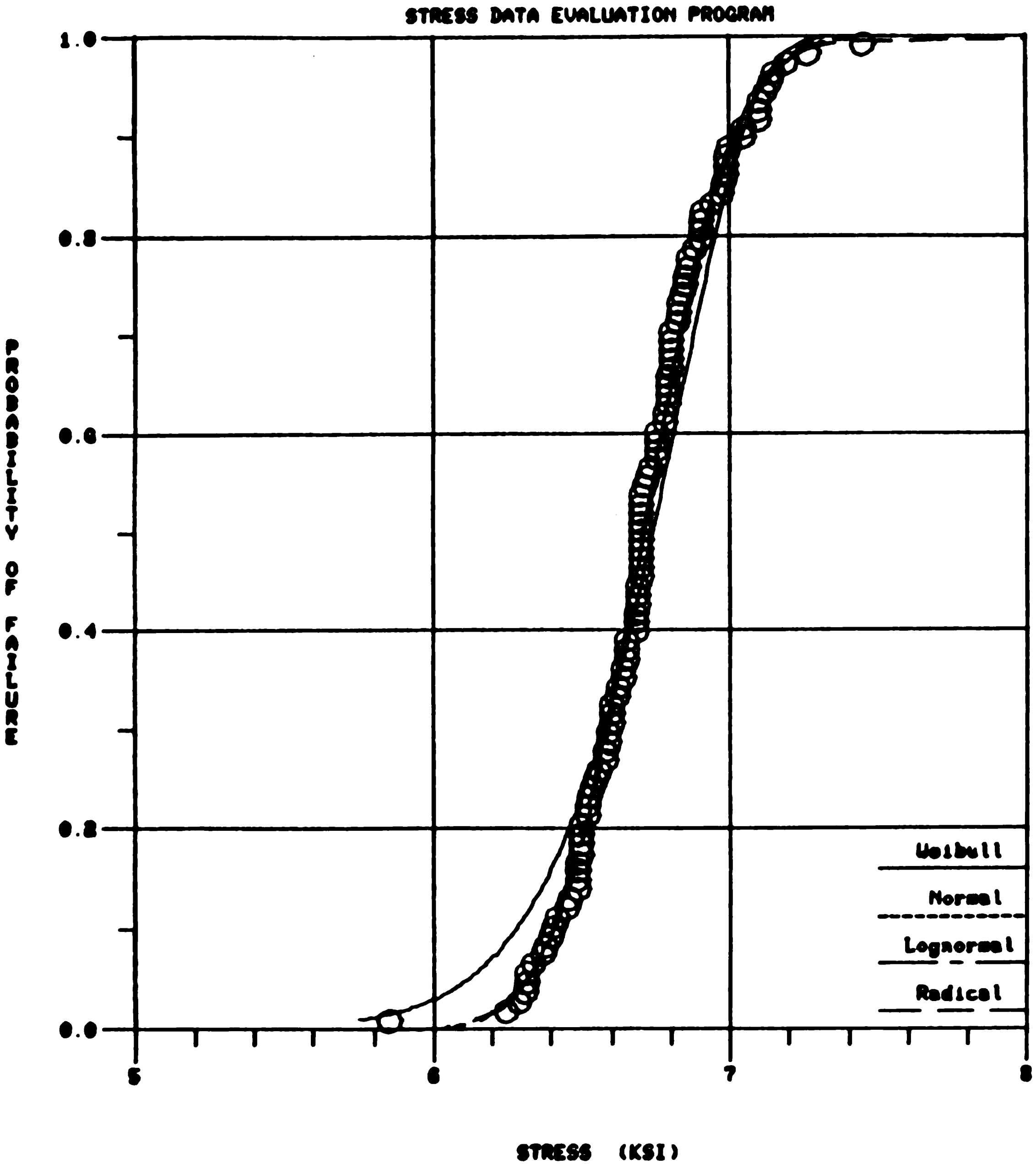
NON-PARAMETRI SOLN: A AND B ALLOWABLES
OPTIMAL: A(99.95,N=300)B(90.95,N= 30)
GIVEN DATA:
A(99.26,N= 30)B(90.96,N= 30)

WEIBULL:
DESIGN A= 2.571
DESIGN B= 3.987

NORMAL:
DESIGN A= 2.721
DESIGN B= 4.028

LOGNORMAL:
DESIGN A= 3.308
DESIGN B= 4.171

Figure 16 Tension Test Results for Kevlar Material



RMS ERROR

	Normal	Lognormal	Weibull	Radical
R1	.0267	.0257	.0584	.0220
R2	.0262	.0252	.0577	.0224
R3	.0256	.0246	.0566	.0232

		90% CONFIDENCE INTERVAL	
MEAN (KSI)=	6.712	6.672	6.752
S DEV (KSI)=	.244	.219	.275

WEIBULL PARAMETERS

		90% CONFIDENCE INTERVAL	
SLOPE M=	27.008	23.393	30.643
CHAR VALUE=	6.829	6.787	6.871
99% ORIGIN=	5.771	5.616	(95% CL)
ORIGIN=	.000		

RADICAL PARAMETERS

A=	6.806	EXP B (N)=	18.060
B=	1.764	EXP(R)=	12.000
C=	-1.807		
SIGI=	4.939	SIGF=	8.571

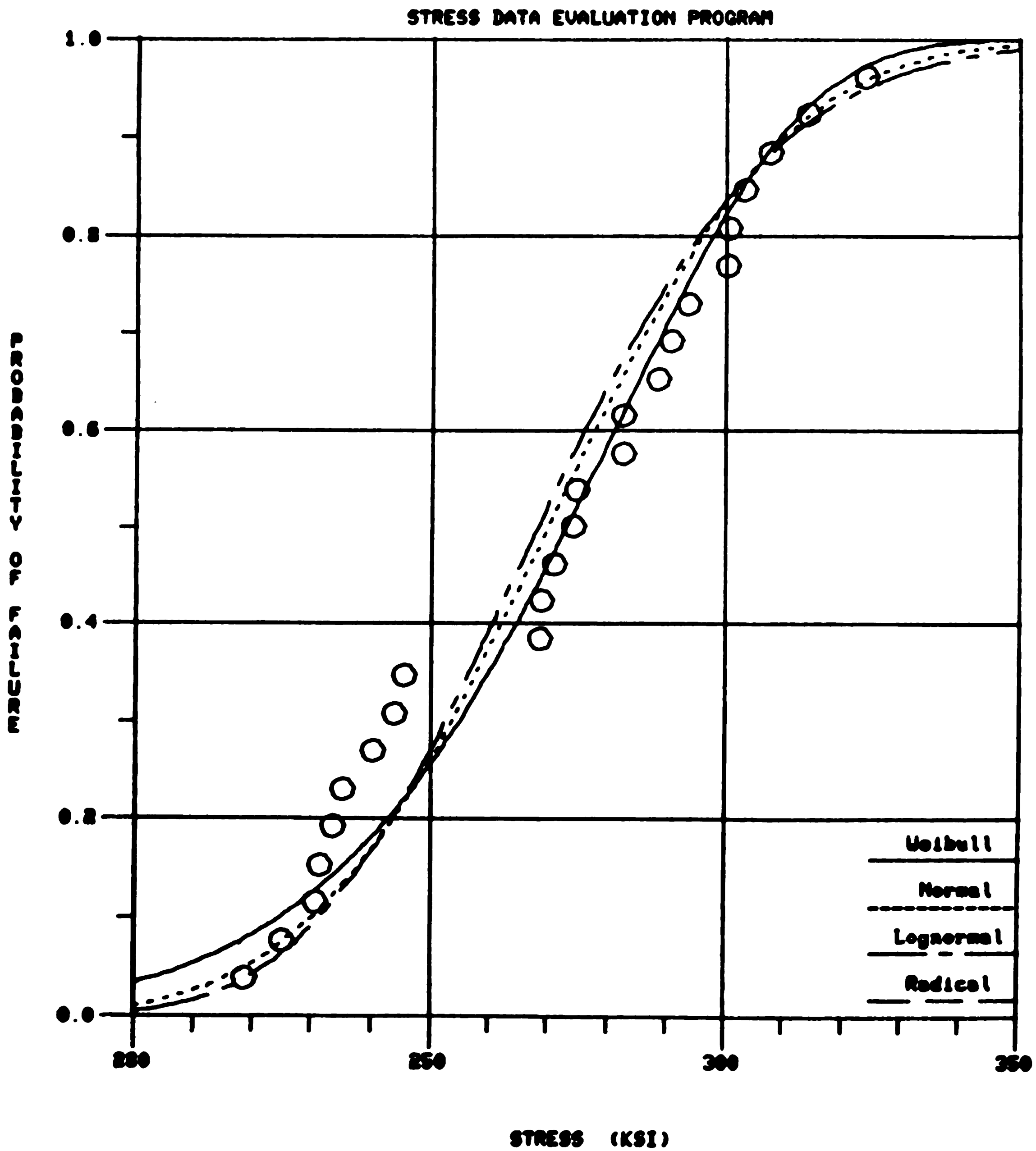
NON-PARAMETRI SOLN: A AND B ALLOWABLES
OPTIMAL: A(99.95,N=300)B(90.95,N= 30)
GIVEN DATA:
A(99.66,N=107)B(90.XX,N=107)

WEIBULL:
DESIGN A= 5.818
DESIGN B= 6.194

NORMAL:
DESIGN A= 6.061
DESIGN B= 6.342

LOGNORMAL:
DESIGN A= 6.086
DESIGN B= 6.347

Figure 17 Statistical Data Evaluation Composite (CYCOM) Tension



RMS ERROR

	Normal	Lognormal	Weibull	Radical
R1	.0585	.0668	.0455	.0431
R2	.0598	.0679	.0463	.0418
R3	.0623	.0703	.0480	.0400

		90% CONFIDENCE INTERVAL	
MEAN (KSI)=	280.946	259.000	290.966
S DEV (KSI)=	31.000	25.780	41.773

WEIBULL PARAMETERS

		90% CONFIDENCE INTERVAL	
SLOPE M=	9.673	7.418	12.000
CHAR VALUE=	283.527	273.439	294.139
99% ORIGIN=	180.289	149.834	(95% CL)
ORIGIN=	.000		

RADICAL PARAMETERS

A=	317.341	EXP B (N)=	8.000
B=	104.400	EXP(R)=	18.000
C=	-126.725	SIGF=	481.201
SIGI=	190.016		

NON-PARAMETRI SOLN: A AND B ALLOWABLES

OPTIMAL: A(99,95,N=300)B(90,95,N= 30)
GIVEN DATA:
A(89,82,N= 25)B(60,93,N= 25)

WEIBULL:
DESIGN A= 149.834
DESIGN B= 206.233

NORMAL:
DESIGN A= 171.757
DESIGN B= 212.806

LOGNORMAL:
DESIGN A= 195.446
DESIGN B= 216.370

Figure 18 Statistical Data Evaluation (Graphite Composite)

Conclusion

1. Exploratory Data Procedure should be applied prior to acceptance of statistical model used in the allowable computation.

2. Quantile Box Plot provides an excellent summary of test data results in addition to location outliers and multi-modality in the sample.

3. The authors recommend using the Weibull distribution function for obtaining the allowables, if outliers (higher ordered values) or multi-modality, do not exist in data set.

4. An extreme value distribution (Weibull) provides a degree of security if pooled samples do not represent general data population. (a common occurrence).

5. Normal distribution is recommended for allowable computation if outliers exist in data set at upper tail region.

6. In multi-modality case, the Maximum Penalized Likelihood Method is suggested for the allowable determination.

7. At present the authors recommend pooling all samples made available even though significant difference test indicated otherwise. If a sample contains modality then this data set would be excluded.

8. The development of tables 1 and 2 required a considerable amount of effort (most complete tabulation for the Weibull function). With the aid of computer code listed in appendix A and tabulated P_s^* values, the reader should be able to obtain a simple and accurate computation of the allowables when the Weibull function is considered.

9. Non-parametric procedures are always the most desirable in obtaining the allowables for the given sample if properly applied.

10. A sufficiently large number of sources in obtaining test data is more important in determining allowables than size of individual sample.

Literature Cited

[1] Hahn, G. J., and Shapiro, S. S., "Statistical Models in Engineering", John Wiley and Sons, Inc., New York, London and Sidney, 1967

[2] Good, I. J., and Gaskins, R. A., "Density Estimation and Bump - Hunting by the Penalized Likelihood Method Explified by Scattering and Meteorita Data", Journal of the American Statistical Assoc., Vol. 75, No. 369, March 1980

[3] Freudenthal, A. M.. and Gumbel, E. J., "On the Statistical Interpretation of Fatigue Tests", Proc. Royal Soc. (A) 216, 309, 1953

[4] Parzen, Emmanuel, "Entropy Interpretation of Tests for Normality by Shapiro-Wilk Statistics", presented at the Twenty-Eighth Conference on the Design of Experiments in Army Research, Development and Testing, 20-22 Oct. 1982, Monterrey, California

[5] Parzen, Emmanuel, "A Density-Quantile Function Perspective on Robust Estimation", Robustness in Statistics, Academic Press, New York, 1979

[6] Huber, P. J., "Robust Estimation of a Location Parameter", Annuals of Mathematical Statistics, 35, 73-101, 1964

[7] Andrews, D. F.; Bickel, P. J.; Hampel, F. R.; Huber, P. J.; Roger, W. H.; Tukey, J. W., "Robust Estimates of Location", Princeton University Press, 1972

[8] Tukey, J. W., "Exploratory Data Analysis", Addison-Wesley Publishing Company, 1977

[9] Harter, L. H., and Moore, A. H., "Maximum Likelihood Estimation of the Parameter of the Gamma and Weibull Population from Complete and Censored Samples", Journal of the American Statistical Association, Vol. 61, 842-851, 1966

[10] Thoman, D. R., Bain, L. J., Antle, C. E., "Inference on the Parameters of the Weibull Distribution", Technometrics, Vol. 11, 445-459, May, 1969

[11] Bradley, James V., "Distribution - Free Statistical Tests", Prentice-Hall, Inc., Englewood Cliffs, New Jersey

[12] Walsh, J. E., Handbook of Non-parametric Statistics, Vol. II, 326, D. Van Nostrand Company, Inc., New Jersey

```
C                                   ^     ^
C         WEIBULL PARAMETERS M AND U
C
C-----------------------------------------------------C
C
C     READS IN AMOUNT OF DATA AND THE DATA
C
          DIMENSION T(200)
          IMPLICIT DOUBLE PRECISION (A-H,O-Z)
          READ(5,1)NPTS                   <-------------Sample Size
          READ(5,1)(T(I), I=1,NPTS)    <---------Data
C                  ^
C     DETERMINES M BY MAXIMUM LIKELIHOOD
C
          A=0.0
          N=0.0
          DA=1.0
  2       XNUM=0.0
          XMINUS=0.0
          PROD1=0.0
          PROD2=0.0
          A=A+DA
C
C     SUMATIONS CALCULATED FIRST
C
          DO 10 I=1,NPTS
          XNUM=T(I)**A+XNUM
          XMINUS=T(I)**A*LOG(T(I))+XMINUS
 10       PROD2=LOG(T(I))+PROD2
          A1=NPTS*XNUM/(NPTS*XMINUS-XNUM*PROD2)
          IF(A-A1)2,5,3
  3       N=N+1
          IF(N-5)4,4,5
  4       A=A-DA
          DA=1./10.**N
          A=A-DA
          GO TO 2                                     ^
  5       WRITE(6,1)A         <--------------m
C                  ^
C     DETERMINES U 'HAT'
C
          BSUM=0.0
          DO 6 I=1,NPTS
  6       BSUM=T(I)**A+BSUM
          B=(1./NPTS*BSUM)**(1./A)                     ^
          WRITE(6,1)B    <-------------------u
  1       FORMAT( )
          END
```

Appendix A   Computer Code for Determining Weibull Parameters

CHOICE OF RESPONSE SURFACE DESIGN AND ALPHABETIC OPTIMALITY

George E. P. Box

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

ABSTRACT

It is argued that the specification of problems of experimental design (and in particular, of response surface design) should depend on scientific context. The specification for a widely developed theory of "alphabetic optimality" for response surface applications is analyzed and found to be unduly limiting. Ways in which designs might be chosen to satisfy a set of criteria of greater scientific relevance are suggested. Detailed consideration is given to regions of operability and interest, to the design information function, to sensitivity of criteria to size and shape of the region, and to the effect of bias. Problems are discussed of checking for lack of fit, sequential assembly, orthogonal blocking, estimation of error, estimation of transformations, robustness to bad values, using minimum numbers of points, and employing simple data patterns.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

ice of Response Surface Design and Alphabetic Optimality

George E. P. Box

University of Wisconsin - Madison

1. INTRODUCTION

There seems no doubt that of all the activities in which the statistician can engage, that of designing experiments is by far the most important, since it is here that the actual mode of generation of scientific data is decided.

The importance of practice in guiding the development of the theory of experimental design [45] is clearly seen from the time of its invention. Fisher was engaged by Russell [16] on a temporary basis at Rothamsted Experimental Station in 1919 "to examine our data and elicit further information that we had missed." Records were available from the ongoing Broadbalk experiment in which particular combinations of fertilizers had been consistently applied to 13 plots for a period of almost 70 years. In his analysis ([22], [24]), Fisher attempted to relate yield to fertilizer combination, to weather, and in particular, to rainfall. The method he used was multiple regression with distributed lag models, involving an ingenious employment of orthogonal polynomials which led to important advances in the theory of regression analysis, and in particular its distribution theory.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

With only the crudest of computational aids, the work must have been burdensome, making it all the more frustrating to discover that, however ingenious the analysis, the inherent nature of the data ensured that the answers to many questions were inaccessible. A comprehension of the logical problems in drawing conclusions from such analyses led naturally to speculation on how some of the difficulties might be overcome by appropriate design. These ideas were further stimulated by the Analysis of Variance, which Fisher introduced in 1923 with W.A. Mackenzie [23] for the elucidation of what was clearly a most unsatisfactory design which he had had no part in choosing. Thereafter, as Fisher gradually acquired more influence in the setting up of field trials, the principles of replication, randomization and their application to randomized blocks, latin squares and factorial designs quickly evolved out of the actual planning, running, and analysis of a series of experimental designs of increasing complexity and beauty.

The practical context of scientific experimentation continued to produce important theoretical advances when Yates came to Rothamsted in 1931, leading in particular to important developments in the design and analysis of complex factorial designs and their associated systems of confounding ([44], [46]) and to the introduction of incomplete block designs.

My own experience with experimental design began during the Second World War. I worked at the Chemical Defense Experimental Station in England with a group of medical research workers who were attempting, using animals and volunteers, to find ways to combat the effects of poison gas and other toxic agents. At this time it was believed that these agents might be used not only against the military, but also against the civilian population. It was important therefore that our work should progress as rapidly as possible. I found myself a part of evolving investigations which employed sequences of experiments which I designed and whose nature needed to adapt to changing needs at different stages of the study. The designs employed were randomized blocks, balanced incomplete blocks, latin squares, and factorials. Later, during my eight years as a statistician with Imperial Chemical Industries, my role was again as a member of various scientific teams tackling evolving problems with sequences of designs. Many of the problems were similar to those I had previously encountered, and again employed the (by now) standard designs of Fisher and Yates. However, some investigations directly concerned with the improvement of chemical processes at the lab, pilot plant, and full scale, seemed to require additional methods, which however, still drew on the fundamental principles laid down by the originators of experimental design. This led to the development of what has come to be called response surface methodology. See for example [4], [14], [15], [30], [31], and [39].

Suppose some response η of interest is believed to be locally approximated by a polynomial of low degree in k continuous experimental variables $\underset{\sim}{x} = (x_1, x_2, \ldots, x_k)'$. To fit such a function we need appropriate experimental designs. Let us call a design suitable for estimating a general polynomial of degree d a d̲th order design in k variables. Thus a design suitable for fitting the function

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2$$

would be a second order design in two variables.

One route for choosing such designs, which has generated an enormous amount of mathematical research over the last twenty or so years, we shall refer to as the "alphabetic optimality" approach. For reasons I will explain, I have reservations about the usefulness of this approach so far as response surface designs are concerned. For completeness, a brief summary of some of the main ideas are set out below ([32], [33], [34], [35], [36], [42], [43]).

2. SOME ASPECTS OF OPTIMAL DESIGN THEORY FOR CONTINUOUS EXPERIMENTAL VARIABLES

Consider a response η which is supposed to be an exactly known function $\eta = \underset{\sim}{x}'\underset{\sim}{\beta}$ linear in p coefficients $\underset{\sim}{\beta}$, where $\underset{\sim}{x} = \{f_1(\underset{\sim}{\chi}), f_2(\underset{\sim}{\chi}), \ldots, f_p(\underset{\sim}{\chi})\}'$ is a vector of p functions of k experimental variables $\underset{\sim}{\chi}$. Suppose a design is to be run defining n sets of k experimental conditions given by the n x k design matrix $\{\underset{\sim}{\chi}_u\}$ and yielding n observations $\{y_u\}$, so that

$$\eta_u = \underset{\sim}{x}_u'\underset{\sim}{\beta} \qquad (u = 1, 2, \ldots, n)$$

where $y_u - \eta_u = \varepsilon_u$ is distributed $N(0,\sigma^2)$ and the n x p matrix $\underset{\sim}{X} = \{\underset{\sim}{x}_u'\}$.

The elements of $\{c_{ij}\} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}$ are proportional to the variances and covariances of the least squares estimates $\hat{\underset{\sim}{\beta}}$. Within this specification, the problem of experimental design is that of choosing the design $\{\underset{\sim}{\chi}_u\}$ so that the elements c_{ij} are to our liking. Because there are $\frac{1}{2}p(p+1)$ of these, simplification is desirable.

A motivation for simplification is provided by considering the confidence region[1] for $\underset{\sim}{\beta}$

---

[1] Obviously there are also parallel fiducial and Bayesian rationalizations.

$$(\underset{\sim}{\beta} - \hat{\underset{\sim}{\beta}})'\underset{\sim}{X}'\underset{\sim}{X}(\underset{\sim}{\beta} - \hat{\underset{\sim}{\beta}}) = \text{constant}$$

defining an ellipsoid in p parameters. The eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ of $(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$ are proportional to the squared lengths of the p principal axes of this ellipsoid. Suppose their maximum, arithmetic mean, and geometric mean are indicated by λ_{max}, $\bar{\lambda}$, and $\tilde{\lambda}$. Then it is illuminating to consider the transformation of the $\frac{1}{2}p(p+1)$ elements c_{ij} to a corresponding number of items as follows:

(i) $D = |X'X| = \tilde{\lambda}^{-p}$ (so that $D^{-1/2} = \tilde{\lambda}^{p/2}$ is proportional to the volume of the confidence ellipsoid).

(ii) $\underset{\sim}{H}$, a vector of p - 1 homogeneous functions of degree zero in the λ's, which measure the non-sphericity or state of ill-conditioning of the ellipsoid. In particular we might choose, for two of these, $H_1 = \bar{\lambda}/\tilde{\lambda}$ and $H_2 = \lambda_{max}/\tilde{\lambda}$, both of which would take the value unity for a spherical region.

(iii) $\frac{1}{2}p(p-1)$ independent direction cosines which determine the orientation of the orthogonal axes of the ellipsoid.

It is traditionally assumed that the $\frac{1}{2}p(p-1)$ elements concerned with orientation of the ellipsoid are of no interest, and attention has been concentrated on particular criteria which measure in some way or another the sizes of the eigenvalues, measuring some combination of size and sphericity of the confidence ellipsoid. Among these criteria are

$$D = |X'X| = \Pi\lambda_i^{-1} = \tilde{\lambda}^{-p}$$

$$A = \Sigma\lambda_i = \text{tr}(\underset{\sim}{X}'\underset{\sim}{X})^{-1} = \sigma^{-2}\Sigma\text{var}(\hat{\beta}_i) = p\tilde{\lambda}H_1$$

$$E = \max\{\lambda_i\} = \tilde{\lambda}H_2$$

The desirability of a design, as measured by the D, A, and E criteria, increases as $\bar{\lambda}$, $\bar{\lambda}H_1$, and $\bar{\lambda}H_2$ respectively, are decreased. But in practical situations, each of these criteria will take smaller and hence more desirable values as the ranges for the experimental variables χ are taken larger and larger. To cope with this problem it is usually assumed that the experimental variables $\underset{\sim}{\chi}_u$ may vary only within some exactly known region in the space of $\underset{\sim}{\chi}$, but not outside it. I will call this permissible region $\underset{\sim\sim}{RO}$.

Another characteristic of the problem which makes its study mathematically difficult is the necessary discreteness of the number of runs which can be made at any given location. In a technically brilliant paper [37], Kiefer and Wolfowitz dealt with this obstacle by introducing a continuous design measure ξ which determines the proportion of runs which should ideally be made at each of a number of points in the $\underset{\sim}{\chi}$ space. Realizable designs which most nearly approximated the optimal distribution could then be used in practice.

A further important result of Kiefer and Wolfowitz linked the problem of estimating $\underset{\sim}{\beta}$ with that of estimating the response η via the property of "G-optimality." G-optimal designs were defined as those which minimized the maximum value of $V(\hat{y}_x)$ within $\underset{\sim\sim}{RO}$. The authors were then able to show, for their measure designs, the equivalence of G- and D-optimality. Furthermore, they showed that, for such a design, within the region $\underset{\sim\sim}{RO}$, the maximum value of $n \cdot Var(\hat{y}_x)/\sigma^2$ was p, and that this value was actually attained at each of the design points.

For illustration we consider a second order measure-design in two variables; that is, a design appropriate for the fitting of the second degree polynomial of equation (1). Such a design which is both D- and G- optimal for a square region $\underset{\sim\sim}{RO}$ with vertices $(\pm 1, \pm 1)$ was given by Fedorov [21] (see also Herzberg [27]). The design places 14.6% of the measure at each of the four vertices, 8.0% at each of the midpoints of the edges, and 9.6% at the origin. The design is set out in Figure 4(b).

While this approach has generated much interesting mathematics, it does not, I believe, solve the problem of choosing good response surface designs. In the hope of stimulating new initiative, I have set out below what I believe is the scientific context for response surface studies and indicated some possible lines of development.

3. THE RESPONSE SURFACE CONTEXT

As an example suppose it is desired to study some chemical system, with the object of obtaining a higher value for a response η such as yield which is initially believed to be some function $\eta = g(\underset{\sim}{\chi})$ of k continuous input variables $\underset{\sim}{\chi} = (\chi_1, \chi_2, \ldots, \chi_k)'$ such as reaction time, temperature, or concentration. As is illustrated in Figure 1, it is usually known initially that the system can be operated at some point $\underset{\sim}{\chi}_0$ in the space of $\underset{\sim}{\chi}$ and is expected to be capable of operating over some much more extensive region $\underset{\sim}{O}$ called the <u>operability region</u>, which[2] however is usually unknown or poorly known. Response surface methods are employed when the nature of the true response function $\eta = g(\underset{\sim}{\chi})$ is also unknown[3] or is inaccessible.

---

[2] One secondary object of the investigation may be to find out more about the operability region O.

[3] Occasionally the true functional form $\eta = g(\underset{\sim}{\chi})$ may be known, or at least conjectured, from knowledge of physical mechanisms. Typically however $g(\underset{\sim}{\chi})$ will then appear as a solution of a set of differential equations which are nonlinear in a number of parameters which may represent physical constants. Problems of nonlinear experimental design then arise which are of considerable interest although they have received comparatively little attention (see for example [13], [18], [25]).

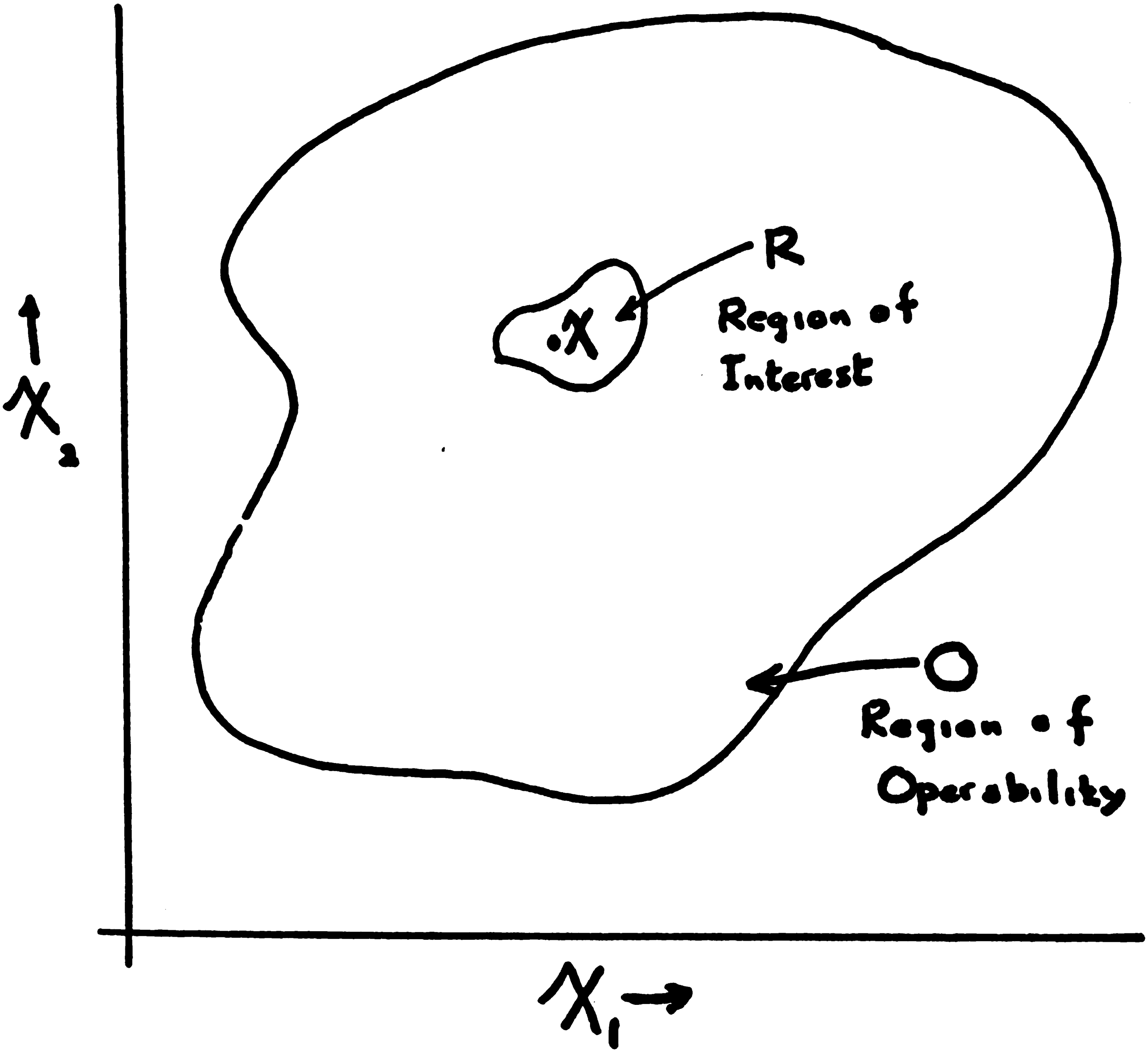


Figure 1. The current region of interest $\tilde{R}$ and the region of operability $\tilde{O}$ in the space of two continuous experimental variables X_1 and X_2.

Suppose that over some (typically much less extensive) immediate region of interest $\underset{\sim}{R}$ in the neighborhood of $\underset{\sim}{\chi}_0$ it is guessed that a "graduating" function, such as a dth degree polynomial in x,

$$\eta_{\underset{\sim}{\chi}} \doteq \underset{\sim}{x}'\underset{\sim}{\beta}$$

might provide a locally adequate approximation to the true function $\eta_{\underset{\sim}{\chi}} = g(\underset{\sim}{\chi})$ where as before $\underset{\sim}{x}$ is a p-dimensional vector of suitably transformed input variables $\underset{\sim}{x}' = \{f_1(\underset{\sim}{\chi}), f_2(\underset{\sim}{\chi}), \ldots, f_p(\underset{\sim}{\chi})\}$, and $\underset{\sim}{\beta}$ is a vector of coefficients occurring linearly that may be adjusted to approximate the unknown true response function $\eta_{\underset{\sim}{\chi}} = g(\underset{\sim}{\chi})$. Then progress may be achieved by using a sequence of such approximations. For example when a first degree polynomial approximation could be employed it might, via the method of steepest ascent, be used to find a new region of interest R_1 where, say, the yield was higher. Also a maximum in many variables is often represented by some rather complicated ridge system[4] and a second degree polynomial approximation when suitably analysed might be used to elucidate, describe, and exploit such a system.

Thus we are typically involved in using a sequence of designs, each making use of information gleaned from earlier experiments -- a characteristic typical of a much wider field of scientific investigation. This provides the

---

[4] Empirical evidence suggests this. Also, integration of sets of differential equations which describe the kinetics of chemical systems almost invariably leads to ridge systems ([4], [15], [26], [41]). See also the discussion of Figure 8.

opportunity to progressively improve not only the objective function η directly, but also the mode of gathering information about it. For example, at the ith stage, a design performed in a region $\underset{\sim}{R}_i$ may suggest that a new region $\underset{\sim}{R}_{i+1}$ is worthy of investigation (either because it can be expected to give higher values of η or because it may throw light on other important aspects of the function). But this new region may be different not only in (a) its location in the space of $\underset{\sim}{\chi}$, but (b) in its shape also (for instance because of information fed back from previous data on transformations of χ's individually or jointly), and (c) in the identity of its component space (because of feedback from the results themselves, indicating that certain variables should be dropped, and/or that new variables should be added). Thus in any realistic view of the process of investigation the dimensions, identity, location and metrics of measurement of regions of interest in the experimental space are all iteratively evolving. The problem of choosing suitable experimental designs in such a context is a difficult one. Some properties ([5], [8]) of a response surface design, any, all or some of which might in different circumstances be of importance in the above context are given in Table 1.

The design information function

Associated with requirements (1) and (2) of Table 1, consider the design variance function [11]

$$V_{\underset{\sim}{\chi}} = n \cdot \mathrm{Var}(\hat{y}_{\underset{\sim}{\chi}})/\sigma^2 = n\underset{\sim}{x}'(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{x}$$

or equivalently the Information Function

$$I_{\underset{\sim}{\chi}} = V_{\underset{\sim}{\chi}}^{-1} .$$

The design should:

(i) generate a satisfactory distribution of information throughout the region of interest, R;

(ii) ensure that the fitted value at x, $\hat{y}(x)$ be as close as possible to the true value at x, $\eta(x)$;

(iii) give good detectability of lack of fit;

(iv) allow transformations to be estimated;

(v) allow experiments to be performed in blocks;

(vi) allow designs of increasing order to be built up sequentially;

(vii) provide an internal estimate of error;

(viii) be insensitive to wild observations and to violation of the usual normal theory assumptions;

(ix) require a minimum number of experimental points;

(x) provide simple data patterns that allow ready visual appreciation;

(xi) ensure simplicity of calculation;

(xii) behave well when errors occur in the settings of the predictor variables, the x's;

(xiii) not require an impractically large number of predictor variable levels;

(xiv) provide a check on the 'constancy of variance' assumption.

TABLE 1. SOME ATTRIBUTES OF DESIGNS OF POTENTIAL IMPORTANCE

It is evident that if we were to make the unrealistic assumption (made in alphabetic optimality) that the graduating function $\eta = \underset{\sim}{x}'\underset{\sim}{\beta}$ is capable of <u>exactly</u> representing the true function $g(\underset{\sim}{\chi})$, then the information function would tell us all we could know about the design's ability to estimate η. For illustration, information functions and associated information contours for a 2^2 factorial used as a first order design and for a 3^2 factorial used as a second order design are shown in Figures 2 and 3, for standard variables x_1 and x_2.

4. APPLICABILITY OF ALPHABETIC OPTIMALITY

The information function for Fedorov's second order D/G-optimal design over the permissible $\underset{\sim}{RO}$ region $(\pm1, \pm1)$, referred to earlier, is shown in Figure 4. For illustration, this is related to the two experimental variables χ_1 = temp in ^{0}C and χ_2 = time in hours. Thus, in this particular example, $x_1 = (\chi_1 - 180)/10$, $x_2 = \chi_2 - 4$ and the $\underset{\sim}{RO}$ region would permit experimentation within the limits $\chi_1 = 170 - 190$ °C and $\chi_2 = 3 - 5$ hours, but not outside these limits. In the response surface context a number of questions arise concerning the appropriateness of the specification set out in Section 2 of this paper for alphabetic optimality. These concern

(i) Formulation in terms of the $\underset{\sim}{RO}$ region
(ii) Distribution of information over a wider region
(iii) Sensitivity of criteria to size and shape of the $\underset{\sim}{RO}$ region
(iv) Ignoring of bias.

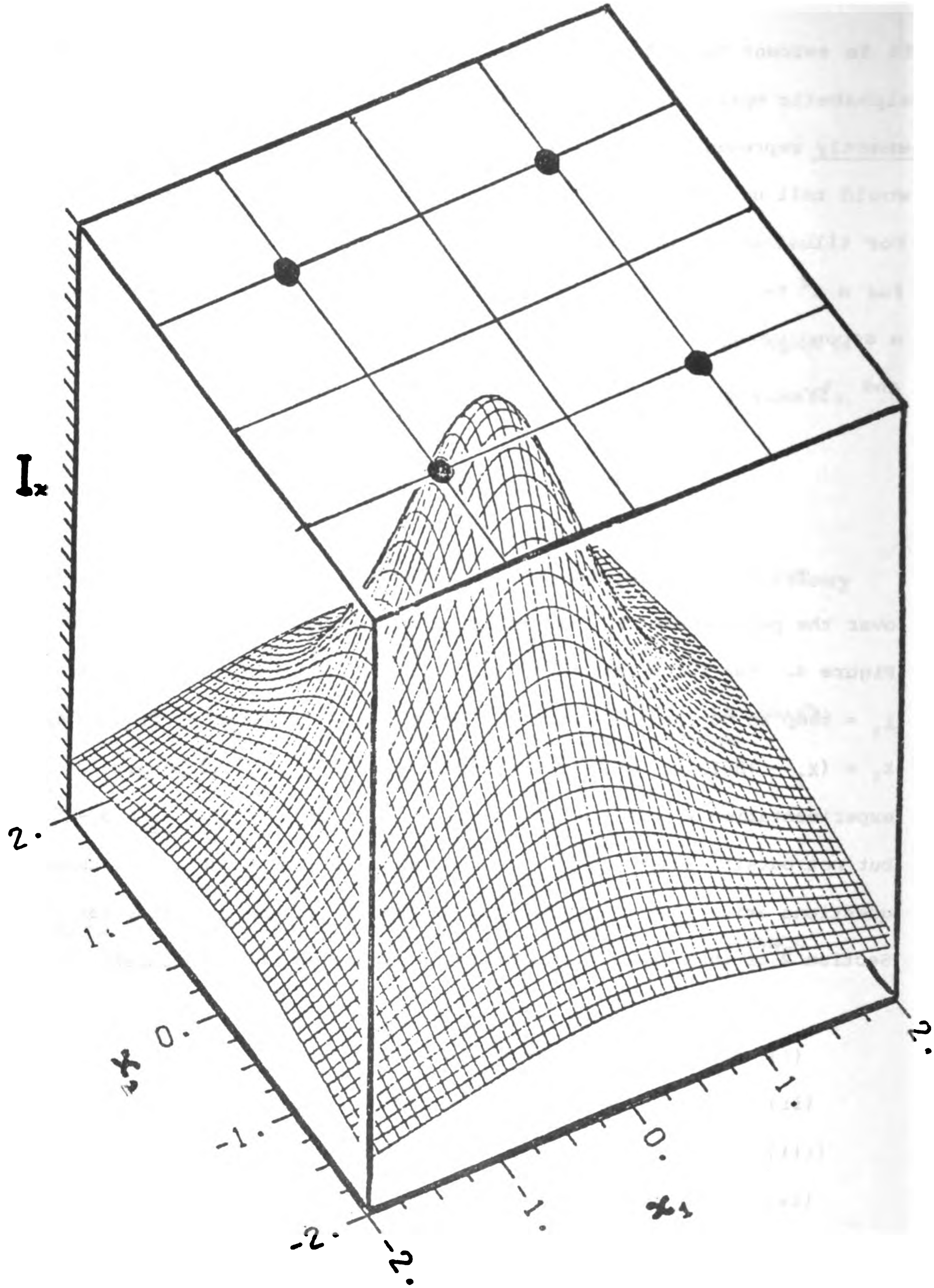


Figure 2(a) Information surface for a 2^2 factorial used as a first order design.

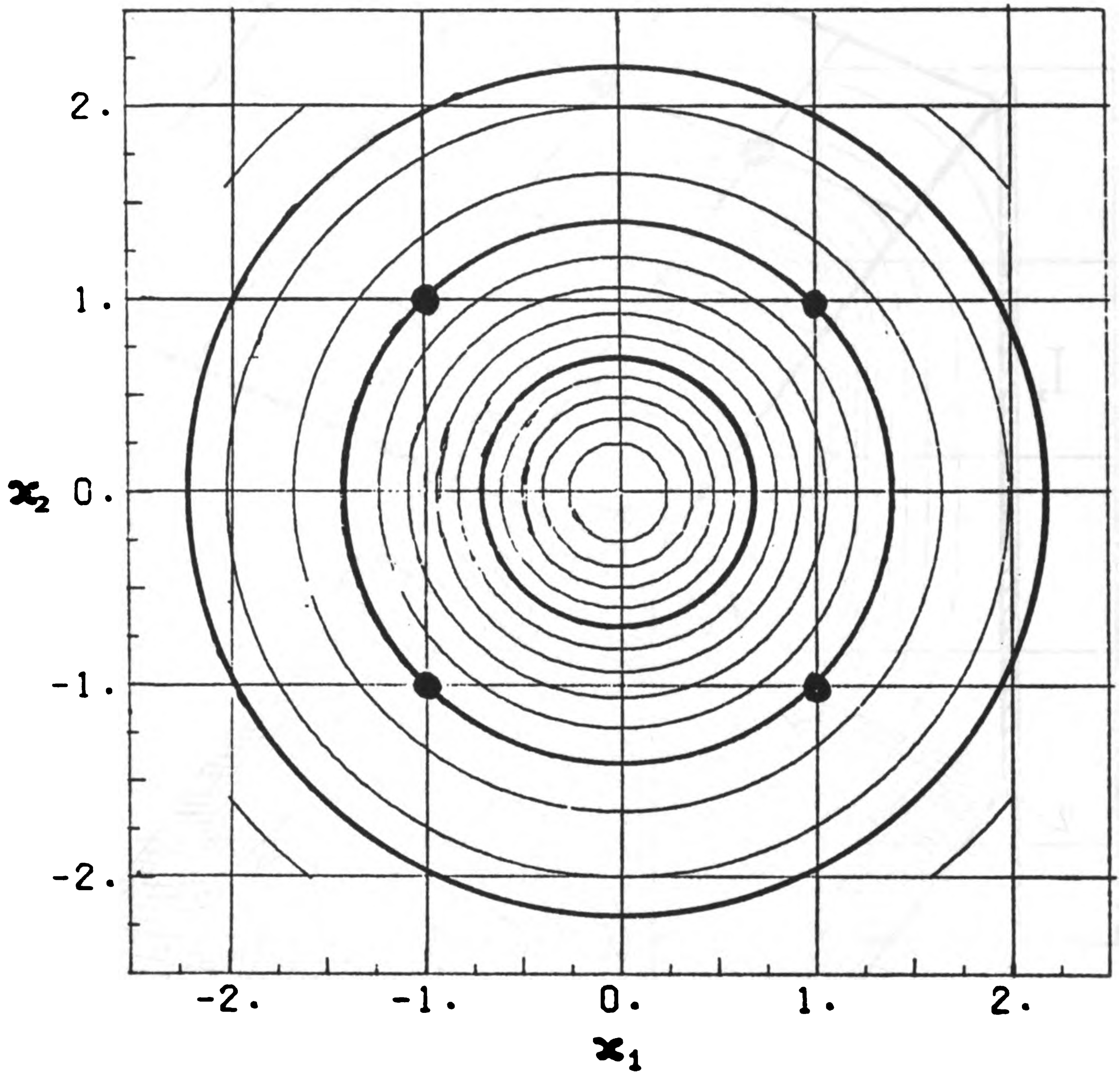


Figure 2(b) Information contours for a 2^2 factorial used as a first order design.

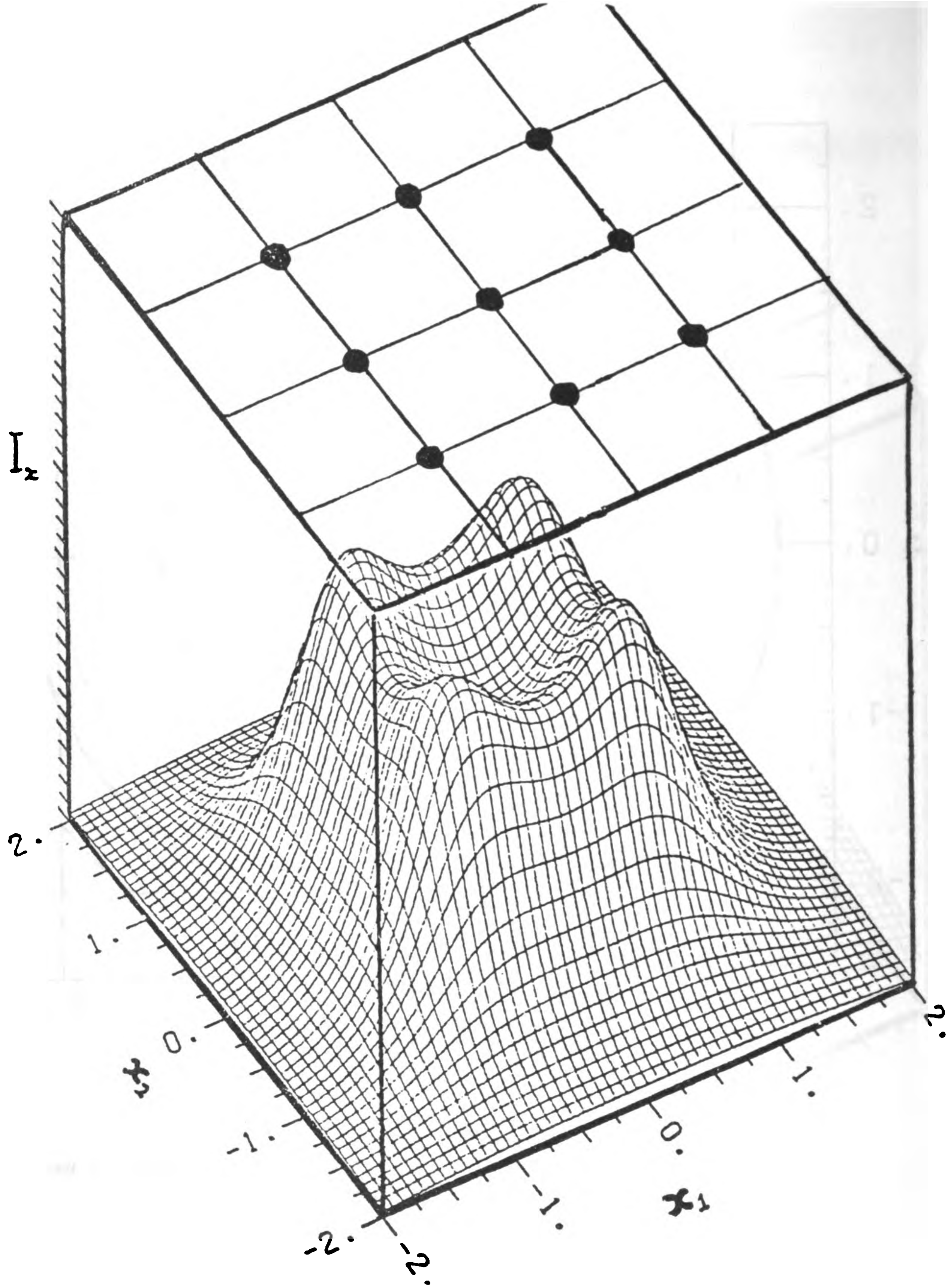


Figure 3(a) Information surface for a 3^2 factorial used as a second order design.

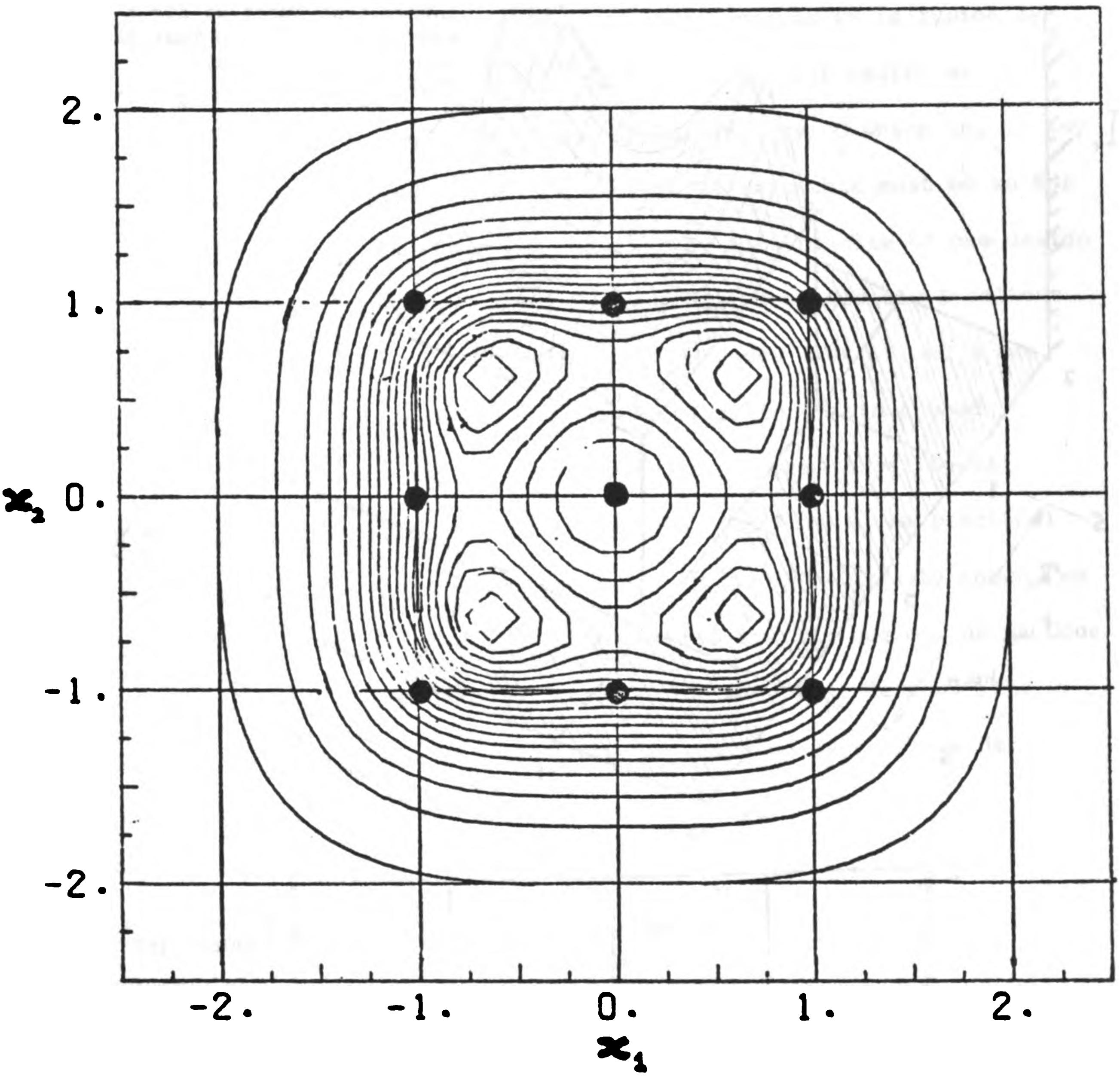


Figure 3(b) Information contours for a 3^2 factorial used as a second order design.

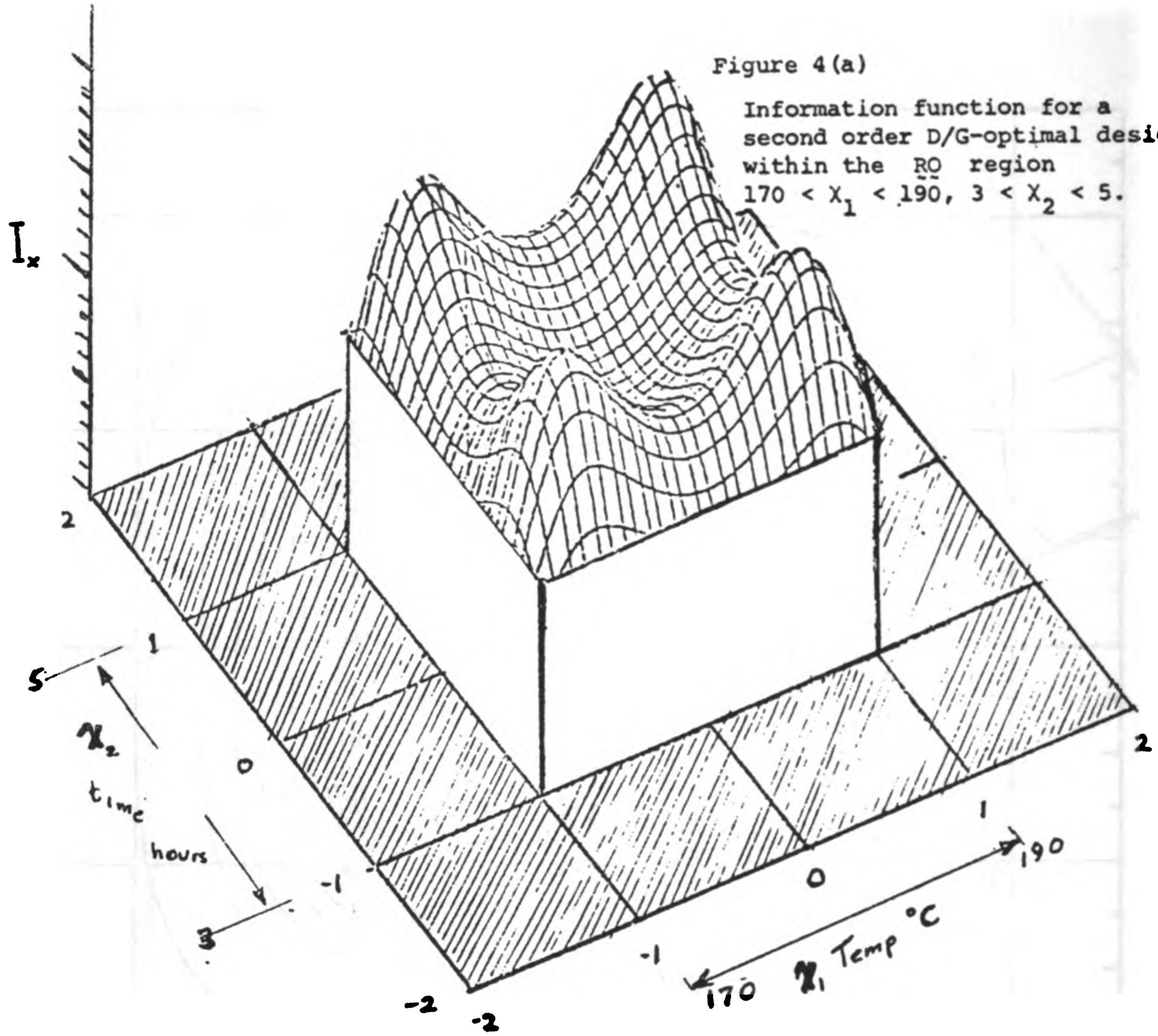


Figure 4(a)

Information function for a second order D/G-optimal design within the RO region $170 < X_1 < 190$, $3 < X_2 < 5$.

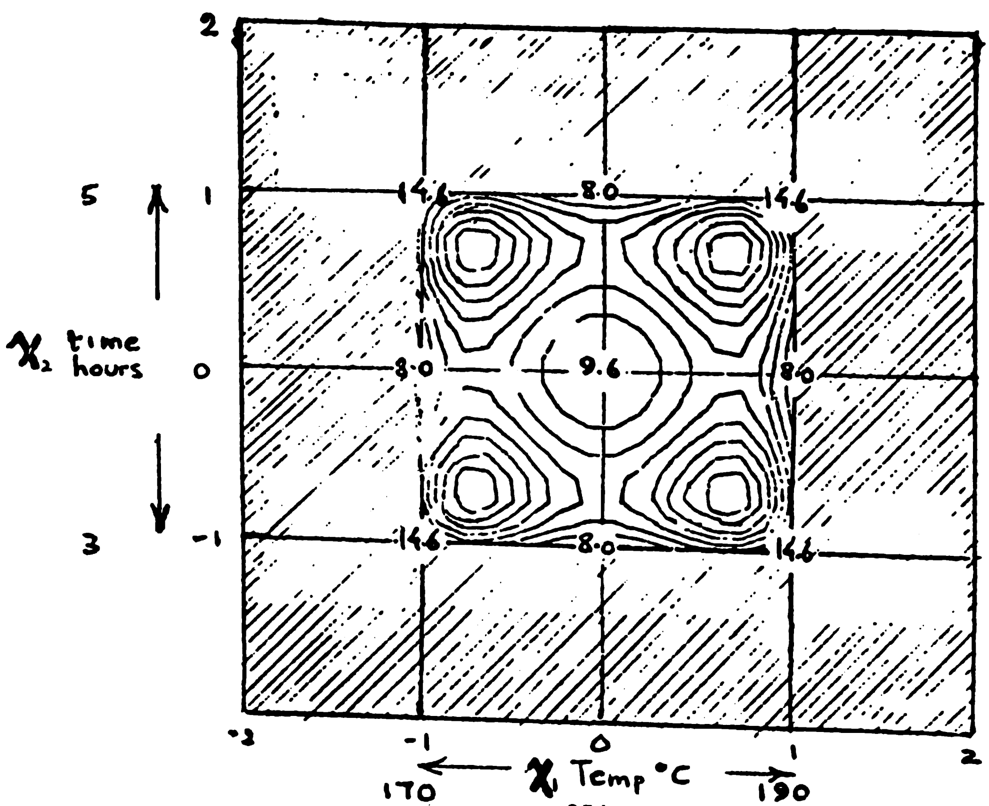


Figure 4(b)

Information contour plot and % distribution of the design measure.

Formulation in terms of the RO region

As has been pointed out, in response surface studies it is typically true that at any given stage of an investigation the current region of interest $\underset{\sim}{R}$ is much smaller than the region of operability $\underset{\sim}{O}$ which is, in any case, usually unknown. In particular, it is obvious that this must be so for any investigation in which we allow the possibility that results of one design may allow progress to a different unexplored region. Consequently I believe that formulation in terms of an $\underset{\sim}{RO}$ region which assumes that $\underset{\sim}{R}$ and $\underset{\sim}{O}$ are identical is artificial and limiting. In particular, to obtain a good approximation <u>within $\underset{\sim}{R}$</u> one may very well wish to put some experimental points <u>outside $\underset{\sim}{R}$</u> and so long as they are within $\underset{\sim}{O}$ there is no practical reason why we should not. Also since typically $\underset{\sim}{R}$ is only vaguely known, we will want to consider the information function over a wider region, as is done for example in Figure 5 for Fedorov's second order D-optimal design. The information function for this design may now be compared over this wider region with that for the 3^2 factorial in Figure 3.

Distribution of information over a wider region

In the response surface context, the coefficients $\underset{\sim}{\beta}$ of a graduating function $\eta_{\underset{\sim}{x}} \doteq \underset{\sim}{x}'\underset{\sim}{\beta}$ acting as they do merely as adjustments to a kind of mathematical french curve are not usually of individual interest except insofar as they affect η, in which case only the G-optimality criterion among those considered is of direct interest. For response surface studies however, it is far from clear how desirable is the property of G-optimality itself.

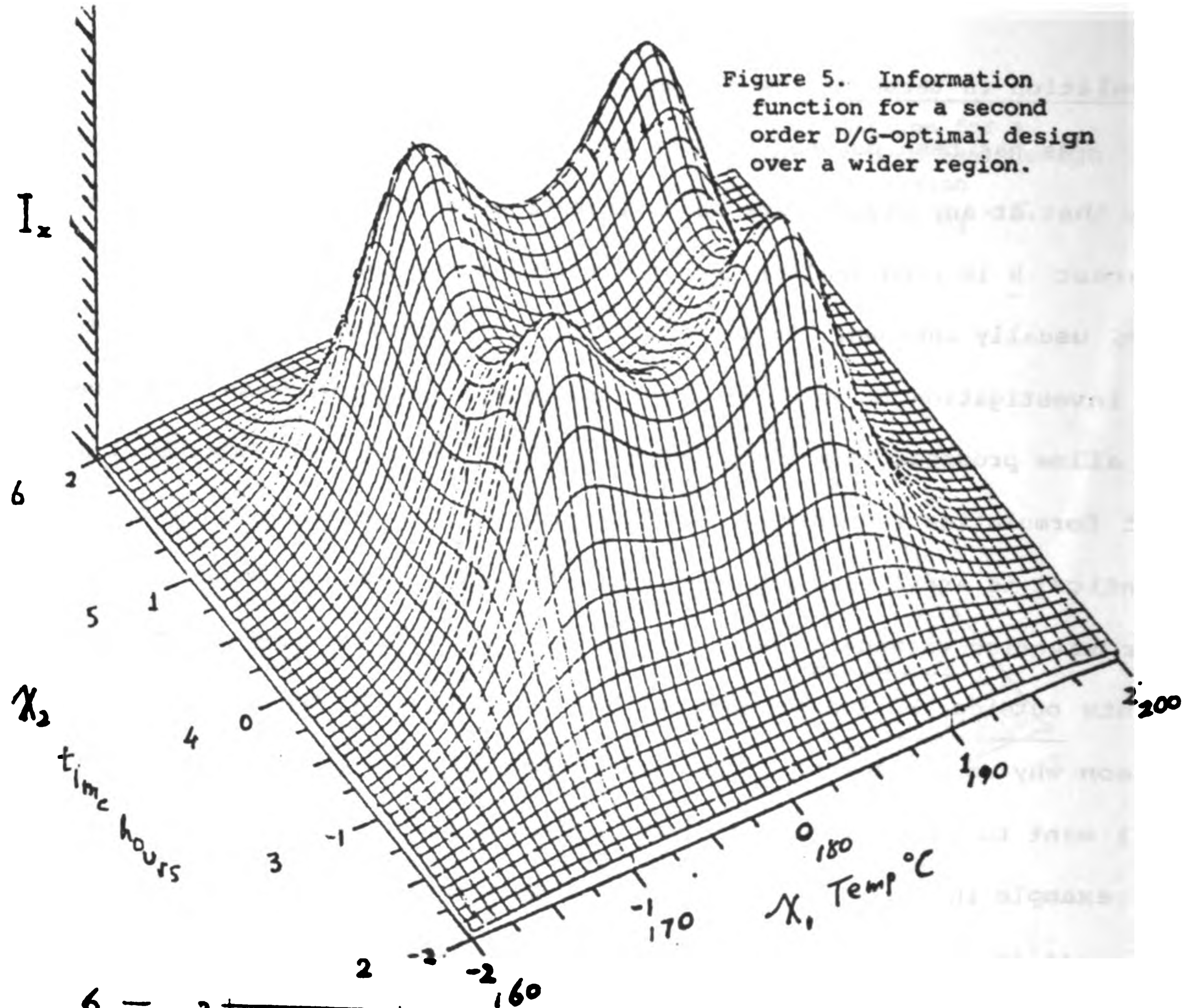
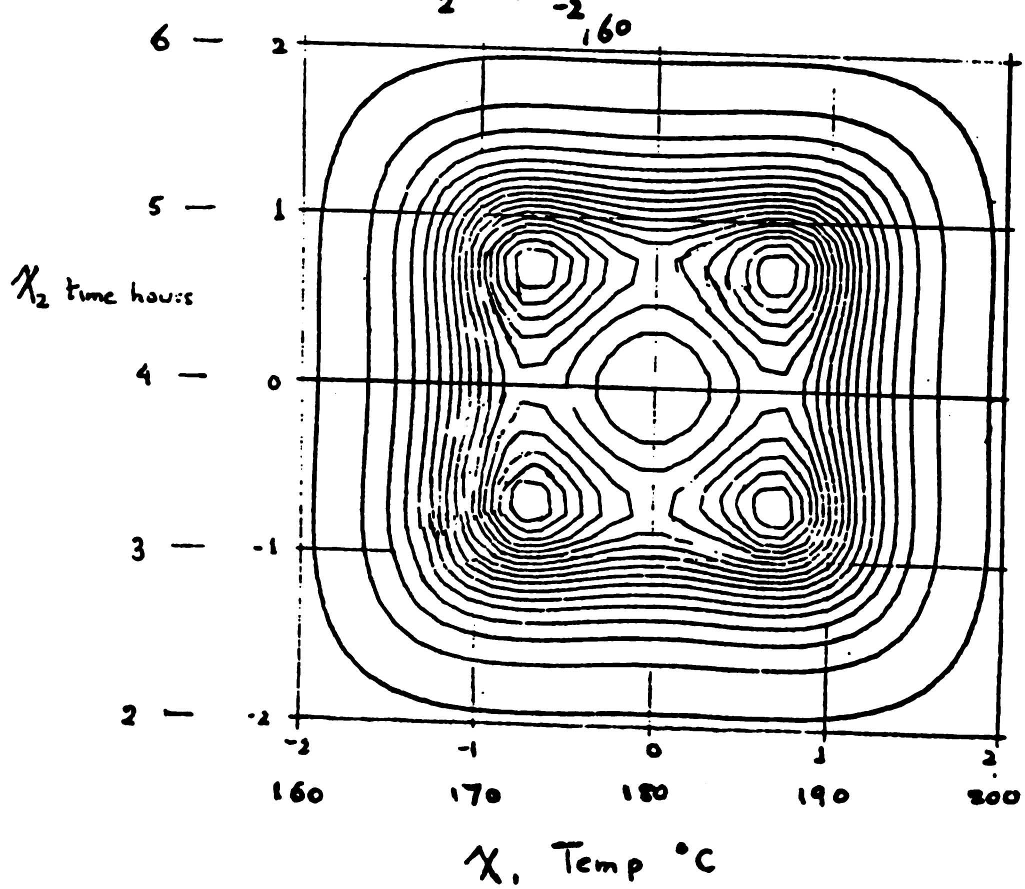


Figure 5. Information function for a second order D/G-optimal design over a wider region.



For instance, the profiles of Figure 6 made by taking sections of the surfaces of Figure 3 and Figure 5 suggest that neither the G/D-optimal design nor the 3^2 design are universally superior one to the other. In some subregions one design is slightly better, and in others the other design is slightly better. Both information functions, and particularly that of the G/D-optimal design, show a tendency to sag in the middle. This happens for the G/D-optimal design because the G-optimality characteristic guarantees that (maximized) minima for $I_{\underset{\sim}{x}}$, each equal to 1/p, occur at every design point, which must include the center point. However, this sagging information pattern of the second order design is not of course a characteristic of the first order design of Figure 2 which is also D/G-optimal but contains no center point. If the idea of the desirability of designs possessing a particular kind of information profile is basic, then it seems unsatisfactory that the nature of that profile should depend so very much on the order of the design. Indeed, the relevance of the minimax criterion which produces G-optimality is arguable. It follows from the Kiefer-Wolfowitz theorem that a second order design for the (±1,±1) region whose information function did not sag in the middle would necessarily not be D-optimal. But as we have seen, D-optimality is only one of many single-valued criteria that might be used in attempts to describe some important characteristic of the $\underset{\sim}{X}'\underset{\sim}{X}$ matrix. Others for example would be A-optimality and E-optimality, and these would yield different information profiles. But I would argue that since the information function itself is the most direct measure of desirability so far as the single issue of variance properties is concerned, our best course is to choose our design directly by picking a suitable information function, and not indirectly by finding some extremum for A, E, D, or other arbitrary criterion.

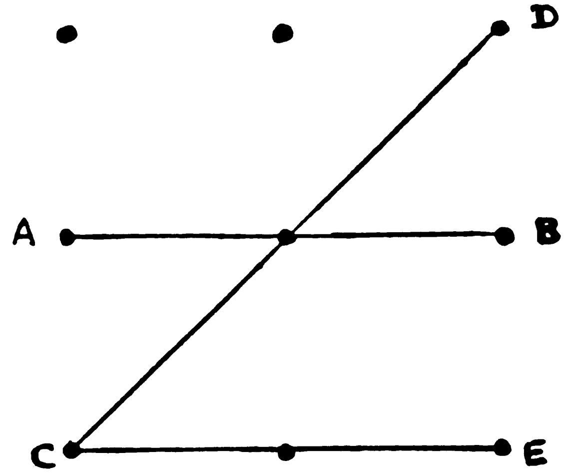
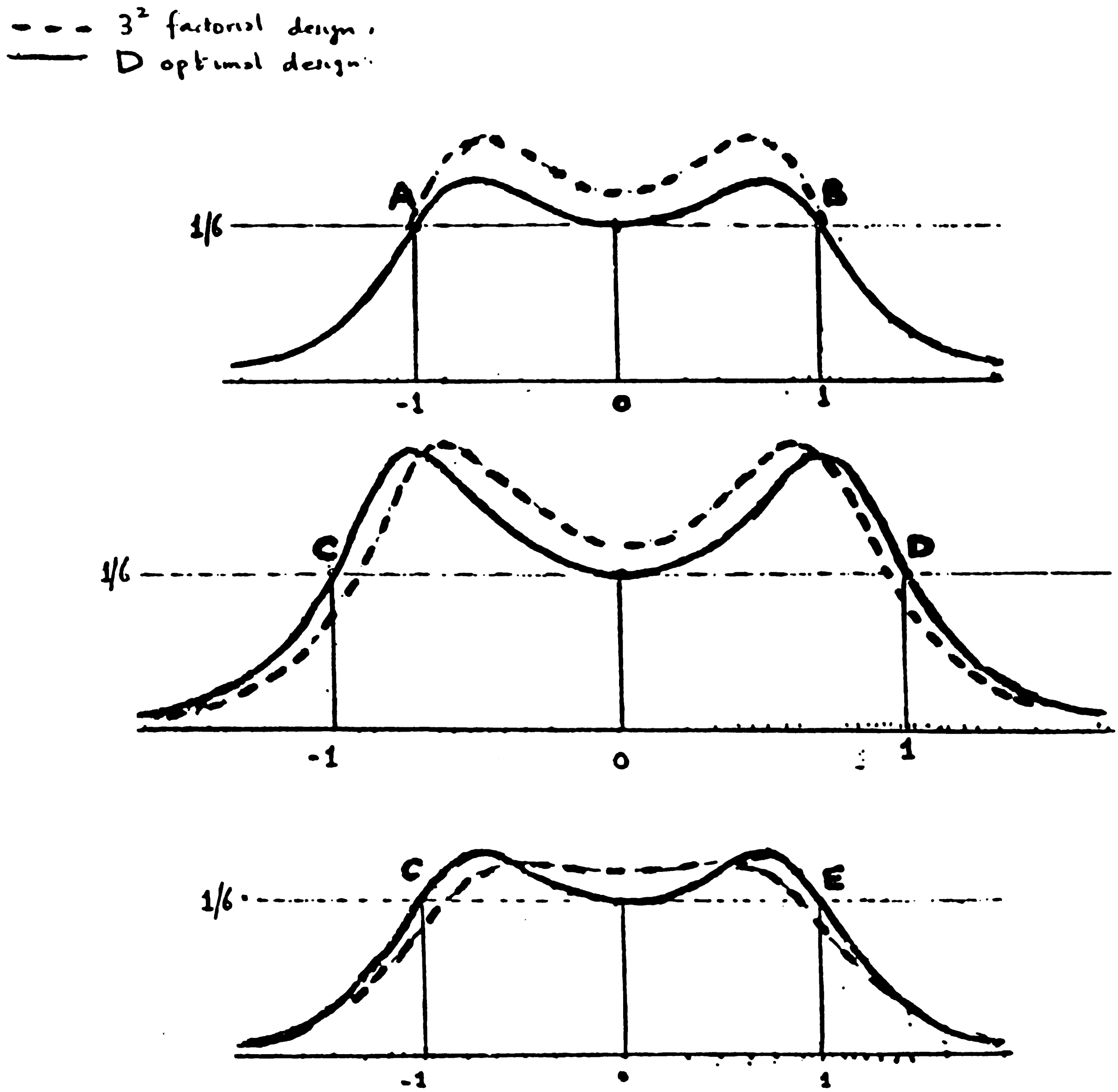


Figure 6. Profiles of $I_{\underset{\sim}{x}}$ for the second order D/G-optimal design and for the 3^2 factorial design.



Sensitivity of criteria to size and shape

In the process of scientific investigation, the investigator and the statistician must do a great deal of guesswork. In matching the region of interest $\tilde{R}$ and the degree of complexity of the approximating function, they must try to take into account, for example, that a more flexible second degree approximating polynomial can be expected to be adequate over a larger region $\tilde{R}$ than a first degree approximation. Obviously different experimenters would have different ideas of appropriate locations and ranges for experimental variables. In particular, ranges could easily differ from one experimenter to another by a factor of two or more[5]. In view of this, extreme sensitivity of design criteria to scaling is disturbing[6]. For example, suppose each dimension of a dth order experimental design is increased[7] by a factor c. Then the D criterion is increased by a factor of c^q where

[5] Over a sequence of designs, initial bad choices of scale and location would tend to be corrected, of course.

[6] In particular, designs can only be fairly compared if they are first scaled to be of the "same size." But how is size to be measured? It was suggested in [14] that designs should be judged as being of the same size when their marginal second moments $\Sigma(x_{iu} - \bar{x}_u)^2/n$ were identical. This convention is not entirely satisfactory, but will of course give very different results from those which assume design points to be all included in the same region RO. It is important to be aware that the apparent superiority of one design over another will often disappear if the method of scaling the design is changed. In particular this applies to comparisons such as those made by Nalimov *et al* [40] and Lucas [38].

[7] A measure of efficiency of a design criterion (see for example [3], [17]) is motivated by considering the ratio of the number of runs necessary to achieve the optimal design to the number of runs required for the suboptimal design to obtain the same value of the criterion (supposing fractional numbers of runs to be allowed). In particular for the D criterion, this measure of D-efficiency is $(D/D_{opt})^{1/k}$. Equivalently here, to illustrate scale sensitivity, we concentrate attention on the factor c by which each scale would need to be inflated to achieve the same value of the D criterion.

$$q = \frac{2k(k+d)!}{(k+1)!(d-1)!}$$

Equivalently a confidence region of the same volume as that for a D-optimal design can be achieved for a design of given D value by increasing the scale for each variable by a factor of $c = (D_{opt}/D)^{1/q}$, thus increasing the volume occupied by the design in the $\underset{\sim}{x}$ space, by a factor $c^k = (D_{opt}/D)^{k/q}$. For example the D value for the 3^2 factorial design of Figure 3 is 0.98×10^{-2} as compared with a D value of 1.14×10^{-2} for the D-optimal design. For (k = 2, d = 2), we find q = 16, and $c = (1.14/0.98)^{1/16} = 1.009$. Thus the same value of D (the same volume of a confidence region for the β's) as is obtained for the D-optimal design would be obtained from a 3^2 design if each side of the square region were increased by less than 1%. Equivalently, the area of the region would be increased by less than 2%. Using the scaling that was used in Figure 4 for illustration, we should have to change the temperature by 20.18 °C instead of 20 °C, and the time by two hours and one minute instead of two hours, for the 3^2 factorial to give the same D value as the D/G-optimal design. Obviously no experimenter can guess to anything approaching this accuracy what are suitable ranges over which to vary these factors.

Obviously choice of region and choice of information function are closely interlinked. For example, <u>any</u> set of N = k+1 points in k-space which have no coplanarities is obviously a D-optimal first order design for <u>some</u>[8] ellipsoidal region. Furthermore the information function for a design of order d is a smooth function whose harmonic average over the n experimental points (which can presumably be regarded as representative of the region of

[8] Namely for that region enclosed within the information contour $I_{\underset{\sim}{x}} = 1/p$ which must pass through all the k+1 experimental points.

interest) is always 1/p wherever we place the points. Thus the problem of design is not so much a question of choosing the design to increase total information as spreading the total information around in the manner desired.

Rotatable Designs

A route for simplication different from alphabetic optimality occurs when, after suitable transformation of the inputs $\underset{\sim}{x}$ to standardized variables $\underset{\sim}{x}$ nothing is known about the orientation in the $\underset{\sim}{x}$ space of the response surface we wish to study. It was argued by Box and Hunter [11] that we should then employ designs having the property that the variance of $\hat{y}$ is a function only of $\rho = (\underset{\sim}{x}'\underset{\sim}{x})^{1/2}$ so that

$$V_{\underset{\sim}{x}} = V_\rho \quad \text{and} \quad I_{\underset{\sim}{x}} = I_\rho \ .$$

For a first order design, rotatability implies orthogonality and vice versa, and completely decides the information function. For second and higher order designs, a requirement of rotatability fixes many moment properties of the design, but V_ρ and hence I_ρ are still to some extent at our choice, and can be changed by changing certain moment ratios [11]. In particular, for a second order design, V_ρ depends on the single moment ratio $\lambda = (n/3)\Sigma x_i^4/(\Sigma x_i^2)^2$. For illustration, Figure 7 shows the information function for a second order rotatable design with $\lambda = .75$ consisting of 8 points arranged in a regular octagon with 4 points at the center.

The truth seems to be that at any particular phase of an investigation the scientific decision that most contributes to the outcome of that phase is the choice of the current region of interest (involving choice of variables, locations, ranges, and transformations) -- this is a choice that does not really involve statistics. After this decision is made, (and given the assumption that the model fits perfectly so that only the variance properties

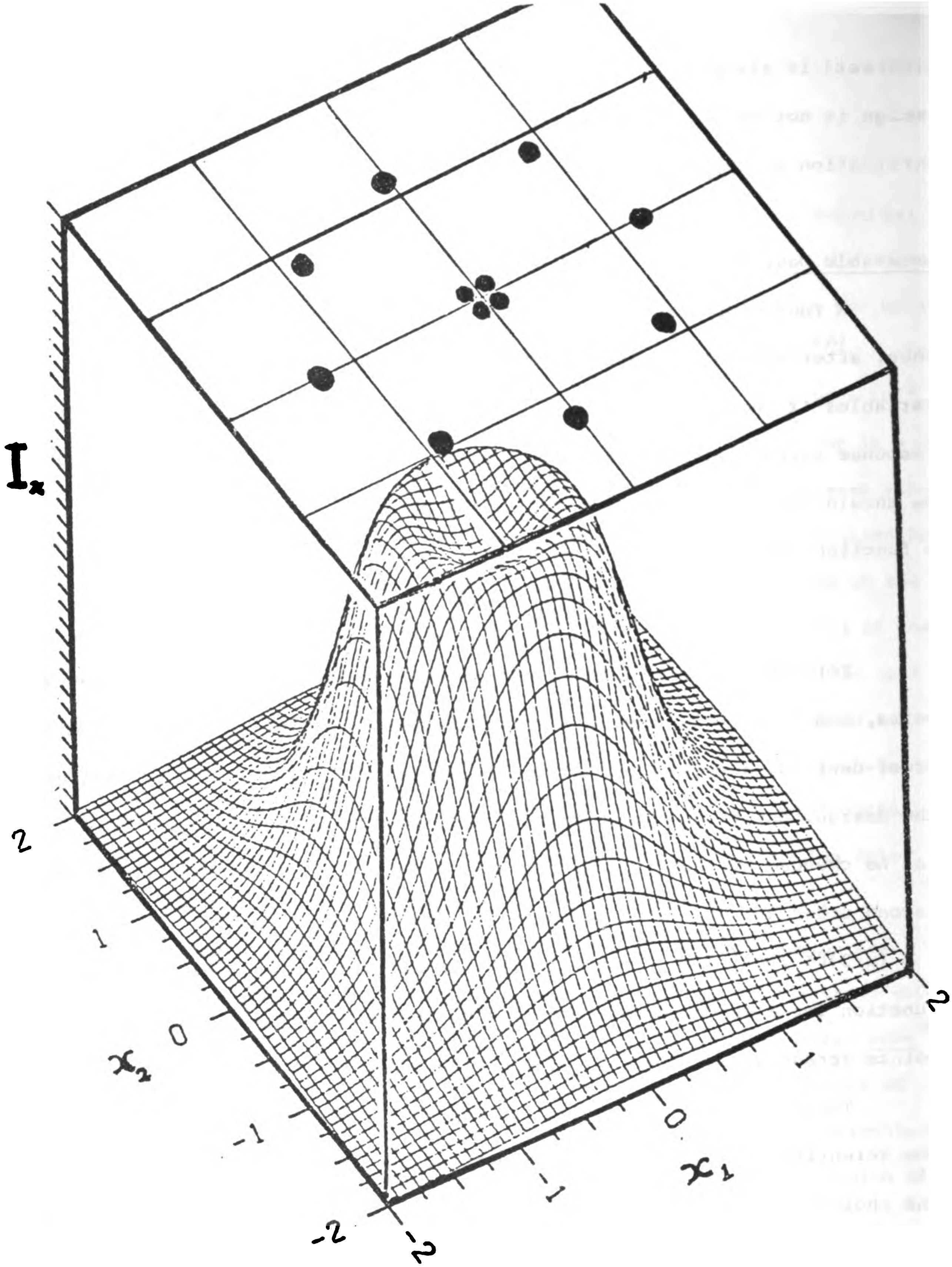


Figure 7(a) **Information function for a second order rotatable design consisting of 8 points on a circle plus 4 center points.**

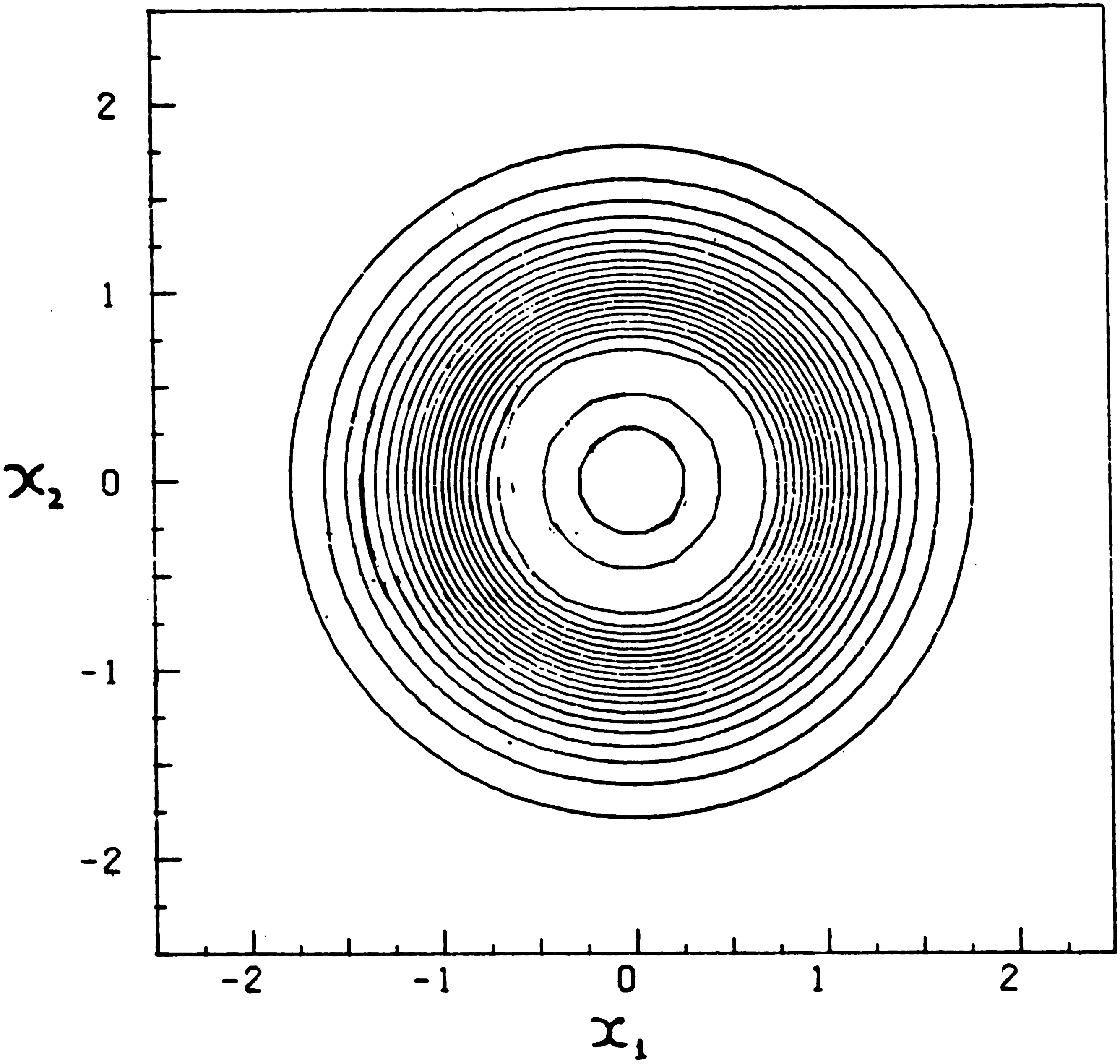


Figure 7(b) Information contours for the same design.

of the design are of interest) any set of experiments that cover this region in some reasonably uniform way is likely to do quite well. I cannot see that the various optimality criteria are particularly relevant to this choice, although there would certainly be no harm in considering them, together with many other factors briefly discussed later.

