PROCEEDINGS OF THE TWENTY-FIFTH 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. 80-2

June 1980

Conference on the Design of Experiments in Army Research Development and Testing.

PROCEEDINGS OF THE (TWENTY-FIFTH) CONFERENCE ON THE DESIGN OF EXPERIMENTS

Sponsored by the Army Mathematics Steering Committee

HOST

U. S. Army Natick Research and Development Command

Natick, Massachusetts

17-19 October 1979

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

FOREWORD

The 25th Conference on the Design of Experiments in Army Research, Development and Testing (DOE) was held 17-19 October 1979 at the U. S. Army Natick Research and Development Command in Natick, Massachusetts. This was the second time in the history of these conferences that this Army base has provided the facilities to conduct one of these scientific meetings. The fourth conference in the series was held here. At that time, the base was called the Quartermaster Research and Engineering Center.

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

Several features in this year's agenda pointed out the special significance of this, the Silver Anniversary of these meetings. The program was dedicated to Dr. Francis G. Dressel, formerly Professor of Mathematics at Duke University. For the past 25 years Dr. Dressel has coordinated the conference programming and local arrangements,

and has edited proceedings of this and other AMSC sponsored conferences. The DOE Program Committee invited Dr. Dressel's friends to join in expressing their appreciation for this loyal service.

The Program was to begin with an address by Dr. Frank E. Grubbs, formerly of the U. S. Army Ballistic Research Laboratory, and a renowned statistician. His talk was entitled, "A Quarter Century of Army Design of Experiments Conferences". This was to be followed by an address entitled, "Summarizing the Results of a Series of Experiments" to be given by Dr. William G. Cochran, Professor of Statistics Emeritus, Harvard University. Unfortunately, both of these gentlemen were unable to attend the conference. Their addresses were read, respectively, by Dr. Dressel and Dr. Herman Chernoff, Professor of Mathematics at the Massachusetts Institute of Technology. (Dr. Grubbs was also unable to serve as Master of Ceremonies at the banquet and to make the presentation of the Samuel S. Wilks Memorial Medal. These duties were taken over by Dr. Robert Launer of the Army Research Office.) The other invited speakers and their topics are noted below.

Speaker and Affiliation	Title of Address
Mr. Al L. May Pillsbury Research Labs	DESIGNED EXPERIMENTS IN SENSORY TESTING
Dr. Ray E. Schafer Hughes Aircraft Company	COMPUTER AIDED HYPOTHESIS TESTS - THE BIRNBAUM TEST

Professor Warren Stewart Mathematics Research Center and University of Wisconsin	NEW ALGORITHMS FOR NONLINEAR LEAST SQUARES AND BAYESIAN PARAMETER ESTIMATION
Professor Marvin Zelen Harvard School of Public Sciences	ETHICS AND STRATEGY IN THERAPEUTIC INVESTIGATIONS

The members of the AMSC take this opportunity to express their thanks to the speakers and other research workers who participated in the meeting; to Colonel H. F. Penny, Commanding Officer of the U. S. Army Natick Research and Development Command, for making available the excellent facilities of his organization for the conference; and to Mr. Donald Kass who so ably handled the details of the local arrangements for this meeting. The AMSC is making available most of the papers presented at this meeting in the present form in order to encourage wider use of modern statistical principles of the design of experiments in research, development and testing work of concern to the Army.

TABLE OF CONTENTS

1979
THE TWENTY-FIFTH ANNIVERSARY OF THE CONFERENCES ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH, DEVELOPMENT AND TESTING

THIS PAMPHLET CONTAINS THE NAMES OF INVITED SPEAKERS
SELECTED FROM THE PROGRAMS OF
THE FIRST TWENTY-FIVE MEETINGS IN THIS SERIES

FOLLOWED BY A LIST OF
RECIPIENTS OF THE WILKS MEMORIAL AWARD

These Conferences Are Sponsored By
The Army Mathematics Steering Committee

U. S. Army Research Office
Research Triangle Park, North Carolina

CONFERENCES ON THE DESIGN OF EXPERIMENTS IN
ARMY RESEARCH, DEVELOPMENT AND TESTING

List of Invited Speakers at
the First Twenty-Five Meetings

First Conference: 19-21 October 1955, Diamond Ordnance Fuze Laboratory and National Bureau of Standards

Professor W. G. Cochran

THE PHILOSOPHY UNDERLYING THE DESIGN OF EXPERIMENTS

Dr. Churchill Eisenhart

THE PRINCIPLE OF RANDOMIZATION IN THE DESIGN OF EXPERIMENTS

Dr. M. E. Terry

FINDING OPTIMUM CONDITIONS BY EXPERIMENTATION

Professor John Tukey (Chairman)

PANEL DISCUSSION ON HOW AND WHERE DO STATISTICIANS FIT IN. (THE OTHERS ON THIS PANEL WERE: MR. CUTHBERT DANIEL, MS. BESSE DAY, DR. CHURCHILL EISENHART, DR. M. E. TERRY, AND PROFESSOR S. S. WILKS).

Dr. W. J. Youden

DESIGN OF EXPERIMENTS IN INDUSTRIAL RESEARCH AND DEVELOPMENT

Second Conference: 17-19 October 1956, Diamond Ordnance Fuze Laboratory and the National Bureau of Standards

Dr. C. A. Bennett

THE PREDESIGN PHASE OF LARGE SAMPLE EXPERIMENTS

Professor R. A. Bradley

RECENT RESEARCH IN STATISTICAL PROBLEMS IN SUBJECTIVE TESTING

Professor B. G. Greenberg

APPLICATION OF ORDER STATISTICS IN MEDICAL EXPERIMENTS

Professor G. E. Nicholson, Jr.

THE PLANNING OF EXPERIMENTS IN THE PRESENCE OF VARIATION

Dr. M. B. Wilk

DERIVED LINEAR MODELS IN THE ANALYSIS OF VARIANCE

Dr. Jerome Cornfield

CHOICE OF ERROR IN THE DESIGN OF EXPERIMENTS

Third Conference: 16-18 October 1957, Diamond Ordnance Fuze Laboratory and the National Bureau of Standards

Professor Benjamin Epstein

LIFE TESTING

Sir R. A. Fisher

PRACTICAL PROBLEMS IN EXPERIMENTAL DESIGN

Professor H. O. Hartley

CHANGES IN THE OUTLOOK OF STATISTICS BROUGHT ABOUT BY MODERN COMPUTERS

Dr. A. W. Marshall

EXPERIMENTATION BY SIMULATION AND MONTE CARLO

Fourth Conference: 22-24 October 1958, The Quartermaster Research and Engineering Center, Natick

Mr. C. I. Bliss

SOME STATISTICAL ASPECTS OF PREFERENCE STUDIES

Professor A. C. Cohen

SIMPLIFIED COMPUTATIONAL PROCEDURES FOR ESTIMATING PARAMETERS OF A NORMAL DISTRIBUTION FROM RESTRICTED SAMPLES

Dr. A. W. Kimball

ERRORS OF THE THIRD KIND IN STATISTICAL CONSULTING

Professor C. F. Kossack

THE AASHO ROAD TEST AS AN EXAMPLE OF LARGE SCALE TESTS

Mr. L. H. C. Tippett

STATISTICAL METHODS APPLIED TO THE TEXTILE INDUSTRY

Fifth Conference: 4-6 November 1959, The U. S. Army Biological Warfare Laboratories, Ft. Detrick

Dr. Joseph Berkson

THE MEASURE OF DEATH

Dr. H. A. David

THE METHOD OF PAIRED COMPARISONS

Dr. D. B. DeLury

SAMPLING IN BIOLOGICAL POPULATIONS

Dr. W. J. Dixon

MEDICAL HEALTH STATISTICS

Dr. N. E. Golovin

PREDICTION OF THE RELIABILITY OF COMPLEX SYSTEMS

Dr. Richard Weiss

THE ARMY RESEARCH AND DEVELOPMENT PROGRAM AS IT RELATES TO THE CIVIL ECONOMY

Sixth Conference: 19-21 October 1960, The Ballistic Research Laboratory

Dr. James R. Duffett

RELIABILITY

Professor F. J. Anscombe

EXAMINATION OF RESIDUALS

Dr. W. S. Connor

DEVELOPMENTS IN THE DESIGN OF EXPERIMENTS

Dr. J. E. Jackson

MULTIVARIATE ANALYSIS ILLUSTRATED BY NIKE-HERCULES:
I. SEPARATION OF PRODUCT AND MEASUREMENT VARIABILITY
II. ACCEPTANCE SAMPLING

Professor G. E. P. Box (Chairman)

PANEL DISCUSSION ON COMMON PITFALLS IN THE DESIGN OF EXPERIMENTS. (OTHERS ON THE PANEL WERE MR. CUTHBERT DANIEL, DR. J. S. HUNTER, DR. W. J. YOUDEN AND DR. MARVIN ZELEN).

Seventh Conference: 18-20 October 1961, U. S. Signal Research and Development Laboratory, Ft. Monmouth

Dr. G. A. Watterson

TIME SERIES AND SPECTRAL ANALYSIS

Dr. J. M. Hammersley

MONTE CARLO METHODS

Dr. R. L. Anderson

DESIGNS FOR ESTIMATING VARIANCE COMPONENTS

Dr. G. S. Watson

HAZARD ANALYSIS

Professor Robert M. Thrall (Chairman)

PANEL DISCUSSION ON SIMULATION. (OTHERS ON THE PANEL WERE COL. A. W. DEQUOY, DR. JOHN HAMMERSLEY, MR. JOHN H. MOSS AND DR. GUSTAVE ROBSON).

Eighth Conference: 24-26 October 1962, Walter Reed Army Institute of Research

Professor Egon S. Pearson

A STATISTICIAN'S PLACE IN ASSESSING THE LIKELY OPERATIONAL PERFORMANCE OF ARMY WEAPONS AND EQUIPMENT

Dr. Marvin A. Schneiderman

A GENERAL SURVEY OF SCREENING THEORY

Professor Herman Chernoff

OPTIMAL DESIGN EXPERIMENTS

Dr. R. P. Abelson

AN EXPERIMENTAL DESIGN FOR DECISIONS UNDER UNCERTAINTY

Dr. H. C. Batson

BIO-ASSAY

Dr. Harold F. Dorn (Chairman)

PANEL DISCUSSION ON DIET AND HEART DISEASE. (OTHERS ON THE PANEL WERE MR. JEROME CORNFIELD, AND DR. GEORGE V. MANN).

Ninth Conference: 23-25 October 1963, U. S. Army Missile Command

Professor Solomon Kullback

COMMUNICATION THEORY

Professor Frank Proschan

THE CONCEPT OF MONOTONE HAZARD RATE IN SYSTEMS REALIABILITY

Dr. Churchill Eisenhart

REALISTIC EVALUATION OF THE PRECISION AND ACCURACY OF INSTRUMENT CALIBRATION SYSTEMS

Professor H. O. Hartley

NONLINEAR ESTIMATION

Professor D. B. Duncan

ON THE SIMULTANEOUS ESTIMATION OF A MISSILE TRAJECTORY AND THE ERROR VARIANCE COMPONENTS INCLUDING THE ERROR POWER SPECTRA OF SEVERAL TRACKING SYSTEMS

Professor Boyd Harshbarger (Chairman)

PANEL DISCUSSION ON WHAT TYPE OF STATISTICIANS ARE NEEDED IN RESEARCH AND DEVELOPMENT LABORATORIES. (OTHERS ON THE PANEL WERE DR. E. L. COX, DR. CHURCHILL EISENHART, MR. JOHN L. MCDANIEL, DR. PAUL R. RIDER, DR. WILLIAM WOLMAN AND DR. DONALD A GARDINER).

Tenth Conference: 4-6 November 1964, The Army Research Office, Washington, DC

MAJ GEN Leslie E. Simon (Ret'd)

THE STIMULUS OF S. S. WILKS TO ARMY STATISTICS

Professor Oscar Kempthorne

DEVELOPMENT OF THE DESIGN OF EXPERIMENTS OVER THE PAST TEN YEARS

Professor H. O. Hartley and Professor A. W. Wortham

ASSESSMENT AND CORRECTION OF DEFICIENCIES IN PERT ANALYSIS

Dr. Churchill Eisenhart

SAM WILKS AS I REMEMBER HIM

Dr. W. J. Youden

AN OPERATIONS RESEARCH YARN AND OTHER COMMENTS

Professor John W. Tukey

THE FUTURE OF PROCESSES OF DATA ANALYSIS

Dr. M. G. Kendall

STATISTICS AND MANAGEMENT

Professor Gerald J. Lieberman (Chairman)

PANEL DISCUSSION ON REGRESSION ANALYSIS. (OTHERS ON THE PANEL WERE PROFESSORS ROBERT BECHHOFER, G. E. P. BOX, JACK C. KIEFER AND INGRAM OLKIN).

Eleventh Conference: 20-22 October 1965, U. S. Army Munition Command, Dover. Held on the campus of Stevens Institute of Technology in Hoboken, New Jersey

Dr. Joan R. Rosenblatt

CONFIDENCE LIMITS FOR THE REALIABILITY OF COMPLEX SYSTEMS

Professor J. Stuart Hunter

NONLINEAR MODELS: ESTIMATION AND DESIGN

Professor William C. Guenther

TARGET COVERAGE PROBLEMS

Professor H. O. Hartley

MAXIMUM LIKELIHOOD ESTIMATES FOR THE GENERAL MIXED ANALYSIS OF VARIANCE MODEL

Professor R. E. Bechhofer

PANEL DISCUSSION ON SELECTING THE BEST TREATMENT. (THE OTHER PANEL MEMBER WAS PROFESSOR SHANTI S. GUPTA).

Twelfth Conference: 19-21 October 1966, Harry Diamond Labs, and the National Bureau of Standards

Professor Brian W. Conolly
OPERATIONS RESEARCH

Dr. John Mandel
STATISTICS AS A DIAGNOSTIC TOOL IN DATA ANALYSIS

Professor W. G. Cochran
PLANNING AND ANALYSIS OF OBSERVATIONAL STUDIES

Professor Norman L. Johnson
SAMPLE CENSORING

Thirteenth Conference: 1-3 November 1967, The U. S. Army Mobility Equipment Development Center and the U. S. Army Engineer Topographic Laboratories

Professor Francis J. Anscombe
REGRESSION ANALYSIS

Professor K. A. Brownlee
SOME COMMENTS ON MATCHING

Professor I. J. Good
SOME STATISTICAL METHODS IN MACHINE INTELLIGENCE RESEARCH

Dr. Frank Proschan
MAXIMUM LIKELIHOOD ESTIMATION OF RELIABILITY

Dr. M. B. Wilk
DATA ANALYSIS

Fourteenth Conference: 23-25 October 1968, U. S. Army Edgewood Arsenal

LT GEN William B. Bunker
BROADENING THE HORIZONS OF EXPERIMENTAL DESIGN

Professor Rolf E. Bargmann
STRUCTURE AND CLASSIFICATION OF PATTERNS

Professor Acheson J. Duncan

BULK SAMPLING

Professor Emanuel Parzen

TIME SERIES

Dr. Walter D. Foster (Chairman)

PANEL DISCUSSION ON BULK SAMPLING. (OTHERS ON THE PANEL WERE PROFESSORS ACHESON J. DUNCAN AND BOYD HARSHBARGER AND MESSRS. HENRY ELLNER, GENE RAY LOWRIMORE, JOSEPH MANDELSON AND VERNON H. RECHMEYER).

Fifteenth Conference: 22-24 October 1969, U. S. Army Missile Command

Dr. John E. Condon

RELIABILITY AND QUALITY ASSURANCE

Dr. Nancy R. Mann

SYSTEMS RELIABILITY

Dr. Clifford J. Maloney

A PROBABILITY APPROACH TO CATASTROPHIC THREAT

Professor Richard G. Krutchkoff

THE EMPIRICAL BAYES APPROACH TO THE DESIGN AND ANALYSIS OF EXPERIMENTS

Dr. S. C. Saunders

ON CONFIDENCE LIMITS FOR THE PERFORMANCE OF A SYSTEM WHEN FEW FAILURES ARE ENCOUNTERED

Sixteenth Conference: 21-23 October 1970, U. S. Army Logistics Management Center

Professor Solomon Kullback

MINIMUM DISCRIMINATION INFORMATION ESTIMATION AND APPLICATION

Dr. Richard J. Kaplan

FIELD TESTING

Professor Gary G. Koch

THE ANALYSIS OF COMPLEX CONTINGENCY TABLE DATA FROM GENERAL EXPERIMENTAL DESIGNS AND SAMPLE SURVEYS

Professor A. Clifford Cohen

ESTIMATION IN TRUNCATED POISSON DISTRIBUTIONS WITH CONCOMITANT INTERVALS AND TRUNCATION POINTS

Professor Dana Quade

NONPARAMETRIC ANALYSIS OF COVARIANCE

Seventeenth Conference: 27-29 October 1971, Walter Reed Army Institute of Research

Professor Marvin Zelen

THE ROLE OF MATHEMATICAL SCIENCES IN BIOMEDICAL RESEARCH

Professor Bernard G. Greenberg

RANDOMIZED RESPONSE: A NEW SURVEY TOOL TO COLLECT DATA OF A PERSONAL NATURE

Dr. Geoffrey H. Ball

CLASSIFICATION AND CLUSTERING TECHNIQUES IN DATA ANALYSIS

Professor K. S. Banerjee

HOTELLING'S WEIGHING DESIGNS

Dr. John J. Gart

THE COMPARISON OF PROPORTIONS: A REVIEW OF SIGNIFICANCE TESTS, CONFIDENCE INTERVALS AND ADJUSTMENTS FOR STRATIFICATION

Eighteenth Conference: 25-27 October 1972, U. S. Army Test and Evaluation Command

Professor John Tukey

EXPLORATORY DATA ANALYSIS

Professor G. S. Watson

ORIENTATION ANALYSIS

Professor J. Stuart Hunter

SEQUENTIAL FACTORIAL ESTIMATION

Professor G. E. P. Box

FORECASTING AND CONTROL

Professor Raymond H. Myers

DUAL RESPONSE SURFACE ANALYSIS

Nineteenth Conference: 24-26 October 1973, Headquarters, U. S. Army Armament Command & U. S. Army Management Engineering Training Agency

Professor Jerome Cornfield

BAYESIAN STATISTICS

Professor S. S. Gupta

RANKING AND SELECTION PROCEDURES FOR MULTIVARIATE NORMAL POPULATIONS

Professor H. L. Gray

GENERALIZED JACKKNIFE TECHNIQUES

Professor Frank Proschan

RELIABILITY GROWTH

Professor S. C. Saunders

ACCELERATED LIFE TESTING

Professor W. A. Thompson, Jr.

RELIABILITY OF MULTIPLE COMPONENT SYSTEMS

Twentieth Conference: 23-25 October 1974, U. S. Army Operational Test & Evaluation Agency and U. S. Army Engineer Center at Ft. Belvoir

Dr. Churchill Eisenhart

SAMUEL S. WILKS AND THE ARMY DESIGN CONFERENCES

Professor Solomon Kullback

MULTIDIMENSIONAL CONTINGENCY TABLES

Professor Herbert Solomon
MULTIVARIANT DATA ANALYSIS

Professor H. A. David
ORDER STATISTICS

Professor Gerald Lieberman
RELIABILITY

Professor Robert Bechhofer
RANKING AND SELECTION PROCEDURES

Drs. Marion R. Bryson and William Mallios
MAXIMUM INFORMATION FROM EXPERIMENTS

Twenty-First Conference: 22-24 October 1975, Walter Reed Army Medical Center and the Armed Forces Institute of Pathology

Professor Frederick Mosteller
SUCCESS IN SOCIAL AND MEDICAL EXPERIMENTATION

Professor Edmund A. Gehan
NONRANDOMIZED CLINICAL TRIALS

Professor Paul Meier
RANDOMIZED CLINICAL TRIALS

Professor Seymour Geisser
PREDICTIVE SAMPLE REUSE

Professor Edmond Murphy
NORMALITY AND DISEASE

Twenty-Second Conference: 20-22 October 1976, Harry Diamond Labs, Adelphi, Maryland

Professor J. Stuart Hunter
THE MEASUREMENT PROCESS

Professor Benjamin S. Blanchard

MANAGEMENT OF RELIABILITY

Dr. Carl N. Morris

STEIN'S ESTIMATOR, ITS GENERALIZATIONS AND ITS APPLICATIONS

Professor Robert V. Hogg

ON ROBUST STATISTICAL PROCEDURES

Professor Nozer D. Singpurwalla

ACCELERATED LIFE TESTING

Twenty-Third Conference: 19-21 October 1977, U. S. Army Combat Developments Experimentation Command. Held at the Naval Postgraduate School, Monterey, California

Professor G. E. P. Box

TIME SERIES MODELING

Professor Norman Breslow

CENSORED DATA

Professor Donald P. Gaver

MODELING AND ESTIMATION OF COMPLEX SYSTEM AVAILABILITY

Professor H. O. Hartley (Keynote)

ANALYSIS OF UNBALANCED EXPERIMENTS

Professor Rupert Miller

THE JACKKNIFE: SURVEY AND APPLICATIONS

Twenty-Fourth Conference: 2-6 October 1978, Mathematics Research Center, University of Wisconsin-Madison

Professor Ralph Bradley

SOME APPROACHES TO STATISTICAL ANALYSIS OF WEATHER MODIFICATION

Mr. Bernard Davis (for Professor Richard E. Barlow)

RECENT ADVANCES IN GRAPHICAL TECHNIQUES FOR ANALYZING FAILURE DATA

Professor Norman Draper (Keynote)

RIDGE REGRESSION

Professor Brian Joiner

STATISTICAL CONSULTING

Professor Grace Wahba

DESIGN PROBLEMS IN RECOVERING FUNCTIONS OF TWO OR SEVERAL VARIABLES

Twenty-Fifth Conference: 17-19 October 1979, U. S. Army Natick Research and Development Command

Professor William G. Cochran (Keynote)

SUMMARIZING THE RESULTS OF A SERIES OF EXPERIMENTS

Dr. Frank E. Grubbs

A QUARTER CENTURY OF ARMY DESIGN OF EXPERIMENTS CONFERENCES

Mr. Al L. May

DESIGNED EXPERIMENTS IN SENSORY TESTING

Dr. Ray E. Schafer

COMPUTER AIDED HYPOTHESIS TESTS - THE BIRNBAUM TEST

Professor Warren Stewart

NEW ALGORITHMS FOR NONLINEAR LEAST SQUARES AND BAYESIAN PARAMETER ESTIMATION

Professor Marvin Zelen

ETHICS AND STRATEGY IN THERAPEUTIC INVESTIGATIONS

SAMUEL S. WILKS MEMORIAL MEDAL

The Samuel S. Wilks Memorial Medal Award, initiated in 1964 by the U. S. Army and American Statistical Association jointly, is administered by the American Statistical Association, a non-profit educational and scientific society founded in 1839. The Wilks Award is given each year to a statistician and is based primarily on his contributions to the advancement of scientific or technical knowledge in Army statistics, ingenious application of such knowledge, or successful activity in the fostering of cooperative scientific matters which coincidentally benefit the Army, the Department of Defense, the U. S. Government, and our country generally.

The Award consists of a medal, with a profile of Professor Wilks and the name of the Award on one side, the seal of the American Statistical Association and name of the recipient on the reverse, and a citation and honorarium related to the magnitude of the Award funds. The annual Army Design of Experiments Conferences, at which the Award is given each year, are sponsored by the Army Mathematics Steering Committee on behalf of the Office of the Chief of Research and Development, Department of the Army.

The funds for the S. S. Wilks Memorial Award were donated by Philip G. Rust, retired industrialist, Thomasville, Georgia.

RECIPIENTS OF THIS AWARD

NAME	DESIGN CONFERENCE (and year) PRESENTED
Dr. Frank E. Grubbs *Ballistic Research Laboratory*	*Tenth (1964)*
Professor John Tukey *Princeton University*	*Eleventh (1965)*
Major General Leslie E. Simon *United States Army (Ret'd)*	*Twelfth (1966)*
Professor W. G. Cochran *Harvard University*	*Thirteenth (1967)*
Professor Jerzy Neyman *University of California-Berkeley*	*Fourteenth (1968)*
Dr. W. J. Youden *National Bureau of Standards (Ret'd)*	*Fifteenth (1969)*

Professor George W. Snedecor Iowa State University (Ret'd)	Sixteenth (1970)
Professor R. W. Dodge Rutgers University and Bell Telephone Laboratories	Seventeenth (1971)
Professor G. E. P. Box University of Wisconsin	Eighteenth (1972)
Professor H. O. Hartley Texas A&M University	Nineteenth (1973)
Mr. Cuthbert Daniel Consultant	Twentieth (1974)
Professor Herbert Solomon Stanford University	Twenty-First (1975)
Professor Solomon Kullback George Washington University (Ret'd)	Twenty-Second (1976)
Dr. Churchill Eisenhart National Bureau of Standards	Twenty-Third (1977)
Professor William H. Kruskal University of Chicago	Twenty-Fourth (1978)

SAMUEL S. WILKS MEMORIAL MEDAL

The Samuel S. Wilks Memorial Medal Award, initiated in 1964 by the U. S. Army and American Statistical Association jointly, is administered by the American Statistical Association, a non-profit educational and scientific society founded in 1839. The Wilks Award is given each year to a statistician and is based primarily on his contributions to the advancement of scientific or technical knowledge in Army statistics, ingenious application of such knowledge, or successful activity in the fostering of cooperative scientific matters which coincidentally benefit the Army, the Department of Defense, the U. S. Government, and our country generally.

The Award consists of a medal, with a profile of Professor Wilks and the name of the Award on one side, the seal of the American Statistical Association and name of the recipient on the reverse, and a citation and honorarium related to the magnitude of the Award funds. The annual Army Design of Experiments Conferences, at which the Award is given each year, are sponsored by the Army Mathematics Steering Committee on behalf of the Office of the Chief of Research and Development, Department of the Army.

The funds for the S. S. Wilks Memorial Award were donated by Philip G. Rust, retired industrialist, Thomasville, Georgia.

RECIPIENTS OF THIS AWARD

NAME	DESIGN CONFERENCE (and year) PRESENTED
Dr. Frank E. Grubbs *Ballistic Research Laboratory*	*Tenth (1964)*
Professor John Tukey *Princeton University*	*Eleventh (1965)*
Major General Leslie E. Simon *United States Army (Ret'd)*	*Twelfth (1966)*
Professor W. G. Cochran *Harvard University*	*Thirteenth (1967)*
Professor Jerzy Neyman *University of California-Berkeley*	*Fourteenth (1968)*
Dr. W. J. Youden *National Bureau of Standards (Ret'd)*	*Fifteenth (1969)*

Professor George W. Snedecor Iowa State University (Ret'd)	Sixteenth (1970)
Professor R. W. Dodge Rutgers University and Bell Telephone Laboratories	Seventeenth (1971)
Professor G. E. P. Box University of Wisconsin	Eighteenth (1972)
Professor H. O. Hartley Texas A&M University	Nineteenth (1973)
Mr. Cuthbert Daniel Consultant	Twentieth (1974)
Professor Herbert Solomon Stanford University	Twenty-First (1975)
Professor Solomon Kullback George Washington University (Ret'd)	Twenty-Second (1976)
Dr. Churchill Eisenhart National Bureau of Standards	Twenty-Third (1977)
Professor William H. Kruskal University of Chicago	Twenty-Fourth (1978)

A G E N D A

THE TWENTY-FIFTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH, DEVELOPMENT AND TESTING

17-19 October 1979

Host: The Natick Research and Development Command

Held: Auditorium, Administration Building, Natick Research Laboratories, Natick, Massachusetts

***** Wednesday, 17 October *****

0815-0915 REGISTRATION -- Lobby, Administration Building

0915-0930 CALLING OF THE CONFERENCE TO ORDER -- Lobby

Donald Kass, Chairman of Local Arrangements, Natick Research and Development Command

WELCOMING REMARKS

COL Robert J. Cuthbertson, Natick Research and Development Command

0930-1200 GENERAL SESSION I

CHAIRMAN - Douglas B. Tang, Chief, Department of Biostatistics/ Applied Mathematics, Walter Reed Army Institute of Research, Washington, DC

0930-1030 Special Silver Anniversary Address

A QUARTER CENTURY OF ARMY DESIGN OF EXPERIMENTS CONFERENCES

Frank E. Grubbs, Program Committee Chairman, Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland

1030-1100 BREAK

1100-1200 GENERAL SESSION I (Continued)

KEYNOTE ADDRESS

SUMMARIZING THE RESULTS OF A SERIES OF EXPERIMENTS

William G. Cochran, Department of Statistics, Harvard University, Cambridge, Massachusetts

1200-1330 LUNCH

1330-1500 TECHNICAL SESSION I -- ROBUSTNESS AND OUTLIERS

CHAIRMAN - Francis E. Dressel, Army Research Office, Research Triangle Park, North Carolina

APPLICATION OF ROBUST FILTERING AND SMOOTHING TO TRACKING DATA

William S. Agee and Robert H. Turner, Analysis and Computation Division, White Sands Missile Range, New Mexico

ROBUST REGRESSION ANALYSIS IN PREDICTING CERAMIC STRUCTURAL FAILURE

Donald M. Neal, US Army Materials & Mechanics Research Center, Watertown, Massachusetts

COMPOUND FREQUENCY DISTRIBUTIONS

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

1500-1530 BREAK

1530-1630 GENERAL SESSION II

CHAIRMAN - Donald Kass, Natick Research and Development Command, Natick, Massachusetts

ETHICS AND STRATEGY IN THERAPEUTIC INVESTIGATIONS

Marvin Zelen, Department of Statistical Sciences, Harvard School of Public Sciences, Boston, Massachusetts

1830-1930 SOCIAL HOUR -- Maridon Restaurant

1930 BANQUET -- Maridon Restaurant

***** Thursday, 18 October *****

0900-1030 TECHNICAL SESSION II -- STATISTICAL THEORY

CHAIRMAN - Grady Miller, US Army Materiel Systems Analysis Activity, Aberdeen Proving Ground, Maryland

ON SIMULTANEOUS INFERENCE FOLLOWING A SIGNIFICANT KRUSKAL-WALLIS TEST

Andrew P. Soms, Mathematics Research Center, University of Wisconsin-Madison

SOME BAYESIAN ALTERNATIVES TO SIGNIFICANCE TESTING

Thomas Leonard, Mathematics Research Center, University of Wisconsin-Madison

VARIANCE REDUCTION IN MONTE CARLO SIMULATION

Mark Brown, Florida State University; Herbert Solomon, Stanford University; and Michael A. Stephens, Simon Fraser University, Burnaby, B.C.

1030-1100 BREAK

1100-1200 TECHNICAL SESSION III -- DATA ANALYSIS

CHAIRMAN - Carl Bates, US Army Concepts Analysis Agency, Bethesda, Maryland

SENSITIVITY OF TOLERANCE LIMITS TO SMALL SAMPLE SIZES

James R. Knaub, Jr., US Army Materiel Test and Evaluation Directorate, White Sands Missile Range, New Mexico

THE PERIODIC NATURE OF EXPERIMENTALLY MEASURED DATA

Michael Hacskaylo, Night Vision and Electro-Optics Laboratory, Ft. Belvoir, Virginia

1200-1330 LUNCH

1330-1530 GENERAL SESSION III

CHAIRMAN - Edward Ross, Natick Research and Development Command, Natick, Massachusetts

DESIGNED EXPERIMENTS IN SENSORY TESTING

Alfred T. May, Pillsbury Research Labs, Minneapolis, Minnesota

COMPUTER AIDED HYPOTHESIS TESTS - THE BIRBAUM TEST

Ray E. Schafer, Hughes Aircraft Company, Fullerton, California

1530-1600 BREAK

1600-1700 TECHNICAL SESSION IV -- LEAST SQUARES

CHAIRMAN - Maxson Greenland, Chemical Systems Laboratory, Aberdeen Proving Ground, Maryland

VARIABLE TRANSFORMATIONS IN NONLINEAR LEAST SQUARES PROBLEMS

Aivars Celmins, Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland

ANALYSIS OF DATA WITH THE NONLINEAR LEAST CHI SQUARE ALGORITHM

Richard L. Moore, US Army Armament Research and Development Command, Dover, New Jersey

***** Friday, 19 October *****

0800-0900 TECHNICAL SESSION V -- BIOMEDICAL APPLICATIONS

CHAIRMAN - William E. Baker, Probability and Statistics Branch, Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland

ERROR-TIME RESPONSE PERFORMANCE OF NAIVE SUBJECTS

Michael Hacskaylo and Joseph E. Swistak, Night Vision and Electro-Optics Laboratory, Ft. Belvoir, Virginia

IMAGE INTERPRETATION PERFORMANCE IN FOUR STANDARD TYPES OF AEROGRAPHIC FILM

Ronald L. Johnson, US Army Mobility Equipment Research and Development Command, Ft. Belvoir, Virginia

0900-0930 GENERAL SESSION IV

CHAIRMAN - Frank E. Grubbs, Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland

OPEN MEETING OF THE AMSC SUBCOMMITTEE ON PROBABILITY AND STATISTICS

Douglas B. Tang, Chief, Department of Biostatistics/Applied Math, Walter Reed Army Institute of Research, Washington, DC

0930-1000	BREAK
1000-1100	GENERAL SESSION IV (Continued)

NEW ALGORITHMS FOR NON-LINEAR LEAST SQUARES AND BAYESIAN PARAMETER ESTIMATION

Warren Stewart, Mathematics Research Center, University of Wisconsin-Madison

1100 ADJOURN

* * * * * * * * * *

PROGRAM COMMITTEE

Carl Bates	Walter Foster	J. Richard Moore
George E. P. Box	Frank E. Grubbs	Douglas Tang
Larry Crow	Donald Kass	Malcolm Taylor
Francis Dressel	Robert Launer	Michael White

SPECIAL SILVER ANNIVERSARY ADDRESS: A QUARTER CENTURY OF ARMY DESIGN OF EXPERIMENTS CONFERENCES

Frank E. Grubbs
Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland

Welcome to the 25th Conference on the Design of Experiments in Army Research, Development and Testing. A whole quarter of a century has passed since Sam Wilks recommended that the Army start this series of conferences, and what an excellent idea he had, with all the vision for the future of Army statistics. As we all know, Wilks was a very remarkable man: a gentleman, a good leader, an outstanding scholar and research statistician, a man who also had very vital interests in applications, and he liked to see people work together. Sam travelled much for the Department of Defense and he consulted widely on all probable areas of statistical application for the Government. He missed none of these conferences, and we remember him so well in his role of selecting many of the key statisticians in the universities to participate in these conferences, as we met at the Cosmos Club in Washington, and drank and dined with Sam. I think we have a better pay-off from these statistical conferences than the other DOD conferences, because of the close interface with university statisticians, including, of course, the eminent statistician who gives the keynote address next. To the memory of Sam Wilks we owe so much, and therefore in 1964 we devoted the Army Design of Experiments Conferences to Sam's memory.

This particular conference, the landmark 25th, is also dedicated to our good friend, fellow mathematician and statistician, teacher, and excellent administrator, Dr. Francis G. Dressel. How would the design of experiments conferences ever have survived if

it had not been for Francis ?, who carried always the big part of the load. I am asking him to make a few remarks next. I am glad Francis has trained Bob Launer so well too!

Now how in the world did I ever get saddled with a "Special Silver Anniversary Address", including the fancy title that came from, I might say, a former friend? It is probably because some of my colleagues saw me enjoying being too much a "free lancer" at these conferences, so they thought! I must stretch the exact title of my talk a bit to cover more statistics. Back in the mid-Thirties, I was a timid, very illiterate Southerner, trying to eke out a living by teaching engineering math at (now) Auburn University, and it became starkly clear that we lowly instructors would hardly ever be promoted unless we got a Ph D! But it was also made quite clear to us that getting a Ph D would not make us a better teacher! A quick OR study (not so-called then) convinced us that we should seek something in applied math, as an outlet, and statistics was the subject to study, for it was needed and spreading fast too, for example to our agricultural experiment station problems. I had a friend, who got a Ph D at Cornell University, and had a good job at our experiment station, so that he introduced me to the analysis of variance, which seemed to be a misnomer, and he even alarmed me with the idea of the analysis of covariance! I later heard that Karl Pearson was the greatest statistician of all, but that one R. A. Fisher was not sprinkling holy water on all the things Karl Pearson had done. My friend wanted to know if I had read any of the reputable journals on the subject of statistics, and I hadn't, of course. In fact, in about 1934 a paper by a young genius on the distribution of quadratic forms in a normal system, with applications to the analysis of covariance had appeared in a (strange) journal called

the Proceedings of the Cambridge Philosophoical Society, written by our keynote speaker of today. Later, in the early 1940's when I had been in uniform at the BRL during World War II, we had trouble justifying a journal called "Biometrika" at first, and there was no way the Army would ever approve the Annals of Eugenics in our Library even if it contained statistical papers!

In any event, back in the mid- to late 30's it did seem that statistics would be a good choice of graduate study with an outlet. But where in the United States could one study statistics? He certainly could not do so down South, and in fact there were only two, or maybe three, places to go for statistics courses - Iowa State University, the University of Iowa, and perhaps the University of Michigan. The latter was mostly an acturial school with well-known voices such as Menge and Glover, with Cecil C. Craig and Paul Dwyer coming along, and perhaps most interesting of all an athletic statistician and actuary, Harry Clyde Carver, who would challenge his graduate students to beat him at any sport of their own choice. If Carver won, there would be a stiff final exam and no A's! One had to beat Carver at his own game!

We selected Michigan, for Iowa State University seemed too far away, and Alan T. Craig of Iowa was scheduled to give the basic graduate statistics course at Michigan in the summer of 1937, and what a good start to learn to throw dice, et al! That summer, I tried to learn what a random variable was. I had known Clifford Cohen for years back at Auburn, and at Michigan, Clifford of all things, had elected to write his dissertation on the very obscure subject of truncated sample theory! But how in the world could there ever be much interest, let alone wide applications, of such an odd topic? A colleague, who had treaded this mill before, assured me that "Clifford Cohen was a very

smart man - he chose a topic no one else would ever work on!" The advice continued, "Don't ever pick a 'hot' topic, for a genius will beat you to it, and you'll never get your degree"! The non-statistical graduate students in mathematics frowned at anyone studying statistics, as it was not as important as topology or even differential geometry either. One of them was somewhat friendly though, and on occasion would drop by our "flat" with his wife. He would pick up and continue my wife's needlepoint with much enthusiasm, but didn't care for statistics then, or any part of it! His name - Jimmy Savage (!). And it goes to show you what can happen toat pure, rigorous mathematician, once he is "bitten by the bug" or otherwise the clever ideas of the Reverend Thomas Bayes!

It was not easy to to find and settle upon a dissertation topic without some guidance, but all the professors already had too many graduate students, and they had passed along topics to some they never heard from, so that I had "better look around in the library". No one then told me, for example, that concerning truncated sample theory this would develop into the field of order statistics, and moreover blossom into reliability , life-testing, et al, and it was in fact many, many years before that did occur. You see, no advice I had been given really sunk in, for I decided to work on outliers, and the international situation had gotten so gloomy that writing a dissertation would not be done very quickly anyway, So, being a reserve officer, I was intro- to the Army.

The physical and engineering sciences were just beginning to make some uses of statistics, although Walter Shewhart had made applications of statistical quality control. In 1941 at Michigan I got ahold of a new book, "An Engineers's Manual of Statistical Methods" by one Major Leslie E. Simon, and as I read it and was enlightened

by the book, some correspondence developed, for here was an authority who had the vision, the wisdom and the courage, of all things, at the time of a very low cycle on the Bayesian revolution (so that we later kidded Les Simon that he had the unmitigated gall) to publish in the back of his book some I_Q Charts to estimate the fraction of defectives in a lot by using Bayes' equally likely hypothesis! Today, there seem to be no 100% classical statisticians, so that we can chalk up another win for Les.* And this introduced me to the Army's Ballistic Research Laboratories, on active duty in uniform as a Lieutenant, where there was never to be an end to all kinds of knotty statistical problems.

As pointed out in Les Simon's book, An Engineers Manual of Statistical Methods, Dr. L. S. Dederick had worked out the probability distribution of the sample range (largest minus smallest observation) back in 1926, and had partially tabulated its distribution, but wouldn't submit it for publication! Sam Wilks had on occasion consulted with the personnel of the Ballistic Section of BRL at Aberdeen Proving Ground on various statistical problems. Also, since the dispersion of shots on a target, as from rifle firings, was often measured by the "extreme spread", or bivariate range, this little nasty statistical distribution had eluded statisticians, and Mr. Philip G. Rust, an industrialist and "rifle accuracy bug", established by sampling shot patterns the distribution of the extreme spread for small sample sizes. Also, on the train from Washington to Wilmington, Phil Rust had told Sam Wilks about it and had suggested that he look into the

* In a panel discussion on Bayesian methods on reliability one time, I stated that statistically I was 50% classical. 25% fiducial and 25% Bayesian, but Frank Proschan promptly branded me as a hermaphrodite.

theory of the probability distribution of the extreme spread in order to study its properties, as it was widely used in ballistics and rifle accuracy competititons also.

When I arrived at the BRL at Aberdeen Proving Ground in 1941, John von Neumann, Robert H. Kent, H. R. Bellinson, and B. I. Hart had just worked out and published in the Annals of Mathematical Statistics the distribution of the mean square successive difference, and the mean square successive difference to the variance, and B. I. Hart had calculated percentage points of both.

The "real world data", coming out of ballistic testing of all kinds, often defied any good or "normal" analysis, and were loaded with outliers! There was thus an applicable dissertation topic! In the mid Forties, there existed a critical need to do something about speeding up the production of firing tables, as about 100 female "computers" were always busy running computations on those big, heavy desk electric Friden or Monroe calculators.

Leslie E. Simon valued brains to solve the Army's problems in ballistics, and he had established a scientific advisory committee with some of the best brains in the physical sciences in the country. What a wonderful and stimulating place to work, less much time to be in uniform and fight the "battle of Aberdeen"!

Back in the early Thirties, Simon was Chief of Manufacture at Picatinny Arsenal, and had cultivated the interest and expertise of Walter Shewhart to apply the principles of statistical

quality control to the manufacture of ammunition at Picatinny Arsenal. At the BRL in World War II, Simon saw the pressing need to pass on the principles of statistical quality control to industry in the production of ammunition and weapons for the US Army, and he also worked with Harold Dodge of the Bell Telephone Laboratories to start computations of the standard sampling inspection tables for the Army Ordnance Corps, later put into Military Standard 105A.

Sam Wilks had long been aware of the need for well-designed experiments and hence suggested that the Army start a series of annual conferences to promote statistical methods. Sam suggested that the Design of Experiments Conferences should have three types of sessions: First, there would be some special invited papers by well-known authorities on the philosophy and general principles of statistical design of experiments, then there would be some technical papers presented by Army statisticians, and finally there would be clinical sessions with suggestions from the experts - and we still stick to this format today. These conferences had their beginning 19-21 October 1955 at the Diamond Ordnance Fuze Laboratory and National Bureau of Standards in Washington, D. C. We note that Sam's conferences were Army Wide, and attracted DOD interest, while a conference the Ballistic Research Laboratories put on a year earlier (1954) on the use of statistical methods was primarily for Army Ordnance personnel.

Within Ordnance and the Army, Leslie E. Simon certainly was the great stimulus to the advancement of statistical methods, for

at the BRL Les was not only its Director, but he also prepared a large number of papers on engineering statistics or statistical engineering - what ever you want to call it. Moreover, there was a pressing need for these very papers to acquaint industry to the methods of quality control and statistics in connection with the World War II effort. And Les helped promote the short courses on statistical methods in industry. There was a great deal of interest during this period concerning the concept of "economical lot sizes", and also the concept of producing very large "homogeneous lots" so that for ammunition at least we could get rid of the situation where at a field artillery battery site there existed a mixture of rounds from several or many lots with different levels of muzzle velocity and degrees of surface-finish roughness.

In the mid-nineteen Forties, a very significant and World-Wide development occurred due to an idea of our imminent and esteemed Scientific Advisory Committee member, John von Neumann. He had suggested the construction of the ENIAC or Electronic Numerical Integrator and Calculator /- a digital computer - at the BRL. We saw the handwriting on the wall: The ENIAC could be used to Monte Carlo anything to death, in addition to the more straight-forward computations of mathematical and statistically tractable functions. And so many statistical problems were planned! First, however, there had to be some calculations on the distributions of outliers, and the ENIAC staff was looking for work! It was then that I learned about priorities and the real importance of any statistical problem to the country!

Once they got the ENIAC wired for my outlier problem, the Atomic Energy Commission called on Gen Simon and Johnny von Neumann to use the ENIAC to obtain an optimum solution to the problem of imploding the core of a nuclear warhead, with the result, of course, that even though our Computing Laboratory had begged for work, my suggestion and the start of some extensive calculations to keep them busy immediately got a vanishing priority!

In those days of a great scientific effort at the BRL and many of the Country's great physicists, chemical physicists, mathematicians, engineers, et al, I felt just like Les Simon had always said, "The engineers would call him a statistician, and the statisticians would call him an engineer"! Indeed, we were trying to apply statistical techniques to many knotty physical problems for which there was a phisical model that applied well. Perhaps I would have been much better off in agriculture! We survived some way or the other and hence got away from the use of primarily the probable error which was never to be deleted from firing tables though!

Now, getting back to the uses of Army Statistics, which led up to the Design of Experiments Conferences, we record that a very good account of the statistics in the Army from the very beginning has been prepared by our good friend Clifford Maloney (The American Statistician, June 1962), who traced various statistical interests in one way or the other from very early times - he started out with Daniel Bernoulli in 1777. As pointed out by Maloney, there certainly was much vital interest in medical

statistics of the Army from the beginning, and at West Point the graduates needed to know something about the dispersion of shots on a target, and least squares and the adjustment of data. In fact, in early times, the best engineers in the USA were really coming from West Point. Then again, there was always considerable interest in the sensitivity of explosives to impact or shock, the sensitivity of primers and other items of Ammunition, which no doubt brought about the so-called "Bruceton Method" of sensitivity analysis, and later developed into the Dixon-Mood "Up and Down" technique at Princeton, and since has been widely used. Naturally, Dixon and Mood were students of Sam Wilks, who again enters the general picture! So you see, the Army did indeed have the most natural needs and demands for the application of statistical methods, and Sam Wilks was the first university professor to recognize this vital development for the good of all concerned, as he was always in touch with so many important applications.

I think that the Army Design of Experiments (DOE) Conferences Sam Wilks started have performed the the vital task of fulfilling the need for cross-fertilization of statistical theory and practice, even though these conferences occurred only once a year. It is through the Army DOE conferences that we have become acquainted with each other, discussed common statistical problems, presented solutions to others, learned a lot from the eminent university statisticians and gotten their best suggestions during the clinical sessions. Moreover, this has all stimulated Army

satisticians to perform very good work and publish a number of useful results for others to apply. If it had not been for Wilks' vision and the DOE conferences, we would have been off to ourselves,no doubt,working very much more inefficiently.

Now all of you have the nice little pamphlet prepared by Bob Launer for this, the Twenty-fifth anniversary of the Army DOE conferences, and we note that the series got off with a bang in 1954 as Bill Cochran led with the philosophy underlying the the design of experiments. Churchhill Eisenhart spoke on principles of randomization (Isn't that still an unsettled topic ?) and John Tukey headed a panel on "Where do statisticians fit in?"). Jack Youden aided in his most interesting way of talking about the design of experiments in industrial research and develpoment. The application of order statistics and problems in subjective testing came into the Second DOE conference, and we were fortunate to have R. A. Fisher at the Third conference! Also at the Third conference Ho Hartley spoke on changes in the outlook of statistics brought about by modern computers, and Ben Epstein, who at one time even worked at Frankford Arsenal, covered what was to become a very important Army field "life-testing" - and later reliability and reliability growth. Here at Natick for the Fourth conference, it was appropriate to have L. H. C. Tippett discuss statistical methods in the textile industry, and the Fifth conference taught me a lot about smoking and lung cancer (now forgotten!) because of the lively debates between Joe Berkson of the Mayo Clinic and Jerzy Neyman (both smoking, I believe!)

The theme or title of these conferences was stretched many times, I am reminded, to include many important topics of the day or time, and this was necessary and good too! In fact, I note that Egon Pearson gave the keynote address of the Eighth conference on a statistician's place in assessing the likely operational performance of Army weapons and equipment, or the need for statistics in military operations research and weapon systems analysis. In fact, the Army has a parallel series of conferences, started in 1961, known as the Army Operations Research Symposia. I found that the field of operations research was being staffed primarily by mathematicians, physicists, engineers and others, but not enough statisticians, who could aid in their modelling problems of stochastic processes. For example, for probability of hitting problems there was often the need to have simple approximations to the distribution of quadratic forms in normal variables, and techniques like the Wilson-Hilferty transformation of Chi-square to approximate normality and the Polya-Wilson approximation to those darn cut-off normal integrals that were found to be very useful. Moreover, we also saw that the theories of life-testing would apply to Lanchester type combat theory. Because of the critical need for the evaluation of weapon systems, and later many other military operations research topics, the Army OR symposia have attracted a large number of "high brass" type visitors. Statistical topics have been often discussed at the Army OR conferences and OR topics at the Army DOE conferences. Forget titles!

In the Spring of 1964, some six months before the Tenth conference, we got the shock of our lives with the untimely passing of Sam Wilks. For the Tenth conference, Les Simon came forth with an excellent and informative paper on the stimulus of S. S. Wilks to Army Statistics, and the high importance of the DOE conferences to Army statistical endeavors.

Fortunately, the DOE conferences have proceeded to cover the water front, and stimulate and train more statisticians.

Now although I have mentioned many of the key benefits and much in the way of significant progress that has resulted from the DOE conferences, let me now jump to a look at the whole series, or the view from an operations research eye. To begin with, it becomes quite clear that we have learned a lot about modelling processes (stochastic) or fitting models to data in order to make more general predictions, or to summarize. "Models" ? Yes! And this reminds me of what George Box is quoted as saying, "All models are wrong, (but) some (even) work"! How true this is! Aren't models competitive, and haven't we found that the situation doesn't exist for which only one model is right and all the others wrong? In fact, we are often lucky that any of several competitive models may serve the purpose at hand very well. Yes, I think we have learned how to model many important Army areas of application, and this has also brought about model development or better theories.

Obviously, the great benefit to the Army from the DOE

conferences has been the expert counseling of in-house statisticians by the eminent university statisticians who have so kindly given of their time and experience. One has only to look at the little booklet of featured speakers to be very highly impressed with the caliber of the talent. We greatly appreciate this, as their help, including the clinical session suggestions, has been outstanding, and for very difficult areas of application.

This brings to mind another point. The US Army is a very large and diversified organization. In case you need some converting on this point, just attend one of the Army Science Conferences held biennally at West Point. In addition to our little corners of application we have discussed over the years, at the Army Science conferences, they have presented papers on, for example, sampling the polar ice caps - which brought up many statistical problems of note - or even the extraction and analysis of snake venom! What are the main controversies about? You guessed it: the instrumentation, the measurements and their interpretation. Army investigators has grown increasingly aware of errors of measurement, precision and accuracy, and even just how to define these illusive concepts. And so have others. For example, Committee E-11 of the American Society for Testing and Materials has for some 20-25 years been working on the problem of standardizing the views of engineers, chemists, etc., on the subjects of precision and accuracy, and come forth with a recommended practice. I still don't see an end to this effort, for

there are enough "divinely endowed", stubbornly statistical minds to bring about nothing but impasse's. (Incidentally, I know that I alone am right though, and they needn't think they can sway me to a compromise!). As a passing remark, there's a full time job for a young, competent statistician for NASA, the FAA, and such agencies, in connection with sampling the atmosphere in order to establish temperature profiles, ozone content profiles, etc., by knowing the capability of their instrumentation for the first time.

We have learned much about the statistical design and analysis of scientific type experiments, and the construction of designs - the latter, I think! Furthermore, I see evidence very frequently of some "fancy" experimental designs that Army investigators are using, with very sophisticated analyses, too. On the front cover of the program of the Tenth Conference, there is a 10 x 10 Graeco-Latin Square, and no one yet has pointed out an error in it! When the nice, balanced experiments have been violated in one way or the other, speakers like Hoh Hartley (who regrets that his duties as President of ASA keeps him away today) have come along to help or straighten us out. We have used linear models mostly, but have been hit by nonlinear models at times, and George Box has on several occasions given us his unique approach to time series analysis.

There have been many advances over the years in the analysis of contingency tables, and count data generally. We have had many contributors on this subject speak to us, and the several

approaches presented to us, including Kullback's information theory approach, and some of the recent work of Feinberg in our preceding two-day tutorial course. Perhaps the US Army Operational Test and Evaluation Agency has made much use of contingency table analyses, and have benefitted from them. I will continue to try and sort out that problem, and I note that my experience has been primarily in connection with the comparison of two or more binomial type proportions, and irrespective of Fisher's fixed marginals, and stuff like that, I still don't want to confuse the issue by imbedding the comparison of binomial p's in a contingency table analysis. Maybe the real experts have other views.

Hasn't the field of reliability and related applications hit us with a big bang, to say the least ? And the high-level "brass" or managers have shown the greatest of interest in it too. Remember, I remarked that Clifford Cohen in the late 1930's wrote his dissertation on the obscure subject of truncated sample theory? Well, finally the area came to life and how! Although the normal distribution was the "universal" one in the past, it didn't "take" with the reliability analysts at all, and they aimed for the exponential distribution. At the 1977 Monterey Conference (23rd), a paper was given by Herback, Green and Blumenthal on the "curse" of the exponential model, and they quote:

"The exponential is wrong,
But works like a song.
Beware the Weibull:
It's incorrigible" - Anon

Remember George Box - All models are wrong!

There is some heavy interest in reliability growth, and the whole field of reliability will continue and will continue to expand. There are now so many methods or recommendations for obtaining confidence bounds on system reliability that an appointed committee has not been able to standardize on a technique for DOD. It might be said that sample order statistics are of much importance to the Army nowadays, and often even help to take care of the outlier problem. Finally, reliability analysts have worked on estimation and other properties of the two- and three- parameter Weibull models so much that this has actually aided in the spread of the Weibull distribution to many other areas of application than reliability. Perhaps this is because of the robustness of the Weibull model in representing a variety of shapes.

A very old statistical problem is that of bio-assay types of analysis, and it borders on the estimation of risks and safety levels in any number of other fields. There are many papers on the subject of quantal response, "sensitivity analyses", explosive sensitivity (Bruceton), ballistic limit, Up and Down method, etc. which have been aired in these conferences. Quantal response investigations, and especially the estimation of both high and low percentage points (of unknown distributions), does indeed cover a very important statistical effort for the Army, and it will continue to expand also. Maybe this is an area for which the use of physical models is needed in addition to statistical analyses, or at least a combination of both. We will face more

and more safety type problems, for which there will be heavy demands for statistical treatment, while we used to avoid them altogether. Let me mention one nasty little problem having to do with armor protection in tanks, or penetration mechanics, and hence for safety of tank crews as a result of armor thickness determination and design. It is also a statistical problem the Army continues to need help on to estimate the parameters for zero chance of penetration. Do you like continuous distributions that slowly change to a series of binomial and continuous models of some kind? And even approach a binomial distribution with parameter zero? In this case, we start firing at a piece of armor plate of a certain thickness, and for the high-striking velocities we will (usually) get 100% penetrations of the projectiles through the plate, and there will be a "residual" velocity distribution for the projectiles or pieces of projectiles which have penetrated the plate and come off the back. But as the striking velocity is decreased, then the proportion of projectiles penetrating the plate will decrease, ultimately to zero for low-striking velocities, and thus we say that a safety level exist somewhere, or at least we would like to know just where, for example, only 1 in a 1000 of the projectiles would penetrate. The curve or residual velocity versus striking velocity gets very steep near the bottom, obviously, and it's a challenge to ballisticians and statisticians to deal with the precise and highest striking velocity for which zero penetrations occur. What I am also indicating is that there are many problems of interest for which statisticians and physical scientists must work as team members, and the DOE conferences guarantee just that.

Let us not forget the field of sampling inspection or acceptance sampling inspection, and the DOD's use of standard sampling inspection tables and practices. These are important activities that the Army initiated with the original help of Harold Dodge, and our DOE Proceedings include a number of papers on the subject. This is really the area of statistics that taught us much about operating characteristic curves, or power curves, and the determination of sample sizes, and the like. Thus, many statistical areas of interest spill over into other topics, and so the process continues.

Now I have made my little choices of some of the benefits and topics of value that we have been priveleged to be part of in our twenty-five years of Army Design of Experiments Conferences. Perhaps you can expand or improve on what I have covered and hence make more sense out of things. I invite you to do so. In any event, it certainly seems quite clear that these conferences have been very "cost-effective" to the Army.

I think it was Cliff Maloney who once suggested that there should be published a volume of the best papers of the proceedings of these conferences. This assignment I was given made me look through the whole shelf-wide proceedings, and I agree that there are certain of the papers which indeed should be brought together in some kind of memoirs.

Maybe we can now get Francis Dressell to make a remark or two, as we have dedicated this the landmark 25th Conference to him. Francis!

SUMMERIZING THE RESULTS OF A SERIES OF EXPERIMENTS

William G. Cochran
Professor of Statistics Emeritus, Harvard University

I first met this problem in the thirties in agriculture. I wrote a paper on it (1), and later a more ambitious paper with Yates (2), in which a number of examples were worked. We tried to see in what respects the analysis of a group of experiments resembled and in what respects it differed from the analysis of a single experiment.

