

Applications of Statistics

Paruchuri R. Krishnaiah
Editor

North-Holland

APPLICATIONS OF STATISTICS

PROCEEDINGS OF THE SYMPOSIUM HELD AT
WRIGHT STATE UNIVERSITY, DAYTON, OHIO,
14-18 JUNE 1976

Edited by

PARUCHURI R. KRISHNAIAH

*Department of Mathematics and Statistics
University of Pittsburgh, Pittsburgh, Pa., U.S.A.*



1977

NORTH-HOLLAND PUBLISHING COMPANY—AMSTERDAM • NEW YORK • OXFORD

©NORTH-HOLLAND PUBLISHING COMPANY-1977

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

North-Holland ISBN 0444 85034 1

Published by:
NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM • NEW YORK • OXFORD

Distributors for the U.S.A. and Canada:

Elsevier North-Holland, Inc.
52 Vanderbilt Avenue
New York, N.Y. 10017

Library of Congress Cataloging in Publication Data

Symposium on Applications of Statistics, Wright State
University, 1976.
Applications of statistics.

Includes bibliographical references.

1. Mathematical statistics. I. Krishnaiah,
Paruchuri R. II. Title.
QA276.A1S86 1976 501'.82 77-14481
ISBN 0-444-85034-1

PRINTED IN THE NETHERLANDS

APPLICATIONS OF STATISTICS

PREFACE

Statistical techniques play an important role in drawing meaningful conclusions from non-deterministic data in many disciplines. Therefore, there is a great need to improve the interaction between statisticians and scientists in other disciplines who are interested in applying statistical methodology. A Symposium on Applications of Statistics was organized under the sponsorship of the Air Force Flight Dynamics Laboratory to increase such interaction and to stimulate research in statistical sciences. This symposium was held at Wright State University, Dayton, Ohio, during the period June 14-18, 1976. Professor J. Neyman, who has made many pioneering contributions to the field of statistics, delivered the Inaugural Address entitled "Experimentation with Weather Control and Statistical Problems Generated by It". The program consisted of sessions with invited papers and contributed papers as well as clinical sessions. In the clinical sessions scientists and engineers in the U. S. Air Force presented practical problems that require solution.

This book consists of the invited papers presented at the symposium. In these papers, distinguished scientists discuss the current state-of-the-art on a very broad spectrum of topics in the field of applications of statistics. Some of the areas of applications covered include acoustics, cancer research, cluster analysis, communication, econometrics, hydrology, meteorology, model building, pattern recognition, pharmacokinetics, psychometrics, reduction of dimensionality, reliability, stability of structures, and turbulence. Unfortunately, the contributed papers and the presentations made at the clinical sessions are not included in this book due to the limitation of space, but their titles are listed at the end of the book.

I wish to thank Dr. R. Kahal (Office of the Assistant Secretary, Research and Development, U. S. Air Force) for making the Opening Remarks. Also, I wish to express my appreciation to Dr. D. Zonars (Air Force Flight Dynamics Laboratory) and Dr. J. V. Murray (Wright State University) for welcoming the delegates. Thanks are due to Professors R. E. Bargmann, A. P. Basu, C. B. Bell, R. Elashoff, K. S. Fu, S. S. Gupta, T. Jayachandran, R. H. Jones, M. Kabrisky, G. Kallianpur, F. Kozin, C. Maneri, M. R. Paruchuri, M. L. Puri, A. Ramakrishnan, C. R. Rao, J. R. Van Ryzin, D. G. Shankland, N. Timm, C. P. Tsokos, V. M. R. Tummala and Drs. J. S. Bendat, S. Dubey, I. I. Gringorten, H. Hughes, R. Launer, W. R. Riggan and E. A. Robinson for presiding over different sessions. Thanks are also due to Professors R. E. Bargmann, L. Kanal, C. G. Khatri, S. Kotz, K. V. Mardia, F. Proschan, J. S. Rustagi, S. Saunders, S. Sclove, B. K. Shah and V. R. R. Uppuluri for refereeing the papers. I am grateful to Drs. H. L. Harter and I. N. Shimi for their help in the organization of the symposium and to Professors J. C. Lee and C. Maneri for their help in making the local arrangements. I wish to thank Colonel M. Duggins for his encouragement and Mrs. R. Kish for typing the manuscript patiently.