Ignoring of bias

All models are wrong; some models are useful. This aphorism is particularly true for empirical functions such as polynomials that make no claim to do more than locally graduate the true function. For chemical examples some idea of the adequacy of such approximations can be gained by studying surfaces produced by chemical kinetic models. An example[8] taken from [10] is shown in Figure 8. See also [15].

One conclusion I reached from many such studies was that approximations would not need to be very good for response surface methods to work. Thus within region A of Figure 8 the locally monotonic function could be crudely approximated by a plane which could indicate a useful path of ascent. Also valuable information might be obtained about a ridge such as that in region B, even though the underlying surface was not exactly quadratic. Notice however

---

[8] This surface was generated (see [10] for details) by considering the yield of the product B in a consecutive reaction $A \xrightarrow{k_1} B \xrightarrow{k_2} C$ following first order kinetics with temperature sensitivity given by the Arrhenius relation $\ln k_i = \ln \alpha_i + \beta_i/T$, where temperature T is measured in degrees Kelvin, using plausible values for the constants $\alpha_1, \alpha_2, \beta_1, \beta_2$.

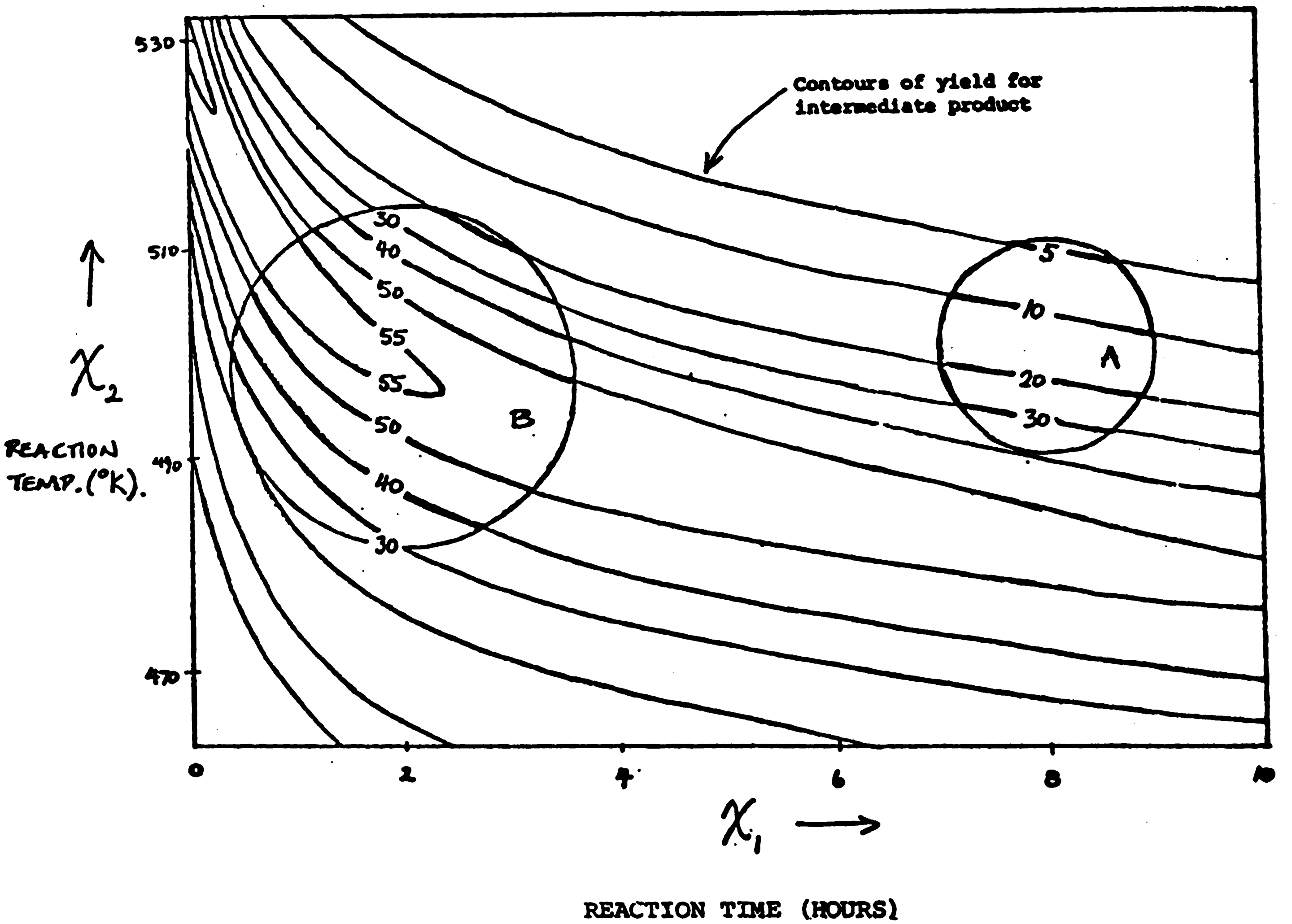


Figure 8. Contours of a theoretical response surface in reaction time and reaction temperature for a first order consecutive reaction, with plausible values substituted for kinetic constants.

that in the light of such examples any theory of experimental design which depended on the exactness of such approximations should be regarded with some skepticism.

5. TAKING ACCOUNT OF BIAS

If $\hat{y} = \underset{\sim}{x}'\hat{\underset{\sim}{\beta}}$ is the fitted value using the empirical approximation, then its total error ε is

$$\hat{y} - \eta = \{\hat{y} - E(\hat{y})\} + \{E(\hat{y}) - \eta\}$$

$$\varepsilon \quad = \quad \varepsilon_V \quad + \quad \varepsilon_B \ .$$

Thus the error ε contains a random part ε_V and a systematic, or bias, part ε_B, and we must expect that ε_B will <u>not</u> be negligible. Since all the theory previously discussed makes the assumption that ε_B is zero, we must consider whether the resulting designs are robust to this kind of discrepancy. The optimality criteria discussed earlier which assume the response function to be exact usually produce a substantial proportion of experimental points on the boundary of $\underset{\sim}{R}\underset{\sim}{O}$. In the context of possible bias, this is not reassuring, since it is at these points that the approximating function will be most strained.

The explicit recognition that bias will certainly be present does however seem to provide a more rational means for approaching the scaling problem ([6], [7]). To see this, consider again the formulation given earlier in terms of a region of interest $\underset{\sim}{R}$ and a larger region $\underset{\sim}{O}$ of operability. If we were to assume (unrealistically) that the approximation remained exact however widely the points were spread, and if some measure of variance reduction were the only consideration, then to obtain most accurate estimation within $\underset{\sim}{R}$, the size of the design would have to be increased to the boundaries of the operability region $\underset{\sim}{O}$. But in fact of course the wider the points were spread, the less applicable would be the approximating function, and the bigger the bias error. This suggests that we should seek restriction of the spread of the experimental points not by artificial limitation to some region $\underset{\sim}{R}\underset{\sim}{O}$, but by balancing off the competing requirements of variance on the one hand, which is reduced as the spread of the points is increased, and bias on the other hand, which is increased as the spread of the points is increased.

The mean square error associated with estimating $\eta_{\underset{\sim}{x}}$ by $\hat{y}_{\underset{\sim}{x}}$ standardized for the number, n, of design points and the error variance σ^2, can be written as the sum of a variance component and a squared bias component

$$n \cdot E(\hat{y}_{\underset{\sim}{x}} - \eta_{\underset{\sim}{x}})^2/\sigma^2 = n \cdot V(\hat{y}_{\underset{\sim}{x}})/\sigma^2 + n\{E(\hat{y}_{\underset{\sim}{x}}) - \eta_{\underset{\sim}{x}}\}^2/\sigma^2 ,$$

or

$$M_{\underset{\sim}{x}} = V_{\underset{\sim}{x}} + B_{\underset{\sim}{x}} .$$

For illustration, an example is taken from a forthcoming book with N.R. Draper and J.S. Hunter [10]. Figure 9 shows a situation as it might exist for a single variable when a straight line approximating function is to be used. The diagram shows what might be the true underlying function which would of course be obscured by experimental error. Suppose the region of interest $\underset{\sim}{R}$ is scaled so that $-x_0 < x < x_0$ and in particular consider the two designs

(a) (-2/3, 0, 2/3) and (b) (-4/3, 0, 4/3).

One way [6] to obtain overall measures of variance and squared bias over any specified region of interest $\underset{\sim}{R}$ is by averaging $V_{\underset{\sim}{x}}$ and $B_{\underset{\sim}{x}}$ over $\underset{\sim}{R}$ to provide the quantities

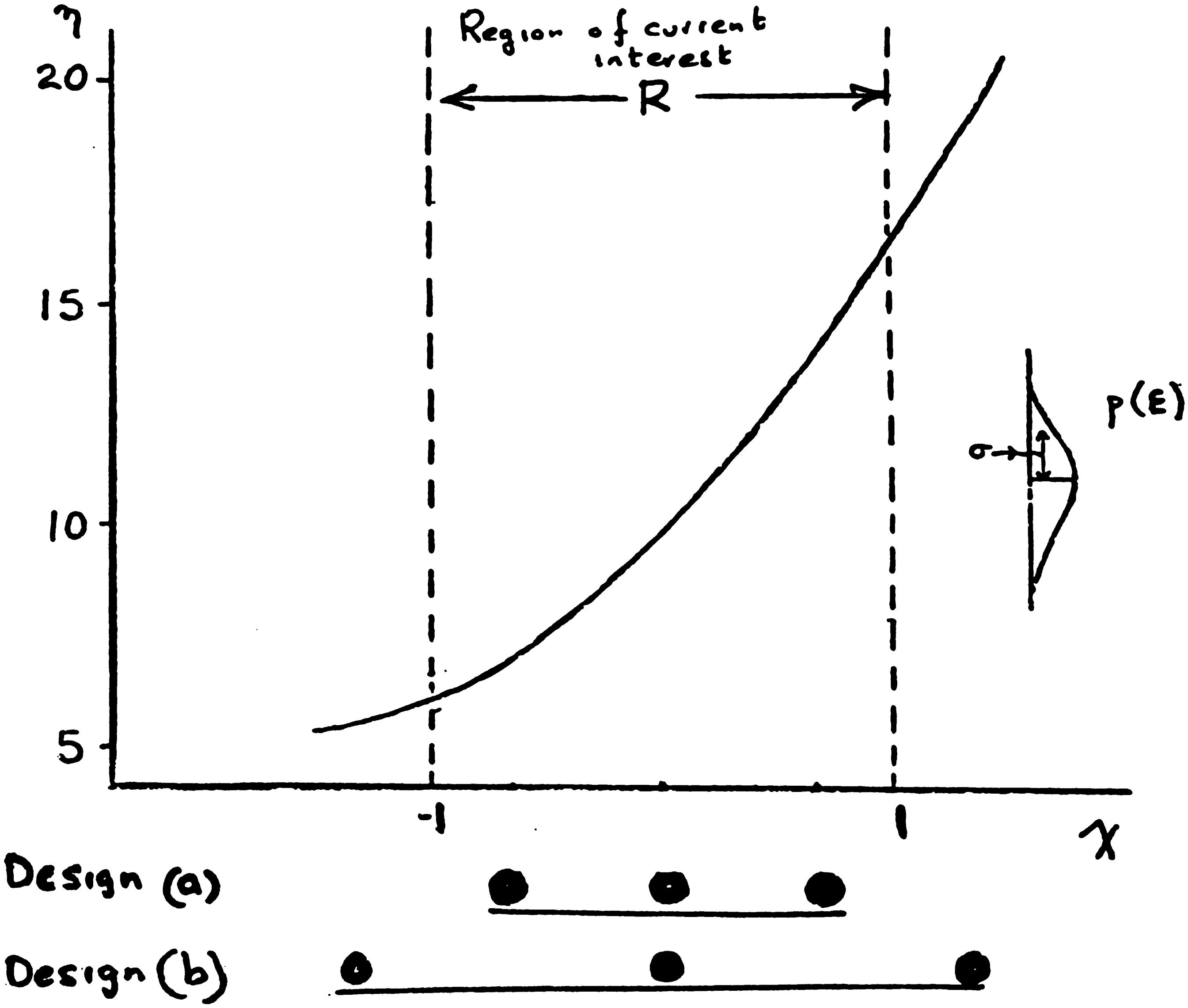


Figure 9. Two possible designs for fitting a straight line over a region of interest $\underset{\sim}{R}$ $-1 < x < 1$.

$$V = \int_{\underset{\sim}{R}} V_{\underset{\sim}{x}} d\underset{\sim}{x} / \int_{\underset{\sim}{R}} d\underset{\sim}{x} \quad \text{and} \quad B = \int_{\underset{\sim}{R}} B_{\underset{\sim}{x}} d\underset{\sim}{x} / \int_{\underset{\sim}{R}} d\underset{\sim}{x} .$$

Denoting the integrated (over $\underset{\sim}{R}$) mean square error by M, we can then write

$$M = V + B.$$

For the previous example, V, B, and M are plotted against x_0 in Figure 10. We see how V becomes very large if the spread of the design is made very small, while if the design is made very large, V slowly approaches its minimum value of unity. The average squared bias B, on the other hand, has a minimum value when x_0 is about 0.7, and increases for larger or smaller designs. A rather flat minimum for M = V + B occurs near $x_0 = 0.79$. Thus in this manner the design which minimizes average mean squared error M is not very different from the design which minimizes average squared bias B, but <u>is</u> extremely different from that which minimizes average variance V.

<u>Choice of alternative model</u>

A difficulty in all this is that in practice we do not know the nature of the true function $\eta_{\underset{\sim}{x}}$. Progress may be made however by supposing that $\eta_{\underset{\sim}{x}}$ is t. some satisfactory approximation represented by a polynomial model of higher degree d_2. Suppose then that a polynomial model of degree d_1 is fitted to n data values to give

$$\hat{\underset{\sim}{y}}_{\underset{\sim}{x}} = \underset{\sim}{X}_1 \underset{\sim}{b}_1$$

while the true model is in fact a polynomial of degree d_2, so that

$$\underset{\sim}{\eta}_{\underset{\sim}{x}} = \underset{\sim}{X}_1 \underset{\sim}{\beta}_1 + \underset{\sim}{X}_2 \underset{\sim}{\beta}_2$$

We also need to know something about the <u>relative</u> magnitudes of systematic and random errors that we could expect to meet in practical cases. It was argued in [6] that an investigator might typically employ a

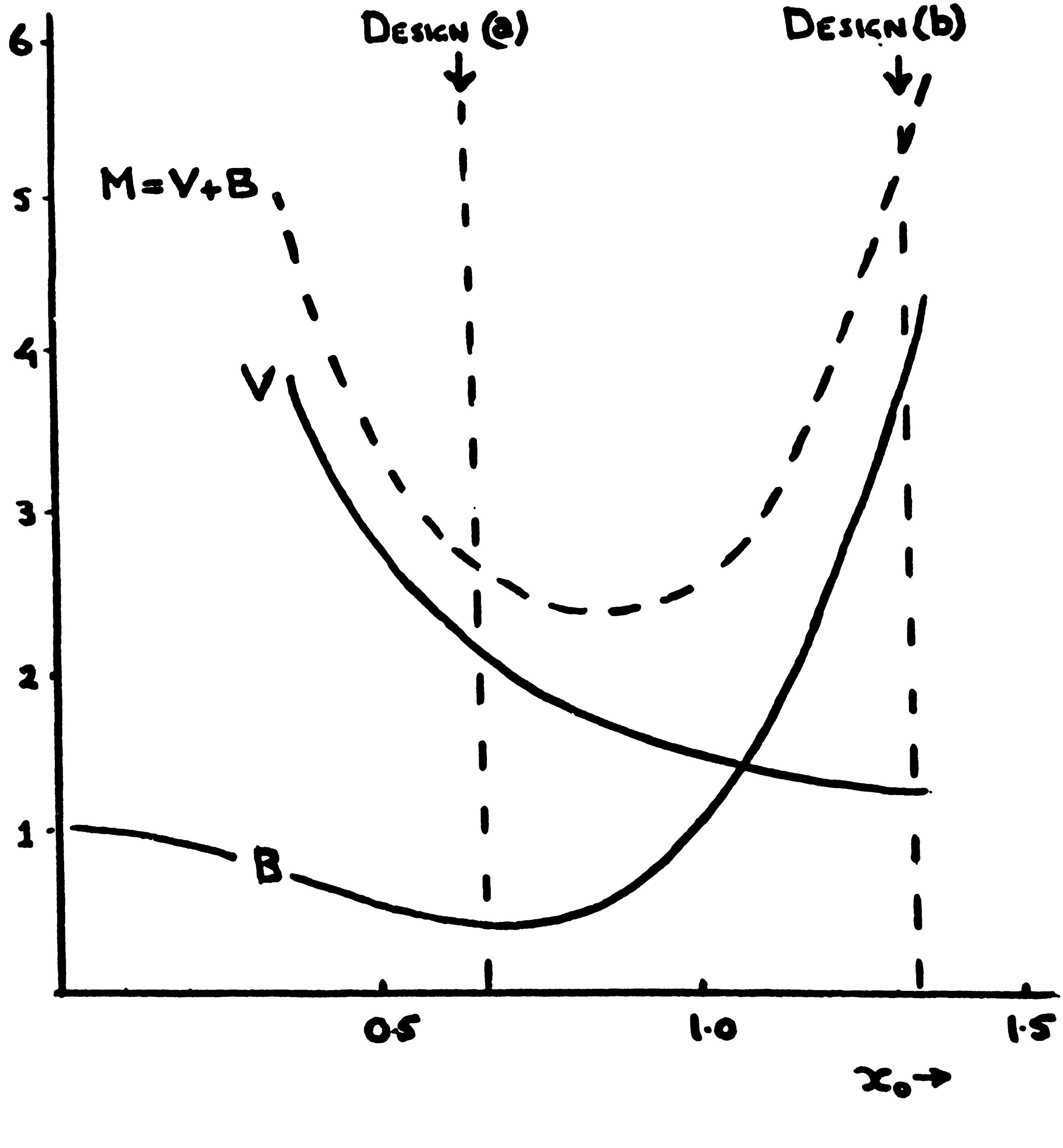


Figure 10. The behavior of integrated variance V, integrated squared bias B, and their total M = V + B, for three-point designs $(-x_0, 0, x_0 0)$.

fitted approximating function such as a straight line when he believed that the average departure from the truth induced by this approximating function was no worse than that induced by the process of fitting. This would suggest that the experimenter would tend to choose the size of his region $\underset{\sim}{R}$, and the degree of his approximating function in such a way that the integrated random error and the integrated systematic error were about equal. Thus we might suppose that a situation of particular interest is that where B is roughly equal to V. Examples that we studied seemed to show that designs that minimized M with the constraint V = B were close to those which minimized B. Consequently we suggested that, if a simplification were to be made in the design problem, it might almost be better to ignore the effects of sampling variation rather than those of bias.

However this may be, there seems no doubt that, in making a table of useful designs, a component in our thinking should be the characteristics of the designs which minimized squared bias against feared alternatives. As a factor in our final choice, this should certainly receive as much attention as the indications supplied by, say, D-optimality.

For illustration particular examples of designs in three dimensions which minimize integrated squared bias when $\underset{\sim}{R}$ is a sphere of unit radius are shown in Figure 11(a) for $d_1 = 1$ and $d_2 = 2$ (a first order design robust to second order effects) and in Figure 11(b) for $d_1 = 2$ and $d_2 = 3$ (a second order design robust to third order effects). The former is the familiar 2^3 factorial scaled so that the points are 0.71 units from the center. The latter is a rotatable composite design with "cube" points at a distance 0.86 from the center, and "star" points at a distance 0.83 from the center.

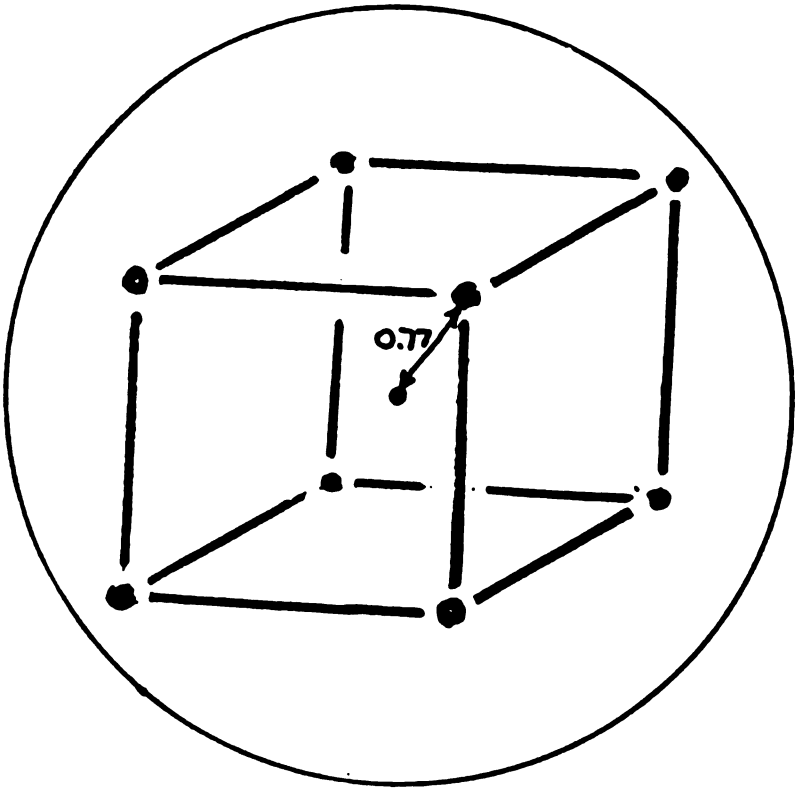


Figure 11(a) A first order (two-level factorial) design in three factors which minimizes squared bias from second order terms when the region of interest is a sphere of unit radius.

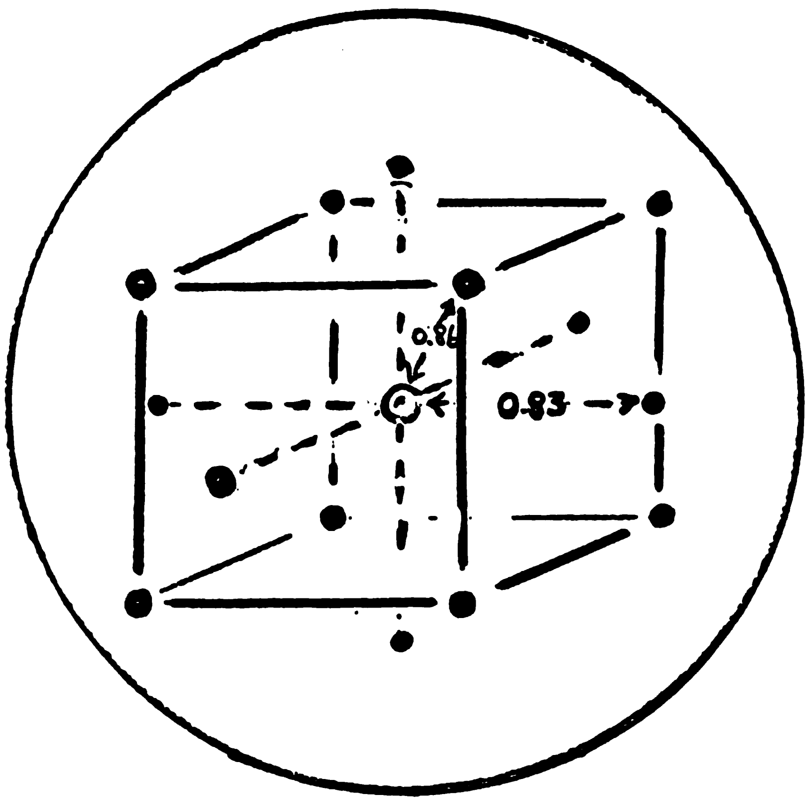


Figure 11(b) A second order composite rotatable design which minimizes squared bias from third order terms when the weight function is uniform over a spherical region of interest of unit radius.

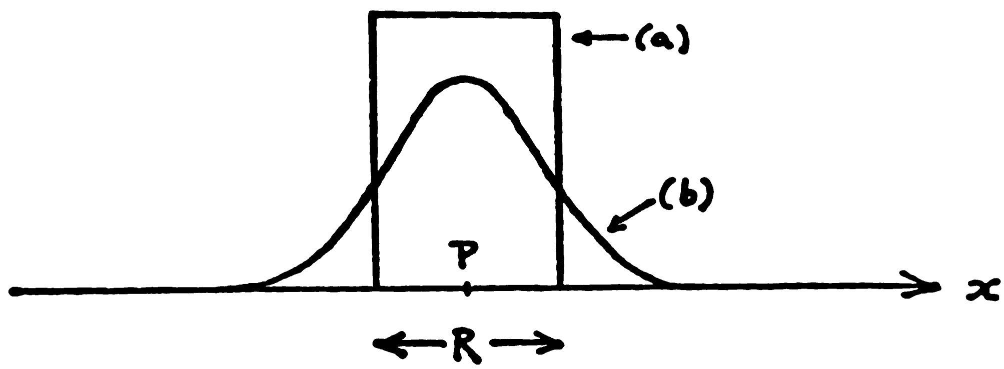


Figure 12: Two possible weight functions for $k = 1$:
(a) "Uniform over R" type indicating uniform interest over R, no interest outside R;
(b) Normal Distribution shape, giving greater weight to points nearer P.

Obviously in practice because of the inevitable inexactness of choosing scales exact dimensions of the designs should not be taken too seriously, but these examples illustrate the fact that as soon as we take account of bias, design points are not chosen on the boundary of R.

Choice of designs which minimize bias

Before considering the problem of choosing minimum bias designs it is desirable to generalize slightly the previous formulation. Although it avoids limiting the location of the design points in an artificial way the idea of a region of interest $\underset{\sim}{R}$ within a larger operability region $\underset{\sim}{O}$ is still not entirely satisfactory because it implies that we have equal interest at all points within $\underset{\sim}{R}$. A more general formulation [7] which subsumes that we have been discussing employs a weight function $w(\underset{\sim}{x})$ which extends over the operability region $\underset{\sim}{O}$ so that $\int_{\underset{\sim}{O}} w(\underset{\sim}{x})d\underset{\sim}{x} = 1$. The weighted mean square error M can now be split into a weighted variance part V and a weighted squared bias part B so that again $M = V + B$, with

$$M = \int_{\underset{\sim}{O}} w(\underset{\sim}{x})E\{\hat{y}(\underset{\sim}{x}) - \eta(\underset{\sim}{x})\}^2 d\underset{\sim}{x}$$

$$V = \int_{\underset{\sim}{O}} w(\underset{\sim}{x})E\{\hat{y}(\underset{\sim}{x}) - E(\hat{y}(\underset{\sim}{x}))\}^2 d\underset{\sim}{x}$$

$$B = \int_{\underset{\sim}{O}} w(\underset{\sim}{x})\{E(\hat{y}(\underset{\sim}{x}) - \eta(\underset{\sim}{x})\}^2 d\underset{\sim}{x} .$$

Two possible weight functions for $k = 1$ [20] are shown in Figure 12.

Suppose as before the fitted function is a polynomial $\underset{\sim}{X}_1\underset{\sim}{b}_1$ of degree d_1 while the true model is a polynomial $\underset{\sim}{X}_1\underset{\sim}{\beta}_1 + \underset{\sim}{X}_2\underset{\sim}{\beta}_2$ of degree d_2 and define moment matrices for the design and for the weight function by

$$M_{11} = n^{-1}X_1'X_1, \qquad M_{12} = n^{-1}X_1'X_2$$

$$\mu_{11} = \int_O w(x)x_1x_1'dx, \qquad \mu_{12} = \int_O w(x)x_1x_2'dx.$$

Then [6] a necessary and sufficient condition for the squared bias B to be minimized is that

$$M_{11}^{-1}M_{12} = \mu_{11}^{-1}\mu_{12}$$

and hence a sufficient condition is that all the moments of the design up to and including order $d_1 + d_2$, are equal to all the corresponding moments of the weight function.

6. SOME OTHER CONSIDERATIONS IN DESIGN CHOICE

There is insufficient space to discuss here all of the items in Table 1 that, in one circumstance or another, it might be necessary to take into account, but mention will be made of a few.

Lack of Fit (iii), Sequential Assembly (vi), Blocking (v), Estimation of Error (vii), Transformation Estimation (iv)

While the adequacy of a particular approximating function to explore a region of current interest is always to some extent a matter of guesswork, simple approximations requiring fewer runs for their elucidation will usually be preferred to more complicated ones. This leads to a strategy of building up from simpler models, rather than down from more complicated ones. A practical procedure is then: to employ the simplest approximating function which it is hoped may be adequate; to allow for checking its adequacy of fit (see also [1], [2], [6], and [19]); to switch to a more elaborate approximating function when this appears necessary. The implication for designs is (a) that they should provide for checking model adequacy, (b) that they should be capable of sequential assembly -- a design of order d should be augmentable to one of order d + 1, (c) since conditions may change slightly from one set of runs to another, especially affecting level, the pieces of the design should form orthogonal blocks.

For illustration, Figure 13 shows the sequential assembly of a design arranged in three orthogonal blocks, each of six runs, labeled I, II, and III. Block I is a first order design but also provides a check for overall curvature (obtained by contrasting the average response of the center points with the average response on the cube). A single contrast of the center response is available as a gross check on previous information about experimental error. If after analyzing the results from Block I there are doubts about the adequacy of a first degree polynomial model, Block II may be performed. It uses the complementary simplex, and the two parts together form a first order design (I+II) with much greater ability to detect lack of fit due to second order terms provided by additional orthogonal contrasts estimating the two-factor interactions. The addition of Block III produces a composite design (I+II+III) which allows a full second degree approximating equation to be fitted if this appears to be desirable. The complete design also provides orthogonal checking contrasts for lack of quadraticity in each of the three directions ([9], [12]). These contrasts can also be regarded as checking the need for transformation in each of the X's. Finally if it were

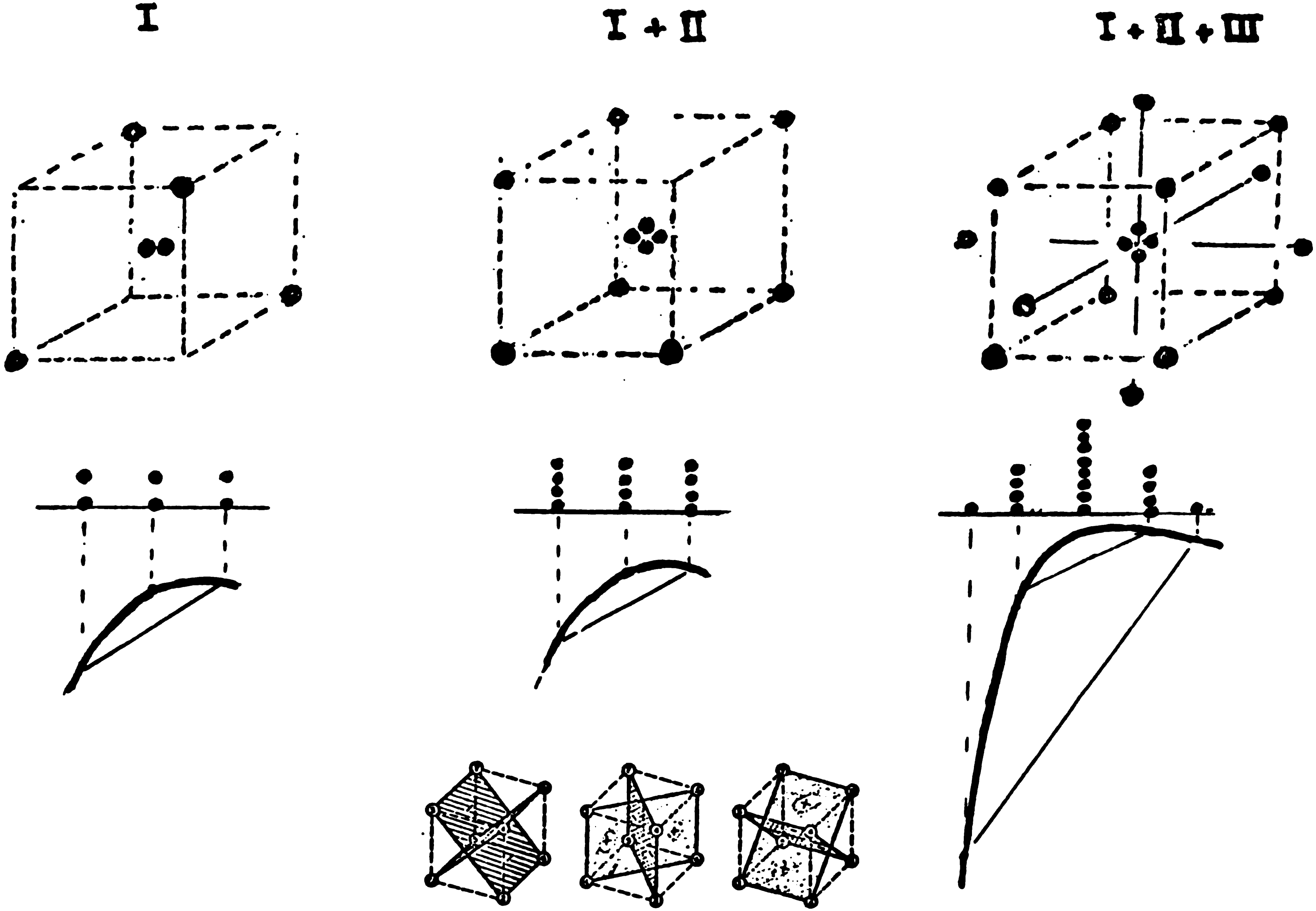


Figure 13. An example of sequential assembly, showing checks of linearity and quadraticity.

decided that more information about experimental error was desirable, the replication of the star in a further Block IV could furnish this, and also provide some increase in the robustness of the design to wild observations.

Robustness

Approaches to the robust design of experiments have been recently reviewed by Herzberg [28]; see also [29]. In particular, Box and Draper [8] suggested that the effects of wild observations could be minimized by making $r = \Sigma r_{uu}^2$ small, where $\underset{\sim}{R} = \{r_{tu}\} = \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'$. This is equivalent to minimizing $\Sigma r_{uu}^2 - p^2/n = \mathrm{Var}\{V(\hat{y})\}$ which takes the value zero when $V(\hat{y}_u) = p/n$ $(u = 1,2,\ldots,n)$. Thus G-optimal designs are optimally robust in this sense.

Size of the experimental design

A good experimental design is one which focuses experimental effort on what is judged important in the particular current experimental context. Suppose that, in addition to estimating the p parameters of the assumed model form, it is concluded that $f > 0$ contrasts are needed to check adequacy of fit, $b > 0$ further contrasts for blocking, and that an estimate of experimental error is needed having $e > 0$ degrees of freedom. To obtain independent estimates of all items of interest we then require a design containing at least $p + f + b + e$ runs. However the importance of checking fit, blocking, and obtaining an independent estimate of error will differ in different circumstances, and the minimum value of n will thus correspondingly differ. But this minimum design will in any case only be adequate if σ^2 is below some critical value. When σ^2 is larger designs larger than the minimum design will be needed to obtain estimates of sufficient precision. In this circumstance rather than merely replicate the minimum design, opportunity may be taken to employ a higher order design allowing the fitting of a more elaborate approximating function which can then cover a wider experimental region. Notice that even when σ^2 is small designs for which n is larger than p are not necessarily wasteful. This depends on whether the additional degrees of freedom are genuinely used to achieve the experimenter's current objectives.

Simple Data Patterns

It has sometimes been argued that we may as well choose points randomly to cover the "design region" or employ some algorithm that distributes them evenly even though this does not result in a simple data pattern such as is achieved by factorials and composite response surface designs. In favor of this idea it has been urged that the fitting of a function by least squares to a haphazard set of points is no longer a problem for modern computational devices. This is true, but overlooks an important attribute of designs which form simple patterns. The statistician's task as a member of a scientific team is a dual one, involving inductive criticism and deductive estimation. The latter involves deducing in the light of the data the consequences of given assumptions (estimating the fitted function), and this can certainly be done with haphazard designs. But the former involves the question (a) of what function should be fitted in the first place, and (b) of how to examine residuals from the fitted function in an attempt to understand deviations from the initial model, in particular in relation to the independent variables, and so to be led to appropriate model modification.

Designs such as factorials and composite response surface designs employ patterns of experimental points that allow many such comparisons to be made, both for the original observations and for the residuals from any fitted function. For example, consider a 3^2 factorial design used to elucidate the effects of temperature and concentration on some response such as yield. Intelligent inductive criticism is greatly enhanced by the possibility of being able to plot the original data and residuals against temperature for each individual level of concentration, and against concentration for each individual level of temperature.

7. CONCLUSION

(i) We must look for good design criteria which measure characteristics of the experimental arrangement in which the scientist might sensibly be interested. Because the importance of various characteristics will differ in different situations, tables of such criteria for particular designs would encourage good judgment to be used in matching the design to the scientific context. Optimum levels of these criteria can be useful as bench marks in judging the efficiencies of a particular design with respect to these various criteria.

(ii) However good designs must in practice be good compromises, and it is doubtful how useful single criterion optimal designs are in locating such compromises. An optimal design is represented by a point in the multi-dimensional space of the coordinates of the design and a series of different criteria will give a series of such extremal points which can be very differently located. Obviously knowledge of the location of such extrema may tell us almost nothing about the location of good compromises. For this we would need to study the joint behaviour of the criterion functions at levels close to their extremal values. One limited but useful step would be to further investigate which criteria are in accord, (such as G-optimality and robustness to wild observations) and which in conflict (such as variance and bias).

(iii) It is true that the problem of experimental design is full of scientific arbitrariness -- no two investigators would choose the same variables, start their experiments in the same place, change variables over the same regions, and so on -- but science works not by uniqueness but by employing iterative techniques which tend to converge. Clearly we must learn to live with scientific arbitrariness, or else we are in a world of make believe. But we can make the problems worse, not better, by introducing arbitrariness for purely mathematical reasons.

REFERENCES

[1] Atkinson, A.C. (1972). Planning experiments to detect inadequate regression models. Biometrika, 59, 275-293.

[2] Atkinson, A.C. (1973). Multifactor second order designs for cuboidal regions. Biometrika, 60, 15-19.

[3] Atwood, C.L. (1969). Optimal and efficient designs of experiments. Ann. Math. Statist., 40, 1570-1602.

[4] Box, G.E.P. (1954). The exploration and exploitation of response surfaces: Some general considerations and examples. Biometrics, 10, 16-60.

[5] Box, G.E.P. (1968). Response surfaces. Article under 'Experimental Design' in The International Encyclopedia of the Social Sciences, Ed. D.L. Sills, 254-259. New York: MacMillan and Free Press.

[6] Box, G.E.P. and Draper, N.R. (1959). A basis for the selection of a response surface design. J. Am. Statist. Assoc., 54, 622-654.

[7] Box, G.E.P. and Draper, N.R. (1963). The choice of a second order rotatable design. Biometrika, 50, 335-352.

[8] Box, G.E.P. and Draper, N.R. (1975). Robust designs. Biometrika, 62, 347-352.

[9] Box, G.E.P. and Draper, N.R. (1982). Measures of lack of fit for response surface designs and predictor variable transformations. To appear in Technometrics.

ok
[10] Box, G.E.P., Draper, N.R., and Hunter, J.S. Empirical Model Building with Response Surfaces. Book in preparation.

ok
[11] Box, G.E.P. and Hunter, J.S. (1957). Multi-factor experimental designs for exploring response surfaces. Ann. Math. Statist., 28, 195-241.

ok
[12] Box, G.E.P., Hunter, W.G., and Hunter, J.S. (1978). Statistics for Experimenters. New York: John Wiley & Sons, Inc.

[13] Box, G.E.P. and Lucas, H.L. (1959). Design of Experiments in nonlinear situations. Biometrika, 46, 77-90.
ok

[14] Box, G.E.P. and Wilson, K.B. (1951). On the experimental attainment of optimum conditions. J. R. Statist. Soc., Ser. B, 13, 1-45.

[15] Box, G.E.P. and Youle, P.V. (1955). The exploration and exploitation of response surfaces: An example of the link between the fitted surface and the basic mechanism of the system. Biometrics, 11, 287-323.

[32] Kiefer, J. (1958). On the nonrandomized optimality and the randomized nonoptimality of symmetrical designs. Ann. Math. Statist., 29, 675-699.

[33] Kiefer, J. (1959). Optimum experimental designs. J. R. Statist. Soc., Ser. B, 21, 272-319.

[34] Kiefer, J. (1975). Optimal design: variation in structure and performance under change of criterion. Biometrika, 62, 277-288.

[35] Kiefer, J. and Studden, W.J. (1976). Optimal designs for large degree polynomial regression. Ann. Statist., 4, 1113-1123.

[36] Kiefer, J. and Wolfowitz, J. (1959). Optimum designs in regression problems. Ann. Math. Statist., 30, 271-294.

ok [37] Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. Canadian J. Math., 12, 363-366.

* [38] Lucas, J.M. (1976). Which response surface design is best: a performance comparison of several types of quadratic response surface designs in symmetric regions. Technometrics, 18, 411-417.

ok [39] Mead, R. and Pike, D.J. (1975). A review of response surface methodology from a biometric viewpoint. Biometrics, 31, 803-851.

[40] Nalimov, V.V., Golikovo, T.I., and Mikeshina, N.G. (1970). On practical use of the concept of D-optimality. Technometrics, 12, 799-812.

[41] Pinchbeck, P. H. (1957). The kinetic implications of an empirically fitted yield surface for the vapour phase oxidation of naphthalene to phthalic anhydride. Chem. Engr. Science, 6, 105-111.

[42] Silvey, S.D. and Titterington, D.M. (1973). A geometrical approach to optimal design theory. Biometrika, 60, 21-32.

[43] Titterington, D.M. (1975). Optimal design: some geometrical aspects of D-optimality. Biometrika, 62, 313-320.

[44] Yates, F. (1937). The design and analysis of factorial experiments. Imperial Bureau of Soil Science.

[45] Yates, F. (1967). A fresh look at the basic principles of the design and analysis of experiments. In Proc. 5th Berkeley Symp. Math. Statist. Prob., IV, 777-790. Berkeley and Los Angeles: University of California Press.

[46] Yates, F. (1970). Experimental Design: Selected Papers of Frank Yates. Darien, Connecticut: Hafner.

[16] Box, J.F. (1978). R.A. Fisher: The Life of a Scientist. New York: John Wiley & Sons, Inc.

[17] Box, M.J. (1968). The occurrence of replications in the optimal designs of experiments to estimate parameters in non-linear models. J. R. Statist. Soc., Ser. B, 30, 290-302.

[18] Cochran, W. G. (1973). Experiments for nonlinear functions. J. Am. Statist. Assoc., 68, 771-781.

[19] DeBaun, R. M. (1956). Block effects in the determination of optimum conditions. Biometrics, 12, 20-22.

[20] Draper, N. R. and Lawrence, W. E. (1967). Sequential designs for spherical weight functions. Technometrics, 9, 517-529.

[21] Fedorov, V.V. (1972). Theory of Optimal Experiments. New York and London: Academic Press.

[22] Fisher, R.A. (1921). Studies in crop variation. I. An examination of the yield of dressed grain from Broadbalk. J. Agric. Sci., 11, 107-135.

[23] Fisher, R.A. (1923). Studies in crop variation. II. The manurial response of different potato varieties. J. Agric. Sci., 13, 311-320.

[24] Fisher, R.A. (1924). The influence of rainfall on the yield of wheat at Rothamsted. Phil. Trans., B, 213, 89-142.

[25] Fisher, R. A. (1935). The Design of Experiments. Oliver and Boyd.

[26] Franklin, N. L., Pinchbeck, P. H., and Popper, F. (1956). A statistical approach to catalyst development, Part 1. The effect of process variables on the vapour phase oxidation of naphthalene. Trans. Inst. Chem. Engrs., 34, 280-293.

[27] Herzberg, A.M. (1979). Are theoretical designs applicable? Operations Research Verfahren/Methods of Operations Research, 30, 68-76.

[28] Herzberg, A.M. (1981). The robust design of experiments: a review. Technical Summary Report 2218, Math. Research Center, University of Wisconsin at Madison.

[29] Herzberg, A.M. and Andrews, D.F. (1976). Some considerations in the optimal design of experiments in non-optimal situations. J. R. Statist. Soc., Ser. B, 38, 284-289.

[30] Herzberg, A.M. and Cox, D.R. (1969). Recent work on the design of experiments: A bibliography and a review. J. R. Statist. Soc., Ser. A, 132, 29-67.

[31] Hill, W.J. and Hunter, W.G. (1966). A review of response surface methodology: a literature survey. Technometrics, 8, 571-590.

AMBUSHED BY A LURKING VARIABLE

Barry A. Bodt
Jerry Thomas

USA Ballistic Research Laboratory

ABSTRACT

In the formal study of design and analysis of experiments, it is often overlooked that a simple and straight-forward design can become complicated during analysis. Presented here is a specific case in which the design was readily apparent but where difficulties subsequently arose. Analysis, plagued by nonhomogeneity of variance and the suspicion of a lurking variable, is discussed.

INTRODUCTION

Answers to questions concerning the performance of a MLRS (Multiple Launcher Rocket System) bomblet were desired. The M42 is a small shaped-charge bomblet (figure 1), designed to detonate on impact causing a jet, comprised primarily of copper, to penetrate the armor which it has impacted. Many bomblets are placed within a time-fused rocket, which is flown over the target area. A charge within the rocket is ignited, causing the skin of the rocket to peel away. This allows the undetonated bomblets to be sprayed over the target area; as the bomblets fall to the ground, a portion of them will impact the target.

DESIGN

There were three questions about the performance of this munition to be answered. First, is there a difference in bomblet performance among vendors? In this study, performance of the bomblet was taken to be penetration depth of the jet into the target. This question is self-explanatory and we will only note that there were three vendors considered. Second, does the dispersing process have an effect on bomblet performance? Dispersing is the process by which the bomblets are delivered from the rocket to the target. In particular, the customer was concerned with the ignition of the charge within the rocket which causes the skin of the rocket to peel away. When this charge is ignited, the bomblets are subjected to a certain amount of force. The above question then becomes how does this force affect bomblet performance. In order to answer this question, one half of the bomblets went through the dispersing simulation before testing for penetration depth. Third, how does Standoff affect bomblet performance? Standoff is the distance above the target at which the bomblet is detonated. The customer was interested in bomblet performance where detonation occurs at four different heights above the target.

To answer these questions, an experimental design was developed (figure 2). A 2x3x4 factorial design with response, Penetration Depth, and with factors, Dispensing, Vendor, and Standoff was chosen. In consideration of available bomblets, six observations per cell were used. This design was then suggested to the customer who then contracted a third party to run the experiment.

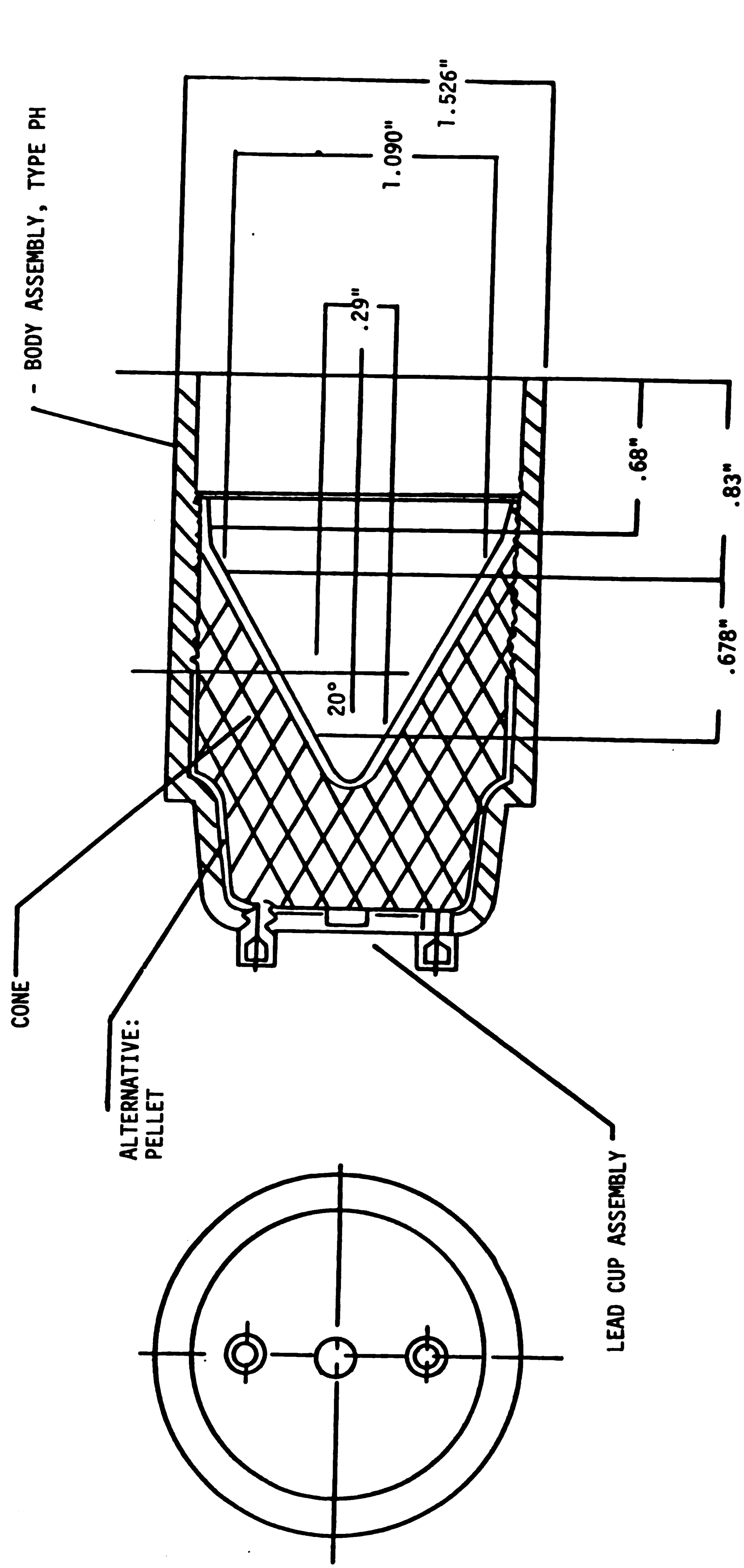

BODY ASSEMBLY, TYPE PH
1.526"
1.090"
.29"
.68"
.83"
.678"
20°
CONE
ALTERNATIVE:
PELLET
LEAD CUP ASSEMBLY

Fig. 1

STANDOFF

		.72"	3.86"	7.72"	15.44"
DISPERSED BOMBLETS	VENDOR 1	6 REPS			
	VENDOR 2				
	VENDOR 3				
NON-DISPERSED BOMBLETS	VENDOR 1				
	VENDOR 2				
	VENDOR 3				

FIGURE 2.

In examining the data, irregularities in the values caused us concern with respect to the usual model assumptions of normality, homogeneity of variance, and additivity. Prior to performing an analysis of variance, testing of those assumptions was begun. To test for normality, a Shapiro-Wilk test was run on the observations within cells. At the .05 significance level we found the results not inconsistent with the assumption of normality. Turning then to the question of homogeneity of variance, a plot (figure 3) of the cell means against the cell variances was constructed. When examining this graph, it was fairly obvious that conditions were somewhat less than ideal. Various corrective measures using the common transformations were unsuccessful in obtaining homogeneity of variance. Thus, efforts were begun to determine the cause of heterogeneity of variance.

A more critical look at the data revealed that within many of the cells representing dispersed bomblets there seemed to be two populations of data, a group of high values and a group of low values. Subjectively we flagged the lower values. Graphically (figure 4) we compared the means of the lower values and the means of the higher values within a given cell. On the plot, the symbol at the approximate coordinates (.75,.75) represents the mean of the lower values from vendor 1 at the first standoff. Noting the obvious difference between the mean of the lower and upper values within a given cell, we began to feel that maybe there were in fact two populations of data. It was at this point that we first suspected the existence of a lurking variable.

In mid-stream we were asked to look at the effect of a new variable, Damage, which is a measure of 'out of round' of the bomblet. It was previously conjectured that the dispersing process may affect bomblet performance. Damage was an attempt at a more precise explanation of the possible effect of dispersing. In explaining how this measurement was taken, it is necessary that the testing sequence and apparatus first be described. First, bomblets are disarmed and, noting each bomblet position, loaded into a rocket-like canister comprised of five bomblet-holding packs (figure 5). The dispersing simulation involves exploding a charge within the canister causing bomblets to be sprayed over the test area. The bomblets are then gathered and measured for Damage, which is the absolute difference of two perpendicular measurements of bomblet diameter. After this simulation, the bomblets are armed and detonated at various heights over a plate of armor for the penetration depth data. Looking at this variable, Damage, led us to find our lurking variable.

Investigation of Damage brought out the following observations. First, those bomblets positioned in packs one and two during the dispersing simulation sustained a higher level of Damage than did those positioned in packs three through five. Second, those bomblets positioned in packs one and two during the dispersing simulation showed poorer penetration than did those in packs three through five. Third, high levels of Damage sustained by the bomblets adversely affected penetration performance. These observations are supported graphically by representative figures 6 and 7.

In figure 6, the symbol at the approximate coordinates (1.,3.75) represents the mean Damage sustained by bomblets, positioned in pack one during the dispersing simulation and then fired at the 7.72 inch standoff. The symbol at the approximate coordinates (1.,1.) represents the mean penetration depth achieved by those same bomblets. Note that in each graph the highest level of Damage is sustained by bomblets from pack one and that the level of damage decreases for bomblets from higher packs. Also the lowest mean penetration depth is exhibited by bomblets from pack one and generally increases for bomblets from higher packs. The apparent relationship between Damage and Penetration Depth was important, but not totally unexpected. More interesting and more important was the relationship of Pack to both Damage and Penetration Depth.

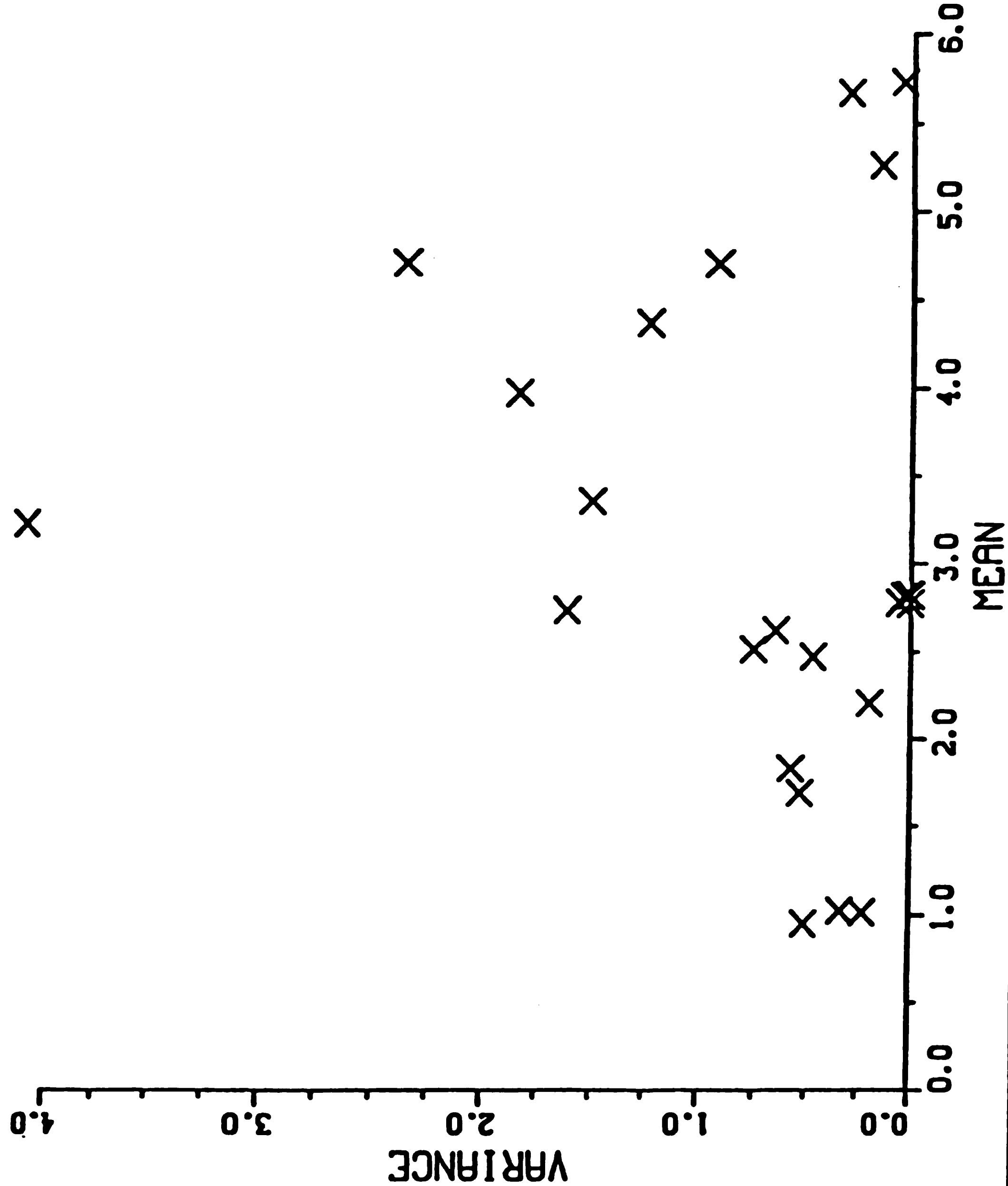


FIGURE 3.

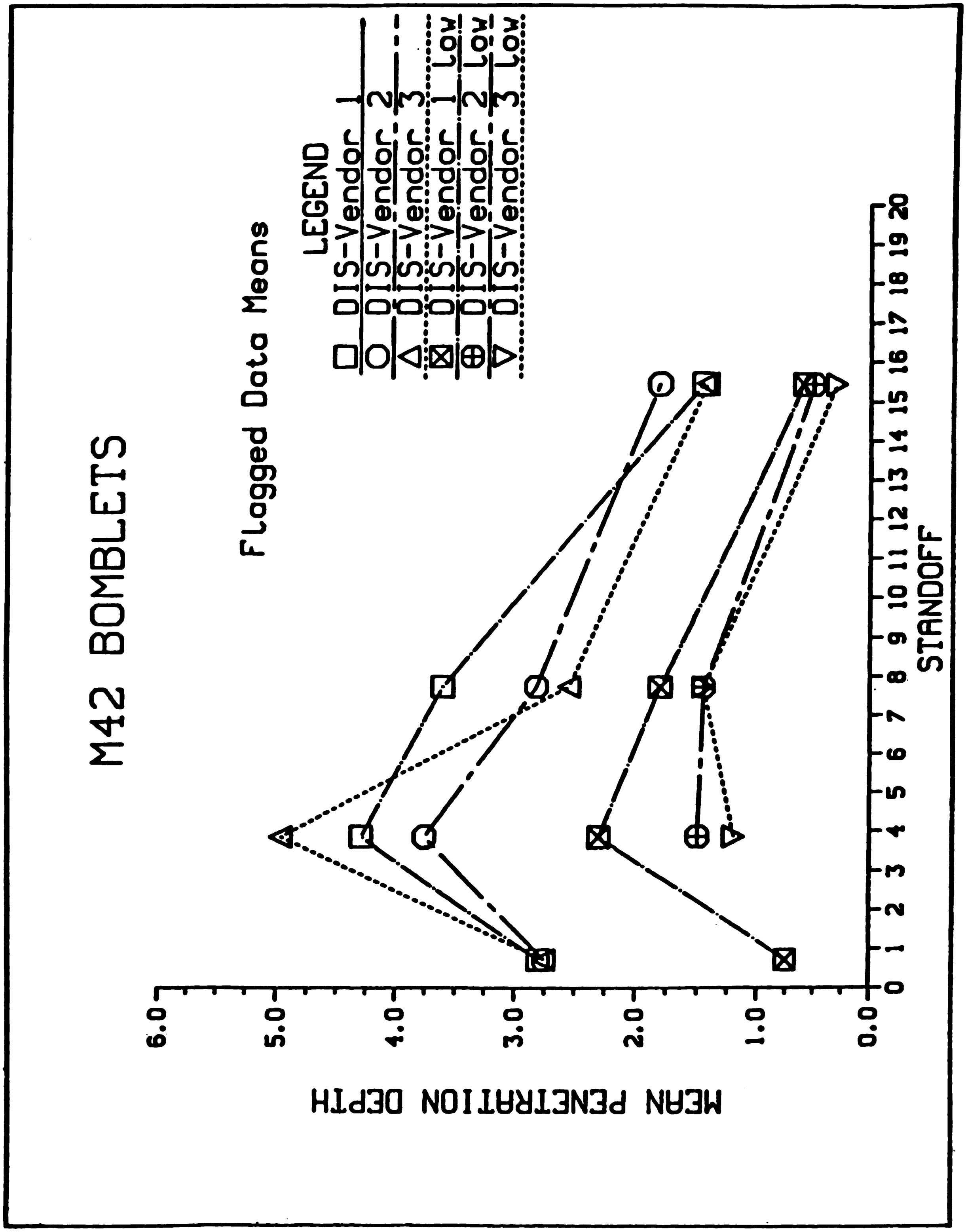

M42 BOMBLETS
Flagged Data Means
LEGEND
DIS-Vendor 1
DIS-Vendor 2
DIS-Vendor 3
DIS-Vendor 1 Low
DIS-Vendor 2 Low
DIS-Vendor 3 Low
MEAN PENETRATION DEPTH
0.0
1.0
2.0
3.0
4.0
5.0
6.0
STANDOFF
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

FIGURE 4.

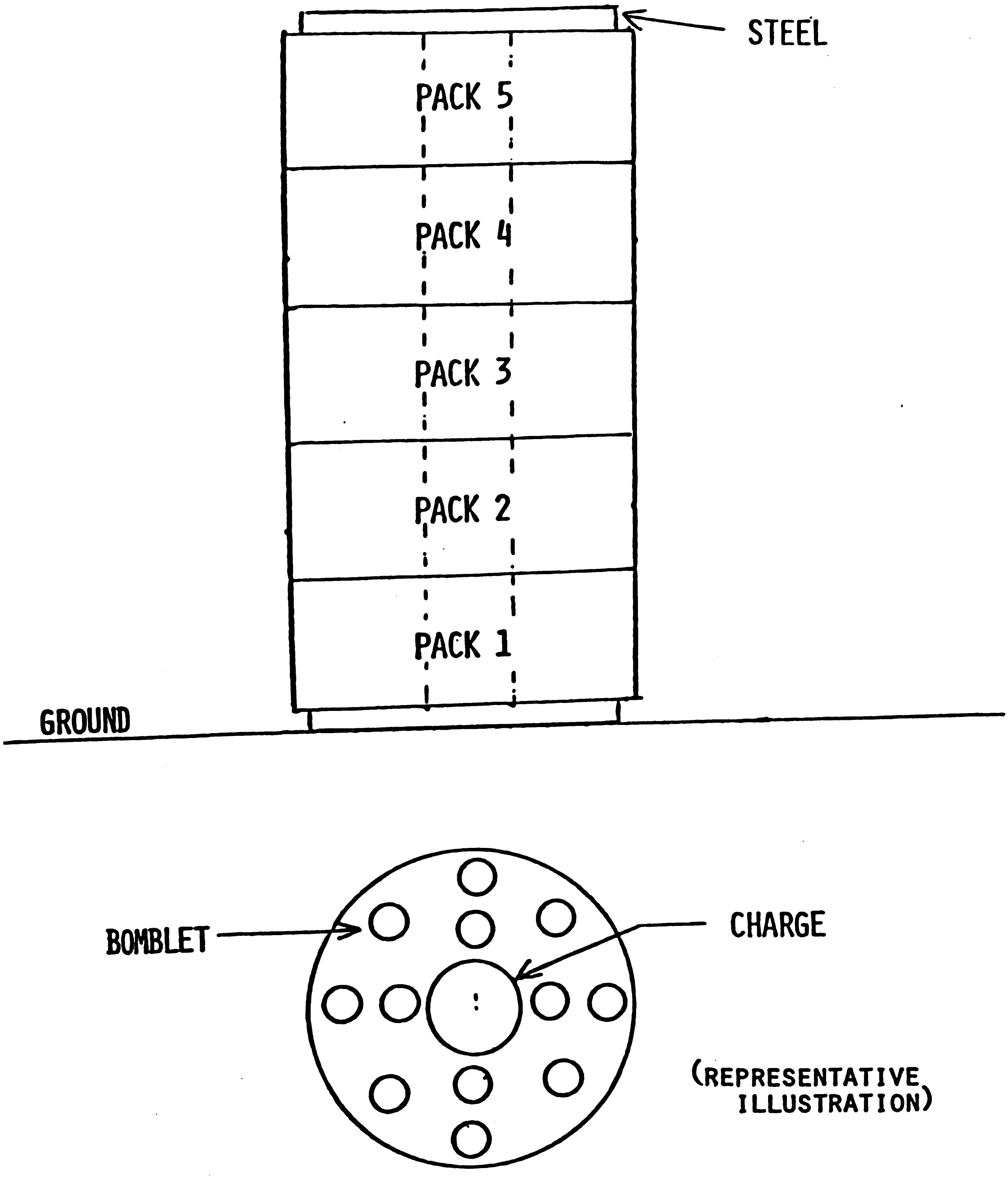

STEEL
PACK 5
PACK 4
PACK 3
PACK 2
PACK 1
GROUND
BOMBLET
CHARGE
(REPRESENTATIVE ILLUSTRATION)

FIGURE 5.

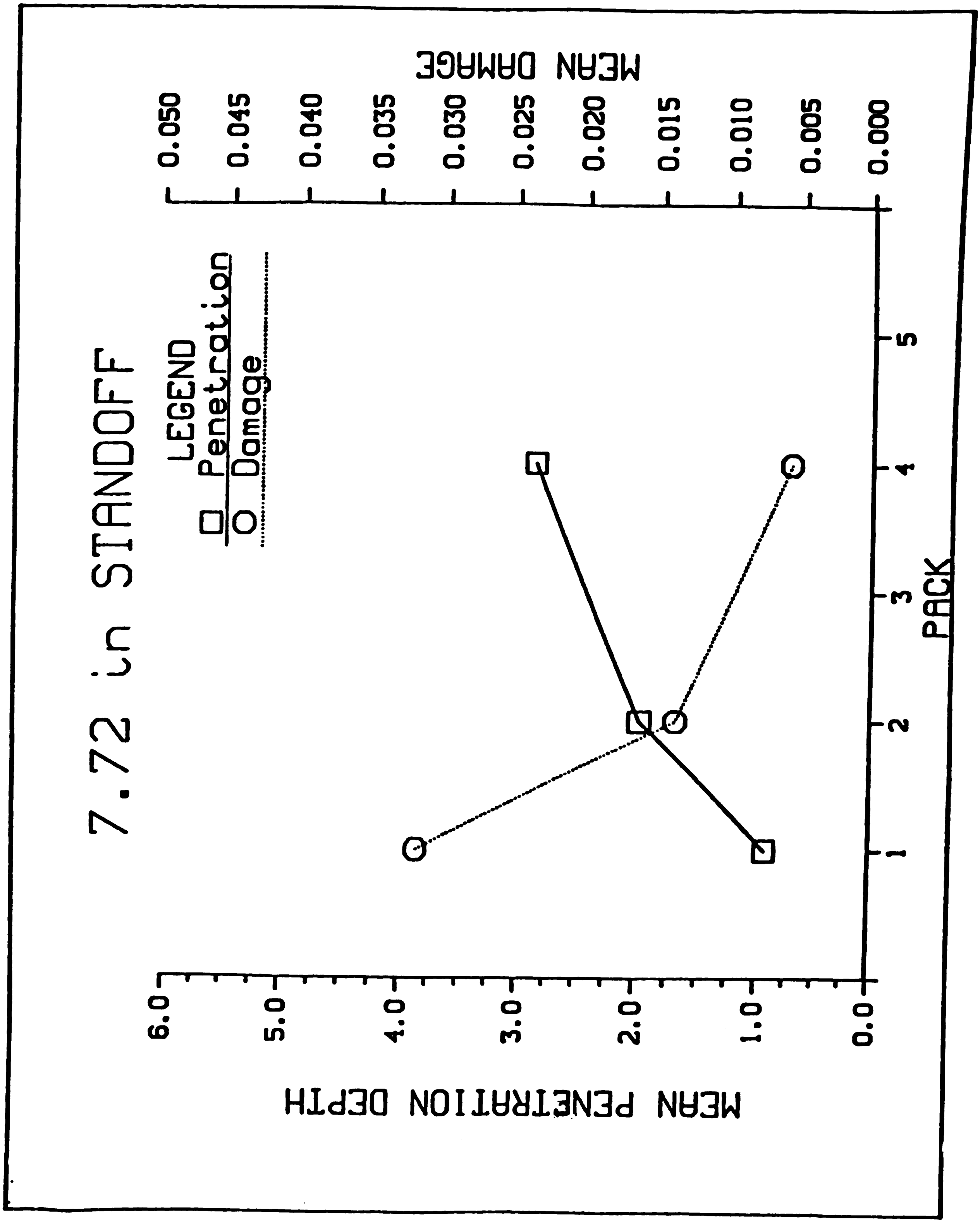

7.72 in STANDOFF
LEGEND
□ Penetration
○ Damage
MEAN DAMAGE
0.050
0.045
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
MEAN PENETRATION DEPTH
6.0
5.0
4.0
3.0
2.0
1.0
0.0
1
2
3
4
5
PACK

FIGURE 6.

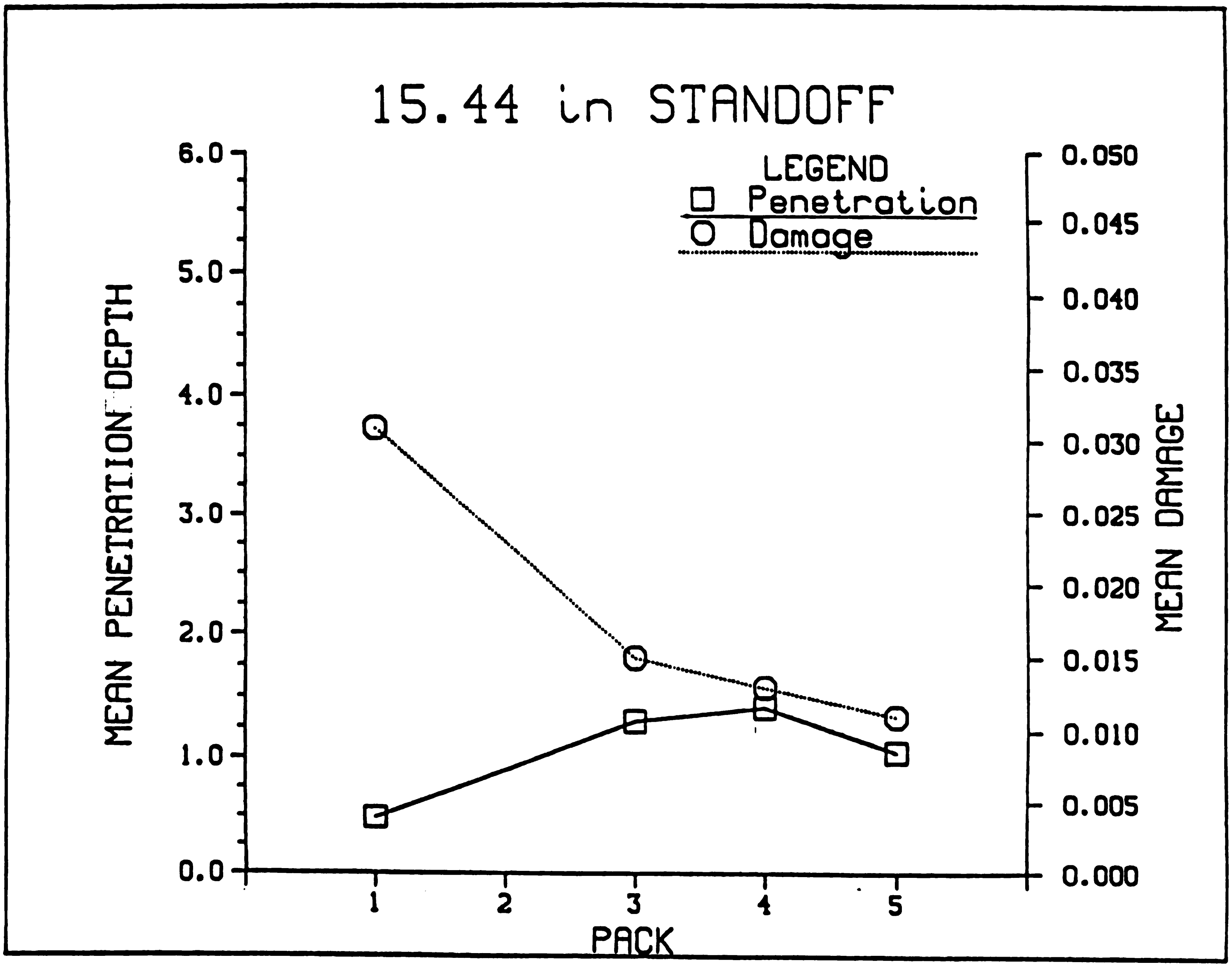

15.44 in STANDOFF
LEGEND
Penetration
Damage
MEAN PENETRATION DEPTH
6.0
5.0
4.0
3.0
2.0
1.0
0.0
MEAN DAMAGE
0.050
0.045
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
1
2
3
4
5
PACK

FIGURE 7.

At one point early in the analysis we flagged bomblets showing lower penetration depths as possibly coming from a different population. The relationship between Pack and those penetration depths being flagged is pointed out further in figure 8. Of fourteen bomblets positioned in pack one during the dispersing simulation, eleven were flagged for low penetration. Of fifteen bomblets positioned in pack two, nine were flagged for low penetration. Finally, of twenty seven bomblets flagged, twenty had been positioned in packs one or two during the dispersing simulation. Due to its unexpected effect on bomblet performance, Pack was determined to be our lurking variable.

Why did Pack have an effect on penetration depth? One possibility was proposed by a systems analyst familiar with MLRS munitions. In figure 5, note that steel plates were bolted to the top of pack five and to the bottom of pack one. Rather than being suspended in air, the test apparatus rested on the ground. When the charge within the canister was ignited, the shell of the canister, the bomblets, and the steel plate on pack five were blown out away from the center of the canister. The bottom steel plate remained stationary, pinned by the force of the explosion and the ground. Many bomblets from the lower packs caromed off this hard fixed surface, causing more severe deformation to themselves.

CONCLUSION

In conclusion, some information, not addressed here, could still be extracted from these experimental data, but problems created by the lurking variable hindered the intended complete analysis. It is interesting to note that heterogeneity of variance played a hero's role in this analysis, since investigation of this problem aided in the discovery of the lurking variable, Pack. Also, proper design made it possible to draw some conclusions in the face of unexpected circumstances. Finally, as suggested by Professor G.E.P. Box during this presentation, this example illustrates that statistical analysis can accomplish much more than hypothesis testing by lending insight to the physical environment, in this case by pointing out possible inadequacies in the test apparatus.

RATIO OF FLAGGED DATA POINTS TO THE NUMBER OF POINTS IN A CELL

	STANDOFF				
PACK	BUILT IN	3.86"	7.72"	15.44	TOTAL
1	0/3	2/2	3/3	6/6	11/14
2	0/0	2/6	7/9	0/0	9/15
3	0/3	0/3	0/0	2/6	2/12
4	0/3	0/4	2/6	0/3	2/16
5	1/9	1/3	0/0	1/3	3/15
TOTAL	1/18	5/18	12/18	9/18	

20 OF 27 FLAGGED DATA POINTS CAME FROM PACKS 1 AND 2

FIGURE 8.

The Use of Box-Jenkins Methodology in Forecasting DARCOM's Central Procurement Workload

Charles A. Correia

Army Materiel Systems Analysis Activity
Army Procurement Research Office
Fort Lee, Virginia 23801

ABSTRACT. This paper addresses the development of a time series model to forecast the central procurement workload in the U.S. Army Materiel Development and Readiness Command (DARCOM). The Box-Jenkins approach of Identification, Estimation and Diagnostic Checking is used to build a seasonal auto-regressive integrated moving average model (SARIMA) to forecast quarterly procurement actions (QPA), the procurement workload indicator. Models are developed using 60 and 61 data points from FY 65 through FY 79 to include 7T. Forecasted values are compared to actual values for FY 80, 81 and 82.

I. Introduction. The user of the Box-Jenkins methodology for time series forecasting is required to exercise judgment in the choice of a particular model from a general class of autoregressive integrated moving average (ARIMA) models:

$$(1-\phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1-B)^d X_t = (1-\theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)\,\varepsilon_t$$

where ε_t has mean zero, fixed variance and ε_t and ε_u uncorrelated for $t \neq u$ and B is the backshift operator defined as $B^m X_t = X_{t-m}$.

For seasonal time series of period s the model is generalized to

$$(1-\phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1-\phi_{1,s} B^s - \dots - \phi_{P_s,s} B^{P_s s})(1-B)^d (1-B^s)^{d_s} X_t = (1-\theta_1 B - \dots - \theta_q B^q)(1-\theta_{1,s} B^s - \dots - \theta_{q_s,s} B^{q_s s})\,\varepsilon_t$$

The task is to select a specific model from the general class by choosing appropriate values for p, d, and q. These values are determined by examining the sample autocorrelations and partial autocorrelations. The coefficients ϕ and θ are estimated by a nonlinear optimization algorithm and the residuals of the proper fitted model should resemble the properties of a white noise process. The computer program used for determining the optimal parameter values is a FORTRAN program called ERSF, Estimation of Rational Distributed Log Structural Form Models, developed by K. D. Wall, School of Engineering, University of Virginia.

II. Model Development.

Models are developed using 60 and 61 data points from FY 65 through FY 79 to include 7T. One model is fitted for 60 points and two for 61 points. The correlogram of both series indicate non-stationarity and seasonality as shown in figure 1.

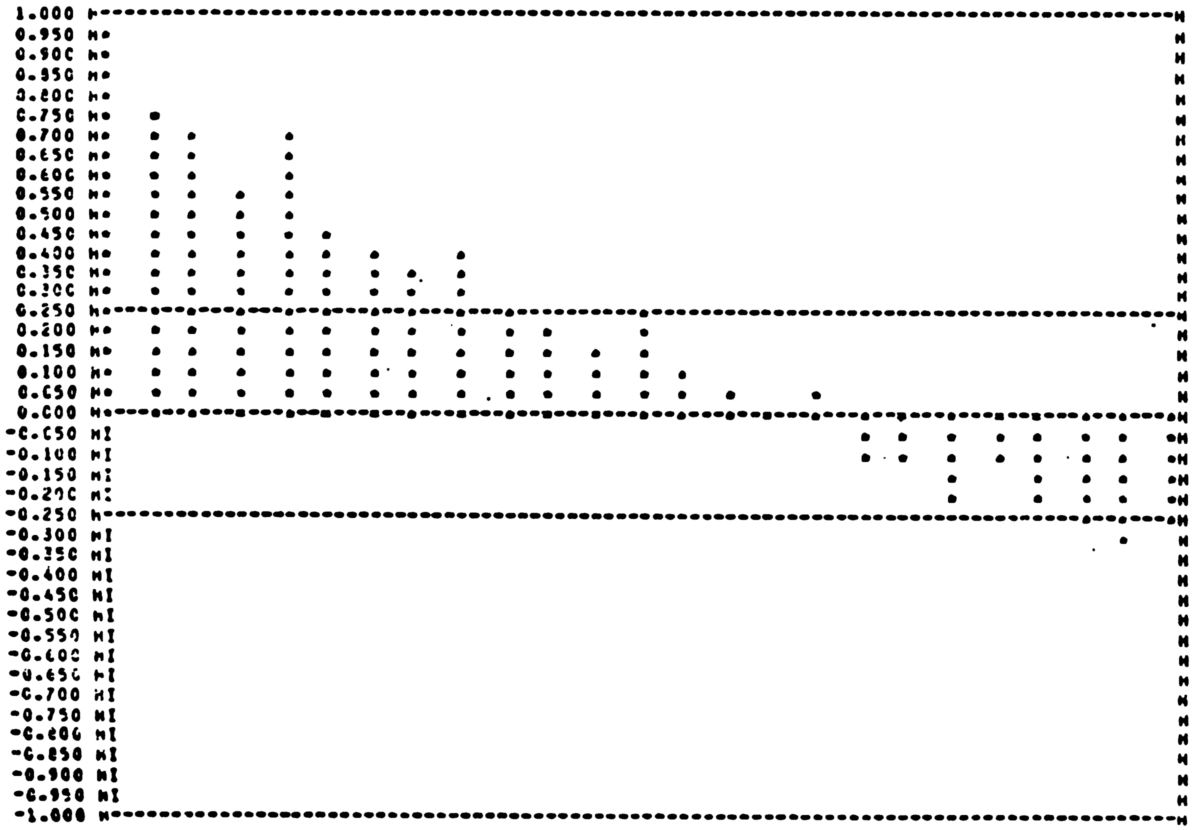


Figure 1. Autocorrelations of Original Data

Forecasted values are compared to actual values for FY 80, 81, and 82. The mean absolute percentage error (MAPE) is used as a measure of forecast accuracy rather than the mean squared error (MSE). The MAPE has a more intuitive interpretation than the MSE, and hence more understandable to non-statisticians.