The need to summarize results of a series of experiments on the same treatments arises in two types of application. The first type may be described as exploratory; a number of experiments on the relative performance of something or of two treatments have been carried out, and we are trying to answer the question; what is the present state of knowledge about the relative merits of the two treatments? For instance, the recent academy study of saccharin started with the experiments in which large doses were given to rats; these were the prime experiments. To cite a second example, Yates and Crowther realized at the beginning of World War II that Britain would have to import most of her fertilizers during the war and would be short of fertilizers. Accordingly, they summarized the experiments (4) about the responses of the common farm crops to fertilizers in order to answer the question: What is the present state of knowledge about the effects of fertilizers and to provide material for an intelligent rationing system for fertilizers?

As another example, I was in a group that studied two common methods of surgery for duodenal ulcer--vagotomy (cutting the vagus nerves) plus a radical antrectomy (which removes the lower portion of the stomach) versus vagotomy plus the milder pyloroplasty (which widens the outlet of the stomach to provide better drainage).

We found four experiments that appeared to have been carefully done and properly randomized. We could have come across a number of comparisons that were well done but not randomized--the type sometimes called observational studies. Since often we cannot use randomization and have to make a comparison without it, I would have been interested in including the observational studies so as to learn whether they agreed with the randomized studies and if not, why not? But the medical members of our team had been too well brought up by statisticians, and refused to look at anything but randomized experiments. In this type of surgery, we may expect the experiment to be of different designs and perphaps differing numbers of replications.

The second type of application occurs commonly in agriculture. It differs from the first in two ways. It is known that the relative performance of a treatment (variety of a crop or fertilizer) is likely to vary both from field to field within a year and from year to year. Thus experiments are likely to be repeated in different fields and for a number of years. Secondly, there is a better chance that the experiments, being jointly planned, are of the same design and number of replications. For instance, when the growing of sugarbeets was introduced into Britain after World War I, the government conducted 3×3×3 factorials (ultimately 30 per year) at the leading centers for a number of years.

The objective of the experiments may be a series of decisions as to which varieties of a crop look promising and should be kept for further testing, which varieties should be discarded, and which varieties having been fully tested, should be part of an approved list and have their seed made available to farmers. As an example, Patterson and Silvey (5) have described the trials of varieties of cereals that Britain has conducted in recent years, the

designs being incomplete blocks. This kind of screening program is not confined to agriculture. It may be used in seeking the best drugs or vaccines for some purpose in medicine, or in seeking persons best capable of doing some task. In 1963, Federer gave a bibliography of some 500 papers on screening programs.

2. Miscellaneous Experiments in Exploratory Work

I'll start with exploratory experiments done by different people at different places and times. Since these experiments were not planned as a coordinated series, we must expect them to differ in designs, and in numbers of replications. First we must think of the question: of what population, if any, can these experiments be considered something approaching a random sample? Is this population relevant to future applications of any conclusions that we draw? In some cases we may reluctantly conclude that the experiments do not sample any population of interest to us, and decide not to prepare any summary. In some cases the experiments are so variable that some must be thrown out before any summary is attempted. The way in which the experiments were done also affects the nature of the population that they sample. The nature of the experiments also affects the kind of population that they sample. In the National Academy study of saccharin to which I referred, the doses in the laboratory experiments were so large that the estimates of the effects of more normal doses depended to a substantial extent on the kind of model used in extrapolating the experimental results. In experiments comparing two methods of surgery, the experiments may be confined, for ethical or logical reasons, to the kind of patients whose doctors state beforehand that they can safely take either method of surgery. Otherwise, it is difficult to interpret the results of the experiments. This restriction affects the character of the population to whom the conclusions apply.

In agriculture, as I have stated, we have to contend with variations in both space and time: But in other fields of application there may be no strong reasons to consider time as a separate source of variation, even though the experiments will have presumably been done at different times. So in considering a summary of miscellaneous exploratory experiments, I shall combine time and space and speak of treatments x places.

experiments interactions for the ith Treatment in the jth experiment. We may also expect experiments to have different variances σ_j^2 per observation and to differ in number of replications n_j.

For the jth experiment, a model that seems reasonable with a quantitative response is that the mean of the ith Treatment in the jth experiment is

$$\bar{y}_{ij} = \mu + t_i + \gamma_{ij} + \bar{e}_{ij}$$

where γ_{ij} is the treatments × experiments interaction and the variance of the error term e_{ij} is σ_j^2/n_{ij} ($i = 1,2,\ldots,t$; $j = 1,2,\ldots,k$).

In a combined analysis of these means, a reasonable first step is to form a two-way treatments × experiment table of these means. If all treatments are present in all experiments, an analysis of variance into the following components should be easy.

	df
Experiments	(k-1)
Treatments	(t-1)
Treatments × Experiments	(t-1)(k-1)
Pooled error	

The purpose is to test the interaction. If some treatments are missing from some experiments, a least squares analysis appropriate to missing data is used. In this case the Treatments line is Treatments, adjusted for experiments.

The pooled error in the analysis of variance of the treatment means is $(1/k)(s^2/n_{ij})$, or if the s_j^2 seem to be homogeneous, $s^2(1/n_j)$. We will want to examine whether the s_j^2 appear to be heterogeneous, since this affects the F-test of the ratio treatments × experiments/pooled error. For this we can

use Bartlett's test, or if the data seem nonnormal and we want a more robust test, we can use Levene's test, based on the absolute value of the deviations that lead to the s^2, that appears to be less affected by nonnormality. If the s^2 seem markedly heterogeneous, the F-test of the interactions against the pooled error is not exact, but assingning a number of df to the pooled error by Satterthwaite's approximation should provide an approximate test.

The next step is to reach one of three decisions about the Treatments × experiments interactions. (i) that it is negligible, (ii) that it is not negligible but has no discernable structure. By this I mean that although the effects of the treatments vary from experiment to experiment, we have no information for making different predictions in different parts of the population and must draw single overall conclusions about the effects of the treatments. (iii) The third case is that in which the interaction is of a nature that we think we understand, and is large enough so that different treatments win in different parts of the population that can be described. In this case we expect to recommend different treatments for different parts of the population.

Consider first case (i) in which we judge that the treatments × experiments interactions are negligible. If the experiments differ in number of replications and in their error variances, a question to be considered is: Should the treatment means in individual experiments be weighted in forming the overall means, so as to give more influence to the more accurate experiments? If so, what should the weights be? If the error variances σ_j^2 were known, the weights should presumably be $w_j = \sigma_j^2/n_j$, but the variances are only estimated, unless the σ_j^2 appear to be equal so that weights n_j can be

used for the treatment in the jth experiment. Various authors have worked on this problem of weighting with fallible weights.

The first step is to find out if there is much gain in accuracy from the use of weighted means. If the s_j^2 appear to be homogeneous, and the weights are the known values $n_j/\bar{s}_j^2$, this can be done, because the ratio of the variance of the weighted to the unweighted mean of the y_{ij} is $(\sum w_j)(\sum 1/w_j)/k^2$. For instance, if one-third of the experiments each have n_j with relative values 1, 1/2, 1/4, the relative value of the variance of the weighted to the unweighted mean is 36/49 = 0.73. The situation is less favorable to the weighting if the s_j^2 differ, so that we have to use something like estimated weights n_j/s_j^2. Under normality, the maximun likelihood estimate of the overall mean μ_i is

$$\sum_j \frac{n_j(f_j - 1)}{f_j s_j^2 + n_j(\bar{y}_{ij} - \hat{\mu}_i)^2} (\bar{y}_{ij} - \hat{\mu}_i) = 0 \quad .$$

This has to be found iteratively. In this type of estimate, an experiment with low s_j^2 and apparently high precision is prevented from dominating the overall mean if it disagrees markedly from the value suggested by the other experiments, since the term $n_j(\bar{y}_{ij} - \hat{\mu}_i)^2$ will be large, and will decrease the weight given to this experiment.

Some years ago, C. R. Rao (7) brought out a new method of estimating variances and variance components called the MINQUE (minimum norm quadratic unbiased estimator). Since I have been interested in this problem for over 40 years, I asked J.N.K. Rao of Carleton University and P.S.R.S. Rao of the University of Rochester if the MINQUE method would lead to improved estimates

of the weighted mean. Both men looked into the problem--J.N.K. Rao in the case with no treatments $\times$ experiments interactions which is now being considered and P.S. Rao in the case in which we assume a random treatments $\times$ experiments interaction with variance σ_γ^2, which also has to be estimated. Both men discovered what I had suspected in working with MINQUE--that if one is trying to produce an improved method of estimating variance components, it may not be wise to make the estimates unbiased. With unbiased methods one may get variance component estimates that sometimes take negative values and have large variances. Both men produced adjustments to MINQUE that are essentially positive. J.N.K. Rao's method (8) uses non-iterative weights rather similar to the maximum likelihood weights. The weights are

$$w_j = n_j(f_j + 1)/[f_j s_j^2 + n_j(\bar{y}_{ij} - \bar{y}_i)^2] ,$$

where $\bar{y}_i$ is the unweighted mean of the $\bar{y}_{ij}$. Some limited Monte Carlo studies have shown that the weighting does better than the maximum likelihood estimates of the treatments means except when differences in the error variances are extreme. This estimate also does better than MINQUE and better than the simple weights $w_j = n_j/s_j^2$ and is probably the best found thus far.

For estimating the gain in accuracy from the use of erroneous weights like these, the previous figures for the relative accuracy of weighted to unweighted means must be reduced, because of sampling error in the weights. The dampening factor depends both on the average df with which s_j^2 are estimated, and on the amount of heterogeneity in the weights. For the previous example with weights proportional to 1, 1/2 and 1/3 in thirds, and 1.36 if the weights are known, the dampening factor is approximately $(\bar{f} + 6)/(\bar{f} + 8)$, where $\bar{f}$ is the average number of df in s_j^2.

Thus if the s_j^2 have 6 df on the average, the relative efficiency of weighted to unweighted means is estimated as (12)(1.36)/(14) = 1.18--a rather modest gain from weighting. Before resorting to weighting, check also that weighted means apply to the same population as unweighted means. For example, if the weights tend to be high when the mean yields of the experiments are also high, we may conclude that the results for weighted means apply to a population having a higher mean yield than our actual population and decide not to use weighted means.

For comparison between the estimated means of the treatments, we need standard errors. With unweighted means, the estimate of their standard error is $\sqrt{\Sigma(s_j^2/n_j)}/k$. With the experimental error variances of the individual experiments taken as homogeneous, the estimated variance of the mean weighted as n_j is $s^2/\Sigma n_j$. For Rao's estimate with fallible weights, Rao (8) has given a rough estimate of the variance of this weighted mean, which also implies a dampening factor for the fact that fallible weights are being used. The jacknife estimate is another possibility.

When the treatments × experiments interaction is significant, we need to see if we can understand the nature of the interaction. For this, a two-way treatments × experiments table of residuals is helpful. Sometimes there is no winner; different treatments appear to win in different parts of the population, but either we do not fully understand the interaction or do not wish to use it in a recommendation. Sometimes there are two distinguishable parts of the population in which the ranking of the treatment is different, and we understand why. Student (10) cites an example. After a long series of experiments, the Irish Department of Agriculture introduced Spratt-Archer barley as the best suited to the country. In one county the farmers refused

to grow it, claiming that their native barley was superior. In order to convince these farmers, the Department of Agriculture made some special comparisons in this county of the native barley versus Spratt-Archer. To their surprise, the native barley was superior. The reason also became clear. This barley is a quick-starting variety. Now in this county, farming is rather lackadaisical, so that the weeds flourish. The weeds tended to smother the Spratt-Archer barley, which starts slowly, but the native barley, starting quickly, could smother the weeds. Another maxim from this example is make sure the experiments sample the population to which their results will be applied.

If there are two parts which have k_1 and k_2 experiments, the following breakdown of the interaction is relevant

		df
Treatments	(Part I - Part II)	(t-1)
Treatments	Part I experiments	$(t-1)(k_1-1)$
Treatments	Part II experiments	$(t-1)(k_2-1)$

In this breakdown, we expect the first term to be large and the other parts small. In addition, we need to analyze parts I and II separately, in order to see if there are definite treatments differences in each part.

If the interaction is significant and is assumed to be random, the variance of a treatment mean in an individual experiment is $(\sigma_\gamma^2 + \sigma^2/n_j)$, which moves nearer equality because of the term σ_γ^2 but also means that an extra parameter has to be estimated if weighted means are contemplated. In a Monte Carlo study by P.S.R.S. Rao, Kaplan, and Cochran (9) several types of weighted means including a revised MINQUE were included but the unweighted

mean proved very hard to beat, as might be expected, unless σ_γ^2 is small and the variation in the σ_j^2 is extreme. Use of the unweighted mean has the advantage that an unbiased estimate mean of the variance of the overall mean of a treatment is $\Sigma(y_{ij} - \bar{y})^2 / k(k-1)$.

If the original observations are in proportions, remember that a decision, e.g. whether a single overall mean has enough advantage overall over the other means to recommend it, or whether two means should be recommended for different parts of the population, must be made in proportions. If the combined analysis is made in some other scale, such as angles or logits, because it is thought nearer to normality or in some ways more suitable, remember that means in the original proportional scale will be slightly biased when we transfer back. Quenouille (11) has given approximate corrections for this bias, which do not appear to be well known. Let s^2 be our estimate of the variance of $\bar{z}$ (where z denotes the transformed scale), that is, the mean in the transformed scale. If an angular transformation is used, Quenouille's correction for bias in the transformed mean is to increase $\sin^2\bar{z}$ by $\frac{1}{2}(1 - e^{-2\bar{z}})\cos(2\bar{z})$. If logits are used with equal weights, the usual procedure is to take $p = e^z / (1 + e^z)$ when transforming back to p. Quenouille's correction for bias is to add $(n-1)s^2/2n$ to $\bar{z}$ before taking $e^z / (1 + e^z)$.

3. Variations in Both Time and Space

This situation is likely to occur primarily in agriculture. Since the experiments are likely to be jointly planned, they may have the same designs and number of replications, the same experiment being repeated at the same place for four or five years. As mentioned, the number of

years will commonly be limited to at most four or five, since a larger number slows up any recommendations. But the experiments may not have the same numbers of replication - more may have been added in later years. In varietal trials, a new variety may be added in the second or third years, so that different treatments may have different numbers of years at any given time. However, unless the numbers of replications differ greatly, a preliminary analysis of the treatment means will usually be adequate and is fairly easy, although there are extra complications and full least squares may have to be used if some treatments are only present in the later years.

It will usually be necessary to treat the treatments x years variation as random, with variance σ^2_{ty}, even if it does not act like a random variate. A good deal is known about the influence of weather on crops, and we may have found, for instance, that in a good year the best treatments have a greater advantage, so that the treatments x years interaction is definitely not random. But a superior treatment before recommendation, must be superior, on the average, over a span of years, taking σ^2_{ty} into account, since we cannot recommend a different treatment for different years.

The preliminary analysis of variance and the expected values of the means squares are shown below. I have treated the treatments x places interactions as random as well as treatments x years, since this is usually the assumption that has to be made if it is a question of recommending the overall use of one treatment.

	df	Expected value of mean squares
Treatments	(t-1)	$\bar{\sigma}^2 + n\sigma^2_{tpy} + np\sigma^2_{ty} + ny\sigma^2_{tp} + npy\sigma^2_{t}$
T × Y	(t-1)(y-1)	$\bar{\sigma}^2 + n\sigma^2_{tpy} + np\sigma^2_{ty}$
T × P	(t-1)(p-1)	$\bar{\sigma}^2 + n\sigma^2_{tpy} + ny\sigma^2_{tp}$
T × P × Y	(t-1)(p-1)(y-1)	$\bar{\sigma}^2 + n\sigma^2_{tpy}$
Pooled error		$\bar{\sigma}^2$

In presenting the expected values, I have taken the simplest case, in which all experiments are of the same size and design, the symbols n, t, p, and y standing for number of replications, number of treatments, number of places, and number of years. The symbol $\bar{\sigma}^2$ is, of course, the true pooled error variance. The MS_{tpy} is tested against error, and if F is about 1, this mean square may be combined with the pooled error. The expected values are written as if treatments are also random, with variance σ^2_t. If the effects of treatments are fixed, as they usually are, replace σ^2_t by what is usually called $s^2_t = \Sigma(t-\bar{t})^2/(n-1)$.

From the expected values it is clear that the treatments x years and treatments x places interactions are tested by an appropriate F test (approximate if σ^2_j varies from experiment to experiment) against the mean square for the tpy three-factor interaction, and that an unbiased estimate of σ^2_{ty} is $(MS_{ty} - MS_{tpy})/np$. For the main effects of Treatments, no single line in the analysis of variance is a proper error. An unbiased estimate of the error variance for the error of a treatment mean, if interactions are present and random, is

$$MS_{tp} + MS_{ty} - MS_{tpy}$$

and an approximate F test of the treatments mean square may be made by taking $F = MS_t/(MS_{tp} + MS_{ty} - MS_{tpy})$, with Satterthwaite's approximation used to ascribe a number of df to the denominator. However, in a small Monte Carlo study of experiments, Hudson and Krutchkoff (13) found, somewhat surprisingly, that a rival $F = (MS_t + MS_{tpy})/(MS_{tp} + MS_{ty})$ using Satterthwaite, had somewhat better power and recommended it, although it did not approximate the 5% and 1% levels of F when the null hypothesis was true.

Since whether we recommend one treatment, two treatments or suspend judgement for some reason depends mainly on how the treatments vary in effects from place to place, the two-way table of treatments and places deserves careful study. The treatments x places interaction is sometimes heterogeneous; some comparisons of some treatments have a higher mean square interaction than others. Subdivisions of the treatments and places and the treatments x places sum of squares should be tried.

Thus, as we have seen, the summary of a series of experiments calls mainly for experience in the analysis of variance, which we now have. It is well to adopt something of the attitude in exploratory analysis and be on the lookout for anything unexpected, since the nature of the tp interaction is often a hard thing to puzzle out.

REFERENCES

(1) W. G. Cochran (1937). Problem arising in the analysis of a series of similar experiments. Supp. J. Roy. Statist. Soc., 4, 102-118.

(2) F. Yates and W. G. Cochran (1938). The analysis of groups of experiments, J. Agric. Sci., 28, 550-580.

(3) W. G. Cochran et al. (1977). Experiments in surgical treatment of duodenal ulcer. In Costs, Risks and Benefits of Surgery, Bunker, Barnes and Mosteller (eds.), Oxford University Press.

(4) F. Yates and E. M. Crowther (1941). Fertilizer policy in wartime. The fertilizer requirements of arable crops, Emp. J. Exp. Agric., 9, 77-97.

(5) H. D. Patterson and V. Silvey (1979). Statutory and recommended list trials of crop varieties in The United Kingdoms, J. Roy. Stat. Soc. (in press).

(6) F. E. Satterthwaite (1946). An approximate distribution of estimates of variance components, Biometrics Bull., 2, 110-114.

(7) C. R. Rao (1970). Estimation of heteroscedastic variances in linear models, J. Amer. Statist. Assoc., 65, 161-72.

(8) J. N. K. Rao (1979). Estimating the common mean of possibly different normal populations. A simulation study, unpublished manuscript.

(9) P. S. R. S. Rao, J. Kaplan and W. G. Cochran (1979). Estimators for the one-way random effects model with unequal error variances, unpublished manuscript.

(10) Student (1931). Agriculture field experiments, Nature, 14 March, 404.

(11) M. H. Quenouille (1953). The design and analysis of experiments. Charles Griffin and Co. Ltd., London

(12) W. T. Federer (1963). Procedures and designs useful for screening material in selection and allocation, with a bibliography, Biometrics, 19, 553-587.

(13) J. D. Hudson and R. G. Krutchkoff. A Monte Carlo investigation of the size and power of tests employing Satterthwaite's synthetic mean squares. Biometrika, 55, 431-433.

ROBUST FILTERING AND SMOOTHING OF TRACKING DATA

WILLIAM S. AGEE and ROBERT H. TURNER
Mathematical Services Branch
Analysis and Computation Division
White Sands Missile Range, NM 88002

ABSTRACT

Robust methods provide a fresh approach to the problem of treatment of wild observations in filtering and smoothing problems. The robust M-estimates of regression are extended to filtering and fixed lag smoothing employing a pseudo-density of the observations in a conditional mean derivation of the filter and fixed lag smoother. These robust methods have been applied to simulated and real tracking data to obtain improved estimation performance in the presence of wild observations.

INTRODUCTION

Robust filtering and smoothing are a natural extension of the robust M-estimates of regression developed by Huber [1]. The robust M-estimates provide a natural treatment of outlying observations and have been extremely successful in dealing with outliers in other data reduction problems [2] and [3]. The extension of the M-estimate methods provides a fresh approach to the problems caused by outliers in filtering and smoothing applications. Robust methods for estimation are designed to perform well when observations from contaminating distributions are present. The conventional estimation techniques of least squares, maximum likelihood, minimum variance, etc. may become useless when the observations are contaminated by gross outliers or wild data points. When using these estimation methods, outliers are often treated by testing the residuals. If it is decided that a residual is statistically too large, the corresponding observation is declared an outlier and is not processed. These hypothesis testing methods are often successful if only a small number of outliers are present but breakdown for larger proportions of outlying observations. Also, in order for outliers to be detected, they must be relatively large compared to the measurement noise. The detection methods based on testing of residuals are relatively insensitive to small outliers which leads to an inflation of the mean square estimation error. Thus, methods for treating outliers should be evaluated on their ability to achieve a small mean square estimation error as well as their ability to offer protection from gross outliers.

Very little development has appeared on the application of robust estimation techniques to filtering and smoothing. The most notable work in this direction is that of Masreliez and Martin [4]. Their development of the application of M-estimates to the Kalman filter is mainly theoretical. The emphasis here is on the development of some practical results on the application of M-estimates to filtering and smoothing. We have applied these methods to filtering and smoothing of tracking data from trajectory measurement systems at WSMR. Using simulated tracking data we have also performed extensive Monte Carlo evaluation of filtering methods based on M-estimates to determine the conditions for which we can expect to benefit from the application of these methods.

M-ESTIMATES FOR REGRESSION

Given scalar observations y_i, $i = 1, N$ of a linear model

$$y_i = X_i\theta + e_i \tag{1}$$

where X_i is a row vector of known independent variables and e_i is a random error term we want to estimate the unknown p-vector, θ. The M-estimate of θ minimizes

$$\sum_{i=1}^{N} \rho\left((y_i - X_i\theta)/s\right), \tag{2}$$

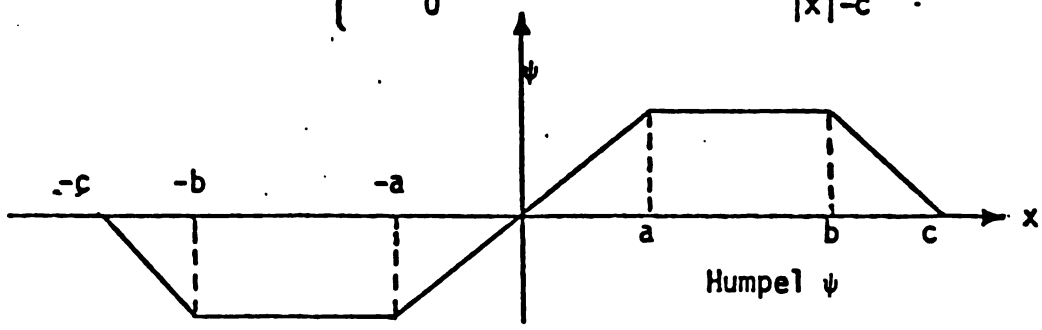
where $\rho(\cdot)$ is a specified function and s is a robust measure of dispersion of the residuals, $y_i - X_i\theta$. Minimizing (2) by differentiating with respect to θ gives

$$\sum_{i=1}^{N} X_i^T \psi\left((y_i - X_i\hat{\theta})/s\right) = 0 \tag{3}$$

where ψ-is the derivative of ρ and $\hat{\theta}$ is the M-estimate of θ. (3) is the analog of the normal equations in least squares estimation. θ is computed iteratively by applying a weighted least squares algorithm to (3). For details see [2].

Rather than specifying the function ρ, M-estimates are usually described by specifying the function ψ. Several ψ functions have been proposed in the literature. The only ψ functions considered here are variations of the one proposed by Hampel [5]. The Hampel ψ function with breakpoints a, b, c, denoted by Ha(a,b,c) is given by

$$\psi(x) = \begin{cases} x & |x| \le a \\ a \cdot \mathrm{sgn}(x) & a \le |x| \le b \\ a(x - c\cdot\mathrm{sgn}(x))/(b - c) & b \le |x| \le c \\ 0 & |x| \ge c \end{cases} \tag{4}$$



Humpel ψ

The M-estimates can also be applied to regression problems having vector observations and to nonlinear regression problems. If the probability density function p of the observations is related to ψ by $p'/p = -\psi$, the resulting M-estimate is maximum likelihood. For any ψ function we call $e^{-\rho}$ a pseudo-density and derive filters in some conventional ways with the density function of the observations replaced with a pseudo-density.

APPROPRIATE NON-GAUSSIAN FILTERING

Assume that the state $x(k)$ of the process being observed is governed by the discrete linear model,

$$x(k+1) = \phi(k+1,k)\, x(k) + u(k) \tag{5}$$

where the state vector $x(k)$ is an n-vector, $u(k)$ is a Gaussian state noise vector with zero mean and covariance $Q(k)$. $\phi(k+1,k)$ is an $n \times n$ state transition matrix. Scalar observations, $z(k)$, of the process are given by

$$z(k) = H(k)x(k) + v(k) \tag{6}$$

where $H(k)$ is a row vector and $v(k)$ is a measurement noise error which may be contaminated by outliers.

In order to derive robust filters corresponding to the M-estimates of regression, we use the results of Masreliez [6] on approximate non-Gaussian filtering. Masreliez obtains an approximate conditional mean of $p(x(k)|Z^k)$ where Z^k is the collection of observations, $Z^k = \{z(1), z(2), ---, z(k)\}$. Using Bayes rule $p(x(k)|Z^k)$ is given by

$$p(x(k)|Z^k) = \frac{p(z(k)|x(k))p(x(k)|Z^{k-1})}{p(z(k)|Z^{k-1})} \tag{7}$$

In order to derive a useful approximation to the conditional mean, $E[x(k)|Z^k]$, of (7), Masreliez assumes that $p(x(k)|Z^{k-1})$ is Gaussian with mean $\hat{x}(k|k-1)$ and covariance $P(k|k-1)$. The resulting approximate conditional mean, $x(k|k)$ is given by

$$\hat{x}(k|k) = \hat{x}(k|k-1) + P(k|k-1)H^T(k)g(z(k)), \tag{8}$$

where $g(z(k))$ is the scalar

$$g(z(k)) = -\bar{p}^1(z(k)|Z^{k-1})\frac{\partial}{\partial z(k)}p(z(k)|Z^{k-1}) \tag{9}$$

Masreliez also derives the second moment, $P(k|k)$, of $p(x(k)|Z^k)$. The result is

$$P(k|k) = P(k|k-1) - P(k|k-1)H^T(k)G(z(k))H(k)P(k|k-1), \tag{10}$$

with

$$G(z(k)) = \frac{\partial g(z(k))}{\partial z(k)} \tag{11}$$

A second method for approximate non-Gaussian filtering is the marginal maximum likelihood filter. In this case we find the estimate $x(k|k)$ which maximizes (7). In this derivation we also assume that $p(x(k)|Z^{k-1})$ is Gaussian. The resulting equation for $\hat{x}(k|k)$ is

$$\hat{x}(k|k) = \hat{x}(k|k-1) - P(x|k-1)H^T(k)\bar{p}^1(z(k)|\hat{x}(k|k))\frac{\partial p(z(k)|\hat{x}(k|k))}{\partial z(k)} \tag{12}$$

The estimate given by (12) is of the same form as the conditional mean estimate given by (8). The difference in the two estimates is that the right hand side of (12) depends on $x(k|k)$ while the right hand side of (8) depends only on the predicted estimate, $x(k|k-1)$. Thus, (12) requires iteration to obtain the estimate and if in the first iteration of (12), we substitute $\hat{x}(k|k-1)$ for $\hat{x}(k|k)$ on the right hand side, the resulting estimate on the first iteration will be identical to the conditional mean estimate of (8). Thus, the maximum likelihood filter may be regarded as a correction to the conditional mean filter.

ROBUST FILTER EQUATIONS

The robust filtering equations corresponding to M-estimates are obtained by replacing the density function, $p(z(k)|Z^{k-1})$, in (9) or (12) by a pseudo-density, $e^{-\rho}$, where $\rho' = \psi\left(\frac{z(k) - H(k)\hat{x}(k|k-1)}{s}\right)$ is a desired influence function. This substitution results in the following equations for the conditional mean robust filter.

$$\hat{x}(k|k) = \hat{x}(k|k-1) + \frac{P(k|k-1)H^T(k)}{s_k}\psi\left(\frac{z(k) - H(k)\hat{x}(k|k-1)}{s_k}\right) \tag{13}$$

$$P(k|k) = P(k|k-1) - \left(\psi'\left(\frac{r(k)}{s_k}\right)\Big/ s_k^2\right)P(k|k-1)H^T(k)H(k)P(k|k-1), \tag{14}$$

where ψ' is the derivative of ψ and $r(k) = z(k) - H(k)\hat{x}(k|k-1)$. The filter equations are completed by the usual Kalman filter equations for the predicted moments.

In order to insure the robustness of the filter described by (13) and (14), the dispersion s_k of the predicted residuals must be specified so that it is insensitive to outliers. We used the MAD estimate of s_k computed from past residuals as

$$s_k = \underset{j=0,N-1}{\text{median}} \left| z(k-j) - H^T(k-j)\hat{x}(k-j|k-j-1) \right| / .6745 \tag{15}$$

where N is a suitably chosen integer.

The robust maximum likelihood filter is obtained by replacing the density $p(z(k)|Z^{k-1})$ in (12) by a pseudo-density. The resulting filter is given by

$$\hat{x}(k|k) = \bar{x}(k|k-1) + \frac{P(k|k-1)H^T(k)}{s_k}\psi\left(\frac{z(k) - H(k)\hat{x}(k|k)}{s_k}\right) \tag{16}$$

In (16) we use $\bar{x}(k|k-1)$ to denote the mean of $p(x(k)|Z^{k-1})$. We use (14) to compute $P(k|k)$ and use (13) to compute $\hat{x}(k|k)$. Several simple methods are available for the iterative solution of (16). The simplest of these is to use

$$\hat{x}^{(\alpha+1)}(k|k) = \bar{x}(k|k-1) + \frac{P(k|k-1)H^T(k)}{s_k}\psi\left(\frac{z(k) - H(k)\hat{x}^{(\alpha)}(k|k)}{s_k}\right), \tag{17}$$

starting with $\hat{x}^{(0)}(k|k) = \bar{x}(k|k-1)$ so that the first estimate $\hat{x}^{(1)}(k|k)$ is the conditional mean estimate, $\bar{x}(k|k)$.

EVALUATION OF THE ROBUST FILTER

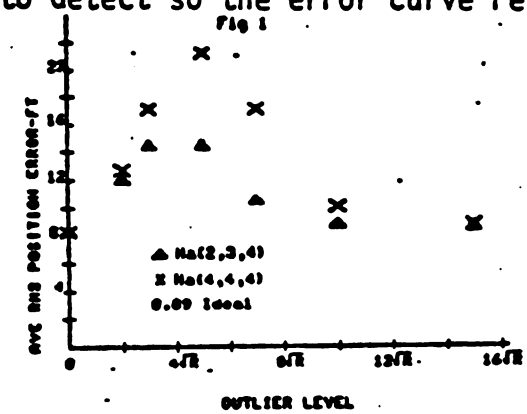
Evaluation of the robust filtering methods described above was done with a view toward eventual application to trajectory estimation. The emphasis in the evaluation was on simulated rather than real trajectory data. This allows a quantitative determination of any advantages in the

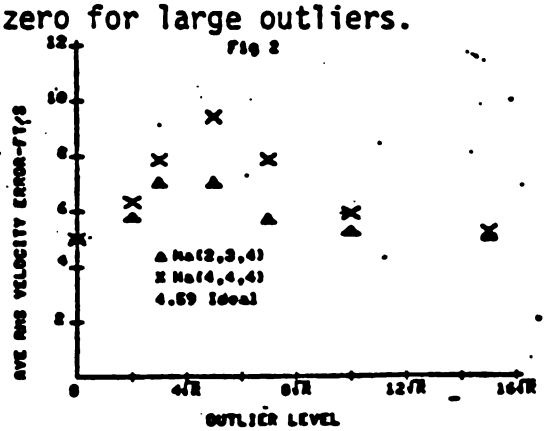
use of robust filtering in the presence of outliers and also any loss in efficiency using robust methods when no outliers are present. The simulated trajectory is that of a constant velocity, level flying aircraft. The measurements are of position in each of three cartesian components with added measurement noise which is contaminated by outliers. The filter model assumes the trajectory to have constant acceleration in each coordinate. The filter for each coordinate has a small acceleration state noise, $Q(k) = 5$. The outlier contamination is controlled by a two state Markov chain with a transition possible at each measurement time. An outlier is added to the measurement if the Markov chain is in state two and no outlier is added if the chain is in state one. The transition probabilities, Pij, are used to determine the percentage of outliers contaminating the measurements and also the length of runs of outliers in the measurements. The magnitude of the outlier contamination is $C \cdot \sqrt{R}$, i.e., a constant multiple of the measurement noise standard deviation.

Using the simulated trajectory data a Monte Carlo evaluation of robust filtering was performed. The rms estimation error was computed pointwise for position, velocity, and acceleration using a sample size of twenty-five. The plots of the rms errors for each of the conditions tested requires far too much space to present here. Instead, these results are summarized by time averages of the rms error in position and velocity for each of the conditions tested.

Figure 1 compares the average rms position error for two filters using the Hampel ψ functions Ha(2, 3, 4) and Ha (4, 4, 4). Figure 2 gives the rms velocity error comparison for the same two filters. Also indicated in Figures 1 and 2 are the ideal rms error values which were obtained with an ordinary Kalman filter with no outliers present and using a known measurement covariance, $R_k = 400.$ The Monte Carlo evaluation of figures 1 and 2 was made with a measurement noise standard deviation of $\sqrt{R_k} = 20$ ft. R_k was unknown to the filter.

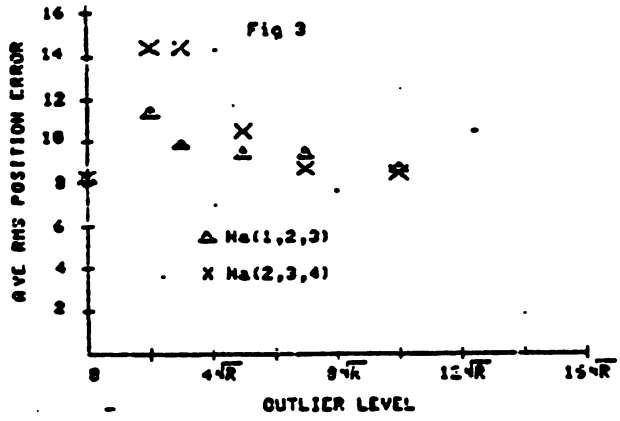
We note from figures 1 and 2 that neither of the robust filters lose much efficiency from the ideal values when no outliers are present. The error curves in figures 1 and 2 behave as expected. Since outliers small in relation to the measurement noise are hardest to detect, the error curve rises sharply. Outliers large relative to the measurement noise are easy to detect so the error curve returns to zero for large outliers.

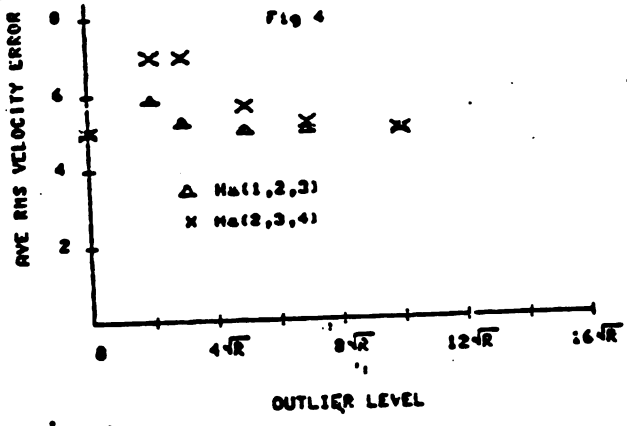




We see from figures 1 and 2 that Ha (2, 3, 4) has a significantly smaller mean square error than Ha (4, 4, 4). Except for the way in which the dispersion of the residuals is measured, i.e., the MAD estimate in (15), Ha (4, 4, 4) is a conventional way of handling outliers in a Kalman filtering application. Using Ha (4, 4, 4) any observation whose predicted

residual is greater than 4 · s is not processed and any observations whose predicted residual is less than 4 · s is processed as an ordinary Kalman filter observation. The above Monte Carlo evaluation was made with Markov chain probabilities $P_{21} = .05$ and $P_{12} = .5$ which gives an outlier probability of .088 and an average outlier run length of three. In order to reduce the average rms errors, we pull in the breakpoints of the Hampel ψ function. Figures 3 and 4 compare the average rms errors in position and velocity for the Ha (1, 2, 3) and Ha (2, 3, 4).





Figures 3 and 4 were made using the same outlier proportions and measurement noise standard deviations as figures 1 and 2.

The iterated filter, i.e., the approximate maximum likelihood filter given by (14), (16), and (17) was also evaluated under the same conditions as the conditional mean filter. Comparison of the average mean square errors for the conditional mean and maximum likelihood filters showed no discernible differences.

ROBUST SMOOTHING

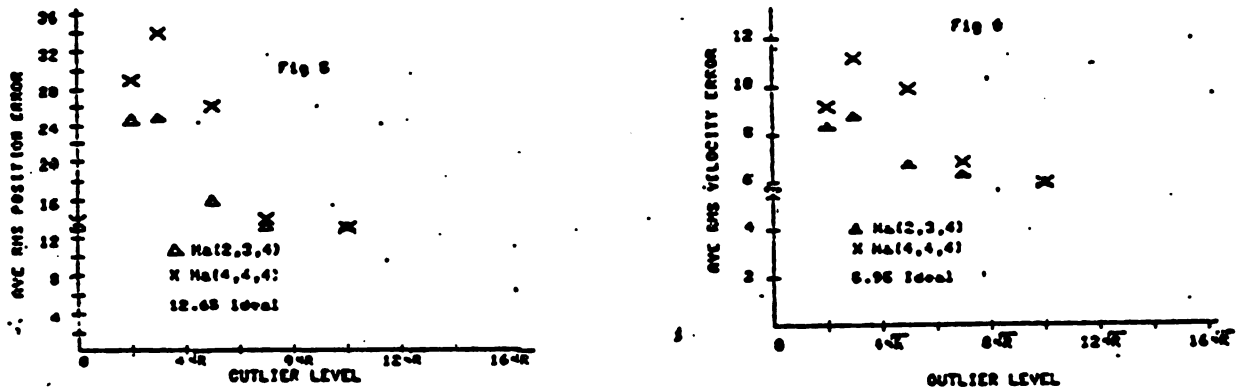
A simplified robust, fixed lag smoother, was derived in a similar manner to the robust conditional mean filter derivation. We derive an approximate conditional mean, $E[x(k)|Z^{k+n}]$, of the density $p(x(k)|Z^{k+n})$ using the same methods and assumptions used by Masreliez and used in the derivation of the robust filter. The robust smoothed estimate is given by

$$\hat{x}(k|k+n) = \hat{x}(k|k)+P(k|k-1)\sum_{j=1}^{n}\frac{\phi^T(t_{k+j},t_k)H^T(k+j)}{s_{k+j}}\psi\left(\frac{Z(k+j)-H(k+j)\hat{x}(k+j|k-1)}{s_{k+j}}\right) \quad (18)$$

In obtaining (18) the fixed lag smoother has been greatly simplified by leaving out the state noise in the forward interval. $x(k|k)$ in (18) is the robust filtered estimate described by (13) and (14). s_{k+j} is a robust measure of dispersion of the residuals, $z(k+j)-H(k+j)\hat{x}(k+j|k-1)$. Several possibilities exist for computing a useful dispersion measure, s_{k+j}. The simplist method and the one used to obtain the smoother evaluation given below is to make s_{k+j} a constant $s_{k+j} = s_k$ and then compute s_k by (15).

A Monte Carlo evaluation of the simplified fixed lag smoother was performed using the same simulated trajectory as was used for the filter evaluation. Measurement noise having a standard deviation of 50 feet was added to the simulated positions. The measurement noise standard deviation was unknown to the smoother. The forward smoothing interval had a length of n = 20 which represents a one second smoothing time. The outlier proportions and run lengths were the same as for the filter evaluation. A sample size of ten was used for the Monte Carlo evaluation of the smoother.

Figures 5 and 6 display the average rms position and velocity estimation errors obtained using the robust, fixed lag smoother with the Hampel ψ functions, Ha (2, 3, 4) and Ha (4, 4, 4). Also noted in figures 5 and 6 is the ideal average rms values which were obtained using the smoother with no outliers and a known measurement covariance, $R_k = 2500$. The robust smoother using Ha (4, 4, 4) is representative, except for the measurement of s_k, of a conventional way of handling outliers in an optimal smoothing application. We note that either of the smoothers offers good protection from very large outliers but that Ha (2, 3, 4) results in a significantly smaller estimation error when small outliers are present.



ROBUST FILTERING USING GAUSSIAN MIXTURES

Robust estimation methods deweight and/or reject outlying observations by specifying observational densities or pseudo-densities having long, heavy tails. We can also realize a long, heavy tailed density by using a Gaussian mixture. Suppose we replace $p(z(k)|x(k))$ in (7) or (12) by the Gaussian mixture pseudo-density,

$$p(z(k)|x(k)) = \sum_i \alpha_i N(z(k) - H(k)x(k) - a_k^{(i)}, R_k) \tag{19}$$

where

$$N(z(k) - H(k)x(k) - a_k^{(i)}, R_k) = \left(1/\sqrt{2\pi R_k}\right) EXP\left\{-\left(z(k) - H(k)x(k) - a_k^{(i)}\right)^2 / 2R_k\right\} \tag{20}$$

We do not require that $\Sigma\alpha_i = 1$. Thus, we have individual Gaussians centered at $a_k^{(i)}$ and each having standard deviation R_k. The sum in (19) may be infinite. The locations, $a_k^{(i)}$, and the amplitudes, α_i, are free parameters. Using (19) we obtain $p(z(k)|Z^{k-1})$ as

$$p(z(k)|Z^{k-1}) = \sum_i \alpha_i N\left(z(k) - H(k)\hat{x}(k|k-1) - a_k^{(i)}, H(k)P(k|k-1)H^T(k) + R_k\right) \quad (21)$$

In obtaining (21) we have again assumed that $p(x(k)|Z^{k-1})$ is Gaussian.

Substituting (21) into (8) and (9) with the variance of the residuals, $H(k)P(k\ k-1)H^T(k)+R_k$, replaced with an estimated value, s_k^2, gives

$$\hat{x}(k|k) = \hat{x}(k|k-1) + \frac{P(k|k-1)H^T(k)}{s_k^2}\left(z(k) - H(k)\hat{x}(k|k-1) - \bar{a}_k\right) \quad (22)$$

In (22) $\bar{a}_k$ is the weighted average

$$\bar{a}_k = \sum_i W_i a_k^{(i)}, \quad (23)$$

where the weights W_i are given by

$$W_i = \frac{\alpha_i N\left(z(k) - H(k)\hat{x}(k|k-1) - a_k^{(i)}, s_k^2\right)}{\sum_j \alpha_j N\left(z(k) - H(k)\hat{x}(k|k-1) - a_k^{(j)}, s_k^2\right)} \quad (24)$$

The filter equations (22), (23), and (24) are identical in form to the simplified pseudo-Bayes filter by Ackerson and Fu [7] for adaptive filtering when the mean of the measurement noise is unknown. Using (10) and (11) the conditional covariance is

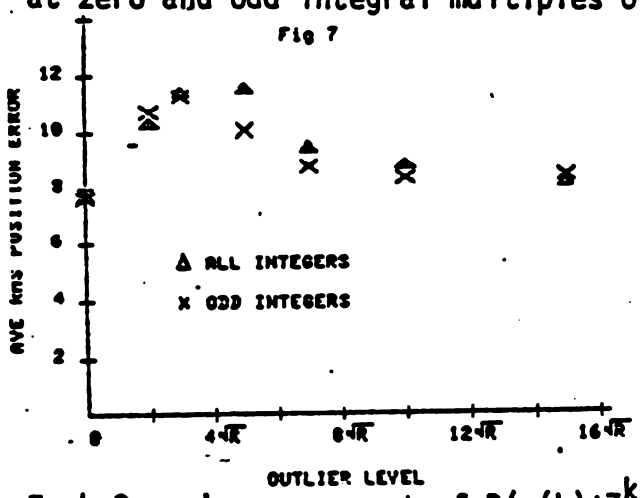
$$P(k|k) = P(k|k-1) - P(k|k-1)H(k)\left(1/s_k^2 - \frac{\overline{\left(a_k^{(i)} - \bar{a}_k\right)^2}}{s_k^4}\right)H^T(k)P(k|k-1), \quad (25)$$

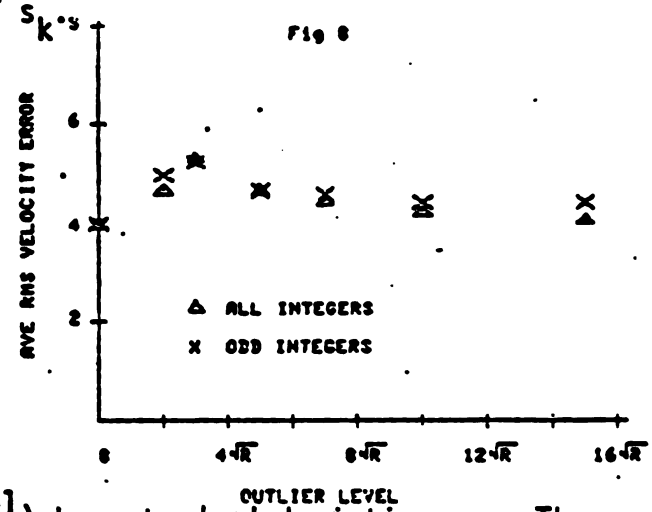
where

$$\overline{\left(a_k^{(i)} - \bar{a}_k\right)^2} = \sum_i W_i\left(a_k^{(i)} - \bar{a}_k\right)^2 \quad (26)$$

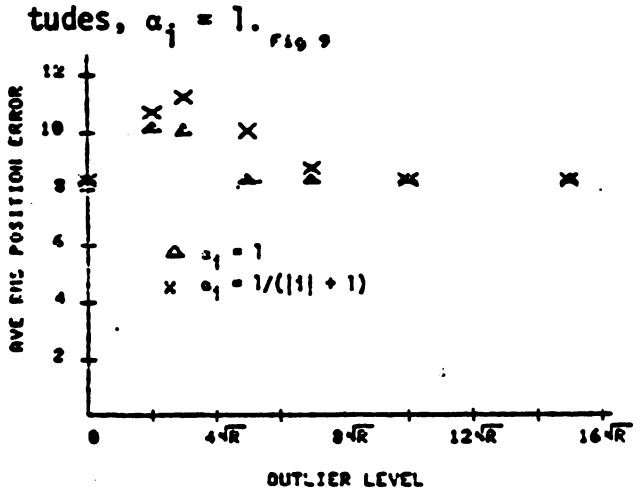
Although the above sums may be theoretically infinite, we only need to calculate the relatively few terms in the sums which have significant values of the weights, W_i. Thus, we compute only those terms in the sums for which $|z(k)-H(k)\hat{x}(k|k-1)-a_k^{(i)}|/s_k \leq 4$. With this simplification the amount of extra computation required to implement this robust filter is relatively small. The locations, $a_k^{(i)}$, produce a smooth pseudo-density if they are chosen as zero and odd integral multiples of s_k, $a_k^{(0)} = 0$, $a_k^{(i)} = \mathrm{sgn}(i)(2|i| - 1)s_k$, $|i| \geq 1$. We have also tested the filter with $a_k^{(i)} = i \cdot k$, $|i| \geq 0$. The value of the residual dispersion, s_k, is still computed by (15). Several different choices of the amplitudes have been tested. The most extensive testing has been done with $\alpha_i = 1$ and $\alpha_i = 1/(|i| + 1)$.

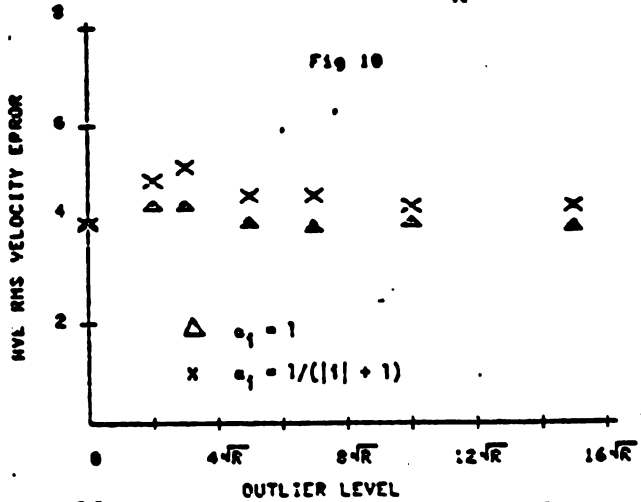
Some robust filters using the Gaussian mixture formulation were also evaluated via Monte Carlo testing. These filters were tested using the same simulated trajectory data and under the same outlier and measurement noise conditions as the robust filters using Hampel ψ functions. The sample size for Monte Carlo was twenty-five, the Markov transition probabilities were $P_{21} = .05$ and $P_{12} = .5$ and the measurement noise standard deviation, which was unknown to the filter and estimated by (15), was $\sqrt{R_k}$ = 20 feet. Figures 7 and 8 present the average rms position and velocity errors for a Gaussian mixture filter with observations contaminated by various magnitudes of outliers. The Gaussian mixture filters used in generating Figures 7 and 8 used magnitudes of the Gaussians, $\alpha_i = 1/(|i| + 1)$. Two different Gaussian mixture filters are represented in Figures 7 and 8, one with Gaussians at all integral multiples of s_k and one with Gaussians at zero and odd integral multiples of s_k.





Each Gaussian component of $P(z(k)|Z^{k-1})$ has standard deviation s_k. There is very little difference in the estimation errors obtained for the two filters of Figures 7 and 8. The filter with Gaussians at only the odd multiples of s_k is computationally less complex. Figures 9 and 10 give the results of the Monte Carlo evaluation of a Gaussian mixture filter which places Gaussians at zero and odd integral multiples of s_k with amplitudes, $\alpha_i = 1$.





This filter appears to give somewhat smaller errors than the other Gaussian mixture filters and also slightly smaller errors that the robust filter which uses Ha(1, 2, 3). We note that using $\alpha_i = 1$ with Gaussians at all integral multiples of s_k does not result in a useful filter since it has a zero influence function and therefore does not produce any error correction.

CONCLUSIONS

Two methods based on M-estimates have been presented for robust filtering and smoothing, one using the Hampel ψ function with various breakpoints and the other which models the observation error as a Gaussian mixture. These robust filtering methods were subjected to a Monte Carlo evaluation using simulated trajectory data from an aircraft tracking application. The results of this evaluation show that both of these robust filtering methods give a significant reduction in average rms estimation error for small outliers compared to a more conventional way of treating outliers in an optimal filtering application. The tests also suggest that the Gaussian mixture robust filter methods offer the most promise for application and future testing. Further evaluation of robust filtering methods under more severe trajectory applications are necessary.

REFERENCES

1. Huber, Peter J., (1973) Robust Regression: Asymptotics, Conjectures, and Monte Carlo, Ann. Statist., 1, 799-821.

2. Agee, W.S., and Turner, R. H., (1979), Application of Robust Regression to Trajectory Data Reduction, Robustness in Statistics, ed. by R. L. Launer and G. Wilkinson, Academic Press.

3. Agee, W. S., and Turner, R. H., (1978), Application of Robust Statistical Methods to Data Reduction, Analysis and Computation Div. Tech. Rpt. No. 65, White Sands Missile Range.

4. Masreliez, C. Johan and Martin, R. Douglas, (June 1977), Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter, IEEE Trans. Automatic Control, AC-22, 361-371.

5. Hampel, Frank R., (June 1974). The Influence Curve and its Role in Robust Estimation, J. Amer. Statist. Assoc., 69, 383-393.

6. Masreliez, C. Johan, (Feb. 1975), Approximate Non-Gaussian Filtering with Linear State and Observation Relations, IEEE Trans. Automatic Control, AC-20, 107-110.

7. Ackerson, G.A., and Fu, K.S., (Feb. 1970), On State Estimation in Switching Environments, IEEE Trans. Automatic Control, AC-15, 10-17.

COMPOUND FREQUENCY DISTRIBUTIONS

A METHOD FOR ESTIMATING STATISTICAL PARAMETERS FROM AN ADULTERATED SAMPLE

Donald W. Rankin

Army Materiel Test and Evaluation Directorate
US Army White Sands Missile Range
White Sands Missile Range, New Mexico 88002

ABSTRACT. When a sample is contaminated by extraneous "outliers", computation of the higher statistical moments may contain large errors. The proposed method treats these "outliers" as members of another "unwanted" population, and assumes that they perturb the distribution minimally near the maximum ordinate (mode).

The distribution is studied only near this maximum ordinate. A simple curve (a parabola, say) is fit by the method of least squares and the various derivatives are evaluated at this maximum ordinate. Not only the usual statistical parameters (mean, variance), but also the proportional number of "outliers" turn out to be expressible as simple functions of these derivatives.

I. THE PROBLEM

Statistical analysis usually requires that certain a priori assumptions be made; e.g., a certain population is normally distributed. From time to time, however, a test will reveal that a sample has been drawn which is incompatible with the basic assumptions.

An example which quickly comes to mind is the distribution of aerial bombing scores. For many samples, the assumption of a normally distributed population appears to be invalid -- frequency in the "tails" is far too high.

Before abandoning the postulation of normality, let us address the problem from a different point of view. Keeping the example of aerial bombing scores, suppose that in a sample of, say, 100 bombing runs, the bombardier misidentified the intended aiming point on ten occasions. It is obvious that only 90 scores will be drawn from the "correct" population (i.e., the population for which we have postulated normality), while the remaining ten will come from populations with displaced means. To choose terminology, we shall say that the sample is adulterated by the ten runs from unwanted sources.

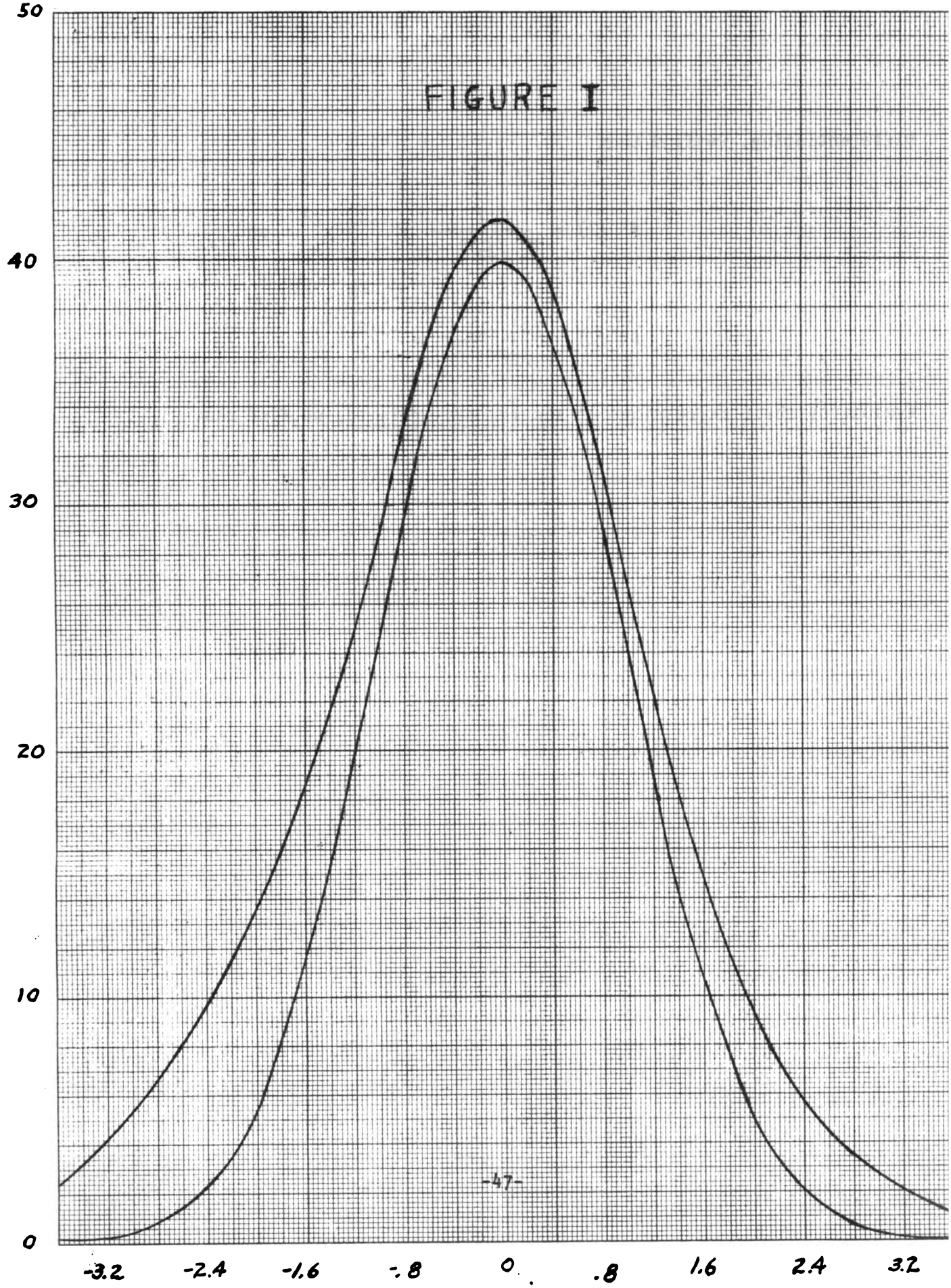
How the density function is affected by mixing different populations is seen in Figure I. The lower curve is simply the normal curve in which $\sigma = 1$ and $N = 100$. Adding adulteration from two extraneous populations ($\sigma = 1, \mu = -2, N = 20$, and $\sigma = 1, \mu = 2, N = 10$) yields the density function for the compound frequency distribution, illustrated by the upper curve. It is apparent at once that the "tails" are abnormally thick. It should also be noted that the density function is deformed least near the mean.