I am indebted to Professor J. Neyman for delivering the Inaugural Address. Special thanks are due to the contributors to this volume and to the North-Holland Publishing Company for their excellent cooperation. I wish to express my appreciation to the Department of Statistics at Carnegie-Mellon University, the Department of Mathematics and Statistics at University of Pittsburgh and the Air Force Flight Dynamics Laboratory for providing the necessary facilities to edit this volume. Last but not least, I wish to thank my wife, Indira, for her constant encouragement in the organization of the symposium.

P. R. Krishnaiah

CONTENTS

PREFACE

v

EXPERIMENTATION WITH WEATHER CONTROL AND STATISTICAL PROBLEMS GENERATED BY IT*

Jerzy Neyman

I. Introduction	1
II. First Decade, 1951 - 1960	2
1. Randomized Experiments vs. Non-Randomized Cloud Seeding Operations	2
2. Efforts to Learn the New Field of Study	3
3. Biblical "Seven Fat Years" Followed by "Seven Lean Years"	4
4. Convincing Domestic Freezer Experiment	6
5. First Disenchantment: "Report of the Advisory Committee on Weather Control"	7
6. Some Findings in the Theory of Statistics	8
III. Second Period of my Involvement, 1964 to Present	9
7. Arrival of Seven Annual Reports on Grossversuch III	9
8. Special Weather Modification Conference at the Fifth Berkeley Symposium on Mathematical Statistics and Probability	11
9. Our Use of Grossversuch III Data	11
10. Difficulties in Using Old Experiments to Verify New Hypotheses	14
11. Verification of Generality of the Findings at Grossversuch III	17
12. Hypothetical Mechanism of Far Away Losses of Rain at the Arizona Experiment	20
Acknowledgements	22
References	23

ON ENTROPY MAXIMIZATION PRINCIPLE

Hirotsugu Akaike

1. Introduction	27
2. A Critical View of Conventional Procedures	27
3. Entropy Maximization Principle	29
4. Classical Procedures and the Entropy Maximization Principle	30
5. Some Realizations of the Principle	31
6. Conclusion	39
Appendix	40
References	40

STOCHASTIC PROBLEMS IN THE STABILITY OF STRUCTURES

S. T. Ariaratnam and D. S. F. Tam

1. Introduction	43
2. Formulation	43
3. Approximation to Markov Process and Stability Equations	46
4. Application	48
5. Conclusions	53
Acknowledgement	53
References	53

*Inaugural Address

A MINIMAX APPROACH TO COMPONENT ANALYSIS

Rolf E. Bargmann and Fred D. Baker

1. Introduction	55
2. Rank Reduction to Control Maximum Error	57
3. Numerical Procedures	59
4. Representation	62
5. Applications	65
References	69

COHERENCE ANALYSIS OF MULTIPLE INPUT/OUTPUT DATA

J. S. Bendat

1. Models for Arbitrary Inputs	71
2. Multiple and Partial Coherence Functions	71
3. Ordered Conditioned Records	74
4. Models for Conditioned Inputs	74
5. Acoustics and Vibration Applications	74
References	78

MULTIVARIATE RELIABILITY CLASSES

Henry W. Block

1. Introduction	79
2. Multivariate Exponential Distributions and Extensions	80
3. Multivariate Increasing Failure Rate	81
4. Multivariate Increasing Failure Rate Average	86
5. Other Multivariate Classes and Estimation Procedures	87
References	87

CHARACTERIZING A LATENT TRAIT DISTRIBUTION

R. Darrell Bock and William A. Thrash

1. Background	89
2. The Error Function of the Spatial Ability Test	91
3. The Test Score Distribution in the Presence of Major-Gene Effects and Measurement Error	92
4. Maximum Likelihood Estimation of p , μ_1 , μ_2 , and σ	93
5. Example	95
References	96

APPROXIMATIONS TO THE DISTRIBUTIONS OF THE LIKELIHOOD RATIO STATISTICS FOR TESTING THE HYPOTHESES ON COVARIANCE MATRICES AND MEAN VECTORS SIMULTANEOUSLY