A. Model I

The first model addressed is fitted to 60 data points. When a seasonal fourth difference is taken the correlagram in figure 2 results, showing two autocorrelations significantly different from zero, r_1 and r_2. An MA(2), ARMA (1,1) and AR(1) were then fitted with only the residuals of the AR(1) exhibiting the characteristics of a white noise process. The autocorrelation of the residuals of the following model

$$(1-B^4)\ (1-0.67205B)\ X_t = \varepsilon_t$$

is shown in figure 3. These appear to be random and hence possibly a correct model for forecasting purposes.

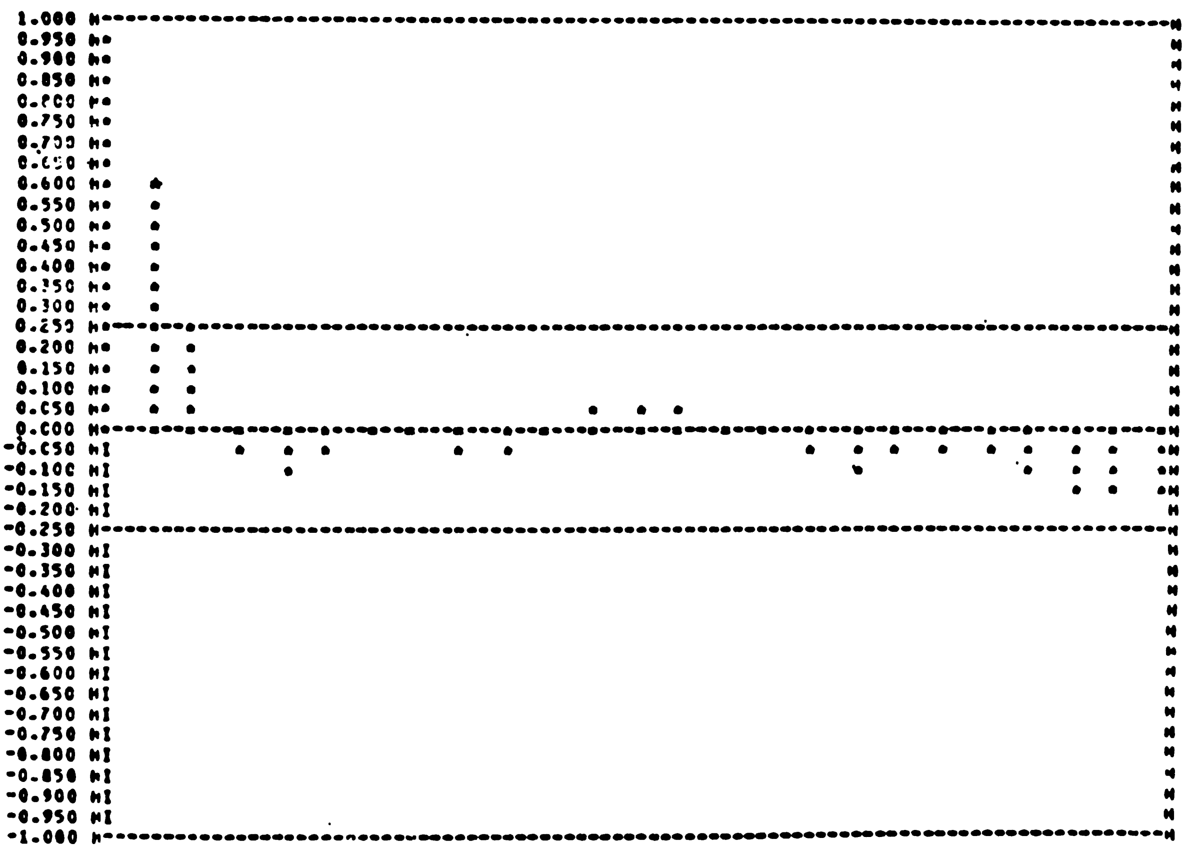


Figure 2. Autocorrelations of Seasonal Fourth Difference

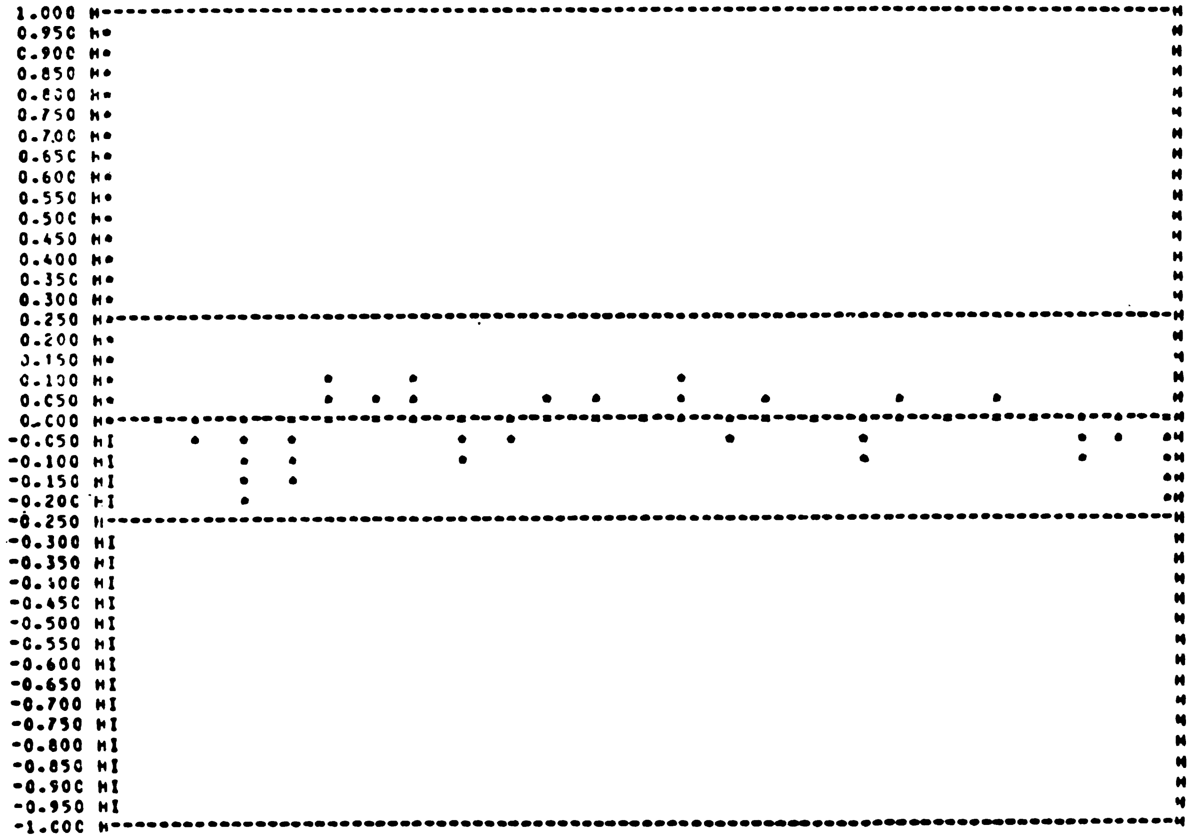


Figure 3. Autocorrelations of Residuals of Model 1

B. Model II

The 7T data point (the quarter which occurred when the fiscal year was changed from 1 July - 30 June to 1 Oct - 30 Sept) was added to the time series and a new model was fitted. A first difference was taken and a seasonal parameter added resulting in the correlogram of figure 4. The pattern of the autocorrelations is that of an MA(3). The autocorrelations of the residuals of the model

$$(1-B)(1-0.51982B^4)\ X_t = (1.0.2895B-0.13664B^2 - 0.60893B^3)\ \epsilon_t$$

are shown in figure 5. Although the autocorrelations of the residuals are all within one standard error a trend still appears to exist since r_1 through r_{13} are all positive. This trend caused some doubt as to whether the residuals were truly random, and so a third model was attempted.

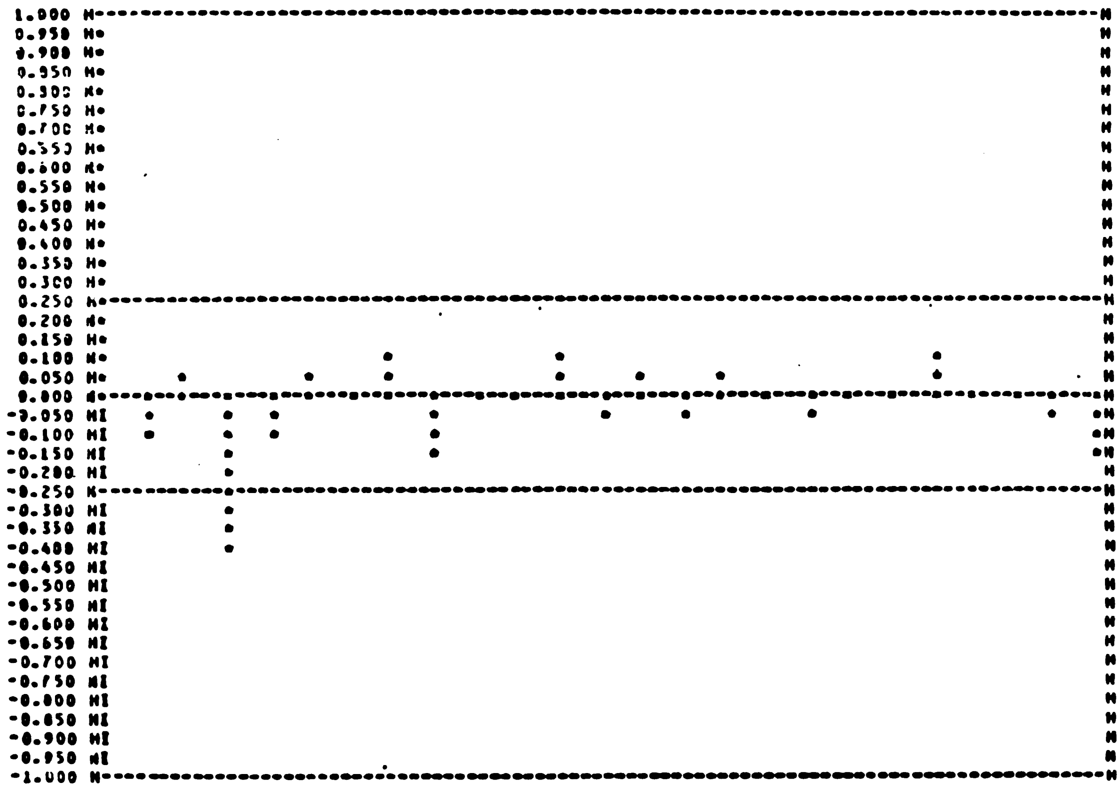


Figure 4. Autocorrelations of First Difference and Seasonal Parameter

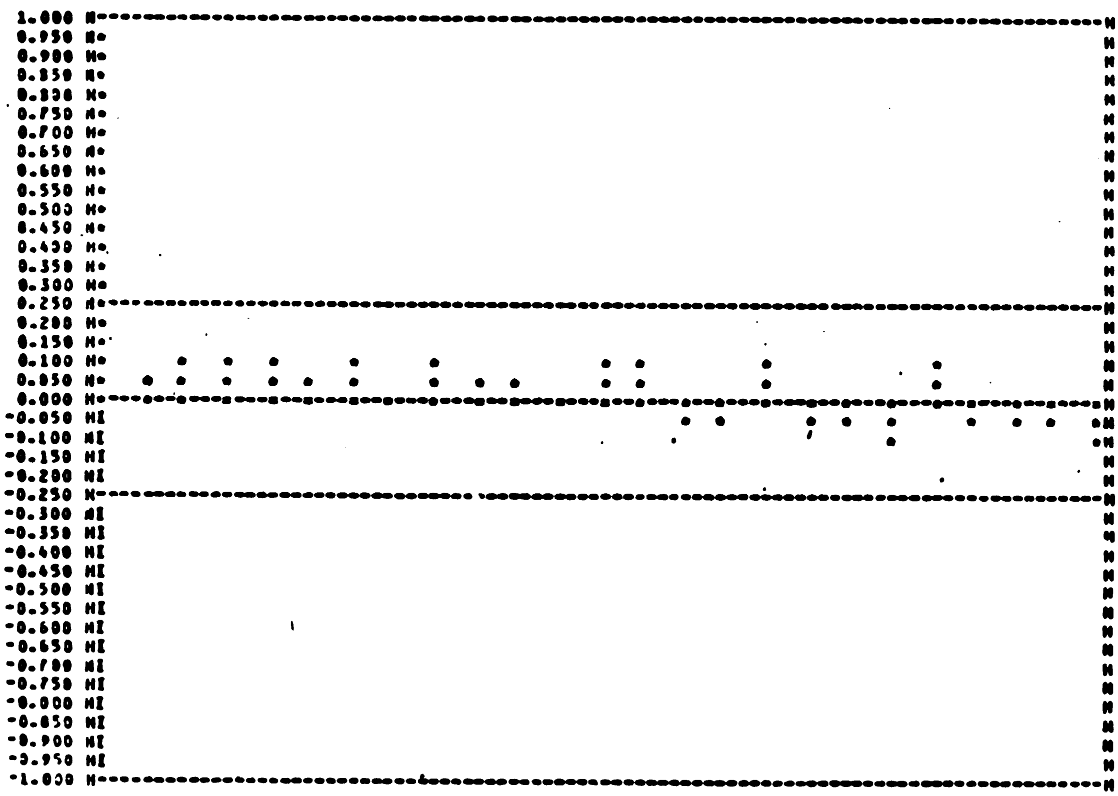


Figure 5. Autocorrelations of Residuals of Model 2

C. Model III

In an attempt to eliminate the trend of model 2 a fourth difference instead of the first difference is taken but the seasonal parameter along with the three moving average terms are kept. This resulted in the correlogram of figure 6. When a fourth moving average term is added, the autocorrelations of the residuals exhibit a random pattern as shown in figure 7. The resulting model is

$$(1-B^4)(1-0.60162B^4)\, X_t = (1+0.81815B + 0.87187B^2 + 0.66825B^3 - 0.29190B^4)\, \varepsilon_t$$

III. Conclusion

The three models along with their forecasts are compared to the actual values for fiscal years 1980, 1981 and 1982, in tables 1, 2 and 3. The mean absolute percentage error for model 2 is the lowest. The forecasts from model 2 were chosen to predict future workload back in fiscal year 1979.

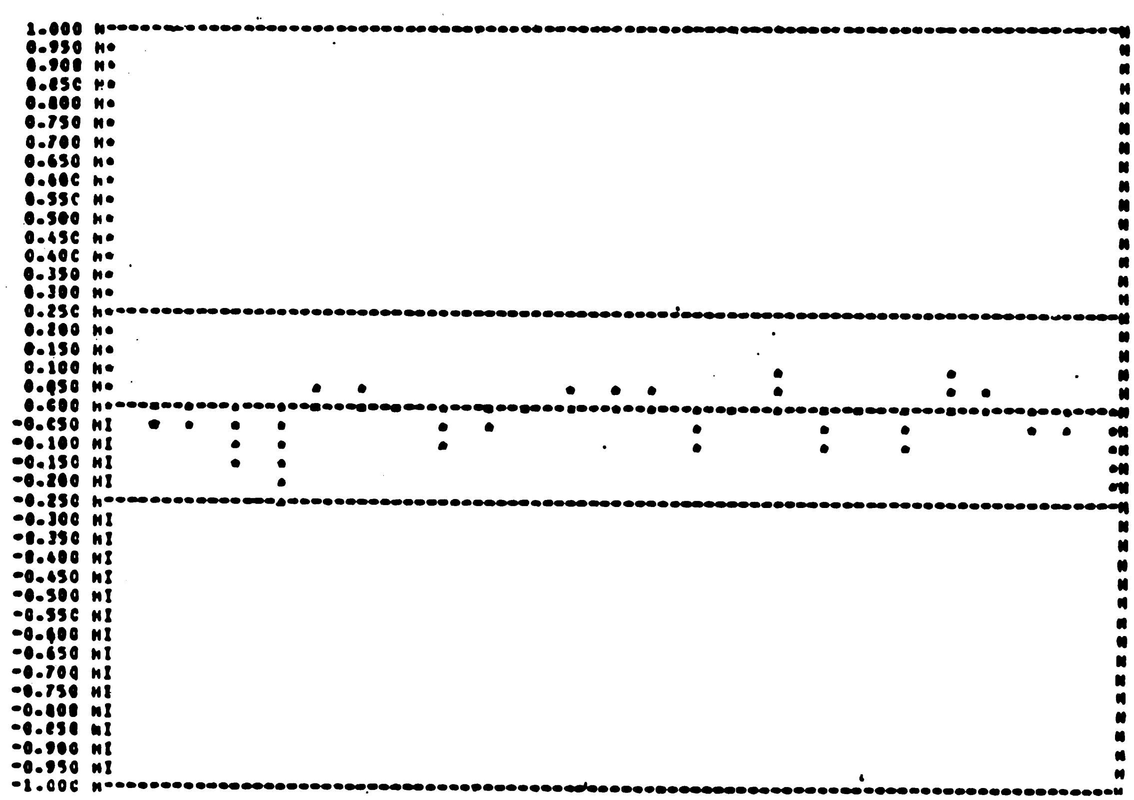


Figure 6. Autocorrelations of Seasonal Fourth Difference of Model 2

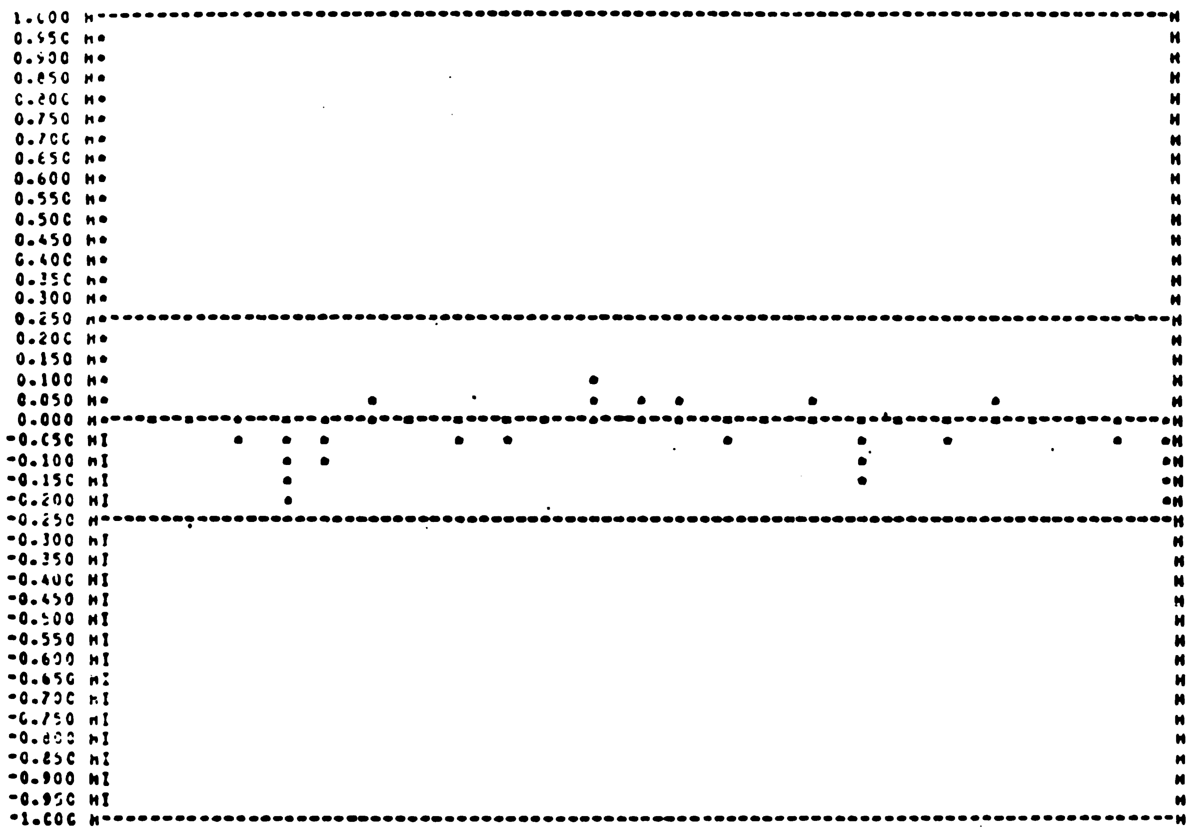


Figure 7. Autocorrelations of Residual of Model 3

Table 1

Performance of Model 1

MODEL 1 $(1-B^4)(1-0.67205B) Y_t = \varepsilon_t$

	FY'80			FY'81		
	FORECAST	ACTUAL	PE	FORECAST	ACTUAL	PE
1st Qtr	23,322	27,455	15.05	22,784	29,455	22.65
2nd Qtr	35,210	35,034	.50	35,572	34,938	1.81
3rd Qtr	40,198	36,864	9.04	39,955	37,182	7.46
4th Qtr	40,721	36,562	11.38	40,884	38,802	5.36
TOTAL	139,451	135,915		139,195	140,377	

	FY'82		
	FORECAST	ACTUAL	PE
1st Qtr	22,674	30,330	25.24
2nd Qtr	35,646	35,040	1.73
3rd Qtr	39,905	41,282	3.34
4th Qtr	40,918	45,739	10.54
TOTAL	139,143	153,194	

MAPE = 9.50

Table 2

Performance of Model 2

MODEL 2 $(1-B)(1-0.51982B^4)\,Y_t = (1.0.2895B - 0.13664B^2 - 0.60893B^3)\,\varepsilon_t$

	FY'80			FY'81		
	FORECAST	ACTUAL	PE	FORECAST	ACTUAL	PE
1st Qtr	32,364	27,455	17.88	34,507	29,455	17.15
2nd Qtr	32,854	35,034	6.22	34,762	34,938	0.50
3rd Qtr	39,199	36,864	6.33	38,060	37,182	2.36
4th Qtr	38,435	36,562	5.12	37,663	38,802	2.94
TOTAL	142,852	135,915		144,992	140,377	

	FY'82		
	FORECAST	ACTUAL	PE
1st Qtr	35,621	30,330	17.44
2nd Qtr	35,753	35,040	2.03
3rd Qtr	37,468	41,282	9.24
4th Qtr	37,261	45,739	18.54
TOTAL	146,103	153,194	

MAPE = 8.81

Table 3

Performance of Model 3

MODEL 3 $$(1-B^4)(1-0.60162B^4)\,Y_t = (1 + 0.81815B + 0.87187B^2 + 0.66825B^3 - 0.29190B^4)\,\varepsilon_t$$

	FY'80			FY'81		
	FORECAST	ACTUAL	PE	FORECAST	ACTUAL	PE
1st Qtr	31,814	27,455	15.88	35,335	29,455	19.96
2nd Qtr	34,899	35,034	.004	35,779	34,938	2.41
3rd Qtr	43,464	36,864	17.90	44,712	37,182	20.25
4th Qtr	43,116	36,562	17.92	45,039	38,802	16.07
TOTAL	153,293	135,915		160,865	140,377	

FY'82

	FORECAST	ACTUAL	PE
1st Qtr	37,453	30,330	23.48
2nd Qtr	36,308	35,040	1.02
3rd Qtr	41,463	41,282	10.13
4th Qtr	46,196	45,739	1.00
TOTAL	165,420	153,194	

MAPE = 12.17

COMPARISON OF CEP ESTIMATORS FOR
ELLIPTICAL NORMAL ERRORS

AUDREY E. TAUB
MARLIN A. THOMAS

Strategic Systems Department
Naval Surface Weapons Center
Dahlgren, Virginia

Introduction

A common parameter for describing the accuracy of a weapon is the circular probable error, generally referred to as CEP. CEP is simply the bivariate analog of the probable error of a single variable and measures the radius of a mean centered circle which includes 50% of the bivariate probability mass. In the case of circular normal errors where the error variances are the same in both directions, CEP can be expressed in terms of the common standard deviation, and estimators are easily formulated and compared. In the case of elliptical normal errors, CEP cannot be expressed in closed form, and hence, estimators are less easily formulated. The problem addressed herein is the comparison of CEP estimators for the elliptical case based on some of the commonly used CEP approximations.

It will be instructive to first review the case of circular normal errors. In general, it will be assumed that the errors in the X and Y directions are independent with mean zero and variances σ_x^2 and σ_y^2, respectively. Under the circular normal assumption, $\sigma_x^2 = \sigma_y^2 = \sigma^2$ and the bivariate distribution of errors is given by

$$f_c(x,y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}, \quad -\infty < x,y < \infty. \tag{1}$$

The distribution of $R = (X^2 + Y^2)^{\frac{1}{2}}$ is easily derived and found to be

$$g_C(r) = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2}, \quad r > 0 . \tag{2}$$

This is the well known Rayleigh distribution with cumulative distribution function

$$P(R < r) = G_C(r) = 1 - e^{-r^2/2\sigma^2} . \tag{3}$$

By definition, $G_C(CEP) = .5$, and the solution of equation (3) yields the well-known expression

$$CEP = [-2\ell n\ (.50)]^{\frac{1}{2}} \sigma = 1.1774\sigma . \tag{4}$$

Four estimators for CEP in the circular case were formulated and compared by Moranda (1959).

Consider now the case of elliptical normal errors. Here the variances are unequal, and the bivariate distribution of errors is given by

$$f_E(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left[(x/\sigma_x)^2 + (y/\sigma_y)^2\right]}, \quad -\infty < x,y < \infty . \tag{5}$$

For this case, the distribution of the radial error R was derived by Chew and Boyce (1961) and has form

$$g_E(r) = \frac{r}{\sigma_x\sigma_y} e^{-a r^2} I_o(b r^2) \tag{6}$$

where

$$a = \frac{\sigma_y^2 + \sigma_x^2}{(2\sigma_x\sigma_y)^2} , \quad b = \frac{\sigma_y^2 - \sigma_x^2}{(2\sigma_x\sigma_y)^2}$$

and I_o is the modified Bessel function of the first kind and zero order, i.e.,

$$I_o(x) = \frac{1}{\pi} \int_0^{\pi} e^{-x \cos\theta} \, d\theta .$$

The cumulative distribution function for R is denoted by

$$P(R < r) = G_E(r) = \int_0^r g_E(t) \, dt, \qquad (7)$$

but it cannot be expressed in closed form. Hence, the radius of the 50% circle for the elliptical case cannot be expressed by a simple formula as it was in the circular case. One has to solve $G_E(CEP) = .5$ by numerical methods or by referring to tables prepared by Harter (1960), DiDonato and Jarnagin (1962), and others. To avoid using these tables or numerical procedures for CEP evaluation, a number of approximations have been developed over the years. Five of these approximations have been chosen for examination. They are designated below as CEP_1 through CEP_5:

$$CEP_1 = 1.1774 \left(\frac{\sigma_x^2 + \sigma_y^2}{2} \right)^{1/2} \qquad (8)$$

$$CEP_2 = 1.1774 \left(\frac{\sigma_x + \sigma_y}{2} \right) \qquad (9)$$

$$CEP_3 = \left(2 \, \chi^2_{\nu,.50}/\nu \right)^{1/2} \left(\frac{\sigma_x^2 + \sigma_y^2}{2} \right)^{1/2} \qquad (10)$$

$$\nu = \frac{\left(\sigma_x^2 + \sigma_y^2 \right)^2}{\sigma_x^4 + \sigma_y^4}$$

$$CEP_4 = .565\ \sigma_{max} + .612\ \sigma_{min},\ \sigma_{min}/\sigma_{max} \geq .25 \quad (11)$$

$$= .667\ \sigma_{max} + .206\ \sigma_{min},\ \sigma_{min}/\sigma_{max} < .25$$

$$CEP_5 = \left[2^{\frac{1}{3}}\left(1 - \frac{2}{9\nu}\right)\right]^{\frac{3}{2}} \left(\frac{\sigma_x^2 + \sigma_y^2}{2}\right)^{\frac{1}{2}} . \quad (12)$$

CEP_1 and CEP_2 were taken from Groves (1961), CEP_3 was established by Grubbs (1964), CEP_4 is a piece-wise linear combination of the standard deviations, and CEP_5 was also established by Grubbs (1964) using a Wilson-Hilferty transformation of the chi-square in CEP_3. Plots of each approximation versus the true CEP as a function $\sigma_{min}/\sigma_{max}$ are shown in Figures 1 through 5. These give a fairly good indication of how well each performs. It is seen that CEP_1 deteriorates rapidly as we depart from the circular case (for which CEP_1 degenerates to 1.1774σ), CEP_2 is reasonably good if the ratio $\sigma_{min}/\sigma_{max}$ is not less than about .2, CEP_3 appears good for all ratios, and CEP_4 and CEP_5 appear good to a lesser extent for all ratios.

If these approximations were used only as approximations for assumed values of the error variances (as one does in wargaming and round requirement studies), then there would be no estimation problem. However, in many cases, weapons analysts are using estimates of the variances in these approximations (based on sample data) to form estimates of CEP. Hence, the problem now becomes an estimation problem instead of an approximation problem. In particular, the problem addressed in this paper is that of comparing the five estimators for CEP formed by replacing the population variances in equations (8) through (12) with sample variances $S_x^2 = \Sigma X_i^2/n$ and $S_y^2 = \Sigma Y_i^2/n$. (In these expressions, X_i and Y_i are the recorded errors in the X and Y directions, respectively, for the ith impact and n is the number of sample impacts.) These estimators will be referred to as $\hat{CEP}_1$ through $\hat{CEP}_5$ in the discussion which follows.

Methodology

Measures of comparison employed in this study were the mean squared error (MSE), expected confidence interval length, and confidence interval confidence. With regard to the former, the MSE of an estimator $\hat{\theta}$ for a parameter θ is defined in the usual sense, i.e.,

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = V(\hat{\theta}) + B^2(\hat{\theta})$$

where E represents expectation, V represents variance and B represents bias. It was chosen because it accounts for bias as well as variance and all five estimators are biased for CEP except in the degenerate circular case. With regard to the second measure, it was chosen because it too accounts for bias as well as variance but in the sense of interval estimation vice point. These computations were based on approximate distributions of CEP estimators and did not provide intervals with specified confidence in all cases. Hence, the third measure was included to estimate the true confidence.

The computation of these measures was straightforward but not simple due to the complexity of the estimators. Recall that 1, 3 and 5 each involve radicals of linear combinations of sample variances and estimators 2 and 4 involve linear combinations of sample standard deviations. Hence, the sampling distributions were approximated. The approximations were achieved by matching the variance of estimators 3 and 5 with the variance of the chi-square distribution and by matching the variance of estimators 2 and 4 with the variance of the chi distribution. Estimator 1 was simply approximated by a chi-square with $2n$ degrees of freedom. This distribution is exact only in the circular case and was included to show how poorly it becomes when eccentricity of the distribution increases. The approximations are shown in Figure 6 and are discussed in more detail in the next paragraph.

Figure 6 provides a summary of approximate distributions for each $\hat{CEP}_i$ and defines several multiplicative factors, K_i, eccentricity c and degrees of freedom ν and ν'. ν^* does not have a simple form and is described below.

Because estimators 3 and 5 are of the same general form, the distribution of the squares of both was approximated by matching the variance of

$$\frac{\nu' \ (S_x^2 + S_y^2)}{(\sigma_x^2 + \sigma_y^2)} \tag{13}$$

with $2\nu'$, the variance of a chi-square with ν' degrees of freedom. It was found that $\nu' = n\nu$ where $\nu = \frac{(c^2 + 1)^2}{c^4 + 1}$. Expression (13) can be rewritten as

$$\frac{\nu' \ \hat{CEP}_i^2}{\sigma_y^2 \ K_i\left(\frac{c^2 + 1}{2}\right)} \qquad \text{where } i = 3 \text{ or } 5 \tag{14}$$

to conform to the expressions in Figure 6.

Estimators 2 and 4, representing linear combinations of the standard deviations, were approximated by matching the variance of a chi with ν^* degrees of freedom with the variance of

$$\frac{(\nu^*)^{1/2} \ (S_x + S_y)}{\sigma_x + \sigma_y} \tag{15}$$

The variance of expression (15) was found to be

$$\nu^* \ (1 - H^2(n)) \ \frac{1+c^2}{(1+c)^2} \tag{16}$$

and the variance of a chi with ν^* degrees of freedom is

$$\nu^*(1 - H^2(\nu^*))$$

where $H(x) = \sqrt{\frac{2}{x}} \frac{\Gamma\left(\frac{x+1}{2}\right)}{\Gamma\left(\frac{x}{2}\right)}$

Upon equating the two, we find that ν^* satisfies

$$H(\nu^*) = \left\{1 - (1 - H^2(n)) \frac{1+c^2}{(1+c)^2}\right\}^{\frac{1}{2}}$$

and can be obtained from a table of inverse solutions of the H function.

The approximate distributions allow one to derive approximate mean squared errors for the $\widehat{CEP}_i$ estimators which are given in Figure 7. The K_i coefficients are defined as before and Bias($\widehat{CEP}_i$) is defined as

$$E(\widehat{CEP}_i) - \text{True } CEP_i \text{ for } i = 1, 2, \ldots, 5.$$

In general, σ_y is a scale factor representing the maximum σ value; however, in the examples given here σ_y is always equal to 1. Note that MSE($\widehat{CEP}_2$) and MSE($\widehat{CEP}_4$) can be expressed in exact rather than approximate form.

Since a point estimate may not provide adequate information, approximate 95% confidence intervals were constructed for each estimator using the distributions discussed above. The approximate 100(1-α)% confidence limits for CEP are given by

$$\left[\frac{\widehat{CEP}_i}{\left(\chi^2_{\nu_i,1-\alpha/2}/\nu_i\right)^{\frac{1}{2}}}, \frac{\widehat{CEP}_i}{\left(\chi^2_{\nu_i,\alpha/2}/\nu_i\right)^{\frac{1}{2}}}\right]$$

where $\widehat{CEP}_i$ is the ith estimator and ν_i equals the degrees of freedom associated with $\widehat{CEP}_i$. Expected confidence interval widths can then be computed and used as measures of comparison between estimators. Clearly, if one could compute exact 95% confidence intervals, comparison of interval widths would be straightforward. However, only approximate intervals can be obtained and the confidence associated with each interval must be computed before a complete evaluation can be made.

Confidence was estimated using 10,000 Monte Carlo replicates for samples of size 5, 10 and 20 and measuring the percentage of time the true CEP fell within the interval. Confidence and expected confidence interval widths were then jointly examined.

Results

The object of this study was to examine and evaluate the behavior of several candidate CEP estimators over a wide range of conditions. Sample sizes ranged from 5 to 400 and eccentricities ranged from $c = 1$, the circular case to $c = 20$, a highly elliptical case. Extreme values of the sample size and eccentricity may be infrequently encountered but were included for completeness. Clearly, an estimator behaving poorly under circumstances unlikely to be observed should not be disregarded as a viable candidate.

Prior to determining approximate distributions and mean squared error (MSE) approximations for the estimators, a Monte Carlo simulation was developed for computing the variance, bias, average squared error (ASE) and standard error for each estimator at each of three sample sizes ($n = 5, 10, 20$). The simulated ASE's were used as a check against MSE approximations which were subsequently computed.

Upon comparing the simulated ASE's against results of the MSE approximations for sample sizes 5, 10, and 20, it became evident that MSE approximations were inadequate for estimators 3 and 5. In fact, in the mid-range of the eccentricity, c, the MSE for 3 and 5 differed from the simulated values of ASE by as much as three times the standard error. For this reason, the simulated ASE values are presented in Figure 8 while the approximate MSE values, found suitable for larger sample sizes, are shown in Figure 9.

Despite some fluctuation at $c = .05$, Figures 8 and 9 show estimators 2 through 5 producing fairly close results. As the sample size increased, estimator 4 exhibited the smallest mean squared error and appeared to be the most satisfactory point estimator.

Figures 10 and 11 contain expected confidence interval width and confidence interval confidence, respectively. If the computation of exact 95% confidence intervals were possible, a straightforward selection of the estimator producing the narrowest width could be made. However, the approximate confidence intervals have varying levels of confidence associated with them, all of which underestimate or overestimate the desired 95% level. It appears that the wider lengths are associated with higher confidence and the narrower widths with the lower confidences so that a true comparison is not really possible. However, it is evident that estimators 2 through 5 do not distinguish themselves as being far superior or grossly inferior to one another. This is essentially the same result obtained from the MSE comparisons.

In summary, unless c is very small, estimators 2 through 5 produce reasonably close results. If confidence intervals are not desired, estimator 4 would be an acceptable choice. Otherwise, estimator 3 is recommended due to ease of confidence interval computability.

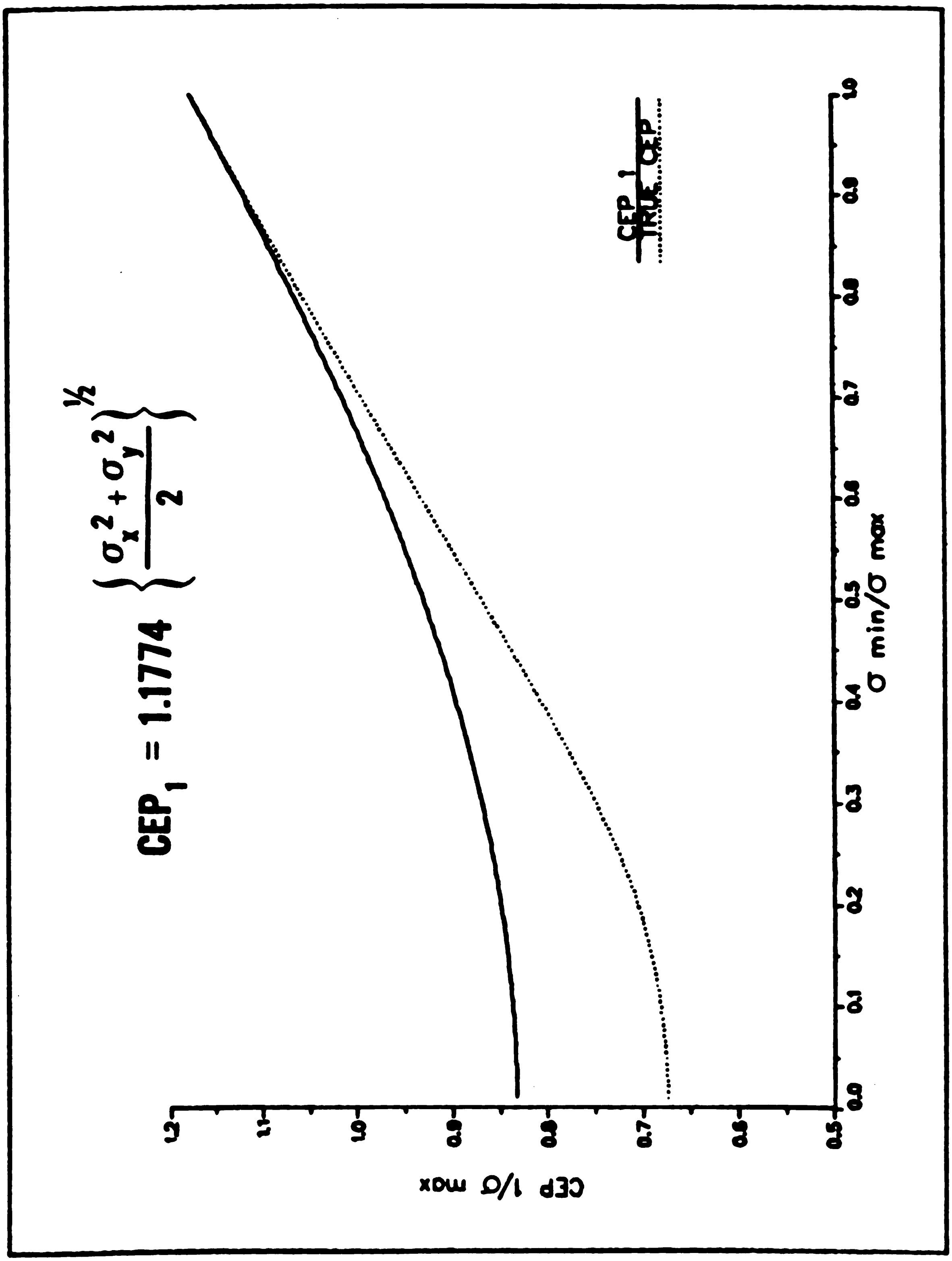
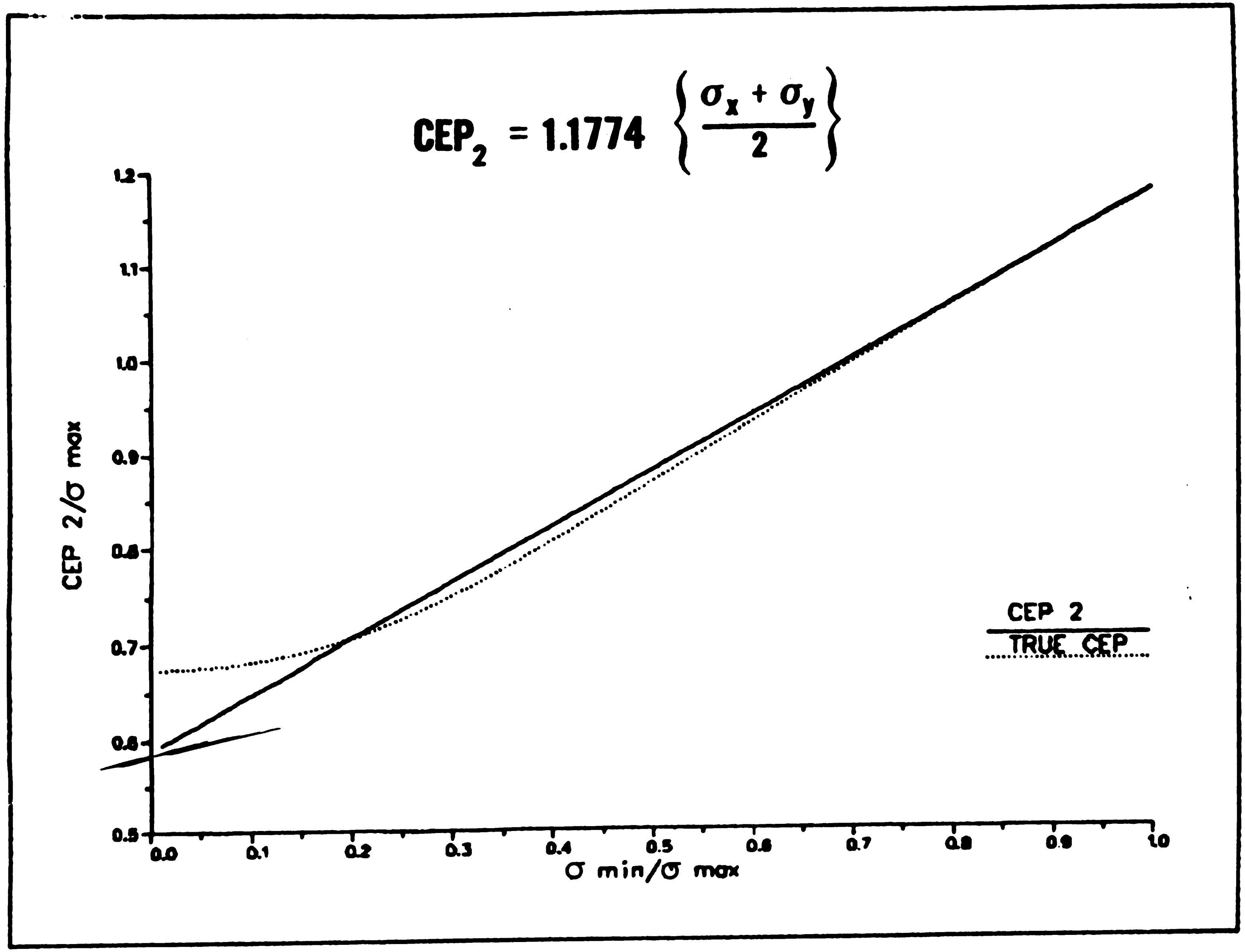
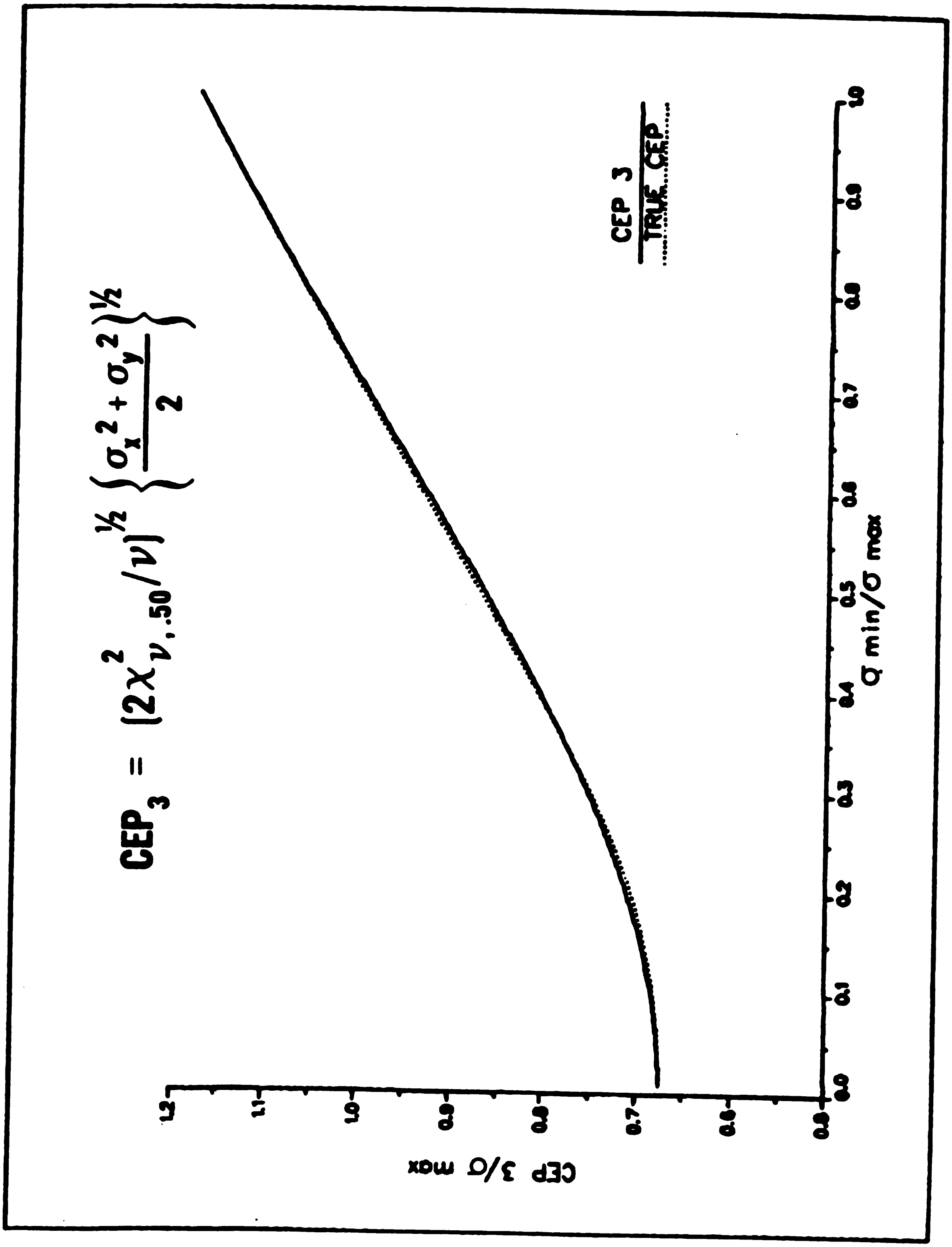

$CEP_1 = 1.1774 \left\{\frac{\sigma_x^2 + \sigma_y^2}{2}\right\}^{1/2}$
CEP 1
TRUE CEP
σ min/σ max
CEP 1/σ max
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2

Figure 1



$CEP_2 = 1.1774 \left\{ \frac{\sigma_x + \sigma_y}{2} \right\}$
CEP 2/σ max
σ min/σ max
CEP 2
TRUE CEP
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

Figure 2



$CEP_3 = [2\chi^2_{\nu,.50}/\nu]^{1/2} \left\{\frac{\sigma_x^2+\sigma_y^2}{2}\right\}^{1/2}$
CEP 3
TRUE CEP
CEP 3/σ max
σ min/σ max
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0.6
0.7
0.8
0.9
1.0
1.1
1.2

Figure 3

$$CEP_4 = .565\,\sigma_{max} + .612\,\sigma_{min}\ , \quad \frac{\sigma_{min}}{\sigma_{max}} \geq .25$$

$$ = .667\,\sigma_{max} + .206\,\sigma_{min}\ , \quad \frac{\sigma_{min}}{\sigma_{max}} < .25$$

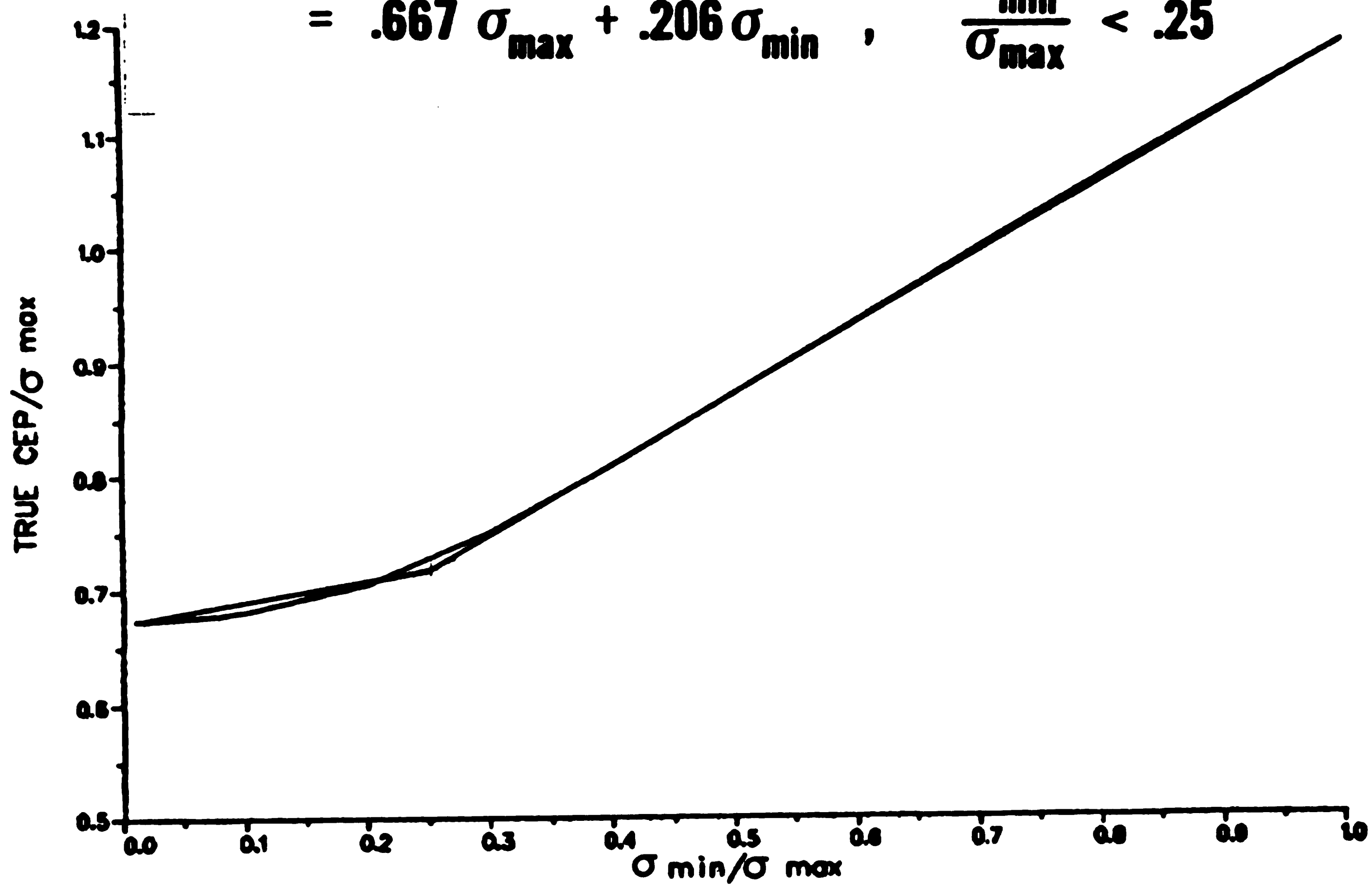


Figure 4

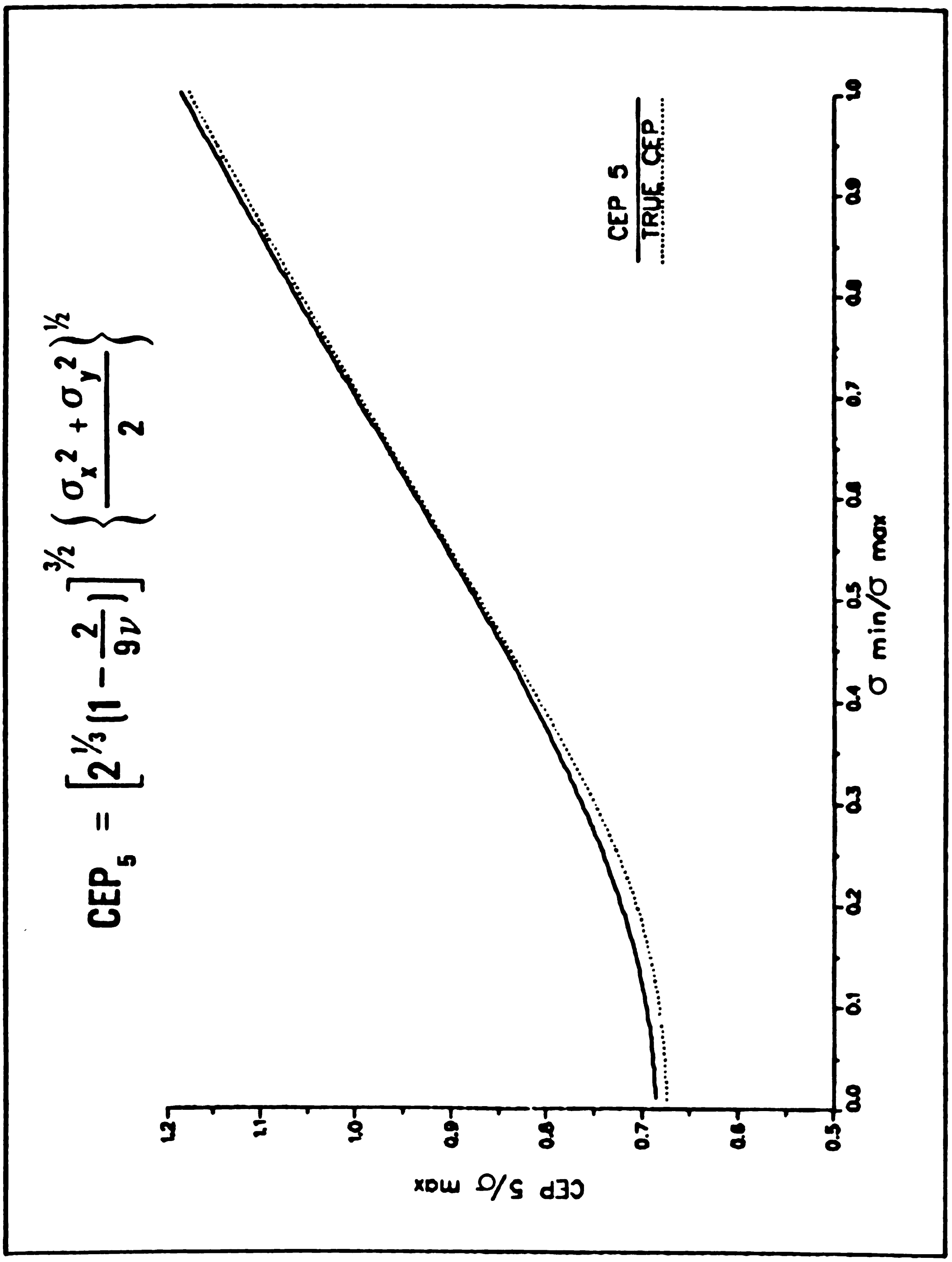

$CEP_5 = \left[2^{1/3}\left(1 - \frac{2}{9\nu}\right)\right]^{3/2} \left\{\frac{\sigma_x^2 + \sigma_y^2}{2}\right\}^{1/2}$
CEP 5
TRUE CEP
CEP 5/σ max
σ min/σ max
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

Figure 5

Approximate Distributions of Estimators

(1) $$\frac{2n\,\hat{CEP}_1^2}{\sigma_Y^2\,K_1^2\left(\frac{c^2+1}{2}\right)} \sim \chi^2_{2n}$$

(2) $$\frac{\nu^*\,\hat{CEP}_2^2}{\sigma_y^2\,K_1^2\left(\frac{c+1}{2}\right)^2} \sim \chi^2_{\nu^*}$$

(3) $$\frac{\nu'\,\hat{CEP}_3^2}{\sigma_y^2\,K_3^2\left(\frac{c^2+1}{2}\right)} \sim \chi^2_{\nu'}$$

(4) $$\frac{\nu^*\,\hat{CEP}_4^2}{\sigma_y^2\,(g_1+g_2c)^2} \sim \chi^2_{\nu^*}$$

(5) $$\frac{\nu'\,\hat{CEP}_5^2}{\sigma_y^2\,K_5^2\left(\frac{c^2+1}{2}\right)} \sim \chi^2_{\nu'}$$

$K_1 = 1.1774$ $\qquad K_3 = (2\,\chi_{\nu,.50}/\nu)^{\frac{1}{2}}$ $\qquad K_5 = \left[2^{\frac{1}{3}}\left(1-\frac{2}{9\nu}\right)\right]^{\frac{3}{2}}$

$c = \sigma_x/\sigma_y \leq 1$ $\qquad \nu = \frac{(c^2+1)^2}{c^4+1}$ $\qquad \nu' = n\nu$

$g_1 = .565$, $g_2 = .612$ when $c \geq .25$

$g_1 = .667$, $g_2 = .206$ when $c < .25$

Figure 6

MEAN SQUARED ERRORS

$$MSE(\hat{CEP}_1) \approx \sigma_y^2 K_1^2 \left(\frac{C^2+1}{2}\right)\left[1-H^2(2n)\right]+\left[Bias(\hat{CEP}_1)\right]^2$$

$$MSE(\hat{CEP}_2) = \sigma_y^2 K_1^2 \left(\frac{C^2+1}{4}\right)\left[1-H^2(n)\right]+\left[Bias(\hat{CEP}_2)\right]^2 \quad *$$

$$MSE(\hat{CEP}_3) \approx \sigma_y^2 K_3^2 \left(\frac{C^2+1}{2}\right)\left[1-H^2(\nu')+\left[Bias(\hat{CEP}_3)\right]^2$$

$$\nu' = n\nu \qquad \nu = \frac{(C^2+1)^2}{C^4+1}$$

$$MSE(\hat{CEP}_4) = \sigma_y^2 (g_1^2+g_2^2 C^2)\left[1-H^2(n)\right]+\left[Bias(\hat{CEP}_4)\right]^2 \quad *$$

$$MSE(\hat{CEP}_5) \approx \sigma_y^2 K_5^2 \left(\frac{C^2+1}{2}\right)\left[1-H^2(\nu')\right]+\left[Bias(\hat{CEP}_5)\right]^2$$

* Exact

Figure 7

AVERAGE SQUARED ERROR

N = 5

C	ASE 1	ASE 2	ASE 3	ASE 4	ASE 5	S.E.
1.0	.068	.069	.070	.069	.069	.003
.75	.056	.053	.053	.053	.053	.002
.50	.056	.042	.041	.041	.042	.002
.35	.063	.037	.038	.038	.039	.002
.20	.073	.035	.039	.040	.041	.002
.05	.079	.041	.044	.043	.045	.002

N = 10

C	ASE 1	ASE 2	ASE 3	ASE 4	ASE 5	S.E.
1.0	.034	.035	.035	.035	.035	.002
.75	.029	.027	.027	.027	.027	.002
.50	.031	.021	.021	.021	.021	.001
.35	.038	.019	.019	.019	.020	.001
.20	.048	.018	.020	.021	.021	.001
.05	.053	.022	.022	.022	.023	.001

N = 20

C	ASE 1	ASE 2	ASE 3	ASE 4	ASE 5	S.E.
1.0	.017	.017	.017	.017	.017	.001
.75	.014	.013	.013	.013	.013	.001
.50	.017	.011	.010	.010	.010	.001
.35	.024	.010	.009	.009	.010	.001
.20	.034	.009	.010	.011	.011	.001
.05	.039	.013	.011	.011	.012	.001

Figure 8

MEAN SQUARED ERROR

N = 50

C	MSE 1	MSE 2	MSE 3	MSE 4	MSE 5
1.0	.0069	.0069	.0069	.0069	.0070
.75	.0055	.0054	.0057	.0054	.0057
.50	.0077	.0044	.0051	.0042	.0052
.35	.0143	.0041	.0049	.0037	.0050
.20	.0235	.0036	.0047	.0036	.0050
.05	.0275	.0072	.0045	.0042	.0048

N = 100

C	MSE 1	MSE 2	MSE 3	MSE 4	MSE 5
1.0	.0035	.0035	.0035	.0035	.0035
.75	.0029	.0027	.0028	.0027	.0029
.50	.0057	.0023	.0026	.0021	.0026
.35	.0126	.0022	.0024	.0019	.0026
.20	.0220	.0018	.0023	.0018	.0026
.05	.0261	.0053	.0023	.0021	.0025

N = 400

C	MSE 1	MSE 2	MSE 3	MSE 4	MSE 5
1.0	.0009	.0009	.0009	.0009	.0009
.75	.0009	.0007	.0007	.0007	.0007
.50	.0042	.0007	.0007	.0005	.0007
.35	.0113	.0008	.0006	.0005	.0007
.20	.0209	.0005	.0006	.0005	.0008
.05	.0251	.0039	.0006	.0005	.0007

Figure 9

CONFIDENCE INTERVAL LENGTHS

N = 5

C	CL 1	CL 2	CL 3	CL 4	CL 5
1.0	1.213	1.160	1.305	1.145	1.317
.75	1.071	1.028	1.175	1.014	1.186
.50	.950	.923	1.118	.918	1.131
.35	.897	.889	1.129	.917	1.144
.20	.856	.876	1.149	.985	1.168
.05	.841	.912	1.176	1.054	1.196

N = 10

C	CL 1	CL 2	CL 3	CL 4	CL 5
1.0	.792	.776	.817	.768	.823
.75	.698	.685	.735	.676	.741
.50	.622	.614	.697	.601	.705
.35	.587	.586	.693	.586	.702
.20	.564	.573	.694	.636	.706
.05	.553	.579	.695	.662	.707

N = 20

C	CL 1	CL 2	CL 3	CL 4	CL 5
1.0	.537	.533	.545	.531	.549
.75	.475	.472	.492	.467	.496
.50	.423	.421	.465	.411	.470
.35	.400	.401	.460	.393	.466
.20	.384	.388	.455	.429	.462
.05	.378	.387	.453	.441	.461

Figure 10

SIMULATED CONFIDENCE LEVELS

N = 5

C	PROB 1	PROB 2	PROB 3	PROB 4	PROB 5
1.0	.950	.947	.963	.946	.963
.75	.941	.947	.965	.945	.965
.50	.894	.941	.968	.944	.967
.35	.830	.937	.963	.939	.961
.20	.753	.932	.952	.923	.950
.05	.714	.930	.950	.936	.948

N = 10

C	PROB 1	PROB 2	PROB 3	PROB 4	PROB 5
1.0	.947	.945	.955	.944	.955
.75	.935	.944	.958	.943	.959
.50	.876	.941	.967	.943	.966
.35	.789	.939	.967	.941	.965
.20	.689	.939	.959	.931	.955
.05	.640	.931	.951	.943	.948

N = 20

C	PROB 1	PROB 2	PROB 3	PROB 4	PROB 5
1.0	.952	.952	.956	.951	.956
.75	.938	.951	.960	.951	.960
.50	.858	.945	.970	.948	.969
.35	.724	.942	.970	.947	.968
.20	.567	.946	.961	.937	.955
.05	.506	.912	.950	.946	.946

Figure 11

WIND VARIABILITY IN THE BOUNDARY LAYER AND ITS ASSOCIATION WITH TURBULENCE, RED AND WHITE NOISE

Oskar M. Essenwanger
Aerophysics Group
Research Directorate
US Army Missile Laboratory
US Army Missile Command
Redstone Arsenal, AL 35898

ABSTRACT. In the boundary layer, wind fluctuations during short time intervals in the magnitude of seconds are commonly interpreted in terms of turbulence theory. According to Kolmogorov, Obukhov and Corrsin, this requires a -5/3 slope in a diagram of the power spectrum versus the wave length in double logarithmic coordinates. The turbulence characteristics fade with averaging time of the wind measurements.

Investigations of one- and six-second tower measurements at Redstone Arsenal and Otis Air Force Base revealed, however, that these small scale fluctuations more frequently show associations with white or red noise rather than turbulence characteristics in the spectrum. It will be illustrated that the turbulence slope is a special case of red noise.

I. INTRODUCTION. In the past two decades wind sensors with high sensitivity and short time responses were developed. These instruments made it possible to measure wind fluctuations on the scale of seconds or even for shorter time intervals. In the boundary layer these short time fluctuations are mostly analyzed or interpreted in terms of frictional or convective turbulence theory. It is sometimes overlooked, however, that these fluctuations may be random (white) noise or simply include persistence which produces a red noise spectrum.

In the subsequent sections a detailed analysis of power spectra of the wind is performed for observations from one- and six-second tower measurements at Redstone Arsenal, Alabama, and Otis AFB, Mass., by Gill u-v-w anemometers (see Gill, 1975). It is delineated that the slope of some spectra (in double logarithmic coordinates) differs from the expected -5/3 law of turbulence (inertial subrange), but it follows a pattern of persistence.

The slope of power spectra for red noise was studied. It can be shown that a first lag correlation of 0.85 and exponential red noise produce a slope of -5/3 in the spectra for medium and short waves, although for the empirical data in this investigation the first lag correlation producing a -5/3 slope was slightly higher (see Figure 3). Red noise in the data series can produce the same slope as turbulence, although turbulence and red noise spectra are not identical and differ in the region of long waves and at the very tail end of short wave length.

Finally, the relationship in time averaging of wind components for data from Redstone and Otis is analyzed. As expected, the variance of the data decreases with increasing average time. The first lag correlations for the averaged data sets, however, do not display a uniform pattern.

II. THE POWER SPECTRUM OF TURBULENCE. It is customary to present the time fluctuation of the U-component of the wind (in direction of the mean flow) by:

$$U = U_s + U_t \tag{1}$$

where U_s is the "stationary" part and U_t the "turbulent fluctuation." The other wind components, V and W, follow in analogy.

The stationary part is usually the mean value, $U_s = \overline{U}$. If $\overline{U}$ changes in time, the series must be considered as nonstationary and U_s may represent some changing quantity, although the change may be slow. Nonstationary time series may show complex patterns, although the change in U_s would not affect the part of a power spectrum towards waves of short duration. Techniques to separate U_s from U_t have been developed (e.g., Essenwanger and Billions, 1965). For this investigation, U_s is considered to be constant over a homogeneous time period.

The turbulent fluctuation requires (by definition of $U_s = \overline{U}$)

$$\Sigma\, U_t/N = \overline{U}_t = 0 \tag{2}$$

but the variance:

$$\Sigma\, U_t^2/N = \sigma_{U_t}^2 \neq 0 \tag{3}$$

In addition, for turbulence in the inertial subrange, the energy spectrum as a function of the (standardized) wave number K follows:

$$E(K) = \alpha\varepsilon^{2/3}\, K^{-5/3}$$

according to Kolmogorov-Obukhov-Corrsin (e.g., Tennekes and Lumley, 1973, p. 266; Priestly, 1959, p. 61; Hinze, 1959, p. 194, etc.). This leads to a squared standardized amplitude L_j of a Fourier series for the time series of turbulence data:

$$\ln E(K) = \ln L_j = \text{const} -(5/3) \ln K. \tag{4}$$

(For L_j of the power spectrum, see Tukey, 1949; Blackman and Tukey, 1958; or Cooley and Tukey, 1965.) Thus, L_j will show a slope of -5/3 in double logarithmic coordinates with L_j as the ordinate and ln K as the abscissa. Equation (4) is referred to in this study whenever the slope angle is discussed.

Wind observations recorded at one-second time intervals with the u-v-w Gill anemometers (see Gill, 1975) were made in 1974 on towers at Redstone Arsenal, Alabama, and the power spectrum of the ΔU component for a combination of levels was established. Figure 1 provides an example for the power spectrum ΔU of the windshear from 18- to 30-foot heights, where ΔU is the difference of the U-components from two respective heights. The slope of the power spectrum (obtained by regression analysis) shows -1.63, which is close to -5/3. The measurements over a one-hour time period fulfilled the definition of stationary data.

Later recordings, at six-second time intervals, were obtained (1981) from Otis AFB, but the power spectra of the ΔU windshears did not delineate a -5/3 slope for this set of data. A detailed analysis of the U-components themselves revealed that the wind components at the individual levels do not follow a power spectrum of -5/3. Consequently, it may not be expected that the power spectra of the windshear show a -5/3 slope. An example for the power spectrum (double logarithmic coordinates) of the U-component at an altitude of 10 feet for the tower data set under investigation is illustrated in Figure 2.

At first it was considered that the difference in the slope angle could be attributed to the sampling rate. Some change in the angle may be expected for an assumption of red noise (see Section III), but turbulence in the inertial subrange would require that the slope should still be around -5/3.

The power spectrum (Figure 2) is truncated at waves of 12-second duration compared with Figure 1, which extends to two seconds, but the slope should not have changed that significantly. Corresponding with Figure 1, spectral values from wave length corresponding to 30 seconds to 12 seconds do not produce a -5/3 slope in Figure 2.

One may suggest that the 25 November 1981 data stem from a nonstationary time series. Indeed, there is a break in $\overline{U}$ at about the middle of the time series. However, as pointed out previously, fluctuation of U_S would affect the amplitudes of waves of longer duration and should not alter the tail end of the spectrum for waves with short duration. Thus, nonstationary behavior is not reflected in changes of the wave spectrum of shorter length. A split of the data into the two different periods provided the same pattern as illustrated in Figure 2.

A further possibility for the lower slope angle in Figure 2 data is the assumption that the data do not show turbulence in the inertial subrange, or the fluctuations are governed by different laws. As an alternative hypothesis, a white or red noise pattern will be studied next.

III. WHITE AND RED NOISE. It is common knowledge that meteorological time series show persistence. Persistence in signal processing is usually called red noise, while white noise indicates independence in the data sample. White noise leads to a power spectrum where all (squared) amplitudes are equal except for random fluctuations. Red noise can be expressed by several models, although two primary forms have been customary in geoscience (e.g., Taubenheim, 1969).

In various texts the autocorrelogram in turbulence analysis is assumed to be Gaussian noise (e.g., Tennekes and Lumley, 1973; Hinze, 1959, etc.).

FIG. 1

POWER SPECTRUM
(19 AUGUST 1974, 1 SEC RECORDING)
Δu = LEVEL 3 – LEVEL 2
MAX LAG 60 SECONDS

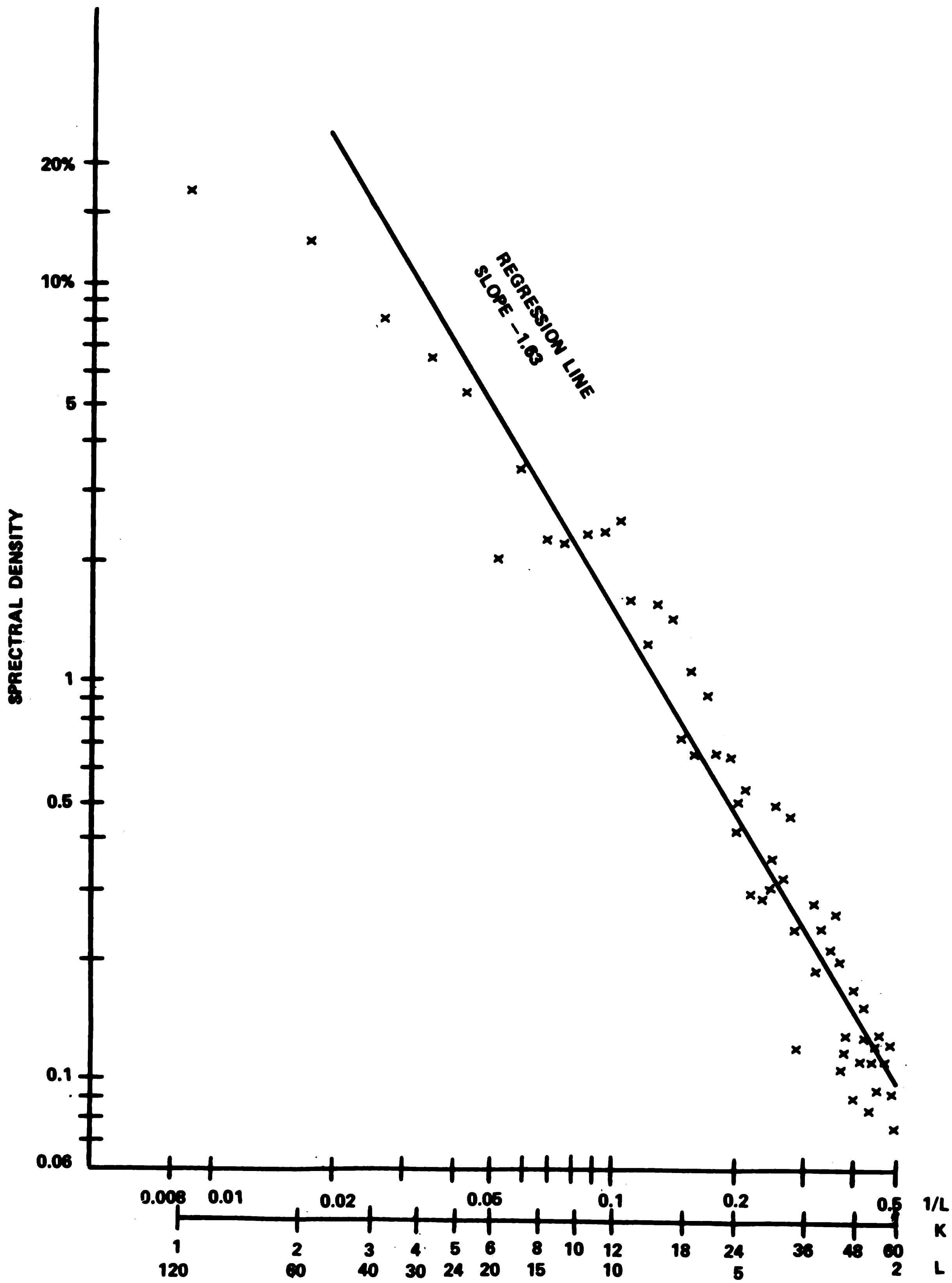
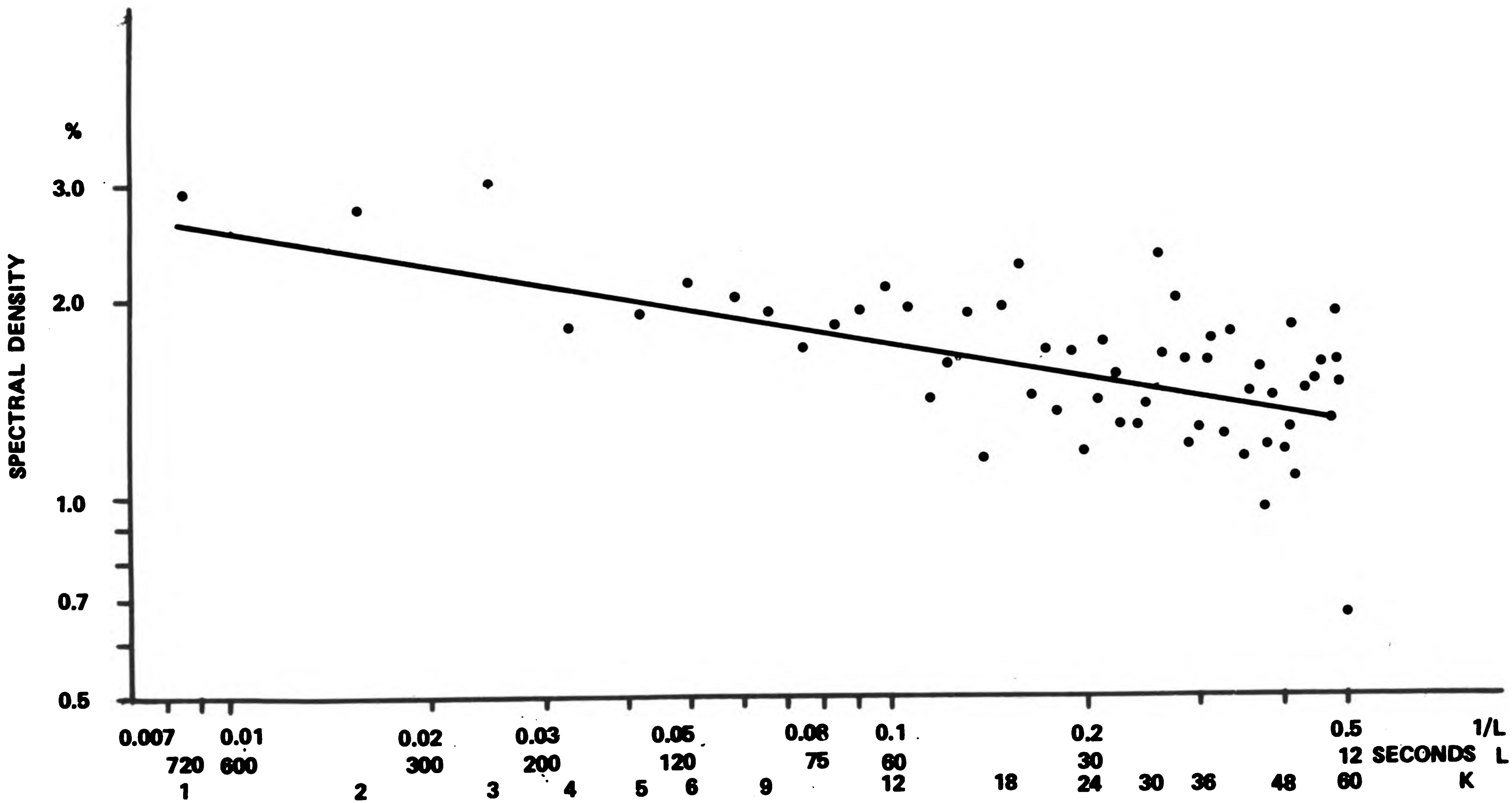


FIG. 2

POWER SPECTRUM
(25 NOVEMBER 1981, 6 SEC RECORDING)
Δu, = LEVEL 3 – LEVEL 2
MAX. LAG 60 SECONDS



This form was ruled out by the author for the investigated data because of dissimilarity in the correlogram and spectrum, and the use of the exponential red noise model by other authors such as Pasquill (1962), Fichtl and McVehil (1970), or Hanna (1979). (See also Stewart, 1981.)

The form chosen here is the exponential red noise which is identical with a first order Markov chain. In terms of an autocorrelogram we can write:

$$\rho_k = \rho_c^k \qquad (5)$$

where ρ_c is the first lag correlation and $k = 0,\ldots,m$. This series is identical with:

$$\rho_t = \exp(-bt) \qquad (6)$$

where $t \geq 0$, $b = -\ln \rho_c > 0$, $\rho_c > 0$ (see Box and Jenkins, 1970, or Essenwanger, 1980). The corresponding power spectrum is:

$$L_j = (2/b) / [1 + \pi^2 j^2/b] \qquad (7)$$

(see Taubenheim, 1969, etc.). Gilman et al. (1963) derived:

$$L_k = [(1 - \rho_c^2) / (1 + \rho_c^2 - 2\,\rho_c \cos k\pi/m] / m \qquad (8)$$

which is a good approximation for $\rho_c \leq 0.9$.

The power spectrum of white noise can be stated as:

$$L_k = \text{const} \qquad (9)$$

and is produced from $\rho_0 = 1$, $\rho_k = 0$ for all $k > 0$.

Figure 3 represents the power spectrum of (exponential) red noise for a variety of first lag correlations from $\rho_c = 0.9$ down to 0.1 for a normalized maximum lag of $m = 60$. The power spectra show a variety of slope angles ranging from -1.78 for $\rho_c = 0.9$ to about -0.20 for $\rho_c = 0.1$. Thus, the power spectrum exhibited in Figure 2 could come primarily from data associated with red noise of $\rho_c \simeq 0.2$. If the red noise data concept is the generating background of the wind fluctuation, then the slope in the power spectrum would change with a change in the sampling time interval.