If the mean of the extraneous bits of data is displaced by much more than 2σ, the effect on the center of the sample distribution is virtually nil. Can the parameters of the desired population be recovered by studying the sample distribution only near its center?

II. A SOLUTION

The probability density function of a standardized ($\mu = 0, \sigma = 1$) normal distribution is given by

$$\phi(Z) = \frac{1}{\sqrt{2\pi}} e^{-1/2 \; Z^2}$$



Differentiating, we find that

$$\phi'(Z) = \frac{-Z}{\sqrt{2\pi}} e^{-1/2\, Z^2} \quad \text{and}$$

$$\phi''(Z) = \frac{(Z^2 - 1)}{\sqrt{2\pi}} e^{-1/2\, Z^2}$$

Evaluating these expressions at the maximum ordinate, we find that

$$Z = 0$$

$$\phi(0) = \frac{1}{\sqrt{2\pi}}$$

$$\phi'(0) = 0$$

$$\phi''(0) = \frac{-1}{\sqrt{2\pi}}$$

Expressing $\phi(Z)$ in series form,

$$\phi(Z) = \frac{1}{\sqrt{2\pi}} \left[1 - \frac{Z^2}{2} + \frac{Z^4}{8} - \frac{Z^6}{48} + \frac{Z^8}{384} - \cdots\right]$$

It is easy to see that for small values of Z -- say $|Z|<0.4$ -- the first two terms form a sufficiently good approximation. Note that for

$$\phi(Z) = \frac{1}{\sqrt{2\pi}} \left[1 - \frac{Z^2}{2}\right] \quad \text{and } Z = 0,$$

$$\phi'(Z) = -\frac{Z}{\sqrt{2\pi}}$$

$$\phi''(Z) = -\frac{1}{\sqrt{2\pi}}$$

$$\phi(0) = \frac{1}{\sqrt{2\pi}}$$

$$\phi'(0) = 0$$

$$\phi''(0) = -\frac{1}{\sqrt{2\pi}}$$

In other words, at the point of the maximum ordinate (Z = 0), the approximating parabola and its pertinent derivatives take on exactly the same values as the probability density function.

Since $\phi(Z)$ is a probability function,

$$\int_{-\infty}^{\infty} \phi(Z)\, dZ = 1.$$

In an actual case, a sample of size N will be drawn, consisting of N values of the form y_i.

Each y_i can be thought of as a deviation about a certain origin; i.e., as an abscissa. Since each y_i occurs with frequency 1, we have immediately

$$f(y_i) = 1 \text{ and } \sum_{i=1}^{N} f(y_i) = N.$$

Let us make two assumptions -- first that the desired population (call it the "Z population") is normally distributed -- second that the extraneous members of the sample are clustered about points far enough removed from the mean of the Z population that the frequency distribution is minimally perturbed near the center.

Let N_z denote the number of members of the sample which are drawn from the desired population. Let μ and σ be, respectively, the mean and standard deviation of the y_i's of this Z population. Then

$Z = \frac{y - \mu}{\sigma}$ is the standardized variable,

allowing our previously developed expressions in Z to hold. Note that

$$\int_{-\infty}^{\infty} \phi(Z)\, dZ = \int_{-\infty}^{\infty} \frac{1}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right) dy = 1.$$

If we could identify the members of the Z population, we could determine

$$N_z = \sum_{z} f(y_i)$$

merely by counting. This cannot be done. But we note that

$$\int_{-\infty}^{\infty} \frac{N_z}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right) dy = N_z$$

choosing to regard y (without a subscript) as a continuous variable, and thus considering f(y) as an ordinate rather than as a frequency, we find,

$$f(y) = \frac{N_z}{\sigma} \phi(Z) = \frac{N_z}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

$$f'(y) = \frac{-N_z(y-\mu)}{\sigma^3\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

$$f''(y) = \frac{-N_z}{\sigma^3\sqrt{2\pi}} \left[1 - \left(\frac{y-\mu}{\sigma}\right)^2\right] e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

Evaluating these at the maximum ordinate ($y = \mu$), we find that

$$f(\bar{y}) = \frac{N_z}{\sigma\sqrt{2\pi}}$$

$$f'(\bar{y}) = 0$$

$$f''(\bar{y}) = \frac{-N_z}{\sigma^3\sqrt{2\pi}}$$

Solving for N_z and σ yields two fundamental expressions.

$$\sigma^2 = \frac{-f(\bar{y})}{f''(\bar{y})}$$

$$N_z = \sqrt{-\frac{2\pi f^3(\bar{y})}{f''(\bar{y})}}$$

For actual computation, the form

$$N_z = \sigma f(\bar{y})\sqrt{2\pi}$$

may be preferable.

It remains to evaluate the mean ($\bar{y}$ is the best estimate of μ) and the values of the function and its second derivative at that point. For this we use the approximating parabola (least squares fit *).

$$f(y) = K_o + K_1 y + K_2 y^2$$

$$f'(y) = K_1 + 2K_2 y$$

$$f''(y) = 2K_2$$

Since $f'(\bar{y}) = 0$,

$$\bar{y} = \frac{-K_1}{2K_2}$$

and $$f(\bar{y}) = K_o - \frac{K_1^2}{4K_2} .$$

Of course $f''(\bar{y}) = 2K_2$.

The reader is reminded that the form of the distribution is selected from criteria <u>other</u> than the appearance of the raw data. For example, a distribution which follows (Appendix A, Table 1, Figure II) <u>appears</u> to be tri-modal, due to the presence of many extraneous outliers.

*See Appendix A

APPENDIX A

A SIMPLIFIED METHOD FOR FITTING A PARABOLA (LEAST SQUARES)

The application of the method of least squares is greatly facilitated by transforming the independent variable so that its transform has a mean of 0 and an increment of 1. By way of illustration, in Table 1, y is the independent variable, f its frequency and x its transform.

$$x = \frac{y - 24}{3} \quad , \qquad \sigma_y = 3\,\sigma_x$$

In terms of the transformed variable x, since $\sum_{-n}^{n} x_i^{2m-1} = 0$

(m is any positive integer),

$$g(x) = A_o + A_1 x + A_2 x^2$$

$$A_o = \frac{\begin{vmatrix} \Sigma f & \Sigma x^2 \\ \Sigma x^2 f & \Sigma x^4 \end{vmatrix}}{\begin{vmatrix} \Sigma 1 & \Sigma x^2 \\ \Sigma x^2 & \Sigma x^4 \end{vmatrix}}$$

$$A_1 = \frac{\Sigma xf}{\Sigma x^2}$$

$$A_2 = \frac{\begin{vmatrix} \Sigma 1 & \Sigma f \\ \Sigma x^2 & \Sigma x^2 f \end{vmatrix}}{\begin{vmatrix} \Sigma 1 & \Sigma x^2 \\ \Sigma x^2 & \Sigma x^4 \end{vmatrix}}$$

Table 2 contains numbers useful in computation.

TABLE 1

A HYPOTHETICAL SAMPLE DISTRIBUTION

v	f(y)	x	xf	x^2f
0	0	-8	0	0
3	2	-7	-14	98
6	4	-6	-24	144
9	3	-5	-15	75
12	0	-4	0	0
15	3	-3	-9	27
18	2	-2	-4	8
21	4	-1	-4	4
24	2	0	0	0
27	5	1	5	5
30	2	2	4	8
33	2	3	6	18
36	0	4	0	0
39	2	5	10	50
42	1	6	6	36
45	2	7	14	98
48	1	8	8	64
$\sum$	35		-17	635

TABLE 2

SUMMED SQUARES AND FOURTH POWERS WITH OTHER USEFUL NUMBERS

i	$\sum_{-i}^{i} 1$	$\sum_{-i}^{i} x_i^2$	$\sum_{-i}^{i} x_i^4$	*D
2	5	10	34	70
3	7	28	196	588
4	9	60	708	2772
5	11	110	1958	9438
6	13	182	4550	26026
7	15	280	9352	61880
8	17	408	17544	131784
9	19	570	30666	257754
10	21	770	50666	471086
11	23	1012	79948	814660
12	25	1300	121420	1345500

$$*D = \begin{vmatrix} \sum 1 & \sum x^2 \\ \sum x^2 & \sum x^4 \end{vmatrix}$$

π = 3.141 592 654

2π = 6.283 185 307

$\sqrt{2\pi}$ = 2.506 628 275

Suppose it is desired to fit a parabola to the 9 central points (Table 1).

$i = 4, \Sigma f = 20, \Sigma xf = -2, \Sigma x^2 f = 70$.

$$A_o = \frac{\begin{vmatrix} 20 & 60 \\ 70 & 708 \end{vmatrix}}{2772} = \frac{9960}{2772} = 3.593\ 074$$

$$A_1 = -2/60 = -0.033\ 333$$

$$A_2 = \frac{\begin{vmatrix} 9 & 20 \\ 60 & 70 \end{vmatrix}}{2772} = \frac{-570}{2772} = -0.205\ 628$$

$$\bar{x} = \frac{-A_1}{2A_2} = -0.081\ 053$$

$$\bar{y} = 23.756\ 84$$

$$g(\bar{x}) = A_o - \frac{A_1^2}{4A_2} = 3.594\ 424$$

$$g''(\bar{x}) = -0.411\ 255$$

$$\sigma_x = \sqrt{\frac{g(\bar{x})}{g''(\bar{x})}} = 2.95637$$

$$\sigma_y = 8.86911$$

$$N_z = \sigma_x\, g(\bar{x})\, \sqrt{2\pi} = 26.64 \text{ or } 27$$

APPENDIX B

THE ERROR INDUCED BY ASSUMING A PARABOLA TO APPROXIMATE A NORMAL CURVE

With no prior knowledge of the value of σ, it usually happens that rather large values of $\frac{y-\mu}{\sigma}$ are used to fit a parabola (1.38 in illustration in Appendix A). When this happens, the computed values of N_z and σ will be too large.

Table 3 shows the error induced by the approximation $1 - \frac{z^2}{2} = e^{-z^2/2}$. The tabular values give the error as a proportion of $g(\bar{x})$, the maximum ordinate ($\sigma = 1$).

If the mean and variance can be estimated (and the first parabola fit will provide a rough estimate), the error at each value of the independent variable can be computed. If these error terms are subtracted from the corresponding frequencies, a parabola fit to these "corrected frequencies" will exactly reproduce the desired parameters with no residual error (except that induced by errors in the estimation of μ and σ).

Although the correction function is exact, it should be remembered that on the normal curve there are inflection points at $\pm\sigma$, outside of which the parabola and normal curve diverge very rapidly. This makes corrections computed for points outside σ dependent upon accurate estimates of μ and σ. For example, a 5% error in the estimate of σ will result in a correction error at 0.5 σ of only 0.3% of the maximum ordinate, but 3.2% at 1.2 σ -- a tenfold increase.

The histogram of the data from Table 1, with the best-fitting normal curve and associated parabola is illustrated in Figure II.

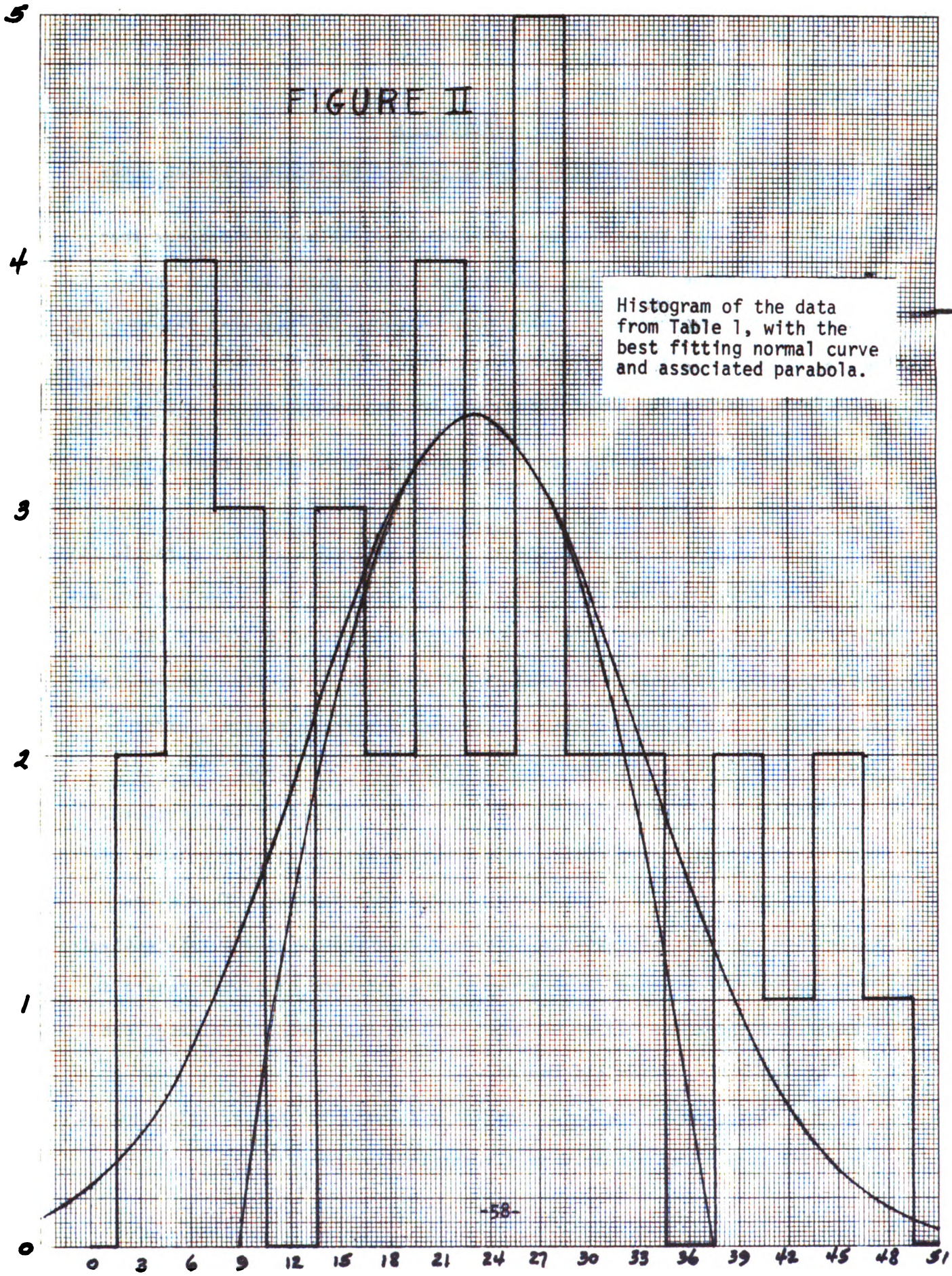


TABLE 3

$$e^{-\frac{1}{2}Z^2} - \left(1 - \frac{1}{2}Z^2\right)$$

Z	$e^{-\frac{1}{2}Z^2}$	$1 - \frac{Z^2}{2}$	Diff	Z	$e^{-\frac{1}{2}Z^2}$	$1 - \frac{Z^2}{2}$	Diff
0.00	1.0000	1.000	.00000	0.70	.78270	.7550	.02770
0.02	.99980	.9998	.00000	0.72	.77167	.7408	.03087
0.04	.99920	.9992	.00000	0.74	.76048	.7262	.03428
0.06	.99820	.9982	.00000	0.76	.74916	.7112	.03796
0.08	.99681	.9968	.00001	0.78	.73771	.6958	.04191
0.10	.99501	.9950	.00001	0.80	.72615	.6800	.04615
0.12	.99283	.9928	.00003	0.82	.71448	.6638	.05068
0.14	.99025	.9902	.00005	0.84	.70272	.6472	.05552
0.16	.98728	.9872	.00008	0.86	.69087	.6302	.06067
0.18	.98393	.9838	.00013	0.88	.67896	.6128	.06616
0.20	.98020	.9800	.00020	0.90	.66698	.5950	.07198
0.22	.97609	.9758	.00029	0.92	.65495	.5768	.07815
0.24	.97161	.9712	.00041	0.94	.64288	.5582	.08468
0.26	.96676	.9662	.00056	0.96	.63078	.5392	.09158
0.28	.96156	.9608	.00076	0.98	.61866	.5198	.09886
0.30	.95600	.9550	.00100	1.00	.60653	.5000	.10653
0.32	.95009	.9488	.00129	1.02	.59440	.4798	.11460
0.34	.94384	.9422	.00164	1.04	.58228	.4592	.12308
0.36	.93725	.9352	.00205	1.06	.57018	.4382	.13198
0.38	.93034	.9278	.00254	1.08	.55811	.4168	.14131
0.40	.92312	.9200	.00312	1.10	.54608	.3950	.15108
0.42	.91558	.9118	.00378	1.12	.53409	.3728	.16129
0.44	.90774	.9032	.00454	1.14	.52215	.3502	.17195
0.46	.89960	.8942	.00540	1.16	.51028	.3272	.18308
0.48	.89119	.8848	.00639	1.18	.49848	.3038	.19468
0.50	.88250	.8750	.00750	1.20	.48675	.2800	.20675
0.52	.87354	.8648	.00874	1.22	.47511	.2558	.21931
0.54	.86433	.8542	.01013	1.24	.46357	.2312	.23237
0.56	.85488	.8432	.01168	1.26	.45212	.2062	.24592
0.58	.84518	.8318	.01338	1.28	.44078	.1808	.25998
0.60	.83527	.8200	.01527	1.30	.42956	.1550	.27456
0.62	.82514	.8078	.01734	1.32	.41845	.1288	.28965
0.64	.81481	.7952	.01961	1.34	.40747	.1022	.30527
0.66	.80429	.7822	.02209	1.36	.39661	.0752	.32141
0.68	.79358	.7688	.02478	1.38	.38589	.0478	.33809
0.70	.78270	.7550	.02770	1.40	.37531	.0200	.35531

Interpolation in Table 3 can be avoided by choosing values of σ and μ such that

$$\frac{1}{\sigma_x} = 0.02\ R$$

and

$$\frac{\mu}{\sigma} = 0.02\ R'$$

(R and R' are integers).

This is a reasonable procedure; since the first estimate of σ_x probably is too high, it is legitimate to choose that lower value which serves the purpose. Table 4 is presented for convenience in choosing σ_x.

Application of the correction technique to the example of Appendix A is shown in Table 5 and following.

TABLE 4

VALUES OF σ_x WHICH YIELD CLASS INTERVALS OF .02R

$\frac{1}{\sigma}$	σ	$\frac{1}{\sigma}$	σ
.02	50	.32	3.125
.04	25	.34	2.94118
.06	16.66667	.36	2.77778
.08	12.5	.38	2.63158
.10	10	.40	2.5
.12	8.33333	.42	2.38095
.14	7.14286	.44	2.27273
.16	6.25	.46	2.17391
.18	5.55556	.48	2.08333
.20	5	.50	2
.22	4.54545	.52	1.92308
.24	4.16667	.54	1.85185
.26	3.84615	.56	1.78571
.28	3.57143	.58	1.72414
.30	3.33333	.60	1.66667

TABLE 5

CORRECTION OF A HYPOTHETICAL FREQUENCY DISTRIBUTION

y	f(y)	x	$\frac{x - \bar{x}}{\sigma_x}$	Factor	Δf	f+Δf	x(f+Δf)	$x^2(f+\Delta f)$
12	0	-4	-1.42					
15	3	-3	-1.06	.13198	-0.47	2.53	-7.59	22.77
18	2	-2	-.70	.02770	-0.10	1.90	-3.80	7.60
21	4	-1	-.34	.00164	-0.01	3.99	-3.99	3.99
24	2	0	.02	.00000	-0.00	2.		
27	5	1	.38	.00254	-0.01	4.99	4.99	4.99
30	2	2	.74	.03428	-0.12	1.88	3.76	7.52
33	2	3	1.10	.15108	-0.54	1.46	4.38	13.14
36	0	4	1.46					
$\sum$	20	0			-1.25	18.75	-2.25	60.01

From Appendix A, we have

$\bar{x}$ = -0.081 053

$g(\bar{x})$ = 3.594424

σ_x = 2.95637

$1/\sigma_x$ = 0.338

for correction, set $1/\sigma_x$ = 0.36

Then 0.36 $\bar{x}$ = -0.029, set 0.36 $\bar{x}$ = -0.02

It is observed that for $x = \pm 4$, the entering argument lies outside the range of Table 3. This merely means that the original curve fit spanned too much data. The values for $x = \pm 4$ are simply dropped from further calculations.

Fitting a parabola to the "corrected" frequencies, we find

$$B_o = \frac{\begin{vmatrix} 18.75 & 28 \\ 60.01 & 196 \end{vmatrix}}{588} = \frac{1994.72}{588} = 3.39238$$

$$B_1 = \frac{-2.25}{28} = -0.08036$$

$$B_2 = \frac{\begin{vmatrix} 7 & 18.75 \\ 28 & 60.01 \end{vmatrix}}{588} = \frac{-104.93}{588} = -0.17845$$

$\bar{x} = -0.22515$

$\bar{y} = 23.32455$

$g(\bar{x}) = 3.40143$

$g''(\bar{x}) = -0.35690$

$\sigma_x = 3.08713$

$\sigma_y = 9.26138$

$N = 26.32$ or 26

$\frac{1}{\sigma_x} = 0.324$, let $\frac{1}{\sigma_x} = 0.32$

$0.32\ \bar{x} = -0.072$, let $0.32\ \bar{x} = -0.08$

The consequence here of dropping the data for $x = \pm 4$ is that rarity, too small an estimate of σ. Another iteration seems in order. See Tables 6, 7 and following

TABLE 6

FURTHER CORRECTION OF A HYPOTHETICAL FREQUENCY DISTRIBUTION

y	f	$\frac{x - \bar{x}}{\sigma_x}$	Factor	Δg	x	$F = f + \Delta g$	x F	x^2 F
15	3	-.88	.06616	-0.23	-3	2.77	-8.31	24.93
18	2	-.56	.01168	-0.04	-2	1.96	-3.92	7.84
21	4	-.24	.00041	0.	-1	4.	-4.	4.
24	2	0.08	.00001	0.	0	2.	0	0
27	5	0.40	.00312	-0.01	1	4.99	4.99	4.99
30	2	0.72	.03087	-0.11	2	1.89	3.78	7.56
33	2	1.04	.12308	-0.42	3	1.58	4.74	14.22
Σ	20			-0.81		19.19	-2.72	63.54

$$C_o = \frac{\begin{vmatrix} 19.19 & 28 \\ 63.54 & 196 \end{vmatrix}}{588} = \frac{1982.12}{588} = 3.37095$$

$$C_1 = \frac{-2.72}{28} = -0.09714$$

$$C_2 = \frac{\begin{vmatrix} 7 & 19.19 \\ 28 & 63.54 \end{vmatrix}}{588} = \frac{-92.54}{588} = -0.15738$$

$\bar{x}$ = -0.30862

$\bar{y}$ = 23.07413

$g(\bar{x})$ = 3.38594

$g''(\bar{x})$ = -0.31476

σ_x = 3.27981

σ_y = 9.83943

N = 27.84 or 28

$\frac{1}{\sigma_x}$ = 0.305 let $\frac{1}{\sigma_x}$ = 0.30

$0.30\ \bar{x} = -0.093$, let $0.3\bar{x} = -0.10$

TABLE 7

FINAL CORRECTION OF A HYPOTHETICAL FREQUENCY DISTRIBUTION

y	f	$\frac{x - \bar{x}}{\sigma_x}$	Factor	Δg	ϕ	x	$x\phi$	$x^2\phi$
15	3	-.80	.04615	-0.16	2.84	-3	-8.52	25.56
18	2	-.50	.00750	-0.03	1.97	-2	-3.94	7.88
21	4	-.20	.00020	0	4.	-1	-4.	4
24	2	.10	.00001	0	2.	0	0	0
27	5	.40	.00312	-0.01	4.99	1	4.99	4.99
30	2	.70	.02770	-0.09	1.91	2	3.82	7.64
33	2	1.00	.10653	-0.36	1.64	3	4.92	14.76
$\sum$	20			-0.65	19.35		-2.73	64.83

$$D_o = \frac{\begin{vmatrix} 19.35 & 28 \\ 64.83 & 196 \end{vmatrix}}{588} = 3.36286$$

$$D_1 = \frac{-2.73}{28} = -0.09750$$

$$D_2 = \frac{\begin{vmatrix} 7 & 19.35 \\ 28 & 64.83 \end{vmatrix}}{588} = -0.14964$$

$\bar{x} = -0.32578$

$\bar{y} = 23.02267$

$g(\bar{x}) = 3.37874$

$g''(\bar{x}) = -0.29929$

$\sigma_x = 3.35996$

$\sigma_y = 10.07988$

$N = 28.46$ or 28

Check:

$$\frac{1}{\sigma_x} = 0.298$$

$0.298\ \bar{x} = -0.097$

APPENDIX C

EXAMPLES

To test the method,three samples of size 150 were drawn from Rand's Table* of Gaussian deviates and were adulterated as described.

Problem 1

Sample starts with line 2206, Page 45

First 100 numbers unbiased

Next 35 numbers biased +2.5

Next 15 numbers biased -2.0

Class interval was chosen as 0.3, yielding the sample distribution displayed in Table 8. For the population, $\mu = 0$, $\sigma = 1$. The "pure" portion of the sample (N = 100) estimates these parameters as $\mu = 0.015$, $\sigma = 0.943$. The total sample provides the useless estimates $\mu = 0.332$, $\sigma = 1.579$

The curve-fitting technique provides a first approximation of $\mu = 0.059$, $\sigma = 1.025$, $N = 109$. After applying the corrections, the method yields $\mu = 0.062$, $\sigma = 0.920$, $N = 99$. Computations are shown in Tables 9, 10 and following.

*The RAND Corporation, A Million Random Digits with 100,000 Normal Deviates, Free Press, 1955.

TABLE 8

FREQUENCY DISTRIBUTION, PROBLEM 1

y Mid-Point	x	Frequency Unbiased Portion	Total Sample	y Mid-Point	x	Frequency Unbiased Portion	Total Sample
-4.5	-15		1	0.3	1	14	14
-4.2	-14		0	0.6	2	10	11
-3.9	-13		0	0.9	3	8	8
-3.6	-12		1	1.2	4	8	14
-3.3	-11		0	1.5	5	3	7
-3.0	-10		2	1.8	6	0	4
-2.7	-9		0	2.1	7	2	5
-2.4	-8		2	2.4	8	1	4
-2.1	-7	2	3	2.7	9		5
-1.8	-6	3	4	3.0	10		2
-1.5	-5	3	7	3.3	11		2
-1.2	-4	6	8	3.6	12		2
-0.9	-3	7	8	3.9	13		2
-0.6	-2	9	9	4.2	14		0
-0.3	-1	12	12	4.5	15		0
0.0	0	12	12	4.8	16		1

TABLE 9

CURVE FIT, PROBLEM 1

y	f	x	xf	x^2f
-.9	8	-3	-24	72
-.6	9	-2	-18	36
-.3	12	-1	-12	12
0	12	0	0	0
.3	14	1	14	14
.6	11	2	22	44
.9	8	3	24	72
Σ	74		6	250

$$A_o = \frac{\begin{vmatrix} 74 & 28 \\ 250 & 196 \end{vmatrix}}{588} = 12.76190$$

$A_1 = 6/28 = 0.21429$

$$A_2 = \frac{\begin{vmatrix} 7 & 74 \\ 28 & 250 \end{vmatrix}}{588} = -0.54762$$

$\bar{x} = -A_1/2A_2 = 0.19565$

$f(\bar{x}) = A_o - A_1^2/4A_2 = 12.78287$

$\sigma_x = 3.41633$

To compute corrections

$1/\sigma_x = .293$, Use 0.32

$0.32\ \bar{x} = .063$, Use 0.06

$\bar{y} = 0.05870$

$\sigma_{\bar{y}} = 1.02490$

$N_z = 109$

TABLE 10

CORRECTED CURVE FIT, PROBLEM 1

y	f	$\frac{x - \bar{x}}{\sigma_x}$	Factor	Δf	g	x	xg	x^2g
-.9	8	-1.02	.11460	-1.46	6.54	-3	-19.62	58.86
-.6	9	- .70	.02770	-0.35	8.65	-2	-17.30	34.60
-.3	12	- .38	.00254	-0.03	11.97	-1	-11.97	11.97
0.	12	- .06	.00000	0.	12.	0	0	0
.3	14	.26	.00056	-0.01	13.99	1	13.99	13.99
.6	11	.58	.01338	-0.17	10.83	2	21.66	43.32
.9	8	.90	.07198	-0.92	7.08	3	21.24	63.72
$\sum$	74			-2.94	71.06		8.00	226.46

$$B_o = \frac{\begin{vmatrix} 71.06 & 28 \\ 226.46 & 196 \end{vmatrix}}{588} = 12.90286$$

$B_1 = 8.00/28 = 0.28571$

$$B_2 = \frac{\begin{vmatrix} 7 & 71.06 \\ 28 & 226.46 \end{vmatrix}}{588} = -0.68786$$

$\bar{x} = 0.20768$

$\bar{y} = 0.06231$

$g(\bar{x}) = 12.93253$

$\sigma_x = 3.06604$

$\sigma_y = 0.91981$

$N = 99.4$ or 99

Check:

$\frac{1}{\sigma_x} = 0.326$

$0.326\ \bar{x} = 0.068$

PROBLEM 2

Sample starts with line 5622, Page 113. All other conditions identical to Problem 1. Results are shown in Table 11, 12 and 13 and following. The "pure" portion of the sample yields $\mu = 0$, $\sigma = 0.913$. The total sample gives $\mu = 0.426$, $\sigma = 1.728$.

The curve-fitting technique provides a first approximation of $\mu = -0.101$, $\sigma = 1.011$, $N = 112$. After applying the corrections, the method yields $\mu = -0.106$, $\sigma = 0.908$, $N = 102$.

PROBLEM 3

Sample starts with line 8371, Page 168. All other conditions identical to Problems 1 and 2. The "pure" portion of the sample yields $\mu = 0.102$, $\sigma = 0.961$. The total sample gives $\mu = 0.386$, $\sigma = 1.508$.

The curve-fitting technique provides a first approximation of $\mu = 0.123$, $\sigma = 1.230$, $N = 136$. After applying the corrections, the method yields $\mu = 0.132$, $\sigma = 1.109$, $N = 123$. Results are shown in Tables 14, 15 and 16. It should be observed that this particular sample contains 5 "bad" data bits in the interval $|x| \leq 0.6$, more than 8% of the sample. The method cannot identify these points, with the result that the computed values of N and σ tend to be too large, although they are still better estimates of the true parameters than those obtainable from the entire sample.

TABLE 11

FREQUENCY DISTRIBUTION, PROBLEM 2

y Mid-Point	x	Frequency Unbiased Portion	Frequency Total Sample	y Mid-Point	x	Frequency Unbiased Portion	Frequency Total Sample
-3.9	-13		2	0.9	3	6	8
-3.6	-12		1	1.2	4	11	12
-3.3	-11		0	1.5	5	2	3
-3.0	-10		1	1.8	6	3	7
-2.7	-9		0	2.1	7	1	3
-2.4	-8	1	4	2.4	8		4
-2.1	-7	0	2	2.7	9		1
-1.8	-6	1	1	3.0	10		1
-1.5	-5	3	3	3.3	11		7
-1.2	-4	9	12	3.6	12		4
-0.9	-3	6	8	3.9	13		4
-0.6	-2	11	11	4.2	14		0
-0.3	-1	18	18	4.5	15		1
0.	0	11	12	4.8	16		0
0.3	1	8	9	5.1	17		0
0.6	2	9	10	5.4	18		1

TABLE 12

CURVE FIT, PROBLEM 2

y Mid Point	x	f	xf	x^2f
-0.9	-3	8	-24	72
-0.6	-2	11	-22	44
-0.3	-1	18	-18	18
0.	0	12	0	0
0.3	1	9	9	9
0.6	2	10	20	40
0.9	3	8	24	72
Σ		76	-11	255

$$A_o = \frac{\begin{vmatrix} 76 & 28 \\ 255 & 196 \end{vmatrix}}{588} = 13.19048$$

$A_1 = -11/28 = -0.39286$

$$A_2 = \frac{\begin{vmatrix} 7 & 76 \\ 28 & 255 \end{vmatrix}}{588} = -0.58333$$

$\bar{x} = -0.33673$

$f(\bar{x}) = 13.25662$

$\sigma_x = 3.37088$

To compute corrections,

$\frac{1}{\sigma_x} = 0.297$, Use 0.32

$0.32\bar{x} = -0.108$, Use -0.10

$\bar{y} = -0.10102$

$\sigma_y = 1.01126$

$N = 112$

TABLE 13

CORRECTED CURVE FIT, PROBLEM 2

y	f	$\frac{x - \bar{x}}{\sigma_x}$	Factor	Δf	g	x	xg	x^2g
-.9	8	-.86	.06067	-0.80	7.20	-3	-21.60	64.80
-.6	11	-.54	.01013	-0.13	10.87	-2	-21.74	43.48
-.3	18	-.22	.00029	0.	18.	-1	-18.	18.
0.	12	.10	.00001	0.	12.	0	0	0
.3	9	.42	.00378	-0.05	8.95	1	8.95	8.95
.6	10	.74	.03428	-0.45	9.55	2	19.10	38.20
.9	8	1.06	.13198	-1.75	6.25	3	18.75	56.25
$\sum$	76			-3.18	72.82		-14.54	229.68

$$B_o = \frac{\begin{vmatrix} 72.82 & 28 \\ 229.68 & 196 \end{vmatrix}}{588} = 13.33619$$

$B_1 = -14.54/28 = -0.51929$

$$B_2 = \frac{\begin{vmatrix} 7 & 72.82 \\ 28 & 229.68 \end{vmatrix}}{588} = -0.73333$$

$\bar{x} = -0.35406$

$\bar{y} = -0.10622$

$g(\bar{x}) = 13.42812$

$\sigma_x = 3.02581$

$\sigma_y = 0.90774$

$N = 101.8$ or 102

Check:

$\frac{1}{\sigma_x} = 0.330$

$.330\ \bar{x} = -0.117$

TABLE 14

FREQUENCY DISTRIBUTION, PROBLEM 3

y Mid Point	x	Frequency Unbiased Portion	Frequency Total Sample	y Mid Point	x	Frequency Unbiased Portion	Frequency Total Sample
-3.0	-10		2	1.2	4	6	9
-2.7	- 9		1	1.5	5	3	4
-2.4	- 8		1	1.8	6	2	4
-2.1	- 7	1	4	2.1	7	1	6
-1.8	- 6	3	3	2.4	8	0	6
-1.5	- 5	4	5	2.7	9	2	6
-1.2	- 4	3	6	3.0	10		4
-0.9	- 3	8	9	3.3	11		5
-0.6	- 2	9	11	3.6	12		0
-0.3	- 1	10	11	3.9	13		1
0.0	0	13	14	4.2	14		0
0.3	1	13	13	4.5	15		0
0.6	2	12	13	4.8	16		2
0.9	3	10	10	5.1	17		0

TABLE 15

CURVE FIT, PROBLEM 3

y	f	x	xf	x^2f	$\frac{x-\mu}{\sigma_x}$	Factor	Δf
-.9	9	-3	-27	81	-.96	.09158	-1.21
-.6	11	-2	-22	44	-.68	.02478	-0.33
-.3	11	-1	-11	11	-.40	.00312	-0.04
0.	14	0	0	0	-.12	.00003	0.
.3	13	1	13	13	.16	.00008	0.
.6	13	2	26	52	.44	.00454	-0.06
.9	10	3	30	90	.72	.03087	-0.41
$\sum$	81		9	291			-2.05

$$A_o = \frac{\begin{vmatrix} 81 & 28 \\ 291 & 196 \end{vmatrix}}{588} = 13.14286$$

$$A_1 = 9/28 = 0.32143$$

$$A_2 = \frac{\begin{vmatrix} 7 & 81 \\ 28 & 291 \end{vmatrix}}{588} = -0.39286$$

$\bar{x} = 0.40909$

$f(\bar{x}) = 13.20860$

$\sigma_x = 4.10012$

$\bar{y} = 0.12273$

$\sigma_y = 1.23003$

$N = 136$

To compute Δf, use

$\frac{1}{\sigma_x} = 0.28$

$\mu/\sigma = 0.12$

TABLE 16

CORRECTED CURVE FIT, PROBLEM 3

y	f	$\phi = f + \Delta f$	x	$x\phi$	$x^2\phi$
-.9	9	7.79	-3	-23.37	70.11
-.6	11	10.67	-2	-21.34	42.68
-.3	11	10.96	-1	-10.96	10.96
0.	14	14.	0	0.	0.
.3	13	13.	1	13.	13.
.6	13	12.94	2	25.88	51.76
.9	10	9.59	3	28.77	86.31
Σ		78.95		11.98	274.82

$$B_o = \frac{\begin{vmatrix} 78.95 & 28 \\ 274.82 & 196 \end{vmatrix}}{588} = 13.23000$$

$B_1 = 11.98/28 = 0.42786$

$$B_2 = \frac{\begin{vmatrix} 7 & 78.95 \\ 28 & 274.82 \end{vmatrix}}{588} = -0.48786$$

$\bar{x} = 0.43851$

$f(\bar{x}) = 13.32381$

$\sigma_x = 3.69533$

$\bar{y} = 0.13155$

$\sigma_y = 1.10860$

$N = 123$

Check:

$\frac{1}{\sigma_x} = 0.271$

$\mu/\sigma = 0.119$

Is the result affected by varying the span of data used? Let us rework Problem 3, using all the data up to $|y| = 1.8$. Even though the estimates of σ and μ are poor, iteration of the method quickly puts data for $|y| > 1.2$ outside the range of the correction table, and so drops them from the calculations. The remaining 9 points, when properly corrected, should virtually duplicate the results earlier obtained from fitting a curve to 7 points. The results compare as follows:

9-point fit: $\mu = 0.162$, $\sigma = 1.096$, $N = 123$

7-point fit: $\mu = 0.132$, $\sigma = 1.109$ $N = 123$

TABLE 17

13-POINT CURVE FIT, PROBLEM 3

y	f	x	xf	x^2f	$\frac{x-\mu}{\sigma_x}$	Factor	Δf
-1.8	3	-6	-18	108	-1.36	.32141	-4.03
-1.5	5	-5	-25	125	-1.14	.17195	-2.16
-1.2	6	-4	-24	96	-.92	.07815	-0.98
-0.9	9	-3	-27	81	-.70	.02770	-0.35
-0.6	11	-2	-22	44	-.48	.00639	-0.08
-0.3	11	-1	-11	11	-.26	.00056	-0.01
0	14	0	0	0	-.04	.00000	0
0.3	13	1	13	13	.18	.00013	0
0.6	13	2	26	52	.40	.00312	-0.04
0.9	10	3	30	90	.62	.01734	-0.22
1.2	9	4	36	144	.84	.05552	-0.70
1.5	4	5	20	100	1.06	.13198	-1.66
1.8	4	6	24	144	1.28	.25998	-3.26
Σ	112		22	1008			-13.49

$A_o = \frac{\begin{vmatrix} 112 & 182 \\ 1008 & 4550 \end{vmatrix}}{26026} = 12.53147$

$A_1 = 22/182 = 0.12088$

$A_2 = \frac{\begin{vmatrix} 13 & 112 \\ 182 & 1008 \end{vmatrix}}{26026} = -0.27972$

$\bar{x} = 0.21607$

$f(\bar{x}) = 12.54453$

$\sigma_x = 4.73533$

$\frac{1}{\sigma_x} = 0.211$, Use 0.22

$0.22\ \bar{x} = 0.048$, Use 0.04

$\bar{y} = 0.06482$

$\sigma_y = 1.42060$

$N = 149$

TABLE 18

13-POINT CURVE FIT, CONTINUED, PROBLEM 3

x	ϕ = f+Δf	xφ	$x^2\phi$	$\frac{x-\mu}{\sigma_x}$	Factor	Δφ	x	F = f+Δφ	xF	x^2F
-6	-1.03	6.18	-37.08	-1.62			-6			
-5	2.84	-14.20	71.	-1.36	.32141	-4.15	-5	0.85	-4.25	21.25
-4	5.02	-20.08	80.32	-1.10	.15108	-1.95	-4	4.05	-16.20	64.80
-3	8.65	-25.95	77.85	-.84	.05552	-0.72	-3	8.28	-24.84	74.52
-2	10.92	-21.84	43.68	-.58	.01338	-0.17	-2	10.83	-21.66	43.32
-1	10.99	-10.99	10.99	-.32	.00129	-0.02	-1	10.98	-10.98	10.98
0	14.	0.	0.	-.06	.00000	0	0	14.	0.	0.
1	13.	13.	13.	.20	.00020	0	1	13.	13.	13.
2	12.96	25.92	51.84	.46	.00540	-0.07	2	12.93	25.86	51.72
3	9.78	29.34	88.02	.72	.03087	-0.40	3	9.60	28.80	86.40
4	8.30	33.20	132.80	.98	.09886	-1.28	4	7.72	30.88	123.52
5	2.34	11.70	58.50	1.24	.23237	-3.00	5	1.00	5.	25.
6	0.74	4.44	26.64	1.50			6			
Σ	98.51	30.72	617.56			-11.76		93.24	25.61	514.51

$$B_o = \frac{\begin{vmatrix} 98.51 & 182 \\ 617.56 & 4550 \end{vmatrix}}{26026} = 12.90343$$

$B_1 = 30.72/182 = 0.16879$

$$B_2 = \frac{\begin{vmatrix} 13 & 98.51 \\ 182 & 617.56 \end{vmatrix}}{26026} = -0.3804095$$

$\bar{x} = 0.22185$

$f(\bar{x}) = 12.92215$

$\sigma_x = 4.12123$

$\bar{y} = 0.06656$

$\sigma_y = 1.23637$

$N = 133$

$\frac{1}{\sigma_x} = 0.243$, Use 0.26

$0.26\,\mu = 0.058$, Use 0.06

$$C_o = \frac{\begin{vmatrix} 93.24 & 110 \\ 514.51 & 1958 \end{vmatrix}}{9438} = 13.34688$$

$C_1 = 25.61/110 = 0.23282$

$$C_2 = \frac{\begin{vmatrix} 11 & 93.24 \\ 110 & 514.51 \end{vmatrix}}{9438} = -0.48705$$

$\bar{x} = 0.23901$

$f(\bar{x}) = 13.37470$

$\sigma_x = 3.70544$

$\bar{y} = 0.07170$

$\sigma_y = 1.11163$

$N = 124$

$\frac{1}{\sigma_x} = 0.270$, Use 0.28

$.28\,\bar{x} = 0.067$, Use 0.06

TABLE 19

9 POINT CURVE FIT, PROBLEM 3

x	$\frac{x-\mu}{\sigma_x}$	Factor	ΔF	$g = f+\Delta F$	xg	x^2g	$\frac{x-\mu}{\sigma_x}$	Factor	Δg
-4	-1.18	.19468	-2.60	3.40	-13.60	54.40	-1.26	.24592	-3.29
-3	-0.90	.07198	-0.96	8.04	-24.12	72.36	-..98	.09886	-1.32
-2	-0.62	.01734	-0.23	10.77	-21.54	43.08	- .70	.02770	-0.37
-1	-0.34	.00164	-0.02	10.98	-10.98	10.98	- .42	.00378	-0.05
0	-0.06	.00000	0.	14.	0.	0.	- .14	.00005	0.
1	0.22	.00029	0.	13.	13.	13.	.14	.00005	0.
2	0.50	.00750	-0.10	12.90	25.80	51.60	.42	.00378	-0.05
3	0.78	.04191	-0.56	9.44	28.32	84.96	.70	.02770	-0.37
4	1.06	.13198	-1.77	7.23	28.92	115.68	.98	.09886	-1.32
5	1.34	Not Used							
$\sum$			-6.24	89.76	25.80	446.06			-6.77

$$D_o = \frac{\begin{vmatrix} 89.76 & 60 \\ 446.06 & 708 \end{vmatrix}}{2772} = 13.27074$$

$D_1 = 25.80/60 = 0.43000$

$$D_2 = \frac{\begin{vmatrix} 9 & 89.76 \\ 60 & 446.06 \end{vmatrix}}{2772} = -0.49461$$

$\bar{x} = 0.43469$

$f(\bar{x}) = 13.36419$

$\sigma_x = 3.67557$

$\bar{y} = 0.13041$

$\sigma_y = 1.10267$

$N = 123$

$\frac{1}{\sigma_x} = 0.272$, Use 0.28

$.28\ \bar{x} = 0.122$, Use 0.14

TABLE 20

9 POINT CURVE FIT, CONTINUED, PROBLEM 3

x	G = f+Δg	xG	x^2G
-4	2.71	-10.84	43.36
-3	7.68	-23.04	69.12
-2	10.63	-21.26	42.52
-1	10.95	-10.95	10.95
0	14.	0	0
1	13.	13.	13.
2	12.95	25.90	51.80
3	9.63	28.89	86.67
4	7.68	30.72	122.88
$\sum$	89.23	32.42	440.30

$$E_o = \frac{\begin{vmatrix} 89.23 & 60 \\ 440.30 & 708 \end{vmatrix}}{2772} = 13.26004$$

$E_1 = 32.42/60 = 0.54033$

$$E_2 = \frac{\begin{vmatrix} 9 & 89.23 \\ 60 & 440.30 \end{vmatrix}}{2772} = -0.50184$$

$\bar{x} = 0.53835$

$f(\bar{x}) = 13.40549$

$\sigma_x = 3.65463$

$\bar{y} = 0.16151$

$\sigma_y = 1.09639$

$N = 123$

Check:

$\frac{1}{\sigma_x} = 0.274$

$.274\ \bar{x} = 0.148$

THE 1979 SAMUEL S. WILKS MEMORIAL MEDAL

Frank E. Grubbs

The Samuel S. Wilks Memorial Medal Award was initiated in 1964 by the US Army and the American Statistical Association, and has been administered for the Army by the American Statistical Association, a non-profit, educational and scientific society founded 140 years ago in 11839. The Wilks Medal and Award is given each year to a statistician - and a top-notch one! - and is based primarily on his contributions to the advancement of scientific or technical knowledge in Army statistics, ingenious application of such knowledge, or successful activity in the fostering of cooperative scientific matters which coincidentally benefit the Army, the Department of Defense, the US Government, and our country generally. The Award consists of a medal, with a profile of Professor Wilks and the name of the Award on one side, the seal of the American Statistical Association and the name of the recipient on the reverse side, and a citation and honorarium related to the magnitude of the Award funds, which were generously donated by Phillip G. Rust of the Winnstead Plantation, Thomasville, Georgia. Mr. Rust originally stimulated the interest of Sam Wilks in distributional properties of the "extreme spread" (bivariate range), a measure of the "accuracy" of rifle shot on a target.

These annual Army Design of Experiments Conference, at which the Wilks Medal is awarded each year, are sponsored by the Army Mathematics Steering Committee on behalf of the Office of the Chief of Research, Development and Acquisition, Department of the Army.

Previous recipients of the Samuel S. Wilks Memorial Medal include John W. Tukey of Princeton University (1965), Major General Leslie E. Simon (retired, 1966), William G. Cochran of Harvard University (1967), Jerzy Neyman of the university of California, Berkeley (1968), Jack Youden (deceased) formerly of the National Bureau of Standards (1969), George W. Snedecor (deceased) formerly of Iowa State University (1970), Harold Dodge (deceased) formerly of the Bell Telephone Laboratories (1971), George E. P. Box of the University of Wisconsin (1972), H. O. Hartley (1973), this year's President of the American Statistical Association, Cuthbert Daniel, private statistical consultant (1974), Herbert Solomon of Stanford University (1975), Solomon Kullback of George Washington University (1976), Churchill Eisenhart of the National Bureau of Standards (1977), and William Kruskal of the University of Chicago (1978).

This brings us up to this year, for which the competition for the Wilk's Medal turned out to be keen indeed, and as usual the "best man won". The members of the 1979 Wilk's Memorial Medal Committee consisted of individuals skilled in the art of arguing their points and getting their best views in the minds of others! They were: Chruchill Eisenhart, Fred Frishman, Frank Grubbs (Chairman), Bill Kruskal, Jeff Kurkjian, and Frank Proschan. They had the job of concentrating on some 12 deserving candidates from many nominees, and coming up with their best selection according to the Wilk's Medal criteria.

The 1979 Wilks Medalist, like Sam Wilks, was born in Texas. He received his B. A. in Physics (with highest honours) in 1934 from the University of Texas, and later his Ph. D. in Mathematics from Princeton University in 1940, in the first cohort of Wilks disciples. Between 1934 and 1940 he taught as a graduate assistant or instructor at Brown University, the University of Texas, and Princeton University, while also during that period he made the transition from Physics through Applied Mathematics to the great field of Mathematical Statistics. After obtaining his Ph. D. in 1940, he returned to the University of Texas as Instructor in Applied Mathematics and Astronomy, interrupting his academic career to join the Bureau of Labor Statistics in 1942, and in 1944 a project of the Applied Mathematics Panel of the National Defense Research Council (I believe under Sam Wilks). Then from 1945 until 1948 he was Professor of Mathematics and Statistics at Iowa State University, and then joined the RAND Corporation, where he served as Deputy Chief of the Mathematics Division until 1955.

At this point in time the entrepreneur emerged and the 1979 Wilks Medalist founded the General Analysis Corporation, served as its President until 1960, when it merged with CEIR, INc. He then became a Vice-President of CEIR and Manager of its Western Division until 1964, when he next went to Washington as Assistant Commissioner of Educational Statistics in the Office of Education. He was then Director of the National Center of Educational Statistics until 1967, after which he returned to the West Coast as Professor of Administration and Director of the Public Policy Research Organization at the University of California, Irvine.

There is much, much to say about this scholar and gentleman concerning his great contributions to the field of statistics generally, although they should be recorded elsewhere. We should note, however, that he wrote the best key graduate book for a solid statistics course, and one which has trained many good statisticians. (Introduction to the Theory of Statistics). He has been a prolific publisher of technical papers on statistics, operations research, education and public policy research. He has been Presidents of both the Institute of Mathematical Statistics and the Operations Research Society of America.

A long-term friend and colleague, George W. Brown, told me that the 1979 Wilks Medalist is an "extraordinarily and deceptively quiet man", so that I don't think he would win an award as the "most talkative statistician"! Yet, he has exerted major direct and indirect influences on an enormous number of individuals. And he has had many separate careers actually, including roles as a professor, a think-tank researcher, an operations research analyst, an administrator, an entrepreneur and manager, a pioneering public servant, founding director of an important research organization, and distinguished consultant and advosor to universities and government organizations.

By now it should be unmistakebly clear that we are referring to none other than Alexander M. Mood.

The citation for Alex Mood reads:
"To Alexander M. Mood for his many significant contributions to the theory of statistics, an outstanding textbook on the subject, his extensive applications to operations research and systems analysis, and unique statistical assessments of education and public policy research."

Variance Reduction in Monte Carlo Simulation

Mark Brown
Florida State University

Herbert Solomon
Stanford University

Michael A. Stephens
Simon Fraser University, B. C.

1. Introduction.

Monte Carlo simulation is employed in a large variety of problems. Frequently, one is interested in the expectation of a function $g(X_1,\ldots,X_N)$ where $< X_i,\ i \geq 1 >$ is i.i.d. with known distribution F and N is a stopping time (often a constant). The procedure followed is to generate a large number of samples $(X_1^{(i)},\ldots,X_{N_i}^{(i)})$, $i = 1,2,\ldots,M$, and estimate the expectation of interest by

$$\frac{1}{M}\sum_{i=1}^{M} g(X_1^{(i)},\ldots,X_{N_i}^{(i)}) .$$

An interesting aspect of the simulation estimation problem is that F is known. Thus functions of the form $\ell(F,X_1,\ldots,X_N)$ can be employed as estimators, while in statistical estimation problem with F unknown ℓ cannot be computed from the data and is thus not considered to be an estimator. Thus the class of estimators is considerably wider in Monte Carlo problems.

One approach available to reduce the variance of the Monte Carlo estimator is to find a function $\ell(F,X_1,\ldots,X_N)$ with the same expectation as g, and with smaller variance. Then ℓ rather than g is averaged over the M samples. Of course, $\ell = E_F g$ fits this description but were

Partially supported under U.S. Army Research Office Grant DAAG-29-77-G-0031 and issued as Technical Report No. 35.

it directly computable one would not need to simulate in the first place. Thus an important requirement of ℓ is that it be simply computable.

We illustrate the above remarks by considering the problem of Monte Carlo estimation of $M(t) = EN(t)$, the expected number of renewals in $[0,t]$ for a renewal process with known interarrival time distribution F. Several unbiased estimators which compete favorably with the naive estimator, $N(t)$, are presented and studied.

We believe that our approach and methodology, although only applied to renewal function estimation in this paper, can be useful in a large variety of Monte Carlo simulation problems.

2. Assume that $< X_i, i \geq 1 >$ is i.i.d. with cdf F where $F(0) = 0$. Define $S_0 = 0$, $S_n = \sum_1^n X_i$, $n = 1,2,\ldots, N(t) = \max\{n: S_n \leq t\}$, and $M(t) = EN(t)$, $t \geq 0$. Sometimes we consider the point $t = 0$ as a renewal epoch. In this case we use $N_0(t) = N(t)+1$ and $M_0(t) = M(t)+1$. The renewal age at time t is defined by $A(t) = t-S_{N(t)}$; $\Pr(A(t) = t) = \bar{F}(t)$ and $dF^{(x)}_{A(t)} = \bar{F}(x)dM(t-x)$ for $0 \leq x < t$, thus $dF^{(x)}_{A(t)} = \bar{F}(x)dM_0(t-x)$ for $0 \leq x \leq t$.

Define

$$\delta_i = \begin{cases} 1 & \text{if } S_i \leq t \\ 0 & \text{if } S_i > t \,. \end{cases}$$

Then $N(t) = \sum_1^\infty \delta_i$ and $M(t) = E\sum_1^\infty \delta_i = \sum_1^\infty F^{(i)}(t)$, where $F^{(i)}$ is the i^{th} convolution of F.

To estimate $F^{(i)}(t) = E\delta_i$ we will use

$$E(\delta_i | X_1, \ldots, X_{i-1}) = E(\delta_i | S_{i-1}) = F(t-S_{i-1}) .$$

We then estimate $M(t)$ by:

$$(1) \qquad M_F(t) = \sum_{i=1}^{\infty} F(t-S_{i-1}) = \sum_{i=1}^{N(t)+1} F(t-S_{i-1}) .$$

Since $\mathrm{Var}(F(t-S_{i-1})) = \mathrm{Var}[E(\delta_i | S_{i-1})] \leq \mathrm{Var}\ \delta_i$, we have replaced each component, δ_i, by a component with the same expectation and smaller variance. Intuitively we would expect that if we reduce the variability at each stage (given the past) then we should reduce the variability of the overall estimator. However, the computation of variance involves covariance terms, and if these are increased while variances are decreased there can conceivably be an increase in variance. Theorem 1 (below) demonstrates that $M_F(t)$ does indeed have lower variance than $M(t)$.

Theorem 1. $M_F(t)$ is an unbiased estimator of $M(t)$ and $\mathrm{Var}\ N(t) - \mathrm{Var}\ M_F(t) = E[2M(A(t)) - F(A(t))] \geq 0$, with strict equality if $F(t) > 0$.

Before proving theorem 1 we comment that the reduction in variance is unsatisfactorily small for large t. If $\mu_2 = EX^2 < \infty$ then $E[2M(A(t)) - F(A(t))] = O(1)$, thus $\mathrm{Var}\ N(t)$ and $\mathrm{Var}\ M_F(t)$ are of the form $\gamma t + O(1)$ with common γ, and we improve only the asymptotically negligible $O(1)$ term. Estimators considered in later sections do considerably better for large t.