T. C. Chang, P. R. Krishnaiah and J. C. Lee

1. Introduction	97
2. Likelihood Ratio Test Statistic for the Homogeneity of Real Multivariate Normal Populations	97
3. Likelihood Ratio Statistic for Testing the Hypotheses that the Mean Vector and Covariance Matrix are Equal to Specified Values	102
4. Likelihood Ratio Test Statistic for the Homogeneity of Complex Multivariate Normal Populations	103
5. Likelihood Ratio Statistic for Testing the Hypothesis that the Covariance Matrix and Mean Vector of Complex Multivariate Normal are Equal to Specified Values	107
References	108

AN ALTERNATING RENEWAL PROCESS WITH AN ABSORBING STATE

C. L. Chiang and J. P. Hsu

1. Introduction	109
2. Identities and Exact Formulas	112
3. Probability Generating Functions	115
4. Properized Probability Distributions	116
5. Asymptotic Behavior	117
6. An Example	120
References	121

STATISTICAL PROBLEMS ARISING IN THE THEORY OF AGGREGATION

John S. Chipman

1. Introduction	123
2. Aggregation as a Means of Improving Estimation	123
3. Limitations of James-Stein Estimation as Applied to Aggregative Models	127
4. Application to a Study of the Relation Between Internal and External Prices in Germany, 1963 - 1973	133
Appendix	136
References	139

METHODOLOGICAL DEVELOPMENTS IN SOME APPLICATIONS OF CLUSTERING

A. Cohen, R. Gnanadesikan, J. R. Kettenring and J. M. Landwehr

1. Introduction	141
2. Descriptions of the Four Examples	142
3. Problems and Methodological Developments at the Analytical Methods Stage	143
4. Problems and Methodological Developments at the Output Stage	149
5. Concluding Remarks	157
References	161

SOURCE CORRELATION EFFECTS ON STRUCTURAL RESPONSE

Stephen H. Crandall and Alexander P. Kulvets

1. Introduction	163
2. One-Dimensional Structures	164
3. Two-Dimensional Structures	173
Acknowledgement	182
References	182

SOME APPLICATIONS OF STOCHASTIC LANGUAGES

K. S. Fu

1. Introduction	183
2. Review of Stochastic Languages	183
3. Application to Communication and Coding	189
4. Application to Syntactic Pattern Recognition	194
5. Application to Error-Correcting Parsing	198
6. Conclusions and Remarks	199
References	199

STATISTICAL PROBLEMS IN THE SIZE EFFECT ON MATERIAL STRENGTH

H. Leon Harter

1. Introduction	201
2. Effect of Length	202
3. Effect of Cross-Section	204
4. Effect of Volume	206
5. Strength of Fiber Bundles	208
6. Strength of Composites	209
7. Size Effect in Fatigue	213
8. Unsolved Problems	215
References	216

BISPECTRA OF ATMOSPHERIC AND WIND TUNNEL TURBULENCE

K. N. Helland, K. S. Lii and M. Rosenblatt

1. Introduction	223
2. Definition of Bispectrum	223
3. Data Acquisition	226
4. Computation and Statistical Analysis	227
5. Results of the Computation	229
6. Concluding Remarks	247
Acknowledgement	248
References	248

URN MODELS: A USEFUL TOOL IN APPLIED STATISTICS

N. L. Johnson and Samuel Kotz

1. Introduction	249
2. Biological Models	250
3. Learning Models	254
4. Sampling Schemes	257
5. Capture-Recapture Problems	258
6. Occupancy Distributions	260
7. Some Other Applications	262
References	263

STRUCTURAL EQUATION MODELS IN THE SOCIAL SCIENCES:
SPECIFICATION, ESTIMATION AND TESTING*Karl G. Jöreskog*

1. Introduction	265
2. General Theory	266
3. Applications	273
4. Conclusion	284
References	286

THE FLABBINESS AND INSTABILITY OF REGRESSION ANALYSIS AND
COMPUTATIONAL METHODS FOR IMPROVEMENT*Harriet Kagiwada, R. Kalaba and K. Mease*