Figure 4 provides an example of the autocorrelogram of data taken at Redstone (19 May 1974) at one-second sampling time and the corresponding series of the same data at six-second time intervals. In the latter, every sixth correlation coefficient appears. While the first lag correlation for one-seconds is $\rho_1 = 0.97$, the value drops to $\rho_1 = 0.80$ for six seconds. A corresponding change in the slope of the power spectrum is observed. Under the assumption of exponential red noise, the change in angles is very small. It amounts to about four degrees. (From $\rho_c = 0.97$ to $\rho_c = 0.8$, the slope reduces from -1.85 to -1.56, or in angles from 61.6° to 57.3°.)

FIG. 3

POWER SPECTRA OF EXPONTENTIAL
RED NOISE

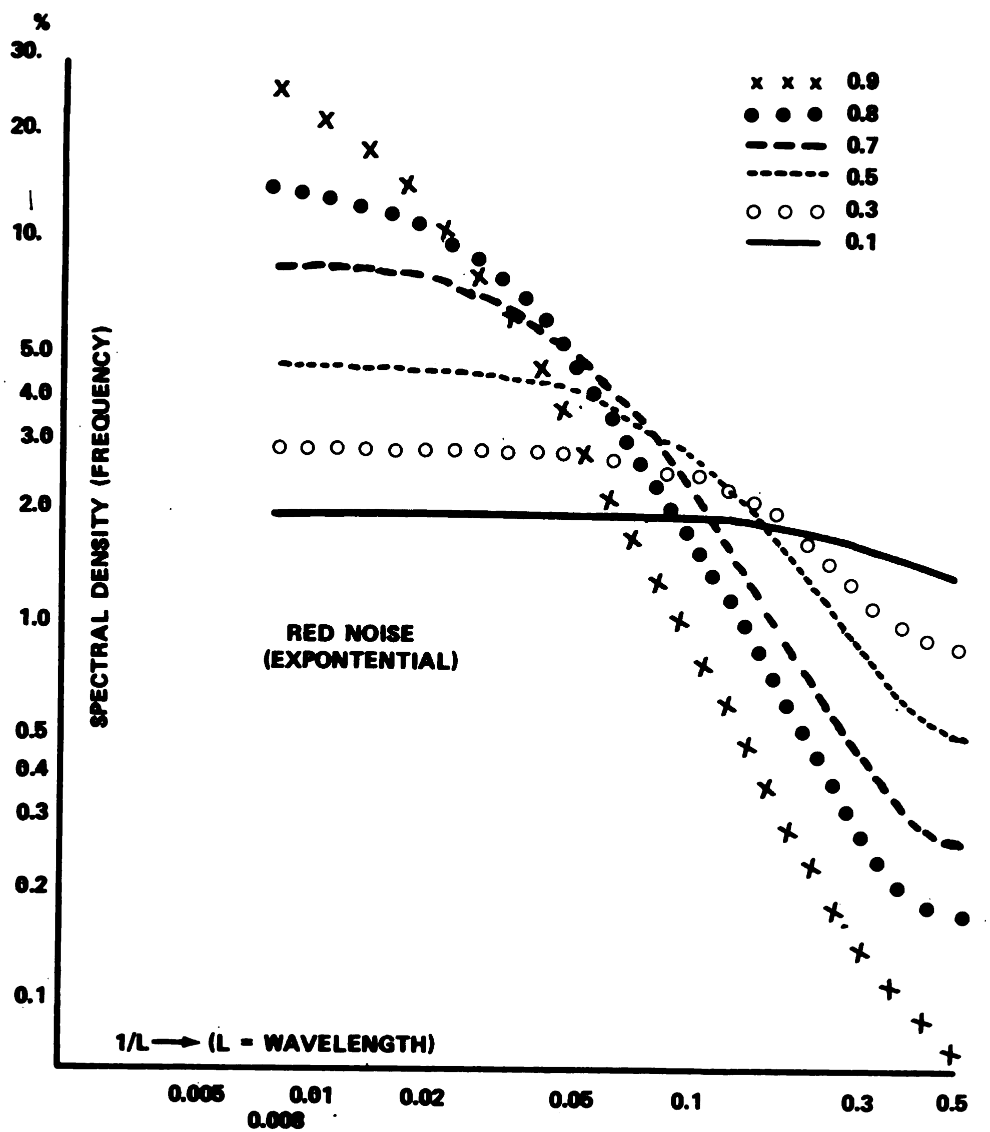


FIG. 4

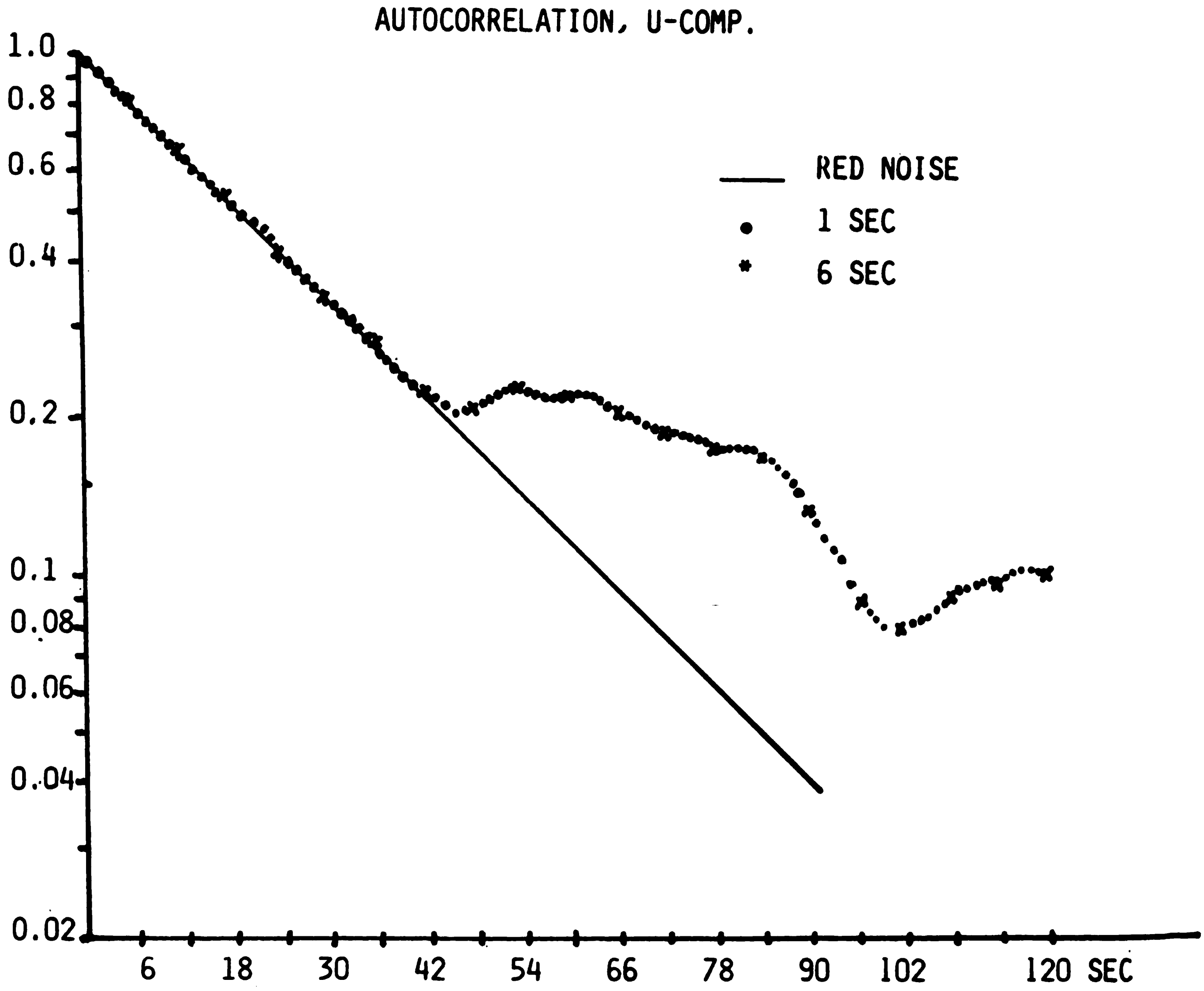

AUTOCORRELATION, U-COMP.
RED NOISE
1 SEC
6 SEC
1.0
0.8
0.6
0.4
0.2
0.1
0.08
0.06
0.04
0.02
6
18
30
42
54
66
78
90
102
120 SEC

Figure 4 supports the conjecture that exponential red noise is a primary factor in the 1974 data series. Only correlation coefficients ρ_k for $k > 50$ deviate from the pattern. These coefficients $k > 50$ are not significantly different from zero, however, although they appear dissimilar to the exponential red noise line.

Figure 5 exhibits the empirical relationship between the slope angle and the first lag correlation ρ_c of the U-component and ΔU of observed wind data. These 37 samples were prepared from power spectra of the 19 August 1974 (Redstone) and 25 November 1981 (Otis AFB) data. The regression line deviates only slightly from the expected line of pure red noise (Figure 3). Six samples (dots) of the 37 samples represent the slopes of time difference windshears (data 19 August 1974):

$$\Delta_t S = [(\Delta_t U)^2 + (\Delta_t V)^2]^{\frac{1}{2}} \qquad (10)$$

where Δ_t denotes the time difference. In order to create independent data sets, the original data series was reduced by accepting only every t-th value. The autocorrelogram for $\Delta_t = 8$ seconds showed a $\rho_1 = 0.65$, which slowly decreased to $\rho_{60} = 0.48$. Thus, the very slow decline produces the outlier ($\rho_c = 0.65$, slope angle -5°) and can be explained as a truncated but non-exponential autocorrelogram. However, the physical interpretation of the high persistence in this particular data series needs further investigation.

The relationship in Figure 5 and a comparison with Figure 3 render the conclusion that ρ_c is about 0.85 if the slope should agree with the required slope of turbulence in the inertial subrange. Consequently, a slope angle of -5/3 can also be produced by a data series whose persistence factor agrees with an exponential red noise model of ρ_c between 0.8 to 0.9.

In turbulence analysis an integral scale is defined:

$$\overline{R}_s = \int_0^\infty \rho(t)dt \qquad (11)$$

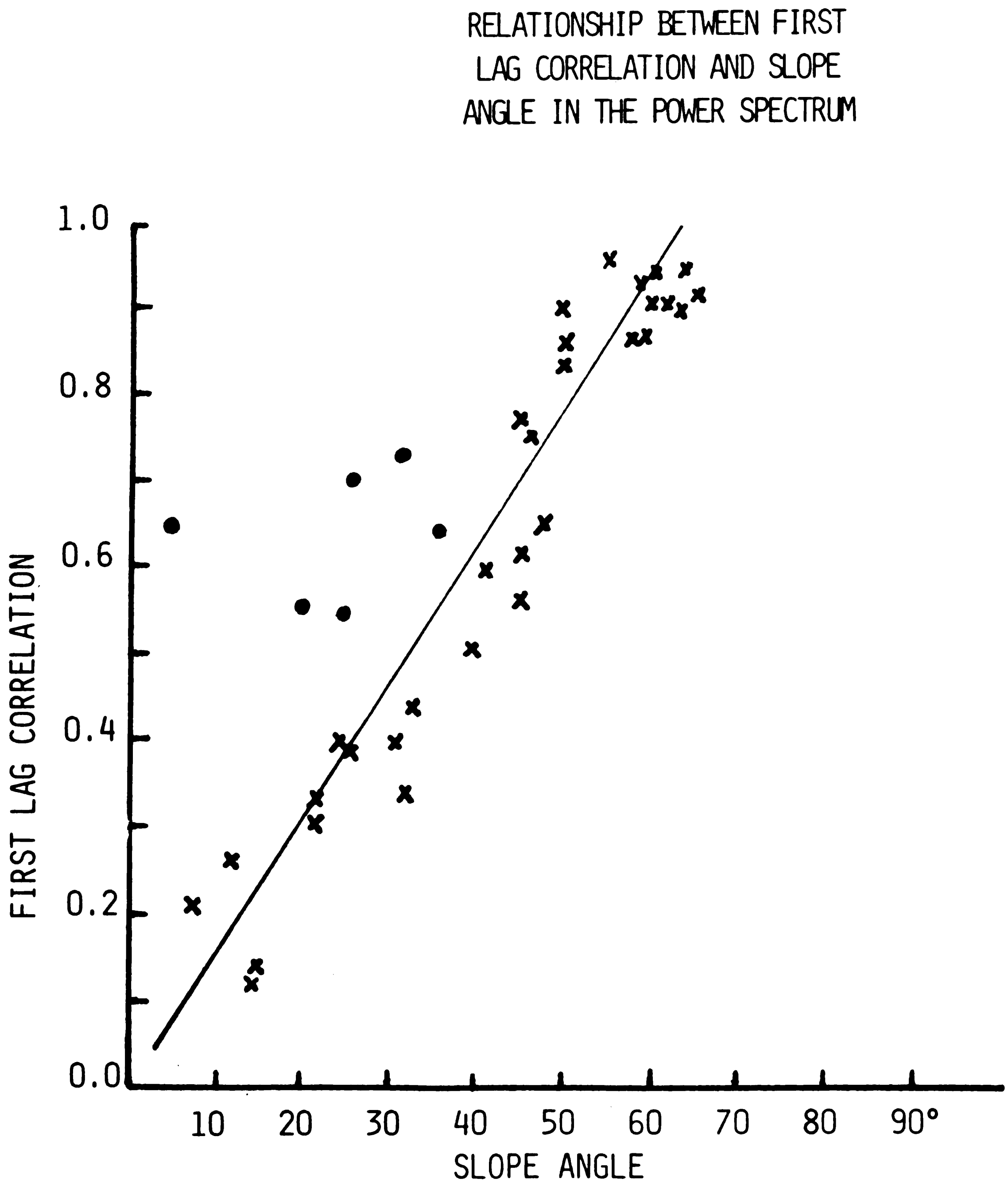
For time series records at a given point $\overline{R}_s$ is the Eulerian integral scale (e.g., Tennekes and Lumley, 1972, p. 275, etc.). It can be related to the Eulerian length scale of turbulence:

$$L_E = \overline{U}\,\overline{R}_s \qquad (12)$$

(e.g., Ivanov and Klinov, 1962, Tennekes and Lumley, 1972, Weber et al., 1982, etc.).

In the exponential model $\overline{R}_s$ is a unique function of ρ_c, the first lag correlation, and ρ_c is related to the slope. If turbulence were to follow the exponential red noise model, it implies that turbulence in the inertial subrange requires a special value of $\overline{R}_s$ and the length scale of turbulence L_E would change only with a change of $\overline{U}$. If $\overline{U}$ is constant, however, the slope

FIG. 5


RELATIONSHIP BETWEEN FIRST
LAG CORRELATION AND SLOPE
ANGLE IN THE POWER SPECTRUM
1.0
0.8
0.6
0.4
0.2
0.0
FIRST LAG CORRELATION
10
20
30
40
50
60
70
80
90°
SLOPE ANGLE

of the power spectrum is a function of $\overline{R}_s$ alone, which implies that it is a function of ρ_c, and may differ from -5/3. Thus the power spectrum may show other turbulence than the inertial subrange, a variation of the slope according to the type of stability conditions (e.g., Tennekes and Lumley, 1972, Weber et al., 1982, etc.), or no turbulence at all. In the latter case, the fluctuations of the wind velocity follow a random pattern with persistence (red noise).

The associations illustrated in Figures 3, 4, and 5 leave another interpretation open. Let us return to equation (5) and write:

$$\rho_M = \rho_c^M = \varepsilon \tag{13}$$

Using M as a scaling factor, the exponential autocorrelogram can be standardized, e.g., assume that $\varepsilon = 0.001$. In a first case of $\rho_c = 0.9$, we find $M_1 = 65.5$. Under an assumption of $\rho_c = 0.5$, we derive $M_2 = 10.0$. Consequently, $M_1 = 6.5\ M_2$. Let us assume that we have sampled our records every six seconds and have found a first lag correlation of $\rho_c = 0.5$. The autocorrelogram for (roughly) one-second sampling would increase ρ_c to 0.9 and provide a slope of -5/3. Since we have not altered the data set, $\overline{U}$ would not have changed. In this case, the Eulerian length scale of turbulence L_E would have changed because $\overline{R}_s$ for $\rho_c = 0.5$ is smaller than $\overline{R}_s$ for 0.9. It must be taken into account, however, that Δt is different in both cases (equation 11). Otherwise, the slope angle changes as an effect of the sampling rate.

The last example illustrates that it is not sufficient to study only the slope of the power spectrum in turbulence analysis. Other turbulence characteristics must be added in order to come to valid conclusions.

IV. INSTANTANEOUS AND STANDARD INSTRUMENTATION. Redstone wind data taken at one-second sampling intervals by the u-v-w Gill anemometer and at Otis AFB can be considered as instantaneous observations, although at Otis AFB sampling occurred every sixth second. There is a trend toward development of even more sensitive instrumentation in the future, using electro-optical wind measuring devices.

The difficulties in the interpretation of the slope angle which were delineated in the previous section are not limited to instantaneous measurements. Standard instrumentation measurements resemble the application of an averaging process to instantaneous values. It may be appropriate to examine briefly the effects.

In the averaging process, waves of short length are truncated compared with the original spectrum. This is not an adverse effect by itself. Under the assumption of exponential red noise, the first lag correlation would determine the slope.

TABLE 1

EFFECT OF AVERAGING

DATA: 19 AUGUST 1974

(1 SEC SAMPLING, LEVEL 6 FT)

SEC	ρ_U	ρ_V	σ_U^2	σ_V^2
1	0.98	0.98	1.28	0.50
4	0.95	0.92	1.25	0.48
8	0.90	0.85	1.21	0.46
16	0.81	0.74	1.15	0.43
24	0.76	0.65	1.07	0.40
32	0.68	0.61	1.04	0.38
48	0.61	0.55	0.92	0.33
64	0.57	0.51	0.86	0.31

MEAN $\overline{U}$ = - 0.60 $\overline{V}$ = 0.07 M/SEC

TABLE 2

EFFECT OF AVERAGING

(Tower P)

DATA: 25 NOVEMBER 1981

(6 SEC SAMPLING, LEVEL 10 FT)

AVERAGE	FIRST				SECOND			
	ρ_U	ρ_V	σ_U^2	σ_V^2	ρ_U	ρ_V	σ_U^2	σ_V^2
1	0.57	0.40	2.85	1.66	0.44	0.25	2.23	1.34
5	0.59	0.53	1.63	0.74	0.40	0.40	1.04	0.46
10	0.61	0.57	1.30	0.55	0.37	0.44	0.70	0.31
15	0.60	0.59	1.16	0.49	0.39	0.35	0.54	0.27
20	0.60	0.65	1.05	0.43	0.28	0.39	0.50	0.22
30	0.60	0.69	0.90	0.38				

MEAN $\overline{U}$ = 6.08 $\overline{V}$ = 0.0 M/SEC $\overline{U}$ = 6.41 $\overline{V}$ = 0.0 M/SEC

It is not trivial to predict the behavior of the first lag correlation as a function of the averaging interval. The first lag correlation for averages depends on the structure in the covariance matrix of the data series and the decrease of the variance by smoothing. Table 1 provides an example of a declining first lag correlation with increase of the averaging interval, but the decline is not always found. Table 2 exhibits a complex pattern. In three columns of Table 2, the first lag correlation increases with increasing averaging time, while one column shows a decrease. Since decrease and increase do not simply relate to the magnitude of the first lag correlation of the original data series, no simple prediction model is applicable.

Tables 1 and 2 confirm that the variance decreases with increasing averaging interval in accordance with the central limit theorem.

The slope in the power spectrum for the averaged data series changes in accord with the magnitude of ρ_c (see also Figure 5). In the exponential red noise model, $\overline{R}_s$ depends on ρ_c, and the previously discussed problems apply. It is possible, however, that averaged values represent fluctuations of the wind following a random process with persistence. In this case, the slope angle need not be -5/3. More investigations are necessary, however, before final conclusions can be made.

V. CONCLUSIONS. It is customary to consider the fluctuation of the wind in the boundary layer during short time intervals (e.g., seconds) to be in agreement with turbulence in the inertial subrange. While this association was confirmed for tower data at Redstone Arsenal, other data from Otis AFB delineated a different slope in the power spectrum (in double logarithmic coordinates). The detailed investigation revealed that these small-scale fluctuations are more likely to be produced by a random process with persistence (exponential red noise). This concept would explain that the angle in the power spectrum is not -5/3, but is a function of the first lag correlation (Figure 5).

A second interpretation is possible. The relationship between slope angle and first lag correlation is based on an exponential red noise model. The structure of turbulence in the inertial subrange is an exponential red noise model with a first lag correlation around 0.85 for fluctuations of short duration.

It was further shown that averaging of the instantaneous wind measurements may lead to a change in slope angle, although no simple relationship with the exhibited behavior of the first lag correlation could be derived. The averaging process would tie together instantaneous measurements and observations on less sensitive instruments.

ACKNOWLEDGMENT

The author expresses his appreciation to Dr. Dorathy A. Stewart and Mr. Larry Levitt for critical comments and the editorial review of the manuscript. Mrs. Alexa M. Mims deserves most of the credit for the preparation of the computer programs. Finally, Mrs. Mona White deserves credit for her patience and expeditious typing of the manuscript. Appreciation goes, too, to Major James C. Weyman, Air Force Geoph. Lab., for the data set from Otis AFB

REFERENCES

Blackman, R. B. and J. W. Tukey, 1958. The Measurement of Power Spectra. Dover Publ. Inc., New York, pp 190.

Box, E. P. and G. M. Jenkins, 1970. Time Series Analysis, Forecasting and Control. Holden Day, San Francisco, pp 553.

Cooley, J. W. and J. W. Tukey, 1965. An Algorithm for the Machine Calculation of Complex Fourier Series. Math. Comput., Vol. 19, p. 297-301.

Essenwanger, O. M., 1980. On Red Noise and Quasi-Periodicity in the Time Series of Atmospheric Temperature. Statistical Climatology, p. 165-181. Editors: S. Ikeda, E. Suzuki, E. Uchida and M. M. Yoshino. Elsevier, Amsterdam, pp 388.

Fichtl, G. H. and G. E. McVehil, 1970. Longitudinal and Lateral Spectra of Turbulence in the Atmospheric Boundary Layer at Kennedy Space Center. Journ. Appl. Meteorol., Vol 9, p. 51-63.

Gill, G. C., 1975. Development and Use of the Gill UVW Anemometer. Boundary-Layer Meteorol., Vol. 8, p. 475-495.

Hanna, S. R., 1979. Some Statistics of Lagrangian and Eulerian Wind Fluctuations. Journ. Appl. Meteorol., Vol. 18, p. 518-525.

Hinze, J. O., 1959. Turbulence, An Introduction to Its Mechanism and Theory, McGraw Hill, New York, pp 586.

Ivanov, V. N. and F. Ya. Klinov, 1962. Some Characteristics of a Turbulent Velocity Field in the Lower 300 Meter Layer of the Atmosphere. Bull. Acad. Science, USSR, p. 1026-1030.

Pasquill, F., 1962. Atmospheric Diffusion. Van Nostrand, New York, pp 297.

Priestley, C. H.B., 1959. Turbulent Transfer in the Lower Atmosphere. University of Chicago Press, Chicago, pp 130.

Stewart, D. A., 1981. A Survey of Atmospheric Turbulence Characteristics. Technical Report RR-81-6, US Army Missile Laboratory, Redstone Arsenal, AL 35898, pp 23.

Taubenheim, J., 1969. Statistische Auswertung Geophysikalischer und Meteorologischer Daten. Akadem. Verlagsgesellschaft, Leipzig, pp 386.

Tennekes, H. and J. L. Lumley, 1973. A First Course in Turbulence. MIT Press, Cambridge, Mass., pp 300.

Tukey, J. W., 1949. The Sampling Theory of Power Spectrum Estimates. NAVEXOS-P-735, Woods Hole. Office of Naval Research, p. 47-67.

Weber, H. H., J. S. Irwin and W. B. Petersen, J. J. Mathis, Jr., J. P. Kohler, 1982. Spectral Scales in the Atmospheric Boundary Layer. Journ. Appl. Meteorol., Vol. 21, p. 1622-1632.

A NEED FOR A METHODOLOGY FOR PRIORITIZATION OF MISSION AREA DEFICIENCIES

RICHARD T. MARUYAMA

METHODOLOGY BRANCH
STUDIES AND ANALYSIS DIRECTORATE
DEPUTY CHIEF OF STAFF FOR COMBAT DEVELOPMENTS
US ARMY TRAINING AND DOCTRINE COMMAND
FT MONROE, VIRGINIA

Abstract . The US Army Training and Doctrine Command (TRADOC) represents the battlefield user in developing doctrine, training, force structure, and materiel requirements for the future. To ensure these requirements stem from an overall battlefield concept and are based on sound analysis, TRADOC conducts Mission Area Analysis (MAA) in the twelve areas outlined in figure 1. Each area is assigned to a TRADOC center or school for analyses and the prioritization of resulting deficiencies.

Once all the analyses are completed and each proponent has prioritized the deficiencies within the mission area, TRADOC must integrate and prioritize the twelve deficiency lists into a single ordered list of battlefield deficiencies. This single list will guide the development of programs and the allocation of resources toward correcting deficiencies in order of their importance.

The difficulty is in developing a prioritization methodology which is sufficiently structured and rigorous to produce consistent results from year to year, while being sufficiently simple and well defined to be understood and accepted by the decision makers who will use it.

1. Introduction. The Army's Training and Doctrine Command (TRADOC) and Materiel Development and Readiness Command (DARCOM) are forging a linkage which will help compress the materiel development cycle while enabling the Army to project requirements further into the future. The Army's Long Range Research Development and Acquisition (RDA) planning system and TRADOC's Mission Area Analysis (MAA) process combine to provide a roadmap of how to get to the Army of the future. They provide a means to consider the future implications of current decisions and a way to couple these actions with the Planning, Programing, Budgeting, and Execution System for resource allocation.

The current Long Range RDA process, while still in its infancy, facilitates timely and systematic modernization. It recognizes that modernization must be coordinated throughout a total system that includes materiel, training, personnel, logistics, doctrine, tactics and related system requirements. It understands that these components are interrelated; solutions in one area could well cause deficiencies in another. Only a comprehensive approach to the total system will produce equipment that meshes with the force structure, training, and doctrine. To implement the process, however, DARCOM must understand the needs of the future battlefield, and that is where TRADOC comes into the picture.

Mission Area Analysis (MAA) allows the synthesizing of information gained through many individual studies and analyses into a single, internally consistent framework. To facilitate the detailed analyses of the Army's ability to execute its wartime missions, the overall battlefield concept is divided into 12 mission areas. These mission areas serve as the basis for measuring the capabilities of the force programed in the current Program Objectives Memorandum (POM) to fight a successful battle against a projected threat. Each mission area was assigned to a TRADOC center/school for analyses and the prioritization of resulting deficiencies.

Figure 1 shows the TRADOC mission area structure and proponent for each area.

MISSION AREA	PROPONENT
CLOSE COMBAT (HEAVY)	US ARMY ARMOR CENTER, FT KNOX, KY
CLOSE COMBAT (LIGHT)	US ARMY INFANTRY CENTER, FT BENNING, GA
AVIATION	US ARMY AVIATION CENTER, FT RUCKER, AL
AIR DEFENSE	US ARMY AIR DEFENSE CENTER, FT BLISS, TX
COMBAT SUPPORT ENGINEERING, & MINE WARFARE	US ARMY ENGINEER CENTER, FT BELVOIR, VA
COMBAT SERVICE SUPPORT	US ARMY LOGISTICS CENTER, FT LEE, VA
FIRE SUPPORT	US ARMY FIELD ARTILLERY CENTER, FT SILL, OK
BATTLEFIELD THEATER NUCLEAR WARFARE	US ARMY COMBINED ARMS CENTER, FT LEAVENWORTH, KS
NUCLEAR, BIOLOGICAL, CHEMICAL	US ARMY CHEMICAL SCHOOL, FT MCCLELLAN, AL
COMMAND & CONTROL	US ARMY COMBINED ARMS CENTER, FT LEAVENWORTH, KS
COMMUNICATIONS	US ARMY SIGNAL CENTER, FT GORDON, GA
INTELLIGENCE & ELECTRONIC WARFARE	US ARMY INTELLIGENCE CENTER, FT HUACHUCA, AZ

Once the MAAs are complete, work begins to integrate the lists of deficiencies from each mission area and prioritize them into a single ordered list of battlefield deficiencies. This single list will guide the development of programs and the allocation of resources toward correcting deficiencies in order of their importance.

II. Methodology.

The TRADOC prioritization process consists of four phases (see Table 1).

TABLE 1

4 PHASE APPROACH

I. ESTABLISH LIST OF MAJOR DEFICIENCIES BY MISSION AREAS

II. PRIORITIZE DEFICIENCIES WITHIN MISSION AREAS

III. INTEGRATE 12 MISSION AREA DEFICIENCY LISTS

IV. AGGREGATE RESULTS INTO ONE PRIORITIZED LIST

a. Phase I - Establish lists of major deficiencies by mission area (see figure 1).

A strawman list of major deficiencies was developed for each of the 12 TRADOC mission areas by a HQ TRADOC panel comprised of DCSCD, DCSDOC and DCST representatives. Each strawman list was then forwarded to the appropriate mission area proponent for review and input. The mission area proponents submitted recommended corrective actions (three to five) for each identified major deficiency. The number of corrective actions was arbitrarily fixed, at 3 to 5 with the intent being only to capture the thrust of significant corrective actions. Mission area proponents then provided the integrating centers and HQ TRADOC with a copy of the revised deficiency list with recommended corrective actions for final review. The headquarters review included a screen for consistency in describing deficiencies. The result of Phase I is 12 separate lists of major deficiencies. Figure 2 shows the development of the 12 separate deficiency lists for each mission area.

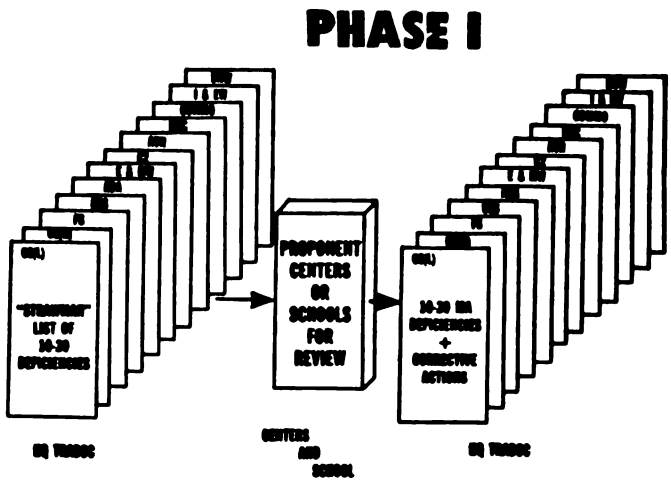


FIGURE 2

b. Phase II - Prioritize deficiencies within each mission area.

Each mission area proponent then prioritized their list of major deficiencies obtained in Phase I. This assessment was conducted considering the Army's programed forces using systems scheduled for fielding or fielded by 1987.

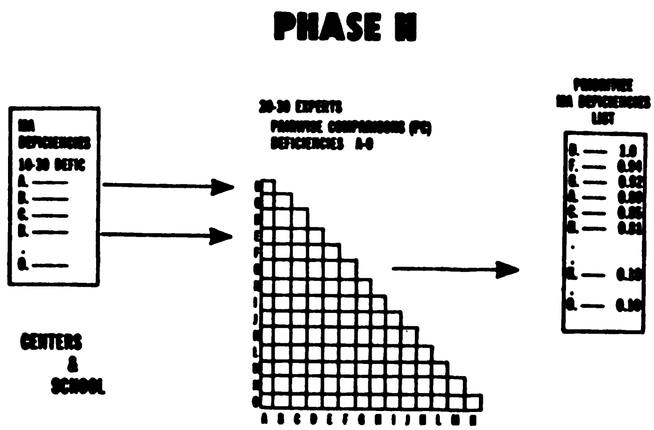
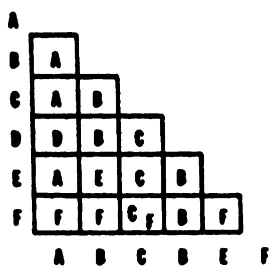


FIGURE 3

The technique of pairwise comparisons was used to prioritize the list of major deficiencies. It involves asking mission area experts (approximately 30) to independently consider the list of mission area deficiencies, compare the deficiencies two at a time, and sequentially determine their relative importance. Figure 3 shows the pairwise comparison process used to prioritize each mission area deficiency list.

A sample of the survey form for the pairwise comparisons is shown in figure 4. The ranking of a particular deficiency was determined by the number of times (frequency count) it was judged to be most important in accordance with the above criteria. Individual judgments were treated equally. The deficiency frequency counts from each mission area expert was aggregated and then normalized between 0.0 and 1.0. This process produced a cardinally ranked list of deficiencies (i.e., the order as well as the interval between each deficiency was established). The list of cardinally ranked deficiencies along with the completed survey forms was then returned to HQ TRADOC.

PAIRWISE COMPARISON EXAMPLE



A	3	0.75	3
B	1	0.25	tie 5
C	2.5	0.625	4
D	4	1.000	1
E	1	0.25	tie 5
F	3.5	0.875	2

FIGURE 4

c. Phase III - Integration of deficiency lists across mission areas. (see Table II)

The prioritized lists of major deficiencies obtained in Phase II was then integrated across mission areas. TRADOC conducted four separate sessions with general officers (GO) from proponent schools and centers. The composition of each panel reflected a broad coverage of expertise. Each GO panel then integrated the 12 mission area lists, two at a time, using the pairwise comparisons technique. Each GO panel member was asked to consider the top ranked deficiency in each mission area list and compare them two at a time to determine their relative importance as was done in Phase II. The lowest ranked deficiency in each list was similarly compared. Based on a frequency count of these deficiencies, the order of integration for the mission area lists was established (see figure 5). The list judged to have the single most important deficiency became the base list and the list with the second most important deficiency was integrated into the base list. The top ranked deficiency in the second list will be fixed on the base list in comparison with its importance to the other deficiencies on the base list. Its position on the base list was determined by the consensus of the GO panel. The lowest ranked deficiency in the second list was fixed on the base list in the same manner. The interval on the base list between the two fixed deficiencies forms the region of integration. The remaining deficiencies on the second list were then mathematically transformed into the base list. The GO panel screened the resultant list for any incongruencies or major discrepancies and made appropriate adjustments. Consensus among the panel members was required for adjustments to be made. This list became the new base list. Using the above procedure which is graphically portrayed in figure 6, the next ordered list was merged into the new base list until all 12 lists were integrated. Four prioritized lists of major deficiencies across mission areas (one from each panel) emerged from this phase.

TABLE II

PHASE III — INTEGRATION OF DEFICIENCIES ACROSS ALL MISSION AREAS

- PUBLISH PRIORITIZED DEFICIENCIES — READ AHEAD BOOK
- ESTABLISH FOUR GO PANELS — BROAD COVERAGE OF EXPERTISE
- PROCESS CONTROLLED BY A DECISION ANALYSIS TEAM
- INTEGRATE MA DEFICIENCIES LISTS — TWO LISTS AT A TIME
- RESULT: ONE PRIORITIZED LIST OF DEFICIENCIES ACROSS ALL MA FROM EACH PANEL

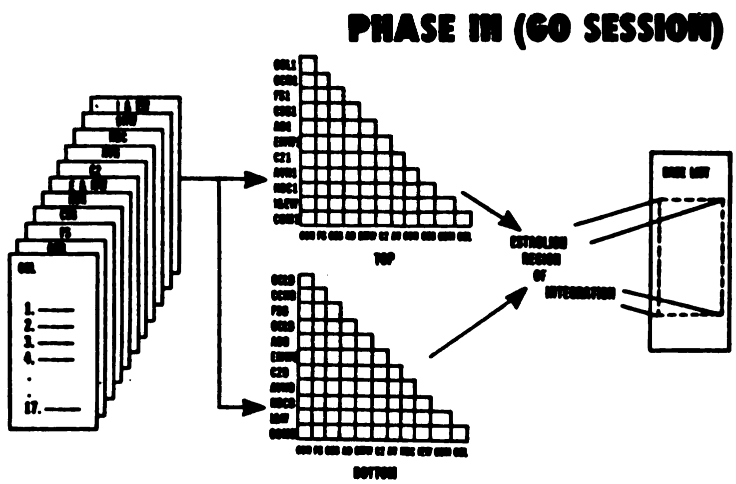


FIGURE 5

PHASE III

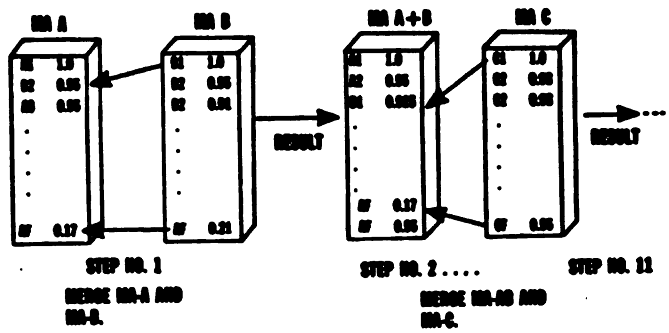


FIGURE 6

d. Phase IV - Aggregation of the four GO panel results.

A fifth GO panel was convened at HQ TRADOC to aggregate the results of the other GO sessions. This panel consisted of general officers from HQ TRADOC, Combined Arms Center (CAC), Ft Leavenworth, KS, Logistics Center (LOGC), Ft Lee, VA, Soldier Support Center (SSC), Ft Ben Harrison, IN, and the US Army Forces Command (FORSCOM). The four integrated lists of major deficiencies were reviewed and mathematically merged by aggregating the pairwise comparison results from the four GO sessions. The aggregated top and bottom deficiencies established the order of merge and the region of integration during this process. This list was screened for discrepancies and final adjustments were made based on the consensus of the panel. Phase IV was the final step in the overall prioritization effort, the result being a single prioritized list of major deficiencies across all TRADOC mission areas (see figure 6).

PHASE IV

AGGREGATED HQ TRADOC LIST

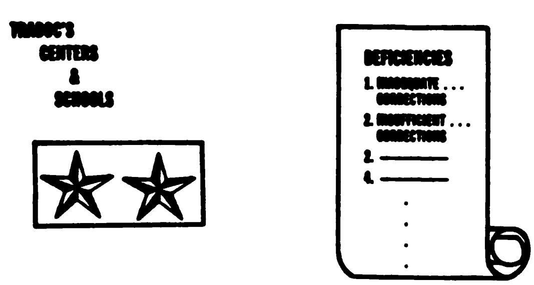


FIGURE 7

III. Summary

The TRADOC integration and prioritization process occurs annually. The current methodology is new and, hence, asks for improvements in overcoming several shortfalls. The technique does not allow for weighting either individual deficiencies or mission area lists to give consideration to the fact that not all mission areas are equally deficient nor equally significant. Additionally, an artificial ceiling of 20 deficiencies per mission area had to be established to limit the numbers of required comparisons to an acceptable level. In reality, many MAAs produced deficiencies numbering in the hundreds.

Contributors desiring additional information or wishing to comment on proposed improvements to the integration and prioritization process are encouraged to contact the author by phoning (804) 727-3004 or by writing HQ TRADOC, ATTN: ATCD-AM, Ft Monroe. VA 23669.

UNBIASED RANDOM INTEGRATION METHODS WITH EXACTNESS FOR LOW ORDER POLYNOMIALS

Andrew F. Siegel and Fanny O'Brien
Department of Statistics
Princeton University

ABSTRACT:

When a definite integral cannot be evaluated exactly, we turn to computer-based methods for approximations. There are many different kinds of such procedures, but they divide roughly into two classes: deterministic and random. We explore the use of methods that combine these two approaches, preserving the unbiasedness and error estimation advantages of random methods, but at the same time maintaining the closer approximation generally found in deterministic methods.

INTRODUCTION:

Consider estimation of the integral

$$I_f = \int_{-1}^{1} f(x)\,dx.$$

A classical deterministic approach might be to try Simpson's rule, which uses the approximation

$$\frac{1}{3}\,[f(-1) + 4f(0) + f(1)].$$

This is the integral of the interpolating quadratic approximation to f that takes on the same values at -1, 0, and 1. The advantage is exactness for quadratics (and cubics too, by symmetry.) The disadvantage of this approach is that the error assessment involves a higher order derivative of the function f, which may be difficult to find (especially for a function which was difficult to integrate analytically in the first place.)

A second approach might be the standard random method known as simple Monte Carlo. To match this with the previous approach, let us also do three function evaluations, and estimate the integral as

$$\frac{2}{3}\ [f(X_1) + f(X_2) + f(X_3)]$$

where the function is evaluated at three values that are independently and uniformly distributed random variables in (-1, 1). The advantage of this approach is unbiasedness: no matter what f is, so long as it is integrable, the average (in the sense of expectation) of this approximation will be the true integral of f. The practical advantage of this is that the approximation can be repeated several times, generating new independent random variables each time, and from these results the average and standard error can be calculated. This gives an error assesment without need for further mathematical analysis (which might not be tractable.) The disadvantage of this approach is that it is generally not as exact as Simpson's rule and the higher order deterministic methods.

Methods that combine the advantages (but not the disadvantages) of each method have been available since the work of Ermakov and Zolotukhin (1960). This and other random integration procedures are discussed by Hammersley and Handscomb (1964). Related work has been done by Bogues, Corbett, and Patterson (1981), Cranley and Patterson (1970, 1976), Haber (1969, 1970), and Quackenbush (1969). We have proposed (Siegel and Zambuto, 1982) the use of symmetric quadrature designs of $2k+1$ points which achieve unbiasedness together with exactness for polynomials of degree $2k+1$ for this problem.

THE THREE-POINT INTEGRATION RULE:

The symmetric random unbiased 3-point integration rule is unique, and is given by

$$I_f(\xi) = \frac{1}{3\xi^2}\ [f(-\xi) + 2(3\xi^2-1)f(0) + f(\xi)]$$

where ξ has the distribution of the cube root of a uniformly distributed random variable in (0, 1).

This may be thought of as a systematically random (as opposed to a completely random) adaptation of Simpson's rule. The function is evaluated at three points: $-\xi$, 0, and ξ, in a symmetric but partially random design. The weights are chosen properly in order to assure that the approximation is equal to the integral of the quadratic function that agrees with f at these three points; this assures exactness for quadratics (as does Simpson's rule.) Finally, there remains a degree of freedom: ξ must be specified. As it turns out, there is only one distribution for ξ that will preserve the unbiasedness property of simple Monte Carlo: the cube root of a uniform. It is like applying Simpson's rule with a special random scaling of the design points at which the function is evaluated. Simpson's rule represents only one extreme ($\xi = 1$) of this continuum of possibilities which must be sampled carefully in order to obtain an unbiased estimate.

EXAMPLE:

Consider the integral

$$I_f = \int_{-1}^{1} \cos\left(\frac{\pi x}{2}\right) dx = \frac{4}{\pi} \sim 1.27324\ldots$$

The integral can be evaluated exactly in this case, but we will use numerical methods in order to gain insight into the way the procedures work. This function goes from zero to one and back to zero; hence the simple Monte Carlo estimates can range from 0 to 2. The symmetric random design for this problem yields the following integral estimate:

$$I_f(\xi) = 2 - \frac{2}{3\xi^2}\left[1 - \cos\left(\frac{\pi\xi}{2}\right)\right]$$

This is a much flatter function! It takes on values from 1.1775... to 1.3333... When these values are sampled randomly it is clear that they will fall much closer to the true value (1.27324...) than simple Monte Carlo would. Using the cube root of a uniform for ξ guarantees that the results will be correct on the average, and that any standard error or confidence interval will be asymptotically correct. Simpson's rule, representing $\xi = 1$, always yields the estimate 1.333..., one of the extremes that the random rule can achieve (although in general, Simpson's estimate need not be an extreme.)

TWO DIMENSIONS:

There is a corresponding simple formula for the two-dimensional integral

$$\int_{-1}^{1}\int_{-1}^{1} f(\xi,\eta)\, d\xi\, d\eta$$

given by

$$J_f(\xi,\eta) = \frac{f(\xi,\eta)+f(-\eta,\xi)+f(\eta,-\xi)+f(-\xi,-\eta) - 4f(0,0)}{3(\xi^2 + \eta^2)/2} + 4f(0,0)$$

This formula provides an exact answer if f is a polynomial in ξ and η of degree at most 3. It will give an unbiased estimate if (ξ,η) is sampled from the density

$$\frac{3}{2}(\xi^2 + \eta^2) \qquad \xi \text{ and } \eta \text{ in } (0,1).$$

ACKNOWLEDGMENTS:

This work was supported in part by the U.S. Army Research Office Contracts DAAG-29-79-C-0205 and DAAG29-82-K-0178, and by the U.S. Department of Energy Contract DE-AC02-81ER10841.

REFERENCES:

Bogues, K., Corbett, R.M., and Patterson, T.N.L. (1981). An Implementation of the Method of Ermakov and Zolotukhin for Multidimensional Integration and Interpolation. Numerische Mathematik 37, 49-60.

Cranley, R., and Patterson, T.N.L. (1970). A Regression Method for the Monte Carlo Evaluation of Multi-Dimensional Integrals. Numerische Mathematik 16, 58-72.

Cranley, R., and Patterson, T.N.L. (1976). Randomization of Number Theoretic Methods for Multiple Integration. SIAM Journal on Numerical Analysis 13, 904-914.

Ermakov, S.M., and Zolotukhin, V.G. (1960). Polynomial Approximations and the Monte Carlo Method. Teor. Veroyatnost. i Primenen 5, 473-476, translated as Theory of Probability and its Applications 5, 428-431.

Haber, S. (1969). Stochastic Quadrature Formulas. Mathematics of Computation 23, 751-764.

Haber, S. (1970). Numerical Evaluation of Multiple Integrals. SIAM Review 12, 481-526.

Hammersley, J.M., and Handscomb, D.C. (1964). Monte Carlo Methods. London: Methuen.

Quackenbush, R.W. (1969). Monte Carlo Quadrature with Exactness for Polynomials. Ph.D. Thesis, Stevens Institute of Technology.

Siegel, A.F., and Zambuto, F. (1982). Quadrature Designs for Unbiased Estimation of the Integral of Any Function that are Exact for Low Order Polynomials. Technical Report No. 226, Series 2, Department of Statistics, Princeton University.

ESTIMATING THE VARIANCE OF THE LOSS EXCHANGE RATIO

Eugene Dutoit
OR/SA Branch
Directorate of Combat Developments
US Army Infantry School
Fort Benning, Georgia 31905

ABSTRACT. A measure of force effectiveness that is often used in military analysis of combat is the Loss Exchange Ratio (LER). In many cases, the input required to calculate this measaure is obtained by replicating a stochastic wargame model by using a computer or a manual exercise. It would be useful to determine a confidence interval about this measure of force effectiveness. This confidence interval would enable the analyst to examine problems concerning the precision of the measure, and compute the replication requirement for a stated degree of precision. Hypothesis testing could be done to compare the LERs of different alternative weapon systems introduced into the force. Two methods of solution are developed and proposed and an example is given.

I. Acknowledgements. I want to thank the following people for submitting information and comments that were relevant to this problem. It is this kind of after-conference communication that results in the pay-offs for presenting technical and clinical papers at these conferences.

(a) Mark Adams (USAMERADCOM, Ft Belvoir, VA) and Gordon Holterman (Systems Cost and Automation Center; Ft Lee, VA) for both referencing George Fishman's text Principles of Discrete Event Simulation (pages 55-61).

(b) Larry Crow (USAAMSAA; Aberdeen Proving Ground, Maryland) for referencing C. R. Rao's text Linear Statistical Inference and Its Applications, pages 319-321.

(c) John Farmer (ATCT-MA; Ft Hood, Texas) for referencing Finney's text Statistical Method in Biological Assay, pages 27-29 and also for including some of his personal notes.

These references are cited in this paper.

II. Introduction. A measure of force effectiveness that is often used in military analysis is the Loss Exchange Ratio. This measure is defined as the ratio of red casualties (R) to blue casualties (B):

$$LER = R/B \qquad (1)$$

The LER shows an operational advantage to the blue force if $R > B$. For this discussion the values of R and B are obtained by replicating a stochastic wargame model. For each replication of the wargame the paried values of R and B are recorded. The average LER, (LER), is computed as:

$$\hat{LER} = \frac{\bar{R}}{\bar{B}} \qquad (2)$$

Because the generators of these average values are the results of a stochastic wargame, it would be useful to determine a confidence interval about the measure. The confidence interval could be used to answer the following questions:

1. Is the LER $\geq$ 1?

2. What sample size is required to estimate the LER with some stated degree of precision?

3. Are various measures of LER statistically different from each other at some selected level of significance?

In order to determine the confidence interval for the LER it is necessary to compute the variance of the estimate.

III. Error Propagation. It is well known (reference Beers) that if a function of the form

$$y = f(x_1, x_2, x_3, \ldots, x_k) \qquad (3)$$

exists, then the variance of this function can be written as:

$$var(y) = \sum_{i=1}^{k} \left(\frac{\partial f}{\partial x_i}\right)^2 var(x_i) + 2\sum_{\substack{i,j \\ i \neq j}}^{k} \left(\frac{\partial f}{\partial x_i}\right)\left(\frac{\partial f}{\partial x_j}\right) R_{ij} (var(x_i))^{1/2} (var(x_j))^{1/2} \qquad (4)$$

where

$var(x_i)$ = variance of the i^{th} variable

R_{ij} = correlation between the i^{th} and j^{th} variable.

The general form of the LER is shown as equation (2); therefore, the variance of this form can be written as:

$$VAR(\hat{LER}) = VAR\left(\frac{\bar{R}}{\bar{B}}\right) \qquad (5)$$

Applying equation (4) to compute the variance of the (LER) we obtain the following:

$$Var(\hat{LER}) = (1/\bar{B})^2 var(\bar{R}) + (-\bar{R}/\bar{B}^2)^2 var(\bar{B}) + 2(1/\bar{B})(-\bar{R}/\bar{B}^2) R_{\bar{R}\bar{B}} \sqrt{var(\bar{R})}\sqrt{var(\bar{B})} \qquad (6)$$

The following is also true:

$$\left.\begin{aligned} &VAR(R) = S^2_R/n; \text{ where } n = \text{the number of replications} \\ &VAR(B) = S^2_B/n \\ &R_{\bar{R},\bar{B}} = R_{R,B}; \text{ as pointed out in Beers (Reference 1, page 31)} \end{aligned}\right\} \quad (7)$$

Substituting the set of relationships (7) into equation (6) gives an expression for VAR(LER).

$$VAR(\widehat{LER}) = \frac{1}{n}[(1/\bar{B})^2 S_R^2 + (-\bar{R}/\bar{B}^2)^2 S_B^2 + 2(1/\bar{B})(-\bar{R}/\bar{B}^2) R S_R S_B] \quad (8)$$

This expression is equivalent to the one given by Fishman (Reference 3, page 59) for "large" value of n. The appropriate 100(1 - α) confidence interval (C.I.) for the LER could be computed as:

$$100(1 - \alpha)\ C.I.\ (LER) = \widehat{LER} \pm t\sqrt{VAR(\widehat{LER})} \quad (9)$$

Where t is Student's t with (n - 1) degrees of freedom. Equation (9) is supported, in part, by Rao (Reference 5, pages 319 through 321) who points out that in practical applications, distributions of the form we are studying are <u>asymptotically</u> normal if the co-variance and partial derivatives are continuous.

IV. <u>Fieller's Theorem</u>. Goldstein's text <u>Biostatistics</u> (reference 4, page 184) gives a situation where it is necessary to compute a confidence interval for a ratio. The problem from the Goldstein text is given below. Note the correspondence between the biological experiment and the force-on-force simulation.

"Quite often in biological experimentation one wishes to estimate a ratio from a set of observations on the numerator (y) and another set of observations on the denominator (x). Now these may be paired observations, each item in a sample supplying a value of y and a value of x, so that there may be some degree of correlation between the two...Suppose the protein content of cells per unit of DNA is to be determined. If the cells in question are growing in replicate bottles, we may determine both DNA and protein on the contents of each bottle...in (this) case, replicate estimates of the desired ratio will be available..."

Using this example, the replicate bottles become the replicated force-on-force simulations and the correlated values of both DNA and protein within each replicate bottle become the number of red and blue casualties respectively. Goldstein then points out, "that the appropriate limits of the true ratio R whose estimate is $\bar{y}/\bar{x}$, are given by Fieller's Theorem as roots of a quadratic equation:

$$- [\bar{x}^2 - t^2(S_x^2/n)]R^2 + 2[\bar{x}\bar{y} - t^2 r S_x S_y/n]R - [\bar{y}^2 - t^2(S_y^2/n)] = 0 \quad (10)$$

where r = sample correlation coefficient
n = number of paired observations in the sample

S^2x, S^2y = sample variances
t = two-tailed value of Student's t with $(n - 1)$ degrees of freedom"

Finney (reference 2, page 27 through 29) gives a discussion of Fieller's Theorem and its application to finding fiducial limits to a ratio of two means. Solving equation (10) for R and using the notation consistent with the LER and force-on-force simulation (i.e., $\bar{x} \equiv \bar{B}$ and $\bar{y} \equiv \bar{R}$) we obtain:

$$R_{U,L} = \frac{\overline{BR} - \frac{t^2 r S_B S_R}{n} \pm \sqrt{\left(\overline{BR} - \frac{t^2 r S_B S_R}{n}\right)^2 - [\bar{B}^2 - t^2(\frac{S_B^2}{n})][\bar{R}^2 - t^2(\frac{S_R^2}{n})]}}{\bar{B}^2 - t^2(\frac{S_B^2}{n})} \quad (11)$$

Fishman's (reference 3, pages 59 through 61) section on confidence intervals uses a quadistic form similar to equations (10) and (11) to derive a confidence interval for R which is cited as work done by Fieller and is suggested for use in simulation by Crane and Iglehart.

V. Example. The following numerical example is based on the data used by Goldstein for his explanation of replicate bottles of DNA and protein cited in section IV of this paper. The column headings have been changed to red and blue casualties, respectively.

TABLE 1.

Red and Blue Casualties for each Battle Replication
(adopted from Goldstein)

Replication	Number Red Casualties	Number Blue Casualties
1	12	5
2	14	7
3	12	3
4	12	3
5	13	8
6	13	6
7	13	4

The summary statistics are:

$N = 7$

$\bar{R} = 12.71$ casualties

$S_R = .756$ casualties

$\bar{B} = 5.14$ casualties

$S_B = 1.952$ casualties

$R = .710$

and the esimate of LER ($\hat{LER}$) = 2.47.

a. Error Propagation. Applying these data to equation (8) gives a value of the VAR ($\hat{LER}$) equal to .101. The appropriate value of t(for a 95% confidence interval with 6 degrees of freedom) is 2.447. The 95% confidence limits for the LER (using equation (9)) are 3.3 and 1.7. Note that the lower confidence limit (1.7) is greater than 0.

b. Fieller's Method. Applying these same summary data to equation (11) and solving for both roots of the quadistic gives values of 3.7 and 1.9, respectively. Note that the lower confidence limit (1.9) is also greater than 0.

VI. Conclusions. Although this effort represents a limited study, the following conclusions are emerging.

a. Error Progation and Fieller's method appear to give "reasonably" consistent results.

b. Fieller's method is the preferred way to compute a confidence interval about a ratio. This conclusion is based on some of the existing literature (Finney, Fishman, and Goldstein).

c. The error propagation method should increase in accuracy as "n" becomes large (Fishman and Rao); however, Fieller's Theorem should be more appropriate for the smaller numbers of replications that are used in force-on-force simulation.

References

1. Beers, Y; Introduction to the Theory of Error, Addison-Wesley, 1957.
2. Finney, D. J.; Statistical Method in Biological Assay, Charles Griffen and Co limited.
3. Fishman, G; Principals of Discrete Event Simulation, Wiley, 1978.
4. Goldstein, A; Biostatistics, An Introductory Text, Mac Millan Company, 1968.
5. Rao, C. R., Linear Statistical Inference and Its Applications, Wiley, 1965.

AN EXAMPLE OF SOFTWARE VALIDATION USING A FACTORIAL DESIGN

Joseph M. Tessmer
Department of Energy
Office of the Strategic Petroleum Reserve

ABSTRACT

This paper reports on the effects of the Department of Energy to test and evaluate, "validate", a large computer model, which represents the world petroleum distribution system. The evaluation technique employed is a complete 2^4 factorial design. The main effects, as well as all 2nd order effects are estimated. The technique provided criterial insights into the nature of the software identifying errors and assisting in the development, of a methodology, for reexamining candidate crude mixes, of oil stored by the government, for use during possible petroleum interruptions.

DISCLAIMER

The assumptions, procedures, analysis, conclusions, and recommendations contained in this paper are solely those of the author and do not represent any official policy of the Department of Energy or U.S. Government.

A1 Factorial Design Methodology

An experiment was performed to measure the effect of four sets of input factors on the oil market, as represented by a large linear programing model supplied by a private contractor. Two levels for each set of input factors were chosen and all 16 possible combinations of these input factors were used as model input to the model. This procedure, a 2^4 factorial design was chosen since it is economical, easy to use and provides a great deal of valuable information. Specifically a two (2) level factorial design has the following advantages:

1. If sets of input factors are varied one set at a time with the remaining factors held constant, it is necessary to assume that the effect would be the same at other settings of the other sets of input factors. Factorial designs avoid the assumption.

2. If the effects of input factors act additively, a factorial design estimates those effects with more precision. If the effects of the input factors do not act additively, factorial designs can detect and estimate the interactions which measures the non-additivity.

3. Factorial designs require relatively few runs per set of input factors studied and can indicate major trends and determine promising direction for further investigation. To obtain the same precision of the estimate of the effects measured, in this effort forty runs would have had to be run using the traditional, one factor at a time approach rather than the sixteen used in the experiment.

4. If a more thorough local exploration is needed, it can be suitably augmented to form composite designs.

5. These designs and their corresponding fractional designs may be used as building blocks so that the degree of complexity of the finally constructed design can match the sophistication of the problem.

To perform a 2^4 factorial design the two levels (or versions) for four (4) sets of input factors were selected and all sixteen (16) possible combinations were executed. The four sets of input factors and their levels (or versions) are listed on the following page.

Input Factor		Levels	
1,	Composition of the SPR	1a,	The SPR is filled with 100% light and sweet crude represented by Ekofisk (Sweet).
		1b,	The SPR is filled with 100% heavy and sour crude represented by Arab Heavy (Sour).
2,	Crude pattern or availability	2a,	Historical 1978 BAU case (BAU)
		2b,	As 2a above with a 50% closure of the Persian Gulf with a uniform SPR drawdown of 3MMBD (50% P.G.)*
3,	Refinery configuration	3a,	Worldwide 1978 refinery capacity (1978)
		3b,	Worldwide estimated 1985 refinery capacity (1985)
4,	Product price elasticities	4a,	The proposed set of elasticities compiled for this project by the contractor (CON)
		4b,	As 4a except a quite different set of elasticities developed by an alternate contractor for major products in the US (ALT)

* The selection of this crude oil disruption does not represent the policy of the Department of Energy and was used solely to evaluate the reaction of the model to changes in the world crude pattern.

These input factors combine to produce the following design or validation matrix:

Design Matrix

RUN NUMBER	SPR COMP	CRUDE PATTERN	REF CONFIG	ELAS
1	1a	2a	3a	4a
2	1a	2a	3a	4b
3	1a	2a	3b	4a
4	1a	2a	3b	4b
5	1a	2b	3a	4a
6	1a	2b	3a	4b
7	1a	2b	3b	4a
8	1a	2b	3b	4b
9	1b	2a	3a	4a
10	1b	2a	3a	4b
11	1b	2a	3b	4a
12	1b	2a	3b	4b
13	1b	2b	3a	4a
14	1b	2b	3a	4b
15	1b	2b	3b	4a
16	1b	2b	3b	4b

Table A-1

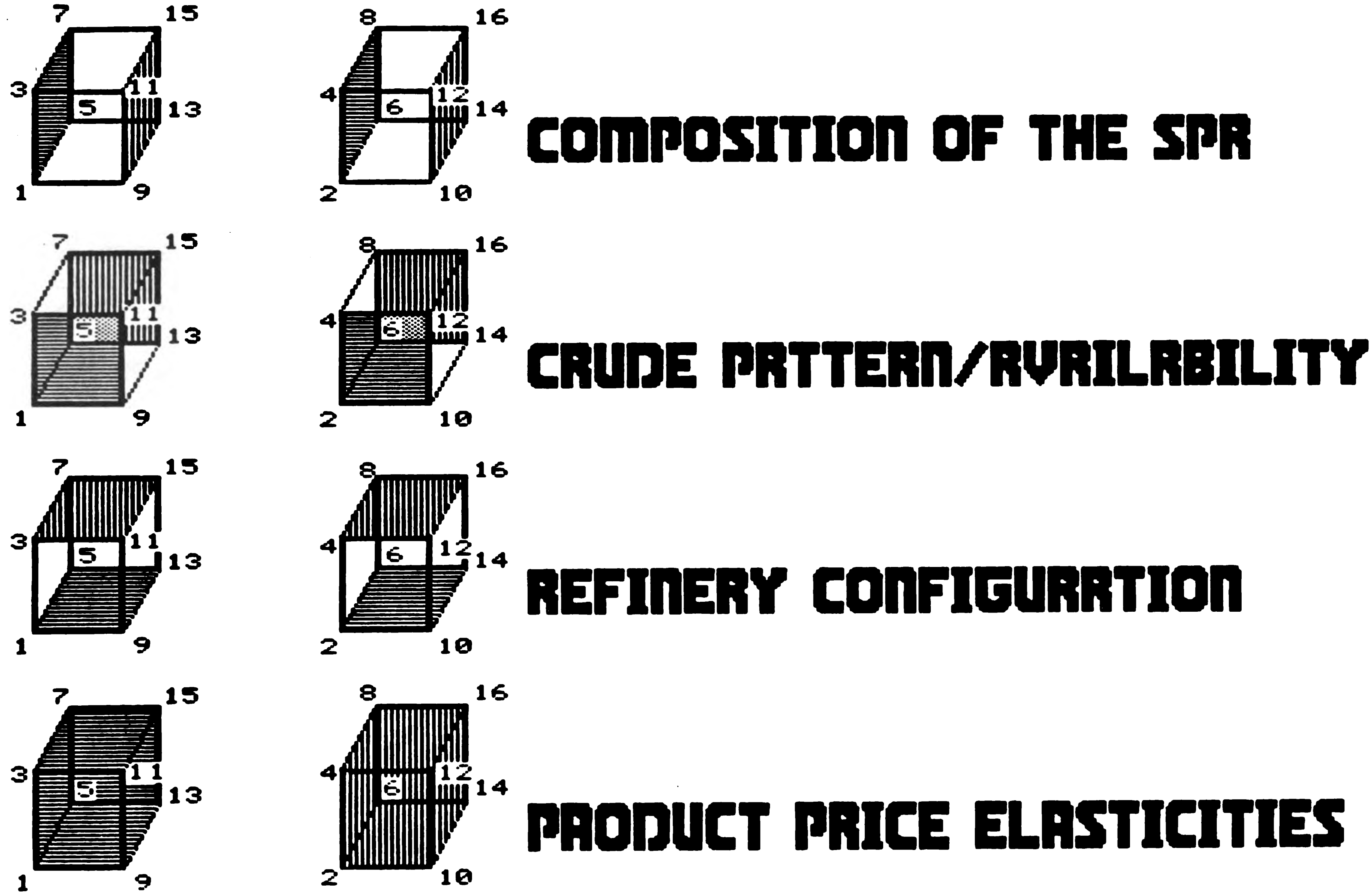
The interpertation of the runs in Table A-1 is easily illustrated by run number 6 which assumes that the SPR is filled with 100% light and sweet crude oil. There is a major oil interruption consisting of a 50% closure of the Persian Gulf and the SPR in being withdrawn at the rate of 3MMBD to reduce the effects of the crude oil shortfall. The refinery configuration during this period represents the 1978 time frame. Finally the set of elasticities for the demand of products, within the US developed by the alternate contractor (ALT) are assumed. The other fifteen runs are interperted similarily.

The sixteen runs of the design matrix, may be visualized geometrically as two cubes. One possible visualization appears in figure A-1 on the following page. The run number is at each vertex.

STRATEGIC PETROLEUM RESERVE

figure A-1

GEOMETRICAL REPRESENTATION OF THE 2^4 FACTORIAL DESIGN MAIN EFFECTS



A2 Calculation of Main Effects

The "main effect" of a set of input factors is the change in the response as we move from the "a" case to the "b" case version of that set of input factors. To examine the effect of the composition of the SPR a table of eight pair of column vectors was constructed (see table A-2). Aside from experimental error, the difference between the first column vector of the pair and the second column vector in the pair is due to a change, in the composition of the SPR. The average of these eight differences (one difference for each pair of column vectors) is the main effect due to the composition of the SPR. Table A-2 contrasts the composition of the reserve. If the columns are rearanged so that the run numbers are in a assending order, one obtains the table contrasting the product price elasticities. Similar rearangements yield tables contrasting the other two sets of input factor.

Geometrically speaking, using Figure A-1 the main effects are calculated from corresponding vertices from the two cubes as described below.

Input factor	
Composition of SPR	Left side of both cubes vs the right side of both cubes
Crude pattern/availability	The front of both cubes vs the backs of both cubes
Refinery configuration	The bottom of both cubes vs the tops of both cubes.
Product price elasticities	The left cube vs the right cube.

Composition of the SPR Sweet/Sour

Crude Pattern	1978 BAU		1978 BAU		1978 BAU		1978 BAU		50% P.G.		50% P.G.		50% P.G.		50% P.G.	
Ref Config.	78		78		85		85		78		78		85		85	
HLAS	CON		ALT		CON		ALT		CON		ALT		CON		ALT	
Run #	1	9	2	10	3	11	4	12	5	13	6	14	7	15	8	16
$ Arab Light/BBL	12.7	12.7	12.7	12.7	12.7	12.7	12.7	12.7	27.9	28	25.2	25.4	28.1	28.7	25.6	26.2
U.S. Consumer Bill (bill $)	97	97	97	97	95	95	95	95	171	171	146	146	171	172	145	146
U.S. Import Bill (bill $)	42.5	42.4	42.5	42.4	53.4	53.4	53.4	53.4	36.3	33.3	24.4	26.3	57.5	57.7	41.8	47.2
Revised U.S. Import Bill	41.3	41.3	41.3	41.3	41.1	41.1	41.1	41.1	28.0	29.6	17.5	19.8	27.2	29.0	16.2	18.2
Product MMBD																
Total	18.8	18.8	18.8	18.8	18.9	18.9	18.9	18.9	16.5	16.6	15.5	15.8	16.3	16.5	15.3	[illegible]
U.S. Gasoline	7.4	7.4	7.4	7.4	7.4	7.4	7.4	7.4	6.8	6.8	5.5	5.6	6.7	6.7	5.4	5.4
U.S. Dist	4.6	4.6	4.6	4.6	4.6	4.6	4.6	4.6	3.9	3.9	4.1	4.0	4.5	3.8	4.1	4.0
U.S. Resid	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	2.4	2.5	2.8	2.8	2.4	2.5	2.7	2.9
U.S. Crude Imports MMBD	5.4	5.4	5.4	5.4	5.5	5.5	5.5	5.5	1.6	1.6	0.9	1.1	1.5	1.5	0.8	1.1
U.S. Sweet Crude Imports	2.8	2.8	2.8	2.8	2.9	2.9	2.9	2.9	0.9	1.3	0.4	1.0	0.9	1.4	0.5	0.9

Table A-2

5.A3 2nd-Order Interaction effects

Suppose that one is interested in examining the effects of two sets of input factors; for example, refinery configuration and crude pattern. Then the sixteen runs of the factorial design can be grouped into four sets of four runs each. Each run in the group would have the same value for the input factors studied, although other input factors would vary within each group. Assume that for the business as usual crude pattern with the 1978 refinery configuration, the average value for the output variable being studied is 100. This will be the base point. Also assume that the main effects for the crude pattern and the refinery configuration are 25 and 10 respectively. This means that, on the average, changing from a BAU crude pattern to a 50% closure of the Persian Gulf will increase the output variable under study by 25. Likewise a change from the 1978 refinery configuration to the 1985 refinery configuration, will on the average, increase this same output variable by 10. If the input factors act additively, then the average value of the output variable with both the 1985 refinery configuration and a 50 percent closure of the Persian Gulf would be
100 + 25 + 10 = 135.

This artificial case is represented by the upper diagram in figure A-2. Note that the quantity

$$(b + c - a - d)/2 = (110 + 125 - 100 - 135)/2 = 0$$

i.e. there is no interaction.

Suppose that the input factors do not act additively, and the base point of 100 and main effects are the same. Then the resulting measurements could be described by the lower diagram in figure A-2. The input factors are now said to interact. By convention a measure of this interaction is

$$(b + c - a - d)/2 = (145 + 160 - 100 - 135)/2 = 35$$

This is the second order interaction and is called the refinery configuration X crude pattern interaction.

Like the main effect, the 2nd order interaction is the difference between two averages, eight of the sixteen results being included in one average and eight in the other. Analogous explanations are easily constructed for all other 2nd order interaction effects.

5.A4 Higher-Order interaction effects and the Standard Error.

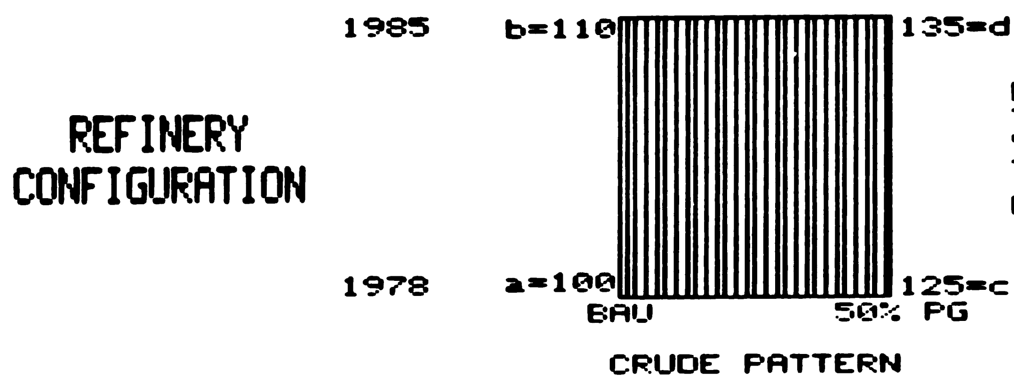
Similar procedures to those above can be given for deriving the third and fourth-order interactions. Due to the similarity of response functions it is reasonable to assume that higher-ordered interactions are negligible and measure differences arising principally from experimental error. Thus the mean, of the sum of squares, of these interactions give an estimated value for the variance of an effect, having five degrees of freedom. The square root of this value is an estimate of the standard error.

The level of statistical significance chosen for this study was $p=.01$. In order to select the statistically significant main effects and second order interactions multiply the standard error by $t_{1-p/2}=4.032$ any main effect or interaction greater than this product is considered statistically significant.

STRATEGIC PETROLEUM RESERVE

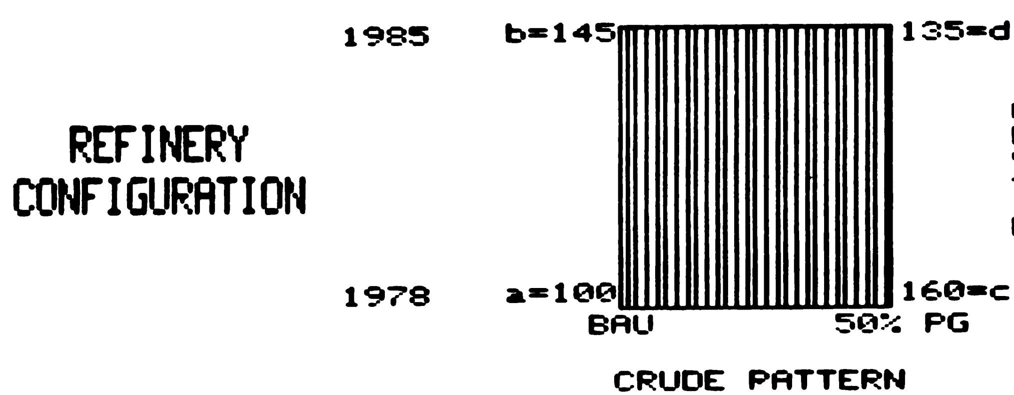
figure A-2

INTERPRETATION OF 2-WAY DIAGRAMS



NO 2nd ORDER INTERACTION. THE BEHAVIOR OF THE OUTPUT VARIABLE IS EXPLAINED BY THE MAIN EFFECTS.

NOTE: (110+125-100-135)/2 = 0



AN IMPORTANT 2nd ORDER INTERACTION EFFECT. THE BEHAVIOR OF THE OUTPUT VARIABLE IS NOT FULLY EXPLAINED BY THE MAIN EFFECTS.

NOTE: (145+160-100-135)/2 = 35

5.B1 Summary of significant effects

Table B-0 on the following page is a summary of all detected main effects and 2nd order interactions which were significant at the p=0.01 level. In addition sections B2, B3 and B4 provide detailed analysis of the import bill and total product consumption. The analysis of the remaining output variables is straight forward and available from the author upon request. Each column and subsection of the analysis reports on the univariate analysis of variance of the selected output variable. No attempt was made to perform a multivariate analysis of variance since model is completely deterministic and not stochastic.

5.B1 Summary of significant effects

Table B-0 on the following page is a summary of all detected main effects and 2nd order interactions which were significant at the p=0.01 level. In addition sections B2, B3 and B4 provide detailed analysis of the import bill and total product consumption. The analysis of the remaining output variables is straight forward and available from the author upon request. Each column and subsection of the analysis reports on the univariate analysis of variance of the selected output variable. No attempt was made to perform a multivariate analysis of variance since model is completely deterministic and not stochastic.

SUMMARY OF EFFECTS SIGNIFICANT AT p LESS THAN .01

	Price of Arab Light	Consumer Bill	Import Bill	Revised Import Bill	Total Product Consumption	Gasoline Consumption	Distillate Consumption	Resid Consumption	Total Crude Imports	Sweet Crude Imports
Main Effects										
Product price elasticities	-1.28	-12.75	-5.6	-5.3	-466	-636		181	-258	-204
Refinery configuration	.27	- 1.00	16.0	-0.6	-67					
Crude pattern/availability	14.19	62.50	-7.3	18.0	-2817	-1275	-579	-409	-4197	-1921
Composition of the SPR				1.0	118			52		231
2nd Order Interactions										
Product price elasticities X Refinery configuration										
Product price elasticities X Crude pattern/availability	-1.28	-12.75	-5.6	-5.3	-466	-636		181	-258	-204
Product price elasticities X Composition of the SPR										
Refinery configuration X Crude pattern/availability	.27	1.00	5.0	-0.4	89					
Refinery configuration X Composition of the SPR										
Crude pattern/availability X Composition of the SPR				1.0	118			45		231

Table B-0

B2 Analysis of the U.S. import bill

This section has been included to demonstrate the use of factorial designs to find errors within complex computer models. A change of $21.0 billion in the U.S. import bill due to a change in refinery configuration is unreasonable. This result prompted an investigation into the method of calculating the U.S. import bill and the correction of the appropriate code. The result of the revised U.S. import bill immediately follow this section.

There seemed to be appreciable 2nd order interactions between product price elasticities, refinery configurations, and the crude pattern when the model estimates the U.S. import bill. Therefore, the first two sets of input factors had to be evaluated jointly with the crude patterns. The two-way diagrams of figure B-1 indicate the nature of these interactions.

During a 50 percent closure of the Persian Gulf with a uniform drawdown of 3 MMBD from the SPR, the model estimated that if the ALT set of elasticities are correct, rather than those compiled by the contractor (CON), the import bill will be $11,297 million less. The change in product price elasticities had no effect on the original estimate of the U.S. import bill in the BAU case.

In a business as usual environment, the model estimates that a change in the refinery configurations from the 1978 configuration to the estimated configuration in 1985 increases the import bill by $10.9 billion. With a 50 percent closure of the Persian Gulf and a uniform drawdown of the SPR of 3 MMBD, the effect of the change in refinery configuration widens to $21.0 billion and all levels are lower.

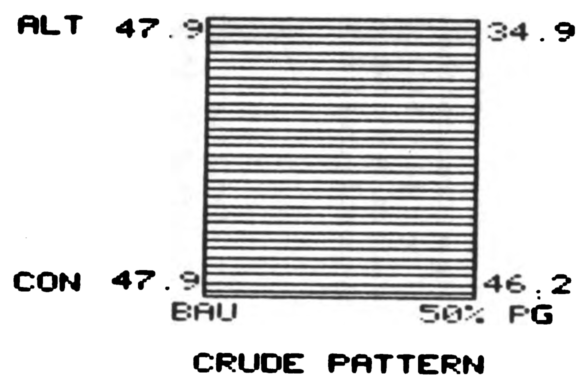
The SPR composition did not have a statistically significant effect when the model esimated the U.S. import bill over the levels of input factors tested.

STRATEGIC PETROLEUM RESERVE

figure B-1

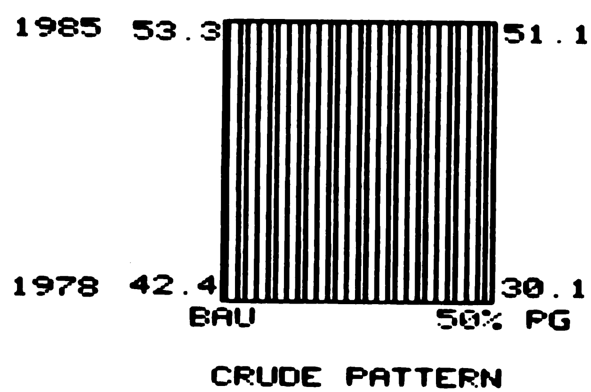
U.S. IMPORT BILL
2nd ORDER INTERACTION

PRODUCT PRICE ELASTICITIES



THE U.S. IMPORT BILL IN BILLIONS OF DOLLARS AS A FUNCTION OF PRODUCT PRICE ELASTICITIES AND CRUDE PATTERN.

REFINERY CONFIGURATION



THE U.S. IMPORT BILL IN BILLIONS OF DOLLARS AS A FUNCTION OF REFINERY CONFIGURATION AND CRUDE PATTERN.

U.S. import bill in millions of dollars

Main Effects	estimate
Product price elasticities	-5648 *
Refinery configuration	15971 *
Crude pattern/availability	-7340 *
Composition of the SPR	558

2nd Order Interactions	estimate
Product price elasticities X Refinery configuration	-902
Product price elasticities X Crude pattern/availability	-5648 *
Product price elasticities X Composition of the SPR.	1258
Refinery configuration X Crude pattern/availability	5041 *
Refinery configuration X Composition of the SPR	839
Crude pattern/availability X Composition of the SPR.	563
estimated standard error	787
level of statistical significance at p less than .01	3146

* significant effects at p less than .01

Table B-3

B3 Revised Analysis of the U.S. import bill

The previous conclusions of the import bill raised some serious questions on the techniques used within the model to estimate the bill. Further investigations lead to the discovery of an error within the software which was responsible for $21.0 billion increase. The error has subsequently been corrected and the analysis of the corrected version of the import bill appears below.

There are perceptible 2nd order interactions between product price elasticities, the composition of the SPR, the refinery configuration, and the crude pattern when the revised model estimates the U.S. import bill. Therefore, the first three sets of input factors must be evaluated jointly with the crude pattern. The two way diagrams of figure B-2 indicates the nature of these interactions.

During a 50% closure of the Persian Gulf with a uniform drawdown of the SPR, the model estimates that if the ALT elasticities are correct, rather than elasticities set CON, the U.S. import bill will be $10.6 billion less. Under the same interruption, the effect of a totally sour SPR rather than a totally sweet SPR will increase the import bill by $2 billion. With the business as usual crude patterns, neither the change in product price elasticities nor a change in the composition of the SPR has an effect on the estimated U.S. import bill. The model estimates that upgrading the refinery configuration decreases the import bill by $0.2 billion with a business as usual crude pattern. During an interruption, this same upgrading may reduce the import bill by $1.1 billion.

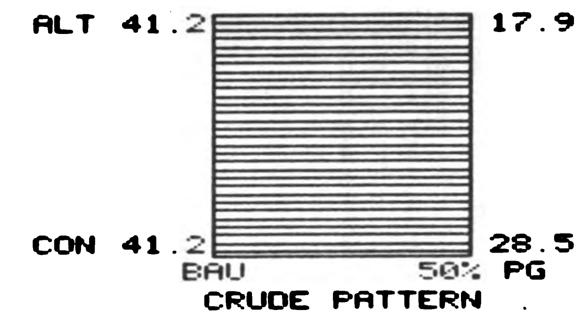
In comparing the revised estimate of the import bill with the initial estimate of the import bill, two changes are most apparent. First, the initial estimates of the import bill estimates that the average change due to upgrading the refinery configuration increased the import bill by $16.0 billion. An increase beyond reason, especially when one expects a reduction. The more reasonable result of an average decrease of 0.6 billion was estimated in the revised run. Secondly, the composition of the SPR does not have a statistically significant effect over the range of input factors tested in the original estimate of import bill, but it does in the revised version.

STRATEGIC PETROLEUM RESERVE

figure B-2

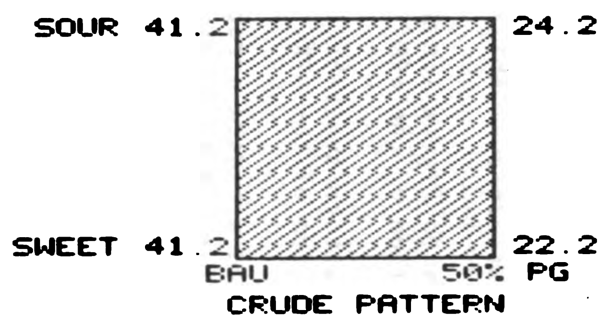
REVISED IMPORT BILL
2nd ORDER INTERACTION

PRODUCT PRICE ELASTICITIES



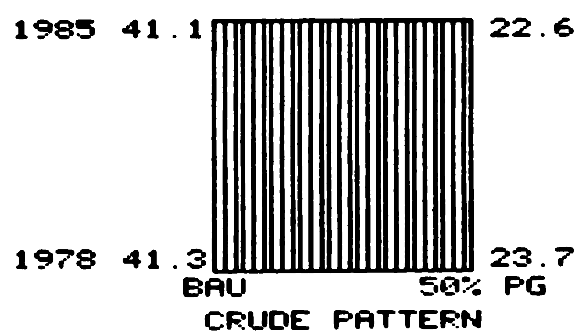
THE REVISED IMPORT BILL IN BILLIONS OF DOLLARS AS A FUNCTION OF PRODUCT PRICE ELASTICITIES AND THE CRUDE PATTERN.