Proof of Theorem 1. Express $M_F(t)$ as

$$F(t) + \int_0^t F(t-x)dN(x) = \int_0^t F(t-x)dN_0(x) .$$

Then

$$EM_F(t) = \int_0^t F(t-x)dM_0(x) = \int_0^t 1dM_0(x) - \int_0^t \bar{F}(t-x)dM_0(x)$$

$$= M_0(t) - \int_0^t dF_{A(t)}^{(t-x)} = M_0(t)-1 = M(t) .$$

Now,

$$EM_F^2(t) = \int_0^t F^2(t-x)dM_0(x)$$

$$+ 2 \iint_{r<s} F(t-r)F(t-s)dM_0(r)dM_0(s-r) .$$

We evaluate this expression in several steps:

(i) $$\int_0^t F^2(t-x)dM_0(x) = \int_0^t F(t-x)dM_0(x) - \int_0^t F(t-x)\bar{F}(t-x)dM_0(x)$$

$$= M(t) - EF(A(t)) .$$

(ii) $$F(t-r)F(t-s) = 1-\bar{F}(t-r) - \bar{F}(t-s) + \bar{F}(t-r)\bar{F}(t-s) .$$

(iii) $$2 \iint_{r<s} 1dM_0(r)dM_0(s-r) = 2 \int_0^t M(t-r)dM_0(r) = 2M(t) + 2M^{(2)}(t) .$$

(iv) $$-2 \iint_{r<s} \bar{F}(t-r)dM_0(r)dM_0(s-r) = -2 \int_{r=0}^t \bar{F}(t-r)M(t-r)dM_0(r)$$

$$= -2EM(A(t)) .$$

$$(v) \qquad -2 \iint_{r<s} \bar{F}(t-s)dM_0(r)dM_0(s-r) = -2\int_{r=0}^{t} F(t-r)dM_0(r)$$

$$= -2M(t) .$$

$$(vi) \qquad 2 \iint_{r<s} \bar{F}(t-r)\bar{F}(t-s)dM_0(r)dM_0(s-r) = 2\int_{r=0}^{t} F(t-r)\bar{F}(t-r)dM_0(r)$$

$$= 2EF(A(t)) .$$

Combining (i)-(vi) we obtain:

$$(2) \qquad EM_F^2(t) = M(t) + 2M^{(2)}(t) - E(2M(A(t)) - F(A(t))) .$$

Furthermore

$$(3) \qquad EN^2(t) = E[\int_0^t 1dN(t)]^2 = M(t) + 2 \iint_{r<s} dM(r)dM(s-r)$$

$$= M(t) + 2M^{(2)}(t) .$$

Thus from (2) and (3):

$$\text{Var } N(t) - \text{Var } M_F(t) = E[2M(A(t)) - F(A(t))] .$$

Since

$$M(s) = \sum_{i=1}^{\infty} F^{(i)}(s), \quad 2M(s) - F(s) = F(s) + 2\sum_{i=2}^{\infty} F^{(i)}(s) \geq 0 ;$$

thus $E[2M(A(t)) - F(A(t))] \geq 0$ for all t and is strictly positive for $F(t) > 0$.

3. In this section we assume that F is continuous. The cumulative hazard H is defined by $H(t) = -\log \bar{F}(t)$. When F is absolutely continuous with density f then $H(t) = \int_0^t h(y)dy$ where h is the hazard function, $h(t) = \frac{f(t)}{\bar{F}(t)}$.

Our next estimator is based on the intuitive idea that $E(dN(s)|\text{past}) = dH(A(s))$. Thus instead of using $N(t) = \int_0^t dN(s)$ we try

$$M_H(t) = \int_0^t dH(A(s)) = \sum_1^{N(t)} H(X_i) + H(A(t)) = \sum_1^{N(t)+1} H_i$$

where $H_i = H[(t-S_{i-1}) \wedge X_i]$ (where $a \wedge b = \min(a,b)$).

Note that $N(t) = \sum_1^\infty \delta_i$ while $M_H(t) = \sum_1^\infty H_i$. Thus δ_i is replaced by H_i, and $E(\delta_i|S_{i-1}) = E(H_i|S_{i-1}) = F(t-S_{i-1})$.

The process $M_H(t)$ is a cumulative process in the sense of Smith [3]. Thus (Smith [3])

$$\text{Var } M_H(t) \sim \frac{t}{\mu} E[H(X) - (\frac{EH(X)}{\mu})X]^2 ,$$

where $\mu = EX$. But $H(X) = -\log \bar{F}(X)$ is exponentially distributed with parameter 1, thus:

$$E[H(X) - \frac{EH(X)}{EX} X]^2 = 1 + \frac{\sigma^2}{\mu^2} - \frac{2\sigma\rho}{\mu} ,$$

where ρ is the correlation coefficient between X and $H(X)$ and σ^2 is the variance of X. Thus $M_H(t)$ is asymptotically better than $N(t)$ for $\rho > \mu/2\sigma$, asymptotically worse than $N(t)$ for $\rho < \mu/2\sigma$.

In general if we have two unbiased estimators of a parameter, T_1 and T_2, with covariance matrix A, then the minimum variance unbiased estimator of the form $\alpha T_1 + (1-\alpha)T_2$ is the one with

$$\alpha = \frac{\sum_{j=1}^{2} A_{1j}^{-1}}{\sum_{i=1}^{2}\sum_{j=1}^{2} A_{ij}^{-1}} .$$

The variance of this estimator is

$$\frac{1}{\sum_{i,j} A_{ij}^{-1}} .$$

The idea now is to let A be the asymptotic covariance matrix of

$$\left(\frac{N(t)}{\sqrt{t}} , \frac{M_F(t)}{\sqrt{t}}\right)$$

and to employ the above result to obtain an unbiased estimator which improves on both $M_H(t)$ and $N(t)$ for large t. We already know the $O(t)$ terms for Var $N(t)$ and Var $M_H(t)$. We only need the leading term for Cov$(N(t), M_H(t))$. This is given in lemma 1 below.

Lemma 1. If σ^2 is finite then

$$\text{Cov}(N(t), M_H(t)) = \frac{t}{\mu}\left(\frac{\sigma^2}{\mu^2} - \frac{\sigma\rho}{\mu}\right) + o(t) .$$

Proof.

$$\mathrm{Var}(N(t) - M_H(t)) = \mathrm{Var} \sum_1^\infty (\delta_i - H(t - S_{i-1} \wedge X_i))$$

$$= \sum_1^\infty E\ \mathrm{Var}[\delta_i - H(t - S_{i-1} \wedge X_i) | S_{i-1}] = E \sum_1^\infty F(t - S_{i-1}) = EN(t) = M(t) .$$

Thus

$$M(t) = \mathrm{Var}(N(t) - M_H(t)) = \mathrm{Var}\ N(t) + \mathrm{Var}(M_H(t)) - 2\mathrm{Cov}(N(t), M_H(t)) ,$$

and therefore

$$\mathrm{Cov}(N(t), M_H(t)) = \frac{1}{2} [\mathrm{Var}\ N(t) + \mathrm{Var}\ M_H(t) - M(t)]$$

$$= \frac{t}{2\mu} [\frac{\sigma^2}{\mu^2} + \frac{\sigma^2}{\mu^2} + 1 - \frac{2\rho\sigma}{\mu} - 1 + o(1)]$$

$$= \frac{t}{\mu} (\frac{\sigma^2}{\mu^2} - \frac{\sigma\rho}{\mu}) + o(t). \quad \|$$

Now

$$A = \frac{1}{\mu} \begin{pmatrix} \frac{\sigma^2}{\mu^2} & \frac{\sigma^2}{\mu^2} - \frac{\rho}{\mu} \\ \frac{\sigma^2}{\mu^2} - \frac{\sigma\rho}{\mu} & \frac{\sigma^2}{\mu^2} + 1 - \frac{2\sigma\rho}{\mu} \end{pmatrix},$$

$$\alpha = \frac{\sum_{j=1}^{2} A_{1j}^{-1}}{\sum_{i=1}^{2} \sum_{j=1}^{2} A_{ij}^{-1}} = 1 - \frac{\sigma\rho}{\mu}$$

and

$$\frac{1}{\sum_{i,j} A^{-1}_{ij}} = \frac{\sigma^2}{\mu^3}(1-\rho^2) .$$

Note that the asymptotic relative savings in variance is ρ^2 the square of the correlation coefficient between X and $H(X)$. Summarizing:

Theorem 2. The estimator

$$M^*(t) = (1 - \frac{\sigma\rho}{\mu})N(t) + \frac{\sigma\rho}{\mu} M_H(t)$$

is an unbiased for $M(t)$ with variance

$$\frac{t\sigma^2}{\mu^3}(1-\rho^2) + o(t)$$

(ρ is the correlation coefficient between X and $H(X)$). It follows that:

$$\frac{\text{Var } N(t) - \text{Var } M^*(t)}{\text{Var } N(t)} = \rho^2 + o(1) .$$

Example: Let $H(x) = x^2$, $\bar{F}(x) = e^{-x^2}$. Then,

$$\mu = \int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-x^2} dx = \frac{\sqrt{\pi}}{2} ;$$

$$EX^2 = 2\int_0^\infty x e^{-x^2} dx = 1 ,$$

thus

$$\sigma^2 = 1 - \frac{\pi}{4} = \frac{4-\pi}{4} ; \quad \rho = \frac{1}{\sigma}[\int 2x^4 e^{-x^2} dx - \mu] = \frac{\mu}{2\sigma} = \frac{1}{2}\sqrt{\frac{\pi}{4-\pi}} , \quad \rho^2 = \frac{\pi}{4(4-\pi)} = .915 .$$

Thus in this case (Weibull with shape parameter 2) the unbiased estimator $M^*(t)$ has an asymptotic relative reduction in risk over $N(t)$ of 91.5 percent. ‖

Integration by parts shows that

$$\rho = \frac{1}{\sigma}\int_0^\infty H(x)\bar{F}(x)dx \ ;$$

since $H(x) = -\log \bar{F}(x)$ the integral can probably be given an enthropy interpretation. Also $\rho = \frac{1}{\sigma} E\tilde{H}(X)$ where $\tilde{H}(x) = \int_0^x H(z)dz$. This is true since

$$\int_0^\infty H(x)\bar{F}(x)dx = \int_0^\infty H(x)EI_{X>x}\,dx = E\int_0^\infty H(x)I_{X>x}dx = E\int_0^X H(x)dx = E\tilde{H}(X).$$

Note that both ρ and $\frac{\rho\sigma}{\mu}$ are invariant under a change of time scale, $t \to ct$, $c > 0$.

4. In section 3 we estimated $M(t)$ by a weighted average of $N(t) = \sum_1^\infty \delta_i$ and $M_F(t) = \sum_1^{N(t)+1} H((t-S_{i-1}) \wedge X_i)$. Now we apply the same idea but stagewise. At stage i, having observed $X_1,\ldots,X_{i-1}$, $N(t)$ adds the component $\delta_i = I_{X_i \le t-S_{i-1}}$, while $M_F(t)$ adds $H_i = H((t-S_{i-1}) \wedge X_i)$. Each of δ_i, H_i are conditionally (given S_{i-1}) unbiased for $F(t-S_{i-1})$ and unconditionally unbiased for $F^{(i)}(t)$. The approach we now follow is to use the weighted average of δ_i and H_i which has smallest conditional variance given $X_1,\ldots,X_{i-1}$.

Define $F_i = F(t-S_{i-1})$, $C_i = H(t-S_{i-1})$. Then:

$$\mathrm{Var}(\delta_i | S_{i-1}) = F_i - F_i^2$$

$$\mathrm{Cov}(\delta_i, H_i | S_{i-1}) = \bar{F}_i (F_i - C_i)$$

$$\mathrm{Var}(H_i | S_{i-1}) = F_i + \bar{F}_i (F_i - 2C_i) .$$

The minimum conditional variance (given $X_1, \ldots, X_{i-1}$) unbiased linear combination is then:

$$L_i = (1 - \frac{C_i \bar{F}_i}{F_i})\delta_i + \frac{C_i \bar{F}_i}{F_i} H_i .$$

The corresponding estimator of $M(t)$ is:

$$M_L(t) = N(t) - \sum_{1}^{N(t)+1} \frac{H(t-S_{i-1})\bar{F}(t-S_{i-1})}{F(t-S_{i-1})} (\delta_i - H_i) .$$

We do not know how $M_L(t)$ compares with the other estimators we have looked at. The variance of an estimator of the form ΣK_i is $\Sigma \mathrm{Var}\, K_i + 2 \sum_{i<j} \mathrm{Cov}(K_i, K_j)$; L_i was chosen from among a class of estimators ΣK_i to minimize $\Sigma \mathrm{Var}\, K_i$. However we know very little about $\mathrm{Cov}(L_i, L_j)$. This latter quantity must be shown to be suitably small in order to demonstrate that $M_L(t)$ has desirable variance properties.

5. We next consider an unbiased estimator with asymptotic variance $O(1)$. Thus it asymptotically enjoys a 100 percent reduction in variance over $N(t)$.

As is well known $N(t)+1$ is a stopping time and thus by Wald's identity:

$$ES_{N(t)+1} = E\sum_{1}^{N(t)+1} X_i = \mu(M(t)+1) .$$

Thus

$$\hat{M}(t) = \frac{S_{N(t)+1}}{\mu} - 1$$

is unbiased for $M(t)$. Now $Var(S_{N(t)+1}) = Var(t+Z(t)) = Var\ Z(t)$, where $Z(t)$ is the forward recurrence time at t. If $\mu_3 = EX^3 < \infty$ then $Var\ Z(t)$ converges to

$$\frac{\mu_3}{3\mu} - \frac{\mu_2^2}{4\mu^2} = \frac{4\mu\mu_3 - 3\mu_2^2}{12\mu^2}$$

as $t \to \infty$. Thus

$$Var\ \hat{M}(t) \to \frac{4\mu\mu_3 - 3\mu_2^2}{12\mu^4},$$

and is thus $O(1)$.

References

[1] Brown, M. and Ross, S.M. (1972). "Asymptotic properties of cumulative processes". Siam Journal of Applied Mathematics, 22, 93-105.

[2] Feller, W. (1970). An Introduction to Probability Theory and Its Applications, Vol. II, Second Edition, John Wiley, N.Y.

[3] Smith, W.L. (1955). "Regenerative stochastic processes". Proc. Roy. Soc. A, 232, 6-31.

SMALL SAMPLE SIZE EFFECTS ON TOLERANCE LIMITS, EXCEEDANCES

JAMES R. KNAUB, JR.
ARMY MATERIEL TEST AND EVALUATION DIRECTORATE
US ARMY WHITE SANDS MISSILE RANGE
WHITE SANDS MISSILE RANGE, NEW MEXICO

ABSTRACT - Tolerance Limits, exceedances and includances, are useful indications of the adequacy of a sample size. However, for very small sample sizes, such results become very sensitive and may require a thorough analysis before concluding whether sample sizes are adequate. Some measures of sensitivity are investigated in this paper.

1. Introduction - When dealing with missile systems or any other materiel which consists of a relatively small number of very expensive items which will not survive a test, one would not like to destroy any more materiel than necessary, so sample sizes are made small. Time may also be a factor in keeping sample sizes small. When testing to specified test objectives, however, small sample sizes cause large uncertainties in the results obtained. In hypothesis testing, for example, a small sample size means that the power of the test is low, and therefore one's ability to discriminate between an untrue null hypothesis and a true alternative hypothesis may be low even when the two hypotheses are very different. Although the power of a test is very important in

missile work, as it shows the sensitivity of results to small sample sizes, it is often ignored. An example is the use of Wilcoxon's Rank Sum Test which is often used with no power calculated, even though it is easily obtainable from a paper written by E. L. Lehmann in 1953 (Lehmann, "The Power of Rank Tests," Annals of Mathematical Statistics, 24 (1953), 23-43).

Any measure of sensitivity which provides the likelihood of confusing one result for another would be analogous, to a degree, to power. The central question is, "Is the sample size sufficient to reduce to an acceptable level, the risk of saying that more is known than actually is known?"

From Gumbel, Statistics of Extremes, pages 97 and 103-104 (see also Hoel, Introduction to Mathematical Statistics, pages 274-277), the following equation is produced (due to S.S. Wilks):

$$P = 1-n\gamma^{n-1} + (n-1)\gamma^{n}$$

where $P \equiv P$[at least 100γ% of the population is between the smallest and largest observation of the sample of size n].
A first approximation (which appears to be low for small sample sizes) is given as

$$n \approx \frac{1}{2} + \frac{\sqrt{2P}}{1-\gamma}$$

The previous equation can be solved iteratively from here using small increments.

This is useful information. However, for very small sample sizes, perhaps a measure of sensitivity as described

earlier may be desired. To make one further analogy, when sample sizes are very large, point estimates may be sufficient, however, when sample sizes are small, lower confidence limits are needed in order to avoid misleading decision makers. What is to follow is based upon the theory of exceedances. (See Bradley, Distribution-Free Statistical Tests, pages 216-218.)

2. A Measure of Sensitivity - Consider an initial sample X, made up of n observations, ranked from r equal 1 to n, from which one wishes to make predictions about the next m observations in a sample, Y. If b represents the number of observations in Y which have values lower than the rth lowest value in X, then (from Bradley, pages 216-218),

$$P[\text{exactly } b \text{ of the } Y\text{'s will be } < X_r] = \frac{\binom{r-1+b}{r-1}\binom{n+m-r-b}{n-r}}{\binom{n+m}{n}}.$$

Letting $p_1 = \lim_{m\to\infty} \frac{b_1}{m}$ and $p_o = \lim_{m\to\infty} \frac{b_o}{m}$, where $p_o<p_1$, one can investigate the case of n, a finite sample size, but n+m, the infinite population size. Now let β be used to represent a measure of sensitivity for tolerance limits, in some ways analogous to, but not the same as, the probability of a type II error. Here,

$$\beta \equiv P[100p_1\% \text{ or more } Y\text{'s} < X_r] : P[100p_o\% \text{ or fewer } Y\text{'s} < X_r].$$

β, here, is not a probability, but a ratio of probabilities.

It shows the relative probability of having an unacceptable ($100p_1$%) percent of the population values for a certain parameter (e.g., missile lethality), or more, fall below the rth lowest value in the X sample, as compared to a hoped for $100p_o$% or fewer members of the population falling in that category. Therefore, the smaller β, the more sensitive the results are to implying more is known than can be known from a certain sample size. (i.e., the smaller β, the better.)

The following briefly sketches the derivation of β:

$$\beta = \lim_{m\to\infty} \frac{\sum_{j=b_1}^{m} \binom{r-1+j}{r-1}\binom{n+m-r-j}{n-r}}{\sum_{i=o}^{b_o} \binom{r-1+i}{r-1}\binom{n+m-r-i}{n-r}}$$

$$= \lim_{m\to\infty} \frac{\sum_{j=b_1}^{m} [(r-1+j)(r-2+j)\dots(j+1)(n+m-r-j)(n+m-r-j-1)\dots(m-j+1)]}{\sum_{i=o}^{b_o} [(r-1+i)(r-2+i)\dots(i+1)(n+m-r-i)(n+m-r-i-1)\dots(m-i+1)]}$$

For m large, $b_1 \doteq p_1 m$ and $b_o \doteq p_o m$, so sum in the numerator from $(p_1 m)^{r-1}[m(1-p_1)]^{n-r}$ to $m^{n-1}(n-r)!$, and in the denominator from $(r-1)!m^{n-r}$ to $(p_o m)^{r-1}[m(1-p_o)]^{n-r}$. Since there are an infinite number of terms, these summations become integrals. For r=1, n=5, the numerator becomes $\lim_{m\to\infty} \int_{p_1 m}^{m} (4+m-j)(3+m-j)(2+m-j)(1+m-j)dj$, which makes the integral approach $m \int_{p_1 m}^{m} (m-j)^4 dj$. For r=2, n=5, the integral is $\int_{p_1 m}^{m} (j+1)(3+m-j)(2+m-j)(1+m-j)dj \to \int_{p_1 m}^{m} j(m-j)^3 dj$.

In general then, the integrals will be of the form

$$\int j^{r-1}(m-j)^{n-r}dj.$$

The only difference between numerator and denominator are the limits of the integration.

For r=1, n=5, one has

$$\beta = \frac{1-p_1+2(p_1^2-1)+2(1-p_1^3)+p_1^4-1+\frac{1}{5}(1-p_1^5)}{p_o-2p_o^2+2p_o^3-p_o^4+\frac{1}{5}p_o^5}.$$

As an example, letting $p_1=0.3$ and $p_o=0.2$, $\beta(r=1, n=5, p_o=0.2, p_1=0.3)\approx 0.42$. This means that the probability that 30% or more of the population will have values below the lowest in the sample of size 5 is 42% of the probability that 20% or less of the population will be that low. If one does not wish to tolerate having more than 20% of the population that low, then there is cause for a great deal of alarm, especially if 30% is an unreasonable alternative.

One can determine the exact probability of having 100a% or more of the population fall below the rth value of the sample, and also the exact probability of having 100a% or less of the population fall there. Call the first probability x and the second y. Then, one has

$x+y = 1$, and

$$\beta(p_o=p_1=100a\%) = \frac{x}{y}$$

Therefore, for such a β, x and y,

$$x = \frac{\beta}{1+\beta} \quad \text{and} \quad y = \frac{1}{1+\beta}$$

As an example, $\beta(r=1, n=5, p_o=0.2, p_1=0.2) \approx 0.55$, $P[20\%$ or more of the Y's $< X_1] \approx 0.36$, and $P[20\%$ or less of the Y's $< X_1] \approx 0.64$.

For r=2, n=5, one has

$$\beta = \frac{\frac{1}{2}(1-p_1^2) + p_1^3 - 1 + \frac{3}{4}(1-p_1^4) + \frac{1}{5}(p_1^5-1)}{\frac{1}{2}p_o^2 - p_o^3 + \frac{3}{4}p_o^4 - \frac{1}{5}p_o^5}.$$

As an example, for $p_o=0.4$ tested against the alternative that $p_1=0.5$, $\beta \approx 0.28$.

As can be seen from the above, one may calculate, a priori, what ranges of values of n and r may be used in order to have an adequate degree of faith in the accuracy of results. A major advantage in this approach is that one may use r>1. Often one can not obtain an adequate sample size without a large probability that at least one of the observations will be lower than a value toward which one would like to test.

3. A More Subjective Approach to Sensitivity. When planning sample sizes for a test, there is an additional approach which may be helpful. Consider Danziger and Davis, "Tables of Distribution-Free Tolerance Limits," Annals of Mathematical Statistics, 35 (1964), 1361-1365. From Danziger and Davis, "The probability that N_o of the Y_i's lie above X_r is given by:

$$P(N_o) = \binom{N_o+n-r}{N_o}\binom{N-N_o+r-1}{N-N_o} / \binom{N+n}{n}."$$

Here, i denotes a ranking of the Y sample, N is what has been labelled m, and N_o is the number of Y_i's above X_r, where b

has been used earlier to represent the number of them below it.

From the table by Danziger and Davis, with probability at least 0.75, at least 0.752 of a population will lie above the second lowest value in a sample of ten. With probability at least 0.75, however, the proportion of the population above the third lowest in a sample of ten is only at least 0.644. Therefore, if X_2-X_3 is very small, then a sample of size ten is not adequate in that it can not distinguish very well between 0.752 and 0.644 as lower probability bounds to the proportion of the population values above such a point. Examining the table by Danziger and Davis shows that, in the cases shown, this form of sensitivity, as are most forms of sensitivity, is extremely sample size dependent.

As a practical example of how this approach could be used, consider the case of determining whether targets are detected by a certain range. Imagine that the same number of targets are detected before a critical range in each of two tests using equal sample sizes. In one case, say all (or many) of the detections before the critical range were just barely before it, but in the other case, they were very early. Thus, a slight change in critical range would cause a large change in the proportion of the former population values believed to be above this point, but no change in the latter case. (This situation is expected when dealing with rank procedures.)

4. Conclusion - When analyzing data, one must be particularly careful in dealing with small sample sizes. It is highly desirable to obtain every bit of information possible from such samples, but it is equally desirable to avoid claiming more knowledge of the population than the sample can actually provide. Balancing cost considerations against information obtainable can only be accomplished by careful consideration of all aspects. Tolerance limits and exceedances have a role in such considerations.

THE PERIODIC NATURE OF EXPERIMENTALLY MEASURED DATA

Michael Hacskaylo
US Army Electronics Research and Development Command
Night Vision and Electro-Optics Laboratory
Infrared Technology Division
Fort Belvoir, Virginia 22060

ABSTRACT. There is a phenomenon that appears to exist in the portrayal of measured data for nearly all types of experiments. It is observed that when experimental data points are plotted as a function of a variable, the data points are portrayed periodically rather than randomly about the theoretical curve. This is in contrast to an expected random distribution of the measured data. Although not always well defined, two distinct periodic characteristics can be pointed out: The first is a sinusoidal characteristic that appears to be cyclic in 10-12 measured events, and the second is the "cusp" characteristic that appears to be cyclic in about every three to four measurements. A method, which is based on a heuristic equation that relates a "periodic" probability of the arithmetic deviation with the associated measurement, generates data points which are in agreement with some published experimental values.

I. INTRODUCTION. There is a phenomenon that appears to exist in the portrayal of measured data for nearly all types of experiments. The phenomenon is that when experimental data points are plotted as a function of a variable, the points are manifested in a periodic fashion about some smoothly drawn curve. The periodic appearance of the data points is in contrast to an expected random distribution of the measured data. The periodicity, although not always well defined, can be seen upon examination of such plots in various technical journals. Two distinct periodic characteristics can be pointed out. The first is a sinusoidal characteristic that appears to be cyclic in 10-12 measured events (1, 2, 3, 4). The second is the "cusp" characteristic (5) that appears to be cyclic in about every three measurements, (6, 7, 8, 9) whereas other curves appear to be a combination of both characteristics (10). The cited references are specific examples selected from the literature that clearly exhibit the periodicity. However, most other figures show data points, that exhibit the periodic deviations of the data points less clearly, but nearly all of the experimental plotted data points show the phenomenon regardless of the physical parameters that were measured. This phenomemon is based on observation devoid of *a priori* concepts of the randomness of events and associated measurements. It is thus concluded that there is a non-random relationship between the experimentally measured data and the sequence of measuring events. In order to develop a mathematical treatment for the periodicity of experimentally measured data, the following postulated is made: If for a well-ordered experiment, the experimentally measured data are taken in an identical manner at equal increments of the independent parameter, the data points will be periodic as a function of the (measuring) events about the average curve. Thus, the experimentally measured data can be determined from the probability of the associated measuring event, in conjunction with

the equation for the experiment, it is, in principle, possible to predict the experimental data points as a function of measurements.

The purpose of this paper is to present a heuristic method that generates a periodic form for data points which are in general agreement with published experimental values. Three examples in which the "periodic" data points exhibit close agreement with published experimental data points will be shown.

II. DEVELOPMENT. The deviation of a datum point from theoretical curve (considered to be the mean) is, by definition, the arithmetic deviation. The arithmetic deviation of a measured value of a physical property of an experiment is dependent upon the experimental procedures, and if enough measurements are made, the distribution of the arithmetic deviations will follow the Gaussian curve. From the Gaussian distribution curve, it can be readily seen that the arithmetic deviation is proportional to the standard deviation. This proportionality factor is the only variable in the Gaussian distribution equation and thus is a key parameter for the comprehension of this paper. The proportionality factor, derived as a function of the probability of the Gaussian distribution equation, will also be derived from a heuristically developed periodic "pseudo-Gaussian" distribution for the generation of periodic distribution of the data points.

The Gaussian distribution equation is

$$p_m = 1/(\sigma\sqrt{2\pi})\exp[-(m-M)^2/2\sigma^2] \qquad (1)$$

where p_m is the probability of the measured value, m is the measured value, M is the mean and σ is the standard deviation.

Let $$m = M \pm k\sigma \qquad (2)$$

where $k\sigma$ is the arithmetic deviation and k, the multiplier of σ, is defined as the arithmetic deviation coefficient. By substitution,

$$p_m = 1/(\sigma\sqrt{2\pi})\exp(-k^2/2). \qquad (3)$$

Now p_m is normalized so that the area under the curve is unity and thus $p_m = 0.399/\sigma$ when $m = M$, the peak of the Gaussian. However, by letting $P_c = p_m(\sigma\sqrt{2\pi})$, a comparative probability is obtained such that $P_c = 1$ when $k = 0$, i.e., $m = M$.

Now $$P_c = \exp(-k^2/2) \qquad (4)$$

and $$k = \pm\sqrt{-2 \ln P_c} \ . \qquad (5)$$

Eq. (5) can be rewritten as

$$k = \pm\sqrt{2I} \qquad (6)$$

where $I = -\ln P_c$ and is a form of the self-information equation (11) of the measurement which is associated with the arithmetic deviation coefficient. Thus, if the comparative probability associated with a measurement can be determined, the arithmetic deviation coefficient, k, of that measurement can be calculated from Eq. (5). Since σ is a constant, k is proportional to the arithmetic deviation.

Since the portrayal of the experimentally measured data indicates that the measured values are a periodic function of the measurement and since the comparative equation developed from the Gaussian equation is not periodic, a heuristic periodic expression which closely matches the Gaussian distribution (to about $\pm 2.5\sigma$) was developed for the probability of the measured value as a function of measurement.

The empirical equation is

$$P_N = \cos^2\theta_N \tag{7}$$

where P_N is the probability associated with measurement N, $N = 1,2,3, \ldots$ and

$$\theta_N = 30\ K_N + 8\sqrt{|K_N|} \tag{8}$$

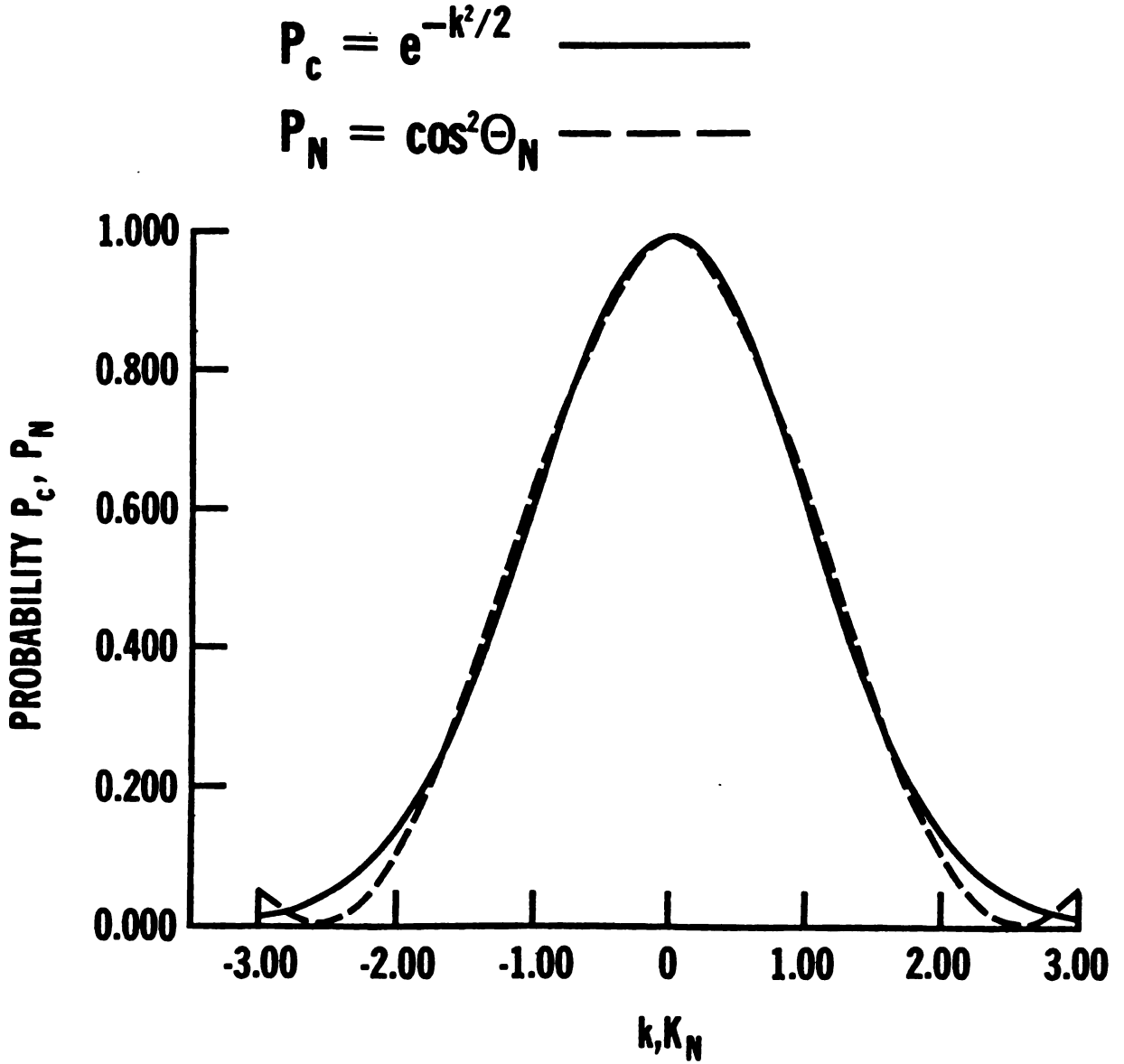
where $-\infty < K_N < +\infty$ and is dependent upon N. This dependence will be discussed later.

Comparisons of the probability values of Eqns. (4) and (7) are shown in Fig. 1 for K_N and k in the range $-3.00 \leq k = K_N \leq +3.00$. The comparison shows Eq. (7) is an approximation to the Gaussian to about $\pm 1.75\sigma$ and a reasonable approximation from $\pm 1.75\sigma$ to about $\pm 2.5\sigma$. When $P_N = 0$, $K_N = \pm 2.5723\ldots$, and for the comparable value of $\pm 2.5723\sigma$ nearly 99% of the measured data points would be included. An extension of Fig. 1 would show that P_c would asymptotically approach zero, whereas P_N would oscillate in a cosine squared manner periodically in about every 5.2 K_N but the period would increase with increasing (and decreasing) K_N.

The term K_N is selected such that $K_N = \alpha_1 + (N-1)(\alpha_2-\alpha_1)$ where α_1 is heuristically selected for P_1 associated with N=1, the first measured event, and α_2 is similarly selected for P_2 associated with N=2, the second measured event. Note that K_N is not necessarily an integer associated with the corresponding Nth measurement and thus $(\alpha_2-\alpha_1)$, may nor may not be an integer.

In the empirical expression for P_N (Eq. 7) the standard deviation coefficient per measurement does not appear. However, since P_c and P_N serve comparable functions, and since k can be determined from P_c, it is postulated that the arithmetic deviation coefficient per measurement, k_N, can be determined from P_N by a similar expression as Eq. 5. The expression is

$$k_N = \pm\sqrt{-2 \ln P_N}\ , \tag{9}$$



$\pm k, \pm K_N$	P_c	P_N
0.000	1.000	1.000
0.250	0.969	0.960
0.500	0.883	0.876
0.750	0.755	0.759
1.000	0.607	0.621
1.250	0.458	0.475
1.500	0.325	0.332
1.750	0.216	0.205
2.000	0.135	0.103
2.250	0.080	0.033
2.500	0.044	0.002
2.750	0.023	0.010
3.000	0.011	0.057

Fig. 1. The Gaussian comparative probability, P_c and the heuristic measurement probability, P_N plotted as functions of k and K_N, respectively.

where k_N is a function of K_N through Eqs. (9) and (8) in a transcendental fashion.

The measured value, m, is now expressed as

$$m = M + k_N\sigma_E \tag{10}$$

for each measurement N, where σ_E is a constant and defined as the experimental standard deviation. Thus if M and σ_E are known, m can be determined as a function of measuring events.

The polarity of k_N is selected as either + or - for the cusp portrayal of the data points, whereas the polarity is alternating in sign (+ and -) corresponding to the alternating periods of P_N for the sinusoidal portrayals.

Eq. (10) is also a simplified form of an equation for a specific physical phenomenon when the experimentally measured data are plotted as a function of the independent parameter. Formally, each measured value will be

$$Y = F(X) \pm k_N\sigma_E \tag{11}$$

where Y is the measured value, F(X) is the equation governing the physical phenomenon and $k_N\sigma_E$ is the arithmetic deviation. Thus assuming for any experiment the equation F(X) and the dependent constants are known, the evaluation of F(X) as a function of the independent variable would be the mean and corresponds to the theoretical curve. The arithmetic deviation, hence Y, can be made as a function of measuring events if an *a priori* determination of σ_E can be made.

III. EXAMPLES. Eq. (11) shows that experimentally measured data are periodic as a function of measurement events about some smooth curve governed by the physical equation. If F(X) is well-behaved and if σ_E can be *a priori* determined (at least heuristically if not by other means) the experimental data points should be, in principle, predictive. The predictiveness of the method was not subjected to experimental verification. The reason is obvious since there can be no *a priori* determination of σ_E and k_N for Eq. (11) for an experiment. However, the published data points of three experiments have been closely duplicated bv the described method. Three examples indicate *a posteriori* verification that experimentally measured data are periodic as well as predictive. (The calculations were accomplished with a pocket calculator, and thus the duplication of the data points were not optimized.)

The first example is a relative easy one: The sinusoidal data portrayal of Ref. 1. The equation of the line was determined to be

$$F(X) = i(t) = 9.40 - 0.49t \tag{12}$$

where i(t) is the relative transient current and t is in units of equal increments of time corresponding to the time of 1.37 nsec (average) between each measured datum point. The first point was at 2.74 nsec (or the 3rd experimental point) which selected for the N=1. Note that the logarithmic representation of the ordinate was portrayed in linear units. The terms α_1, α_2 and σ_E for this example were heuristically chosen which generated the points that were in agreement with the experimental points. The terms and generated points are listed in Table I and the curve is shown in Fig. 2. The generated data points are nearly in exact agreement.

Table I. Constants and "periodic" data points generated as a function of measurement for the sinusoidal example for Ref. 1.

$$F(X) = i(t) = 9.40 - 0.49t$$

$$\alpha_1 = 0.00, \quad \alpha_2 = 1.00, \sigma_E, = 0.05$$

N	K_N	P_N	k_N	$k_N\sigma_E$	i(t)	$i(t)+k_N\sigma_E$
1	0	1.000	0.00	0.00	8.06	8.06
2	1	0.607	-0.98	-0.05	7.39	7.34
3	2	0.103	-2.15	-0.11	6.71	6.60
4	3	0.057	-2.37	-0.12	6.04	5.82
5	4	0.517	-1.15	-0.06	5.37	5.33
6	5	0.956	-0.30	-0.02	4.70	4.68
7	6	0.888	+0.49	+0.03	4.03	4.06
8	7	0.393	+1.37	+0.07	3.36	3.43
9	8	0.016	+2.87	+0.14	2.69	2.83
10	9	0.165	+1.90	+0.10	2.02	2.12

The second example is to duplicate Ref. 6 for the cusp periodicity. The equation of the line is $F(X) = R(t) = 98.0$ where R(t) is the counting rate and is a constant value as a function of time. The measurements were made in units of equal increments of time of 200 seconds (average) between each datum point. The first point was at 600 seconds (or the fourth experimental point) which was selected for N_1. Again α_1, α_2 and σ_E were heuristically chosen. The constants and the calculated values are listed in Table II and the "predictive" points are shown in Fig. 3(a) by the solid dots. The open circles are the relative positions of the experimental data points as estimated from Ref. 6 with which the predictive points are not in agreement. The first twelve predictive points are nearly in exact agreement with the experimental points numbers 4 through 15. However, the 13th predictive point must be moved to the 18th experimental point for the cusp periodicity to agree with increasing measurements. With the exceptions of experimental points 16, 17 and 30, the other points are either in nearly exact or close agreement.

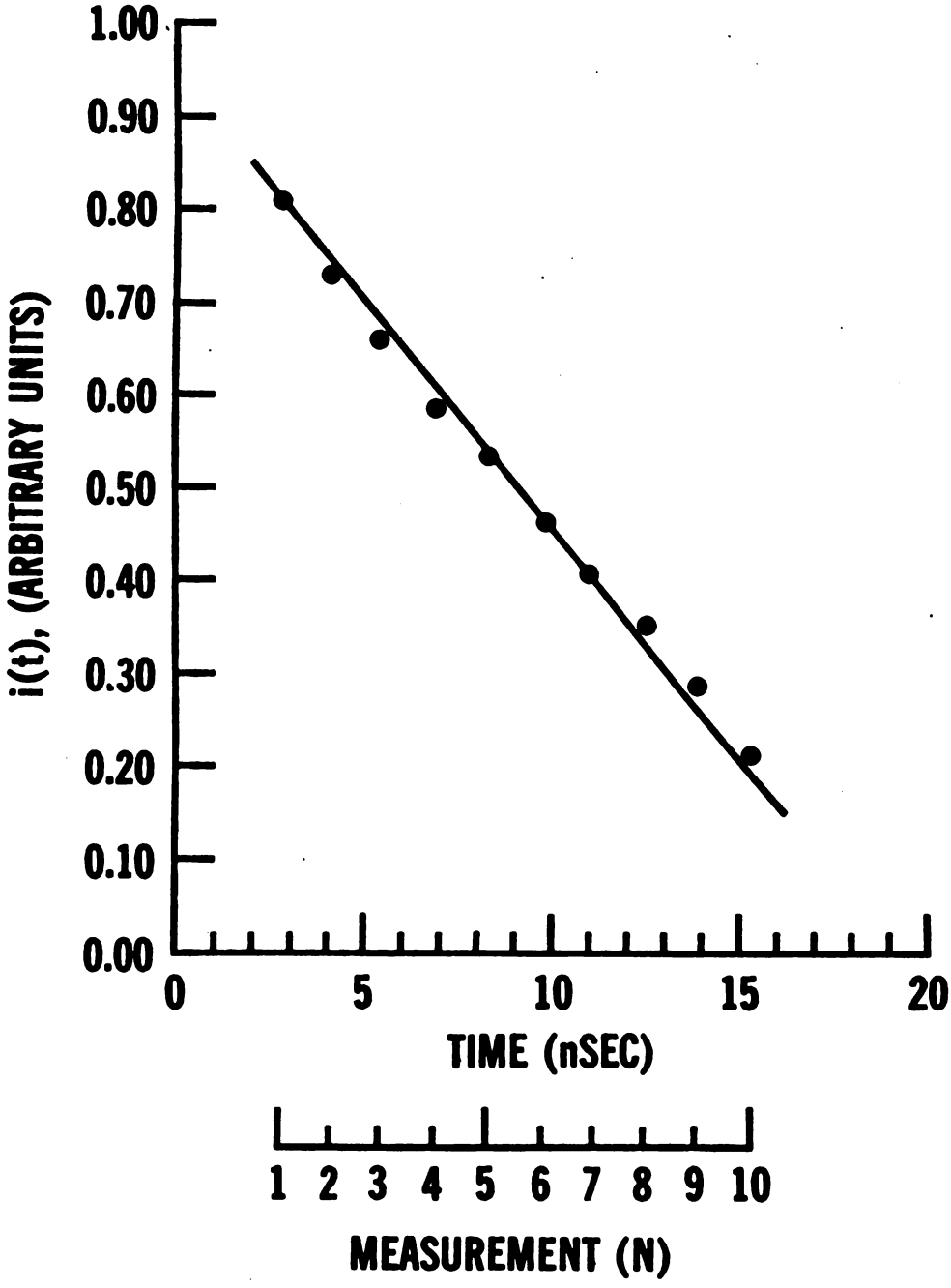


Fig. 2. Duplication of the sinusoidal portrayal of experimental points.

Table II. Constants and "periodic" data points generated as a function of measurement for the cusp example for Ref. 6.

$$F(X) = R(t) = 98.0$$

$$\alpha_1 = 1.00,\ \alpha_2 = 3.00,\ \sigma_E = 0.8$$

N	K_N	P_N	k_N	$k_N\sigma_E$	$R(t)+k_N\sigma_E$
1	1	0.607	-0.98	-0.78	97.22
2	3	0.067	-2.39	-1.91	96.09
3	5	0.956	-0.30	-0.24	97.76
4	7	0.393	-1.37	-1.10	96.90
5	9	0.165	-1.90	-1.52	96.48
6	11	0.996	-0.09	-0.07	97.93
7	13	0.268	-1.62	-1.30	96.70
8	15	0.265	-1.63	-1.30	96.70
9	17	0.997	-0.07	-0.05	97.95
10	19	0.180	-1.85	-1.48	96.52
11	21	0.357	-1.44	-1.15	96.85
12	23	0.979	-0.21	-0.17	97.83
13	25	0.117	-2.07	-1.66	96.34
14	27	0.440	-1.28	-1.02	96.98
15	29	0.949	-0132	-0.26	97.74
16	31	0.071	-2.30	-1.84	96.16
17	33	0.517	-1.15	-0.92	97.08
18	35	0.911	-0.43	-0.34	97.66
19	37	0.039	-2.55	-2.04	95.96
20	39	0.586	-1.03	-0.82	97.18
21	41	0.869	-0.53	-0.42	97.58
22	43	0.017	-2.85	-2.28	95.72
23	45	0.640	-0.93	-0.74	97.46
24	47	0.824	-0.62	-0.50	97.50
25	49	0.005	-3.26	-2.61	95.39
26	51	0.706	-0.84	-0.67	97.33
27	53	0.776	-0.71	-0.57	97.43
28	55	0.001	-4.22	-3.38	94.62
29	57	0.756	-0.75	-0.60	97.40

The third example is to duplicate Ref. 10 for a curve of experimental points that are less regular than Refs. 1-4 in the periodic portrayal. The smooth curve of Ref. 10 could not be fitted to an exponential or power equation. However, for F(X) the point by point estimated values for each equal increment of t = 200 sec are listed in Table III. The first point was at 200 seconds (the second experimental point) which was selected for N_1. The parameters α_1,α_2 and σ_E were heuristically selected. The constants and calculated values are listed in Table III, and the points are shown in Fig. 3(b) by the solid dots. The polarity of the 17th through 29th data points were reversed from the expected polarity to exhibit agreement with the reconstructed experimental values. The open circles are the experimental values as estimated from Ref. 10 with which the predictive points are not in agreement. The positions of the periodicities of the two sets of data are in close agreement with the exceptions at the 8th, 20th and 25th points. The values of most of the data points, 21 of the 29, are nearly identical, and with the exception of predictive point No. 8, the other seven are in reasonably close agreement.

IV. DISCUSSION. The subject of this paper encounters a paradox in the distribution of the measured values of a physical parameter of an experiment. The binomial distribution is a mathematical treatment of random events whose measured values are discrete, whereas the Gaussian distribution is a mathematical treatment of random events whose measured values are not discrete and cannot be exactly duplicated. The measured values of a measurement are considered to be independent of the sequence of the measurements and the values randomly distributed about some mean. The independence and randomness of the measured values from event to event "vary in an irregular manner that defies all attempts at prediction" (12). However, since it was pointed out in this paper that most plots of data points are portrayed in some periodic manner, the data points, in principle, should be predictive. The predictiveness was demonstrated in this paper, albeit ex post facto. The data points for three experiments were closely duplicated after heuristically determining three constants. It must be stated that the mean (the equation of the experiment) was known. Thus, in this context, the measured data points are predictive, hence the paradox.

The Gaussian distribution is one (of two primary concepts) equation used in this paper. The treatment of the equation to obtain the comparative probability eliminated the dependence of the probability from all parameters except one: The arithmetic deviation coefficient. Solving for the coefficient, it was found that the coefficient was a function of the probability which was identified as a form of the self-information equation. The significance of the relationship between the coefficient and the information equation cannot be developed at this time, but that the Gaussian distribution equation is a form of the information equation can be recognized.

The second concept is the heuristic equation relating the probability of a measured value as a function of sequential measurement. If a well ordered experiment is performed such that upon equal increments of an independent parameter, the measured value is recorded in the identical manner, including

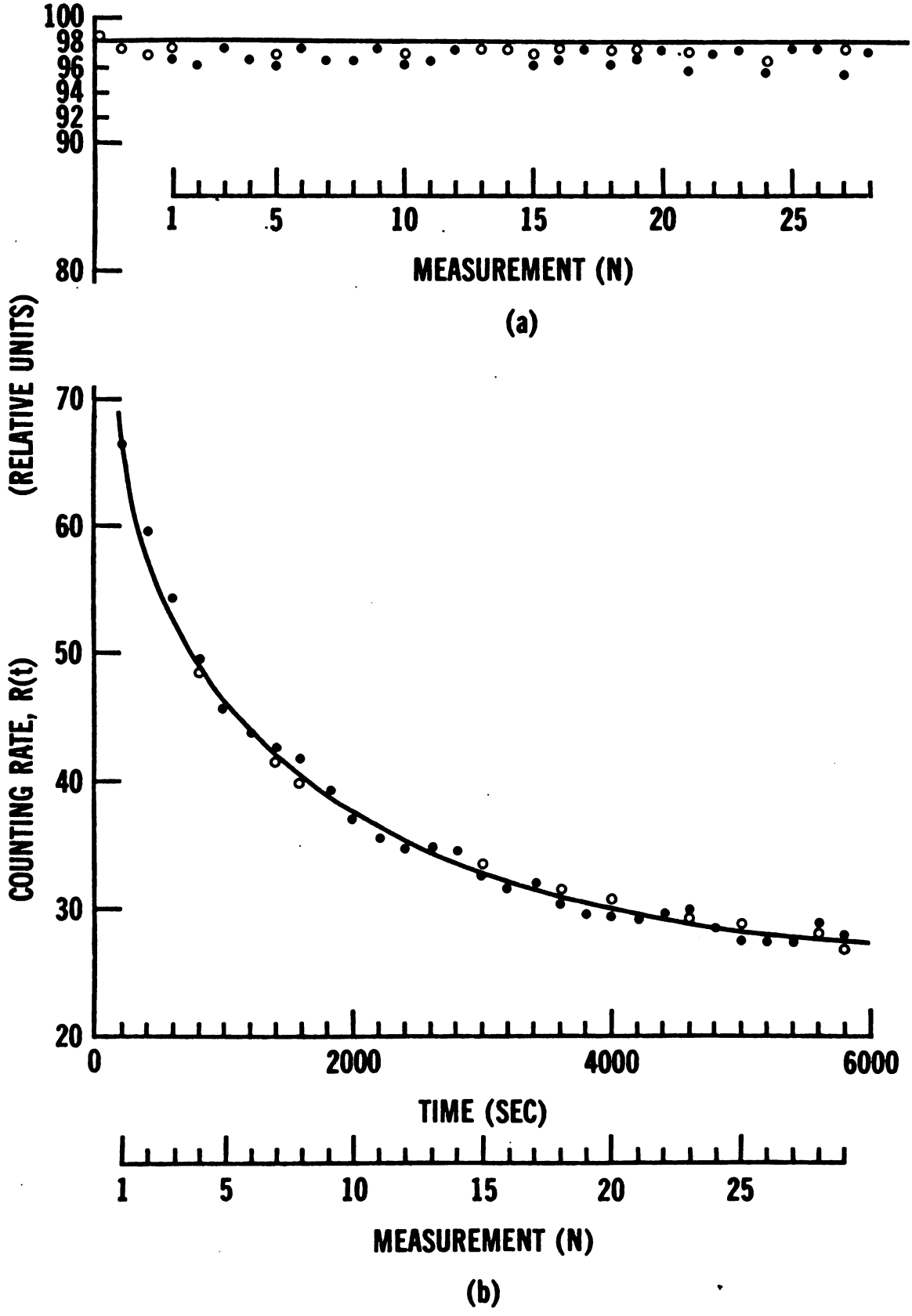


Fig. 3. Duplications of (a) the cusp portrayal and (b) the cusp-sinusoidal portrayal of experimental data points.

Table III. Constants and "periodic" data points generated as a function of measurement for the cusp-sinusoidal example for Ref. 9.

$$F(X) = R(t)$$

$$\alpha_1 = 0.00, \quad \alpha_2 = 2.00, \sigma_E = 0.3$$

N	K_N	P_N	k_N	$k_N\sigma_E$	R(t)	$R(t)+k_N\sigma_E$
1	0	1.000	0.000	0.00	66.2	66.2
2	2	0.103	+2.13	+0.64	59.0	59.6
3	4	0.517	+1.15	+0.35	53.7	54.1
4	6	0.888	-0.49	-0.15	49.3	49.1
5	8	0.016	-2.87	-0.86	46.6	45.7
6	10	0.676	-0.89	-0.27	44.2	43.9
7	12	0.783	+0.70	+0.21	42.0	42.2
8	14	0.000001	+5.20	+1.56	40.3	41.9
9	16	0.655	+0.92	+0.28	38.7	39.0
10	18	0.688	-0.86	-0.26	37.4	37.1
11	20	0.010	-3.03	-0.91	36.2	35.3
12	22	0.854	-0.56	-0.17	35.1	34.9
13	24	0.601	+1.01	+0.30	34.2	34.5
14	26	0.035	+2.59	+0.78	33.4	34.2
15	28	0.908	+0.44	+0.13	32.6	32.7
16	30	0.521	-1.14	-0.34	31.9	31.6
17	32	0.069	+2.31	+0.69	31.3	32.0
18	34	0.947	+0.33	+0.10	30.7	30.8
19	36	0.448	-1.27	-0.50	30.2	29.7
20	38	0.109	-2.10	-0.63	29.9	29.3
21	40	0.973	-0.23	-0.07	29.4	29.3
22	42	0.382	+1.39	+0.42	29.0	29.4
23	44	0.153	+1.94	+0.58	28.7	29.3
24	46	0.990	+0.14	+0.04	28.4	28.4
25	48	0.322	-1.51	-0.45	28.2	27.7
26	50	0.200	-1.79	-0.54	28.0	27.5
27	52	0.998	+0.06	+0.02	27.8	27.8
28	54	0.269	+1.62	+0.49	27.6	28.1
29	56	0.248	+1.67	+0.50	27.4	27.9

time lapses, the data points appear to be portrayed in a periodic fashion in about either every 3-4 or every 10-12 measurements. The probability of the measured value is a function of the measurements for $-3.00 \leq k=K_N \leq +3.00$ and follows a cosine-squared law that nearly duplicates the Gaussian distribution probability. The probability dependence on measurements cannot be derived from first principles but is heuristically justified by the fact it enables the "predictive" values to agree with the experimental data.

The duplication of the experimental data points that are portrayed in a sinusoidal fashion is in itself not too profound. In fact, for Refs. 1-4, the points can be readily duplicated by nearly any periodic mathematical treatment. However, the heuristic method, when applied to the duplication of the points that are portrayed in cusps (as in Refs. 6-9) and combination of cusps and sinusoidal periodicities (as in Ref. 10), the method must be considered as being significant. The significance is emphasized when form a cursory examination some of the data points appear to be random but are duplicated by the method. Obviously, all of the published data points cannot be duplicated by the limited treatment of the heuristic method described in this paper.

It is postulated that the sinusoidal portrayal of data points in which the measured parameter does not have a fixed bound or physical barrier. Examples of the sinusoidal characteristic would be the measure of electron trapping as a function of time (Ref. 1), and the determination of the beam displacement of a reflected electromagnetic wave as a function of distance (Ref. 3).

It is postulated that the cusp-type portrayal of data points is characteristic of an experiment in which the measured parameter (the dependent variable) has a fixed bound or physical barrier. Examples of the cusp characteristic would be the "no influence" effect of source strength as a function of time (Ref. 6), and the measurement of the length of a meter stick by the eclipsing of the position of a light source by the end of the stick.

There are two baffling considerations that were required for obtaining agreement of the points in two of the examples. The first is that two experimental points had to be by-passed in Ref 7, and the second is that a change of polarity for the arithmetic deviation coefficient for Ref. 10. These considerations may be resolved if the constants of the heuristic equation were evaluated simultaneously with the experimental data points of an experiment performed in a human factors evaluation laboratory. In such a laboratory, the constants may be determined from the nature of the experiment and the experimental procedures, and thus in conjunction with the equation governing the experiment, the measured data points may be indeed predictive.

V. ACKNOWLEDGEMENT. The author wishes to acknowledge Mr. Raymond J. Stefanik for valuable discussions and comments.

VI, REFERENCES.

1. J. W. Mayer, K. Zanio, and M. Martini, "CdTe and Trapping Analysis," Proc. Int. Symp. on Cadmium Telluride, P. Siefert and A. Cornet, eds. (Strausbourg, Fr. 1972) See Fig. 6.

2. W. Elser, J. L. W. Pohlmann, P. R. Boyd, Mol. Cryst. Liq. Cryst. 20, 77-86, (1973). See Fig. 2.

3. L. A. A. Read, M. Wong, and G. E. Reesor, J. Opt. Soc. Am. 68 (No. 3) 319-322 (1978). See Figs. 3 and 4.

4. T. Witkowicz, IEEE Trans. Electron Devices, 25, 269-271 (1978). See Fig. 2, curve 1.

5. J. J. Hill, private communication, Rockhurst College. The cusps are also known as "hillocks."

6. P. Sieffert, J. Berger, C. Scharager, A. Cornet, R. Stuck, R. O. Bell, H. B. Serreze, and F. W. Wald, IEEE Trans. Nucl. Sci. 23, 159-170 (1976). See Fig. 7, top curve.

7. C. H. Graham, "Discrimination That Depends On Wavelength," Vision and Visual Perception, C. H. Graham, ed. (John Wiley and Sons, Inc., New York, 2nd Printing, 964), See Fig. 12.13.

8. T. Witkowicz, Loc. Cit, See Fig. 2, curve 2.

9. C. W. White, et al, J. Appl. Phys. 50, 3261 (1979). #19.4 Annealed Curve

10. P. Sieffert, P., et al, Loc. Cit, See Fig. 7, bottom curve.

11. R. C. Gallager, Information Theory and Reliable Communication, (John Wiley and Sons, Inc., New York, 1968). Secs. 1.2 and 2.2.

12. C. Eisenhart, and M. Zelen, "Elements of Probability," in Handbook of Physics, E. U. Condon and H. Odishaw, eds. (McGraw-Hill Book Company, New York, 2nd Edition, 1965) p 1-163, chap. 12.

VARIABLE TRANSFORMATION IN NONLINEAR LEAST SQUARES MODEL FITTING

Aivars Celmiņš
Interior Ballistics Division
U.S. Army Ballistic Research Laboratory
Aberdeen Proving Ground, Maryland

ABSTRACT. The numerical treatment of nonlinear model fitting problems often can be simplified by manipulating the model equations. Algebraic manipulations, including nonlinear transformations of model parameters, do not change the numerical result of the adjustment. Therefore, such manipulations can be a powerful method to improve the performance of solution algorithms. Nonlinear transformations of the observations, on the other hand, do change the numerical results unless the normal equations are transformed accordingly. The latter transformation has been neglected by previous authors and this article provides a complete set of formulas that are needed to implement transformations of observations. The transformations are, however, in general less useful than parameter transformations but may have applications in particular situations.

1. INTRODUCTION. A mathematical model fitting problem arises when one compares real observations with theoreical predictions. The observations always contain observational inaccuracies and, likewise, the theory of the prediction can be inadequate. If discrepancies between observations and predictions are unacceptably large for a particular situation then one is faced with the task to adjust in a rational manner either the observations, or the theory, or both so that an acceptable mathematical description of the event can be established. The problem can be subdivided conveniently into three subtasks, each of which requires a different approach and background information.