1. Introduction	289
2. Dynamic Programming Solution for Fixed Values of λ and C	292
3. Successive Approximations	294
4. Analytic Continuation	294
5. Numerical Experiments	295
6. Discussion	297
References	298

ON HIERARCHICAL CLASSIFIER THEORY AND INTERACTIVE DESIGN

Laveen N. Kanal

1. Introduction	301
2. Multistage Decision Schemes and Hierarchical Classifiers	302
3. Automated Optimal Tree Design	305
4. MIPACS	306
5. Cloud Pattern Classification with MIPACS	315
References	321

TRANSFORMATIONS YIELDING RELIABILITY MODELS BASED ON
INDEPENDENT RANDOM VARIABLES: A SURVEY

N. Langberg, F. Proschan and A. J. Quinzi

1. Introduction	323
2. Distributions with Exponential and Proportional Hazard Minima	323
3. Additive Families of Distributions	327
4. Preserving System Life Length and Failure Patterns	329
5. Applications to the Theory of Competing Risks, Life Testing, and Censored Data Problems	332
References	336

STATISTICAL PROBLEMS IN MENTAL TESTING

Frederic M. Lord

Text	339
Appendix	347
References	348

SELF-SELECTIVITY PROBLEMS IN ECONOMETRIC MODELS

G. S. Maddala

1. Introduction	351
2. A Simple Model with no Explanatory Variables	352
3. Models with Explanatory Variables	359
4. Conclusion	365
References	366

SOME APPLICATIONS OF COMPLEX DISTRIBUTIONS TO THE TESTING
OF STATISTICAL HYPOTHESIS

Kenneth S. Miller

1. Introduction	367
2. Distributions Associated with the Complex Gaussian	368
3. Some Simple Applications to Hypothesis Testing	372
4. The Multiple Cell Problem	376
References	380

THE WEIGHTED DISTRIBUTIONS: A SURVEY OF THEIR APPLICATIONS

G. P. Patil and C. Radhakrishna Rao

1. Introduction	383
2. Notation and Terminology	383

PART I: Discrete Models and Size-Biased Probability Sampling	385
3. The Analysis of Family Data	385
4. Alcoholism and Family Size	388
5. Aerial Survey and Visibility Bias	389

6. Forest Disease and Line Transect Sampling	390
7. Discrete Waiting Time Paradox	392
PART II. Continuous Models and Size-Biased Probability Sampling	393
8. Renewal Theory and Its Applications	393
9. Waiting Time Paradox	394
10. Cell Cycle Analysis and Pulse Labeling	395
11. Efficacy of Early Screening for Disease	396
12. Forest Products Research	398
13. Particle Size Distributions by Thin Section Methods	398
14. Aerial Survey in Traffic Research	399
15. Rejection Technique in Random Variables Generation	400
PART III. Weighted Distributions as Mass-Size Distributions in Geology and Economics	401
16. Weight-Size Distributions in a Breakage Process	401
17. Moment Distributions in Economics	402
References	403
AN INVESTIGATION OF THE PROPERTIES OF SOME TWO SAMPLE TWO-STAGE DESIGNS	
<i>Dale Preston and Robert M. Elashoff</i>	
1. Introduction	407
2. The Basic Statistical Model	407
3. An Alternative Statistical Model	412
4. The Minimization Algorithm	413
5. Results	416
6. Discussion	427
7. Summary	431
8. Acknowledgements	432
9. References	432
PATTERN COMPARISON PROBLEMS	
<i>J. M. Richardson</i>	
1. Introduction	433
2. Preliminaries	433
3. General Statement of the Problem	435
4. Estimation of the Rate of Transformation of an Image	436
5. More Complex Comparison Problems	441
6. Discussion	445
References	445
DECONVOLUTION OF TIME SERIES AS APPLIED TO SPEECH	
<i>Enders A. Robinson</i>	
1. Introduction	447
2. Acoustic Tube Model	449
3. Deep Source Model	454
4. Surface-Source Model	455
5. Normal Equations	456
6. Deconvolution of Speech	458
7. Conclusion	460
References	460