COMPOSITION OF THE SPR



THE REVISED IMPORT BILL IN BILLIONS OF DOLLARS AS A FUNCTION OF THE COMPOSITION OF THE SPR AND THE CRUDE PATTERN.

REFINERY CONFIGURATION



THE REVISED IMPORT BILL IN BILLIONS OF DOLLARS AS A FUNCTION OF THE REFINERY CONFIGURATION AND THE CRUDE PATTERN.

Revised U.S. import bill in billions of dollars

Main Effects	estimate
Product price elasticities	-5.26 *
Refinery configuration	-0.64 *
Crude pattern/availability	18.01 *
Composition of the SPR	0.96 *

2nd Order Interactions	estimate
Product price elasticities X Refinery configuration	-0.19
Product price elasticities X Crude pattern/availability	-5.26 *
Product price elasticities X Composition of the SPR.	-0.11
Refinery configuration X Crude pattern/availability	-0.44 *
Refinery configuration X Composition of the SPR	-0.01
Crude pattern/availability X Composition of the SPR.	0.96 *
estimated standard error	.11
level of statistical significance at p less than .01	.42

* significant effects at p less than .01

Table B-4

B4 Analysis of the Total U.S. consumption of products

There are important 2nd order interactions between elasticities, the composition of the SPR, refinery configurations, and the crude pattern which effect the estimates of total consumption of products. Therefore, each of the first three sets of input factors must be evaluated jointly with the input factors representing the crude pattern. The three two-way diagrams depicting the nature of the interactions appear in figure B-3.

During a 50 percent closure of the Persian Gulf, the use of the ALT elasticities rather than the CON elasticities decreases the estimated total U.S. consumption of products by 933 MBD. Under the same interruption, a sour SPR produces an estimated increase in the total consumption of products by 237 MBD over use of a sweet SPR.

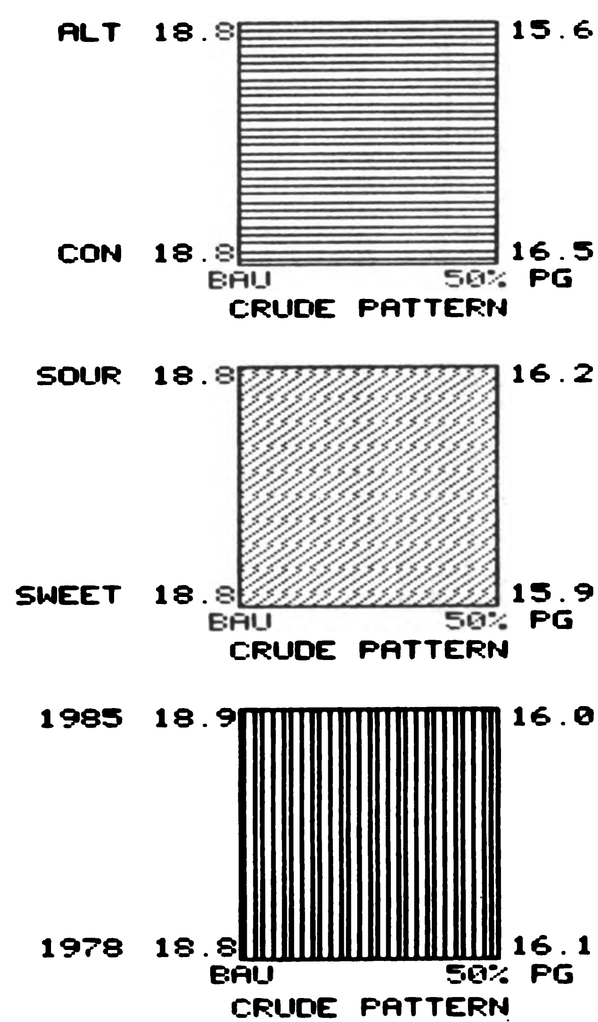
When using the BAU crude pattern, the model estimates that neither variable has a statistically significant effect on the total U.S. consumption of products.

With a crude pattern representing the BAU case, upgrading the refinery configuration from 1978 to 1985 increases the total U.S. consumption of products by about 22 MBD, whereas the same upgrading during a 50 percent closure of the Persian Gulf with a uniform drawdown of the SPR of 3 MMBD decreases the total U.S. consumption of products by about 155 MBD. This is the only example in the study of a crossed pattern and is worthy of further investigation. This crossed pattern is contrary to anticipated results.

STRATEGIC PETROLEUM RESERVE

figure B-3

TOTAL U.S. CONSUMPTION OF PRODUCTS IN MMBD
2nd ORDER INTERACTION



PRODUCT PRICE ELASTICITIES

TOTAL U.S. CONSUMPTION OF PRODUCTS AS A FUNCTION OF PRODUCT PRICE ELASTICITIES AND CRUDE PATTERN.

COMPOSITION OF THE SPR

TOTAL U.S. CONSUMPTION OF PRODUCTS AS A FUNCTION OF THE COMPOSITION OF THE SPR AND CRUDE PATTERN.

REFINERY CONFIGURATION

TOTAL U.S. CONSUMPTION OF PRODUCTS AS A FUNCTION OF THE REFINERY CONFIGURATION AND CRUDE PATTERN.

The U.S. total product consumption in MMBD

Main Effects	estimate
Product price elasticities	-466 *
Refinery configuration	-67 *
Crude pattern/availability	-2817 *
Composition of the SPR	118 *

2nd Order Interactions	estimate
Product price elasticities X Refinery configuration	-20
Product price elasticities X Crude pattern/availability	-466 *
Product price elasticities X Composition of the SPR.	24
Refinery configuration X Crude pattern/availability	89 *
Refinery configuration X Composition of the SPR	-8
Crude pattern/availability X Composition of the SPR.	118 *
estimated standard error	16
level of statistical significance at p less than .01	66

* significant effects at p less than .01

Table B-5

C1 Results

The corresponding, results of the experiment were:

a. Significant differences were detected in the United States crude oil and product import patterns associated with drawdown of the two SPR types, but relatively little differences in the level and pattern of product demands supplied. Of the products examined, only the consumption of residual fuel oil was significantly effected by a change in the composition of the SPR.

b. Significant differences in all examined variables were detected as a result of changing the availability of crude oil.

c. Changes in the refinery configuration create small, but significant effects in several measured variables. However, changes in the refinery configuration should not effect the recommended mix of crude oil to be stored in the reserve since the 2nd order refinery configuration X composition of the SPR interaction is not statistically significant for the variables tested.

d. Significant differences in the pattern of US demands, for several products, resulted from the application of the alternate sets of product demand price elasticities. However, changes of the elasticities should not effect the recommended mix of crude oils to be stored in the reserve since the 2nd order elasticity X composition of the SPR interaction is not statistically significant for the variables tested.

Recommendation

The factorial design detected errors within the US import bill and several other output variables. It was recommended that this model be corrected before it was put into production use.

Post Script

In addition to the procedure described in this paper, other checks of the quality of the software were made including an analysis of the estimated results for the year 1978 against historical data. The contractor incorporated the recommended corrections and in the fall of 1982 the corrected model produced results which were diametrically opposite of earlier runs.

6. Bibliography

Box, G.P.E. and D.R. Cox (1964). An analysis of transformations, J. Roy. Stat. Soc., Series B, 26,211.

Box, G.P.E. and N.R. Draper (1969). Evolutionary Operation: A Statistical Method for Process Improvement, Wiley.

Box G.P.E. and J.S. Hunter (1961) The 2^{k-p} factorial design, Technometric, 3,311,449.

Box, G.P.E., J.S. Hunter and W.G. Hunter (1978) Statistic for Experimenters, Wiley

Cochran, W.G. and G.M. Cox (1957).Experimental Design, 2nd ed. Wiley

Daniel, C (1976). Applications of Statistics to Industrial Experimentation, Wiley.

Daniel, C (1959). Use of half - normal plot initerpreting factorial two-level experiments, Techonomtrics, 1,149

Davies, O.L. (Ed.) (1971). The Design and Analysis of Industrial Experiments, Hafner (Macmillan).

Hunter, J.S. (1966). Inverse Yates algorithum, Technometrics.

Hunter, W.G. (1976). Some ideas about teaching design of experiments, with 2^5 examples of experiments conducted by students, Am. Stat., 31,12.

Hunter, W.G. and E. Chacko (1971). Increasing industrial productivity in developing countries, Int. Rev. Div., 13,3,11.

Jenkins G.M. (1969). A systems study of a petrochemical plan, J.Syst Enh., 1, 90.

Maske, D.M. (1970). High - rate, fine-mesh screening of combined wastewater flows, J Water Pollut Control Fed., 42, 1476.

Plackett, R.L. and J. P. Burman (1946). The design of optimum multifactorial experiments, Biometrics, 33,305.

Yates, F (1937).The Design and Analysis of Factorial Experiments, Bulletin 35, Imperial Bureau of Soil Science, Harpenden, Herts England, Hafner (Macmillan).

Yates F. (1970). Selected Papers, Hafner (Macmillan)

Youden, W.J. (1972). Enduring values, Technometrics, 14,1.

Nonparametric Probability Density Estimation for Data Analysis in Several Dimensions[1]

David W. Scott

Rice University
Houston, TX 77251

1. Introduction

Our purpose in this paper is to illustrate how nonparametric probability density estimates, in particular the corresponding contour curves, are a useful adjunct to scatter diagrams when performing a preliminary examination of a set of random data in several dimensions. For a preliminary approach we generally want to perform fairly simple tasks with free-form techniques to uncover structures and features of interest in the data. Such procedures are often graphical and unlike summary statistics seldom lead to much compression of the data. Tukey (1977) presents a wealth of such procedures. One which well illustrates the power and flexibility of these preliminary procedures is the running median smoothing algorithm for time series data (with resmoothing of the rough and the like). Other graphical techniques for multivariate data are presented in Tukey and Tukey (1981).

For preliminary viewing of one-dimensional data, both scatter diagrams and frequency curves such as histograms are widely and successfully employed to examine clustering, tail behavior, and skewness of

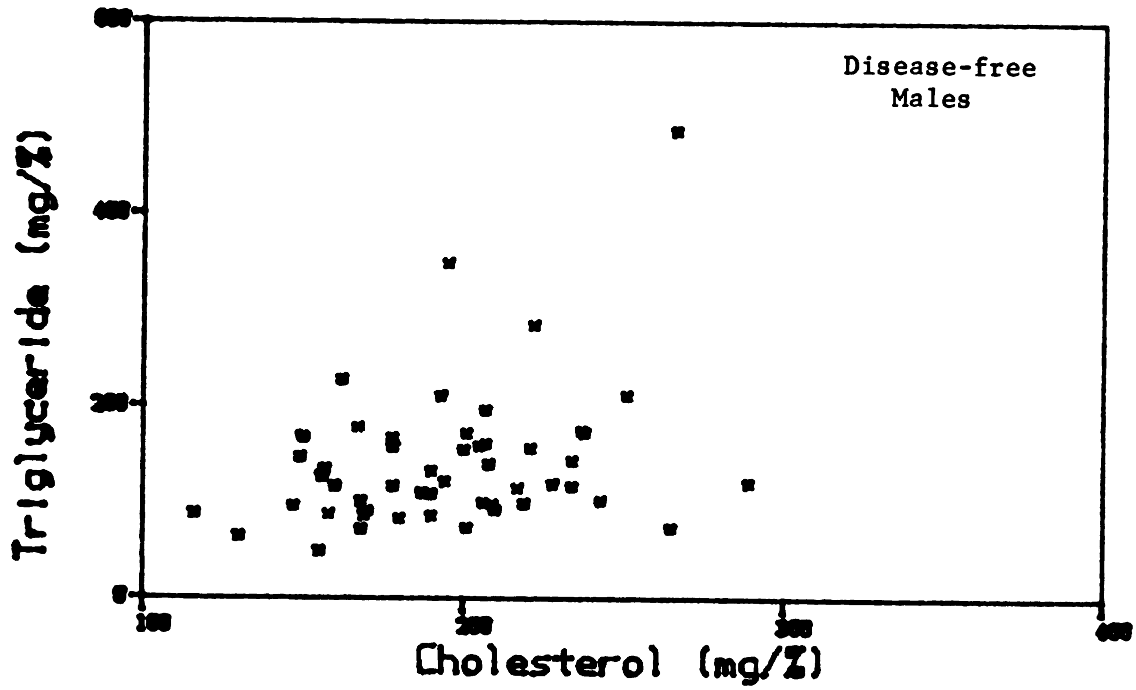
[1]This research was supported in part by the Army Research Office under DAAG-29-82-K-0014 and by NASA/Lockheed under PO-0200100079.

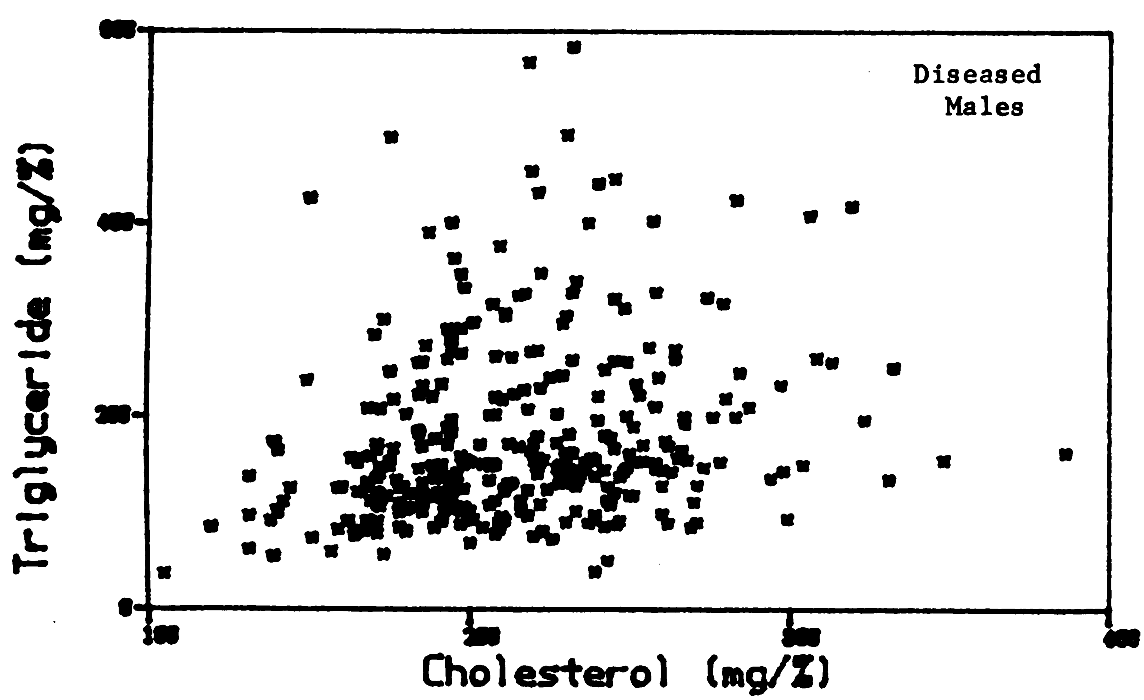
data. For bivariate data, scatter diagrams are in practice widely preferred to bivariate frequency curves. Scatter diagrams of three dimensional data may be realized by viewing a projection of the data on a rotating plane represented by the screen on a computer graphics terminal. For higher dimensions carefully selected projections may also be viewed, and sophisticated techniques have been developed, and are evolving, for choosing good projections (Friedman and Tukey, 1974). Apparently the success of frequency curves in one dimension has not readily extended to higher dimensions. It is an open question as to the number of dimensions that may be successfully visualized with a nonparametric density estimator under various conditions (sample size, for example). It is our purpose to illustrate the power of preliminary frequency curves as an adjunct to viewing scatter diagrams.

2. Bivariate Data

We shall examine a data set which contains information on the status of the coronary arteries of 371 men suspected of having heart disease, having experienced episodes of severe chest pain. These data have been more fully described and analyzed; see Gotto, et al. (1977) and Scott, et al. (1978). After visual examination of the coronary arteries by angiography, 51 men were determined to be free of significant coronary artery disease. It was of interest to compare the levels of blood fats, plasma cholesterol and plasma triglyceride concentrations, between the group of 51 disease-free males and the group of 320 diseased males. The scatter diagrams of these two data sets are displayed in Figure 1. Patients with elevated levels of cholesterol and

Figure 1. Scatter Diagrams





triglyceride are evident among the diseased males. This observation is difficult to evaluate in light of the large difference in sample sizes. However, it is unlikely that a larger sample of 320 disease-free males would result in a scatter diagram similar to that of the 320 diseased males.

To obtain a nonparametric density contour plot we computed a bivariate product kernel estimate (Epanechnikov, 1969) given by

$$f(x,y) = \frac{1}{nh_x h_y} \sum_{i=1}^{n} K\left(\frac{x_i - x}{h_x}\right) K\left(\frac{y_i - y}{h_y}\right) \qquad (1)$$

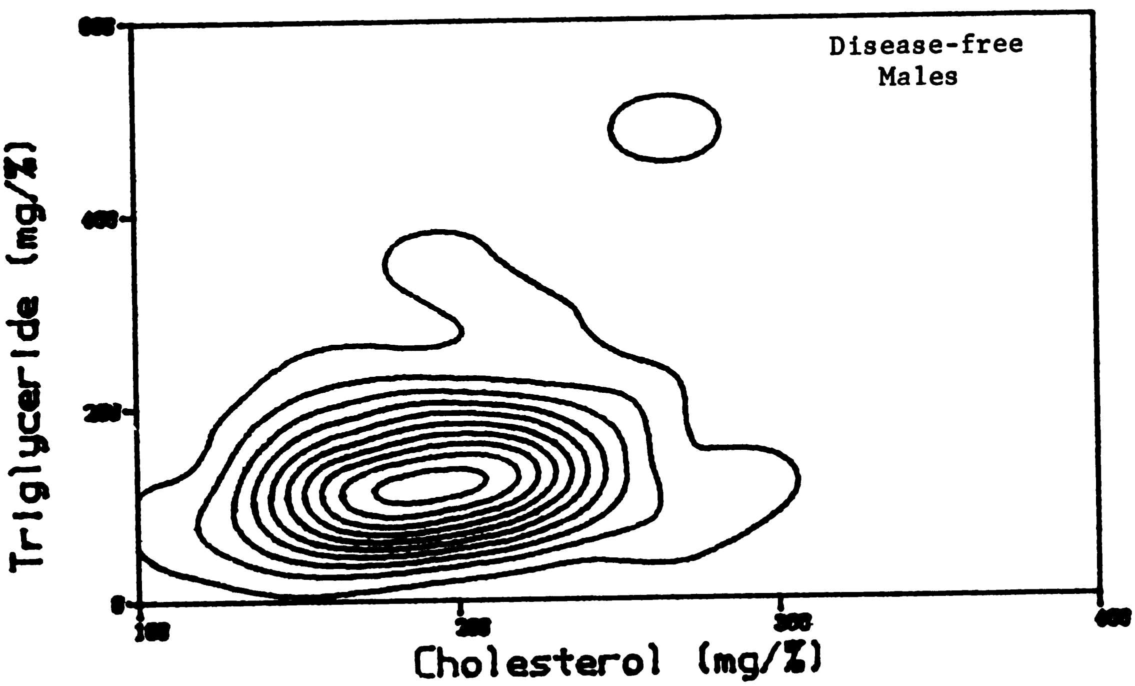
using a quartic (biweight) kernel

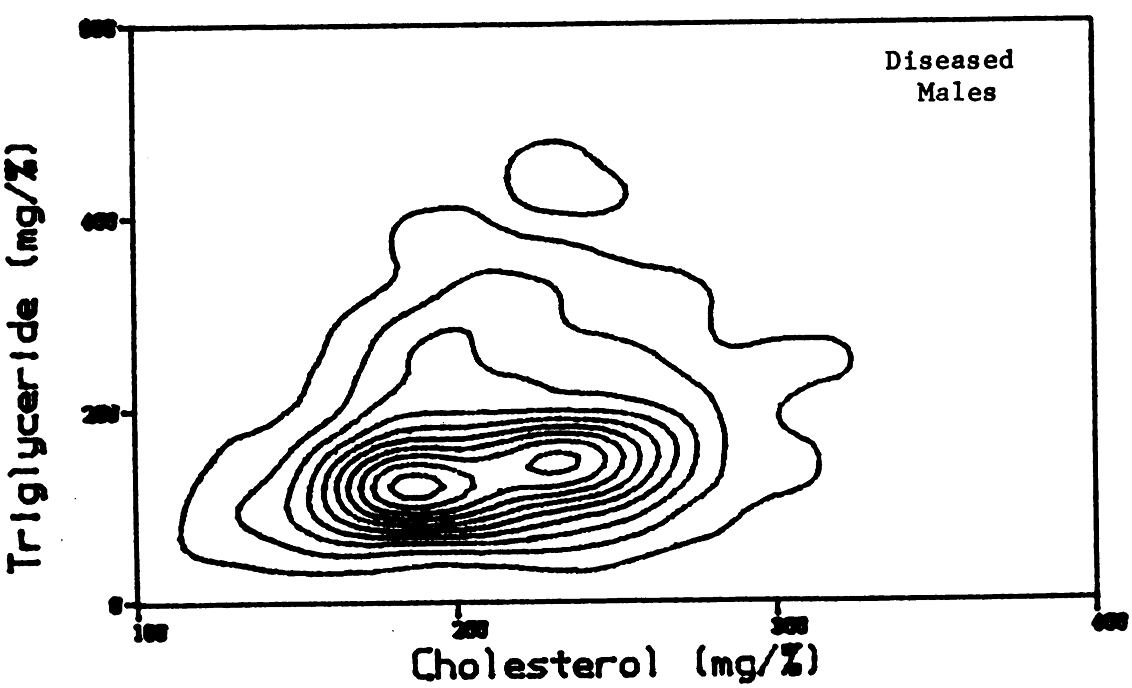
$$K(z) = \frac{15}{16}(1-z^2)^2 \, I_{[-1,1]}(z) \qquad (2)$$

and preliminary values of the smoothing parameters given by $h_x = 2\,s_x\,n^{-1/6}$ where s_x represents a trimmed and pooled estimate of the standard deviation for the two groups with a similar expression for h_y. Density values were computed over a grid of 150 by 90 points. When applied to the data for the diseased males, the contour plot reveals a striking bimodal feature, as shown in Figure 2. The contours of equal probability are at the ten levels 0.05 to 0.95 in increments of 0.10 as a fraction of the respective maximal modal levels. The density function of the disease-free males could be well approximated by a bivariate Normal form. Its mode coincides with the left of the two modes in the density function of the diseased males.

The contour plots have helped emphasize a feature in the scatter diagram that might have gone unnoticed. The contour plots also aid in compensating for the difference in sample sizes. The discovery of the bimodal feature led to formulation of a complex cholesterol-triglyceride

Figure 2. Bivariate Density Contours





interaction in the model for estimating the risk of coronary artery disease. Clinically, the difference of 50 mg/% between the two modes in Figure 2 for the diseased males is greater than the reduction in cholesterol by dietary intervention (which usually achieves proportional reductions in the range of 10 to 15 percent).

3. Trivariate Data

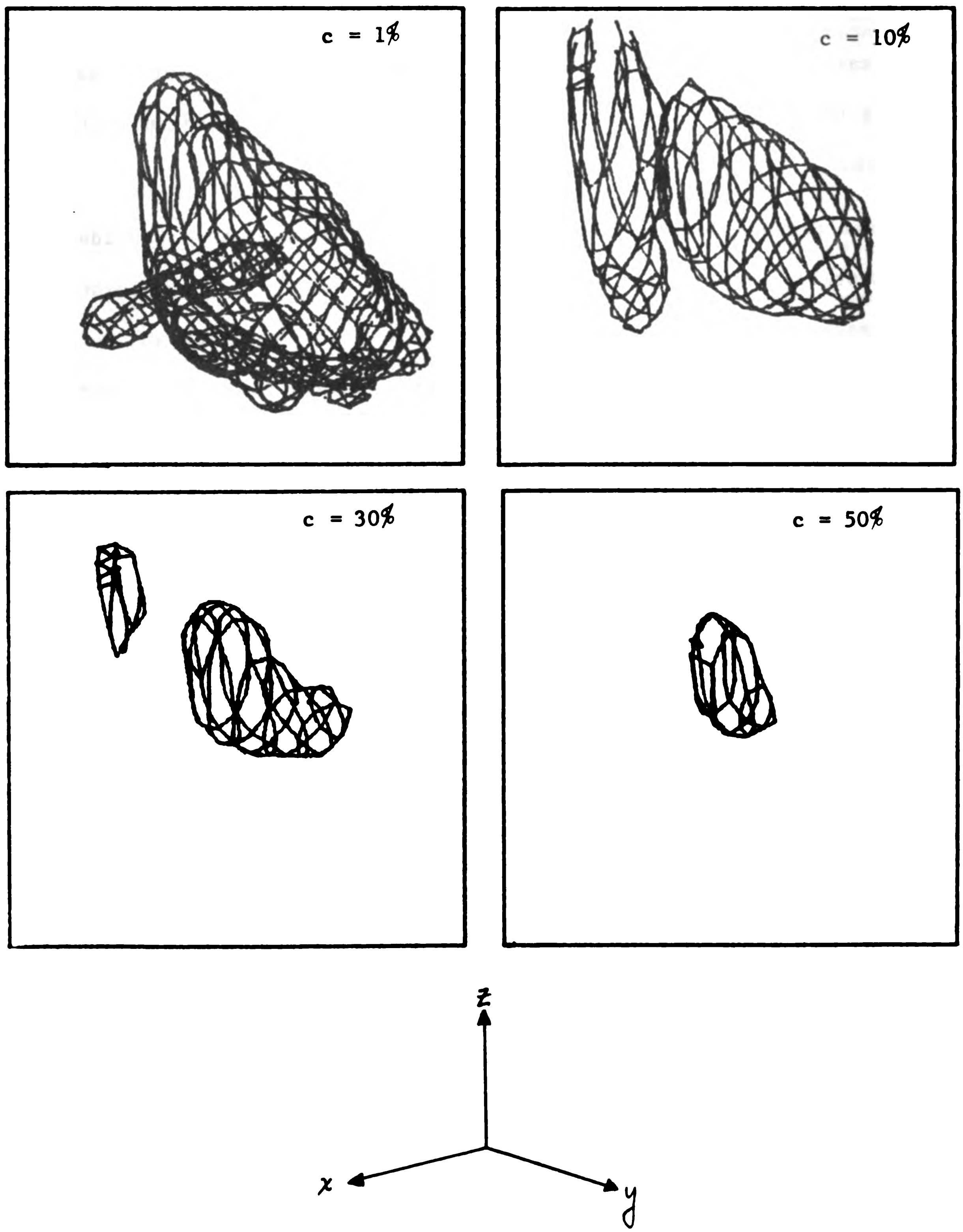
The data presented in this section were obtained by processing four-channel Landsat data measured over North Dakota during the summer growing season of 1977 and were furnished by Dick Heydorn of NASA/Houston and Chuck Sorensen of Lockheed/Houston. The sample contains approximately 21,000 points, each representing a 1.1 acre pixel, covering a 5 by 6 nautical mile section. On each pass over an individual pixel by the Landsat satellite, the four channel readings were combined into a single value that measures the "greenness" of the pixel at that time. The greenness of a pixel was plotted as a function of time from the five passes during the growing season. Finally, Badhwar's (1982) growth model was fitted to this curve. This model has three parameters which are contained in each trivariate data point. The first variable (x) gives the time the "crop" (if any) ripened. The second variable (y) measures the approximate time to ripen. And the third variable (z) measures the level of "greenness" at the time of ripening. Although it is natural to group these data by actual type of ground cover for classification procedures, we have not done so here.

It is not possible to present a satisfactory picture of a three-dimensional scatter diagram of these data for this article. However, on an AED512 terminal with 512 by 512 resolution, a projection of these data onto the screen typically displayed only 4000 points, the rest being "hidden" behind displayed points. Viewed from several different angles, various shapes and features in the data were easily perceived. Color was used to indicate the level of the variable perpendicular to the screen.

We can present density contours of an estimate f(x,y,z). Consider an equiprobable contour at level c; that is, consider those points (x,y,z) satisfying the equation f(x,y,z) = c. The solution of this equation for a smooth density estimate f is a smooth surface (or surfaces) in $\mathbb{R}^3$. This surface may be displayed by intersecting it with a series of planes displaced equal distances along the co-ordinate axes, in the following, along only the x and y axes. In Figure 3, we display the surface for c = 1% of the maximal mode value. Comparing Figure 3 to the corresponding scatter diagram on the same projection plane reveals how surprisingly little of the data space is enclosed in this contour. In the scatter diagram our eyes focused on rays of points that seemed interesting but represented only a small fraction of the data. Also notable in Figure 3 is a cylindrical shape disjoint and behind the larger surface. This feature was also clearly visible in the scatter diagram and represents acres in which sugar beets were grown. Apparently the method by which sugar beets are harvested leads to a singularity in the estimation of the growth model parameters with $y \approx 0$.

Expanding the scale by a factor of 2 while retaining the same center as in the c = 1% picture, we show the contour shapes at levels c = 10%, 30%, and 50% of modal height. Notice how each contour shape

Figure 3. Trivariate Density Contours



"fits" inside the preceding one. Also observe how multimodal features appear in this space. Three modes are shown in this sequence. On a color graphics terminal, we may simultaneously view these and other contours by using different colors to draw each contour.

Again, the density plots have complemented and added to our understanding of these data. It is easier to see inside the data cloud with this representation and also makes rotation of the data cloud less important.

4. Computational Considerations

A new algorithm and density estimator were developed to display the trivariate contour plots and we hope to report on it in another paper (Scott, 1983b). Speed is an important factor in an interactive environment. The kernel method used in the bivariate case becomes excruciatingly slow when presented with 21,000 points in three dimensions. In real time, a few minutes were required on a Vax 11/780 to compute the bivariate kernel contours for 320 points on a 150 by 90 mesh. To generate the pictures in Figure 3, we evaluated the density on a 30 by 30 by 30 mesh for 21,000 points. A straightforward kernel estimator would have required several hours to compute!

The histogram estimator is extremely efficient computationally, but very inefficient statistically -- and relatively more inefficient in higher dimensions than kernel methods. One recent discovery indicates that the frequency polygon may be a good choice of a nonparametric density estimator since it is computationally equivalent to a histogram but statistically similar to a kernel estimate (Scott, 1983a). However, the frequency polygon in several dimensions suffers from sensitivity to choice of cell boundaries. The new algorithm addresses this problem and is asymptotically equivalent to a certain kernel estimate. Other fast preliminary estimates in one and two dimensions may be obtained by numerical approximation of kernel estimates in place of statistical approximation, which we prefer.

5. Where Do We Go?

We do not really know for how many dimensions nonparametric density estimates will be useful and feasible. Scatter diagrams have been used in a highly interactive environment to visualize nine-dimensional data (Tukey, Friedman, and Fisherkeller, 1976). Many possible strategies may be envisioned for using color and motion to examine data in more than three dimensions. We expect much progress in this area. But for larger and larger data sets requiring sophisticated analysis, we believe that density-based methods will be both efficient and effective.

REFERENCES

Badhwar, G.D., J.G. Carnes, and W.W. Austin (1982), "Use of Landsat-Derived Temporal Profiles for Corn-Soybean Feature Extraction and Classification," Remote Sensing of Environment, 12, 57-79.

Epanechnikov, V.A. (1969), "Nonparametric Estimates of a Multivariate Probability Density," Th. Prob. and Appl., 14, 153-158.

Friedman, J.H. and J.W. Tukey (1974), "A Projection Pursuit Algorithm for Exploratory Data Analysis," IEEE Trans. Comp. C-23, 881-890.

Gotto, A.M., G.A. Gorry, J.R. Thompson, J.S. Cole, R. Trost, D. Yeshurun, M.E. DeBakey (1977), "Relationship Between Plasma Lipid Concentration and Coronary Artery Disease in 496 Patients," Circulation, 56:5, 875-883.

Scott, D.W. (1983a), "Optimal Frequency Polygons: Theory and Application," submitted.

Scott, D.W. (1983b), "Average Shifted Histograms," working paper.

Scott, D.W., A.M. Gotto, J.S. Cole, and G.A. Gorry (1978), "Plasma Lipids as Collateral Risk Factors in Coronary Artery Disease - A Study of 371 Males with Chest Pain," J. Chronic Diseases, 31, 337-345.

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

Tukey, J.W., J.H. Friedman, and M.A. Fisherkeller (1976), "PRIM-9, an Interactive Multidimensional Data Display and Analysis System," Proc. 4th Inter. Congress for Stereology, Sept. 4-9, 1975, Gaithersburg, Maryland.

Tukey, P.A. and J.W. Tukey (1981), "Graphical Display of Data Sets in 3 or More Dimensions," in Interpreting Multivariate Data, V. Barnett, ed., John Wiley & Sons, New York.

LOGISTIC SUPPORTABILITY

One afternoon of the 28th Conference on the Design of Experiments in Army Research Development and Testing was devoted to the important area of Logistic Supportability. First on the agenda was the presentation by OTEA. It was entitled "Improving the Test and Evaluation of Integrated Logistics Support in OT," and is published in these proceedings in the format of a slide presentation. The DARCOM presentation came next and carried the title "Supportability - Requirements, Design Test and Evaluation." Unfortunately, no printed matter was submitted for publication concerning this address. The final presentation, "Logistic Supportability Testing and Evaluation During OT," was given by TRADOC personnel. Their report directly follows that made by OTEA.

INTEGRATED LOGISTIC SUPPORTABILITY TESTING IN OT

FORTY YEARS AGO

"I DON'T KNOW WHAT THE HELL THIS LOGISTICS IS THAT MARSHALL IS ALWAYS TALKING ABOUT, BUT I WANT SOME OF IT"

ADM ERNEST J. KING, 1942

PANEL MEMBERS

MR WALTER HOLLIS, DUSA/OR

MR RUSSELL R. SHOREY, OASD/MRA&L

MG WILLIAM HUNZEKER, USA LOG CENTER

DR R. ELY, THE JOHNS HOPKINS UNIVERSITY
APPLIED PHYSICS LABORATORY

BG KENNETH A. JOLEMORE, DCSLOG/S&M

COL W. B. MOATS, AFTEC

GOAL

TO IDENTIFY IMPROVED METHODS FOR THE TESTING AND EVALUATION OF INTEGRATED LOGISTIC SUPPORTABILITY FOR ALL SYSTEMS UNDERGOING OPERATIONAL TESTING.

LANGHORNE WITHERS

AREAS COMMON TO ALL THREE PRESENTATIONS

EACH PRESENTATION:

- ADDRESSES THE SAME 12 ILS ELEMENTS
- ASSUMES THAT THE MAJOR PROVISIONS OF CARLUCCI INITIATIVES 9 & 12 WILL BE IMPLEMENTED

  INITIATIVE 9: IMPROVED SYSTEM SUPPORT AND READINESS

  - READINESS IS COMPARABLE IN IMPORTANCE TO REDUCED UNIT COST AND ACQUISITION TIME
  - RESOURCES TO ACHIEVE READINESS WILL RECEIVE SAME EMPHASIS AS THOSE REQUIRED TO ACHIEVE SCHEDULE OR PERFORMANCE OBJECTIVES

  INITIATIVE 12: PROVIDE FRONT END FUNDING FOR TEST HARDWARE

  - VISIBILITY AND PRIORITY IN PLANNING DOCUMENTS
  - DEMONSTRATION OF SYSTEM SUPPORTABILITY BEFORE PRODUCTION AND FIELDING
- RECOGNIZES THAT TEST DISCIPLINES FOR LOGISTIC SUPPORTABILITY TESTING NEED TO BE IMPROVED

DISCIPLINES FOR LOGISTIC SUPPORTABILITY TESTING
NEED TO BE IMPROVED

ILS ELEMENTS

1. DESIGN INFLUENCE
2. MAINTENANCE PLANNING
3. MANPOWER & PERSONNEL
4. SUPPLY SUPPORT
5. SUPPORT EQUIPMENT TO INCLUDE TMDE
6. TRAINING & TRAINING DEVICES
7. TECHNICAL DATA
8. COMPUTER RESOURCES SUPPORT
9. PACKAGING, HANDLING, STORAGE
10. TRANSPORTATION/TRANSPORTABILITY
11. FACILITIES
12. STANDARDIZATION AND INTEROPERABILITY

SOURCE: AR700-127

SUPPORTABILITY TEST AND EVALUATION

INITIATIVES AND THRUSTS

LTC DAVIS

RECOGNIZED PROBLEMS

- INADEQUATE SYSTEM SUPPORT PACKAGES (SSP)
- SEVERE LIMITATIONS ON ILS T&E

KEY ELEMENTS OF ARMY IMPLEMENTATION

- SYSTEM READINESS OBJECTIVES (SRO)
- TECHNOLOGICAL INNOVATIONS/SELF-SUSTAINABILITY
- "FAST TRACK" INTEGRATED LOGISTIC SUPPORT (ILS) AND RAM GUIDELINES
- STREAMLINE FUNCTIONAL PROCESSES
- SPECIAL MANAGEMENT FOR "FAST TRACK" PROGRAMS
- ORGANIZATIONAL RESPONSIBILITIES AND REGULATORY CHANGES
- SUPPORTABILITY TESTING GUIDELINES

AR 700-127 REGULATORY CHANGES (CURRENT)

- ACCELERATED ACQUISITION PROGRAM ILS GUIDELINES
- LIFE CYCLE ILS (DISPLACED SYSTEMS)
- SYSTEM READINESS OBJECTIVES (SRO)
- SUPPORTABILITY TESTING AND EVALUATION
- CONTRACTOR SUPPORT

TEST AND EVALUATION GUIDELINES

EMPHASIS:

- PROVIDE FOR BEST EFFORT IN SUPPORTABILITY ASPECTS OF DEVELOPMENT AND OPERATIONAL TESTING
- KEY SYSTEM SUPPORT TEST DESIGN AND EVALUATION TO IDENTIFIED ISSUES AND CONCERNS
- PROVIDE METHODOLOGY TO TREAT TESTING ARTIFICIALITIES
- MINIMIZE TESTING REDUNDANCY AND WIDEN DATA BASE
- PROVIDE FOR ANALYTICAL PROJECTION OF SYSTEM MANPOWER AND RESOURCE IMPACTS (TRADOC MALA METHODOLOGY - COMPLETE)

TASKS:

- INCORPORATE SUPPORTABILITY TESTING GUIDELINES FOR USER TESTING ACTIVITIES
- DEVELOP AND PUBLISH SUPPORTABILITY TESTING GUIDELINES
- DEVELOP REVISED DEFINITIONS AND POLICY GUIDELINES FOR SYSTEM SUPPORT PACKAGE (SSP) CONCEPT
- PROVIDE ADDITIONAL MANAGEMENT GUIDELINES FOR DESIGNING IN RELIABILITY AND SUPPORTABILITY

TEST AND EVALUATION ISSUES AND OBJECTIVES

- ACHIEVEMENT OF READINESS GOALS AND OBJECTIVES *
- ADEQUACY AND IMPACT OF:
  - PUBLICATIONS
  - TOOLS
  - TMDE
  - FACILITIES
  - HANDLING EQUIPMENT
  - PLL/ASL/BII, ETC
  - TRAINING/TRAINING DEVICES
  - PERSONNEL (NUMBERS, SKILLS)
  - DESIGN FACTORS (RAM, BITE, TRANSPORTABILITY)
  - ANCILLARY EQUIPMENT (TRUCKS, GENERATORS, RADIOS)
- CONSIDERATION OF:
  - MAN-MACHINE INTERFACE (HUMAN FACTORS)
  - SAFETY

* PRINCIPAL ISSUE

SYSTEM SUPPORT PACKAGE (SSP)

- SUPPORT ELEMENTS FOR TESTING AND EVALUATION
- TAILORED TO SUPPORTABILITY TEST ISSUES
- INCLUDES:
  - TECHNICAL PUBLICATIONS
  - SPARES AND REPAIR PARTS
  - TRAINING DEVICES AND INSTRUCTIONAL MODULES
  - SPECIAL AND COMMON TOOLS
  - TMDE
  - TRAINED PERSONNEL
  - TRANSPORTATION AND HANDLING EQUIPMENT
  - CALIBRATION PROCEDURES AND EQUIPMENT
  - MOBILE SUPPORT FACILITIES
  - ANCILLARY EQUIPMENT REQUIRED TO ROUND-OUT OPERATIONAL SYSTEM CONFIGURATION
- MATERIEL DEVELOPER RESPONSIBLE FOR PROPER CONSTITUTION AND TIMELY DELIVERY IN COORDINATION WITH COMBAT DEVELOPERS, TRAINING DEVELOPERS, TESTERS AND EVALUATORS

DELIVERY IN COORDINATION ... FOR PROPER CONSTITUTION AND TIMELY
... WITH COMBAT DEVELOPERS, TRAINING DEVELOPERS,

GOALS

- TIMELY DELIVERY OF SYSTEM SUPPORT ELEMENTS
- FORMULATION OF IMPROVED TEST AND EVALUATION TECHNIQUES AND PROCEDURES
- SUPPORTABLE AND AFFORDABLE MATERIEL SYSTEMS IN THE 1980'S AND BEYOND

INTEGRATED LOGISTIC SUPPORTABILITY TESTING IN OT

STEVE FRENCH

HISTORICAL TEST LIMITATIONS

- INADEQUATE SUPPORT PACKAGES
- IMMATURE HARDWARE/SOFTWARE
- INSUFFICIENT NUMBERS OF EQUIPMENT TO TEST
- SHORT TEST PERIODS
- INCONSISTENT TEST DISCIPLINE
- RANGE/SCHEDULE/SORTIE LIMITATIONS
- COST

SUPPORT PACKAGE LIMITATIONS WHICH IMPACTED OTEA IERS

ILS INADEQUACY	PATRIOT OT II	DIVAD CHECK TEST	DGM IOT&E	TDF IOT&E	AAH OT II	PLRS OT II	CH47 OT II	TTC39 OT II	ROLAND	XM1 OT III	TPQ 37 OT III	HELLFIRE OT II	VIPER
ANCILLARY EQUIPMENT	X	X	X									NONE	NONE
BITE	X	X						X					
MAINTENANCE	X	X	X	X	X	X	X						
MANUALS	X	X	X		X	X	X		X	X			
SUPPLY ORGANIZATION	X	X				X		X	X		X		
TMDE	X	X	X		X	X			X	X			
MAINTENANCE TRAINING	X	X	X	X				X					
TOOLS & EQUIPMENT	X	X	X	X	X		X						
SPARE PARTS	X	X	X	X									

EVALUATION PROBLEM FOR OTEA

WITHIN THE BOUNDS OF OPERATIONAL TEST AND EVALUATION, DEVELOP A METHODOLOGY WHICH MEASURES HOW TESTED ILS ELEMENTS CONTRIBUTE TO "DOWNTIME" IN THE WARTIME READINESS MEASURE

INHERENT TEST LIMITATIONS

- LACK OF REALISTIC COMBAT DAMAGE TO EQUIPMENT
- LACK OF REALISTIC COMBAT DAMAGE TO SUPPORT FACILITIES
- LIMITED SAMPLE OF GENERAL SUPPORT MAINTENANCE ACTIONS
- LIMITED LOADING OF DS/GS MAINTENANCE ORGANIZATIONS
- INABILITY TO FORECAST SUPPORT SYSTEMS POTENTIAL AS IT IS MATURING
- LIMITED ABILITY TO TEST TMDE/BIT SOFTWARE

ARMY PROBLEM

IT NEEDS TO BE RECOGNIZED THAT ALTHOUGH OTEA CAN MAKE POSITIVE CONTRIBUTIONS TO THE T&E OF ILS, PRIMARILY AT THE UNIT AND DS LEVELS, THE NEED FOR FURTHER EVALUATION CALLS FOR EXPERTISE AND RESOURCES OUTSIDE THIS AGENCY

OTEA CONTRIBUTION TO ARMY ILS EVALUATION

	UNIT	DIRECT SUPT.	GENERAL SUPT.	DEPOT	CONUS	LOC	THEATER
DESIGN INFLUENCE	x	✓					
MAINT. PLAN	x	✓					
MANPOWER	x	x	✓				
SUPPLY SUPT.	x	✓					
SUPT. EQUIP.	x	✓					
TRAINING	x	✓					
TECH. DATA	x	✓					
COMPUTER RESOURCES	✓						
PACK. HAND. AND STOR.	x	✓					
TRANSPORT	x	✓					
FACILITIES	x	✓					
STANDARDIZATION AND INTEROPERABILITY	✓						

x SUBSTANTIAL EVALUATION
✓ PARTIAL EVALUATION

x SUBSTANTIAL EVALUATION
✓ PARTIAL EVALUATION

INTEGRATED LOGISTIC SUPPORTABILITY TESTING IN OT

BILL DUNN

IMPROVING LOGISTIC SUPPORTABILITY IN OT

IMPROVEMENTS TO LOGISTIC SUPPORTABILITY TESTING AND EVALUATION:

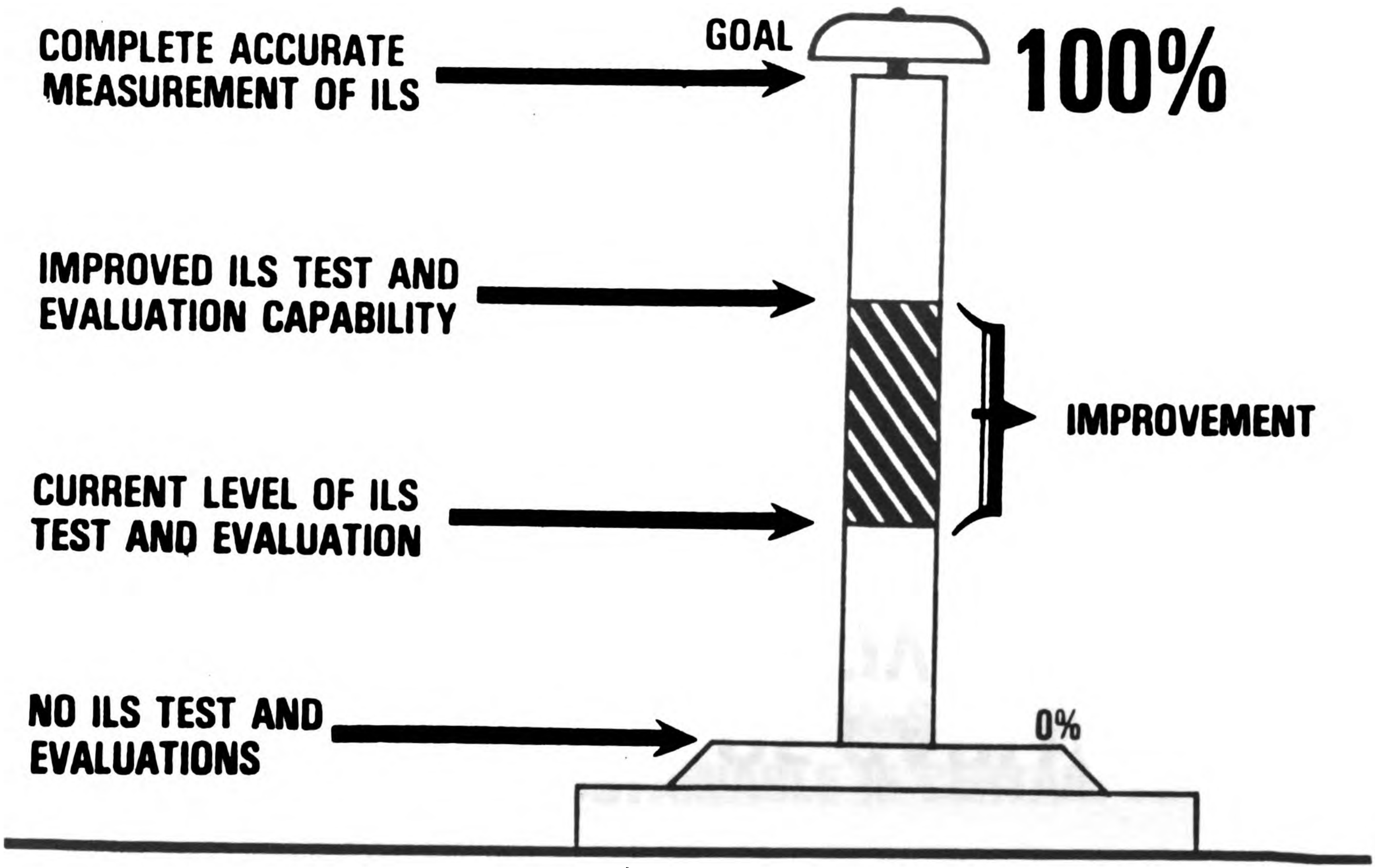
- RECOGNIZING OPPORTUNITIES AND CAPITALIZING ON PROVISIONS OF CARLUCCI INITIATIVES
- USING A MIXTURE OF TEST & SIMULATION DATA WHEN NECESSARY
- RELATING LOGISTIC SUPPORTABILITY MEASURES TO WARTIME SRO's
- ANALYZING THE IMPACT OF PROBLEMS ASSOCIATED WITH THE 12 ILS ELEMENTS ON DOWNTIME
- IDENTIFYING INCREASED LEVELS OF READINESS POTENTIALLY OBTAINABLE IF CERTAIN FIXES ARE MADE

POTENTIALLY OBTAINABLE IF CERTAIN FIXES ARE MADE

IMPACT OF CARLUCCI INITIATIVES

- ADDED IMPORTANCE OF READINESS
- IMPROVED SUPPORT PACKAGES DUE TO FRONT END FUNDING
- MORE TEST HARDWARE

EXPECTED IMPACT OF CARLUCCI INITIATIVES, IMPROVED TEST DISCIPLINE, AND EVALUATION METHODOLOGY



CAPITALIZING ON CARLUCCI INITIATIVES

- PARTICIPATION IN THE INTEGRATED LOGISTIC SUPPORT MANAGEMENT TEAM IN ORDER TO EVALUATE THE ADEQUACY OF SSP CONTENTS AND DELIVERY SCHEDULE

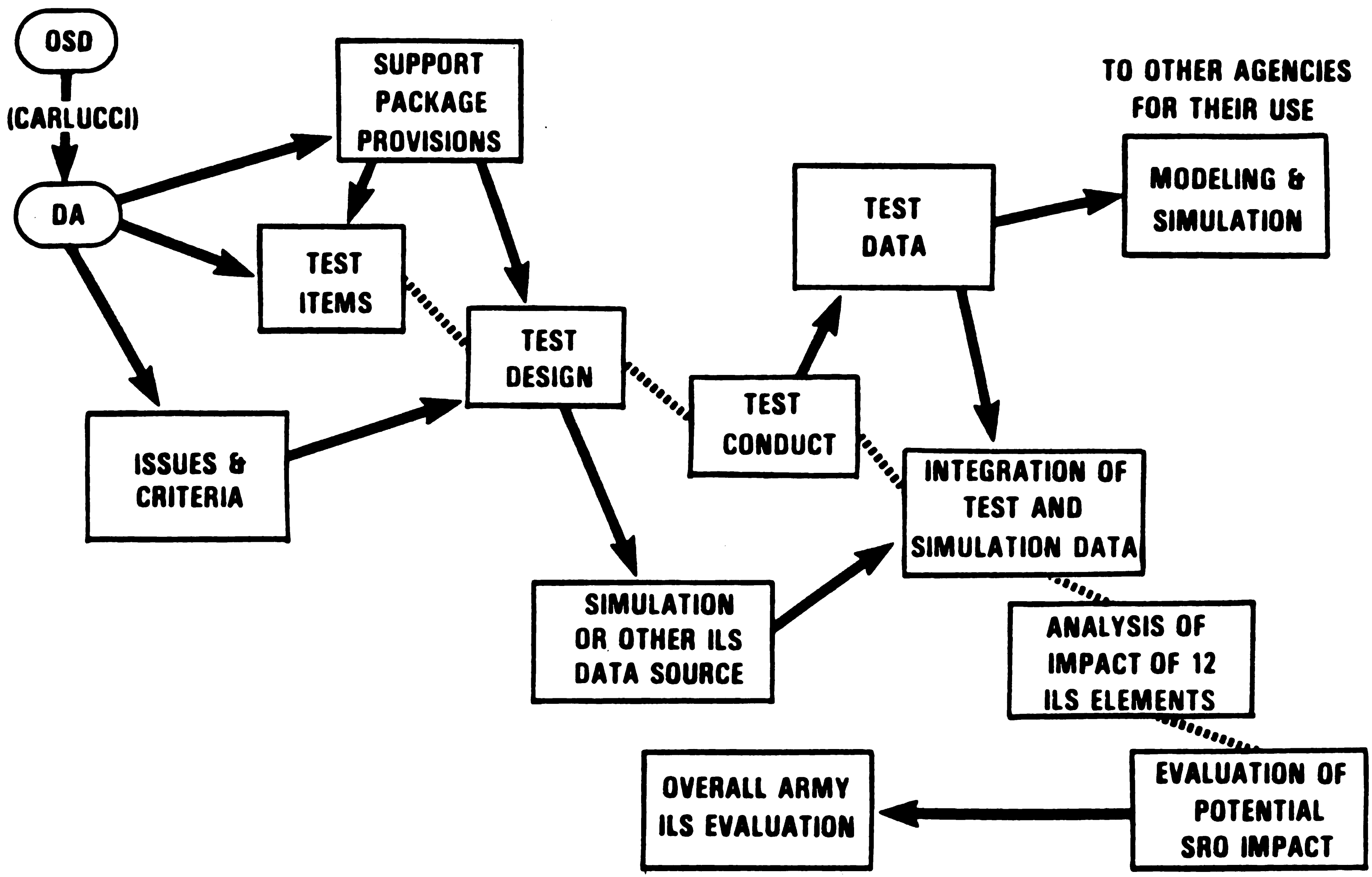
- USE LEVERAGE PROVIDED BY INCREASED IMPORTANCE OF READINESS TO IMPROVE THE SUPPORTABILITY TEST AND EVALUATION

IMPROVEMENTS

USE LEVERAGE TO OBTAIN:

- DEPUTY TEST DIRECTOR FOR LOGISTICS
- BETTER OMS/MP
- ADDITIONAL TESTING RESOURCES

APPROACH TO ILS TEST AND EVALUATION



BASIC INTENT

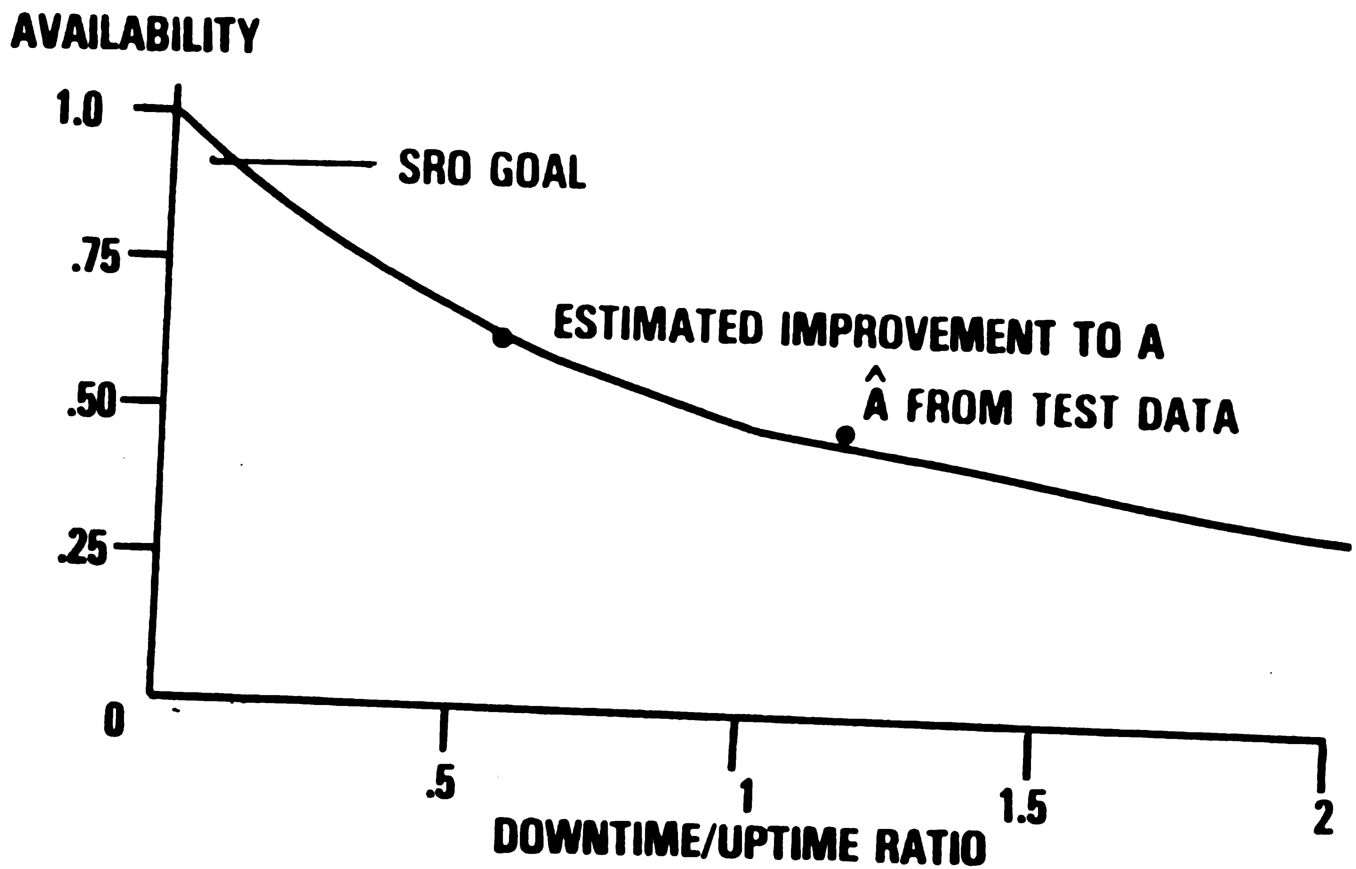
TO MEASURE THE PERCENTAGE OF DOWNTIME DUE TO EACH OF THE 12 ILS ELEMENTS

IMPACT ON SRO OF ILS ELEMENTS

ILS ELEMENT	DOWNTIME	PERCENT OF TOTAL TEST DOWNTIME	MAJOR CONTRIBUTOR
1. DESIGN INFLUENCE	8	2%	
2. MAINTENANCE PLANNING	12	3%	
3. MANPOWER & PERSONNEL	00	20%	X
4. SUPPLY SUPPORT	100	25%	X
⋮ ⋮	⋮	⋮	⋮
12. STANDARDIZATION/INTEROP	8	2	
	400	100%	

$$SRO = \frac{600}{600 + 400} = .60$$

IMPACT ON SRO



TEST AND EVALUATION OF THE TWELVE ILS ELEMENTS

- TWELVE ILS ELEMENTS WILL BE:
  - DESIGNED INTO OT
  - ANALYZED
  - LINKED TO THE SRO
- EACH ELEMENT WILL BE PRESENTED CONSIDERING:
  - PLAN: TEST DESIGN CONSIDERATIONS TO BE EXAMINED
  - TYPICAL MEASURES OF EFFECTIVENESS: EXAMPLES OF ANALYTIC METHODS
  - IMPACT ON SRO: INFLUENCE OF THE ILS ELEMENT ON DOWNTIME

INTEGRATED LOGISTIC SUPPORTABILITY TESTING IN OT

DOUGLAS McGOWEN

DESIGN INFLUENCE

- PLAN
  - MAINTENANCE AND SUPPLY DEMANDS
  - MAINTAINABILITY
  - HUMAN FACTORS
- TYPICAL MOE's
  - MEAN TIME BETWEEN UNSCHEDULED MAINTENANCE ACTIONS
  - MEAN TIME BETWEEN REMOVALS OF DIRECT EXCHANGE COMPONENTS
- IMPACT ON SRO
  - INFLUENCES CORRECTIVE MAINTENANCE TIME WHEN ADDITIONAL MAINTENANCE REQUIRED BECAUSE OF DIFFICULTY IN REMOVING COMPONENTS OR ACCESS TO COMPONENTS

DESIGN INFLUENCE (CONT)

- IMPACT ON SRO (CONT.)
  - INFLUENCES PREVENTIVE MAINTENANCE TIME BY INACCESSABLE LUBE POINTS, HARD TO REACH FILTERS, PM REQUIRED TOO FREQUENTLY
  - INFLUENCES ALDT BY LARGE DEMAND RATE ON SUPPLY ITEMS
  - INFLUENCES ALDT BECAUSE MAC, RPSTL DOES NOT PROVIDE SUFFICIENT TOOLS AT THE REQUIRED LEVEL, BII LIST DOES NOT CONTAIN APPROPRIATE ITEMS
  - INFLUENCES PREVENTIVE MAINTENANCE BECAUSE LUBE POINTS HARD TO REACH, SCHEDULED MAINTENANCE OCCURING TOO FREQUENTLY

MAINTENANCE PLANNING

- PLAN
  - PLANNED CONTRACTOR SUPPORT
  - OPERATIONAL READINESS FLOAT
  - USE OF CONTACT TEAMS
- TYPICAL MOEs
  - TURN AROUND TIME ON CONTRACTOR REPAIRED ITEMS
  - DEMAND SATISFACTION ON CONTRACTOR REPAIRED ITEMS
  - PERCENT TIME FLOAT ITEM AVAILABLE
- IMPACT ON SRO
  - INFLUENCES ALDT IF CONTRACTOR PART NOT AVAILABLE

MANPOWER AND PERSONNEL

- PLAN
  - MAINTENANCE ORGANIZATIONS/SUPPORT ORGANIZATIONS
  - NUMBER OF OPERATOR AND SUPPORT PERSONNEL BY MOS
  - PEAK MAINTENANCE WORKLOADS
- TYPICAL MOEs
  - MAINTENANCE RATIO AT EACH LEVEL
  - MEAN TIME TO REPAIR
  - MEAN TIME BETWEEN UNSCHEDULED MAINTENANCE ACTIONS
  - PERCENT REPAIR ACTIONS AT EACH LEVEL
  - PERCENT MAINTENANCE ACTIONS THAT COULD NOT BE COMPLETED BY MOS
- IMPACT ON SRO
  - INFLUENCES CORRECTIVE MAINTENANCE TIME IF PERSONNEL NOT APPORTIONED AT CORRECT LEVEL OR MOS
  - INFLUENCES PREVENTIVE MAINTENANCE TIME IF SCHEDULED MAINTENANCE OCCURS TOO FREQUENTLY OR IS EXCESSIVE IN LENGTH

SUPPLY SUPPORT

- PLAN
  - PLL/ASL WITH WARTIME STOCKAGE LEVELS
  - REORDER POINTS
  - ORDER-SHIP TIMES
  - LOCATION/MOBILITY
  - COMBAT ASL/ITEM ESSENTIALITY
  - BULK SUPPLIES INCLUDING POL
- TYPICAL MOE's
  - DEMAND SATISFACTION
  - MOBILITY INDEX
  - CONSUMPTION RATES
  - DELIVERY RATES
  - STORAGE CAPACITY
- IMPACT ON SRO
  - INFLUENCES ALDT WHEN STOCKOUTS OCCUR.
  - PLL/ASL NOT MOBILE

SUPPORT EQUIPMENT TO INCLUDE TMDE

- PLAN
  - EXAMINE EFFECT OF TOOLS, TMDE, ATE, CALIBRATION, BIT, BITE, PERSONNEL
  - REALISTIC SUPPORT EQUIPMENT SCENARIOS
  - INVESTIGATE PLANNED INCREASES IN BOIP
  - CONSIDERATION FOR ADDITIONAL MAINTENANCE SOFTWARE TESTING FOR TMDE, BIT AND BITE

(CONT)

- TYPICAL MOEs
  - PERCENT CONTRIBUTION OF DIAGNOSTIC TIME TO TOTAL MAINTENANCE TIME
  - FOUR PROBABILITIES OF BIT (AND TMDE) PERFORMANCE
    - PROBABILITY OF INDICATING GOOD WHEN ACTUALLY GOOD
    - PROBABILITY OF INDICATING GOOD WHEN ACTUALLY BAD
    - PROBABILITY OF INDICATING BAD WHEN ACTUALLY GOOD
    - PROBABILITY OF INDICATING BAD WHEN ACTUALLY BAD

SUPPORT EQUIPMENT TO INCLUDE TMDE (CONT)

- TYPICAL MOEs (CONT)
  - AUTOMATIC TEST EQUIPMENT
    - MEAN TIME TO INITIALIZE TPS
    - MEAN TIME TO TEST AND ISOLATE FAULT
    - MEAN TIME TO REPAIR FAULT
    - PROBABILITY OF RETEST OK
    - PROBABILITY OF NO DEFECT FOUND
- IMPACT ON SRO
  - BIT, TMDE, AND TOOLS INFLUENCE CORRECTIVE MAINTENANCE TIME
  - TOOLS INFLUENCE PREVENTIVE MAINTENANCE TIME
  - CALIBRATION, SUPPORT EQUIPMENT, AND ATE INFLUENCE ALDT

TRAINING AND TRAINING DEVICES

- PLAN
  - TRAINING PLANS
  - PRESELECTED MOS's
  - TRANSFER OF KNOWLEDGE, SKILLS AND TASKS
  - INSERTED MAINTENANCE ACTIONS
- TYPICAL MOEs
  - PERCENT CRITICAL MAINTENANCE TASKS SUCCESSFULLY COMPLETED (TEST)
  - PERCENT CRITICAL MAINTENANCE TASKS SUCCESSFULLY COMPLETED (INSERTED)
- IMPACT ON SRO
  - INFLUENCES CORRECTIVE MAINTENANCE TIME DUE TO EXCESSIVE REPAIR TIME

TECHNICAL DATA

- PLAN
  - TECHNICAL MANUALS
  - CALIBRATION INSTRUCTIONS
  - PACKAGING, HANDLING, STORAGE
  - FIELD MANUALS
- TYPICAL MOEs
  - PERCENT CORRECTIVE DIAGNOSTIC PROCEDURES IN MANUAL
  - PROBABILITY OF CORRECT CALIBRATION
- IMPACT ON SRO
  - INFLUENCES CORRECTIVE MAINTENANCE TIME BECAUSE OF LONGER DIAGNOSTIC TIMES

- INFLUENCES CORRECTIVE MAINTENANCE TIME BECAUSE OF LONGER DIAGNOSTIC TIMES

COMPUTER RESOURCES SUPPORT

- PLAN
  - LIMITED TO ILS CONSIDERATIONS
  - LIMITED TO SOFTWARE MEDIA IN SUPPORT SYSTEM
  - SCHEDULED MAINTENANCE
  - LOCATION OF SPARES AND MAINTAINERS
- TYPICAL MOEs
  - MEAN TIME BETWEEN MAINTENANCE FOR SOFTWARE MEDIA
  - PERCENT OF DOWNTIME DUE TO SOFTWARE MEDIA
- IMPACT ON SRO
  - INFLUENCES PREVENTIVE AND CORRECTIVE MAINTENANCE

PACKAGING, HANDLING, STORAGE

- PLAN
  - INITIAL INSPECTION
  - REPRESENTATION OF TACTICAL STORAGE AND SUPPLY
- TYPICAL MOEs
  - PERCENT OF REPAIR PARTS UNUSEABLE DUE TO DAMAGE
  - PERCENT OF DOWN TIME DUE TO DAMAGE
- IMPACT ON SRO
  - CORRECTIVE MAINTENANCE INFLUENCED BY DAMAGE
  - ALDT INFLUENCED BY STORAGE

TRANSPORTATION

TRANSPORTATION AND TRANSPORTABILITY

- PLAN
  - RESEARCH PLANNED TRANSPORTATION METHODS
  - INCLUDE ON-OFF LOADING, TOWING, AND FIELD RECOVERY
  - PROVISIONS FOR TRANSPORT OF SUPPORT EQUIPMENT/ACCESSORIES
  - CONVERSION FROM TRANSPORT TO OPERATION (SET-UP
  - CONVERSION FROM OPERATION TO TRANSPORT (TEARDOWN)
- TYPICAL MOEs
  - PERCENT OF DOWNTIME DUE TO DAMAGE INCURRED BY TRANSPORT
  - MEAN TIME TO SET-UP AND TEARDOWN
  - MEAN TIME TO PREPARE EQUIPMENT FOR AIRCRAFT/SLINGLOAD
- IMPACT ON SRO
  - CORRECTIVE MAINTENANCE INFLUENCED BY DAMAGE
  - ALDT INFLUENCED BY DELAYS/ADDITIONAL SPARES

FACILITIES

- PLAN
  - SIZE
  - LOCATION
  - TYPES—CALIBRATION/MAINTENANCE/SUPPLY
- TYPICAL MOEs
  - MAXIMUM VOLUME UTILIZED IN PEAK PERIOD
  - PERCENT DOWNTIME DUE TO LACK OF FACILITIES
- IMPACT ON SRO
  - ALDT INFLUENCED BY MAINTENANCE QUEUEING OR SUPPLY

STANDARDIZATION AND INTEROPERABILITY

- PLAN
  - LIMITED TO ILS CONSIDERATIONS
  - COMMONALITY OF POWER SOURCES, SUPPORT EQUIPMENT
  - INTERCHANGEABILITY
  - STANDARD TERMINOLOGY, PRACTICES
  - RESEARCH ALL POSSIBLE ILS INTERFACES
- TYPICAL MOEs
  - PERCENTAGE OF POWER SOURCES, SUPPORT EQUIPMENT NON-INTERCHANGEABLE
  - MEAN TIME TO INITIALIZE MAINTENANCE COMMUNICATIONS LINKS
- IMPACT ON SRO
  - ALDT INFLUENCED BY DELAYS DUE TO STANDARDIZATION/ INTEROPERABILITY

PLL/ASL AND ALDT

REQUIREMENTS TO SET UP THE PLL/ASL

- PLL/ASL LISTING
- WARTIME STOCKAGE LEVELS
- COMBAT ASL/ITEM ESSENTIALITY
- LOCATION/MOBILITY INFORMATION
- REORDER POINTS
- WEIGHT/VOLUME/SIZE OF PLL/ASL
- DIRECT EXCHANGE ITEMS
- SUFFICIENT ITEMS FOR TEST CONTINUITY

PLL/ASL AND ALDT

- CONSTRUCT TIME LINE WITH QUANTITY VS TIME
- INDICATE WAR TIME STOCKAGE LEVEL AND REORDER LEVEL
- PLOT STOCKAGE LEVEL AFTER EACH DEMAND
- MONITOR DEMAND DURING REORDER PERIOD
- PLOT STOCKAGE LEVEL AFTER ORDER-SHIP TIME
- INDICATE POINTS WHERE DEMAND OCCURRED WITH ZERO STOCK BALANCE

ESTIMATING ORDER-SHIP TIMES

- DETERMINE MISSION CRITICAL COMPONENTS
- SELECT SIMILAR SYSTEM OR COMPONENTS
- USE NSN AND APPROPRIATE CODES TO GET ORDER-SHIP TIMES (EUROPEAN, AVERAGE, CONUS) FROM LOGISTICS CONTROL ACTIVITY

- AND THEN—
  - EXAMINE THE EFFECTS OF VARIABILITY IN ALDT ON SRO

PLL/ASL AND ALDT

FILLS IN THE FOLLOWING DATA "GAPS"

- LESS THAN 100 PERCENT INITIAL STOCKAGE OF THE PLL/ASL
- ADMINISTRATIVE AND LOGISTICS DELAY TIME
- SEPARATION OF PLL/ASL FROM END ITEM

THIS METHOD ALLOWS:

- CONTRACTOR REPAIR ON SITE
- TEST CONTINUATION
- OFF SYSTEM REPAIR

ILS TEST AND EVALUATION INITIATIVES

- ACTIVE PARTICIPATION IN ILSMT
- IDENTIFICATION AND RESEARCH OF ILS BIBLIOGRAPHY
- NEED FOR DEVELOPMENT OF PEAK/INTENSE PERIODS IN OMS/MP
- IDENTIFICATION OF ADDITIONAL MAINTENANCE TESTING
- IDENTIFICATION AND DESIGN OF WORKAROUND FOR SSP INADEQUACIES OR TEST ARTIFICIALITIES
- ESTABLISHMENT OF DEPUTY TEST DIRECTOR FOR LOGISTICS
- FORMALIZATION OF PLL/ASL TEST DISCIPLINE
- EVALUATION METHODOLOGY MEASURING SUPPORTABILITY IMPACT ON SRO

SITUATION

SITUATION

- INCREASED EMPHASIS ON ILS IMPACTS THE TEST COMMUNITY BY REQUIRING ADDITIONAL MANPOWER TO:

  PARTICIPATE IN ILSMT

  ACCOMMODATE EXPANDED TEST DESIGN SCOPE

  MANAGE ILS TEST EXECUTION AND DATA COLLECTION

  IDENTIFY, UNDERSTAND, AND INTEGRATE SIMULATION DATA WITH TEST DATA

  ACCOMMODATE EXPANDED EVALUATION SCOPE

- PROBLEM

  THE SOURCE AND FUNDING FOR THIS INCREASE IN PERSONNEL HAS NOT BEEN IDENTIFIED OR APPROVED

[illegible]

Susa
rmy
[illegible]
Lea

Ler
rmy T
-MA)
Good

ane
rmy L
-MT)
Lee,

bert E
rmy TR
-TEM)
Sand

Mason
rmy Tra
-P)
Monroe

ABSTRA
ortabil
the ad
uage dev
compre
ity duri

Present
ation o
rehensiv
cient,
perspe

Followi
ation of
essed wit

ON ARMY OPERATIONAL TEST AND EVALUATION

OF LOGISTICS SUPPORTABILITY

Donald G. Reich
Assistant Deputy Chief of Staff
for Test and Evaluation
US Army Training and Doctrine Command
Fort Monroe, Virginia

ACKNOWLEDGEMENT: The paper presented herein was a TRADOC team effort including participation of the following individuals:

Ms. Susan Wright
US Army Combined Arms Center
(ATZL-TIE)
Fort Leavenworth, KS

Major Lennart G. Krals
US Army TRADOC Combined Arms Test Activity
(ATCT-MA)
Fort Hood, Texas

Mr. Gene Harvey
US Army Logistics Center
(ATCL-ME)
Fort Lee, VA

Mr. Bert E. Wing
US Army TRADOC Systems Analysis Activity
(ATAA-TEM)
White Sands, NM

Mr. Mason E. McClanahan
US Army Training and Doctrine Command
(ATTE-P)
Fort Monroe, VA

ABSTRACT: Army policy requires testing and evaluation of the logistics supportability of a system in the acquisition process. The thrust is to assess both the adequacy of system design for support and the adequacy of the support package developed with the system. While the policy is generally sound, the lack of a comprehensive Army methodology for test and evaluation of logistics supportability during operational testing has weakened implementation of the policy.

Presented first is a review of the current status of operational test and evaluation of logistics supportability. The key ingredients to effective and comprehensive logistics supportability test and evaluation are identified. For each ingredient, problem areas with proposed solutions are discussed, thus providing a TRADOC perspective of where we are and where we should be.

Following the review is TRADOC proposed methodology for operational testing and evaluation of logistics supportability. All phases of operational testing are addressed with incorporation of proposed solutions for problem areas.

LOGISTICS SUPPORTABILITY (LOGS)
OPERATIONAL TEST AND EVALUATION (OT&E)
METHODOLOGY

1. PURPOSE. This paper provides a discussion of key ingredients, current problem areas and proposed solutions for accomplishment of effective LOGS test and evaluation during operational testing (OT). For the purposes of this paper, LOGS is defined as:

Logistics Supportability (LOGS). The characteristics of the system (materiel and crew) and the related support elements (support concept, support materiel, and support personnel) as they contribute to the retention and restoration of the materiel system in an operational effective status. Therefore, LOGS is the way these three elements affect and are affected by the materiel system.

2. Comprehensive effective OT&E of the supportability of a system can be divided into three categories: the decision process, analysis and resources. Within each of these categories, there are key ingredients to an effective logistics evaluation. Below is a discussion which provides a list of dilemmas with proposed solutions for each key ingredient. Proposed LOGS operational test and evaluation methodology providing specific guidance is at Annex A. It is assumed that waivers will not be granted to any of the key ingredients. When a waiver is granted, it becomes a dilemma in itself.

a. Decision process.

(1) Recognition of LOGS as a mission oriented issue.

DILEMMA - Decision makers often perceive operational effectiveness and combat power issues as having higher visibility and being the most mission oriented, therefore, are highest priority. Thus, in the battle for resources, LOGS issues are often sacrificed, since supportability issues which require greater time and manpower than effectiveness or combat power.

SOLUTION - All LOGS issues must be critically reviewed by the decision maker at each decision point with data provided showing LOGS inpact on combat power.

(2) Acquisition schedule to accommodate effective LOGS OT&E.

DILEMMA - The acquisition cycle is decision milestone driven. Therefore with OT following developmental tests (DT), as slips occur in DT, the time available for OT decreases. Thus effective LOGS OT&E is compromised.

SOLUTION - Existing policies need to be enforced. Additional policies needed are as follows:

(a) When slip occurs in critical program milestones (e.g. DT start or complete), subsequent milestones will be slipped accordingly.

(b) When critical LOGS elements (e.g., system peculiar Test Measurement Diagnostic Equipment (TMDE) and generators, DS and GS support package, etc.) will not be tested in OT II of the system, OT II testing of these support items will be accomplished prior to their type classification. This currently happens for training devices since they have a separate requirement document. However, other items encompassed by the system requirement document do not undergo this necessary testing.

(3) Test and logistics test resources waivers held to a minimum.

DILEMMA - Current waiver procedures are confusing. Policy (AR 71-3, AR 700-127, etc.) provides a myriad of waiver approval procedures and authorities dependent on resource, test, and system category of concern and situation. Procedures and guidelines are not clear or consistent for all situations.

SOLUTION - Have only one procedure for waiving tests and test resources with well defined procedures and guidelines.

b. Analysis.

(1) LOGS OT&E methodology understood and consistently applied across the Army community.

DILEMMA - There is no DA published baseline methodology on LOGS OT&E to guide and harmonize the community. Recently drafted DA pamphlet on supportability test and evaluation provides broad OT guidance but refers to AR 71-3 and DA Pam 71-3 for specifics. These specifics have not been incorporated. TRADOC developed and implemented a detailed methodology in coordination with all the Army community in 1977. However, this methodology has not been since updated or sanctioned for all to use by DA.

SOLUTION - A methodology be developed and incorporated in a DA pamphlet.

(2) Critical supportability issues and criteria clearly and appropriately defined.

DILEMMA - Issues and criteria have not effectively encompassed operational readiness of system and logistics burden emplaced by system.

SOLUTIONS:

(a) Set three critical LOGS issues, (1) operational readiness, (2) logistics burden, and (3) *system support package deficiencies.

*System support package (SSP) - A composite package of support elements in initial issue quantities for a materiel system in the operational (deployed) environment. For OT it is a combination of test support packages provided by materiel, combat and training developers. At OT II content is scaled commensurate with the force slice being played to evaluate organizational, direct support and general support capabilities.

(b) Insure requirements documents provide a sound basis for defining critical criteria.

(3) Appropriate evaluation techniques defined and available.

DILEMMA - Current methodologies in TRADOC or DARCOM have not provided analytical techniques for establishing logistics burden criteria and evaluating operational readiness or logistics burden achievement.

SOLUTIONS:

(a) Provide modeling and analysis routines for developing criteria for logistics burden. Recently published TRADOC/DARCOM Pam 71-11 provides techniques for developing system readiness objective (SRO) requirements which also apply to criteria.

(b) Provide modeling and analytical routines for evaluating achievement of readiness measures and logistics burden.

(c) Computerize modeling and analytical routines and make available to all OT issue-criteria developers, testers and evaluators.

c. Resources.

(1) Operator, crew, maintenance and supply personnel in the appropriate specialty and trained as planned when fielded.

DILEMMA - Same as addressed for decision process key ingredients.

SOLUTION - Same as addressed for decision process key ingredients. Conduct of OT I is critical to determining personnel skill and training requirements for development during following phase.

(2) System and support gear (hardware and *software) suitably mature for test.

DILEMMA - Same as addressed for decision process key ingredients.

SOLUTION - Same as addressed for decision process key ingredients. Conduct of OT I is critical to identifying materiel system and support gear deficiencies for improvement during next phase so that sufficiently mature items are provided for OT II.

(3) Sufficient slice of the force to place proper logistics demands on the support structure.

DILEMMA - OT test players and data collectors are provided from active FORSCOM units which have other missions thus constraining their availability for test.

SOLUTIONS:

(a) The evaluator must utilize to the maximum other data sources.

(b) Work with DT testers to incorporate representative trained troops in their test as maintenance personnel or perhaps have DT off system repairables shipped to the OT site for maintenance.

(c) Work with materiel developers to have representative trained troops used in their logistics demonstrations and PTEAR.

(d) Require both OT II and FOE be performed on a routine basis and combine data sets to degree possible for a more comprehensive evaluation prior to fielding.

*Support software - Includes entire set of programs, procedures, and related documentation such as technical manuals and computer programs necessary for supply, maintenance and training.

(e) Evaluation models addressed above must provide for expansion beyond the unit and support structure sizes tested to provide a representative slice of force structure.

(f) Consider combining several small scale OT occurring in same year into a single larger scale test.

(g) Explore use of DA sample data collection program, National Training Center (NTC) and field exercies (e.g., REFORGER) to expand supportability data base during initial fielding.

(4) Sufficient test time to properly exercise the critical support elements.

DILEMMA - While our test guidelines use RAM as a standard for determining the test length, there is no assurance that all critical LOGS elements will be functioned, including both *primary and **secondary logistics systems.

SOLUTIONS:

(a) Have simulated maintenance during OT such that a suitable percentage of critical tasks are accomplished at each maintenance level through GS.

(b) Conduct sample data collection programs on a routine basis with initial fielding and provide results to OT evaluators and testers.

*Primary support system - Personnel, maintenance and supply system which responds directly to the materiel system under consideration.

**Secondary support system - Maintenance and supply system which responds to the primary support system.

(c) Same as (f) for key ingredient concerning force slice.

(5) Adequately defined and implemented tactical scenario and support concept for test.

DILEMMA - Test support packages provided by materiel, combat and training developers for OT are often incomplete or do not properly define scenario and support concepts.

SOLUTIONS:

(a) Improved quality control of support packages.

(b) Improved waiver approval procedures as described above for the decision process key element.

(c) Readiness for test reviews conducted and results brought to decision makers attention.

(6) Effective data collection system to satisfy issues and criteria.

DILEMMA - Data collection system is fragmented due to lack of standardization, coordination, and communication across the Army OT community resulting in duplicated efforts.

SOLUTIONS:

(a) Use all data sources.

(b) Simulated maintenance actions during OT.

(c) Develop a standard set of programs for OT to assess repair parts and POL consumption, stock number analysis of tools, TMDE, and special equipment, assess off-line recoverable repairables and training, and validating and verifying manuals, and maintenance allocation charts. ADP resources should be used to the maximum degree for this methodology.