First, one has to chose a model. Normally, this requires supporting information from engineering, physics, geometry, etc., which may suggest or postulate a reasonable mathematical description of the observable event. We shall assume in this article that the model is formulated as a system of equations containing observations and, possibly, also some undetermined model parameters.

Once the model is selected, one can compare predicted values of observable quantities with corresponding observations. The comparison provides the basis for a rational adjustment of the observations and/or of the model. This subtask of the problem is a purely mathematical part of model fitting and it belongs to the category of ill-posed problems. Its mathematical/numerical treatment is independent of the other two subtasks, i.e., of applications. We shall be concerned with this part of the problem in the present article.

After the adjustments have been carried out, one has to validate the mathematical model, unless it has been prescribed, e.g., by the geometry of the event. The validation involves typically, but not necessarily, a statistical analysis of the discrepancies between observations and predictions. The result of the validation process may be a new formulation of model equations and subsequent fitting, i.e., a repetition of the whole task until some validation criterion is satisified. We shall not discuss this part of the problem, noticing only that the results of the second subtask provide the data basis necessary for a validation.

If the model equations are not linear then the model fitting problem generally leads to systems of complicated simultaneous equations and corresponding numerical difficulties may arise. Often the numerical treatment can be simplified by a reformulation of the model equations, particularly by introduction of new variables through variable transformations. Such manipulations have been suggested in textbooks[1-7] and are routinely used in applications. Examples of recently published applications where variable transformations have been used are references 8, 9, and 10.

A closer investigation of variable transformations in model fitting problems suggests that the formulations should be used more cautiously than some of the texts suggest. Therefore, we shall present in this article an investigation of some consequences of the transformations and draw conclusions about their usefulness for the simplification of the numerical treatment of model fitting problems.

In Section 2 we shall formulate the mathematical model fitting problem in general terms and discuss the effects that can be anticipated from manipulations of model equations. In Section 3 we shall specialize the considerations to nonlinear least squares problems and produce explicit formulas that are needed in such problems. Some examples will be presented in Section 4, and Section 5 will summarize the conclusions that can be drawn from the theoretical discussions and from examples.

2. GENERAL ASPECTS OF MATHEMATICAL MODEL FITTING. Let the model equations be

$$A(X)\theta = 0, \tag{2.1}$$

where $X \varepsilon R^n$ is the vector of all observations, $\theta \varepsilon R^p$ is a model parameter vector, and $A(X)$ is an operator that operates on θ and has a range R^r. We assume that the following relations hold between the dimensions n, r, and p:

$$n \geq r > p \geq 0. \tag{2.2}$$

By permitting the dimension p to be zero, we include in our considerations also cases in which the model equations do not contain free parameters. Then equation (2.1) reduces to $A(X)=0$.

Typical for applications are cases in which the r equation (2.1) for θ are independent and, because of (2.2), do not have a solution. Then one replaces the model equations by another system

$$\tilde{A}(X)\theta = 0, \tag{2.3}$$

chosing the operator $\tilde{A}(X)$ such that it approximates $A(X)$ and has a solution. The determination of $\tilde{A}(X)$ can be considered as the central part of the model fitting problem.

In order to have a measure for the approximation we introduce a metric for the operators. Let $\rho[\tilde{A}(X), A(X)]$ be a metric. Then one can formulate the mathematical model fitting as the following constrained minimization problem:

$$\tilde{A}(X)\theta = 0 \ , \quad W\{\rho[\tilde{A}(X), A(X)] = \min., \tag{2.4}$$

where $W\{\rho\}$ is generally a convex object function. The choice of the metric ρ and of the object function $W\{\rho\}$ determines the type of the model fitting, e.g., least squares, maximum norm, etc.

We shall now discuss the selection of an approximate operator $\tilde{A}(X)$. First, we notice that the model operators $A(X)$ and $\tilde{A}(X)$ are generally needed and defined only within a finite neighborhood of the observations X. Therefore, assumptions about properties of the operators need to be made for that neighborhood only. Let the neighborhood consist of all points $Z = X+C$, whereby C is restricted component-wise by

$$\gamma_i \leq C_i \leq \Gamma_i \ , \ i = 1, 2, \ldots, n. \tag{2.5}$$

The intervals (γ_i, Γ_i) normally contain zero, but exceptions are possible and do occur in applications. Second, we assume that within the neighborhood (2.5) $A(Z)$ is a continuous function of Z. Then a reasonable choice of $\tilde{A}(X)$ is

$$\tilde{A}(X) = A(X+C). \tag{2.6}$$

The choice achieves a natural parametrization of the approximation. The approximation parameter is the vector $C \in R^n$ and the operator $\tilde{A}(X)$ depends continuously on the parameter within the restrictions (2.5).

The parametrized model fitting problem can be formulated as follows:

$$\left.\begin{aligned} &A(X+C)\theta = 0, \\ &W\{\rho[A(X+C), A(X)]\} = \min. \end{aligned}\right\} \qquad (2.7)$$

The quantities to be determined by equation (2.7) are the approximation parameter C and the model parameter θ. We assume that the solution vector C is within the limits specified by equation (2.5).

We will need in the sequel some differentiability properties for the model operator. As far as X is concerned, we assume the properties to hold within the neighborhood (2.5). With respect to θ we assume that a similar neighborhood exists in the vicinity of the solution of equation (2.7) in which $A(X)\theta$ is a continuous function of θ. The differentiability assumptions are that $A(X+C)\theta$ is twice differentiable with respect to all its n+p arguments within the cartesian product space of the neighborhoods of X and θ. We also assume that within that space

$$\text{rank } \frac{\partial A}{\partial X} = r, \qquad (2.8)$$

and define

$$\rho[A(Z), A(X)] = ||Z - X||. \qquad (2.9)$$

ρ is a metric within the neighborhood in which (2.8) holds. We also assume that the model equations do not contain redundant parameters. The assumption may be expressed as the requirement

$$\text{rank } \frac{\partial A(X)\theta}{\partial \theta} = p. \qquad (2.10)$$

With the specialization (2.9), the model fitting problem becomes

$$\left.\begin{aligned} &A(X+C)\theta = 0, \\ &W\{\rho[A(X+C), A(X)]\} = W\{||C||\} = \min. \end{aligned}\right\} \qquad (2.11)$$

Equation (2.11) is an abstract formulation of common model fitting problems. The difference C between the observations X and the "corrected observations" X+C is called the residual vector. In the formulation (2.11) we require that a norm of the residual vector be minimized, subject to model equations which have to be satisfied at X+C. The model parameter vector θ is not essential in this formulation. The number of model parameters may be zero and it is normally orders of magnitudes smaller than the number of approximation parameters, i.e., residuals. The determination of θ can be, of course, in some applica-

tions more important than the determination of C, but this is not always the case.

A least squares model fitting problem is a special case of (2.11), characterized by a particular choice of the norm in the definition (2.5), and of the object function $W\{\rho\}$. The least squares metric is

$$\rho[A(Z), A(X)] = ||Z-X|| = [(Z-X)^T R^{-1}(Z-X)]^{\frac{1}{2}} , \qquad (2.12)$$

where R is an estimate of the variance-covariance matrix of the observations. The least squares object function is

$$W\{\rho\} = \rho^2 . \qquad (2.13)$$

Therefore, the least squares model fitting problem is defined by

$$\left.\begin{aligned} &A(X+c)t = 0, \\ &W = ||c||^2 = c^T R^{-1} c = \min. \end{aligned}\right\} \qquad (2.14)$$

In equation (2.14) we have used c and t instead of C and θ, respectively, thus indicating the least squares values of both parameter vectors.

The use of R^{-1} as a norm matrix in the definition (2.11) makes the norm $||C||$ and W dimensionless, which is very convenient when fitting results are compared. If the variance-covariance matrix R is known exactly, then the solution of equation (2.14) is a maximum likelihood solution of the approximation problem[11]. The same maximum likelihood solution is obtained if R approximates the variance-covariance matrix up to an unknown factor. In applications one has to be content with an estimate of R. Then often the off-diagonal elements are assumed to be zero as a matter-of-course. Because the results of the model fitting depend on R, such assumptions should not be made without having reasons that zero is a better approximation than a non-zero value. The theoretical treatment is not complicated by the assumption that R is not diagonal, nor are the numerical complications unsurmountable. Realistic estimates of R are, however, important for the interpretation of the results, and for the validation of the fitting.

We solve the optimization problem (2.11) or (2.14) using Lagrange multiplier technique, and call the multipliers correlates, as usual in adjustment problems. Let $K \varepsilon R^r$ be a correlate vector and let the modified object function be

$$\tilde{W} = \frac{1}{2} W\{||C||\} - K^T A(X+C)\theta . \qquad (2.15)$$

Necessary conditions for the solution of the optimization problem are obtained by setting zero the partial derivatives of $\tilde{W}$ with respect to the unknown C, θ, and K. This yields the following set of normal equations.

$$\frac{1}{2}\frac{\partial}{\partial C}\ W\{||C||\} - \frac{\partial}{\partial C}\ [K^T A(X+C)\theta] = 0\ , \qquad (2.16a)$$

$$\frac{\partial}{\partial \theta}\ [K^T A(X+C)\theta] = 0, \qquad (2.16b)$$

$$A(X+C)\theta = 0. \qquad (2.16c)$$

The solution of the model fitting problem (2.11) is among the solutions of equations (2.16). On the other hand, one cannot guarantee that a particular solution of the normal equations corresponds to the absolute minimum solution of equation (2.11), nor is the uniqueness of the solution given. An investigation of these complications is not the subject of this paper. Mostly, such problems can be, and are taken care of by ad hoc measures based on background information from the application. Therefore, we simplify our present theoretical discussion by assuming in this section that a numerical solution of equations (2.16) can be obtained, and that it has been verified as the absolute minimum solution of equation (2.11).

In least squares problems, the first term $\partial W/\partial C$ in equation (2.16a) is linear with respect to C. Nonlinear expressions which could be possibly simplified by algebraic manipulations may occur in the second term in equation (2.16a), and in equations (2.16b) and (2.16c). The structure of these terms strongly depend on the form in which the model equations (2.16c) are cast, and it is obvious that simplifications can be achieved by proper formulations. Particularly, one does not have to insist that each model equation be solved for a "dependent" observation. Such a form is assumed in most textbooks on data reduction and postulated in computer programs for data reduction problems. Quite often an implicit formulation of the equations (2.16c) can be simpler, producing also simpler expressions for the derivatives in equations (2.16a) and (2.16b). The solution of the problem (2.11) is, of course, independent of the particular form in which the model equations are cast. This remark is trivial in the present context, and it is a consequence of the formulation of the model fitting problem by equation (2.11). Reference 12 reports about numerous unsuccessful attempts to achieve a similar invariance statement when the problem was formulated differently.

The aforementioned manipulations of the model operator $A(X)\theta$ can also include nonlinear transformations of the parameter θ. Such transformations do not affect the definition of the metric ρ, because the metric of the operator is independent of the operand. Therefore,

the transformations do not affect the first term in equation (2.16a) either and are a powerful tool for the simplification of the rest of the equations. An example in which nonlinear parameter transformations are used to linearize the model equations is reported in reference 9. In Section 4 we shall give other examples.

The formal procedure of replacing parameters is as follows: Suppose that one wants to replace the parameter θ by σ whereby both parameters are related by a nonsingular function

$$\theta = w(\sigma). \tag{2.17}$$

(Regularity of the transformation need to be assumed only within a neighborhood of the solution.) Let the model equations be in terms of σ

$$\bar{A}(X)\sigma = 0. \tag{2.18}$$

The operator $\bar{A}$ can be obtained from A always by the definition

$$\bar{A}(X)\sigma \equiv A(X)w(\sigma), \tag{2.19}$$

however, often one can find other equivalent formulations that are simpler. The metric $\bar{\rho}$ associated with $\bar{A}$ is defined as in equation (2.9)

$$\bar{\rho}\,[\bar{A}(Z), \bar{A}(X)] = ||Z-X||. \tag{2.20}$$

With this definition and the same object function $W\{\rho\}$ as before one obtains the normal equations

$$\frac{1}{2}\frac{\partial}{\partial \bar{C}} W\{||\bar{C}||\} - \frac{\partial}{\partial \bar{C}}[\bar{K}^T\bar{A}(X+\bar{C})\sigma] = 0, \tag{2.21a}$$

$$\frac{\partial}{\partial \sigma}[\bar{K}^T\bar{A}(X+\bar{C})\sigma] = 0, \tag{2.21b}$$

$$\bar{A}(X+\bar{C})\sigma = 0. \tag{2.21c}$$

The solution vectors of equations (2.16) and equations (2.21) are related by

$$C = \bar{C}\,, \quad \theta = w(\sigma) \tag{2.22}$$

The vectors K and $\bar{K}$ can be computed from these values using formulas given in the next section.

The relation (2.22) is again a simple consequence of the formulation (2.11) of the model fitting problem. Bender[9] proves the correspondence (2.22) for a particular transformation and application, and indicates that previous developers of software for such problems were not aware of the relation.

If the solution of the model fitting task has been found from equation (2.21) in terms of σ, but the parameter vector θ is of interest, then one needs in addition to equation (2.22) another formula for the accuracy of θ. Let us assume that the solution algorithm for equation (2.21) has also provided information about the accuracy of σ in form of an estimate V_σ of the variance-covariance matrix of the components of σ. (In Section 3 we shall give formulas for V_σ in least squares problems.) Then an estimate of the variance-covariance matrix V_θ of the components of θ can be obtained by applying the linearized law of variance propagation to the relation (2.22). The result is

$$V_\theta = \frac{\partial w}{\partial \sigma} V_\sigma \left(\frac{\partial w}{\partial \sigma}\right)^T . \qquad (2.23)$$

More complicated are consequences of such manipulations of the model equations that involve transformations of the observations. This is so because the transformations now affect the definition of the norm ρ. Next, we shall consider such transformations.

Let a transformation of observations be

$$Y = v(X) \qquad (2.24)$$

with the inverse

$$X = u(Y) .$$

We assume that the transformation is regular within the neighborhood (2.5), including the solution X+C, and that the function u(Y) is there twice differentiable. The model equations (2.1) are replaced by equivalent (usually simpler) equations

$$\hat{A}(Y)\theta = 0. \qquad (2.25)$$

The operator $\hat{A}(Y)\theta$ can be obtained, e.g., by the definition

$$\hat{A}(Y)\theta \equiv A(u(Y))\theta, \tag{2.26}$$

but, as in the case of parameter transformations, usually other equivalent formulations can be found that are simpler.

When we formulate the model fitting problem in terms of Y, we have to keep in mind that the goal is to minimize the distance C between the actual observations X and their corrected values X+C. In least squares problems, only such a minimization yields under conditions a maximum likelihood solution. Then the minimization problem (2.11) is

$$\left.\begin{aligned} &Y = v(X), \\ &\hat{A}(Y+B)\theta = 0, \\ &W\{||u(Y+B) - X||\} = \min. \end{aligned}\right\} \tag{2.27}$$

The normal equations for the problem (2.27) are

$$\frac{1}{2}\frac{\partial}{\partial B} W\{||u(Y+B)-X||\} - \frac{\partial}{\partial B}\,[K^T\hat{A}(Y+B)\theta] = 0, \tag{2.28a}$$

$$\frac{\partial}{\partial \theta}\,[K^T\hat{A}(Y+B)\theta] = 0, \tag{2.28b}$$

$$\hat{A}(Y+B)\theta = 0. \tag{2.28c}$$

The first term in equation (2.28a) is not linear with respect to the unknown B unless the transformation (2.24) is linear. Therefore, a nonlinear transformation that produces an operator $\hat{A}(Y)\theta$ which is simpler than the original operator $A(X)\theta$, introduces nonlinear terms in equation (2.28a). The new nonlinearities may offset the advantages gained by a simplification of the other terms in the equations.

We shall pursue this point further in the next section and show in detail how the normal equations and algorithms are affected by transformations of observations specifically in least squares problems.

3. LEAST SQUARES MODEL FITTING. We consider in this section the effects of variable transformations on least squares model fitting problems. We shall first derive the basic equations for nonlinear least squares problems in terms of the original observations, and then show how the equations are affected by a transformation of the

observations. We simplify our notation by defining a vector function $F(X,\theta)$ by

$$F(X,\theta) \equiv A(X)\theta. \tag{3.1}$$

Then the model equation (2.1) is

$$F(X,\theta) = 0, \tag{3.2}$$

and the least squares model fitting problem (2.14) is

$$\left.\begin{aligned} &F(X+c,t) = 0, \\ &||c||^2 = c^T R^{-1} c = \min. \end{aligned}\right\} \tag{3.3}$$

In the sequel we will use subscripts to denote derivatives. Also, because derivatives of $F(X+c,t)$ with respect to c are identical to derivatives with respect to X we shall use the subscript X for both. Thus, e.g.,

$$F_X(X+c,t) = \frac{\partial}{\partial X} F(X+c,t) \equiv \frac{\partial}{\partial c} F(X+c,t)$$

and

$$[K^T F(X+c,t)]_{Xt} = \frac{\partial^2}{\partial X \partial t} [K^T F(X+c,t)] \equiv \frac{\partial^2}{\partial c \partial t} [K^T F(X+c,t)]$$

are matrices with the dimensions rxn and nxp, respectively.

Using this notation, the normal equations corresponding to the problem (3.3) are

$$R^{-1}c - k^T F_X(X+c,t) = 0, \tag{3.4a}$$

$$k^T F_t(X+c,t) = 0, \tag{3.4b}$$

$$F(X+c,t) = 0. \tag{3.4c}$$

The normal equations are in general nonlinear with respect to c and t. Therefore, their numerical solution will require some kind of iteration. We obtain second order iteration equations for equations (3.4) by expanding the normal equations at an approximation to the solution and keeping the linear terms of the expansion. Let the approximation to the solution be C, K, and T, and that the corresponding corrections be ε, κ, and τ. Then the expansion yields the following Newton equations for the corrections:

$$[I-R(K^TF)_{XX}]\varepsilon - RF_X^T\cdot(K+\kappa) - R(K^TF)_{Xt}\tau = - C, \qquad (3.5a)$$

$$(K^TF)_{tX}\varepsilon + F_t^T\cdot(K+\kappa) + (K^TF)_{tt}\tau = 0, \qquad (3.5b)$$

$$F_X\varepsilon + F_t\tau = - F. \qquad (3.5c)$$

The arguments of F and its derivatives in equations (3.5) are X+C and T.

Newton-Raphson iteration equations can be established by suitable manipulations of equations (3.5)[8,13,14,15]. A set of such iteration equations are given in the Appendix. Most authors simplify equations (3.5) by neglecting all terms that contain second order derivatives[1,11,16,17]. This yields so-called Gauss-Newton procedures that have theoretically only linear convergence and that also may have other peculiarities[13].

The final step in a model fitting problem is to obtain variance estimates of the solution in terms of the estimated variances of the observations. We shall restrict ourselves in this article to the estimation of the accuracies of the least squares value t of the parameter vector, and show how the estimation formulas change due to transformations of observables. We shall use the linearized variance propagation formula for the estimates. Estimates of the accuracies of the corrected observations x=X+c can be obtained by analogous processes.

The formulas can be derived from the linear terms of an expansion of the normal equations (3.5) at the solution[13]. Let dx, dk, and dt be the differentials of the solution vectors x=X+c, k and t, respectively. Then the expansion yields

$$(I-R(k^TF)_{XX}]dx - RF_X^Tdk - R(k^TF)_{Xt}dt = dX, \qquad (3.6a)$$

$$(k^TF)_{tX}dx + F_t^Tdk + (k^TF)_{tt}dt = 0, \qquad (3.6b)$$

$$F_Xdx + F_tdt = 0. \qquad (3.6c)$$

The arguments of F and its derivatives in equations (3.6) are x and t.

By manipulations of equations (3.6) that can be done in various ways[13,18] one obtains linear relations between dt and dX, and between dx and dX, respectively. Let the former relation be

$$N\,dt = S\,dX. \tag{3.7}$$

(Explicit formulas for N and S are given in the Appendix.) Then the estimated variance-covariance matrix V_t of the parameter vector t is

$$V_t = N^{-1} S\, R\, S^T (N^{-1})^T. \tag{3.8}$$

It is obvious from the derivation of equation (3.8) that V_t which itself is only a linearized approximation depends on second order derivatives of F. (The formulas in the Appendix show explicitly this dependency.) Neglect of the second order derivative terms renders a formula that is theoretically less than first order accurate. Therefore, such a neglect has to be justified in each application by providing estimates of the magnitudes of the neglected terms. Of the cited references, only in references 13, 14, 15, and 18 complete first order formulas are used.

Next, we introduce variable transformations into the least squares model fitting problem. We can restrict ourselves to transformations of observations because, as shown in Section 2, transformations of model parameters have the same effects as simple algebraic manipulations of the model equations.

Let, as in Section 2, the transformation be given by

$$Y = v(X) \tag{3.9}$$

with the inverse

$$X = u(Y).$$

In terms of Y, the least squares model fitting problem is defined by

$$Y = v(X), \tag{3.10a}$$

$$H(Y+b,t) = 0, \tag{3.10b}$$

$$||u(Y+b)-X||^2 = [u(Y+b)-X]^T R^{-1} [u(Y+b)-X] = \min. \qquad (3.10c)$$

Equation (3.10b) is a model equation, equivalent to equation (3.2) and expressed in terms of Y.

The normal equations for the problem (3.10) are

$$(u_y(Y+B)]^T R^{-1} [u(Y+b)-X] - k^T H_y(Y+b,t) = 0 \qquad (3.11a)$$

$$k^T H_t(Y+b,t) = 0, \qquad (3.11b)$$

$$H(Y+b,t) = 0. \qquad (3.11c)$$

Corresponding Newton equations for corrections β, κ, and τ of approximate solutions B, K, and T, respectively, are

$$[I-Q\Xi]\beta - QH_y^T \cdot (K+\kappa) - Q(K^T H)_{yt}\tau = -\Delta, \qquad (3.12a)$$

$$(K^T H)_{ty}\beta + H_t^T \cdot (K+\kappa) + (K^T H)_{tt}\tau = 0, \qquad (3.12b)$$

$$H_y\beta + H_t\tau = -H, \qquad (3.12c)$$

where

$$Q = v_x R v_x^T = (u_y)^{-1} R (u_y^T)^{-1}, \qquad (3.13)$$

$$\Delta = v_x \cdot [u(Y+B)-X] = v_x \cdot C = (u_y)^{-1} \cdot C, \qquad (3.14)$$

$$\Xi = (K^T H)_{yy} - (u^T R^{-1} C)_{yy}. \qquad (3.15)$$

The arguments of the functions H and u in equations (3.12) through (3.15) are Y+B and T, and the last term in equation (3.15) is differentiated assuming C=u(Y+B)-X to be constant. The term is a symmetric nxn matrix containing second order derivatives of the transformation function u(Y).

A comparison of equations (3.12) with equations (3.5) shows that the important changes in the Newton equations due to the transformation (3.9) are in equations (3.12a). The rest of equations (3.12) is formally identical to the corresponding terms in equations (3.5), if $F(X,\theta)$ is is replaced by $H(Y,\theta)$. In equations (3.12a) we see three other replacements: the estimated variance-covariance matrix R is replaced by Q, the right hand side -C is replaced by $-\Delta$, and the term $(K^TF)_{xx}$ is replaced by Ξ.

The replacement of R by Q corresponds to an application of the linearized variance propagation formula to the transformation (3.9). The replacement of the right hand sides is a linearized transformation of the residuals C into the Y-space. If the transformation (3.9) is linear, then only these two replacements occur. If, however, the transformation is nonlinear, then the last term in equation (3.15) does not vanish and, because it contains second order derivatives of $u(Y)$, it can be quite complicated. This complication can offset algorithmic advantages gained by a simplification of other terms in the Newton equations.

Iteration algorithms and formulas for the variances of the solution again can be obtained by manipulations of the Newton equations. Explicit formulas are given in the Appendix. We notice that second order Newton-Raphson algorithms necessarily contain second order derivatives of the model function H as well as of the transformation function $u(Y)$. The coding of the second order derivatives can, of course, be avoided if first order Gauss-Newton algorithms are used. However, variance estimates of the solution can be calculated to a first order accuracy only if all the second order derivatives are available.

The author has carried out numerical experiments to determine whether a solution of equations (3.11) instead of equations (3.4) has algorithmic advantages. The experiments were done with the utility programs described in reference 15. The programs permit one to carry out the calculations either in terms of X, or in terms of Y, and to use either Newton-Raphson, or Gauss-Newton algorithms. The experiments were inconclusive. In some examples the algorithms converged better when the problem was formulated in X, in other examples a formulation in $Y=v(X)$ produced better algorithms. However, the differences in performance were never significant. This result is in strong contrast to similar experiments involving transformations of parameters. In those experiments, a suitable parameter transformation often had a dramatic effect on the performance of the solution algorithm. Some examples are given in the next section.

Another possible benefit from nonlinear transformations of observations could be a simpler problem formulation. The complexity of the normal equations is thereby of secondary importance, if one uses an available general utility program for their solution. However, the model equations must be made available to the utility program,

which means that the equations must be programmed. Then one has the choice to program either the function $F(X,\theta)$ with its first and second order derivatives, or the two functions $H(Y,\theta)$ and $u(Y)$ with their derivatives. If the transformation is nonlinear, then normally the programming of H and u will not be simpler than the programming of F. An exception may be the situation where the same transformation u(Y) (e.g., polar-cartesian) is used for several problems with different model functions $H(Y,\theta)$, so that u(Y) has to be programmed only once.

We may conclude that in general a transformation of observations offer little or no advantages over a formulation of the model equations in terms of the original observations. There are, however, other useful applications of such transformations. First, a graphical display of the results can be clearer in terms of Y then in terms of X. Second, and more importantly, the transformations can be a convenient method to derive a "falsified" problem that can be solved easily and that provides initial approximations to the unknown least squares solution vectors. One can falsify the problem, e.g., by using a nonlinear transformation but linearizing its effects on the problem formulation. A simple and effective falsification is to replace the problem (3.10) by

$$Y = v(X), \tag{3.16a}$$

$$H(Y+b,t) = 0, \tag{3.16b}$$

$$b^T[u_y^T(Y)R^{-1}u_y(Y)]b = \min. \tag{3.16c}$$

The formulation is identical to the correct formulation (3.10) only if the transformation is linear, but the normal equations for the false problem (3.16) are simple:

$$\tilde{Q}^{-1}b - K^T H_y(Y+b,t) = 0, \tag{3.17a}$$

$$K^T H_t(Y+b,t) = 0, \tag{3.17b}$$

$$H(Y+b,t) = 0, \tag{3.17c}$$

where

$$\tilde{Q} = [u_{y(Y)}]^{-1}R[u_y^T(Y)]^{-1}. \tag{3.18}$$

This system can be much simpler and easier to solve than equations (3.4) or the equivalent equations (3.11). Its solution is, however, not the least squares solution but an approximate solution of unknown quality.

Initial approximations to the solution also can be obtained by other falsifications in addition to the one described, or instead of it. Such falsifications are, e.g., assumptions that certain observations are error free, that some correlations are zero, that some model parameters have prescribed values, etc.

4. EXAMPLES. The first example is a case involving transformation between polar and cartesian coordinates. We shall compare results that are obtained using the approach of the previous section with results that are obtained by following suggestions by other authors. In data processing literature one finds different suggestions. The simplest one is to treat the problem after transformation as if the transformed quantities were observed. It is clear from the discussions in Section 2 that such an approach does not produce the least squares solution, i.e., it does not minimize $W\{||c||\}$, even if the transformation is linear. The most sophisiticated suggestion[1,8,10] is to apply the transformation (3.18) to R, i.e., to solve the system (3.17). As we have seen in the previous section, this approach yields the least squares solution only if the transformation Y=v(X) is linear. The following example illustrates the practical consequences of such a problem falsification.

Let the observations be distances r_i and azimuth angles ϕ_i, and let the model equations represent a straight line in cartesian coordinates. Then the model equations are in terms of the original observations

$$F(r,\phi;a,b) = \begin{cases} r_1\sin\phi_1 - a - br_1\cos\phi_1 = 0 \\ r_2\sin\phi_2 - a - br_2\cos\phi_2 = 0 \\ \cdots\cdots\cdots\cdots\cdots \\ r_n\sin\phi_n - a - br_n\cos\phi_n = 0 \end{cases} \qquad (4.1)$$

The transformation of the observations into cartesian coordinates are

$$x_i = r_i \cos\phi_i \ ,$$

$$y_i = r_i \sin\phi_i \ , \ i = 1, 2, \ldots, n \ , \qquad (4.2)$$

and the model equations are in terms of the transformed observations

$$H(X,Y;a,b) = \begin{cases} y_1 - a - bx_1 = 0 \\ y_2 - a - bx_2 = 0 \\ \cdots\cdots\cdots\cdots \\ y_n - a - bx_n = 0 \end{cases} \qquad (4.3)$$

The Jacobian matrix of the transformation is

$$J = \begin{pmatrix} J_1 & & & 0 \\ & J_2 & & \\ & & \ddots & \\ 0 & & & J_n \end{pmatrix} , \qquad (4.4)$$

where

$$J_i = \frac{\partial(x_i,y_i)}{\partial(r_i,\phi_i)} = \begin{pmatrix} \cos\phi_i & -r_i\sin\phi_i \\ \sin\phi_i & r_i\cos\phi_i \end{pmatrix} \qquad (4.5)$$

We assume for simplicity that all observations are independent with estimated standard errors e_{ri} and $e_{\phi i}$, respectively. Then the estimated variance-covariance matrix R is the diagonal matrix

$$R = \begin{pmatrix} e_{r1}^2 & & & & & & \\ & e_{\phi 1}^2 & & & & 0 & \\ & & e_{r2}^2 & & & & \\ & & & e_{\phi 2}^2 & & & \\ & & & & \ddots & & \\ & 0 & & & & \ddots & \\ & & & & & & e_{rn}^2 \\ & & & & & & & e_{\phi n}^2 \end{pmatrix} \qquad (4.6)$$

The transformed variance-covariance matrix $\tilde{Q}$ is according to equation (3.18) the block diagonal matrix

$$\tilde{Q} = JRJ^T = \begin{pmatrix} Q_1 & & & \\ & Q_2 & & 0 \\ & & \ddots & \\ & 0 & & Q_n \end{pmatrix} \tag{4.7}$$

where

$$Q_i = \begin{pmatrix} e_{ri}^2\cos^2\phi_i + e_{\phi i}^2 r_i^2 \sin^2\phi_i & (e_{ri}^2 - e_{\phi i}^2 r_i^2)\sin\phi_i\cos\phi_i \\ (e_{ri}^2 - e_{\phi i}^2 r_i^2)\sin\phi_i\cos\phi_i & e_{ri}^2\sin^2\phi_i + e_{\phi i}^2 r_i^2\cos^2\phi_i \end{pmatrix} \tag{4.8}$$

For a numerical example we take the ten points listed in Table I as observations and assume that their standard errors are

$$e_{ri} = 0.048,\ e_{\phi i} = 27.5^\circ,\ i = 1, 2, \ldots, n. \tag{4.9}$$

We made three adjustments. First, the r,ϕ-data were used together with the model equations (4.1). In the second adjustment, the x,y-data were used together with the model equations (4.3) and the transformation function (4.2) in a utility program[15] based on the normal equations (3.11). The results of both adjustments were identical, as they should be, and they are listed in Table II. The listed standard errors of the parameters are the square roots of the diagonal elements of V_t, computed with formula (3.8). The correlation coefficient c_{ab} is the off-diagonal element of the correlation matrix C_t, defined by

$$C_t = D_t^{-\frac{1}{2}} V_t D_t^{-\frac{1}{2}} \tag{4.10}$$

where D_t is the diagonal matrix of V_t. The standard error of weight one is defined by

$$m_o = \left[\frac{1}{n-p} C^T R^{-1} C\right]^{\frac{1}{2}} = \left[\frac{1}{n-p} W\right]^{\frac{1}{2}} . \tag{4.11}$$

Figure 1a shows the result of the adjustment in the ϕ,r-plane, i.e., in the plane of the original observations. The accuracies of the observations are indicated by error ellipses around the observed points. The adjustment is indicated by connecting the observed points with the corresponding corrected locations on the fitted curve. The figure shows that all adjustments are in the direction of largest uncertainties.

Figure 1b shows the same result in the x,y-plane. The accuracies of the transformed observations are again indicated by error ellipses, corresponding to the transformed variance-covariance matrices $\tilde{Q}_i$. In this presentation the adjustments seem to be in directions other than those with largest uncertainties. This is typical for nonlinear transformations of observations. The object of the fitting is to minimize residuals of the original observations. The presentation in the x,y-plane is distorted by the nonlinearity of the transformation.

In a third adjustment we used the x,y-data, the model equation (4.3), and the variance-covariance matrix $\tilde{Q}$, defined by equation (4.7). The treatment, suggested by Deming[1] and other authors, was described in Section 3, equations (3.16) through (3.18), as a falsification of the problem. The numerical results of this adjustment are listed in Table II. They are different from the previous results, and the increase of m_o indicates that the solution is not optimal. We notice also that the correlation coefficient c_{ab} has changed its magnitude and sign.

Figure 2b shows the results of the adjustment in the x,y-plane. It indicates that the adjustment would indeed be optimal, if x,y were the observations and $\tilde{Q}$ was their variance-covariance matrix. However, when the same results are plotted in the ϕ,z-plane, Figure 2a, then it becomes obvious that the adjustment has not achieved the goal to minimize the residuals of the original observations ϕ,r. The treatment of transformations of observations in this form is a falsification of the problem. The results are approximations to the least squares solution, but since the quality of the approximations are not known, they may be useful only as initial approximations for a least squares algorithm. However, in a case like this example, an initial approximation could be simpler obtained, e.g., graphically by drawing a straight line in the x,y-plane through the observations.

Next, we present an example for the linearization of parameters. Let the model equation be

$$y - Ax^B \exp\left(\frac{C}{x}\right) = 0, \qquad (4.12)$$

where x and y are observations and A, B, and C are model parameters. An equivalent model formulation is

$$\ln y - a - b \ln x - \frac{c}{x} = 0. \tag{4.13}$$

In equation (4.13) the parameters a, b, and c enter linearly. One can expect a much better performance of solution algorithms if equation (4.13) is used. The parameter transformation is in this example

$$\left.\begin{aligned} A &= e^{a}, \\ B &= b, \\ C &= c, \end{aligned}\right\} \tag{4.14}$$

and the Jacobian matrix, needed in equation (2.23) is

$$\frac{\partial(A,B,C)}{\partial(a,b,c)} = \begin{pmatrix} e^{a} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{4.15}$$

Another example is the trigonometric model

$$y - A \cos \frac{x-B}{C} = 0. \tag{4.16}$$

An equivalent model is

$$y - a \sin(cx) - b \cos(cx). \tag{4.17}$$

The corresponding parameter transformation is

$$\left.\begin{aligned} a &= A \sin(B/C), \\ b &= A \cos(B/C), \\ c &= 1/C, \end{aligned}\right\} \tag{4.18}$$

with the Jacobian matrix

$$\frac{\partial(A,B,C)}{\partial(a,b,c)} = \left[\frac{\partial(a,b,c)}{\partial(A,B,C)} \right]^{-1} =$$

$$= \begin{pmatrix} \sin(B/C) & (A/C)\cos(B/C) & -(AB/C^2)\cos(B/C) \\ \cos(B/C) & -(A/C)\sin(B/C) & (AB/C^2)\sin(B/C) \\ 0 & 0 & -1/C^2 \end{pmatrix}^{-1}, \qquad (4.19)$$

In this example, the model (4.17) is linear only with respect to two parameters. However, the difference of numerical treatments of the problem is dramatic if one uses equation (4.16) or equation (4.17), respectively. In numerical experiments we found that in order to achieve convergence, one had to start with parameter values A,B,C within few percent of their least squares values. Using the parameters a,b,c and the model equation (4.17), one achieves fast convergence, e.g., with the initial values a=b=0.

5. SUMMARY AND CONCLUSIONS. Manipulations of model equations that produce simpler but equivalent equations can greatly facilitate the preparation of the problems (e.g., computer programming) for utility routines. The manipulations can also improve the performance of numerical algorithms. If the manipulations are merely algebraic and/or involve nonlinear transformations of the model parameters, then their application is straight forward and their implementation simple. If, however, the manipulations include transformations of observations, then one has to transform also the normal equations correspondingly. Neglect of this transformation falsifies the problem and produces results that are of unknown quality and equally reliable as, e.g., a graphical construction of a fitting curve. A correct implementation of transformations of observations requires the programming of the transformation function, including its first and second order derivatives. It also does not improve the performances of algorithms. Therefore, in most cases, it is more efficient to formulate the model equations in terms of the original observations, thereby avoiding the programming of the transformation function.

The need for second order derivatives of the model equations has been often overlooked. In order to avoid the programming of these derivatives, most authors suggest to use a first order Gauss-Newton algorithm for the solution of the normal equations, instead of a second order Newton-Raphson algorithm. The performance of the former may be often comparable to the latter, because even with more iterations, the computing effort can be less due to the simpler equations. Second order derivatives of the model equations (and of the transformation function) are, however, needed to compute the linear terms in formulas

for variance estimates of the results. Their neglect cannot be justified cursory by the argument that linearized model equations are already second order accurate and, therefore, their second order derivatives are not needed. It can be shown that the linearized normal equations do contain these derivatives and, therefore, are needed in the linearized variance propagation formula. Formulas for variance estimates that do not contain second order derivatives are less than first order accurate.

TABLE I. OBSERVATIONS ϕ AND r AND CORRESPONDING CARTESIAN COORDINATES

ϕ	r	x	y
206.6°	0.559	-0.50	-.025
26.6°	1.342	1.20	0.60
26.6°	2.236	2.00	1.00
26.6°	3.354	3.00	1.50
26.6°	4.472	4.00	2.00
123.7°	1.803	-1.00	1.50
92.9°	1.952	-0.10	1.95
68.2°	2.693	1.00	2.50
52.4°	4.100	2.50	3.25
42.0°	6.727	5.00	4.50

TABLE II. ADJUSTMENT RESULTS

Case 1 and 2 (Original and Transformed Problem)

$a = 0.381 \pm 0.298$

$b = 1.141 \pm 0.744$ $\qquad c_{ab} = 0.015065$

$m_o = 1.24541$

Case 3 (Falsified Problem)

$a = 0.680 \pm 0.407$

$b = 1.837 \pm 0.259$ $\qquad c_{ab} = -0.568659$

$m_o = 1.75646$

The standard error of weight one, m_o, is not included in the standard errors of the parameters.

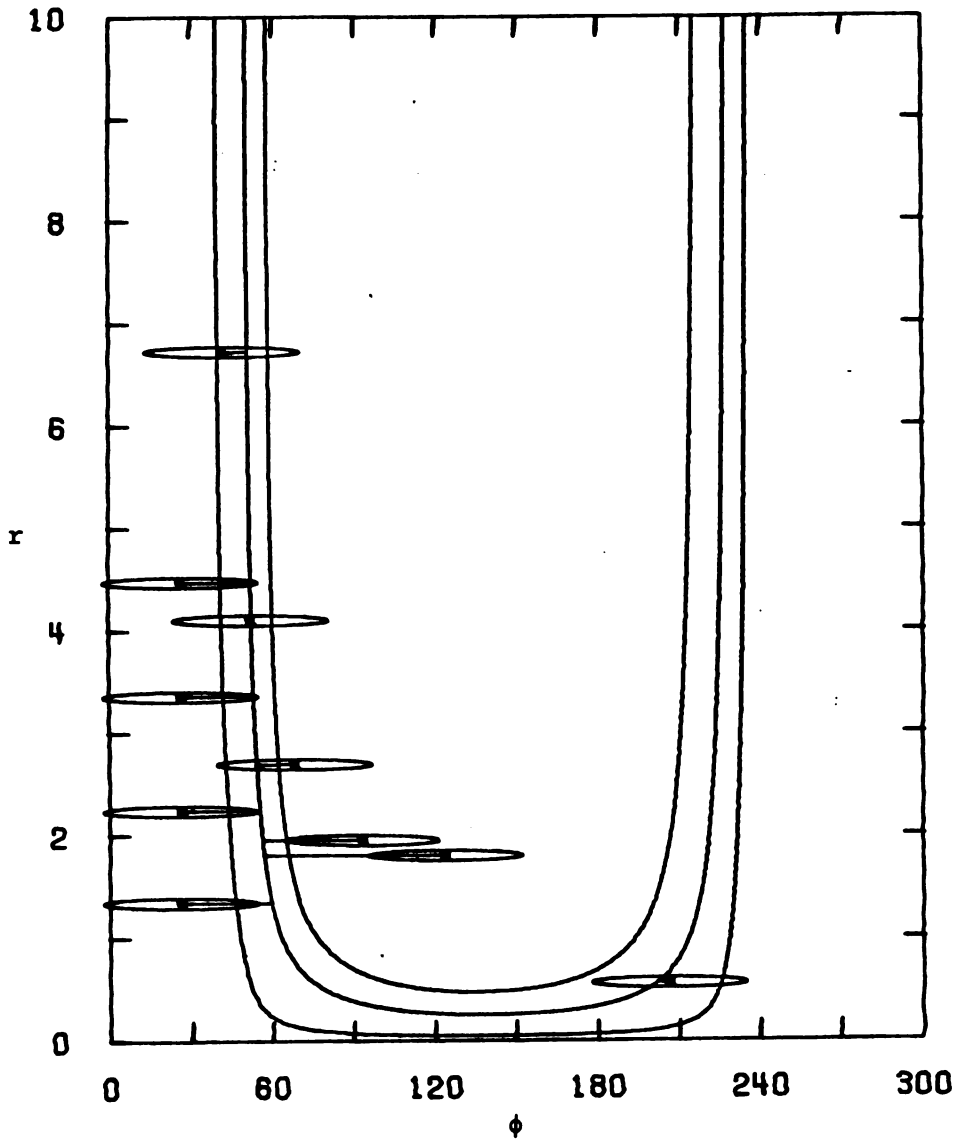


Figure 1a. Adjustment in ϕ,r-space.

The data are shown with their one standard error ellipses and the adjusted curve is shown with one standard error confidence limits. The same results are shown by Figure 1b in the cartesian x,y-plane.

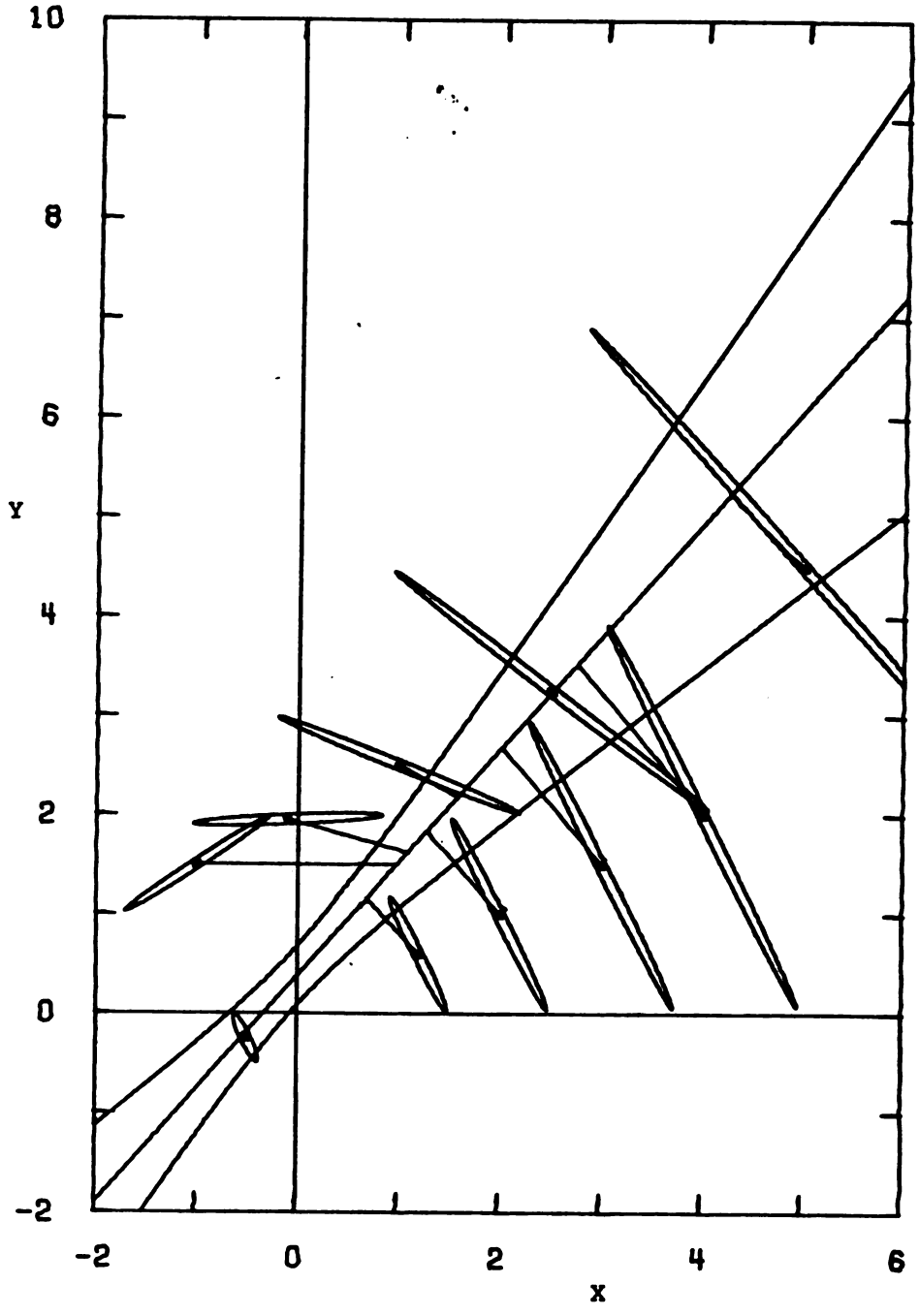


Figure 1b. Adjustment in ϕ,r-space.

The transformed data are shown with their one standard error ellipses and the adjusted line is shown with one standard error confidence limits. The same results are shown by Figure 1a in the ϕ,r-place of observations.

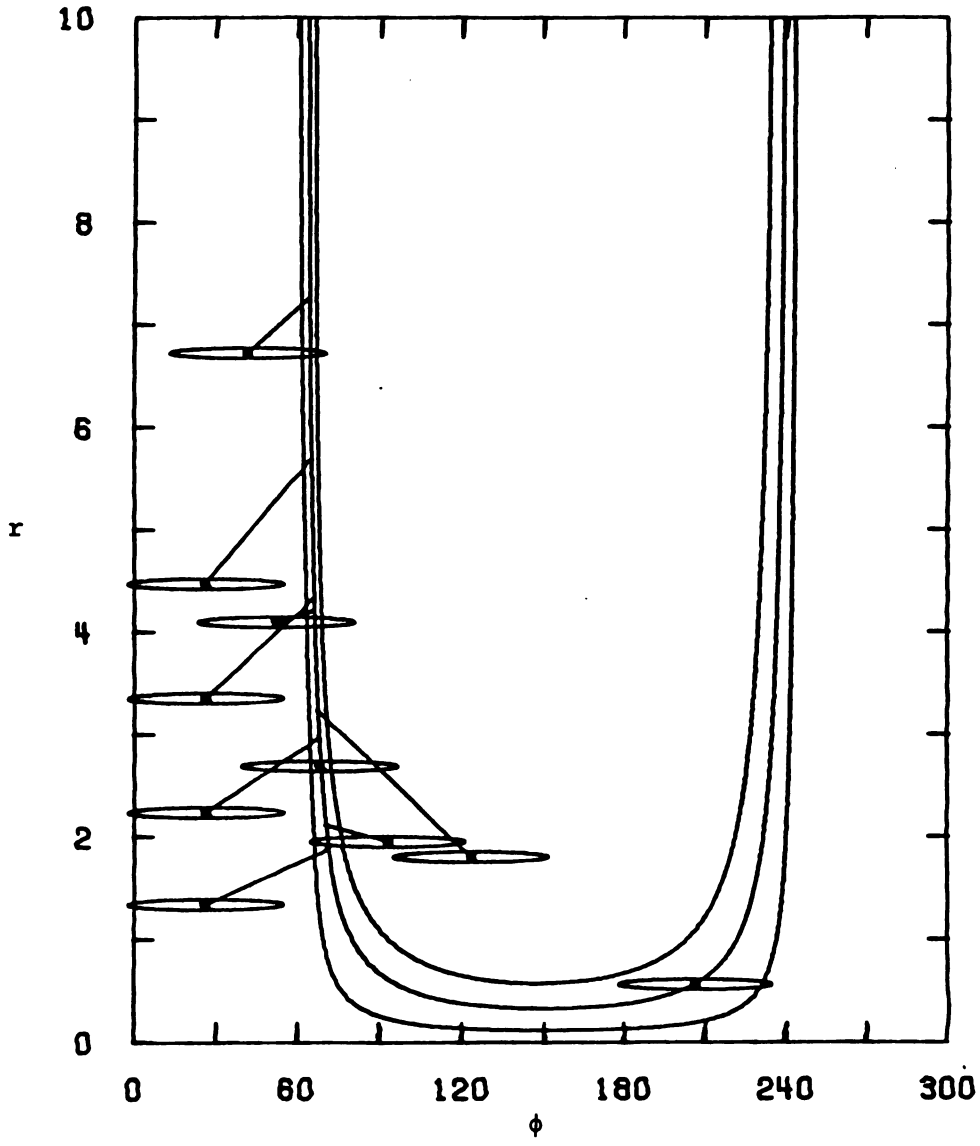


Figure 2a. Falsified Adjustment in x,y Space.

The data are shown with their one standard error ellipses and the adjusted curve is shown with one standard error confidence limits. The same results are shown in Figure 2b in the cartesian x,y-plane.

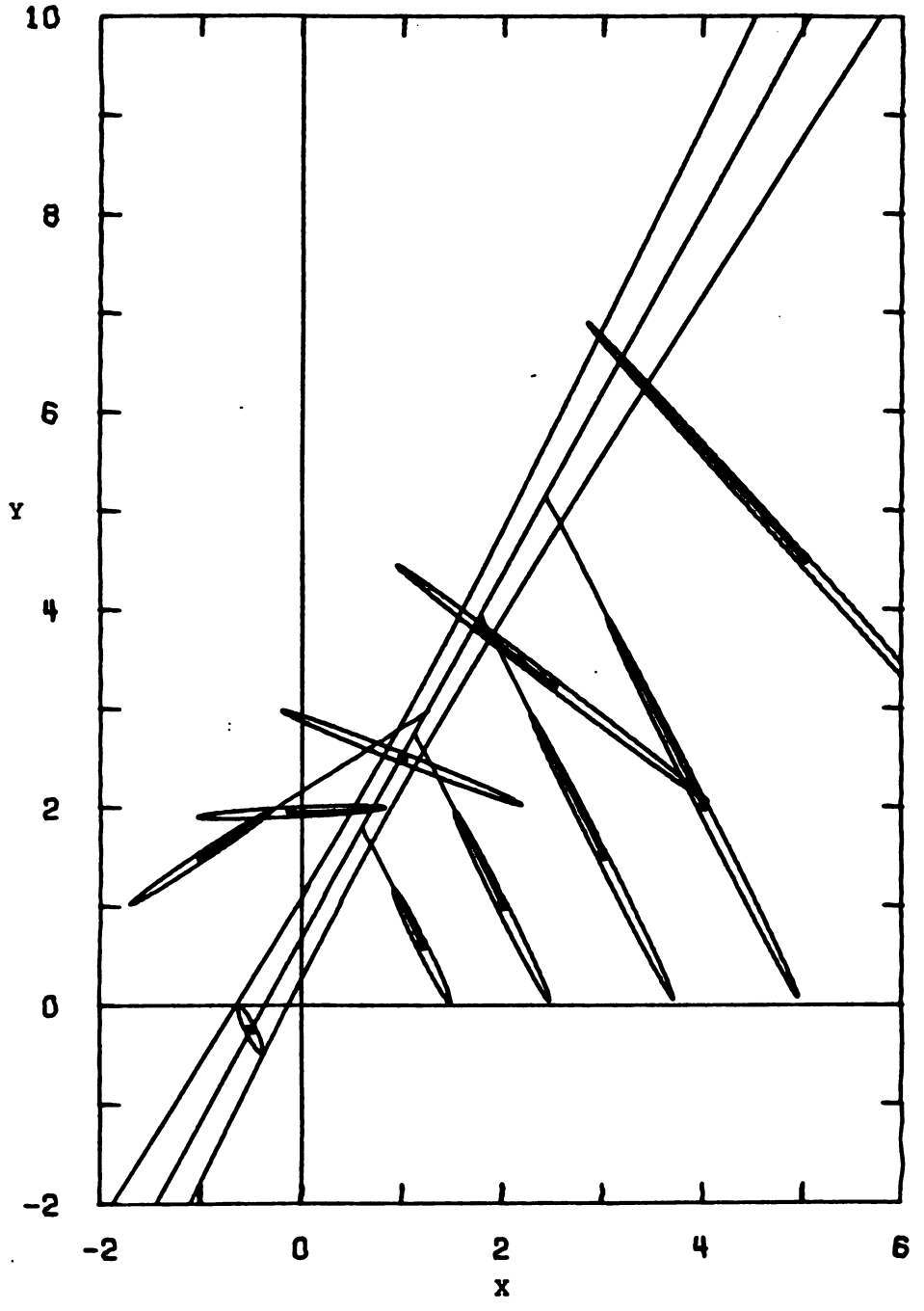


Figure 2b. Falsified Adjustment in x,y-space.

The transformed data are shown with their one standard error ellipses and the adjusted line is shown with one standard error confidence limits. The same results are shown in Figure 2a in the ϕ,r-plane of observations.

APPENDIX

We provide a set of iteration formulas that are derived from the Newton equation (3.6) by algebraic manipulations. First, we define the following matrices:

$$G = (F_x R F_x^T)^{-1} \tag{A.1}$$

$$A = R F_x^T G F_x - I \tag{A.2}$$

$$\Gamma = [I + AR(K^T F)_{XX}]^{-1} \tag{A.3}$$

$$E_o = \Gamma \cdot [AC - R F_x^T G F_x] \tag{A.4}$$

$$E_1 = \Gamma \cdot [R F_x^T G F_t + AR(K^T F)_{Xt}] \tag{A.5}$$

$$D_o = (K^T F)_{tX} - F_t^T G F_x R (K^T F)_{XX} \tag{A.6}$$

$$D_1 = (K^T F)_{tt} - F_t^T G F_x R (K^T F)_{Xt} \tag{A.7}$$

$$N = F_t^T G F_t - D_1 + D_0 E_1 \tag{A.8}$$

The iteration equations are

$$N\tau = F_t^T G(F_x C - F) + D_0 E_0 \tag{A.9}$$

$$K + \kappa = G(F_x C - F) + G[F_t + F_x R(K^T F)_{Xt}]\tau - G F_x R (K^T F)_{XX} \varepsilon \tag{A.10}$$

$$\varepsilon = E_0 - E_1 \tau \quad . \tag{A.11}$$

Numerical experiments have shown that the convergence of the iteration is enhanced if the equations are used in a subiteration mode by iterating alternatively on the parameters and residuals, respectively. For parameter subiteration only equations (A.9) and (A.10) are used, assuming $\varepsilon \equiv 0$. For residual subiteration one sets $\tau \equiv 0$ and uses equations (A.10) and (A.11).

In the variance formula (3.8) one uses N, defined by equation (A.8) and

$$S = F_t^T G F_x + D_0 \Gamma A. \qquad (A.12)$$

Another equivalent set of Newton-Raphson iteration equations are given in reference 13. None of the sets are numerically superior to the other, and both require subiterations of parameters and residuals for efficiency.

Gauss-Newton iteration equations can be obtained from Newton-Raphson iteration equations by setting all second order derivatives zero. The convergence of Gauss-Newton algorithms is inferior, but in some applications they have a larger domain of convergence.

Iteration equations for least squares problems with transformations of observations can be obtained from the formulas in this Appendix by substituting

Q for R

Δ for C

and

Ξ for $(K^T F)_{XX}$.

Expressions for Q, Δ, and Ξ in terms of the model and the transformation functions are given in Section 3, equations (3.13), (3.14), and (3.15).

REFERENCES

1. W. Edwards Deming, Statistical Adjustment of Data, John Wiley and Sons, London, 1944.

2. John E. Freund, Mathematical Statistics, Prentice Hall, Englewood Cliffs, NJ, 1962.

3. Karel Rektorys, ed., Survey of Applicable Mathematics, The M.I.T. Press, Cambridge, MA, 1969.

4. Yonathan Bard, Nonlinear Parameter Estimation, Academic Press, New York, NY, 1974.

5. Yo Lun Chou, Statistical Analysis, Holt, Rinehart, and Winston, New York, NY, 1975.

6. A.S.C. Ehrenberg, Data Reduction, John Wiley and Sons, New York, NY, 1975.

7. Thomas H. Wonnacott and Ronald F. Wonnacott, Introductory Statistics, 3rd ed., John Wiley and Sons, New York, NY, 1977.

8. R.M. Passi, Use of Nonlinear Least Squares in Meteorological Applications, Journal of Applied Meteorology, 16 (1977), 827-832 and 17 (1978), 1579-1580.

9. Heike von Benda, Zur Gitterkonstantenbestimung mit Ausgleichsmetoden, Zeitschrift für Kristallographie, 149 (1979), 205-209.

10. Ph. Wahl, Analysis of Fluorescence Anisotrophy Decays by a Least Square Method, Biophysical Chemistry 10 (1979), 91-104.