STATISTICAL ANALYSIS OF COMPARTMENTAL MODELS WITH APPLICATIONS TO PHARMACOKINETICS AND BIOAVAILABILITY

Jagdish S. Rustagi and Umed Singh

1. Introduction	461
2. Compartmental Models	462
3. Difference Equation Approach	466
4. Statistical Considerations	472
5. Illustrations	473
6. Discussion	476
Acknowledgements	477
References	478

THE RELIABILITY OF STRUCTURES SUBJECT TO FATIGUE UNDER INSPECTION AND REPAIR

Sam C. Saunders

1. Introduction	481
2. Residual Strength for a Component	482
3. The Structural Overload Model for a Component	482
4. The Crack Growth Model for a Component	483
5. Crack Detection and Inspection	486
6. Component Failure under Inspection and Repair	489
7. Generalization to Structures with Jointly Distributed Loads	492
References	496

SYSTEM FAILURES AND STOCHASTIC HAZARD RATE

I. N. Shimi

1. Introduction	497
2. The Model	498
3. The Main Result	501
4. Examples	503
References	505

A SURVEY OF SOME SMOOTHING PROBLEMS AND THE METHOD OF GENERALIZED CROSS-VALIDATION FOR SOLVING THEM

Grace Wahba

1. Introduction	507
2. Ridge Regression	509
3. Spline Smoothing	515
4. Density Estimation	517
5. The Approximate Solution of Linear Operator Equations when the Data are Noisy	519
Notes Added in Proof	521
References	522

STATISTICAL PROBLEMS IN WATER RESOURCES

Vujica Yevjevich

1. Character of Stochastic Processes	525
2. Sources of Stochasticity	526
3. Complexities of Short-Interval Runoff Time Series	529
4. Problem of Stationarity of Annual Hydrologic and Climatologic Time Series	534
5. Combination of Statistical and Physical Information	535

6. Discrimination of Changes in Climatologic and Hydrologic Time Processes	535
7. New Methods for Statistical Inference	537
8. Design of Experiments	537
9. Properties of Derived Variables	538
10. Multivariate Analysis	539
References	540
TITLES OF CONTRIBUTED PAPERS	541
TITLES OF PRESENTATIONS AT THE CLINICAL SESSIONS	543

EXPERIMENTATION WITH WEATHER CONTROL
AND STATISTICAL PROBLEMS GENERATED BY IT*

JERZY NEYMAN

Statistical Laboratory, University of California
Berkeley, California

CONTENTS

I. INTRODUCTION

II. FIRST DECADE, 1951 - 1960

1. Randomized Experiments vs. Non-Randomized Cloud Seeding Operations
2. Efforts to Learn the New Field of Study
3. Biblical "Seven Fat Years" Followed by "Seven Lean Years"
4. Convincing Domestic Freezer Experiment
5. First Disenchantment: "Report of the Advisory Committee on Weather Control"
6. Some Findings in the Theory of Statistics

III. SECOND PERIOD OF MY INVOLVEMENT, 1964 TO PRESENT

7. Arrival of Seven Annual Reports on Grossversuch III
8. Special Weather Modification Conference at the Fifth Berkeley Symposium on Mathematical Statistics and Probability
9. Our Use of the Grossversuch III data
10. Difficulties in Using Old Experiments to Verify New Hypotheses
11. Verification of Generality of the Findings at Grossversuch III
12. Hypothetical Mechanisms of the Far Away Losses of Rain at the Arizona Experiment

I. INTRODUCTION

I feel honored by being asked to give this inaugural address at the present Symposium attended by so many distinguished scholars, some of them specialists in mathematical statistics, and others, specialists in the various domains of applications.

I must confess that the pleasure I experience by being so honored is augmented by a particular circumstance of personal character. While thinking of what I might present that would be likely to be interesting to the distinguished audience, I suddenly realized that the present year, 1976, marks the quarter of a century

*Inaugural Address

since the first report on weather modification has been issued by our Statistical Laboratory in Berkeley. In fact, this first report of ours [1] is dated December 15, 1951. The work we did was in response to the request of the Division of Water Resources, Department of Public Works of the State of California. Since that time, we worked on the subject of weather control almost continuously. Thus, to me personally, the present occasion is very much like a celebration of the silver jubilee of continuing efforts in an important domain of study!