(d) Obtain from Soldier Support Center the profile of the qualifications for MOS test player personnel. This is very critical when a small personnel sample is being used since it is possible to obtain people not in the profile but having the MOS.

(e) Data collectors should be carefully chosen and trained with specific skills and therefore, they would not solely perform the function of data collection but be a system evaluator.

(7) Personnel adept and qualified in logistics supportability on both the test and evaluation teams as well as in those agencies responsible for defining the issues and criteria.

DILEMMA - Probably the most serious deficiency in LOGS testing and evaluation is the lack of trained logistics testers and evaluators. Not only are individuals assigned to a testing or evaluation agency without training in techniques of testing or evaluation, but they are not trained in the specific area of logistics supportability.

SOLUTION - Match SC personnel with the system to be tested and train them in LOGS management techniques, LOGS quantitative techniques, and logistics support analysis. These courses are readily available and should be mandatory for all logistics testers and evaluators prior to assignment. Likewise, civilian logistics test and evaluation personnel need similar training.

3. CONCLUSION.

In conclusion, the current regulation guidelines provide a system for addressing LOGS issues; however, total adequacy of LOGS evaluation is resource dependent and trade-offs generally reduce the credibility of operational test results for logistics support, and unless these problems are solved, the perception of inadequate LOGS testing and evaluation will persist.

ANNEX A

TEST AND EVALUATION
LOGISTICS SUPPORTABILITY (LOGS) IN OT

1. INTRODUCTION. The test and evaluation of logistics supportability is addressed in numerous test related documents within the US Army. One of the dilemmas main addressed in the paper is that there is no DA operational test and evaluation procedure documented and/or implemented to assure that logistics is comprehensively evaluated. The methodology proposed herein to fill this Dept of the Army void is a refinement of the methodology developed by and presently being used within TRADOC. It is assumed that the remaining dilemmas discussed in the main paper have been resolved.

2. DEFINITIONS. For the purposes of this methodology, the following definitions apply:

a. Logistics Supportability (LOGS). The characteristics of the system (materiel and crew) and the related support elements (support concept, support materiel, and support personnel) as they contribute to the retention and restoration of the materiel system in an operational effective status. Therefore, LOGS is the way the three elements, support concept, support materiel and support personnel, affect and are affected by the materiel system.

b. System Support Package (SSP). A composite package of support elements in initial issue quantities planned for a materiel system in the operational (deployed) environment. For OT it is a combination of test support packages provided by materiel, combat and training developers. In its preliminary form, it is provided before and evaluated during developmental and operational testing and evaluation to validate the organizational, direct support, and general support maintenance capabilities. For logistic supportability testing, it normally includes:

(1) Support and test equipment.

(2) Trained personnels (including the training programs, materials, devices, and ammunition needed to develop those skills).

(3) Supply support.

(4) Technical logistic data.

(5) Facilities.

(6) Computer resources.

(7) Maintenance support.

(8) The logistics concept.

c. Primary Support System. Personnel, maintenance and supply system which responds directly to the materiel system.

d. Secondary Support System. Maintenance and supply system which responds to the primary support system.

e. Logistics Support Concept. The overall "how" the logistics system is set up and administered to support the materiel system. The support concepts provide the organizational structure and responsibilities for accomplishing the maintenance and supply functions at each level. This includes the identification and allocation of hardware, software and support personnel to each supply and maintenance level.

f. Logistics Support Materiel. Those hardware and software items needed for supply, maintenance and training support. Logistics support hardware includes, test measurement and diagnostic equipment, special and common tools, repair parts, resupply and rearm vehicles, training devices and similar equipment assigned to supply, maintenance and training units. Logistics support software includes the entire set of programs, procedures, and related documentation such as technical manuals, lubrication orders, computer programs, etc., necessary for maintenance, supply and training.

g. Logistics Support Personnel. Selection criteria and training required for operator, crew, maintenance and supply personnel. Personnel selection entails definition of duty requirements and the skills and characteristics needed to retain the system in or when failure occurs, restore it to an operationally effective condition. The amount and type of training is a function of the system complexity and the designated military occupation specialty (MOS).

h. Materiel System. Mission item being acquired for which LOGS is to be evaluated. Characteristics include all factors of design which affect logistics support. Examples are, design for maintainability, human factors and safety affect the efficiency and speed of maintenance operations. Hardware requirements for special handling, training devices, frequency of calibration, TMDE, transportability, resupply equipment affect end item logistics. Design for standardization within the Army and rationalization, standardization and interoperability (RSI) with other services and NATO is also considered.

3. EVALUATION PROCESS. To assure that a materiel system in the acquisition cycle can be fully supported when fielded, logistics supportability assessments should be conducted during every phase of the acquisition cycle. Geniune logistics supportability assessments do not just happen as a result of operational testing. Evaluation planning is key to timely and effective operational LOGS assessment as with any other system performance parameter. The first step in the evaluation process is the identification of issues with associated criteria which must be addressed by the decisioin milestone of the

acquisition cycle. The combat developer or training developer provides the issues and criteria to the evaluator. Subsequently, the evaluator must identify, for a given program, the evaluation approach or methodology, the analysis to be performed, the data required and the sources of the data. This evaluation planning information is documented in the independent evaluation plan (IEP). Required operational tests (OT) are performed by the tester. Once the required data becomes available, the evaluator performs the prescribed analyses to evaluate each issue and develop his overall assessment of the system and testing done to date. Hence, evaluation of LOGS involves the following steps:

a. Identification of critical issues and criteria.

b. Evaluation planning.

c. Perform studies/conduct testing/collect data.

d. Evaluation.

4. ISSUES AND CRITERIA. The first step in effective LOGS evaluation is identification of issues and criteria which must be addressed and data provided for an adequate resolution by time of the decision milestone. There are three critical supportabilitiy issues to which decision makers need answers before making the production decision as follows:

a. Does the system, when supported in accordance with the approved logistics concept, achieve required operational readiness?

This issue examines both the design of the system for effective support and the ability of the support system to rapidly respond to the system need for maintenance. Operational availability or other system readiness objective (SRO) measure will be the principal criteria for this issue.

TRADOC/DARCOM Pam 71-11 provides guidance on development of these values and also may be used for guidance in evaluating the SRO.

b. Does the system impose excessive burden on any of its support elements?

This issue examines demands placed on critical support elements and resources required and available to meet these needs. Unit quantity rather than single system support burden is of dominant concern. In those cases where the SRO (issue "a" above) is not achieved, two levels of evaluation will be made. One is the resources required for the achieved SRO. The other is estimated additional resources necessary to achieve the required SRO. Criteria will include measures such as: 2 MOS XX man-years at organizational level and 2000 gal MOGAS per day.

c. Are there any deficiencies in the system support package?

This issue examines the completeness, appropriateness, accuracy, and adequacy of logistic elements in the test support packages provided for OT. This issues is subdivided into four key areas which encompass the AR 700-127 Integrated Logistics Support elements of concern. These four areas are logistics concept, support materiel, support personnel and materiel system characteristics. Criteria for this issue are generally subjective requiring judgment or direct observation by testers. For example, maintenance tasks and resources allocated at proper level.

d. Issue and criteria dendritic. A complete dendritic of the three issues is provided at Appendix A.

5. EVALUATION PLANNING.

a. Materiel Acquisition Phases. To assure that a materiel system in the acquisition process can be effectively supported when fielded, supportability assessments occur during every phase of the acquisition process. However, OT&E evaluations occur in three phases with emphasis as follows:

(1) Demonstration and Validation Phase. DT I and OT I are conducted during this phase using advanced development, breadboard or brassboard prototypes. Generally, contractor or developer taining is provided and concentration is on the developmental system. Early doctrine, organization and logistic concepts are available. Generally, there is competition between alternative system concepts to continue to next phase. The evaluation focus is on:

(a) Analyses of supportability merits of the competing system.

(b) Identify system and support concept modification needed.

(c) Identify special requirements for personnel, TMDE, and training to support development in the next phase.

(d) Refine OT I issues and criteria and identify any new ones for OT II.

(2) Full Scale Engineering Development Phase. This is the phase leading to full production decision with that rare exception where Low Rate Initial Production (LRIP) is authorized. DT II and OT II tests occur. Answers to the three critical issues (readiness, burden and system support package) are needed.

(3) Production and Deployment Phase. Evaluation conducted in this phase concentrates on answering unanswered or unresolved issues and verification of support deficiencies from the previous phase.

b. Analytical methods. For each phase of the materiel acquisition process, once the LOGS issues and criteria have been finalized, the next step is to establish the evaluation scheme to ascertain the analyses techniques to be utilized and and identify the data requirements and sources.

(1) Operational Readiness. Readiness is a function of system utilization, maintenance requirements and administrative and logistics downtime. Normally, the readiness parameter for Army systems will be operational availability. The equation for operational availability is as follows:

$$Ao = \frac{OT + ST}{OT + ST + TCM + TPM + TALDT}$$

Where:

OT = operating time during a given calendar time period.

ST = Standby time (not operating, but assumed operable) during that period.

TCM = Total corrective maintenance downtime in clock hours during that period.

TPM = Total preventive maintenance downtime in clock hours during the OT period.

TALDT = Total administrative and logistics downtime spent waiting for parts, maintenance personnel, or transportation during the time stated period. Furthermore, TALDT is a function of (1) operational mission reliability, (2) percentage of operational mission failures requiring parts, (3) probability of required part being on the Prescribed Load List and Authorized Stockage List, (4) probability of required parts being in stock and (5) delay times encountered at various levels of maintenance. Predicted values of the above parameters are used to develop the Ao value. Once the system is sufficiently matured and the logistics support developed, actual system data may replace predictions to give a better quantitative estimate of the equipment and the support system to achieve the SRO. Although Ao cannot be measured directly in an operational test, testing can generate data on the maintenance requirement, operational mission reliability, percentage of operational mission failures requiring parts, and the probability of required parts being on the stockage lists at various levels of maintenance. Substituting this data and utilizing the remaining data used in developing the requirements, one could develop a more valid estimate of the SRO. Sensitivity analysis could indicate needed improvements such as the pay off for improving logistics support characteristics of the system or refining the PLL or ASL.

(2) Logistics burden. The purpose of these analyses is to determine the strengths and weaknesses of planned support. Basically, maintenance, supply and transportation demands placed on the support system are compared against the the resources provided in the planned support system. Consider the examples at Appendix B as follows

(a) In the manpower analysis example, although the reliability requirement is nearly met, the reliability degradation translates into approximately three times as many failures as projected. Accordingly, if no reliability improvement is obtained, an additional 43 maintenance personnel will be required to support the system.

(b) In the fuel consumption analysis, meeting half the criteria for fuel consumption translates into double the requirement tankers, bladders and manpower supporing at the resupply point than programed for the system. Without redesign or adding these needed assets, a fifty percent reduction in operational readiness (i.e., operational availability) will result.

(c) The TMDE analysis demonstrates how the difference between force slice tested and to be supported in the field impacts burden. While no problem with queing at the TMDE device was observed in test, considering the force slice to be fielded shows a queing problem will result. The analysis also suggests how consideration of combat losses can impact the analysis conclusions. While the first analysis concludes operational readiness (i.e., operational availability) will be significantly degraded because of the queing problem, considration of combat losses may result in determining the allocation of TMDE adequate. Expected systems lost in the first three days of battle could remove the queing problem.

(d) At present there is no documentation or regulatory guidance requiring a specific logistics analysis to be performed. It would also be beneficial to document in future DA pamphlets several examples of typical maintenance, supply and transportation burden analyses. However, it should be realized that there is no "cookbook" method; evaluations and analysis techniques must be tailored to the specific system. Analytical techniques should be kept simple and provide timely response. Modeling such as Maintenance and Logistics Analysis (MALA) analysis is too complex and not responsive.

(3) Support Package Deficiencies. Test support packages provided by the materiel, combat and training developers define the support structure and its operation and provide the personnel, materiel and software upon which it functions. Errors and inadequacies are generally found in these elements during conduct of test. Subjective analysis is then made of these errors and inadequacies to determine their level of severity with regard to functions of the support system. Further, because of acquisition program constraints or strategy, elements may not be available at time of test. These must be identified and assessed as to impact on other areas. These deficiencies may or may not have impact on operational readiness or logistics burden analysis

above. For example, unnecessary tools would not affect either, where as, inability to transport in C130 or C141 aircraft affects burden due to commitment status of C5A aircraft. The analysis should result in categorization of the errors and inadequacies as follows:

(a) Significant deficiency - makes system unacceptable for deployment or correction involves more than the most routine engineering. Verification of correction needed prior to full production decision.

(b) Other deficiency - impacts system supportability but does not constitute a significant deficiency. Verification of correction required before or at time of initial fielding.

(c) Shortcoming - doesn't significantly impact system supportability but correction should be made if possible. Verification of correction not required.

c. Data Sources. Once the issues and criteria have been defined together with analysis techniques for their evaluation, the next step is to identify data sources from which appropriate data can be obtained to support the evaluation planned. Potential sources include but are not limited to the following:

(1) OT will be the dominant data source for the independent operational evaluation of supportability. OT is conducted with representative user operators, crews and units in as realistic an environment as possible. Operations are tactical scenario driven. The support system (personnel, equipment, software, procedures and organization) is as close as possible to that for the system when fielded. However, because of limitations, other sources should be investigated to determine if they provide a realistic source for data not achieveable in the OT.

(2) Consideration should be given to the possibility of combining several

OT scheduled to occur in the same year into a single large scale test. This may allow for a longer test with less total impact on FORSCOM resources. The larger force slice for test would provide a more realistic evaluation of the total support structure when fielding occurs.

(3) DT, while a technically oriented test, can provide data of benefit to the OT evaluation. Some areas include transportability analysis, TMDE and calibration equipment functional accuracy and reliability, RSI compatibility, component interchangeability, and technical accuracy of documentation. Data on hardware failure frequency which places demands on the support system can also be provided. However, because of differences in technical and operational environments, demands may be significantly different from the OT. It may be possible to bring in trained representative troops to perform maintenance tasks in DT which because of test duration are not expected to occur in OT. Likewise a possibility exists to have DT ship off system repairables to OT for repair. Thus, DT can fill an OT data void on personnel selection, training, manual and support equipment adequacy, human factors and system maintainability for those tasks.

(4) FDTE are user tests conducted to address issues concerning doctrine, organization and training. Some are conducted on systems during the acquisition process. As such, these may be appropriate data sources for issues concerning the logistics concept, personnel selection or training. When conducted, these may expand the OT data base.

(5) Logistics Demonstration (LD) and Preliminary Teardown Analysis (PTEAR) conducted by the materiel developer may also be a valid source similar to DT. The LD is a special experiment to address technical logistics issues not satisfied by other tests. The PTEAR provides data on manual accuracy by actually going through the procedures and performing the various maintenance

tasks on the system.

(6) Skill Performance Aids (SPA) verification is conducted by the training developer on those systems requiring SPAS manuals. Readability is determined by using a statistically valid sample of the representative MOS soldiers to use the manual. For other manuals, desk audits are performed by the combat and training developers. Conducted before OT II, these efforts serve to assure manuals received are more mature and representative of that to be fielded. Conducted after the test, they may serve to verify correction of manual deficiences found in test.

(7) Sample Data Collection Programs are conducted by DARCOM commands responsible for readiness. As new systems are fielded, data collection teams monitor the system and collect specific data as defined in a data collection plan. These programs vary in length from six months to in excess of one year. These are good data sources for further expansion of the LOGS data base. Opportunity to submit data elements and other involved in this effort is available through DARCOM.

(8) The National Training Center (NTC) may be a valid data source for supportability evaluation during early fielding. Data collected from this source would provide for evaluation of supportability under varied tactical scenarios. It should be noted that NTC is not to be a data collection agency. However, it may be possible to obtain supportability dat they routinely collect or to find work around solutions through coordination with NTC staff.

(9) Field exercises such as REFORGER may also provide valid source for supportability data during initial fielding. This source should be explored to determine ability to provide suitable supportability data either through their routine data collection or special collection effort such as could be funded by an FDTE.

6. OPERATIONAL TEST (OT).

a. General.

(1) OT provides the mechanism where all the key supportability elements are brought together with the materiel system under conditions of employment most representative of that expected when fielded. These user tests are conducted during specific phase of the materiel acquisition process for systems to support scheduled decision reviews. Tests are scheduled to occur after sufficient developmental testing (DT) is done to demonstrate required technical maturity of the system and its support elements. As such, OT provides the "proof of the pudding" for operational effectiveness and suitability (including supportability) of the system. For the purposes of this methodology, follow-on evaluation (FOE) is considered to be an operational test, although not so named.

(2) Testing LOGS can only be accomplished to the degree that the various elements reflect the anticipated application in the field. In this respect, the maturity of test support packages and the test prototype will determine the extent to which the logistics elements can be implemented/exercised and valid data generated. The test support packages define the logistics concept for supporting materiel. All elements (hardware, software and personnel) should be exercised in a realistic environment as possible to include implementation of the support concepts as defined by the logistics concept to obtain logistics supportability data.

(3) Typically, there are three phases during which operational testing can occur to validate the supportability of the system in the acquisition

cycle. These are Demonstration and Validation Phase (OT I), Full Scale Engineering Development Phase (OT II) and Production and Deployment Phase (FOE). The latter two phases (OT II and FOE) are when system and support maturity are such that comprehensive and effective LOGS testing should be achieveable. However, the first phase (OT I) is vitally important to aiding the materiel, combat and training developers in delivering this mature system and support in the latter stages. It is here that the system and support concept is first introduced to the environment (operational procedures, people, associated equipment, organization and battlefield conditions) in which it must operate and be supported.

b. Testing Guidelines.

(1) OT I. The thrust of OT I testing is to provide data upon which a mature system and support elements can be developed and provided in the following phase. Changes in the logistics system and materiel design are most cost effective when identified early. As with all testing, the OT I responds to those issues and criteria assigned by the evaluator in determining the data needs and sources. Guidelines for testing in this phase are as follows:

(a) Number of Systems - Since they will be immature (i.e., breadboard, brassboard, advanced development prototype), quantity is not generally critical. However, consideration must be given to doctrine and organizational concepts. For many systems, one or two prototypes will be sufficient, while others may require 5 to 10 in order to establish interface (e.g., communications systems).

(b) Employment - System should be introduced into assigned unit and operated in accordance with preliminary doctrine, organization, mission profile, and tactical scenario concepts.

(c) Length of Test - Based on tentative RAM requirements should provide sufficient iterations to validate mission profile, and failure definition/scoring criteria and operation by a sufficient sample of operators and crews (i.e., "as much as possible".

(d) Logistic Concept - Preliminary concept should be implemented.

(e) Logistics Support Materiel - Draft manuals, tools and test equipment should be provided at organizational level of maintenance. As much support materiel as possible should be provided to address DS support (especially for critical DS maintenance tasks). GS tasks should be observed for their complexity and equipment requirements.

(f) Logistics Support Personnel - User personnel of the planned MOS will be used for operator and organizational maintenance (DS level if possible). Training will generally be contractor or developer provided.

(g) Readiness to Test Reviews - The OT test agency reviews the SSP elements, operational test readiness statements (OTRS) and safety release prior to start of test. When deficiencies are found such that critical issues (including LOGS issues) cannot be addressed, decision makers are informed with recommended course of action.

(h) Obtain from Soldier Support Center, the profile of the qualifications for MOS test player personnel.

(i) Specific skill requirements should be identified to agencies suppling data collectors. These skills along with training should provide an individual who doubles as data collector and on spot logistics evaluator.

(2) OT II. This is normally the final test prior to the full production decision, therefore, comprehensive testing of the logistics system should be accomplished. Guidelines are:

(a) Number of Systems - At least three (3). Doctrine and organization may demand more for employment purposes.

(b) Employment - Tactical scenario oriented employing system in accordance with approved doctrine, organization and mission profile. Generally includes force-on-force combat operations.

(c) Length of Test - Based primarily on RAM and scheduled maintenance requirements as follows:

1. A minimum of three test items will each accumulate test time equal to at least 1.5 times the minimum acceptable value (MAV) for reliability and operate past the scheduled organizational, DS and GS maintenance points.

2. Total test time will be sufficient for statistical decision risk levels specified in the IEP.

3. Perform simulated maintenance actions as needed to accomplish 100% of organizational tasks, 60-75% of DS and 40-60% of GS when combined with tasks required in test.

*(d) Logistics Concept - Fully defined and implemented through GS level of supply and maintenance.

*(e) Logistics Support Materiel - All logistics support hardware and software should be available and utilized.

*(f) Logistics Support Personnel - All operator,maintenance and supply MOS personnel selected and trained in accordance with the TRADOC approved training program.

(g) Materiel System Characteristics - Prototype of sufficient maturity that characteristics that impact logistics represent design to be fielded.

(h) On system and off system replace and repair data will be collected.

(i) Readiness to Trest Reviews - The OT test agency reviews the SSP elements, operational test readiness statements (OTRS) and safety release prior to start of test. When deficiencies are found such that critical issues (including LOGS issues) cannot be addressed, decision makers are informed with recommended course of action.

(j) Obtain from Soldier Support Center, the profile of the qualifications for MOS test player personnel.

(k) Specific skill requirements should be identified to agencies suppling data collectors. These sills along with training should provide an individual who doubles as data collector and on spot logistics evaluator.

*These areas should be the same as that planned for initial fielding except preculiar spare parts actually available may be reduced below stockage levels as long as developer is in a position to timely resupply to keep test on schedule. Example, if contractor support at DS and GS levels is planned for first three years fielded, then the test should include DS and GS by contractor and not the Army standard system. Likewise, if the standard Army system is to be used when first fielded, that is the system to be employed in the test.

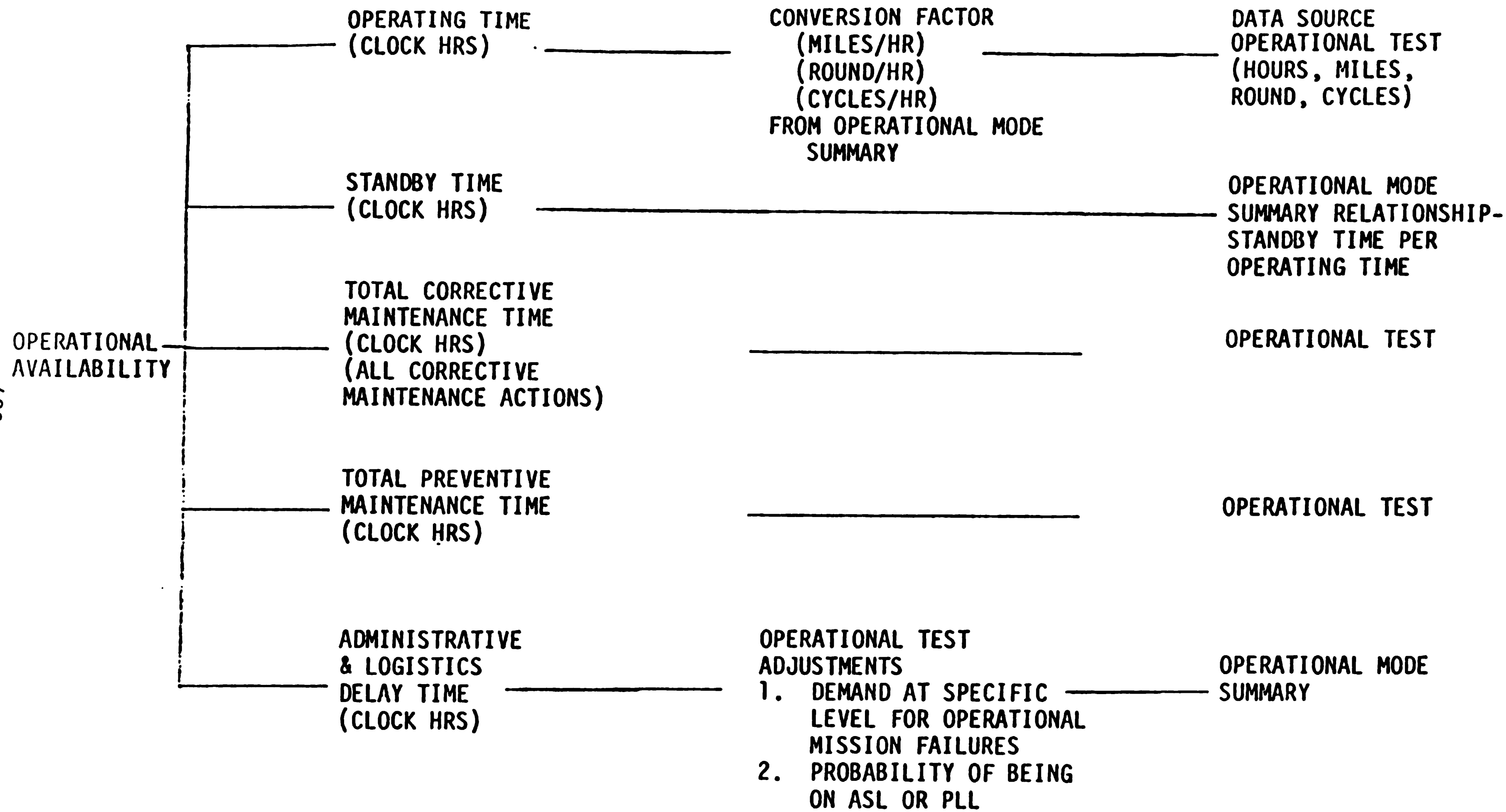
(3) OT IIA, OT III and FOE. Theses tests answers those issues and criteria not addressed or unresolved during and verifies correction of deficiencies found in OT II. Therefore, the guidelines are the same as for OT II when LOGS critical issues apply.

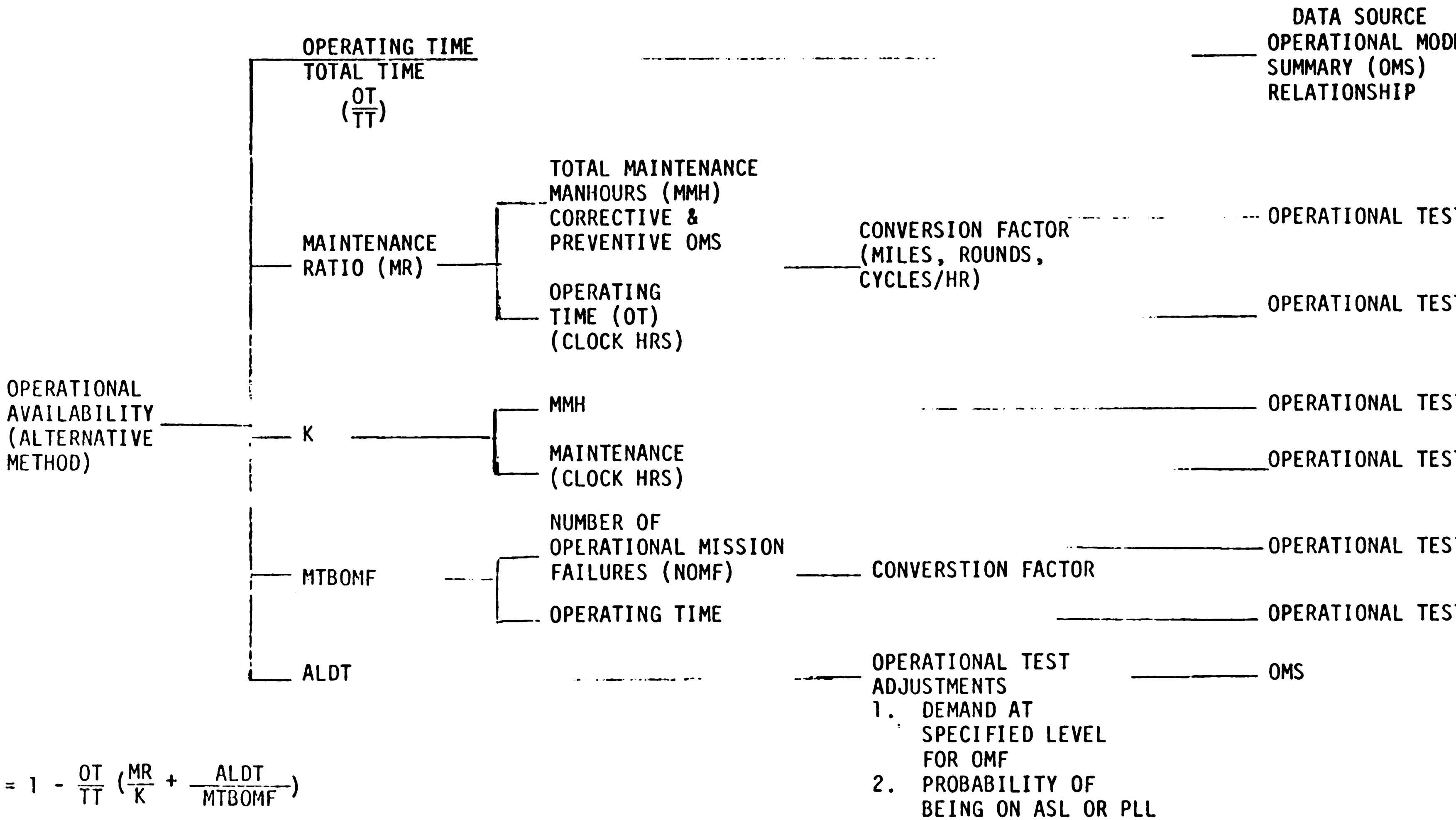
7. <u>EVALUATION.</u> Once testing is completed and data gathered from other sources, the evaluation can be completed. Analyses as planned in the IEP should be conducted. The evaluation should consider each issue for both positive and negative impacts. Changes to the logistics and materiel systems are almost inevitable. Any suggested changes should be thoroughly examined, since solving one problem often creates another. Conclusions addressing the

overall satisfaction with logistics should be stated and viable alternatives should be proposed where appropriate. The impacts of any deficiencies or proposed changes should be quantified.

APPENDIX A

LOGS OT&E DENTRITIC


OPERATIONAL AVAILABILITY
OPERATING TIME (CLOCK HRS)
CONVERSION FACTOR (MILES/HR) (ROUND/HR) (CYCLES/HR) FROM OPERATIONAL MODE SUMMARY
DATA SOURCE
OPERATIONAL TEST (HOURS, MILES, ROUND, CYCLES)
STANDBY TIME (CLOCK HRS)
OPERATIONAL MODE SUMMARY RELATIONSHIP-STANDBY TIME PER OPERATING TIME
TOTAL CORRECTIVE MAINTENANCE TIME (CLOCK HRS) (ALL CORRECTIVE MAINTENANCE ACTIONS)
OPERATIONAL TEST
TOTAL PREVENTIVE MAINTENANCE TIME (CLOCK HRS)
OPERATIONAL TEST
ADMINISTRATIVE & LOGISTICS DELAY TIME (CLOCK HRS)
OPERATIONAL TEST ADJUSTMENTS
1. DEMAND AT SPECIFIC LEVEL FOR OPERATIONAL MISSION FAILURES
2. PROBABILITY OF BEING ON ASL OR PLL
OPERATIONAL MODE SUMMARY

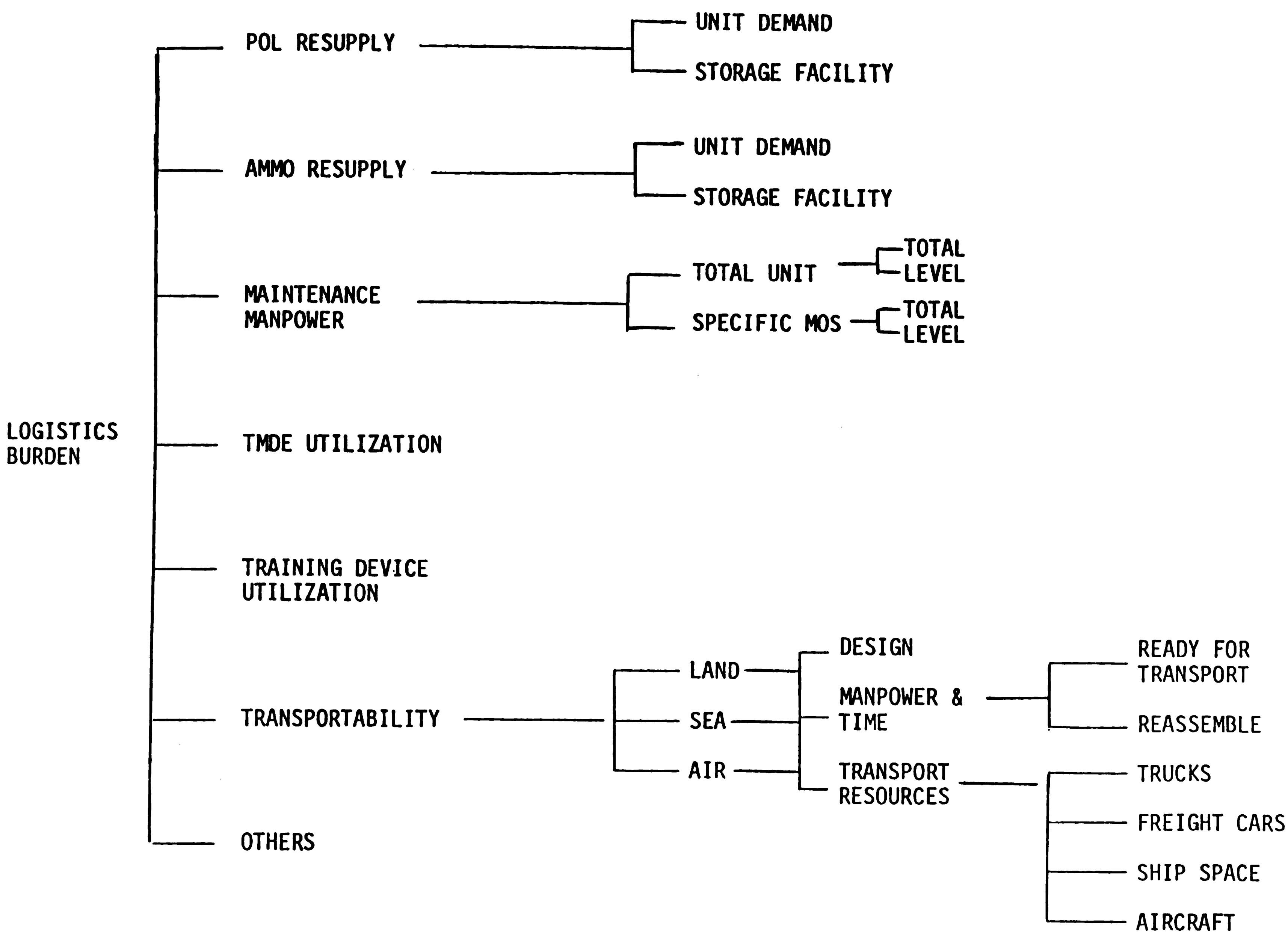


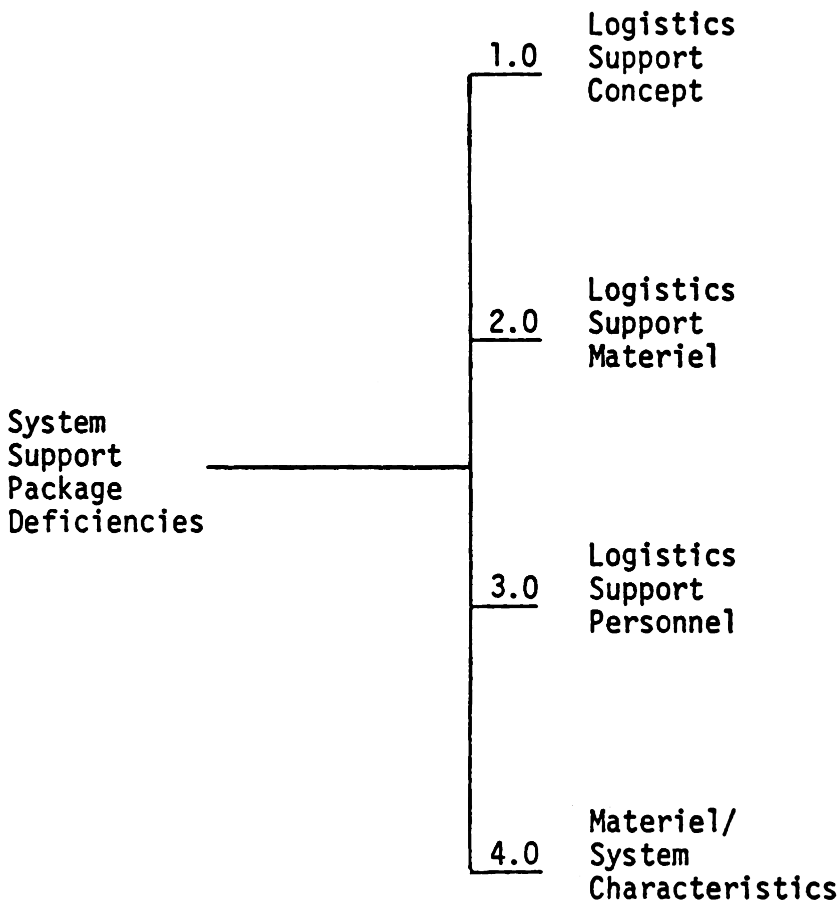
$$A_o = 1 - \frac{OT}{TT}\left(\frac{MR}{K} + \frac{ALDT}{MTBOMF}\right)$$

$$K = \frac{MMH}{MCH}$$

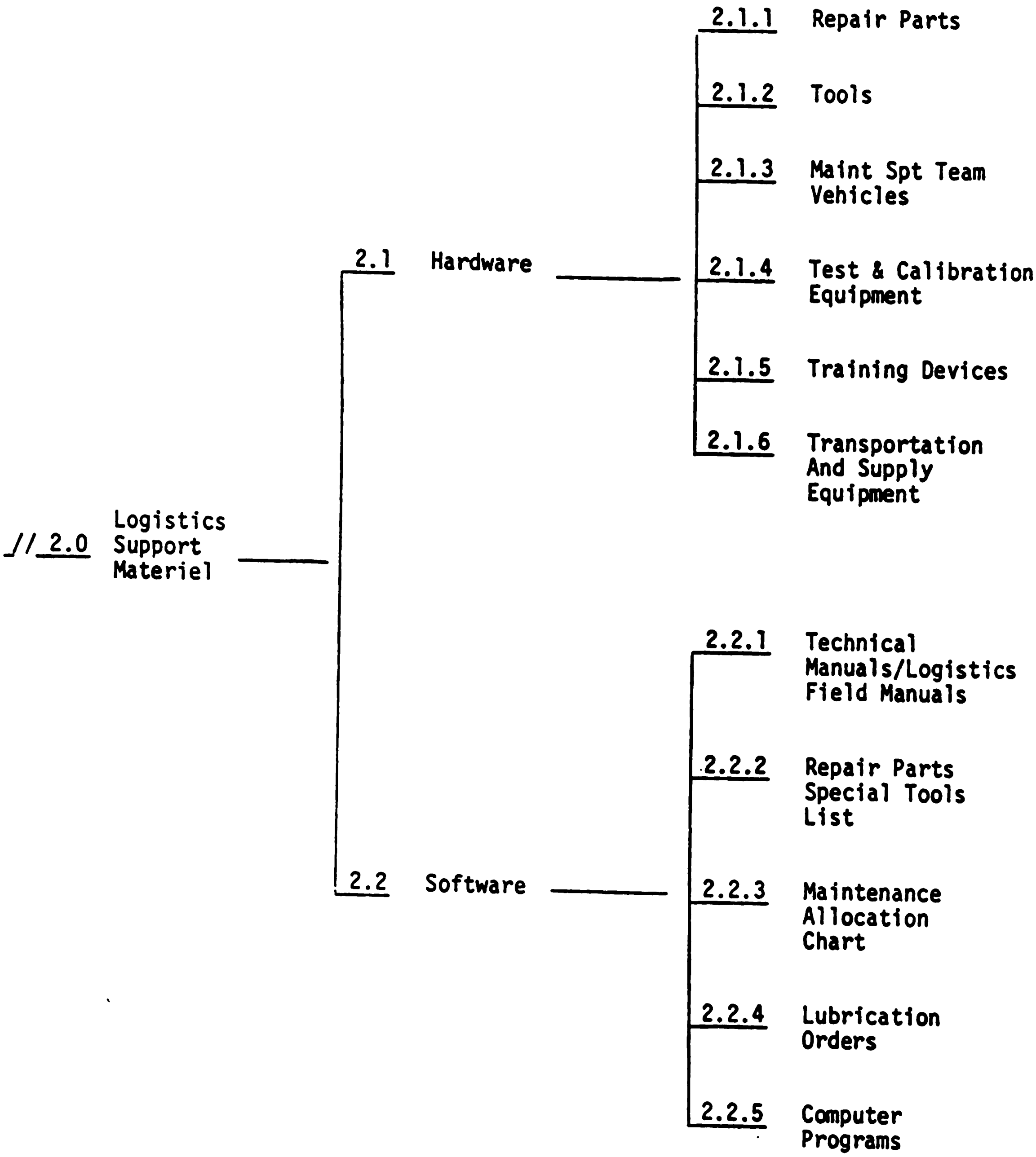
$$MR = \frac{MMH}{OT}$$

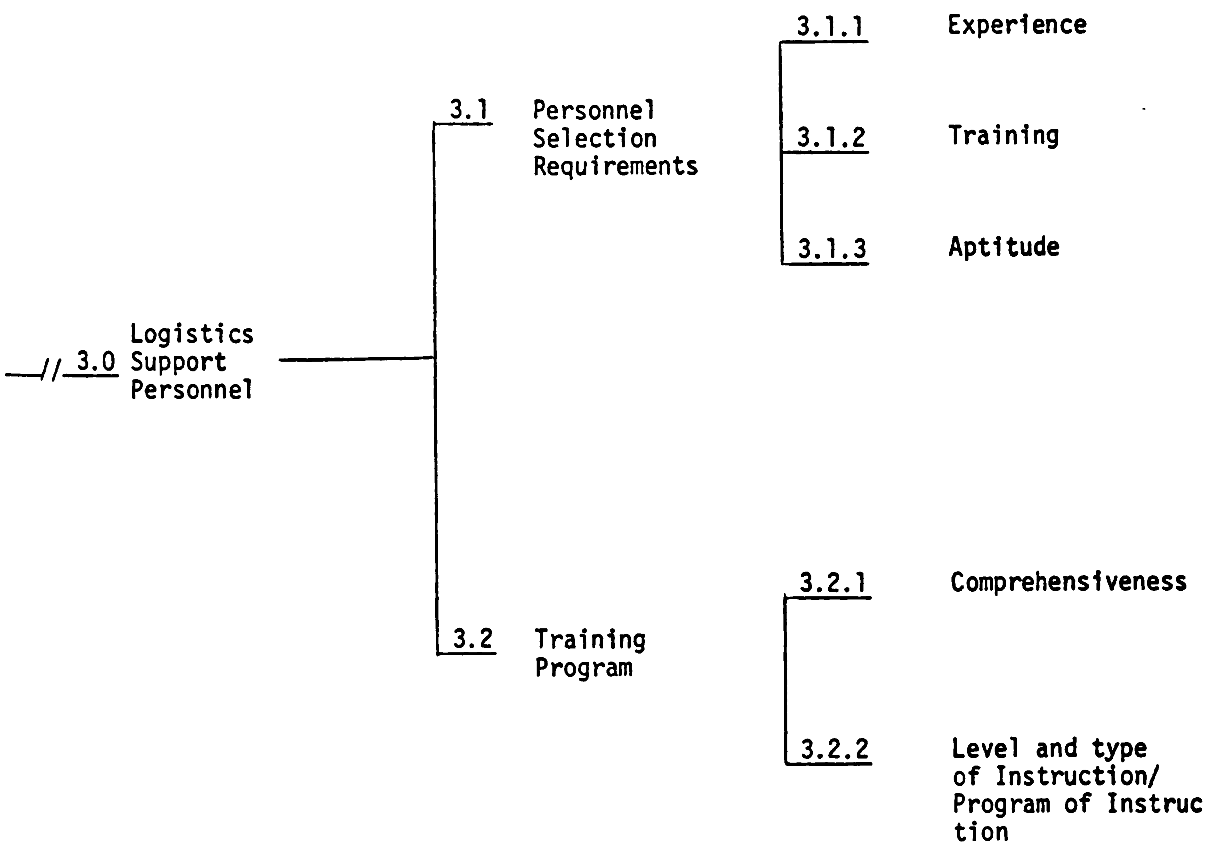
$$MTBOMF = \frac{NOMF}{OT}$$

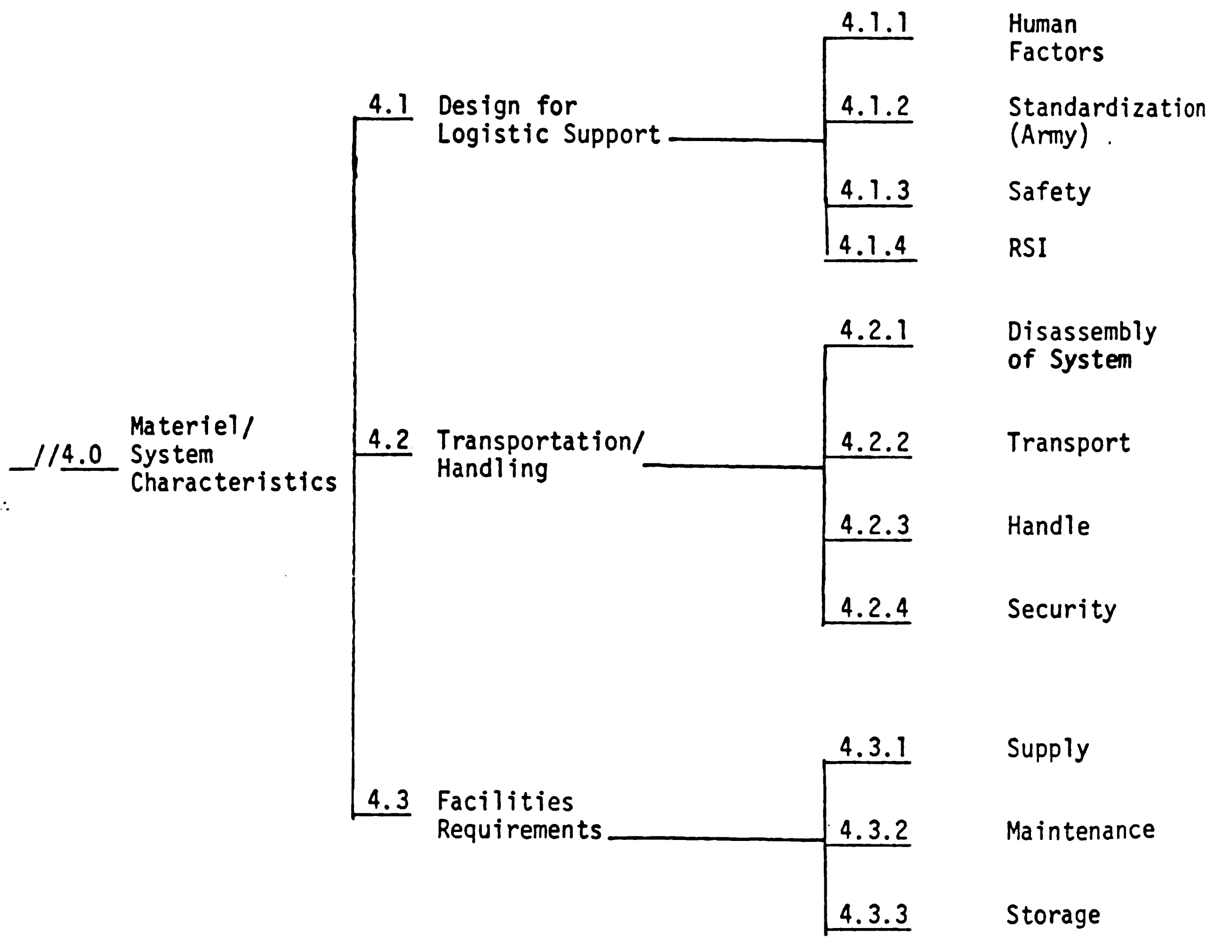

LOGISTICS BURDEN
POL RESUPPLY
UNIT DEMAND
STORAGE FACILITY
AMMO RESUPPLY
UNIT DEMAND
STORAGE FACILITY
MAINTENANCE MANPOWER
TOTAL UNIT
TOTAL
LEVEL
SPECIFIC MOS
TOTAL
LEVEL
TMDE UTILIZATION
TRAINING DEVICE UTILIZATION
TRANSPORTABILITY
LAND
SEA
AIR
DESIGN
MANPOWER & TIME
READY FOR TRANSPORT
REASSEMBLE
TRANSPORT RESOURCES
TRUCKS
FREIGHT CARS
SHIP SPACE
AIRCRAFT
OTHERS


System Support Package Deficiencies
1.0
Logistics Support Concept
2.0
Logistics Support Materiel
3.0
Logistics Support Personnel
4.0
Materiel/ System Characteristics

- 1.0 Logistics Support Concept
  - 1.1 Organization
    - 1.1.1 Crew/Operator
    - 1.1.2 Organizational
    - 1.1.3 Direct Support
    - 1.1.4 General Support
    - 1.1.5 Depot
  - 1.2 Allocation of Mission Responsibilities
    - 1.2.1 Maint/Calibration
    - 1.2.2 Repair Parts
    - 1.2.4 POL
    - 1.2.4 Ammunition Resupply
    - 1.2.5 Joint Services Interface
    - 1.2.6 Transportation
  - 1.3 Allocation of Mission Related Hardware/Software
    - 1.3.1 Mission Related Hardware/Software Required
    - 1.3.2 Amount of Mission Related Hardware/Software Available
    - 1.3.3 Location of Mission Related Hardware/Software
  - 1.4 Allocation of Supply/Maintenance Personnel
    - 1.4.1 Crew/Operator
    - 1.4.2 Organizational
    - 1.4.3 Direct Support
    - 1.4.4 General Support
    - 1.4.5 Depot


2.0 Logistics Support Materiel
2.1 Hardware
2.1.1 Repair Parts
2.1.2 Tools
2.1.3 Maint Spt Team Vehicles
2.1.4 Test & Calibration Equipment
2.1.5 Training Devices
2.1.6 Transportation And Supply Equipment
2.2 Software
2.2.1 Technical Manuals/Logistics Field Manuals
2.2.2 Repair Parts Special Tools List
2.2.3 Maintenance Allocation Chart
2.2.4 Lubrication Orders
2.2.5 Computer Programs


3.0 Logistics Support Personnel
3.1 Personnel Selection Requirements
3.1.1 Experience
3.1.2 Training
3.1.3 Aptitude
3.2 Training Program
3.2.1 Comprehensiveness
3.2.2 Level and type of Instruction/ Program of Instruc tion


__//4.0 Materiel/ System Characteristics
4.1 Design for Logistic Support
4.1.1 Human Factors
4.1.2 Standardization (Army)
4.1.3 Safety
4.1.4 RSI
4.2 Transportation/ Handling
4.2.1 Disassembly of System
4.2.2 Transport
4.2.3 Handle
4.2.4 Security
4.3 Facilities Requirements
4.3.1 Supply
4.3.2 Maintenance
4.3.3 Storage

APPENDIX B

LOGISTICS BURDEN ANALYSIS EXAMPLES

1. MANPOWER.

2. PETROLEUM, OIL AND LUBRICANTS (POL).

3. TEST MEASUREMENT AND DIAGNOSTIC EQUIPMENT (TMDE).

MANPOWER ANALYSIS EXAMPLE

Density - 6000 Items

Mission Length - 12 Hours

Combat Usage Rate - 2400 Hrs/Year

	Rel	MTBF	Failure/Years
Rqmt	.97	500	4.8
Data	.93	175	13.7

Increased burden increases in Failure/Year 8.9

Additional Failures/Years - 6000 Items X 8.9 = 53,400 Failures/Year

2 Manhours/Failure X 53,400 Failure/Year = 166,400 Manhours/Year

EQUIVALENT to additional 43 Repairman

Further analysis could be performed to determine the level(s) of maintenance (i.e., MOS) at which the shortage occurred. Addtionally, this shortfall may also indicate the need for additional tools, TMDE and in the case of contract team, vehicles.

POL BURDEN EXAMPLE

GENERATOR SET XYZ

Data Required for Unit Demand Analysis

ITEM	DESCRIPTION	SOURCE
POL CONSUMPTION PER SYS	10 GAL/HR	OPERATIONAL TEST
MISSION LENGTH	24 HR/DAY OPN	OPERATIONAL MODE SUMMARY
NO. OF GEN IN UNIT	20 GENERATORS	ORGN CONCEPT TSP ELEMENT
FUEL TANKER CAP	600 GAL	TOE AND TANKER DESCRIPTION

Unit Demand Analysis

10 GAL/HR X 24 HRS/DAY = 240 GAL/DAY/SYSTEM

240 GAL/DAY/SYSTEM X 20 SYSTEMS/UNIT = 4,800 GAL/DAY/UNIT

4,800 GAL/DAY/UNIT : 600 GAL/TANKER = 8 TANKER LOADS/DAY/UNIT

Criteria

5 GAL/HR/SYSTEM, 2400 GAL/DAY OR 4 TANKER LOADS/DAY/UNIT

ISSUE CRITERIA NOT MET: ACHIEVEMENT WAS TWICE THE CRITERIA THUS ANALYSIS OF IMPACT ON POL STORAGE FACILITY IS NEEDED

POL BURDEN EXAMPLE (CONT).

Additional Data Required for Storage Facility Impact Analysis

ITEM	DESCRIPTION	SOURCE
TANKER TURN AROUND TIME	4 HRS	OPERATIONAL TEST
NO. OF UNITS SUPPORTED	10 UNITS	LOG CONCEPT ELEMENT OF TEST SUPPORT PACKAGE
STORAGE BLADDER CAP	10,000 GAL	TOE & ITEM DESCRIPTION
TANKER OPERATIONAL AVAILABILITY	80%	FIELD OPERATIONAL READINESS REPORTS
REQUIRED SUP ON HAND	20 DAY	STORAGE FACILITY STD
MANPOWER STD-BLADDERS	3 OPERATORS/10 BLADDERS	ITEM DESCRIPTION TOE
TANKER	4 OPERATORS/TRUCK (I.E., 2 CREWS FOR 24 HRS)	ITEM DESCRIPTION

Supply Facility Impact Analysis

NUMBER OF TANKERS:

8 TANKER LOADS PER DAY X 10 UNITS SUPPORTED = 80 TANKER LOADS/DAY DEMAND

24 HR/DAY : 4 HR TURNAROUND TIME/TANKER = 6 TANKER LOADS/TANKER/DAY

80 TANKER LOADS/DAY REQUIRED : 6 TANKER LOADS/TANKER/DAY = 13 1/3 TANKERS REQUIRED

13.3 TANKERS REQUIRED X 1.2 OPERATIONAL AVAILABILITY FACTOR = 16 TANKERS REQUIRED IN UNI

(3 ARE DX SUPPLY ITEMS IN UNIT)

NUMBER OF BLADDERS REQUIRED:

4,800 GAL/DAY/UNIT X 10 UNITS = 48,000 GAL/DAY/SUPPLIED

48,000 GAL/DAY/SUPPLIED X 20 DAY SUPPLY REQUIRED = 960,000 GAL STORAGE REQUIRED

960,000 GAL STORAGE REQ : 10,000 GAL BLADDER CAPACITY = 96 BLADDERS REQUIRED

MANPOWER REQUIRED:

13 TANKERS OPERATION X 4 OPERATOR/TANKER = 52 TANKER OPERATORS

96 STORAGE BLADDERS : 10 BLADDERS/3 PEOPLE X 3 PEOPLE = 29 BLADDER OPERATORS

TOTAL MANPOWER = 81 PERSONS REQUIRED

Additional Resources Required from Not Having Met Criteria

ACHIEVEMENT WAS TWICE THE CRITERIA, THEREFORE, TWICE AS MANY RESOURCES REQUIRED AS WOULD HAVE BEEN IF CRITERIA HAD BEEN MET. THUS ADDITIONAL RESOURCES ARE:

TANKERS = 8 BLADDERS = 48 MANPOWER = 40

IF THESE RESOURCES ARE NOT ADDED, ONLY 1/2 OF THE TOTAL FLEET WILL BE MAINTAINED COMMITABLE WITH FUEL OR WILL BE ABLE TO OPERATE 1/2 THE 24 HR DAY. WHILE OPERATIONAL AVAILABILITY (Ao) DOES NOT INCLUDE MEASUREMENT OF FUEL AVAILABILITY, THIS LACK OF RESOURCES CARRIES THE SAME IMPACT ON Ao AS MAINTENANCE REQUIREMENTS. IN EFFECT, THE Ao FOR THIS GENERATOR IS 50% BEFORE MAINTENANCE BECOMES INVOLVED TO FURTHER DEGRADE Ao. 1/2 THE REQUIRED FUEL RESERVE WILL BE AVAILABLE.

The Secondary Logistics System

ALTHOUGH NOT ANALYZED, IF THE ADDITIONAL RESOURCES ARE ADDED, THEN IT WILL HAVE THE IMPACT OF DOUBLING MAINTENANCE AND SUPPLY RESOURCES REQUIRED TO SUPPORT TRUCKS AND BLADDERS FOR POL SUPPLY TO THE GENERATOR.

TMDE UTILIZATION EXAMPLE

Data Required

ITEM	DESCRIPTION	SOURCE
NO. OF SYS TO BE SPT/UNIT	200	TOE
NO. OF SYS SPT IN TEST	4	OPERATIONAL TEST
DAILY UTILIZATION TIME (TMDE)	2 HRS/DAY	OPERATIONAL TEST
Ao FOR SYSTEM	.90	OPERATIONAL TEST
Ao FOR TMDE	.80	DEVELOPMENTAL & OPERATIONAL TEST
NO. OF TMDE IN TEST	1	OPERATIONAL TEST
NO. OF TMDE PLANNED FOR UNIT	4	TOE

Utilization Analysis

AVERAGE UTILIZATION PER SYSTEM:

2 HR/DAY : 4 SYSTEMS TESTED = 1/2 HR/DAY/SYSTEM

200 SYSTEMS IN UNIT X .90 AVAILABILITY FACTOR = 180 SYS OPERATIONAL/DAY/UNIT

180 SYS OPERATIONAL X 1/2 HR/DAY/SYSTEM = 90 HR/DAY TMDE UTILIZATION

TMDE REQUIRED:

24 HR/DAY X .80 TMDE AVAILABILITY = 19.2 HRS OPERATION AVAILABLE/DAY

90 HR/DAY TMDE UTILIZATION : 19.2 HRS AVAILABLE/DAY = 4.7 TMDE ITEMS REQUIRED

IMPACT OF FIELDING WITH 4 TMDE:

1. WITHOUT COMBAT LOSSES BEING CONSIDERED 4 TMDE CAN SUPPORT:

4 TMDE AUTHORIZED X .80 Ao = 3.2 TMDE AVAILABLE

3.2 TMDE AVAILABLE X 24 HR/DAY = 76.8 HR OPERATION/DAY

76.8 HRS OPERATION/DAY : 1/2 HR TMDE DEMAND/DAY/SYSTEM = 153.6 SYS SUPPORTED

WITH 180 SYSTEMS OPERATION PER DAY, THIS SHOWS A BACK LOG OF SYSTEMS WILL DEVELOP AT THE TMDE STATIONS AND BUILD UNIT ONLY 154 ARE OPERATIONAL. THUS DEGRADING OPERATIONAL AVAILABILITY OF THE SYSTEM.

2. CONSIDERING COMBAT LOSSES:

EXPECTED % LOST PER DAY = 20%

THIS IMPLIES THAT AT THE END OF FIRST COMBAT DAY:

180 SYSTEM X .80 = 144 SYSTEMS + 20 IN MAINTENANCE = 164 SYSTEMS IN FLEET FOR SECOND DAY WITH

147 BEING AVAILABLE (164 X .90)

THUS, AFTER THE FIRST DAY OF BATTLE, THE TMDE WILL BE ABLE TO ACCOMMODATE THE UNIT SUPPORTED. CONSIDERING THIS, 4 TMDE ITEMS MAY BE ADEQUATE.

ABSTRA

issing dat
onal fac
ained the
sing est
operty).

action i
timable
averagi
points
resolut
ithout
ssentiall
only t
rts ar
e effe
e least
t necess
ete fo
ete on
ys to
hor s
ctional

I.

sign o
ler e
approp
ding
, 19
rred,
be a
inst
sign
ssing

Hop
ctoria
simpl
erest
nceive
nes
ed f

THE PERVERSITY OF MISSING POINTS IN THE 2^4 DESIGN

Carl T. Russell
US Army Cold Regions Test Center
Fort Greely, Alaska

ABSTRACT. The author would like better to understand the impact of missing data on estimability (and variance) in 2^n factorials and 2^{n-k} fractional factorials. Hoping to find generalizable results, the author examined the 2^4 design to determine what points could be deleted without losing estimability of main effects and 2-factor interactions (resolution V property). He was guided by a result of P. W. M. John which shows that if a fraction is missing from a 2^n design, then estimable effects are those estimable from half replicates, and the least squares estimates are obtained by averaging the estimates from half replicates. In particular, if one or two points are missing from the 2^4 factorial, then the remaining design is of resolution V, and the least squares estimates can be written down easily (without explicitly solving the normal equations). Likewise, there are essentially six ways a quarter replicate can be deleted from the 2^4 design, and only two of those leave designs of resolution V. However, if only three points are deleted, the remaining design is always of resolution V, estimable effects are not necessarily those estimable from half replicates, and the least squares estimates of effects estimable from half replicates are not necessarily averages of estimates from half replicates. The only way to delete four points and fail to have a remaining design of resolution V is to delete one of the fractions mentioned above. Moreover, there are numerous ways to delete five points but still retain a design of resolution V. The author seeks insight to what is going on with missing points in the 2^4 factional, hopefully insight which can be generalized to other designs.

I. INTRODUCTION. Factorial designs are frequently exploited in the design of field tests of military materiel. I suspect that they can be better exploited. For example, field tests can be run in blocks consisting of appropriately chosen fractional factorials to reduce the bias due to confounding which is common in much traditional field test design (see Russell, 1981, 1982). Unfortunately, execution of a field test seldom proceeds as planned, and rather large amounts of missing data are common. I would like to be able to produce experimental designs which are in some sense robust against data loss. In particular, I would hate inadvertently to use a design with nice theoretical properties which could easily be demolished by missing data.

Hoping to gain a better understanding of the impact of data loss in factorial designs, I began an empirical study of what I anticipated would be a simple case, the 2^4 design. (The 2^4 design is also of great practical interest, since an experiment in four factors each at two levels can be conceived and displayed easily but still provides substantial analytical richness.) The study was limited to considering what points could be deleted from the 2^4 design without losing estimability of main effects and

2-factor interactions (resolution V property) and to considering the structure of the least squares (LS) estimates obtained. In fact, the only non-trivial cases considered thoroughly were the cases in which three points were deleted. Although I was able to obtain some insight in the three-point cases, that insight was limited and incomplete. Since the four- and five-point cases appear to be more complicated, this paper deals mostly with the pathology which results when three points are deleted from the 2^4 design.

II. NOTATION, ANTICIPATED RESULTS, AND ACTUAL RESULTS. The full 2^4 design was conceived as a labelled test point matrix (TPM) in four factors A, B, C, and D, where the presence of a particular lower-case letter in a cell label indicated that the corresponding factor was at high level in that cell. The labelled TPM was

(1)	c	d	cd
a	ac	ad	acd
b	bc	bd	bcd
ab	abc	abd	abcd

Potential data to be obtained from this design were modelled as

$$\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{B} + \underset{\sim}{e}$$

where the design matrix was

$$\underset{\sim}{X} = \begin{bmatrix} 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & -1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & -1 & 1 & -1 & 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} \underset{\sim}{x}'_{(1)} \\ \underset{\sim}{x}'_{a} \\ \underset{\sim}{x}'_{b} \\ \underset{\sim}{x}'_{ab} \\ \underset{\sim}{x}'_{c} \\ \underset{\sim}{x}'_{ac} \\ \underset{\sim}{x}'_{bc} \\ \underset{\sim}{x}'_{abc} \\ \underset{\sim}{x}'_{d} \\ \underset{\sim}{x}'_{ad} \\ \underset{\sim}{x}'_{bd} \\ \underset{\sim}{x}'_{abd} \\ \underset{\sim}{x}'_{cd} \\ \underset{\sim}{x}'_{acd} \\ \underset{\sim}{x}'_{bcd} \\ \underset{\sim}{x}'_{abcd} \end{bmatrix}$$

and the vector of parameters was

$$\underset{\sim}{B} = (\mu,\alpha,\beta,\alpha\beta,\gamma,\alpha\gamma,\beta\gamma,\delta,\alpha\delta,\beta\delta,\gamma\delta)'.$$

For convenience, the random vector $\underset{\sim}{e}$ was assumed to have an identity matrix as its dispersion matrix. Deleting points from the TPM (for example, deleting {(1), a} and placing "X" at appropriate positions in the TPM) resulted in deletion of the corresponding rows (for example, $\underset{\sim}{x}'_{(1)}$ and $\underset{\sim}{x}'_{a}$) from the design matrix to obtain a reduced design matrix $\underset{\sim}{R}$, and estimability of all effects[1] was determined by checking $\underset{\sim}{R}'\underset{\sim}{R}$ for singularity. When $\underset{\sim}{R}'\underset{\sim}{R}$ was nonsingular, the LS estimate of any specified effect was obtained from the appropriate row of $(\underset{\sim}{R}'\underset{\sim}{R})^{-1}\underset{\sim}{R}$ and portrayed in terms of the test points by writing the weights for each remaining test point in the corresponding TPM: for example, with {(1), a} missing from the TPM, the LS estimate of the A effect was

$32\hat{\alpha} =$

X	-3	-3	-2
X	3	3	2
-3	-2	-2	-1
3	2	2	1

$$= -3b+3ab-3c+3ac-2bc+2abc-3d+3ad-2bd+2abd-2cd+2acd-bcd+abcd.$$

I was guided in this study by a result of P. W. M. John (1971, pages 161-163) which shows that if a fraction is missing a 2^n design, then estimable effects are those estimable from half replicates, and the LS estimates are obtained by averaging the estimates from half replicates. In particular, if one or two points are missing from the 2^4 factorial, then the remaining design is of resolution V, and the LS estimates can be written down explicitly without solving the normal equations. For example, {(1), a} defines the 2^{4-2} fraction with defining contrast

$$I = -B = -C = BC = -D = BD = CD = -BCD$$

so that if {(1), a} is deleted from the TPM, the remaining design contains the half replicates I = B, I =C, I = -BC, I = D, I = -BD, I = -CD, and I = BCD. The main effect A is estimable in the four half replicates defined by two or three factors (since it is aliased with 3- or 4-factor interactions in those half replicates), and its LS estimate is obtained from

[1]Henceforth "effect" will refer to the mean (I), a main effect (A, B, C, or D) or a 2-factor interaction (AB, AC, AD, BC, BD, or CD).

$32\hat{\alpha} =$

8α

X	-1		-1
X	1		1
-1		-1	
1		1	

I = -BC

\+

8α

X		-1	-1
X		1	1
-1	-1		
1	1		

I = -BD

\+

8α

X	-1	-1	
X	1	1	
	-1	-1	
	1	1	

I = -CD

\+

8α

X	-1	-1	
X	1	1	
-1			-1
1			1

I = BCD

=

32α

X	-3	-3	-2
X	3	3	2
-3	-2	-2	-1
3	2	2	1

This computation[2] thus provides a structural explanation for the rather strange looking estimate which pops out of the normal equations.

There are essentially six ways in which a quarter replicate can be deleted from the 2^4 design to obtain what is called the three-quarter replicate defined by the defining contrast for the deleted quarter replicate. These are characterized by the numbers of 1-, 2-, 3-, and 4-factor interactions in the defining contrasts. Only two yield designs of resolution V: one is represented by any fraction with 1-, 3-, and 4-factor interactions in its defining contrast, and the other is represented by a design with a 2-factor interaction and two 3-factor interactions in its defining contrast. The other four designs fail to be resolution V since not all effects are estimable from half replicates contained in the remaining design.

Any three points in the 2^4 design are contained in exactly one quarter replicate, and a design obtained by deleting three points from the 2^4 design contains just those half replicates which are contained in the corresponding three-quarter replicate obtained by deleting the quarter replicate containing the three points. Since many of those three-quarter replicates are not

[2]The notation introduced in this computation will be used repeatedly in what follows. LS estimates are written as sums of estimates from appropriate fractions. Each fraction is represented as a TPM having appropriate weights at all points in the fraction and having blanks or "X"s at all points not in the fraction. The defining contrast for the fraction is written below the TPM and the expectation of the estimate is written above the TPM.

of resolution V, I therefore anticipated finding many ways to destroy the resolution V property by deleting three points from the TPM. This anticipation turned out to be incorrect. All designs obtained from the 2^4 design by deleting three points from the TPM are of resolution V: there are estimable effects which are not estimable from remaining half replicates. Moreover, in the two cases where all effects are estimable from half replicates, the least squares estimates are not averages of estimates from half replicates.

Examination of the various ways four points can be deleted from the 2^4 design showed that the only way to delete four points from the 2^4 design and fail to have a remaining design of resolution V was to delete one of the quarter replicates for which the remaining three-quarter replicate is not of resolution V. Moreover, there are numerous ways to delete five points and still retain a design of resolution V.

III. DESIGNS OBTAINED BY DELETING THREE POINTS FROM THE 2^4 DESIGN.

There are 560 ways to delete three points from the sixteen points in the TPM for the 2^4 design. Since any three points are contained in exactly one quarter replicate, these 560 ways can be classified into six cases by relabelling factors and factor levels so that the quarter replicate containing the deleted points also contains (1) and (1) is not deleted. These six cases are described in Table 1.

TABLE 1. Classification of Designs Remaining After Three Points Are Deleted From the 2^4 Design.

Case	Points Deleted	Defining Contrasts*	No. Ways Obtained**
1	a, b, ab	I = -C = -D = CD	96
2	a, bc, abc	I = -D = BC = -BCD	192
3	a, bcd, abcd	I = BC = BD = CD	64
4	ab, ac, bc	I = -D = -ABC = ABCD***	64
5	ab, cd, abcd	I = AB = CD = ABCD	48
6	ab, acd, bcd	I = -ABC = -ABD = CD***	96

*Defining contrast for the quarter replicate which contains the points deleted.
**Number of ways this design can be obtained by relabelling factors and factor levels so that the defining contrast contains (1) and (1) is not deleted.
***These two contrasts define three-quarter replicates of resolution V.

The method used for this reduction to six cases (relabelling factors and factor levels) changes signs and interchanges labels among main effects and among 2-factor interactions but does not interchange main effects with

2-factor interactions. A further reduction to three cases can be accomplished by introducing four new factor labels W, X, Y, and Z with W = A, X = -AB, Y = -AC, and Z = -AD. This induces a relabelling of the TPM

from

(1)	c	d	cd
a	ac	ad	acd
b	bc	bd	bcd
ab	abc	abd	abcd

to

(1)	y	z	yz
wxyz	wxz	wxy	wx
x	xy	xz	xyz
wyz	wz	wy	w

and produces a formal correspondence between the six cases in pairs or two (Table 2).

TABLE 2. Formal Correspondence Between Three-Point Cases in Pairs of Two

Correspondence	ABCD Notation	WXYZ Notation
Case 1 ~ Case 3		
Case	1	3
Points Deleted	a, b, ab	wxyz, x, wyz
Defining Contrast	I= -C = -D = CD	I = WY = WZ = YZ
Case 2 ~ Case 5		
Case	2	5
Points Deleted	a, bc, abc	wxyz, xy, wz
Defining Contrast	I = -D = BC = -BCD	I = WZ = XY = WXYZ
Case 4 ~ Case 6		
Case	4	6
Points Deleted	ab, ac, bc	wyz, wxz, xy
Defining Contrast	I= -D = -ABC = ABCD	I = WZ = -WXY = -XYZ

In cases 1 and 3, one effect is estimable from three half replicates, six are estimable from one half replicate, and four are not estimable from half replicates. In cases 2 and 5, four effects are estimable from two half replicates, three are estimable from one half replicate, and four are not estimable from half replicates. In cases 4 and 6, two effects (one the mean effect) are estimable from two half replicates and the other nine effects are estimable from one half replicate. Because of the formal correspondence between the two cases in each pair, structure of LS estimates need only be studied for the first case in each pair (cases 1, 2, and 4): LS estimates for the other case in each pair (cases 3, 5, and 6) can be obtained simply by relabelling the weighted TPMs for the first case using the WXYZ notation.

A. Case 1 and Case 3 (Represented by Case 1). For case 1, the three half replicates contained in the remaining design are I = C, I = D, and I = -CD. AB is estimable from all three half replicates, A and B are estimable only from I = CD, AC and BC are estimable only from I = D, AD and BD are estimable only from I = C, and I, C, D, and CD are not estimable from half replicates. The LS estimates of effects estimable from half replicates are the estimates from half replicates or their average (in the case of AB).

$24\hat{\alpha}$ = (24α; I = -CD)

		-3	-3
X		3	3
X		-3	-3
X		3	3

$24\widehat{\alpha\gamma}$ = (24αγ; I = D)

		3	-3
X		-3	3
X		3	-3
X		-3	3

$24\widehat{\alpha\delta}$ = (24αδ; I = C)

	3		-3
X	-3		3
X	3		-3
X	-3		3

$24\hat{\beta}$ = (24β; I = -CD)

		-3	-3
X		3	3
X		-3	-3
X		3	3

$24\widehat{\beta\gamma}$ = (24βγ; I = D)

		3	-3
X		3	-3
X		-3	3
X		-3	3

$24\widehat{\beta\delta}$ = (24βδ; I = C)

	3		-3
X	3		-3
X	-3		3
X	-3		3

$24\widehat{\alpha\beta}$ = (8αβ; I = -CD)

	1	1	
X	-1	-1	
X	-1	-1	
X	1	1	

\+ (8αβ; I = D)

		1	1
X		-1	-1
X		-1	-1
X		1	1

\+ (8αβ; I = C)

	1		1
X	-1		-1
X	-1		-1
X	1		1

= (24αβ)

	2	2	2
X	-2	-2	-2
X	-2	-2	-2
X	2	2	2

Although the effects C, D, and CD are not estimable from half replicates, each is estimable from a quarter replicate which is contained in the remaining design and contains (1); such quarter replicates will be referred to as (1)-quarter replicates. The (1)-quarter replicates from which C, D, and CD are estimable are the following.

Effect Estimable	Defining Contrast for (1)-Quarter Replicate	Points in (1) Quarter Replicate
CD	I = -A = -B = AB	(1), c, d, cd
C	I = AB = AD = BD	(1), c, abd, abcd
D	I = AB = AC = BC	(1), d, abc, abcd

Inconveniently, the estimates obtained from these (1)-quarter replicates are not the LS (minimum variance) estimates. Instead, the LS estimate of any effect not estimable form a half replicate can be obtained by estimating it in a (1)-quarter replicate where it is aliased with ±AB (the effect estimated with smallest variance among those effects estimable form half replicates) then correcting for the bias using the least square estimates of the effects from half replicates. For example, C is aliased with -AB in four (1)-quarter replicates.

Points In (1)-Quarter Replicate	Defining Contrast for (1)-Quarter Replicate	Alias Chain Containing C
(1), d, ac, acd	I = -B = AC = -ABC	C + A - AB - BC
(1), d, bc, bcd	I = -A = BC = -ABC	C + B - AB - AC
(1), ac, abd, bcd	I = BD = -ACD = -ABC	C - AB - AD + BCD
(1), bc, abd, acd	I = AD = -BCD = -ABC	C - AB - BD + ACD

Since 3-factor interactions are assumed to be zero, the four (identical!) estimates are as follows.

$24\hat{\gamma}$ = [$24(\gamma+\alpha-\alpha\beta-\beta\gamma)$] + [$-24\alpha$] + [$24\alpha\beta$] + [$24\beta\gamma$]

$24(\gamma+\alpha-\alpha\beta-\beta\gamma)$

-6		-6	
X	6		6
X			
X			

I=-B=AC=-ABC

-24α

	3	3	
X	-3	-3	
X	3	3	
X	-3	-3	

$24\alpha\beta$

	2	2	2
X	-2	-2	-2
X	-2	-2	-2
X	2	2	2

$24\beta\gamma$

		3	-3
X		3	-3
X		-3	3
X		-3	3

= [$24(\gamma+\beta-\alpha\beta-\alpha\gamma)$] + [$-24\beta$] + [$24\alpha\beta$] + [$24\alpha\gamma$]

$24(\gamma+\beta-\alpha\beta-\alpha\gamma)$

-6		-6	
X			
X	6		6
X			

I=-A=BC=-ABC

-24β

	3	3	
X	3	3	
X	-3	-3	
X	-3	-3	

$24\alpha\beta$

	2	2	2
X	-2	-2	-2
X	-2	-2	-2
X	2	2	2

$24\alpha\gamma$

		3	-3
X		-3	3
X		3	-3
X		-3	3

$24(\gamma-\alpha\beta-\alpha\delta)$

$=$

-6			
X	6		
X			6
X		-6	

I=BD=-ACD=-ABC

$+$

$24\alpha\beta$

	2	2	2
X	-2	-2	-2
X	-2	-2	-2
X	2	2	2

$+$

$24\alpha\delta$

	3		-3
X	-3		3
X	3		-3
X	-3		3

$24(\gamma-\alpha\beta-\beta\delta)$

$=$

-6			
X			6
X	6		
X		-6	

I=AD=-BCD=-ABC

$+$

$24\alpha\beta$

	2	2	2
X	-2	-2	-2
X	-2	-2	-2
X	2	2	2

$+$

$24\beta\delta$

	3		-3
X	3		-3
X	-3		3
X	-3		3

24γ

$=$

-6	5	2	-1
X	1	-2	1
X	1	-2	1
X	-1	-4	5

The very strange looking LS estimate for the effect of C which comes from solving the normal equations can therefore be explained in this case by a reasonable rule. The same rule works for D and CD as well as I (which is not even estimable from a (1)-quarter replicate). For example,

$24\hat{\delta}=$

$24(\delta-\alpha\beta-\alpha\gamma)$

-6			
X		6	
X			6
X	-6		

I=BC=-ABD=-ACD

$+$

$24\alpha\beta$

	2	2	2
X	-2	-2	-2
X	-2	-2	-2
X	2	2	2

$+$

$24\alpha\gamma$

		3	-3
X		-3	3
X		3	-3
X		-3	3

$=$

24δ

-6	2	5	-1
X	-2	1	1
X	-2	1	1
X	-4	-1	5

$24\widehat{\gamma\delta} = $ [24(γδ−α+αβ)] + [24α] + [−24αβ] = [24γδ]

$24(\gamma\delta-\alpha+\alpha\beta)$

6		6	
X	-6	-6	
X			
X			

I=-B=-ACD=ABCD

+

24α

	-3	-3	
X	3	3	
X	-3	-3	
X	3	3	

+

$-24\alpha\beta$

	-2	-2	-2
X	2	2	2
X	2	2	2
X	-2	-2	-2

=

$24\gamma\delta$

6	-5	-5	4
X	-1	-1	2
X	-1	-1	2
X	1	1	-2

$24\hat{\mu} = $ [24(μ+αβ)] + [−24αβ] = [24μ]

$24(\mu+\alpha\beta)$

6			6
X			
X			
X	6	6	

I=AB=-ACD=-BCD

+

$-24\alpha\beta$

	-2	-2	-2
X	2	2	2
X	2	2	2
X	-2	-2	-2

=

24μ

6	-2	-2	4
X	2	2	2
X	2	2	2
X	4	4	-2

The variances of these LS estimates are as follows.

Effect Estimated	Variance of LS Estimate
I, C, D, CD	5/24 = 0.21
A, B, AC, BC, AD, BD	3/24 = 0.12
AB	2.24 = 0.08

B. Case 2 and Case 5 (Represented by Case 2). For case 2, the three half replicates contained in the remaining design are I = D, I = -BC, and I = BCD. A and AD are estimable from both I = -BC and I = BCD, AB and AC are estimable from both I = D and I = BCD; BC, D, and I are each estimable from only one half replicate, and B, C, BD, and CD are not estimable from half replicates. As in case 1, the LS estimate of an effect estimable from half replicates is the average of the estimates from half replicates.

$16\hat{\mu} = $

16μ

	2	2	
X	2	2	
2	X		2
2	X		2

I=BCD

$16\hat{\delta} = $

16δ

	-2		2
X	-2		2
-2	X	2	
-2	X	2	

I=-BC

$16\widehat{\beta\gamma} = $

$16\beta\gamma$

		2	-2
X		2	-2
	X	-2	2
	X	-2	2

I=D

$16\hat{\alpha} =$

8α

	-1	-1	
X	1	1	
-1	X		-1
1	X		1

I=BCD

+

8α

	-1		-1
X	1		1
-1	X	-1	
1	X	1	

I=-BC

=

16α

	-2	-1	-1
X	2	1	1
-2	X	-1	-1
2	X	1	1

$16\widehat{\alpha\delta} =$

$8\alpha\delta$

	1	-1	
X	-1	1	
1	X		-1
-1	X		1

I=BCD

+

$8\alpha\delta$

	1		-1
X	-1		1
1	X	-1	
-1	X	1	

I=-BC

=

$16\alpha\delta$

	2	-1	-1
X	-2	1	1
2	X	-1	-1
-2	X	1	1

$16\widehat{\alpha\beta} =$

$8\alpha\beta$

	1	1	
X	-1	-1	
-1	X		-1
1	X		1

I=BCD

+

$8\alpha\beta$

		1	1
X		-1	-1
	X	-1	-1
	X	1	1

I=D

=

$16\alpha\beta$

	1	2	1
X	-1	-2	-1
-1	X	-1	-2
1	X	1	2

$16\widehat{\alpha\gamma} =$

$8\alpha\gamma$

	-1	1	
X	1	-1	
1	X		-1
-1	X		1

I=BCD

+

$8\alpha\gamma$

		1	-1
X		-1	1
	X	1	-1
	X	-1	1

I=D

=

$16\alpha\gamma$

	-1	2	-1
X	1	-2	1
1	X	1	-2
-1	X	-1	2

In a similar manner to that of case 1, the LS estimates of B, C, BD, and CD can be obtained by estimating each from a (1)-quarter replicate where it is aliased with effects estimable from more than one half replicate, then correcting for bias. However, the (1)-quarter replicate used must alias the effect of interest with <u>two</u> effects estimable from two half replicates: there are four such (1)-quarter replicates for each of B, C, BD, and CD, and for each of these effects, all four (1)-quarter replicates yield the same estimate. Just one of the (1)-quarter replicates in each set aliases the effect with a higher-order interaction (in this case, the 3-factor interaction ±BCD), and the resulting LS estimates are as follows.