11. H.J. Britt and R.H. Luecke, The Estimation of Parameters in Nonlinear, Implicit Models, Technometrics 15 (1973), 233-247.

12. P.A.D. DeMaine, Automatic Curve Fitting, I, Test Methods, Computers and Chemistry, 2 (1978), 1-6.

13. Allen J. Pope, Two Approaches to Nonlinear Least Squares Adjustments, The Canadian Surveyor, 28 (1974), 663-669.

14. Robert E. Barieau and B.J. Dalton, Nonlinear Regression and the Principle of Least Squares, Bureau of Mines Report of Investigations 6900, 1967.

15. Aivars Celmiņš, A Manual for General Least Squares Model Fitting, Ballistic Research Laboratory Technical Report ARBRL-TR-02167, 1979.

16. E. Stark and E. Mikhail, Least Squares and Non-Linear Functions, Photogrammetric Engineering, 1973, 405-412.

17. William H. Sachs, Implicit Multifunctional Nonlinear Regression Analysis, Technometrics 18 (1976), 161-173.

18. Aivars Celmiņš, Least Squares Adjustment with Finite Residuals for Non-Linear Constraints and Partially Correlated Data, Ballistic Research Laboratory Report, BRL-R-1658, 1973.

ANALYSIS OF DATA WITH THE NONLINEAR LEAST CHI SQUARE ALGORITHM

Richard L. Moore
US Army Armament Research and Development Command
System Evaluation Office
Dover, NJ 07801

ABSTRACT. This paper reports on the use of the least chi-square algorithm for fitting data to non-linear functions of the parameters. A well known computer program of the National Institutes Health, SAAM-27, has been modified to use this algorithm. Comparison of the ordinary least-squares algorithm with the new algorithm have been made on four different problems as follows: Pressure waves in gun chambers, control of aircraft yaw, a biomedical kinetic reaction involving four measured components, and a very non-linear nuclear reactor kinetics problem. The preliminary results indicate that the least chi-square algorithm is practicable, that the computing time is increased for short problems, but evens out for long problems.

The least chi-square algorithm appears to be less failure prone than least squares and a test has been inserted in the program to preclude any iterations which might tend toward maximizing the autocorrelations as could occur when their initial value is large.

I. INTRODUCTION. It is accepted procedure in analysing the goodness-of-fit of experimental data to a theory which is nonlinear in the adjustable parameters to estimate whether the residuals are consistent with being drawn from a normally distributed population in a random sequence. A common statistic to test the random sequence hypothesis is the sum of the squares of the normalized autocorrelation coefficients frequently called the Box-Pearce test. However, if these tests indicate a lack of agreement with the hypothesis, no rationale has been available to modify the parameters to obtain a better fit. A solution to this problem has been provided by the use of a least chi square algorithm which estimates the parameters which give the greatest probability that the residuals arise from a population with variance σ_e^2 , and are sampled from a random sequence.

II. SUMMARY OF MATHEMATICS.

Following the notation of Aitken (1) and as previously derived by Moore (2, 3), we define the following:

The transpose of a vector or matrix is indicated by a ' on the symbol u'.

u is the vector of observed values.
y^* is the vector of theoretical values corresponding to u.
θ^* is the vector of the estimates of the unknown parameters.
P^* is the matrix of the partials of y^* with respect to θ^*.

$$V_1^{-1} = \begin{bmatrix} 0\,1\,0\,\ldots\,0 \\ 0\,0\,1\,0\,\ldots\,0 \\ 0\,0\,0\,1\,0\,\ldots\,0 \\ \ldots \end{bmatrix}; \quad V_2^{-1} = \begin{bmatrix} 0\,0\,1\,0\,\ldots\,0 \\ 0\,0\,0\,1 \\ 0\,0\,0 \end{bmatrix}$$

$$V_j^{-1} = \begin{bmatrix} 0\,0\,\ldots\,0\,1\,\ldots \\ 0\,0\,0\,\ldots\ldots\,0\,1\,\ldots \\ \ldots \end{bmatrix}$$

$$\alpha_j = \frac{2r_j V_j^{-1}}{(d)'(d)/\sigma_e^2 - 2\sum_{i=1}^{s} (r_i)^2 V_i^{-1}}$$

$$\Gamma = 1 + \sum_{i=1}^{s} \alpha_i r_i V_i^{-1}$$

The term V_j^{-1} is the inverse of the variance of r_j^2.

$$\chi_T^2 = \sigma_e^{-2}\, d'd + \sum_{j=1}^{s} r_j^2 / V_j$$

The calculaton of $\{\delta\theta\}$ is done from the following expression which is the same as the usual non-linear least square interation except for the matrix Γ .

$$[\delta\theta^*] = [P^{*\prime}\,\Gamma\,P^*]^{-1}\,P^{*\prime}\,\Gamma\,u^*.$$

In the usual case, Γ equals I.

This expression has been programed into the Simulation And Analysis Modeling (SAAM-27) (4, 5) program of Berman et al, by post-multipying $P^{*\prime}$, by Γ , and letting the program proceed from that point, with the data for $P^{*\prime}$ being replaced in the memory by the product $P^{*\prime}\Gamma$. The usual iteration continues from this point. The computer program resulting from this change has been designated for control purposes as SAACH, and has been tested on the CDC 6600 at ARRADCOM, Dover, to determine the following questions:

1. How much change is there in the final parameter estimates?

2. What change, if any, is there in the number of iterations?

3. What change is there in the time per iteration?

III. EXAMPLES. Four problems of different origin which use different mathematical models have been run on the SAACH program to answer the above questions. In the first example: Gun Chamber Pressure Waves, the mathematical model used is the superposition of two pressure waves generated by analytic models in the program, with the adjustment of up to eight parameters to obtain the best fit to observed data. In the second example, an aircraft control system simulation, the mathematical model is a set of four linear differential equations, simulating the Yaw Damper system on an aircraft. These equations were solved by a special procedure developed for SAAM-27 by Berman et al. (6), with up to four adjustable parameters. In the third example, a biomedical problem furnished as a test case by Miss Rita Straub of Brookhaven National Laboratory, the mathematical model was a set of seven coupled linear differential equations with five adjustable parameters; this was solved by the same method as used in the second case. In the fourth and final example: KEWB Kinetics, a simulation of the nuclear reactor transients of the Kinetic Experiment Water Boiler, the mathematical model was an extremely non-linear set of coupled differential equations as described by Hetrick and Gamble (7). These equation were integrated by the fourth order Runge-Kutte integration procedure of SAAM-27, with only one adjustable parameter.

III-1 Gun Chamber Pressure Waves. Unusual pressure waves suggestive of an acoustic wave superposed on the normal gun chamber pressure-time curve, have occurred in tests of the XM211 propellant charges at zone 3 for the M101 projectile in the 155mm gun, (Knutelski, (8)). Analysis of these waves was initiated by Mr. B. G. Knutelski of the Large Caliber Weapom Systems Laboratory using the SAAM-27 program. A parallel analysis was made by the author using the SAACH program. In order to have as little bias as possible injected into the comparison, the Knutelski model, data, and procedure was followed as closely as possible. The resulting data fit was later improved by using more data and improved models. The history of this analysis is important because it illustrates the problems which arise when no prior knowledge is available about the best-fitting model. (This example is the only one of the four examples for which prior knowledge was not available.)

The first case was run using the data shown as asterisks in Figure 1. This figure shows the theoretical fit by the following model:

$$P = P(1)\sin\{2\pi(P(2)t + P(3))\} + P(4)\sin\{2\pi(P(5)t + P(6))\} .$$

Fig. 2 shows the theoretical fit by the same model as above using the Least Chi Program (SAACH) with five autocorrelation coefficients (BGK 1.101). Table 1 indicates the number of iterations to convergence and the final values of the parameters (the initial values were the same). The value of the sum of the squares (X_1^2) is given for comparison, as well as the autocorrelation coefficients up to rank 5. Case BGK had slightly lower values of sums of squares, but the chi square was much smaller for BGK 1.101. (The symbol X will be used for the greek letter Chi for the rest of this report.)

Because not all the data points available were used in this preliminary analysis, additional data were obtained and entered into the computer using the same model and same initial conditions as in the previous runs. In this case (BGK 3.002) the least squares iteration stopped at seven iterations; as shown in Figure 3, the fit was poor and the convergence obviously false. The least chi square iteration, BGK 3.102, using the same data terminated at 14 iterations with an obviously better fit (Fig 4), but yet not a good eyeball fit. The results of both cases are also shown in Table 1. The autocorrelation coefficients are large for case 3.102, and indicate the general lack of fit.

Because of this obvious lack of fit in BGK 3.102, the model was changed to the following,

$$P = P_1 \exp\{ - (t-t_1)^2/2\sigma_1^2\}$$
$$+P_2 \exp\{ - (t-t_2)^2/2\sigma_2^2 \} \times \sin\{2\pi f (t-t_3) + \pi/2\}$$

Three parallel cases were computed once the fit was good enough to permit iteration. Because of computing difficulties which arose when trying to converge on six or seven parameters, the iteration was initially restricted to four parameters: Once the fit was good and had converged using these four parameters, their final values were used as initial values for a six-parameter fit. Finally, all eight parameters were allowed to vary.

The results of this series of analysis are plotted in Figs 5, 6, and 7. The case numbers are BGK-3.30356301-0, 3.30356511-5 and 3.30356511-10 respectively. The first has no autocorrelation coefficients; the second, 5; and the third, 10. The parameters for these cases are given in Table 2, (note that the last three digits only of the identifier are used here). The estimated errors are the estimated standard deivations based on the value of the sum of the squares. In the case of 511-5 and -10, the value of X_T^2 was used rather than the sum of squares. The statistical validity of this procedure has not yet been established.

Results shown in Figures 5, 6, and 7 indicate that the apparent fit to the data is best for the case of five autocorrelations, (Fig 6). In this figure the autocorrelations were weighted higher than in Fig 7, where ten autocorrelations were used, and of course much higher than in Figure 5, where no weight was given to the autocorrelations.

It is clear from Table 2 that ordinary least squares, case 301 indicates a small fractional standard deviation as compared to the other two cases, but yet the fit to the data is not as good as seen from its plot, (Figure 5).

The last row of Table 2, gives the values of σ_e^2 , the experimental variances assumed for these cases. These were arbitrary numbers in this case, because the precision of the measurement system is probably much greater than the value given i.e., the variances should be smaller. However, if smaller values were used, such as when case 511-10 is compared to 511-5, the weight on the sum of the squares is greater but the goodness of fit appears to decrease.

Table 2 also shows the effect of least chi-square in terms of number of iterations, and computing time. When five autocorrelations were used, as in case 511-5, only a small increase in number of iterations is found and a moderate increase in computing time as compared to 301. If ten autocorrelations were used, as in 511-10, the number of iterations increased, and the time increase was 1.8 times greater, giving about double the increase in time for double the number of autocorrelation coefficients.

Table 3 shows the autocorrelations up to order 20 for the three cases. The values of X_1^2, X_2^2, and X_T^2 for the number of autocorrelations used (0, 5, 10) is shown in the last rows of this table.

The difference in the assumed experimental variances accounts for most of the difference between X_1^2 for cases 511-5 and 511-10. If the experimental variances had been the same, X_1^2 would have been 62.96 or 57.62. Case 511-5 appears (in the figures) to fit better because the first five autocorrelations as well as most of the later autocorrelations are smaller.

III-2 Aircraft Control Systems. The block diagram of a typical problem of this type is shown in Fig 8. To optimize the design four parameters may be adjusted to give the best fit to a desired response curve. These parameters are δ_r, K_ψ, τ_1, and K_β. These correspond to the parameters L(0,4), L(4,1), L(4,2) and L(4.3). A previous analysis of this example using SAAM-23 was available. As a result, a completely unbiased comparison of least squares and least chi squares prodecures was difficult to ensure. Two different approaches were used on this example. First, the "data"--corresponding to the desired curve--was used "as is" for comparison with the calculated response. Second, a vector of a random sequence of normally distributed errors from a population with variance of $(.033)^2$ was added to the data vector to simulate the effects of sampling error; this may be considered to represent an allowable error or tolerance in fitting the curve.

In the first approach, the cases to be compared are 2-6 and 4. Case 1-6 was a reference run which adjusted four parameters, and started near to the final values. It iterated three times and took 23.6 sec to complete. A similar case, 2-6 used the same starting point and used six autocorrelations coefficients. It failed to improve the fit in but one iteration, primarily because it attempted to increase the autocorrelations in its attempts to improve the fit. (Several cases of this type were found which led to a modification in the least chi-square algorithm, to be discussed later). The data on the parameters, autocorrelations, and chi square are given in the first column of Table 4. The fit to these data are shown in Fig 9.

In an attempt to understand this problem various strategies were tried, but the results were nearly all similar. Case 4 and 4.1 are typical. In these comparisons the number of adjustable parameters was reduced to three, the value of the experimental variance was increased to 1.0 and the rank of the autocorrelations was taken as 5, 6, and 12. In the least squares case, case 4, as indicated in Table 8, the calculation converged in six iterations, taking 11.2 sec. In case 4.1, with the same initial point, the three attempts iterated for six iterations, for the same time, 11.2 sec, but because the autocorrelations were large, and the value of X_2^2 was large compared to X_1^2, they all eventually diverged from a good fit. These results indicated that the algorithm was not reliable when the autocorrelations were large at the outset. To correct this, an internal algorithm will be added so that value of α which is $1/(X_1^2 - 2X_2^2)$ will not be allowed to be greater than .5. It is believed that this change will prevent situations of this kind from arising in the future, but the effect of this change has not yet been fully tested. To determine whether the least chi-square technique is valid for the Yaw Damper calculation, the second approach, the addition of Monte Carlo errors to the desired response curve, was used as a test.

For this second approach, a random sequence was added to each of the data points. The value of σ_e^2 was set at $(.033)^2$, six autocorrelations were used for the problem which was identified as CONRLM 4.011-6. Another run was used on the same data with the standard least squares algorithm. Fig. 10, (CONRLM 4.012) shows the fit obtained for the data and is typical of the results. Table 4 shows the number of iterations for each case. It took 4 iterations for the ordinary algorithm to converge, and only two for the least chi-sq. algorithm with six autocorrelation coefficients (CONRLM 4.011-6). The time for one iteration was 8.2 and 8.5 sec respectively. (Part of the increase in time for the least chi-square case was due to several attempts in both iterations to improve the fit by reducing the step size.) As shown in Table 4 the parameters L(0,4), L(4,2) and L(4,3) appear to be different by significant amounts, and the difference in the "significance" of the two results is considerable. (The autocorrelations for case 4.011-6 appear well within the random range.)

III-3 Brookhaven Example. A sample test case was received from Miss Rita Straub of Brookhaven National Laboratory. The exact nature of the problem was unspecified but from the form of the differential equations given in table 5, it appears to be a kinetic problem in which the material in component one decays into components two to five, and component two may change into component one. Component seven is composed of components three, four, and five. Although the "S" and "K" parameters may actually be unknown, they were assumed

known, because the present version of the program will not iterate either type of linear parameters with the least chi-square algorithm. The data were available for the amount of components 1, 3, 4, and 6 as a function of time, (where component 6 is the sum of components 1, 2, and 5).

Both the run with no autocorrelations and the run with 5 correlations (KJE 1.0023-5), took 7 iterations to converge. The results for the two cases are compared in Table 6. Since the value of X_1^2, (31.24) is large compared to X_2^2 (3.04), the major emphasis in this case was on reducing the sum of squares, and thus it is similar to the case run with no weight on the autocorrelations. As would be expected, there is only a small difference between the final values of the parameters of the two cases. Figs 10, 11, and 12 show the graphs of the data fit to the components 1, 3, and 4. (Component 6 shows an exact fit to data points and therefore a graph of this component is not provided.)

III-4. Reactor Kinetics Example. This example illustrates two things: First the use of the least chi-square algorithm, and second a good fit between data and a physically incorrect model. Hetrick and Gamble (7) proposed a non-linear feed-back term proportional to the energy in the reactivity of the KEWB reactor to describe the effect of void on reactor shutdown. Although this model gives a good fit, later experiments (9) where the void amount was inferred from measurements and where the thermal effects on reactivity were also carefully measured, showed that shutdown was due to thermal, not void effects. In the simulation, the effect of the energy on void formation was simulated by the parameter L(11, 1). The functions correspond, in numerical order, to the functions used in the simulation: (1) Nuclear reactor power level, (2) Mean temperature, (3) Mean void volume, (4)-(9) Delayed neutron groups, (10) Not used, (11) Energy released to that time. The result of the iterations is shown in Figure 13, a logarithm plot of theoretical and experimental nuclear power. In Table 7, three different cases are shown:

Case 1.003-0 was ordinary least-squares. The values of the autocorrelations and chi-squares are shown for comparison with the other two cases. Case 1.005-3 used three autocorrelations with a small value of the experimental variance thus resulting in a large value of X_1^2. Both case 1.007-6 and 1.003-0 use 1×10^7 for the experimental variance thereby reducing the emphasis on the sum of the squares of the errors. All of these runs took four iterations to converge.

Cases 1.003-0 and 1.005-3 give almost exactly the same results. On comparing 1.003-0 with 1.007-6, a difference is found in the value of the adjustable parameter L(11,1). The value of chisquare total is smaller for 1.007-6, and thus this result would be chosen over that of the other case.

The value of the chi-square for the autocorrelation (X_2^2) is much smaller for case 1.007-6, although X_1^2 is slightly larger for the same case--thus illustrating the trade-off between getting the minimum as in ordinary least squares, and reducing the autocorrelations as in least chi-squares. The data for Case 1.003 show the values for R(1) to R(6) for comparison purposes. The data show that the sum of squares does not increase from one to other appreciably, but X_2^2, the Box-Pearce statistic, does change appreciably. Each of the calculations gives a total chi square which is too large to be consistent with the residuals being drawn from a random sample.

III-5. Comparison of Computing Time. Table 8 summarizes the comparison of the number of iterations to converge, and the computing time required. As seen in the previous discussions, the number of iterations was usually about the same, except for two cases--the case 4.1 under the Yaw Damper, where the iteration with least chi square failed to properly converge, and for the XM211 Pressure Curves where the ordinary least squares took more iterations or failed to converge As seen in the last column, for all the cases except the case 4.1 under the Yaw Damper, the computing time is comparable, with a tendency for the computing time to be longer for least chi square than for least squares. The relative difference is greater when the original total computing time is short. This just means that, as would be expected, it takes a larger fraction of the computing time to compute the matrix Γ and post multiply it into $P^{*'}$ for cases where the time of iteration is short.

IV CONCLUSIONS. Based on four different types of non-linear theoretical models for data analysis, our results indicate that:

(1) Least chi-square is practicable for non-linear analysis.

(2) The computing line for least chi square is longer for the models which use less computing time, but because the convergence of this iterative procedure is somewhat better, the number of iterations (and particularly the number of "tries" per iteration) is reduced, thus keeping the total computing time about the same. Models with longer integrating time would expect to benefit more from least chi-square.

(3) With one exception as given below, the least chi-square procedure appears to be less prone to failure to converge.

(4) When the autocorrelation are large and their weighted sum is large compared to the chi-square for the residuals, the iteration tends to produce a maximum value of the autocorrelations. A test has been devised to prevent this situation from occuring.

(At the time of final editing and review of this paper, an error was discovered in the programming of the calculation of the variance of the autocorrelations, V_l . The error amounts to only a few percent but would make it difficult to reproduce the present results. It is believed that the main thrust of the results of this paper remain valid.)

REFERENCES

1. Aitken, A,C,. Proc. Roy. Soc. Edinb. A 55, 42-47 (1934) and R.L. Plackett, Principles of Regression Analysis, Oxford, Clarendon Press, 1960.

2. Moore, R.L., Proc. 1975 Army Numerical Analysis Conference, ARO Report 75-3.

3. Moore, R.L., Proc. 23d Army Conference on Design of Experiments, ARO Report 1977.

4. Berman M., E. Shahn, and M.F. Weiss, Biophys J. 2, 275, 1962.

5. Berman, Mones, and M.F. Weiss, USERS MANUAL FOR SAAM (Simulation, Analysis and Modeling) Version SAAM 27, April 15, 1977, Math Research Branch, National Institute of Arthritis and Metabolic Diseases, National Institutes of Health, Bethesda, MD.

6. Model 10 of reference 5.

7. Hetrick, D., and J. Gamble, Transient Reactivity during Power Excursions in a Water Boiler Reactors, Atomics International, Feb 9, 1958.

8. Knutelski, B. G., private communication.

9. Dunenfeld, M. S., R.K. Stett, Summary Review of the Kinetics Experiments on Water Boilers, Atomic International, SR 7087, Feb 1, 1963.

CASE	BGK	BGK 1.101	BGK 3.002	BGK 3.102
NUMBER OF ITERATIONS	14	11	7	14
P(1)	15.95	15.89	15.0	19.48
P(2)	209.9	203.7	332.6	308.00
P(3)	.722	.767	-4.16	2.97
P(4)	1.719	1.700	+5.67	1.38
P(5)	2.93×10^3	3.00×10^3	3.06×10^3	3.27×10^3
P(6)	-1.885	-2.196	-1.80	-6.0
R(I) 1	.614	.586	-	.843
2	.275	.252	-	.646
3	-.077	-.076	-	.428
4	-.186	-.175	-	.219
5	-.209	-.192	-	.046
X_1^2	15.5	16.1	-	27.6
X_T^2	30.018	25.0	-	81.6
SIGNIFICANCE	---	.44		3.5

Table 1. Results of computer runs on XM211 Pressure Oscilations

UNITS	SYMBOL	COMPUTER SYMBOL	CASE 301	CASE 511-5	CASE 511-10
10 kpsi	P_1	L(1,1)	20.04 ± 1.28%	19.7 ± 1.68%	19.95 ± 3.75%
10 kpsi	P_2	L(3,1)	4.146 ± 52%	3.33 ± 46%	3.998 ± 148.%
sec	t_1	L(2,1)	.057 ± .07%	.0568 ± .28%	.0568 ± .22%
sec	t_2	L(4,1)	.056 ± .12%	.0564 ± .14%	.0565 ± .46%
sec	t_3	L(3,3)	.0557 ± .7%	.0558 ± .50%	.0558 ± 2.1%
sec	σ_1	L(2,2)	.0027 ± 2.7%	.00279 ± 6.8%	.00270 ± 8.86%
sec	σ_2	L(4,2)	.00051 ± 36%	.000597 ± 31.%	.000519 ± 112.%
Hz	f	L(3,2)	327.6 ± 57%	344.4 ± 43.%	361.08 ± 17.6%
No of Iterations			18	19	21
Computing Time (sec)			51.2	58.0	63.2
Experimental Variance (σ_e^2)			NOT APPLICABLE	$(.45)^2$	$(.317)^2$

TABLE 2. Parameters Fitting Pressure Curve

CASE	301	511-5	511-10
ORDER			
1	.717	.680	.696
2	.477	.427	.443
3	.286	.247	.248
4	.058	.027	.013
5	-.113	-.136	-.160
6	-.246	-.258	-.291
7	-.303	-.315	-.346
8	-.322	-.334	-.361
9	-.307	-.312	-.342
10	-.245	-.240	-.276
11	-.130	-.104	-.155
12	-.011	.031	-.035
13	.069	.130	.048
14	.137	.217	.123
15	.110	.203	.106
16	.075	.168	.085
17	-.028	.057	.0004
18	-.141	-.068	-.090
19	-.202	-.149	-.130
20	-.235	-.205	-.145
X_1^2	57.9	62.96	116.2
X_2^2	-	33.01	62.4
X_{TOT}^2	-	95.97	178.6

Table 3. Autocorrelations and Chi-Square for final model of XM211 Pressure Oscillations.
X_2^2 based on the first 5 Autocorrelations for Case 511-5, and the first 10, for case 511-10

CASE		2-6	4.012-0	4.011-6
NO. of ITER.		1	4	2
L(0, 4)		18.6	17.70	18.58
L(4, 4)		53.1	53.11	53.02
L(4, 2)		.605	1.094	0.605
L(4, 3)		10.27	6.204	10.26
R(I)	1	.769	-.111	-.110
	2	.431	-.230	-.232
	3	.144	.013	.014
	4	-.040	.061	.068
	5	-.159	-.124	-.124
	6	---	.029	.030
Sum of sqs		.00275	.03147	.03138
X_1^2		.011	28.90	28.82
X_2^2		24.78	2.77	2.49
X_T^2		24.79	31.67	31.31
Signif		-.959	.273	-.087
σ_e^2		.250	$(.033)^2$	$(.033)^2$

Table 4, Results of Yaw Damper Calculations

$$\frac{dF(1)}{dt} = L(1,1)\,F(1) + L(1,2)\,F(2)$$

$$\frac{dF(2)}{dt} = L(2,1)\,F(1) + L(1,2)\,F(2)$$

$$\frac{dF(3)}{dt} = L(3,1)\,F(1) - ((L(1,3) + L(0,3))F(3)$$

$$\frac{dF(4)}{dt} = L(4,1)\,F(1) - (L(7,4) + L(0,4))F(4)$$

$$\frac{dF(5)}{dt} = - (L(7,5) + L(0,5))F(5)$$

$$\frac{dF(7)}{dt} = L(7,3)\,F(3) + L(7,4)\,F(4) + L(7,5)\,F(5) - L(0,7)\,F(7)$$

and $L(1,1) = L(2,1) + L(3,1) + L(4,1) + L(5,1)$

$QC(1) = K(1)\,F(1)$
$QC(6) = K(6)\,(S(1)\,F(1) + S(2)\,F(2) + S(5)\,F(5))$

Table 5 Differential equations for Brookhaven problem

ITEM/CASE	KJE 1.0021-0	KJE 1.0023-5
L(1, 2)	.2155	.2199
L(2, 1)	.4527	.4442
L(3, 1)	.0431	.0431
L(4, 1)	.0252	.0251
L(5, 1)	.0743	.0824
R(1)	.122	-.086
R(2)	.065	-.052
R(3)	.009	.034
R(4)	.227	-.223
R(5)	.179	-.168
X_1^2	31.22	31.24
X_2^2	3.99	3.04
X_{TOT}	35.21	34.28
Significance	-.026	-.086

Table 6. Results of Brookhaven example calculation. Autocorrelation, X_2^2, and X_T^2 for case KJE 1.0021-0 computed for comparison.

ITEM/CASE	KWB 1.003-0	KWB 1.005-3	KWB 1.007-6
L(11,1)	5.318×10^{-4}	5.3183×10^{-4}	5.262×10^{-4}
R(1)	.782	.782	.786
R(2)	.44	.44	.453
R(3)	.098	.098	-.122
R(4)	-.204	-	-.170
R(5)	-.316	-	-.274
R(6)	-.235	-	.290
X_1^2	121.14	1.2×10^9	121.89
X_2^2	36.71		35.0
X_T^2	157.85		156.89

TABLE 7. Results of Kinetic Experiment Water Boiler Calculations

CASE	ADJUST PARAM	AUTO-CORR RANK	NO OF ITER	TIME (SEC)
KEWB KINETICS				
1.003	1	0	4	84.
1.005-3	1	3	4	81.
1.007-6	1	6	4	89.
YAW DAMPER				
1=6	4	0	3	23.6
4	3	0	6	11.2
4.1	3	5,6,12	6	11.2 (FAILED)
4.011	4	6	2	8.5
4.012	4	0	4	8.2
BROOKHAVEN				
1.0021	5	0	5*	14.9
1.0023-5	5	5	7	21.5 (15.2)
XM211 PRESSURE				
BGK	6	0	14	6.5
BGK1.101	6	5	11	11.
3.002	6	0	7	11.31-/FAILED
3.102-5	6	5	15	23.
3.102-5	6	5	11*	18.2

TABLE 8. Comparison of Computing Time.

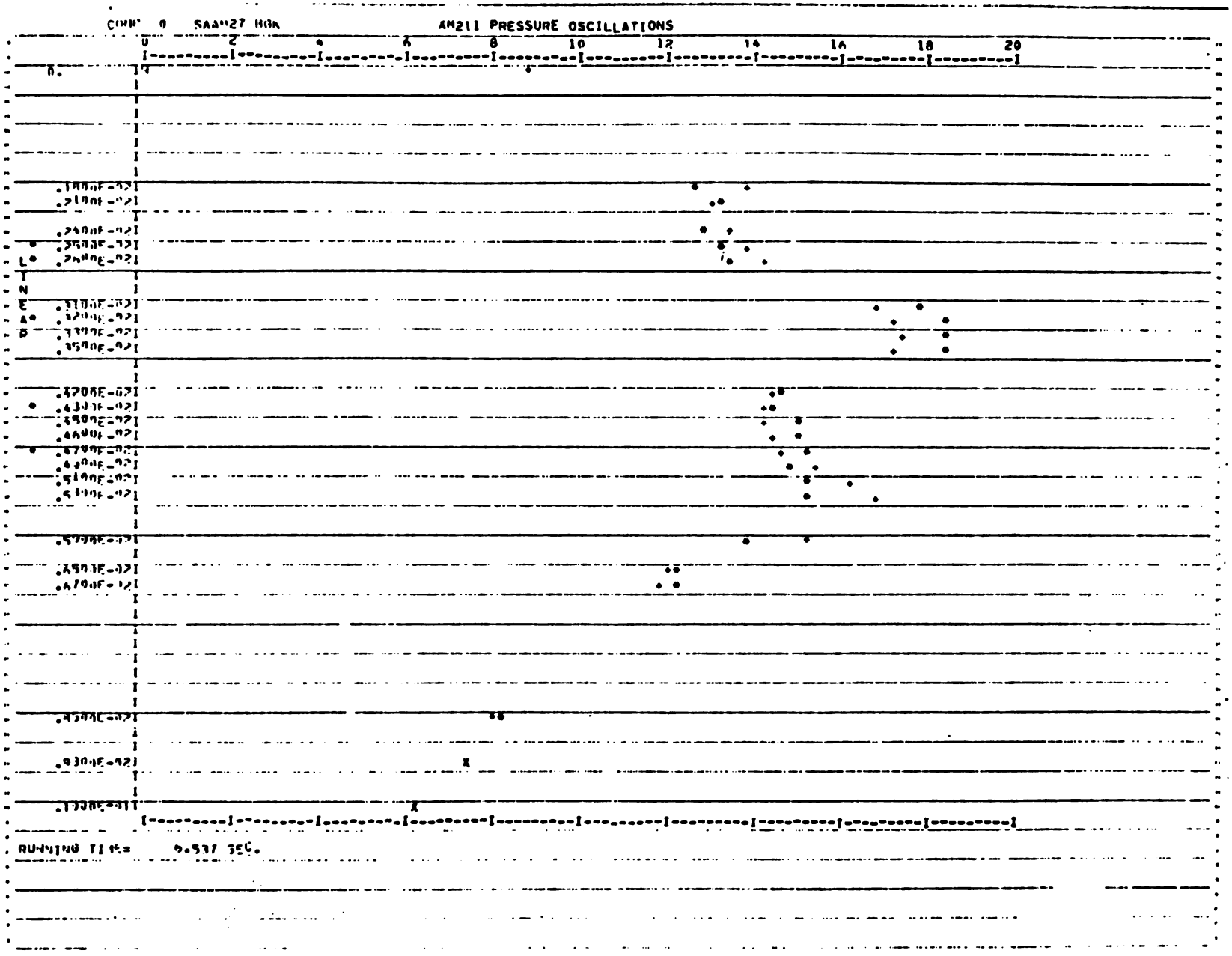


Figure 1 - Fit to Partial Data of Pressure Curve of XM211 Propellant Charge by Ordinary Least Squares.

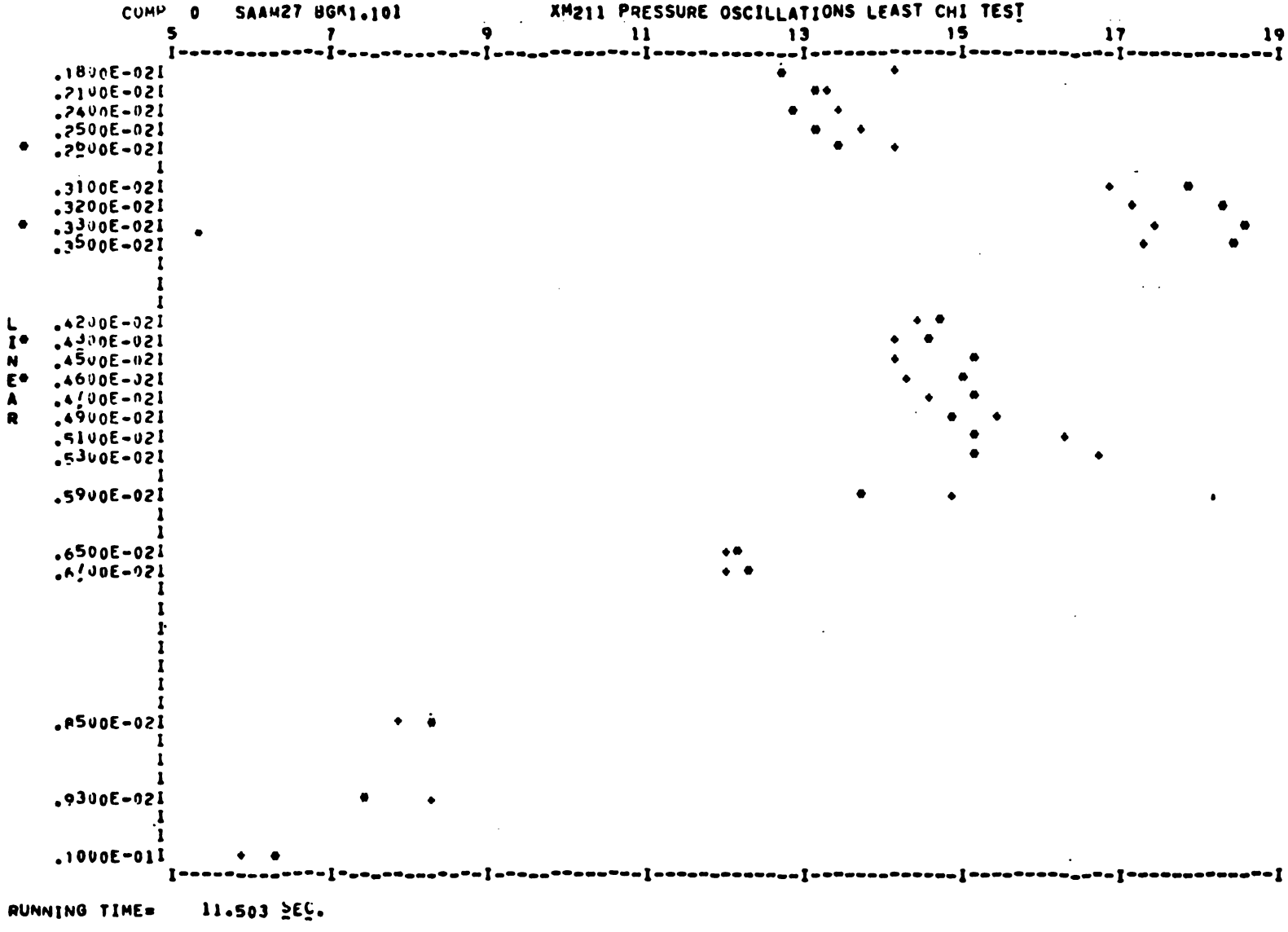


Figure 2 - Fit to Partial Data of Pressure Curve of XM211 Propellant Charge by Least Chi Squares.

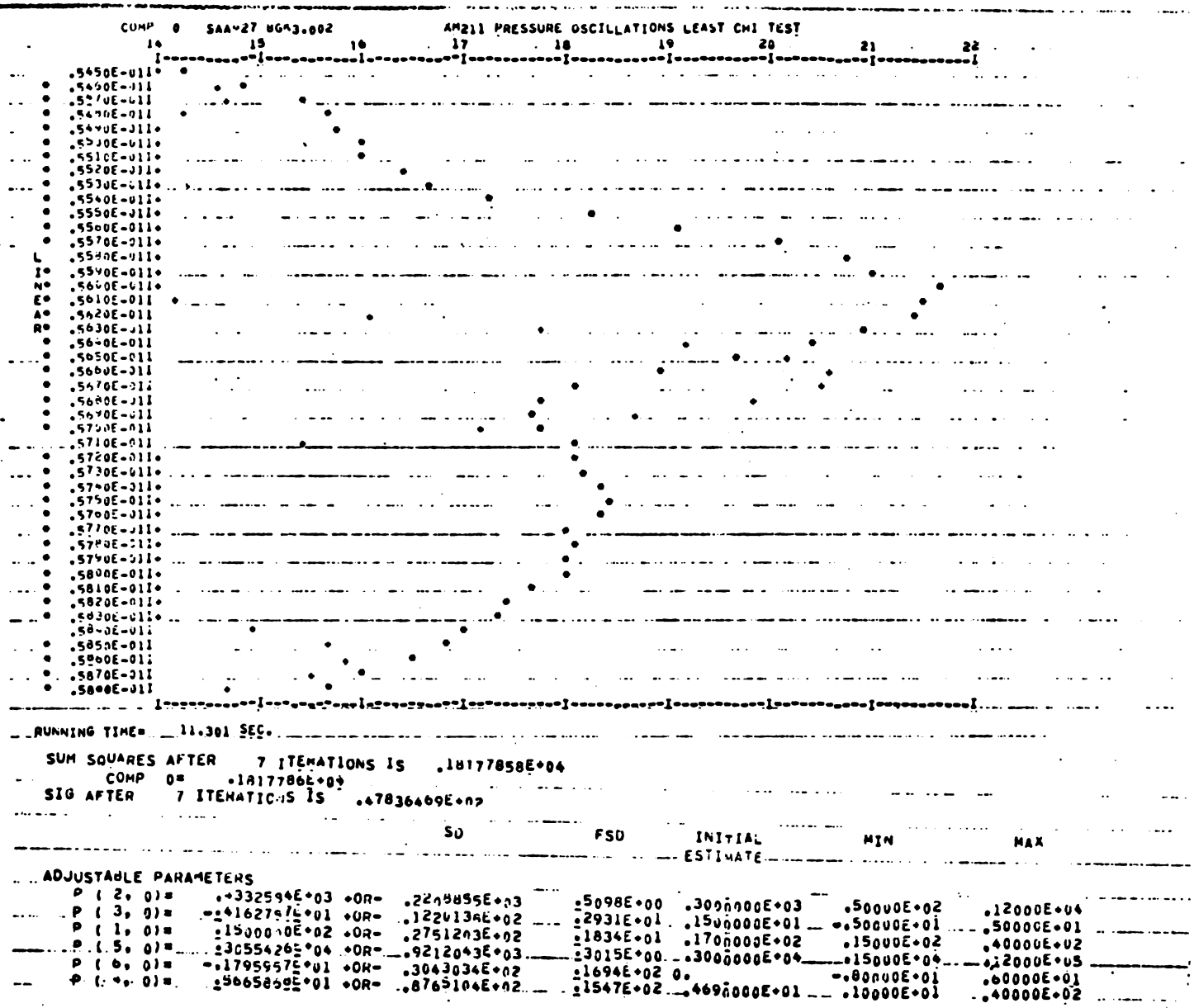


Figure 3 - Fit to Full Data of Pressure Curve of XM211 Propellant Charge by Ordinary Least Squares.

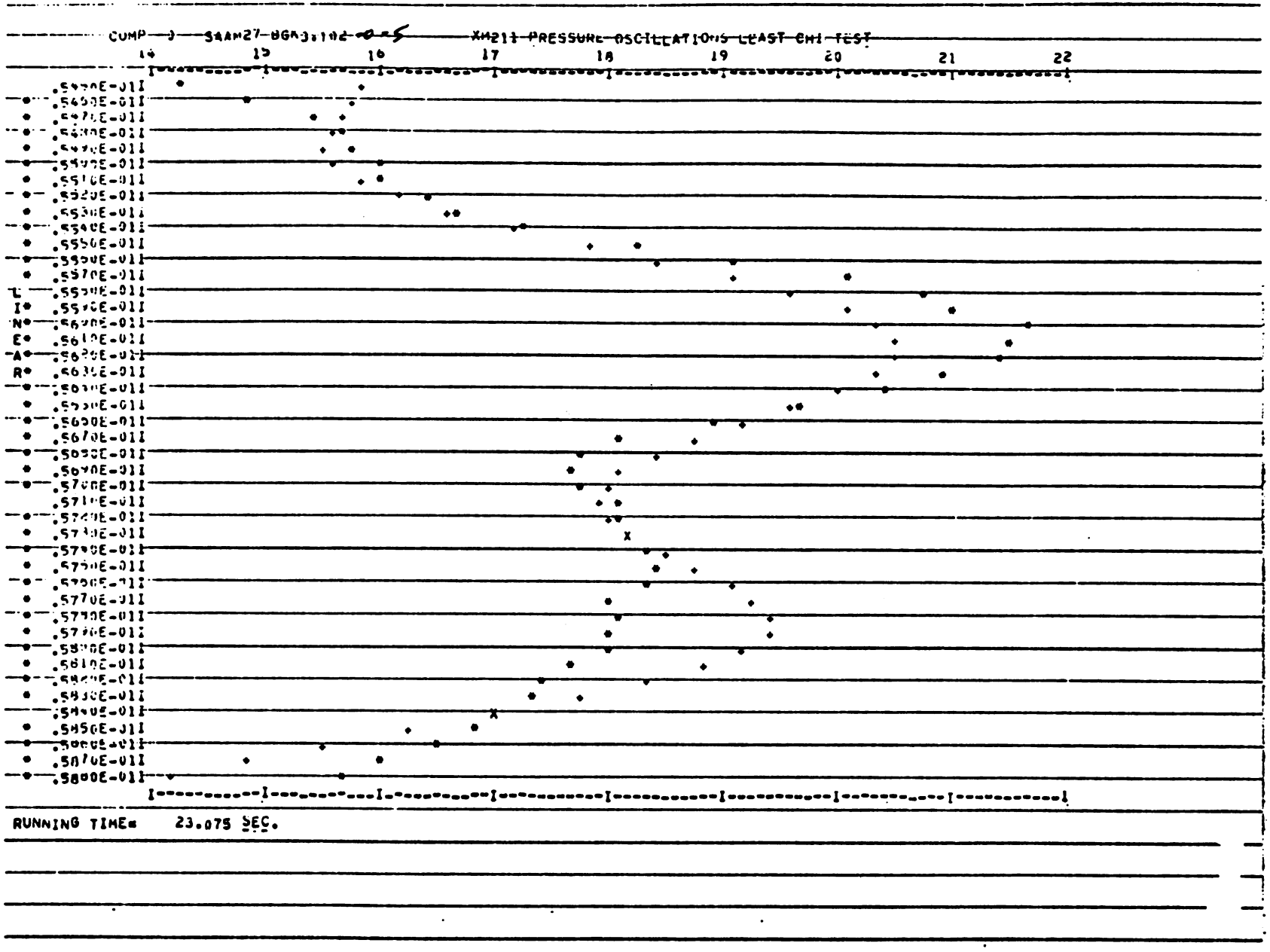


Figure 4 - Fit to Full Data of Pressure Curve of XM211 Propellant Charge by Least Chi Squares.

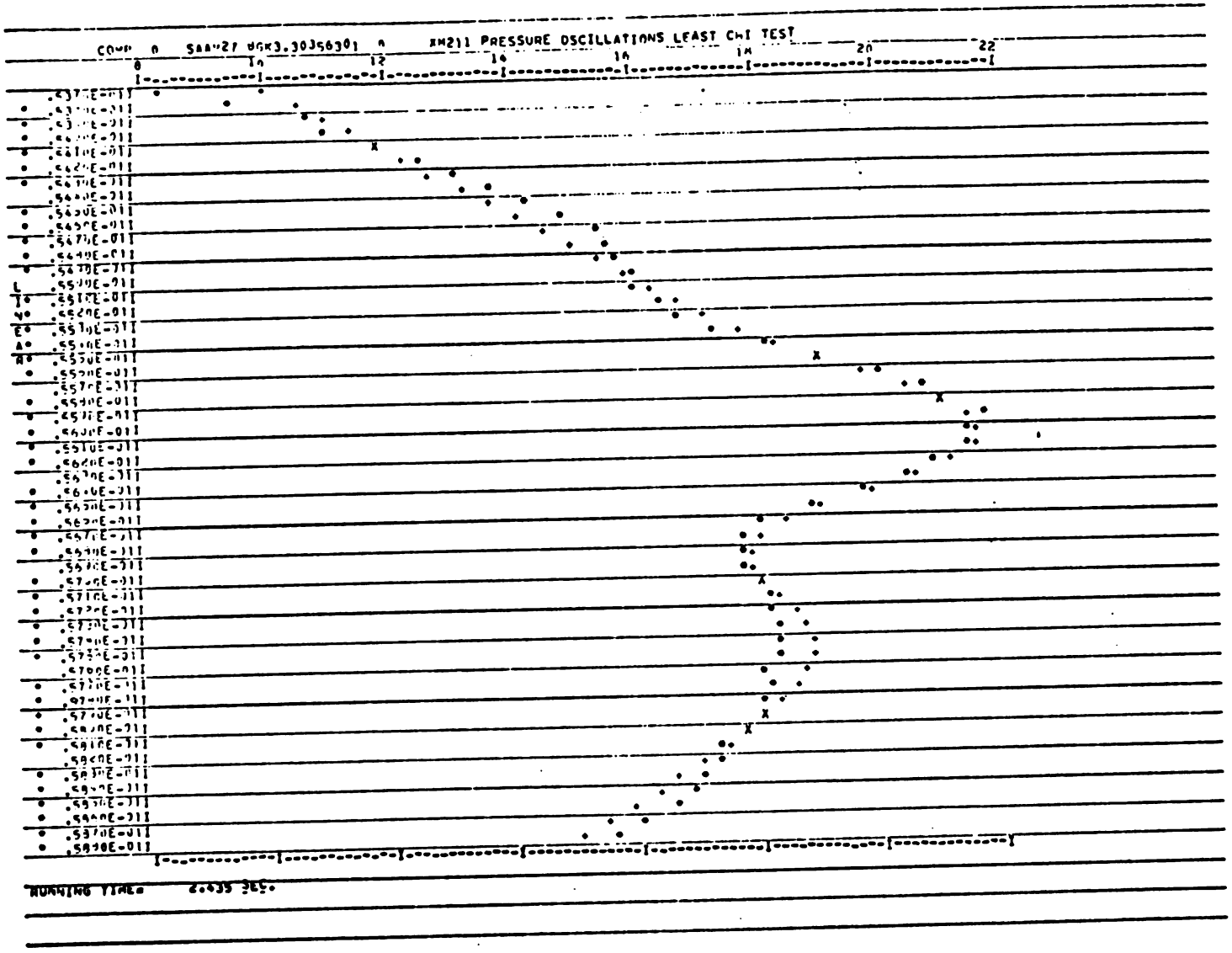


Figure 5 - Pressure as a function of time. "+"s indicate theory for case 301.

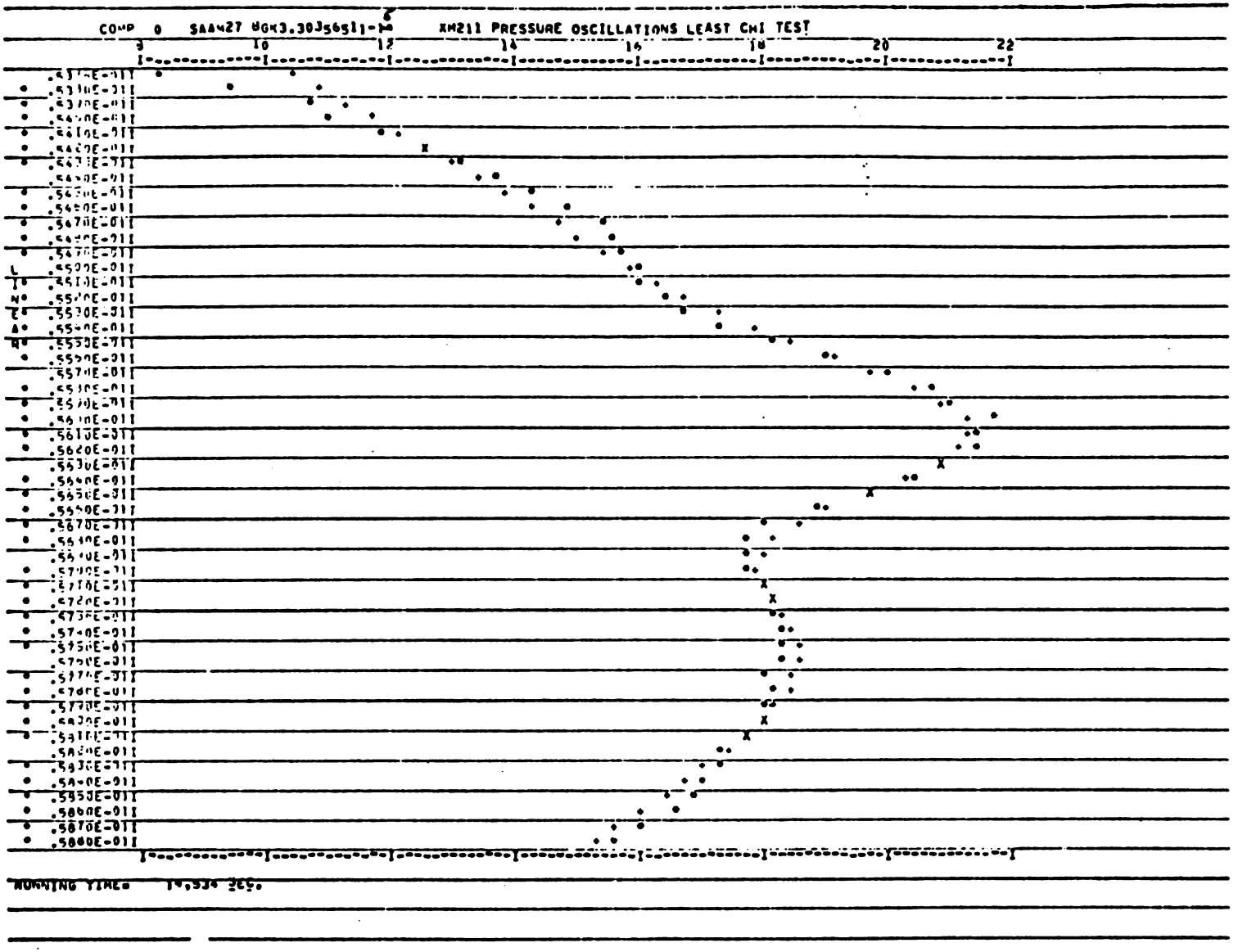


Figure 6 - Pressure as a function of time. "+"s indicate theory for case 511-5

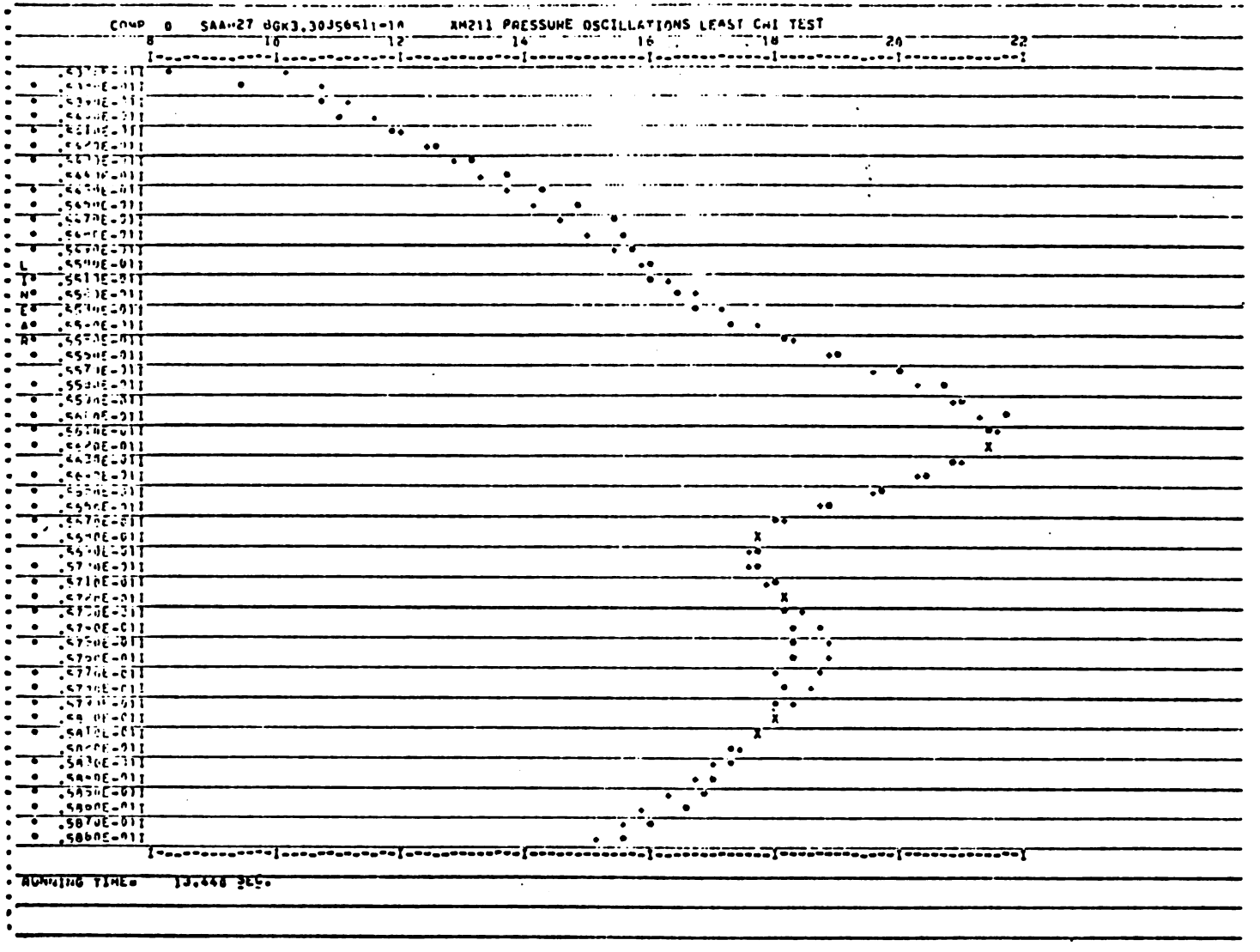


Figure 7 - Pressure as a function of time. "+"s indicate theory for case 511-10

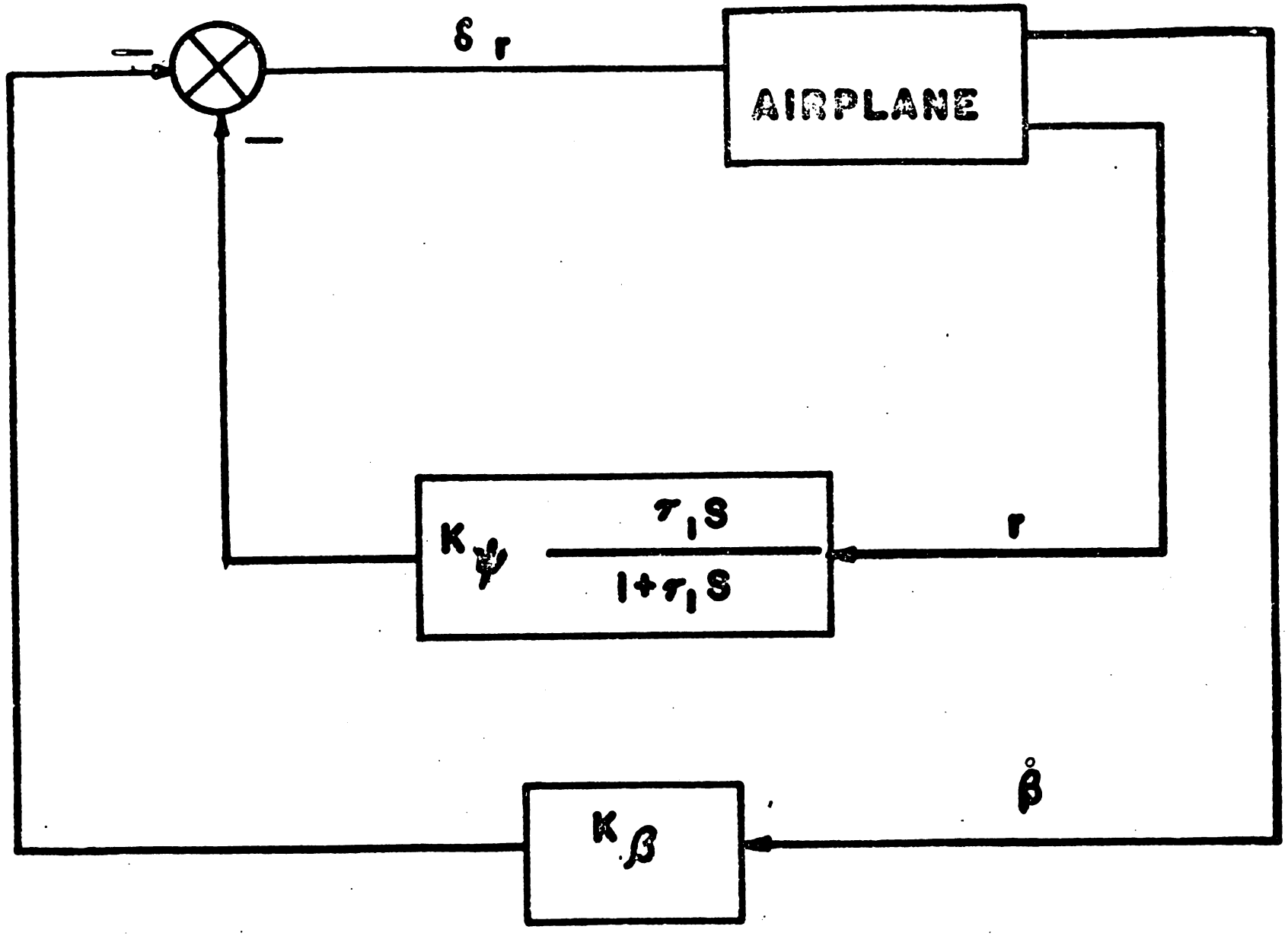


Figure 8 - Yaw Damper Block Diagram

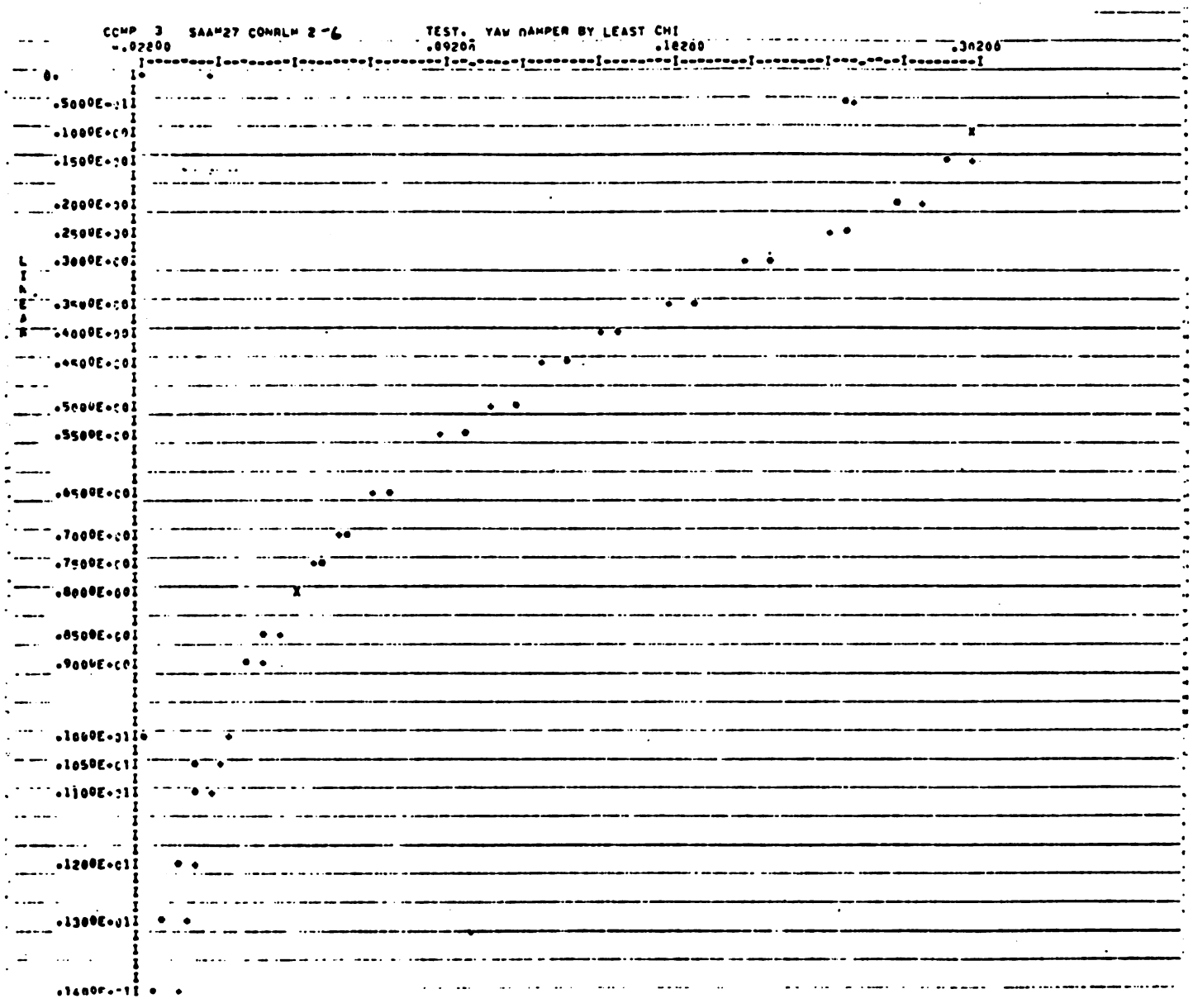


Figure 9 - Yaw Damper Response (Case 2-6)