During the past quarter of a century a large number of problems came to our attention. Quite a few of them were "substantive" in character, as was the question of the Division of Water Resources in Sacramento. This question was whether the enthusiastic claims of success voiced by the many commercial cloud seeding operators in California should be taken seriously. The specific question was whether reliable evidence exists that cloud seeding represents a realistic resource of water much needed in California, particularly in the southern part. Here, then, we had to deal with a "substantive" problem concerned with rainfall or snowfall.

However, as is almost invariably the case, the substantive problem relating to a natural phenomenon generated a number of theoretical statistical problems which we tried to solve. Then, there were repeated "feed-backs", etc. In this process, certain new statistical methodologies were developed and also certain substantive findings occurred. We believe, or should I say we have the illusion, that these substantive findings are quite important. To my personal regret, the cloud seeding community does not like our results, that is, apart from a few exceptional individuals. Just too bad!

It seems to me that a reasonable plan for the present paper might consist of a necessarily brief sketch of three distinct periods of our research, in a summary of our findings, and in pointing out certain regrettable phenomena occurring in the weather modification literature. Hopefully this plan would provide a reasonably comprehensive picture of the present state of the domain as seen from our Statistical Laboratory in Berkeley. Curiously, some of our substantive findings resulted not from some kind of systematic search, but unexpectedly. Then systematic work was needed to confirm the finding.

II. FIRST DECADE, 1951 - 1960

1. Randomized Experiments vs. Non-Randomized Cloud Seeding Operations.

As I see it, the greatest achievement of R. A. Fisher consists in inventing, in developing and in propagandizing the idea that, when working with variable material, the randomization of experiments is a necessary condition of their reliability. A very interesting historical sketch of this subject was recently published by Cochran [2]. One of the interesting details is that some of the difficulties with

non-randomized trials were noted more than two centuries ago. The individual in question is described by Cochran as Arthur Young, "the great English experimentalist." The year given is 1764! One of the difficulties pointed out is that the experimentalists tend to have personal preferences to the treatments they study. The unavoidable results are first self-deception and later deception of others. As Cochran states, the general remarks of Arthur Young are as much relevant now as they were in 1764.

2. *Efforts to Learn the New Field of Study.* In the early 1950's, our efforts were directed towards learning the field of study which at the time was new to us, and towards convincing the authorities in Sacramento that, because of the lack of randomization of the commercial cloud seeding operations, the successes claimed by the operators should be taken with a degree of skepticism [3, 4].

It will be noticed that the reports quoted were authored or coauthored by several individuals who, at the time, were in early stages of their respective careers, but now hold responsible positions in various institutions. In alphabetical order they are: Arnold Court (now Professor at California State University, Northridge), Terry A. Jeeves (now with the Westinghouse Company), Lucien M. LeCam (perhaps the "most mathematical" leader in the field of mathematical statistics, deeply interested in applied fields such as immunotherapy of cancer), Elizabeth L. Scott (our former Department Chairman and a respected colleague), George Steck (now with Sandia Corporation), B. V. Sukhatme (now Professor at Iowa State University), and several others.

With the kind help of Edward M. Vernon, then Chief Regional Forecaster of the U. S. Weather Bureau in San Francisco, it did not take us too long to learn some of the background, sufficient to initiate certain studies of our own. The point was that, while the idea of the necessity of randomization as a basic prerequisite to reliability of experimentation was clear to me, as well as to my colleagues, the prospects of success in our efforts to convince the authorities depended very much on producing some "tangible" evidence invalidating the evaluations of the commercial cloud seeders.

The methodology of evaluating non-randomized cloud seeding operations used at the time, and also now, is the so-called "historical regression" methodology. Consider two not very distant areas, one being the area in which the operator contracted to increase the precipitation, called "target", and the other called "control". The presumption was, and continues to be, that cloud seeding over the target cannot possibly affect the precipitation over the control.

Using precipitation data available for a few years before the beginning of the cloud seeding era, one establishes the linear regression line of the target precipitation, say \underline{Y} , on the control precipitation, say \underline{X} . (Frequently, this is done with \underline{X} and \underline{Y} not expressed in inches of actual precipitation, but in terms of some