$16\hat{\beta} =$

$16(\beta-\alpha\gamma-\alpha\delta)$

-4			
X			-4
	X		4
4	X		

I=CD=-ABC=-ABD

+

$16\alpha\gamma$

	-1	2	-1
X	1	-2	1
1	X	1	-2
-1	X	-1	2

+

$16\alpha\delta$

	2	-1	-1
X	-2	1	1
2	X	-1	-1
-2	X	1	1

=

16β

-4	1	1	-2
X	-1	-1	-2
3	X	0	1
1	X	0	3

$16\hat{\gamma} =$

$16(\gamma-\alpha\beta-\alpha\delta)$

-4			
X	4		
	X		4
	X	-4	

I=BD=-ABC=-ACD

+

$16\alpha\beta$

	1	2	1
X	-1	-2	-1
-1	X	-1	-2
1	X	1	2

+

$16\alpha\delta$

	2	-1	-1
X	-2	1	1
2	X	-1	-1
-2	X	1	1

=

16γ

-4	3	1	0
X	1	-1	0
1	X	-2	1
-1	X	-2	3

$16\widehat{\beta\delta} =$

$16(\beta\delta-\alpha+\alpha\gamma)$

4			
X		-4	
	X	4	
-4	X		

I=-C=-ABD=ABCD

+

16α

	-2	-1	-1
X	2	1	1
-2	X	-1	-1
2	X	1	1

+

$-16\alpha\gamma$

	1	-2	1
X	-1	2	-1
-1	X	-1	2
1	X	1	-2

=

$16\beta\delta$

4	-1	-3	0
X	1	-1	0
-3	X	2	1
-1	X	2	-1

$16\widehat{\gamma\delta}$ = [$16(\gamma\delta-\alpha+\alpha\beta)$] + [$16\alpha$] + [$-16\alpha\beta$] = [$16\gamma\delta$]

$16(\gamma\delta-\alpha+\alpha\beta)$

4			4
X	-4	-4	
	X		
	X		

I=-B=-ACD=ABCD

16α

	-2	-1	-1
X	2	1	1
-2	X	-1	-1
2	X	1	1

$-16\alpha\beta$

	1	-2	-1
X	-1	2	1
1	X	1	2
-1	X	-1	-2

$16\gamma\delta$

4	-1	-3	2
X	-3	-1	2
-1	X	0	1
1	X	0	-1

The variances of these LS estimates are as follows.

Effect Estimated	Variance of LS Estimate
B, C, BD, CD	6/32 = 0.19
I, D, BC	4/32 = 0.12
A, AB, AC, AD	3.32 = 0.09

In both case 1 and case 2 (therefore also case 3 and case 5), the scheme given for obtaining the LS estimates of effects not estimable from half replicates uses a (1)-quarter replicate and estimates from all three half replicates contained in the remaining design. However, using a (1)-quarter replicate and estimates from all three half replicates contained in the remaining design is not enough. This and more can be seen by considering all estimates for effect C in case 2 which are obtained by estimating C in a (1)-quarter replicate, then correcting for bias. For case 2 (actually for cases 1-6) there are sixteen (1)-quarter replicates in all. In case 2, four of the (1)-quarter replicates yield the LS estimate of C. These are: I = BD = -ABC = -ACD, I = -A = CD = -ACD, I = -A = -C = AC, and I = -A = -B = AB. The remaining twelve (1)-quarter replicates yield a total of eight different estimates of C having two different variances. These estimates are listed below together with the (1)-quarter replicate used (beneath the TPM) and the alias chain for C in that (1)-quarter replicate (above the TPM).

C-AD-BD+ABC

-4	2	1	1
X	2	-1	-1
2	X	-3	1
-2	X	-1	3

I=-B=AB=-ACD

C-AB+D-ABCD

-4	3	2	-1
X	1	-2	1
1	X	-3	2
-1	X	-1	2

I=CD=-ABC=-ABD

C-AC-BC+ABC

-4	3		1
X	1		-1
1	X	-1	
-1	X	-3	4

I=-A=-B=AB

C-AB-I+ABC

-4	3		1
X	1		-1
1	X	-1	
-1	X	-3	4

I=-C=AB=-ABC

C-AD-I+ACD

-4	4	1	-1
X		-1	1
	X	-1	1
	X	-3	3

I=-C=AD=-ACD

C+A+D+ACD

-4	4	1	-1
X		-1	1
	X	-1	1
	X	-3	3

I=AC=AD=CD

C-AC+BCD-ABCD

-4	3	2	-1
X	1	-2	1
1	X	-3	2
-1	X	-1	2

I=-A=BD=-ABD

C+A+BCD+ABD

-4	2	1	1
X	2	-1	-1
2	X	-3	1
-2	X	-1	3

I=AC=BD=ABCD

C-BC+ACD-ABCD

-4	4	2	-2
X		2	2
	X	-2	2
	X	-2	2

I=-B=AD=-ABD

C+D+ABC+ABD

-4	2		2
X	2		-2
2	X	-2	
-2	X	-2	4

I=AB=CD=ABCD

C-I+ABD-ABCD

-4	2	2	
X	2	-2	
2	X	-4	2
-2	X		2

I=-C=-ABD=ABCD

C+ABC+ACD+BCD

-4	4		
X			
	X		
	X	-4	4

I=AB=AD=BD

Each estimate in the first row estimates C in a (1)-quarter replicate where it is aliased with an effect from two half replicates and a second effect estimated from the remaining half replicate, but neither is the LS estimate (each has variance 7/32 = 0.22); thus merely involving all three half replicates in the estimation of C is not enough. Each estimate in the second row estimates C in a (1)-quarter replicate where is is aliased with both an effect estimated from two half replicates and a second effect estimated from one of those same half replicates; this row contains two copies each of two different estimates. Each estimate in the third row estimates C in a (1)-quarter replicate where it is aliased only with a single effect estimable from two half replicates; each estimate in the third row is the same as one of the estimates in the first row. All estimates in the first three rows have the same variance, 7/32 = 0.22). Each of the four different estimates in the last two rows has variance 8/32 = 0.25. Each of the three estimates in the fourth row estimates C in a (1)-quarter replicate where it is aliased only with a single effect estimable from one half replicate, and the estimate in the last row estimates C in the (1)-quarter replicate where it is estimable.

<u>C. Case 4 and Case 6 (Represented by Case 4)</u>. For case 4, the three half replicates contained in the remaining design are I = D, I = ABC, and I = -ABCD. D and I are estimable from I = ABC and I = -ABCD. A, B, and C are estimable from I = -ABCD; AC, BC, and AB are estimable from I = D; and AD, BD, and CD are estimable from I = ABC. Thus all effects are estimable from half replicates. However, none of the estimates obtained by averaging estimates from half replicates is the LS estimate. The LS estimate of D can be obtained by estimating D in the (1)-quarter replicate where it is estimable and forming a strangely weighted average with the estimates from the two half replicates where D is estimable.

$56\hat{\delta}$ = 8δ (I=AB=AC=BC) + 24δ (I=-ABCD) + 24δ (I=ABC) = 56δ

8δ, I=AB=AC=BC:

-2		2	
	X		
	X		
X	-2		2

24δ, I=-ABCD:

	-3	3	
-3	X		3
-3	X		3
X	-3	3	

24δ, I=ABC:

	-3		3
-3	X	3	
-3	X	3	
X	-3		3

56δ:

-2	-6	5	3
-6	X	3	3
-6	X	3	3
X	-8	3	5

On the other hand, I is not estimable from any (1)-quarter replicate; I can be estimated by estimating I in any of the three (1)-quarter replicates where it is aliased with one of the three effects estimable only from I = D, correcting for the bias, and forming a strangely weighted average with the estimates from the two half replicates where I is estimable. For example, if the (1)-quarter replicate I = AB = -ACD = -BCD is used, the LS estimate can be obtained as follows.

$56\hat{\mu} =$

$8(\mu+\alpha\beta)$

2			2
	X		
	X		
X	2	2	

I=AB=-ACD=-BCD

+

$-8\alpha\beta$

		-1	-1
	X	1	1
	X	1	1
X		-1	-1

I=D

+

24μ

	3	3	
3	X		3
3	X		3
X	3	3	

I=-ABCD

+

24μ

	3		3
3	X	3	
3	X	3	
X	3		3

I=ABC

=

56μ

2	6	2	4
6	X	4	4
6	X	4	4
X	8	4	2

A, B, C, AD, BD, and CD are each estimable from a (1)-quarter replicate contained in the design, and each can be estimated by forming a strangely weighted average of the estimate from that (1)-quarter replicate, the estimate from a half replicate, and <u>an estimate of zero</u> (the estimate of -ABC-ABCD from I = D). Those LS estimates follow.

$56\hat{\alpha} =$

16α

-4			
4	X		
	X		-4
X			4

I=BD=CD=BC

+

40α

	-5	-5	
5	X		5
-5	X		-5
X	5	5	

I=-ABCD

+

0

		1	-1
	X	-1	1
	X	-1	1
X		1	-1

I=D

=

56α

-4	-5	-4	-1
9	X	-1	6
-5	X	-1	-8
X	5	6	3

$56\hat{\beta} =$

16β

-4			
	X		-4
4	X		
X			4

I=AC=AD=BC

+

40β

	-5	-5	
-5	X		-5
5	X		5
X	5	5	

I=-ABCD

+

0

		1	-1
	X	-1	1
	X	-1	1
X		1	-1

I=D

=

56β

-4	-5	-4	-1
-5	X	-1	-8
9	X	-1	6
X	5	6	3

$56\hat{\gamma}$ =

16γ

-4	4		
	X		
	X		
X		-4	4

I=AB=AD=BD

\+

40γ

	5	-5	
-5	X		5
-5	X		5
X	5	-5	

I=-ABCD

\+

0

		1	-1
	X	-1	1
	X	-1	1
X		1	-1

I=D

=

56γ

-4	9	-4	-1
-5	X	-1	6
-5	X	-1	6
X	5	-8	3

$56\widehat{\alpha\delta}$ =

$16\alpha\delta$

4		-4	
-4	X	4	
	X		
X			

I=-B=-C=BC

\+

$40\alpha\delta$

	5		-5
-5	X	5	
5	X	-5	
X	-5		5

I=ABC

\+

0

		1	-1
	X	-1	1
	X	-1	1
X		1	-1

I=D

=

$56\alpha\delta$

4	5	-3	-6
-9	X	8	1
5	X	-6	1
X	-5	1	4

$56\widehat{\beta\delta}$ =

$16\beta\delta$

4		-4	
	X		
-4	X	4	
X			

I=-A=-C=AC

\+

$40\beta\delta$

	5		-5
5	X	-5	
-5	X	5	
X	-5		5

I=ABC

\+

0

		1	-1
	X	-1	1
	X	-1	1
X		1	-1

I=D

=

$56\beta\delta$

4	5	-3	-6
5	X	-6	1
-9	X	8	1
X	-5	1	4

$56\widehat{\gamma\delta}$ =

$16\gamma\delta$

4	-4	-4	4
	X		
	X		
X			

I=-A=-B=AB

\+

$40\gamma\delta$

	-5		5
5	X	-5	
5	X	-5	
X	-5		5

I=ABC

\+

0

		1	-1
	X	-1	1
	X	-1	1
X		1	-1

I=D

=

$56\gamma\delta$

4	-9	-3	8
5	X	-6	1
5	X	-6	1
X	-5	1	4

Finally, none of AB, AC, and AD is estimable from a (1)-quarter replicate, but each can be estimated by forming a strangely weighted average of the estimate from I = D (where each is estimable), an estimate from I = ABC (where each is aliased with a main effect estimable from a (1)-quarter replicate), and an estimate from I = -ABCD (where each is aliased with a 2-factor interaction estimable from a (1)-quarter replicate), then correcting for bias using estimates from appropriate (1)-quarter replicates. The LS estimates follow.

$56\widehat{\alpha\beta}$ =

$40\alpha\beta$

		5	5
	X	-5	-5
	X	-5	-5
X		5	5

I=D

\+ $8(\alpha\beta+\gamma)$

	1		1
-1	X	-1	
-1	X	-1	
X	1		1

I=ABC

\+ -8γ

2	-2		
	X		
	X		
X		2	-2

I=AB=AD=BD

\+ $8(\alpha\beta-\gamma\delta)$

	1	1	
-1	X		-1
-1	X		-1
X	1	1	

I=-ABCD

\+ $8\gamma\delta$

2	-2	-2	2
	X		
	X		
X			

I=-A=-B=AB

= $56\alpha\beta$

4	-2	4	8
-2	X	-6	-6
-2	X	-6	-6
X	2	8	4

$56\widehat{\alpha\gamma}$ =

$40\alpha\gamma$

		5	-5
	X	-5	5
	X	5	-5
X		-5	5

I=D

\+ $8(\alpha\gamma+\beta)$

	-1		-1
-1	X	-1	
1	X	1	
X	1		1

I=ABC

\+ -8β

2			
	X		2
-2	X		
X			-2

I=AC=AD=CD

\+ $8(\alpha\gamma-\beta\delta)$

	-1	1	
-1	X		1
1	X		-1
X	1	-1	

I=-ABCD

\+ $8\beta\delta$

2		-2	
	X		
-2	X	2	
X			

I=-A=-C=AC

= $56\alpha\gamma$

4	-2	4	-6
-2	X	-6	8
-2	X	8	-6
X	2	-6	4

$56\widehat{\beta\gamma}=$

$40\beta\gamma$

		5	-5
	X	5	-5
	X	-5	5
X		-5	5

I=D

+

$8(\beta\gamma+\alpha)$

	-1		-1
1	X	1	
-1	X	-1	
X	1		1

I=ABC

+

-8α

2			
-2	X		
	X		2
X			-2

I=BC=BD=CD

+

$8(\beta\gamma-\alpha\delta)$

	-1	1	
1	X		-1
-1	X		1
X	1	-1	

I=-ABCD

+

$8\alpha\delta$

2		-2	
-2	X	2	
	X		
X			

I=-B=-C=BC

=

$56\beta\gamma$

4	-2	4	-6
-2	X	8	-6
-2	X	-6	8
X	2	-6	4

Of course, all these LS estimates improved on the LS estimates which could be obtained if the entire quarter replicate I = -D = -ABC = ABCD were deleted to obtain a three-quarter replicate. For D and I the variance of the LS estimate decreased from 3/32 = 0.094 to 5/56 = 0.089, and for each of the other effects the variance of the LS estimate decreased from 1/8 = 0.125 to 3/28 = 0.107. Unfortunately, the structural decompositions of LS estimates obtained in case 4 were not nearly so appealing as those obtained for cases 1 and 2. They bore little resemblance to the decompositions in cases 1 and 2, the ad hoc weights needed for the averages had no obvious justification, and the estimates of zero needed for six of the estimates were counter to intuition. Possibly there are some other decompositions for case 4 which resemble more closely the decompositions for cases 1 and 2, or possibly there is some unifying principle which has been missed.

IV. DESIGNS OBTAINED BY DELETING FOUR POINTS FROM THE 2^4 DESIGN. Four points can be deleted from the 2^4 design by deleting a quarter replicate or by deleting a set of four points which do not form a quarter replicate. As discussed in Section II, there are essentially six ways to delete a quarter replicate, and four of those leave a design which is not of resolution V. The sledgehammer approach was used to examine the ways that a set of four points which do not form a quarter replicate can be deleted: appropriate points different from (1) were deleted one at a time from the six cases in Table 1, and the resulting matrix $\tilde{R}'\tilde{R}$ was checked for singularity. None of the exhaustive list of designs checked (Table 3) yielded a singular matrix $\tilde{R}'\tilde{R}$, which shows that the only way to delete four points from the 2^4 design and fail to have a resulting design of resolution V is to consider one of the three-quarter replicates which is not of resolution V. An attempt was

TABLE 3.--Designs Derived by Deleting an Additional Point Other That (1) From the Designs in Table 1 (All Are of Resolution V)

Case	Basic Points Deleted	Additional Point Deleted	Case	Basic Points Deleted	Additional Point Deleted
1-1	a, b, ab	c	4-1	ab, ac, bc	a
1-2		ac	4-2		d
1-3		cd	4-3		ad
1-4		abc	4-4		abc
1-5		acd	4-5		abd
1-6		abcd	4-6		abcd
2-1	a, bc, abc	b	5-1	ab, cd, abcd	a
2-2		d	5-2		ac
2-3		ab	5-3		abc
2-4		ad			
2-5		bd			
2-6		abd			
2-7		bcd			
2-8		abcd			
3-1	a, bcd, abcd	b	6-1	ab, abc, bcd	a
3-2		ab	6-2		c
3-3		bc	6-3		ac
3-4		abc	6-4		cd
			6-5		abc
			6-6		abcd

made to describe for a few of these designs the structure of the LS estimates obtained, but the attempt was discontinued without success when it appeared as anticipated that the LS estimates from these designs were even more obstinate than the three-point cases considered in detail. Table 4 gives the variances of the LS estimates for each case in Table 3; these were obtained as the diagonal entries of $(\underset{\sim}{R}'\underset{\sim}{R})^{-1}$. Reclassification of the cases in Table 4 by numbers of estimates having particular variances might lead to a considerable reduction in cases for detailed study, but I have not yet had time to attempt such a reduction.

V. DESIGNS OBTAINED BY DELETING FIVE POINTS FROM THE 2^4 DESIGN. Only a few designs obtained by deleting five points from the 2^4 design were examined. First, the sledgehammer approach was continued for designs derived from three-point case 1. Appropriate points different from (1) were deleted one at a time from cases 1-1 through 1-6 in Table 3 (obviously redundant subcases were excluded from consideration), and the resulting matrix $\underset{\sim}{R}'\underset{\sim}{R}$ was checked for singularity. Most of the resulting designs were of resolution V, and none of the resolution V designs contained a quarter replicate within the points deleted. All six designs which failed to be of

TABLE 4.--Variances of the LS Estimates for Each Design in Table 3

Case	Effect Estimated										
	I	A	B	AB	C	AC	BC	D	AD	BD	CD
1-1	1/4	7/32	7/32	1/8	15/32	1/8	1/8	1/4	7/32	7/32	15/32
1-2	1/4	7/32	7/32	1/8	7/32	1/8	1/8	1/4	7/32	7/32	7/32
1-3	3/8	1/8	1/8	1/8	7/32	7/32	7/32	7/32	7/32	7/32	3/8
1-4	3/8	7/32	7/32	1/8	7/32	1/8	1/8	3/8	7/32	7/32	7/32
1-5	1/4	1/8	1/8	1/8	7/32	7/32	7/32	7/32	7/32	7/32	1/4
1-6	1/4	1/8	1/8	1/8	15/32	7/32	7/32	15/32	7/32	7/32	1/4
2-1	1/4	7/32	15/32	1/8	7/32	1/8	1/8	1/4	7/32	15/32	7/32
2-2	1/4	1/8	7/32	7/32	7/32	7/32	1/4	1/8	1/8	15/32	15/32
2-3	1/4	7/32	7/32	1/8	7/32	1/8	1/8	1/4	7/32	7/32	7/32
2-4	1/4	1/8	7/32	7/32	7/32	7/32	1/4	1/8	1/8	7/32	7/32
2-5	1/8	7/64	3/16	7/64	1/4	7/64	3/16	3/16	7/64	1/4	3/16
2-6	1/8	7/64	3/16	7/64	1/4	7/64	3/16	3/16	7/64	1/4	3/16
2-7	1/4	1/8	7/32	7/32	7/32	7/32	1/4	1/8	1/8	7/32	7/32
2-8	1/4	1/8	15/32	7/32	15/32	7/32	1/4	1/8	1/8	7/32	7/32
3-1	1/4	1/8	1/8	1/8	7/32	7/32	15/32	7/32	7/32	15/32	1/4
3-2	1/4	1/8	1/8	1/8	7/32	7/32	7/32	7/32	7/32	7/32	1/4
3-3	3/8	1/8	7/32	7/32	7/32	7/32	3/8	1/8	1/8	7/32	7/32
3-4	1/4	1/8	7/32	7/32	7/32	7/32	1/4	1/8	1/8	7/32	7/32
4-1	1/4	15/32	7/32	1/8	7/32	1/8	1/8	1/4	15/32	7/32	7/32
4-2	3/32	1/8	1/8	1/8	1/8	1/8	1/8	15/128	15/128	15/128	15/128
4-3	1/8	7/64	7/64	3/16	7/64	3/16	1/4	7/64	1/4	3/16	3/16
4-4	3/8	7/32	7/32	1/8	7/32	1/8	1/8	3/8	7/32	7/32	7/32
4-5	1/8	3/16	3/16	1/4	1/4	3/16	3/16	7/64	7/64	7/64	7/64
4-6	3/32	15/128	15/128	1/8	15/128	1/8	1/8	15/128	1/8	1/8	1/8
5-1	1/4	1/8	1/8	1/8	7/32	15/32	7/32	7/32	15/32	7/32	1/4
5-2	1/8	7/64	7/64	3/16	7/64	1/4	3/16	7/64	3/16	1/4	3/16
5-3	1/4	7/32	7/32	1/4	1/8	7/32	7/32	1/8	7/32	7/32	1/8
6-1	1/4	1/8	1/8	1/8	7/32	15/32	7/32	7/32	15/32	7/32	1/4
6-2	3/32	1/8	1/8	1/8	15/128	15/128	15/128	1/8	1/8	1/8	15/128
6-3	1/8	3/8	1/4	3/8	3/8	1/4	3/8	7/64	7/64	7/64	7/64
6-4	3/8	1/8	1/8	1/8	7/32	7/32	7/32	7/32	7/32	7/32	3/8
6-5	1/8	3/8	3/8	1/4	7/64	7/64	7/64	1/4	3/16	3/16	7/64
6-6	1/4	1/8	1/8	1/8	15/32	7/32	7/32	15/32	7/32	7/32	1/4

resolution V contained a quarter replicate within the points deleted: for four of these designs the deleted quarter replicate defined one of the three-quarter replicates not of resolution V (that is, the defining contrast for the deleted quarter replicate corresponded to one of cases 1, 2, 3, or 5 in Table 1); however, for two of these designs the deleted quarter replicate defined a three-quarter replicate of resolution V (corresponding to cases 4 or 6 in Table 1). Table 5 lists the designs checked and gives for all designs which were not of resolution V the defining contrast for the quarter replicate contained within the points deleted as well as the corresponding case number from Table 1.

TABLE 5.--Designs Derived by Deleting an Additional Point Other Than (1) From Cases 1-1 Through 1-6 in Table 3

Case	Additional Point Deleted	Resolution V	Defining Contrast for Quarter Replicate Contained Within Points Deleted	Corresponding Case Number From Table 1
1-1-1	d	yes		
1-1-2	ac	no	I = -D = -BC = BCD	Case 2
1-1-3	ad	yes		
1-1-4	cd	yes		
1-1-5	abc	no	I = -D = ABC = -ABCD	Case 4
1-1-6	abd	yes		
1-1-7	acd	yes		
1-1-8	abcd	yes		
1-2-1	ad	yes		
1-2-2	cd	yes		
1-2-3	abc	no	I = A = -D = AD	Case 1
1-2-4	abd	yes		
1-2-5	acd	yes		
1-2-6	abcd	yes		
1-3-1	abc	yes		
1-3-2	acd	no	I = -BC = -BD = CD	Case 3
1-3-3	abcd	no	I = CD = ABC = ABD	Case 6
1-4-1	abd	yes		
1-4-2	acd	yes		
1-4-3	abcd	yes		
1-5-1	abcd	no	I = A = CD = ACD	Case 2

Based on the results for designs derived from case 1, twelve final designs were checked. These were the designs resulting when (1) was deleted from cases 4-1 through 4-6 and 6-1 through 6-6 in Table 3. These final checks were equivalent to examining both the resolution V three-quarter replicate defined by I = -D = -ABC = ABCD (case 4 in Table 1) and the resolution V three-quarter replicate defined by I = CD = -ABC = -ABD (case 6 in Table 1) to determine the result of deleting each remaining point one at a time. The TPM for each of these three-quarter replicates is given below with points in the defining quarter replicate labelled by "X" and each remaining point labelled by "+" (if deleting that point yields a resolution V design) or "-" (if deleting that point fails to yield a resolution V design).

X	-	+	+
-	X	+	+
-	X	+	+
X	-	+	+

I=-D=-ABC=ABCD

X	+	+	-
-	+	+	X
-	+	+	X
X	+	+	-

I=CD=-ABC=-ABD

In each case, the designs which fail to be of resolution V are those where all five points are contained in a half replicate defined by a main effect or a 2-factor interaction.

From the cases studied, a promising conjecture for designs resulting when five points are deleted from the 2^4 design in that any such design is of resolution V if either of the following holds:

i) The points deleted contain no quarter replicate.

ii) The points deleted contain a quarter replicate having exactly one main effect or exactly one 2-factor interaction (but not both) in its defining contrast and not all five points deleted are contained in any half replicate defined by a main effect or a 2-factor interaction.

Unfortunately, I have neither been able to prove this conjecture algebraically nor had the time to apply the sledgehammer approach to designs derived from cases 2, 3, and 5 in Table 1.

VI. SUMMARY AND QUESTIONS. In this paper, laborious computational methods have been used to investigate the designs resulting when less than six points are deleted from the 2^4 design. Clearly, deleting six or more points from the 2^4 design leaves a remaining design which is not of resolution V. I had anticipated that there would be many ways to delete as few as three points from the 2^4 design and fail to have a remaining design of resolution V. Instead, this paper showed that there is no way to delete three points from the 2^4 design and fail to have a design of resolution V. Furthermore, there are many ways to delete four or five points from the 2^4 design and retain a remaining design of resolution V. (I conjecture that most designs obtained be deleting four or five points at random from the 2^4 design are of resolution V, but I have not yet attempted the combinatorics involved.) In the cases where the least squares estimates were studied in detail, however, the least squares estimates obstinately resisted unified structural representation.

Like many Army statisticians, I am somewhat isolated from the statistical community: there are few convenient persons with whom I can discuss statistical problems, and I lack ready access to an extensive statistical library. Fortunately, this Conference provides an opportunity in the form of Clinical Papers for statisticians like myself to present incompletely solved statistical problems for expert discussion.

I would like better to understand the impact of missing data on estimability and variance in 2^n factorials and 2^{n-k} fractional factorials. In the context of this paper I have five questions:

1. Are there blunders in this paper which invalidate some or all of the results?

2. Are there known results which encompass some or all of the results of this paper?

3. Are there (hopefully generalizable) methods which might yield the results of this paper more easily?

4. Have I overlooked some unifying principle which ties together the results of this paper?

5. Are the structural decompositions of least squares estimates attempted for the three-point cases in this paper

    i) the wrong ones, in the sense that other more usable versions exist?
    ii) naive, in the sense that one should not hope to explain simply the solution of the normal equations for fairly arbitrary designs?

VII. REFERENCES.

John, Peter W. M. (1971). Statistical Design and Analysis of Experiments. New York: Macmillan.

Russell, Carl T. (1981). "The Potential Utility of Crossing a Fractional Factorial with a Full Factorial in the Design of Field Tests." Raleigh: US Army Research Office Report No. 81-2, Proceedings of the Twenty-Sixth Conference on Design of Experiments, 335-346.

Russell, Carl T. (1982). "Selling a Complicated Experimental Design to the Field Test Operator." Raleigh: US Army Research Office Report No. 82-2, Proceedings of the Twenty-Seventh Conference on Design of Experiments, 293-308.

SENSITIVITY AND ASYMPTOTIC PROPERTIES OF BAYESIAN RELIABILITY ESTIMATES

Bernard Harris and Andrew P. Soms

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

ABSTRACT

In decision theoretic approaches to estimation problems, loss functions of the type $L(\theta,\hat{\theta}) = |\theta-\hat{\theta}|^{\alpha}$, $\alpha > 0$ are often employed; often $\alpha = 2$. In many applications of reliability and life testing, such loss functions are inappropriate. Alternative loss functions which appear to be more suited to the intended application are proposed and Bayesian estimates of the exponential parameter are obtained for these.

Asymptotic expansions of such estimators are given and compared with estimators given in the literature.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041 and the Office of Naval Research under Contract No. N00014-79-C-0321.

SENSITIVITY AND ASYMPTOTIC PROPERTIES OF BAYESIAN RELIABILITY ESTIMATES

Bernard Harris
&
Andrew P. Soms

1. INTRODUCTION

In decision theoretic approaches to estimation problems, the error of estimation is typically measured by loss functions of the form

$$L(\theta,\hat{\theta}) = |\theta-\hat{\theta}|^{\nu}, \quad \nu > 0 ,$$

where θ is the unknown parameter and $\hat{\theta}$ is the corresponding estimator. However, as the following heuristic argument will demonstrate, such a loss function is oten inappropriate for many applications of reliability and life testing.

Let R, $0 \le R < 1$, be the reliability of some device. If $\hat{R}$ is an estimator of R and the true value of R is .5, then an error of the magnitude $|\hat{R}-R| = .1$ would be unlikely to affect any administrative decision concerning the feasibility of the device, since a device whose reliability is between .4 and .6 would not usually be regarded as satisfactory. On the other hand, if $R = .90$, then one device in ten fails and the estimate $\hat{R} = .99$ would suggest that only one device in 100 would fail. Thus, one might be inclined to conclude that one-tenth as many replacements were needed as were in fact required and consequently seriously misjudge maintenance and replacement costs. A similar but opposite error in judgement occurs when $R = .99$ and $\hat{R} = .90$. Consequently, it appears to be desirable to concentrate on errors of estimation for R close to unity.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041 and the Office of Naval Research under Contract No. N00014-79-C-0321.

Considerations such as those described above suggest that loss functions suitable for reliability applications might have forms such as

$$L(R,\hat{R}) = |R-\hat{R}|^{\nu}/(1-R)^{\kappa}, \quad 0 \le R < 1, \; 0 \le \hat{R} < 1, \quad \nu > 0, \; \kappa > 0,$$

or

$$L(R,\hat{R}) = \left|\frac{1}{1-R} - \frac{1}{1-\hat{R}}\right|^{\nu}, \quad 0 \le R < 1, \quad 0 \le \hat{R} < 1, \quad \nu > 0.$$

In order to study the effects of using such loss functions, we consider a Bayesian model for exponential life testing and examine the asymptotic behavior of the estimates thus obtained.

Specifically let $X_1, X_2, \ldots, X_N$ be independent identically distributed observations from an exponential distribution with probability density function

$$f(x;\theta) = \theta e^{-\theta x}, \quad x > 0, \quad \theta > 0. \tag{1}$$

Then the reliability R_T is defined as

$$R_T = e^{-\theta T}, \tag{2}$$

where $T > 0$ is a constant assigned in advance of the experiment and known as the <u>mission time</u>. With no loss of generality, we can set $T = 1$, since this corresponds to selecting the scale with respect to which the observations $X_1, X_2, \ldots, X_N$ are measured.

In Section 2, we obtain Bayesian estimators for specific loss functions of the type described previously, when R has a beta distribution prior. We compare this with the Bayesian estimator of R that is obtained by assigning a gamma prior to θ, a choice of Bayesian model quite often found in the literature. Obviously, since R is a function of θ, this induces a prior on R; however, the priors assigned to θ in such studies do not induce the beta family of priors on R.

In Section 3, the asymptotic behavior of these estimators is given. These asymptotic expressions facilitate studying the sensitivity of the estimators to the changes in the loss functions.

Section 4 is devoted to comparing the results obtained herein with previous estimation techniques.

Several appendices which provide the technical details needed to establish the existence of the estimators and the calculation of the asymptotic expansions are included at the end of the paper.

2. BAYESIAN ESTIMATION FOR UNCENSORED AND TYPE II CENSORED DATA

Let $X_{(1)} \le X_{(2)} \le \cdots \le X_{(N)}$ be the order statistics from a random sample disbributed by (1). Assume that only the first n order statistics have been observed, $1 \le n \le N$. Then, it is well-known that the total time on test statistic Y, defined by

$$Y = \sum_{i=1}^{n} X_{(i)} + (N-n)X_{(n)} \tag{3}$$

is a sufficient statistic and its probability density function is

$$f_Y(y;\theta) = \theta^n e^{-\theta y} y^{n-1}/\Gamma(n), \quad y > 0, \theta > 0. \tag{4}$$

To represent (4) in terms of the reliability R, we reparametrize, obtaining

$$f_Y(y;R) = (-\log R)^n R^y y^{n-1}/\Gamma(n), \quad 0 < R < 1, \quad y > 0. \tag{5}$$

Let

$$\tau(R;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} R^{\alpha-1} (1-R)^{\beta-1}, \quad 0 < R < 1, \alpha > 0, \beta > 0, \tag{6}$$

be the prior distribution on R. Then the joint distribution of R and Y is given by

$$f(y,R;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(n)} R^{\alpha+y-1}(1-R)^{\beta-1}(-\log R)^n y^{n-1}, \tag{7}$$

$0 < R < 1, \quad 0 < y, \alpha > 0, \quad \beta > 0, \quad 1 \le n \le N$.

We now obtain the marginal distribution of Y. Expanding $(1-R)^{\beta-1}$ in a binomial series, we obtain

$$f_1(y;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(n)} y^{n-1} \int_0^1 R^{\alpha+y-1}(-\log R)^n \sum_{i=0}^{\infty} \binom{\beta-1}{i}(-1)^i R^i dR. \tag{8}$$

In (8), it is understood that if β is a positive integer, the series terminates; that is, all terms with $i > \beta - 1$ are zero.

The interchange of summation and integration in (8) is justifiable. Hence, we can write (8) as

$$f_1(y;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(n)} y^{n-1} \sum_{i=0}^{\infty} (-1)^i \binom{\beta-1}{i} \int_0^1 R^{\alpha+y+i-1}(-\log R)^n \, dR$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(n)} y^{n-1} \sum_{i=0}^{\infty} (-1)^i \binom{\beta-1}{i} \int_0^{\infty} \theta^n e^{-\theta(\alpha+y+i)} d\theta$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(n)} y^{n-1} \sum_{i=0}^{\infty} (-1)^i \binom{\beta-1}{i} \frac{\Gamma(n+1)}{(\alpha+y+i)^{n+1}}$$

$$= \frac{n\,\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{n-1} \sum_{i=0}^{\infty} (-1)^i \binom{\beta-1}{i} / (\alpha+y+i)^{n+1}. \tag{9}$$

In particular, the integration given in (9) is valid over a larger range of the parameters than specified in (7). Specifically

$$\int_0^1 R^{\alpha+y-1} (1-R)^{\beta-1}(-\log R)^n \, dR = \sum_{i=0}^{\infty} (-1)^i \binom{\beta-1}{i} \frac{\Gamma(n+1)}{(\alpha+y+i)^{n+1}}, \tag{10}$$

whenever $\alpha + y > 0$, $\beta + n > 0$. These facts will be subsequently utilized. The reader is referred to Appendix I for various details relevant to these remarks and calculations.

Some particular cases of (10) are worth noting. If $\beta = 0$, we have

$$\int_0^1 R^{\alpha+y-1} (1-R)^{-1}(-\log R)^n \, dR = \Gamma(n+1)\zeta(n+1,\alpha+y) \tag{11}$$

where $\zeta(r,s)$ denotes the generalized Riemann zeta function and $\zeta(r.1) = \zeta(r)$ is the Riemann zeta function.

Combining the results of (7) and (9), we can write down the posterior distribution of R given $Y = y$ as

$$f(R|Y=y) = \frac{R^{\alpha+y-1}\ (1-R)^{\beta-1}(-\log R)^n}{\Gamma(n+1)\ \sum_{i=0}^{\infty}(-1)^i\binom{\beta-1}{i}/(\alpha+y+i)^{n+1}} \quad . \tag{12}$$

This enables us to calculate the Bayes estimators for a number of the loss functions of the type described previously.

For $L_1(R,\hat{R}) = (R-\hat{R})^2$, the Bayes estimator

$$\hat{R}_{\alpha,\beta}(L_1) = E(R|Y=y) = \frac{\int_0^1 R^{\alpha+y}\ (1-R)^{\beta-1}(-\log R)^n\, dR}{\int_0^1 R^{\alpha+y-1}(1-R)^{\beta-1}(-\log R)^n\, dR} \tag{13}$$

or equivalently,

$$\hat{R}_{\alpha,\beta}(L_1) = \frac{\sum_{i=0}^{\infty}(-1)^i\binom{\beta-1}{i}/(\alpha+y+i+1)^{n+1}}{\sum_{i=0}^{\infty}(-1)^i\binom{\beta-1}{i}/(\alpha+y+i)^{n+1}} \quad . \tag{14}$$

In particular, for $\beta = 1$

$$\hat{R}_{\alpha,\beta}(L_1) = \left(1 - \frac{1}{\alpha+y+1}\right)^{n+1} \quad . \tag{15}$$

Similarly, for $L_{2,\nu}(R,\hat{R}) = (R-\hat{R})^2/(1-R)^{\nu}$, $\nu > 0$, the Bayes estimator is

$$\hat{R}_{\alpha,\beta,\nu}(L_2) = \frac{E\{R/(1-R)^{\nu}|Y=y\}}{E\{1/(1-R)^{\nu}|Y=y\}} \quad , \tag{16}$$

or

$$\hat{R}_{\alpha,\beta,\nu}(L_2) \quad \frac{\sum_{i=0}^{\infty} (-1)^i \binom{\beta-\nu-1}{i} / (\alpha+y+i+1)^{n+1}}{\sum_{i=0}^{\infty} (-1)^i \binom{\beta-\nu-1}{i} / (\alpha+y+i)^{n+1}} , \quad \beta-\nu+n>0 \quad . \tag{17}$$

Finally, we consider the loss function $L_3(R,\hat{R}) = ((1-R)^{-1}-(1-\hat{R})^{-1})^2$. Here the Bayes estimator is given by

$$\hat{R}_{\alpha,\beta}(L_3) = 1 - (E\{(1-R)^{-1}|Y=y\})^{-1} \quad . \tag{18}$$

Since

$$E\{(1-R)^{-1}|Y-y\} = \frac{\int_0^1 R^{\alpha+y-1}(1-R)^{\beta-2}\,(-\log R)^n\,dR}{\Gamma(n+1)\,\sum_{i=0}^{\infty}(-1)^i\binom{\beta-1}{i}/(\alpha+y+i)^{n+1}} , \tag{19}$$

and since the integral (19) converges whenever $\beta + n - 1 > 0$, we have

$$E\{(1-R)^{-1}|Y-y\} = \frac{\sum_{i=0}^{\infty} (-1)^i\binom{\beta-2}{i}/(\alpha+y+i)^{n+1}}{\sum_{i=0}^{\infty} (-1)^i\binom{\beta-1}{i}/(\alpha+y+i)^{n+1}} \quad . \tag{20}$$

Then, we can write

$$\hat{R}_{\alpha,\beta}(L_3) = \frac{\sum_{i=0}^{\infty}(-1)^i\binom{\beta-2}{i}/(\alpha+y+i)^{n+1} - \sum_{i=0}^{\infty}(-1)^i\binom{\beta-1}{i}/(\alpha+y+i)^{n+1}}{\sum_{i=0}^{\infty}(-1)^i\binom{\beta-2}{i}/(\alpha+y+i)^{n+1}} \quad . \tag{21}$$

Applying the Pascal triangle identity, we have

$$\hat{R}_{\alpha,\beta}(L_3) = \frac{\sum_{i=0}^{\infty} (-1)^{i-1}\binom{\beta-2}{i-1}/(\alpha+y+i)^{n+1}}{\sum_{i=0}^{\infty} (-1)^{i}\binom{\beta-2}{i}/(\alpha+y+i)^{n+1}}$$

$$= \frac{\sum_{i=0}^{\infty} (-1)^{i}\binom{\beta-2}{i}/(\alpha+y+i+1)^{n+1}}{\sum_{i=0}^{\infty} (-1)^{i}\binom{\beta-2}{i}/(\alpha+y+i)^{n+1}} \quad . \tag{22}$$

Comparing (14), (17) and (22), we have

$$\hat{R}_{\alpha,\beta,\nu}(L_2) = \hat{R}_{\alpha,\beta-\nu}(L_1) \tag{23}$$

and

$$\hat{R}_{\alpha,\beta}(L_3) = \hat{R}_{\alpha,\beta-1}(L_1) \ . \tag{24}$$

Thus, we have shown that the three families of estimators have the same form. Hence, it is possible to study all of them simultaneously and a single asymptotic expansion, given in the appendix, permits simultaneous analysis of their asymptotic properties.

One caution should be noted. We have listed the posterior means whenever the posterior means exist. However, in some of these instances the Bayes risk will be infinite. Specifically, for the loss function L_2, the Bayes risk exists whenever $\beta + n - 2 > 0$.

In the statistical literature, one frequently finds the following Bayes estimator of R (see for example, N. R. Mann, R. E. Schafer, N. D. Singpurwalla [6], p. 398). This is the estimator obtained by using the loss function

$L_1(R,\hat{R})$ and assigning a gamma distribution prior to θ, the parameter in the exponential density (1). A brief sketch of the calculation of the estimator follows. The prior is given by

$$\tau(\theta:\gamma,\delta) = \delta^{\gamma}\,\theta^{\gamma-1}\,e^{-\delta\theta}/\Gamma(\gamma)\ , \tag{25}$$

where $\delta,\gamma > 0$ and $\theta > 0$. Then, calculating the posterior distribution of θ, when the data is given by the total time on test statistic (3), it follows that the estimator

$$\hat{R}(\gamma,\delta) = \left(\frac{y+\delta}{y+\delta+1}\right)^{n+\gamma}\ , \tag{26}$$

the conditional expected value of R given $Y = y$. Subsequently, this will be compared with (14), (17) and (22).

<u>Remark</u>. To the authors, one of the more interesting properties of these estimators is the role played by the Riemann zeta function and the generalized Riemann zeta function. From (11), we have that

$$\hat{R}_{\alpha,0}(L_1) = \zeta(n+1,\ \alpha+y+1)\,/\,\zeta(n+1,\alpha+y)$$

$$= \hat{R}_{\alpha,\beta,\beta}(L_2) = \hat{R}_{\alpha,1}(L_3)\ .$$

Further, from (21), we see that

$$\hat{R}_{\alpha,1}(L_3) = (\zeta(n+1,\alpha+y) - (\alpha+y)^{-n-1})\,/\,\zeta(n+1,\alpha+y)$$

and if $\alpha = 1$, $y = 0$,

$$\hat{R}_{1,1}(L_3) = (\zeta(n+1) - 1)\,/\,\zeta(n+1)\ .$$

Specifically, if in addition $n = 1$,

$$\hat{R}_{1,1}(L_3) = (\pi^2/6 - 1)/(\pi^2/6) = 1 - \frac{6}{\pi^2} .$$

We conclude this section with some observations concerning elementary properties of the estimators given by (14), (17) and (22). With no loss of generality, we can examine specifically (13) and (14). From (13), it is immediately evident that $0 < \hat{R} < 1$. We now show that $\hat{R}$ is an increasing function of α (or y) and a decreasing function of β, which one would naturally expect to be the case, given the respective roles played by α and β in the model. Therefore, we calculate

$$\frac{\partial \hat{R}}{\partial \beta} = \frac{\int_0^1 R^{\alpha+y}(1-R)^{\beta-1} \log (1-R)(-\log R)^n \, dR}{\int_0^1 R^{\alpha+y-1}(1-R)^{\beta-1} (-\log R)^n \, dR}$$

$$- \frac{\int_0^1 R^{\alpha+y}(1-R)^{\beta-1} (-\log R)^n dR \int_0^1 S^{\alpha+y-1}(1-S)^{\beta-1} \log (1-S)(-\log S)^n \, dS}{\left(\int_0^1 R^{\alpha+y-1} (1-R)^{\beta-1} (-\log R)^n \, dR\right)^2}$$

Thus, it suffices to consider

$$\int_0^1 R^{\alpha+y} (1-R)^{\beta-1} \log (1-R)(-\log R)^n \, dR \int_0^1 S^{\alpha+y-1}(1-S)^{\beta-1}(-\log S)^n \, dS$$

$$- \int_0^1 R^{\alpha+y} (1-R)^{\beta-1}(-\log R)^n \, dR \int_0^1 S^{\alpha+y-1}(1-S)^{\beta-1}\log(1-S)(-\log S)^n \, dS$$

$$= \int_0^1\int_0^1 R^{\alpha+y} (1-R)^{\beta-1}(-\log R)^n S^{\alpha+y}(1-S)^{\beta-1} (-\log S)^n \left[\frac{\log(1-R)-\log(1-S)}{S}\right] dR\,dS,$$

from which it follows readily that $\partial\hat{R}/\partial\beta < 0$. The verification that $\partial\hat{R}/\partial\alpha > 0$ ($\partial\hat{R}/\partial y > 0$) is completely parallel and is omitted.

3. ASYMPTOTIC COMPARISONS OF ESTIMATORS

From the results of Appendix III (see A.82), we have that

$$\hat{R}_{\alpha,\beta}(L_1) \sim e^{-\frac{1}{\bar{y}} + \frac{1}{n\bar{y}^2}(\alpha+\frac{1}{2}) - \frac{(\beta-1)e^{-1/\bar{y}}}{n\bar{y}^2(1-e^{-1/\bar{y}})}} \left(1 - \frac{1}{n\bar{y}}\right) , \qquad (27)$$

$$\hat{R}_{\alpha,\beta}(L_{2,\nu}) \sim e^{-\frac{1}{\bar{y}} + \frac{1}{n\bar{y}^2}(\alpha+\frac{1}{2}) - \frac{(\beta-\nu-1)e^{-1/\bar{y}}}{n\bar{y}^2(1-e^{-1/\bar{y}})}} \left(1 - \frac{1}{n\bar{y}}\right) , \qquad (28)$$

$$R_{\alpha,\beta}(L_3) \sim e^{-\frac{1}{\bar{y}} + \frac{1}{n\bar{y}^2}(\alpha+\frac{1}{2}) - \frac{(\beta-2)e^{-1/\bar{y}}}{n\bar{y}^2(1-e^{-1/\bar{y}})}} \left(1 - \frac{1}{n\bar{y}}\right) . \qquad (29)$$

As is evident from the above, the differences are small. For large n,

$$\hat{R}_{\alpha,\beta}(L_1)/\hat{R}_{(\alpha,\beta)}(L_{2,\nu}) \sim 1 + \frac{\nu e^{-1/\bar{y}}}{n\bar{y}^2(1-e^{-1/\bar{y}})} \qquad (30)$$

and in general $\hat{R}_{(\alpha,\beta)}(L_i)/\hat{R}_{\alpha,\beta}(L_j) \sim 1 + 0(n^{-1})$, $1 \le i, j \le 3$.

Naturally, since the maximum likelihood estimator is $e^{-1/\bar{y}}$, one expects the estimators above to be asymptotically equivalent to it. Note further that

$$\nu e^{-1/\bar{y}}/\bar{y}^2(1-e^{-1/\bar{y}})$$

is bounded for all $\bar{y} > 0$.

Similarly, for the "traditional" Bayesian estimator (26), the comparable representation is

$$\hat{R}(\gamma,\delta) \sim e^{-\frac{1}{\bar{y}} - \frac{\gamma}{n\bar{y}} + \frac{2\delta+1}{n\bar{y}^2}} \qquad (31)$$

and $\hat{R}(\gamma,\delta)/\hat{R}_{\alpha,\beta}(L_i) = 1 + 0(n^{-1})$, $i = 1,2,3$. Thus, it appears that the differences are about as small as could reasonably be expected.

4. SOME PROPOSED ESTIMATORS OF R

The minimum variance unbiased estimator of R has been studied by E. L. Pugh [8], A. P. Basu [2] and S. Zacks and M. Even [12]. In our notation, this is given by

$$R^* = \begin{cases} (1 - 1/n\bar{y})^{n-1} & n\bar{y} > 1 \\ 0 & n\bar{y} \leq 1 \end{cases} \tag{32}$$

The asymptotic representation for R^* is easily seen to be

$$R^* = e^{-\frac{1}{\bar{y}} + \frac{1}{n\bar{y}} - \frac{1}{2n\bar{y}^2}} (1 + 0(n^{-2})) .$$

We can also adapt the estimators of the exponential parameter given by G. M. El-Sayyad [4] to correct them in a naive manner to reliability estimators. Naturally, the optimality hypotheses used therein no longer apply,

$$R_1^* = \exp\left(-\frac{\{\Gamma(n-\rho)/\Gamma(n-2\rho)\}^{1/\rho}}{n\bar{y}}\right) , \quad \rho > 0 \tag{33}$$

$$= e^{-\frac{1}{\bar{y}} - (\frac{3\rho+1}{2n\bar{y}})} (1 + 0(n-2)) \tag{34}$$

El-Sayyad also provides an argument which leads to the well-known estimator

$$\hat{\theta} = e^{\psi(n)}/n\bar{y} ,$$

where $\psi(n)$ is known as the Psi function or digamma function. Fron this, one deduces

$$\hat{R} = e^{e^{\psi(n)}/n\bar{y}} \tag{35}$$

and from well-known asymptotic properties of $\psi(n)$, (see for example, M. Abramowitz and I. A. Stegun [1]),

$$\hat{R} = e^{\frac{1}{\bar{y}} - \frac{1}{2n\bar{y}}}(1 + O(n^{-2})) . \tag{36}$$

The Psi function arises naturally in estimators as the scale transformation invariant estimator for squared error loss. In this connection see T. S. Ferguson [5].

El-Sayyad also obtained some Bayesian point estimators for the exponential parameter using the gamma prior and some loss functions which are generalizations of the squared error loss function.

In the notation of (26), his estimators for θ are

$$\hat{\theta}_1 = (\delta+n\bar{y})^{-1} \{\Gamma(\gamma+n+\alpha+\beta)/\Gamma(\gamma+n+\beta)\}^{1/\beta} \tag{37}$$

and

$$\hat{\theta}_2 = e^{\psi(n+\gamma)}/(\delta+n\bar{y}), \tag{38}$$

where α,β are nonnegative parameters in the loss functions used by El-Sayyad. These yield estimators of R as follows,

$$\hat{R}_i = e^{-\hat{\theta}_i}, \quad i = 1,2 \tag{39}$$

and asymptotically, we have,

$$\hat{R}_1 = e^{-\frac{1}{\bar{y}} - \frac{\gamma+\alpha}{n\bar{y}} - \frac{(\beta-1)}{2n\bar{y}} + \frac{\delta}{n\bar{y}^2}} (1 + O(n^{-2})), \tag{40}$$

and

$$\hat{R}_2 = e^{-\frac{1}{\bar{y}} - \frac{\gamma}{n\bar{y}} + \frac{1}{2n\bar{y}} + \frac{\delta}{n\bar{y}^2}} (1 + O(n^{-2})) . \tag{41}$$

Various Bayesian estimators for the exponential parameter were suggested by S. K. Bhattacharyya [3]. In one of these, he considered the range of $\tau = 1/\theta$ to be finite and used the uniform prior

$$g(\tau) = \begin{cases} (\beta-\alpha)^{-1}, & 0 < \alpha \le \tau \le \beta , \\ 0 \quad , & \text{otherwise.} \end{cases} \tag{42}$$

He also considered the inverted gamma density

$$g(\tau) = \frac{e^{-\mu/\tau} (\frac{\mu}{\tau})^{\nu+1}}{\mu\ \Gamma(\nu)} \quad , \quad 0 < \tau < \infty, \quad \mu,\nu > 0. \tag{43}$$

Letting

$$\gamma(n,x) = \int_0^x e^{-t}\, t^{n-1}\, dt \ ,$$

the Bayes estimate for (42) is

$$\hat{R}_3 = \frac{\gamma^*(n-1,n\bar{y}+1)}{\gamma^*(n-1,n\bar{y})} \frac{1}{(1 + 1/n\bar{y})^{n-1}} \ , \tag{44}$$

where $\gamma^*(n,\nu) = \gamma(n,\frac{\nu}{\alpha}) - \gamma\ (n, \frac{\nu}{\beta})$.

For the prior density given by (43), the Bayesian estimator was snown to be

$$\hat{R}_4 = (1 + \frac{1}{n\bar{y}+\mu})^{-n-\nu} \quad , \tag{45}$$

which asymptotically behaves like

$$\hat{R}_4 = e^{-\frac{1}{\bar{y}} - \frac{\mu}{n\bar{y}^2} - \frac{1}{2n\bar{y}^2} + \frac{\nu}{n\bar{y}}} \ (1 + 0(n^{-2})). \tag{46}$$

He also calculated the Bayesian estimator for the exponential prior obtaining,

$$\hat{R}_5 = \left(K_{n-1}\left(2\sqrt{\frac{n\bar{y}+1}{\lambda}}\right) / K_{n-1}\left(2\sqrt{\frac{n\bar{y}}{\lambda}}\right)\right) \left(1 + \frac{1}{n\bar{y}}\right)^{-\frac{(n-1)}{2}} , \qquad (47)$$

where the prior is given by

$$g(\tau) = \lambda^{-1} e^{-\tau/\lambda} , \qquad 0 < \tau < \infty , \quad \lambda > 0 . \qquad (48)$$

It can be shown that (47) may be approximated by

$$\hat{R}_5 \sim \left(1 + \frac{1}{n\bar{y}}\right)^{-(n-1)} \qquad (49)$$

for large n. Formula 9.7.8 p. 378 in M. Abramowitz and I. A. Stegun [1] may be used to obtain more detailed asymptotic information.

In the papers by S. K. Sinha and I. Guttman [10], [11] the improper prior

$$g(\tau) = \tau^{-1}, \quad 0 < \tau < \infty \qquad (50)$$

is assigned to $\tau = 1/\theta$, obtaining as the Bayes estimator

$$\hat{R} = (1 + 1/n\bar{y})^{-n} , \qquad (51)$$

which yields the asymptotic expression

$$\hat{R} = e^{-\frac{1}{\bar{y}} - \frac{1}{2n\bar{y}^2}} (1 + O(n^{-2})) . \qquad (52)$$

V. M. Rao Tummala and P. T. Sathe [9] employed gamma priors and obtained "minimum expected loss estimators" for the reliability. They compare these estimators with the Bayes and maximum likelihood estimators for quadratic loss functions in estimating θ .

APPENDIX I

In this appendix, we study the series given in (9) and establish the remarks following (9) and (10). This can also serve as a justification for the interchange of operations used in deriving (9). Specifically, we establish the following

Theorem A.1. Let

$$I(\alpha,\beta,\tau) = \sum_{i=0}^{\infty} (-1)^i \binom{\beta-1}{i} (\alpha+i)^{-(\tau+1)}, \tag{A.1}$$

where $\alpha > 0$, and τ and β are real, and where

$$\binom{\beta-1}{i} = \begin{cases} (\beta-1)(\beta-2)\cdots(\beta-i)/i! & i = 1,2,\ldots \\ 1\,, & i = 0\,. \end{cases} \tag{A.2}$$

Then $I(\alpha,\beta,\tau)$ converges whenever β is a positive integer. If β is not a positive integer, $I(\alpha,\beta,\tau)$ converges if and only if $\beta + \tau > 0$.

Proof. Clearly, if β is a positive integer, the series (A.1) terminates and convergence is trivially verified. Hence assume that β is not a positive integer. Write

$$I(\alpha,\beta,\tau) = \sum_{i<\beta} \binom{\beta-1}{i}(-1)^i(\alpha+i)^{-(\tau+1)}$$
$$+ \sum_{i\geq\beta} \binom{\beta-1}{i}(-1)^i(\alpha+i)^{-(\tau+1)}\,. \tag{A.3}$$

The first sum has a finite (possibly zero) number of terms. Hence, to study convergence, it suffices to restrict attention to the second sum, which we write as

$$\sum_{i\geq\beta} C_\beta \frac{\gamma(\gamma+1)\cdots(\gamma+i-r)(\alpha+i)^{-(\tau+1)}}{i!}, \tag{A.4}$$

where

$$r = 1, \quad C_\beta = 1, \gamma = 1-\beta, \quad \text{if} \quad \beta \leq 0$$

$$r = [\beta]+1, \quad C_\beta = (-1)^{r-1}(\beta-1)\cdots(\beta-r+1), \; \gamma = (r-\beta) \quad \text{if} \quad \beta > 0. \tag{A.5}$$

Further, in (A.4), vacuous products are interpreted as unity and γ is always positive.

We now determine when (A.4) converges and obtain estimates for the tail of (A.4), when it is convergent. Hence, it suffices to consider

$$I_M(\alpha,\gamma,r,\tau) = \sum_{i\geq M} \frac{\gamma(\gamma+1)\cdots(\gamma+i-r)(\alpha+i)^{-(\tau+1)}}{i!}, \tag{A.6}$$

which has all terms positive. Note that the sign of (A.4) is determined solely by C_β.

Now

$$\log\left[\left(\prod_{j=0}^{i-r} (\gamma+j)\right) (\alpha+i)^{-(\tau+1)}/i!\right] \leq$$

$$(\gamma+i-r)\log(\gamma+i-r)-(\gamma+i-r)-\gamma\log\gamma+\gamma+\frac{1}{2}\log[(\gamma)(\gamma+i-r)]$$

$$-(\tau+1)\log(\alpha+i) - i\log i + i - \frac{1}{2}\log i - \frac{1}{2}\log(2\Pi)$$

$$=(\gamma+i-r)[\log i + \log(1+\frac{\gamma-r}{i})] + r - \gamma\log\gamma + \frac{1}{2}\log\gamma + \frac{1}{2}\log i$$

$$+\frac{1}{2}\log(1+\frac{\gamma-r}{i}) - (\tau+1)\log i - (\tau+1)\log(1+\frac{\alpha}{i}) - i\log i$$

$$-\frac{1}{2}\log i - \frac{1}{2}\log(2\Pi). \tag{A.7}$$

For M sufficiently large, we have for all $i \geq M$,

$$\frac{1}{2} < \left(1 + \frac{\gamma-r}{i}\right) \leq e^{\frac{\gamma-r}{i}} < e$$

and

$$1 < \left(1 + \frac{\alpha}{i}\right) < \frac{3}{2} \ .$$

Thus,

$$\left(\prod_{j=0}^{i-r} (\gamma+j)\right) (\alpha+i)^{-(\tau+1)}/i! \leq d(\tau)\left(\frac{\gamma}{2\pi}\right)^{\frac{1}{2}} \gamma^{-\gamma} e^{2\gamma-r+\frac{1}{2}} \, i^{\gamma-r-\tau-1} \ , \quad \text{(A.8)}$$

where

$$d(\tau) = \begin{cases} 1, & \tau \geq -1 \\ \left(\frac{3}{2}\right)^{-(\tau+1)}, & \tau < -1 \end{cases} \ . \quad \text{(A.9)}$$

Therefore, we have shown that

$$I_M \leq c \sum_{i \geq M} i^{\gamma-r-\tau-1} \ , \quad \text{(A.10)}$$

for some positive constant $c = c(\gamma,r,\tau)$. From (A.10), we can immediately observe that I_M converges absolutely whenever $\gamma-r-\tau<0$. Thus, from (A.5), we have that if $\beta > 0$, $I(\alpha,\beta,\tau)$ converges whenever $\beta+\tau>0$. Similarly, if $\beta \leq 0$, the same conclusion holds.

To establish the converse, note that

$$\log \left[\left(\prod_{j=0}^{i-\gamma} (\gamma+j)\right) (\alpha+i)^{-(\tau+1)}/i!\right] \geq$$

$$(\gamma+i-r)[\log i + \log(1 + \frac{\gamma-r}{i})] + r - \gamma \log \gamma + \frac{1}{2} \log \gamma + \frac{1}{2} \log i$$

$$+ \frac{1}{2} \log (1 + \frac{\gamma-r}{i}) + \frac{1}{12}[(\gamma+i-r)^{-1} - \gamma^{-1}] - (\tau+1) \log(1 + \frac{\alpha}{i})$$

$$-(\tau+1) \log i - i \log i - \frac{1}{2} \log i - \frac{1}{2} \log (2\Pi) - \frac{1}{12i} .$$

Now, for M sufficiently large and all $i > M$

$$(\gamma+i-r)^{-1} > 0 \ ,$$

$$(12i)^{-1} \le 1/12$$

and

$$(1 + \frac{\gamma-r}{i}) \ge e^{(\gamma-r)/(i+(\gamma-r))} \ .$$

Consequently,

$$\left(\prod_{j=0}^{i-r} (\gamma+j) \right) (\alpha+i)^{-(\tau+1)}/i! \ge$$

$$(2\pi)^{-\frac{1}{2}} e^{\gamma - \frac{1}{12} - \frac{1}{12\gamma}} i^{\gamma-r-\tau-1} \gamma^{-\gamma + \frac{1}{12}} d_1(\tau) \ , \qquad \text{(A.11)}$$

where $d_1(\tau)$ is defined by (A.9).

Thus

$$I_M \ge k \sum_{i \ge M} i^{\gamma-r-\tau-1} \ , \qquad \text{(A.12)}$$

where $k = k(\gamma,\tau,r)$ is a positive constant. Thus I_M diverges whenever $\gamma-r-\tau \ge 0$ and thus diverges whenever $\beta+\tau \le 0$, establishing the theorem.

Remark. The above analysis permits us to readily estimate the tail of the series whenever $\beta+\tau > 0$, since

$$\sum_{i>M} i^{-(\beta+\tau+1)} < \int_{M-1}^{\infty} x^{-(\beta+\tau+1)} dx = \frac{(M-1)^{-(\beta+\tau)}}{(\beta+\tau)} \qquad \text{(A.13)}$$

APPENDIX II

In this appendix, we obtain a complete asymptotic expansion $(n\to\infty)$ of

$$I_n(\alpha,\beta,y) = \int_0^1 R^{\alpha+y-1}(1-R)^{\beta-1}(-\log R)^n dR \quad ,$$

where $y = n\bar{y}$, $\bar{y} > 0$. In particular, we establish the following theorem.

Theorem A.2. For $\bar{y} > 0$, $\alpha > 0$, as $n \to \infty$, for any $\delta > 0$, $k = 2,3,\ldots$

$$e^{M_n(\theta_n^*)} I_n(\alpha,\beta,y) = \sqrt{2\pi/\gamma_n}\left(\sum_{r=0}^{k} K_{n,k}^{(r)}(\theta_n^*)\mu_r/\gamma_n^{r/2}\right) + O\left(n^{-\frac{k+1}{6}-\frac{1}{2}+\delta}\right) \quad , \tag{A.14}$$

where

$$M_n(\theta) = \theta(\alpha+n\bar{y}) - n\log\theta - (\beta-1)\log(1-e^{-\theta}) \tag{A.15}$$

and θ_n^* is the largest positive root of $M_n'(\theta) = 0$. Further,

$$\gamma_n = n\left[\frac{1}{\theta_n^{*2}} - \frac{\beta-1}{n}\left(\left(1-e^{-\theta_n^*}\right)^{-2} - \left(1-e^{-\theta_n^*}\right)^{-1}\right)\right] \quad , \tag{A.16}$$

$$\mu_r = \begin{cases} r!/(\frac{r}{2})!\ 2^{r/2}, & r \text{ even} \\ 0 \quad , & r \text{ odd} \end{cases} \tag{A.17}$$

and

$$K_{nk}^{(r)}(\theta) = \Sigma(-1)^{\Sigma\ell_i} M_n^{(3)}(\theta)^{\ell_3} \cdots\cdots M_n^{(k)}(\theta)^{\ell_k}/\ell_3!\cdots\ell_k!\ 3!^{\ell_3}\cdots k!^{\ell_k} \quad , \tag{A.18}$$

where the sum extends over all $\ell_3,\ldots,\ell_k \geq 0$ with $\sum_{j=3}^{k} j\ell_j = r$.

In order to prove Theorem A.2, it is desirable to introduce a number of preliminary lemmas.

Lemma A.1. For $j = 1,2,\ldots,$

$$M_n^{(j)}(\theta) = (\alpha+n\bar{y})\delta_{1j} + \frac{(-1)^j(j-1)!n}{\theta^j} - (\beta-1)\sum_{\ell=0}^{j} b_{\ell j}(1-e^{-\theta})^{-\ell} \quad , \tag{A.19}$$

where $\delta_{11} = 1$, $\delta_{1j} = 0$, $j \neq 1$ and

$$\begin{cases} b_{01} = -1,\ b_{11} = 1,\ b_{\ell j} = 0,\ \ell > j \text{ or } \ell < 0 , \\ b_{\ell j} = \ell b_{\ell,j-1} - (\ell-1)b_{\ell-1,j-1},\ 0 < \ell < j,\ j > 2 . \end{cases} \tag{A.20}$$

Alternatively, we can write

$$M_n^{(j)}(\theta) = (\alpha+n\bar{y})\delta_{1j} + \frac{(-1)^j(j-1)!n}{\theta^j} - (\beta-1)\sum_{\ell=1}^{j} a_{\ell j}e^{-\ell\theta}(1-e^{-\theta})^{-j} , \tag{A.21}$$

where

$$\begin{cases} a_{11} = 1,\ a_{0j} = 0,\ a_{jj} = 0,\ j > 1,\ a_{\ell j} = 0,\ \ell > j \\ a_{\ell,j+1} = -\ell a_{\ell j} - (j-\ell+1)a_{\ell-1,j} ,\ 1 < \ell < j,\ j > 2 . \end{cases} \tag{A.22}$$

Further, for $j > 2$

$$\frac{n(j-1)!}{\theta^j} - |\beta-1|e^{-\theta}(1-e^{-\theta})^{-j}(j-1)! < |M_n^{(j)}(\theta)| < \frac{n(j-1)!}{\theta^j}$$

$$+ |\beta-1|e^{-\theta}(1-e^{-\theta})^{-j}(j-1)! \tag{A.23}$$

and for n sufficiently large, for each fixed θ

$$\frac{n(j-1)!}{2\theta^j} < |M_n^{(j)}(\theta)| < \frac{2n(j-1)!}{\theta^j} . \tag{A.24}$$

<u>Proof</u>. (A.19), (A.20), (A.21) and (A.22) are easily verified. From (A.22), it is evident that $\operatorname{sgn} a_{\ell j} = (-1)^{j+1},\ j > 1,\ 1 < \ell < j$. Further

$$\sum_{\ell=1}^{j} a_{\ell,j+1} = \sum_{\ell=1}^{j} (-\ell a_{\ell j}) - \sum_{\ell=1}^{j} (j-\ell+1)a_{\ell-1,j}$$

$$= -\sum_{\ell=0}^{j} [\ell a_{\ell j} + (j-\ell)a_{\ell j}] = -j \sum_{\ell=0}^{j} a_{\ell j} .$$

Thus

$$\sum_{\ell=0}^{j} a_{\ell j} = (-1)^{j+1}c(j-1)!$$

for some constant c. Since $a_{11} = 1$, it follows that $c = 1$, and

$$\sum_{\ell=0}^{j} a_{\ell j} = (-1)^{j+1}(j-1)! . \tag{A.25}$$

Thus, from (A.25), we have

$$\left|\sum_{\ell=1}^{j} a_{\ell j} e^{-\ell\theta}(1-e^{-\theta})^{-j}\right| < e^{-\theta}(1-e^{-\theta})^{-j}(j-1)! \quad . \qquad (A.26)$$

(A.23 and (A.24) follow directly from (A.26), establishing the lemma.

Lemma A.2. A sequence of asymptotic $(n\to\infty)$ estimates of θ_n^* is provided by

$$\theta_n^* = \bar{\theta}_i + 0(n^{-i}),\ i = 1,2,\ldots \qquad (A.27)$$

where

$$\bar{\theta}_1 = 1/\bar{y} \quad , \qquad (A.28)$$

and

$$\bar{\theta}_{i+1} = \bar{\theta}_i + \frac{\bar{\theta}_i^2}{n}\left(\frac{n}{\bar{\theta}_i} + \frac{\beta-1}{1-e^{-\bar{\theta}_i}} - n\bar{y} - (\alpha+\beta-1)\right) \quad . \qquad (A.29)$$

Proof. For every $\varepsilon > 0$, there is an n sufficiently large so that $M_n(\theta)$ is strictly convex on (ε,∞). Thus, $M_n'(\theta) = 0$ has a unique root θ_n^* on (ε,∞). (A.27), (A.28) and (A.29) follow readily.

Specifically,

$$\begin{aligned}\theta_n^* &= \frac{1}{\bar{y}} + 0(n^{-1})\\ &= \frac{1}{\bar{y}} + \frac{1}{n\bar{y}^{-2}}\left(\frac{\beta-1}{1-e^{-1/\bar{y}}} - (\alpha+\beta-1)\right) + 0(n^{-2})\\ &= \frac{1}{\bar{y}} + \frac{1}{n\bar{y}^{-2}}\left(\frac{\beta-1}{1-e^{-1/\bar{y}}} - (\alpha+\beta-1)\right) + \frac{1}{n^2\bar{y}^{-3}}\left(\frac{\beta-1}{1-e^{-1/\bar{y}}} - (\alpha+\beta-1)\right)^2\\ &\quad - \frac{(\beta-1)e^{-1/\bar{y}}}{n^2\bar{y}^{-4}(1-e^{-1/\bar{y}})^2}\left(\frac{\beta-1}{1-e^{-1/\bar{y}}} - (\alpha+\beta-1)\right) + 0(n^{-3}) \quad . \end{aligned} \qquad (A.30)$$

Remark: Applying lemma A.2, and letting $\theta = \theta_n^*$ in (A.23), we can leduce

$$n(j-1)!/2\theta_n^{*j} < |M_n^{(j)}(\theta_n^*)| < 2n(j-1)!/\theta_n^{*j} \quad . \qquad (A.31)$$

Lemma A.3. For $0 < r < k,\ k > 2$,

$$K_{nk}^{(0)}(\theta_n^*) = 1 \qquad (A.32)$$

$$K_{nk}^{(r)}(\theta_n^*) = 0, \; r = 1,2 \tag{A.33}$$

$$|K_{nk}^{(r)}(\theta_n^*)| \le k(2n)^{r/3}\, \bar{y}^{-r}, \; r = 3,4,\ldots,k \quad , \tag{A.34}$$

for n sufficiently large.

<u>Proof</u>. From (A.18), (A.32) and (A.33) follows trivially. Therefore, we need only establish (A.34). From (A.24) and (A.18), for n sufficiently large, $3 \le r \le k$,

$$|K_{nk}^{(r)}(\theta)| \le \Sigma\Big(\frac{2n(2!)}{\theta^3}\Big)^{\ell_3}\cdots\Big(\frac{2n(k-1)!}{\theta^k}\Big)^{\ell_k}\Big/\ell_3!\cdots\ell_k!\; 3!^{\ell_3}\cdots k!^{\ell_k}$$

$$= \Sigma\Big(\frac{(2n)^{\Sigma\ell_n}}{\theta^{\Sigma j\ell_j}}\Big/\ell_3!\cdots\ell_k!\; 3^{\ell_3}\cdots k^{\ell_k} \quad . \tag{A.35}$$

Since

$$\sum_{j=3}^{k} j\ell_j = r \quad ,$$

we have

$$3\sum_{j=3}^{k} \ell_j \le r$$

and

$$\sum_{j=3}^{k} \ell_j \le r/3 \quad .$$

Thus

$$|K_{nk}^{(r)}(\theta)| \le \frac{(2n)^{r/3}}{\theta^r}\, \Sigma 1/\ell_3!\cdots\ell_k!\; 3^{\ell_3}\cdots k^{\ell_n} \quad .$$

Now

$$\Sigma 1/\ell_3!\cdots\ell_k!\; 3^{\ell_3}\cdots k^{\ell_k} \le \sum_{m=1}^{[r/3]} \frac{1}{m!} \sum_{\ell_3,\ldots,\ell_k \ge 0} \frac{m!}{\ell_3!\cdots\ell_k!}\Big(\frac{1}{3}\Big)^{\ell_3}\cdots\Big(\frac{1}{k}\Big)^{\ell_k}$$

$$= \sum_{m=1}^{[r/3]} \frac{1}{m!}\Big(\frac{1}{3} + \cdots + \frac{1}{k}\Big)^m$$

$$\le \sum_{m=1}^{\infty} \frac{1}{m!}(\log k)^m = k \quad .$$

Thus

$$|K_{nk}^{(r)}(\theta)| < k(2n)^{r/3}/\theta^r \quad .$$

From (A.31), the conclusion follows readily upon replacing θ by θ_n^* in (A.35).

We now proceed to the proof of the theorem.

__Proof__. Let $R = e^{-\theta}$ obtaining

$$I_n(\alpha,\beta,y) = \int_0^\infty e^{-\theta(\alpha+n\bar{y})}\left(1-e^{-\theta}\right)^{\beta-1}\theta^n d\theta \quad . \tag{A.36}$$

Let

$$g(\theta) = e^{-\theta(\alpha+n\bar{y})}\left(1-e^{-\theta}\right)^{\beta-1}\theta^n \quad . \tag{A.37}$$

Choose $\tau_1(n)$, $\tau_2(n)$ so that

$$0 < \tau_1(n) < n/(\alpha+n\bar{y}) < \tau_2(n) < \infty \quad . \tag{A.38}$$

For $\beta > 1$,

$$\int_0^{\tau_1(n)} g(\theta)d\theta < \int_0^{\tau_1(n)} e^{-\theta(\alpha+n\bar{y})}\theta^n d\theta < (\tau_1(n))^{n+1} e^{-\tau_1(n)(\alpha+n\bar{y})} \quad . \tag{A.39}$$

For $\beta < 1$, since $1-e^{-\theta} > \theta(1-e^{-\tau_1})/\tau_1$, $0 < \theta < \tau_1$,

$$\int_0^{\tau_1(n)} g(\theta)d\theta < \left(\frac{1-e^{\tau_1(n)}}{\tau_1(n)}\right)^{\beta-1} \int_0^{\tau_1(n)} e^{-\theta(\alpha+n\bar{y})}\theta^{n+\beta-1}d\theta \quad ,$$

whenever

$$\tau_1(n) < \frac{n+\beta-1}{\alpha+n\bar{y}} \quad ,$$

we have

$$\int_0^{\tau_1(n)} g(\theta)d\theta < \left(\frac{1-e^{-\tau_1(n)}}{\tau_1(n)}\right) \tau_1(n)^{n+\beta} e^{-\tau_1(n)(\alpha+n\bar{y})} \quad . \tag{A.40}$$

Proceeding similarly, we have,

$$\int_{\tau_2(n)}^\infty g(\theta)d\theta < f(\beta) \int_{\tau_2(n)}^\infty e^{-\theta(\alpha+ny)}\theta^n d\theta \quad ,$$

where

$$f(\beta) = 1 \quad , \quad \beta > 1$$

$$f(\beta) = \left(1-e^{-\tau_2(n)}\right)^{\beta-1} , \ \beta < 1 \quad .$$

From F. W. J. Olver [7], p. 70, we have

$$\int_{\tau_2(n)}^{\infty} e^{-\theta(\alpha+n\bar{y})}\theta^n d\theta = (\alpha+n\bar{y})^{-n-1} \int_{(\alpha+n\bar{y})\tau_2(n)}^{\infty} e^{-u}u^n du$$
$$< \frac{e^{-(\alpha+n\bar{y})\tau_2(n)}(\tau_2(n))^{n+1}}{(\alpha+n\bar{y})\tau_2(n)-n} .$$

Hence

$$\int_{\tau_2(n)}^{\infty} g(\theta)d\theta < \frac{f(\beta)e^{-(\alpha+n\bar{y})\tau_2(n)}(\tau_2(n))^{n+1}}{(\alpha+n\bar{y})\tau_2(n)-n} . \qquad (A.41)$$

From (A.15) and (A.36), we have that

$$I_n(\alpha,\beta,y) = \int_0^{\infty} e^{-M_n(\theta)} d\theta . \qquad (A.42$$

Now write, for $k > 2$

$$M_n(\theta) = M_n(\theta_n^*) + \frac{(\theta-\theta_n^*)^2}{2!} M_n''(\theta_n^*) + \cdots + \frac{(\theta-\theta_n^*)^k}{k!} M_n^{(k)}\theta_n^*$$
$$+ W_{nk}(\theta) , \qquad (A.43)$$

where

$$|W_{nk}(\theta)| < \frac{|\theta-\theta_n^*|^{k+1}}{(k+1)!} |M_n^{(k+1)}(\xi_n)| \qquad (A.44)$$

and ξ_n lies between θ and θ_n^*.

From (A.30) or (A.15), it is easily seen that $\theta^* = n/(\alpha+n\bar{y}) + o(1)$ and thus for sufficiently large n,

$$\tau_1(n) < \theta_n^* < \tau_2(n) .$$

Consequently for $\theta \in (\tau_1(n), \tau_2(n))$, from (A.24) and (A.44),

$$|W_{nk}(\theta)| < \frac{(\tau_2(n)-\tau_1(n))^{k+1}}{(k+1)!} \left(\frac{2nk!}{\xi_n^{k+1}}\right) . \qquad (A..45)$$

Now write $g(\theta)$ as

$$g(\theta) = e^{-M_n(\theta_n^*) - \frac{\gamma_n}{2}(\theta-\theta_n^*)^2} \sum_{r=0}^{\infty} \frac{(-1)^r}{r!} \left(\sum_{\ell=3}^{k} \frac{(\theta-\theta_n^*)^{\ell}}{\ell!} M_n^{(\ell)}(\theta_n^*)\right)^r .$$
$$\cdot\, e^{-W_{nk}(\theta)} \qquad (A.46)$$

where $\gamma_n = M_n''(\theta_n^*)$. Then write

$$\sum_{r=0}^{\infty} \frac{(-1)^r}{r!} \left(\sum_{\ell=3}^{k} \frac{(\theta-\theta_n^*)^\ell}{\ell!} M_n^{(\ell)}(\theta_n^*) \right)^r$$

$$= \sum_{r=0}^{\infty} \frac{(-1)^r}{r!} \sum_{\ell_3,\ldots,\ell_k > 0} \frac{r!}{\ell_3!\ldots\ell_k!} \frac{(\theta-\theta_n^*)^{3\ell_3+\ldots+k\ell_k} M_n^{(3)}(\theta_n^*)^{\ell_3}\ldots M_n^{(k)}(\theta_n^*)^{\ell_k}}{3!^{\ell_3}\ldots k!^{\ell_k}}$$

$$= 1 + \sum_{\ell=3}^{\infty} \sum_{\ell_3,\ldots,\ell_k}^{\infty} \frac{(-1)^{\sum_3^k \ell_j} (\theta-\theta_n^*)^\ell \prod_{j=3}^{k} \left(M_n^{(j)}(\theta_n^*) \right)^{\ell_j}}{\prod_{j=3}^{k} \ell_j! \prod_{j=3}^{k} j!^{\ell_j}} , \qquad \text{(A.47)}$$

where the sum is over $\ell_3,\ldots,\ell_k > 0$ with $\Sigma j\ell_j = \ell$. Thus, we can write (A.47) as

$$1 + \sum_{\ell=3}^{k} K_{nk}^{(\ell)}(\theta_n^*)(\theta-\theta_n^*)^\ell + \sum_{\ell=k+1}^{\infty} K_{nk}^{(\ell)}(\theta_n^*)(\theta-\theta_n^*)^\ell , \qquad \text{(A.48)}$$

where

$$K_{nk}^{(\ell)} = \sum_{\ell_3,\ldots,\ell_k} (-1)^{\Sigma \ell j} \prod_{j=3}^{k} \left(M_n^{(j)}(\theta_n^*) \right)^{\ell_j} \Big/ \prod_{j=3}^{k} \ell_j! \prod_{j=3}^{k} j!^{\ell_j} . \qquad \text{(A.49)}$$

Accordingly, we define

$$h_k(\theta) = e^{-M_n(\theta_n^*) - \frac{\gamma_n}{2}(\theta-\theta_n^*)^2} \left(1 + \sum_{\ell=3}^{k} K_{nk}^{(\ell)}(\theta-\theta_n^*)^\ell \right) , \qquad \text{(A.50)}$$

and consider

$$e^{M_n(\theta_n^*)} \left| \int_0^\infty g(\theta)d\theta - \int_{-\infty}^{\infty} h_k(\theta)d\theta \right| = R_k . \qquad \text{(A.51)}$$

Then

$$R_k < e^{M_n(\theta_n^*)}\{\int_{-\infty}^{\tau_1(n)} h_k(\theta)d\theta + \int_0^{\tau_1(n)} g(\theta)d\theta + \int_{\tau_1(n)}^{\tau_2(n)} |h_k(\theta)-g(\theta)|d\theta$$

$$+ \int_{\tau_2(n)}^{\infty} g(\theta)d\theta + \int_{\tau_2(n)}^{\infty} h_k(\theta)d\theta\} \quad . \tag{A.52}$$

From (A.34), we have

$$|1 + \sum_{\ell=3}^{k} K_{nk}^{(\ell)}(\theta-\theta_n^*)^\ell| < 1 + k \sum_{\ell=3}^{k} (2n)^{\ell/3} \bar{y}^{-\ell}(\theta-\theta_n^*)^\ell \quad .$$

Thus for $\tau_2(n) > \theta_n^*$,

$$e^{M_n(\theta_n^*)} \int_{\tau_2(n)}^{\infty} h_k(\theta)d\theta = \int_{\tau_2(n)}^{\infty} e^{-\frac{\gamma_n}{2}(\theta-\theta_n^*)^2}\left(1 + \sum_{\ell=3}^{k} K_{nk}^{(\ell)}(\theta-\theta_n^*)^\ell\right)d\theta \quad .$$

Setting $w = \frac{\gamma_n}{2}(\theta-\theta_n^*)^2$, we obtain

$$e^{M_n(\theta_n^*)} \int_{\tau_2(n)}^{\infty} h_k(\theta)d\theta = \frac{1}{\sqrt{2\gamma_n}} \int_{\frac{\gamma_n}{2}(\tau_2(n)-\theta_n^*)^2}^{\infty} e^{-w}\left(w^{-\frac{1}{2}} + \sum_{\ell=3}^{k} K_{nk}^{(\ell)}\left(\frac{2}{\gamma_n}\right)^{\ell/2} w^{\frac{\ell-1}{2}}\right)dw \quad .$$

Then, applying (A.31) with $j = 2$, and (A.41)

$$\left|\frac{1}{\sqrt{2\gamma_n}} \int_{\frac{\gamma_n}{2}(\tau_2(n)-\theta_n^*)^2}^{\infty} e^{-w}(w^{-\frac{1}{2}} + \sum_{\ell=3}^{k} K_{nk}^{(\ell)}\left(\frac{2}{\gamma_n}\right)^{\ell/2} w^{\frac{\ell-1}{2}} dw\right|$$

$$< \frac{1}{n^{1/2}\bar{y}} \int_{\frac{\gamma_n}{2}(\tau_2(n)-\theta_n^*)^2}^{\infty} e^{-w}\left(w^{-\frac{1}{2}} + k\sum_{\ell=3}^{k} \frac{2^{4\ell/3}}{n^{\ell/6}} w^{\frac{\ell-1}{2}}\right)dw$$

$$< \frac{1}{n^{1/2}\bar{y}}\left[e^{-\frac{\gamma_n}{2}(\tau_2-\theta_n^*)^2}\left(\left[\frac{\gamma_n}{2}(\tau_2(n)-\theta_n^*)^2\right]^{-\frac{1}{2}}\right.\right.$$

$$+ k \sum_{\ell=3}^{k} \frac{2^{4\ell/3}}{n^{\ell/6}} \frac{\left[\gamma_n(\tau_2(n)-\theta_n^*)^2\right]^{\frac{\ell+1}{2}}}{\left[\frac{\gamma_n}{2}(\tau_2(n)-\theta_n^*)^2 - \frac{\ell-1}{2}\right]}\Bigg)\Bigg] ,$$

the last inequality following from inequalities for the incomplete gamma function on pages 66 and 70 of F. J. Olver [7].