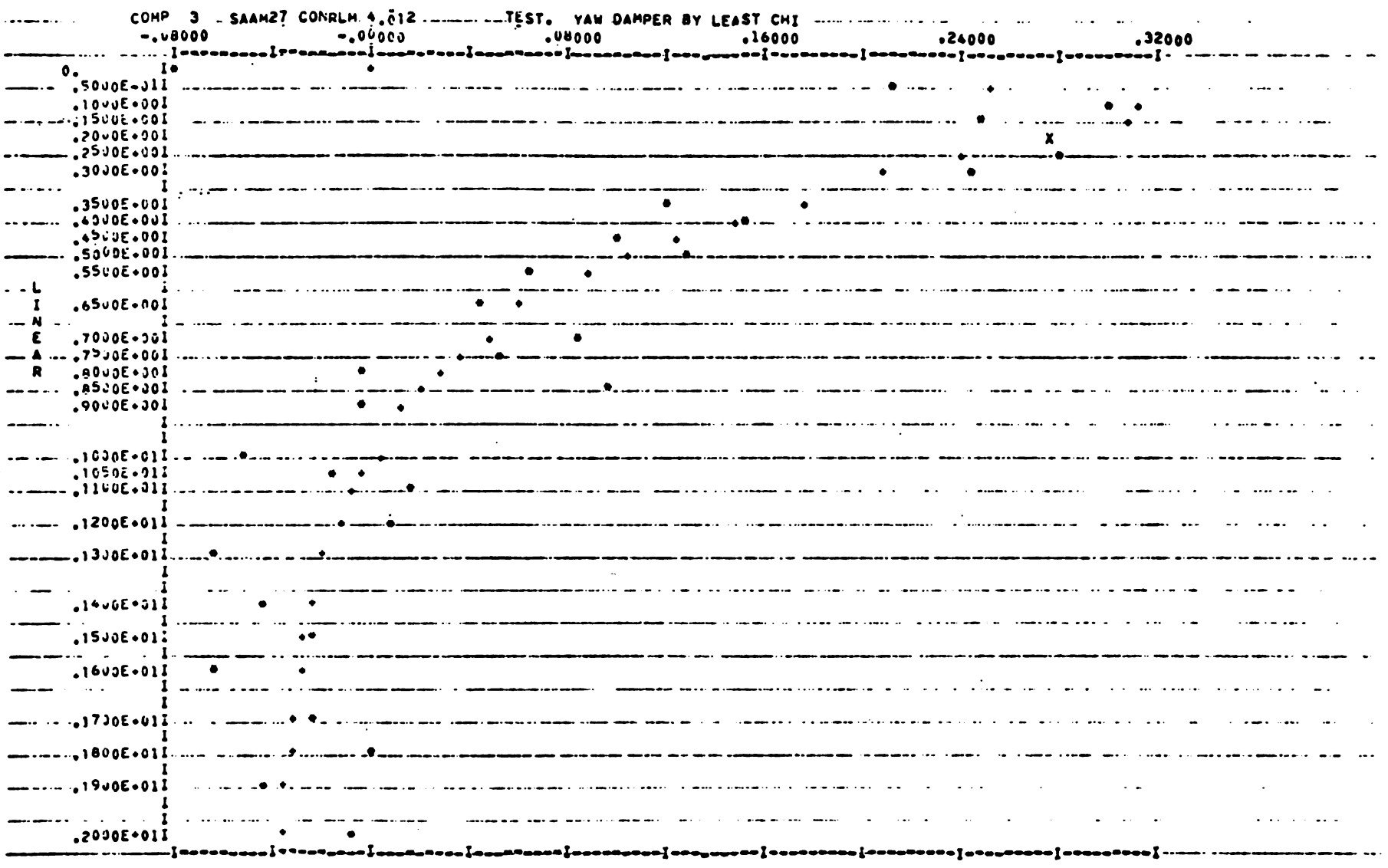


Figure 10 - Yaw Damper Response with Random Errors Added (Case 4.012)

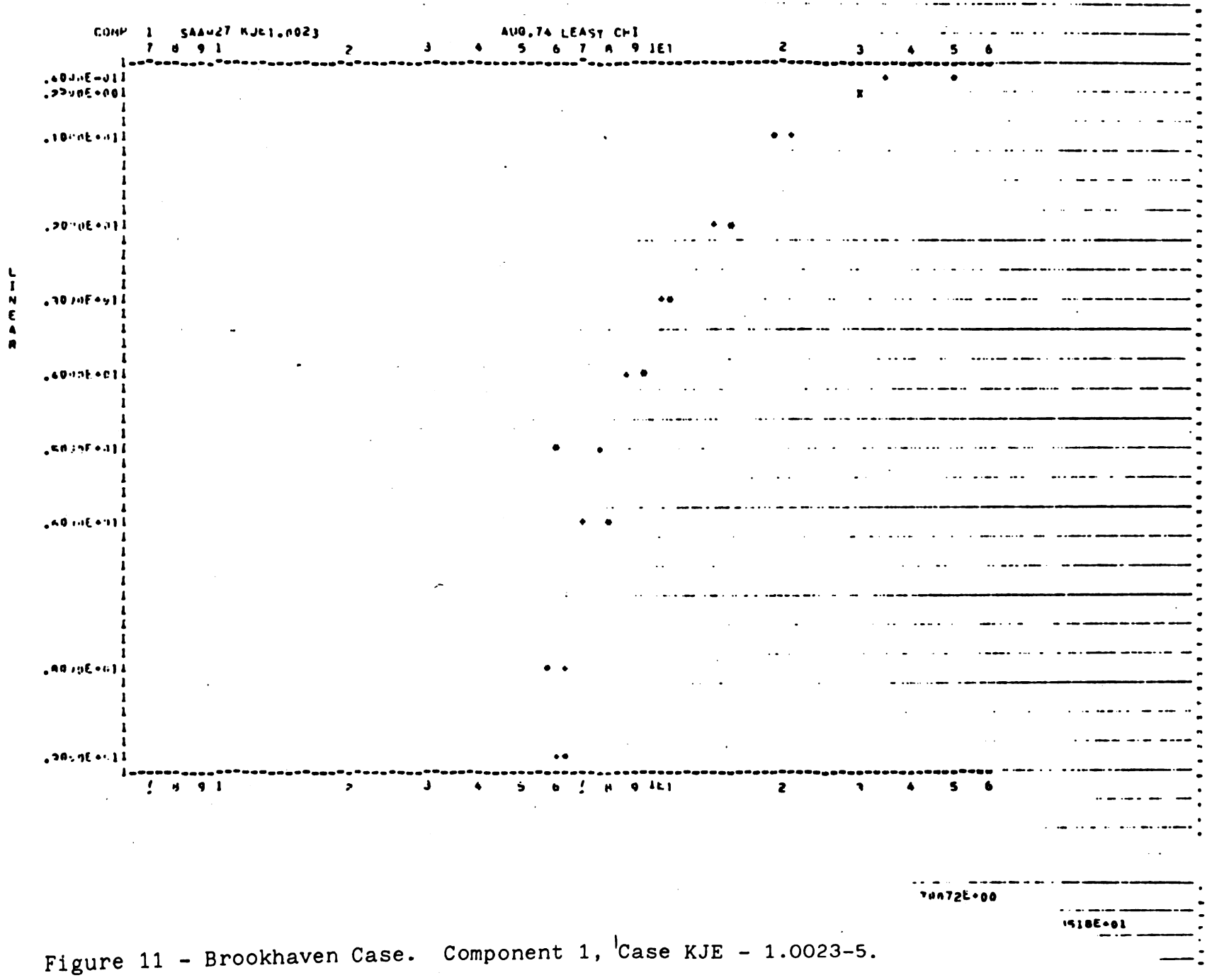


Figure 11 - Brookhaven Case. Component 1, Case KJE - 1.0023-5.

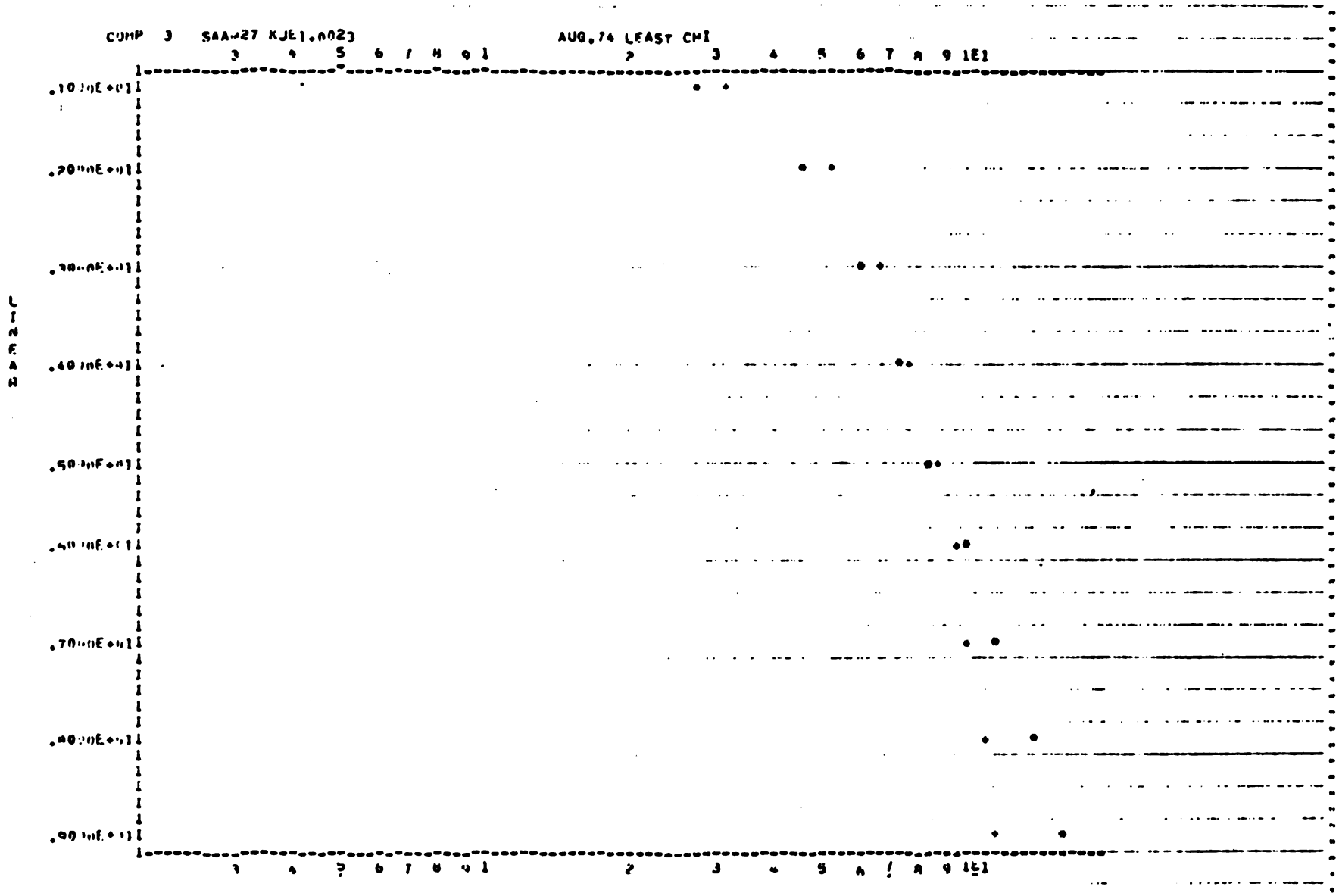


Figure 12 - Brookhaven Case. Component 3, Case KJE 1.0023-5.

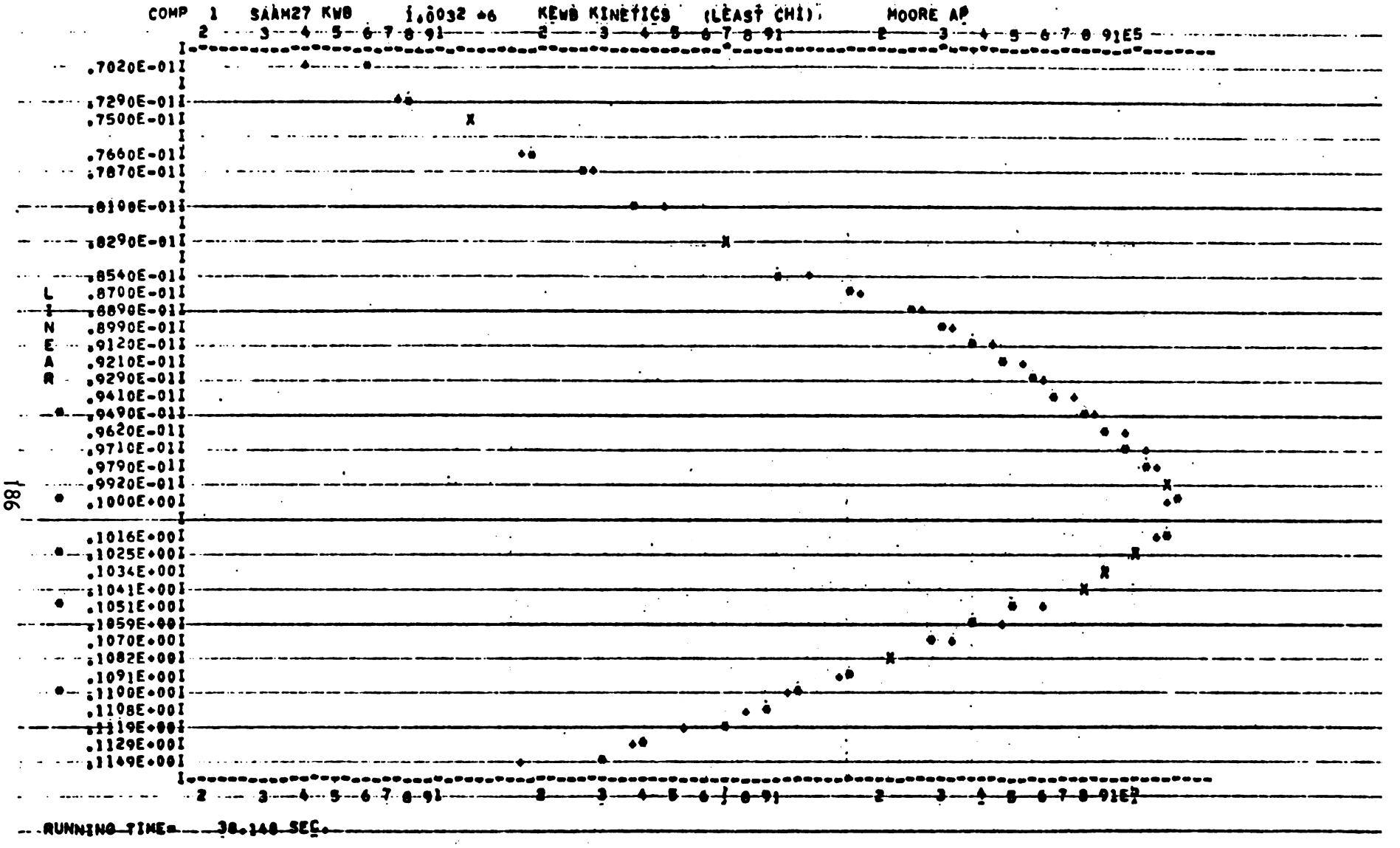


Figure 13 - KEWB Kinetics, Case KWB 1.0032-6, Power as a function of time.

ON THE POWER OF BIRNBAUM'S TEST

Ray E. Schafer
Hughes Aircraft Co., Fullerton, CA 02634

ABSTRACT

Z. W. Birnbaum has proposed a hypothesis test procedure which, under fairly general conditions, does not require explicit knowledge of the critical values of the test statistic. In this paper we investigate the power of the test in a variety of situations. In particular we have considered situations in which the underlying observations have normal and chi-square related distributions. We show that the asymptotic power of this test is identical to the classical test using the same statistic and that the Birnbaum test achieves its asymptotic power very rapidly.

The normal case is considered both for complete and censored samples.

1.0 INTRODUCTION

The classical hypothesis testing problem involves the sampling distribution of the test statistic (say S). For example, to test

H_0: DF (distribution function) is N (μ_o, 1)

versus

H_1: DF is N(μ_1, 1) $\mu_1 > \mu_0$

where $N(\mu, \sigma^2)$ means normal with mean μ and variance σ^2.

the test statistic is $S = \bar{X}$, the sample mean, and the CRITICAL VALUE, say S*, is

$$S^* = \bar{X}^* = \mu_0 + Z_{1-\alpha}/\sqrt{n}.$$ Here $Z_{1-\alpha}$ is the $(1-\alpha)$ quantile of the standard normal distribution and n is the random sample size. The sampling distribution of $S = \bar{X}$ is well tabled so that it is easy to obtain $S^* = \bar{X}^*$ the critical value.

In many situations however the sampling distribution of S is analytically intractable. For example consider the Weibull DF:

$$F_X(x) = 1 - \exp\left[-(x/b)^c\right], \quad b, c, x > 0,$$
$$= 0 \text{ elsewhere}.$$

The sampling of the maximum likelihood estimate of c, say $\hat{c}$, is intractable. However, the sampling distribution of $\hat{c}/c$, while intractable has a distribution free of b and c. Thus, the DF of $\hat{c}/c$ could be obtained (indeed was obtained by Thoman, Bain and Antle, 1969, Technometrics 11, 445-460) by Monte Carlo methods. The Monte Carlo approach is quite expensive; involving some 10,000 to 50,000 $\hat{c}/c$'s for each n.

Z. W. Birnbaum ("Computers and Unconventional Test Statistics," 1974, Reliability and Biometry, Eds.: F. Proschan and R. J. Serfling, SIAM, 441-458 and "Testing for Intervals of Increased Mortability," 1975, Reliability and Fault Free Analysis, Eds.: Richard E. Barlow, Jerry B. Fussell and Nozer, D.

Singpurwalla, SIAM 413-426) has proposed a remarkedly simple procedure in the situation of testing

H_0: DF of S = $G_0(s)$

versus

H_1: DF of S = $G_1(s)$

where G is analytically intractable. Birnbaum's test avoids the expense of a large Monte Carlo simulation. Here we will investigate the power of the Birnbaum test (B.T.) in a variety of situations.

2.0 B. T. DESCRIPTION

We discuss in this section a right (upper) tail hypothesis test. Obvious modifications lead to left-tail and two-tail situations.

Let $F(x, \theta)$ be the distribution function for random variable X and let S_n be a test statistic for θ, based on a sample of size n. Suppose $G_0(s_n)$ and $G_1(s_n)$ are the distribution functions for S_n when $\theta = \theta_0$ and $\theta = \theta_1$ respectively. The B.T. requires that, for all real S_n $G_1(s_n) \leq G_0(s_n)$. and for at least one S_n $G_1(s_n) \neq G_0(s_n)$. Suppose that a random sample of n observations, $x_1, \ldots, x_n$, is available which has been used to calculate a single value of S_n say s_n^*, and we wish to choose between $H_0: \theta = \theta_0$ and $H_1: \theta = \theta_1$ on the basis of this observation.

Birnbaum has shown that, if it is possible to obtain a random sequence of N observations of S_n using only $F(x, \theta_0)$, a hypothesis test may be performed by selecting a number, γ $0 < \gamma < 1$, and observing the number, M, of these N

observations which are greater than s_n^* with rejection of H_0 if and only if $M/N \leq \gamma$. Birnbaum shows that the size of his test is

$$\alpha = (1 + [N\gamma])/(1 + N)$$

where $[N\gamma]$ is the greatest integer $\leq N\gamma$.

In practice, one performs a B.T. of exact size α as follows. Choose $0 < \gamma < 1$, the size desired, and choose an integer N such that $(N + 1)$ is a positive integer. Generate N observations of S_n under H_0 and observe M/N, Reject H_0 if M/N

$$\leq (\alpha(N + 1) - 1)/N = \gamma.$$

3.0 CASES INVESTIGATED

The advantage of the B.T. is that a knowledge of $G_0(s_n)$, the sampling distribution of S_n under H_0 is not required. The N values of S_n may be generated by Monte Carlo methods directly from $F(x, \theta_0)$. The cost of the B.T. is a function of both n and N, and may be substantial (although much cheaper than a "full" Monte Carlo simulation of $G_0(s_n)$) if calculation of s_n must be performed by iterative methods. Hence, it is important to know how the power of the B.T. varies with n and N.

We have investigated the power of the B.T. for test sizes $\alpha = 0.01$, 0.05 and 0.10: for sample sizes n = 5, 10, 20, and 50 and for N = 9, 19, 39, 99, 199 and 499. It should be noted that, for $\alpha = 0.01$, no B.T. exists with N = 9, 19, 39 and for $\alpha = 0.05$, no B.T. exists with N = 9. In fact, no B.T. of size α can be obtained unless $(N+1)^{-1} \leq \alpha$.

We have restricted our investigation to maximum likelihood estimators (or functions thereof) and to the following distributions and parameters.

a) Normal distribution - tests for the mean, θ (known σ), with complete samples, and tests for the mean (unknown σ) with Type II censored samples.

b) The general class of tests for which S_n has a chi-square distribution This includes tests for σ (known or unknown θ) in the normal distribution and tests for the mean of a one-parameter exponential distribution.

4.0 POWER OF THE BIRNBAUM TEST

Intuitively, it is clear that letting $N \to \infty$ is tantamount to obtaining the exact sampling distribution of S_n: hence, the asymptotic power with respect to N should be identical to the classical power based on the same statistic.

The B. T. power under H_1 is

$$1 - \beta(N) = P(M/N \leq \gamma \mid H_1) .$$

$$= \int_{-\infty}^{\infty} \sum_{j=0}^{[N\gamma]} \binom{N}{j} (1 - G_0(s_n))^j \, G_0(s_n)^{N-j} d\, G_1(s_n).$$

$$= \int_0^1 \sum_{j=0}^{[N\gamma]} \binom{N}{j} (1-u)^j \, u^{N-j} d\, J_1(u)$$

where $J_1(u) = G_1(G_0^{-1}(u))$. Birnbaum has proven that

$$\lim_{N \to \infty} (1 - \beta(N)) = 1 - J_1(1 - \gamma).$$

It is easy to show that this is equal to the power of the classical test based on the same test statistic.

5.0 METHOD OF CALCULATIONS

A statistician is seldom interested in knowing the power of a test to even three decimal places. Clearly, the requirement of extreme accuracy in power determination increases the cost of computation. We have chosen to relax the accuracy requirement so that more distributions and sample sizes could be studied.

All of the results were obtained on a CDC Cyber-173 computer. Where power was obtained by numerical integration, the trapezoid rule was used with 200 equal intervals over the domain of integration. Where the limits of integration extended to $\pm\infty$, the heuristic limits used were the 0.0001 and 0.9999 quantiles ($\pm$3.895, for example, with the normal distribution). As a check on the numerical accuracy of the integrations, both "tails" were evaluated. That is, we determined $1 - \beta$ and β separately. In every case, the sum of the two was in the domain (0.995, 1.005).

Where Monte Carlo methods were employed, the random number generating algorithm was the multiplicative congruential method suggested by Knuth using modulo 2^{48} arguments. For the Monte Carlo simulations:

a) If it was necessary to determine classical power by simulation, 10,000 observations of S_n when $\theta=\theta_0$ were obtained and utilized to estimate the 0.90, 0.95, and 0.99 quantiles of the distribution under H_0. Then 10,000 observations of S_n when $\theta=\theta_1$ were generated and compared with these quantiles.

b) To determine the power of the B.T. by simulation, a single observation of S_n ($\theta=\theta_1$) was obtained and compared with N observations of S_n ($\theta=\theta_0$). The number M of these N observations greater than $S_n(\theta=\theta_1)$ was recorded, and if $M/N \leq (\alpha(N+1)-1)/N$, the null hypothesis was rejected. This complete procedure was repeated 2500 times.

It appears that a B.T. using N=199 or greater could substitute for a classical test on the mean of a normal distribution with virtually no loss in power. This may give a practicing statician some confidence in using the B.T. for problems where the distribution of S_n is not obtainable, or obtainable only at great expense.

6.0 NORMAL MEANS (KNOWN σ)

Of course, no one would ever use a B.T. in place of a Neyman-Pearson test for a hypothesis about the mean of a normal distribution with known σ. The p.d.f. of S_n (i.e., the sample mean) under both the null and alternative hypotheses and hence the power, is known analytically. But such an artificial case is valuable for studying the B.T. for precisely this reason. We may observe the relative power of the B.T. in comparison to the classical test as a function of N, n, and the classical power, to get a "feel" for the behavior of the B.T. as a function of sample size.

Let X be $N(\theta, \sigma^2)$. We chose as the null hypothesis N(0, 1) and as alternatives θ=0.1, 0.2, 0.5, 1.0, and 2.0. The power of the B.T. was obtained analytically by numerical integration.

The efficiency of the B.T. Relative to the classical test was remarkably high, even for small N. Over the entire domain of samples sizes, n, the B.T. power was never less than 85% of the classical power. For N = 199 and 499, the B.T. power was never less than 95% of the classical power. As an example, for N = 199, we obtained:

	n = 5		n = 10	
	θ=0.5	θ=1.0	θ=0.5	θ=1.0
B.T. Power	0.29	0.71	0.46	0.93
Classical Power	0.300	0.723	0.475	0.935
	n = 20		n = 50	
B.T. Power	0.71	0.99	0.97	1.000
Classical Power	0.723	0.998	0.971	1.000

7.0 NORMAL MEANS (UNKNOWN σ) WITH CENSORED SAMPLES

The classical power for hypothesis tests on normal means with the standard deviation unknown, but constant, is available for complete random samples through tables of the non-central t-distribution. Here, we examine Type II censored samples, where no such power distributions are available. Given a censored sample.

$$X_{(1)}, X_{(2)}, \ldots, X_{(r)} \qquad r < n.$$

the sample mean, $\bar{x}'$ and standard deviation $s'^2 = \sum (x_{(i)} - \bar{x}')^2/r$ are calculated. An auxiliary function λ is needed. The value of λ depends only on r/n and on $s'^2/(\bar{x}' - x_{(r)})^2$. The M. L. E. for θ is $\bar{x}' - \lambda(\bar{x}' - x_{(r)})$.

Hypothesis tests of the form $H_0: \theta = \theta_0$ vs. $H_1: \theta > \theta_0$ were examined for the cases $(\theta_1 - \theta_0)/\sigma = 0.1, 0.2, 0.5, 1.0$ with sample sizes 5, 10, 20, 50 and censoring at $0.8n$ for $n = 5, 10, 20$ and at $0.4n$ for $n = 10, 20, 50$. Both the classical power and B.T. power were obtained by Monte Carlo simulation, as described previously.

The B.T. with N = 199 or 499 will provide essentially equivalent to that of the classical test for all of the parameters and sample sizes examined.

8.0 THE CHI-SQUARE CASE

Many hypothesis testing situations involve test statistics which have a χ^2 distribution. e.g., tests on the variance of a normal distribution (known or unknown mean) and tests on the mean of an exponential distribution. The power of the classical χ^2 test is available in the literature from tables of the non-central χ^2 distribution. But we have explored this case for the same reason that the normal distribution was examined – the B.T. power may be obtained analytically and its behavior with respect to N may lend credence to the assertion that the B.T. is essentially as powerful as the classical test for a variety of probability distributions.

Many hypothesis tests involving the χ^2 distribution are equivalent to

$$H_0: S_n \xrightarrow{d} \chi^2_m$$

$$H_1: aS_n \xrightarrow{d} \chi^2_m$$

where S_n is the test statistic, m is degrees of freedom and $0<a<1$ for a right-tail test. $a > 1$ for a left-tail test. We examined right-tail tests for m = 5, 10, 20 and for $a = 2/3, 1/2, 1/3, 1/4$. $\alpha = 0.10, 0.05, 0.01$.

The B.T. power was remarkably close to the classical power, even for N = 9. In fact, the smallest ratio of B.T. power to classical power for all the combinations investigated was 0.87. A sample of results for N = 199 and $\alpha = 0.05$ is:

	m = 5				m = 10			
a	1/4	1/3	1/2	2/3	1/4	1/3	1/2	2/3
B.T. Power	0.72	0.59	0.34	0.19	0.91	0.80	0.51	0.27
Classical Power	0.736	0.595	0.354	0.194	0.918	0.807	0.518	0.272

	m = 20			
a	1/4	1/3	1/2	2/3
B.T. Power	0.99	0.96	0.73	0.39
Classical Power	0.993	0.960	0.735	0.401

Hence, the χ^2 data support earlier conclusions that the B.T. with a reasonable value for N, say 199, is essentially as powerful as a classical test.

9.0 CONCLUSIONS

We have investigated the power of the B.T. with respect to the power of the corresponding classical test in a variety of situations. These situations included complete and Type II censored samples for the commonly used test sizes and frequently used sample sizes.

It seems clear that the B.T. offers cost savings when the sampling distribution of the test statistic is unknown and must be obtained by expensive methods.

In addition to the fact that the asymptotic power of the B.T., as $N \to \infty$, is equal to the power of the classical test based on the same statistic, the B.T. has some interesting characteristics. In all cases the relative power of the B.T. was quite large even for N as small as 9 and generally for $N \geq 199$ the power was 95% of the asymptotic power or greater. Also, generally, the relative power of the B.T. increased as the alternate hypothesis got further away from the null hypothesis. Finally the B.T. relative power increased with N.

ERROR-TIME RESPONSE PERFORMANCE OF NAIVE SUBJECTS

Michael Hacskaylo and Joseph E. Swistak
Night Vision and Electro-Optics Laboratory
Fort Belvoir, Virginia 22060

Abstract. An experiment was conducted in which the error and time response performance for designating the angular location of a single flash of light on a circular pattern of lights were measured. Fourteen naive subjects were instructed to record as accurately and rapidly as possible the angular position of an activated light. They were allowed only one attempt for each of six consecutive trials. The data are presented in terms of mean time of each response per trial and mean error per trial. The mean error, as a function of mean time, appears to be bounded by an error-time response equation; $E = -20.88 \log(t/15.37)$, where E is the mean angular error in degrees and t is the mean time in seconds. Surprisingly, the subjects responses as measured in either time or error did not follow classical reaction time or learning patterns. That is, while time of response remained fairly consistent from trial to trial, the lowest error occurred on the first trial while maximum error consistently occurred on the fourth trial. Based upon the six trial limit used in the experiment, it is believed that the naive subjects, first trial performance is the best for designating the angular location of a single flash of light.

Introduction. The philosophy which tank crewmen have always adopted has been "make your first shot count because you may not get a second chance." This philosophy has become more acute with the recent advent of "SMART" weapons which ride beams of light to a target. A system was designed which would allow tank crewmen to detect and radially demarcate the source of designation by a coherent light source. The system is designed to operate by having a tank crewman observe a circular array of lights on a panel. When the tank is illuminated by a laser beam, a corresponding azimuthal light is activated. The crewman would interpret and record the azimuthal position for appropriate tank action. The effectiveness of the crew would depend upon (1) the speed and accuracy with which the azimuth is read out, and (2) the panel configuration used to display the azimuthal information. The panel used in this experiment was designed from a technical consideration based on the circular representation of equally spaced light bulbs (Fitts and Seeger, 1953).

Method.

Subjects. Fourteen U. S. Army enlisted men of various ranks were randomly selected from a large group of individuals to serve as subjects. None had prior training in tanks or tank related equipment and none had prior experience with the display panel being tested. The fourteen subjects were then randomly assigned to one of two groups comprised of seven subjects each.

Apparatus. The stimulus display panel consisted of a 9cm diameter ring of 36 equally spaced light bulbs. This panel, portrayed in Figure 1, was positioned on the front panel of a box 20cm long, 10cm high and 5cm deep. The light bulbs were angularly marked from zero to 360 in ten degree increments in a clockwise direction with zero at the top.

Responses were recorded on a response panel. This consisted of a 12cm circle drawn on a 20x25cm sheet of plain paper. The circle was divided into quadrants and marked into degrees as follows: Zero degrees (0°) was marked at the top. In a clockwise direction, each quadrant was successively marked 90°, 180°, 270°, and again at the top, 360°. A pencil was used for marking angular positions with an "X" on the circle.

Procedure. Each subject was briefed individually prior to his participation in the experiment. They were brought into a room which contained the stimulus display panel, a bench, chair and associated equipment required to activate the lights of the panel. Each subject was briefed as follows:

"As accurately and as rapidly as possible, determine the angular location of a light when it comes on and mark with an "X", that position on the circle on the sheet of paper in front of you. The sheet of paper was referred to as the response panel for purposes of the study. Each subject was allowed two familarization trials to be sure they understood the instructions. Each subject was then given six trials. The sequence of lights for trials one to six are presented in Figure 2.

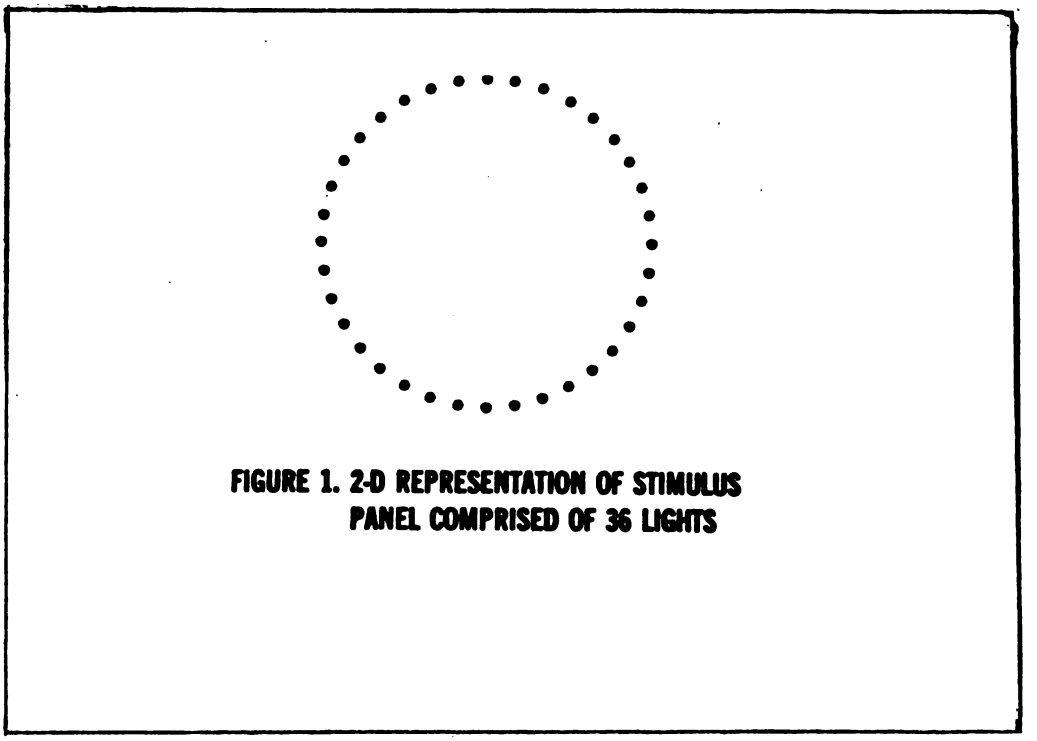


FIGURE 1. 2-D REPRESENTATION OF STIMULUS PANEL COMPRISED OF 36 LIGHTS

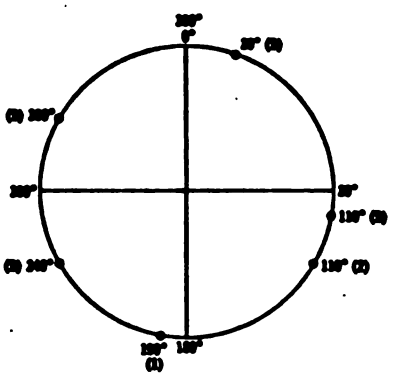


FIGURE 2. SAMPLE RESPONSE PANEL SHOWING THE ORDER IN WHICH THE LIGHTS WERE FLASHED AND THE ASSOCIATED DEGREES OF ARC FROM THE ORIGIN.

Each subject saw this same sequence. A new response panel was supplied for each trial. The time interval from when the light came on to when the subject marked the panel was measured by a stop watch to 0.01 seconds. The stop watch was controlled by the experimenter and it was assumed that the reaction time error introduced was fairly constant.

Upon completion of a set of six trials, the subject was dismissed. The subjects tested versus those not tested were kept in separate rooms until all seven subjects in a group were finished. One group (A) of seven subjects was tested on one day, the other group (B) of seven subjects was tested on the following day.

The angular positions marked on the response panels were scored in degrees by using a transparent template graduated to 0.5 degrees and superimposed on the marked response panel. The accuracy of the marked position was then measured to $\pm$ 0.5 degrees which was the resolution of the scoring template.

Results and Discussion.

The mean time of response for each trial are presented in Figure 3.

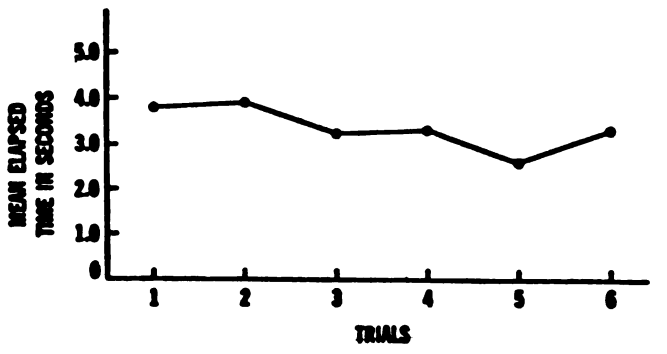


FIGURE 3. MEAN RESPONSE TIME TO DESIGNATE ANGULAR LOCATION OF A FLASH OF LIGHT ON SIX CONSECUTIVE TRIALS.

These mean times and associated standard deviations are presented in Table 1. In general, it appears that the mean time to respond did not vary significantly over the six trials. There does appear to be a gradual decrease in response time from trial one to trial five, but an increase on trial six. There was no readily obvious reason for this increase on the sixth trial, i.e., no subject took an inordinate amount of time It would also seem that fatique could not be a factor with only six trials having elapsed.

TRIAL NO.	1	2	3	4	5	6
MEAN TIME (SEC)	3.79	3.94	3.21	3.34	2.60	3.32
STANDARD DEVIATION	1.41	2.03	1.31	1.49	1.00	1.23

TABLE 1. MEAN TIME AND RELATED STANDARD DEVIATION REQUIRED TO DETECT AND DESIGNATE THE ANGULAR LOCATION OF A LIGHT FLASHED WITHIN A 360° ARRAY OF LIGHTS ON SIX CONSECUTIVE TRIALS.

The mean error in degrees for each trial are presented in Figure 4. The numerical values and associated standard deviations are presented in Table 2. Trial one had the smallest angular error. The amount of error

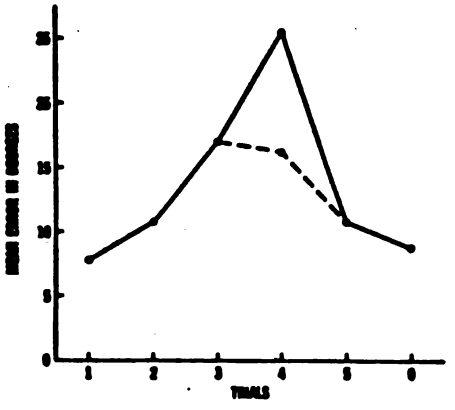


FIGURE 4. MEAN ERROR IN ESTIMATES OF ANGULAR LOCATION OF A FLASH OF LIGHT ON SIX CONSECUTIVE TRIALS.

then increased with the greatest error occurring on the fourth trial, after which, the error decreased. The error on the sixth trial was very near the error on the first trial. A large part of the error on trial four can be attributed to the performance of two subjects who had

TRIAL NO.	1	2	3	4	5	6
MEAN ERROR (DEGREES)	7.8	11.9	16.9	25.5	10.8	8.6
STANDARD DEVIATION	6.2	6.6	9.4	26.5	10.2	6.6

TABLE 2. MEAN ERROR AND RELATED STANDARD DEVIATION ASSOCIATED WITH THE ANGULAR DETECTION AND DESIGNATION OF A LIGHT FLASHED WITHIN A 360° ARRAY OF LIGHTS ON SIX CONSECUTIVE TRIALS.

errors of 94.5 and 68.0 degrees, on that trial. However, even with these two values removed from the data, the mean degrees of error for trial four remains at 16.2. If this level of error is the more accurate, then it can be said that the third and fourth trials were the worst in terms of performance, and the dotted portion of Figure 4 would more aptly represent the performance on this task. The mean error, as a function of mean time, (Figure 5) appears to be bounded by an error-time response equation: $E = -20.88 \log(t/15.37)$, where E is the mean time in seconds. This curve provides somewhat of an upward estimate of angular error given an elapsed period of time for a response - the greater the time, the lower the error.

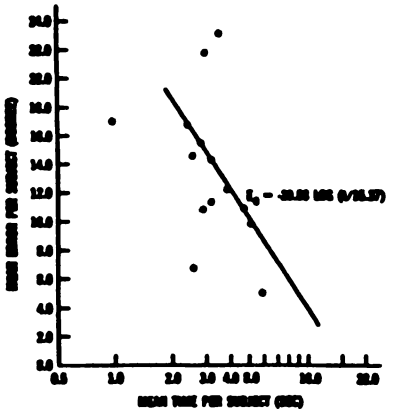


FIGURE 5. CURVE DELINEATING THE UPPER BOUNDS OF THE ERROR-TIME RESPONSE.

However, regardless of which curve in Figure 4 most accurately represent the performance which could be expected on the task described in this study, it is interesting to note that neither set of data follows the expected learning pattern described in learning literature. Normally, one could expect accuracy to be poorest on the first trial, rapidly improve on the next few trials and then continue to improve at a slower rate until some maximum level was reached. The number of trials required for asymptotic performance to occur would depend upon the degree of difficulty of the task. The present task should have required 6-8 trials. It appears that asymptotic performance was being approached on the sixth trial, but what is truly interesting is that the performance on the first trial was actually better than on the sixth.

In terms of an untrained subject being able to determine angular direction of designation, the first attempt he made would be the most accurate of his initial six attempts.

The conclusion of this study must be that the initial attempt by an untrained gunner would be at least as accurate as one who is starting to asymptote.

References

Fitts, P. M. and Seeger, C. M., S-R Compatibility: Spatial Characteristics of Stimulus and Response Codes, J. Experimental Psychology, 1953, 46, 199-210.

IMAGE INTERPRETATION PERFORMANCE ON FOUR STANDARD TYPES OF AEROGRAPHIC FILM

RONALD L. JOHNSON and PAUL J. SCHOOL

US Army Mobility Equipment Research and Development Command, Ft. Belvoir, Virginia

ABSTRACT

This study involved trained operational image interpreters who analyzed highly controlled aerial imagery from which the effects of type of film upon target detection were determined. One-hundred-and-one operational image interpreters generated the following mean target detection probabilities: Color Infrared - 58.6%, Color - 55.4%, Panchromatic - 44.7%, and Black and White Infrared - 43.4%. At the 0.05 significance level, target detections were affected by film type as follows: Color Infrared differed significantly from both Panchromatic and Black and White Infrared films. Color differed from Black and White Infrared. The combined mean of target detection for Color Infrared and Color differed significantly (0.01 level) from the combined mean for Panchromatic and Black and White Infrared. Therefore, use of Color and Color Infrared imagery results is significantly more accurate day image interpretation.

1.0 INTRODUCTION

The purpose of this study was to quantitatively determine the target acquisition capabilities by image interpreters (II) of the following film types: Aerochrome MS 2448 (Color), Aerochrome Infrared 2443 (Color Infrared), Plus X Panchromatic 2042 (Black and White), and Infrared Aerographic 2424 (Black and White Infrared). Image interpretation is defined 1/ as the examination of images of objects on film for the purpose of identifying the objects and deducing their significance.

Approximately 90 percent of the intelligence gathered in World War II was derived from aerial photography. The requirement for accurate imagery intelligence is escalating as weapon systems and tactics develop and become more refined. To obtain this information, it is becoming increasingly more common to use color, color infrared, and black and white infrared film. Strandberg 2/ states " color aerial photography offers much promise in the gathering of imagery intelligence, because humans have the capability of discriminating between an almost infinite number of different colors, but at most, only a few hundred different shades of gray".

2.0 TEST SITE

An 820 acre site was selected. This site is used for equipment evaluation by the US Army. Military equipment and camouflage devices such as nets were randomly located throughout the study area. The soil contained a high moisture content and the color was reddish-tan. The brush was gray and brown in color. Included within this site were building complexes, open fields, dense woods, and clumps of green grass. The forest composition was a mixture of oak and pine.

3.0 TEST IMAGERY

Photographic images consisted of a 13 frame series of 9" X 9" positives taken with 60% forward overlap. One frame series was acquired for each of the four standard types of aerial film. Film and filter characteristics are summarized in Table 1.

TABLE 1

FILM CHARACTERISTICS

FILM TYPE	RESOLVING POWER (T.O.C.)* 1000:1	1.6:1	SPECTRAL SENSITIVITY RANGE (NM)	FILTER CUT-OFF	RESULTANT SPECTRAL SENSITIVTY (NM)
Plus X Panchromatic 2042 (Black & White)	100	50	250-700	Zeiss Yellow 490 NM	490-700
Infrared Aerographic 2424 (Black & White)	80	40	400-900	Zeiss Orange 550 NM	550-900
Aerochrome MS 2448 (Color)	80	40	400-700	Zeiss Clear Activig No cut off	400-700
Aerochrome Infrared 2443 (Color)	63	32	400-900	Zeiss Orange 550-NM	550-900

*Target Object Contract

A KC-4B camera system with a 6 inch focal length lens was used. In all cases except black and white infrared film, standard film/filter combinations were employed. A zeiss orange filter was used with the black and white infrared film instead of a red filter; to increase the spectral response, and therefore, the information content of this film type. All imagery was gathered during four overflights (one per film type) at an altitude of 1500 feet above ground between the hours of 1100 and 1400. Therefore, the sun angle effect was negligible. The photographs were taken in February, and the weather was clear and sunny. Each 9 inch photograph covered a land area of approximately 124 acres. The total number of targets present or detected for each strip of imagery were determined by three senior image interpreters. They performed detailed and exhaustive analysis upon the imagery. The results of which are presented in Table 2.

TABLE 2

TOTAL MILITARY TARGETS DETECTED FOR EACH TYPE OF FILM

FILM TYPE	NUMBER OF TARGETS
Plus X Panchromatic 2042 (Black and White)	46
Infrared Aerographic 2424 (Black and White)	50
Aeochrome MS 2448 (Color)	44
Aerochrome Infrared 2443 (Color)	47

Variations in the number of targets detected between film types (ground truth) are not significant. They were apparently due to slightly different flight lines flown by the photgraphic aircraft.

4.0 METHODS OF PROCEDURE

The Pseudo-Isochromatic Plates for Testing Color Perception, developed by the American Optical Corporation, were given to each II in order to insure that the interpreters were not color deficient. A total of 101 operational US Marine Corps image interpreters participated in this study. The participants were randomly divided into four groups, one group for each type of film. The assumption was made that the four groups, due to the Central Limit Theorem, contained interpreters of equal ability. Each II was instructed to perform detailed image analysis to detect military targets such as jeeps, trucks, etc., and was allotted 45 minutes to analyze a selected film strip. Each II viewed only one strip of film, and consequently only one film type.

5.0 RESULTS

The percentage of military targets detected by the image interpreters for each of the four types of film was calculated. These data, along with the associated standard deviations, 95% confidence intervals, and sample sizes are presented in Table 3.

TABLE 3

STATISTICAL SUMMARY FOR MEAN PERCENTAGE OF TARGETS DETECTED BY FILM TYPE

FILM TYPE	MEAN PERCENTAGE OF TARGETS DETECTED	STANDARD DEVIATION	95% CONFIDENCE LEVEL		SAMPLE SIZE (# of observers)
			Lower	Upper	
Plus X Panchromatic 2042 (Black & White)	44.7	16.2	38.1	51.2	26
Infrared Aerographic 2424 (Black & White)	43.4	18.4	35.9	50.8	26
Aerochrome MS 2448 (Color)	55.4	22.2	46.1	64.8	24
Aerochrome Infrared 2443 (Color)	58.6	17.6	51.4	65.9	25

An analysis-of-variance 4/ of the mean values shown in Table 3 was performed and the results are presented in Table 4.

TABLE 4

ONE WAY ANALYSIS-OF-VARIANCE FOR MEAN PERCENTAGE OF TARGETS DETECTED FROM FOUR TYPES OF AERIAL FILM

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F-RATIO
Types of Aerial Film	4,428.8864	3	1,476.2955	*4.2351
Within Types of Aerial Film	33,812.6291	97	348.5838	
Total	38,412.5125	100		

Critical F
* 0.05, 3, 97 = 3.27

The data presented in Table 4 revealed significant effects between the mean percentages of targets detected and the type of aerial film. The degree of this relationship was determined by individual comparison employing the t statistic. These results are presented in Table 5.

TABLE 5

INDIVIDUAL COMPARISONS UPON THE MEAN PERCENTAGES OF TARGETS DETECTED

	Plus X Panchro-Matic 2042 B/W	Degrees of Freedom	Infrared Aerographic 2424 B/W	Degrees of Freedom	Aerochrome MS 2448 Color	Degrees of Freedom	Aerochrome Infrared 2443 Color	Degree of Freedo
Plus X Panchro-matic 2042 B/W								
Infrared Aerographic 2424 B/W	0.279	51						
Aerochrome MS 2448 Color	1.938	43	2.0085*	46				
Aerochrome Infrared 2443 Color	2.932*	50	3.028*	50	0.553	45		

Critical value for significance: $t_{0.05,43} = 2.017$; $t_{0.05,45} = 2.15$; $t_{0.05,46} = 2.014$; $t_{0.05,50} = 2.010$; $t_{0.05,51} = 2.009$.
* Indicates significance $\leq$ 0.05 level.

The data in Table 5 indicates that at the significance level $\alpha = 0.05$, the mean detection of targets on the Color and Color Infrared film, was greater when compared to the mean detection of targets in the Black and White Infrared film type. The mean detection of targets by image interpreters on color infrared film was also significantly greater (≤ 0.05) when compared with the mean detection of targets on Panchromatic film. An even more significant relationship ($\alpha = 0.01$) was found when the means of the Color and Color Infrared films were combined and compared with the combined means of the Panchromatic and Black and White Infrared. Table 6 contains the means for the number of targets detected, as well as the 95% confidence intervals and sample size. Table 7 contains the results of the analysis-of-variance performed on the data of Table 6.

6.0 DISCUSSION

From these results, and assuming the interpreters to be of equal experience levels, it was statistically ($\alpha = 0.05$) determined that the use of Aerochrome Infrared 2443 film resulted in a greater mean percentage of target detections than that of either Plus X Panchromatic 2042 or Infrared Aerographic 2424. Aerochrome MS 2448 film also allowed statistically ($\alpha = 0.05$) greater mean number of targets detected than that of Infrared Aerographic 2424. The mean number of targets detected was combined for the two color films and also for the two black and white films. They were then statistically compared against each other; it was determined that they differed at the 0.01 level. The task involved in this study was basically one of searching an unknown area of film for the detection of military targets, some of which were embeded in trees. The resulting mean percentage for target detection on color films complements some of the conclusions of a US Naval Technical Bulletin 5/ which states that color photograph provides the most benefits in a area being searched for unknown or unlocated targets. The bulletin also states that the detecting of partially hidden targets is aided by the use of color imagery which provides details within the shadows. Strandberg stated that atmospheric haze reduces the advantages of color film over black and white when high obliques or horizon-to-horizon panoramics are taken. Therefore, both color and black and white imagery may be required. It is interesting to note that 64% of the image interpreters who analyzed the Aerochrome Infrared 2443 film stated that, with the exception of a brief session in school, they have not had further experience with it. Forty-two percent of the image interpreters made a similar statement concerning the use of color film. Accordingly, it may be that given additional experience with these films, the mean percentage of targets detected would show an even greater disparity between color and black and white aerial film than the results

TABLE 6

STATISTICAL SUMMARY FOR PERCENTAGE OF TARGETS DETECTED FROM THE COMBINED MEANS OF COLOR AND BLACK AND WHITE AERIAL FILM

Film Type	Mean Percentage of Target Detected	95% Confidence Interval Lower	Upper	Standard Deviation	Sample Size
Combined Color	57.1	50.5	63.6	19.8	49
Combined Black & White	44.0	38.5	49.5	17.2	52

Table 7, below, contains the results of the analysis-of-variance performed in the data of Table 6.

TABLE 7

ONE WAY ANALYSIS-OF-VARIANCE FOR PERCENTAGE OF TARGETS DETECTED FROM THE COMBINED MEANS OF COLOR AND BLACK AND WHITE AERIAL FILM

Source	Sums of Squares	Degrees of Freedom	Mean Square	
Between Combined Color and Combined Black and White Film	4,282.0657	1	4282.0657	12.4833*
Within Types of Aerial Film	33,959.4467	99	343.0247	
Total	38,241.5125	100		

Critical Value: $F_{.01,1,99} = 8.29$

* indicates significance $\leq$ 0.01 level

obtained from this study. Therefore, the quality of day imagery intelligency should increase with the increased use of Aerochrome MS 2448 and Aerochrome Infrared 2443 film by image interpreters. The use of Infrared Aerochrome 2424 film for gathering of day imagery intelligence is not justified by the results of this study. Should Aerochrome Infrared 2443 not be available, or night photography using artificial light sources is desired, Infrared Aerographic 2424 film may be of value.