Now let $n^{-\frac{1}{3}+\delta} < \tau_2(n)-\theta_n^* = \theta_n^* - \tau_1(n) < n^{-\frac{1}{3}-\delta}$, $0 < \delta < \frac{1}{12}$. Then for n sufficiently large, $3 \leq \ell \leq k$,

$$\frac{k2^{4\ell/3}}{n^{\ell/6}}\left(\frac{\gamma_n}{2}(\tau_2(n)-\theta_n^*)^2\right)^{(\ell+1)/2} \Big/ \left(\frac{\gamma_n}{2}(\tau_2(n)-\theta_n^*)^2 - (\ell-1)/2\right)$$

$$< \frac{2k2^{4\ell/3}}{n^{\ell/6}}\left(\frac{\gamma_n}{2}(\tau_2(n)-\theta_n^*)^2\right)^{(\ell+1)/2}\left(\frac{\gamma_n}{2}(\tau_2(n)-\theta_n^*)^2\right)$$

$$= \frac{2^{(4\ell+3)/3}k}{n^{\ell/6}}\left(\frac{\gamma_n}{2}(\tau_2(n)-\theta_n^*)^2\right)^{\frac{\ell-1}{2}}$$

$$< kn^{-\ell/6}2^{(4\ell+3)/3}\left[(n\bar{y}^2)n^{-\frac{2}{3}}\right]^{\frac{\ell-1}{2}} = k2^{(4\ell+3)/3}n^{-1/6} . \qquad \text{(A.53)}$$

Similarly,

$$\left[\frac{\gamma_n}{2}(\tau_2(n)-\theta_n^*)^2\right]^{-\frac{1}{2}} < \left[\frac{n\bar{y}^2}{4}n^{-1+2\delta}\right]^{-\frac{1}{2}}$$

$$= n^{-\delta}\bar{y}^2/4 .$$

Consequently, there is a constant $c_{k,\bar{y}}$ such that

$$\left|e^{M_n(\theta_n^*)}\int_{\tau_2(n)}^{\infty} h_k(\theta)d\theta\right| < \frac{c_{k,\bar{y}}\, e^{-n^{2\delta}\bar{y}^2/4}}{n^{\frac{1}{2}+\delta}} . \qquad \text{(A.54)}$$

Similar calculations establish the same bound for

$$e^{M_n(\theta_n^*)} \int_0^{\tau_1(n)} h_k(\theta)d\theta \ .$$

We now consider $e^{M_n(\theta_n^*)} \int_{\tau_1(n)}^{\tau_2(n)} (h_k(\theta)-g(\theta))d\theta$. From (A.46) and (A.50),

$$e^{M_n(\theta_n^*)}(g(\theta) - h_k(\theta)) = e^{-\frac{\gamma_n}{2}(\theta-\theta_n^*)^2} \Big((1 + \sum_{\ell+3}^{k} K_{nk}^{(\ell)}(\theta-\theta_n^*)^{\ell}\Big) \ .$$

$$(e^{-W_{nk}(\theta)} - 1) + e^{-W_{nk}(\theta)} \sum_{\ell=k+1}^{\infty} K_{nk}^{(\ell)}(\theta-\theta_n^*)^{\ell} \ . \tag{A.55}$$

From (A.44) and (A.31), for sufficiently large n,

$$|W_{nk}(\theta)| < \frac{\big(\tau_2(n)-\tau_1(n)\big)^{k+1}}{(k+1)!} 4nk!\bar{y} \ ,$$

since $|\xi_n - \frac{1}{y}| < \tau_2(n) - \tau_1(n) < 2n^{-1/3}$. Thus

$$|e^{-W_{nk}(\theta)} - 1| = O\big((\tau_2(n)-\tau_1(n))^{k+1}n\big)$$

and

$$|W_{nk}| = O((\tau_2(n) - \tau_1(n))^{k+1}n)$$

as $n \to \infty$. Further from (A.34),

$$|\sum_{\ell=k+1}^{\infty} K_{nk}^{(\ell)}(\theta-\theta_n^*)^{\ell}| < k \sum_{\ell=k+1}^{\infty} (2n)^{\ell/3}\bar{y}^{\ell}(\tau_2(n) - \tau_1(n))^{\ell}$$

$$= \frac{k(2n)^{(k+1)/3}\ \bar{y}^{(k+1)}(\tau_2(n) - \tau_1(n))^{(k+1)}}{1 - (2n)^{1/3}\bar{y}(\tau_2(n) - \tau_1(n))} \ . \tag{A.56}$$

Since $n(\tau_2(n) - \tau_1(n))^3 \to 0$ as $n \to \infty$,

$$|\sum_{\ell=k+1}^{\infty} K_{nk}^{(\ell)}(\theta-\theta_n^*)^{\ell}| < 2k(2n)^{(k+1)/3}\ \bar{y}^{(k+1)}(\tau_2(n) - \tau_1(n))^{k+1} \tag{A.57}$$

for sufficiently large n. Similarly

$$|1 + \sum_{\ell=3}^{k} K_{nk}^{(\ell)}(\theta-\theta_n^*)^\ell|$$

$$< 1 + \sum_{\ell=3}^{k} k(2n)^{\ell/3-\ell}\bar{y}^{\ell}(\tau_2(n) - \tau_1(n))^\ell$$

$$= 1 + 0\left(\tau_2(n) - \tau_1(n))^\ell n^{\ell/3}\right) . \qquad (A.58)$$

Thus

$$\left|1 + \sum_{\ell=3}^{k} K_{nk}^{(\ell)}(\theta-\theta_n^*)^\ell\right|\left(0(\tau_2(n) - \tau_1(n)^{k+1}n\right)$$

$$= 0\left((\tau_2(n) - \tau_1(n)^{k+1}n\right) . \qquad (A.59)$$

Hence

$$\int_{\tau_1(n)}^{\tau_2(n)} |g(\theta) - h(\theta)|d\theta < \int_{\tau_1(n)}^{\tau_2(n)} e^{-\frac{\gamma_n}{2}(\theta-\theta_n^*)^2}\Big[2k(2n)^{(k+1)/3}\,\bar{y}^{\,k+1} \cdot$$

$$\cdot\ (\tau_2(n) - \tau_1(n))^{k+1} + 0(\tau_2-\tau_1)^{k+1}n)\Big]\, d\theta$$

$$= 0\left((\tau_2(n)-\tau_1(n))^{k+1}n^{(k+1)/3}\right)\int_{\tau_1(n)}^{\tau_2(n)} e^{-\frac{\gamma_n}{2}(\theta-\theta_n^*)^2} d\theta$$

$$= 0\left((\tau_2(n)-\tau_1(n))^{k+1}n^{(k+1)/3}\right)\frac{1}{\sqrt{\gamma_n}}\int_{\tau_1(n)\sqrt{\gamma_n}}^{\tau_2(n)\sqrt{\gamma_n}} e^{-(\theta-\theta_n^*)^2/2}d\theta$$

$$= 0\left((\tau_2(n)-\tau_1(n))^{(k+1)}n^{(k+1)/3-\frac{1}{2}}\right)$$

$$= 0\left(n^{-\frac{k+1}{6}-\frac{1}{2}+(k+1)\delta}\right) . \qquad (A.60)$$

From (A.37) and (A.40)

$$\int_0^{\tau_1(n)} g(\theta)d\theta < (\tau_1(n))^{n+d_\beta}\, e^{-(\alpha+n\bar{y})\tau_1(n)}\, r(\tau_1(n)) ,$$

where $d_\beta = \begin{cases} 1 , & \beta > 1 \\ \beta , & \beta < 1 \end{cases}$

and

$$r(\tau_1(n)) = \begin{cases} 1 & \beta > 1 \\ \left(1-e^{-\tau_1(n)}\right)/\tau_1(n), & \beta < 1 \end{cases} .$$

Thus

$$\int_0^{\tau_1(n)} g(\theta)d\theta/e^{-M_n(\theta_n^*)}$$

$$= \left(\frac{\tau_1(n)}{\theta^*(n)}\right)^n e^{(\theta_n^*-\tau_1(n))(\alpha+n\bar{y})} (\tau_1(n))^{d_\beta} \frac{r(\tau_1(n))}{\left(1-e^{-\theta_n^*}\right)^{\beta-1}} .$$

Since

$$\frac{1}{2\bar{y}} < \tau_1(n) < \frac{2}{\bar{y}} \tag{A.61}$$

for n sufficiently large, we can write

$$(\tau_1(n))^{d_\beta} r(\tau_1(n))/\left(1-e^{-\theta_n^*}\right)^{\beta-1} < d_1(\bar{y},\beta) , \tag{A.62}$$

where $d_1(\bar{y},\beta)$ does not depend on n. Similarly,

$$\int_{\tau_2(n)}^{\infty} g(\theta)d\theta < \frac{f_1(\beta)e^{-(\alpha+n\bar{y})\tau_2(n)}}{(\alpha+n\bar{y})\tau_2(n)-n} , \tag{A.63}$$

where

$$f_1(\beta) = \tau_2(n)f(\beta) ,$$

and from (A.61), for n sufficiently large

$$f_1(\beta) < d_2(\bar{y},\beta) ,$$

where $d_2(\bar{y},\beta)$ does not depend on n.

Note further that

$$(\alpha+n\bar{y})\tau_2(n) - n > (\alpha+n\bar{y})\left(\theta_n^*+n^{-\frac{1}{3}-\delta}\right) - n$$

$$= (\alpha+n\bar{y})\left(\frac{1}{\bar{y}} + 0\left(n^{-\frac{1}{3}}\right)\right) - n$$

$$= 0(n^{2/3}) \quad . \qquad (A.64)$$

Therefore, combining (A.62), (A.63), and (A.64), we can write

$$M_n(\theta_n^*) + \log(\int_0^{\tau_1(n)} g(\theta)d\theta) < n \log\left(\frac{\tau_1(n)}{\theta^*(n)}\right) - (\theta_n^*-\tau_1(n))(\alpha+n\bar{y})$$

$$+ \log d_1(\bar{y},\beta)$$

and

$$M_n(\theta_n^*) + \log\left(\int_{\tau_2(n)}^{\infty} g(\theta)d\theta)\right) < n \log\left(\frac{\tau_2(n)}{\theta^*(n)}\right) - (\theta_n^*-\tau_2(n))(\alpha+n\bar{y})$$

$$+ \log d_2(\bar{y},\beta) + \frac{2}{3}\log n + \log c \quad , \qquad (A.65)$$

where c is a suitable constant. Accordingly, we consider the expression, for $i = 1,2$,

$$n \log\left(\frac{\tau_i(n)}{\theta^*(n)}\right) - (\theta_n^*-\tau_i(n))(\alpha+n\bar{y})$$

$$= n \log\left(1 + \frac{\tau_i(n)-\theta^*(n)}{\theta_n^*}\right) - (\theta_n^*-\tau_i(n))(\alpha+n\bar{y})$$

$$= n\left(\frac{\tau_i(n)-\theta^*(n)}{\theta_n^*} - \frac{(\tau_i(n)-\theta^*(n))^2}{2\theta_n^*}\right) - (\theta_n^*-\tau_i(n))(\alpha+n\bar{y})$$

$$+ 0((\tau_i-\theta^*(n))^3) \quad .$$

Now, since $\theta_n^* = \frac{1}{\bar{y}} + 0(n^{-1})$, we have

$$n \log\left(\frac{\tau_i(n)}{\theta^*(n)}\right) - (\theta_n^*-\tau_i(n))(\alpha+n\bar{y})$$

$$= n((\tau_i(n)-\theta^*(n))\left(\bar{y}+0(n^{-1}\right)$$

$$- n\left(\frac{\tau_i(n)-\theta^*(n))^2}{2}\left(\frac{1}{\bar{y}} + 0(n^{-1})\right)\right.$$

$$\left. - \alpha(\theta_n^*-\tau_i(n) - n\bar{y}(\theta_n^*-\tau_i(n))\right)$$

$$+ 0((\tau_i-\theta^*n)^3)$$

$$= -n\,\frac{\left(\tau_i(n)-\theta^*(n)\right)^2}{2\bar{y}} + 0\left(n^{-\frac{1}{3}-\delta}\right) \quad . \tag{A.66}$$

Thus

$$\int_0^{\tau_1(n)} g(\theta)d\theta/e^{-M_n(\theta_n^*)} < e^{-\frac{n^{\frac{1}{3}-2\delta}}{4\bar{y}}} \tag{A.67}$$

and

$$\int_{\tau_2(n)}^{\infty} g(\theta)d\theta/e^{-M_n(\theta_n^*)} < e^{-\frac{n^{\frac{1}{3}-3\delta}}{4\bar{y}}} \tag{A.68}$$

for n sufficiently large.

Combining (A.52), (A.54), (A.60), (A.67) and (A.68), we have

$$e^{M_n(\theta_n^*)}\left|\int_0^{\infty} g(\theta)d\theta - \int_{-\infty}^{\infty} h_k(\theta)d\theta\right| = 0\left(n^{-\frac{k+1}{6}-\frac{1}{2}+\delta}\right) \quad .$$

Since

$$\int_{-\infty}^{\infty} h_k(\theta)d\theta = e^{-M_n(\theta_n^*)}\int_{-\infty}^{\infty} e^{-\frac{\gamma_n}{2}(\theta-\theta_n^*)^2}\left(1+\sum_{\ell=3}^{k} K_{nk}^{(\ell)}(\theta-\theta_n^*)^{\ell}\right)d\theta$$

$$= \frac{e^{-M_n(\theta_n^*)}}{\sqrt{\gamma_n}}\int_{-\infty}^{\infty} e^{-\frac{(\theta-\theta_n^*)^2}{2}}\left(1+\sum_{\ell=3}^{k} K_{nk}^{(\ell)}\frac{(\theta-\theta_n^*)^{\ell}}{\gamma_n^{\ell/2}}\right)d\theta$$

$$= \sqrt{\frac{2\pi}{\gamma_n}}\, e^{-M_n(\theta_n^*)}\sum_{r=0}^{k} K_{nk}^{(\ell)}\mu_r/\gamma_n^{r/2} \quad , \tag{A.69}$$

thus establishing the theorem.

APPENDIX III

From the results of Appendix II, the various estimates of section 2 all have asymptotic estimates given by ratios of the form,

$$\hat{R}_{\alpha,\beta,y} = \int_{-\infty}^{\infty} h_k(\theta,\alpha+1,\beta)d\theta / \int_{-\infty}^{\infty} h_k(\theta;\alpha,\beta)d\theta \quad . \qquad (A.70)$$

We now proceed to obtain an asymptotic expansion for this ratio.

Since $h_k(\theta)$ has the form

$$e^{-M_n(\theta_n^*) - \frac{\gamma_n}{2}(\theta-\theta_n^*)^2} \sum_{r=0}^{k} K_r(\theta_n^*)(\theta-\theta_n^*)^r$$

we have

$$\int_{-\infty}^{\infty} h_k(\theta)d\theta$$

$$= e^{-M_n(\theta_n^*)} \sqrt{\frac{2\pi}{\gamma_n}} \sum_{r=0}^{k} K_{nk}^{(r)}(\theta_n^*)\mu_r/\gamma_n^{r/2} \quad ,$$

where μ_r are the central moments of the normal distribution with variance unity. Now $M_n(\theta_n^*)$, γ_n, and θ_n^* depend on α. Thus, from (A.14) and (A.70)

$$\hat{R}_{\alpha,\beta,y} \sim e^{-M_n(\theta_n^*(\alpha+1,\beta)) + M_n(\theta_n^*(\alpha,\beta))} \sqrt{\frac{\gamma_n(\alpha,\beta)}{\gamma_n(\alpha+1,\beta)}} \cdot$$

$$\frac{\sum_{r=0}^{k} K_{nk}^{(r)}(\theta_n^*(\alpha+1,\beta)\mu_r/(\gamma_n(\alpha+1,\beta)^{r/2}}{\sum_{r=0}^{k} K_{nk}^{(r)}(\theta_n^*(\alpha,\beta)\mu_r/\gamma_n(\alpha,\beta))^{r/2}} \left(1 + 0\left(n^{-\frac{k-7}{6}+\delta}\right)\right)$$

(A.71)

We now employ the results of Appendix II to evaluate (A.71). From (A.15)

$$M_n(\theta_n^*(\alpha,\beta)) - M_n(\theta_n^*(\alpha+1,\beta)) =$$

$$(\theta_n^*(\alpha,\beta) - \theta_n^*(\alpha+1,\beta))(\alpha+n\bar{y}) - \theta_n^*(\alpha+1,\beta)$$

$$+ n \log \frac{\theta_n^*(\alpha+1,\beta)}{\theta_n^*(\alpha,\beta)} + (\beta-1) \log\left(\frac{1-e^{-\theta_n^*(\alpha+1,\beta)}}{1-e^{-\theta_n^*(\alpha,\beta)}}\right) . \qquad \text{(A.72)}$$

Using (A.30), direct calculations establish

$$\theta_n^*(\alpha,\beta) - \theta_n^*(\alpha+1,\beta) =$$

$$\frac{1}{n\bar{y}^2} + \frac{1}{n^2\bar{y}^3}\left(\frac{2(\beta-1)}{1-e^{-1/\bar{y}}} + 1 - 2(\alpha+\beta)\right)$$

$$- \frac{(\beta-1)e^{-1/\bar{y}}}{n^2\bar{y}^4\left(1-e^{-1/\bar{y}}\right)^2} + O(n^{-3}) \quad . \qquad \text{(A.73)}$$

Similarly,

$$\frac{\theta_n^*(\alpha,\beta)}{\theta_n^*(\alpha+1,\beta)} = 1 + \frac{1}{n\bar{y}} + \frac{1}{n^2\bar{y}^2}\left(\frac{\beta-1}{1-e^{-1/\bar{y}}} - (\alpha+\beta-1)\right)$$

$$- \frac{(\beta-1)e^{-1/y}}{n^2\bar{y}^3\left(1-e^{-1/\bar{y}}\right)^2} + O(n^{-3}) \quad . \qquad \text{(A.74)}$$

Thus,

$$n \log(\theta_n^*(\alpha,\beta)/\theta_n^*(\alpha+1,\beta)) =$$

$$\frac{1}{\bar{y}} + \frac{1}{n\bar{y}^2}\left(\frac{\beta-1}{1-e^{-1/\bar{y}}} - (\alpha+\beta-1)\right) - \frac{(\beta-1)e^{-1/\bar{y}}}{n\bar{y}^3\left(1-e^{-1/\bar{y}}\right)^2} - \frac{1}{2n\bar{y}^2}$$

$$+ O(n^{-2}) \quad . \qquad \text{(A.75)}$$

Now write

$$1-e^{-\theta_n^*(\alpha,\beta)} = (1 - e^{-1/\bar{y}}) + \frac{e^{-1/\bar{y}}}{n\bar{y}^2}\left(\frac{(\beta-1)}{1-e^{-1/\bar{y}}} - (\alpha+\beta-1)\right) + O(n^{-2}) \qquad (A.76)$$

and replacing α by $\alpha+1$ in (A.76) produces $1-e^{-\theta_n^*(\alpha+1,\beta)}$.

Rewrite (A.76) as

$$(1-e^{-1/\bar{y}})\left(1 + \frac{e^{-1/\bar{y}}}{n(1-e^{-1/\bar{y}})\bar{y}^2}\left(\frac{\beta-1}{1-e^{-1/\bar{y}}} - (\alpha+\beta-1)\right) + O(n^{-2})\right) .$$

Then

$$\log\left(1-e^{-\theta_n^*(\alpha,\beta)}\right) = \log(1-e^{-1/\bar{y}}) + \frac{e^{-1/\bar{y}}}{n(1-e^{-1/\bar{y}})\bar{y}^2}\left(\frac{\beta-1}{1-e^{-1/\bar{y}}} - (\alpha+\beta-1)\right) + O(n^{-2}) .$$

Thus,

$$\log\left\{\frac{\left(1-e^{-\theta_n^*(\alpha+1,\beta)}\right)}{\left(1-e^{-\theta_n^*(\alpha,\beta)}\right)}\right\} = \frac{-e^{-1/\bar{y}}}{n\bar{y}^2\left(1-e^{-1/\bar{y}}\right)} + O(n^{-2}) . \qquad (A.77)$$

Consequently,

$$M_n(\theta_n^*(\alpha,\beta) - M_n(\theta_n^*(\alpha+1,\beta)) =$$

$$-\frac{1}{\bar{y}} + \frac{1}{n\bar{y}^2}(\alpha + \tfrac{1}{2}) - \frac{(\beta-1)e^{-1/\bar{y}}}{n\bar{y}^2(1-e^{-1/\bar{y}})} + O(n^{-2}) . \qquad (A.78)$$

We now evaluate $\gamma_n(\alpha,\beta)/\gamma_n(\alpha+1,\beta)$. Since

$$n\gamma_n(\alpha,\beta) = \frac{1}{(\theta_n^*(\alpha,\beta))^2} - \frac{\beta-1}{n}\left[\left(1-e^{-\theta_n^*(\alpha,\beta)}\right)^{-2} - \left(1-e^{-\theta_n^*(\alpha,\beta)}\right)^{-1}\right] ,$$

from (A.30)

$$(\theta_n^*(\alpha,\beta))^{-2} = \bar{y}^2\ 1 + \frac{2}{n\bar{y}}\left(\frac{\beta-1}{1-e^{-1/\bar{y}}} - (\alpha+\beta-1)\right)^{-1} + O(n^{-2}) ,$$

and

$$\frac{\beta-1}{n}\left(\left(1-e^{-\theta_n^*(\alpha,\beta)}\right)^{-2} - \left(1-e^{-\theta_n^*(\alpha,\beta)}\right)^{-1}\right)$$

$$= \frac{\beta-1}{n}\left[\left(1-e^{-1/\bar{y}}\right)^{-2} - \left(1-e^{-1/\bar{y}}\right)^{-1}\right] + 0(n^{-2}) \quad ,$$

we set

$$\gamma_n(\alpha,\beta)/\gamma_n(\alpha+1,\beta) = 1 - \frac{2}{n\bar{y}} + 0(n^{-2})$$

and

$$\left(\gamma_n(\alpha,\beta)/\gamma_n(\alpha+1,\beta)\right)^{\frac{1}{2}} = 1 - \frac{1}{n\bar{y}} + 0(n^{-2}) \quad . \tag{A.79}$$

Finally, we set $k = 5$, obtaining

$$\sum_{r=0}^{k} K_{nk}^{(r)}(\theta_n^*(\alpha,\beta))\mu_r/\left(\gamma_n(\alpha,\beta)\right)^{r/2}$$

$$= 1 - M_n^{(4)}(\theta_n^*(\alpha,\beta))/8\gamma_n^2(\alpha,\beta) \quad . \tag{A.80}$$

From (A.19), (A.30) and (A.79), we see that

$$\frac{1-M_n^{(4)}\left(\theta_n^*(\alpha+1,\beta)\right)/8\gamma_n^2(\alpha+1,\beta)}{1-M_n^{(4)}\left(\theta_n^*(\alpha,\beta)\right)/8\gamma_n^2(\alpha,\beta)} = 1 + 0(n^{-2}) \quad . \tag{A.81}$$

Combining (A.71), (A.76), (A.77), (A.78), (A.79) and (A.80), we have

$$\hat{R}_{\alpha,\beta,y} = e^{-\frac{1}{\bar{y}} + \frac{1}{n\bar{y}^2}(\alpha+\frac{1}{2}) - \frac{(\beta-1)^{-1/\bar{y}}}{n\bar{y}^2(1-e^{-1/\bar{y}})}}\left(1 - \frac{1}{n\bar{y}}\right)\left(1 + 0(n^{-2})\right) \quad . \tag{A.82}$$

REFERENCES

[1] Abramowitz, M. and Stegun, I. A., Editors, "Handbook of Mathematical Functions," National Bureau of Standards, Applied Mathematics Series. 55, Washington, 1964.

[2] Basu, A. P., "Estimates of Reliability for Some Distributions Useful in Life Testing," Technometrics, 6 (1964), 215-219.

[3] Bhattacharya, S. K., "Bayesian Approach to Life Testing and Reliability Estimation," J. Amer. Statist. Assoc., 62, (1967), 48-62.

[4] El-Sayyad, B. M., "Estimation of the Parameter of an Exponential Distribution," J. Roy. Statist. Soc. Ser. B, 29 (1967), 525-532.

[5] Ferguson, T. S., "Mathematical Statistics, A Decision Theoretic Approach," Academic Press, New York, 1967.

[6] Mann, N. R., Schafer, R. E. and Singpurwalla, N. D., "Methods for Statistical Analysis of Reliability and Life Data," John Wiley and Sons, New York, 1974.

[7] Olver, F. W. J., "Asymptotics and Special Functions," Academic Press, New York, 1974.

[8] Pugh, E. L., "The Best Estimate of Reliability in the Exponential Case," Oper. Res., 11 (1963), 57-61.

[9] Rao Tummala, V. M. and Sathe, P. T., "Minimum Expected Loss Estimators of Reliability and Parameters of Certain Lifetime Distributions," IEEE Trans. on Reliability, R-27 (1978), 283-285.

[10] Sinha, S. K. and Guttman, I., "Bayesian Inference About the Reliability Functions for the Exponential Distributions," Commun. Statist. - Theor. Meth., A5, (1976), 471-479.

[11] Sinha, S. K. and Guttman, I., "Bayesian Inference About the Reliability Function for Exponential Distributions," University of Manitoba Technical Report No. 58, 1974.

[12] Zacks, S. and Even, M., "The Efficiencies in Small Samples of the Maximum Likelihood and Best Unbiased Estimators of Reliability Functions," J. Amer. Statist. Assoc., 61 (1966), 1033-1051.

BH/APS/jvs

A NOTE ON THE VARIABLE KERNEL ESTIMATOR OF THE HAZARD FUNCTION FROM RANDOMLY CENSORED DATA*

Martin A. Tanner

Department of Statistics
University of Wisconsin
Madison, Wisconsin 53706

ABSTRACT

In a recent paper (Tanner and Wong (1982b)), a family of data-based nonparametric hazard estimators was introduced. Several of these estimators were studied in an extensive simulation experiment. The estimator which allows for variable bandwidth was found to have a superior performance. In this note, sufficient conditions for the variable kernel estimator to be strongly consistent are presented.

---

*Support for this research has been provided in part by Research Grant No. MCS-8101836 of the National Science Foundation. The author wishes to express his sincere appreciation to professors R. R. Bahadur, S. M. Stigler, and W. H. Wong for their valuable suggestions.

1. INTRODUCTION

Let (T_i, C_i), $i = 1,\ldots,n$ be independent and identically distributed pairs of non-negative random variables. Assume that T_i and C_i are independent for all i. Denote by $S_T(f_T)$ and $S_C(f_C)$ the survivor (density) functions of T_i and C_i, respectively. (See Miller (1981).) In the random censorship model we observe the pairs $(Y_i, \tilde{\delta}_i)$, $i = 1,\ldots,n$ where

$$Y_i = \min(T_i, C_i)$$

$$\tilde{\delta}_i = I(T_i < C_i)$$

$$\ell = \sum_{i=1}^{n} \tilde{\delta}_i \quad .$$

The problem is to estimate the hazard function $h(z) = f_T(z)/S_T(z)$.

Define R_k as the distance from the point z to the kth nearest of $Y_{i_1},\ldots,Y_{i_\ell}$, where $\tilde{\delta}_{i_1} = \tilde{\delta}_{i_2} = \ldots = \tilde{\delta}_{i_\ell} = 1$ (assume $k \leq \ell$). R_k, then, is the distance to the kth closest failure neighbor from z. Define J to be the index of the largest order statistic of the ℓ failure points which precedes the interval $[z - R_k, z + R_k]$. (If $z - R_k < Y_{(i_1)}$, let $J = i_1 - 1$.) Let $J' = \min(n, J + k)$ and let δ_i be the indicator random variable associated with $Y_{(i)}$.

The <u>variable kernel estimator</u> of $h(z)$ is defined as

$$\tilde{h}_n(z) = \frac{1}{2R_k} \sum_{i=1}^{n} \frac{\delta_i}{n-i+1} K\left(\frac{z-Y_{(i)}}{2R_k}\right) \quad . \qquad (1)$$

This estimator has the appealing feature that the configuration of the data plays a role in determining the degree of smoothing. In data sparse (dense) regions, R_k will be large (small) and the kernel will be flat (peaked).

In an extensive simulation study, Tanner and Wong (1982b) compare a data-based 3-parameter nonparametric estimator, which incorporates the kth nearest failure neighbor distance, to a data-based 1-parameter nonparametric estimator with constant bandwidth. (The theoretical properties of the 1-parameter estimator are discussed in detail in Tanner and Wong (1982a), while Yandell (1983) and Ramlau-Hansen (1983) examine a truncated 1-parameter kernel estimator.) The performance of the data-based 3-parameter estimator is shown to be superior to that of the 1-parameter estimator. Our ultimate goal is to establish the theoretical properties of this fully data-adaptive estimator. However, this is a difficult problem. We regard the present paper as solving a significant component problem. One must understand how these estimators behave when the parameters are chosen deterministically as a prerequisite to the analysis of the behavior of the data-adaptive procedure.

Several authors (Fix and Hodges (1951), Loftsgaarden and Quesenberry (1965), Wagner (1975), Moore and Yackel (1977) and Mack and Rosenblatt (1979)) have discussed the theoretical properties of the variable kernel estimator of the density function and the special case nearest neighbor estimator. We point out that the estimation of the hazard is a somewhat more difficult problem, since formula (1) depends on both the order statistics of the sample and the ordering induced by estimating the hazard at a point and sorting the data to obtain the kth nearest failure neighbor of this point. For this reason, direct application of previous techniques yields intractable formulas.

2. CONSISTENCY OF $\tilde{h}_n$

We assume that the survivor and density functions are continuous in a neighborhood around the point of interest. We begin with some lemmas. In Lemma 2.1, we present the density of R_k. We use this result in Lemma 2.2 to show that R_k converges almost surely to zero. Lemma 2.3 enables us to use Proposition 31 of Aalen (1978) to prove almost sure convergence of $\tilde{h}_n(z)$.

LEMMA 2.1. Let R_k represent the distance between the point x and its kth nearest failure point. Let $p = P(T_i > C_i)$, $G(r) = \int_{|x-y|<r} f_T(y)S_C(y)dy$, $F(r) = (1-p)G(r)$, $G'(r) = f_T(x-r)S_C(x-r) + f_T(x+r)S_C(x+r)$ and $F'(r) = (1-p)G'(r)$. Then the density of R_k is

$$f_{R_k} = n\binom{n-1}{k-1} F(r)^{k-1}(1-F(r))^{n-k}F'(r) \quad .$$

PROOF. The probability of m censored observations in a sample of size n is given as

$$P(m) = \binom{n}{m} p^m(1-p)^{n-m} \quad .$$

In addition, given that m observations in a sample of size n have been censored, the density of R_k is given as

$$P(r|m) = (n-m)\binom{n-m-1}{k-1} G(r)^{k-1}(1-G(r))^{n-m-k}\, G'(r) .$$

The result now follows by direct calculation.

LEMMA 2.2. Let $k = k(n) = [n^{\alpha}]$, $0 < \alpha < 1$, and let R_k be defined as above. Then $R_k \xrightarrow{a.s.} 0$.

PROOF. Given $\delta' > 0$, by Lemma 2.1 and repeated application of integration by parts it is easy to show that

$$P(R_k > \delta') \le \sum_{i=0}^{k-1} \binom{n}{i} \delta^i (1-\delta)^{n-i}$$

From Chernoff (1952), it can be shown that this quantity is bounded by $2^{-nA(n)}$, where $A(n)$ equals

$$-\left[\log_2(\delta^{\delta}(1-\delta)^{1-\delta}\right] + \log_2\left[p^p(1-p)^{1-p}\right] + \log_2\left(\frac{1-\delta}{\delta}\right)^p - \log_2\left(\frac{1-\delta}{\delta}\right)^{\delta},$$

with $p = k/n$. It is now straightforward to show that

$$-nA(n) = -n\left\{\left[\varepsilon - \left|\log_2\left(1 - \frac{1}{n^{1-\alpha}}\right)\right|\right] - \frac{1}{n^{1-\alpha}}\left[(1-\alpha)\log_2(nc) - \left|\log_2\left(1 - \frac{1}{n^{1-\alpha}}\right)\right|\right]\right\}.$$

For n sufficiently large we have $-nA(n) < -n\varepsilon'$, for some positive $\varepsilon' < \varepsilon$, and the result follows.

LEMMA 2.3. Let R_k and $k = k(n)$ be defined as above, with $1/2 < \alpha < 1$. Then

$$\frac{n^{1/2}}{\log(n)} R_{k(n)} \xrightarrow{a.s.} \infty$$

PROOF. The result will follow if we can show that for all $\varepsilon > 0$,

$$\sum_{n=2}^{\infty} P\left(\frac{n^{1/2}}{\log(n)} R_{k(n)} \le \varepsilon\right) < +\infty .$$

Now

$$P\left(\frac{n^{1/2}}{\log(n)} R_{k(n)} \le \varepsilon\right) = \int_0^{F(\varepsilon_n)} n\binom{n-1}{k-1} t^{k-1}(1-t)^{n-k}dt,$$

where $\varepsilon_n = \varepsilon \dfrac{\log(n)}{n^{1/2}}$. One can show that the result will follow if

$$\sum_{n=2}^{\infty} \int_0^{\varepsilon_n} n\binom{n-1}{k-1} t^{k-1}(1-t)^{n-k}dt < +\infty .$$

Proceeding analogously to Lemma 2.2

$$\int_0^{\varepsilon_n} n\binom{n-1}{k-1} t^{k-1}(1-t)^{n-k}dt = \sum_{i=k}^{n} \binom{n}{i} \varepsilon_n^i(1-\varepsilon_n)^{n-i} \le 2^{-nA(n)} ,$$

where, for $p = k/n$,

$$A(n) = -\log_2 \varepsilon_n^p - \log_2(1-\varepsilon_n)^{1-p} + \log_2(p)^p + \log_2(1-p)^{1-p}$$

$$= \frac{1}{n^{1-\alpha}} \log_2\left(\frac{n^{1/2+\alpha-1}}{\varepsilon \log(n)}\right) + \left(1 - \frac{1}{n^{1-\alpha}}\right) \log_2\left(\frac{1 - \frac{1}{n^{1-\alpha}}}{1 - \frac{\varepsilon \log n}{n^{1/2}}}\right)$$

Hence for $1/2 < \alpha < 1$ and sufficiently large n, $-nA(n) < -n^{\alpha'}$, where $0 < \alpha' < \alpha$, and the result follows.

THEOREM 2.1. Let $k = k(n) = [n^{\alpha}]$, $1/2 < \alpha < 1$, and R_k be defined as above, let $K(\cdot)$ be a function of bounded variation with compact support on the interval $[-1,+1]$, let h be continuous at z, then $\tilde{h}_n(z) \xrightarrow{a.s.} h(z)$.

PROOF. Let $A_n = \{\sup_{m\geq n}|\tilde{h}_m(z) - h(z)| > \varepsilon\}$ for $\varepsilon > 0$. Now choose δ such that $(z - 2\delta) \geq 0$, then

$$A_n = \left\{A_n \cap \left\{\sup_{m\geq n} R_{k(m)} > \delta\right\}\right\} \cup \left\{A_n \cap \left\{\sup_{m\geq n} R_{k(m)} < \delta\right\}\right\} .$$

Now

$$P(A_n) \leq P\left(A_n \cap \left\{\sup_{m\geq n} R_{k(m)} > \delta\right\}\right) + P\left(A_n \cap \left\{\sup_{m\geq n} R_{k(m)} < \delta\right\}\right)$$

and by Lemma 2.2, $R_k \xrightarrow{a.s.} 0$. Hence we need only consider the event $\{A_n \cap \{\sup_{m\geq n} R_{k(m)} < \delta\}\}$. Now by the triangle inequality one can show that

$$\left\{A_n \cap \left\{\sup_{m\geq n} R_{k(m)} < \delta\right\}\right\} \subseteq \left\{A_n' \cap \left\{\sup_{m\geq n} R_{k(m)} < \delta\right\}\right\} \cup \left\{A_n'' \cap \left\{\sup_{m\geq n} R_{k(m)} < \delta\right\}\right\} ,$$

where,

$$A_n' = \left\{\sup_{m\geq n}\left(\left|\frac{1}{2R_{k(m)}}\int_{|u|\leq 1} K(u)d\hat{H}_m(z-2R_{k(m)}u) - \frac{1}{2R_{k(m)}}\int_{|u|\leq 1} K(u)dH(z-2R_{k(m)}u)\right|\right) \geq \frac{\varepsilon}{2}\right\},$$

and

$$A_n'' = \left\{\sup_{m\geq n}\left(\left|\frac{1}{2R_{k(m)}}\int_{|u|\leq 1} K(u)dH(z - 2R_{k(m)}u) - h(z)\right|\right) \geq \frac{\varepsilon}{2}\right\},$$

Regarding the first event, using an application of integration by parts, one can show

$$\left\{A'_n \cap \left\{ \sup_{m\geq n} R_{k(m)} < \delta \right\}\right\} \subseteq \left\{ \sup_{m\geq n} \left(\frac{c}{2R_{k(m)}} \sup_{y\in[z-2\delta,z+2\delta]} |\hat{H}_m(y)-H(y)| \right) \geq \frac{\varepsilon}{4} \right\},$$

since $K(\cdot)$ is assumed to be a function of bounded variation with compact support. But by Proposition 3i of Aalen (1978) and Lemma 2.3, we have that

$$\lim_{n\to\infty} P\left(\sup_{m\geq n} \left(c \sup_{y\in[z-2\delta,z+2\delta]} |\hat{H}_m(y) - H(y)|/2R_{k(m)} \right) \geq \frac{\varepsilon}{4} \right) = 0.$$

Regarding the second event, it is immediate that

$$\left\{A''_n \cap \left\{ \sup_{m\geq n} R_{k(m)} > \delta \right\}\right\} \subseteq \{A''_n\} .$$

Therefore, if the function

$$f(\alpha) = \begin{cases} 0 & \alpha = 0 \\ \left| \int_{|u|\leq 1} K(u)h(z-2\alpha u)\,du - h(z) \right| & \alpha > 0 \end{cases}$$

is continuous at z, then $\lim_{n\to\infty} P(A''_n) = 0$, since $R_k \xrightarrow{a.s.} 0$. Now $f(\alpha)$ can be shown to be dominated by

$$\max_{|u|\leq 1} |h(z-2\alpha u) - h(z)| \int_{|u|\leq 1} |K(u)|\,du.$$

If we let $\alpha \to 0$, the result follows.

REFERENCES

[1] Aalen, O. (1978). Nonparametric estimation of partial transition probabilities in multiple decrement models. Annals of Statistics. **6** 534-545.

[2] Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations. Annals of Mathematical Statistics, **23** 493-507.

[3] Fix, E., and Hodges, J. L. (1951). Discriminatory analysis. Nonparametric discrimination: consistency properties, Report 4, Project Number 21-49-004, USAF School of Aviation Medicine, Randolf Field, Texas.

[4] Loftsgaarden, D. O., and Quesenberry, C. P. (1965). A nonparametric estimate of a multivariate density function. Annals of Mathematical Statistics. **36** 1049-1051.

[5] Mack, Y. P., and Rosenblatt, M. (1979). Multivariate k-nearest neighbor density estimates. Journal of Multivariate Analysis. **9** 1-15.

[6] Miller, R. G. (1981). Survival Analysis. Wiley, New York.

[7] Moore, D. S., and Yackel, J. W. (1977). Large sample properties of nearest neighbor density function estimators. In Statistical Decision Theory and Related Topics (S. S. Gupta and D. S. Moore, Eds.) Academic Press, New York.

(8) Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. Annals of Statistics, (To Appear).

(9) Tanner, M. A. and Wong, W. H. (1982a). The estimation of the hazard function from randomly censored data by the kernel method. Technical Report 134. Department of Statistics, University of Chicago.

(10) Tanner, M. A. and Wong, W. H. (1982b). Data-based nonparametric estimation of the hazard function with applications to exploratory analysis and model diagnostics. Journal of the American Statistical Association, (To Appear).

(11) Wagner, T. J. (1975). Nonparametric estimates of probability densities. IEE Transactions on Information Theory. IT-21, 438-440.

(12) Yandell, B. S. (1983). Non-parametric inference for rates and densities with censored serial data. Annals of Statistics, (To Appear).

THE EARLY INFLUENCE OF W. EDWARDS DEMING ON THE DEVELOPMENT OF STATISTICAL QUALITY CONTROL IN THE UNITED STATES AND IN JAPAN

Nancy R. Mann
Department of Biomathematics
UCLA
Los Angeles, California 90024

The first time that I heard a detailed account of W. Edwards Deming's early experiences with quality control, I was in Washington to attend an "International Conference on Fatigue Failure of Engineering Structures" and to give a paper there. It was 1969 and the 69th year of his life.

Some time before the conference, I received an invitation to join Dr. Deming and another conference speaker who was from the University of Lisbon. The invitation was for dinner at the Cosmos Club on the first evening of the event. I accepted with some feeling of anticipation, and then when I arrived in town, touched base by phone to find when I should meet them for the occasion.

At that time I discovered two problems concerning my projected fellow dinner guest. First, he was not the Portuguese professor whom Ed Deming had met some years before, but a younger man with the same unusual last name. Second, he had stopped in New York, on his way to Washington from Portugal, to consult with a co-investigator at Columbia University; had thereupon contracted food poisoning, or perhaps stomach flu; and had rushed back home to Lisbon to seek treatment or comfort or both. He didn't make it to the conference, though he did later submit a paper for the published proceedings.

So, it didn't matter that this was the wrong man; he didn't come to dinner anyhow. That left the two of us to eat and converse, once we met in the Ladies' Parlor, just inside the ladies' entrance to the Cosmos Club. (In those days, my consciousness of male chauvinism was languishing comfortably, yet to be raised, so I paid little attention to this quaint arrangement.) I might add that the ladies' entrance is still there at the Cosmos Club, but its use has been abandoned for reasons of security.

Mealtime provided a chance for me to find out how W. Edwards Deming, who was originally trained in mathematics and physics, had made such an impact on the discipline of statistical quality control and had had so much infuence in its application in this country and in Japan.

Recently, I have refreshed my memory and filled in details in conversations with him during several Saturday and Sunday afternoons in his home-based office in Washington. These took place between his trips to South Africa, British Columbia, the Netherlands, Japan, Korea and most of the major and many of the minor cities of the United States. I have also been aided in the following by documentation provided (directly or indirectly) by Churchill Eisenhart, Allen Wallis, Holbrook Working and THE MAINICHI DAILY NEWS of Tokyo, the issue of November 10, 1965.

It's useful to begin the story in March of 1938, shortly before the time

Ed Deming left his position as a mathematical physicist at the U. S. Department of Agriculture (USDA) to join the U.S. Census Bureau. At that time, he arranged for Dr. Walter Shewhart of the Bell Telephone Laboratories to deliver a series of four lectures on "Statistical Method from the Viewpoint of Quality Control" at the USDA Graduate School. These lectures were published by the Graduate School in 1939 "with the editorial assistance of W. Edwards Deming".

Shewhart, in his 1931 book, "Economic Control of Quality of Manufactured Product," had given his criteria for determining whether a given set of numerical data was in a state of statistical control -- and had given also the particulars of his corresponding control-chart techniques.

In a 1981 interview published in MILITARY SCIENCE AND TECHNOLOGY, Volume 1, Issue No. 3, Ed Deming discussed Shewhart's important contribution.

> "Dr. Shewhart first saw the fact that random variation represents the ultimate state of a system, that when you have achieved that state, then you have an identifiable process, and until then you do not -- you have chaos in a small degree or to a high degree.
>
> "Causes of nonrandom variation are called assignable causes or special causes. And those are usually chargeable to particular, local conditions that the workers can recognize and eliminate. And then you have left random variation that defines the system, and from then on only the management can improve it. That was Shewhart's great contribution."

In his first book, Shewhart was concerned with the application of his methods and techniques in controlling the quality of industrial production processes. In the USDA lectures and the book derived from them, however, he not only reviewed his earlier work and the developments during the intervening years, but also devoted one full lecture (chapter) to their application to the results of measurement of physical properties and constants, and one lecture (chapter) to the "specification of accuracy and precision" of measurement processes generally.

The editing of the Shewhart book, along with earlier work with Harold Dodge at Bell Labs and Captain Leslie Simon (later Lt. General) at Aberdeen Proving Ground, had a profound effect on W. Edwards Deming. The ideas that resulted from this exposure are central to his total philosophy of dealing with problems of production.

He first made use of the material in the two chapters in the Shewhart book on measurement and precision in consulting he did some few years later for the U.S War Department during World War II. Shewhart's general theory, however, he applied shortly after he became familiar with it. This is explained by Dr. Churchill Eisenhart, Senior Research Fellow at the National Bureau of Standards, in notes he wrote recently on Deming accomplishments. The notes were to be read on the occasion of the presentation of a fourth honorary doctorate to Dr. Deming, this by the University of Maryland on January 8, 1983. Many of the facts were obtained from "Revolution in U.S. Government Statistics, 1926 - 1976," a 1978 U.S. Government Printing Office publication by Joseph Duncan and William Shelton.

"In neither of his books, nor in his other related publications, did Shewhart mention or hint that his statistical quality control procedures could be applied equally well to routine clerical operations, with comparable beneficial effect. This is obvious once one thinks of it, and think of it Deming did. Statistical quality control procedures were applied, at his suggestion, in the clerical operations of the 1940 population census, for example in the coding and card-punching operations. The procedure worked very well. During the learning period, the error rate of a card puncher was high; but with training and experience, a good card puncher's error rate dropped markedly and exhibited statistical control at a low level. At first, the work of all card punchers received 100% verification or correction. Later 39% qualified for only sample verification.

"Work subject only to sample verification flowed through the process six times faster than otherwise. Deming and Leon Geoffrey, in an article in the September 1941 issue of the JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, estimated that the introduction of quality control saved the Bureau $263,000, which was transferred to other work, and also contributed to earlier publication. Use of statistical quality control procedures has been a standard practice at the Bureau of the Census ever since."

The next relevant scenario began early in 1942, soon after war was declared against the United States by Japan, Germany, Italy, and their allies. By then Ed Deming was well established at the Census Bureau, but was also working half time as Consultant to the Secretary of War. Recent Wilks Award winner, W. Allen Wallis, now Undersecretary of State for Economic Affairs and then on the faculty of Stanford University, tells of those times in the June 1980 issue of the JOURNAL OF THE AMERICAN

STATISTICAL ASSOCIATION:

"The atmosphere there that spring was satirized by a squib in the student paper saying, 'It is rumored that in the outside world there is a war and a shortage of Coca Cola.' As one of several statisticians - Holbrook Working, Eugene Grant, Quinn McNemar, Harold Bacon - seeking some way that we at Stanford could contribute to the war effort, I wrote on April 17, 1942, to a friend in Washington, W. Edwards Deming of the Census Bureau, that 'those of us teaching statistics in various departments here are trying to work out a curriculum adapted to the immediate statistical requirements of the war. It seems probable that a good many students with research training might by training in statistics become more useful for war than in their present work, or might increase their usefulness within their present fields.....'

"Deming responded on April 24 with four single-spaced pages on the letterhead of the Chief of Ordnance, War Department. After some explanatory background on the theme that, 'the only useful function of a statistician is to make predictions, and thus to provide a basis for action', he wrote:

"'Here is my idea. Time and materials are at a premium, and there is no time to be lost. There is no royal short cut to producing a highly trained statistician, but I do firmly believe that the most important principles of application can be expounded in a very short time to engineers and others. I have done it and have seen it done. You could accomplish a great deal by holding a school in the Shewhart methods some time in the near future. I would suggest a concentrated effort -- a short course followed by a long course. The short course would be a two-day sesson for executives and industrial people who want to find out some of the main principles and advantages of a statistical program in industry. It would be a sort of popularization, four lectures by noted industrial people who have seen statistical methods used and can point out some of their advantages. The long course would extend over a period of weeks, or, if given evenings, over a longer period. It would be attended by the people who actually intend to use statistical methods on the job. In many cases they would be delegated by the men who had attended the short course.'

"'I would suggest that both courses be thrown open to engineers, inspectors, and industrial people with or without mathematical and statistical training. Naturally, any person who has had considerable statistical training would be in a position to get much more out of the course, but few would be in this fortunate position...'

"On May 1, I was able to write Deming that, 'Your letter arrived a few hours ago...The specific suggestions struck home so well that Holbrook Working (Chairman of the University Committee on Statistics) has already talked with two or three of the key people and arranged a general meeting of everyone in statistics'; on May 21 the first letter about the course went to firms supplying Army ordnance in the western region; and the first course was given in July 1942 at Stanford."

A short article by Holbrook Working, published in SCIENCE in November, 1942 describes this effort. Working, after some preliminaries and his description of the Deming letter, went on as follows.

"The suggestion posed two problems: that of providing for the requisite instruction, and that of bringing to the course men actually in a position to apply the methods.

"Suitable machinery for organizing and financing the suggested course was already in existence in the engineering science and management War Training Program, financed by the Office of Education. The institutional director of the program at Stanford took up the plan with enthusiasm. Aided by active support from the Ordnance Department, through its San Francisco District Office, he was able to bring together, in early July, less than six weeks after the original suggestion had been received, a group of twenty-nine key men from industries holding war contracts and from procurement agencies of various branches of the armed services. These men, with three others, entered upon an intensive ten-day course with classes running eight hours a day. All thirty-two men completed the course."

Dr. Working went on to describe a second course, offered in Los Angeles in September, 1942, and then discussed the personnel involved in the instruction.

"Two Professors, Eugene L. Grant and Holbrook Working from different departments of Stanford University and Dr. W. Edwards Deming from the Census Bureau worked together in each course. A fourth man on the staff for each course was drawn from industry to present the point of view of a man meeting, from day to day, the practical problems of applying the methods under discussion.

Dr. Churchill Eisenhart, in his notes on Deming accomplishments, describes subsequent events.

"The course was such a success that early in 1943, Working was chosen to head the now famous major national program that put on intensive 8-day courses in statistical quality control throughout the country, under the auspices of the Office of Production Research and Development of the U.S. Office of Education. Deming was the teacher of 23 of these courses. By March 1945, they had been attended by more than 1900 persons from 678 industrial concerns in the United States and 13 in Canada. Many of the 'students' in the earlier of these went out to serve as 'instructors' in part-time courses that brought the message to an additional 31,000 persons in American and Canadian industry, and 2500 persons attended other part-time courses in statistical quality control. The program had an enormously beneficial effect on the quality and volume of American and Canadian war production; and 'prepared the soil' for the establishment of the American Society for Quality Control (ASQC) in February 1946, in the founding of which Deming also played an important role."

Ed Deming agrees that he did, indeed, play an important role in the founding of the ASQC.

"Wherever I taught I told the people, 'Nothing will happen if you don't keep working together. And you've learned only a little. I know only a little. You must keep on working and meeting together. Get someone to send out postal cards, and persuade someone to let you use a room for an evening.' And they did it. It was that nucleus upon which congealed the ASQC."

In the 1981 MILITARY SCIENCE AND TECHNOLOGY interview, Ed Deming stated in a few words the principal reason that the brilliant successes in using statistical quality control methodology to increase quality and productivity, later to be exhibited on a grand scale in Japan, were not realized in this country.

> "The courses were well received by engineers, but management paid no attention to them. Management did not understand that they had to get behind quality control and carry out their obligations from the top down."

In our recent conversations, he expanded on this theme, discussing first the random variation that defines a process, a manufacturing process, say, in an industrial setting.

> "In the wartime courses we taught people that there is variation in all things and that the readings that one takes from a manufacturing process must exhibit stable randomness, or they don't have any meaning as far as defining the process. Any instabilities can help to point out specific times or locations of local problems. Once these local problems are removed, then there is a process that will persist until somebody changes it.
>
> "Changing the process is management's responsibility. And that we failed to teach. Professor Working thought that it would be a good idea to include management in the courses, so we devoted one afternoon to letting the people bring their management. Well, some did come, but most did not. And I don't think we had anything wonderful

to tell them. We had no stories to tell them.

> "By 1950, these simple methods that we had taught were working all right, but nothing astounding happened. Not that they weren't accomplishing something, but it was only a small part of what could be accomplished. The big gains come from changes in the system, the responsibility of management.

In Japan, management did take responsibility for putting statistical quality control methods to work. The story of how that happened begins in 1946. In that year, Dr. Deming made a trip around the world under the auspices of the Economic and Scientific Section of the U.S. Department of War. While he was in India, working with Mahalanobis, the famous Indian statistician who had founded the prestigious Indian Statistical Institute, he got orders to continue on to Japan. He described those times to me.

> "I stayed in Japan for two months to assist with studies of nutrition, agricultural production, housing, fisheries, etc. In that way I became friends with and learned from some of the great Japanese statisticians. Statistics was well established in Japan."

He is not aware of how there came to be so many learned statisticians in Japan those many years ago, but he remembered that a Dr. Seito had been studying statistics at University College in London when he was there a few years earlier.

> "In 1948, I went again to Japan, this time for the Department of Defense, to do more of what I had done before. I made an effort to talk whenever possible with Japanese statisticians. I would go to the Post Exchange, where I had privileges, and buy food. Then I would

lug it to the Army operated Dai-Ichi Hotel where I had a very small room. If I said it was 8 feet by 9 feet, I wouldn't be far off. Then I would arrange for a private dining room in the hotel and serve the food to my Japanese friends.

"Any food tasted good to them, I'm sure. We'd sit around the table and talk. I had no vision of what was to happen. I just told them that they were important to the country in the reconstruction of Japan. This idea was new to them.

"Now, there is a sub-plot involving a Mr. Ken-ichi Koyanagi, who had earlier been in jail for 8 years -- ostensibly for being a Communist. Whether he had been under house arrest or actually in jail, I don't know. Probably all there was to it was that he had a mind of his own and wouldn't go along with the war lords. I say this because when it came time for him to get a visa later to come to this country, there was no great problem.

"His major in the university was German literature. Most people who rise in management in Japan never have studied Management Science, thank goodness. It's better that they don't.

"In 1947 he formed the Union of Japanese Scientists and Engineers (JUSE) consisting then of about 7 men, their purpose being the reconstruction of Japan. Mr. Koyanagi held the group together. And Dr. Nishibori, who was in the original group and later Chairman of Japan's equivalent of our Atomic Energy Commission, told me that they had nothing much to talk about. They would just eat and drink. Suddenly one night, Dr. Nishibori had the bright idea that statistical methods could help in the reconstruction of Japan. This would be a way of helping that wouldn't require new equipment, which they had no means of obtaining.

"One of the principal problems of Japanese industry at that time was that the captive markets of China and Korea that they had had prior to the war, were now lost to them. And they needed to trade so that they could import food.

"Came in 1949 a letter asking me to teach statistical methods in industry. I couldn't go at that time, though I wished to. I had too many projects going, so I kept stalling. I finally did go in June of 1950 under the auspices of the Supreme Commander of Alled Powers."

THE MAINICHI DAILY NEWS OF TOKYO, on the occasion of the presentation of the Deming prizes on November 10, 1965, described the visit and the conditions in Japan immediately following the war.

"The scholastic contact between Japan and Dr. Deming dated back to April 1950 when Ken-ichi Koyanagi managing director of the JUSE, wrote to Dr. Deming, then in the U.S. asking for lectures on statistical quality control when he visited Japan later in the year. He readily accepted the plea.

"At that time, few Japanese realized the significance of quality control. In the prewar years, there were, indeed, some Japanese scholars and engineers who were engaged in the study of quality control, and some of them attempted to put it into practice. But no company dared to carry out the wholesale introduction of the revolutionary idea.

"After the war, the nation's industry was quick to rise again, but the quality of its products were all but inferior. Faced with enormous demand, manufacturers were all busy in turning out as many products as possible, and no one cared about quality.

"The concept of quality control made inroads into the Japanese industries in the form of an Occupation Forces order to communication equipment manufacturers. When they started to employ the modern production formula, some private organizations paid a deep concern. Soon they stepped into the field and started dissemination activities.

"Independent from these organizations, the JUSE also launched an educational service of quality control in 1948. A series of lectures was sponsored on the subject of statistical analysis of small samples. Several Japanese experts gathered to form a research group, primarily aimed at collecting necessary literature. But these activities had a discouraging result: there was little experience and material available. Still under occupation, Japan was in no position to obtain enough literature and material related to quality control.

"Then came the offer from Dr. Deming to the joy and surprise of all the people concerned. In his first lecture meeting in Tokyo in mid-1950, 230 scholars and statisticians gathered, impressed by the exciting concept of statistical quality control uttered by the U.S. scholar. In another lecture meeting in Fukuoka, 110 were present.

"Dr. Deming called on the students to come out of their studies and, with courage and confidence, go into factories, to keep contact with, and teach, business managers and engineers, and to promote their theoretical research on the application of statistical methods"

He recalls:

"I lectured in English, but I had a wonderful translator, Mr. Hisamachi Kano. His father was a banker, and as a child, he lived in New York, London and Paris, so he learned English and French as he was growing up. He probably learned Japanese at home."

"His English was absolutely perfect, with every kind of idiom. I was very fortunate because I had him with me for the duration of every visit for a period of over ten years.

Dr. Deming described to me the fateful events that involved Japanese higher management in the educational process and provided the critical impetus for changing the image of Japanese products.

"They were wonderful students, but on the first day of the lectures a horrible thought came to me, 'Nothing will happen in Japan; it'll be a farce unless I talk to top management.' By that time I had some idea of what top management must do. There are many duties to be performed that only the top people can do: consumer research for example, work with vendors just for example. I knew that I must reach top management. Otherwise it would just be another flop as it was in the states.

I immediately talked to American friends who knew the right Japanese and before long, I was talking to Mr. Icharo Ishikawa, who had formed the great Kei-dan-ren, the Japanese association of top management.

"I had 3 sessions with Mr. Ishikawa; and at the end of the third session, he understood what I needed to do. He sent telegrams to the 45 or so top level men to come to the Industry Club the next Tuesday at 5 o'clock to hear Dr. Deming. And they all came.

"I did the best I could. I gave them encouragement. That was the main thing. I told them that they could produce quality for the consumer, partly industrial, partly household, for the Western world, in return for food. Conditions were such that they would have to do that.

"They thought that they could not because they had such a terrible reputation when it came to quality. But they knew what good quality was. Ask anybody in our Navy, and they'll tell you that. What they made for military purposes was superb. But for consumer goods, they'd never tried. They didn't know what it was to stand back of any goods. At that time a Japanese item wouldn't last very long; there was no endurance.

"I told them, 'Those days are over. You can produce quality. You have a method for doing it. You've learned what quality is. You must carry out consumer research, look toward the future and produce goods that will have a market years from now and stay in business. You have to do it to eat. You can send quality out and get food back. The city of Chicago does it. The people of Chicago do not produce their own food. They make things and ship them out. Switzerland does not produce all their own food, nor does England.'"

"Incoming materials were terrible, off gauge and off color, nothing right. And I urged them to work with the vendors and to work on instrumentation. A lot of what I urged them to do came very naturally to the Japanese, though they were not doing it. I said, 'You don't need to receive the junk that comes in. You can never produce quality with that stuff. But with process controls that your engineers are learning about, specifications as loose as possible, consumer research, redesign of products, you can. Don't just make it and try to sell it. But redesign it and then again bring the process under control. The cycle goes on and on continually, with ever increasing quality.'

"I knew the problems because I'd been at Aberdeen Proving Ground, working there for the War Department, with people in industry. And look at the Census Bureau. It was one of the largest organizations to be immersed in quality.

"One of the big problems of management is to define quality and realize that there are several facets. One is what you're trying to do for the future, whatever quality you're aiming at. Should your purchasing agent continue to buy this kind of paint, or should he switch? But also, how about turning out product today? What is the plant manager's job today?

"Now only the management can work on that problem of defining quality. It's a complicated problem with no easy solutions, but it's management's responsibility.

"I tried to explain these things to them, and apparently they understood. They wanted more conferences, so we had more. It was a terrifying experience for me because I was new at it. I was a technical man.

"I told them they would capture markets the world over within 5 years. They beat that prediction. Within 4 years buyers all over the world were screaming for Japanese products.

"I was back in Japan in 6 months, in Janury of 1951. They already had had many brilliant successes, brilliant fires, just as they had had here during the war. But that's not quality; those are just dividends. The top management showed me what they were doing. Mr. Nishimura, President of the Furukawa Electric Company, was himself working to evaluate the process that produced insulated wire. He brought control charts to show me, and he was able to reduce the amount of rework to 10% of what it had been.

Mr. Tanabe, President of the Tanabe Pharmaceutical Company, was working himself in quality control. In a few months he was producing 3 times as much para aminosalyscilic acid as before, with the same machinery,, by just improving the system.

"But you cannot improve the system until you've achieved statistical control. Then engineers and chemists can see that it will stay this way until they make some changes.

"Now six months later here were these members of top management showing me what they had done. Six months after that trip, I was there again, and a year later there again. They were working hard, and they were getting results. I made it clear to them in those first conferences that this must be company wide. 'Everybody in the company has a job to do to improve quality. And as you improve quality, your productivity will go up. Your customers will be happy, and you'll have something to sell.'

"I also told them 'This movement must be nationwide. You must teach other companies, teach your competitors, move along together. As you learn, tell others.' I didn't have to tell them that. That was the natural Japanese way of working But I did tell them anyway.

"By the time I'd made several trips to Japan, Juse was able to teach hundreds of people. They had courses for people outside of Tokyo in the daytime and courses in Tokyo in the evening for people who were working there during the day. There were also courses for management. They trained almost 20,000 engineers in rudimentary statistical methods within 10 years. These courses today are booked up seven months ahead.

Clearly, the Japanese appreciate what Ed Deming and statistical quality control have done to change their destiny. The MAINICHI DAILY NEWS describes the history of the Deming Prize, which symbolizes this appreciation.

"The Deming Prize was created in 1951 by the Union of Japanese Scientists and Engineers (JUSE) to commemorate the friendship and contribution of Dr. Deming to the whole spectrum of Japanese industry. The prize has played a significant role to give an impetus to industry in its dazzling growth.

"The Deming Prize is a silver medal. Designed by Professor

Kiyoshi Unno of Tokyo University of Fine Arts and some other artists, the medal bears an engraved profile of Dr. Deming.

"The Deming Prize is divided into three categories. The prize for research and education is awarded to those who made excellent researches in theory and application of quality control. Another prize for application is given to corporations or plants which attained outstanding results in practicing quality control. The third prize is provided for smaller enterprises.

"The prize has been awarded annually ever since 1951. The Deming Prize Committee is responsible for the selection of the winners from among a number of candidates. Parallel with the progress in Japan in the concept of quality control, the selection standard has been rising year after year, and the race for the laurels has become keen. It is said that most corporate candidates are spending years in streamlining and reinforcing their quality control setup under the guidance of specially invited experts before they apply for the prize."

Business Week, on page 21 of a special advertising section on "Japan: Quality Control and Innovation" of July 20, 1981, lists the winners of the Deming Prize for Application for the years 1954 through 1980 and discusses its impact.

"Each year the competition grows in intensity, as more and more companies volunteer to undergo the close scrutiny required. For the firm that wins the Prize, and those that gain one of the associated awards, however, the rewards are significant, in profits as well as prestige. For other companies, the ceremony is a time for self-reckoning. The innovations in quality-control honored in any year usually soon become national norms."

ANALYSIS OF FERROGRAPHIC ENGINE WEAR DATA USING QUALITY CONTROL TECHNIQUES

Robert L. Launer and Edward A. Saibel
U.S. Army Research Office
Research Triangle Park, N.C. 27709

1. Background.

It is generally accepted that wear is the leading factor in engine and gear failure. There are many types of wear, some of which are: adhesion, abrasion, corrosion, erosion, fretting, cavitation, fatigue, melting, ablution and delamination. Each of these results in its characteristic form of wear particle, the identification of which is sometimes difficult. There are many methods for indentifying these particles and for monitoring their development over time. One such method is ferrography.

Ferrography is a technique developed by Seifert and Wescott for separating wear particles from the lubricant matrix and depositing these on a glass slide, arranged or sorted by particle size [4, 7]. This slide is then examined microscopically. An indirect measure of wear is obtained by measuring the amount of light which is transmitted through the glass slide, subject to the amount of particles which have been deposited. The transmittance is reported as the percentage of the area within the field of view which is covered by the deposited particles. Measurements are made in areas on the slide corresponding to the large particles and to the small particles. The two measurements are called by workers in the field, A_L and A_s respectively. The particles are deposited by dripping the engine or transmission oil onto an inclined glass slide which is immersed in a magnetic field. The larger particles are thus deposited first and the smallest particles, last. A good survey of this method is presented in [6].

2. Statement of Objective.

Our objective is to produce an easy to use method for improving the amount of information which can be obtained in Ferrography without an increase in time, effort and instrumentation. As things stand now, optical measurements are made from the ferrogram deposit and an index of wear severity, I_s is calculated using an arbitrary relationship

$$I_s = A_L^2 - A_s^2 \tag{1}$$

where A_L is percent area covered by particles at the entry deposit, (particles greater than 5 μm) and A_s is the percent area covered by particles at 50 mm from the exit of the Ferrogram, (particles ranging from 1 to 2 μm.) [6]. This index of severity, proposed by V. Wescott, is attractive because of its conciseness and the ease with which it is calculated. Since it contains only information obtained directly from the Ferrogram, this measure is apparently germane and relevant.

3. Brief Discussion of Current Methodology.

As a direct measure of wear, I_s is difficult to interpret. Let $\dot{I}_s$, $\dot{A}_s$ and $\dot{A}_L$ represent the time derivatives of I_s, A_s, and A_L. Notice that,

$$\frac{\partial I_s}{\partial A_s} = -2A_s < o, \text{ and } \quad \frac{\partial I_s}{\partial A_L} = 2A_L > o,$$

so that $\dot{I}_s = 2A_L\dot{A}_L - 2A_S\dot{A}_S$. Therefore, a net positive change in I_s can result from either an increase in A_L or a decrease in A_s. In general, simultaneous increases and/or decreases in A_L and A_s in differing amounts may result in either an increase or a decrease in I_s. In the following, we propose a change in this index which will produce a direct measure of the ferrogram information which is easy to compute and interpret.

A Ferrogram is an indirect measure of engine wear at a specific time so that, for practical purposes, it can be considered a monitoring process. The onset of failure is signalled by a fairly abrupt increase in A_L or A_s or both. Early failures are indicated by premature deviations from the normal values or trends in one of these parameters. It would be very useful to devise a scale for plotting Ferrogram values with automatic warning limits so that interpretation of individual cases could be reduced to a minimum. If this were accomplished with a preliminary sample or other past history (such as factory test data) to establish benchmarks, we will have described a quality control monitoring process.

Ferrogram measurements exhibit unpredictable variation which demands a statistical analysis for proper interpretation. Although the statistical distributions of A_L and A_s are somewhat normal in appearance [5], we suggest that several repeated measurements be taken of each value from each Ferrogram, yielding average $\bar{A}_L$ and $\bar{A}_s$, so that the assumption of normality may be justified by invoking the central limit theorom [3]. Since these measurements are taken from the same Ferrogram, there is the possibility of correlation between them. $\bar{A}_L$ and $\bar{A}_s$ are, however, related to the extreme values of the available measurements from the Ferrogram and are, therefore, related to the extreme order statistics. Since the extreme order statistics are asymptotically independent [2], $\bar{A}_L$ and $\bar{A}_s$ are assumed to be independent. We will, nevertheless, present a method which will allow for correlation between them.

4. Proposal for an Improved Method.

Let

x = sample average, $\bar{A}_L$, at time t

y = sample average, $\bar{A}_s$, at time t

$\mu_x(t)$ = expected value of $\bar{A}_L$ at time t

$\mu_y(t)$ = expected value of $\bar{A}_s$ at time t

σ_x^2 = variance of $\bar{A}_L$

σ_y^2 = variance of $\bar{A}_s$

ρ = correlation between $\bar{A}_L$ and $\bar{A}_s$

Under the assumptions stated previously, the joint statistical distribution of x and y is

$$f(x,y) = (2\pi\sigma_x\sigma_y\sqrt{1-\rho^2})^{-1} \exp\left[-g(x,y)/2(1-\rho^2)\right] \tag{2}$$

where

$$g(x,y) = \left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 . \tag{3}$$

The appropriate regions for monitoring sample values (x,y) are the ellipses of equal probability density, for the probability

$$\alpha = \exp\left[-a^2/2\right], \tag{4}$$

$$\iint_A f(x,y)\,dx\,dy = 1 - \exp\left[-a^2/2\right], \tag{5}$$

where A is the region enclosed by the ellipse [1].

$$g(x,y) = a^2 \ (1-\rho^2) \quad . \tag{6}$$

In order to standardize the graphical representation of the sample values, it is recommended that the ellipses (6) be transformed to unit circles as follows. When x and y are independent the ellipse (6) becomes

$$\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 = a^2 \tag{7}$$

Let

$$ar = (x-\mu_x)/\sigma_x \quad \text{and} \quad as = (y-\mu_y)/\sigma_y \tag{8}$$

Then (7) becomes

$$r^2 + s^2 = 1 \ . \tag{9}$$

The new index of severity, J_s, is made by transforming the data thus:

$$u = \frac{x-\mu_x}{\sigma_x} \quad \text{and} \quad u = \frac{y-\mu_y}{\sigma_y} \tag{10}$$

Then,

$$J_s = u^2 + v^2 \tag{11}$$

is the proposed new index of severity. This value should be compared to an extreme upper tail percentage point of the central chi-square distribution with 2 degrees of freedom. For example, the probabilities =.01 and .005 correspond to the values 9.2 and 10.6, respectively. If there is correlation present these values are reduced, with the lower bounds 6.6 and 7.9 corresponding to perfect correlation. On the other hand, the onset of failure is marked by instability of the distribution of the particle sizes with A_L, A_S or both rapidly becoming very large, depending on the underlying cause of the failure. This implies that these measures change separately or independently, so that the deflation of these values due to correlation would tend to emphasize them during the onset of failure. The appropriate critical values in either case, then, could be obtained from the chi-square distribution with 2 degrees of freedom. We suggest using the value 10 (or 9 if the user is conservative) for the critical value of J_s.

5. <u>A Numerical Example.</u>

The foregoing development might appear somewhat complicated, although J_s is only slightly more complicated than (1). We maintain that J_s contains more engine history and therefore more information on which to base automated decisions. We further suggest that J_s and the related preceding formulation can be easily computed with a handheld computer or even programmed for a microcomputer. The following example will illustrate the point.

Suppose that it has been determined that a certain helicopter engine is characterized by $\mu_x(t) = 15 + .00625t$, $\mu_y(t)=6+.003t$, $\sigma_x^2=11.1$, $\sigma_y^2=2.75$, and t is engine operation time in hours. Suppose further that the engine Ferrogram measurements at 600, 650, and 700 hours are $(A_L, A_S) = (24.8, 6.1)$ (24.1,8.6), and (25.7, 12.7). First we note that $\mu_x(600)=18.75$ and $\mu_y(600)=7.8$. Then

$$J_s = \left(\frac{24.1 - 18.75}{3.33}\right)^2 + \left(\frac{8.6 - 7.8}{1.66}\right)^2 = 4.35$$

The other values of J_s and the values of I_s are calculated similarly and are given in the table below. Notice that while the successive values of I_s decrease, the third value of J_s exceeds the critical value, which is a signal of impending failure.

t	J_s	I_s
600	4.35	578
650	2.43	507
700	11.28	499

REFERENCES

1. Abramowitz, Milton and Irene A. Stegun; Handbook of Mathematical Functions; Dover Publications, Inc. (1965).

2. David, Herbert A.; Order Statistics; John Wiley & Sons, 2nd Edition (1981).

3. Parzen, Emanuel; Modern Probability Theory and Its Applications; John Wiley & Sons, (1960).

4. Seifert, W. W., and V. C. Wescott; "A Method for the Study of Wear Particles in Lubricating Oil", Wear, 21 (1972).

5. Engine and Transmission Lubricating System Diagnostics; TEDECO, Bulletin QDM80, (1982).

6. Wescott, Vernon C., "Ferrographic Analysis Recovering Wear Particles - from Machines to Humans"; Naval Research Reviews, V XXX, No. 3, (1977).

7. Wescott, V. C. and W. W. Seifert; "Investigation of Iron Content of Lubricating Oils Using Ferrograph and Emission Spectrometer", Wear, 23 (1973).

ATTENDEES

28TH ARMY DESIGN OF EXPERIMENTS CONFERENCE
20-22 October 1982

and

TUTORIAL ON NON-PARAMETRIC STATISTICS
18-19 October 1982

	NAME	ORGANIZATION	TUTORIAL	CONFERENCE
1.	ADAMS, MARK S.	MERADCOM		X
2.	ADELSON, L.			X
3.	AGEE, WILLIAM	White Sands		X
4.	ASHLEY, WILLIAM L., III	TACOM	X	X
5.	BAKER, WILLIAM E.	ARADCOM	X	X
6.	BATES, CARL	CAA		X
7.	BAUER, C.	NPS		X
8.	BECHHOFER, ROBERT	Cornell Univ		X
9.	BELL, R.	AMSAA		X
10.	BISSINGER, BARNEY	Penn State Univ	X	X
11.	BODT, BARRY A.	BRL	X	X
12.	BOEHNE, R. C.	SSL, CDEC		X
13.	BOX, GEORGE E. P.	Univ of Wisconsin/ Madison		X
14.	BREIMAN, LEO	Univ of California/ Berkeley		X
15.	BRYSON, MARION R.	CDEC		X
16.	BURGE, BOB	WRAIT		C
17.	CASTRO, OSCAR J,	White Sands	X	X

	NAME	ORGANIZATION	TUTORIAL	CONFERENCE
18.	CONOVER, WILLIAM J.	Texas Tech Univ	X	X
19.	COOK, CHARLES H.	MICOM		X
20.	CORREIA, CHARLES A.	Mat Sys Anal Acty, Ft Lee		X
21.	CROW, LARRY H.	Mat Sys Anal Acty, Aberdeen	X	X
22.	CRUESS, DAVID F.	Uniformed Services Univ of Health Sci	X	X
23.	CURRIER, EDWARD F., CPT	TACOM	X	X
24.	DAVIS, H.	SSL, CDEC		X
25.	DAVIS, L.	OTEA		
26.	DiGIORGIO, EMILIO	NPS		X
27.	DOWLING, J.	SSL, CDEC	X	X
28.	DEUTSCHER, WAYNE LTC	Military Asst to Mr. Hollis		X
29.	DUNN, BILL H.	OTEA	X	X
30.	DUTOIT, EUGENE F.	Infantry School	X	X
31.	EFRON, BRADLEY	Stanford Univ		X
32.	ELSMORE, TIMOTHY F.	Walter Reed Inst of Res	X	X
33.	ELY, R. L.	John Hopkins Univ		X
34.	ESSENWANGER, OSKAR M.	MICOM		X
35.	FARMER, JOHN H.	TCATA	X	X
36.	FERNANDEZ	NPS		
37.	FOSTER, JAMES	SSL, CDEC	X	X
38.	FRENCH, STEPHEN A.	OTEA		X
39.	FURMAN, ERIC	SSL, CDEC	X	X
40.	GAVER, DON	NPS		X

	NAME	ORGANIZATION	TUTORIAL	CONFERENCE
41.	GILDENGORIN, VIRGINIA L.	LAIR		X
42.	GRAHAM, BRETON C.	CAA	X	X
43.	HARRIS, BERNARD	MRC		X
44.	HARVEY, X.			X
45.	HOLLIS, WALTER W.	Dep Under Secy of the Army		X
46.	HOLTERMAN, GORDON C.	Log Mgt Ctr	X	X
47.	HOPPE, GEORGE W.	Natl Guard Bureau	X	X
48.	HUNZEKER, WILLIAM, MG	USA Logistics Ctr		X
49.	IRVINE, NELSON	SSL, CDEC	X	X
50.	IRWIN, ROBERT P.	CECOM	X	X
51.	JACKSON, A.	Waterways Exp Sta	X	X
52.	JAYACHANDRAN, TOKE	NPS		X
53.	JOHNSON, RONALD L.	MERADCOM		X
54.	JOLEMORE, KENNETH A., BG	DA DCSLOG		X
55.	KIRBY, DONALD G., MAJ	TCATA	X	X
56.	KLUGE, P.		X	X
57.	KNAUB, JIM	Log Ctr	X	X
58.	KNISS, JIM	AMSAA		X
59.	KYSOR, K.	HEL	X	X
60.	LAPOINT, STEVE	White Sands	X	X
61.	LARSON, HAROLD	NPS		X
62.	LAUNER, ROBERT L.	ARO	X	X
63.	LEE, CLAYTON R.	Log Ctr	X	X
64.	LEHMANN, WILLIAM L.	SSL, CDEC	X	X

	NAME	ORGANIZATION	TUTORIAL	CONFERENCE
65.	LENNOX, WILLARD J.	MRI	X	X
66.	LEONARD, TOM	MRC		X
67.	LUNG, H. (Steve) R.	Elec Prov Ground		X
68.	MAAR, JAMES R.	Research Group, NSA	X	X
69.	MAHER, MARY ANNE	White Sands	X	X
70.	MANN, NANCY R.	UCLA		X
71.	MARDO, JOHN G.	ARRADCOM	X	X
72.	MARUYAMA, RICHARD T.	TRADOC		X
73.	McAFEE, Walter	Elec R&D Comd	X	X
74.	McCLANAHAN, MASON E.	TRADOC		X
75.	McGOWEN, DOUGLAS J.	OTEA	X	X
76.	McLAUGHLIN, GEORGE J.	Def Res Establishment, Valcartier, Quebec, Canada	X	X
77.	MERRITT, TERRY	Log Eval Agency		X
78.	MERVILLE, D.	SSL, CDEC	X	X
79.	MOATS, W. B., COL	Log Directorate AF T&E Comd, Kirtland AFB		X
80.	MOORE, J. RICHARD	BRL		X
81.	NEAL, DONALD	USA Mat & Mechs Res Ctr		X
82.	NIVISON, R. B.			X
83.	NORDSTROM, A.	DARCOM		X
84.	OBAL, JOHN	NPS		X
85.	PARSONS, R.	SSL, CDEC		X
86.	PARZEN, EMANUEL	Texas A&M Univ		X
87.	POWELL, GERARD M.	Natick R&D Labs	X	X

	NAME	ORGANIZATION	TUTORIAL	CONFERENCE
88.	QUINZI, ANTHONY J.	TRASANA	X	X
89.	READ, BOB C.	NPS		X
90.	REICH, DONALD G.	TRADOC		X
91.	RICHARDSON, G.	SSL, CDEC	X	X
92.	ROGERS, JEFF	NPS		X
93.	RUSSELL, CARL T.	Cold Regions Test Ctr	X	X
94.	SAIBEL, EDWARD	ARO		X
95.	SANCHEZ	NPS		X
96.	SCOTT, DAVID W.	Rice Univ		X
97.	SELIG, SEYMOUR M.	Ofc of Nav Res, Dahlgren		X
98.	SHOREY, RUSSELL R.	Director for Weapon Spt		X
99.	SIEGEL, ANDREW F.	Univ of Washington		X
100.	SOVEY, J.			X
101.	STEVENSON, TODD E.	T&E Comd		X
102.	STEWART, PERRY C..	Log Eval Ctr		X
103.	STRATTON, W. F.	MRSA		X
104.	STUART, PAUL, CPT	TCATA	X	X
105.	SWINGLE, DONALD M.	Consultant, Las Cruces, NM	X	X
106.	SYRCOS, GEORGE P.	Navigation Lab Ft Monmouth	X	X
107.	TANG, DOUGLAS	WRAIR	X	X
108.	TANNER, MARTIN A.	Univ of Wisconsin		X
109.	TARTER, MICHAEL	Univ. of California/ Berkeley		X

	NAME	ORGANIZATION	TUTORIAL	CONFERENCE
110.	TAUB, AUDREY F.	Naval Surface Wpns Ctr, Dahlgren, VA		X
111.	TAYLOR, MALCOLM S.	BRL	X	X
112.	TESSMER, JOSEPH M.	Dept of Energy/SPRO	X	X
113.	THOMAS, JERRY	BRL	X	X
114.	THOMAS, MARLIN A.	Naval Surface Wpns Ctr, Dahlgren, VA		X
115.	THOMPSON, WILLIAM R., COL	OTEA		X
116.	THRASHER, PAUL H.	White Sands	X	X
117.	TURNAGE, GERALD W.	Waterways Exp Stn		X
118.	VILLARONGA, RAUL, COL	Log Eval Agency		X
119.	WEINSTEIN, JOE	SSL, CDEC	X	X
120.	WHITNEY, GARY	CECOM		X
121.	WILEY, GORDON S.	SSL, CDEC		X
122.	WING, BURT			X
123.	WINTER, RALPH	SSL, CDEC		X
124.	WITHERS, LANG	OTEA		X
125.	WRIGHT, SUSAN J.	USACAC		X
126.	ZANGMEISTER, BOB	NPS		X

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARO Report 83-2	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) PROCEEDINGS OF THE TWENTY-EIGHTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH, DEVELOPMENT AND TESTING		5. TYPE OF REPORT & PERIOD COVERED Interim Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s)		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Army Mathematics Steering Committee on Behalf of the Chief of Research, Development and Acquisition		12. REPORT DATE June 1983 13. NUMBER OF PAGES 606
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited. The findings in this report are to be construed as Official Department of the Army position, unless so designated by other authorized documents.

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

18. SUPPLEMENTARY NOTES

This is a technical report from the Twenty-Eighth Conference on the Design of Experiments in Army Research, Development and Testing. It contains most of the papers presented at that meeting. These articles treat various Army statistical and design problems.

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

goodness of fit tests
Smirnov-type statistics
combining small samples
countermine systems
target tracking systems
camouflage
estimating mean life
"A" and "B" allowables
response surface design and alphabetic optimality
design and analysis of experiments
time series
circular probable error
red and white noise
mission area deficiencies
random integration methods
loss exchange ratio
factorial desians
nonparametric probability density estimation
logistical supportability
factorial designs
Bayesian reliability estimates
variable kernel estimators
W. Edwards Deming
quality control

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

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)