7.0 SUMMARY

We quantitatively compared the target detection capabilities of 101, US Marine trained, operational II's. They analyzed the same targets photographed with the following four types of film:

Aerochrome MS 2448
Aerochrome Infrared 2443
Infrared Aerographic 2424
Plus X Panchromatic 2042

We found:

a. Aerochrome infrared images resulted in significantly greater ($\alpha \leq 0.05$) mean percentages of targets detected than Plus X Panchromatic and Infrared aerographic images.

b. The mean percentage of targets detected with aerochrome images was significantly (≤ 0.05) greater than that of the Infrared Aerographic film.

c. The combined mean number of targets detected with image types Aerochrome MS and Aerochrome Infrared was significantly ($\alpha = 0.01$) greater than that of the combined mean number of targets detected from Plus X Panchromatic and Infrared Aerographic imagery.

From the above results it is concluded that the accuracy of day imagery intelligence will increase with usage of Aerochrome MS and Aerochrome Infrared imagery by II's.

REFERENCES

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

2. Carl Strandberg, The Language of Color, the Sensation of Color, Manual of Color Aerial Photography, American Society of Photogrammetry, First Edition.

3. Eastman Kodak Comapny, Kodak Data for Aerial Photography, 34th Edition, 1971.

4. Dixon and Massey, Introduction to Statistical Analysis, 3rd Edition, McGraw Hill, 1969.

5. The Naval Reconnaissance and Technical Support Center, An Investigation Into the Role of Color Photography in Multi-Sensor Reconnaissance, Technical Bulletin NAVAIR 10-35TB-315, January 1970.

NEW ALGORITHMS FOR NONLINEAR LEAST SQUARES AND BAYESIAN PARAMETER ESTIMATION

Warren E. Stewart and Jan P. Sørensen
Chemical Engineering Department
and Mathematics Research Center
University of Wisconsin
Madison, Wisconsin 53706

New algorithms are described for Bayesian estimation of parameters in nonlinear models of multiple-response systems. Modal and interval estimates are provided for the parameter vector $\underset{\sim}{\theta}$ of the predictor model, and for the variance-covariance matrix $\underset{\sim}{\sigma}$ of a Normal error distribution. Allowance is made for gaps (missing values of responses), such as commonly occur in practice. Two chemical examples are analyzed.

INTRODUCTION

Realistic models of multivariate phenomena often relate several predicted responses to a common set of parameters. Multiresponse experiments are required to establish such models, but frequently yield irregular data which are difficult to analyze by classical methods.

Bayes' theorem is a good starting point for parameter estimation in these situations. The multivariate error distribution can be estimated concurrently, whereas it has to be prescribed when least-squares methods are used. Thus, the Bayesian approach allows more objective parameter estimates, if sufficient data are provided. An excellent general account of this approach is given by Box and Tiao (1973).

Bayesian inference deals with a data array $\{y_{ui}\} \equiv \underset{\sim}{y}$, a model for $E(\underset{\sim}{y})$ with parameter vector $\underset{\sim}{\theta}$, and an error distribution model. If a Normal error model is used, with variance-covariance matrix $\underset{\sim}{\sigma}$, the unknown elements of $\underset{\sim}{\sigma}$ will appear as additional parameters. The full set of parameters can be

estimated optimally by maximizing the posterior density $p(\underset{\sim}{\theta},\underset{\sim}{\sigma}|\underset{\sim}{y})$; confidence regions can also be calculated from this function.

In certain cases, the posterior density can be integrated analytically to obtain the marginal density $p(\underset{\sim}{\theta}|\underset{\sim}{y})$. Box and Draper (1965) accomplished this for multivariate Normal error distributions and rectangular data structures (Table 1a). For block-rectangular structures (Table 1b), $p(\underset{\sim}{\theta}|\underset{\sim}{y})$ is the product of the Box-Draper densities for the individual rectangles. More complicated data structures often occur, however, such as that in Table 1c, for which $p(\underset{\sim}{\theta}|\underset{\sim}{y})$ cannot be expressed in closed form. Therefore, in this paper we use the full posterior density $p(\underset{\sim}{\theta},\underset{\sim}{\sigma}|\underset{\sim}{y})$, which has a closed form for any finite data structure.

Inspection of the parameter estimates and residuals often suggests alternatives to the postulated model. Therefore, parameter estimation should not be viewed as an end in itself, but should be followed by critical examination of the model and investigation of any promising alternatives. Interesting predictions or unresolved differences between models will naturally lead to further experiments.

PROBLEM FORMULATION

Consider a set of independent experiments, $u = 1,\ldots,n$, in which a table $\{y_{ui}\}$ of observed responses have been obtained at known settings $\{\underset{\sim}{x}_u\}$ of the independent variables. There are m linearly independent kinds of observations; thus the index i ranges from 1 to m, but in each experiment some values may be missing as in Tables 1b and 1c.

The observations in the uth experiment are regarded as a sample from a population of the form

$$y_{ui} = f_i(\underset{\sim}{x}_u, \underset{\sim}{\theta}) + \varepsilon_{ui} \,. \tag{1}$$

The functions $f_i(\underset{\sim}{x}_u, \underset{\sim}{\theta})$ are models for the expected responses $E(y_{ui}|\underset{\sim}{\theta})$. The residuals ε_{ui} in the uth experiment are treated as a random sample from an m-variate Normal distribution; this gives the probability density (Wilks, 1962)

$$p(\underset{\sim}{\varepsilon}_u|\underset{\sim}{\sigma}) = (2\pi)^{-m_u/2} |\underset{\sim}{\sigma}_u|^{-1/2} \exp(-\tfrac{1}{2}\, \underset{\sim}{\varepsilon}_u^T\, \underset{\sim}{\sigma}_u^{-1}\, \underset{\sim}{\varepsilon}_u). \tag{2}$$

Here $\underset{\sim}{\varepsilon}_u$ is the column vector of error variables $\varepsilon_{u1},\ldots,\varepsilon_{um}$ with dummy zeroes inserted where observations are missing. Correspondingly, $\underset{\sim}{\sigma}_u$ is obtained from the full variance-covariance matrix, $\underset{\sim}{\sigma} = \{\sigma_{ij}\}$, by substituting dummy elements δ_{ij} whenever observation y_{ui} or y_{uj} is missing. Here δ_{ij} is unity when $i=j$, and zero otherwise.

The joint error density model for the set of n experiments follows directly from Equation (2):

$$p(\underset{\sim}{\varepsilon}|\underset{\sim}{\theta},\underset{\sim}{\sigma}) = \prod_{u=1}^{n} (2\pi)^{-m_u/2} |\underset{\sim}{\sigma}_u|^{-1/2} \exp(-\tfrac{1}{2}\, \underset{\sim}{\varepsilon}_u^T\, \underset{\sim}{\sigma}_u^{-1}\, \underset{\sim}{\varepsilon}_u). \tag{3}$$

Insertion of Equation (1) gives the corresponding density in observation space:

$$p(\underset{\sim}{y}|\underset{\sim}{\theta},\underset{\sim}{\sigma}) = \Big[\prod_{u=1}^{n} (2\pi)^{-m_u/2} |\underset{\sim}{\sigma}_u|^{-1/2}\Big] \cdot \exp\{-\tfrac{1}{2} \sum_{u=1}^{n} \sum_{i=1}^{m} \sum_{j=1}^{m} \sigma_u^{ij} [y_{ui} - f_{ui}(\underset{\sim}{\theta})][y_{uj} - f_{uj}(\underset{\sim}{\theta})]\}. \tag{4}$$

Here the functions $f_{ui}(\underset{\sim}{\theta})$ stand for $f_i(\underset{\sim}{x}_u, \underset{\sim}{\theta})$ evaluated at the known settings $\underset{\sim}{x}_u$ of the independent variables. The σ_u^{ij} are the elements of the precision matrices $\underset{\sim}{\sigma}_u^{-1}$. The right-hand term may also be regarded, by Bayes' theorem, as the likelihood function for $\underset{\sim}{\theta}$ and $\underset{\sim}{\sigma}$ when evaluated with given observations $\underset{\sim}{y}$.

The usual factorization of the prior density $p(\underset{\sim}{\theta}, \underset{\sim}{\sigma})$ is assumed,

$$p(\underset{\sim}{\theta}, \underset{\sim}{\sigma}) = p(\underset{\sim}{\theta})\, p(\underset{\sim}{\sigma}) \tag{5}$$

and a locally uniform density $p(\underset{\sim}{\theta})$ is assumed in the region of appreciable likelihood. The latter assumption requires some care in the parametrization of the model. The prior density of $\underset{\sim}{\sigma}$ is taken from Box and Draper (1965):

$$p(\underset{\sim}{\sigma}) \propto |\underset{\sim}{\sigma}|^{-(m+1)/2} . \tag{6}$$

Bayes' theorem then gives the posterior density

$$\begin{aligned} p(\underset{\sim}{\theta}, \underset{\sim}{\sigma} | \underset{\sim}{y}) &= p(\underset{\sim}{\theta}, \underset{\sim}{\sigma})\, p(\underset{\sim}{y} | \underset{\sim}{\theta}, \underset{\sim}{\sigma}) \\ &= c|\underset{\sim}{\sigma}|^{-(m+1)/2} \left[\prod_{u=1}^{n} |\underset{\sim}{\sigma}_u|^{-1/2} \right] \\ &\cdot \exp\left\{ -\tfrac{1}{2} \sum_{u=1}^{n} \sum_{i=1}^{m} \sum_{j=1}^{m} \sigma_u^{ij} [y_{ui} - f_{ui}(\underset{\sim}{\theta})][y_{uj} - f_{uj}(\underset{\sim}{\theta})] \right\} \end{aligned} \tag{7}$$

in which c is a proportionality constant. All that the data reveal about the parameters $\underset{\sim}{\theta}$ and $\underset{\sim}{\sigma}$ is contained in this density function.

Point estimates of $\underset{\sim}{\theta}$ and $\underset{\sim}{\sigma}$ are obtainable by maximizing the posterior density just described, or by minimizing the function

$$\begin{aligned} S(\underset{\sim}{\psi}) \equiv S(\underset{\sim}{\theta}, \underset{\sim}{\sigma}) &= -2 \ln p(\underset{\sim}{\theta}, \underset{\sim}{\sigma} | \underset{\sim}{y}) + 2 \ln c \\ &= (m+1) \ln |\underset{\sim}{\sigma}| + \sum_{u=1}^{n} \ln |\underset{\sim}{\sigma}_u| \\ &+ \sum_{u=1}^{n} \sum_{i=1}^{m} \sum_{j=1}^{m} \sigma_u^{ij} [y_{ui} - f_{ui}(\underset{\sim}{\theta})][y_{uj} - f_{uj}(\underset{\sim}{\theta})] \end{aligned} \tag{8}$$

over the permitted region of $\underset{\sim}{\theta}$ and $\underset{\sim}{\sigma}$. Here $\underset{\sim}{\psi}$ is a column array of the model parameters $\theta_1, \ldots, \theta_p$ and the independent elements of $\underset{\sim}{\sigma}$.

The latter are taken from the lower triangle of $\underset{\sim}{\sigma}$ in row order, i.e. $\psi_{p+k} = \sigma_{ij}$ with $i \geq j$ and $k = j + i(i-1)/2$. Thus, the total number of parameters is $q = p + m(m+1)/2$.

If the matrix $\underset{\sim}{\sigma}$ were believed to be known, i.e., if a sharply focussed prior density $p(\underset{\sim}{\sigma})$ were assumed, then $S(\underset{\sim}{\psi})$ would reduce to $S(\underset{\sim}{\theta})$ and we would have a least-squares estimation problem with just p parameters. In practice, one seldom knows $\underset{\sim}{\sigma}$ accurately; hence, the full Bayesian solution is recommended.

PARAMETER ESTIMATION ALGORITHMS

Several algorithms are described here for obtaining summary information from Equation (8). These algorithms are part of a Fortran IV package available from the authors.

1. Counting Algorithm

Before analyzing S we count Equations (1) to see which parameters can plausibly be estimated from the data. We first try to match each parameter σ_{kj} in $\underset{\sim}{\psi}$ with an observation pair (y_{uk}, y_{uj}) of a replicate experiment (i.e., an experiment which has the same expected response values as a prior experiment in the data set). If this process cannot be completed for a given k, we then try to match each remaining error parameter σ_{kj}, and each model parameter θ_r in the function pairs $[f_{uk}(\underset{\sim}{\theta}), f_{uj}(\underset{\sim}{\theta})]$, with a non-replicate observation pair (y_{uk}, y_{uj}). Finally, any remaining model parameters θ_r are matched with remaining non-replicate observations. If the matching can be completed for all elements of $\underset{\sim}{\psi}$, we proceed with the estimation. Otherwise, the full set of parameters cannot be estimated from the data.

The counting algorithm is a logical Gaussian elimination. This test is a useful diagnostic, but is not infallible, since the actual rank of the estimation equations depends on the numerical values of $\underset{\sim}{x}$, $\underset{\sim}{y}$, and $\underset{\sim}{\psi}$.

2. Minimization Algorithm

A modified Newton method is used to find a minimum of $S(\underset{\sim}{\psi})$. Let $\underset{\sim}{\psi}_0$ be the value of $\underset{\sim}{\psi}$ at the start of an iteration. A correction vector $(\underset{\sim}{\psi}_1 - \underset{\sim}{\psi}_0)$ is computed by minimizing the local quadratic expansion (see Appendix A for derivative expressions)

$$\tilde{S}(\underset{\sim}{\psi}) = S(\underset{\sim}{\psi}_0) + \left.\frac{\partial S}{\partial \underset{\sim}{\psi}}\right|_0 (\underset{\sim}{\psi} - \underset{\sim}{\psi}_0) + \tfrac{1}{2} (\underset{\sim}{\psi} - \underset{\sim}{\psi}_0)^T \left.\frac{\partial^2 S}{\partial \underset{\sim}{\psi} \partial \underset{\sim}{\psi}}\right|_0 (\underset{\sim}{\psi} - \underset{\sim}{\psi}_0) \qquad (9)$$

over a user-specified rectangular region around $\underset{\sim}{\psi}_0$. The region is chosen small enough to ensure that $\tilde{S}(\underset{\sim}{\psi})$ is a good approximation to the function $S(\underset{\sim}{\psi})$ of Equation (9). A search is then made for a minimum of S in the interval of positive definite $\underset{\sim}{\sigma}$ on the line from $\underset{\sim}{\psi}_0$ through $\underset{\sim}{\psi}_1$; this gives the starting point for the next iteration. The calculation continues until two successive line-minima agree within confidence intervals calculated from Equation (14) for each parameter.

3. Response-Independence Test

Box and co-workers (1973) have pointed out the need to test the responses for linear independence. Preferably, one should perform this test on the residuals $[y_{ui} - f_{ui}(\theta)]$, which might become linearly dependent in certain regions of $\underset{\sim}{\theta}$. In the present procedure, such linear dependence is readily detected during the inversion of $\underset{\sim}{\sigma}$ at the start of each iteration. The calculation can continue if all pivot elements (Stewart, 1973) found in this inversion are greater than a specified fraction, say 0.1, of the corresponding elements σ_{ii} .

4. Confidence Regions

Equation (8) gives the simple form

$$p(\underset{\sim}{\psi}|\underset{\sim}{y}) \propto \exp[-\tfrac{1}{2} S(\underset{\sim}{\psi})] \qquad (10)$$

for the posterior density function, or "confidence density". Use of Equation (9) gives the approximation

$$p(\underset{\sim}{\psi}|\underset{\sim}{y}) \propto \exp[-\tfrac{1}{2}(\underset{\sim}{\psi} - \hat{\underset{\sim}{\psi}})^T \hat{\underset{\sim}{A}}(\underset{\sim}{\psi} - \hat{\underset{\sim}{\psi}})], \qquad (11)$$

valid in the neighborhood of the minimum point $\hat{\underset{\sim}{\psi}}$. Here $\hat{\underset{\sim}{A}}$ is the $q \times q$ matrix (positive definite since S is at a minimum) with elements

$$\hat{A}_{km} = \tfrac{1}{2} \left. \frac{\partial^2 S}{\partial \psi_k \partial \psi_m} \right|_{\hat{\underset{\sim}{\psi}}} \qquad (12)$$

computed as described in the Appendix. Thus, near the optimum, the parameters are Normally distributed with variance-covariance matrix $\hat{\underset{\sim}{A}}^{-1}$. If Equation (11) is used as an approximation for all values of $\underset{\sim}{\psi}$, then the confidence intervals for Normal distributions can be applied. For example, the ellipsoidal region

$$(\underset{\sim}{\psi} - \hat{\underset{\sim}{\psi}})^T \hat{\underset{\sim}{A}}(\underset{\sim}{\psi} - \hat{\underset{\sim}{\psi}}) \leq \chi^2(q,\alpha) \tag{13}$$

roughly approximates the $100(1 - \alpha)$ percent highest-posterior-density region, or joint confidence region, for $\underset{\sim}{\psi}$ based on the given data. The intervals

$$(|\psi_k - \hat{\psi}_k| / \sqrt{2\, A^{kk}}) \leq \text{erfc}^{-1}(\alpha) \tag{14}$$

roughly approximate the $100(1 - \alpha)$ percent confidence intervals for the individual parameters. For symmetric 95 percent confidence intervals ($\alpha = 0.05$), $\text{erfc}^{-1}(\alpha)$ has the value 1.96.

Equation (14) is more reliable than (13), since the integration used to obtain it is less affected by the tails of the posterior density function. More accurate intervals can be obtained, but with greater effort, by numerical integration of Equation (7) or (10).

RESULTS FOR RECTANGULAR DATA STRUCTURES

If every experiment gives a full set of observations $y_{u1},\ldots,y_{um}$, then Equation (7) takes the form

$$p(\underset{\sim}{\theta},\underset{\sim}{\sigma}|\underset{\sim}{y}) \propto |\underset{\sim}{\sigma}|^{-(m+n+1)/2} \exp[-\tfrac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \sigma^{ij} v_{ij}(\underset{\sim}{\theta})] \qquad (15)$$

in which

$$v_{ij}(\underset{\sim}{\theta}) = \sum_{u=1}^{n} [y_{ui} - f_{ui}(\underset{\sim}{\theta})][y_{uj} - f_{uj}(\underset{\sim}{\theta})] \; . \qquad (16)$$

Integration of Equation (15) over the region of positive definite $\underset{\sim}{\sigma}$ gives the marginal density function

$$p(\underset{\sim}{\theta}|\underset{\sim}{y}) \propto |\underset{\sim}{v}(\underset{\sim}{\theta})|^{-n/2} \qquad (17)$$

as shown by Box and Draper (1965). We wish to compare the estimates based on this function with those obtained from the full density function of Equation (15).

Setting $p(\underset{\sim}{\theta}|\underset{\sim}{y})$ stationary with respect to its parameters gives

$$\frac{\partial \ln |\underset{\sim}{v}(\underset{\sim}{\theta})|}{\partial \theta_k} \equiv \sum_i \sum_j v^{ij} \frac{\partial v_{ij}}{\partial \theta_k} = 0 \qquad k = 1,\ldots,p \qquad (18)$$

when use is made of the Laplace expansion of $|\underset{\sim}{v} + \underset{\sim}{dv}|$. Here the v^{ij} are the elements of the matrix $\underset{\sim}{v}^{-1}$.

Setting $p(\underset{\sim}{\theta},\underset{\sim}{\sigma}|\underset{\sim}{y})$ stationary with respect to its parameters gives, after use of Equation (15),

$$-2 \frac{\partial \ln p(\underset{\sim}{\theta},\underset{\sim}{\sigma}|\underset{\sim}{y})}{\partial \theta_k} \equiv \sum_i \sum_j \sigma^{ij} \frac{\partial v_{ij}(\underset{\sim}{\theta})}{\partial \theta_k} = 0 \qquad k = 1,\ldots,p \qquad (19)$$

$$-2 \frac{\partial \ln p(\underset{\sim}{\theta},\underset{\sim}{\sigma}|\underset{\sim}{y})}{\partial \sigma^{rs}} \equiv (m+n+1) \frac{\partial}{\partial \sigma^{rs}} (\ln |\underset{\sim}{\sigma}|) + \frac{\partial}{\partial \sigma^{rs}} \sum_i \sum_j \sigma^{ij} v_{ij}(\underset{\sim}{\theta})$$

$$\equiv (2 - \delta_{rs})[-(m+n+1)\,\sigma_{rs} + v_{rs}(\underset{\sim}{\theta})] = 0 \qquad (20)$$

$$r = 1,\ldots,m \qquad s = 1,\ldots,r \; .$$

Equation (20) gives, at the stationary point,

$$\hat{\sigma}_{rs} = \frac{v_{rs}(\hat{\underset{\sim}{\theta}})}{m+n+1} . \tag{21}$$

Hence,

$$\hat{\sigma}^{rs} = (m + n + 1)\, v^{rs}(\hat{\underset{\sim}{\theta}}) . \tag{22}$$

Insertion of Equation (22) into (19) gives Equation (18) at the stationary point of $p(\underset{\sim}{\theta},\underset{\sim}{\sigma}|\underset{\sim}{y})$. Hence, for rectangular data structures, the same values of $\hat{\underset{\sim}{\theta}}$ and $\hat{\underset{\sim}{\sigma}}$ are obtained whether one maximizes $p(\underset{\sim}{\theta},\underset{\sim}{\sigma}|\underset{\sim}{y})$ or $p(\underset{\sim}{\theta}|\underset{\sim}{y})$. Of course, the marginal confidence regions for $\underset{\sim}{\theta}$ can be estimated more directly in the latter case. The normal equations based on $p(\underset{\sim}{\theta}|\underset{\sim}{y})$, given by Stewart and Sørensen (1976), are convenient for this purpose.

The covariance estimates in Equation (21) are maximum-density values, and thus differ from the expectation values $E(\sigma_{rs}|\underset{\sim}{y})$ unless $n-m-p$ is very large. If expectation estimates of the σ_{rs} are desired, one can compute them as the corresponding moments of the normalized posterior density $p(\underset{\sim}{\theta},\underset{\sim}{\sigma}|\underset{\sim}{y})$.

EXAMPLE 1. Kinetics of a Three-Component System

Consider the chemical conversion of initially pure species 1 to species 2 and 3 in a batch isothermal reactor. Simulated data for the system are given in Table 1, reproduced from Box and Draper (1965); here y_{ui} is the yield of species i in experiment u. The system is modelled by the differential equations

$$\frac{df_1}{dt} = -k_1 f_1$$

$$\frac{df_2}{dt} = k_1 f_1 - k_2 f_2$$

$$\frac{df_3}{dt} = k_2 f_2$$

which have the solution

$$f_1 = \exp(-k_1 t)$$

$$f_2 = [\exp(-k_1 t) - \exp(-k_2 t)]k_1/(k_2 - k_1)$$

$$f_3 = 1 - f_1 - f_2$$

under the indicated initial conditions. As noted by Box and Draper, it is natural to regard the parameters $\theta_i = \ln k_i$ as uniformly distributed a priori.

There are three responses y_{ui} per experiment. Only two would be linearly independent if the yields were mass-balanced (i.e., if the yields in each row added up to unity). The data in Table 2 are clearly not mass-balanced, so we use all three columns of responses.

The replicates in Table 2 allow preliminary estimation of the parameters σ_{ij}, by the relation

$$s_{ij} = \frac{1}{2n_R} \sum_{r=1}^{n_R} (y_{ri} - y'_{ri})(y_{rj} - y'_{rj}) .$$

Here y_{ri} and y'_{ri} are the observations of response i in the first and second tests of replicate pair r, and n_R is the number of such pairs. This procedure gives

$$\{s_{ij}\} = \begin{Bmatrix} 0.00102 & -0.00128 & 0.00025 \\ -0.00128 & 0.00351 & 0.00024 \\ 0.00025 & 0.00024 & 0.00101 \end{Bmatrix}$$

as a preliminary expectation estimate of $\underset{\sim}{\sigma}$. This is a well-conditioned matrix, so our choice $m = 3$ was correct.

The parameter vector $\underset{\sim}{\psi}$ for the present example consists of θ_1, θ_2, and the six elements on and below the diagonal of $\underset{\sim}{\sigma}$. To test the convergence of the estimation from a poor initial guess, the calculation was started from the initial value shown in Table 3. Convergence was obtained in eight iterations, to the point estimates and 95 percent confidence intervals given there.

A second calculation was made with the same initial values, but with second-order θ-derivatives neglected. Convergence was obtained to the same point estimates in nine iterations. The confidence intervals differed slightly, as shown in Table 3.

A third calculation was made by minimizing the determinant $|\underset{\sim}{v}(\theta)|$. Box and Draper (1965) did this by a search procedure; we used the modified Newton algorithm of Stewart and Sørensen (1976), but neglected the second-order θ-derivatives of the functions $f_{ui}(\underset{\sim}{\theta})$. Convergence was obtained in seven iterations, to the same point estimates $\hat{\theta}_i$. The point estimates for the σ_{ij}, computed from Equation (21), also agreed exactly with the two preceding solutions. The one-parameter confidence intervals (computed in this case only for θ_1 and θ_2) are wider than before, and are considered more accurate since in this case the σ_{ij} have been integrated out exactly (Box and Draper, 1965).

EXAMPLE 2. Kinetics of a Five-Component System

Fuguitt and Hawkins (1945, 1947) did extensive experiments on the liquid-phase thermal reactions of α-pinene and its decomposition products. The following products, in order of boiling point, were identified.

A.	α-Pinene	$C_{10}H_{16}$
B.	α - and β-Pyronene	$C_{10}H_{16}$
C.	Dipentene	$C_{10}H_{16}$
D.	allo-Ocimene	$C_{10}H_{16}$
E.	Dimer	$C_{20}H_{32}$

The reaction conditions and yields are reported in Table 4.

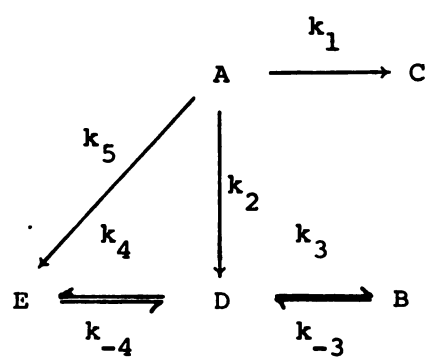
We have normalized the yields to obtain exact mass balances; this makes the yields linearly dependent, and accordingly we have omitted species D. The remaining species are grouped as cumulative distillation fractions:

A, A+B, A+B+C, and E. Each of these responses represents essentially the total mass fraction distilling above or below a particular temperature. The yields of B originally reported in tests 1-15 have been deleted, since they were interpolated values rather than observations (Fuguitt and Hawkins, 1947; Box and co-workers, 1973).

There are numerous gaps in the data. α-Pinene (A) was reported in experiments 1-16, but was considered negligible in the remaining experiments. Pyronenes (B) were reported only in experiments 16-31; they proved difficult to isolate except at small concentrations of α-pinene. Only the dimer fraction (E) was reported in the experiments with allo-ocimene (D) or dimer (E) as feed. The simplified reaction scheme proposed by Fuguitt and Hawkins (1947) implies that α-pinene (A) and dipentene (C) would not be formed in the latter experiments, but that the other three species would be present.

The first eight experiments were used for parameter estimation according to Equation (17) with $m = 3$ by Box and co-workers (1973), and by the present authors (1976). The full 41 experiments could not be so analyzed because of their irregular structure; therefore, only rough estimates were obtainable for several of the reaction parameters. With Equation (8), on the other hand, all 41 experiments can be analyzed.

We postulate the following reaction scheme,



with the following differential equations for the concentrations:

$$\frac{d\phi_A}{dt} = -(k_1 + k_2)\ \phi_A - 2k_5\ \phi_A^2$$

$$\frac{d\phi_B}{dt} = -k_{-3}\ \phi_B + k_3\ \phi_D$$

$$\frac{d\phi_C}{dt} = k_1\ \phi_A$$

$$\frac{d\phi_D}{dt} = k_2\ \phi_A + k_{-3}\ \phi_B - k_3\ \phi_D - 2k_4\ \phi_D^2 + 2k_{-4}\ \phi_E$$

$$\frac{d\phi_E}{dt} = k_5\ \phi_A^2 + k_4\ \phi_D^2 - k_{-4}\ \phi_E$$

Here we have assumed equal densities for the reaction mixture and all species. The ϕ_i are molar concentrations relative to the molar density of pure liquid α-pinene at the reaction temperature. The resulting initial ϕ_i values for the pure reactants are: 1.0 for α-pinene, 1.0 for allo-ocimene, and 0.5 for dimer. The rate coefficients are represented as Arrhenius functions,

$$\ln (k_i) = \theta_i - (1/T - 1/T_B)\ \theta_{i+5} \qquad i = 1,\ldots,5$$

$$\ln (k_3/k_{-3}) = -\theta_{11}/T_B + (1/T - 1/T_B)\ \theta_{13}$$

$$\ln (k_4/k_{-4}) = -\theta_{12}/T_B + (1/T - 1/T_B)\ \theta_{14}$$

with k_i values in min^{-1}, T in Kelvins, and a base temperature T_B of 478.5 K.

The data and parameters were paired to check the feasibility of the estimation. This indicated a sufficient amount of data for estimation of all parameters except σ_{21}. However, the replicate comparisons (u = 18-19,20-21, 22-23,24-25) involving y_{u2} all give duplication of y_{u3}; furthermore each of these comparisons gives a duplication of either y_{u2} or y_{u4}. With these results, we find that neither σ_{32} nor σ_{42} can be estimated; indeed, an attempt to estimate them was terminated by the linear independence test described above. Thereafter, σ_{21}, σ_{32}, and σ_{42} were all fixed at zero, and the remaining parameters were estimated by minimization of S.

Initial values of the θ-parameters were chosen from the results of Fuguitt and Hawkins (1945, 1947), Box and co-workers (1973), and the present authors (1976). Initial variance estimates σ_{ii} were calculated from replicate data available in Table 4, and zeros were inserted initially as covariances.

The model was integrated, for each experiment, by the method of Guertin et al (1977) with 6 mesh points. The coefficients in Equation (9) were computed as described in the Appendix, with first-order sensitivities $\partial\phi_{ui}/\partial\theta_k$ computed by the method of Stewart and Sørensen (1976).

A first minimization, with reaction 5 omitted, converged within 20 iterations. This gave $\hat{S} = 41.06$ with parameter estimates as shown in Table 5. The confidence intervals show the θ's to be estimated quite precisely. The σ_{ij} are estimated less precisely, as anticipated from the limited number of data on several combinations of responses. The deviations of the data from the fitted model are shown in Table 6.

A second minimization of S was done with the full 5-reaction model. This calculation converged to a very flat minimum at $\hat{S} = 34.09$, with parameter estimates as shown in Table 5. The deviations of the data from this fitted model are also shown in Table 6.

The 5-reaction model is better able to describe the polymer yields from α-pinene at short times, as can be seen in Table 6. We can also test the significance of the added parameter θ_5 by use of the confidence intervals. Table 5 gives $\theta_5 = -11.945 \pm 0.698$, based on Equation (14); this implies the limits $(1 \pm 0.698)\exp(-11.945)$ for k_5 with the alternate prior $p(k_5) = c$. Hence, the 95% confidence interval for k_5 does not include zero.

On the other hand, Equations (9) and (13) give the following approximate expression for the 95% confidence region of the 20 fitted parameters of the 5-reaction model:

$$S(\underset{\sim}{\psi}) - 34.09 < \chi^2_{20}(0.05) = 31.41.$$

All $\underset{\sim}{\psi}$ values such that $S(\underset{\sim}{\psi}) < 65.50$ lie within this estimated 95% joint confidence region. By this criterion, the model with $k_5 = 0$ is acceptable. However, as indicated earlier, Equation (14) is more reliable than (13). From this, and a study of the residuals, we conclude that the 5-reaction model is to be preferred.

ACKNOWLEDGMENT

This work was sponsored by the National Science Foundation through Grant ENG76-24368; by the United States Army under Contract DAAG29-75-C-0024; and by the Wisconsin Alumni Research Foundation through funds administrated by the Graduate School of the University of Wisconsin.

BIBLIOGRAPHY

[1] Box, G. E. P. and Draper, N. R. (1965).
The Bayesian estimation of common parameters from several responses.
Biometrika, 52, 355-365.

[2] Box, G. E. P., Hunter, W. G., MacGregor, J. F., and Erjavec, J. (1973).
Some problems associated with the analysis of multiresponse data.
Technometrics, 15, 33-51.

[3] Box, G. E. P. and Tiao, G. C. (1973).
Bayesian Inference in Statistical Analysis. Reading, Massachusetts:
Addison-Wesley, Inc.

[4] Fuguitt, R. E. and Hawkins, J. E. (1945).
The liquid-phase thermal isomerization of α-pinene.
J. A. C. S., 67 , 242-245.

[5] Fuguitt, R. E. and Hawkins, J. E. (1947).
Rate of the thermal isomerization of α-pinene in the liquid phase.
J. A. C. S., 69 , 319-322.

[6] Guertin, E. W., Sørensen, J. P., and Stewart, W. E. (1977).
Exponential collocation of stiff reactor models.
Computers and Chemical Engineering, 1 , 197-203.

[7] Stewart, G. W. (1973)
Introduction to Matrix Computations. New York: Academic Press, Inc.

[8] Stewart, W. E. and Sørensen, J. P. (1976).
Sensitivity and regression of multicomponent reactor models.
Fourth International Symposium on Chemical Reaction Engineering.
Frankfurt: DECHEMA, pp. I-12 to I-20.

[9] Wilks, S. S. (1962).
Mathematical Statistics. New York: John Wiley and Sons, Inc.

APPENDIX: DERIVATIVES OF S .

The matrices $\underset{\sim}{\sigma}_u$ are real and symmetric; furthermore, S is defined only when these matrices are positive definite. The following derivative relations then hold:

$$\frac{\partial \ln |\underset{\sim}{\sigma}_u|}{\partial \sigma_{uij}} = (2 - \delta_{ij})\sigma_u^{ij} \qquad j \leq i \tag{A1}$$

$$\frac{\partial \sigma_u^{ij}}{\partial \sigma_{uk\ell}} = -\tfrac{1}{2}(2 - \delta_{k\ell})[\sigma_u^{ik}\sigma_u^{\ell j} + \sigma_u^{i\ell}\sigma_u^{kj}] \qquad \ell \leq k. \tag{A2}$$

The relations for second derivatives follow by combination of (A1) and (A2):

$$\frac{\partial^2 \ln |\underset{\sim}{\sigma}_u|}{\partial \sigma_{uk\ell}\partial \sigma_{uij}} = -\tfrac{1}{2}(2 - \delta_{ij})(2 - \delta_{k\ell})[\sigma_u^{ik}\sigma_u^{\ell j} + \sigma_u^{i\ell}\sigma_u^{kj}] \qquad j \leq i,\ \ell \leq k \tag{A3}$$

$$\frac{\partial^2 \sigma_u^{ij}}{\partial \sigma_{ust}\partial \sigma_{uk\ell}} = \tfrac{1}{4}(2 - \delta_{k\ell})(2 - \delta_{st})$$
$$[(\sigma_u^{is}\sigma_u^{tk} + \sigma_u^{it}\sigma_u^{sk})\sigma_u^{\ell j} + \sigma_u^{ik}(\sigma_u^{\ell s}\sigma_u^{tj} + \sigma_u^{\ell t}\sigma_u^{sj})$$
$$+ (\sigma_u^{is}\sigma_u^{t\ell} + \sigma_u^{it}\sigma_u^{s\ell})\sigma_u^{kj} + \sigma_u^{i\ell}(\sigma_u^{ks}\sigma_u^{tj} + \sigma_u^{kt}\sigma_u^{sj})]$$
$$\ell \leq k\ ,\quad t \leq s\ . \tag{A4}$$

As indicated earlier, if response h is absent from experiment u , the elements σ_{uhj} and σ_{ujh} are replaced by the constant dummy values δ_{hj} . Note also that the symmetry of $\underset{\sim}{\sigma}_u$ has been used to express these derivatives in terms of elements on and below the diagonal.

The derivatives required for Equation (9) are obtained as follows:

$$-\tfrac{1}{2}\frac{\partial S}{\partial \theta_r} = -\sum_u \sum_i \sum_{j \leq i} \tfrac{1}{2}(2 - \delta_{ij})\sigma_u^{ij}\frac{\partial}{\partial \theta_r}(\varepsilon_{ui}\varepsilon_{uj}) \tag{A5}$$

$$\tfrac{1}{2}\frac{\partial^2 S}{\partial \theta_r \partial \theta_v} = \sum_u \sum_i \sum_{j \leq i} \tfrac{1}{2}(2 - \delta_{ij})\sigma_u^{ij}\frac{\partial^2}{\partial \theta_r \partial \theta_v}(\varepsilon_{ui}\varepsilon_{uj}) \tag{A6}$$

$$\tfrac{1}{2}\frac{\partial^2 S}{\partial\theta_r \partial\sigma_{k\ell}} = \sum_u \sum_i \sum_{j\le i} \tfrac{1}{2}(2-\delta_{ij})\frac{\partial\sigma_u^{ij}}{\partial\sigma_{ukl}}\frac{\partial}{\partial\theta_r}(\varepsilon_{ui}\,\varepsilon_{uj}) \qquad \text{(A7)}$$

$$-\tfrac{1}{2}\frac{\partial S}{\partial\sigma_{st}} = -\tfrac{1}{2}(m+1)\frac{\partial \ln|\underset{\sim}{\sigma}|}{\partial\sigma_{st}} - \sum_{u=1}^{n}\tfrac{1}{2}\frac{\partial \ln|\underset{\sim}{\sigma}_u|}{\partial\sigma_{st}}$$
$$- \sum_u \sum_i \sum_{j\le i} \tfrac{1}{2}(2-\delta_{ij})\frac{\partial\sigma_u^{ij}}{\partial\sigma_{st}}\,\varepsilon_{ui}\,\varepsilon_{uj} \qquad \text{(A8)}$$

$$\tfrac{1}{2}\frac{\partial^2 S}{\partial\sigma_{st}\partial\sigma_{k\ell}} = \tfrac{1}{2}(m+1)\frac{\partial^2 \ln|\underset{\sim}{\sigma}|}{\partial\sigma_{st}\partial\sigma_{k\ell}} + \sum_u \tfrac{1}{2}\frac{\partial^2 \ln|\underset{\sim}{\sigma}_u|}{\partial\sigma_{ust}\partial\sigma_{ukl}}$$
$$+ \sum_u \sum_i \sum_{j\le i} \tfrac{1}{2}(2-\delta_{ij})\frac{\partial^2\sigma_u^{ij}}{\partial\sigma_{ust}\partial\sigma_{ukl}}\,\varepsilon_{ui}\,\varepsilon_{uj} \qquad \text{(A9)}$$

Equations (A6), (A7), and (A9) evaluated at $\underset{\sim}{\theta}_0$ and $\underset{\sim}{\sigma}_0$ provide the coefficient matrix $\underset{\sim}{A}$ of the normal equations. Equations (A5) and (A8) give the right-hand column vector.

The residuals ε_{ui} and ε_{uj} are expressed as functions of $\underset{\sim}{\theta}$ by use of Equation (1). The θ-derivative in Equation (A6) is expanded to give:

$$\frac{\partial^2}{\partial\theta_r\partial\theta_v}(\varepsilon_{ui}\,\varepsilon_{uj}) = \frac{\partial\varepsilon_{ui}}{\partial\theta_r}\frac{\partial\varepsilon_{uj}}{\partial\theta_v} + \frac{\partial\varepsilon_{uj}}{\partial\theta_r}\frac{\partial\varepsilon_{ui}}{\partial\theta_v}$$
$$+ \varepsilon_{ui}\frac{\partial^2\varepsilon_{uj}}{\partial\theta_r\partial\theta_v} + \varepsilon_{uj}\frac{\partial^2\varepsilon_{ui}}{\partial\theta_r\partial\theta_v} \qquad \text{(A10)}$$

The second-derivative terms are unimportant if the data are well fitted; compare Solutions 1 and 2 in Table 3.

If the experiments have different weights w_u as in Table 4, then $\varepsilon_{ui}\,\varepsilon_{uj}$ and its derivatives should be multiplied by w_u throughout the development. As usual, the matrix $\underset{\sim}{\sigma}$ is defined for experiments of unit weight.

Table 1. Examples of Data Structures with m = 4 and n = 8

	1a. Rectangular				1b. Block-rectangular				1c. Irregular			
u	y_{u1}	y_{u2}	y_{u3}	y_{u4}	y_{u1}	y_{u2}	y_{u3}	y_{u4}	y_{u1}	y_{u2}	y_{u3}	y_{u4}
1	+	+	+	+	+	+			+		+	+
2	+	+	+	+	+	+			+	+	+	+
3	+	+	+	+	+	+				+	+	
4	+	+	+	+			+	+	+			
5	+	+	+	+			+	+		+		+
6	+	+	+	+			+	+				+
7	+	+	+	+			+	+	+	+		+
8	+	+	+	+			+	+			+	+

Table 2. Data for Example 1, from Box and Draper (1973)

t_u	y_{u1}	y_{u2}	y_{u3}
0.5	0.959	0.025	0.028
0.5	0.914	0.061	0.000
1.	0.855	0.152	0.068
1.	0.785	0.197	0.096
2.	0.628	0.130	0.090
2.	0.617	0.249	0.118
4.	0.480	0.184	0.374
4.	0.423	0.298	0.358
8.	0.166	0.147	0.651
8.	0.205	0.050	0.684
16.	0.034	0.000	0.899
16.	0.054	0.047	0.991

Table 3. Parameter Values for Example 1

Parameter	Initial Value	Solution 1 Eqs. (8,14)*	Solution 2 Eqs. (8,14)*	Solution 3 Eqs. (18,21)*
θ_1	-2.3026	-1.5723±0.0567	-1.5723±0.0558	-1.5723±0.0800
θ_2	0.	-0.7023±0.1374	-0.7023±0.1346	-0.7023±0.1931
σ_{11}	0.01	$(0.76\pm0.52)\ 10^{-3}$	$(0.76\pm0.53)\ 10^{-3}$	$0.76\ 10^{-3}$
σ_{21}	0.	$-(0.50\pm0.63)\ 10^{-3}$	$-(0.50\pm0.63)\ 10^{-3}$	$-0.50\ 10^{-3}$
σ_{22}	0.01	$(1.86\pm1.28)\ 10^{-3}$	$(1.86\pm1.29)\ 10^{-3}$	$1.86\ 10^{-3}$
σ_{31}	0.	$(0.32\pm0.41)\ 10^{-3}$	$(0.32\pm0.41)\ 10^{-3}$	$0.32\ 10^{-3}$
σ_{32}	0.	$(0.40\pm0.62)\ 10^{-3}$	$(0.40\pm0.62)\ 10^{-3}$	$0.40\ 10^{-3}$
σ_{33}	0.01	$(0.77\pm0.54)\ 10^{-3}$	$(0.77\pm0.54)\ 10^{-3}$	$0.77\ 10^{-3}$

*All intervals are 95% highest posterior density regions. In Solution 3, the intervals are computed from the normal equations with "residual mean square" $|\underset{\sim}{v}(\hat{\theta})|/(n-2)$ and $n-2 = 10$ residual degrees of freedom. In Solution 1, the second-derivative terms of Equation (A10) are included.

Table 4. Data for Example 2, from Fuguitt and Hawkins (1945,1947)

Expt. u	Feed	T, C.	t_u, min	w_u**	Normalized yields, weight percent y_{u1} (A)	y_{u2} (A+B)	y_{u3} (A+B+C)	y_{u4} (E)
1	A	189.5	1230.	1	88.3	***	96.2	2.2
2*	A	189.5	1230.	1	88.2	***	95.7	1.3
3	A	189.5	3060.	2	76.4	***	92.7	2.8
4	A	189.5	4920.	2	64.8	***	88.9	5.8
5	A	189.5	7800.	2	50.3	***	84.7	9.3
6	A	189.5	10680.	2	37.5	***	82.0	12.0
7	A	189.5	15030.	2	25.9	***	77.1	17.0
8	A	189.5	22620.	2	14.0	***	73.9	21.0
9	A	204.5	440.	2	86.6	***	95.3	.6
10	A	204.5	825.	2	75.0	***	91.5	1.6
11	A	204.5	1200.	2	66.0	***	88.8	3.4
12	A	204.5	1500.	2	59.4	***	86.4	5.1
13	A	204.5	2040.	2	48.9	***	83.0	8.3
14	A	204.5	3060.	2	32.8	***	77.8	13.8
15	A	204.5	6060.	2	11.5	***	70.4	22.5
16	A	189.5	36420.	2	4.5	7.4	70.5	25.7
17	A	204.5	16020.	2	-	3.1	66.2	28.6
18	A	225.0	3000.	1	-	3.0	66.0	28.0
19*	A	225.0	3000.	1	-	4.0	66.0	28.0
20	A	245.0	630.	1	-	4.0	65.0	27.0
21*	A	245.0	630.	1	-	5.0	65.0	27.0
22	A	265.0	120.	1	-	7.0	65.0	23.0
23*	A	265.0	120.	1	-	7.0	65.0	24.0
24	A	285.0	30.	1	-	11.0	66.0	19.0
25*	A	285.0	30.	1	-	9.0	66.0	19.0
26	D	189.5	1020.	1	-	-	-	80.0
27	D	189.5	3990.	1	-	-	-	87.3
28*	D	189.5	3990.	1	-	-	-	87.3
29	D	189.5	6780.	1	-	-	-	87.5
30	D	189.5	8220.	1	-	-	-	86.5
31	D	189.5	13260.	1	-	-	-	88.5
32	D	189.5	14760.	1	-	-	-	89.8
33	D	204.5	3480.	1	-	-	-	87.5
34	D	204.5	5700.	1	-	-	-	86.8
35	E	189.5	8880.	1	-	-	-	91.9
36*	E	189.5	8880.	1	-	-	-	92.0
37	E	189.5	14340.	1	-	-	-	89.8
38	E	189.5	23400.	1	-	-	-	89.7
39*	E	189.5	23400.	1	-	-	-	88.5
40	E	204.5	5700.	1	-	-	-	88.4
41	E	204.5	8100.	1	-	-	-	87.9

* Replicate of the preceding test.

** w_u is the number of independent tests combined to obtain each observation y_{ui}.

*** Originally reported but not observed; see text.

- No value reported.

Table 5. Parameters for α-Pinene Conversion

Parameter	Estimates* for 4-Reaction Model	Estimates* for 5-Reaction Model
θ_1	-8.331 ± .024	-8.333 ± .025
θ_2	-8.898 ± .029	-8.961 ± .054
θ_3	-8.242 ± .341	-8.196 ± .325
θ_4	-5.389 ± .081	-5.438 ± .087
θ_5		-11.945 ± .698
θ_6	19814. ± 428.	19785. ± 457.
θ_7	20828. ± 474.	20890. ± 536.
θ_8	17336. ± 4079.	17212. ± 4203.
θ_9	10321. ± 915.	10322. ± 918.
θ_{10}		19957. **
θ_{11}	269. ± 83.	279. ± 83.
θ_{12}	-1976. ± 64.	-1985. ± 63.
θ_{13}	-336. ± 950.	-259. ± 958.
θ_{14}	-3873. ± 1624.	-3781. ± 1555.
σ_{11}	.696 ± .419	.784 ± .492
σ_{21}	.000 **	.000 **
σ_{22}	.391 ± .359	.376 ± .348
σ_{31}	.358 ± .412	.426 ± .456
σ_{32}	.000 **	.000 **
σ_{33}	.706 ± .426	.732 ± .444
σ_{41}	-.248 ± .344	-.294 ± .354
σ_{42}	.000 **	.000 **
σ_{43}	-.504 ± .317	-.493 ± .314
σ_{44}	.744 ± .304	.654 ± .282

*95% highest posterior density intervals calculated from Equation (14).

**Posterior estimates were not obtained for these parameters.

Table 6. Final Residuals $\varepsilon_{ui}(\hat{\theta})$ for Example 2.

	4-Reaction Model				5-Reaction Model			
Expt., u	ε_{u1} (A)	ε_{u2} (A+B)	ε_{u3} (A+B+C)	ε_{u4} (E)	ε_{u1} (A)	ε_{u2} (A+B)	ε_{u3} (A+B+C)	ε_{u4} (E)
1	-1.32	-	-.37	2.00	-1.22	-	-.26	1.69
2	-1.42	-	-.87	1.10	-1.32	-	-.76	.79
3	.26	-	.24	.88	.43	-	.43	.45
4	.28	-	-.15	1.10	.45	-	.06	.72
5	.38	-	-.04	.22	.48	-	.12	-.04
6	-1.13	-	.70	-.81	-1.12	-	.78	-.96
7	-.32	-	-.26	-.17	-.43	-	-.29	-.18
8	.66	-	.89	-1.06	.47	-	.74	-.92
9	.88	-	.21	.30	1.00	-	.35	-.11
10	.10	-	-.17	.14	.24	-	.04	-.38
11	.31	-	-.07	.16	.45	-	.15	-.34
12	.27	-	-.51	.23	.38	-	-.29	-.23
13	-.04	-	-.86	.42	.01	-	-.69	.08
14	-1.44	-	-1.56	.90	-1.52	-	-1.49	.75
15	-.47	-	-1.47	.63	-.70	-	-1.61	.77
16	.60	.78	.98	-.36	.44	.72	.74	-.14
17	-	-.12	-.67	.34	-	-.07	-.87	.50
18	-	-.81	.51	-.48	-	-.76	.38	-.39
19	-	.19	.51	-.48	-	.24	.38	-.39
20	-	-.89	.29	-.56	-	-.88	.22	-.47
21	-	.11	.29	-.56	-	.13	.22	-.47
22	-	-.54	-.31	-.37	-	-.58	-.32	-.28
23	-	-.54	-.31	.63	-	-.58	-.32	.72
24	-	1.54	.49	-.15	-	1.51	.58	-.20
25	-	-.46	.49	-.15	-	-.49	.58	-.20
26	-	-	-	1.12	-	-	-	1.95
27	-	-	-	-.92	-	-	-	-.61
28	-	-	-	-.92	-	-	-	-.61
29	-	-	-	-1.31	-	-	-	-1.16
30	-	-	-	-2.37	-	-	-	-2.27
31	-	-	-	-.41	-	-	-	-.42
32	-	-	-	.90	-	-	-	.86
33	-	-	-	.67	-	-	-	.72
34	-	-	-	-.31	-	-	-	-.40
35	-	-	-	1.26	-	-	-	1.17
36	-	-	-	1.36	-	-	-	1.27
37	-	-	-	.24	-	-	-	.16
38	-	-	-	.91	-	-	-	.80
39	-	-	-	-.29	-	-	-	-.40
40	-	-	-	.42	-	-	-	.27
41	-	-	-	.51	-	-	-	.35

ROSTER OF REGISTRANTS

Dr. Francis Dressel
Army Research Office
AV 955-3331

Robert Launer
Army Research Office
AV 955-3331

LTC James Adams
US Army Operation Test
Evaluation Agency
Falls Church, VA
(703)756-2163

Robert C. McMillan
US Army Mobility Equipment
Research and Development Command
(703)664-5133

Dr. Grady W. Miller
US Army Materiel Systems Analysis Activity
APG, MD
AV 283-4291 or 5882

Virginia Gildengorin
Letterman Army Institute of Research
Presidio, San Francisco, CA 94129
AV 561-5548

Leon E. Klarman
US Army Natick Research &
Development Command
Kansas Street, Natick, MA
AV 955-2111

Ray Bell
AMSAA

Carl B. Bates
USACAA
8120 Woodmont Avenue
Bethesda, Maryland 20014
(202)295-0529

Thomas J. Kitchell
US Army Concepts Analysis Agency
8120 Woodmont Avenue
Bethesda, Maryland 20014
AV 295-1534

Barnard H. Bissinger
Pennsylvania State University
Middletown, PA 17057
(717)783-6081

Dale A. Madden
Food & Drug Administration
Rockville, Maryland 20857
(301)443-3240

William E. Baker
Ballistic Research Lab

Richard T. Maruyama
Ballistic Research Lab

C. Maxson Greenland
Chemical Systems Lab
Aberdeen Proving Ground
Maryland 21010

Jerry Thomas
Ballistic Research Lab

Richard L. Moore
HQ ARRADCOM, Bldg. 151
Dover, NJ 07801
AV 880-5791

James Knaub, Jr.
US Army White Sands Missile Range
New Mexico
AV 258-3496

William S. Agee
White Sands Missile Range
New Mexico
AV 258-2361

Barbara Edelman
US Army Natick Research &
Development Command
Kansas Street, Natick, MA
Food Sciences Laboratory
AV 955-2720

Keith R. Eberhardt
Nat'l Bureau of Standards
Bldg. 101-Room A337
Washington, DC (301)921-3651

TWENTY-FIFTH CONFERENCE
ON THE DESIGN OF EXPERIMENTS

Bernard Harris
Mathematics Research Center
University of Wisconsin
610 Walnut Street
Madison, Wisconsin

Harold Ascher
Naval Research Lab
Washington, DC
AV 297-3745

Harry Feingold
David Taylor Naval &
Ship R&D Center

Al May
Pillsbury Research Labs
Minneapolis, Minnesota (612)330-4999

Dr. J. Richard Moore
USA BRL
Aberdeen Proving Ground
Maryland 21005
AV 283-3605

Dr. Kali Banerjee
USA BRL
Aberdeen Proving Ground
Maryland 21005
AV 283-3605

Ronald L. Johnson
US Army Mobility Research &
Development Command
Camouflage & Topographic Lab
Ft. Belvoir, VA 22060

Edmund Inselmann
USACAC & Ft. Leavenworth
Kansas 66027
AV 552-4465

Michael Hacskaylo
Night Vision and Electro-
Optics Laboratory
Fort Belvoir, VA 22044
(703)664-3063

Larry Hatch
NSA
Fort Meade, MD
AV 796-6461

Jared B. Jobe
US Army Natick R&D Command
US Army Institute of
Environmental Medicine
Kansas Street, Natick, MA
(617)653-1000 X2854

Herman Chernoff
Mass. Institute of Technolog

Alexander Mood
University of California

Donald F. Kass
US Army Natick R&D Command
Kansas Street, Natick, MA
Management Information Syste
AV 955-2111

Donald M. Neal
Army Materials & Mechanics
Research Center
Watertown, MA 02172

William E. Jackson
AFRRI - Defense Nuclear Agen
Bethesda, MD (202)295-1173

Sheldon Levin
AFRRI-Defense Nuclear Agency
Bethesda, MD (202)295-1173

Badrig Kurkjian
University of Alabama

Douglas B. Tang
Walter Reed Army Institute
of Research
Washington, DC
AV 576-2011

TWENTY-FIFTH CONFERENCE
ON THE DESIGN OF EXPERIMENTS

J. Robert Burge
Walter Reed Army Institute
of Research
Washington, DC

Edward W. Ross
US Army Natick R&D Command
Kansas Street, Natick, MA
O/TD
AV955-2384

Gertrude Weintraub
ARRADCOM
AV 880-4561

Malcolm Taylor
Aberdeen Proving Ground
Maryland 21005

Doris Sherman
US Army Natick R&D Command
Kansas Street, Natick, MA
Food Sciences Laboratory
AV 955-2720

James Sweeney
US Army Natick R&D Command
Kansas Street, Natick, MA

Anita Ontiveros
USA TRASANA
White Sands Missile Range
New Mexico AV 258-3823

Beatrice S. Orleans
Naval Sea Systems Command

Dr. Chia P. Wang
Food Engineering Laboratory
US Army Natick R&D Command
Kansas Street, Natick, MA
AV 955-2222

Dr. James B. Sampson
USARIEM
US Army Natick R&D Command
Kansas Street, Natick, MA
AV 955-2854

Andrew Soms
Mathematics Research Ctre.
610 Walnut Street
Madison, WI 53706
(608)263-5550

Marvin Zelen
Sidney Farber Cancer Institute
44 Binney Street
Boston, MA
(617)732-3012

Dr. J. Walter Giffee
Food Sciences Laboratory
US Army Natick R&D Command
Kansas Street, Natick, MA
AV 955-2478

Dr. Herbert A. Hollender
Food Engineering Laboratory
US Army Natick R&D Command
Kansas Street, Natick, MA
AV 955-2778

Dr. Robert A. Prosser
Clothing, Equipment & Materials
Engineering Laboratory
US Army Natick R&D Command
Kansas Street, Natick, MA
AV 955-2967

Ray E. Schafer
Hughes Aircraft Co.
Box 3310
Fullerton, CA 92634
(714)732-4865

Aivars Celmins
Ballistic Research Lab
APG, MD 21005
(301)278-3688

TWENTY-FIFTH CONFERENCE
ON THE DESIGN OF EXPERIMENTS

Thomas Leonard
Mathematics Research Center
University of Wisconsin
610 Walnut Street
Madison, Wisconsin
(608)263-5508

Warren Stewart
Mathematics Research Center
University of Wisconsin
610 Walnut Street
Madison, Wisconsin
(608)233-4613

Glen Higbee
US Army Medical Research
Institute of Infectious Diseases
Frederick, MD
(301)663-2640

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: ARO Report 80-2	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE *(and Subtitle)*: PROCEEDINGS OF THE TWENTY-FIFTH 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: May 1980
	13. NUMBER OF PAGES: 242
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

DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited. The findings in this report are not 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 resulting from the Twenty-Fifth Conference on the Design of Experiments in Army Research, Development and Testing. It contains most of the papers presented at that meeting. These treat various Army statistical and design problems.

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

design of experiments
robust filtering
tracking data
compound frequency distributions
variance reduction
Monte Carlo simulation
sensitivity of tolerance limits
periodic nature of measured data
computer aided hypothesis tests
nonlinear least squares
nonlinear least chi square algorithm
error-time response performances of naive subjects
image interpretation
nonlinear least squares and Bayesian parameter estimation

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

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*